Supplement for:

Dispersion Parameter Extension of Generalized Linear Mixed Model Asymptotics

AISHWARYA BHASKARAN AND MATT P. WAND

University of Technology Sydney

S.1 Introduction

This supplement contains derivational details concerning the article's results. In Section S.2 we provide a proof of Theorem 1. Section S.3 provides illustration of how, for a special case, the (A3) moment condition can be reduced to a simpler moment condition.

S.2 Proof of Theorem 1

In this section we provide a proof of Theorem 1. We start by setting up notation for some key quantities that arise throughout the proof, as well as useful generic mathematical notation. The main body of the proof involves asymptotic approximation of the Fisher information matrix and its inverse. Much of this was achieved in Appendix A of Jiang *et al.* (2022). However, the dispersion parameter extension leads to new Fisher information entries. We then apply Lemma 2 of Jiang *et al.* (2022) to establish an asymptotic equivalence result between the matrix square roots of two relevant approximations to the inverse Fisher information matrix. The final steps required to establish Theorem 1 are then carried out.

S.2.1 Notation

We divide the notation into two parts: (1) that for key quantities specific to the model at hand and (2) generic mathematical notation that aids the proof.

S.2.1.1 Notation for Some Key Quantities

Throughout this proof we let

$$\psi \equiv 1/\phi$$

denote the *reciprocal* dispersion parameter. Working with ψ, rather than ϕ, in Fisher information approximations involves simpler expressions in the derivation of the asymptotic joint normality result for the model parameters. The transformation from ψ to $\phi = 1/\psi$, using the Multivariate Delta Method, is carried out after such a result is established.

For each $1 \leq i \leq m$ and $1 \leq j \leq n_i$ let $\boldsymbol{X}_{ij} \equiv (\boldsymbol{X}_{\mathrm{A}ij}^T, \boldsymbol{X}_{\mathrm{B}ij}^T)^T$ and $\boldsymbol{X}_i \equiv (\boldsymbol{X}_{i1}, \ldots, \boldsymbol{X}_{in_i})$. Let $\boldsymbol{Y}_i$, $1 \leq i \leq m$, be defined analogously.

Define $\mathcal{G}_{\mathrm{A}i}$ and $\mathcal{H}_{\mathrm{AA}i}$, for each $1 \leq i \leq m$, as follows

$$\mathcal{G}_{\mathrm{A}i} \equiv \sum_{j=1}^{n_i} \{Y_{ij} - b'\big((\boldsymbol{\beta}_{\mathrm{A}} + \boldsymbol{U}_i)^T \boldsymbol{X}_{\mathrm{A}ij} + \boldsymbol{\beta}_{\mathrm{B}}^T \boldsymbol{X}_{\mathrm{B}ij}\big)\} \boldsymbol{X}_{\mathrm{A}ij},$$

$$\mathcal{H}_{\mathrm{AA}i} \equiv \sum_{j=1}^{n_i} b''\big((\boldsymbol{\beta}_{\mathrm{A}} + \boldsymbol{U}_i)^T \boldsymbol{X}_{\mathrm{A}ij} + \boldsymbol{\beta}_{\mathrm{B}}^T \boldsymbol{X}_{\mathrm{B}ij}\big) \boldsymbol{X}_{\mathrm{A}ij} \boldsymbol{X}_{\mathrm{A}ij}^T,$$

In a similar vein, define $\mathcal{H}'_{\text{AAA}i}$ to be the $d_{\text{A}} \times d_{\text{A}} \times d_{\text{A}}$ array with (r,s,t) entry equal to

$$\sum_{j=1}^{n_i} b'''\big((\boldsymbol{\beta}_{\text{A}} + \boldsymbol{U}_i)^T \boldsymbol{X}_{\text{A}ij} + \boldsymbol{\beta}_{\text{B}}^T \boldsymbol{X}_{\text{B}ij}\big)(\boldsymbol{X}_{\text{A}ij})_r (\boldsymbol{X}_{\text{A}ij})_s (\boldsymbol{X}_{\text{A}ij})_t$$

S.2.1.2 Generic Mathematical Notation

For a generic $d \times 1$ vector $\boldsymbol{v}$ we define $\boldsymbol{v}^{\otimes 2} \equiv \boldsymbol{v}\boldsymbol{v}^T$. We also let diag$(\boldsymbol{v})$ denote the $d \times d$ diagonal matrix with the entries of $\boldsymbol{v}$ along the diagonal. For a matrix $\boldsymbol{M}$ let $\|\boldsymbol{M}\|_F = \{\text{tr}(\boldsymbol{M}^T\boldsymbol{M})\}^{1/2}$ denote its Frobenius norm.

For f a smooth real-valued function of the d-variate argument $\boldsymbol{x} \equiv (x_1, \ldots, x_d)$, let $\nabla f(\boldsymbol{x})$ denote the $d \times 1$ vector with ith entry $\partial f(\boldsymbol{x})/\partial x_i$, $\nabla^2 f(\boldsymbol{x})$ denote the $d \times d$ matrix with (i,j) entry $\partial^2 f(\boldsymbol{x})/(\partial x_i \partial x_j)$ and $\nabla^3 f(\boldsymbol{x})$ denote the $d \times d \times d$ array with (i,j,k) entry $\partial^3 f(\boldsymbol{x})/(\partial x_i \partial x_j \partial x_k)$.

If $\boldsymbol{\mathcal{A}}$ is a $d_1 \times d_2 \times d_3$ array and $\boldsymbol{M}$ is a $d_1 \times d_2$ matrix then we let

$$\boldsymbol{\mathcal{A}} \bigstar \boldsymbol{M} \quad \text{denote the } d_3 \times 1 \text{ vector with } t\text{th entry given by} \quad \sum_{r=1}^{d_1}\sum_{s=1}^{d_2} (\boldsymbol{\mathcal{A}})_{rst} \boldsymbol{M}_{rs}.$$

S.2.2 Fisher Information Approximation

The Fisher information corresponding to the parameter vector

$$(\boldsymbol{\beta}_{\text{A}}, \boldsymbol{\beta}_{\text{B}}, \text{vech}(\boldsymbol{\Sigma}), \psi) \tag{S.1}$$

is a symmetric matrix having $d_{\text{A}} + d_{\text{B}} + \frac{1}{2} d_{\text{A}}(d_{\text{A}} + 1) + 1$ rows and columns. Appendix A of Jiang *et al.* (2022) provides an adequate approximation of the $\big(\boldsymbol{\beta}_{\text{A}}, \boldsymbol{\beta}_{\text{B}}, \text{vech}(\boldsymbol{\Sigma})\big)$ diagonal block. The dispersion parameter extension requires similar approximations for the ψ diagonal block and the $\big(\boldsymbol{\beta}_{\text{A}}, \boldsymbol{\beta}_{\text{B}}, \text{vech}(\boldsymbol{\Sigma}), \psi\big)$ off-diagonal block. The sub-blocks of the $\big(\boldsymbol{\beta}_{\text{A}}, \boldsymbol{\beta}_{\text{B}}, \text{vech}(\boldsymbol{\Sigma}), \psi\big)$ off-diagonal block correspond to each of

$$\big(\boldsymbol{\beta}_{\text{A}}, \boldsymbol{\beta}_{\text{B}}, \text{vech}(\boldsymbol{\Sigma})\big), \quad \big(\boldsymbol{\beta}_{\text{A}}, \boldsymbol{\beta}_{\text{B}}, \psi\big) \quad \text{and} \quad \big(\text{vech}(\boldsymbol{\Sigma}), \psi\big)$$

The first of these is treated in Jiang *et al.* (2022). The second and third of these are treated in Sections S.2.2.5 and S.2.2.6 respectively.

S.2.2.1 Higher Order Approximation of Multivariate Integral Ratios

Our main tool for approximation of the Fisher information matrix entries for generalized linear mixed models is higher order Laplace-type approximation of multivariate integral ratios. Appendix A of Miyata (2004) provides such a result, which states that for smooth real-valued d-variate functions $\mathfrak{g}$, $\mathfrak{c}$ and $\mathfrak{h}$ we have

$$\begin{aligned} \frac{\int_{\mathbb{R}^d} \mathfrak{g}(\boldsymbol{x})\mathfrak{c}(\boldsymbol{x}) \exp\{-n\mathfrak{h}(\boldsymbol{x})\}\, d\boldsymbol{x}}{\int_{\mathbb{R}^d} \mathfrak{c}(\boldsymbol{x}) \exp\{-n\mathfrak{h}(\boldsymbol{x})\}\, d\boldsymbol{x}} &= \mathfrak{g}(\boldsymbol{x}^*) + \frac{\nabla \mathfrak{g}(\boldsymbol{x}^*)^T \{\nabla^2 \mathfrak{h}(\boldsymbol{x}^*)\}^{-1} \nabla \mathfrak{c}(\boldsymbol{x}^*)}{n\mathfrak{c}(\boldsymbol{x}^*)} \\ &\quad + \frac{\text{tr}[\{\nabla^2 \mathfrak{h}(\boldsymbol{x}^*)\}^{-1} \nabla^2 \mathfrak{g}(\boldsymbol{x}^*)]}{2n} - \frac{\nabla \mathfrak{g}(\boldsymbol{x}^*)^T \Big[\nabla^3 \mathfrak{h}(\boldsymbol{x}^*) \bigstar \{\nabla^2 \mathfrak{h}(\boldsymbol{x}^*)\}^{-1}\Big]}{2n} + O(n^{-2}) \end{aligned} \tag{S.2}$$

where

$$\boldsymbol{x}^* \equiv \underset{\boldsymbol{x} \in \mathbb{R}^d}{\text{argmin}}\, \mathfrak{h}(\boldsymbol{x}).$$

S.2.2.2 The U_i^* Quantity and Its Approximation

For all required Fisher information approximations for the ith group, the h function appearing in (S.2) corresponds to the stochastic function

$$\mathfrak{h}_i(\boldsymbol{u}) \equiv -\frac{\psi}{n}\sum_{j=1}^{n_i}\left\{Y_{ij}\boldsymbol{u}^T\boldsymbol{X}_{\mathrm{A}ij} - b\big((\boldsymbol{\beta}_\mathrm{A}+\boldsymbol{u})^T\boldsymbol{X}_{\mathrm{A}ij}+\boldsymbol{\beta}_\mathrm{B}^T\boldsymbol{X}_{\mathrm{B}ij}\big)\right\} \tag{S.3}$$

and its minimum is denoted by the random vector

$$\boldsymbol{U}_i^* \equiv \underset{\boldsymbol{u}\in\mathbb{R}^d}{\mathrm{argmin}}\,\mathfrak{h}_i(\boldsymbol{u}).$$

Taylor series expansion, similar to that given in Appendix A.3.1 of Jiang *et al.* (2022), followed by asymptotic series inversion leads to the three-term approximation:

$$\boldsymbol{U}_i^* = \boldsymbol{U}_i + \mathcal{H}_{\mathrm{AA}i}^{-1}\mathcal{G}_{\mathrm{A}i} - \tfrac{1}{2}\mathcal{H}_{\mathrm{AA}i}^{-1}\Big\{\mathcal{H}'_{\mathrm{AAA}i}\bigstar\big(\mathcal{H}_{\mathrm{AA}i}^{-1}\mathcal{G}_{\mathrm{A}i}\mathcal{G}_{\mathrm{A}i}^T\mathcal{H}_{\mathrm{AA}i}^{-1}\big)\Big\} + O_P(n^{-3/2})\mathbf{1}_{d_\mathrm{A}}. \tag{S.4}$$

S.2.2.3 The $\big(\boldsymbol{\beta}_\mathbf{A}, \boldsymbol{\beta}_\mathbf{B}, \mathrm{vech}(\boldsymbol{\Sigma})\big)$ Diagonal Block

For the case of the dispersion parameter being fixed rather than estimated, Jiang *et al.* (2022) derives an approximation to the Fisher information of $\big(\boldsymbol{\beta}_\mathrm{A}, \boldsymbol{\beta}_\mathrm{B}, \mathrm{vech}(\boldsymbol{\Sigma})\big)$. Appendix A.5 provides the resultant approximation. For the extension to $\phi = 1/\psi$ being estimated, this approximation corresponds to the $\big(\boldsymbol{\beta}_\mathrm{A}, \boldsymbol{\beta}_\mathrm{B}, \mathrm{vech}(\boldsymbol{\Sigma})\big)$ diagonal block of the Fisher information matrix for the extended parameter vector (S.1).

S.2.2.4 The ψ Diagonal Block

The ith contribution to the score is

$$\frac{\partial\log\{p_{\boldsymbol{Y}_i|\boldsymbol{X}_i}(\boldsymbol{Y}_i|\boldsymbol{X}_i)\}}{\partial\psi} = \sum_{j=1}^{n_i}\left[\Big\{Y_{ij}\big((\boldsymbol{\beta}_\mathrm{A})^T\boldsymbol{X}_{\mathrm{A}ij}+(\boldsymbol{\beta}_\mathrm{B})^T\boldsymbol{X}_{\mathrm{B}ij}\big)+\mathfrak{c}(Y_{ij})\Big\} - \frac{dd(1/\psi)}{d\psi}\right] + \frac{\displaystyle\int_{\mathbb{R}^{d_\mathrm{A}}}\mathfrak{g}_i(\boldsymbol{u})\mathfrak{c}(\boldsymbol{u})\exp\{\psi\mathfrak{g}_i(\boldsymbol{u})\}\,d\boldsymbol{u}}{\displaystyle\int_{\mathbb{R}^{d_\mathrm{A}}}\mathfrak{c}(\boldsymbol{u})\exp\{\psi\mathfrak{g}_i(\boldsymbol{u})\}\,d\boldsymbol{u}}.$$

The derivative of the ith contribution to the score is

$$\frac{\partial^2\log\{p_{\boldsymbol{Y}_i|\boldsymbol{X}_i}(\boldsymbol{Y}_i|\boldsymbol{X}_i)\}}{\partial\psi^2} = -\frac{n_i\,d^2d\big(1/\psi\big)}{d\psi^2} + \mathcal{Q}_{1i} - \mathcal{Q}_{2i}^2 \tag{S.5}$$

where

$$\mathcal{Q}_{1i} \equiv \frac{\displaystyle\int_{\mathbb{R}^{d_\mathrm{A}}}\mathfrak{g}_i^2(\boldsymbol{u})\mathfrak{c}(\boldsymbol{u})\exp\{-n\mathfrak{h}_i(\boldsymbol{u})\}\,d\boldsymbol{u}}{\displaystyle\int_{\mathbb{R}^{d_\mathrm{A}}}\mathfrak{c}(\boldsymbol{u})\exp\{-n\mathfrak{h}_i(\boldsymbol{u})\}\,d\boldsymbol{u}} \tag{S.6}$$

and

$$\mathcal{Q}_{2i} \equiv \frac{\displaystyle\int_{\mathbb{R}^{d_\mathrm{A}}}\mathfrak{g}_i(\boldsymbol{u})\mathfrak{c}(\boldsymbol{u})\exp\{-n\mathfrak{h}_i(\boldsymbol{u})\}\,d\boldsymbol{u}}{\displaystyle\int_{\mathbb{R}^{d_\mathrm{A}}}\mathfrak{c}(\boldsymbol{u})\exp\{-n\mathfrak{h}_i(\boldsymbol{u})\}\,d\boldsymbol{u}} \tag{S.7}$$

with

$$\mathfrak{c}(\boldsymbol{u}) \equiv \exp\left(-\tfrac{1}{2}\boldsymbol{u}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{u}\right), \quad \mathfrak{g}_i(\boldsymbol{u}) \equiv \sum_{j=1}^{n_i}\left\{Y_{ij}\boldsymbol{u}^T\boldsymbol{X}_{\mathrm{A}ij} - b\big((\boldsymbol{\beta}_\mathrm{A}+\boldsymbol{u})^T\boldsymbol{X}_{\mathrm{A}ij}+\boldsymbol{\beta}_\mathrm{B}^T\boldsymbol{X}_{\mathrm{B}ij}\big)\right\}$$

and $\mathfrak{h}_i$ is as given by (S.3). Application of (S.2) to each of (S.6) and (S.7) and use of (S.4) leads to the following three-term approximations to $\mathcal{Q}_{1i}$ and $\mathcal{Q}_{2i}$:

$$\mathcal{Q}_{1i} = \mathfrak{g}_i(\boldsymbol{U}_i)^2 + \mathcal{G}_{\mathrm{A}i}^T \mathcal{H}_{\mathrm{AA}i}^{-1} \mathcal{G}_{\mathrm{A}i}\, \mathfrak{g}_i(\boldsymbol{U}_i) - \frac{d_{\mathrm{A}}\, \mathfrak{g}_i(\boldsymbol{U}_i)}{\psi} + O_P(n^{1/2})$$

and

$$\mathcal{Q}_{2i} = \mathfrak{g}_i(\boldsymbol{U}_i) + \tfrac{1}{2}\mathcal{G}_{\mathrm{A}i}^T \mathcal{H}_{\mathrm{AA}i}^{-1} \mathcal{G}_{\mathrm{A}i} - \frac{d_{\mathrm{A}}}{2\psi} + O_P(n^{-1/2}).$$

Therefore

$$\begin{aligned}
\mathcal{Q}_{1i} - \mathcal{Q}_{2i}^2 &= \mathfrak{g}_i(\boldsymbol{U}_i)^2 + \mathcal{G}_{\mathrm{A}i}^T \mathcal{H}_{\mathrm{AA}i}^{-1} \mathcal{G}_{\mathrm{A}i}\, \mathfrak{g}_i(\boldsymbol{U}_i) - \frac{d_{\mathrm{A}} \mathfrak{g}_i(\boldsymbol{U}_i)}{\psi} + O_P(n^{1/2}) \\
&\quad - \left\{ \mathfrak{g}_i(\boldsymbol{U}_i) + \tfrac{1}{2}\mathcal{G}_{\mathrm{A}i}^T \mathcal{H}_{\mathrm{AA}i}^{-1} \mathcal{G}_{\mathrm{A}i} - \frac{d_{\mathrm{A}}}{2\psi} + O_P(n^{-1/2}) \right\}^2 \\
&= \mathfrak{g}_i(\boldsymbol{U}_i)^2 + \mathcal{G}_{\mathrm{A}i}^T \mathcal{H}_{\mathrm{AA}i}^{-1} \mathcal{G}_{\mathrm{A}i}\, \mathfrak{g}_i(\boldsymbol{U}_i) - \frac{d_{\mathrm{A}} \mathfrak{g}_i(\boldsymbol{U}_i)}{\psi} + O_P(n^{1/2}) \\
&\quad - \left\{ \mathfrak{g}_i(\boldsymbol{U}_i)^2 + \mathcal{G}_{\mathrm{A}i}^T \mathcal{H}_{\mathrm{AA}i}^{-1} \mathcal{G}_{\mathrm{A}i}\, \mathfrak{g}_i(\boldsymbol{U}_i) - \frac{d_{\mathrm{A}} \mathfrak{g}_i(\boldsymbol{U}_i)}{\psi} \right\} + O_P(n^{1/2}) \\
&= O_P(n^{1/2})
\end{aligned}$$

and we arrive at the approximation

$$\frac{\partial^2 \log\{p_{\boldsymbol{Y}_i|\boldsymbol{X}_i}(\boldsymbol{Y}_i|\boldsymbol{X}_i)\}}{\partial \psi^2} = -\frac{n_i\, d^2 d(1/\psi)}{d\psi^2} + O_P(n^{1/2}).$$

Therefore, the ψ diagonal block of the Fisher information is

$$mn\left(\frac{d^2 d(1/\psi)}{d\psi^2}\right) + O_P(mn^{1/2}).$$

S.2.2.5 The $(\boldsymbol{\beta}_{\mathbf{A}}, \boldsymbol{\beta}_{\mathbf{B}}, \psi)$ Off-Diagonal Block

Let

$$\boldsymbol{\beta} \equiv \begin{bmatrix} \boldsymbol{\beta}_{\mathrm{A}} \\ \boldsymbol{\beta}_{\mathrm{B}} \end{bmatrix}$$

denote the full fixed effects vector. As established in the appendix of Wand (2007), the ith contribution to the partial derivative, with respect to ψ, of the $\boldsymbol{\beta}$ score is

$$\frac{\partial \nabla_{\boldsymbol{\beta}} \log\{p_{\boldsymbol{Y}_i|\boldsymbol{X}_i}(\boldsymbol{Y}_i|\boldsymbol{X}_i)\}}{\partial \psi} = \boldsymbol{X}_i^T[\boldsymbol{Y}_i - E\{E(\boldsymbol{Y}_i|\boldsymbol{U}_i)|\boldsymbol{Y}_i\}].$$

Noting that

$$\begin{aligned}
E\big[\boldsymbol{X}_i^T\{\boldsymbol{Y}_i - E\{E(\boldsymbol{Y}_i|\boldsymbol{U}_i)|\boldsymbol{Y}_i\}|\boldsymbol{X}_i\big] &= \boldsymbol{X}_i^T\Big(E(\boldsymbol{Y}_i) - E[E\{E(\boldsymbol{Y}_i|\boldsymbol{U}_i)|\boldsymbol{Y}_i\}]\Big) \\
&= \boldsymbol{X}_i^T\big\{E(\boldsymbol{Y}_i) - E(\boldsymbol{Y}_i)\big\} = \boldsymbol{O}
\end{aligned}$$

it is apparent that the $(\boldsymbol{\beta}_{\mathrm{A}}, \boldsymbol{\beta}_{\mathrm{B}}, \psi)$ off-diagonal block of the Fisher information has all entries being exactly zero.

S.2.2.6 The $\big(\text{vech}(\boldsymbol{\Sigma}), \psi\big)$ Off-Diagonal Block

The ith contribution to the second order partial derivative with respect to vech($\boldsymbol{\Sigma}$) and ψ is

$$\begin{aligned}\nabla_{\text{vech}(\boldsymbol{\Sigma})}\frac{\partial \log\{p_{\boldsymbol{Y}_i|\boldsymbol{X}_i}(\boldsymbol{Y}_i|\boldsymbol{X}_i)\}}{\partial\psi} &= \nabla_{\text{vech}(\boldsymbol{\Sigma})}\left[\frac{\displaystyle\int_{\mathbb{R}^{d_{\text{A}}}}\mathfrak{g}_i(\boldsymbol{u})\mathfrak{c}(\boldsymbol{u})\exp\{-n\mathfrak{h}_i(\boldsymbol{u})\}\,d\boldsymbol{u}}{\displaystyle\int_{\mathbb{R}^{d_{\text{A}}}}\mathfrak{c}(\boldsymbol{u})\exp\{-n\mathfrak{h}_i(\boldsymbol{u})\}\,d\boldsymbol{u}}\right] \\ &= \tfrac{1}{2}\Big(\boldsymbol{\mathcal{Q}}_{3i} - \mathcal{Q}_{2i}\boldsymbol{\mathcal{Q}}_{4i}\Big)\end{aligned} \tag{S.8}$$

where

$$\boldsymbol{\mathcal{Q}}_{3i} \equiv \frac{\displaystyle\int_{\mathbb{R}^{d_{\text{A}}}}\mathfrak{g}_i(\boldsymbol{u})\boldsymbol{D}_{d_{\text{A}}}^T\text{vec}(\boldsymbol{\Sigma}^{-1}\boldsymbol{u}\boldsymbol{u}^T\boldsymbol{\Sigma}^{-1})\mathfrak{c}(\boldsymbol{u})\exp\{-n\mathfrak{h}_i(\boldsymbol{u})\}\,d\boldsymbol{u}}{\displaystyle\int_{\mathbb{R}^{d_{\text{A}}}}\mathfrak{c}(\boldsymbol{u})\exp\{-n\mathfrak{h}_i(\boldsymbol{u})\}\,d\boldsymbol{u}} \tag{S.9}$$

and

$$\boldsymbol{\mathcal{Q}}_{4i} \equiv \frac{\displaystyle\int_{\mathbb{R}^{d_{\text{A}}}}\boldsymbol{D}_{d_{\text{A}}}^T\text{vec}(\boldsymbol{\Sigma}^{-1}\boldsymbol{u}\boldsymbol{u}^T\boldsymbol{\Sigma}^{-1})\mathfrak{c}(\boldsymbol{u})\exp\{-n\mathfrak{h}_i(\boldsymbol{u})\}\,d\boldsymbol{u}}{\displaystyle\int_{\mathbb{R}^{d_{\text{A}}}}\mathfrak{c}(\boldsymbol{u})\exp\{-n\mathfrak{h}_i(\boldsymbol{u})\}\,d\boldsymbol{u}}. \tag{S.10}$$

Application of (S.2) to each of (S.9) and (S.10) and use of (S.4) leads to the following approximations to $\boldsymbol{\mathcal{Q}}_{3i}$ and $\boldsymbol{\mathcal{Q}}_{4i}$:

$$\boldsymbol{\mathcal{Q}}_{3i} = \boldsymbol{D}_{d_{\text{A}}}^T\text{vec}\Big(\boldsymbol{\Sigma}^{-1}(\boldsymbol{U}_i\boldsymbol{U}_i^T + 2\mathcal{H}_{\text{AA}i}^{-1}\mathcal{G}_{\text{A}i}\boldsymbol{U}_i^T)\boldsymbol{\Sigma}^{-1}\Big)\mathfrak{g}_i(\boldsymbol{U}_i) + O_P(1)\boldsymbol{1}_{d_{\text{A}}^{\boxplus}}$$

and

$$\boldsymbol{\mathcal{Q}}_{4i} = \boldsymbol{D}_{d_{\text{A}}}^T\text{vec}\Big(\boldsymbol{\Sigma}^{-1}(\boldsymbol{U}_i\boldsymbol{U}_i^T + 2\mathcal{H}_{\text{AA}i}^{-1}\mathcal{G}_{\text{A}i}\boldsymbol{U}_i^T)\boldsymbol{\Sigma}^{-1}\Big) + O_P(n^{-1})\boldsymbol{1}_{d_{\text{A}}^{\boxplus}}$$

where $d_{\text{A}}^{\boxplus} \equiv d_{\text{A}}(d_{\text{A}}+1)/2$. Substitution of these approximations into (S.8) then gives

$$\begin{aligned}\nabla_{\text{vech}(\boldsymbol{\Sigma})}\frac{\partial \log\{p_{\boldsymbol{Y}_i|\boldsymbol{X}_i}(\boldsymbol{Y}_i|\boldsymbol{X}_i)\}}{\partial\psi} &= \tfrac{1}{2}\boldsymbol{D}_{d_{\text{A}}}^T\text{vec}\Big(\boldsymbol{\Sigma}^{-1}(\boldsymbol{U}_i\boldsymbol{U}_i^T + 2\mathcal{H}_{\text{AA}i}^{-1}\mathcal{G}_{\text{A}i}\boldsymbol{U}_i^T)\boldsymbol{\Sigma}^{-1}\Big)\mathfrak{g}_i(\boldsymbol{U}_i) \\ &\quad -\tfrac{1}{2}\Big(\mathfrak{g}_i(\boldsymbol{U}_i) + O_P(1)\Big)\Big\{\boldsymbol{D}_{d_{\text{A}}}^T\text{vec}\Big(\boldsymbol{\Sigma}^{-1}(\boldsymbol{U}_i\boldsymbol{U}_i^T \\ &\qquad +2\mathcal{H}_{\text{AA}i}^{-1}\mathcal{G}_{\text{A}i}\boldsymbol{U}_i^T)\boldsymbol{\Sigma}^{-1}\Big) + O_P(n^{-1})\boldsymbol{1}_{d_{\text{A}}^{\boxplus}}\Big\} \\ &= O_P(1)\boldsymbol{1}_{d_{\text{A}}^{\boxplus}}.\end{aligned}$$

It follows that the $\big(\text{vech}(\boldsymbol{\Sigma}), \psi\big)$ off-diagonal block of the Fisher information matrix is

$$-E\left[\sum_{i=1}^m \nabla_{\text{vech}(\boldsymbol{\Sigma})}\frac{\partial \log\{p_{\boldsymbol{Y}_i|\boldsymbol{X}_i}(\boldsymbol{Y}_i|\boldsymbol{X}_i)\}}{\partial\psi}\right] = O_P(m)\boldsymbol{1}_{d_{\text{A}}^{\boxplus}}.$$

S.2.2.7 Assembly of Fisher Information Sub-Block Approximations

From the Fisher information sub-block approximations obtained in the previous six subsubsections we have

$$
I\Big(\boldsymbol{\beta}_{\rm A},\boldsymbol{\beta}_{\rm B},{\rm vech}(\boldsymbol{\Sigma}),\psi\Big)
$$

$$
=\begin{bmatrix}
m\boldsymbol{\Sigma}^{-1}+O_P(mn^{-1})\mathbf{1}_{d_{\rm A}}^{\otimes 2} & O_P(m)\mathbf{1}_{d_{\rm A}}\mathbf{1}_{d_{\rm B}}^T & O_P(mn^{-1})\mathbf{1}_{d_{\rm A}}\mathbf{1}_{d_{\rm A}^{\boxplus}}^T & \mathbf{0}_{d_{\rm A}} \\
O_P(m)\mathbf{1}_{d_{\rm B}}\mathbf{1}_{d_{\rm A}}^T & \dfrac{mn\boldsymbol{\Lambda}_{\boldsymbol{\beta}_{\rm B}}^{-1}}{\phi}+o_P(mn)\mathbf{1}_{d_{\rm B}}^{\otimes 2} & O_P(m)\mathbf{1}_{d_{\rm B}}\mathbf{1}_{d_{\rm A}^{\boxplus}}^T & \mathbf{0}_{d_{\rm B}} \\
O_P(mn^{-1})\mathbf{1}_{d_{\rm A}^{\boxplus}}\mathbf{1}_{d_{\rm A}}^T & O_P(m)\mathbf{1}_{d_{\rm A}^{\boxplus}}\mathbf{1}_{d_{\rm B}}^T & \begin{array}{c}\dfrac{m\boldsymbol{D}_{d_{\rm A}}^T(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1})\boldsymbol{D}_{d_{\rm A}}}{2}\\ +O_P(mn^{-1})\mathbf{1}_{d_{\rm A}^{\boxplus}}^{\otimes 2}\end{array} & O_P(m)\mathbf{1}_{d_{\rm A}^{\boxplus}} \\
\mathbf{0}_{d_{\rm A}}^T & \mathbf{0}_{d_{\rm B}}^T & O_P(m)\mathbf{1}_{d_{\rm A}^{\boxplus}}^T & \begin{array}{c} mn\left(\dfrac{d^2d(1/\psi)}{d\psi^2}\right)\\ +O_P(mn^{1/2})\end{array}
\end{bmatrix}.
$$

S.2.3 Inverse Fisher Information Approximation

Note that

$$
I\Big(\boldsymbol{\beta}_{\rm A},\boldsymbol{\beta}_{\rm B},{\rm vech}(\boldsymbol{\Sigma}),\psi\Big)=\begin{bmatrix}\boldsymbol{A}_{11} & \boldsymbol{A}_{12}\\ \boldsymbol{A}_{12}^T & A_{22}\end{bmatrix}
$$

where

$$
\boldsymbol{A}_{11}\equiv\begin{bmatrix}
m\boldsymbol{\Sigma}^{-1}+O_P(mn^{-1})\mathbf{1}_{d_{\rm A}}^{\otimes 2} & O_P(m)\mathbf{1}_{d_{\rm A}}\mathbf{1}_{d_{\rm B}}^T & O_P(mn^{-1})\mathbf{1}_{d_{\rm A}}\mathbf{1}_{d_{\rm A}^{\boxplus}}^T \\
O_P(m)\mathbf{1}_{d_{\rm B}}\mathbf{1}_{d_{\rm A}}^T & \dfrac{mn\boldsymbol{\Lambda}_{\boldsymbol{\beta}_{\rm B}}^{-1}}{\phi}+o_P(mn)\mathbf{1}_{d_{\rm B}}^{\otimes 2} & O_P(m)\mathbf{1}_{d_{\rm B}}\mathbf{1}_{d_{\rm A}^{\boxplus}}^T \\
O_P(mn^{-1})\mathbf{1}_{d_{\rm A}^{\boxplus}}\mathbf{1}_{d_{\rm A}}^T & O_P(m)\mathbf{1}_{d_{\rm A}^{\boxplus}}\mathbf{1}_{d_{\rm B}}^T & \begin{array}{c}\dfrac{m\boldsymbol{D}_{d_{\rm A}}^T(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1})\boldsymbol{D}_{d_{\rm A}}}{2}\\ +O_P(mn^{-1})\mathbf{1}_{d_{\rm A}^{\boxplus}}^{\otimes 2}\end{array}
\end{bmatrix}, \qquad \text{(S.11)}
$$

$$
A_{22}\equiv mnE\left\{\frac{d^2d(\frac{1}{\psi})}{d\psi^2}\right\}+O_P(mn^{1/2})\quad\text{and}\quad \boldsymbol{A}_{12}\equiv\big[\mathbf{0}_{d_{\rm A}}^T\;\;\mathbf{0}_{d_{\rm B}}^T\;\;O_P(m)\mathbf{1}_{d_{\rm A}^{\boxplus}}^T\big]^T.
$$

Let

$$
I\Big(\boldsymbol{\beta}_{\rm A},\boldsymbol{\beta}_{\rm B},{\rm vech}(\boldsymbol{\Sigma}),\psi\Big)^{-1}=\begin{bmatrix}\boldsymbol{A}^{11} & \boldsymbol{A}^{12}\\ (\boldsymbol{A}^{12})^T & A^{22}\end{bmatrix}.
$$

The upper left block of $I\Big(\boldsymbol{\beta}_{\rm A},\boldsymbol{\beta}_{\rm B},{\rm vech}(\boldsymbol{\Sigma}),\psi\Big)^{-1}$ is

$$
\boldsymbol{A}^{11}=\boldsymbol{A}_{11}^{-1}+\frac{\boldsymbol{A}_{11}^{-1}\boldsymbol{A}_{12}\boldsymbol{A}_{12}^T\boldsymbol{A}_{11}^{-1}}{A_{22}-\boldsymbol{A}_{12}^T\boldsymbol{A}_{11}^{-1}\boldsymbol{A}_{12}}. \qquad \text{(S.12)}
$$

The leading terms of $\boldsymbol{A}_{11}^{-1}$ are provided by equation (A27) of Jiang *et al.* (2022). Since

$$
\boldsymbol{A}_{12}^T\boldsymbol{A}_{11}^{-1}\boldsymbol{A}_{12}=O_P(m)
$$

we have

$$A_{22} - \boldsymbol{A}_{12}^T \boldsymbol{A}_{11}^{-1} \boldsymbol{A}_{12} \equiv mn \left(\frac{d^2 d(1/\psi)}{d\psi^2} \right) + O_P(mn^{1/2})$$

and so

$$\frac{1}{A_{22} - \boldsymbol{A}_{12}^T \boldsymbol{A}_{11}^{-1} \boldsymbol{A}_{12}} = O_P(m^{-1}n^{-1}). \tag{S.13}$$

Also,

$$\text{the lower right block of } \boldsymbol{A}_{11}^{-1} \boldsymbol{A}_{12} \boldsymbol{A}_{12}^T \boldsymbol{A}_{11}^{-1} = O_P(1) \mathbf{1}_{d_{\mathrm{A}}^{\boxplus}}^{\otimes 2}$$

and all other entries of $\boldsymbol{A}_{11}^{-1} \boldsymbol{A}_{12} \boldsymbol{A}_{12}^T \boldsymbol{A}_{11}^{-1}$ are exactly zero. Hence, in view of (S.13), the second term of (S.12) is a matrix with all entries either $O_P(m^{-1}n^{-1})$ or zero. Consequently, $\boldsymbol{A}^{11}$ equals the right-hand side of (A27) of Jiang *et al.* (2022) except for a rearrangement of the entries to concur with the $\big(\boldsymbol{\beta}_{\mathrm{A}}, \boldsymbol{\beta}_{\mathrm{B}}, \mathrm{vech}(\boldsymbol{\Sigma})\big)$ parameter ordering rather than $\big(\boldsymbol{\beta}_{\mathrm{A}}, \mathrm{vech}(\boldsymbol{\Sigma}), \boldsymbol{\beta}_{\mathrm{B}}\big)$.

The lower right entry of

$$I\Big(\boldsymbol{\beta}_{\mathrm{A}}, \boldsymbol{\beta}_{\mathrm{B}}, \mathrm{vech}(\boldsymbol{\Sigma}), \psi\Big)^{-1}$$

is

$$A^{22} = \frac{1}{A_{22}} + \frac{\boldsymbol{A}_{12}^T (\boldsymbol{A}_{11} - \boldsymbol{A}_{12} \boldsymbol{A}_{12}^T / A_{22})^{-1} \boldsymbol{A}_{12}}{A_{22}^2}. \tag{S.14}$$

Note that

$$\frac{1}{A_{22}} = \left(\frac{d^2 d(1/\psi)}{d\psi^2} \right)^{-1} (mn)^{-1} + O_P(m^{-1}n^{-3/2}).$$

Also

$$\text{the lower right block of } \boldsymbol{A}_{12} \boldsymbol{A}_{12}^T / A_{22} = O_P(mn^{-1}) \mathbf{1}_{d_{\mathrm{A}}^{\boxplus}}^{\otimes 2}$$

and all other entries of $\boldsymbol{A}_{12} \boldsymbol{A}_{12}^T / A_{22}$ are exactly zero. It follows that the expression on the right-hand side of (S.11) also holds for $\boldsymbol{A}_{11} - \boldsymbol{A}_{12} \boldsymbol{A}_{12}^T / A_{22}$ and, using equation (A27) of Jiang *et al.* (2022),

$$\frac{\boldsymbol{A}_{12}^T (\boldsymbol{A}_{11} - \boldsymbol{A}_{12} \boldsymbol{A}_{12}^T / A_{22})^{-1} \boldsymbol{A}_{12}}{A_{22}^2} = O_P(m^{-1}n^{-2}).$$

Hence, the second term on the right-hand side of (S.14) is asymptotically negligible compared with the first term and we have

$$A^{22} = \left(\frac{d^2 d(1/\psi)}{d\psi^2} \right)^{-1} (mn)^{-1} + O_P(m^{-1}n^{-3/2}).$$

The off-diagonal block of the inverse Fisher information matrix is

$$\boldsymbol{A}^{12} = -(\boldsymbol{A}_{11} - \boldsymbol{A}_{12} \boldsymbol{A}_{12}^T / A_{22})^{-1} \boldsymbol{A}_{12} / A_{22} = \begin{bmatrix} \mathbf{0}_{d_{\mathrm{A}}}^T & \mathbf{0}_{d_{\mathrm{B}}}^T & O_P(m^{-1}n^{-1}) \mathbf{1}_{d_{\mathrm{A}}^{\boxplus}}^T \end{bmatrix}^T$$

and so we have

$$\begin{aligned} I\big(\boldsymbol{\beta}_{\mathrm{A}}, \boldsymbol{\beta}_{\mathrm{B}}, \mathrm{vech}(\boldsymbol{\Sigma}), \psi\big)^{-1} &= I\big(\boldsymbol{\beta}_{\mathrm{A}}, \boldsymbol{\beta}_{\mathrm{B}}, \mathrm{vech}(\boldsymbol{\Sigma}), \psi\big)_{\infty}^{-1} \\ &+ \frac{1}{mn} \begin{bmatrix} O_P(1) \mathbf{1}_{d_{\mathrm{A}}}^{\otimes 2} & O_P(1) \mathbf{1}_{d_{\mathrm{A}}} \mathbf{1}_{d_{\mathrm{B}}}^T & O_P(1) \mathbf{1}_{d_{\mathrm{A}}} \mathbf{1}_{d_{\mathrm{A}}^{\boxplus}}^T & \mathbf{0}_{d_{\mathrm{A}}} \\ O_P(1) \mathbf{1}_{d_{\mathrm{B}}} \mathbf{1}_{d_{\mathrm{A}}}^T & o_P(1) \mathbf{1}_{d_{\mathrm{B}}}^{\otimes 2} & O_P(1) \mathbf{1}_{d_{\mathrm{B}}} \mathbf{1}_{d_{\mathrm{A}}^{\boxplus}}^T & \mathbf{0}_{d_{\mathrm{B}}} \\ O_P(1) \mathbf{1}_{d_{\mathrm{A}}^{\boxplus}} \mathbf{1}_{d_{\mathrm{A}}}^T & O_P(1) \mathbf{1}_{d_{\mathrm{A}}^{\boxplus}} \mathbf{1}_{d_{\mathrm{B}}}^T & O_P(1) \mathbf{1}_{d_{\mathrm{A}}^{\boxplus}}^{\otimes 2} & O_P(1) \mathbf{1}_{d_{\mathrm{A}}^{\boxplus}} \\ \mathbf{0}_{d_{\mathrm{A}}}^T & \mathbf{0}_{d_{\mathrm{B}}}^T & O_P(1) \mathbf{1}_{d_{\mathrm{A}}^{\boxplus}}^T & O_P(n^{-1/2}) \end{bmatrix} \end{aligned} \tag{S.15}$$

where

$$I(\boldsymbol{\beta}_{\mathrm{A}}, \boldsymbol{\beta}_{\mathrm{B}}, \mathrm{vech}(\boldsymbol{\Sigma}), \psi)_{\infty}^{-1} \equiv \begin{bmatrix} \dfrac{\boldsymbol{\Sigma}}{m} & \boldsymbol{O} & \boldsymbol{O} & \boldsymbol{O} \\ \boldsymbol{O} & \dfrac{\phi \boldsymbol{\Lambda}_{\boldsymbol{\beta}_{\mathrm{B}}}}{mn} & \boldsymbol{O} & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{O} & \dfrac{2\boldsymbol{D}_{d_{\mathrm{A}}}^{+}(\boldsymbol{\Sigma} \otimes \boldsymbol{\Sigma})\boldsymbol{D}_{d_{\mathrm{A}}}^{+T}}{m} & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{O} & \boldsymbol{O} & \dfrac{\left(\dfrac{d^2 d(1/\psi)}{d\psi^2}\right)^{-1}}{mn} \end{bmatrix}.$$

S.2.4 Asymptotic Equivalence of $\{I(\boldsymbol{\beta}_{\mathbf{A}}, \boldsymbol{\beta}_{\mathbf{B}}, \mathrm{vech}(\boldsymbol{\Sigma}), \psi)^{-1}\}^{1/2}$ and $\{I(\boldsymbol{\beta}_{\mathbf{A}}, \boldsymbol{\beta}_{\mathbf{B}}, \mathrm{vech}(\boldsymbol{\Sigma}), \psi)_{\infty}^{-1}\}^{1/2}$

Our aim in this subsection is to establish asymptotic equivalence between

$$\{I(\boldsymbol{\beta}_{\mathrm{A}}, \boldsymbol{\beta}_{\mathrm{B}}, \mathrm{vech}(\boldsymbol{\Sigma}), \psi)^{-1}\}^{1/2} \quad \text{and} \quad \{I(\boldsymbol{\beta}_{\mathrm{A}}, \boldsymbol{\beta}_{\mathrm{B}}, \mathrm{vech}(\boldsymbol{\Sigma}), \psi)_{\infty}^{-1}\}^{1/2}$$

in the following sense:

$$\left\| \{I(\boldsymbol{\beta}_{\mathrm{A}}, \boldsymbol{\beta}_{\mathrm{B}}, \mathrm{vech}(\boldsymbol{\Sigma}), \psi)_{\infty}^{-1}\}^{-1/2} \{I(\boldsymbol{\beta}_{\mathrm{A}}, \boldsymbol{\beta}_{\mathrm{B}}, \mathrm{vech}(\boldsymbol{\Sigma}), \psi)^{-1}\}^{1/2} - \boldsymbol{I} \right\|_{F} \xrightarrow{P} 0. \tag{S.16}$$

Without loss of generality, we change the ordering of the parameters from $(\boldsymbol{\beta}_{\mathrm{A}}, \boldsymbol{\beta}_{\mathrm{B}}, \mathrm{vech}(\boldsymbol{\Sigma}), \psi)$ to $(\boldsymbol{\beta}_{\mathrm{A}}, \mathrm{vech}(\boldsymbol{\Sigma}), \boldsymbol{\beta}_{\mathrm{B}}, \psi)$ and note that

$$I(\boldsymbol{\beta}_{\mathrm{A}}, \mathrm{vech}(\boldsymbol{\Sigma}), \boldsymbol{\beta}_{\mathrm{B}}, \psi)_{\infty}^{-1} = \frac{1}{m} \begin{bmatrix} \boldsymbol{K} & \boldsymbol{O} \\ \boldsymbol{O} & \frac{1}{n}\boldsymbol{L} \end{bmatrix}$$

where

$$\boldsymbol{K} \equiv \begin{bmatrix} \boldsymbol{\Sigma} & \boldsymbol{O} \\ \boldsymbol{O} & 2\boldsymbol{D}_{d_{\mathrm{A}}}^{+}(\boldsymbol{\Sigma} \otimes \boldsymbol{\Sigma})\boldsymbol{D}_{d_{\mathrm{A}}}^{+T} \end{bmatrix} \quad \text{and} \quad \boldsymbol{L} \equiv \begin{bmatrix} \phi \boldsymbol{\Lambda}_{\boldsymbol{\beta}_{\mathrm{B}}} & \boldsymbol{O} \\ \boldsymbol{O} & \left(\dfrac{d^2 d(1/\psi)}{d\psi^2}\right)^{-1} \end{bmatrix}.$$

Also,

$$I(\boldsymbol{\beta}_{\mathrm{A}}, \mathrm{vech}(\boldsymbol{\Sigma}), \boldsymbol{\beta}_{\mathrm{B}}, \psi)^{-1} = \frac{1}{m} \begin{bmatrix} \boldsymbol{K} + O_p(n^{-1})\mathbf{1}_{d_{\mathrm{A}}+d_{\mathrm{A}}^{\boxplus}}^{\otimes 2} & O_p(n^{-1})\mathbf{1}_{d_{\mathrm{A}}+d_{\mathrm{A}}^{\boxplus}}\mathbf{1}_{d_{\mathrm{B}}+1}^{T} \\ O_P(n^{-1})\mathbf{1}_{d_{\mathrm{B}}+1}^{T}\mathbf{1}_{d_{\mathrm{A}}+d_{\mathrm{A}}^{\boxplus}} & \frac{1}{n}\boldsymbol{L} + o_p(n^{-1})\mathbf{1}_{d_{\mathrm{B}}+1}^{\otimes 2} \end{bmatrix}$$

and (S.16) follows from Lemma 2 of Jiang *et al.* (2022).

S.2.5 Final Steps

Using steps analogous to those given in Appendix A.8 of Jiang *et al.* (2022) we have

$$\sqrt{m} \begin{bmatrix} \widehat{\boldsymbol{\beta}}_{\mathrm{A}} - \boldsymbol{\beta}_{\mathrm{A}}^{0} \\ \sqrt{n}\left(\widehat{\boldsymbol{\beta}}_{\mathrm{B}} - \boldsymbol{\beta}_{\mathrm{B}}^{0}\right) \\ \mathrm{vech}(\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}^{0}) \\ \sqrt{n}(\widehat{\psi} - \psi^{0}) \end{bmatrix} \xrightarrow{\mathcal{D}} N \left(\begin{bmatrix} \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \\ 0 \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Sigma}^{0} & \boldsymbol{O} & \boldsymbol{O} & \boldsymbol{O} \\ \boldsymbol{O} & \phi^{0}\boldsymbol{\Lambda}_{\boldsymbol{\beta}_{\mathrm{B}}} & \boldsymbol{O} & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{O} & 2\boldsymbol{D}_{d_{\mathrm{A}}}^{+}(\boldsymbol{\Sigma}^{0} \otimes \boldsymbol{\Sigma}^{0})\boldsymbol{D}_{d_{\mathrm{A}}}^{+T} & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{O} & \boldsymbol{O} & \left(\dfrac{d^2 d(1/\psi)}{d\psi^2}\right)_{\psi=\psi^{0}}^{-1} \end{bmatrix} \right).$$

Application of the Multivariate Delta Method (e.g. Agresti, 2013, Section 16.1.3) with the mapping

$$\left(x_1, \ldots, x_{d_{\mathrm{A}}+d_{\mathrm{B}}+d_{\mathrm{A}}^{\boxplus}}, x_{d_{\mathrm{A}}+d_{\mathrm{B}}+d_{\mathrm{A}}^{\boxplus}+1}\right) \mapsto \left(x_1, \ldots, x_{d_{\mathrm{A}}+d_{\mathrm{B}}+d_{\mathrm{A}}^{\boxplus}}, 1/x_{d_{\mathrm{A}}+d_{\mathrm{B}}+d_{\mathrm{A}}^{\boxplus}+1}\right)$$

leads to Theorem 1.

S.3 Illustration of Simplification of (A3)

Assumption (A3) of Theorem 1 is that

$$E\left[\frac{E\Big[\max\big(1,\|\boldsymbol{X}\|\big)^8\,\max\big\{1,b''\big((\boldsymbol{\beta}_{\text{A}}+\boldsymbol{U})^T\boldsymbol{X}_{\text{A}}+\boldsymbol{\beta}_{\text{B}}^T\boldsymbol{X}_{\text{B}}\big)\big\}^4\Big|\boldsymbol{U}\Big]}{\min\big\{1,\lambda_{\min}\big(E\{\boldsymbol{X}_{\text{A}}\boldsymbol{X}_{\text{A}}^T\,b''\big((\boldsymbol{\beta}_{\text{A}}+\boldsymbol{U})^T\boldsymbol{X}_{\text{A}}+\boldsymbol{\beta}_{\text{B}}^T\boldsymbol{X}_{\text{B}}\big)|\boldsymbol{U}\}\big)\big\}^2}\right]<\infty \tag{S.17}$$

for all $\boldsymbol{\beta}_{\text{A}}\in\mathbb{R}^{d_{\text{A}}}$, $\boldsymbol{\beta}_{\text{B}}\in\mathbb{R}^{d_{\text{B}}}$ and $\boldsymbol{\Sigma}$ a $d_{\text{A}}\times d_{\text{A}}$ symmetric and positive definite matrix.

In this section we prove that, for the special case:

$$d_{\text{A}}=1,\quad \boldsymbol{X}_{\text{A}}=1\quad\text{and}\quad b=\exp \tag{S.18}$$

corresponding to Poisson responses, assumption (A3) is implied by the moment generating function existence condition

$$E\{\exp(\boldsymbol{t}^T\boldsymbol{X}_{\text{B}})\}<\infty\text{ for all }\boldsymbol{t}\in\mathbb{R}^{d_{\text{B}}}. \tag{S.19}$$

To justify the sufficiency of (S.19) first note that, for the (S.18) special case, the numerator of the random variable inside the outermost expectation of (S.17) equals

$$E\Big[\max\big\{1,(1+\|\boldsymbol{X}_{\text{B}}\|^2)^4\big\}\,\max\big\{1,\exp\big(\beta_{\text{A}}+U+\boldsymbol{\beta}_{\text{B}}^T\boldsymbol{X}_{\text{B}}\big)\big\}^4|U\Big].$$

Then application of the Cauchy-Schwartz inequality for conditional expectations gives

$$\begin{aligned}&E\Big[\max\big\{1,(1+\|\boldsymbol{X}_{\text{B}}\|^2)^4\big\}\,\max\big\{1,\exp\big(\beta_{\text{A}}+U+\boldsymbol{\beta}_{\text{B}}^T\boldsymbol{X}_{\text{B}}\big)\big\}^4\Big|U\Big]\\&\le\Big(E\Big[\max\big\{1,(1+\|\boldsymbol{X}_{\text{B}}\|^2)^8\big\}\Big|U\Big]\Big)^{1/2}\Big(E\big[\max\big\{1,\exp\big(\beta_{\text{A}}+U+\boldsymbol{\beta}_{\text{B}}^T\boldsymbol{X}_{\text{B}}\big)\big\}^8|U\big]\Big)^{1/2}\\&\le\Big[1+E\big\{\big(1+\|\boldsymbol{X}_{\text{B}}\|^2\big)^8\big\}\Big]^{1/2}\Big(1+E\big[\exp\big\{8(\beta_{\text{A}}+U+\boldsymbol{\beta}_{\text{B}}^T\boldsymbol{X}_{\text{B}})\big\}|U\big]\Big)^{1/2}\\&=\Big[1+E\big\{\big(1+\|\boldsymbol{X}_{\text{B}}\|^2\big)^8\big\}\Big]^{1/2}\Big[1+\exp(8\beta_{\text{A}})E\big\{\exp(8\boldsymbol{\beta}_{\text{B}}^T\boldsymbol{X}_{\text{B}})\big\}\exp(8U)\Big]^{1/2}.\end{aligned}$$

The denominator of the random variable inside the outermost expectation of (S.17) is

$$\begin{aligned}&\min\big\{1,E\{\exp\big(\beta_{\text{A}}+U+\boldsymbol{\beta}_{\text{B}}^T\boldsymbol{X}_{\text{B}}\big)|U\}\big\}^2\\&\qquad=\min\big\{1,\exp(2\beta_{\text{A}})[E\{\exp(\boldsymbol{\beta}_{\text{B}}^T\boldsymbol{X}_{\text{B}})\}]^2\exp(2U)\}.\end{aligned}$$

Noting that, for all $x\in\mathbb{R}$ and $a,b>0$,

$$\frac{\{1+a\exp(8x)\}^{1/2}}{\min\{1,b\exp(2x)\}}\le1+\frac{a^{1/2}\exp(2x)}{b}+a^{1/2}\exp(4x)+\frac{\exp(-2x)}{b}$$

the random variable inside the outermost expectation of (S.17) is bounded above by

$$\begin{aligned}\Big[1+E\big\{\big(1+\|\boldsymbol{X}_{\text{B}}\|^2\big)^8\big\}\Big]^{1/2}\Bigg(1&+\frac{\exp(2\beta_{\text{A}})[E\{\exp(8\boldsymbol{\beta}_{\text{B}}^T\boldsymbol{X}_{\text{B}})\}]^{1/2}\exp(2U)}{[E\{\exp(\boldsymbol{\beta}_{\text{B}}^T\boldsymbol{X}_{\text{B}})\}]^2}\\&+\exp(4\beta_{\text{A}})[E\{\exp(8\boldsymbol{\beta}_{\text{B}}^T\boldsymbol{X}_{\text{B}})\}]^{1/2}\exp(4U)\\&+\frac{\exp(-2U)}{\exp(2\beta_{\text{A}})[E\{\exp(\boldsymbol{\beta}_{\text{B}}^T\boldsymbol{X}_{\text{B}})\}]^2}\Bigg).\end{aligned} \tag{S.20}$$

Since $U\sim N(0,(\sigma^2)^0)$ we have $E\{\exp(tU)\}=\exp\{\frac{1}{2}t^2(\sigma^2)^0\}$ for all $t\in\mathbb{R}$. Hence, under assumption (S.19), the expectation of (S.20) is finite which implies that (S.17) holds for the (S.18) special case.

References

Agresti, A. (2013). *Categorical Data Analysis*. Hoboken, New Jersey: John Wiley & Sons.

Miyata, Y. (2004). Fully exponential Laplace approximation using asymptotic modes. *Journal of the American Statistical Association*, **99**, 1037–1049.

Wand, M.P. (2007). Fisher information for generalised linear mixed models. *Journal of Multivariate Analysis*, **98**, 1412–1416.