Supplement for:

Precise Asymptotics for Linear Mixed Models with Crossed Random Effects

Jiming Jiang[1], Matt P. Wand[2] and Swarnadip Ghosh[3]

[1]*University of California, Davis,* [2]*University of Technology Sydney and* [3]*Radix Trading*

S.1 Derivation of Result 1

In this section we provide a derivation of Result 1, starting with notation.

S.1.1 Notation

For any matrix $\boldsymbol{M}$ let

$$\boldsymbol{M}^{\otimes 2} \equiv \boldsymbol{M}\boldsymbol{M}^\top \quad \text{and} \quad \|\boldsymbol{M}\|_F \equiv \{\text{tr}(\boldsymbol{M}^\top\boldsymbol{M})\}^{1/2}.$$

The latter definition is often called the *Frobenius norm* of $\boldsymbol{M}$.

The matrix $\boldsymbol{V}(\boldsymbol{\Sigma},\boldsymbol{\Sigma}',\sigma^2)$ given by (3) is central to the derivations. Throughout this appendix, we omit the dependence on the covariance matrix parameters by simply writing it as $\boldsymbol{V}$. Define the following partitioning of the inverse of $\boldsymbol{V}$:

$$\boldsymbol{V}^{-1} = \begin{bmatrix} \boldsymbol{V}^{11} & \boldsymbol{V}^{12} & \cdots & \boldsymbol{V}^{1m} \\ \boldsymbol{V}^{21} & \boldsymbol{V}^{22} & \cdots & \boldsymbol{V}^{2m} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{V}^{m1} & \boldsymbol{V}^{m2} & \cdots & \boldsymbol{V}^{mm} \end{bmatrix} \quad \text{where} \quad \boldsymbol{V}^{i\underset{\sim}{i}} \ \text{is} \ \left(\sum_{i'=1}^{m'} n_{ii'}\right) \times \left(\sum_{i'=1}^{m'} n_{\underset{\sim}{i}i'}\right).$$

If $\mathcal{P}$ is a logical proposition then $I(\mathcal{P}) = 1$ if $\mathcal{P}$ is true. Otherwise, $I(\mathcal{P}) = 0$.

S.1.2 Lemmas

The upcoming Fisher information approximations rely on four lemmas, which we present here.

S.1.2.1 A Lemma that Provides a Simple Kronecker Product Form

Lemma 1. *Let $\boldsymbol{A}_d$ be a symmetric $d \times d$ matrix with (r, s)th entry denoted by A_{rs}. Also, let $\boldsymbol{B}_d$ be the $\frac{1}{2}d(d+1) \times \frac{1}{2}d(d+1)$ matrix with entries determined according to the following table:*

entry of $\text{vech}(\boldsymbol{A}_d)\text{vech}(\boldsymbol{A}_d)^\top$	*entry of* $\boldsymbol{B}_d$ *in the same position*
$A_{rr}A_{tt}$	A_{rt}^2
$A_{rr}A_{tu}, t \neq u$	$2A_{rt}A_{ru}$
$A_{rs}A_{tu}, r \neq s, t \neq u$	$2(A_{rt}A_{su} + A_{ru}A_{st})$

Table S.1: *Definition of the matrix $\boldsymbol{B}_d$, a function of a $d \times d$ symmetric matrix $\boldsymbol{A}_d$.*

Then

$$\boldsymbol{B}_d = \boldsymbol{D}_d^\top(\boldsymbol{A}_d \otimes \boldsymbol{A}_d)\boldsymbol{D}_d.$$

S.1.2.2 Three Lemmas Stating Key Matrix Identities

The following three lemmas state some matrix identities which play key roles in the derivation of Result 1.

Lemma 2. *Let $\lambda > 0$, $\boldsymbol{A}$ be a invertible $d \times d$ matrix and $\boldsymbol{X}$, $\dot{\boldsymbol{X}}$ and $\ddot{\boldsymbol{X}}$ each be an $n \times d$ matrix, where $n, d \in \mathbb{N}$. Then, assuming that all required matrix inverses exist,*

$$\dot{\boldsymbol{X}}^\top(\boldsymbol{X}\boldsymbol{A}\boldsymbol{X}^\top + \lambda\boldsymbol{I})^{-1}\ddot{\boldsymbol{X}} = (1/\lambda)\dot{\boldsymbol{X}}^\top\{\boldsymbol{I} - \boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top\}\ddot{\boldsymbol{X}}$$
$$+\dot{\boldsymbol{X}}^\top\boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\{\boldsymbol{A} + \lambda(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\}^{-1}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top\ddot{\boldsymbol{X}}.$$

Lemma 2 has the following immediate corollary:

Corollary 2.1. *If λ, $\boldsymbol{A}$, $\boldsymbol{X}$, $\dot{\boldsymbol{X}}$ and $\ddot{\boldsymbol{X}}$ are as defined in Lemma 2 then, under the Lemma 2 assumptions, we have:*

(a) $\boldsymbol{X}^\top(\boldsymbol{X}\boldsymbol{A}\boldsymbol{X}^\top + \lambda\boldsymbol{I})^{-1}\ddot{\boldsymbol{X}} = \{\boldsymbol{A} + \lambda(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\}^{-1}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top\ddot{\boldsymbol{X}}$.

(b) $\dot{\boldsymbol{X}}^\top(\boldsymbol{X}\boldsymbol{A}\boldsymbol{X}^\top + \lambda\boldsymbol{I})^{-1}\boldsymbol{X} = \dot{\boldsymbol{X}}^\top\boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\{\boldsymbol{A} + \lambda(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\}^{-1}$.

(c) $\boldsymbol{X}^\top(\boldsymbol{X}\boldsymbol{A}\boldsymbol{X}^\top + \lambda\boldsymbol{I})^{-1}\boldsymbol{X} = \{\boldsymbol{A} + \lambda(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\}^{-1}$.

The following related matrix identity is also important:

Lemma 3. *If λ, $\boldsymbol{A}$ and $\boldsymbol{X}$ are as defined in Lemma 2 then, assuming all required matrix inverses exist,*

$$\boldsymbol{X}^\top(\boldsymbol{X}\boldsymbol{A}\boldsymbol{X}^\top + \lambda\boldsymbol{I})^{-2}\boldsymbol{X} = \{\boldsymbol{A} + \lambda(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\}^{-1}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\{\boldsymbol{A} + \lambda(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\}^{-1}.$$

In addition, the derivation of Result 1 makes use of:

Lemma 4. *Let $\boldsymbol{A}$ and $\boldsymbol{B}$ be $d \times d$ matrices such that each of*

$$\begin{bmatrix} \boldsymbol{A} & \boldsymbol{B} \\ \boldsymbol{B} & \boldsymbol{A} \end{bmatrix} \text{ and } \boldsymbol{A} + \boldsymbol{B} \text{ are invertible.}$$

Then

$$\begin{bmatrix} \boldsymbol{I}_d \\ \boldsymbol{I}_d \end{bmatrix}^\top \begin{bmatrix} \boldsymbol{A} & \boldsymbol{B} \\ \boldsymbol{B} & \boldsymbol{A} \end{bmatrix}^{-1} \begin{bmatrix} \boldsymbol{I}_d \\ \boldsymbol{I}_d \end{bmatrix} = 2(\boldsymbol{A} + \boldsymbol{B})^{-1}.$$

S.1.2.3 Lemmas for Limits of Forms Arising in the Fisher Information Matrix

Here we provide three convergence in probability lemmas that are key to dealing with particular forms that arise in the Fisher information matrix.

First we present Lemma 5 which identifies some key convergence in probability limits related to predictor summation quantities about the $\boldsymbol{V}^{-1}$ matrix. Let $\boldsymbol{X}_\circ$ be a $d \times 1$ random vector and let

$$\boldsymbol{X}_{ii'j}, \quad 1 \le i \le m, \quad 1 \le i' \le m', \quad 1 \le j \le n_{ii'}, \tag{S.1}$$

be independent and identically distributed random vectors having the same distribution as $\boldsymbol{X}_\circ$. Then define for $1 \le i \le m$ and $1 \le i' \le m'$:

$$\boldsymbol{X}_{ii'} \equiv \begin{bmatrix} \boldsymbol{X}_{ii'1}^\top \\ \vdots \\ \boldsymbol{X}_{ii'n_{ii'}}^\top \end{bmatrix}, \quad \boldsymbol{X} \equiv \underset{1\le i\le m}{\text{stack}}\left(\overset{\blacktriangle}{\boldsymbol{X}}_i\right) \text{ where } \overset{\blacktriangle}{\boldsymbol{X}}_i \equiv \underset{1\le i'\le m'}{\text{stack}}(\boldsymbol{X}_{ii'}). \tag{S.2}$$

Next, let

$$
\begin{aligned}
\boldsymbol{Q}_{mm'} \equiv \operatorname*{blockdiag}_{1\le i\le m}&\left\{\operatorname*{blockmatrix}_{1\le i',\underset{\sim}{i'}\le m'}(\boldsymbol{X}_{ii'}\boldsymbol{M}\boldsymbol{X}_{i\underset{\sim}{i'}}^{\top})\right\}\\
&+\operatorname*{blockmatrix}_{1\le i,\underset{\sim}{i}\le m}\left\{\operatorname*{blockdiag}_{1\le i'\le m'}(\boldsymbol{X}_{ii'}\boldsymbol{M}'\boldsymbol{X}_{\underset{\sim}{i}i'}^{\top})\right\}+\lambda\boldsymbol{I}
\end{aligned}
\tag{S.3}
$$

where

$$\boldsymbol{M} \text{ and } \boldsymbol{M}' \text{ are } d\times d \text{ symmetric positive definite matrices and } \lambda>0. \tag{S.4}$$

Partition $\boldsymbol{Q}_{mm'}^{-1}$ as follows

$$
\boldsymbol{Q}_{mm'}^{-1}=\begin{bmatrix}
\boldsymbol{Q}_{mm'}^{11} & \boldsymbol{Q}_{mm'}^{12} & \cdots & \boldsymbol{Q}_{mm'}^{1m}\\
\boldsymbol{Q}_{mm'}^{21} & \boldsymbol{Q}_{mm'}^{22} & \cdots & \boldsymbol{Q}_{mm'}^{2m}\\
\vdots & \vdots & \ddots & \vdots\\
\boldsymbol{Q}_{mm'}^{m1} & \boldsymbol{Q}_{mm'}^{m2} & \cdots & \boldsymbol{Q}_{mm'}^{mm}
\end{bmatrix}
\quad\text{where}\quad \boldsymbol{Q}_{mm'}^{i\underset{\sim}{i}}\quad\text{is}\quad \left(\sum_{i'=1}^{m'} n_{ii'}\right)\times\left(\sum_{i'=1}^{m'} n_{\underset{\sim}{i}i'}\right). \tag{S.5}
$$

Introduce the following assumptions:

(A4) All entries of $\boldsymbol{X}_\circ$ are not degenerate at zero and have finite second moment.

(A5) Each of $n_{ii'}$, $1\le i\le m,\ 1\le i'\le m'$, diverge to ∞.

Lemma 5. *Let $\boldsymbol{X}_\circ$ be a $d\times 1$ random vector for which (A4) holds. For $m,m'\in\mathbb{N}$ define $\boldsymbol{X}$, $\overset{\blacktriangle}{\boldsymbol{X}}_i$, $\boldsymbol{Q}_{mm'}$ and $\boldsymbol{Q}_{mm'}^{i\underset{\sim}{i}}$, $1\le i\le m$, $1\le \underset{\sim}{i}\le m'$, according to (S.1)–(S.5). Under (A5) we have the following results for fixed $m,m'\in\mathbb{N}$:*

(a) $\boldsymbol{X}^{\top}\boldsymbol{Q}_{mm'}^{-1}\boldsymbol{X}\xrightarrow{P}\left(\frac{1}{m}\boldsymbol{M}+\frac{1}{m'}\boldsymbol{M}'\right)^{-1}$.

(b) For all $1\le i\le m$, $\overset{\blacktriangle}{\boldsymbol{X}}{}_i^{\top}\boldsymbol{Q}_{mm'}^{ii}\overset{\blacktriangle}{\boldsymbol{X}}_i\xrightarrow{P}\boldsymbol{M}^{-1}-\frac{1}{mm'}\boldsymbol{M}^{-1}\boldsymbol{M}'\left(\frac{1}{m}\boldsymbol{M}+\frac{1}{m'}\boldsymbol{M}'\right)^{-1}$.

(c) If $m\ge 2$ *then for all* $1\le i,\underset{\sim}{i}\le m$ *such that* $i\ne\underset{\sim}{i}$,

$$\overset{\blacktriangle}{\boldsymbol{X}}{}_i^{\top}\boldsymbol{Q}_{mm'}^{i\underset{\sim}{i}}\overset{\blacktriangle}{\boldsymbol{X}}_{\underset{\sim}{i}}\xrightarrow{P}-\frac{1}{mm'}\boldsymbol{M}^{-1}\boldsymbol{M}'\left(\frac{1}{m}\boldsymbol{M}+\frac{1}{m'}\boldsymbol{M}'\right)^{-1}.$$

(d) $\left(\sum_{i=1}^{m}\sum_{i'=1}^{m'}n_{ii'}\right)^{-1}\operatorname{tr}(\boldsymbol{Q}_{mm'}^{-2})\xrightarrow{P}1/\lambda^2.$

Let $\overset{\star}{\boldsymbol{X}}_\circ$ be a $\overset{\star}{d}\times 1$ random vector and let

$$\overset{\star}{\boldsymbol{X}}_{ii'j},\quad 1\le i\le m,\quad 1\le i'\le m',\quad 1\le j\le n_{ii'}, \tag{S.6}$$

be independent and identically distributed random vectors having the same distribution as $\overset{\star}{\boldsymbol{X}}_\circ$. Then define for $1\le i\le m$ and $1\le i'\le m'$:

$$\overset{\star}{\boldsymbol{X}}_{ii'}\equiv\begin{bmatrix}\overset{\star}{\boldsymbol{X}}{}_{ii'1}^{\top}\\ \vdots\\ \overset{\star}{\boldsymbol{X}}{}_{ii'n_i}^{\top}\end{bmatrix}\quad\text{and}\quad \overset{\star}{\boldsymbol{X}}\equiv\operatorname*{stack}_{1\le i\le m}\left\{\operatorname*{stack}_{1\le i'\le m'}(\overset{\star}{\boldsymbol{X}}_{ii'})\right\}. \tag{S.7}$$

Lemma 6. *Let $\boldsymbol{X}_\circ$ be a $d\times 1$ random vector and $\overset{\star}{\boldsymbol{X}}_\circ$ be a $\overset{\star}{d}\times 1$ random vector such that (A4) holds for both $\boldsymbol{X}_\circ$ and $\overset{\star}{\boldsymbol{X}}_\circ$. Define $\boldsymbol{X}$ according to (S.1)–(S.2), $\boldsymbol{Q}_{mm'}$ according to (S.2)–(S.4) and $\overset{\star}{\boldsymbol{X}}$ according to (S.6)–(S.7). Under (A5) we have the following results for all fixed $m,m'\in\mathbb{N}$:*

(a) $\left(\sum_{i=1}^{m}\sum_{i'=1}^{m'} n_{ii'}\right)^{-1} \overset{\star}{\boldsymbol{X}}{}^{\top}\boldsymbol{Q}_{mm'}^{-1}\overset{\star}{\boldsymbol{X}} \xrightarrow{P} (1/\lambda)\Big[\textit{lower right } \overset{\star}{d} \times \overset{\star}{d} \textit{ block of } \big\{E\big([\boldsymbol{X}_\circ\ \overset{\star}{\boldsymbol{X}}_\circ{}^{\top}]^{\otimes 2}\big)\big\}^{-1}\Big]^{-1}.$

(b) $\boldsymbol{X}^{\top}\boldsymbol{Q}_{mm'}^{-1}\overset{\star}{\boldsymbol{X}} \xrightarrow{P} \left(\frac{1}{m}\boldsymbol{M} + \frac{1}{m'}\boldsymbol{M}'\right)^{-1}\{E(\boldsymbol{X}_\circ^{\otimes 2})\}^{-1}E(\boldsymbol{X}_\circ\overset{\star}{\boldsymbol{X}}{}_\circ^{\top}).$

S.1.3 Fisher Information Matrix Approximation

The Fisher information matrix of the full vector of unique parameters, corresponding to the conditional log-likelihood (4), is denoted by

$$I\big(\boldsymbol{\beta}_{\mathrm{A}}, \boldsymbol{\beta}_{\mathrm{B}}, \mathrm{vech}(\boldsymbol{\Sigma}), \mathrm{vech}(\boldsymbol{\Sigma}'), \sigma^2\big). \tag{S.8}$$

We now obtain approximations to each of the sub-blocks of (S.8).

From (A1), m' has the same order of magnitude as m. Therefore, remainder terms such as $o_P(mm'n)$ can be also written as $o_P(m^2n)$. Throughout this derivation we follow the convention of expressing all remainder terms that involve m and m' in terms of m only.

S.1.3.1 The $(\boldsymbol{\beta}_{\mathbf{A}}, \boldsymbol{\beta}_{\mathbf{A}})$ Diagonal Block

The $(\boldsymbol{\beta}_{\mathrm{A}}, \boldsymbol{\beta}_{\mathrm{A}})$ diagonal block is $\boldsymbol{X}_{\mathrm{A}}^{\top}\boldsymbol{V}^{-1}\boldsymbol{X}_{\mathrm{A}}$. From (A3) and Lemma 5(a), we have for all fixed $m, m' \in \mathbb{N}$ and as $n \to \infty$,

$$\boldsymbol{X}_{\mathrm{A}}^{\top}\boldsymbol{V}^{-1}\boldsymbol{X}_{\mathrm{A}} \xrightarrow{P} \left(\frac{\boldsymbol{\Sigma}}{m} + \frac{\boldsymbol{\Sigma}'}{m'}\right)^{-1}.$$

Therefore, under (A1) and (A3), the $(\boldsymbol{\beta}_{\mathrm{A}}, \boldsymbol{\beta}_{\mathrm{A}})$ diagonal block of the Fisher information matrix is

$$\left(\frac{\boldsymbol{\Sigma}}{m} + \frac{\boldsymbol{\Sigma}'}{m'}\right)^{-1} + o_P(m)\mathbf{1}_{d_{\mathrm{A}}}^{\otimes 2}.$$

S.1.3.2 The $(\boldsymbol{\beta}_{\mathbf{B}}, \boldsymbol{\beta}_{\mathbf{B}})$ Diagonal Block

The $(\boldsymbol{\beta}_{\mathrm{B}}, \boldsymbol{\beta}_{\mathrm{B}})$ diagonal block is $\boldsymbol{X}_{\mathrm{B}}^{\top}\boldsymbol{V}^{-1}\boldsymbol{X}_{\mathrm{B}}$. Under (A2)–(A3), and applying Lemma 6(a) with $\boldsymbol{X} = \boldsymbol{X}_{\mathrm{A}}$ and $\overset{\star}{\boldsymbol{X}} = \boldsymbol{X}_{\mathrm{B}}$ we have

$$\boldsymbol{X}_{\mathrm{B}}^{\top}\boldsymbol{V}^{-1}\boldsymbol{X}_{\mathrm{B}} = \frac{mm'n\boldsymbol{C}_{\boldsymbol{\beta}_{\mathrm{B}}}^{-1}}{\sigma^2} + o_P(mm'n)\mathbf{1}_{d_{\mathrm{B}}}^{\otimes 2}.$$

S.1.3.3 The $\big(\mathrm{vech}(\boldsymbol{\Sigma}), \mathrm{vech}(\boldsymbol{\Sigma})\big)$ Diagonal Block

From results given in e.g. Section 4.3 of Wand (2002), the $(\boldsymbol{\Sigma}_{rs}, \boldsymbol{\Sigma}_{tu})$ entry of the $\big(\mathrm{vech}(\boldsymbol{\Sigma}), \mathrm{vech}(\boldsymbol{\Sigma})\big)$ diagonal block of the Fisher information matrix is

$$\tfrac{1}{2}\mathrm{tr}\left(\boldsymbol{V}^{-1}\frac{\partial\boldsymbol{V}}{\partial(\boldsymbol{\Sigma})_{rs}}\boldsymbol{V}^{-1}\frac{\partial\boldsymbol{V}}{\partial(\boldsymbol{\Sigma})_{tu}}\right).$$

Then note that

$$\frac{\partial\boldsymbol{V}}{\partial(\boldsymbol{\Sigma})_{rs}} = \boldsymbol{L}_r\boldsymbol{L}_s^{\top} + I(r \neq s)\boldsymbol{L}_s\boldsymbol{L}_r^{\top} \quad \text{where} \quad \boldsymbol{L}_r \equiv \underset{1\le i\le m}{\mathrm{blockdiag}}\big(\overset{\blacktriangle}{\boldsymbol{X}}_{\mathrm{A}i}\boldsymbol{e}_r\big), \quad \overset{\blacktriangle}{\boldsymbol{X}}_{\mathrm{A}i} \equiv \underset{1\le i'\le m'}{\mathrm{stack}}(\boldsymbol{X}_{\mathrm{A}ii'}) \tag{S.9}$$

and $\boldsymbol{e}_r$ denotes the $d_{\mathrm{A}} \times 1$ matrix with the rth entry 1 and all other entries 0. Noting the $\mathrm{tr}(\boldsymbol{AB}) = \mathrm{tr}(\boldsymbol{BA})$ identity for all compatible matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ and introducing the notation

$$T_{rstu} \equiv \mathrm{tr}\big\{(\boldsymbol{L}_r^{\top}\boldsymbol{V}^{-1}\boldsymbol{L}_s)^{\top}(\boldsymbol{L}_t^{\top}\boldsymbol{V}^{-1}\boldsymbol{L}_u)\big\}.$$

we then have the following simplifications of the various sub-types of the $(\boldsymbol{\Sigma}_{rs}, \boldsymbol{\Sigma}_{tu})$ Fisher information blocks:

$$
\begin{array}{ll}
(\boldsymbol{\Sigma}_{rr}, \boldsymbol{\Sigma}_{tt}): & \frac{1}{2}T_{rtrt} \\
(\boldsymbol{\Sigma}_{rr}, \boldsymbol{\Sigma}_{tu}),\ t \neq u: & \frac{1}{2}\Big(T_{rurt} + T_{rtru}\Big) \\
(\boldsymbol{\Sigma}_{rs}, \boldsymbol{\Sigma}_{tt}),\ r \neq s: & \frac{1}{2}\Big(T_{rtst} + T_{strt}\Big) \\
(\boldsymbol{\Sigma}_{rs}, \boldsymbol{\Sigma}_{tu}),\ r \neq s, t \neq u: & \frac{1}{2}\Big(T_{rust} + T_{rtsu} + T_{surt} + T_{stru}\Big).
\end{array}
\tag{S.10}
$$

Since

$$
\boldsymbol{L}_r^\top \boldsymbol{V}^{-1}\boldsymbol{L}_s = \Big[\boldsymbol{e}_r^\top \overset{\blacktriangle}{\boldsymbol{X}}{}_{\mathrm{A}i}^\top \boldsymbol{V}^{i\underset{\sim}{i}} \overset{\blacktriangle}{\boldsymbol{X}}_{\mathrm{A}\underset{\sim}{i}} \boldsymbol{e}_s\Big]_{1\le i\le m, 1\le \underset{\sim}{i}\le m},
$$

we then have

$$
\begin{aligned}
T_{rstu} &= \sum_{i=1}^m \sum_{\underset{\sim}{i}=1}^m \Big(\boldsymbol{e}_r^\top \overset{\blacktriangle}{\boldsymbol{X}}{}_{\mathrm{A}i}^\top \boldsymbol{V}^{i\underset{\sim}{i}} \overset{\blacktriangle}{\boldsymbol{X}}_{\mathrm{A}\underset{\sim}{i}} \boldsymbol{e}_s\Big)\Big(\boldsymbol{e}_t^\top \overset{\blacktriangle}{\boldsymbol{X}}{}_{\mathrm{A}i}^\top \boldsymbol{V}^{i\underset{\sim}{i}} \overset{\blacktriangle}{\boldsymbol{X}}_{\mathrm{A}\underset{\sim}{i}} \boldsymbol{e}_u\Big) \\
&= \sum_{i=1}^m \Big(\boldsymbol{e}_r^\top \overset{\blacktriangle}{\boldsymbol{X}}{}_{\mathrm{A}i}^\top \boldsymbol{V}^{ii} \overset{\blacktriangle}{\boldsymbol{X}}_{\mathrm{A}i} \boldsymbol{e}_s\Big)\Big(\boldsymbol{e}_t^\top \overset{\blacktriangle}{\boldsymbol{X}}{}_{\mathrm{A}i}^\top \boldsymbol{V}^{ii} \overset{\blacktriangle}{\boldsymbol{X}}_{\mathrm{A}i} \boldsymbol{e}_u\Big) \\
&\qquad + \sum\sum_{i \neq \underset{\sim}{i}} \Big(\boldsymbol{e}_r^\top \overset{\blacktriangle}{\boldsymbol{X}}{}_{\mathrm{A}i}^\top \boldsymbol{V}^{i\underset{\sim}{i}} \overset{\blacktriangle}{\boldsymbol{X}}_{\mathrm{A}\underset{\sim}{i}} \boldsymbol{e}_s\Big)\Big(\boldsymbol{e}_t^\top \overset{\blacktriangle}{\boldsymbol{X}}{}_{\mathrm{A}i}^\top \boldsymbol{V}^{i\underset{\sim}{i}} \overset{\blacktriangle}{\boldsymbol{X}}_{\mathrm{A}\underset{\sim}{i}} \boldsymbol{e}_u\Big).
\end{aligned}
$$

Lemma 5 (b)–(c) implies that for any fixed $m \in \{2, 3, \ldots\}$ and $m' \in \mathbb{N}$ we have, as $n \to \infty$,

$$
\begin{aligned}
T_{rstu} \xrightarrow{P}\ & m\Big(\boldsymbol{\Sigma}^{-1} - \tfrac{1}{mm'}\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}'\left(\tfrac{1}{m}\boldsymbol{\Sigma} + \tfrac{1}{m'}\boldsymbol{\Sigma}'\right)^{-1}\Big)_{rs}\Big(\boldsymbol{\Sigma}^{-1} - \tfrac{1}{mm'}\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}'\left(\tfrac{1}{m}\boldsymbol{\Sigma} + \tfrac{1}{m'}\boldsymbol{\Sigma}'\right)^{-1}\Big)_{tu} \\
& + \frac{m(m-1)}{(mm')^2}\Big(\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}'\left(\tfrac{1}{m}\boldsymbol{\Sigma} + \tfrac{1}{m'}\boldsymbol{\Sigma}'\right)^{-1}\Big)_{rs}\Big(\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}'\left(\tfrac{1}{m}\boldsymbol{\Sigma} + \tfrac{1}{m'}\boldsymbol{\Sigma}'\right)^{-1}\Big)_{tu}.
\end{aligned}
$$

Now suppose that m and m' diverge according to (A1). Then straightforward steps show that

$$
T_{rstu} = m\big(\boldsymbol{\Sigma}^{-1}\big)_{rs}\big(\boldsymbol{\Sigma}^{-1}\big)_{tu} + O_P(1). \tag{S.11}
$$

In view of (S.10) and (S.11), under (A1) and (A2), the entries of the $\big(\mathrm{vech}(\boldsymbol{\Sigma}), \mathrm{vech}(\boldsymbol{\Sigma})\big)$ diagonal block have the following leading term behavior:

$$
\begin{array}{ll}
(\boldsymbol{\Sigma}_{rr}, \boldsymbol{\Sigma}_{tt}): & \frac{1}{2}m(\boldsymbol{\Sigma}^{-1})_{rt}^2 + O_P(1) \\
(\boldsymbol{\Sigma}_{rr}, \boldsymbol{\Sigma}_{tu}),\ t \neq u: & m(\boldsymbol{\Sigma}^{-1})_{rt}(\boldsymbol{\Sigma}^{-1})_{ru} + O_P(1) \\
(\boldsymbol{\Sigma}_{rs}, \boldsymbol{\Sigma}_{tt}),\ r \neq s: & m(\boldsymbol{\Sigma}^{-1})_{rt}(\boldsymbol{\Sigma}^{-1})_{st} + O_P(1) \\
(\boldsymbol{\Sigma}_{rs}, \boldsymbol{\Sigma}_{tu}),\ r \neq s, t \neq u: & m\big\{(\boldsymbol{\Sigma}^{-1})_{rt}(\boldsymbol{\Sigma}^{-1})_{su} + (\boldsymbol{\Sigma}^{-1})_{ru}(\boldsymbol{\Sigma}^{-1})_{st}\big\} + O_P(1).
\end{array}
$$

Application of Lemma 1 then leads to the following succinct expression for the $\big(\mathrm{vech}(\boldsymbol{\Sigma}), \mathrm{vech}(\boldsymbol{\Sigma})\big)$ Fisher information block:

$$
\tfrac{1}{2}m\boldsymbol{D}_{d_{\mathrm{A}}}^\top(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1})\boldsymbol{D}_{d_{\mathrm{A}}} + O_P(1)\boldsymbol{1}_{d_{\mathrm{A}}(d_{\mathrm{A}}+1)/2}^{\otimes 2}.
$$

S.1.3.4 The $\big(\mathrm{vech}(\boldsymbol{\Sigma}'), \mathrm{vech}(\boldsymbol{\Sigma}')\big)$ Diagonal Block

The conditional log-likelihood is unaffected by the interchanging of $\boldsymbol{\Sigma}$ and $\boldsymbol{\Sigma}'$. Hence, noting the conclusion of the previous subsection, the $\big(\mathrm{vech}(\boldsymbol{\Sigma}'), \mathrm{vech}(\boldsymbol{\Sigma}')\big)$ diagonal block of the Fisher information is

$$
\tfrac{1}{2}m'\boldsymbol{D}_{d_{\mathrm{A}}}^\top\big((\boldsymbol{\Sigma}')^{-1} \otimes (\boldsymbol{\Sigma}')^{-1}\big)\boldsymbol{D}_{d_{\mathrm{A}}} + O_P(1)\boldsymbol{1}_{d_{\mathrm{A}}(d_{\mathrm{A}}+1)/2}^{\otimes 2}.
$$

S.1.3.5 The (σ^2, σ^2) Diagonal Block

Appealing again to Section 4.3 of Wand (2002), the (σ^2, σ^2) diagonal block of the Fisher information matrix is

$$\tfrac{1}{2}\mathrm{tr}\left(\boldsymbol{V}^{-1}\frac{\partial \boldsymbol{V}}{\partial\sigma^2}\boldsymbol{V}^{-1}\frac{\partial \boldsymbol{V}}{\partial\sigma^2}\right) = \tfrac{1}{2}\mathrm{tr}(\boldsymbol{V}^{-2}) = \frac{mm'n}{2\sigma^4} + o_P(m^{-2}n^{-1}),$$

with the last equality following from Lemma 5(d).

S.1.3.6 The $(\boldsymbol{\beta}_{\mathbf{A}}, \boldsymbol{\beta}_{\mathbf{B}})$ Off-Diagonal Block

The $(\boldsymbol{\beta}_{\mathrm{A}}, \boldsymbol{\beta}_{\mathrm{B}})$ diagonal block is $\boldsymbol{X}_{\mathrm{A}}^{\top}\boldsymbol{V}^{-1}\boldsymbol{X}_{\mathrm{B}}$. From (A3) and Lemma 6(b), we have for all fixed $m, m' \in \mathbb{N}$ and as $n \to \infty$,

$$\boldsymbol{X}_{\mathrm{A}}^{\top}\boldsymbol{V}^{-1}\boldsymbol{X}_{\mathrm{B}} \xrightarrow{P} \left(\frac{\boldsymbol{\Sigma}}{m} + \frac{\boldsymbol{\Sigma}'}{m'}\right)^{-1}\{E(\boldsymbol{X}_{\mathrm{A}\circ}^{\otimes 2})\}^{-1}E(\boldsymbol{X}_{\mathrm{A}\circ}^{\top}\boldsymbol{X}_{\mathrm{B}\circ}).$$

Therefore, under (A1) and (A3), the $(\boldsymbol{\beta}_{\mathrm{A}}, \boldsymbol{\beta}_{\mathrm{B}})$ diagonal block of the Fisher information matrix is

$$\left(\frac{\boldsymbol{\Sigma}}{m} + \frac{\boldsymbol{\Sigma}'}{m'}\right)^{-1}\{E(\boldsymbol{X}_{\mathrm{A}\circ}^{\otimes 2})\}^{-1}E(\boldsymbol{X}_{\mathrm{A}\circ}^{\top}\boldsymbol{X}_{\mathrm{B}\circ}) + o_P(m).$$

S.1.3.7 The $\big((\boldsymbol{\beta}_{\mathbf{A}}, \boldsymbol{\beta}_{\mathbf{B}}), (\mathrm{vech}(\boldsymbol{\Sigma}), \mathrm{vech}(\boldsymbol{\Sigma}'), \sigma^2)\big)$ Off-Diagonal Block

From e.g. Section 4.3 of Wand (2002), the

$$\big((\boldsymbol{\beta}_{\mathrm{A}}, \boldsymbol{\beta}_{\mathrm{B}}), (\mathrm{vech}(\boldsymbol{\Sigma}), \mathrm{vech}(\boldsymbol{\Sigma}'), \sigma^2)\big)$$

off-diagonal block is a matrix having all entries equal to zero. In other words, the fixed effects parameters and the covariance matrix parameters are exactly orthogonal in Gaussian response linear mixed models.

S.1.3.8 The $\big(\mathrm{vech}(\boldsymbol{\Sigma}), \mathrm{vech}(\boldsymbol{\Sigma}')\big)$ Off-Diagonal Block

We commence with the special case of $d_{\mathrm{A}} = 1$, $n_{ii'} = n$ and $\boldsymbol{X}_{\mathrm{A}ii'} = \mathbf{1}_n$ for all $1 \le i \le m$, $1 \le i' \le m'$. In this case $\mathbf{1}_{mm'n}$ is an eigenvector of $\boldsymbol{V}$ with corresponding eigenvalue $m'n\Sigma + mn\Sigma' + \sigma^2$. This implies that $\mathbf{1}_{mm'n}$ is also an eigenvector of $\boldsymbol{V}^{-1}$ with the just-mentioned eigenvalue reciprocated. Relatively straightforward manipulations then lead to the following expression for the (Σ, Σ') entry of the Fisher information matrix:

$$\tfrac{1}{2}\big[\{\Sigma(m'/m) + \Sigma' + \sigma^2/(mn)\}\{\Sigma + \Sigma'(m/m') + \sigma^2/(m'n)\}\big]^{-1} \tag{S.12}$$

which is $O(1)$ under (A1).

Next we treat the general d_{A}, $n_{ii'}$ and $\boldsymbol{X}_{\mathrm{A}ii'}$ situation with $m \in \mathbb{N}$ and $m' = 1$. From e.g. Section 4.3 of Wand (2002), the $(\boldsymbol{\Sigma}_{rr}, \boldsymbol{\Sigma}'_{tt})$ entry of the $\big(\mathrm{vech}(\boldsymbol{\Sigma}), \mathrm{vech}(\boldsymbol{\Sigma}')\big)$ off-diagonal block of the Fisher information matrix is

$$\tfrac{1}{2}\mathrm{tr}\left(\boldsymbol{V}^{-1}\frac{\partial \boldsymbol{V}}{\partial(\boldsymbol{\Sigma})_{rr}}\boldsymbol{V}^{-1}\frac{\partial \boldsymbol{V}}{\partial(\boldsymbol{\Sigma}')_{tt}}\right) \tag{S.13}$$

where, noting the current $m' = 1$ special case,

$$\frac{\partial \boldsymbol{V}}{\partial(\boldsymbol{\Sigma})_{rr}} = \underset{1\le i\le m}{\mathrm{blockdiag}}\big(\boldsymbol{X}_{\mathrm{A}i1}\boldsymbol{e}_r\boldsymbol{e}_r^{\top}\boldsymbol{X}_{\mathrm{A}i1}^{\top}\big) \text{ and } \frac{\partial \boldsymbol{V}}{\partial(\boldsymbol{\Sigma}')_{tt}} = \underset{1\le i,\utilde{i}\le m}{\mathrm{blockmatrix}}\big(\boldsymbol{X}_{\mathrm{A}i1}\boldsymbol{e}_t\boldsymbol{e}_t^{\top}\boldsymbol{X}_{\mathrm{A}\utilde{i}1}^{\top}\big). \tag{S.14}$$

Substitution of (S.14) into (S.13) and algebraic manipulations such as those involving the $\mathrm{tr}(\boldsymbol{AB}) = \mathrm{tr}(\boldsymbol{BA})$ identity lead to

$$
\begin{aligned}
\mathrm{tr}\left(\boldsymbol{V}^{-1}\frac{\partial \boldsymbol{V}}{\partial(\boldsymbol{\Sigma})_{rr}}\boldsymbol{V}^{-1}\frac{\partial \boldsymbol{V}}{\partial(\boldsymbol{\Sigma}')_{tt}}\right) &= \sum_{i=1}^{m}\sum_{\underset{\sim}{i}=1}^{m}\sum_{\underset{\sim}{i}^*=1}^{m}\big(\boldsymbol{e}_t^\top \boldsymbol{X}_{\mathrm{A}i1}^\top \boldsymbol{V}^{i\underset{\sim}{i}}\boldsymbol{X}_{\mathrm{A}\underset{\sim}{i}1}\boldsymbol{e}_r\big)\big(\boldsymbol{e}_r^\top \boldsymbol{X}_{\mathrm{A}\underset{\sim}{i}1}^\top \boldsymbol{V}^{\underset{\sim}{i}\underset{\sim}{i}^*}\boldsymbol{X}_{\mathrm{A}\underset{\sim}{i}^*1}\boldsymbol{e}_t\big)\\
&= \sum_{i=1}^{m}\big(\boldsymbol{e}_r^\top \boldsymbol{X}_{\mathrm{A}i1}^\top \boldsymbol{V}^{ii}\boldsymbol{X}_{\mathrm{A}i1}\boldsymbol{e}_t\big)^2\\
&\quad + \sum\sum_{i\neq \underset{\sim}{i}^*}\big(\boldsymbol{e}_t^\top \boldsymbol{X}_{\mathrm{A}i1}^\top \boldsymbol{V}^{ii}\boldsymbol{X}_{\mathrm{A}i1}\boldsymbol{e}_r\big)\big(\boldsymbol{e}_r^\top \boldsymbol{X}_{\mathrm{A}i1}^\top \boldsymbol{V}^{i\underset{\sim}{i}^*}\boldsymbol{X}_{\mathrm{A}\underset{\sim}{i}^*1}\boldsymbol{e}_t\big)\\
&\quad + \sum\sum_{i\neq \underset{\sim}{i}}\big(\boldsymbol{e}_t^\top \boldsymbol{X}_{\mathrm{A}i1}^\top \boldsymbol{V}^{i\underset{\sim}{i}}\boldsymbol{X}_{\mathrm{A}\underset{\sim}{i}1}\boldsymbol{e}_r\big)\big(\boldsymbol{e}_r^\top \boldsymbol{X}_{\mathrm{A}\underset{\sim}{i}1}^\top \boldsymbol{V}^{\underset{\sim}{i}\underset{\sim}{i}}\boldsymbol{X}_{\mathrm{A}\underset{\sim}{i}1}\boldsymbol{e}_t\big)\\
&\quad + \sum\sum\sum_{i\neq \underset{\sim}{i}\neq \underset{\sim}{i}^*}\big(\boldsymbol{e}_t^\top \boldsymbol{X}_{\mathrm{A}i1}^\top \boldsymbol{V}^{i\underset{\sim}{i}}\boldsymbol{X}_{\mathrm{A}\underset{\sim}{i}1}\boldsymbol{e}_r\big)\big(\boldsymbol{e}_r^\top \boldsymbol{X}_{\mathrm{A}\underset{\sim}{i}1}^\top \boldsymbol{V}^{\underset{\sim}{i}\underset{\sim}{i}^*}\boldsymbol{X}_{\mathrm{A}\underset{\sim}{i}^*1}\boldsymbol{e}_t\big).
\end{aligned}
$$

Lemma 5(b) and 5(c) then imply that

$$
\begin{aligned}
\tfrac{1}{2}\mathrm{tr}\left(\boldsymbol{V}^{-1}\frac{\partial \boldsymbol{V}}{\partial(\boldsymbol{\Sigma})_{rr}}\boldsymbol{V}^{-1}\frac{\partial \boldsymbol{V}}{\partial(\boldsymbol{\Sigma}')_{tt}}\right) \xrightarrow{P}\; & \tfrac{1}{2}m\Big(\boldsymbol{M}^{-1}-\tfrac{1}{m}\boldsymbol{M}^{-1}\boldsymbol{M}'\left(\tfrac{1}{m}\boldsymbol{M}+\boldsymbol{M}'\right)\Big)_{rt}^2\\
&+\tfrac{1}{2}m(m-1)\Big(\boldsymbol{M}^{-1}-\tfrac{1}{m}\boldsymbol{M}^{-1}\boldsymbol{M}'\left(\tfrac{1}{m}\boldsymbol{M}+\boldsymbol{M}'\right)\Big)_{tr}\\
&\times\Big(-\tfrac{1}{m}\boldsymbol{M}^{-1}\boldsymbol{M}'\left(\tfrac{1}{m}\boldsymbol{M}+\boldsymbol{M}'\right)\Big)_{rt}\\
&+\tfrac{1}{2}m(m-1)\Big(-\tfrac{1}{m}\boldsymbol{M}^{-1}\boldsymbol{M}'\left(\tfrac{1}{m}\boldsymbol{M}+\boldsymbol{M}'\right)\Big)_{tr}\\
&\times\Big(\boldsymbol{M}^{-1}-\tfrac{1}{m}\boldsymbol{M}^{-1}\boldsymbol{M}'\left(\tfrac{1}{m}\boldsymbol{M}+\boldsymbol{M}'\right)\Big)_{rt}\\
&+\tfrac{1}{2}m(m-1)^2\Big(-\tfrac{1}{m}\boldsymbol{M}^{-1}\boldsymbol{M}'\left(\tfrac{1}{m}\boldsymbol{M}+\boldsymbol{M}'\right)\Big)_{tr}\\
&\times\Big(-\tfrac{1}{m}\boldsymbol{M}^{-1}\boldsymbol{M}'\left(\tfrac{1}{m}\boldsymbol{M}+\boldsymbol{M}'\right)\Big)_{rt}\\
=\;& \tfrac{1}{2m}\Big(\left(\tfrac{1}{m}\boldsymbol{M}+\boldsymbol{M}'\right)^{-1}\Big)_{rt}^2
\end{aligned}
$$

after several algebraic steps and cancellations. The $r \neq s$ and $t \neq u$ cases are similar. This confirms that (S.12) also holds in general, with the exception of m' being set to 1. For $m' \geq 2$ similar arguments can be used to show that the summations in (S.13) lead to convergents analogous to those in the $d_{\mathrm{A}} = 1$, $n_{ii'} = n$ and $\boldsymbol{X}_{\mathrm{A}ii'} = \boldsymbol{1}_n$ case and a matrix with order $O(1)\boldsymbol{1}_{d_{\mathrm{A}}(d_{\mathrm{A}}+1)/2}^{\otimes 2}$ under (A1) eventuates.

S.1.3.9 The $\big(\mathrm{vech}(\boldsymbol{\Sigma}), \sigma^2\big)$ Off-Diagonal Block

We commence with the special case of $d_{\mathrm{A}} = 1$, $n_{ii'} = n$ and $\boldsymbol{X}_{\mathrm{A}ii'} = \boldsymbol{1}_n$ for all $1 \leq i \leq m$, $1 \leq i' \leq m'$. Using the eigenvalue and eigenvector properties described near the beginning of Section S.1.3.8, relatively straightforward manipulations then lead to the following expression for the (Σ, σ^2) entry of the Fisher information matrix:

$$
\begin{aligned}
&\frac{m'(1-1/m)\{\Sigma+\Sigma'(m/m')+\sigma^2/(m'n)\}^2}{2mn\{\Sigma(m'/m)+\Sigma'+\sigma^2/(mn)\}^2\{\Sigma+\sigma^2/(m'n)\}^2}\\
&+\frac{m'}{2m^2n\{\Sigma(m'/m)+\Sigma'+\sigma^2/(mn)\}^2}
\end{aligned}
\qquad \text{(S.15)}
$$

which is $O(n^{-1})$ under (A1).

Now consider the general d_{A}, $n_{ii'}$ and $\boldsymbol{X}_{\mathrm{A}ii'}$ situation with $m \in \mathbb{N}$ and $m' = 1$. Results in e.g. Section 4.3 of Wand (2002) imply that the $(\boldsymbol{\Sigma}_{rr}, \sigma^2)$ entry of the $\big(\mathrm{vech}(\boldsymbol{\Sigma}), \sigma^2\big)$ off-diagonal block of the Fisher information matrix is

$$\tfrac{1}{2}\mathrm{tr}\left(\boldsymbol{V}^{-2} \underset{1\le i\le m}{\mathrm{blockdiag}}\left(\boldsymbol{X}_{\mathrm{A}i1}\boldsymbol{e}_r\boldsymbol{e}_r^\top\boldsymbol{X}_{\mathrm{A}i1}^\top\right)\right) = \tfrac{1}{2}\sum_{i=1}^m\sum_{\underset{\sim}{i}=1}^m \boldsymbol{e}_r^\top\big(\boldsymbol{V}^{\underset{\sim}{i}i}\boldsymbol{X}_{\mathrm{A}i1}\big)^\top\big(\boldsymbol{V}^{\underset{\sim}{i}i}\boldsymbol{X}_{\mathrm{A}i1}\big)\boldsymbol{e}_r. \qquad \text{(S.16)}$$

For the $m = m' = 1$ special case of (S.16), Lemma 3 leads to

$$\begin{aligned}
n_{11}\boldsymbol{X}_{\mathrm{A}11}^\top\boldsymbol{V}_{11}^{-2}\boldsymbol{X}_{\mathrm{A}11} &= n_{11}\boldsymbol{e}_r^\top\boldsymbol{X}_{\mathrm{A}11}^\top\{\boldsymbol{X}_{\mathrm{A}11}(\boldsymbol{\Sigma}+\boldsymbol{\Sigma}')\boldsymbol{X}_{\mathrm{A}11}^\top + \sigma^2\boldsymbol{I}\}^{-2}\boldsymbol{X}_{\mathrm{A}11}\boldsymbol{e}_r \\
&= \boldsymbol{e}_r^\top\big\{\boldsymbol{\Sigma}+\boldsymbol{\Sigma}'+\sigma^2(\boldsymbol{X}_{\mathrm{A}11}^\top\boldsymbol{X}_{\mathrm{A}11})^{-1}\big\}^{-1}\left(\tfrac{1}{n_{11}}\boldsymbol{X}_{\mathrm{A}11}^\top\boldsymbol{X}_{\mathrm{A}11}\right)^{-1} \\
&\quad \times\big\{\boldsymbol{\Sigma}+\boldsymbol{\Sigma}'+\sigma^2(\boldsymbol{X}_{\mathrm{A}11}^\top\boldsymbol{X}_{\mathrm{A}11})^{-1}\big\}^{-1}\boldsymbol{e}_r \\
&\xrightarrow{P} \boldsymbol{e}_r^\top(\boldsymbol{\Sigma}+\boldsymbol{\Sigma}')^{-1}E(\boldsymbol{X}_\circ^\top\boldsymbol{X}_\circ)(\boldsymbol{\Sigma}+\boldsymbol{\Sigma}')^{-1}\boldsymbol{e}_r^\top.
\end{aligned}$$

Hence, the $(\boldsymbol{\Sigma}_{rr}, \sigma^2)$ entry of the Fisher information is

$$\frac{\big((\boldsymbol{\Sigma}+\boldsymbol{\Sigma}')^{-1}E(\boldsymbol{X}_\circ^\top\boldsymbol{X}_\circ)(\boldsymbol{\Sigma}+\boldsymbol{\Sigma}')^{-1}\big)_{rr}\{1+o_P(1)\}}{2n_{11}}$$

which extends (S.15) for $d_{\mathrm{A}} \in \mathbb{N}$ and general predictors for $m = m' = 1$. Treatment of the $(\boldsymbol{\Sigma}_{rs}, \sigma^2)$ entries for $r \neq s$ is similar and also leads to $O_P(n^{-1})$ leading term behavior under (A2).

For the $(m, m') = (2, 1)$ case, with assistance from Lemmas 2 and 3, $2n_{11}$ multiplied by the $(i, \underset{\sim}{i}) = (1, 1)$ term on the right-hand side of (S.16) equals

$$\begin{aligned}
&n_{11}\boldsymbol{e}_r^\top\boldsymbol{X}_{\mathrm{A}11}^\top(\boldsymbol{V}^{11})^2\boldsymbol{X}_{\mathrm{A}11}\boldsymbol{e}_r \\
&\quad= n_{11}\boldsymbol{e}_r^\top\boldsymbol{X}_{\mathrm{A}11}^\top\Bigg(\text{upper left } n_{11}\times n_{11} \text{ block of} \\
&\qquad\begin{bmatrix}\boldsymbol{X}_{\mathrm{A}11}(\boldsymbol{\Sigma}+\boldsymbol{\Sigma}')\boldsymbol{X}_{\mathrm{A}11}^\top+\sigma^2\boldsymbol{I} & \boldsymbol{X}_{\mathrm{A}11}\boldsymbol{\Sigma}'\boldsymbol{X}_{\mathrm{A}21}^\top \\ \boldsymbol{X}_{\mathrm{A}21}\boldsymbol{\Sigma}'\boldsymbol{X}_{\mathrm{A}11}^\top & \boldsymbol{X}_{\mathrm{A}21}(\boldsymbol{\Sigma}+\boldsymbol{\Sigma}')\boldsymbol{X}_{\mathrm{A}21}^\top+\sigma^2\boldsymbol{I}\end{bmatrix}^{-1}\Bigg)^2\boldsymbol{X}_{\mathrm{A}11}\boldsymbol{e}_r \\
&\quad= n_{11}\boldsymbol{e}_r^\top\boldsymbol{X}_{\mathrm{A}11}^\top\Big[\boldsymbol{X}_{\mathrm{A}11}(\boldsymbol{\Sigma}+\boldsymbol{\Sigma}')\boldsymbol{X}_{\mathrm{A}11}^\top+\sigma^2\boldsymbol{I} \\
&\qquad-\boldsymbol{X}_{\mathrm{A}11}\boldsymbol{\Sigma}'\boldsymbol{X}_{\mathrm{A}21}^\top\big\{\boldsymbol{X}_{\mathrm{A}21}(\boldsymbol{\Sigma}+\boldsymbol{\Sigma}')\boldsymbol{X}_{\mathrm{A}21}^\top+\sigma^2\boldsymbol{I}\big\}^{-1}\boldsymbol{X}_{\mathrm{A}21}\boldsymbol{\Sigma}'\boldsymbol{X}_{\mathrm{A}11}^\top\Big]^{-2}\boldsymbol{X}_{\mathrm{A}11}\boldsymbol{e}_r \\
&\quad= n_{11}\boldsymbol{e}_r^\top\boldsymbol{X}_{\mathrm{A}11}^\top\Big[\boldsymbol{X}_{\mathrm{A}11}(\boldsymbol{\Sigma}+\boldsymbol{\Sigma}')\boldsymbol{X}_{\mathrm{A}11}^\top+\sigma^2\boldsymbol{I} \\
&\qquad-\boldsymbol{X}_{\mathrm{A}11}\boldsymbol{\Sigma}'\big\{\boldsymbol{\Sigma}+\boldsymbol{\Sigma}'+\sigma^2(\boldsymbol{X}_{\mathrm{A}21}^\top\boldsymbol{X}_{\mathrm{A}21})^{-1}\boldsymbol{I}\big\}^{-1}\boldsymbol{\Sigma}'\boldsymbol{X}_{\mathrm{A}11}^\top\Big]^{-2}\boldsymbol{X}_{\mathrm{A}11}\boldsymbol{e}_r \\
&\quad= \boldsymbol{e}_r^\top\Big[\boldsymbol{\Sigma}+\boldsymbol{\Sigma}'-\boldsymbol{\Sigma}'\big\{\boldsymbol{\Sigma}+\boldsymbol{\Sigma}'+\sigma^2(\boldsymbol{X}_{\mathrm{A}21}^\top\boldsymbol{X}_{\mathrm{A}21})^{-1}\big\}^{-1}\boldsymbol{\Sigma}'+\sigma^2(\boldsymbol{X}_{\mathrm{A}11}^\top\boldsymbol{X}_{\mathrm{A}11})^{-1}\Big]^{-1} \\
&\qquad\times\left(\tfrac{1}{n_{11}}\boldsymbol{X}_{\mathrm{A}11}^\top\boldsymbol{X}_{\mathrm{A}11}\right)^{-1} \\
&\qquad\times\Big[\boldsymbol{\Sigma}+\boldsymbol{\Sigma}'-\boldsymbol{\Sigma}'\big\{\boldsymbol{\Sigma}+\boldsymbol{\Sigma}'+\sigma^2(\boldsymbol{X}_{\mathrm{A}21}^\top\boldsymbol{X}_{\mathrm{A}21})^{-1}\big\}^{-1}\boldsymbol{\Sigma}'+\sigma^2(\boldsymbol{X}_{\mathrm{A}11}^\top\boldsymbol{X}_{\mathrm{A}11})^{-1}\Big]^{-1}\boldsymbol{e}_r \\
&\quad\xrightarrow{P} \boldsymbol{e}_r^\top\Big\{\boldsymbol{\Sigma}+\boldsymbol{\Sigma}'-\boldsymbol{\Sigma}'\big(\boldsymbol{\Sigma}+\boldsymbol{\Sigma}'\big)^{-1}\boldsymbol{\Sigma}'\Big\}^{-1}E(\boldsymbol{X}_\circ^\top\boldsymbol{X}_\circ)\Big\{\boldsymbol{\Sigma}+\boldsymbol{\Sigma}'-\boldsymbol{\Sigma}'\big(\boldsymbol{\Sigma}+\boldsymbol{\Sigma}'\big)^{-1}\boldsymbol{\Sigma}'\Big\}^{-1}\boldsymbol{e}_r.
\end{aligned}$$

Similar arguments lead to

$$n_{21}\boldsymbol{e}_r^\top\boldsymbol{X}_{\mathrm{A}21}^\top(\boldsymbol{V}^{22})^2\boldsymbol{X}_{\mathrm{A}21}\boldsymbol{e}_r$$

having the same convergence in probability limit. In addition, and again using Lemmas 2 and 3,

$$
\begin{aligned}
&n_{21}\boldsymbol{e}_r^\top \boldsymbol{X}_{\mathrm{A}11}^\top (\boldsymbol{V}^{21})^\top \boldsymbol{V}^{21}\boldsymbol{X}_{\mathrm{A}11}\boldsymbol{e}_r \\
&\quad = n_{21}\boldsymbol{e}_r^\top \boldsymbol{X}_{\mathrm{A}11}^\top \Bigg(\text{the transposed lower left } n_{12}\times n_{11} \text{ block of} \\
&\qquad \begin{bmatrix} \boldsymbol{X}_{\mathrm{A}11}(\boldsymbol{\Sigma}+\boldsymbol{\Sigma}')\boldsymbol{X}_{\mathrm{A}11}^\top + \sigma^2\boldsymbol{I} & \boldsymbol{X}_{\mathrm{A}11}\boldsymbol{\Sigma}'\boldsymbol{X}_{\mathrm{A}21}^\top \\ \boldsymbol{X}_{\mathrm{A}21}\boldsymbol{\Sigma}'\boldsymbol{X}_{\mathrm{A}11}^\top & \boldsymbol{X}_{\mathrm{A}21}(\boldsymbol{\Sigma}+\boldsymbol{\Sigma}')\boldsymbol{X}_{\mathrm{A}21}^\top + \sigma^2\boldsymbol{I} \end{bmatrix}^{-1} \Bigg)^{\otimes 2} \boldsymbol{X}_{\mathrm{A}21}\boldsymbol{e}_r \\
&\quad = n_{21}\boldsymbol{e}_r^\top \boldsymbol{X}_{\mathrm{A}11}^\top \{\boldsymbol{X}_{\mathrm{A}11}(\boldsymbol{\Sigma}+\boldsymbol{\Sigma}')\boldsymbol{X}_{\mathrm{A}11}^\top + \sigma^2\boldsymbol{I}\}^{-1}\boldsymbol{X}_{\mathrm{A}11}\boldsymbol{\Sigma}'\boldsymbol{X}_{\mathrm{A}21}^\top \\
&\qquad \times\Big[\boldsymbol{X}_{\mathrm{A}21}(\boldsymbol{\Sigma}+\boldsymbol{\Sigma}')\boldsymbol{X}_{\mathrm{A}21}^\top + \sigma^2\boldsymbol{I} \\
&\qquad -\boldsymbol{X}_{\mathrm{A}21}\boldsymbol{\Sigma}'\boldsymbol{X}_{\mathrm{A}11}^\top \{\boldsymbol{X}_{\mathrm{A}11}(\boldsymbol{\Sigma}+\boldsymbol{\Sigma}')\boldsymbol{X}_{\mathrm{A}11}^\top + \sigma^2\boldsymbol{I}\}^{-1}\boldsymbol{X}_{\mathrm{A}11}\boldsymbol{\Sigma}'\boldsymbol{X}_{\mathrm{A}21}^\top\Big]^{-2} \\
&\qquad \times\boldsymbol{X}_{\mathrm{A}21}\boldsymbol{\Sigma}'\boldsymbol{X}_{\mathrm{A}11}^\top \{\boldsymbol{X}_{\mathrm{A}11}(\boldsymbol{\Sigma}+\boldsymbol{\Sigma}')\boldsymbol{X}_{\mathrm{A}11}^\top + \sigma^2\boldsymbol{I}\}^{-1}\boldsymbol{X}_{\mathrm{A}11}\boldsymbol{e}_r \\
&\quad = \boldsymbol{e}_r^\top\{\boldsymbol{\Sigma}+\boldsymbol{\Sigma}'+\sigma^2(\boldsymbol{X}_{\mathrm{A}11}^\top\boldsymbol{X}_{\mathrm{A}11})^{-1}\}^{-1}\boldsymbol{\Sigma}' \\
&\qquad \times\Big[\boldsymbol{\Sigma}+\boldsymbol{\Sigma}'+\sigma^2(\boldsymbol{X}_{\mathrm{A}21}^\top\boldsymbol{X}_{\mathrm{A}21})^{-1} - \boldsymbol{\Sigma}'\{\boldsymbol{\Sigma}+\boldsymbol{\Sigma}'+\sigma^2(\boldsymbol{X}_{\mathrm{A}11}^\top\boldsymbol{X}_{\mathrm{A}11})^{-1}\}^{-1}\boldsymbol{\Sigma}'\Big]^{-1} \\
&\qquad \times\big(\tfrac{1}{n_{21}}\boldsymbol{X}_{\mathrm{A}21}^\top\boldsymbol{X}_{\mathrm{A}21}\big)^{-1} \\
&\qquad \times\Big[\boldsymbol{\Sigma}+\boldsymbol{\Sigma}'+\sigma^2(\boldsymbol{X}_{\mathrm{A}21}^\top\boldsymbol{X}_{\mathrm{A}21})^{-1} - \boldsymbol{\Sigma}'\{\boldsymbol{\Sigma}+\boldsymbol{\Sigma}'+\sigma^2(\boldsymbol{X}_{\mathrm{A}11}^\top\boldsymbol{X}_{\mathrm{A}11})^{-1}\}^{-1}\boldsymbol{\Sigma}'\Big]^{-1} \\
&\qquad \times\boldsymbol{\Sigma}'\{\boldsymbol{\Sigma}+\boldsymbol{\Sigma}'+\sigma^2(\boldsymbol{X}_{\mathrm{A}11}^\top\boldsymbol{X}_{\mathrm{A}11})^{-1}\}^{-1}\boldsymbol{e}_r \\
&\quad \xrightarrow{P} \boldsymbol{e}_r^\top(\boldsymbol{\Sigma}+\boldsymbol{\Sigma}')^{-1}\boldsymbol{\Sigma}'\Big\{\boldsymbol{\Sigma}+\boldsymbol{\Sigma}'-\boldsymbol{\Sigma}'(\boldsymbol{\Sigma}+\boldsymbol{\Sigma}')^{-1}\boldsymbol{\Sigma}'\Big\}^{-1}E(\boldsymbol{X}_\circ^\top\boldsymbol{X}_\circ) \\
&\qquad \times\Big\{\boldsymbol{\Sigma}+\boldsymbol{\Sigma}'-\boldsymbol{\Sigma}'(\boldsymbol{\Sigma}+\boldsymbol{\Sigma}')^{-1}\boldsymbol{\Sigma}'\Big\}^{-1}\boldsymbol{\Sigma}'(\boldsymbol{\Sigma}+\boldsymbol{\Sigma}')^{-1}\boldsymbol{e}_r.
\end{aligned}
$$

Similar steps lead to $n_{11}\boldsymbol{e}_r^\top\boldsymbol{X}_{\mathrm{A}21}^\top(\boldsymbol{V}^{12})^\top\boldsymbol{V}^{12}\boldsymbol{X}_{\mathrm{A}21}\boldsymbol{e}_r$ having the same convergence in probability limit. On combining these results we obtain the $(\boldsymbol{\Sigma}_{rr},\sigma^2)$ entry of the Fisher information for $(m,m')=(2,1)$ having leading term behavior:

$$
\begin{aligned}
&\tfrac{1}{2}\left(\tfrac{1}{n_{11}}+\tfrac{1}{n_{21}}\right)\Bigg(\Big\{\boldsymbol{\Sigma}+\boldsymbol{\Sigma}'-\boldsymbol{\Sigma}'(\boldsymbol{\Sigma}+\boldsymbol{\Sigma}')^{-1}\boldsymbol{\Sigma}'\Big\}^{-1} \\
&\times E(\boldsymbol{X}_\circ^\top\boldsymbol{X}_\circ)\Big\{\boldsymbol{\Sigma}+\boldsymbol{\Sigma}'-\boldsymbol{\Sigma}'(\boldsymbol{\Sigma}+\boldsymbol{\Sigma}')^{-1}\boldsymbol{\Sigma}'\Big\}^{-1}\Bigg)_{rr}\{1+o_P(1)\} \\
&+\tfrac{1}{2}\left(\tfrac{1}{n_{11}}+\tfrac{1}{n_{21}}\right)\Bigg((\boldsymbol{\Sigma}+\boldsymbol{\Sigma}')^{-1}\boldsymbol{\Sigma}'\Big\{\boldsymbol{\Sigma}+\boldsymbol{\Sigma}'-\boldsymbol{\Sigma}'(\boldsymbol{\Sigma}+\boldsymbol{\Sigma}')^{-1}\boldsymbol{\Sigma}'\Big\}^{-1}E(\boldsymbol{X}_\circ^\top\boldsymbol{X}_\circ) \\
&\times\Big\{\boldsymbol{\Sigma}+\boldsymbol{\Sigma}'-\boldsymbol{\Sigma}'(\boldsymbol{\Sigma}+\boldsymbol{\Sigma}')^{-1}\boldsymbol{\Sigma}'\Big\}^{-1}\boldsymbol{\Sigma}'(\boldsymbol{\Sigma}+\boldsymbol{\Sigma}')^{-1}\Bigg)_{rr}\{1+o_P(1)\}
\end{aligned}
$$

which, under (A2), has $O_P(n^{-1})$ leading term behavior. Similar arguments lead to the $O_P(n^{-1})$ property holding for the $(\boldsymbol{\Sigma}_{rs},\sigma^2)$ entries of the Fisher information matrix for $r\neq s$ when $(m,m')=(2,1)$.

For higher m and m', similar arguments can be used to show that the summations in $(\mathrm{vech}(\boldsymbol{\Sigma}),\sigma^2)$ Fisher information block lead to convergents that are analogous to those in the $d_{\mathrm{A}}=1$, $n_{ii'}=n$ and $\boldsymbol{X}_{\mathrm{A}ii'}=\boldsymbol{1}_n$ case and the block satisfies $O_P(n^{-1})\boldsymbol{1}_{d_{\mathrm{A}}(d_{\mathrm{A}}+1)/2}$ under (A1) and (A2).

This very low order of magnitude of the $(\mathrm{vech}(\boldsymbol{\Sigma}),\sigma^2)$ off-diagonal block of the Fisher information matrix is more than enough for asymptotic orthogonality between $\boldsymbol{\Sigma}$ and σ^2. A larger order of magnitude, such as $O_P(1)\boldsymbol{1}_{d_{\mathrm{A}}(d_{\mathrm{A}}+1)/2}$, would still be sufficient.

S.1.3.10 The $\big(\text{vech}(\mathbf{\Sigma}'),\sigma^2\big)$ Off-Diagonal Block

In the special case of $d_{\text{A}}=1$, $n_{ii'}=n$ and $\boldsymbol{X}_{\text{A}ii'}=\mathbf{1}_n$ for all $1\le i\le m$, $1\le i'\le m'$ use of the eigenvalue and eigenvector properties described near the commencement of Section S.1.3.8 lead to the (Σ',σ^2) entry of the Fisher information matrix having exact expression

$$\frac{m(1-1/m')\{\Sigma(m'/m)+\Sigma'+\sigma^2/(mn)\}^2}{2m'n\{\Sigma+\Sigma'(m/m')+\sigma^2/(m'n)\}^2\{\Sigma'+\sigma^2/(mn)\}^2}\\+\frac{m}{2(m')^2n\{\Sigma+\Sigma'(m/m')+\sigma^2/(m'n)\}^2}$$

which has the same form as (S.15) but with the roles of (M,m) and (M',m') reversed. Symmetry considerations dictate that the same happens in the general setting and the $\big(\text{vech}(\mathbf{\Sigma}'),\sigma^2\big)$ off-diagonal block is $O(n^{-1})\mathbf{1}_{d_{\text{A}}/(d_{\text{A}}+1)/2}$.

S.1.3.11 Assembly of the Fisher Information Sub-Block Approximations

The Fisher information sub-block approximations obtained in the previous nine sub-subsections lead to

$$I\big(\boldsymbol{\beta}_{\text{A}},\boldsymbol{\beta}_{\text{B}},\text{vech}(\mathbf{\Sigma}),\text{vech}(\mathbf{\Sigma}'),\sigma^2\big)=$$

$$\begin{bmatrix}
\begin{matrix}\left(\dfrac{\mathbf{\Sigma}}{m}+\dfrac{\mathbf{\Sigma}'}{m'}\right)^{-1}\\+o_P(m)\mathbf{1}_{d_{\text{A}}}^{\otimes 2}\end{matrix} & O_P(m)\mathbf{1}_{d_{\text{A}}}\mathbf{1}_{d_{\text{B}}}^{\top} & \boldsymbol{O} & \boldsymbol{O} & \boldsymbol{O}\\
O_P(m)\mathbf{1}_{d_{\text{B}}}\mathbf{1}_{d_{\text{A}}}^{\top} & \begin{matrix}\dfrac{mm'n\boldsymbol{C}_{\boldsymbol{\beta}_{\text{B}}}^{-1}}{\sigma^2}\\+o_P(m^2n)\mathbf{1}_{d_{\text{B}}}^{\otimes 2}\end{matrix} & \boldsymbol{O} & \boldsymbol{O} & \boldsymbol{O}\\
\boldsymbol{O} & \boldsymbol{O} & \begin{matrix}\dfrac{m\boldsymbol{D}_{d_{\text{A}}}^{\top}(\mathbf{\Sigma}^{-1}\otimes\mathbf{\Sigma}^{-1})\boldsymbol{D}_{d_{\text{A}}}}{2}\\+o_P(m)\mathbf{1}_{d_{\text{A}}^{\boxplus}}^{\otimes 2}\end{matrix} & O_P(1)\mathbf{1}_{d_{\text{A}}^{\boxplus}}^{\otimes 2} & O_P(n^{-1})\mathbf{1}_{d_{\text{A}}^{\boxplus}}\\
\boldsymbol{O} & \boldsymbol{O} & O_P(1)\mathbf{1}_{d_{\text{A}}^{\boxplus}}^{\otimes 2} & \begin{matrix}\dfrac{m'\boldsymbol{D}_{d_{\text{A}}}^{\top}\big((\mathbf{\Sigma}')^{-1}\otimes(\mathbf{\Sigma}')^{-1}\big)\boldsymbol{D}_{d_{\text{A}}}}{2}\\+o_P(m)\mathbf{1}_{d_{\text{A}}^{\boxplus}}^{\otimes 2}\end{matrix} & O_P(n^{-1})\mathbf{1}_{d_{\text{A}}^{\boxplus}}\\
\boldsymbol{O} & \boldsymbol{O} & O_P(n^{-1})\mathbf{1}_{d_{\text{A}}^{\boxplus}}^{\top} & O_P(n^{-1})\mathbf{1}_{d_{\text{A}}^{\boxplus}}^{\top} & \begin{matrix}\dfrac{mm'n}{2\sigma^4}\\+o_P(m^2n)\end{matrix}
\end{bmatrix}$$

where $d_{\text{A}}^{\boxplus}\equiv\frac{1}{2}d_{\text{A}}(d_{\text{A}}+1)$.

S.1.4 Inverse Fisher Information Matrix Approximation

First note that, since $I\big(\boldsymbol{\beta}_{\text{A}},\boldsymbol{\beta}_{\text{B}},\text{vech}(\mathbf{\Sigma}),\text{vech}(\mathbf{\Sigma}'),\sigma^2\big)$ is block diagonal, its inversion involves the individual inversions of the $\big(\boldsymbol{\beta}_{\text{A}},\boldsymbol{\beta}_{\text{B}}\big)$ and $\big(\text{vech}(\mathbf{\Sigma}),\text{vech}(\mathbf{\Sigma}'),\sigma^2\big)$ blocks. These two inversions involve application of well-known block matrix inversion formulae and keeping track of the various terms that arise and their orders of magnitude. For example, if the sub-blocks of the $\big(\boldsymbol{\beta}_{\text{A}},\boldsymbol{\beta}_{\text{B}}\big)$ block are denoted as follows:

$$\begin{bmatrix}\boldsymbol{A}_{11} & \boldsymbol{A}_{12}\\ \boldsymbol{A}_{12}^{\top} & \boldsymbol{A}_{22}\end{bmatrix}\quad\text{where}\quad \boldsymbol{A}_{11}\text{ is }d_{\text{A}}\times d_{\text{A}},$$

then the upper left $d_{\text{A}}\times d_{\text{A}}$ block of the required inverse matrix is

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

Appendix A.6 of Jiang *et al.* (2022) contains a detailed account of this approach for related setting. Analogous steps for the current setting lead to

$$I\big(\boldsymbol{\beta}_{\mathrm{A}},\boldsymbol{\beta}_{\mathrm{B}},\mathrm{vech}(\boldsymbol{\Sigma}),\mathrm{vech}(\boldsymbol{\Sigma}'),\sigma^2\big)^{-1}=I\big(\boldsymbol{\beta}_{\mathrm{A}},\boldsymbol{\beta}_{\mathrm{B}},\mathrm{vech}(\boldsymbol{\Sigma}),\mathrm{vech}(\boldsymbol{\Sigma}'),\sigma^2\big)_{\infty}^{-1}$$

$$+\frac{1}{m}\begin{bmatrix} o_P(1)\boldsymbol{1}_{d_{\mathrm{A}}}^{\otimes 2} & O_P(m^{-1}n^{-1})\boldsymbol{1}_{d_{\mathrm{A}}}\boldsymbol{1}_{d_{\mathrm{B}}}^{\top} & \boldsymbol{O} & \boldsymbol{O} & \boldsymbol{O}\\ O_P(m^{-1}n^{-1})\boldsymbol{1}_{d_{\mathrm{B}}}\boldsymbol{1}_{d_{\mathrm{A}}}^{\top} & o_P(m^{-1}n^{-1})\boldsymbol{1}_{d_{\mathrm{B}}}^{\otimes 2} & \boldsymbol{O} & \boldsymbol{O} & \boldsymbol{O}\\ \boldsymbol{O} & \boldsymbol{O} & o_P(1)\boldsymbol{1}_{d_{\mathrm{A}}^{\boxplus}}^{\otimes 2} & O_P(m^{-1})\boldsymbol{1}_{d_{\mathrm{A}}^{\boxplus}}\boldsymbol{1}_{d_{\mathrm{A}}^{\boxplus}}^{\otimes 2} & O_P(m^{-2}n^{-1})\boldsymbol{1}_{d_{\mathrm{A}}^{\boxplus}}\\ \boldsymbol{O} & \boldsymbol{O} & O_P(m^{-1})\boldsymbol{1}_{d_{\mathrm{A}}^{\boxplus}}\boldsymbol{1}_{d_{\mathrm{A}}^{\boxplus}}^{\otimes 2} & o_P(1)\boldsymbol{1}_{d_{\mathrm{A}}^{\boxplus}}^{\otimes 2} & O_P(m^{-2}n^{-1})\boldsymbol{1}_{d_{\mathrm{A}}^{\boxplus}}\\ \boldsymbol{O} & \boldsymbol{O} & O_P(m^{-2}n^{-1})\boldsymbol{1}_{d_{\mathrm{A}}^{\boxplus}}^{\top} & O_P(m^{-2}n^{-1})\boldsymbol{1}_{d_{\mathrm{A}}^{\boxplus}}^{\top} & o_P(m^{-2}n^{-1}) \end{bmatrix}$$

where

$$\begin{aligned}&I\big(\boldsymbol{\beta}_{\mathrm{A}},\boldsymbol{\beta}_{\mathrm{B}},\mathrm{vech}(\boldsymbol{\Sigma}),\mathrm{vech}(\boldsymbol{\Sigma}'),\sigma^2\big)_{\infty}^{-1}\\ &=\begin{bmatrix}\dfrac{\boldsymbol{\Sigma}}{m}+\dfrac{\boldsymbol{\Sigma}'}{m'} & \boldsymbol{O} & \boldsymbol{O} & \boldsymbol{O} & \boldsymbol{O}\\ \boldsymbol{O} & \dfrac{\sigma^2\boldsymbol{C}_{\boldsymbol{\beta}_{\mathrm{B}}}}{mm'n} & \boldsymbol{O} & \boldsymbol{O} & \boldsymbol{O}\\ \boldsymbol{O} & \boldsymbol{O} & \dfrac{2\boldsymbol{D}_{d_{\mathrm{A}}}^{+}(\boldsymbol{\Sigma}\otimes\boldsymbol{\Sigma})\boldsymbol{D}_{d_{\mathrm{A}}}^{+\top}}{m} & \boldsymbol{O} & \boldsymbol{O}\\ \boldsymbol{O} & \boldsymbol{O} & \boldsymbol{O} & \dfrac{2\boldsymbol{D}_{d_{\mathrm{A}}}^{+}(\boldsymbol{\Sigma}'\otimes\boldsymbol{\Sigma}')\boldsymbol{D}_{d_{\mathrm{A}}}^{+\top}}{m'} & \boldsymbol{O}\\ \boldsymbol{O} & \boldsymbol{O} & \boldsymbol{O} & \boldsymbol{O} & \dfrac{2\sigma^4}{mm'n}\end{bmatrix}.\end{aligned}$$

S.1.5 Asymptotic Normality of the Maximum Likelihood Estimators

Let

$$\boldsymbol{\beta}\equiv(\boldsymbol{\beta}_{\mathrm{A}},\boldsymbol{\beta}_{\mathrm{B}})\quad\text{and}\quad\boldsymbol{\psi}\equiv\big(\mathrm{vech}(\boldsymbol{\Sigma}),\mathrm{vech}(\boldsymbol{\Sigma}'),\sigma^2\big).$$

As alluded to in Section S.1.3.7, the Fisher information has the block diagonal form:

$$I(\boldsymbol{\beta},\boldsymbol{\psi})=\begin{bmatrix} I(\boldsymbol{\beta},\boldsymbol{\psi})_{\boldsymbol{\beta\beta}} & \boldsymbol{O}\\ \boldsymbol{O} & I(\boldsymbol{\beta},\boldsymbol{\psi})_{\boldsymbol{\psi\psi}}\end{bmatrix}. \tag{S.17}$$

where $I(\boldsymbol{\beta},\boldsymbol{\psi})_{\boldsymbol{\beta\beta}}$ is the upper left $(d_{\mathrm{A}}+d_{\mathrm{B}})\times(d_{\mathrm{A}}+d_{\mathrm{B}})$ block of $I(\boldsymbol{\beta},\boldsymbol{\psi})$ and $I(\boldsymbol{\beta},\boldsymbol{\psi})_{\boldsymbol{\psi\psi}}$ is defined similarly. Then, under (A1)–(A3) and some additional regularity conditions

$$\{I(\boldsymbol{\beta}^0,\boldsymbol{\psi}^0)^{-1}\}^{-1/2}\begin{bmatrix}\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}^0\\ \widehat{\boldsymbol{\psi}}-\boldsymbol{\psi}^0\end{bmatrix}\xrightarrow{D}N(\boldsymbol{0},\boldsymbol{I}). \tag{S.18}$$

Justification for (S.18) is given in Section S.1.8.

S.1.6 Convergence Results for Matrix Square Root Discrepancies

We now deal with the problem of proving that matrix square roots of the exact inverse Fisher information matrix and its convergent

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

are also sufficiently close to each other as m, m' and n diverge. Using the notation from (S.17), we treat the fixed effects and covariance parameter diagonal blocks separately. To this end, define

$$I(\boldsymbol{\beta},\boldsymbol{\psi})^{-1}_{\boldsymbol{\beta\beta},\infty} \equiv \begin{bmatrix} \dfrac{\boldsymbol{\Sigma}}{m}+\dfrac{\boldsymbol{\Sigma}'}{m'} & \boldsymbol{O} \\ \boldsymbol{O} & \dfrac{\sigma^2\boldsymbol{C}_{\boldsymbol{\beta}_{\mathrm{B}}}}{mm'n} \end{bmatrix}$$

and

$$I(\boldsymbol{\beta},\boldsymbol{\psi})^{-1}_{\boldsymbol{\psi\psi},\infty} \equiv \begin{bmatrix} \dfrac{2\boldsymbol{D}^{+}_{d_{\mathrm{A}}}(\boldsymbol{\Sigma}\otimes\boldsymbol{\Sigma})\boldsymbol{D}^{+\top}_{d_{\mathrm{A}}}}{m} & \boldsymbol{O} & \boldsymbol{O} \\ \boldsymbol{O} & \dfrac{2\boldsymbol{D}^{+}_{d_{\mathrm{A}}}(\boldsymbol{\Sigma}'\otimes\boldsymbol{\Sigma}')\boldsymbol{D}^{+\top}_{d_{\mathrm{A}}}}{m'} & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{O} & \dfrac{2\sigma^4}{mm'n} \end{bmatrix}.$$

Next note that

$$m'I(\boldsymbol{\beta},\boldsymbol{\psi})^{-1}_{\boldsymbol{\beta\beta}} = \begin{bmatrix} \boldsymbol{K}+o_P(\mathbf{1}^{\otimes 2}_{d_{\mathrm{A}}}) & O_P\big((mn)^{-1}\big)\mathbf{1}_{d_{\mathrm{A}}}\mathbf{1}^{\top}_{d_{\mathrm{B}}} \\ O_P\big((mn)^{-1}\big)\mathbf{1}_{d_{\mathrm{B}}}\mathbf{1}^{\top}_{d_{\mathrm{A}}} & \frac{1}{m}\boldsymbol{L}+o_P\big((mn)^{-1}\big)\mathbf{1}^{\otimes 2}_{d_{\mathrm{B}}} \end{bmatrix} \text{ and } m'I(\boldsymbol{\beta},\boldsymbol{\psi})^{-1}_{\boldsymbol{\beta\beta},\infty} = \begin{bmatrix} \boldsymbol{K} & \boldsymbol{O} \\ \boldsymbol{O} & \frac{1}{m}\boldsymbol{L} \end{bmatrix}$$

where

$$\boldsymbol{K} \equiv (m'/m)\boldsymbol{\Sigma}+\boldsymbol{\Sigma}' \quad \text{and} \quad \boldsymbol{L} \equiv \frac{\sigma^2\boldsymbol{C}_{\boldsymbol{\beta}_{\mathrm{B}}}}{n}.$$

Then application of Lemma 2 of Jiang *et al.* (2022) as $m\to\infty$ implies that

$$\left\|\{I\big(\boldsymbol{\beta},\boldsymbol{\psi}\big)^{-1}_{\boldsymbol{\beta\beta},\infty}\}^{-1/2}\{I\big(\boldsymbol{\beta},\boldsymbol{\psi}\big)^{-1}_{\boldsymbol{\beta\beta}}\}^{1/2}-\boldsymbol{I}\right\|_F \xrightarrow{P} 0. \tag{S.19}$$

The establishment of

$$\left\|\{I\big(\boldsymbol{\beta},\boldsymbol{\psi}\big)^{-1}_{\boldsymbol{\psi\psi},\infty}\}^{-1/2}\{I\big(\boldsymbol{\beta},\boldsymbol{\psi}\big)^{-1}_{\boldsymbol{\psi\psi}}\}^{1/2}-\boldsymbol{I}\right\|_F \xrightarrow{P} 0 \tag{S.20}$$

is very similar.

S.1.7 Final Steps for the Derivation of Result 1

Let

$$\boldsymbol{\theta} \equiv (\boldsymbol{\beta},\boldsymbol{\psi}) = \big(\boldsymbol{\beta}_{\mathrm{A}},\boldsymbol{\beta}_{\mathrm{B}},\mathrm{vech}(\boldsymbol{\Sigma}),\mathrm{vech}(\boldsymbol{\Sigma}'),\sigma^2\big)$$

be the full parameter vector. In terms of this new notation, (S.18) is

$$\big\{I\big(\boldsymbol{\theta}^0\big)^{-1}\big\}^{-1/2}(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}^0) \xrightarrow{D} N(\mathbf{0},\boldsymbol{I}) \tag{S.21}$$

where

$$\widehat{\boldsymbol{\theta}} = \big[(\widehat{\boldsymbol{\beta}}_{\mathrm{A}})^{\top}\ (\widehat{\boldsymbol{\beta}}_{\mathrm{B}})^{\top}\ \mathrm{vech}(\widehat{\boldsymbol{\Sigma}})^{\top}\ \mathrm{vech}(\widehat{\boldsymbol{\Sigma}}')^{\top}\ \widehat{\sigma}^2\big]^{\top}$$

and

$$\boldsymbol{\theta}^0 = \big[(\boldsymbol{\beta}^0_{\mathrm{A}})^{\top}\ (\boldsymbol{\beta}^0_{\mathrm{B}})^{\top}\ \mathrm{vech}(\boldsymbol{\Sigma}^0)^{\top}\ \mathrm{vech}(\boldsymbol{\Sigma}^0)^{\top}\ \mathrm{vech}((\boldsymbol{\Sigma}')^0)^{\top}\ (\sigma^2)^0\big]^{\top}.$$

It follows from (S.21) that, for all $(d_{\mathrm{A}}+d_{\mathrm{B}}+2d^{\boxplus}_{\mathrm{A}}+1)\times 1$ vectors $\boldsymbol{a}\neq\mathbf{0}$, we have

$$\boldsymbol{a}^{\top}\{I\big(\boldsymbol{\theta}^0\big)^{-1}\}^{-1/2}(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}^0) \xrightarrow{D} N(0,\boldsymbol{a}^{\top}\boldsymbol{a}).$$

As a consequence

$$\boldsymbol{a}^{\top}\{I\big(\boldsymbol{\theta}^0\big)^{-1}_{\infty}\}^{-1/2}(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}^0)+r_{mm'n}(\boldsymbol{a}) \xrightarrow{D} N(0,\boldsymbol{a}^{\top}\boldsymbol{a}) \tag{S.22}$$

where

$$r_{mm'n}(\boldsymbol{a}) \equiv \boldsymbol{a}^\top[\{I(\boldsymbol{\theta}^0)^{-1}\}^{-1/2} - \{I(\boldsymbol{\theta}^0)^{-1}_{\infty}\}^{-1/2}](\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^0)$$
$$= \boldsymbol{a}^\top[\boldsymbol{I} - \{I(\boldsymbol{\theta}^0)^{-1}_{\infty}\}^{-1/2}\{I(\boldsymbol{\theta}^0)^{-1}\}^{1/2}]\{I(\boldsymbol{\theta}^0)^{-1}\}^{1/2}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^0)$$
$$= \Big(\big[\{I(\boldsymbol{\theta}^0)^{-1}_{\infty}\}^{-1/2}\{I(\boldsymbol{\theta}^0)^{-1}\}^{1/2} - \boldsymbol{I}\big]^\top \boldsymbol{a}\Big)^\top \boldsymbol{Z}_{mm'n}$$

and $\boldsymbol{Z}_{mm'n} \xrightarrow{D} N(\boldsymbol{0}, \boldsymbol{I}_{d_{\mathrm{A}}+d_{\mathrm{B}}+2d^{\boxplus}_{\mathrm{A}}+1})$. Then note that

$$\Big\|\big[\{I(\boldsymbol{\theta}^0)^{-1}_{\infty}\}^{-1/2}\{I(\boldsymbol{\theta}^0)^{-1}\}^{1/2} - \boldsymbol{I}\big]^\top \boldsymbol{a}\Big\|_F \le \big\|\{I(\boldsymbol{\theta}^0)^{-1}_{\infty}\}^{-1/2}\{I(\boldsymbol{\theta}^0)^{-1}\}^{1/2} - \boldsymbol{I}\big\|_F \, \|\boldsymbol{a}\|_F.$$

As a consequence of (S.19) and (S.20) we have

$$\big\|\{I(\boldsymbol{\theta}^0)^{-1}_{\infty}\}^{-1/2}\{I(\boldsymbol{\theta}^0)^{-1}\}^{1/2} - \boldsymbol{I}\big\|_F \xrightarrow{P} 0 \tag{S.23}$$

and so

$$\big[\{I(\boldsymbol{\theta}^0)^{-1}_{\infty}\}^{-1/2}\{I(\boldsymbol{\theta}^0)^{-1}\}^{1/2} - \boldsymbol{I}\big]\boldsymbol{a} \xrightarrow{P} 0.$$

Application of Slutsky's Theorem then gives $r_{mm'n}(\boldsymbol{a}) \xrightarrow{P} 0$. From (S.22) and another application of Slutsky's Theorem we have

$$\boldsymbol{a}^\top \{I(\boldsymbol{\theta}^0)^{-1}_{\infty}\}^{-1/2}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^0) \xrightarrow{D} N(0, \boldsymbol{a}^\top \boldsymbol{a}).$$

Result 1 then follows from the Cramér-Wold Device.

S.1.8 Justification of (S.18)

We now provide justification for the asymptotic normality statement (S.18) concerning the maximum likelihood estimators and the Fisher information matrix.

As in Section S.1.7 let

$$\boldsymbol{\theta} \equiv (\boldsymbol{\beta}, \boldsymbol{\psi}) = \big(\boldsymbol{\beta}_{\mathrm{A}}, \boldsymbol{\beta}_{\mathrm{B}}, \mathrm{vech}(\boldsymbol{\Sigma}), \mathrm{vech}(\boldsymbol{\Sigma}'), \sigma^2\big)$$

be the full parameter vector. The score vector is

$$\nabla_{\boldsymbol{\theta}}\ell(\boldsymbol{\theta}) = \begin{bmatrix} \boldsymbol{X}_{\mathrm{A}}^\top \boldsymbol{V}^{-1}(\boldsymbol{Y} - \boldsymbol{X}_{\mathrm{A}}\boldsymbol{\beta}_{\mathrm{A}} - \boldsymbol{X}_{\mathrm{B}}\boldsymbol{\beta}_{\mathrm{B}}) \\ \boldsymbol{X}_{\mathrm{B}}^\top \boldsymbol{V}^{-1}(\boldsymbol{Y} - \boldsymbol{X}_{\mathrm{A}}\boldsymbol{\beta}_{\mathrm{A}} - \boldsymbol{X}_{\mathrm{B}}\boldsymbol{\beta}_{\mathrm{B}}) \\ \frac{1}{2}\underset{(r,s)\in\mathcal{I}_{d_{\mathrm{A}}}}{\mathrm{stack}} \Big\{\mathrm{tr}\Big(\boldsymbol{V}^{-1}\breve{\boldsymbol{L}}_{(r,s)}\boldsymbol{V}^{-1}(\boldsymbol{Y} - \boldsymbol{X}_{\mathrm{A}}\boldsymbol{\beta}_{\mathrm{A}} - \boldsymbol{X}_{\mathrm{B}}\boldsymbol{\beta}_{\mathrm{B}})^{\otimes 2} - \boldsymbol{V}^{-1}\breve{\boldsymbol{L}}_{(r,s)}\Big)\Big\} \\ \frac{1}{2}\underset{(r,s)\in\mathcal{I}_{d_{\mathrm{A}}}}{\mathrm{stack}} \Big\{\mathrm{tr}\Big(\boldsymbol{V}^{-1}\breve{\boldsymbol{L}}'_{(r,s)}\boldsymbol{V}^{-1}(\boldsymbol{Y} - \boldsymbol{X}_{\mathrm{A}}\boldsymbol{\beta}_{\mathrm{A}} - \boldsymbol{X}_{\mathrm{B}}\boldsymbol{\beta}_{\mathrm{B}})^{\otimes 2} - \boldsymbol{V}^{-1}\breve{\boldsymbol{L}}'_{(r,s)}\Big)\Big\} \\ \frac{1}{2}\mathrm{tr}\Big(\boldsymbol{V}^{-2}(\boldsymbol{Y} - \boldsymbol{X}_{\mathrm{A}}\boldsymbol{\beta}_{\mathrm{A}} - \boldsymbol{X}_{\mathrm{B}}\boldsymbol{\beta}_{\mathrm{B}})^{\otimes 2} - \boldsymbol{V}^{-1}\Big) \end{bmatrix}$$

where

$$\mathcal{I}_{d_{\mathrm{A}}} \equiv \{(1,1),(2,1),\ldots,(d_{\mathrm{A}},1),(2,2),(3,2),\ldots,(d_{\mathrm{A}},2),\ldots,(d_{\mathrm{A}},d_{\mathrm{A}})\}$$

corresponds to positions on and below the diagonal of a $d_{\mathrm{A}} \times d_{\mathrm{A}}$ matrix with the vech operator ordering,

$$\breve{\boldsymbol{L}}_{(r,s)} \equiv \boldsymbol{L}_r\boldsymbol{L}_s^\top + I(r \ne s)\boldsymbol{L}_s\boldsymbol{L}_r^\top$$

with $\boldsymbol{L}_r$ as defined by (S.9), and

$$\breve{\boldsymbol{L}}'_{(r,s)} \equiv \underset{1\le i,\underset{\sim}{i}\le m}{\mathrm{blockmatrix}}\left\{\underset{1\le i'\le m'}{\mathrm{blockdiag}}\Big(\boldsymbol{X}_{\mathrm{A}ii'}\Big(\boldsymbol{e}_r\boldsymbol{e}_s^\top + I(r\ne s)\boldsymbol{e}_s\boldsymbol{e}_r^\top\Big)\boldsymbol{X}_{\mathrm{A}\underset{\sim}{i}i'}^\top\Big)\right\}.$$

Let

$$\boldsymbol{Z} \equiv \left[\underset{1\le i\le m}{\text{blockdiag}} \left\{ \underset{1\le i'\le m'}{\text{stack}} (\boldsymbol{X}_{\text{A}ii'}) \right\} \quad \underset{1\le i\le m}{\text{stack}} \left\{ \underset{1\le i'\le m'}{\text{blockdiag}} (\boldsymbol{X}_{\text{A}ii'}) \right\} \right],$$

$$\boldsymbol{U}_{\text{all}} \equiv \begin{bmatrix} \underset{1\le i\le m}{\text{stack}} (\boldsymbol{U}_i) \\ \underset{1\le i'\le m'}{\text{stack}} (\boldsymbol{U}'_{i'}) \end{bmatrix} \quad \text{and} \quad \boldsymbol{G} \equiv \begin{bmatrix} \boldsymbol{I}_m \otimes \boldsymbol{\Sigma} & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{I}_{m'} \otimes \boldsymbol{\Sigma}' \end{bmatrix}.$$

Next, define

$$\boldsymbol{z} \equiv \begin{bmatrix} \boldsymbol{G} & \boldsymbol{O} \\ \boldsymbol{O} & \sigma^2 \boldsymbol{I} \end{bmatrix}^{-1/2} \begin{bmatrix} \boldsymbol{U}_{\text{all}} \\ \boldsymbol{Y} - \boldsymbol{X}_{\text{A}}\boldsymbol{\beta}_{\text{A}} - \boldsymbol{X}_{\text{B}}\boldsymbol{\beta}_{\text{B}} - \boldsymbol{Z}\boldsymbol{U}_{\text{all}} \end{bmatrix} \text{ and } \boldsymbol{V}^{1/2}_{\text{loose}} \equiv [\boldsymbol{Z}\ \boldsymbol{I}] \begin{bmatrix} \boldsymbol{G} & \boldsymbol{O} \\ \boldsymbol{O} & \sigma^2 \boldsymbol{I} \end{bmatrix}^{1/2}.$$

The relationship

$$\boldsymbol{V}^{1/2}_{\text{loose}} \big(\boldsymbol{V}^{1/2}_{\text{loose}}\big)^{\top} = \boldsymbol{V}$$

is the reason for the $\boldsymbol{V}^{1/2}_{\text{loose}}$ notation since, loosely (i.e. ignoring transposes), it is a matrix square root of $\boldsymbol{V}$. Noting that

$$\boldsymbol{Y} - \boldsymbol{X}_{\text{A}}\boldsymbol{\beta}_{\text{A}} - \boldsymbol{X}_{\text{B}}\boldsymbol{\beta}_{\text{B}} = \boldsymbol{V}^{1/2}_{\text{loose}}\boldsymbol{z}$$

we can re-write the score vector as

$$\nabla_{\boldsymbol{\theta}} \ell(\boldsymbol{\theta}) = \begin{bmatrix} \underset{1\le r\le d_{\text{A}}}{\text{stack}} (\boldsymbol{w}_{\text{A}r}^{\top}\boldsymbol{z}) \\ \underset{1\le r\le d_{\text{B}}}{\text{stack}} (\boldsymbol{w}_{\text{B}r}^{\top}\boldsymbol{z}) \\ \frac{1}{2} \underset{(r,s)\in\mathcal{I}_{d_{\text{A}}}}{\text{stack}} \left\{ \text{tr}\big(\boldsymbol{W}_{(r,s)}(\boldsymbol{z}^{\otimes 2} - \boldsymbol{I})\big) \right\} \\ \frac{1}{2} \underset{(r,s)\in\mathcal{I}_{d_{\text{A}}}}{\text{stack}} \left\{ \text{tr}\big(\boldsymbol{W}'_{(r,s)}(\boldsymbol{z}^{\otimes 2} - \boldsymbol{I})\big) \right\} \\ \frac{1}{2}\text{tr}\big(\boldsymbol{W}_{\sigma^2}(\boldsymbol{z}^{\otimes 2} - \boldsymbol{I})\big) \end{bmatrix}$$

where

$$\begin{aligned} \boldsymbol{w}_{\text{A}r} &\equiv r\text{th column of } (\boldsymbol{V}^{1/2}_{\text{loose}})^{\top}\boldsymbol{V}^{-1}\boldsymbol{X}_{\text{A}}, \quad 1 \le r \le d_{\text{A}}, \\ \boldsymbol{w}_{\text{B}r} &\equiv r\text{th column of } (\boldsymbol{V}^{1/2}_{\text{loose}})^{\top}\boldsymbol{V}^{-1}\boldsymbol{X}_{\text{B}}, \quad 1 \le r \le d_{\text{B}}, \\ \boldsymbol{W}_{(r,s)} &= (\boldsymbol{V}^{1/2}_{\text{loose}})^{\top}\boldsymbol{V}^{-1}\,\breve{\boldsymbol{L}}_{(r,s)}\boldsymbol{V}^{-1}\boldsymbol{V}^{1/2}_{\text{loose}}, \quad (r,s) \in \mathcal{I}_{d_{\text{A}}}, \\ \boldsymbol{W}'_{(r,s)} &= (\boldsymbol{V}^{1/2}_{\text{loose}})^{\top}\boldsymbol{V}^{-1}\,\breve{\boldsymbol{L}}'_{(r,s)}\boldsymbol{V}^{-1}\boldsymbol{V}^{1/2}_{\text{loose}}, \quad (r,s) \in \mathcal{I}_{d_{\text{A}}} \\ \text{and} \quad \boldsymbol{W}_{\sigma^2} &= (\boldsymbol{V}^{1/2}_{\text{loose}})^{\top}\boldsymbol{V}^{-2}\boldsymbol{V}^{1/2}_{\text{loose}}. \end{aligned}$$

Let

$$\boldsymbol{s}(m,n) \equiv \left[m\boldsymbol{1}_{d_{\text{A}}} \quad m^2 n \boldsymbol{1}_{d_{\text{B}}} \quad m\boldsymbol{1}_{\frac{1}{2}d_{\text{A}}(d_{\text{A}}+1)} \quad m\boldsymbol{1}_{\frac{1}{2}d_{\text{B}}(d_{\text{B}}+1)} \quad m^2 n \right]^{\top}$$

be a vector of sample size quantities that accounts for the $m = O(m')$ and $m' = O(m)$ assumptions. Then define

$$\boldsymbol{a}_{\text{norm}} \equiv \text{diag}\{\boldsymbol{s}(m,n)\}^{1/2} I\big(\boldsymbol{\theta}^0\big)^{-1/2}\boldsymbol{a}.$$

Letting $\boldsymbol{n}$ denote the matrix of $n_{ii'}$ values, note that

$$\boldsymbol{a}^{\top} I\big(\boldsymbol{\theta}^0\big)^{-1/2} \nabla_{\boldsymbol{\theta}} \ell(\boldsymbol{\theta}^0) = \boldsymbol{a}_{\text{norm}}^{\top} \text{diag}\{\boldsymbol{s}(m,n)\}^{-1/2} \nabla_{\boldsymbol{\theta}} \ell(\boldsymbol{\theta}^0) = \sum_{t=1}^{N_{\text{mart}}} \xi_t(m, m', \boldsymbol{n})$$

where, for $1 \le t \le N_{\text{mart}}$,

$$
\begin{aligned}
\xi_t(m,m',\boldsymbol{n}) \equiv\ & (\boldsymbol{a}_{\text{norm}})_1 m^{-1/2}(\boldsymbol{w}^0_{\text{A}1})_t(\boldsymbol{z})_t + \ldots + (\boldsymbol{a}_{\text{norm}})_{d_{\text{A}}} m^{-1/2}(\boldsymbol{w}^0_{\text{A}d_{\text{A}}})_t(\boldsymbol{z})_t \\
& +(\boldsymbol{a}_{\text{norm}})_{d_{\text{A}}+1}(m^2n)^{-1/2}(\boldsymbol{w}^0_{\text{B}1})_t(\boldsymbol{z})_t + \ldots + (\boldsymbol{a}_{\text{norm}})_{d_{\text{A}}+d_{\text{B}}}(m^2n)^{-1/2}(\boldsymbol{w}^0_{\text{B}d_{\text{B}}})_t(\boldsymbol{z})_t \\
& +\tfrac{1}{2}(\boldsymbol{a}_{\text{norm}})_{d_{\text{A}}+d_{\text{B}}+1} m^{-1/2}\big(\boldsymbol{W}^0_{(1,1)}(\boldsymbol{z}^{\otimes 2}-\boldsymbol{I})\big)_{tt} \\
& +\ldots+\tfrac{1}{2}(\boldsymbol{a}_{\text{norm}})_{d_{\text{A}}+d_{\text{B}}+\frac{1}{2}d_{\text{A}}(d_{\text{A}}+1)} m^{-1/2}\big(\boldsymbol{W}^0_{(d_{\text{A}},d_{\text{A}})}(\boldsymbol{z}^{\otimes 2}-\boldsymbol{I})\big)_{tt} \\
& +\tfrac{1}{2}(\boldsymbol{a}_{\text{norm}})_{d_{\text{A}}+d_{\text{B}}+\frac{1}{2}d_{\text{A}}(d_{\text{A}}+1)+1} m^{-1/2}\big((\boldsymbol{W}')^0_{(1,1)}(\boldsymbol{z}^{\otimes 2}-\boldsymbol{I})\big)_{tt} \\
& +\ldots+\tfrac{1}{2}(\boldsymbol{a}_{\text{norm}})_{d_{\text{A}}+d_{\text{B}}+\frac{1}{2}d_{\text{A}}(d_{\text{A}}+1)+\frac{1}{2}d_{\text{B}}(d_{\text{B}}+1)} m^{-1/2}\big((\boldsymbol{W}')^0_{(d_{\text{B}},d_{\text{B}})}(\boldsymbol{z}^{\otimes 2}-\boldsymbol{I})\big)_{tt} \\
& +\tfrac{1}{2}(\boldsymbol{a}_{\text{norm}})_{d_{\text{A}}+d_{\text{B}}+\frac{1}{2}d_{\text{A}}(d_{\text{A}}+1)+\frac{1}{2}d_{\text{B}}(d_{\text{B}}+1)+1}(m^2n)^{-1/2}\big(\boldsymbol{W}^0_{\sigma^2}(\boldsymbol{z}^{\otimes 2}-\boldsymbol{I})\big)_{tt}
\end{aligned}
$$

and $N_{\text{mart}} \equiv m + m' + n_{\bullet\bullet}$. In the definition of $\xi_t(m,m',\boldsymbol{n})$, the notation $\boldsymbol{w}^0_{\text{A}r}$ signifies that each of the model parameters that appear in the definition of $\boldsymbol{w}_{\text{A}r}$ is set to their true values. A similar convention applies to the $\boldsymbol{w}^0_{\text{B}r}$, $\boldsymbol{W}^0_{(r,s)}$, $(\boldsymbol{W}')^0_{(r,s)}$ and $\boldsymbol{W}^0_{\sigma^2}$. Let $\mathcal{X}$ denote the full set of predictor random variables in $\boldsymbol{X}_{\text{A}}$ and $\boldsymbol{X}_{\text{B}}$. For $1 \le t \le m$, let

$$\mathcal{F}_t(m,m'\boldsymbol{n}) \text{ denote the } \sigma\text{-field generated by } \mathcal{X}, \boldsymbol{U}_1,\ldots,\boldsymbol{U}_t.$$

For $m \le t \le m+m'$, let

$$\mathcal{F}_t(m,m',\boldsymbol{n}) \text{ denote the } \sigma\text{-field generated by } \mathcal{X}, \boldsymbol{U}_1,\ldots,\boldsymbol{U}_m,\boldsymbol{U}'_1,\ldots,\boldsymbol{U}'_t.$$

For $m+m'+1 \le t \le N_{\text{mart}}$, let

$$
\begin{aligned}
\mathcal{F}_t(m,m',\boldsymbol{n}) \text{ denote the } \sigma\text{-field generated by } & \mathcal{X}, \boldsymbol{U}_1,\ldots,\boldsymbol{U}_m,\boldsymbol{U}'_1,\ldots,\boldsymbol{U}'_{m'}, \\
& (\boldsymbol{Y}-\boldsymbol{X}_{\text{A}}\boldsymbol{\beta}^0_{\text{A}}-\boldsymbol{X}_{\text{B}}\boldsymbol{\beta}^0_{\text{B}}-\boldsymbol{Z}\boldsymbol{U}_{\text{all}})_{t-m-m'}.
\end{aligned}
$$

Then

$$\big(\xi_t(m,m',\boldsymbol{n}), \mathcal{F}_t(m,m',\boldsymbol{n})\big), \quad 1 \le t \le N_{\text{mart}},$$

is an array of martingale differences.

According to Theorem 3.2 of Hall & Heyde (1980),

$$\boldsymbol{a}^\top I\big(\boldsymbol{\theta}^0\big)^{-1/2}\nabla_{\boldsymbol{\theta}}\ell(\boldsymbol{\theta}^0) = \sum_{t=1}^{N_{\text{mart}}}\xi_t(m,m',\boldsymbol{n}) \xrightarrow{D} N(0,\boldsymbol{a}^\top\boldsymbol{a}) \tag{S.24}$$

if the $\xi_t(m,m',\boldsymbol{n})$ satisfy

$$
\begin{gathered}
\max_{1\le t\le N_{\text{mart}}}\Big|\xi_t(m,m',\boldsymbol{n})\Big| \xrightarrow{P} 0, \quad \sum_{t=1}^{N_{\text{mart}}}\xi_t(m,m',\boldsymbol{n})^2 \xrightarrow{P} \boldsymbol{a}^\top\boldsymbol{a} \\
\text{and} \quad E\left(\max_{1\le t\le N_{\text{mart}}}\xi_t(m,m',\boldsymbol{n})^2\right) \text{ is bounded in } (m,m',\boldsymbol{n}).
\end{gathered}
\tag{S.25}
$$

Arguments similar to those given in Jiang (1996) and Jiang *et al.* (2023) can be used to establish (S.25) under conditions such as (A1)–(A3). The pathway used in these references involves studying the asymptotic behaviors of the norms

$$\|\boldsymbol{w}_{\text{A}r}\|^2 = \big(\boldsymbol{X}_{\text{A}}^\top\boldsymbol{V}^{-1}\boldsymbol{X}_{\text{A}}\big)_{rr}, \quad 1 \le r \le d_{\text{A}}, \qquad \|\boldsymbol{w}_{\text{B}r}\|^2 = \big(\boldsymbol{X}_{\text{B}}^\top\boldsymbol{V}^{-1}\boldsymbol{X}_{\text{B}}\big)_{rr}, \quad 1 \le r \le d_{\text{A}},$$

$$\|\boldsymbol{W}_{(r,s)}\|^2_F = \text{tr}\big((\boldsymbol{V}^{-1}\breve{\boldsymbol{L}}_{(r,s)})^2\big), \quad \|\boldsymbol{W}'_{(r,s)}\|^2_F = \text{tr}\big((\boldsymbol{V}^{-1}\breve{\boldsymbol{L}}'_{(r,s)})^2\big), \quad \|\boldsymbol{W}_{\sigma^2}\|^2_F = \text{tr}\big(\boldsymbol{V}^{-2}\big)$$

for $(r,s)\in\mathcal{I}_{d_{\rm A}}$, as well as the maximum eigenvalues of the $\boldsymbol{W}_{(r,s)}$, $\boldsymbol{W}'_{(r,s)}$ and $\boldsymbol{W}_{\sigma^2}$ matrices. From Section S.1.3, the $\|\boldsymbol{w}_{{\rm A}r}\|^2$, $\|\boldsymbol{W}_{(r,s)}\|_F^2$ and $\|\boldsymbol{W}'_{(r,s)}\|_F^2$ quantities are each $O_P(m)$ under (A1). The $\|\boldsymbol{w}_{{\rm B}r}\|^2$ and $\|\boldsymbol{W}_{\sigma^2}\|_F^2$ quantities are $O_P(m^2n)$ under (A1). The maximum eigenvalue quantities have similar asymptotic behaviors.

The conditions in (S.25) follow from results such as

$$m^{-1}E\big(\|\boldsymbol{w}_{{\rm A}r}\|^2\big)=O(1)\quad\text{and}\quad(m^2n)^{-1}E\big(\|\boldsymbol{W}_{\sigma^2}\|_F^2\big)=O(1). \tag{S.26}$$

In the case of crossed random intercepts, these matrix norm expectations follow quickly from the Section S.1.3 results. For the general crossed random effects model (1) the $\boldsymbol{V}$ matrix is random and some additional regularity conditions are required to ensure that expectations, such as those appearing in (S.26), have the correct orders of magnitude and, in turn, provide (S.24). Assuming these regularity conditions, the Cramér-Wold Device leads to

$$I\big(\boldsymbol{\theta}^0\big)^{-1/2}\nabla_{\boldsymbol{\theta}}\ell(\boldsymbol{\theta}^0)\xrightarrow{D}N(\boldsymbol{0},\boldsymbol{I}).$$

Standard likelihood theory arguments then lead to

$$I\big(\boldsymbol{\theta}^0\big)^{1/2}(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}^0)\xrightarrow{D}N(\boldsymbol{0},\boldsymbol{I}).$$

S.1.9 Proofs of Lemmas

The derivation of Result 1 heavily depends on Lemmas 1–6. We now get to proving them.

S.1.9.1 Proof of Lemma 1

Let $\boldsymbol{e}_r$ denote the $d\times 1$ matrix with rth entry 1 and all other entries 0. Then note that

$$A_{rs}=\text{vech}(\boldsymbol{e}_r\boldsymbol{e}_s^\top)^\top\text{vech}(\boldsymbol{A}_d)\quad\text{for all } r\geq s.$$

Therefore

$$A_{rs}A_{tu}=\text{vech}(\boldsymbol{e}_r\boldsymbol{e}_s^\top)^\top\text{vech}(\boldsymbol{A}_d)\text{vech}(\boldsymbol{A}_d)^\top\text{vech}(\boldsymbol{e}_t\boldsymbol{e}_u^\top)\quad\text{for all } r\geq s,\ t\geq u.$$

Next, note that

$$\boldsymbol{D}_d\text{vech}(\boldsymbol{e}_r\boldsymbol{e}_s^\top)=\boldsymbol{D}_d\text{vech}\big(\boldsymbol{e}_r\boldsymbol{e}_s^\top+I(r\neq s)\boldsymbol{e}_s\boldsymbol{e}_r^\top\big)=\text{vec}\big(\boldsymbol{e}_r\boldsymbol{e}_s^\top+I(r\neq s)\boldsymbol{e}_s\boldsymbol{e}_r^\top\big)\quad\text{for all } r\geq s.$$

Use of the $\text{vec}(\boldsymbol{a}\boldsymbol{b}^\top)=\boldsymbol{b}\otimes\boldsymbol{a}$ identity then gives

$$\boldsymbol{D}_d\text{vech}(\boldsymbol{e}_r\boldsymbol{e}_s^\top)=\boldsymbol{e}_r\otimes\boldsymbol{e}_s+I(r\neq s)(\boldsymbol{e}_s\otimes\boldsymbol{e}_r).$$

We then have for all $r\geq s$ and $t\geq u$

$$\begin{aligned}\text{vech}(\boldsymbol{e}_r\boldsymbol{e}_s^\top)^\top\boldsymbol{D}_d^\top(\boldsymbol{A}\otimes\boldsymbol{A})\boldsymbol{D}_d\text{vech}(\boldsymbol{e}_t\boldsymbol{e}_u^\top)&=\{\boldsymbol{e}_r\otimes\boldsymbol{e}_s+I(r\neq s)(\boldsymbol{e}_s\otimes\boldsymbol{e}_r)\}^\top(\boldsymbol{A}\otimes\boldsymbol{A})\\&\qquad\times\{\boldsymbol{e}_t\otimes\boldsymbol{e}_u+I(t\neq u)(\boldsymbol{e}_u\otimes\boldsymbol{e}_t)\}\\&=(\boldsymbol{e}_r^\top\boldsymbol{A}\boldsymbol{e}_t)(\boldsymbol{e}_s^\top\boldsymbol{A}\boldsymbol{e}_u)+I(t\neq u)(\boldsymbol{e}_r^\top\boldsymbol{A}\boldsymbol{e}_u)(\boldsymbol{e}_s^\top\boldsymbol{A}\boldsymbol{e}_t)\\&\quad+I(r\neq s)(\boldsymbol{e}_s^\top\boldsymbol{A}\boldsymbol{e}_t)(\boldsymbol{e}_r^\top\boldsymbol{A}\boldsymbol{e}_u)\\&\quad+I(r\neq s)I(t\neq u)(\boldsymbol{e}_s^\top\boldsymbol{A}\boldsymbol{e}_u)(\boldsymbol{e}_r^\top\boldsymbol{A}\boldsymbol{e}_t)\\&=\begin{cases}A_{rt}^2, & r=s,t=u,\\ 2A_{rt}A_{ru}, & r=s,t>u,\\ 2(A_{rt}A_{su}+A_{ru}A_{st}), & r>s,t>u\end{cases}\\&=\text{vech}(\boldsymbol{e}_r\boldsymbol{e}_s^\top)^\top\boldsymbol{B}_d\text{vech}(\boldsymbol{e}_t\boldsymbol{e}_u^\top).\end{aligned}$$

Therefore, the $r\geq s$ and $t\geq u$ entries of $\boldsymbol{B}_d$ match those of $\boldsymbol{D}_d^\top(\boldsymbol{A}\otimes\boldsymbol{A})\boldsymbol{D}_d$. However, if the roles of r and s are reversed then each of the expressions involving A_{vw} forms is unaffected and the $r\geq s$ ordering restriction can be removed. The $t\geq u$ ordering restriction can be removed for the same reason and Lemma 1 is established.

S.1.9.2 Proof of Lemma 2

We start with a statement of *Woodbury's matrix identity* (Woodbury, 1950). For invertible matrices $\boldsymbol{S}$ $(n \times n)$ and $\boldsymbol{T}$ $(d \times d)$ and additional matrices $\boldsymbol{U}(n \times d)$ and $\boldsymbol{V}$ $(d \times n)$,

$$(\boldsymbol{S} + \boldsymbol{U}\boldsymbol{T}\boldsymbol{V})^{-1} = \boldsymbol{S}^{-1} - \boldsymbol{S}^{-1}\boldsymbol{U}(\boldsymbol{T}^{-1} + \boldsymbol{V}\boldsymbol{S}^{-1}\boldsymbol{U})^{-1}\boldsymbol{V}\boldsymbol{S}^{-1}. \tag{S.27}$$

Application of (S.27) with

$$\boldsymbol{S} = \lambda\boldsymbol{I}_n, \quad \boldsymbol{T} = \boldsymbol{A}, \quad \boldsymbol{U} = \boldsymbol{X} \quad \text{and} \quad \boldsymbol{V} = \boldsymbol{X}^T$$

leads to

$$(\boldsymbol{X}\boldsymbol{A}\boldsymbol{X}^\top + \lambda\boldsymbol{I})^{-1} = (1/\lambda)\boldsymbol{I}_n - (1/\lambda^2)\boldsymbol{X}(\boldsymbol{A}^{-1} + \boldsymbol{X}^\top\boldsymbol{X}/\lambda)^{-1}\boldsymbol{X}^\top. \tag{S.28}$$

Therefore,

$$\begin{aligned}
\dot{\boldsymbol{X}}^\top(\boldsymbol{X}\boldsymbol{A}\dot{\boldsymbol{X}}^\top + \lambda\boldsymbol{I})^{-1}\ddot{\boldsymbol{X}} &= (1/\lambda)\dot{\boldsymbol{X}}^\top\ddot{\boldsymbol{X}} - (1/\lambda^2)\dot{\boldsymbol{X}}^\top\boldsymbol{X}(\boldsymbol{A}^{-1} + \boldsymbol{X}^\top\boldsymbol{X}/\lambda)^{-1}\boldsymbol{X}^\top\ddot{\boldsymbol{X}} \\
&= (1/\lambda)\dot{\boldsymbol{X}}^\top\ddot{\boldsymbol{X}} - (1/\lambda^2)\dot{\boldsymbol{X}}^\top\boldsymbol{X}[(1/\lambda)\boldsymbol{X}^\top\boldsymbol{X}\{\boldsymbol{I}_d + \lambda(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{A}^{-1}\}]^{-1}\boldsymbol{X}^\top\ddot{\boldsymbol{X}} \\
&= (1/\lambda)\dot{\boldsymbol{X}}^\top\ddot{\boldsymbol{X}} - (1/\lambda^2)\dot{\boldsymbol{X}}^\top\boldsymbol{X}\{\boldsymbol{I}_d + \lambda(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{A}^{-1}\}^{-1}\lambda(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top\ddot{\boldsymbol{X}} \\
&= (1/\lambda)\dot{\boldsymbol{X}}^\top\ddot{\boldsymbol{X}} - (1/\lambda)\dot{\boldsymbol{X}}^\top\boldsymbol{X}\{\boldsymbol{I}_d + \lambda(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{A}^{-1}\}^{-1}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top\ddot{\boldsymbol{X}}.
\end{aligned}$$

Next we apply Woodbury's matrix identity (S.27) to $\{\boldsymbol{I}_d + \lambda(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{A}^{-1}\}^{-1}$ with

$$\boldsymbol{S} = \boldsymbol{I}_d, \quad \boldsymbol{T} = \boldsymbol{A}^{-1}, \quad \boldsymbol{U} = (\boldsymbol{X}^\top\boldsymbol{X})^{-1} \quad \text{and} \quad \boldsymbol{V} = \lambda\boldsymbol{I}_d$$

to obtain

$$\{\boldsymbol{I}_d + \lambda(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{A}^{-1}\}^{-1} = \boldsymbol{I}_d - (\boldsymbol{X}^\top\boldsymbol{X})^{-1}\{\boldsymbol{A} + \lambda(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\}^{-1}\lambda.$$

Plugging this into the above set of equations we have

$$\begin{aligned}
\dot{\boldsymbol{X}}^\top(\boldsymbol{X}\boldsymbol{A}\boldsymbol{X}^\top + \lambda\boldsymbol{I})^{-1}\ddot{\boldsymbol{X}} &= (1/\lambda)\dot{\boldsymbol{X}}^\top\dot{\boldsymbol{X}} - (1/\lambda)\dot{\boldsymbol{X}}^\top\dot{\boldsymbol{X}}(\boldsymbol{X}^\top\dot{\boldsymbol{X}})^\top(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top\ddot{\boldsymbol{X}} \\
&\quad + \dot{\boldsymbol{X}}^\top\boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\{\boldsymbol{A} + \lambda(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\}^{-1}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top\ddot{\boldsymbol{X}} \\
&= (1/\lambda)\dot{\boldsymbol{X}}^\top\{\boldsymbol{I}_n - \boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top\}\ddot{\boldsymbol{X}} \\
&\quad + \dot{\boldsymbol{X}}^\top\boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\{\boldsymbol{A} + \lambda(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\}^{-1}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top\ddot{\boldsymbol{X}}
\end{aligned}$$

and the lemma is proven.

S.1.9.3 Proof of Lemma 3

It follows from (S.28) that

$$\begin{aligned}
\boldsymbol{X}^\top(\boldsymbol{X}\boldsymbol{A}\boldsymbol{X}^\top + \lambda\boldsymbol{I})^{-2}\boldsymbol{X} &= (1/\lambda^2)\boldsymbol{X}^\top\boldsymbol{X} - (2/\lambda^3)\boldsymbol{X}^\top\boldsymbol{X}(\boldsymbol{A}^{-1} + \boldsymbol{X}^\top\boldsymbol{X}/\lambda)^{-1}\boldsymbol{X}^\top\boldsymbol{X} \\
&\quad + (1/\lambda^4)\boldsymbol{X}^\top\boldsymbol{X}(\boldsymbol{A}^{-1} + \boldsymbol{X}^\top\boldsymbol{X}/\lambda)^{-1} \\
&\quad \times \boldsymbol{X}^\top\boldsymbol{X}(\boldsymbol{A}^{-1} + \boldsymbol{X}^\top\boldsymbol{X}/\lambda)^{-1}\boldsymbol{X}^\top\boldsymbol{X}.
\end{aligned} \tag{S.29}$$

Steps similar to those given in the proof of Lemma 2 lead to

$$(\boldsymbol{A}^{-1} + \boldsymbol{X}^\top\boldsymbol{X}/\lambda)^{-1}\boldsymbol{X}^\top\boldsymbol{X} = \lambda\boldsymbol{I} - \lambda^2(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\{\boldsymbol{A} + \lambda(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\}^{-1}. \tag{S.30}$$

Triple substitution of (S.30) into (S.29) and simplification yields the stated result.

S.1.9.4 Proof of Lemma 4

Lemma 4 follows quickly from the following identity:

$$\begin{bmatrix} \boldsymbol{A} & \boldsymbol{B} \\ \boldsymbol{B} & \boldsymbol{A} \end{bmatrix} \begin{bmatrix} \boldsymbol{I}_d \\ \boldsymbol{I}_d \end{bmatrix} = \begin{bmatrix} \boldsymbol{I}_d \\ \boldsymbol{I}_d \end{bmatrix} (\boldsymbol{A}+\boldsymbol{B}).$$

S.1.9.5 Proof of Lemma 5

Proof of Lemma 5(a)–(c)

In the special case of $d = 1$, $n_{ii'} = n$ and $\boldsymbol{X}_{ii'} = \mathbf{1}_n$ for all $1 \le i \le m$, $1 \le i' \le m'$. Long-winded, but straightforward, algebraic arguments based on the eigenvalue and eigenvector properties described near the commencement of Section S.1.3.8 lead to the exact expression

$$\mathbf{1}_{m'n}^{\top}\boldsymbol{Q}_{mm'}^{ii}\mathbf{1}_{m'n} = \\ \frac{mm'M'\{(m-1)M'+m'M\}+m'\{(m-1)M'+mM'+m'M\}(\lambda/n)+m'(\lambda/n)^2}{(mM'+\lambda/n)\{m'M(m'M+mM')+(mM'+2m'M)(\lambda/n)+(\lambda/n)^2\}}$$

for all $1 \le i \le m$. This result leads to

$$\lim_{n\to\infty}\left(\mathbf{1}_{m'n}^{\top}\boldsymbol{Q}_{mm'}^{ii}\mathbf{1}_{m'n}\right) = \frac{1}{M} - \frac{M'}{mm'M}\left(\frac{M}{m}+\frac{M'}{m'}\right)^{-1} \quad \text{for all} \quad m, m' \in \mathbb{N} \tag{S.31}$$

which proves Lemma 5(b) in this scalar case. Similar calculations lead to

$$\lim_{n\to\infty}\left(\mathbf{1}_{mm'n}^{\top}\boldsymbol{Q}_{mm'}^{-1}\mathbf{1}_{mm'n}\right) = \left(\frac{M}{m}+\frac{M'}{m'}\right)^{-1} \quad \text{for all} \quad m, m' \in \mathbb{N}. \tag{S.32}$$

The result

$$\lim_{n\to\infty}\left(\mathbf{1}_{m'n}^{\top}\boldsymbol{Q}_{mm'}^{i\underset{\sim}{i}}\mathbf{1}_{m'n}\right) = -\frac{M'}{mm'M}\left(\frac{M}{m}+\frac{M'}{m'}\right)^{-1} \quad \text{for} \quad i \ne \underset{\sim}{i} \tag{S.33}$$

follows by subtraction and symmetric considerations.

Next, consider the general $d \in \mathbb{N}$ and unrestricted $n_{ii'}$ setting, but with $m = 2$ and $m' = 1$. Then

$$\boldsymbol{Q}_{21} = \begin{bmatrix} \boldsymbol{X}_{11}(\boldsymbol{M}+\boldsymbol{M}')\boldsymbol{X}_{11}^{\top}+\lambda\boldsymbol{I} & \boldsymbol{X}_{11}\boldsymbol{M}'\boldsymbol{X}_{21}^{\top} \\ \boldsymbol{X}_{21}\boldsymbol{M}'\boldsymbol{X}_{11}^{\top} & \boldsymbol{X}_{21}(\boldsymbol{M}+\boldsymbol{M}')\boldsymbol{X}_{21}^{\top}+\lambda\boldsymbol{I} \end{bmatrix}$$

and so, using Corollary 2.1.(c),

$$\begin{aligned}\boldsymbol{X}_{11}^{\top}\boldsymbol{Q}_{21}^{11}\boldsymbol{X}_{11} &= \boldsymbol{X}_{11}^{\top}\big[\boldsymbol{X}_{11}(\boldsymbol{M}+\boldsymbol{M}')\boldsymbol{X}_{11}^{\top} \\ &\quad -\boldsymbol{X}_{11}\boldsymbol{M}'\boldsymbol{X}_{21}^{\top}\{\boldsymbol{X}_{21}(\boldsymbol{M}+\boldsymbol{M}')\boldsymbol{X}_{21}^{\top}+\lambda\boldsymbol{I}\}^{-1}\boldsymbol{X}_{21}\boldsymbol{M}'\boldsymbol{X}_{11}^{\top}+\lambda\boldsymbol{I}\big]^{-1}\boldsymbol{X}_{11} \\ &= \big[\boldsymbol{M}+\boldsymbol{M}'-\boldsymbol{M}'\{\boldsymbol{M}+\boldsymbol{M}'+\lambda(\boldsymbol{X}_{21}^{\top}\boldsymbol{X}_{21})^{-1}\}^{-1}\boldsymbol{M}'+\lambda(\boldsymbol{X}_{11}^{\top}\boldsymbol{X}_{11})^{-1}\big]^{-1} \\ &\xrightarrow{P} \{\boldsymbol{M}+\boldsymbol{M}'-\boldsymbol{M}'(\boldsymbol{M}+\boldsymbol{M}')^{-1}\boldsymbol{M}'\}^{-1} = \Big[\boldsymbol{M}+\boldsymbol{M}'\{\boldsymbol{I}-(\boldsymbol{M}+\boldsymbol{M}')^{-1}\boldsymbol{M}'\}\Big]^{-1}.\end{aligned}$$

Noting that $\boldsymbol{I}-(\boldsymbol{M}+\boldsymbol{M}')^{-1}\boldsymbol{M}' = (\boldsymbol{M}+\boldsymbol{M}')^{-1}\boldsymbol{M}$ we then have the convergence in probability limit $\big\{\boldsymbol{M}+\boldsymbol{M}'(\boldsymbol{M}+\boldsymbol{M}')^{-1}\boldsymbol{M}\big\}^{-1}$. Application of Woodbury's matrix identity (S.27) with $\boldsymbol{S} = \boldsymbol{M}$, $\boldsymbol{U} = \boldsymbol{M}'$, $\boldsymbol{T} = (\boldsymbol{M}+\boldsymbol{M}')^{-1}$ and $\boldsymbol{V} = \boldsymbol{M}$ leads to the limit

$$\boldsymbol{M}^{-1}-\boldsymbol{M}^{-1}\boldsymbol{M}'(\boldsymbol{M}+2\boldsymbol{M}')^{-1} = \boldsymbol{M}^{-1}-\tfrac{1}{2}\boldsymbol{M}^{-1}\boldsymbol{M}'(\tfrac{1}{2}\boldsymbol{M}+\boldsymbol{M}')^{-1}$$

which verifies Lemma 5(b) for $m=2$, $m'=1$ and $i=1$. The proof for $i=2$ is very similar. Then note that, using Corollary 2.1(c),

$$\begin{aligned}
\boldsymbol{X}_{11}^\top \boldsymbol{Q}_{21}^{12} \boldsymbol{X}_{21} &= -\boldsymbol{X}_{11}^\top \big[\boldsymbol{X}_{11}(\boldsymbol{M}+\boldsymbol{M}')\boldsymbol{X}_{11}^\top \\
&\quad -\boldsymbol{X}_{11}\boldsymbol{M}'\boldsymbol{X}_{21}^\top\{\boldsymbol{X}_{21}(\boldsymbol{M}+\boldsymbol{M}')\boldsymbol{X}_{21}^\top+\lambda\boldsymbol{I}\}^{-1}\boldsymbol{X}_{21}\boldsymbol{M}'\boldsymbol{X}_{11}^\top\big]^{-1} \\
&\quad \times\boldsymbol{X}_{11}\boldsymbol{M}'\boldsymbol{X}_{21}^\top\{\boldsymbol{X}_{21}(\boldsymbol{M}+\boldsymbol{M}')\boldsymbol{X}_{21}^\top+\lambda\boldsymbol{I}\}^{-1}\boldsymbol{X}_{21} \\
&\xrightarrow{P} -\big\{\boldsymbol{M}+\boldsymbol{M}'-\boldsymbol{M}'(\boldsymbol{M}+\boldsymbol{M}')^{-1}\boldsymbol{M}'\big\}^{-1}\boldsymbol{M}'(\boldsymbol{M}+\boldsymbol{M}')^{-1} \\
&= -\tfrac{1}{2}\boldsymbol{M}^{-1}\boldsymbol{M}'(\tfrac{1}{2}\boldsymbol{M}+\boldsymbol{M}')^{-1}.
\end{aligned}$$

Hence Lemma 5(c) holds for $m=2$ and $m'=1$. Lemma 5(a) for $m=2$ and $m'=1$ follows from summation of the Lemma 5(b)–(c) results. This completes verification of Lemma 5(a)–(c) for $m=2$ and $m'=1$.

Next we prove Lemma 5 for all $m\ge 2$ and $m'=1$ via induction on m. Let

$$\begin{aligned}
\boldsymbol{Q}_{m+1,1} &= \begin{bmatrix} \boldsymbol{Q}_{m1} & \boldsymbol{R}_m \\ \boldsymbol{R}_m^\top & \boldsymbol{S}_m \end{bmatrix} \text{ where } \boldsymbol{S}_m \equiv \boldsymbol{X}_{m+1,1}(\boldsymbol{M}+\boldsymbol{M}')\boldsymbol{X}_{m+1,1}^\top+\lambda\boldsymbol{I}, \\
\boldsymbol{R}_m &\equiv \boldsymbol{X}_{1:m,1}\boldsymbol{M}'\boldsymbol{X}_{m+1,1}^\top \quad \text{and} \quad \boldsymbol{X}_{1:m,1}\equiv \underset{1\le i\le m}{\text{stack}}(\boldsymbol{X}_{i1}).
\end{aligned} \tag{S.34}$$

We then have, with use of Corollary 2.1.(c),

$$\begin{aligned}
\boldsymbol{X}_{m+1,1}^\top \boldsymbol{Q}_{m+1,1}^{m+1,m+1}\boldsymbol{X}_{m+1,1} &= \boldsymbol{X}_{m+1,1}^\top\Big(\boldsymbol{S}_m-\boldsymbol{R}_m^\top\boldsymbol{Q}_{m1}^{-1}\boldsymbol{R}_m\Big)^{-1}\boldsymbol{X}_{m+1,1} \\
&= \boldsymbol{X}_{m+1,1}^\top\Big\{\boldsymbol{X}_{m+1,1}(\boldsymbol{M}+\boldsymbol{M}')\boldsymbol{X}_{m+1,1}^\top \\
&\quad -\boldsymbol{X}_{m+1,1}\boldsymbol{M}'\boldsymbol{X}_{1:m,1}^\top\boldsymbol{Q}_{m1}^{-1}\boldsymbol{X}_{1:m,1}\boldsymbol{M}'\boldsymbol{X}_{m+1,1}^\top+\lambda\boldsymbol{I}\Big\}^{-1}\boldsymbol{X}_{m+1,1} \\
&= \Big\{\boldsymbol{M}+\boldsymbol{M}'-\boldsymbol{M}'\boldsymbol{X}_{1:m,1}^\top\boldsymbol{Q}_{m1}^{-1}\boldsymbol{X}_{1:m,1}\boldsymbol{M}'+\lambda(\boldsymbol{X}_{m+1,1}^\top\boldsymbol{X}_{m+1,1})^{-1}\Big\}^{-1} \\
&\xrightarrow{P} \Big\{\boldsymbol{M}+\boldsymbol{M}'-\boldsymbol{M}'\Big(\frac{1}{m}\boldsymbol{M}+\boldsymbol{M}'\Big)^{-1}\boldsymbol{M}'\Big\}^{-1} \\
&= \boldsymbol{M}^{-1}-\frac{1}{m+1}\boldsymbol{M}^{-1}\boldsymbol{M}'\Big(\frac{1}{m+1}\boldsymbol{M}+\boldsymbol{M}'\Big)^{-1}.
\end{aligned}$$

Analogous arguments for other partitions of $\boldsymbol{Q}_{m+1,1}$ lead to the same convergence in probability limit for $\boldsymbol{X}_{i,1}^\top\boldsymbol{Q}_{m+1,1}^{i,i}\boldsymbol{X}_{i,1}$ for each $1\le i\le m+1$. Therefore, by induction, Lemma 5(b) holds for all $m\ge 2$ and $m'=1$.

Next, let $\boldsymbol{Q}_{m+1,1}^{1:m,m+1}\equiv \underset{1\le i\le m}{\text{stack}}\big(\boldsymbol{Q}_{m+1,1}^{i,m+1}\big)$ and note that

$$\begin{aligned}
\boldsymbol{X}_{1:m,1}^\top\boldsymbol{Q}_{m+1,1}^{1:m,m+1}\boldsymbol{X}_{m+1,1} &= -\boldsymbol{X}_{1:m,1}^\top\boldsymbol{Q}_{m1}^{-1}\boldsymbol{R}_m\Big(\boldsymbol{S}_m-\boldsymbol{R}_m^\top\boldsymbol{Q}_{m1}^{-1}\boldsymbol{R}_m\Big)^{-1}\boldsymbol{X}_{m+1,1} \\
&= -\boldsymbol{X}_{1:m,1}^\top\boldsymbol{Q}_{m1}^{-1}\boldsymbol{X}_{1:m,1}\boldsymbol{M}'\boldsymbol{X}_{m+1,1}^\top\big\{\boldsymbol{X}_{m+1,1}(\boldsymbol{M}+\boldsymbol{M}')\boldsymbol{X}_{m+1,1}^\top \\
&\quad -\boldsymbol{X}_{m+1,1}\boldsymbol{M}'\boldsymbol{X}_{1:m,1}^\top\boldsymbol{Q}_{m1}^{-1}\boldsymbol{X}_{1:m,1}\boldsymbol{M}'\boldsymbol{X}_{m+1,1}^\top+\lambda\boldsymbol{I}\big\}^{-1}\boldsymbol{X}_{m+1,1} \\
&= -\boldsymbol{X}_{1:m,1}^\top\boldsymbol{Q}_{m1}^{-1}\boldsymbol{X}_{1:m,1}\boldsymbol{M}'\big\{\boldsymbol{M}+\boldsymbol{M}'-\boldsymbol{M}'\boldsymbol{X}_{1:m,1}^\top\boldsymbol{Q}_{m1}^{-1}\boldsymbol{X}_{1:m,1}\boldsymbol{M}' \\
&\quad +\lambda(\boldsymbol{X}_{m+1,1}^\top\boldsymbol{X}_{m+1,1})^{-1}\big\}^{-1}
\end{aligned}$$

$$\xrightarrow{P} -\Big(\frac{1}{m}\boldsymbol{M}+\boldsymbol{M}'\Big)^{-1}\boldsymbol{M}'\Big\{\boldsymbol{M}+\boldsymbol{M}'-\boldsymbol{M}'\Big(\frac{1}{m}\boldsymbol{M}+\boldsymbol{M}'\Big)^{-1}\boldsymbol{M}'\Big\}^{-1}$$
$$= -\Big(\frac{1}{m}\boldsymbol{M}+\boldsymbol{M}'\Big)^{-1}\boldsymbol{M}'\Big\{\boldsymbol{M}^{-1}-\frac{1}{m+1}\boldsymbol{M}^{-1}\boldsymbol{M}'\Big(\frac{1}{m+1}\boldsymbol{M}+\boldsymbol{M}'\Big)^{-1}\Big\}$$
$$= -\frac{m}{m+1}\boldsymbol{M}^{-1}\boldsymbol{M}'\Big(\frac{1}{m+1}\boldsymbol{M}+\boldsymbol{M}'\Big)^{-1}.$$

However,

$$\boldsymbol{X}_{1:m,1}^{\top}\boldsymbol{Q}_{m+1,1}^{1:m,m+1}\boldsymbol{X}_{m+1,1}=\sum_{i=1}^{m}\boldsymbol{X}_{i1}^{\top}\boldsymbol{Q}_{m+1,1}^{i,m+1}\boldsymbol{X}_{m+1,1} \tag{S.35}$$

and each term in the summation on the right-hand side of (S.35) has the same distribution and, therefore, the same convergence in probability limit. Hence,

$$\boldsymbol{X}_{i1}^{\top}\boldsymbol{Q}_{m+1,1}^{i,m+1}\boldsymbol{X}_{m+1,1}\xrightarrow{P} -\frac{1}{m+1}\boldsymbol{M}^{-1}\boldsymbol{M}'\Big(\frac{1}{m+1}\boldsymbol{M}+\boldsymbol{M}'\Big)^{-1},\quad 1\le i\le m.$$

Analogous arguments for other partitions of $\boldsymbol{Q}_{m+1}$ lead to

$$\boldsymbol{X}_{i1}^{\top}\boldsymbol{Q}_{m+1,1}^{i,\underset{\sim}{i}}\boldsymbol{X}_{\underset{\sim}{i}1}\xrightarrow{P} -\frac{1}{m+1}\boldsymbol{M}^{-1}\boldsymbol{M}'\Big(\frac{1}{m+1}\boldsymbol{M}+\boldsymbol{M}'\Big)^{-1},\quad 1\le i\ne \underset{\sim}{i}\le m+1,$$

and by induction, Lemma 5(b) and (c) hold for $m\ge 2$ and $m'=1$.

To establish Lemma 5(a) for $m\ge 2$ and $m'=1$ we sum the results just derived for Lemma 5(b) and (c):

$$\begin{aligned}\Big\{\underset{1\le i\le m+1}{\text{stack}}(\boldsymbol{X}_{i1})\Big\}^{\top}\boldsymbol{Q}_{m+1,1}^{-1}\underset{1\le i\le m+1}{\text{stack}}(\boldsymbol{X}_{i1}) &= \sum_{i=1}^{m+1}\sum_{\underset{\sim}{i}=1}^{m+1}\boldsymbol{X}_{i,1}^{\top}\boldsymbol{Q}_{m+1,1}^{i,\underset{\sim}{i}}\boldsymbol{X}_{\underset{\sim}{i},1}\\
&\xrightarrow{P}(m+1)\Big\{\boldsymbol{M}^{-1}-\frac{1}{m+1}\boldsymbol{M}^{-1}\boldsymbol{M}'\Big(\frac{1}{m+1}\boldsymbol{M}+\boldsymbol{M}'\Big)^{-1}\Big\}\\
&\quad+m(m+1)\Big\{-\frac{1}{m+1}\boldsymbol{M}^{-1}\boldsymbol{M}'\Big(\frac{1}{m+1}\boldsymbol{M}+\boldsymbol{M}'\Big)^{-1}\Big\}\\
&=\Big(\frac{\boldsymbol{M}}{m+1}+\boldsymbol{M}'\Big)^{-1}.\end{aligned}$$

Induction then leads to Lemma 5(a) holding for $m\ge 2$ and $m'=1$. This completes verification of Lemma 5 for $m\ge 2$ and $m'=1$.

For the $m=1$ and $m'=2$ case the matrix of interest is

$$\boldsymbol{Q}_{12}^{11}=\boldsymbol{Q}_{12}^{-1}$$

where

$$\begin{aligned}\boldsymbol{Q}_{12}&=\begin{bmatrix}\boldsymbol{X}_{11}(\boldsymbol{M}+\boldsymbol{M}')\boldsymbol{X}_{11}^{\top} & \boldsymbol{X}_{11}\boldsymbol{M}\boldsymbol{X}_{12}^{\top}\\ \boldsymbol{X}_{12}\boldsymbol{M}\boldsymbol{X}_{11}^{\top} & \boldsymbol{X}_{12}(\boldsymbol{M}+\boldsymbol{M}')\boldsymbol{X}_{12}^{\top}\end{bmatrix}+\lambda\boldsymbol{I}\\
&=\begin{bmatrix}\boldsymbol{X}_{11} & \boldsymbol{O}\\ \boldsymbol{O} & \boldsymbol{X}_{12}\end{bmatrix}\begin{bmatrix}\boldsymbol{M}+\boldsymbol{M}' & \boldsymbol{M}\\ \boldsymbol{M} & \boldsymbol{M}+\boldsymbol{M}'\end{bmatrix}\begin{bmatrix}\boldsymbol{X}_{11} & \boldsymbol{O}\\ \boldsymbol{O} & \boldsymbol{X}_{12}\end{bmatrix}^{\top}+\lambda\boldsymbol{I}.\end{aligned}$$

Noting that

$$\overset{\blacktriangle}{\boldsymbol{X}}_1\equiv\begin{bmatrix}\boldsymbol{X}_{11}\\ \boldsymbol{X}_{12}\end{bmatrix}=\begin{bmatrix}\boldsymbol{X}_{11} & \boldsymbol{O}\\ \boldsymbol{O} & \boldsymbol{X}_{12}\end{bmatrix}\begin{bmatrix}\boldsymbol{I}_d\\ \boldsymbol{I}_d\end{bmatrix}$$

we have

$$\overset{\blacktriangle}{\boldsymbol{X}}_1^\top \boldsymbol{Q}_{12}^{11} \overset{\blacktriangle}{\boldsymbol{X}}_1 = \begin{bmatrix} \boldsymbol{I}_d \\ \boldsymbol{I}_d \end{bmatrix}^\top \begin{bmatrix} \boldsymbol{X}_{11} & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{X}_{12} \end{bmatrix}^\top \left\{ \begin{bmatrix} \boldsymbol{X}_{11} & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{X}_{12} \end{bmatrix} \begin{bmatrix} \boldsymbol{M}+\boldsymbol{M}' & \boldsymbol{M} \\ \boldsymbol{M} & \boldsymbol{M}+\boldsymbol{M}' \end{bmatrix} \begin{bmatrix} \boldsymbol{X}_{11} & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{X}_{12} \end{bmatrix}^\top + \lambda \boldsymbol{I} \right\}^{-1}$$

$$\times \begin{bmatrix} \boldsymbol{X}_{11} & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{X}_{12} \end{bmatrix} \begin{bmatrix} \boldsymbol{I}_d \\ \boldsymbol{I}_d \end{bmatrix}$$

$$= \begin{bmatrix} \boldsymbol{I}_d \\ \boldsymbol{I}_d \end{bmatrix}^\top \left\{ \begin{bmatrix} \boldsymbol{M}+\boldsymbol{M}' & \boldsymbol{M} \\ \boldsymbol{M} & \boldsymbol{M}+\boldsymbol{M}' \end{bmatrix} + \lambda \begin{bmatrix} (\boldsymbol{X}_{11}^\top \boldsymbol{X}_{11})^{-1} & \boldsymbol{O} \\ \boldsymbol{O} & (\boldsymbol{X}_{12}^\top \boldsymbol{X}_{12})^{-1} \end{bmatrix} \right\}^{-1} \begin{bmatrix} \boldsymbol{I}_d \\ \boldsymbol{I}_d \end{bmatrix}$$

$$\xrightarrow{P} \begin{bmatrix} \boldsymbol{I}_d \\ \boldsymbol{I}_d \end{bmatrix}^\top \begin{bmatrix} \boldsymbol{M}+\boldsymbol{M}' & \boldsymbol{M} \\ \boldsymbol{M} & \boldsymbol{M}+\boldsymbol{M}' \end{bmatrix}^{-1} \begin{bmatrix} \boldsymbol{I}_d \\ \boldsymbol{I}_d \end{bmatrix} = (\boldsymbol{M} + \tfrac{1}{2}\boldsymbol{M}')^{-1}$$

where the last equality is due to Lemma 4.

For the $m = 2$ and $m' = 2$ case the matrix of interest is

$$\boldsymbol{Q}_{22}^{11} = \text{the top left } (n_{11} + n_{12}) \times (n_{11} + n_{12}) \text{ block of } \begin{bmatrix} \boldsymbol{Q}_{12} & \boldsymbol{R}_{12} \\ \boldsymbol{R}_{12}^\top & \widetilde{\boldsymbol{Q}}_{12} \end{bmatrix}^{-1}$$

where

$$\boldsymbol{R}_{12} = \begin{bmatrix} \boldsymbol{X}_{11}\boldsymbol{M}'\boldsymbol{X}_{12}^\top & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{X}_{21}\boldsymbol{M}'\boldsymbol{X}_{22}^\top \end{bmatrix} = \begin{bmatrix} \boldsymbol{X}_{11} & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{X}_{12} \end{bmatrix} \begin{bmatrix} \boldsymbol{M}' & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{M}' \end{bmatrix} \begin{bmatrix} \boldsymbol{X}_{21} & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{X}_{22} \end{bmatrix}^\top$$

and

$$\widetilde{\boldsymbol{Q}}_{12} = \begin{bmatrix} \boldsymbol{X}_{21}(\boldsymbol{M}+\boldsymbol{M}')\boldsymbol{X}_{21}^\top + \lambda\boldsymbol{I} & \boldsymbol{X}_{21}\boldsymbol{M}\boldsymbol{X}_{22}^\top \\ \boldsymbol{X}_{22}\boldsymbol{M}\boldsymbol{X}_{21}^\top & \boldsymbol{X}_{22}(\boldsymbol{M}+\boldsymbol{M}')\boldsymbol{X}_{22}^\top + \lambda\boldsymbol{I} \end{bmatrix}$$

$$= \begin{bmatrix} \boldsymbol{X}_{21} & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{X}_{22} \end{bmatrix} \begin{bmatrix} \boldsymbol{M}+\boldsymbol{M}' & \boldsymbol{M} \\ \boldsymbol{M} & \boldsymbol{M}+\boldsymbol{M}' \end{bmatrix} \begin{bmatrix} \boldsymbol{X}_{21} & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{X}_{22} \end{bmatrix}^\top + \lambda\boldsymbol{I}.$$

Therefore,

$$\overset{\blacktriangle}{\boldsymbol{X}}_1^\top \boldsymbol{Q}_{22}^{11} \overset{\blacktriangle}{\boldsymbol{X}}_1 = \overset{\blacktriangle}{\boldsymbol{X}}_1^\top \left(\boldsymbol{Q}_{12} - \boldsymbol{R}_{12}\widetilde{\boldsymbol{Q}}_{12}^{-1}\boldsymbol{R}_{12}^\top \right)^{-1} \overset{\blacktriangle}{\boldsymbol{X}}_1$$

$$= \begin{bmatrix} \boldsymbol{I}_d \\ \boldsymbol{I}_d \end{bmatrix}^\top \begin{bmatrix} \boldsymbol{X}_{11} & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{X}_{12} \end{bmatrix}^\top \left\{ \begin{bmatrix} \boldsymbol{X}_{11} & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{X}_{12} \end{bmatrix} \begin{bmatrix} \boldsymbol{M}+\boldsymbol{M}' & \boldsymbol{M} \\ \boldsymbol{M} & \boldsymbol{M}+\boldsymbol{M}' \end{bmatrix} \begin{bmatrix} \boldsymbol{X}_{11} & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{X}_{12} \end{bmatrix}^\top \right.$$

$$\left. - \begin{bmatrix} \boldsymbol{X}_{11} & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{X}_{12} \end{bmatrix} \begin{bmatrix} \boldsymbol{M}' & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{M}' \end{bmatrix} \begin{bmatrix} \boldsymbol{X}_{21} & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{X}_{22} \end{bmatrix}^\top \widetilde{\boldsymbol{Q}}_{12}^{-1} \begin{bmatrix} \boldsymbol{X}_{21} & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{X}_{22} \end{bmatrix} \begin{bmatrix} \boldsymbol{M}' & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{M}' \end{bmatrix} \begin{bmatrix} \boldsymbol{X}_{11} & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{X}_{12} \end{bmatrix}^\top \right\}^{-1}$$

$$\times \begin{bmatrix} \boldsymbol{X}_{11} & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{X}_{12} \end{bmatrix} \begin{bmatrix} \boldsymbol{I}_d \\ \boldsymbol{I}_d \end{bmatrix}$$

$$= \begin{bmatrix} \boldsymbol{I}_d \\ \boldsymbol{I}_d \end{bmatrix}^\top \left\{ \begin{bmatrix} \boldsymbol{M}+\boldsymbol{M}' & \boldsymbol{M} \\ \boldsymbol{M} & \boldsymbol{M}+\boldsymbol{M}' \end{bmatrix} - \begin{bmatrix} \boldsymbol{M}' & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{M}' \end{bmatrix} \boldsymbol{\Psi} \begin{bmatrix} \boldsymbol{M}' & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{M}' \end{bmatrix} \right.$$

$$\left. + \lambda \begin{bmatrix} (\boldsymbol{X}_{11}^\top \boldsymbol{X}_{11})^{-1} & \boldsymbol{O} \\ \boldsymbol{O} & (\boldsymbol{X}_{12}^\top \boldsymbol{X}_{12})^{-1} \end{bmatrix} \right\}^{-1} \begin{bmatrix} \boldsymbol{I}_d \\ \boldsymbol{I}_d \end{bmatrix}$$

where

$$\boldsymbol{\Psi} \equiv \begin{bmatrix} \boldsymbol{X}_{21} & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{X}_{22} \end{bmatrix}^{\top} \widetilde{\boldsymbol{Q}}_{12}^{-1} \begin{bmatrix} \boldsymbol{X}_{21} & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{X}_{22} \end{bmatrix}.$$

Now note that

$$\begin{bmatrix} \boldsymbol{M}' & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{M}' \end{bmatrix} \boldsymbol{\Psi} \begin{bmatrix} \boldsymbol{M}' & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{M}' \end{bmatrix} = \begin{bmatrix} \boldsymbol{M}'\boldsymbol{X}_{21}^{\top}\widetilde{\boldsymbol{Q}}_{12}^{[1,1]}\boldsymbol{X}_{21}\boldsymbol{M}' & \boldsymbol{M}'\boldsymbol{X}_{21}^{\top}\widetilde{\boldsymbol{Q}}_{12}^{[1,2]}\boldsymbol{X}_{22}\boldsymbol{M}' \\ \boldsymbol{M}'\boldsymbol{X}_{22}^{\top}\widetilde{\boldsymbol{Q}}_{12}^{[2,1]}\boldsymbol{X}_{21}\boldsymbol{M}' & \boldsymbol{M}'\boldsymbol{X}_{22}^{\top}\widetilde{\boldsymbol{Q}}_{22}^{[2,2]}\boldsymbol{X}_{22}\boldsymbol{M}' \end{bmatrix}$$

where

$$\begin{bmatrix} \widetilde{\boldsymbol{Q}}_{12}^{[1,1]} & \widetilde{\boldsymbol{Q}}_{12}^{[1,2]} \\ \widetilde{\boldsymbol{Q}}_{12}^{[2,1]} & \widetilde{\boldsymbol{Q}}_{12}^{[2,2]} \end{bmatrix}$$

is the partition of $\widetilde{\boldsymbol{Q}}_{12}$ such that the sub-blocks have dimensions as follows:

$$\widetilde{\boldsymbol{Q}}_{12}^{[1,1]} \text{ is } n_{21} \times n_{21}, \quad \widetilde{\boldsymbol{Q}}_{12}^{[1,2]} \text{ is } n_{21} \times n_{22}, \quad \widetilde{\boldsymbol{Q}}_{12}^{[2,1]} \text{ is } n_{22} \times n_{21} \quad \text{and} \quad \widetilde{\boldsymbol{Q}}_{12}^{[2,2]} \text{ is } n_{22} \times n_{22}.$$

We then have

$$\begin{aligned}
\overset{\blacktriangle}{\boldsymbol{X}}_1^{\top}\boldsymbol{Q}_{22}^{11}\overset{\blacktriangle}{\boldsymbol{X}}_1 &= \begin{bmatrix} \boldsymbol{I}_d \\ \boldsymbol{I}_d \end{bmatrix}^{\top} \begin{bmatrix} \begin{matrix} \boldsymbol{M}+\boldsymbol{M}'-\boldsymbol{M}'\boldsymbol{X}_{21}^{\top}\widetilde{\boldsymbol{Q}}_{12}^{[1,1]}\boldsymbol{X}_{21}\boldsymbol{M}' \\ +\lambda(\boldsymbol{X}_{11}^{\top}\boldsymbol{X}_{11})^{-1} \end{matrix} & \boldsymbol{M}-\boldsymbol{M}'\boldsymbol{X}_{21}^{\top}\widetilde{\boldsymbol{Q}}_{12}^{[1,2]}\boldsymbol{X}_{22}\boldsymbol{M}' \\ \boldsymbol{M}-\boldsymbol{M}'\boldsymbol{X}_{22}^{\top}\widetilde{\boldsymbol{Q}}_{12}^{[2,1]}\boldsymbol{X}_{21}\boldsymbol{M}' & \begin{matrix} \boldsymbol{M}+\boldsymbol{M}'-\boldsymbol{M}'\boldsymbol{X}_{22}^{\top}\widetilde{\boldsymbol{Q}}_{12}^{[2,2]}\boldsymbol{X}_{22}\boldsymbol{M}' \\ +\lambda(\boldsymbol{X}_{12}^{\top}\boldsymbol{X}_{12})^{-1} \end{matrix} \end{bmatrix}^{-1} \begin{bmatrix} \boldsymbol{I}_d \\ \boldsymbol{I}_d \end{bmatrix} \\
&= \begin{bmatrix} \boldsymbol{I}_d \\ \boldsymbol{I}_d \end{bmatrix}^{\top} \begin{bmatrix} \widetilde{\boldsymbol{A}} & \widetilde{\boldsymbol{B}} \\ \widetilde{\boldsymbol{B}} & \widetilde{\boldsymbol{A}} \end{bmatrix}^{-1} \begin{bmatrix} \boldsymbol{I}_d \\ \boldsymbol{I}_d \end{bmatrix}
\end{aligned}$$

where

$$\widetilde{\boldsymbol{A}} \equiv \boldsymbol{M}+\boldsymbol{M}'-\tfrac{1}{2}\boldsymbol{M}'\left\{\boldsymbol{X}_{21}^{\top}\widetilde{\boldsymbol{Q}}_{12}^{[1,1]}\boldsymbol{X}_{21}+\boldsymbol{X}_{22}^{\top}\widetilde{\boldsymbol{Q}}_{12}^{[2,2]}\boldsymbol{X}_{22}\right\}\boldsymbol{M}'\{1+o_P(1)\}+\lambda(\boldsymbol{X}_{11}^{\top}\boldsymbol{X}_{11})^{-1}$$

and

$$\widetilde{\boldsymbol{B}} \equiv \boldsymbol{M}-\tfrac{1}{2}\boldsymbol{M}'\left\{\boldsymbol{X}_{21}^{\top}\widetilde{\boldsymbol{Q}}_{12}^{[1,2]}\boldsymbol{X}_{22}+\boldsymbol{X}_{22}^{\top}(\widetilde{\boldsymbol{Q}}_{12}^{[1,2]})^{\top}\boldsymbol{X}_{21}\right\}\boldsymbol{M}'\{1+o_P(1)\}$$

with the $\{1+o_P(1)\}$ factors being justified due to each of $\boldsymbol{X}_{11}$, $\boldsymbol{X}_{11}$, $\boldsymbol{X}_{21}$ and $\boldsymbol{X}_{22}$ containing random samples from the same distribution. Application of Lemma 4 leads to, with $\widetilde{\boldsymbol{X}} \equiv \begin{bmatrix}\boldsymbol{X}_{21}^{\top} \ \boldsymbol{X}_{22}^{\top}\end{bmatrix}^{\top}$ being the $\widetilde{\boldsymbol{Q}}_{12}$ version of the $\boldsymbol{X}$ matrix from Lemma 5(b) but for $\widetilde{\boldsymbol{Q}}_{12}$ rather than $\boldsymbol{Q}_{12}$, the result

$$\begin{aligned}
\overset{\blacktriangle}{\boldsymbol{X}}_1^{\top}\boldsymbol{Q}_{22}^{11}\overset{\blacktriangle}{\boldsymbol{X}}_1 &= 2\Big[\boldsymbol{M}+\boldsymbol{M}'+\boldsymbol{M}-\tfrac{1}{2}\boldsymbol{M}'\big(\widetilde{\boldsymbol{X}}^{\top}\widetilde{\boldsymbol{Q}}_{12}^{-1}\widetilde{\boldsymbol{X}}\big)\boldsymbol{M}'\{1+o_P(1)\}+\lambda(\boldsymbol{X}_{11}^{\top}\boldsymbol{X}_{11})^{-1}\{1+o_P(1)\}\Big]^{-1} \\
&\xrightarrow{P} 2\{2\boldsymbol{M}+\boldsymbol{M}'-\tfrac{1}{2}\boldsymbol{M}'(\boldsymbol{M}+\tfrac{1}{2}\boldsymbol{M}')^{-1}\boldsymbol{M}'\}^{-1} = \boldsymbol{M}^{-1}-\tfrac{1}{4}\boldsymbol{M}^{-1}\boldsymbol{M}'\big(\tfrac{1}{2}\boldsymbol{M}+\tfrac{1}{2}\boldsymbol{M}'\big)^{-1}.
\end{aligned}$$

which verifies Lemma 5(b) for the $(m,m')=(2,2)$ case. Induction on m can be used to show that Lemma 5(b) holds for general $m \in \mathbb{N}$ and $m'=2$.

It is apparent from these derivations in the $m' \in \{1,2\}$ cases that the behaviors of the summations that lead to the limits given by (S.31)–(S.33) in the $d=1$ and balanced cell counts situation also lead to the analogous matrix forms for general $m' \in \mathbb{N}$.

Proof of Lemma 5(d)

In the special case of $d=1$, $n_{ii'}=n$ and $\boldsymbol{X}_{ii'}=\boldsymbol{1}_n$ for all $1 \le i \le m$, $1 \le i' \le m'$. The eigenvalue and eigenvector properties described near the start of Section S.1.3.8 are such that relatively straightforward manipulations produce the exact expression

$$\frac{1}{mm'n}\,\mathrm{tr}(\boldsymbol{Q}_{mm'}^{-2}) = \frac{1}{\lambda^2} - \frac{M/(m'n)}{\lambda^2\{M+\lambda/(m'n)\}} - \frac{M'/(mn)}{\lambda^2\{M'+\lambda/(mn)\}}$$
$$+\frac{MM'/(mm'n)}{\lambda^2\{M+\lambda/(m'n)\}\{M(m'/m)+M'+\lambda/(mn)\}}$$
$$+\frac{MM'/(mm'n)}{\lambda^2\{M'+\lambda/(mn)\}\{M+M'(m/m')+\lambda/(m'n)\}}$$
$$-\frac{M'/\{(mn)^2\}}{\lambda\{M'+\lambda/(mn)\}^2} - \frac{M/\{(m'n)^2\}}{\lambda\{M+\lambda/(m'n)\}^2}$$
$$+\frac{MM'/\{m'(mn)^2\}}{\lambda\{M+M'(m/m')+\lambda/(m'n)\}\{M'+\lambda/(mn)\}^2}$$
$$+\frac{MM'/\{m'(mn)^2\}}{\lambda\{M(m'/m)+M'+\lambda/(mn)\}^2\{M+\lambda/(m'n)\}}$$
$$+\frac{MM'/\{m(m'n)^2\}}{\lambda\{M(m'/m)+M'+\lambda/(mn)\}\{M+\lambda/(m'n)\}^2}$$
$$+\frac{MM'/\{m(m'n)^2\}}{\lambda\{M+M'(m/m')+\lambda/(m'n)\}^2\{M'+\lambda/(mn)\}}.$$

Hence, under (A5),

$$\frac{1}{mm'n}\,\mathrm{tr}(\boldsymbol{Q}_{mm'}^{-2}) = \left(\sum_{i=1}^{m}\sum_{i'=1}^{m'} n_{ii'}\right)^{-1} \mathrm{tr}(\boldsymbol{Q}_{mm'}^{-2}) \to \frac{1}{\lambda^2} \tag{S.36}$$

for all $1 \le i \le m$, $1 \le i' \le m'$.

Next consider the case of $d_{\mathrm{A}} \in \mathbb{N}$ and $m = m' = 1$. Then

$$\boldsymbol{Q}_{11}^2 = \lambda^2 \boldsymbol{I}_{n_{11}} + \boldsymbol{X}_{11}\boldsymbol{\Omega}_1\boldsymbol{X}_{11}^\top \quad \text{where} \quad \boldsymbol{\Omega}_1 \equiv (\boldsymbol{M}+\boldsymbol{M}')\boldsymbol{X}_{11}^\top\boldsymbol{X}_{11}(\boldsymbol{M}+\boldsymbol{M}') + 2\lambda(\boldsymbol{M}+\boldsymbol{M}').$$

Application of Woodbury's matrix identity (S.27) with

$$\boldsymbol{S} = \lambda^2\boldsymbol{I}_{n_{11}}, \quad \boldsymbol{U} \equiv \boldsymbol{X}_{11}\boldsymbol{\Omega}_1, \quad \boldsymbol{T} = \boldsymbol{I}_{d_{\mathrm{A}}} \quad \text{and} \quad \boldsymbol{V} \equiv \boldsymbol{X}_{11}^\top$$

then gives

$$\boldsymbol{Q}_{11}^{-2} = \lambda^{-2}\boldsymbol{I}_{n_{11}} - \lambda^{-4}\boldsymbol{X}_{11}\boldsymbol{\Omega}_1(\boldsymbol{I} + \lambda^{-2}\boldsymbol{X}_{11}^\top\boldsymbol{X}_{11}\boldsymbol{\Omega}_1)^{-1}\boldsymbol{X}_{11}^\top$$

and so

$$\frac{1}{n_{11}}\mathrm{tr}(\boldsymbol{Q}_{11}^{-2}) = \frac{1}{\lambda^2} - \frac{1}{n_{11}\lambda^4}\mathrm{tr}\Big((\boldsymbol{I} + \lambda^{-2}\boldsymbol{X}_{11}^\top\boldsymbol{X}_{11}\boldsymbol{\Omega}_1)^{-1}\boldsymbol{X}_{11}^\top\boldsymbol{X}_{11}\boldsymbol{\Omega}_1\Big)$$
$$= \frac{1}{\lambda^2} - \frac{1}{n_{11}\lambda^2}\mathrm{tr}\Big(\big\{\boldsymbol{\Omega}_1 + \lambda^2(\boldsymbol{X}_{11}^\top\boldsymbol{X}_{11})^{-1}\big\}^{-1}\boldsymbol{\Omega}_1\Big) \xrightarrow{P} \frac{1}{\lambda^2}.$$

For the $d_{\mathrm{A}} \in \mathbb{N}$, $m \in \mathbb{N}$ and $m' = 1$ extension we note, as given earlier in (S.34), that

$$\boldsymbol{Q}_{m+1,1} = \begin{bmatrix} \boldsymbol{Q}_{m1} & \boldsymbol{R}_m \\ \boldsymbol{R}_m^\top & \boldsymbol{S}_m \end{bmatrix} \text{ where } \boldsymbol{S}_m \equiv \boldsymbol{X}_{m+1,1}(\boldsymbol{M}+\boldsymbol{M}')\boldsymbol{X}_{m+1,1}^\top + \lambda\boldsymbol{I},$$

$$\boldsymbol{R}_m \equiv \boldsymbol{X}_{1:m,1}\boldsymbol{M}'\boldsymbol{X}_{m+1,1}^\top \quad \text{and} \quad \boldsymbol{X}_{1:m,1} \equiv \underset{1\le i\le m}{\mathrm{stack}}(\boldsymbol{X}_{i1})$$

which gives

$$\boldsymbol{Q}_{m+1,1}^2 = \begin{bmatrix} \boldsymbol{Q}_{m1}^2 + \boldsymbol{R}_m\boldsymbol{R}_m^\top & \boldsymbol{Q}_{m1}\boldsymbol{R}_m + \boldsymbol{R}_m\boldsymbol{S}_m \\ (\boldsymbol{Q}_{m1}\boldsymbol{R}_m + \boldsymbol{R}_m\boldsymbol{S}_m)^\top & \boldsymbol{S}_m^2 + \boldsymbol{R}_m^\top\boldsymbol{R}_m \end{bmatrix}.$$

Then the lower right $n_{m+1,1} \times n_{m+1,1}$ block of $\boldsymbol{Q}_{m1}^{-2}$ equals

$$\begin{aligned}
&\big\{\boldsymbol{S}_m^2 + \boldsymbol{R}_m^\top \boldsymbol{R}_m - (\boldsymbol{Q}_{m1}\boldsymbol{R}_m + \boldsymbol{R}_m\boldsymbol{S}_m)^\top (\boldsymbol{Q}_{m1}^2 + \boldsymbol{R}_m\boldsymbol{R}_m^\top)^{-1}(\boldsymbol{Q}_{m1}\boldsymbol{R}_m + \boldsymbol{R}_m\boldsymbol{S}_m)\big\}^{-1} \\
&\qquad = \Big(\lambda^2 \boldsymbol{I}_{n_{m+1,1}} + \boldsymbol{X}_{m+1,1}^\top \boldsymbol{\Omega}_2 \boldsymbol{X}_{m+1,1}\Big)^{-1}
\end{aligned}$$

where

$$\begin{aligned}
\boldsymbol{\Omega}_2 \equiv\ & 2\lambda(\boldsymbol{M}+\boldsymbol{M}') + (\boldsymbol{M}+\boldsymbol{M}')\boldsymbol{X}_{m+1,1}^\top \boldsymbol{X}_{m+1,1}(\boldsymbol{M}+\boldsymbol{M}') + \boldsymbol{M}'\boldsymbol{X}_{1:m,1}^\top \boldsymbol{X}_{1:m,1}\boldsymbol{M}' \\
& -\boldsymbol{\Omega}_3^\top (\boldsymbol{Q}_{m1}^2 + \boldsymbol{R}_m\boldsymbol{R}_m^\top)^{-1}\boldsymbol{\Omega}_3
\end{aligned}$$

with

$$\boldsymbol{\Omega}_3 \equiv (\boldsymbol{Q}_{m1} + \lambda\boldsymbol{I})\boldsymbol{X}_{1:m,1}\boldsymbol{M}' + \boldsymbol{X}_{1:m,1}\boldsymbol{M}'\boldsymbol{X}_{m+1,1}^\top \boldsymbol{X}_{m+1,1}(\boldsymbol{M}+\boldsymbol{M}').$$

Another application of (S.27) with

$$\boldsymbol{S} = \lambda^2 \boldsymbol{I}_{n_{m+1,1}}, \quad \boldsymbol{U} \equiv \boldsymbol{X}_{m+1,1}\boldsymbol{\Omega}_2, \quad \boldsymbol{T} = \boldsymbol{I}_{d_{\mathrm{A}}} \quad \text{and} \quad \boldsymbol{V} \equiv \boldsymbol{X}_{m+1,1}^\top$$

then gives the lower right $n_{m+1,1} \times n_{m+1,1}$ block of $\boldsymbol{Q}_{m1}^{-2}$ equalling

$$\lambda^{-2}\boldsymbol{I}_{n_{m+1,1}} - \lambda^{-4}\boldsymbol{X}_{m+1,1}\boldsymbol{\Omega}_2(\boldsymbol{I} + \lambda^{-2}\boldsymbol{X}_{m+1,1}^\top \boldsymbol{X}_{m+1,1}\boldsymbol{\Omega}_2)^{-1}\boldsymbol{X}_{m+1,1}^\top$$

and so

$$\begin{aligned}
&\frac{1}{n_{m+1,1}}\mathrm{tr}\Big(\text{lower right } n_{m+1,1} \times n_{m+1,1} \text{ block of } \boldsymbol{Q}_{m1}^{-2}\Big) \\
&\quad = \frac{1}{\lambda^2} - \frac{1}{n_{m+1,1}\lambda^4}\mathrm{tr}\Big((\boldsymbol{I} + \lambda^{-2}\boldsymbol{X}_{m+1,1}^\top \boldsymbol{X}_{m+1,1}\boldsymbol{\Omega}_2)^{-1}\boldsymbol{X}_{m+1,1}^\top \boldsymbol{X}_{m+1,1}\boldsymbol{\Omega}_2\Big) \\
&\quad = \frac{1}{\lambda^2} - \frac{1}{n_{m+1,1}\lambda^2}\mathrm{tr}\Big(\big\{\boldsymbol{\Omega}_2 + \lambda^2(\boldsymbol{X}_{m+1,1}^\top \boldsymbol{X}_{m+1,1})^{-1}\big\}^{-1}\boldsymbol{\Omega}_2\Big) \xrightarrow{P} \frac{1}{\lambda^2}.
\end{aligned}$$

By induction on m we then have, under (A5),

$$\left(\sum_{i=1}^m n_{i1}\right)^{-1} \mathrm{tr}(\boldsymbol{Q}_{m1}^{-2}) \xrightarrow{P} \frac{1}{\lambda^2} \quad \text{for all } m \in \mathbb{N}.$$

For higher m', similar arguments can be used to show that the summations in $\mathrm{tr}(\boldsymbol{Q}_{mm'}^{-2})$ lead to convergents that are analogous to those in the $d_{\mathrm{A}} = 1$, $n_{ii'} = n$ and $\boldsymbol{X}_{\mathrm{A}ii'} = \boldsymbol{1}_n$ case and Lemma 5(d) holds.

S.1.9.6 Proof of Lemma 6

First we prove Lemma 6 for $m = m' = 1$, for which the $\boldsymbol{Q}$ matrix reduces to

$$\boldsymbol{Q}_{11} = \boldsymbol{X}_{11}(\boldsymbol{M}+\boldsymbol{M}')\boldsymbol{X}_{11}^\top + \lambda\boldsymbol{I}.$$

Then, from Lemma 2,

$$\begin{aligned}
\overset{\star}{\boldsymbol{X}}{}_{11}^\top \boldsymbol{Q}_{11}^{-1}\overset{\star}{\boldsymbol{X}}_{11} &= \overset{\star}{\boldsymbol{X}}{}_{11}^\top \big\{\boldsymbol{X}_{11}(\boldsymbol{M}+\boldsymbol{M}')\boldsymbol{X}_{11}^\top + \lambda\boldsymbol{I}\big\}^{-1}\overset{\star}{\boldsymbol{X}}_{11} \\
&= (1/\lambda)\overset{\star}{\boldsymbol{X}}{}_{11}^\top\{\boldsymbol{I} - \boldsymbol{X}_{11}(\boldsymbol{X}_{11}^\top \boldsymbol{X}_{11})^{-1}\boldsymbol{X}_{11}^\top\}\overset{\star}{\boldsymbol{X}}_{11} \\
&\quad + \overset{\star}{\boldsymbol{X}}{}_{11}^\top \boldsymbol{X}_{11}(\boldsymbol{X}_{11}^\top \boldsymbol{X}_{11})^{-1}\{\boldsymbol{M}+\boldsymbol{M}' + \lambda(\boldsymbol{X}_{11}^\top \boldsymbol{X}_{11})^{-1}\}^{-1}(\boldsymbol{X}_{11}^\top \boldsymbol{X}_{11})^{-1}\boldsymbol{X}_{11}^\top \overset{\star}{\boldsymbol{X}}_{11}.
\end{aligned}$$

Hence

$$\tfrac{1}{n_{11}}\overset{\star}{\boldsymbol{X}}_{11}^{\top}\boldsymbol{Q}_{11}^{-1}\overset{\star}{\boldsymbol{X}}_{11} = (1/\lambda)\left\{\left(\tfrac{1}{n_{11}}\overset{\star}{\boldsymbol{X}}_{11}^{\top}\overset{\star}{\boldsymbol{X}}_{11}\right) - \left(\tfrac{1}{n_{11}}\overset{\star}{\boldsymbol{X}}_{11}^{\top}\boldsymbol{X}_{11}\right)\left(\tfrac{1}{n_{11}}\boldsymbol{X}_{11}^{\top}\boldsymbol{X}_{11}\right)^{-1}\left(\tfrac{1}{n_{11}}\boldsymbol{X}_{11}^{\top}\overset{\star}{\boldsymbol{X}}_{11}\right)\right\}$$

$$+\tfrac{1}{n_{11}}\left(\tfrac{1}{n_{11}}\overset{\star}{\boldsymbol{X}}_{11}^{\top}\boldsymbol{X}_{11}\right)\left(\tfrac{1}{n_{11}}\boldsymbol{X}_{11}^{\top}\boldsymbol{X}_{11}\right)^{-1}\left\{\boldsymbol{M}+\boldsymbol{M}'+\lambda(\boldsymbol{X}_{11}^{\top}\boldsymbol{X}_{11})^{-1}\right\}^{-1}$$

$$\times\left(\tfrac{1}{n_{11}}\boldsymbol{X}_{11}^{\top}\boldsymbol{X}_{11}\right)^{-1}\left(\tfrac{1}{n_{11}}\boldsymbol{X}_{11}^{\top}\overset{\star}{\boldsymbol{X}}_{11}\right)$$

$$\xrightarrow{P}(1/\lambda)\Big[E\Big(\overset{\star}{\boldsymbol{X}}{}_{\circ}^{\otimes 2}\Big) - E\Big(\overset{\star}{\boldsymbol{X}}_{\circ}\boldsymbol{X}_{\circ}^{\top}\Big)\Big\{E\Big(\boldsymbol{X}_{\circ}^{\otimes 2}\Big)\Big\}^{-1}E\Big(\boldsymbol{X}_{\circ}\overset{\star}{\boldsymbol{X}}{}_{\circ}^{\top}\Big)\Big]$$

$$= (1/\lambda)\left[\text{lower right } \overset{\star}{d}\times\overset{\star}{d} \text{ block of } \big\{E\big([\boldsymbol{X}_{\circ}\ \overset{\star}{\boldsymbol{X}}{}_{\circ}{}^{\top}]^{\otimes 2}\big)\big\}^{-1}\right]^{-1}.$$

Thus, Lemma 6 (a) holds for $m=m'=1$.

To establish Lemma 6(b) for $m=m'=1$ we apply Corollary 2.1(a) to obtain

$$\begin{aligned}\boldsymbol{X}_{11}^{\top}\boldsymbol{Q}_{11}^{-1}\overset{\star}{\boldsymbol{X}}_{11} &= \boldsymbol{X}_{11}^{\top}\big\{\boldsymbol{X}_{11}(\boldsymbol{M}+\boldsymbol{M}')\boldsymbol{X}_{11}^{\top}+\lambda\boldsymbol{I}\big\}^{-1}\overset{\star}{\boldsymbol{X}}_{11}\\ &= \{\boldsymbol{M}+\boldsymbol{M}'+\lambda(\boldsymbol{X}_{11}^{\top}\boldsymbol{X}_{11})^{-1}\}^{-1}\left(\tfrac{1}{n_{11}}\boldsymbol{X}_{11}^{\top}\boldsymbol{X}_{11}\right)^{-1}\tfrac{1}{n_{11}}\boldsymbol{X}_{11}^{\top}\overset{\star}{\boldsymbol{X}}_{11}\\ &\xrightarrow{P}(\boldsymbol{M}+\boldsymbol{M}')^{-1}\big\{E(\boldsymbol{X}_{\circ}^{\otimes 2})\big\}^{-1}E(\boldsymbol{X}_{\circ}\overset{\star}{\boldsymbol{X}}{}_{\circ}^{\top}).\end{aligned}$$

Therefore, Lemma 6 is proven for $m=m'=1$.

Next we prove that the lemma holds for all $m\geq 1$ and $m'=1$ via induction on m. Let $\boldsymbol{Q}_{m1}$ denote the $m'=1$ version of (S.3) and consider the partition of $\boldsymbol{Q}_{m+1,1}$ given by (S.34). Also let

$$\overset{\star}{\boldsymbol{X}}_{1:m,1}\equiv\underset{1\leq i\leq m}{\text{stack}}(\overset{\star}{\boldsymbol{X}}_{i1})\quad\text{and}\quad\overset{\star}{\boldsymbol{X}}_{1:m+1,1}\equiv\underset{1\leq i\leq m+1}{\text{stack}}(\overset{\star}{\boldsymbol{X}}_{i1})=\begin{bmatrix}\overset{\star}{\boldsymbol{X}}_{1:m,1}\\ \overset{\star}{\boldsymbol{X}}_{m+1,1}\end{bmatrix}.$$

Then $\overset{\star}{\boldsymbol{X}}{}_{1:m+1,1}^{\top}\boldsymbol{Q}_{m+1,1}^{-1}\overset{\star}{\boldsymbol{X}}_{1:m+1,1}$ equals

$$\begin{aligned}&\overset{\star}{\boldsymbol{X}}{}_{1:m,1}^{\top}(\boldsymbol{Q}_{m1}-\boldsymbol{R}_m\boldsymbol{S}_m^{-1}\boldsymbol{R}_m)^{-1}\overset{\star}{\boldsymbol{X}}_{1:m,1}\\ &-\overset{\star}{\boldsymbol{X}}{}_{1:m,1}^{\top}(\boldsymbol{Q}_{m1}-\boldsymbol{R}_m\boldsymbol{S}_m^{-1}\boldsymbol{R}_m)^{-1}\boldsymbol{R}_m\boldsymbol{S}_m^{-1}\overset{\star}{\boldsymbol{X}}_{m+1,1}\\ &-\overset{\star}{\boldsymbol{X}}{}_{m+1,1}^{\top}\boldsymbol{S}_m^{-1}\boldsymbol{R}_m^{\top}(\boldsymbol{Q}_{m1}-\boldsymbol{R}_m\boldsymbol{S}_m^{-1}\boldsymbol{R}_m)^{-1}\overset{\star}{\boldsymbol{X}}_{1:m,1}\\ &+\overset{\star}{\boldsymbol{X}}{}_{m+1,1}^{\top}(\boldsymbol{S}_m-\boldsymbol{R}_m^{\top}\boldsymbol{Q}_{m1}^{-1}\boldsymbol{R}_m)^{-1}\overset{\star}{\boldsymbol{X}}_{m+1,1}\\ =&\overset{\star}{\boldsymbol{X}}{}_{1:m,1}^{\top}\Big[\boldsymbol{Q}_{m1}-\boldsymbol{X}_{1:m,1}\boldsymbol{M}'\boldsymbol{X}_{m+1,1}^{\top}\big\{\boldsymbol{X}_{m+1,1}(\boldsymbol{M}+\boldsymbol{M}')\boldsymbol{X}_{m+1,1}^{\top}+\lambda\boldsymbol{I}\big\}^{-1}\boldsymbol{X}_{m+1,1}\boldsymbol{M}'\boldsymbol{X}_{1:m,1}^{\top}\Big]^{-1}\overset{\star}{\boldsymbol{X}}_{1:m,1}\\ &-\overset{\star}{\boldsymbol{X}}{}_{1:m,1}^{\top}\Big[\boldsymbol{Q}_{m1}-\boldsymbol{X}_{1:m,1}\boldsymbol{M}'\boldsymbol{X}_{m+1,1}^{\top}\big\{\boldsymbol{X}_{m+1,1}(\boldsymbol{M}+\boldsymbol{M}')\boldsymbol{X}_{m+1,1}^{\top}+\lambda\boldsymbol{I}\big\}^{-1}\boldsymbol{X}_{m+1,1}\boldsymbol{M}'\boldsymbol{X}_{1:m,1}^{\top}\Big]^{-1}\\ &\times\boldsymbol{X}_{1:m,1}\boldsymbol{M}'\boldsymbol{X}_{m+1,1}^{\top}\big\{\boldsymbol{X}_{m+1,1}(\boldsymbol{M}+\boldsymbol{M}')\boldsymbol{X}_{m+1,1}^{\top}+\lambda\boldsymbol{I}\big\}^{-1}\overset{\star}{\boldsymbol{X}}_{m+1,1}\\ &-\overset{\star}{\boldsymbol{X}}{}_{m+1,1}^{\top}\big\{\boldsymbol{X}_{m+1,1}(\boldsymbol{M}+\boldsymbol{M}')\boldsymbol{X}_{m+1,1}^{\top}+\lambda\boldsymbol{I}\big\}^{-1}\boldsymbol{X}_{m+1,1}\boldsymbol{M}'\boldsymbol{X}_{1:m,1}^{\top}\\ &\times\Big[\boldsymbol{Q}_{m1}-\boldsymbol{X}_{1:m,1}\boldsymbol{M}'\boldsymbol{X}_{m+1,1}^{\top}\big\{\boldsymbol{X}_{m+1,1}(\boldsymbol{M}+\boldsymbol{M}')\boldsymbol{X}_{m+1,1}^{\top}+\lambda\boldsymbol{I}\big\}^{-1}\boldsymbol{X}_{m+1,1}\boldsymbol{M}'\boldsymbol{X}_{1:m,1}^{\top}\Big]^{-1}\overset{\star}{\boldsymbol{X}}_{1:m,1}\\ &+\overset{\star}{\boldsymbol{X}}{}_{m+1,1}^{\top}\big\{\boldsymbol{X}_{m+1,1}(\boldsymbol{M}+\boldsymbol{M}')\boldsymbol{X}_{m+1,1}^{\top}-\boldsymbol{X}_{m+1,1}\boldsymbol{M}'\boldsymbol{X}_{1:m,1}^{\top}\boldsymbol{Q}_{m1}^{-1}\boldsymbol{X}_{1:m,1}\boldsymbol{M}'\boldsymbol{X}_{m+1,1}^{\top}+\lambda\boldsymbol{I}\big\}^{-1}\overset{\star}{\boldsymbol{X}}_{m+1,1}.\end{aligned}$$

Thus,

$$\overset{\star}{\boldsymbol{X}}{}^{\top}_{1:m+1,1}\boldsymbol{Q}^{-1}_{m+1,1}\overset{\star}{\boldsymbol{X}}_{1:m+1,1} = \mathfrak{T}_1 - \mathfrak{T}_2 - \mathfrak{T}_2^{\top} + \mathfrak{T}_3 + \mathfrak{T}_4$$

where

$$\begin{aligned}
\mathfrak{T}_1 &\equiv \overset{\star}{\boldsymbol{X}}{}^{\top}_{1:m,1}\big(\boldsymbol{Q}_{m1} + \boldsymbol{\Gamma}_1\big)^{-1}\overset{\star}{\boldsymbol{X}}_{1:m,1},\\
\mathfrak{T}_2 &= \overset{\star}{\boldsymbol{X}}{}^{\top}_{1:m,1}\big(\boldsymbol{Q}_{m1} + \boldsymbol{\Gamma}_1\big)^{-1}\boldsymbol{X}_{1:m,1}\boldsymbol{M}'\boldsymbol{\Gamma}_2(\boldsymbol{X}^{\top}_{m+1,1}\boldsymbol{X}_{m+1,1})^{-1}\boldsymbol{X}^{\top}_{m+1,1}\overset{\star}{\boldsymbol{X}}_{m+1,1}\\
\mathfrak{T}_3 &= (1/\lambda)\overset{\star}{\boldsymbol{X}}{}^{\top}_{m+1,1}\big\{\boldsymbol{I} - \boldsymbol{X}_{m+1,1}(\boldsymbol{X}^{\top}_{m+1,1}\boldsymbol{X}_{m+1,1})^{-1}\boldsymbol{X}^{\top}_{m+1,1}\big\}\overset{\star}{\boldsymbol{X}}_{m+1,1}\\
\mathfrak{T}_4 &= \overset{\star}{\boldsymbol{X}}{}^{\top}_{m+1,1}\boldsymbol{X}_{m+1,1}(\boldsymbol{X}^{\top}_{m+1,1}\boldsymbol{X}_{m+1,1})^{-1}\\
&\quad\times\big\{\boldsymbol{M} + \boldsymbol{M}' - \boldsymbol{M}'\boldsymbol{X}^{\top}_{1:m,1}\boldsymbol{Q}^{-1}_{m1}\boldsymbol{X}_{1:m,1}\boldsymbol{M}' + \lambda(\boldsymbol{X}^{\top}_{1:m,1}\boldsymbol{X}_{1:m,1})^{-1}\big\}^{-1}\\
&\quad\times(\boldsymbol{X}^{\top}_{m+1,1}\boldsymbol{X}_{m+1,1})^{-1}\boldsymbol{X}^{\top}_{m+1,1}\overset{\star}{\boldsymbol{X}}_{m+1,1},\\
\boldsymbol{\Gamma}_1 &\equiv \boldsymbol{X}_{1:m,1}\boldsymbol{M}'\boldsymbol{\Gamma}_2\boldsymbol{M}'\boldsymbol{X}^{\top}_{1:m,1} \quad\text{and}\quad \boldsymbol{\Gamma}_2 \equiv -\big\{\boldsymbol{M} + \boldsymbol{M}' + \lambda(\boldsymbol{X}^{\top}_{m+1,1}\boldsymbol{X}_{m+1,1})^{-1}\big\}^{-1}.
\end{aligned}$$

Application of Woodbury's matrix identity (S.27) to $\big(\boldsymbol{Q}_{m1} + \boldsymbol{\Gamma}_1\big)^{-1}$ with $\boldsymbol{S} = \boldsymbol{Q}_{m1}$, $\boldsymbol{U} = \boldsymbol{X}_{1:m,1}\boldsymbol{M}'$, $\boldsymbol{V} = \boldsymbol{M}'\boldsymbol{X}^{\top}_{1:m,1}$ and $\boldsymbol{T} = \boldsymbol{\Gamma}_2$ leads to

$$\begin{aligned}
\mathfrak{T}_1 &= \overset{\star}{\boldsymbol{X}}{}^{\top}_{1:m,1}\boldsymbol{Q}^{-1}_{m1}\overset{\star}{\boldsymbol{X}}_{1:m,1}\\
&\quad-\overset{\star}{\boldsymbol{X}}{}^{\top}_{1:m,1}\boldsymbol{Q}^{-1}_{m1}\boldsymbol{X}_{1:m,1}\boldsymbol{M}'\big\{\boldsymbol{\Gamma}_2 + \boldsymbol{M}'\boldsymbol{X}^{\top}_{1:m,1}\boldsymbol{Q}^{-1}_{m1}\boldsymbol{X}_{1:m,1}\boldsymbol{M}'\big\}^{-1}\boldsymbol{M}'\boldsymbol{X}^{\top}_{1:m,1}\boldsymbol{Q}^{-1}_{m1}\overset{\star}{\boldsymbol{X}}_{1:m,1}\\
&= \frac{n_{11} + \ldots + n_{m1}}{\lambda}\Big[\text{lower right } \overset{\star}{d} \times \overset{\star}{d} \text{ block of } \big\{E\big([\boldsymbol{X}_{\circ}\,\overset{\star}{\boldsymbol{X}}_{\circ}{}^{\top}]^{\otimes 2}\big)\big\}^{-1}\Big]^{-1}\{1 + o_P(1)\}
\end{aligned}$$

by Lemma 6 and the inductive hypothesis. Similarly, the first three factors of $\mathfrak{T}_2$ are

$$\begin{aligned}
\overset{\star}{\boldsymbol{X}}{}^{\top}_{1:m,1}\big(\boldsymbol{Q}_{m1} + \boldsymbol{\Gamma}_1\big)^{-1}\boldsymbol{X}_{1:m,1} &= \overset{\star}{\boldsymbol{X}}{}^{\top}_{1:m,1}\boldsymbol{Q}^{-1}_{m1}\boldsymbol{X}_{1:m,1} - \overset{\star}{\boldsymbol{X}}{}^{\top}_{1:m,1}\boldsymbol{Q}^{-1}_{m1}\boldsymbol{X}_{1:m,1}\\
&\quad\times\boldsymbol{M}'\big\{\boldsymbol{\Gamma}_2 + \boldsymbol{M}'\boldsymbol{X}^{\top}_{1:m,1}\boldsymbol{Q}^{-1}_{m1}\boldsymbol{X}_{1:m,1}\boldsymbol{M}'\big\}^{-1}\boldsymbol{M}'\boldsymbol{X}^{\top}_{1:m,1}\boldsymbol{Q}^{-1}_{m1}\boldsymbol{X}_{1:m,1}
\end{aligned}$$

which soon leads to $\mathfrak{T}_2$ having all entries being $O_P(m)$. Next, we have

$$\mathfrak{T}_3 = \big(n_{m+1,1}/\lambda\big)\Big[\text{lower right } \overset{\star}{d} \times \overset{\star}{d} \text{ block of } \big\{E\big([\boldsymbol{X}_{\circ}\,\overset{\star}{\boldsymbol{X}}_{\circ}{}^{\top}]^{\otimes 2}\big)\big\}^{-1}\Big]^{-1}\{1 + o_P(1)\}$$

and $\mathfrak{T}_4$ having all entries being $O_P(1)$. Combining these results for $\mathfrak{T}_1$, $\mathfrak{T}_2$, $\mathfrak{T}_3$ and $\mathfrak{T}_4$ leads to

$$\left(\sum_{i=1}^{m} n_{i1}\right)^{-1}\overset{\star}{\boldsymbol{X}}{}^{\top}_{1:m+1,1}\boldsymbol{Q}^{-1}_{m+1,1}\overset{\star}{\boldsymbol{X}}_{1:m+1,1} \xrightarrow{P} \frac{1}{\lambda}\Big[\text{lower right } \overset{\star}{d} \times \overset{\star}{d} \text{ block of } \big\{E\big([\boldsymbol{X}_{\circ}\,\overset{\star}{\boldsymbol{X}}_{\circ}{}^{\top}]^{\otimes 2}\big)\big\}^{-1}\Big]^{-1}$$

which proves Lemma 6 (a) for all $m \in \mathbb{N}$ and $m' = 1$. The proof of Lemma 6 (b) for all $m \in \mathbb{N}$ and $m' = 1$ involves a similar set of arguments.

Now we turn our attention to establishing Lemma 6 (a) for $m = 1$ and $m' = 2$. Noting that

$$\boldsymbol{Q}_{12} = \begin{bmatrix} \boldsymbol{X}_{11} & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{X}_{12} \end{bmatrix}\begin{bmatrix} \boldsymbol{M} + \boldsymbol{M}' & \boldsymbol{M} \\ \boldsymbol{M} & \boldsymbol{M} + \boldsymbol{M}' \end{bmatrix}\begin{bmatrix} \boldsymbol{X}_{11} & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{X}_{12} \end{bmatrix}^{\top} + \lambda\boldsymbol{I}.$$

and

$$\overset{\star}{\boldsymbol{X}} = \begin{bmatrix} \overset{\star}{\boldsymbol{X}}_{11} \\ \overset{\star}{\boldsymbol{X}}_{12} \end{bmatrix} = \begin{bmatrix} \overset{\star}{\boldsymbol{X}}_{11} & \boldsymbol{O} \\ \boldsymbol{O} & \overset{\star}{\boldsymbol{X}}_{12} \end{bmatrix}\begin{bmatrix} \boldsymbol{I}_d \\ \boldsymbol{I}_d \end{bmatrix}$$

we have

$$\overset{\star}{\boldsymbol{X}}^\top \boldsymbol{Q}_{12}^{-1} \overset{\star}{\boldsymbol{X}} = \begin{bmatrix} \boldsymbol{I}_{\overset{\star}{d}} \\ \boldsymbol{I}_{\overset{\star}{d}} \end{bmatrix}^\top \begin{bmatrix} \overset{\star}{\boldsymbol{X}}_{11} & \boldsymbol{O} \\ \boldsymbol{O} & \overset{\star}{\boldsymbol{X}}_{12} \end{bmatrix}^\top \left\{ \begin{bmatrix} \boldsymbol{X}_{11} & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{X}_{12} \end{bmatrix} \begin{bmatrix} \boldsymbol{M} + \boldsymbol{M}' & \boldsymbol{M} \\ \boldsymbol{M} & \boldsymbol{M} + \boldsymbol{M}' \end{bmatrix} \begin{bmatrix} \boldsymbol{X}_{11} & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{X}_{12} \end{bmatrix}^\top + \lambda \boldsymbol{I} \right\}^{-1}$$

$$\times \begin{bmatrix} \overset{\star}{\boldsymbol{X}}_{11} & \boldsymbol{O} \\ \boldsymbol{O} & \overset{\star}{\boldsymbol{X}}_{12} \end{bmatrix} \begin{bmatrix} \boldsymbol{I}_{\overset{\star}{d}} \\ \boldsymbol{I}_{\overset{\star}{d}} \end{bmatrix}$$

$$= \mathfrak{T}_5 + \mathfrak{T}_6$$

where

$$\mathfrak{T}_5 = (1/\lambda) \begin{bmatrix} \boldsymbol{I}_{\overset{\star}{d}} \\ \boldsymbol{I}_{\overset{\star}{d}} \end{bmatrix}^\top \begin{bmatrix} \overset{\star}{\boldsymbol{X}}_{11} & \boldsymbol{O} \\ \boldsymbol{O} & \overset{\star}{\boldsymbol{X}}_{12} \end{bmatrix}^\top$$

$$\times \left\{ \boldsymbol{I} - \begin{bmatrix} \boldsymbol{X}_{11} & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{X}_{12} \end{bmatrix} \left(\begin{bmatrix} \boldsymbol{X}_{11} & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{X}_{12} \end{bmatrix}^\top \begin{bmatrix} \boldsymbol{X}_{11} & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{X}_{12} \end{bmatrix} \right)^{-1} \begin{bmatrix} \boldsymbol{X}_{11} & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{X}_{12} \end{bmatrix}^\top \right\} \begin{bmatrix} \overset{\star}{\boldsymbol{X}}_{11} & \boldsymbol{O} \\ \boldsymbol{O} & \overset{\star}{\boldsymbol{X}}_{12} \end{bmatrix} \begin{bmatrix} \boldsymbol{I}_{\overset{\star}{d}} \\ \boldsymbol{I}_{\overset{\star}{d}} \end{bmatrix}$$

$$= \frac{n_{11}}{\lambda} \left\{ \left(\frac{1}{n_{11}} \overset{\star}{\boldsymbol{X}}_{11}^\top \overset{\star}{\boldsymbol{X}}_{11} \right) - \left(\frac{1}{n_{11}} \overset{\star}{\boldsymbol{X}}_{11}^\top \boldsymbol{X}_{11} \right) \left(\frac{1}{n_{11}} \boldsymbol{X}_{11}^\top \boldsymbol{X}_{11} \right)^{-1} \left(\frac{1}{n_{11}} \boldsymbol{X}_{11}^\top \overset{\star}{\boldsymbol{X}}_{11} \right) \right\}$$

$$+ \frac{n_{12}}{\lambda} \left\{ \left(\frac{1}{n_{12}} \overset{\star}{\boldsymbol{X}}_{12}^\top \overset{\star}{\boldsymbol{X}}_{12} \right) - \left(\frac{1}{n_{12}} \overset{\star}{\boldsymbol{X}}_{11}^\top \boldsymbol{X}_{12} \right) \left(\frac{1}{n_{12}} \boldsymbol{X}_{12}^\top \boldsymbol{X}_{12} \right)^{-1} \left(\frac{1}{n_{12}} \boldsymbol{X}_{12}^\top \overset{\star}{\boldsymbol{X}}_{12} \right) \right\}$$

$$= \frac{n_{11} + n_{12}}{\lambda} \left[\text{lower right } \overset{\star}{d} \times \overset{\star}{d} \text{ block of } \left\{ E\big([\boldsymbol{X}_\circ \, \overset{\star}{\boldsymbol{X}}_\circ^\top]^{\otimes 2} \big) \right\}^{-1} \right]^{-1} \{1 + o_P(1)\}.$$

and

$$\mathfrak{T}_6 = \begin{bmatrix} \big(\overset{\star}{\boldsymbol{X}}_{11}^\top \boldsymbol{X}_{11}\big) \big(\boldsymbol{X}_{11} \boldsymbol{X}_{11}\big)^{-1} \\ \big(\overset{\star}{\boldsymbol{X}}_{12}^\top \boldsymbol{X}_{12}\big) \big(\boldsymbol{X}_{12} \boldsymbol{X}_{12}\big)^{-1} \end{bmatrix} \begin{bmatrix} \boldsymbol{M} + \boldsymbol{M}' + \lambda (\boldsymbol{X}_{11}^\top \boldsymbol{X}_{11})^{-1} & \boldsymbol{M} \\ \boldsymbol{M} & \boldsymbol{M} + \boldsymbol{M}' + \lambda (\boldsymbol{X}_{12}^\top \boldsymbol{X}_{12})^{-1} \end{bmatrix}^{-1}$$

$$\times \begin{bmatrix} \big(\overset{\star}{\boldsymbol{X}}_{11}^\top \boldsymbol{X}_{11}\big) \big(\boldsymbol{X}_{11} \boldsymbol{X}_{11}\big)^{-1} \\ \big(\overset{\star}{\boldsymbol{X}}_{12}^\top \boldsymbol{X}_{12}\big) \big(\boldsymbol{X}_{12} \boldsymbol{X}_{12}\big)^{-1} \end{bmatrix}.$$

Since each of the entries of $\mathfrak{T}_6$ is $O_P(1)$ we have

$$\frac{1}{n_{11} + n_{12}} \overset{\star}{\boldsymbol{X}}^\top \boldsymbol{Q}_{12}^{-1} \overset{\star}{\boldsymbol{X}} \xrightarrow{P} (1/\lambda) \left[\text{lower right } \overset{\star}{d} \times \overset{\star}{d} \text{ block of } \left\{ E\big([\boldsymbol{X}_\circ \, \overset{\star}{\boldsymbol{X}}_\circ^\top]^{\otimes 2} \big) \right\}^{-1} \right]^{-1}$$

which verifies Lemma 6(a) for $m = 1$ and $m' = 2$. An analogous pattern continues for higher m and m' which leads to the Lemma 6(a) result holding generally.

For Lemma 6(b) in the $m = 1$ and $m' = 2$ case we instead have, using Corollary 2.1(a) and

Lemma 4,

$$
\begin{aligned}
\boldsymbol{X}^\top \boldsymbol{Q}_{12}^{-1} \overset{\bigstar}{\boldsymbol{X}} &= \begin{bmatrix} \boldsymbol{I}_d \\ \boldsymbol{I}_d \end{bmatrix}^\top \begin{bmatrix} \boldsymbol{X}_{11} & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{X}_{12} \end{bmatrix}^\top \left\{ \begin{bmatrix} \boldsymbol{X}_{11} & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{X}_{12} \end{bmatrix} \begin{bmatrix} \boldsymbol{M}+\boldsymbol{M}' & \boldsymbol{M} \\ \boldsymbol{M} & \boldsymbol{M}+\boldsymbol{M}' \end{bmatrix} \begin{bmatrix} \boldsymbol{X}_{11} & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{X}_{12} \end{bmatrix}^\top + \lambda \boldsymbol{I} \right\}^{-1} \\
&\quad \times \begin{bmatrix} \overset{\bigstar}{\boldsymbol{X}}_{11} & \boldsymbol{O} \\ \boldsymbol{O} & \overset{\bigstar}{\boldsymbol{X}}_{12} \end{bmatrix} \begin{bmatrix} \boldsymbol{I}_{\overset{\bigstar}{d}} \\ \boldsymbol{I}_{\overset{\bigstar}{d}} \end{bmatrix} \\
&= \begin{bmatrix} \boldsymbol{I}_d \\ \boldsymbol{I}_d \end{bmatrix}^\top \begin{bmatrix} \boldsymbol{M}+\boldsymbol{M}'+\lambda(\boldsymbol{X}_{11}^\top \boldsymbol{X}_{11})^{-1} & \boldsymbol{M} \\ \boldsymbol{M} & \boldsymbol{M}+\boldsymbol{M}'+\lambda(\boldsymbol{X}_{12}^\top \boldsymbol{X}_{12})^{-1} \end{bmatrix}^{-1} \\
&\quad \times \begin{bmatrix} \left(\frac{1}{n_{11}} \boldsymbol{X}_{11}^\top \boldsymbol{X}_{11}\right)^{-1} \left(\frac{1}{n_{11}} \boldsymbol{X}_{11}^\top \overset{\bigstar}{\boldsymbol{X}}_{11}\right) \\ \left(\frac{1}{n_{12}} \boldsymbol{X}_{11}^\top \boldsymbol{X}_{12}\right)^{-1} \left(\frac{1}{n_{12}} \boldsymbol{X}_{12}^\top \overset{\bigstar}{\boldsymbol{X}}_{12}\right) \end{bmatrix} \\
&\xrightarrow{P} \begin{bmatrix} \boldsymbol{I}_d \\ \boldsymbol{I}_d \end{bmatrix}^\top \begin{bmatrix} \boldsymbol{M}+\boldsymbol{M}' & \boldsymbol{M} \\ \boldsymbol{M} & \boldsymbol{M}+\boldsymbol{M}' \end{bmatrix}^{-1} \begin{bmatrix} \boldsymbol{I}_d \\ \boldsymbol{I}_d \end{bmatrix} \left\{E(\boldsymbol{X}_\circ^{\otimes 2})\right\}^{-1} E(\boldsymbol{X}_\circ \overset{\bigstar}{\boldsymbol{X}}_\circ^\top) \\
&= (\boldsymbol{M} + \tfrac{1}{2}\boldsymbol{M}')^{-1} \left\{E(\boldsymbol{X}_\circ^{\otimes 2})\right\}^{-1} E(\boldsymbol{X}_\circ \overset{\bigstar}{\boldsymbol{X}}_\circ^\top)
\end{aligned}
$$

which verifies Lemma 6(b) for $m = 1$ and $m' = 2$.

For general m and m', note that the behavior of $\boldsymbol{X}^\top \boldsymbol{Q}_{mm'}^{-1} \overset{\bigstar}{\boldsymbol{X}}$ mimics that of the $\boldsymbol{X}^\top \boldsymbol{Q}_{mm'}^{-1} \boldsymbol{X}$ special case, with the $\left\{E(\boldsymbol{X}_\circ^{\otimes 2})\right\}^{-1} E(\boldsymbol{X}_\circ \overset{\bigstar}{\boldsymbol{X}}_\circ^\top)$ factor being the only difference in the convergence in probability limit. The summations that provide the Lemma 5(a) result have analogous behaviors in this extended case and lead to Lemma 6(b) holding generally.

Additional References

Hall, P. & Heyde, C. C. (1980). *Martingale Central Limit Theory and Its Application.* New York: Academic Press.

Jiang, J., Jiang, W., Paul, D., Zhang, Y. & Zhao, H. (2023). High-dimensional asymptotic behaviour of inference based on GWAS summary statistic. *Statistica Sinica*, **33**, 1555–1576.

Magnus, J. R. & Neudecker, H. (1999). *Matrix Differential Calculus with Applications in Statistics and Econometrics, Revised Edition.* Chichester U.K.: Wiley.

Wand, M. P. (2002). Vector differential calculus in statistics. *The American Statistician*, **56**, 55–62.

Woodbury, M. A. (1950). Inverting modified matrices. Statistical Research Group, Memorandum Report Number 42, Princeton University, Princeton, New Jersey, U.S.A.