Supplement for:

Second Term Improvements to Generalised Linear Mixed Model Asymptotics

By Luca Maestrini[1], Aishwarya Bhaskaran[2] and Matt P. Wand[2]

[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.16 provides some additional results concerning from the logistic mixed model simulation exercise described in § 4.

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

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

and, assuming that A is invertible,

$$\left\{D_d^T(A\otimes A)D_d\right\}^{-1}=D_d^+(A^{-1}\otimes A^{-1})D_d^{+T}. \tag{S.5}$$

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

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

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

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

S.3 Multivariate Derivative Notation

For f a smooth real-valued function of the d-variate argument $x\equiv(x_1,\ldots,x_d)$, let $\nabla f(x)$ denote the $d\times 1$ vector with rth 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 $f:\mathbb{R}^d\to\mathbb{R}$. Then, for sufficiently smooth f, the three-term Taylor expansion of

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

is

$$f(x+h)=f(x)+\sum_{r=1}^d\{\nabla f(x)\}_r h_r+\tfrac{1}{2}\sum_{r=1}^d\sum_{s=1}^d\{\nabla^2 f(x)\}_{rs}h_rh_s+\ldots. \tag{S.8}$$

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

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

From this it is clear that

$$(\nabla\alpha)(x+h)=(\nabla\alpha)(x)+\{(\nabla^2\alpha)(x)\}h+\tfrac{1}{2}\{(\nabla^3\alpha)(x)\}\bigstar(hh^T)+\ldots. \tag{S.9}$$

where the $\bigstar$ notation is as defined by (6).

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,

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

where

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

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

$$S_{\mathrm{A}i} \equiv \nabla_{\beta_{\mathrm{A}}} \log p_{Y_i|X_i}(Y_i|X_i), \quad S_{\mathrm{B}i} \equiv \nabla_{\beta_{\mathrm{B}}} \log p_{Y_i|X_i}(Y_i|X_i)$$

and

$$S_{\mathrm{C}i} \equiv \nabla_{\mathrm{vech}(\Sigma)} \log p_{Y_i|X_i}(Y_i|X_i)$$

denote the ith contribution to the scores with respect to each of β_{A}, β_{B} and $\mathrm{vech}(\Sigma)$. Then it is straightforward to show that the exact scores are

$$S_{\mathrm{A}i} = \frac{\int_{\mathbb{R}^{d_{\mathrm{R}}}} g_{iA}(u) c_S(u) \exp\{-nh_i(u)\}\, du}{\int_{\mathbb{R}^{d_{\mathrm{R}}}} c_S(u) \exp\{-nh_i(u)\}\, du}, \tag{S.11}$$

$$S_{\mathrm{B}i} = \frac{\int_{\mathbb{R}^{d_{\mathrm{R}}}} g_{iB}(u) c_S(u) \exp\{-nh_i(u)\}\, du}{\int_{\mathbb{R}^{d_{\mathrm{R}}}} c_S(u) \exp\{-nh_i(u)\}\, du} \tag{S.12}$$

and

$$S_{\mathrm{C}i} = \frac{\int_{\mathbb{R}^{d_{\mathrm{R}}}} g_{iC}(u) c_S(u) \exp\{-nh_i(u)\}\, du}{\int_{\mathbb{R}^{d_{\mathrm{R}}}} c_S(u) \exp\{-nh_i(u)\}\, du} - \tfrac{1}{2} D_{d_{\mathrm{R}}}^T \mathrm{vec}(\Sigma^{-1}) \tag{S.13}$$

where

$$\begin{aligned}
c_S(u) &\equiv \exp(-\tfrac{1}{2} u^T \Sigma^{-1} u), \quad g_{iA}(u) \equiv \Sigma^{-1} u,\\
g_{iB}(u) &\equiv \frac{1}{\phi} \sum_{j=1}^{n_i} X_{\mathrm{B}ij} \{Y_{ij} - b'\big((\beta_{\mathrm{A}} + u)^T X_{\mathrm{A}ij} + \beta_{\mathrm{B}}^T X_{\mathrm{B}ij}\big)\},\\
g_{iC}(u) &\equiv \tfrac{1}{2} D_{d_{\mathrm{R}}}^T (\Sigma^{-1} \otimes \Sigma^{-1}) \mathrm{vec}(uu^T)\\
\text{and} \quad h_i(u) &\equiv -\frac{1}{n\phi} \sum_{j=1}^{n_i} \big\{Y_{ij} u^T X_{\mathrm{A}ij} - b\big((\beta_{\mathrm{A}} + u)^T X_{\mathrm{A}ij} + \beta_{\mathrm{B}}^T X_{\mathrm{B}ij}\big)\big\}.
\end{aligned}$$

An integration by parts step is used to obtain the $S_{\mathrm{A}i}$ expression.

In the upcoming sections we obtain asymptotic approximations of $S_{\mathrm{A}i}$ $S_{\mathrm{B}i}$ and $S_{\mathrm{C}i}$. Key quantities for these approximations are

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

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 $\mathcal{G}_{\mathrm{A}i}$, $\mathcal{G}_{\mathrm{B}i}$, $\mathcal{H}_{\mathrm{AA}i}$, $\mathcal{H}_{\mathrm{AB}i}$, and $\mathcal{H}_{\mathrm{BB}i}$ as follows:

$$\mathcal{G}_{\mathrm{A}i} \equiv \sum_{j=1}^{n_i} \{Y_{ij} - b'\big((\beta_{\mathrm{A}} + U_i)^T X_{\mathrm{A}ij} + \beta_{\mathrm{B}}^T X_{\mathrm{B}ij}\big)\} X_{\mathrm{A}ij},$$

$$\mathcal{G}_{\mathrm{B}i} \equiv \sum_{j=1}^{n_i} \{Y_{ij} - b'\big((\beta_{\mathrm{A}} + U_i)^T X_{\mathrm{A}ij} + \beta_{\mathrm{B}}^T X_{\mathrm{B}ij}\big)\} X_{\mathrm{B}ij},$$

$$\mathcal{H}_{\mathrm{AA}i} \equiv \sum_{j=1}^{n_i} b''\big((\beta_{\mathrm{A}} + U_i)^T X_{\mathrm{A}ij} + \beta_{\mathrm{B}}^T X_{\mathrm{B}ij}\big) X_{\mathrm{A}ij} X_{\mathrm{A}ij}^T,$$

$$\mathcal{H}_{\text{AB}i} \equiv \sum_{j=1}^{n_i} b''\big((\beta_\text{A} + U_i)^T X_{\text{A}ij} + \beta_\text{B}^T X_{\text{B}ij}\big) X_{\text{A}ij} X_{\text{B}ij}^T$$

$$\text{and} \quad \mathcal{H}_{\text{BB}i} \equiv \sum_{j=1}^{n_i} b''\big((\beta_\text{A} + U_i)^T X_{\text{A}ij} + \beta_\text{B}^T X_{\text{B}ij}\big) X_{\text{B}ij} X_{\text{B}ij}^T.$$

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

$$\sum_{j=1}^{n_i} b'''\big((\beta_\text{A} + U_i)^T X_{\text{A}ij} + \beta_\text{B}^T X_{\text{B}ij}\big)(X_{\text{A}ij})_r (X_{\text{A}ij})_s (X_{\text{A}ij})_t$$

and $\mathcal{H}'_{\text{AAB}i}$ to be the $d_\text{R} \times d_\text{R} \times d_\text{B}$ array with (r, s, t) entry equal to

$$\sum_{j=1}^{n_i} b'''\big((\beta_\text{A} + U_i)^T X_{\text{A}ij} + \beta_\text{B}^T X_{\text{B}ij}\big)(X_{\text{A}ij})_r (X_{\text{A}ij})_s (X_{\text{B}ij})_t$$

where

$$d_\text{B} \equiv d_\text{F} - d_\text{R}.$$

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

$$\begin{gathered} E\big(\mathcal{G}_{\text{A}i} | X_i, U_i\big) = 0, \quad E\big(\mathcal{G}_{\text{B}i} | X_i, U_i\big) = 0, \\ E\big(\mathcal{G}_{\text{A}i}^{\otimes 2} | X_i, U_i\big) = \phi \mathcal{H}_{\text{AA}i}, \quad E\big(\mathcal{G}_{\text{A}i} \mathcal{G}_{\text{B}i}^T | X_i, U_i\big) = \phi \mathcal{H}_{\text{AB}i} \quad \text{and} \quad E\big(\mathcal{G}_{\text{B}i}^{\otimes 2} | X_i, U_i\big) = \phi \mathcal{H}_{\text{BB}i} \end{gathered} \tag{S.14}$$

where, throughout this supplement,

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

Also note that

$$\begin{gathered} \mathcal{G}_{\text{A}i} = O_p(n^{1/2}) 1_{d_\text{R}}, \quad \mathcal{G}_{\text{B}i} = O_p(n^{1/2}) 1_{d_\text{B}}, \quad \mathcal{H}_{\text{AA}i} = O_p(n) 1_{d_\text{R}}^{\otimes 2}, \\ \mathcal{H}_{\text{BB}i} = O_p(n) 1_{d_\text{B}}^{\otimes 2}, \quad \mathcal{H}_{\text{AB}i} = O_p(n) 1_{d_\text{R}} 1_{d_\text{B}}^T \end{gathered}$$

and that all entries of $\mathcal{H}'_{\text{AAA}i}$ and $\mathcal{H}'_{\text{AAB}i}$ are $O_p(n)$.

S.8 Approximation of U_i^*

Use of (S.10) to approximate $S_{\text{A}i}$, $S_{\text{B}i}$ and $S_{\text{C}i}$ 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

$$\nabla \mathcal{C}_i(u) \equiv -\sum_{j=1}^{n_i} \Big\{ Y_{ij} - b'\big((\beta_\text{A} + u)^T X_{\text{A}ij} + \beta_\text{B}^T X_{\text{B}ij}\big) \Big\} X_{\text{A}ij}.$$

Then, from (S.9) we have

$$\begin{aligned} \nabla \mathcal{C}_i(U_i^*) &= \nabla \mathcal{C}_i(U_i + U_i^* - U_i) \\ &= \nabla \mathcal{C}_i(U_i) + \{\nabla^2 \mathcal{C}_i(U_i)\}(U_i^* - U_i) \\ &\quad + \tfrac{1}{2} \left\{ \nabla^3 \mathcal{C}_i(U_i) \right\} \bigstar \{(U_i^* - U_i)(U_i^* - U_i)^T\} + \ldots. \end{aligned}$$

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

$$\begin{aligned}\nabla^2\mathcal{C}_i(u) &= \sum_{j=1}^{n_i} b''\big((\beta_{\rm A}+u)^T X_{{\rm A}ij} + \beta_{\rm B}^T X_{{\rm B}ij}\big) X_{{\rm A}ij} X_{{\rm A}ij}^T \\ &= \left[\sum_{j=1}^{n_i} b''\big((\beta_{\rm A}+u)^T X_{{\rm A}ij} + \beta_{\rm B}^T X_{{\rm B}ij}\big)(X_{{\rm A}ij})_r (X_{{\rm A}ij})_s\right]_{1\le r,s\le d_{\rm R}}.\end{aligned}$$

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

$$\nabla^3\mathcal{C}_i(u) = \left[\sum_{j=1}^{n_i} b'''\big((\beta_{\rm A}+u)^T X_{{\rm A}ij} + \beta_{\rm B}^T X_{{\rm B}ij}\big)(X_{{\rm A}ij})_r (X_{{\rm A}ij})_s (X_{{\rm A}ij})_t\right]_{1\le r,s,t\le d_{\rm R}}.$$

We then have

$$\nabla\mathcal{C}_i(U_i^*) = -\mathcal{G}_{{\rm A}i} + \mathcal{H}_{{\rm AA}i}(U_i^* - U_i) + \tfrac{1}{2}\mathcal{H}'_{{\rm AAA}i}\bigstar\{(U_i^*-U_i)(U_i^*-U_i)^T\} + \ldots$$

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

$$\mathcal{H}_{{\rm AA}i}^{-1}\mathcal{G}_{{\rm A}i} = (U_i^*-U_i) + \tfrac{1}{2}\mathcal{H}_{{\rm AA}i}^{-1}\Big[\mathcal{H}'_{{\rm AAA}i}\bigstar\{(U_i^*-U_i)(U_i^*-U_i)^T\}\Big] + \ldots. \tag{S.15}$$

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

$$y \equiv \mathcal{H}_{{\rm AA}i}^{-1}\mathcal{G}_{{\rm A}i} \quad \text{and} \quad x \equiv U_i^* - U_i.$$

Then, in keeping with the displayed equation just before (9.43) of Pace & Salvan (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,

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

Then

$$y^a = x^a + A^a_{rs}x^{rs} + \ldots.$$

where

$$\begin{aligned}A^a_{rs}x^{rs} &= \text{the } a\text{th entry of } \tfrac{1}{2}\mathcal{H}_{{\rm AA}i}^{-1}\Big[\mathcal{H}'_{{\rm AAA}i}\bigstar\{(U_i^*-U_i)(U_i^*-U_i)^T\}\Big] \\ &= \text{the } a\text{th entry of } \tfrac{1}{2}\mathcal{H}_{{\rm AA}i}^{-1}\Big\{\mathcal{H}'_{{\rm AAA}i}\bigstar(xx^T)\Big\}.\end{aligned}$$

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

$$\begin{aligned}x^a &= y^a - A^a_{rs}y^{rs} + \ldots \\ &= y^a - \text{the } a\text{th entry of } \tfrac{1}{2}\mathcal{H}_{{\rm AA}i}^{-1}\Big\{\mathcal{H}'_{{\rm AAA}i}\bigstar(yy^T)\Big\} + \ldots \\ &= y^a - \text{the } a\text{th entry of } \tfrac{1}{2}\mathcal{H}_{{\rm AA}i}^{-1}\Big\{\mathcal{H}'_{{\rm AAA}i}\bigstar\Big(\mathcal{H}_{{\rm AA}i}^{-1}\mathcal{G}_{{\rm A}i}\Big)^{\otimes 2}\Big\} + \ldots.\end{aligned}$$

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

$$U_i^* = U_i + \mathcal{H}_{{\rm AA}i}^{-1}\mathcal{G}_{{\rm A}i} - \tfrac{1}{2}\mathcal{H}_{{\rm AA}i}^{-1}\Big\{\mathcal{H}'_{{\rm AAA}i}\bigstar\Big(\mathcal{H}_{{\rm AA}i}^{-1}\mathcal{G}_{{\rm A}i}\mathcal{G}_{{\rm A}i}^T\mathcal{H}_{{\rm AA}i}^{-1}\Big)\Big\} + O_p(n^{-3/2})1_{d_{\rm R}}. \tag{S.16}$$

S.9 Score Asymptotic Approximation

We are now ready to obtain approximations of the scores $S_{\mathrm{A}i}$, $S_{\mathrm{B}i}$ and $S_{\mathrm{C}i}$ with accuracies that are sufficient for the two-term asymptotic covariance matrices of (7).

S.9.1 Approximation of $S_{\mathbf{A}i}$

For each $1 \le r \le d_{\mathrm{R}}$, let e_r denote the $d_{\mathrm{R}} \times 1$ vector having rth entry equal to 1 and zeroes elsewhere.

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

For each $1 \le r \le d_{\mathrm{R}}$, the contribution to the rth entry of $S_{\mathrm{A}i}$ from the first term on the right-hand side of (S.10) is the rth entry of $\Sigma^{-1}U_i^*$. In view of (S.16) we obtain the following contribution to $S_{\mathrm{A}i}$:

$$\Sigma^{-1}U_i + \Sigma^{-1}\mathcal{H}_{\mathrm{AA}i}^{-1}\mathcal{G}_{\mathrm{A}i} - \tfrac{1}{2}\Sigma^{-1}\mathcal{H}_{\mathrm{AA}i}^{-1}\Big\{\mathcal{H}'_{\mathrm{AAA}i}\bigstar\Big(\mathcal{H}_{\mathrm{AA}i}^{-1}\mathcal{G}_{\mathrm{A}i}\mathcal{G}_{\mathrm{A}i}^T\mathcal{H}_{\mathrm{AA}i}^{-1}\Big)\Big\} + O_p(n^{-3/2})1_{d_{\mathrm{R}}}.$$

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 rth entry of $S_{\mathrm{A}i}$ from the second term on the right-hand side of (S.10) is

$$-\phi e_r^T\Sigma^{-1}\left\{\sum_{j=1}^{n_i} b''\big((\beta_{\mathrm{A}} + U_i^*)^T X_{\mathrm{A}ij} + \beta_{\mathrm{B}}^T X_{\mathrm{B}ij}\big)X_{\mathrm{A}ij}X_{\mathrm{A}ij}^T\right\}^{-1}\Sigma^{-1}U_i^*. \tag{S.17}$$

Substitution of (S.16) into (S.17) then leads to the following contribution to $S_{\mathrm{A}i}$:

$$-\phi\Sigma^{-1}\mathcal{H}_{\mathrm{AA}i}^{-1}\Sigma^{-1}U_i + O_p(n^{-3/2})1_{d_{\mathrm{R}}}.$$

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

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

$$-\frac{\phi}{2}\Sigma^{-1}\mathcal{H}_{\mathrm{AA}i}^{-1}\Big(\mathcal{H}'_{\mathrm{AAA}i}\bigstar\mathcal{H}_{\mathrm{AA}i}^{-1}\Big) + O_p(n^{-3/2})1_{d_{\mathrm{R}}}.$$

S.9.1.5 The Resultant Score Approximation

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

$$\begin{aligned} S_{\mathrm{A}i} = {} & \Sigma^{-1}U_i + \Sigma^{-1}\mathcal{H}_{\mathrm{AA}i}^{-1}\mathcal{G}_{\mathrm{A}i} - \frac{1}{2}\Sigma^{-1}\mathcal{H}_{\mathrm{AA}i}^{-1}\Big\{\mathcal{H}'_{\mathrm{AAA}i}\bigstar\Big(\mathcal{H}_{\mathrm{AA}i}^{-1}\mathcal{G}_{\mathrm{A}i}\mathcal{G}_{\mathrm{A}i}^T\mathcal{H}_{\mathrm{AA}i}^{-1}\Big)\Big\} \\ & -\phi\Sigma^{-1}\mathcal{H}_{\mathrm{AA}i}^{-1}\Sigma^{-1}U_i - \frac{\phi}{2}\Sigma^{-1}\mathcal{H}_{\mathrm{AA}i}^{-1}\Big(\mathcal{H}'_{\mathrm{AAA}i}\bigstar\mathcal{H}_{\mathrm{AA}i}^{-1}\Big) + O_p(n^{-3/2})1_{d_{\mathrm{R}}}. \end{aligned} \tag{S.18}$$

S.9.2 Approximation of $S_{\mathbf{B}i}$

For each $1 \le r \le d_{\mathrm{R}}$, let e_r denote the $d_{\mathrm{B}} \times 1$ vector having rth entry equal to 1 and zeroes elsewhere.

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

The contribution to $S_{\mathrm{B}i}$ from the first term on the right-hand side of (S.10) is

$$g_{iB}(U_i^*) = \frac{1}{\phi}\sum_{j=1}^{n_i} X_{\mathrm{B}ij}\big\{Y_{ij} - b'\big((\beta_{\mathrm{A}} + U_i^*)^T X_{\mathrm{A}ij} + \beta_{\mathrm{B}}^T X_{\mathrm{B}ij}\big)\big\}. \tag{S.19}$$

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

$$\begin{aligned}
&b'\big((\beta_{\mathrm{A}} + U_i^*)^T X_{\mathrm{A}ij} + \beta_{\mathrm{B}}^T X_{\mathrm{B}ij}\big)\\
&\quad = b'\big((\beta_{\mathrm{A}} + U_i)^T X_{\mathrm{A}ij} + \beta_{\mathrm{B}}^T X_{\mathrm{B}ij}\big)\\
&\qquad + X_{\mathrm{A}ij}^T (U_i^* - U_i) b''\big((\beta_{\mathrm{A}} + U_i)^T X_{\mathrm{A}ij} + \beta_{\mathrm{B}}^T X_{\mathrm{B}ij}\big)\\
&\qquad + \frac{1}{2} X_{\mathrm{A}ij}^T (U_i^* - U_i)^{\otimes 2} X_{\mathrm{A}ij} b'''\big((\beta_{\mathrm{A}} + U_i)^T X_{\mathrm{A}ij} + \beta_{\mathrm{B}}^T X_{\mathrm{B}ij}\big) + O_p(n^{-3/2})\\
&\quad = b'\big((\beta_{\mathrm{A}} + U_i)^T X_{\mathrm{A}ij} + \beta_{\mathrm{B}}^T X_{\mathrm{B}ij}\big)\\
&\qquad + X_{\mathrm{A}ij}^T \left[\mathcal{H}_{\mathrm{AA}i}^{-1}\mathcal{G}_{\mathrm{A}i} - \frac{1}{2}\mathcal{H}_{\mathrm{AA}i}^{-1}\Big\{\mathcal{H}'_{\mathrm{AAA}i}\bigstar\Big(\mathcal{H}_{\mathrm{AA}i}^{-1}\mathcal{G}_{\mathrm{A}i}\mathcal{G}_{\mathrm{A}i}^T\mathcal{H}_{\mathrm{AA}i}^{-1}\Big)\Big\}\right]\\
&\qquad\qquad \times b''\big((\beta_{\mathrm{A}} + U_i)^T X_{\mathrm{A}ij} + \beta_{\mathrm{B}}^T X_{\mathrm{B}ij}\big)\\
&\qquad + \frac{1}{2} X_{\mathrm{A}ij}^T \Big(\mathcal{H}_{\mathrm{AA}i}^{-1}\mathcal{G}_{\mathrm{A}i}\Big)^{\otimes 2} X_{\mathrm{A}ij} b'''\big((\beta_{\mathrm{A}} + U_i)^T X_{\mathrm{A}ij} + \beta_{\mathrm{B}}^T X_{\mathrm{B}ij}\big) + O_p(n^{-3/2}).
\end{aligned}$$

Substitution of this result into (S.19) leads to the first term of $S_{\mathrm{B}i}$ equalling

$$\begin{aligned}
&\frac{1}{\phi}\Big(\mathcal{G}_{\mathrm{B}i} - \mathcal{H}_{\mathrm{AB}i}^T\mathcal{H}_{\mathrm{AA}i}^{-1}\mathcal{G}_{\mathrm{A}i}\Big) + \frac{1}{2\phi}\mathcal{H}_{\mathrm{AB}i}^T\mathcal{H}_{\mathrm{AA}i}^{-1}\Big\{\mathcal{H}'_{\mathrm{AAA}i}\bigstar\Big(\mathcal{H}_{\mathrm{AA}i}^{-1}\mathcal{G}_{\mathrm{A}i}\mathcal{G}_{\mathrm{A}i}^T\mathcal{H}_{\mathrm{AA}i}^{-1}\Big)\Big\}\\
&\qquad - \frac{1}{2\phi}\Big\{\mathcal{H}'_{\mathrm{AAB}i}\bigstar\Big(\mathcal{H}_{\mathrm{AA}i}^{-1}\mathcal{G}_{\mathrm{A}i}\mathcal{G}_{\mathrm{A}i}^T\mathcal{H}_{\mathrm{AA}i}^{-1}\Big)\Big\} + O_p(n^{-1/2})1_{d_{\mathrm{R}}}.
\end{aligned}$$

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

Noting that

$$\nabla\{e_r^T g_{iB}(u)\} = -\frac{1}{\phi}\sum_{j=1}^{n_i} b''\big((\beta_{\mathrm{A}} + u)^T X_{\mathrm{A}ij} + \beta_{\mathrm{B}}^T X_{\mathrm{B}ij}\big) e_r^T X_{\mathrm{A}ij} X_{\mathrm{B}ij}^T \tag{S.20}$$

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 $S_{\mathrm{B}i}$ is

$$\begin{aligned}
&\left\{\sum_{j=1}^{n_i} b''\big((\beta_{\mathrm{A}} + U_i^*)^T X_{\mathrm{A}ij} + \beta_{\mathrm{B}}^T X_{\mathrm{B}ij}\big) X_{\mathrm{A}ij} X_{\mathrm{B}ij}^T\right\}^T\\
&\qquad \times \left\{\sum_{j=1}^{n_i} b''\big((\beta_{\mathrm{A}} + U_i^*)^T X_{\mathrm{A}ij} + \beta_{\mathrm{B}}^T X_{\mathrm{B}ij}\big) X_{\mathrm{A}ij} X_{\mathrm{A}ij}^T\right\}^{-1} \Sigma^{-1} U_i^*.
\end{aligned}$$

Substitution of (S.16) then leads to the contribution to $S_{\mathrm{B}i}$ from the second term of (S.10) equalling

$$\mathcal{H}_{\mathrm{AB}i}^T\mathcal{H}_{\mathrm{AA}i}^{-1}\Sigma^{-1}U_i + O_p(n^{-1/2})1_{d_{\mathrm{R}}}.$$

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

The rth entry of the contribution to $S_{\mathrm{B}i}$ from the third term of (S.10) is

$$\frac{1}{2n}\sum_{s=1}^{d_{\mathrm{B}}}\sum_{t=1}^{d_{\mathrm{B}}}\big[\nabla^2\{e_r^T g_{iB}(U_i^*)\}\big]_{st}\big[\{\nabla^2 h_i(U_i^*)\}^{-1}\big]_{st}$$

$$=\frac{\phi}{2}\sum_{s=1}^{d_{\mathrm{B}}}\sum_{t=1}^{d_{\mathrm{B}}}\big[\nabla^2\{e_r^T g_{iB}(U_i)\}\big]_{st}\big(\mathcal{H}_{\mathrm{AA}i}^{-1}\big)_{st}+O_p(n^{-1/2}).$$

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

$$-\frac{1}{\phi}\sum_{j=1}^{n_i}b'''\big((\beta_{\mathrm{A}}+U_i)^T X_{\mathrm{A}ij}+\beta_{\mathrm{B}}^T X_{\mathrm{B}ij}\big)(e_r^T X_{\mathrm{A}ij})(e_s^T X_{\mathrm{A}ij})(e_t^T X_{\mathrm{B}ij})=-\frac{1}{\phi}\big(\mathcal{H}'_{\mathrm{AAB}i}\big)_{rst}.$$

Noting (6), the contribution to $S_{\mathrm{B}i}$ from the third term of (S.10) is

$$-\frac{1}{2}\mathcal{H}'_{\mathrm{AAB}i}\bigstar\mathcal{H}_{\mathrm{AA}i}^{-1}+O_p(n^{-1/2})1_{d_{\mathrm{R}}}.$$

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

With the aid of (S.20), the contribution to $S_{\mathrm{B}i}$ from the fourth term of (S.10) is

$$\frac{1}{2n\phi}\left\{\sum_{j=1}^{n_i}b''\big((\beta_{\mathrm{A}}+U_i^*)^T X_{\mathrm{A}ij}+\beta_{\mathrm{B}}^T X_{\mathrm{B}ij}\big)X_{\mathrm{A}ij}X_{\mathrm{B}ij}^T\right\}^T\{\nabla^2 h_i(U_i^*)\}^{-1}\left[\nabla^3 h_i(U_i^*)\bigstar\{\nabla^2 h_i(U_i^*)\}^{-1}\right]$$

$$=\frac{1}{2}\mathcal{H}_{\mathrm{AB}i}^T\mathcal{H}_{\mathrm{AA}i}^{-1}\Big(\mathcal{H}'_{\mathrm{AAA}i}\bigstar\mathcal{H}_{\mathrm{AA}i}^{-1}\Big)+O_p(n^{-1/2})1_{d_{\mathrm{R}}}.$$

S.9.2.5 The Resultant Score Approximation

On combining each of the contributions, we obtain

$$\begin{aligned}S_{\mathrm{B}i}=&\frac{1}{\phi}\Big(\mathcal{G}_{\mathrm{B}i}-\mathcal{H}_{\mathrm{AB}i}^T\mathcal{H}_{\mathrm{AA}i}^{-1}\mathcal{G}_{\mathrm{A}i}\Big)+\frac{1}{2\phi}\mathcal{H}_{\mathrm{AB}i}^T\mathcal{H}_{\mathrm{AA}i}^{-1}\Big\{\mathcal{H}'_{\mathrm{AAA}i}\bigstar\Big(\mathcal{H}_{\mathrm{AA}i}^{-1}\mathcal{G}_{\mathrm{A}i}\mathcal{G}_{\mathrm{A}i}^T\mathcal{H}_{\mathrm{AA}i}^{-1}\Big)\Big\}\\&-\frac{1}{2\phi}\Big\{\mathcal{H}'_{\mathrm{AAB}i}\bigstar\Big(\mathcal{H}_{\mathrm{AA}i}^{-1}\mathcal{G}_{\mathrm{A}i}\mathcal{G}_{\mathrm{A}i}^T\mathcal{H}_{\mathrm{AA}i}^{-1}\Big)\Big\}+\mathcal{H}_{\mathrm{AB}i}^T\mathcal{H}_{\mathrm{AA}i}^{-1}\Sigma^{-1}U_i\\&-\frac{1}{2}\mathcal{H}'_{\mathrm{AAB}i}\bigstar\mathcal{H}_{\mathrm{AA}i}^{-1}+\frac{1}{2}\mathcal{H}_{\mathrm{AB}i}^T\mathcal{H}_{\mathrm{AA}i}^{-1}\Big(\mathcal{H}'_{\mathrm{AAA}i}\bigstar\mathcal{H}_{\mathrm{AA}i}^{-1}\Big)+O_p(n^{-1/2})1_{d_{\mathrm{R}}}.\end{aligned}\tag{S.21}$$

S.9.3 Approximation of $S_{\mathrm{C}i}$

For each $1\le r\le\frac{1}{2}d_{\mathrm{R}}(d_{\mathrm{R}}+1)$ let e_r denote the $d_{\mathrm{R}}(d_{\mathrm{R}}+1)/2\times 1$ vector with 1 in the rth position and zeroes elsewhere.

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

For each $1\le r\le\frac{1}{2}d_{\mathrm{R}}(d_{\mathrm{R}}+1)$, the rth entry of the contribution to $S_{\mathrm{C}i}$ from the first term of (S.10) is

$$e_r^T g_{iC}(U_i^*)=\frac{1}{2}e_r^T D_{d_{\mathrm{R}}}^T(\Sigma^{-1}\otimes\Sigma^{-1})\mathrm{vec}\big((U_i^*)^{\otimes 2}\big).$$

Since

$$
\begin{aligned}
(U_i^*)^{\otimes 2} &= \Big[U_i + \mathcal{H}_{\mathrm{AA}i}^{-1}\mathcal{G}_{\mathrm{A}i} - \tfrac{1}{2}\mathcal{H}_{\mathrm{AA}i}^{-1}\Big\{\mathcal{H}'_{\mathrm{AAA}i}\bigstar\Big(\mathcal{H}_{\mathrm{AA}i}^{-1}\mathcal{G}_{\mathrm{A}i}\mathcal{G}_{\mathrm{A}i}^T\mathcal{H}_{\mathrm{AA}i}^{-1}\Big)\Big\} + O_p(n^{-3/2})1_{d_{\mathrm{R}}}\Big]^{\otimes 2} \\
&= U_i^{\otimes 2} + U_i\mathcal{G}_{\mathrm{A}i}^T\mathcal{H}_{\mathrm{AA}i}^{-1} + \mathcal{H}_{\mathrm{AA}i}^{-1}\mathcal{G}_{\mathrm{A}i}U_i^T + \mathcal{H}_{\mathrm{AA}i}^{-1}\mathcal{G}_{\mathrm{A}i}\mathcal{G}_{\mathrm{A}i}^T\mathcal{H}_{\mathrm{AA}i}^{-1} \\
&\quad -\tfrac{1}{2}U_i\Big\{\mathcal{H}'_{\mathrm{AAA}i}\bigstar\Big(\mathcal{H}_{\mathrm{AA}i}^{-1}\mathcal{G}_{\mathrm{A}i}\mathcal{G}_{\mathrm{A}i}^T\mathcal{H}_{\mathrm{AA}i}^{-1}\Big)\Big\}^T\mathcal{H}_{\mathrm{AA}i}^{-1} \\
&\quad -\tfrac{1}{2}\mathcal{H}_{\mathrm{AA}i}^{-1}\Big\{\mathcal{H}'_{\mathrm{AAA}i}\bigstar\Big(\mathcal{H}_{\mathrm{AA}i}^{-1}\mathcal{G}_{\mathrm{A}i}\mathcal{G}_{\mathrm{A}i}^T\mathcal{H}_{\mathrm{AA}i}^{-1}\Big)\Big\}U_i^T + O_p(n^{-3/2})1_{d_{\mathrm{R}}}^{\otimes 2},
\end{aligned}
$$

and noting (S.3) and (S.6), the contribution to $S_{\mathrm{C}i}$ from the first term of (S.10) is

$$
\begin{aligned}
&\frac{1}{2}D_{d_{\mathrm{R}}}^T\mathrm{vec}\Big(\Sigma^{-1}\Big[U_iU_i^T + 2\mathcal{H}_{\mathrm{AA}i}^{-1}\mathcal{G}_{\mathrm{A}i}U_i^T + \mathcal{H}_{\mathrm{AA}i}^{-1}\mathcal{G}_{\mathrm{A}i}\mathcal{G}_{\mathrm{A}i}^T\mathcal{H}_{\mathrm{AA}i}^{-1} \\
&\qquad -\mathcal{H}_{\mathrm{AA}i}^{-1}\Big\{\mathcal{H}'_{\mathrm{AAA}i}\bigstar\Big(\mathcal{H}_{\mathrm{AA}i}^{-1}\mathcal{G}_{\mathrm{A}i}\mathcal{G}_{\mathrm{A}i}^T\mathcal{H}_{\mathrm{AA}i}^{-1}\Big)\Big\}U_i^T\Big]\Sigma^{-1}\Big) + O_p(n^{-3/2})1_{d_{\mathrm{R}}(d_{\mathrm{R}}+1)/2.}
\end{aligned}
$$

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

Noting that, for each $1 \le r \le d_{\mathrm{R}}(d_{\mathrm{R}}+1)/2$,

$$
\big[\nabla\{e_r^Tg_{iC}(u)\}\big]^T = \frac{1}{2}e_r^TD_{d_{\mathrm{R}}}^T(\Sigma^{-1}\otimes\Sigma^{-1})\big\{(u\otimes I)+(I\otimes u)\big\} \tag{S.22}
$$

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 $S_{\mathrm{C}i}$ is

$$
\begin{aligned}
&-\frac{\phi}{2}D_{d_{\mathrm{R}}}^T(\Sigma^{-1}\otimes\Sigma^{-1})\big\{(U_i^*\otimes I)+(I\otimes U_i^*)\big\} \\
&\qquad\times\left\{\sum_{j=1}^{n_i}b''\big((\beta_{\mathrm{A}}+U_i^*)^TX_{\mathrm{A}ij}+\beta_{\mathrm{B}}^TX_{\mathrm{B}ij}\big)X_{\mathrm{A}ij}X_{\mathrm{A}ij}^T\right\}^{-1}\Sigma^{-1}U_i^* \\
&= -\frac{\phi}{2}D_{d_{\mathrm{R}}}^T(\Sigma^{-1}\otimes\Sigma^{-1})\big\{(U_i\otimes I)+(I\otimes U_i)\big\}\mathcal{H}_{\mathrm{AA}i}^{-1}\Sigma^{-1}U_i + O_p(n^{-3/2})1_{d_{\mathrm{R}}(d_{\mathrm{R}}+1)/2} \\
&= -\phi D_{d_{\mathrm{R}}}^T\mathrm{vec}\Big(\Sigma^{-1}\mathcal{H}_{\mathrm{AA}i}^{-1}\Sigma^{-1}U_iU_i^T\Sigma^{-1}\Big) + O_p(n^{-3/2})1_{d_{\mathrm{R}}(d_{\mathrm{R}}+1)/2}.
\end{aligned}
$$

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 $S_{\mathrm{C}i}$ 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 $\mathrm{vec}^{-1}(b)$ is the $d \times d$ matrix such that $\mathrm{vec}\big(\mathrm{vec}^{-1}(b)\big) = b$.

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

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

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

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

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

For each $1 \le r \le \frac{1}{2}d_{\text{R}}(d_{\text{R}}+1)$, the rth entry of the contribution to $S_{\text{C}i}$ from the third term of (S.10) is

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

Next note from (S.22) that

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

Using Lemma 1 we then have

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

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

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

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:

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

Now note that

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

The rth entry of the leading term of the contribution to $S_{\text{C}i}$ from the third term on the right-hand side of (S.10) is

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

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

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

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 rth entry of the contribution to S_{Ci} from the fourth term on the right-hand side of (S.10) is

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

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

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

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_{Ci} is

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

S.10 Score Outer Product Conditional Moments Approximation

The ith term of the Fisher information matrix of $\big(\beta,\text{vech}(\Sigma)\big)$ 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 and keeping track of orders of magnitude.

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

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

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

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

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

$$E(S_{Bi}^{\otimes 2}|X_i) = \frac{1}{\phi}E\Big(\mathcal{H}_{\text{BB}i} - \mathcal{H}_{\text{AB}i}^T\mathcal{H}_{\text{AA}i}^{-1}\mathcal{H}_{\text{AB}i}\big|X_i\Big) + O_p(1)1_{d_\text{B}}^{\otimes 2}. \tag{S.25}$$

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

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

$$\begin{aligned}E(S_{\text{C}i}^{\otimes 2}|X_i) &= \frac{1}{2}D_{d_\text{R}}^T(\Sigma^{-1}\otimes\Sigma^{-1})D_{d_\text{R}} + \tfrac{\phi}{2}D_{d_\text{R}}^T(\Sigma^{-1}\otimes\Sigma^{-1})E\Big[2(U_iU_i^T)\otimes\mathcal{H}_{\text{AA}i}^{-1} \\ &\quad +\text{vec}(U_iU_i^T-\Sigma)\text{vec}\Big(\mathcal{H}_{\text{AA}i}^{-1}\Sigma^{-1}\Big\{\Sigma - U_iU_i^T - \Sigma\Big(\mathcal{H}'_{\text{AAA}i}\bigstar\mathcal{H}_{\text{AA}i}^{-1}\Big)U_i^T\Big\}\Big)^T \\ &\quad +\text{vec}\Big(\mathcal{H}_{\text{AA}i}^{-1}\Sigma^{-1}\Big\{\Sigma - U_iU_i^T - \Sigma\Big(\mathcal{H}'_{\text{AAA}i}\bigstar\mathcal{H}_{\text{AA}i}^{-1}\Big)U_i^T\Big\}\Big)\text{vec}(U_iU_i^T-\Sigma)^T\Big|X_i\Big] \\ &\quad \times(\Sigma^{-1}\otimes\Sigma^{-1})D_{d_\text{R}} + O_p(n^{-2})1_{d_\text{R}(d_\text{R}+1)/2}^{\otimes 2}.\end{aligned} \tag{S.26}$$

S.10.4 Approximation of $E(S_{\mathbf{A}i}S_{\mathbf{B}i}^T|X_i)$

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

$$\begin{aligned}E(S_{\text{A}i}S_{\text{B}i}^T|X_i) = \Sigma^{-1}E\Big\{&U_iU_i^T\Sigma^{-1}\mathcal{H}_{\text{AA}i}^{-1}\mathcal{H}_{\text{AB}i} - U_i\Big(\mathcal{H}'_{\text{AAB}i}\bigstar\mathcal{H}_{\text{AA}i}^{-1}\Big)^T \\ &+U_i\Big(\mathcal{H}'_{\text{AAA}i}\bigstar\mathcal{H}_{\text{AA}i}^{-1}\Big)^T\mathcal{H}_{\text{AA}i}^{-1}\mathcal{H}_{\text{AB}i}\Big|X_i\Big\} + O_p(n^{-1})1_{d_\text{R}}1_{d_\text{B}}^T.\end{aligned} \tag{S.27}$$

S.10.5 Approximation of $E(S_{\mathbf{A}i}S_{\mathbf{C}i}^T|X_i)$

An important aspect of the $E(S_{\text{A}i}S_{\text{C}i}^T|X_i)$ approximation is that, even though

$$S_{\text{A}i} = O_p(1)1_{d_\text{R}} \quad \text{and} \quad S_{\text{C}i} = O_p(1)1_{d_\text{R}(d_\text{R}+1)/2}$$

we can establish that

$$E(S_{\text{A}i}S_{\text{C}i}^T|X_i) = O_p(n^{-1})1_{d_\text{R}}1_{d_\text{R}(d_\text{R}+1)/2}^T, \tag{S.28}$$

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

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

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

S.10.6 Approximation of $E(S_{\mathbf{B}i}S_{\mathbf{C}i}^T|X_i)$

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

$$\begin{aligned}E\big(S_{\text{B}i}S_{\text{C}i}^T|X_i\big) &= \frac{1}{2}E\Big[\Big\{\mathcal{H}_{\text{AB}i}^T\mathcal{H}_{\text{AA}i}^{-1}\Sigma^{-1}U_i - \mathcal{H}'_{\text{AAB}i}\bigstar\mathcal{H}_{\text{AA}i}^{-1} \\ &\quad +\mathcal{H}_{\text{AB}i}^T\mathcal{H}_{\text{AA}i}^{-1}\Big(\mathcal{H}'_{\text{AAA}i}\bigstar\mathcal{H}_{\text{AA}i}^{-1}\Big)\Big\}\text{vec}(U_iU_i^T-\Sigma)^T\Big|X_i\Big](\Sigma^{-1}\otimes\Sigma^{-1})D_{d_\text{R}} \\ &\quad +O_p(n^{-1})1_{d_\text{B}}1_{d_\text{R}(d_\text{R}+1)/2}^T.\end{aligned} \tag{S.29}$$

S.11 The Fisher Information Matrix

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

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

$$M_{12} \equiv \sum_{i=1}^{m} \begin{bmatrix} E(S_{\text{A}i}S_{\text{C}i}^T|X_i) \\ E(S_{\text{B}i}S_{\text{C}i}^T|X_i) \end{bmatrix} \quad \text{and} \quad M_{22} \equiv \sum_{i=1}^{m} E(S_{\text{C}i}^{\otimes 2}|X_i).$$

The results of the previous section lead to high-order asymptotic approximation of the matrix $I\big(\beta, \text{vech}(\Sigma)\big)$. 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

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

correspond to the $d_{\text{F}} \times d_{\text{F}}$ and $\frac{1}{2}d_{\text{R}}(d_{\text{R}}+1) \times \frac{1}{2}d_{\text{R}}(d_{\text{R}}+1)$ diagonal blocks of

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

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 $\text{Cov}\big(\widehat{\beta}|\mathcal{X}\big)$

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

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

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

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

where

$$\mathcal{K}_{\text{AA}} \equiv \frac{n}{m}\sum_{i=1}^{m} E\Big\{U_iU_i^T\Sigma^{-1}\mathcal{H}_{\text{AA}i}^{-1} + \mathcal{H}_{\text{AA}i}^{-1}\Sigma^{-1}U_iU_i^T - \mathcal{H}_{\text{AA}i}^{-1} + \mathcal{H}_{\text{AA}i}^{-1}\Big(\mathcal{H}'_{\text{AAA}i}\bigstar\mathcal{H}_{\text{AA}i}^{-1}\Big)U_i^T + U_i\Big(\mathcal{H}'_{\text{AAA}i}\bigstar\mathcal{H}_{\text{AA}i}^{-1}\Big)^T\mathcal{H}_{\text{AA}i}^{-1}\Big|X_i\Big\},$$

$$\mathcal{K}_{\text{AB}} \equiv \frac{1}{m}\sum_{i=1}^{m} E\Big\{U_iU_i^T\Sigma^{-1}\mathcal{H}_{\text{AA}i}^{-1}\mathcal{H}_{\text{AB}i} - U_i\Big(\mathcal{H}'_{\text{AAB}i}\bigstar\mathcal{H}_{\text{AA}i}^{-1}\Big)^T + U_i\Big(\mathcal{H}'_{\text{AAA}i}\bigstar\mathcal{H}_{\text{AA}i}^{-1}\Big)^T\mathcal{H}_{\text{AA}i}^{-1}\mathcal{H}_{\text{AB}i}\Big|X_i\Big\}$$

and $$\mathcal{K}_{\text{BB}} \equiv \frac{1}{mn}\sum_{i=1}^{m} E\Big(\mathcal{H}_{\text{BB}i} - \mathcal{H}_{\text{AB}i}^T\mathcal{H}_{\text{AA}i}^{-1}\mathcal{H}_{\text{AB}i}\Big|X_i\Big)$$

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

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

Therefore

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

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

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

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

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

The upper right $d_{\text{R}} \times d_{\text{B}}$ block of $\big(M_{11} - M_{12}M_{22}^{-1}M_{12}^T\big)^{-1}$ is

$$\begin{aligned} &-\Big\{\frac{\Sigma}{m} + O_p(m^{-1}n^{-1})1_{d_{\text{R}}}^{\otimes 2}\Big\}\Big\{m\Sigma^{-1}\mathcal{K}_{\text{AB}} + O_p(mn^{-1})1_{d_{\text{R}}}1_{d_{\text{B}}}^T\Big\}\Big\{\frac{mn}{\phi}\mathcal{K}_{\text{BB}} + O_p(m)1_{d_{\text{B}}}^{\otimes 2}\Big\}^{-1} \\ &= -\frac{\phi\mathcal{K}_{\text{AB}}\mathcal{K}_{\text{BB}}^{-1}}{mn} + O_p(m^{-1}n^{-2})1_{d_{\text{R}}}1_{d_{\text{B}}}^T. \end{aligned}$$

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

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

Therefore,

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

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

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

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

$$\text{the lower right } \tfrac{1}{2}d_{\mathrm{R}}(d_{\mathrm{R}}+1) \times \tfrac{1}{2}d_{\mathrm{R}}(d_{\mathrm{R}}+1) \text{ block of } I\big(\beta, \mathrm{vech}(\Sigma)\big)^{-1} = \big(M_{22} - M_{12}^T M_{11}^{-1} M_{12}\big)^{-1}.$$

From (S.26)

$$\sum_{i=1}^m E(S_{\mathrm{C}i}S_{\mathrm{C}i}^T|X_i) = \frac{m}{2}D_{d_{\mathrm{R}}}^T(\Sigma^{-1}\otimes\Sigma^{-1})D_{d_{\mathrm{R}}} - \frac{m\phi}{n}D_{d_{\mathrm{R}}}^T(\Sigma^{-1}\otimes\Sigma^{-1})\mathcal{K}_{\mathrm{CC}}(\Sigma^{-1}\otimes\Sigma^{-1})D_{d_{\mathrm{R}}} + O_p(mn^{-2})1^{\otimes 2}_{d_{\mathrm{R}}(d_{\mathrm{R}}+1)/2}$$

with the following $O_p(1)$ matrix:

$$\begin{aligned}\mathcal{K}_{\mathrm{CC}} = \frac{n}{m}\sum_{i=1}^m E\Big[&\frac{1}{2}\mathrm{vec}(\Sigma - U_iU_i^T)\mathrm{vec}\left(\mathcal{H}_{\mathrm{AA}i}^{-1}\Sigma^{-1}\Big\{\Sigma - U_iU_i^T - \Sigma\Big(\mathcal{H}'_{\mathrm{AAA}i}\bigstar\mathcal{H}_{\mathrm{AA}i}^{-1}\Big)U_i^T\Big\}\right)^T \\ &+\frac{1}{2}\mathrm{vec}\left(\mathcal{H}_{\mathrm{AA}i}^{-1}\Sigma^{-1}\Big\{\Sigma - U_iU_i^T - \Sigma\Big(\mathcal{H}'_{\mathrm{AAA}i}\bigstar\mathcal{H}_{\mathrm{AA}i}^{-1}\Big)U_i^T\Big\}\right)\mathrm{vec}(\Sigma - U_iU_i^T)^T \\ &-(U_iU_i^T)\otimes(\mathcal{H}_{\mathrm{AA}i}^{-1})\Big|X_i\Big].\end{aligned}$$

Next, note that

$$M_{11}^{-1} = \begin{bmatrix} O_p(m^{-1})1^{\otimes 2}_{d_{\mathrm{R}}} & O_p\{(mn)^{-1}\}1_{d_{\mathrm{R}}}1_{d_{\mathrm{B}}}^T \\ O_p\{(mn)^{-1}\}1_{d_{\mathrm{B}}}1_{d_{\mathrm{R}}}^T & O_p\{(mn)^{-1}\}1^{\otimes 2}_{d_{\mathrm{B}}}\end{bmatrix}. \tag{S.31}$$

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

$$\left\{\sum_{i=1}^m E(S_{\mathrm{B}i}S_{\mathrm{C}i}^T|X_i)\right\}^T \left\{\sum_{i=1}^m E(S_{\mathrm{B}i}S_{\mathrm{B}i}^T|X_i)\right\}^{-1} \sum_{i=1}^m E(S_{\mathrm{B}i}S_{\mathrm{C}i}^T|X_i) = \frac{\phi m}{n}D_{d_{\mathrm{R}}}^T(\Sigma^{-1}\otimes\Sigma^{-1})\mathcal{K}_{\mathrm{BC}}\mathcal{K}_{\mathrm{BB}}^{-1}\mathcal{K}_{\mathrm{BC}}(\Sigma^{-1}\otimes\Sigma^{-1})D_{d_{\mathrm{R}}} + o_p(m/n)1^{\otimes 2}_{d_{\mathrm{R}}(d_{\mathrm{R}}+1)/2}$$

where

$$\begin{aligned}\mathcal{K}_{\mathrm{BC}} \equiv \frac{1}{2m}\sum_{i=1}^m E\Big[\Big\{&\mathcal{H}_{\mathrm{AB}i}^T\mathcal{H}_{\mathrm{AA}i}^{-1}\Sigma^{-1}U_i - \mathcal{H}'_{\mathrm{AAB}i}\bigstar\mathcal{H}_{\mathrm{AA}i}^{-1} \\ &+\mathcal{H}_{\mathrm{AB}i}^T\mathcal{H}_{\mathrm{AA}i}^{-1}\Big(\mathcal{H}'_{\mathrm{AAA}i}\bigstar\mathcal{H}_{\mathrm{AA}i}^{-1}\Big)\Big\}\Big|X_i\Big]\mathrm{vec}(U_iU_i^T - \Sigma)^T\end{aligned}$$

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

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

then

$$\text{Cov}\big(\text{vech}(\widehat{\Sigma})|\mathcal{X}\big) = \frac{1}{m}\left(\mathcal{A} - \frac{\phi}{n}\mathcal{B}\right)^{-1} + o_p\{(mn)^{-1}\}1^{\otimes 2}_{d_{\text{R}}(d_{\text{R}}+1)/2}$$

$$= \frac{1}{m}\mathcal{A}^{-1} + \frac{\phi}{mn}\mathcal{A}^{-1}\mathcal{B}\mathcal{A}^{-1} + o_p\{(mn)^{-1}\}1^{\otimes 2}_{d_{\text{R}}(d_{\text{R}}+1)/2}.$$

From (S.5),

$$\mathcal{A}^{-1} = 2D^{+}_{d_{\text{R}}}(\Sigma^0 \otimes \Sigma^0)D^{+T}_{d_{\text{R}}}.$$

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

$$\text{Cov}\big(\text{vech}(\widehat{\Sigma})|\mathcal{X}\big) = \frac{2D^{+}_{d_{\text{R}}}(\Sigma^0 \otimes \Sigma^0)D^{+T}_{d_{\text{R}}}}{m} + \frac{4\phi D^{+}_{d_{\text{R}}}\{\mathcal{K}^0_{\text{CC}} + (\mathcal{K}^0_{\text{BC}})^T(\mathcal{K}^0_{\text{BB}})^{-1}\mathcal{K}^0_{\text{BC}}\}D^{+T}_{d_{\text{R}}}}{mn} + O_p(m^{-1}n^{-2})1^{\otimes 2}_{d_{\text{R}}(d_{\text{R}}+1)/2}. \tag{S.32}$$

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 $\text{vech}\big(\widehat{\Sigma}\big)$ 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 $\mathcal{K}_{\text{AA}}$ quantity is

$$\mathcal{K}_{\text{AA}} \equiv \frac{1}{m}\sum_{i=1}^{m} E\Big[U_iU_i^T\Sigma^{-1}\big(\tfrac{1}{n}\mathcal{H}_{\text{AA}i}\big)^{-1} + \big(\tfrac{1}{n}\mathcal{H}_{\text{AA}i}\big)^{-1}\Sigma^{-1}U_iU_i^T - \big(\tfrac{1}{n}\mathcal{H}_{\text{AA}i}\big)^{-1}$$
$$+\big(\tfrac{1}{n}\mathcal{H}_{\text{AA}i}\big)^{-1}\Big\{\big(\tfrac{1}{n}\mathcal{H}'_{\text{AAA}i}\big)\bigstar\big(\tfrac{1}{n}\mathcal{H}_{\text{AA}i}\big)^{-1}\big)\Big\}U_i^T$$
$$+U_i\Big\{\big(\tfrac{1}{n}\mathcal{H}'_{\text{AAA}i}\big)\bigstar\big(\tfrac{1}{n}\mathcal{H}_{\text{AA}i}\big)^{-1}\Big\}^T\big(\tfrac{1}{n}\mathcal{H}_{\text{AA}i}\big)^{-1}\Big|X_i\Big].$$

Since

$$E\big(\tfrac{1}{n}\mathcal{H}_{\text{AA}i}|X_i\big) \xrightarrow{P} \Omega_{\text{AA}}(U_i) \quad \text{and} \quad E\big(\tfrac{1}{n}\mathcal{H}'_{\text{AAA}i}|X_i\big) \xrightarrow{P} \Omega'_{\text{AAA}}(U_i)$$

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

$$\mathcal{K}_{\text{AA}} \xrightarrow{P} \frac{1}{m}\sum_{i=1}^{m} E\Big[U_iU_i^T\Sigma^{-1}\Omega_{\text{AA}}(U_i)^{-1} + \Omega_{\text{AA}}(U_i)^{-1}\Sigma^{-1}U_iU_i^T - \Omega_{\text{AA}}(U_i)^{-1}$$
$$+\Omega_{\text{AA}}(U_i)^{-1}\Big\{\Omega'_{\text{AAA}}(U_i)\bigstar\Omega_{\text{AA}}(U_i)^{-1}\Big\}U_i^T$$
$$+U_i\Big\{\Omega'_{\text{AAA}}(U_i)\bigstar\Omega_{\text{AA}}(U_i)^{-1}\Big\}^T\Omega_{\text{AA}}(U_i)^{-1}\Big]$$
$$= E\Big[UU^T\Sigma^{-1}\Omega_{\text{AA}}(U)^{-1} + \Omega_{\text{AA}}(U)^{-1}\Sigma^{-1}UU^T - \Omega_{\text{AA}}(U)^{-1}$$
$$+\Omega_{\text{AA}}(U)^{-1}\Big\{\Omega'_{\text{AAA}}(U)\bigstar\Omega_{\text{AA}}(U)^{-1}\Big\}U^T + U\Big\{\Omega'_{\text{AAA}}(U)\bigstar\Omega_{\text{AA}}(U)^{-1}\Big\}^T\Omega_{\text{AA}}(U)^{-1}\Big]$$
$$= \Lambda_{\text{AA}}$$

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

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

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

$$\Omega_{\text{AA}}(U) = E(X_{\text{A}}X_{\text{A}}^T), \quad \Omega_{\text{AB}}(U) = E(X_{\text{A}}X_{\text{B}}^T), \quad \Omega_{\text{BB}}(U) = E(X_{\text{B}}X_{\text{B}}^T)$$

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

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

For the Gaussian response situation

$$\Lambda_{\text{AA}} = E(X_{\text{A}}X_{\text{A}}^T)^{-1}, \quad \Lambda_{\text{AB}} = E(X_{\text{A}}X_{\text{A}}^T)^{-1}E(X_{\text{A}}X_{\text{B}}^T)$$

and

$$E\{\Psi_6(U)\} = E(X_{\text{B}}X_{\text{B}}^T) - E(X_{\text{A}}X_{\text{B}}^T)^T E(X_{\text{A}}X_{\text{A}}^T)^{-1}E(X_{\text{A}}X_{\text{B}}^T).$$

Therefore,

$$\left[\begin{array}{cc} \Lambda_{\text{AA}}^{-1} & \Lambda_{\text{AA}}^{-1}\Lambda_{\text{AB}} \\ \Lambda_{\text{AB}}^T\Lambda_{\text{AA}}^{-1} & \Lambda_{\text{AB}}^T\Lambda_{\text{AA}}^{-1}\Lambda_{\text{AB}} + E\{\Psi_6(U)\} \end{array}\right] = E\left[\begin{array}{cc} X_{\text{A}}X_{\text{A}}^T & X_{\text{A}}X_{\text{B}}^T \\ X_{\text{B}}X_{\text{A}}^T & X_{\text{B}}X_{\text{B}}^T \end{array}\right] = E(XX^T).$$

Hence, for the Gaussian special case

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

This result generalises the two-term expansion of $\text{Var}\big(\widehat{\beta}_{\text{A}}|\mathcal{X}\big)$ provided in § 3.5 of McCulloch *et al.* (2008) for the $d_{\text{R}} = d_{\text{B}} = 1$ and $X_{\text{A}} = 1$ special case.

S.14.2 The $\text{Cov}\big(\text{vech}(\widehat{\Sigma})|\mathcal{X}\big)$ 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 $\text{Cov}\big(\text{vech}(\widehat{\Sigma})|\mathcal{X}\big)$ is

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

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

$$\psi_4(U) = D_{d_{\text{R}}}^{+}\text{vec}\Big(\{E(X_{\text{A}}X_{\text{A}}^T)\}^{-1}\Sigma^{-1}(\Sigma - UU^T)\Big)$$

and

$$\Psi_8(U) = D_{d_{\text{R}}}^{+}\big[(UU^T) \otimes \{E(X_{\text{A}}X_{\text{A}}^T)\}^{-1}\big]D_{d_{\text{R}}}^{+T}.$$

We immediately have

$$E\{\Psi_8(U)\} = D_{d_{\text{R}}}^{+}\big[\Sigma \otimes \{E(X_{\text{A}}X_{\text{A}}^T)\}^{-1}\big]D_{d_{\text{R}}}^{+T}.$$

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

$$E\Big\{\text{vech}(\Sigma - UU^T)\psi_4(U)^T + \psi_4(U)\text{vech}(\Sigma - UU^T)^T\Big\} = 4D_{d_{\text{R}}}^{+}\big[\Sigma \otimes \{E(X_{\text{A}}X_{\text{A}}^T)\}^{-1}\big]D_{d_{\text{R}}}^{+T}.$$

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

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

S.15 Confidence Interval Construction Details

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

$$\begin{aligned}\widehat{\Omega}_{\mathrm{AA}}(u) &\equiv \frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n_i} b''\Big((\widehat{\beta}_{\mathrm{A}}+u)^T X_{\mathrm{A}ij}+\widehat{\beta}_{\mathrm{B}}^T X_{\mathrm{B}ij}\Big) X_{\mathrm{A}ij}X_{\mathrm{A}ij}^T,\\ \widehat{\Omega}_{\mathrm{AB}}(u) &\equiv \frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n_i} b''\Big((\widehat{\beta}_{\mathrm{A}}+u)^T X_{\mathrm{A}ij}+\widehat{\beta}_{\mathrm{B}}^T X_{\mathrm{B}ij}\Big) X_{\mathrm{A}ij}X_{\mathrm{B}ij}^T\\ \text{and}\quad \widehat{\Omega}_{\mathrm{BB}}(u) &\equiv \frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n_i} b''\Big((\widehat{\beta}_{\mathrm{A}}+u)^T X_{\mathrm{A}ij}+\widehat{\beta}_{\mathrm{B}}^T X_{\mathrm{B}ij}\Big) X_{\mathrm{B}ij}X_{\mathrm{B}ij}^T.\end{aligned} \tag{S.34}$$

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

$$\begin{aligned}\widehat{E}\{\Psi_6(U)\} &\equiv E\Big\{\widehat{\Omega}_{\mathrm{BB}}(U)-\widehat{\Omega}_{\mathrm{AB}}(U)^T\widehat{\Omega}_{\mathrm{AA}}(U)^{-1}\widehat{\Omega}_{\mathrm{AB}}(U)\big|\mathcal{X}\Big\}\\ &= |2\pi\widehat{\Sigma}|^{-1/2}\int_{d_{\mathrm{R}}}\Big\{\widehat{\Omega}_{\mathrm{BB}}(u)-\widehat{\Omega}_{\mathrm{AB}}(u)^T\widehat{\Omega}_{\mathrm{AA}}(u)^{-1}\widehat{\Omega}_{\mathrm{AB}}(u)\Big\}\exp\Big(-\tfrac{1}{2}u^T\widehat{\Sigma}u\Big)\,du.\end{aligned} \tag{S.35}$$

In the last expression of (S.35) integration is applied element-wise to each entry of the matrix inside the integral. The natural studentisations of

$$\Lambda_{\mathrm{AA}},\quad \Lambda_{\mathrm{AB}},\quad \Delta,\quad E\{\Psi_8(U)\}\quad\text{and}\quad E\{\Psi_9(U)\} \tag{S.36}$$

are analogous to that for $E\{\Psi_6(U)\}$. The studentisations for the quantities in (S.36) depend on the functions defined by (S.34) as well as similar sample counterparts of $\Omega'_{\mathrm{AAA}}(U)$ and $\Omega'_{\mathrm{AAB}}(U)$. Next define

$$\widehat{\text{Asy.Cov}}(\widehat{\beta}) = \frac{1}{m}\begin{bmatrix}\widehat{\Sigma} & O\\ O & O\end{bmatrix} + \frac{\widehat{\phi}}{mn}\begin{bmatrix}\widehat{\Lambda}_{\mathrm{AA}}^{-1} & \widehat{\Lambda}_{\mathrm{AA}}^{-1}\widehat{\Lambda}_{\mathrm{AB}}\\ \widehat{\Lambda}_{\mathrm{AB}}^T\widehat{\Lambda}_{\mathrm{AA}}^{-1} & \widehat{\Lambda}_{\mathrm{AB}}^T\widehat{\Lambda}_{\mathrm{AA}}^{-1}\widehat{\Lambda}_{\mathrm{AB}}+\widehat{E}\{\Psi_6(U)\}\end{bmatrix}^{-1} \tag{S.37}$$

and

$$\begin{aligned}\widehat{\text{Asy.Cov}}\big(\text{vech}(\widehat{\Sigma})\big) &= \frac{2D_{d_{\mathrm{R}}}^{+}(\widehat{\Sigma}\otimes\widehat{\Sigma})D_{d_{\mathrm{R}}}^{+T}}{m}\\ &\quad + \frac{\widehat{\phi}}{mn}\Big(2\widehat{E}\{\Psi_9(U)\}-4\widehat{E}\{\Psi_8(U)\}+\widehat{\Delta}^T\big[\widehat{E}\{\Psi_6(U)\}\big]^{-1}\widehat{\Delta}\Big).\end{aligned} \tag{S.38}$$

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 kth entry of β^0. Then approximate $100(1-\alpha)\%$ confidence intervals for $(\beta^0)_k$ based on (S.37) are

$$(\widehat{\beta})_k \pm \Phi^{-1}(1-\tfrac{1}{2}\alpha)\sqrt{\Big\{\widehat{\text{Asy.Cov}}(\widehat{\beta})\Big\}_{kk}},\quad 1\le k\le d_{\mathrm{F}}. \tag{S.39}$$

The confidence intervals in (S.39) are analogous to those given in § 4 of Jiang *et al.* (2022). For $1 \le k \le d_{\mathrm{R}}$, (S.39) provides second term improvements of the Jiang *et al.* (2022) confidence intervals. For $d_{\mathrm{R}}+1 \le k \le d_{\mathrm{F}}$ both sets of confidence intervals are identical.

Confidence intervals for the entries of Σ^0 have expressions analogous to (S.39). 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 $d_{\mathrm{R}} = 2$ the parameters are

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

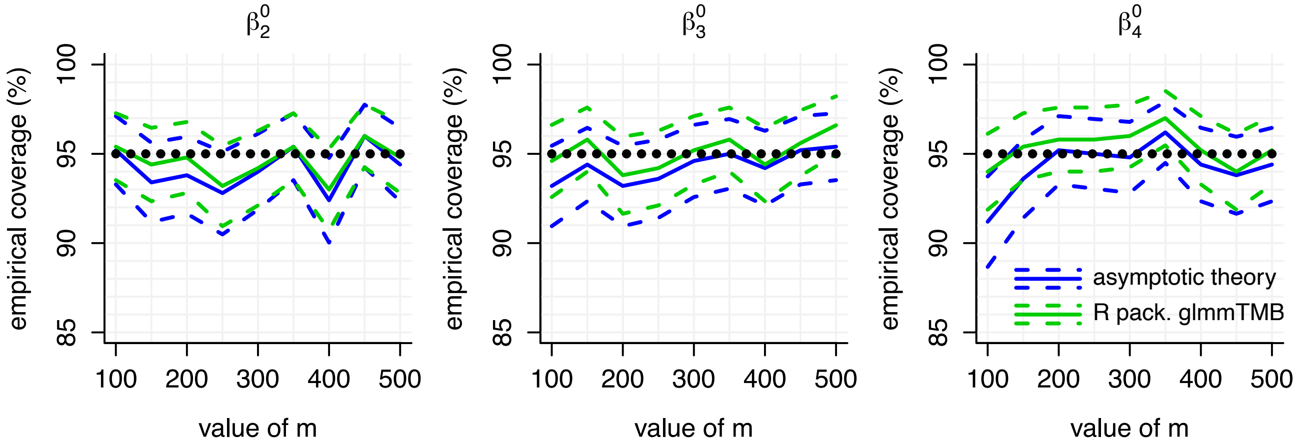

Figure S.1: *Empirical coverage of confidence intervals from the simulation exercise described in § 4. Each panel corresponds to a fixed effect model parameter that is not impacted by second term asymptotic improvements. The advertised coverage level is fixed at 95% and is indicated by a horizontal dotted 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 $m/10$.*

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 $\widehat{\sigma}_1$, $\widehat{\sigma}_2$ and $\widehat{\rho}$, 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

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

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 $(\widehat{\omega}_1, \widehat{\omega}_2, \widehat{\omega}_3)$.

S.16 Additional Simulation Results

The simulation exercise described in § 4 is necessarily limited and subject to constraints with regards to reporting of results. In this section we present an additional figure from that simulation study and 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.

S.16.1 Additional § 4 Simulation Figure

Figure S.1 corresponds to the simulation exercise described in § 4 and compares the empirical coverages of confidence intervals with advertised levels of 95% for the parameters of (7) that are not affected

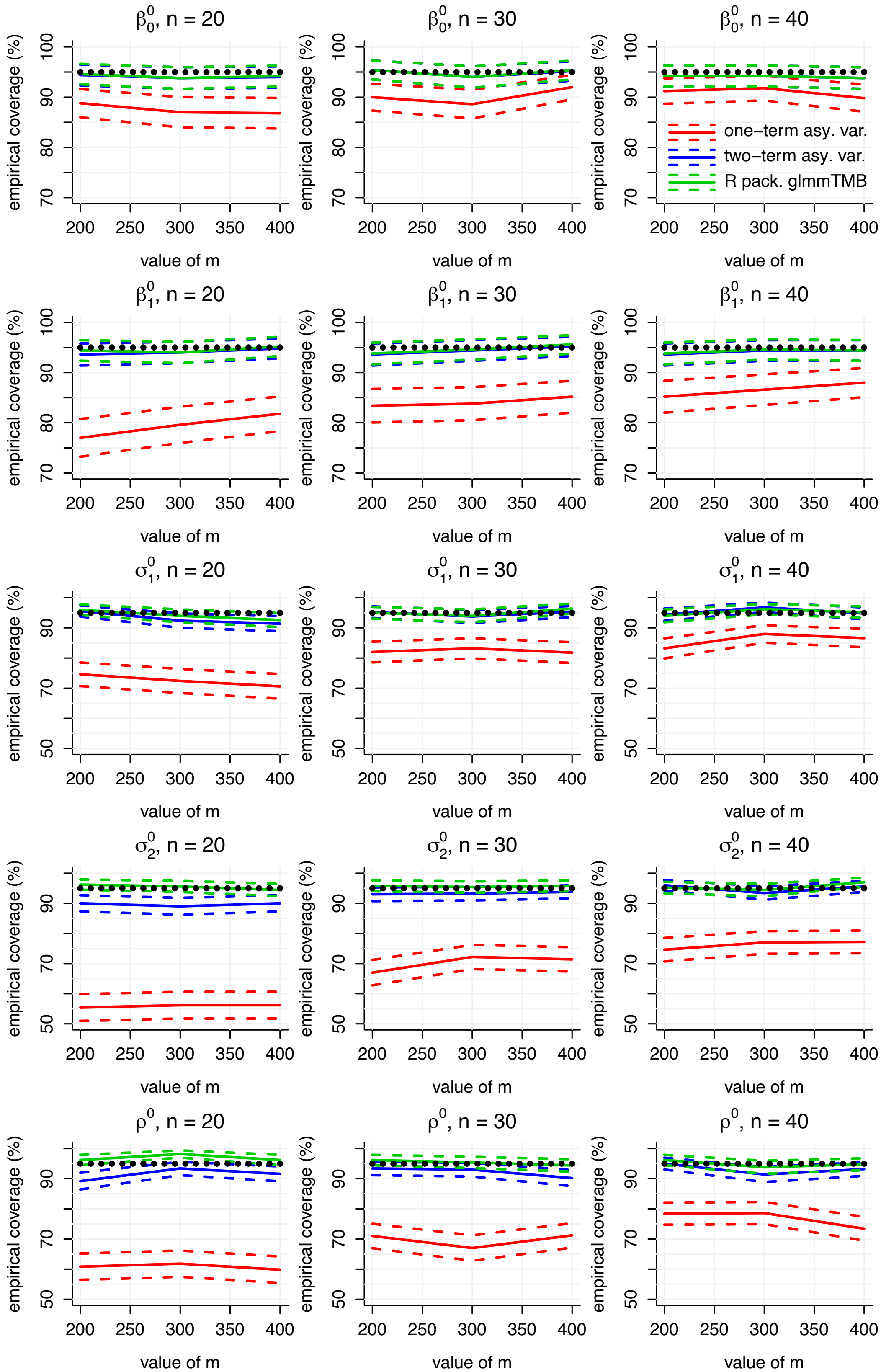

Figure S.2: *Empirical coverage of confidence intervals from the simulation exercise described in the text. Each panel corresponds to a model parameter that is impacted by second term asymptotic improvements. The advertised coverage level is fixed at 95% and is indicated by a horizontal dotted line in each panel. The solid curves show 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.*

by second term improvement. It is clear from Figure S.1 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.

S.16.2 Extension of the § 4 Simulation Exercise

We extended the simulation exercise of § 4 to allow for m and n to vary in an unconstrained manner. The extension involved replacement of the § 4 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.2 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.2 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.

References

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

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.