UTS UNIVERSITY OF TECHNOLOGY SYDNEY

Likelihood Theory and Methods for Generalized Linear Mixed Models

by Aishwarya Bhaskaran

Thesis submitted in fulfilment of the requirements for the degree of

Doctor of Philosophy in Mathematics

under the supervision of **Prof. Matt P. Wand** and **Dr. Joanna Wang**

University of Technology Sydney
Faculty of Science

Submitted in September 2022

Certificate of original authorship

I, **Aishwarya Bhaskaran**, declare that this thesis, is submitted in fulfilment of the requirements for the award of Doctor of Philosophy in Mathematics, in the School of Mathematical and Physical Sciences at the University of Technology Sydney.

This thesis is wholly my own work unless otherwise referenced or acknowledged. In addition, I certify that all information sources and literature used are indicated in the thesis.

This document has not been submitted for qualifications at any other academic institution.

This research is supported by an Australian Government Research Training Program.

Signature:

Production Note:
Signature removed prior to publication.

Date: September 23, 2022

Acknowledgements

First and foremost, I would like to express my sincere gratitude to my supervisor Matt Wand. This journey was definitely a challenging one and I would like to thank you for all your patience, guidance and constant encouragement throughout the research process. I have indeed learnt a lot from your valuable insights and advice. Thank you so much for always looking out for my well-being as well. This milestone would not have been possible without your help and I am truly grateful.

Next, I would like to thank all my dear friends whom I have met throughout my life. Thank you all for constantly cheering me up, being such great listening ears and for always encouraging me to go the extra mile. I am really glad to have met each and every one of you.

Last but definitely not the least, I would like to give special thanks to my dear family. Thank you for always supporting me in my decisions and for offering me advice when I needed it. Thanks for also being so understanding especially these past three years and for always making sure that I am alright. I really appreciate everything that you all have done and will continue to do for me.

List of papers/publications

The following list of paper and publications awards relate to the work presented in this thesis:

- Jiang, J., Wand, M.P. and Bhaskaran, A. (2022), Usable and precise asymptotics for generalized linear mixed model analysis and design. Journal of the Royal Statistical Society, Series B, 84: 55-82. DOI: 10.1111/rssb.12473.
- Bhaskaran, A. and Wand, M.P.(2023), Dispersion parameter extension of precise generalized linear mixed model asymptotics. Statistics and Probability Letters, 193, Article 109691.

Notation

In this chapter, we introduce acronyms that are frequently used throughout this thesis.

Acronyms

Table 1: Table with acronyms used in the thesis with their meanings.

Acronym	**Meaning**
GLMM	Generalized linear mixed model
GVA	Gaussian variational approximation
MLE	Maximum likelihood estimation
TAP	Thouless-Anderson-Palmer

Contents

Abstract

Generalized linear mixed models are an essential group of models for analysing many present-day complex data sets, especially those that contain non-normal and correlated response data. Despite the large volume of research concerning this group of models, there is very little theory concerning the statistical properties of maximum likelihood estimators for generalized linear mixed models. Existing theoretical results available for the asymptotic variance-covariance matrix for such estimators contain limits and expectations over the response distribution, hence such results are not in ready-to-use forms when carrying out tasks such as constructing studentized confidence intervals or optimal design determination. In this thesis, we derive precise asymptotic results for likelihood-based generalized linear mixed model analysis. The novel asymptotic normality results are derived for both cases involving either a canonical or noncanonical link function. In our approach, we derive the exact leading term behaviour of the Fisher information matrix when both the number of groups and number of observations within each group diverge. This leads to asymptotic normality results with explicit and simple studentizable forms. The implications of these results in optimal design theory is also explored, leading to simpler and more direct determination of approximate locally D-optimal designs. Towards the end of this thesis, a Thouless-Anderson-Palmer approach is introduced for modern statistical inference for generalized linear mixed models. Such methods have proven to provide accurate approximations to problems arising in machine learning contexts. However, statistical applications such as generalized linear mixed model analysis have not been investigated. Thus, we derive results for implementing the Thouless-Anderson-Palmer frequentist variational approach to generalized linear mixed models and analyse the accuracy of its variational estimates.

Chapter 1

Introduction and Background

This chapter serves as an introduction for the thesis and also presents the necessary background information and theory required for the remaining chapters in this thesis. Note that the results, definitions and work presented in this chapter are not novel.

1.1 Introduction

In recent developments in statistical analysis, generalized linear mixed models (GLMMs) have become an essential group of models for analysing many present-day complex data sets, causing it to become a rapidly growing area of research. These models have been deemed useful and practical when accounting for overdispersion is necessary, a common occurrence when working with outcomes that have underlying Poisson or binomial distributions. In addition, GLMMs can also be applied widely in areas such as longitudinal data analysis and disease mapping (Breslow and Clayton, 1993).

A popular approach for fitting GLMMs is maximum likelihood estimation (MLE). Problems however arise as while the log-likelihood of a GLMM can be expressed mathematically, it involves integration over the random effects component of the model, which cannot be evaluated as closed form integrals. In some cases, to counter this hindrance caused by integral intractability and to accurately evaluate these integrals, standard quadrature techniques such as Gauss-Hermite quadrature can be used. Many software packages are also available to fit these GLMMs. One such example is the package `lme4` (Bates et al., 2015) in the R computing environment (R Core Team, 2022) and contains the function `glmer()` which is used to fit GLMMs.

While estimation by maximum likelihood for GLMMs is well and widely established, asymptotic normality results that can be used for practical purposes such as constructing confidence intervals and Wald tests via Studentization are currently unavailable in the existing generalized linear mixed model literature.

With regards to possible developments in methodology used for GLMMs, one may consider the Thouless-Anderson-Palmer (TAP) paradigm, which was developed in statistical physics literature for spin glass models (Thouless et al., 1977). Recently, there has been a realisation that it provides accurate approximations to problems arising in machine learning learning contexts. The TAP approach is also able to overcome issues involving intractable integrals, a common problem that hinders frequentist inference carried out on GLMMs. However, statistical applications such as longitudinal data analysis and multilevel models analysis, which may benefit from using TAP methodology, have not been investigated.

1.2 Thesis Aim

This thesis aims to address some of the gaps currently present in the immense statistical literature available for GLMMs in the frequentist setting. We firstly aim to develop novel asymptotic theory for maximum quasi-likelihood estimators for GLMMs. The goal is to develop such theory for both cases where one may choose to either use a canonical link or noncanonical link as part of the GLMM, dependent on the type of data being used. Following that, we aim to assess the efficacy of the confidence intervals constructed using the newly derived asymptotic normality distribution for maximum quasi-likelihood estimators for GLMMs. We also aim to explore applications for the asymptotic results, such as its implications in optimal design theory. Last but not least, we aim to develop theory for implementing the TAP frequentist variational approach to generalized linear mixed models and analyse the accuracy of its variational estimates.

1.3 Outline

Following the introductory chapter, Chapter 2 introduces lemmas that serve as essential statistical tools for carrying out the asymptotic derivations present in this thesis. Chapter 3 then starts off by dealing with the Gaussian generalized linear mixed model and presents a theorem concerning the joint asymptotic normality of all of the maximum

quasi-likelihood estimators for such a model. The theorem is then extended to the generalized linear mixed model case in Chapter 4. Consequences and applications of the novel asymptotic normality results are then investigated and discussed in Chapter 5. Chapter 6 then turns to newly developed methodology for GLMMs and develops theory for usage of the TAP variational method for GLMMs and investigates if there are improvements in the statistical approximations. Since only canonical links are considered in Chapters 4 and 6, Chapter 7 further extends the novel asymptotic results and TAP variational approximation derivations to cater for noncanonical links as well. The thesis ends off with a discussion and conclusion based on the work presented so far.

1.4 Matrix Theory

In this section, we present some background on the matrix theory required for the derivations present in this thesis.

1.4.1 Difference Between Two Matrix Inverses

For any two equal-sized invertible matrices $\boldsymbol{A}$ and $\boldsymbol{B}$, we have,

$$\boldsymbol{A}^{-1} - \boldsymbol{B}^{-1} = \boldsymbol{A}^{-1}(\boldsymbol{B} - \boldsymbol{A})\boldsymbol{B}^{-1}. \tag{1.1}$$

1.4.2 Other Useful Matrix Identities

Let $\boldsymbol{M}$ and $\boldsymbol{N}$ be invertible square matrices of the same size. Using an iterative application of the Sherman-Morrison-Woodbury formula, we have,

$$(\boldsymbol{M} - \boldsymbol{N})^{-1} = \sum_{k=0}^{\infty} (\boldsymbol{M}^{-1}\boldsymbol{N})^k \boldsymbol{M}^{-1}, \tag{1.2}$$

for matrices such that the spectral radius of $\boldsymbol{M}^{-1}\boldsymbol{N}$ is less than 1. Let $\boldsymbol{A}$ and $\boldsymbol{B}$ be invertible square matrices and $\boldsymbol{I}$ be an identity matrix with all the matrices having the same size. Using (1.2) and setting $\boldsymbol{M} = \boldsymbol{I}$ and $\boldsymbol{N} = -\boldsymbol{B}^{-1}\boldsymbol{A}^{-1}$ results in the following

matrix identity

$$\begin{aligned}(\boldsymbol{I}+\boldsymbol{AB})^{-1}\boldsymbol{A} &= \left[(\boldsymbol{AB})\{(\boldsymbol{AB})^{-1}+\boldsymbol{I}\}\right]^{-1}\boldsymbol{A} \\ &= (\boldsymbol{I}+\boldsymbol{B}^{-1}\boldsymbol{A}^{-1})^{-1}\boldsymbol{B}^{-1}\boldsymbol{A}^{-1}\boldsymbol{A} \\ &= (\boldsymbol{I}+\boldsymbol{B}^{-1}\boldsymbol{A}^{-1})^{-1}\boldsymbol{B}^{-1} \\ &= (\boldsymbol{I}+\boldsymbol{B}^{-1}\boldsymbol{A}^{-1}+\dots)\boldsymbol{B}^{-1}\end{aligned}. \tag{1.3}$$

Let $\boldsymbol{A}$, $\boldsymbol{B}$ and $\boldsymbol{C}$ be invertible square matrices and $\boldsymbol{I}$ be an identity matrix with all the matrices having the same size. Using (1.2) and setting $\boldsymbol{M}=\boldsymbol{I}$ and $\boldsymbol{N}=-\boldsymbol{B}^{-1}\boldsymbol{A}^{-1}$ results in the following matrix identity

$$\begin{aligned}(\boldsymbol{I}+\boldsymbol{AB})^{-1}\boldsymbol{C} &= \left[(\boldsymbol{AB})\{(\boldsymbol{AB})^{-1}+\boldsymbol{I}\}\right]^{-1}\boldsymbol{C} \\ &= (\boldsymbol{I}+\boldsymbol{B}^{-1}\boldsymbol{A}^{-1})^{-1}\boldsymbol{B}^{-1}\boldsymbol{A}^{-1}\boldsymbol{C} \\ &= (\boldsymbol{I}+\boldsymbol{B}^{-1}\boldsymbol{A}^{-1}+\dots)\boldsymbol{B}^{-1}\boldsymbol{A}^{-1}\boldsymbol{C}.\end{aligned} \tag{1.4}$$

Let $\boldsymbol{A}$ and $\boldsymbol{B}$ be invertible square matrices and $\boldsymbol{I}$ be an identity matrix with all the matrices having the same size. Using (1.2) and setting $\boldsymbol{M}=\boldsymbol{A}$ and $\boldsymbol{N}=-\boldsymbol{B}^{-1}$ results in the following matrix identity

$$\begin{aligned}\boldsymbol{B}(\boldsymbol{I}+\boldsymbol{AB})^{-1} &= (\boldsymbol{B}^{-1})^{-1}(\boldsymbol{I}+\boldsymbol{AB})^{-1} \\ &= \left\{(\boldsymbol{I}+\boldsymbol{AB})(\boldsymbol{B}^{-1})\right\}^{-1} \\ &= (\boldsymbol{B}^{-1}+\boldsymbol{A})^{-1} \\ &= \left\{\boldsymbol{A}-(-\boldsymbol{B}^{-1})\right\}^{-1} \\ &= \boldsymbol{A}^{-1}-\boldsymbol{A}^{-1}\boldsymbol{B}^{-1}\boldsymbol{A}^{-1}+\dots\end{aligned} \tag{1.5}$$

1.4.3 Block Matrix Inversion

The following definition instructs how a block matrix can be inverted.

Result 1. *Consider the following matrix which has been partitioned into four blocks*

$$\begin{bmatrix}\boldsymbol{A} & \boldsymbol{B} \\ \boldsymbol{C} & \boldsymbol{D}\end{bmatrix}$$

where $\boldsymbol{A}$ and $\boldsymbol{D}$ are square blocks of arbitrary size and blocks $\boldsymbol{B}$ and $\boldsymbol{C}$ are conformable such that the matrix can be properly partitioned. The matrix can then be inverted

blockwise as follows

$$\begin{bmatrix} \boldsymbol{A} & \boldsymbol{B} \\ \boldsymbol{C} & \boldsymbol{D} \end{bmatrix}^{-1} = \begin{bmatrix} \boldsymbol{A}^{-1} + \boldsymbol{A}^{-1}\boldsymbol{B}(\boldsymbol{D} - \boldsymbol{C}\boldsymbol{A}^{-1}\boldsymbol{B})^{-1}\boldsymbol{C}\boldsymbol{A}^{-1} & -\boldsymbol{A}^{-1}\boldsymbol{B}(\boldsymbol{D} - \boldsymbol{C}\boldsymbol{A}^{-1}\boldsymbol{B})^{-1} \\ -(\boldsymbol{D} - \boldsymbol{C}\boldsymbol{A}^{-1}\boldsymbol{B})^{-1}\boldsymbol{C}\boldsymbol{A}^{-1} & (\boldsymbol{D} - \boldsymbol{C}\boldsymbol{A}^{-1}\boldsymbol{B})^{-1} \end{bmatrix}. \tag{1.6}$$

By permuting the blocks, an equivalent result is

$$\begin{bmatrix} \boldsymbol{A} & \boldsymbol{B} \\ \boldsymbol{C} & \boldsymbol{D} \end{bmatrix}^{-1} = \begin{bmatrix} (\boldsymbol{A} - \boldsymbol{B}\boldsymbol{D}^{-1}\boldsymbol{C})^{-1} & -(\boldsymbol{A} - \boldsymbol{B}\boldsymbol{D}^{-1}\boldsymbol{C})^{-1}\boldsymbol{B}\boldsymbol{D}^{-1} \\ -\boldsymbol{D}^{-1}\boldsymbol{C}(\boldsymbol{A} - \boldsymbol{B}\boldsymbol{D}^{-1}\boldsymbol{C})^{-1} & \boldsymbol{D}^{-1} + \boldsymbol{D}^{-1}\boldsymbol{C}(\boldsymbol{A} - \boldsymbol{B}\boldsymbol{D}^{-1}\boldsymbol{C})^{-1}\boldsymbol{B}\boldsymbol{D}^{-1} \end{bmatrix}. \tag{1.7}$$

Note that matrices $\boldsymbol{A}$ and $\boldsymbol{D} - \boldsymbol{C}\boldsymbol{A}^{-1}\boldsymbol{B}$ must be invertible when using the block matrix inversion result presented in (1.6). Likewise, matrices $\boldsymbol{D}$ and $\boldsymbol{A} - \boldsymbol{B}\boldsymbol{D}^{-1}\boldsymbol{C}$ must be invertible when using the block matrix inversion result presented in (1.7).

1.4.4 The vec and vech Operators

Let $\boldsymbol{A}$ be a $m \times n$ matrix and let a_{ij} represent the element in the matrix located in the *i-th* row and *j-th* column. For any matrix $\boldsymbol{A}$, vec($\boldsymbol{A}$) is defined as the $mn \times 1$ vector which is constructed from the columns in $\boldsymbol{A}$ being stacked on top each other, one column after the other, from left to right. If $\boldsymbol{A}$ is a square $d \times d$ matrix, then vech($\boldsymbol{A}$) is defined as a $\frac{1}{2}d(d+1) \times 1$ vector, where the entries including and below the diagonal of $\boldsymbol{A}$, are stacked on top each other, one column after the other, from left to right.

For example, if $\boldsymbol{A}$ is a 3×3 square matrix, then we have the following matrix, vec and vech operators, where

$$\boldsymbol{A} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$$

with

$$\text{vec}(\boldsymbol{A}) = \begin{bmatrix} a_{11} \\ a_{22} \\ a_{31} \\ a_{12} \\ a_{22} \\ a_{32} \\ a_{13} \\ a_{23} \\ a_{33} \end{bmatrix} \quad \text{and} \quad \text{vech}(\boldsymbol{A}) = \begin{bmatrix} a_{11} \\ a_{21} \\ a_{31} \\ a_{22} \\ a_{32} \\ a_{33} \end{bmatrix}.$$

1.4.4.1 The Commutation Matrix

The *commutation matrix* of order d, denoted by $\boldsymbol{K}_d$, allows for the conversion between the operators $\text{vec}(\boldsymbol{A})$ and $\text{vec}(\boldsymbol{A}^T)$. It is the $d^2 \times d^2$ matrix containing only zeroes and ones such that

$$\boldsymbol{K}_d \text{vec}(\boldsymbol{A}) = \text{vec}(\boldsymbol{A}^T)$$

for all $d \times d$ matrices $\boldsymbol{A}$. The following useful property regarding commutation matrices also exists (Magnus and Neudecker, 1999):

$$\boldsymbol{K}_d^T = \boldsymbol{K}_d^{-1} = \boldsymbol{K}_d. \tag{1.8}$$

1.4.4.2 The Duplication Matrix

The *duplication matrix* of order d, denoted by $\boldsymbol{D}_d$, allows for the conversion between the operators $\text{vec}(\boldsymbol{A})$ and $\text{vech}(\boldsymbol{A})$. $\boldsymbol{D}_d$ is the unique $d^2 \times \frac{1}{2}d(d+1)$ matrix of zeros and ones such that

$$\text{vec}(\boldsymbol{A}) = \boldsymbol{D}_d \text{vech}(\boldsymbol{A})$$

for all $d \times d$ matrices $\boldsymbol{A}$. Finally, the Moore-Penrose inverse of $\boldsymbol{D}_d$ is defined as

$$\boldsymbol{D}_d^+ = (\boldsymbol{D}_d^T \boldsymbol{D}_d)^{-1} \boldsymbol{D}_d^T.$$

1.4.5 Kronecker Products and Related Properties

Let $\boldsymbol{A}$ be a $m \times n$ matrix and let $\boldsymbol{B}$ be a $p \times q$ matrix. The *Kronecker product* of matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ is denoted as $\boldsymbol{A} \otimes \boldsymbol{B}$ and it is the $mp \times nq$ matrix defined by

$$\begin{bmatrix} a_{11}\boldsymbol{B} & \dots & a_{1n}\boldsymbol{B} \\ \vdots & & \vdots \\ a_{m1}\boldsymbol{B} & \dots & a_{mn}\boldsymbol{B} \end{bmatrix}.$$

Let $\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}$ and $\boldsymbol{D}$ be square matrices. Then some of the properties regarding Kronecker products are as follows:

$$\begin{aligned} \boldsymbol{A} \otimes (\boldsymbol{B} + \boldsymbol{C}) &= \boldsymbol{A} \otimes \boldsymbol{B} + \boldsymbol{A} \otimes \boldsymbol{C}, \\ (\boldsymbol{A} \otimes \boldsymbol{B})(\boldsymbol{C} \otimes \boldsymbol{D}) &= \boldsymbol{AC} \otimes \boldsymbol{BD}, \\ (\boldsymbol{A} \otimes \boldsymbol{B})^{-1} &= \boldsymbol{A}^{-1} \otimes \boldsymbol{B}^{-1}, \\ \text{vec}(\boldsymbol{ABC}) &= (\boldsymbol{C}^T \otimes \boldsymbol{A})\text{vec}(\boldsymbol{B}) \quad \text{and} \\ \text{tr}(\boldsymbol{ABCD}) &= \text{vec}(\boldsymbol{D})^T(\boldsymbol{A} \otimes \boldsymbol{C}^T)\text{vec}(\boldsymbol{B}^T). \end{aligned}$$

1.4.6 Vector and Matrix Norms

In this subsection, we present a few vector and matrix norms and their properties.

1.4.6.1 Euclidean Norm

Let $\boldsymbol{x} \in \mathbb{R}^d$. Then, the *Euclidean Norm* of $\boldsymbol{x}$ is

$$\|\boldsymbol{x}\|_2 = \sqrt{\boldsymbol{x}^T\boldsymbol{x}}.$$

1.4.6.2 Frobenius Norm

The *Frobenius norm* of a general matrix $\boldsymbol{A}$ is

$$\|\boldsymbol{A}\|_F \equiv \sqrt{\text{trace}(\boldsymbol{A}^T\boldsymbol{A})}.$$

1.4.6.3 Spectral Norm

For a general symmetric matrix $\boldsymbol{M}$, let us define the following:

$$\lambda_{\min}(\boldsymbol{M}) \equiv \text{smallest eigenvalue of } \boldsymbol{M} \quad \text{and} \quad \lambda_{\max}(\boldsymbol{M}) \equiv \text{largest eigenvalue of } \boldsymbol{M}.$$

The *spectral norm* of a general matrix $\boldsymbol{A}$ is such that

$$\|\boldsymbol{A}\|_S \equiv \sqrt{\lambda_{\max}(\boldsymbol{A}^T\boldsymbol{A})}.$$

If $\boldsymbol{A}$ is symmetric then

$$\|\boldsymbol{A}\|_S = \lambda_{\max}(\boldsymbol{A}).$$

Also, if $\boldsymbol{A}$ is symmetric and positive definite, then the spectral decomposition of $\boldsymbol{A}$ is

$$\boldsymbol{A} = \boldsymbol{U}\text{diag}(\boldsymbol{\lambda})\boldsymbol{U}^T$$

where $\boldsymbol{U}^T\boldsymbol{U} = \boldsymbol{I}$ and $\boldsymbol{\lambda}$ is the vector containing the eigenvalues of $\boldsymbol{A}$. We then have

$$\boldsymbol{A}^{-1} = \boldsymbol{U}\text{diag}(\mathbf{1}/\boldsymbol{\lambda})\boldsymbol{U}^T$$

and therefore

$$\|\boldsymbol{A}^{-1}\|_S = 1/\lambda_{\min}(\boldsymbol{A}).$$

The spectral norm also possesses the following sub-multiplicity property

$$\|\boldsymbol{AB}\|_S \leq \|\boldsymbol{A}\|_S\|\boldsymbol{B}\|_S$$

for any pair of matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ such that the matrix product $\boldsymbol{AB}$ is defined.

Finally, suppose that $\boldsymbol{A}$ is a $d \times d$ matrix and $\mathbf{1}_d$ is the vector of ones. Then, using the sub-multiplicity property of the spectral norm, we can claim that

$$\|\mathbf{1}_d^T\boldsymbol{A}\mathbf{1}_d\|_S \leq \|\mathbf{1}_d^T\|_S\|\boldsymbol{A}\|_S\|\mathbf{1}_d\|_S.$$

Then

$$\|\mathbf{1}_d^T\|_S = \sqrt{\text{largest eigenvalue of } \mathbf{1}_d\mathbf{1}_d^T} = \sqrt{d}$$

and

$$\|\mathbf{1}_d\|_S = \sqrt{\text{largest eigenvalue of } \mathbf{1}_d^T\mathbf{1}_d} = \sqrt{d}.$$

Hence

$$\|\mathbf{1}_d^T \boldsymbol{A} \mathbf{1}_d\|_S \leq d\|\boldsymbol{A}\|_S.$$

1.4.7 Eigenvalue Bound Results

Let the eigenvalues of a $d \times d$ matrix $\boldsymbol{M}$ be denoted by

$$\lambda_1(\boldsymbol{M}), \ldots, \lambda_d(\boldsymbol{M}).$$

Theorem 8.1.5 of Golub and Van Loan (2013) states that, for any $d \times d$ matrices $\boldsymbol{A}$ and $\boldsymbol{E}$ such that $\boldsymbol{A}$ and $\boldsymbol{A} + \boldsymbol{E}$ are symmetric, we have

$$\lambda_j(\boldsymbol{A}) + \lambda_{\min}(\boldsymbol{E}) \leq \lambda_j(\boldsymbol{A} + \boldsymbol{E}) \leq \lambda_j(\boldsymbol{A}) + \lambda_{\max}(\boldsymbol{E}) \quad \text{for all} \quad 1 \leq j \leq d.$$

In particular, by choosing λ_j to correspond to $\lambda_{\min}$, we have

$$\lambda_{\min}(\boldsymbol{A} + \boldsymbol{E}) \geq \lambda_{\min}(\boldsymbol{A}) + \lambda_{\min}(\boldsymbol{E}). \tag{1.9}$$

1.4.7.1 Matrix Identities from Harville (1977)

For a general linear model

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}\boldsymbol{\alpha} + \boldsymbol{\varepsilon},$$

assume that $\boldsymbol{y}$ is a $n \times 1$ response vector, $\boldsymbol{X}$ and $\boldsymbol{Z}$ are $n \times p$ and $n \times q$ matrices respectively, $\boldsymbol{\beta}$ is a $p \times 1$ vector of unobservable fixed effects, $\boldsymbol{\alpha}$ is a $q \times 1$ vector of unobservable random effects and $\boldsymbol{\varepsilon}$ is a $n \times 1$ vector of unobservable random errors. In addition, the following properties apply where $E(\boldsymbol{\alpha}) = \mathbf{0}$, $E(\boldsymbol{\varepsilon}) = \mathbf{0}$ and $E(\boldsymbol{\alpha}^T\boldsymbol{\varepsilon}) = \mathbf{0}$. Also let $\boldsymbol{D} = \text{Cov}(\boldsymbol{\alpha}), \boldsymbol{R} = \text{Cov}(\boldsymbol{\varepsilon})$ and $\boldsymbol{V} = \boldsymbol{Z}\boldsymbol{D}\boldsymbol{Z}^T + \boldsymbol{R}$ such that $\text{Cov}(\boldsymbol{y}) = \boldsymbol{V}$, where Cov( ) denotes the covariance function. Harville (1977) then provides the following matrix identities.

$$\boldsymbol{V}^{-1} \equiv \boldsymbol{R}^{-1} - \boldsymbol{R}^{-1}\boldsymbol{Z}\boldsymbol{D}(\boldsymbol{I} + \boldsymbol{Z}^T\boldsymbol{R}^{-1}\boldsymbol{Z}\boldsymbol{D})^{-1}\boldsymbol{Z}^T\boldsymbol{R}^{-1}, \tag{1.10a}$$

$$\boldsymbol{Z}^T\boldsymbol{V}^{-1} \equiv (\boldsymbol{I} + \boldsymbol{Z}^T\boldsymbol{R}^{-1}\boldsymbol{Z}\boldsymbol{D})^{-1}\boldsymbol{Z}^T\boldsymbol{R}^{-1}. \tag{1.10b}$$

1.4.8 Vector Differential Calculus

Let f be a scalar-valued function with a vector $\boldsymbol{x} \in \mathbb{R}^d$ as its argument. Then the $1 \times d$ *derivative vector* of f, denoted by $\mathbf{D}f(\boldsymbol{x})$, has entries

$$\frac{\partial f(\boldsymbol{x})}{\partial x_i}$$

where $1 \leq i \leq d$. The $d \times d$ *Hessian matrix* of f is then defined as follows

$$\mathbf{H}f(\boldsymbol{x}) = \mathbf{D}\{\mathbf{D}f(\boldsymbol{x})^T\}$$

with (i, j) entry equal to

$$\frac{\partial^2 f(\boldsymbol{x})}{\partial x_i \partial x_j}$$

where $1 \leq i \leq d$ and $1 \leq j \leq d$. When vectors and matrices are involved, it is more appropriate to use vector differential calculus rather than ordinary scalar differential calculus. In order to compute the derivative vector, we use the following definition (Magnus and Neudecker, 1999):

Theorem 1. *First Identification Theorem: If* $\boldsymbol{a}$ *and* $\boldsymbol{x}$ *are* $1 \times d$ *vectors such that*

$$df(\boldsymbol{x}) = \boldsymbol{a}\, d\boldsymbol{x}$$

then

$$\boldsymbol{a} = \mathbf{D}f(\boldsymbol{x}).$$

The Hessian matrix can then be computed as follows:

Theorem 2. *Second Identification Theorem: If* $\boldsymbol{x}$ *is a* $1 \times d$ *vector and* $\boldsymbol{A}$ *is a* $d \times d$ *matrix such that*

$$d^2 f(\boldsymbol{x}) = (d\boldsymbol{x})^T\, \boldsymbol{A}\, d\boldsymbol{x}$$

then

$$\boldsymbol{A} = \mathbf{H}f(\boldsymbol{x}).$$

Further rules regarding vector differential calculus are provided in Magnus and Neudecker (1999) and Wand (2002).

1.5 Key Integral Results

In this section, we present key integral results used in this thesis.

1.5.1 Useful Integral Results

Let

$$a_1, a_2, a_3, b > 0$$

be strictly positive real numbers such that

$$a_1 a_2 > b.$$

Then, from Wolfram Research Inc. (2022), we have the following integral results:

$$\begin{aligned}
\int_0^\infty \frac{dx}{a_1 + x^2} &= \frac{\pi}{2\sqrt{a_1}}, \\
\int_0^\infty \frac{dx}{(a_1 + x^2)^2} &= \frac{\pi}{4a_1\sqrt{a_1}}, \\
\int_0^\infty \frac{dx}{(a_1 + x^2)(a_2 + x^2)} &= \frac{\pi}{2\sqrt{a_1 a_2}(\sqrt{a_1} + \sqrt{a_2})}, \\
\int_0^\infty \frac{dx}{(a_1 + x^2)^2(a_2 + x^2)} &= \frac{\pi(2\sqrt{a_1} + \sqrt{a_2})}{4a_1\sqrt{a_1 a_2}(\sqrt{a_1} + \sqrt{a_2})^2}, \\
\int_0^\infty \frac{x^2\, dx}{(a_1 + x^2)(a_2 + x^2)} &= \frac{\pi}{2(\sqrt{a_1} + \sqrt{a_2})}, \\
\int_0^\infty \frac{dx}{(a_1 + x^2)(a_2 + x^2)(a_3 + x^2)} &= \frac{\pi(\sqrt{a_1} + \sqrt{a_2} + \sqrt{a_3})}{2\sqrt{a_1 a_2 a_3}(\sqrt{a_1} + \sqrt{a_2})(\sqrt{a_1} + \sqrt{a_3})(\sqrt{a_2} + \sqrt{a_3})}
\end{aligned}$$

and

$$\begin{aligned}
&\int_0^\infty \frac{x^2\, dx}{(a_1 + x^2)(a_2 + x^2) - b} \\
&= \frac{\pi}{\sqrt{2}\left\{\sqrt{a_1 + a_2 + \sqrt{(a_1 - a_2)^2 + 4b}} + \sqrt{a_1 + a_2 - \sqrt{(a_1 - a_2)^2 + 4b}}\right\}}.
\end{aligned}$$

Also note that,

$$\frac{\pi(2\sqrt{a_1} + \sqrt{a_2})}{4a_1\sqrt{a_1 a_2}(\sqrt{a_1} + \sqrt{a_2})^2} < \frac{2\pi(\sqrt{a_1} + \sqrt{a_2})}{4a_1\sqrt{a_1 a_2}(\sqrt{a_1} + \sqrt{a_2})^2} = \frac{\pi}{2a_1\sqrt{a_1 a_2}(\sqrt{a_1} + \sqrt{a_2})}$$

which leads to the bound

$$\int_0^\infty \frac{dx}{(a_1 + x^2)^2(a_2 + x^2)} < \frac{\pi}{2a_1\sqrt{a_1 a_2}(\sqrt{a_1} + \sqrt{a_2})}. \tag{1.11}$$

1.5.2 Integral Form of the Matrix Square Root

The integral form of the square root of a matrix, for a matrix $\boldsymbol{A}$ having no eigenvalues on $\mathbb{R}^-$, is given by Higham (2008) as

$$\boldsymbol{A}^{1/2} = \frac{2}{\pi}\int_0^\infty \boldsymbol{A}(\boldsymbol{A} + t^2\boldsymbol{I})^{-1}\, dt. \tag{1.12}$$

1.6 Key Expectation Results

In this section, we present some key expectation results used frequently in the derivations in this thesis.

1.6.1 Law of Total Expectation

Theorem 3. *Let X and Y be random variables defined on the same probability space and where the expected value of X, $E(X)$, is defined. Then the law of total expectation is defined as follows*

$$E(X) = E\{E(X|Y)\}.$$

1.6.2 Jensen's Inequality

Theorem 4. *Let $f : \mathbb{R} \to \mathbb{R}$ be a convex function and X be a random variable such that its expected value, $E(X)$, is finite. Then, Jensen's inequality states the following*

$$f(E(X)) \leq E(f(X)).$$

1.6.3 Markov's Inequality

Theorem 5. *Let X be a random variable and a be a scalar. Markov's inequality states that if X is non-negative and $a > 0$, then the probability that X is greater than or equal to a is at most the expectation of X divided by a,*

$$P(X \geq a) \leq \frac{E(X)}{a}.$$

1.6.4 Cauchy-Schwarz Inequality

Theorem 6. *Let X and Y be random variables. Then the Cauchy-Schwarz inequality is defined as follows*

$$|E(XY)|^2 \leq E(X^2)E(Y^2).$$

1.7 Exponential Families

Exponential families consist of a set of probability distributions that can be written in certain parametric forms. These exponential family forms can be used to provide several parametric families of distributions with alternative parametrizations, in terms of natural parameters, which possess useful algebraic properties. Several commonly used probability distributions fall under the umbrella of exponential family density functions, including but not restricted to the normal, binomial, Poisson and gamma distributions.

1.7.1 One-Parameter Exponential Families

In this subsection, we present the class of *one-parameter exponential family* probability distributions.

Definition 1. *The class of one-parameter exponential family density, or probability mass, functions have generic form*

$$p(y;\eta) = \exp\{y\eta - b(\eta) + c(y)\}\, h(y) \tag{1.13}$$

where η is the natural parameter and the functions $b(\cdot), c(\cdot)$ and $h(\cdot)$ are defined according to the desired response distribution.

Explicit examples of the functions $b(\cdot), c(\cdot)$ and $h(\cdot)$ for the binomial and Poisson family of distributions have been provided in Table 1.1.

Family	$b(\eta)$	$c(y)$	$h(y)$
Binomial	$\log(1+e^{\eta})$	0	$I(y \in \{0,1\})$
Poisson	e^{η}	$-\log(y!)$	$I(y \in \{0\} \cup \mathbb{N})$

Table 1.1: Examples of one-parameter exponential families and their b, c and h functions.

Here, $I(\mathcal{P}) = 1$ if the condition $\mathcal{P}$ is true and $I(\mathcal{P}) = 0$ if $\mathcal{P}$ is false.

If the random variable Y has density, or probability mass, function as in (1.13), then $E(Y) = b'(\eta)$ and $\text{Var}(Y) = b''(\eta)$. A common modelling extension is to account for overdispersion. Overdispersion occurs when the variability present in the data is larger than what the proposed statistical model can account for. To model the variance flexibly, a dispersion parameter $\phi > 0$ is introduced and $\log\{p(y;\eta)\}$ is replaced by a *quasi-likelihood* function as shown in Definition 2.

Definition 2. *When accounting for overdispersion, the quasi-likelihood for the class of one-parameter exponential family density, or probability mass, functions have generic form*

$$\{y\eta - b(\eta) + c(y)\}/\phi + d(y,\phi) \tag{1.14}$$

where η is the natural parameter, $\phi > 0$ is the dispersion parameter and the functions $b(\cdot), c(\cdot)$ and $d(\cdot,\cdot)$ are defined according to the desired response distribution.

Now, if the random variable Y has quasi-likelihood as in (1.14), then $E(Y) = b'(\eta)$ and $\text{Var}(Y) = \phi b''(\eta)$.

Now, let $\boldsymbol{Y}$ be a vector of independent observations, $\boldsymbol{X}$ be a matrix of known covariates and $\boldsymbol{\beta}$ be a vector of unknown regression coefficients (fixed effects). In generalized linear models, it is common to model the mean function $\boldsymbol{\mu} = E(\boldsymbol{Y})$ as some non-linear function of the linear predictor or natural parameter such that

$$g(\boldsymbol{\mu}) = \boldsymbol{\eta}$$

where

$$\boldsymbol{\eta} = \boldsymbol{X}\boldsymbol{\beta}.$$

Here, g is known as the *link* function. If $g^{-1} = b'$, then g is called the *canonical link*

function and the following useful relationship exists

$$\boldsymbol{\mu} = g^{-1}(\boldsymbol{\eta}) = b'(\boldsymbol{\eta}).$$

Selecting the link function to be a canonical link leads to simpler likelihood expressions and provides useful sufficient statistics.

However, one may choose to use a *noncanonical link* function if, for example, using a noncanonical link leads to a better data fit. Some examples of noncanonical links are the probit link (Φ^{-1}) for binary regression and the log link (log) for Gamma regression. The following two definitions provide exponential family forms written in terms of the natural parameter, η, when noncanonical links are used (Fan et al., 1995).

Definition 3. *When using noncanonical links, the class of one-parameter exponential family density, or probability mass, functions have a generic form in terms of η as follows*

$$p(y;\eta) = \exp\left[y(g \circ b')^{-1}(\eta) - \{b \circ (g \circ b')^{-1}\}(\eta) + c(y)\right] h(y) \tag{1.15}$$

where η is the natural parameter, g is the link function and the functions $b(\cdot), c(\cdot)$ and $h(\cdot)$ are defined according to the desired response distribution.

If the random variable Y has density, or probability mass, function as in (1.15), then $E(Y) = g^{-1}(\eta)$ and $\text{Var}(Y) = \{b'' \circ (b')^{-1} \circ g^{-1}\}(\eta)$. As mentioned earlier in this section, one can apply a modelling extension to account for overdispersion which leads to the expression for $\log\{p(y;\eta)\}$ as in (1.15) being replaced by a quasi-likelihood function as shown in Definition 4.

Definition 4. *When accounting for overdispersion and using noncanonical links, the quasi-likelihood for the class of one-parameter exponential family density, or probability mass, functions have a generic form in terms of η as follows*

$$\left[y(g \circ b')^{-1}(\eta) - \{b \circ (g \circ b')^{-1}\}(\eta) + c(y)\right] / \phi + d(y, \phi) \tag{1.16}$$

where η is the natural parameter, $\phi > 0$ is the dispersion parameter and the functions $b(\cdot), c(\cdot)$ and $d(\cdot,\cdot)$ are defined according to the desired response distribution.

Now, if the random variable Y has quasi-likelihood as in (1.16), then $E(Y) = b'(\eta)$ and $\text{Var}(Y) = \phi\{b'' \circ (b')^{-1} \circ g^{-1}\}(\eta)$.

1.7.2 Two-Parameter Exponential Families

In this subsection, we present the class of *two-parameter exponential family* probability distributions.

Definition 5. *The class of two-parameter exponential family density, or probability mass, functions have generic form*

$$p(y;\eta,\phi) = \exp\left[\{y\eta - b(\eta) + c(y)\}/\phi + d(y,\phi)\right] h(y)$$

where η is the natural parameter and the functions $b(\cdot), c(\cdot), d(\cdot,\cdot)$ and $h(\cdot)$ are defined according to the desired response distribution.

Explicit examples of the functions $b(\cdot), c(\cdot), d(\cdot,\cdot)$ and $h(\cdot)$ for the Gaussian and gamma family of distributions have been provided in Table 1.2.

Family	$b(\eta)$	$c(y)$	$d(y,\phi)$	$h(y)$
Gaussian	$\frac{1}{2}\eta^2$	$-\frac{1}{2}y^2$	$-\log(2\pi\phi)$	1
Gamma	$-\log(-\eta)$	$\log(y)$	$-\log(\phi\Gamma(1/\phi)) - \log(y)$	$I(y>0)$

Table 1.2: Examples of two-parameter exponential families and their b, c, d and h functions.

1.8 Generalized Linear Mixed Models

We start off with classical linear models, where the mean of the response, often required to be normally distributed, can be expressed as a linear combination of the unknown model parameters and the predictor variables. In other words, a linear model has the mean

$$E(\boldsymbol{Y}) = \boldsymbol{X}\boldsymbol{\beta}$$

where $\boldsymbol{Y}$ is a vector of independent observations, $\boldsymbol{X}$ is a matrix of known covariates and $\boldsymbol{\beta}$ is a vector of unknown regression coefficients (fixed effects). These models, however, fall short when the observations are correlated or when the mean of the response cannot be written as a linear function of the covariates.

Generalized linear mixed models serve as an extension of linear models in two distinct ways. Firstly, GLMMs allow for the modelling of correlated data through the inclusion of random effects. Secondly, the mean, μ, is linked to the linear predictor through

a known function, g, known as the link function. With these additional properties, GLMMs have become an essential group of models for analysing many present-day complex data sets, which contain non-normal and correlated response data.

A summary of the various linear model effects structures ranging from linear models to generalized linear mixed models is further detailed in Table 1.3.

Model	(conditional) Mean Response
Linear Models (LMs)	$\boldsymbol{X\beta}$
Linear Mixed Models (LMMs)	$\boldsymbol{X\beta} + \boldsymbol{ZU}, \quad \boldsymbol{U} \sim (\boldsymbol{0}, \boldsymbol{\Sigma})$
Generalized Linear Models (GLMs)	$g^{-1}(\boldsymbol{X\beta})$
Generalized Linear Mixed Models (GLMMs)	$g^{-1}(\boldsymbol{X\beta} + \boldsymbol{ZU}), \quad \boldsymbol{U} \sim (\boldsymbol{0}, \boldsymbol{\Sigma})$

Table 1.3: Summary of various linear model effects structures.

To specify the structure of a generalized linear mixed model, we first define the conditional distribution of the response, Y_{ij}, given its associated random effect $\boldsymbol{U}_i$. Let there be m groups and n_i observations within each group. We also assume the random effects to be independent normally distributed variables and the response Y_{ij}, conditional on the random effects, to be from an exponential family f. Then the generalized linear mixed model has the following generic form:

$$Y_{ij}|\boldsymbol{U}_i \overset{\text{ind.}}{\sim} f_{Y_{ij}|\boldsymbol{U}_i}(y_{ij}|\boldsymbol{u}_i), \quad \boldsymbol{U}_i \overset{\text{ind.}}{\sim} N(\boldsymbol{0}, \boldsymbol{\Sigma}) \tag{1.17}$$

where $\overset{\text{ind.}}{\sim}$ means 'independently distributed as' and with natural parameter

$$\eta_{ij} = \boldsymbol{X}_{ij}^T\boldsymbol{\beta} + \boldsymbol{Z}_{ij}^T\boldsymbol{U}_i,$$

for $1 \leq i \leq m$ and $1 \leq j \leq n_i$. Here, $\boldsymbol{X}_{ij}$ is a $d_F \times 1$ vector of predictors having a fixed effects coefficient vector $\boldsymbol{\beta}$ and $\boldsymbol{Z}_{ij}$ is a $d_R \times 1$ vector of predictors having a $d_R \times 1$ random effects coefficient vector $\boldsymbol{U}_i$. For this generalized linear mixed model, the conditional mean of Y_{ij} is

$$E(Y_{ij}|\boldsymbol{U}_i) = \mu_{ij},$$

and there is a known link function, g, linking together the conditional mean and natural parameter such that

$$g(\mu_{ij}) = \eta_{ij} = \boldsymbol{X}_{ij}^T\boldsymbol{\beta} + \boldsymbol{Z}_{ij}^T\boldsymbol{U}_i.$$

Estimation of model parameters in generalized linear mixed models can be carried out using maximum likelihood estimation, with more details being provided in the next

section.

A detailed overview of the usefulness and difficulties of GLMM-based analysis can be found in **?** and Jiang and Nguyen (2021). Inferential methods for GLMMs other than the maximum likelihood approach, such as generalized estimating equations and penalized quasi-likelihood have also been discussed in these books.

1.9 Maximum Likelihood for Generalized Linear Mixed Models

In this section, the *maximum likelihood* approach to estimating model parameters in a generalized linear mixed model is presented.

1.9.1 The Likelihood Function

We begin by describing the *likelihood function* for a general statistical model. Consider a statistical model, parametrized by a vector of model parameters $\boldsymbol{\theta}$, with probability density function $f(\boldsymbol{y};\boldsymbol{\theta})$, where $\boldsymbol{y}$ is a vector of random variables. Here, if $\boldsymbol{\theta}$ is assumed to be known, then $f(\boldsymbol{y};\boldsymbol{\theta})$ is viewed to be the probability density function for $\boldsymbol{y}$. On the other hand, when $\boldsymbol{y}$ represents a vector of known observations and $\boldsymbol{\theta}$ is unknown, then $f(\boldsymbol{y})$ is simply a function of $\boldsymbol{\theta}$. This is known as the likelihood function of $\boldsymbol{y}$ and is usually represented as $\mathcal{L}(\boldsymbol{\theta};\boldsymbol{y})$ to emphasise that $\boldsymbol{\theta}$ is unknown and $\boldsymbol{y}$ is known. Note that mathematically, we have that

$$f(\boldsymbol{y};\boldsymbol{\theta}) = \mathcal{L}(\boldsymbol{\theta};\boldsymbol{y}).$$

Now let us consider the likelihood for a generic generalized linear mixed model. By letting $\boldsymbol{\theta} = (\boldsymbol{\beta}, \boldsymbol{\Sigma})$ and using the model description in (1.17), the likelihood can be written as follows

$$\mathcal{L}(\boldsymbol{\beta}, \boldsymbol{\Sigma}; \boldsymbol{y}) = \prod_{i=1}^{m} \int_{\mathbb{R}^{d_R}} \prod_{j=1}^{n_i} f_{Y_{ij}|\boldsymbol{U}_i}(y_{ij}|\boldsymbol{u}_i) f_{\boldsymbol{U}_i}(\boldsymbol{u}_i) d\boldsymbol{u}_i. \tag{1.18}$$

These likelihood functions form the basis for maximum likelihood estimation which is explained in the next subsection.

1.9.2 Maximum Likelihood Estimation

The maximum likelihood estimation method estimates the values of model parameters such that under the fitted statistical model, the observed data is most probable. In order to find these maximum likelihood estimates, one would need to find the values of $\boldsymbol{\theta}$ that maximise the likelihood $\mathcal{L}(\boldsymbol{\theta}; \boldsymbol{y})$, where the maximization is carried out within the permissible range of values for $\boldsymbol{\theta}$. For example, if one of the elements of $\boldsymbol{\theta}$ represents a variance or covariance parameter, then its range of permissible values is restricted to non-negative values. This aspect of maximum likelihood estimation is critical for estimating variances and covariances of random effects variables.

Note that finding the values of $\boldsymbol{\theta}$ that maximise the likelihood, $\mathcal{L}(\boldsymbol{\theta}; \boldsymbol{y})$, is equivalent to finding the values of $\boldsymbol{\theta}$ that maximise the *log-likelihood*, $\log \mathcal{L}(\boldsymbol{\theta}; \boldsymbol{y})$, since the log function is a monotonic increasing function. The log-likelihood, commonly denoted as $\ell(\boldsymbol{\theta})$, is often a more convenient mathematical expression to work with. Hence the maximum likelihood estimator for $\boldsymbol{\theta}^0$, the true value of the parameter $\boldsymbol{\theta}$, in a general statistical model can now be expressed as follows

$$\widehat{\boldsymbol{\theta}} = \underset{\boldsymbol{\theta}}{\operatorname{argmax}}\ \ell(\boldsymbol{\theta}). \tag{1.19}$$

Now, we aim to define the maximum likelihood estimators for generalized linear mixed models. Taking the log function on both sides of (1.18), we obtain the log-likelihood of a generic generalized linear mixed model as follows

$$\ell(\boldsymbol{\beta}, \boldsymbol{\Sigma}) = \sum_{i=1}^{m} \left(\log \int_{\mathbb{R}^{d_R}} \prod_{j=1}^{n_i} f_{Y_{ij}|\boldsymbol{U}_i}(y_{ij}|\boldsymbol{u}_i) f_{\boldsymbol{U}_i}(\boldsymbol{u}_i) d\boldsymbol{u}_i \right). \tag{1.20}$$

Then, for any $\boldsymbol{\beta}$ $(d_F \times 1)$ and $\boldsymbol{\Sigma}$ $(d_R \times d_R)$ that is symmetric and positive definite, the *maximum likelihood estimator* of $(\boldsymbol{\beta}^0, \boldsymbol{\Sigma}^0)$, the true values of the parameters $\boldsymbol{\beta}$ and $\boldsymbol{\Sigma}$, is

$$(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\Sigma}}) = \underset{\boldsymbol{\beta}, \boldsymbol{\Sigma}}{\operatorname{argmax}}\ \ell(\boldsymbol{\beta}, \boldsymbol{\Sigma}). \tag{1.21}$$

Note that *maximum quasi-likelihood estimators* for GLMMs can also be defined if one works with quasi-likelihoods such as those in (1.14) or (1.16).

1.9.3 Asymptotic Properties of Maximum Likelihood Estimators for Generalized Linear Mixed Models

Under certain regularity conditions, the maximum likelihood estimates of $\boldsymbol{\theta}^0$ obtained using (1.19) are consistent and asymptotically normally distributed according to the theorem below (Knight, 2000; **?**).

Theorem 7. *Let $\boldsymbol{Y}$ be a vector of independently and identically distributed random variables and let $\boldsymbol{\theta}$ denote the vector of model parameters used to parametrize a statistical model, such that the density function $f(\boldsymbol{y};\boldsymbol{\theta})$ satisfies the following regularity conditions:*

1. *The true value $\boldsymbol{\theta}^0$ of $\boldsymbol{\theta}$ is interior to the parameter space $\boldsymbol{\Theta}$, which has finite dimension and is compact;*
2. *The set $A = \{\boldsymbol{y} : f(\boldsymbol{y};\boldsymbol{\theta}) > \mathbf{0}\}$ does not depend on $\boldsymbol{\theta}$;*
3. *$f(\boldsymbol{y};\boldsymbol{\theta})$ is three times continuously differentiable with respect to $\boldsymbol{\theta}$ for all $\boldsymbol{y}$ in A;*
4. *$E[\ell'(\boldsymbol{\theta})] = \mathbf{0}$ for all $\boldsymbol{\theta}$ and $Var[\ell'(\boldsymbol{\theta})] = -E[\ell''(\boldsymbol{\theta})] = I(\boldsymbol{\theta})$ where $\mathbf{0} < I(\boldsymbol{\theta}) < \boldsymbol{\infty}$ for all $\boldsymbol{\theta}$;*
5. *For each $\boldsymbol{\theta}$ and $\boldsymbol{\delta} > \mathbf{0}$, there exists $|\ell'''(\boldsymbol{t};\boldsymbol{y})| < M(\boldsymbol{y})$ for $|\theta - \boldsymbol{t}| \leq \boldsymbol{\delta}$ where $E[M(\boldsymbol{Y})] < \boldsymbol{\infty}$.*

Then, the following asymptotic normality result for maximum likelihood estimators exists

$$\widehat{\boldsymbol{\theta}} \overset{asy.}{\sim} N(\boldsymbol{\theta}^0, I(\boldsymbol{\theta}^0)^{-1}) \tag{1.22}$$

where $\overset{asy.}{\sim}$ means 'asymptotically distributed as' and with mean equal to the vector of true model parameters $\boldsymbol{\theta}^0$ and asymptotic variance-covariance matrix equal to the inverse of the Fisher information matrix*, $I(\boldsymbol{\theta}^0)$.*

As stated in the theorem above, there are two ways to derive the Fisher information matrix. The first approach involves computing the first derivative of $\ell(\boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$. Then the Fisher information matrix can be defined as follows where

$$I(\boldsymbol{\theta}) = E\left\{\frac{\partial \ell}{\partial \boldsymbol{\theta}}\left(\frac{\partial \ell}{\partial \boldsymbol{\theta}}\right)^T\right\}.$$

Alternatively, one could compute the second derivative of $\ell(\boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$ and

use the following definition of the Fisher information matrix where

$$I(\boldsymbol{\theta}) = -E\left\{\frac{\partial^2 \ell}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T}\right\}.$$

Despite the large volume of research concerning generalized linear mixed models, there is very little theory concerning the statistical properties of maximum likelihood estimators or maximum quasi-likelihood estimators for these models.

Recent related literature published includes Hall et al. (2011), who derive precise asymptotic normality results for estimators for models that fall under the generalized linear mixed models framework. In their case, these results are derived for Gaussian variational approximation (GVA) estimators for a single-predictor Poisson mixed model. Let m be the number of groups or subjects and let n be the number of observations within each group. The asymptotic results are derived for the case where both m and n in the model diverge. The final results obtained give rise to asymptotically valid statistical inference where Gaussian variational approximations are concerned. In this case, the aim of this thesis differs as we aim to derive asymptotic normality results for maximum likelihood estimators and for the general class of GLMMs.

Nie (2007) presents properties of maximum likelihood estimators in generalized linear and non-linear mixed effects models. In this article, the convergence rates of the asymptotic variances of these maximum likelihood estimators were investigated in three cases. In terms of the notation presented in Section 1.8, one of these cases caters to when both m and n tend to infinity while the other two cases concentrate on either m or n diverging towards infinity while the other quantity remains finite. By using the Fisher information matrix, the convergence rates of the MLEs were determined by finding out the orders of the leading terms and the remainder terms in the asymptotic variances for each estimator. However, the leading terms involved were not explicitly derived. This leaves a gap in terms of deriving the asymptotic distributions explicitly for such GLMMs by investigating the exact expressions of the leading terms involved in the asymptotic variances of the estimators.

The work presented in this thesis addresses this gap in the current statistical literature for generalized linear mixed models.

1.10 Asymptotics

In this section, we present the statistical tools used to carry out the asymptotic derivations present in this thesis.

1.10.1 Convergence of Random Variables

In this subsection, we look at two different types of convergence for sequences of random variables. We also consider the properties of these sequences of random variables when various algebraic operations are applied.

1.10.1.1 Convergence in Probability

The notion of *convergence in probability* for a sequence of random variables deals with the convergence of the random variables themselves and is defined below,

Definition 6. *Let $\{X_n\}$ be a sequence of random variables. Then $\{X_n\}$ converges in probability to the random variable X as $n \to \infty$, or $X_n \xrightarrow{P} X$, if for all $\varepsilon > 0$,*

$$\lim_{n\to\infty} P(|X_n - X| > \varepsilon) = 0.$$

It is common for the limiting random variable X to be a constant c, for which we then have $X_n \xrightarrow{P} c$.

1.10.1.2 Convergence in Distribution

The notion of *convergence in distribution* for a sequence of random variables deals with the convergence of the distribution functions of the random variables and is defined below,

Definition 7. *Let $\{X_n\}$ be a sequence of random variables. Then $\{X_n\}$ converges in distribution to the random variable X as $n \to \infty$, or $X_n \xrightarrow{\mathcal{D}} X$, if*

$$\lim_{n\to\infty} P(X_n \leq x) = P(X_n \leq x) = F(x)$$

for each point $x \in \mathbb{R}$ at which $F(x)$ is continuous, where $F(x)$ is the cumulative distribution function of the random variable X.

1.10.1.3 Continuous Mapping Theorem

In probability theory, the *continuous mapping theorem* states that continuous functions of sequences of random variables preserve limits. A formal definition is provided below.

Theorem 8. *Let X be a random variable and $\{X_n\}$ be a sequence of random variables. If g is a continuous function, then the continuous mapping theorem states the following*

$$X_n \xrightarrow{\mathcal{D}} X \quad \textit{implies} \quad g(X_n) \xrightarrow{\mathcal{D}} g(X) \quad \textit{and} \quad X_n \xrightarrow{P} X \quad \textit{implies} \quad g(X_n) \xrightarrow{P} g(X).$$

1.10.1.4 Slutsky's Theorem

Slutsky's theorem, which is partly derived using the continuous mapping theorem, provides useful results when dealing with algebraic operations involving two sequences of random variables, where one sequence converges in distribution to a random variable while the other sequence converges in probability to a constant.

Theorem 9. *Let both $\{X_n\}$ and $\{Y_n\}$ be sequences of random variables. If $X_n \xrightarrow{\mathcal{D}} X$ and $Y_n \xrightarrow{P} c$, then the following properties of algebraic operations involving both $\{X_n\}$ and $\{Y_n\}$ exist*

$$X_n + Y_n \xrightarrow{\mathcal{D}} X + c, \quad X_n Y_n \xrightarrow{\mathcal{D}} Xc \quad \textit{and} \quad X_n / Y_n \xrightarrow{\mathcal{D}} X/c.$$

1.10.1.5 Cramér-Wold Device

The *Cramér-Wold device* is a useful result that can be used to prove the joint convergence of random variables. A formal definition is provided below.

Theorem 10. *Let $\boldsymbol{X}$ be a random variable and $\{\boldsymbol{X}_n\}$ be a sequence of random variables, where $\boldsymbol{X}, \boldsymbol{X}_n \in \mathbb{R}^d$. If $\boldsymbol{a}^T \boldsymbol{X}_n \xrightarrow{\mathcal{D}} \boldsymbol{a}^T \boldsymbol{X}$ for all $\boldsymbol{a} \in \mathbb{R}^d$, then $\boldsymbol{X}_n \xrightarrow{\mathcal{D}} \boldsymbol{X}$.*

1.10.2 Stochastic Order Notation

It is convenient to have notation that represent sequences of random variables that converge in probability to zero or sequences of random variables that are bounded in probability (van der Vaart, 1998). We make use of stochastic order notation such as $o_P(1)$ and $O_P(1)$ for such purposes. The notation $o_P(1)$, used for the notion of convergence in probability to zero, is formally defined below

Definition 8. *Let $\{X_n\}$ be a sequence of random variables. It is convenient to write $X_n = o_P(1)$ to represent that X_n converges in probability to zero, or $X_n \xrightarrow{P} 0$, if for every $\varepsilon > 0$ we have,*

$$P(|X_n| > \varepsilon) \to 0, \quad \text{as } n \to \infty.$$

Sequences that are bounded in probability can be represented using the $O_P(1)$ notation, which is formally defined below

Definition 9. *Let $\{X_n\}$ be a sequence of random variables. It is convenient to write $X_n = O_P(1)$ to represent that X_n is bounded in probability, if for every $\varepsilon > 0$, there exists $M_\varepsilon > 0$ such that,*

$$P(|X_n| > M_\varepsilon) < \varepsilon, \quad \text{for all } n.$$

Note that for a sequence of random variables $\{X_n\}$, if $X_n = o_P(1)$, then $X_n = O_P(1)$ as well. Using Definitions 8 and 9, we also have the following general results, where for sequences of random variables $\{X_n\}$, $\{Y_n\}$ and $\{R_n\}$, we have,

$$X_n = o_P(R_n) \quad \text{if and only if} \quad X_n = Y_n R_n \quad \text{and} \quad Y_n = o_P(1).$$

and

$$X_n = O_P(R_n) \quad \text{if and only if} \quad X_n = Y_n R_n \quad \text{and} \quad Y_n = O_P(1).$$

Lastly, there are useful rules of calculus concerning o_P and O_P symbols. Some of

these rules are presented here:

$$o_P(1) + o_P(1) = o_P(1),$$
$$o_P(1) + O_P(1) = O_P(1),$$
$$o_P(1)O_P(1) = o_P(1).$$

Let $\{a_n\}$ and $\{b_n\}$ be sequences of positive real numbers. Then we also have the following rules:

$$o_P(a_n)o_P(b_n) = o_P(a_n b_n),$$
$$o_P(a_n)O_P(b_n) = o_P(a_n b_n),$$
$$o_P(a_n) + o_P(b_n) = o_P(\max\{a_n, b_n\}).$$

1.10.3 Other Tools for Working with Asymptotic Expansions

The *stochastic Taylor formula*, *inversion formula for an asymptotic series* and the *Laplace expansion for evaluating an integral* are useful tools for working with asymptotic approximations and expansions. We will highlight and present the inversion formula for an asymptotic series while the details regarding the other tools can be found in Pace and Salvan (1997).

1.10.3.1 Inversion of Asymptotic Series

In this subsubsection, we present an approach for inverting a univariate asymptotic series. Details regarding the derivation of the inversion formula for both the univariate and multivariate cases can be found in Pace and Salvan (1997).

Let $y = f(x), x \in \mathbb{R}$ be a real smooth function which admits the following power series expansion

$$y = x + a_1 x^2 + a_2 x^3 + \dots. \tag{1.23}$$

Assume that the terms in (1.23) depend on an asymptotic parameter n. Specifically, let $x = O(n^{-\alpha})$, $\alpha > 0$ and let $a_i = O(1), i = 1, 2, \dots$. Suppose that we invert the function $y = f(x)$ as $x = g(y)$ and wished to express $g(y)$, in the neighbourhood of $y = 0$, as a power series expansion as follows

$$x = y + b_1 x^2 + b_2 x^3 + O(n^{-4\alpha}), \tag{1.24}$$

where b_1 and b_2 can be expressed in terms of constants a_1 and a_2. Ignoring terms of

order $O(n^{-4\alpha})$, the final expressions for b_1 and b_2 are as follows

$$b_1 = -a_1, \quad b_2 = -(a_2 - 2a_1^2).$$

1.11 Frequentist Variational Approximations

Variational approximations have roots in variational calculus and serve as an approach for performing approximate inference on model parameters in complex statistical models. This class of methods is commonly used in Bayesian inference and in recent years, it has become a popular alternative to existing methods such as Markov chain Monte Carlo and Laplace approximation methods. It is shown that the same ideas can also be transferred to frequentist contexts (Ormerod and Wand, 2010). In this section, we will delve into how variational approximations can be used in frequentist contexts.

In frequentist inferential problems, variational approximation methods mainly benefit inference carried out on statistical models where the vector of observations $\boldsymbol{y}$ is conditioned on a latent variable vector $\boldsymbol{u}$. In the context of generalized linear mixed models, the vector of latent variables essentially corresponds to the vector of random effects as shown in (1.17).

Let $\boldsymbol{\theta}$ be a vector of model parameters. When conditioning over the vector of latent variables $\boldsymbol{u}$ is present, the log-likelihood for a general statistical model parametrized by $\boldsymbol{\theta}$ is as follows

$$\ell(\boldsymbol{\theta}) = \log p(\boldsymbol{y}; \boldsymbol{\theta}) = \log \int p(\boldsymbol{y}|\boldsymbol{u}; \boldsymbol{\theta}) p(u; \boldsymbol{\theta}) d\boldsymbol{u}. \tag{1.25}$$

However, the integral in (1.25) may be intractable. Thus, $\ell(\boldsymbol{\theta})$ may not have a closed form and maximum likelihood estimation is hindered. The variational approximation method works around the intractability issue to provide a variational approximation to the maximum likelihood estimation approach, explained further below.

Let us define $q(\boldsymbol{u})$ to be an arbitrary density function in $\boldsymbol{u}$. Then the expression for

the log-likelihood satisfies the following mathematical steps

$$\begin{aligned}\ell(\boldsymbol{\theta}) &= \ell(\boldsymbol{\theta}) \int q(\boldsymbol{u}) d\boldsymbol{u} \\ &= \int q(\boldsymbol{u}) \ell(\boldsymbol{\theta}) d\boldsymbol{u} \\ &= \int q(\boldsymbol{u}) \log \left\{ \frac{p(\boldsymbol{y}, \boldsymbol{u}; \boldsymbol{\theta}) / q(\boldsymbol{u})}{p(\boldsymbol{u}|\boldsymbol{y}; \boldsymbol{\theta}) / q(\boldsymbol{u})} \right\} d\boldsymbol{u} \\ &= \int q(\boldsymbol{u}) \log \left\{ \frac{p(\boldsymbol{y}, \boldsymbol{u}; \boldsymbol{\theta})}{q(\boldsymbol{u})} \right\} d\boldsymbol{u} + \int q(\boldsymbol{u}) \log \left\{ \frac{q(\boldsymbol{u})}{p(\boldsymbol{u}|\boldsymbol{y}; \boldsymbol{\theta})} \right\} d\boldsymbol{u} \\ &\geq \underline{\ell}(\boldsymbol{\theta}; q)\end{aligned} \tag{1.26}$$

where

$$\underline{\ell}(\boldsymbol{\theta}; q) \equiv \int q(\boldsymbol{u}) \log \left\{ \frac{p(\boldsymbol{y}, \boldsymbol{u}; \boldsymbol{\theta})}{q(\boldsymbol{u})} \right\} d\boldsymbol{u}. \tag{1.27}$$

The inequality exists since

$$\int q(\boldsymbol{u}) \log \left\{ \frac{q(\boldsymbol{u})}{p(\boldsymbol{u}|\boldsymbol{y}; \boldsymbol{\theta})} \right\} d\boldsymbol{u} \geq 0 \tag{1.28}$$

across all densities q. Equality in (1.28) occurs if and only if

$$q(\boldsymbol{u}) = p(\boldsymbol{u}|\boldsymbol{y}; \boldsymbol{\theta})$$

almost everywhere. The integral expression in (1.28) is known as the *Kullback-Leibler divergence* between q and $p(\cdot|y)$.

One may now select a density q, where $q(\boldsymbol{u})$ approximates $p(\boldsymbol{u}|\boldsymbol{y}; \boldsymbol{\theta})$, such that $\underline{\ell}(\boldsymbol{\theta}; q)$ is more tractable than $\ell(\boldsymbol{\theta})$. One must also simultaneously aim to minimize the Kullback-Leibler divergence between $q(\boldsymbol{u})$ and $p(\boldsymbol{u}|\boldsymbol{y}; \boldsymbol{\theta})$, which can be achieved by maximizing $\underline{\ell}(\boldsymbol{\theta}; q)$, as shown in (1.26).

Suppose that q is restricted to a family of parametric densities $\{q(\boldsymbol{u}; \boldsymbol{\xi}) : \boldsymbol{\xi} \in \boldsymbol{\Xi}\}$ parametrized by a vector of variational parameters $\boldsymbol{\xi}$. Then, the expression for $\underline{\ell}(\boldsymbol{\theta}; q)$ in (1.27) becomes

$$\underline{\ell}(\boldsymbol{\theta}, \boldsymbol{\xi}; q) \equiv \int q(\boldsymbol{u}; \boldsymbol{\xi}) \log \left\{ \frac{p(\boldsymbol{y}, \boldsymbol{u}; \boldsymbol{\theta})}{q(\boldsymbol{u}; \boldsymbol{\xi})} \right\} d\boldsymbol{u}. \tag{1.29}$$

One now maximises over the vector of model parameters $\boldsymbol{\theta}$ and the vector of variational parameters $\boldsymbol{\xi}$, in order to maximise the approximate log-likelihood, $\underline{\ell}(\boldsymbol{\theta}, \boldsymbol{\xi}; q)$, and to minimize the Kullback-Leibler divergence between $q(\boldsymbol{u}; \boldsymbol{\xi})$ and $p(\boldsymbol{u}|\boldsymbol{y}; \boldsymbol{\theta})$ respectively.

This leads to the following altered maximization problem where

$$(\underline{\widehat{\boldsymbol{\theta}}}, \underline{\widehat{\boldsymbol{\xi}}}) = \underset{\boldsymbol{\theta}, \boldsymbol{\xi}}{\operatorname{argmax}}\ \ell(\boldsymbol{\theta}, \boldsymbol{\xi}; q).$$

Then, $\underline{\widehat{\boldsymbol{\theta}}}$ is the variational approximation to the maximum likelihood estimator $\widehat{\boldsymbol{\theta}}$ as defined in (1.21).

When q is chosen to be a Gaussian density function, then this particular class of variational approximation methods is known as Gaussian variational approximations. In GVA, the variational parameters are the mean and variance (or covariance matrix) parameters of the approximating normal distribution.

1.11.1 Thouless-Anderson-Palmer Variational Approach

The TAP variational framework, which builds on the GVA approach, has recently surfaced in statistical literature and potentially provides more accurate approximations as compared to the GVA approach.

The *Thouless-Anderson-Palmer* paradigm (Thouless et al., 1977) was first developed in statistical physics literature and gained traction as the authors provided TAP equations as an alternative approach to the solution for certain spin glass models. The work was further built on in Plefka (1982) where it is shown that the power expansion of the Gibbs potential of the infinite-ranged Ising spin glass model of Sherrington and Kirkpatrick (Sherrington and Kirkpatrick, 1975) up to the second order in the exchange couplings leads to the TAP equations.

In machine learning contexts, with the help of Plefka's expansion, it has been shown that better approximations arise by minimizing the TAP free energy, instead of the mean field free energy typically used in variational inference (Fan et al., 2021).

Recent theoretical work by Professor Song Mei from University of California, Berkeley, U.S.A and Professor Iain Johnstone from Stanford University, U.S.A, compares the estimates obtained from maximum likelihood estimation, GVA and TAP variational approaches for GLMMs. Let $\boldsymbol{\theta}$ be a vector of model parameters. Then, under some conditions, recent work by Professors Song Mei and Iain Johnstone (private communication) shows that

$$\|\boldsymbol{\theta}_{\text{GVA}} - \boldsymbol{\theta}_{\text{MLE}}\|_2 \approx Cn^{-2} \quad \text{and} \quad \|\boldsymbol{\theta}_{\text{TAP}} - \boldsymbol{\theta}_{\text{MLE}}\|_2 \approx Cn^{-3},$$

where C denotes a constant independent of $\boldsymbol{\theta}$. Note that the approximation error between the TAP and MLE estimates is smaller than the approximation error between the GVA and MLE estimates.

Despite the potential improvement in the accuracy of approximations that the TAP variational method can provide, statistical applications such as longitudinal data analysis and multilevel models analysis, which may benefit from using TAP methodology, have not been investigated.

Therefore, towards the end of this thesis, we apply the Thouless-Anderson-Palmer methodology to GLMMs and evaluate the approach via simulation studies.

Chapter 2

Preliminary Lemmas and Their Proofs

Detailed asymptotic analysis is necessary to obtain the asymptotic normality theorems for the maximum likelihood estimators for Gaussian response linear mixed models and maximum quasi-likelihood estimators for generalized linear mixed models in Chapters 3, 4 and 7 respectively. In the process of doing so, we deal with working with population limits of predictor-dependent sample mean quantities and establishing matrix norm asymptotic negligibility between matrix square roots of inverse Fisher information matrices and their simpler asymptotic block diagonal forms. Currently, there are no results available to deal with both these tasks in a simple manner.

Therefore, in this chapter, we introduce three novel lemmas that will act as essential tools required to solve these two tasks.

The appendix contains the proofs for the lemmas introduced in this chapter.

2.1 Lemma 1

Certain population quantities appear in the asymptotic normality theorems in Chapters 3, 4 and 7 respectively. These population quantities correspond to the convergence in probability limit of two particular predictor-dependent sample mean quantities each. In this section, we isolate the problem of deriving the population leading term of the first predictor-dependent sample mean quantity in the form of Lemma 1.

Lemma 1. *Let* $\boldsymbol{X} \equiv (\boldsymbol{X}_A^T, \boldsymbol{X}_B^T)^T$ *and* $\boldsymbol{X}_{ij} \equiv (\boldsymbol{X}_{Aij}^T, \boldsymbol{X}_{Bij}^T)^T, 1 \leq i \leq m, 1 \leq j \leq n_i$ *be independent and identically distributed* $(d_A + d_B) \times 1$ *random vectors, with* $d_A \geq 1$ *being the number of entries of* $\boldsymbol{X}_A$ *and the* $\boldsymbol{X}_{Aij}$*s and* $d_B \geq 1$ *being the number of entries of* $\boldsymbol{X}_B$ *and the* $\boldsymbol{X}_{Bij}$*s, with* $d = d_A + d_B$*. Also, let* $\boldsymbol{U}$ *and* $\boldsymbol{U}_1, \ldots, \boldsymbol{U}_m$ *be independent and identically distributed* $N(\boldsymbol{0}, \boldsymbol{I})$ $(d_A \times 1)$ *random vectors, distributed independently of* $\boldsymbol{X}$ *and the* $\boldsymbol{X}_{ij}$*s, where* $\boldsymbol{\Sigma}$ *is symmetric and positive definite. Let* f *be a Borel measurable, positive real-valued function on* $\mathbb{R}^{d_A+d_B}$ *and assume that*

$$E\left\{|X_k\, X_{k'}| f(\boldsymbol{X}, \boldsymbol{U})\right\} < \infty \text{ for all } 1 \leq k, k' \leq d$$

where X_k *is the* k^{th} *row of* $\boldsymbol{X}$*. Then*

$$\frac{1}{mn} \sum_{i=1}^{m} \sum_{j=1}^{n} \boldsymbol{X}_{ij} \boldsymbol{X}_{ij}^T E\left\{f(\boldsymbol{X}_{ij}, \boldsymbol{U}) | \boldsymbol{X}_{ij}\right\} = E\left\{\boldsymbol{X}\boldsymbol{X}^T f(\boldsymbol{X}, \boldsymbol{U})\right\} + o_P(1)\mathbf{1}_d \mathbf{1}_d^T$$

where

$$n \equiv \frac{1}{m} \sum_{i=1}^{m} n_i.$$

2.2 Lemma 2

In this section, we isolate the problem of deriving the population leading term of the second predictor-dependent sample mean quantity in the form of Lemma 2. Using both Lemmas 1 and 2 lead to a full expression for the population leading term in the main Fisher information block, represented by $\boldsymbol{\Sigma}_{\boldsymbol{\beta}_\mathrm{B}}$, $\boldsymbol{\Lambda}_{\boldsymbol{\beta}_\mathrm{B}}$ and $\boldsymbol{\Lambda}^*_{\boldsymbol{\beta}_1}$ in Chapters 3, 4 and 7 respectively.

Lemma 2. *Let* $\boldsymbol{X} \equiv (\boldsymbol{X}_A^T, \boldsymbol{X}_B^T)^T$ *and* $\boldsymbol{X}_{ij} \equiv (\boldsymbol{X}_{Aij}^T, \boldsymbol{X}_{Bij}^T)^T, 1 \leq i \leq m, 1 \leq j \leq n_i$ *be independent and identically distributed* $(d_A + d_B) \times 1$ *random vectors, with* $d_A \geq 1$ *being the number of entries of* $\boldsymbol{X}_A$ *and the* $\boldsymbol{X}_{Aij}$*s and* $d_B \geq 1$ *being the number of entries of* $\boldsymbol{X}_B$ *and the* $\boldsymbol{X}_{Bij}$*s, with* $d = d_A + d_B$*.. Also, let* $\boldsymbol{U}$ *and* $\boldsymbol{U}_1, \ldots, \boldsymbol{U}_m$ *be independent and identically distributed random vectors, distributed independently of* $\boldsymbol{X}$ *and the* $\boldsymbol{X}_{ij}$*s. Let* f *be a Borel measurable, positive real-valued function on* $\mathbb{R}^{d_A+d_B}$ *and assume that*

$$E\left[\frac{E\left[\max\{1, \|\boldsymbol{X}\|\}^8 \max\{1, f(\boldsymbol{X}, \boldsymbol{U})\}^4 | \boldsymbol{U}\right]}{\min\{1, \lambda_{\min}(E\{\boldsymbol{X}_A \boldsymbol{X}_A^T f(\boldsymbol{X}, \boldsymbol{U}|\boldsymbol{U})\})\}^2}\right] < \infty.$$

If m *and* n_i *satisfy assumptions that the number of groups* m *diverges to* ∞ *and that the within-group sample sizes* n_i *diverge to* ∞ *in such a way that* $n_i/n \to C_i$ *for constants*

$0 < C_i < \infty, 1 \leq i \leq m$, *then*

$$E\Bigg[\frac{1}{mn}\sum_{i=1}^{m}\left\{\sum_{j=1}^{n}\boldsymbol{X}_{Bij}\boldsymbol{X}_{Aij}^{T}f(\boldsymbol{X}_{ij},U_i)\right\}\left\{\sum_{j=1}^{n}\boldsymbol{X}_{Aij}\boldsymbol{X}_{Aij}^{T}f(\boldsymbol{X}_{ij},U_i)\right\}^{-1}$$
$$\times\left\{\sum_{j=1}^{n}\boldsymbol{X}_{Bij}\boldsymbol{X}_{Aij}^{T}f(\boldsymbol{X}_{ij},U_i)\right\}^{T}\Bigg|\boldsymbol{X}_{11},\ldots,\boldsymbol{X}_{mn_m}\Bigg]$$
$$\xrightarrow{P} E\left[E\left(\boldsymbol{X}_B\boldsymbol{X}_A^{T}f(\boldsymbol{X},U)|U\right)\left\{E\left(\boldsymbol{X}_A\boldsymbol{X}_A^{T}f(\boldsymbol{X},U)|U\right)\right\}^{-1}E\left(\boldsymbol{X}_B\boldsymbol{X}_A^{T}f(\boldsymbol{X},U)|U\right)^{T}\right].$$

2.3 Lemma 3

The asymptotic normality theorems in Chapters 3, 4 and 7 involve replacement of the matrix square root of the inverse Fisher information matrix by the matrix square root of the asymptotic expression for the inverse Fisher information matrix. This is due to the remainder terms (calculated as a difference between the inverse Fisher information matrix and its asymptotic counterpart) having an asymptotically negligible effect on the relevant matrix square roots. Lemma 3 provides a formalization of this state of affairs, which is used in the final steps of the derivation in the asymptotic normality theorems in Chapters 3, 4 and 7.

Lemma 3. *Define the sequences of matrices*

$$\boldsymbol{M}_n \equiv \begin{bmatrix}\boldsymbol{K}+Q_n\mathbf{1}_p^{\otimes 2} & R_n\mathbf{1}_p\mathbf{1}_q^{T}\\ R_n\mathbf{1}_q\mathbf{1}_p^{T} & \frac{1}{n}\boldsymbol{L}+T_n\mathbf{1}_q^{\otimes 2}\end{bmatrix} \text{ and } \boldsymbol{M}_{n,\infty}\equiv\begin{bmatrix}\boldsymbol{K} & \mathbf{0}\\ \mathbf{0} & \frac{1}{n}\boldsymbol{L}\end{bmatrix}$$

where $\boldsymbol{K}$ *and* $\boldsymbol{L}$ *are* $p\times p$ *and* $q\times q$ *symmetric positive definite matrices and* Q_n, R_n *and* T_n *are sequences of random variables satisfying* $Q_n = o_P(1), R_n = O_P(n^{-1})$ *and* $T_n = o_P(n^{-1})$. *Also note that* $\nu^{\otimes 2}\equiv\nu\nu^{T}$. *Then, as* $n\to\infty$,

$$\left\|\boldsymbol{M}_{n,\infty}^{-1/2}\boldsymbol{M}_n^{1/2}-\boldsymbol{I}\right\|_F\xrightarrow{P}0.$$

2.4 Appendix

2.4.1 Proof of Lemma 1

Let

$$\overline{\boldsymbol{G}} \equiv \frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n}\boldsymbol{X}_{ij}\boldsymbol{X}_{ij}^{T}E\{f(\boldsymbol{X}_{ij},\boldsymbol{U})|\boldsymbol{X}_{ij}\}.$$

Firstly, note that a more explicit form of $\overline{\boldsymbol{G}}$ is:

$$\overline{\boldsymbol{G}} = \frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n}\boldsymbol{X}_{ij}\boldsymbol{X}_{ij}^{T}(2\pi)^{-d_{\mathrm{A}}/2}|\boldsymbol{\Sigma}|^{-1/2}\int_{\mathbb{R}^{d_{\mathrm{A}}}} f(\boldsymbol{X}_{ij},\boldsymbol{u})\exp(-\frac{1}{2}\boldsymbol{u}^{T}\boldsymbol{\Sigma}^{-1}\boldsymbol{u})d\boldsymbol{u}.$$

Next, let

$$N \equiv mn = \sum_{i=1}^{m} n_i,$$

and $\boldsymbol{G}_l$, $1 \le l \le N$, be the following "single subscript" re-labelling of the terms inside the double summation of $\overline{\boldsymbol{G}}$. Then we have the following terms where

$$\begin{aligned}
\boldsymbol{G}_1 &= \boldsymbol{X}_{11}\boldsymbol{X}_{11}^{T}(2\pi)^{-d_{\mathrm{A}}/2}|\boldsymbol{\Sigma}|^{-1/2}\int_{\mathbb{R}^{d_{\mathrm{A}}}} f(\boldsymbol{X}_{ij},\boldsymbol{u})\exp(-\frac{1}{2}\boldsymbol{u}^{T}\boldsymbol{\Sigma}^{-1}\boldsymbol{u})d\boldsymbol{u},\\
\boldsymbol{G}_2 &= \boldsymbol{X}_{12}\boldsymbol{X}_{12}^{T}(2\pi)^{-d_{\mathrm{A}}/2}|\boldsymbol{\Sigma}|^{-1/2}\int_{\mathbb{R}^{d_{\mathrm{A}}}} f(\boldsymbol{X}_{ij},\boldsymbol{u})\exp(-\frac{1}{2}\boldsymbol{u}^{T}\boldsymbol{\Sigma}^{-1}\boldsymbol{u})d\boldsymbol{u},\\
&\vdots\\
\boldsymbol{G}_N &= \boldsymbol{X}_{mn_m}\boldsymbol{X}_{mn_m}^{T}(2\pi)^{-d_{\mathrm{A}}/2}|\boldsymbol{\Sigma}|^{-1/2}\int_{\mathbb{R}^{d_{\mathrm{A}}}} f(\boldsymbol{X}_{ij},\boldsymbol{u})\exp(-\frac{1}{2}\boldsymbol{u}^{T}\boldsymbol{\Sigma}^{-1}\boldsymbol{u})d\boldsymbol{u}.
\end{aligned}$$

Then,

$$\overline{\boldsymbol{G}} = \frac{1}{N}\sum_{l=1}^{N}\boldsymbol{G}_l$$

is a sample mean of N independent and identically distributed $d \times d$ random matrices with the common distribution

$$\boldsymbol{G} \equiv \boldsymbol{X}\boldsymbol{X}^{T}E\{f(\boldsymbol{X},\boldsymbol{U})|\boldsymbol{X}\}.$$

Using the law of total expectation, the mean of $\boldsymbol{G}$ is

$$\begin{aligned}
E(\boldsymbol{G}) &= E[\boldsymbol{X}\boldsymbol{X}^{T}E\{f(\boldsymbol{X},\boldsymbol{U})|\boldsymbol{X}\}]\\
&= E[E\{\boldsymbol{X}\boldsymbol{X}^{T}f(\boldsymbol{X},\boldsymbol{U})|\boldsymbol{X}\}]\\
&= E\{\boldsymbol{X}\boldsymbol{X}^{T}f(\boldsymbol{X},\boldsymbol{U})\}.
\end{aligned}$$

Lastly, we need to impose the following first order moment conditions on the entries of $\boldsymbol{G}$ where

$$E\{|\boldsymbol{X}_k\boldsymbol{X}_{k'}|f(\boldsymbol{X},\boldsymbol{U})\} < \infty \text{ for all } 1 \le k, k' \le d, \tag{2.1}$$

where $\boldsymbol{X}_k$ is the kth entry of $\boldsymbol{X}$. Therefore, under moment conditions involving the entries of $\boldsymbol{G}$ as stated in (2.1), we have,

$$\overline{\boldsymbol{G}} = \frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n}\boldsymbol{X}_{ij}\boldsymbol{X}_{ij}^T E\{f(\boldsymbol{X}_{ij},\boldsymbol{U})|\boldsymbol{X}_{ij}\} = E\{\boldsymbol{X}\boldsymbol{X}^T f(\boldsymbol{X},\boldsymbol{U})\} + o_P(1)\mathbf{1}_d\mathbf{1}_d^T.$$

2.4.2 Proof of Lemma 2

2.4.2.1 A Fundamental Inequality for the Spectral Norm of a Vectorised Matrix

Let $\boldsymbol{A}$ be a $d_1 \times d_2$ matrix. This subsubsection looks into the relationship between

$$\|\boldsymbol{A}\|_S \quad \text{and} \quad \|\text{vec}(\boldsymbol{A})\|_S.$$

We start with the following inequality:

$$\|\boldsymbol{A}\|_S \leq \|\boldsymbol{A}\|_F \leq \sqrt{\text{rank}(\boldsymbol{A})}\,\|\boldsymbol{A}\|_S. \tag{2.2}$$

However, since $\text{rank}(\boldsymbol{A}) \leq \max(d_1, d_2)$, we then obtain the following

$$\|\boldsymbol{A}\|_S \leq \|\boldsymbol{A}\|_F \leq \max\left(\sqrt{d_1}, \sqrt{d_2}\right)\|\boldsymbol{A}\|_S. \tag{2.3}$$

Next note that

$$\|\boldsymbol{A}\|_F = \sqrt{\text{tr}(\boldsymbol{A}^T\boldsymbol{A})} = \sqrt{\text{vec}(\boldsymbol{A})^T\text{vec}(\boldsymbol{A})} = \|\text{vec}(\boldsymbol{A})\|_F. \tag{2.4}$$

Replacement of $\boldsymbol{A}$ by $\text{vec}(\boldsymbol{A})$ in the first inequality of (2.2) gives

$$\|\text{vec}(\boldsymbol{A})\|_S \leq \|\text{vec}(\boldsymbol{A})\|_F.$$

The equality in (2.4) then gives

$$\|\text{vec}(\boldsymbol{A})\|_S \leq \|\boldsymbol{A}\|_F.$$

Application of the second inequality of (2.3) leads to

$$\|\text{vec}(\boldsymbol{A})\|_S \leq \max\left(\sqrt{d_1}, \sqrt{d_2}\right)\|\boldsymbol{A}\|_S \tag{2.5}$$

for all $d_1 \times d_2$ matrices $\boldsymbol{A}$.

2.4.2.2 Notational Definitions

Let us define the following sample and population type quantities as follows:

$$\widehat{\mathcal{N}}_i(\boldsymbol{U}) \equiv \frac{1}{n_i}\sum_{j=1}^{n}\boldsymbol{X}_{\text{B}ij}\boldsymbol{X}_{\text{A}ij}^T f(\boldsymbol{X}_{ij},\boldsymbol{U}), \quad \widehat{\mathcal{D}}_i(\boldsymbol{U}) \equiv \frac{1}{n_i}\sum_{j=1}^{n}\boldsymbol{X}_{\text{A}ij}\boldsymbol{X}_{\text{A}ij}^T f(\boldsymbol{X}_{ij},\boldsymbol{U}),$$
$$\mathcal{N}(\boldsymbol{U}) \equiv E\{\boldsymbol{X}_{\text{B}}\boldsymbol{X}_{\text{A}}^T f(\boldsymbol{X},\boldsymbol{U})|\boldsymbol{U}\} \quad \text{and} \quad \mathcal{D}(\boldsymbol{U}) \equiv E\{\boldsymbol{X}_{\text{A}}\boldsymbol{X}_{\text{A}}^T f(\boldsymbol{X},\boldsymbol{U})|\boldsymbol{U}\}.$$

Next, for $t \in [0,1]$, let

$$\mathcal{N}_{it}^{\dagger}(\boldsymbol{U}) \equiv (1-t)\mathcal{N}(\boldsymbol{U}) + t\widehat{\mathcal{N}}_i(\boldsymbol{U}) \quad \text{and} \quad \mathcal{D}_{it}^{\dagger}(\boldsymbol{U}) \equiv (1-t)\mathcal{D}(\boldsymbol{U}) + t\widehat{\mathcal{D}}_i(\boldsymbol{U}).$$

Throughout this proof, we also let

$$\mathcal{X}_i \equiv \{\boldsymbol{X}_{i1}, \ldots, \boldsymbol{X}_{in_i}\}.$$

In addition, if $\boldsymbol{S}$ is a $d_{\mathrm{B}} \times d_{\mathrm{A}}$ matrix and $\boldsymbol{T}$ is a $d_{\mathrm{A}} \times d_{\mathrm{A}}$ symmetric matrix, define

$$\mathcal{R}\left(\begin{bmatrix}\boldsymbol{S}\\ \boldsymbol{T}\end{bmatrix}\right) \equiv \mathrm{vec}(\boldsymbol{S}\boldsymbol{T}^{-1}\boldsymbol{S}^T)^T. \tag{2.6}$$

In the next subsubsection, we wish to find an explicit expression for

$$\nabla_{\mathrm{vec}(\boldsymbol{S}),\,\mathrm{vec}(\boldsymbol{T})}\mathcal{R}\left(\begin{bmatrix}\boldsymbol{S}\\ \boldsymbol{T}\end{bmatrix}\right). \tag{2.7}$$

2.4.2.3 Derivation of (2.7)

Differentiation with Respect to $\boldsymbol{S}$

Throughout this subsubsubsection it is assumed that the differential operator d is with respect to $\boldsymbol{S}$. Now, noting that $\boldsymbol{T}^T = \boldsymbol{T}$,

$$\begin{aligned}
d\,\mathrm{vec}(\boldsymbol{S}\boldsymbol{T}^{-1}\boldsymbol{S}^T) &= \mathrm{vec}\Big(d(\boldsymbol{S}\boldsymbol{T}^{-1}\boldsymbol{S}^T)\Big)\\
&= \mathrm{vec}\Big(d(\boldsymbol{S}\boldsymbol{T}^{-1})\boldsymbol{S}^T + (\boldsymbol{S}\boldsymbol{T}^{-1})d\boldsymbol{S}^T\Big)\\
&= \mathrm{vec}\Big((d\boldsymbol{S})\boldsymbol{T}^{-1}\boldsymbol{S}^T\Big) + \mathrm{vec}\Big(\boldsymbol{S}\boldsymbol{T}^{-1}(d\boldsymbol{S})^T\Big)\\
&= \mathrm{vec}\Big((d\boldsymbol{S})\boldsymbol{T}^{-1}\boldsymbol{S}^T\Big) + \boldsymbol{K}_{d_{\mathrm{B}}}\mathrm{vec}\Big((d\boldsymbol{S})\boldsymbol{T}^{-1}\boldsymbol{S}^T\Big)\\
&= (\boldsymbol{I}_{d_{\mathrm{B}}^2} + \boldsymbol{K}_{d_{\mathrm{B}}})\mathrm{vec}\Big((d\boldsymbol{S})\boldsymbol{T}^{-1}\boldsymbol{S}^T\Big)\\
&= (\boldsymbol{I}_{d_{\mathrm{B}}^2} + \boldsymbol{K}_{d_{\mathrm{B}}})\mathrm{vec}\Big(\boldsymbol{I}_{d_{\mathrm{B}}}(d\boldsymbol{S})\boldsymbol{T}^{-1}\boldsymbol{S}^T\Big)\\
&= (\boldsymbol{I}_{d_{\mathrm{B}}^2} + \boldsymbol{K}_{d_{\mathrm{B}}})\big\{(\boldsymbol{T}^{-1}\boldsymbol{S}^T)^T \otimes \boldsymbol{I}_{d_{\mathrm{B}}}\big\}\mathrm{vec}(d\boldsymbol{S})\\
&= (\boldsymbol{I}_{d_{\mathrm{B}}^2} + \boldsymbol{K}_{d_{\mathrm{B}}})\big\{(\boldsymbol{S}\boldsymbol{T}^{-1}) \otimes \boldsymbol{I}_{d_{\mathrm{B}}}\big\}d\mathrm{vec}(\boldsymbol{S}).
\end{aligned}$$

Hence,

$$\mathsf{D}_{\mathrm{vec}(\boldsymbol{S})}\mathrm{vec}(\boldsymbol{S}\boldsymbol{T}^{-1}\boldsymbol{S}^T) = (\boldsymbol{I}_{d_{\mathrm{B}}^2} + \boldsymbol{K}_{d_{\mathrm{B}}})\big\{(\boldsymbol{S}\boldsymbol{T}^{-1}) \otimes \boldsymbol{I}_{d_{\mathrm{B}}}\big\}.$$

Since $\boldsymbol{K}_{d_{\mathrm{B}}}^T = \boldsymbol{K}_{d_{\mathrm{B}}}$ as in (1.8), we have,

$$\nabla_{\mathrm{vec}(\boldsymbol{S})}\mathrm{vec}(\boldsymbol{S}\boldsymbol{T}^{-1}\boldsymbol{S}^T)^T = \big\{(\boldsymbol{T}^{-1}\boldsymbol{S}^T) \otimes \boldsymbol{I}_{d_{\mathrm{B}}}\big\}(\boldsymbol{I}_{d_{\mathrm{B}}^2} + \boldsymbol{K}_{d_{\mathrm{B}}}).$$

Differentiation with Respect to $\boldsymbol{T}$

Throughout this subsubsubsection it is assumed that the differential operator d is with respect to $\boldsymbol{T}$. Now, noting that $\boldsymbol{T}^T = \boldsymbol{T}$,

$$\begin{aligned} d\,\mathrm{vec}(\boldsymbol{S}\boldsymbol{T}^{-1}\boldsymbol{S}^T) &= \mathrm{vec}\Big(\boldsymbol{S}(d\boldsymbol{T}^{-1})\boldsymbol{S}^T\Big) \\ &= -\mathrm{vec}\Big(\boldsymbol{S}\boldsymbol{T}^{-1}(d\boldsymbol{T})\boldsymbol{T}^{-1}\boldsymbol{S}^T\Big) \\ &= -\{(\boldsymbol{T}^{-1}\boldsymbol{S}^T)^T \otimes (\boldsymbol{S}\boldsymbol{T}^{-1})\}\mathrm{vec}(d\boldsymbol{T}) \\ &= -\{(\boldsymbol{S}\boldsymbol{T}^{-T}) \otimes (\boldsymbol{S}\boldsymbol{T}^{-1})\}d\mathrm{vec}(\boldsymbol{T}). \end{aligned}$$

Hence,

$$\mathsf{D}_{\mathrm{vec}(\boldsymbol{T})}\mathrm{vec}(\boldsymbol{S}\boldsymbol{T}^{-1}\boldsymbol{S}^T) = -(\boldsymbol{S}\boldsymbol{T}^{-T}) \otimes (\boldsymbol{S}\boldsymbol{T}^{-1}).$$

Therefore, we have,

$$\nabla_{\mathrm{vec}(\boldsymbol{T})}\mathrm{vec}(\boldsymbol{S}\boldsymbol{T}^{-1}\boldsymbol{S}^T)^T = -(\boldsymbol{T}^{-1}\boldsymbol{S}^T) \otimes (\boldsymbol{T}^{-1}\boldsymbol{S}^T).$$

Combination of the Two Gradient Vectors

On combining the results of the previous two subsubsubsections, we get an explicit expression for (2.7) as follows

$$\nabla_{\mathrm{vec}(\boldsymbol{S}),\,\mathrm{vec}(\boldsymbol{T})}\mathcal{R}\left(\begin{bmatrix}\boldsymbol{S}\\ \boldsymbol{T}\end{bmatrix}\right) = \begin{bmatrix}\{(\boldsymbol{T}^{-1}\boldsymbol{S}^T) \otimes \boldsymbol{I}_{d_\mathrm{B}}\}(\boldsymbol{I}_{d_\mathrm{B}^2} + \boldsymbol{K}_{d_\mathrm{B}}) \\ -(\boldsymbol{T}^{-1}\boldsymbol{S}^T) \otimes (\boldsymbol{T}^{-1}\boldsymbol{S}^T)\end{bmatrix}. \tag{2.8}$$

2.4.2.4 Expression for (2.6) with Lagrange Form of Remainder

Using (2.8), a Taylor series expansion of $\mathcal{R}$ with the Lagrange form of the remainder is

$$\begin{aligned} \mathcal{R}\left(\begin{bmatrix}\boldsymbol{S}\\ \boldsymbol{T}\end{bmatrix}\right) &= \mathcal{R}\left(\begin{bmatrix}\boldsymbol{S}_0 + \boldsymbol{S} - \boldsymbol{S}_0\\ \boldsymbol{T}_0 + \boldsymbol{T} - \boldsymbol{T}_0\end{bmatrix}\right) \\ &= \mathcal{R}\left(\begin{bmatrix}\boldsymbol{S}_0\\ \boldsymbol{T}_0\end{bmatrix}\right) + \begin{bmatrix}\mathrm{vec}(\boldsymbol{S} - \boldsymbol{S}_0)\\ \mathrm{vec}(\boldsymbol{T} - \boldsymbol{T}_0)\end{bmatrix}^T \nabla_{\mathrm{vec}(\boldsymbol{S}),\,\mathrm{vec}(\boldsymbol{T})}\mathcal{R}\left(\begin{bmatrix}\boldsymbol{S}_t^\dagger\\ \boldsymbol{T}_t^\dagger\end{bmatrix}\right) \\ &= \mathcal{R}\left(\begin{bmatrix}\boldsymbol{S}_0\\ \boldsymbol{T}_0\end{bmatrix}\right) + \begin{bmatrix}\mathrm{vec}(\boldsymbol{S} - \boldsymbol{S}_0)\\ \mathrm{vec}(\boldsymbol{T} - \boldsymbol{T}_0)\end{bmatrix}^T \begin{bmatrix}[\{(\boldsymbol{T}_t^\dagger)^{-1}(\boldsymbol{S}_t^\dagger)^T\} \otimes \boldsymbol{I}_{d_\mathrm{B}}](\boldsymbol{I}_{d_\mathrm{B}^2} + \boldsymbol{K}_{d_\mathrm{B}})\\ -\{(\boldsymbol{T}_t^\dagger)^{-1}(\boldsymbol{S}_t^\dagger)^T\} \otimes \{(\boldsymbol{T}_t^\dagger)^{-1}(\boldsymbol{S}_t^\dagger)^T\}\end{bmatrix} \end{aligned}$$

where

$$\begin{bmatrix}\boldsymbol{S}_t^\dagger\\ \boldsymbol{T}_t^\dagger\end{bmatrix} \equiv (1-t)\begin{bmatrix}\boldsymbol{S}_0\\ \boldsymbol{T}_0\end{bmatrix} + t\begin{bmatrix}\boldsymbol{S}\\ \boldsymbol{T}\end{bmatrix} = \begin{bmatrix}(1-t)\boldsymbol{S}_0 + t\boldsymbol{S}\\ (1-t)\boldsymbol{T}_0 + t\boldsymbol{T}\end{bmatrix}$$

and $t \in [0,1]$. It follows that

$$\left\{\mathcal{R}\left(\begin{bmatrix}\boldsymbol{S}\\ \boldsymbol{T}\end{bmatrix}\right)-\mathcal{R}\left(\begin{bmatrix}\boldsymbol{S}_0\\ \boldsymbol{T}_0\end{bmatrix}\right)\right\}^T = (\boldsymbol{I}_{d_{\mathrm{B}}^2}+\boldsymbol{K}_{d_{\mathrm{B}}})[\{\boldsymbol{S}_t^\dagger(\boldsymbol{T}_t^\dagger)^{-1}\}\otimes \boldsymbol{I}_{d_{\mathrm{B}}}]\mathrm{vec}(\boldsymbol{S}-\boldsymbol{S}_0) - [\{\boldsymbol{S}_t^\dagger(\boldsymbol{T}_t^\dagger)^{-1}\}\otimes\{\boldsymbol{S}_t^\dagger(\boldsymbol{T}_t^\dagger)^{-1}\}]\mathrm{vec}(\boldsymbol{T}-\boldsymbol{T}_0). \tag{2.9}$$

2.4.2.5 Spectral Norm Bounding of (2.9)

It follows from (2.9) that

$$\begin{aligned}\left\|\mathcal{R}\left(\begin{bmatrix}\boldsymbol{S}\\ \boldsymbol{T}\end{bmatrix}\right)-\mathcal{R}\left(\begin{bmatrix}\boldsymbol{S}_0\\ \boldsymbol{T}_0\end{bmatrix}\right)\right\|_S &\le \|\boldsymbol{I}_{d_{\mathrm{B}}^2}+\boldsymbol{K}_{d_{\mathrm{B}}}\|_S\,\|[\{\boldsymbol{S}_t^\dagger(\boldsymbol{T}_t^\dagger)^{-1}\}\otimes\boldsymbol{I}_{d_{\mathrm{B}}}\|_S\,\|\boldsymbol{S}-\boldsymbol{S}_0\|_F\\ &\quad+\|\{\boldsymbol{S}_t^\dagger(\boldsymbol{T}_t^\dagger)^{-1}\}\otimes\{\boldsymbol{S}_t^\dagger(\boldsymbol{T}_t^\dagger)^{-1}\}\|_S\,\|\boldsymbol{T}-\boldsymbol{T}_0\|_F\\ &\le \big(\|\boldsymbol{I}_{d_{\mathrm{B}}^2}\|_S+\|\boldsymbol{K}_{d_{\mathrm{B}}}\|_S\big)\|\boldsymbol{S}_t^\dagger(\boldsymbol{T}_t^\dagger)^{-1}\|_S\|\boldsymbol{I}_{d_{\mathrm{B}}}\|_S\|\boldsymbol{S}-\boldsymbol{S}_0\|_F\\ &\quad+\big\{\|\boldsymbol{S}_t^\dagger(\boldsymbol{T}_t^\dagger)^{-1}\|_S\big\}^2\|\boldsymbol{T}-\boldsymbol{T}_0\|_F\\ &\le 2\|\boldsymbol{S}_t^\dagger\|_S\,\|(\boldsymbol{T}_t^\dagger)^{-1}\|_S\|\boldsymbol{S}-\boldsymbol{S}_0\|_F+\big\{\|\boldsymbol{S}_t^\dagger\|_S\,\|(\boldsymbol{T}_t^\dagger)^{-1}\|_S\big\}^2\|\boldsymbol{T}-\boldsymbol{T}_0\|_F.\end{aligned}$$

Now, in terms of the notation given in the previous subsubsections, our goal is to show that

$$E\left\{\frac{1}{m}\sum_{i=1}^m\frac{n_i}{n}\mathcal{R}\left(\begin{bmatrix}\widehat{\mathcal{N}}_i(\boldsymbol{U}_i)\\ \widehat{\mathcal{D}}_i(\boldsymbol{U}_i)\end{bmatrix}\right)\middle|\mathcal{X}_i\right\}\xrightarrow{P}E\left\{\mathcal{R}\left(\begin{bmatrix}\mathcal{N}(\boldsymbol{U})\\ \mathcal{D}(\boldsymbol{U})\end{bmatrix}\right)\right\}. \tag{2.10}$$

in order to prove Lemma 2.

2.4.2.6 Strategy for Proving (2.10)

Result (2.10) is a consequence of

$$E\left\|E\left\{\frac{1}{m}\sum_{i=1}^m\frac{n_i}{n}\mathcal{R}\left(\begin{bmatrix}\widehat{\mathcal{N}}_i(\boldsymbol{U}_i)\\ \widehat{\mathcal{D}}_i(\boldsymbol{U}_i)\end{bmatrix}\right)\middle|\mathcal{X}_i\right\}-E\left\{\mathcal{R}\left(\begin{bmatrix}\mathcal{N}(\boldsymbol{U})\\ \mathcal{D}(\boldsymbol{U})\end{bmatrix}\right)\right\}\right\|_S\to 0 \tag{2.11}$$

as $m,n\to\infty$, If we bring the second inner expectation inside the $i=1,\ldots,m$ summation and replace the $\boldsymbol{U}$ of this term by $\boldsymbol{U}_i$, then we can replace (2.11) by

$$E\left\|\frac{1}{m}\sum_{i=1}^m\frac{n_i}{n}E\left\{\mathcal{R}\left(\begin{bmatrix}\widehat{\mathcal{N}}_i(\boldsymbol{U}_i)\\ \widehat{\mathcal{D}}_i(\boldsymbol{U}_i)\end{bmatrix}\right)-\mathcal{R}\left(\begin{bmatrix}\mathcal{N}(\boldsymbol{U}_i)\\ \mathcal{D}(\boldsymbol{U}_i)\end{bmatrix}\right)\middle|\mathcal{X}_i\right\}\right\|_S. \tag{2.12}$$

By noting that the left-hand side of (2.12) is bounded above by

$$\frac{1}{mn}\sum_{i=1}^m n_iE\left\|E\left\{\mathcal{R}\left(\begin{bmatrix}\widehat{\mathcal{N}}_i(\boldsymbol{U}_i)\\ \widehat{\mathcal{D}}_i(\boldsymbol{U}_i)\end{bmatrix}\right)-\mathcal{R}\left(\begin{bmatrix}\mathcal{N}(\boldsymbol{U}_i)\\ \mathcal{D}(\boldsymbol{U}_i)\end{bmatrix}\right)\middle|\mathcal{X}_i\right\}\right\|_S,$$

it is sufficient to prove that

$$\frac{1}{mn}\sum_{i=1}^{m} n_i E\left\| E\left\{ \mathcal{R}\left(\begin{bmatrix} \widehat{\mathcal{N}}_i(\boldsymbol{U}_i) \\ \widehat{\mathcal{D}}_i(\boldsymbol{U}_i) \end{bmatrix} \right) - \mathcal{R}\left(\begin{bmatrix} \mathcal{N}(\boldsymbol{U}_i) \\ \mathcal{D}(\boldsymbol{U}_i) \end{bmatrix} \right) \middle| \mathcal{X}_i \right\} \right\|_S \to 0 \tag{2.13}$$

as $m, n \to \infty$. Next, for each $1 \le i \le m$, define the event

$$\mathcal{A}_i \equiv \left\{ \|\widehat{\mathcal{N}}_i(\boldsymbol{U}_i) - \mathcal{N}(\boldsymbol{U}_i)\|_S \le 1,\ \lambda_{\min}\big(\widehat{\mathcal{D}}_i(\boldsymbol{U}_i)\big) \ge \tfrac{1}{2}\lambda_{\min}\big(\mathcal{D}(\boldsymbol{U}_i)\big) \right\}. \tag{2.14}$$

Now, note that

$$\begin{aligned} &\frac{1}{mn}\sum_{i=1}^{m} n_i E\left\| E\left\{ \mathcal{R}\left(\begin{bmatrix} \widehat{\mathcal{N}}_i(\boldsymbol{U}_i) \\ \widehat{\mathcal{D}}_i(\boldsymbol{U}_i) \end{bmatrix} \right) - \mathcal{R}\left(\begin{bmatrix} \mathcal{N}(\boldsymbol{U}_i) \\ \mathcal{D}(\boldsymbol{U}_i) \end{bmatrix} \right) \middle| \mathcal{X}_i \right\} \right\|_S \\ &\le \frac{1}{mn}\sum_{i=1}^{m} n_i E\left\| E\left[\left\{ \mathcal{R}\left(\begin{bmatrix} \widehat{\mathcal{N}}_i(\boldsymbol{U}_i) \\ \widehat{\mathcal{D}}_i(\boldsymbol{U}_i) \end{bmatrix} \right) - \mathcal{R}\left(\begin{bmatrix} \mathcal{N}(\boldsymbol{U}_i) \\ \mathcal{D}(\boldsymbol{U}_i) \end{bmatrix} \right) \right\} I(\mathcal{A}_i) \middle| \mathcal{X}_i \right] \right\|_S \\ &\quad + \frac{1}{mn}\sum_{i=1}^{m} n_i E\left\| E\left[\left\{ \mathcal{R}\left(\begin{bmatrix} \widehat{\mathcal{N}}_i(\boldsymbol{U}_i) \\ \widehat{\mathcal{D}}_i(\boldsymbol{U}_i) \end{bmatrix} \right) - \mathcal{R}\left(\begin{bmatrix} \mathcal{N}(\boldsymbol{U}_i) \\ \mathcal{D}(\boldsymbol{U}_i) \end{bmatrix} \right) \right\} I(\mathcal{A}_i^C) \middle| \mathcal{X}_i \right] \right\|_S . \end{aligned} \tag{2.15}$$

Then, to prove Lemma 2, it is sufficient to prove that

$$\frac{1}{mn}\sum_{i=1}^{m} n_i E\left\| E\left[\left\{ \mathcal{R}\left(\begin{bmatrix} \widehat{\mathcal{N}}_i(\boldsymbol{U}_i) \\ \widehat{\mathcal{D}}_i(\boldsymbol{U}_i) \end{bmatrix} \right) - \mathcal{R}\left(\begin{bmatrix} \mathcal{N}(\boldsymbol{U}_i) \\ \mathcal{D}(\boldsymbol{U}_i) \end{bmatrix} \right) \right\} I(\mathcal{A}_i) \middle| \mathcal{X}_i \right] \right\|_S \to 0 \tag{2.16}$$

as $m, n, \to \infty$ and

$$\frac{1}{mn}\sum_{i=1}^{m} n_i E\left\| E\left[\left\{ \mathcal{R}\left(\begin{bmatrix} \widehat{\mathcal{N}}_i(\boldsymbol{U}_i) \\ \widehat{\mathcal{D}}_i(\boldsymbol{U}_i) \end{bmatrix} \right) - \mathcal{R}\left(\begin{bmatrix} \mathcal{N}(\boldsymbol{U}_i) \\ \mathcal{D}(\boldsymbol{U}_i) \end{bmatrix} \right) \right\} I(\mathcal{A}_i^C) \middle| \mathcal{X}_i \right] \right\|_S \to 0 \tag{2.17}$$

as $m, n \to \infty$.

2.4.2.7 Proof of Result (2.16)

Throughout this subsubsection we are considering

$$(\boldsymbol{U}_i, \mathcal{X}_i) \text{ such that } \mathcal{A}_i \text{ occurs}, 1 \le i \le m. \tag{2.18}$$

Since

$$\mathcal{N}_{it}^{\dagger}(\boldsymbol{U}_i) \equiv (1-t)\mathcal{N}(\boldsymbol{U}_i) + t\widehat{\mathcal{N}}_i(\boldsymbol{U}_i) = \mathcal{N}(\boldsymbol{U}_i) + t\{\widehat{\mathcal{N}}_i(\boldsymbol{U}_i) - \mathcal{N}(\boldsymbol{U}_i)\}$$

we have

$$\|\mathcal{N}_{it}^{\dagger}(\boldsymbol{U}_i)\|_S \le \|\mathcal{N}(\boldsymbol{U}_i)\|_S + t\|\widehat{\mathcal{N}}_i(\boldsymbol{U}_i) - \mathcal{N}(\boldsymbol{U}_i)\|_S,$$

and it follows that, for $t \in [0,1]$ and under (2.18), we have,

$$\|\mathcal{N}_{it}^{\dagger}(\boldsymbol{U}_i)\|_S \le \|\mathcal{N}(\boldsymbol{U}_i)\|_S + 1. \tag{2.19}$$

Next, note that

$$\begin{aligned}\|\mathcal{D}_{it}^{\dagger}(\boldsymbol{U})^{-1}\|_S &= 1\Big/\lambda_{\min}\big(\mathcal{D}_{it}^{\dagger}(\boldsymbol{U})\big)\\ &= 1\Big/\lambda_{\min}\big((1-t)\mathcal{D}(\boldsymbol{U}_i)+t\widehat{\mathcal{D}}_i(\boldsymbol{U}_i)\big)\\ &\leq 1\Big/\Big\{(1-t)\lambda_{\min}\big(\mathcal{D}(\boldsymbol{U}_i)\big)+t\lambda_{\min}\big(\widehat{\mathcal{D}}_i(\boldsymbol{U}_i)\big)\Big\}.\end{aligned} \tag{2.20}$$

Under (2.18), we have,

$$\begin{aligned}\|\mathcal{D}_{it}^{\dagger}(\boldsymbol{U}_i)^{-1}\|_S &\leq 1\Big/\Big\{(1-t)\lambda_{\min}\big(\mathcal{D}(\boldsymbol{U}_i)\big)+t\lambda_{\min}\big(\widehat{\mathcal{D}}_i(\boldsymbol{U}_i)\big)\Big\}\\ &\leq 1\Big/\Big\{(1-t)\lambda_{\min}\big(\mathcal{D}(\boldsymbol{U}_i)\big)+\tfrac{t}{2}\lambda_{\min}\big(\mathcal{D}(\boldsymbol{U}_i)\big)\Big\}\\ &= \frac{1}{(1-t/2)\lambda_{\min}\big(\mathcal{D}(\boldsymbol{U}_i)\big)}\\ &= \left(\frac{2}{2-t}\right)\|\mathcal{D}(\boldsymbol{U}_i)^{-1}\|_S.\end{aligned}$$

Since

$$\sup_{t\in[0,1]}\left(\frac{2}{2-t}\right)=2$$

and $t\in[0,1]$, under (2.18), we have,

$$\|\mathcal{D}_{it}^{\dagger}(\boldsymbol{U}_i)^{-1}\|_S \leq 2\|\mathcal{D}(\boldsymbol{U}_i)^{-1}\|_S. \tag{2.21}$$

Substituting (2.19) and (2.21) into the above discrepancy in (2.16) involving the $\mathcal{R}$, $\widehat{\mathcal{D}}_i$, $\widehat{\mathcal{N}}_i$, $\mathcal{D}$ and $\mathcal{N}$ functions we have,

$$\begin{aligned}&\left\|\left\{\mathcal{R}\left(\begin{bmatrix}\widehat{\mathcal{N}}_i(\boldsymbol{U}_i)\\ \widehat{\mathcal{D}}_i(\boldsymbol{U}_i)\end{bmatrix}\right)-\mathcal{R}\left(\begin{bmatrix}\mathcal{N}(\boldsymbol{U}_i)\\ \mathcal{D}(\boldsymbol{U}_i)\end{bmatrix}\right)\right\}I(\mathcal{A}_i)\right\|_S\\ &\quad\leq 4\big(\|\mathcal{N}(\boldsymbol{U}_i)\|_S+1\big)\,\|\mathcal{D}(\boldsymbol{U}_i)^{-1}\|_S\|\widehat{\mathcal{N}}_i(\boldsymbol{U}_i)-\mathcal{N}(\boldsymbol{U}_i)\|_F\\ &\qquad+\Big\{2\big(\|\mathcal{N}(\boldsymbol{U}_i)\|_S+1\big)\,\|\mathcal{D}(\boldsymbol{U}_i)^{-1}\|_S\Big\}^2\|\widehat{\mathcal{D}}_i(\boldsymbol{U}_i)-\mathcal{D}(\boldsymbol{U}_i)\|_F\\ &\quad=4\Big\{\mathcal{W}(\boldsymbol{U}_i)\|\widehat{\mathcal{N}}_i(\boldsymbol{U}_i)-\mathcal{N}(\boldsymbol{U}_i)\|_F+\mathcal{W}(\boldsymbol{U}_i)^2\|\widehat{\mathcal{D}}_i(\boldsymbol{U}_i)-\mathcal{D}(\boldsymbol{U}_i)\|_F\Big\}\end{aligned}$$

where

$$\mathcal{W}(\boldsymbol{U}_i)\equiv\big(\|\mathcal{N}(\boldsymbol{U}_i)\|_S+1\big)\,\|\mathcal{D}(\boldsymbol{U}_i)^{-1}\|_S=\frac{\|\mathcal{N}(\boldsymbol{U}_i)\|_S+1}{\lambda_{\min}\big(\mathcal{D}(\boldsymbol{U}_i)\big)}.$$

The left-hand side of (2.16) can now be re-written as

$$\begin{aligned}
&\frac{1}{mn}\sum_{i=1}^{m} n_i E\left\| E\left[\left\{\mathcal{R}\left(\begin{bmatrix}\widehat{\mathcal{N}}_i(\boldsymbol{U}_i)\\ \widehat{\mathcal{D}}_i(\boldsymbol{U}_i)\end{bmatrix}\right) - \mathcal{R}\left(\begin{bmatrix}\mathcal{N}(\boldsymbol{U}_i)\\ \mathcal{D}(\boldsymbol{U}_i)\end{bmatrix}\right)\right\} I(\mathcal{A}_i)\middle|\mathcal{X}_i\right]\right\|_S \\
&\leq \frac{1}{mn}\sum_{i=1}^{m} n_i E\left[E\left\|\left\{\mathcal{R}\left(\begin{bmatrix}\widehat{\mathcal{N}}_i(\boldsymbol{U}_i)\\ \widehat{\mathcal{D}}_i(\boldsymbol{U}_i)\end{bmatrix}\right) - \mathcal{R}\left(\begin{bmatrix}\mathcal{N}(\boldsymbol{U}_i)\\ \mathcal{D}(\boldsymbol{U}_i)\end{bmatrix}\right)\right\} I(\mathcal{A}_i)\middle|\mathcal{X}_i\right\|_S\right] \\
&\leq \frac{4}{mn}\sum_{i=1}^{m} n_i E\left(E\left[\left\{\mathcal{W}(\boldsymbol{U}_i)\|\widehat{\mathcal{N}}_i(\boldsymbol{U}_i) - \mathcal{N}(\boldsymbol{U}_i)\|_F + \mathcal{W}(\boldsymbol{U}_i)^2\|\widehat{\mathcal{D}}_i(\boldsymbol{U}_i) - \mathcal{D}(\boldsymbol{U}_i)\|_F\right\}\middle|\mathcal{X}_i\right]\right) \\
&\leq \frac{4}{mn}\sum_{i=1}^{m} n_i E\left(E\left[\mathcal{W}(\boldsymbol{U}_i)\|\widehat{\mathcal{N}}_i(\boldsymbol{U}_i) - \mathcal{N}(\boldsymbol{U}_i)\|_F\middle|\mathcal{X}_i\right]\right) \\
&\quad + \frac{4}{mn}\sum_{i=1}^{m} n_i E\left(E\left[\mathcal{W}(\boldsymbol{U}_i)^2\|\widehat{\mathcal{D}}_i(\boldsymbol{U}_i) - \mathcal{D}(\boldsymbol{U}_i)\|_F\middle|\mathcal{X}_i\right]\right) \\
&= \frac{4}{mn}\sum_{i=1}^{m} n_i E\left\{\mathcal{W}(\boldsymbol{U}_i)\|\widehat{\mathcal{N}}_i(\boldsymbol{U}_i) - \mathcal{N}(\boldsymbol{U}_i)\|_F\right\} \\
&\quad + \frac{4}{mn}\sum_{i=1}^{m} n_i E\left\{\mathcal{W}(\boldsymbol{U}_i)^2\|\widehat{\mathcal{D}}_i(\boldsymbol{U}_i) - \mathcal{D}(\boldsymbol{U}_i)\|_F\right\}.
\end{aligned} \tag{2.22}$$

For the first term in the final expression of (2.22), note that,

$$\begin{aligned}
E\left\{\mathcal{W}(\boldsymbol{U}_i)\|\widehat{\mathcal{N}}_i(\boldsymbol{U}_i) - \mathcal{N}(\boldsymbol{U}_i)\|_F\right\} &= E\Big[E\left\{\mathcal{W}(\boldsymbol{U}_i)\|\widehat{\mathcal{N}}_i(\boldsymbol{U}_i) - \mathcal{N}(\boldsymbol{U}_i)\|_F\middle|\boldsymbol{U}_i\right\}\Big] \\
&= E\Big[\mathcal{W}(\boldsymbol{U}_i)E\left\{\|\widehat{\mathcal{N}}_i(\boldsymbol{U}_i) - \mathcal{N}(\boldsymbol{U}_i)\|_F\middle|\boldsymbol{U}_i\right\}\Big].
\end{aligned} \tag{2.23}$$

From a conditional version of the Cauchy-Schwarz inequality,

$$\begin{aligned}
E\left\{\|\widehat{\mathcal{N}}_i(\boldsymbol{U}_i) - \mathcal{N}(\boldsymbol{U}_i)\|_F\middle|\boldsymbol{U}_i\right\} &= \left(\left[E\left\{\|\widehat{\mathcal{N}}_i(\boldsymbol{U}_i) - \mathcal{N}(\boldsymbol{U}_i)\|_F\middle|\boldsymbol{U}_i\right\}\right]^2\right)^{1/2} \\
&\leq \left[E\left\{\|\widehat{\mathcal{N}}_i(\boldsymbol{U}_i) - \mathcal{N}(\boldsymbol{U}_i)\|_F^2\middle|\boldsymbol{U}_i\right\}\right]^{1/2}.
\end{aligned} \tag{2.24}$$

Using (2.24), observe that

$$\left\|\widehat{\mathcal{N}}_i(\boldsymbol{U}_i) - \mathcal{N}(\boldsymbol{U}_i)\right\|_F^2 = \sum_{k=1}^{d_{\mathrm{B}}}\sum_{k'=1}^{d_{\mathrm{A}}}[\{\widehat{\mathcal{N}}_i(\boldsymbol{U}_i) - \mathcal{N}(\boldsymbol{U}_i)\}_{kk'}]^2.$$

Then note that

$$\left[\{\widehat{\mathcal{N}}_i(\boldsymbol{U}_i) - \mathcal{N}(\boldsymbol{U}_i)\}_{kk'}\right]^2 = \left(\frac{1}{n_i}\sum_{j=1}^{n_i}\left[X_{\mathrm{B}ijk}X_{\mathrm{A}ijk'}f(\boldsymbol{X}_{ij},\boldsymbol{U}_i) - E\{X_{\mathrm{B}k}X_{\mathrm{A}k'}f(\boldsymbol{X},\boldsymbol{U}_i)|\boldsymbol{U}_i\}\right]\right)^2$$

from which it follows that

$$
\begin{aligned}
&E\Big(\Big[\{\widehat{\mathcal{N}}_i(\boldsymbol{U}_i) - \mathcal{N}(\boldsymbol{U}_i)\}_{kk'}\Big]^2 \Big| \boldsymbol{U}_i\Big) \\
&= E\left[\left(\frac{1}{n_i} \sum_{j=1}^{n_i} \left[X_{\mathrm{B}ijk} X_{\mathrm{A}ijk'} f(\boldsymbol{X}_{ij}, \boldsymbol{U}_i) - E\{X_{\mathrm{B}k} X_{\mathrm{A}k'} f(\boldsymbol{X}, \boldsymbol{U}_i) | \boldsymbol{U}_i\} \right] \right)^2 \Big| \boldsymbol{U}_i \right] \\
&= \mathrm{Var}\left[\left(\frac{1}{n_i} \sum_{j=1}^{n_i} \left[X_{\mathrm{B}ijk} X_{\mathrm{A}ijk'} f(\boldsymbol{X}_{ij}, \boldsymbol{U}_i) - E\{X_{\mathrm{B}k} X_{\mathrm{A}k'} f(\boldsymbol{X}, \boldsymbol{U}_i) | \boldsymbol{U}_i\} \right] \right) \Big| \boldsymbol{U}_i \right] \\
&= \mathrm{Var}\left[\left(\frac{1}{n_i} \sum_{j=1}^{n_i} X_{\mathrm{B}ijk} X_{\mathrm{A}ijk'} f(\boldsymbol{X}_{ij}, \boldsymbol{U}_i) \right) \Big| \boldsymbol{U}_i \right] \\
&= \frac{1}{n_i^2} \sum_{j=1}^{n_i} \mathrm{Var}\{X_{\mathrm{B}ijk} X_{\mathrm{A}ijk'} f(\boldsymbol{X}_{ij}, \boldsymbol{U}_i) | \boldsymbol{U}_i\} \\
&= \frac{1}{n_i} \mathrm{Var}\{X_{\mathrm{B}k} X_{\mathrm{A}k'} f(\boldsymbol{X}, \boldsymbol{U}) | \boldsymbol{U}\} \\
&= \frac{1}{n_i} \left(E\{X_{\mathrm{B}k}^2 X_{\mathrm{A}k'}^2 f(\boldsymbol{X}, \boldsymbol{U}) | \boldsymbol{U}\} - [E\{X_{\mathrm{B}k} X_{\mathrm{A}k'} f(\boldsymbol{X}, \boldsymbol{U}) | \boldsymbol{U}\}]^2 \right) \\
&\leq \frac{1}{n_i} E\{X_{\mathrm{B}k}^2 X_{\mathrm{A}k'}^2 f(\boldsymbol{X}, \boldsymbol{U}) | \boldsymbol{U}\}.
\end{aligned}
$$

This implies that

$$
\left[E\left\{ \|\widehat{\mathcal{N}}_i(\boldsymbol{U}_i) - \mathcal{N}(\boldsymbol{U}_i)\|_F^2 | \boldsymbol{U}_i \right\} \right]^{1/2} \leq \left[\frac{1}{n_i} E\left\{ \sum_{k=1}^{d_B} \sum_{k'=1}^{d_A} X_{\mathrm{B}k}^2 X_{\mathrm{A}k'}^2 f(\boldsymbol{X}, \boldsymbol{U}) | \boldsymbol{U} \right\} \right]^{1/2}. \tag{2.25}
$$

Substituting (2.25) into (2.24) and substituting (2.24) into (2.23) leads to

$$
\begin{aligned}
&\frac{4}{mn} \sum_{i=1}^{m} n_i E\left\{ \mathcal{W}(\boldsymbol{U}_i) \|\widehat{\mathcal{N}}_i(\boldsymbol{U}_i) - \mathcal{N}(\boldsymbol{U}_i)\|_F \right\} \\
&\leq \frac{4}{mn} \left(\sum_{i=1}^{m} \sqrt{n_i} \right) E\left[\mathcal{W}(\boldsymbol{U}_i) \left\{ E\left(\sum_{k=1}^{d_\mathrm{A}} \sum_{k'=1}^{d_\mathrm{B}} X_{\mathrm{A}k}^2 X_{\mathrm{B}k'}^2 f(\boldsymbol{X}, \boldsymbol{U}) | \boldsymbol{U} \right) \right\}^{1/2} \right] \\
&\leq \frac{4}{\sqrt{\min_{1 \leq i \leq m} (n_i)}} E\left[\mathcal{W}(\boldsymbol{U}) \left\{ E\Big(\|\boldsymbol{X}_\mathrm{B} \boldsymbol{X}_\mathrm{A}^T\|_F^2 f(\boldsymbol{X}, \boldsymbol{U}) | \boldsymbol{U} \Big) \right\}^{1/2} \right] \\
&= \frac{4}{\sqrt{\min_{1 \leq i \leq m} (n_i)}} E\left[\mathcal{W}(\boldsymbol{U}) \left\{ E\Big(\|\boldsymbol{X}_\mathrm{A}\|^2 \|\boldsymbol{X}_\mathrm{B}\|^2 f(\boldsymbol{X}, \boldsymbol{U}) | \boldsymbol{U} \Big) \right\}^{1/2} \right].
\end{aligned} \tag{2.26}
$$

The final expression in (2.26) convergences to zero provided that we assume the moment condition

$$
E\left[\mathcal{W}(\boldsymbol{U}) \left\{ E\Big(\|\boldsymbol{X}_\mathrm{A}\|^2 \|\boldsymbol{X}_\mathrm{B}\|^2 f(\boldsymbol{X}, \boldsymbol{U}) | \boldsymbol{U} \Big) \right\}^{1/2} \right] < \infty. \tag{2.27}
$$

If the spectral norm in $\mathcal{W}(\boldsymbol{U}_i)$ is replaced by the Frobenius norm, then (2.27) leads to

$$E\left[\frac{\left\{\|E\{\boldsymbol{X}_{\mathrm{B}}\boldsymbol{X}_{\mathrm{A}}^T f(\boldsymbol{X},\boldsymbol{U})|\boldsymbol{U}\}\|_F+1\right\}\left\{E\Big(\|\boldsymbol{X}_{\mathrm{A}}\|^2\,\|\boldsymbol{X}_{\mathrm{B}}\|^2 f(\boldsymbol{X},\boldsymbol{U})|\boldsymbol{U}\Big)\right\}^{1/2}}{\lambda_{\min}\Big(E\{\boldsymbol{X}_{\mathrm{A}}\boldsymbol{X}_{\mathrm{A}}^T f(\boldsymbol{X},\boldsymbol{U})|\boldsymbol{U}\}\Big)}\right]<\infty. \tag{2.28}$$

Now note that,

$$\begin{aligned}\|E\{\boldsymbol{X}_{\mathrm{B}}\boldsymbol{X}_{\mathrm{A}}^T f(\boldsymbol{X},\boldsymbol{U})|\boldsymbol{U}\}\|_F &\le E\{\|\boldsymbol{X}_{\mathrm{B}}\boldsymbol{X}_{\mathrm{A}}^T\|_F\, f(\boldsymbol{X},\boldsymbol{U})|\boldsymbol{U}\}\\ &= E\{\|\boldsymbol{X}_{\mathrm{B}}\|\|\boldsymbol{X}_{\mathrm{A}}\|\, f(\boldsymbol{X},\boldsymbol{U})|\boldsymbol{U}\}.\end{aligned}$$

This means that we can replace (2.28) by

$$E\left[\frac{\left\{E\{\|\boldsymbol{X}_{\mathrm{A}}\|\,\|\boldsymbol{X}_{\mathrm{B}}\|\, f(\boldsymbol{X},\boldsymbol{U})|\boldsymbol{U}\}+1\right\}\left\{E\Big(\|\boldsymbol{X}_{\mathrm{A}}\|^2\,\|\boldsymbol{X}_{\mathrm{B}}\|^2 f(\boldsymbol{X},\boldsymbol{U})|\boldsymbol{U}\Big)\right\}^{1/2}}{\lambda_{\min}\Big(E\{\boldsymbol{X}_{\mathrm{A}}\boldsymbol{X}_{\mathrm{A}}^T f(\boldsymbol{X},\boldsymbol{U})|\boldsymbol{U}\}\Big)}\right]<\infty. \tag{2.29}$$

Next we treat part of the second term in the final expression of (2.22). Note that,

$$\begin{aligned}E\left\{\mathcal{W}(\boldsymbol{U}_i)^2\|\widehat{\mathcal{D}}_i(\boldsymbol{U}_i)-\mathcal{D}(\boldsymbol{U}_i)\|_F\right\} &= E\Big[E\left\{\mathcal{W}(\boldsymbol{U}_i)^2\|\widehat{\mathcal{D}}_i(\boldsymbol{U}_i)-\mathcal{D}(\boldsymbol{U}_i)\|_F\big|\boldsymbol{U}_i\right\}\Big]\\ &= E\Big[\mathcal{W}(\boldsymbol{U}_i)^2 E\left\{\|\widehat{\mathcal{D}}_i(\boldsymbol{U}_i)-\mathcal{D}(\boldsymbol{U}_i)\|_F\big|\boldsymbol{U}_i\right\}\Big].\end{aligned} \tag{2.30}$$

By using a conditional version of the Cauchy-Schwarz inequality, we obtain

$$\begin{aligned}E\left\{\|\widehat{\mathcal{D}}_i(\boldsymbol{U}_i)-\mathcal{D}(\boldsymbol{U}_i)\|_F\big|\boldsymbol{U}_i\right\} &= \left(\left[E\left\{\|\widehat{\mathcal{D}}_i(\boldsymbol{U}_i)-\mathcal{D}(\boldsymbol{U}_i)\|_F\big|\boldsymbol{U}_i\right\}\right]^2\right)^{1/2}\\ &\le \left[E\left\{\|\widehat{\mathcal{D}}_i(\boldsymbol{U}_i)-\mathcal{D}(\boldsymbol{U}_i)\|_F^2\big|\boldsymbol{U}_i\right\}\right]^{1/2}.\end{aligned} \tag{2.31}$$

Using (2.31), note that

$$\left\|\widehat{\mathcal{D}}_i(\boldsymbol{U}_i)-\mathcal{D}(\boldsymbol{U}_i)\right\|_F^2=\sum_{k=1}^{d_{\mathrm{A}}}\sum_{k'=1}^{d_{\mathrm{A}}}[\{\widehat{\mathcal{D}}_i(\boldsymbol{U}_i)-\mathcal{D}(\boldsymbol{U}_i)\}_{kk'}]^2.$$

Then note that

$$\left[\{\widehat{\mathcal{D}}_i(\boldsymbol{U}_i)-\mathcal{D}(\boldsymbol{U}_i)\}_{kk'}\right]^2=\left(\frac{1}{n_i}\sum_{j=1}^{n_i}\left[X_{\mathrm{A}ijk}X_{\mathrm{A}ijk'}f(\boldsymbol{X}_{ij},\boldsymbol{U}_i)-E\{X_{\mathrm{A}k}X_{\mathrm{A}k'}f(\boldsymbol{X},\boldsymbol{U}_i)|\boldsymbol{U}_i\}\right]\right)^2.$$

Then we have,

$$
\begin{aligned}
&E\left\{\|\widehat{\mathcal{D}}_i(\boldsymbol{U}_i)-\mathcal{D}(\boldsymbol{U}_i)\|_F\big|\boldsymbol{U}_i\right\}\\
&\quad\le\left[E\left\{\|\widehat{\mathcal{D}}_i(\boldsymbol{U}_i)-\mathcal{D}(\boldsymbol{U}_i)\|_F^2\big|\boldsymbol{U}_i\right\}\right]^{1/2}\\
&\quad=\left\{E\left(\sum_{k=1}^{d_{\mathrm{A}}}\sum_{k'=1}^{d_{\mathrm{A}}}[\{\widehat{\mathcal{D}}_i(\boldsymbol{U}_i)-\mathcal{D}(\boldsymbol{U}_i)\}_{kk'}]^2\Big|\boldsymbol{U}_i\right)\right\}^{1/2}\\
&\quad=\left[\sum_{k=1}^{d_{\mathrm{A}}}\sum_{k'=1}^{d_{\mathrm{A}}}E\left\{\left(\frac{1}{n_i}\sum_{j=1}^{n_i}\left[X_{\mathrm{A}ijk}X_{\mathrm{A}ijk'}f(\boldsymbol{X}_{ij},\boldsymbol{U}_i)-E\{X_{\mathrm{A}k}X_{\mathrm{A}k'}f(\boldsymbol{X},\boldsymbol{U}_i)|\boldsymbol{U}_i\}\right]\right)^2\Big|\boldsymbol{U}_i\right\}\right]^{1/2}\\
&\quad=\left[\frac{1}{n_i}\sum_{k=1}^{d_{\mathrm{A}}}\sum_{k'=1}^{d_{\mathrm{A}}}\mathrm{Var}\{X_{\mathrm{A}k}X_{\mathrm{A}k'}f(\boldsymbol{X},\boldsymbol{U})|\boldsymbol{U}\}\right]^{1/2}\\
&\quad\le\left[\frac{1}{n_i}\sum_{k=1}^{d_{\mathrm{A}}}\sum_{k'=1}^{d_{\mathrm{A}}}E\{X_{\mathrm{A}k}^2X_{\mathrm{A}k'}^2f(\boldsymbol{X},\boldsymbol{U})|\boldsymbol{U}\}\right]^{1/2}.
\end{aligned}
\tag{2.32}
$$

By substituting (2.32) into (2.30), we have that

$$
\begin{aligned}
&\frac{4}{mn}\sum_{i=1}^{m}n_iE\left\{\mathcal{W}(\boldsymbol{U}_i)^2\|\widehat{\mathcal{D}}_i(\boldsymbol{U}_i)-\mathcal{D}(\boldsymbol{U}_i)\|_F\right\}\\
&\quad\le\frac{4}{mn}\left(\sum_{i=1}^{m}\sqrt{n_i}\right)E\left[\mathcal{W}(\boldsymbol{U})^2\left\{E\left(\sum_{k=1}^{d_{\mathrm{A}}}\sum_{k'=1}^{d_{\mathrm{A}}}X_{\mathrm{A}k}^2X_{\mathrm{A}k'}^2f(\boldsymbol{X},\boldsymbol{U})|\boldsymbol{U}\right)\right\}^{1/2}\right]\\
&\quad\le\frac{4}{\sqrt{\min_{1\le i\le m}(n_i)}}E\left[\mathcal{W}(\boldsymbol{U})^2\left\{E\Big(\|\boldsymbol{X}_{\mathrm{A}}\boldsymbol{X}_{\mathrm{A}}^T\|_F^2f(\boldsymbol{X},\boldsymbol{U})|\boldsymbol{U}\Big)\right\}^{1/2}\right]\\
&\quad=\frac{4}{\sqrt{\min_{1\le i\le m}(n_i)}}E\left[\mathcal{W}(\boldsymbol{U})^2\left\{E\Big(\|\boldsymbol{X}_{\mathrm{A}}\|^4f(\boldsymbol{X},\boldsymbol{U})|\boldsymbol{U}\Big)\right\}^{1/2}\right]
\end{aligned}
$$

which convergences to zero provided that we assume the moment condition

$$
E\left[\mathcal{W}(\boldsymbol{U})^2\left\{E\Big(\|\boldsymbol{X}_{\mathrm{A}}\|^4f(\boldsymbol{X},\boldsymbol{U})|\boldsymbol{U}\Big)\right\}^{1/2}\right]<\infty. \tag{2.33}
$$

If the spectral norm in $\mathcal{W}(\boldsymbol{U}_i)$ is replaced by the Frobenius norm, then (2.33) leads to

$$
E\left[\frac{\left\{\|E\{\boldsymbol{X}_{\mathrm{A}}\boldsymbol{X}_{\mathrm{A}}^Tf(\boldsymbol{X},\boldsymbol{U})|\boldsymbol{U}\}\|_F+1\right\}^2\left\{E\Big(\|\boldsymbol{X}_{\mathrm{A}}\|^4f(\boldsymbol{X},\boldsymbol{U})|\boldsymbol{U}\Big)\right\}^{1/2}}{\lambda_{\min}\Big(E\{\boldsymbol{X}_{\mathrm{A}}\boldsymbol{X}_{\mathrm{A}}^Tf(\boldsymbol{X},\boldsymbol{U})|\boldsymbol{U}\}\Big)^2}\right]<\infty. \tag{2.34}
$$

Now note that

$$
\begin{aligned}
\|E\{\boldsymbol{X}_{\mathrm{A}}\boldsymbol{X}_{\mathrm{A}}^Tf(\boldsymbol{X},\boldsymbol{U})|\boldsymbol{U}\}\|_F&\le E\{\|\boldsymbol{X}_{\mathrm{A}}\boldsymbol{X}_{\mathrm{A}}^T\|_Ff(\boldsymbol{X},\boldsymbol{U})|\boldsymbol{U}\}\\
&=E\{\|\boldsymbol{X}_{\mathrm{A}}\|^2f(\boldsymbol{X},\boldsymbol{U})|\boldsymbol{U}\}.
\end{aligned}
$$

This means that we can replace (2.34) by

$$E\left[\frac{\left[E\{\|\boldsymbol{X}_{\mathrm{A}}\|^2 f(\boldsymbol{X},\boldsymbol{U})|\boldsymbol{U}\}+1\right]^2\left\{E\Big(\|\boldsymbol{X}_{\mathrm{A}}\|^4 f(\boldsymbol{X},\boldsymbol{U})|\boldsymbol{U}\Big)\right\}^{1/2}}{\lambda_{\min}\Big(E\{\boldsymbol{X}_{\mathrm{A}}\boldsymbol{X}_{\mathrm{A}}^T f(\boldsymbol{X},\boldsymbol{U})|\boldsymbol{U}\}\Big)^2}\right]<\infty. \tag{2.35}$$

2.4.2.8 Proof of Result (2.17)

In this subsubsection we aim to prove (2.17), which is

$$\frac{1}{mn}\sum_{i=1}^{m} n_i E\left\|E\left[\left\{\mathcal{R}\left(\begin{bmatrix}\widehat{\mathcal{N}}_i(\boldsymbol{U}_i)\\ \widehat{\mathcal{D}}_i(\boldsymbol{U}_i)\end{bmatrix}\right)-\mathcal{R}\left(\begin{bmatrix}\mathcal{N}(\boldsymbol{U}_i)\\ \mathcal{D}(\boldsymbol{U}_i)\end{bmatrix}\right)\right\} I(\mathcal{A}_i^C)\middle|\mathcal{X}_i\right]\right\|_S \to 0$$

as $m,n\to\infty$. Note that throughout this subsubsection, we are considering

$$(\boldsymbol{U}_i,\mathcal{X}_i) \text{ such that } \mathcal{A}_i^C \text{ occurs}, 1\le i\le m. \tag{2.36}$$

We first start with

$$\begin{aligned}
&E\left\|E\left[\left\{\mathcal{R}\left(\begin{bmatrix}\widehat{\mathcal{N}}_i(\boldsymbol{U}_i)\\ \widehat{\mathcal{D}}_i(\boldsymbol{U}_i)\end{bmatrix}\right)-\mathcal{R}\left(\begin{bmatrix}\mathcal{N}(\boldsymbol{U}_i)\\ \mathcal{D}(\boldsymbol{U}_i)\end{bmatrix}\right)\right\} I(\mathcal{A}_i^C)\middle|\mathcal{X}_i\right]\right\|_S\\
&\le E\left(E\left[\left\|\mathcal{R}\left(\begin{bmatrix}\widehat{\mathcal{N}}_i(\boldsymbol{U}_i)\\ \widehat{\mathcal{D}}_i(\boldsymbol{U}_i)\end{bmatrix}\right)\right\|_S I(\mathcal{A}_i^C)\middle|\mathcal{X}_i\right]\right)+E\left(E\left[\left\|\mathcal{R}\left(\begin{bmatrix}\mathcal{N}(\boldsymbol{U}_i)\\ \mathcal{D}(\boldsymbol{U}_i)\end{bmatrix}\right)\right\|_S I(\mathcal{A}_i^C)\middle|\mathcal{X}_i\right]\right)\\
&= E\left\{\left\|\mathcal{R}\left(\begin{bmatrix}\widehat{\mathcal{N}}_i(\boldsymbol{U}_i)\\ \widehat{\mathcal{D}}_i(\boldsymbol{U}_i)\end{bmatrix}\right)\right\|_S I(\mathcal{A}_i^C)\right\}+E\left\{\left\|\mathcal{R}\left(\begin{bmatrix}\mathcal{N}(\boldsymbol{U}_i)\\ \mathcal{D}(\boldsymbol{U}_i)\end{bmatrix}\right)\right\|_S I(\mathcal{A}_i^C)\right\}\\
&\le \left[E\left\{\left\|\mathcal{R}\left(\begin{bmatrix}\widehat{\mathcal{N}}_i(\boldsymbol{U}_i)\\ \widehat{\mathcal{D}}_i(\boldsymbol{U}_i)\end{bmatrix}\right)\right\|_S^2\right\}\right]^{1/2} P(\mathcal{A}_i^C)^{1/2}+\left[E\left\{\left\|\mathcal{R}\left(\begin{bmatrix}\mathcal{N}(\boldsymbol{U}_i)\\ \mathcal{D}(\boldsymbol{U}_i)\end{bmatrix}\right)\right\|_S^2\right\}\right]^{1/2} P(\mathcal{A}_i^C)^{1/2}.
\end{aligned}$$

To deal with the following expression

$$E\left\{\left\|\mathcal{R}\left(\begin{bmatrix}\widehat{\mathcal{N}}_i(\boldsymbol{U}_i)\\ \widehat{\mathcal{D}}_i(\boldsymbol{U}_i)\end{bmatrix}\right)\right\|_S^2\right\},$$

first recall that

$$\begin{aligned}
\left\|\mathcal{R}\left(\begin{bmatrix}\widehat{\mathcal{N}}_i(\boldsymbol{U}_i)\\ \widehat{\mathcal{D}}_i(\boldsymbol{U}_i)\end{bmatrix}\right)\right\|_S &\le \sqrt{d_B}\left\|\widehat{\mathcal{N}}_i(\boldsymbol{U}_i)\widehat{\mathcal{D}}_i(\boldsymbol{U}_i)^{-1}\widehat{\mathcal{N}}_i(\boldsymbol{U}_i)^T\right\|_S\\
&= \sqrt{d_B}\left\|\left\{\frac{1}{n_i}\sum_{j=1}^{n}\boldsymbol{X}_{\mathrm{B}ij}\boldsymbol{X}_{\mathrm{A}ij}^T f(\boldsymbol{X}_{ij},\boldsymbol{U}_i)\right\}\left\{\frac{1}{n_i}\sum_{j=1}^{n}\boldsymbol{X}_{\mathrm{A}ij}\boldsymbol{X}_{\mathrm{A}ij}^T f(\boldsymbol{X}_{ij},\boldsymbol{U}_i)\right\}^{-1}\right.\\
&\quad\left.\times\left\{\frac{1}{n_i}\sum_{j=1}^{n}\boldsymbol{X}_{\mathrm{A}ij}\boldsymbol{X}_{\mathrm{B}ij}^T f(\boldsymbol{X}_{ij},\boldsymbol{U}_i)\right\}\right\|_S.
\end{aligned}$$

Now we appeal to the second-last displayed equation on page 1093 of Chipman (1964), which is referred to as the *generalized Schwarz inequality*, to justify:

$$\left\| \mathcal{R}\left(\begin{bmatrix} \widehat{\mathcal{N}}_i(\boldsymbol{U}_i) \\ \widehat{\mathcal{D}}_i(\boldsymbol{U}_i) \end{bmatrix} \right) \right\|_S^2 \leq \sqrt{d_B} \left\| \frac{1}{n_i} \sum_{j=1}^{n} \boldsymbol{X}_{\mathrm{B}ij} \boldsymbol{X}_{\mathrm{B}ij}^T f(\boldsymbol{X}_{ij}, \boldsymbol{U}_i) \right\|_S^2 .$$

Hence,

$$\begin{aligned} E\left\{ \left\| \mathcal{R}\left(\begin{bmatrix} \widehat{\mathcal{N}}_i(\boldsymbol{U}_i) \\ \widehat{\mathcal{D}}_i(\boldsymbol{U}_i) \end{bmatrix} \right) \right\|_S^2 \right\} &\leq \frac{d_B}{n_i} \sum_{j=1}^{n} E\big\{ \left\| \boldsymbol{X}_{\mathrm{B}ij} \boldsymbol{X}_{\mathrm{B}ij}^T f(\boldsymbol{X}_{ij}, \boldsymbol{U}_i) \right\|_F^2 \big\} \\ &= d_B E\big\{ \|\boldsymbol{X}_{\mathrm{B}}\|^4 f(\boldsymbol{X}, \boldsymbol{U})^2 \big\}. \end{aligned}$$

Next, to deal with

$$E\left\{ \left\| \mathcal{R}\left(\begin{bmatrix} \mathcal{N}(\boldsymbol{U}_i) \\ \mathcal{D}(\boldsymbol{U}_i) \end{bmatrix} \right) \right\|_S^2 \right\}$$

note that

$$\begin{aligned} \left\| \mathcal{R}\left(\begin{bmatrix} \mathcal{N}(\boldsymbol{U}_i) \\ \mathcal{D}(\boldsymbol{U}_i) \end{bmatrix} \right) \right\|_S &\leq \sqrt{d_B} \left\| \mathcal{N}(\boldsymbol{U}_i) \mathcal{D}(\boldsymbol{U}_i)^{-1} \mathcal{N}(\boldsymbol{U}_i)^T \right\|_S \\ &\leq \sqrt{d_B} \left\| \mathcal{N}(\boldsymbol{U}_i) \right\|_S \left\| \mathcal{D}(\boldsymbol{U}_i)^{-1} \right\|_S \left\| \mathcal{N}(\boldsymbol{U}_i) \right\|_S \\ &= \frac{\sqrt{d_B} \left\| \mathcal{N}(\boldsymbol{U}_i) \right\|_S^2}{\lambda_{\min}\big(\mathcal{D}(\boldsymbol{U}_i)\big)}. \end{aligned}$$

Hence,

$$E\left\{ \left\| \mathcal{R}\left(\begin{bmatrix} \mathcal{N}(\boldsymbol{U}_i) \\ \mathcal{D}(\boldsymbol{U}_i) \end{bmatrix} \right) \right\|_S^2 \right\} \leq d_B E\left\{ \frac{\left\| \mathcal{N}(\boldsymbol{U}) \right\|_S^4}{\lambda_{\min}\big(\mathcal{D}(\boldsymbol{U})\big)^2} \right\}.$$

Finally, note that

$$\begin{aligned} \left\| \mathcal{N}(\boldsymbol{U}) \right\|_S^4 &\leq \left\| E\{ \boldsymbol{X}_{\mathrm{B}} \boldsymbol{X}_{\mathrm{A}}^T f(\boldsymbol{X}, \boldsymbol{U}) | \boldsymbol{U} \} \right\|_F^4 \\ &\leq E\{ \left\| \boldsymbol{X}_{\mathrm{B}} \boldsymbol{X}_{\mathrm{A}}^T \right\|_F^4 f(\boldsymbol{X}, \boldsymbol{U}) | \boldsymbol{U} \} \\ &= E\{ \|\boldsymbol{X}_{\mathrm{A}}\|^4 \|\boldsymbol{X}_{\mathrm{B}}\|^4 f(\boldsymbol{X}, \boldsymbol{U}) | \boldsymbol{U} \} \end{aligned}$$

which implies that

$$E\left\{ \frac{\left\| \mathcal{N}(\boldsymbol{U}) \right\|_S^4}{\lambda_{\min}\big(\mathcal{D}(\boldsymbol{U})\big)^2} \right\} \leq E\left\{ \frac{E\{ \|\boldsymbol{X}_{\mathrm{A}}\|^4 \|\boldsymbol{X}_{\mathrm{B}}\|^4 f(\boldsymbol{X}, \boldsymbol{U}) | \boldsymbol{U} \}}{\lambda_{\min}\big(\mathcal{D}(\boldsymbol{U})\big)^2} \right\}.$$

From(2.36), note that the following holds

$$\begin{aligned} P(\mathcal{A}_i^C) \leq\ & P\big(\|\widehat{\mathcal{N}}_i(\boldsymbol{U}_i) - \mathcal{N}(\boldsymbol{U}_i)\|_S > 1 \big) \\ & + P\Big(\big| \lambda_{\min}\big(\widehat{\mathcal{D}}_i(\boldsymbol{U}_i)\big) - \lambda_{\min}\big(\mathcal{D}(\boldsymbol{U}_i)\big) \big| > \tfrac{1}{2} \lambda_{\min}\big(\mathcal{D}(\boldsymbol{U}_i)\big) \Big). \end{aligned} \tag{2.37}$$

Using Markov's inequality and the Cauchy-Schwarz inequality, we have

$$\begin{aligned} P\big(\|\widehat{\mathcal{N}}_i(\boldsymbol{U}_i) - \mathcal{N}(\boldsymbol{U}_i)\|_S > 1\big) &\leq P\big(\|\widehat{\mathcal{N}}_i(\boldsymbol{U}_i) - \mathcal{N}(\boldsymbol{U}_i)\|_S^2 > 1\big) \\ &\leq E\{\|\widehat{\mathcal{N}}_i(\boldsymbol{U}_i) - \mathcal{N}(\boldsymbol{U}_i)\|_S^2\} \\ &\leq E\{\|\widehat{\mathcal{N}}_i(\boldsymbol{U}_i) - \mathcal{N}(\boldsymbol{U}_i)\|_F^2\} \\ &= E\big[E\{\|\widehat{\mathcal{N}}_i(\boldsymbol{U}_i) - \mathcal{N}(\boldsymbol{U}_i)\|_F^2|\boldsymbol{U}_i\}\big] \\ &\leq \frac{1}{n_i}\sum_{k=1}^{d_B}\sum_{k'=1}^{d_A} E\left[E\left\{X_{\mathrm{B}k}^2 X_{\mathrm{A}k'}^2 f(\boldsymbol{X},\boldsymbol{U})|\boldsymbol{U}\right\}\right] \\ &= \frac{1}{n_i} E\left[E\left\{\|\boldsymbol{X}_{\mathrm{A}}\|^2\|\boldsymbol{X}_{\mathrm{B}}\|^2 f(\boldsymbol{X},\boldsymbol{U})|\boldsymbol{U}\right\}\right]. \end{aligned} \tag{2.38}$$

Using Markov's inequality again as well as Theorem 8.1.4 (Wielandt-Hoffman) of Golub and Van Loan (2013), we obtain the following expression for the second term in (2.37),

$$\begin{aligned} &P\Big(\big|\lambda_{\min}\big(\widehat{\mathcal{D}}_i(\boldsymbol{U}_i)\big) - \lambda_{\min}\big(\mathcal{D}(\boldsymbol{U}_i)\big)\big| > \tfrac{1}{2}\lambda_{\min}(\mathcal{D}(\boldsymbol{U}_i))\Big) \\ &\quad= E\Big\{P\Big(\big|\lambda_{\min}\big(\widehat{\mathcal{D}}_i(\boldsymbol{U}_i)\big) - \lambda_{\min}\big(\mathcal{D}(\boldsymbol{U}_i)\big)\big| > \tfrac{1}{2}\lambda_{\min}(\mathcal{D}(\boldsymbol{U}_i))\big|\boldsymbol{U}_i\Big)\Big\} \\ &\quad= E\Big\{P\Big(\big\{\lambda_{\min}\big(\widehat{\mathcal{D}}_i(\boldsymbol{U}_i)\big) - \lambda_{\min}\big(\mathcal{D}(\boldsymbol{U}_i)\big)\big\}^2 > \tfrac{1}{4}\lambda_{\min}(\mathcal{D}(\boldsymbol{U}_i))^2\big|\boldsymbol{U}_i\Big)\Big\} \\ &\quad\leq 4E\left(\frac{E\Big[\big\{\lambda_{\min}\big(\widehat{\mathcal{D}}_i(\boldsymbol{U}_i)\big) - \lambda_{\min}\big(\mathcal{D}(\boldsymbol{U}_i)\big)\big\}^2\Big|\boldsymbol{U}_i\Big]}{\lambda_{\min}(\mathcal{D}(\boldsymbol{U}_i))^2}\right) \\ &\quad\leq 4E\left(\frac{E\Big\{\big\|\widehat{\mathcal{D}}_i(\boldsymbol{U}_i) - \mathcal{D}(\boldsymbol{U}_i)\big\|_F^2\Big|\boldsymbol{U}_i\Big\}}{\lambda_{\min}(\mathcal{D}(\boldsymbol{U}_i))^2}\right). \end{aligned}$$

From earlier calculations, we have that

$$\begin{aligned} E\Big\{\big\|\widehat{\mathcal{D}}_i(\boldsymbol{U}_i) - \mathcal{D}(\boldsymbol{U}_i)\big\|_F^2\Big|\boldsymbol{U}_i\Big\} &\leq \frac{1}{n_i}\sum_{k=1}^{d_{\mathrm{A}}}\sum_{k'=1}^{d_{\mathrm{A}}} E\{X_{\mathrm{A}k}^2 X_{\mathrm{A}k'}^2 f(\boldsymbol{X},\boldsymbol{U})|\boldsymbol{U}\} \\ &= \frac{1}{n_i} E\Big(\|\boldsymbol{X}_{\mathrm{A}}\|^4 f(\boldsymbol{X},\boldsymbol{U})|\boldsymbol{U}\Big). \end{aligned}$$

Therefore,

$$P\Big(\big|\lambda_{\min}\big(\widehat{\mathcal{D}}_i(\boldsymbol{U}_i)\big) - \lambda_{\min}\big(\mathcal{D}(\boldsymbol{U}_i)\big)\big| > \tfrac{1}{2}\lambda_{\min}(\mathcal{D}(\boldsymbol{U}_i))\Big) \leq \frac{4}{n_i} E\left[\frac{E\{\|\boldsymbol{X}_{\mathrm{A}}\|^4 f(\boldsymbol{X},\boldsymbol{U})|\boldsymbol{U}\}}{\lambda_{\min}(\mathcal{D}(\boldsymbol{U}))^2}\right]. \tag{2.39}$$

Combining (2.38) and (2.39), we obtain

$$P(\mathcal{A}_i^C) \leq \frac{1}{n_i}\left(E\Big[E\{\|\boldsymbol{X}_{\mathrm{A}}\|\,\|\boldsymbol{X}_{\mathrm{B}}\| f(\boldsymbol{X},\boldsymbol{U})|\boldsymbol{U}\}\Big] + 4E\left[\frac{E\{\|\boldsymbol{X}_{\mathrm{A}}\|^4 f(\boldsymbol{X},\boldsymbol{U})|\boldsymbol{U}\}}{\lambda_{\min}(\mathcal{D}(\boldsymbol{U}))^2}\right]\right).$$

Using each of the several bounds established in this subsubsection, we have

$$
\begin{aligned}
&\frac{1}{mn}\sum_{i=1}^{m} n_i E\left\| E\left[\left\{\mathcal{R}\left(\begin{bmatrix}\widehat{\mathcal{N}}_i(\boldsymbol{U}_i)\\ \widehat{\mathcal{D}}_i(\boldsymbol{U}_i)\end{bmatrix}\right)-\mathcal{R}\left(\begin{bmatrix}\mathcal{N}(\boldsymbol{U}_i)\\ \mathcal{D}(\boldsymbol{U}_i)\end{bmatrix}\right)\right\} I(\mathcal{A}_i^C)\middle|\mathcal{X}_i\right]\right\|_S \\
&\quad\le \frac{\sqrt{d_B}}{mn}\sum_{i=1}^{m}\left[n_i\left(\left[E\left\{\|\boldsymbol{X}_{\mathrm{B}}\|^4 f(\boldsymbol{X},\boldsymbol{U})^2\right\}\right]^{1/2}+\left[E\left\{\frac{E\{\|\boldsymbol{X}_{\mathrm{A}}\|^4\|\boldsymbol{X}_{\mathrm{B}}\|^4 f(\boldsymbol{X},\boldsymbol{U})|\boldsymbol{U}\}}{\lambda_{\min}\big(\mathcal{D}(\boldsymbol{U})\big)^2}\right\}\right]^{1/2}\right)\right. \\
&\quad\left.\times\left\{\frac{1}{n_i}\left(E\Big[E\{\|\boldsymbol{X}_{\mathrm{A}}\|^2\|\boldsymbol{X}_{\mathrm{B}}\|^2 f(\boldsymbol{X},\boldsymbol{U})|\boldsymbol{U}\}\Big]+4E\left[\frac{E\{\|\boldsymbol{X}_{\mathrm{A}}\|^4 f(\boldsymbol{X},\boldsymbol{U})|\boldsymbol{U}\}}{\lambda_{\min}(\mathcal{D}(\boldsymbol{U}))^2}\right]\right)\right\}^{1/2}\right] \\
&\quad\le \frac{\sqrt{d_B}}{\sqrt{\min\limits_{1\le i\le m}(n_i)}}\left[\left(\left[E\left\{\|\boldsymbol{X}_{\mathrm{B}}\|^4 f(\boldsymbol{X},\boldsymbol{U})^2\right\}\right]^{1/2}\right)\right. \\
&\quad\left.\times\left\{\left(E\Big[E\{\|\boldsymbol{X}_{\mathrm{A}}\|\,\|\boldsymbol{X}_{\mathrm{B}}\| f(\boldsymbol{X},\boldsymbol{U})|\boldsymbol{U}\}\Big]+4E\left[\frac{E\{\|\boldsymbol{X}_{\mathrm{A}}\|^4 f(\boldsymbol{X},\boldsymbol{U})|\boldsymbol{U}\}}{\lambda_{\min}(\mathcal{D}(\boldsymbol{U}))^2}\right]\right)\right\}^{1/2}\right] \\
&\quad+\frac{\sqrt{d_B}}{\sqrt{\min\limits_{1\le i\le m}(n_i)}}\left[\left(\left[E\left\{\frac{E\{\|\boldsymbol{X}_{\mathrm{A}}\|^4\|\boldsymbol{X}_{\mathrm{B}}\|^4 f(\boldsymbol{X},\boldsymbol{U})|\boldsymbol{U}\}}{\lambda_{\min}\big(\mathcal{D}(\boldsymbol{U})\big)^2}\right\}\right]^{1/2}\right)\right. \\
&\quad\left.\times\left\{\left(E\Big[E\{\|\boldsymbol{X}_{\mathrm{A}}\|\,\|\boldsymbol{X}_{\mathrm{B}}\| f(\boldsymbol{X},\boldsymbol{U})|\boldsymbol{U}\}\Big]+4E\left[\frac{E\{\|\boldsymbol{X}_{\mathrm{A}}\|^4 f(\boldsymbol{X},\boldsymbol{U})|\boldsymbol{U}\}}{\lambda_{\min}(\mathcal{D}(\boldsymbol{U}))^2}\right]\right)\right\}^{1/2}\right],
\end{aligned}
$$

which converges to zero under Theorems 11, 12 and 14 sample size and moment assumptions.

2.4.2.9 Summary of Moment Assumptions

The moment assumptions used to prove (2.10), are as follows:

$$
\begin{aligned}
&\text{(MA1)}\quad E\left\{\|\boldsymbol{X}_{\mathrm{B}}\|^4 f(\boldsymbol{X},\boldsymbol{U})^2\right\}<\infty,\\
&\text{(MA2)}\quad E\left[\frac{\left\{E\Big(\|\boldsymbol{X}_{\mathrm{A}}\|^2\|\boldsymbol{X}_{\mathrm{B}}\|^2 f(\boldsymbol{X},\boldsymbol{U})|\boldsymbol{U}\Big)\right\}^{1/2}\{E\{\|\boldsymbol{X}_{\mathrm{A}}\|\,\|\boldsymbol{X}_{\mathrm{B}}\| f(\boldsymbol{X},\boldsymbol{U})|\boldsymbol{U}\}+1\}}{\lambda_{\min}\Big(E\{\boldsymbol{X}_{\mathrm{A}}\boldsymbol{X}_{\mathrm{A}}^T f(\boldsymbol{X},\boldsymbol{U})|\boldsymbol{U}\}\Big)}\right]<\infty,\\
&\text{(MA3)}\quad E\left[\frac{\left\{E\Big(\|\boldsymbol{X}_{\mathrm{A}}\|^4 f(\boldsymbol{X},\boldsymbol{U})|\boldsymbol{U}\Big)\right\}^{1/2}\left[E\{\|\boldsymbol{X}_{\mathrm{A}}\|^2 f(\boldsymbol{X},\boldsymbol{U})|\boldsymbol{U}\}+1\right]^2}{\lambda_{\min}\Big(E\{\boldsymbol{X}_{\mathrm{A}}\boldsymbol{X}_{\mathrm{A}}^T f(\boldsymbol{X},\boldsymbol{U})|\boldsymbol{U}\}\Big)^2}\right]<\infty,\\
&\text{(MA4)}\quad E\Big[E\{\|\boldsymbol{X}_{\mathrm{A}}\|^2\|\boldsymbol{X}_{\mathrm{B}}\|^2 f(\boldsymbol{X},\boldsymbol{U})|\boldsymbol{U}\}\Big]<\infty,\\
&\text{(MA5)}\quad E\left\{\frac{E\{\|\boldsymbol{X}_{\mathrm{A}}\|^4 f(\boldsymbol{X},\boldsymbol{U})|\boldsymbol{U}\}}{\lambda_{\min}\Big(E\{\boldsymbol{X}_{\mathrm{A}}\boldsymbol{X}_{\mathrm{A}}^T f(\boldsymbol{X},\boldsymbol{U})|\boldsymbol{U}\}\Big)^2}\right\}<\infty,\\
&\text{(MA6)}\quad E\left\{\frac{E\{\|\boldsymbol{X}_{\mathrm{A}}\|^4\|\boldsymbol{X}_{\mathrm{B}}\|^4 f(\boldsymbol{X},\boldsymbol{U})|\boldsymbol{U}\}}{\lambda_{\min}\Big(E\{\boldsymbol{X}_{\mathrm{A}}\boldsymbol{X}_{\mathrm{A}}^T f(\boldsymbol{X},\boldsymbol{U})|\boldsymbol{U}\}\Big)^2}\right\}<\infty.
\end{aligned}
$$

2.4.2.10 Succinct Expression for Moment Assumptions

We now attempt to find an alternative way to express the moment assumptions in Section 2.4.2.9 in a unified fashion.

Inequality to Re-write the Numerator of (MA2)

To re-express the moment assumption (MA2), we intend to apply the following inequality to the numerator of (MA2):

$$(x+1)y < 1 + x^2 + y^2 \quad \text{for all } x, y \in \mathbb{R}. \tag{2.40}$$

One way to prove (2.40) is to write it as

$$x^2 + y^2 - y - xy + 1 > 0 \quad \text{for all } x, y \in \mathbb{R}. \tag{2.41}$$

If the left-hand side is then written as

$$A\,x^2 + B\,y^2 + C\,x + D\,y + E\,xy + F$$

where

$$A = 1, \quad B = 1, \quad C = 0, \quad D = -1, \quad E = -1, \quad F = 1.$$

The discriminant-type quantity for the quadratic expression is

$$4AB - E^2 = 4 - (-1)^2 = 3 > 0$$

and since $A > 0$, the left-hand side of (2.41) has a minimum at

$$(x, y) = (DE - 2BC, CE - 2AD)/(4AB - E^2) = (1/3, 2/3).$$

Substitution of this point into the left-hand side of (2.41) leads to

$$(1/9) + (4/9) - (2/3) - (2/9) + 1 = (1/9) + (4/9) - (6/9) - (2/9) + (9/9) = 2/3 > 0.$$

Thus, we have established (2.40).

Inequality to Re-write the Numerator of (MA3)

Next, to re-express the moment assumption (MA3), we intend to apply the following inequality to the numerator of (MA3):

$$(x+1)^2 y < 1 + x^2 + x^4 + 2y^2 \quad \text{for all } x, y \in \mathbb{R}. \tag{2.42}$$

We first start by expanding the term on the left-hand side of (2.42) which leads to

$$\begin{aligned} (x+1)^2 y &= (x^2 + 2x + y)y \\ &= (x^2 + 1)y + 2xy. \end{aligned} \tag{2.43}$$

Also note that, by using the inequality $(x-y)^2 \geq 0$, we obtain the following

$$2xy \leq x^2 + y^2. \tag{2.44}$$

Hence, the result in (2.42) can be obtained using the inequalities in (2.40), (2.43) and (2.44).

Re-expressing the Complete Set of Moment Conditions

Now we write

$$M_{p_1p_2} = M_{p_1p_2}(\boldsymbol{U}) \equiv E\big(\|\boldsymbol{X}_{\mathrm{A}}\|^{p_1}\,\|\boldsymbol{X}_{\mathrm{B}}\|^{p_2} f(\boldsymbol{X},\boldsymbol{U})|\boldsymbol{U}\big)\}.$$

Then the application of (2.40) to (MA2) and application of (2.42) to (MA3) leads to

$$M_{22}^{1/2}(M_{11}+1) \leq 1 + M_{11}^2 + M_{22}$$

and

$$M_{40}^{1/2}(M_{20}+1)^2 \leq 1 + M_{20}^2 + M_{20}^4 + 2M_{40}.$$

Next, noting that

$$1 \geq \min(1,x),\ x \geq \min(1,x) \quad \text{and} \quad x^2 \geq \{\min(1,x)\}^2 \quad \text{for all } x>0,$$

the following inequalities can be applied to the denominators of (MA2)–(MA6):

$$\frac{1}{1} \leq \frac{1}{\{\min(1,x)\}^2}, \quad \frac{1}{x} \leq \frac{1}{\{\min(1,x)\}^2} \quad \text{and} \quad \frac{1}{x^2} \leq \frac{1}{\{\min(1,x)\}^2} \quad \text{for all } x>0.$$

Combining all of these facts leads to the following alternative set of moment assumptions:

$$E\left[\frac{\big\{E\big(\|\boldsymbol{X}_{\mathrm{A}}\|^{p_1}\,\|\boldsymbol{X}_{\mathrm{B}}\|^{p_2} f(\boldsymbol{X},\boldsymbol{U})|\boldsymbol{U}\big)\big\}^{p_3}}{\big[\min\big\{1,\lambda_{\min}\big(E\{\boldsymbol{X}_{\mathrm{A}}\boldsymbol{X}_{\mathrm{A}}^T f(\boldsymbol{X},\boldsymbol{U})|\boldsymbol{U}\}\big)\big\}\big]^2}\right] < \infty$$

for each of $(p_1,p_2,p_3) \in \{(0,0,0),(4,0,1),(2,2,1),(4,4,1),(1,1,2),(2,0,2),(2,0,4)\}$.

2.4.2.11 A Sufficient Condition for the Moment Assumptions

Let $\boldsymbol{X} = (\boldsymbol{X}_{\mathrm{A}}, \boldsymbol{X}_{\mathrm{B}})$. Then we have $\|\boldsymbol{X}\| > \max(\|\boldsymbol{X}_{\mathrm{A}}\|, \|\boldsymbol{X}_{\mathrm{B}}\|)$. Hence,

$$\begin{aligned}
\big\{E\big(\|\boldsymbol{X}_{\mathrm{A}}\|^{p_1}\,\|\boldsymbol{X}_{\mathrm{B}}\|^{p_2} f(\boldsymbol{X},\boldsymbol{U})|\boldsymbol{U}\big)\big\}^{p_3} &\leq \big\{E\big(\|\boldsymbol{X}\|^{p_1+p_2} f(\boldsymbol{X},\boldsymbol{U})|\boldsymbol{U}\big)\big\}^{p_3} \\
&\leq E\big(\|\boldsymbol{X}\|^{(p_1+p_2)p_3} f(\boldsymbol{X},\boldsymbol{U})^{p_3}|\boldsymbol{U}\big) \\
&\leq E\left(\max\{1,\|\boldsymbol{X}\|\}^{(p_1+p_2)p_3} \max\{1, f(\boldsymbol{X},\boldsymbol{U})\}^{p_3}\,|\boldsymbol{U}\right) \\
&\leq E\big(\max\{1,\|\boldsymbol{X}\|\}^{8} \max\{1, f(\boldsymbol{X},\boldsymbol{U})\}^{4}\,|\boldsymbol{U}\big)
\end{aligned}$$

since $p_3 \leq 4$ and $(p_1+p_2)p_3 \leq 8$ over the set of values that (p_1,p_2,p_3) takes in Section 2.4.2.10.

It follows that the condition

$$E\left[\frac{E\big(\max\{1,\|\boldsymbol{X}\|\}^8\max\{1,f(\boldsymbol{X},\boldsymbol{U})\}^4\,|\boldsymbol{U}\big)}{\big[\min\big\{1,\lambda_{\min}\big(E\{\boldsymbol{X}_{\mathrm{A}}\boldsymbol{X}_{\mathrm{A}}^T f(\boldsymbol{X},\boldsymbol{U})|\boldsymbol{U}\}\big)\big\}\big]^2}\right]<\infty \tag{2.45}$$

implies each of (MA2)-(MA6). Also, recalling that,

$$\frac{1}{\{\min(1,x)\}^2}\geq 1$$

we have

$$\frac{E\big(\max\{1,\|\boldsymbol{X}\|\}^8\max\{1,f(\boldsymbol{X},\boldsymbol{U})\}^4\,|\boldsymbol{U}\big)}{\big[\min\big\{1,\lambda_{\min}\big(E\{\boldsymbol{X}_{\mathrm{A}}\boldsymbol{X}_{\mathrm{A}}^T f(\boldsymbol{X},\boldsymbol{U})|\boldsymbol{U}\}\big)\big\}\big]^2}\geq E\{\|\boldsymbol{X}_B\|^8 f(\boldsymbol{X},\boldsymbol{U})^4|\boldsymbol{U}\}.$$

Hence, moment condition (2.45) implies that

$$E\left[E\{\|\boldsymbol{X}_B\|^8 f(\boldsymbol{X},\boldsymbol{U})^4|\boldsymbol{U}\}\right]<\infty,$$

which is equivalent to

$$E\left[\{\|\boldsymbol{X}_B\|^4 f(\boldsymbol{X},\boldsymbol{U})^2\}^2\right]<\infty. \tag{2.46}$$

However, since for any random variable Z, it is the case that

$$E(Z^2)<\infty \quad \text{implies} \quad E(|Z|)<\infty$$

condition (2.46) implies (MA1). Thus, we can claim that (2.45) implies each of (MA1)-(MA6).

2.4.3 Proof of Lemma 3

2.4.3.1 Matrix Extension of Results Concerning Integrals of Half-Cauchy Forms

Using (1.12), premultiplication on both sides of the equation by $\left(\frac{2}{\pi}\boldsymbol{A}^{1/2}\right)^{-1}$ leads to

$$\frac{\pi}{2}\boldsymbol{I}=\int_0^\infty \boldsymbol{A}^{1/2}(\boldsymbol{A}+x^2\boldsymbol{I})^{-1}\,dx. \tag{2.47}$$

Next, note that by using (2.47), we have,

$$\begin{aligned}\int_0^\infty \left(\boldsymbol{I}+x^2\boldsymbol{A}\right)^{-1}dx &= \int_0^\infty \left\{\boldsymbol{A}\left(\boldsymbol{A}^{-1}+x^2\boldsymbol{I}\right)\right\}^{-1}dx\\ &= \int_0^\infty \left(\boldsymbol{A}^{-1}+x^2\boldsymbol{I}\right)^{-1}\boldsymbol{A}^{-1}\,dx\\ &= \boldsymbol{A}^{1/2}\int_0^\infty \boldsymbol{A}^{-1/2}\left(\boldsymbol{A}^{-1}+x^2\boldsymbol{I}\right)^{-1}dx\boldsymbol{A}^{-1}\\ &= \boldsymbol{A}^{1/2}\left(\frac{\pi}{2}\boldsymbol{I}\right)\boldsymbol{A}^{-1}\\ &= \frac{\pi}{2}\boldsymbol{A}^{-1/2}.\end{aligned}$$

Therefore,

$$\frac{\pi}{2}\boldsymbol{I} = \int_0^\infty \left(\boldsymbol{I} + x^2\boldsymbol{A}\right)^{-1} \boldsymbol{A}^{1/2}\, dx. \tag{2.48}$$

Combining both (2.47) and (2.48), for $\boldsymbol{A} \in \mathbb{C}^{n\times n}$ with no eigenvalues on $\mathbb{R}^-$, we have,

$$\frac{\pi}{2}\boldsymbol{I} = \int_0^\infty \boldsymbol{A}^{1/2}(\boldsymbol{A} + x^2\boldsymbol{I})^{-1}\, dx = \int_0^\infty \left(\boldsymbol{I} + x^2\boldsymbol{A}\right)^{-1} \boldsymbol{A}^{1/2}\, dx. \tag{2.49}$$

From these results, we have,

$$\boldsymbol{I}_p = \frac{4}{\pi^2}\int_0^\infty \int_0^\infty (\boldsymbol{I}_p + t^2\boldsymbol{K})^{-1}\boldsymbol{K}(\boldsymbol{K} + u^2\boldsymbol{I}_p)^{-1}\, dt\, du \tag{2.50}$$

and

$$\boldsymbol{I}_q = \frac{4}{\pi^2}\int_0^\infty \int_0^\infty \left\{\boldsymbol{I}_q + t^2\left(\frac{1}{n}\boldsymbol{L}\right)\right\}^{-1}\left(\frac{1}{n}\boldsymbol{L}\right)\left\{\left(\frac{1}{n}\boldsymbol{L}\right) + u^2\boldsymbol{I}_q\right\}^{-1} dt\, du. \tag{2.51}$$

2.4.3.2 Derivation of Integrand Expressions

An Integral Expression for $\boldsymbol{M}_{n,\infty}^{-1/2} = \left(\boldsymbol{M}_{n,\infty}^{-1}\right)^{1/2}$

Note that

$$\boldsymbol{M}_{n,\infty} \equiv \begin{bmatrix} \boldsymbol{K} & \boldsymbol{O} \\ \boldsymbol{O} & \frac{1}{n}\boldsymbol{L} \end{bmatrix}$$

and

$$(\boldsymbol{M}_{n,\infty})^{-1} \equiv \begin{bmatrix} \boldsymbol{K}^{-1} & \boldsymbol{O} \\ \boldsymbol{O} & n\boldsymbol{L}^{-1} \end{bmatrix}.$$

Application of (2.49) to $\boldsymbol{M}_{n,\infty}^{-1}$ leads to

$$\begin{aligned}
\boldsymbol{M}_{n,\infty}^{-1/2} &= \left(\boldsymbol{M}_{n,\infty}^{-1}\right)^{1/2} \\
&= \frac{2}{\pi}\int_0^\infty \boldsymbol{M}_{n,\infty}^{-1}\left(\boldsymbol{M}_{n,\infty}^{-1} + t^2\boldsymbol{I}\right)^{-1} dt \\
&= \frac{2}{\pi}\int_0^\infty \begin{bmatrix} \boldsymbol{K}^{-1} & \boldsymbol{O} \\ \boldsymbol{O} & n\boldsymbol{L}^{-1} \end{bmatrix}\begin{bmatrix} \boldsymbol{K}^{-1} + t^2\boldsymbol{I}_p & \boldsymbol{O} \\ \boldsymbol{O} & n\boldsymbol{L}^{-1} + t^2\boldsymbol{I}_q \end{bmatrix}^{-1} dt \\
&= \frac{2}{\pi}\int_0^\infty \begin{bmatrix} \boldsymbol{K}^{-1} & \boldsymbol{O} \\ \boldsymbol{O} & n\boldsymbol{L}^{-1} \end{bmatrix}\begin{bmatrix} \left(\boldsymbol{K}^{-1} + t^2\boldsymbol{I}_p\right)^{-1} & \boldsymbol{O} \\ \boldsymbol{O} & \left(n\boldsymbol{L}^{-1} + t^2\boldsymbol{I}_q\right)^{-1} \end{bmatrix} dt \\
&= \frac{2}{\pi}\int_0^\infty \begin{bmatrix} \boldsymbol{K}^{-1} & \boldsymbol{O} \\ \boldsymbol{O} & n\boldsymbol{L}^{-1} \end{bmatrix}\begin{bmatrix} \boldsymbol{K}\left(\boldsymbol{I}_p + t^2\boldsymbol{K}\right)^{-1} & \boldsymbol{O} \\ \boldsymbol{O} & \left(\frac{1}{n}\boldsymbol{L}\right)\left\{\boldsymbol{I}_q + t^2\left(\frac{1}{n}\boldsymbol{L}\right)\right\}^{-1} \end{bmatrix} dt \\
&= \frac{2}{\pi}\int_0^\infty \begin{bmatrix} \left(\boldsymbol{I}_p + t^2\boldsymbol{K}\right)^{-1} & \boldsymbol{O} \\ \boldsymbol{O} & \left\{\boldsymbol{I}_q + t^2\left(\frac{1}{n}\boldsymbol{L}\right)\right\}^{-1} \end{bmatrix} dt.
\end{aligned}$$

An Integral Expression for $\boldsymbol{M}_n^{1/2}$

Note that

$$\boldsymbol{M}_n \equiv \begin{bmatrix} \boldsymbol{K} + Q_n \mathbf{1}_p^{\otimes 2} & R_n \mathbf{1}_p \mathbf{1}_q^T \\ R_n \mathbf{1}_q \mathbf{1}_p^T & \frac{1}{n}\boldsymbol{L} + T_n \mathbf{1}_q^{\otimes 2} \end{bmatrix}$$

where $Q_n = o_p(1), R_n = O_p(n^{-1})$ and $T_n = o_p(n^{-1})$. For all n sufficiently large so that negative eigenvalues are avoided, application of (2.49) to $\boldsymbol{M}_n$ leads to

$$\begin{aligned} \boldsymbol{M}_n^{1/2} &= \frac{2}{\pi}\int_0^\infty \boldsymbol{M}_n \left(\boldsymbol{M}_n + u^2 \boldsymbol{I}\right)^{-1} du \\ &= \frac{2}{\pi}\int_0^\infty \begin{bmatrix} \boldsymbol{K} + Q_n \mathbf{1}_p^{\otimes 2} & R_n \mathbf{1}_p \mathbf{1}_q^T \\ R_n \mathbf{1}_q \mathbf{1}_p^T & \frac{1}{n}\boldsymbol{L} + T_n \mathbf{1}_q^{\otimes 2} \end{bmatrix} \begin{bmatrix} \boldsymbol{K} + u^2 \boldsymbol{I}_p + Q_n \mathbf{1}_p^{\otimes 2} & R_n \mathbf{1}_p \mathbf{1}_q^T \\ R_n \mathbf{1}_q \mathbf{1}_p^T & \frac{1}{n}\boldsymbol{L} + u^2 \boldsymbol{I}_q + T_n \mathbf{1}_q^{\otimes 2} \end{bmatrix}^{-1} du. \end{aligned}$$

An Integral Expression for $\boldsymbol{M}_{n,\infty}^{-1/2}\boldsymbol{M}_n^{1/2}$

Firstly, note that

$$\begin{aligned} &\begin{bmatrix} \left(\boldsymbol{I}_p + t^2 \boldsymbol{K}\right)^{-1} & \boldsymbol{O} \\ \boldsymbol{O} & \left\{\boldsymbol{I}_q + t^2\left(\frac{1}{n}\boldsymbol{L}\right)\right\}^{-1} \end{bmatrix} \begin{bmatrix} \boldsymbol{K} + Q_n \mathbf{1}_p^{\otimes 2} & R_n \mathbf{1}_p \mathbf{1}_q^T \\ R_n \mathbf{1}_q \mathbf{1}_p^T & \frac{1}{n}\boldsymbol{L} + T_n \mathbf{1}_q^{\otimes 2} \end{bmatrix} \\ &\quad = \begin{bmatrix} \left(\boldsymbol{I}_p + t^2 \boldsymbol{K}\right)^{-1}\left(\boldsymbol{K} + Q_n \mathbf{1}_p^{\otimes 2}\right) & R_n \left(\boldsymbol{I}_p + t^2 \boldsymbol{K}\right)^{-1} \mathbf{1}_p \mathbf{1}_q^T \\ R_n \left\{\boldsymbol{I}_q + t^2\left(\frac{1}{n}\boldsymbol{L}\right)\right\}^{-1} \mathbf{1}_q \mathbf{1}_p^T & \left\{\boldsymbol{I}_q + t^2\left(\frac{1}{n}\boldsymbol{L}\right)\right\}^{-1} \frac{1}{n}\boldsymbol{L} + T_n \mathbf{1}_q^{\otimes 2} \end{bmatrix}. \end{aligned}$$

Then

$$\begin{aligned} &\boldsymbol{M}_{n,\infty}^{-1/2}\boldsymbol{M}_n^{1/2} \\ &= \frac{4}{\pi^2}\int_0^\infty \int_0^\infty \begin{bmatrix} \left(\boldsymbol{I}_p + t^2 \boldsymbol{K}\right)^{-1}\left(\boldsymbol{K} + Q_n \mathbf{1}_p^{\otimes 2}\right) & R_n \left(\boldsymbol{I}_p + t^2 \boldsymbol{K}\right)^{-1} \mathbf{1}_p \mathbf{1}_q^T \\ R_n \left\{\boldsymbol{I}_q + t^2\left(\frac{1}{n}\boldsymbol{L}\right)\right\}^{-1} \mathbf{1}_q \mathbf{1}_p^T & \left\{\boldsymbol{I}_q + t^2\left(\frac{1}{n}\boldsymbol{L}\right)\right\}^{-1} \frac{1}{n}\boldsymbol{L} + T_n \mathbf{1}_q^{\otimes 2} \end{bmatrix} \\ &\quad \times \begin{bmatrix} \boldsymbol{K} + u^2 \boldsymbol{I}_p + Q_n \mathbf{1}_p^{\otimes 2} & R_n \mathbf{1}_p \mathbf{1}_q^T \\ R_n \mathbf{1}_q \mathbf{1}_p^T & \frac{1}{n}\boldsymbol{L} + u^2 \boldsymbol{I}_q + T_n \mathbf{1}_q^{\otimes 2} \end{bmatrix}^{-1} dt\, du. \end{aligned}$$

An Integral Expression for $\boldsymbol{M}_{n,\infty}^{-1/2}\boldsymbol{M}_n^{1/2} - \boldsymbol{I}$

For any $u > 0$ and values of $n, \boldsymbol{K}$ and $\boldsymbol{L}$, define the $(p+q) \times (p+q)$ matrix,

$$\boldsymbol{H}_1(u; n, \boldsymbol{K}, \boldsymbol{L}) \equiv \begin{bmatrix} \boldsymbol{K} + u^2 \boldsymbol{I}_p + Q_n \mathbf{1}_p^{\otimes 2} & R_n \mathbf{1}_p \mathbf{1}_q^T \\ R_n \mathbf{1}_q \mathbf{1}_p^T & \frac{1}{n}\boldsymbol{L} + u^2 \boldsymbol{I}_q + T_n \mathbf{1}_q^{\otimes 2} \end{bmatrix}^{-1}. \tag{2.52}$$

Then, for any $(t,u)\in\mathbb{R}^2$, define the $(p+q)\times(p+q)$ matrix as follows,

$$\begin{aligned}
&\boldsymbol{H}_2(t,u;n,\boldsymbol{K},\boldsymbol{L})\\
&\equiv\begin{bmatrix}(\boldsymbol{I}_p+t^2\boldsymbol{K})^{-1}(\boldsymbol{K}+Q_n\mathbf{1}_p^{\otimes 2}) & R_n(\boldsymbol{I}_p+t^2\boldsymbol{K})^{-1}\mathbf{1}_p\mathbf{1}_q^T\\ R_n\{\boldsymbol{I}_q+t^2(\frac{1}{n}\boldsymbol{L})\}^{-1}\mathbf{1}_q\mathbf{1}_p^T & \{\boldsymbol{I}_q+t^2(\frac{1}{n}\boldsymbol{L})\}^{-1}(\frac{1}{n}\boldsymbol{L}+T_n\mathbf{1}_q^{\otimes 2})\end{bmatrix}\boldsymbol{H}_1(u;n,\boldsymbol{K},\boldsymbol{L})\\
&\quad-\begin{bmatrix}(\boldsymbol{I}_p+t^2\boldsymbol{K})^{-1}\boldsymbol{K}(\boldsymbol{K}+u^2\boldsymbol{I}_p)^{-1} & \boldsymbol{O}\\ \boldsymbol{O} & \{\boldsymbol{I}_q+t^2(\frac{1}{n}\boldsymbol{L})\}^{-1}(\frac{1}{n}\boldsymbol{L})\{(\frac{1}{n}\boldsymbol{L})+u^2\boldsymbol{I}_q\}^{-1}\end{bmatrix}.
\end{aligned}\tag{2.53}$$

Then from (2.50) and (2.51) and according to the definitions above,

$$\boldsymbol{M}_{n,\infty}^{-1/2}\boldsymbol{M}_n^{1/2}-\boldsymbol{I}_{p+q}=\frac{4}{\pi^2}\int_0^\infty\int_0^\infty\boldsymbol{H}_2(t,u;n,\boldsymbol{K},\boldsymbol{L})\,dt\,du.\tag{2.54}$$

Throughout the rest of this subsubsection, we aim to find a more explicit expression for $\boldsymbol{H}_2(t,u;n,\boldsymbol{K},\boldsymbol{L})$.

Inversion of $\boldsymbol{H}_1(u;n,\boldsymbol{K},\boldsymbol{L})$ Using Block Matrix Inversion

The upper left $p\times p$ block of $\boldsymbol{H}_1(u;n,\boldsymbol{K},\boldsymbol{L})^{-1}$ is

$$\left\{\boldsymbol{K}+u^2\boldsymbol{I}_p+Q_n\mathbf{1}_p^{\otimes 2}-R_n^2\mathbf{1}_p\mathbf{1}_q^T\left(\frac{1}{n}\boldsymbol{L}+u^2\boldsymbol{I}_q+T_n\mathbf{1}_q^{\otimes 2}\right)^{-1}\mathbf{1}_q\mathbf{1}_p^T\right\}^{-1}.$$

The upper right $p\times q$ block of $\boldsymbol{H}_1(u;n,\boldsymbol{K},\boldsymbol{L})^{-1}$ is

$$\begin{aligned}
&-R_n\left\{\boldsymbol{K}+u^2\boldsymbol{I}_p+Q_n\mathbf{1}_p^{\otimes 2}-R_n^2\mathbf{1}_p\mathbf{1}_q^T\left(\frac{1}{n}\boldsymbol{L}+u^2\boldsymbol{I}_q+T_n\mathbf{1}_q^{\otimes 2}\right)^{-1}\mathbf{1}_q\mathbf{1}_p^T\right\}^{-1}\\
&\quad\times\mathbf{1}_p\mathbf{1}_q^T\left(\frac{1}{n}\boldsymbol{L}+u^2\boldsymbol{I}_q+T_n\mathbf{1}_q^{\otimes 2}\right)^{-1}.
\end{aligned}$$

The lower left $q\times p$ block of $\boldsymbol{H}_1(u;n,\boldsymbol{K},\boldsymbol{L})^{-1}$ is

$$\begin{aligned}
&-R_n\left(\frac{1}{n}\boldsymbol{L}+u^2\boldsymbol{I}_q+T_n\mathbf{1}_q^{\otimes 2}\right)^{-1}\mathbf{1}_q\mathbf{1}_p^T\\
&\quad\times\left\{\boldsymbol{K}+u^2\boldsymbol{I}_p+Q_n\mathbf{1}_p^{\otimes 2}-R_n^2\mathbf{1}_p\mathbf{1}_q^T\left(\frac{1}{n}\boldsymbol{L}+u^2\boldsymbol{I}_q+T_n\mathbf{1}_q^{\otimes 2}\right)^{-1}\mathbf{1}_q\mathbf{1}_p^T\right\}^{-1}.
\end{aligned}$$

The lower right $q\times q$ block of $\boldsymbol{H}_1(u;n,\boldsymbol{K},\boldsymbol{L})^{-1}$ is

$$\left\{\frac{1}{n}\boldsymbol{L}+u^2\boldsymbol{I}_q+T_n\mathbf{1}_q^{\otimes 2}-R_n^2\mathbf{1}_q\mathbf{1}_p^T\left(\boldsymbol{K}+u^2\boldsymbol{I}_p+Q_n\mathbf{1}_p^{\otimes 2}\right)^{-1}\mathbf{1}_p\mathbf{1}_q^T\right\}^{-1}.$$

The Upper Left $p \times p$ Block of $\boldsymbol{H}_2(t,u;n,\boldsymbol{K},\boldsymbol{L})$

The upper left $p \times p$ block of $\boldsymbol{H}_2(t,u;n,\boldsymbol{K},\boldsymbol{L})$ is

$$
\begin{aligned}
&(\boldsymbol{I}_p + t^2\boldsymbol{K})^{-1}(\boldsymbol{K} + Q_n\mathbf{1}_p^{\otimes 2})\left\{\boldsymbol{K} + u^2\boldsymbol{I}_p + Q_n\mathbf{1}_p^{\otimes 2} - R_n^2\mathbf{1}_p\mathbf{1}_q^T\left(\frac{1}{n}\boldsymbol{L} + u^2\boldsymbol{I}_q + T_n\mathbf{1}_q^{\otimes 2}\right)^{-1}\mathbf{1}_q\mathbf{1}_p^T\right\}^{-1} \\
&\quad - R_n^2(\boldsymbol{I}_p + t^2\boldsymbol{K})^{-1}\mathbf{1}_p\mathbf{1}_q^T\left(\frac{1}{n}\boldsymbol{L} + u^2\boldsymbol{I}_q + T_n\mathbf{1}_q^{\otimes 2}\right)^{-1}\mathbf{1}_q\mathbf{1}_p^T \\
&\quad \times\left\{\boldsymbol{K} + u^2\boldsymbol{I}_p + Q_n\mathbf{1}_p^{\otimes 2} - R_n^2\mathbf{1}_p\mathbf{1}_q^T\left(\frac{1}{n}\boldsymbol{L} + u^2\boldsymbol{I}_q + T_n\mathbf{1}_q^{\otimes 2}\right)^{-1}\mathbf{1}_q\mathbf{1}_p^T\right\}^{-1} \\
&\quad - (\boldsymbol{I}_p + t^2\boldsymbol{K})^{-1}\boldsymbol{K}\left(\boldsymbol{K} + u^2\boldsymbol{I}_p\right)^{-1} \\
&= \left\{(\boldsymbol{I}_p + t^2\boldsymbol{K})^{-1}(\boldsymbol{K} + Q_n\mathbf{1}_p^{\otimes 2}) - R_n^2(\boldsymbol{I}_p + t^2\boldsymbol{K})^{-1}\mathbf{1}_p\mathbf{1}_q^T\left(\frac{1}{n}\boldsymbol{L} + u^2\boldsymbol{I}_q + T_n\mathbf{1}_q^{\otimes 2}\right)^{-1}\mathbf{1}_q\mathbf{1}_p^T\right\} \\
&\qquad \times\left\{\boldsymbol{K} + u^2\boldsymbol{I}_p + Q_n\mathbf{1}_p^{\otimes 2} - R_n^2\mathbf{1}_p\mathbf{1}_q^T\left(\frac{1}{n}\boldsymbol{L} + u^2\boldsymbol{I}_q + T_n\mathbf{1}_q^{\otimes 2}\right)^{-1}\mathbf{1}_q\mathbf{1}_p^T\right\}^{-1} \\
&\qquad - (\boldsymbol{I}_p + t^2\boldsymbol{K})^{-1}\boldsymbol{K}\left(\boldsymbol{K} + u^2\boldsymbol{I}_p\right)^{-1} \\
&= (\boldsymbol{I}_p + t^2\boldsymbol{K})^{-1}\Bigg[\left\{\boldsymbol{K} + Q_n\mathbf{1}_p^{\otimes 2} - R_n^2\mathbf{1}_p\mathbf{1}_q^T\left(\frac{1}{n}\boldsymbol{L} + u^2\boldsymbol{I}_q + T_n\mathbf{1}_q^{\otimes 2}\right)^{-1}\mathbf{1}_q\mathbf{1}_p^T\right\} \\
&\qquad \times\left\{\boldsymbol{K} + u^2\boldsymbol{I}_p + Q_n\mathbf{1}_p^{\otimes 2} - R_n^2\mathbf{1}_p\mathbf{1}_q^T\left(\frac{1}{n}\boldsymbol{L} + u^2\boldsymbol{I}_q + T_n\mathbf{1}_q^{\otimes 2}\right)^{-1}\mathbf{1}_q\mathbf{1}_p^T\right\}^{-1} \\
&\qquad\quad - \boldsymbol{K}\left(\boldsymbol{K} + u^2\boldsymbol{I}_p\right)^{-1}\Bigg] \\
&= (\boldsymbol{I}_p + t^2\boldsymbol{K})^{-1}\boldsymbol{K}\Bigg[\left\{\boldsymbol{K} + u^2\boldsymbol{I}_p + Q_n\mathbf{1}_p^{\otimes 2} - R_n^2\mathbf{1}_p\mathbf{1}_q^T\left(\frac{1}{n}\boldsymbol{L} + u^2\boldsymbol{I}_q + T_n\mathbf{1}_q^{\otimes 2}\right)^{-1}\mathbf{1}_q\mathbf{1}_p^T\right\}^{-1} \\
&\qquad - \left(\boldsymbol{K} + u^2\boldsymbol{I}_p\right)^{-1}\Bigg] + (\boldsymbol{I}_p + t^2\boldsymbol{K})^{-1}\left\{Q_n\mathbf{1}_p^{\otimes 2} - R_n^2\mathbf{1}_p\mathbf{1}_q^T\left(\frac{1}{n}\boldsymbol{L} + u^2\boldsymbol{I}_q + T_n\mathbf{1}_q^{\otimes 2}\right)^{-1}\mathbf{1}_q\mathbf{1}_p^T\right\} \\
&\qquad \times\left\{\boldsymbol{K} + u^2\boldsymbol{I}_p + Q_n\mathbf{1}_p^{\otimes 2} - R_n^2\mathbf{1}_p\mathbf{1}_q^T\left(\frac{1}{n}\boldsymbol{L} + u^2\boldsymbol{I}_q + T_n\mathbf{1}_q^{\otimes 2}\right)^{-1}\mathbf{1}_q\mathbf{1}_p^T\right\}^{-1}.
\end{aligned}
$$

Using (1.1), the first term in the expression can be re-written as

$$
\begin{aligned}
&(\boldsymbol{I}_p + t^2\boldsymbol{K})^{-1}\boldsymbol{K}\left\{\boldsymbol{K} + u^2\boldsymbol{I}_p + Q_n\mathbf{1}_p^{\otimes 2} - R_n^2\mathbf{1}_p\mathbf{1}_q^T\left(\frac{1}{n}\boldsymbol{L} + u^2\boldsymbol{I}_q + T_n\mathbf{1}_q^{\otimes 2}\right)^{-1}\mathbf{1}_q\mathbf{1}_p^T\right\}^{-1} \\
&\quad \times\left\{R_n^2\mathbf{1}_p\mathbf{1}_q^T\left(\frac{1}{n}\boldsymbol{L} + u^2\boldsymbol{I}_q + T_n\mathbf{1}_q^{\otimes 2}\right)^{-1}\mathbf{1}_q\mathbf{1}_p^T - Q_n\mathbf{1}_p^{\otimes 2}\right\}\left(\boldsymbol{K} + u^2\boldsymbol{I}_p\right)^{-1}.
\end{aligned}
$$

Putting together the results given so far in this subsubsection, we have that the upper left $p \times p$ block of $\boldsymbol{H}_2(t, u; n, \boldsymbol{K}, \boldsymbol{L})$ is:

$$
\begin{aligned}
&(\boldsymbol{I}_p + t^2\boldsymbol{K})^{-1}\boldsymbol{K}\left\{\boldsymbol{K} + u^2\boldsymbol{I}_p + Q_n\mathbf{1}_p^{\otimes 2} - R_n^2\mathbf{1}_p\mathbf{1}_q^T\left(\frac{1}{n}\boldsymbol{L} + u^2\boldsymbol{I}_q + T_n\mathbf{1}_q^{\otimes 2}\right)^{-1}\mathbf{1}_q\mathbf{1}_p^T\right\}^{-1} \\
&\quad\times\left\{R_n^2\mathbf{1}_p\mathbf{1}_q^T\left(\frac{1}{n}\boldsymbol{L} + u^2\boldsymbol{I}_q + T_n\mathbf{1}_q^{\otimes 2}\right)^{-1}\mathbf{1}_q\mathbf{1}_p^T - Q_n\mathbf{1}_p^{\otimes 2}\right\}(\boldsymbol{K} + u^2\boldsymbol{I}_p)^{-1} \\
&\quad-(\boldsymbol{I}_p + t^2\boldsymbol{K})^{-1}\left\{R_n^2\mathbf{1}_p\mathbf{1}_q^T\left(\frac{1}{n}\boldsymbol{L} + u^2\boldsymbol{I}_q + T_n\mathbf{1}_q^{\otimes 2}\right)^{-1}\mathbf{1}_q\mathbf{1}_p^T - Q_n\mathbf{1}_p^{\otimes 2}\right\} \\
&\quad\times\left\{\boldsymbol{K} + u^2\boldsymbol{I}_p + Q_n\mathbf{1}_p^{\otimes 2} - R_n^2\mathbf{1}_p\mathbf{1}_q^T\left(\frac{1}{n}\boldsymbol{L} + u^2\boldsymbol{I}_q + T_n\mathbf{1}_q^{\otimes 2}\right)^{-1}\mathbf{1}_q\mathbf{1}_p^T\right\}^{-1}.
\end{aligned}
$$

The Lower Left $q \times p$ Block of $\boldsymbol{H}_2(t, u; n, \boldsymbol{K}, \boldsymbol{L})$

Noting that the lower left $q \times p$ block of $\boldsymbol{I}$ is a zero matrix, the lower left $q \times p$ block of $\boldsymbol{H}_2(t, u; n, \boldsymbol{K}, \boldsymbol{L})$ is

$$
\begin{aligned}
&R_n\left\{\boldsymbol{I}_q + t^2\left(\frac{1}{n}\boldsymbol{L}\right)\right\}^{-1}\mathbf{1}_q\mathbf{1}_p^T\left\{\boldsymbol{K} + u^2\boldsymbol{I}_p + Q_n\mathbf{1}_p^{\otimes 2} - R_n^2\mathbf{1}_p\mathbf{1}_q^T\left(\frac{1}{n}\boldsymbol{L} + u^2\boldsymbol{I}_q + T_n\mathbf{1}_q^{\otimes 2}\right)^{-1}\mathbf{1}_q\mathbf{1}_p^T\right\}^{-1} \\
&\quad- R_n\left\{\boldsymbol{I}_q + t^2\left(\frac{1}{n}\boldsymbol{L}\right)\right\}^{-1}\left(\frac{1}{n}\boldsymbol{L} + T_n\mathbf{1}_q^{\otimes 2}\right)\left(\frac{1}{n}\boldsymbol{L} + u^2\boldsymbol{I}_q + T_n\mathbf{1}_q^{\otimes 2}\right)^{-1}\mathbf{1}_q\mathbf{1}_p^T \\
&\quad\times\left\{\boldsymbol{K} + u^2\boldsymbol{I}_p + Q_n\mathbf{1}_p^{\otimes 2} - R_n^2\mathbf{1}_p\mathbf{1}_q^T\left(\frac{1}{n}\boldsymbol{L} + u^2\boldsymbol{I}_q + T_n\mathbf{1}_q^{\otimes 2}\right)^{-1}\mathbf{1}_q\mathbf{1}_p^T\right\}^{-1} \\
&= R_n\left\{\boldsymbol{I}_q + t^2\left(\frac{1}{n}\boldsymbol{L}\right)\right\}^{-1}\left\{\boldsymbol{I}_q - \left(\frac{1}{n}\boldsymbol{L} + T_n\mathbf{1}_q^{\otimes 2}\right)\left(\frac{1}{n}\boldsymbol{L} + u^2\boldsymbol{I}_q + T_n\mathbf{1}_q^{\otimes 2}\right)^{-1}\right\}\mathbf{1}_q\mathbf{1}_p^T \\
&\quad\times\left\{\boldsymbol{K} + u^2\boldsymbol{I}_p + Q_n\mathbf{1}_p^{\otimes 2} - R_n^2\mathbf{1}_p\mathbf{1}_q^T\left(\frac{1}{n}\boldsymbol{L} + u^2\boldsymbol{I}_q + T_n\mathbf{1}_q^{\otimes 2}\right)^{-1}\mathbf{1}_q\mathbf{1}_p^T\right\}^{-1} \\
&= R_n u^2\left\{\boldsymbol{I}_q + t^2\left(\frac{1}{n}\boldsymbol{L}\right)\right\}^{-1}\left(\frac{1}{n}\boldsymbol{L} + u^2\boldsymbol{I}_q + T_n\mathbf{1}_q^{\otimes 2}\right)^{-1}\mathbf{1}_q\mathbf{1}_p^T \\
&\quad\times\left\{\boldsymbol{K} + u^2\boldsymbol{I}_p + Q_n\mathbf{1}_p^{\otimes 2} - R_n^2\mathbf{1}_p\mathbf{1}_q^T\left(\frac{1}{n}\boldsymbol{L} + u^2\boldsymbol{I}_q + T_n\mathbf{1}_q^{\otimes 2}\right)^{-1}\mathbf{1}_q\mathbf{1}_p^T\right\}^{-1}.
\end{aligned}
$$

The Upper Right $p \times q$ Block of $\boldsymbol{H}_2(t, u; n, \boldsymbol{K}, \boldsymbol{L})$

Noting that the upper right $p \times q$ block of $\boldsymbol{I}$ is a zero matrix, the upper right $p \times q$

block of $\boldsymbol{H}_2(t,u;n,\boldsymbol{K},\boldsymbol{L})$ is

$$
\begin{aligned}
&- R_n(\boldsymbol{I}_p + t^2\boldsymbol{K})^{-1}(\boldsymbol{K} + Q_n\mathbf{1}_p^{\otimes 2}) \\
&\quad \times \Big\{\boldsymbol{K} + u^2\boldsymbol{I}_p + Q_n\mathbf{1}_p^{\otimes 2} - R_n^2\mathbf{1}_p\mathbf{1}_q^T\Big(\frac{1}{n}\boldsymbol{L} + u^2\boldsymbol{I}_q + T_n\mathbf{1}_q^{\otimes 2}\Big)^{-1}\mathbf{1}_q\mathbf{1}_p^T\Big\}^{-1} \\
&\quad \times \mathbf{1}_p\mathbf{1}_q^T\Big(\frac{1}{n}\boldsymbol{L} + u^2\boldsymbol{I}_q + T_n\mathbf{1}_q^{\otimes 2}\Big)^{-1} \\
&\quad + R_n(\boldsymbol{I}_p + t^2\boldsymbol{K})^{-1}\mathbf{1}_p\mathbf{1}_q^T\Big\{\frac{1}{n}\boldsymbol{L} + u^2\boldsymbol{I}_q + T_n\mathbf{1}_q^{\otimes 2} - R_n^2\mathbf{1}_q\mathbf{1}_p^T(\boldsymbol{K} + u^2\boldsymbol{I}_p + Q_n\mathbf{1}_p^{\otimes 2})^{-1}\mathbf{1}_p\mathbf{1}_q^T\Big\}^{-1} \\
&= R_n(\boldsymbol{I}_p + t^2\boldsymbol{K})^{-1}\Big[\mathbf{1}_p\mathbf{1}_q^T\Big\{\frac{1}{n}\boldsymbol{L} + u^2\boldsymbol{I}_q + T_n\mathbf{1}_q^{\otimes 2} - R_n^2\mathbf{1}_q\mathbf{1}_p^T(\boldsymbol{K} + u^2\boldsymbol{I}_p + Q_n\mathbf{1}_p^{\otimes 2})^{-1}\mathbf{1}_p\mathbf{1}_q^T\Big\}^{-1} \\
&\qquad - (\boldsymbol{K} + Q_n\mathbf{1}_p^{\otimes 2})\Big\{\boldsymbol{K} + u^2\boldsymbol{I}_p + Q_n\mathbf{1}_p^{\otimes 2} - R_n^2\mathbf{1}_p\mathbf{1}_q^T\Big(\frac{1}{n}\boldsymbol{L} + u^2\boldsymbol{I}_q + T_n\mathbf{1}_q^{\otimes 2}\Big)^{-1}\mathbf{1}_q\mathbf{1}_p^T\Big\}^{-1}\mathbf{1}_p\mathbf{1}_q^T \\
&\qquad \times \left(\frac{1}{n}\boldsymbol{L} + u^2\boldsymbol{I}_q + T_n\mathbf{1}_q^{\otimes 2}\right)^{-1}\Big].
\end{aligned}
$$

The Lower Right $q \times q$ Block of $\boldsymbol{H}_2(t,u;n,\boldsymbol{K},\boldsymbol{L})$

The lower right $q \times q$ block of $\boldsymbol{H}_2(t,u;n,\boldsymbol{K},\boldsymbol{L})$ is

$$
\begin{aligned}
&\left\{\boldsymbol{I}_q + t^2\left(\frac{1}{n}\boldsymbol{L}\right)\right\}^{-1}\left(\frac{1}{n}\boldsymbol{L} + T_n\mathbf{1}_q^{\otimes 2}\right) \\
&\quad \times \left\{\frac{1}{n}\boldsymbol{L} + u^2\boldsymbol{I}_q + T_n\mathbf{1}_q^{\otimes 2} - R_n^2\mathbf{1}_q\mathbf{1}_p^T\left(\boldsymbol{K} + u^2\boldsymbol{I}_p + Q_n\mathbf{1}_p^{\otimes 2}\right)^{-1}\mathbf{1}_p\mathbf{1}_q^T\right\}^{-1} - R_n^2\left\{\boldsymbol{I}_q + t^2\left(\frac{1}{n}\boldsymbol{L}\right)\right\}^{-1} \\
&\quad \times \mathbf{1}_q\mathbf{1}_p^T\left\{\boldsymbol{K} + u^2\boldsymbol{I}_p + Q_n\mathbf{1}_p^{\otimes 2} - R_n^2\mathbf{1}_p\mathbf{1}_q^T\left(\frac{1}{n}\boldsymbol{L} + u^2\boldsymbol{I}_q + T_n\mathbf{1}_q^{\otimes 2}\right)^{-1}\mathbf{1}_q\mathbf{1}_p^T\right\}^{-1} \\
&\quad \times \mathbf{1}_p\mathbf{1}_q^T\left(\frac{1}{n}\boldsymbol{L} + u^2\boldsymbol{I}_q + T_n\mathbf{1}_q^{\otimes 2}\right)^{-1} - \left\{\boldsymbol{I}_q + t^2\left(\frac{1}{n}\boldsymbol{L}\right)\right\}^{-1}\left(\frac{1}{n}\boldsymbol{L}\right)\left\{\left(\frac{1}{n}\boldsymbol{L}\right) + u^2\boldsymbol{I}_q\right\}^{-1} \\
&= \left\{\boldsymbol{I}_q + t^2\left(\frac{1}{n}\boldsymbol{L}\right)\right\}^{-1} \\
&\qquad \times \Bigg[\left(\frac{1}{n}\boldsymbol{L} + T_n\mathbf{1}_q^{\otimes 2}\right)\left\{\frac{1}{n}\boldsymbol{L} + u^2\boldsymbol{I}_q + T_n\mathbf{1}_q^{\otimes 2} - R_n^2\mathbf{1}_q\mathbf{1}_p^T\left(\boldsymbol{K} + u^2\boldsymbol{I}_p + Q_n\mathbf{1}_p^{\otimes 2}\right)^{-1}\mathbf{1}_p\mathbf{1}_q^T\right\}^{-1} \\
&\qquad - \left(\frac{1}{n}\boldsymbol{L}\right)\left\{\left(\frac{1}{n}\boldsymbol{L}\right) + u^2\boldsymbol{I}_q\right\}^{-1} \\
&\qquad - R_n^2\mathbf{1}_q\mathbf{1}_p^T\left\{\boldsymbol{K} + u^2\boldsymbol{I}_p + Q_n\mathbf{1}_p^{\otimes 2} - R_n^2\mathbf{1}_p\mathbf{1}_q^T\left(\frac{1}{n}\boldsymbol{L} + u^2\boldsymbol{I}_q + T_n\mathbf{1}_q^{\otimes 2}\right)^{-1}\mathbf{1}_q\mathbf{1}_p^T\right\}^{-1} \\
&\qquad \times \mathbf{1}_p\mathbf{1}_q^T\left(\frac{1}{n}\boldsymbol{L} + u^2\boldsymbol{I}_q + T_n\mathbf{1}_q^{\otimes 2}\right)^{-1}\Bigg].
\end{aligned}
$$

Therefore, the lower right $q \times q$ block of $\boldsymbol{H}_2(t, u; n, \boldsymbol{K}, \boldsymbol{L})$ is

$$
\begin{aligned}
&= \left\{ \boldsymbol{I}_q + t^2 \left(\frac{1}{n} \boldsymbol{L} \right) \right\}^{-1} \\
&\quad \times \left(\left(\frac{1}{n} \boldsymbol{L} \right) \left[\left\{ \frac{1}{n} \boldsymbol{L} + u^2 \boldsymbol{I}_q + T_n \mathbf{1}_q^{\otimes 2} - R_n^2 \mathbf{1}_q \mathbf{1}_p^T \left(\boldsymbol{K} + u^2 \boldsymbol{I}_p + Q_n \mathbf{1}_p^{\otimes 2} \right)^{-1} \mathbf{1}_p \mathbf{1}_q^T \right\}^{-1} \right. \right. \\
&\qquad \left. \left. - \left\{ \left(\frac{1}{n} \boldsymbol{L} \right) + u^2 \boldsymbol{I}_q \right\}^{-1} \right] + T_n \mathbf{1}_q^{\otimes 2} \left\{ \frac{1}{n} \boldsymbol{L} + u^2 \boldsymbol{I}_q \right. \right. \\
&\quad \left. \left. + T_n \mathbf{1}_q^{\otimes 2} - R_n^2 \mathbf{1}_q \mathbf{1}_p^T \left(\boldsymbol{K} + u^2 \boldsymbol{I}_p + Q_n \mathbf{1}_p^{\otimes 2} \right)^{-1} \mathbf{1}_p \mathbf{1}_q^T \right\}^{-1} \right. \\
&\quad - R_n^2 \mathbf{1}_q \mathbf{1}_p^T \left\{ \boldsymbol{K} + u^2 \boldsymbol{I}_p + Q_n \mathbf{1}_p^{\otimes 2} - R_n^2 \mathbf{1}_p \mathbf{1}_q^T \left(\frac{1}{n} \boldsymbol{L} + u^2 \boldsymbol{I}_q + T_n \mathbf{1}_q^{\otimes 2} \right)^{-1} \mathbf{1}_q \mathbf{1}_p^T \right\}^{-1} \\
&\quad \left. \times \mathbf{1}_p \mathbf{1}_q^T \left(\frac{1}{n} \boldsymbol{L} + u^2 \boldsymbol{I}_q + T_n \mathbf{1}_q^{\otimes 2} \right)^{-1} \right).
\end{aligned}
$$

Application of (1.1) to the first term in the preceding expression leads to

$$
\begin{aligned}
&\left\{ \boldsymbol{I}_q + t^2 \left(\frac{1}{n} \boldsymbol{L} \right) \right\}^{-1} \left(\frac{1}{n} \boldsymbol{L} \right) \\
&\quad \times \left\{ \frac{1}{n} \boldsymbol{L} + u^2 \boldsymbol{I}_q + T_n \mathbf{1}_q^{\otimes 2} - R_n^2 \mathbf{1}_q \mathbf{1}_p^T \left(\boldsymbol{K} + u^2 \boldsymbol{I}_p + Q_n \mathbf{1}_p^{\otimes 2} \right)^{-1} \mathbf{1}_p \mathbf{1}_q^T \right\}^{-1} \\
&\quad \times \left\{ R_n^2 \mathbf{1}_q \mathbf{1}_p^T \left(\boldsymbol{K} + u^2 \boldsymbol{I}_p + Q_n \mathbf{1}_p^{\otimes 2} \right)^{-1} \mathbf{1}_p \mathbf{1}_q^T - T_n \mathbf{1}_q^{\otimes 2} \right\} \left\{ \left(\frac{1}{n} \boldsymbol{L} \right) + u^2 \boldsymbol{I}_q \right\}^{-1}.
\end{aligned}
$$

The next term of the lower right $q \times q$ block of $\boldsymbol{H}_2(t, u; n, \boldsymbol{K}, \boldsymbol{L})$ is

$$
\begin{aligned}
&T_n \left\{ \boldsymbol{I}_q + t^2 \left(\frac{1}{n} \boldsymbol{L} \right) \right\}^{-1} \mathbf{1}_q^{\otimes 2} \\
&\quad \times \left\{ \frac{1}{n} \boldsymbol{L} + u^2 \boldsymbol{I}_q + T_n \mathbf{1}_q^{\otimes 2} - R_n^2 \mathbf{1}_q \mathbf{1}_p^T \left(\boldsymbol{K} + u^2 \boldsymbol{I}_p + Q_n \mathbf{1}_p^{\otimes 2} \right)^{-1} \mathbf{1}_p \mathbf{1}_q^T \right\}^{-1}.
\end{aligned}
$$

Hence, the final term of the lower right $q \times q$ block of $\boldsymbol{H}_2(t, u; n, \boldsymbol{K}, \boldsymbol{L})$ is

$$
\begin{aligned}
&- R_n^2 \left\{ \boldsymbol{I}_q + t^2 \left(\frac{1}{n} \boldsymbol{L} \right) \right\}^{-1} \mathbf{1}_q \mathbf{1}_p^T \\
&\quad \times \left\{ \boldsymbol{K} + u^2 \boldsymbol{I}_p + Q_n \mathbf{1}_p^{\otimes 2} - R_n^2 \mathbf{1}_p \mathbf{1}_q^T \left(\frac{1}{n} \boldsymbol{L} + u^2 \boldsymbol{I}_q + T_n \mathbf{1}_q^{\otimes 2} \right)^{-1} \mathbf{1}_q \mathbf{1}_p^T \right\}^{-1} \\
&\quad \times \mathbf{1}_p \mathbf{1}_q^T \left(\frac{1}{n} \boldsymbol{L} + u^2 \boldsymbol{I}_q + T_n \mathbf{1}_q^{\otimes 2} \right)^{-1}.
\end{aligned}
$$

Putting all these expressions together, we have,

$$\begin{aligned}
&\left\{\boldsymbol{I}_q + t^2\left(\frac{1}{n}\boldsymbol{L}\right)\right\}^{-1}\left[\left(\frac{1}{n}\boldsymbol{L}\right)\left\{\frac{1}{n}\boldsymbol{L} + u^2\boldsymbol{I}_q + T_n\mathbf{1}_q^{\otimes 2} - R_n^2\mathbf{1}_q\mathbf{1}_p^T\left(\boldsymbol{K} + u^2\boldsymbol{I}_p + Q_n\mathbf{1}_p^{\otimes 2}\right)^{-1}\mathbf{1}_p\mathbf{1}_q^T\right\}^{-1}\right.\\
&\quad\times\left\{R_n^2\mathbf{1}_q\mathbf{1}_p^T\left(\boldsymbol{K} + u^2\boldsymbol{I}_p + Q_n\mathbf{1}_p^{\otimes 2}\right)^{-1}\mathbf{1}_p\mathbf{1}_q^T - T_n\mathbf{1}_q^{\otimes 2}\right\}\left\{\left(\frac{1}{n}\boldsymbol{L}\right) + u^2\boldsymbol{I}_q\right\}^{-1}\\
&\quad+ T_n\mathbf{1}_q^{\otimes 2}\left\{\frac{1}{n}\boldsymbol{L} + u^2\boldsymbol{I}_q + T_n\mathbf{1}_q^{\otimes 2} - R_n^2\mathbf{1}_q\mathbf{1}_p^T\left(\boldsymbol{K} + u^2\boldsymbol{I}_p + Q_n\mathbf{1}_p^{\otimes 2}\right)^{-1}\mathbf{1}_p\mathbf{1}_q^T\right\}^{-1}\\
&\quad- R_n^2\mathbf{1}_q\mathbf{1}_p^T\left\{\boldsymbol{K} + u^2\boldsymbol{I}_p + Q_n\mathbf{1}_p^{\otimes 2} - R_n^2\mathbf{1}_p\mathbf{1}_q^T\left(\frac{1}{n}\boldsymbol{L} + u^2\boldsymbol{I}_q + T_n\mathbf{1}_q^{\otimes 2}\right)^{-1}\mathbf{1}_q\mathbf{1}_p^T\right\}^{-1}\\
&\quad\times\left.\mathbf{1}_p\mathbf{1}_q^T\left(\frac{1}{n}\boldsymbol{L} + u^2\boldsymbol{I}_q + T_n\mathbf{1}_q^{\otimes 2}\right)^{-1}\right].
\end{aligned}$$

2.4.3.3 Succinct Expressions for the Components in (2.53)

Define

$$\begin{aligned}
\boldsymbol{\Gamma}_{1n}(u; T_n, \boldsymbol{K}, \boldsymbol{L}) &\equiv \left(\frac{1}{n}\boldsymbol{L} + u^2\boldsymbol{I}_q + T_n\mathbf{1}_q^{\otimes 2}\right)^{-1}\\
\boldsymbol{\Gamma}_{2n}(u; Q_n, R_n, \boldsymbol{K}, \boldsymbol{L}) &\equiv R_n^2\mathbf{1}_p\mathbf{1}_q^T\left(\boldsymbol{\Gamma}_{1n}(u; T_n, \boldsymbol{K}, \boldsymbol{L})\right)\mathbf{1}_q\mathbf{1}_p^T - Q_n\mathbf{1}_p^{\otimes 2}\\
\boldsymbol{\Gamma}_{3n}(u; Q_n, R_n, T_n, \boldsymbol{K}, \boldsymbol{L}) &\equiv R_n^2\mathbf{1}_q\mathbf{1}_p^T\left(\boldsymbol{K} + u^2\boldsymbol{I}_p + Q_n\mathbf{1}_p^{\otimes 2}\right)^{-1}\mathbf{1}_p\mathbf{1}_q^T - T_n\mathbf{1}_q^{\otimes 2}\\
\boldsymbol{\Gamma}_{4n}(u; Q_n, R_n, \boldsymbol{K}, \boldsymbol{L}) &\equiv \left\{\boldsymbol{K} + u^2\boldsymbol{I}_p - \left(\boldsymbol{\Gamma}_{2n}(u; Q_n, R_n, \boldsymbol{K}, \boldsymbol{L})\right)\right\}^{-1} \quad \text{and}\\
\boldsymbol{\Gamma}_{5n}(u; Q_n, R_n, T_n, \boldsymbol{K}, \boldsymbol{L}) &\equiv \left\{\frac{1}{n}\boldsymbol{L} + u^2\boldsymbol{I}_q - \left(\boldsymbol{\Gamma}_{3n}(u; Q_n, R_n, T_n, \boldsymbol{K}, \boldsymbol{L})\right)\right\}^{-1}.
\end{aligned}$$

From now on, we write our expressions for $\boldsymbol{H}_2(t, u; n, \boldsymbol{K}, \boldsymbol{L})$ in terms of $\boldsymbol{\Gamma}_{1n}(u), \ldots, \boldsymbol{\Gamma}_{5n}(u)$ and suppress all other arguments. Then,

$$\begin{aligned}
\boldsymbol{H}_2(t, u; n, \boldsymbol{K}, \boldsymbol{L})_{11} &= \left(\boldsymbol{I}_p + t^2\boldsymbol{K}\right)^{-1}\left\{\boldsymbol{K}\boldsymbol{\Gamma}_{4n}(u)\boldsymbol{\Gamma}_{2n}(u)\left(\boldsymbol{K} + u^2\boldsymbol{I}_p\right)^{-1} - \boldsymbol{\Gamma}_{2n}(u)\boldsymbol{\Gamma}_{4n}(u)\right\}\\
\boldsymbol{H}_2(t, u; n, \boldsymbol{K}, \boldsymbol{L})_{21} &= R_n u^2\left\{\boldsymbol{I}_q + t^2\left(\frac{1}{n}\boldsymbol{L}\right)\right\}^{-1}\boldsymbol{\Gamma}_{1n}(u)\mathbf{1}_q\mathbf{1}_p^T\boldsymbol{\Gamma}_{4n}(u)\\
\boldsymbol{H}_2(t, u; n, \boldsymbol{K}, \boldsymbol{L})_{12} &= R_n(\boldsymbol{I}_p + t^2\boldsymbol{K})^{-1}\left\{\mathbf{1}_p\mathbf{1}_q^T\boldsymbol{\Gamma}_{5n}(u) - (\boldsymbol{K} + Q_n\mathbf{1}_p^{\otimes 2})\boldsymbol{\Gamma}_{4n}(u)\mathbf{1}_p\mathbf{1}_q^T\boldsymbol{\Gamma}_{1n}(u)\right\} \quad \text{and}\\
\boldsymbol{H}_2(t, u; n, \boldsymbol{K}, \boldsymbol{L})_{22} &= \left\{\boldsymbol{I}_q + t^2\left(\frac{1}{n}\boldsymbol{L}\right)\right\}^{-1}\left[\left(\frac{1}{n}\boldsymbol{L}\right)\boldsymbol{\Gamma}_{5n}(u)\boldsymbol{\Gamma}_{3n}(u)\left\{\frac{1}{n}\boldsymbol{L} + u^2\boldsymbol{I}_q\right\}^{-1}\right.\\
&\qquad\left. + T_n\mathbf{1}_q^{\otimes 2}\boldsymbol{\Gamma}_{5n}(u) - R_n^2\mathbf{1}_q\mathbf{1}_p^T\boldsymbol{\Gamma}_{4n}(u)\mathbf{1}_p\mathbf{1}_q^T\boldsymbol{\Gamma}_{1n}(u)\right].
\end{aligned}$$

2.4.3.4 Simplification of Integrals

It follows from the results in Section 2.4.3.1 that

$$\int_0^\infty \left(\boldsymbol{I}_p + t^2\boldsymbol{K}\right)^{-1} dt = \frac{\pi}{2}\boldsymbol{K}^{-1/2}.$$

Also, using (2.49),

$$\int_0^\infty \left\{\boldsymbol{I}_q + t^2\left(\frac{1}{n}\boldsymbol{L}\right)\right\}^{-1} dt = \frac{\pi}{2}\left(\frac{1}{n}\boldsymbol{L}\right)^{-1/2} = \frac{\pi n^{1/2}}{2}\boldsymbol{L}^{-1/2}.$$

2.4.3.5 Explicit Expressions for (2.54)

Define

$$\frac{\pi}{2}\left(\boldsymbol{M}_{n,\infty}^{-1/2}\boldsymbol{M}_n^{1/2} - \boldsymbol{I}_{p+q}\right) = \begin{bmatrix} \boldsymbol{K}^{-1/2}\int_0^\infty \boldsymbol{F}_{11n}(u;\boldsymbol{K},\boldsymbol{L})du & \boldsymbol{K}^{-1/2}\int_0^\infty \boldsymbol{F}_{12n}(u;\boldsymbol{K},\boldsymbol{L})du \\ \boldsymbol{L}^{-1/2}\int_0^\infty \boldsymbol{F}_{21n}(u;\boldsymbol{K},\boldsymbol{L})du & \boldsymbol{L}^{-1/2}\int_0^\infty \boldsymbol{F}_{22n}(u;\boldsymbol{K},\boldsymbol{L})du \end{bmatrix} \tag{2.55}$$

where

$$\begin{aligned}
\boldsymbol{F}_{11n}(u;\boldsymbol{K},\boldsymbol{L}) &\equiv \boldsymbol{K}\boldsymbol{\Gamma}_{4n}(u)\boldsymbol{\Gamma}_{2n}(u)\left(\boldsymbol{K}+u^2\boldsymbol{I}_p\right)^{-1} - \boldsymbol{\Gamma}_{2n}(u)\boldsymbol{\Gamma}_{4n}(u) \\
\boldsymbol{F}_{21n}(u;\boldsymbol{K},\boldsymbol{L}) &\equiv n^{1/2}R_n u^2\boldsymbol{\Gamma}_{1n}(u)\mathbf{1}_q\mathbf{1}_p^T\boldsymbol{\Gamma}_{4n}(u) \\
\boldsymbol{F}_{12n}(u;\boldsymbol{K},\boldsymbol{L}) &\equiv R_n\{\mathbf{1}_p\mathbf{1}_q^T\boldsymbol{\Gamma}_{5n}(u) - (\boldsymbol{K}+Q_n\mathbf{1}_p^{\otimes 2})\boldsymbol{\Gamma}_{4n}(u)\mathbf{1}_p\mathbf{1}_q^T\boldsymbol{\Gamma}_{1n}(u)\} \quad \text{and} \\
\boldsymbol{F}_{22n}(u;\boldsymbol{K},\boldsymbol{L}) &\equiv n^{1/2}\Bigg[\left(\frac{1}{n}\boldsymbol{L}\right)\boldsymbol{\Gamma}_{5n}(u)\boldsymbol{\Gamma}_{3n}(u)\left\{\frac{1}{n}\boldsymbol{L}+u^2\boldsymbol{I}_q\right\}^{-1} \\
&\qquad + T_n\mathbf{1}_q^{\otimes 2}\boldsymbol{\Gamma}_{5n}(u) - R_n^2\mathbf{1}_q\mathbf{1}_p^T\boldsymbol{\Gamma}_{4n}(u)\mathbf{1}_p\mathbf{1}_q^T\boldsymbol{\Gamma}_{1n}(u)\Bigg].
\end{aligned}$$

2.4.3.6 Convergence in Probability Limits of the Functions in (2.55)

In Appendix 2.4.4, we establish that

$$\operatorname*{plim}_{n\to\infty}\int_0^\infty \boldsymbol{F}_{kk'n}(u;\boldsymbol{K},\boldsymbol{L})\,du = \boldsymbol{O}, \quad k,k' = 1,2.$$

for $u > 0$. Hence, the lemma in Section 2.3 is proven.

2.4.4 Multivariate Integral Limits for the Matrix Square Root Result

2.4.4.1 Overview of this Appendix

In this appendix we deal with the problem of establishing that

$$\operatorname*{plim}_{n\to\infty} \int_0^\infty \boldsymbol{F}_{kk'n}(u;\boldsymbol{K},\boldsymbol{L})\,du = \boldsymbol{O}, \quad k,k'=1,2.$$

The approach involves spectral norm bounds on the integrands $\boldsymbol{F}_{kk'n}(u;\boldsymbol{K},\boldsymbol{L})$ uniformly over $u>0$ and exact integral results over the positive half-line for functions of x with factors of the form $1/(a_j+x^2)$, $a_j>0$.

2.4.4.2 Computing Spectral Norms

Bounding of $\|\boldsymbol{\Gamma}_{1n}(u)\|_S$

Recall that

$$\boldsymbol{\Gamma}_{1n}(u) \equiv \left(\frac{1}{n}\boldsymbol{L} + u^2\boldsymbol{I}_q + T_n\mathbf{1}_q^{\otimes 2}\right)^{-1}.$$

Then, noting that $\boldsymbol{\Gamma}_{1n}(u)$ is symmetric and positive definite,

$$\|\boldsymbol{\Gamma}_{1n}(u)\|_S = \lambda_{\max}\left\{\left(\frac{1}{n}\boldsymbol{L} + u^2\boldsymbol{I}_q + T_n\mathbf{1}_q^{\otimes 2}\right)^{-1}\right\} = 1/\lambda_{\min}\left(\frac{1}{n}\boldsymbol{L} + u^2\boldsymbol{I}_q + T_n\mathbf{1}_q^{\otimes 2}\right).$$

Application of (1.9) leads to

$$\begin{aligned}\lambda_{\min}\left(\frac{1}{n}\boldsymbol{L} + u^2\boldsymbol{I}_q + T_n\mathbf{1}_q^{\otimes 2}\right) &\geq \lambda_{\min}\left(\frac{1}{n}\boldsymbol{L}\right) + \lambda_{\min}\left(u^2\boldsymbol{I}_q\right) + \lambda_{\min}\left(T_n\mathbf{1}_q^{\otimes 2}\right)\\ &= \frac{1}{n}\lambda_{\min}\left(\boldsymbol{L}\right) + u^2\lambda_{\min}\left(\boldsymbol{I}_q\right) + T_n\lambda_{\min}\left(\mathbf{1}_q^{\otimes 2}\right)\\ &\geq \frac{1}{n}\lambda_{\min}\left(\boldsymbol{L}\right) + u^2 + \min\left(T_n,0\right).\end{aligned}$$

Since $T_n = o_P(n^{-1})$, for every $0<\varepsilon\leq 1$ we can choose $n_1\in\mathbb{N}$ such that, for all $n>n_1$, $|T_n| < \frac{1}{2n}\lambda_{\min}(\boldsymbol{L})$ with probability exceeding $1-\varepsilon$. For all such large n we then have

$$T_n > -\frac{1}{2n}\lambda_{\min}\left(\boldsymbol{L}\right).$$

Also,

$$0 > -\frac{1}{2n}\lambda_{\min}\left(\boldsymbol{L}\right)$$

and therefore we have,

$$\min\left(T_n,0\right) > -\frac{1}{2n}\lambda_{\min}\left(\boldsymbol{L}\right).$$

Therefore, for all such large n, we then have

$$\lambda_{\min}\left(\frac{1}{n}\boldsymbol{L}+u^2\boldsymbol{I}_q+T_n\mathbf{1}_q^{\otimes 2}\right)>\frac{1}{2n}\lambda_{\min}\left(\boldsymbol{L}\right)+u^2 \quad \text{for all} \quad u>0.$$

This leads to

$$\|\boldsymbol{\Gamma}_{1n}(u)\|_S<\frac{1}{\frac{1}{2n}\lambda_{\min}\left(\boldsymbol{L}\right)+u^2} \quad \text{for all } n>n_1 \text{ and } u>0$$

with probability exceeding $1-\varepsilon$.

Bounding of $\|\boldsymbol{\Gamma}_{2n}(u)\|_S$

Recall that

$$\begin{aligned}\boldsymbol{\Gamma}_{2n}(u)&\equiv R_n^2\mathbf{1}_p\mathbf{1}_q^T\boldsymbol{\Gamma}_{1n}(u)\mathbf{1}_q\mathbf{1}_p^T-Q_n\mathbf{1}_p^{\otimes 2}\\&=\left\{R_n^2\mathbf{1}_q^T\boldsymbol{\Gamma}_{1n}(u)\mathbf{1}_q-Q_n\right\}\mathbf{1}_p^{\otimes 2}.\end{aligned}$$

Also note that for n large enough and all $u>0$

$$\|\boldsymbol{\Gamma}_{1n}(u)\|_S<\frac{1}{\frac{1}{2n}\lambda_{\min}\left(\boldsymbol{L}\right)+u^2}<\frac{1}{\frac{1}{2n}\lambda_{\min}\left(\boldsymbol{L}\right)}=\frac{2n}{\lambda_{\min}\left(\boldsymbol{L}\right)}.$$

Then

$$\begin{aligned}\|\boldsymbol{\Gamma}_{2n}(u)\|_S&=\|\left\{R_n^2\mathbf{1}_q^T\boldsymbol{\Gamma}_{1n}(u)\mathbf{1}_q-Q_n\right\}\mathbf{1}_p^{\otimes 2}\|_S\\&\le\|R_n^2\mathbf{1}_q^T\boldsymbol{\Gamma}_{1n}(u)\mathbf{1}_q-Q_n\|_S\|\mathbf{1}_p^{\otimes 2}\|_S\\&\le p\left\{R_n^2\|\mathbf{1}_q^T\boldsymbol{\Gamma}_{1n}(u)\mathbf{1}_q\|_S+\|Q_n\|_S\right\}\\&\le p\left\{qR_n^2\|\boldsymbol{\Gamma}_{1n}(u)\|_S+|Q_n|\right\}\\&<p\left\{\frac{2qnR_n^2}{\lambda_{\min}\left(\boldsymbol{L}\right)}+|Q_n|\right\}.\end{aligned}$$

In summary,

$$\|\boldsymbol{\Gamma}_{2n}(u)\|_S<p\left\{\frac{2qnR_n^2}{\lambda_{\min}\left(\boldsymbol{L}\right)}+|Q_n|\right\} \quad \text{for all } u>0$$

with probability exceeding $1-\varepsilon$.

Bounding of $\|\boldsymbol{\Gamma}_{3n}(u)\|_S$

Recall that

$$\boldsymbol{\Gamma}_{3n}(u)\equiv R_n^2\mathbf{1}_q\mathbf{1}_p^T\left(\boldsymbol{K}+u^2\boldsymbol{I}_p+Q_n\mathbf{1}_p^{\otimes 2}\right)^{-1}\mathbf{1}_p\mathbf{1}_q^T-T_n\mathbf{1}_q^{\otimes 2}.$$

Then

$$\begin{aligned}\|\boldsymbol{\Gamma}_{3n}(u)\|_S&\le R_n^2\|\mathbf{1}_q\|_S\|\mathbf{1}_p^T\|_S\|\left(\boldsymbol{K}+u^2\boldsymbol{I}_p+Q_n\mathbf{1}_p^{\otimes 2}\right)^{-1}\|_S\|\mathbf{1}_p\|_S\|\mathbf{1}_q^T\|_S+|T_n|\|\mathbf{1}_q^{\otimes 2}\|_S\\&=\frac{pqR_n^2}{\lambda_{\min}\left(\boldsymbol{K}+u^2\boldsymbol{I}_p+Q_n\mathbf{1}_p^{\otimes 2}\right)}+q|T_n|.\end{aligned}$$

Now,

$$\begin{aligned}\lambda_{\min}\left(\boldsymbol{K}+u^2\boldsymbol{I}_p+Q_n\mathbf{1}_p^{\otimes 2}\right) &\geq \lambda_{\min}(\boldsymbol{K})+\lambda_{\min}(u^2\boldsymbol{I}_p)+\lambda_{\min}\left(Q_n\mathbf{1}_p^{\otimes 2}\right)\\ &\geq \lambda_{\min}(\boldsymbol{K})+u^2+\min\left(Q_n,0\right).\end{aligned}$$

Since $Q_n = o_P(1)$, for every $0<\varepsilon\leq 1$ we can choose $n_3\in\mathbb{N}$ such that, for all $n>n_3$, $|Q_n|<\frac{1}{2}\lambda_{\min}\left(\boldsymbol{K}\right)$ with probability exceeding $1-\varepsilon$. For all such large n we then have

$$Q_n > -\frac{1}{2}\lambda_{\min}\left(\boldsymbol{K}\right).$$

Also,

$$0 > -\frac{1}{2}\lambda_{\min}\left(\boldsymbol{K}\right)$$

and therefore we have,

$$\min\left(Q_n,0\right) > -\frac{1}{2}\lambda_{\min}\left(\boldsymbol{K}\right).$$

Therefore, for all such large n, we then have

$$\lambda_{\min}\left(\boldsymbol{K}+u^2\boldsymbol{I}_p+Q_n\mathbf{1}_p^{\otimes 2}\right) > \frac{1}{2}\lambda_{\min}\left(\boldsymbol{K}\right)+u^2.$$

Therefore,

$$\|\boldsymbol{\Gamma}_{3n}(u)\|_S < \frac{pqR_n^2}{\frac{1}{2}\lambda_{\min}\left(\boldsymbol{K}\right)+u^2}+q|T_n| \quad \text{for all } n>n_3 \text{ and } u>0.$$

Bounding of $\|\boldsymbol{\Gamma}_{4n}(u)\|_S$

Recall that

$$\boldsymbol{\Gamma}_{4n}(u)\equiv\left\{\boldsymbol{K}+u^2\boldsymbol{I}_p-\boldsymbol{\Gamma}_{2n}(u)\right\}^{-1}.$$

Hence,

$$\begin{aligned}\|\boldsymbol{\Gamma}_{4n}(u)\|_S &= \lambda_{\max}\left\{\left(\boldsymbol{K}+u^2\boldsymbol{I}_p-\boldsymbol{\Gamma}_{2n}(u)\right)^{-1}\right\}\\ &= 1/\lambda_{\min}\left(\boldsymbol{K}+u^2\boldsymbol{I}_p-\boldsymbol{\Gamma}_{2n}(u)\right).\end{aligned}$$

Then, from (1.9) we have

$$\begin{aligned}\lambda_{\min}\left(\boldsymbol{K}+u^2\boldsymbol{I}_p-\boldsymbol{\Gamma}_{2n}(u)\right) &\geq \lambda_{\min}\left(\boldsymbol{K}+u^2\boldsymbol{I}_p\right)+\lambda_{\min}\left(-\boldsymbol{\Gamma}_{2n}(u)\right)\\ &\geq \lambda_{\min}\left(\boldsymbol{K}\right)+\lambda_{\min}\left(u^2\boldsymbol{I}_p\right)-\lambda_{\min}\left(\boldsymbol{\Gamma}_{2n}(u)\right)\\ &\geq \lambda_{\min}\left(\boldsymbol{K}\right)+u^2-\lambda_{\max}\left(\boldsymbol{\Gamma}_{2n}(u)\right)\\ &= \lambda_{\min}\left(\boldsymbol{K}\right)+u^2-\|\boldsymbol{\Gamma}_{2n}(u)\|_S\\ &\geq \lambda_{\min}\left(\boldsymbol{K}\right)+u^2-p\left\{\frac{2qnR_n^2}{\lambda_{\min}\left(\boldsymbol{L}\right)}+|Q_n|\right\}.\end{aligned}$$

Since $R_n = O_P(n^{-1})$ and $Q_n = o_P(1)$, for every for every $0<\varepsilon\leq 1$ we can choose $n_4\in\mathbb{N}$ such that, for all $n>n_4$,

$$p\left\{\frac{2qnR_n^2}{\lambda_{\min}\left(\boldsymbol{L}\right)}+|Q_n|\right\} < \frac{1}{2}\lambda_{\min}\left(\boldsymbol{K}\right).$$

This is equivalent to

$$-p\left\{\frac{2qnR_n^2}{\lambda_{\min}(\boldsymbol{L})}+|Q_n|\right\}>-\frac{1}{2}\lambda_{\min}(\boldsymbol{K}).$$

This allows us to claim that for all $n > n_4$,

$$\lambda_{\min}\left(\boldsymbol{K}+u^2\boldsymbol{I}_p-\boldsymbol{\Gamma}_{2n}(u)\right)>\frac{1}{2}\lambda_{\min}(\boldsymbol{K})+u^2 \quad \text{for all } u>0.$$

Hence, we have,

$$\|\boldsymbol{\Gamma}_{4n}(u)\|_S<\frac{1}{\frac{1}{2}\lambda_{\min}(\boldsymbol{K})+u^2} \quad \text{for all } n>n_4 \text{ and } u>0$$

with probability exceeding $1-\varepsilon$.

Bounding of $\|\boldsymbol{\Gamma}_{5n}(u)\|_S$

Recall that

$$\boldsymbol{\Gamma}_{5n}(u)\equiv\left\{\frac{1}{n}\boldsymbol{L}+u^2\boldsymbol{I}_q-\boldsymbol{\Gamma}_{3n}(u)\right\}^{-1}.$$

Then,

$$\|\boldsymbol{\Gamma}_{5n}(u)\|_S=1/\lambda_{\min}\left(\frac{1}{n}\boldsymbol{L}+u^2\boldsymbol{I}_q-\boldsymbol{\Gamma}_{3n}(u)\right).$$

Next,

$$\begin{aligned}\lambda_{\min}\left(\frac{1}{n}\boldsymbol{L}+u^2\boldsymbol{I}_q-\boldsymbol{\Gamma}_{3n}(u)\right)&\geq\lambda_{\min}\left(\frac{1}{n}\boldsymbol{L}\right)+u^2\lambda_{\min}(\boldsymbol{I}_q)-\lambda_{\min}(\boldsymbol{\Gamma}_{3n}(u))\\&\geq\frac{1}{n}\lambda_{\min}(\boldsymbol{L})+u^2-\lambda_{\max}(\boldsymbol{\Gamma}_{3n}(u))\\&=\frac{1}{n}\lambda_{\min}(\boldsymbol{L})+u^2-\|\boldsymbol{\Gamma}_{3n}(u)\|_S\\&>\frac{1}{n}\lambda_{\min}(\boldsymbol{L})+u^2-\frac{pqR_n^2}{\frac{1}{2}\lambda_{\min}(\boldsymbol{K})+u^2}-q|T_n|.\end{aligned}$$

Since $R_n=O_P(n^{-1})$ and $T_n=o_P(n^{-1})$, for every for every $0<\varepsilon\leq 1$ we can choose $n_5\in\mathbb{N}$ such that for all sufficiently large n,

$$\frac{pqR_n^2}{\frac{1}{2}\lambda_{\min}(\boldsymbol{K})+u^2}+q|T_n|<\frac{1}{2n}\lambda_{\min}(\boldsymbol{L}).$$

This is equivalent to

$$-\frac{pqR_n^2}{\frac{1}{2}\lambda_{\min}(\boldsymbol{K})+u^2}-q|T_n|>-\frac{1}{2n}\lambda_{\min}(\boldsymbol{L}).$$

This allows us to claim that for all $n > n_5$,

$$\lambda_{\min}\left(\frac{1}{n}\boldsymbol{L}+u^2\boldsymbol{I}_q-\boldsymbol{\Gamma}_{3n}(u)\right)>\frac{1}{2n}\lambda_{\min}(\boldsymbol{L})+u^2.$$

Therefore, we have,

$$\|\boldsymbol{\Gamma}_{5n}(u)\|_S < \frac{1}{\frac{1}{2n}\lambda_{\min}(\boldsymbol{L}) + u^2} \quad \text{for all } n > n_5 \text{ and } u > 0$$

with probability exceeding $1 - \varepsilon$.

Bounding of $\|\boldsymbol{F}_{11n}(u; \boldsymbol{K}, \boldsymbol{L})\|_S$

Recall that

$$\boldsymbol{F}_{11n}(u; \boldsymbol{K}, \boldsymbol{L}) \equiv \boldsymbol{K}\boldsymbol{\Gamma}_{4n}(u)\boldsymbol{\Gamma}_{2n}(u)\left(\boldsymbol{K} + u^2\boldsymbol{I}_p\right)^{-1} - \boldsymbol{\Gamma}_{2n}(u)\boldsymbol{\Gamma}_{4n}(u).$$

Hence

$$\|\boldsymbol{F}_{11n}(u; \boldsymbol{K}, \boldsymbol{L})\|_S \leq \|\boldsymbol{K}\|_S\|\boldsymbol{\Gamma}_{4n}(u)\|_S\|\boldsymbol{\Gamma}_{2n}(u)\|_S\|\left(\boldsymbol{K} + u^2\boldsymbol{I}_p\right)^{-1}\|_S + \|\boldsymbol{\Gamma}_{2n}(u)\|_S\|\boldsymbol{\Gamma}_{4n}(u)\|_S.$$

Now,

$$\|\boldsymbol{K}\|_S = \lambda_{\max}(\boldsymbol{K})$$

and

$$\|\left(\boldsymbol{K} + u^2\boldsymbol{I}_p\right)^{-1}\|_S = \frac{1}{\lambda_{\min}(\boldsymbol{K} + u^2\boldsymbol{I}_p)} \leq \frac{1}{\lambda_{\min}(\boldsymbol{K}) + u^2}.$$

Hence, for all sufficiently large n and $u > 0$,

$$\|\boldsymbol{F}_{11n}(u; \boldsymbol{K}, \boldsymbol{L})\|_S \leq \left\{\frac{\lambda_{\max}(\boldsymbol{K})}{\lambda_{\min}(\boldsymbol{K}) + u^2} + 1\right\}\|\boldsymbol{\Gamma}_{2n}(u)\|_S\|\boldsymbol{\Gamma}_{4n}(u)\|_S.$$

Therefore, for all sufficiently large n and $u > 0$,

$$\|\boldsymbol{F}_{11n}(u; \boldsymbol{K}, \boldsymbol{L})\|_S < p\left\{\frac{2qnR_n^2}{\lambda_{\min}(\boldsymbol{L})} + |Q_n|\right\}\left\{\frac{\lambda_{\max}(\boldsymbol{K})}{\lambda_{\min}(\boldsymbol{K}) + u^2} + 1\right\}\left\{\frac{1}{\frac{1}{2}\lambda_{\min}(\boldsymbol{K}) + u^2}\right\} \tag{2.56}$$

with probability exceeding $1 - \varepsilon$.

Bounding of $\|\boldsymbol{F}_{21n}(u; \boldsymbol{K}, \boldsymbol{L})\|_S$

Recall that

$$\boldsymbol{F}_{21n}(u; \boldsymbol{K}, \boldsymbol{L}) \equiv n^{1/2}R_n u^2\boldsymbol{\Gamma}_{1n}(u)\mathbf{1}_q\mathbf{1}_p^T\boldsymbol{\Gamma}_{4n}(u).$$

Then,

$$\begin{aligned}\|\boldsymbol{F}_{21n}(u; \boldsymbol{K}, \boldsymbol{L})\|_S &\leq n^{1/2}|R_n|u^2\|\boldsymbol{\Gamma}_{1n}(u)\|_S\|\mathbf{1}_q\|_S\|\mathbf{1}_p^T\|_S\|\boldsymbol{\Gamma}_{4n}(u)\|_S \\ &= \sqrt{pq}n^{1/2}|R_n|u^2\|\boldsymbol{\Gamma}_{1n}(u)\|_S\|\boldsymbol{\Gamma}_{4n}(u)\|_S.\end{aligned}$$

Hence, for all sufficiently large n we have,

$$\|\boldsymbol{F}_{21n}(u; \boldsymbol{K}, \boldsymbol{L})\|_S < \frac{\sqrt{pq}n^{1/2}|R_n|u^2}{\left\{\frac{1}{2n}\lambda_{\min}(\boldsymbol{L}) + u^2\right\}\left\{\frac{1}{2}\lambda_{\min}(\boldsymbol{K}) + u^2\right\}} \quad \text{for all } u > 0 \tag{2.57}$$

with probability exceeding $1-\varepsilon$.

Bounding of $\|\boldsymbol{F}_{12n}(u;\boldsymbol{K},\boldsymbol{L})\|_S$

Recall that

$$\boldsymbol{F}_{12n}(u;\boldsymbol{K},\boldsymbol{L}) \equiv R_n\{\mathbf{1}_p\mathbf{1}_q^T\boldsymbol{\Gamma}_{5n}(u) - (\boldsymbol{K}+Q_n\mathbf{1}_p^{\otimes 2})\boldsymbol{\Gamma}_{4n}(u)\mathbf{1}_p\mathbf{1}_q^T\boldsymbol{\Gamma}_{1n}(u)\}.$$

Then

$$\begin{aligned}\|\boldsymbol{F}_{12n}(u;\boldsymbol{K},\boldsymbol{L})\|_S &\le |R_n|\Big\{\|\mathbf{1}_p\|_S\|\mathbf{1}_q^T\|_S\|\boldsymbol{\Gamma}_{5n}(u)\|_S \\ &\quad + \left(\|\boldsymbol{K}\|_S + |Q_n|\|\mathbf{1}_p^{\otimes 2}\|_S\right)\|\boldsymbol{\Gamma}_{4n}(u)\|_S\|\mathbf{1}_p\|_S\|\mathbf{1}_q^T\|_S\|\boldsymbol{\Gamma}_{1n}(u)\|_S\Big\} \\ &= |R_n|\left\{\sqrt{pq}\|\boldsymbol{\Gamma}_{5n}(u)\|_S + \sqrt{pq}\left(\lambda_{\max}(\boldsymbol{K})+p|Q_n|\right)\|\boldsymbol{\Gamma}_{4n}(u)\|_S\|\boldsymbol{\Gamma}_{1n}(u)\|_S\right\} \\ &= \sqrt{pq}|R_n|\left\{\|\boldsymbol{\Gamma}_{5n}(u)\|_S + \left(\lambda_{\max}(\boldsymbol{K})+p|Q_n|\right)\|\boldsymbol{\Gamma}_{4n}(u)\|_S\|\boldsymbol{\Gamma}_{1n}(u)\|_S\right\}.\end{aligned}$$

For all sufficiently large n and $u>0$, we have,

$$\begin{aligned}\|\boldsymbol{F}_{12n}(u;\boldsymbol{K},\boldsymbol{L})\|_S < \sqrt{pq}|R_n|\Bigg\{&\frac{1}{\frac{1}{2n}\lambda_{\min}(\boldsymbol{L})+u^2} \\ &+ \left(\lambda_{\max}(\boldsymbol{K})+p|Q_n|\right)\left(\frac{1}{\frac{1}{2}\lambda_{\min}(\boldsymbol{K})+u^2}\right)\left(\frac{1}{\frac{1}{2n}\lambda_{\min}(\boldsymbol{L})+u^2}\right)\Bigg\}\end{aligned} \tag{2.58}$$

with probability exceeding $1-\varepsilon$.

Bounding of $\|\boldsymbol{F}_{22n}(u;\boldsymbol{K},\boldsymbol{L})\|_S$

Recall that

$$\begin{aligned}\boldsymbol{F}_{22n}(u;\boldsymbol{K},\boldsymbol{L}) \equiv n^{1/2}\Bigg[&\left(\frac{1}{n}\boldsymbol{L}\right)\boldsymbol{\Gamma}_{5n}(u)\boldsymbol{\Gamma}_{3n}(u)\left\{\frac{1}{n}\boldsymbol{L}+u^2\boldsymbol{I}_q\right\}^{-1} \\ &+ T_n\mathbf{1}_q^{\otimes 2}\boldsymbol{\Gamma}_{5n}(u) - R_n^2\mathbf{1}_q\mathbf{1}_p^T\boldsymbol{\Gamma}_{4n}(u)\mathbf{1}_p\mathbf{1}_q^T\boldsymbol{\Gamma}_{1n}(u)\Bigg].\end{aligned}$$

We then have

$$\begin{aligned}\|\boldsymbol{F}_{22n}(u;\boldsymbol{K},\boldsymbol{L})\|_S &\le n^{1/2}\Bigg[\left\|\frac{1}{n}\boldsymbol{L}\right\|_S\|\boldsymbol{\Gamma}_{5n}(u)\|_S\|\boldsymbol{\Gamma}_{3n}(u)\|_S\left\|\left\{\frac{1}{n}\boldsymbol{L}+u^2\boldsymbol{I}_q\right\}^{-1}\right\|_S \\ &\quad + |T_n|\|\mathbf{1}_q^{\otimes 2}\|_S\|\boldsymbol{\Gamma}_{5n}(u)\|_S + R_n^2\|\mathbf{1}_q\|_S\|\mathbf{1}_p^T\|_S\|\boldsymbol{\Gamma}_{4n}(u)\|_S\|\mathbf{1}_p\|_S\|\mathbf{1}_q^T\|_S\|\boldsymbol{\Gamma}_{1n}(u)\|_S\Bigg] \\ &= n^{1/2}\Bigg[\frac{1}{n}\lambda_{\max}(\boldsymbol{L})\|\boldsymbol{\Gamma}_{5n}(u)\|_S\|\boldsymbol{\Gamma}_{3n}(u)\|_S\left\|\left\{\frac{1}{n}\boldsymbol{L}+u^2\boldsymbol{I}_q\right\}^{-1}\right\|_S \\ &\quad + q|T_n|\|\boldsymbol{\Gamma}_{5n}(u)\|_S + pqR_n^2\|\boldsymbol{\Gamma}_{4n}(u)\|_S\|\boldsymbol{\Gamma}_{1n}(u)\|_S\Bigg].\end{aligned} \tag{2.59}$$

Next, note that

$$\left\| \left\{ \frac{1}{n}\boldsymbol{L} + u^2 \boldsymbol{I}_q \right\}^{-1} \right\|_S \leq 1/\lambda_{\min}\left(\frac{1}{n}\boldsymbol{L} + u^2 \boldsymbol{I}_q \right).$$

Since

$$\lambda_{\min}\left(\frac{1}{n}\boldsymbol{L} + u^2 \boldsymbol{I}_q \right) \geq \frac{1}{n}\lambda_{\min}(\boldsymbol{L}) + u^2$$

we have the bound

$$\left\| \left\{ \frac{1}{n}\boldsymbol{L} + u^2 \boldsymbol{I}_q \right\}^{-1} \right\|_S \leq \frac{1}{\frac{1}{n}\lambda_{\min}(\boldsymbol{L}) + u^2} \quad \text{for all } u > 0.$$

It follows that for all sufficiently large n, the first term on the right-hand side of (2.59) is bounded above by

$$\begin{aligned} &\frac{n^{-1/2}\lambda_{\max}(\boldsymbol{L})}{\frac{1}{n}\lambda_{\min}(\boldsymbol{L}) + u^2} \|\boldsymbol{\Gamma}_{5n}(u)\|_S \|\boldsymbol{\Gamma}_{3n}(u)\|_S \\ &\quad < \left\{ \frac{n^{-1/2}\lambda_{\max}(\boldsymbol{L})}{\frac{1}{n}\lambda_{\min}(\boldsymbol{L}) + u^2} \right\} \left\{ \frac{1}{\frac{1}{2n}\lambda_{\min}(\boldsymbol{L}) + u^2} \right\} \left\{ \frac{pqR_n^2}{\frac{1}{2}\lambda_{\min}(\boldsymbol{K}) + u^2} + q|T_n| \right\}. \end{aligned}$$

It follows that for all sufficiently large n, the second term on the right-hand side of (2.59) is bounded above by

$$\frac{qn^{1/2}|T_n|}{\frac{1}{2n}\lambda_{\min}(\boldsymbol{L}) + u^2}.$$

Finally, it follows that for all sufficiently large n, the third term on the right-hand side of (2.59) is bounded above by

$$\frac{pqn^{1/2}R_n^2}{\left(\frac{1}{2n}\lambda_{\min}(\boldsymbol{L}) + u^2\right)\left(\frac{1}{2}\lambda_{\min}(\boldsymbol{K}) + u^2\right)}.$$

Combining all of these bounds, we have

$$\begin{aligned} \|\boldsymbol{F}_{22n}(u; \boldsymbol{K}, \boldsymbol{L})\|_S < &\left\{ \frac{n^{-1/2}\lambda_{\max}(\boldsymbol{L})}{\frac{1}{n}\lambda_{\min}(\boldsymbol{L}) + u^2} \right\} \left\{ \frac{1}{\frac{1}{2n}\lambda_{\min}(\boldsymbol{L}) + u^2} \right\} \left\{ \frac{pqR_n^2}{\frac{1}{2}\lambda_{\min}(\boldsymbol{K}) + u^2} + q|T_n| \right\} \\ &+ \frac{qn^{1/2}|T_n|}{\frac{1}{2n}\lambda_{\min}(\boldsymbol{L}) + u^2} + \frac{pqn^{1/2}R_n^2}{\left(\frac{1}{2n}\lambda_{\min}(\boldsymbol{L}) + u^2\right)\left(\frac{1}{2}\lambda_{\min}(\boldsymbol{K}) + u^2\right)} \end{aligned} \tag{2.60}$$

for all sufficiently large n and $u > 0$, with probability exceeding $1 - \varepsilon$.

2.4.4.3 Verifying Convergence in Probability Limits of the Functions in (2.55)

The Convergence in Probability Limit of $\int_0^\infty \boldsymbol{F}_{11n}(u;\boldsymbol{\kappa},\boldsymbol{K})\,du$

Noting that

$$\left\|\int_0^\infty \boldsymbol{F}_{11n}(u;\boldsymbol{\kappa},\boldsymbol{K})\,du\right\|_S \leq \int_0^\infty \|\boldsymbol{F}_{11n}(u;\boldsymbol{\kappa},\boldsymbol{K})\|_S\,du$$

and from (2.56) we have, for all sufficiently large n

$$\begin{aligned}
&\int_0^\infty \|\boldsymbol{F}_{11n}(u;\boldsymbol{\kappa},\boldsymbol{K})\|_S\,du \\
&\quad< p\left\{\frac{2qnR_n^2}{\lambda_{\min}(\boldsymbol{L})}+|Q_n|\right\}\int_0^\infty\left\{\frac{\lambda_{\max}(\boldsymbol{K})}{\lambda_{\min}(\boldsymbol{K})+u^2}+1\right\}\left\{\frac{1}{\frac{1}{2}\lambda_{\min}(\boldsymbol{K})+u^2}\right\}du \\
&\quad< p\left\{\frac{2qnR_n^2}{\lambda_{\min}(\boldsymbol{L})}+|Q_n|\right\}\left\{\int_0^\infty\frac{\lambda_{\max}(\boldsymbol{K})}{\left(\frac{1}{2}\lambda_{\min}(\boldsymbol{K})+u^2\right)^2}\,du+\int_0^\infty\frac{1}{\frac{1}{2}\lambda_{\min}(\boldsymbol{K})+u^2}\,du\right\} \\
&\quad= p\left\{\frac{2qnR_n^2}{\lambda_{\min}(\boldsymbol{L})}+|Q_n|\right\}\left(\frac{\pi\lambda_{\max}(\boldsymbol{K})}{\lambda_{\min}(\boldsymbol{K})\sqrt{2\lambda_{\min}(\boldsymbol{K})}}+\frac{\pi}{\sqrt{2\lambda_{\min}(\boldsymbol{K})}}\right)
\end{aligned}$$

with probability exceeding $1-\varepsilon$. Since $R_n = O_P(n^{-1})$, $Q_n = o_P(1)$ and ε is arbitrary, we must have

$$\left\|\int_0^\infty \boldsymbol{F}_{11n}(u;\boldsymbol{\kappa},\boldsymbol{K})\,du\right\|_S \xrightarrow{P} 0 \quad \text{as } n\to\infty$$

and therefore,

$$\int_0^\infty \boldsymbol{F}_{11n}(u;\boldsymbol{\kappa},\boldsymbol{K})\,du \xrightarrow{P} \boldsymbol{O} \quad \text{as } n\to\infty.$$

The Convergence in Probability Limit of $\int_0^\infty \boldsymbol{F}_{21n}(u;\boldsymbol{\kappa},\boldsymbol{K})\,du$

Noting that

$$\left\|\int_0^\infty \boldsymbol{F}_{21n}(u;\boldsymbol{\kappa},\boldsymbol{K})\,du\right\|_S \leq \int_0^\infty \|\boldsymbol{F}_{21n}(u;\boldsymbol{\kappa},\boldsymbol{K})\|_S\,du$$

and from (2.57) we have, for all sufficiently large n

$$\begin{aligned}
&\int_0^\infty \|\boldsymbol{F}_{21n}(u;\boldsymbol{\kappa},\boldsymbol{K})\|_S\,du \\
&\quad< \sqrt{pq}n^{1/2}|R_n|\int_0^\infty\frac{u^2}{\left\{\frac{1}{2n}\lambda_{\min}(\boldsymbol{L})+u^2\right\}\left\{\frac{1}{2}\lambda_{\min}(\boldsymbol{K})+u^2\right\}}\,du \\
&\quad= \frac{\sqrt{pq}\pi n^{1/2}|R_n|}{\sqrt{2}\left(\sqrt{\lambda_{\min}(\boldsymbol{K})}+\sqrt{\lambda_{\min}(\boldsymbol{L})/n}\right)}
\end{aligned}$$

with probability exceeding $1-\varepsilon$. Since $R_n = O_P(n^{-1})$ and ε is arbitrary, we must have

$$\left\| \int_0^\infty \boldsymbol{F}_{21n}(u;\boldsymbol{\kappa},\boldsymbol{K})\,du \right\|_S \xrightarrow{P} 0 \quad \text{as } n\to\infty$$

and therefore,

$$\int_0^\infty \boldsymbol{F}_{21n}(u;\boldsymbol{\kappa},\boldsymbol{K})\,du \xrightarrow{P} \boldsymbol{O} \quad \text{as } n\to\infty.$$

The Convergence in Probability Limit of $\int_0^\infty \boldsymbol{F}_{12n}(u;\boldsymbol{\kappa},\boldsymbol{K})\,du$

Noting that

$$\left\| \int_0^\infty \boldsymbol{F}_{12n}(u;\boldsymbol{\kappa},\boldsymbol{K})\,du \right\|_S \le \int_0^\infty \|\boldsymbol{F}_{12n}(u;\boldsymbol{\kappa},\boldsymbol{K})\|_S\,du$$

and from (2.58) we have, for all sufficiently large n

$$\begin{aligned}
&\int_0^\infty \|\boldsymbol{F}_{12n}(u;\boldsymbol{\kappa},\boldsymbol{K})\|_S\,du \\
&\quad< \sqrt{pq}|R_n|\left\{ \int_o^\infty \frac{1}{\frac{1}{2n}\lambda_{\min}(\boldsymbol{L})+u^2}\,du \right. \\
&\qquad \left. + (\lambda_{\max}(\boldsymbol{K})+p|Q_n|)\int_0^\infty \left(\frac{1}{\frac{1}{2}\lambda_{\min}(\boldsymbol{K})+u^2}\right)\left(\frac{1}{\frac{1}{2n}\lambda_{\min}(\boldsymbol{L})+u^2}\right)du \right\} \\
&\quad= \sqrt{pq}|R_n|\left[\frac{\pi}{2\sqrt{\frac{1}{2n}\lambda_{\min}(\boldsymbol{L})}} \right. \\
&\qquad \left. + (\lambda_{\max}(\boldsymbol{K})+p|Q_n|)\left\{ \frac{\pi}{2\sqrt{\left(\frac{1}{2}\lambda_{\min}(\boldsymbol{K})\right)\left(\frac{1}{2n}\lambda_{\min}(\boldsymbol{L})\right)}\left(\sqrt{\frac{1}{2}\lambda_{\min}(\boldsymbol{K})}+\sqrt{\frac{1}{2n}\lambda_{\min}(\boldsymbol{L})}\right)} \right\} \right] \\
&\quad= \sqrt{pq}|R_n|\left[\frac{n^{1/2}\pi}{\sqrt{2\lambda_{\min}(\boldsymbol{L})}} \right. \\
&\qquad \left. + (\lambda_{\max}(\boldsymbol{K})+p|Q_n|)\left\{ \frac{\sqrt{2}n^{1/2}\pi}{\sqrt{\lambda_{\min}(\boldsymbol{K})\lambda_{\min}(\boldsymbol{L})}\left(\sqrt{\lambda_{\min}(\boldsymbol{K})}+\sqrt{\frac{1}{n}\lambda_{\min}(\boldsymbol{L})}\right)} \right\} \right]
\end{aligned}$$

with probability exceeding $1-\varepsilon$. Since $R_n = O_P(n^{-1})$, $Q_n = o_P(1)$ and ε is arbitrary, we must have

$$\left\| \int_0^\infty \boldsymbol{F}_{12n}(u;\boldsymbol{\kappa},\boldsymbol{K})\,du \right\|_S \xrightarrow{P} 0 \quad \text{as } n\to\infty$$

and therefore,

$$\int_0^\infty \boldsymbol{F}_{12n}(u;\boldsymbol{\kappa},\boldsymbol{K})\,du \xrightarrow{P} \boldsymbol{O} \quad \text{as } n\to\infty.$$

The Convergence in Probability Limit of $\int_0^\infty \boldsymbol{F}_{22n}(u;\boldsymbol{\kappa},\boldsymbol{K})\,du$

Noting that

$$\left\|\int_0^\infty \boldsymbol{F}_{22n}(u;\boldsymbol{\kappa},\boldsymbol{K})\,du\right\|_S \le \int_0^\infty \|\boldsymbol{F}_{22n}(u;\boldsymbol{\kappa},\boldsymbol{K})\|_S\,du$$

and from (2.60) we have, for all sufficiently large n

$$\begin{aligned}
&\int_0^\infty \|\boldsymbol{F}_{22n}(u;\boldsymbol{\kappa},\boldsymbol{K})\|_S\,du \\
&\quad< \lambda_{\max}(\boldsymbol{L})pqn^{-1/2}R_n^2\int_0^\infty \frac{1}{\left(\frac{1}{n}\lambda_{\min}(\boldsymbol{L})+u^2\right)\left(\frac{1}{2n}\lambda_{\min}(\boldsymbol{L})+u^2\right)\left(\frac{1}{2}\lambda_{\min}(\boldsymbol{K})+u^2\right)}\,du \\
&\qquad+ n^{-1/2}\lambda_{\max}(\boldsymbol{L})q|T_n|\int_0^\infty \frac{1}{\left(\frac{1}{n}\lambda_{\min}(\boldsymbol{L})+u^2\right)\left(\frac{1}{2n}\lambda_{\min}(\boldsymbol{L})+u^2\right)}\,du \\
&\qquad+ qn^{1/2}|T_n|\int_0^\infty \frac{1}{\frac{1}{2n}\lambda_{\min}(\boldsymbol{L})+u^2}\,du \\
&\qquad+ pqn^{1/2}R_n^2\int_0^\infty \frac{1}{\left(\frac{1}{2n}\lambda_{\min}(\boldsymbol{L})+u^2\right)\left(\frac{1}{2}\lambda_{\min}(\boldsymbol{K})+u^2\right)}\,du \\
&\quad< \lambda_{\max}(\boldsymbol{L})pqn^{-1/2}R_n^2\int_0^\infty \frac{1}{\left(\frac{1}{2n}\lambda_{\min}(\boldsymbol{L})+u^2\right)^2\left(\frac{1}{2}\lambda_{\min}(\boldsymbol{K})+u^2\right)}\,du \\
&\qquad+ n^{-1/2}\lambda_{\max}(\boldsymbol{L})q|T_n|\int_0^\infty \frac{1}{\left(\frac{1}{2n}\lambda_{\min}(\boldsymbol{L})+u^2\right)^2}\,du \\
&\qquad+ qn^{1/2}|T_n|\int_0^\infty \frac{1}{\frac{1}{2n}\lambda_{\min}(\boldsymbol{L})+u^2}\,du \\
&\qquad+ pqn^{1/2}R_n^2\int_0^\infty \frac{1}{\left(\frac{1}{2n}\lambda_{\min}(\boldsymbol{L})+u^2\right)\left(\frac{1}{2}\lambda_{\min}(\boldsymbol{K})+u^2\right)}\,du \\
&\quad< \frac{2\sqrt{2}pq\pi n\lambda_{\max}(\boldsymbol{L})R_n^2}{\lambda_{\min}(\boldsymbol{L})\sqrt{\lambda_{\min}(\boldsymbol{L})\lambda_{\min}(\boldsymbol{K})}\left(\sqrt{\frac{1}{2}\lambda_{\min}(\boldsymbol{L})}+\sqrt{\lambda_{\min}(\boldsymbol{K})}\right)} \\
&\qquad+ \frac{q\pi n\lambda_{\max}(\boldsymbol{K})|T_n|}{\sqrt{2}\lambda_{\min}(\boldsymbol{L})\sqrt{\lambda_{\min}(\boldsymbol{L})}}+\frac{q\pi n|T_n|}{\sqrt{2\lambda_{\min}(\boldsymbol{L})}} \\
&\qquad+ \frac{pq\pi n\sqrt{2}R_n^2}{\sqrt{\lambda_{\min}(\boldsymbol{L})\lambda_{\min}(\boldsymbol{K})}\left(\sqrt{\frac{1}{n}\lambda_{\min}(\boldsymbol{L})}+\sqrt{\lambda_{\min}(\boldsymbol{K})}\right)}
\end{aligned}$$

with probability exceeding $1-\varepsilon$. Since $R_n = O_P(n^{-1})$, $T_n = o_P(n^{-1})$ and ε is arbitrary, it follows that

$$\left\|\int_0^\infty \boldsymbol{F}_{22n}(u;\boldsymbol{\kappa},\boldsymbol{K})\,du\right\|_S \xrightarrow{P} 0 \quad \text{as } n\to\infty$$

and therefore,

$$\int_0^\infty \boldsymbol{F}_{22n}(u;\boldsymbol{\kappa},\boldsymbol{K})\,du \xrightarrow{P} \boldsymbol{O} \quad \text{as } n\to\infty.$$

2.4.4.4 Conclusion for Multivariate Integral Limits for the Matrix Square Root Result

Hence, we have shown that

$$\operatorname*{plim}_{n\to\infty} \int_0^\infty \boldsymbol{F}_{kk'n}(u; \boldsymbol{K}, \boldsymbol{L})\, du = \boldsymbol{O}, \quad k, k' = 1, 2$$

for $u > 0$.

Chapter 3

Usable Asymptotic Normality Results and Inference for Gaussian Response Linear Mixed Models

Even though estimation by maximum likelihood for linear mixed models and generalized linear mixed models is well established, asymptotic normality results that can be used to construct confidence intervals and Wald tests via Studentization are currently unavailable in the existing linear mixed model and generalized linear mixed model literature.

Asymptotic normality results for maximum likelihood estimators for Gaussian response linear mixed models have been presented in literature such as McCulloch et al. (2008), Miller (1973), Miller (1977), Jiang (1996) and Jiang and Nguyen (2021). As discussed in Section 1.9.3, Nie (2007) investigated an extension to a generalized linear mixed model setting, but did not give explicit forms void of limits or expectations with respect to the response. Hence, the existing *generalized* linear mixed model literature lacks asymptotic covariance results that are amenable to practical purposes such as confidence interval construction.

Other recent related literature by Lyu and Welsh (2022) derive explicit asymptotic normality results for both maximum likelihood estimators and restricted maximum likelihood estimators for model parameters in a nested regression model (random intercept model) for clustered data. The authors, as done in this thesis, considered the scenario where both the number of independent clusters and number of observations within each cluster go to infinity. When restricted to Gaussian responses, the work by

Some of the content of this chapter is published in: Bhaskaran, A. and Wand, M.P.(2023), Dispersion parameter extension of precise generalized linear mixed model asymptotics. Statistics and Probability Letters, 193, Article 109691.

Lyu and Welsh (2022) is closely related to the work in this chapter. However, while Lyu and Welsh (2022) consider regression models with a random intercept, this chapter considers Gaussian response linear mixed models that allow for both a random intercept and slope to be included. On the other hand, Westfall (1986) considers a linear mixed model set-up with vectors of fixed effects, nonerror random effects and error random effects and develop asymptotic distribution theory for the corresponding analysis of variance estimates. In contrast, this chapter presents the explicit first order (leading term) asymptotic approximations used for the asymptotic normality results in this chapter, serving as a prequel to the results presented for the generalized linear mixed model case in Chapter 4.

In this chapter, we aim to derive asymptotic normality results that are directly usable for asymptotically valid confidence intervals and Wald tests for analysis concerning linear mixed models with Gaussian responses. The main theorem in this chapter concerns the joint asymptotic normality of all maximum likelihood estimators for a Gaussian response mixed model and elegantly shows faster rates of convergence of fixed effects that are not accompanied by a random effect as compared to fixed effects that have partnering random effects.

The results presented in this chapter will then be extended for generalized linear mixed models with multivariate fixed and random effects in Chapter 4.

3.1 Model Description

In this section, we study Gaussian response linear mixed models of the following form, for observations of the random triples $(\boldsymbol{X}_{\mathrm{A}ij}, \boldsymbol{X}_{\mathrm{B}ij}, Y_{ij}), 1 \leq i \leq m, 1 \leq j \leq n_i$,

$$Y_{ij}|\boldsymbol{X}_{\mathrm{A}ij}, \boldsymbol{X}_{\mathrm{B}ij}, \boldsymbol{U}_i \text{ are independent } N((\boldsymbol{\beta}^0_{\mathrm{A}} + \boldsymbol{U}_i)^T \boldsymbol{X}_{\mathrm{A}ij} + (\boldsymbol{\beta}^0_{\mathrm{B}})^T \boldsymbol{X}_{\mathrm{B}ij}, \sigma^2_\varepsilon). \tag{3.1}$$

The $\boldsymbol{U}_i$ are $d_{\mathrm{A}} \times 1$ unobserved random vectors for each $1 \leq i \leq m$. The $\boldsymbol{X}_{\mathrm{A}ij}$ are $d_{\mathrm{A}} \times 1$ random vectors corresponding to predictors that are partnered by both a fixed effect and a random effect. The $\boldsymbol{X}_{\mathrm{B}ij}$ are $d_{\mathrm{B}} \times 1$ random vectors are predictors that have a fixed effect only. Let $\boldsymbol{X}_{ij} \equiv (\boldsymbol{X}^T_{\mathrm{A}ij}, \boldsymbol{X}^T_{\mathrm{B}ij})^T$ denote the combined predictor vectors such that $d_{\mathrm{A}} + d_{\mathrm{B}} = d$. We also assumed that the $\boldsymbol{X}_{ij}$ and $\boldsymbol{U}_i$, for $1 \leq i \leq m$ and $1 \leq j \leq n_i$ are independent, with the $\boldsymbol{X}_{ij}$ each having the same distribution as the $(d_{\mathrm{A}} + d_{\mathrm{B}}) \times 1$ random vector $\boldsymbol{X} = (\boldsymbol{X}^T_{\mathrm{A}}, \boldsymbol{X}^T_{\mathrm{B}})^T$ and the $\boldsymbol{U}_i$ are independent $N(0, (\boldsymbol{\Sigma})^0)$, each having the same distribution as the random vector $\boldsymbol{U}$.

Then, for any $\boldsymbol{\beta}_{\mathrm{A}}(d_{\mathrm{A}} \times 1)$, $\boldsymbol{\beta}_{\mathrm{B}}(d_{\mathrm{B}} \times 1)$, σ_ε^2 and $\boldsymbol{\Sigma}(d_{\mathrm{A}} \times d_{\mathrm{A}})$, the maximum likelihood estimator of $(\boldsymbol{\beta}_{\mathrm{A}}^0, \boldsymbol{\beta}_{\mathrm{B}}^0, \boldsymbol{\Sigma}^0, (\sigma^2)^0)$ is,

$$(\widehat{\boldsymbol{\beta}}_{\mathrm{A}}, \widehat{\boldsymbol{\beta}}_{\mathrm{B}}, \widehat{\boldsymbol{\Sigma}}, \widehat{\sigma}^2) = \underset{\boldsymbol{\beta}_{\mathrm{A}}, \boldsymbol{\beta}_{\mathrm{B}}, \boldsymbol{\Sigma}, \sigma_\varepsilon^2}{\operatorname{argmax}} \ell(\boldsymbol{\beta}_{\mathrm{A}}, \boldsymbol{\beta}_{\mathrm{B}}, \boldsymbol{\Sigma}, \sigma_\varepsilon^2)$$

where the conditional log-likelihood is

$$\begin{aligned}
&\ell(\boldsymbol{\beta}_{\mathrm{A}}, \boldsymbol{\beta}_{\mathrm{B}}, \boldsymbol{\Sigma}, \sigma_\varepsilon^2) \\
&= \sum_{i=1}^{m} \sum_{j=1}^{n} \left[\left\{ Y_{ij} \left(\boldsymbol{\beta}_{\mathrm{A}}^T \boldsymbol{X}_{\mathrm{A}ij} + \boldsymbol{\beta}_{\mathrm{B}}^T \boldsymbol{X}_{\mathrm{B}ij} \right) - \frac{1}{2} Y_{ij}^2 \right\} / \sigma_\varepsilon^2 - \frac{1}{2} \log(2\pi\sigma_\varepsilon^2) \right] - \frac{m}{2} \log |2\pi\boldsymbol{\Sigma}| \\
&\quad + \sum_{i=1}^{m} \log \int_{\mathbb{R}^{d_{\mathrm{A}}}} \exp \left(\sum_{j=1}^{n} \left[Y_{ij} \boldsymbol{u}^T \boldsymbol{X}_{\mathrm{A}ij} - \frac{1}{2} \{ (\boldsymbol{\beta}_{\mathrm{A}} + u)^T \boldsymbol{X}_{\mathrm{A}ij} + \boldsymbol{\beta}_{\mathrm{B}}^T \boldsymbol{X}_{\mathrm{B}ij} \} \right] / \sigma_\varepsilon^2 - \frac{1}{2} \boldsymbol{u}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{u} \right) d\boldsymbol{u}.
\end{aligned}$$

3.2 Notation Required for Fisher Information Calculations

Let

$$\boldsymbol{X}_i = \begin{bmatrix} \boldsymbol{X}_{i1}^T \\ \vdots \\ \boldsymbol{X}_{in_i}^T \end{bmatrix} \quad \text{and } \boldsymbol{Y}_i = \begin{bmatrix} Y_{i1} \\ \vdots \\ Y_{in} \end{bmatrix}.$$

Also define

$$n \equiv \frac{1}{m} \sum_{i=1}^{m} n_i = \text{ average of the within-group sample sizes,}$$

and

$$\boldsymbol{\Sigma}_{\boldsymbol{\beta}_{\mathrm{B}}} = \text{lower right } d_{\mathrm{B}} \times d_{\mathrm{B}} \text{ block of } E \left(\begin{bmatrix} \boldsymbol{X}_{\mathrm{A}} \boldsymbol{X}_{\mathrm{A}}^T & \boldsymbol{X}_{\mathrm{A}} \boldsymbol{X}_{\mathrm{B}}^T \\ \boldsymbol{X}_{\mathrm{B}} \boldsymbol{X}_{\mathrm{A}}^T & \boldsymbol{X}_{\mathrm{B}} \boldsymbol{X}_{\mathrm{B}}^T \end{bmatrix} \right)^{-1}.$$

3.3 Asymptotic Normality Theorem

The main theoretical contribution of this chapter is an asymptotic normality theorem for the maximum likelihood estimators for a Gaussian response mixed model as described in Section 3.1.

The theorem relies on the following assumptions:

(A1) The number of groups m diverges to ∞.

(A2) The within-group sample sizes n_i diverge to ∞ in such a way that $n_i/n \to C_i$ for constants $0 < C_i < \infty$, $1 \leq i \leq m$.

(A3) The distribution of $\boldsymbol{X}$ is such that

$$E(\|\boldsymbol{X}\|^8) < \infty$$

and none of the entries in $\boldsymbol{X}_{\mathrm{A}}$ are zero degenerate random variables.

Theorem 11. *Assume that conditions (A1) - (A3) hold. Then we have the following*

$$\sqrt{m}\begin{bmatrix} \widehat{\boldsymbol{\beta}}_A - \boldsymbol{\beta}_A^0 \\ \sqrt{n}\left(\widehat{\boldsymbol{\beta}}_B - \boldsymbol{\beta}_B^0\right) \\ \mathrm{vech}(\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}^0) \\ \sqrt{n}\left(\widehat{\sigma}^2 - (\sigma^2)^0\right) \end{bmatrix} \xrightarrow{\mathcal{D}} N\left(\begin{bmatrix} \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \\ 0 \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Sigma}^0 & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & (\sigma^2)^0\boldsymbol{\Sigma}_{\boldsymbol{\beta}_B} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & 2\boldsymbol{D}_{d_A}^{+}\left(\boldsymbol{\Sigma}^0 \otimes \boldsymbol{\Sigma}^0\right)\boldsymbol{D}_{d_A}^{+T} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & 2\{(\sigma^2)^0\}^2 \end{bmatrix}\right),$$

where $\boldsymbol{D}_{d_A}^{+}$ is the Moore-Penrose inverse of $\boldsymbol{D}_{d_A}$.

The proof for Theorem 11 is in the appendix.

3.4 Appendix

3.4.1 Linear Mixed Models with Multivariate Fixed and Random Effects

Let

$$\boldsymbol{X}_i = \begin{bmatrix} \boldsymbol{X}_{\mathrm{A}i} & \boldsymbol{X}_{\mathrm{B}i} \end{bmatrix}, \quad \boldsymbol{\beta} = \begin{bmatrix} \boldsymbol{\beta}_{\mathrm{A}} \\ \boldsymbol{\beta}_{\mathrm{B}} \end{bmatrix} \text{ and } \quad \boldsymbol{\theta} = \begin{bmatrix} \mathrm{vec}(\boldsymbol{\Sigma}) \\ \sigma_\varepsilon^2 \end{bmatrix}.$$

Then the model presented in Section 3.1 can be rewritten in the following form

$$\boldsymbol{y}_i \overset{\text{ind.}}{\sim} N(\boldsymbol{X}_i\boldsymbol{\beta}, \boldsymbol{V}_{\boldsymbol{\theta}i}), \tag{3.2}$$

with

$$\boldsymbol{V}_{\boldsymbol{\theta}i} = \boldsymbol{X}_{\mathrm{A}i}\boldsymbol{\Sigma}\boldsymbol{X}_{\mathrm{A}i}^T + \sigma_\varepsilon^2 I_{n_i},$$

where $\boldsymbol{V}_{\boldsymbol{\theta}i}$ is the $n_i \times n_i$ covariance matrix of $\boldsymbol{y}_i$ parametrized by $\boldsymbol{\theta}$. The log-likelihood of $(\boldsymbol{\beta}, \boldsymbol{\theta})$ can then be expressed as

$$\ell(\boldsymbol{\beta}, \boldsymbol{\theta}) = -\frac{1}{2}\sum_{i=1}^{m}\left\{\log|\boldsymbol{V}_{\boldsymbol{\theta}i}| + (\boldsymbol{y}_i - \boldsymbol{X}_i\boldsymbol{\beta})^T\boldsymbol{V}_{\boldsymbol{\theta}i}^{-1}(\boldsymbol{y}_i - \boldsymbol{X}_i\boldsymbol{\beta}) + n_i\log(2\pi)\right\}.$$

The model in (3.2) is a special case of a general Gaussian variance regression model presented in Section 4.3 of Wand (2002). Using steps similar to those in Wand (2002), an expression for the Fisher information matrix for a Gaussian response linear mixed model with multivariate fixed and random effects can be obtained. Note that the model in (3.2) belongs to a common class of submodels with

$$\boldsymbol{V}_{\boldsymbol{\theta}i} = \sum_{h=1}^{c} \theta_h \boldsymbol{K}_h, \quad \boldsymbol{\theta} = [\theta_1, \ldots, \theta_c], \tag{3.3}$$

for a set of $n_i \times n_i$ matrices $\boldsymbol{K}_1, \ldots, \boldsymbol{K}_c$, which leads to considerable simplifications in obtaining the expression for the Fisher information matrix. An alternative expression for equation (3.3) is as follows, where

$$\text{vec}(\boldsymbol{V}_{\boldsymbol{\theta}i}) = \boldsymbol{\mathcal{K}}_i \boldsymbol{\theta}, \quad \boldsymbol{\mathcal{K}} = [\text{vec}(\boldsymbol{K}_1)| \ldots |\text{vec}(\boldsymbol{K}_c)].$$

It remains to solve for $\boldsymbol{\mathcal{K}}_i$. By making use of the fourth property in Subsection 1.4.5, it can be done as follows

$$\begin{aligned}
\text{vec}(\boldsymbol{V}_{\boldsymbol{\theta}i}) &= \text{vec}(\boldsymbol{X}_{\text{A}i} \boldsymbol{\Sigma} \boldsymbol{X}_{\text{A}i}^T + \sigma_\varepsilon^2 I_n) \\
&= \text{vec}(\boldsymbol{X}_{\text{A}i} \boldsymbol{\Sigma} \boldsymbol{X}_{\text{A}i}^T) + \text{vec}(\sigma_\varepsilon^2 I_n) \\
&= (\boldsymbol{X}_{\text{A}i} \otimes \boldsymbol{X}_{\text{A}i}) \text{vec}(\boldsymbol{\Sigma}) + \text{vec}(\boldsymbol{I}_{n_i}) \sigma_\varepsilon^2 \\
&= (\boldsymbol{X}_{\text{A}i} \otimes \boldsymbol{X}_{\text{A}i}) \boldsymbol{D}_{d_\text{A}} \text{vech}(\boldsymbol{\Sigma}) + \text{vec}(\boldsymbol{I}_{n_i}) \sigma_\varepsilon^2 \\
&= \begin{bmatrix} (\boldsymbol{X}_{\text{A}i} \otimes \boldsymbol{X}_{\text{A}i}) \boldsymbol{D}_{d_\text{A}} & \text{vec}(\boldsymbol{I}_{n_i}) \end{bmatrix} \begin{bmatrix} \text{vech}(\boldsymbol{\Sigma}) \\ \sigma_\varepsilon^2 \end{bmatrix} \\
&= \boldsymbol{\mathcal{K}}_i \boldsymbol{\theta},
\end{aligned}$$

with

$$\boldsymbol{\mathcal{K}}_i = \begin{bmatrix} (\boldsymbol{X}_{\text{A}i} \otimes \boldsymbol{X}_{\text{A}i}) \boldsymbol{D}_{d_\text{A}} & \text{vec}(\boldsymbol{I}_{n_i}) \end{bmatrix}.$$

The expression for the full Fisher information matrix is subsequently

$$\begin{aligned}
I(\boldsymbol{\beta}_\text{A}, \boldsymbol{\beta}_\text{B}, \text{vech}(\boldsymbol{\Sigma}), \sigma_\varepsilon^2) &= \begin{bmatrix} I(\boldsymbol{\beta}_\text{A}, \boldsymbol{\beta}_\text{B}) & 0 \\ 0 & I(\text{vech}(\boldsymbol{\Sigma}), \sigma_\varepsilon^2) \end{bmatrix} \\
&= \sum_{i=1}^{m} \begin{bmatrix} \boldsymbol{X}_i^T \boldsymbol{V}_{\boldsymbol{\theta}i}^{-1} \boldsymbol{X}_i & 0 \\ 0 & \frac{1}{2} \boldsymbol{\mathcal{K}}_i^T (\boldsymbol{V}_{\boldsymbol{\theta}i}^{-1} \otimes \boldsymbol{V}_{\boldsymbol{\theta}i}^{-1}) \boldsymbol{\mathcal{K}}_i \end{bmatrix} \\
&= \begin{bmatrix} \sum_{i=1}^{m} \boldsymbol{X}_i^T \boldsymbol{V}_{\boldsymbol{\theta}i}^{-1} \boldsymbol{X}_i & 0 \\ 0 & \frac{1}{2} \sum_{i=1}^{m} \boldsymbol{\mathcal{K}}_i^T (\boldsymbol{V}_{\boldsymbol{\theta}i}^{-1} \otimes \boldsymbol{V}_{\boldsymbol{\theta}i}^{-1}) \boldsymbol{\mathcal{K}}_i \end{bmatrix}.
\end{aligned} \tag{3.4}$$

To find an explicit expression for $I(\boldsymbol{\beta}_\text{A}, \boldsymbol{\beta}_\text{B}, \text{vech}(\boldsymbol{\Sigma}), \sigma_\varepsilon^2)$, it suffices to find explicit expressions for the block matrices $I(\boldsymbol{\beta}_\text{A}, \boldsymbol{\beta}_\text{B})$ and $I(\text{vech}(\boldsymbol{\Sigma}), \sigma_\varepsilon^2)$ separately.

3.4.2 Expression for Top Left Block of Fisher Information Matrix

Expanding the current expression for $I(\boldsymbol{\beta}_{\mathrm{A}}, \boldsymbol{\beta}_{\mathrm{B}})$ leads to

$$\begin{aligned} I(\boldsymbol{\beta}_{\mathrm{A}}, \boldsymbol{\beta}_{\mathrm{B}}) &= \sum_{i=1}^{m} \boldsymbol{X}_i^T \boldsymbol{V}_{\boldsymbol{\theta} i}^{-1} \boldsymbol{X}_i \\ &= \sum_{i=1}^{m} \begin{bmatrix} \boldsymbol{X}_{\mathrm{A}i} & \boldsymbol{X}_{\mathrm{B}i} \end{bmatrix}^T (\boldsymbol{X}_{\mathrm{A}i} \boldsymbol{\Sigma} \boldsymbol{X}_{\mathrm{A}i}^T + \sigma_\varepsilon^2 \boldsymbol{I}_{n_i})^{-1} \begin{bmatrix} \boldsymbol{X}_{\mathrm{A}i} & \boldsymbol{X}_{\mathrm{B}i} \end{bmatrix} \\ &= \sum_{i=1}^{m} \begin{bmatrix} \boldsymbol{X}_{\mathrm{A}i}^T \\ \boldsymbol{X}_{\mathrm{B}i}^T \end{bmatrix} (\boldsymbol{X}_{\mathrm{A}i} \boldsymbol{\Sigma} \boldsymbol{X}_{\mathrm{A}i}^T + \sigma_\varepsilon^2 \boldsymbol{I}_{n_i})^{-1} \begin{bmatrix} \boldsymbol{X}_{\mathrm{A}i} & \boldsymbol{X}_{\mathrm{B}i} \end{bmatrix} \\ &= \sum_{i=1}^{m} \begin{bmatrix} \boldsymbol{X}_{\mathrm{A}i}^T (\boldsymbol{X}_{\mathrm{A}i} \boldsymbol{\Sigma} \boldsymbol{X}_{\mathrm{A}i}^T + \sigma_\varepsilon^2 \boldsymbol{I}_{n_i})^{-1} \boldsymbol{X}_{\mathrm{A}i} & \boldsymbol{X}_{\mathrm{A}i}^T (\boldsymbol{X}_{\mathrm{A}i} \boldsymbol{\Sigma} \boldsymbol{X}_{\mathrm{A}i}^T + \sigma_\varepsilon^2 \boldsymbol{I}_{n_i})^{-1} \boldsymbol{X}_{\mathrm{B}i} \\ \boldsymbol{X}_{\mathrm{B}i}^T (\boldsymbol{X}_{\mathrm{A}i} \boldsymbol{\Sigma} \boldsymbol{X}_{\mathrm{A}i}^T + \sigma_\varepsilon^2 \boldsymbol{I}_{n_i})^{-1} \boldsymbol{X}_{\mathrm{A}i} & \boldsymbol{X}_{\mathrm{B}i}^T (\boldsymbol{X}_{\mathrm{A}i} \boldsymbol{\Sigma} \boldsymbol{X}_{\mathrm{A}i}^T + \sigma_\varepsilon^2 \boldsymbol{I}_{n_i})^{-1} \boldsymbol{X}_{\mathrm{B}i} \end{bmatrix}. \end{aligned} \tag{3.5}$$

It remains to evaluate and simplify the expressions in the matrix, which can be done by making use of the matrix identities from Harville (1977), highlighted in Subsubsection 1.4.7.1.

3.4.2.1 Top Left Block of (3.5)

Now, we solve a part of the expression of the top left block of $I(\boldsymbol{\beta}_{\mathrm{A}}, \boldsymbol{\beta}_{\mathrm{B}})$, specifically

$$\boldsymbol{X}_{\mathrm{A}i}^T (\boldsymbol{X}_{\mathrm{A}i} \boldsymbol{\Sigma} \boldsymbol{X}_{\mathrm{A}i}^T + \sigma_\varepsilon^2 \boldsymbol{I}_{n_i})^{-1} \boldsymbol{X}_{\mathrm{A}i}.$$

Using (1.10b) and (1.3) by setting $\boldsymbol{A} = \frac{1}{\sigma_\varepsilon^2} \boldsymbol{X}_{\mathrm{A}i}^T \boldsymbol{X}_{\mathrm{A}i}$ and $\boldsymbol{B} = \boldsymbol{\Sigma}$, we have,

$$\begin{aligned} \boldsymbol{X}_{\mathrm{A}i}^T (\boldsymbol{X}_{\mathrm{A}i} \boldsymbol{\Sigma} \boldsymbol{X}_{\mathrm{A}i}^T + \sigma_\varepsilon^2 \boldsymbol{I}_{n_i})^{-1} \boldsymbol{X}_{\mathrm{A}i} &= \left\{ \boldsymbol{I}_{d_{\mathrm{A}}} + \boldsymbol{X}_{\mathrm{A}i}^T (\sigma_\varepsilon^2 \boldsymbol{I}_{n_i})^{-1} \boldsymbol{X}_{\mathrm{A}i} \boldsymbol{\Sigma} \right\}^{-1} \boldsymbol{X}_{\mathrm{A}i}^T (\sigma_\varepsilon^2 \boldsymbol{I}_{n_i})^{-1} \boldsymbol{X}_{\mathrm{A}i} \\ &= \left(\boldsymbol{I}_{d_{\mathrm{A}}} + \frac{1}{\sigma_\varepsilon^2} \boldsymbol{X}_{\mathrm{A}i}^T \boldsymbol{X}_{\mathrm{A}i} \boldsymbol{\Sigma} \right)^{-1} \left(\frac{1}{\sigma_\varepsilon^2} \boldsymbol{X}_{\mathrm{A}i}^T \boldsymbol{X}_{\mathrm{A}i} \right) \\ &= \left\{ \boldsymbol{I}_{d_{\mathrm{A}}} - \boldsymbol{\Sigma}^{-1} \left(\frac{1}{\sigma_\varepsilon^2} \boldsymbol{X}_{\mathrm{A}i}^T \boldsymbol{X}_{\mathrm{A}i} \right)^{-1} + \ldots \right\} \boldsymbol{\Sigma}^{-1} \\ &= \left\{ \boldsymbol{I}_{d_{\mathrm{A}}} - \sigma_\varepsilon^2 \boldsymbol{\Sigma}^{-1} \left(\boldsymbol{X}_{\mathrm{A}i}^T \boldsymbol{X}_{\mathrm{A}i} \right)^{-1} + \ldots \right\} \boldsymbol{\Sigma}^{-1} \\ &= \boldsymbol{\Sigma}^{-1} - \sigma_\varepsilon^2 \boldsymbol{\Sigma}^{-1} \left(\boldsymbol{X}_{\mathrm{A}i}^T \boldsymbol{X}_{\mathrm{A}i} \right)^{-1} \boldsymbol{\Sigma}^{-1} + \ldots \\ &= \boldsymbol{\Sigma}^{-1} + O_P(n^{-1}) \mathbf{1}_{d_{\mathrm{A}}} \mathbf{1}_{d_{\mathrm{A}}}^T. \end{aligned}$$

Therefore, the top left block of $I(\boldsymbol{\beta}_{\mathrm{A}}, \boldsymbol{\beta}_{\mathrm{B}})$ can be computed as

$$\sum_{i=1}^{m} \left(\boldsymbol{\Sigma}^{-1} + O_P(n^{-1}) \mathbf{1}_{d_{\mathrm{A}}} \mathbf{1}_{d_{\mathrm{A}}}^T \right) = m \boldsymbol{\Sigma}^{-1} + O_P(mn^{-1}) \mathbf{1}_{d_{\mathrm{A}}} \mathbf{1}_{d_{\mathrm{A}}}^T.$$

3.4.2.2 Top Right Block of (3.5)

Now, we solve a part of the expression of the top right block of $I(\boldsymbol{\beta}_{\mathrm{A}}, \boldsymbol{\beta}_{\mathrm{B}})$, specifically

$$\boldsymbol{X}_{\mathrm{A}i}^T (\boldsymbol{X}_{\mathrm{A}i} \boldsymbol{\Sigma} \boldsymbol{X}_{\mathrm{A}i}^T + \sigma_\varepsilon^2 \boldsymbol{I}_{n_i})^{-1} \boldsymbol{X}_{\mathrm{B}i}.$$

Using (1.10b) and (1.3) by setting $\boldsymbol{A} = \frac{1}{\sigma_\varepsilon^2}\boldsymbol{X}_{\mathrm{A}i}^T\boldsymbol{X}_{\mathrm{A}i}$, $\boldsymbol{B} = \boldsymbol{\Sigma}$ and $\boldsymbol{C} = \frac{1}{\sigma_\varepsilon^2}\boldsymbol{X}_{\mathrm{A}i}^T\boldsymbol{X}_{\mathrm{B}i}$ we have,

$$\begin{aligned}
\boldsymbol{X}_{\mathrm{A}i}^T(\boldsymbol{X}_{\mathrm{A}i}\boldsymbol{\Sigma}\boldsymbol{X}_{\mathrm{A}i}^T + \sigma_\varepsilon^2\boldsymbol{I}_{n_i})^{-1}\boldsymbol{X}_{\mathrm{B}i} &= \left\{\boldsymbol{I}_{d_\mathrm{A}} + \boldsymbol{X}_{\mathrm{A}i}^T(\sigma_\varepsilon^2\boldsymbol{I}_{n_i})^{-1}\boldsymbol{X}_{\mathrm{A}i}\boldsymbol{\Sigma}\right\}^{-1}\boldsymbol{X}_{\mathrm{A}i}^T(\sigma_\varepsilon^2\boldsymbol{I}_{n_i})^{-1}\boldsymbol{X}_{\mathrm{B}i} \\
&= \left(\boldsymbol{I}_{d_\mathrm{A}} + \frac{1}{\sigma_\varepsilon^2}\boldsymbol{X}_{\mathrm{A}i}^T\boldsymbol{X}_{\mathrm{A}i}\boldsymbol{\Sigma}\right)^{-1}\left(\frac{1}{\sigma_\varepsilon^2}\boldsymbol{X}_{\mathrm{A}i}^T\boldsymbol{X}_{\mathrm{B}i}\right) \\
&= \boldsymbol{\Sigma}^{-1}(\boldsymbol{X}_{\mathrm{A}i}^T\boldsymbol{X}_{\mathrm{A}i})^{-1}(\boldsymbol{X}_{\mathrm{A}i}^T\boldsymbol{X}_{\mathrm{B}i}) \\
&\quad - \sigma_\varepsilon^2\boldsymbol{\Sigma}^{-1}(\boldsymbol{X}_{\mathrm{A}i}^T\boldsymbol{X}_{\mathrm{A}i})^{-1}\boldsymbol{\Sigma}^{-1}(\boldsymbol{X}_{\mathrm{A}i}^T\boldsymbol{X}_{\mathrm{A}i})^{-1}(\boldsymbol{X}_{\mathrm{A}i}^T\boldsymbol{X}_{\mathrm{B}i}) + \ldots \\
&= \boldsymbol{\Sigma}^{-1}(\boldsymbol{X}_{\mathrm{A}i}^T\boldsymbol{X}_{\mathrm{A}i})^{-1}(\boldsymbol{X}_{\mathrm{A}i}^T\boldsymbol{X}_{\mathrm{B}i}) + O_P(n^{-1})\mathbf{1}_{d_\mathrm{A}}\mathbf{1}_{d_\mathrm{B}}^T.
\end{aligned}$$

Hence, the top right block of $I(\boldsymbol{\beta}_\mathrm{A}, \boldsymbol{\beta}_\mathrm{B})$ can be computed as

$$\begin{aligned}
&\sum_{i=1}^m \left\{\boldsymbol{\Sigma}^{-1}(\boldsymbol{X}_{\mathrm{A}i}^T\boldsymbol{X}_{\mathrm{A}i})^{-1}(\boldsymbol{X}_{\mathrm{A}i}^T\boldsymbol{X}_{\mathrm{B}i}) + O_P(n^{-1})\mathbf{1}_{d_\mathrm{A}}\mathbf{1}_{d_\mathrm{B}}^T\right\} \\
&= \sum_{i=1}^m \left\{\boldsymbol{\Sigma}^{-1}(\boldsymbol{X}_{\mathrm{A}i}^T\boldsymbol{X}_{\mathrm{A}i})^{-1}(\boldsymbol{X}_{\mathrm{A}i}^T\boldsymbol{X}_{\mathrm{B}i})\right\} + O_P(mn^{-1})\mathbf{1}_{d_\mathrm{A}}\mathbf{1}_{d_\mathrm{B}}^T. \\
&= O_P(m)\mathbf{1}_{d_\mathrm{A}}\mathbf{1}_{d_\mathrm{B}}^T.
\end{aligned}$$

3.4.2.3 Bottom Left Block of (3.5)

Subsequently, the expression for the bottom left block of $I(\boldsymbol{\beta}_\mathrm{A}, \boldsymbol{\beta}_\mathrm{B})$ can be easily computed as

$$\begin{aligned}
&\left[\sum_{i=1}^m \left\{\boldsymbol{\Sigma}^{-1}(\boldsymbol{X}_{\mathrm{A}i}^T\boldsymbol{X}_{\mathrm{A}i})^{-1}(\boldsymbol{X}_{\mathrm{A}i}^T\boldsymbol{X}_{\mathrm{B}i})\right\} + O_P(mn^{-1})\mathbf{1}_{d_\mathrm{A}}\mathbf{1}_{d_\mathrm{B}}^T\right]^T \\
&= \sum_{i=1}^m \left\{(\boldsymbol{X}_{\mathrm{B}i}^T\boldsymbol{X}_{\mathrm{A}i})(\boldsymbol{X}_{\mathrm{A}i}^T\boldsymbol{X}_{\mathrm{A}i})^{-1}\boldsymbol{\Sigma}^{-1}\right\} + O_P(mn^{-1})\mathbf{1}_{d_\mathrm{B}}\mathbf{1}_{d_\mathrm{A}}^T \\
&= O_P(m)\mathbf{1}_{d_\mathrm{B}}\mathbf{1}_{d_\mathrm{A}}^T.
\end{aligned}$$

3.4.2.4 Bottom Right Block of (3.5)

Now, we solve a part of the expression of the bottom right block of $I(\boldsymbol{\beta}_\mathrm{A}, \boldsymbol{\beta}_\mathrm{B})$, specifically

$$\boldsymbol{X}_{\mathrm{B}i}^T(\boldsymbol{X}_{\mathrm{A}i}\boldsymbol{\Sigma}\boldsymbol{X}_{\mathrm{A}i}^T + \sigma_\varepsilon^2\boldsymbol{I}_{n_i})^{-1}\boldsymbol{X}_{\mathrm{B}i}.$$

Using (1.10a) and (1.5) by setting $\boldsymbol{A} = \frac{1}{\sigma_\varepsilon^2}\boldsymbol{X}_{\mathrm{A}i}^T\boldsymbol{X}_{\mathrm{A}i}$ and $\boldsymbol{B} = \boldsymbol{\Sigma}$, we have,

$$\begin{aligned}
&\boldsymbol{X}_{\mathrm{B}i}^T\left\{\boldsymbol{X}_{\mathrm{A}i}\boldsymbol{\Sigma}\boldsymbol{X}_{\mathrm{A}i}^T + \sigma_\varepsilon^2\boldsymbol{I}_{n_i}\right\}^{-1}\boldsymbol{X}_{\mathrm{B}i} \\
&= \boldsymbol{X}_{\mathrm{B}i}^T\left[(\sigma_\varepsilon^2\boldsymbol{I}_{n_i})^{-1} - (\sigma_\varepsilon^2\boldsymbol{I}_{n_i})^{-1}\boldsymbol{X}_{\mathrm{A}i}\boldsymbol{\Sigma}\left\{\boldsymbol{I}_{d_\mathrm{A}} + \boldsymbol{X}_{\mathrm{A}i}^T(\sigma_\varepsilon^2\boldsymbol{I}_{n_i})^{-1}\boldsymbol{X}_{\mathrm{A}i}\boldsymbol{\Sigma}\right\}^{-1}\boldsymbol{X}_{\mathrm{A}i}^T(\sigma_\varepsilon^2\boldsymbol{I}_{n_i})^{-1}\right]\boldsymbol{X}_{\mathrm{B}i} \\
&= \boldsymbol{X}_{\mathrm{B}i}^T\left\{\frac{1}{\sigma_\varepsilon^2}\boldsymbol{I}_{n_i} - \left(\frac{1}{\sigma_\varepsilon^2}\right)^2\boldsymbol{X}_{\mathrm{A}i}\boldsymbol{\Sigma}\left(\boldsymbol{I}_{d_\mathrm{A}} + \frac{1}{\sigma_\varepsilon^2}\boldsymbol{X}_{\mathrm{A}i}^T\boldsymbol{X}_{\mathrm{A}i}\boldsymbol{\Sigma}\right)^{-1}\boldsymbol{X}_{\mathrm{A}i}^T\right\}\boldsymbol{X}_{\mathrm{B}i} \\
&= \frac{1}{\sigma_\varepsilon^2}\boldsymbol{X}_{\mathrm{B}i}^T\boldsymbol{X}_{\mathrm{B}i} - \frac{1}{\sigma_\varepsilon^2}\boldsymbol{X}_{\mathrm{B}i}^T\boldsymbol{X}_{\mathrm{A}i}\left(\boldsymbol{X}_{\mathrm{A}i}^T\boldsymbol{X}_{\mathrm{A}i}\right)^{-1}\boldsymbol{X}_{\mathrm{A}i}^T\boldsymbol{X}_{\mathrm{B}i} \\
&\quad + \boldsymbol{X}_{\mathrm{B}i}^T\boldsymbol{X}_{\mathrm{A}i}\left(\boldsymbol{X}_{\mathrm{A}i}^T\boldsymbol{X}_{\mathrm{A}i}\right)^{-1}\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{X}_{\mathrm{A}i}^T\boldsymbol{X}_{\mathrm{A}i}\right)^{-1}\boldsymbol{X}_{\mathrm{A}i}^T\boldsymbol{X}_{\mathrm{B}i} + \ldots \\
&= \frac{1}{\sigma_\varepsilon^2}\boldsymbol{X}_{\mathrm{B}i}^T\boldsymbol{X}_{\mathrm{B}i} - \frac{1}{\sigma_\varepsilon^2}\boldsymbol{X}_{\mathrm{B}i}^T\boldsymbol{X}_{\mathrm{A}i}\left(\boldsymbol{X}_{\mathrm{A}i}^T\boldsymbol{X}_{\mathrm{A}i}\right)^{-1}\boldsymbol{X}_{\mathrm{A}i}^T\boldsymbol{X}_{\mathrm{B}i} + O_P(1)\mathbf{1}_{d_\mathrm{B}}\mathbf{1}_{d_\mathrm{B}}^T.
\end{aligned}$$

Hence, the bottom right block of $I(\boldsymbol{\beta}_\mathrm{A}, \boldsymbol{\beta}_\mathrm{B})$ can be computed as

$$\begin{aligned}
&\sum_{i=1}^m\left\{\frac{1}{\sigma_\varepsilon^2}\boldsymbol{X}_{\mathrm{B}i}^T\boldsymbol{X}_{\mathrm{B}i} - \frac{1}{\sigma_\varepsilon^2}\boldsymbol{X}_{\mathrm{B}i}^T\boldsymbol{X}_{\mathrm{A}i}\left(\boldsymbol{X}_{\mathrm{A}i}^T\boldsymbol{X}_{\mathrm{A}i}\right)^{-1}\boldsymbol{X}_{\mathrm{A}i}^T\boldsymbol{X}_{\mathrm{B}i} + O_P(1)\mathbf{1}_{d_\mathrm{B}}\mathbf{1}_{d_\mathrm{B}}^T\right\} \\
&= \frac{1}{\sigma_\varepsilon^2}\sum_{i=1}^m\boldsymbol{X}_{\mathrm{B}i}^T\boldsymbol{X}_{\mathrm{B}i} - \frac{1}{\sigma_\varepsilon^2}\sum_{i=1}^m\boldsymbol{X}_{\mathrm{B}i}^T\boldsymbol{X}_{\mathrm{A}i}\left(\boldsymbol{X}_{\mathrm{A}i}^T\boldsymbol{X}_{\mathrm{A}i}\right)^{-1}\boldsymbol{X}_{\mathrm{A}i}^T\boldsymbol{X}_{\mathrm{B}i} + O_P(m)\mathbf{1}_{d_\mathrm{B}}\mathbf{1}_{d_\mathrm{B}}^T \\
&= \frac{1}{\sigma_\varepsilon^2}\sum_{i=1}^m\sum_{j=1}^n\boldsymbol{X}_{\mathrm{B}ij}\boldsymbol{X}_{\mathrm{B}ij}^T - \frac{1}{\sigma_\varepsilon^2}\sum_{i=1}^m\left(\sum_{j=1}^n\boldsymbol{X}_{\mathrm{B}ij}\boldsymbol{X}_{\mathrm{A}ij}^T\right)\left(\sum_{j=1}^n\boldsymbol{X}_{\mathrm{A}ij}\boldsymbol{X}_{\mathrm{A}ij}^T\right)^{-1}\left(\sum_{j=1}^n\boldsymbol{X}_{\mathrm{A}ij}\boldsymbol{X}_{\mathrm{B}ij}^T\right) \\
&\quad + O_P(m)\mathbf{1}_{d_\mathrm{B}}\mathbf{1}_{d_\mathrm{B}}^T.
\end{aligned}$$

The bottom right block of $I(\boldsymbol{\beta}_\mathrm{A}, \boldsymbol{\beta}_\mathrm{B})$ can then be re-expressed as follows

$$\begin{aligned}
&\frac{mn}{\sigma_\varepsilon^2}\left(\frac{1}{mn}\sum_{i=1}^m\sum_{j=1}^n\boldsymbol{X}_{\mathrm{B}ij}\boldsymbol{X}_{\mathrm{B}ij}^T\right) \\
&\quad - \frac{mn}{\sigma_\varepsilon^2}\left\{\frac{1}{mn}\sum_{i=1}^m\left(\sum_{j=1}^n\boldsymbol{X}_{\mathrm{B}ij}\boldsymbol{X}_{\mathrm{A}ij}^T\right)\left(\sum_{j=1}^n\boldsymbol{X}_{\mathrm{A}ij}\boldsymbol{X}_{\mathrm{A}ij}^T\right)^{-1}\left(\sum_{j=1}^n\boldsymbol{X}_{\mathrm{A}ij}\boldsymbol{X}_{\mathrm{B}ij}^T\right)\right\} \\
&\quad + O_P(m)\mathbf{1}_{d_\mathrm{B}}\mathbf{1}_{d_\mathrm{B}}^T.
\end{aligned}$$

Both of the leading terms in the previous expression are of order $O_P(mn)\mathbf{1}_{d_\mathrm{B}}\mathbf{1}_{d_\mathrm{B}}^T$ but it remains to determine the overall order of magnitude of the bottom right block of $I(\boldsymbol{\beta}_\mathrm{A}, \boldsymbol{\beta}_\mathrm{B})$. In order to do this, we shall treat both of the leading terms separately.

Treatment of First Leading Term

Note that for a Gaussian mixed model with multivariate fixed and random effects as described in Section 3.1,

$$b''\left((\boldsymbol{\beta}_\mathrm{A} + \boldsymbol{U}_i)^T\boldsymbol{X}_{\mathrm{A}ij} + \boldsymbol{\beta}_\mathrm{B}^T\boldsymbol{X}_{\mathrm{B}ij}\right) = 1, \quad 1 \le i \le m, \quad 1 \le j \le n_i.$$

Then from using Lemma 1, we have

$$\frac{1}{mn}\sum_{i=1}^m\sum_{j=1}^n\boldsymbol{X}_{\mathrm{B}ij}\boldsymbol{X}_{\mathrm{B}ij}^T = E(\boldsymbol{X}_\mathrm{B}\boldsymbol{X}_\mathrm{B}^T) + o_P(1)\mathbf{1}_{d_\mathrm{B}}\mathbf{1}_{d_\mathrm{B}}^T.$$

Finally we obtain

$$\frac{mn}{\sigma_\varepsilon^2}\left(\frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n}\boldsymbol{X}_{\mathrm{B}ij}\boldsymbol{X}_{\mathrm{B}ij}^T\right)=\frac{mn}{\sigma_\varepsilon^2}E(\boldsymbol{X}_{\mathrm{B}}\boldsymbol{X}_{\mathrm{B}}^T)+o_P(mn)\mathbf{1}_{d_{\mathrm{B}}}\mathbf{1}_{d_{\mathrm{B}}}^T. \tag{3.6}$$

Treatment of Second Leading Term

Note that for a Gaussian mixed model with multivariate fixed and random effects as described in Section 3.1,

$$f(\boldsymbol{X}_{ij},\boldsymbol{U}_i)=1,\quad 1\le i\le m,\quad 1\le j\le n_i,$$

where $f(\boldsymbol{X}_{ij},\boldsymbol{U}_i)$ is as defined in Lemma 2. Assuming that the conditions and assumptions in Lemma 2 are met, in the Gaussian mixed model case, we can simplify the left side of Lemma 2 as follows

$$\begin{aligned}
&E\Bigg[\frac{1}{mn}\sum_{i=1}^{m}\left\{\sum_{j=1}^{n}\boldsymbol{X}_{\mathrm{B}ij}\boldsymbol{X}_{\mathrm{A}ij}^T f(\boldsymbol{X}_{ij},\boldsymbol{U}_i)\right\}\left\{\sum_{j=1}^{n}\boldsymbol{X}_{\mathrm{A}ij}\boldsymbol{X}_{\mathrm{A}ij}^T f(\boldsymbol{X}_{ij},\boldsymbol{U}_i)\right\}^{-1}\times\\
&\left\{\sum_{j=1}^{n}\boldsymbol{X}_{\mathrm{B}ij}\boldsymbol{X}_{\mathrm{A}ij}^T f(\boldsymbol{X}_{ij},\boldsymbol{U}_i)\right\}^{T}\Bigg|\boldsymbol{X}_{11},\ldots,\boldsymbol{X}_{mn_m}\Bigg]\\
&=E\left\{\frac{1}{mn}\sum_{i=1}^{m}\left(\sum_{j=1}^{n}\boldsymbol{X}_{\mathrm{B}ij}\boldsymbol{X}_{\mathrm{A}ij}^T\right)\left(\sum_{j=1}^{n}\boldsymbol{X}_{\mathrm{A}ij}\boldsymbol{X}_{\mathrm{A}ij}^T\right)^{-1}\left(\sum_{j=1}^{n}\boldsymbol{X}_{\mathrm{B}ij}\boldsymbol{X}_{\mathrm{A}ij}^T\right)^{T}\Bigg|\boldsymbol{X}_{11},\ldots,\boldsymbol{X}_{mn_m}\right\}\\
&=\frac{1}{mn}\sum_{i=1}^{m}\left(\sum_{j=1}^{n}\boldsymbol{X}_{\mathrm{B}ij}\boldsymbol{X}_{\mathrm{A}ij}^T\right)\left(\sum_{j=1}^{n}\boldsymbol{X}_{\mathrm{A}ij}\boldsymbol{X}_{\mathrm{A}ij}^T\right)^{-1}\left(\sum_{j=1}^{n}\boldsymbol{X}_{\mathrm{A}ij}\boldsymbol{X}_{\mathrm{B}ij}^T\right).
\end{aligned}$$

We can also simplify the right hand side of Lemma 2 as follows

$$\begin{aligned}
&E\left[E\left(\boldsymbol{X}_{\mathrm{B}}\boldsymbol{X}_{\mathrm{A}}^T f(\boldsymbol{X},\boldsymbol{U})|\boldsymbol{U}\right)\left\{E\left(\boldsymbol{X}_{\mathrm{A}}\boldsymbol{X}_{\mathrm{A}}^T f(\boldsymbol{X},\boldsymbol{U})|\boldsymbol{U}\right)\right\}^{-1}E\left(\boldsymbol{X}_{\mathrm{B}}\boldsymbol{X}_{\mathrm{A}}^T f(\boldsymbol{X},\boldsymbol{U})|\boldsymbol{U}\right)^{T}\right]\\
&=E\left[E\left(\boldsymbol{X}_{\mathrm{B}}\boldsymbol{X}_{\mathrm{A}}^T\right)\left\{E\left(\boldsymbol{X}_{\mathrm{A}}\boldsymbol{X}_{\mathrm{A}}^T\right)\right\}^{-1}E\left(\boldsymbol{X}_{\mathrm{B}}\boldsymbol{X}_{\mathrm{A}}^T\right)^{T}\right]\\
&=E\left(\boldsymbol{X}_{\mathrm{B}}\boldsymbol{X}_{\mathrm{A}}^T\right)\left\{E\left(\boldsymbol{X}_{\mathrm{A}}\boldsymbol{X}_{\mathrm{A}}^T\right)\right\}^{-1}E\left(\boldsymbol{X}_{\mathrm{A}}\boldsymbol{X}_{\mathrm{B}}^T\right).
\end{aligned}$$

Now we have a simplified version of Lemma 2 for the Gaussian response linear mixed model which states

$$\begin{aligned}
&\frac{1}{mn}\sum_{i=1}^{m}\left(\sum_{j=1}^{n}\boldsymbol{X}_{\mathrm{B}ij}\boldsymbol{X}_{\mathrm{A}ij}^T\right)\left(\sum_{j=1}^{n}\boldsymbol{X}_{\mathrm{A}ij}\boldsymbol{X}_{\mathrm{A}ij}^T\right)^{-1}\left(\sum_{j=1}^{n}\boldsymbol{X}_{\mathrm{A}ij}\boldsymbol{X}_{\mathrm{B}ij}^T\right)\\
&\xrightarrow{P}E\left(\boldsymbol{X}_{\mathrm{B}}\boldsymbol{X}_{\mathrm{A}}^T\right)\left\{E\left(\boldsymbol{X}_{\mathrm{A}}\boldsymbol{X}_{\mathrm{A}}^T\right)\right\}^{-1}E\left(\boldsymbol{X}_{\mathrm{A}}\boldsymbol{X}_{\mathrm{B}}^T\right).
\end{aligned}\tag{3.7}$$

Equation (3.7) can also be written as

$$\begin{aligned}
&\frac{1}{mn}\sum_{i=1}^{m}\left(\sum_{j=1}^{n}\boldsymbol{X}_{\mathrm{B}ij}\boldsymbol{X}_{\mathrm{A}ij}^T\right)\left(\sum_{j=1}^{n}\boldsymbol{X}_{\mathrm{A}ij}\boldsymbol{X}_{\mathrm{A}ij}^T\right)^{-1}\left(\sum_{j=1}^{n}\boldsymbol{X}_{\mathrm{A}ij}\boldsymbol{X}_{\mathrm{B}ij}^T\right)\\
&=E\left(\boldsymbol{X}_{\mathrm{B}}\boldsymbol{X}_{\mathrm{A}}^T\right)\left\{E\left(\boldsymbol{X}_{\mathrm{A}}\boldsymbol{X}_{\mathrm{A}}^T\right)\right\}^{-1}E\left(\boldsymbol{X}_{\mathrm{A}}\boldsymbol{X}_{\mathrm{B}}^T\right)+o_P(1)\mathbf{1}_{d_{\mathrm{B}}}\mathbf{1}_{d_{\mathrm{B}}}^T.
\end{aligned}\tag{3.8}$$

Therefore we have,

$$\frac{mn}{\sigma_\varepsilon^2}\left\{\frac{1}{mn}\sum_{i=1}^{m}\left(\sum_{j=1}^{n}\boldsymbol{X}_{\mathrm{B}ij}\boldsymbol{X}_{\mathrm{A}ij}^T\right)\left(\sum_{j=1}^{n}\boldsymbol{X}_{\mathrm{A}ij}\boldsymbol{X}_{\mathrm{A}ij}^T\right)^{-1}\left(\sum_{j=1}^{n}\boldsymbol{X}_{\mathrm{A}ij}\boldsymbol{X}_{\mathrm{B}ij}^T\right)\right\}$$
$$=\frac{mn}{\sigma_\varepsilon^2}E\left(\boldsymbol{X}_{\mathrm{B}}\boldsymbol{X}_{\mathrm{A}}^T\right)\left\{E\left(\boldsymbol{X}_{\mathrm{A}}\boldsymbol{X}_{\mathrm{A}}^T\right)\right\}^{-1}E\left(\boldsymbol{X}_{\mathrm{A}}\boldsymbol{X}_{\mathrm{B}}^T\right)+o_P(mn)\mathbf{1}_{d_{\mathrm{B}}}\mathbf{1}_{d_{\mathrm{B}}}^T.$$

Hence, from equations (3.6) and (3.7) we have the following expression for the bottom right block of $I(\boldsymbol{\beta}_{\mathrm{A}},\boldsymbol{\beta}_{\mathrm{B}})$

$$\sum_{i=1}^{m}\boldsymbol{X}_{\mathrm{B}i}^T(\boldsymbol{X}_{\mathrm{A}i}\boldsymbol{\Sigma}\boldsymbol{X}_{\mathrm{A}i}^T+\sigma_\varepsilon^2\boldsymbol{I}_{n_i})^{-1}\boldsymbol{X}_{\mathrm{B}i}$$
$$=\frac{mn}{\sigma_\varepsilon^2}\left[E\left(\boldsymbol{X}_{\mathrm{B}}\boldsymbol{X}_{\mathrm{B}}^T\right)-E\left(\boldsymbol{X}_{\mathrm{B}}\boldsymbol{X}_{\mathrm{A}}^T\right)\left\{E\left(\boldsymbol{X}_{\mathrm{A}}\boldsymbol{X}_{\mathrm{A}}^T\right)\right\}^{-1}E\left(\boldsymbol{X}_{\mathrm{A}}\boldsymbol{X}_{\mathrm{B}}^T\right)\right]+o_P(mn)\mathbf{1}_{d_{\mathrm{B}}}\mathbf{1}_{d_{\mathrm{B}}}^T.$$

Consider the following

$$E\left(\begin{bmatrix}\boldsymbol{X}_{\mathrm{A}}\boldsymbol{X}_{\mathrm{A}}^T & \boldsymbol{X}_{\mathrm{A}}\boldsymbol{X}_{\mathrm{B}}^T\\ \boldsymbol{X}_{\mathrm{B}}\boldsymbol{X}_{\mathrm{A}}^T & \boldsymbol{X}_{\mathrm{B}}\boldsymbol{X}_{\mathrm{B}}^T\end{bmatrix}\right)^{-1}=\begin{bmatrix}E(\boldsymbol{X}_{\mathrm{A}}\boldsymbol{X}_{\mathrm{A}}^T) & E(\boldsymbol{X}_{\mathrm{A}}\boldsymbol{X}_{\mathrm{B}}^T)\\ E(\boldsymbol{X}_{\mathrm{B}}\boldsymbol{X}_{\mathrm{A}}^T) & E(\boldsymbol{X}_{\mathrm{B}}\boldsymbol{X}_{\mathrm{B}}^T)\end{bmatrix}^{-1}.$$

Then let

$$\boldsymbol{\Sigma}_{\boldsymbol{\beta}_{\mathrm{B}}}=\text{lower right } d_{\mathrm{B}}\times d_{\mathrm{B}} \text{ block of } E\left(\begin{bmatrix}\boldsymbol{X}_{\mathrm{A}}\boldsymbol{X}_{\mathrm{A}}^T & \boldsymbol{X}_{\mathrm{A}}\boldsymbol{X}_{\mathrm{B}}^T\\ \boldsymbol{X}_{\mathrm{B}}\boldsymbol{X}_{\mathrm{A}}^T & \boldsymbol{X}_{\mathrm{B}}\boldsymbol{X}_{\mathrm{B}}^T\end{bmatrix}\right)^{-1}$$
$$=\left[E\left(\boldsymbol{X}_{\mathrm{B}}\boldsymbol{X}_{\mathrm{B}}^T\right)-E\left(\boldsymbol{X}_{\mathrm{B}}\boldsymbol{X}_{\mathrm{A}}^T\right)\left\{E\left(\boldsymbol{X}_{\mathrm{A}}\boldsymbol{X}_{\mathrm{A}}^T\right)\right\}^{-1}E\left(\boldsymbol{X}_{\mathrm{A}}\boldsymbol{X}_{\mathrm{B}}^T\right)\right]^{-1}.$$

The expression for the bottom right block of $I(\boldsymbol{\beta}_{\mathrm{A}},\boldsymbol{\beta}_{\mathrm{B}})$ can then be re-defined as

$$\frac{mn\boldsymbol{\Sigma}_{\boldsymbol{\beta}_{\mathrm{B}}}^{-1}}{\sigma_\varepsilon^2}+o_P(mn)\mathbf{1}_{d_{\mathrm{B}}}\mathbf{1}_{d_{\mathrm{B}}}^T.$$

Putting the expressions obtained for the sub-blocks of $I(\boldsymbol{\beta}_{\mathrm{A}},\boldsymbol{\beta}_{\mathrm{B}})$ together, we have

$$I(\boldsymbol{\beta}_{\mathrm{A}},\boldsymbol{\beta}_{\mathrm{B}})=\begin{bmatrix}m\boldsymbol{\Sigma}^{-1}+O_P(mn^{-1})\mathbf{1}_{d_{\mathrm{A}}}\mathbf{1}_{d_{\mathrm{A}}}^T & O_P(m)\mathbf{1}_{d_{\mathrm{A}}}\mathbf{1}_{d_{\mathrm{B}}}^T\\ O_P(m)\mathbf{1}_{d_{\mathrm{B}}}\mathbf{1}_{d_{\mathrm{A}}}^T & \frac{mn\boldsymbol{\Sigma}_{\boldsymbol{\beta}_{\mathrm{B}}}^{-1}}{\sigma_\varepsilon^2}+o_P(mn)\mathbf{1}_{d_{\mathrm{B}}}\mathbf{1}_{d_{\mathrm{B}}}^T\end{bmatrix}. \tag{3.9}$$

3.4.3 Expression for Bottom Right Block of Fisher Information Matrix

Using similar steps to Section 3.4.2, an explicit expression for $I(\text{vech}(\boldsymbol{\Sigma}), \sigma_\varepsilon^2)$ can be obtained. Expanding the current expression for $I(\text{vech}(\boldsymbol{\Sigma}), \sigma_\varepsilon^2)$ leads to

$$
\begin{aligned}
&I(\text{vech}(\boldsymbol{\Sigma}), \sigma_\varepsilon^2) \\
&= \frac{1}{2}\sum_{i=1}^{m} \boldsymbol{\mathcal{K}}_i^T (\boldsymbol{V}_{\boldsymbol{\theta}i}^{-1} \otimes \boldsymbol{V}_{\boldsymbol{\theta}i}^{-1}) \boldsymbol{\mathcal{K}}_i \\
&= \frac{1}{2}\sum_{i=1}^{m} \begin{bmatrix} (\boldsymbol{X}_{\text{A}i} \otimes \boldsymbol{X}_{\text{A}i})\, \boldsymbol{D}_{d_\text{A}} & \text{vec}(\boldsymbol{I}_{n_i}) \end{bmatrix}^T \left\{ (\boldsymbol{X}_{\text{A}i} \boldsymbol{\Sigma} \boldsymbol{X}_{\text{A}i}^T + \sigma_\varepsilon^2 \boldsymbol{I}_{n_i})^{-1} \otimes (\boldsymbol{X}_{\text{A}i} \boldsymbol{\Sigma} \boldsymbol{X}_{\text{A}i}^T + \sigma_\varepsilon^2 \boldsymbol{I}_{n_i})^{-1} \right\} \times \\
&\qquad \begin{bmatrix} (\boldsymbol{X}_{\text{A}i} \otimes \boldsymbol{X}_{\text{A}i})\, \boldsymbol{D}_{d_\text{A}} & \text{vec}(\boldsymbol{I}_{n_i}) \end{bmatrix} \\
&= \begin{bmatrix} \frac{1}{2}\sum_{i=1}^{m} \boldsymbol{D}_{d_\text{A}}^T \left(\boldsymbol{X}_{\text{A}i}^T \otimes \boldsymbol{X}_{\text{A}i}^T \right) \left\{ (\boldsymbol{X}_{\text{A}i} \boldsymbol{\Sigma} \boldsymbol{X}_{\text{A}i}^T + \sigma_\varepsilon^2 \boldsymbol{I}_{n_i})^{-1} \otimes (\boldsymbol{X}_{\text{A}i} \boldsymbol{\Sigma} \boldsymbol{X}_{\text{A}i}^T + \sigma_\varepsilon^2 \boldsymbol{I}_{n_i})^{-1} \right\} (\boldsymbol{X}_{\text{A}i} \otimes \boldsymbol{X}_{\text{A}i})\, \boldsymbol{D}_{d_\text{A}} & \\
\quad \frac{1}{2}\sum_{i=1}^{m} \{\text{vec}(\boldsymbol{I}_{n_i})\}^T \left\{ (\boldsymbol{X}_{\text{A}i} \boldsymbol{\Sigma} \boldsymbol{X}_{\text{A}i}^T + \sigma_\varepsilon^2 \boldsymbol{I}_{n_i})^{-1} \otimes (\boldsymbol{X}_{\text{A}i} \boldsymbol{\Sigma} \boldsymbol{X}_{\text{A}i}^T + \sigma_\varepsilon^2 \boldsymbol{I}_{n_i})^{-1} \right\} (\boldsymbol{X}_{\text{A}i} \otimes \boldsymbol{X}_{\text{A}i})\, \boldsymbol{D}_{d_\text{A}} & \\
= & \\
\qquad \frac{1}{2}\sum_{i=1}^{m} \boldsymbol{D}_{d_\text{A}}^T \left(\boldsymbol{X}_{\text{A}i}^T \otimes \boldsymbol{X}_{\text{A}i}^T \right) \left\{ (\boldsymbol{X}_{\text{A}i} \boldsymbol{\Sigma} \boldsymbol{X}_{\text{A}i}^T + \sigma_\varepsilon^2 \boldsymbol{I}_{n_i})^{-1} \otimes (\boldsymbol{X}_{\text{A}i} \boldsymbol{\Sigma} \boldsymbol{X}_{\text{A}i}^T + \sigma_\varepsilon^2 \boldsymbol{I}_{n_i})^{-1} \right\} \text{vec}(\boldsymbol{I}_{n_i}) & \\
\quad \frac{1}{2}\sum_{i=1}^{m} \{\text{vec}(\boldsymbol{I}_{n_i})\}^T \left\{ (\boldsymbol{X}_{\text{A}i} \boldsymbol{\Sigma} \boldsymbol{X}_{\text{A}i}^T + \sigma_\varepsilon^2 \boldsymbol{I}_{n_i})^{-1} \otimes (\boldsymbol{X}_{\text{A}i} \boldsymbol{\Sigma} \boldsymbol{X}_{\text{A}i}^T + \sigma_\varepsilon^2 \boldsymbol{I}_{n_i})^{-1} \right\} \text{vec}(\boldsymbol{I}_{n_i}) & \end{bmatrix}.
\end{aligned}
\tag{3.10}
$$

Once again, it remains to evaluate and simplify the expressions in the matrix, which can be done by making use of the matrix identities from Harville (1977) highlighted in Section 1.4.7.1.

3.4.3.1 Top Left Block of (3.10)

Now, we solve a part of the expression of the top left block of $I(\text{vech}(\boldsymbol{\Sigma}), \sigma_\varepsilon^2)$, specifically

$$
\boldsymbol{D}_{d_\text{A}}^T \left(\boldsymbol{X}_{\text{A}i}^T \otimes \boldsymbol{X}_{\text{A}i}^T \right) \left\{ (\boldsymbol{X}_{\text{A}i} \boldsymbol{\Sigma} \boldsymbol{X}_{\text{A}i}^T + \sigma_\varepsilon^2 \boldsymbol{I}_{n_i})^{-1} \otimes (\boldsymbol{X}_{\text{A}i} \boldsymbol{\Sigma} \boldsymbol{X}_{\text{A}i}^T + \sigma_\varepsilon^2 \boldsymbol{I}_{n_i})^{-1} \right\} (\boldsymbol{X}_{\text{A}i} \otimes \boldsymbol{X}_{\text{A}i})\, \boldsymbol{D}_{d_\text{A}}.
$$

Using the properties of Kronecker products listed under Subsection 1.4.5 we have,

$$
\begin{aligned}
&\boldsymbol{D}_{d_\text{A}}^T \left(\boldsymbol{X}_{\text{A}i}^T \otimes \boldsymbol{X}_{\text{A}i}^T \right) \left\{ (\boldsymbol{X}_{\text{A}i} \boldsymbol{\Sigma} \boldsymbol{X}_{\text{A}i}^T + \sigma_\varepsilon^2 \boldsymbol{I}_{n_i})^{-1} \otimes (\boldsymbol{X}_{\text{A}i} \boldsymbol{\Sigma} \boldsymbol{X}_{\text{A}i}^T + \sigma_\varepsilon^2 \boldsymbol{I}_{n_i})^{-1} \right\} (\boldsymbol{X}_{\text{A}i} \otimes \boldsymbol{X}_{\text{A}i})\, \boldsymbol{D}_{d_\text{A}} \\
&= \boldsymbol{D}_{d_\text{A}}^T \left[\left\{ \boldsymbol{X}_{\text{A}i}^T (\boldsymbol{X}_{\text{A}i} \boldsymbol{\Sigma} \boldsymbol{X}_{\text{A}i}^T + \sigma_\varepsilon^2 \boldsymbol{I}_{n_i})^{-1} \boldsymbol{X}_{\text{A}i} \right\} \otimes \left\{ \boldsymbol{X}_{\text{A}i}^T (\boldsymbol{X}_{\text{A}i} \boldsymbol{\Sigma} \boldsymbol{X}_{\text{A}i}^T + \sigma_\varepsilon^2 \boldsymbol{I}_{n_i})^{-1} \boldsymbol{X}_{\text{A}i} \right\} \right] \boldsymbol{D}_{d_\text{A}} \\
&= \boldsymbol{D}_{d_\text{A}}^T \left[\{ \boldsymbol{\Sigma}^{-1} + O_P(n^{-1}) \boldsymbol{1}_{d_\text{A}} \boldsymbol{1}_{d_\text{A}}^T \} \otimes \{ \boldsymbol{\Sigma}^{-1} + O_P(n^{-1}) \boldsymbol{1}_{d_\text{A}} \boldsymbol{1}_{d_\text{A}}^T \} \right] \boldsymbol{D}_{d_\text{A}} \\
&= \boldsymbol{D}_{d_\text{A}}^T \left(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1} \right) \boldsymbol{D}_{d_\text{A}} + O_P(n^{-1}) \boldsymbol{1}_{d_\text{A}(d_\text{A}+1)/2} \boldsymbol{1}_{d_\text{A}(d_\text{A}+1)/2}^T.
\end{aligned}
$$

Therefore, the top left block of $I(\text{vech}(\boldsymbol{\Sigma}), \sigma_\varepsilon^2)$ can be computed as

$$
\begin{aligned}
&\frac{1}{2}\sum_{i=1}^{m} \left\{ \boldsymbol{D}_{d_\text{A}}^T \left(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1} \right) \boldsymbol{D}_{d_\text{A}} + O_P(n^{-1}) \boldsymbol{1}_{d_\text{A}(d_\text{A}+1)/2} \boldsymbol{1}_{d_\text{A}(d_\text{A}+1)/2}^T \right\} \\
&= \frac{m}{2} \left\{ \boldsymbol{D}_{d_\text{A}}^T \left(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1} \right) \boldsymbol{D}_{d_\text{A}} \right\} + O_P(mn^{-1}) \boldsymbol{1}_{d_\text{A}(d_\text{A}+1)/2} \boldsymbol{1}_{d_\text{A}(d_\text{A}+1)/2}^T.
\end{aligned}
$$

3.4.3.2 Top Right Block of (3.10)

Now, we solve a part of the expression of the top right block of $I(\text{vech}(\boldsymbol{\Sigma}), \sigma_\varepsilon^2)$, specifically

$$\boldsymbol{D}_{d_\text{A}}^T \left(\boldsymbol{X}_{\text{A}i}^T \otimes \boldsymbol{X}_{\text{A}i}^T\right) \left\{(\boldsymbol{X}_{\text{A}i}\boldsymbol{\Sigma}\boldsymbol{X}_{\text{A}i}^T + \sigma_\varepsilon^2 \boldsymbol{I}_{n_i})^{-1} \otimes (\boldsymbol{X}_{\text{A}i}\boldsymbol{\Sigma}\boldsymbol{X}_{\text{A}i}^T + \sigma_\varepsilon^2 \boldsymbol{I}_{n_i})^{-1}\right\} \text{vec}(\boldsymbol{I}_{n_i}).$$

It follows that

$$\begin{aligned}
&\boldsymbol{D}_{d_\text{A}}^T \left(\boldsymbol{X}_{\text{A}i}^T \otimes \boldsymbol{X}_{\text{A}i}^T\right) \left\{(\boldsymbol{X}_{\text{A}i}\boldsymbol{\Sigma}\boldsymbol{X}_{\text{A}i}^T + \sigma_\varepsilon^2 \boldsymbol{I}_{n_i})^{-1} \otimes (\boldsymbol{X}_{\text{A}i}\boldsymbol{\Sigma}\boldsymbol{X}_{\text{A}i}^T + \sigma_\varepsilon^2 \boldsymbol{I}_{n_i})^{-1}\right\} \text{vec}(\boldsymbol{I}_{n_i}) \\
&= \boldsymbol{D}_{d_\text{A}}^T \left\{\boldsymbol{X}_{\text{A}i}^T(\boldsymbol{X}_{\text{A}i}\boldsymbol{\Sigma}\boldsymbol{X}_{\text{A}i}^T + \sigma_\varepsilon^2 \boldsymbol{I}_{n_i})^{-1} \otimes \boldsymbol{X}_{\text{A}i}^T(\boldsymbol{X}_{\text{A}i}\boldsymbol{\Sigma}\boldsymbol{X}_{\text{A}i}^T + \sigma_\varepsilon^2 \boldsymbol{I}_{n_i})^{-1}\right\} \text{vec}(\boldsymbol{I}_{n_i}) \\
&= \boldsymbol{D}_{d_\text{A}}^T \text{vec}\left[\boldsymbol{X}_{\text{A}i}^T(\boldsymbol{X}_{\text{A}i}\boldsymbol{\Sigma}\boldsymbol{X}_{\text{A}i}^T + \sigma_\varepsilon^2 \boldsymbol{I}_{n_i})^{-1}\boldsymbol{I}_{n_i} \left\{\boldsymbol{X}_{\text{A}i}^T(\boldsymbol{X}_{\text{A}i}\boldsymbol{\Sigma}\boldsymbol{X}_{\text{A}i}^T + \sigma_\varepsilon^2 \boldsymbol{I}_{n_i})^{-1}\right\}^T\right] \\
&= \boldsymbol{D}_{d_\text{A}}^T \text{vec}\left[\boldsymbol{X}_{\text{A}i}^T(\boldsymbol{X}_{\text{A}i}\boldsymbol{\Sigma}\boldsymbol{X}_{\text{A}i}^T + \sigma_\varepsilon^2 \boldsymbol{I}_{n_i})^{-1} \left\{\boldsymbol{X}_{\text{A}i}^T(\boldsymbol{X}_{\text{A}i}\boldsymbol{\Sigma}\boldsymbol{X}_{\text{A}i}^T + \sigma_\varepsilon^2 \boldsymbol{I}_{n_i})^{-1}\right\}^T\right] \\
&= \boldsymbol{D}_{d_\text{A}}^T \text{vec}\Bigg[\left\{\sigma_\varepsilon^2\boldsymbol{\Sigma}^{-1}(\boldsymbol{X}_{\text{A}i}^T\boldsymbol{X}_{\text{A}i})^{-1} + \sigma_\varepsilon^2\boldsymbol{\Sigma}^{-1}(\boldsymbol{X}_{\text{A}i}^T\boldsymbol{X}_{\text{A}i})^{-1}\boldsymbol{\Sigma}^{-1}(\boldsymbol{X}_{\text{A}i}^T\boldsymbol{X}_{\text{A}i})^{-1} + \dots\right\}^{-1} \\
&\quad \times \left(\frac{1}{\sigma_\varepsilon^2}\boldsymbol{X}_{\text{A}i}^T\right)\left(\frac{1}{\sigma_\varepsilon^2}\boldsymbol{X}_{\text{A}i}\right)\left\{\sigma_\varepsilon^2\boldsymbol{\Sigma}^{-1}(\boldsymbol{X}_{\text{A}i}^T\boldsymbol{X}_{\text{A}i})^{-1} + \sigma_\varepsilon^2\boldsymbol{\Sigma}^{-1}(\boldsymbol{X}_{\text{A}i}^T\boldsymbol{X}_{\text{A}i})^{-1}\boldsymbol{\Sigma}^{-1}\right. \\
&\quad \left. \times (\boldsymbol{X}_{\text{A}i}^T\boldsymbol{X}_{\text{A}i})^{-1} + \dots\right\}^{-1}\Bigg].
\end{aligned}$$

By expanding and simplifying the last line of the previous expression, we have,

$$\begin{aligned}
&\boldsymbol{D}_{d_\text{A}}^T \text{vec}\Bigg[\sigma_\varepsilon^2\boldsymbol{\Sigma}^{-1}(\boldsymbol{X}_{\text{A}i}^T\boldsymbol{X}_{\text{A}i})^{-1}\left(\frac{1}{\sigma_\varepsilon^2}\boldsymbol{X}_{\text{A}i}^T\right)\left(\frac{1}{\sigma_\varepsilon^2}\boldsymbol{X}_{\text{A}i}\right)\sigma_\varepsilon^2(\boldsymbol{X}_{\text{A}i}^T\boldsymbol{X}_{\text{A}i})^{-1}\boldsymbol{\Sigma}^{-1} \\
&\quad - \sigma_\varepsilon^2\boldsymbol{\Sigma}^{-1}(\boldsymbol{X}_{\text{A}i}^T\boldsymbol{X}_{\text{A}i})^{-1}\boldsymbol{\Sigma}^{-1}(\boldsymbol{X}_{\text{A}i}^T\boldsymbol{X}_{\text{A}i})^{-1}\left(\frac{1}{\sigma_\varepsilon^2}\boldsymbol{X}_{\text{A}i}^T\right)\left(\frac{1}{\sigma_\varepsilon^2}\boldsymbol{X}_{\text{A}i}\right)\sigma_\varepsilon^2(\boldsymbol{X}_{\text{A}i}^T\boldsymbol{X}_{\text{A}i})^{-1}\boldsymbol{\Sigma}^{-1} \\
&\quad - \sigma_\varepsilon^2\boldsymbol{\Sigma}^{-1}(\boldsymbol{X}_{\text{A}i}^T\boldsymbol{X}_{\text{A}i})^{-1}\left(\frac{1}{\sigma_\varepsilon^2}\boldsymbol{X}_{\text{A}i}^T\right)\left(\frac{1}{\sigma_\varepsilon^2}\boldsymbol{X}_{\text{A}i}\right)\sigma_\varepsilon^2(\boldsymbol{X}_{\text{A}i}^T\boldsymbol{X}_{\text{A}i})^{-1}\boldsymbol{\Sigma}^{-1}(\boldsymbol{X}_{\text{A}i}^T\boldsymbol{X}_{\text{A}i})^{-1}\boldsymbol{\Sigma}^{-1} \\
&\quad - \sigma_\varepsilon^2\boldsymbol{\Sigma}^{-1}(\boldsymbol{X}_{\text{A}i}^T\boldsymbol{X}_{\text{A}i})^{-1}\boldsymbol{\Sigma}^{-1}(\boldsymbol{X}_{\text{A}i}^T\boldsymbol{X}_{\text{A}i})^{-1}\left(\frac{1}{\sigma_\varepsilon^2}\boldsymbol{X}_{\text{A}i}^T\right)\left(\frac{1}{\sigma_\varepsilon^2}\boldsymbol{X}_{\text{A}i}\right)\sigma_\varepsilon^2(\boldsymbol{X}_{\text{A}i}^T\boldsymbol{X}_{\text{A}i})^{-1}\boldsymbol{\Sigma}^{-1} \\
&\quad \times (\boldsymbol{X}_{\text{A}i}^T\boldsymbol{X}_{\text{A}i})^{-1}\boldsymbol{\Sigma}^{-1} + \dots\Bigg] \\
&= \boldsymbol{D}_{d_\text{A}}^T \text{vec}\{\boldsymbol{\Sigma}^{-1}(\boldsymbol{X}_{\text{A}i}^T\boldsymbol{X}_{\text{A}i})^{-1}\boldsymbol{\Sigma}^{-1} - \sigma_\varepsilon^2\boldsymbol{\Sigma}^{-1}(\boldsymbol{X}_{\text{A}i}^T\boldsymbol{X}_{\text{A}i})^{-1}\boldsymbol{\Sigma}^{-1}(\boldsymbol{X}_{\text{A}i}^T\boldsymbol{X}_{\text{A}i})^{-1}\boldsymbol{\Sigma}^{-1} \\
&\quad - \sigma_\varepsilon^2\boldsymbol{\Sigma}^{-1}(\boldsymbol{X}_{\text{A}i}^T\boldsymbol{X}_{\text{A}i})^{-1}\boldsymbol{\Sigma}^{-1}(\boldsymbol{X}_{\text{A}i}^T\boldsymbol{X}_{\text{A}i})^{-1}\boldsymbol{\Sigma}^{-1} \\
&\quad + (\sigma_\varepsilon^2)^2\boldsymbol{\Sigma}^{-1}(\boldsymbol{X}_{\text{A}i}^T\boldsymbol{X}_{\text{A}i})^{-1}\boldsymbol{\Sigma}^{-1}(\boldsymbol{X}_{\text{A}i}^T\boldsymbol{X}_{\text{A}i})^{-1}\boldsymbol{\Sigma}^{-1}(\boldsymbol{X}_{\text{A}i}^T\boldsymbol{X}_{\text{A}i})^{-1}\boldsymbol{\Sigma}^{-1} + \dots\} \\
&= \boldsymbol{D}_{d_\text{A}}^T \left[\text{vec}\left\{\boldsymbol{\Sigma}^{-1}(\boldsymbol{X}_{\text{A}i}^T\boldsymbol{X}_{\text{A}i})^{-1}\boldsymbol{\Sigma}^{-1}\right\} - 2\text{vec}\left\{\sigma_\varepsilon^2\boldsymbol{\Sigma}^{-1}(\boldsymbol{X}_{\text{A}i}^T\boldsymbol{X}_{\text{A}i})^{-1}\boldsymbol{\Sigma}^{-1}(\boldsymbol{X}_{\text{A}i}^T\boldsymbol{X}_{\text{A}i})^{-1}\boldsymbol{\Sigma}^{-1}\right\} + \dots\right] \\
&= \boldsymbol{D}_{d_\text{A}}^T \text{vec}\left\{\boldsymbol{\Sigma}^{-1}(\boldsymbol{X}_{\text{A}i}^T\boldsymbol{X}_{\text{A}i})^{-1}\boldsymbol{\Sigma}^{-1}\right\} + O_P(n^{-2})\mathbf{1}_{d_\text{A}(d_\text{A}+1)/2} \\
&= O_P(n^{-1})\mathbf{1}_{d_\text{A}(d_\text{A}+1)/2}.
\end{aligned}$$

Therefore, the top right block of $I(\text{vech}(\boldsymbol{\Sigma}), \sigma_\varepsilon^2)$ can be computed as

$$\frac{1}{2}\sum_{i=1}^{m} \boldsymbol{D}_{d_\text{A}}^T \text{vec}\{\boldsymbol{\Sigma}^{-1}(\boldsymbol{X}_{\text{A}i}^T\boldsymbol{X}_{\text{A}i})^{-1}\boldsymbol{\Sigma}^{-1}\} + O_P(mn^{-2})\mathbf{1}_{d_\text{A}(d_\text{A}+1)/2} = O_P(mn^{-1})\mathbf{1}_{d_\text{A}(d_\text{A}+1)/2}.$$

3.4.3.3 Bottom Left Block of (3.10)

Now, we solve a part of the expression of the bottom left block of $I(\text{vech}(\boldsymbol{\Sigma}), \sigma_\varepsilon^2)$, specifically

$$\{\text{vec}(\boldsymbol{I}_{n_i})\}^T \left\{(\boldsymbol{X}_{\text{A}i}\boldsymbol{\Sigma}\boldsymbol{X}_{\text{A}i}^T + \sigma_\varepsilon^2\boldsymbol{I}_{n_i})^{-1} \otimes (\boldsymbol{X}_{\text{A}i}\boldsymbol{\Sigma}\boldsymbol{X}_{\text{A}i}^T + \sigma_\varepsilon^2\boldsymbol{I}_{n_i})^{-1}\right\} (\boldsymbol{X}_{\text{A}i} \otimes \boldsymbol{X}_{\text{A}i})\, \boldsymbol{D}_{d_\text{A}}.$$

Using the properties for Kronecker products highlighted in Subsection 1.4.5, we have,

$$\begin{aligned}
&\{\text{vec}(\boldsymbol{I}_{n_i})\}^T \left\{(\boldsymbol{X}_{\text{A}i}\boldsymbol{\Sigma}\boldsymbol{X}_{\text{A}i}^T + \sigma_\varepsilon^2\boldsymbol{I}_{n_i})^{-1} \otimes (\boldsymbol{X}_{\text{A}i}\boldsymbol{\Sigma}\boldsymbol{X}_{\text{A}i}^T + \sigma_\varepsilon^2\boldsymbol{I}_{n_i})^{-1}\right\} (\boldsymbol{X}_{\text{A}i} \otimes \boldsymbol{X}_{\text{A}i})\, \boldsymbol{D}_{d_\text{A}} \\
&= \left[\boldsymbol{D}_{d_\text{A}}^T \left(\boldsymbol{X}_{\text{A}i}^T \otimes \boldsymbol{X}_{\text{A}i}^T\right) \left\{(\boldsymbol{X}_{\text{A}i}\boldsymbol{\Sigma}\boldsymbol{X}_{\text{A}i}^T + \sigma_\varepsilon^2\boldsymbol{I}_{n_i})^{-1} \otimes (\boldsymbol{X}_{\text{A}i}\boldsymbol{\Sigma}\boldsymbol{X}_{\text{A}i}^T + \sigma_\varepsilon^2\boldsymbol{I}_{n_i})^{-1}\right\} \text{vec}(\boldsymbol{I}_{n_i})\right]^T \\
&= \left[\boldsymbol{D}_{d_\text{A}}^T \text{vec}\{\boldsymbol{\Sigma}^{-1}(\boldsymbol{X}_{\text{A}i}^T\boldsymbol{X}_{\text{A}i})^{-1}\boldsymbol{\Sigma}^{-1}\} + O_P(n^{-2})\mathbf{1}_{d_\text{A}(d_\text{A}+1)/2}\right]^T \\
&= \left[\boldsymbol{D}_{d_\text{A}}^T \text{vec}\{\boldsymbol{\Sigma}^{-1}(\boldsymbol{X}_{\text{A}i}^T\boldsymbol{X}_{\text{A}i})^{-1}\boldsymbol{\Sigma}^{-1}\}\right]^T + O_P(n^{-2})\mathbf{1}_{d_\text{A}(d_\text{A}+1)/2}^T \\
&= O_P(n^{-1})\mathbf{1}_{d_\text{A}(d_\text{A}+1)/2}^T.
\end{aligned}$$

Therefore, the bottom left block of $I(\text{vech}(\boldsymbol{\Sigma}), \sigma_\varepsilon^2)$ can be computed as

$$\frac{1}{2}\sum_{i=1}^m \left[\boldsymbol{D}_{d_\text{A}}^T \text{vec}\{\boldsymbol{\Sigma}^{-1}(\boldsymbol{X}_{\text{A}i}^T\boldsymbol{X}_{\text{A}i})^{-1}\boldsymbol{\Sigma}^{-1}\}\right]^T + O_P(mn^{-2})\mathbf{1}_{d_\text{A}(d_\text{A}+1)/2}^T = O_P(mn^{-1})\mathbf{1}_{d_\text{A}(d_\text{A}+1)/2}^T.$$

3.4.3.4 Bottom Right Block of (3.10)

Now, we solve a part of the expression of the bottom right block of $I(\text{vech}(\boldsymbol{\Sigma}), \sigma_\varepsilon^2)$, specifically

$$\{\text{vec}(\boldsymbol{I}_{n_i})\}^T \left\{(\boldsymbol{X}_{\text{A}i}\boldsymbol{\Sigma}\boldsymbol{X}_{\text{A}i}^T + \sigma_\varepsilon^2\boldsymbol{I}_{n_i})^{-1} \otimes (\boldsymbol{X}_{\text{A}i}\boldsymbol{\Sigma}\boldsymbol{X}_{\text{A}i}^T + \sigma_\varepsilon^2\boldsymbol{I}_{n_i})^{-1}\right\} \{\text{vec}(\boldsymbol{I}_{n_i})\}.$$

Using (1.10a) and (1.5), we have

$$\begin{aligned}
&\{\text{vec}(\boldsymbol{I}_{n_i})\}^T \left\{(\boldsymbol{X}_{\text{A}i}\boldsymbol{\Sigma}\boldsymbol{X}_{\text{A}i}^T + \sigma_\varepsilon^2\boldsymbol{I}_{n_i})^{-1} \otimes (\boldsymbol{X}_{\text{A}i}\boldsymbol{\Sigma}\boldsymbol{X}_{\text{A}i}^T + \sigma_\varepsilon^2\boldsymbol{I}_{n_i})^{-1}\right\} \{\text{vec}(\boldsymbol{I}_{n_i})\} \\
&= \text{tr}\left\{(\boldsymbol{X}_{\text{A}i}\boldsymbol{\Sigma}\boldsymbol{X}_{\text{A}i}^T + \sigma_\varepsilon^2\boldsymbol{I}_{n_i})^{-1}(\boldsymbol{X}_{\text{A}i}\boldsymbol{\Sigma}\boldsymbol{X}_{\text{A}i}^T + \sigma_\varepsilon^2\boldsymbol{I}_{n_i})^{-1}\right\} \\
&= \text{tr}\left[\left\{\frac{1}{\sigma_\varepsilon^2}\boldsymbol{I}_{n_i} - \frac{1}{\sigma_\varepsilon^2}\boldsymbol{X}_{\text{A}i}(\boldsymbol{X}_{\text{A}i}^T\boldsymbol{X}_{\text{A}i})^{-1}\boldsymbol{X}_{\text{A}i}^T + \ldots\right\}\left\{\frac{1}{\sigma_\varepsilon^2}\boldsymbol{I}_{n_i} - \frac{1}{\sigma_\varepsilon^2}\boldsymbol{X}_{\text{A}i}(\boldsymbol{X}_{\text{A}i}^T\boldsymbol{X}_{\text{A}i})^{-1}\boldsymbol{X}_{\text{A}i}^T + \ldots\right\}\right] \\
&= \text{tr}\Bigg\{\left(\frac{1}{\sigma_\varepsilon^2}\right)^2 \boldsymbol{I}_{n_i} - 2\left(\frac{1}{\sigma_\varepsilon^2}\right)^2 \boldsymbol{X}_{\text{A}i}(\boldsymbol{X}_{\text{A}i}^T\boldsymbol{X}_{\text{A}i})^{-1}\boldsymbol{X}_{\text{A}i}^T \\
&\quad + \left(\frac{1}{\sigma_\varepsilon^2}\right)^2 \boldsymbol{X}_{\text{A}i}(\boldsymbol{X}_{\text{A}i}^T\boldsymbol{X}_{\text{A}i})^{-1}\boldsymbol{X}_{\text{A}i}^T\boldsymbol{X}_{\text{A}i}(\boldsymbol{X}_{\text{A}i}^T\boldsymbol{X}_{\text{A}i})^{-1}\boldsymbol{X}_{\text{A}i}^T\Bigg\} + \ldots \\
&= \left(\frac{1}{\sigma_\varepsilon^2}\right)^2 \text{tr}(\boldsymbol{I}_{n_i}) - \left(\frac{1}{\sigma_\varepsilon^2}\right)^2 \text{tr}\left\{\boldsymbol{X}_{\text{A}i}(\boldsymbol{X}_{\text{A}i}^T\boldsymbol{X}_{\text{A}i})^{-1}\boldsymbol{X}_{\text{A}i}^T\right\} + \ldots \\
&= \frac{n}{(\sigma_\varepsilon^2)^2} + O_P(1).
\end{aligned}$$

Therefore, the bottom right block of $I(\text{vech}(\boldsymbol{\Sigma}), \sigma_\varepsilon^2)$ can be computed as

$$\frac{1}{2}\sum_{i=1}^m \left\{\frac{n}{(\sigma_\varepsilon)^2} + O_P(1)\right\} = \frac{mn}{2(\sigma_\varepsilon)^2} + O_P(m).$$

Putting the expressions obtained for the sub-blocks of $I(\text{vech}(\boldsymbol{\Sigma}), \sigma_\varepsilon^2)$ together, we have

$$I(\text{vech}(\boldsymbol{\Sigma}), \sigma_\varepsilon^2)$$
$$= \begin{bmatrix} \frac{m}{2}\left\{\boldsymbol{D}_{d_\text{A}}^T\left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right)\boldsymbol{D}_{d_\text{A}}\right\} + O_P(mn^{-1})\mathbf{1}_{d_\text{A}(d_\text{A}+1)/2}\mathbf{1}_{d_\text{A}(d_\text{A}+1)/2}^T & O_P(mn^{-1})\mathbf{1}_{d_\text{A}(d_\text{A}+1)/2} \\ O_P(mn^{-1})\mathbf{1}_{d_\text{A}(d_\text{A}+1)/2}^T & \frac{mn}{2(\sigma_\varepsilon^2)^2} + O_P(m) \end{bmatrix}. \tag{3.11}$$

3.4.4 The Inverse of the Fisher Information Matrix

Since $I(\boldsymbol{\beta}_\text{A}, \boldsymbol{\beta}_\text{B}, \text{vech}(\boldsymbol{\Sigma}), \sigma_\varepsilon^2)$ is a block diagonal matrix, we can invert $I(\boldsymbol{\beta}_\text{A}, \boldsymbol{\beta}_\text{B})$ and $I(\text{vech}(\boldsymbol{\Sigma}), \sigma_\varepsilon^2)$ separately and put the expressions together to find $I(\boldsymbol{\beta}_\text{A}, \boldsymbol{\beta}_\text{B}, \text{vech}(\boldsymbol{\Sigma}), \sigma_\varepsilon^2)^{-1}$ as shown below.

$$I(\boldsymbol{\beta}_\text{A}, \boldsymbol{\beta}_\text{B}, \text{vech}(\boldsymbol{\Sigma}), \sigma_\varepsilon^2)^{-1} = \begin{bmatrix} I(\boldsymbol{\beta}_\text{A}, \boldsymbol{\beta}_\text{B})^{-1} & 0 \\ 0 & I(\text{vech}(\boldsymbol{\Sigma}), \sigma_\varepsilon^2)^{-1} \end{bmatrix}$$

3.4.4.1 Expression for Top Left Block of Inverse Fisher Information Matrix

Firstly, let us partition $I(\boldsymbol{\beta}_\text{A}, \boldsymbol{\beta}_\text{B})$ as follows

$$I(\boldsymbol{\beta}_\text{A}, \boldsymbol{\beta}_\text{B}) = \begin{bmatrix} \boldsymbol{A}_{11} & \boldsymbol{A}_{12} \\ \boldsymbol{A}_{21} & \boldsymbol{A}_{22} \end{bmatrix} \text{ where } \boldsymbol{A}_{21} = \boldsymbol{A}_{12}^T.$$

The expressions for $\boldsymbol{A}_{11}$, $\boldsymbol{A}_{12}$, $\boldsymbol{A}_{21}$ and $\boldsymbol{A}_{22}$ are currently as follows

$$\begin{aligned} \boldsymbol{A}_{11} &= m\boldsymbol{\Sigma}^{-1} + O_P(mn^{-1})\mathbf{1}_{d_\text{A}}\mathbf{1}_{d_\text{A}}^T, \\ \boldsymbol{A}_{12} &= O_P(m)\mathbf{1}_{d_\text{A}}\mathbf{1}_{d_\text{B}}^T, \\ \boldsymbol{A}_{21} &= O_P(m)\mathbf{1}_{d_\text{B}}\mathbf{1}_{d_\text{A}}^T, \\ \boldsymbol{A}_{22} &= \frac{mn\boldsymbol{\Sigma}_{\boldsymbol{\beta}_\text{B}}^{-1}}{\sigma_\varepsilon^2} + o_P(mn)\mathbf{1}_{d_\text{B}}\mathbf{1}_{d_\text{B}}^T. \end{aligned}$$

Let $I(\boldsymbol{\beta}_\text{A}, \boldsymbol{\beta}_\text{B})^{-1}$ assume the following form

$$\begin{aligned} I(\boldsymbol{\beta}_\text{A}, \boldsymbol{\beta}_\text{B})^{-1} &= \begin{bmatrix} \boldsymbol{A}_{11} & \boldsymbol{A}_{12} \\ \boldsymbol{A}_{21} & \boldsymbol{A}_{22} \end{bmatrix}^{-1} \\ &= \begin{bmatrix} \boldsymbol{A}^{11} & \boldsymbol{A}^{12} \\ \boldsymbol{A}^{21} & \boldsymbol{A}^{22} \end{bmatrix} \quad \text{where } \boldsymbol{A}^{21} = (\boldsymbol{A}^{12})^T. \end{aligned}$$

Firstly note that

$$\begin{aligned} \boldsymbol{A}_{11}^{-1} &= \left\{m\boldsymbol{\Sigma}^{-1} + O_P(mn^{-1})\mathbf{1}_{d_\text{A}}\mathbf{1}_{d_\text{A}}^T\right\}^{-1} \\ &= \left[m\boldsymbol{\Sigma}^{-1}\left\{\boldsymbol{I}_{d_\text{A}} + O_P(n^{-1})\mathbf{1}_{d_\text{A}}\mathbf{1}_{d_\text{A}}^T\right\}\right]^{-1} \\ &= m\boldsymbol{\Sigma}^{-1}\left\{\boldsymbol{I}_{d_\text{A}} - O_P(n^{-1})\mathbf{1}_{d_\text{A}}\mathbf{1}_{d_\text{A}}^T + \dots\right\} \\ &= m^{-1}\boldsymbol{\Sigma} + O_P(m^{-1}n^{-1})\mathbf{1}_{d_\text{A}}\mathbf{1}_{d_\text{A}}^T. \end{aligned} \tag{3.12}$$

Also note that

$$\begin{aligned}
\boldsymbol{A}_{22}^{-1} &= \left\{ \frac{mn\boldsymbol{\Sigma}_{\boldsymbol{\beta}_{\mathrm{B}}}^{-1}}{\sigma_{\varepsilon}^{2}} + o_P(mn)\mathbf{1}_{d_{\mathrm{B}}}\mathbf{1}_{d_{\mathrm{B}}}^{T} \right\}^{-1} \\
&= \left[\frac{mn\boldsymbol{\Sigma}_{\boldsymbol{\beta}_{\mathrm{B}}}^{-1}}{\sigma_{\varepsilon}^{2}} \left\{ \boldsymbol{I}_{d_{\mathrm{B}}} + o_P(1)\boldsymbol{\Sigma}_{\boldsymbol{\beta}_{\mathrm{B}}}\mathbf{1}_{d_{\mathrm{B}}}\mathbf{1}_{d_{\mathrm{B}}}^{T} \right\} \right]^{-1} \\
&= \frac{\sigma_{\varepsilon}^{2}\boldsymbol{\Sigma}_{\boldsymbol{\beta}_{\mathrm{B}}}}{mn} \left\{ \boldsymbol{I}_{d_{\mathrm{B}}} + o_P(1)\boldsymbol{\Sigma}_{\boldsymbol{\beta}_{\mathrm{B}}}\mathbf{1}_{d_{\mathrm{B}}}\mathbf{1}_{d_{\mathrm{B}}}^{T} \right\}^{-1} \\
&= \frac{\sigma_{\varepsilon}^{2}\boldsymbol{\Sigma}_{\boldsymbol{\beta}_{\mathrm{B}}}}{mn} \left\{ \boldsymbol{I}_{d_{\mathrm{B}}} + o_P(1)\boldsymbol{\Sigma}_{\boldsymbol{\beta}_{\mathrm{B}}}\mathbf{1}_{d_{\mathrm{B}}}\mathbf{1}_{d_{\mathrm{B}}}^{T} + \dots \right\} \\
&= \frac{\sigma_{\varepsilon}^{2}\boldsymbol{\Sigma}_{\boldsymbol{\beta}_{\mathrm{B}}}}{mn} + o_P(m^{-1}n^{-1})\mathbf{1}_{d_{\mathrm{B}}}\mathbf{1}_{d_{\mathrm{B}}}^{T}.
\end{aligned} \tag{3.13}$$

Using the result for carrying out block matrix inversion presented in Section 1, $\boldsymbol{A}^{11}$, $\boldsymbol{A}^{12}$, $\boldsymbol{A}^{21}$ and $\boldsymbol{A}^{22}$ can be calculated as follows. Firstly we have,

$$\boldsymbol{A}^{11} = \boldsymbol{A}_{11}^{-1} + \boldsymbol{A}_{11}^{-1}\boldsymbol{A}_{12}(\boldsymbol{A}_{22} - \boldsymbol{A}_{21}\boldsymbol{A}_{11}^{-1}\boldsymbol{A}_{12})^{-1}\boldsymbol{A}_{21}\boldsymbol{A}_{11}^{-1}.$$

Note that

$$(\boldsymbol{A}_{22} - \boldsymbol{A}_{21}\boldsymbol{A}_{11}^{-1}\boldsymbol{A}_{12})^{-1} = \boldsymbol{A}_{22}^{-1} + \boldsymbol{A}_{22}^{-1}\boldsymbol{A}_{21}\boldsymbol{A}_{11}^{-1}\boldsymbol{A}_{12}\boldsymbol{A}_{22}^{-1} + \dots$$

It follows that

$$\boldsymbol{A}_{11}^{-1}\boldsymbol{A}_{12}(\boldsymbol{A}_{22} - \boldsymbol{A}_{21}\boldsymbol{A}_{11}^{-1}\boldsymbol{A}_{12})^{-1}\boldsymbol{A}_{21}\boldsymbol{A}_{11}^{-1} = O_P(m^{-1}n^{-1})\mathbf{1}_{d_{\mathrm{A}}}\mathbf{1}_{d_{\mathrm{A}}}^{T}.$$

Therefore by making use of (3.12), we get,

$$\boldsymbol{A}^{11} = m^{-1}\boldsymbol{\Sigma} + O_P(m^{-1}n^{-1})\mathbf{1}_{d_{\mathrm{A}}}\mathbf{1}_{d_{\mathrm{A}}}^{T}.$$

Next we have,

$$\boldsymbol{A}^{12} = -(\boldsymbol{A}_{11} - \boldsymbol{A}_{12}\boldsymbol{A}_{22}^{-1}\boldsymbol{A}_{21})^{-1}\boldsymbol{A}_{12}\boldsymbol{A}_{22}^{-1}.$$

Note that

$$(\boldsymbol{A}_{11} - \boldsymbol{A}_{12}\boldsymbol{A}_{22}^{-1}\boldsymbol{A}_{21})^{-1} = \boldsymbol{A}_{11}^{-1} + \boldsymbol{A}_{11}^{-1}\boldsymbol{A}_{12}\boldsymbol{A}_{22}^{-1}\boldsymbol{A}_{21}\boldsymbol{A}_{11}^{-1} + \dots$$

Therefore, by making use of (3.12) and (3.13), we have

$$\boldsymbol{A}^{12} = O_P(m^{-1}n^{-1})\mathbf{1}_{d_{\mathrm{A}}}\mathbf{1}_{d_{\mathrm{B}}}^{T}.$$

Note that

$$\boldsymbol{A}^{21} = (\boldsymbol{A}^{12})^{T}.$$

Therefore, we have

$$\boldsymbol{A}^{21} = O_P(m^{-1}n^{-1})\mathbf{1}_{d_{\mathrm{B}}}\mathbf{1}_{d_{\mathrm{A}}}^{T}.$$

Finally, making use of (3.13), we get

$$\boldsymbol{A}^{22} = \boldsymbol{A}_{22}^{-1} + \boldsymbol{A}_{22}^{-1}\boldsymbol{A}_{21}(\boldsymbol{A}_{11} - \boldsymbol{A}_{12}\boldsymbol{A}_{22}^{-1}\boldsymbol{A}_{21})^{-1}\boldsymbol{A}_{12}\boldsymbol{A}_{22}^{-1}.$$

Also note that

$$\boldsymbol{A}_{22}^{-1}\boldsymbol{A}_{21}(\boldsymbol{A}_{11}-\boldsymbol{A}_{12}\boldsymbol{A}_{22}^{-1}\boldsymbol{A}_{21})^{-1}\boldsymbol{A}_{12}\boldsymbol{A}_{22}^{-1}=O_P(m^{-1}n^{-2})\mathbf{1}_{d_\mathrm{B}}\mathbf{1}_{d_\mathrm{B}}^T.$$

This leads to

$$\begin{aligned}\boldsymbol{A}^{22}&=\frac{\sigma_\varepsilon^2\boldsymbol{\Sigma}_{\boldsymbol{\beta}_\mathrm{B}}}{mn}+o_P(m^{-1}n^{-1})\mathbf{1}_{d_\mathrm{B}}\mathbf{1}_{d_\mathrm{B}}^T+O_P(m^{-1}n^{-2})\mathbf{1}_{d_\mathrm{B}}\mathbf{1}_{d_\mathrm{B}}^T\\&=\frac{\sigma_\varepsilon^2\boldsymbol{\Sigma}_{\boldsymbol{\beta}_\mathrm{B}}}{mn}+o_P(m^{-1}n^{-1})\mathbf{1}_{d_\mathrm{B}}\mathbf{1}_{d_\mathrm{B}}^T.\end{aligned}$$

Using the expressions for $\boldsymbol{A}^{11}$, $\boldsymbol{A}^{12}$, $\boldsymbol{A}^{21}$ and $\boldsymbol{A}^{22}$, we have the following expression for $I(\boldsymbol{\beta}_\mathrm{A},\boldsymbol{\beta}_\mathrm{B})^{-1}$

$$I(\boldsymbol{\beta}_\mathrm{A},\boldsymbol{\beta}_\mathrm{B})^{-1}=\begin{bmatrix}m^{-1}\boldsymbol{\Sigma}+O_P(m^{-1}n^{-1})\mathbf{1}_{d_\mathrm{A}}\mathbf{1}_{d_\mathrm{A}}^T & O_P(m^{-1}n^{-1})\mathbf{1}_{d_\mathrm{A}}\mathbf{1}_{d_\mathrm{B}}^T\\ O_P(m^{-1}n^{-1})\mathbf{1}_{d_\mathrm{B}}\mathbf{1}_{d_\mathrm{A}}^T & \frac{\sigma_\varepsilon^2\boldsymbol{\Sigma}_{\boldsymbol{\beta}_\mathrm{B}}}{mn}+o_P(m^{-1}n^{-1})\mathbf{1}_{d_\mathrm{B}}\mathbf{1}_{d_\mathrm{B}}^T\end{bmatrix}.$$

3.4.4.2 Expression for Bottom Right Block of Inverse Fisher Information Matrix

Similarly, let us partition $I(\mathrm{vech}(\boldsymbol{\Sigma}),\sigma_\varepsilon^2)$ as follows

$$I(\mathrm{vech}(\boldsymbol{\Sigma}),\sigma_\varepsilon^2)=\begin{bmatrix}\boldsymbol{B}_{11} & \boldsymbol{B}_{12}\\ \boldsymbol{B}_{21} & \boldsymbol{B}_{22}\end{bmatrix}\text{ where }\boldsymbol{B}_{21}=\boldsymbol{B}_{12}^T.$$

The expressions for $\boldsymbol{B}_{11}$, $\boldsymbol{B}_{12}$, $\boldsymbol{B}_{21}$ and $\boldsymbol{B}_{22}$ are currently as follows

$$\begin{aligned}\boldsymbol{B}_{11}&=\frac{m}{2}\left\{\boldsymbol{D}_{d_\mathrm{A}}^T\left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right)\boldsymbol{D}_{d_\mathrm{A}}\right\}+O_P(mn^{-1})\mathbf{1}_{d_\mathrm{A}(d_\mathrm{A}+1)/2}\mathbf{1}_{d_\mathrm{A}(d_\mathrm{A}+1)/2}^T,\\ \boldsymbol{B}_{12}&=O_P(mn^{-1})\mathbf{1}_{d_\mathrm{A}(d_\mathrm{A}+1)/2},\\ \boldsymbol{B}_{21}&=O_P(mn^{-1})\mathbf{1}_{d_\mathrm{A}(d_\mathrm{A}+1)/2}^T,\\ \boldsymbol{B}_{22}&=\frac{mn}{2(\sigma_\varepsilon^2)^2}+O_P(m).\end{aligned}$$

Let $I(\mathrm{vech}(\boldsymbol{\Sigma}),\sigma_\varepsilon^2)^{-1}$ assume the following form

$$\begin{aligned}I(\mathrm{vech}(\boldsymbol{\Sigma}),\sigma_\varepsilon^2)^{-1}&=\begin{bmatrix}\boldsymbol{B}_{11} & \boldsymbol{B}_{12}\\ \boldsymbol{B}_{21} & \boldsymbol{B}_{22}\end{bmatrix}^{-1}\\&=\begin{bmatrix}\boldsymbol{B}^{11} & \boldsymbol{B}^{12}\\ \boldsymbol{B}^{21} & \boldsymbol{B}^{22}\end{bmatrix}\quad\text{where }\boldsymbol{B}^{21}=(\boldsymbol{B}^{12})^T.\end{aligned}$$

Firstly note that

$$\begin{aligned}
\boldsymbol{B}_{11}^{-1} &= \left[\frac{m}{2}\left\{\boldsymbol{D}_{d_{\mathrm{A}}}^{T}\left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right)\boldsymbol{D}_{d_{\mathrm{A}}}\right\}+O_P(mn^{-1})\mathbf{1}_{d_{\mathrm{A}}(d_{\mathrm{A}}+1)/2}\mathbf{1}_{d_{\mathrm{A}}(d_{\mathrm{A}}+1)/2}^{T}\right]^{-1} \\
&= \left[\frac{m}{2}\left\{\boldsymbol{D}_{d_{\mathrm{A}}}^{T}\left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right)\boldsymbol{D}_{d_{\mathrm{A}}}\right\}\left\{\boldsymbol{I}_{d_{\mathrm{A}}(d_{\mathrm{A}}+1)/2}+O_P(n^{-1})\mathbf{1}_{d_{\mathrm{A}}(d_{\mathrm{A}}+1)/2}\mathbf{1}_{d_{\mathrm{A}}(d_{\mathrm{A}}+1)/2}^{T}\right\}\right]^{-1} \\
&= \frac{2}{m}\left\{\boldsymbol{I}_{d_{\mathrm{A}}(d_{\mathrm{A}}+1)/2}+O_P(n^{-1})\mathbf{1}_{d_{\mathrm{A}}(d_{\mathrm{A}}+1)/2}\mathbf{1}_{d_{\mathrm{A}}(d_{\mathrm{A}}+1)/2}^{T}\right\}^{-1}\left\{\boldsymbol{D}_{d_{\mathrm{A}}}^{T}\left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right)\boldsymbol{D}_{d_{\mathrm{A}}}\right\}^{-1} \\
&= \frac{2}{m}\left\{\boldsymbol{I}_{d_{\mathrm{A}}(d_{\mathrm{A}}+1)/2}+O_P(n^{-1})\mathbf{1}_{d_{\mathrm{A}}(d_{\mathrm{A}}+1)/2}\mathbf{1}_{d_{\mathrm{A}}(d_{\mathrm{A}}+1)/2}^{T}+\ldots\right\}\boldsymbol{D}_{d_{\mathrm{A}}}^{+}\left(\boldsymbol{\Sigma}\otimes\boldsymbol{\Sigma}\right)\boldsymbol{D}_{d_{\mathrm{A}}}^{+T} \\
&= \frac{2\boldsymbol{D}_{d_{\mathrm{A}}}^{+}\left(\boldsymbol{\Sigma}\otimes\boldsymbol{\Sigma}\right)\boldsymbol{D}_{d_{\mathrm{A}}}^{+T}}{m}+O_P(m^{-1}n^{-1})\mathbf{1}_{d_{\mathrm{A}}(d_{\mathrm{A}}+1)/2}\mathbf{1}_{d_{\mathrm{A}}(d_{\mathrm{A}}+1)/2}^{T}.
\end{aligned} \tag{3.14}$$

Also note that

$$\begin{aligned}
\boldsymbol{B}_{22}^{-1} &= \left\{\frac{mn}{2(\sigma_{\varepsilon}^{2})^{2}}+O_P(m)\right\}^{-1} \\
&= \left[\frac{mn}{2(\sigma_{\varepsilon}^{2})^{2}}\left\{1+O_P(n^{-1})\right\}\right]^{-1} \\
&= \frac{2(\sigma_{\varepsilon}^{2})^{2}}{mn}\left\{1+O_P(n^{-1})+\ldots\right\} \\
&= \frac{2(\sigma_{\varepsilon}^{2})^{2}}{mn}+O_P(m^{-1}n^{-2}).
\end{aligned} \tag{3.15}$$

Using the result for carrying out block matrix inversion presented in Section 1, $\boldsymbol{B}^{11}$, $\boldsymbol{B}^{12}$, $\boldsymbol{B}^{21}$ and $\boldsymbol{B}^{22}$ can be calculated as follows. Firstly, we have,

$$\boldsymbol{B}^{11}=\boldsymbol{B}_{11}^{-1}+\boldsymbol{B}_{11}^{-1}\boldsymbol{B}_{12}(\boldsymbol{B}_{22}-\boldsymbol{B}_{21}\boldsymbol{B}_{11}^{-1}\boldsymbol{B}_{12})^{-1}\boldsymbol{B}_{21}\boldsymbol{B}_{11}^{-1}.$$

Note that

$$(\boldsymbol{B}_{22}-\boldsymbol{B}_{21}\boldsymbol{B}_{11}^{-1}\boldsymbol{B}_{12})^{-1}=\boldsymbol{B}_{22}^{-1}+\boldsymbol{B}_{22}^{-1}\boldsymbol{B}_{21}\boldsymbol{A}_{11}^{-1}\boldsymbol{B}_{12}\boldsymbol{B}_{22}^{-1}+\ldots$$

It follows that

$$\boldsymbol{B}_{11}^{-1}\boldsymbol{B}_{12}(\boldsymbol{B}_{22}-\boldsymbol{B}_{21}\boldsymbol{B}_{11}^{-1}\boldsymbol{B}_{12})^{-1}\boldsymbol{B}_{21}\boldsymbol{B}_{11}^{-1}=O_P(m^{-1}n^{-3})\mathbf{1}_{d_{\mathrm{A}}(d_{\mathrm{A}}+1)/2}\mathbf{1}_{d_{\mathrm{A}}(d_{\mathrm{A}}+1)/2}^{T}.$$

Therefore by making use of (3.14), we get,

$$\boldsymbol{B}^{11}=\frac{2\boldsymbol{D}_{d_{\mathrm{A}}}^{+}\left(\boldsymbol{\Sigma}\otimes\boldsymbol{\Sigma}\right)\boldsymbol{D}_{d_{\mathrm{A}}}^{+T}}{m}+O_P(m^{-1}n^{-1})\mathbf{1}_{d_{\mathrm{A}}(d_{\mathrm{A}}+1)/2}\mathbf{1}_{d_{\mathrm{A}}(d_{\mathrm{A}}+1)/2}^{T}.$$

Next we have,

$$\boldsymbol{B}^{12}=-(\boldsymbol{B}_{11}-\boldsymbol{B}_{12}\boldsymbol{B}_{22}^{-1}\boldsymbol{B}_{21})^{-1}\boldsymbol{B}_{12}\boldsymbol{B}_{22}^{-1}.$$

Note that

$$(\boldsymbol{B}_{11}-\boldsymbol{B}_{12}\boldsymbol{B}_{22}^{-1}\boldsymbol{B}_{21})^{-1}=\boldsymbol{B}_{11}^{-1}+\boldsymbol{B}_{11}^{-1}\boldsymbol{B}_{12}\boldsymbol{B}_{22}^{-1}\boldsymbol{B}_{21}\boldsymbol{B}_{11}^{-1}+\ldots$$

Therefore, by making use of (3.14) and (3.15), we have

$$\boldsymbol{B}^{12}=O_P(m^{-1}n^{-2})\mathbf{1}_{d_{\mathrm{A}}(d_{\mathrm{A}}+1)/2}.$$

Subsequently

$$\boldsymbol{B}^{21} = (\boldsymbol{B}^{12})^T.$$

Therefore we have

$$\boldsymbol{B}^{21} = O_P(m^{-1}n^{-2})\mathbf{1}^T_{d_{\mathrm{A}}(d_{\mathrm{A}}+1)/2}.$$

Finally, making use of (3.15), we get

$$\boldsymbol{B}^{22} = \boldsymbol{B}_{22}^{-1} + \boldsymbol{B}_{22}^{-1}\boldsymbol{B}_{21}(\boldsymbol{B}_{11} - \boldsymbol{B}_{12}\boldsymbol{B}_{22}^{-1}\boldsymbol{B}_{21})^{-1}\boldsymbol{B}_{12}\boldsymbol{B}_{22}^{-1}.$$

Note that

$$\boldsymbol{B}_{22}^{-1}\boldsymbol{B}_{21}(\boldsymbol{B}_{11} - \boldsymbol{B}_{12}\boldsymbol{B}_{22}^{-1}\boldsymbol{B}_{21})^{-1}\boldsymbol{B}_{12}\boldsymbol{B}_{22}^{-1} = O_P(m^{-1}n^{-4}).$$

This leads to

$$\begin{aligned}\boldsymbol{B}^{22} &= \frac{2(\sigma_\varepsilon^2)^2}{mn} + O_P(m^{-1}n^{-2}) + O_P(m^{-1}n^{-4}) \\ &= \frac{2(\sigma_\varepsilon^2)^2}{mn} + O_P(m^{-1}n^{-2}).\end{aligned}$$

Using the expressions for $\boldsymbol{B}^{11}$, $\boldsymbol{B}^{12}$, $\boldsymbol{B}^{21}$ and $\boldsymbol{B}^{22}$, we have the following expression for $I(\mathrm{vech}(\boldsymbol{\Sigma}), \sigma_\varepsilon^2)^{-1}$

$$\begin{aligned}&I(\mathrm{vech}(\boldsymbol{\Sigma}), \sigma_\varepsilon^2)^{-1} \\ &= \begin{bmatrix} \frac{2\boldsymbol{D}^+_{d_{\mathrm{A}}}(\boldsymbol{\Sigma}\otimes\boldsymbol{\Sigma})\boldsymbol{D}^{+T}_{d_{\mathrm{A}}}}{m} + O_P(m^{-1}n^{-1})\mathbf{1}_{d_{\mathrm{A}}(d_{\mathrm{A}}+1)/2}\mathbf{1}^T_{d_{\mathrm{A}}(d_{\mathrm{A}}+1)/2} & O_P(m^{-1}n^{-2})\mathbf{1}_{d_{\mathrm{A}}(d_{\mathrm{A}}+1)/2} \\ O_P(m^{-1}n^{-2})\mathbf{1}^T_{d_{\mathrm{A}}(d_{\mathrm{A}}+1)/2} & \frac{2(\sigma_\varepsilon^2)^2}{mn} + O_P(m^{-1}n^{-2}) \end{bmatrix}.\end{aligned}$$

The resultant expression for $I(\boldsymbol{\beta}_{\mathrm{A}}, \boldsymbol{\beta}_{\mathrm{B}}, \mathrm{vech}(\boldsymbol{\Sigma}), \sigma_\varepsilon^2)$ is

$$\begin{aligned}&I(\boldsymbol{\beta}_{\mathrm{A}}, \boldsymbol{\beta}_{\mathrm{B}}, \mathrm{vech}(\boldsymbol{\Sigma}), \sigma_\varepsilon^2)^{-1} \\ &= \begin{bmatrix} I(\boldsymbol{\beta}_{\mathrm{A}}, \boldsymbol{\beta}_{\mathrm{B}})^{-1} & 0 \\ 0 & I(\mathrm{vech}(\boldsymbol{\Sigma}), \sigma_\varepsilon^2)^{-1} \end{bmatrix} \\ &= I(\boldsymbol{\beta}_{\mathrm{A}}, \boldsymbol{\beta}_{\mathrm{B}}, \mathrm{vech}(\boldsymbol{\Sigma}), \sigma_\varepsilon^2)^{-1}_\infty \\ &\quad + \frac{1}{mn}\begin{bmatrix} O_P(1)\mathbf{1}_{d_{\mathrm{A}}}\mathbf{1}^T_{d_{\mathrm{A}}} & O_P(1)\mathbf{1}_{d_{\mathrm{A}}}\mathbf{1}^T_{d_{\mathrm{B}}} & \mathbf{0} & \mathbf{0} \\ O_P(1)\mathbf{1}_{d_{\mathrm{B}}}\mathbf{1}^T_{d_{\mathrm{A}}} & o_P(1)\mathbf{1}_{d_{\mathrm{B}}}\mathbf{1}^T_{d_{\mathrm{B}}} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & O_P(1)\mathbf{1}_{d_{\mathrm{A}}(d_{\mathrm{A}}+1)/2}\mathbf{1}^T_{d_{\mathrm{A}}(d_{\mathrm{A}}+1)/2} & O_P(n^{-1})\mathbf{1}_{d_{\mathrm{A}}(d_{\mathrm{A}}+1)/2} \\ \mathbf{0} & \mathbf{0} & O_P(n^{-1})\mathbf{1}^T_{d_{\mathrm{A}}(d_{\mathrm{A}}+1)/2} & O_P(n^{-1}) \end{bmatrix},\end{aligned}$$

where

$$I(\boldsymbol{\beta}_{\mathrm{A}}, \boldsymbol{\beta}_{\mathrm{B}}, \mathrm{vech}(\boldsymbol{\Sigma}), \sigma_\varepsilon^2)^{-1}_\infty = \begin{bmatrix} \frac{\boldsymbol{\Sigma}}{m} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \frac{\sigma_\varepsilon^2\Sigma_{\boldsymbol{\beta}_{\mathrm{B}}}}{mn} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \frac{2\boldsymbol{D}^+_{d_{\mathrm{A}}}(\boldsymbol{\Sigma}\otimes\boldsymbol{\Sigma})\boldsymbol{D}^{+T}_{d_{\mathrm{A}}}}{m} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \frac{2(\sigma_\varepsilon^2)^2}{mn} \end{bmatrix}.$$

3.4.5 Derivation of the Final Asymptotic Normality Result for Gaussian Response Linear Mixed Models

For a matrix $\boldsymbol{M}$ let

$$\|\boldsymbol{M}\|_F = \sqrt{\mathrm{tr}(\boldsymbol{M}^T\boldsymbol{M})}$$

denote the *Frobenius norm* of $\boldsymbol{M}$.

Consider working with the order $(\boldsymbol{\beta}_{\mathrm{A}}, \mathrm{vech}(\boldsymbol{\Sigma}), \boldsymbol{\beta}_{\mathrm{B}}, \sigma^2_\varepsilon)$ rather than the order in $(\boldsymbol{\beta}_{\mathrm{A}}, \boldsymbol{\beta}_{\mathrm{B}}, \mathrm{vech}(\boldsymbol{\Sigma}), \sigma^2_\varepsilon)$, when working with the Fisher information matrix for the derivation of the final asymptotic normality result. From standard results concerning asymptotic normality of maximum likelihood estimators we have

$$\{I\left(\boldsymbol{\beta}^0_{\mathrm{A}}, \mathrm{vech}(\boldsymbol{\Sigma}^0), \boldsymbol{\beta}^0_{\mathrm{B}}, (\sigma^2)^0\right)^{-1}\}^{-1/2}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^0) \xrightarrow{\mathcal{D}} N(\boldsymbol{0}, \boldsymbol{I})$$

where $\widehat{\boldsymbol{\theta}} = \left[\widehat{\boldsymbol{\beta}}^T_{\mathrm{A}}\ \mathrm{vech}(\widehat{\boldsymbol{\Sigma}})^T\ \widehat{\boldsymbol{\beta}}^T_{\mathrm{B}}\ \widehat{\sigma}^2\right]^T$ and $\boldsymbol{\theta}^0 = [(\boldsymbol{\beta}^0_{\mathrm{A}})^T\ \{\mathrm{vech}(\boldsymbol{\Sigma}^0)\}^T\ \boldsymbol{\beta}^0_{\mathrm{B}})^T\ ((\sigma^2)^0)^T]^T$. Therefore, for all $(d_{\mathrm{A}} + d_{\mathrm{A}}(d_{\mathrm{A}}+1)/2 + d_{\mathrm{B}} + 1) \times 1$ vectors $\boldsymbol{a} \neq 0$ we have

$$\boldsymbol{a}^T\{I\left(\boldsymbol{\beta}^0_{\mathrm{A}}, \mathrm{vech}(\boldsymbol{\Sigma}^0), \boldsymbol{\beta}^0_{\mathrm{B}}, (\sigma^2)^0\right)^{-1}\}^{-1/2}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^0) \xrightarrow{\mathcal{D}} N(\boldsymbol{0}, \boldsymbol{a}^T\boldsymbol{a}).$$

Note that

$$\begin{aligned}
&\boldsymbol{a}^T\{I\left(\boldsymbol{\beta}^0_{\mathrm{A}}, \mathrm{vech}(\boldsymbol{\Sigma}^0), \boldsymbol{\beta}^0_{\mathrm{B}}, (\sigma^2)^0\right)^{-1}\}^{-1/2}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^0)\\
&= \boldsymbol{a}^T\Big[\{I\left(\boldsymbol{\beta}^0_{\mathrm{A}}, \mathrm{vech}(\boldsymbol{\Sigma}^0), \boldsymbol{\beta}^0_{\mathrm{B}}, (\sigma^2)^0\right)^{-1}_{\infty}\}^{-1/2} + \{I\left(\boldsymbol{\beta}^0_{\mathrm{A}}, \mathrm{vech}(\boldsymbol{\Sigma}^0), \boldsymbol{\beta}^0_{\mathrm{B}}, (\sigma^2)^0\right)^{-1}\}^{-1/2}\\
&\qquad - \{I\left(\boldsymbol{\beta}^0_{\mathrm{A}}, \mathrm{vech}(\boldsymbol{\Sigma}^0), \boldsymbol{\beta}^0_{\mathrm{B}}, (\sigma^2)^0\right)^{-1}_{\infty}\}^{-1/2}\Big](\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^0)\\
&= \boldsymbol{a}^T\{I\left(\boldsymbol{\beta}^0_{\mathrm{A}}, \mathrm{vech}(\boldsymbol{\Sigma}^0), \boldsymbol{\beta}^0_{\mathrm{B}}, (\sigma^2)^0\right)^{-1}_{\infty}\}^{-1/2}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^0)\\
&\quad + \boldsymbol{a}^T\Big[\{I\left(\boldsymbol{\beta}^0_{\mathrm{A}}, \mathrm{vech}(\boldsymbol{\Sigma}^0), \boldsymbol{\beta}^0_{\mathrm{B}}, (\sigma^2)^0\right)^{-1}\}^{-1/2}\\
&\qquad - \{I\left(\boldsymbol{\beta}^0_{\mathrm{A}}, \mathrm{vech}(\boldsymbol{\Sigma}^0), \boldsymbol{\beta}^0_{\mathrm{B}}, (\sigma^2)^0\right)^{-1}_{\infty}\}^{-1/2}\Big](\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^0).
\end{aligned}$$

As a consequence

$$\boldsymbol{a}^T\{I\left(\boldsymbol{\beta}^0_{\mathrm{A}}, \mathrm{vech}(\boldsymbol{\Sigma}^0), \boldsymbol{\beta}^0_{\mathrm{B}}, (\sigma^2)^0\right)^{-1}_{\infty}\}^{-1/2}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^0) + r_{mn}(\boldsymbol{a}) \xrightarrow{\mathcal{D}} N(\boldsymbol{0}, \boldsymbol{a}^T\boldsymbol{a}) \tag{3.16}$$

with

$$\begin{aligned}
&r_{mn}(\boldsymbol{a})\\
&= \boldsymbol{a}^T\Big[\{I\left(\boldsymbol{\beta}^0_{\mathrm{A}}, \mathrm{vech}(\boldsymbol{\Sigma}^0), \boldsymbol{\beta}^0_{\mathrm{B}}, (\sigma^2)^0\right)^{-1}\}^{-1/2} - \{I\left(\boldsymbol{\beta}^0_{\mathrm{A}}, \mathrm{vech}(\boldsymbol{\Sigma}^0), \boldsymbol{\beta}^0_{\mathrm{B}}, (\sigma^2)^0\right)^{-1}_{\infty}\}^{-1/2}\Big](\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^0)\\
&= \boldsymbol{a}^T\Big[\boldsymbol{I} - \{I\left(\boldsymbol{\beta}^0_{\mathrm{A}}, \mathrm{vech}(\boldsymbol{\Sigma}^0), \boldsymbol{\beta}^0_{\mathrm{B}}, (\sigma^2)^0\right)^{-1}_{\infty}\}^{-1/2}\{I\left(\boldsymbol{\beta}^0_{\mathrm{A}}, \mathrm{vech}(\boldsymbol{\Sigma}^0), \boldsymbol{\beta}^0_{\mathrm{B}}, (\sigma^2)^0\right)^{-1}\}^{1/2}\Big]\\
&\quad \times \{I\left(\boldsymbol{\beta}^0_{\mathrm{A}}, \mathrm{vech}(\boldsymbol{\Sigma}^0), \boldsymbol{\beta}^0_{\mathrm{B}}, (\sigma^2)^0\right)^{-1}\}^{-1/2}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^0)\\
&= \left(-\Big[\{I\left(\boldsymbol{\beta}^0_{\mathrm{A}}, \mathrm{vech}(\boldsymbol{\Sigma}^0), \boldsymbol{\beta}^0_{\mathrm{B}}, (\sigma^2)^0\right)^{-1}_{\infty}\}^{-1/2}\{I\left(\boldsymbol{\beta}^0_{\mathrm{A}}, \mathrm{vech}(\boldsymbol{\Sigma}^0), \boldsymbol{\beta}^0_{\mathrm{B}}, (\sigma^2)^0\right)^{-1}\}^{1/2} - \boldsymbol{I}\Big]^T\boldsymbol{a}\right)^T\boldsymbol{Z}
\end{aligned}$$

where $\boldsymbol{Z} \sim N\left(\boldsymbol{0}, \boldsymbol{I}_{d_{\mathrm{A}}+d_{\mathrm{A}}(d_{\mathrm{A}}+1)/2+d_{\mathrm{B}}+1}\right)$. Next, note that using the matrix norm properties $\|-\boldsymbol{A}\| = \|\boldsymbol{A}\|$ and $\|\boldsymbol{AB}\| \leq \|\boldsymbol{A}\|\|\boldsymbol{B}\|$ for any matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ and the fact that $\|\boldsymbol{M}^T\|_F = \|\boldsymbol{M}\|$ for any matrix $\boldsymbol{M}$, we have

$$\begin{aligned}
&\left\| - \left[\{I\left(\boldsymbol{\beta}_{\mathrm{A}}^0, \mathrm{vech}(\boldsymbol{\Sigma}^0), \boldsymbol{\beta}_{\mathrm{B}}^0, (\sigma^2)^0\right)_\infty^{-1}\}^{-1/2}\{I\left(\boldsymbol{\beta}_{\mathrm{A}}^0, \mathrm{vech}(\boldsymbol{\Sigma}^0), \boldsymbol{\beta}_{\mathrm{B}}^0, (\sigma^2)^0\right)^{-1}\}^{1/2} - \boldsymbol{I}\right]^T \boldsymbol{a}\right\|_F \\
&\leq \left\|\left[\{I\left(\boldsymbol{\beta}_{\mathrm{A}}^0, \mathrm{vech}(\boldsymbol{\Sigma}^0), \boldsymbol{\beta}_{\mathrm{B}}^0, (\sigma^2)^0\right)_\infty^{-1}\}^{-1/2}\{I\left(\boldsymbol{\beta}_{\mathrm{A}}^0, \mathrm{vech}(\boldsymbol{\Sigma}^0), \boldsymbol{\beta}_{\mathrm{B}}^0, (\sigma^2)^0\right)^{-1}\}^{1/2} - \boldsymbol{I}\right]^T\right\|_F \|\boldsymbol{a}\|_F \\
&= \left\|\left[\{I\left(\boldsymbol{\beta}_{\mathrm{A}}^0, \mathrm{vech}(\boldsymbol{\Sigma}^0), \boldsymbol{\beta}_{\mathrm{B}}^0, (\sigma^2)^0\right)_\infty^{-1}\}^{-1/2}\{I\left(\boldsymbol{\beta}_{\mathrm{A}}^0, \mathrm{vech}(\boldsymbol{\Sigma}^0), \boldsymbol{\beta}_{\mathrm{B}}^0, (\sigma^2)^0\right)^{-1}\}^{1/2} - \boldsymbol{I}\right]\right\|_F \|\boldsymbol{a}\|_F.
\end{aligned} \tag{3.17}$$

Our next aim is to establish that

$$\left\|\left[\{I\left(\boldsymbol{\beta}_{\mathrm{A}}^0, \mathrm{vech}(\boldsymbol{\Sigma}^0), \boldsymbol{\beta}_{\mathrm{B}}^0, (\sigma^2)^0\right)_\infty^{-1}\}^{-1/2}\{I\left(\boldsymbol{\beta}_{\mathrm{A}}^0, \mathrm{vech}(\boldsymbol{\Sigma}^0), \boldsymbol{\beta}_{\mathrm{B}}^0, (\sigma^2)^0\right)^{-1}\}^{1/2} - \boldsymbol{I}\right]\right\|_F \xrightarrow{P} 0 \tag{3.18}$$

Recall that

$$\begin{aligned}
&I(\boldsymbol{\beta}_{\mathrm{A}}, \mathrm{vech}(\boldsymbol{\Sigma}), \boldsymbol{\beta}_{\mathrm{B}}, \sigma_\varepsilon^2)^{-1} \\
&= I(\boldsymbol{\beta}_{\mathrm{A}}, \mathrm{vech}(\boldsymbol{\Sigma}), \boldsymbol{\beta}_{\mathrm{B}}, \sigma_\varepsilon^2)_\infty^{-1} \\
&\quad + \frac{1}{m}\begin{bmatrix}
O_P(n^{-1})\mathbf{1}_{d_{\mathrm{A}}}\mathbf{1}_{d_{\mathrm{A}}}^T & \boldsymbol{0} & O_P(n^{-1})\mathbf{1}_{d_{\mathrm{A}}}\mathbf{1}_{d_{\mathrm{B}}}^T & \boldsymbol{0} \\
\boldsymbol{0} & O_P(n^{-1})\mathbf{1}_{d_{\mathrm{A}}(d_{\mathrm{A}}+1)/2}\mathbf{1}_{d_{\mathrm{A}}(d_{\mathrm{A}}+1)/2}^T & \boldsymbol{0} & O_P(n^{-2})\mathbf{1}_{d_{\mathrm{A}}(d_{\mathrm{A}}+1)/2} \\
O_P(n^{-1})\mathbf{1}_{d_{\mathrm{B}}}\mathbf{1}_{d_{\mathrm{A}}}^T & \boldsymbol{0} & o_P(n^{-1})\mathbf{1}_{d_{\mathrm{B}}}\mathbf{1}_{d_{\mathrm{B}}}^T & \boldsymbol{0} \\
\boldsymbol{0} & O_P(n^{-2})\mathbf{1}_{d_{\mathrm{A}}(d_{\mathrm{A}}+1)/2}^T & \boldsymbol{0} & O_P(n^{-2})
\end{bmatrix}
\end{aligned}$$

where

$$I(\boldsymbol{\beta}_{\mathrm{A}}, \mathrm{vech}(\boldsymbol{\Sigma}), \boldsymbol{\beta}_{\mathrm{B}}, \sigma_\varepsilon^2)_\infty^{-1} = \frac{1}{m}\begin{bmatrix}
\boldsymbol{\Sigma} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} \\
\boldsymbol{0} & 2\boldsymbol{D}_{d_{\mathrm{A}}}^+ (\boldsymbol{\Sigma} \otimes \boldsymbol{\Sigma}) \boldsymbol{D}_{d_{\mathrm{A}}}^{+T} & \boldsymbol{0} & \boldsymbol{0} \\
\boldsymbol{0} & \boldsymbol{0} & \frac{\sigma_\varepsilon^2 \Sigma_{\boldsymbol{\beta}_{\mathrm{B}}}}{n} & \boldsymbol{0} \\
\boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \frac{2(\sigma_\varepsilon^2)^2}{n}
\end{bmatrix},$$

so that

$$\left\{I\left(\boldsymbol{\beta}_{\mathrm{A}}^0, \mathrm{vech}(\boldsymbol{\Sigma}^0), \boldsymbol{\beta}_{\mathrm{B}}^0, (\sigma^2)^0\right)_\infty^{-1}\right\}^{-1/2}\left\{I\left(\boldsymbol{\beta}_{\mathrm{A}}^0, \mathrm{vech}(\boldsymbol{\Sigma}^0), \boldsymbol{\beta}_{\mathrm{B}}^0, (\sigma^2)^0\right)^{-1}\right\}^{1/2} = \boldsymbol{M}_{n,\infty}^{-1/2}\boldsymbol{M}_n^{1/2}$$

where

$$\boldsymbol{M}_{n,\infty} = m\left\{I\left(\boldsymbol{\beta}_{\mathrm{A}}^0, \mathrm{vech}(\boldsymbol{\Sigma}^0), \boldsymbol{\beta}_{\mathrm{B}}^0, (\sigma^2)^0\right)_\infty^{-1}\right\} \text{ and } \boldsymbol{M}_n = \left\{I\left(\boldsymbol{\beta}_{\mathrm{A}}^0, \mathrm{vech}(\boldsymbol{\Sigma}^0), \boldsymbol{\beta}_{\mathrm{B}}^0, (\sigma^2)^0\right)^{-1}\right\}.$$

Therefore, we can apply Lemma 3 with the following

$$p = d_{\mathrm{A}} + d_{\mathrm{A}}(d_{\mathrm{A}}+1)/2, \quad \boldsymbol{K} = \begin{bmatrix}
\boldsymbol{\Sigma} & \boldsymbol{0} \\
\boldsymbol{0} & 2\boldsymbol{D}_{d_{\mathrm{A}}}^+ (\boldsymbol{\Sigma} \otimes \boldsymbol{\Sigma}) \boldsymbol{D}_{d_{\mathrm{A}}}^{+T}
\end{bmatrix}, \quad q = d_{\mathrm{B}} + 1$$

and

$$\boldsymbol{L} = \begin{bmatrix}
\frac{\sigma_\varepsilon^2 \Sigma_{\boldsymbol{\beta}_{\mathrm{B}}}}{n} & \boldsymbol{0} \\
\boldsymbol{0} & \frac{2(\sigma_\varepsilon^2)^2}{n}
\end{bmatrix}$$

in order to show that (3.18) holds. It then follows from (3.17) and (3.18) that

$$\left[\{I\left(\boldsymbol{\beta}_{\mathrm{A}}^0, \mathrm{vech}(\boldsymbol{\Sigma}^0), \boldsymbol{\beta}_{\mathrm{B}}^0, (\sigma^2)^0\right)_\infty^{-1}\}^{-1/2}\{I\left(\boldsymbol{\beta}_{\mathrm{A}}^0, \mathrm{vech}(\boldsymbol{\Sigma}^0), \boldsymbol{\beta}_{\mathrm{B}}^0, (\sigma^2)^0\right)^{-1}\}^{1/2} - \boldsymbol{I}\right]\boldsymbol{a} \xrightarrow{P} 0.$$

Application of Slutsky's Theorem then gives $r_{mn}(\boldsymbol{a}) \xrightarrow{P} 0$. From (3.16) and another application of Slutsky's Theorem we have

$$\boldsymbol{a}^T\{I\left(\boldsymbol{\beta}_{\mathrm{A}}^0, \mathrm{vech}(\boldsymbol{\Sigma}^0), \boldsymbol{\beta}_{\mathrm{B}}^0, (\sigma^2)^0\right)_{\infty}^{-1}\}^{-1/2}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^0) \xrightarrow{\mathcal{D}} N(\mathbf{0}, \boldsymbol{a}^T\boldsymbol{a}).$$

It then follows from the Cramér-Wold Device and the Continuous Mapping Theorem that

$$\sqrt{m}\begin{bmatrix} \widehat{\boldsymbol{\beta}}_{\mathrm{A}} - \boldsymbol{\beta}_{\mathrm{A}}^0 \\ \sqrt{n}\left(\widehat{\boldsymbol{\beta}}_{\mathrm{B}} - \boldsymbol{\beta}_{\mathrm{B}}^0\right) \\ \mathrm{vech}(\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}^0) \\ \sqrt{n}\left(\widehat{\sigma}^2 - (\sigma^2)^0\right) \end{bmatrix} \xrightarrow{\mathcal{D}} N\left(\begin{bmatrix} \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \\ 0 \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Sigma}^0 & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & (\sigma^2)^0\Sigma_{\boldsymbol{\beta}_{\mathrm{B}}} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & 2\boldsymbol{D}_{d_{\mathrm{A}}}^{+}\left(\boldsymbol{\Sigma}^0 \otimes \boldsymbol{\Sigma}^0\right)\boldsymbol{D}_{d_{\mathrm{A}}}^{+T} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & 2((\sigma^2)^0)^2 \end{bmatrix}\right)$$

as shown in Theorem 11.

Chapter 4

Usable Asymptotic Normality Results and Inference for Generalized Linear Mixed Models

In this chapter, we aim to derive asymptotic normality results beyond those that have been derived in Chapter 3 for Gaussian response linear mixed models. The main theorem in this chapter concerns the joint asymptotic normality of all of the maximum quasi-likelihood estimators for a generalized linear mixed model. Once again, it elegantly shows faster rates of convergence for fixed effects that are not accompanied by a random effect compared to fixed effects that have a partnering random effect. The results derived in this chapter can also be used for the construction of asymptotically valid confidence intervals and Wald tests for generalized linear mixed model analysis, which will be discussed in the next chapter.

In addition, we extend this theorem under certain circumstances to dispersion parameters, introduced to account for overdispersion, as well. For the class of two-parameter exponential families, maximum likelihood estimation is possible for all model parameters including the dispersion parameter. Thus, we extend Theorem 12 and derive the asymptotic normality results for the maximum likelihood estimator for a dispersion parameter in the Gaussian and Gamma response cases.

The appendix contains the proofs for the theorem introduced in this chapter.

The content of this chapter is published as: [1] Jiang, J., Wand, M.P. and Bhaskaran, A. (2022), Usable and precise asymptotics for generalized linear mixed model analysis and design. Journal of the Royal Statistical Society, Series B, 84: 55-82. DOI: 10.1111/rssb.12473. [2] Bhaskaran, A. and Wand, M.P.(2023), Dispersion parameter extension of precise generalized linear mixed model asymptotics. Statistics and Probability Letters, 193, Article 109691.

4.1 Model Description

Consider the following density, or probability mass, function for the class of one-parameter exponential families

$$p(y;\eta) = \exp\{y\eta - b(\eta) + c(y)\}\, h(y) \tag{4.1}$$

where η is the natural parameter. For example, the Bernoulli probability mass function has $b(x) = \log(1+e^x), c(x) = 0$ and $h(x) = I(x \in \{0,1\})$. Whereas for the Poisson mass function, $b(x) = e^x, c(x) = -\log(x!)$ and $h(x) = I(x \in \{0\} \cup \mathbb{N})$. Here, $I(\mathcal{P}) = 1$ if the condition $\mathcal{P}$ is true and $I(\mathcal{P}) = 0$ if $\mathcal{P}$ is false. If the random variable Y has density, or probability mass, function as in (4.1), then $E(Y) = b'(\eta)$ and $\mathrm{Var}(Y) = b''(\eta)$. To account for overdispersion in data and to allow one to model the variance flexibly, a common modelling extension is implemented such that $\mathrm{Var}(Y) = \phi b''(\eta)$, where $\phi > 0$ represents the dispersion parameter. This involves the replacement of $\log\{p(y;\eta)\}$ by the following quasi-likelihood function

$$\{y\eta - b(\eta) + c(y)\}/\phi + d(y,\phi) \tag{4.2}$$

where $d(y,\phi)$ is a function of y and ϕ only. Note that for ordinary binomial and Poisson response models, ϕ is fixed at 1. For Gaussian and gamma response models, (4.2) corresponds to the expression of $\log\{p(y;\eta)\}$ for a two-parameter exponential family density function and ordinary likelihood applies. In this section, we study generalized linear mixed models of the following form, for observations of the random triples $(\boldsymbol{X}_{\mathrm{A}ij}, \boldsymbol{X}_{\mathrm{B}ij}, Y_{ij})$, $1 \le i \le m, 1 \le j \le n_i$,

$$\begin{aligned} &Y_{ij}|\boldsymbol{X}_{\mathrm{A}ij}, \boldsymbol{X}_{\mathrm{B}ij}, \boldsymbol{U}_i \text{ are independent having quasi-likelihood function (4.2) with} \\ &\text{natural parameter } (\boldsymbol{\beta}^0_{\mathrm{A}} + \boldsymbol{U}_i)^T \boldsymbol{X}_{\mathrm{A}ij} + (\boldsymbol{\beta}^0_{\mathrm{B}})^T X_{\mathrm{B}ij} \text{ such that the } \boldsymbol{U}_i \text{ are} \\ &\text{independent } N(\boldsymbol{0}, \boldsymbol{\Sigma}^0) \text{ random vectors.} \end{aligned} \tag{4.3}$$

The $\boldsymbol{U}_i$ are $d_{\mathrm{A}} \times 1$ unobserved random vectors for each $1 \le i \le m$. The $\boldsymbol{X}_{\mathrm{A}ij}$ are $d_{\mathrm{A}} \times 1$ random vectors corresponding to predictors that are partnered by both a fixed effect and a random effect. The $\boldsymbol{X}_{\mathrm{B}ij}$ are $d_{\mathrm{B}} \times 1$ random vectors which are predictors that have a fixed effect only. Let $\boldsymbol{X}_{ij} \equiv (\boldsymbol{X}^T_{\mathrm{A}ij}, \boldsymbol{X}^T_{\mathrm{B}ij})^T$ denote the combined predictor vectors such that $d_{\mathrm{A}} + d_{\mathrm{B}} = d$. We also assumed that the $\boldsymbol{X}_{ij}$ and $\boldsymbol{U}_i$, for $1 \le i \le m$ and $1 \le j \le n_i$, are independent. The $\boldsymbol{X}_{ij}$ are each assumed as having the same distribution as the $(d_{\mathrm{A}} + d_{\mathrm{B}}) \times 1$ random vector $\boldsymbol{X} = (\boldsymbol{X}^T_{\mathrm{A}}, \boldsymbol{X}^T_{\mathrm{B}})^T$ and the $\boldsymbol{U}_i$ are assumed to be independent $N(\boldsymbol{0}, \boldsymbol{\Sigma}^0)$, each having the same distribution as the random vector $\boldsymbol{U}$.

Then, for any $\boldsymbol{\beta}_{\mathrm{A}}(d_{\mathrm{A}} \times 1)$, $\boldsymbol{\beta}_{\mathrm{B}}(d_{\mathrm{B}} \times 1)$ and $\boldsymbol{\Sigma}(d_{\mathrm{A}} \times d_{\mathrm{A}})$ that is symmetric, positive definite and conditional on the $\boldsymbol{X}_{ij}$ data, the maximum likelihood estimator of $(\boldsymbol{\beta}_{\mathrm{A}}^0, \boldsymbol{\beta}_{\mathrm{B}}^0, \boldsymbol{\Sigma}^0)$ is,

$$(\widehat{\boldsymbol{\beta}}_{\mathrm{A}}, \widehat{\boldsymbol{\beta}}_{\mathrm{B}}, \widehat{\boldsymbol{\Sigma}}) = \underset{\boldsymbol{\beta}_{\mathrm{A}}, \boldsymbol{\beta}_{\mathrm{B}}, \boldsymbol{\Sigma}}{\operatorname{argmax}}\, \ell(\boldsymbol{\beta}_{\mathrm{A}}, \boldsymbol{\beta}_{\mathrm{B}}, \boldsymbol{\Sigma})$$

where $\ell(\boldsymbol{\beta}_{\mathrm{A}}, \boldsymbol{\beta}_{\mathrm{B}}, \boldsymbol{\Sigma})$ is the conditional log-likelihood and has the expression

$$\begin{aligned}\ell(\boldsymbol{\beta}_{\mathrm{A}}, \boldsymbol{\beta}_{\mathrm{B}}, \boldsymbol{\Sigma}) = \sum_{i=1}^{m}\sum_{j=1}^{n}\left[\left\{Y_{ij}\left(\boldsymbol{\beta}_{\mathrm{A}}^T\boldsymbol{X}_{\mathrm{A}ij} + \boldsymbol{\beta}_{\mathrm{B}}^T\boldsymbol{X}_{\mathrm{B}ij}\right) + c(Y_{ij})\right\}/\phi + d(Y_{ij},\phi)\right] - \frac{m}{2}\log|2\pi\boldsymbol{\Sigma}| \\ + \sum_{i=1}^{m}\log\int_{\mathbb{R}^{d_{\mathrm{A}}}}\exp\left[\sum_{j=1}^{n}\left\{Y_{ij}\boldsymbol{u}^T\boldsymbol{X}_{\mathrm{A}ij} - b((\boldsymbol{\beta}_{\mathrm{A}}+u)^T\boldsymbol{X}_{\mathrm{A}ij} + \boldsymbol{\beta}_{\mathrm{B}}^T\boldsymbol{X}_{\mathrm{B}ij})\right\}/\phi \right. \\ \left. - \frac{1}{2}\boldsymbol{u}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{u}\right]d\boldsymbol{u}.\end{aligned}$$

Note that one-parameter exponential family densities have a variety of desirable properties, so that the regularity conditions for Theorem 7 are met. For example, in the class of one-parameter exponential families, the support of $f(\boldsymbol{y}|\boldsymbol{\theta})$ does not depend on $\boldsymbol{\theta}$. Additional assumptions, for instance, having $\boldsymbol{\Sigma}^0$ being positive definite, ensures that the true values for all model parameters are interior to the parameter space. Thus, the model description above ensures that the regularity conditions required of the density function is met for the convergence in distribution result for MLEs. Similar explanations apply for the model descriptions in Chapters 6 and 7 as well.

4.2 Notation

Define

$$n \equiv \frac{1}{m}\sum_{i=1}^{m} n_i = \text{ average of the within-group sample sizes,}$$

$$\boldsymbol{\Omega}_{\boldsymbol{\beta}_{\mathrm{B}}}(\boldsymbol{U}) \equiv E\left\{b''\left((\boldsymbol{\beta}_{\mathrm{A}}^0 + \boldsymbol{U})^T\boldsymbol{X}_{\mathrm{A}} + (\boldsymbol{\beta}_{\mathrm{B}}^0)^T\boldsymbol{X}_{\mathrm{B}}\right)\begin{bmatrix}\boldsymbol{X}_{\mathrm{A}}\boldsymbol{X}_{\mathrm{A}}^T & \boldsymbol{X}_{\mathrm{A}}\boldsymbol{X}_{\mathrm{B}}^T \\ \boldsymbol{X}_{\mathrm{B}}\boldsymbol{X}_{\mathrm{A}}^T & \boldsymbol{X}_{\mathrm{B}}\boldsymbol{X}_{\mathrm{B}}^T\end{bmatrix}\Big|\boldsymbol{U}\right\}$$

and

$$\boldsymbol{\Lambda}_{\boldsymbol{\beta}_{\mathrm{B}}} \equiv \left(E\left[\left\{\text{lower right } d_{\mathrm{B}} \times d_{\mathrm{B}} \text{ block of } \boldsymbol{\Omega}_{\boldsymbol{\beta}_{\mathrm{B}}}(\boldsymbol{U})^{-1}\right\}^{-1}\right]\right)^{-1}.$$

4.3 Asymptotic Normality Theorem

The main theoretical contribution of this chapter is an asymptotic normality theorem for the maximum quasi-likelihood estimators for a generalized linear mixed model as described in Section 4.1.

The theorem relies on the following assumptions:

(A4) The number of groups m diverges to ∞.

(A5) The within-group sample sizes n_i diverge to ∞ in such a way that $n_i/n \to C_i$ for constants $0 < C_i < \infty$, $1 \leq i \leq m$. Also, $n/m \to 0$ as m and n diverge.

(A6) The distribution of $\boldsymbol{X}$ is such that

$$E\left[\frac{E\left[\max\{1, \|\boldsymbol{X}\|\}^8 \max\{1, b''\left((\boldsymbol{\beta}_{\mathrm{A}} + \boldsymbol{U})^T \boldsymbol{X}_{\mathrm{A}} + \boldsymbol{\beta}_{\mathrm{B}}^T \boldsymbol{X}_{\mathrm{B}}\right)\}^4 | \boldsymbol{U}\right]}{\min\{1, \lambda_{\min}(E\{\boldsymbol{X}_{\mathrm{A}} \boldsymbol{X}_{\mathrm{A}}^T b''\left((\boldsymbol{\beta}_{\mathrm{A}} + \boldsymbol{U})^T \boldsymbol{X}_{\mathrm{A}} + \boldsymbol{\beta}_{\mathrm{B}}^T \boldsymbol{X}_{\mathrm{B}}\right) | \boldsymbol{U}\})\}^2}\right] < \infty$$

for all $\boldsymbol{\beta}_{\mathrm{A}} \in \mathbb{R}^{d_{\mathrm{A}}}$, $\boldsymbol{\beta}_{\mathrm{B}} \in \mathbb{R}^{d_{\mathrm{B}}}$ and $\boldsymbol{\Sigma}$ that is a $d_{\mathrm{A}} \times d_{\mathrm{A}}$ symmetric and positive definite matrix.

Theorem 12. *Assume that conditions (A4) - (A6) hold. Then we have the following*

$$\sqrt{m}\begin{bmatrix} \widehat{\boldsymbol{\beta}}_A - \boldsymbol{\beta}_A^0 \\ \sqrt{n}\left(\widehat{\boldsymbol{\beta}}_B - \boldsymbol{\beta}_B^0\right) \\ \mathrm{vech}(\widehat{\boldsymbol{\Sigma}}) - \mathrm{vech}(\boldsymbol{\Sigma}^0) \end{bmatrix} \xrightarrow{\mathcal{D}} N\left(\begin{bmatrix} \boldsymbol{0} \\ \boldsymbol{0} \\ \boldsymbol{0} \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Sigma}^0 & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \phi\boldsymbol{\Lambda}_{\boldsymbol{\beta}_B} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & 2\boldsymbol{D}_{d_A}^{+}\left(\boldsymbol{\Sigma}^0 \otimes \boldsymbol{\Sigma}^0\right)\boldsymbol{D}_{d_A}^{+T} \end{bmatrix}\right).$$

Proof of Theorem 12 is in the appendices. Some remarks concerning Theorem 12 are:

1. Firstly, note that the asymptotic variances of the maximum quasi-likelihood estimator of the fixed effects that are accompanied by random effects and the maximum quasi-likelihood estimator of the variance and covariance parameters of the random effects, both have a convergence rate of m^{-1}. On the other hand, the asymptotic variance of the estimator of the fixed effects unaccompanied by random effects has a much faster convergence rate of $(mn)^{-1}$.

2. The off-diagonal entries of the asymptotic variance-covariance matrix of the maximum quasi-likelihood estimators of the model parameters are zero matrices. Hence, this implies asymptotic orthogonality between the $\boldsymbol{\beta}_{\mathrm{A}}$, $\boldsymbol{\beta}_{\mathrm{B}}$ and $\boldsymbol{\Sigma}$ model parameters.

3. In existing literature, results such as Theorem 3 of Nie (2007) contains limits and expectations over the response distribution in their Fisher information approximations. In Theorem 12, we avoid such limits and expectations over the response distribution. The $\Lambda_{\boldsymbol{\beta}_B}$ matrix only involves expectations over the simpler random effects distribution. Therefore the results in this thesis, including Theorem 12, provide expressions that are easy to implement for practical tasks such as confidence interval construction, making them "usable" in practice compared to other theoretical asymptotic normality results (rather than results that make use of observed Fisher information) available for maximum likelihood estimators for generalized linear mixed models.

4.4 Dispersion Parameter Extension

In Theorem 12, we treat the dispersion parameter ϕ as being fixed. When considering the Gaussian and Gamma response cases, all of the model parameters in (4.2), including ϕ, can be estimated using ordinary maximum likelihood. Theorem 12 can then be extended for maximum likelihood estimation of ϕ^0 and involves the addition of

$$\sqrt{mn}\left(\hat{\phi} - \phi^0\right) \xrightarrow{\mathcal{D}} N(0, v(\phi^0))$$

where $v(x) \equiv 2x^2$ for Gaussian responses and $v(x) \equiv x^4/\{\text{trigamma}(1/x) - x\}$. Proof of the asymptotic variance expression for the maximum likelihood estimator of the dispersion parameter in the Gaussian response case is in Chapter 3 and proof of the asymptotic variance expression for the maximum likelihood estimator of the dispersion parameter in the Gamma response case is in the appendix for this chapter.

Note that from our results, for the Gaussian and gamma response cases, we can conclude that exact orthogonality exists between ϕ and $(\boldsymbol{\beta}_A, \boldsymbol{\beta}_B)$ and asymptotic orthogonality exists between ϕ and $\boldsymbol{\Sigma}$. Thus, the covariance matrices of Theorem 12 still hold for $\widehat{\boldsymbol{\beta}}_A, \widehat{\boldsymbol{\beta}}_B$ and $\text{vech}(\widehat{\boldsymbol{\Sigma}})$. On the other hand, when implementing the quasi-likelihood extension of the Binomial and Poisson response cases, ϕ cannot be estimated via maximum quasi-likelihood and is typically estimated via a method of moments approach. Note that the values of the maximum quasi-likelihood estimates of $\boldsymbol{\beta}_A$, $\boldsymbol{\beta}_B$ and $\boldsymbol{\Sigma}$ asymptotically do not depend on ϕ. This can be deduced from the likelihood equations formed using the first-order asymptotic approximations of the scores that only have ϕ as a constant. Hence, Theorem 12 is unaffected by the estimation of ϕ for the response cases under the umbrella of one-parameter exponential families too.

4.5 Appendix

4.5.1 Multivariate Extension of (2.6) of Tierney et al. (1989)

To carry out the derivations in the next appendix, one has to deal with solving ratios of intractable integrals. In this appendix, we show how to deal with such ratios of intractable integrals containing d-variate arguments by working with its equivalent multi-term Laplace's method expansion instead. This is accomplished by considering the multivariate extension of (2.6) of Tierney et al. (1989), which follows from results in Appendix A of Miyata (2004).

4.5.1.1 Overview

For smooth real-valued functions g, c and h, Equation (2.6) of Tierney et al. (1989) states that

$$\frac{\int_{-\infty}^{\infty} b_N(x)\exp\{-nh(x)\}\,dx}{\int_{-\infty}^{\infty} b_D(x)\exp\{-nh(x)\}\,dx} = g(x^*) + \frac{b_D'(x^*)g'(x^*)}{nb_D(x^*)h''(x^*)} + \frac{g''(x^*)}{2nh''(x^*)} - \frac{g'(x^*)h'''(x^*)}{2nh''(x^*)^2} + O(n^{-2}). \tag{4.4}$$

where

$$g \equiv b_N/b_D \tag{4.5}$$

and

$$x^* \equiv \text{value of } x \text{ that minimises } h \text{ over } \mathbb{R}.$$

Now consider the first two equations in Appendix A of Miyata (2004). Suppose that in the right-hand side of the first equation we set $\boldsymbol{\Theta} = \mathbb{R}^d$, replace the ρ symbol by b_D, replace the h_n symbol by h and replace the integral dummy variable $\boldsymbol{\theta}$ by $\boldsymbol{x}$. Then we get

$$\frac{\int_{\mathbb{R}^d} g(\boldsymbol{x})b_D(\boldsymbol{x})\exp\{-nh(x)\}\,d\boldsymbol{x}}{\int_{\mathbb{R}^d} b_D(\boldsymbol{x})\exp\{-nh(\boldsymbol{x})\}\,d\boldsymbol{x}}.$$

If the function b_N is defined according to (7.3.6), then this quantity becomes

$$\frac{\int_{\mathbb{R}^d} b_N(\boldsymbol{x})\exp\{-nh(x)\}\,d\boldsymbol{x}}{\int_{\mathbb{R}^d} b_D(\boldsymbol{x})\exp\{-nh(\boldsymbol{x})\}\,d\boldsymbol{x}},$$

which is equivalent to both the right-hand sides of the first and second equations in Appendix A of Miyata (2004). So the asymptotic expansion is given by the right-hand side of the second equation in Appendix A of Miyata (2004) with $\widehat{\boldsymbol{\theta}}$ replaced by $\boldsymbol{x}^*$ and ρ replaced b_D.

4.5.1.2 Multivariate Derivative Notation

Let f be a smooth real-valued d-variate function of the d-variate argument $\boldsymbol{x} \equiv (x_1, \ldots, x_d)$. The gradient vector of f is

$$\nabla f(x_1, \ldots, x_d) = d \times 1 \text{ vector with } i\text{th entry } \frac{\partial f(x_1, \ldots, x_d)}{\partial x_i}.$$

The Hessian matrix of f is

$$\nabla^2 f(x_1, \ldots, x_d) = d \times d \text{ matrix with } (i,j)\text{th entry } \frac{\partial^2 f(x_1, \ldots, x_d)}{\partial x_i \partial x_j}.$$

The third derivatives three-dimensional array of f is

$$\nabla^3 f(x_1, \ldots, x_d) = d \times d \times d \text{ array with } (i,j,k)\text{th entry } \frac{\partial^3 f(x_1, \ldots, x_d)}{\partial x_i \partial x_j \partial x_k}.$$

4.5.1.3 Check of the Miyata (2004) Appendix A Result for the Univariate Case

We first make sure that the right-hand side of the second equation in Appendix A of Miyata (2004) matches (2.6) of Tierney et al. (1989) when $d = 1$. Recall that we have to replace ρ by b_D. We first deal with the term that is the one just before the $O(n^{-2})$ term. The author points out that this term is, in fact (with some trivial re-arrangement),

$$\frac{1}{2n}\text{tr}[\{\nabla^2 h(\boldsymbol{x}^*)\}^{-1}\nabla^2 g(\boldsymbol{x}^*)]$$

In the univariate case this becomes

$$\frac{g''(x^*)}{2n\, h''(x^*)}$$

which is one of the terms in (7.3.6).

Consider the second term involving the $\sum_{ij}$ symbol. Next note that Miyata (2004) uses h^{ij} to denote the components of $\{\nabla^2 h(\boldsymbol{x}^*)\}^{-1}$. In the univariate case this is simply $1/h''(x^*)$. In this univariate case the summation over ij collapses to a scalar and the first component is

$$\frac{1}{n} g'(x^*)\{1/h''(x^*)\} b_D'(x^*)/b_D(x^*) = \frac{b_D'(x^*) g'(x^*)}{n b_D(x^*) h''(x^*)},$$

which is also one of the terms in (7.3.6).

It remains to show that the second component of the main $\sum_{ij}$ expression reduces to

$$-\frac{g'(x^*)h'''(x^*)}{2nh''(x^*)^2}.$$

For a general set of model parameters $\boldsymbol{\theta}$, let $h_{j_1 \ldots j_d}(\hat{\boldsymbol{\theta}})$ denote the dth partial derivative $\partial^d h_n(\boldsymbol{\theta})/\partial\theta_{j1} \ldots \partial\theta_{jd}$ with respect to $\boldsymbol{\theta}$ evaluated $\hat{\boldsymbol{\theta}}$. Then from this definition of $h_{j_1 \ldots j_d}$ provided in Miyata (2004), it is apparent that, in the $d = 1$ case, we can replace h_{rsj} by h_{111} and then set it to $h'''(x^*)$. Also, in this $d = 1$ case the summation of rs collapses to a scalar. Combining we get

$$\frac{1}{n} g'(x^*)\{1/h''(x^*)\}(-\tfrac{1}{2})\{1/h''(x^*)\}h'''(x^*) = -\frac{g'(x^*)h'''(x^*)}{2nh''(x^*)^2}$$

as required.

In summary, the result in Appendix A of Miyata (2004) does reduce to (2.6) of Tierney et al. (1989) in the $d = 1$ case.

4.5.1.4 The Multivariate Case

The second equation of Appendix A of Miyata (2004) gives an expansion for

$$\frac{\int_{\mathbb{R}^d} g(\boldsymbol{x}) b_D(\boldsymbol{x}) \exp\{-nh(x)\}\, d\boldsymbol{x}}{\int_{\mathbb{R}^d} b_D(\boldsymbol{x}) \exp\{-nh(\boldsymbol{x})\}\, d\boldsymbol{x}}.$$

It is relatively easy to show that

$$\frac{\int_{\mathbb{R}^d} g(\boldsymbol{x}) b_D(\boldsymbol{x}) \exp\{-nh(x)\}\, d\boldsymbol{x}}{\int_{\mathbb{R}^d} b_D(\boldsymbol{x}) \exp\{-nh(\boldsymbol{x})\}\, d\boldsymbol{x}} = g(\boldsymbol{x}^*) + \frac{\nabla g(\boldsymbol{x}^*)^T \{\nabla^2 h(\boldsymbol{x}^*)\}^{-1} \nabla b_D(\boldsymbol{x}^*)}{n b_D(\boldsymbol{x}^*)} + \frac{\mathrm{tr}[\{\nabla^2 h(\boldsymbol{x}^*)\}^{-1} \nabla^2 g(\boldsymbol{x}^*)]}{2n} + \Upsilon(\boldsymbol{x}^*) + O(n^{-2})$$

for smooth real-valued d-variate functions g, c and h, where $\Upsilon(\boldsymbol{x}^*)$ denotes the term involving third order derivatives and therefore is more challenging when it comes to getting succinct matrix algebraic expressions.

In terms of the subscript derivative notation used in Miyata (2004):

$$\Upsilon(\boldsymbol{x}^*) = -\frac{1}{2n} \sum_{i=1}^{d} \sum_{j=1}^{d} g_i h^{ij} \sum_{k=1}^{d} \sum_{\ell=1}^{d} h^{k\ell} h_{k\ell j}$$

where

$$g_i = i\text{th entry of } \nabla g(\boldsymbol{x}^*),$$
$$h^{ij} = (i,j) \text{ entry of } \{\nabla^2 h(\boldsymbol{x}^*)\}^{-1}$$
$$\text{and} \quad h_{ijk} = (i,j,k) \text{ entry of } \nabla^3 h(\boldsymbol{x}^*).$$

Note that if $\boldsymbol{\omega}(\boldsymbol{x}^*)$ is the $d \times 1$ vector with kth entry equal to

$$\sum_{i=1}^{d} \sum_{j=1}^{d} h^{ij} h_{ijk} = \sum_{i=1}^{d} \sum_{j=1}^{d} [\{\nabla^2 h(\boldsymbol{x}^*)\}^{-1}]_{ij} \{\nabla^3 h(\boldsymbol{x}^*)\}_{ijk}$$

then we have

$$\Upsilon(\boldsymbol{x}^*) = -\frac{\nabla g(\boldsymbol{x}^*)^T\{\nabla^2 h(\boldsymbol{x}^*)\}^{-1}\boldsymbol{\omega}(\boldsymbol{x}^*)}{2n}.$$

We could also define

$$\nabla^3 h(\boldsymbol{x})_{[k]} \equiv d \times d \text{ matrix with } (i,j) \text{ entry equal to the } (i,j,k) \text{ entry of } \nabla^3 h(\boldsymbol{x})$$

and then the kth entry of $\boldsymbol{\omega}(\boldsymbol{x}^*)$ is

$$\text{tr}[\{\nabla^2 h(\boldsymbol{x}^*)\}^{-1}\nabla^3 h(\boldsymbol{x}^*)_{[k]}].$$

4.5.1.5 Final Expression for the Multivariate Extension of (2.6) of Tierney et al. (1989)

Putting everything together from the previous subsubsection, the multivariate extension of (2.6) of Tierney et al. (1989) is:

$$\begin{aligned}\frac{\int_{\mathbb{R}^d} g(\boldsymbol{x}) b_D(\boldsymbol{x}) \exp\{-nh(x)\}\, d\boldsymbol{x}}{\int_{\mathbb{R}^d} b_D(\boldsymbol{x}) \exp\{-nh(\boldsymbol{x})\}\, d\boldsymbol{x}} = g(\boldsymbol{x}^*) &+ \frac{\nabla g(\boldsymbol{x}^*)^T\{\nabla^2 h(\boldsymbol{x}^*)\}^{-1}\nabla b_D(\boldsymbol{x}^*)}{n b_D(\boldsymbol{x}^*)} \\ &+ \frac{\text{tr}[\{\nabla^2 h(\boldsymbol{x}^*)\}^{-1}\nabla^2 g(\boldsymbol{x}^*)]}{2n} \\ &- \frac{\nabla g(\boldsymbol{x}^*)^T\{\nabla^2 h(\boldsymbol{x}^*)\}^{-1}\boldsymbol{\omega}(\boldsymbol{x}^*)}{2n} + O(n^{-2})\end{aligned} \tag{4.6}$$

where

$$\boldsymbol{\omega}(\boldsymbol{x}) \equiv \begin{bmatrix} \text{tr}\big[\{\nabla^2 h(\boldsymbol{x})\}^{-1}\nabla^3 h(\boldsymbol{x})_{[1]}\big] \\ \vdots \\ \text{tr}\big[\{\nabla^2 h(\boldsymbol{x})\}^{-1}\nabla^3 h(\boldsymbol{x})_{[d]}\big] \end{bmatrix}$$

and

$$\nabla^3 h(\boldsymbol{x})_{[k]} \equiv d \times d \text{ matrix with } (i,j) \text{ entry equal to the } (i,j,k) \text{ entry of } \nabla^3 h(\boldsymbol{x}).$$

4.5.2 Proof of Theorem 12

This appendix contains the details for the derivations leading up to Theorem 12.

4.5.2.1 Constructing the Fisher Information Matrix

In order to compute the asymptotic covariance matrix for the maximum quasi-likelihood estimators, we would first need to compute the Fisher information matrix for the model

parameters as per the model description in (4.3). To do so, let

$$\boldsymbol{S}_i \equiv \begin{bmatrix} \boldsymbol{S}_{\mathrm{A}i} \\ \boldsymbol{S}_{\mathrm{B}i} \\ \boldsymbol{S}_{\mathrm{C}i} \end{bmatrix} = \begin{bmatrix} \nabla_{\boldsymbol{\beta}_{\mathrm{A}}} \log p_{\boldsymbol{Y}_i|\boldsymbol{X}_i}(\boldsymbol{Y}_i|\boldsymbol{X}_i) \\ \nabla_{\boldsymbol{\beta}_{\mathrm{B}}} \log p_{\boldsymbol{Y}_i|\boldsymbol{X}_i}(\boldsymbol{Y}_i|\boldsymbol{X}_i) \\ \nabla_{\mathrm{vech}(\boldsymbol{\Sigma})} \log p_{\boldsymbol{Y}_i|\boldsymbol{X}_i}(\boldsymbol{Y}_i|\boldsymbol{X}_i) \end{bmatrix} \tag{4.7}$$

denote the ith contribution to the scores for each of the model parameters. Then the Fisher information matrix can be computed as

$$I(\boldsymbol{\beta}_{\mathrm{A}}, \boldsymbol{\beta}_{\mathrm{B}}, \mathrm{vech}(\boldsymbol{\Sigma})) = \sum_{i=1}^{m} E(\boldsymbol{S}_i \boldsymbol{S}_i^T | \boldsymbol{X}_i).$$

The next few sections then focus on obtaining the expressions for the scores and the quadratic conditional expectations that are required to construct the final Fisher information matrix.

4.5.2.2 Expression for Conditional Density Function

The expression for $p_{\boldsymbol{Y}_i|\boldsymbol{X}_i}(\boldsymbol{Y}_i|\boldsymbol{X}_i)$ as per the model description in (4.3) is as follows

$$\begin{aligned}
&p_{\boldsymbol{Y}_i|\boldsymbol{X}_i}(\boldsymbol{Y}_i|\boldsymbol{X}_i) \\
&= \int_{\mathbb{R}^{d_{\mathrm{A}}}} \prod_{j=1}^{n_i} \left\{ p(Y_{ij}|\boldsymbol{X}_{\mathrm{A}ij}, \boldsymbol{X}_{\mathrm{B}ij}, \boldsymbol{U}_i) \right\} p(\boldsymbol{U}_i) d\boldsymbol{U}_i \\
&= \int_{\mathbb{R}^{d_{\mathrm{A}}}} \exp\left\{ \sum_{j=1}^{n} \Big(\Big[Y_{ij} \left\{ (\boldsymbol{\beta}_{\mathrm{A}} + \boldsymbol{u})^T \boldsymbol{X}_{\mathrm{A}ij} + \boldsymbol{\beta}_{\mathrm{B}}^T \boldsymbol{X}_{\mathrm{B}ij} \right\} - b\left((\boldsymbol{\beta}_{\mathrm{A}} + \boldsymbol{u})^T \boldsymbol{X}_{\mathrm{A}ij} + \boldsymbol{\beta}_{\mathrm{B}}^T \boldsymbol{X}_{\mathrm{B}ij} \right) \right. \\
&\qquad \left. + c(Y_{ij}) \Big] / \phi + d(Y_{ij}, \phi) \Big) \right\} \times (2\pi)^{-d_{\mathrm{A}}/2} |\boldsymbol{\Sigma}|^{-1/2} \exp(-\frac{1}{2} \boldsymbol{u}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{u}) d\boldsymbol{u} \\
&= \int_{\mathbb{R}^{d_{\mathrm{A}}}} |2\pi\boldsymbol{\Sigma}|^{-1/2} \exp\left\{ \sum_{j=1}^{n} \Big(\Big[Y_{ij} \left\{ (\boldsymbol{\beta}_{\mathrm{A}} + \boldsymbol{u})^T \boldsymbol{X}_{\mathrm{A}ij} + \boldsymbol{\beta}_{\mathrm{B}}^T \boldsymbol{X}_{\mathrm{B}ij} \right\} \right. \right. \\
&\qquad \left. \left. - b\left((\boldsymbol{\beta}_{\mathrm{A}} + \boldsymbol{u})^T \boldsymbol{X}_{\mathrm{A}ij} + \boldsymbol{\beta}_{\mathrm{B}}^T \boldsymbol{X}_{\mathrm{B}ij} \right) + c(Y_{ij}) \Big] / \phi + d(Y_{ij}, \phi) \Big) - \frac{1}{2} \boldsymbol{u}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{u} \right\} d\boldsymbol{u} \\
&= |2\pi\boldsymbol{\Sigma}|^{-1/2} \exp\left(\sum_{j=1}^{n} \left[\left\{ Y_{ij} \left(\boldsymbol{\beta}_{\mathrm{A}}^T \boldsymbol{X}_{\mathrm{A}ij} + \boldsymbol{\beta}_{\mathrm{B}}^T \boldsymbol{X}_{\mathrm{B}ij} \right) + c(Y_{ij}) \right\} / \phi + d(Y_{ij}, \phi) \right] \right) \\
&\qquad \times \int_{\mathbb{R}^{d_{\mathrm{A}}}} \exp\left[\sum_{j=1}^{n} \left\{ Y_{ij} \boldsymbol{u}^T \boldsymbol{X}_{\mathrm{A}ij} - b\left((\boldsymbol{\beta}_{\mathrm{A}} + \boldsymbol{u})^T \boldsymbol{X}_{\mathrm{A}ij} + \boldsymbol{\beta}_{\mathrm{B}}^T \boldsymbol{X}_{\mathrm{B}ij} \right) \right\} / \phi - \frac{1}{2} \boldsymbol{u}^T \Sigma^{-1} \boldsymbol{u} \right] d\boldsymbol{u}.
\end{aligned}$$

4.5.2.3 Introduction of Useful Notation and its Properties

Here we introduce expressions that will be useful in summarising the derivations throughout this appendix. Firstly,

$$\mathbf{1}_{d_{\mathrm{A}}^{\boxplus}} \equiv \mathbf{1}_{d_{\mathrm{A}}(d_{\mathrm{A}}+1)/2}$$

and

$$\mathbf{1}_{d_{\mathrm{A}}}^{\otimes 2} \equiv \mathbf{1}_{d_{\mathrm{A}}}\mathbf{1}_{d_{\mathrm{A}}}^{T}, \quad \mathbf{1}_{d_{\mathrm{B}}}^{\otimes 2} \equiv \mathbf{1}_{d_{\mathrm{B}}}\mathbf{1}_{d_{\mathrm{B}}}^{T} \quad \text{and} \quad \mathbf{1}_{d_{\mathrm{A}}^{\boxplus}}^{\otimes 2} \equiv \mathbf{1}_{d_{\mathrm{A}}^{\boxplus}}\mathbf{1}_{d_{\mathrm{A}}^{\boxplus}}^{T}.$$

Then note that

$$\mathcal{G}_{\mathrm{A}i} \equiv \sum_{j=1}^{n}\{Y_{ij} - b'\big((\boldsymbol{\beta}_{\mathrm{A}} + \boldsymbol{U}_i)^T\boldsymbol{X}_{\mathrm{A}ij} + (\boldsymbol{\beta}_{\mathrm{B}})^T\boldsymbol{X}_{\mathrm{B}ij}\big)\}\boldsymbol{X}_{\mathrm{A}ij},$$

$$\mathcal{G}_{\mathrm{B}i} \equiv \sum_{j=1}^{n}\{Y_{ij} - b'\big((\boldsymbol{\beta}_{\mathrm{A}} + \boldsymbol{U}_i)^T\boldsymbol{X}_{\mathrm{A}ij} + (\boldsymbol{\beta}_{\mathrm{B}})^T\boldsymbol{X}_{\mathrm{B}ij}\big)\}\boldsymbol{X}_{\mathrm{B}ij},$$

$$\mathcal{H}_{\mathrm{AA}i} \equiv \sum_{j=1}^{n} b''\big((\boldsymbol{\beta}_{\mathrm{A}} + \boldsymbol{U}_i)^T\boldsymbol{X}_{\mathrm{A}ij} + (\boldsymbol{\beta}_{\mathrm{B}})^T\boldsymbol{X}_{\mathrm{B}ij}\big)\boldsymbol{X}_{\mathrm{A}ij}\boldsymbol{X}_{\mathrm{A}ij}^T,$$

$$\mathcal{H}_{\mathrm{AB}i} = \sum_{j=1}^{n} b''\big((\boldsymbol{\beta}_{\mathrm{A}} + \boldsymbol{U}_i)^T\boldsymbol{X}_{\mathrm{A}ij} + (\boldsymbol{\beta}_{\mathrm{B}})^T\boldsymbol{X}_{\mathrm{B}ij}\big)\boldsymbol{X}_{\mathrm{A}ij}\boldsymbol{X}_{\mathrm{B}ij}^T,$$

$$\mathcal{H}_{\mathrm{BB}i} \equiv \sum_{j=1}^{n} b''\big((\boldsymbol{\beta}_{\mathrm{A}} + \boldsymbol{U}_i)^T\boldsymbol{X}_{\mathrm{A}ij} + (\boldsymbol{\beta}_{\mathrm{B}})^T\boldsymbol{X}_{\mathrm{B}ij}\big)\boldsymbol{X}_{\mathrm{B}ij}\boldsymbol{X}_{\mathrm{B}ij}^T.$$

$\mathcal{H}'_{\mathrm{AAA}i}$ is the $d_{\mathrm{A}} \times d_{\mathrm{A}} \times d_{\mathrm{A}}$ array with (r,s,t) entry equal to

$$\sum_{j=1}^{n} b'''\big((\boldsymbol{\beta}_{\mathrm{A}} + \boldsymbol{U}_i)^T\boldsymbol{X}_{\mathrm{A}ij} + (\boldsymbol{\beta}_{\mathrm{B}})^T\boldsymbol{X}_{\mathrm{B}ij}\big)(\boldsymbol{X}_{\mathrm{A}ij})_r(\boldsymbol{X}_{\mathrm{A}ij})_s(\boldsymbol{X}_{\mathrm{A}ij})_t,$$

and $\mathcal{H}'_{\mathrm{AAA}i[t]}$ is the $d_{\mathrm{A}} \times d_{\mathrm{A}}$ matrix with (r,s) entry equal to the (r,s,t) entry of $\mathcal{H}'_{\mathrm{AAA}i}$. Note that the expressions listed above have the following probabilistic orders where

$$\mathcal{G}_{\mathrm{A}i} = O_P(n^{1/2})\mathbf{1}_{d_{\mathrm{A}}}, \quad \mathcal{G}_{\mathrm{B}i} = O_P(n^{1/2})\mathbf{1}_{d_{\mathrm{B}}},$$

$$\mathcal{H}_{\mathrm{AA}i} = O_P(n)\mathbf{1}_{d_{\mathrm{A}}}^{\otimes 2}, \quad \mathcal{H}_{\mathrm{AB}i} = O_P(n)\mathbf{1}_{d_{\mathrm{A}}}\mathbf{1}_{d_{\mathrm{B}}}^{T} \quad \text{and} \quad \mathcal{H}_{\mathrm{BB}i} = O_P(n)\mathbf{1}_{d_{\mathrm{B}}}^{\otimes 2}.$$

The Conditional Expectation of $\mathcal{G}_{Ai}$ and $\mathcal{G}_{Bi}$ Given $(\boldsymbol{X}_i, \boldsymbol{U}_i)$

$$\begin{aligned}
E(\mathcal{G}_{Ai}|\boldsymbol{X}_i, \boldsymbol{U}_i) &= E\left(\left[\sum_{j=1}^{n}\left\{Y_{ij} - b'\left(\left(\boldsymbol{\beta}_{\mathrm{A}} + \boldsymbol{U}_i\right)^T \boldsymbol{X}_{\mathrm{A}ij} + \boldsymbol{\beta}_{\mathrm{B}}^T \boldsymbol{X}_{\mathrm{B}ij}\right)\right\} \boldsymbol{X}_{\mathrm{A}ij}\right] |\boldsymbol{X}_i, \boldsymbol{U}_i\right) \\
&= \sum_{j=1}^{n}\left\{E\left(Y_{ij}|\boldsymbol{X}_i, \boldsymbol{U}_i\right) - b'\left(\left(\boldsymbol{\beta}_{\mathrm{A}} + \boldsymbol{U}_i\right)^T \boldsymbol{X}_{\mathrm{A}ij} + \boldsymbol{\beta}_{\mathrm{B}}^T \boldsymbol{X}_{\mathrm{B}ij}\right)\right\} \boldsymbol{X}_{\mathrm{A}ij} \\
&= \mathbf{0}.
\end{aligned}$$

Similarly we also have that

$$E(\mathcal{G}_{Bi}|\boldsymbol{X}_i, U_i) = \mathbf{0}.$$

The Conditional Expectations of $\mathcal{G}_{Ai}\,\mathcal{G}_{Ai}^T$, $\mathcal{G}_{Ai}\,\mathcal{G}_{Bi}^T$ and $\mathcal{G}_{Bi}\,\mathcal{G}_{Bi}^T$ Given $(\boldsymbol{X}_i, \boldsymbol{U}_i)$

$$\begin{aligned}
&E(\mathcal{G}_{Ai}\,\mathcal{G}_{Ai}^T|\boldsymbol{X}_i, \boldsymbol{U}_i) \\
&= E\Bigg(\left[\sum_{j=1}^{n}\left\{Y_{ij} - b'\left(\left(\boldsymbol{\beta}_{\mathrm{A}} + \boldsymbol{U}_i\right)^T \boldsymbol{X}_{\mathrm{A}ij} + \boldsymbol{\beta}_{\mathrm{B}}^T \boldsymbol{X}_{\mathrm{B}ij}\right)\right\} \boldsymbol{X}_{\mathrm{A}ij}\right] \\
&\quad \times \left[\sum_{j=1}^{n}\left\{Y_{ij'} - b'\left(\left(\boldsymbol{\beta}_{\mathrm{A}} + \boldsymbol{U}_i\right)^T \boldsymbol{X}_{\mathrm{A}ij'} + \boldsymbol{\beta}_{\mathrm{B}}^T \boldsymbol{X}_{\mathrm{B}ij'}\right)\right\} \boldsymbol{X}_{\mathrm{A}ij'}\right]^T |\boldsymbol{X}_i, \boldsymbol{U}_i\Bigg) \\
&= \sum_{j\neq j'}\sum E\Bigg(\left[\left\{Y_{ij} - b'\left(\left(\boldsymbol{\beta}_{\mathrm{A}} + \boldsymbol{U}_i\right)^T \boldsymbol{X}_{\mathrm{A}ij} + \boldsymbol{\beta}_{\mathrm{B}}^T \boldsymbol{X}_{\mathrm{B}ij}\right)\right\} \boldsymbol{X}_{\mathrm{A}ij}\right] \\
&\quad \times \left[\left\{Y_{ij'} - b'\left(\left(\boldsymbol{\beta}_{\mathrm{A}} + \boldsymbol{U}_i\right)^T \boldsymbol{X}_{\mathrm{A}ij'} + \boldsymbol{\beta}_{\mathrm{B}}^T \boldsymbol{X}_{\mathrm{B}ij'}\right)\right\} \boldsymbol{X}_{\mathrm{A}ij'}\right]^T |\boldsymbol{X}_i, \boldsymbol{U}_i\Bigg) \\
&\quad + \sum_{j=1}^{n} E\Bigg(\left[\left\{Y_{ij} - b'\left(\left(\boldsymbol{\beta}_{\mathrm{A}} + \boldsymbol{U}_i\right)^T \boldsymbol{X}_{\mathrm{A}ij} + \boldsymbol{\beta}_{\mathrm{B}}^T \boldsymbol{X}_{\mathrm{B}ij}\right)\right\} \boldsymbol{X}_{\mathrm{A}ij}\right] \\
&\quad \times \left[\left\{Y_{ij} - b'\left(\left(\boldsymbol{\beta}_{\mathrm{A}} + \boldsymbol{U}_i\right)^T \boldsymbol{X}_{\mathrm{A}ij} + \boldsymbol{\beta}_{\mathrm{B}}^T \boldsymbol{X}_{\mathrm{B}ij}\right)\right\} \boldsymbol{X}_{\mathrm{A}ij}\right]^T |\boldsymbol{X}_i, \boldsymbol{U}_i\Bigg) \\
&= \sum_{j\neq j'}\sum E\Bigg\{\left[\left\{Y_{ij} - b'\left(\left(\boldsymbol{\beta}_{\mathrm{A}} + \boldsymbol{U}_i\right)^T \boldsymbol{X}_{\mathrm{A}ij} + \boldsymbol{\beta}_{\mathrm{B}}^T \boldsymbol{X}_{\mathrm{B}ij}\right)\right\} \boldsymbol{X}_{\mathrm{A}ij}|\boldsymbol{X}_i, \boldsymbol{U}_i\right] \\
&\quad \times E\left(\left[\left\{Y_{ij'} - b'\left(\left(\boldsymbol{\beta}_{\mathrm{A}} + \boldsymbol{U}_i\right)^T \boldsymbol{X}_{\mathrm{A}ij'} + \boldsymbol{\beta}_{\mathrm{B}}^T \boldsymbol{X}_{\mathrm{B}ij'}\right)\right\} \boldsymbol{X}_{\mathrm{A}ij'}\right]^T |\boldsymbol{X}_i, \boldsymbol{U}_i\right)\Bigg\} \\
&\quad + \sum_{j=1}^{n} \mathrm{Var}\left(Y_{ij}\boldsymbol{X}_{\mathrm{A}ij}|\boldsymbol{X}_i, \boldsymbol{U}_i\right)
\end{aligned}$$

The equation above simplifies to

$$\begin{aligned}
&\sum_{j\neq j'}\sum \left[\left\{E\left(Y_{ij}|\boldsymbol{X}_i,\boldsymbol{U}_i\right)-b'\left(\left(\boldsymbol{\beta}_{\mathrm{A}}+\boldsymbol{U}_i\right)^T\boldsymbol{X}_{\mathrm{A}ij}+\boldsymbol{\beta}_{\mathrm{B}}^T\boldsymbol{X}_{\mathrm{B}ij}\right)\right\}\boldsymbol{X}_{\mathrm{A}ij}\right] \\
&\quad\times\left[\left\{E\left(Y_{ij'}|\boldsymbol{X}_i,\boldsymbol{U}_i\right)-b'\left(\left(\boldsymbol{\beta}_{\mathrm{A}}+\boldsymbol{U}_i\right)^T\boldsymbol{X}_{\mathrm{A}ij'}+\boldsymbol{\beta}_{\mathrm{B}}^T\boldsymbol{X}_{\mathrm{B}ij'}\right)\right\}\boldsymbol{X}_{\mathrm{A}ij'}\right]^T \\
&\quad+\sum_{j=1}^{n}\boldsymbol{X}_{\mathrm{A}ij}\mathrm{Var}\left(Y_{ij}|\boldsymbol{X}_i,\boldsymbol{U}_i\right)\boldsymbol{X}_{\mathrm{A}ij}^T \\
&=\phi\sum_{j=1}^{n}b''\left(\left(\boldsymbol{\beta}_{\mathrm{A}}+\boldsymbol{U}_i\right)^T\boldsymbol{X}_{\mathrm{A}ij}+\boldsymbol{\beta}_{\mathrm{B}}^T\boldsymbol{X}_{\mathrm{B}ij}\right)\boldsymbol{X}_{\mathrm{A}ij}\boldsymbol{X}_{\mathrm{A}ij}^T \\
&=\phi\mathcal{H}_{\mathrm{AA}i}.
\end{aligned}$$

Similarly we also have that

$$E(\mathcal{G}_{Ai}\,\mathcal{G}_{Bi}^T|\boldsymbol{X}_i,U_i)=\phi\mathcal{H}_{\mathrm{AB}i}\text{ and }E(\mathcal{G}_{Bi}\,\mathcal{G}_{Bi}^T|\boldsymbol{X}_i,U_i)=\phi\mathcal{H}_{\mathrm{BB}i}.$$

4.5.2.4 Computing an Asymptotic Approximation for the First Entry in (4.7)

To overcome the intractability of the ratio of integrals present when deriving the scores with respect to each of the model parameters, we will work with an asymptotic approximation of the ratio of integrals by using a multi-term Laplace's method expansion as described in the appendix in Subsection 4.5.1. Hence, the ith contribution to the score of $\boldsymbol{\beta}_{\mathrm{A}}$ is

$$\begin{aligned}
\boldsymbol{S}_{\mathrm{A}i}&=\nabla_{\boldsymbol{\beta}_{\mathrm{A}}}\log p_{\boldsymbol{Y}_i|\boldsymbol{X}_i}(\boldsymbol{Y}_i|\boldsymbol{X}_i) \\
&=\frac{\int_{\mathbb{R}^{d_{\mathrm{A}}}}\boldsymbol{b}_N^{1\mathrm{st}}(\boldsymbol{u})\exp\{-nh_N(\boldsymbol{u})\}\,d\boldsymbol{u}}{\int_{\mathbb{R}^{d_{\mathrm{A}}}}b_D^{1\mathrm{st}}(\boldsymbol{u})\exp\{-nh_N(\boldsymbol{u})\}\,d\boldsymbol{u}}
\end{aligned}$$

where

$$\begin{aligned}
\boldsymbol{b}_N^{1\mathrm{st}}(\boldsymbol{u})&\equiv\exp\Big(-\tfrac{1}{2}\boldsymbol{u}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{u}\Big)\frac{1}{\phi}\sum_{j=1}^{n}\Big\{Y_{ij}-b'\big((\boldsymbol{\beta}_{\mathrm{A}}+\boldsymbol{u})^T\boldsymbol{X}_{\mathrm{A}ij}+(\boldsymbol{\beta}_{\mathrm{B}})^T\boldsymbol{X}_{\mathrm{B}ij}\big)\Big\}\boldsymbol{X}_{\mathrm{A}ij}, \\
b_D^{1\mathrm{st}}(\boldsymbol{u})&\equiv\exp\Big(-\tfrac{1}{2}\boldsymbol{u}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{u}\Big)\quad\text{and} \\
h_N(\boldsymbol{u})&\equiv-\frac{1}{n\phi}\sum_{j=1}^{n}\Big\{Y_{ij}\boldsymbol{u}^T\boldsymbol{X}_{\mathrm{A}ij}-b\big((\boldsymbol{\beta}_{\mathrm{A}}+\boldsymbol{u})^T\boldsymbol{X}_{\mathrm{A}ij}+(\boldsymbol{\beta}_{\mathrm{B}})^T\boldsymbol{X}_{\mathrm{B}ij}\big)\Big\}.
\end{aligned}$$

Now define

$$\begin{aligned}
\boldsymbol{U}_i^*&\equiv\text{value of }\boldsymbol{u}\text{ that minimises }h_N(\boldsymbol{u}) \\
&=\text{value of }\boldsymbol{u}\text{ such that }\nabla_{\boldsymbol{u}}h_N(\boldsymbol{u})=\boldsymbol{0} \\
&=\text{value of }\boldsymbol{u}\text{ such that }\sum_{j=1}^{n}\Big\{Y_{ij}-b'\big((\boldsymbol{\beta}_{\mathrm{A}}+\boldsymbol{u})^T\boldsymbol{X}_{\mathrm{A}ij}+(\boldsymbol{\beta}_{\mathrm{B}})^T\boldsymbol{X}_{\mathrm{B}ij}\big)\Big\}\boldsymbol{X}_{\mathrm{A}ij}=\boldsymbol{0}.
\end{aligned}$$

However,

$$\boldsymbol{b}_N^{\text{1st}}(\boldsymbol{U}_i^*) = \mathbf{0}.$$

This violates the condition in Hsu (1948), in the sense that $\boldsymbol{b}_N^{\text{1st}}(\boldsymbol{U}_i^*) \neq \mathbf{0}$ is required in order for the Laplace approximation to hold. To counter this issue, firstly note that the numerator of $\boldsymbol{S}_{\text{A}i}$ is

$$\int_{\mathbb{R}^{d_\text{A}}} \{\nabla_{\boldsymbol{u}} s(\boldsymbol{u})\}\, t(\boldsymbol{u})\, d\boldsymbol{u} \tag{4.8}$$

where

$$s(\boldsymbol{u}) \equiv \exp\left(\frac{n}{\phi}\left[\frac{1}{n}\sum_{j=1}^{n}\left\{Y_{ij}\boldsymbol{u}^T\boldsymbol{X}_{\text{A}ij} - b\big((\boldsymbol{\beta}_\text{A}+\boldsymbol{u})^T\boldsymbol{X}_{\text{A}ij} + (\boldsymbol{\beta}_\text{B})^T\boldsymbol{X}_{\text{B}ij}\big)\right\}\right]\right)$$

and

$$t(\boldsymbol{u}) \equiv \exp\Big(-\tfrac{1}{2}\boldsymbol{u}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{u}\Big).$$

Using an $\mathbb{R}^{d_\text{A}}$ extension of integration by parts, we can replace (4.8) by

$$-\int_{\mathbb{R}^{d_\text{A}}} s(\boldsymbol{u})\, \{\nabla_{\boldsymbol{u}} t(\boldsymbol{u})\}\, d\boldsymbol{u}.$$

We then obtain that

$$\begin{aligned} dt(\boldsymbol{u}) &= \exp\Big(-\tfrac{1}{2}\boldsymbol{u}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{u}\Big)d\left\{-\tfrac{1}{2}\boldsymbol{u}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{u}\right\} \\ &= -\exp\Big(-\tfrac{1}{2}\boldsymbol{u}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{u}\Big)\tfrac{1}{2}\boldsymbol{u}^T\boldsymbol{\Sigma}^{-1}d\boldsymbol{u}, \end{aligned}$$

which leads to

$$\nabla_{\boldsymbol{u}} t(\boldsymbol{u}) = -\exp\Big(-\tfrac{1}{2}\boldsymbol{u}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{u}\Big)\boldsymbol{\Sigma}^{-1}\boldsymbol{u}.$$

Now by rewriting the numerator of $\boldsymbol{S}_{\text{A}i}$, we have,

$$\boldsymbol{S}_{\text{A}i} = \frac{\int_{\mathbb{R}^{d_\text{A}}} \boldsymbol{b}_N(\boldsymbol{u}) \exp\{-nh_N(\boldsymbol{u})\}\, d\boldsymbol{u}}{\int_{\mathbb{R}^{d_\text{A}}} b_D(\boldsymbol{u}) \exp\{-nh_N(\boldsymbol{u})\}\, d\boldsymbol{u}}$$

where

$$\begin{aligned} \boldsymbol{b}_N(\boldsymbol{u}) &\equiv \exp\Big(-\tfrac{1}{2}\boldsymbol{u}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{u}\Big)\boldsymbol{\Sigma}^{-1}\boldsymbol{u} \\ b_D(\boldsymbol{u}) &\equiv \exp\Big(-\tfrac{1}{2}\boldsymbol{u}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{u}\Big) \quad \text{and} \\ h_N(\boldsymbol{u}) &\equiv -\frac{1}{n\phi}\sum_{j=1}^{n}\left\{Y_{ij}\boldsymbol{u}^T\boldsymbol{X}_{\text{A}ij} - b\left((\boldsymbol{\beta}_\text{A}+\boldsymbol{u})^T\boldsymbol{X}_{\text{A}ij} + (\boldsymbol{\beta}_\text{B})^T\boldsymbol{X}_{\text{B}ij}\right)\right\}. \end{aligned}$$

Expansion of $\boldsymbol{U}_i^*$

Here we find an asymptotic expression for $\boldsymbol{U}_i^*$. We have that

$$\begin{aligned}
0 &= \nabla_{\boldsymbol{u}} h_N(\boldsymbol{u}) \\
&= \sum_{j=1}^{n} \left\{ Y_{ij} - b'\left((\boldsymbol{\beta}_{\mathrm{A}} + \boldsymbol{U}_i^*)^T \boldsymbol{X}_{\mathrm{A}ij} + (\boldsymbol{\beta}_{\mathrm{B}})^T \boldsymbol{X}_{\mathrm{B}ij} \right) \right\} \boldsymbol{X}_{\mathrm{A}ij} \\
&= \sum_{j=1}^{n} \left\{ Y_{ij} - b'\left((\boldsymbol{\beta}_{\mathrm{A}}^0 + \boldsymbol{U}_i)^T \boldsymbol{X}_{\mathrm{A}ij} + (\boldsymbol{\beta}_{\mathrm{B}}^0)^T \boldsymbol{X}_{\mathrm{B}ij} \right) \right\} \boldsymbol{X}_{\mathrm{A}ij} \\
&\quad - \sum_{j=1}^{n} b''\left((\boldsymbol{\beta}_{\mathrm{A}} + \boldsymbol{U}_i)^T \boldsymbol{X}_{\mathrm{A}ij} + (\boldsymbol{\beta}_{\mathrm{B}})^T \boldsymbol{X}_{\mathrm{B}ij} \right) \boldsymbol{X}_{\mathrm{A}ij} \boldsymbol{X}_{\mathrm{A}ij}^T (\boldsymbol{U}_i^* - \boldsymbol{U}_i) + r_{it} \\
&= \mathcal{G}_{\mathrm{A}i} - \mathcal{H}_{\mathrm{AA}i} (\boldsymbol{U}_i^* - \boldsymbol{U}_i) + r_{it}
\end{aligned}$$

where r_{it} is the Lagrange form of the remainder and is a quadratic form in $\boldsymbol{U}_i^* - \boldsymbol{U}_i$ and a smooth function of $\boldsymbol{U}_{it}^\dagger \equiv (1-t)\boldsymbol{U}_i + t\boldsymbol{U}_i^*$ for some $t \in [0,1]$. Inversion of this asymptotic series leads to

$$\boldsymbol{U}_i^* = \boldsymbol{U}_i + \mathcal{H}_{\mathrm{AA}i}^{-1} \mathcal{G}_{\mathrm{A}i} + O_P(n^{-1}) \mathbf{1}_{d_{\mathrm{A}}}.$$

The $\boldsymbol{e}_k$ Notation

The following notation is useful for the upcoming calculations. For each $1 \le k_A \le d_{\mathrm{A}}$, $1 \le k_B \le d_{\mathrm{B}}$, $1 \le k_C \le d_{\mathrm{A}}^{\boxplus}$, let

$\boldsymbol{e}_{k_A}$ = the $d_{\mathrm{A}} \times 1$ vector with 1 in the k_Ath entry and all other entries equal to zero,

$\boldsymbol{e}_{k_B}$ = the $d_{\mathrm{B}} \times 1$ vector with 1 in the k_Bth entry and all other entries equal to zero

and

$\boldsymbol{e}_{k_C}$ = the $d_{\mathrm{A}}^{\boxplus} \times 1$ vector with 1 in the k_Cth entry and all other entries equal to zero.

The First Term of $\boldsymbol{S}_{Ai}$

For each $1 \le k_A \le d_{\mathrm{A}}$, the k_Ath entry of the first term of $\boldsymbol{S}_{\mathrm{A}i}$ depends on the function

$$g(\boldsymbol{u}) = \frac{\boldsymbol{e}_{k_A}^T \boldsymbol{b}_N(\boldsymbol{u})}{b_D(\boldsymbol{u})} = \boldsymbol{e}_{k_A}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{u},$$

and is

$$g(\boldsymbol{U}_i^*) = \boldsymbol{e}_{k_A}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{U}_i^* = \boldsymbol{e}_{k_A}^T \boldsymbol{\Sigma}^{-1} \left(\boldsymbol{U}_i + \mathcal{H}_{\mathrm{AA}i}^{-1} \mathcal{G}_{\mathrm{A}i} \right) + O_P(n^{-1}).$$

Therefore, the first term of $\boldsymbol{S}_{\mathrm{A}i}$ is

$$\boldsymbol{\Sigma}^{-1} \left(\boldsymbol{U}_i + \mathcal{H}_{\mathrm{AA}i}^{-1} \mathcal{G}_{\mathrm{A}i} \right) + O_P(n^{-1}) \mathbf{1}_{d_{\mathrm{A}}}.$$

The Second Term of $\boldsymbol{S}_{Ai}$

For each $1 \leq k_A \leq d_{\mathrm{A}}$, the k_Ath entry of the second term of $\boldsymbol{S}_{\mathrm{A}i}$ depends on

$$\nabla g(\boldsymbol{u}) = \boldsymbol{\Sigma}^{-1}\boldsymbol{e}_{k_A}$$

and

$$\begin{aligned} db_D(\boldsymbol{u}) &= d\exp(-\frac{1}{2}\boldsymbol{u}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{u}) = -\exp(-\frac{1}{2}\boldsymbol{u}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{u})\boldsymbol{u}^T\boldsymbol{\Sigma}^{-1}d\boldsymbol{u}, \\ \nabla b_D(\boldsymbol{u}) &= -b_D(\boldsymbol{u})\boldsymbol{\Sigma}^{-1}\boldsymbol{u}. \end{aligned}$$

It also depends on $\nabla^2 h_N(\boldsymbol{u})$ which can be evaluated as follows

$$\begin{aligned} dh_N(\boldsymbol{u}) &= -\frac{1}{n\phi}\sum_{j=1}^{n} d\left\{Y_{ij}\boldsymbol{u}^T\boldsymbol{X}_{\mathrm{A}ij} - b'\left((\boldsymbol{\beta}_{\mathrm{A}}+\boldsymbol{u})^T\boldsymbol{X}_{\mathrm{A}ij} + (\boldsymbol{\beta}_{\mathrm{B}})^T\boldsymbol{X}_{\mathrm{B}ij}\right)\right\} \\ &= -\frac{1}{n\phi}\sum_{j=1}^{n} d\left\{Y_{ij}\boldsymbol{u}^T\boldsymbol{X}_{\mathrm{A}ij} - b'\left((\boldsymbol{\beta}_{\mathrm{A}}+\boldsymbol{u})^T\boldsymbol{X}_{\mathrm{A}ij} + (\boldsymbol{\beta}_{\mathrm{B}})^T\boldsymbol{X}_{\mathrm{B}ij}\right)\right\}^T \\ &= -\frac{1}{n\phi}\sum_{j=1}^{n} \left\{Y_{ij}d(\boldsymbol{u})^T\boldsymbol{X}_{\mathrm{A}ij} - b'\left((\boldsymbol{\beta}_{\mathrm{A}}+\boldsymbol{u})^T\boldsymbol{X}_{\mathrm{A}ij} + (\boldsymbol{\beta}_{\mathrm{B}})^T\boldsymbol{X}_{\mathrm{B}ij}\right)d(\boldsymbol{u})^T\boldsymbol{X}_{\mathrm{A}ij}\right\}^T \\ &= -\frac{1}{n\phi}\sum_{j=1}^{n} d\left\{Y_{ij} - b'\left((\boldsymbol{\beta}_{\mathrm{A}}+\boldsymbol{u})^T\boldsymbol{X}_{\mathrm{A}ij} + (\boldsymbol{\beta}_{\mathrm{B}})^T\boldsymbol{X}_{\mathrm{B}ij}\right)\right\}\boldsymbol{X}_{\mathrm{A}ij}^T d\boldsymbol{u}, \\ d^2h_N(\boldsymbol{u}) &= \frac{1}{n\phi}\sum_{j=1}^{n}(d\boldsymbol{u})^T\boldsymbol{X}_{\mathrm{A}ij}\left\{b''\left((\boldsymbol{\beta}_{\mathrm{A}}+\boldsymbol{u})^T\boldsymbol{X}_{\mathrm{A}ij} + (\boldsymbol{\beta}_{\mathrm{B}})^T\boldsymbol{X}_{\mathrm{B}ij}\right)\right\}\boldsymbol{X}_{\mathrm{A}ij}^T d\boldsymbol{u}, \\ \nabla^2 h_N(\boldsymbol{u}) &= \frac{1}{n\phi}\sum_{j=1}^{n}\left\{b''\left((\boldsymbol{\beta}_{\mathrm{A}}+\boldsymbol{u})^T\boldsymbol{X}_{\mathrm{A}ij} + (\boldsymbol{\beta}_{\mathrm{B}})^T\boldsymbol{X}_{\mathrm{B}ij}\right)\right\}\boldsymbol{X}_{\mathrm{A}ij}\boldsymbol{X}_{\mathrm{A}ij}^T. \end{aligned}$$

By using a stochastic Taylor series approximation, one can show that

$$\{\nabla^2 h_N(\boldsymbol{U}_i^*)\}^{-1} = n\phi\mathcal{H}_{\mathrm{AA}i}^{-1} + O_P(n^{-1/2})\mathbf{1}_{d_{\mathrm{A}}}^{\otimes 2}.$$

Hence, for each $1 \leq k_A \leq d_{\mathrm{A}}$, the k_Ath entry of the second term of $\boldsymbol{S}_{\mathrm{A}i}$ is

$$\begin{aligned} &\frac{\nabla g(\boldsymbol{U}_i^*)^T\{\nabla^2 h_N(\boldsymbol{U}_i^*)\}^{-1}\nabla b_D(\boldsymbol{U}_i^*)}{nb_D(\boldsymbol{U}_i^*)} \\ &= \frac{-\boldsymbol{e}_{k_A}^T\boldsymbol{\Sigma}^{-1}\left\{n\phi\mathcal{H}_{\mathrm{AA}i}^{-1} + O_P(n^{-1/2})\mathbf{1}_{d_{\mathrm{A}}}^{\otimes 2}\right\}b_D(\boldsymbol{U}_i^*)\boldsymbol{\Sigma}^{-1}\boldsymbol{U}_i^*}{nb_D(\boldsymbol{U}_i^*)} \\ &= -\frac{1}{n}\boldsymbol{e}_{k_A}^T\boldsymbol{\Sigma}^{-1}\left\{n\phi\mathcal{H}_{\mathrm{AA}i}^{-1} + O_P(n^{-1/2})\mathbf{1}_{d_{\mathrm{A}}}^{\otimes 2}\right\}\boldsymbol{\Sigma}^{-1}\left\{\boldsymbol{U}_i + \mathcal{H}_{\mathrm{AA}i}^{-1}\mathcal{G}_{\mathrm{A}i} + O_P(n^{-1}\mathbf{1}_{d_{\mathrm{A}}})\right\} \\ &= -\phi\boldsymbol{e}_{k_A}^T\boldsymbol{\Sigma}^{-1}\mathcal{H}_{\mathrm{AA}i}^{-1}\boldsymbol{\Sigma}^{-1}\boldsymbol{U}_i + O_P(n^{-3/2}). \end{aligned}$$

Then the leading term behaviour of the second term of $\boldsymbol{S}_{\mathrm{B}i}$ is as follows

$$-\phi\boldsymbol{\Sigma}^{-1}\mathcal{H}_{\mathrm{AA}i}^{-1}\boldsymbol{\Sigma}^{-1}\boldsymbol{U}_i + O_P(n^{-3/2})\mathbf{1}_{d_{\mathrm{A}}} = O_P(n^{-1})\mathbf{1}_{d_{\mathrm{A}}}.$$

The Third Term of $\boldsymbol{S}_{Ai}$

For each $1 \leq k_A \leq d_{\mathrm{A}}$, the k_Ath entry of the third term of $\boldsymbol{S}_{\mathrm{A}i}$ depends on

$$\nabla^2 g(\boldsymbol{u}) = \nabla^2 (\boldsymbol{e}_{k_A}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{u}) = \boldsymbol{O},$$

the $d_{\mathrm{A}} \times d_{\mathrm{A}}$ matrix of zeroes. Since this matrix appears in a trace expression, the third term of $\boldsymbol{S}_{\mathrm{A}i}$ is $\mathbf{0}_{d_{\mathrm{A}}}$.

The Fourth Term of $\boldsymbol{S}_{Ai}$

The contribution from the fourth term of $\boldsymbol{S}_{\mathrm{A}i}$ is $O_P(1)\mathbf{1}_{d_{\mathrm{A}}}$ but does not have a concise matrix algebraic expression. It is of lower order compared to the leading term of $\boldsymbol{S}_{\mathrm{A}i}$.

Overall Leading Term Expression for $\boldsymbol{S}_{Ai}$

Putting the terms of $\boldsymbol{S}_{\mathrm{A}i}$ together, we can assert that

$$\boldsymbol{S}_{\mathrm{A}i} = \boldsymbol{\Sigma}^{-1} \left(\boldsymbol{U}_i + \mathcal{H}_{\mathrm{AA}i}^{-1} \mathcal{G}_{\mathrm{A}i} \right) + O_P(n^{-1}) \mathbf{1}_{d_{\mathrm{A}}}.$$

4.5.2.5 Computing an Asymptotic Approximation for the Second Entry in (4.7)

The ith contribution to the score of $\boldsymbol{\beta}_{\mathrm{B}}$ is

$$\begin{aligned} \boldsymbol{S}_{\mathrm{B}i} &= \nabla_{\boldsymbol{\beta}_{\mathrm{B}}} \log p_{\boldsymbol{Y}_i | \boldsymbol{X}_i}(\boldsymbol{Y}_i | \boldsymbol{X}_i) \\ &= \frac{\int_{\mathbb{R}^{d_{\mathrm{A}}}} b_N(\boldsymbol{u}) \exp\{-n h_N(\boldsymbol{u})\} \, d\boldsymbol{u}}{\int_{\mathbb{R}^{d_{\mathrm{A}}}} b_D(\boldsymbol{u}) \exp\{-n h_N(\boldsymbol{u})\} \, d\boldsymbol{u}} \end{aligned}$$

where

$$\begin{aligned} \boldsymbol{b}_N(\boldsymbol{u}) &\equiv \exp\left(-\tfrac{1}{2} \boldsymbol{u}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{u}\right) \frac{1}{\phi} \sum_{j=1}^{n} \left[\left\{ Y_{ij} - b'\left((\boldsymbol{\beta}_{\mathrm{A}} + \boldsymbol{u})^T \boldsymbol{X}_{\mathrm{A}ij} + (\boldsymbol{\beta}_{\mathrm{B}})^T \boldsymbol{X}_{\mathrm{B}ij} \right) \right\} \boldsymbol{X}_{\mathrm{B}ij} \right] \\ b_D(\boldsymbol{u}) &\equiv \exp\left(-\tfrac{1}{2} \boldsymbol{u}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{u}\right) \quad \text{and} \\ h_N(\boldsymbol{u}) &\equiv -\frac{1}{\phi n} \sum_{j=1}^{n} \left\{ Y_{ij} \boldsymbol{u}^T \boldsymbol{X}_{\mathrm{A}ij} - b\left((\boldsymbol{\beta}_{\mathrm{A}} + \boldsymbol{u})^T \boldsymbol{X}_{\mathrm{A}ij} + (\boldsymbol{\beta}_{\mathrm{B}})^T \boldsymbol{X}_{\mathrm{B}ij} \right) \right\}. \end{aligned}$$

The First Term of $\boldsymbol{S}_{Bi}$

For each $1 \leq k_B \leq d_{\mathrm{B}}$, the k_Bth entry of the first term of $\boldsymbol{S}_{\mathrm{B}i}$ depends on the function

$$g(\boldsymbol{u}) = \frac{\boldsymbol{e}_{k_B}^T \boldsymbol{b}_N(\boldsymbol{u})}{b_D(\boldsymbol{u})} = \frac{1}{\phi} \sum_{j=1}^{n} \boldsymbol{e}_{k_B}^T \left[\left\{ Y_{ij} - b'\left((\boldsymbol{\beta}_{\mathrm{A}} + \boldsymbol{u})^T \boldsymbol{X}_{\mathrm{A}ij} + (\boldsymbol{\beta}_{\mathrm{B}})^T \boldsymbol{X}_{\mathrm{B}ij} \right) \right\} \boldsymbol{X}_{\mathrm{B}ij} \right]$$

and is

$$\begin{aligned}
g(\boldsymbol{U}_i^*) &= \frac{1}{\phi}\sum_{j=1}^{n} \boldsymbol{e}_{k_B}^T \left[\left\{Y_{ij} - b'\left((\boldsymbol{\beta}_{\mathrm{A}} + \boldsymbol{U}_i^*)^T \boldsymbol{X}_{\mathrm{A}ij} + (\boldsymbol{\beta}_{\mathrm{B}})^T \boldsymbol{X}_{\mathrm{B}ij}\right)\right\} \boldsymbol{X}_{\mathrm{B}ij}\right] \\
&= \frac{1}{\phi}\sum_{j=1}^{n} \boldsymbol{e}_{k_B}^T \left[\left\{Y_{ij} - b'\left((\boldsymbol{\beta}_{\mathrm{A}} + \boldsymbol{U}_i + \boldsymbol{U}_i^* - \boldsymbol{U}_i)^T \boldsymbol{X}_{\mathrm{A}ij} + \boldsymbol{\beta}_{\mathrm{B}}^T \boldsymbol{X}_{\mathrm{B}ij}\right)\right\} \boldsymbol{X}_{\mathrm{B}ij}\right] \\
&= \frac{1}{\phi}\sum_{j=1}^{n} \boldsymbol{e}_{k_B}^T \Big[\left\{Y_{ij} - b'\left((\boldsymbol{\beta}_{\mathrm{A}} + \boldsymbol{U}_i)^T \boldsymbol{X}_{\mathrm{A}ij} + \boldsymbol{\beta}_{\mathrm{B}}^T \boldsymbol{X}_{\mathrm{B}ij}\right)\right\} \boldsymbol{X}_{\mathrm{B}ij} \\
&\qquad - b''\left((\boldsymbol{\beta}_{\mathrm{A}} + \boldsymbol{U}_i)^T \boldsymbol{X}_{\mathrm{A}ij} + \boldsymbol{\beta}_{\mathrm{B}}^T \boldsymbol{X}_{\mathrm{B}ij}\right) \boldsymbol{X}_{\mathrm{B}ij}\boldsymbol{X}_{\mathrm{A}ij}^T(\boldsymbol{U}_i^* - \boldsymbol{U}_i) + O_P(n^{-1})\mathbf{1}_{d_{\mathrm{B}}}\Big] \\
&= \frac{1}{\phi}\boldsymbol{e}_{k_B}^T \left(\mathcal{G}_{\mathrm{B}i} - \mathcal{H}_{\mathrm{AB}i}^T \mathcal{H}_{\mathrm{AA}i}^{-1} \mathcal{G}_{\mathrm{A}i}\right) + O_P(1).
\end{aligned}$$

Therefore, the first term of $\boldsymbol{S}_{\mathrm{B}i}$ is

$$\frac{1}{\phi}\left(\mathcal{G}_{\mathrm{B}i} - \mathcal{H}_{\mathrm{AB}i}^T \mathcal{H}_{\mathrm{AA}i}^{-1} \mathcal{G}_{\mathrm{A}i}\right) + O_P(1)\mathbf{1}_{d_{\mathrm{B}}}.$$

The Second Term of $\boldsymbol{S}_{Bi}$

For each $1 \le k_B \le d_{\mathrm{B}}$, the k_Bth entry of the second term of $\boldsymbol{S}_{\mathrm{B}i}$ depends on

$$\begin{aligned}
dg(\boldsymbol{u}) &= \frac{1}{\phi}\sum_{j=1}^{n} \boldsymbol{e}_{k_B}^T \boldsymbol{X}_{\mathrm{B}ij} \left[d\left\{Y_{ij} - b'\left((\boldsymbol{\beta}_{\mathrm{A}} + \boldsymbol{u})^T \boldsymbol{X}_{\mathrm{A}ij} + (\boldsymbol{\beta}_{\mathrm{B}})^T \boldsymbol{X}_{\mathrm{B}ij}\right)\right\}\right] \\
&= -\frac{1}{\phi}\sum_{j=1}^{n} \boldsymbol{e}_{k_B}^T \boldsymbol{X}_{\mathrm{B}ij} b''\left((\boldsymbol{\beta}_{\mathrm{A}} + \boldsymbol{u})^T \boldsymbol{X}_{\mathrm{A}ij} + (\boldsymbol{\beta}_{\mathrm{B}})^T \boldsymbol{X}_{\mathrm{B}ij}\right) \boldsymbol{X}_{\mathrm{A}ij}^T (d\boldsymbol{u}), \\
\nabla g(\boldsymbol{u}) &= -\frac{1}{\phi}\sum_{j=1}^{n} b''\left((\boldsymbol{\beta}_{\mathrm{A}} + \boldsymbol{u})^T \boldsymbol{X}_{\mathrm{A}ij} + (\boldsymbol{\beta}_{\mathrm{B}})^T \boldsymbol{X}_{\mathrm{B}ij}\right) \boldsymbol{X}_{\mathrm{A}ij}\boldsymbol{X}_{\mathrm{B}ij}^T \boldsymbol{e}_{k_B}
\end{aligned}$$

and

$$\nabla b_D(\boldsymbol{u}) = -b_D(\boldsymbol{u})\boldsymbol{\Sigma}^{-1}\boldsymbol{u}.$$

It also depends on

$$\nabla^2 h_N(\boldsymbol{u}) = \frac{1}{n\phi}\sum_{j=1}^{n} \left\{b''\left((\boldsymbol{\beta}_{\mathrm{A}} + \boldsymbol{u})^T \boldsymbol{X}_{\mathrm{A}ij} + (\boldsymbol{\beta}_{\mathrm{B}})^T \boldsymbol{X}_{\mathrm{B}ij}\right)\right\} \boldsymbol{X}_{\mathrm{A}ij}\boldsymbol{X}_{\mathrm{A}ij}^T.$$

By using a stochastic Taylor series approximation, note that

$$\{\nabla^2 h_N(\boldsymbol{U}_i^*)\}^{-1} = n\phi\mathcal{H}_{\mathrm{AA}i}^{-1} + O_P(n^{-1/2})\mathbf{1}_{d_{\mathrm{A}}}^{\otimes 2}.$$

Hence, for each $1 \leq k_B \leq d_{\mathrm{B}}$, the k_Bth entry of the second term of $\boldsymbol{S}_{\mathrm{B}i}$ is

$$\begin{aligned}
&\frac{\nabla g(\boldsymbol{U}_i^*)^T\{\nabla^2 h_N(\boldsymbol{U}_i^*)\}^{-1}\nabla b_D(\boldsymbol{U}_i^*)}{nb_D(\boldsymbol{U}_i^*)} \\
&= \frac{\frac{1}{\phi}\boldsymbol{e}_{k_B}^T\sum_{j=1}^{n} b''\left((\boldsymbol{\beta}_{\mathrm{A}}+\boldsymbol{U}_i^*)^T\boldsymbol{X}_{\mathrm{A}ij}+(\boldsymbol{\beta}_{\mathrm{B}})^T\boldsymbol{X}_{\mathrm{B}ij}\right)\boldsymbol{X}_{\mathrm{B}ij}\boldsymbol{X}_{\mathrm{A}ij}^T\left\{n\phi\mathcal{H}_{\mathrm{AA}i}^{-1}+O_P(n^{-1/2})\mathbf{1}_{d_{\mathrm{A}}}^{\otimes 2}\right\}}{nb_D(\boldsymbol{U}_i^*)} \\
&\quad\times\frac{b_D(\boldsymbol{U}_i^*)\boldsymbol{\Sigma}^{-1}\boldsymbol{U}_i^*}{nb_D(\boldsymbol{U}_i^*)} \\
&= \frac{1}{n\phi}\boldsymbol{e}_{k_B}^T\mathcal{H}_{\mathrm{AB}i}^T\left\{n\phi\mathcal{H}_{\mathrm{AA}i}^{-1}+O_P(n^{-1/2})\mathbf{1}_{d_{\mathrm{A}}}^{\otimes 2}\right\}\boldsymbol{\Sigma}^{-1}\left\{\boldsymbol{U}_i+\mathcal{H}_{\mathrm{AA}i}^{-1}\mathcal{G}_{\mathrm{A}i}+O_P(n^{-1}\mathbf{1}_{d_{\mathrm{A}}})\right\} \\
&= \boldsymbol{e}_{k_B}^T\mathcal{H}_{\mathrm{AB}i}^T\mathcal{H}_{\mathrm{AA}i}^{-1}\boldsymbol{\Sigma}^{-1}\boldsymbol{U}_i+O_P(n^{-1/2}).
\end{aligned}$$

Then the leading term behaviour of the second term of $\boldsymbol{S}_{\mathrm{B}i}$ is as follows

$$\mathcal{H}_{\mathrm{AB}i}^T\mathcal{H}_{\mathrm{AA}i}^{-1}\boldsymbol{\Sigma}^{-1}\boldsymbol{U}_i+O_P(n^{-1/2})\mathbf{1}_{d_{\mathrm{B}}}=O_P(1)\mathbf{1}_{d_{\mathrm{B}}}.$$

The Third Term of $\boldsymbol{S}_{Bi}$

The contribution from the third term of $\boldsymbol{S}_{\mathrm{B}i}$ is $O_P(1)\mathbf{1}_{d_{\mathrm{B}}}$ but does not have a concise matrix algebraic expression. It is of lower order compared to the leading term of $\boldsymbol{S}_{\mathrm{B}i}$.

The Fourth Term of $\boldsymbol{S}_{Bi}$

Similarly, the contribution from the fourth term of $\boldsymbol{S}_{\mathrm{B}i}$ is also $O_P(1)\mathbf{1}_{d_{\mathrm{B}}}$ and does not have a concise matrix algebraic expression. It is of lower order compared to the leading term of $\boldsymbol{S}_{\mathrm{B}i}$.

Overall Leading Term Expression for $\boldsymbol{S}_{Bi}$

Combining all four asymptotic approximations of the terms of $\boldsymbol{S}_{\mathrm{B}i}$ together, we have that

$$\boldsymbol{S}_{\mathrm{B}i}=\frac{1}{\phi}\left(\mathcal{G}_{\mathrm{B}i}-\mathcal{H}_{\mathrm{AB}i}^T\mathcal{H}_{\mathrm{AA}i}^{-1}\mathcal{G}_{\mathrm{A}i}\right)+O_P(1)\mathbf{1}_{d_{\mathrm{B}}}.$$

4.5.2.6 Computing an Asymptotic Approximation for the Third Entry in (4.7)

First note that,

$$\begin{aligned} d\log|\boldsymbol{\Sigma}| &= \frac{d|\boldsymbol{\Sigma}|}{|\boldsymbol{\Sigma}|} \\ &= \frac{|\boldsymbol{\Sigma}|\mathrm{tr}\left(\boldsymbol{\Sigma}^{-1}d\boldsymbol{\Sigma}\right)}{|\boldsymbol{\Sigma}} \\ &= \mathrm{vec}\left(\boldsymbol{\Sigma}^{-1}\right)^T \mathrm{vec}\left(d\boldsymbol{\Sigma}\right) \\ &= \mathrm{vec}\left(\boldsymbol{\Sigma}^{-1}\right)^T d\mathrm{vec}\left(\boldsymbol{\Sigma}\right) \\ &= \mathrm{vec}\left(\boldsymbol{\Sigma}^{-1}\right)^T d\boldsymbol{D}_{d_\mathrm{A}}\mathrm{vech}\left(\boldsymbol{\Sigma}\right) \\ &= \mathrm{vec}\left(\boldsymbol{\Sigma}^{-1}\right)^T \boldsymbol{D}_{d_\mathrm{A}} d\mathrm{vech}\left(\boldsymbol{\Sigma}\right). \end{aligned}$$

Next note that,

$$\begin{aligned} d\boldsymbol{u}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{u} &= -\boldsymbol{u}^T\boldsymbol{\Sigma}^{-1}\left(d\boldsymbol{\Sigma}\right)\boldsymbol{\Sigma}^{-1}\boldsymbol{u} \\ &= -\mathrm{tr}\left\{\boldsymbol{\Sigma}^{-1}\boldsymbol{u}\boldsymbol{u}^T\boldsymbol{\Sigma}^{-1}(d\boldsymbol{\Sigma})\right\} \\ &= -\mathrm{vec}\left(\boldsymbol{\Sigma}^{-1}\boldsymbol{u}\boldsymbol{u}^T\boldsymbol{\Sigma}^{-1}\right)^T \mathrm{vec}\left(d\boldsymbol{\Sigma}\right) \\ &= -\left\{\left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right)\mathrm{vec}\left(\boldsymbol{u}\boldsymbol{u}^T\right)\right\}^T \boldsymbol{D}_{d_\mathrm{A}} d\mathrm{vech}(\boldsymbol{\Sigma}). \end{aligned}$$

With further calculations, the ith contribution to the score of $\mathrm{vech}(\boldsymbol{\Sigma})$ can be computed as

$$\begin{aligned} \boldsymbol{S}_{\mathrm{C}i} &= \nabla_{\mathrm{vech}(\boldsymbol{\Sigma})}\, \log p_{\boldsymbol{Y}_i|\boldsymbol{X}_i}(\boldsymbol{Y}_i|\boldsymbol{X}_i) \\ &= -\frac{1}{2}\boldsymbol{D}_{d_\mathrm{A}}^T\mathrm{vec}(\boldsymbol{\Sigma}^{-1}) + \frac{1}{2}\,\frac{\int_{\mathbb{R}^{d_\mathrm{A}}} b_N(\boldsymbol{u})\exp\{-nh_N(\boldsymbol{u})\}\,d\boldsymbol{u}}{\int_{\mathbb{R}^{d_\mathrm{A}}} b_D(\boldsymbol{u})\exp\{-nh_N(\boldsymbol{u})\}\,d\boldsymbol{u}} \end{aligned}$$

where

$$\begin{aligned} b_N(\boldsymbol{u}) &\equiv \exp\Big(-\tfrac{1}{2}\boldsymbol{u}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{u}\Big)\boldsymbol{D}_{d_\mathrm{A}}^T(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1})\,\mathrm{vec}(\boldsymbol{u}\boldsymbol{u}^T), \\ b_D(\boldsymbol{u}) &\equiv \exp\Big(-\tfrac{1}{2}\boldsymbol{u}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{u}\Big) \quad \text{and} \\ h_N(\boldsymbol{u}) &\equiv -\frac{1}{n\phi}\sum_{j=1}^{n}\Big\{Y_{ij}\boldsymbol{u}^T\boldsymbol{X}_{\mathrm{A}ij} - b\big((\boldsymbol{\beta}_\mathrm{A}+\boldsymbol{u})^T\boldsymbol{X}_{\mathrm{A}ij} + (\boldsymbol{\beta}_\mathrm{B})^T\boldsymbol{X}_{\mathrm{B}ij}\big)\Big\}. \end{aligned}$$

The First Term of the Integral Ratio Component of $\boldsymbol{S}_{Ci}$

For each $1 \le k_C \le d_\mathrm{A}^{\boxplus}$, the k_Cth entry of the first term of the integral ratio component of $\boldsymbol{S}_{\mathrm{C}i}$ depends on the function

$$g(\boldsymbol{u}) = \frac{\boldsymbol{e}_{k_C}^T\boldsymbol{b}_N(\boldsymbol{u})}{b_D(\boldsymbol{u})} = \boldsymbol{e}_{k_C}^T\boldsymbol{D}_{d_\mathrm{A}}^T(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1})\mathrm{vec}(\boldsymbol{u}\boldsymbol{u}^T)$$

and is

$$\begin{aligned}
g(\boldsymbol{U}_i^*) &= \boldsymbol{e}_{k_C}^T \boldsymbol{D}_{d_\mathrm{A}}^T (\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1}) \mathrm{vec}\left\{ \boldsymbol{U}_i^* (\boldsymbol{U}_i^*)^T \right\} \\
&= \boldsymbol{e}_{k_C}^T \boldsymbol{D}_{d_\mathrm{A}}^T (\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1}) \mathrm{vec}\Big[\left\{ \boldsymbol{U}_i + \mathcal{H}_{\mathrm{AA}i}^{-1} \mathcal{G}_{\mathrm{A}i} + O_P(n^{-1} \mathbf{1}_{d_\mathrm{A}}) \right\} \\
&\quad \times \left\{ \boldsymbol{U}_i + \mathcal{H}_{\mathrm{AA}i}^{-1} \mathcal{G}_{\mathrm{A}i} + O_P(n^{-1} \mathbf{1}_{d_\mathrm{A}}) \right\}^T \Big] \\
&= \boldsymbol{e}_{k_C}^T \boldsymbol{D}_{d_\mathrm{A}}^T (\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1}) \mathrm{vec}\left(\boldsymbol{U}_i \boldsymbol{U}_i^T + \boldsymbol{U}_i \mathcal{G}_{\mathrm{A}i}^T \mathcal{H}_{\mathrm{AA}i}^{-1} + \mathcal{H}_{\mathrm{AA}i}^{-1} \mathcal{G}_{\mathrm{A}i} \boldsymbol{U}_i^T \right) + O_P(n^{-1}).
\end{aligned}$$

Therefore, the first term of the integral ratio component of $\boldsymbol{S}_{\mathrm{C}i}$ is

$$\begin{aligned}
&\boldsymbol{D}_{d_\mathrm{A}}^T (\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1}) \mathrm{vec}\left(\boldsymbol{U}_i \boldsymbol{U}_i^T + \boldsymbol{U}_i \mathcal{G}_{\mathrm{A}i}^T \mathcal{H}_{\mathrm{AA}i}^{-1} + \mathcal{H}_{\mathrm{AA}i}^{-1} \mathcal{G}_{\mathrm{A}i} \boldsymbol{U}_i^T \right) + O_P(n^{-1}) \mathbf{1}_{d_\mathrm{A}^{\boxplus}} \\
&= \boldsymbol{D}_{d_\mathrm{A}}^T \mathrm{vec}\left\{ \boldsymbol{\Sigma}^{-1} \left(\boldsymbol{U}_i \boldsymbol{U}_i^T + \boldsymbol{U}_i \mathcal{G}_{\mathrm{A}i}^T \mathcal{H}_{\mathrm{AA}i}^{-1} + \mathcal{H}_{\mathrm{AA}i}^{-1} \mathcal{G}_{\mathrm{A}i} \boldsymbol{U}_i^T \right) \boldsymbol{\Sigma}^{-1} \right\} + O_P(n^{-1}) \mathbf{1}_{d_\mathrm{A}^{\boxplus}}.
\end{aligned}$$

The Second Term of the Integral Ratio Component of $\boldsymbol{S}_{Ci}$

For each $1 \leq k_C \leq d_\mathrm{A}^{\boxplus}$, the k_Cth entry of the second term of the integral ratio component of $\boldsymbol{S}_{\mathrm{C}i}$ depends on the following function. By making use of the first property in Theorem 12 of Magnus and Neudecker (1999), we have

$$\begin{aligned}
dg(\boldsymbol{u}) &= \boldsymbol{e}_{k_C}^T \boldsymbol{D}_{d_\mathrm{A}}^T (\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1}) d\mathrm{vec}\left(\boldsymbol{u}\boldsymbol{u}^T \right) \\
&= \boldsymbol{e}_{k_C}^T \boldsymbol{D}_{d_\mathrm{A}}^T (\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1}) \mathrm{vec}\left\{ (d\boldsymbol{u})\boldsymbol{u}^T + \boldsymbol{u}(d\boldsymbol{u})^T \right\} \\
&= \boldsymbol{e}_{k_C}^T \boldsymbol{D}_{d_\mathrm{A}}^T (\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1}) \mathrm{vec}\left\{ \boldsymbol{I}_{d_\mathrm{A}} (d\boldsymbol{u})\boldsymbol{u}^T \right\} + \boldsymbol{e}_{k_C}^T \boldsymbol{D}_{d_\mathrm{A}}^T (\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1}) \mathrm{vec}\left\{ \boldsymbol{u}(d\boldsymbol{u})^T \boldsymbol{I}_{d_\mathrm{A}} \right\} \\
&= \boldsymbol{e}_{k_C}^T \boldsymbol{D}_{d_\mathrm{A}}^T (\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1}) \left(\boldsymbol{u} \otimes \boldsymbol{I}_{d_\mathrm{A}} \right) \mathrm{vec}(d\boldsymbol{u}) \\
&\quad + \boldsymbol{e}_{k_C}^T \boldsymbol{D}_{d_\mathrm{A}}^T (\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1}) \left(\boldsymbol{I}_{d_\mathrm{A}} \otimes \boldsymbol{u} \right) \mathrm{vec}\left\{ (d\boldsymbol{u})^T \right\} \\
&= \boldsymbol{e}_{k_C}^T \boldsymbol{D}_{d_\mathrm{A}}^T (\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1}) \left\{ \left(\boldsymbol{u} \otimes \boldsymbol{I}_{d_\mathrm{A}} \right) + \left(\boldsymbol{I}_{d_\mathrm{A}} \otimes \boldsymbol{u} \right) \right\} d\boldsymbol{u} \\
&= \boldsymbol{e}_{k_C}^T \boldsymbol{D}_{d_\mathrm{A}}^T \left\{ (\boldsymbol{\Sigma}^{-1}\boldsymbol{u} \otimes \boldsymbol{\Sigma}^{-1}) + (\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1}\boldsymbol{u}) \right\} d\boldsymbol{u} \\
&= \boldsymbol{e}_{k_C}^T \left\{ \boldsymbol{D}_{d_\mathrm{A}}^T (\boldsymbol{\Sigma}^{-1}\boldsymbol{u} \otimes \boldsymbol{\Sigma}^{-1}) + \left(\boldsymbol{K}_{d_\mathrm{A}} \boldsymbol{D}_{d_\mathrm{A}} \right)^T (\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1}\boldsymbol{u}) \right\} d\boldsymbol{u} \\
&= \boldsymbol{e}_{k_C}^T \left\{ \boldsymbol{D}_{d_\mathrm{A}}^T (\boldsymbol{\Sigma}^{-1}\boldsymbol{u} \otimes \boldsymbol{\Sigma}^{-1}) + \boldsymbol{D}_{d_\mathrm{A}}^T \boldsymbol{K}_{d_\mathrm{A}}^T (\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1}\boldsymbol{u}) \right\} d\boldsymbol{u} \\
&= \boldsymbol{e}_{k_C}^T \left\{ \boldsymbol{D}_{d_\mathrm{A}}^T (\boldsymbol{\Sigma}^{-1}\boldsymbol{u} \otimes \boldsymbol{\Sigma}^{-1}) + \boldsymbol{D}_{d_\mathrm{A}}^T \boldsymbol{K}_{d_\mathrm{A}} (\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1}\boldsymbol{u}) \right\} d\boldsymbol{u} \\
&= \boldsymbol{e}_{k_C}^T \left\{ \boldsymbol{D}_{d_\mathrm{A}}^T (\boldsymbol{\Sigma}^{-1}\boldsymbol{u} \otimes \boldsymbol{\Sigma}^{-1}) + \boldsymbol{D}_{d_\mathrm{A}}^T (\boldsymbol{\Sigma}^{-1}\boldsymbol{u} \otimes \boldsymbol{\Sigma}^{-1}) \right\} d\boldsymbol{u} \\
&= 2\boldsymbol{e}_{k_C}^T \left\{ \boldsymbol{D}_{d_\mathrm{A}}^T (\boldsymbol{\Sigma}^{-1}\boldsymbol{u} \otimes \boldsymbol{\Sigma}^{-1}) \right\} d\boldsymbol{u}, \\
\nabla g(\boldsymbol{u}) &= 2(\boldsymbol{\Sigma}^{-1}\boldsymbol{u} \otimes \boldsymbol{\Sigma}^{-1})^T \boldsymbol{D}_{d_\mathrm{A}} \boldsymbol{e}_{k_C}
\end{aligned}$$

and

$$\nabla b_D(\boldsymbol{u}) = -b_D(\boldsymbol{u}) \boldsymbol{\Sigma}^{-1} \boldsymbol{u}.$$

It also depends on

$$\nabla^2 h_N(\boldsymbol{u}) = \frac{1}{n\phi} \sum_{j=1}^{n} \left\{ b'' \left((\boldsymbol{\beta}_\mathrm{A} + \boldsymbol{u})^T \boldsymbol{X}_{\mathrm{A}ij} + (\boldsymbol{\beta}_\mathrm{B})^T \boldsymbol{X}_{\mathrm{B}ij} \right) \right\} \boldsymbol{X}_{\mathrm{A}ij} \boldsymbol{X}_{\mathrm{A}ij}^T.$$

By using a stochastic Taylor series approximation, note that

$$\{\nabla^2 h_N(\boldsymbol{U}_i^*)\}^{-1} = n\phi\mathcal{H}_{\mathrm{AA}i}^{-1} + O_P(n^{-1/2})\mathbf{1}_{d_\mathrm{A}}^{\otimes 2}.$$

Hence, for each $1 \leq k_C \leq d_\mathrm{A}^{\boxplus}$, the k_Cth entry of the second term of the integral component of $\boldsymbol{S}_{\mathrm{C}i}$ is

$$\begin{aligned}
&\frac{\nabla g(\boldsymbol{U}_i^*)^T\{\nabla^2 h_N(\boldsymbol{U}_i^*)\}^{-1}\nabla b_D(\boldsymbol{U}_i^*)}{nb_D(\boldsymbol{U}_i^*)} \\
&= \frac{-2\boldsymbol{e}_{k_C}^T\boldsymbol{D}_{d_\mathrm{A}}^T(\boldsymbol{\Sigma}^{-1}\boldsymbol{U}_i^*\otimes\boldsymbol{\Sigma}^{-1})\left\{n\phi\mathcal{H}_{\mathrm{AA}i}^{-1} + O_P(n^{-1/2})\mathbf{1}_{d_\mathrm{A}}^{\otimes 2}\right\} b_D(\boldsymbol{U}_i^*)\boldsymbol{\Sigma}^{-1}\boldsymbol{U}_i^*}{nb_D(\boldsymbol{U}_i^*)} \\
&= -2\phi\boldsymbol{e}_{k_C}^T\boldsymbol{D}_{d_\mathrm{A}}^T\left[\boldsymbol{\Sigma}^{-1}\left\{\boldsymbol{U}_i + \mathcal{H}_{\mathrm{AA}i}^{-1}\mathcal{G}_{\mathrm{A}i} + O_P(n^{-1}\mathbf{1}_{d_\mathrm{A}})\right\}\otimes\boldsymbol{\Sigma}^{-1}\right]\left\{\mathcal{H}_{\mathrm{AA}i}^{-1} + O_P(n^{-3/2})\mathbf{1}_{d_\mathrm{A}}^{\otimes 2}\right\}\boldsymbol{\Sigma}^{-1} \\
&\quad\times\left\{\boldsymbol{U}_i + \mathcal{H}_{\mathrm{AA}i}^{-1}\mathcal{G}_{\mathrm{A}i} + O_P(n^{-1}\mathbf{1}_{d_\mathrm{A}})\right\} \\
&= -2\phi\boldsymbol{e}_{k_C}^T\boldsymbol{D}_{d_\mathrm{A}}^T\left(\boldsymbol{\Sigma}^{-1}\boldsymbol{U}_i\otimes\boldsymbol{\Sigma}^{-1}\right)\mathcal{H}_{\mathrm{AA}i}^{-1}\boldsymbol{\Sigma}^{-1}\boldsymbol{U}_i + O_P(n^{-3/2}) \\
&= -2\phi\boldsymbol{e}_{k_C}^T\boldsymbol{D}_{d_\mathrm{A}}^T\left(\boldsymbol{\Sigma}^{-1}\boldsymbol{U}_i\otimes\boldsymbol{\Sigma}^{-1}\right)\mathrm{vec}\left(\mathcal{H}_{\mathrm{AA}i}^{-1}\boldsymbol{\Sigma}^{-1}\boldsymbol{U}_i\right) + O_P(n^{-3/2}) \\
&= -2\phi\boldsymbol{e}_{k_C}^T\boldsymbol{D}_{d_\mathrm{A}}^T\mathrm{vec}\left(\boldsymbol{\Sigma}^{-1}\mathcal{H}_{\mathrm{AA}i}^{-1}\boldsymbol{\Sigma}^{-1}\boldsymbol{U}_i\boldsymbol{U}_i^T\boldsymbol{\Sigma}^{-1}\right) + O_P(n^{-3/2}).
\end{aligned}$$

Then the leading term behaviour of the second term of the integral component of $\boldsymbol{S}_{\mathrm{C}i}$ is as follows

$$-2\phi\boldsymbol{D}_{d_\mathrm{A}}^T\mathrm{vec}\left(\boldsymbol{\Sigma}^{-1}\mathcal{H}_{\mathrm{AA}i}^{-1}\boldsymbol{\Sigma}^{-1}\boldsymbol{U}_i\boldsymbol{U}_i^T\boldsymbol{\Sigma}^{-1}\right) + O_P(n^{-3/2})\mathbf{1}_{d_\mathrm{A}^{\boxplus}} = O_P(n^{-1})\mathbf{1}_{d_\mathrm{A}^{\boxplus}}.$$

The Third Term of the Integral Ratio Component of $\boldsymbol{S}_{Ci}$

The leading term behaviour of the third term of the integral component of $\boldsymbol{S}_{\mathrm{C}i}$ is

$$\phi\boldsymbol{D}_{d_\mathrm{A}}^T\mathrm{vec}\left(\boldsymbol{\Sigma}^{-1}\mathcal{H}_{\mathrm{AA}i}^{-1}\boldsymbol{\Sigma}^{-1}\right) + O_P(n^{-3/2})\mathbf{1}_{d_\mathrm{A}^{\boxplus}} = O_P(n^{-1})\mathbf{1}_{d_\mathrm{A}^{\boxplus}}.$$

The Fourth Term of the Integral Ratio Component of $\boldsymbol{S}_{Ci}$

The contribution from the fourth term of $\boldsymbol{S}_{\mathrm{B}i}$ is also $O_P(n^{-1})\mathbf{1}_{d_\mathrm{A}^{\boxplus}}$ but does not have a concise matrix algebraic expression. It is of lower order compared to the leading term of the integral ratio component of $\boldsymbol{S}_{\mathrm{C}i}$.

Overall Leading Term Expression for $\boldsymbol{S}_{Ci}$

Combining all four asymptotic approximations of the terms of the integral ratio component of $\boldsymbol{S}_{\mathrm{C}i}$ together, we have that

$$\begin{aligned}
\boldsymbol{S}_{\mathrm{C}i} = \frac{1}{2}\boldsymbol{D}_{d_\mathrm{A}}^T&\left[\mathrm{vec}\left\{\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{U}_i\boldsymbol{U}_i^T + \boldsymbol{U}_i\mathcal{G}_{\mathrm{A}i}^T\mathcal{H}_{\mathrm{AA}i}^{-1} + \mathcal{H}_{\mathrm{AA}i}^{-1}\mathcal{G}_{\mathrm{A}i}\boldsymbol{U}_i^T\right)\boldsymbol{\Sigma}^{-1}\right\} - \mathrm{vec}\left(\boldsymbol{\Sigma}^{-1}\right)\right] \\
&+ O_P(n^{-1})\mathbf{1}_{d_\mathrm{A}^{\boxplus}}.
\end{aligned}$$

4.5.2.7 The Quadratic Conditional Expectations of the Scores

In this subsubsection we find the conditional expectations required to compute the Fisher information matrix of $(\boldsymbol{\beta}_{\mathrm{A}}, \boldsymbol{\beta}_{\mathrm{B}}, \mathrm{vech}(\boldsymbol{\Sigma}))$.

The Expectation of $\boldsymbol{S}_{Ai}\boldsymbol{S}_{Ai}^T$ Given $\boldsymbol{X}_i$

From the previous sections, we have the following approximation

$$\boldsymbol{S}_{\mathrm{A}i} = \boldsymbol{\Sigma}^{-1}\left(\boldsymbol{U}_i + \mathcal{H}_{\mathrm{AA}i}^{-1}\mathcal{G}_{\mathrm{A}i}\right) + O_P(n^{-1})\mathbf{1}_{d_{\mathrm{A}}}.$$

Therefore,

$$\begin{aligned}
&E\left(\boldsymbol{S}_{\mathrm{A}i}\boldsymbol{S}_{\mathrm{A}i}^T|\boldsymbol{X}_i\right)\\
&= E\left[\left\{\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{U}_i + \mathcal{H}_{\mathrm{AA}i}^{-1}\mathcal{G}_{\mathrm{A}i}\right) + O_P(n^{-1})\mathbf{1}_{d_{\mathrm{A}}}\right\}\left\{\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{U}_i + \mathcal{H}_{\mathrm{AA}i}^{-1}\mathcal{G}_{\mathrm{A}i}\right) + O_P(n^{-1})\mathbf{1}_{d_{\mathrm{A}}}\right\}^T|\boldsymbol{X}_i\right]\\
&= \boldsymbol{\Sigma}^{-1}E\left(\boldsymbol{U}_i\boldsymbol{U}_i^T\right)\boldsymbol{\Sigma}^{-1} + E\left(\boldsymbol{\Sigma}^{-1}\boldsymbol{U}_i\mathcal{G}_{\mathrm{A}i}^T\mathcal{H}_{\mathrm{AA}i}^{-1}\boldsymbol{\Sigma}^{-1}|\boldsymbol{X}_i\right) + E\left(\boldsymbol{\Sigma}^{-1}\mathcal{H}_{\mathrm{AA}i}^{-1}\mathcal{G}_{\mathrm{A}i}\boldsymbol{U}_i^T\boldsymbol{\Sigma}^{-1}|\boldsymbol{X}_i\right)\\
&\quad + E\left(\boldsymbol{\Sigma}^{-1}\mathcal{H}_{\mathrm{AA}i}^{-1}\mathcal{G}_{\mathrm{A}i}\mathcal{G}_{\mathrm{A}i}^T\mathcal{H}_{\mathrm{AA}i}^{-1}\boldsymbol{\Sigma}^{-1}|\boldsymbol{X}_i\right) + O_P(n^{-1})\mathbf{1}_{d_{\mathrm{A}}}^{\otimes 2}\\
&= \boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}\boldsymbol{\Sigma}^{-1} + E\left\{\boldsymbol{\Sigma}^{-1}\boldsymbol{U}_iE\left(\mathcal{G}_{\mathrm{A}i}^T|\boldsymbol{X}_i,\boldsymbol{U}_i\right)\mathcal{H}_{\mathrm{AA}i}^{-1}\boldsymbol{\Sigma}^{-1}|\boldsymbol{X}_i\right\}\\
&\quad + E\left\{\boldsymbol{\Sigma}^{-1}\mathcal{H}_{\mathrm{AA}i}^{-1}E\left(\mathcal{G}_{\mathrm{A}i}|\boldsymbol{X}_i,\boldsymbol{U}_i\right)\boldsymbol{U}_i^T\boldsymbol{\Sigma}^{-1}|\boldsymbol{X}_i\right\}\\
&\quad + E\left\{\boldsymbol{\Sigma}^{-1}\mathcal{H}_{\mathrm{AA}i}^{-1}E\left(\mathcal{G}_{\mathrm{A}i}\mathcal{G}_{\mathrm{A}i}^T|\boldsymbol{X}_I,\boldsymbol{U}_i\right)\mathcal{H}_{\mathrm{AA}i}^{-1}\boldsymbol{\Sigma}^{-1}|\boldsymbol{X}_i\right\} + O_P(n^{-1})\mathbf{1}_{d_{\mathrm{A}}}^{\otimes 2}\\
&= \boldsymbol{\Sigma}^{-1} + \phi E\left\{\boldsymbol{\Sigma}^{-1}\mathcal{H}_{\mathrm{AA}i}^{-1}\boldsymbol{\Sigma}^{-1}|\boldsymbol{X}_i\right\} + O_P(n^{-1})\mathbf{1}_{d_{\mathrm{A}}}^{\otimes 2}\\
&= \boldsymbol{\Sigma}^{-1} + O_P(n^{-1})\mathbf{1}_{d_{\mathrm{A}}}^{\otimes 2}.
\end{aligned}$$

The Expectation of $\boldsymbol{S}_{Bi}\boldsymbol{S}_{Bi}^T$ Given $\boldsymbol{X}_i$

From the previous sections, we have the following approximation

$$\boldsymbol{S}_{\mathrm{B}i} = \frac{1}{\phi}\left(\mathcal{G}_{\mathrm{B}i} - \mathcal{H}_{\mathrm{AB}i}^T\mathcal{H}_{\mathrm{AA}i}^{-1}\mathcal{G}_{\mathrm{A}i}\right) + O_P(1)\mathbf{1}_{d_{\mathrm{B}}}.$$

Therefore,

$$\begin{aligned}
&E\left(\boldsymbol{S}_{\mathrm{B}i}\boldsymbol{S}_{\mathrm{B}i}^T|\boldsymbol{X}_i\right)\\
&= E\left[\left\{\frac{1}{\phi}\left(\mathcal{G}_{\mathrm{B}i} - \mathcal{H}_{\mathrm{AB}i}^T\mathcal{H}_{\mathrm{AA}i}^{-1}\mathcal{G}_{\mathrm{A}i}\right) + O_P(1)\mathbf{1}_{d_{\mathrm{B}}}\right\}\left\{\frac{1}{\phi}\left(\mathcal{G}_{\mathrm{B}i} - \mathcal{H}_{\mathrm{AB}i}^T\mathcal{H}_{\mathrm{AA}i}^{-1}\mathcal{G}_{\mathrm{A}i}\right) + O_P(1)\mathbf{1}_{d_{\mathrm{B}}}\right\}^T|\boldsymbol{X}_i\right]\\
&= \frac{1}{\phi^2}\Big\{E\left(\mathcal{G}_{\mathrm{B}i}\mathcal{G}_{\mathrm{B}i}^T|\boldsymbol{X}_i\right) - E\left(\mathcal{G}_{\mathrm{B}i}\mathcal{G}_{\mathrm{A}i}^T\mathcal{H}_{\mathrm{AA}i}^{-1}\mathcal{H}_{\mathrm{AB}i}|\boldsymbol{X}_i\right) - E(\mathcal{H}_{\mathrm{AB}i}^T\mathcal{H}_{\mathrm{AA}i}^{-1}\mathcal{G}_{\mathrm{A}i}\mathcal{G}_{\mathrm{B}i}^T|\boldsymbol{X}_i)\\
&\quad + E\left(\mathcal{H}_{\mathrm{AB}i}^T\mathcal{H}_{\mathrm{AA}i}^{-1}\mathcal{G}_{\mathrm{A}i}\mathcal{G}_{\mathrm{A}i}^T\mathcal{H}_{\mathrm{AA}i}^{-1}\mathcal{H}_{\mathrm{AB}i}|\boldsymbol{X}_i\right)\Big\} + O_P(1)\mathbf{1}_{d_{\mathrm{B}}}^{\otimes 2}\\
&= \frac{1}{\phi^2}\Big[E\left\{E\left(\mathcal{G}_{\mathrm{B}i}\mathcal{G}_{\mathrm{B}i}^T|\boldsymbol{X}_i,\boldsymbol{U}_i\right)|\boldsymbol{X}_i\right\} - E\left\{E\left(\mathcal{G}_{\mathrm{B}i}\mathcal{G}_{\mathrm{A}i}^T|\boldsymbol{X}_i,\boldsymbol{U}_i\right)\mathcal{H}_{\mathrm{AA}i}^{-1}\mathcal{H}_{\mathrm{AB}i}|\boldsymbol{X}_i\right\}\\
&\quad - E\left\{\mathcal{H}_{\mathrm{AB}i}^T\mathcal{H}_{\mathrm{AA}i}^{-1}E(\mathcal{G}_{\mathrm{A}i}\mathcal{G}_{\mathrm{B}i}^T|\boldsymbol{X}_i,\boldsymbol{U}_i)|\boldsymbol{X}_i\right\}\\
&\quad + E\left\{\mathcal{H}_{\mathrm{AB}i}^T\mathcal{H}_{\mathrm{AA}i}^{-1}E\left(\mathcal{G}_{\mathrm{A}i}\mathcal{G}_{\mathrm{A}i}^T|\boldsymbol{X}_i,\boldsymbol{U}_i\right)\mathcal{H}_{\mathrm{AA}i}^{-1}\mathcal{H}_{\mathrm{AB}i}|\boldsymbol{X}_i\right\}\Big] + O_P(1)\mathbf{1}_{d_{\mathrm{B}}}^{\otimes 2}.
\end{aligned}$$

Simplifying the previous expression, we have,

$$\begin{aligned}
&\frac{1}{\phi^2}\Big\{\phi E\left(\mathcal{H}_{\mathrm{BB}i}|\boldsymbol{X}_i\right) - \phi E\left(\mathcal{H}_{\mathrm{AB}i}^T\mathcal{H}_{\mathrm{AA}i}^{-1}\mathcal{H}_{\mathrm{AB}i}|\boldsymbol{X}_i\right) - \phi E\left(\mathcal{H}_{\mathrm{AB}i}^T\mathcal{H}_{\mathrm{AA}i}^{-1}\mathcal{H}_{\mathrm{AB}i}|\boldsymbol{X}_i\right) \\
&\qquad + \phi E\left(\mathcal{H}_{\mathrm{AB}i}^T\mathcal{H}_{\mathrm{AA}i}^{-1}\mathcal{H}_{\mathrm{AB}i}|\boldsymbol{X}_i\right)\Big\} + O_P(1)\mathbf{1}_{d_{\mathrm{B}}}^{\otimes 2} \\
&= \frac{1}{\phi}E\left(\mathcal{H}_{\mathrm{BB}i} - \mathcal{H}_{\mathrm{AB}i}^T\mathcal{H}_{\mathrm{AA}i}^{-1}\mathcal{H}_{\mathrm{AB}i}|\boldsymbol{X}_i\right) + O_P(1)\mathbf{1}_{d_{\mathrm{B}}}^{\otimes 2}.
\end{aligned}$$

The Expectation of $\boldsymbol{S}_{Ci}\boldsymbol{S}_{Ci}^T$ Given $\boldsymbol{X}_i$

From the previous sections, we have the following approximation

$$\begin{aligned}
\boldsymbol{S}_{\mathrm{C}i} &= \frac{1}{2}\boldsymbol{D}_{d_{\mathrm{A}}}^T\left[\mathrm{vec}\left\{\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{U}_i\boldsymbol{U}_i^T + \boldsymbol{U}_i\mathcal{G}_{\mathrm{A}i}^T\mathcal{H}_{\mathrm{AA}i}^{-1} + \mathcal{H}_{\mathrm{AA}i}^{-1}\mathcal{G}_{\mathrm{A}i}\boldsymbol{U}_i^T\right)\boldsymbol{\Sigma}^{-1}\right\} - \mathrm{vec}\left(\boldsymbol{\Sigma}^{-1}\right)\right] \\
&\quad + O_P(n^{-1})\mathbf{1}_{d_{\mathrm{A}}^{\boxplus}} \\
&= \frac{1}{2}\boldsymbol{D}_{d_{\mathrm{A}}}^T\left\{\mathrm{vec}\left(\boldsymbol{\Sigma}^{-1}\boldsymbol{U}_i\boldsymbol{U}_i^T\boldsymbol{\Sigma}^{-1}\right) - \mathrm{vec}\left(\boldsymbol{\Sigma}^{-1}\right)\right\} \\
&\quad + \frac{1}{2}\boldsymbol{D}_{d_{\mathrm{A}}}^T\left[\mathrm{vec}\left\{\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{U}_i\mathcal{G}_{\mathrm{A}i}^T\mathcal{H}_{\mathrm{AA}i}^{-1} + \mathcal{H}_{\mathrm{AA}i}^{-1}\mathcal{G}_{\mathrm{A}i}\boldsymbol{U}_i^T\right)\boldsymbol{\Sigma}^{-1}\right\}\right] + O_P(n^{-1})\mathbf{1}_{d_{\mathrm{A}}^{\boxplus}} \\
&= \frac{1}{2}\boldsymbol{D}_{d_{\mathrm{A}}}^T\left\{\left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right)\mathrm{vec}\left(\boldsymbol{U}_i\boldsymbol{U}_i^T\right) - \mathrm{vec}\left(\boldsymbol{\Sigma}^{-1}\right)\right\} \\
&\quad + \frac{1}{2}\boldsymbol{D}_{d_{\mathrm{A}}}^T\left\{\left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right)\mathrm{vec}\left(\boldsymbol{U}_i\mathcal{G}_{\mathrm{A}i}^T\mathcal{H}_{\mathrm{AA}i}^{-1} + \mathcal{H}_{\mathrm{AA}i}^{-1}\mathcal{G}_{\mathrm{A}i}\boldsymbol{U}_i^T\right)\right\} + O_P(n^{-1})\mathbf{1}_{d_{\mathrm{A}}^{\boxplus}}.
\end{aligned}$$

We will deal with each term arising in $E\left(\boldsymbol{S}_{\mathrm{C}i}\boldsymbol{S}_{\mathrm{C}i}^T|\boldsymbol{X}_i\right)$ separately. The first term in $E\left(\boldsymbol{S}_{\mathrm{C}i}\boldsymbol{S}_{\mathrm{C}i}^T|\boldsymbol{X}_i\right)$ is

$$\frac{1}{4}\boldsymbol{D}_{d_{\mathrm{A}}}^T E\big[\{(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1})\mathrm{vec}(\boldsymbol{U}_i\boldsymbol{U}_i^T) - \mathrm{vec}(\boldsymbol{\Sigma}^{-1})\}\{(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1})\mathrm{vec}(\boldsymbol{U}_i\boldsymbol{U}_i^T) - \mathrm{vec}(\boldsymbol{\Sigma}^{-1})\}^T\big]\boldsymbol{D}_{d_{\mathrm{A}}}.$$

Next note that

$$\begin{aligned}
&E\left[\{(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1})\mathrm{vec}(\boldsymbol{U}_i\boldsymbol{U}_i^T) - \mathrm{vec}(\boldsymbol{\Sigma}^{-1})\}\{(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1})\mathrm{vec}(\boldsymbol{U}_i\boldsymbol{U}_i^T) - \mathrm{vec}(\boldsymbol{\Sigma}^{-1})\}^T\right] \\
&\quad = E\left\{(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1})\mathrm{vec}(\boldsymbol{U}_i\boldsymbol{U}_i^T)\mathrm{vec}(\boldsymbol{U}_i\boldsymbol{U}_i^T)^T(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1})\right\} \\
&\quad - E\left\{(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1})\mathrm{vec}(\boldsymbol{U}_i\boldsymbol{U}_i^T)\mathrm{vec}(\boldsymbol{\Sigma}^{-1})^T\right\} - E\left\{\mathrm{vec}(\boldsymbol{\Sigma}^{-1})\mathrm{vec}(\boldsymbol{U}_i\boldsymbol{U}_i^T)^T(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1})\right\} \\
&\quad + \mathrm{vec}(\boldsymbol{\Sigma}^{-1})\mathrm{vec}(\boldsymbol{\Sigma}^{-1})^T.
\end{aligned} \tag{4.9}$$

Then note that

$$E\{\mathrm{vec}(\boldsymbol{U}_i\boldsymbol{U}_i^T)\} = \mathrm{vec}\big(E(\boldsymbol{U}_i\boldsymbol{U}_i^T)\big) = \mathrm{vec}(\boldsymbol{\Sigma}).$$

With this we can simplify the following

$$\begin{aligned}
&- E\left\{(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1})\mathrm{vec}(\boldsymbol{U}_i\boldsymbol{U}_i^T)\mathrm{vec}(\boldsymbol{\Sigma}^{-1})^T\right\} \\
&= -(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1})\mathrm{vec}(\boldsymbol{\Sigma})\mathrm{vec}(\boldsymbol{\Sigma}^{-1})^T \\
&= -\mathrm{vec}(\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}\boldsymbol{\Sigma}^{-1})\mathrm{vec}(\boldsymbol{\Sigma}^{-1})^T \\
&= -\mathrm{vec}(\boldsymbol{\Sigma}^{-1})\mathrm{vec}(\boldsymbol{\Sigma}^{-1})^T
\end{aligned}$$

and

$$\begin{aligned}
&- E\left\{\text{vec}(\boldsymbol{\Sigma}^{-1})\text{vec}(\boldsymbol{U}_i\boldsymbol{U}_i^T)^T(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1})\right\}\\
&= -\text{vec}(\boldsymbol{\Sigma}^{-1})\text{vec}(\boldsymbol{\Sigma})^T(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1})\\
&= -\text{vec}\left(\boldsymbol{\Sigma}^{-1}\right)\text{vec}(\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}\boldsymbol{\Sigma}^{-1})^T\\
&= -\text{vec}(\boldsymbol{\Sigma}^{-1})\text{vec}(\boldsymbol{\Sigma}^{-1})^T.
\end{aligned}$$

These calculations lead to the right-hand side of (4.9) to simplify as

$$(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1})E\big\{\text{vec}(\boldsymbol{U}_i\boldsymbol{U}_i^T)\text{vec}(\boldsymbol{U}_i\boldsymbol{U}_i^T)^T\big\}(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}) - \text{vec}(\boldsymbol{\Sigma}^{-1})\text{vec}(\boldsymbol{\Sigma}^{-1})^T. \quad (4.10)$$

Next, we appeal to Theorem 4.3 (iv) Magnus and Neudecker (1979) to get

$$\begin{aligned}
E\big\{\text{vec}(\boldsymbol{U}_i\boldsymbol{U}_i^T)\text{vec}(\boldsymbol{U}_i\boldsymbol{U}_i^T)^T\big\} &= E\big\{(\boldsymbol{U}_i\otimes\boldsymbol{U}_i)(\boldsymbol{U}_i\otimes\boldsymbol{U}_i)^T\big\}\\
&= \text{Cov}(\boldsymbol{U}_i\otimes\boldsymbol{U}_i) + E(\boldsymbol{U}_i\otimes\boldsymbol{U}_i)\{E(\boldsymbol{U}_i\otimes\boldsymbol{U}_i)\}^T\\
&= (\boldsymbol{I}_{d_{\text{A}}^2} + \boldsymbol{K}_{d_{\text{A}}})(\boldsymbol{\Sigma}\otimes\boldsymbol{\Sigma}) + E\{\text{vec}(\boldsymbol{U}_i\boldsymbol{U}_i^T)\}E\{\text{vec}(\boldsymbol{U}_i\boldsymbol{U}_i^T)\}^T\\
&= (\boldsymbol{I}_{d_{\text{A}}^2} + \boldsymbol{K}_{d_{\text{A}}})(\boldsymbol{\Sigma}\otimes\boldsymbol{\Sigma}) + \text{vec}(\boldsymbol{\Sigma})\text{vec}(\boldsymbol{\Sigma})^T.
\end{aligned} \quad (4.11)$$

Substitution of (4.11) into (4.10) leads to

$$\begin{aligned}
&(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1})(\boldsymbol{I}_{d_{\text{A}}^2} + \boldsymbol{K}_{d_{\text{A}}})(\boldsymbol{\Sigma}\otimes\boldsymbol{\Sigma})(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1})\\
&\quad + (\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1})\text{vec}(\boldsymbol{\Sigma})\big\{(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1})\text{vec}(\boldsymbol{\Sigma})\big\}^T - \text{vec}(\boldsymbol{\Sigma}^{-1})\text{vec}(\boldsymbol{\Sigma}^{-1})^T\\
&= (\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1})(\boldsymbol{I}_{d_{\text{A}}^2} + \boldsymbol{K}_{d_{\text{A}}})(\boldsymbol{I}_d\otimes\boldsymbol{I}_d) + \text{vec}(\boldsymbol{\Sigma}^{-1})\text{vec}(\boldsymbol{\Sigma}^{-1})^T - \text{vec}(\boldsymbol{\Sigma}^{-1})\text{vec}(\boldsymbol{\Sigma}^{-1})^T\\
&= (\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1})(\boldsymbol{I}_{d_{\text{A}}^2} + \boldsymbol{K}_{d_{\text{A}}})(\boldsymbol{I}_d\otimes\boldsymbol{I}_d)\\
&= (\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}) + (\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1})\boldsymbol{K}_{d_{\text{A}}}(\boldsymbol{I}_d\otimes\boldsymbol{I}_d)\\
&= (\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}) + (\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1})(\boldsymbol{I}_d\otimes\boldsymbol{I}_d)\boldsymbol{K}_{d_{\text{A}}}\\
&= (\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}) + (\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1})\boldsymbol{K}_{d_{\text{A}}}\\
&= (\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}) + \boldsymbol{K}_{d_{\text{A}}}(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1})\\
&= (\boldsymbol{I}_{d_{\text{A}}^2} + \boldsymbol{K}_{d_{\text{A}}})(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}).
\end{aligned}$$

Therefore, we have the first term of $E(\boldsymbol{S}_{\text{C}i}\boldsymbol{S}_{\text{C}i}^T|\boldsymbol{X}_i)$ being equal to

$$\begin{aligned}
&\frac{1}{4}\boldsymbol{D}_{d_{\text{A}}}^T(\boldsymbol{I}_{d_{\text{A}}^2} + \boldsymbol{K}_{d_{\text{A}}})(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1})\boldsymbol{D}_{d_{\text{A}}}\\
&= \frac{1}{4}\boldsymbol{D}_{d_{\text{A}}}^T(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1})\boldsymbol{D}_{d_{\text{A}}} + \tfrac{1}{4}(\boldsymbol{K}_{d_{\text{A}}}\boldsymbol{D}_{d_{\text{A}}})^T(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1})\boldsymbol{D}_{d_{\text{A}}}\\
&= \frac{1}{2}\boldsymbol{D}_{d_{\text{A}}}^T(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1})\boldsymbol{D}_{d_{\text{A}}}.
\end{aligned}$$

We can then show that the second term of $E(\boldsymbol{S}_{\text{C}i}\boldsymbol{S}_{\text{C}i}^T|\boldsymbol{X}_i)$ is

$$\begin{aligned}
&\frac{1}{4}\boldsymbol{D}_{d_{\text{A}}}^T E\big[\{(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1})\text{vec}(\boldsymbol{U}_i\boldsymbol{U}_i^T) - \text{vec}(\boldsymbol{\Sigma}^{-1})\}\\
&\qquad\times\{(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1})\text{vec}(\boldsymbol{U}_i\mathcal{G}_{\text{A}i}^T\mathcal{H}_{\text{AA}i}^{-1} + \mathcal{H}_{\text{AA}i}^{-1}\mathcal{G}_{\text{A}i}\boldsymbol{U}_i^T)\}^T\big|\boldsymbol{X}_i]\boldsymbol{D}_{d_{\text{A}}}\\
&= \frac{1}{4}\boldsymbol{D}_{d_{\text{A}}}^T E\Bigg(\{(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1})\text{vec}(\boldsymbol{U}_i\boldsymbol{U}_i^T) - \text{vec}(\boldsymbol{\Sigma}^{-1})\}\\
&\qquad\times\left[(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1})\text{vec}\left\{\boldsymbol{U}_i\mathsf{E}\left(\mathcal{G}_{\text{A}i}^T|\boldsymbol{X}_i,\boldsymbol{U}_i\right)\mathcal{H}_{\text{AA}i}^{-1} + \mathcal{H}_{\text{AA}i}^{-1}E\left(\mathcal{G}_{\text{A}i}|\boldsymbol{X}_i,\boldsymbol{U}_i\right)\boldsymbol{U}_i^T\right\}\right]^T\big|\boldsymbol{X}_i\Bigg)\boldsymbol{D}_{d_{\text{A}}}\\
&= \boldsymbol{O}.
\end{aligned}$$

Similarly, we can then show that the third term of $E(\boldsymbol{S}_{\mathrm{C}i}\boldsymbol{S}_{\mathrm{C}i}^T|\boldsymbol{X}_i)$ is

$$\begin{aligned}
&\frac{1}{4}\boldsymbol{D}_{d_\mathrm{A}}^T E\big[\{(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1})\mathrm{vec}(\boldsymbol{U}_i\mathcal{G}_{\mathrm{A}i}^T\mathcal{H}_{\mathrm{AA}i}^{-1}+\mathcal{H}_{\mathrm{AA}i}^{-1}\mathcal{G}_{\mathrm{A}i}\boldsymbol{U}_i^T)\}\\
&\qquad\times\{(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1})\mathrm{vec}(\boldsymbol{U}_i\boldsymbol{U}_i^T)-\mathrm{vec}(\boldsymbol{\Sigma}^{-1})\}^T\big|\boldsymbol{X}_i\big]\boldsymbol{D}_{d_\mathrm{A}}\\
&=\frac{1}{4}\boldsymbol{D}_{d_\mathrm{A}}^T E\Bigg(\big[(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1})\mathrm{vec}\left\{\boldsymbol{U}_i\mathsf{E}\left(\mathcal{G}_{\mathrm{A}i}^T|\boldsymbol{X}_i,\boldsymbol{U}_i\right)\mathcal{H}_{\mathrm{AA}i}^{-1}+\mathcal{H}_{\mathrm{AA}i}^{-1}E\left(\mathcal{G}_{\mathrm{A}i}|\boldsymbol{X}_i,\boldsymbol{U}_i\right)\boldsymbol{U}_i^T\right\}\big]\\
&\qquad\times\{(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1})\mathrm{vec}(\boldsymbol{U}_i\boldsymbol{U}_i^T)-\mathrm{vec}(\boldsymbol{\Sigma}^{-1})\}^T\big|\boldsymbol{X}_i\Bigg)\boldsymbol{D}_{d_\mathrm{A}}\\
&=\boldsymbol{O}.
\end{aligned}$$

Lastly, the fourth term of $E(\boldsymbol{S}_{\mathrm{C}i}\boldsymbol{S}_{\mathrm{C}i}^T|\boldsymbol{X}_i)$ can be shown to simplify as follows

$$\begin{aligned}
&\frac{1}{4}\boldsymbol{D}_{d_\mathrm{A}}^T E\big[\{(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1})\mathrm{vec}(\boldsymbol{U}_i\mathcal{G}_{\mathrm{A}i}^T\mathcal{H}_{\mathrm{AA}i}^{-1}+\mathcal{H}_{\mathrm{AA}i}^{-1}\mathcal{G}_{\mathrm{A}i}\boldsymbol{U}_i^T)\}\\
&\qquad\times\{(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1})\mathrm{vec}(\boldsymbol{U}_i\mathcal{G}_{\mathrm{A}i}^T\mathcal{H}_{\mathrm{AA}i}^{-1}+\mathcal{H}_{\mathrm{AA}i}^{-1}\mathcal{G}_{\mathrm{A}i}\boldsymbol{U}_i^T)\}^T\big|\boldsymbol{X}_i\big]\boldsymbol{D}_{d_\mathrm{A}}\\
&=\frac{1}{4}\boldsymbol{D}_{d_\mathrm{A}}^T E\big[(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1})\{\mathrm{vec}(\boldsymbol{U}_i\mathcal{G}_{\mathrm{A}i}^T\mathcal{H}_{\mathrm{AA}i}^{-1})+\mathrm{vec}(\mathcal{H}_{\mathrm{AA}i}^{-1}\mathcal{G}_{\mathrm{A}i}\boldsymbol{U}_i^T)\}\\
&\qquad\times\{\mathrm{vec}(\boldsymbol{U}_i\mathcal{G}_{\mathrm{A}i}^T\mathcal{H}_{\mathrm{AA}i}^{-1})+\mathrm{vec}(\mathcal{H}_{\mathrm{AA}i}^{-1}\mathcal{G}_{\mathrm{A}i}\boldsymbol{U}_i^T)\}^T(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1})\big|\boldsymbol{X}_i\big]\boldsymbol{D}_{d_\mathrm{A}}\\
&=\frac{1}{4}\boldsymbol{D}_{d_\mathrm{A}}^T E\big[(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1})\{(\mathcal{H}_{\mathrm{AA}i}^{-1}\otimes\boldsymbol{U}_i)\mathrm{vec}(\mathcal{G}_{\mathrm{A}i}^T)+(\boldsymbol{U}_i\otimes\mathcal{H}_{\mathrm{AA}i}^{-1})\mathrm{vec}(\mathcal{G}_{\mathrm{A}i})\}\\
&\qquad\times\{(\mathcal{H}_{\mathrm{AA}i}^{-1}\otimes\boldsymbol{U}_i)\mathrm{vec}(\mathcal{G}_{\mathrm{A}i}^T)+(\boldsymbol{U}_i\otimes\mathcal{H}_{\mathrm{AA}i}^{-1})\mathrm{vec}(\mathcal{G}_{\mathrm{A}i})\}^T(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1})\big|\boldsymbol{X}_i\big]\boldsymbol{D}_{d_\mathrm{A}}\\
&=\frac{1}{4}\boldsymbol{D}_{d_\mathrm{A}}^T E\big[(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1})\{(\mathcal{H}_{\mathrm{AA}i}^{-1}\otimes\boldsymbol{U}_i)+(\boldsymbol{U}_i\otimes\mathcal{H}_{\mathrm{AA}i}^{-1})\}\mathcal{G}_{\mathrm{A}i}\mathcal{G}_{\mathrm{A}i}^T\\
&\qquad\times\{(\mathcal{H}_{\mathrm{AA}i}^{-1}\otimes\boldsymbol{U}_i)+(\boldsymbol{U}_i\otimes\mathcal{H}_{\mathrm{AA}i}^{-1})\}^T(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1})\big|\boldsymbol{X}_i\big]\boldsymbol{D}_{d_\mathrm{A}}\\
&=\frac{1}{4}\boldsymbol{D}_{d_\mathrm{A}}^T E\big[(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1})\{(\mathcal{H}_{\mathrm{AA}i}^{-1}\otimes\boldsymbol{U}_i)+(\boldsymbol{U}_i\otimes\mathcal{H}_{\mathrm{AA}i}^{-1})\}E(\mathcal{G}_{\mathrm{A}i}\mathcal{G}_{\mathrm{A}i}^T|\boldsymbol{X}_i,\boldsymbol{U}_i)\\
&\qquad\times\{(\mathcal{H}_{\mathrm{AA}i}^{-1}\otimes\boldsymbol{U}_i)+(\boldsymbol{U}_i\otimes\mathcal{H}_{\mathrm{AA}i}^{-1})\}^T(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1})\big|\boldsymbol{X}_i\big]\boldsymbol{D}_{d_\mathrm{A}}\\
&=\frac{\phi}{4}\boldsymbol{D}_{d_\mathrm{A}}^T E\big[(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1})\{(\mathcal{H}_{\mathrm{AA}i}^{-1}\otimes\boldsymbol{U}_i)+(\boldsymbol{U}_i\otimes\mathcal{H}_{\mathrm{AA}i}^{-1})\}\mathcal{H}_{\mathrm{AA}i}\\
&\qquad\times\{(\mathcal{H}_{\mathrm{AA}i}^{-1}\otimes\boldsymbol{U}_i)+(\boldsymbol{U}_i\otimes\mathcal{H}_{\mathrm{AA}i}^{-1})\}^T(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1})\big|\boldsymbol{X}_i\big]\boldsymbol{D}_{d_\mathrm{A}}\\
&=O_P(n^{-1})\mathbf{1}_{d_\mathrm{A}^{\boxplus}}\mathbf{1}_{d_\mathrm{A}^{\boxplus}}^T.
\end{aligned}$$

Putting together all the terms of $E(\boldsymbol{S}_{\mathrm{C}i}\boldsymbol{S}_{\mathrm{C}i}^T|\boldsymbol{X}_i)$, we have that

$$E(\boldsymbol{S}_{\mathrm{C}i}\boldsymbol{S}_{\mathrm{C}i}^T|\boldsymbol{X}_i)=\frac{1}{2}\boldsymbol{D}_{d_\mathrm{A}}^T(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1})\boldsymbol{D}_{d_\mathrm{A}}+O_P(n^{-1})\mathbf{1}_{d_\mathrm{A}^{\boxplus}}\mathbf{1}_{d_\mathrm{A}^{\boxplus}}^T.$$

The Expectation of $\boldsymbol{S}_{Ai}\boldsymbol{S}_{Bi}^T$ Given $\boldsymbol{X}_i$

From the previous sections, we have the following approximations

$$\begin{aligned}
\boldsymbol{S}_{\mathrm{A}i}&=\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{U}_i+\mathcal{H}_{\mathrm{AA}i}^{-1}\mathcal{G}_{\mathrm{A}i}\right)+O_P(n^{-1})\mathbf{1}_{d_\mathrm{A}},\\
\boldsymbol{S}_{\mathrm{B}i}&=\frac{1}{\phi}\left(\mathcal{G}_{\mathrm{B}i}-\mathcal{H}_{\mathrm{ABi}}^T\mathcal{H}_{\mathrm{AA}i}^{-1}\mathcal{G}_{\mathrm{A}i}\right)+O_P(1)\mathbf{1}_{d_\mathrm{B}}.
\end{aligned}$$

Therefore, by using the law of total expectation,

$$
\begin{aligned}
&E\left(\boldsymbol{S}_{\mathrm{A}i}\boldsymbol{S}_{\mathrm{B}i}^{T}\right)\\
&=E\left[\left\{\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{U}_i+\mathcal{H}_{\mathrm{AA}i}^{-1}\mathcal{G}_{\mathrm{A}i}\right)+O_P(n^{-1})\mathbf{1}_{d_{\mathrm{A}}}\right\}\left\{\frac{1}{\phi}\left(\mathcal{G}_{\mathrm{B}i}-\mathcal{H}_{\mathrm{AB}i}^{T}\mathcal{H}_{\mathrm{AA}i}^{-1}\mathcal{G}_{\mathrm{A}i}\right)+O_P(1)\mathbf{1}_{d_{\mathrm{B}}}\right\}^{T}|\boldsymbol{X}_i\right]\\
&=\frac{1}{\phi}\Bigg\{E(\boldsymbol{\Sigma}^{-1}\boldsymbol{U}_i\mathcal{G}_{\mathrm{B}i}^{T}|\boldsymbol{X}_i)+E(\boldsymbol{\Sigma}^{-1}\mathcal{H}_{\mathrm{AA}i}^{-1}\mathcal{G}_{\mathrm{A}i}\mathcal{G}_{\mathrm{B}i}^{T}|\boldsymbol{X}_i)-E\left(\boldsymbol{\Sigma}^{-1}\boldsymbol{U}_i\mathcal{G}_{\mathrm{A}i}^{T}\mathcal{H}_{\mathrm{AA}i}^{-1}\mathcal{H}_{\mathrm{AB}i}|\boldsymbol{X}_i\right)\\
&\quad-E\left(\boldsymbol{\Sigma}^{-1}\mathcal{H}_{\mathrm{AA}i}^{-1}\mathcal{G}_{\mathrm{A}i}\mathcal{G}_{\mathrm{A}i}^{T}\mathcal{H}_{\mathrm{AA}i}^{-1}\mathcal{H}_{\mathrm{AB}i}|\boldsymbol{X}_i\right)\Bigg\}+O_P(1)\mathbf{1}_{d_{\mathrm{A}}}\mathbf{1}_{d_{\mathrm{B}}}^{T}\\
&=\frac{1}{\phi}\left\{E\left(\boldsymbol{\Sigma}^{-1}\mathcal{H}_{\mathrm{AA}i}^{-1}\mathcal{H}_{\mathrm{AB}i}|\boldsymbol{X}_i\right)-E\left(\boldsymbol{\Sigma}^{-1}\mathcal{H}_{\mathrm{AA}i}^{-1}\mathcal{H}_{\mathrm{AB}i}|\boldsymbol{X}_i\right)\right\}+O_P(1)\mathbf{1}_{d_{\mathrm{A}}}\mathbf{1}_{d_{\mathrm{B}}}^{T}\\
&=O_P(1)\mathbf{1}_{d_{\mathrm{A}}}\mathbf{1}_{d_{\mathrm{B}}}^{T}
\end{aligned}
$$

The Expectation of $\boldsymbol{S}_{Ai}\boldsymbol{S}_{Ci}^{T}$ Given $\boldsymbol{X}_i$

From the previous sections, we have the following approximations where

$$\boldsymbol{S}_{\mathrm{A}i}=\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{U}_i+\mathcal{H}_{\mathrm{AA}i}^{-1}\mathcal{G}_{\mathrm{A}i}\right)+O_P(n^{-1})\mathbf{1}_{d_{\mathrm{A}}}$$

and

$$
\begin{aligned}
\boldsymbol{S}_{\mathrm{C}i}&=\frac{1}{2}\boldsymbol{D}_{d_{\mathrm{A}}}^{T}\left[\mathrm{vec}\left\{\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{U}_i\boldsymbol{U}_i^{T}+\boldsymbol{U}_i\mathcal{G}_{\mathrm{A}i}^{T}\mathcal{H}_{\mathrm{AA}i}^{-1}+\mathcal{H}_{\mathrm{AA}i}^{-1}\mathcal{G}_{\mathrm{A}i}\boldsymbol{U}_i^{T}\right)\boldsymbol{\Sigma}^{-1}\right\}-\mathrm{vec}\left(\boldsymbol{\Sigma}^{-1}\right)\right]\\
&\quad+O_P(n^{-1})\mathbf{1}_{d_{\mathrm{A}}^{\boxplus}}\\
&=\frac{1}{2}\boldsymbol{D}_{d_{\mathrm{A}}}^{T}\left\{\mathrm{vec}\left(\boldsymbol{\Sigma}^{-1}\boldsymbol{U}_i\boldsymbol{U}_i^{T}\boldsymbol{\Sigma}^{-1}\right)-\mathrm{vec}\left(\boldsymbol{\Sigma}^{-1}\right)\right\}\\
&\quad+\frac{1}{2}\boldsymbol{D}_{d_{\mathrm{A}}}^{T}\left[\mathrm{vec}\left\{\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{U}_i\mathcal{G}_{\mathrm{A}i}^{T}\mathcal{H}_{\mathrm{AA}i}^{-1}+\mathcal{H}_{\mathrm{AA}i}^{-1}\mathcal{G}_{\mathrm{A}i}\boldsymbol{U}_i^{T}\right)\boldsymbol{\Sigma}^{-1}\right\}\right]+O_P(n^{-1})\mathbf{1}_{d_{\mathrm{A}}^{\boxplus}}\\
&=\frac{1}{2}\boldsymbol{D}_{d_{\mathrm{A}}}^{T}\left\{\left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right)\mathrm{vec}\left(\boldsymbol{U}_i\boldsymbol{U}_i^{T}\right)-\mathrm{vec}\left(\boldsymbol{\Sigma}^{-1}\right)\right\}\\
&\quad+\frac{1}{2}\boldsymbol{D}_{d_{\mathrm{A}}}^{T}\left\{\left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right)\mathrm{vec}\left(\boldsymbol{U}_i\mathcal{G}_{\mathrm{A}i}^{T}\mathcal{H}_{\mathrm{AA}i}^{-1}+\mathcal{H}_{\mathrm{AA}i}^{-1}\mathcal{G}_{\mathrm{A}i}\boldsymbol{U}_i^{T}\right)\right\}+O_P(n^{-1})\mathbf{1}_{d_{\mathrm{A}}^{\boxplus}}.
\end{aligned}
$$

We will deal with each term arising in $E\left(\boldsymbol{S}_{\mathrm{A}i}\boldsymbol{S}_{\mathrm{C}i}^{T}|\boldsymbol{X}_i\right)$ separately. The first term in $E\left(\boldsymbol{S}_{\mathrm{A}i}\boldsymbol{S}_{\mathrm{C}i}^{T}|\boldsymbol{X}_i\right)$ can be simplified as follows

$$
\begin{aligned}
&\frac{1}{2}E\left[\boldsymbol{\Sigma}^{-1}\boldsymbol{U}_i\left\{(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1})\mathrm{vec}(\boldsymbol{U}_i\boldsymbol{U}_i^{T})-\mathrm{vec}(\boldsymbol{\Sigma}^{-1})\right\}^{T}\boldsymbol{D}_{d_{\mathrm{A}}}|\boldsymbol{X}_i\right]\\
&=\frac{1}{2}E\left\{\boldsymbol{\Sigma}^{-1}\boldsymbol{U}_i\mathrm{vec}(\boldsymbol{U}_i\boldsymbol{U}_i^{T})^{T}(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1})\boldsymbol{D}_{d_{\mathrm{A}}}-\boldsymbol{\Sigma}^{-1}\boldsymbol{U}_i\mathrm{vec}(\boldsymbol{\Sigma}^{-1})^{T}\boldsymbol{D}_{d_{\mathrm{A}}}|\boldsymbol{X}_i\right\}\\
&=\boldsymbol{O}.
\end{aligned}
$$

The second term in $E\left(\boldsymbol{S}_{\mathrm{A}i}\boldsymbol{S}_{\mathrm{C}i}^{T}|\boldsymbol{X}_i\right)$ can be shown to be

$$
\begin{aligned}
&\frac{1}{2}E\left[\boldsymbol{\Sigma}^{-1}\mathcal{H}_{\mathrm{AA}i}^{-1}\mathcal{G}_{\mathrm{A}i}\left\{(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1})\mathrm{vec}(\boldsymbol{U}_i\boldsymbol{U}_i^{T})-\mathrm{vec}(\boldsymbol{\Sigma}^{-1})\right\}^{T}\boldsymbol{D}_{d_{\mathrm{A}}}|\boldsymbol{X}_i\right]\\
&=\frac{1}{2}E\left[\boldsymbol{\Sigma}^{-1}\mathcal{H}_{\mathrm{AA}i}^{-1}E\left(\mathcal{G}_{\mathrm{A}i}|\boldsymbol{X}_i,\boldsymbol{U}_i\right)\left\{(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1})\mathrm{vec}(\boldsymbol{U}_i\boldsymbol{U}_i^{T})-\mathrm{vec}(\boldsymbol{\Sigma}^{-1})\right\}^{T}\boldsymbol{D}_{d_{\mathrm{A}}}|\boldsymbol{X}_i\right]\\
&=\boldsymbol{O}.
\end{aligned}
$$

Similarly, the third term in $E\left(\boldsymbol{S}_{\mathrm{A}i}\boldsymbol{S}_{\mathrm{C}i}^T|\boldsymbol{X}_i\right)$ can be shown to be

$$\begin{aligned}
&\frac{1}{2}E\left[\boldsymbol{\Sigma}^{-1}\boldsymbol{U}_i\left\{\mathrm{vec}(\boldsymbol{U}_i\mathcal{G}_{\mathrm{A}i}^T\mathcal{H}_{\mathrm{AA}i}^{-1}+\mathcal{H}_{\mathrm{AA}i}^{-1}\mathcal{G}_{\mathrm{A}i}\boldsymbol{U}_i^T)\right\}^T(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1})\boldsymbol{D}_{d_{\mathrm{A}}}|\boldsymbol{X}_i\right]\\
&=\frac{1}{2}E\Bigg(\boldsymbol{\Sigma}^{-1}\boldsymbol{U}_i\left[\mathrm{vec}\left\{\boldsymbol{U}_iE(\mathcal{G}_{\mathrm{A}i}^T|\boldsymbol{X}_i,\boldsymbol{U}_i)\mathcal{H}_{\mathrm{AA}i}^{-1}+\mathcal{H}_{\mathrm{AA}i}^{-1}E(\mathcal{G}_{\mathrm{A}i}|\boldsymbol{X}_i,\boldsymbol{U}_i)\boldsymbol{U}_i^T\right\}\right]^T\\
&\quad\times(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1})\boldsymbol{D}_{d_{\mathrm{A}}}|\boldsymbol{X}_i\Bigg)\\
&=\boldsymbol{O}.
\end{aligned}$$

Lastly, the fourth term in $E\left(\boldsymbol{S}_{\mathrm{A}i}\boldsymbol{S}_{\mathrm{C}i}^T|\boldsymbol{X}_i\right)$ simplifies as follows

$$\begin{aligned}
&\frac{1}{2}E\left[\boldsymbol{\Sigma}^{-1}\mathcal{H}_{\mathrm{AA}i}^{-1}\mathcal{G}_{\mathrm{A}i}\left\{\mathrm{vec}(\boldsymbol{U}_i\mathcal{G}_{\mathrm{A}i}^T\mathcal{H}_{\mathrm{AA}i}^{-1}+\mathcal{H}_{\mathrm{AA}i}^{-1}\mathcal{G}_{\mathrm{A}i}\boldsymbol{U}_i^T)\right\}^T(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1})\boldsymbol{D}_{d_{\mathrm{A}}}|\boldsymbol{X}_i\right]\\
&=\frac{1}{2}E\left[\boldsymbol{\Sigma}^{-1}\mathcal{H}_{\mathrm{AA}i}^{-1}\mathcal{G}_{\mathrm{A}i}\left\{\mathrm{vec}(\boldsymbol{U}_i\mathcal{G}_{\mathrm{A}i}^T\mathcal{H}_{\mathrm{AA}i}^{-1})+\mathrm{vec}(\mathcal{H}_{\mathrm{AA}i}^{-1}\mathcal{G}_{\mathrm{A}i}\boldsymbol{U}_i^T)\right\}^T(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1})\boldsymbol{D}_{d_{\mathrm{A}}}|\boldsymbol{X}_i\right]\\
&=\frac{1}{2}E\left[\boldsymbol{\Sigma}^{-1}\mathcal{H}_{\mathrm{AA}i}^{-1}\mathcal{G}_{\mathrm{A}i}\left\{(\mathcal{H}_{\mathrm{AA}i}^{-1}\otimes\boldsymbol{U}_i)\mathrm{vec}(\mathcal{G}_{\mathrm{A}i}^T)+(\boldsymbol{U}_i\otimes\mathcal{H}_{\mathrm{AA}i}^{-1})\mathrm{vec}(\mathcal{G}_{\mathrm{A}i})\right\}^T(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1})\boldsymbol{D}_{d_{\mathrm{A}}}|\boldsymbol{X}_i\right]\\
&=\frac{1}{2}E\left[\boldsymbol{\Sigma}^{-1}\mathcal{H}_{\mathrm{AA}i}^{-1}\mathcal{G}_{\mathrm{A}i}\mathcal{G}_{\mathrm{A}i}^T\left\{(\mathcal{H}_{\mathrm{AA}i}^{-1}\otimes\boldsymbol{U}_i)+(\boldsymbol{U}_i\otimes\mathcal{H}_{\mathrm{AA}i}^{-1})\right\}^T(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1})\boldsymbol{D}_{d_{\mathrm{A}}}|\boldsymbol{X}_i\right]\\
&=\frac{1}{2}E\left[\boldsymbol{\Sigma}^{-1}\mathcal{H}_{\mathrm{AA}i}^{-1}E(\mathcal{G}_{\mathrm{A}i}\mathcal{G}_{\mathrm{A}i}^T|\boldsymbol{X}_i,\boldsymbol{U}_i)\left\{(\mathcal{H}_{\mathrm{AA}i}^{-1}\otimes\boldsymbol{U}_i)+(\boldsymbol{U}_i\otimes\mathcal{H}_{\mathrm{AA}i}^{-1})\right\}^T(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1})\boldsymbol{D}_{d_{\mathrm{A}}}|\boldsymbol{X}_i\right]\\
&=\frac{\phi}{2}E\left[\boldsymbol{\Sigma}^{-1}\mathcal{H}_{\mathrm{AA}i}^{-1}\mathcal{H}_{\mathrm{AA}i}\left\{(\mathcal{H}_{\mathrm{AA}i}^{-1}\otimes\boldsymbol{U}_i)+(\boldsymbol{U}_i\otimes\mathcal{H}_{\mathrm{AA}i}^{-1})\right\}^T(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1})\boldsymbol{D}_{d_{\mathrm{A}}}|\boldsymbol{X}_i\right]\\
&=\frac{\phi}{2}E\left[\boldsymbol{\Sigma}^{-1}\left\{(\mathcal{H}_{\mathrm{AA}i}^{-1}\otimes\boldsymbol{U}_i)+(\boldsymbol{U}_i\otimes\mathcal{H}_{\mathrm{AA}i}^{-1})\right\}^T(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1})\boldsymbol{D}_{d_{\mathrm{A}}}|\boldsymbol{X}_i\right]\\
&=O_P(n^{-1})\mathbf{1}_{d_{\mathrm{A}}}\mathbf{1}_{d_{\mathrm{A}}^{\boxplus}}^T.
\end{aligned}$$

Putting together all the terms of $E(\boldsymbol{S}_{\mathrm{A}i}\boldsymbol{S}_{\mathrm{C}i}^T|\boldsymbol{X}_i)$, we have that

$$E(\boldsymbol{S}_{\mathrm{A}i}\boldsymbol{S}_{\mathrm{C}i}^T|\boldsymbol{X}_i)=O_P(n^{-1})\mathbf{1}_{d_{\mathrm{A}}}\mathbf{1}_{d_{\mathrm{A}}^{\boxplus}}^T.$$

The Expectation of $\boldsymbol{S}_{Bi}\boldsymbol{S}_{Ci}^T$ Given $\boldsymbol{X}_i$

From the previous sections, we have the following approximations

$$\boldsymbol{S}_{\mathrm{B}i}=\frac{1}{\phi}\left(\mathcal{G}_{\mathrm{B}i}-\mathcal{H}_{\mathrm{AB}i}^T\mathcal{H}_{\mathrm{AA}i}^{-1}\mathcal{G}_{\mathrm{A}i}\right)+O_P(1)\mathbf{1}_{d_{\mathrm{B}}}$$

and

$$\begin{aligned}
\boldsymbol{S}_{\mathrm{C}i}&=\frac{1}{2}\boldsymbol{D}_{d_{\mathrm{A}}}^T\left[\mathrm{vec}\left\{\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{U}_i\boldsymbol{U}_i^T+\boldsymbol{U}_i\mathcal{G}_{\mathrm{A}i}^T\mathcal{H}_{\mathrm{AA}i}^{-1}+\mathcal{H}_{\mathrm{AA}i}^{-1}\mathcal{G}_{\mathrm{A}i}\boldsymbol{U}_i^T\right)\boldsymbol{\Sigma}^{-1}\right\}-\mathrm{vec}\left(\boldsymbol{\Sigma}^{-1}\right)\right]\\
&\quad+O_P(n^{-1})\mathbf{1}_{d_{\mathrm{A}}^{\boxplus}}\\
&=\frac{1}{2}\boldsymbol{D}_{d_{\mathrm{A}}}^T\left\{\mathrm{vec}\left(\boldsymbol{\Sigma}^{-1}\boldsymbol{U}_i\boldsymbol{U}_i^T\boldsymbol{\Sigma}^{-1}\right)-\mathrm{vec}\left(\boldsymbol{\Sigma}^{-1}\right)\right\}\\
&\quad+\frac{1}{2}\boldsymbol{D}_{d_{\mathrm{A}}}^T\left[\mathrm{vec}\left\{\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{U}_i\mathcal{G}_{\mathrm{A}i}^T\mathcal{H}_{\mathrm{AA}i}^{-1}+\mathcal{H}_{\mathrm{AA}i}^{-1}\mathcal{G}_{\mathrm{A}i}\boldsymbol{U}_i^T\right)\boldsymbol{\Sigma}^{-1}\right\}\right]+O_P(n^{-1})\mathbf{1}_{d_{\mathrm{A}}^{\boxplus}}\\
&=\frac{1}{2}\boldsymbol{D}_{d_{\mathrm{A}}}^T\left\{\left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right)\mathrm{vec}\left(\boldsymbol{U}_i\boldsymbol{U}_i^T\right)-\mathrm{vec}\left(\boldsymbol{\Sigma}^{-1}\right)\right\}\\
&\quad+\frac{1}{2}\boldsymbol{D}_{d_{\mathrm{A}}}^T\left\{\left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right)\mathrm{vec}\left(\boldsymbol{U}_i\mathcal{G}_{\mathrm{A}i}^T\mathcal{H}_{\mathrm{AA}i}^{-1}+\mathcal{H}_{\mathrm{AA}i}^{-1}\mathcal{G}_{\mathrm{A}i}\boldsymbol{U}_i^T\right)\right\}+O_P(n^{-1})\mathbf{1}_{d_{\mathrm{A}}^{\boxplus}}.
\end{aligned}$$

We will deal with each term arising in $E\left(\boldsymbol{S}_{\mathrm{B}i}\boldsymbol{S}_{\mathrm{C}i}^T|\boldsymbol{X}_i\right)$ separately. The first term in $E\left(\boldsymbol{S}_{\mathrm{B}i}\boldsymbol{S}_{\mathrm{C}i}^T|\boldsymbol{X}_i\right)$ can be simplified as follows

$$\begin{aligned}
&\frac{1}{2\phi}E\left[\mathcal{G}_{\mathrm{B}i}\left\{(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1})\mathrm{vec}(\boldsymbol{U}_i\boldsymbol{U}_i^T)-\mathrm{vec}(\boldsymbol{\Sigma}^{-1})\right\}^T\boldsymbol{D}_{d_{\mathrm{A}}}|\boldsymbol{X}_i\right]\\
&=\frac{1}{2\phi}E\left[E(\mathcal{G}_{\mathrm{B}i}|\boldsymbol{X}_i,\boldsymbol{U}_i)\left\{(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1})\mathrm{vec}(\boldsymbol{U}_i\boldsymbol{U}_i^T)-\mathrm{vec}(\boldsymbol{\Sigma}^{-1})\right\}^T\boldsymbol{D}_{d_{\mathrm{A}}}|\boldsymbol{X}_i\right]\\
&=\boldsymbol{O}.
\end{aligned}$$

. The second term in $E\left(\boldsymbol{S}_{\mathrm{B}i}\boldsymbol{S}_{\mathrm{C}i}^T|\boldsymbol{X}_i\right)$ can be shown to be

$$\begin{aligned}
&-\frac{1}{2\phi}E\left[\mathcal{H}_{\mathrm{AB}i}^T\mathcal{H}_{\mathrm{AA}i}^{-1}\mathcal{G}_{\mathrm{A}i}\left\{(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1})\mathrm{vec}(\boldsymbol{U}_i\boldsymbol{U}_i^T)-\mathrm{vec}(\boldsymbol{\Sigma}^{-1})\right\}^T\boldsymbol{D}_{d_{\mathrm{A}}}|\boldsymbol{X}_i\right]\\
&=-\frac{1}{2\phi}E\left[\mathcal{H}_{\mathrm{AB}i}^T\mathcal{H}_{\mathrm{AA}i}^{-1}E\left(\mathcal{G}_{\mathrm{A}i}|\boldsymbol{X}_i,\boldsymbol{U}_i\right)\left\{(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1})\mathrm{vec}(\boldsymbol{U}_i\boldsymbol{U}_i^T)-\mathrm{vec}(\boldsymbol{\Sigma}^{-1})\right\}^T\boldsymbol{D}_{d_{\mathrm{A}}}|\boldsymbol{X}_i\right]\\
&=\boldsymbol{O}.
\end{aligned}$$

The third term in $E\left(\boldsymbol{S}_{\mathrm{B}i}\boldsymbol{S}_{\mathrm{C}i}^T|\boldsymbol{X}_i\right)$ can be shown to be

$$\begin{aligned}
&\frac{1}{2\phi}E\left[\mathcal{G}_{\mathrm{B}i}\left\{\mathrm{vec}(\boldsymbol{U}_i\mathcal{G}_{\mathrm{A}i}^T\mathcal{H}_{\mathrm{AA}i}^{-1}+\mathcal{H}_{\mathrm{AA}i}^{-1}\mathcal{G}_{\mathrm{A}i}\boldsymbol{U}_i^T)\right\}^T(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1})\boldsymbol{D}_{d_{\mathrm{A}}}|\boldsymbol{X}_i\right]\\
&=\frac{1}{2\phi}E\left[\mathcal{G}_{\mathrm{B}i}\left\{\mathrm{vec}(\boldsymbol{U}_i\mathcal{G}_{\mathrm{A}i}^T\mathcal{H}_{\mathrm{AA}i}^{-1})+\mathrm{vec}(\mathcal{H}_{\mathrm{AA}i}^{-1}\mathcal{G}_{\mathrm{A}i}\boldsymbol{U}_i^T)\right\}^T(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1})\boldsymbol{D}_{d_{\mathrm{A}}}|\boldsymbol{X}_i\right]\\
&=\frac{1}{2\phi}E\left[\mathcal{G}_{\mathrm{B}i}\left\{(\mathcal{H}_{\mathrm{AA}i}^{-1}\otimes\boldsymbol{U}_i)\mathrm{vec}(\mathcal{G}_{\mathrm{A}i}^T)+(\boldsymbol{U}_i\otimes\mathcal{H}_{\mathrm{AA}i}^{-1})\mathrm{vec}(\mathcal{G}_{\mathrm{A}i})\right\}^T(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1})\boldsymbol{D}_{d_{\mathrm{A}}}|\boldsymbol{X}_i\right]\\
&=\frac{1}{2\phi}E\left[\mathcal{G}_{\mathrm{B}i}\mathcal{G}_{\mathrm{A}i}^T\left\{(\mathcal{H}_{\mathrm{AA}i}^{-1}\otimes\boldsymbol{U}_i)+(\boldsymbol{U}_i\otimes\mathcal{H}_{\mathrm{AA}i}^{-1})\right\}^T(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1})\boldsymbol{D}_{d_{\mathrm{A}}}|\boldsymbol{X}_i\right]\\
&=\frac{1}{2\phi}E\left[E(\mathcal{G}_{\mathrm{B}i}\mathcal{G}_{\mathrm{A}i}^T|\boldsymbol{X}_i,\boldsymbol{U}_i)\left\{(\mathcal{H}_{\mathrm{AA}i}^{-1}\otimes\boldsymbol{U}_i)+(\boldsymbol{U}_i\otimes\mathcal{H}_{\mathrm{AA}i}^{-1})\right\}^T(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1})\boldsymbol{D}_{d_{\mathrm{A}}}|\boldsymbol{X}_i\right]\\
&=\frac{1}{2\phi^2}E\left[\mathcal{H}_{\mathrm{AB}i}^T\left\{(\mathcal{H}_{\mathrm{AA}i}^{-1}\otimes\boldsymbol{U}_i)+(\boldsymbol{U}_i\otimes\mathcal{H}_{\mathrm{AA}i}^{-1})\right\}^T(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1})\boldsymbol{D}_{d_{\mathrm{A}}}|\boldsymbol{X}_i\right]\\
&=O_P(1)\mathbf{1}_{d_{\mathrm{B}}}\mathbf{1}_{d_{\mathrm{A}}^{\boxplus}}^T.
\end{aligned}$$

Lastly, the fourth term in $E\left(\boldsymbol{S}_{\mathrm{B}i}\boldsymbol{S}_{\mathrm{C}i}^T|\boldsymbol{X}_i\right)$ simplifies as follows

$$\begin{aligned}
&-\frac{1}{2\phi}E\left[\mathcal{H}_{\mathrm{AB}i}^T\mathcal{H}_{\mathrm{AA}i}^{-1}\mathcal{G}_{\mathrm{A}i}\left\{\mathrm{vec}(\boldsymbol{U}_i\mathcal{G}_{\mathrm{A}i}^T\mathcal{H}_{\mathrm{AA}i}^{-1}+\mathcal{H}_{\mathrm{AA}i}^{-1}\mathcal{G}_{\mathrm{A}i}\boldsymbol{U}_i^T)\right\}^T(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1})\boldsymbol{D}_{d_{\mathrm{A}}}|\boldsymbol{X}_i\right]\\
&=-\frac{1}{2\phi}E\left[\mathcal{H}_{\mathrm{AB}i}^T\mathcal{H}_{\mathrm{AA}i}^{-1}\mathcal{G}_{\mathrm{A}i}\left\{\mathrm{vec}(\boldsymbol{U}_i\mathcal{G}_{\mathrm{A}i}^T\mathcal{H}_{\mathrm{AA}i}^{-1})+\mathrm{vec}(\mathcal{H}_{\mathrm{AA}i}^{-1}\mathcal{G}_{\mathrm{A}i}\boldsymbol{U}_i^T)\right\}^T(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1})\boldsymbol{D}_{d_{\mathrm{A}}}|\boldsymbol{X}_i\right]\\
&=-\frac{1}{2\phi}E\left[\mathcal{H}_{\mathrm{AB}i}^T\mathcal{H}_{\mathrm{AA}i}^{-1}\mathcal{G}_{\mathrm{A}i}\left\{(\mathcal{H}_{\mathrm{AA}i}^{-1}\otimes\boldsymbol{U}_i)\mathrm{vec}(\mathcal{G}_{\mathrm{A}i}^T)+(\boldsymbol{U}_i\otimes\mathcal{H}_{\mathrm{AA}i}^{-1})\mathrm{vec}(\mathcal{G}_{\mathrm{A}i})\right\}^T(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1})\boldsymbol{D}_{d_{\mathrm{A}}}|\boldsymbol{X}_i\right]\\
&=-\frac{1}{2\phi}E\left[\mathcal{H}_{\mathrm{AB}i}^T\mathcal{H}_{\mathrm{AA}i}^{-1}\mathcal{G}_{\mathrm{A}i}\mathcal{G}_{\mathrm{A}i}^T\left\{(\mathcal{H}_{\mathrm{AA}i}^{-1}\otimes\boldsymbol{U}_i)+(\boldsymbol{U}_i\otimes\mathcal{H}_{\mathrm{AA}i}^{-1})\right\}^T(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1})\boldsymbol{D}_{d_{\mathrm{A}}}|\boldsymbol{X}_i\right]\\
&=-\frac{1}{2\phi}E\left[\mathcal{H}_{\mathrm{AB}i}^T\mathcal{H}_{\mathrm{AA}i}^{-1}E(\mathcal{G}_{\mathrm{A}i}\mathcal{G}_{\mathrm{A}i}^T|\boldsymbol{X}_i,\boldsymbol{U}_i)\left\{(\mathcal{H}_{\mathrm{AA}i}^{-1}\otimes\boldsymbol{U}_i)+(\boldsymbol{U}_i\otimes\mathcal{H}_{\mathrm{AA}i}^{-1})\right\}^T(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1})\boldsymbol{D}_{d_{\mathrm{A}}}|\boldsymbol{X}_i\right]\\
&=-\frac{1}{2\phi^2}E\left[\mathcal{H}_{\mathrm{AB}i}^T\mathcal{H}_{\mathrm{AA}i}^{-1}\mathcal{H}_{\mathrm{AA}i}\left\{(\mathcal{H}_{\mathrm{AA}i}^{-1}\otimes\boldsymbol{U}_i)+(\boldsymbol{U}_i\otimes\mathcal{H}_{\mathrm{AA}i}^{-1})\right\}^T(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1})\boldsymbol{D}_{d_{\mathrm{A}}}|\boldsymbol{X}_i\right]\\
&=\frac{1}{2\phi^2}E\left[\mathcal{H}_{\mathrm{AB}i}^T\left\{(\mathcal{H}_{\mathrm{AA}i}^{-1}\otimes\boldsymbol{U}_i)+(\boldsymbol{U}_i\otimes\mathcal{H}_{\mathrm{AA}i}^{-1})\right\}^T(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1})\boldsymbol{D}_{d_{\mathrm{A}}}|\boldsymbol{X}_i\right]\\
&=O_P(1)\mathbf{1}_{d_{\mathrm{B}}}\mathbf{1}_{d_{\mathrm{A}}^{\boxplus}}^T.
\end{aligned}$$

Putting together all the terms of $E(\boldsymbol{S}_{\mathrm{B}i}\boldsymbol{S}_{\mathrm{C}i}^T|\boldsymbol{X}_i)$, we have that

$$E(\boldsymbol{S}_{\mathrm{A}i}\boldsymbol{S}_{\mathrm{C}i}^T|\boldsymbol{X}_i) = O_P(1)\mathbf{1}_{d_{\mathrm{B}}}\mathbf{1}_{d_{\mathrm{A}}^{\boxplus}}^T.$$

4.5.2.8 Treating the Leading Term of the (2,2)-Entry of the Fisher Information Matrix

Note that the leading term of $\sum_{i=1}^m E(\boldsymbol{S}_{\mathrm{B}i}\boldsymbol{S}_{\mathrm{B}i}^T|\boldsymbol{X}_i)$ is

$$\begin{aligned}
&\frac{1}{\phi}\sum_{i=1}^m E\left(\mathcal{H}_{\mathrm{BB}i} - \mathcal{H}_{\mathrm{AB}i}^T\mathcal{H}_{\mathrm{AA}i}^{-1}\mathcal{H}_{\mathrm{AB}i}|\boldsymbol{X}_i\right)\\
&= \frac{mn}{\phi}E\left\{\frac{1}{mn}\sum_{i=1}^m\left(\mathcal{H}_{\mathrm{BB}i} - \mathcal{H}_{\mathrm{AB}i}^T\mathcal{H}_{\mathrm{AA}i}^{-1}\mathcal{H}_{\mathrm{AB}i}|\boldsymbol{X}_i\right)\right\} \qquad (4.12)\\
&= \frac{mn}{\phi}E\left\{\frac{1}{mn}\sum_{i=1}^m\left(\mathcal{H}_{\mathrm{BB}i}|\boldsymbol{X}_i\right) - \frac{1}{mn}\sum_{i=1}^m\left(\mathcal{H}_{\mathrm{AB}i}^T\mathcal{H}_{\mathrm{AA}i}^{-1}\mathcal{H}_{\mathrm{AB}i}|\boldsymbol{X}_i\right)\right\}.
\end{aligned}$$

Using Lemma 1 from Chapter 2 with $f(\boldsymbol{X}_{ij},\boldsymbol{U}_i) = b''\big((\boldsymbol{\beta}_{\mathrm{A}}+\boldsymbol{U}_i)^T\boldsymbol{X}_{\mathrm{A}ij} + (\boldsymbol{\beta}_{\mathrm{B}})^T\boldsymbol{X}_{\mathrm{B}ij}\big)$, we have that the first term in (4.12) can be re-expressed as follows

$$\begin{aligned}
&\frac{mn}{\phi}E\left\{\frac{1}{mn}\sum_{i=1}^m\left(\mathcal{H}_{\mathrm{BB}i}|\boldsymbol{X}_i\right)\right\}\\
&= \frac{mn}{\phi}E\left[\frac{1}{mn}\sum_{i=1}^m\sum_{j=1}^n \boldsymbol{X}_{\mathrm{B}ij}\boldsymbol{X}_{\mathrm{B}ij}^T E\left\{b''\big((\boldsymbol{\beta}_{\mathrm{A}}+\boldsymbol{U}_i)^T\boldsymbol{X}_{\mathrm{A}ij} + (\boldsymbol{\beta}_{\mathrm{B}})^T\boldsymbol{X}_{\mathrm{B}ij}\big)|\boldsymbol{X}_i\right\}\right] \qquad (4.13)\\
&= \frac{mn}{\phi}E\left\{\boldsymbol{X}_{\mathrm{B}}\boldsymbol{X}_{\mathrm{B}}^T b''\big((\boldsymbol{\beta}_{\mathrm{A}}+\boldsymbol{U})^T\boldsymbol{X}_{\mathrm{A}} + \boldsymbol{\beta}_{\mathrm{B}}^T\boldsymbol{X}_{\mathrm{B}}\big)\right\} + o_P(mn)\mathbf{1}_{d_{\mathrm{B}}}^{\otimes 2}.
\end{aligned}$$

Now, using Lemma 2 from Chapter 2 with $f(\boldsymbol{X}_{ij},\boldsymbol{U}_i) = b''\big((\boldsymbol{\beta}_{\mathrm{A}}+\boldsymbol{U}_i)^T\boldsymbol{X}_{\mathrm{A}ij}+(\boldsymbol{\beta}_{\mathrm{B}})^T\boldsymbol{X}_{\mathrm{B}ij}\big)$, we have that the second term in (4.12) can be re-expressed as follows

$$\begin{aligned}
&\frac{mn}{\phi}E\left\{\frac{1}{mn}\sum_{i=1}^m\left(\mathcal{H}_{\mathrm{AB}i}^T\mathcal{H}_{\mathrm{AA}i}^{-1}\mathcal{H}_{\mathrm{AB}i}|\boldsymbol{X}_i\right)\right\}\\
&= \frac{mn}{\phi}E\Bigg[\frac{1}{mn}\sum_{i=1}^m\left\{\sum_{j=1}^n \boldsymbol{X}_{\mathrm{B}ij}\boldsymbol{X}_{\mathrm{A}ij}^T b''\big((\boldsymbol{\beta}_{\mathrm{A}}+\boldsymbol{U}_i)^T\boldsymbol{X}_{\mathrm{A}ij} + (\boldsymbol{\beta}_{\mathrm{B}})^T\boldsymbol{X}_{\mathrm{B}ij}\big)\right\}\\
&\qquad\times\left\{\sum_{j=1}^n \boldsymbol{X}_{\mathrm{A}ij}\boldsymbol{X}_{\mathrm{A}ij}^T b''\big((\boldsymbol{\beta}_{\mathrm{A}}+\boldsymbol{U}_i)^T\boldsymbol{X}_{\mathrm{A}ij} + (\boldsymbol{\beta}_{\mathrm{B}})^T\boldsymbol{X}_{\mathrm{B}ij}\big)\right\}^{-1}\\
&\qquad\times\left\{\sum_{j=1}^n \boldsymbol{X}_{\mathrm{B}ij}\boldsymbol{X}_{\mathrm{A}ij}^T b''\big((\boldsymbol{\beta}_{\mathrm{A}}+\boldsymbol{U}_i)^T\boldsymbol{X}_{\mathrm{A}ij} + (\boldsymbol{\beta}_{\mathrm{B}})^T\boldsymbol{X}_{\mathrm{B}ij}\big)\right\}^T\Bigg|\boldsymbol{X}_i\Bigg]
\end{aligned}$$

The previous expression simplifies to

$$\frac{mn}{\phi}E\Bigg(E\left\{\boldsymbol{X}_{\mathrm{B}}\boldsymbol{X}_{\mathrm{A}}^{T}b''\big((\boldsymbol{\beta}_{\mathrm{A}}+\boldsymbol{U})^{T}\boldsymbol{X}_{\mathrm{A}}+\boldsymbol{\beta}_{\mathrm{B}}^{T}\boldsymbol{X}_{\mathrm{B}}\big)\right\}\left[E\left\{\boldsymbol{X}_{\mathrm{A}}\boldsymbol{X}_{\mathrm{A}}^{T}b''\big((\boldsymbol{\beta}_{\mathrm{A}}+\boldsymbol{U})^{T}\boldsymbol{X}_{\mathrm{A}}+\boldsymbol{\beta}_{\mathrm{B}}^{T}\boldsymbol{X}_{\mathrm{B}}\big)\right\}\right]^{-1}$$
$$\times E\left\{\boldsymbol{X}_{\mathrm{B}}\boldsymbol{X}_{\mathrm{A}}^{T}b''\big((\boldsymbol{\beta}_{\mathrm{A}}+\boldsymbol{U})^{T}\boldsymbol{X}_{\mathrm{A}}+\boldsymbol{\beta}_{\mathrm{B}}^{T}\boldsymbol{X}_{\mathrm{B}}\big)\right\}^{T}\Bigg)+o_P(mn)\mathbf{1}_{d_{\mathrm{B}}}^{\otimes 2}. \tag{4.14}$$

Combining (4.13) and (4.5.2.8), we have

$$\frac{1}{\phi}\sum_{i=1}^{m}E\left(\mathcal{H}_{\mathrm{BB}i}-\mathcal{H}_{\mathrm{AB}i}^{T}\mathcal{H}_{\mathrm{AA}i}^{-1}\mathcal{H}_{\mathrm{AB}i}|\boldsymbol{X}_i\right)$$
$$=\frac{mn}{\phi}E\Bigg(E\left\{\boldsymbol{X}_{\mathrm{B}}\boldsymbol{X}_{\mathrm{B}}^{T}b''\big((\boldsymbol{\beta}_{\mathrm{A}}+\boldsymbol{U})^{T}\boldsymbol{X}_{\mathrm{A}}+\boldsymbol{\beta}_{\mathrm{B}}^{T}\boldsymbol{X}_{\mathrm{B}}\big)\right\}-E\left\{\boldsymbol{X}_{\mathrm{B}}\boldsymbol{X}_{\mathrm{A}}^{T}b''\big((\boldsymbol{\beta}_{\mathrm{A}}+\boldsymbol{U})^{T}\boldsymbol{X}_{\mathrm{A}}+\boldsymbol{\beta}_{\mathrm{B}}^{T}\boldsymbol{X}_{\mathrm{B}}\big)\right\}$$
$$\times\left[E\left\{\boldsymbol{X}_{\mathrm{B}}\boldsymbol{X}_{\mathrm{A}}^{T}b''\big((\boldsymbol{\beta}_{\mathrm{A}}+\boldsymbol{U})^{T}\boldsymbol{X}_{\mathrm{A}}+\boldsymbol{\beta}_{\mathrm{B}}^{T}\boldsymbol{X}_{\mathrm{B}}\big)\right\}\right]^{-1}E\left\{\boldsymbol{X}_{\mathrm{B}}\boldsymbol{X}_{\mathrm{A}}^{T}b''\big((\boldsymbol{\beta}_{\mathrm{A}}+\boldsymbol{U})^{T}\boldsymbol{X}_{\mathrm{A}}+\boldsymbol{\beta}_{\mathrm{B}}^{T}\boldsymbol{X}_{\mathrm{B}}\big)\right\}^{T}\Bigg)$$
$$+o_P(mn)\mathbf{1}_{d_{\mathrm{B}}}^{\otimes 2}.$$

Now since

$$\boldsymbol{\Omega}_{\boldsymbol{\beta}_{\mathrm{B}}}(\boldsymbol{U})\equiv E\left\{b''\left((\boldsymbol{\beta}_{\mathrm{A}}^{0}+\boldsymbol{U})^{T}\boldsymbol{X}_{\mathrm{A}}+(\boldsymbol{\beta}_{\mathrm{B}}^{0})^{T}\boldsymbol{X}_{\mathrm{B}}\right)\begin{bmatrix}\boldsymbol{X}_{\mathrm{A}}\boldsymbol{X}_{\mathrm{A}}^{T} & \boldsymbol{X}_{\mathrm{A}}\boldsymbol{X}_{\mathrm{B}}^{T}\\ \boldsymbol{X}_{\mathrm{B}}\boldsymbol{X}_{\mathrm{A}}^{T} & \boldsymbol{X}_{\mathrm{B}}\boldsymbol{X}_{\mathrm{B}}^{T}\end{bmatrix}\Big|\boldsymbol{U}\right\}$$

and

$$\boldsymbol{\Lambda}_{\boldsymbol{\beta}_{\mathrm{B}}}\equiv\left(E\left[\left\{\text{lower right } d_{\mathrm{B}}\times d_{\mathrm{B}}\text{ block of }\boldsymbol{\Omega}_{\boldsymbol{\beta}_{\mathrm{B}}}(\boldsymbol{U})^{-1}\right\}^{-1}\right]\right)^{-1},$$

then we have,

$$\frac{1}{\phi}\sum_{i=1}^{m}E\left(\mathcal{H}_{\mathrm{BB}i}-\mathcal{H}_{\mathrm{AB}i}^{T}\mathcal{H}_{\mathrm{AA}i}^{-1}\mathcal{H}_{\mathrm{AB}i}|\boldsymbol{X}_i\right)=\frac{mn\boldsymbol{\Lambda}_{\boldsymbol{\beta}_{\mathrm{B}}}^{-1}}{\phi}+o_P(mn)\mathbf{1}_{d_{\mathrm{B}}}^{\otimes 2}.$$

4.5.2.9 The Fisher Information Matrix

Putting together the expressions for the quadratic conditional expectations of the scores from the earlier sections, we have

$$I(\boldsymbol{\beta}_{\mathrm{A}},\boldsymbol{\beta}_{\mathrm{B}},\mathrm{vech}(\boldsymbol{\Sigma}))$$
$$=\begin{bmatrix}m\boldsymbol{\Sigma}^{-1}+O_P(mn^{-1})\mathbf{1}_{d_{\mathrm{A}}}^{\otimes 2} & O_P(m)\mathbf{1}_{d_{\mathrm{A}}}\mathbf{1}_{d_{\mathrm{B}}}^{T} & O_P(mn^{-1})\mathbf{1}_{d_{\mathrm{A}}}\mathbf{1}_{d_{\mathrm{A}}^{\boxplus}}^{T}\\ O_P(m)\mathbf{1}_{d_{\mathrm{B}}}\mathbf{1}_{d_{\mathrm{A}}}^{T} & \frac{mn\boldsymbol{\Lambda}_{\boldsymbol{\beta}_{\mathrm{B}}}^{-1}}{\phi}+o_P(mn)\mathbf{1}_{d_{\mathrm{B}}}^{\otimes 2} & O_P(m)\mathbf{1}_{d_{\mathrm{B}}}\mathbf{1}_{d_{\mathrm{A}}^{\boxplus}}^{T}\\ O_P(mn^{-1})\mathbf{1}_{d_{\mathrm{A}}^{\boxplus}}\mathbf{1}_{d_{\mathrm{A}}}^{T} & O_P(m)\mathbf{1}_{d_{\mathrm{A}}^{\boxplus}}\mathbf{1}_{d_{\mathrm{B}}}^{T} & \frac{m\boldsymbol{D}_{d_{\mathrm{A}}}^{T}(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1})\boldsymbol{D}_{d_{\mathrm{A}}}}{2}+O_P(mn^{-1})\mathbf{1}_{d_{\mathrm{A}}^{\boxplus}}^{\otimes 2}\end{bmatrix}.$$

4.5.2.10 The Inverse of the Fisher Information Matrix

To invert the Fisher information matrix, we choose to work with the $(\boldsymbol{\beta}_{\mathrm{A}},\mathrm{vech}(\boldsymbol{\Sigma}),\boldsymbol{\beta}_{\mathrm{B}})$ ordering instead of $(\boldsymbol{\beta}_{\mathrm{A}},\boldsymbol{\beta}_{\mathrm{B}},\mathrm{vech}(\boldsymbol{\Sigma}))$. A trivial rearrangement of the matrix entries

leads to

$$I(\boldsymbol{\beta}_{\mathrm{A}}, \boldsymbol{\beta}_{\mathrm{B}}, \mathrm{vech}(\boldsymbol{\Sigma}))$$
$$= \begin{bmatrix} m\boldsymbol{\Sigma}^{-1} + O_P(mn^{-1})\mathbf{1}_{d_{\mathrm{A}}}^{\otimes 2} & O_P(mn^{-1})\mathbf{1}_{d_{\mathrm{A}}}\mathbf{1}_{d_{\mathrm{A}}^{\boxplus}}^T & O_P(m)\mathbf{1}_{d_{\mathrm{A}}}\mathbf{1}_{d_{\mathrm{B}}}^T \\ O_P(mn^{-1})\mathbf{1}_{d_{\mathrm{A}}^{\boxplus}}\mathbf{1}_{d_{\mathrm{A}}}^T & \frac{m\boldsymbol{D}_{d_{\mathrm{A}}}^T(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1})\boldsymbol{D}_{d_{\mathrm{A}}}}{2} + O_P(mn^{-1})\mathbf{1}_{d_{\mathrm{A}}^{\boxplus}}^{\otimes 2} & O_P(m)\mathbf{1}_{d_{\mathrm{A}}^{\boxplus}}\mathbf{1}_{d_{\mathrm{B}}}^T \\ O_P(m)\mathbf{1}_{d_{\mathrm{B}}}\mathbf{1}_{d_{\mathrm{A}}}^T & O_P(m)\mathbf{1}_{d_{\mathrm{B}}}\mathbf{1}_{d_{\mathrm{A}}^{\boxplus}}^T & \frac{mn\boldsymbol{\Lambda}_{\boldsymbol{\beta}_{\mathrm{B}}}^{-1}}{\phi} + o_P(mn)\mathbf{1}_{d_{\mathrm{B}}}^{\otimes 2} \end{bmatrix}.$$

Now, let us partition $I(\boldsymbol{\beta}_{\mathrm{A}}, \mathrm{vech}(\boldsymbol{\Sigma}), \boldsymbol{\beta}_{\mathrm{B}})$ as follows

$$I(\boldsymbol{\beta}_{\mathrm{A}}, \mathrm{vech}(\boldsymbol{\Sigma}), \boldsymbol{\beta}_{\mathrm{B}}) = \begin{bmatrix} \boldsymbol{A}_{11} & \boldsymbol{A}_{12} \\ \boldsymbol{A}_{21} & \boldsymbol{A}_{22} \end{bmatrix} \text{ where } \boldsymbol{A}_{21} = \boldsymbol{A}_{12}^T.$$

The expressions for $\boldsymbol{A}_{11}$, $\boldsymbol{A}_{12}$, $\boldsymbol{A}_{21}$ and $\boldsymbol{A}_{22}$ are currently as follows

$$\boldsymbol{A}_{11} = \begin{bmatrix} m\boldsymbol{\Sigma}^{-1} + O_P(mn^{-1})\mathbf{1}_{d_{\mathrm{A}}}^{\otimes 2} & O_P(mn^{-1})\mathbf{1}_{d_{\mathrm{A}}}\mathbf{1}_{d_{\mathrm{A}}^{\boxplus}}^T \\ O_P(mn^{-1})\mathbf{1}_{d_{\mathrm{A}}^{\boxplus}}\mathbf{1}_{d_{\mathrm{A}}}^T & \frac{m\boldsymbol{D}_{d_{\mathrm{A}}}^T(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1})\boldsymbol{D}_{d_{\mathrm{A}}}}{2} + O_P(mn^{-1})\mathbf{1}_{d_{\mathrm{A}}^{\boxplus}}^{\otimes 2} \end{bmatrix},$$

$$\boldsymbol{A}_{12} = O_P(m)\begin{bmatrix} \mathbf{1}_{d_{\mathrm{A}}}\mathbf{1}_{d_{\mathrm{B}}}^T \\ \mathbf{1}_{d_{\mathrm{A}}^{\boxplus}}\mathbf{1}_{d_{\mathrm{B}}}^T \end{bmatrix},$$

$$\boldsymbol{A}_{21} = O_P(m)\begin{bmatrix} \mathbf{1}_{d_{\mathrm{B}}}\mathbf{1}_{d_{\mathrm{A}}}^T & \mathbf{1}_{d_{\mathrm{B}}}\mathbf{1}_{d_{\mathrm{A}}^{\boxplus}}^T \end{bmatrix},$$

$$\boldsymbol{A}_{22} = \frac{mn\boldsymbol{\Lambda}_{\boldsymbol{\beta}_{\mathrm{B}}}^{-1}}{\phi} + o_P(mn)\mathbf{1}_{d_{\mathrm{B}}}^{\otimes 2}.$$

Let $I(\boldsymbol{\beta}_{\mathrm{A}}, \mathrm{vech}(\boldsymbol{\Sigma}), \boldsymbol{\beta}_{\mathrm{B}})^{-1}$ assume the following form

$$I(\boldsymbol{\beta}_{\mathrm{A}}, \mathrm{vech}(\boldsymbol{\Sigma}), \boldsymbol{\beta}_{\mathrm{B}})^{-1} = \begin{bmatrix} \boldsymbol{A}_{11} & \boldsymbol{A}_{12} \\ \boldsymbol{A}_{21} & \boldsymbol{A}_{22} \end{bmatrix}^{-1} = \begin{bmatrix} \boldsymbol{A}^{11} & \boldsymbol{A}^{12} \\ \boldsymbol{A}^{21} & \boldsymbol{A}^{22} \end{bmatrix} \quad \text{where } \boldsymbol{A}^{21} = (\boldsymbol{A}^{12})^T.$$

Firstly note that

$$\boldsymbol{A}_{11}^{-1} = \begin{bmatrix} \frac{\boldsymbol{\Sigma}}{m} + O_P(m^{-1}n^{-1})\mathbf{1}_{d_{\mathrm{A}}}^{\otimes 2} & O_P(m^{-1}n^{-1})\mathbf{1}_{d_{\mathrm{A}}}\mathbf{1}_{d_{\mathrm{A}}^{\boxplus}}^T \\ O_P(m^{-1}n^{-1})\mathbf{1}_{d_{\mathrm{A}}^{\boxplus}}\mathbf{1}_{d_{\mathrm{A}}}^T & \frac{2\boldsymbol{D}_{d_{\mathrm{A}}}^+(\boldsymbol{\Sigma}\otimes\boldsymbol{\Sigma})\boldsymbol{D}_{d_{\mathrm{A}}}^{+\ T}}{m} + O_P(m^{-1}n^{-1})\mathbf{1}_{d_{\mathrm{A}}^{\boxplus}}^{\otimes 2} \end{bmatrix}.$$

Also note that

$$\boldsymbol{A}_{22}^{-1} = \frac{\phi\boldsymbol{\Lambda}_{\boldsymbol{\beta}_{\mathrm{B}}}}{mn} + o_P(m^{-1}n^{-1})\mathbf{1}_{d_{\mathrm{B}}}^{\otimes 2}.$$

Using the result for carrying out block matrix inversion under Result 1, the quantities $\boldsymbol{A}^{11}, \boldsymbol{A}^{12}, \boldsymbol{A}^{21}$ and $\boldsymbol{A}^{22}$ can be calculated as follows. Firstly we have

$$\boldsymbol{A}^{11} = \boldsymbol{A}_{11}^{-1} + \boldsymbol{A}_{11}^{-1}\boldsymbol{A}_{12}(\boldsymbol{A}_{22} - \boldsymbol{A}_{21}\boldsymbol{A}_{11}^{-1}\boldsymbol{A}_{12})^{-1}\boldsymbol{A}_{21}\boldsymbol{A}_{11}^{-1}.$$

Note that

$$(\boldsymbol{A}_{22} - \boldsymbol{A}_{21}\boldsymbol{A}_{11}^{-1}\boldsymbol{A}_{12})^{-1} = \boldsymbol{A}_{22}^{-1} + \boldsymbol{A}_{22}^{-1}\boldsymbol{A}_{21}\boldsymbol{A}_{11}^{-1}\boldsymbol{A}_{12}\boldsymbol{A}_{22}^{-1} + \ldots$$

It follows that

$$\boldsymbol{A}_{11}^{-1}\boldsymbol{A}_{12}(\boldsymbol{A}_{22}-\boldsymbol{A}_{21}\boldsymbol{A}_{11}^{-1}\boldsymbol{A}_{12})^{-1}\boldsymbol{A}_{21}\boldsymbol{A}_{11}^{-1}=O_P(m^{-1}n^{-1})\begin{bmatrix}\mathbf{1}_{d_{\mathrm{A}}}^{\otimes 2} & \mathbf{1}_{d_{\mathrm{A}}}\mathbf{1}_{d_{\mathrm{A}}^{\boxplus}}^{T}\\ \mathbf{1}_{d_{\mathrm{A}}^{\boxplus}}\mathbf{1}_{d_{\mathrm{A}}}^{T} & \mathbf{1}_{d_{\mathrm{A}}^{\boxplus}}^{\otimes 2}\end{bmatrix}.$$

Therefore, it follows that

$$\boldsymbol{A}^{11}=\begin{bmatrix}\frac{\boldsymbol{\Sigma}}{m}+O_P(m^{-1}n^{-1})\mathbf{1}_{d_{\mathrm{A}}}^{\otimes 2} & O_P(m^{-1}n^{-1})\mathbf{1}_{d_{\mathrm{A}}}\mathbf{1}_{d_{\mathrm{A}}^{\boxplus}}^{T}\\ O_P(m^{-1}n^{-1})\mathbf{1}_{d_{\mathrm{A}}^{\boxplus}}\mathbf{1}_{d_{\mathrm{A}}}^{T} & \frac{2\boldsymbol{D}_{d_{\mathrm{A}}}^{+}(\boldsymbol{\Sigma}\otimes\boldsymbol{\Sigma})\boldsymbol{D}_{d_{\mathrm{A}}}^{+\ T}}{m}+O_P(m^{-1}n^{-1})\mathbf{1}_{d_{\mathrm{A}}^{\boxplus}}^{\otimes 2}\end{bmatrix}.$$

Next we have,

$$\boldsymbol{A}^{12}=-(\boldsymbol{A}_{11}-\boldsymbol{A}_{12}\boldsymbol{A}_{22}^{-1}\boldsymbol{A}_{21})^{-1}\boldsymbol{A}_{12}\boldsymbol{A}_{22}^{-1}.$$

Note that

$$(\boldsymbol{A}_{11}-\boldsymbol{A}_{12}\boldsymbol{A}_{22}^{-1}\boldsymbol{A}_{21})^{-1}=\boldsymbol{A}_{11}^{-1}+\boldsymbol{A}_{11}^{-1}\boldsymbol{A}_{12}\boldsymbol{A}_{22}^{-1}\boldsymbol{A}_{21}\boldsymbol{A}_{11}^{-1}+\ldots$$

Therefore we have

$$\boldsymbol{A}^{12}=O_P(m^{-1}n^{-1})\begin{bmatrix}\mathbf{1}_{d_{\mathrm{A}}}\mathbf{1}_{d_{\mathrm{B}}}^{T}\\ \mathbf{1}_{d_{\mathrm{A}}^{\boxplus}}\mathbf{1}_{d_{\mathrm{B}}}^{T}\end{bmatrix}.$$

Subsequently,

$$\boldsymbol{A}^{21}=(\boldsymbol{A}^{12})^{T}.$$

Therefore we have

$$\boldsymbol{A}^{21}=O_P(m^{-1}n^{-1})\begin{bmatrix}\mathbf{1}_{d_{\mathrm{B}}}\mathbf{1}_{d_{\mathrm{A}}}^{T} & \mathbf{1}_{d_{\mathrm{B}}}\mathbf{1}_{d_{\mathrm{A}}^{\boxplus}}^{T}\end{bmatrix}.$$

Finally, we obtain

$$\boldsymbol{A}^{22}=\boldsymbol{A}_{22}^{-1}+\boldsymbol{A}_{22}^{-1}\boldsymbol{A}_{21}(\boldsymbol{A}_{11}-\boldsymbol{A}_{12}\boldsymbol{A}_{22}^{-1}\boldsymbol{A}_{21})^{-1}\boldsymbol{A}_{12}\boldsymbol{A}_{22}^{-1}.$$

Note that

$$\boldsymbol{A}_{22}^{-1}\boldsymbol{A}_{21}(\boldsymbol{A}_{11}-\boldsymbol{A}_{12}\boldsymbol{A}_{22}^{-1}\boldsymbol{A}_{21})^{-1}\boldsymbol{A}_{12}\boldsymbol{A}_{22}^{-1}=O_P(m^{-1}n^{-2})\mathbf{1}_{d_{\mathrm{B}}}^{\otimes 2}.$$

This leads to

$$\boldsymbol{A}^{22}=\frac{\phi\boldsymbol{\Lambda}_{\boldsymbol{\beta}_{\mathrm{B}}}}{mn}+o_P(m^{-1}n^{-1})\mathbf{1}_{d_{\mathrm{B}}}^{\otimes 2}.$$

Using the expressions for $\boldsymbol{A}^{11}$, $\boldsymbol{A}^{12}$, $\boldsymbol{A}^{21}$ and $\boldsymbol{A}^{22}$, we have the following expression for $I(\boldsymbol{\beta}_{\mathrm{A}},\mathrm{vech}(\boldsymbol{\Sigma}),\boldsymbol{\beta}_{\mathrm{B}})^{-1}$ where

$$\begin{aligned}&I(\boldsymbol{\beta}_{\mathrm{A}},\mathrm{vech}(\boldsymbol{\Sigma}),\boldsymbol{\beta}_{\mathrm{B}})^{-1}\\ &=\begin{bmatrix}\frac{\boldsymbol{\Sigma}}{m}+O_P(m^{-1}n^{-1})\mathbf{1}_{d_{\mathrm{A}}}^{\otimes 2} & O_P(m^{-1}n^{-1})\mathbf{1}_{d_{\mathrm{A}}}\mathbf{1}_{d_{\mathrm{A}}^{\boxplus}}^{T} & O_P(m^{-1}n^{-1})\mathbf{1}_{d_{\mathrm{A}}}\mathbf{1}_{d_{\mathrm{B}}}^{T}\\ O_P(m^{-1}n^{-1})\mathbf{1}_{d_{\mathrm{A}}^{\boxplus}}\mathbf{1}_{d_{\mathrm{A}}}^{T} & \frac{2\boldsymbol{D}_{d_{\mathrm{A}}}^{+}(\boldsymbol{\Sigma}\otimes\boldsymbol{\Sigma})\boldsymbol{D}_{d_{\mathrm{A}}}^{+\ T}}{m}+O_P(m^{-1}n^{-1})\mathbf{1}_{d_{\mathrm{A}}^{\boxplus}}^{\otimes 2} & O_P(m^{-1}n^{-1})\mathbf{1}_{d_{\mathrm{A}}^{\boxplus}}\mathbf{1}_{d_{\mathrm{B}}}^{T}\\ O_P(m^{-1}n^{-1})\mathbf{1}_{d_{\mathrm{B}}}\mathbf{1}_{d_{\mathrm{A}}}^{T} & O_P(m^{-1}n^{-1})\mathbf{1}_{d_{\mathrm{B}}}\mathbf{1}_{d_{\mathrm{A}}^{\boxplus}}^{T} & \frac{\phi\boldsymbol{\Lambda}_{\boldsymbol{\beta}_{\mathrm{B}}}}{mn}+o_P(m^{-1}n^{-1})\mathbf{1}_{d_{\mathrm{B}}}^{\otimes 2}\end{bmatrix}.\end{aligned}$$

The expression for the inverse of the Fisher information matrix can be also written as follows

$$
\begin{aligned}
&I(\boldsymbol{\beta}_{\mathrm{A}}, \mathrm{vech}(\boldsymbol{\Sigma}), \boldsymbol{\beta}_{\mathrm{B}})^{-1} \\
&= I(\boldsymbol{\beta}_{\mathrm{A}}, \mathrm{vech}(\boldsymbol{\Sigma}), \boldsymbol{\beta}_{\mathrm{B}})_{\infty}^{-1} + \frac{1}{mn} \begin{bmatrix} O_P(1)\mathbf{1}_{d_{\mathrm{A}}}^{\otimes 2} & O_P(1)\mathbf{1}_{d_{\mathrm{A}}}\mathbf{1}_{d_{\mathrm{A}}^{\boxplus}}^T & O_P(1)\mathbf{1}_{d_{\mathrm{A}}}\mathbf{1}_{d_{\mathrm{B}}}^T \\ O_P(1)\mathbf{1}_{d_{\mathrm{A}}^{\boxplus}}\mathbf{1}_{d_{\mathrm{A}}}^T & O_P(1)\mathbf{1}_{d_{\mathrm{A}}^{\boxplus}}^{\otimes 2} & O_P(1)\mathbf{1}_{d_{\mathrm{A}}^{\boxplus}}\mathbf{1}_{d_{\mathrm{B}}}^T \\ O_P(1)\mathbf{1}_{d_{\mathrm{B}}}\mathbf{1}_{d_{\mathrm{A}}}^T & O_P(1)\mathbf{1}_{d_{\mathrm{B}}}\mathbf{1}_{d_{\mathrm{A}}^{\boxplus}}^T & o_P(1)\mathbf{1}_{d_{\mathrm{B}}}^{\otimes 2} \end{bmatrix},
\end{aligned}
$$

where

$$
I(\boldsymbol{\beta}_{\mathrm{A}}, \mathrm{vech}(\boldsymbol{\Sigma}), \boldsymbol{\beta}_{\mathrm{B}})_{\infty}^{-1} = \begin{bmatrix} \frac{\boldsymbol{\Sigma}}{m} & \boldsymbol{O} & \boldsymbol{O} \\ \boldsymbol{O} & \frac{2\boldsymbol{D}_{d_{\mathrm{A}}}^{+}(\boldsymbol{\Sigma}\otimes\boldsymbol{\Sigma})\boldsymbol{D}_{d_{\mathrm{A}}}^{+\,T}}{m} & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{O} & \frac{\phi\boldsymbol{\Lambda}_{\boldsymbol{\beta}_{\mathrm{B}}}}{mn} \end{bmatrix}.
$$

4.5.2.11 Derivation of the Final Asymptotic Normality Result for Generalized Response Linear Mixed Models

For a matrix $\boldsymbol{M}$ let

$$
\|\boldsymbol{M}\|_F = \sqrt{\mathrm{tr}(\boldsymbol{M}^T\boldsymbol{M})}
$$

denote the *Frobenius norm* of $\boldsymbol{M}$.

For regular likelihood situations, from standard results concerning asymptotic normality of maximum likelihood estimators we have

$$
\{I\left(\boldsymbol{\beta}_{\mathrm{A}}^0, \mathrm{vech}(\boldsymbol{\Sigma}^0), \boldsymbol{\beta}_{\mathrm{B}}^0\right)^{-1}\}^{-1/2}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^0) \xrightarrow{\mathcal{D}} N(\mathbf{0}, \boldsymbol{I})
$$

where $\widehat{\boldsymbol{\theta}} = \left[\widehat{\boldsymbol{\beta}}_{\mathrm{A}}^T\ \mathrm{vech}(\widehat{\boldsymbol{\Sigma}})^T\ \widehat{\boldsymbol{\beta}}_{\mathrm{B}}^T\right]^T$ and $\boldsymbol{\theta}^0 = [(\boldsymbol{\beta}_{\mathrm{A}}^0)^T\ \{\mathrm{vech}(\boldsymbol{\Sigma}^0)\}^T\ (\boldsymbol{\beta}_{\mathrm{B}}^0)^T]^T$. On the other hand, general quasi-likelihood situations require asymptotic normality theory as treated in, for example, Section 5.3 of the book *Asymptotic Statistics* (see van der Vaart (1998)). Therefore, for all $(d_{\mathrm{A}} + d_{\mathrm{A}}^{\boxplus} + d_{\mathrm{B}}) \times 1$ vectors $\boldsymbol{a} \neq 0$ we have

$$
\boldsymbol{a}^T\{I\left(\boldsymbol{\beta}_{\mathrm{A}}^0, \mathrm{vech}(\boldsymbol{\Sigma}^0), \boldsymbol{\beta}_{\mathrm{B}}^0\right)^{-1}\}^{-1/2}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^0) \xrightarrow{\mathcal{D}} N(\mathbf{0}, \boldsymbol{a}^T\boldsymbol{a}).
$$

Note that

$$
\begin{aligned}
&\boldsymbol{a}^T\{I\left(\boldsymbol{\beta}_{\mathrm{A}}^0, \mathrm{vech}(\boldsymbol{\Sigma}^0), \boldsymbol{\beta}_{\mathrm{B}}^0\right)^{-1}\}^{-1/2}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^0) \\
&= \boldsymbol{a}^T\Big[\{I\left(\boldsymbol{\beta}_{\mathrm{A}}^0, \mathrm{vech}(\boldsymbol{\Sigma}^0), \boldsymbol{\beta}_{\mathrm{B}}^0\right)_{\infty}^{-1}\}^{-1/2} + \{I\left(\boldsymbol{\beta}_{\mathrm{A}}^0, \mathrm{vech}(\boldsymbol{\Sigma}^0), \boldsymbol{\beta}_{\mathrm{B}}^0\right)^{-1}\}^{-1/2} \\
&\quad - \{I\left(\boldsymbol{\beta}_{\mathrm{A}}^0, \mathrm{vech}(\boldsymbol{\Sigma}^0), \boldsymbol{\beta}_{\mathrm{B}}^0\right)_{\infty}^{-1}\}^{-1/2}\Big](\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^0) \\
&= \boldsymbol{a}^T\{I\left(\boldsymbol{\beta}_{\mathrm{A}}^0, \mathrm{vech}(\boldsymbol{\Sigma}^0), \boldsymbol{\beta}_{\mathrm{B}}^0\right)_{\infty}^{-1}\}^{-1/2}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^0) \\
&\quad + \boldsymbol{a}^T\Big[\{I\left(\boldsymbol{\beta}_{\mathrm{A}}^0, \mathrm{vech}(\boldsymbol{\Sigma}^0), \boldsymbol{\beta}_{\mathrm{B}}^0\right)^{-1}\}^{-1/2} - \{I\left(\boldsymbol{\beta}_{\mathrm{A}}^0, \mathrm{vech}(\boldsymbol{\Sigma}^0), \boldsymbol{\beta}_{\mathrm{B}}^0\right)_{\infty}^{-1}\}^{-1/2}\Big](\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^0).
\end{aligned}
$$

As a consequence

$$\boldsymbol{a}^T\{I\left(\boldsymbol{\beta}_{\mathrm{A}}^0, \mathrm{vech}(\boldsymbol{\Sigma}^0), \boldsymbol{\beta}_{\mathrm{B}}^0\right)_{\infty}^{-1}\}^{-1/2}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^0) + r_{mn}(\boldsymbol{a}) \xrightarrow{\mathcal{D}} N(\boldsymbol{0}, \boldsymbol{a}^T\boldsymbol{a}) \tag{4.15}$$

with

$$\begin{aligned} r_{mn}(\boldsymbol{a}) &= \boldsymbol{a}^T\Big[\{I\left(\boldsymbol{\beta}_{\mathrm{A}}^0, \mathrm{vech}(\boldsymbol{\Sigma}^0), \boldsymbol{\beta}_{\mathrm{B}}^0\right)^{-1}\}^{-1/2} - \{I\left(\boldsymbol{\beta}_{\mathrm{A}}^0, \mathrm{vech}(\boldsymbol{\Sigma}^0), \boldsymbol{\beta}_{\mathrm{B}}^0\right)_{\infty}^{-1}\}^{-1/2}\Big](\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^0) \\ &= \boldsymbol{a}^T\Big[\boldsymbol{I} - \{I\left(\boldsymbol{\beta}_{\mathrm{A}}^0, \mathrm{vech}(\boldsymbol{\Sigma}^0), \boldsymbol{\beta}_{\mathrm{B}}^0\right)_{\infty}^{-1}\}^{-1/2}\{I\left(\boldsymbol{\beta}_{\mathrm{A}}^0, \mathrm{vech}(\boldsymbol{\Sigma}^0), \boldsymbol{\beta}_{\mathrm{B}}^0\right)^{-1}\}^{1/2}\Big] \\ &\quad \times \{I\left(\boldsymbol{\beta}_{\mathrm{A}}^0, \mathrm{vech}(\boldsymbol{\Sigma}^0), \boldsymbol{\beta}_{\mathrm{B}}^0\right)^{-1}\}^{-1/2}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^0) \\ &= \left(-\Big[\{I\left(\boldsymbol{\beta}_{\mathrm{A}}^0, \mathrm{vech}(\boldsymbol{\Sigma}^0), \boldsymbol{\beta}_{\mathrm{B}}^0\right)_{\infty}^{-1}\}^{-1/2}\{I\left(\boldsymbol{\beta}_{\mathrm{A}}^0, \mathrm{vech}(\boldsymbol{\Sigma}^0), \boldsymbol{\beta}_{\mathrm{B}}^0\right)^{-1}\}^{1/2} - \boldsymbol{I}\Big]^T\boldsymbol{a}\right)^T\boldsymbol{Z} \end{aligned}$$

where $\boldsymbol{Z} \sim N\left(\boldsymbol{0}, \boldsymbol{I}_{d_{\mathrm{A}}+d_{\mathrm{A}}^{\boxplus}+d_{\mathrm{B}}}\right)$. Next note that using the matrix norm properties $\|-\boldsymbol{A}\| = \|\boldsymbol{A}\|$ and $\|\boldsymbol{AB}\| \leq \|\boldsymbol{A}\|\|\boldsymbol{B}\|$ for any matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ and the fact that $\|\boldsymbol{M}^T\|_F = \|\boldsymbol{M}\|$ for any matrix $\boldsymbol{M}$, we have

$$\begin{aligned} &\Big\| -\Big[\{I\left(\boldsymbol{\beta}_{\mathrm{A}}^0, \mathrm{vech}(\boldsymbol{\Sigma}^0), \boldsymbol{\beta}_{\mathrm{B}}^0\right)_{\infty}^{-1}\}^{-1/2}\{I\left(\boldsymbol{\beta}_{\mathrm{A}}^0, \mathrm{vech}(\boldsymbol{\Sigma}^0), \boldsymbol{\beta}_{\mathrm{B}}^0\right)^{-1}\}^{1/2} - \boldsymbol{I}\Big]^T\boldsymbol{a}\Big\|_F \\ &\leq \Big\|\Big[\{I\left(\boldsymbol{\beta}_{\mathrm{A}}^0, \mathrm{vech}(\boldsymbol{\Sigma}^0), \boldsymbol{\beta}_{\mathrm{B}}^0\right)_{\infty}^{-1}\}^{-1/2}\{I\left(\boldsymbol{\beta}_{\mathrm{A}}^0, \mathrm{vech}(\boldsymbol{\Sigma}^0), \boldsymbol{\beta}_{\mathrm{B}}^0\right)^{-1}\}^{1/2} - \boldsymbol{I}\Big]^T\Big\|_F\|\boldsymbol{a}\|_F \\ &= \Big\|\Big[\{I\left(\boldsymbol{\beta}_{\mathrm{A}}^0, \mathrm{vech}(\boldsymbol{\Sigma}^0), \boldsymbol{\beta}_{\mathrm{B}}^0\right)_{\infty}^{-1}\}^{-1/2}\{I\left(\boldsymbol{\beta}_{\mathrm{A}}^0, \mathrm{vech}(\boldsymbol{\Sigma}^0), \boldsymbol{\beta}_{\mathrm{B}}^0\right)^{-1}\}^{1/2} - \boldsymbol{I}\Big]\Big\|_F\|\boldsymbol{a}\|_F. \end{aligned} \tag{4.16}$$

Our next aim is to establish that

$$\Big\|\Big[\{I\left(\boldsymbol{\beta}_{\mathrm{A}}^0, \mathrm{vech}(\boldsymbol{\Sigma}^0), \boldsymbol{\beta}_{\mathrm{B}}^0\right)_{\infty}^{-1}\}^{-1/2}\{I\left(\boldsymbol{\beta}_{\mathrm{A}}^0, \mathrm{vech}(\boldsymbol{\Sigma}^0), \boldsymbol{\beta}_{\mathrm{B}}^0\right)^{-1}\}^{1/2} - \boldsymbol{I}\Big]\Big\|_F \xrightarrow{P} 0 \tag{4.17}$$

Recall that

$$\begin{aligned} &I(\boldsymbol{\beta}_{\mathrm{A}}, \mathrm{vech}(\boldsymbol{\Sigma}), \boldsymbol{\beta}_{\mathrm{B}})^{-1} \\ &= I(\boldsymbol{\beta}_{\mathrm{A}}, \mathrm{vech}(\boldsymbol{\Sigma}), \boldsymbol{\beta}_{\mathrm{B}})_{\infty}^{-1} + \frac{1}{mn}\begin{bmatrix} O_P(1)\boldsymbol{1}_{d_{\mathrm{A}}}^{\otimes 2} & O_P(1)\boldsymbol{1}_{d_{\mathrm{A}}}\boldsymbol{1}_{d_{\mathrm{A}}^{\boxplus}}^T & O_P(1)\boldsymbol{1}_{d_{\mathrm{A}}}\boldsymbol{1}_{d_{\mathrm{B}}}^T \\ O_P(1)\boldsymbol{1}_{d_{\mathrm{A}}^{\boxplus}}\boldsymbol{1}_{d_{\mathrm{A}}}^T & O_P(1)\boldsymbol{1}_{d_{\mathrm{A}}^{\boxplus}}^{\otimes 2} & O_P(1)\boldsymbol{1}_{d_{\mathrm{A}}^{\boxplus}}\boldsymbol{1}_{d_{\mathrm{B}}}^T \\ O_P(1)\boldsymbol{1}_{d_{\mathrm{B}}}\boldsymbol{1}_{d_{\mathrm{A}}}^T & O_P(1)\boldsymbol{1}_{d_{\mathrm{B}}}\boldsymbol{1}_{d_{\mathrm{A}}^{\boxplus}}^T & o_P(1)\boldsymbol{1}_{d_{\mathrm{B}}}^{\otimes 2} \end{bmatrix}, \end{aligned}$$

where

$$I(\boldsymbol{\beta}_{\mathrm{A}}, \mathrm{vech}(\boldsymbol{\Sigma}), \boldsymbol{\beta}_{\mathrm{B}})_{\infty}^{-1} = \begin{bmatrix} \frac{\boldsymbol{\Sigma}}{m} & \boldsymbol{O} & \boldsymbol{O} \\ \boldsymbol{O} & \frac{2\boldsymbol{D}_{d_{\mathrm{A}}}^{+}(\boldsymbol{\Sigma}\otimes\boldsymbol{\Sigma})\boldsymbol{D}_{d_{\mathrm{A}}}^{+\,T}}{m} & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{O} & \frac{\phi\boldsymbol{\Lambda}_{\boldsymbol{\beta}_{\mathrm{B}}}}{mn} \end{bmatrix}.$$

so that

$$\left\{I\left(\boldsymbol{\beta}_{\mathrm{A}}^0, \mathrm{vech}(\boldsymbol{\Sigma}^0), \boldsymbol{\beta}_{\mathrm{B}}^0\right)_{\infty}^{-1}\right\}^{-1/2}\left\{I\left(\boldsymbol{\beta}_{\mathrm{A}}^0, \mathrm{vech}(\boldsymbol{\Sigma}^0), \boldsymbol{\beta}_{\mathrm{B}}^0\right)^{-1}\right\}^{1/2} = \boldsymbol{M}_{n,\infty}^{-1/2}\boldsymbol{M}_n^{1/2}$$

with

$$\boldsymbol{M}_{n,\infty} = m\left\{I\left(\boldsymbol{\beta}_{\mathrm{A}}^0, \mathrm{vech}(\boldsymbol{\Sigma}^0), \boldsymbol{\beta}_{\mathrm{B}}^0\right)_{\infty}^{-1}\right\} \text{ and } \boldsymbol{M}_n = m\left\{I\left(\boldsymbol{\beta}_{\mathrm{A}}^0, \mathrm{vech}(\boldsymbol{\Sigma}^0), \boldsymbol{\beta}_{\mathrm{B}}^0\right)^{-1}\right\}.$$

Therefore, Lemma 3 from Chapter 2 applies with

$$p = d_{\mathrm{A}} + d_{\mathrm{A}}^{\boxplus}, \quad \boldsymbol{K} = \begin{bmatrix} \boldsymbol{\Sigma} & \mathbf{0} \\ \mathbf{0} & 2\boldsymbol{D}_{d_{\mathrm{A}}}^{+}\left(\boldsymbol{\Sigma}\otimes\boldsymbol{\Sigma}\right)\boldsymbol{D}_{d_{\mathrm{A}}}^{+T} \end{bmatrix}, \quad q = d_{\mathrm{B}} \quad \text{and} \quad \boldsymbol{L} = \frac{\phi\boldsymbol{\Lambda}_{\boldsymbol{\beta}_{\mathrm{B}}}}{n}.$$

Therefore (4.17) holds. It then follows from (4.16) and (4.17) that

$$\left[\{I\left(\boldsymbol{\beta}_{\mathrm{A}}^{0}, \mathrm{vech}(\boldsymbol{\Sigma}^{0}), \boldsymbol{\beta}_{\mathrm{B}}^{0}\right)_{\infty}^{-1}\}^{-1/2}\{I\left(\boldsymbol{\beta}_{\mathrm{A}}^{0}, \mathrm{vech}(\boldsymbol{\Sigma}^{0}), \boldsymbol{\beta}_{\mathrm{B}}^{0}\right)^{-1}\}^{1/2} - \boldsymbol{I}\right]\boldsymbol{a} \xrightarrow{P} 0.$$

Application of Slutsky's Theorem then gives $r_{mn}(\boldsymbol{a}) \xrightarrow{P} 0$. From (4.15) and another application of Slutsky's Theorem we have

$$\boldsymbol{a}^{T}\{I\left(\boldsymbol{\beta}_{\mathrm{A}}^{0}, \mathrm{vech}(\boldsymbol{\Sigma}^{0}), \boldsymbol{\beta}_{\mathrm{B}}^{0}\right)_{\infty}^{-1}\}^{-1/2}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^{0}) \xrightarrow{\mathcal{D}} N(\mathbf{0}, \boldsymbol{a}^{T}\boldsymbol{a}).$$

It then follows from the Cramér-Wold Device and the Continuous Mapping Theorem that

$$\sqrt{m}\begin{bmatrix} \widehat{\boldsymbol{\beta}}_{\mathrm{A}} - \boldsymbol{\beta}_{\mathrm{A}}^{0} \\ \sqrt{n}\left(\widehat{\boldsymbol{\beta}}_{\mathrm{B}} - \boldsymbol{\beta}_{\mathrm{B}}^{0}\right) \\ \mathrm{vech}(\widehat{\boldsymbol{\Sigma}}) - \mathrm{vech}(\boldsymbol{\Sigma}^{0}) \end{bmatrix} \xrightarrow{\mathcal{D}} N\left(\begin{bmatrix} \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Sigma}^{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \phi\boldsymbol{\Lambda}_{\boldsymbol{\beta}_{\mathrm{B}}} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & 2\boldsymbol{D}_{d_{\mathrm{A}}}^{+}\left(\boldsymbol{\Sigma}^{0}\otimes\boldsymbol{\Sigma}^{0}\right)\boldsymbol{D}_{d_{\mathrm{A}}}^{+T} \end{bmatrix}\right).$$

4.5.3 The Reciprocal Dispersion Parameter Fisher Information Block for Gamma Responses

In this appendix, we derive the block of the Fisher information matrix for the parameter $\psi \equiv 1/\phi$ where ϕ is the dispersion parameter. We start with the general response situation and, later, focus on the Gamma case. With notational simplicity in mind we treat the $d_{\mathrm{A}} = d_{\mathrm{B}} = 1$ case with $\boldsymbol{\Sigma} = \sigma^{2}$ and $n_i = n$, $1 \le i \le m$. These restrictions will not affect the ψ contribution to the Fisher information matrix.

4.5.3.1 The Conditional Density Function

The conditional density function for the ith group is

$$\begin{aligned} p_{\boldsymbol{Y}_i|\boldsymbol{X}_i}(\boldsymbol{Y}_i|\boldsymbol{X}_i) &= \frac{1}{\sqrt{2\pi\sigma^{2}}}\int_{-\infty}^{\infty}\exp\Bigg[-\frac{u^{2}}{2\sigma^{2}} + \sum_{j=1}^{n}\Big\{\psi\Big(Y_{ij}(\beta_0 + \beta_1 X_{ij} + u) \\ &\qquad - b(\beta_0 + \beta_1 X_{ij} + u) + c(Y_{ij})\Big) + d(Y_{ij}, \psi)\Big\}\Bigg]\, du \\ &= (2\pi\sigma^{2})^{-1/2}\exp\left(\sum_{j=1}^{n}\left[\psi\left\{Y_{ij}(\beta_0 + \beta_1 X_{ij}) + c(Y_{ij})\right\} + d(Y_{ij}, \psi)\right]\right) \\ &\qquad \times \int_{-\infty}^{\infty}\exp\left[-\frac{u^{2}}{2\sigma^{2}} + \psi\sum_{j=1}^{n}\left\{Y_{ij}u - b(\beta_0 + \beta_1 X_{ij} + u)\right\}\right]\, du. \end{aligned}$$

In the Gamma response case,

$$c(y) = \log(y) \quad \text{and} \quad d(y,\psi) = \psi\log(\psi) - \log\{\Gamma(\psi)\} - \log(y).$$

Therefore

$$\frac{\partial d(y,\psi)}{\partial \psi} = \mathfrak{q}(\psi) \quad \text{and} \quad \frac{\partial^2 d(y,\psi)}{\partial \psi^2} = (1/\psi) - \text{trigamma}(\psi),$$

where

$$\mathfrak{q}(x) \equiv 1 + \log(x) - \text{digamma}(x).$$

Note that the model is such that

$$Y_{ij}|\boldsymbol{X}_i, U_i \overset{\text{ind.}}{\sim} \text{Gamma}\big(\psi, \psi/b'(\beta_0 + \beta_1 X_{ij} + U_i)\big). \tag{4.18}$$

Since $b'(x) = -1/x$ the statement (4.18) is equivalent to

$$Y_{ij}|\boldsymbol{X}_i, U_i \overset{\text{ind.}}{\sim} \text{Gamma}\big(\psi, -\psi(\beta_0 + \beta_1 X_{ij} + U_i)\big). \tag{4.19}$$

Using the result that

$$X \sim \text{Gamma}(\kappa,\lambda) \quad \text{implies} \quad E\{\log(X)\} = \text{digamma}(\kappa) - \log(\lambda)$$

and using $b(x) = -\log(-x)$, we have,

$$\begin{aligned} E\{c(Y_{ij})|U_i, \boldsymbol{X}_i\} &= E\{\log(Y_{ij})|U_i, \boldsymbol{X}_i\} \\ &= \text{digamma}(\psi) - \log[\{-(\beta_0 + \beta_1 X_{ij} + U_i)\}\psi] \\ &= -\log\{-(\beta_0 + \beta_1 X_{ij} + U_i)\} + \text{digamma}(\psi) - \log(\psi) \\ &= b(\beta_0 + \beta_1 X_{ij} + U_i) + \text{digamma}(\psi) - \log(\psi). \end{aligned}$$

Therefore, if we define

$$\mathcal{A}_i \equiv \sum_{j=1}^{n} b(\beta_0 + \beta_1 X_{ij} + U_i)$$

then

$$\sum_{j=1}^{n} E\{\log(Y_{ij})|U_i, \boldsymbol{X}_i\} = \mathcal{A}_i - n\{\log(\psi) - \text{digamma}(\psi)\}.$$

Also, if we define

$$\mathfrak{q}(x) \equiv 1 + \log(x) - \text{digamma}(x)$$

then

$$E\{\log(Y_{ij})|U_i, \boldsymbol{X}_i\} = b(\beta_0 + \beta_1 X_{ij} + U_i) + 1 - \mathfrak{q}(\psi)$$

and

$$\sum_{j=1}^{n} E\{\log(Y_{ij})|U_i, \boldsymbol{X}_i\} = \mathcal{A}_i + n\{1 - \mathfrak{q}(\psi)\}.$$

In the Gamma case

$$d(y,\psi) = \psi\log(\psi) - \log\{\Gamma(\psi)\} - \log(y) \quad \text{implies} \quad \frac{\partial d(y,\psi)}{\partial\psi} = \mathfrak{q}(\psi).$$

Therefore,

$$\sum_{j=1}^{n}\frac{\partial}{\partial\psi}d(Y_{ij},\psi) = n\mathfrak{q}(\psi).$$

Next consider the problem of obtaining an expression for $E(Y_{ij}^2|\boldsymbol{X}_i,U_i)$.

$$\begin{aligned}
E(Y_{ij}^2|\boldsymbol{X}_i,U_i) &= \mathrm{Var}(Y_{ij}|\boldsymbol{X}_i,U_i) + \{E(Y_{ij}|\boldsymbol{X}_i,U_i)\}^2 \\
&= \frac{1}{\psi(\beta_0+\beta_1X_{ij}+U_i)^2} + \left(\frac{1}{-(\beta_0+\beta_1X_{ij}+U_i)}\right)^2 \\
&= \left(\frac{1}{\psi}+1\right)\frac{1}{(\beta_0+\beta_1X_{ij}+U_i)^2} \\
&= \left(\frac{1}{\psi}+1\right)\{b'(\beta_0+\beta_1X_{ij}+U_i)\}^2.
\end{aligned}$$

Later on we also need expressions for

$$E\{Y_{ij}\log(Y_{ij})|\boldsymbol{X}_i,U_i\} \quad \text{and} \quad E[\{\log(Y_{ij})\}^2|\boldsymbol{X}_i,U_i].$$

As a prelude to obtaining these expressions we consider

$$X \sim \mathrm{Gamma}(\kappa,\lambda).$$

Then note that

$$E\{X\log(X)\} = \{\kappa\,\mathrm{digamma}(\kappa) + 1 - \kappa\log(\lambda)\}/\lambda$$

and

$$E[\{\log(X)\}^2] = \mathrm{trigamma}(\kappa) + \{\mathrm{digamma}(\kappa) - \log(\lambda)\}^2.$$

The first of these results leads to (with $\kappa=\psi$ and $\lambda = -\psi(\beta_0+\beta_1X_{ij}+U_i)$),

$$\begin{aligned}
&E\{Y_{ij}\log(Y_{ij})|\boldsymbol{X}_i,U_i\} \\
&= \frac{1}{\psi\{-(\beta_0+\beta_1X_{ij}+U_i)\}}\Big(\psi\mathrm{digamma}(\psi) + 1 - \psi\log\left[\psi\{-(\beta_0+\beta_1X_{ij}+U_i)\}\right]\Big) \\
&= \frac{1}{\psi\{-(\beta_0+\beta_1X_{ij}+U_i)\}}\Big[\psi\mathrm{digamma}(\psi) + 1 - \psi\log(\psi) - \psi\log\left\{-(\beta_0+\beta_1X_{ij}+U_i)\right\}\Big] \\
&= \frac{1}{\psi\{-(\beta_0+\beta_1X_{ij}+U_i)\}}\Big\{\psi\mathrm{digamma}(\psi) + 1 - \psi\log(\psi) + \psi b(\beta_0+\beta_1X_{ij}+U_i)\Big\} \\
&= b'(\beta_0+\beta_1X_{ij}+U_i)\Big\{\frac{1}{\psi} + \mathrm{digamma}(\psi) - \log(\psi) + b(\beta_0+\beta_1X_{ij}+U_i)\Big\} \\
&= b'(\beta_0+\beta_1X_{ij}+U_i)\Big\{\frac{1}{\psi} + 1 - \mathfrak{q}(\psi) + b(\beta_0+\beta_1X_{ij}+U_i)\Big\}.
\end{aligned}$$

The second of these results leads to (with $\kappa = \psi$ and $\lambda = -\psi(\beta_0 + \beta_1 X_{ij} + U_i)$)

$$\begin{aligned}
E[\{\log(Y_{ij})\}^2|\boldsymbol{X}_i, U_i] &= \text{trigamma}(\psi) + \Big(\text{digamma}(\psi) - \log\left[\psi\{-(\beta_0 + \beta_1 X_{ij} + U_i)\}\right]\Big)^2 \\
&= \text{trigamma}(\psi) + \big[\text{digamma}(\psi) - \log(\psi) - \log\{-(\beta_0 + \beta_1 X_{ij} + U_i)\}\big]^2 \\
&= \text{trigamma}(\psi) + \big\{1 - \mathfrak{q}(\psi) + b(\beta_0 + \beta_1 X_{ij} + U_i)\big\}^2 \\
&= \text{trigamma}(\psi) + \big\{b(\beta_0 + \beta_1 X_{ij} + U_i) + 1 - \mathfrak{q}(\psi)\big\}^2.
\end{aligned}$$

4.5.3.2 The Score of the Reciprocal Dispersion Parameter

The score of ψ is

$$\begin{aligned}
S_{3i} &\equiv \frac{\partial \log p_{\boldsymbol{Y}_i|\boldsymbol{X}_i}(\boldsymbol{Y}_i|\boldsymbol{X}_i)}{\partial \psi} \\
&= \sum_{j=1}^{n}\Big[\{Y_{ij}(\beta_0 + \beta_1 X_{ij}) + \log(Y_{ij})\} + \mathfrak{q}(\psi)\Big] \\
&\quad + \frac{\displaystyle\int_{-\infty}^{\infty} \exp\left[-\tfrac{u^2}{2\sigma^2} + \psi\sum_{j=1}^{n}\{Y_{ij}u - b(\beta_0 + \beta_1 X_{ij} + u)\}\right]\sum_{j=1}^{n}\{Y_{ij}u - b(\beta_0 + \beta_1 X_{ij} + u)\}\,du}{\displaystyle\int_{-\infty}^{\infty} \exp\left[-\tfrac{u^2}{2\sigma^2} + \psi\sum_{j=1}^{n}\{Y_{ij}u - b(\beta_0 + \beta_1 X_{ij} + u)\}\right]du} \\
&= \sum_{j=1}^{n}\Big[\{Y_{ij}(\beta_0 + \beta_1 X_{ij}) + \log(Y_{ij})\} + \mathfrak{q}(\psi)\Big] + \frac{\displaystyle\int_{-\infty}^{\infty} b_N(u)\,\exp\{-nh_N(u)\}\,du}{\displaystyle\int_{-\infty}^{\infty} b_D(u)\,\exp\{-nh_N(u)\}\,du}
\end{aligned}$$

where

$$h_N(u) \equiv -\frac{\psi}{n}\sum_{j=1}^{n}\Big\{Y_{ij}u - b(\beta_0 + \beta_1 X_{ij} + u)\Big\},$$

$$b_N(u) \equiv \sum_{j=1}^{n}\Big\{Y_{ij}u - b(\beta_0 + \beta_1 X_{ij} + u)\Big\}\exp\left(-\frac{u^2}{2\sigma^2}\right) \quad \text{and} \quad b_D(u) \equiv \exp\left(-\frac{u^2}{2\sigma^2}\right).$$

The First Term of the S_{3i} Integral

The first term on the right-hand side of (2.6) of Tierney et al. (1989) is

$$g(U_i^*)$$

where

$$g(u) \equiv b_N(u)/b_D(u) = \sum_{j=1}^{n}\Big\{Y_{ij}u - b(\beta_0 + \beta_1 X_{ij} + u)\Big\}.$$

Using steps similar to those given for approximating the score of $\boldsymbol{\beta}_{\text{B}}$ in Subsubsection 4.5.2.5, we get

$$g(U_i^*) = U_i Y_{i\bullet} - \mathcal{A}_i + O_P(1)$$

where

$$Y_{i\bullet} \equiv \sum_{j=1}^{n} Y_{ij} \quad \text{and} \quad \mathcal{A}_i \equiv \sum_{j=1}^{n} b(\beta_0 + \beta_1 X_{ij} + U_i).$$

Also define

$$\mathcal{A}'_i \equiv \sum_{j=1}^{n} b'(\beta_0 + \beta_1 X_{ij} + U_i).$$

4.5.3.3 Computing the Fisher Information Block for the Reciprocal Dispersion Parameter

Define

$$\mathcal{B}_i \equiv \sum_{j=1}^{n} \Big\{ Y_{ij}(\beta_0 + \beta_1 X_{ij}) + \log(Y_{ij}) + \mathfrak{q}(\psi) \Big\}.$$

Then

$$S_{3i} = \mathcal{B}_i + U_i Y_{i\bullet} - \mathcal{A}_i + O_P(1).$$

Therefore,

$$E(S_{3i}^2 | \boldsymbol{X}_i) \approx T_A + T_B + T_C + T_D + T_E + T_F$$

where

$$\begin{aligned} T_A &= E(\mathcal{B}_i^2 | \boldsymbol{X}_i), \\ T_B &= E(U_i^2 Y_{i\bullet}^2 | \boldsymbol{X}_i) \\ T_C &= E(\mathcal{A}_i^2 | \boldsymbol{X}_i), \\ T_D &= 2E(\mathcal{B}_i U_i Y_{i\bullet} | \boldsymbol{X}_i), \\ T_E &= -2E(\mathcal{A}_i \mathcal{B}_i | \boldsymbol{X}_i), \\ \text{and} \quad T_F &= -2E(U_i \mathcal{A}_i Y_{i\bullet} | \boldsymbol{X}_i). \end{aligned}$$

Treatment of T_A

In the Gamma case

$$b'(x) = -1/x \quad \text{and} \quad c(x) = \log(x).$$

Also,

$$\frac{\partial d(Y_{ij}, \psi)}{\partial \psi} = \mathfrak{q}(\psi).$$

Hence

$$\mathcal{B}_i = \sum_{j=1}^{n} \big\{ Y_{ij}(\beta_0 + \beta_1 X_{ij}) + \log(Y_{ij}) + \mathfrak{q}(\psi) \big\}$$

and

$$\mathcal{B}_i^2 = \sum_{j=1}^{n}\sum_{j'=1}^{n}\left[\{Y_{ij}(\beta_0+\beta_1 X_{ij})+\log(Y_{ij})\}+\mathfrak{q}(\psi)\right]\left[\{Y_{ij'}(\beta_0+\beta_1 X_{ij})+\log(Y_{ij'})\}+\mathfrak{q}(\psi)\right]$$

$$= \sum_{j=1}^{n}\left[\{Y_{ij}(\beta_0+\beta_1 X_{ij})+\log(Y_{ij})\}+\mathfrak{q}(\psi)\right]^2$$

$$+\sum_{j\neq j'}^{n}\left[\{Y_{ij}(\beta_0+\beta_1 X_{ij})+\log(Y_{ij})\}+\mathfrak{q}(\psi)\right]\left[\{Y_{ij'}(\beta_0+\beta_1 X_{ij'})+\log(Y_{ij'})\}+\mathfrak{q}(\psi)\right].$$

Therefore

$$T_A = E(\mathcal{B}_i^2|\boldsymbol{X}_i) = \mathfrak{r}_1(\boldsymbol{X}_i)+\mathfrak{r}_2(\boldsymbol{X}_i)$$

where

$$\mathfrak{r}_1(\boldsymbol{X}_i) \equiv \sum_{j\neq j'}^{n} E\left(\left[Y_{ij}(\beta_0+\beta_1 X_{ij})+\log(Y_{ij})+\mathfrak{q}(\psi)\right]\left[Y_{ij'}(\beta_0+\beta_1 X_{ij'})+\log(Y_{ij'})+\mathfrak{q}(\psi)\right]\middle|\boldsymbol{X}_i\right)$$

and

$$\mathfrak{r}_2(\boldsymbol{X}_i) \equiv \sum_{j=1}^{n} E\left[\left\{Y_{ij}(\beta_0+\beta_1 X_{ij})+\log(Y_{ij})+\mathfrak{q}(\psi)\right\}^2\middle|\boldsymbol{X}_i\right].$$

Treatment of $\mathfrak{r}_1(\boldsymbol{X}_i)$

The (j,j')th term in the $\mathfrak{r}_1(\boldsymbol{X}_i)$ summation is

$$E\left\{E\left(\left[Y_{ij}(\beta_0+\beta_1 X_{ij})+\log(Y_{ij})+\mathfrak{q}(\psi)\right]\left[Y_{ij'}(\beta_0+\beta_1 X_{ij'})+\log(Y_{ij'})+\mathfrak{q}(\psi)\right]\middle|\boldsymbol{X}_i,U_i\right)\middle|\boldsymbol{X}_i\right\}$$

$$= E\left\{E\left(\left[Y_{ij}(\beta_0+\beta_1 X_{ij})+\log(Y_{ij})+\mathfrak{q}(\psi)\right]\middle|\boldsymbol{X}_i,U_i\right)\right.$$

$$\left.\times E\left(\left[Y_{ij'}(\beta_0+\beta_1 X_{ij'})+\log(Y_{ij'})+\mathfrak{q}(\psi)\right]\middle|\boldsymbol{X}_i,U_i\right)\middle|\boldsymbol{X}_i\right\}$$

Next note that

$$E\left(\left[Y_{ij}(\beta_0+\beta_1 X_{ij})+\log(Y_{ij})+\mathfrak{q}(\psi)\right]\middle|\boldsymbol{X}_i,U_i\right)$$

$$= b'(\beta_0+\beta_1 X_{ij}+U_i)(\beta_0+\beta_1 X_{ij})+b(\beta_0+\beta_1 X_{ij}+U_i)+1-\mathfrak{q}(\psi)+\mathfrak{q}(\psi)$$

$$= b'(\beta_0+\beta_1 X_{ij}+U_i)(\beta_0+\beta_1 X_{ij}+U_i)-U_i b'(\beta_0+\beta_1 X_{ij}+U_i)+b(\beta_0+\beta_1 X_{ij}+U_i)+1$$

$$= b(\beta_0+\beta_1 X_{ij}+U_i)-U_i b'(\beta_0+\beta_1 X_{ij}+U_i).$$

Hence,

$$\mathfrak{r}_1(\boldsymbol{X}_i) = \sum_{j\neq j'}^{n} E\Big[\{b(\beta_0+\beta_1 X_{ij}+U_i)-U_i b'(\beta_0+\beta_1 X_{ij}+U_i)\}$$

$$\times\{b(\beta_0+\beta_1 X_{ij'}+U_i)-U_i b'(\beta_0+\beta_1 X_{ij'}+U_i)\}|\boldsymbol{X}_i\Big].$$

Treatment of $\mathfrak{r}_2(\boldsymbol{X}_i)$

The jth term in the $\mathfrak{r}_2(\boldsymbol{X}_i)$ summation is

$$E\left(E\left[\left\{Y_{ij}(\beta_0+\beta_1X_{ij})+\log(Y_{ij})+\mathfrak{q}(\psi)\right\}^2\middle|\boldsymbol{X}_i,U_i\right]\middle|\boldsymbol{X}_i\right).$$

Next we expand out the inner conditional expectation:

$$\begin{aligned}
&E\left[\left\{Y_{ij}(\beta_0+\beta_1X_{ij})+\log(Y_{ij})+\mathfrak{q}(\psi)\right\}^2\middle|\boldsymbol{X}_i,U_i\right]\\
&=(\beta_0+\beta_1X_{ij})^2E\left(Y_{ij}^2|\boldsymbol{X}_i,U_i\right)+E\left[\{\log(Y_{ij})\}^2|\boldsymbol{X}_i,U_i\right]+\mathfrak{q}(\psi)^2\\
&\qquad+2(\beta_0+\beta_1X_{ij})E\left\{Y_{ij}\log(Y_{ij})|\boldsymbol{X}_i,U_i\right\}\\
&\qquad+2(\beta_0+\beta_1X_{ij})\mathfrak{q}(\psi)E\left(Y_{ij}|\boldsymbol{X}_i,U_i\right)+2\mathfrak{q}(\psi)\,E\left\{\log(Y_{ij})|\boldsymbol{X}_i,U_i\right\}\\
&=(\beta_0+\beta_1X_{ij})^2\left(\frac{1}{\psi}+1\right)\{b'(\beta_0+\beta_1X_{ij}+U_i)\}^2\\
&\quad+\text{trigamma}(\psi)+\left\{b(\beta_0+\beta_1X_{ij}+U_i)+1-\mathfrak{q}(\psi)\right\}^2\\
&\quad+\mathfrak{q}(\psi)^2+2(\beta_0+\beta_1X_{ij})b'(\beta_0+\beta_1X_{ij}+U_i)\left\{\frac{1}{\psi}+1-\mathfrak{q}(\psi)+b(\beta_0+\beta_1X_{ij}+U_i)\right\}\\
&\qquad+2(\beta_0+\beta_1X_{ij})\mathfrak{q}(\psi)b'(\beta_0+\beta_1X_{ij}+U_i)+2\mathfrak{q}(\psi)\{b(\beta_0+\beta_1\,X_{ij}+U_i)+1-\mathfrak{q}(\psi)\}.
\end{aligned}$$

The first of these terms is

$$\begin{aligned}
&\left(\frac{1}{\psi}+1\right)(\beta_0+\beta_1X_{ij}+U_i-U_i)^2\{b'(\beta_0+\beta_1X_{ij}+U_i)\}^2\\
&\quad=\left(\frac{1}{\psi}+1\right)\{(\beta_0+\beta_1X_{ij}+U_i)^2-2U_i(\beta_0+\beta_1X_{ij}+U_i)+U_i^2\}\{b'(\beta_0+\beta_1X_{ij}+U_i)\}^2.\\
&\quad=\left(\frac{1}{\psi}+1\right)\left[1+2U_ib'(\beta_0+\beta_1X_{ij}+U_i)+U_i^2\{b'(\beta_0+\beta_1X_{ij}+U_i)\}^2\right].
\end{aligned}$$

The sixth of these terms is

$$2(\beta_0+\beta_1X_{ij}+U_i-U_i)\mathfrak{q}(\psi)b'(\beta_0+\beta_1X_{ij}+U_i)=-2\mathfrak{q}(\psi)\{1+U_ib'(\beta_0+\beta_1X_{ij}+U_i)\}.$$

With similar steps, the fifth of these terms is

$$-2\left\{\frac{1}{\psi}+1-\mathfrak{q}(\psi)+b(\beta_0+\beta_1X_{ij}+U_i)\right\}\{1+U_ib'(\beta_0+\beta_1X_{ij}+U_i)\}.$$

The sum of the fifth and sixth terms is then

$$-2\left\{\frac{1}{\psi}+1+b(\beta_0+\beta_1X_{ij}+U_i)\right\}\{1+U_ib'(\beta_0+\beta_1X_{ij}+U_i)\}.$$

Assembling the above results we have

$$
\begin{aligned}
T_A = {} & n\,\text{trigamma}(\psi) + n\{\mathfrak{q}(\psi)\}^2 \\
& + \sum_{j\neq j'}^{n} E\Big[\{b(\beta_0+\beta_1 X_{ij}+U_i) - U_i b'(\beta_0+\beta_1 X_{ij}+U_i)\}\{b(\beta_0+\beta_1 X_{ij'}+U_i) \\
& \quad -U_i b'(\beta_0+\beta_1 X_{ij'}+U_i)\}|\boldsymbol{X}_i\Big] \\
& + \left(\frac{1}{\psi}+1\right)\sum_{j=1}^{n} E\Big(\big[1+2U_i b'(\beta_0+\beta_1 X_{ij}+U_i) + U_i^2\{b'(\beta_0+\beta_1 X_{ij}+U_i)\}^2\big]\Big|\boldsymbol{X}_i\Big) \\
& + \sum_{j=1}^{n} E\Big(\big\{b(\beta_0+\beta_1 X_{ij}+U_i)+1-\mathfrak{q}(\psi)\big\}^2\Big|\boldsymbol{X}_i\Big) \\
& - 2\sum_{j=1}^{n} E\left[\left\{\left(\frac{1}{\psi}+1\right)+b(\beta_0+\beta_1 X_{ij}+U_i)\right\}\{1+U_i b'(\beta_0+\beta_1 X_{ij}+U_i)\}\Big|\boldsymbol{X}_i\right] \\
& + 2\mathfrak{q}(\psi)\sum_{j=1}^{n} E\Big[\{b(\beta_0+\beta_1\, X_{ij}+U_i)+1-\mathfrak{q}(\psi)\}\Big|\boldsymbol{X}_i\Big].
\end{aligned}
$$

A next useful step (for cancellation purposes) is to expand out the second, third, fourth, fifth and sixth terms of this expression for T_A.

Expansion of the second term of T_A

For the second term of T_A we have

$$
\begin{aligned}
& \sum_{j\neq j'}^{n} E\Big[\{b(\beta_0+\beta_1 X_{ij}+U_i) - U_i b'(\beta_0+\beta_1 X_{ij}+U_i)\} \\
& \qquad \times\{b(\beta_0+\beta_1 X_{ij'}+U_i) - U_i b'(\beta_0+\beta_1 X_{ij'}+U_i)\}|\boldsymbol{X}_i\Big] \\
& = \sum_{j\neq j'}^{n} E\Big\{b(\beta_0+\beta_1 X_{ij}+U_i)b(\beta_0+\beta_1 X_{ij'}+U_i)|\boldsymbol{X}_i\Big\} \\
& \quad - \sum_{j\neq j'}^{n} E\Big\{U_i b(\beta_0+\beta_1 X_{ij}+U_i)b'(\beta_0+\beta_1 X_{ij'}+U_i)|\boldsymbol{X}_i\Big\} \\
& \quad - \sum_{j\neq j'}^{n} E\Big\{U_i b'(\beta_0+\beta_1 X_{ij}+U_i)b(\beta_0+\beta_1 X_{ij'}+U_i)|\boldsymbol{X}_i\Big\} \\
& \quad + \sum_{j\neq j'}^{n} E\Big\{U_i^2 b'(\beta_0+\beta_1 X_{ij}+U_i)b'(\beta_0+\beta_1 X_{ij'}+U_i)|\boldsymbol{X}_i\Big\}.
\end{aligned}
$$

Expansion of the third term of T_A

For the third term of T_A we have

$$\left(\frac{1}{\psi}+1\right)\sum_{j=1}^{n}E\Big(\big[1+2U_ib'(\beta_0+\beta_1X_{ij}+U_i)+U_i^2\{b'(\beta_0+\beta_1X_{ij}+U_i)\}^2\big]\Big|\boldsymbol{X}_i\Big)$$

$$=n\left(\frac{1}{\psi}+1\right)+2\left(\frac{1}{\psi}+1\right)\sum_{j=1}^{n}E\Big\{U_ib'(\beta_0+\beta_1X_{ij}+U_i)\Big|\boldsymbol{X}_i\Big\}$$

$$+\left(\frac{1}{\psi}+1\right)\sum_{j=1}^{n}E\Big[U_i^2\{b'(\beta_0+\beta_1X_{ij}+U_i)\}^2\Big|\boldsymbol{X}_i\Big].$$

Expansion of the fourth term of T_A

For the fourth term of T_A we have

$$\sum_{j=1}^{n}E\Big(\big\{b(\beta_0+\beta_1X_{ij}+U_i)+1-\mathfrak{q}(\psi)\big\}^2\Big|\boldsymbol{X}_i\Big)$$

$$=\sum_{j=1}^{n}E\Big(\{b(\beta_0+\beta_1X_{ij}+U_i)\}^2\Big|\boldsymbol{X}_i\Big)+2\{1-\mathfrak{q}(\psi)\}\sum_{j=1}^{n}E\Big(b(\beta_0+\beta_1X_{ij}+U_i)\Big|\boldsymbol{X}_i\Big)$$
$$+n\{1-\mathfrak{q}(\psi)\}^2$$

$$=\sum_{j=1}^{n}E\Big(\{b(\beta_0+\beta_1X_{ij}+U_i)\}^2\Big|\boldsymbol{X}_i\Big)+2\sum_{j=1}^{n}E\Big(b(\beta_0+\beta_1X_{ij}+U_i)\Big|\boldsymbol{X}_i\Big)$$
$$-2\mathfrak{q}(\psi)\sum_{j=1}^{n}E\Big\{b(\beta_0+\beta_1X_{ij}+U_i)\Big|\boldsymbol{X}_i\Big\}+n\{\mathfrak{q}(\psi)\}^2-2n\mathfrak{q}(\psi)+n.$$

Expansion of the fifth term of T_A

For the fifth term of T_A we have

$$-2\sum_{j=1}^{n}E\left[\left\{\left(\frac{1}{\psi}+1\right)+b(\beta_0+\beta_1X_{ij}+U_i)\right\}\{1+U_ib'(\beta_0+\beta_1X_{ij}+U_i)\}\Big|\boldsymbol{X}_i\right]$$

$$=-2n\left(\frac{1}{\psi}+1\right)-2\left(\frac{1}{\psi}+1\right)\sum_{j=1}^{n}E\left\{U_ib'(\beta_0+\beta_1X_{ij}+U_i)\Big|\boldsymbol{X}_i\right\}$$

$$-2\sum_{j=1}^{n}E\left\{b(\beta_0+\beta_1X_{ij}+U_i)\Big|\boldsymbol{X}_i\right\}-2\sum_{j=1}^{n}E\left\{U_ib(\beta_0+\beta_1X_{ij}+U_i)b'(\beta_0+\beta_1X_{ij}+U_i)\Big|\boldsymbol{X}_i\right\}.$$

Expansion of the sixth term of T_A

For the sixth term of T_A we have

$$
\begin{aligned}
&2\mathfrak{q}(\psi)\sum_{j=1}^{n} E\Big[\{b(\beta_0+\beta_1 X_{ij}+U_i)+1-\mathfrak{q}(\psi)\}\Big|\boldsymbol{X}_i\Big] \\
&= 2\mathfrak{q}(\psi)\sum_{j=1}^{n} E\big[b(\beta_0+\beta_1 X_{ij}+U_i)\big|\boldsymbol{X}_i\big] + 2n\mathfrak{q}(\psi)\{1-\mathfrak{q}(\psi)\} \\
&= 2\mathfrak{q}(\psi)\sum_{j=1}^{n} E\big[b(\beta_0+\beta_1 X_{ij}+U_i)\big|\boldsymbol{X}_i\big] + 2n\mathfrak{q}(\psi) - 2n\{\mathfrak{q}(\psi)\}^2.
\end{aligned}
$$

The fully expanded version of T_A is as follows:

$$
\begin{aligned}
T_A = n &\left\{\text{trigamma}(\psi)-\frac{1}{\psi}\right\} \\
&+\sum_{j\neq j'}^{n} E\Big\{b(\beta_0+\beta_1 X_{ij}+U_i)b(\beta_0+\beta_1 X_{ij'}+U_i)|\boldsymbol{X}_i\Big\} \\
&-\sum_{j\neq j'}^{n} E\Big\{U_i b(\beta_0+\beta_1 X_{ij}+U_i)b'(\beta_0+\beta_1 X_{ij'}+U_i)|\boldsymbol{X}_i\Big\} \\
&-\sum_{j\neq j'}^{n} E\Big\{U_i b'(\beta_0+\beta_1 X_{ij}+U_i)b(\beta_0+\beta_1 X_{ij'}+U_i)|\boldsymbol{X}_i\Big\} \\
&+\sum_{j\neq j'}^{n} E\Big\{U_i^2 b'(\beta_0+\beta_1 X_{ij}+U_i)b'(\beta_0+\beta_1 X_{ij'}+U_i)|\boldsymbol{X}_i\Big\} \\
&+\left(\frac{1}{\psi}+1\right)\sum_{j=1}^{n} E\Big[U_i^2\{b'(\beta_0+\beta_1 X_{ij}+U_i)\}^2\Big|\boldsymbol{X}_i\Big] \\
&+\sum_{j=1}^{n} E\Big(\{b(\beta_0+\beta_1 X_{ij}+U_i)\}^2\Big|\boldsymbol{X}_i\Big) \\
&-2\sum_{j=1}^{n} E\left\{U_i b(\beta_0+\beta_1 X_{ij}+U_i)b'(\beta_0+\beta_1 X_{ij}+U_i)\Big|\boldsymbol{X}_i\right\}.
\end{aligned}
$$

Treatment of T_B

We have

$$
E\left(U_i^2 Y_{i\bullet}^2\Big|\boldsymbol{X}_i\right) = E\left\{E\left(U_i^2 Y_{i\bullet}^2\Big|\boldsymbol{X}_i, U_i\right)\Big|\boldsymbol{X}_i\right\} = E\left\{U_i^2\, E\left(Y_{i\bullet}^2\Big|\boldsymbol{X}_i, U_i\right)\Big|\boldsymbol{X}_i\right\}.
$$

Recall that

$$Y_{i\bullet} = \sum_{j=1}^{n} Y_{ij} \quad \text{implying that} \quad Y_{i\bullet}^2 = \sum_{j \neq j'}^{n} Y_{ij} Y_{ij'} + \sum_{j=1}^{n} Y_{ij}^2.$$

Therefore

$$\begin{aligned}
&E(Y_{i\bullet}^2 | \boldsymbol{X}_i, U_i) \\
&= \sum_{j \neq j'}^{n} E(Y_{ij} Y_{ij'} | \boldsymbol{X}_i, U_i) + \sum_{j=1}^{n} E(Y_{ij}^2 | \boldsymbol{X}_i, U_i) \\
&= \sum_{j \neq j'}^{n} b'(\beta_0 + \beta_1 X_{ij} + U_i) b'(\beta_0 + \beta_1 X_{ij'} + U_i) + \left(\frac{1}{\psi} + 1\right) \sum_{j=1}^{n} \{b'(\beta_0 + \beta_1 X_{ij} + U_i)\}^2.
\end{aligned}$$

This implies that

$$\begin{aligned}
T_B = &\sum_{j \neq j'}^{n} E\big\{U_i^2 b'(\beta_0 + \beta_1 X_{ij} + U_i) b'(\beta_0 + \beta_1 X_{ij'} + U_i) | \boldsymbol{X}_i\big\} \\
&+ \left(\frac{1}{\psi} + 1\right) \sum_{j=1}^{n} E\big[U_i^2 \{b'(\beta_0 + \beta_1 X_{ij} + U_i)\}^2 | \boldsymbol{X}_i\big].
\end{aligned}$$

Treatment of T_C

We have

$$T_C = E(\mathcal{A}_i^2 | \boldsymbol{X}_i)$$

which implies that

$$T_C = \sum_{j=1}^{n} \sum_{j'=1}^{n} E\{b(\beta_0 + \beta_1 X_{ij} + U_i) b(\beta_0 + \beta_1 X_{ij'} + U_i) | \boldsymbol{X}_i\}.$$

Breaking this up into "diagonal" and "off-diagonal" components, we get

$$\begin{aligned}
T_C = &\sum_{j \neq j'}^{n} E\big\{b(\beta_0 + \beta_1 X_{ij} + U_i) b(\beta_0 + \beta_1 X_{ij'} + U_i) | \boldsymbol{X}_i\big\} \\
&+ \sum_{j=1}^{n} E\big[\{b(\beta_0 + \beta_1 X_{ij} + U_i)\}^2 | \boldsymbol{X}_i\big].
\end{aligned}$$

Treatment of T_D

First recall that

$$2E(\mathcal{B}_i U_i Y_{i\bullet} | \boldsymbol{X}_i)$$

where

$$\mathcal{B}_i \equiv \sum_{j=1}^{n} \Bigg[\{Y_{ij}(\beta_0 + \beta_1 X_{ij}) + \log(Y_{ij})\} + \mathfrak{q}(\psi) \Bigg].$$

Therefore

$$\begin{aligned} 2\mathcal{B}_i U_i Y_{i\bullet} &= 2U_i \sum_{j=1}^{n} \sum_{j'=1}^{n} \Bigg[\{Y_{ij}(\beta_0 + \beta_1 X_{ij}) + \log(Y_{ij})\} + \mathfrak{q}(\psi) \Bigg] Y_{ij'} \\ &= 2U_i \sum_{j \neq j'}^{n} \Bigg[\{Y_{ij}(\beta_0 + \beta_1 X_{ij}) + \log(Y_{ij})\} + \mathfrak{q}(\psi) \Bigg] Y_{ij'} \\ &\quad + 2U_i \sum_{j=1}^{n} \Bigg[\left\{ Y_{ij}^2(\beta_0 + \beta_1 X_{ij}) + Y_{ij} \log(Y_{ij}) \right\} + Y_{ij} \mathfrak{q}(\psi) \Bigg]. \end{aligned}$$

It follows that

$$2E(\mathcal{B}_i U_i Y_{i\bullet} | \boldsymbol{X}_i) = \mathfrak{r}_3(\boldsymbol{X}_i) + \mathfrak{r}_4(\boldsymbol{X}_i)$$

where

$$\mathfrak{r}_3(\boldsymbol{X}_i) \equiv 2 \sum_{j \neq j'}^{n} E\left(U_i \Bigg[\{Y_{ij}(\beta_0 + \beta_1 X_{ij}) + \log(Y_{ij})\} + \mathfrak{q}(\psi) \Bigg] Y_{ij'} \Bigg| \boldsymbol{X}_i \right)$$

and

$$\mathfrak{r}_4(\boldsymbol{X}_i) \equiv 2 \sum_{j=1}^{n} E\left(U_i \Bigg[\left\{ Y_{ij}^2(\beta_0 + \beta_1 X_{ij}) + Y_{ij} \log(Y_{ij}) \right\} + Y_{ij} \mathfrak{q}(\psi) \Bigg] \Bigg| \boldsymbol{X}_i \right).$$

Treatment of $\mathfrak{r}_3(\boldsymbol{X}_i)$

The (j, j')th term in the $\mathfrak{r}_3(\boldsymbol{X}_i)$ summation is

$$\begin{aligned} &2E \left\{ E \left(U_i \Bigg[Y_{ij}(\beta_0 + \beta_1 X_{ij}) + \log(Y_{ij}) + \mathfrak{q}(\psi) \Bigg] Y_{ij'} \Bigg| \boldsymbol{X}_i, U_i \right) \Bigg| \boldsymbol{X}_i \right\} \\ &= 2E \left\{ U_i \, E \left(\Bigg[Y_{ij}(\beta_0 + \beta_1 X_{ij}) + \log(Y_{ij}) + \mathfrak{q}(\psi) \Bigg] Y_{ij'} \Bigg| \boldsymbol{X}_i, U_i \right) \Bigg| \boldsymbol{X}_i \right\} \\ &= 2E \left\{ U_i \, E \left(\Bigg[Y_{ij}(\beta_0 + \beta_1 X_{ij}) + \log(Y_{ij}) + \mathfrak{q}(\psi) \Bigg] \Bigg| \boldsymbol{X}_i, U_i \right) E \left(Y_{ij'} \Bigg| \boldsymbol{X}_i, U_i \right) \Bigg| \boldsymbol{X}_i \right\}. \end{aligned}$$

As stated earlier in this document,

$$E\left(\Big[Y_{ij}(\beta_0 + \beta_1 X_{ij}) + \log(Y_{ij}) + \mathfrak{q}(\psi) \Big] \Big| \boldsymbol{X}_i, U_i \right) = b(\beta_0 + \beta_1 X_{ij} + U_i) - U_i b'(\beta_0 + \beta_1 X_{ij} + U_i).$$

Also,

$$E\left(Y_{ij'}\middle|\boldsymbol{X}_i, U_i\right) = b'(\beta_0 + \beta_1 X_{ij'} + U_i)$$

from which it follows that the (j, j')th term in the $\mathfrak{r}_3(\boldsymbol{X}_i)$ summation is

$$2E\left[U_i\big\{b(\beta_0 + \beta_1 X_{ij} + U_i) - U_i b'(\beta_0 + \beta_1 X_{ij} + U_i)\big\}b'(\beta_0 + \beta_1 X_{ij'} + U_i)\middle|\boldsymbol{X}_i\right].$$

Hence,

$$\mathfrak{r}_3(\boldsymbol{X}_i) = 2\sum_{j\neq j'}^{n} E\left[U_i\big\{b(\beta_0 + \beta_1 X_{ij} + U_i) - U_i b'(\beta_0 + \beta_1 X_{ij} + U_i)\big\}b'(\beta_0 + \beta_1 X_{ij'} + U_i)\middle|\boldsymbol{X}_i\right]$$

Treatment of $\mathfrak{r}_4(\boldsymbol{X}_i)$

The jth term in the $\mathfrak{r}_4(\boldsymbol{X}_i)$ summation is

$$\begin{aligned}
&2E\left(E\left[U_i\Big\{Y_{ij}^2(\beta_0 + \beta_1 X_{ij}) + Y_{ij}\log(Y_{ij}) + Y_{ij}\mathfrak{q}(\psi)\Big\}\middle|\boldsymbol{X}_i, U_i\right]\middle|\boldsymbol{X}_i\right)\\
&= 2E\left(U_i E\left[\Big\{Y_{ij}^2(\beta_0 + \beta_1 X_{ij}) + Y_{ij}\log(Y_{ij}) + Y_{ij}\mathfrak{q}(\psi)\Big\}\middle|\boldsymbol{X}_i, U_i\right]\middle|\boldsymbol{X}_i\right).
\end{aligned}$$

Next note that

$$\begin{aligned}
E\left[Y_{ij}^2(\beta_0 + \beta_1 X_{ij})\middle|\boldsymbol{X}_i, U_i\right] &= \left(\frac{1}{\psi} + 1\right)(\beta_0 + \beta_1 X_{ij})\{b'(\beta_0 + \beta_1 X_{ij} + U_i)\}^2\\
&= \left(\frac{1}{\psi} + 1\right)(\beta_0 + \beta_1 X_{ij} + U_i)\{b'(\beta_0 + \beta_1 X_{ij} + U_i)\}^2\\
&\quad - \left(\frac{1}{\psi} + 1\right)U_i\{b'(\beta_0 + \beta_1 X_{ij} + U_i)\}^2\\
&= -\left(\frac{1}{\psi} + 1\right)\left[b'(\beta_0 + \beta_1 X_{ij} + U_i) + U_i\{b'(\beta_0 + \beta_1 X_{ij} + U_i)\}^2\right].
\end{aligned}$$

Also, remember that

$$E\{Y_{ij}\log(Y_{ij})|\boldsymbol{X}_i, U_i\} = b'(\beta_0 + \beta_1 X_{ij} + U_i)\Big\{\frac{1}{\psi} + 1 - \mathfrak{q}(\psi) + b(\beta_0 + \beta_1 X_{ij} + U_i)\Big\}$$

and

$$\mathfrak{q}(\psi)E\{Y_{ij}|\boldsymbol{X}_i, U_i\} = \mathfrak{q}(\psi)b'(\beta_0 + \beta_1 X_{ij} + U_i),$$

which means that

$$E\{Y_{ij}\log(Y_{ij}) + \mathfrak{q}(\psi)Y_{ij}|\boldsymbol{X}_i, U_i\} = b'(\beta_0 + \beta_1 X_{ij} + U_i)\Big\{\frac{1}{\psi} + 1 + b(\beta_0 + \beta_1 X_{ij} + U_i)\Big\}.$$

This means that

$$
\begin{aligned}
&E\left[\left\{Y_{ij}^2(\beta_0+\beta_1X_{ij})+Y_{ij}\log(Y_{ij})+Y_{ij}\mathfrak{q}(\psi)\right\}\middle|\boldsymbol{X}_i,U_i\right]\\
&\quad=-\left(\frac{1}{\psi}+1\right)b'(\beta_0+\beta_1X_{ij}+U_i)-\left(\frac{1}{\psi}+1\right)U_i\{b'(\beta_0+\beta_1X_{ij}+U_i)\}^2\\
&\qquad+\left(\frac{1}{\psi}+1\right)b'(\beta_0+\beta_1X_{ij}+U_i)+b(\beta_0+\beta_1X_{ij}+U_i)b'(\beta_0+\beta_1X_{ij}+U_i)\\
&\qquad=b(\beta_0+\beta_1X_{ij}+U_i)b'(\beta_0+\beta_1X_{ij}+U_i)-\left(\frac{1}{\psi}+1\right)U_i\{b'(\beta_0+\beta_1X_{ij}+U_i)\}^2.
\end{aligned}
$$

Hence, the jth term in the $\mathfrak{r}_4(\boldsymbol{X}_i)$ summation is

$$
2E\Big\{U_ib(\beta_0+\beta_1X_{ij}+U_i)b'(\beta_0+\beta_1X_{ij}+U_i)\Big|\boldsymbol{X}_i\Big\}-2\left(\frac{1}{\psi}+1\right)E\Big[U_i^2\{b'(\beta_0+\beta_1X_{ij}+U_i)\}^2\Big|\boldsymbol{X}_i\Big].
$$

Putting all of this together, we arrive at

$$
\begin{aligned}
T_D=&\,2\sum_{j\neq j'}^{n}E\left[U_i\big\{b(\beta_0+\beta_1X_{ij}+U_i)-U_ib'(\beta_0+\beta_1X_{ij}+U_i)\big\}b'(\beta_0+\beta_1X_{ij'}+U_i)\middle|\boldsymbol{X}_i\right]\\
&+2\sum_{j=1}^{n}E\Big\{U_ib(\beta_0+\beta_1X_{ij}+U_i)b'(\beta_0+\beta_1X_{ij}+U_i)\Big|\boldsymbol{X}_i\Big\}\\
&-2\left(\frac{1}{\psi}+1\right)\sum_{j=1}^{n}E\Big[U_i^2\{b'(\beta_0+\beta_1X_{ij}+U_i)\}^2\Big|\boldsymbol{X}_i\Big].
\end{aligned}
$$

However, note that the first term is

$$
\begin{aligned}
&2\sum_{j\neq j'}^{n}E\left[U_i\big\{b(\beta_0+\beta_1X_{ij}+U_i)-U_ib'(\beta_0+\beta_1X_{ij}+U_i)\big\}b'(\beta_0+\beta_1X_{ij'}+U_i)\middle|\boldsymbol{X}_i\right]\\
&\quad=2\sum_{j\neq j'}^{n}E\left\{U_ib(\beta_0+\beta_1X_{ij}+U_i)b'(\beta_0+\beta_1X_{ij'}+U_i)\middle|\boldsymbol{X}_i\right\}\\
&\qquad-2\sum_{j\neq j'}^{n}E\left\{U_i^2b'(\beta_0+\beta_1X_{ij}+U_i)b'(\beta_0+\beta_1X_{ij'}+U_i)\middle|\boldsymbol{X}_i\right\}.
\end{aligned}
$$

Therefore

$$\begin{aligned}T_D = 2\sum_{j\neq j'}^{n} E\left\{U_i b(\beta_0+\beta_1 X_{ij}+U_i)b'(\beta_0+\beta_1 X_{ij'}+U_i)\Big|\boldsymbol{X}_i\right\}\\ -2\sum_{j\neq j'}^{n} E\left\{U_i^2 b'(\beta_0+\beta_1 X_{ij}+U_i)b'(\beta_0+\beta_1 X_{ij'}+U_i)\Big|\boldsymbol{X}_i\right\}\\ +2\sum_{j=1}^{n} E\Big\{U_i b(\beta_0+\beta_1 X_{ij}+U_i)b'(\beta_0+\beta_1 X_{ij}+U_i)\Big|\boldsymbol{X}_i\Big\}\\ -2\left(\frac{1}{\psi}+1\right)\sum_{j=1}^{n} E\Big\{U_i^2\{b'(\beta_0+\beta_1 X_{ij}+U_i)\}^2\Big|\boldsymbol{X}_i\Big\}.\end{aligned}$$

Treatment of T_E

Recall that

$$T_E \equiv -2E(\mathcal{A}_i\mathcal{B}_i|\boldsymbol{X}_i)$$

where

$$\mathcal{A}_i \equiv \sum_{j=1}^{n} b(\beta_0+\beta_1 X_{ij}+U_i) \quad \text{and} \quad \mathcal{B}_i \equiv \sum_{j=1}^{n}\Big\{Y_{ij}(\beta_0+\beta_1 X_{ij})+\log(Y_{ij})+\mathfrak{q}(\psi)\Big\}.$$

Therefore

$$\begin{aligned}T_E &= -2\sum_{j=1}^{n}\sum_{j'=1}^{n} E\Big[b(\beta_0+\beta_1 X_{ij}+U_i)\Big\{Y_{ij'}(\beta_0+\beta_1 X_{ij'})+\log(Y_{ij'})+\mathfrak{q}(\psi)\Big\}\Big|\boldsymbol{X}_i\Big]\\ &= -2\sum_{j=1}^{n}\sum_{j'=1}^{n} E\left(E\Big[b(\beta_0+\beta_1 X_{ij}+U_i)\Big\{Y_{ij'}(\beta_0+\beta_1 X_{ij'})+\log(Y_{ij'})+\mathfrak{q}(\psi)\Big\}\Big|\boldsymbol{X}_i,U_i\Big]\Big|\boldsymbol{X}_i\right)\\ &= -2\sum_{j=1}^{n}\sum_{j'=1}^{n} E\left(b(\beta_0+\beta_1 X_{ij}+U_i)E\Big[Y_{ij'}(\beta_0+\beta_1 X_{ij'})+\log(Y_{ij'})+\mathfrak{q}(\psi)\Big|\boldsymbol{X}_i,U_i\Big]\Big|\boldsymbol{X}_i\right)\\ &= -2\sum_{j=1}^{n}\sum_{j'=1}^{n} E\left(b(\beta_0+\beta_1 X_{ij}+U_i)\{b(\beta_0+\beta_1 X_{ij'}+U_i)-U_i b'(\beta_0+\beta_1 X_{ij'}+U_i)\}\Big|\boldsymbol{X}_i\right)\\ &= -2\sum_{j=1}^{n}\sum_{j'=1}^{n} E\Big\{b(\beta_0+\beta_1 X_{ij}+U_i)b(\beta_0+\beta_1 X_{ij'}+U_i)\Big|\boldsymbol{X}_i\Big\}\\ &\quad +2\sum_{j=1}^{n}\sum_{j'=1}^{n} E\Big\{U_i b(\beta_0+\beta_1 X_{ij}+U_i)b'(\beta_0+\beta_1 X_{ij'}+U_i)\Big|\boldsymbol{X}_i\Big\}\end{aligned}$$

Simplifying the previous expression, we have,

$$
\begin{aligned}
&-2\sum_{j\neq j'}^{n} E\big[b(\beta_0+\beta_1 X_{ij}+U_i)b(\beta_0+\beta_1 X_{ij'}+U_i)\Big|\boldsymbol{X}_i\big] \\
&\quad +2\sum_{j\neq j'}^{n} E\big[U_i b(\beta_0+\beta_1 X_{ij}+U_i)b'(\beta_0+\beta_1 X_{ij'}+U_i)\Big|\boldsymbol{X}_i\big] \\
&\quad -2\sum_{j=}^{n} E\Big[\{b(\beta_0+\beta_1 X_{ij}+U_i)\}^2\Big|\boldsymbol{X}_i\Big] \\
&\quad +2\sum_{j=1}^{n} E\big[U_i b(\beta_0+\beta_1 X_{ij}+U_i)b'(\beta_0+\beta_1 X_{ij}+U_i)\Big|\boldsymbol{X}_i\big].
\end{aligned}
$$

Treatment of T_F

Using steps similar to those given above for T_B we have

$$T_F = -2\,E(U_i\mathcal{A}_i\mathcal{A}_i'|\boldsymbol{X}_i).$$

Therefore,

$$T_F = -2\sum_{j=1}^{n}\sum_{j'=1}^{n} E\{U_i b(\beta_0+\beta_1 X_{ij}+U_i)b'(\beta_0+\beta_1 X_{ij'}+U_i)|\boldsymbol{X}_i\}.$$

Breaking this up into "diagonal" and "off-diagonal" components, we get

$$
\begin{aligned}
T_F = &-2\sum_{j\neq j'}^{n} E\{U_i b(\beta_0+\beta_1 X_{ij}+U_i)b'(\beta_0+\beta_1 X_{ij'}+U_i)|\boldsymbol{X}_i\} \\
&-2\sum_{j=1}^{n} E\{U_i b(\beta_0+\beta_1 X_{ij}+U_i)b'(\beta_0+\beta_1 X_{ij}+U_i)|\boldsymbol{X}_i\}.
\end{aligned}
$$

The Sum of T_E and T_F

Inspection of the fully expanded versions of T_E and T_F reveals that

$$
\begin{aligned}
T_E+T_F = &-2\sum_{j\neq j'}^{n} E\big[b(\beta_0+\beta_1 X_{ij}+U_i)b(\beta_0+\beta_1 X_{ij'}+U_i)\Big|\boldsymbol{X}_i\big] \\
&-2\sum_{j=1}^{n} E\Big[\{b(\beta_0+\beta_1 X_{ij}+U_i)\}^2\Big|\boldsymbol{X}_i\Big].
\end{aligned}
$$

Combining T_A, T_B, T_C, T_D, T_E and T_F

We now combine the T_A, T_B, T_C, T_D, T_E and T_F terms to get the full approximation of $E(S_{3i}^2|\boldsymbol{X}_i)$.

$$
\begin{aligned}
&E(S_{3i}^2|\boldsymbol{X}_i)\\
&\approx T_A + T_B + T_C + T_D + T_E + T_F\\
&= n\left\{\text{trigamma}(\psi) - \frac{1}{\psi}\right\} + \sum_{j\neq j'}^{n} E\Big\{b(\beta_0+\beta_1 X_{ij}+U_i)b(\beta_0+\beta_1 X_{ij'}+U_i)|\boldsymbol{X}_i\Big\}\\
&\quad - \sum_{j\neq j'}^{n} E\Big\{U_i b(\beta_0+\beta_1 X_{ij}+U_i)b'(\beta_0+\beta_1 X_{ij'}+U_i)|\boldsymbol{X}_i\Big\}\\
&\quad - \sum_{j\neq j'}^{n} E\Big\{U_i b'(\beta_0+\beta_1 X_{ij}+U_i)b(\beta_0+\beta_1 X_{ij'}+U_i)|\boldsymbol{X}_i\Big\}\\
&\quad + \sum_{j\neq j'}^{n} E\Big\{U_i^2 b'(\beta_0+\beta_1 X_{ij}+U_i)b'(\beta_0+\beta_1 X_{ij'}+U_i)|\boldsymbol{X}_i\Big\}\\
&\quad + \left(\frac{1}{\psi}+1\right)\sum_{j=1}^{n} E\Big[U_i^2\{b'(\beta_0+\beta_1 X_{ij}+U_i)\}^2\Big|\boldsymbol{X}_i\Big] + \sum_{j=1}^{n} E\Big(\{b(\beta_0+\beta_1 X_{ij}+U_i)\}^2\Big|\boldsymbol{X}_i\Big)\\
&\quad -2\sum_{j=1}^{n} E\left\{U_i b(\beta_0+\beta_1 X_{ij}+U_i)b'(\beta_0+\beta_1 X_{ij}+U_i)\Big|\boldsymbol{X}_i\right\}\\
&\quad + \sum_{j\neq j'}^{n} E\big\{U_i^2 b'(\beta_0+\beta_1 X_{ij}+U_i)b'(\beta_0+\beta_1 X_{ij'}+U_i)|\boldsymbol{X}_i\big\}\\
&\quad + \left(\frac{1}{\psi}+1\right)\sum_{j=1}^{n} E\big[U_i^2\{b'(\beta_0+\beta_1 X_{ij}+U_i)\}^2|\boldsymbol{X}_i\big]\\
&\quad + \sum_{j\neq j'}^{n} E\big\{b(\beta_0+\beta_1 X_{ij}+U_i)b(\beta_0+\beta_1 X_{ij'}+U_i)|\boldsymbol{X}_i\big\} + \sum_{j=1}^{n} E\big[\{b(\beta_0+\beta_1 X_{ij}+U_i)\}^2|\boldsymbol{X}_i\big]\\
&\quad +2\sum_{j\neq j'}^{n} E\left\{U_i b(\beta_0+\beta_1 X_{ij}+U_i)b'(\beta_0+\beta_1 X_{ij'}+U_i)\Big|\boldsymbol{X}_i\right\}\\
&\quad -2\sum_{j\neq j'}^{n} E\left\{U_i^2 b'(\beta_0+\beta_1 X_{ij}+U_i)b'(\beta_0+\beta_1 X_{ij'}+U_i)\Big|\boldsymbol{X}_i\right\}\\
&\quad +2\sum_{j=1}^{n} E\Big\{U_i b(\beta_0+\beta_1 X_{ij}+U_i)b'(\beta_0+\beta_1 X_{ij}+U_i)\Big|\boldsymbol{X}_i\Big\}\\
&\quad -2\left(\frac{1}{\psi}+1\right)\sum_{j=1}^{n} E\Big\{U_i^2\{b'(\beta_0+\beta_1 X_{ij}+U_i)\}^2\Big|\boldsymbol{X}_i\Big\}\\
&\quad -2\sum_{j\neq j'}^{n} E\big[b(\beta_0+\beta_1 X_{ij}+U_i)b(\beta_0+\beta_1 X_{ij'}+U_i)\Big|\boldsymbol{X}_i\big] - 2\sum_{j=1}^{n} E\Big[\{b(\beta_0+\beta_1 X_{ij}+U_i)\}^2\Big|\boldsymbol{X}_i\Big].
\end{aligned}
$$

Many of the terms cancel with each other, and we are left with

$$E(S_{3i}^2|\boldsymbol{X}_i) \approx n\left\{\text{trigamma}(\psi) - \frac{1}{\psi}\right\}.$$

4.5.3.4 Asymptotic Normality and Variance Results for the Maximum Likelihood Estimator of the Reciprocal Dispersion Parameter

We can show that (using results given in Wand (2007))) ψ is orthogonal to the other model parameters. Therefore, results in the previous sections of this document lead to

$$\sqrt{mn}(\widehat{\psi} - \psi) \xrightarrow{\mathcal{D}} N\left(0, \frac{1}{\text{trigamma}(\psi) - \frac{1}{\psi}}\right)$$

and we have

$$\text{Asy.Var}(\widehat{\psi}) = \frac{1}{mn\left\{\text{trigamma}(\psi) - \frac{1}{\psi}\right\}}.$$

4.5.3.5 Asymptotic Normality and Variance Results for the Maximum Likelihood Estimator of the Dispersion Parameter

If

$$\phi = 1/\psi = g(\psi) \quad \text{and} \quad g(x) = x^{-1}$$

then using the delta method leads to

$$\sqrt{mn}\{g(\widehat{\psi}) - g(\psi)\} \xrightarrow{\mathcal{D}} N\left(0, \frac{g'(\psi)^2}{\text{trigamma}(\psi) - \frac{1}{\psi}}\right).$$

Noting that

$$g'(x) = -x^{-2} \quad \text{and} \quad g'(x)^2 = x^{-4}$$

, we have

$$\frac{g'(\psi)^2}{\text{trigamma}(\psi) - \frac{1}{\psi}} = \frac{\psi^{-4}}{\text{trigamma}(\psi) - \frac{1}{\psi}} = \frac{\phi^4}{\text{trigamma}(1/\phi) - \phi}.$$

Hence

$$\sqrt{mn}(\widehat{\phi} - \phi) \xrightarrow{\mathcal{D}} N\left(0, \frac{\phi^4}{\text{trigamma}(1/\phi) - \phi}\right).$$

Lastly, we have that,

$$\text{Asy.Var}(\widehat{\phi}) = \frac{\phi^4}{mn\left\{\text{trigamma}(1/\phi) - \phi\right\}}.$$

Chapter 5

Consequences and Applications of Asymptotic Normality Results

In this chapter, we discuss the consequences and applications of the novel asymptotic normality results presented under Theorem 12.

Firstly, we present how Theorem 12 can be used to carry out asymptotically valid statistical inference in generalized linear mixed model analysis. This is done using confidence intervals constructed via the studentization process. Following that, to assess the efficacy of the Theorem 12-based confidence intervals, we ran two simulation studies and investigated the performance of our confidence intervals when the samples are finite.

Next, we move onto the implications of Theorem 12 on optimal design theory. Optimal designs contain values of the covariates in the design matrix such that these designs give the smallest standard errors of the estimators of the model parameters. These in turn lead to narrower confidence intervals and result in higher powers for hypothesis tests as compared to non-optimal designs. However, when dealing with generalized linear mixed models, choosing optimal designs can be complicated with most optimality criteria being based on the Fisher information matrix, which is computationally expensive to compute. Therefore, we then demonstrate how the derivations leading to Theorem 12 that involve large sample approximations of the Fisher information can be used in approximate optimal design settings.

This chapter is broken up into two main parts. Section 5.1 details how Theorem 12-based confidence intervals can be constructed using the studentization process. Next, Section 5.2 applies the derivations leading up to Theorem 12 in the design setting and briefly demonstrates how approximate locally D-optimal designs can be constructed

Some of the content of this chapter is published in: Jiang, J., Wand, M.P. and Bhaskaran, A. (2022), Usable and precise asymptotics for generalized linear mixed model analysis and design. Journal of the Royal Statistical Society, Series B, 84: 55-82. DOI: 10.1111/rssb.12473.

using the D-optimality criterion as presented in Theorem 13.

5.1 Asymptotically Valid Inference

In the first part of this section, we present details regarding the construction of asymptotically valid confidence intervals using Theorem 12. It is followed up by two simulation studies that are used to assess the efficacy of these asymptotically valid confidence intervals.

5.1.1 Construction of Asymptotically Valid Confidence Intervals

The asymptotic normality results for maximum quasi-likelihood estimators presented in Theorem 12 can be used to construct asymptotically valid $100(1-\alpha)\%$ confidence intervals. The confidence intervals for $\boldsymbol{\beta}_{\mathrm{A}}^0, \boldsymbol{\beta}_{\mathrm{B}}^0$ and $\boldsymbol{\Sigma}^0$ are developed using the studentization process, which involve replacing the true quantities in the asymptotic variances in Theorem 12 by their consistent estimators.

$100(1-\alpha)\%$ Confidence Interval for the Entries of $\boldsymbol{\beta}_A^0$

For $\widehat{\boldsymbol{\beta}}_{\mathrm{A}}$, its asymptotic covariance matrix only involves $\boldsymbol{\Sigma}^0$. Hence, studentization simply involves replacing $\boldsymbol{\Sigma}^0$ by $\widehat{\boldsymbol{\Sigma}}$ which leads to the following asymptotic normality result

$$\sqrt{m}\widehat{\boldsymbol{\Sigma}}^{-1/2}\left(\widehat{\boldsymbol{\beta}}_{\mathrm{A}} - \boldsymbol{\beta}_{\mathrm{A}}^0\right) \xrightarrow{\mathcal{D}} N(\mathbf{0}, \boldsymbol{I}).$$

Let $\widehat{\sigma}_k^2$ denote the kth diagonal entry of $\widehat{\boldsymbol{\Sigma}}$, $1 \leq k \leq d_{\mathrm{A}}$. It follows that the asymptotically valid $100(1-\alpha)\%$ confidence interval for the kth entry of $\boldsymbol{\beta}_{\mathrm{A}}^0$ is

$$(\widehat{\boldsymbol{\beta}}_{\mathrm{A}})_k \pm \Phi^{-1}\left(1 - \frac{\alpha}{2}\right)\sqrt{\frac{\widehat{\sigma}_k^2}{m}}, \tag{5.1}$$

where Φ represents the standard normal cumulative distribution function.

$100(1-\alpha)\%$ Confidence Intervals for the Entries of $\boldsymbol{\beta}_B^0$

Constructing asymptotically valid confidence intervals for the entries of $\boldsymbol{\beta}_{\mathrm{B}}^0$ is less straightforward compared to constructing confidence intervals for $\boldsymbol{\beta}_{\mathrm{A}}^0$ or $\boldsymbol{\Sigma}^0$. Studentiza-

tion involves replacing $\boldsymbol{\Lambda}_{\boldsymbol{\beta}_{\mathrm{B}}}$ by $\widehat{\boldsymbol{\Lambda}}_{\boldsymbol{\beta}_{\mathrm{B}}}$ which leads to the following asymptotic normality result

$$\sqrt{mn}\left(\phi\widehat{\boldsymbol{\Lambda}}_{\boldsymbol{\beta}_{\mathrm{B}}}\right)^{-1/2}\left(\widehat{\boldsymbol{\beta}}_{\mathrm{B}}-\boldsymbol{\beta}_{\mathrm{B}}^{0}\right)\xrightarrow{\mathcal{D}}N(\mathbf{0},\boldsymbol{I}),$$

where $\widehat{\boldsymbol{\Lambda}}_{\boldsymbol{\beta}_{\mathrm{B}}}$ is defined as follows

$$\widehat{\boldsymbol{\Lambda}}_{\boldsymbol{\beta}_{\mathrm{B}}}\equiv\left[\left|2\pi\widehat{\boldsymbol{\Sigma}}\right|^{-1/2}\int_{\mathbb{R}^{d_{\mathrm{A}}}}\left\{\text{lower right } d_{\mathrm{B}}\times d_{\mathrm{B}} \text{ block of } \widehat{\boldsymbol{\Omega}}_{\boldsymbol{\beta}_{\mathrm{B}}}(\boldsymbol{u})^{-1}\right\}^{-1}\exp\left(-\frac{1}{2}\boldsymbol{u}^{T}\widehat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{u}\right)d\boldsymbol{u}\right]^{-1}$$

and

$$\widehat{\boldsymbol{\Omega}}_{\boldsymbol{\beta}_{\mathrm{B}}}(\boldsymbol{u})\equiv\frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n}\left\{b''\left((\widehat{\boldsymbol{\beta}}_{\mathrm{A}}+\boldsymbol{u})^{T}\boldsymbol{X}_{\mathrm{A}ij}+(\widehat{\boldsymbol{\beta}}_{\mathrm{B}})^{T}\boldsymbol{X}_{\mathrm{B}ij}\right)\begin{bmatrix}\boldsymbol{X}_{\mathrm{A}ij}\boldsymbol{X}_{\mathrm{A}ij}^{T} & \boldsymbol{X}_{\mathrm{A}ij}\boldsymbol{X}_{\mathrm{B}ij}^{T}\\ \boldsymbol{X}_{\mathrm{B}ij}\boldsymbol{X}_{\mathrm{A}ij}^{T} & \boldsymbol{X}_{\mathrm{B}ij}\boldsymbol{X}_{\mathrm{B}ij}^{T}\end{bmatrix}\right\}.$$

Then, the asymptotically valid $100(1-\alpha)\%$ confidence interval for the kth entry of $\boldsymbol{\beta}_{\mathrm{B}}^{0}$ is

$$(\widehat{\boldsymbol{\beta}}_{\mathrm{B}})_{k}\pm\Phi^{-1}\left(1-\frac{\alpha}{2}\right)\sqrt{\frac{\left(\phi\widehat{\boldsymbol{\Lambda}}_{\boldsymbol{\beta}_{\mathrm{B}}}\right)_{kk}}{m}}. \tag{5.2}$$

$100(1-\alpha)\%$ Confidence Intervals for the Entries of $\boldsymbol{\Sigma}^{0}$

For $\widehat{\boldsymbol{\Sigma}}$, in a similar manner to the case involving $\widehat{\boldsymbol{\beta}}_{\mathrm{A}}$, its asymptotic covariance matrix only consists of $\boldsymbol{\Sigma}^{0}$. We apply the studentization process and replace $\boldsymbol{\Sigma}^{0}$ by $\widehat{\boldsymbol{\Sigma}}$ which leads to the following asymptotic normality result

$$\sqrt{m}\left\{2\boldsymbol{D}_{d_{\mathrm{A}}}^{+}\left(\widehat{\boldsymbol{\Sigma}}\otimes\widehat{\boldsymbol{\Sigma}}\right)\boldsymbol{D}_{d_{\mathrm{A}}}^{+\,T}\right\}^{-1/2}\operatorname{vech}\left(\widehat{\boldsymbol{\Sigma}}-\boldsymbol{\Sigma}^{0}\right)\xrightarrow{\mathcal{D}}N(\mathbf{0},\boldsymbol{I}).$$

It then follows that the asymptotically valid $100(1-\alpha)\%$ confidence interval for the (k,k)th entry of $\boldsymbol{\Sigma}^{0}$ is

$$\widehat{\sigma}_{k}^{2}\pm\Phi^{-1}\left(1-\frac{\alpha}{2}\right)\sqrt{\frac{2\left(\widehat{\sigma}_{k}^{2}\right)^{2}}{m}}. \tag{5.3}$$

5.1.2 Simulation Study

Two simulation studies were run to assess the efficacy of the confidence intervals constructed using the asymptotic normality results presented in Theorem 12. In this study, confidence intervals for the following $d_{\mathrm{A}}=d_{\mathrm{B}}=1$ Poisson mixed model were

generated

$$Y_{ij}|X_{ij}, U_i \overset{\text{ind.}}{\sim} \text{Poisson}\left(\exp\left(\beta_0^0 + \beta_B^0 X_{ij} + U_i\right)\right),$$
$$U_i \overset{\text{ind.}}{\sim} N\left(0, (\sigma^2)^0\right),\ 1 \le i \le m,\ 1 \le j \le n,$$

with $\phi = 1$. To simplify the notation involved, $\boldsymbol{\beta}_{\text{A}}, \boldsymbol{\beta}_{\text{B}}$ and $\boldsymbol{\Sigma}$ have been replaced by the scalar parameter symbols β_0, β_B and σ^2. The values for the true parameter vector $(\beta_0^0, \beta_B^0, (\sigma^2)^0)$ were chosen from the following possible set of pre-determined values

$$\{(-0.3, 0.2, 0.5), (2.2, -0.1, 0.16), (1.2, 0.4, 0.1), (0.02, 1.3, 1), (-0.3, 0.2, 0.1)\},$$

such that the data was well-behaved and led to fewer singularity issues. The distribution for X_{ij} was also chosen to be either $N(0, 1)$ or Uniform$(-1, 1)$. The number of groups in the simulated data, m, varied over the set $\{100, 200, \ldots, 1000\}$ and the number of observations present within each group, n, was fixed at $m/10$. A total of 1000 replications were simulated for every possible combination of the true parameter vector, chosen X_{ij} distribution and value of (m, n) pair.

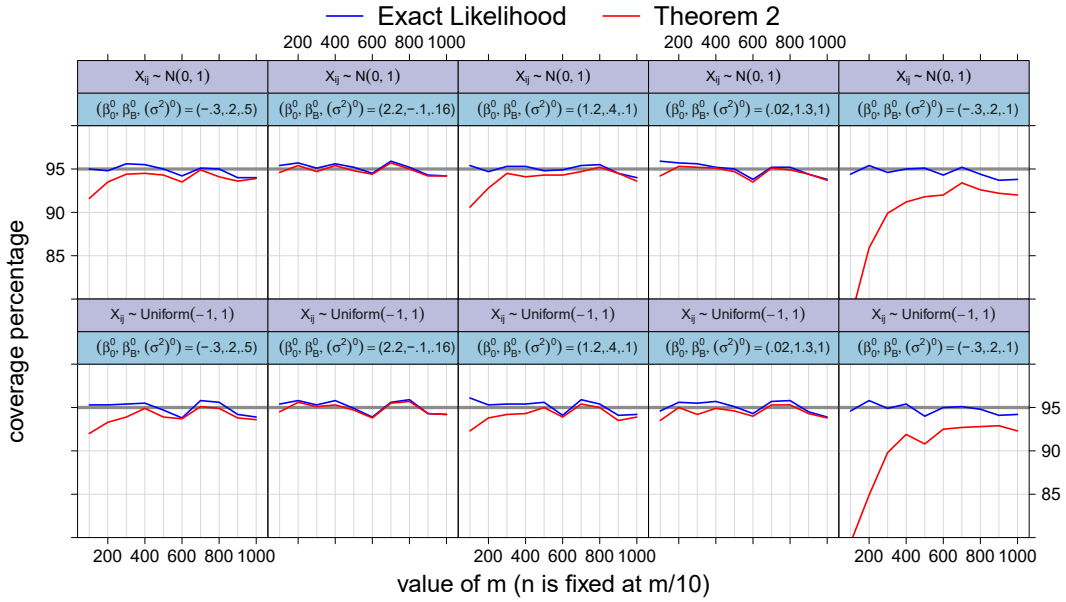

Figure 5.1: Actual coverage percentage of nominally 95% confidence intervals for β_0^0 in a $d_A = d_B = 1$ Poisson mixed model. The confidence intervals are obtained using the exact observed Fisher information computations provided by the function `glmer()` in the `R` package `lme4` (blue lines) and Theorem 12 with studentization according to (5.1) (red lines). The nominal percentage is shown as a thick grey horizontal line. The percentages are based on 1000 replications. The values of m are $100, 200, \ldots, 1000$. The value of n is fixed at $m/10$.

Using the `glmer()` function in the R package lme4 (Bates et al., 2015), maximum likelihood estimates of β_0^0, β_B^0 and $(\sigma^2)^0$ were obtained for every sample of simulated data generated. 95% confidence intervals based on (5.1) and (5.2) were computed using the maximum likelihood estimates obtained. 95% confidence intervals based on exact observed Fisher information were also obtained using `glmer()`. The coverage percentages corresponding to the percentage of times the true value landed in the confidence intervals were also calculated for both approaches.

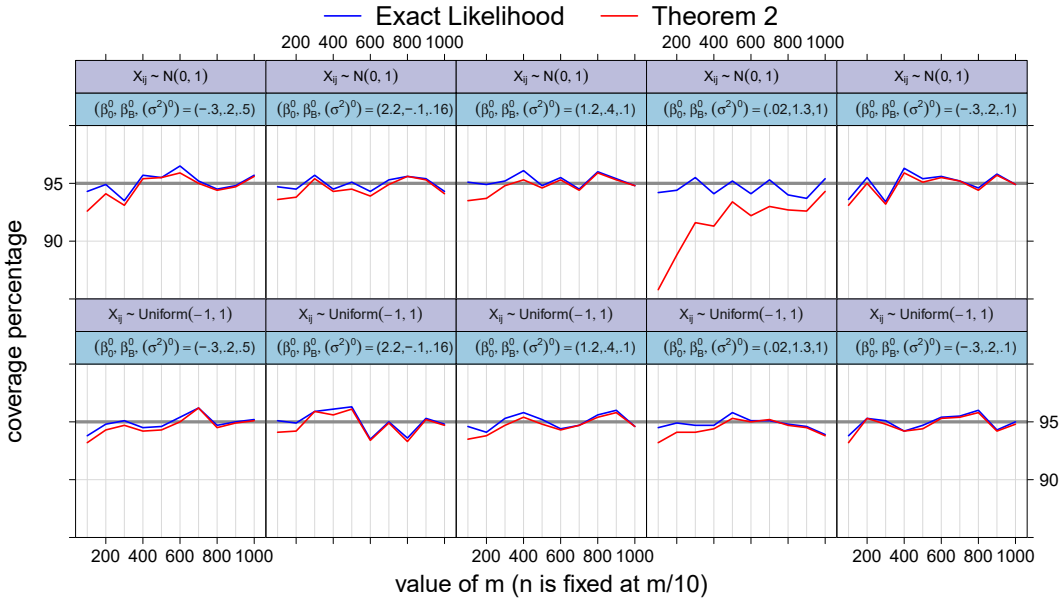

Figure 5.2: Actual coverage percentage of nominally 95% confidence intervals for β_B^0 in a $d_A = d_B = 1$ Poisson mixed model. The confidence intervals are obtained using the exact observed Fisher information computations provided by the function `glmer()` in the `R` package `lme4` (blue lines) and Theorem 12 with studentization according to (5.2) (red lines). The nominal percentage is shown as a thick grey horizontal line. The percentages are based on 1000 replications. The values of m are $100, 200, \ldots, 1000$. The value of n is fixed at $m/10$.

Figure 5.1 displays the coverage percentages for the 95% confidence intervals computed using the two approaches for various simulation settings. Both the Theorem 12-based approach and the approach based on exact observed Fisher information give almost identical coverage percentages for all response distribution and sample size combinations across all values of m, for the first four true parameter vector settings. For the last setting, while the Theorem 12-based asymptotic approach does not perform well for smaller values of m, the asymptotic properties it is based on gives similar coverage

percentages to that of the exact observed Fisher information approach once the value of m exceeds 500 and continues to get larger. This suggests that the asymptotic variance of $\widehat{\beta}_0$ being σ^2/m is a very good approximation to the variance of $\widehat{\beta}_0$ that arises from using exact observed Fisher information, especially with larger values of m.

In Figure 5.2, once again, both the Theorem 12-based approach and the approach based on exact observed Fisher information, for most of the true parameter vector settings, give almost identical coverage percentages for all response distribution and sample size combinations across all values of m. For the fourth setting, when the X_{ij}s are generated from a standard normal distribution, the Theorem 12-based approach does not perform well for smaller values of m as compared to the exact observed Fisher information approach. However, both approaches have similar performances for larger values of m. This suggests that the asymptotic variance of $\widehat{\beta}_B$ being Λ_{β_B}/mn is a very good approximation to the variance of $\widehat{\beta}_B$ that arises from using exact observed Fisher information.

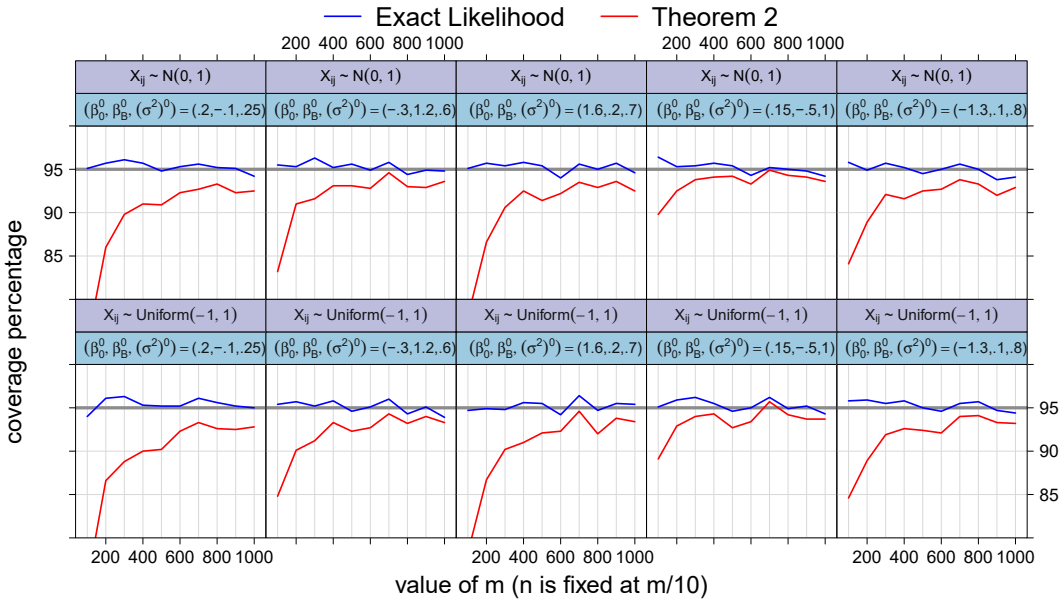

Figure 5.3: Actual coverage percentage of nominally 95% confidence intervals for β_0^0 in a $d_A = d_B = 1$ logistic mixed model. The confidence intervals are obtained using the exact observed Fisher information computations provided by the function `glmer()` in the `R` package `lme4` (blue lines) and Theorem 12 with studentization according to (5.1) (red lines). The nominal percentage is shown as a thick grey horizontal line. The percentages are based on 1000 replications. The values of m are $100, 200, \ldots, 1000$. The value of n is fixed at $m/10$.

In the next study, confidence intervals for the following $d_{\mathrm{A}} = d_{\mathrm{B}} = 1$ logistic mixed model were generated

$$Y_{ij}|X_{ij}, U_i \overset{\text{ind.}}{\sim} \text{Bernoulli}\left(\text{expit}\left(\beta_0^0 + \beta_B^0 X_{ij} + U_i\right)\right),$$
$$U_i \overset{\text{ind.}}{\sim} N\left(0, (\sigma^2)^0\right), \ 1 \le i \le m, \ 1 \le j \le n,$$

with $\phi = 1$.

The values for the true parameter vector $\left(\beta_0^0, \beta_B^0, \left(\sigma^2\right)^0\right)$ were chosen from the following possible set of pre-determined values

$$\{(0.2, -0.1, 0.25), (-0.3, 1.2, 0.6), (1.6, 0.2, 0.7), (0.15, -0.5, 1), (-1.3, 0.1, 0.8)\}.$$

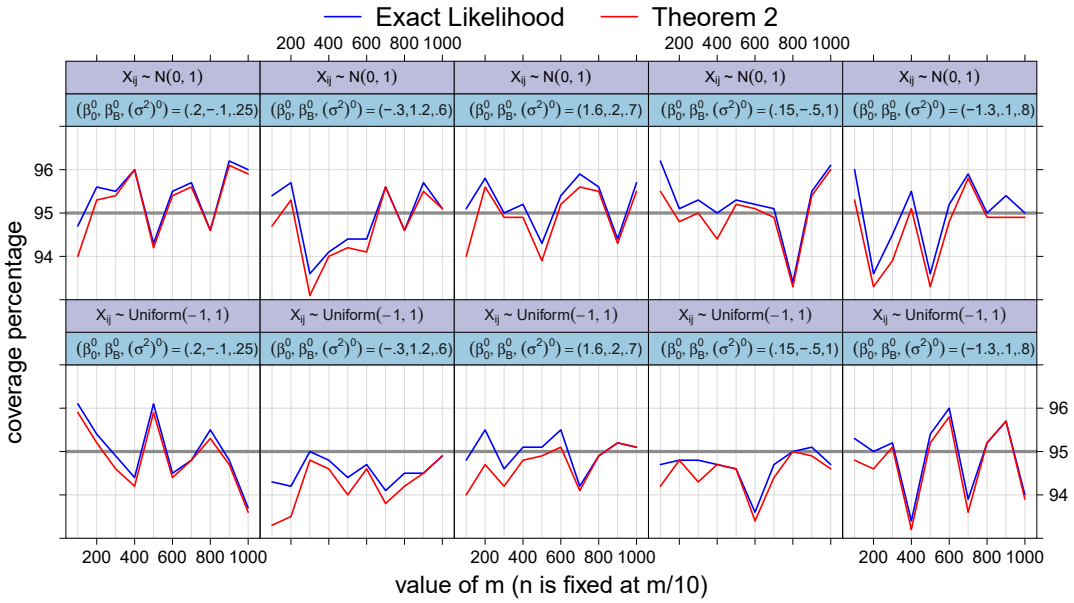

Figure 5.4: Actual coverage percentage of nominally 95% confidence intervals for $\boldsymbol{\beta}_B^0$ in a $d_A = d_B = 1$ logistic mixed model. The confidence intervals are obtained using the exact observed Fisher information computations provided by the function `glmer()` in the `R` package `lme4` (blue lines) and Theorem 12 with studentization according to (5.2) (red lines). The nominal percentage is shown as a thick grey horizontal line. The percentages are based on 1000 replications. The values of m are $100, 200, \ldots, 1000$. The value of n is fixed at $m/10$.

The remaining simulation settings match that of the Poisson simulation study conducted. Once again, the maximum likelihood estimates of β_0^0, β_B^0 and $(\sigma^2)^0$ were

obtained and 95% confidence intervals based on both Theorem 12 and exact observed Fisher information were computed.

In Figure 5.3, using exact observed Fisher information leads to better coverage percentages across all values of m. While the confidence intervals constructed using Theorem 12 do not perform as well for lower values of m, their performances match those of exact observed Fisher information when the values of m grows beyond $m = 500$.

In Figure 5.4, both approaches give almost identical coverage percentages for all response distribution and sample size combinations across all values of m, for all five true parameter vector settings. Note that the asymptotic variance of $\widehat{\beta}_B$ has a convergence rate of $(mn)^{-1}$ while the asymptotic variance of $\widehat{\beta}_0$ has a slower convergence rate of m^{-1}. This attributes to why the coverage percentages computed from Theorem 12-based confidence intervals achieve percentages closer to 95% for smaller values of m and n in Figure 5.4 as compared to the results shown in Figure 5.3 in the simple logistic mixed model simulation study.

Across all four figures, the simulation results indicate that sometimes, certain true values of the model parameters and the chosen distribution of $\boldsymbol{X}$ may result in Theorem 12 performing worse in comparison to the other simulation cases presented. This is particularly evident in Figures 5.3 and 5.4. To explain this, we need to look beyond the first-order asymptotic covariances presented in this thesis and analyse the second-order asymptotic covariances instead (Maestrini et al., 2023). Let $\mathcal{X} \equiv \{\boldsymbol{X}_{ij} : 1 \leq i \leq m,\ 1 \leq j \leq n_i\}$. As an example, for the $d_{\mathrm{A}} = 1$, $d_{\mathrm{B}} = 1$ Poisson quasi-likelihood special case of (4.3), a solution to the two-term asymptotic covariance problem can be expressed relatively simply with parameters

$$\boldsymbol{\beta}_A = \beta_0, \quad \boldsymbol{\beta}_B = \beta_1 \quad \text{and} \quad \boldsymbol{\Sigma} = \sigma^2 \quad \text{and predictor variable} \quad \boldsymbol{X} = \begin{bmatrix} 1 \\ X \end{bmatrix}$$

for a scalar random variable X. Define

$$a_1\big(\beta_0, \beta_1, \sigma^2\big) \equiv e^{\beta_0 + \sigma^2/2}\big[E(X^2 e^{\beta_1 X})E(e^{\beta_1 X}) - \{E(X e^{\beta_1 X})\}^2\big]$$

and

$$a_2(\beta_1, \sigma^2) \equiv \frac{e^{\sigma^2} E\big(X^2 e^{\beta_1 X}\big) E\big(e^{\beta_1 X}\big) + \big(1 - e^{\sigma^2}\big) E\big\{\big(X e^{\beta_1 X}\big)\big\}^2}{E\big(e^{\beta_1 X}\big)}.$$

Then the two-term asymptotic covariance matrix of $(\widehat{\beta}_0, \widehat{\beta}_1)$ is

$$\text{Cov}\left(\left[\begin{array}{c}\widehat{\beta}_0\\ \widehat{\beta}_1\end{array}\right]\middle|\mathcal{X}\right) = \frac{1}{m}\left[\begin{array}{cc}(\sigma^2)^0 & 0\\ 0 & 0\end{array}\right] + \frac{\phi\{1+o_p(1)\}}{a_1\big(\beta_0^0, \beta_1^0, (\sigma^2)^0\big)\, mn}\left[\begin{array}{cc}a_2\big(\beta_1^0, (\sigma^2)^0\big) & -E\big(Xe^{\beta_1^0 X}\big)\\ -E\big(Xe^{\beta_1^0 X}\big) & E\big(e^{\beta_1^0 X}\big)\end{array}\right]. \tag{5.4}$$

As is apparent from (5.4), the differences between one-term and two-term asymptotic variances depend on m, n, σ_0^2 and particular moments of the X distribution in a complicated way. Theoretically speaking, it is possible to make these differences arbitrarily small or large by appropriate choices of m, n, σ_0^2 and the distribution of X. The 4th column of Figure 5.2 contrasts with the other columns in terms of the empirical coverage because the one-term and two-term approximations differ more, with exact likelihood being closer to the two-term approximation.

Therefore, for certain chosen values of the true model parameters and distribution of $\boldsymbol{X}$, the first order asymptotic variance terms are insufficient to obtain good coverage percentages as the second term in the asymptotic variance is also significant, especially for small values of m and n. In these cases, the confidence intervals should be constructed using two-term asymptotic variances instead rather than the asymptotic variances presented in Theorem 12, which will lead to better coverage percentages.

With regards to computing these confidence intervals, the Theorem 12 and studentization based approach provides the analyst with a quicker and simpler option, especially for large m. When computing asymptotically valid confidence intervals using the exact approach, numerical integration is required to compute the ratios of integrals involved when computing the exact observed Fisher information matrix. Note that for $d_A > 1$, multivariate numerical integration is needed for the exact approach. On the other hand, for constructing asymptotically valid confidence intervals for β_0^0 using Theorem 12, when m is in the several hundreds or thousands, the closed form confidence interval arising from Theorem 12 and studentization is an attractive alternative to the numerical integration-based exact approach. For constructing asymptotically valid confidence intervals for β_B^0 using Theorem 12, the logistic case requires simple numerical integration as compared to the exact approach while the Poisson case does not require any numerical integration.

5.2 Approximate Optimal Design

In this section, we demonstrate how the derivations leading to Theorem 12 involving large sample approximations of the Fisher information can be used in approximate optimal design settings.

5.2.1 Background and Model Description

In previous sections, we assumed that the data was observed in accordance to the model described in (4.3). Now, let us consider the use of a generalized linear mixed model as in (4.3) with $d_A = 1$, $\boldsymbol{X}_A = 1$, $\boldsymbol{\beta}_{\mathrm{A}} = \beta_0$, $\boldsymbol{\Sigma} = \sigma^2$ and with the same number of observations in each group. Also consider the case where the data is yet to be observed. This simplifies to a random intercept generalized linear mixed model as follows

$$\begin{aligned} &Y_{ij}|U_i \text{ are independent having quasi-likelihood function (4.2) with} \\ &\text{natural parameter } \beta_0^0 + (\boldsymbol{\beta}_{\mathrm{B}}^0)^T \boldsymbol{x}_{\mathrm{B}ij} + U_i \text{ such that the } U_i \text{ are independent} \\ &N(0, (\sigma^2)^0) \text{ random variables,} \end{aligned} \tag{5.5}$$

where $1 \leq i \leq m$ and $1 \leq j \leq N$. The unique values of the non-random $\boldsymbol{x}_{\mathrm{B}ij}$ predictor vectors in the ith group can be viewed as a finite set of points in $\mathbb{R}^{d_{\mathrm{B}}}$ and can be denoted as $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_s$. Here, s denotes the number of unique predictor vectors.

In optimal design, with the help of an optimality criterion, one is firstly required to select the possible values of $\boldsymbol{x}$ at which the observations of Y_{ij} will be made. One also has to determine the fraction of occurrences of independent observations made at each value of $\boldsymbol{x}$ (Russell, 2018). Each $\boldsymbol{x}$ used in the design is a support point. Hence, there are s unique support points included in the design.

Now let $\mathcal{X} \subseteq \mathbb{R}^{d_B}$ denote the set to which the support points are restricted to. For example, if $d_B = 2$ with the first predictor being binary and the second predictor being a proportion then $\mathcal{X} = \{(x_1, x_2) : x_1 \in \{0, 1\}, 0 \leq x_2 \leq 1\}$. Also, denote the number of independent observations made at $\boldsymbol{x}_k$, $1 \leq k \leq s$, as n_k and define

$$\delta_k \equiv \frac{n_k}{N}, \quad 1 \leq k \leq s,$$

where $N \equiv n_1 + \cdots + n_s$. Hence, the δ_k are known as design weights and represent the fraction of data in the ith group associated to each support point. Note that $\delta_1 + \ldots \delta_s = 1$. Our working assumption throughout the rest of this section is that the

asymptotically valid D-optimal designs are such that each of the m groups have exactly the same support points and design weights.

In this thesis, we restrict our attention to approximate optimal designs, which are common when the design weights are represented as decimals or fractions if there are recurring decimals. For example, a design could contain 4 support points with the design weights for $(\delta_1, \delta_2, \delta_3, \delta_4)$ being $\{0.127, 0.2378, 0.452, 0.1832\}$. This is an example of an ideal design that is rarely exactly attainable. In such cases, exact designs with all $n_k \in \mathbb{Z}^+$ are not always possible, especially for small values of N, and approximate designs are achievable instead.

Out of the optimality-criteria available, we also restrict our attention to D-optimality. This involves choosing the design that maximises the determinant of the Fisher information matrix. The derivations leading to Theorem 12 involve large sample expressions for the Fisher information for the class of generalized linear mixed models. By using the D-optimality criterion, these analogous large sample approximations of the Fisher information allow for approximate locally optimal design determination. Since the approximations of the Fisher information matrix are asymptotic approximations, we only considered designed experiments for which large sample sizes are feasible.

In addition, note that in non-Gaussian generalized linear mixed models, the Fisher information matrix contains entries dependent on the model parameters. Hence, we work with designs that maximise the determinant of the Fisher information matrix with fixed values for the model parameters, known as locally D-optimal designs.

5.2.2 Approximate Locally D-Optimal Design Determination

Define

$$n \equiv \frac{1}{s}\sum_{k=1}^{s} n_k = \text{ average of the support point replication sizes within each group.}$$

The theorem relies on the following assumption:

(A7) The design sample sizes n_k diverge to ∞ in such a way that $n_k/(sn) \to \delta_k$ for constants $0 \leq \delta_1 \leq 1, 1 \leq k \leq s$.

Theorem 13. *Consider the random intercept generalized linear mixed model described in* (5.5) *with design weights* δ_k *and corresponding support points* $x_k \in \mathcal{X} \subseteq \mathbb{R}^{d_B}, 1 \leq k \leq s$. *Assume that condition (A7) holds. Then, based on the exact leading term behaviour of*

the determinant of the Fisher information matrix, approximate locally D-optimal designs at the parameter vector $(\beta_0, \boldsymbol{\beta}_B, \sigma^2)$ *are those for which*

$$\left| \int_{-\infty}^{\infty} \left\{ \textit{lower right } d_B \times d_B \textit{ block of } \left(\sum_{k=1}^{s} \delta_k b''(\beta_0 + \boldsymbol{\beta}_B^T \boldsymbol{x}_k + u) \begin{bmatrix} 1 & \boldsymbol{x}_k^T \\ \boldsymbol{x}_k & \boldsymbol{x}_k \boldsymbol{x}_k^T \end{bmatrix} \right)^{-1} \right\}^{-1} \times \exp\{-u^2/(2\sigma^2)\} du \right|$$

is maximal over $\{\delta_k : \delta_k \geq 0, \sum_{k=1}^{s} \delta_k = 1, 1 \leq k \leq s\}$ *and* $\{\boldsymbol{x}_k \in \mathcal{X} : 1 \leq k \leq s\}$.

The proof of Theorem 13 is in the appendix. Some remarks regarding Theorem 13 are as follows:

1. Existing literature that consider the selection of optimal designs using the D-optimality criterion for classes of generalized linear mixed models similar to those considered in this thesis include Waite and Woods (2015) and Zhang et al. (2017). Waite and Woods (2015) used generalized estimating equations to obtain approximations of the mixed model Fisher Information Matrix. On the other hand, Zhang et al. (2017) explored three methods to approximate the Fisher information matrix, namely importance sampling, Laplace approximation and joint sampling. In contrast, Theorem 13 facilitates approximate locally D-optimal design determination in a more direct manner and is based on the precise leading term behaviour of the Fisher information matrix.

2. In this chapter, the theory and discussion has been restricted to D-optimality. Other optimality criteria such as A-optimality, which requires minimization of the trace of the inverse of the information matrix, also benefit from our precise asymptotic approximations of the Fisher information matrix for generalized linear mixed models.

3. Note that in the Gaussian case, $b''(x) = 1$, and the determinant in Theorem 13 is proportional to

$$\left| \sum_{k=1}^{s} \delta_k \begin{bmatrix} 1 & \boldsymbol{x}_k^T \\ \boldsymbol{x}_k & \boldsymbol{x}_k \boldsymbol{x}_k^T \end{bmatrix} \right|. \tag{5.6}$$

   Since the expression above does not depend on any of the model parameters, designs that maximize (5.6) are globally D-optimal.

4. In the Poisson linear mixed model case, $b''(x) = \exp(x)$ leads to the simplification

of Theorem 13. The D-optimality criterion then reduces to

$$\left|\sum_{k=1}^{s}\delta_k \exp(\boldsymbol{\beta}_B^T\boldsymbol{x}_k)\begin{bmatrix}1 & \boldsymbol{x}_k^T\\ \boldsymbol{x}_k & \boldsymbol{x}_k\boldsymbol{x}_k^T\end{bmatrix}\right| \Big/ \sum_{k=1}^{s}\delta_k \exp(\boldsymbol{\beta}_B^T\boldsymbol{x}_k).$$

Note that generating approximate locally D-optimal designs for the Poisson linear mixed model case is only dependent on $\boldsymbol{\beta}_B$.

5. When considering logistic mixed models, $b''(x) = 1/[2\{1+\cosh(x)\}]$ and Theorem 13 does not simplify further. Hence, in the logistic mixed model case, approximate locally D-optimal designs depend on $\beta_0, \boldsymbol{\beta}_B$ and σ^2. Although Theorem 13 does not admit an explicit form in this case, each of the entries of the approximate Fisher information matrix can be computed using univariate numerical integration.

5.2.3 Illustration of Theorem 13

In this section, we illustrate the use of Theorem 13 in determining the approximate optimal design when $d_B = 2$ and both predictors are binary, taking on values of either 0 or 1. In this scenario, there are at most $s = 4$ support points with the only possible support points being $\boldsymbol{x}_k \in \{(0,0),(0,1),(1,0),(1,1)\}$. Since all the possible support points are known, one would only need to maximise the expression in Theorem 13 over the design weights. Figure 5.5 shows the approximate locally D-optimal designs, for the situation where $\beta_0 = -0.3$, $\boldsymbol{\beta}_B = (1.7, 2.1)$ and the values of σ take on a value from $\{0.6, 0.76, 0.97, 1.24, 1.57, 2.00\}$, by displaying the optimal design weights for the possible corresponding support points. To obtain Figure 5.5, we used code similar to that provided in Section 4.5 of Russell (2018), which is based on the `optim()` function in R (R Core Team (2022)) and Nelder-Mead searches with 100 random initial values.

We noted that the choices for the initial values of the design weights did not impact the results. From Figure 5.5, we see that for the two lowest values of σ, the optimal designs have only three support points, with the point $(1,1)$ being excluded from the design. However, as the value of σ increases, the corresponding design weight for the support point $(1,1)$ becomes positive and larger.

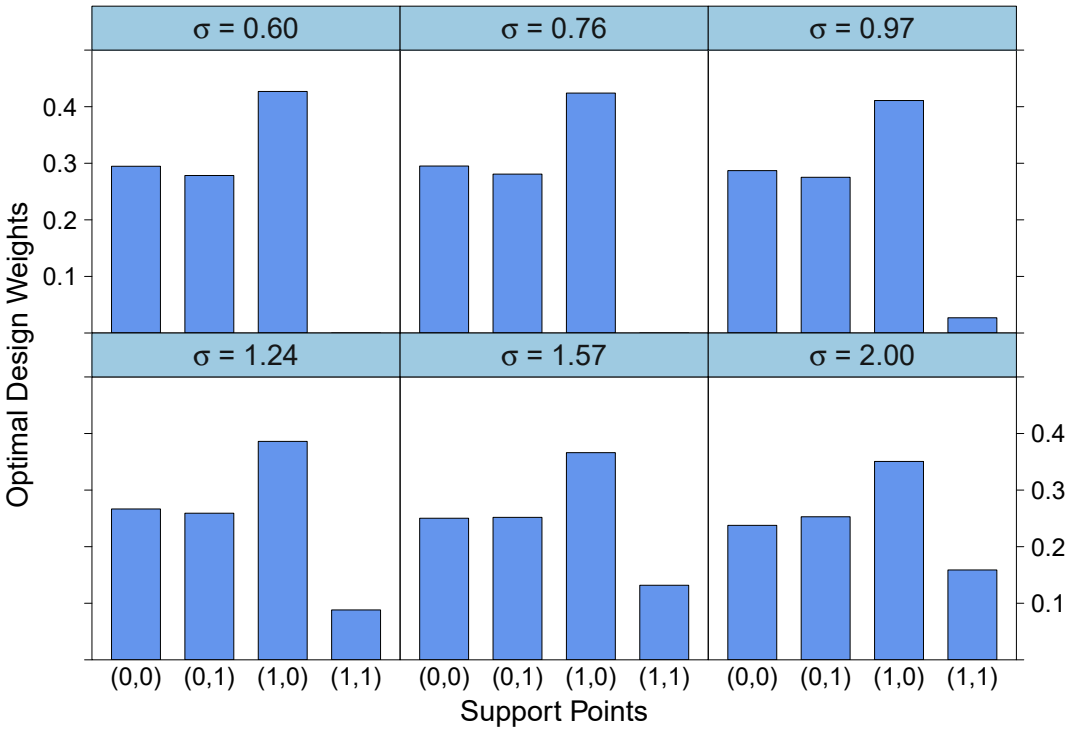

Figure 5.5: Approximate locally D-optimal designs for logistic mixed models with two binary predictors when $\beta_0 = -0.3$, $\boldsymbol{\beta}_B = (1.7, 2.1)$ and the values of σ take on a value from $\{0.6, 0.76, 0.97, 1.24, 1.57, 2.00\}$.

5.3 Appendix

This appendix contains the details for the derivations leading up to Theorem 13.

5.3.1 Model Description

Our working assumption throughout this section is that the asymptotically D-optimal designs have exactly the same support points and design weights for each of the m groups. Let $\boldsymbol{x}_k$, $1 \leq k \leq s$, be the support points that are common to each group. For each $1 \leq i \leq m$, let

$$n_k = \text{number of } \boldsymbol{x}_k \text{ values in the design, } \; 1 \leq k \leq s.$$

Following that, the full data for the ith group can be expressed as:

$$\begin{gathered}\left(\boldsymbol{x}_1, Y_{i1}^{[1]}\right), \left(\boldsymbol{x}_1, Y_{i2}^{[1]}\right), \ldots, \left(\boldsymbol{x}_1, Y_{in_1}^{[1]}\right), \\ \left(\boldsymbol{x}_2, Y_{i1}^{[2]}\right), \left(\boldsymbol{x}_2, Y_{i2}^{[2]}\right), \ldots, \left(\boldsymbol{x}_2, Y_{in_2}^{[2]}\right), \\ \vdots \\ \left(\boldsymbol{x}_s, Y_{i1}^{[s]}\right), \left(\boldsymbol{x}_s, Y_{is}^{[s]}\right), \ldots, \left(\boldsymbol{x}_s, Y_{in_s}^{[s]}\right).\end{gathered}$$

Then, the conditional quasi-probability mass function or quasi-density function of $Y_{ij}^{[k]}$ given U_i is

$$p_{Y_{ij}^{[k]}|U_i}(y|U_i = u_i) = \exp\left[\frac{1}{\phi}\left\{y(\beta_0 + \boldsymbol{\beta}_B^T \boldsymbol{x}_k + u_i) - b(\beta_0 + \boldsymbol{\beta}_B^T \boldsymbol{x}_k + u_i)\right\} + c(y)\right],$$

where $U_i \sim N(0, \sigma^2)$, $1 \le i \le m$, $1 \le j \le n_k$ and $1 \le k \le s$. Also, for each $1 \le i \le m$, conditional on U_i the

$$Y_{ij}^{[k]}, \quad 1 \le j \le n_k,\ 1 \le k \le s$$

are independent. Therefore, the quasi-likelihood of $(\beta_0, \boldsymbol{\beta}_B, \sigma^2)$ is

$$\mathcal{L}(\beta_0, \boldsymbol{\beta}_B, \sigma^2) = \prod_{i=1}^{m} \int_{-\infty}^{\infty} \left\{\prod_{k=1}^{s}\prod_{j=1}^{n_k} p_{Y_{ij}^{[k]}|U_i}(Y_{ij}^{[k]}|U_i = u)\right\} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\{-(u^2)/(2\sigma^2)\}\, du,$$

and the conditional quasi-log-likelihood is

$$\begin{aligned}\ell(\beta_0, \boldsymbol{\beta}_B, \sigma^2) &= \sum_{i=1}^{m} \log \int_{-\infty}^{\infty} \left\{\prod_{k=1}^{s}\prod_{j=1}^{n_k} p_{Y_{ij}^{[k]}|U_i}(Y_{ij}^{[k]}|U_i = u)\right\} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\{-(u^2)/(2\sigma^2)\}\, du \\ &= \sum_{i=1}^{m} \log\left(\frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} \exp\left[\sum_{k=1}^{s}\sum_{j=1}^{n_k}\left\{\frac{1}{\phi} Y_{ij}^{[k]}(\beta_0 + \boldsymbol{\beta}_B^T \boldsymbol{x}_k + u)\right.\right.\right. \\ &\qquad \left.\left.\left. -\frac{1}{\phi} b(\beta_0 + \boldsymbol{\beta}_B^T \boldsymbol{x}_k + u) + c(Y_{ij}^{[k]})\right\} - (u^2)/(2\sigma^2)\right] du\right).\end{aligned}$$

Note that,

$$\begin{aligned}&\int_{-\infty}^{\infty} \exp\left[\sum_{k=1}^{s}\sum_{j=1}^{n_k} \frac{1}{\phi}\left\{Y_{ij}^{[k]}(\beta_0 + \boldsymbol{\beta}_B^T \boldsymbol{x}_k + u) - b(\beta_0 + \boldsymbol{\beta}_B^T \boldsymbol{x}_k + u)\right\} - \frac{u^2}{2\sigma^2}\right] du \\ &= \exp\left[\sum_{k=1}^{s}\sum_{j=1}^{n_k} \frac{1}{\phi}\left\{Y_{ij}^{[k]}(\beta_0 + \boldsymbol{\beta}_B^T \boldsymbol{x}_k)\right\}\right] \\ &\qquad \times \int_{-\infty}^{\infty} \exp\left[\sum_{k=1}^{s}\sum_{j=1}^{n_k} \frac{1}{\phi}\left\{Y_{ij}^{[k]} u - b(\beta_0 + \boldsymbol{\beta}_B^T \boldsymbol{x}_k + u)\right\} - \frac{u^2}{2\sigma^2}\right] du.\end{aligned}$$

Hence,

$$\log \int_{-\infty}^{\infty} \exp\left[\sum_{k=1}^{s}\sum_{j=1}^{n_k} \frac{1}{\phi}\left\{Y_{ij}^{[k]}(\beta_0 + \boldsymbol{\beta}_B^T \boldsymbol{x}_k + u) - b(\beta_0 + \boldsymbol{\beta}_B^T \boldsymbol{x}_k + u)\right\} - \frac{u^2}{2\sigma^2}\right] du$$

$$= \sum_{k=1}^{s}\sum_{j=1}^{n_k} \frac{1}{\phi} Y_{ij}^{[k]}(\beta_0 + \boldsymbol{\beta}_B^T \boldsymbol{x}_k)$$

$$+ \log \int_{-\infty}^{\infty} \exp\left[\sum_{k=1}^{s}\sum_{j=1}^{n_k} \frac{1}{\phi}\left\{Y_{ij}^{[k]} u - b(\beta_0 + \boldsymbol{\beta}_B^T \boldsymbol{x}_k + u)\right\} - \frac{u^2}{2\sigma^2}\right] du.$$

Therefore, by rewriting the conditional quasi-log-likelihood, we have,

$$\ell(\beta_0, \boldsymbol{\beta}_B, \sigma^2) = -\frac{m}{2}\log(2\pi\sigma^2) + \sum_{i=1}^{m}\sum_{k=1}^{s}\sum_{j=1}^{n_k} \frac{1}{\phi} Y_{ij}^{[k]}(\beta_0 + \boldsymbol{\beta}_B^T \boldsymbol{x}_k)$$

$$+ \sum_{i=1}^{m} \log \int_{-\infty}^{\infty} \exp\left[\sum_{k=1}^{s}\sum_{j=1}^{n_k} \frac{1}{\phi}\left\{Y_{ij}^{[k]} u - b(\beta_0 + \boldsymbol{\beta}_B^T \boldsymbol{x}_k + u)\right\} - \frac{u^2}{2\sigma^2}\right] du$$

$$+ C,$$

where C denotes a constant term independent of $(\beta_0, \boldsymbol{\beta}_B, \sigma^2)$.

5.3.2 Asymptotic Assumption for Support Point Sample Sizes

Define

$$n \equiv \frac{1}{s}\sum_{k=1}^{s} n_k = \text{ average of the support point replication sizes within each group.}$$

For the upcoming asymptotic analysis we assume:

(A4) The design sample sizes n_k diverge to ∞ in such a way that $n_k/(sn) \to \delta_k$ for constants $0 < \delta_k < 1$, $1 \le k \le s$.

The δ_k correspond to the so-called *design weights*.

5.3.3 Useful Notation

Let $\boldsymbol{v}$ be a generic $d \times 1$ vector. Then for $r = 0, 1, 2$ we define

$$\boldsymbol{v}^{\otimes r} \equiv \begin{cases} 1 & \text{for} \quad r = 0 \\ \boldsymbol{v} & \text{for} \quad r = 1 \\ \boldsymbol{v}\boldsymbol{v}^T & \text{for} \quad r = 2. \end{cases}$$

Note that, according to this notation, for all $(r, r') \in \{(0,0),(1,0),(1,1)\}$,

$$\boldsymbol{v}^{\otimes r}(\boldsymbol{v}^{\otimes r'})^T = \boldsymbol{v}^{\otimes(r+r')}.$$

For $r = 0, 1$ and $1 \le i \le m$, also define

$$\widetilde{\mathcal{G}}_{ri} \equiv \sum_{k=1}^{s}\sum_{j=1}^{n_k} \boldsymbol{x}_k^{\otimes r}\{Y_{ij}^{[k]} - b'(\beta_0 + \boldsymbol{\beta}_B^T\boldsymbol{x}_k + U_i)\} \quad \text{and} \quad \widetilde{\mathcal{H}}_{ri} \equiv \sum_{k=1}^{s} n_k \boldsymbol{x}_k^{\otimes r} b''(\beta_0 + \boldsymbol{\beta}_B^T\boldsymbol{x}_k + U_i).$$

Also, let $\widetilde{\mathcal{H}}'_{ir}$ be the same as $\widetilde{\mathcal{H}}_{ir}$ but with b'' replaced by b'''.

5.3.4 Key Moment Results

In this subsection, we present some key moment results required for the asymptotic derivations in the subsections to follow.

The Expectation of $\widetilde{\mathcal{G}}_{ri}$ Given U_i

Note that for $r, r' = 0, 1$:

$$\begin{aligned}
&E[\boldsymbol{x}_k^{\otimes r}(\boldsymbol{x}_k^{\otimes r'})^T\{Y_{ij}^{[k]} - b'(\beta_0 + \boldsymbol{\beta}_B^T\boldsymbol{x}_k + U_i)\}|U_i] \\
&\quad = \boldsymbol{x}_k^{\otimes r}(\boldsymbol{x}_k^{\otimes r'})^T E\{Y_{ij}^{[k]} - b'(\beta_0 + \boldsymbol{\beta}_B^T\boldsymbol{x}_k + U_i)|U_i\} \\
&\quad = \boldsymbol{x}_k^{\otimes r}(\boldsymbol{x}_k^{\otimes r'})^T E\{b'(\beta_0 + \boldsymbol{\beta}_B^T\boldsymbol{x}_k + U_i) - b'(\beta_0 + \boldsymbol{\beta}_B^T\boldsymbol{x}_k + U_i)\} \\
&\quad = \boldsymbol{O}.
\end{aligned}$$

Hence

$$E[\boldsymbol{x}_k^{\otimes r}(\boldsymbol{x}_k^{\otimes r'})^T\{Y_{ij}^{[k]} - b'(\beta_0 + \boldsymbol{\beta}_B^T\boldsymbol{x}_k + U_i)\}|U_i] = \boldsymbol{O}. \tag{5.7}$$

It follows immediately from (5.7) that

$$E(\widetilde{\mathcal{G}}_{0i}|U_i) = 0, \quad E(\widetilde{\mathcal{G}}_{1i}|U_i) = \boldsymbol{0} \quad \text{and} \quad E(\widetilde{\mathcal{G}}_{2i}|U_i) = \boldsymbol{O}.$$

The Expectation of $\widetilde{\mathcal{G}}_{ri}\,\widetilde{\mathcal{G}}_{r'i}^T$ Given U_i

Note that for all $(r, r') \in \{(0,0),(1,0),(1,1)\}$,

$$\begin{aligned}
E(\widetilde{\mathcal{G}}_{ri}\,\widetilde{\mathcal{G}}_{r'i}^T|U_i) &= E\Bigg(\left[\sum_{k=1}^{s}\sum_{j=1}^{n_k} \boldsymbol{x}_k^{\otimes r}\{Y_{ij}^{[k]} - b'(\beta_0 + \boldsymbol{\beta}_B^T\boldsymbol{x}_k + U_i)\}\right] \\
&\qquad \times \left[\sum_{k'=1}^{s}\sum_{j'=1}^{n_{k'}} \boldsymbol{x}_{k'}^{\otimes r'}\{Y_{ij'}^{[k']} - b'(\beta_0 + \boldsymbol{\beta}_B^T\boldsymbol{x}_{k'} + U_i)\}\right]^T \Bigg| U_i\Bigg) \\
&= \sum_{k=1}^{s}\sum_{k'=1}^{s}\sum_{j=1}^{n_k}\sum_{j'=1}^{n_{k'}} \boldsymbol{x}_k^{\otimes r}(\boldsymbol{x}_{k'}^{\otimes r'})^T E\Big[\{Y_{ij}^{[k]} - b'(\beta_0 + \boldsymbol{\beta}_B^T\boldsymbol{x}_k + U_i)\} \\
&\qquad \times\{Y_{ij'}^{[k']} - b'(\beta_0 + \boldsymbol{\beta}_B^T\boldsymbol{x}_{k'} + U_i)\}|U_i\Big].
\end{aligned} \tag{5.8}$$

First consider those terms for which

$$k = k' \quad \text{and} \quad j = j'. \tag{5.9}$$

Such terms have the form

$$\boldsymbol{x}_k^{\otimes r}(\boldsymbol{x}_k^{\otimes r'})^T \mathrm{Var}\left[\{Y_{ij}^{[k]} - b'(\beta_0 + \boldsymbol{\beta}_B^T \boldsymbol{x}_k + U_i)\}|U_i\right] = \phi b''(\beta_0 + \boldsymbol{\beta}_B^T \boldsymbol{x}_k + U_i).$$

Therefore, the contribution to (5.8) from the terms satisfying (5.9) is

$$\phi \sum_{k=1}^{s} n_k \boldsymbol{x}_k^{\otimes(r+r')} b''(\beta_0 + \boldsymbol{\beta}_B^T \boldsymbol{x}_k + U_i) = \phi \widetilde{\mathcal{H}}_{(r+r')i}.$$

Next consider those terms for which

$$k = k' \quad \text{and} \quad j \neq j'. \tag{5.10}$$

Since

$$\begin{aligned} &E[\{Y_{ij}^{[k]} - b'(\beta_0 + \boldsymbol{\beta}_B^T \boldsymbol{x}_k + U_i)\}\{Y_{ij'}^{[k]} - b'(\beta_0 + \boldsymbol{\beta}_B^T \boldsymbol{x}_k + U_i)\}|U_i] \\ &\quad = E[\{Y_{ij}^{[k]} - b'(\beta_0 + \boldsymbol{\beta}_B^T \boldsymbol{x}_k + U_i)\}|U_i]\, E[\{Y_{ij'}^{[k]} - b'(\beta_0 + \boldsymbol{\beta}_B^T \boldsymbol{x}_k + U_i)\}|U_i] \\ &\quad = 0, \end{aligned}$$

the contribution to (5.8) from the terms satisfying (5.10) is 0. Next consider those terms for which

$$k \neq k'. \tag{5.11}$$

Then $Y_{ij}^{[k]}$ and $Y_{ij}^{[k']}$ must be distinct random variables, which implies that

$$E[\{Y_{ij}^{[k]} - b'(\beta_0 + \boldsymbol{\beta}_B^T \boldsymbol{x}_k + U_i)\}\{Y_{ij}^{[k']} - b'(\beta_0 + \boldsymbol{\beta}_B^T \boldsymbol{x}_{k'} + U_i)\}|U_i] = 0 \quad \text{for all } k \neq k'.$$

Therefore, for $(r, r') \in \{(0,0), (1,0), (1,1)\}$ we have

$$E(\widetilde{\mathcal{G}}_{ri}\widetilde{\mathcal{G}}_{r'i}^T|U_i) = \phi \widetilde{\mathcal{H}}_{(r+r')i}.$$

Lastly, there is the issue when $(r, r') = (0, 1)$. Note the special case

$$E(\widetilde{\mathcal{G}}_{1i}\widetilde{\mathcal{G}}_{0i}^T|U_i) = \phi \widetilde{\mathcal{H}}_{1i}.$$

Taking the transpose on each side of this equation we obtain

$$E(\widetilde{\mathcal{G}}_{0i}\widetilde{\mathcal{G}}_{1i}^T|U_i) = \phi \widetilde{\mathcal{H}}_{1i}^T.$$

Then, the full set of results for $E(\widetilde{\mathcal{G}}_{ri}\widetilde{\mathcal{G}}_{r'i}^T|U_i)$ is:

$$\begin{aligned} E(\widetilde{\mathcal{G}}_{0i}\widetilde{\mathcal{G}}_{0i}^T|U_i) = \phi \widetilde{\mathcal{H}}_{0i}, \quad E(\widetilde{\mathcal{G}}_{0i}\widetilde{\mathcal{G}}_{1i}^T|U_i) = \phi \widetilde{\mathcal{H}}_{1i}^T, \\ E(\widetilde{\mathcal{G}}_{1i}\widetilde{\mathcal{G}}_{0i}^T|U_i) = \phi \widetilde{\mathcal{H}}_{1i}, \quad E(\widetilde{\mathcal{G}}_{1i}\widetilde{\mathcal{G}}_{1i}^T|U_i) = \phi \widetilde{\mathcal{H}}_{2i}. \end{aligned}$$

But since $\widetilde{\mathcal{G}}_{0i}$ is a scalar, these results can be simplified to

$$E(\widetilde{\mathcal{G}}_{0i}^2|U_i) = \phi \widetilde{\mathcal{H}}_{0i}, \quad E(\widetilde{\mathcal{G}}_{0i}\widetilde{\mathcal{G}}_{1i}|U_i) = \phi \widetilde{\mathcal{H}}_{1i} \quad \text{and} \quad E(\widetilde{\mathcal{G}}_{1i}\widetilde{\mathcal{G}}_{1i}^T|U_i) = \phi \widetilde{\mathcal{H}}_{2i}. \tag{5.12}$$

5.3.5 The Fisher Information Matrix

To construct an asymptotic approximation of the Fisher information matrix, we require asymptotic expansions of the scores and their quadratic expectations. Hence, we follow steps similar to those detailed under the proof for Theorem 12 for these computations as well. Putting together the resultant expressions from these computations, we have the following expression for the Fisher information

$$
\begin{aligned}
&I(\beta_0, \boldsymbol{\beta}_B, \sigma^2) \\
&= \begin{bmatrix}
\frac{m}{\sigma^2} + O(mn^{-1}) & O(m)\mathbf{1}_{d_B}^T & O(mn^{-1}) \\
O(m)\mathbf{1}_{d_B} & \frac{1}{\phi}\sum_{i=1}^{m} E\left(\widetilde{\mathcal{H}}_{2i} - \frac{\widetilde{\mathcal{H}}_{1i}^{\otimes 2}}{\widetilde{\mathcal{H}}_{0i}}\right) + O(m)\mathbf{1}_{d_B}^{\otimes 2} & O(m)\mathbf{1}_{d_B} \\
O(mn^{-1}) & O(m)\mathbf{1}_{d_B}^T & \frac{m}{2\sigma^4} + O(mn^{-1})
\end{bmatrix}.
\end{aligned}
$$

5.3.6 The Asymptotic D-Optimality Criterion

Next, we change the ordering of the parameters from $(\beta_0, \boldsymbol{\beta}_B, \sigma^2)$ to $(\beta_0, \sigma^2, \boldsymbol{\beta}_B)$. Then partition $I(\beta_0, \sigma^2, \boldsymbol{\beta}_1)$ according to

$$
I(\beta_0, \sigma^2, \boldsymbol{\beta}_B) = \begin{bmatrix} \widetilde{\boldsymbol{A}}_{11} & \widetilde{\boldsymbol{A}}_{12}^T \\ \widetilde{\boldsymbol{A}}_{12} & \widetilde{\boldsymbol{A}}_{22} \end{bmatrix}
$$

where

$$
\widetilde{\boldsymbol{A}}_{11} \equiv \begin{bmatrix} \frac{m}{\sigma^2} + O(mn^{-1}) & O(mn^{-1}) \\ O(mn^{-1}) & \frac{m}{2\sigma^4} + O(mn^{-1}) \end{bmatrix}, \quad \widetilde{\boldsymbol{A}}_{12} \equiv O(m)[\mathbf{1}_{d_B}\ \mathbf{1}_{d_B}],
$$

and

$$
\widetilde{\boldsymbol{A}}_{22} \equiv \frac{m}{\phi} E\left(\widetilde{\mathcal{H}}_{21} - \frac{\widetilde{\mathcal{H}}_{11}^{\otimes 2}}{\widetilde{\mathcal{H}}_{01}}\right) + O(m)\mathbf{1}_{d_B}^{\otimes 2}.
$$

Now, we apply a standard result concerning the determinant of a 2×2 block-partitioned matrix (e.g. Harville, 2008; Theorem 13.3.8) to obtain

$$
|I(\beta_0, \boldsymbol{\beta}_B, \sigma^2)| = \left|\widetilde{\boldsymbol{A}}_{11}\right| \left|\widetilde{\boldsymbol{A}}_{22} - \widetilde{\boldsymbol{A}}_{12}^T \widetilde{\boldsymbol{A}}_{11}^{-1} \widetilde{\boldsymbol{A}}_{12}\right|.
$$

It is easily verified that $\left|\widetilde{\boldsymbol{A}}_{11}\right| = m^2/(2\sigma^6) + O(mn^{-1})$ and $\widetilde{\boldsymbol{A}}_{12}^T \widetilde{\boldsymbol{A}}_{11}^{-1} \widetilde{\boldsymbol{A}}_{12} = O(m)\mathbf{1}_{d_B}^{\otimes 2}$. It follows that

$$
\frac{2\phi^{d_B}\sigma^6}{m^{d_B+2}} \left|I(\beta_0, \boldsymbol{\beta}_B, \sigma^2)\right| = \left|E\left(\widetilde{\mathcal{H}}_{21} - \frac{\widetilde{\mathcal{H}}_{11}^{\otimes 2}}{\widetilde{\mathcal{H}}_{01}}\right) + O(1)\mathbf{1}_{d_B}^{\otimes 2}\right| = \left|\boldsymbol{\Psi}_n + O(1)\mathbf{1}_{d_B}^{\otimes 2}\right| \tag{5.13}
$$

where $\boldsymbol{\Psi}_n \equiv E\left(\widetilde{\mathcal{H}}_{21} - \widetilde{\mathcal{H}}_{11}^{\otimes 2}/\widetilde{\mathcal{H}}_{01}\right)$. Since $\widetilde{\mathcal{H}}_{01} = O_P(n)$, $\widetilde{\mathcal{H}}_{11} = O_P(n)\mathbf{1}_{d_B}$ and $\widetilde{\mathcal{H}}_{21} = O_P(n)\mathbf{1}_{d_B}^{\otimes 2}$, we have $\boldsymbol{\Psi}_n = O(n)\mathbf{1}_{d_B}^{\otimes 2}$. Let $\lambda_1(\boldsymbol{M}), \ldots, \lambda_{d_B}(\boldsymbol{M})$ denote the eigenvalues of a generic $d_B \times d_B$ matrix $\boldsymbol{M}$. Then

$$\left|\boldsymbol{\Psi}_n + O(1)\mathbf{1}_{d_B}^{\otimes 2}\right| = \prod_{j=1}^{d_B} \lambda_j\left(\boldsymbol{\Psi}_n + O(1)\mathbf{1}_{d_B}^{\otimes 2}\right).$$

As a consequence of Theorem 8.1.4 (Wielandt-Hoffman) of Golub and Van Loan (2013),

$$\lambda_j\left(\boldsymbol{\Psi}_n + O(1)\mathbf{1}_{d}^{\otimes 2}\right) = \lambda_j\left(\boldsymbol{\Psi}_n\right) + O(1)$$

for each $1 \leq j \leq d_B$. Hence

$$\left|\boldsymbol{\Psi}_n + O(1)\mathbf{1}_{d_B}^{\otimes 2}\right| = |\boldsymbol{\Psi}_n| + O(1)\sum_{j=1}^{d_B} |\boldsymbol{\Psi}_n| / \lambda_j(\boldsymbol{\Psi}_n). \tag{5.14}$$

To obtain the order of magnitude of the $\lambda_j(\boldsymbol{\Psi}_n)$ we appeal to Theorem 8.1.3 (Gershgorin) of Golub and Van Loan (2013). Since all entries of $\boldsymbol{\Psi}_n$ are $O(n)$, the same is true for the lower and upper limits of each of the Gershgorin discs of $\boldsymbol{\Psi}_n$. Since each eigenvalue of $\boldsymbol{\Psi}_n$ is inside at least one Gershgorin disc, we have $\lambda_j(\boldsymbol{\Psi}_n) = O(n)$, $1 \leq j \leq d_B$. It follows from this fact and (5.14) that

$$\left|\boldsymbol{\Psi}_n + O(1)\mathbf{1}_{d_B}^{\otimes 2}\right| = |\boldsymbol{\Psi}_n| \{1 + o(1)\}.$$

In view of (5.13), the determinant of $\left|I(\beta_0, \boldsymbol{\beta}_B, \sigma^2)\right|$ is proportional to a quantity with leading term $|\boldsymbol{\Psi}_n|$ as $n \to \infty$. Recalling that $n_k = ns\delta_k$ and dividing through by ns we can assert that approximate locally D-optimal designs, based on the exact leading term behaviour of the determinant of the Fisher information matrix, are those which maximize

$$\left| E\left[\sum_{k=1}^{s} \delta_k \boldsymbol{x}_k^{\otimes r} b''(\beta_0 + \boldsymbol{\beta}_B^T \boldsymbol{x}_k + U) - \frac{\left(\sum_{k=1}^{s} \delta_k \boldsymbol{x}_k b''(\beta_0 + \boldsymbol{\beta}_B^T \boldsymbol{x}_k + U)\right)^{\otimes 2}}{\sum_{k=1}^{s} \delta_k b''(\beta_0 + \boldsymbol{\beta}_B^T \boldsymbol{x}_k + U)}\right]\right|, \; U \sim N(0, \sigma^2),$$

over the design weights δ_k and support points $\boldsymbol{x}_k$, $1 \leq k \leq s$.

The following quantity can then be used to obtain asymptotic D-optimal designs:

$$\left|\sum_{k=1}^{s} \delta_k E\{b''(\beta_0 + \boldsymbol{\beta}_B^T \boldsymbol{x}_k + U)\}\boldsymbol{x}_k^{\otimes 2} - E\left\{\frac{\left(\sum_{k=1}^{s} \delta_k b''(\beta_0 + \boldsymbol{\beta}_B^T \boldsymbol{x}_k + U)\boldsymbol{x}_k\right)^{\otimes 2}}{\sum_{k=1}^{s} \delta_k b''(\beta_0 + \boldsymbol{\beta}_B^T \boldsymbol{x}_k + U)}\right\}\right| \tag{5.15}$$

where $U \sim N(0, \sigma^2)$. An equivalent integral form is:

$$\left| \begin{array}{l} \displaystyle\sum_{k=1}^{s} \delta_k \boldsymbol{x}_k^{\otimes 2} \int_{-\infty}^{\infty} b''(\beta_0 + \boldsymbol{\beta}_B^T \boldsymbol{x}_k + u) \exp\{-u^2/(2\sigma^2)\}\, du \\ \qquad - \displaystyle\int_{-\infty}^{\infty} \frac{\left(\displaystyle\sum_{k=1}^{s} \delta_k b''(\beta_0 + \boldsymbol{\beta}_B^T \boldsymbol{x}_k + u) \boldsymbol{x}_k \right)^{\otimes 2} \exp\{-u^2/(2\sigma^2)\}\, du}{\displaystyle\sum_{k=1}^{s} \delta_k b''(\beta_0 + \boldsymbol{\beta}_B^T \boldsymbol{x}_k + u)} \end{array} \right| . \tag{5.16}$$

5.3.7 Alternative Final Asymptotic D-optimality Criterion

Consider the matrix-valued function $\boldsymbol{\Omega}(u)$ given by

$$\boldsymbol{\Omega}(u) \equiv \sum_{k=1}^{s} \delta_k b''(\beta_0 + \boldsymbol{\beta}_B^T + u) \begin{bmatrix} 1 & \boldsymbol{x}_k \\ \boldsymbol{x}_k^T & \boldsymbol{x}_k \boldsymbol{x}_k^T \end{bmatrix} .$$

Then

$$\boldsymbol{\Omega}(u) = \begin{bmatrix} \displaystyle\sum_{k=1}^{s} \delta_k b''(\beta_0 + \boldsymbol{\beta}_B^T + u) & \displaystyle\sum_{k=1}^{s} \delta_k \boldsymbol{x}_k b''(\beta_0 + \boldsymbol{\beta}_B^T + u) \\ \displaystyle\sum_{k=1}^{s} \delta_k \boldsymbol{x}_k^T b''(\beta_0 + \boldsymbol{\beta}_B^T + u) & \displaystyle\sum_{k=1}^{s} \delta_k \boldsymbol{x}_k \boldsymbol{x}_k^T b''(\beta_0 + \boldsymbol{\beta}_B^T + u) \end{bmatrix}$$

and the lower right $d_B \times d_B$ block of $\boldsymbol{\Omega}(u)^{-1}$ is

$$\left\{ \sum_{k=1}^{s} \delta_k \boldsymbol{x}_k \boldsymbol{x}_k^T b''(\beta_0 + \boldsymbol{\beta}_B^T + u) - \frac{\left(\displaystyle\sum_{k=1}^{s} \delta_k \boldsymbol{x}_k b''(\beta_0 + \boldsymbol{\beta}_B^T + u) \right)^{\otimes 2}}{\displaystyle\sum_{k=1}^{s} \delta_k b''(\beta_0 + \boldsymbol{\beta}_B^T + u)} \right\}^{-1} .$$

The inverse of this function is the function of u that is multiplied by $\exp\{-u^2/(2\sigma^2)\}$ in the determinant of (5.16). Therefore, an alternative expression for (5.16) is

$$\left| \int_{-\infty}^{\infty} \left\{ \text{lower right } d \times d \text{ block of } \left(\sum_{k=1}^{s} \delta_k b''(\beta_0 + \boldsymbol{\beta}_1^T \boldsymbol{x}_k + u) \begin{bmatrix} 1 & \boldsymbol{x}_k \\ \boldsymbol{x}_k^T & \boldsymbol{x}_k \boldsymbol{x}_k^T \end{bmatrix} \right)^{-1} \right\}^{-1} \times \exp\{-u^2/(2\sigma^2)\}\, du \right|$$

and is the form of the asymptotic D-optimality criterion presented in Theorem 13.

5.3.8 Special Distribution Cases

Poisson Special Case

In the Poisson special case, $b'' = \exp$ and therefore, we have,

$$b''(\beta_0 + \boldsymbol{\beta}_B^T \boldsymbol{x}_k + U) = \exp(U + \beta_0)\, \exp(\boldsymbol{\beta}_B^T \boldsymbol{x}_k).$$

We can show that $\exp(U + \beta_0)$ comes out as a multiplicative factor in (5.15). This is the only random factor in the expectation expression. Therefore, an equivalent D-optimality criterion is

$$\left| \sum_{k=1}^{s} \delta_k \exp(\boldsymbol{\beta}_B^T \boldsymbol{x}_k \} \boldsymbol{x}_k^{\otimes 2} - \frac{\left(\sum_{k=1}^{s} \delta_k \exp(\boldsymbol{\beta}_B^T \boldsymbol{x}_k) \boldsymbol{x}_k \right)^{\otimes 2}}{\sum_{k=1}^{s} \delta_k \exp(\boldsymbol{\beta}_B^T \boldsymbol{x}_k)} \right|. \tag{5.17}$$

Define

$$\boldsymbol{\Lambda} \equiv \sum_{k=1}^{s} \delta_k \exp(\boldsymbol{\beta}_B^T \boldsymbol{x}_k) \begin{bmatrix} 1 & \boldsymbol{x}_k^T \\ \boldsymbol{x}_k & \boldsymbol{x}_k \boldsymbol{x}_k^T \end{bmatrix}.$$

Then

$$\boldsymbol{\Lambda} = \begin{bmatrix} \sum_{k=1}^{s} \delta_k \exp(\boldsymbol{\beta}_B^T \boldsymbol{x}_k) & \sum_{k=1}^{s} \delta_k \exp(\boldsymbol{\beta}_B^T \boldsymbol{x}_k) \boldsymbol{x}_k^T \\ \sum_{k=1}^{s} \delta_k \exp(\boldsymbol{\beta}_B^T \boldsymbol{x}_k) \boldsymbol{x}_k & \sum_{k=1}^{s} \delta_k \exp(\boldsymbol{\beta}_B^T \boldsymbol{x}_k) \boldsymbol{x}_k^{\otimes 2} \end{bmatrix}.$$

From the result concerning the determinant of a 2×2 block partitioned matrix, we obtain

$$|\boldsymbol{\Lambda}| = \left| \left[\sum_{k=1}^{s} \delta_k \exp(\boldsymbol{\beta}_B^T \boldsymbol{x}_k) \right] \right| \left| \sum_{k=1}^{s} \delta_k \exp(\boldsymbol{\beta}_B^T \boldsymbol{x}_k \} \boldsymbol{x}_k^{\otimes 2} - \frac{\left(\sum_{k=1}^{s} \delta_k \exp(\boldsymbol{\beta}_B^T \boldsymbol{x}_k) \boldsymbol{x}_k \right)^{\otimes 2}}{\sum_{k=1}^{s} \delta_k \exp(\boldsymbol{\beta}_B^T \boldsymbol{x}_k)} \right|$$

It follows that (5.17) is equivalent to

$$|\boldsymbol{\Lambda}| \Big/ \sum_{k=1}^{s} \delta_k \exp(\boldsymbol{\beta}_1^T \boldsymbol{x}_k).$$

In other words, an equivalent asymptotic D-optimality criterion for Poisson mixed models is

$$\left| \sum_{k=1}^{s} \delta_k \exp(\boldsymbol{\beta}_B^T \boldsymbol{x}_k) \begin{bmatrix} 1 & \boldsymbol{x}_k^T \\ \boldsymbol{x}_k & \boldsymbol{x}_k^{\otimes 2} \end{bmatrix} \right| \Big/ \sum_{k=1}^{s} \delta_k \exp(\boldsymbol{\beta}_B^T \boldsymbol{x}_k).$$

Note that, in the Poisson case, this means that the asymptotic locally D-optimal designs do not depend on β_0 or σ^2. They only depend on $\boldsymbol{\beta}_B$. Also, for Poisson regression models, the D-optimality criterion is

$$\left| \sum_{k=1}^{s} \delta_k \exp(\boldsymbol{\beta}_B^T \boldsymbol{x}_k) \begin{bmatrix} 1 & \boldsymbol{x}_k^T \\ \boldsymbol{x}_k & \boldsymbol{x}_k^{\otimes 2} \end{bmatrix} \right|.$$

Therefore, asymptotic D-optimality for Poisson mixed models involves maximisation of a quantity that is similar, but not identical, to that for Poisson regression models.

Logistic Special Case

In the logistic mixed model special case, after taking out multiplicative factors, the quantity (5.16) becomes:

$$\left| \sum_{k=1}^{s} \delta_k \boldsymbol{x}_k^{\otimes 2} \int_{-\infty}^{\infty} \frac{\exp\{-u^2/(2\sigma^2)\}\, du}{1+\cosh(\beta_0 + \boldsymbol{\beta}_B^T \boldsymbol{x}_k + u)} - \int_{-\infty}^{\infty} \frac{\left(\sum_{k=1}^{s} \delta_k \boldsymbol{x}_k / \{1+\cosh(\beta_0 + \boldsymbol{\beta}_B^T \boldsymbol{x}_k + u)\} \right)^{\otimes 2} \exp\{-u^2/(2\sigma^2)\}\, du}{\sum_{k=1}^{s} \delta_k / \{1+\cosh(\beta_0 + \boldsymbol{\beta}_B^T \boldsymbol{x}_k + u)\}} \right|.$$

Chapter 6

Thouless-Anderson-Palmer Enhancement of Generalized Linear Mixed Models

Frequentist inference for GLMMs is hindered by integral intractability problems. In machine learning contexts, Thouless-Anderson-Palmer approaches, highlighted in Subsection 1.11.1 in Chapter 1, can not only help overcome issues involving intractable integrals but theoretically also provide better approximations. However, statistical applications such as longitudinal data analysis and multilevel models analysis have not been investigated at all. Therefore, in this chapter, the goal is to apply the Thouless-Anderson-Palmer frequentist variational approach to generalized linear mixed models with canonical links.

Firstly, the TAP enhancement approach is explained and we obtain a result detailing the explicit form of the TAP approximate negative log-likelihood expression for GLMMs with canonical links, which can then be locally minimized to obtain TAP estimates of the true model parameters.

We then carry out simulation studies to investigate the use of the TAP enhancement approach in practical settings and compare it to the popular Gaussian variational approximation approach for simplified Poisson generalized linear mixed model set-ups.

This chapter is broken up into several parts. Section 6.1 details the model set-up used throughout sections 6.2 to 6.4. Section 6.2 then provides an explicit expression for the GVA log-likelihood. Following that, Section 6.3 provide details regarding the TAP enhancement approach which builds on the GVA approach. An explicit result

for the TAP negative approximate log-likelihood for GLMMs with canonical links is then provided in Section 6.4. Lastly, Section 6.5 delves into two simulation studies constructed for Poisson GLMMs and compares the quality of estimates obtained from the TAP approach against the GVA approach.

6.1 Model Description

Consider the use of a simple canonical link generalized linear mixed model as follows where

$$\begin{aligned} &Y_{ij}|X_{ij}, U_i \text{ are independent having density function (4.1) with} \\ &\text{natural parameter } \beta_0^0 + \beta_1^0 X_{ij} + U_i \text{ such that the } U_i \text{ are independent} \\ &N\left(0, \left(\sigma^2\right)^0\right) \text{ random variables.} \end{aligned} \tag{6.1}$$

Here, the values of (X_{ij}, Y_{ij}) are observed for $1 \leq i \leq m$ and $1 \leq j \leq n$. We have assumed that the X_{ij} and U_i are independent random variables. In addition, the X_{ij} are each assumed as having the same distribution as the random vector X. The U_i are the unobserved random effects variables and are assumed to be having the same distribution as the random vector U.

Let $\boldsymbol{\beta} \equiv (\beta_0, \beta_1)$ be the vector of fixed parameters. Then the model parameters for this set-up are $(\boldsymbol{\beta}, \sigma^2)$.

6.2 The Gaussian Variational Approximate Log-Likelihood

Using the model description in (6.1), one can then obtain $\ell(\boldsymbol{\beta}, \sigma^2)$, the conditional log-likelihood of $(\boldsymbol{\beta}, \sigma^2)$, where

$$\begin{aligned} \ell\left(\boldsymbol{\beta}, \sigma^2\right) = &\sum_{i=1}^{m}\sum_{j=1}^{n}\left\{Y_{ij}\left(\beta_0 + \beta_1 X_{ij}\right) + c(Y_{ij})\right\} - \frac{m}{2}\log(2\pi\sigma^2) \\ &+ \sum_{i=1}^{m}\log\int_{-\infty}^{\infty}\exp\left\{\sum_{j=1}^{n}\left(Y_{ij}u - b(\beta_0 + \beta_1 X_{ij} + u)\right) - \frac{u^2}{2\sigma^2}\right\}du. \end{aligned} \tag{6.2}$$

Maximum likelihood estimation is hindered due to the m intractable integrals arising in the expression in (6.2). However, each of the m integrals can be re-written to overcome

this obstacle by re-expressing the ith integral as follows (Hall et al., 2011)

$$\int_{-\infty}^{\infty} \exp\left\{\sum_{j=1}^{n}(Y_{ij}u - b(\beta_0+\beta_1 X_{ij}+u)) - \frac{u^2}{2\sigma^2}\right\} \frac{e^{-(1/2)(u-\mu_i)^2\lambda_i}/\sqrt{2\pi\lambda_i}}{e^{-(1/2)(u-\mu_i)^2\lambda_i}/\sqrt{2\pi\lambda_i}} du$$
$$= \sqrt{2\pi\lambda_i} E_{\tilde{U}_i}\left[\exp\left\{\sum_{j=1}^{n}\left(Y_{ij}\tilde{U}_i - b(\beta_0+\beta_1 X_{ij}+\tilde{U}_i)\right) - \frac{\tilde{U}_i^2}{2\sigma^2} + \frac{(\tilde{U}_i-\mu_i)^2}{2\lambda_i}\right\}\right]$$

where $E_{\tilde{U}_i}$ denotes the expectation with respect to the random variable $\tilde{U}_i \sim N(\mu_i, \lambda_i)$ with $\lambda_i > 0$, for $1 \le i \le m$. Here, $(\boldsymbol{\mu}, \boldsymbol{\lambda})$ are known as the variational parameters where $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_m)$ and $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_m)$. Using Jensen's inequality, one can then obtain the following lower bound

$$\log E_{\tilde{U}_i}\left[\exp\left\{\sum_{j=1}^{n}\left(Y_{ij}\tilde{U}_i - b(\beta_0+\beta_1 X_{ij}+\tilde{U}_i)\right) - \frac{\tilde{U}_i^2}{2\sigma^2} + \frac{(\tilde{U}_i-\mu_i)^2}{2\lambda_i}\right\}\right]$$
$$\ge E_{\tilde{U}_i}\left\{\sum_{j=1}^{n}\left(Y_{ij}\tilde{U}_i - b(\beta_0+\beta_1 X_{ij}+\tilde{U}_i)\right) - \frac{\tilde{U}_i^2}{2\sigma^2} + \frac{(\tilde{U}_i-\mu_i)^2}{2\lambda_i}\right\},$$

which is now tractable. Then, the Gaussian variational approximation to $\ell(\boldsymbol{\beta}, \sigma^2)$ is derived as,

$$\begin{aligned}
&\underline{\ell}_{\mathrm{GVA}}(\boldsymbol{\beta}, \sigma^2, \boldsymbol{\mu}, \boldsymbol{\lambda}) \\
&= \sum_{i=1}^{m}\sum_{j=1}^{n}\{Y_{ij}(\beta_0+\beta_1 X_{ij}) + c(Y_{ij})\} - \frac{m}{2}\log(2\pi\sigma^2) + \sum_{i=1}^{m}\log(\sqrt{2\pi\lambda_i}) \\
&\quad + \sum_{i=1}^{m} E_{\tilde{U}_i}\left\{\sum_{j=1}^{n}\left(Y_{ij}\tilde{U}_i - b(\beta_0+\beta_1 X_{ij}+\tilde{U}_i)\right) - \frac{\tilde{U}_i^2}{2\sigma^2} + \frac{(\tilde{U}_i-\mu_i)^2}{2\lambda_i}\right\} \qquad (6.3) \\
&= \sum_{i=1}^{m} E_{\tilde{U}_i}\left[\sum_{j=1}^{n}\left\{Y_{ij}(\beta_0+\beta_1 X_{ij}+\tilde{U}_i) - b(\beta_0+\beta_1 X_{ij}+\tilde{U}_i) + c(Y_{ij})\right\}\right. \\
&\quad \left. - \frac{\tilde{U}_i^2}{2\sigma^2} - \frac{1}{2}\log(2\pi\sigma^2)\right] + \frac{1}{2}\sum_{i=1}^{m}\{1+\log(2\pi\lambda_i)\}.
\end{aligned}$$

Note that

$$\ell(\boldsymbol{\beta}, \sigma^2) \ge \underline{\ell}_{\mathrm{GVA}}(\boldsymbol{\beta}, \sigma^2, \boldsymbol{\mu}, \boldsymbol{\lambda})$$

for all vectors $\boldsymbol{\mu}$ and $\boldsymbol{\lambda}$.

6.3 Overview of Thouless-Anderson-Palmer Enhancement

In this section, we provide details on how the TAP enhancement approach builds upon the GVA approach. Firstly, for each $1 \leq i \leq m$, define the following data vectors:

$$\boldsymbol{Y}_i \equiv (Y_{i1}, \dots Y_{in}) \text{ and } \boldsymbol{X}_i \equiv (X_{i1}, \dots X_{in}) .$$

The GVA negative log-likelihood can then be expressed as follows (Johnstone, 2022)

$$-\underline{\ell}_{\text{GVA}}(\boldsymbol{\beta}, \sigma^2, \boldsymbol{\mu}, \boldsymbol{\lambda}) = -\frac{1}{2}\sum_{i=1}^{m}\{1 + \log(2\pi\lambda_i)\} + \sum_{i=1}^{m} E_{\tilde{U}_i}\left\{\Psi_i(\tilde{U}_i)|\boldsymbol{Y}_i, \boldsymbol{X}_i\right\},$$

where

$$\Psi_i(\tilde{U}_i) = \sum_{j=1}^{n}\left\{-Y_{ij}(\beta_0 + \beta_1 X_{ij} + \tilde{U}_i) + b(\beta_0 + \beta_1 X_{ij} + \tilde{U}_i) - c(Y_{ij})\right\} + \frac{\tilde{U}_i^2}{2\sigma^2} + \frac{1}{2}\log(2\pi\sigma^2).$$

The TAP enhancement approach theoretically obtains better approximations than the GVA approach by enhancing the expression in $-\underline{\ell}_{\text{GVA}}(\boldsymbol{\beta}, \sigma^2, \boldsymbol{\mu}, \boldsymbol{\lambda})$ through the addition of the Onsager's correction term to $-\underline{\ell}_{\text{GVA}}(\boldsymbol{\beta}, \sigma^2, \boldsymbol{\mu}, \boldsymbol{\lambda})$, which was first introduced in Thouless et al. (1977). The Onsager's correction term is defined as follows

$$-1/2\sum_{i=1}^{m}\xi_i\left(\mu_i, \lambda_i; \boldsymbol{X}_i, \boldsymbol{Y}_i, \beta_0, \beta_1, \sigma^2\right),$$

with

$$\begin{aligned}&\xi_i\left(\mu_i, \lambda_i; \boldsymbol{Y}_i, \boldsymbol{X}_i, \beta_0, \beta_1, \sigma^2\right)\\ &\equiv \text{Var}\{\Psi_i(\tilde{U}_i)|\boldsymbol{Y}_i, \boldsymbol{X}_i\} - \lambda_i\left[E\{\Psi_i'(\tilde{U}_i)|\boldsymbol{Y}_i, \boldsymbol{X}_i\}\right]^2 - \frac{\lambda_i^2}{2}\left[E\{\Psi_i''(\tilde{U}_i)|\boldsymbol{Y}_i, \boldsymbol{X}_i\}\right]^2,\end{aligned} \tag{6.4}$$

where (6.4) was obtained based on personal communication with Professor Song Mei (University of California, Berkeley, U.S.A) and Professor Iain Johnstone (Stanford University, U.S.A). They derived a working expression for the main quantity in the Onsager's correction term for density functions from exponential families.

Following that, the TAP approximate negative log-likelihood can be obtained as,

$$\begin{aligned}-\underline{\ell}_{\text{TAP}}(\boldsymbol{\beta}, \sigma^2, \boldsymbol{\mu}, \boldsymbol{\lambda}) &= -\underline{\ell}_{\text{GVA}}(\boldsymbol{\beta}, \sigma^2, \boldsymbol{\mu}, \boldsymbol{\lambda}) - \frac{1}{2}\sum_{i=1}^{m} \xi_i \left(\mu_i, \lambda_i; \boldsymbol{X}_i, \boldsymbol{Y}_i, \beta_0, \beta_1, \sigma^2\right) \\ &= -\frac{1}{2}\sum_{i=1}^{m}\{1 + \log(2\pi\lambda_i)\} + \sum_{i=1}^{m} E\left\{\Psi_i(\tilde{U}_i)|\boldsymbol{Y}_i, \boldsymbol{X}_i\right\} \\ &\quad - \frac{1}{2}\sum_{i=1}^{m}\Bigg(\text{Var}\{\Psi_i(\tilde{U}_i)|\boldsymbol{Y}_i, \boldsymbol{X}_i\} - \lambda_i\left[E\{\Psi_i'(\tilde{U}_i)|\boldsymbol{Y}_i, \boldsymbol{X}_i\}\right]^2 \\ &\quad - \frac{\lambda_i^2}{2}\left[E\{\Psi_i''(\tilde{U}_i)|\boldsymbol{Y}_i, \boldsymbol{X}_i\}\right]^2\Bigg).\end{aligned} \tag{6.5}$$

Now define

$$Y_{i\bullet} \equiv \sum_{j=1}^{n} Y_{ij} \quad \text{and} \quad \mathcal{A}_i(\tilde{U}_i) \equiv \sum_{j=1}^{n} b(\beta_0 + \beta_1 X_{ij} + \tilde{U}_i).$$

Then, we can re-express $\Psi_i(\tilde{U}_i)$ as

$$\Psi_i(\tilde{U}_i) = -\sum_{j=1}^{n} Y_{ij}(\beta_0 + \beta_1 X_{ij} + \tilde{U}_i) + \mathcal{A}_i(\tilde{U}_i) - \sum_{j=1}^{n} c(Y_{ij}) + \frac{\tilde{U}_i^2}{2\sigma^2} + \frac{1}{2}\log(2\pi\sigma^2).$$

It follows that

$$\Psi_i'(\tilde{U}_i) = -Y_{i\bullet} + \mathcal{A}_i'(\tilde{U}_i) + \frac{\tilde{U}_i}{\sigma^2}$$

and

$$\Psi_i''(\tilde{U}_i) = \mathcal{A}_i''(\tilde{U}_i) + \frac{1}{\sigma^2}.$$

Also note that

$$\mathcal{A}_i'(\tilde{U}_i) = \sum_{j=1}^{n} b'(\beta_0 + \beta_1 X_{ij} + \tilde{U}_i) \quad \text{and} \quad \mathcal{A}_i''(\tilde{U}_i) = \sum_{j=1}^{n} b''(\beta_0 + \beta_1 X_{ij} + \tilde{U}_i).$$

6.4 The Thouless-Anderson-Palmer Approximate Negative Log-Likelihood

In this section, we present a result detailing the Thouless-Anderson-Palmer approximate negative log-likelihood expression for canonical link GLMMs after having solved for the equation in (6.5). First, for $p, q \in \{0, 1, 2\}, r > 0$ and $s, t \in \mathbb{R}$, define

$$\mathcal{Q}(p, q, r, s, t) \equiv (2\pi)^{-1/2}\int_{-\infty}^{\infty}(s + rx)^p b^{(q)}(t + rx)\exp\left(-\frac{x^2}{2}\right)dx. \tag{6.6}$$

Also, for $r > 0$ and $s, t \in \mathbb{R}$, define

$$\mathcal{R}(r, s, t) \equiv (2\pi)^{-1/2} \int_{-\infty}^{\infty} b(s + rx) b(t + rx) \exp\left(-\frac{x^2}{2}\right) dx. \tag{6.7}$$

We then have the following result.

Result 2. *Consider the model set-up as in* (6.1)*. Then the Thouless-Anderson-Palmer approximate negative likelihood is*

$$-\underline{\ell}_{TAP}(\boldsymbol{\beta}, \sigma^2, \boldsymbol{\mu}, \boldsymbol{\lambda}) = -\underline{\ell}_{GVA}(\boldsymbol{\beta}, \sigma^2, \boldsymbol{\mu}, \boldsymbol{\lambda}) - \frac{1}{2} \sum_{i=1}^{m} \xi_i \left(\mu_i, \lambda_i; \boldsymbol{X}_i, \boldsymbol{Y}_i, \beta_0, \beta_1, \sigma^2\right),$$

where

$$\underline{\ell}_{GVA}(\boldsymbol{\beta}, \sigma^2, \boldsymbol{\mu}, \boldsymbol{\lambda}) = \sum_{i=1}^{m} E_{\tilde{U}_i} \Bigg[\sum_{j=1}^{n} \Big\{ Y_{ij}(\beta_0 + \beta_1 X_{ij} + \tilde{U}_i) - b(\beta_0 + \beta_1 X_{ij} + \tilde{U}_i) + c(Y_{ij}) \Big\} - \frac{\tilde{U}_i^2}{2\sigma^2} - \frac{1}{2} \log(2\pi\sigma^2) \Bigg] + \frac{1}{2} \sum_{i=1}^{m} \{1 + \log(2\pi\lambda_i)\}$$

and

$$\begin{aligned}
&\xi_i \left(\mu_i, \lambda_i; \boldsymbol{Y}_i, \boldsymbol{X}_i, \beta_0, \beta_1, \sigma^2\right) \\
&= \sum_{j=1}^{n} \Bigg\{ \left(2Y_{i\bullet}\mu_i - \frac{\lambda_i + \mu_i^2}{\sigma^2}\right) \mathcal{Q}(0, 0, \sqrt{\lambda_i}, \mu_i, \beta_0 + \beta_1 X_{ij} + \mu_i) \\
&\quad + 2\lambda_i \left(Y_{i\bullet} - \frac{\mu_i}{\sigma^2}\right) \mathcal{Q}(0, 1, \sqrt{\lambda_i}, \mu_i, \beta_0 + \beta_1 X_{ij} + \mu_i) \\
&\quad - 2Y_{i\bullet} \mathcal{Q}(1, 0, \sqrt{\lambda_i}, \mu_i, \beta_0 + \beta_1 X_{ij} + \mu_i) \\
&\quad - \frac{\lambda_i^2}{\sigma^2} \mathcal{Q}(0, 2, \sqrt{\lambda_i}, \mu_i, \beta_0 + \beta_1 X_{ij} + \mu_i) + \frac{1}{\sigma^2} \mathcal{Q}(2, 0, \sqrt{\lambda_i}, \mu_i, \beta_0 + \beta_1 X_{ij} + \mu_i) \\
&\quad - \mathcal{Q}(0, 0, \sqrt{\lambda_i}, \mu_i, \beta_0 + \beta_1 X_{ij} + \mu_i)^2 - \lambda_i \mathcal{Q}(0, 1, \sqrt{\lambda_i}, \mu_i, \beta_0 + \beta_1 X_{ij} + \mu_i)^2 \\
&\quad - \frac{\lambda_i^2}{2} \mathcal{Q}(0, 2, \sqrt{\lambda_i}, \mu_i, \beta_0 + \beta_1 X_{ij} + \mu_i)^2 \\
&\quad + \sum_{j'=1}^{n} \mathcal{R}(\sqrt{\lambda_i}, \beta_0 + \beta_1 X_{ij} + \mu_i, \beta_0 + \beta_1 X_{ij'} + \mu_i) \Bigg\}.
\end{aligned}$$

The proof for Result 2 is provided in the appendix under Subsection 6.6.1.

6.5 Thouless-Anderson-Palmer Enhancement for Poisson Generalized Linear Mixed Models

In the Poisson case, when $b(x) = \exp(x)$, $Q(p, q, r, s, t)$ and $R(r, s, t)$ admit exact expressions when evaluating Result 2. However, for general b functions, numerical integration is required for evaluating (6.6) and (6.7). Hence in this section, we work with a Poisson generalized linear mixed model and capitalize on having exact expressions to work with. Here we consider the model set-up of Hall et al. (2011) where

$$\begin{aligned} &Y_{ij}|X_{ij}, U_i \text{ independent Poisson with mean } \exp(\beta_0^0 + \beta_1^0 X_{ij} + U_i), \\ &\text{such that the } U_i \text{ are independent } N(0, (\sigma^2)^0). \end{aligned} \tag{6.8}$$

In Subsection 6.5.1, we detail the expression for the Gaussian variational approximate log-likelihood for the model set-up in (6.8). Similarly, in Subsection 6.5.2, the expression for the Thouless-Anderson-Palmer approximate negative log-likelihood is presented. Subsection 6.5.3 moves on to study the optimisation issues present when using the TAP approach by exploring a simpler version of the set-up in (6.8) with $m = 1$. Lastly, in Subsection 6.5.4, a full simulation study is carried out to assess and compare the accuracy of the estimates of the model parameters across the GVA and TAP approaches for the model set-up in (6.8).

6.5.1 The Gaussian Variational Approximate Log-Likelihood for Simulation Set-Up

Substituting $b(x) = \exp(x)$ and $c(x) = -\log(x!)$ into (7.7), the Gaussian variational approximate log-likelihood for the model set-up in (6.8) is

$$\begin{aligned} \underline{\ell}_{\text{GVA}}(\boldsymbol{\beta}, \sigma^2, \boldsymbol{\mu}, \boldsymbol{\lambda}) &= \sum_{i=1}^{m} E_{\tilde{U}_i}\Bigg[\sum_{j=1}^{n}\Big\{Y_{ij}(\beta_0 + \beta_1 X_{ij} + \tilde{U}_i) - e^{\beta_0+\beta_1 X_{ij}+\tilde{U}_i} - \log(Y_{ij}!)\Big\} \\ &\qquad - \frac{\tilde{U}_i^2}{2\sigma^2} - \frac{1}{2}\log(2\pi\sigma^2)\Bigg] + \frac{1}{2}\sum_{i=1}^{m}\{1 + \log(2\pi\lambda_i)\} \\ &= \sum_{i=1}^{m}\sum_{j=1}^{n}\Big\{Y_{ij}(\beta_0 + \beta_1 X_{ij} + \mu_i) - e^{\beta_0+\beta_1 X_{ij}+\mu_i+\frac{1}{2}\lambda_i}\Big\} \\ &\qquad - \sum_{i=1}^{m}\frac{\mu_i^2 + \lambda_i}{2\sigma^2} - \frac{m}{2}\log(\sigma^2) + \frac{1}{2}\sum_{i=1}^{m}\log(\lambda_i) + C, \end{aligned}$$

with vectors $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_m)$ and $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_m)$ as the variational parameters. Also, note that C denotes the constant term independent of any model or variational parameters.

6.5.2 The Thouless-Anderson-Palmer Negative Approximate Log-Likelihood for Simulation Set-Up

We start out with the following result:

Result 3. *Consider the model set-up as in* (6.8)*. Then the Onsager's correction term in the Thouless-Anderson-Palmer negative approximate likelihood is*

$$
\begin{aligned}
&-\frac{1}{2}\sum_{i=1}^{m}\xi_i\left(\mu_i, \lambda_i; \boldsymbol{X}_i, \boldsymbol{Y}_i, \beta_0, \beta_1, \sigma^2\right) \\
&= -\frac{1}{2}\sum_{i=1}^{m}\left[\left\{\exp(\lambda_i) - 1 - \lambda_i - \frac{1}{2}\lambda_i^2\right\}\exp(2\mu_i + \lambda_i)\left\{\sum_{j=1}^{n}\exp(\beta_0 + \beta_1 X_{ij})\right\}^2\right].
\end{aligned}
$$

The details for the derivations leading to Result 3 is in the appendix under Subsection 6.6.3. By using Results 2 and 3 together, we have that the TAP negative approximate log-likelihood for the model set-up in (6.8) is

$$
\begin{aligned}
&-\underline{\ell}_{\mathrm{TAP}}(\boldsymbol{\beta}, \sigma^2, \boldsymbol{\mu}, \boldsymbol{\lambda}) \\
&= \sum_{i=1}^{m}\sum_{j=1}^{n}\left\{-Y_{ij}(\beta_0 + \beta_1 X_{ij} + \mu_i) + e^{\beta_0 + \beta_1 X_{ij} + \mu_i + \frac{1}{2}\lambda_i}\right\} + \sum_{i=1}^{m}\frac{\mu_i^2 + \lambda_i}{2\sigma^2} + \frac{m}{2}\log(\sigma^2) \\
&\quad - \frac{1}{2}\sum_{i=1}^{m}\log(\lambda_i) - \frac{1}{2}\sum_{i=1}^{m}\left[\left\{\exp(\lambda_i) - 1 - \lambda_i - \frac{1}{2}\lambda_i^2\right\}\exp(2\mu_i + \lambda_i)\left\{\sum_{j=1}^{n}\exp(\beta_0 + \beta_1 X_{ij})\right\}^2\right] \\
&\quad + C,
\end{aligned}
$$

where C denotes a constant term independent of any model or variational parameters.

6.5.3 Optimisation Issues

In this section, we look into the problem of obtaining a local minimum for the Thouless-Anderson-Palmer approximate negative log-likelihood. We have chosen to focus on a simplified version of the model set-up in (6.8) to highlight the optimisation problems encountered.

6.5.3.1 A Simplified Version of the Optimisation Problem

First, consider the case where

$$\beta_1 = 0, \quad \sigma^2 = \sigma^2_{\text{fixed}}, \quad m = 1.$$

This situation corresponds to the following simplified Poisson mixed model:

$$Y_{1j}|U_1 \sim \text{Poisson}\left(e^{\beta_0+U_1}\right), \quad 1 \leq j \leq n, \quad U_1 \sim N(0, \sigma^2_{\text{fixed}}). \tag{6.9}$$

It follows that in this case, the Thouless-Anderson-Palmer approximate negative log-likelihood can be simplified to

$$\begin{aligned}-\underline{\ell}_{\text{TAP}}(\beta_0, \mu_1, \lambda_1) = &-(\beta_0 + \mu_1)\sum_{j=1}^{n} Y_{ij} + ne^{\beta_0+\mu_1+\frac{1}{2}\lambda_1} + \frac{\mu_1^2 + \lambda_1}{2\sigma^2_{\text{fixed}}} - \frac{1}{2}\log(\lambda_1) \\ &- \frac{1}{2}\left\{\exp(\lambda_1) - 1 - \lambda_1 - \frac{1}{2}\lambda_1^2\right\} n^2 \exp(2\mu_1 + \lambda_1 + 2\beta_0) + C,\end{aligned}$$

where C denotes a constant term independent of any model or variational parameters.

To facilitate the optimization process, we compute the partial derivatives required for minimising the Thouless-Anderson-Palmer approximate negative log-likelihood. The first order partial derivatives are as follows:

$$\begin{aligned}\frac{-\partial\underline{\ell}_{\text{TAP}}(\beta_0, \mu_1, \lambda_1)}{\partial\beta_0} = &-\sum_{j=1}^{n} Y_{ij} + n\exp\left(\beta_0 + \mu_1 + \frac{1}{2}\lambda_1\right) \\ &- n^2\left\{\exp(\lambda_1) - 1 - \lambda_1 - \frac{1}{2}\lambda_1^2\right\}\exp(2\mu_1 + \lambda_1 + 2\beta_0), \\ \frac{-\partial\underline{\ell}_{\text{TAP}}(\beta_0, \mu_1, \lambda_1)}{\partial\mu_1} = &-\sum_{j=1}^{n} Y_{ij} + n\exp\left(\beta_0 + \mu_1 + \frac{1}{2}\lambda_1\right) + \frac{\mu_1}{\sigma^2_{\text{fixed}}} \\ &- n^2\left\{\exp(\lambda_1) - 1 - \lambda_1 - \frac{1}{2}\lambda_1^2\right\}\exp(2\mu_1 + \lambda_1 + 2\beta_0) \quad \text{and} \\ \frac{-\partial\underline{\ell}_{\text{TAP}}(\beta_0, \mu_1, \lambda_1)}{\partial\lambda_1} = &\frac{n}{2}\exp\left(\beta_0 + \mu_1 + \frac{1}{2}\lambda_1\right) + \frac{1}{2\sigma^2_{\text{fixed}}} - \frac{1}{2\lambda_1} \\ &- \frac{n^2}{2}\left\{\exp(\lambda_1) - 1 - \lambda_1\right\}\exp(2\mu_1 + \lambda_1 + 2\beta_0) \\ &- \frac{n^2}{2}\left\{\exp(\lambda_1) - 1 - \lambda_1 - \frac{1}{2}\lambda_1^2\right\}\exp(2\mu_1 + \lambda_1 + 2\beta_0) \\ = &\frac{n}{2}\exp\left(\beta_0 + \mu_1 + \frac{1}{2}\lambda_1\right) + \frac{1}{2\sigma^2_{\text{fixed}}} - \frac{1}{2\lambda_1} \\ &- \frac{n^2}{2}\left\{2\exp(\lambda_1) - 2 - 2\lambda_1 - \frac{1}{2}\lambda_1^2\right\}\exp(2\mu_1 + \lambda_1 + 2\beta_0).\end{aligned}$$

Next, the second order partial derivatives in the diagonal of the Hessian matrix are as follows:

$$\begin{aligned}
\frac{-\partial \underline{\ell}_{\text{TAP}}(\beta_0, \mu_1, \lambda_1)}{\partial \beta_0^2} &= n \exp\left(\beta_0 + \mu_1 + \frac{1}{2}\lambda_1\right) \\
&\quad - 2n^2 \left\{\exp(\lambda_1) - 1 - \lambda_1 - \frac{1}{2}\lambda_1^2\right\} \exp(2\mu_1 + \lambda_1 + 2\beta_0), \\
\frac{-\partial \underline{\ell}_{\text{TAP}}(\beta_0, \mu_1, \lambda_1)}{\partial \mu_1^2} &= n \exp\left(\beta_0 + \mu_1 + \frac{1}{2}\lambda_1\right) + \frac{1}{\sigma^2_{\text{fixed}}} \\
&\quad - 2n^2 \left\{\exp(\lambda_1) - 1 - \lambda_1 - \frac{1}{2}\lambda_1^2\right\} \exp(2\mu_1 + \lambda_1 + 2\beta_0) \quad \text{and} \\
\frac{-\partial \underline{\ell}_{\text{TAP}}(\beta_0, \mu_1, \lambda_1)}{\partial \lambda_1^2} &= \frac{n}{4} \exp\left(\beta_0 + \mu_1 + \frac{1}{2}\lambda_1\right) + \frac{1}{2\lambda_1^2} \\
&\quad - \frac{n^2}{2} \left\{4\exp(\lambda_1) - 4 - 3\lambda_1 - \frac{1}{2}\lambda_1^2\right\} \exp(2\mu_1 + \lambda_1 + 2\beta_0).
\end{aligned}$$

Lastly, the second order partial derivatives in the off-diagonals of the Hessian matrix are as follows:

$$\begin{aligned}
\frac{-\partial \underline{\ell}_{\text{TAP}}(\beta_0, \mu_1, \lambda_1)}{\partial \beta_0 \partial \mu_1} &= n \exp\left(\beta_0 + \mu_1 + \frac{1}{2}\lambda_1\right) \\
&\quad - 2n^2 \left\{\exp(\lambda_1) - 1 - \lambda_1 - \frac{1}{2}\lambda_1^2\right\} \exp(2\mu_1 + \lambda_1 + 2\beta_0), \\
\frac{-\partial \underline{\ell}_{\text{TAP}}(\beta_0, \mu_1, \lambda_1)}{\partial \beta_0 \partial \lambda_1} &= \frac{n}{2} \exp\left(\beta_0 + \mu_1 + \frac{1}{2}\lambda_1\right) \\
&\quad - n^2 \left\{2\exp(\lambda_1) - 2 - 2\lambda_1 - \frac{1}{2}\lambda_1^2\right\} \exp(2\mu_1 + \lambda_1 + 2\beta_0) \quad \text{and} \\
\frac{-\partial \underline{\ell}_{\text{TAP}}(\beta_0, \mu_1, \lambda_1)}{\partial \mu_1 \partial \lambda_1} &= \frac{n}{2} \exp\left(\beta_0 + \mu_1 + \frac{1}{2}\lambda_1\right) \\
&\quad - n^2 \left\{2\exp(\lambda_1) - 2 - 2\lambda_1 - \frac{1}{2}\lambda_1^2\right\} \exp(2\mu_1 + \lambda_1 + 2\beta_0).
\end{aligned}$$

6.5.3.2 Simplified Simulation Study

A simplified simulation study with the model set-up in (6.9) was run with the following settings where

$$\beta_0^0 = -0.2, \quad \sigma_{\text{fixed}} = 0.3, \quad \text{and} \quad n = 20$$

with 10000 replications. The search for a local minimiser of $-\underline{\ell}_{\text{TAP}}(\beta_0, \mu_1, \lambda_1)$ in a three-dimensional space can be challenging. Hence, we use the following strategy involving the `optim()` function in `R`:

1. First, initialise values for $(\beta_0, \mu_1, \lambda_1)$.

2. Next, based on the initial values specified in (1), carry out a large number of Nelder-Mead iterations, via `optim()`, to search for a local minimiser of $-\underline{\ell}_{\text{TAP}}(\beta_0, \mu_1, \lambda_1)$.

3. Lastly, use the results from step 2 to obtain starting values for the Broyden-Fletcher-Goldberg- Shanno quasi-Newton approach, via `optim()`, to improve on the result from implementing step 2.

Whilst this strategy may seem reasonable, it turns out that it is prone to erratic behaviour if the initial value in step 1 is a poor choice. For the simplified model set-up in (6.9), the expressions for the partial derivatives can be analysed to develop stationary point equations and determine suitable starting values.

6.5.3.3 Results and Conclusion

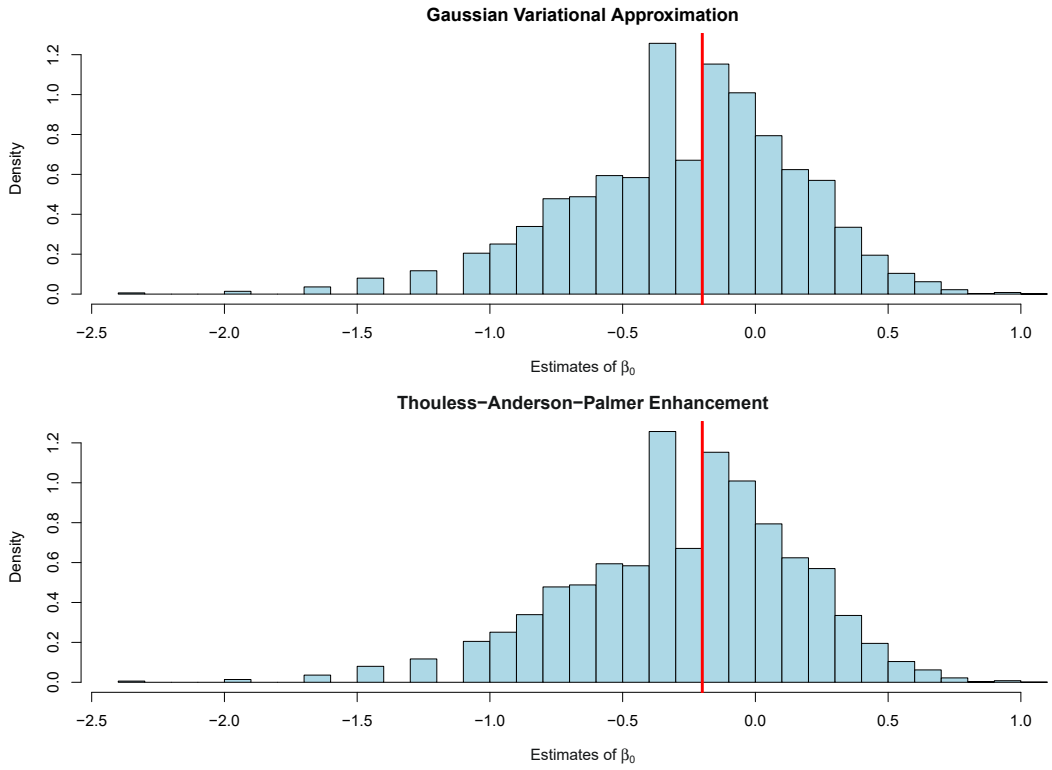

Figure 6.1: Histograms of the Gaussian variational approximation estimates of β_0 and the Thouless-Anderson-Palmer enhancement estimates of β_0.

After having found finite local minima of the Thouless-Anderson-Palmer approximate log-likelihood surface for this simulation study, the estimates of β_0 were gathered from implementing both the GVA and TAP approaches. Figure 6.1 displays the histograms of the estimates of β_0 obtained from the GVA approach and TAP enhancement approach. Note that the vertical red line is situated at the true value of $\beta_0^0 = -0.2$. We see that for both approximation approaches, the estimates are distributed about the true value.

We then computed the absolute error values for the estimates. For a generic estimate $\widehat{\beta}_0$, the absolute error is computed as

$$|\widehat{\beta}_0 - \beta_0^0|.$$

Applying this definition to the vectors of the estimates of β_0 obtained from both the GVA and TAP enhancement approaches, we obtain vectors of the GVA absolute errors and the TAP enhancement absolute errors of length 10000 each. Next, we obtained the pairwise differences with ordering as follows:

Pairwise difference = (TAP enhancement absolute error) − (GVA absolute error).

Figure 6.2 shows a histogram of the pairwise differences.

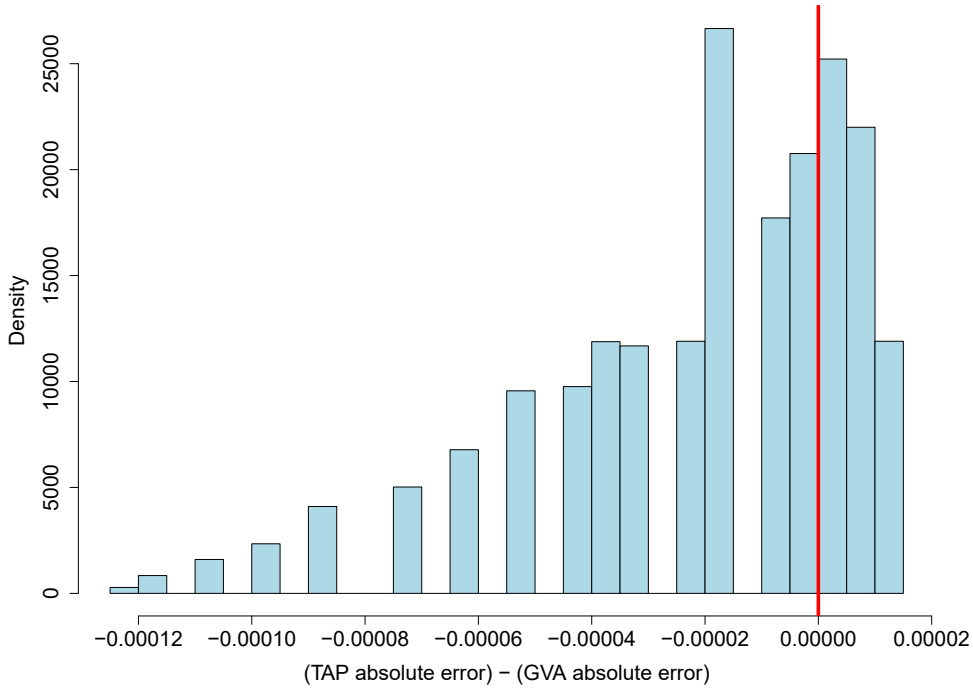

Figure 6.2: Histogram of the pairwise differences between the TAP enhancement absolute errors and GVA absolute errors.

Figure 6.2 suggests a statistically significant improvement due to using the TAP enhancement. However, the improvement seems to be quite slight from a practical standpoint. On the positive side, the results from this simulation study shows that with some care it is possible to get estimates of Poisson mixed model parameters that are improved by the TAP enhancement. On the negative side, the numerical problem is quite challenging even for the m = 1 and $\sigma^2 = \sigma^2_{\text{fixed}}$ case. In the more practically relevant case where m is in the hundreds and σ^2 is estimated instead of being fixed, obtaining local minima for the Thouless-Anderson-Palmer approximate negative log-likelihood could be quite challenging. We will investigate this further in the next subsection.

6.5.4 Simulation Study

A full simulation study was run to investigate and compare the accuracy of approximations from the GVA approach and the TAP enhancement approach. In this simulation study, a Poisson linear mixed model was used following the set-up in (6.8). The values for the true parameter vector $(\beta_0^0, \beta_1^0, (\sigma^2)^0)$ were chosen from the following possible set of pre-determined values

$$\{(-0.3, 0.2, 0.5), (2.2, -0.1, 0.16), (1.2, 0.4, 0.1), (0.02, 1.3, 1), (-0.3, 0.2, 0.1)\}.$$

and the distribution of the X_{ij} was taken to be either $N(0,1)$ or Uniform $(-1,1)$. The number of groups in the simulated data, m, varied over the set $\{100, 200, \ldots, 1000\}$ and the number of observations present within each group, n, was fixed at $m/10$. 100 replications were simulated for every possible combination of the true parameter vector, chosen X_{ij} distribution and value of (m, n) pair. For each sample, estimates for the true parameter vector $(\beta_0^0, \beta_1^0, (\sigma^2)^0)$, were obtained using similar steps to those outlined in Section 6.5.3.2 for the TAP enhancement approach. Estimates for $(\beta_0^0, \beta_1^0, (\sigma^2)^0)$ via the GVA approach were also obtained. For generic estimates $\widehat{\beta}_0$, $\widehat{\beta}_1$ and $\widehat{\sigma}^2$, their absolute error values were computed as

$$|\widehat{\beta}_0 - \beta_0^0|, \quad |\widehat{\beta}_1 - \beta_1^0| \quad \text{and} \quad |\widehat{\sigma}^2 - (\sigma^2)^0|$$

respectively. Applying this definition to the vectors of the estimates of β_0, β_1 and σ^2 obtained from both the GVA and TAP enhancement approaches, we obtain vectors of the GVA absolute errors and the TAP enhancement absolute errors of length 100 each. Like in the previous simulation study, we obtained the pairwise differences with

ordering as follows:

Pairwise difference = (TAP enhancement absolute error) − (GVA absolute error).

Grouped boxplots for each model parameter were then produced to graphically demonstrate the mean and spread for each vector of pairwise differences.

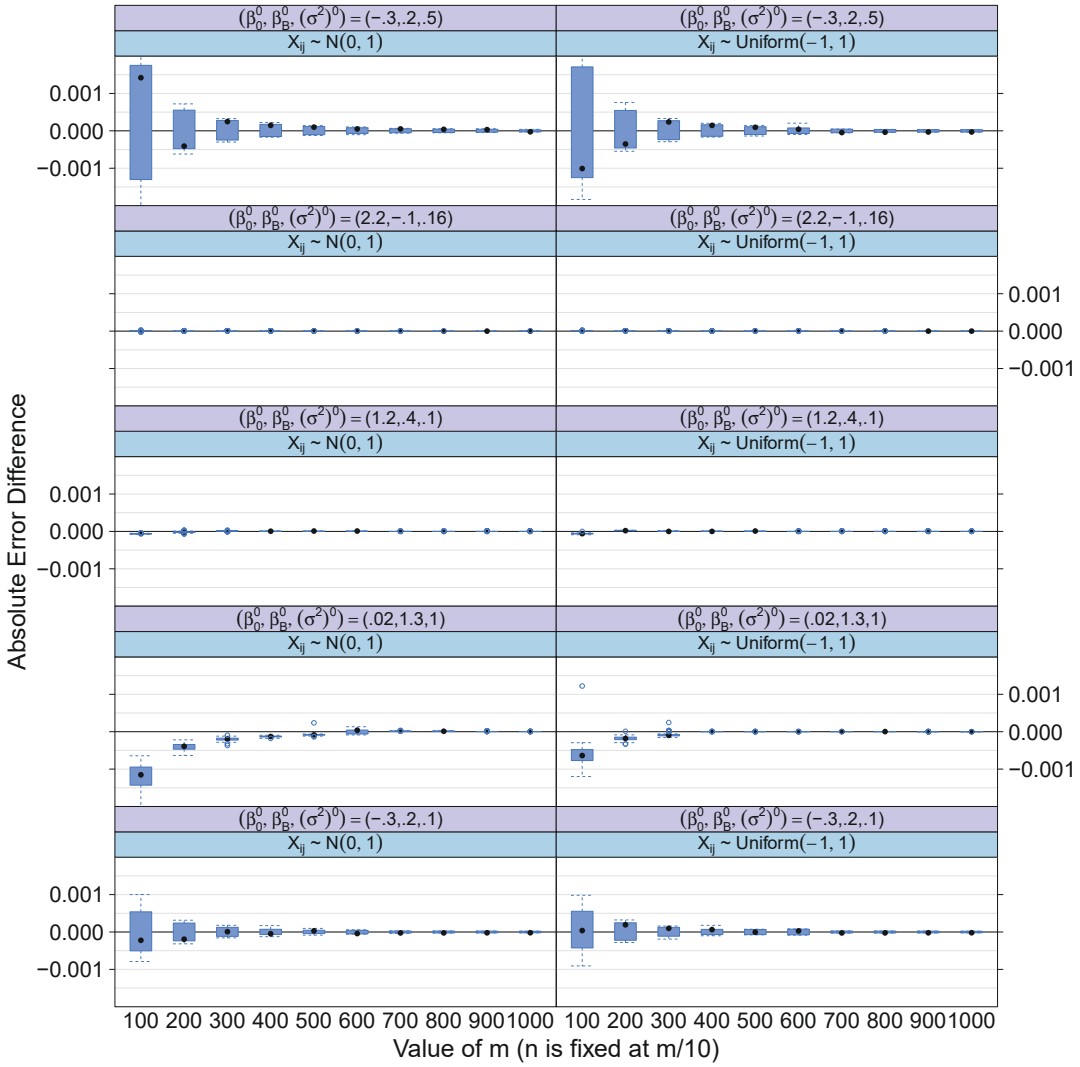

Figure 6.3: Grouped boxplots representing the pairwise difference in absolute errors (TAP enhancement absolute error - GVA absolute error) in estimating β_0. The values of m are 100, 200,..., 1000 while the value of n is fixed at $m/10$.

Figure 6.3 shows that for values of m in the lower hundreds, if there is a difference in accuracy in estimating β_0 between the GVA approach and TAP enhancement approach,

the TAP enhancement approach seems to give slightly better estimates in most of those cases. There seems to be a more significant improvement when the true value β_0^0 is close to 0 coupled with a large $(\sigma^2)^0$ value. However, as the value of m increases, the difference in absolute errors between the TAP enhancement approach and GVA approach diminishes.

Inspection of Figure 6.4 reveals almost no observable difference in the accuracy of estimates of β_1 between the GVA approach and TAP enhancement approach, for all values of m.

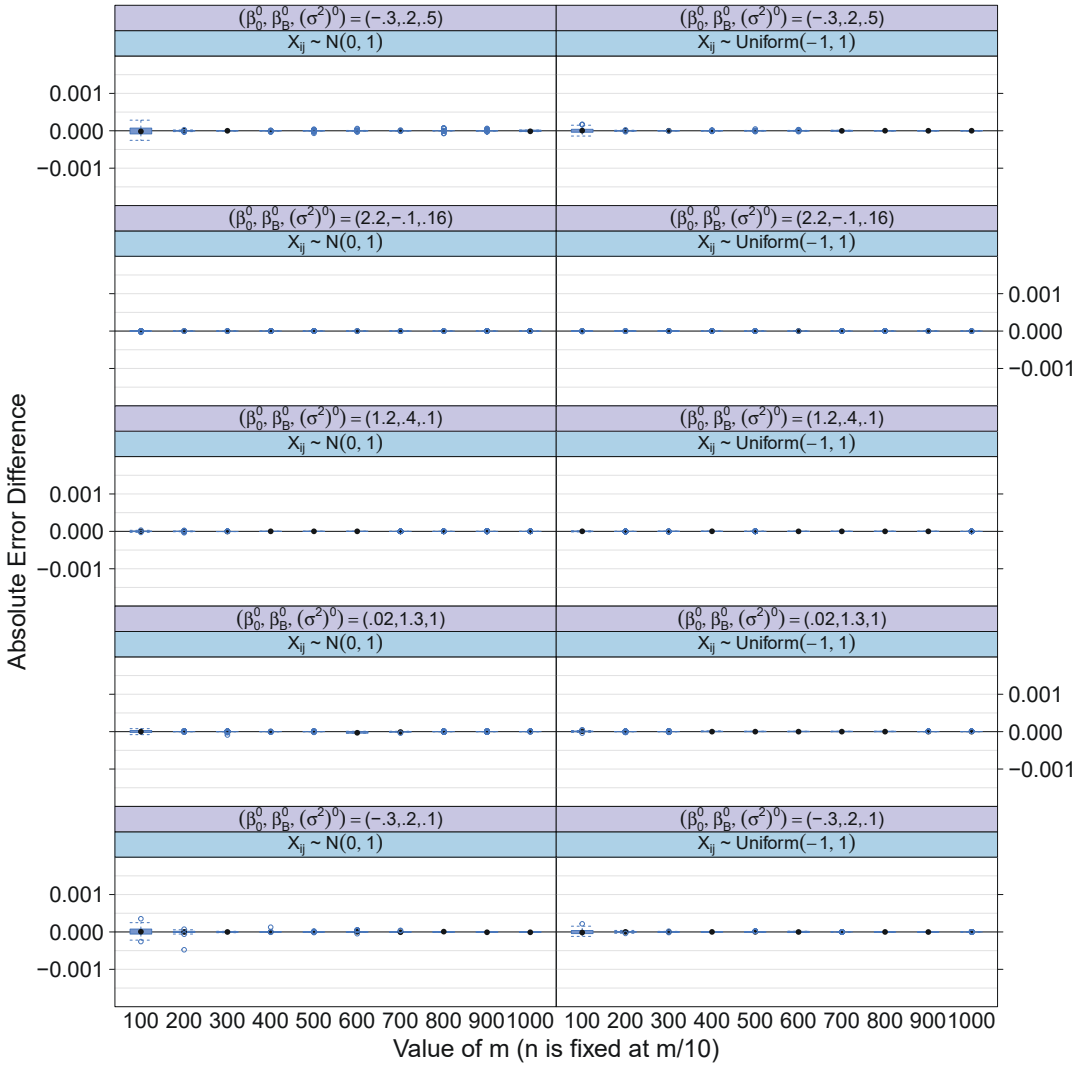

Figure 6.4: Grouped boxplots representing the pairwise difference in absolute errors (TAP enhancement absolute error - GVA absolute error) in estimating β_1. The values of m are 100, 200,..., 1000 while the value of n is fixed at $m/10$.

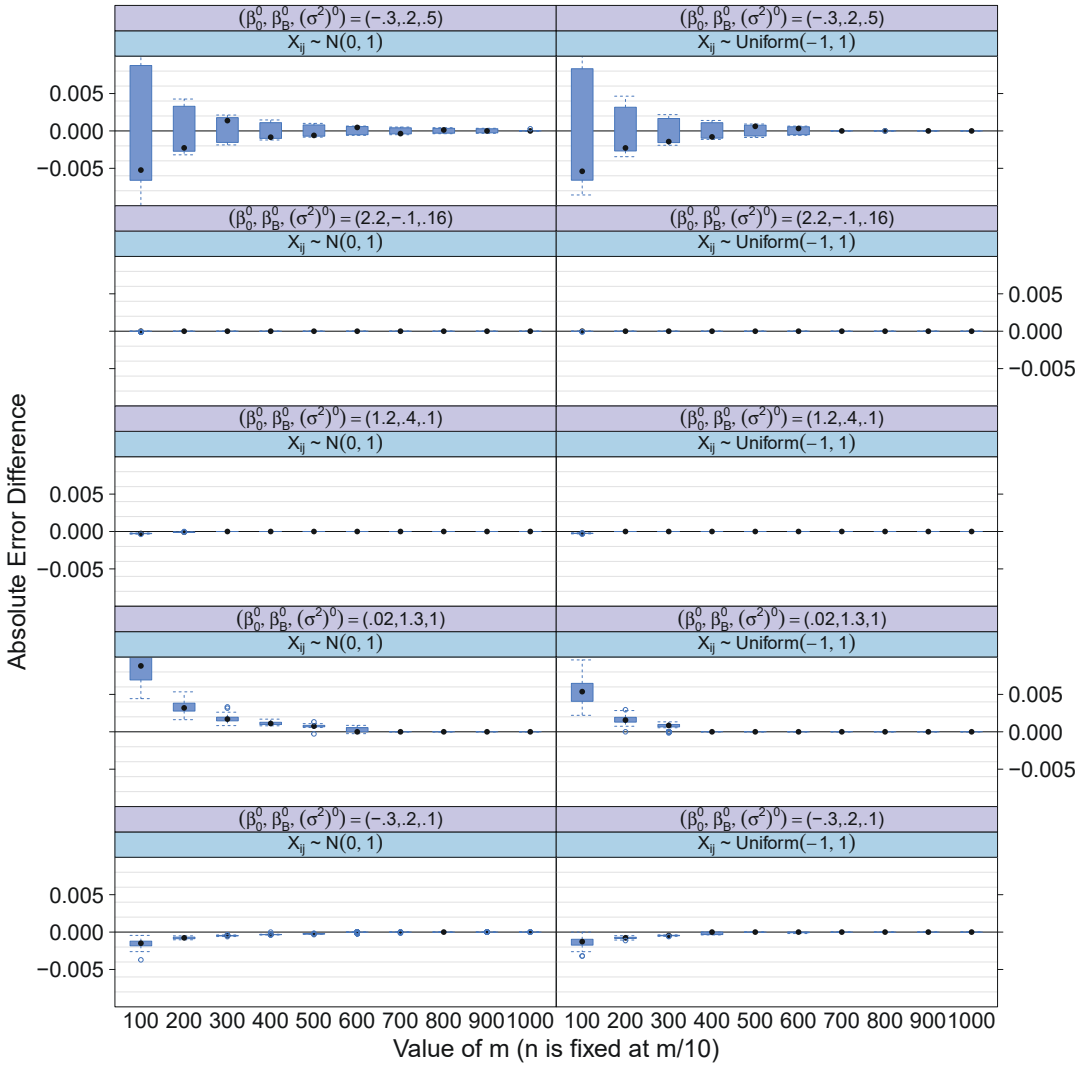

Figure 6.5: Grouped boxplots representing the pairwise difference in absolute errors (TAP enhancement absolute error - GVA absolute error) in estimating σ^2. The values of m are 100, 200,..., 1000 while the value of n is fixed at $m/10$.

In Figure 6.5, for values of m in the lower hundreds, there are 6 subplots where a visible difference in the accuracy in estimating σ^2 between the GVA approach and TAP enhancement approach is present. The TAP enhancement approach seems to give slightly better estimates in 4 out of 6 such cases. In contrast to Figure 6.3, the TAP estimates are worse when the true value β_0^0 is close to 0 coupled with a large $(\sigma^2)^0$ value. Once again, as the value of m increases, there is no significant difference in the absolute errors between the TAP enhancement approach and GVA approach.

To conclude, the Thouless-Anderson-Palmer enhancement approach suggests a slight yet statistically significant improvement as compared to using the Gaussian variational approximation approach for small datasets ($m \leq 200$). However, although the TAP estimates for β_0^0 fare better than the GVA estimates when the true value β_0^0 is close to 0 coupled with a large $(\sigma^2)^0$ value, that is not the case for the TAP estimates of $(\sigma^2)^0$. It is also worth exploring how the TAP approach performs against the GVA approach for generalized linear mixed models with response distributions other than the Poisson family.

6.6 Appendix

6.6.1 Proof of Result 2

This appendix contains the details for the derivations leading up to Result 2.

6.6.1.1 Main Quantity in Onsager's Correction Term

Note that the main quantity in the Onsager's correction term is

$$\begin{aligned}&\xi_i\left(\mu_i, \lambda_i; \boldsymbol{Y}_i, \boldsymbol{X}_i, \beta_0, \beta_1, \sigma^2\right)\\&= \text{Var}\{\Psi_i(\tilde{U}_i)|\boldsymbol{Y}_i, \boldsymbol{X}_i\} - \lambda_i\left[E\{\Psi_i'(\tilde{U}_i)|\boldsymbol{Y}_i, \boldsymbol{X}_i\}\right]^2 - \frac{\lambda_i^2}{2}\left[E\{\Psi_i''(\tilde{U}_i)|\boldsymbol{Y}_i, \boldsymbol{X}_i\}\right]^2.\end{aligned} \tag{6.10}$$

This quantity consists of three terms which we will work with in the next three subsubsections.

6.6.1.2 An Explicit Expression for the First Term in (6.10)

In this subsubsection, we find an explicit expression for $\text{Var}\{\Psi_i(\tilde{U}_i)|\boldsymbol{Y}_i, \boldsymbol{X}_i\}$. Firstly, note that

$$\begin{aligned}&\text{Var}\{\Psi_i(\tilde{U}_i)|\boldsymbol{Y}_i, \boldsymbol{X}_i\}\\&= \text{Var}\left(\frac{\tilde{U}_i^2}{2\sigma^2} - Y_{i\bullet}\tilde{U}_i + \mathcal{A}_i(\tilde{U}_i)\Big|\boldsymbol{Y}_i, \boldsymbol{X}_i\right)\\&= \text{Var}\left(\frac{\tilde{U}_i^2}{2\sigma^2} - Y_{i\bullet}\tilde{U}_i\Big|\boldsymbol{Y}_i, \boldsymbol{X}_i\right) + 2\text{Cov}\left(\frac{\tilde{U}_i^2}{2\sigma^2} - Y_{i\bullet}\tilde{U}_i, \mathcal{A}_i(\tilde{U}_i)\Big|\boldsymbol{Y}_i, \boldsymbol{X}_i\right) + \text{Var}\left(\mathcal{A}_i(\tilde{U}_i)\Big|\boldsymbol{Y}_i, \boldsymbol{X}_i\right).\end{aligned}$$

Now we treat each term in the expression for $\text{Var}\{\Psi_i(\tilde{U}_i)|\boldsymbol{Y}_i, \boldsymbol{X}_i\}$ individually.

Treatment of the $Var\left(\frac{\tilde{U}_i^2}{2\sigma^2} - Y_{i\bullet}\tilde{U}_i \middle| \boldsymbol{Y}_i, \boldsymbol{X}_i\right)$ *Term*

We first obtain an expression for

$$\mathrm{Var}(\mathfrak{a}\tilde{U}_i^2 + \mathfrak{b}\tilde{U}_i + \mathfrak{c}) = \mathrm{Var}(\mathfrak{a}\tilde{U}_i^2 + \mathfrak{b}\tilde{U}_i)$$

for general $\mathfrak{a}, \mathfrak{b}, \mathfrak{c} \in \mathbb{R}$. Then, note that

$$\begin{aligned}
\mathrm{Var}_{\tilde{U}_i}\left(\tilde{U}_i^2\right) &= E\left(\tilde{U}_i^4\right) - \left\{E\left(\tilde{U}_i^2\right)\right\}^2 \\
&= (\mu_i^4 + 6\mu_i^2\lambda_i + 3\lambda_i^2) - (\mu_i^2 + \lambda_i)^2 \\
&= 2\lambda_i^2 + 4\mu_i^2\lambda_i \\
&= 2\lambda_i(2\mu_i^2 + \lambda_i)
\end{aligned}$$

and

$$\begin{aligned}
\mathrm{Cov}_{\tilde{U}_i}(\tilde{U}_i, \tilde{U}_i^2) &= E\left(\tilde{U}_i^3\right) - E\left(\tilde{U}_i\right)E\left(\tilde{U}_i^2\right) \\
&= (\mu_i + 3\mu_i\lambda_i) - (\mu_i)(\mu_i^2 + \lambda_i) \\
&= 2\mu_i\lambda_i.
\end{aligned}$$

Then we have

$$\begin{aligned}
\mathrm{Var}(\mathfrak{a}\tilde{U}_i^2 + \mathfrak{b}\tilde{U}_i) &= \mathfrak{a}\mathrm{Var}(\tilde{U}_i^2) + 2\mathfrak{a}\mathfrak{b}\mathrm{Cov}(\tilde{U}_i^2, \tilde{U}_i) + \mathfrak{b}^2\mathrm{Var}(\tilde{U}_i) \\
&= 2\mathfrak{a}^2\lambda_i(2\mu_i^2 + \lambda_i) + 4\mathfrak{a}\mathfrak{b}\lambda_i\mu_i + \mathfrak{b}^2\lambda_i \\
&= 4\mathfrak{a}^2\lambda_i\mu_i^2 + 2\mathfrak{a}^2\lambda_i^2 + 4\mathfrak{a}\mathfrak{b}\lambda_i\mu_i + \mathfrak{b}^2\lambda_i.
\end{aligned}$$

Setting

$$\mathfrak{a} = \frac{1}{2\sigma^2} \quad \text{and} \quad b = -Y_{i\bullet},$$

we get

$$\mathrm{Var}\left(\frac{\tilde{U}_i^2}{2\sigma^2} - Y_{i\bullet}\tilde{U}_i \middle| \boldsymbol{Y}_i, \boldsymbol{X}_i\right) = \frac{2\lambda_i\mu_i^2 + \lambda_i^2}{2\sigma^4} - \frac{2Y_{i\bullet}\lambda_i\mu_i}{\sigma^2} + \lambda_i(Y_{i\bullet})^2. \tag{6.11}$$

Treatment of the $2\mathrm{Cov}\left(\frac{\tilde{U}_i^2}{2\sigma^2} - Y_{i\bullet}\tilde{U}_i, \mathcal{A}_i(\tilde{U}_i) \middle| \boldsymbol{Y}_i, \boldsymbol{X}_i\right)$ *Term*

Next, we obtain an expression for

$$\mathrm{Cov}(\mathfrak{a}\tilde{U}_i^2 + \mathfrak{b}\tilde{U}_i, \mathcal{A}_i(\tilde{U}_i)|\boldsymbol{Y}_i, \boldsymbol{X}_i) = \mathfrak{a}\mathrm{Cov}(\tilde{U}_i^2, \mathcal{A}_i(\tilde{U}_i)|\boldsymbol{Y}_i, \boldsymbol{X}_i) + \mathfrak{b}\mathrm{Cov}(\tilde{U}_i, \mathcal{A}_i(\tilde{U}_i)|\boldsymbol{Y}_i, \boldsymbol{X}_i)$$

for general $\mathfrak{a}, \mathfrak{b} \in \mathbb{R}$. Next note that, for $k \in \{1, 2\}$,

$$\begin{aligned}
&\mathrm{Cov}\left(\tilde{U}_i^k, \mathcal{A}_i(\tilde{U}_i)|\boldsymbol{Y}_i, \boldsymbol{X}_i\right) \\
&= \sum_{j=1}^{n} \mathrm{Cov}\left(\tilde{U}_i^k, b(\beta_0 + \beta_1 X_{ij} + \tilde{U}_i)|\boldsymbol{X}_i\right) \\
&= \sum_{j=1}^{n} E\left(\tilde{U}_i^k b(\beta_0 + \beta_1 X_{ij} + \tilde{U}_i)|\boldsymbol{X}_i\right) - E(\tilde{U}_i^k)\sum_{j=1}^{n} E\left(b(\beta_0 + \beta_1 X_{ij} + \tilde{U}_i)|\boldsymbol{X}_i\right).
\end{aligned}$$

Therefore,

$$\begin{aligned}
&\mathrm{Cov}\left(\tilde{U}_i, \mathcal{A}_i(\tilde{U}_i)|\boldsymbol{X}_i\right) \\
&= \sum_{j=1}^{n} E\left(\tilde{U}_i b(\beta_0 + \beta_1 X_{ij} + \tilde{U}_i)|\boldsymbol{X}_i\right) - \mu_i \sum_{j=1}^{n} E\left(b(\beta_0 + \beta_1 X_{ij} + \tilde{U}_i)|\boldsymbol{X}_i\right),
\end{aligned}$$

and

$$\begin{aligned}
&\mathrm{Cov}\left(\tilde{U}_i^2, \mathcal{A}_i(\tilde{U}_i)|\boldsymbol{X}_i\right) \\
&= \sum_{j=1}^{n} E\left(\tilde{U}_i^2 b(\beta_0 + \beta_1 X_{ij} + \tilde{U}_i)|\boldsymbol{X}_i\right) - (\lambda_i + \mu_i^2) \sum_{j=1}^{n} E\left(b(\beta_0 + \beta_1 X_{ij} + \tilde{U}_i)|\boldsymbol{X}_i\right).
\end{aligned}$$

It follows from these results that

$$\begin{aligned}
&\mathrm{Cov}(\mathfrak{a}\tilde{U}_i^2 + \mathfrak{b}\tilde{U}_i, \mathcal{A}_i(\tilde{U}_i)|\boldsymbol{Y}_i, \boldsymbol{X}_i) \\
&= 2\mathfrak{a} \sum_{j=1}^{n} E\left(\tilde{U}_i^2 b(\beta_0 + \beta_1 X_{ij} + \tilde{U}_i)|\boldsymbol{X}_i\right) + 2\mathfrak{b} \sum_{j=1}^{n} E\left(\tilde{U}_i b(\beta_0 + \beta_1 X_{ij} + \tilde{U}_i)|\boldsymbol{X}_i\right) \\
&\quad - 2\left\{\mathfrak{a}(\lambda_i + \mu_i^2) + \mathfrak{b}\mu_i\right\} \sum_{j=1}^{n} E\left(b(\beta_0 + \beta_1 X_{ij} + \tilde{U}_i)|\boldsymbol{X}_i\right).
\end{aligned}$$

Setting

$$\mathfrak{a} = \frac{1}{2\sigma^2} \quad \text{and} \quad b = -Y_{i\bullet},$$

we obtain

$$\begin{aligned}
&2\mathrm{Cov}\left(\frac{\tilde{U}_i^2}{2\sigma^2} - Y_{i\bullet}\tilde{U}_i, \mathcal{A}_i(\tilde{U}_i)\Big|\boldsymbol{Y}_i, \boldsymbol{X}_i\right) \\
&= \frac{1}{\sigma^2} \sum_{j=1}^{n} E\left(\tilde{U}_i^2 b(\beta_0 + \beta_1 X_{ij} + \tilde{U}_i)|\boldsymbol{X}_i\right) - 2Y_{i\bullet} \sum_{j=1}^{n} E\left(\tilde{U}_i b(\beta_0 + \beta_1 X_{ij} + \tilde{U}_i)|\boldsymbol{X}_i\right) \\
&\quad + \left(2Y_{i\bullet}\mu_i - \frac{\lambda_i + \mu_i^2}{\sigma^2}\right) \sum_{j=1}^{n} E\left(b(\beta_0 + \beta_1 X_{ij} + \tilde{U}_i)|\boldsymbol{X}_i\right).
\end{aligned} \tag{6.12}$$

Treatment of the Var $\left(\mathcal{A}_i(\tilde{U}_i)\middle|\boldsymbol{Y}_i, \boldsymbol{X}_i\right)$ *Term*

It follows that

$$
\begin{aligned}
&\text{Var}\left(\mathcal{A}_i(\tilde{U}_i)\middle|\boldsymbol{Y}_i, \boldsymbol{X}_i\right) \\
&= \text{Var}\left(\sum_{j=1}^{n} b(\beta_0 + \beta_1 X_{ij} + \tilde{U}_i)\middle|\boldsymbol{X}_i\right) \\
&= \sum_{j=1}^{n} \text{Var}\left(b(\beta_0 + \beta_1 X_{ij} + \tilde{U}_i)|\boldsymbol{X}_i\right) + \sum_{j\neq j'} \text{Cov}\left(b(\beta_0 + \beta_1 X_{ij} + \tilde{U}_i), b(\beta_0 + \beta_1 X_{ij'} + \tilde{U}_i)|\boldsymbol{X}_i\right) \\
&= \sum_{j=1}^{n} E\left(b(\beta_0 + \beta_1 X_{ij} + \tilde{U}_i^2\middle|\boldsymbol{X}_i\right) - \sum_{j=1}^{n}\left\{E\left(b(\beta_0 + \beta_1 X_{ij} + \tilde{U}_i)\middle|\boldsymbol{X}_i\right)\right\}^2 \\
&\quad + \sum_{j\neq j'} E\left(b(\beta_0 + \beta_1 X_{ij} + \tilde{U}_i) b(\beta_0 + \beta_1 X_{ij'} + \tilde{U}_i)|\boldsymbol{X}_i\right) \\
&\quad - \sum_{j\neq j'}\left\{E\left(b(\beta_0 + \beta_1 X_{ij} + \tilde{U}_i)|\boldsymbol{X}_i\right) E\left(b(\beta_0 + \beta_1 X_{ij'} + \tilde{U}_i)|\boldsymbol{X}_i\right)\right\}
\end{aligned}
$$

The previous expression then simplifies to

$$
\begin{aligned}
&\sum_{j=1}^{n} E\left(b(\beta_0 + \beta_1 X_{ij} + \tilde{U}_i)^2\middle|\boldsymbol{X}_i\right) + \sum_{j\neq j'} E\left(b(\beta_0 + \beta_1 X_{ij} + \tilde{U}_i) b(\beta_0 + \beta_1 X_{ij'} + \tilde{U}_i)|\boldsymbol{X}_i\right) \\
&\quad - \left\{\sum_{j=1}^{n} E\left(b(\beta_0 + \beta_1 X_{ij} + \tilde{U}_i)\middle|\boldsymbol{X}_i\right)\right\}^2 .
\end{aligned} \tag{6.13}
$$

The Resultant Var$\{\Psi_i(\tilde{U}_i)|\boldsymbol{Y}_i, \boldsymbol{X}_i\}$ *Expression*

Putting together the expressions from (6.11), (6.12) and (6.13), we have

$$
\begin{aligned}
&\text{Var}\{\Psi_i(\tilde{U}_i)|\boldsymbol{Y}_i, \boldsymbol{X}_i\} \\
&= \frac{2\lambda_i\mu_i^2 + \lambda_i^2}{2\sigma^4} - \frac{2Y_{i\bullet}\lambda_i\mu_i}{\sigma^2} + \lambda_i(Y_{i\bullet})^2 + \frac{1}{\sigma^2}\sum_{j=1}^{n} E\left(\tilde{U}_i^2 b(\beta_0 + \beta_1 X_{ij} + \tilde{U}_i)|\boldsymbol{X}_i\right) \\
&\quad - 2Y_{i\bullet}\sum_{j=1}^{n} E\left(\tilde{U}_i b(\beta_0 + \beta_1 X_{ij} + \tilde{U}_i)|\boldsymbol{X}_i\right) \\
&\quad + \left(2Y_{i\bullet}\mu_i - \frac{\lambda_i + \mu_i^2}{\sigma^2}\right)\sum_{j=1}^{n} E\left(b(\beta_0 + \beta_1 X_{ij} + \tilde{U}_i)|\boldsymbol{X}_i\right) \\
&\quad + \sum_{j=1}^{n} E\left(b(\beta_0 + \beta_1 X_{ij} + \tilde{U}_i)^2\middle|\boldsymbol{X}_i\right) + \sum_{j\neq j'} E\left(b(\beta_0 + \beta_1 X_{ij} + \tilde{U}_i) b(\beta_0 + \beta_1 X_{ij'} + \tilde{U}_i)|\boldsymbol{X}_i\right) \\
&\quad - \left\{\sum_{j=1}^{n} E\left(b(\beta_0 + \beta_1 X_{ij} + \tilde{U}_i)\middle|\boldsymbol{X}_i\right)\right\}^2 .
\end{aligned}
$$

However, note that,

$$\sum_{j=1}^{n} E\left(b(\beta_0+\beta_1 X_{ij}+\tilde{U}_i)^2\Big|\boldsymbol{X}_i\right)+\sum_{j\neq j'} E\left(b(\beta_0+\beta_1 X_{ij}+\tilde{U}_i)b(\beta_0+\beta_1 X_{ij'}+\tilde{U}_i)|\boldsymbol{X}_i\right)$$
$$=\sum_{j=1}^{n}\sum_{j'=1}^{n} E\left(b(\beta_0+\beta_1 X_{ij}+\tilde{U}_i)b(\beta_0+\beta_1 X_{ij'}+\tilde{U}_i)|\boldsymbol{X}_i\right).$$

Therefore,

$$\begin{aligned}
&\mathrm{Var}\{\Psi_i(\tilde{U}_i)|\boldsymbol{Y}_i,\boldsymbol{X}_i\}\\
&=\frac{2\lambda_i\mu_i^2+\lambda_i^2}{2\sigma^4}-\frac{2Y_{i\bullet}\lambda_i\mu_i}{\sigma^2}+\lambda_i(Y_{i\bullet})^2+\frac{1}{\sigma^2}\sum_{j=1}^{n}E\left(\tilde{U}_i^2 b(\beta_0+\beta_1 X_{ij}+\tilde{U}_i)|\boldsymbol{X}_i\right)\\
&\quad-2Y_{i\bullet}\sum_{j=1}^{n}E\left(\tilde{U}_i b(\beta_0+\beta_1 X_{ij}+\tilde{U}_i)|\boldsymbol{X}_i\right)+\left(2Y_{i\bullet}\mu_i-\frac{\lambda_i+\mu_i^2}{\sigma^2}\right)\sum_{j=1}^{n}E\left(b(\beta_0+\beta_1 X_{ij}+\tilde{U}_i)|\boldsymbol{X}_i\right)\\
&\quad+\sum_{j=1}^{n}\sum_{j'=1}^{n}E\left(b(\beta_0+\beta_1 X_{ij}+\tilde{U}_i)b(\beta_0+\beta_1 X_{ij'}+\tilde{U}_i)|\boldsymbol{X}_i\right)-\left\{\sum_{j=1}^{n}E\left(b(\beta_0+\beta_1 X_{ij}+\tilde{U}_i)\Big|\boldsymbol{X}_i\right)\right\}^2.
\end{aligned}$$

6.6.1.3 An Explicit Expression for the Second Term in (6.10)

In this subsubsection, we find an explicit expression for $-\lambda_i\left[E\{\Psi_i'(\tilde{U}_i)|\boldsymbol{Y}_i,\boldsymbol{X}_i\}\right]^2$. First note that

$$\Psi_i'(\tilde{U}_i)=-Y_{i\bullet}+\mathcal{A}_i'(\tilde{U}_i)+\frac{\tilde{U}_i}{\sigma^2}.$$

Hence,

$$E\{\Psi_i'(\tilde{U}_i)|\boldsymbol{Y}_i,\boldsymbol{X}_i\}=-Y_{i\bullet}+\sum_{j=1}^{n}E\left(b'(\beta_0+\beta_1 X_{ij}+\tilde{U}_i)|\boldsymbol{X}_i\right)+\frac{\mu_i}{\sigma^2}.$$

Then we have,

$$\begin{aligned}
-\lambda_i\left[E\{\Psi_i'(\tilde{U}_i)|\boldsymbol{Y}_i,\boldsymbol{X}_i\}\right]^2&=\lambda_i\left\{-Y_{i\bullet}+\sum_{j=1}^{n}E\left(b'(\beta_0+\beta_1 X_{ij}+\tilde{U}_i)|\boldsymbol{X}_i\right)+\frac{\mu_i}{\sigma^2}\right\}^2\\
&=-\lambda_i(Y_{i\bullet})^2-\lambda_i\left\{\sum_{j=1}^{n}E\left(b'(\beta_0+\beta_1 X_{ij}+\tilde{U}_i)|\boldsymbol{X}_i\right)\right\}^2\\
&\quad-\frac{\lambda_i\mu_i^2}{\sigma^4}+2\lambda_i Y_{i\bullet}\left\{\sum_{j=1}^{n}E\left(b'(\beta_0+\beta_1 X_{ij}+\tilde{U}_i)|\boldsymbol{X}_i\right)\right\}\\
&\quad+\frac{2Y_{i\bullet}\lambda_i\mu_i}{\sigma^2}-\frac{2\mu_i\lambda_i}{\sigma^2}\left\{\sum_{j=1}^{n}E\left(b'(\beta_0+\beta_1 X_{ij}+\tilde{U}_i)|\boldsymbol{X}_i\right)\right\}.
\end{aligned}$$

6.6.1.4 An Explicit Expression for the Third Term in (6.10)

In this subsubsection, we find an explicit expression for $-\frac{\lambda_i^2}{2}\left[E\{\Psi_i''(\tilde{U}_i)|\boldsymbol{Y}_i,\boldsymbol{X}_i\}\right]^2$. Note that

$$\Psi_i''(\tilde{U}_i) = \mathcal{A}_i''(\tilde{U}_i) + \frac{1}{\sigma^2}.$$

Therefore,

$$E\{\Psi_i''(\tilde{U}_i)|\boldsymbol{Y}_i,\boldsymbol{X}_i\} = \sum_{j=1}^{n} E\left(b''(\beta_0+\beta_1 X_{ij}+\tilde{U}_i)|\boldsymbol{X}_i\right) + \frac{1}{\sigma^2}.$$

Following that, we have,

$$\begin{aligned}
-\frac{\lambda_i^2}{2}\left[E\{\Psi_i''(\tilde{U}_i)|\boldsymbol{Y}_i,\boldsymbol{X}_i\}\right]^2 &= -\frac{\lambda_i^2}{2}\left(\sum_{j=1}^{n} E\left(b''(\beta_0+\beta_1 X_{ij}+\tilde{U}_i)|\boldsymbol{X}_i\right) + \frac{1}{\sigma^2}\right)^2 \\
&= -\frac{\lambda_i^2}{2}\left(\sum_{j=1}^{n} E\left(b''(\beta_0+\beta_1 X_{ij}+\tilde{U}_i)|\boldsymbol{X}_i\right)\right)^2 \\
&\quad - \frac{\lambda_i^2}{\sigma^2}\sum_{j=1}^{n} E\left(b''(\beta_0+\beta_1 X_{ij}+\tilde{U}_i)|\boldsymbol{X}_i\right) - \frac{\lambda_i^2}{2\sigma^4}.
\end{aligned}$$

6.6.1.5 The Resultant Expression for the Main Quantity in the Onsager's Correction Term

Based on the results from the previous three subsections, we have the following

$$\begin{aligned}
&\xi_i\left(\mu_i,\lambda_i;\boldsymbol{Y}_i,\boldsymbol{X}_i,\beta_0,\beta_1,\sigma^2\right) \\
&= \frac{2\lambda_i\mu_i^2+\lambda_i^2}{2\sigma^4} - \frac{2Y_{i\bullet}\lambda_i\mu_i}{\sigma^2} + \lambda_i(Y_{i\bullet})^2 + \frac{1}{\sigma^2}\sum_{j=1}^{n} E\left(\tilde{U}_i^2 b(\beta_0+\beta_1 X_{ij}+\tilde{U}_i)|\boldsymbol{X}_i\right) \\
&\quad - 2Y_{i\bullet}\sum_{j=1}^{n} E\left(\tilde{U}_i b(\beta_0+\beta_1 X_{ij}+\tilde{U}_i)|\boldsymbol{X}_i\right) + \left(2Y_{i\bullet}\mu_i - \frac{\lambda_i+\mu_i^2}{\sigma^2}\right)\sum_{j=1}^{n} E\left(b(\beta_0+\beta_1 X_{ij}+\tilde{U}_i)|\boldsymbol{X}_i\right) \\
&\quad + \sum_{j=1}^{n}\sum_{j'=1}^{n} E\left(b(\beta_0+\beta_1 X_{ij}+\tilde{U}_i)b(\beta_0+\beta_1 X_{ij'}+\tilde{U}_i)|\boldsymbol{X}_i\right) - \left\{\sum_{j=1}^{n} E\left(b(\beta_0+\beta_1 X_{ij}+\tilde{U}_i)\Big|\boldsymbol{X}_i\right)\right\}^2 - \lambda_i(Y_{i\bullet})^2 \\
&\quad - \lambda_i\left\{\sum_{j=1}^{n} E\left(b'(\beta_0+\beta_1 X_{ij}+\tilde{U}_i)|\boldsymbol{X}_i\right)\right\}^2 - \frac{\lambda_i\mu_i^2}{\sigma^4} + 2\lambda_i Y_{i\bullet}\left\{\sum_{j=1}^{n} E\left(b'(\beta_0+\beta_1 X_{ij}+\tilde{U}_i)|\boldsymbol{X}_i\right)\right\} + \frac{2Y_{i\bullet}\lambda_i\mu_i}{\sigma^2} \\
&\quad - \frac{2\mu_i\lambda_i}{\sigma^2}\left\{\sum_{j=1}^{n} E\left(b'(\beta_0+\beta_1 X_{ij}+\tilde{U}_i)|\boldsymbol{X}_i\right)\right\} - \frac{\lambda_i^2}{2\sigma^4} - \frac{\lambda_i^2}{2}\left(\sum_{j=1}^{n} E\left(b''(\beta_0+\beta_1 X_{ij}+\tilde{U}_i)|\boldsymbol{X}_i\right)\right)^2 \\
&\quad - \frac{\lambda_i^2}{\sigma^2}\sum_{j=1}^{n} E\left(b''(\beta_0+\beta_1 X_{ij}+\tilde{U}_i)|\boldsymbol{X}_i\right).
\end{aligned}$$

By simplifying this expression further, we have,

$$
\begin{aligned}
&\xi_i\left(\mu_i, \lambda_i; \boldsymbol{Y}_i, \boldsymbol{X}_i, \beta_0, \beta_1, \sigma^2\right) \\
&= \frac{1}{\sigma^2} \sum_{j=1}^{n} E\left(\tilde{U}_i^2 b(\beta_0 + \beta_1 X_{ij} + \tilde{U}_i) | \boldsymbol{X}_i\right) - 2Y_{i\bullet} \sum_{j=1}^{n} E\left(\tilde{U}_i b(\beta_0 + \beta_1 X_{ij} + \tilde{U}_i) | \boldsymbol{X}_i\right) \\
&\quad + \left(2Y_{i\bullet}\mu_i - \frac{\lambda_i + \mu_i^2}{\sigma^2}\right) \sum_{j=1}^{n} E\left(b(\beta_0 + \beta_1 X_{ij} + \tilde{U}_i) | \boldsymbol{X}_i\right) \\
&\quad + \sum_{j=1}^{n} \sum_{j'=1}^{n} E\left(b(\beta_0 + \beta_1 X_{ij} + \tilde{U}_i) b(\beta_0 + \beta_1 X_{ij'} + \tilde{U}_i) | \boldsymbol{X}_i\right) \\
&\quad - \left\{\sum_{j=1}^{n} E\left(b(\beta_0 + \beta_1 X_{ij} + \tilde{U}_i) \Big| \boldsymbol{X}_i\right)\right\}^2 - \lambda_i \left\{\sum_{j=1}^{n} E\left(b'(\beta_0 + \beta_1 X_{ij} + \tilde{U}_i) | \boldsymbol{X}_i\right)\right\}^2 \\
&\quad + 2\lambda_i \left(Y_{i\bullet} - \frac{\mu_i}{\sigma^2}\right) \left\{\sum_{j=1}^{n} E\left(b'(\beta_0 + \beta_1 X_{ij} + \tilde{U}_i) | \boldsymbol{X}_i\right)\right\} \\
&\quad - \frac{\lambda_i^2}{2} \left(\sum_{j=1}^{n} E\left(b''(\beta_0 + \beta_1 X_{ij} + \tilde{U}_i) | \boldsymbol{X}_i\right)\right)^2 - \frac{\lambda_i^2}{\sigma^2} \sum_{j=1}^{n} E\left(b''(\beta_0 + \beta_1 X_{ij} + \tilde{U}_i) | \boldsymbol{X}_i\right).
\end{aligned}
$$

For any $n \times 1$ random vector $\boldsymbol{X}_i = (X_{i1}, \ldots, X_{in})$ and $p, q \in \mathbb{Z}^+$, define

$$
\mathcal{B}(p, q, \beta_0, \beta_1, \mu_i, \lambda_i, \boldsymbol{X}_i) \equiv \sum_{j=1}^{n} E\left(\tilde{U}_i^p b^{(q)}(\beta_0 + \beta_1 X_{ij} + \tilde{U}_i) | \boldsymbol{X}_i\right),
$$

for $1 \leq i \leq m$. Also define

$$
\mathcal{C}(\beta_0, \beta_1, \mu_i, \lambda_i, \boldsymbol{X}_i) \equiv \sum_{j=1}^{n} \sum_{j'=1}^{n} E\left(b(\beta_0 + \beta_1 X_{ij} + \tilde{U}_i) b(\beta_0 + \beta_1 X_{ij'} + \tilde{U}_i) | \boldsymbol{X}_i\right),
$$

for $1 \leq i \leq m$. We then have

$$
\begin{aligned}
&\xi_i\left(\mu_i, \lambda_i; \boldsymbol{Y}_i, \boldsymbol{X}_i, \beta_0, \beta_1, \sigma^2\right) \\
&= \left(2Y_{i\bullet}\mu_i - \frac{\lambda_i + \mu_i^2}{\sigma^2}\right) \mathcal{B}(0, 0, \beta_0, \beta_1, \mu_i, \lambda_i, \boldsymbol{X}_i) \\
&\quad + 2\lambda_i \left(Y_{i\bullet} - \frac{\mu_i}{\sigma^2}\right) \mathcal{B}(0, 1, \beta_0, \beta_1, \mu_i, \lambda_i, \boldsymbol{X}_i) - 2Y_{i\bullet} \mathcal{B}(1, 0, \beta_0, \beta_1, \mu_i, \lambda_i, \boldsymbol{X}_i) \\
&\quad - \frac{\lambda_i^2}{\sigma^2} \mathcal{B}(0, 2, \beta_0, \beta_1, \mu_i, \lambda_i, \boldsymbol{X}_i) + \frac{1}{\sigma^2} \mathcal{B}(2, 0, \beta_0, \beta_1, \mu_i, \lambda_i, \boldsymbol{X}_i) \\
&\quad - \mathcal{B}(0, 0, \beta_0, \beta_1, \mu_i, \lambda_i, \boldsymbol{X}_i)^2 - \lambda_i \mathcal{B}(0, 1, \beta_0, \beta_1, \mu_i, \lambda_i, \boldsymbol{X}_i)^2 \\
&\quad - \frac{\lambda_i^2}{2} \mathcal{B}(0, 2, \beta_0, \beta_1, \mu_i, \lambda_i, \boldsymbol{X}_i)^2 + \mathcal{C}(\beta_0, \beta_1, \mu_i, \lambda_i, \boldsymbol{X}_i).
\end{aligned} \tag{6.14}
$$

6.6.2 Expressing the Main Quantity in the Onsager's Correction Term Using Integral Families

The aim of this final section in this appendix is to obtain an expression for the quantity $\xi_i\left(\mu_i, \lambda_i; \boldsymbol{Y}_i, \boldsymbol{X}_i, \beta_0, \beta_1, \sigma^2\right)$ that does not involve any expectation operators. Rather, we use specially tailored families of integrals for the desired expressions.

We first define the integral family definitions that we will work with. For $p, q \in \{0, 1, 2\}, r > 0$ and $s, t \in \mathbb{R}$, define

$$\mathcal{Q}(p, q, r, s, t) \equiv (2\pi)^{-1/2} \int_{-\infty}^{\infty} (s + rx)^p b^{(q)}(t + rx) \exp\left(-\frac{x^2}{2}\right) dx. \tag{6.15}$$

Also, for $r > 0$ and $s, t \in \mathbb{R}$, define

$$\mathcal{R}(r, s, t) \equiv (2\pi)^{-1/2} \int_{-\infty}^{\infty} b(s + rx) b(t + rx) \exp\left(-\frac{x^2}{2}\right) dx. \tag{6.16}$$

In the Poisson special case, since $b(x) = \exp(x)$, $\mathcal{Q}(p, q, r, s, t)$ and $\mathcal{R}(r, s, t)$ admit exact expressions. However, for general b functions, we are stuck with expressions as in (6.15) and (6.16). Thus, numerical integration is required for evaluation purposes. Then note that in (6.14),

$$\begin{aligned}
&\mathcal{B}(p, q, \beta_0, \beta_1, \mu_i, \lambda_i, \boldsymbol{X}_i) \\
&= \sum_{j=1}^{n} E\left(\tilde{U}_i^p b^{(q)}(\beta_0 + \beta_1 X_{ij} + \tilde{U}_i) | \boldsymbol{X}_i\right) \\
&= \sum_{j=1}^{n} \int_{-\infty}^{\infty} u^p b^{(q)}(\beta_0 + \beta_1 X_{ij} + u)(2\pi\lambda_i)^{-1/2} \exp\left\{-\frac{(u - \mu_i)^2}{2\lambda_i}\right\} du \\
&= \sum_{j=1}^{n} (2\pi)^{-1/2} \int_{-\infty}^{\infty} (\mu_i + \sqrt{\lambda_i} z)^p b^{(q)}(\beta_0 + \beta_1 X_{ij} + \mu_i + \sqrt{\lambda_i} z) \exp\left(-\frac{z^2}{2}\right) dz \\
&= \sum_{j=1}^{n} \mathcal{Q}(p, q, \sqrt{\lambda_i}, \mu_i, \beta_0 + \beta_1 X_{ij} + \mu_i).
\end{aligned}$$

Also note that

$$\begin{aligned}
&\mathcal{C}(\beta_0, \beta_1, \mu_i, \lambda_i, \boldsymbol{X}_i) \\
&= \sum_{j=1}^{n} E\left(b(\beta_0 + \beta_1 X_{ij} + \tilde{U}_i) b(\beta_0 + \beta_1 X_{ij'} + \tilde{U}_i) | \boldsymbol{X}_i\right) \\
&= \sum_{j=1}^{n} \sum_{j'=1}^{n} \int_{-\infty}^{\infty} b(\beta_0 + \beta_1 X_{ij} + u) b(\beta_0 + \beta_1 X_{ij'} + u)(2\pi\lambda_i)^{-1/2} \exp\left\{-\frac{(u - \mu_i)^2}{2\lambda_i}\right\} du \\
&= \sum_{j=1}^{n} \sum_{j'=1}^{n} (2\pi)^{-1/2} \int_{-\infty}^{\infty} b(\beta_0 + \beta_1 X_{ij} + \mu_i + \sqrt{\lambda_i} z) b(\beta_0 + \beta_1 X_{ij'} + \mu_i + \sqrt{\lambda_i} z) \\
&\qquad \times \exp\left(-\frac{z^2}{2}\right) dz \\
&= \sum_{j=1}^{n} \sum_{j'=1}^{n} \mathcal{R}(\sqrt{\lambda_i}, \beta_0 + \beta_1 X_{ij} + \mu_i, \beta_0 + \beta_1 X_{ij'} + \mu_i).
\end{aligned}$$

Hence, the final expression for $\xi_i\left(\mu_i, \lambda_i; \boldsymbol{Y}_i, \boldsymbol{X}_i, \beta_0, \beta_1, \sigma^2\right)$ used in Result 2 is as follows

$$
\begin{aligned}
&\xi_i\left(\mu_i, \lambda_i; \boldsymbol{Y}_i, \boldsymbol{X}_i, \beta_0, \beta_1, \sigma^2\right) \\
&= \sum_{j=1}^{n} \Bigg\{ \left(2Y_{i\bullet}\mu_i - \frac{\lambda_i + \mu_i^2}{\sigma^2}\right) \mathcal{Q}(0, 0, \sqrt{\lambda_i}, \mu_i, \beta_0 + \beta_1 X_{ij} + \mu_i) \\
&\quad + 2\lambda_i \left(Y_{i\bullet} - \frac{\mu_i}{\sigma^2}\right) \mathcal{Q}(0, 1, \sqrt{\lambda_i}, \mu_i, \beta_0 + \beta_1 X_{ij} + \mu_i) \\
&\quad - 2Y_{i\bullet}\mathcal{Q}(1, 0, \sqrt{\lambda_i}, \mu_i, \beta_0 + \beta_1 X_{ij} + \mu_i) \\
&\quad - \frac{\lambda_i^2}{\sigma^2}\mathcal{Q}(0, 2, \sqrt{\lambda_i}, \mu_i, \beta_0 + \beta_1 X_{ij} + \mu_i) + \frac{1}{\sigma^2}\mathcal{Q}(2, 0, \sqrt{\lambda_i}, \mu_i, \beta_0 + \beta_1 X_{ij} + \mu_i) \\
&\quad - \mathcal{Q}(0, 0, \sqrt{\lambda_i}, \mu_i, \beta_0 + \beta_1 X_{ij} + \mu_i)^2 - \lambda_i \mathcal{Q}(0, 1, \sqrt{\lambda_i}, \mu_i, \beta_0 + \beta_1 X_{ij} + \mu_i)^2 \\
&\quad - \frac{\lambda_i^2}{2}\mathcal{Q}(0, 2, \sqrt{\lambda_i}, \mu_i, \beta_0 + \beta_1 X_{ij} + \mu_i)^2 \\
&\quad + \sum_{j'=1}^{n} \mathcal{R}(\sqrt{\lambda_i}, \beta_0 + \beta_1 X_{ij} + \mu_i, \beta_0 + \beta_1 X_{ij'} + \mu_i) \Bigg\}.
\end{aligned}
$$

6.6.3 Proof of Result 3

This appendix contains the details leading to Result 3. In the Poisson case, $b(x) = \exp(x)$. Hence for $p, q \in \{0, 1, 2\}$, each term in (6.14) can be simplified greatly, leading to a reduced expression for $\xi_i\left(\mu_i, \lambda_i; \boldsymbol{Y}_i, \boldsymbol{X}_i, \beta_0, \beta_1, \sigma^2\right)$.

6.6.3.1 Simplifications in Poisson Case

When $p = 0, q \in \{0, 1, 2\}$,

$$
\begin{aligned}
\mathcal{B}(0, q, \beta_0, \beta_1, \mu_i, \lambda_i, \boldsymbol{X}_i) &= \sum_{j=1}^{n} E\left(b^{(q)}(\beta_0 + \beta_1 X_{ij} + \tilde{U}_i)|\boldsymbol{X}_i\right) \\
&= \sum_{j=1}^{n} E\left\{\exp(\beta_0 + \beta_1 X_{ij} + \tilde{U}_i)|\boldsymbol{X}_i\right\} \\
&= E\left\{\exp(\tilde{U}_i)\right\} \sum_{j=1}^{n} \exp(\beta_0 + \beta_1 X_{ij}) \\
&= \exp(\mu_i + \frac{1}{2}\lambda_i) \sum_{j=1}^{n} \exp(\beta_0 + \beta_1 X_{ij}).
\end{aligned}
\tag{6.17}
$$

When $p = 1, q = 0$,

$$\begin{aligned}
\mathcal{B}(1, 0, \beta_0, \beta_1, \mu_i, \lambda_i, \boldsymbol{X}_i) &= \sum_{j=1}^{n} E\left(\tilde{U}_i b(\beta_0 + \beta_1 X_{ij} + \tilde{U}_i) | \boldsymbol{X}_i\right) \\
&= \sum_{j=1}^{n} E\left\{\tilde{U}_i \exp(\beta_0 + \beta_1 X_{ij} + \tilde{U}_i) | \boldsymbol{X}_i\right\} \\
&= E\left\{\tilde{U}_i \exp(\tilde{U}_i)\right\} \sum_{j=1}^{n} \exp(\beta_0 + \beta_1 X_{ij}) \\
&= (\mu_i + \lambda_i) \exp(\mu_i + \frac{1}{2}\lambda_i) \sum_{j=1}^{n} \exp(\beta_0 + \beta_1 X_{ij}).
\end{aligned} \tag{6.18}$$

When $p = 2, q = 0$,

$$\begin{aligned}
\mathcal{B}(2, 0, \beta_0, \beta_1, \mu_i, \lambda_i, \boldsymbol{X}_i) &= \sum_{j=1}^{n} E\left(\tilde{U}_i^2 b(\beta_0 + \beta_1 X_{ij} + \tilde{U}_i) | \boldsymbol{X}_i\right) \\
&= \sum_{j=1}^{n} E\left\{\tilde{U}_i^2 \exp(\beta_0 + \beta_1 X_{ij} + \tilde{U}_i) | \boldsymbol{X}_i\right\} \\
&= E\left\{\tilde{U}_i^2 \exp(\tilde{U}_i)\right\} \sum_{j=1}^{n} \exp(\beta_0 + \beta_1 X_{ij}) \\
&= \left\{\lambda_i + (\mu_i + \lambda_i)^2\right\} \exp(\mu_i + \frac{1}{2}\lambda_i) \sum_{j=1}^{n} \exp(\beta_0 + \beta_1 X_{ij}).
\end{aligned} \tag{6.19}$$

Substituting (6.17), (6.18) and (6.19) into (6.14), several terms cancel with each other and we have

$$\begin{aligned}
&\xi_i\left(\mu_i, \lambda_i; \boldsymbol{Y}_i, \boldsymbol{X}_i, \beta_0, \beta_1, \sigma^2\right) \\
&= \left\{\exp(\lambda_i) - 1 - \lambda_i - \frac{1}{2}\lambda_i^2\right\} \exp(2\mu_i + \lambda_i) \left\{\sum_{j=1}^{n} \exp(\beta_0 + \beta_1 X_{ij})\right\}^2,
\end{aligned} \tag{6.20}$$

leading to Result 3.

Chapter 7

Extensions to Noncanonical Link Generalized Linear Mixed Models

In some situations, for a better data fit, using a noncanonical link might be beneficial over using a canonical link. To cater for these situations, we present an extension of the asymptotic normality results derived in Chapter 4 for generalized linear mixed models with noncanonical links. Thus, this chapter presents the final asymptotic normality theorem in this thesis, that concerns the joint asymptotic normality of all of the maximum quasi-likelihood estimators for a generalized linear mixed model with noncanonical links. As in Chapter 4, it elegantly shows faster rates of convergence for fixed effects that are not accompanied by a random effect compared to fixed effects that have a partnering random effect.

Finally, to wrap up the thesis, we build on the theory presented in Chapter 6 and present some details concerning the Thouless-Palmer-Anderson enhancement approach for improving statistical inference for generalized linear mixed models when noncanonical links are involved.

This chapter starts off by providing asymptotic normality results concerning generalized linear mixed models with noncanonical links in Section 7.1. Section 7.1.1 presents the model being used while Section 7.1.2 presents the notation required for the asymptotic normality theorem presented in Section 7.1.3. The chapter concludes with Section 7.2 which presents an introduction into the usage of the Thouless-Anderson-Palmer enhancement approach when noncanonical links are involved. Section 7.2.1 provides the model description for this section. Section 7.2.2 then provides an explicit expression for the GVA log-likelihood. Finally, Section 7.2.3 provide some introductory details

regarding the TAP enhancement approach for noncanonical links.

7.1 Asymptotic Normality Results Involving Noncanonical Links

In this section, we present asymptotic normality results for maximum quasi-likelihood estimators for a generalized linear mixed model with noncanonical links.

7.1.1 Model Description

To accommodate the use of noncanonical links, consider the following density, or probability mass, function for the class of one-parameter exponential families as in Fan et al. (1995) where

$$p(y;\eta) = \exp\left[y(g\circ b')^{-1}(\eta) - \{b\circ(g\circ b')^{-1}\}(\eta) + c(y)\right] h(y) \tag{7.1}$$

where g is the link function and η is the natural parameter. Here, $I(\mathcal{P}) = 1$ if the condition $\mathcal{P}$ is true and $I(\mathcal{P}) = 0$ if $\mathcal{P}$ is false. If the random variable Y has density, or probability mass, function as in (7.1), then $E(Y) = g^{-1}(\eta)$ and $\mathrm{Var}(Y) = \{b''\circ(b')^{-1}\circ g^{-1}\}(\eta)$. To account for overdispersion in the data and to allow one to model the variance flexibly, a common modelling extension is implemented such that $\mathrm{Var}(Y) = \phi\{b''\circ(b')^{-1}\circ g^{-1}\}(\eta)$, where $\phi > 0$ represents the dispersion parameter. This involves replacement of $\log\{p(y:\eta)\}$ by the following quasi-likelihood function

$$\left[y(g\circ b')^{-1}(\eta) - \{b\circ(g\circ b')^{-1}\}(\eta) + c(y)\right]/\phi + d(y,\phi) \tag{7.2}$$

where $d(y,\phi)$ is a function of y and ϕ only. Note that for ordinary Binomial and Poisson response models, ϕ is fixed at 1. For Gaussian and Gamma response models, (7.2) corresponds to the expression of $\log\{p(y;\eta)\}$ for a two-parameter exponential family density function and ordinary likelihood applies. In this section, we study generalized linear mixed models of the following form, for observations of the random pairs $(\boldsymbol{X}_{ij}, Y_{ij})$, $1 \le i \le m, 1 \le j \le n_i$,

$$\begin{aligned} &Y_{ij}|\boldsymbol{X}_{ij}, \boldsymbol{U}_i \text{ are independent having quasi-likelihood function (7.2) with} \\ &\text{natural parameter } \beta_0 + \boldsymbol{\beta}_1^T\boldsymbol{X}_{ij} + U_i \text{ such that the } U_i \text{ are independent} \\ &N(0,\sigma^2) \text{ random vectors.} \end{aligned} \tag{7.3}$$

The U_i are unobserved random effects variables. We assumed that the $\boldsymbol{X}_{ij}$ and U_i, for $1 \leq i \leq m$ and $1 \leq j \leq n_i$, are totally independent, with the $\boldsymbol{X}_{ij}$ each having the same distribution as the $d \times 1$ random vector $\boldsymbol{X}$ and the U_i each having the same distribution as the random variable U.

Then, for any $\beta_0 \in \mathbb{R}$, $\boldsymbol{\beta}_1 \in \mathbb{R}^d$ and $\sigma^2 > 0$, the conditional log-quasi-likelihood is the joint density function of the Y_{ij}, given the $\boldsymbol{X}_{ij}$, as a function of the parameters $(\beta_0^0, \boldsymbol{\beta}_1^0, (\sigma^2)^0)$ is,

$$(\widehat{\beta}_0, \widehat{\boldsymbol{\beta}}_1, \widehat{\sigma}^2) = \underset{\beta_0, \boldsymbol{\beta}_1, \sigma^2}{\operatorname{argmax}} \, \ell(\beta_0, \boldsymbol{\beta}_1, \sigma^2)$$

where $\ell(\beta_0, \boldsymbol{\beta}_1, \sigma^2)$ is the conditional log-likelihood and has the expression

$$\begin{aligned} \ell(\beta_0, \boldsymbol{\beta}_1, \sigma^2) = & -\frac{m}{2}\log(2\pi\sigma^2) + \sum_{i=1}^{m}\sum_{j=1}^{n}\left(c(Y_{ij})/\phi + d(Y_{ij}, \phi)\right) \\ & + \sum_{i=1}^{m} \log \int_{-\infty}^{\infty} \exp\left(\sum_{j=1}^{n} \Big[Y_{ij}(g \circ b')^{-1}(\beta_0 + \boldsymbol{\beta}_1^T \boldsymbol{X}_{ij} + u) \right. \\ & \left. - \{b \circ (g \circ b')^{-1}\}(\beta_0 + \boldsymbol{\beta}_1^T \boldsymbol{X}_{ij} + u) \Big]/\phi - \frac{u^2}{2\sigma^2} \right) du. \end{aligned} \tag{7.4}$$

7.1.2 Notation

Define

$$n \equiv \frac{1}{m}\sum_{i=1}^{m} n_i = \text{ average of the within-group sample sizes.}$$

Also define

$$\Omega^*_{\boldsymbol{\beta}_1}(U) \equiv E\left\{ w(\beta_0 + \boldsymbol{\beta}_1 + U) \begin{bmatrix} 1 & \boldsymbol{X}^T \\ \boldsymbol{X} & \boldsymbol{X}\boldsymbol{X}^T \end{bmatrix} \middle| U \right\},$$

where

$$w \equiv \frac{\left\{(g^{-1})'\right\}^2}{\{b'' \circ (b')^{-1} \circ g^{-1}\}}$$

and let

$$\Lambda^*_{\boldsymbol{\beta}_1} \equiv \left(E\left[\left\{ \text{lower right } d \times d \text{ block of } \Omega^*_{\boldsymbol{\beta}_1}(U)^{-1} \right\}^{-1} \right] \right)^{-1}.$$

7.1.3 Asymptotic Normality Theorem

The main theoretical contribution of this chapter is an asymptotic normality theorem for the maximum quasi-likelihood estimators for a generalized linear mixed model with noncanonical links as described in Section 7.1.1.

The theorem relies on the following assumptions:

(A8) The number of groups m diverges to ∞.

(A9) The within-group sample sizes n_i diverge to ∞ in such a way that $n_i/n \to C_i$ for constants $0 < C_i < \infty$, $1 \leq i \leq m$. Also, $n/m \to 0$ as m and n diverge.

(A10) The distribution of $\boldsymbol{X}$ is such that

$$E\left[\frac{E\left[\max\{1, \|\boldsymbol{X}\|\}^8 \max\{1, w(\beta_0 + \boldsymbol{\beta}_1^T \boldsymbol{X} + U)\}^4 | U\right]}{\min\{1, \lambda_{\min}(E\{w(\beta_0 + \boldsymbol{\beta}_1^T \boldsymbol{X} + U)|U\})\}^2}\right] < \infty$$

for all $\beta_0 \in \mathbb{R}$, $\boldsymbol{\beta}_1 \in \mathbb{R}^d$ and $\sigma^2 > 0$.

Theorem 14. *Assume that conditions (A8) - (A10) hold. Then we have the following*

$$\sqrt{m}\begin{bmatrix} \widehat{\beta}_0 - \beta_0^0 \\ \sqrt{n}\left(\widehat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1^0\right) \\ \widehat{\sigma^2} - (\sigma^2)^0 \end{bmatrix} \xrightarrow{\mathcal{D}} N\left(\begin{bmatrix} 0 \\ \mathbf{0} \\ 0 \end{bmatrix}, \begin{bmatrix} (\sigma^2)^0 & \mathbf{0} & 0 \\ \mathbf{0} & \phi \boldsymbol{\Lambda}^*_{\boldsymbol{\beta}_1} & \mathbf{0} \\ 0 & \mathbf{0} & 2\left\{(\sigma^2)^0\right\}^2 \end{bmatrix}\right).$$

The proof of Theorem 14 is in the appendix. A remark concerning Theorem 14 is as follows:

1. When working with binary responses, there are two common alternative noncanonical link functions used to the canonical logit link, namely the probit link and the complementary log-log link. Both noncanonical links map the mean response restricted to the $(0, 1)$ interval to the $(-\infty, \infty)$ interval. Also in both cases, $w(x)$ in Section 7.1.2 can be expressed as an explicit function to work with. Note that $b(x) = \log\{1 + \exp(x)\}$ for binary responses. Then the usage of a probit link where $g(x) = \Phi^{-1}(x)$ leads to $w(x)$ taking on the following expression

$$w = \frac{\varphi^2}{\Phi(1 - \Phi)},$$

where φ represents the standard normal density function.

Meanwhile, the usage of a complementary log-log link where

$$g(x) = \ln\{-\ln(1-x)\}$$

results in

$$w(x) = \frac{\exp(2x)}{\exp\{\exp(x)\} - 1}.$$

7.2 Thouless-Anderson-Palmer Approach Involving Noncanonical Links

In this section, we present a primer for those who wish to carry out the Thouless-Anderson-Palmer approach for generalized linear mixed models with noncanonical links.

7.2.1 Model Description

Now, consider the use of a simple noncanonical link generalized linear mixed model as follows

$$\begin{aligned} &Y_{ij}|X_{ij}, U_i \text{ are independent having density function (7.1) with} \\ &\text{natural parameter } \beta_0^0 + \beta_1^0 X_{ij} + U_i \text{ such that the } U_i \text{ are independent} \\ &N\left(0, \left(\sigma^2\right)^0\right) \text{ random variables.} \end{aligned} \tag{7.5}$$

Here, the values of (X_{ij}, Y_{ij}) are observed for $1 \leq i \leq m$ and $1 \leq j \leq n$. We have assumed that the X_{ij} and U_i are independent random variables. In addition, the X_{ij} are each assumed as having the same distribution as the random vector X. The U_i are the unobserved random effects variables and are assumed to be having the same distribution as the random vector U.

Let $\boldsymbol{\beta} = (\beta_0, \beta_1)$ be the vector of fixed parameters. Then the model parameters for this set-up are $(\boldsymbol{\beta}, \sigma^2)$. Following that $\ell(\boldsymbol{\beta}, \sigma^2)$, the conditional log-likelihood of $(\boldsymbol{\beta}, \sigma^2)$,

is

$$\ell(\boldsymbol{\beta}, \sigma^2) = -\frac{m}{2}\log(2\pi\sigma^2) + \sum_{i=1}^{m}\sum_{j=1}^{n} c(Y_{ij}) + \sum_{i=1}^{m}\log\int_{-\infty}^{\infty}\exp\Bigg(\sum_{j=1}^{n}\Big[Y_{ij}(g\circ b')^{-1}(\beta_0+\beta_1 X_{ij}+u) - \{b\circ(g\circ b')^{-1}\}(\beta_0+\beta_1 X_{ij}+u)\Big] - \frac{u^2}{2\sigma^2}\Bigg)du. \tag{7.6}$$

7.2.2 The Gaussian Variational Approximate Log-Likelihood

Following steps similar to those in Section 6.2, the Gaussian variational approximation to $\ell(\boldsymbol{\beta}, \sigma^2)$ is derived as,

$$\begin{aligned}&\underline{\ell}_{\mathrm{GVA}}(\boldsymbol{\beta}, \sigma^2, \boldsymbol{\mu}, \boldsymbol{\lambda})\\ &= \sum_{i=1}^{m} E_{\tilde{U}_i}\Bigg[\sum_{j=1}^{n}\Big\{Y_{ij}(g\circ b')^{-1}(\beta_0+\beta_1 X_{ij}+\tilde{U}_i) - \{b\circ(g\circ b')^{-1}\}(\beta_0+\beta_1 X_{ij}+\tilde{U}_i)\\ &\quad + c(Y_{ij})\Big\} - \frac{\tilde{U}_i^2}{2\sigma^2} - \frac{1}{2}\log(2\pi\sigma^2)\Bigg] + \frac{1}{2}\sum_{i=1}^{m}\{1+\log(2\pi\lambda_i)\},\end{aligned} \tag{7.7}$$

where $E_{\tilde{U}_i}$ denotes the expectation with respect to the random variable $\tilde{U}_i \sim N(\mu_i, \lambda_i)$ with $\lambda_i > 0$, for $1 \le i \le m$. The variational parameters are $(\boldsymbol{\mu}, \boldsymbol{\lambda})$ where $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_m)$ and $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_m)$. Note that

$$\ell(\boldsymbol{\beta}, \sigma^2) \ge \underline{\ell}_{\mathrm{GVA}}(\boldsymbol{\beta}, \sigma^2, \boldsymbol{\mu}, \boldsymbol{\lambda})$$

for all vectors $\boldsymbol{\mu}$ and $\boldsymbol{\lambda}$.

7.2.3 Overview of Thouless-Anderson-Palmer Enhancement

Next, we provide details on how the TAP enhancement approach builds upon the GVA approach when using noncanonical links. Firstly, for each $1 \le i \le m$, define the following data vectors:

$$\boldsymbol{Y}_i \equiv (Y_{i1}, \ldots Y_{in}) \text{ and } \boldsymbol{X}_i \equiv (X_{i1}, \ldots X_{in}).$$

The Gaussian variational approximate negative log-likelihood can then be expressed as follows

$$-\underline{\ell}_{\text{GVA}}(\boldsymbol{\beta}, \sigma^2, \boldsymbol{\mu}, \boldsymbol{\lambda}) = -\frac{1}{2}\sum_{i=1}^{m}\{1 + \log(2\pi\lambda_i)\} + \sum_{i=1}^{m} E_{\tilde{U}_i}\left\{\Psi_i(\tilde{U}_i)|\boldsymbol{Y}_i, \boldsymbol{X}_i\right\},$$

where

$$\Psi_i(\tilde{U}_i) = \sum_{j=1}^{n}\left[-Y_{ij}(g \circ b')^{-1}(\beta_0 + \beta_1 X_{ij} + \tilde{U}_i) + \{b \circ (g \circ b')^{-1}\}(\beta_0 + \beta_1 X_{ij} + \tilde{U}_i) - c(Y_{ij})\right] + \frac{\tilde{U}_i^2}{2\sigma^2} + \frac{1}{2}\log(2\pi\sigma^2).$$

Using theory detailed in Section 6.3, the TAP approximate negative log-likelihood can be obtained as,

$$\begin{aligned}
-\underline{\ell}_{\text{TAP}}(\boldsymbol{\beta}, \sigma^2, \boldsymbol{\mu}, \boldsymbol{\lambda}) &= -\underline{\ell}_{\text{GVA}}(\boldsymbol{\beta}, \sigma^2, \boldsymbol{\mu}, \boldsymbol{\lambda}) - \frac{1}{2}\sum_{i=1}^{m}\xi_i\left(\mu_i, \lambda_i; \boldsymbol{X}_i, \boldsymbol{Y}_i, \beta_0, \beta_1, \sigma^2\right) \\
&= -\frac{1}{2}\sum_{i=1}^{m}\{1 + \log(2\pi\lambda_i)\} + \sum_{i=1}^{m} E\left\{\Psi_i(\tilde{U}_i)|\boldsymbol{Y}_i, \boldsymbol{X}_i\right\} \\
&\quad - \frac{1}{2}\sum_{i=1}^{m}\left(\text{Var}\{\Psi_i(\tilde{U}_i)|\boldsymbol{Y}_i, \boldsymbol{X}_i\} - \lambda_i\left[E\{\Psi_i'(\tilde{U}_i)|\boldsymbol{Y}_i, \boldsymbol{X}_i\}\right]^2\right. \\
&\quad \left. - \frac{\lambda_i^2}{2}\left[E\{\Psi_i''(\tilde{U}_i)|\boldsymbol{Y}_i, \boldsymbol{X}_i\}\right]^2\right).
\end{aligned} \tag{7.8}$$

Now define

$$\mathcal{E}_i(\tilde{U}_i) \equiv \sum_{j=1}^{n} Y_{ij}(g \circ b')^{-1}(\beta_0 + \beta_1 X_{ij} + \tilde{U}_i) \quad \text{and}$$

$$\mathcal{F}_i(\tilde{U}_i) \equiv \sum_{j=1}^{n}\{b \circ (g \circ b')^{-1}\}(\beta_0 + \beta_1 X_{ij} + \tilde{U}_i).$$

Then, we can re-express $\Psi_i(\tilde{U}_i)$ as

$$\Psi_i(\tilde{U}_i) = -\mathcal{E}_i(\tilde{U}_i) + \mathcal{F}_i(\tilde{U}_i) - \sum_{j=1}^{n} c(Y_{ij}) + \frac{\tilde{U}_i^2}{2\sigma^2} + \frac{1}{2}\log(2\pi\sigma^2).$$

It follows that

$$\Psi_i'(\tilde{U}_i) = -\mathcal{E}_i'(\tilde{U}_i) + \mathcal{F}_i'(\tilde{U}_i) + \frac{\tilde{U}_i}{\sigma^2}$$

and

$$\Psi_i''(\tilde{U}_i) = -\mathcal{E}_i''(\tilde{U}_i) + \mathcal{F}_i''(\tilde{U}_i) + \frac{1}{\sigma^2}.$$

By solving for and the substituting explicit expressions for $\Psi_i(\tilde{U}_i)$, $\Psi_i'(\tilde{U}_i)$ and $\Psi_i''(\tilde{U}_i)$

into (7.8), the TAP approximate negative log-likelihood can be obtained. This expression can then be optimized to find a local minima and obtain TAP estimates of the model parameters.

7.3 Appendix

This appendix contained the details for the derivations leading up to Theorem 14.

7.3.1 Constructing the Fisher Information Matrix

In order to compute the asymptotic covariance matrix for the maximum quasi-likelihood estimators, we would first need to compute the Fisher information matrix for the model parameters as per the model description in 7.1.1. To do so, let

$$\boldsymbol{S}_i \equiv \begin{bmatrix} \boldsymbol{S}_{0i} \\ \boldsymbol{S}_{1i} \\ \boldsymbol{S}_{2i} \end{bmatrix} = \begin{bmatrix} \frac{\partial}{\partial \beta_0} \log p_{\boldsymbol{Y}_i|\boldsymbol{X}_i}(\boldsymbol{Y}_i|\boldsymbol{X}_i) \\ \nabla_{\boldsymbol{\beta}_1} \log p_{\boldsymbol{Y}_i|\boldsymbol{X}_i}(\boldsymbol{Y}_i|\boldsymbol{X}_i) \\ \frac{\partial}{\partial \sigma^2} \log p_{\boldsymbol{Y}_i|\boldsymbol{X}_i}(\boldsymbol{Y}_i|\boldsymbol{X}_i) \end{bmatrix} \tag{7.9}$$

denote the ith contribution to the scores for each of the model parameters. Then the Fisher information matrix can be computed as

$$I(\beta_0, \boldsymbol{\beta}_1, \sigma^2) = \sum_{i=1}^{m} E(\boldsymbol{S}_i \boldsymbol{S}_i^T | \boldsymbol{X}_i).$$

The next few sections then focus on obtaining the expressions for the scores and the quadratic conditional expectations that are required to construct the final Fisher information matrix.

7.3.2 Expression for Conditional Density Function

The expression for $p_{\boldsymbol{Y}_i|\boldsymbol{X}_i}(\boldsymbol{Y}_i|\boldsymbol{X}_i)$ as per the model description in (7.3) is

$$\begin{aligned}
&p_{\boldsymbol{Y}_i|\boldsymbol{X}_i}(\boldsymbol{Y}_i|\boldsymbol{X}_i) \\
&= \int_{-\infty}^{\infty} \prod_{j=1}^{n_i} \left\{ p(Y_{ij}|\boldsymbol{X}_{ij}, U_i) \right\} p(U_i) dU_i \\
&= \int_{-\infty}^{\infty} \exp \left\{ \sum_{j=1}^{n} \Big(\Big[Y_{ij} (g \circ b')^{-1} \left(\beta_0 + \boldsymbol{\beta}_1^T \boldsymbol{X}_{ij} + u \right) - \left\{ b \circ (g \circ b')^{-1} \right\} \left(\beta_0 + \boldsymbol{\beta}_1^T \boldsymbol{X}_{ij} + u \right) \right. \\
&\quad \left. + c(Y_{ij}) \Big] / \phi + d(Y_{ij}, \phi) \Big) \right\} \times (2\pi\sigma^2)^{-1/2} \exp \left(-\frac{u^2}{2\sigma^2} \right) du
\end{aligned}$$

The previous expression can be further simplified as follows

$$\begin{aligned}
&\int_{-\infty}^{\infty} (2\pi\sigma^2)^{-1/2} \exp\left\{ \sum_{j=1}^{n} \Big(\Big[Y_{ij}(g \circ b')^{-1} \left(\beta_0 + \boldsymbol{\beta}_1^T \boldsymbol{X}_{ij} + u\right) \right. \right. \\
&\quad \left. \left. - \{b \circ (g \circ b')^{-1}\} \left(\beta_0 + \boldsymbol{\beta}_1^T \boldsymbol{X}_{ij} + u\right) + c(Y_{ij}) \Big] / \phi + d(Y_{ij}, \phi) \Big) - \frac{u^2}{2\sigma^2} \right\} du \\
&= (2\pi\sigma^2)^{-1/2} \exp\left\{ \sum_{j=1}^{n} (c(Y_{ij})/\phi + + d(Y_{ij}, \phi)) \right\} \\
&\quad \times \int_{-\infty}^{\infty} \exp\left[\sum_{j=1}^{n} \Big\{ Y_{ij}(g \circ b')^{-1} \left(\beta_0 + \boldsymbol{\beta}_1^T \boldsymbol{X}_{ij} + u\right) - \{b \circ (g \circ b')^{-1}\} \left(\beta_0 + \boldsymbol{\beta}_1^T \boldsymbol{X}_{ij} + u\right) \Big\} / \phi \right. \\
&\qquad \left. - \frac{u^2}{2\sigma^2} \right] du.
\end{aligned}$$

7.3.3 Deriving Expressions for the Expectation and Variance of the Response Variable

Next, using the Bartlett identities, we can obtain expressions for $E(Y)$ and $\text{Var}(Y)$. Firstly let,

$$a_1(\eta) \equiv (g \circ b')^{-1}(\eta) \quad \text{and} \quad a_2(\eta) \equiv \{b \circ (g \circ b')^{-1}\}(\eta).$$

Then note that the noncanonical extension of the one-parameter exponential family of density, or probability mass, functions takes on the following form

$$\begin{aligned}
p(y; \eta, \phi) &\propto \exp\left[\frac{y(g \circ b')^{-1}(\eta) - \{b \circ (g \circ b')^{-1}\}(\eta) + c(y)}{\phi} \right] \\
&= \exp\left\{ \frac{a_1(\eta) y - a_2(\eta) + c(y)}{\phi} \right\}
\end{aligned}$$

where η is the natural parameter. Then, we have the following

$$\log p(y; \eta, \phi) = \frac{a_1(\eta) y - a_2(\eta) + c(y)}{\phi} + C$$

where C is a constant independent of η. Let $\ell(\eta) \equiv \log p(y; \eta, \phi)$. Then,

$$\frac{\partial \ell(\eta)}{\partial \eta} = \frac{a_1'(\eta) y - a_2'(\eta)}{\phi}. \tag{7.10}$$

The first Bartlett identity states that

$$E\left(\frac{\partial \ell(\eta)}{\partial \eta} \right) = 0.$$

Hence, by substituting (7.10) into the first Bartlett identity, we have,

$$E\left\{ \frac{a_1'(\eta) y - a_2'(\eta)}{\phi} \right\} = 0.$$

This leads to

$$\frac{a_1'(\eta)E(y) - a_2'(\eta)}{\phi} = 0.$$

Hence, now we have an initial expression for $E(Y)$ which is as follows

$$E(Y) = \frac{a_2'(\eta)}{a_1'(\eta)}. \tag{7.11}$$

Now note that,

$$\frac{\partial^2 \ell}{\partial \eta^2} = \frac{a_1''(\eta)(y) - a_2''(\eta)}{\phi}. \tag{7.12}$$

The second Bartlett identity states that

$$E\left(\frac{\partial^2 \ell}{\partial \eta^2}\right) + E\left\{\left(\frac{\partial \ell}{\partial \eta}\right)^2\right\} = 0.$$

Using the first Bartlett identity, we can re-write the second Bartlett identity as follows

$$E\left(\frac{\partial^2 \ell}{\partial \eta^2}\right) + \mathrm{Var}\left(\frac{\partial \ell}{\partial \eta}\right) = 0.$$

Substituting (7.10) and (7.12) into the second Bartlett identity, we have,

$$E\left\{\frac{a_1''(\eta)(y) - a_2''(\eta)}{\phi}\right\} + \mathrm{Var}\left\{\frac{a_1'(\eta)y - a_2'(\eta)}{\phi}\right\} = 0. \tag{7.13}$$

This leads to

$$\frac{a_1''(\eta)E(y) - a_2''(\eta)}{\phi} + \frac{\left(a_1'(\eta)\right)^2}{\phi^2}\mathrm{Var}(y) = 0.$$

Therefore, the initial expression for $\mathrm{Var}(Y)$ is as follows

$$\begin{aligned}
\mathrm{Var}(Y) &= \left[\frac{a_2''(\eta) - a_1''(\eta)\left\{\frac{a_2'(\eta)}{a_1'(\eta)}\right\}}{\phi}\right]\left[\frac{\phi^2}{\left\{a_1'(\eta)\right\}^2}\right] \\
&= \frac{\phi\left\{a_1'(\eta)a_2''(\eta) - a_1''(\eta)a_2'(\eta)\right\}}{\left\{a_1'(\eta)\right\}^3} \\
&= \frac{\phi\left\{a_2'(\eta)/a_1'(\eta)\right\}'}{a_1'(\eta)}.
\end{aligned} \tag{7.14}$$

Now let us simplify the expressions for $E(Y)$ and $\mathrm{Var}(Y)$. Note that

$$\begin{aligned}
a_2(\eta) &= \{b \circ (g \circ b')^{-1}\}(\eta) = (b \circ a_1)(\eta) \\
a_2'(\eta) &= (b \circ a_1)'(\eta) = (b' \circ a_1)(\eta) \times a_1'(\eta).
\end{aligned}$$

Hence, the expression for $E(Y)$ simplifies to

$$\begin{aligned}
E(Y) &= \frac{a_2'(\eta)}{a_1'(\eta)} \\
&= (b' \circ a_1)(\eta) \\
&= \{b' \circ (g \circ b')^{-1}\}(\eta) \\
&= \{b' \circ (b')^{-1} \circ g^{-1}\}(\eta) \\
&= g^{-1}(\eta).
\end{aligned}$$

The expression for $\text{Var}(Y)$ simplifies to

$$\begin{aligned}\text{Var}(Y) &= \frac{\phi\{a_2'(\eta)/a_1'(\eta)\}'}{a_1'(\eta)} \\ &= \frac{\phi\{(b' \circ a_1)'(\eta)\}}{a_1'(\eta)} \\ &= \frac{\phi\{(b'' \circ a_1)(\eta)\}\{a_1'(\eta)\}}{a_1'(\eta)} \\ &= \phi\{b'' \circ (g \circ b')^{-1}\}(\eta) \\ &= \phi\{b'' \circ (b')^{-1} \circ g^{-1}\}(\eta).\end{aligned}$$

7.3.4 Introduction of Useful Notation and Its Properties

Let $\boldsymbol{v}$ be a generic $d \times 1$ vector. Then for $r = 0, 1, 2$ we define

$$\boldsymbol{v}^{\otimes r} \equiv \begin{cases} 1 & \text{for} \quad r = 0 \\ \boldsymbol{v} & \text{for} \quad r = 1 \\ \boldsymbol{v}\boldsymbol{v}^T & \text{for} \quad r = 2. \end{cases}$$

For $r \in \{0, 1, 2\}$, let

$$\mathcal{G}_{ri}^* \equiv \sum_{j=1}^{n} \boldsymbol{X}_{ij}^{\otimes r} \left\{ Y_{ij} a_1'(\beta_0 + \boldsymbol{\beta}_1^T \boldsymbol{X}_{ij} + U_i) - a_2'(\beta_0 + \boldsymbol{\beta}_1^T \boldsymbol{X}_{ij} + U_i) \right\} \quad \text{and}$$

$$\mathcal{H}_{ri}^* \equiv \sum_{j=1}^{n} \boldsymbol{X}_{ij}^{\otimes r} \left\{ a_2''(\beta_0 + \boldsymbol{\beta}_1^T \boldsymbol{X}_{ij} + U_i) - Y_{ij} a_1''(\beta_0 + \boldsymbol{\beta}_1^T \boldsymbol{X}_{ij} + U_i) \right\}.$$

Note that the expressions listed above have the following probabilistic orders where

$$\mathcal{G}_{ri}^* = O_P(n^{1/2}) \mathbf{1}_d^{\otimes r} \quad \text{and} \quad \mathcal{H}_{ri}^* = O_P(n) \mathbf{1}_d^{\otimes r}.$$

7.3.5 Key Conditional Moment Results

We now compute the conditional expectations of $\mathcal{G}_{0i}^*$ and $\mathcal{G}_{1i}^*$ given $(\boldsymbol{X}_i, U_i)$. Note that,

$$\begin{aligned}&E(\mathcal{G}_{0i}^* | \boldsymbol{X}_i, U_i) \\ &= E\left[\sum_{j=1}^{n} \left\{ Y_{ij} a_1'(\beta_0 + \boldsymbol{\beta}_1^T \boldsymbol{X}_{ij} + U_i) - a_2'(\beta_0 + \boldsymbol{\beta}_1^T \boldsymbol{X}_{ij} + U_i) \right\} \middle| \boldsymbol{X}_i, U_i \right] \\ &= \sum_{j=1}^{n} \left\{ E(Y_{ij} | \boldsymbol{X}_i, U_i) a_1'(\beta_0 + \boldsymbol{\beta}_1^T \boldsymbol{X}_{ij} + U_i) - a_2'(\beta_0 + \boldsymbol{\beta}_1^T \boldsymbol{X}_{ij} + U_i) \right\} \\ &= \sum_{j=1}^{n} \left\{ \frac{a_2'(\beta_0 + \boldsymbol{\beta}_1^T \boldsymbol{X}_{ij} + U_i)}{a_1'(\beta_0 + \boldsymbol{\beta}_1^T \boldsymbol{X}_{ij} + U_i)} a_1'(\beta_0 + \boldsymbol{\beta}_1^T \boldsymbol{X}_{ij} + U_i) - a_2'(\beta_0 + \boldsymbol{\beta}_1^T \boldsymbol{X}_{ij} + U_i) \right\} \\ &= 0.\end{aligned}$$

Similarly, we have that

$$E(\mathcal{G}_{1i}^*|\boldsymbol{X}_i, U_i) = \mathbf{0}.$$

Next, we compute the conditional expectations of ${\mathcal{G}_{0i}^*}^2, \mathcal{G}_{0i}^*\mathcal{G}_{1i}^*$ and ${\mathcal{G}_{1i}^*}^2$ given $(\boldsymbol{X}_i, U_i)$. Note that,

$$\begin{aligned}
&E({\mathcal{G}_{0i}^*}^2|\boldsymbol{X}_i, U_i) \\
&= E\Bigg(\left[\sum_{j=1}^{n} \left\{ Y_{ij} a_1'(\beta_0 + \boldsymbol{\beta}_1^T \boldsymbol{X}_{ij} + U_i) - a_2'(\beta_0 + \boldsymbol{\beta}_1^T \boldsymbol{X}_{ij} + U_i) \right\} \right] \\
&\quad \times \left[\sum_{j'=1}^{n} \left\{ Y_{ij'} a_1'(\beta_0 + \boldsymbol{\beta}_1^T \boldsymbol{X}_{ij'} + U_i) - a_2'(\beta_0 + \boldsymbol{\beta}_1^T \boldsymbol{X}_{ij'} + U_i) \right\} \right] \Big| \boldsymbol{X}_i, U_i \Bigg) \\
&= \sum_{j \neq j'} \sum E\Big[\left\{ Y_{ij} a_1'(\beta_0 + \boldsymbol{\beta}_1^T \boldsymbol{X}_{ij} + U_i) - a_2'(\beta_0 + \boldsymbol{\beta}_1^T \boldsymbol{X}_{ij} + U_i) \right\} \\
&\quad \times \left\{ Y_{ij'} a_1'(\beta_0 + \boldsymbol{\beta}_1^T \boldsymbol{X}_{ij'} + U_i) - a_2'(\beta_0 + \boldsymbol{\beta}_1^T \boldsymbol{X}_{ij'} + U_i) \right\} \Big| \boldsymbol{X}_i, U_i \Big] \\
&\quad + \sum_{j=1}^{n} E\Big[\left\{ Y_{ij} a_1'(\beta_0 + \boldsymbol{\beta}_1^T \boldsymbol{X}_{ij} + U_i) - a_2'(\beta_0 + \boldsymbol{\beta}_1^T \boldsymbol{X}_{ij} + U_i) \right\} \\
&\quad \times \left\{ Y_{ij} a_1'(\beta_0 + \boldsymbol{\beta}_1^T \boldsymbol{X}_{ij} + U_i) - a_2'(\beta_0 + \boldsymbol{\beta}_1^T \boldsymbol{X}_{ij} + U_i) \right\} \Big| \boldsymbol{X}_i, U_i \Big] \\
&= \sum_{j \neq j'} \sum \Bigg(E\left[\left\{ Y_{ij} a_1'(\beta_0 + \boldsymbol{\beta}_1^T \boldsymbol{X}_{ij} + U_i) - a_2'(\beta_0 + \boldsymbol{\beta}_1^T \boldsymbol{X}_{ij} + U_i) \right\} \Big| \boldsymbol{X}_i, U_i \right] \\
&\quad \times E\left[\left\{ Y_{ij'} a_1'(\beta_0 + \boldsymbol{\beta}_1^T \boldsymbol{X}_{ij'} + U_i) - a_2'(\beta_0 + \boldsymbol{\beta}_1^T \boldsymbol{X}_{ij'} + U_i) \right\} \Big| \boldsymbol{X}_i, U_i \right] \Bigg) \\
&\quad + \sum_{j=1}^{n} \mathrm{Var}\left\{ Y_{ij} a_1'(\beta_0 + \boldsymbol{\beta}_1^T \boldsymbol{X}_{ij} + U_i) \Big| \boldsymbol{X}_i, U_i \right\} \\
&= \sum_{j=1}^{n} a_1'(\beta_0 + \boldsymbol{\beta}_1^T \boldsymbol{X}_{ij} + U_i)^2 \mathrm{Var}\left(Y_{ij} \Big| \boldsymbol{X}_i, U_i \right).
\end{aligned}$$

Now let $\eta_{ij} \equiv \beta_0 + \boldsymbol{\beta}_1^T \boldsymbol{X}_{ij} + U_i$. By substituting the expression for $\mathrm{Var}(Y)$ in line 2 of (7.14), we have,

$$\begin{aligned}
E({\mathcal{G}_{0i}^*}^2|\boldsymbol{X}_i, U_i) &= \phi \sum_{j=1}^{n} a_1'(\eta_{ij})^2 \left[\frac{\{a_1'(\eta_{ij}) a_2''(\eta_{ij}) - a_1''(\eta_{ij}) a_2'(\eta_{ij})\}}{\{a_1'(\eta_{ij})\}^3} \right] \\
&= \phi \sum_{j=1}^{n} \left\{ a_2''(\eta_{ij}) - \frac{a_1''(\eta_{ij}) a_2'(\eta_{ij})}{a_1'(\eta_{ij})} \right\} \\
&= \phi \sum_{j=1}^{n} \left\{ a_2''(\eta_{ij}) - a_1''(\eta_{ij}) E(Y_{ij}|\boldsymbol{X}_i, U_i) \right\} \\
&= \phi E(\mathcal{H}_{0i}^*|\boldsymbol{X}_i, U_i).
\end{aligned}$$

Similarly, we have that

$$E(\mathcal{G}_{0i}^*\mathcal{G}_{1i}^*|\boldsymbol{X}_i, U_i) = \phi E(\mathcal{H}_{1i}^*|\boldsymbol{X}_i, U_i) \quad \text{and} \quad E({\mathcal{G}_{1i}^*}^2|\boldsymbol{X}_i, U_i) = \phi E(\mathcal{H}_{2i}^*|\boldsymbol{X}_i, U_i).$$

Now let us simplify the expression for $E(\mathcal{H}^*_{ri}|\boldsymbol{X}_i, U_i)$ for $r \in \{0, 1, 2\}$. Note that,

$$\begin{aligned}
E(\mathcal{H}^*_{ri}|\boldsymbol{X}_i, U_i) &= \sum_{j=1}^{n} \boldsymbol{X}_{ij}^{\otimes r} \left\{ a_2''(\eta_{ij}) - a_1''(\eta_{ij}) E(Y_{ij}|\boldsymbol{X}_i, U_i) \right\} \\
&= \sum_{j=1}^{n} \boldsymbol{X}_{ij}^{\otimes r} \left\{ a_2''(\eta_{ij}) - \frac{a_1''(\eta_{ij}) a_2'(\eta_{ij})}{a_1'(\eta_{ij})} \right\} \\
&= \sum_{j=1}^{n} \boldsymbol{X}_{ij}^{\otimes r} a_1'(\eta_{ij})^2 \left[\frac{\{a_1'(\eta_{ij}) a_2''(\eta_{ij}) - a_1''(\eta_{ij}) a_2'(\eta_{ij})\}}{\{a_1'(\eta_{ij})\}^3} \right] \\
&= \sum_{j=1}^{n} \boldsymbol{X}_{ij}^{\otimes r} a_1'(\eta_{ij})^2 \{b'' \circ (g \circ b')^{-1}\}(\eta_{ij}).
\end{aligned}$$

The term $a_1'(\eta_{ij})^2$ can be simplified as follows

$$\begin{aligned}
a_1'(\eta_{ij})^2 &= \left[\left\{ (g \circ b')^{-1} \right\}' \right]^2 (\eta_{ij}) \\
&= \left\{ \frac{1}{(g \circ b')' \circ (g \circ b')^{-1}} \right\}^2 (\eta_{ij}) \\
&= \left[\frac{1}{\{(g \circ b') b''\} \circ \{(b')^{-1} \circ g^{-1}\}} \right]^2 (\eta_{ij}) \\
&= \left[\frac{1}{(g' \circ g^{-1}) \{b'' \circ (b')^{-1} \circ g^{-1}\}} \right]^2 (\eta_{ij}) \\
&= \left[\frac{(g^{-1})'}{\{b'' \circ (b')^{-1} \circ g^{-1}\}} \right]^2 (\eta_{ij}).
\end{aligned}$$

Hence,

$$\begin{aligned}
E(\mathcal{H}^*_{ri}|\boldsymbol{X}_i, U_i) &= \sum_{j=1}^{n} \boldsymbol{X}_{ij}^{\otimes r} \left[\frac{\left\{ (g^{-1})' \right\}^2}{\{b'' \circ (b')^{-1} \circ g^{-1}\}} \right] (\eta_{ij}) \\
&= \sum_{j=1}^{n} \boldsymbol{X}_{ij}^{\otimes r} w(\eta_{ij}),
\end{aligned} \tag{7.15}$$

where

$$w \equiv \frac{\left\{ (g^{-1})' \right\}^2}{b'' \circ (b')^{-1} \circ g^{-1}}.$$

7.3.6 Computing an Asymptotic Approximation for the First Entry in (7.9)

Once again, to overcome the intractability of the ratio of integrals present when deriving the scores with respect to each of the model parameters, we will work with an asymptotic approximation of the ratio of integrals by using a multi-term Laplace's method expansion.

For smooth real-valued functions b, g and h, Equation (2.6) of Tierney et al. (1989) states that

$$\frac{\int_{-\infty}^{\infty} b_N(x)\exp\{-nh(x)\}\,dx}{\int_{-\infty}^{\infty} b_D(x)\exp\{-nh(x)\}\,dx} = g(x^*) + \frac{b_D'(x^*)g'(x^*)}{nb_D(x^*)h''(x^*)} + \frac{g''(x^*)}{2nh''(x^*)} - \frac{g'(x^*)h'''(x^*)}{2nh''(x^*)^2} + O(n^{-2}).$$

where

$$g \equiv b_N/b_D$$

and

$$x^* \equiv \text{value of } x \text{ that minimises } h \text{ over } \mathbb{R}.$$

Hence, the ith contribution to the score of β_0 can be expressed as

$$\begin{aligned} S_{0i} &\equiv \frac{\partial \log p_{\boldsymbol{Y}_i|\boldsymbol{X}_i}(\boldsymbol{Y}_i|\boldsymbol{X}_i)}{\partial\beta_0} \\ &= \frac{\int_{-\infty}^{\infty} b_N^{1st}(u)\exp\{-nh_N(u)\}du}{\int_{-\infty}^{\infty} b_D^{1st}(u)\exp\{-nh_N(u)\}du} \end{aligned}$$

where

$$b_N^{1st}(u) = \exp\left(-\frac{u^2}{2\sigma^2}\right)\frac{1}{\phi}\sum_{j=1}^{n}\{Y_{ij}a_1'(\beta_0+\boldsymbol{\beta}_1^T\boldsymbol{X}_{ij}+u) - a_2'(\beta_0+\boldsymbol{\beta}_1^T\boldsymbol{X}_{ij}+u)\},$$

$$b_D^{1st}(u) = \exp\left(-\frac{u^2}{2\sigma^2}\right) \quad \text{and}$$

$$h_N(u) = -\frac{1}{n\phi}\sum_{j=1}^{n}\left\{Y_{ij}a_1(\beta_0+\boldsymbol{\beta}_1^T\boldsymbol{X}_{ij}+u) - a_2(\beta_0+\boldsymbol{\beta}_1^T\boldsymbol{X}_{ij}+u)\right\}.$$

Now define

$$\begin{aligned} U_i^* &\equiv \text{value of } u \text{ that minimises } h_N(u) \\ &= \text{value of } u \text{ such that } \tfrac{d}{du}h_N(u) = 0 \\ &= \text{value of } u \text{ such that } \sum_{j=1}^{n}\left\{Y_{ij}a_1'(\beta_0+\boldsymbol{\beta}_1^T\boldsymbol{X}_{ij}+u) - a_2'(\beta_0+\boldsymbol{\beta}_1^T\boldsymbol{X}_{ij}+u)\right\} = 0. \end{aligned}$$

However,

$$b_N^{1st}(U_i^*) = 0.$$

This violates the necessary condition in Hsu (1948), in the sense that $b_N^{1st}(U_i^*) \neq 0$ is required in order for the Laplace approximation to hold. To counter this issue, firstly note that the numerator of S_{0i} is

$$\int_{-\infty}^{\infty} s'(u)t(u)du$$

where

$$s(u) \equiv \exp\left(\frac{n}{\phi}\left[\frac{1}{n}\sum_{j=1}^{n}\left\{Y_{ij}a_1(\beta_0+\boldsymbol{\beta}_1^T\boldsymbol{X}_{ij}+u) - a_2(\beta_0+\boldsymbol{\beta}_1^T\boldsymbol{X}_{ij}+u)\right\}\right]\right)$$

and

$$t(u) \equiv \exp\left(-\frac{u^2}{2\sigma^2}\right).$$

Application of integration by parts leads to the integral being equal to

$$-\int_{-\infty}^{\infty} s(u)t'(u)du.$$

Note that

$$t'(u) = (-u/\sigma^2)\exp\left(-\frac{u^2}{2\sigma^2}\right).$$

Now by rewriting the numerator of S_{0i}, we have,

$$\begin{aligned} S_{0i} &\equiv \frac{\partial \log p_{\boldsymbol{Y}_i|\boldsymbol{X}_i}(\boldsymbol{Y}_i|\boldsymbol{X}_i)}{\partial \beta_0} \\ &= \frac{\int_{-\infty}^{\infty} b_N(u)\exp\{-nh_N(u)\}du}{\int_{-\infty}^{\infty} b_D(u)\exp\{-nh_N(u)\}du} \end{aligned}$$

where

$$\begin{aligned} b_N(u) &= (u/\sigma^2)\exp\left(-\frac{u^2}{2\sigma^2}\right), \\ b_D(u) &= \exp\left(-\frac{u^2}{2\sigma^2}\right) \quad \text{and} \\ h_N(u) &= -\frac{1}{n\phi}\sum_{j=1}^{n}\left\{Y_{ij}a_1(\beta_0 + \boldsymbol{\beta}_1^T\boldsymbol{X}_{ij} + u) - a_2(\beta_0 + \boldsymbol{\beta}_1^T\boldsymbol{X}_{ij} + u)\right\}. \end{aligned}$$

Expansion of U_i^*

Here we find an asymptotic expression for U_i^*. We have that

$$\begin{aligned} 0 &= \frac{d}{du}h_N(\boldsymbol{u}) \\ &= \sum_{j=1}^{n}\left\{Y_{ij}a_1'(\beta_0 + \boldsymbol{\beta}_1^T\boldsymbol{X}_{ij} + U_i^*) - a_2'(\beta_0 + \boldsymbol{\beta}_1^T\boldsymbol{X}_{ij} + U_i^*)\right\} \\ &= \sum_{j=1}^{n}\left\{Y_{ij}a_1'(\beta_0 + \boldsymbol{\beta}_1^T\boldsymbol{X}_{ij} + U_i) - a_2'(\beta_0 + \boldsymbol{\beta}_1^T\boldsymbol{X}_{ij} + U_i)\right\} \\ &\quad - (U_i^* - U_i)\sum_{j=1}^{n}\left\{a_2''(\beta_0 + \boldsymbol{\beta}_1^T\boldsymbol{X}_{ij} + U_i) - Y_{ij}a_1''(\beta_0 + \boldsymbol{\beta}_1^T\boldsymbol{X}_{ij} + U_i)\right\} + r_{it} \\ &= \mathcal{G}_{0i}^* - (U_i^* - U_i)\mathcal{H}_{0i}^* + r_{it} \end{aligned}$$

where r_{it} is the Lagrange form of the remainder and is a quadratic form in $U_i^* - U_i$ and a smooth function of $U_{it}^\dagger \equiv (1-t)U_i + tU_i^*$ for some $t \in [0,1]$. Inversion of this asymptotic series leads to

$$U_i^* = U_i + \frac{\mathcal{G}_{0i}^*}{\mathcal{H}_{0i}^*} + O_P(n^{-1}).$$

7.3.6.1 The First Term of the First Score

The first term of S_{0i} is

$$g(U_i^*) = \frac{b_N(U_i^*)}{b_D(U_i^*)} = \frac{U_i^*}{\sigma^2} = \frac{1}{\sigma^2}\left(U_i + \frac{\mathcal{G}_{0i}^*}{\mathcal{H}_{0i}^*}\right) + O_P(n^{-1}).$$

7.3.6.2 The Other Terms of the First Score

The second, third and fourth terms of S_{0i} are $O_P(n^{-1})$, 0 and $O_P(n^{-1})$ as in the canonical case in Chapter 4.

7.3.6.3 Overall Leading Term Expression for the First Score

The first term of S_{0i} is $O_P(1)$, the second term of S_{0i} is $O_P(n^{-1})$, the third term of S_{0i} is 0 and the fourth term of S_{0i} is $O_P(n^{-1})$. Putting these together we obtain the following asymptotic expansion for S_{0i} such that

$$S_{0i} = \frac{1}{\sigma^2}\left(U_i + \frac{\mathcal{G}_{0i}^*}{\mathcal{H}_{0i}^*}\right) + O_P(n^{-1}).$$

7.3.7 Computing an Asymptotic Approximation for the Second Entry in (7.9)

The ith contribution to the score of $\boldsymbol{\beta}_1$ is

$$\begin{aligned} S_{1i} &\equiv \nabla_{\boldsymbol{\beta}_1} \log p_{\boldsymbol{Y}_i|\boldsymbol{X}_i}(\boldsymbol{Y}_i|\boldsymbol{X}_i) \\ &= \frac{\int_{-\infty}^{\infty} b_N(u)\exp\{-nh_N(u)\}du}{\int_{-\infty}^{\infty} b_D(u)\exp\{-nh_N(u)\}du} \end{aligned}$$

where

$$b_N(u) = \exp\left(-\frac{u^2}{2\sigma^2}\right)\frac{1}{\phi}\sum_{j=1}^{n}\{Y_{ij}a_1'(\beta_0 + \boldsymbol{\beta}_1^T\boldsymbol{X}_{ij} + u) - a_2'(\beta_0 + \boldsymbol{\beta}_1^T\boldsymbol{X}_{ij} + u)\}\boldsymbol{X}_{ij},$$

$$b_D(u) = \exp\left(-\frac{u^2}{2\sigma^2}\right) \text{ and}$$

$$h_N(u) = -\frac{1}{n\phi}\sum_{j=1}^{n}\left\{Y_{ij}a_1(\beta_0 + \boldsymbol{\beta}_1^T\boldsymbol{X}_{ij} + u) - a_2(\beta_0 + \boldsymbol{\beta}_1^T\boldsymbol{X}_{ij} + u)\right\}.$$

7.3.7.1 The First Term of the Second Score

The first term of S_{1i} is

$$\begin{aligned} g(U_i^*) &= \frac{b_N(U_i^*)}{b_D(U_i^*)} \\ &= \frac{1}{\phi}\sum_{j=1}^{n}\{Y_{ij}a_1'(\beta_0+\boldsymbol{\beta}_1^T\boldsymbol{X}_{ij}+U_i^*) - a_2'(\beta_0+\boldsymbol{\beta}_1^T\boldsymbol{X}_{ij}+U_i^*)\}\boldsymbol{X}_{ij}. \end{aligned}$$

Next note that,

$$\begin{aligned} &Y_{ij}a_1'(\beta_0+\boldsymbol{\beta}_1^T\boldsymbol{X}_{ij}+U_i^*) - a_2'(\beta_0+\boldsymbol{\beta}_1^T\boldsymbol{X}_{ij}+U_i^*) \\ &= Y_{ij}a_1'(\beta_0+\boldsymbol{\beta}_1^T\boldsymbol{X}_{ij}+U_i) - a_2'(\beta_0+\boldsymbol{\beta}_1^T\boldsymbol{X}_{ij}+U_i) \\ &\quad + (U_i^*-U_i)\left\{Y_{ij}a_1''(\beta_0+\boldsymbol{\beta}_1^T\boldsymbol{X}_{ij}+U_i) - a_2''(\beta_0+\boldsymbol{\beta}_1^T\boldsymbol{X}_{ij}+U_i)\right\} + O_P(n^{-1}). \end{aligned}$$

Therefore,

$$g(U_i^*) = \frac{1}{\phi}\left(\mathcal{G}_{1i}^* - \frac{\mathcal{G}_{0i}^*\mathcal{H}_{1i}^*}{\mathcal{H}_{0i}^*}\right) + O_P(1)\mathbf{1}_d.$$

7.3.7.2 The Other Terms of the Second Score

The second, third and fourth terms are $O_P(1)\mathbf{1}_d$ as in the canonical case in Chapter 4.

7.3.7.3 Overall Leading Term Expression for the Second Score

Putting the terms of the score together, we obtain the following asymptotic expansion for S_{1i} where

$$S_{1i} = \frac{1}{\phi}\left(\mathcal{G}_{1i}^* - \frac{\mathcal{G}_{0i}^*\mathcal{H}_{1i}^*}{\mathcal{H}_{0i}^*}\right) + O_P(1)\mathbf{1}_d.$$

7.3.8 Computing an Asymptotic Approximation for the Third Entry in (7.9)

The ith contribution to the score of σ^2 is

$$\begin{aligned} S_{2i} &\equiv \frac{\partial \log p_{\boldsymbol{Y}_i|\boldsymbol{X}_i}(\boldsymbol{Y}_i|\boldsymbol{X}_i)}{\partial\sigma^2} \\ &= -\frac{1}{2\sigma^2} + \frac{\int_{-\infty}^{\infty} b_N(u)\exp\{-nh_N(u)\}du}{\int_{-\infty}^{\infty} b_D(u)\exp\{-nh_N(u)\}du} \end{aligned}$$

where

$$b_N(u) \equiv \frac{u^2}{2\sigma^4}\exp\left(-\frac{u^2}{2\sigma^2}\right),$$
$$b_D(u) \equiv \exp\left(-\frac{u^2}{2\sigma^2}\right) \text{ and}$$
$$h_N(u) \equiv -\frac{1}{n\phi}\sum_{j=1}^{n}\left\{Y_{ij}a_1(\beta_0+\boldsymbol{\beta}_1^T\boldsymbol{X}_{ij}+u)-a_2(\beta_0+\boldsymbol{\beta}_1^T\boldsymbol{X}_{ij}+u)\right\}.$$

7.3.8.1 The First Term of the Third Score

The first term of S_{2i} is

$$\begin{aligned} g(U_i^*) = \frac{b_N(U_i^*)}{b_D(U_i^*)} &= \frac{(U_i^*)^2}{2\sigma^4} \\ &= \frac{1}{2\sigma^4}\left(U_i+\frac{\mathcal{G}_{0i}^*}{\mathcal{H}_{0i}^*}+O_P(n^{-1})\right)^2 \\ &= \frac{1}{2\sigma^4}\left(U_i^2+\frac{2U_i\mathcal{G}_{0i}^*}{\mathcal{H}_{0i}^*}+O_P(n^{-1})\right). \end{aligned}$$

7.3.8.2 The Other Terms of the Third Score

The second, third and fourth terms of S_{2i} are all $O_P(n^{-1})$ as in the canonical case in Chapter 4.

7.3.8.3 Overall Leading Term Expression for the Third Score

The first term of S_{2i} is $O_P(1)$, the second term of S_{2i} is $O_P(n^{-1})$, the third term of S_{2i} is $O_P(n^{-1})$ and the fourth term of S_{2i} is $O_P(n^{-1})$. Putting these together we obtain the following asymptotic expansion for S_{2i} such that

$$S_{2i} = -\frac{1}{2\sigma^2}+\frac{U_i^2}{2\sigma^4}+\frac{U_i\mathcal{G}_{0i}^*}{\sigma^4\mathcal{H}_{0i}^*}+O_P(n^{-1}).$$

7.3.9 The Quadratic Conditional Expectations of the Scores

In this section we find the conditional expectations required for the diagonal entries of the Fisher information matrix of $(\beta_0, \boldsymbol{\beta}_1, \sigma^2)$.

7.3.9.1 The Conditional Expectation of the Square of the First Score

From the previous sections, we have the following approximation where

$$S_{0i} = \frac{U_i}{\sigma^2} + O_P(n^{-1/2}).$$

Therefore,

$$E(S_{0i}^2|\boldsymbol{X}_i) = E\left\{\left(\frac{U_i}{\sigma^2} + O_P(n^{-1/2})\right)^2 \Big| \boldsymbol{X}_i\right\} = \frac{1}{\sigma^2} + O_P(n^{-1}).$$

7.3.9.2 The Conditional Expectation of the Square of the Second Score

We have the following approximation

$$\phi S_{1i} = \mathcal{G}_{1i}^* - \frac{\mathcal{G}_{0i}^*\mathcal{H}_{1i}^*}{\mathcal{H}_{0i}^*} + O_P(1)\mathbf{1}_d.$$

Therefore,

$$\begin{aligned}
\phi^2 E(S_{1i}S_{1i}^T|\boldsymbol{X}_i) &= E\left\{\left(\mathcal{G}_{1i}^* - \frac{\mathcal{G}_{0i}^*\mathcal{H}_{1i}^*}{\mathcal{H}_{0i}^*} + O_P(1)\mathbf{1}_d\right)^{\otimes 2} \Big| \boldsymbol{X}_i\right\} \\
&= E({\mathcal{G}_{1i}^*}^{\otimes 2}|\boldsymbol{X}_i) + E\left(\frac{{\mathcal{G}_{0i}^*}^2{\mathcal{H}_{1i}^*}^{\otimes 2}}{{\mathcal{H}_{0i}^*}^2}\Big|\boldsymbol{X}_i\right) - E\left(\frac{\mathcal{G}_{0i}^*\mathcal{G}_{1i}^*{\mathcal{H}_{1i}^*}^T}{\mathcal{H}_{0i}^*}\Big|\boldsymbol{X}_i\right) \\
&\quad - E\left(\frac{\mathcal{G}_{0i}^*\mathcal{H}_{1i}^*{\mathcal{G}_{1i}^*}^T}{\mathcal{H}_{0i}^*}\Big|\boldsymbol{X}_i\right) + O_P(1)\mathbf{1}_d^{\otimes 2}.
\end{aligned}$$

We will now solve for the conditional expectations occurring in $\phi^2 E(S_{1i}S_{1i}^T|\boldsymbol{X}_i)$. Firstly,

$$\begin{aligned}
E({\mathcal{G}_{1i}^*}^{\otimes 2}|\boldsymbol{X}_i) &= E\left\{E({\mathcal{G}_{1i}^*}^{\otimes 2}|\boldsymbol{X}_i, U_i)|\boldsymbol{X}_i\right\} \\
&= E\left\{\phi E(\mathcal{H}_{2i}^*|\boldsymbol{X}_i, U_i)|\boldsymbol{X}_i\right\} \\
&= \phi E\left\{E(\mathcal{H}_{2i}^*|\boldsymbol{X}_i, U_i)|\boldsymbol{X}_i\right\}.
\end{aligned}$$

Next,

$$\begin{aligned}
&E\left(\frac{{\mathcal{G}_{0i}^*}^2{\mathcal{H}_{1i}^*}^{\otimes 2}}{{\mathcal{H}_{0i}^*}^2}\Big|\boldsymbol{X}_i\right) \\
&= E\left\{E\left(\frac{{\mathcal{G}_{0i}^*}^2{\mathcal{H}_{1i}^*}^{\otimes 2}}{{\mathcal{H}_{0i}^*}^2}\Big|\boldsymbol{X}_i, U_i\right)\Big|\boldsymbol{X}_i\right\} \\
&= E\left(E\left[\frac{{\mathcal{G}_{0i}^*}^2\left\{E(\mathcal{H}_{1i}^*|\boldsymbol{X}_i, U_i) + \mathcal{H}_{1i}^* - E(\mathcal{H}_{1i}^*|\boldsymbol{X}_i, U_i)\right\}^{\otimes 2}}{\left\{E(\mathcal{H}_{0i}^*|\boldsymbol{X}_i, U_i) + \mathcal{H}_{0i}^* - E(\mathcal{H}_{0i}^*|\boldsymbol{X}_i, U_i)\right\}^2}\Big|\boldsymbol{X}_i, U_i\right]\Big|\boldsymbol{X}_i\right).
\end{aligned}$$

The previous expression can be further evaluated as follows

$$
\begin{aligned}
&E\left(E\left[\frac{\mathcal{G}_{0i}^{*\,2}\left\{E(\mathcal{H}_{1i}^{*}|\boldsymbol{X}_i,U_i)+\mathcal{H}_{1i}^{*}-E(\mathcal{H}_{1i}^{*}|\boldsymbol{X}_i,U_i)\right\}^{\otimes 2}}{\left\{E(\mathcal{H}_{0i}^{*}|\boldsymbol{X}_i,U_i)\right\}^{2}\left\{1-\frac{E(\mathcal{H}_{0i}^{*}|\boldsymbol{X}_i,U_i)-\mathcal{H}_{0i}^{*}}{E(\mathcal{H}_{0i}^{*}|\boldsymbol{X}_i,U_i)}\right\}^{2}}\middle|\boldsymbol{X}_i,U_i\right]\middle|\boldsymbol{X}_i\right)\\
&=E\Bigg(E\Bigg[\frac{\mathcal{G}_{0i}^{*\,2}\left\{E(\mathcal{H}_{1i}^{*}|\boldsymbol{X}_i,U_i)+\mathcal{H}_{1i}^{*}-E(\mathcal{H}_{1i}^{*}|\boldsymbol{X}_i,U_i)\right\}^{\otimes 2}}{\left\{E(\mathcal{H}_{0i}^{*}|\boldsymbol{X}_i,U_i)\right\}^{2}}\\
&\qquad\times\left\{1-\frac{E(\mathcal{H}_{0i}^{*}|\boldsymbol{X}_i,U_i)-\mathcal{H}_{0i}^{*}}{E(\mathcal{H}_{0i}^{*}|\boldsymbol{X}_i,U_i)}\right\}^{2}\Bigg|\boldsymbol{X}_i,U_i\Bigg]\Bigg|\boldsymbol{X}_i\Bigg)\\
&=E\left(E\left[\frac{\mathcal{G}_{0i}^{*\,2}\left\{E(\mathcal{H}_{1i}^{*}|\boldsymbol{X}_i,U_i)\right\}^{\otimes 2}}{\left\{E(\mathcal{H}_{0i}^{*}|\boldsymbol{X}_i,U_i)\right\}^{2}}\middle|\boldsymbol{X}_i,U_i\right]\middle|\boldsymbol{X}_i\right)+\mathcal{R}_i\\
&=\phi E\left[\frac{E(\mathcal{H}_{0i}^{*}|\boldsymbol{X}_i,U_i)\left\{E(\mathcal{H}_{1i}^{*}|\boldsymbol{X}_i,U_i)\right\}^{\otimes 2}}{\left\{E(\mathcal{H}_{0i}^{*}|\boldsymbol{X}_i,U_i)\right\}^{2}}\middle|\boldsymbol{X}_i\right]+\mathcal{R}_i\\
&=\phi E\left[\frac{\left\{E(\mathcal{H}_{1i}^{*}|\boldsymbol{X}_i,U_i)\right\}^{\otimes 2}}{E(\mathcal{H}_{0i}^{*}|\boldsymbol{X}_i,U_i)}\middle|\boldsymbol{X}_i\right]+\mathcal{R}_i
\end{aligned}
$$

where $\mathcal{R}_i$ comprises of remainder terms that are of lower order as compared to the leading terms in the final steps of the derivation. Now note that,

$$
\begin{aligned}
&E\left(\frac{\mathcal{G}_{0i}^{*}\mathcal{G}_{1i}^{*}\mathcal{H}_{1i}^{*\,T}}{\mathcal{H}_{0i}^{*}}\middle|\boldsymbol{X}_i\right)\\
&=E\left\{E\left(\frac{\mathcal{G}_{0i}^{*}\mathcal{G}_{1i}^{*}\mathcal{H}_{1i}^{*\,T}}{\mathcal{H}_{0i}^{*}}\middle|\boldsymbol{X}_i,U_i\right)\middle|\boldsymbol{X}_i\right\}\\
&=E\left(E\left[\frac{\mathcal{G}_{0i}^{*}\mathcal{G}_{1i}^{*}\left\{E(\mathcal{H}_{1i}^{*}|\boldsymbol{X}_i,U_i)+\mathcal{H}_{1i}^{*}-E(\mathcal{H}_{1i}^{*}|\boldsymbol{X}_i,U_i)\right\}^{T}}{\left\{E(\mathcal{H}_{0i}^{*}|\boldsymbol{X}_i,U_i)+\mathcal{H}_{0i}^{*}-E(\mathcal{H}_{0i}^{*}|\boldsymbol{X}_i,U_i)\right\}}\middle|\boldsymbol{X}_i,U_i\right]\middle|\boldsymbol{X}_i\right)\\
&=E\left(E\left[\frac{\mathcal{G}_{0i}^{*}\mathcal{G}_{1i}^{*}\left\{E(\mathcal{H}_{1i}^{*}|\boldsymbol{X}_i,U_i)+\mathcal{H}_{1i}^{*}-E(\mathcal{H}_{1i}^{*}|\boldsymbol{X}_i,U_i)\right\}^{T}}{\left\{E(\mathcal{H}_{0i}^{*}|\boldsymbol{X}_i,U_i)\right\}\left\{1-\frac{E(\mathcal{H}_{0i}^{*}|\boldsymbol{X}_i,U_i)-\mathcal{H}_{0i}^{*}}{E(\mathcal{H}_{0i}^{*}|\boldsymbol{X}_i,U_i)}\right\}}\middle|\boldsymbol{X}_i,U_i\right]\middle|\boldsymbol{X}_i\right)\\
&=E\Bigg(E\Bigg[\frac{\mathcal{G}_{0i}^{*}\mathcal{G}_{1i}^{*}\left\{E(\mathcal{H}_{1i}^{*}|\boldsymbol{X}_i,U_i)+\mathcal{H}_{1i}^{*}-E(\mathcal{H}_{1i}^{*}|\boldsymbol{X}_i,U_i)\right\}^{T}}{\left\{E(\mathcal{H}_{0i}^{*}|\boldsymbol{X}_i,U_i)\right\}}\\
&\qquad\times\left\{1-\frac{E(\mathcal{H}_{0i}^{*}|\boldsymbol{X}_i,U_i)-\mathcal{H}_{0i}^{*}}{E(\mathcal{H}_{0i}^{*}|\boldsymbol{X}_i,U_i)}\right\}\Bigg|\boldsymbol{X}_i,U_i\Bigg]\Bigg|\boldsymbol{X}_i\Bigg)\\
&=E\left[E\left\{\frac{\mathcal{G}_{0i}^{*}\mathcal{G}_{1i}^{*}E(\mathcal{H}_{1i}^{*}|\boldsymbol{X}_i,U_i)^{T}}{\left\{E(\mathcal{H}_{0i}^{*}|\boldsymbol{X}_i,U_i)\right\}}\middle|\boldsymbol{X}_i,U_i\right\}\middle|\boldsymbol{X}_i\right]+\mathcal{R}_i\\
&=\phi E\left[\frac{\left\{E(\mathcal{H}_{1i}^{*}|\boldsymbol{X}_i,U_i)\right\}^{\otimes 2}}{E(\mathcal{H}_{0i}^{*}|\boldsymbol{X}_i,U_i)}\middle|\boldsymbol{X}_i\right]+\mathcal{R}_i.
\end{aligned}
$$

Similarly,

$$
E\left(\frac{\mathcal{G}_{0i}^{*}\mathcal{H}_{1i}^{*}\mathcal{G}_{1i}^{*\,T}}{\mathcal{H}_{0i}^{*}}\middle|\boldsymbol{X}_i\right)=\phi E\left[\frac{\left\{E(\mathcal{H}_{1i}^{*}|\boldsymbol{X}_i,U_i)\right\}^{\otimes 2}}{E(\mathcal{H}_{0i}^{*}|\boldsymbol{X}_i,U_i)}\middle|\boldsymbol{X}_i\right]+\mathcal{R}_i,
$$

where $\mathcal{R}_i$ comprises of remainder terms that are of lower order as compared to the leading term in the expression. Combining the results so far in this subsection, we obtain

$$E(S_{1i}S_{1i}^T|\boldsymbol{X}_i) = \frac{1}{\phi}E\left[E(\mathcal{H}_{2i}^*|\boldsymbol{X}_i,U_i) - \frac{\{E(\mathcal{H}_{1i}^*|\boldsymbol{X}_i,U_i)\}^{\otimes 2}}{E(\mathcal{H}_{0i}^*|\boldsymbol{X}_i,U_i)}\middle|\boldsymbol{X}_i\right] + \mathcal{R}.$$

7.3.9.3 The Conditional Expectation of the Square of the Third Score

We have the following approximation

$$S_{2i} = -\frac{1}{2\sigma^2} + \frac{U_i^2}{2\sigma^4} + \frac{U_i\mathcal{G}_{0i}^*}{\sigma^4\mathcal{H}_{0i}^*} + O_P(n^{-1}).$$

Therefore,

$$\begin{aligned}
4\sigma^8 E(S_{2i}^2|\boldsymbol{X}_i) &= E\left\{\left(U_i^2 - \sigma^2 + \frac{2U_i\mathcal{G}_{0i}^*}{\mathcal{H}_{0i}^*} + O_P(n^{-1})\right)^2\middle|\boldsymbol{X}_i\right\} \\
&= E(U_i^4) + \sigma^4 + 4E\left(\frac{U_i^2{\mathcal{G}_{0i}^*}^2}{{\mathcal{H}_{0i}^*}^2}\middle|\boldsymbol{X}_i\right) - 2\sigma^2E(U_i^2) \\
&\quad + 2E\left(\frac{U_i^3\mathcal{G}_{0i}^*}{\mathcal{H}_{0i}^*}\middle|\boldsymbol{X}_i\right) - 4\sigma^2\left(\frac{U_i\mathcal{G}_{0i}^*}{\mathcal{H}_{0i}^*}\middle|\boldsymbol{X}_i\right) + O_P(n^{-1}).
\end{aligned}$$

Now note that,

$$\begin{aligned}
&E\left(\frac{U_i^2{\mathcal{G}_{0i}^*}^2}{{\mathcal{H}_{0i}^*}^2}\middle|\boldsymbol{X}_i\right) \\
&= E\left\{E\left(\frac{U_i^2{\mathcal{G}_{0i}^*}^2}{{\mathcal{H}_{0i}^*}^2}\middle|\boldsymbol{X}_i,U_i\right)\middle|\boldsymbol{X}_i\right\} \\
&= E\left(E\left[\frac{U_i^2{\mathcal{G}_{0i}^*}^2}{\{E(\mathcal{H}_{0i}^*|\boldsymbol{X}_i,U_i) + \mathcal{H}_{0i}^* - E(\mathcal{H}_{0i}^*|\boldsymbol{X}_i,U_i)\}^2}\middle|\boldsymbol{X}_i,U_i\right]\middle|\boldsymbol{X}_i\right) \\
&= E\left(E\left[\frac{U_i^2{\mathcal{G}_{0i}^*}^2}{\{E(\mathcal{H}_{0i}^*|\boldsymbol{X}_i,U_i)\}^2\left\{1 - \frac{E(\mathcal{H}_{0i}^*|\boldsymbol{X}_i,U_i) - \mathcal{H}_{0i}^*}{E(\mathcal{H}_{0i}^*|\boldsymbol{X}_i,U_i)}\right\}^2}\middle|\boldsymbol{X}_i,U_i\right]\middle|\boldsymbol{X}_i\right) \\
&= E\left(E\left[\left\{\frac{U_i^2{\mathcal{G}_{0i}^*}^2}{\{E(\mathcal{H}_{0i}^*|\boldsymbol{X}_i,U_i)\}^2}\right\}\left\{1 - \frac{E(\mathcal{H}_{0i}^*|\boldsymbol{X}_i,U_i) - \mathcal{H}_{0i}^*}{E(\mathcal{H}_{0i}^*|\boldsymbol{X}_i,U_i)}\right\}^2\middle|\boldsymbol{X}_i,U_i\right]\middle|\boldsymbol{X}_i\right) \\
&= E\left(E\left[\frac{U_i^2{\mathcal{G}_{0i}^*}^2}{\{E(\mathcal{H}_{0i}^*|\boldsymbol{X}_i,U_i)\}^2}\middle|\boldsymbol{X}_i,U_i\right]\middle|\boldsymbol{X}_i\right) + \mathcal{R}_i \\
&= \phi E\left[\frac{U_i^2E(\mathcal{H}_{0i}^*|\boldsymbol{X}_i,U_i)}{\{E(\mathcal{H}_{0i}^*|\boldsymbol{X}_i,U_i)\}^2}\middle|\boldsymbol{X}_i\right] + \mathcal{R}_i \\
&= \phi E\left[\frac{U_i^2}{E(\mathcal{H}_{0i}^*|\boldsymbol{X}_i,U_i)}\middle|\boldsymbol{X}_i\right] + \mathcal{R}_i \\
&= O_P(n^{-1}).
\end{aligned}$$

Also, for $k = 1, 3$,

$$E\left(\frac{U_i^k \mathcal{G}_{0i}^*}{\mathcal{H}_{0i}^*}\middle| \boldsymbol{X}_i\right) = 0.$$

Hence,

$$\begin{aligned} 4\sigma^8 E(S_{2i}^2|\boldsymbol{X}_i) &= E(U_i^4) + \sigma^4 - 2\sigma^2 E(U_i^2) + O_P(n^{-1}) \\ &= 3\sigma^4 + \sigma^4 - 2\sigma^4 + O_P(n^{-1}) \\ &= 2\sigma^4 + O_P(n^{-1}). \end{aligned}$$

Therefore we have,

$$E(S_{2i}^2|\boldsymbol{X}_i) = \frac{1}{2\sigma^4} + O_P(n^{-1}).$$

7.3.10 The Fisher Information Matrix

The conditional expectations for the off-diagonal entries of the Fisher information matrix of $(\beta_0, \boldsymbol{\beta}_1, \sigma^2)$ expressed in terms of order notation are the same as in the canonical case. Putting together the expressions for the quadratic conditional expectations of the scores from the earlier three subsections, we have

$$\begin{aligned} &I(\beta_0, \boldsymbol{\beta}_1, \sigma^2) \\ &= \begin{bmatrix} \frac{m}{\sigma^2} + O_P(mn^{-1}) & O_P(m)\mathbf{1}_d^T & O_P(mn^{-1}) \\ O_P(m)\mathbf{1}_d & \frac{m}{\phi} E\left[E(\mathcal{H}_{2i}^*|\boldsymbol{X}_i, U_i) - \frac{\{E(\mathcal{H}_{1i}^*|\boldsymbol{X}_i, U_i)\}^{\otimes 2}}{E(\mathcal{H}_{0i}^*|\boldsymbol{X}_i, U_i)}\middle|\boldsymbol{X}_i\right] + \mathcal{R} & O_P(m)\mathbf{1}_d \\ O_P(mn^{-1}) & O_P(m)\mathbf{1}_d^T & \frac{m}{2\sigma^4} + O_P(mn^{-1}) \end{bmatrix}. \end{aligned}$$

However note that by using (7.15), we can rewrite the expression below as follows,

$$\begin{aligned} E(S_{1i}S_{1i}^T|\boldsymbol{X}_i) &= \frac{1}{\phi} E\left[E(\mathcal{H}_{2i}^*|\boldsymbol{X}_i, U_i) - \frac{\{E(\mathcal{H}_{1i}^*|\boldsymbol{X}_i, U_i)\}^{\otimes 2}}{E(\mathcal{H}_{0i}^*|\boldsymbol{X}_i, U_i)}\middle|\boldsymbol{X}_i\right] + \mathcal{R} \\ &= \frac{1}{\phi} E\left[\sum_{j=1}^n \boldsymbol{X}_{ij}^{\otimes 2} w(\eta_{ij}) - \frac{\left\{\sum_{j=1}^n \boldsymbol{X}_{ij} w(\eta_{ij})\right\}^{\otimes 2}}{\sum_{j=1}^n w(\eta_{ij})}\middle|\boldsymbol{X}_i\right] + \mathcal{R}. \end{aligned}$$

It follows that the leading term in $(2, 2)$ block of $I(\beta_0, \boldsymbol{\beta}_1, \sigma^2)$ is,

$$\frac{mn}{\phi}\left(\frac{1}{mn}\sum_{i=1}^m E\left[\sum_{j=1}^n \boldsymbol{X}_{ij}^{\otimes 2} w(\eta_{ij}) - \frac{\left\{\sum_{j=1}^n \boldsymbol{X}_{ij} w(\eta_{ij})\right\}^{\otimes 2}}{\sum_{j=1}^n w(\eta_{ij})}\middle|\boldsymbol{X}_i\right]\right) = \frac{mn\boldsymbol{\Sigma}_{\boldsymbol{\beta}_1}}{\phi}$$

where

$$\begin{aligned}\boldsymbol{\Sigma}_{\boldsymbol{\beta}_1} &\equiv \frac{1}{mn}\sum_{i=1}^{m} E\left[\sum_{j=1}^{n}\boldsymbol{X}_{ij}^{\otimes 2}w(\eta_{ij}) - \frac{\left\{\sum_{j=1}^{n}\boldsymbol{X}_{ij}w(\eta_{ij})\right\}^{\otimes 2}}{\sum_{j=1}^{n}w(\eta_{ij})}\middle|\boldsymbol{X}_i\right] \\ &= \frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n}\boldsymbol{X}_{ij}^{\otimes 2}E\{w(\eta_{ij})|\boldsymbol{X}_i\} \\ &\quad - E\left[\frac{1}{mn}\sum_{i=1}^{m}\left\{\sum_{j=1}^{n}\boldsymbol{X}_{ij}w(\eta_{ij})\right\}\left\{\sum_{j=1}^{n}w(\eta_{ij})\right\}^{-1}\left\{\sum_{j=1}^{n}\boldsymbol{X}_{ij}w(\eta_{ij})\right\}^{T}\right].\end{aligned} \tag{7.16}$$

Using Lemma 1 from Chapter 2 with $f(\boldsymbol{X}_{ij}, \boldsymbol{U}_i) = w(\eta_{ij})$, we have that the first term in (7.16) can be re-expressed as follows

$$\frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n}\boldsymbol{X}_{ij}^{\otimes 2}E\{w(\eta_{ij})|\boldsymbol{X}_i\} = E\left\{\boldsymbol{X}\boldsymbol{X}^T w(\beta_0+\boldsymbol{\beta}_1+U)\right\} + o_P(1)\mathbf{1}_d^{\otimes 2}.$$

Now, using Lemma 2 from Chapter 2 with $f(\boldsymbol{X}_{ij}, \boldsymbol{U}_i) = w(\eta_{ij})$, we have that the second term in (7.16) can be re-expressed as follows

$$\begin{aligned}&E\left[\frac{1}{mn}\sum_{i=1}^{m}\left\{\sum_{j=1}^{n}\boldsymbol{X}_{ij}w(\eta_{ij})\right\}\left\{\sum_{j=1}^{n}w(\eta_{ij})\right\}^{-1}\left\{\sum_{j=1}^{n}\boldsymbol{X}_{ij}w(\eta_{ij})\right\}^{T}\right] \\ &= E\left(E\left\{\boldsymbol{X}w(\beta_0+\boldsymbol{\beta}_1+U)|U\right\}\left[E\left\{w(\beta_0+\boldsymbol{\beta}_1+U)|U\right\}\right]^{-1}E\left\{\boldsymbol{X}w(\beta_0+\boldsymbol{\beta}_1+U)|U\right\}^{T}\right) \\ &\quad + o_P(1)\mathbf{1}_d^{\otimes 2}.\end{aligned}$$

Now,

$$\boldsymbol{\Sigma}_{\boldsymbol{\beta}_1} = E\left[E\left\{\boldsymbol{X}\boldsymbol{X}^T w(\beta_0+\boldsymbol{\beta}_1+U)|U\right\} - \frac{E\left\{\boldsymbol{X}w(\beta_0+\boldsymbol{\beta}_1+U)|U\right\}^{\otimes 2}}{E\left\{w(\beta_0+\boldsymbol{\beta}_1+U)|U\right\}}\right] + o_P(1)\mathbf{1}_d^{\otimes 2}.$$

Let

$$\Omega^*_{\boldsymbol{\beta}_1}(U) \equiv E\left\{w(\beta_0+\boldsymbol{\beta}_1+U)\begin{bmatrix}1 & \boldsymbol{X}^T\\ \boldsymbol{X} & \boldsymbol{X}\boldsymbol{X}^T\end{bmatrix}\middle|U\right\}.$$

The inverse of the lower right $d\times d$ block of $\Omega_{\boldsymbol{\beta}_1}(U)^{-1}$ is

$$E\left\{\boldsymbol{X}\boldsymbol{X}^T w(\beta_0+\boldsymbol{\beta}_1+U)|U\right\} - \frac{E\left\{\boldsymbol{X}w(\beta_0+\boldsymbol{\beta}_1+U)|U\right\}^{\otimes 2}}{E\left\{w(\beta_0+\boldsymbol{\beta}_1+U)|U\right\}}.$$

Therefore, we conclude that that the $(2,2)$ block of $I(\beta_0, \boldsymbol{\beta}_1, \sigma^2)$ is as follows

$$\frac{mn(\Lambda^*_{\boldsymbol{\beta}_1})^{-1}}{\phi} + o_P(mn)\mathbf{1}_d^{\otimes 2}$$

where

$$\Lambda^*_{\boldsymbol{\beta}_1} \equiv \left(E\left[\left\{\text{lower right } d\times d \text{ block of } \Omega^*_{\boldsymbol{\beta}_1}(U)^{-1}\right\}^{-1}\right]\right)^{-1}.$$

Hence, the final Fisher information matrix can be expressed as

$$I(\beta_0, \boldsymbol{\beta}_1, \sigma^2) = \begin{bmatrix} \frac{m}{\sigma^2} + O_P(mn^{-1}) & O_P(m)\mathbf{1}_d^T & O_P(mn^{-1}) \\ O_P(m)\mathbf{1}_d & \frac{mn(\Lambda^*_{\boldsymbol{\beta}_1})^{-1}}{\phi} + o_P(mn)\mathbf{1}_d^{\otimes 2} & O_P(m)\mathbf{1}_d \\ O_P(mn^{-1}) & O_P(m)\mathbf{1}_d^T & \frac{m}{2\sigma^4} + O_P(mn^{-1}) \end{bmatrix}.$$

7.3.11 The Inverse of the Fisher Information Matrix

To invert the Fisher information matrix, we choose to work with the $(\beta_0, \sigma^2, \boldsymbol{\beta}_1)$ ordering instead of $(\beta_0, \boldsymbol{\beta}_1, \sigma^2)$. A trivial rearrangement of the matrix entries leads to

$$I(\beta_0, \boldsymbol{\beta}_1, \sigma^2) = \begin{bmatrix} \frac{m}{\sigma^2} + O_P(mn^{-1}) & O_P(mn^{-1}) & O_P(m)\mathbf{1}_d^T \\ O_P(mn^{-1}) & \frac{m}{2\sigma^4} + O_P(mn^{-1}) & O_P(m)\mathbf{1}_d^T \\ O_P(m)\mathbf{1}_d & O_P(m)\mathbf{1}_d & \frac{mn(\mathbf{\Lambda}^*_{\boldsymbol{\beta}_1})^{-1}}{\phi} + o_P(mn)\mathbf{1}_d^{\otimes 2} \end{bmatrix}.$$

Following similar steps to that in Subsubsection 4.5.2.10, we obtain the following expression for $I(\beta_0, \sigma^2, \boldsymbol{\beta}_1)^{-1}$ where

$$\begin{aligned} &I(\beta_0, \sigma^2, \boldsymbol{\beta}_1)^{-1} \\ &= \begin{bmatrix} \frac{\sigma^2}{m} + O_P(m^{-1}n^{-1}) & O_P(m^{-1}n^{-1}) & O_P(m^{-1}n^{-1})\mathbf{1}_d^T \\ O_P(m^{-1}n^{-1}) & \frac{2\sigma^4}{m} + O_P(m^{-1}n^{-1}) & O_P(m^{-1}n^{-1})\mathbf{1}_d^T \\ O_P(m^{-1}n^{-1})\mathbf{1}_d & O_P(m^{-1}n^{-1})\mathbf{1}_d & \frac{\phi\mathbf{\Lambda}^*_{\boldsymbol{\beta}_1}}{mn} + o_P(m^{-1}n^{-1})\mathbf{1}_d^{\otimes 2} \end{bmatrix}. \end{aligned}$$

The expression for the inverse of the Fisher information matrix can be also written as follows

$$I(\beta_0, \sigma^2, \boldsymbol{\beta}_1)^{-1} = I(\beta_0, \sigma^2, \boldsymbol{\beta}_1)^{-1} + \frac{1}{mn}\begin{bmatrix} O_P(1) & O_P(1) & O_P(1)\mathbf{1}_d^T \\ O_P(1) & O_P(1) & O_P(1)\mathbf{1}_d^T \\ O_P(1)\mathbf{1}_d & O_P(1)\mathbf{1}_d & o_P(1)\mathbf{1}_d^{\otimes 2} \end{bmatrix}.$$

where

$$I(\beta_0, \sigma^2, \boldsymbol{\beta}_1)^{-1} = \begin{bmatrix} \frac{\sigma^2}{m} & 0 & \mathbf{0} \\ 0 & \frac{2\sigma^4}{m} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \frac{\phi\mathbf{\Lambda}^*_{\boldsymbol{\beta}_1}}{mn} \end{bmatrix}.$$

7.3.12 Final Asymptotic Normality Result

The derivations leading to the final asymptotic normality result follows steps similar to those in Subsubsection 4.5.2.11j. It follows that

$$\sqrt{m}\begin{bmatrix} \widehat{\beta}_0 - \beta_0^0 \\ \sqrt{n}\left(\widehat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1^0\right) \\ \widehat{\sigma^2} - (\sigma^2) \end{bmatrix} \xrightarrow{\mathcal{D}} N\left(\begin{bmatrix} 0 \\ \mathbf{0} \\ 0 \end{bmatrix}, \begin{bmatrix} (\sigma^2)^0 & \mathbf{0} & 0 \\ \mathbf{0} & \phi\mathbf{\Lambda}^*_{\boldsymbol{\beta}_1} & \mathbf{0} \\ 0 & \mathbf{0} & 2\left\{(\sigma^2)^0\right\}^2 \end{bmatrix}\right).$$

Chapter 8

Discussion and Conclusion

The aim of this thesis was to address some of the current gaps present in the literature available for GLMMs. Specifically, we aimed to develop asymptotic theory for maximum likelihood and maximum quasi-likelihood estimators for GLMMs. Once these results were derived, we wished to assess the efficacy of the studentized confidence results constructed based upon these asymptotic normality results and also explore the implications of such results on optimal design theory. Lastly, with regards to methodology for GLMMs, we aimed to implement the Thouless-Anderson-Palmer approach and analyse the accuracy of the variational estimates obtained.

In order to carry out detailed asymptotic analysis on maximum likelihood estimators and maximum quasi-likelihood estimators for GLMMs, we first required results for two important tasks. The first of which required deriving population limits of particular predictor-dependent sample mean quantities. The second task involved establishing matrix norm asymptotic negligibility between matrix square roots of inverse Fisher information matrices and their simpler asymptotic block diagonal forms. Currently, there are no results available to deal with either of these tasks in a simple manner. Hence in Chapter 2, we provided the necessary tools and results in the forms of Lemmas 1, 2 and 3.

Next, in Chapter 3, we derived asymptotic normality results for GLMMs involving Gaussian responses. Existing theory for Gaussian linear mixed models by Wand (2002) and Harville (1977) provided base expressions for the required Fisher information matrix. Thereafter, the leading terms in the entries of the Fisher information matrix were retained, as the number of groups and the number of observations within each group diverged. (This approach was repeated throughout Chapters 4 and 7 in this thesis

as well). The resulting theorem as a result of our work concerns the joint asymptotic normality of all maximum likelihood estimators for a Gaussian response mixed model and elegantly shows faster rates of convergence of fixed effects unaccompanied by random effects as compared to fixed effects that have partnering random effects.

The results in Chapter 3 were extended to the class of all GLMMs, including a model extension for overdispersion, in Chapter 4. However, frequentist inference for GLMMs is hindered by the existence of intractable integrals due to the inclusion of random effects in these models. To overcome this obstacle, we used a multi-term Laplace's method expansion for ratios of intractable integrals (Miyata, 2004; Tierney et al., 1989). The resulting asymptotic normality theorem concerns the joint asymptotic normality of all of the maximum quasi-likelihood estimators, for fixed values of the dispersion parameter, for a generalized linear mixed model. Once again, the results derived in this chapter show faster rates of convergence for fixed effects that are not accompanied by random effects as compared to fixed effects accompanied by random effects. We also noted that for the class of two-parameter exponential families, maximum likelihood estimation is possible for all model parameters including the dispersion parameter. Based on this, we also derived asymptotic normality results for the maximum likelihood estimator for the dispersion parameter in the Gaussian and Gamma response cases.

Chapter 5 presents the consequences and applications of the asymptotic normality results derived in Chapter 4. First, we present how studentized confidence intervals can be constructed based on our asymptotic normality results in order to carry out asymptotically valid inference. The efficacy of the confidence intervals were then assessed, which showed that the Theorem 12-based approach in Section 4.3 as an attractive alternative to the exact observed Fisher information approach. The Theorem 12-based approach required simpler or no numerical integration at all compared to the exact observed Fisher information approach and gave similar coverage properties, especially for larger values of the number of groups and number of observations within each group. Next, we looked into the implications of Theorem 12 on optimal design theory. Most optimality criteria are based on the Fisher information matrix, which is computationally expensive to evaluate. Hence, we presented a simple approach to constructing approximate locally D-optimal designs based on large sample approximations of the Fisher information matrix.

In Chapter 6, we tackle the implementation of the Thouless-Anderson-Palmer approach for GLMMs. First, we presented a general result for deriving the Thouless-Anderson-Palmer approximate negative log-likelihood for GLMMs. This expression can then be locally minimized to obtain TAP estimates of the true model parameters. Then,

we carried out several simulation studies using a simple Poisson linear mixed model and analysed the Thouless-Anderson-Palmer variational estimates obtained against the estimates obtained from implementing Gaussian variational approximation. Based on the simulation studies, the Thouless-Anderson-Palmer enhancement approach suggests a slight yet statistically significant improvement as compared to using the Gaussian variational approximation approach, especially for small datasets.

Last but not least, in Chapter 7, we developed theory to consider the usage of noncanonical links for both the development of asymptotic normality results for maximum quasi-likelihood estimators for GLMMs and also the implementation of the Thouless-Anderson-Palmer approach for GLMMs.

In conclusion, this thesis presents important theoretical and methodological work that concerns the asymptotic distributions of maximum quasi-likelihood estimators for GLMMs and the implementation of the Thouless-Anderson-Palmer variational method for GLMMs. We believe that the work in this thesis will make a significant and novel contribution to the area of GLMMs, which have been a mainstay of regression-type statistical analyses in important areas such as longitudinal data analysis, multilevel modelling, panel data analysis and small area estimation.

Potential future work involves deriving second-order asymptotic approximations of the Fisher information matrix. In this thesis, we only retained the leading terms in the entries of the Fisher information matrix, hence leading to a first-order asymptotic approximation. Deriving second-order asymptotic approximations can give more accurate expressions for the asymptotic variance-covariance matrix for the maximum quasi-likelihood estimators for GLMMs, which is especially useful when implementing such results for smaller finite samples. For example, with second-order approximations, better coverages can be achieved when constructing studentized confidence intervals for smaller values for the number of groups and number of observations within each group as compared to those in the simulation study in Subsection 5.1.2. Deriving second-order approximations will also allow us to determine approximate locally D-optimal designs when considering multivariate random effects, which is not met by the theory presented in this thesis.

The techniques used in this thesis could also be used to study statistical models other than GLMMs. An example of a potential class of such models would be generalized additive mixed models. This class of models extends the GLMM framework by allowing for continuous predictors impacting the mean response to be modelled by additive non-parametric functions. This provides additional flexibility for modelling the actual

relationship between the response (specified by an exponential family distribution) and predictors, It also potentially provides better fits to the data as well. The techniques used in this thesis could be used to derive the first-order asymptotic approximations of the ratios of intractable integrals that arise when calculating the scores required for the Fisher information matrix, and subsequently the asymptotic variance-covariance matrix.

Last but not least, we can explore how the TAP approach performs against the GVA approach for GLMMs with response distributions other than the Poisson family.

References

Bates, D., Mächler, M., Bolker, B., and Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software, Articles*, 67:1–48.

Breslow, N. E. and Clayton, D. G. (1993). Approximate Inference in Generalized Linear Mixed Models. *Journal of the American Statistical Association*, 88:9–25.

Chipman, J. S. (1964). On Least Squares with Insufficient Observations. *Journal of the American Statistical Association*, 59(308):1078–1111.

Fan, J., Heckman, N. E., and Wand, M. P. (1995). Local Polynomial Kernel Regression for Generalized Linear Models and Quasi-Likelihood Functions. *Journal of the American Statistical Association*, 90:141–150.

Fan, Z., Mei, S., and Montanari, A. (2021). TAP Free Energy, Spin Glasses and Variational Inference. *The Annals of Probability*, 49:1 – 45.

Golub, G. and Van Loan, C. (2013). *Matrix Computations*. Johns Hopkins Studies in the Mathematical Sciences. Johns Hopkins University Press.

Hall, P., Pham, T., Wand, M. P., and Wang, S. S. J. (2011). Asymptotic Normality and Valid Inference for Gaussian Variational Approximation. *Annals of Statistics*, 39:2502–2532.

Harville, D. A. (1977). Maximum Likelihood Approaches to Variance Component Estimation and to Related Problems. *Journal of the American Statistical Association*, 72:320–338.

Higham, N. J. (2008). *Functions of Matrices: Theory and Computation*. Society for Industrial and Applied Mathematics.

Hsu, L. C. (1948). Approximations to a Class of Double Integrals of Functions of Large Numbers. *American Journal of Mathematics*, 70:698–708.

Jiang, J. (1996). REML Estimation: Asymptotic Behavior and Related Topics. *The Annals of Statistics*, 24:255–286.

Jiang, J. and Nguyen, T. (2021). *Linear and Generalized Linear Mixed Models and Their Applications, Second Edition*. New York:Springer.

Johnstone, I. (2022). Expectation Propagation in Mixed Models. Presentation in the conference "Statistics in the Big Data Era". Simons Institute, California, U.S.A, June 2022.

Knight, K. (2000). *Mathematical Statistics*. Chapman and Hall/CRC.

Lyu, Z. and Welsh, A. (2022). Increasing Cluster Size Asymptotics for Nested Error Regression Models. *Journal of Statistical Planning and Inference*, 217:52–68.

Maestrini, L., Bhaskaran, A., and Wand, M. P. (2023). Second term improvement to generalised linear mixed model asymptotics.

Magnus, J. and Neudecker, H. (1999). *Matrix Differential Calculus with Applications in Statistics and Econometrics*. John Wiley & Sons.

Magnus, J. R. and Neudecker, H. (1979). The Commutation Matrix: Some Properties and Applications. *The Annals of Statistics*, 7:381 – 394.

McCulloch, C., Searle, S., and Neuhaus, J. (2008). *Generalized, Linear, and Mixed Models, Second Edition.* New York: John Wiley & Sons.

Miller, J. J. (1973). Asymptotic properties of maximum likelihood estimates in the mixed model of the analysis of variance. *Technical Report No. 12, Department of Statistics*.

Miller, J. J. (1977). Asymptotic properties of maximum likelihood estimates in the mixed model of the analysis of variance. *The Annals of Statistics*, 5:746–762.

Miyata, Y. (2004). Fully Exponential Laplace Approximations Using Asymptotic Modes. *Journal of the American Statistical Association*, 99:1037–1049.

Nie, L. (2007). Convergence Rate of MLE in Generalized Linear and Nonlinear Mixed-Effects Models: Theory and Applications. *Journal of Statistical Planning and Inference*, 137:1787–1804.

Ormerod, J. T. and Wand, M. P. (2010). Explaining Variational Approximations. *The American Statistician*, 64:140–153.

Pace, L. and Salvan, A. (1997). *Principles of Statistical Inference from a Neo-Fisherian Perspective*. World Scientific.

Plefka, T. (1982). Convergence Condition of the TAP Equation for the Infinite-Ranged Ising Spin Glass Model. *Journal of Physics A: Mathematical and General*, 15:1971–1978.

R Core Team (2022). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Russell, K. (2018). *Design of Experiments for Generalized Linear Models*. Chapman and Hall/CRC.

Sherrington, D. and Kirkpatrick, S. (1975). Solvable model of a spin-glass. *Phys. Rev. Lett.*, 35:1792–1796.

Thouless, D. J., Anderson, P. W., and Palmer, R. G. (1977). Solution of 'Solvable Model of a Spin Glass'. *The Philosophical Magazine: A Journal of Theoretical Experimental and Applied Physics*, 35:593–601.

Tierney, L., Kass, R. E., and Kadane, J. B. (1989). Fully Exponential Laplace Approximations to Expectations and Variances of Nonpositive Functions. *Journal of the American Statistical Association*, 84:710–716.

van der Vaart, A. W. (1998). *Asymptotic Statistics*. Cambridge, U.K.: Cambridge University Press.

Waite, T. W. and Woods, D. C. (2015). Designs for Generalized Linear Models with Random Block Effects Via Information Matrix Approximations. *Biometrika*, 102:677–693.

Wand, M. (2007). Fisher Information for Generalised Linear Mixed Models. *Journal of Multivariate Analysis*, 98:1412–1416.

Wand, M. P. (2002). Vector Differential Calculus in Statistics. *The American Statistician*, 56:55–62.

Westfall, P. H. (1986). Asymptotic Normality of the Anova Estimates of Components of Variance in the Nonnormal, Unbalanced Hierarchal Mixed Model. *The Annals of Statistics*, 14:1572 – 1582.

Wolfram Research Inc. (2022). *Mathematica, Version 12.0*. Champaign, Illinois, U.S.A.

Zhang, W., Mandal, A., and Stufken, J. (2017). Approximations of the Information Matrix for a Panel Mixed Logit Model. *Journal of Statistical Theory and Practice*, 11:269–295.