

Supplement for:
**Second Term Improvements to Generalised
 Linear Mixed Model Asymptotics**

BY LUCA MAESTRINI¹, AISHWARYA BHASKARAN² AND MATT P. WAND²

¹*The Australian National University and* ²*University of Technology Sydney*

S.1 Introduction

The purpose of this supplement is to provide detailed derivational steps for the main results of § 3.3 and further details on our simulation exercise. § S.2–S.5 provide relevant results concerning matrix algebra and multivariate calculus. In § S.6–S.9 we focus on the scores of the model parameters and their high-order asymptotic approximations. § S.10 and § S.11 are concerned with approximation of the Fisher information matrix. The final stages of the derivations of the § 3.3 results are given in § S.12 and § S.14. § S.15 provides derivational details concerning the sample size illustration described in § 4. § S.16 reports some additional simulation results involving confidence intervals for the parameters of a logistic mixed model.

S.2 Matrix Algebraic Results

The derivation of the results in § 3.3 benefits from particular matrix results, which are summarized in this section.

For each $d \in \mathbb{N}$ the $d^2 \times \frac{1}{2}d(d+1)$ matrix D_d and $d^2 \times d^2$ matrix K_d are constant matrices containing zeroes and ones such that

$$D_d \text{vech}(A) = \text{vec}(A) \quad \text{for all symmetric } d \times d \text{ matrices } A$$

and

$$K_d \text{vec}(B) = \text{vec}(B^T) \quad \text{for all } d \times d \text{ matrices } B.$$

Examples are

$$D_2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad \text{and} \quad K_2 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

The D_d are called *duplication* matrices, whilst the K_d are called *commutation* matrices. As stated in § 2, the Moore-Penrose inverse of D_d is $D_d^+ = (D_d^T D_d)^{-1} D_d^T$. Chapter 3 of Magnus & Neudecker (1999) contains several results concerning these families of matrices, a few of which are relevant to the derivation of (7). For convenience, we list them here.

Theorem 9(c) in Chapter 3 of Magnus & Neudecker (1999) implies that for any $d \times d$ matrix A and $d \times 1$ vector b , we have

$$K_d(A \otimes b) = b \otimes A. \tag{S.1}$$

Theorem 12(a) in the same chapter asserts that

$$K_d D_d = D_d \tag{S.2}$$

and implies that, for any $d \times d$ matrix A ,

$$D_d^T \text{vec}(A) = D_d^T \text{vec}(A^T). \tag{S.3}$$

Also, Theorem 13(b) and Theorem 13(d) provide for a $d \times d$ matrix A

$$D_d D_d^+(A \otimes A) D_d^{+T} = (A \otimes A) D_d^{+T} \quad (\text{S.4})$$

and, assuming that A is invertible,

$$\{D_d^T(A \otimes A) D_d\}^{-1} = D_d^+(A^{-1} \otimes A^{-1}) D_d^{+T}. \quad (\text{S.5})$$

Lastly, we state two matrix identities that are used in the derivations. For matrices A , B and C such that ABC is defined, we have

$$\text{vec}(ABC) = (C^T \otimes A) \text{vec}(B). \quad (\text{S.6})$$

For conformable matrices A , B , C and D , we have

$$(A \otimes B)(C \otimes D) = (AC) \otimes (BD). \quad (\text{S.7})$$

S.3 Multivariate Derivative Notation

For f a smooth real-valued function of the d -variate argument $x \equiv (x_1, \dots, x_d)$, let $\nabla f(x)$ denote the $d \times 1$ vector with r th entry $\partial f(x)/\partial x_r$, $\nabla^2 f(x)$ denote the $d \times d$ matrix with (r, s) entry $\partial^2 f(x)/(\partial x_r \partial x_s)$ and $\nabla^3 f(x)$ denote the $d \times d \times d$ array with (r, s, t) entry $\partial^3 f(x)/(\partial x_r \partial x_s \partial x_t)$.

S.4 Three-Term Taylor Series Expansion of Gradient Vectors

Consider $f : \mathbb{R}^d \rightarrow \mathbb{R}$. Then, for sufficiently smooth f , the three-term Taylor expansion of

$$f(x+h) \quad \text{where } x \equiv (x_1, \dots, x_d) \text{ and } h \equiv (h_1, \dots, h_d)$$

is

$$f(x+h) = f(x) + \sum_{r=1}^d \{\nabla f(x)\}_r h_r + \frac{1}{2} \sum_{r=1}^d \sum_{s=1}^d \{\nabla^2 f(x)\}_{rs} h_r h_s + \dots \quad (\text{S.8})$$

Now consider $\alpha : \mathbb{R}^d \rightarrow \mathbb{R}$ and its gradient function $\nabla \alpha : \mathbb{R}^d \rightarrow \mathbb{R}^d$. If (S.8) is applied to each entry of $(\nabla \alpha)(x+h)$ then we have

$$\begin{bmatrix} \{(\nabla \alpha)(x+h)\}_1 \\ \vdots \\ \{(\nabla \alpha)(x+h)\}_d \end{bmatrix} = \begin{bmatrix} \{(\nabla \alpha)(x)\}_1 + \sum_{r=1}^d \{\nabla^2 \alpha(x)\}_{r1} h_r + \frac{1}{2} \sum_{r=1}^d \sum_{s=1}^d \{\nabla^3 \alpha(x)\}_{rs1} h_r h_s \\ \vdots \\ \{(\nabla \alpha)(x)\}_d + \sum_{r=1}^d \{\nabla^2 \alpha(x)\}_{rd} h_r + \frac{1}{2} \sum_{r=1}^d \sum_{s=1}^d \{\nabla^3 \alpha(x)\}_{rds} h_r h_s \end{bmatrix} + \dots$$

From this it is clear that

$$(\nabla \alpha)(x+h) = (\nabla \alpha)(x) + \{(\nabla^2 \alpha)(x)\}h + \frac{1}{2} \{(\nabla^3 \alpha)(x)\} \star (hh^T) + \dots \quad (\text{S.9})$$

where the \star notation is as defined in Section 3.2.

S.5 Higher Order Approximation of Multivariate Integral Ratios

The main tool for approximation of the Fisher information matrix of (3) 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 g , c and h ,

$$\begin{aligned} \frac{\int_{\mathbb{R}^d} g(x) c(x) \exp\{-nh(x)\} dx}{\int_{\mathbb{R}^d} c(x) \exp\{-nh(x)\} dx} &= g(x^*) + \frac{\nabla g(x^*)^T \{\nabla^2 h(x^*)\}^{-1} \nabla c(x^*)}{nc(x^*)} \\ &+ \frac{\text{tr}[\{\nabla^2 h(x^*)\}^{-1} \nabla^2 g(x^*)]}{2n} - \frac{\nabla g(x^*)^T \{\nabla^2 h(x^*)\}^{-1} [\nabla^3 h(x^*) \star \{\nabla^2 h(x^*)\}^{-1}]}{2n} + O(n^{-2}) \end{aligned} \quad (\text{S.10})$$

where

$$x^* \equiv \underset{x \in \mathbb{R}^d}{\operatorname{argmin}} h(x).$$

S.6 Exact Score Expressions

For $1 \leq i \leq m$, let $p_{Y_i|X_i}$ denote the conditional density function, or probability mass function, of Y_i given X_i . Then let

$$S_{Ai} \equiv \nabla_{\beta_A} \log p_{Y_i|X_i}(Y_i|X_i), \quad S_{Bi} \equiv \nabla_{\beta_B} \log p_{Y_i|X_i}(Y_i|X_i)$$

and

$$S_{Ci} \equiv \nabla_{\operatorname{vech}(\Sigma)} \log p_{Y_i|X_i}(Y_i|X_i)$$

denote the i th contribution to the scores with respect to each of β_A , β_B and $\operatorname{vech}(\Sigma)$. Then it is straightforward to show that the exact scores are

$$S_{Ai} = \frac{\int_{\mathbb{R}^{d_R}} g_{iA}(u) c_S(u) \exp\{-nh_i(u)\} du}{\int_{\mathbb{R}^{d_R}} c_S(u) \exp\{-nh_i(u)\} du}, \quad (\text{S.11})$$

$$S_{Bi} = \frac{\int_{\mathbb{R}^{d_R}} g_{iB}(u) c_S(u) \exp\{-nh_i(u)\} du}{\int_{\mathbb{R}^{d_R}} c_S(u) \exp\{-nh_i(u)\} du} \quad (\text{S.12})$$

and

$$S_{Ci} = \frac{\int_{\mathbb{R}^{d_R}} g_{iC}(u) c_S(u) \exp\{-nh_i(u)\} du}{\int_{\mathbb{R}^{d_R}} c_S(u) \exp\{-nh_i(u)\} du} - \frac{1}{2} D_{d_R}^T \operatorname{vec}(\Sigma^{-1}) \quad (\text{S.13})$$

where

$$c_S(u) \equiv \exp\left(-\frac{1}{2} u^T \Sigma^{-1} u\right), \quad g_{iA}(u) \equiv \Sigma^{-1} u,$$

$$g_{iB}(u) \equiv \frac{1}{\phi} \sum_{j=1}^{n_i} X_{Bij} \{Y_{ij} - b'((\beta_A + u)^T X_{Aij} + \beta_B^T X_{Bij})\},$$

$$g_{iC}(u) \equiv \frac{1}{2} D_{d_R}^T (\Sigma^{-1} \otimes \Sigma^{-1}) \operatorname{vec}(uu^T)$$

$$\text{and } h_i(u) \equiv -\frac{1}{n\phi} \sum_{j=1}^{n_i} \{Y_{ij} u^T X_{Aij} - b((\beta_A + u)^T X_{Aij} + \beta_B^T X_{Bij})\}.$$

An integration by parts step is used to obtain the S_{Ai} expression.

In the upcoming sections we obtain asymptotic approximations of S_{Ai} , S_{Bi} and S_{Ci} . Key quantities for these approximations are

$$U_i^* \equiv \underset{u \in \mathbb{R}^{d_R}}{\operatorname{argmin}} h_i(u), \quad 1 \leq i \leq m.$$

S.7 Definitions of Key Summation Quantities

Our derivation of (7) involves manipulations of particular summation quantities, which are defined in this section. At the end of this section we state some important moment-type relationships between the quantities.

For each $1 \leq i \leq m$, define \mathcal{G}_{Ai} , \mathcal{G}_{Bi} , \mathcal{H}_{AAi} , \mathcal{H}_{ABi} , and \mathcal{H}_{BBi} as follows:

$$\begin{aligned}\mathcal{G}_{Ai} &\equiv \sum_{j=1}^{n_i} \{Y_{ij} - b'((\beta_A + U_i)^T X_{Aij} + \beta_B^T X_{Bij})\} X_{Aij}, \\ \mathcal{G}_{Bi} &\equiv \sum_{j=1}^{n_i} \{Y_{ij} - b'((\beta_A + U_i)^T X_{Aij} + \beta_B^T X_{Bij})\} X_{Bij}, \\ \mathcal{H}_{AAi} &\equiv \sum_{j=1}^{n_i} b''((\beta_A + U_i)^T X_{Aij} + \beta_B^T X_{Bij}) X_{Aij} X_{Aij}^T, \\ \mathcal{H}_{ABi} &\equiv \sum_{j=1}^{n_i} b''((\beta_A + U_i)^T X_{Aij} + \beta_B^T X_{Bij}) X_{Aij} X_{Bij}^T \\ \text{and } \mathcal{H}_{BBi} &\equiv \sum_{j=1}^{n_i} b''((\beta_A + U_i)^T X_{Aij} + \beta_B^T X_{Bij}) X_{Bij} X_{Bij}^T.\end{aligned}$$

In a similar vein, define \mathcal{H}'_{AAAi} to be the $d_R \times d_R \times d_R$ array with (r, s, t) entry equal to

$$\sum_{j=1}^{n_i} b'''((\beta_A + U_i)^T X_{Aij} + \beta_B^T X_{Bij}) (X_{Aij})_r (X_{Aij})_s (X_{Aij})_t$$

and \mathcal{H}'_{AABi} to be the $d_R \times d_R \times d_B$ array with (r, s, t) entry equal to

$$\sum_{j=1}^{n_i} b'''((\beta_A + U_i)^T X_{Aij} + \beta_B^T X_{Bij}) (X_{Aij})_r (X_{Aij})_s (X_{Bij})_t$$

where

$$d_B \equiv d_F - d_R.$$

The following relationships are of fundamental importance for the derivation of (7):

$$\begin{aligned}E(\mathcal{G}_{Ai}|X_i, U_i) &= 0, \quad E(\mathcal{G}_{Bi}|X_i, U_i) = 0, \\ E(\mathcal{G}_{Ai}^{\otimes 2}|X_i, U_i) &= \phi \mathcal{H}_{AAi}, \quad E(\mathcal{G}_{Ai} \mathcal{G}_{Bi}^T|X_i, U_i) = \phi \mathcal{H}_{ABi} \quad \text{and} \quad E(\mathcal{G}_{Bi}^{\otimes 2}|X_i, U_i) = \phi \mathcal{H}_{BBi}\end{aligned}\tag{S.14}$$

where, throughout this supplement,

$$v^{\otimes 2} \equiv vv^T \quad \text{for any column vector } v.$$

Also note that

$$\begin{aligned}\mathcal{G}_{Ai} &= O_p(n^{1/2})1_{d_R}, \quad \mathcal{G}_{Bi} = O_p(n^{1/2})1_{d_B}, \quad \mathcal{H}_{AAi} = O_p(n)1_{d_R}^{\otimes 2}, \\ \mathcal{H}_{BBi} &= O_p(n)1_{d_B}^{\otimes 2}, \quad \mathcal{H}_{ABi} = O_p(n)1_{d_R}1_{d_B}^T\end{aligned}$$

and that all entries of \mathcal{H}'_{AAAi} and \mathcal{H}'_{AABi} are $O_p(n)$.

S.8 Approximation of U_i^*

Use of (S.10) to approximate S_{Ai} , S_{Bi} and S_{Ci} requires approximation of U_i^* . Introduce the notation $\mathcal{C}_i(u) \equiv n\phi h_i(u)$. Then U_i^* satisfies

$$\nabla \mathcal{C}_i(U_i^*) = 0$$

where

$$\nabla \mathcal{C}_i(u) \equiv - \sum_{j=1}^{n_i} \left\{ Y_{ij} - b'((\beta_A + u)^T X_{Aij} + \beta_B^T X_{Bij}) \right\} X_{Aij}.$$

Then, from (S.9) we have

$$\begin{aligned} \nabla \mathcal{C}_i(U_i^*) &= \nabla \mathcal{C}_i(U_i + U_i^* - U_i) \\ &= \nabla \mathcal{C}_i(U_i) + \{ \nabla^2 \mathcal{C}_i(U_i) \} (U_i^* - U_i) \\ &\quad + \frac{1}{2} \{ \nabla^3 \mathcal{C}_i(U_i) \} \star \{ (U_i^* - U_i)(U_i^* - U_i)^T \} + \dots \end{aligned}$$

Next we seek explicit expressions for $\nabla^2 \mathcal{C}_i(u)$ and $\nabla^3 \mathcal{C}_i(u)$. Standard vector calculus arguments lead to

$$\begin{aligned} \nabla^2 \mathcal{C}_i(u) &= \sum_{j=1}^{n_i} b''((\beta_A + u)^T X_{Aij} + \beta_B^T X_{Bij}) X_{Aij} X_{Aij}^T \\ &= \left[\sum_{j=1}^{n_i} b''((\beta_A + u)^T X_{Aij} + \beta_B^T X_{Bij}) (X_{Aij})_r (X_{Aij})_s \right]_{1 \leq r, s \leq d_R}. \end{aligned}$$

Then, the three-dimension array of all third order partial derivatives of $\mathcal{C}_i(u)$ is

$$\nabla^3 \mathcal{C}_i(u) = \left[\sum_{j=1}^{n_i} b'''((\beta_A + u)^T X_{Aij} + \beta_B^T X_{Bij}) (X_{Aij})_r (X_{Aij})_s (X_{Aij})_t \right]_{1 \leq r, s, t \leq d_R}.$$

We then have

$$\nabla \mathcal{C}_i(U_i^*) = -\mathcal{G}_{Ai} + \mathcal{H}_{AAi}(U_i^* - U_i) + \frac{1}{2} \mathcal{H}'_{AAAi} \star \{ (U_i^* - U_i)(U_i^* - U_i)^T \} + \dots$$

and so $\nabla \mathcal{C}_i(U_i^*) = 0$ is equivalent to

$$\mathcal{H}_{AAi}^{-1} \mathcal{G}_{Ai} = (U_i^* - U_i) + \frac{1}{2} \mathcal{H}_{AAi}^{-1} \left[\mathcal{H}'_{AAAi} \star \{ (U_i^* - U_i)(U_i^* - U_i)^T \} \right] + \dots \quad (\text{S.15})$$

We now invert (S.15) using the set-up given around equations (9.43) and (9.44) of Pace & Salvati (1997). To match the notation given there, set

$$y \equiv \mathcal{H}_{AAi}^{-1} \mathcal{G}_{Ai} \quad \text{and} \quad x \equiv U_i^* - U_i.$$

Then, in keeping with the displayed equation just before (9.43) of Pace & Salvati (1997) and using their superscript and subscript conventions, we have

$$y^a \equiv \text{the } a\text{th entry of } y \quad \text{and} \quad x^a \equiv \text{the } a\text{th entry of } x.$$

Also,

$$x^{rs} \equiv x^r x^s = \text{the } (r, s) \text{ entry of } x x^T = \text{the } (r, s) \text{ entry of } (U_i^* - U_i)^{\otimes 2}.$$

Then

$$y^a = x^a + A_{rs}^a x^{rs} + \dots$$

where

$$\begin{aligned} A_{rs}^a x^{rs} &= \text{the } a\text{th entry of } \frac{1}{2} \mathcal{H}_{AAi}^{-1} \left[\mathcal{H}'_{AAAi} \star \{ (U_i^* - U_i)(U_i^* - U_i)^T \} \right] \\ &= \text{the } a\text{th entry of } \frac{1}{2} \mathcal{H}_{AAi}^{-1} \left\{ \mathcal{H}'_{AAAi} \star (x x^T) \right\}. \end{aligned}$$

From equations (9.43) and (9.44) of Pace & Salvan (1997),

$$\begin{aligned}
x^a &= y^a - A_{rs}^a y^{rs} + \dots \\
&= y^a - \text{the } a\text{th entry of } \frac{1}{2} \mathcal{H}_{AAi}^{-1} \left\{ \mathcal{H}'_{AAAi} \star (yy^T) \right\} + \dots \\
&= y^a - \text{the } a\text{th entry of } \frac{1}{2} \mathcal{H}_{AAi}^{-1} \left\{ \mathcal{H}'_{AAAi} \star \left(\mathcal{H}_{AAi}^{-1} \mathcal{G}_{Ai} \right)^{\otimes 2} \right\} + \dots
\end{aligned}$$

This results in the following three-term approximation of U_i^* :

$$U_i^* = U_i + \mathcal{H}_{AAi}^{-1} \mathcal{G}_{Ai} - \frac{1}{2} \mathcal{H}_{AAi}^{-1} \left\{ \mathcal{H}'_{AAAi} \star \left(\mathcal{H}_{AAi}^{-1} \mathcal{G}_{Ai} \mathcal{G}_{Ai}^T \mathcal{H}_{AAi}^{-1} \right) \right\} + O_p(n^{-3/2}) 1_{d_R}. \quad (\text{S.16})$$

S.9 Score Asymptotic Approximation

We are now ready to obtain approximations of the scores S_{Ai} , S_{Bi} and S_{Ci} with accuracies that are sufficient for the two-term asymptotic covariance matrices of (7).

S.9.1 Approximation of S_{Ai}

For each $1 \leq r \leq d_R$, let e_r denote the $d_R \times 1$ vector having r th entry equal to 1 and zeroes elsewhere.

S.9.1.1 The (S.10) First Term Contribution

For each $1 \leq r \leq d_R$, the contribution to the r th entry of S_{Ai} from the first term on the right-hand side of (S.10) is the r th entry of $\Sigma^{-1} U_i^*$. In view of (S.16) we obtain the following contribution to S_{Ai} :

$$\Sigma^{-1} U_i + \Sigma^{-1} \mathcal{H}_{AAi}^{-1} \mathcal{G}_{Ai} - \frac{1}{2} \Sigma^{-1} \mathcal{H}_{AAi}^{-1} \left\{ \mathcal{H}'_{AAAi} \star \left(\mathcal{H}_{AAi}^{-1} \mathcal{G}_{Ai} \mathcal{G}_{Ai}^T \mathcal{H}_{AAi}^{-1} \right) \right\} + O_p(n^{-3/2}) 1_{d_R}.$$

S.9.1.2 The (S.10) Second Term Contribution

Noting that

$$\nabla \{ e_r^T g_{iA}(u) \} = e_r^T \Sigma^{-1} \quad \text{and} \quad \nabla c_S(u) = -c_S(u) \Sigma^{-1} u,$$

the contribution to the r th entry of S_{Ai} from the second term on the right-hand side of (S.10) is

$$-\phi e_r^T \Sigma^{-1} \left\{ \sum_{j=1}^{n_i} b'' \left((\beta_A + U_i^*)^T X_{Aij} + \beta_B^T X_{Bij} \right) X_{Aij} X_{Aij}^T \right\}^{-1} \Sigma^{-1} U_i^*. \quad (\text{S.17})$$

Substitution of (S.16) into (S.17) then leads to the following contribution to S_{Ai} :

$$-\phi \Sigma^{-1} \mathcal{H}_{AAi}^{-1} \Sigma^{-1} U_i + O_p(n^{-3/2}) 1_{d_R}.$$

S.9.1.3 The (S.10) Third Term Contribution

Noting that $\nabla^2 \{ e_r^T g_{iA}(u) \} = O$, the contribution to S_{Ai} from the third term on the right-hand side of (S.10) is 0.

S.9.1.4 The (S.10) Fourth Term Contribution

Via arguments similar to those given in § S.9.1.2, the contribution to S_{Ai} from the fourth term on the right-hand side of (S.10) is

$$-\frac{\phi}{2} \Sigma^{-1} \mathcal{H}_{AAi}^{-1} \left(\mathcal{H}'_{AAAi} \star \mathcal{H}_{AAi}^{-1} \right) + O_p(n^{-3/2}) 1_{d_R}.$$

S.9.1.5 The Resultant Score Approximation

On combining the results of § S.9.1.1–S.9.1.4, we obtain

$$S_{A_i} = \Sigma^{-1}U_i + \Sigma^{-1}\mathcal{H}_{AA_i}^{-1}\mathcal{G}_{A_i} - \frac{1}{2}\Sigma^{-1}\mathcal{H}_{AA_i}^{-1}\left\{\mathcal{H}'_{AAA_i}\star\left(\mathcal{H}_{AA_i}^{-1}\mathcal{G}_{A_i}\mathcal{G}_{A_i}^T\mathcal{H}_{AA_i}^{-1}\right)\right\} \\ - \phi\Sigma^{-1}\mathcal{H}_{AA_i}^{-1}\Sigma^{-1}U_i - \frac{\phi}{2}\Sigma^{-1}\mathcal{H}_{AA_i}^{-1}\left(\mathcal{H}'_{AAA_i}\star\mathcal{H}_{AA_i}^{-1}\right) + O_p(n^{-3/2})1_{d_R}. \quad (\text{S.18})$$

S.9.2 Approximation of S_{B_i}

For each $1 \leq r \leq d_R$, let e_r denote the $d_B \times 1$ vector having r th entry equal to 1 and zeroes elsewhere.

S.9.2.1 The (S.10) First Term Contribution

The contribution to S_{B_i} from the first term on the right-hand side of (S.10) is

$$g_{iB}(U_i^*) = \frac{1}{\phi} \sum_{j=1}^{n_i} X_{Bij} \{Y_{ij} - b'((\beta_A + U_i^*)^T X_{Aij} + \beta_B^T X_{Bij})\}. \quad (\text{S.19})$$

Next note that, with (S.16) as a basis,

$$\begin{aligned} & b'((\beta_A + U_i^*)^T X_{Aij} + \beta_B^T X_{Bij}) \\ &= b'((\beta_A + U_i)^T X_{Aij} + \beta_B^T X_{Bij}) \\ & \quad + X_{Aij}^T (U_i^* - U_i) b''((\beta_A + U_i)^T X_{Aij} + \beta_B^T X_{Bij}) \\ & \quad + \frac{1}{2} X_{Aij}^T (U_i^* - U_i)^{\otimes 2} X_{Aij} b'''((\beta_A + U_i)^T X_{Aij} + \beta_B^T X_{Bij}) + O_p(n^{-3/2}) \\ &= b'((\beta_A + U_i)^T X_{Aij} + \beta_B^T X_{Bij}) \\ & \quad + X_{Aij}^T \left[\mathcal{H}_{AA_i}^{-1} \mathcal{G}_{A_i} - \frac{1}{2} \mathcal{H}_{AA_i}^{-1} \left\{ \mathcal{H}'_{AAA_i} \star \left(\mathcal{H}_{AA_i}^{-1} \mathcal{G}_{A_i} \mathcal{G}_{A_i}^T \mathcal{H}_{AA_i}^{-1} \right) \right\} \right] \\ & \quad \times b''((\beta_A + U_i)^T X_{Aij} + \beta_B^T X_{Bij}) \\ & \quad + \frac{1}{2} X_{Aij}^T \left(\mathcal{H}_{AA_i}^{-1} \mathcal{G}_{A_i} \right)^{\otimes 2} X_{Aij} b'''((\beta_A + U_i)^T X_{Aij} + \beta_B^T X_{Bij}) + O_p(n^{-3/2}). \end{aligned}$$

Substitution of this result into (S.19) leads to the first term of S_{B_i} equalling

$$\begin{aligned} & \frac{1}{\phi} \left(\mathcal{G}_{B_i} - \mathcal{H}_{AB_i}^T \mathcal{H}_{AA_i}^{-1} \mathcal{G}_{A_i} \right) + \frac{1}{2\phi} \mathcal{H}_{AB_i}^T \mathcal{H}_{AA_i}^{-1} \left\{ \mathcal{H}'_{AAA_i} \star \left(\mathcal{H}_{AA_i}^{-1} \mathcal{G}_{A_i} \mathcal{G}_{A_i}^T \mathcal{H}_{AA_i}^{-1} \right) \right\} \\ & \quad - \frac{1}{2\phi} \left\{ \mathcal{H}'_{AAB_i} \star \left(\mathcal{H}_{AA_i}^{-1} \mathcal{G}_{A_i} \mathcal{G}_{A_i}^T \mathcal{H}_{AA_i}^{-1} \right) \right\} + O_p(n^{-1/2})1_{d_R}. \end{aligned}$$

S.9.2.2 The (S.10) Second Term Contribution

Noting that

$$\nabla \{e_r^T g_{iB}(u)\} = -\frac{1}{\phi} \sum_{j=1}^{n_i} b''((\beta_A + u)^T X_{Aij} + \beta_B^T X_{Bij}) e_r^T X_{Aij} X_{Bij}^T \quad (\text{S.20})$$

and recalling that $\nabla c_S(u) = -c_S(u)\Sigma^{-1}u$, the contribution from the second term on the right-hand side of (S.10) to S_{B_i} is

$$\left\{ \sum_{j=1}^{n_i} b'' \left((\beta_A + U_i^*)^T X_{Aij} + \beta_B^T X_{Bij} \right) X_{Aij} X_{Bij}^T \right\}^T \\ \times \left\{ \sum_{j=1}^{n_i} b'' \left((\beta_A + U_i^*)^T X_{Aij} + \beta_B^T X_{Bij} \right) X_{Aij} X_{Aij}^T \right\}^{-1} \Sigma^{-1} U_i^*.$$

Substitution of (S.16) then leads to the contribution to S_{B_i} from the second term of (S.10) equalling

$$\mathcal{H}_{ABi}^T \mathcal{H}_{AAi}^{-1} \Sigma^{-1} U_i + O_p(n^{-1/2}) 1_{d_R}.$$

S.9.2.3 The (S.10) Third Term Contribution

The r th entry of the contribution to S_{B_i} from the third term of (S.10) is

$$\frac{1}{2n} \sum_{s=1}^{d_B} \sum_{t=1}^{d_B} [\nabla^2 \{e_r^T g_{iB}(U_i^*)\}]_{st} [\{\nabla^2 h_i(U_i^*)\}^{-1}]_{st} \\ = \frac{\phi}{2} \sum_{s=1}^{d_B} \sum_{t=1}^{d_B} [\nabla^2 \{e_r^T g_{iB}(U_i)\}]_{st} (\mathcal{H}_{AAi}^{-1})_{st} + O_p(n^{-1/2}).$$

However, the (s, t) entry of $\nabla^2 \{e_r^T g_{iB}(U_i)\}$ is

$$-\frac{1}{\phi} \sum_{j=1}^{n_i} b''' \left((\beta_A + U_i)^T X_{Aij} + \beta_B^T X_{Bij} \right) (e_r^T X_{Aij}) (e_s^T X_{Aij}) (e_t^T X_{Bij}) = -\frac{1}{\phi} (\mathcal{H}'_{AABi})_{rst}.$$

Noting (6), the contribution to S_{B_i} from the third term of (S.10) is

$$-\frac{1}{2} \mathcal{H}'_{AABi} \star \mathcal{H}_{AAi}^{-1} + O_p(n^{-1/2}) 1_{d_R}.$$

S.9.2.4 The (S.10) Fourth Term Contribution

With the aid of (S.20), the contribution to S_{B_i} from the fourth term of (S.10) is

$$\frac{1}{2n\phi} \left\{ \sum_{j=1}^{n_i} b'' \left((\beta_A + U_i^*)^T X_{Aij} + \beta_B^T X_{Bij} \right) X_{Aij} X_{Bij}^T \right\}^T \{\nabla^2 h_i(U_i^*)\}^{-1} [\nabla^3 h_i(U_i^*) \star \{\nabla^2 h_i(U_i^*)\}^{-1}] \\ = \frac{1}{2} \mathcal{H}_{ABi}^T \mathcal{H}_{AAi}^{-1} \left(\mathcal{H}'_{AAAi} \star \mathcal{H}_{AAi}^{-1} \right) + O_p(n^{-1/2}) 1_{d_R}.$$

S.9.2.5 The Resultant Score Approximation

On combining each of the contributions, we obtain

$$S_{B_i} = \frac{1}{\phi} \left(\mathcal{G}_{B_i} - \mathcal{H}_{ABi}^T \mathcal{H}_{AAi}^{-1} \mathcal{G}_{A_i} \right) + \frac{1}{2\phi} \mathcal{H}_{ABi}^T \mathcal{H}_{AAi}^{-1} \left\{ \mathcal{H}'_{AAAi} \star \left(\mathcal{H}_{AAi}^{-1} \mathcal{G}_{A_i} \mathcal{G}_{A_i}^T \mathcal{H}_{AAi}^{-1} \right) \right\} \\ - \frac{1}{2\phi} \left\{ \mathcal{H}'_{AABi} \star \left(\mathcal{H}_{AAi}^{-1} \mathcal{G}_{A_i} \mathcal{G}_{A_i}^T \mathcal{H}_{AAi}^{-1} \right) \right\} + \mathcal{H}_{ABi}^T \mathcal{H}_{AAi}^{-1} \Sigma^{-1} U_i \\ - \frac{1}{2} \mathcal{H}'_{AABi} \star \mathcal{H}_{AAi}^{-1} + \frac{1}{2} \mathcal{H}_{ABi}^T \mathcal{H}_{AAi}^{-1} \left(\mathcal{H}'_{AAAi} \star \mathcal{H}_{AAi}^{-1} \right) + O_p(n^{-1/2}) 1_{d_R}. \quad (\text{S.21})$$

S.9.3 Approximation of S_{C_i}

For each $1 \leq r \leq \frac{1}{2}d_R(d_R + 1)$ let e_r denote the $d_R(d_R + 1)/2 \times 1$ vector with 1 in the r th position and zeroes elsewhere.

S.9.3.1 The (S.10) First Term Contribution

For each $1 \leq r \leq \frac{1}{2}d_R(d_R + 1)$, the r th entry of the contribution to S_{C_i} from the first term of (S.10) is

$$e_r^T g_{iC}(U_i^*) = \frac{1}{2} e_r^T D_{d_R}^T (\Sigma^{-1} \otimes \Sigma^{-1}) \text{vec}((U_i^*)^{\otimes 2}).$$

Since

$$\begin{aligned} (U_i^*)^{\otimes 2} &= \left[U_i + \mathcal{H}_{AAi}^{-1} \mathcal{G}_{Ai} - \frac{1}{2} \mathcal{H}_{AAi}^{-1} \left\{ \mathcal{H}'_{AAAi} \star \left(\mathcal{H}_{AAi}^{-1} \mathcal{G}_{Ai} \mathcal{G}_{Ai}^T \mathcal{H}_{AAi}^{-1} \right) \right\} + O_p(n^{-3/2}) 1_{d_R} \right]^{\otimes 2} \\ &= U_i^{\otimes 2} + U_i \mathcal{G}_{Ai}^T \mathcal{H}_{AAi}^{-1} + \mathcal{H}_{AAi}^{-1} \mathcal{G}_{Ai} U_i^T + \mathcal{H}_{AAi}^{-1} \mathcal{G}_{Ai} \mathcal{G}_{Ai}^T \mathcal{H}_{AAi}^{-1} \\ &\quad - \frac{1}{2} U_i \left\{ \mathcal{H}'_{AAAi} \star \left(\mathcal{H}_{AAi}^{-1} \mathcal{G}_{Ai} \mathcal{G}_{Ai}^T \mathcal{H}_{AAi}^{-1} \right) \right\}^T \mathcal{H}_{AAi}^{-1} \\ &\quad - \frac{1}{2} \mathcal{H}_{AAi}^{-1} \left\{ \mathcal{H}'_{AAAi} \star \left(\mathcal{H}_{AAi}^{-1} \mathcal{G}_{Ai} \mathcal{G}_{Ai}^T \mathcal{H}_{AAi}^{-1} \right) \right\} U_i^T + O_p(n^{-3/2}) 1_{d_R}^{\otimes 2}, \end{aligned}$$

and noting (S.3) and (S.6), the contribution to S_{C_i} from the first term of (S.10) is

$$\begin{aligned} \frac{1}{2} D_{d_R}^T \text{vec} \left(\Sigma^{-1} \left[U_i U_i^T + 2 \mathcal{H}_{AAi}^{-1} \mathcal{G}_{Ai} U_i^T + \mathcal{H}_{AAi}^{-1} \mathcal{G}_{Ai} \mathcal{G}_{Ai}^T \mathcal{H}_{AAi}^{-1} \right. \right. \\ \left. \left. - \mathcal{H}_{AAi}^{-1} \left\{ \mathcal{H}'_{AAAi} \star \left(\mathcal{H}_{AAi}^{-1} \mathcal{G}_{Ai} \mathcal{G}_{Ai}^T \mathcal{H}_{AAi}^{-1} \right) \right\} U_i^T \right] \Sigma^{-1} \right) + O_p(n^{-3/2}) 1_{d_R(d_R+1)/2}. \end{aligned}$$

S.9.3.2 The (S.10) Second Term Contribution

Noting that, for each $1 \leq r \leq d_R(d_R + 1)/2$,

$$[\nabla \{ e_r^T g_{iC}(u) \}]^T = \frac{1}{2} e_r^T D_{d_R}^T (\Sigma^{-1} \otimes \Sigma^{-1}) \{ (u \otimes I) + (I \otimes u) \} \quad (\text{S.22})$$

and keeping in mind that $\nabla c_S(u) = -c_S(u) \Sigma^{-1} u$, the contribution from the second term on the right-hand side of (S.10) to S_{C_i} is

$$\begin{aligned} & -\frac{\phi}{2} D_{d_R}^T (\Sigma^{-1} \otimes \Sigma^{-1}) \{ (U_i^* \otimes I) + (I \otimes U_i^*) \} \\ & \quad \times \left\{ \sum_{j=1}^{n_i} b''((\beta_A + U_i^*)^T X_{Aij} + \beta_B^T X_{Bij}) X_{Aij} X_{Aij}^T \right\}^{-1} \Sigma^{-1} U_i^* \\ & = -\frac{\phi}{2} D_{d_R}^T (\Sigma^{-1} \otimes \Sigma^{-1}) \{ (U_i \otimes I) + (I \otimes U_i) \} \mathcal{H}_{AAi}^{-1} \Sigma^{-1} U_i + O_p(n^{-3/2}) 1_{d_R(d_R+1)/2} \\ & = -\phi D_{d_R}^T \text{vec} \left(\Sigma^{-1} \mathcal{H}_{AAi}^{-1} \Sigma^{-1} U_i U_i^T \Sigma^{-1} \right) + O_p(n^{-3/2}) 1_{d_R(d_R+1)/2}. \end{aligned}$$

The last step makes use of (S.1), (S.3) and (S.6).

S.9.3.3 The (S.10) Third Term Contribution

The derivation of the (S.10) third term contribution to S_{C_i} benefits from notation and a result concerning the inverse of the vec operator. For $d \in \mathbb{N}$, if b is a $d^2 \times 1$ vector then $\text{vec}^{-1}(b)$ is the $d \times d$ matrix such that $\text{vec}(\text{vec}^{-1}(b)) = b$.

Lemma 1. Let $d \in \mathbb{N}$, a be a $d \times 1$ vector and b be a $d^2 \times 1$ vector. Then

$$(a^T \otimes I)b = \text{vec}^{-1}(b)a \quad \text{and} \quad (I \otimes a^T)b = \text{vec}^{-1}(b)^T a.$$

Lemma 1 is a relatively simple consequence of (S.6). To prove the first part of Lemma 1, note that its right-hand side is

$$\text{vec}^{-1}(b)a = \text{vec}(\text{vec}^{-1}(b)a) = \text{vec}(I \text{vec}^{-1}(b)a) = (a^T \otimes I)\text{vec}(\text{vec}^{-1}(b)) = (a^T \otimes I)b.$$

The proof of the second part of Lemma 1 is similar.

For each $1 \leq r \leq \frac{1}{2}d_{\mathbb{R}}(d_{\mathbb{R}} + 1)$, the r th entry of the contribution to S_{C_i} from the third term of (S.10) is

$$\frac{1}{2n} \text{tr} [\{\nabla^2 h_i(U_i)\}^{-1} \nabla^2 \{e_r^T g_{iC}(U_i^*)\}].$$

Next note from (S.22) that

$$d\{e_r^T g_{iC}(u)\} = \frac{1}{2} e_r^T D_{d_{\mathbb{R}}}^T (\Sigma^{-1} \otimes \Sigma^{-1}) \{(u \otimes I) + (I \otimes u)\} du.$$

Using Lemma 1 we then have

$$\begin{aligned} 2d^2 \{e_r^T g_{iC}(u)\} &= e_r^T D_{d_{\mathbb{R}}}^T (\Sigma^{-1} \otimes \Sigma^{-1}) \{(du \otimes I) + (I \otimes du)\} du \\ &= \left[\{(\Sigma^{-1} \otimes \Sigma^{-1}) D_{d_{\mathbb{R}}} e_r\}^T (du \otimes I) + \{(\Sigma^{-1} \otimes \Sigma^{-1}) D_{d_{\mathbb{R}}} e_r\}^T (I \otimes du) \right] du \\ &= (du)^T \left[\text{vec}^{-1} \left((\Sigma^{-1} \otimes \Sigma^{-1}) D_{d_{\mathbb{R}}} e_r \right) + \text{vec}^{-1} \left((\Sigma^{-1} \otimes \Sigma^{-1}) D_{d_{\mathbb{R}}} e_r \right)^T \right] du. \end{aligned}$$

From the second identification theorem of matrix differential calculus (e.g. Magnus & Neudecker, 1999) we then have

$$\nabla^2 \{e_r^T g_{iC}(u)\} = \frac{1}{2} \text{vec}^{-1} \left((\Sigma^{-1} \otimes \Sigma^{-1}) D_{d_{\mathbb{R}}} e_r \right) + \frac{1}{2} \text{vec}^{-1} \left((\Sigma^{-1} \otimes \Sigma^{-1}) D_{d_{\mathbb{R}}} e_r \right)^T$$

which does not depend on u . Therefore $\nabla^2 g_k(U_i^*)$ is a symmetric matrix that depends only on Σ , which we denote as follows:

$$Q(\Sigma; r) \equiv \frac{1}{2} \text{vec}^{-1} \left((\Sigma^{-1} \otimes \Sigma^{-1}) D_{d_{\mathbb{R}}} e_r \right) + \frac{1}{2} \text{vec}^{-1} \left((\Sigma^{-1} \otimes \Sigma^{-1}) D_{d_{\mathbb{R}}} e_r \right)^T.$$

Now note that

$$\begin{aligned} \frac{1}{2n} \text{tr} [\{\nabla^2 h_i(U_i^*)\}^{-1} \nabla^2 \{e_r^T g_{iC}(U_i^*)\}] &= \text{tr} [\{\nabla^2 h_i(U_i^*)\}^{-1} Q(\Sigma; r)] \\ &= \frac{1}{2n} \text{tr} [\{\nabla^2 h_i(U_i)\}^{-1} Q(\Sigma; r)] + O_p(n^{-3/2}) \\ &= \frac{\phi}{2} \text{tr} \{ \mathcal{H}_{AA_i}^{-1} Q(\Sigma; r) \} + O_p(n^{-3/2}). \end{aligned}$$

The r th entry of the leading term of the contribution to S_{C_i} from the third term on the right-hand side

of (S.10) is

$$\begin{aligned}
& \frac{\phi}{2} \text{tr} \{ \mathcal{H}_{AAi}^{-1} Q(\Sigma; r) \} \\
&= \frac{\phi}{4} \text{vec}(\mathcal{H}_{AAi}^{-1})^T \text{vec} \left(\text{vec}^{-1} \left((\Sigma^{-1} \otimes \Sigma^{-1}) D_{d_R} e_r \right) \right) \\
&\quad + \frac{\phi}{4} \text{vec}(\mathcal{H}_{AAi}^{-1})^T \text{vec} \left(\text{vec}^{-1} \left((\Sigma^{-1} \otimes \Sigma^{-1}) D_{d_R} e_r \right)^T \right) \\
&= \frac{\phi}{4} \text{vec}(\mathcal{H}_{AAi}^{-1})^T \text{vec} \left(\text{vec}^{-1} \left((\Sigma^{-1} \otimes \Sigma^{-1}) D_{d_R} e_r \right) \right) \\
&\quad + \frac{\phi}{4} \text{vec}(\mathcal{H}_{AAi}^{-1})^T K_{d_R} \text{vec} \left(\text{vec}^{-1} \left((\Sigma^{-1} \otimes \Sigma^{-1}) D_{d_R} e_r \right) \right) \\
&= \frac{\phi}{4} \text{vec}(\mathcal{H}_{AAi}^{-1})^T (\Sigma^{-1} \otimes \Sigma^{-1}) D_{d_R} e_r + \frac{\phi}{4} \text{vec}(\mathcal{H}_{AAi}^{-1})^T K_{d_R} (\Sigma^{-1} \otimes \Sigma^{-1}) D_{d_R} e_r \\
&= \frac{\phi}{4} \text{vec}(\mathcal{H}_{AAi}^{-1})^T (\Sigma^{-1} \otimes \Sigma^{-1}) D_{d_R} e_r + \frac{\phi}{4} \text{vec}(\mathcal{H}_{AAi}^{-1})^T (\Sigma^{-1} \otimes \Sigma^{-1}) K_{d_R} D_{d_R} e_r \\
&= \frac{\phi}{4} \text{vec}(\mathcal{H}_{AAi}^{-1})^T (\Sigma^{-1} \otimes \Sigma^{-1}) D_{d_R} e_r + \frac{\phi}{4} \text{vec}(\mathcal{H}_{AAi}^{-1})^T (\Sigma^{-1} \otimes \Sigma^{-1}) D_{d_R} e_r \\
&= \frac{\phi}{2} e_r^T D_{d_R}^T (\Sigma^{-1} \otimes \Sigma^{-1}) \text{vec}(\mathcal{H}_{AAi}^{-1}) = \frac{\phi}{2} e_r^T D_{d_R}^T \text{vec}(\Sigma^{-1} \mathcal{H}_{AAi}^{-1} \Sigma^{-1}).
\end{aligned}$$

Hence, contribution to S_{C_i} from the third term on the right-hand side of (S.10) is

$$\frac{\phi}{2} D_{d_R}^T \text{vec}(\Sigma^{-1} \mathcal{H}_{AAi}^{-1} \Sigma^{-1}) + O_p(n^{-3/2}) 1_{d_R(d_R+1)/2}.$$

S.9.3.4 The (S.10) Fourth Term Contribution

For each $1 \leq r \leq \frac{1}{2} d_R(d_R + 1)$, the r th entry of the contribution to S_{C_i} from the fourth term on the right-hand side of (S.10) is

$$-\frac{1}{2n} [\nabla \{ e_r^T g_{iC}(U_i^*) \}]^T \{ \nabla^2 h_i(U_i^*) \}^{-1} [\nabla^3 h_i(U_i^*) \star \{ \nabla^2 h_i(U_i^*) \}^{-1}].$$

Noting (S.22) and using (S.7), it follows that the contribution to S_{C_i} from the fourth term on the right-hand side of (S.10) is

$$\begin{aligned}
& -\frac{1}{4n} D_{d_R}^T (\Sigma^{-1} \otimes \Sigma^{-1}) \{ (U_i^* \otimes I) + (I \otimes U_i^*) \} \{ \nabla^2 h_i(U_i^*) \}^{-1} [\nabla^3 h_i(U_i^*) \star \{ \nabla^2 h_i(U_i^*) \}^{-1}] \\
&= -\frac{1}{4n} D_{d_R}^T [\{ (\Sigma^{-1} U_i^*) \otimes \Sigma^{-1} \} + \{ \Sigma^{-1} \otimes (\Sigma^{-1} U_i^*) \}] \\
&\quad \times \{ \nabla^2 h_i(U_i^*) \}^{-1} [\nabla^3 h_i(U_i^*) \star \{ \nabla^2 h_i(U_i^*) \}^{-1}] \\
&= -\frac{\phi}{4} D_{d_R}^T [\{ (\Sigma^{-1} U_i) \otimes \Sigma^{-1} \} + \{ \Sigma^{-1} \otimes (\Sigma^{-1} U_i) \}] \mathcal{H}_{AAi}^{-1} \left(\mathcal{H}'_{AAAi} \star \mathcal{H}_{AAi}^{-1} \right) + O_p(n^{-3/2}) 1_{d_R(d_R+1)/2} \\
&= -\frac{\phi}{2} D_{d_R}^T \text{vec} \left(\Sigma^{-1} \left\{ \mathcal{H}_{AAi}^{-1} \left(\mathcal{H}'_{AAAi} \star \mathcal{H}_{AAi}^{-1} \right) U_i^T \right\} \Sigma^{-1} \right) + O_p(n^{-3/2}) 1_{d_R(d_R+1)/2}
\end{aligned}$$

where the last step follows from application of (S.3) and (S.6).

S.9.3.5 The Resultant Score Approximation

The resultant approximation of S_{C_i} is

$$\begin{aligned}
S_{C_i} = & \frac{1}{2} D_{d_R}^T \text{vec} \left(\Sigma^{-1} \left[U_i U_i^T - \Sigma + 2 \mathcal{H}_{AAi}^{-1} \mathcal{G}_{Ai} U_i^T + \mathcal{H}_{AAi}^{-1} \mathcal{G}_{Ai} \mathcal{G}_{Ai}^T \mathcal{H}_{AAi}^{-1} \right. \right. \\
& \left. \left. + \phi \mathcal{H}_{AAi}^{-1} - 2 \phi \mathcal{H}_{AAi}^{-1} \Sigma^{-1} U_i U_i^T \right. \right. \\
& \left. \left. - \mathcal{H}_{AAi}^{-1} \left\{ \mathcal{H}'_{AAAi} \star \left(\mathcal{H}_{AAi}^{-1} \mathcal{G}_{Ai} \mathcal{G}_{Ai}^T \mathcal{H}_{AAi}^{-1} \right) \right\} U_i^T \right. \right. \\
& \left. \left. - \phi \mathcal{H}_{AAi}^{-1} \left(\mathcal{H}'_{AAAi} \star \mathcal{H}_{AAi}^{-1} \right) U_i^T \right] \Sigma^{-1} \right) + O_p(n^{-3/2}) 1_{d_R(d_R+1)/2}.
\end{aligned} \tag{S.23}$$

S.10 Score Outer Product Conditional Moments Approximation

The i th term of the Fisher information matrix of $(\beta, \text{vech}(\Sigma))$ is a 3×3 block partitioned matrix with the blocks corresponding to the various moments of pairwise outer products, conditional on X_i . The relevant approximations involve repeated use of (S.14) and keeping track of orders of magnitude.

S.10.1 Approximation of $E(S_{A_i}^{\otimes 2} | X_i)$

Using (S.18), (S.14) and standard algebraic steps we have

$$\begin{aligned}
E(S_{A_i}^{\otimes 2} | X_i) = & \Sigma^{-1} \\
& + \phi \Sigma^{-1} E \left(\mathcal{H}_{AAi}^{-1} - U_i U_i^T \Sigma^{-1} \mathcal{H}_{AAi}^{-1} - \mathcal{H}_{AAi}^{-1} \Sigma^{-1} U_i U_i^T \mid X_i \right) \Sigma^{-1} \\
& - \phi \Sigma^{-1} E \left\{ \mathcal{H}_{AAi}^{-1} \left(\mathcal{H}'_{AAAi} \star \mathcal{H}_{AAi}^{-1} \right) U_i^T + U_i \left(\mathcal{H}'_{AAAi} \star \mathcal{H}_{AAi}^{-1} \right)^T \mathcal{H}_{AAi}^{-1} \mid X_i \right\} \Sigma^{-1} \\
& + O_p(n^{-2}) 1_{d_R}^{\otimes 2}.
\end{aligned} \tag{S.24}$$

S.10.2 Approximation of $E(S_{B_i}^{\otimes 2} | X_i)$

From (S.21) and (S.14) we obtain

$$E(S_{B_i}^{\otimes 2} | X_i) = \frac{1}{\phi} E \left(\mathcal{H}_{BBi} - \mathcal{H}_{ABi}^T \mathcal{H}_{AAi}^{-1} \mathcal{H}_{ABi} \mid X_i \right) + O_p(1) 1_{d_B}^{\otimes 2}. \tag{S.25}$$

S.10.3 Approximation of $E(S_{C_i}^{\otimes 2} | X_i)$

After some long-winded, but relatively straightforward, matrix algebra that involves application of (S.14) we have from (S.23) that

$$\begin{aligned}
E(S_{C_i}^{\otimes 2} | X_i) = & \frac{1}{2} D_{d_R}^T (\Sigma^{-1} \otimes \Sigma^{-1}) D_{d_R} + \frac{\phi}{2} D_{d_R}^T (\Sigma^{-1} \otimes \Sigma^{-1}) E \left[2(U_i U_i^T) \otimes \mathcal{H}_{AAi}^{-1} \right. \\
& \left. + \text{vec}(U_i U_i^T - \Sigma) \text{vec} \left(\mathcal{H}_{AAi}^{-1} \Sigma^{-1} \left\{ \Sigma - U_i U_i^T - \Sigma \left(\mathcal{H}'_{AAAi} \star \mathcal{H}_{AAi}^{-1} \right) U_i^T \right\} \right)^T \right. \\
& \left. + \text{vec} \left(\mathcal{H}_{AAi}^{-1} \Sigma^{-1} \left\{ \Sigma - U_i U_i^T - \Sigma \left(\mathcal{H}'_{AAAi} \star \mathcal{H}_{AAi}^{-1} \right) U_i^T \right\} \right) \text{vec}(U_i U_i^T - \Sigma)^T \mid X_i \right] \\
& \times (\Sigma^{-1} \otimes \Sigma^{-1}) D_{d_R} + O_p(n^{-2}) 1_{d_R(d_R+1)/2}^{\otimes 2}.
\end{aligned} \tag{S.26}$$

S.10.4 Approximation of $E(S_{A_i} S_{B_i}^T | X_i)$

Multiplication of (S.18) by the transpose of (S.21), taking expectations conditional on X_i and use of (S.14) leads to

$$\begin{aligned}
E(S_{A_i} S_{B_i}^T | X_i) = & \Sigma^{-1} E \left\{ U_i U_i^T \Sigma^{-1} \mathcal{H}_{AAi}^{-1} \mathcal{H}_{ABi} - U_i \left(\mathcal{H}'_{AABi} \star \mathcal{H}_{AAi}^{-1} \right)^T \right. \\
& \left. + U_i \left(\mathcal{H}'_{AAAi} \star \mathcal{H}_{AAi}^{-1} \right)^T \mathcal{H}_{AAi}^{-1} \mathcal{H}_{ABi} \mid X_i \right\} + O_p(n^{-1}) 1_{d_R} 1_{d_B}^T.
\end{aligned} \tag{S.27}$$

S.10.5 Approximation of $E(S_{A_i}S_{C_i}^T|X_i)$

An important aspect of the $E(S_{A_i}S_{C_i}^T|X_i)$ approximation is that, even though

$$S_{A_i} = O_p(1)1_{d_R} \quad \text{and} \quad S_{C_i} = O_p(1)1_{d_R(d_R+1)/2}$$

we can establish that

$$E(S_{A_i}S_{C_i}^T|X_i) = O_p(n^{-1})1_{d_R}1_{d_R(d_R+1)/2}^T, \quad (\text{S.28})$$

which indicates a degree of asymptotic orthogonality between β_A and Σ . An illustrative cancellation, involving the leading terms of each score, is

$$E\{\Sigma^{-1}U_i\Sigma^{-1}(U_iU_i^T - \Sigma)\Sigma^{-1}|X_i\}D_{d_R} = \Sigma^{-1}U_i\Sigma^{-1}\{E(U_iU_i^T|X_i) - \Sigma\}\Sigma^{-1}D_{d_R} = O.$$

As will be shown in § S.12, approximation (S.28) is sufficient for (7).

S.10.6 Approximation of $E(S_{B_i}S_{C_i}^T|X_i)$

Multiplication of (S.21) by the transpose of (S.23), and similar arguments, leads to

$$\begin{aligned} E(S_{B_i}S_{C_i}^T|X_i) &= \frac{1}{2}E\left[\left\{\mathcal{H}_{ABi}^T\mathcal{H}_{AAi}^{-1}\Sigma^{-1}U_i - \mathcal{H}'_{AABi}\star\mathcal{H}_{AAi}^{-1}\right.\right. \\ &\quad \left.\left. + \mathcal{H}_{ABi}^T\mathcal{H}_{AAi}^{-1}\left(\mathcal{H}'_{AAAi}\star\mathcal{H}_{AAi}^{-1}\right)\right\}\text{vec}(U_iU_i^T - \Sigma)^T|X_i\right](\Sigma^{-1} \otimes \Sigma^{-1})D_{d_R} \\ &\quad + O_p(n^{-1})1_{d_B}1_{d_R(d_R+1)/2}^T. \end{aligned} \quad (\text{S.29})$$

S.11 The Fisher Information Matrix

The Fisher information matrix of $(\beta, \text{vech}(\Sigma))$ is

$$\begin{aligned} I(\beta, \text{vech}(\Sigma)) &= \begin{bmatrix} M_{11} & M_{12} \\ M_{12}^T & M_{22} \end{bmatrix} \quad \text{where} \quad M_{11} \equiv \sum_{i=1}^m \begin{bmatrix} E(S_{A_i}^{\otimes 2}|X_i) & E(S_{A_i}S_{B_i}^T|X_i) \\ E(S_{B_i}S_{A_i}^T|X_i) & E(S_{B_i}^{\otimes 2}|X_i) \end{bmatrix}, \\ M_{12} &\equiv \sum_{i=1}^m \begin{bmatrix} E(S_{A_i}S_{C_i}^T|X_i) \\ E(S_{B_i}S_{C_i}^T|X_i) \end{bmatrix} \quad \text{and} \quad M_{22} \equiv \sum_{i=1}^m E(S_{C_i}^{\otimes 2}|X_i). \end{aligned}$$

The results of the previous section lead to high-order asymptotic approximation of the matrix $I(\beta, \text{vech}(\Sigma))$. In the next section we show that inversion of this approximate Fisher information matrix leads to two-term covariance matrix approximations for the maximum likelihood estimators.

S.12 Approximation of Covariance Matrices of Estimators

The dominant terms in the approximation of

$$\text{Cov}(\hat{\beta}|\mathcal{X}) \quad \text{and} \quad \text{Cov}(\text{vech}(\hat{\Sigma})|\mathcal{X})$$

correspond to the $d_F \times d_F$ and $\frac{1}{2}d_R(d_R + 1) \times \frac{1}{2}d_R(d_R + 1)$ diagonal blocks of

$$I(\beta, \text{vech}(\Sigma))^{-1}.$$

We now treat each of these in turn in the upcoming subsections, which make extensive use of block matrix inversions. If a matrix is partitioned into four blocks A, B, C and D , then

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix}^{-1} = \begin{bmatrix} A^{-1} + A^{-1}B(D - CA^{-1}B)^{-1}CA^{-1} & -A^{-1}B(D - CA^{-1}B)^{-1} \\ -(D - CA^{-1}B)^{-1}CA^{-1} & (D - CA^{-1}B)^{-1} \end{bmatrix}$$

or, equivalently,

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

Another result that is repeatedly used in the following subsections is

$$(A - B)^{-1} = \sum_{k=0}^{\infty} (A^{-1}B)^k A^{-1}$$

for A and B invertible matrices of the same size and such that the spectral radius of $A^{-1}B$ is less than 1.

S.12.1 Two-Term Approximation of $\text{Cov}(\hat{\beta}|\mathcal{X})$

The dominant terms of $\text{Cov}(\hat{\beta}|\mathcal{X})$ correspond to

$$\text{the upper left } d_f \times d_f \text{ block of } I(\beta, \text{vech}(\Sigma))^{-1} = (M_{11} - M_{12}M_{22}^{-1}M_{12}^T)^{-1}.$$

Based on (S.24), (S.25) and (S.27) we have

$$M_{11} = \begin{bmatrix} m\Sigma^{-1} - \frac{\phi m}{n}\Sigma^{-1}\mathcal{K}_{AA}\Sigma^{-1} + O_p(mn^{-2})1_{d_R}^{\otimes 2} & m\Sigma^{-1}\mathcal{K}_{AB} + O_p(mn^{-1})1_{d_R}1_{d_B}^T \\ m\Sigma^{-1}\mathcal{K}_{AB}^T + O_p(mn^{-1})1_{d_B}1_{d_R}^T & \frac{mn}{\phi}\mathcal{K}_{BB} + O_p(m)1_{d_B}^{\otimes 2} \end{bmatrix}$$

where

$$\begin{aligned} \mathcal{K}_{AA} \equiv & \frac{n}{m} \sum_{i=1}^m E \left\{ U_i U_i^T \Sigma^{-1} \mathcal{H}_{AAi}^{-1} + \mathcal{H}_{AAi}^{-1} \Sigma^{-1} U_i U_i^T - \mathcal{H}_{AAi}^{-1} \right. \\ & \left. + \mathcal{H}_{AAi}^{-1} \left(\mathcal{H}'_{AAAi} \star \mathcal{H}_{AAi}^{-1} \right) U_i^T + U_i \left(\mathcal{H}'_{AAAi} \star \mathcal{H}_{AAi}^{-1} \right)^T \mathcal{H}_{AAi}^{-1} \middle| X_i \right\}, \end{aligned}$$

$$\begin{aligned} \mathcal{K}_{AB} \equiv & \frac{1}{m} \sum_{i=1}^m E \left\{ U_i U_i^T \Sigma^{-1} \mathcal{H}_{AAi}^{-1} \mathcal{H}_{ABi} - U_i \left(\mathcal{H}'_{AABi} \star \mathcal{H}_{AAi}^{-1} \right)^T \right. \\ & \left. + U_i \left(\mathcal{H}'_{AAAi} \star \mathcal{H}_{AAi}^{-1} \right)^T \mathcal{H}_{AAi}^{-1} \mathcal{H}_{ABi} \middle| X_i \right\} \end{aligned}$$

$$\text{and } \mathcal{K}_{BB} \equiv \frac{1}{mn} \sum_{i=1}^m E \left(\mathcal{H}_{BBi} - \mathcal{H}_{ABi}^T \mathcal{H}_{AAi}^{-1} \mathcal{H}_{ABi} \middle| X_i \right)$$

are matrices with all entries being $O_p(1)$. As consequences of (S.26), (S.28) and (S.29) we have

$$M_{22}^{-1} = O_p(m^{-1})1_{d_R(d_R+1)/2}^{\otimes 2} \quad \text{and} \quad M_{12} = \begin{bmatrix} O_p(mn^{-1})1_{d_R}1_{d_R(d_R+1)/2}^T \\ O_p(m)1_{d_B}1_{d_R(d_R+1)/2}^T \end{bmatrix}. \quad (\text{S.30})$$

Therefore

$$M_{12}M_{22}^{-1}M_{12}^T = \begin{bmatrix} O_p(mn^{-2})1_{d_R}^{\otimes 2} & O_p(mn^{-1})1_{d_R}1_{d_B}^T \\ O_p(mn^{-1})1_{d_B}1_{d_R}^T & O_p(m)1_{d_B}^{\otimes 2} \end{bmatrix}.$$

From these results for M_{11} and $M_{12}M_{22}^{-1}M_{12}^T$, it follows that

$$\begin{aligned} M_{11} - M_{12}M_{22}^{-1}M_{12}^T = & \\ \begin{bmatrix} m\Sigma^{-1} - \frac{\phi m}{n}\Sigma^{-1}\mathcal{K}_{AA}\Sigma^{-1} + O_p(mn^{-2})1_{d_R}^{\otimes 2} & m\Sigma^{-1}\mathcal{K}_{AB} + O_p(mn^{-1})1_{d_R}1_{d_B}^T \\ m\Sigma^{-1}\mathcal{K}_{AB}^T + O_p(mn^{-1})1_{d_B}1_{d_R}^T & \frac{mn}{\phi}\mathcal{K}_{BB} + O_p(m)1_{d_B}^{\otimes 2} \end{bmatrix} & \end{aligned}$$

The upper left $d_R \times d_R$ block of $(M_{11} - M_{12}M_{22}^{-1}M_{12}^T)^{-1}$ is

$$\begin{aligned} & \left\{ m\Sigma^{-1} - \frac{\phi m}{n}\Sigma^{-1}(\mathcal{K}_{AA} + \mathcal{K}_{AB}\mathcal{K}_{BB}^{-1}\mathcal{K}_{AB}^T)\Sigma^{-1} + O_p(mn^{-2})1_{d_R}^{\otimes 2} \right\}^{-1} \\ &= \frac{1}{m} \left\{ I - \frac{\phi}{n}(\mathcal{K}_{AA} + \mathcal{K}_{AB}\mathcal{K}_{BB}^{-1}\mathcal{K}_{AB}^T)\Sigma^{-1} + O_p(n^{-2})1_{d_R}^{\otimes 2} \right\}^{-1}\Sigma \\ &= \frac{\Sigma}{m} + \frac{\phi(\mathcal{K}_{AA} + \mathcal{K}_{AB}\mathcal{K}_{BB}^{-1}\mathcal{K}_{AB}^T)}{mn} + O_p(m^{-1}n^{-2})1_{d_R}^{\otimes 2}. \end{aligned}$$

The upper right $d_R \times d_B$ block of $(M_{11} - M_{12}M_{22}^{-1}M_{12}^T)^{-1}$ is

$$\begin{aligned} & - \left\{ \frac{\Sigma}{m} + O_p(m^{-1}n^{-1})1_{d_R}^{\otimes 2} \right\} \left\{ m\Sigma^{-1}\mathcal{K}_{AB} + O_p(mn^{-1})1_{d_R}1_{d_B}^T \right\} \left\{ \frac{mn}{\phi}\mathcal{K}_{BB} + O_p(m)1_{d_B}^{\otimes 2} \right\}^{-1} \\ &= -\frac{\phi\mathcal{K}_{AB}\mathcal{K}_{BB}^{-1}}{mn} + O_p(m^{-1}n^{-2})1_{d_R}1_{d_B}^T. \end{aligned}$$

The lower right $d_B \times d_B$ block of $(M_{11} - M_{12}M_{22}^{-1}M_{12}^T)^{-1}$ is

$$\frac{\phi\mathcal{K}_{BB}^{-1}}{mn} + O_p(m^{-1}n^{-2})1_{d_B}^{\otimes 2}.$$

Therefore,

$$\begin{aligned} \text{Cov}(\widehat{\beta}|\mathcal{X}) &= \frac{1}{m} \begin{bmatrix} \Sigma^0 & O \\ O & O \end{bmatrix} + \frac{\phi}{mn} \begin{bmatrix} (\mathcal{K}_{AA}^0)^{-1} & (\mathcal{K}_{AA}^0)^{-1}\mathcal{K}_{AB}^0 \\ (\mathcal{K}_{AB}^0)^T(\mathcal{K}_{AA}^0)^{-1} & \mathcal{K}_{BB}^0 + (\mathcal{K}_{AB}^0)^T(\mathcal{K}_{AA}^0)^{-1}\mathcal{K}_{AB}^0 \end{bmatrix}^{-1} \\ & \quad + O_p(m^{-1}n^{-2})1_{d_F}^{\otimes 2} \end{aligned}$$

where, for example, \mathcal{K}_{AA}^0 is the \mathcal{K}_{AA} quantity with β set to β^0 and Σ set to Σ^0 .

S.12.2 Two-Term Approximation of $\text{Cov}(\text{vech}(\widehat{\Sigma})|\mathcal{X})$

The dominant terms of $\text{Cov}(\text{vech}(\widehat{\Sigma})|\mathcal{X})$ correspond to

$$\begin{aligned} & \text{the lower right } \frac{1}{2}d_R(d_R + 1) \times \frac{1}{2}d_R(d_R + 1) \text{ block of } I(\beta, \text{vech}(\Sigma))^{-1} \\ &= (M_{22} - M_{12}^T M_{11}^{-1} M_{12})^{-1}. \end{aligned}$$

From (S.26)

$$\begin{aligned} \sum_{i=1}^m E(S_{Ci}S_{Ci}^T|X_i) &= \frac{m}{2}D_{d_R}^T(\Sigma^{-1} \otimes \Sigma^{-1})D_{d_R} - \frac{m\phi}{n}D_{d_R}^T(\Sigma^{-1} \otimes \Sigma^{-1})\mathcal{K}_{CC}(\Sigma^{-1} \otimes \Sigma^{-1})D_{d_R} \\ & \quad + O_p(mn^{-2})1_{d_R(d_R+1)/2}^{\otimes 2} \end{aligned}$$

with the following $O_p(1)$ matrix:

$$\begin{aligned} \mathcal{K}_{CC} &= \frac{n}{m} \sum_{i=1}^m E \left[\frac{1}{2} \text{vec}(\Sigma - U_i U_i^T) \text{vec} \left(\mathcal{H}_{AAi}^{-1} \Sigma^{-1} \left\{ \Sigma - U_i U_i^T - \Sigma \left(\mathcal{H}'_{AAAi} \star \mathcal{H}_{AAi}^{-1} \right) U_i^T \right\} \right)^T \right. \\ & \quad \left. + \frac{1}{2} \text{vec} \left(\mathcal{H}_{AAi}^{-1} \Sigma^{-1} \left\{ \Sigma - U_i U_i^T - \Sigma \left(\mathcal{H}'_{AAAi} \star \mathcal{H}_{AAi}^{-1} \right) U_i^T \right\} \right) \text{vec}(\Sigma - U_i U_i^T)^T \right. \\ & \quad \left. - (U_i U_i^T) \otimes (\mathcal{H}_{AAi}^{-1}) \middle| X_i \right]. \end{aligned}$$

Next, note that

$$M_{11}^{-1} = \begin{bmatrix} O_p(m^{-1})1_{d_R}^{\otimes 2} & O_p\{(mn)^{-1}\}1_{d_R}1_{d_B}^T \\ O_p\{(mn)^{-1}\}1_{d_B}1_{d_R}^T & O_p\{(mn)^{-1}\}1_{d_B}^{\otimes 2} \end{bmatrix}. \quad (\text{S.31})$$

Given the orders of magnitude in (S.30) and (S.31), from expansion of $M_{12}^T M_{11}^{-1} M_{12}$ it is apparent that its dominant $O_p(m/n)$ contribution is from

$$\begin{aligned} & \left\{ \sum_{i=1}^m E(S_{B_i} S_{C_i}^T | X_i) \right\}^T \left\{ \sum_{i=1}^m E(S_{B_i} S_{B_i}^T | X_i) \right\}^{-1} \sum_{i=1}^m E(S_{B_i} S_{C_i}^T | X_i) \\ &= \frac{\phi m}{n} D_{d_R}^T (\Sigma^{-1} \otimes \Sigma^{-1}) \mathcal{K}_{BC} \mathcal{K}_{BB}^{-1} \mathcal{K}_{BC} (\Sigma^{-1} \otimes \Sigma^{-1}) D_{d_R} + o_p(m/n) 1_{d_R(d_R+1)/2}^{\otimes 2} \end{aligned}$$

where

$$\begin{aligned} \mathcal{K}_{BC} \equiv & \frac{1}{2m} \sum_{i=1}^m E \left[\left\{ \mathcal{H}_{AB_i}^T \mathcal{H}_{AA_i}^{-1} \Sigma^{-1} U_i - \mathcal{H}'_{AAB_i} \star \mathcal{H}_{AA_i}^{-1} \right. \right. \\ & \left. \left. + \mathcal{H}_{AB_i}^T \mathcal{H}_{AA_i}^{-1} \left(\mathcal{H}'_{AAA_i} \star \mathcal{H}_{AA_i}^{-1} \right) \right\} \middle| X_i \right] \text{vec}(U_i U_i^T - \Sigma)^T \end{aligned}$$

is a matrix with all entries being $O_p(1)$. Hence, if we let $\mathcal{A} \equiv \frac{1}{2} D_{d_R}^T ((\Sigma^0)^{-1} \otimes (\Sigma^0)^{-1}) D_{d_R}$ and

$$\mathcal{B} \equiv D_{d_R}^T ((\Sigma^0)^{-1} \otimes (\Sigma^0)^{-1}) \{ \mathcal{K}_{CC}^0 + (\mathcal{K}_{BC}^0)^T (\mathcal{K}_{BB}^0)^{-1} \mathcal{K}_{BC}^0 \} ((\Sigma^0)^{-1} \otimes (\Sigma^0)^{-1}) D_{d_R}$$

then

$$\begin{aligned} \text{Cov}(\text{vech}(\widehat{\Sigma}) | \mathcal{X}) &= \frac{1}{m} \left(\mathcal{A} - \frac{\phi}{n} \mathcal{B} \right)^{-1} + o_p\{(mn)^{-1}\} 1_{d_R(d_R+1)/2}^{\otimes 2} \\ &= \frac{1}{m} \mathcal{A}^{-1} + \frac{\phi}{mn} \mathcal{A}^{-1} \mathcal{B} \mathcal{A}^{-1} + o_p\{(mn)^{-1}\} 1_{d_R(d_R+1)/2}^{\otimes 2}. \end{aligned}$$

From (S.5),

$$\mathcal{A}^{-1} = 2D_{d_R}^+ (\Sigma^0 \otimes \Sigma^0) D_{d_R}^{+T}.$$

To simplify $\mathcal{A}^{-1} \mathcal{B} \mathcal{A}^{-1}$, we use (S.4) and (S.7) to obtain

$$\begin{aligned} \text{Cov}(\text{vech}(\widehat{\Sigma}) | \mathcal{X}) &= \frac{2D_{d_R}^+ (\Sigma^0 \otimes \Sigma^0) D_{d_R}^{+T}}{m} + \frac{4\phi D_{d_R}^+ \{ \mathcal{K}_{CC}^0 + (\mathcal{K}_{BC}^0)^T (\mathcal{K}_{BB}^0)^{-1} \mathcal{K}_{BC}^0 \} D_{d_R}^{+T}}{mn} \\ &\quad + O_p(m^{-1} n^{-2}) 1_{d_R(d_R+1)/2}^{\otimes 2}. \end{aligned} \quad (\text{S.32})$$

S.13 Population Forms of Covariance Matrix Second Terms

In the previous section, the second terms of the asymptotic covariance matrices of $\widehat{\beta}$ and $\text{vech}(\widehat{\Sigma})$ are stochastic. However, under relatively mild moment conditions such as assumption (A3) of Jiang *et al.* (2022), these terms converge in probability to deterministic population forms. In this section we determine these limiting forms.

A re-writing of the \mathcal{K}_{AA} quantity is

$$\begin{aligned} \mathcal{K}_{AA} \equiv & \frac{1}{m} \sum_{i=1}^m E \left[U_i U_i^T \Sigma^{-1} \left(\frac{1}{n} \mathcal{H}_{AA_i} \right)^{-1} + \left(\frac{1}{n} \mathcal{H}_{AA_i} \right)^{-1} \Sigma^{-1} U_i U_i^T - \left(\frac{1}{n} \mathcal{H}_{AA_i} \right)^{-1} \right. \\ & \left. + \left(\frac{1}{n} \mathcal{H}_{AA_i} \right)^{-1} \left\{ \left(\frac{1}{n} \mathcal{H}'_{AAA_i} \right) \star \left(\frac{1}{n} \mathcal{H}_{AA_i} \right)^{-1} \right\} U_i^T \right. \\ & \left. + U_i \left\{ \left(\frac{1}{n} \mathcal{H}'_{AAA_i} \right) \star \left(\frac{1}{n} \mathcal{H}_{AA_i} \right)^{-1} \right\}^T \left(\frac{1}{n} \mathcal{H}_{AA_i} \right)^{-1} \middle| X_i \right]. \end{aligned}$$

Since

$$E\left(\frac{1}{n}\mathcal{H}_{AAi}|X_i\right) \xrightarrow{P} \Omega_{AA}(U_i) \quad \text{and} \quad E\left(\frac{1}{n}\mathcal{H}'_{AAAi}|X_i\right) \xrightarrow{P} \Omega'_{AAA}(U_i)$$

we have, under relatively mild conditions (see e.g. Lemma A1 of Jiang *et al.*, 2022),

$$\begin{aligned} \mathcal{K}_{AA} &\xrightarrow{P} \frac{1}{m} \sum_{i=1}^m E \left[U_i U_i^T \Sigma^{-1} \Omega_{AA}(U_i)^{-1} + \Omega_{AA}(U_i)^{-1} \Sigma^{-1} U_i U_i^T - \Omega_{AA}(U_i)^{-1} \right. \\ &\quad \left. + \Omega_{AA}(U_i)^{-1} \left\{ \Omega'_{AAA}(U_i) \star \Omega_{AA}(U_i)^{-1} \right\} U_i^T \right. \\ &\quad \left. + U_i \left\{ \Omega'_{AAA}(U_i) \star \Omega_{AA}(U_i)^{-1} \right\}^T \Omega_{AA}(U_i)^{-1} \right] \\ &= E \left[U U^T \Sigma^{-1} \Omega_{AA}(U)^{-1} + \Omega_{AA}(U)^{-1} \Sigma^{-1} U U^T - \Omega_{AA}(U)^{-1} \right. \\ &\quad \left. + \Omega_{AA}(U)^{-1} \left\{ \Omega'_{AAA}(U) \star \Omega_{AA}(U)^{-1} \right\} U^T + U \left\{ \Omega'_{AAA}(U) \star \Omega_{AA}(U)^{-1} \right\}^T \Omega_{AA}(U)^{-1} \right] \\ &= \Lambda_{AA} \end{aligned}$$

where Λ_{AA} is as defined in § 3.1. Analogous arguments lead to

$$\mathcal{K}_{AB} \xrightarrow{P} \Lambda_{AB}, \quad \mathcal{K}_{BB} \xrightarrow{P} E\{\Psi_6(U)\}, \quad \mathcal{K}_{BC} \xrightarrow{P} \Delta \quad \text{and} \quad D_{d_R}^+ \mathcal{K}_{CC} D_{d_R}^{+T} \xrightarrow{P} \frac{1}{2} E\{\Psi_9(U) - 2\Psi_8(U)\}$$

where $\Psi_8(U)$, $\Psi_9(U)$ and Λ_{AB} are as defined in § 3.1. It follows that the deterministic forms of the order $(mn)^{-1}$ terms match those stated in (7).

S.14 The Gaussian Response Special Case

For the Gaussian response special case of (3) the two-term covariance matrix expressions simplify considerably. The main reason is that, for the Gaussian case, $b''(x) = 1$ and $b'''(x) = 0$. These facts imply that

$$\Omega_{AA}(U) = E(X_A X_A^T), \quad \Omega_{AB}(U) = E(X_A X_B^T), \quad \Omega_{BB}(U) = E(X_B X_B^T)$$

and all entries of the three-dimensional arrays $\Omega'_{AAA}(U)$ and $\Omega'_{AAB}(U)$ are exactly zero.

S.14.1 The Cov($\widehat{\beta}|\mathcal{X}$) Approximation

For the Gaussian response situation

$$\Lambda_{AA} = E(X_A X_A^T)^{-1}, \quad \Lambda_{AB} = E(X_A X_A^T)^{-1} E(X_A X_B^T)$$

and

$$E\{\Psi_6(U)\} = E(X_B X_B^T) - E(X_A X_B^T)^T E(X_A X_A^T)^{-1} E(X_A X_B^T).$$

Therefore,

$$\begin{bmatrix} \Lambda_{AA}^{-1} & \Lambda_{AA}^{-1} \Lambda_{AB} \\ \Lambda_{AB}^T \Lambda_{AA}^{-1} & \Lambda_{AB}^T \Lambda_{AA}^{-1} \Lambda_{AB} + E\{\Psi_6(U)\} \end{bmatrix} = E \begin{bmatrix} X_A X_A^T & X_A X_B^T \\ X_B X_A^T & X_B X_B^T \end{bmatrix} = E(XX^T).$$

Hence, for the Gaussian special case

$$\text{Cov}(\widehat{\beta}|\mathcal{X}) = \frac{1}{m} \begin{bmatrix} \Sigma^0 & O \\ O & O \end{bmatrix} + \frac{\phi\{E(XX^T)\}^{-1}\{1 + o_p(1)\}}{mn}.$$

This result generalises the two-term expansion of $\text{Var}(\widehat{\beta}_A|\mathcal{X})$ provided in § 3.5 of McCulloch *et al.* (2008) for the $d_R = d_B = 1$ and $X_A = 1$ special case.

S.14.2 The Cov(vch($\widehat{\Sigma}$)| \mathcal{X}) Approximation

As shown in, for example, § 4.3 of Wand (2002) there is exact orthogonality between β and Σ in the Gaussian case. This means that $\Delta = O$ and, hence, the second term of Cov(vch($\widehat{\Sigma}$)| \mathcal{X}) is

$$\frac{2\phi}{mn} E \left\{ \text{vech}(\Sigma - UU^T) \psi_4(U)^T + \psi_4(U) \text{vech}(\Sigma - UU^T)^T - 2\Psi_8(U) \right\} \quad (\text{S.33})$$

where $\psi_4(U)$ and $\Psi_8(U)$ simplify to

$$\psi_4(U) = D_{d_R}^+ \text{vec} \left(\{E(X_A X_A^T)\}^{-1} \Sigma^{-1} (\Sigma - UU^T) \right)$$

and

$$\Psi_8(U) = D_{d_R}^+ [(UU^T) \otimes \{E(X_A X_A^T)\}^{-1}] D_{d_R}^{+T}.$$

We immediately have

$$E\{\Psi_8(U)\} = D_{d_R}^+ [\Sigma \otimes \{E(X_A X_A^T)\}^{-1}] D_{d_R}^{+T}.$$

The reduction of the other expectations in (S.33) is less immediate and benefits from Theorem 4.3(iv) of Magnus & Neudecker (1979) as well as (S.2). However, such a pathway leads to

$$E \left\{ \text{vech}(\Sigma - UU^T) \psi_4(U)^T + \psi_4(U) \text{vech}(\Sigma - UU^T)^T \right\} = 4D_{d_R}^+ [\Sigma \otimes \{E(X_A X_A^T)\}^{-1}] D_{d_R}^{+T}.$$

On combining the components of (S.33) we arrive at

$$\text{Cov}(\text{vech}(\widehat{\Sigma})|\mathcal{X}) = \frac{2D_{d_R}^+ (\Sigma^0 \otimes \Sigma^0) D_{d_R}^{+T}}{m} + \frac{4\phi D_{d_R}^+ [\Sigma^0 \otimes \{E(X_A X_A^T)\}^{-1}] D_{d_R}^{+T} \{1 + o_p(1)\}}{mn}.$$

S.15 Derivations for the Sample Size Illustration

We provide supporting calculations for the sample size illustration of Section 4 involving the Poisson-response mixed model (8). The model has a random intercept and a random slope, $d_F = d_R = 2$ and all the fixed effects are paired with a random effect, and so all the ‘‘B’’ quantities defined in Section 3.1 are null in this case. Therefore, deriving the explicit expression of the lower-right element of Λ_{AA} is sufficient for obtaining the two-term asymptotic variance of $\hat{\beta}_1$ required for the sample size calculations of our illustrative example. Notice that Λ_{AA} depends on $\Omega_{AA}(U)$, $\psi_2(U)$ and $\Psi_7(U)$, which are defined in Section 3.1. The vector $\psi_2(U)$ itself depends on $\Omega_{AA}(U)$ and $\Omega'_{AAA}(U)$. In the following we will obtain the explicit expressions of these quantities for the model under examination.

S.15.1 The Explicit Expression of $\Omega_{AA}(U)$ and its Inverse

First notice that, for a constant t ,

$$\begin{aligned} & \text{if } Z \sim \text{Bernoulli}(p) \\ & \text{then } E(e^{tZ}) = (1-p) + pe^t \quad \text{and} \quad E(Ze^{tZ}) = E(Z^2 e^{tZ}) = E(Z^3 e^{tZ}) = pe^t. \end{aligned} \quad (\text{S.34})$$

Making use of (S.34), it follows that for model (8)

$$\begin{aligned} \Omega_{AA}(U) &= E \left(\exp\{\beta_0^0 + U_0 + (\beta_1^0 + U_1)X\} \begin{bmatrix} 1 & X \\ X & X^2 \end{bmatrix} \middle| U \right) \\ &= e^{\beta_0^0 + U_0} \begin{bmatrix} 1-p + pe^{\beta_1^0 + U_1} & pe^{\beta_1^0 + U_1} \\ pe^{\beta_1^0 + U_1} & pe^{\beta_1^0 + U_1} \end{bmatrix} \end{aligned}$$

and the inverse of $\Omega_{AA}(U)$ has the following explicit expression:

$$\Omega_{AA}(U)^{-1} = \frac{e^{-\beta_0^0 - U_0}}{p(1-p)} \begin{bmatrix} p & -p \\ -p & (1-p)e^{-\beta_1^0 - U_1} + p \end{bmatrix}. \quad (\text{S.35})$$

S.15.2 The Explicit Expression of $\psi_2(U)$

The vector $\psi_2(U)$ depends on $\Omega_{AA}(U)$ and $\Omega'_{AAA}(U)$. The $((1, 2), (1, 2), 1)$ slice of $\Omega'_{AAA}(U)$ is

$$[\Omega'_{AAA}(U)]_{(1,2),(1,2),1} = E \left(\exp\{\beta_0^0 + U_0 + (\beta_1^0 + U_1)X\} \left[\begin{array}{cc} 1 & X \\ X & X^2 \end{array} \right] \middle| U \right) = \Omega_{AA}(U).$$

Resorting to (S.34), the $((1, 2), (1, 2), 2)$ component of $\Omega'_{AAA}(U)$ is

$$\begin{aligned} [\Omega'_{AAA}(U)]_{(1,2),(1,2),2} &= E \left(\exp\{\beta_0^0 + U_0 + (\beta_1^0 + U_1)X\} \left[\begin{array}{cc} X & X^2 \\ X^2 & X^3 \end{array} \right] \middle| U \right) \\ &= pe^{\beta_0^0 + \beta_1^0 + U_0 + U_1} \left[\begin{array}{cc} 1 & 1 \\ 1 & 1 \end{array} \right]. \end{aligned}$$

The first component of $\psi_2(U)$ is then

$$\begin{aligned} [\psi_2(U)]_1 &= [\Omega'_{AAA}(U)]_{1,1,1}[\Omega_{AA}(U)^{-1}]_{1,1} + [\Omega'_{AAA}(U)]_{1,2,1}[\Omega_{AA}(U)^{-1}]_{1,2} \\ &\quad + [\Omega'_{AAA}(U)]_{2,1,1}[\Omega_{AA}(U)^{-1}]_{2,1} + [\Omega'_{AAA}(U)]_{2,2,1}[\Omega_{AA}(U)^{-1}]_{2,2} \\ &= \frac{1-p + pe^{\beta_1^0 + U_1}}{1-p} - \frac{2pe^{\beta_1^0 + U_1}}{1-p} + \frac{1-p + pe^{\beta_1^0 + U_1}}{1-p} = 2. \end{aligned}$$

The second component of $\psi_2(U)$ is

$$\begin{aligned} [\psi_2(U)]_2 &= [\Omega'_{AAA}(U)]_{1,1,2}[\Omega_{AA}(U)^{-1}]_{1,1} + [\Omega'_{AAA}(U)]_{1,2,2}[\Omega_{AA}(U)^{-1}]_{1,2} \\ &\quad + [\Omega'_{AAA}(U)]_{2,1,2}[\Omega_{AA}(U)^{-1}]_{2,1} + [\Omega'_{AAA}(U)]_{2,2,2}[\Omega_{AA}(U)^{-1}]_{2,2} \\ &= \frac{pe^{\beta_1^0 + U_1}}{1-p} - \frac{2pe^{\beta_1^0 + U_1}}{1-p} + 1 + \frac{pe^{\beta_1^0 + U_1}}{1-p} = 1. \end{aligned}$$

It follows that

$$\psi_2(U) = \begin{bmatrix} 2 \\ 1 \end{bmatrix}. \quad (\text{S.36})$$

S.15.3 The $(2, 2)$ element of $\Psi_7(U)$

For our scope, only the $(2, 2)$ element of $\Psi_7(U)$ is needed. Let

$$(\Sigma^0)^{-1} = \begin{bmatrix} (\Sigma^0)^{00} & (\Sigma^0)^{01} \\ (\Sigma^0)^{10} & (\Sigma^0)^{11} \end{bmatrix}$$

and notice that

$$\Psi_7(U) = \begin{bmatrix} U_0^2(\Sigma^0)^{00} + U_0U_1(\Sigma^0)^{10} & U_0^2(\Sigma^0)^{01} + U_0U_1(\Sigma^0)^{11} \\ U_1^2(\Sigma^0)^{10} + U_0U_1(\Sigma^0)^{00} & U_1^2(\Sigma^0)^{11} + U_0U_1(\Sigma^0)^{01} \end{bmatrix} \Omega_{AA}(U)^{-1}.$$

From replacement of $\Omega_{AA}(U)^{-1}$ with its explicit expression (S.35) in the previous equation we get that

$$\begin{aligned} [\Psi_7(U)]_{2,2} &= \frac{1}{1-p} [U_1^2\{(\Sigma^0)^{11} - (\Sigma^0)^{10}\} + U_0U_1\{(\Sigma^0)^{01} - (\Sigma^0)^{00}\}] e^{-\beta_0^0 - U_0} \\ &\quad + \frac{1}{p} \{U_1^2(\Sigma^0)^{11} + U_0U_1(\Sigma^0)^{01}\} e^{-\beta_0^0 - \beta_1^0 - U_0 - U_1}. \end{aligned} \quad (\text{S.37})$$

S.15.4 The (2, 2) element of Λ_{AA}

Recall that

$$\Lambda_{AA} \equiv E\left\{\Psi_7(U) + \Psi_7(U)^T - \Omega_{AA}(U)^{-1} + \Omega_{AA}(U)^{-1}\psi_2(U)U^T + U\psi_2(U)^T\Omega_{AA}(U)^{-1}\right\}.$$

Using (S.35), (S.36) and simple algebraic steps, the (2, 2) element of the product $\Omega_{AA}(U)^{-1}\psi_2(U)U^T$ is

$$[\Omega_{AA}(U)^{-1}\psi_2(U)U^T]_{2,2} = \frac{U_1}{p}e^{-\beta_0^0 - \beta_1^0 - U_0 - U_1} - \frac{U_1}{1-p}e^{-\beta_0^0 - U_0}. \quad (\text{S.38})$$

Combining (S.35), (S.37) and (S.38) together we get that the (2, 2) element of Λ_{AA} is

$$\begin{aligned} [\Lambda_{AA}]_{2,2} = E & \left(\frac{2}{1-p} [U_1^2 \{(\Sigma^0)^{11} - (\Sigma^0)^{10}\} + U_0 U_1 \{(\Sigma^0)^{01} - (\Sigma^0)^{00}\}] e^{-\beta_0^0 - U_0} \right. \\ & + \frac{2}{p} \{U_1^2 (\Sigma^0)^{11} + U_0 U_1 (\Sigma^0)^{01}\} e^{-\beta_0^0 - \beta_1^0 - U_0 - U_1} - \frac{1}{p} e^{-\beta_0^0 - \beta_1^0 - U_0 - U_1} \\ & \left. - \frac{1}{1-p} e^{-\beta_0^0 - U_0} + \frac{2U_1}{p} e^{-\beta_0^0 - \beta_1^0 - U_0 - U_1} - \frac{2U_1}{1-p} e^{-\beta_0^0 - U_0} \right). \quad (\text{S.39}) \end{aligned}$$

Obtaining an explicit form for (S.39) involves computing $E(e^{-U_0})$ and $E(e^{-U_0 - U_1})$ with respect to

$$\begin{bmatrix} U_0 \\ U_1 \end{bmatrix} \sim N(0, \Sigma^0).$$

Let $f_2(U; 0, \Sigma^0)$ denote the bivariate Normal density function of $U = [U_0 \ U_1]^T$ with mean vector 0 and covariance matrix Σ^0 . Using the moment generating function of the Multivariate Normal distribution we get

$$E(e^{-U_0}) = \exp\left(\frac{\Sigma_{00}^0}{2}\right) \quad \text{and} \quad E(e^{-U_0 - U_1}) = \exp\left\{\frac{1}{2}(\Sigma_{00}^0 + 2\Sigma_{01}^0 + \Sigma_{11}^0)\right\}. \quad (\text{S.40})$$

From (S.39), other quantities that need to be computed to obtain the explicit expression of $[\Lambda_{AA}]_{2,2}$ are

$$E(U_1 e^{-U_0}), \ E(U_1^2 e^{-U_0}), \ E(U_0 U_1 e^{-U_0}), \ E(U_1 e^{-U_0 - U_1}), \ E(U_1^2 e^{-U_0 - U_1}) \ \text{and} \ E(U_0 U_1 e^{-U_0 - U_1}).$$

For computing these expectations it is useful to work with

$$U_0|U_1 \sim N(\mu_{0|1}, \sigma_{0|1}^2) \quad \text{where} \quad \mu_{0|1} = \frac{\Sigma_{01}^0}{\Sigma_{11}^0} U_1 \quad \text{and} \quad \sigma_{0|1}^2 = \Sigma_{00}^0 - \frac{(\Sigma_{01}^0)^2}{\Sigma_{11}^0}$$

and write the joint density of U_0 and U_1 as $f_2(U; 0, \Sigma^0) = f_1(U_0|U_1; \mu_{0|1}, \sigma_{0|1}^2) f_1(U_1; 0, \Sigma_{11}^0)$, where $f_1(Z; \mu, \sigma^2)$ denotes the density function of a random variable Z with a $N(\mu, \sigma^2)$ distribution. Also notice that

$$\text{if } Z \sim N(\mu, \sigma^2) \text{ then } E(e^{-Z}) = -\mu + \frac{1}{2}\sigma^2 \quad \text{and} \quad E(Ze^{-Z}) = (\mu - \sigma^2)e^{\sigma^2/2 - \mu}, \quad (\text{S.41})$$

and that, for a constant a ,

$$\text{if } Z \sim N(0, \sigma^2) \text{ then } E(Ze^{-aZ}) = -a\sigma^2 e^{a^2\sigma^2/2} \quad \text{and} \quad E(Z^2 e^{-aZ}) = (a^2\sigma^4 + \sigma^2)e^{a^2\sigma^2/2}. \quad (\text{S.42})$$

From application of (S.41) and (S.42) we have

$$\begin{aligned}
E(U_1 e^{-U_0}) &= \int_{-\infty}^{\infty} U_1 \int_{-\infty}^{\infty} e^{-U_0} f_2(U; 0, \Sigma^0) dU_0 dU_1 \\
&= \int_{-\infty}^{\infty} U_1 f_1(U_1; 0, \Sigma_{11}^0) \int_{-\infty}^{\infty} e^{-U_0} f_1(U_0|U_1; \mu_{0|1}, \sigma_{0|1}^2) dU_0 dU_1 \\
&= \int_{-\infty}^{\infty} U_1 f_1(U_1; 0, \Sigma_{11}^0) e^{-\mu_{0|1} + \frac{1}{2}\sigma_{0|1}^2} dU_1 \\
&= \exp\left[\frac{1}{2}\left\{\Sigma_{00}^0 - \frac{(\Sigma_{01}^0)^2}{\Sigma_{11}^0}\right\}\right] \int_{-\infty}^{\infty} U_1 \exp\left(-\frac{\Sigma_{01}^0}{\Sigma_{11}^0} U_1\right) f_1(U_1; 0, \Sigma_{11}^0) dU_1 \\
&= -\Sigma_{01}^0 \exp\left(\frac{\Sigma_{00}^0}{2}\right).
\end{aligned} \tag{S.43}$$

Using similar algebraic steps and by making recursive use of (S.41) and (S.42) we get

$$E(U_1^2 e^{-U_0}) = \{(\Sigma_{01}^0)^2 + \Sigma_{11}^0\} \exp\left(\frac{\Sigma_{00}^0}{2}\right), \tag{S.44}$$

$$E(U_0 U_1 e^{-U_0}) = (\Sigma_{00}^0 \Sigma_{01}^0 + \Sigma_{01}^0) \exp\left(\frac{\Sigma_{00}^0}{2}\right), \tag{S.45}$$

$$E(U_1 e^{-U_0 - U_1}) = -(\Sigma_{01}^0 + \Sigma_{11}^0) \exp\left\{\frac{1}{2}(\Sigma_{00}^0 + 2\Sigma_{01}^0 + \Sigma_{11}^0)\right\}, \tag{S.46}$$

$$E(U_1^2 e^{-U_0 - U_1}) = \{(\Sigma_{01}^0)^2 + 2\Sigma_{01}^0 \Sigma_{11}^0 + (\Sigma_{11}^0)^2 + \Sigma_{11}^0\} \exp\left\{\frac{1}{2}(\Sigma_{00}^0 + 2\Sigma_{01}^0 + \Sigma_{11}^0)\right\} \tag{S.47}$$

and

$$E(U_0 U_1 e^{-U_0 - U_1}) = \{\Sigma_{00}^0 \Sigma_{11}^0 + \Sigma_{00}^0 \Sigma_{01}^0 + \Sigma_{01}^0 \Sigma_{11}^0 + (\Sigma_{01}^0)^2 + \Sigma_{01}^0\} \exp\left\{\frac{1}{2}(\Sigma_{00}^0 + 2\Sigma_{01}^0 + \Sigma_{11}^0)\right\}. \tag{S.48}$$

Now notice that (S.39) can be rewritten as $[\Lambda_{AA}]_{2,2} = \sum_{t=1}^8 T_t$ with

$$T_1 = \frac{2e^{-\beta_0^0}}{1-p} \{(\Sigma^0)^{11} - (\Sigma^0)^{10}\} E(U_1^2 e^{-U_0}), \quad T_2 = \frac{2e^{-\beta_0^0}}{1-p} \{(\Sigma^0)^{01} - (\Sigma^0)^{00}\} E(U_0 U_1 e^{-U_0}),$$

$$T_3 = \frac{2e^{-\beta_0^0 - \beta_1^0}}{p} (\Sigma^0)^{11} E(U_1^2 e^{-U_0 - U_1}), \quad T_4 = \frac{2e^{-\beta_0^0 - \beta_1^0}}{p} (\Sigma^0)^{01} E(U_0 U_1 e^{-U_0 - U_1}),$$

$$T_5 = -\frac{e^{-\beta_0^0 - \beta_1^0}}{p} E(e^{-U_0 - U_1}), \quad T_6 = -\frac{e^{-\beta_0^0}}{1-p} E(e^{-U_0}),$$

$$T_7 = \frac{2e^{-\beta_0^0 - \beta_1^0}}{p} E(U_1 e^{-U_0 - U_1}) \quad \text{and} \quad T_8 = -\frac{2e^{-\beta_0^0}}{1-p} E(U_1 e^{-U_0})$$

By noticing that

$$(\Sigma^0)^{00} = \{\Sigma_{00}^0 \Sigma_{11}^0 - (\Sigma_{01}^0)^2\}^{-1} \Sigma_{11}^0, \quad (\Sigma^0)^{01} = (\Sigma^0)^{10} = -\{\Sigma_{00}^0 \Sigma_{11}^0 - (\Sigma_{01}^0)^2\}^{-1} \Sigma_{01}^0$$

$$\text{and} \quad (\Sigma^0)^{11} = \{\Sigma_{00}^0 \Sigma_{11}^0 - (\Sigma_{01}^0)^2\}^{-1} \Sigma_{00}^0$$

and using (S.43)–(S.48) we have that

$$T_1 = \frac{2}{1-p} \{\Sigma_{00}^0 \Sigma_{11}^0 - (\Sigma_{01}^0)^2\}^{-1} \{\Sigma_{00}^0 (\Sigma_{01}^0)^2 + \Sigma_{00}^0 \Sigma_{11}^0 + (\Sigma_{01}^0)^3 + \Sigma_{01}^0 \Sigma_{11}^0\} e^{-\beta_0^0 + \Sigma_{00}^0/2}, \tag{S.49}$$

$$T_2 = -\frac{2}{1-p} \{ \Sigma_{00}^0 \Sigma_{11}^0 - (\Sigma_{01}^0)^2 \}^{-1} \{ \Sigma_{00}^0 (\Sigma_{01}^0)^2 + (\Sigma_{01}^0)^2 + \Sigma_{00}^0 \Sigma_{01}^0 \Sigma_{11}^0 + \Sigma_{01}^0 \Sigma_{11}^0 \} e^{-\beta_0^0 + \Sigma_{00}^0/2}, \quad (\text{S.50})$$

$$T_3 = \frac{2}{p} \{ \Sigma_{00}^0 \Sigma_{11}^0 - (\Sigma_{01}^0)^2 \}^{-1} \{ \Sigma_{00}^0 (\Sigma_{01}^0)^2 + 2 \Sigma_{00}^0 \Sigma_{01}^0 \Sigma_{11}^0 + \Sigma_{00}^0 (\Sigma_{11}^0)^2 + \Sigma_{00}^0 \Sigma_{11}^0 \} \\ \times e^{-\beta_0^0 - \beta_1^0 + (\Sigma_{00}^0 + 2 \Sigma_{01}^0 + \Sigma_{11}^0)/2}, \quad (\text{S.51})$$

$$T_4 = -\frac{2}{p} \{ \Sigma_{00}^0 \Sigma_{11}^0 - (\Sigma_{01}^0)^2 \}^{-1} \{ \Sigma_{00}^0 \Sigma_{01}^0 \Sigma_{11}^0 + \Sigma_{00}^0 (\Sigma_{01}^0)^2 + (\Sigma_{01}^0)^2 \Sigma_{11}^0 + (\Sigma_{01}^0)^3 + (\Sigma_{01}^0)^2 \} \\ \times e^{-\beta_0^0 - \beta_1^0 + (\Sigma_{00}^0 + 2 \Sigma_{01}^0 + \Sigma_{11}^0)/2}, \quad (\text{S.52})$$

$$T_7 = -\frac{2}{p} (\Sigma_{01}^0 + \Sigma_{11}^0) e^{-\beta_0^0 - \beta_1^0 + (\Sigma_{00}^0 + 2 \Sigma_{01}^0 + \Sigma_{11}^0)/2} \quad \text{and} \quad T_8 = \frac{2}{1-p} \Sigma_{01}^0 e^{-\beta_0^0 + \Sigma_{00}^0/2}. \quad (\text{S.53})$$

From application of (S.40),

$$T_5 = -\frac{1}{p} e^{-\beta_0^0 - \beta_1^0 + (\Sigma_{00}^0 + 2 \Sigma_{01}^0 + \Sigma_{11}^0)/2} \quad \text{and} \quad T_6 = -\frac{1}{1-p} e^{-\beta_0^0 + \Sigma_{00}^0/2}. \quad (\text{S.54})$$

The sum of the expressions in (S.49)–(S.54) provides the explicit form of the (2, 2) element of Λ_{AA} ,

$$[\Lambda_{AA}]_{2,2} = \frac{1}{p} e^{-\beta_0^0 - \beta_1^0 + (\Sigma_{00}^0 + 2 \Sigma_{01}^0 + \Sigma_{11}^0)/2} + \frac{1}{1-p} e^{-\beta_0^0 + \Sigma_{00}^0/2}. \quad (\text{S.55})$$

S.15.5 The Two-Term Asymptotic Variance of $\hat{\beta}_1$

From the explicit form of the (2, 2) element of Λ_{AA} in (S.55) we get the following explicit form for the two-term asymptotic variance of $\hat{\beta}_1$ in model (8):

$$\text{Asy.Var}(\hat{\beta}_1) = \frac{\Sigma_{11}^0}{m} + \frac{\phi}{mn} \left[\frac{1}{p} \exp \left\{ -\beta_0^0 - \beta_1^0 + \frac{1}{2} (\Sigma_{00}^0 + 2 \Sigma_{01}^0 + \Sigma_{11}^0) \right\} + \frac{1}{1-p} \exp \left\{ -\beta_0^0 + \frac{\Sigma_{00}^0}{2} \right\} \right].$$

S.16 Application to Confidence Intervals

We conducted a simulation exercise aimed at illustrating the application of second term improvements to generalized linear mixed model asymptotics to confidence intervals. Data sets were generated from the $d_F = 5$ and $d_R = 2$ logistic mixed model

$Y_{ij} | X_{1ij}, X_{2ij}, X_{3ij}, X_{4ij}, U_{0i}, U_{1i}$ independently distributed as

$$\text{Bernoulli} \left(\frac{1}{1 + \exp[-\{\beta_0^0 + U_{0i} + (\beta_1^0 + U_{1i})X_{1ij} + \beta_2^0 X_{2ij} + \beta_3^0 X_{3ij} + \beta_4^0 X_{4ij}\}]} \right),$$

where the $\begin{bmatrix} U_{0i} \\ U_{1i} \end{bmatrix}$ are independent $N(0, \Sigma^0)$ random vectors, $1 \leq i \leq m$, $1 \leq j \leq n$.

The ‘true’ parameter values were set to

$$(\beta_0^0, \beta_1^0, \beta_2^0, \beta_3^0, \beta_4^0) = (0.35, 0.96, -0.47, 1.06, -1.31) \quad \text{and} \quad \Sigma^0 = \begin{bmatrix} 0.56 & -0.34 \\ -0.34 & 0.89 \end{bmatrix}$$

and the predictor data were generated from independent Uniform distributions on the unit interval. The simulation design is such that the asymptotic variance of $\hat{\beta}_1$, corresponding to the fixed effect of the X_1 predictor, benefits from second term improvement. The true β^0 vector was chosen so that there was a variety of strengths of predictor fixed effects. We selected the Σ^0 matrix to ensure that there was a significant amount of heterogeneity in the random intercepts and slopes. In our reporting of

simulation results we use the following standard deviation and correlation parameterisation: $\sigma_1^0 \equiv (\Sigma_{11}^0)^{1/2}$, $\sigma_2^0 \equiv (\Sigma_{22}^0)^{1/2}$ and $\rho^0 \equiv \Sigma_{12}^0 / (\sigma_1^0 \sigma_2^0)$. To assess potential large sample improvements afforded by the two-term asymptotic covariance expressions at (7) we varied m over the set $\{100, 150, \dots, 500\}$ and fixed n at $m/10$. For each (m, n) pair we then simulated 500 replications and obtained approximate 95% confidence intervals for all model parameters according to the approach described in § 4 of Jiang *et al.* (2022) and the second term improvements arising from (7). The confidence intervals for σ_1^0 and σ_2^0 involved the use of asymptotic normality results for logarithms of these parameters, followed by exponentiation, as outlined in Section S.16.2. Similar remarks apply to ρ^0 but with use of \tanh^{-1} and \tanh functions. The requisite bivariate integrals were obtained using the function `hcubature()` within the R language package `cubature` (Balasubramanian *et al.*, 2023). The point estimates, which were obtained via the R language package `glmmTMB` (Brooks *et al.*, 2023), use Laplace’s method to approximate bivariate integrals.

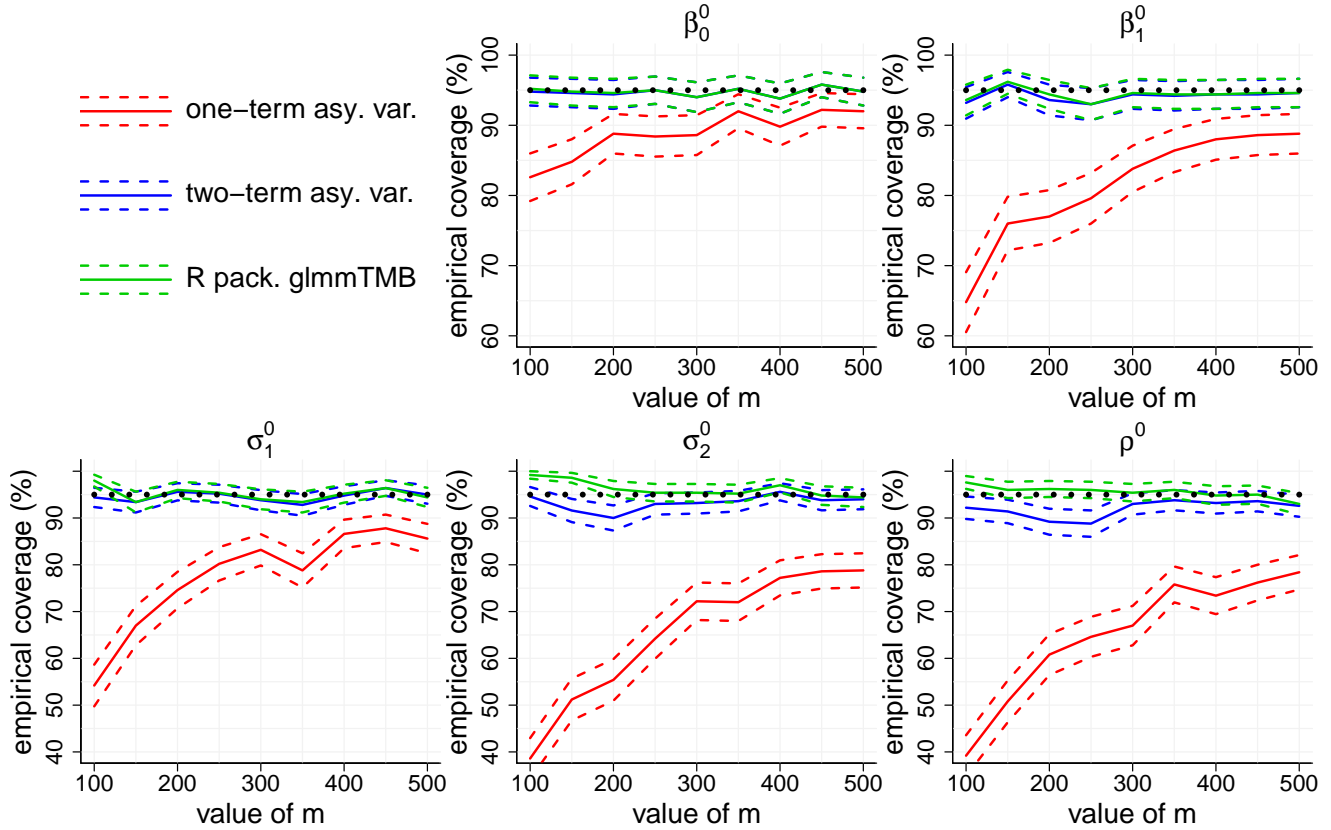


Figure S.1: Empirical coverage of confidence intervals from the simulation exercise described in the text. Each panel corresponds to a model parameter that is impacted by second term asymptotic improvements. The advertised coverage level is fixed at 95% and is indicated by a horizontal dotted line in each panel. The solid curves show, dependent on the number of groups m , the empirical coverage levels for confidence intervals based on each of the three approaches. The dashed curves correspond to plus and minus two standard errors of the sample proportions. The within-group sample size, n , is fixed at $m/10$.

Note that the confidence intervals for β_0^0 , β_1^0 and the entries of Σ^0 differ according to the two asymptotic theory approaches since the estimators of these parameters have order m^{-1} asymptotic variances. The confidence intervals for β_2^0 , β_3^0 and β_4^0 are unaffected by the second term asymptotic improvements since their estimators have order $(mn)^{-1}$ asymptotic variances.

Figure S.1 compares the empirical coverages of confidence intervals with advertised levels of 95% for the one-term asymptotic variances of Jiang *et al.* (2022) and the two-term asymptotic variances that arise from (7). In Figure S.1 we only consider the parameters that are affected by second term improvement. The empirical coverages for the other parameters are provided in Figure S.2. For comparison with existing software products, the empirical coverages for the `glmmTMB` confidence intervals are

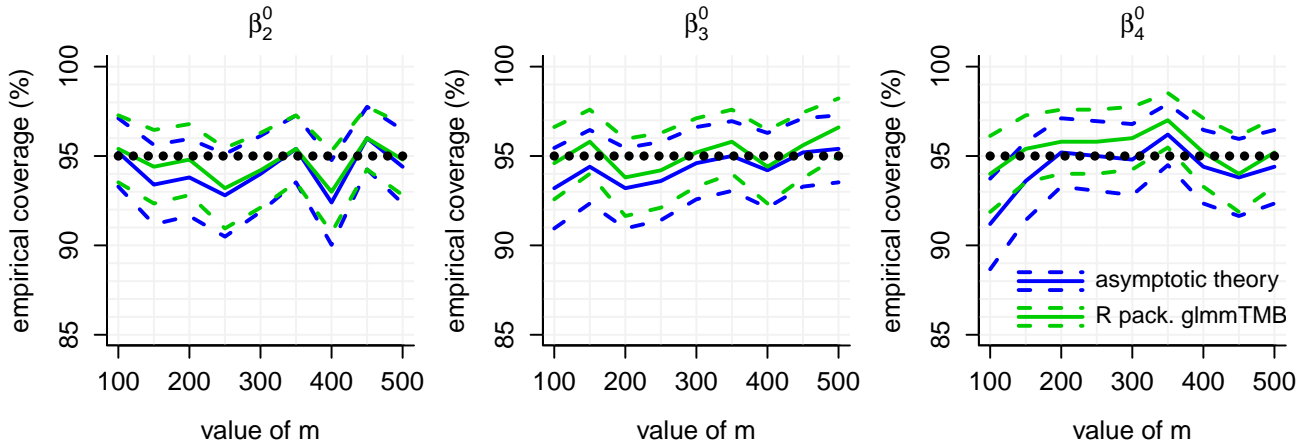


Figure S.2: Empirical coverage of confidence intervals from the simulation exercise described in the text. Each panel corresponds to a fixed effect model parameter that is not impacted by second term asymptotic improvements. The advertised coverage level is fixed at 95% and is indicated by a horizontal dotted line in each panel. The solid curves show, dependent on the number of groups m , the empirical coverage levels for confidence intervals that use a one-term asymptotic variance approximation. The dashed curves correspond to plus and minus two standard errors of the sample proportions. The within-group sample size, n , is fixed at $m/10$.

also shown in Figure S.1. For β_0^0 , β_1^0 and σ_1^0 there is close correspondence between the two-term and glmmTMB confidence intervals. For σ_2^0 and ρ^0 and lower values of m , the two-term confidence intervals are prone to some under-coverage whilst glmmTMB has empirical coverages above the advertised level.

Figure S.2 compares the empirical coverages of confidence intervals with advertised levels of 95% for the parameters of (7) that are not affected by second term improvement. The figure shows that the asymptotic theory variances lead to good coverages for β_2^0 , β_3^0 and β_4^0 , even for lower sample size situations. They are also quite close to those produced by the glmmTMB package.

It is clear from Figure S.1 that our second term improvements lead to much better coverages for lower sample size situations. On the other hand, one-term confidence intervals are trivial to compute whilst the two-term versions require a considerable amount of computing involving numerical integration.

S.16.1 Additional Simulation Results

The simulation exercise described above is necessarily limited and subject to constraints involving, but not confined to, the values of m and n . Here we report on an extension that involves a different set of (m, n) pairs that does not fix n to equal $m/10$. This allows for some understanding of the effect of n on empirical coverage of confidence intervals.

We extended the simulation exercise to allow for m and n to vary in an unconstrained manner. The extension involved replacement of the previous sample size design,

$$(m, n) \in \{(100, 10), (150, 15), (200, 20), (250, 25), (300, 30), (350, 35), (400, 40), (450, 45), (500, 50)\},$$

by

$$(m, n) \in \{200, 300, 400\} \times \{20, 30, 40\}.$$

Figure S.3 summarises the empirical coverage values from this additional simulation exercise.

For fixed m , the empirical coverages are seen to improve as n increases for almost all parameters and m values. However, when n is held fixed then increasing m does not necessarily lead to improvements in empirical coverage. Whilst limited, the simulation results summarised by Figure S.3 suggest that m and n should increase together to improve empirical coverage. This is in keeping with our Section 3.2 assumptions.

It is clear that the one-term asymptotic variance approximation requires much higher sample sizes to achieve advertised coverage performance. For inference concerning β_0^0 and β_1^0 the empirical coverages from the two-term asymptotic variance approximation and the approach used by the R package `glmmTMB` hardly differ. There are some slight differences for inference concerning σ_1^0 , σ_2^0 and ρ^0 . For $n = 20$ the two-term asymptotic variance approach has a tendency to under-cover, whilst the `glmmTMB` approach can sometimes over-cover. Some casual checks suggest that the over-coverage of the `glmmTMB` approach is due to its confidence intervals being noticeably wider.

S.16.2 Confidence Interval Construction Details

For any $u \in \mathbb{R}^{d_R}$, define

$$\begin{aligned}\widehat{\Omega}_{AA}(u) &\equiv \frac{1}{mn} \sum_{i=1}^m \sum_{j=1}^{n_i} b'' \left((\widehat{\beta}_A + u)^T X_{Aij} + \widehat{\beta}_B^T X_{Bij} \right) X_{Aij} X_{Aij}^T, \\ \widehat{\Omega}_{AB}(u) &\equiv \frac{1}{mn} \sum_{i=1}^m \sum_{j=1}^{n_i} b'' \left((\widehat{\beta}_A + u)^T X_{Aij} + \widehat{\beta}_B^T X_{Bij} \right) X_{Aij} X_{Bij}^T \\ \text{and } \widehat{\Omega}_{BB}(u) &\equiv \frac{1}{mn} \sum_{i=1}^m \sum_{j=1}^{n_i} b'' \left((\widehat{\beta}_A + u)^T X_{Aij} + \widehat{\beta}_B^T X_{Bij} \right) X_{Bij} X_{Bij}^T.\end{aligned}\tag{S.56}$$

Then the natural studentisation of $E\{\Psi_6(U)\}$ is

$$\begin{aligned}\widehat{E}\{\Psi_6(U)\} &\equiv E\left\{\widehat{\Omega}_{BB}(U) - \widehat{\Omega}_{AB}(U)^T \widehat{\Omega}_{AA}(U)^{-1} \widehat{\Omega}_{AB}(U) \mid \mathcal{X}\right\} \\ &= |2\pi \widehat{\Sigma}|^{-1/2} \int_{d_R} \left\{ \widehat{\Omega}_{BB}(u) - \widehat{\Omega}_{AB}(u)^T \widehat{\Omega}_{AA}(u)^{-1} \widehat{\Omega}_{AB}(u) \right\} \exp\left(-\frac{1}{2} u^T \widehat{\Sigma} u\right) du.\end{aligned}\tag{S.57}$$

In the last expression of (S.57) integration is applied element-wise to each entry of the matrix inside the integral. The natural studentisations of

$$\Lambda_{AA}, \quad \Lambda_{AB}, \quad \Delta, \quad E\{\Psi_8(U)\} \quad \text{and} \quad E\{\Psi_9(U)\}\tag{S.58}$$

are analogous to that for $E\{\Psi_6(U)\}$. The studentisations for the quantities in (S.58) depend on the functions defined by (S.56) as well as similar sample counterparts of $\Omega'_{AAA}(U)$ and $\Omega'_{AAB}(U)$. Next define

$$\widehat{\text{Asy.Cov}}(\widehat{\beta}) = \frac{1}{m} \begin{bmatrix} \widehat{\Sigma} & O \\ O & O \end{bmatrix} + \frac{\widehat{\phi}}{mn} \begin{bmatrix} \widehat{\Lambda}_{AA}^{-1} & \widehat{\Lambda}_{AA}^{-1} \widehat{\Lambda}_{AB} \\ \widehat{\Lambda}_{AB}^T \widehat{\Lambda}_{AA}^{-1} & \widehat{\Lambda}_{AB}^T \widehat{\Lambda}_{AA}^{-1} \widehat{\Lambda}_{AB} + \widehat{E}\{\Psi_6(U)\} \end{bmatrix}^{-1}\tag{S.59}$$

and

$$\begin{aligned}\widehat{\text{Asy.Cov}}(\text{vech}(\widehat{\Sigma})) &= \frac{2D_{d_R}^+ (\widehat{\Sigma} \otimes \widehat{\Sigma}) D_{d_R}^{+T}}{m} \\ &\quad + \frac{\widehat{\phi}}{mn} \left(2\widehat{E}\{\Psi_9(U)\} - 4\widehat{E}\{\Psi_8(U)\} + \widehat{\Delta}^T [\widehat{E}\{\Psi_6(U)\}]^{-1} \widehat{\Delta} \right).\end{aligned}$$

In the general quasi-likelihood situation, the most common choice for $\widehat{\phi}$ is the method of moments estimator and is often labelled the *Pearson* estimator. For ordinary likelihood settings, such as for Gaussian and Gamma responses, $\widehat{\phi}$ could instead be the maximum likelihood estimator.

Let $(\beta^0)_k$ denote the k th entry of β^0 . Then approximate $100(1 - \alpha)\%$ confidence intervals for $(\beta^0)_k$ based on (S.59) are

$$(\widehat{\beta})_k \pm \Phi^{-1}\left(1 - \frac{1}{2}\alpha\right) \sqrt{\left\{ \widehat{\text{Asy.Cov}}(\widehat{\beta}) \right\}_{kk}}, \quad 1 \leq k \leq d_F.\tag{S.60}$$

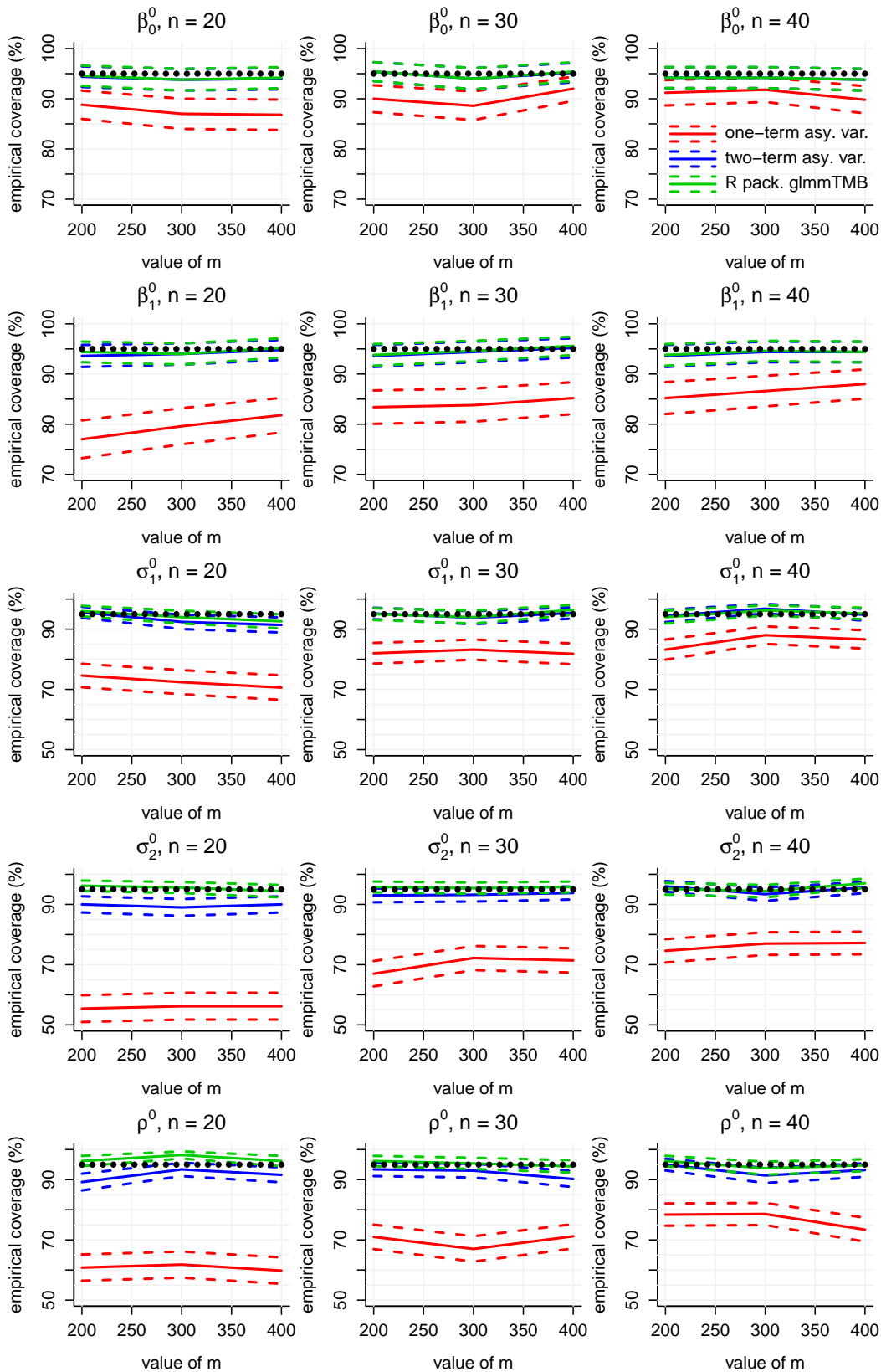


Figure S.3: Empirical coverage of confidence intervals from the simulation exercise described in Section S.16.1. Each panel corresponds to a model parameter that is impacted by second term asymptotic improvements. The advertised coverage level is fixed at 95% and is indicated by a horizontal dotted line in each panel. The solid curves show the empirical coverage levels for confidence intervals based on each of the three approaches. The dashed curves correspond to plus and minus two standard errors of the sample proportions.

The confidence intervals in (S.60) are analogous to those given in § 4 of Jiang *et al.* (2022). For $1 \leq k \leq d_r$, (S.60) provides second term improvements of the Jiang *et al.* (2022) confidence intervals. For $d_r + 1 \leq k \leq d_f$ both sets of confidence intervals are identical.

Confidence intervals for the entries of Σ^0 have expressions analogous to (S.60). However, with interpretability in mind, it is common to instead perform inference for the standard deviation and correlation parameters associated with Σ^0 . In the special case of $d_r = 2$ the parameters are

$$\sigma_1^0 \equiv \sqrt{\Sigma_{11}^0}, \quad \sigma_2^0 \equiv \sqrt{\Sigma_{22}^0} \quad \text{and} \quad \rho^0 \equiv \Sigma_{12}^0 / (\sigma_1^0 \sigma_2^0).$$

In addition, the reparameterisation

$$\omega_1^0 \equiv \log(\sigma_1), \quad \omega_2^0 \equiv \log(\sigma_2) \quad \text{and} \quad \omega_3^0 \equiv \tanh^{-1}(\rho^0)$$

is often used to help counteract the effects of skewness in the sampling distributions of $\hat{\sigma}_1$, $\hat{\sigma}_2$ and $\hat{\rho}$, and ensure that all confidence interval limits are within the relevance parameter spaces. Application of the Multivariate Delta Method (e.g. Agresti, 2013, Section 16.1.3) leads to

$$\widehat{\text{Asy.Cov}} \left(\begin{bmatrix} \hat{\omega}_1 \\ \hat{\omega}_2 \\ \hat{\omega}_3 \end{bmatrix} \right) = \Xi(\hat{\sigma}_1, \hat{\sigma}_2, \hat{\rho})^T \widehat{\text{Asy.Cov}}(\text{vech}(\hat{\Sigma})) \Xi(\hat{\sigma}_1, \hat{\sigma}_2, \hat{\rho})$$

where

$$\Xi(\sigma_1, \sigma_2, \rho) \equiv \frac{1}{2\sigma_1^2\sigma_2^2(1-\rho^2)} \begin{bmatrix} \sigma_2^2(1-\rho^2) & 0 & -\rho\sigma_2^2 \\ 0 & 0 & 2\sigma_1\sigma_2 \\ 0 & \sigma_1^2(1-\rho^2) & -\rho\sigma_1^2 \end{bmatrix}.$$

Routine arguments then lead to confidence intervals for σ_1^0 , σ_2^0 and ρ^0 based on asymptotic normality results for $(\hat{\omega}_1, \hat{\omega}_2, \hat{\omega}_3)$.

References

- Agresti, A. (2013). *Categorical Data Analysis*. Hoboken, New Jersey: John Wiley & Sons.
- Balasubramanian, N., Johnson, S.G., Hahn, T., Bouvier, A. & Ki eu, K. (2023). cubature 2.0.4.6: Adaptive multivariate integration over hypercubes. R package. <https://r-project.org>
- Jiang, J., Wand, M.P. & 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.
- Magnus, J.R. and Neudecker, H. (1979). The commutation matrix: some properties and applications. *The Annals of Statistics*, **7**, 381–394.
- Magnus, J.R. and Neudecker, H. (1999). *Matrix Differential Calculus. Revised Edition*. Chichester, U.K.: John Wiley & Sons.
- Miyata, Y. (2004). Fully exponential Laplace approximation using asymptotic modes. *Journal of the American Statistical Association*, **99**, 1037–1049.
- Pace, L. and Salvan, A. (1997). *Principles of Statistical Inference from a Neo-Fisherian Perspective*. Singapore: World Scientific Publishing Company.
- Wand, M.P. (2002). Vector differential calculus in statistics. *The American Statistician*, **56**, 55–62.