Analyzing repeated measures by employing mixed model/factor analytic hybrid estimators of the covariance matrix

by Charles Steven Todd

A thesis submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Mathematical Sciences
Montana State University

Abstract:
The analysis of repeated measurements is complex because measures taken on the same individual at different times are, in general, correlated. Classic multivariate methods do not require any specific covariance structure, but are inefficient because many covariance parameters must be estimated. In some settings, efficiency can be increased by a priori specification of time-varying covariates or postulation of time series processes with which to model covariances. In many settings, however, neither a priori time-varying covariates nor time series processes are easily specified.

An efficient model will be presented that is based on a mixed model covariance structure. The matrix of time varying covariates is assumed unknown and will be estimated. Two new results will be introduced that simplify the procedure for finding estimators of model parameters. Results of simulations that compare the power of the new method with some existing methods are presented.

ANALYZING REPEATED MEASURES BY EMPLOYING MIXED MODEL/FACTOR ANALYTIC HYBRID ESTIMATORS OF THE COVARIANCE MATRIX

by

Charles Steven Todd

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

of

Doctor of Philosophy

in

Mathematical Sciences

MONTANA STATE UNIVERSITY
Bozeman, Montana

July 1995

APPROVAL

of a thesis submitted by

Charles Steven Todd

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

July 31, 1995 — Date

Robert J. Boik
Chairperson, Graduate Committee

Approved for the Major Department

July 31, 1995 — Date

John Lund
Head, Mathematical Sciences

Approved for the College of Graduate Studies

8/18/95 — Date

Robert Brown
Graduate Dean

STATEMENT OF PERMISSION TO USE

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

Signature Charles S. [signature]

Date 7-31-95

ACKNOWLEDGEMENTS

I would like to thank Robert Boik for his patience and encouragement during the preparation of this dissertation.

I would like to thank Diane, Amanda, Dustin, Sara, Hazel, and Leo for everything.

TABLE OF CONTENTS

TABLE OF CONTENTS — Continued

LIST OF TABLES

ABSTRACT

The analysis of repeated measurements is complex because measures taken on the same individual at different times are, in general, correlated. Classic multivariate methods do not require any specific covariance structure, but are inefficient because many covariance parameters must be estimated. In some settings, efficiency can be increased by a priori specification of time-varying covariates or postulation of time series processes with which to model covariances. In many settings, however, neither a priori time-varying covariates nor time series processes are easily specified.

An efficient model will be presented that is based on a mixed model covariance structure. The matrix of time varying covariates is assumed unknown and will be estimated. Two new results will be introduced that simplify the procedure for finding estimators of model parameters. Results of simulations that compare the power of the new method with some existing methods are presented.

CHAPTER 1

INTRODUCTION

The subject of this thesis is modeling the covariance structure in a repeated measures design. A design in which a response from n independent subjects is observed at each of t occasions is a univariate repeated measures study. These strategies are among the most popular experimental designs in the behavioral and social sciences, as well as in medical studies. The response might be a measurement on blood pressure, cholesterol level, or change due to some social or environmental events. Examples of univariate repeated measures can be found in [6], [25] and [37].

The analysis of repeated measures is typically concerned with estimating and testing location parameters. Although modeling these location parameters is the primary concern, the modeling of the covariances among the repeated measures is also important. The structure of the covariance matrix can influence the accuracy of location estimators as well as the power of any tests. In general, the more covariance parameters that must be estimated, the wider the confidence intervals for location parameters and the lower the power of tests. These concepts are discussed in [8] and [9]. With no a priori information about the covariance structure, all $\frac{1}{2}t(t+1)$ covariance parameters must be estimated. If n is large relative to t, then a priori information about the covariance structure is not important. If n is small relative to t, the covariance parameters can't be simultaneously estimated with much precision, and efficiency is greatly improved by employing a parsimonious covariance model.

The first part of this section describes a repeated measures model. Following this description is a brief synopsis of some common covariance structures and associated analyses.

Chapter 2 presents a novel covariance structure that is a mixed model structure with the modification that the design matrix for the latent variables is unknown. Two theorems are proved in Chapter 3 and the results of these theorems are applied to estimating covariance parameters. The use of the Fisher scoring algorithm and the EM algorithm are also discussed. The results of simulations comparing tests of location parameters, using the structure presented in Chapter 2 and some of the common covariance structures used in practice, are presented in Chapter 4. A discussion of the simulation results and topics for further research are presented in Chapter 5.

A Repeated Measures Model

A linear model for repeated measures of a single response can be written as

$$\mathbf{Y} = \mathbf{X}\mathbf{B} + \mathbf{U}, \tag{1.1}$$

where $\mathbf{Y}$ is an $n \times t$ matrix of responses for n independent subjects; $\mathbf{X}$ is a known $n \times r$ between subjects design matrix; $\mathbf{B}$ is an $r \times t$ matrix of unknown regression coefficients; and $\mathbf{U}$ is an $n \times t$ matrix of random errors. The rows of $\mathbf{U}$, say $\mathbf{u}_1', \ldots, \mathbf{u}_n'$ are assumed to be distributed, independently, as

$$\mathbf{u}_i \sim \mathrm{N}(\mathbf{0},\ \mathbf{\Sigma}_i). \tag{1.2}$$

This implies that

$$\mathrm{vec}(\mathbf{U}) \sim \mathrm{N}\left[\mathbf{0},\ (\mathbf{\Sigma} \otimes \mathbf{I}_n)\right], \tag{1.3}$$

when $\mathbf{\Sigma}_i = \mathbf{\Sigma}$ for all i. Assuming the distribution given in (1.3), the matrix $\mathbf{Y}$ is distributed as

$$\mathrm{vec}(\mathbf{Y}) \sim \mathrm{N}\left[(\mathbf{I}_t \otimes \mathbf{X})\,\mathrm{vec}(\mathbf{B}),\ (\mathbf{\Sigma} \otimes \mathbf{I}_n)\right]. \tag{1.4}$$

Without loss of generality, the design matrix $\mathbf{X}$ is assumed to have full column rank. If $\mathbf{X}$ does not have full column rank, then $\mathbf{XB}$ can be reparameterized as $\mathbf{XB} = \mathbf{X}_*\mathbf{B}_*$ where $\mathbf{X}_*$ does have full column rank.

Interest is in making inferences about the regression coefficients $\mathbf{B}$. In particular, interest will be in main and interaction functions of the form

$$\mathbf{\Phi} = \mathbf{F}'\mathbf{BC}, \tag{1.5}$$

where $\mathbf{F}$ is an $r \times s$ rank s matrix of coefficients for between group contrasts, and $\mathbf{C}$ is a $t \times q$ rank q matrix of coefficients for contrasts among the repeated measures. Without loss of generality, $\mathbf{C}$ is assumed to be orthonormal.

Multivariate Analyses

The classic multivariate model for univariate repeated measures is reviewed in [8] and [9]. In both of these articles, a generalization of a univariate repeated measures study to a multivariate repeated measures study is done. In a multivariate repeated measures study, a p-dimensional response from each subject is observed at each of t occasions. When p equals 1, the result is a univariate repeated measures study.

The classic multivariate model is given by the equation in (1.1), assuming the distribution given in (1.3), where Σ is a positive definite matrix subject to no further restrictions. In this model all $\frac{1}{2}t(t+1)$ covariance parameters must be estimated. Assuming the distribution in (1.3), the maximum likelihood estimator (MLE) of $\mathbf{\Phi}$ in (1.5) is

$$\widehat{\mathbf{\Phi}} = \mathbf{F}'\widehat{\mathbf{B}}\mathbf{C}, \quad \text{where} \quad \widehat{\mathbf{B}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} \tag{1.6}$$

is the MLE of $\mathbf{B}$. By the Gauss-Markov theorem, $\widehat{\mathbf{\Phi}}$ is also the best linear unbiased estimator (BLUE) of $\mathbf{\Phi}$. Given the distribution in (1.3), the distribution of $\widehat{\mathbf{B}}$ is

$$\mathrm{vec}(\widehat{\mathbf{B}}) \sim \mathrm{N}\left[\mathrm{vec}(\mathbf{B}),\ \mathbf{\Sigma} \otimes (\mathbf{X}'\mathbf{X})^{-1}\right]. \tag{1.7}$$

Equations (1.6) and (1.7) imply that the distribution of $\widehat{\mathbf{\Phi}}$ is

$$\mathrm{vec}(\widehat{\mathbf{\Phi}}) \sim \mathrm{N}\left[\mathrm{vec}(\mathbf{\Phi}),\ \mathbf{C}'\mathbf{\Sigma}\mathbf{C} \otimes \mathbf{F}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{F}\right]. \tag{1.8}$$

Inferences concerning $\mathbf{\Phi}$ in (1.5) can be made using $\widehat{\mathbf{\Phi}}$ in (1.6) with the distribution in (1.3). Suppose, for example, interest is in testing

$$\mathrm{H}_0\colon \mathbf{\Phi} = \mathbf{\Phi}_0 \quad \text{against} \quad \mathrm{H}_a\colon \mathbf{\Phi} \neq \mathbf{\Phi}_0. \tag{1.9}$$

Let $\mathbf{B}^* = \mathbf{BC}$. Then the hypotheses of (1.9) are equivalent to

$$\mathrm{H}_0\colon \mathbf{F}'\mathbf{B}^* = \mathbf{\Phi}_0 \quad \text{against} \quad \mathrm{H}_a\colon \mathbf{F}'\mathbf{B}^* \neq \mathbf{\Phi}_0. \tag{1.10}$$

Given the distribution in (1.3), the MLE of $\mathbf{F}'\mathbf{B}^*$ subject to the restriction under the null in (1.10) is $\mathbf{F}'\widehat{\mathbf{B}}_0^*$, where

$$\widehat{\mathbf{B}}_0^* = \widehat{\mathbf{B}}\mathbf{C} - (\mathbf{X}'\mathbf{X})^{-1}\mathbf{F}[\mathbf{F}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{F}]^{-1}(\widehat{\mathbf{\Phi}} - \mathbf{\Phi}_0) \tag{1.11}$$

for $\widehat{\mathbf{B}}$ and $\widehat{\mathbf{\Phi}}$ the unrestricted MLEs in (1.6). Note that $\mathbf{F}'\widehat{\mathbf{B}}_0^* = \mathbf{\Phi}_0$. When $\mathbf{C}'\mathbf{\Sigma}\mathbf{C}$ is known, the likelihood ratio (LR) statistic for testing the hypotheses in (1.10) is

$$L_1 = \frac{\dfrac{\exp\left((-1/2)\,\mathrm{tr}\left[(\mathbf{YC}-\mathbf{X}\widehat{\mathbf{B}}_0^*)'(\mathbf{YC}-\mathbf{X}\widehat{\mathbf{B}}_0^*)(\mathbf{C}'\mathbf{\Sigma}\mathbf{C})^{-1}\right]\right)}{|\mathbf{C}'\mathbf{\Sigma}\mathbf{C}|^{(n/2)}(2\pi)^{(nq/2)}}}{\dfrac{\exp\left((-1/2)\,\mathrm{tr}\left[(\mathbf{YC}-\mathbf{X}\widehat{\mathbf{B}}\mathbf{C})'(\mathbf{YC}-\mathbf{X}\widehat{\mathbf{B}}\mathbf{C})(\mathbf{C}'\mathbf{\Sigma}\mathbf{C})^{-1}\right]\right)}{|\mathbf{C}'\mathbf{\Sigma}\mathbf{C}|^{(n/2)}(2\pi)^{(nq/2)}}}. \tag{1.12}$$

Under the monotonic transformation $G_1 = (-2)\ln(L_1)$, (1.12) simplifies to

$$G_1 = \mathrm{vec}'(\widehat{\mathbf{\Phi}} - \mathbf{\Phi}_0)\left((\mathbf{C}'\mathbf{\Sigma}\mathbf{C})^{-1} \otimes [\mathbf{F}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{F}]^{-1}\right)\mathrm{vec}(\widehat{\mathbf{\Phi}} - \mathbf{\Phi}_0). \tag{1.13}$$

This test statistic is distributed as $G_1 \sim \chi^2(qs, \lambda)$, where

$$\lambda = \mathrm{vec}'(\mathbf{\Phi} - \mathbf{\Phi}_0)\left((\mathbf{C}'\mathbf{\Sigma}\mathbf{C})^{-1} \otimes [\mathbf{F}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{F}]^{-1}\right)\mathrm{vec}(\mathbf{\Phi} - \mathbf{\Phi}_0). \tag{1.14}$$

When $\mathbf{C}'\Sigma\mathbf{C}$ is unknown, one of the standard multivariate statistics, which are listed in [2] and [36] along with their approximate distributions, can be used to test the hypotheses in (1.9). The MLE of $\mathbf{C}'\Sigma\mathbf{C}$ under the distribution in (1.3) is given by $\mathbf{C}'\widehat{\Sigma}\mathbf{C} = \mathbf{E}/n$, where

$$\widehat{\Sigma} = \frac{\mathbf{Y}'[\mathbf{I}_n - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}']\mathbf{Y}}{n} \tag{1.15}$$

is the unrestricted MLE of Σ and

$$\mathbf{E} = (\mathbf{YC} - \mathbf{X}\widehat{\mathbf{B}}\mathbf{C})'(\mathbf{YC} - \mathbf{X}\widehat{\mathbf{B}}\mathbf{C}) = \mathbf{C}'\mathbf{Y}'[\mathbf{I}_n - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}']\mathbf{YC}. \tag{1.16}$$

Under the restriction of H_0 in (1.9), the MLE of $\mathbf{C}'\Sigma\mathbf{C}$ is given by $(\mathbf{E}+\mathbf{H})/n$, where

$$\mathbf{H} = (\widehat{\mathbf{\Phi}} - \mathbf{\Phi}_0)'[\mathbf{F}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{F}]^{-1}(\widehat{\mathbf{\Phi}} - \mathbf{\Phi}_0). \tag{1.17}$$

The standard multivariate statistics are functions of the eigenvalues of $(\mathbf{E}+\mathbf{H})^{-1}\mathbf{H}$. Assuming the distribution in (1.3), $\mathbf{E}$ and $\mathbf{H}$ are independently distributed with Wishart distributions:

$$\mathbf{E} \sim \mathrm{W}_q(m,\ \mathbf{C}'\Sigma\mathbf{C},\ \mathbf{0}) \quad \text{and} \quad \mathbf{H} \sim \mathrm{W}_q(s,\ \mathbf{C}'\Sigma\mathbf{C},\ (\mathbf{C}'\Sigma\mathbf{C})^{-1}\mathbf{\Lambda}), \tag{1.18}$$

where $m = n - r$ and Λ is the non-centrality parameter

$$\mathbf{\Lambda} = (\mathbf{\Phi} - \mathbf{\Phi}_0)'[\mathbf{F}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{F}]^{-1}(\mathbf{\Phi} - \mathbf{\Phi}_0). \tag{1.19}$$

For example, in the LR statistic in (1.12), replacing $\mathbf{C}'\Sigma\mathbf{C}$ with $(\mathbf{E}+\mathbf{H})/n$ in the numerator and $\mathbf{E}/n$ in the denominator gives the Wilks's Lambda statistic. The Lawley-Hotelling trace statistic is given by $T^2 = (n-r)\,\mathrm{tr}(\mathbf{HE}^{-1})$. Using the well known result $\mathrm{tr}(\mathbf{JK}) = \mathrm{vec}'(\mathbf{J}')\,\mathrm{vec}(\mathbf{K})$ for arbitrary conformable matrices $\mathbf{J}$ and $\mathbf{K}$, T^2 can be written as

$$T^2 = \mathrm{vec}'(\widehat{\mathbf{\Phi}} - \mathbf{\Phi}_0)\left((\mathbf{E}/m)^{-1} \otimes [\mathbf{F}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{F}]^{-1}\right)\mathrm{vec}(\widehat{\mathbf{\Phi}} - \mathbf{\Phi}_0). \tag{1.20}$$

Time Series Models

The analysis of repeated measures by employing time series models begins by modeling the covariance structure among the repeated measures in terms of a time series process. Once an estimated covariance matrix has been found, then either traditional multivariate methods or adjusted univariate methods can be used to make inferences concerning $\mathbf{\Phi}$ in (1.5).

Assume the t occasions are equally spaced. Then one example of a time series model is to describe the elements of the error vector for the i^{th} subject, $\mathbf{u}_i = \{u_{ij}\}, j = 1, \ldots, t$, as a p^{th} order autoregressive [AR(p)] process with p fixed and known in advance. This type of model is discussed in [38]. As a generalization of this approach, let the elements of $\mathbf{u}_i$ follow the autoregressive moving average [ARMA(p,q)] process with p and q known in advance. That is,

$$u_i^{(t)} - \gamma_1 u_i^{(t-1)} - \cdots - \gamma_p u_i^{(t-p)} = e_i^{(t)} - \eta_1 e_i^{(t-1)} - \cdots - \eta_q e_i^{(t-q)} , \qquad (1.21)$$

where $p + q \leq t - 1$ and $e_i^{(t)} \sim \text{iid}\,\text{N}(0,\ \sigma^2)$. The ARMA process is assumed to be stationary and invertible, which, along with model (1.21), implies that all elements along any diagonal of Σ assume the same value. Thus, the covariance matrix Σ is a Toeplitz matrix. This type of modeling is discussed in [25] and [33]. A banded, or general autoregressive process was used in [21] to model Σ.

Mixed Models

Let $\mathbf{y}_i'$ be the i^{th} row of $\mathbf{Y}$ in (1.1). That is,

$$\mathbf{Y} = \begin{pmatrix} \mathbf{y}_1' \\ \vdots \\ \mathbf{y}_n' \end{pmatrix}.$$

Then from (1.1), the model for $\mathbf{y}_i$ is

$$\mathbf{y}_i = (\mathbf{I}_t \otimes \mathbf{x}_i')\,\text{vec}(\mathbf{B}) + \mathbf{u}_i, \qquad (1.22)$$

where $\mathbf{x}_i'$ and $\mathbf{u}_i'$ are the i^{th} rows of $\mathbf{X}$ and $\mathbf{U}$, respectively. A mixed model can now be constructed, as was done in [9] and [22], by modeling $\mathbf{u}_i$ as

$$\mathbf{u}_i = \mathbf{Z}_i \mathbf{w}_i + \mathbf{e}_i \quad \text{for } i = 1, \ldots, n, \tag{1.23}$$

where $\mathbf{Z}_i$ is a known $t \times f$ within subject design matrix for the random effects; $\text{rank}\,(\mathbf{Z}_i) = f$; $\mathbf{w}_i$ and $\mathbf{e}_i$ are independently distributed as $\mathbf{w}_i \sim \text{iid}\,\text{N}_f\,(\mathbf{0},\ \boldsymbol{\Delta}_f)$, where $\boldsymbol{\Delta}_f$ is a positive definite matrix, and $\mathbf{e}_i \sim \text{iid}\,\text{N}_t\,(\mathbf{0},\ \sigma^2\mathbf{I}_t)$. Thus, under model (1.23),

$$\mathbf{y}_i \sim \text{N}\left[(\mathbf{I}_t \otimes \mathbf{x}_i')\,\text{vec}(\mathbf{B}),\ \boldsymbol{\Sigma}_i\right], \quad \text{where} \quad \boldsymbol{\Sigma}_i = \mathbf{Z}_i \boldsymbol{\Delta}_f \mathbf{Z}_i' + \sigma^2 \mathbf{I}_t \tag{1.24}$$

for $i = 1, \ldots, n$. The number of covariance parameters to estimate is reduced from $\frac{1}{2}t(t+1)$ to $\frac{1}{2}f(f+1)+1$. The maximum likelihood estimator (MLE) of $\boldsymbol{\Phi}$ in (1.5), assuming model (1.23), is given by

$$\text{vec}(\widehat{\boldsymbol{\Phi}}) = [\mathbf{C} \otimes \mathbf{F}]' \left[\sum_{i=1}^{n} (\widehat{\boldsymbol{\Sigma}}_i^{-1} \otimes \mathbf{x}_i \mathbf{x}_i')\right]^{-1} \sum_{i=1}^{n} (\widehat{\boldsymbol{\Sigma}}_i^{-1} \otimes \mathbf{x}_i)\mathbf{y}_i, \tag{1.25}$$

where $\widehat{\boldsymbol{\Sigma}}_i = \mathbf{Z}_i \hat{\boldsymbol{\Delta}}_f \mathbf{Z}_i' + \hat{\sigma}^2 \mathbf{I}_t$ is the MLE of $\boldsymbol{\Sigma}_i$, and $\hat{\boldsymbol{\Delta}}_f$ and $\hat{\sigma}^2$ are the respective MLEs of $\boldsymbol{\Delta}_c$ and σ^2. As described in [23], if $\mathbf{w}_i$ and $\mathbf{e}_i$ could be observed, then $\hat{\boldsymbol{\Delta}}_f$ and $\hat{\sigma}^2$ would be given by

$$\hat{\boldsymbol{\Delta}}_f = n^{-1} \sum_{i=1}^{n} \mathbf{w}_i \mathbf{w}_i' \tag{1.26}$$

and

$$\hat{\sigma}^2 = (nt)^{-1} \sum_{i=1}^{n} \mathbf{e}_i' \mathbf{e}_i. \tag{1.27}$$

An assumption of the simple, closed-form MLE's in equations (1.26) and (1.27) is that $\mathbf{w}_i$ and $\mathbf{e}_i$ are observable. When modeling the covariance structure under the mixed model in (1.23), this assumption is not met; consequently, an alternative method for finding the MLE's must be used. Two popular iterative algorithms for computing MLE's are the EM and Newton-Raphson algorithms. The application of the EM algorithm in a mixed model setting is discussed in [22], while [26] compares both the EM and Newton-Raphson algorithms in a mixed model setting.

Note that $\mathbf{Y}$ in (1.1), under the mixed model structure in (1.23), can be written as

$$\begin{aligned} \operatorname{vec}(\mathbf{Y}) &= \operatorname{vec}(\mathbf{XB}) + \operatorname{vec}\begin{pmatrix} \mathbf{w}_1'\mathbf{Z}_1' \\ \vdots \\ \mathbf{w}_n'\mathbf{Z}_n' \end{pmatrix} + \varepsilon \\ &= (\mathbf{I}_t \otimes \mathbf{X}) \operatorname{vec}(\mathbf{B}) + \mathbf{I}_{(n,t)} \left(\oplus_{i=1}^{n} \mathbf{Z}_i\right) \mathbf{w} + \varepsilon, \end{aligned} \tag{1.28}$$

where

$$\mathbf{w} = \begin{pmatrix} \mathbf{w}_1 \\ \vdots \\ \mathbf{w}_n \end{pmatrix} \tag{1.29}$$

and $\mathbf{I}_{(n,t)}$ is a vec permutation matrix. Based on asymptotic properties of MLEs for parameters in models of the type in (1.28), which are given in [30], the MLEs $\hat{\mathbf{\Delta}}_f$ and $\hat{\sigma}^2$ are consistent estimators provided the design matrices $(\mathbf{I}_t \otimes \mathbf{X})$ and $\oplus_{i=1}^{n}\mathbf{Z}_i$ in (1.28) satisfy certain conditions. See [17] and [30] for examples. An immediate consequence is that $\widehat{\Sigma}_i$ is also consistent, provided that the mixed model covariance structure is correct. Using consistency, the distribution of $\lim_{n\to\infty} n^{\frac{1}{2}} \operatorname{vec}(\widehat{\mathbf{\Phi}} - \mathbf{\Phi})$ will have mean $\mathbf{0}$ and variance $\Sigma_{\widehat{\mathbf{\Phi}}}$, where

$$\Sigma_{\widehat{\mathbf{\Phi}}} = \lim_{n\to\infty} n\,[\mathbf{C}\otimes\mathbf{F}]' \left[\sum_{i=1}^{n}(\Sigma_i^{-1}\otimes \mathbf{x}_i\mathbf{x}_i')\right]^{-1} [\mathbf{C}\otimes\mathbf{F}]\,;$$

thus,

$$\lim_{n\to\infty} n^{\frac{1}{2}} \operatorname{vec}(\widehat{\mathbf{\Phi}} - \mathbf{\Phi}) \sim \mathrm{N}(\mathbf{0},\ \Sigma_{\widehat{\mathbf{\Phi}}}). \tag{1.30}$$

For large sample sizes, inferences concerning $\mathbf{\Phi}$ in (1.5) can be based on the result in (1.30), with $\widehat{\Sigma}_i$ replacing Σ_i.

In [12] the i^{th} row of $\mathbf{U}$ is modeled as $\mathbf{u}_i = \mathbf{Z}_i\mathbf{w}_i + \mathbf{e}_i$, where $\mathbf{Z}_i$, $\mathbf{w}_i$ and $\mathbf{u}_i$ are defined as they were in the mixed model in (1.23), and the $\mathbf{e}_i$ are assumed to follow an AR(1) process. A scoring method is then used to compute maximum likelihood estimates.

Univariate Analyses

A univariate analysis can be easily understood as a two-way, subjects by occasions, mixed model design. This method makes subjects a random effects blocking factor. This approach is taken in [3], [7], and [31]. The model, as found in [34], is obtained by equating $\mathbf{Z}_i$ in the mixed model in (1.23) to $\mathbf{1}_t$, so that $f = 1$ and

$$\Sigma_i = \Sigma = \mathbf{1}_t \sigma_w^2 \mathbf{1}_t' + \sigma^2 \mathbf{I}_t. \tag{1.31}$$

When this covariance structure is correct, the equation $\mathbf{C}'\Sigma\mathbf{C} = \sigma^2\mathbf{I}_t$ is satisfied and $\mathbf{C}'\Sigma\mathbf{C}$ has the property known as sphericity or circularity. To test $\mathrm{H}_0\colon \mathbf{\Phi} = \mathbf{\Phi}_0$ in (1.9), the standard univariate F statistic can be used. In particular, $\mathrm{tr}(\mathbf{E})/\sigma^2 \sim \chi^2(qm)$ where $\mathbf{E}$ is given in (1.16) and the distribution in (1.3) is assumed. For $\mathbf{H}$ in (1.17), $\mathrm{tr}(\mathbf{H})/\sigma^2 \sim \chi^2(qs,\ \mathrm{tr}(\mathbf{\Lambda})/\sigma^2)$ assuming the distribution in (1.3), where $\mathbf{\Lambda}$ is given in (1.19). The LR test of $\mathrm{H}_0\colon \mathbf{\Phi} = \mathbf{\Phi}_0$ is to reject H_0 for large values of the F statistic, which simplifies to

$$F = \frac{m\,\mathrm{tr}(\mathbf{H})}{s\,\mathrm{tr}(\mathbf{E})}. \tag{1.32}$$

The distribution of F is $F[qs, qm, \mathrm{tr}(\mathbf{\Lambda})/\sigma^2]$.

When sphericity is not satisfied, and assuming a true H_0, approximations to the numerator and denominator degrees of freedom of F in (1.32) were derived in [11]. The ratio of the F statistic in (1.32) is approximately distributed as an F distribution with degrees of freedom $v_1 = \epsilon qs$ and $v_2 = \epsilon qm$ where

$$\epsilon = \frac{\left(\sum_{i=1}^{q} \theta_i\right)^2}{q \sum_{i=1}^{q} \theta_i^2}, \tag{1.33}$$

and θ_i are the eigenvalues of $\mathbf{C}'\boldsymbol{\Sigma}\mathbf{C}$. Under sphericity, $\epsilon = 1$ and the approximation is exact. The parameter ϵ is contained in the closed interval $[q^{-1},\ 1]$, and is estimated by $\hat{\epsilon}$; where $\hat{\epsilon}$ is calculated using the eigenvalues of $\mathbf{C}'\widehat{\boldsymbol{\Sigma}}\mathbf{C}$. An alternative estimator of ϵ, $(\tilde{\epsilon})$, is given in [19]. For near spherical error matrices, results in [18] have shown that $\tilde{\epsilon}$ controls test size better than $\hat{\epsilon}$.

Factor Analytic Structure

The factor analytic covariance structure is briefly discussed in [21], with more detailed discussions given in [4] and [20]. For a factor analytic structure, the error vector for the i^{th} subject, $\mathbf{u}_i$ of (1.22) is modeled as

$$\mathbf{u}_i = \boldsymbol{\Gamma}\mathbf{w}_i + \mathbf{e}_i \quad \text{for } i = 1, \ldots, n, \tag{1.34}$$

where $\boldsymbol{\Gamma}$ is an unknown $t \times f$ factor-loading matrix with rank f; $\mathbf{w}_i$ and $\mathbf{e}_i$ are independently distributed as $\mathbf{w}_i \sim \mathrm{N}_f(\mathbf{0},\ \mathbf{I_f})$ and $\mathbf{e}_i \sim \mathrm{N}_t(\mathbf{0},\ \boldsymbol{\Psi})$; and $\boldsymbol{\Psi}$ a diagonal matrix. The corresponding covariance structure is

$$\operatorname{var}(\mathbf{y}_i) = \boldsymbol{\Sigma} = \boldsymbol{\Gamma}\boldsymbol{\Gamma}' + \boldsymbol{\Psi}. \tag{1.35}$$

Note the resemblance between the factor analytic model in (1.34) and the mixed model in (1.23). The main differences between the two models are for the mixed model, $\mathbf{Z}_i$ is an a priori design matrix, while in the factor analytic model, $\boldsymbol{\Gamma}$ is an unknown factor-loading matrix that must be estimated from the data. Also, in the mixed model, the error vectors $\mathbf{e}_i$ have variance proportional to an identity, represented by $\sigma^2\mathbf{I}_t$, whereas in the factor analytic structure, the $\mathbf{e}_i$ have a diagonal variance matrix $\boldsymbol{\Psi}$; there is no constraint that $\boldsymbol{\Psi}$ be proportional to an identity. The constraint that $\boldsymbol{\Psi}$ be diagonal ensures that, conditional on $\mathbf{w}_i$, the observations for the i^{th} subject, $y_{i1}, y_{i2}, \ldots, y_{it}$ are jointly independent. Thus, $\mathbf{w}_i$ can be interpreted as explaining the dependences among the $\{y_{ij}\}$ $j = 1, \cdots, t$. Denote the MLE of $\mathbf{\Sigma}$ by $\widehat{\mathbf{\Sigma}}$, where

$$\widehat{\mathbf{\Sigma}} = \widehat{\mathbf{\Gamma}}\widehat{\mathbf{\Gamma}}' + \widehat{\mathbf{\Psi}}; \tag{1.36}$$

with $\widehat{\mathbf{\Gamma}}$ and $\widehat{\mathbf{\Psi}}$ being the respective MLE's of $\mathbf{\Gamma}$ and $\mathbf{\Psi}$. For complete or incomplete data, the EM algorithm can be used to compute the MLEs. In practice, the number of underlying factors, f, is unknown. The Akaike information criterion, discussed in [1], is one popular method used to choose f.

Note that when $\mathbf{X} = \mathbf{1}_n$, then from the consistency of $\widehat{\mathbf{\Sigma}}$, the distributions of the random terms in (1.34), and using result 2c.x in [32], it follows that

$$\left(\mathbf{C}'\widehat{\mathbf{\Sigma}}\mathbf{C} \otimes [\mathbf{F}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{F}]\right)^{-1/2} \mathrm{vec}(\widehat{\mathbf{\Phi}} - \mathbf{\Phi}) \xrightarrow{L} \mathrm{N}(\mathbf{0},\ \mathbf{I}_{qs}). \tag{1.37}$$

Applying result 2c.xii in [32] to the random variable in (1.37), it follows that

$$\mathrm{vec}'(\widehat{\mathbf{\Phi}} - \mathbf{\Phi})\left(\mathbf{C}'\widehat{\mathbf{\Sigma}}\mathbf{C} \otimes [\mathbf{F}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{F}]\right)^{-1} \mathrm{vec}(\widehat{\mathbf{\Phi}} - \mathbf{\Phi}), \xrightarrow{L} \chi^2(qs). \tag{1.38}$$

This distributional result implies that, for large samples, inferences such as testing $\mathrm{H}_0\colon \mathbf{\Phi} = \mathbf{\Phi}_0$ for $\mathbf{\Phi}$ in (1.5) can be based on the statistic in (1.38).

CHAPTER 2

A MODIFIED MIXED MODEL FOR THE TRANSFORMED COVARIANCE STRUCTURE

Introduction

For the classic multivariate model in (1.1), the only restriction on the covariance matrix Σ is that it be positive definite. Therefore, the classic model is a sensible model for many repeated measures data sets. However, when the number of subjects, n, is small or the number of repeated measures, t, is large, the cost of using the classic model is the loss of efficiency when making inferences on $\mathbf{\Phi}$ in (1.5). Efficiency can be greatly improved by employing a parsimonious covariance model. The approaches reviewed in Chapter 1 reduce the number of covariance parameters to estimate, but each requires additional information and assumptions.

A time series analysis requires modeling the covariance matrix according to a particular stochastic process. This requires that the researcher have a priori information about which process to use. Also, the assumed stationary process can give rise to a restrictive covariance matrix. Suppose, for example, the ARMA(p,q) time series model is used. This results in a covariance matrix with a Toeplitz structure which would be inappropriate for growth or learning experiments, where the variances along the diagonals are not expected to be the same.

The mixed model described in (1.23) requires that $\mathbf{Z}_i$ be a known $t \times f$ matrix, having full column rank. If f is greater than 1, the rank requirement implies at least one column of $\mathbf{Z}_i$ must contain time-varying covariates. This means the researcher has the often difficult task of specifying these covariates a priori.

Univariate analyses are a sensible procedure if sphericity is satisfied. However, when the subjects are viewed as random blocking factors, the t occasions are not randomly ordered within each block and sphericity is rarely satisfied. The ϵ-adjusted tests adequately control test size but, as shown in [7], they need not be more powerful than classic multivariate analyses.

In the following section a modified mixed model (MMM) is described that also reduces the number of covariance parameters to estimate. Although the MMM requires additional assumptions when compared to the classic multivariate model of (1.1), it will be argued that the MMM has certain advantages.

A Modified Mixed Model

Suppose the variance for the i^{th} subject, Σ_i, is the same for all subjects so that $\Sigma_i = \Sigma$ for all i. Then, to make inferences concerning $\mathbf{\Phi}$ in (1.5), a structure must be specified for the matrix

$$\mathbf{\Upsilon} = \mathbf{C}'\Sigma\mathbf{C}, \tag{2.1}$$

where $\mathbf{C}$ is the $t \times q$ orthonormal matrix of contrast coefficients in (1.5). The structure assumed for $\mathbf{\Upsilon}$ may be indirectly specified by the structure assumed for Σ, or $\mathbf{\Upsilon}$ may be modeled directly. Irregardless of how a model is specified for $\mathbf{\Upsilon}$, inferences concerning $\mathbf{\Phi}$ based on $\widehat{\mathbf{\Phi}}$ in (1.8) will require that $\mathbf{\Upsilon}$ be estimated, which requires a model for $\mathbf{\Upsilon}$.

Let the $n \times t$ random matrix $\mathbf{U}$ in the linear model in (1.1) be normally distributed as in (1.3). This implies that the distribution of $\mathbf{U}^* = \mathbf{UC}$ is

$$\text{vec}(\mathbf{U}^*) \sim \text{N}_{nq}[\mathbf{0},\ (\mathbf{\Upsilon} \otimes \mathbf{I}_n)]. \tag{2.2}$$

Let $\mathbf{u}_i^{*\prime}$ be the i^{th} row of $\mathbf{U}^*$, then the distribution in (2.2) implies the $q \times 1$ vector,

$\mathbf{u}_i^*$ is distributed as

$$\mathbf{u}_i^* \sim \mathrm{N}_q[\mathbf{0},\ \boldsymbol{\Upsilon}] \tag{2.3}$$

for all i. A modified mixed model can now be constructed by modeling the $\mathbf{u}_i^*$ as

$$\mathbf{u}_i^* = \boldsymbol{\Gamma}\mathbf{w}_i + \mathbf{e}_i \quad \text{for} \quad i = 1, \ldots, n, \tag{2.4}$$

where $\boldsymbol{\Gamma}$ is an unknown $q \times f$ design matrix for the random effects; $\text{rank}(\boldsymbol{\Gamma}) = f$; $\mathbf{w}_i$ and $\mathbf{e}_i$ are independently distributed as $\mathbf{w}_i \sim \text{iid}\,\mathrm{N}_f(\mathbf{0},\ \boldsymbol{\Delta}_f)$, where $\boldsymbol{\Delta}_f$ is a positive definite matrix, and $\mathbf{e}_i \sim \text{iid}\,\mathrm{N}_q(\mathbf{0},\ \sigma^2\mathbf{I}_q)$. Note the similarity between the modified mixed model in (2.4) and the mixed model in (1.23). The only differences are in the dimensions of the vectors and matrices; and in the mixed model, the within subjects design matrix $\mathbf{Z}_i$ is known. The latter condition is modified for the model in (2.4), where the design matrix $\boldsymbol{\Gamma}$ is unknown and must be estimated from the data.

Without loss of generality, $\boldsymbol{\Delta}_f = \mathbf{I}_f$ can be assumed. If $\boldsymbol{\Delta}_f \neq \mathbf{I}_f$, write equation (2.4) as

$$\begin{aligned} \mathbf{u}_i^* &= \boldsymbol{\Gamma}\boldsymbol{\Delta}_f^{1/2}\boldsymbol{\Delta}_f^{-1/2}\mathbf{w}_i + \mathbf{e}_i \\ &= \boldsymbol{\Gamma}^*\mathbf{w}_i^* + \mathbf{e}_i, \end{aligned} \tag{2.5}$$

where $\mathbf{w}_i^* = \boldsymbol{\Delta}_f^{-1/2}\mathbf{w}_i$ and $\text{var}(\mathbf{w}_i^*) = \mathbf{I}_f$. Thus, under model (2.4),

$$\mathbf{u}_i^* \sim \text{iid}\,\mathrm{N}_q[\mathbf{0},\ \boldsymbol{\Upsilon}] \quad \text{where} \quad \boldsymbol{\Upsilon} = \boldsymbol{\Gamma}^*\boldsymbol{\Gamma}^{*\prime} + \sigma^2\mathbf{I}_q \tag{2.6}$$

for $i = 1, \ldots, n$.

The main difference between the modified mixed model in (2.4), assuming the distribution in (2.6), and the factor analytic model in (1.34) is that; in the factor analytic model, $\text{var}(\mathbf{e}_i) = \boldsymbol{\Psi}$, where the only restriction on $\boldsymbol{\Psi}$ is that it is diagonal, and in the modified mixed model in (2.4), $\text{var}(\mathbf{e}_i) = \sigma^2\mathbf{I}_t$. One might be tempted to generalize the proposed model so that $\mathbf{e}_i \sim \text{iid}\, \mathrm{N}_t(\mathbf{0},\ \boldsymbol{\Psi}_q)$, where $\boldsymbol{\Psi}_q$ is a diagonal. In this generalized model, $\boldsymbol{\Upsilon}$ has the form

$$\boldsymbol{\Upsilon} = \boldsymbol{\Gamma}\boldsymbol{\Gamma}' + \boldsymbol{\Psi}_q. \tag{2.7}$$

To compare the model for $\boldsymbol{\Upsilon}$ in (2.6) with the more general structure in (2.7), examine inferences regarding $\boldsymbol{\Phi}$ in (1.5). A typical inference regarding $\boldsymbol{\Phi}$ in (1.5) is to test the hypotheses

$$\mathrm{H}_0\colon \mathbf{F}'\mathbf{B}\mathbf{C} = \mathbf{0} \quad \text{against} \quad \mathrm{H}_a\colon \mathbf{F}'\mathbf{B}\mathbf{C} \neq \mathbf{0}. \tag{2.8}$$

A natural restriction to impose on any test of the hypotheses in (2.8) is that the test be invariant under the group $G_{\mathbf{C}}$, where, for any orthonormal matrix $\mathbf{C}$ such that $\mathbf{C}'\mathbf{1}_t = \mathbf{0}$, the group of transformations $G_{\mathbf{C}}$ is

$$G_{\mathbf{C}} = [g;\ g(\mathbf{YC}) = \mathbf{YCQ},\ \mathbf{Q}'\mathbf{Q} = \mathbf{QQ}' = \mathbf{I}_q]\,. \tag{2.9}$$

One appealing property of an invariant test is that the decision rule is independent of the particular coordinate system in which the data matrix, $\mathbf{Y}$, is expressed. A detailed discussion of invariant tests is given in Chapter 6 in [24]. For a test to be invariant under $G_{\mathbf{C}}$, a family of densities must be specified that is invariant to $G_{\mathbf{C}}$. Let $\mathcal{F}$ be a family of probability density functions (pdfs) and write $\mathcal{F}$ as

$$\mathcal{F} = \{f_{\mathbf{y}}(y|\ \boldsymbol{\theta});\ \boldsymbol{\theta} \in \boldsymbol{\Omega}\}\,.$$

Then $\mathcal{F}$ is invariant to $G_{\mathbf{C}}$ if, for any $g \in G_{\mathbf{C}}$ and $\boldsymbol{\theta} \in \boldsymbol{\Omega}$ there exists a function $\overline{g}$ such that

$$\mathbf{y}^* = g(\mathbf{y}) \sim f_{\mathbf{y}}(y^*;\ \overline{g}(\boldsymbol{\theta})) \quad \text{and} \quad \overline{g}(\boldsymbol{\theta}) \in \boldsymbol{\Omega}.$$

Suppose $\mathbf{YC}$ is a $n \times q$ random matrix with mean M and dispersion $\boldsymbol{\Upsilon}$. Let $\mathbf{y}_c = \mathrm{vec}(\mathbf{YC})$ and consider the two families of probability density functions

$$\mathcal{F}_1 = \{f_{\mathbf{y}_c}(y_c|\ \mathrm{M},\ \boldsymbol{\Upsilon});\ \boldsymbol{\Upsilon} = \boldsymbol{\Gamma}\boldsymbol{\Gamma}' + \boldsymbol{\Psi}_q, \mathrm{rank}(\boldsymbol{\Gamma}) = f\} \quad \text{and} \tag{2.10}$$

$$\mathcal{F}_2 = \left\{f_{\mathbf{y}_c}(y_c|\ \mathrm{M},\ \boldsymbol{\Upsilon});\ \boldsymbol{\Upsilon} = \boldsymbol{\Gamma}\boldsymbol{\Gamma}' + \sigma^2\mathbf{I}_q, \mathrm{rank}(\boldsymbol{\Gamma}) = f\right\}. \tag{2.11}$$

In order to determine whether $\mathcal{F}_1$ in (2.10) or $\mathcal{F}_2$ in (2.11) are invariant to $G_{\mathbf{C}}$, consider the following definitions and theorem:

Definition 2.1 *Let $\mathcal{D}_q$ be the group of $q \times q$ positive definite diagonal matrices.*

Definition 2.2 *Let $\mathcal{P}_q$ be the group of $q \times q$ positive semi-definite matrices.*

Definition 2.3 *A positive definite, $t \times t$ matrix Σ is said to have Type I factor analytic (FA) structure with dimension f if*

$$\min_{\substack{\Psi \in \mathcal{D}_t \\ \Sigma - \Psi \in \mathcal{P}_t}} \operatorname{rank}(\Sigma - \Psi) = f.$$

Definition 2.4 *A positive definite, $q \times q$ matrix $\boldsymbol{\Upsilon}$ is said to have Type II FA structure with dimension f if*

$$\min_{\substack{\sigma^2 > 0 \\ \boldsymbol{\Upsilon} - \sigma^2 \mathbf{I} \in \mathcal{P}_q}} \operatorname{rank}(\boldsymbol{\Upsilon} - \sigma^2 \mathbf{I}) = f.$$

Theorem 2.1 *Let $\mathbf{C}$ be a $t \times q$ matrix with rank q, and let $f > 0$. Then*

(1) $\mathbf{C}'\Sigma\mathbf{C}$ *has Type II FA structure with dimension f* $\Rightarrow$ $\mathbf{Q}'\mathbf{C}'\Sigma\mathbf{C}\mathbf{Q}$ *has Type II FA structure with dimension f for all* $\mathbf{Q} \ni \mathbf{Q}'\mathbf{Q} = \mathbf{Q}\mathbf{Q}' = \mathbf{I}$,

and

(2) $\mathbf{C}'\Sigma\mathbf{C}$ *has Type I FA structure with dimension f* $\not\Rightarrow$ $\mathbf{Q}'\mathbf{C}'\Sigma\mathbf{C}\mathbf{Q}$ *has Type I FA structure with dimension f for all* $\mathbf{Q} \ni \mathbf{Q}'\mathbf{Q} = \mathbf{Q}\mathbf{Q}' = \mathbf{I}$.

proof of (1):

Let $\boldsymbol{\Upsilon} = \mathbf{C}'\Sigma\mathbf{C}$. Then, by definition (2.4),

$$\min_{\substack{\sigma^2 > 0 \\ \boldsymbol{\Upsilon} - \sigma^2 \mathbf{I} \in \mathcal{P}_q}} \operatorname{rank}(\boldsymbol{\Upsilon} - \sigma^2 \mathbf{I}) = f.$$

Note that, for any orthogonal matrix $\mathbf{Q}$, $\text{rank}(\mathbf{Q}'\mathbf{A}\mathbf{Q}) = \text{rank}(\mathbf{A})$, therefore, for any orthogonal matrix $\mathbf{Q}$,

$$\min_{\substack{\sigma^2>0 \\ \boldsymbol{\Upsilon}-\sigma^2\mathbf{I}\in\mathcal{P}_q}} \text{rank}\left[\mathbf{Q}'(\boldsymbol{\Upsilon}-\sigma^2\mathbf{I})\mathbf{Q}\right] = f$$

$$\Rightarrow \min_{\substack{\sigma^2>0 \\ \mathbf{Q}'(\boldsymbol{\Upsilon}-\sigma^2\mathbf{I})\mathbf{Q}\in\mathcal{P}_q}} \text{rank}\left[\mathbf{Q}'(\boldsymbol{\Upsilon}-\sigma^2\mathbf{I})\mathbf{Q}\right] = f$$

because $\boldsymbol{\Upsilon}-\sigma^2\mathbf{I}\in\mathcal{P}_q \iff \mathbf{Q}'(\boldsymbol{\Upsilon}-\sigma^2\mathbf{I})\mathbf{Q}\in\mathcal{P}_q$. Note that

$$\min_{\substack{\sigma^2>0 \\ \mathbf{Q}'(\boldsymbol{\Upsilon}-\sigma^2\mathbf{I})\mathbf{Q}\in\mathcal{P}_q}} \text{rank}\left[\mathbf{Q}'(\boldsymbol{\Upsilon}-\sigma^2\mathbf{I})\mathbf{Q}\right] = f$$

$$\Rightarrow \min_{\substack{\sigma^2>0 \\ (\mathbf{Q}'\boldsymbol{\Upsilon}\mathbf{Q}-\sigma^2\mathbf{I})\in\mathcal{P}_q}} \text{rank}\left[\mathbf{Q}'\boldsymbol{\Upsilon}\mathbf{Q}-\sigma^2\mathbf{I})\right] = f,$$

and this implies $\mathbf{Q}'\boldsymbol{\Upsilon}\mathbf{Q}$ has Type II FA structure with dimension f by definition (2.4).

proof of (2):

Suppose $\mathbf{C}'\boldsymbol{\Sigma}\mathbf{C}$ has Type I FA structure with dimension $f \Rightarrow \mathbf{Q}'\mathbf{C}'\boldsymbol{\Sigma}\mathbf{C}\mathbf{Q}$ has Type I FA structure with dimension f for all $\mathbf{Q} \ni \mathbf{Q}'\mathbf{Q} = \mathbf{Q}\mathbf{Q}' = \mathbf{I}$. Decompose $\mathbf{C}'\boldsymbol{\Sigma}\mathbf{C}$ to $\mathbf{V}\mathbf{D}\mathbf{V}'$, where $\mathbf{V}$ is a $q \times q$ orthogonal matrix and $\mathbf{D}$ a diagonal matrix. Let $\mathbf{Q} = \mathbf{V}$, then

$$\mathbf{Q}'\mathbf{C}'\boldsymbol{\Sigma}\mathbf{C}\mathbf{Q} = \mathbf{V}'\mathbf{V}\mathbf{D}\mathbf{V}'\mathbf{V} = \mathbf{D} \in \mathcal{D}_q,$$

and

$$\min_{\substack{\boldsymbol{\Psi}\in\mathcal{D}_q \\ \mathbf{D}-\boldsymbol{\Psi}\in\mathcal{P}_q}} \text{rank}(\mathbf{D}-\boldsymbol{\Psi}) = 0;$$

which contradicts the hypothesis because $f \neq 0$. $\square$

Corollary 2.1 *When $\mathcal{F}_1$ and $\mathcal{F}_2$ are comprised of normal pdfs, then $\mathcal{F}_1$ is not invariant to $G_{\mathbf{C}}$ and $\mathcal{F}_2$ is invariant to $G_{\mathbf{C}}$.*

proof:

Assuming the normal distribution in (1.3), it is readily shown that for $\mathbf{YC}$,

$$\text{vec}(\mathbf{YC}) \sim \text{N}_{nq}\left(\text{vec}(\mathbf{XBC}),\ (\boldsymbol{\Upsilon} \otimes \mathbf{I}_n)\right). \tag{2.12}$$

Thus, because $\mathbf{YCQ}$ is a linear transformation of a normally distributed matrix, this family of normal distributions is invariant to $G_{\mathbf{C}}$, where

$$\overline{g}\,(\mathbf{XBC},\ \boldsymbol{\Upsilon}) = (\mathbf{XBCQ},\ \mathbf{Q}'\boldsymbol{\Upsilon}\mathbf{Q}), \tag{2.13}$$

whenever $\boldsymbol{\Upsilon}$ has the structure given in (2.6). □

To summarize, for the modified mixed model in (2.4), the normal distribution and parameter structure in (2.6) specifies a family of densities that is invariant to $G_{\mathbf{C}}$, and for that reason the structure for $\boldsymbol{\Upsilon}$ in (2.6) is preferred to the more general structure for $\boldsymbol{\Upsilon}$ in (2.7).

CHAPTER 3

ESTIMATING PARAMETERS UNDER A MODIFIED MIXED MODEL

Introduction

Suppose interest is in estimating $\boldsymbol{\Upsilon}$ under the modified mixed model in (2.6). As noted in [4], to find a unique MLE of $\boldsymbol{\Upsilon}$, say $\widehat{\boldsymbol{\Upsilon}}$, where

$$\widehat{\boldsymbol{\Upsilon}} = \widehat{\boldsymbol{\Gamma}}\widehat{\boldsymbol{\Gamma}}' + \hat{\sigma}^2\mathbf{I}_q, \tag{3.1}$$

a specific orthogonal rotation of the latent variables must be given. In general, an orthogonal rotation is given by

$$\mathbf{u}_i^* = \boldsymbol{\Gamma}\mathbf{Q}\mathbf{Q}'\mathbf{w}_i + \mathbf{e}_i = \boldsymbol{\Gamma}^*\mathbf{w}_i^* + \mathbf{e}_i,$$

for $\mathbf{u}_i^*$ in (2.4) and any orthogonal matrix $\mathbf{Q}$ of conformable dimensions. Throughout the remainder of this thesis, it will be assumed that $\boldsymbol{\Gamma}^*$ is a $q \times f$ matrix with zeros above the main diagonal. This condition ensures that $\boldsymbol{\Gamma}^*$ be identifiable under the normal distribution in (2.6).

One common method for finding $\widehat{\boldsymbol{\Upsilon}}$ is to use the invariance property of MLEs under parameter transformations. The invariance property is described in [13]. Using this property, the MLE of $\boldsymbol{\Upsilon}$ is given by

$$\widehat{\boldsymbol{\Upsilon}} = \mathbf{C}'\boldsymbol{\Sigma}_{\widehat{\boldsymbol{\theta}}}\mathbf{C}, \tag{3.2}$$

where $\boldsymbol{\Sigma}_{\widehat{\boldsymbol{\theta}}}$ is the MLE of $\boldsymbol{\Sigma}_{\boldsymbol{\theta}}$, a covariance structure that depends on a $k \times 1$ vector of parameters $\boldsymbol{\theta}$. Note that this method requires estimating all the covariance parameters in $\boldsymbol{\theta}$ when computing $\widehat{\boldsymbol{\Upsilon}}$, which requires a model be specified for $\boldsymbol{\Sigma}_{\boldsymbol{\theta}}$. The modified mixed model in (2.4) specifies a structure for $\mathbf{C}'\Sigma\mathbf{C}$, but does not directly specify a model for Σ. Clearly if Σ has the FA structure in (1.34) with $\Psi = \sigma^2\mathbf{I}$, then $\mathbf{C}'\Sigma\mathbf{C}$ has the modified mixed model structure in (2.4). Nonetheless, the FA structure in (1.34) with $\Psi = \sigma^2\mathbf{I}$ is sufficient for the parameter Υ to have the structure in (2.2) but not necessary. Necessary and sufficient conditions are given in corollary (3.2).

The first part of this section proves a theorem which specifies a model for $\Sigma_{\boldsymbol{\theta}}$ when Υ is structured as a function of $\boldsymbol{\theta}$. The next section proves a theorem which gives results on computing MLEs. The final two sections describe algorithms that can be used for computing MLEs under complete and incomplete data.

A Covariance Structure Result

Theorem 3.2 *Let $\Sigma: t \times t$ be a positive definite matrix and let $\boldsymbol{\Omega}_{11}(\boldsymbol{\theta})$ be a $q \times q$ positive definite matrix. The elements of $\boldsymbol{\Omega}_{11}(\boldsymbol{\theta})$ are functions of $\boldsymbol{\theta}: k \times 1$. Then*

$$\mathbf{C}'\Sigma\mathbf{C} = \boldsymbol{\Omega}_{11}(\boldsymbol{\theta}) \quad \Longleftrightarrow \quad \Sigma = \mathbf{C}^* \begin{pmatrix} \boldsymbol{\Omega}_{11}(\boldsymbol{\theta}) & \boldsymbol{\Omega}_{12} \\ \boldsymbol{\Omega}_{21} & \boldsymbol{\Omega}_{22} \end{pmatrix} \mathbf{C}^{*\prime},$$

where $\mathbf{C}^* = [\mathbf{C} \quad \mathbf{L}]$ is a $t \times t$ orthogonal matrix; $\mathbf{L}$ is a $t \times (t-q)$ orthonormal matrix that is orthogonal to $\mathbf{C}$, the orthonormal matrix in (1.5), and

$$\boldsymbol{\Omega} = \begin{pmatrix} \boldsymbol{\Omega}_{11}(\boldsymbol{\theta}) & \boldsymbol{\Omega}_{12} \\ \boldsymbol{\Omega}_{21} & \boldsymbol{\Omega}_{22} \end{pmatrix} \tag{3.3}$$

is positive definite.

proof of only if:

Suppose $\mathbf{C}'\Sigma\mathbf{C} = \boldsymbol{\Omega}_{11}(\boldsymbol{\theta})$. Then

$$(\mathbf{C}' \otimes \mathbf{C}')\,\mathrm{vec}(\Sigma) = \mathrm{vec}(\boldsymbol{\Omega}_{11}(\boldsymbol{\theta}))$$

and there exists a $t^2 \times q^2$ g-inverse of $(\mathbf{C}' \otimes \mathbf{C}')$, say $(\mathbf{C}' \otimes \mathbf{C}')^g$, which satisfies

$$\text{vec}(\boldsymbol{\Sigma}) = (\mathbf{C}' \otimes \mathbf{C}')^g \, \text{vec}(\boldsymbol{\Omega}_{11}(\boldsymbol{\theta})).$$

Note that $(\mathbf{C} \otimes \mathbf{C})$ is a g-inverse of $(\mathbf{C}' \otimes \mathbf{C}')$. Thus, by result (23) on page 220 in [35], there exists a $t^2 \times q^2$ matrix $\mathbf{W}$ and a $t^2 \times q^2$ matrix $\mathbf{V}$ such that $(\mathbf{C}' \otimes \mathbf{C}')^g$ can be written as

$$(\mathbf{C}' \otimes \mathbf{C}')^g = (\mathbf{C} \otimes \mathbf{C})(\mathbf{C}' \otimes \mathbf{C}')(\mathbf{C} \otimes \mathbf{C}) + (\mathbf{I}_{t^2} - (\mathbf{C} \otimes \mathbf{C})(\mathbf{C}' \otimes \mathbf{C}'))\mathbf{W}$$

$$+\mathbf{V}(\mathbf{I}_{q^2} - (\mathbf{C}' \otimes \mathbf{C}')(\mathbf{C} \otimes \mathbf{C}))$$

$$= (\mathbf{C} \otimes \mathbf{C}) + (\mathbf{I} - (\mathbf{C}\mathbf{C}' \otimes \mathbf{C}\mathbf{C}'))\mathbf{W}$$

because $\mathbf{I}_{q^2} - (\mathbf{C}' \otimes \mathbf{C}')(\mathbf{C} \otimes \mathbf{C}) = \mathbf{0}$. Therefore, $\text{vec}(\boldsymbol{\Sigma})$ can be written as

$$\text{vec}(\boldsymbol{\Sigma}) = \left[(\mathbf{C} \otimes \mathbf{C}) + (\mathbf{I} - (\mathbf{C}\mathbf{C}' \otimes \mathbf{C}\mathbf{C}'))\mathbf{W}\right] \text{vec}(\boldsymbol{\Omega}_{11}(\boldsymbol{\theta}))$$

$$= (\mathbf{C} \otimes \mathbf{C}) \, \text{vec}(\boldsymbol{\Omega}_{11}(\boldsymbol{\theta})) + (\mathbf{I} - (\mathbf{C}\mathbf{C}' \otimes \mathbf{C}\mathbf{C}'))\mathbf{W} \, \text{vec}(\boldsymbol{\Omega}_{11}(\boldsymbol{\theta})).$$

Let $\text{vec}(\mathbf{K}) = (\mathbf{I} - (\mathbf{C}\mathbf{C}' \otimes \mathbf{C}\mathbf{C}'))\mathbf{W} \, \text{vec}(\boldsymbol{\Omega}_{11}(\boldsymbol{\theta}))$, thus

$$\text{vec}(\boldsymbol{\Sigma}) = \text{vec}(\mathbf{C}(\boldsymbol{\Omega}_{11}(\boldsymbol{\theta}))\mathbf{C}') + \text{vec}(\mathbf{K}).$$

It is also true that

$$\text{vec}(\mathbf{K}) = (\mathbf{I} - (\mathbf{C}\mathbf{C}' \otimes \mathbf{C}\mathbf{C}')) \, \text{vec}(\boldsymbol{\Sigma})$$

because

$$(\mathbf{I} - (\mathbf{C}\mathbf{C}' \otimes \mathbf{C}\mathbf{C}')) \, \text{vec}(\boldsymbol{\Sigma})$$

$$= (\mathbf{I} - (\mathbf{C}\mathbf{C}' \otimes \mathbf{C}\mathbf{C}'))(\mathbf{C} \otimes \mathbf{C}) \, \text{vec}(\boldsymbol{\Omega}_{11}(\boldsymbol{\theta})) + \text{vec}(\mathbf{K}) = \text{vec}(\mathbf{K})$$

using

$$(\mathbf{I} - (\mathbf{C}\mathbf{C}' \otimes \mathbf{C}\mathbf{C}'))(\mathbf{C} \otimes \mathbf{C}) = \mathbf{0}$$

and

$$(\mathbf{I} - (\mathbf{CC}' \otimes \mathbf{CC}'))\,\mathrm{vec}(\mathbf{K}) = \mathrm{vec}(\mathbf{K}).$$

Note, by using properties of perpendicular projection operators, the perpendicular projection operator $\mathbf{I}_t - \mathbf{LL}'$ can be factored as $\mathbf{CC}'$, which implies that

$$\begin{aligned}(\mathbf{I} - (\mathbf{CC}' \otimes \mathbf{CC}')) &= \mathbf{I} - (\mathbf{CC}' \otimes \mathbf{I}_t - \mathbf{LL}') \\ &= \mathbf{I} - ((\mathbf{CC}' \otimes \mathbf{I}_t) - (\mathbf{CC}' \otimes \mathbf{LL}')) \\ &= (\mathbf{I}_t \otimes \mathbf{I}_t) - (\mathbf{CC}' \otimes \mathbf{I}_t) + (\mathbf{CC}' \otimes \mathbf{LL}') \\ &= (\mathbf{CC}' \otimes \mathbf{LL}') + (\mathbf{LL}' \otimes \mathbf{I}_t).\end{aligned}$$

Therefore,

$$\mathrm{vec}(\mathbf{K}) = (\mathbf{I} - (\mathbf{CC}' \otimes \mathbf{CC}'))\,\mathrm{vec}(\Sigma) = [(\mathbf{CC}' \otimes \mathbf{LL}') + (\mathbf{LL}' \otimes \mathbf{I}_t)]\,\mathrm{vec}(\Sigma),$$

which gives the following expression for vec(**K**):

$$\begin{aligned}\mathrm{vec}(\mathbf{K}) &= \mathrm{vec}\left(\mathbf{LL}'\Sigma\mathbf{CC}' + \Sigma\mathbf{LL}'\right) \\ &= \mathrm{vec}\left(\mathbf{LL}'\Sigma\mathbf{CC}' + (\mathbf{I} - \mathbf{LL}' + \mathbf{LL}')\Sigma\mathbf{LL}'\right) \\ &= \mathrm{vec}\left(\mathbf{LL}'\Sigma\mathbf{CC}' + \mathbf{CC}'\Sigma\mathbf{LL}' + \mathbf{LL}'\Sigma\mathbf{LL}'\right).\end{aligned}$$

Therefore, Σ can be written as

$$\begin{aligned}\Sigma &= \mathbf{C}(\Omega_{11}(\theta))\mathbf{C}' + \mathbf{CC}'\Sigma\mathbf{LL}' + \mathbf{LL}'\Sigma\mathbf{CC}' + \mathbf{LL}'\Sigma\mathbf{LL}' \\ &= (\mathbf{C}\ \ \mathbf{L})\begin{pmatrix}\Omega_{11}(\theta) & \Omega_{12} \\ \Omega_{21} & \Omega_{22}\end{pmatrix}\begin{pmatrix}\mathbf{C}' \\ \mathbf{L}'\end{pmatrix};\end{aligned}$$

where $\Omega_{12} = \mathbf{C}'\Sigma\mathbf{L}$, $\Omega_{21} = \mathbf{L}'\Sigma\mathbf{C}$, $\Omega_{22} = \mathbf{L}'\Sigma\mathbf{L}$, and

$$\Omega = \begin{pmatrix}\Omega_{11}(\theta) & \Omega_{12} \\ \Omega_{21} & \Omega_{22}\end{pmatrix}$$

is positive definite because $\boldsymbol{\Sigma}$ and $\mathbf{C}^*$ are positive definite. Thus, if $\mathbf{C}'\boldsymbol{\Sigma}\mathbf{C} = \boldsymbol{\Omega}_{11}(\boldsymbol{\theta})$, then

$$\boldsymbol{\Sigma} = (\mathbf{C}\ \ \mathbf{L})\begin{pmatrix} \boldsymbol{\Omega}_{11}(\boldsymbol{\theta}) & \boldsymbol{\Omega}_{12} \\ \boldsymbol{\Omega}_{21} & \boldsymbol{\Omega}_{22} \end{pmatrix}\begin{pmatrix} \mathbf{C}' \\ \mathbf{L}' \end{pmatrix}.$$

proof of if :

Suppose

$$\begin{aligned} \boldsymbol{\Sigma} &= \mathbf{C}^* \begin{pmatrix} \boldsymbol{\Omega}_{11}(\boldsymbol{\theta}) & \boldsymbol{\Omega}_{12} \\ \boldsymbol{\Omega}_{21} & \boldsymbol{\Omega}_{22} \end{pmatrix} \mathbf{C}^{*\prime} \\ &= \mathbf{C}^*\ \boldsymbol{\Omega}\mathbf{C}^{*\prime}. \end{aligned}$$

Note that

$$\begin{aligned} &\mathbf{C}^* \begin{pmatrix} \boldsymbol{\Omega}_{11}(\boldsymbol{\theta}) & \boldsymbol{\Omega}_{12} \\ \boldsymbol{\Omega}_{21} & \boldsymbol{\Omega}_{22} \end{pmatrix} \mathbf{C}^{*\prime} \\ &= \mathbf{C}(\boldsymbol{\Omega}_{11}(\boldsymbol{\theta}))\mathbf{C}' + \mathbf{C}\boldsymbol{\Omega}_{12}\mathbf{L}' + \mathbf{L}\boldsymbol{\Omega}_{21}\mathbf{C}' + \mathbf{L}\boldsymbol{\Omega}_{22}\mathbf{L}'. \end{aligned}$$

Thus

$$\begin{aligned} \mathbf{C}'\boldsymbol{\Sigma}\mathbf{C} &= \mathbf{C}'\left[\mathbf{C}(\boldsymbol{\Omega}_{11}(\boldsymbol{\theta}))\mathbf{C}' + \mathbf{C}\boldsymbol{\Omega}_{12}\mathbf{L}' + \mathbf{L}\boldsymbol{\Omega}_{21}\mathbf{C}' + \mathbf{L}\boldsymbol{\Omega}_{22}\mathbf{L}'\right]\mathbf{C} \\ &= \boldsymbol{\Omega}_{11}(\boldsymbol{\theta}).\square \end{aligned}$$

Corollary 3.2

$$\mathbf{C}'\boldsymbol{\Sigma}\mathbf{C} = \boldsymbol{\Gamma}\boldsymbol{\Gamma}' + \sigma^2\mathbf{I}_q \iff \boldsymbol{\Sigma} = \mathbf{C}^* \begin{pmatrix} \boldsymbol{\Gamma}\boldsymbol{\Gamma}' + \sigma^2\mathbf{I}_t & \boldsymbol{\Omega}_{12} \\ \boldsymbol{\Omega}_{21} & \boldsymbol{\Omega}_{22} \end{pmatrix} \mathbf{C}^{*\prime}$$

where $\boldsymbol{\Gamma}\boldsymbol{\Gamma}' + \sigma^2\mathbf{I}_t - \boldsymbol{\Omega}_{12}\boldsymbol{\Omega}_{22}{}^{-1}\boldsymbol{\Omega}_{21}$ *is positive definite.*

A Marginal Distribution Result

Theorem 3.3 *Let* $\mathbf{A}$ *be a* $t \times t$ *positive definite matrix. Let* $\mathcal{P}_{\boldsymbol{\theta}}$ *be the set of all* $t \times t$ *positive definite matrices that can be written as follows:*

$$\boldsymbol{\Omega} \in \mathcal{P}_{\boldsymbol{\theta}} \Rightarrow \boldsymbol{\Omega} = \begin{pmatrix} \boldsymbol{\Omega}_{11}(\boldsymbol{\theta}) & \boldsymbol{\Omega}_{12} \\ \boldsymbol{\Omega}_{21} & \boldsymbol{\Omega}_{22} \end{pmatrix}, \tag{3.4}$$

where $\Omega_{11}(\boldsymbol{\theta})$ is a known function of a $k \times 1$ vector $\boldsymbol{\theta}$. Then

$$\sup_{\Omega \in \mathcal{P}_{\theta}} f(\Omega) = f(\widehat{\Omega}),$$

where

$$f(\Omega) = \frac{\exp\left[(-1/2)\,\mathrm{tr}\{\mathbf{A}\boldsymbol{\Omega}^{-1}\}\right]}{|\Omega|^{n/2}}; \tag{3.5}$$

$\mathbf{A}$ *is partitioned conformably to* $\boldsymbol{\Omega}$ *as*

$$\mathbf{A} = \begin{pmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{pmatrix}; \tag{3.6}$$

$$\widehat{\Omega} = \begin{pmatrix} \Omega_{11}(\widehat{\boldsymbol{\theta}}) & \widehat{\Omega}_{12} \\ \widehat{\Omega}_{21} & \widehat{\Omega}_{22} \end{pmatrix};$$

$\widehat{\boldsymbol{\theta}}$ *is the maximizer of*

$$\frac{\exp\left[(-1/2)\,\mathrm{trace}\{\mathbf{A}_{11}\boldsymbol{\Omega}_{11}^{-1}(\boldsymbol{\theta})\}\right]}{|\Omega_{11}(\boldsymbol{\theta})|^{n/2}}, \tag{3.7}$$

$$\widehat{\Omega}_{12} = \boldsymbol{\Omega}_{11}(\widehat{\boldsymbol{\theta}})\mathbf{A}_{11}^{-1}\mathbf{A}_{12}, \quad \text{and} \tag{3.8}$$

$$\widehat{\Omega}_{22} = \frac{\mathbf{A}_{22}}{n} - \mathbf{A}_{21}\mathbf{A}_{11}^{-1}\left(\frac{\mathbf{A}_{11}}{n} - \boldsymbol{\Omega}_{11}(\widehat{\boldsymbol{\theta}})\right)\mathbf{A}_{11}^{-1}\mathbf{A}_{12}. \tag{3.9}$$

proof:

The product $\mathbf{A}\boldsymbol{\Omega}^{-1}$ can be written as

$$\mathbf{A}\boldsymbol{\Omega}^{-1} =$$

$$\begin{pmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{pmatrix} \begin{pmatrix} \boldsymbol{\Omega}_{11}^{-1}(\boldsymbol{\theta}) + \boldsymbol{\Omega}_{11}^{-1}(\boldsymbol{\theta})\boldsymbol{\Omega}_{12}\boldsymbol{\Omega}_{22\cdot 1}^{-1}\boldsymbol{\Omega}_{21}\boldsymbol{\Omega}_{11}^{-1}(\boldsymbol{\theta}) & -\boldsymbol{\Omega}_{11}^{-1}(\boldsymbol{\theta})\boldsymbol{\Omega}_{12}\boldsymbol{\Omega}_{22\cdot 1}^{-1} \\ -\boldsymbol{\Omega}_{22\cdot 1}^{-1}\boldsymbol{\Omega}_{21}\boldsymbol{\Omega}_{11}^{-1}(\boldsymbol{\theta}) & \boldsymbol{\Omega}_{22\cdot 1}^{-1} \end{pmatrix},$$

where $\boldsymbol{\Omega}_{22\cdot 1} = \boldsymbol{\Omega}_{22} - \boldsymbol{\Omega}_{21}\boldsymbol{\Omega}_{11}^{-1}(\boldsymbol{\theta})\boldsymbol{\Omega}_{12}$, and the partition of $\boldsymbol{\Omega}^{-1}$ is based on the result on page 260 in [35].

Therefore,

$$\mathrm{tr}\,(\mathbf{A}\boldsymbol{\Omega}^{-1}) = \mathrm{tr}\left(\mathbf{A}_{11}\boldsymbol{\Omega}_{11}^{-1}(\boldsymbol{\theta})\right) +$$

$$\mathrm{tr}\left(\mathbf{A}_{11}\boldsymbol{\Omega}_{11}^{-1}(\boldsymbol{\theta})\boldsymbol{\Omega}_{12}\boldsymbol{\Omega}_{22\cdot1}^{-1}\boldsymbol{\Omega}_{21}\boldsymbol{\Omega}_{11}^{-1}(\boldsymbol{\theta}) - \mathbf{A}_{12}\boldsymbol{\Omega}_{22\cdot1}^{-1}\boldsymbol{\Omega}_{21}\boldsymbol{\Omega}_{11}^{-1}(\boldsymbol{\theta}) - \mathbf{A}_{21}\boldsymbol{\Omega}_{11}^{-1}(\boldsymbol{\theta})\boldsymbol{\Omega}_{12}\boldsymbol{\Omega}_{22\cdot1}^{-1} + \mathbf{A}_{22}\boldsymbol{\Omega}_{22\cdot1}^{-1}\right)$$

$$= \mathrm{tr}\left(\mathbf{A}_{11}\boldsymbol{\Omega}_{11}^{-1}(\boldsymbol{\theta})\right) +$$

$$\mathrm{tr}\left[\left(\boldsymbol{\Omega}_{21}\boldsymbol{\Omega}_{11}^{-1}(\boldsymbol{\theta})\mathbf{A}_{11}\boldsymbol{\Omega}_{11}^{-1}(\boldsymbol{\theta})\boldsymbol{\Omega}_{12} - \boldsymbol{\Omega}_{21}\boldsymbol{\Omega}_{11}^{-1}(\boldsymbol{\theta})\mathbf{A}_{12} - \mathbf{A}_{21}\boldsymbol{\Omega}_{11}^{-1}(\boldsymbol{\theta})\boldsymbol{\Omega}_{12} + \mathbf{A}_{22}\right)\boldsymbol{\Omega}_{22\cdot1}^{-1}\right]$$

$$= \mathrm{tr}\left(\mathbf{A}_{11}\boldsymbol{\Omega}_{11}^{-1}(\boldsymbol{\theta})\right) + \mathrm{tr}\left[(\mathbf{A}_{22} - \boldsymbol{\Delta}'\mathbf{A}_{12} - \mathbf{A}_{21}\boldsymbol{\Delta} + \boldsymbol{\Delta}'\mathbf{A}_{11}\boldsymbol{\Delta})\,\boldsymbol{\Omega}_{22\cdot1}^{-1}\right], \qquad (3.10)$$

where $\boldsymbol{\Delta} = \boldsymbol{\Omega}_{11}^{-1}(\boldsymbol{\theta})\boldsymbol{\Omega}_{12}$. By writing $\mathrm{tr}\,(\mathbf{A}\boldsymbol{\Omega}^{-1})$ as in (3.10), $f(\boldsymbol{\Omega})$ can be factored as

$$f(\boldsymbol{\Omega}) = \frac{\exp\left[(-1/2)\,\mathrm{tr}\{\mathbf{A}_{11}\boldsymbol{\Omega}_{11}^{-1}(\boldsymbol{\theta})\}\right]}{|\boldsymbol{\Omega}_{11}(\boldsymbol{\theta})|^{n/2}}$$

$$\times \frac{\exp\left[(-1/2)\,\mathrm{tr}\{(\mathbf{A}_{22} - \boldsymbol{\Delta}'\mathbf{A}_{12} - \mathbf{A}_{21}\boldsymbol{\Delta} + \boldsymbol{\Delta}'\mathbf{A}_{11}\boldsymbol{\Delta})\,\boldsymbol{\Omega}_{22\cdot1}^{-1}\}\right]}{|\boldsymbol{\Omega}_{22\cdot1}|^{n/2}},$$

where $|\boldsymbol{\Omega}| = |\boldsymbol{\Omega}_{11}(\boldsymbol{\theta})| \times |\boldsymbol{\Omega}_{22\cdot1}|$ by Theorem 8.2.1 in [16].

Let

$$f_1(\boldsymbol{\theta}) = \frac{\exp\left[(-1/2)\,\mathrm{tr}\{\mathbf{A}_{11}\boldsymbol{\Omega}_{11}^{-1}(\boldsymbol{\theta})\}\right]}{|\boldsymbol{\Omega}_{11}(\boldsymbol{\theta})|^{n/2}}$$

and

$$f_2(\boldsymbol{\Delta}, \boldsymbol{\Omega}_{22\cdot1}) = \frac{\exp\left[(-1/2)\,\mathrm{tr}\{(\mathbf{A}_{22} - \boldsymbol{\Delta}'\mathbf{A}_{12} - \mathbf{A}_{21}\boldsymbol{\Delta} + \boldsymbol{\Delta}'\mathbf{A}_{11}\boldsymbol{\Delta})\,\boldsymbol{\Omega}_{22\cdot1}^{-1}\}\right]}{|\boldsymbol{\Omega}_{22\cdot1}|^{n/2}}.$$

Then $f(\boldsymbol{\Omega}) = f_1(\boldsymbol{\theta})f_2(\boldsymbol{\Delta}, \boldsymbol{\Omega}_{22\cdot1})$.

Write $\mathbf{A}$ as a Choleskey decomposition, that is, $\mathbf{A} = \mathbf{U}\mathbf{U}'$. Partition $\mathbf{U}$ conformably with $\mathbf{A}$ as $(\mathbf{U}_1\;\;\mathbf{U}_2)$. That is, $\mathbf{A}_{11} = \mathbf{U}_1'\mathbf{U}_1$, $\mathbf{A}_{12} = \mathbf{U}_1'\mathbf{U}_2$, and $\mathbf{A}_{22} = \mathbf{U}_2'\mathbf{U}_2$. Then $f_2(\boldsymbol{\Delta}\;, \boldsymbol{\Omega}_{22\cdot1})$ can be expressed as

$$f_2(\boldsymbol{\Delta}, \boldsymbol{\Omega}_{22\cdot1}) = \frac{\exp\left[(-1/2)\,\mathrm{tr}\{(\mathbf{U}_2 - \mathbf{U}_1\boldsymbol{\Delta})'(\mathbf{U}_2 - \mathbf{U}_1\boldsymbol{\Delta})\boldsymbol{\Omega}_{22\cdot1}^{-1}\}\right]}{|\boldsymbol{\Omega}_{22\cdot1}|^{n/2}} \qquad (3.11)$$

and, by results given in [2], the maximizers of $f_2(\mathbf{\Delta}, \mathbf{\Omega}_{22\cdot1})$ are

$$\hat{\mathbf{\Delta}} = (\mathbf{U}_1'\mathbf{U}_1)^{-1}\mathbf{U}_1'\mathbf{U}_2 = \mathbf{A}_{11}^{-1}\mathbf{A}_{12}$$

and

$$\widehat{\mathbf{\Omega}}_{22\cdot1} = \frac{\mathbf{U}_2'(\mathbf{I} - \mathbf{U}_1(\mathbf{U}_1'\mathbf{U}_1)^{-1}\mathbf{U}_1')\mathbf{U}_2}{n} = \frac{\mathbf{A}_{22\cdot1}}{n}.$$

Let $\widehat{\boldsymbol{\theta}}$ be the maximizer of f_1, subject to $\mathbf{\Omega}_{11}(\boldsymbol{\theta}) > 0$. Write $\widehat{\mathbf{\Omega}}_{12}$ and $\widehat{\mathbf{\Omega}}_{22}$ as

$$\widehat{\mathbf{\Omega}}_{12} = \mathbf{\Omega}_{11}(\widehat{\boldsymbol{\theta}})\hat{\mathbf{\Delta}} = \mathbf{\Omega}_{11}(\widehat{\boldsymbol{\theta}})\mathbf{A}_{11}^{-1}\mathbf{A}_{12},$$

and

$$\begin{aligned}\widehat{\mathbf{\Omega}}_{22} &= \widehat{\mathbf{\Omega}}_{22\cdot1} + \widehat{\mathbf{\Omega}}_{21}\mathbf{\Omega}_{11}^{-1}(\widehat{\boldsymbol{\theta}})\widehat{\mathbf{\Omega}}_{12} \\ &= \frac{\mathbf{A}_{22}}{n} - \mathbf{A}_{21}\mathbf{A}_{11}^{-1}\left(\frac{\mathbf{A}_{11}}{n} - \mathbf{\Omega}_{11}(\widehat{\boldsymbol{\theta}})\right)\mathbf{A}_{11}^{-1}\mathbf{A}_{12}.\end{aligned}$$

Let

$$\widehat{\mathbf{\Omega}} = \begin{pmatrix} \mathbf{\Omega}_{11}(\widehat{\boldsymbol{\theta}}) & \widehat{\mathbf{\Omega}}_{12} \\ \widehat{\mathbf{\Omega}}_{21} & \widehat{\mathbf{\Omega}}_{22} \end{pmatrix}. \tag{3.12}$$

To show $\widehat{\mathbf{\Omega}} \in \mathcal{P}_{\boldsymbol{\theta}}$, factor $\mathbf{\Omega}_{11}(\widehat{\boldsymbol{\theta}})$ and $\widehat{\mathbf{\Omega}}_{22\cdot1}$ with a Cholesky decomposition so that

$$\mathbf{\Omega}_{11}(\widehat{\boldsymbol{\theta}}) = \mathbf{G}\mathbf{G}' \quad \text{and}$$

$$\widehat{\mathbf{\Omega}}_{22\cdot1} = \mathbf{F}\mathbf{F}'.$$

Then

$$\begin{pmatrix} \mathbf{G} & \mathbf{0} \\ \hat{\mathbf{\Delta}}'\mathbf{G} & \mathbf{F} \end{pmatrix}$$

is a rank t lower triangular matrix with positive diagonal elements. Also, $\widehat{\mathbf{\Omega}}$ in (3.12) can be factored as

$$\widehat{\mathbf{\Omega}} = \begin{pmatrix} \mathbf{G} & \mathbf{0} \\ \hat{\mathbf{\Delta}}'\mathbf{G} & \mathbf{F} \end{pmatrix}\begin{pmatrix} \mathbf{G}' & \mathbf{G}'\hat{\mathbf{\Delta}} \\ \mathbf{0}' & \mathbf{F}' \end{pmatrix}$$

which, by Theorem 12.2.2 in [16], implies $\widehat{\mathbf{\Omega}}$ is positive definite. Moreover, recall that

$$f(\mathbf{\Omega}) = f_1(\boldsymbol{\theta}) \times f_2(\mathbf{\Delta}, \mathbf{\Omega}_{22\cdot1}),$$

where $f_1(\boldsymbol{\theta}) > 0$ and $f_2(\boldsymbol{\Delta}, \boldsymbol{\Omega}_{22\cdot 1}) > 0$ for all $\boldsymbol{\Omega}$ in the domain of f. Thus,

$$\sup_{\boldsymbol{\Omega}\in\mathcal{P}_{\boldsymbol{\theta}}} f(\boldsymbol{\Omega}) \leq \sup_{\boldsymbol{\theta}\in\boldsymbol{\Omega}_{\boldsymbol{\theta}}} f_1(\boldsymbol{\theta}) \sup_{\substack{\|\boldsymbol{\Delta}\|<\infty \\ \boldsymbol{\Omega}_{22\cdot 1}>0}} f_2(\boldsymbol{\Delta}, \boldsymbol{\Omega}_{22\cdot 1})$$

$$= f_1(\widehat{\boldsymbol{\theta}}) \times f_2(\hat{\boldsymbol{\Delta}}, \widehat{\boldsymbol{\Omega}}_{22\cdot 1}) = f(\widehat{\boldsymbol{\Omega}})$$

for $\widehat{\boldsymbol{\Omega}}$ in (3.12). Therefore, $\widehat{\boldsymbol{\Omega}}$ in (3.12) is the maximizer of $f(\boldsymbol{\Omega})$. □

Computing MLEs of Parameters under Complete Data

Suppose an $n \times t$ data matrix $\mathbf{Y}$ is modeled as in (1.1), with the distribution in (1.3), and suppose $\mathbf{C}'\boldsymbol{\Sigma}\mathbf{C} = \boldsymbol{\Omega}_{11}(\boldsymbol{\theta})$. Also, suppose all nt observations are available, so that $\mathbf{Y}$ is complete. To illustrate how MLEs are computed under various conditions, suppose interest is in testing the hypotheses

$$\mathrm{H}_o\colon \mathbf{BC} = \mathbf{0} \quad \text{against} \quad \mathrm{H}_a\colon \mathbf{BC} \neq \mathbf{0} \tag{3.13}$$

for the orthonormal matrix $\mathbf{C}$ in (1.5).

Computing MLEs under the Null Condition

Under the null condition in (3.13),

$$\mathbf{B}(\mathbf{I} - \mathbf{CC}') = \mathbf{B} \Rightarrow \mathbf{B} = \boldsymbol{\Xi}\mathbf{L}', \tag{3.14}$$

where $\mathbf{LL}' = (\mathbf{I} - \mathbf{CC}')$, and $\boldsymbol{\Xi} = \mathbf{BL}$. Thus, under the null condition, the model for $\mathbf{Y}$ in (1.1) can be written as

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\Xi}\mathbf{L}' + \mathbf{U}. \tag{3.15}$$

By applying Theorem (3.2), the covariance parameter $\boldsymbol{\Sigma}$ can be written as $\boldsymbol{\Sigma} = \mathbf{C}^*\boldsymbol{\Omega}\mathbf{C}^{*'}$, where $\boldsymbol{\Omega}$ is given in (3.3). Let $f_{\mathbf{Y}}(\ddot{\mathbf{Y}}|\boldsymbol{\Xi}, \mathbf{C}^*\boldsymbol{\Omega}\mathbf{C}^{*'})$ be the pdf for $\mathbf{Y}$, where $\ddot{\mathbf{Y}}$ is an observed value of $\mathbf{Y}$. Then

$$f_{\mathbf{Y}}(\ddot{\mathbf{Y}}|\boldsymbol{\Xi}, \mathbf{C}^*\boldsymbol{\Omega}\mathbf{C}^{*'}) = \frac{\exp\left\{-\frac{1}{2}\operatorname{tr}\left[(\ddot{\mathbf{Y}} - \mathbf{X}\boldsymbol{\Xi}\mathbf{L}')'(\ddot{\mathbf{Y}} - \mathbf{X}\boldsymbol{\Xi}\mathbf{L}')(\mathbf{C}^*\boldsymbol{\Omega}\mathbf{C}^{*'})^{-1}\right]\right\}}{2\pi^{nt/2}\left|\mathbf{C}^*\boldsymbol{\Omega}\mathbf{C}^{*'}\right|^{n/2}}$$

$$= \frac{\exp\left\{-\frac{1}{2}\,\mathrm{tr}\left[\mathbf{C}^{*'}(\ddot{\mathbf{Y}}-\mathbf{X}\Xi\mathbf{L}')'(\ddot{\mathbf{Y}}-\mathbf{X}\Xi\mathbf{L}')\mathbf{C}^{*}\Omega^{-1}\right]\right\}}{2\pi^{nt/2}|\Omega|^{n/2}} = f_{\mathbf{YC}^*}(\ddot{\mathbf{Y}}\mathbf{C}^*|\Xi,\Omega).$$

By results on conditional distributions given in [2], $f_{\mathbf{YC}^*}(\ddot{\mathbf{Y}}\mathbf{C}^*|\Xi,\Omega)$ can be factored as

$$f_{\mathbf{YC}^*}(\ddot{\mathbf{Y}}\mathbf{C}^*|\Xi,\Omega) = f_{\mathbf{YC}}(\ddot{\mathbf{Y}}\mathbf{C}|\mathbf{0},\Omega_{11}(\boldsymbol{\theta})) \times f_{\mathbf{YL}|\mathbf{YC}}(\ddot{\mathbf{Y}}\mathbf{L}|\mathbf{B}^*,\Omega_{22\cdot1});$$

where

$$\mathrm{vec}(\mathbf{YC}) \sim \mathrm{N}(\mathbf{0},\ \Omega_{11}(\boldsymbol{\theta})\otimes\mathbf{I}_n);$$

$$\mathrm{vec}(\mathbf{YL})|\,\mathrm{vec}(\mathbf{YC}) \sim \mathrm{N}(\mathrm{vec}(\mathbf{X}^*\mathbf{B}^*),\ \Omega_{22\cdot1}\otimes\mathbf{I}_n); \tag{3.16}$$

and

$$\mathbf{X}^* = (\mathbf{X}\ \ \mathbf{YC}),\ \mathbf{B}^* = \begin{pmatrix}\Xi\\ \boldsymbol{\Delta}\end{pmatrix},\ \boldsymbol{\Delta} = \Omega_{11}^{-1}(\boldsymbol{\theta})\Omega_{12},\quad \text{and} \tag{3.17}$$
$$\Omega_{22\cdot1} = \Omega_{22} - \Omega_{21}\Omega_{11}^{-1}(\boldsymbol{\theta})\Omega_{12}.$$

To compute the MLE's of Ω and Ξ:

1. Fit the model $\mathbf{YC} = \mathbf{0} + \mathbf{U}_1$, where

$$\mathrm{vec}(\mathbf{U}_1) \sim \mathrm{N}(\mathbf{0},\ \Omega_{11}(\boldsymbol{\theta})\otimes\mathbf{I}_n),$$

and compute the MLE of $\boldsymbol{\theta}$.

2. Fit the model $\mathbf{YL} = \mathbf{X}^*\mathbf{B}^* + \mathbf{U}^*$, where

$$\mathrm{vec}(\mathbf{U}^*) \sim \mathrm{N}(\mathbf{0},\ \Omega_{22\cdot1}\otimes\mathbf{I}_n),$$

and compute the MLEs of $\Omega_{22\cdot1}$ and $\mathbf{B}^*$:

$$\widehat{\Omega}_{22\cdot1} = \frac{\mathbf{L}'\mathbf{Y}'(\mathbf{I}-\mathbf{H}^*)\mathbf{YL}}{n},\quad \text{where}\quad \mathbf{H}^* = \mathbf{X}^*(\mathbf{X}^{*'}\mathbf{X}^*)^{-1}\mathbf{X}^{*'};$$

and

$$\widehat{\mathbf{B}}^* = \begin{pmatrix}\widehat{\Xi}\\ \widehat{\boldsymbol{\Delta}}\end{pmatrix} = (\mathbf{X}^{*'}\mathbf{X}^*)^{-1}\mathbf{X}^{*'}\mathbf{YL}.$$

Note that

$$\hat{\boldsymbol{\Delta}} = \boldsymbol{\Omega}_{11}^{-1}(\hat{\boldsymbol{\theta}})\hat{\boldsymbol{\Omega}}_{12} \Rightarrow \hat{\boldsymbol{\Omega}}_{12} = \boldsymbol{\Omega}_{11}(\hat{\boldsymbol{\theta}})\hat{\boldsymbol{\Delta}},$$

and

$$\hat{\boldsymbol{\Omega}}_{22\cdot1} = \hat{\boldsymbol{\Omega}}_{22} - \hat{\boldsymbol{\Omega}}_{21}\boldsymbol{\Omega}_{11}^{-1}(\hat{\boldsymbol{\theta}})\hat{\boldsymbol{\Omega}}_{12} \Rightarrow \hat{\boldsymbol{\Omega}}_{22} = \hat{\boldsymbol{\Omega}}_{22\cdot1} + \hat{\boldsymbol{\Omega}}_{21}\boldsymbol{\Omega}_{11}^{-1}(\hat{\boldsymbol{\theta}})\hat{\boldsymbol{\Omega}}_{12}.$$

Let

$$\hat{\boldsymbol{\Omega}} = \begin{pmatrix} \boldsymbol{\Omega}_{11}(\hat{\boldsymbol{\theta}}) & \hat{\boldsymbol{\Omega}}_{12} \\ \hat{\boldsymbol{\Omega}}_{21} & \hat{\boldsymbol{\Omega}}_{22} \end{pmatrix}, \tag{3.18}$$

then by the same approach used in the proof of Theorem (3.3), $\hat{\boldsymbol{\Omega}}$ in (3.18) is the MLE of $\boldsymbol{\Omega}$ under the model in (3.15). Thus, the maximized function $f_{\mathbf{YC}^*}(\ddot{\mathbf{Y}}\mathbf{C}^*|\hat{\boldsymbol{\Xi}}, \hat{\boldsymbol{\Omega}})$ can be expressed as

$$f_{\mathbf{YC}^*}(\ddot{\mathbf{Y}}\mathbf{C}^*|\hat{\boldsymbol{\Xi}}, \hat{\boldsymbol{\Omega}}) = f_{\mathbf{YC}}(\ddot{\mathbf{Y}}\mathbf{C}|\mathbf{0}, \boldsymbol{\Omega}_{11}(\hat{\boldsymbol{\theta}})) \times f_{\mathbf{YL}|\mathbf{YC}}(\ddot{\mathbf{Y}}\mathbf{L}|\hat{\boldsymbol{\Omega}}_{22\cdot1}, \hat{\mathbf{B}}^*). \tag{3.19}$$

<u>**Computing MLEs under the Non-Null Condition**</u>

Suppose $\mathbf{Y}$ is modeled as in (1.1), assuming the distribution in (1.3) with no restrictions other than $\mathbf{C}'\boldsymbol{\Sigma}\mathbf{C} = \boldsymbol{\Omega}_{11}(\boldsymbol{\theta})$. Then, by applying Theorem (3.2) and conditional distribution results, the pdf for $\mathbf{Y}$ can be written as

$$f_{\mathbf{Y}}(\ddot{\mathbf{Y}}|\mathbf{B}, \mathbf{C}^*\boldsymbol{\Omega}\mathbf{C}^{*'}) = f_{\mathbf{YC}^*}(\ddot{\mathbf{Y}}\mathbf{C}^*|\mathbf{B}, \boldsymbol{\Omega})$$

$$= f_{\mathbf{YC}}(\ddot{\mathbf{Y}}\mathbf{C}|\mathbf{BC}, \boldsymbol{\Omega}_{11}(\boldsymbol{\theta})) \times f_{\mathbf{YL}|\mathbf{YC}}(\ddot{\mathbf{Y}}\mathbf{L}|\mathbf{B}^{**}, \boldsymbol{\Omega}_{22\cdot1}); \tag{3.20}$$

where

$$\text{vec}(\mathbf{YC}) \sim \text{N}(\text{vec}(\mathbf{XBC}),\ \boldsymbol{\Omega}_{11}(\boldsymbol{\theta}) \otimes \mathbf{I}_n);$$

$$\text{vec}(\mathbf{YL})|\,\text{vec}(\mathbf{YC}) \sim \text{N}(\text{vec}(\mathbf{X}^*\mathbf{B}^{**}),\ \boldsymbol{\Omega}_{22\cdot1} \otimes \mathbf{I}_n); \tag{3.21}$$

$$\mathbf{B}^{**} = \begin{pmatrix} \mathbf{BL} - \mathbf{BC\Delta} \\ \mathbf{\Delta} \end{pmatrix}$$

and $\mathbf{X}^*$, $\mathbf{\Delta}$, and $\mathbf{\Omega}_{22\cdot 1}$ are given in (3.17).

To compute the MLEs of $\mathbf{\Omega}$, $\mathbf{BC}$, and $\mathbf{B}^{**}$:

1. Fit the model $\mathbf{YC} = \mathbf{XBC} + \mathbf{U}_1$, where

$$\text{vec}(\mathbf{U}_1) \sim \text{N}(\mathbf{0},\ \mathbf{\Omega}_{11}(\boldsymbol{\theta}) \otimes \mathbf{I}_n),$$

and compute the MLEs of $\mathbf{BC}$ and $\boldsymbol{\theta}$, say $\widehat{\mathbf{B}}\mathbf{C}$ and $\widehat{\boldsymbol{\theta}}$, where

$$\widehat{\mathbf{B}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$$

and $\widehat{\boldsymbol{\theta}}$ is the maximizer of the function $f_{\mathbf{YC}}(\ddot{\mathbf{Y}}\mathbf{C}|\widehat{\mathbf{B}}\mathbf{C}, \mathbf{\Omega}_{11}(\boldsymbol{\theta}))$.

2. Fit the model $\mathbf{YL} = \mathbf{X}^*\mathbf{B}^{**} + \mathbf{U}^*$, where

$$\text{vec}(\mathbf{U}^*) \sim \text{N}(\mathbf{0},\ \mathbf{\Omega}_{22\cdot 1} \otimes \mathbf{I}_n),$$

and compute the MLEs of $\mathbf{\Omega}_{22\cdot 1}$ and $\mathbf{B}^{**}$:

$$\widehat{\mathbf{\Omega}}_{22\cdot 1} = \frac{\mathbf{L}'\mathbf{Y}'(\mathbf{I} - \mathbf{H}^*)\mathbf{YL}}{n}, \quad \text{where} \quad \mathbf{H}^* = \mathbf{X}^*(\mathbf{X}^{*\prime}\mathbf{X}^*)^{-1}\mathbf{X}^{*\prime}$$

and

$$\widehat{\mathbf{B}}^{**} = (\mathbf{X}^{*\prime}\mathbf{X}^*)^{-1}\mathbf{X}^{*\prime}\mathbf{YL}.$$

Then, as was shown in the proof of Theorem (3.3), the maximized function $f_{\mathbf{YC}^*}(\ddot{\mathbf{Y}}\mathbf{C}^*|\widehat{\mathbf{B}}, \widehat{\mathbf{\Omega}})$ can be written as

$$f_{\mathbf{YC}^*}(\ddot{\mathbf{Y}}\mathbf{C}^*|\widehat{\mathbf{B}}, \widehat{\mathbf{\Omega}}) = f_{\mathbf{YC}}(\ddot{\mathbf{Y}}\mathbf{C}|\widehat{\mathbf{B}}\mathbf{C}, \mathbf{\Omega}_{11}(\widehat{\boldsymbol{\theta}})) \times f_{\mathbf{YL}|\mathbf{YC}}(\ddot{\mathbf{Y}}\mathbf{L}|\widehat{\mathbf{B}}^{**}, \widehat{\mathbf{\Omega}}_{22\cdot 1}). \qquad (3.22)$$

Theorem 3.4 *Under the model in (1.1) where* $\mathbf{C}'\mathbf{\Sigma}\mathbf{C} = \mathbf{\Omega}_{11}(\boldsymbol{\theta})$*, the LR test statistic for testing the hypotheses in (3.13) is*

$$L = \frac{\sup_{\boldsymbol{\theta}} g(\mathbf{A}_{11}^*|\boldsymbol{\theta})}{\sup_{\boldsymbol{\theta}} g(\mathbf{A}_{11}|\boldsymbol{\theta})}$$

where

$$g\left(\mathbf{A}_{11}|\boldsymbol{\theta}\right) = \frac{\exp\left((-1/2)\,\mathrm{tr}\left[\mathbf{A}_{11}\boldsymbol{\Omega}_{11}^{-1}(\boldsymbol{\theta})\right]\right)}{|\boldsymbol{\Omega}_{11}(\boldsymbol{\theta})|^{(n/2)}}$$

and $\boldsymbol{\Omega}_{11}(\boldsymbol{\theta})$ *is restricted to be positive definite.*

proof

By definition,

$$L = \frac{f_{\mathbf{YC}^*}(\ddot{\mathbf{Y}}\mathbf{C}^*|\widehat{\boldsymbol{\Xi}}, \widehat{\boldsymbol{\Omega}})}{f_{\mathbf{YC}^*}(\ddot{\mathbf{Y}}\mathbf{C}^*|\widehat{\mathbf{B}}, \widehat{\boldsymbol{\Omega}})},$$

and, using the equations in (3.19) and (3.22),

$$\frac{f_{\mathbf{YC}^*}(\ddot{\mathbf{Y}}\mathbf{C}^*|\widehat{\boldsymbol{\Xi}}, \widehat{\boldsymbol{\Omega}})}{f_{\mathbf{YC}^*}(\ddot{\mathbf{Y}}\mathbf{C}^*|\widehat{\mathbf{B}}, \widehat{\boldsymbol{\Omega}})} = \frac{f_{\mathbf{YC}}(\ddot{\mathbf{Y}}\mathbf{C}|\mathbf{0}, \boldsymbol{\Omega}_{11}(\widehat{\boldsymbol{\theta}})) \times f_{\mathbf{YL}|\mathbf{YC}}(\ddot{\mathbf{Y}}\mathbf{L}|\widehat{\mathbf{B}}^*, \widehat{\boldsymbol{\Omega}}_{22\cdot1})}{f_{\mathbf{YC}}(\ddot{\mathbf{Y}}\mathbf{C}|\widehat{\mathbf{B}}\mathbf{C}, \boldsymbol{\Omega}_{11}(\widehat{\boldsymbol{\theta}})) \times f_{\mathbf{YL}|\mathbf{YC}}(\ddot{\mathbf{Y}}\mathbf{L}|\widehat{\mathbf{B}}^{**}, \widehat{\boldsymbol{\Omega}}_{22\cdot1})}$$

$$= \frac{f_{\mathbf{YC}}(\ddot{\mathbf{Y}}\mathbf{C}|\mathbf{0}, \boldsymbol{\Omega}_{11}(\widehat{\boldsymbol{\theta}}))}{f_{\mathbf{YC}}(\ddot{\mathbf{Y}}\mathbf{C}|\widehat{\mathbf{B}}\mathbf{C}, \boldsymbol{\Omega}_{11}(\widehat{\boldsymbol{\theta}}))} \tag{3.23}$$

because $\widehat{\mathbf{B}}^* = \widehat{\mathbf{B}}^{**}$. The proof is completed by noting that

$$f_{\mathbf{YC}}(\ddot{\mathbf{Y}}\mathbf{C}|\mathbf{0}, \boldsymbol{\Omega}_{11}(\widehat{\boldsymbol{\theta}})) = \sup_{\boldsymbol{\theta}} \left(\frac{\exp\left((-1/2)\,\mathrm{tr}\left[\mathbf{C}'\mathbf{Y}'\mathbf{Y}\mathbf{C}\boldsymbol{\Omega}_{11}^{-1}(\boldsymbol{\theta})\right]\right)}{|\boldsymbol{\Omega}_{11}(\boldsymbol{\theta})|^{(n/2)}}\right)$$

$$= \sup_{\boldsymbol{\theta}} g\left(\mathbf{A}_{11}^*|\boldsymbol{\theta}\right),$$

where $\mathbf{A}_{11}^* = \mathbf{C}'\mathbf{Y}'\mathbf{Y}\mathbf{C}$; and

$$f_{\mathbf{YC}}(\ddot{\mathbf{Y}}\mathbf{C}|\widehat{\mathbf{B}}\mathbf{C}, \boldsymbol{\Omega}_{11}(\widehat{\boldsymbol{\theta}})) = \sup_{\boldsymbol{\theta}} \left(\frac{\exp\left((-1/2)\,\mathrm{tr}\left[\mathbf{C}'\mathbf{Y}'(\mathbf{I}-\mathbf{P})\mathbf{Y}\mathbf{C}\boldsymbol{\Omega}_{11}^{-1}(\boldsymbol{\theta})\right]\right)}{|\boldsymbol{\Omega}_{11}(\boldsymbol{\theta})|^{(n/2)}}\right)$$

$$= \sup_{\boldsymbol{\theta}} g\left(\mathbf{A}_{11}|\boldsymbol{\theta}\right),$$

where $\mathbf{A}_{11} = \mathbf{C}'\mathbf{Y}'(\mathbf{I}-\mathbf{P})\mathbf{Y}\mathbf{C}$.

Fisher Scoring Algorithm

If there is no simple, closed form solution for $\widehat{\boldsymbol{\theta}}$ then an iterative algorithm may be of use. One such algorithm is the Fisher scoring algorithm, which is described in [28]. Let $\boldsymbol{\theta}$ be an r-vector of unknown parameters. Denote by $L(\boldsymbol{\theta}|\mathbf{Y})$, the log likelihood function of $\boldsymbol{\theta}$ given an observed $\mathbf{Y}$. The Fisher scoring algorithm is based on an expansion of $L(\boldsymbol{\theta}|\mathbf{Y})$ in a Taylor series around an initial guess, say $\widehat{\boldsymbol{\theta}}$:

$$L(\boldsymbol{\theta}|\mathbf{Y}) = L(\widehat{\boldsymbol{\theta}}|\mathbf{Y}) + \left[\frac{\partial L(\boldsymbol{\theta}|\mathbf{Y})}{\partial \boldsymbol{\theta}'}\bigg|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}}\right](\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}})$$

$$+\frac{1}{2}(\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}})'\left[\frac{\partial^2 L(\boldsymbol{\theta}|\mathbf{Y})}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}'}\bigg|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}}\right](\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}) + o(\|\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}\|^2)$$

$$= L(\widehat{\boldsymbol{\theta}}|\mathbf{Y}) + \mathrm{g}'_{\widehat{\boldsymbol{\theta}}} + \frac{1}{2}(\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}})'\mathbf{H}_{\widehat{\boldsymbol{\theta}}}(\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}) + o(\|\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}\|^2),$$

where

$$\mathrm{g}_{\widehat{\boldsymbol{\theta}}} = \frac{\partial L(\boldsymbol{\theta}|\mathbf{Y})}{\partial \boldsymbol{\theta}}\bigg|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}} \quad \text{and} \quad \mathbf{H}_{\widehat{\boldsymbol{\theta}}} = \frac{\partial^2 L(\boldsymbol{\theta}|\mathbf{Y})}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}'}\bigg|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}}.$$

Let $\mathbf{I}(\widehat{\boldsymbol{\theta}})$ be the information at the value $\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}$:

$$\mathbf{I}(\widehat{\boldsymbol{\theta}}) = -\mathrm{E}\left[\frac{\partial^2 L(\boldsymbol{\theta}|\mathbf{Y})}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}'}\bigg|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}}\right].$$

Then, as noted in [32], the difference $-\mathbf{I}(\widehat{\boldsymbol{\theta}}) - \mathbf{H}_{\widehat{\boldsymbol{\theta}}}$ is $O(\frac{1}{n})$, which implies

$$L(\boldsymbol{\theta}|\mathbf{Y}) \approx L(\widehat{\boldsymbol{\theta}}|\mathbf{Y}) + \mathrm{g}'_{\widehat{\boldsymbol{\theta}}} - \frac{1}{2}(\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}})'\mathbf{I}(\widehat{\boldsymbol{\theta}})(\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}). \qquad (3.24)$$

Set the derivative of $L(\boldsymbol{\theta}|\mathbf{Y})$, expressed as (3.24), to zero and solve for $\boldsymbol{\theta}$. The result is

$$\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}} + \mathbf{I}^{-1}(\widehat{\boldsymbol{\theta}})\mathrm{g}_{\widehat{\boldsymbol{\theta}}}. \qquad (3.25)$$

The right-hand side in (3.25) becomes the new guess and the procedure is repeated. Under certain regularity conditions (see [28]) the Fisher scoring iterations converge to the unique maximum of $L(\boldsymbol{\theta}|\mathbf{Y})$.

Computing MLEs of Covariance Parameters under Incomplete Data

Suppose, as in the previous section, a $n \times t$ data matrix $\mathbf{Y}$ is normally distributed with mean $\mathbf{XB}$ and dispersion $\Sigma \otimes \mathbf{I}_n$, where $\mathbf{C}'\Sigma\mathbf{C} = \Omega_{11}(\boldsymbol{\theta})$. Then, by applying the result of Theorem (3.2), the dispersion of $\mathbf{Y}$ can be written as $\mathbf{C}^*\Omega\mathbf{C}^{*\prime}$, where Ω is given in (3.3). Suppose one or more rows of $\mathbf{Y}$ are partially observed, and that the missing data are missing completely at random (MCAR) as discussed in [27]. The observed variables can be considered a random sample from the matrix $\mathbf{Y}$. To find the MLEs $\widehat{\mathbf{B}}$ and $\widehat{\Omega}$ when $\mathbf{Y}$ is incomplete, the EM algorithm can be used. A detailed discussion of the EM algorithm is given in [14]. The following notation and theory concerning the EM algorithm is taken from [10].

Let $\mathbf{Y}$ be a random matrix and let $\ddot{\mathbf{Y}}$ be a realization of $\mathbf{Y}$. Denote the pdf of $\mathbf{Y}$ by $f(\mathbf{y}|\boldsymbol{\theta})$, where $\boldsymbol{\theta}$ is a vector of unknown parameters and $\mathbf{y} = \text{vec}(\mathbf{Y})$. The pdf can be considered as a random variable [i.e., $u = f(\mathbf{y}|\boldsymbol{\theta})$] or a realization of a random variable [i.e., $\ddot{u} = f(\ddot{\mathbf{y}}|\boldsymbol{\theta})$]. Denote the observable random variables by $\mathbf{y}_{\text{obs}}$ and the unobservable (i.e., missing) random variables as $\mathbf{y}_{\text{miss}}$. The vector $\ddot{\mathbf{y}}_{\text{obs}}$ is observed, while the vector $\ddot{\mathbf{y}}_{\text{miss}}$ is not observed.

The complete data pdf can be factored as

$$f(\mathbf{y}_{\text{obs}}, \mathbf{y}_{\text{miss}}|\boldsymbol{\theta}) = f(\mathbf{y}_{\text{obs}}|\boldsymbol{\theta}) \times f(\mathbf{y}_{\text{miss}}|\mathbf{y}_{\text{obs}}, \boldsymbol{\theta}),$$

which reveals that

$$f(\mathbf{y}_{\text{obs}}|\boldsymbol{\theta}) = \frac{f(\mathbf{y}_{\text{obs}}, \mathbf{y}_{\text{miss}}|\boldsymbol{\theta})}{f(\mathbf{y}_{\text{miss}}|\mathbf{y}_{\text{obs}}, \boldsymbol{\theta})}.$$

The corresponding factoring of the log likelihood function for $\mathbf{y}_{\text{obs}} = \ddot{\mathbf{y}}_{\text{obs}}$ is

$$L(\boldsymbol{\theta}|\ddot{\mathbf{y}}_{\text{obs}}) = L(\boldsymbol{\theta}|\ddot{\mathbf{y}}_{\text{obs}}, \mathbf{y}_{\text{miss}}) - \ln\left[f(\mathbf{y}_{\text{miss}}|\ddot{\mathbf{y}}_{\text{obs}}, \boldsymbol{\theta})\right].$$

To compute the MLE of $\boldsymbol{\theta}$, the observed log likelihood will be maximized with respect to $\boldsymbol{\theta}$ for fixed $\ddot{\mathbf{y}}_{\text{obs}}$.

Let $\boldsymbol{\theta}^{(t)}$ be the approximation to the MLE of $\boldsymbol{\theta}$ at the t^{th} iteration. To get rid of the random variables in the likelihood function, take expectations with respect to the density of $\mathbf{y}_{\text{miss}}$ conditional on $\mathbf{y}_{\text{obs}} = \ddot{\mathbf{y}}_{\text{obs}}$ and assuming $\boldsymbol{\theta}^{(t)}$ as the unknown parameter vector. That is,

$$\begin{aligned} \mathrm{E}\left[L(\boldsymbol{\theta}|\ddot{\mathbf{y}}_{\text{obs}})\right] &= L(\boldsymbol{\theta}|\ddot{\mathbf{y}}_{\text{obs}}) \\ &= Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)}) - H(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)}), \end{aligned}$$

where

$$Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)}) = \int L(\boldsymbol{\theta}|\ddot{\mathbf{y}}_{\text{obs}}, \ddot{\mathbf{y}}_{\text{miss}}) f(\ddot{\mathbf{y}}_{\text{miss}}|\ddot{\mathbf{y}}_{\text{obs}}, \boldsymbol{\theta}^{(t)})\, d\ddot{\mathbf{y}}_{\text{miss}},$$

and

$$H(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)}) = \int \ln\left[f(\ddot{\mathbf{y}}_{\text{miss}}|\ddot{\mathbf{y}}_{\text{obs}}, \boldsymbol{\theta})\right] f(\ddot{\mathbf{y}}_{\text{miss}}|\ddot{\mathbf{y}}_{\text{obs}}, \boldsymbol{\theta}^{(t)})\, d\ddot{\mathbf{y}}_{\text{miss}}.$$

The EM algorithm chooses $\boldsymbol{\theta}^{(t+1)}$ to maximize $Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)})$ with respect to $\boldsymbol{\theta}$. It can be shown that under some fairly general conditions, this strategy yields a sequence of estimates that converges to the MLE. Note that the E step is finding the expectation of the complete data log likelihood function, $L(\boldsymbol{\theta}|\mathbf{y})$, conditional on the observed data and the current estimate of $\boldsymbol{\theta}$. To evaluate the expectation, the conditional density of $\mathbf{y}_{\text{miss}}|\ddot{\mathbf{y}}_{\text{obs}}, \boldsymbol{\theta}^{(t)}$ must be known. The M step is to maximize the expected complete data log likelihood function with respect to $\boldsymbol{\theta}$. When $\mathbf{Y}$ is a normally distributed $n \times t$ random matrix with mean $\mathbf{XB}$ and dispersion $\mathbf{C}^*\boldsymbol{\Omega}\mathbf{C}^{*'} \otimes \mathbf{I}_n$, the MLEs of $\mathbf{B}$ and $\boldsymbol{\Omega}$ are found as follows.

Let $\mathbf{P}$ be the perpendicular projection operator $\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$. Let $\mathbf{W}$ be an $n \times t$ matrix of indicator variables:

$$w_{ij} = \begin{cases} 1 & \text{if the } ij^{\text{th}} \text{ entry in } \mathbf{Y} \text{ is observable,} \\ 0 & \text{if the } ij^{\text{th}} \text{ entry in } \mathbf{Y} \text{ is unobservable.} \end{cases} \tag{3.26}$$

Let $\mathbf{w} = \text{vec}(\mathbf{W})$ and let $\mathbf{D} = \text{diag}(\mathbf{w})$. Then,

$$\mathbf{y}_{\text{obs}} = \mathbf{Dy} \text{ and } \mathbf{y}_{\text{miss}} = (\mathbf{I}_{nt} - \mathbf{D})\mathbf{y}.$$

Note that the nt-vectors $\mathbf{y}_{\text{obs}}$ and $\mathbf{y}_{\text{miss}}$ have zeros replacing the unobservable and observable random variables, respectively. Denote the total number of missing observations by m. That is, $\text{tr}(\mathbf{D}) = nd - m$ and $\text{tr}(\mathbf{I}_{nt} - \mathbf{D}) = m$.

The marginal distribution of $\mathbf{y}_{\text{miss}}$ is

$$\mathbf{y}_{\text{miss}} \sim \text{N}\left[(\mathbf{I}_{nt} - \mathbf{D})(\mathbf{I}_t \otimes \mathbf{X})\boldsymbol{\beta}, (\mathbf{I}_{nt} - \mathbf{D})(\mathbf{C}^*\boldsymbol{\Omega}\mathbf{C}^{*\prime} \otimes \mathbf{I}_n)(\mathbf{I}_{nt} - \mathbf{D})\right].$$

Let $\mathbf{Y}_{\text{obs}}$, $\mathbf{Y}_{\text{miss}}$, $\mathbf{M}_{\text{obs}}$, and $\mathbf{M}_{\text{miss}}$ be the $n \times t$ matrices satisfying

$$\text{vec}(\mathbf{Y}_{\text{obs}}) = \mathbf{y}_{\text{obs}}, \text{vec}(\mathbf{Y}_{\text{miss}}) = \mathbf{y}_{\text{miss}}, \text{vec}(\mathbf{M}_{\text{obs}}) = \text{E}(\mathbf{y}_{\text{obs}}) = \mathbf{D}(\mathbf{I}_t \otimes \mathbf{X})\boldsymbol{\beta}, \text{ and}$$

$$\text{vec}(\mathbf{M}_{\text{miss}}) = \text{E}(\mathbf{y}_{\text{miss}}) = (\mathbf{I}_{nt} - \mathbf{D})(\mathbf{I}_t \otimes \mathbf{X})\boldsymbol{\beta}.$$

Note that

$$\mathbf{Y} = \mathbf{Y}_{\text{obs}} + \mathbf{Y}_{\text{miss}} \text{ and } \mathbf{M} = \mathbf{X}\mathbf{B} = \mathbf{M}_{\text{obs}} + \mathbf{M}_{\text{miss}}.$$

Denote the parameter estimates after the t^{th} iteration by $\boldsymbol{\beta}^{(t)}$ and $\mathbf{C}^*\boldsymbol{\Omega}^{(t)}\mathbf{C}^{*\prime}$, where $\boldsymbol{\beta}^{(t)} = \text{vec}(\mathbf{B}^{(t)})$.

<u>**Conditional Distribution of $\mathbf{Y}_{\text{miss}}$**</u>

The $nt \times nt$ matrices $\mathbf{D}$ and $\mathbf{I}_{nt} - \mathbf{D}$ are each symmetric and idempotent. Let $\mathbf{U}_1$ be an $nt \times (nt - m)$ matrix consisting of the non-zero columns of $\mathbf{D}$. Similarly, let $\mathbf{U}_2$ be an $nt \times m$ matrix consisting of the non-zero columns of $\mathbf{I}_{nt} - \mathbf{D}$. Then,

$$\mathbf{U} = \begin{pmatrix} \mathbf{U}_1 & \mathbf{U}_2 \end{pmatrix} \text{ is an orthogonal matrix,}$$

$$\mathbf{D} = \mathbf{U}_1\mathbf{U}_1', \quad \mathbf{I}_{nt} - \mathbf{D} = \mathbf{U}_2\mathbf{U}_2',$$

$$\mathbf{U}_1'\mathbf{U}_1 = \mathbf{I}_{nt-m}, \text{ and } \mathbf{U}_2'\mathbf{U}_2 = \mathbf{I}_m.$$

Note that

$$\mathbf{z}_1 = \mathbf{U}_1'\mathbf{y}$$

is an $(nt - m)$-vector of observable random variables and

$$\mathbf{z}_2 = \mathbf{U}_2'\mathbf{y}$$

is an m-vector of unobservable random variables. Also,

$$\mathbf{U}_1\mathbf{z}_1 = \mathbf{y}_{\text{obs}}, \quad \mathbf{U}_2\mathbf{z}_2 = \mathbf{y}_{\text{miss}},$$

$$\mathbf{U}_1'\mathbf{y}_{\text{obs}} = \mathbf{z}_1, \text{ and } \mathbf{U}_2'\mathbf{y}_{\text{miss}} = \mathbf{z}_2.$$

The joint density of $\mathbf{z}_1$ and $\mathbf{z}_2$ is

$$\begin{pmatrix}\mathbf{z}_1\\ \mathbf{z}_2\end{pmatrix} \sim \text{N}\left[\begin{pmatrix}\mathbf{U}_1'\,\text{vec}(\mathbf{XB})\\ \mathbf{U}_2'\,\text{vec}(\mathbf{XB})\end{pmatrix}, \begin{pmatrix}\mathbf{U}_1'(\mathbf{C}^*\boldsymbol{\Omega}\mathbf{C}^{*'}\otimes\mathbf{I}_n)\mathbf{U}_1 & \mathbf{U}_1'(\mathbf{C}^*\boldsymbol{\Omega}\mathbf{C}^{*'}\otimes\mathbf{I}_n)\mathbf{U}_2\\ \mathbf{U}_2'(\mathbf{C}^*\boldsymbol{\Omega}\mathbf{C}^{*'}\otimes\mathbf{I}_n)\mathbf{U}_1 & \mathbf{U}_2'(\mathbf{C}^*\boldsymbol{\Omega}\mathbf{C}^{*'}\otimes\mathbf{I}_n)\mathbf{U}_2\end{pmatrix}\right].$$

It follows that the conditional distribution of $\mathbf{z}_2$ given $\mathbf{z}_1 = \ddot{\mathbf{z}}_1$ and $\boldsymbol{\theta} = \boldsymbol{\theta}^{(t)}$ is normal with

$$\text{E}\left[\mathbf{z}_2|\ddot{\mathbf{z}}_1, \boldsymbol{\theta}^{(t)}\right] = \mathbf{U}_2'\,\text{vec}(\mathbf{XB}^{(t)})$$

$$+\ \mathbf{U}_2'(\mathbf{C}^*\boldsymbol{\Omega}^{(t)}\mathbf{C}^{*'}\otimes\mathbf{I}_n)\mathbf{U}_1\left[\mathbf{U}_1'(\mathbf{C}^*\boldsymbol{\Omega}^{(t)}\mathbf{C}^{*'}\otimes\mathbf{I}_n)\mathbf{U}_1\right]^{-1}\left[\ddot{\mathbf{z}}_1 - \mathbf{U}_1'\,\text{vec}(\mathbf{XB}^{(t)})\right],$$

and

$$\text{Var}\left[\mathbf{z}_2|\ddot{\mathbf{z}}_1, \boldsymbol{\theta}^{(t)}\right] = \mathbf{U}_2'(\mathbf{C}^*\boldsymbol{\Omega}^{(t)}\mathbf{C}^{*'}\otimes\mathbf{I}_n)\mathbf{U}_2$$

$$-\ \mathbf{U}_2'(\mathbf{C}^*\boldsymbol{\Omega}^{(t)}\mathbf{C}^{*'}\otimes\mathbf{I}_n)\mathbf{U}_1\left[\mathbf{U}_1'(\mathbf{C}^*\boldsymbol{\Omega}^{(t)}\mathbf{C}^{*'}\otimes\mathbf{I}_n)\mathbf{U}_1\right]^{-1}\mathbf{U}_1'(\mathbf{C}^*\boldsymbol{\Omega}^{(t)}\mathbf{C}^{*'}\otimes\mathbf{I}_n)\mathbf{U}_2.$$

Equivalently, the conditional distribution of $\mathbf{y}_{\text{miss}}$ given $\mathbf{y}_{\text{obs}} = \ddot{\mathbf{y}}_{\text{obs}}$ and $\boldsymbol{\theta} = \boldsymbol{\theta}^{(t)}$ is normal with

$$\text{E}\left[\mathbf{y}_{\text{miss}}|\ddot{\mathbf{y}}_{\text{obs}}, \boldsymbol{\theta}^{(t)}\right] = (\mathbf{I}_{nt} - \mathbf{D})\,\text{vec}(\mathbf{XB}^{(t)})$$

$$+\ (\mathbf{I}_{nt} - \mathbf{D})(\mathbf{C}^*\boldsymbol{\Omega}^{(t)}\mathbf{C}^{*'}\otimes\mathbf{I}_n)\mathbf{U}_1\left[\mathbf{U}_1'(\mathbf{C}^*\boldsymbol{\Omega}^{(t)}\mathbf{C}^{*'}\otimes\mathbf{I}_n)\mathbf{U}_1\right]^{-1}\mathbf{U}_1'\,\text{vec}(\ddot{\mathbf{Y}}_{\text{obs}} - \mathbf{XB}^{(t)}),$$

and

$$\text{Var}\left[\mathbf{y}_{\text{miss}}|\ddot{\mathbf{y}}_{\text{obs}}, \boldsymbol{\theta}^{(t)}\right] = (\mathbf{I}_{nt} - \mathbf{D})(\mathbf{C}^*\boldsymbol{\Omega}^{(t)}\mathbf{C}^{*'}\otimes\mathbf{I}_n)(\mathbf{I}_{nt} - \mathbf{D})$$

$$-\ (\mathbf{I}_{nt}-\mathbf{D})(\mathbf{C}^*\boldsymbol{\Omega}^{(t)}\mathbf{C}^{*'}\otimes\mathbf{I}_n)\mathbf{U}_1\left[\mathbf{U}_1'(\mathbf{C}^*\boldsymbol{\Omega}^{(t)}\mathbf{C}^{*'}\otimes\mathbf{I}_n)\mathbf{U}_1\right]^{-1}\mathbf{U}_1'(\mathbf{C}^*\boldsymbol{\Omega}^{(t)}\mathbf{C}^{*'}\otimes\mathbf{I}_n)(\mathbf{I}_{nt}-\mathbf{D}).$$

To compute the conditional expectation and variance of $\mathbf{y}_{\text{miss}}$, the preceding expression requires computation of an $(nt-m)\times(nt-m)$ inverse. By using the orthogonality of $\mathbf{U}$, an alternative expression can be obtained:

$$\mathbf{U}'\left(\mathbf{C}^*\boldsymbol{\Omega}^{(t)}\mathbf{C}^{*\prime}\otimes\mathbf{I}_n\right)\mathbf{U}=\left[\mathbf{U}'\left([\mathbf{C}^*\boldsymbol{\Omega}^{(t)}\mathbf{C}^{*\prime}]^{-1}\otimes\mathbf{I}_n\right)\mathbf{U}\right]^{-1}\Longrightarrow$$

$$\mathbf{U}_1'\left(\mathbf{C}^*\boldsymbol{\Omega}^{(t)}\mathbf{C}^{*\prime}\otimes\mathbf{I}_n\right)\mathbf{U}_1=\left[\mathbf{U}_1'\left([\mathbf{C}^*\boldsymbol{\Omega}^{(t)}\mathbf{C}^{*\prime}]^{-1}\otimes\mathbf{I}_n\right)\mathbf{U}_1\right.$$

$$-\mathbf{U}_1'\left([\mathbf{C}^*\boldsymbol{\Omega}^{(t)}\mathbf{C}^{*\prime}]^{-1}\otimes\mathbf{I}_n\right)\mathbf{U}_2\left(\mathbf{U}_2'\left([\mathbf{C}^*\boldsymbol{\Omega}^{(t)}\mathbf{C}^{*\prime}]^{-1}\otimes\mathbf{I}_n\right)\mathbf{U}_2\right)^{-1}\times$$
$$\left.\mathbf{U}_2'\left([\mathbf{C}^*\boldsymbol{\Omega}^{(t)}\mathbf{C}^{*\prime}]^{-1}\otimes\mathbf{I}_n\right)\mathbf{U}_1\right]^{-1}\Longrightarrow$$

$$\left[\mathbf{U}_1'\left(\mathbf{C}^*\boldsymbol{\Omega}^{(t)}\mathbf{C}^{*\prime}\otimes\mathbf{I}_n\right)\mathbf{U}_1\right]^{-1}=\mathbf{U}_1'\left([\mathbf{C}^*\boldsymbol{\Omega}^{(t)}\mathbf{C}^{*\prime}]^{-1}\otimes\mathbf{I}_n\right)\mathbf{U}_1$$

$$-\mathbf{U}_1'\left([\mathbf{C}^*\boldsymbol{\Omega}^{(t)}\mathbf{C}^{*\prime}]^{-1}\otimes\mathbf{I}_n\right)\mathbf{U}_2\left[\mathbf{U}_2'\left([\mathbf{C}^*\boldsymbol{\Omega}^{(t)}\mathbf{C}^{*\prime}]^{-1}\otimes\mathbf{I}_n\right)\mathbf{U}_2\right]^{-1}\times$$
$$\mathbf{U}_2'\left([\mathbf{C}^*\boldsymbol{\Omega}^{(t)}\mathbf{C}^{*\prime}]^{-1}\otimes\mathbf{I}_n\right)\mathbf{U}_1.$$

The above expressions require inverting $\mathbf{C}^*\boldsymbol{\Omega}^{(t)}\mathbf{C}^{*\prime}$, a $t\times t$ matrix, and $\mathbf{U}_2'\left([\mathbf{C}^*\boldsymbol{\Omega}^{(t)}\mathbf{C}^{*\prime}]^{-1}\otimes\mathbf{I}_n\right)\mathbf{U}_2$, an $m\times m$ matrix.

Denote the conditional expectation and variance of $\mathbf{Y}_{\text{miss}}$ by

$$\mathrm{E}(\mathbf{Y}_{\text{miss}}|\ddot{\mathbf{Y}}_{\text{obs}},\boldsymbol{\theta}^{(t)})=\mathbf{M}_{\text{miss}}^{(t)},\quad \mathrm{vec}(\mathbf{M}_{\text{miss}}^{(t)})=\boldsymbol{\mu}_{\text{miss}}^{(t)},$$

$$\text{and }\ \mathrm{Disp}(\mathbf{Y}_{\text{miss}}|\ddot{\mathbf{Y}}_{\text{obs}},\boldsymbol{\theta}^{(t)})=\mathbf{C}^*\boldsymbol{\Omega}_{\text{miss}}^{(t)}\mathbf{C}^{*\prime}.$$

Then, the conditional distribution of $\mathbf{y}_{\text{miss}}$ given $\mathbf{y}_{\text{obs}}=\ddot{\mathbf{y}}_{\text{obs}}$ and $\boldsymbol{\theta}=\boldsymbol{\theta}^{(t)}$ is normal with

$$\boldsymbol{\mu}_{\text{miss}}^{(t)}=(\mathbf{I}_{nt}-\mathbf{D})\,\mathrm{vec}(\mathbf{X}\mathbf{B}^{(t)}) \tag{3.27}$$
$$+\,(\mathbf{I}_{nt}-\mathbf{D})(\mathbf{C}^*\boldsymbol{\Omega}^{(t)}\mathbf{C}^{*\prime}\otimes\mathbf{I}_n)\mathbf{D}\left([\mathbf{C}^*\boldsymbol{\Omega}^{(t)}\mathbf{C}^{*\prime}]^{-1}\otimes\mathbf{I}_n\right)(\mathbf{I}_{nt}-\mathbf{P}_2)\mathbf{D}\,\mathrm{vec}(\ddot{\mathbf{Y}}_{\text{obs}}-\mathbf{X}\mathbf{B}^{(t)}),$$

and

$$\mathbf{C}^*\boldsymbol{\Omega}^{(t)}_{\text{miss}}\mathbf{C}^{*\prime} = \text{Disp}\left[\mathbf{Y}_{\text{miss}}|\ddot{\mathbf{Y}}_{\text{obs}}, \theta^{(t)}\right] = (\mathbf{I}_{nt} - \mathbf{D})(\mathbf{C}^*\boldsymbol{\Omega}^{(t)}\mathbf{C}^{*\prime} \otimes \mathbf{I}_n)(\mathbf{I}_{nt} - \mathbf{D})$$
$$- (\mathbf{I}_{nt} - \mathbf{D})(\mathbf{C}^*\boldsymbol{\Omega}^{(t)}\mathbf{C}^{*\prime} \otimes \mathbf{I}_n)\mathbf{D}\left(\left[\mathbf{C}^*\boldsymbol{\Omega}^{(t)}\mathbf{C}^{*\prime}\right]^{-1} \otimes \mathbf{I}_n\right)(\mathbf{I}_{nt} - \mathbf{P}_2)\mathbf{D}\times$$
$$(\mathbf{C}^*\boldsymbol{\Omega}^{(t)}\mathbf{C}^{*\prime} \otimes \mathbf{I}_n)(\mathbf{I}_{nt} - \mathbf{D}), \tag{3.28}$$

where $\mathbf{P}_2$ is the projection operator projecting onto $\Re(\mathbf{U}_2)$ along

$$\mathcal{N}\left[\mathbf{U}_2'\left(\left[\mathbf{C}^*\boldsymbol{\Omega}^{(t)}\mathbf{C}^{*\prime}\right]^{-1} \otimes \mathbf{I}_n\right)\right].$$

That is,

$$\mathbf{P}_2 = \mathbf{U}_2\left[\mathbf{U}_2'\left(\left[\mathbf{C}^*\boldsymbol{\Omega}^{(t)}\mathbf{C}^{*\prime}\right]^{-1} \otimes \mathbf{I}_n\right)\mathbf{U}_2\right]^{-1}\mathbf{U}_2'\left(\left[\mathbf{C}^*\boldsymbol{\Omega}^{(t)}\mathbf{C}^{*\prime}\right]^{-1} \otimes \mathbf{I}_n\right). \tag{3.29}$$

Note that

$$\mathbf{P}_2\mathbf{U}_2 = \mathbf{U}_2 \Rightarrow \mathbf{P}_2\mathbf{U}_2\mathbf{U}_2' = \mathbf{U}_2\mathbf{U}_2',$$

and

$$\mathbf{U}_2\mathbf{U}_2'\mathbf{P}_2 = \mathbf{P}_2 \quad \text{because} \quad \mathbf{U}_2'\mathbf{U}_2 = \mathbf{I}.$$

Therefore,

$$(\mathbf{I} - \mathbf{P}_2)(\mathbf{I} - \mathbf{U}_2\mathbf{U}_2') = (\mathbf{I} - \mathbf{P}_2). \tag{3.30}$$

Write

$$(\mathbf{I}_{nt} - \mathbf{D})(\mathbf{C}^*\boldsymbol{\Omega}^{(t)}\mathbf{C}^{*\prime} \otimes \mathbf{I}_n)\mathbf{D}\left(\left[\mathbf{C}^*\boldsymbol{\Omega}^{(t)}\mathbf{C}^{*\prime}\right]^{-1} \otimes \mathbf{I}_n\right)(\mathbf{I}_{nt} - \mathbf{P}_2)\mathbf{D}$$

as

$$\mathbf{U}_2\mathbf{U}_2'(\mathbf{C}^*\boldsymbol{\Omega}^{(t)}\mathbf{C}^{*\prime} \otimes \mathbf{I}_n)(\mathbf{I} - \mathbf{U}_2\mathbf{U}_2')\left(\left[\mathbf{C}^*\boldsymbol{\Omega}^{(t)}\mathbf{C}^{*\prime}\right]^{-1} \otimes \mathbf{I}_n\right)(\mathbf{I}_{nt} - \mathbf{P}_2)(\mathbf{I} - \mathbf{U}_2\mathbf{U}_2')$$
$$= \mathbf{U}_2\mathbf{U}_2'(\mathbf{C}^*\boldsymbol{\Omega}^{(t)}\mathbf{C}^{*\prime} \otimes \mathbf{I}_n)(\mathbf{I} - \mathbf{U}_2\mathbf{U}_2')\left(\left[\mathbf{C}^*\boldsymbol{\Omega}^{(t)}\mathbf{C}^{*\prime}\right]^{-1} \otimes \mathbf{I}_n\right)(\mathbf{I}_{nt} - \mathbf{P}_2) \tag{3.31}$$

using the result in (3.30). The expression in (3.31) can be written as

$$\begin{aligned}
&\mathbf{U}_2\mathbf{U}_2'(\mathbf{C}^*\boldsymbol{\Omega}^{(t)}\mathbf{C}^{*\prime}\otimes\mathbf{I}_n)(\mathbf{I}-\mathbf{U}_2\mathbf{U}_2')\left(\left[\mathbf{C}^*\boldsymbol{\Omega}^{(t)}\mathbf{C}^{*\prime}\right]^{-1}\otimes\mathbf{I}_n\right)(\mathbf{I}_{nt}-\mathbf{P}_2)\\
&=\mathbf{U}_2\mathbf{U}_2'(\mathbf{C}^*\boldsymbol{\Omega}^{(t)}\mathbf{C}^{*\prime}\otimes\mathbf{I}_n)(\mathbf{I}-\mathbf{U}_2\mathbf{U}_2')(\mathbf{I}_{nt}-\mathbf{P}_2')\left(\left[\mathbf{C}^*\boldsymbol{\Omega}^{(t)}\mathbf{C}^{*\prime}\right]^{-1}\otimes\mathbf{I}_n\right)
\end{aligned}$$

because $\left(\left[\mathbf{C}^*\boldsymbol{\Omega}^{(t)}\mathbf{C}^{*\prime}\right]^{-1}\otimes\mathbf{I}_n\right)(\mathbf{I}_{nt}-\mathbf{P}_2)$ is symmetric. Thus

$$\begin{aligned}
&(\mathbf{I}_{nt}-\mathbf{D})(\mathbf{C}^*\boldsymbol{\Omega}^{(t)}\mathbf{C}^{*\prime}\otimes\mathbf{I}_n)\mathbf{D}\left(\left[\mathbf{C}^*\boldsymbol{\Omega}^{(t)}\mathbf{C}^{*\prime}\right]^{-1}\otimes\mathbf{I}_n\right)(\mathbf{I}_{nt}-\mathbf{P}_2)\mathbf{D}\\
&=\mathbf{U}_2\mathbf{U}_2'(\mathbf{C}^*\boldsymbol{\Omega}^{(t)}\mathbf{C}^{*\prime}\otimes\mathbf{I}_n)(\mathbf{I}-\mathbf{U}_2\mathbf{U}_2')(\mathbf{I}_{nt}-\mathbf{P}_2')\left(\left[\mathbf{C}^*\boldsymbol{\Omega}^{(t)}\mathbf{C}^{*\prime}\right]^{-1}\otimes\mathbf{I}_n\right)\\
&=\mathbf{U}_2\mathbf{U}_2'(\mathbf{C}^*\boldsymbol{\Omega}^{(t)}\mathbf{C}^{*\prime}\otimes\mathbf{I}_n)(\mathbf{I}_{nt}-\mathbf{P}_2')\left(\left[\mathbf{C}^*\boldsymbol{\Omega}^{(t)}\mathbf{C}^{*\prime}\right]^{-1}\otimes\mathbf{I}_n\right)\\
&=\mathbf{U}_2\mathbf{U}_2'(\mathbf{C}^*\boldsymbol{\Omega}^{(t)}\mathbf{C}^{*\prime}\otimes\mathbf{I}_n)\left(\left[\mathbf{C}^*\boldsymbol{\Omega}^{(t)}\mathbf{C}^{*\prime}\right]^{-1}\otimes\mathbf{I}_n\right)(\mathbf{I}_{nt}-\mathbf{P}_2')\\
&=\mathbf{U}_2\mathbf{U}_2'(\mathbf{I}_{nt}-\mathbf{P}_2). \qquad (3.32)
\end{aligned}$$

Therefore, $\boldsymbol{\mu}^{(t)}_{\text{miss}}$ in (3.27) can be expressed as

$$\begin{aligned}
\boldsymbol{\mu}^{(t)}_{\text{miss}}&=\mathbf{U}_2\mathbf{U}_2'\,\text{vec}(\mathbf{X}\mathbf{B}^{(t)})+\mathbf{U}_2\mathbf{U}_2'(\mathbf{I}_{nt}-\mathbf{P}_2)\,\text{vec}(\ddot{\mathbf{Y}}_{\text{obs}}-\mathbf{X}\mathbf{B}^{(t)})\\
&=\mathbf{P}_2\,\text{vec}(\mathbf{X}\mathbf{B}^{(t)}-\ddot{\mathbf{Y}}_{\text{obs}})
\end{aligned}$$

because $\mathbf{U}_2'\,\text{vec}\,\ddot{\mathbf{Y}}_{\text{obs}}=\mathbf{0}$. Let the $t\times 1$ vector $\mathbf{v}_i$ be given by $\mathbf{v}_i=\mathbf{1}_t-\mathbf{w}_i$, where $\mathbf{w}_i'$ is the i^{th} row of $\mathbf{W}$ in (3.26). Let $\mathbf{D}_i=\text{diag}(\mathbf{v}_i)$ and let $\mathbf{V}_i$ be a $t\times m_i$ matrix consisting of the non-zero columns of $\mathbf{D}_i$, where m_i is the number of missing values for case i. Then $\mathbf{P}_2$ in (3.29) can be written as

$$\begin{aligned}
\mathbf{P}_2&=\textstyle\sum_{i=1}^{n}(\mathbf{P}_i\otimes\mathbf{e}_i\mathbf{e}_i'),\quad\text{where} \qquad (3.33)\\
\mathbf{P}_i&=\boldsymbol{\Sigma}^i_{22\cdot1}\left[\mathbf{C}^*\boldsymbol{\Omega}^{(t)}\mathbf{C}^{*\prime}\right]^{-1},\\
\boldsymbol{\Sigma}^i_{22\cdot1}&=\mathbf{V}_i\left(\mathbf{V}_i'\left[\mathbf{C}^*\boldsymbol{\Omega}^{(t)}\mathbf{C}^{*\prime}\right]^{-1}\mathbf{V}_i\right)^{-1}\mathbf{V}_i',
\end{aligned}$$

and $\mathbf{e}_i'$ is the i^{th} row of $\mathbf{I}_n$. Using the expression for $\mathbf{P}_2$ in (3.33), $\boldsymbol{\mu}^{(t)}_{\text{miss}}$ in (3.27) can be written as

$$\boldsymbol{\mu}^{(t)}_{\text{miss}} = \sum_{i=1}^{n}(\mathbf{P}_i\otimes\mathbf{e}_i\mathbf{e}_i')\,\text{vec}(\mathbf{X}\mathbf{B}^{(t)}-\ddot{\mathbf{Y}}_{\text{obs}}) \qquad (3.34)$$

$$\Rightarrow \mathbf{M}_{\text{miss}}^{(t)} = \sum_{i=1}^{n} \mathbf{e}_i \mathbf{e}_i' (\mathbf{X}\mathbf{B}^{(t)} - \ddot{\mathbf{Y}}_{\text{obs}}) \mathbf{P}_i'.$$

Also, using the results in (3.30), (3.32), and the orthonormality of $\mathbf{U}_2$, $\mathbf{C}^*\boldsymbol{\Omega}_{\text{miss}}^{(t)}\mathbf{C}^{*\prime}$ in (3.28) can be expressed as

$$\begin{aligned}\mathbf{C}^*\boldsymbol{\Omega}_{\text{miss}}^{(t)}\mathbf{C}^{*\prime} &= \mathbf{U}_2\left(\mathbf{U}_2'[\mathbf{C}^*\boldsymbol{\Omega}^{(t)}\mathbf{C}^{*\prime}]^{-1}\mathbf{U}_2\right)^{-1}\mathbf{U}_2' \\ &= \textstyle\sum_{i=1}^{n}(\boldsymbol{\Sigma}_{22\cdot 1}^{i} \otimes \mathbf{e}_i\mathbf{e}_i')\end{aligned}$$

for $\boldsymbol{\Sigma}_{22\cdot 1}^{i}$ in (3.33).

The Log Likelihood Function

The complete data log likelihood is

$$\begin{aligned}L(\boldsymbol{\theta}|\mathbf{y}) = &-\left(\tfrac{n}{2}\right)\ln\left(|\mathbf{C}^*\boldsymbol{\Omega}\mathbf{C}^{*\prime}|\right) - \left(\tfrac{nt}{2}\right)\ln(2\pi) \\ &-\left(\tfrac{1}{2}\right)\operatorname{tr}\left[(\mathbf{Y}-\mathbf{X}\mathbf{B})'(\mathbf{Y}-\mathbf{X}\mathbf{B})(\mathbf{C}^*\boldsymbol{\Omega}\mathbf{C}^{*\prime})^{-1}\right].\end{aligned}$$

Writing $\mathbf{Y}$ as $\mathbf{Y} = \mathbf{Y}_{\text{obs}} + \mathbf{Y}_{\text{miss}}$, the tr[ ] term in the likelihood function can be written as

$$\begin{aligned}\operatorname{tr}\left[(\mathbf{Y}-\mathbf{X}\mathbf{B})'(\mathbf{Y}-\mathbf{X}\mathbf{B})(\mathbf{C}^*\boldsymbol{\Omega}\mathbf{C}^{*\prime})^{-1}\right] &= \operatorname{tr}\left[(\mathbf{Y}_{\text{obs}}-\mathbf{X}\mathbf{B})'(\mathbf{Y}_{\text{obs}}-\mathbf{X}\mathbf{B})(\mathbf{C}^*\boldsymbol{\Omega}\mathbf{C}^{*\prime})^{-1}\right] \\ &+ 2\operatorname{tr}\left[\mathbf{Y}_{\text{miss}}'(\mathbf{Y}_{\text{obs}}-\mathbf{X}\mathbf{B})(\mathbf{C}^*\boldsymbol{\Omega}\mathbf{C}^{*\prime})^{-1}\right] + \mathbf{y}_{\text{miss}}'((\mathbf{C}^*\boldsymbol{\Omega}\mathbf{C}^{*\prime})^{-1}\otimes\mathbf{I}_n)\mathbf{y}_{\text{miss}}.\end{aligned}$$

E Step

The E step consists of evaluating the conditional expectation of $L(\boldsymbol{\theta}|\mathbf{Y})$ given $\boldsymbol{\theta} = \boldsymbol{\theta}^{(t)}$ and $\mathbf{Y}_{\text{obs}} = \ddot{\mathbf{Y}}_{\text{obs}}$. The result is

$$\begin{aligned}Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)}) = &-\left(\tfrac{n}{2}\right)\ln\left(|\mathbf{C}^*\boldsymbol{\Omega}\mathbf{C}^{*\prime}|\right) - \left(\tfrac{nt}{2}\right)\ln(2\pi) \\ &-\left(\tfrac{1}{2}\right)\Big\{\operatorname{tr}\left[(\ddot{\mathbf{Y}}_{\text{obs}}-\mathbf{X}\mathbf{B})'(\ddot{\mathbf{Y}}_{\text{obs}}-\mathbf{X}\mathbf{B})\mathbf{C}^*\boldsymbol{\Omega}\mathbf{C}^{*\prime -1}\right] \\ &+2\operatorname{tr}\left[(\ddot{\mathbf{Y}}_{\text{obs}}-\mathbf{X}\mathbf{B})'\mathbf{M}_{\text{miss}}^{(t)}(\mathbf{C}^*\boldsymbol{\Omega}\mathbf{C}^{*\prime})^{-1}\right] + \boldsymbol{\mu}_{\text{miss}}^{(t)\prime}((\mathbf{C}^*\boldsymbol{\Omega}\mathbf{C}^{*\prime})^{-1}\otimes\mathbf{I}_n)\boldsymbol{\mu}_{\text{miss}}^{(t)} \\ &+\operatorname{tr}\left[\mathbf{C}^*\boldsymbol{\Omega}_{\text{miss}}^{(t)}\mathbf{C}^{*\prime}((\mathbf{C}^*\boldsymbol{\Omega}\mathbf{C}^{*\prime})^{-1}\otimes\mathbf{I}_n)\right]\Big\}\end{aligned}$$

. Simplification yields

$$Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)}) = -\left(\tfrac{n}{2}\right)\ln\left(|\mathbf{C}^*\boldsymbol{\Omega}\mathbf{C}^{*\prime}|\right) - \left(\tfrac{nt}{2}\right)\ln(2\pi)$$
$$-\left(\tfrac{1}{2}\right)\mathrm{tr}\left[(\ddot{\mathbf{Y}}_{\text{obs}} + \mathbf{M}_{\text{miss}}^{(t)} - \mathbf{X}\mathbf{B})^{\prime}(\ddot{\mathbf{Y}}_{\text{obs}} + \mathbf{M}_{\text{miss}}^{(t)} - \mathbf{X}\mathbf{B})(\mathbf{C}^*\boldsymbol{\Omega}\mathbf{C}^{*\prime})^{-1}\right]$$
$$-\left(\tfrac{1}{2}\right)\mathrm{tr}\left[\mathrm{T}_t(\mathbf{C}^*\boldsymbol{\Omega}_{\text{miss}}^{(t)}\mathbf{C}^{*\prime})(\mathbf{C}^*\boldsymbol{\Omega}\mathbf{C}^{*\prime})^{-1}\right], \tag{3.35}$$

where T_t is the generalized trace operator of order t, which is discussed in [15]. Note that T_t simplifies to

$$\mathrm{T}_t = \sum_{i=1}^{n} \boldsymbol{\Sigma}_{22\cdot 1}^{i}. \tag{3.36}$$

<u>**M Step**</u>

The M step consists of maximizing the expected log likelihood with respect to $\boldsymbol{\beta}$ and $\boldsymbol{\Omega}$. Equating the derivative of Q with respect to $\boldsymbol{\beta}$ to zero yields

$$\frac{\partial Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)})}{\partial \boldsymbol{\beta}} = \mathbf{0} \Longrightarrow$$
$$-2((\mathbf{C}^*\boldsymbol{\Omega}\mathbf{C}^{*\prime})^{-1} \otimes \mathbf{X})(\ddot{\mathbf{y}}_{\text{obs}} + \boldsymbol{\mu}_{\text{miss}}^{(t)}) + 2((\mathbf{C}^*\boldsymbol{\Omega}\mathbf{C}^{*\prime})^{-1} \otimes \mathbf{X}^{\prime}\mathbf{X})\boldsymbol{\beta} = \mathbf{0} \Longrightarrow$$

$$\mathbf{B}^{(t+1)} = (\mathbf{X}^{\prime}\mathbf{X})^{-1}\mathbf{X}^{\prime}(\ddot{\mathbf{Y}}_{\text{obs}} + \mathbf{M}_{\text{miss}}^{(t)}). \tag{3.37}$$

Substituting $\mathbf{B}^{(t+1)}$ for $\mathbf{B}$ into the expected log likelihood yields

$$Q(\boldsymbol{\theta}|\mathbf{C}^*\boldsymbol{\Omega}^{(t)}\mathbf{C}^{*\prime}, \mathbf{B}^{(t+1)}) = -\left(\frac{n}{2}\right)\ln\left(|\mathbf{C}^*\boldsymbol{\Omega}\mathbf{C}^{*\prime}|\right) - \left(\frac{nt}{2}\right)\ln(2\pi)$$
$$-\left(\frac{1}{2}\right)\mathrm{tr}\left\{\left[(\ddot{\mathbf{Y}}_{\text{obs}} + \mathbf{M}_{\text{miss}}^{(t)})^{\prime}(\mathbf{I}_n - \mathbf{P})(\ddot{\mathbf{Y}}_{\text{obs}} + \mathbf{M}_{\text{miss}}^{(t)}) + \mathrm{T}_t(\mathbf{C}^*\boldsymbol{\Omega}_{\text{miss}}^{(t)}\mathbf{C}^{*\prime})\right](\mathbf{C}^*\boldsymbol{\Omega}\mathbf{C}^{*\prime})^{-1}\right\}$$
$$= -\left(\frac{n}{2}\right)\ln(|\boldsymbol{\Omega}|) - \left(\frac{nt}{2}\right)\ln(2\pi) - \left(\frac{1}{2}\right)\mathrm{tr}\left(\mathbf{A}\boldsymbol{\Omega}^{-1}\right), \tag{3.38}$$

where

$$\mathbf{A} = \mathbf{C}^{*\prime}\left[(\ddot{\mathbf{Y}}_{\text{obs}} + \mathbf{M}_{\text{miss}}^{(t)})^{\prime}(\mathbf{I}_n - \mathbf{P})(\ddot{\mathbf{Y}}_{\text{obs}} + \mathbf{M}_{\text{miss}}^{(t)}) + \mathrm{T}_t(\mathbf{C}^*\boldsymbol{\Omega}_{\text{miss}}^{(t)}\mathbf{C}^{*\prime})\right]\mathbf{C}^*. \tag{3.39}$$

To maximize equation (3.38) with respect to $\boldsymbol{\Omega}$, the results of Theorem (3.3) can be applied with

$$\mathbf{A}_{11} = \mathbf{C}' \left[(\ddot{\mathbf{Y}}_{\text{obs}} + \mathbf{M}^{(t)}_{\text{miss}})'(\mathbf{I}_n - \mathbf{P})(\ddot{\mathbf{Y}}_{\text{obs}} + \mathbf{M}^{(t)}_{\text{miss}}) + \mathrm{T}_t(\mathbf{C}^*\boldsymbol{\Omega}^{(t)}_{\text{miss}}\mathbf{C}^{*\prime}) \right] \mathbf{C}; \quad (3.40)$$

$$\mathbf{A}_{12} = \mathbf{C}' \left[(\ddot{\mathbf{Y}}_{\text{obs}} + \mathbf{M}^{(t)}_{\text{miss}})'(\mathbf{I}_n - \mathbf{P})(\ddot{\mathbf{Y}}_{\text{obs}} + \mathbf{M}^{(t)}_{\text{miss}}) + \mathrm{T}_t(\mathbf{C}^*\boldsymbol{\Omega}^{(t)}_{\text{miss}}\mathbf{C}^{*\prime}) \right] \mathbf{D}; \quad \text{and}$$

$$\mathbf{A}_{22} = \mathbf{D}' \left[(\ddot{\mathbf{Y}}_{\text{obs}} + \mathbf{M}^{(t)}_{\text{miss}})'(\mathbf{I}_n - \mathbf{P})(\ddot{\mathbf{Y}}_{\text{obs}} + \mathbf{M}^{(t)}_{\text{miss}}) + \mathrm{T}_t(\mathbf{C}^*\boldsymbol{\Omega}^{(t)}_{\text{miss}}\mathbf{C}^{*\prime}) \right] \mathbf{D}.$$

Also, under the modified mixed model in (2.4),

$$\boldsymbol{\Omega}_{11}(\boldsymbol{\theta}) = \boldsymbol{\Gamma}\boldsymbol{\Gamma}' + \sigma^2\mathbf{I}_q. \quad (3.41)$$

The Fisher scoring algorithm can be used to maximize the function in (3.7) and thus compute $\boldsymbol{\Omega}_{11}(\boldsymbol{\theta})^{(t+1)}$, while the regression equations in (3.8) can be used to find the remaining terms in $\boldsymbol{\Omega}^{(t+1)}$. The E and M steps are repeated until the sequence $\boldsymbol{\theta}^{(t)}, \boldsymbol{\theta}^{(t+1)}, \boldsymbol{\theta}^{(t+2)}, \cdots$ converges. After completion of the E step, if the function $Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)})$ in (3.35) is increased but not necessarily maximized, the sequence $\boldsymbol{\theta}^{(t)}, \boldsymbol{\theta}^{(t+1)}, \boldsymbol{\theta}^{(t+2)}, \cdots$ will still converge, and the resulting iterations are a generalized EM algorithm (GEM). For example, instead of maximizing the function in (3.38), perform only one or two iterations of the Fisher scoring algorithm so that $Q(\boldsymbol{\theta}|\mathbf{C}^*\boldsymbol{\Omega}^{(t)}\mathbf{C}^{*\prime}, \mathbf{B}^{(t+1)})$ is increased but not maximized with respect to $\boldsymbol{\Omega}^{(t+1)}$. The iterations will still converge to the MLEs, and the process may use less computer time.

CHAPTER 4

COMPARATIVE EVALUATION OF THREE INFERENCE METHODS

Introduction

In this Chapter the performance of a modified mixed model is evaluated and compared, by means of simulation, to the performance of a multivariate model and a factor analytic model. The first section reports the general model, and null and non-null conditions under which the simulations were run. A description of simulation conditions is also given in this section. The second section reports the procedures used to calculate test statistics and p-values when the covariance structure is factor analytic. The third section covers test statistics and p-values under a modified mixed model covariance structure. The fourth section reports simulation results concerning test size and power.

The Simulation Model

The simulations were run under the general model in (1.1) in which the between subjects design matrix is a $n \times 1$ vector of ones, $\mathbf{1}_n$. That is, the $n \times t$ matrix of observations $\mathbf{Y}$ was modeled as

$$\mathbf{Y} = \mathbf{1}_n \boldsymbol{\mu}' + \mathbf{U}, \tag{4.1}$$

where $\boldsymbol{\mu}$ is a $t \times 1$ vector of regression coefficients, and $\mathbf{U}$ is an $n \times t$ matrix of random errors whose distribution is given by

$$\text{vec}(\mathbf{U}) \sim \text{N}_{nt}(\mathbf{0},\ \boldsymbol{\Sigma} \otimes \mathbf{I}_n). \tag{4.2}$$

To shorten the computer time needed for simulations, the data matrix $\mathbf{Y}$ was always complete. Similar results for incomplete data are expected. Simulations were run with the covariance matrix Σ modeled with a factor analytic structure,

$$\Sigma = \Gamma\Gamma' + \sigma^2 \mathbf{I}_t, \tag{4.3}$$

where Γ is a $t \times f$ factor loading matrix, f held at 1 or 2. Simulations were also run under the modified mixed model structure, that is,

$$\mathbf{C}'\Sigma\mathbf{C} = \Gamma^*\Gamma^{*'} + \sigma^2 \mathbf{I}_q, \tag{4.4}$$

where Γ^* is a $q \times f^*$ matrix, f^* held at 1 or 2, and $\mathbf{C}$ a $t \times q$ matrix of coefficients for contrasts among the repeated measures. The number of repeated measures, t, was increased by increments of 2 from 3 to 7, that is,

$$t \in \{3, 5, 7\}.$$

The sample size, n, went from 5 to 200 in the following manner,

$$n \in \{5, 10, 15, 20, 25, 50, 100, 200\}.$$

For each generated data set, statistics were calculated to test

$$\text{H}_o\colon \boldsymbol{\mu}'\mathbf{C} = \mathbf{0} \quad \text{against} \quad \text{H}_a\colon \boldsymbol{\mu}'\mathbf{C} \neq \mathbf{0}, \tag{4.5}$$

where $\mathbf{C}$ is a $t \times (t-1)$ matrix of coefficients for contrasts among the repeated measures. Under the null and non-null conditions in (4.5), and for each n, t, and covariance structure combination, a 1000 data sets were generated. Note that for the simulation runs, $\boldsymbol{\Phi}$ in (1.5) was given the following structure; $\mathbf{F} = 1$, $\mathbf{B} = \boldsymbol{\mu}'$, and $\mathbf{C}$ a $t \times (t-1)$ matrix of coefficients for contrasts. The statistics calculated for each generated data set were the likelihood ratio (LR) statistic and Walds test statistic for both the factor analytic structure in (4.3) and the modified mixed model structure in (4.4); and the Lawley-Hotelling trace statistic.