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*

8th September 2024

Abstract

We obtain an asymptotic normality result that reveals the precise asymptotic behavior of the maximum likelihood estimators of parameters for a very general class of linear mixed models containing cross random effects. In achieving the result, we overcome theoretical difficulties that arise from random effects being crossed as opposed to the simpler nested random effects case. Our new theory is for a class of Gaussian response linear mixed models which includes crossed random slopes that partner arbitrary multivariate predictor effects and does not require the cell counts to be balanced. Statistical utilities include confidence interval construction, Wald hypothesis test and sample size calculations.
Keywords: Asymptotic normality, maximum likelihood estimation, sample size calculations.

1 Introduction

Linear mixed models with crossed random effects are useful for the analysis of regression-type data that are cross-classified according to two or more grouping mechanisms. Baayen *et al.* (2008), for example, use the terms *subjects* and *items* for groupings that are typical in psychology studies. Specific examples discussed in Baayen *et al.* (2008) have subjects corresponding to human participants in a psycholinguistic experiment and items corresponding to words in a particular language. Gao & Owen (2020) and Ghosh *et al.* (2022) is concerned with electronic commerce and related applications involving crossed random effects, and is such that subjects and items correspond to customers and products.

Despite the widespread use of linear mixed models with crossed random effects, theory concerning the asymptotic behaviors of model parameter estimators is scant. This is largely due to the complicated mathematical forms that arise from random effects being crossed. Unlike the nested random effects case, the marginal covariance matrix of the response vector does not have a block diagonal form, which makes theoretical analyses significantly more challenging. For Gaussian response linear mixed models with nested random effects precise asymptotics are relatively straightforward as conveyed by, for example, Section 3.5 of McCulloch *et al.* (2008). Recently Jiang *et al.* (2022) obtained a precise asymptotic normality result for the joint distribution of all model parameters in a generalized linear mixed model with nested random effects. In this article we derive an analogous result for Gaussian response linear mixed models with crossed random effects.

Some early contributions to asymptotic theory for linear mixed models with crossed random effects structures are Hartley & Rao (1967) and Miller (1977). Indeed, the second example in Section 4 of Miller (1977) corresponds to a special case of the class of linear mixed models considered in the present article when his c_{ij} term is omitted. Further details concerning this example are in Sections 6.1 and 6.2 of Miller (1973), and includes an expression for the asymptotic covariance matrix of the maximum likelihood estimator of the vector of variance parameters. Asymptotic normality of the maximum likelihood estimators is also established in Miller (1973, 1977). However, the explicit results in these seminal articles are confined to balanced linear mixed models that are devoid of predictor data. Jiang (1996) focused on restricted maximum likelihood (REML) estimation of variance parameters in a wide class of linear mixed models that include those containing crossed random effects and obtained conditions under which asymptotic normality of the REML estimators hold. The results in Jiang (1996) are expressed in terms of generic Fisher information matrices rather than the explicit asymptotic forms provided by Jiang *et al.* (2022). Lyu, Sisson & Welsh (2024) is a recent article that is also concerned asymptotic normality of estimators in a crossed random effects setting. Connections between Lyu *et al.* (2024) and this paper are described below.

In this article we obtain precise asymptotics, in a similar vein to those of Jiang *et al.* (2022), for Gaussian response linear mixed models with crossed random effects. Our results apply to a wide class of situations that include unbalanced designs, predictor data and multivariate crossed random effects. They reveal that asymptotic covariance matrices of the estimators parameter vectors are quite similar to those that arise for nested random effects despite inherent differences due to effects being crossed. For example, the estimates of fixed effects parameters that are unaccompanied by random effects have the same asymptotic variances regardless of whether the model contains nested or crossed random effects. However, as we shall see, the pathway towards establishing such results for the crossed random effects case is much longer and involved.

The majority of the research in this article was done concurrently with and independently of the Lyu *et al.* (2024) research and we became aware of their article after devising Result 1. The linear mixed model treated by Lyu *et al.* (2024) does not assume that the responses are Gaussian. They also include a random interaction term, which our model does not have. In the case of Gaussian responses and additive crossed random effects, our main result extends the theoretical findings of Lyu *et al.* (2024) in the following two ways: (1) multivariate random slopes are included and (2) unbalanced cell counts are accommodated. Each of (1) and (2) are quite important in practice, but require lengthy matrix algebraic and convergence in probability arguments since the deterministic Kronecker product forms used in Miller (1973) and Lyu *et al.* (2024) no longer apply.

Contemporary data sets for which linear mixed models with crossed random effects provide a useful vehicle for analysis vastly differ in terms of the density of the observations. For some applications, the cell counts arising from subject/item cross-classification are all non-zero. As an example, the illustration given in Section 6 of Menictas *et al.* (2023) for the U.S. National Education Longitudinal Study has $8,564 \times 24 = 205,488$ cells with a few observations per cell. Other data sets, such as those that motivate Ghosh *et al.* (2022), have total number of observations much lower than the number of cells. In this article we focus on dense data situations where the cell counts are non-zero and growing in our asymptotic analyses. Relaxation to various sparse data situations is certainly of interest but, with conciseness and closure in mind, this is left aside in this article's theoretical study.

Generalized linear mixed models with crossed random effects are particularly challenging theoretically and it was not until Jiang (2013) that a consistency proof was established. As pointed out at the end of Section 4.5.7 of Jiang & Nguyen (2021), there is no existing asymptotic distribution theory for maximum likelihood estimators in the non-Gaussian version of such models. We only treat the Gaussian version here.

The linear mixed model with crossed random effects that we study is described in Section 2, as well as maximum likelihood estimation of the model parameters. An asymptotic normality result that reveals the precise asymptotic behavior of all maximum likelihood estimators is given in Section 3. A key finding in Section 3 is that the leading terms are very similar to those arising in nested random effects models. In Section 4 we provide some heuristic arguments that help explain these similarities. Section 5 discusses statistical utility of the new theory. Some concluding remarks are made in Section 6. An online supplement provides derivational details of central result.

2 Model Description and Maximum Likelihood Estimation

Consider the following crossed random effects linear mixed models:

$$\begin{gathered}\boldsymbol{Y}_{ii'}|\boldsymbol{U}_i,\boldsymbol{U}'_{i'},\boldsymbol{X}_{\mathrm{A}ii'},\boldsymbol{X}_{\mathrm{B}ii'}\overset{\text{ind.}}{\sim}N\big((\boldsymbol{\beta}_{\mathrm{A}}^0+\boldsymbol{U}_i+\boldsymbol{U}'_{i'})^T\boldsymbol{X}_{\mathrm{A}ii'}+(\boldsymbol{\beta}_{\mathrm{B}}^0)^T\boldsymbol{X}_{\mathrm{B}ii'},(\sigma^2)^0\boldsymbol{I}\big),\\ \boldsymbol{U}_i\overset{\text{ind.}}{\sim}N(\boldsymbol{0},\boldsymbol{\Sigma}^0),\quad 1\le i\le m,\quad \boldsymbol{U}'_{i'}\overset{\text{ind.}}{\sim}N(\boldsymbol{0},(\boldsymbol{\Sigma}')^0),\quad 1\le i'\le m'\end{gathered}\tag{1}$$

where here, and throughout this article, $\stackrel{\text{ind.}}{\sim}$ stands for "independently distributed as".

The dimensions of the matrices in (1) are:

$$\boldsymbol{Y}_{ii'} \text{ is } n_{ii'}\times 1,\ \boldsymbol{X}_{\mathrm{A}ii'} \text{ is } n_{ii'}\times d_{\mathrm{A}},\ \boldsymbol{\beta}^0_{\mathrm{A}} \text{ is } d_{\mathrm{A}}\times 1,\ \boldsymbol{U}_i \text{ is } d_{\mathrm{A}}\times 1,\ \boldsymbol{U}'_{i'} \text{ is } d_{\mathrm{A}}\times 1$$
$$\boldsymbol{X}_{\mathrm{B}ii'} \text{ is } n_{ii'}\times d_{\mathrm{B}},\ \boldsymbol{\beta}^0_{\mathrm{B}} \text{ is } d_{\mathrm{B}}\times 1,\ \boldsymbol{\Sigma}^0 \text{ is } d_{\mathrm{A}}\times d_{\mathrm{A}} \text{ and } (\boldsymbol{\Sigma}')^0 \text{ is } d_{\mathrm{A}}\times d_{\mathrm{A}}.$$

Here $n_{ii'}$ is the number of response measurements in the (i,i')th cell. If $n_{ii'}=0$ then each of $\boldsymbol{Y}_{ii'}$, $\boldsymbol{X}_{\mathrm{A}ii'}$ and $\boldsymbol{X}_{\mathrm{B}ii'}$ are null. The focus of this article is the precise asymptotic properties of the maximum likelihood estimators of the model parameters when m, m' and the $n_{ii'}$ all diverge to ∞. Therefore, from now onwards, we assume that $n_{ii'}>0$ for all $1\le i\le m$ and $1\le i'\le m'$.

In (1) we assume that the $\boldsymbol{X}_{\mathrm{A}ii'}$ are independent and identically distributed random vectors having the same distribution as $\boldsymbol{X}_{\mathrm{A}\circ}$. Similarly, the $\boldsymbol{X}_{\mathrm{B}ii'}$ are independent and identically distributed random vectors having the same distribution as $\boldsymbol{X}_{\mathrm{B}\circ}$.

The following matrix assembly notation is useful for describing the maximum likelihood estimators and their asymptotic properties. Firstly,

$$\underset{1\le i\le d}{\mathrm{stack}}(\boldsymbol{A}_i)\equiv\begin{bmatrix}\boldsymbol{A}_1\\ \vdots\\ \boldsymbol{A}_d\end{bmatrix}\quad\text{and}\quad \underset{1\le i\le d}{\mathrm{blockdiag}}(\boldsymbol{A}_i)\equiv\begin{bmatrix}\boldsymbol{A}_1 & \boldsymbol{O} & \cdots & \boldsymbol{O}\\ \boldsymbol{O} & \boldsymbol{A}_2 & \cdots & \boldsymbol{O}\\ \vdots & \vdots & \ddots & \vdots\\ \boldsymbol{O} & \boldsymbol{O} & \cdots & \boldsymbol{A}_d\end{bmatrix}$$

for matrices $\boldsymbol{A}_1,\ldots,\boldsymbol{A}_d$. The first of these definitions require that $\boldsymbol{A}_i$, $1\le i\le d$, each have the same number of columns. Next, define

$$\underset{1\le i,\underset{\sim}{i}\le d}{\mathrm{blockmatrix}}(\boldsymbol{B}_{i\underset{\sim}{i}})\equiv\begin{bmatrix}\boldsymbol{B}_{11} & \cdots & \boldsymbol{B}_{1d}\\ \vdots & \ddots & \vdots\\ \boldsymbol{B}_{d1} & \cdots & \boldsymbol{B}_{dd}\end{bmatrix}$$

for matrices $\boldsymbol{B}_{i\underset{\sim}{i}}$, $1\le i,\underset{\sim}{i}\le d$, each having the same numbers of rows and columns. If we then define

$$n_{\bullet\bullet}\equiv\sum_{i=1}^{m}\sum_{i'=1}^{m'}n_{ii'},\quad \boldsymbol{Y}\equiv\underset{1\le i\le m}{\mathrm{stack}}\left\{\underset{1\le i'\le m'}{\mathrm{stack}}(\boldsymbol{Y}_{ii'})\right\},$$

$$\boldsymbol{X}_{\mathrm{A}}\equiv\underset{1\le i\le m}{\mathrm{stack}}\left\{\underset{1\le i'\le m'}{\mathrm{stack}}(\boldsymbol{X}_{\mathrm{A}ii'})\right\}\quad\text{and}\quad \boldsymbol{X}_{\mathrm{B}}\equiv\underset{1\le i\le m}{\mathrm{stack}}\left\{\underset{1\le i'\le m'}{\mathrm{stack}}(\boldsymbol{X}_{\mathrm{B}ii'})\right\}\tag{2}$$

then standard manipulations show that

$$\boldsymbol{Y}|\boldsymbol{X}_{\mathrm{A}},\boldsymbol{X}_{\mathrm{B}}\sim N\Big(\boldsymbol{X}_{\mathrm{A}}\boldsymbol{\beta}^0_{\mathrm{A}}+\boldsymbol{X}_{\mathrm{B}}\boldsymbol{\beta}^0_{\mathrm{B}},\boldsymbol{V}\big(\boldsymbol{\Sigma}^0,(\boldsymbol{\Sigma}')^0,(\sigma^2)^0\big)\Big)$$

where

$$\begin{aligned}\boldsymbol{V}(\boldsymbol{\Sigma},\boldsymbol{\Sigma}',\sigma^2)\equiv\;&\underset{1\le i\le m}{\mathrm{blockdiag}}\left\{\underset{1\le i',\underset{\sim}{i}'\le m'}{\mathrm{blockmatrix}}(\boldsymbol{X}_{\mathrm{A}ii'}\boldsymbol{\Sigma}\boldsymbol{X}^T_{\mathrm{A}i\underset{\sim}{i}'})\right\}\\ &+\underset{1\le i,\underset{\sim}{i}\le m}{\mathrm{blockmatrix}}\left\{\underset{1\le i'\le m'}{\mathrm{blockdiag}}(\boldsymbol{X}_{\mathrm{A}ii'}\boldsymbol{\Sigma}'\boldsymbol{X}^T_{\mathrm{A}\underset{\sim}{i}i'})\right\}+\sigma^2\boldsymbol{I}_{n_{\bullet\bullet}}.\end{aligned}\tag{3}$$

Therefore, the conditional log-likelihood is

$$\begin{aligned}\ell(\boldsymbol{\beta}_{\mathrm{A}},\boldsymbol{\beta}_{\mathrm{B}},\boldsymbol{\Sigma},\boldsymbol{\Sigma}',\sigma^2)=&-\tfrac{1}{2}n_{\bullet\bullet}\log(2\pi)-\tfrac{1}{2}\log\big|\boldsymbol{V}(\boldsymbol{\Sigma},\boldsymbol{\Sigma}',\sigma^2)\big|\\ &-\tfrac{1}{2}(\boldsymbol{Y}-\boldsymbol{X}_{\mathrm{A}}\boldsymbol{\beta}_{\mathrm{A}}-\boldsymbol{X}_{\mathrm{B}}\boldsymbol{\beta}_{\mathrm{B}})^T\boldsymbol{V}(\boldsymbol{\Sigma},\boldsymbol{\Sigma}',\sigma^2)^{-1}(\boldsymbol{Y}-\boldsymbol{X}_{\mathrm{A}}\boldsymbol{\beta}_{\mathrm{A}}-\boldsymbol{X}_{\mathrm{B}}\boldsymbol{\beta}_{\mathrm{B}}).\end{aligned}\tag{4}$$

The maximum likelihood estimator of $\big(\boldsymbol{\beta}^0_{\mathrm{A}},\boldsymbol{\beta}^0_{\mathrm{B}},\boldsymbol{\Sigma}^0,(\boldsymbol{\Sigma}')^0,(\sigma^2)^0\big)$ is

$$(\widehat{\boldsymbol{\beta}}_{\mathrm{A}},\widehat{\boldsymbol{\beta}}_{\mathrm{B}},\widehat{\boldsymbol{\Sigma}},\widehat{\boldsymbol{\Sigma}}',\widehat{\sigma}^2)\equiv\underset{\boldsymbol{\beta}_{\mathrm{A}},\boldsymbol{\beta}_{\mathrm{B}},\boldsymbol{\Sigma},\boldsymbol{\Sigma}',\sigma^2}{\mathrm{argmax}}\ \ell(\boldsymbol{\beta}_{\mathrm{A}},\boldsymbol{\beta}_{\mathrm{B}},\boldsymbol{\Sigma},\boldsymbol{\Sigma}',\sigma^2).$$

3 Asymptotic Normality Result

We now present the article's main centerpiece: an asymptotic normality result that reveals the precise asymptotic behavior of the maximum likelihood estimation of $(\widehat{\boldsymbol{\beta}}_{\rm A}, \widehat{\boldsymbol{\beta}}_{\rm B}, \widehat{\boldsymbol{\Sigma}}, \widehat{\boldsymbol{\Sigma}}', \widehat{\sigma}^2)$ for data corresponding to (1).

Define

$$n \equiv \frac{n_{\bullet\bullet}}{mm'} = \text{average of the within-cell sample sizes}$$

and

$$\boldsymbol{C}_{\boldsymbol{\beta}_{\rm B}} \equiv \text{lower right } d_{\rm B} \times d_{\rm B} \text{ block of } \left\{E\left(\boldsymbol{X}_\circ \boldsymbol{X}_\circ^T\right)\right\}^{-1} \text{ where } \boldsymbol{X}_\circ \equiv \begin{bmatrix} \boldsymbol{X}_{\rm A\circ} \\ \boldsymbol{X}_{\rm B\circ} \end{bmatrix}.$$

Let $\boldsymbol{D}_d$ denote the matrix of zeroes and ones such that $\boldsymbol{D}_d\text{vech}(\boldsymbol{A}) = \text{vec}(\boldsymbol{A})$ for all $d \times d$ symmetric matrices $\boldsymbol{A}$. The Moore-Penrose inverse of $\boldsymbol{D}_d$ is $\boldsymbol{D}_d^+ = (\boldsymbol{D}_d^T \boldsymbol{D}_d)^{-1} \boldsymbol{D}_d^T$.

The result relies on the following assumptions:

- (A1) The cell dimensions m and m' diverge to ∞ in such a way that $m = O(m')$ and $m' = O(m)$.
- (A2) The within-cell sample sizes $n_{ii'}$ diverge to ∞ in such a way that $n_{ii'}/n \to C_{ii'}$ for positive constants $C_{ii'}$, $1 \le i \le m$, $1 \le i' \le m'$, that are bounded above and away from zero. Also, $n/m \to 0$ as m and n diverge.
- (A3) All entries of both $\boldsymbol{X}_{\rm A\circ}$ and $\boldsymbol{X}_{\rm B\circ}$ are not degenerate at zero and have finite second moment.

Result 1. *Assume that (A1)–(A3) and some additional regularity conditions hold. Then*

$$\begin{bmatrix} \left\{\dfrac{\boldsymbol{\Sigma}^0}{m} + \dfrac{(\boldsymbol{\Sigma}')^0}{m'}\right\}^{-1/2} \left(\widehat{\boldsymbol{\beta}}_A - \boldsymbol{\beta}_A^0\right) \\ \left\{\dfrac{(\sigma^2)^0 \boldsymbol{C}_{\boldsymbol{\beta}_B}}{mm'n}\right\}^{-1/2} \left(\widehat{\boldsymbol{\beta}}_B - \boldsymbol{\beta}_B^0\right) \\ \left\{\dfrac{2\boldsymbol{D}_{d_A}^+ (\boldsymbol{\Sigma}^0 \otimes \boldsymbol{\Sigma}^0) \boldsymbol{D}_{d_A}^{+T}}{m}\right\}^{-1/2} \text{vech}\left(\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}^0\right) \\ \left\{\dfrac{2\boldsymbol{D}_{d_A}^+ ((\boldsymbol{\Sigma}')^0 \otimes (\boldsymbol{\Sigma}')^0) \boldsymbol{D}_{d_A}^{+T}}{m'}\right\}^{-1/2} \text{vech}\left(\widehat{\boldsymbol{\Sigma}}' - (\boldsymbol{\Sigma}')^0\right) \\ \left[\dfrac{2\{(\sigma^2)^0\}^2}{mm'n}\right]^{-1/2} \left\{\widehat{\sigma}^2 - (\sigma^2)^0\right\} \end{bmatrix} \xrightarrow{D} N(\boldsymbol{0}, \boldsymbol{I}).$$

Some remarks concerning Result 1 are:

1. Result 1 provides following asymptotic covariance matrices of the maximum likelihood estimators:

$$\text{Asy.Cov}(\widehat{\boldsymbol{\beta}}_{\rm A}) = \frac{\boldsymbol{\Sigma}^0}{m} + \frac{(\boldsymbol{\Sigma}')^0}{m'}, \ \text{Asy.Cov}(\widehat{\boldsymbol{\beta}}_{\rm B}) = \frac{(\sigma^2)^0 \boldsymbol{C}_{\boldsymbol{\beta}_{\rm B}}}{mm'n}, \ \text{Asy.Cov}(\widehat{\boldsymbol{\Sigma}}) = \frac{2\boldsymbol{D}_{d_{\rm A}}^+ (\boldsymbol{\Sigma}^0 \otimes \boldsymbol{\Sigma}^0) \boldsymbol{D}_{d_{\rm A}}^{+T}}{m},$$

$$\text{Asy.Cov}(\widehat{\boldsymbol{\Sigma}}') = \frac{2\boldsymbol{D}_{d_{\rm A}}^+ ((\boldsymbol{\Sigma}')^0 \otimes (\boldsymbol{\Sigma}')^0) \boldsymbol{D}_{d_{\rm A}}^{+T}}{m'} \quad \text{and} \quad \text{Asy.Var}(\widehat{\sigma}^2) = \frac{2\{(\sigma^2)^0\}^2}{mm'n}.$$

There are marked differences in the rates of convergence. For example, the entries of $\widehat{\boldsymbol{\beta}}_{\rm A}$ have order m^{-1} asymptotic variances, whilst those of $\widehat{\boldsymbol{\beta}}_{\rm B}$ have order $(mm'n)^{-1}$ asymptotic variances. Note that $\boldsymbol{\beta}_{\rm A}^0$ and $\boldsymbol{\beta}_{\rm B}^0$ differ in that the former is partnered by crossed random effects in (1).

2. The asymptotic normality results for $\widehat{\boldsymbol{\Sigma}}$ and $\widehat{\boldsymbol{\Sigma}}'$ can be converted to forms that are more amenable to interpretation and confidence interval construction using the Multivariate Delta Method (e.g. Agresti, 2013, Section 16.1.3). For example, if $d_{\text{A}} = 2$ and the entries of $\boldsymbol{\Sigma}$ are parameterized as

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_1^2 & \rho\,\sigma_1\sigma_2 \\ \rho\,\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}$$

then Result 1 implies the following asymptotic normality results for standard transformations of the first standard deviation parameter and correlation parameter:

$$\sqrt{m}\big\{\log(\widehat{\sigma}_1) - \log(\sigma_1^0)\big\} \xrightarrow{D} N(0, \tfrac{1}{2}) \quad \text{and} \quad \sqrt{m}\big\{\tanh^{-1}(\widehat{\rho}) - \tanh^{-1}(\rho^0)\big\} \xrightarrow{D} N(0, 1).$$

Analogous results hold for $\widehat{\sigma}_2$ and $\widehat{\boldsymbol{\Sigma}}'$.

3. There is asymptotic orthogonality between each pair of random vectors within the set

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

4. Outside of Result 1 and Lyu *et al.* (2024), we are not aware of results for linear mixed models with crossed random effects that provide the precise asymptotic covariances given by Result 1 for estimation of fixed effects, even for simplified versions of (1) such as those having $\boldsymbol{X}_{\text{A}ii'} = \boldsymbol{1}_{n_{ii'}}$ and $\boldsymbol{X}_{\text{B}ii'}$ null. In this special case, in which the only fixed effect is the intercept parameter, the $\boldsymbol{\Sigma}/m + \boldsymbol{\Sigma}'/m'$ leading term behaviour is also apparent from Theorem 1 of Lyu *et al.* (2024) when their variable η is in the interior of the positive half-line. The predictor set-ups differ between the two articles, which hinders succinct comparison of the fixed effects results for more general cases.

5. Result 1 extends the results of Miller (1973) and Lyu *et al.* (2024), concerning asymptotic distributions of variance component estimators, to covariance matrices of arbitrary dimension.

6. Under (A1) m and m' diverge to ∞ at the same rate. In some circumstances this assumption may not be realistic and other assumptions concerning m and m' divergence may be more appropriate. The subsequent modification of Result 1 is straightforward. For example, if $m' = o(m)$ then the component concerning $\widehat{\boldsymbol{\beta}}_{\text{A}}$ becomes

$$\left\{\frac{(\boldsymbol{\Sigma}')^0}{m'}\right\}^{-1/2}\left(\widehat{\boldsymbol{\beta}}_{\text{A}} - \boldsymbol{\beta}_{\text{A}}^0\right) \xrightarrow{D} N(\boldsymbol{0}, \boldsymbol{I}) \quad \text{leading to} \quad \text{Asy.Cov}(\widehat{\boldsymbol{\beta}}_{\text{A}}) = \frac{(\boldsymbol{\Sigma}')^0}{m'}.$$

7. The asymptotic covariances for linear mixed models with crossed random effects have forms that are very similar to those with two-level nested random effects. See, for example, the Gaussian special case of Theorem 1 of Jiang, Wand & Bhaskaran (2022). At first glance, this result is somewhat surprising and intriguing since the two types of linear mixed models have fundamental differences. In Section 4 we provide some heuristic arguments that help explain this interesting phenomenon.

8. For the special case $\boldsymbol{X}_{\text{A}\circ} = 1$ and $\boldsymbol{X}_{\text{B}\circ} = X_\circ$ we have

$$\text{Asy.Var}(\widehat{\beta}_B) = \frac{(\sigma^2)^0}{\text{Var}(X_\circ)(\text{total sample size})}.$$

This matches the well-known expression for the asymptotic variance of the slope parameter in the simple linear regression model. Analagous results arise when $\boldsymbol{X}_{\text{B}\circ}$ is multivariate. Despite the presence of crossed random effects, the asymptotic behaviors of the estimators of slope parameters that are unaccompanied by random effects are the same as in the ordinary multiple regression situation. The heuristics in Section 4 provide some insight into this phenomenon.

9. The presence of multivariate random slopes in the crossed random effects model (1) leads to considerable challenges in the establishment of the Result 1 precise asymptotic normality statement. Detailed and delicate arguments, not given here, would be required to obtain sufficient regularity conditions under which Result 1 holds.

10. The establishment of Result 1 requires complicated and long-winded arguments, and are deferred to an online supplement.

4 Heuristics on Nested/Crossed Asymptotics Similarities

We now address the fact that the asymptotic covariance expressions in Result 1 are quite similar to those arising in the two-level nested case. This involves heuristic arguments that show that the fixed effects maximum likelihood estimators admit quite similar forms when sample means are replaced by population means. Throughout this section we write $\boldsymbol{\beta}$ rather than $\boldsymbol{\beta}^0$. A similar convention is used for $\boldsymbol{\Sigma}$, $\boldsymbol{\Sigma}'$ and σ^2. This suppression of the "true value" notation is to aid exposition.

Gaussian response linear mixed models have the following general form:

$$\boldsymbol{Y}|\boldsymbol{U} \sim N\big(\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}\boldsymbol{U}, \boldsymbol{R}\big), \quad \boldsymbol{U} \sim N(\boldsymbol{0}, \boldsymbol{G}). \tag{5}$$

For the crossed random effects model (1)

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

$$\boldsymbol{G} = \text{blockdiag}\big(\boldsymbol{I}_m \otimes \boldsymbol{\Sigma}, \boldsymbol{I}_{m'} \otimes \boldsymbol{\Sigma}'\big) \quad \text{and} \quad \boldsymbol{R} = \sigma^2\boldsymbol{I}$$

where $\boldsymbol{X}_A$ and $\boldsymbol{X}_B$ are given by (2).

The Gaussian version of the class of *nested* linear mixed models studied by Jiang *et al.* (2022) is

$$\begin{aligned} &\boldsymbol{Y}_i|\boldsymbol{U}_i, \boldsymbol{X}_{\text{A}i}, \boldsymbol{X}_{\text{B}i} \overset{\text{ind.}}{\sim} N\big((\boldsymbol{\beta}_{\text{A}} + \boldsymbol{U}_i)^T\boldsymbol{X}_{\text{A}i} + (\boldsymbol{\beta}_{\text{B}})^T\boldsymbol{X}_{\text{B}i}, \sigma^2\boldsymbol{I}\big), \\ &\boldsymbol{U}_i \overset{\text{ind.}}{\sim} N(\boldsymbol{0}, \boldsymbol{\Sigma}), \quad 1 \le i \le m, \end{aligned} \tag{6}$$

which is a special case of (5) with

$$\boldsymbol{X} = \underset{1\le i\le m}{\text{stack}}[\boldsymbol{X}_{\text{A}i} \ \boldsymbol{X}_{\text{B}i}], \quad \boldsymbol{Z} = \underset{1\le i\le m}{\text{blockdiag}}(\boldsymbol{X}_{\text{A}i}), \quad \boldsymbol{G} = \boldsymbol{I}_m \otimes \boldsymbol{\Sigma} \quad \text{and} \quad \boldsymbol{R} = \sigma^2\boldsymbol{I}.$$

In terms of the notation in (5), the fixed effects maximum likelihood estimator has the following generalized least squares form:

$$\widehat{\boldsymbol{\beta}} = (\boldsymbol{X}^T\boldsymbol{V}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{V}^{-1}\boldsymbol{Y} \quad \text{where} \quad \boldsymbol{V} \equiv \boldsymbol{Z}\boldsymbol{G}\boldsymbol{Z}^T + \boldsymbol{R}.$$

If $\mathcal{X}$ denotes the predictor data in the $\boldsymbol{X}$ and $\boldsymbol{Z}$ matrices then the conditional covariance matrix of the fixed effects estimator is

$$\text{Cov}(\widehat{\boldsymbol{\beta}}|\mathcal{X}) = \big(\boldsymbol{X}^T\boldsymbol{V}^{-1}\boldsymbol{X}\big)^{-1}.$$

For the remainder of this section we assume that the data are balanced. In the crossed case this corresponds to $n_{ii'} = n$ for all $1 \le i \le m$ and $1 \le i' \le m'$. For the nested case $n_i = n$ for all $1 \le i \le m$.

4.1 The $\boldsymbol{X} = \boldsymbol{1}$ Special Case

Consider the following special case of (5):

$$\boldsymbol{Y}|\boldsymbol{U} \sim N\big(\boldsymbol{1}\beta_0 + \boldsymbol{Z}\boldsymbol{U}, \boldsymbol{R}\big), \quad \boldsymbol{U} \sim N(\boldsymbol{0}, \boldsymbol{G}).$$

for which $\boldsymbol{X} = \boldsymbol{1}$, which is such that the only fixed effect effect is the intercept parameter β_0.

A further simplification is

$$\boldsymbol{Z} = \begin{cases} \begin{bmatrix} \boldsymbol{I}_m \otimes \boldsymbol{1}_{m'n} & \boldsymbol{1}_m \otimes \boldsymbol{I}_{m'} \otimes \boldsymbol{1}_n \end{bmatrix} & \text{for the crossed case,} \\ \boldsymbol{I}_m \otimes \boldsymbol{1}_n & \text{for the nested case,} \end{cases} \tag{7}$$

which corresponds to the random intercept-only models. Let $\boldsymbol{V}_{\text{cross}}$ and $\boldsymbol{V}_{\text{nest}}$ respectively denote the $\boldsymbol{V}$ matrix for the crossed and nested cases based on the versions of $\boldsymbol{Z}$ given in (7). Bringing in the commonly used notation $\boldsymbol{J}_d \equiv \boldsymbol{1}_d \boldsymbol{1}_d^T$ we then have

$$\boldsymbol{V}_{\text{cross}} = \Sigma(\boldsymbol{I}_m \otimes \boldsymbol{J}_{m'n}) + \Sigma'(\boldsymbol{J}_m \otimes \boldsymbol{I}_{m'} \otimes \boldsymbol{J}_n) + \sigma^2 \boldsymbol{I}_{mm'n} \quad \text{and} \quad \boldsymbol{V}_{\text{nest}} = \Sigma(\boldsymbol{I}_m \otimes \boldsymbol{J}_n) + \sigma^2 \boldsymbol{I}_{mn}$$

where $\Sigma \equiv \boldsymbol{\Sigma}$ and $\Sigma' \equiv \boldsymbol{\Sigma}'$ are scalars in the current random intercept special cases. The following results are key:

$$\boldsymbol{V}_{\text{cross}} \boldsymbol{1} = \lambda_{\text{cross}} \boldsymbol{1} \quad \text{and} \quad \boldsymbol{V}_{\text{nest}} \boldsymbol{1} = \lambda_{\text{nest}} \boldsymbol{1}, \tag{8}$$

where $\boldsymbol{1}$ denotes a vector of ones with appropriate size,

$$\lambda_{\text{cross}} \equiv \Sigma m' n + \Sigma' mn + \sigma^2 \quad \text{and} \quad \lambda_{\text{nest}} \equiv n\Sigma + \sigma^2. \tag{9}$$

The fact that $\boldsymbol{1}$ is an eigenvector of both $\boldsymbol{V}_{\text{cross}}$ and $\boldsymbol{V}_{\text{nest}}$ leads the fixed effects estimators having simpler and similar forms. A key step involves the inverse eigenvalue results

$$\boldsymbol{V}_{\text{cross}}^{-1} \boldsymbol{1} = \big(1/\lambda_{\text{cross}}\big) \boldsymbol{1} \quad \text{and} \quad \boldsymbol{V}_{\text{nest}}^{-1} \boldsymbol{1} = \big(1/\lambda_{\text{nest}}\big) \boldsymbol{1}.$$

We then obtain

$$\widehat{\beta}_0 = (\boldsymbol{1}^T \boldsymbol{V}^{-1} \boldsymbol{1})^{-1} \boldsymbol{1}^T \boldsymbol{V}^{-1} \boldsymbol{Y} = (\boldsymbol{1}^T \boldsymbol{1})^{-1} \boldsymbol{1}^T \boldsymbol{Y} = \text{average of the response data}$$

for both $\boldsymbol{V} = \boldsymbol{V}_{\text{cross}}$ and $\boldsymbol{V} = \boldsymbol{V}_{\text{nest}}$. We also have

$$\text{Var}(\widehat{\beta}_0) = \frac{\lambda}{\text{total sample size}} \tag{10}$$

where $\lambda = \lambda_{\text{cross}}$ in the crossed case and $\lambda = \lambda_{\text{nest}}$ in the nested case. Results (9) and (10) then lead to the exact expressions

$$\text{Var}(\widehat{\beta}_0) = \begin{cases} \dfrac{\Sigma}{m} + \dfrac{\Sigma'}{m'} + \dfrac{\sigma^2}{mm'n} & \text{in the crossed case,} \\ \dfrac{\Sigma}{m} + \dfrac{\sigma^2}{mn} & \text{in the nested case} \end{cases}$$

which are in keeping with the leading term expression in (1) and the analogous result in Jiang *et al.* (2022).

In this subsection, we have seen that the eigenvalue/eigenvector results given by (8) lead to the fixed effects estimator reducing to ordinary least squares form in both cases. Therefore, the β_0 estimators behave quite similarly despite the ostensible differences between the crossed and nested cases.

4.2 Heuristics for the General $\boldsymbol{X}$ Crossed Case

We commence by noting the following exact result:

$$\boldsymbol{V}_{\text{cross}} \boldsymbol{X} = \left[\underset{1 \le i \le m}{\text{stack}} \left[\left\{ \underset{1 \le i' \le m'}{\text{stack}} (\boldsymbol{X}_{\text{A}ii'}) \right\} \left(\boldsymbol{\Sigma} \sum_{i'=1}^{m'} \boldsymbol{X}_{\text{A}ii'}^T \boldsymbol{X}_{\text{A}ii'} + \boldsymbol{\Sigma}' \sum_{i=1}^{m} \boldsymbol{X}_{\text{A}ii'}^T \boldsymbol{X}_{\text{A}ii'} \right) \right] \right.$$
$$\left. \underset{1 \le i \le m}{\text{stack}} \left[\left\{ \underset{1 \le i' \le m'}{\text{stack}} (\boldsymbol{X}_{\text{A}ii'}) \right\} \left(\boldsymbol{\Sigma} \sum_{i'=1}^{m'} \boldsymbol{X}_{\text{A}ii'}^T \boldsymbol{X}_{\text{B}ii'} + \boldsymbol{\Sigma}' \sum_{i=1}^{m} \boldsymbol{X}_{\text{A}ii'}^T \boldsymbol{X}_{\text{B}ii'} \right) \right] \right] + \sigma^2 \boldsymbol{X}.$$

Then results such as

$$\frac{1}{mn}\sum_{i=1}^{m}\boldsymbol{X}_{\mathrm{A}ii'}^T\boldsymbol{X}_{\mathrm{A}ii'}\xrightarrow{P}E(\boldsymbol{X}_{\mathrm{A}\circ}\boldsymbol{X}_{\mathrm{A}\circ}^T)\text{ and }\quad\frac{1}{mn}\sum_{i=1}^{m}\boldsymbol{X}_{\mathrm{A}ii'}^T\boldsymbol{X}_{\mathrm{B}ii'}\xrightarrow{P}E(\boldsymbol{X}_{\mathrm{A}\circ}\boldsymbol{X}_{\mathrm{B}\circ}^T)$$

for all $1\le i'\le m'$ lead to the approximation

$$\boldsymbol{V}_{\text{cross}}\boldsymbol{X}\approx\boldsymbol{X}\boldsymbol{\Lambda}_{\text{cross}}$$

where

$$\boldsymbol{\Lambda}_{\text{cross}}\equiv\begin{bmatrix} n\big(m'\boldsymbol{\Sigma}+m\boldsymbol{\Sigma}'\big)E(\boldsymbol{X}_{\mathrm{A}\circ}\boldsymbol{X}_{\mathrm{A}\circ}^T)+\sigma^2\boldsymbol{I}_{d_{\mathrm{A}}} & n\big(m'\boldsymbol{\Sigma}+m\boldsymbol{\Sigma}'\big)E(\boldsymbol{X}_{\mathrm{A}\circ}\boldsymbol{X}_{\mathrm{B}\circ}^T)\\ \boldsymbol{O} & \sigma^2\boldsymbol{I}_{d_{\mathrm{B}}}\end{bmatrix}.\tag{11}$$

We then have

$$\widehat{\boldsymbol{\beta}}\approx\boldsymbol{\Lambda}_{\text{cross}}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{\Lambda}_{\text{cross}}^{-T}\boldsymbol{X}^T\boldsymbol{Y}\quad\text{and}\quad\mathrm{Cov}(\widehat{\boldsymbol{\beta}}|\mathcal{X})\approx\boldsymbol{\Lambda}_{\text{cross}}(\boldsymbol{X}^T\boldsymbol{X})^{-1}.\tag{12}$$

A simple consequence of (11) and (12) is

$$\begin{aligned}\mathrm{Cov}(\widehat{\boldsymbol{\beta}}_{\mathrm{B}}|\mathcal{X})&\approx\sigma^2\Big\{\text{lower right } d_{\mathrm{B}}\times d_{\mathrm{B}}\text{ block of }(\boldsymbol{X}^T\boldsymbol{X})^{-1}\Big\}\\&\approx\left(\frac{\sigma^2}{mm'n}\right)\Big[\text{lower right } d_{\mathrm{B}}\times d_{\mathrm{B}}\text{ block of }\big\{E\big(\boldsymbol{X}_\circ\boldsymbol{X}_\circ^T\big)\big\}^{-1}\Big].\\&=\left(\frac{\sigma^2}{\text{total sample size}}\right)\Big[\text{lower right } d_{\mathrm{B}}\times d_{\mathrm{B}}\text{ block of }\big\{E\big(\boldsymbol{X}_\circ\boldsymbol{X}_\circ^T\big)\big\}^{-1}\Big].\end{aligned}\tag{13}$$

The asymptotic covariance matrix of $\mathrm{Cov}(\widehat{\boldsymbol{\beta}}_{\mathrm{A}}|\mathcal{X})$ has a similar derivation based on (11) and (12).

4.3 Heuristics for the General $\boldsymbol{X}$ Nested Case

For the nested model (6) we have the exact expression

$$\boldsymbol{V}_{\text{nest}}\boldsymbol{X}=\boldsymbol{X}_A\boldsymbol{\Sigma}\Big(\underset{1\le i\le m}{\text{stack}}\big[\boldsymbol{X}_{\mathrm{A}i}^T\boldsymbol{X}_{\mathrm{A}i}\quad\boldsymbol{X}_{\mathrm{A}i}^T\boldsymbol{X}_{\mathrm{B}i}\big]\Big)+\sigma^2\boldsymbol{X}.$$

As $n\to\infty$ and for each $1\le i\le m$ we have

$$\frac{1}{n}\boldsymbol{X}_{\mathrm{A}i}^T\boldsymbol{X}_{\mathrm{A}i}\xrightarrow{P}E(\boldsymbol{X}_{\mathrm{A}\circ}\boldsymbol{X}_{\mathrm{A}\circ}^T)\quad\text{and}\quad\frac{1}{n}\boldsymbol{X}_{\mathrm{A}i}^T\boldsymbol{X}_{\mathrm{B}i}\xrightarrow{P}E(\boldsymbol{X}_{\mathrm{A}\circ}\boldsymbol{X}_{\mathrm{B}\circ}^T)\quad\text{as }n\to\infty.$$

Therefore

$$\boldsymbol{V}_{\text{nest}}\boldsymbol{X}\approx\boldsymbol{X}\boldsymbol{\Lambda}_{\text{nest}}\quad\text{where}\quad\boldsymbol{\Lambda}_{\text{nest}}\equiv\begin{bmatrix} n\boldsymbol{\Sigma}E(\boldsymbol{X}_{\mathrm{A}\circ}\boldsymbol{X}_{\mathrm{A}\circ}^T)+\sigma^2\boldsymbol{I}_{d_{\mathrm{A}}} & n\boldsymbol{\Sigma}E(\boldsymbol{X}_{\mathrm{A}\circ}\boldsymbol{X}_{\mathrm{B}\circ}^T)\\ \boldsymbol{O} & \sigma^2\boldsymbol{I}_{d_{\mathrm{B}}}\end{bmatrix}$$

which then leads to

$$\widehat{\boldsymbol{\beta}}\approx\boldsymbol{\Lambda}_{\text{nest}}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{\Lambda}_{\text{nest}}^{-T}\boldsymbol{X}^T\boldsymbol{Y}\quad\text{and}\quad\mathrm{Cov}(\widehat{\boldsymbol{\beta}}|\mathcal{X})\approx\boldsymbol{\Lambda}_{\text{nest}}(\boldsymbol{X}^T\boldsymbol{X})^{-1}.$$

The bottom d_{B} rows of $\boldsymbol{\Lambda}_{\text{nest}}$ have the same simple form as $\boldsymbol{\Lambda}_{\text{cross}}$ and we obtain

$$\mathrm{Cov}(\widehat{\boldsymbol{\beta}}_{\mathrm{B}}|\mathcal{X})\approx\left(\frac{\sigma^2}{\text{total sample size}}\right)\Big[\text{lower right } d_{\mathrm{B}}\times d_{\mathrm{B}}\text{ block of }\big\{E\big(\boldsymbol{X}_\circ\boldsymbol{X}_\circ^T\big)\big\}^{-1}\Big]$$

which matches (13) and, indeed, the asymptotic covariance matrix form that arises in ordinary multiple regression.

4.4 Closing Discussion on the Asymptotic Similarities

In this section we have provided heuristic justifications for the fixed effects estimators and their asymptotic covariance matrices having the approximate forms

$$\widehat{\boldsymbol{\beta}} \approx \boldsymbol{\Lambda}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{\Lambda}^{-T}\boldsymbol{X}^T\boldsymbol{Y} \quad \text{and} \quad \text{Cov}(\widehat{\boldsymbol{\beta}}|\mathcal{X}) \approx \boldsymbol{\Lambda}(\boldsymbol{X}^T\boldsymbol{X})^{-1}. \tag{14}$$

for *both* the crossed random effects model (1) and the nested model (6). The common approximate forms in (14) provide a reasonable explanation for the asymptotic covariance matrices in Result 1 having forms similar to the nested case.

The approximate $\widehat{\boldsymbol{\beta}}$ expression in (14) is intriguingly close to the well-known ordinary least expression. In the special case of $\boldsymbol{X}$ being a column vector, $\boldsymbol{\Lambda}$ is scalar and cancels to give the ordinary least squares form. Such reduction occurred in Section 4.1 for the $\boldsymbol{X} = \boldsymbol{1}$ case. However there is no such cancellation in general.

The heuristics in the general $\boldsymbol{X}$ cases involve approximations having generic form

$$\boldsymbol{V}\boldsymbol{X} \approx \boldsymbol{X}\boldsymbol{\Lambda}. \tag{15}$$

In the special case where $\boldsymbol{X} = \boldsymbol{x}$ is a column vector and $\boldsymbol{\Lambda} = \lambda$ is scalar then (15) becomes $\boldsymbol{V}\boldsymbol{x} \approx \boldsymbol{x}\lambda$ which corresponds, approximately, to λ being an eigenvalue of $\boldsymbol{V}$ with eigenvector $\boldsymbol{x}$. For general $\boldsymbol{X}$ and $\boldsymbol{\Lambda}$ (15), with "=" instead of "≈", is an instance of *Sylvester's equation* (e.g. Stewart & Sun, 1990; Chapter V, Section 1.2).

5 Statistical Utility

Result 1 provides a great deal of statistical utility concerning inference and design. Confidence intervals and Wald hypothesis tests based on studentization are immediate consequences. Another is sample size calculations, for which we provide some details in this section.

For illustration of sample size calculations arising from Result 1, consider the following special case of (1):

$$\begin{aligned} Y_{ii'j}|B_{ii'j}, X_{ii'j}, U_i, U'_{i'} &\overset{\text{ind.}}{\sim} N\Big(\beta_0^0 + U_i + U'_{i'} + \beta_1^0 B_{ii'j} + \beta_2^0 X_{ii'j} + \beta_3^0 B_{ii'j}X_{ii'j}, \sigma^2\Big), \\ U_i \overset{\text{ind.}}{\sim} N(0, \Sigma^0), \quad U'_{i'} &\overset{\text{ind.}}{\sim} N\big(0, (\Sigma')^0\big), \quad 1 \le i \le m, \quad 1 \le i' \le m', \quad 1 \le j \le n, \end{aligned} \tag{16}$$

where the $B_{ii'j} \overset{\text{ind.}}{\sim}$ Bernoulli(p) and the $X_{ii'j}$ being independently and identically distributed the same as a general random variable $X_\circ$ having finite second moment. Consider the one-sided hypotheses

$$H_0 : \beta_3^0 = 0 \quad \text{versus} \quad H_1 : \beta_3^0 > 0 \tag{17}$$

corresponding to a possibly positive interaction effect between the two predictors. Let $\Delta > 0$ be a particular alternative value of β_3^0 and let P be the corresponding power. Then Result 1 and standard arguments lead to the following sample size formula:

$$m = \left\lceil \frac{\{\Phi^{-1}(\alpha) + \Phi^{-1}(1-P)\}^2}{(\Delta/\sigma^0)^2 p(1-p)\text{Var}(X_\circ)m'n} \right\rceil \tag{18}$$

where, for any $x \in \mathbb{R}$, $\lceil x \rceil$ denotes the smallest integer greater than or equal to x and Φ^{-1} is the $N(0,1)$ quantile function.

Now consider a psychological study such that model (16) and hypotheses (17) apply with $m' = 25$ items and $n = 1$ observation per subject-item combination. How many subjects should be recruited to potentially detect a smallest meaningful interaction effect of $\Delta = 0.25$ with power 0.9 from a 0.05 level of significance test? If it is further be assumed that $p = \frac{1}{2}$ and $\text{Var}(X) = \frac{1}{12}$ then from (18) we should recruit

$$m = 53 \text{ subjects if the error standard deviation is } \sigma^0 = 0.4.$$

Table 1 below provides the required m values for some other values of σ^0.

In contemporary Gaussian response linear mixed model software, such as the function `lmer()` within the package lme4 (Bates *et al.*, 2015), standard errors are typically obtained using exact observed Fisher information rather than the approximation to the (expected) Fisher information on which (18) is based. This raises the question as to whether the number of subjects chosen according to the Result 1 approximation to the standard error of $\widehat{\beta}_3$ leads to the advertized power for exact Fisher information-based hypothesis tests. We addressed this question by running a simulation study that involved replication of $1,000$ simulated data sets corresponding to (16) with various noise levels according to $\sigma^0 \in \{0.2, 0.4, 0.8, 1.6\}$. The $B_{ii'j}$ and $X_{iid'j}$ data were generated from Bernoulli($\frac{1}{2}$) and Uniform$(0,1)$ distributions, respectively. As above, we set $(m', n, \Delta, \alpha, P) = (20, 1, 0.25, 0.05, 0.9)$ and determined m using (18). For each simulated data set we carried out a test of (17) using calls to `lmer()`, with rejection of H_0 if the t-statistic corresponding to β_3^0 exceeded $\Phi^{-1}(1-\alpha) = \Phi^{-1}(0.95)$. Table 1 shows the empirical estimates of $P = 0.9$ and corresponding 95% confidence intervals. For this example we see that the sample size formula (18) performs well with regards to the actual power delivered.

Error standard deviation (σ^0):	0.2	0.4	0.8	1.6
Minimum number of subjects (m):	14	53	211	842
Empirical estimate of power:	0.889	0.902	0.878	0.885
95% confidence interval of power:	$(0.870, 0.908)$	$(0.884, 0.920)$	$(0.858, 0.898)$	$(0.865, 0.905)$

Table 1: *The results from the illustrative sample size calculation and corresponding empirical power checks for the simulation study described in the text. The number of subjects (m) values correspond to an advertized power of* 0.9.

The example in this section demonstrates the statistical utility of Result 1. We are not aware of previous results in the literature for linear mixed models with crossed random effects that readily provide the sample size formula (18).

6 Concluding Remarks

Result 1 provides the precise leading term behaviours of the maximum likelihood estimators for a general class of linear mixed models containing crossed random effects and enables statistical utilities such as Wald tests for all model parameters and sample size calculations. It complements the recent contributions of Lyu *et al.* (2024) via extensions to random slopes and unbalanced designs. In comparison with the nested random effects situation, the establishment of leading term results in the presence of crossed random effects is lengthy and arduous – even when the responses are Gaussian. The leading terms have similar or identical forms to those arising in nested models, and we have provided some heuristic arguments for this phenomenon. We conjecture that the two-term asymptotic covariance matrices for $\widehat{\boldsymbol{\beta}}_{\mathrm{A}}$, $\widehat{\boldsymbol{\Sigma}}$ and $\widehat{\boldsymbol{\Sigma}}'$ in the Section 2 set-up are similar or identical to those appearing in Section 3.3.1 of Maestrini *et al.* (2024) for the nested case, but such an investigation would require a great deal of additional effort. Lastly, there are questions of what precise asymptotic results, if any, could be obtained for non-Gaussian and sparse data versions of linear mixed models containing crossed random effects. The current article may pave the way for such future endeavours.

Acknowledgements

We are grateful to John Duchi, Iain Johnstone, Ziyang Lyu, Luca Maestrini, Art Owen, Nicola Sartori and Alan Welsh for advice related to this research. This research was supported by the U.S. National Science Foundation grant DMS-2210569 and the Australian Research Council Discovery Project DP230101179. The second author is grateful for hospitality from the Department of Statistical Sciences, University of Padua, Italy, during part of this research.

References

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

Baayen, R.H., Davidson, D.J. & Bates, D.M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, **59**, 390–412.

Bates, D., Maechler, M., Bolker, B. and Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, **67(1)**, 1–48.

Gao, K. & Owen, A.B. (2020). Estimation and inference for very large linear mixed effects models. *Statistica Sinica*, **30**, 1741–1771.

Ghosh, S., Hastie, T. & Owen, A.B. (2022). Backfitting for large scale crossed random effects regressions. *The Annals of Statistics*, **50**, 560–583.

Hartley, H.O. & Rao, J.N.K. (1967). Maximum likelihood estimation for the mixed analysis of variance model. *Biometrika*, **54**, 93–108.

Jiang. J. (1996). REML estimation: asymptotic behavior and related topics. *The Annals of Statistics*, **24**, 255–286.

Jiang, J. (2013). The subset argument and consistency of MLE in GLMM: answer to an open problem and beyond. *The Annals of Statistics*, **41**, 177–195.

Jiang, J. & Nguyen, T. (2021). *Linear and Generalized Linear Mixed Models and Their Applications, Second Edition.* New York: Springer.

Jiang, J., Wand, M.P. & Bhaskaran, A. (2022). Usable and precise asymptotics for generalized linear mixed model analysis and design. *Journal of the Royal Statistical Society, Series B*, **84**, 55–82.

Lyu, Z., Sisson, S.A. & Welsh, A.H. (2024). Increasing dimension asymptotics for two-way crossed mixed effect models. `https://arxiv.org/abs/2401.06446v2`

Maestrini, L., Bhaskaran, A. & Wand, M.P. (2024). Second term improvement to generalised linear mixed model asymptotics. *Biometrika*, **111**, 1077–1084.

Miller, J.J. (1973). Asymptotic properties and computation of maximum likelihood estimates in the mixed model of the analysis of variance. Technical Report No. 12, Department of Statistics, Stanford University, Stanford, California, U.S.A.

Miller, J.J. (1977). Asymptotic properties of maximum likelihood estimates in the mixed model of the analysis of variance. *The Annals of Statistics*, **4**, 746–762.

Stewart, G.W. & Sun, J. (1990). *Matrix Perturbation Theory.* San Diego: Academic Press.

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

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

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 $\boldsymbol{V}$:

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

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 $\text{vech}(\boldsymbol{A}_d)\text{vech}(\boldsymbol{A}_d)^T$	*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

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

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* $n \times d$ *matrices, where* $n, d \in \mathbb{N}$. *Then, assuming all required matrix inverses exist,*

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

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:*

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

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

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

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,*

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

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

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

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

Then

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

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 $1 \le i' \le m'$:

$$\boldsymbol{X}_{ii'} \equiv \begin{bmatrix} \boldsymbol{X}_{ii'1}^T \\ \vdots \\ \boldsymbol{X}_{ii'n_{ii'}}^T \end{bmatrix}, \quad \boldsymbol{X} \equiv \underset{1\le i\le m}{\text{stack}}\,\big(\overset{\blacktriangle}{\boldsymbol{X}}_i\big) \quad \text{where} \quad \overset{\blacktriangle}{\boldsymbol{X}}_i \equiv \underset{1\le i'\le m'}{\text{stack}}\,(\boldsymbol{X}_{ii'}). \tag{S.2}$$

Next, let

$$\boldsymbol{Q}_{mm'} \equiv \underset{1\le i\le m}{\text{blockdiag}}\left\{\underset{1\le i',\underset{\sim}{i}'\le m'}{\text{blockmatrix}}(\boldsymbol{X}_{ii'}\boldsymbol{M}\boldsymbol{X}^T_{i\underset{\sim}{i}'})\right\} + \underset{1\le i,\underset{\sim}{i}\le m}{\text{blockmatrix}}\left\{\underset{1\le i'\le m'}{\text{blockdiag}}(\boldsymbol{X}_{ii'}\boldsymbol{M}'\boldsymbol{X}^T_{\underset{\sim}{i}i'})\right\} + \lambda\boldsymbol{I} \tag{S.3}$$

where

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

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

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

Introduce the following assumptions:

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

(A5) Each of the $n_{ii'}$, $1\le i\le m$, $1\le i'\le m'$, 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 $\boldsymbol{Q}^{i\underset{\sim}{i}}_{mm'}$, $1\le i\le m$, $1\le \underset{\sim}{i}\le m'$, according to (S.1)–(S.5). Under (A5) we have for fixed $m, m'\in\mathbb{N}$:*

(a) $\boldsymbol{X}^T\boldsymbol{Q}^{-1}_{mm'}\boldsymbol{X} \xrightarrow{P} \left(\frac{1}{m}\boldsymbol{M} + \frac{1}{m'}\boldsymbol{M}'\right)^{-1}$.

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

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

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

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

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 $1\le i'\le m'$:

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

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 for (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 for all fixed $m, m'\in\mathbb{N}$:*

(a) $\left(\sum_{i=1}^{m}\sum_{i'=1}^{m'} n_{ii'}\right)^{-1} \overset{\star}{\boldsymbol{X}}^T \boldsymbol{Q}_{mm'}^{-1} \overset{\star}{\boldsymbol{X}} \xrightarrow{P} (1/\lambda)\Big[\textit{lower right } \overset{\star}{d} \times \overset{\star}{d} \textit{ block of } \big\{E\big([\boldsymbol{X}_\circ\ \overset{\star}{\boldsymbol{X}}_\circ{}^T]^{\otimes 2}\big)\big\}^{-1}\Big]^{-1}.$

(b) $\boldsymbol{X}^T \boldsymbol{Q}_{mm'}^{-1} \overset{\star}{\boldsymbol{X}} \xrightarrow{P} \big(\tfrac{1}{m}\boldsymbol{M} + \tfrac{1}{m'}\boldsymbol{M}'\big)^{-1} \{E(\boldsymbol{X}_\circ^{\otimes 2})\}^{-1} E(\boldsymbol{X}_\circ \overset{\star}{\boldsymbol{X}}_\circ^T).$

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

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

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 $o_P(m^2 n)$. 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 $(\boldsymbol{\beta}_{\rm A}, \boldsymbol{\beta}_{\rm A})$ diagonal block is $\boldsymbol{X}_{\rm A}^T \boldsymbol{V}^{-1} \boldsymbol{X}_{\rm A}$. From (A3) and Lemma 5(a), we have for all fixed $m, m' \in \mathbb{N}$ and as $n \to \infty$

$$\boldsymbol{X}_{\rm A}^T \boldsymbol{V}^{-1} \boldsymbol{X}_{\rm A} \xrightarrow{P} \left(\frac{\boldsymbol{\Sigma}}{m} + \frac{\boldsymbol{\Sigma}'}{m'}\right)^{-1}.$$

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

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

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

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

$$\boldsymbol{X}_{\rm B}^T \boldsymbol{V}^{-1} \boldsymbol{X}_{\rm B} = \frac{mm'n\boldsymbol{C}_{\boldsymbol{\beta}_{\rm B}}^{-1}}{\sigma^2} + o_P(mm'n)\boldsymbol{1}_{d_{\rm B}}^{\otimes 2}.$$

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

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

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

Then note that

$$\frac{\partial \boldsymbol{V}}{\partial(\boldsymbol{\Sigma})_{rs}} = \boldsymbol{L}_r\boldsymbol{L}_s^T + I(r \neq s)\boldsymbol{L}_s\boldsymbol{L}_r^T \quad \text{where} \quad \boldsymbol{L}_r \equiv \underset{1\le i\le m}{\text{blockdiag}}\big(\overset{\blacktriangle}{\boldsymbol{X}}_{{\rm A}i}\boldsymbol{e}_r\big), \quad \overset{\blacktriangle}{\boldsymbol{X}}_{{\rm A}i} \equiv \underset{1\le i'\le m'}{\text{stack}}(\boldsymbol{X}_{{\rm A}ii'}) \tag{S.9}$$

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

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

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

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

Since

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

we then have

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

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:

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

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:

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

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

$$
\tfrac{1}{2}m'\boldsymbol{D}_{d_\mathrm{A}}^T\big((\boldsymbol{\Sigma}')^{-1}\otimes(\boldsymbol{\Sigma}')^{-1}\big)\boldsymbol{D}_{d_\mathrm{A}}+O_P(1)\boldsymbol{1}_{d_\mathrm{A}(d_\mathrm{A}+1)/2}^{\otimes 2}.
$$

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

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

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 $(\boldsymbol{\beta}_{\text{A}},\boldsymbol{\beta}_{\text{B}})$ diagonal block is $\boldsymbol{X}_{\text{A}}^T\boldsymbol{V}^{-1}\boldsymbol{X}_{\text{B}}$. From (A3) and Lemma 6(b), we have for all fixed $m,m'\in\mathbb{N}$ and as $n\to\infty$

$$\boldsymbol{X}_{\text{A}}^T\boldsymbol{V}^{-1}\boldsymbol{X}_{\text{B}}\xrightarrow{P}\left(\frac{\boldsymbol{\Sigma}}{m}+\frac{\boldsymbol{\Sigma}'}{m'}\right)^{-1}\{E(\boldsymbol{X}_{\text{A}\circ}^{\otimes 2})\}^{-1}E(\boldsymbol{X}_{\text{A}\circ}^T\boldsymbol{X}_{\text{B}\circ}).$$

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

$$\left(\frac{\boldsymbol{\Sigma}}{m}+\frac{\boldsymbol{\Sigma}'}{m'}\right)^{-1}\{E(\boldsymbol{X}_{\text{A}\circ}^{\otimes 2})\}^{-1}E(\boldsymbol{X}_{\text{A}\circ}^T\boldsymbol{X}_{\text{B}\circ})+o_P(m).$$

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

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

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

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 $\big(\text{vech}(\boldsymbol{\Sigma}),\text{vech}(\boldsymbol{\Sigma}')\big)$ Off-Diagonal Block

We commence with the special case of $d_{\text{A}}=1$, $n_{ii'}=n$ and $\boldsymbol{X}_{\text{A}ii'}=\boldsymbol{1}_n$ for all $1\le i\le m$, $1\le i'\le m'$. In this case $\boldsymbol{1}_{mm'n}$ is an eigenvector of $\boldsymbol{V}$ with corresponding eigenvalue $m'n\Sigma+mn\Sigma'+\sigma^2$. This implies that $\boldsymbol{1}_{mm'n}$ 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 assumption (A1).

Next we treat the general d_{A}, $n_{ii'}$ and $\boldsymbol{X}_{\text{A}ii'}$ 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 $\big(\text{vech}(\boldsymbol{\Sigma}),\text{vech}(\boldsymbol{\Sigma}')\big)$ off-diagonal block of the Fisher information matrix is

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

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

$$\frac{\partial\boldsymbol{V}}{\partial(\boldsymbol{\Sigma})_{rr}}=\underset{1\le i\le m}{\text{blockdiag}}\big(\boldsymbol{X}_{\text{A}i1}\boldsymbol{e}_r\boldsymbol{e}_r^T\boldsymbol{X}_{\text{A}i1}^T\big)\text{ and }\frac{\partial\boldsymbol{V}}{\partial(\boldsymbol{\Sigma}')_{tt}}=\underset{1\le i,\underset{\sim}{i}\le m}{\text{blockmatrix}}\big(\boldsymbol{X}_{\text{A}i1}\boldsymbol{e}_t\boldsymbol{e}_t^T\boldsymbol{X}_{\text{A}\underset{\sim}{i}1}^T\big). \tag{S.14}$$

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

$$
\begin{aligned}
\text{tr}\left(\boldsymbol{V}^{-1}\frac{\partial\boldsymbol{V}}{\partial(\boldsymbol{\Sigma})_{rr}}\boldsymbol{V}^{-1}\frac{\partial\boldsymbol{V}}{\partial(\boldsymbol{\Sigma}')_{tt}}\right) &= \sum_{i=1}^{m}\sum_{\underset{\sim}{i}=1}^{m}\sum_{\underset{\sim}{i}^*=1}^{m}\left(\boldsymbol{e}_t^T\boldsymbol{X}_{\text{A}i1}^T\boldsymbol{V}^{i\underset{\sim}{i}}\boldsymbol{X}_{\text{A}\underset{\sim}{i}1}\boldsymbol{e}_r\right)\left(\boldsymbol{e}_r^T\boldsymbol{X}_{\text{A}\underset{\sim}{i}1}^T\boldsymbol{V}^{\underset{\sim}{i}\underset{\sim}{i}^*}\boldsymbol{X}_{\text{A}\underset{\sim}{i}^*1}\boldsymbol{e}_t\right)\\
&= \sum_{i=1}^{m}\left(\boldsymbol{e}_r^T\boldsymbol{X}_{\text{A}i1}^T\boldsymbol{V}^{ii}\boldsymbol{X}_{\text{A}i1}\boldsymbol{e}_t\right)^2\\
&\quad + \sum\sum_{i\neq\underset{\sim}{i}^*}\left(\boldsymbol{e}_t^T\boldsymbol{X}_{\text{A}i1}^T\boldsymbol{V}^{ii}\boldsymbol{X}_{\text{A}i1}\boldsymbol{e}_r\right)\left(\boldsymbol{e}_r^T\boldsymbol{X}_{\text{A}i1}^T\boldsymbol{V}^{i\underset{\sim}{i}^*}\boldsymbol{X}_{\text{A}\underset{\sim}{i}^*1}\boldsymbol{e}_t\right)\\
&\quad + \sum\sum_{i\neq\underset{\sim}{i}}\left(\boldsymbol{e}_t^T\boldsymbol{X}_{\text{A}i1}^T\boldsymbol{V}^{i\underset{\sim}{i}}\boldsymbol{X}_{\text{A}\underset{\sim}{i}1}\boldsymbol{e}_r\right)\left(\boldsymbol{e}_r^T\boldsymbol{X}_{\text{A}\underset{\sim}{i}1}^T\boldsymbol{V}^{\underset{\sim}{i}i}\boldsymbol{X}_{\text{A}i1}\boldsymbol{e}_t\right)\\
&\quad + \sum\sum\sum_{i\neq\underset{\sim}{i}\neq\underset{\sim}{i}^*}\left(\boldsymbol{e}_t^T\boldsymbol{X}_{\text{A}i1}^T\boldsymbol{V}^{i\underset{\sim}{i}}\boldsymbol{X}_{\text{A}\underset{\sim}{i}1}\boldsymbol{e}_r\right)\left(\boldsymbol{e}_r^T\boldsymbol{X}_{\text{A}\underset{\sim}{i}1}^T\boldsymbol{V}^{\underset{\sim}{i}\underset{\sim}{i}^*}\boldsymbol{X}_{\text{A}\underset{\sim}{i}^*1}\boldsymbol{e}_t\right).
\end{aligned}
$$

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

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

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 $d_{\text{A}} = 1$, $n_{ii'} = n$ and $\boldsymbol{X}_{\text{A}ii'} = \boldsymbol{1}_n$ case and a matrix with order $O(1)\boldsymbol{1}_{d_{\text{A}}(d_{\text{A}}+1)/2}^{\otimes 2}$ under (A1) eventuates.

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

We commence with the special case of $d_{\text{A}} = 1$, $n_{ii'} = n$ and $\boldsymbol{X}_{\text{A}ii'} = \boldsymbol{1}_n$ 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:

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

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

Now consider the general d_{A}, $n_{ii'}$ and $\boldsymbol{X}_{\text{A}ii'}$ 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 $\big(\text{vech}(\boldsymbol{\Sigma}), \sigma^2\big)$ off-diagonal block of the Fisher information matrix is

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

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

$$\begin{aligned} n_{11}\boldsymbol{X}_{\text{A}11}^T\boldsymbol{V}_{11}^{-2}\boldsymbol{X}_{\text{A}11} &= n_{11}\boldsymbol{e}_r^T\boldsymbol{X}_{\text{A}11}^T\{\boldsymbol{X}_{\text{A}11}(\boldsymbol{\Sigma}+\boldsymbol{\Sigma}')\boldsymbol{X}_{\text{A}11}^T+\sigma^2\boldsymbol{I}\}^{-2}\boldsymbol{X}_{\text{A}11}\boldsymbol{e}_r \\ &= \boldsymbol{e}_r^T\big\{\boldsymbol{\Sigma}+\boldsymbol{\Sigma}'+\sigma^2(\boldsymbol{X}_{\text{A}11}^T\boldsymbol{X}_{\text{A}11})^{-1}\big\}^{-1}\left(\tfrac{1}{n_{11}}\boldsymbol{X}_{\text{A}11}^T\boldsymbol{X}_{\text{A}11}\right)^{-1} \\ &\qquad \times\big\{\boldsymbol{\Sigma}+\boldsymbol{\Sigma}'+\sigma^2(\boldsymbol{X}_{\text{A}11}^T\boldsymbol{X}_{\text{A}11})^{-1}\big\}^{-1}\boldsymbol{e}_r \\ &\xrightarrow{P} \boldsymbol{e}_r^T(\boldsymbol{\Sigma}+\boldsymbol{\Sigma}')^{-1}E(\boldsymbol{X}_\circ^T\boldsymbol{X}_\circ)(\boldsymbol{\Sigma}+\boldsymbol{\Sigma}')^{-1}\boldsymbol{e}_r^T. \end{aligned}$$

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

$$\frac{\big((\boldsymbol{\Sigma}+\boldsymbol{\Sigma}')^{-1}E(\boldsymbol{X}_\circ^T\boldsymbol{X}_\circ)(\boldsymbol{\Sigma}+\boldsymbol{\Sigma}')^{-1}\big)_{rr}\{1+o_P(1)\}}{2n_{11}}$$

which extends (S.15) for $d_{\text{A}} \in \mathbb{N}$ 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 assumption (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^T\boldsymbol{X}_{\text{A}11}^T(\boldsymbol{V}^{11})^2\boldsymbol{X}_{\text{A}11}\boldsymbol{e}_r \\ &\quad = n_{11}\boldsymbol{e}_r^T\boldsymbol{X}_{\text{A}11}^T\Bigg(\text{upper left } n_{11}\times n_{11} \text{ block of} \\ &\qquad\qquad \begin{bmatrix} \boldsymbol{X}_{\text{A}11}(\boldsymbol{\Sigma}+\boldsymbol{\Sigma}')\boldsymbol{X}_{\text{A}11}^T+\sigma^2\boldsymbol{I} & \boldsymbol{X}_{\text{A}11}\boldsymbol{\Sigma}'\boldsymbol{X}_{\text{A}21}^T \\ \boldsymbol{X}_{\text{A}21}\boldsymbol{\Sigma}'\boldsymbol{X}_{\text{A}11}^T & \boldsymbol{X}_{\text{A}21}(\boldsymbol{\Sigma}+\boldsymbol{\Sigma}')\boldsymbol{X}_{\text{A}21}^T+\sigma^2\boldsymbol{I}\end{bmatrix}^{-1}\Bigg)^2\boldsymbol{X}_{\text{A}11}\boldsymbol{e}_r \\ &\quad = n_{11}\boldsymbol{e}_r^T\boldsymbol{X}_{\text{A}11}^T\Big[\boldsymbol{X}_{\text{A}11}(\boldsymbol{\Sigma}+\boldsymbol{\Sigma}')\boldsymbol{X}_{\text{A}11}^T+\sigma^2\boldsymbol{I} \\ &\qquad\qquad -\boldsymbol{X}_{\text{A}11}\boldsymbol{\Sigma}'\boldsymbol{X}_{\text{A}21}^T\big\{\boldsymbol{X}_{\text{A}21}(\boldsymbol{\Sigma}+\boldsymbol{\Sigma}')\boldsymbol{X}_{\text{A}21}^T+\sigma^2\boldsymbol{I}\big\}^{-1}\boldsymbol{X}_{\text{A}21}\boldsymbol{\Sigma}'\boldsymbol{X}_{\text{A}11}^T\Big]^{-2}\boldsymbol{X}_{\text{A}11}\boldsymbol{e}_r \\ &\quad = n_{11}\boldsymbol{e}_r^T\boldsymbol{X}_{\text{A}11}^T\Big[\boldsymbol{X}_{\text{A}11}(\boldsymbol{\Sigma}+\boldsymbol{\Sigma}')\boldsymbol{X}_{\text{A}11}^T+\sigma^2\boldsymbol{I} \\ &\qquad\qquad -\boldsymbol{X}_{\text{A}11}\boldsymbol{\Sigma}'\big\{\boldsymbol{\Sigma}+\boldsymbol{\Sigma}'+\sigma^2(\boldsymbol{X}_{\text{A}21}^T\boldsymbol{X}_{\text{A}21})^{-1}\boldsymbol{I}\big\}^{-1}\boldsymbol{\Sigma}'\boldsymbol{X}_{\text{A}11}^T\Big]^{-2}\boldsymbol{X}_{\text{A}11}\boldsymbol{e}_r \\ &\quad = \boldsymbol{e}_r^T\Big[\boldsymbol{\Sigma}+\boldsymbol{\Sigma}'-\boldsymbol{\Sigma}'\big\{\boldsymbol{\Sigma}+\boldsymbol{\Sigma}'+\sigma^2(\boldsymbol{X}_{\text{A}21}^T\boldsymbol{X}_{\text{A}21})^{-1}\big\}^{-1}\boldsymbol{\Sigma}'+\sigma^2(\boldsymbol{X}_{\text{A}11}^T\boldsymbol{X}_{\text{A}11})^{-1}\Big]^{-1} \\ &\qquad \times\left(\tfrac{1}{n_{11}}\boldsymbol{X}_{\text{A}11}^T\boldsymbol{X}_{\text{A}11}\right)^{-1} \\ &\qquad \times\Big[\boldsymbol{\Sigma}+\boldsymbol{\Sigma}'-\boldsymbol{\Sigma}'\big\{\boldsymbol{\Sigma}+\boldsymbol{\Sigma}'+\sigma^2(\boldsymbol{X}_{\text{A}21}^T\boldsymbol{X}_{\text{A}21})^{-1}\big\}^{-1}\boldsymbol{\Sigma}'+\sigma^2(\boldsymbol{X}_{\text{A}11}^T\boldsymbol{X}_{\text{A}11})^{-1}\Big]^{-1}\boldsymbol{e}_r \\ &\quad \xrightarrow{P} \boldsymbol{e}_r^T\Big\{\boldsymbol{\Sigma}+\boldsymbol{\Sigma}'-\boldsymbol{\Sigma}'\big(\boldsymbol{\Sigma}+\boldsymbol{\Sigma}'\big)^{-1}\boldsymbol{\Sigma}'\Big\}^{-1}E(\boldsymbol{X}_\circ^T\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

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

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

$$
\begin{aligned}
&n_{21}\boldsymbol{e}_r^T\boldsymbol{X}_{\mathrm{A}11}^T(\boldsymbol{V}^{21})^T\boldsymbol{V}^{21}\boldsymbol{X}_{\mathrm{A}11}\boldsymbol{e}_r\\
&\quad= n_{21}\boldsymbol{e}_r^T\boldsymbol{X}_{\mathrm{A}11}^T\Bigg(\text{the transposed lower left } n_{12}\times n_{11} \text{ block of}\\
&\qquad\qquad \begin{bmatrix}\boldsymbol{X}_{\mathrm{A}11}(\boldsymbol{\Sigma}+\boldsymbol{\Sigma}')\boldsymbol{X}_{\mathrm{A}11}^T+\sigma^2\boldsymbol{I} & \boldsymbol{X}_{\mathrm{A}11}\boldsymbol{\Sigma}'\boldsymbol{X}_{\mathrm{A}21}^T\\ \boldsymbol{X}_{\mathrm{A}21}\boldsymbol{\Sigma}'\boldsymbol{X}_{\mathrm{A}11}^T & \boldsymbol{X}_{\mathrm{A}21}(\boldsymbol{\Sigma}+\boldsymbol{\Sigma}')\boldsymbol{X}_{\mathrm{A}21}^T+\sigma^2\boldsymbol{I}\end{bmatrix}^{-1}\Bigg)^{\otimes 2}\boldsymbol{X}_{\mathrm{A}21}\boldsymbol{e}_r\\
&\quad= n_{21}\boldsymbol{e}_r^T\boldsymbol{X}_{\mathrm{A}11}^T\{\boldsymbol{X}_{\mathrm{A}11}(\boldsymbol{\Sigma}+\boldsymbol{\Sigma}')\boldsymbol{X}_{\mathrm{A}11}^T+\sigma^2\boldsymbol{I}\}^{-1}\boldsymbol{X}_{\mathrm{A}11}\boldsymbol{\Sigma}'\boldsymbol{X}_{\mathrm{A}21}^T\\
&\qquad\quad\times\Big[\boldsymbol{X}_{\mathrm{A}21}(\boldsymbol{\Sigma}+\boldsymbol{\Sigma}')\boldsymbol{X}_{\mathrm{A}21}^T+\sigma^2\boldsymbol{I}\\
&\qquad\qquad\qquad -\boldsymbol{X}_{\mathrm{A}21}\boldsymbol{\Sigma}'\boldsymbol{X}_{\mathrm{A}11}^T\{\boldsymbol{X}_{\mathrm{A}11}(\boldsymbol{\Sigma}+\boldsymbol{\Sigma}')\boldsymbol{X}_{\mathrm{A}11}^T+\sigma^2\boldsymbol{I}\}^{-1}\boldsymbol{X}_{\mathrm{A}11}\boldsymbol{\Sigma}'\boldsymbol{X}_{\mathrm{A}21}^T\Big]^{-2}\\
&\qquad\quad\times\boldsymbol{X}_{\mathrm{A}21}\boldsymbol{\Sigma}'\boldsymbol{X}_{\mathrm{A}11}^T\{\boldsymbol{X}_{\mathrm{A}11}(\boldsymbol{\Sigma}+\boldsymbol{\Sigma}')\boldsymbol{X}_{\mathrm{A}11}^T+\sigma^2\boldsymbol{I}\}^{-1}\boldsymbol{X}_{\mathrm{A}11}\boldsymbol{e}_r\\
&\quad= \boldsymbol{e}_r^T\{\boldsymbol{\Sigma}+\boldsymbol{\Sigma}'+\sigma^2(\boldsymbol{X}_{\mathrm{A}11}^T\boldsymbol{X}_{\mathrm{A}11})^{-1}\}^{-1}\boldsymbol{\Sigma}'\\
&\qquad\quad\times\Big[\boldsymbol{\Sigma}+\boldsymbol{\Sigma}'+\sigma^2(\boldsymbol{X}_{\mathrm{A}21}^T\boldsymbol{X}_{\mathrm{A}21})^{-1}-\boldsymbol{\Sigma}'\{\boldsymbol{\Sigma}+\boldsymbol{\Sigma}'+\sigma^2(\boldsymbol{X}_{\mathrm{A}11}^T\boldsymbol{X}_{\mathrm{A}11})^{-1}\}^{-1}\boldsymbol{\Sigma}'\Big]^{-1}\\
&\qquad\quad\times\left(\tfrac{1}{n_{21}}\boldsymbol{X}_{\mathrm{A}21}^T\boldsymbol{X}_{\mathrm{A}21}\right)^{-1}\\
&\qquad\quad\times\Big[\boldsymbol{\Sigma}+\boldsymbol{\Sigma}'+\sigma^2(\boldsymbol{X}_{\mathrm{A}21}^T\boldsymbol{X}_{\mathrm{A}21})^{-1}-\boldsymbol{\Sigma}'\{\boldsymbol{\Sigma}+\boldsymbol{\Sigma}'+\sigma^2(\boldsymbol{X}_{\mathrm{A}11}^T\boldsymbol{X}_{\mathrm{A}11})^{-1}\}^{-1}\boldsymbol{\Sigma}'\Big]^{-1}\\
&\qquad\quad\times\boldsymbol{\Sigma}'\{\boldsymbol{\Sigma}+\boldsymbol{\Sigma}'+\sigma^2(\boldsymbol{X}_{\mathrm{A}11}^T\boldsymbol{X}_{\mathrm{A}11})^{-1}\}^{-1}\boldsymbol{e}_r\\
&\quad\xrightarrow{P}\boldsymbol{e}_r^T(\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^T\boldsymbol{X}_\circ)\\
&\qquad\quad\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 $n_{11}\boldsymbol{e}_r^T\boldsymbol{X}_{\mathrm{A}21}^T(\boldsymbol{V}^{12})^T\boldsymbol{V}^{12}\boldsymbol{X}_{\mathrm{A}21}\boldsymbol{e}_r$ 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:

$$
\begin{aligned}
&\tfrac{1}{2}\left(\tfrac{1}{n_{11}}+\tfrac{1}{n_{21}}\right)\Bigg(\Big\{\boldsymbol{\Sigma}+\boldsymbol{\Sigma}'-\boldsymbol{\Sigma}'(\boldsymbol{\Sigma}+\boldsymbol{\Sigma}')^{-1}\boldsymbol{\Sigma}'\Big\}^{-1}\\
&\qquad\qquad\times E(\boldsymbol{X}_\circ^T\boldsymbol{X}_\circ)\Big\{\boldsymbol{\Sigma}+\boldsymbol{\Sigma}'-\boldsymbol{\Sigma}'(\boldsymbol{\Sigma}+\boldsymbol{\Sigma}')^{-1}\boldsymbol{\Sigma}'\Big\}^{-1}\Bigg)_{rr}\{1+o_P(1)\}\\
&+\tfrac{1}{2}\left(\tfrac{1}{n_{11}}+\tfrac{1}{n_{21}}\right)\Bigg((\boldsymbol{\Sigma}+\boldsymbol{\Sigma}')^{-1}\boldsymbol{\Sigma}'\Big\{\boldsymbol{\Sigma}+\boldsymbol{\Sigma}'-\boldsymbol{\Sigma}'(\boldsymbol{\Sigma}+\boldsymbol{\Sigma}')^{-1}\boldsymbol{\Sigma}'\Big\}^{-1}E(\boldsymbol{X}_\circ^T\boldsymbol{X}_\circ)\\
&\qquad\qquad\times\Big\{\boldsymbol{\Sigma}+\boldsymbol{\Sigma}'-\boldsymbol{\Sigma}'(\boldsymbol{\Sigma}+\boldsymbol{\Sigma}')^{-1}\boldsymbol{\Sigma}'\Big\}^{-1}\boldsymbol{\Sigma}'(\boldsymbol{\Sigma}+\boldsymbol{\Sigma}')^{-1}\Bigg)_{rr}\{1+o_P(1)\}
\end{aligned}
$$

which, under assumption (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 $\big(\text{vech}(\boldsymbol{\Sigma}'),\sigma^2\big)$ 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 $\big(\text{vech}(\boldsymbol{\Sigma}'),\sigma^2\big)$ 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

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

$$\begin{bmatrix}
\left(\frac{\boldsymbol{\Sigma}}{m}+\frac{\boldsymbol{\Sigma}'}{m'}\right)^{-1}+o_P(m)\mathbf{1}_{d_{\text{A}}}^{\otimes 2} & O_P(m)\mathbf{1}_{d_{\text{A}}}\mathbf{1}_{d_{\text{B}}}^T & \boldsymbol{O} & \boldsymbol{O} & \boldsymbol{O}\\
O_P(m)\mathbf{1}_{d_{\text{B}}}\mathbf{1}_{d_{\text{A}}}^T & \frac{mm'n\boldsymbol{C}_{\boldsymbol{\beta}_{\text{B}}}^{-1}}{\sigma^2}+o_P(m^2n)\mathbf{1}_{d_{\text{B}}}^{\otimes 2} & \boldsymbol{O} & \boldsymbol{O} & \boldsymbol{O}\\
\boldsymbol{O} & \boldsymbol{O} & \frac{m\boldsymbol{D}_{d_{\text{A}}}^T(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1})\boldsymbol{D}_{d_{\text{A}}}}{2}+o_P(m)\mathbf{1}_{d_{\text{A}}^{\boxplus}}^{\otimes 2} & O_P(1)\mathbf{1}_{d_{\text{A}}^{\boxplus}}^{\otimes 2} & O_P(n^{-1})\mathbf{1}_{d_{\text{A}}^{\boxplus}}\\
\boldsymbol{O} & \boldsymbol{O} & O_P(1)\mathbf{1}_{d_{\text{A}}^{\boxplus}}^{\otimes 2} & \frac{m'\boldsymbol{D}_{d_{\text{A}}}^T\big((\boldsymbol{\Sigma}')^{-1}\otimes(\boldsymbol{\Sigma}')^{-1}\big)\boldsymbol{D}_{d_{\text{A}}}}{2}+o_P(m)\mathbf{1}_{d_{\text{A}}^{\boxplus}}^{\otimes 2} & O_P(n^{-1})\mathbf{1}_{d_{\text{A}}^{\boxplus}}\\
\boldsymbol{O} & \boldsymbol{O} & O_P(n^{-1})\mathbf{1}_{d_{\text{A}}^{\boxplus}}^T & O_P(n^{-1})\mathbf{1}_{d_{\text{A}}^{\boxplus}}^T & \frac{mm'n}{2\sigma^4}+o_P(m^2n)
\end{bmatrix}$$

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 $I\big(\boldsymbol{\beta}_{\text{A}},\boldsymbol{\beta}_{\text{B}},\text{vech}(\boldsymbol{\Sigma}),\text{vech}(\boldsymbol{\Sigma}'),\sigma^2\big)$ is block diagonal, its inversion involves the individual inversions of the $\big(\boldsymbol{\beta}_{\text{A}},\boldsymbol{\beta}_{\text{B}}\big)$ and $\big(\text{vech}(\boldsymbol{\Sigma}),\text{vech}(\boldsymbol{\Sigma}'),\sigma^2\big)$ 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:

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

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

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

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

$$
\begin{aligned}
&I\big(\boldsymbol{\beta}_{\mathrm{A}},\boldsymbol{\beta}_{\mathrm{B}},\mathrm{vech}(\boldsymbol{\Sigma}),\mathrm{vech}(\boldsymbol{\Sigma}'),\sigma^2\big)^{-1} = I\big(\boldsymbol{\beta}_{\mathrm{A}},\boldsymbol{\beta}_{\mathrm{B}},\mathrm{vech}(\boldsymbol{\Sigma}),\mathrm{vech}(\boldsymbol{\Sigma}'),\sigma^2\big)^{-1}_{\infty}\\
&+\frac{1}{m}\begin{bmatrix}
o_P(1)\mathbf{1}_{d_{\mathrm{A}}}^{\otimes 2} & O_P(m^{-1}n^{-1})\mathbf{1}_{d_{\mathrm{A}}}\mathbf{1}_{d_{\mathrm{B}}}^T & \boldsymbol{O} & \boldsymbol{O} & \boldsymbol{O}\\
O_P(m^{-1}n^{-1})\mathbf{1}_{d_{\mathrm{B}}}\mathbf{1}_{d_{\mathrm{A}}}^T & o_P(m^{-1}n^{-1})\mathbf{1}_{d_{\mathrm{B}}}^{\otimes 2} & \boldsymbol{O} & \boldsymbol{O} & \boldsymbol{O}\\
\boldsymbol{O} & \boldsymbol{O} & o_P(1)\mathbf{1}_{d_{\mathrm{A}}^{\boxplus}}^{\otimes 2} & O_P(m^{-1})\mathbf{1}_{d_{\mathrm{A}}^{\boxplus}}\mathbf{1}_{d_{\mathrm{A}}^{\boxplus}}^{\otimes 2} & O_P(m^{-2}n^{-1})\mathbf{1}_{d_{\mathrm{A}}^{\boxplus}}\\
\boldsymbol{O} & \boldsymbol{O} & O_P(m^{-1})\mathbf{1}_{d_{\mathrm{A}}^{\boxplus}}\mathbf{1}_{d_{\mathrm{A}}^{\boxplus}}^{\otimes 2} & o_P(1)\mathbf{1}_{d_{\mathrm{A}}^{\boxplus}}^{\otimes 2} & O_P(m^{-2}n^{-1})\mathbf{1}_{d_{\mathrm{A}}^{\boxplus}}\\
\boldsymbol{O} & \boldsymbol{O} & O_P(m^{-2}n^{-1})\mathbf{1}_{d_{\mathrm{A}}^{\boxplus}}^T & O_P(m^{-2}n^{-1})\mathbf{1}_{d_{\mathrm{A}}^{\boxplus}}^T & o_P(m^{-2}n^{-1})
\end{bmatrix}
\end{aligned}
$$

where

$$
\begin{aligned}
&I\big(\boldsymbol{\beta}_{\mathrm{A}},\boldsymbol{\beta}_{\mathrm{B}},\mathrm{vech}(\boldsymbol{\Sigma}),\mathrm{vech}(\boldsymbol{\Sigma}'),\sigma^2\big)^{-1}_{\infty}\\
&=\begin{bmatrix}
\dfrac{\boldsymbol{\Sigma}}{m}+\dfrac{\boldsymbol{\Sigma}'}{m'} & \boldsymbol{O} & \boldsymbol{O} & \boldsymbol{O} & \boldsymbol{O}\\
\boldsymbol{O} & \dfrac{\sigma^2\boldsymbol{C}_{\boldsymbol{\beta}_{\mathrm{B}}}}{mm'n} & \boldsymbol{O} & \boldsymbol{O} & \boldsymbol{O}\\
\boldsymbol{O} & \boldsymbol{O} & \dfrac{2\boldsymbol{D}_{d_{\mathrm{A}}}^{+}(\boldsymbol{\Sigma}\otimes\boldsymbol{\Sigma})\boldsymbol{D}_{d_{\mathrm{A}}}^{+T}}{m} & \boldsymbol{O} & \boldsymbol{O}\\
\boldsymbol{O} & \boldsymbol{O} & \boldsymbol{O} & \dfrac{2\boldsymbol{D}_{d_{\mathrm{A}}}^{+}(\boldsymbol{\Sigma}'\otimes\boldsymbol{\Sigma}')\boldsymbol{D}_{d_{\mathrm{A}}}^{+T}}{m'} & \boldsymbol{O}\\
\boldsymbol{O} & \boldsymbol{O} & \boldsymbol{O} & \boldsymbol{O} & \dfrac{2\sigma^4}{mm'n}
\end{bmatrix}.
\end{aligned}
$$

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:

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

where $I(\boldsymbol{\beta},\boldsymbol{\psi})_{\boldsymbol{\beta}\boldsymbol{\beta}}$ 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 $I(\boldsymbol{\beta},\boldsymbol{\psi})_{\boldsymbol{\psi}\boldsymbol{\psi}}$ is defined similarly. Then, under (A1)–(A3) and some additional regularity conditions

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

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

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

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

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

and

$$I(\boldsymbol{\beta},\boldsymbol{\psi})^{-1}_{\boldsymbol{\psi}\boldsymbol{\psi},\infty} \equiv \begin{bmatrix} \dfrac{2\boldsymbol{D}^{+}_{d_{\rm A}}(\boldsymbol{\Sigma}\otimes\boldsymbol{\Sigma})\boldsymbol{D}^{+T}_{d_{\rm A}}}{m} & \boldsymbol{O} & \boldsymbol{O} \\ \boldsymbol{O} & \dfrac{2\boldsymbol{D}^{+}_{d_{\rm A}}(\boldsymbol{\Sigma}'\otimes\boldsymbol{\Sigma}')\boldsymbol{D}^{+T}_{d_{\rm A}}}{m'} & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{O} & \dfrac{2\sigma^4}{mm'n} \end{bmatrix}.$$

Next note that

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

where

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

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

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

The establishment

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

is very similar.

S.1.7 Final Steps for the Derivation of Result 1

Let

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

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

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

where

$$\widehat{\boldsymbol{\theta}} = \big[(\widehat{\boldsymbol{\beta}}_{\rm A})^T \ (\widehat{\boldsymbol{\beta}}_{\rm B})^T \ \text{vech}(\widehat{\boldsymbol{\Sigma}})^T \ \text{vech}(\widehat{\boldsymbol{\Sigma}}')^T \ \widehat{\sigma}^2\big]^T$$

and

$$\boldsymbol{\theta}^0 = \big[(\boldsymbol{\beta}^0_{\rm A})^T \ (\boldsymbol{\beta}^0_{\rm B})^T \ \text{vech}(\boldsymbol{\Sigma}^0)^T \ \text{vech}(\boldsymbol{\Sigma}^0)^T \ \text{vech}((\boldsymbol{\Sigma}')^0)^T \ (\sigma^2)^0\big]^T.$$

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

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

As a consequence

$$\boldsymbol{a}^T\{I\big(\boldsymbol{\theta}^0\big)^{-1}_{\infty}\}^{-1/2}(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}^0)+r_{mm'n}(\boldsymbol{a}) \xrightarrow{D} N(0,\boldsymbol{a}^T\boldsymbol{a}) \tag{S.22}$$

where

$$\begin{aligned} r_{mm'n}(\boldsymbol{a}) &\equiv \boldsymbol{a}^T[\{I(\boldsymbol{\theta}^0)^{-1}\}^{-1/2} - \{I(\boldsymbol{\theta}^0)^{-1}_{\infty}\}^{-1/2}](\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^0) \\ &= \boldsymbol{a}^T[\boldsymbol{I} - \{I(\boldsymbol{\theta}^0)^{-1}_{\infty}\}^{-1/2}\{I(\boldsymbol{\theta}^0)^{-1}\}^{1/2}]\{I(\boldsymbol{\theta}^0)^{-1}\}^{1/2}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^0) \\ &= \Big([\{I(\boldsymbol{\theta}^0)^{-1}_{\infty}\}^{-1/2}\{I(\boldsymbol{\theta}^0)^{-1}\}^{1/2} - \boldsymbol{I}]^T\boldsymbol{a}\Big)^T \boldsymbol{Z}_{mm'n} \end{aligned}$$

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

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

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

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

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

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

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 we 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

$$\nabla_{\boldsymbol{\theta}}\ell(\boldsymbol{\theta}) = \begin{bmatrix} \boldsymbol{X}_{\mathrm{A}}^T\boldsymbol{V}^{-1}(\boldsymbol{Y} - \boldsymbol{X}_{\mathrm{A}}\boldsymbol{\beta}_{\mathrm{A}} - \boldsymbol{X}_{\mathrm{B}}\boldsymbol{\beta}_{\mathrm{B}}) \\ \boldsymbol{X}_{\mathrm{B}}^T\boldsymbol{V}^{-1}(\boldsymbol{Y} - \boldsymbol{X}_{\mathrm{A}}\boldsymbol{\beta}_{\mathrm{A}} - \boldsymbol{X}_{\mathrm{B}}\boldsymbol{\beta}_{\mathrm{B}}) \\ \frac{1}{2}\underset{(r,s)\in\mathcal{I}_{d_{\mathrm{A}}}}{\mathrm{stack}}\Big\{\mathrm{tr}\Big(\boldsymbol{V}^{-1}\breve{\boldsymbol{L}}_{(r,s)}\boldsymbol{V}^{-1}(\boldsymbol{Y} - \boldsymbol{X}_{\mathrm{A}}\boldsymbol{\beta}_{\mathrm{A}} - \boldsymbol{X}_{\mathrm{B}}\boldsymbol{\beta}_{\mathrm{B}})^{\otimes 2} - \boldsymbol{V}^{-1}\breve{\boldsymbol{L}}_{(r,s)}\Big)\Big\} \\ \frac{1}{2}\underset{(r,s)\in\mathcal{I}_{d_{\mathrm{A}}}}{\mathrm{stack}}\Big\{\mathrm{tr}\Big(\boldsymbol{V}^{-1}\breve{\boldsymbol{L}}'_{(r,s)}\boldsymbol{V}^{-1}(\boldsymbol{Y} - \boldsymbol{X}_{\mathrm{A}}\boldsymbol{\beta}_{\mathrm{A}} - \boldsymbol{X}_{\mathrm{B}}\boldsymbol{\beta}_{\mathrm{B}})^{\otimes 2} - \boldsymbol{V}^{-1}\breve{\boldsymbol{L}}'_{(r,s)}\Big)\Big\} \\ \frac{1}{2}\mathrm{tr}\Big(\boldsymbol{V}^{-2}(\boldsymbol{Y} - \boldsymbol{X}_{\mathrm{A}}\boldsymbol{\beta}_{\mathrm{A}} - \boldsymbol{X}_{\mathrm{B}}\boldsymbol{\beta}_{\mathrm{B}})^{\otimes 2} - \boldsymbol{V}^{-1}\Big) \end{bmatrix}$$

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,

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

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

$$\breve{\boldsymbol{L}}'_{(r,s)} \equiv \underset{1\le i,\underset{\sim}{i}\le m}{\mathrm{blockmatrix}}\left\{\underset{1\le i'\le m'}{\mathrm{blockdiag}}\Big(\boldsymbol{X}_{\mathrm{A}ii'}\Big(\boldsymbol{e}_r\boldsymbol{e}_s^T + I(r\ne s)\boldsymbol{e}_s\boldsymbol{e}_r^T\Big)\boldsymbol{X}_{\mathrm{A}\underset{\sim}{i}i'}^T\Big)\right\}.$$

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

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

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

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

where

$$\begin{aligned} \boldsymbol{w}_{\text{A}r} &\equiv r\text{th column of } (\boldsymbol{V}^{1/2}_{\text{loose}})^T \boldsymbol{V}^{-1}\boldsymbol{X}_{\text{A}}, \quad 1 \le r \le d_{\text{A}}, \\ \boldsymbol{w}_{\text{B}r} &\equiv r\text{th column of } (\boldsymbol{V}^{1/2}_{\text{loose}})^T \boldsymbol{V}^{-1}\boldsymbol{X}_{\text{B}}, \quad 1 \le r \le d_{\text{B}}, \\ \boldsymbol{W}_{(r,s)} &= (\boldsymbol{V}^{1/2}_{\text{loose}})^T \boldsymbol{V}^{-1} \breve{\boldsymbol{L}}_{(r,s)} \boldsymbol{V}^{-1}\boldsymbol{V}^{1/2}_{\text{loose}}, \quad (r,s) \in \mathcal{I}_{d_{\text{A}}}, \\ \boldsymbol{W}'_{(r,s)} &= (\boldsymbol{V}^{1/2}_{\text{loose}})^T \boldsymbol{V}^{-1} \breve{\boldsymbol{L}}'_{(r,s)} \boldsymbol{V}^{-1}\boldsymbol{V}^{1/2}_{\text{loose}}, \quad (r,s) \in \mathcal{I}_{d_{\text{A}}} \\ \text{and} \quad \boldsymbol{W}_{\sigma^2} &= (\boldsymbol{V}^{1/2}_{\text{loose}})^T \boldsymbol{V}^{-2}\boldsymbol{V}^{1/2}_{\text{loose}}. \end{aligned}$$

Let

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

be a vector of sample size quantities and 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

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

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

$$
\begin{aligned}
\xi_t(m,m',\boldsymbol{n}) \equiv\, & (\boldsymbol{a}_{\rm norm})_1 m^{-1/2}(\boldsymbol{w}^0_{\rm A1})_t(\boldsymbol{z})_t + \ldots + (\boldsymbol{a}_{\rm norm})_{d_{\rm A}} m^{-1/2}(\boldsymbol{w}^0_{{\rm A}d_{\rm A}})_t(\boldsymbol{z})_t \\
& +(\boldsymbol{a}_{\rm norm})_{d_{\rm A}+1}(m^2n)^{-1/2}(\boldsymbol{w}^0_{\rm B1})_t(\boldsymbol{z})_t + \ldots + (\boldsymbol{a}_{\rm norm})_{d_{\rm A}+d_{\rm B}}(m^2n)^{-1/2}(\boldsymbol{w}^0_{{\rm B}d_{\rm B}})_t(\boldsymbol{z})_t \\
& +\tfrac{1}{2}(\boldsymbol{a}_{\rm norm})_{d_{\rm A}+d_{\rm B}+1} m^{-1/2}\big(\boldsymbol{W}^0_{(1,1)}(\boldsymbol{z}^{\otimes 2}-\boldsymbol{I})\big)_{tt} \\
& +\ldots+\tfrac{1}{2}(\boldsymbol{a}_{\rm norm})_{d_{\rm A}+d_{\rm B}+\frac{1}{2}d_{\rm A}(d_{\rm A}+1)} m^{-1/2}\big(\boldsymbol{W}^0_{(d_{\rm A},d_{\rm A})}(\boldsymbol{z}^{\otimes 2}-\boldsymbol{I})\big)_{tt} \\
& +\tfrac{1}{2}(\boldsymbol{a}_{\rm norm})_{d_{\rm A}+d_{\rm B}+\frac{1}{2}d_{\rm A}(d_{\rm A}+1)+1} m^{-1/2}\big((\boldsymbol{W}')^0_{(1,1)}(\boldsymbol{z}^{\otimes 2}-\boldsymbol{I})\big)_{tt} \\
& +\ldots+\tfrac{1}{2}(\boldsymbol{a}_{\rm norm})_{d_{\rm A}+d_{\rm B}+\frac{1}{2}d_{\rm A}(d_{\rm A}+1)+\frac{1}{2}d_{\rm B}(d_{\rm B}+1)} m^{-1/2}\big((\boldsymbol{W}')^0_{(d_{\rm B},d_{\rm B})}(\boldsymbol{z}^{\otimes 2}-\boldsymbol{I})\big)_{tt} \\
& +\tfrac{1}{2}(\boldsymbol{a}_{\rm norm})_{d_{\rm A}+d_{\rm B}+\frac{1}{2}d_{\rm A}(d_{\rm A}+1)+\frac{1}{2}d_{\rm B}(d_{\rm B}+1)+1}(m^2n)^{-1/2}\big(\boldsymbol{W}^0_{\sigma^2}(\boldsymbol{z}^{\otimes 2}-\boldsymbol{I})\big)_{tt}
\end{aligned}
$$

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

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

Then

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

is an array of martingale differences.

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

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

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

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

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

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

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

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 assumption (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

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

Standard likelihood theory arguments then lead to

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

S.1.9 Proofs of Lemmas

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

S.1.9.1 Proof of Lemma 1

It is trivial to show that Lemma 1 holds when $d=1$. We now apply induction on d to prove Lemma 1 in general.

For each $1\le r,s\le d$ let

$$\widetilde{A}_{rs}\equiv a_{d+1-r,d+1-s}$$

Then, according to the $\widetilde{A}_{rs}$ notation the entries of $\boldsymbol{A}$ have their row and column indices starting from the bottom right entry. For example,

$$\boldsymbol{A}_2=\begin{bmatrix}\widetilde{A}_{12}&\widetilde{A}_{21}\\\widetilde{A}_{22}&\widetilde{A}_{11}\end{bmatrix}\quad\text{and}\quad\boldsymbol{A}_3=\begin{bmatrix}\widetilde{A}_{33}&\widetilde{A}_{32}&\widetilde{A}_{31}\\\widetilde{A}_{23}&\widetilde{A}_{12}&\widetilde{A}_{21}\\\widetilde{A}_{13}&\widetilde{A}_{22}&\widetilde{A}_{11}\end{bmatrix}.$$

This labelling is more amenable to use of a key recursion-type formula for $\boldsymbol{D}_d^T(\boldsymbol{A}_d\otimes\boldsymbol{A}_d)\boldsymbol{D}_d$. This is stated next.

Write

$$\boldsymbol{A}_{d+1}=\begin{bmatrix}\widetilde{A}_{d+1,d+1}&\widetilde{\boldsymbol{a}}_{d:1,d+1}^T\\\widetilde{\boldsymbol{a}}_{d:1,d+1}&\boldsymbol{A}_d\end{bmatrix}\quad\text{where}\quad\widetilde{\boldsymbol{a}}_{d:1,d+1}\equiv\begin{bmatrix}\widetilde{A}_{d,d+1}\\\vdots\\\widetilde{A}_{1,d+1}\end{bmatrix}.$$

Then, from Theorem 15 in Chapter 3 of Magnus & Neudecker (1999),

$$\begin{aligned}&\boldsymbol{D}_{d+1}^T(\boldsymbol{A}_{d+1}\otimes\boldsymbol{A}_{d+1})\boldsymbol{D}_{d+1}\\&=\begin{bmatrix}\widetilde{A}_{d+1,d+1}^2&2\widetilde{A}_{d+1,d+1}\widetilde{\boldsymbol{a}}_{d:1,d+1}^T&(\widetilde{\boldsymbol{a}}_{d:1,d+1}\otimes\widetilde{\boldsymbol{a}}_{d:1,d+1})^T\boldsymbol{D}_d\\2\widetilde{A}_{d+1,d+1}\widetilde{\boldsymbol{a}}_{d:1,d+1}&2(\widetilde{A}_{d+1,d+1}\boldsymbol{A}_d+\widetilde{\boldsymbol{a}}_{d:1,d+1}\widetilde{\boldsymbol{a}}_{d:1,d+1}^T)&2(\widetilde{\boldsymbol{a}}_{d:1,d+1}^T\otimes\boldsymbol{A}_d)\boldsymbol{D}_d\\\boldsymbol{D}_d^T(\widetilde{\boldsymbol{a}}_{d:1,d+1}\otimes\widetilde{\boldsymbol{a}}_{d:1,d+1})&2\boldsymbol{D}_d^T(\widetilde{\boldsymbol{a}}_{d:1,d+1}\otimes\boldsymbol{A}_d)&\boldsymbol{D}_d^T(\boldsymbol{A}_d\otimes\boldsymbol{A}_d)\boldsymbol{D}_d\end{bmatrix}.\end{aligned}\tag{S.27}$$

Consider the following partition of $\boldsymbol{B}_{d+1}$ that is conformable with the right-hand side of (S.27):

$$\boldsymbol{B}_{d+1} = \begin{bmatrix} b_{11} & \boldsymbol{b}_{d+1,21}^T & \boldsymbol{b}_{d+1,31}^T \\ \boldsymbol{b}_{d+1,21} & \boldsymbol{B}_{d+1,22} & \boldsymbol{B}_{d+1,32}^T \\ \boldsymbol{b}_{d+1,31} & \boldsymbol{B}_{d+1,32} & \boldsymbol{B}_{d+1,33} \end{bmatrix}. \tag{S.28}$$

The dimensions of the matrices on the right-hand side of (S.28) are as follows:

$$b_{11}\ (1\times 1), \quad \boldsymbol{b}_{d+1,21}\ (d\times 1), \quad \boldsymbol{B}_{d+1,22}\ (d\times d), \quad \boldsymbol{b}_{d+1,31}\ \big(\tfrac{1}{2}d(d+1)\times 1\big),$$

$$\boldsymbol{B}_{d+1,32}\ \big(\tfrac{1}{2}d(d+1)\times d\big) \quad \text{and} \quad \boldsymbol{B}_{d+1,33}\ \big(\tfrac{1}{2}d(d+1)\times \tfrac{1}{2}d(d+1)\big).$$

Since

$$\text{vech}(\boldsymbol{A}_{d+1}) = \begin{bmatrix} \widetilde{A}_{d+1,d+1} \\ \widetilde{\boldsymbol{a}}_{d:1,d+1} \\ \text{vech}(\boldsymbol{A}_d) \end{bmatrix}$$

we have

$$\begin{aligned} &\text{vech}(\boldsymbol{A}_{d+1})\text{vech}(\boldsymbol{A}_{d+1})^T \\ &= \begin{bmatrix} \widetilde{A}^2_{d+1,d+1} & \widetilde{A}_{d+1,d+1}\widetilde{\boldsymbol{a}}^T_{d:1,d+1} & \widetilde{A}_{d+1,d+1}\text{vech}(\boldsymbol{A}_d)^T \\ \widetilde{A}_{d+1,d+1}\widetilde{\boldsymbol{a}}_{d:1,d+1} & \widetilde{\boldsymbol{a}}_{d:1,d+1}\widetilde{\boldsymbol{a}}^T_{d:1,d+1} & \widetilde{\boldsymbol{a}}_{d:1,d+1}\text{vech}(\boldsymbol{A}_d)^T \\ \widetilde{A}_{d+1,d+1}\text{vech}(\boldsymbol{A}_d) & \text{vech}(\boldsymbol{A}_d)\widetilde{\boldsymbol{a}}^T_{d:1,d+1} & \text{vech}(\boldsymbol{A}_d)\text{vech}(\boldsymbol{A}_d)^T \end{bmatrix}. \end{aligned} \tag{S.29}$$

Then proof by induction of Lemma 1 is complete if it is shown that application of the rules in Table S.1 to the entries of each block on the right-hand side of (S.29) leads to corresponding blocks in (S.27). We treat each of the blocks in turn:

The b_{11} Block

From Table S.1, $b_{11} = \widetilde{A}^2_{d+1,d+1}$ which equals the $(1,1)$ entry on the right-hand side of (S.27).

The $\boldsymbol{b}_{d+1,21}$ Block

The rth entry of $\widetilde{\boldsymbol{a}}_{d:1,d+1}$ is $\widetilde{A}_{d+1-r,d+1}$. Therefore, from Table S.1, the rth entry of $\boldsymbol{b}_{d+1,21}$ is

$$2\,\widetilde{A}_{d+1,d+1-r}\widetilde{A}_{d+1,d+1} = \text{the } r\text{th entry of } 2\,\widetilde{A}_{d+1,d+1}\widetilde{\boldsymbol{a}}_{d:1,d+1}.$$

Hence, $\boldsymbol{b}_{d+1,21} = 2\,\widetilde{A}_{d+1,d+1}\widetilde{\boldsymbol{a}}_{d:1,d+1}$, which corresponds to the $(2,1)$ block on the right-hand side of (S.27).

The $\boldsymbol{B}_{d+1,22}$ Block

The (r,s)th entry of $\widetilde{\boldsymbol{a}}_{d:1,d+1}\widetilde{\boldsymbol{a}}^T_{d:1,d+1}$ is $\widetilde{A}_{d+1-r,d+1}\widetilde{A}_{d+1-s,d+1}$. Therefore, from Table S.1, the (r,s)th entry of $\boldsymbol{B}_{d+1,22}$ is

$$\begin{aligned} &2\Big(\widetilde{A}_{d+1-r,d+1-s}\widetilde{A}_{d+1,d+1} + \widetilde{A}_{d+1-r,d+1}\widetilde{A}_{d+1,d+1-s}\Big) \\ &\qquad = \text{the } (r,s)\text{th entry of } 2\Big(\widetilde{A}_{d+1,d+1}\boldsymbol{A}_d + \widetilde{\boldsymbol{a}}_{d:1,d+1}\widetilde{\boldsymbol{a}}^T_{d:1,d+1}\Big). \end{aligned}$$

Therefore, $\boldsymbol{B}_{d+1,22} = 2\big(\widetilde{A}_{d+1,d+1}\boldsymbol{A}_d + \widetilde{\boldsymbol{a}}_{d:1,d+1}\widetilde{\boldsymbol{a}}^T_{d:1,d+1}\big)$, which corresponds to the $(2,2)$ block on the right-hand side of (S.27).

The $\boldsymbol{b}_{d+1,31}$ Block

Use of the result $\boldsymbol{v} \otimes \boldsymbol{w} = \text{vec}(\boldsymbol{w}\boldsymbol{v}^T)$ for column vectors $\boldsymbol{v}$ and $\boldsymbol{w}$, as well as Theorem 14(a) of Chapter 3 of Magnus & Neudecker (1999), leads to the $(3,1)$ block on the right-hand side of (S.27) having the following alternative expression:

$$\text{vech}\big(2\widetilde{\boldsymbol{a}}_{d:1,d+1}\widetilde{\boldsymbol{a}}_{d:1,d+1}^T - \text{diag}(\widetilde{\boldsymbol{a}}_{d:1,d+1} \odot \widetilde{\boldsymbol{a}}_{d:1,d+1})\big). \tag{S.30}$$

Also, note that the $(3,1)$ block of $\text{vech}(\boldsymbol{A}_{d+1})\text{vech}(\boldsymbol{A}_{d+1})^T$ is $\text{vech}(\widetilde{A}_{d+1,d+1}\boldsymbol{A}_d)$ and that the (r,s)th entry of $\widetilde{A}_{d+1,d+1}\boldsymbol{A}_d$ is $\widetilde{A}_{d+1,d+1}\widetilde{A}_{d+1-r,d+1-s}$. Therefore, from Table S.1, $\boldsymbol{b}_{d+1,31} = \text{vech}(\boldsymbol{G}_d)$ where $\boldsymbol{G}_d$ is the $d \times d$ matrix with (r,s)th entry $\widetilde{A}^2_{d+1,d+1-r}$ for $r = s$ and $2\,\widetilde{A}_{d+1,d+1-r}\widetilde{A}_{d+1,d+1-s}$ for $r \neq s$. It follows that $\boldsymbol{G}_d = 2\widetilde{\boldsymbol{a}}_{d:1,d+1}\widetilde{\boldsymbol{a}}_{d:1,d+1}^T - \text{diag}(\widetilde{\boldsymbol{a}}_{d:1,d+1} \odot \widetilde{\boldsymbol{a}}_{d:1,d+1})$ and, hence, $\boldsymbol{b}_{d+1,31}$ equals the expression in (S.30).

The $\boldsymbol{B}_{d+1,32}$ Block

First note that the $(3,2)$ block on the right-hand side of (S.27) can be written as

$$2\boldsymbol{D}_d^T\big[\widetilde{\boldsymbol{a}}_{d:1,d+1} \otimes \widetilde{\boldsymbol{a}}_{d:1,1} \;\cdots\; \widetilde{\boldsymbol{a}}_{d:1,d+1} \otimes \widetilde{\boldsymbol{a}}_{d:1,d}\big]. \tag{S.31}$$

As was done for treatment of the $\boldsymbol{b}_{d+1,31}$ block, we apply $\boldsymbol{v} \otimes \boldsymbol{w} = \text{vec}(\boldsymbol{w}\boldsymbol{v}^T)$ and Theorem 14(a) of Chapter 3 of Magnus & Neudecker (1999) to the columns of (S.31). This leads to the $(3,2)$ block on the right-hand side of (S.27) equalling

$$\begin{aligned} 2\big[&\text{vech}\big(\widetilde{\boldsymbol{a}}_{d:1,d+1}\widetilde{\boldsymbol{a}}_{d:1,1}^T + \widetilde{\boldsymbol{a}}_{d:1,1}\widetilde{\boldsymbol{a}}_{d:1,d+1}^T - \text{diag}(\widetilde{\boldsymbol{a}}_{d:1,d+1} \odot \widetilde{\boldsymbol{a}}_{d:1,1})\big) \\ &\cdots \text{vech}\big(\widetilde{\boldsymbol{a}}_{d:1,d+1}\widetilde{\boldsymbol{a}}_{d:1,d}^T + \widetilde{\boldsymbol{a}}_{d:1,d}\widetilde{\boldsymbol{a}}_{d:1,d+1}^T - \text{diag}(\widetilde{\boldsymbol{a}}_{d:1,d+1} \odot \widetilde{\boldsymbol{a}}_{d:1,d})\big)\big]. \end{aligned} \tag{S.32}$$

Now note that the (3,2) block of $\text{vech}(\boldsymbol{A}_{d+1})\text{vech}(\boldsymbol{A}_{d+1})^T$ is

$$\big[\text{vech}(\widetilde{A}_{d,d+1}\boldsymbol{A}_d) \;\cdots\; \text{vech}(\widetilde{A}_{1,d+1}\boldsymbol{A}_d)\big]$$

Therefore, from Table S.1,

$$\boldsymbol{B}_{d+1,32} = [\text{vech}(\boldsymbol{H}_{d1}) \;\cdots\; \text{vech}(\boldsymbol{H}_{dd})]$$

where, for each $1 \le v \le d$, $\boldsymbol{H}_{dv}$ is the $d \times d$ matrix with (r,s)th entry equal to

$$2\big(\widetilde{A}_{d+1-v,d+1-r}\widetilde{A}_{d+1,d+1-s} + \widetilde{A}_{d+1-v,d+1-s}\widetilde{A}_{d+1,d+1-r}\big) \quad \text{for} \quad r \neq s$$

and

$$2\widetilde{A}_{d+1-v,d+1-r}\widetilde{A}_{d+1,d+1-r} \text{ for } r = s.$$

These facts imply that

$$\boldsymbol{H}_{dv} = 2\left\{\widetilde{\boldsymbol{a}}_{d:1,d+1}\widetilde{\boldsymbol{a}}_{d:1,v}^T + \widetilde{\boldsymbol{a}}_{d:1,v}\widetilde{\boldsymbol{a}}_{d:1,d+1}^T - \text{diag}(\widetilde{\boldsymbol{a}}_{d:1,d+1} \odot \widetilde{\boldsymbol{a}}_{d:1,v})\right\} \text{ for } 1 \le v \le d.$$

Hence, $\boldsymbol{B}_{d+1,32}$ equals the expression in (S.32).

The $\boldsymbol{B}_{d+1,33}$ Block

It follows from the inductive hypothesis that $\boldsymbol{B}_{d+1,33} = \boldsymbol{D}_d^T(\boldsymbol{A}_d \otimes \boldsymbol{A}_d)\boldsymbol{D}_d$, which corresponds to the $(3,3)$ block on the right-hand side of (S.27).

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 $\boldsymbol{V}$ $(d \times n)$, is

$$(\boldsymbol{S} + \boldsymbol{U}\boldsymbol{T}\boldsymbol{V})^{-1} = \boldsymbol{S}^{-1} - \boldsymbol{S}^{-1}\boldsymbol{U}(\boldsymbol{T}^{-1} + \boldsymbol{V}\boldsymbol{S}^{-1}\boldsymbol{U})^{-1}\boldsymbol{V}\boldsymbol{S}^{-1}. \tag{S.33}$$

Application of (S.33) with

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

leads to

$$(\boldsymbol{X}\boldsymbol{A}\boldsymbol{X}^T + \lambda\boldsymbol{I})^{-1} = (1/\lambda)\boldsymbol{I}_n - (1/\lambda^2)\boldsymbol{X}(\boldsymbol{A}^{-1} + \boldsymbol{X}^T\boldsymbol{X}/\lambda)^{-1}\boldsymbol{X}^T. \tag{S.34}$$

Therefore,

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

Next we apply Woodbury's matrix identity (S.33) to $\{\boldsymbol{I}_d + \lambda(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{A}^{-1}\}^{-1}$ with

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

to obtain

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

Plugging this into the above set of equations we have

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

and the lemma is proven.

S.1.9.3 Proof of Lemma 3

It follows from (S.34) that

$$\begin{aligned}
\boldsymbol{X}^T(\boldsymbol{X}\boldsymbol{A}\boldsymbol{X}^T + \lambda\boldsymbol{I})^{-2}\boldsymbol{X} &= (1/\lambda^2)\boldsymbol{X}^T\boldsymbol{X} - (2/\lambda^3)\boldsymbol{X}^T\boldsymbol{X}(\boldsymbol{A}^{-1} + \boldsymbol{X}^T\boldsymbol{X}/\lambda)^{-1}\boldsymbol{X}^T\boldsymbol{X} \\
&\quad + (1/\lambda^4)\boldsymbol{X}^T\boldsymbol{X}(\boldsymbol{A}^{-1} + \boldsymbol{X}^T\boldsymbol{X}/\lambda)^{-1}\boldsymbol{X}^T\boldsymbol{X}(\boldsymbol{A}^{-1} + \boldsymbol{X}^T\boldsymbol{X}/\lambda)^{-1}\boldsymbol{X}^T\boldsymbol{X}
\end{aligned} \tag{S.35}$$

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

$$(\boldsymbol{A}^{-1} + \boldsymbol{X}^T\boldsymbol{X}/\lambda)^{-1}\boldsymbol{X}^T\boldsymbol{X} = \lambda\boldsymbol{I} - \lambda^2(\boldsymbol{X}^T\boldsymbol{X})^{-1}\{\boldsymbol{A} + \lambda(\boldsymbol{X}^T\boldsymbol{X})^{-1}\}^{-1} \tag{S.36}$$

Triple substitution of (S.36) into (S.35) and simplification yields the stated result.

S.1.9.4 Proof of Lemma 4

Standard block matrix inverse results lead to

$$\begin{bmatrix} \boldsymbol{A} & \boldsymbol{B} \\ \boldsymbol{B} & \boldsymbol{A} \end{bmatrix}^{-1} = \begin{bmatrix} (\boldsymbol{A} - \boldsymbol{B}\boldsymbol{A}^{-1}\boldsymbol{B})^{-1} & -(\boldsymbol{A} - \boldsymbol{B}\boldsymbol{A}^{-1}\boldsymbol{B})^{-1}\boldsymbol{B}\boldsymbol{A}^{-1} \\ -(\boldsymbol{A} - \boldsymbol{B}\boldsymbol{A}^{-1}\boldsymbol{B})^{-1}\boldsymbol{B}\boldsymbol{A}^{-1} & (\boldsymbol{A} - \boldsymbol{B}\boldsymbol{A}^{-1}\boldsymbol{B})^{-1} \end{bmatrix}$$

where symmetry of $\boldsymbol{B}$ is being used. Therefore, the left-hand side of the Lemma 4 assertion is

$$2(\boldsymbol{A} - \boldsymbol{B}\boldsymbol{A}^{-1}\boldsymbol{B})^{-1}(\boldsymbol{I} - \boldsymbol{B}\boldsymbol{A}^{-1}). \tag{S.37}$$

Now note that

$$\boldsymbol{B}+\boldsymbol{B}\boldsymbol{A}^{-1}\boldsymbol{B}=\boldsymbol{A}+\boldsymbol{B}-\boldsymbol{A}+\boldsymbol{B}\boldsymbol{A}^{-1}\boldsymbol{B}=\boldsymbol{A}\boldsymbol{B}^{-1}(\boldsymbol{B}+\boldsymbol{B}\boldsymbol{A}^{-1}\boldsymbol{B})-(\boldsymbol{A}-\boldsymbol{B}\boldsymbol{A}^{-1}\boldsymbol{B}) \tag{S.38}$$

which leads to

$$\boldsymbol{A}-\boldsymbol{B}\boldsymbol{A}^{-1}\boldsymbol{B}=(\boldsymbol{A}\boldsymbol{B}^{-1}-\boldsymbol{I})(\boldsymbol{B}+\boldsymbol{B}\boldsymbol{A}^{-1}\boldsymbol{B})$$

and then

$$-(\boldsymbol{A}-\boldsymbol{B}\boldsymbol{A}^{-1}\boldsymbol{B})(\boldsymbol{B}+\boldsymbol{B}\boldsymbol{A}^{-1}\boldsymbol{B})^{-1}=\boldsymbol{I}-\boldsymbol{A}\boldsymbol{B}^{-1}.$$

This implies that

$$(\boldsymbol{A}-\boldsymbol{B}\boldsymbol{A}^{-1}\boldsymbol{B})\big\{\boldsymbol{B}^{-1}-(\boldsymbol{B}+\boldsymbol{B}\boldsymbol{A}^{-1}\boldsymbol{B})^{-1}\big\}=(\boldsymbol{A}-\boldsymbol{B}\boldsymbol{A}^{-1}\boldsymbol{B})\boldsymbol{B}^{-1}+\boldsymbol{I}-\boldsymbol{A}\boldsymbol{B}^{-1}=\boldsymbol{I}-\boldsymbol{B}\boldsymbol{A}^{-1}. \tag{S.39}$$

Combining (S.37) and (S.39), the left-hand side of the Lemma 4 assertion is

$$2\big\{\boldsymbol{B}^{-1}-(\boldsymbol{B}+\boldsymbol{B}\boldsymbol{A}^{-1}\boldsymbol{B})^{-1}\big\}=2(\boldsymbol{A}+\boldsymbol{B})^{-1}.$$

The last step corresponds to *Hua's identity*, which is an established result in matrix algebra and ring theory extensions. This proves the lemma.

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 $\boldsymbol{X}_{ii'}=\boldsymbol{1}_n$ 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

$$\boldsymbol{1}_{m'n}^T\boldsymbol{Q}_{mm'}^{ii}\boldsymbol{1}_{m'n}=$$

$$\frac{mm'M'\{(m-1)M'+m'M\}+m'\{(m-1)M'+mM'+m'M\}(\lambda/n)+m'(\lambda/n)^2}{(mM'+\lambda/n)\{m'M(m'M+mM')+(mM'+2m'M)(\lambda/n)+(\lambda/n)^2\}}$$

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

$$\lim_{n\to\infty}\left(\boldsymbol{1}_{m'n}^T\boldsymbol{Q}_{mm'}^{ii}\boldsymbol{1}_{m'n}\right)=\frac{1}{M}-\frac{M'}{mm'M}\left(\frac{M}{m}+\frac{M'}{m'}\right)^{-1}\quad\text{for all}\quad m,m'\in\mathbb{N} \tag{S.40}$$

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

$$\lim_{n\to\infty}\left(\boldsymbol{1}_{mm'n}^T\boldsymbol{Q}_{mm'}^{-1}\boldsymbol{1}_{mm'n}\right)=\left(\frac{M}{m}+\frac{M'}{m'}\right)^{-1}\quad\text{for all}\quad m,m'\in\mathbb{N}. \tag{S.41}$$

The result

$$\lim_{n\to\infty}\left(\boldsymbol{1}_{m'n}^T\boldsymbol{Q}_{mm'}^{i\underset{\sim}{i}}\boldsymbol{1}_{m'n}\right)=-\frac{M'}{mm'M}\left(\frac{M}{m}+\frac{M'}{m'}\right)^{-1}\quad\text{for}\quad i\ne\underset{\sim}{i} \tag{S.42}$$

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

$$\boldsymbol{Q}_{21}=\begin{bmatrix}\boldsymbol{X}_{11}(\boldsymbol{M}+\boldsymbol{M}')\boldsymbol{X}_{11}^T+\lambda\boldsymbol{I} & \boldsymbol{X}_{11}\boldsymbol{M}'\boldsymbol{X}_{21}^T\\ \boldsymbol{X}_{21}\boldsymbol{M}'\boldsymbol{X}_{11}^T & \boldsymbol{X}_{21}(\boldsymbol{M}+\boldsymbol{M}')\boldsymbol{X}_{21}^T+\lambda\boldsymbol{I}\end{bmatrix}$$

and so, using Corollary 2.1.(c),

$$\begin{aligned}\boldsymbol{X}_{11}^T\boldsymbol{Q}_{21}^{11}\boldsymbol{X}_{11}&=\boldsymbol{X}_{11}^T\big[\boldsymbol{X}_{11}(\boldsymbol{M}+\boldsymbol{M}')\boldsymbol{X}_{11}^T\\&\qquad-\boldsymbol{X}_{11}\boldsymbol{M}'\boldsymbol{X}_{21}^T\{\boldsymbol{X}_{21}(\boldsymbol{M}+\boldsymbol{M}')\boldsymbol{X}_{21}^T+\lambda\boldsymbol{I}\}^{-1}\boldsymbol{X}_{21}\boldsymbol{M}'\boldsymbol{X}_{11}^T+\lambda\boldsymbol{I}\big]^{-1}\boldsymbol{X}_{11}\\&=\big[\boldsymbol{M}+\boldsymbol{M}'-\boldsymbol{M}'\{\boldsymbol{M}+\boldsymbol{M}'+\lambda(\boldsymbol{X}_{21}^T\boldsymbol{X}_{21})^{-1}\}^{-1}\boldsymbol{M}'+\lambda(\boldsymbol{X}_{11}^T\boldsymbol{X}_{11})^{-1}\big]^{-1}\\&\xrightarrow{P}\{\boldsymbol{M}+\boldsymbol{M}'-\boldsymbol{M}'(\boldsymbol{M}+\boldsymbol{M}')^{-1}\boldsymbol{M}'\}^{-1}=\Big[\boldsymbol{M}+\boldsymbol{M}'\{\boldsymbol{I}-(\boldsymbol{M}+\boldsymbol{M}')^{-1}\boldsymbol{M}'\}\Big]^{-1}.\end{aligned}$$

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 equalling $\big\{\boldsymbol{M}+\boldsymbol{M}'(\boldsymbol{M}+\boldsymbol{M}')^{-1}\boldsymbol{M}\big\}^{-1}$. Application of Woodbury's matrix identity (S.33) 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 equalling

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

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),

$$\begin{aligned}
&\boldsymbol{X}_{11}^T\boldsymbol{Q}_{21}^{12}\boldsymbol{X}_{21} = \\
&\quad -\boldsymbol{X}_{11}^T\big[\boldsymbol{X}_{11}(\boldsymbol{M}+\boldsymbol{M}')\boldsymbol{X}_{11}^T - \boldsymbol{X}_{11}\boldsymbol{M}'\boldsymbol{X}_{21}^T\{\boldsymbol{X}_{21}(\boldsymbol{M}+\boldsymbol{M}')\boldsymbol{X}_{21}^T+\lambda\boldsymbol{I}\}^{-1}\boldsymbol{X}_{21}\boldsymbol{M}'\boldsymbol{X}_{11}^T\big]^{-1}\\
&\qquad\quad \times\boldsymbol{X}_{11}\boldsymbol{M}'\boldsymbol{X}_{21}^T\{\boldsymbol{X}_{21}(\boldsymbol{M}+\boldsymbol{M}')\boldsymbol{X}_{21}^T+\lambda\boldsymbol{I}\}^{-1}\boldsymbol{X}_{21}\\
&\quad \xrightarrow{P} -\big\{\boldsymbol{M}+\boldsymbol{M}'-\boldsymbol{M}'(\boldsymbol{M}+\boldsymbol{M}')^{-1}\boldsymbol{M}'\big\}^{-1}\boldsymbol{M}'(\boldsymbol{M}+\boldsymbol{M}')^{-1}\\
&\quad = -\tfrac{1}{2}\boldsymbol{M}^{-1}\boldsymbol{M}'(\tfrac{1}{2}\boldsymbol{M}+\boldsymbol{M}')^{-1}.
\end{aligned}$$

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 $m\geq 2$ and $m'=1$ via induction on m. Let

$$\begin{gathered}
\boldsymbol{Q}_{m+1,1} = \begin{bmatrix} \boldsymbol{Q}_{m1} & \boldsymbol{R}_m \\ \boldsymbol{R}_m^T & \boldsymbol{S}_m \end{bmatrix} \text{ where } \boldsymbol{S}_m \equiv \boldsymbol{X}_{m+1,1}(\boldsymbol{M}+\boldsymbol{M}')\boldsymbol{X}_{m+1,1}^T+\lambda\boldsymbol{I},\\
\boldsymbol{R}_m \equiv \boldsymbol{X}_{1:m,1}\boldsymbol{M}'\boldsymbol{X}_{m+1,1}^T \quad \text{and} \quad \boldsymbol{X}_{1:m,1} \equiv \underset{1\leq i\leq m}{\text{stack}}(\boldsymbol{X}_{i1}).
\end{gathered} \tag{S.43}$$

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

$$\begin{aligned}
\boldsymbol{X}_{m+1,1}^T\boldsymbol{Q}_{m+1,1}^{m+1,m+1}\boldsymbol{X}_{m+1,1} &= \boldsymbol{X}_{m+1,1}^T\Big(\boldsymbol{S}_m - \boldsymbol{R}_m^T\boldsymbol{Q}_{m1}^{-1}\boldsymbol{R}_m\Big)^{-1}\boldsymbol{X}_{m+1,1}\\
&= \boldsymbol{X}_{m+1,1}^T\Big\{\boldsymbol{X}_{m+1,1}(\boldsymbol{M}+\boldsymbol{M}')\boldsymbol{X}_{m+1,1}^T\\
&\qquad -\boldsymbol{X}_{m+1,1}\boldsymbol{M}'\boldsymbol{X}_{1:m,1}^T\boldsymbol{Q}_{m1}^{-1}\boldsymbol{X}_{1:m,1}\boldsymbol{M}'\boldsymbol{X}_{m+1,1}^T+\lambda\boldsymbol{I}\Big\}^{-1}\boldsymbol{X}_{m+1,1}\\
&= \Big\{\boldsymbol{M}+\boldsymbol{M}'-\boldsymbol{M}'\boldsymbol{X}_{1:m,1}^T\boldsymbol{Q}_{m1}^{-1}\boldsymbol{X}_{1:m,1}\boldsymbol{M}'+\lambda(\boldsymbol{X}_{m+1,1}^T\boldsymbol{X}_{m+1,1})^{-1}\Big\}^{-1}\\
&\xrightarrow{P} \Big\{\boldsymbol{M}+\boldsymbol{M}'-\boldsymbol{M}'\Big(\frac{1}{m}\boldsymbol{M}+\boldsymbol{M}'\Big)^{-1}\boldsymbol{M}'\Big\}^{-1}\\
&= \boldsymbol{M}^{-1} - \frac{1}{m+1}\boldsymbol{M}^{-1}\boldsymbol{M}'\Big(\frac{1}{m+1}\boldsymbol{M}+\boldsymbol{M}'\Big)^{-1}.
\end{aligned}$$

Analogous arguments for other partitions of $\boldsymbol{Q}_{m+1,1}$ lead to the same convergence in probability limit for $\boldsymbol{X}_{i,1}^T\boldsymbol{Q}_{m+1,1}^{i,i}\boldsymbol{X}_{i,1}$ for each $1\leq i\leq m+1$. Therefore, by induction, Lemma 5(b) holds for all $m\geq 2$ and $m'=1$.

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

$$\begin{aligned}
\boldsymbol{X}_{1:m,1}^T\boldsymbol{Q}_{m+1,1}^{1:m,m+1}\boldsymbol{X}_{m+1,1} &= -\boldsymbol{X}_{1:m,1}^T\boldsymbol{Q}_{m1}^{-1}\boldsymbol{R}_m\Big(\boldsymbol{S}_m - \boldsymbol{R}_m^T\boldsymbol{Q}_{m1}^{-1}\boldsymbol{R}_m\Big)^{-1}\boldsymbol{X}_{m+1,1}\\
&= -\boldsymbol{X}_{1:m,1}^T\boldsymbol{Q}_{m1}^{-1}\boldsymbol{X}_{1:m,1}\boldsymbol{M}'\boldsymbol{X}_{m+1,1}^T\{\boldsymbol{X}_{m+1,1}(\boldsymbol{M}+\boldsymbol{M}')\boldsymbol{X}_{m+1,1}^T\\
&\qquad -\boldsymbol{X}_{m+1,1}\boldsymbol{M}'\boldsymbol{X}_{1:m,1}^T\boldsymbol{Q}_{m1}^{-1}\boldsymbol{X}_{1:m,1}\boldsymbol{M}'\boldsymbol{X}_{m+1,1}^T+\lambda\boldsymbol{I}\}^{-1}\boldsymbol{X}_{m+1,1}\\
&= -\boldsymbol{X}_{1:m,1}^T\boldsymbol{Q}_{m1}^{-1}\boldsymbol{X}_{1:m,1}\boldsymbol{M}'\{\boldsymbol{M}+\boldsymbol{M}'-\boldsymbol{M}'\boldsymbol{X}_{1:m,1}^T\boldsymbol{Q}_{m1}^{-1}\boldsymbol{X}_{1:m,1}\boldsymbol{M}'\\
&\qquad\qquad +\lambda(\boldsymbol{X}_{m+1,1}^T\boldsymbol{X}_{m+1,1})^{-1}\}^{-1}\\
&\xrightarrow{P} -\Big(\frac{1}{m}\boldsymbol{M}+\boldsymbol{M}'\Big)^{-1}\boldsymbol{M}'\Big\{\boldsymbol{M}+\boldsymbol{M}'-\boldsymbol{M}'\Big(\frac{1}{m}\boldsymbol{M}+\boldsymbol{M}'\Big)^{-1}\boldsymbol{M}'\Big\}^{-1}\\
&= -\Big(\frac{1}{m}\boldsymbol{M}+\boldsymbol{M}'\Big)^{-1}\boldsymbol{M}'\Big\{\boldsymbol{M}^{-1}-\frac{1}{m+1}\boldsymbol{M}^{-1}\boldsymbol{M}'\Big(\frac{1}{m+1}\boldsymbol{M}+\boldsymbol{M}'\Big)^{-1}\Big\}\\
&= -\frac{m}{m+1}\boldsymbol{M}^{-1}\boldsymbol{M}'\Big(\frac{1}{m+1}\boldsymbol{M}+\boldsymbol{M}'\Big)^{-1}.
\end{aligned}$$

However,

$$\boldsymbol{X}_{1:m,1}^T\boldsymbol{Q}_{m+1,1}^{1:m,m+1}\boldsymbol{X}_{m+1,1} = \sum_{i=1}^m \boldsymbol{X}_{i1}^T\boldsymbol{Q}_{m+1,1}^{i,m+1}\boldsymbol{X}_{m+1,1} \tag{S.44}$$

and each term in the summation on the right-hand side of (S.44) has the same distribution and, therefore, the same convergence in probability limit. Hence,

$$\boldsymbol{X}_{i1}^T\boldsymbol{Q}_{m+1,1}^{i,m+1}\boldsymbol{X}_{m+1,1} \xrightarrow{P} -\frac{1}{m+1}\boldsymbol{M}^{-1}\boldsymbol{M}'\Big(\frac{1}{m+1}\boldsymbol{M}+\boldsymbol{M}'\Big)^{-1},\quad 1\le i\le m.$$

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

$$\boldsymbol{X}_{i1}^T\boldsymbol{Q}_{m+1,1}^{i,\underset{\sim}{i}}\boldsymbol{X}_{\underset{\sim}{i}1} \xrightarrow{P} -\frac{1}{m+1}\boldsymbol{M}^{-1}\boldsymbol{M}'\Big(\frac{1}{m+1}\boldsymbol{M}+\boldsymbol{M}'\Big)^{-1},\quad 1\le i\ne \underset{\sim}{i}\le m+1,$$

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):

$$\begin{aligned}
\Big\{\underset{1\le i\le m+1}{\text{stack}}(\boldsymbol{X}_{i1})\Big\}^T\boldsymbol{Q}_{m+1,1}^{-1}\underset{1\le i\le m+1}{\text{stack}}(\boldsymbol{X}_{i1}) &= \sum_{i=1}^{m+1}\sum_{\underset{\sim}{i}=1}^{m+1}\boldsymbol{X}_{i,1}^T\boldsymbol{Q}_{m+1,1}^{i,\underset{\sim}{i}}\boldsymbol{X}_{\underset{\sim}{i},1}\\
&\xrightarrow{P}(m+1)\left\{\boldsymbol{M}^{-1}-\frac{1}{m+1}\boldsymbol{M}^{-1}\boldsymbol{M}'\Big(\frac{1}{m+1}\boldsymbol{M}+\boldsymbol{M}'\Big)^{-1}\right\}\\
&\qquad +m(m+1)\left\{-\frac{1}{m+1}\boldsymbol{M}^{-1}\boldsymbol{M}'\Big(\frac{1}{m+1}\boldsymbol{M}+\boldsymbol{M}'\Big)^{-1}\right\}\\
&= \left(\frac{\boldsymbol{M}}{m+1}+\boldsymbol{M}'\right)^{-1}.
\end{aligned}$$

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

$$\boldsymbol{Q}_{12} = \begin{bmatrix} \boldsymbol{X}_{11}(\boldsymbol{M}+\boldsymbol{M}')\boldsymbol{X}_{11}^T & \boldsymbol{X}_{11}\boldsymbol{M}\boldsymbol{X}_{12}^T \\ \boldsymbol{X}_{12}\boldsymbol{M}\boldsymbol{X}_{11}^T & \boldsymbol{X}_{12}(\boldsymbol{M}+\boldsymbol{M}')\boldsymbol{X}_{12}^T \end{bmatrix} + \lambda \boldsymbol{I}$$

$$= \begin{bmatrix} \boldsymbol{X}_{11} & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{X}_{12} \end{bmatrix} \begin{bmatrix} \boldsymbol{M}+\boldsymbol{M}' & \boldsymbol{M} \\ \boldsymbol{M} & \boldsymbol{M}+\boldsymbol{M}' \end{bmatrix} \begin{bmatrix} \boldsymbol{X}_{11} & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{X}_{12} \end{bmatrix}^T + \lambda \boldsymbol{I}.$$

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

$$\begin{aligned} \overset{\blacktriangle}{\boldsymbol{X}}{}_1^T \boldsymbol{Q}_{12}^{11} \overset{\blacktriangle}{\boldsymbol{X}}_1 &= \begin{bmatrix} \boldsymbol{I}_d \\ \boldsymbol{I}_d \end{bmatrix}^T \begin{bmatrix} \boldsymbol{X}_{11} & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{X}_{12} \end{bmatrix}^T \left\{ \begin{bmatrix} \boldsymbol{X}_{11} & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{X}_{12} \end{bmatrix} \begin{bmatrix} \boldsymbol{M}+\boldsymbol{M}' & \boldsymbol{M} \\ \boldsymbol{M} & \boldsymbol{M}+\boldsymbol{M}' \end{bmatrix} \begin{bmatrix} \boldsymbol{X}_{11} & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{X}_{12} \end{bmatrix}^T + \lambda \boldsymbol{I} \right\}^{-1} \\ &\quad \times \begin{bmatrix} \boldsymbol{X}_{11} & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{X}_{12} \end{bmatrix} \begin{bmatrix} \boldsymbol{I}_d \\ \boldsymbol{I}_d \end{bmatrix} \\ &= \begin{bmatrix} \boldsymbol{I}_d \\ \boldsymbol{I}_d \end{bmatrix}^T \left\{ \begin{bmatrix} \boldsymbol{M}+\boldsymbol{M}' & \boldsymbol{M} \\ \boldsymbol{M} & \boldsymbol{M}+\boldsymbol{M}' \end{bmatrix} + \lambda \begin{bmatrix} (\boldsymbol{X}_{11}^T\boldsymbol{X}_{11})^{-1} & \boldsymbol{O} \\ \boldsymbol{O} & (\boldsymbol{X}_{12}^T\boldsymbol{X}_{12})^{-1} \end{bmatrix} \right\}^{-1} \begin{bmatrix} \boldsymbol{I}_d \\ \boldsymbol{I}_d \end{bmatrix} \\ &\xrightarrow{P} \begin{bmatrix} \boldsymbol{I}_d \\ \boldsymbol{I}_d \end{bmatrix}^T \begin{bmatrix} \boldsymbol{M}+\boldsymbol{M}' & \boldsymbol{M} \\ \boldsymbol{M} & \boldsymbol{M}+\boldsymbol{M}' \end{bmatrix}^{-1} \begin{bmatrix} \boldsymbol{I}_d \\ \boldsymbol{I}_d \end{bmatrix} = (\boldsymbol{M} + \tfrac{1}{2}\boldsymbol{M}')^{-1} \end{aligned}$$

where the last equality is due to Lemma 4.

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

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

where

$$\boldsymbol{R}_{12} = \begin{bmatrix} \boldsymbol{X}_{11}\boldsymbol{M}'\boldsymbol{X}_{12}^T & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{X}_{21}\boldsymbol{M}'\boldsymbol{X}_{22}^T \end{bmatrix} = \begin{bmatrix} \boldsymbol{X}_{11} & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{X}_{12} \end{bmatrix} \begin{bmatrix} \boldsymbol{M}' & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{M}' \end{bmatrix} \begin{bmatrix} \boldsymbol{X}_{21} & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{X}_{22} \end{bmatrix}^T$$

and

$$\widetilde{\boldsymbol{Q}}_{12} = \begin{bmatrix} \boldsymbol{X}_{21}(\boldsymbol{M}+\boldsymbol{M}')\boldsymbol{X}_{21}^T + \lambda\boldsymbol{I} & \boldsymbol{X}_{21}\boldsymbol{M}\boldsymbol{X}_{22}^T \\ \boldsymbol{X}_{22}\boldsymbol{M}\boldsymbol{X}_{21}^T & \boldsymbol{X}_{22}(\boldsymbol{M}+\boldsymbol{M}')\boldsymbol{X}_{22}^T + \lambda\boldsymbol{I} \end{bmatrix}$$

$$= \begin{bmatrix} \boldsymbol{X}_{21} & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{X}_{22} \end{bmatrix} \begin{bmatrix} \boldsymbol{M}+\boldsymbol{M}' & \boldsymbol{M} \\ \boldsymbol{M} & \boldsymbol{M}+\boldsymbol{M}' \end{bmatrix} \begin{bmatrix} \boldsymbol{X}_{21} & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{X}_{22} \end{bmatrix}^T + \lambda \boldsymbol{I}.$$

Therefore,

$$
\begin{aligned}
\overset{\blacktriangle}{\boldsymbol{X}}{}_1^T\boldsymbol{Q}_{22}^{11}\overset{\blacktriangle}{\boldsymbol{X}}{}_1 &= \overset{\blacktriangle}{\boldsymbol{X}}{}_1^T\left(\boldsymbol{Q}_{12}-\boldsymbol{R}_{12}\widetilde{\boldsymbol{Q}}_{12}^{-1}\boldsymbol{R}_{12}^T\right)^{-1}\overset{\blacktriangle}{\boldsymbol{X}}{}_1 \\
&= \begin{bmatrix}\boldsymbol{I}_d\\ \boldsymbol{I}_d\end{bmatrix}^T\begin{bmatrix}\boldsymbol{X}_{11} & \boldsymbol{O}\\ \boldsymbol{O} & \boldsymbol{X}_{12}\end{bmatrix}^T\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}^T\right.\\
&\quad\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}^T\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}^T\right\}^{-1}\\
&\quad\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}^T\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.\\
&\qquad\left.+\lambda\begin{bmatrix}(\boldsymbol{X}_{11}^T\boldsymbol{X}_{11})^{-1} & \boldsymbol{O}\\ \boldsymbol{O} & (\boldsymbol{X}_{12}^T\boldsymbol{X}_{12})^{-1}\end{bmatrix}\right\}^{-1}\begin{bmatrix}\boldsymbol{I}_d\\ \boldsymbol{I}_d\end{bmatrix}
\end{aligned}
$$

where

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

Now note that

$$
\begin{bmatrix}\boldsymbol{M}' & \boldsymbol{O}\\ \boldsymbol{O} & \boldsymbol{M}'\end{bmatrix}\boldsymbol{\Psi}\begin{bmatrix}\boldsymbol{M}' & \boldsymbol{O}\\ \boldsymbol{O} & \boldsymbol{M}'\end{bmatrix}=\begin{bmatrix}\boldsymbol{M}'\boldsymbol{X}_{21}^T\widetilde{\boldsymbol{Q}}_{12}^{[1,1]}\boldsymbol{X}_{21}\boldsymbol{M}' & \boldsymbol{M}'\boldsymbol{X}_{21}^T\widetilde{\boldsymbol{Q}}_{12}^{[1,2]}\boldsymbol{X}_{22}\boldsymbol{M}'\\ \boldsymbol{M}'\boldsymbol{X}_{22}^T\widetilde{\boldsymbol{Q}}_{12}^{[2,1]}\boldsymbol{X}_{21}\boldsymbol{M}' & \boldsymbol{M}'\boldsymbol{X}_{22}^T\widetilde{\boldsymbol{Q}}_{22}^{[2,2]}\boldsymbol{X}_{22}\boldsymbol{M}'\end{bmatrix}
$$

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:

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

We then have

$$
\begin{aligned}
\overset{\blacktriangle}{\boldsymbol{X}}{}_1^T\boldsymbol{Q}_{22}^{11}\overset{\blacktriangle}{\boldsymbol{X}}{}_1 &= \begin{bmatrix}\boldsymbol{I}_d\\ \boldsymbol{I}_d\end{bmatrix}^T\begin{bmatrix}\begin{array}{c}\boldsymbol{M}+\boldsymbol{M}'-\boldsymbol{M}'\boldsymbol{X}_{21}^T\widetilde{\boldsymbol{Q}}_{12}^{[1,1]}\boldsymbol{X}_{21}\boldsymbol{M}'\\ +\lambda(\boldsymbol{X}_{11}^T\boldsymbol{X}_{11})^{-1}\end{array} & \boldsymbol{M}-\boldsymbol{M}'\boldsymbol{X}_{21}^T\widetilde{\boldsymbol{Q}}_{12}^{[1,2]}\boldsymbol{X}_{22}\boldsymbol{M}'\\ \boldsymbol{M}-\boldsymbol{M}'\boldsymbol{X}_{22}^T\widetilde{\boldsymbol{Q}}_{12}^{[2,1]}\boldsymbol{X}_{21}\boldsymbol{M}' & \begin{array}{c}\boldsymbol{M}+\boldsymbol{M}'-\boldsymbol{M}'\boldsymbol{X}_{22}^T\widetilde{\boldsymbol{Q}}_{12}^{[2,2]}\boldsymbol{X}_{22}\boldsymbol{M}'\\ +\lambda(\boldsymbol{X}_{12}^T\boldsymbol{X}_{12})^{-1}\end{array}\end{bmatrix}^{-1}\begin{bmatrix}\boldsymbol{I}_d\\ \boldsymbol{I}_d\end{bmatrix}\\
&= \begin{bmatrix}\boldsymbol{I}_d\\ \boldsymbol{I}_d\end{bmatrix}^T\begin{bmatrix}\widetilde{\boldsymbol{A}} & \widetilde{\boldsymbol{B}}\\ \widetilde{\boldsymbol{B}} & \widetilde{\boldsymbol{A}}\end{bmatrix}^{-1}\begin{bmatrix}\boldsymbol{I}_d\\ \boldsymbol{I}_d\end{bmatrix}
\end{aligned}
$$

where

$$
\widetilde{\boldsymbol{A}}\equiv\boldsymbol{M}+\boldsymbol{M}'-\tfrac{1}{2}\boldsymbol{M}'\left\{\boldsymbol{X}_{21}^T\widetilde{\boldsymbol{Q}}_{12}^{[1,1]}\boldsymbol{X}_{21}+\boldsymbol{X}_{22}^T\widetilde{\boldsymbol{Q}}_{12}^{[2,2]}\boldsymbol{X}_{22}\right\}\boldsymbol{M}'\{1+o_P(1)\}+\lambda(\boldsymbol{X}_{11}^T\boldsymbol{X}_{11})^{-1}
$$

and
$$\widetilde{\boldsymbol{B}} \equiv \boldsymbol{M} - \tfrac{1}{2}\boldsymbol{M}'\left\{\boldsymbol{X}_{21}^T\widetilde{\boldsymbol{Q}}_{12}^{[1,2]}\boldsymbol{X}_{22} + \boldsymbol{X}_{22}^T(\widetilde{\boldsymbol{Q}}_{12}^{[1,2]})^T\boldsymbol{X}_{21}\right\}\boldsymbol{M}'\{1+o_P(1)\}$$
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 $\widetilde{\boldsymbol{X}} \equiv \left[\boldsymbol{X}_{21}^T\ \boldsymbol{X}_{22}^T\right]^T$ 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
$$\begin{aligned}\overset{\blacktriangle}{\boldsymbol{X}}{}_1^T\boldsymbol{Q}_{22}^{11}\overset{\blacktriangle}{\boldsymbol{X}}_1 &= 2\Big[\boldsymbol{M}+\boldsymbol{M}'+\boldsymbol{M}-\tfrac{1}{2}\boldsymbol{M}'\big(\widetilde{\boldsymbol{X}}^T\widetilde{\boldsymbol{Q}}_{12}^{-1}\widetilde{\boldsymbol{X}}\big)\boldsymbol{M}'\{1+o_P(1)\}+\lambda(\boldsymbol{X}_{11}^T\boldsymbol{X}_{11})^{-1}\{1+o_P(1)\}\Big]^{-1}\\ &\xrightarrow{P} 2\{2\boldsymbol{M}+\boldsymbol{M}'-\tfrac{1}{2}\boldsymbol{M}'(\boldsymbol{M}+\tfrac{1}{2}\boldsymbol{M}')^{-1}\boldsymbol{M}'\}^{-1} = \boldsymbol{M}^{-1}-\tfrac{1}{4}\boldsymbol{M}^{-1}\boldsymbol{M}'\big(\tfrac{1}{2}\boldsymbol{M}+\tfrac{1}{2}\boldsymbol{M}'\big)^{-1}.\end{aligned}$$
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.40)–(S.42) 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
$$\begin{aligned}\frac{1}{mm'n}\operatorname{tr}(\boldsymbol{Q}_{mm'}^{-2}) = \frac{1}{\lambda^2} &- \frac{M/(m'n)}{\lambda^2\{M+\lambda/(m'n)\}} - \frac{M'/(mn)}{\lambda^2\{M'+\lambda/(mn)\}}\\ &+\frac{MM'/(mm'n)}{\lambda^2\{M+\lambda/(m'n)\}\{M(m'/m)+M'+\lambda/(mn)\}}\\ &+\frac{MM'/(mm'n)}{\lambda^2\{M'+\lambda/(mn)\}\{M+M'(m/m')+\lambda/(m'n)\}}\\ &-\frac{M'/\{(mn)^2\}}{\lambda\{M'+\lambda/(mn)\}^2}-\frac{M/\{(m'n)^2\}}{\lambda\{M+\lambda/(m'n)\}^2}\\ &+\frac{MM'/\{m'(mn)^2\}}{\lambda\{M+M'(m/m')+\lambda/(m'n)\}\{M'+\lambda/(mn)\}^2}\\ &+\frac{MM'/\{m'(mn)^2\}}{\lambda\{M(m'/m)+M'+\lambda/(mn)\}^2\{M+\lambda/(m'n)\}}\\ &+\frac{MM'/\{m(m'n)^2\}}{\lambda\{M(m'/m)+M'+\lambda/(mn)\}\{M+\lambda/(m'n)\}^2}\\ &+\frac{MM'/\{m(m'n)^2\}}{\lambda\{M+M'(m/m')+\lambda/(m'n)\}^2\{M'+\lambda/(mn)\}}.\end{aligned}$$
Hence, under (A5),
$$\frac{1}{mm'n}\operatorname{tr}(\boldsymbol{Q}_{mm'}^{-2}) = \left(\sum_{i=1}^{m}\sum_{i'=1}^{m'}n_{ii'}\right)^{-1}\operatorname{tr}(\boldsymbol{Q}_{mm'}^{-2}) \to \frac{1}{\lambda^2} \tag{S.45}$$
for all $1\le i\le m$, $1\le i'\le m'$.

Next consider the case of $d_{\text{A}}\in\mathbb{N}$ and $m=m'=1$. Then
$$\boldsymbol{Q}_{11}^2 = \lambda^2\boldsymbol{I}_{n_{11}}+\boldsymbol{X}_{11}\boldsymbol{\Omega}_1\boldsymbol{X}_{11}^T \quad\text{where}\quad \boldsymbol{\Omega}_1\equiv(\boldsymbol{M}+\boldsymbol{M}')\boldsymbol{X}_{11}^T\boldsymbol{X}_{11}(\boldsymbol{M}+\boldsymbol{M}')+2\lambda(\boldsymbol{M}+\boldsymbol{M}').$$
Application of Woodbury's matrix identity (S.33) with
$$\boldsymbol{S}=\lambda^2\boldsymbol{I}_{n_{11}},\quad \boldsymbol{U}\equiv\boldsymbol{X}_{11}\boldsymbol{\Omega}_1,\quad \boldsymbol{T}=\boldsymbol{I}_{d_{\text{A}}}\quad\text{and}\quad \boldsymbol{V}\equiv\boldsymbol{X}_{11}^T$$

then gives

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

and so

$$\begin{aligned}\frac{1}{n_{11}}\text{tr}(\boldsymbol{Q}_{11}^{-2}) &= \frac{1}{\lambda^2} - \frac{1}{n_{11}\lambda^4}\text{tr}\Big((\boldsymbol{I} + \lambda^{-2}\boldsymbol{X}_{11}^T\boldsymbol{X}_{11}\boldsymbol{\Omega}_1)^{-1}\boldsymbol{X}_{11}^T\boldsymbol{X}_{11}\boldsymbol{\Omega}_1\Big) \\ &= \frac{1}{\lambda^2} - \frac{1}{n_{11}\lambda^2}\text{tr}\Big(\big\{\boldsymbol{\Omega}_1 + \lambda^2(\boldsymbol{X}_{11}^T\boldsymbol{X}_{11})^{-1}\big\}^{-1}\boldsymbol{\Omega}_1\Big) \xrightarrow{P} \frac{1}{\lambda^2}.\end{aligned}$$

For the $d_{\text{A}} \in \mathbb{N}$, $m \in \mathbb{N}$ and $m' = 1$ extension we note, as given earlier in (S.43), that

$$\boldsymbol{Q}_{m+1,1} = \begin{bmatrix} \boldsymbol{Q}_{m1} & \boldsymbol{R}_m \\ \boldsymbol{R}_m^T & \boldsymbol{S}_m \end{bmatrix} \text{ where } \boldsymbol{S}_m \equiv \boldsymbol{X}_{m+1,1}(\boldsymbol{M} + \boldsymbol{M}')\boldsymbol{X}_{m+1,1}^T + \lambda\boldsymbol{I},$$

$$\boldsymbol{R}_m \equiv \boldsymbol{X}_{1:m,1}\boldsymbol{M}'\boldsymbol{X}_{m+1,1}^T \quad \text{and} \quad \boldsymbol{X}_{1:m,1} \equiv \underset{1\le i\le m}{\text{stack}}(\boldsymbol{X}_{i1})$$

which gives

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

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

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

where

$$\begin{aligned}\boldsymbol{\Omega}_2 \equiv\, &2\lambda(\boldsymbol{M} + \boldsymbol{M}') + (\boldsymbol{M} + \boldsymbol{M}')\boldsymbol{X}_{m+1,1}^T\boldsymbol{X}_{m+1,1}(\boldsymbol{M} + \boldsymbol{M}') + \boldsymbol{M}'\boldsymbol{X}_{1:m,1}^T\boldsymbol{X}_{1:m,1}\boldsymbol{M}' \\ &-\boldsymbol{\Omega}_3^T(\boldsymbol{Q}_{m1}^2 + \boldsymbol{R}_m\boldsymbol{R}_m^T)^{-1}\boldsymbol{\Omega}_3\end{aligned}$$

with

$$\boldsymbol{\Omega}_3 \equiv (\boldsymbol{Q}_{m1} + \lambda\boldsymbol{I})\boldsymbol{X}_{1:m,1}\boldsymbol{M}' + \boldsymbol{X}_{1:m,1}\boldsymbol{M}'\boldsymbol{X}_{m+1,1}^T\boldsymbol{X}_{m+1,1}(\boldsymbol{M} + \boldsymbol{M}').$$

Another application of (S.33) with

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

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

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

and so

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

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

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

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

$$\boldsymbol{Q}_{11} = \boldsymbol{X}_{11}(\boldsymbol{M} + \boldsymbol{M}')\boldsymbol{X}_{11}^T + \lambda \boldsymbol{I}.$$

Then, from Lemma 2,

$$\begin{aligned}
\overset{\star}{\boldsymbol{X}}{}_{11}^T \boldsymbol{Q}_{11}^{-1} \overset{\star}{\boldsymbol{X}}_{11} &= \overset{\star}{\boldsymbol{X}}{}_{11}^T \{\boldsymbol{X}_{11}(\boldsymbol{M} + \boldsymbol{M}')\boldsymbol{X}_{11}^T + \lambda \boldsymbol{I}\}^{-1} \overset{\star}{\boldsymbol{X}}_{11} \\
&= (1/\lambda) \overset{\star}{\boldsymbol{X}}{}_{11}^T \{\boldsymbol{I} - \boldsymbol{X}_{11}(\boldsymbol{X}_{11}^T \boldsymbol{X}_{11})^{-1}\boldsymbol{X}_{11}^T\} \overset{\star}{\boldsymbol{X}}_{11} \\
&\quad + \overset{\star}{\boldsymbol{X}}{}_{11}^T \boldsymbol{X}_{11}(\boldsymbol{X}_{11}^T \boldsymbol{X}_{11})^{-1} \{\boldsymbol{M} + \boldsymbol{M}' + \lambda(\boldsymbol{X}_{11}^T \boldsymbol{X}_{11})^{-1}\}^{-1} (\boldsymbol{X}_{11}^T \boldsymbol{X}_{11})^{-1} \boldsymbol{X}_{11}^T \overset{\star}{\boldsymbol{X}}_{11}.
\end{aligned}$$

Hence

$$\begin{aligned}
\tfrac{1}{n_{11}} \overset{\star}{\boldsymbol{X}}{}_{11}^T \boldsymbol{Q}_{11}^{-1} \overset{\star}{\boldsymbol{X}}_{11} &= (1/\lambda) \left\{ \left(\tfrac{1}{n_{11}} \overset{\star}{\boldsymbol{X}}{}_{11}^T \overset{\star}{\boldsymbol{X}}_{11} \right) - \left(\tfrac{1}{n_{11}} \overset{\star}{\boldsymbol{X}}{}_{11}^T \boldsymbol{X}_{11} \right) \left(\tfrac{1}{n_{11}} \boldsymbol{X}_{11}^T \boldsymbol{X}_{11} \right)^{-1} \left(\tfrac{1}{n_{11}} \boldsymbol{X}_{11}^T \overset{\star}{\boldsymbol{X}}_{11} \right) \right\} \\
&\quad + \tfrac{1}{n_{11}} \left(\tfrac{1}{n_{11}} \overset{\star}{\boldsymbol{X}}{}_{11}^T \boldsymbol{X}_{11} \right) \left(\tfrac{1}{n_{11}} \boldsymbol{X}_{11}^T \boldsymbol{X}_{11} \right)^{-1} \{\boldsymbol{M} + \boldsymbol{M}' + \lambda(\boldsymbol{X}_{11}^T \boldsymbol{X}_{11})^{-1}\}^{-1} \\
&\quad \times \left(\tfrac{1}{n_{11}} \boldsymbol{X}_{11}^T \boldsymbol{X}_{11} \right)^{-1} \left(\tfrac{1}{n_{11}} \boldsymbol{X}_{11}^T \overset{\star}{\boldsymbol{X}}_{11} \right) \\
&\xrightarrow{P} (1/\lambda) \left[E\left(\overset{\star}{\boldsymbol{X}}{}_{\circ}^{\otimes 2} \right) - E\left(\overset{\star}{\boldsymbol{X}}_{\circ} \boldsymbol{X}_{\circ}^T \right) \left\{ E\left(\boldsymbol{X}_{\circ}^{\otimes 2} \right) \right\}^{-1} E\left(\overset{\star}{\boldsymbol{X}}_{\circ} \overset{\star}{\boldsymbol{X}}{}_{\circ}^T \right) \right] \\
&= (1/\lambda) \left[\text{lower right } \overset{\star}{d} \times \overset{\star}{d} \text{ block of } \left\{ E\left([\boldsymbol{X}_{\circ}\ \overset{\star}{\boldsymbol{X}}{}_{\circ}^T]^{\otimes 2} \right) \right\}^{-1} \right]^{-1}.
\end{aligned}$$

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

$$\begin{aligned}
\boldsymbol{X}_{11}^T \boldsymbol{Q}_{11}^{-1} \overset{\star}{\boldsymbol{X}}_{11} &= \boldsymbol{X}_{11}^T \{\boldsymbol{X}_{11}(\boldsymbol{M} + \boldsymbol{M}')\boldsymbol{X}_{11}^T + \lambda \boldsymbol{I}\}^{-1} \overset{\star}{\boldsymbol{X}}_{11} \\
&= \{\boldsymbol{M} + \boldsymbol{M}' + \lambda(\boldsymbol{X}_{11}^T \boldsymbol{X}_{11})^{-1}\}^{-1} \left(\tfrac{1}{n_{11}} \boldsymbol{X}_{11}^T \boldsymbol{X}_{11} \right)^{-1} \tfrac{1}{n_{11}} \boldsymbol{X}_{11}^T \overset{\star}{\boldsymbol{X}}_{11} \\
&\xrightarrow{P} (\boldsymbol{M} + \boldsymbol{M}')^{-1} \{E(\boldsymbol{X}_{\circ}^{\otimes 2})\}^{-1} E(\boldsymbol{X}_{\circ} \overset{\star}{\boldsymbol{X}}{}_{\circ}^T).
\end{aligned}$$

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.43). 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

$$\overset{\star}{\boldsymbol{X}}{}^T_{1:m+1,1}\boldsymbol{Q}^{-1}_{m+1,1}\overset{\star}{\boldsymbol{X}}_{1:m+1,1} = \overset{\star}{\boldsymbol{X}}{}^T_{1:m,1}(\boldsymbol{Q}_{m1} - \boldsymbol{R}_m\boldsymbol{S}_m^{-1}\boldsymbol{R}_m)^{-1}\overset{\star}{\boldsymbol{X}}_{1:m,1}$$

$$-\overset{\star}{\boldsymbol{X}}{}^T_{1:m,1}(\boldsymbol{Q}_{m1} - \boldsymbol{R}_m\boldsymbol{S}_m^{-1}\boldsymbol{R}_m)^{-1}\boldsymbol{R}_m\boldsymbol{S}_m^{-1}\overset{\star}{\boldsymbol{X}}_{m+1,1}$$

$$-\overset{\star}{\boldsymbol{X}}{}^T_{m+1,1}\boldsymbol{S}_m^{-1}\boldsymbol{R}_m^T(\boldsymbol{Q}_{m1} - \boldsymbol{R}_m\boldsymbol{S}_m^{-1}\boldsymbol{R}_m)^{-1}\overset{\star}{\boldsymbol{X}}_{1:m,1}$$

$$+\overset{\star}{\boldsymbol{X}}{}^T_{m+1,1}(\boldsymbol{S}_m - \boldsymbol{R}_m^T\boldsymbol{Q}_{m1}^{-1}\boldsymbol{R}_m)^{-1}\overset{\star}{\boldsymbol{X}}_{m+1,1}$$

$$= \overset{\star}{\boldsymbol{X}}{}^T_{1:m,1}\Big[\boldsymbol{Q}_{m1} - \boldsymbol{X}_{1:m,1}\boldsymbol{M}'\boldsymbol{X}^T_{m+1,1}\big\{\boldsymbol{X}_{m+1,1}(\boldsymbol{M}+\boldsymbol{M}')\boldsymbol{X}^T_{m+1,1} + \lambda\boldsymbol{I}\big\}^{-1}\boldsymbol{X}_{m+1,1}\boldsymbol{M}'\boldsymbol{X}^T_{1:m,1}\Big]^{-1}\overset{\star}{\boldsymbol{X}}_{1:m,1}$$

$$-\overset{\star}{\boldsymbol{X}}{}^T_{1:m,1}\Big[\boldsymbol{Q}_{m1} - \boldsymbol{X}_{1:m,1}\boldsymbol{M}'\boldsymbol{X}^T_{m+1,1}\big\{\boldsymbol{X}_{m+1,1}(\boldsymbol{M}+\boldsymbol{M}')\boldsymbol{X}^T_{m+1,1} + \lambda\boldsymbol{I}\big\}^{-1}\boldsymbol{X}_{m+1,1}\boldsymbol{M}'\boldsymbol{X}^T_{1:m,1}\Big]^{-1}$$

$$\times\boldsymbol{X}_{1:m,1}\boldsymbol{M}'\boldsymbol{X}^T_{m+1,1}\big\{\boldsymbol{X}_{m+1,1}(\boldsymbol{M}+\boldsymbol{M}')\boldsymbol{X}^T_{m+1,1} + \lambda\boldsymbol{I}\big\}^{-1}\overset{\star}{\boldsymbol{X}}_{m+1,1}$$

$$-\overset{\star}{\boldsymbol{X}}{}^T_{m+1,1}\big\{\boldsymbol{X}_{m+1,1}(\boldsymbol{M}+\boldsymbol{M}')\boldsymbol{X}^T_{m+1,1} + \lambda\boldsymbol{I}\big\}^{-1}\boldsymbol{X}_{m+1,1}\boldsymbol{M}'\boldsymbol{X}^T_{1:m,1}$$

$$\times\Big[\boldsymbol{Q}_{m1} - \boldsymbol{X}_{1:m,1}\boldsymbol{M}'\boldsymbol{X}^T_{m+1,1}\big\{\boldsymbol{X}_{m+1,1}(\boldsymbol{M}+\boldsymbol{M}')\boldsymbol{X}^T_{m+1,1} + \lambda\boldsymbol{I}\big\}^{-1}\boldsymbol{X}_{m+1,1}\boldsymbol{M}'\boldsymbol{X}^T_{1:m,1}\Big]^{-1}\overset{\star}{\boldsymbol{X}}_{1:m,1}$$

$$+\overset{\star}{\boldsymbol{X}}{}^T_{m+1,1}\big\{\boldsymbol{X}_{m+1,1}(\boldsymbol{M}+\boldsymbol{M}')\boldsymbol{X}^T_{m+1,1} - \boldsymbol{X}_{m+1,1}\boldsymbol{M}'\boldsymbol{X}^T_{1:m,1}\boldsymbol{Q}^{-1}_{m1}\boldsymbol{X}_{1:m,1}\boldsymbol{M}'\boldsymbol{X}^T_{m+1,1} + \lambda\boldsymbol{I}\big\}^{-1}\overset{\star}{\boldsymbol{X}}_{m+1,1}$$

$$= \mathfrak{T}_1 - \mathfrak{T}_2 - \mathfrak{T}_2^T + \mathfrak{T}_3 + \mathfrak{T}_4$$

where

$$\mathfrak{T}_1 \equiv \overset{\star}{\boldsymbol{X}}{}^T_{1:m,1}\big(\boldsymbol{Q}_{m1} + \boldsymbol{\Gamma}_1\big)^{-1}\overset{\star}{\boldsymbol{X}}_{1:m,1},$$

$$\mathfrak{T}_2 = \overset{\star}{\boldsymbol{X}}{}^T_{1:m,1}\big(\boldsymbol{Q}_{m1} + \boldsymbol{\Gamma}_1\big)^{-1}\boldsymbol{X}_{1:m,1}\boldsymbol{M}'\boldsymbol{\Gamma}_2(\boldsymbol{X}^T_{m+1,1}\boldsymbol{X}_{m+1,1})^{-1}\boldsymbol{X}^T_{m+1,1}\overset{\star}{\boldsymbol{X}}_{m+1,1}$$

$$\mathfrak{T}_3 = (1/\lambda)\overset{\star}{\boldsymbol{X}}{}^T_{m+1,1}\big\{\boldsymbol{I} - \boldsymbol{X}_{m+1,1}(\boldsymbol{X}^T_{m+1,1}\boldsymbol{X}_{m+1,1})^{-1}\boldsymbol{X}^T_{m+1,1}\big\}\overset{\star}{\boldsymbol{X}}_{m+1,1}$$

$$\mathfrak{T}_4 = \overset{\star}{\boldsymbol{X}}{}^T_{m+1,1}\boldsymbol{X}_{m+1,1}(\boldsymbol{X}^T_{m+1,1}\boldsymbol{X}_{m+1,1})^{-1}$$

$$\times\big\{\boldsymbol{M}+\boldsymbol{M}' - \boldsymbol{M}'\boldsymbol{X}^T_{1:m,1}\boldsymbol{Q}^{-1}_{m1}\boldsymbol{X}_{1:m,1}\boldsymbol{M}' + \lambda(\boldsymbol{X}^T_{1:m,1}\boldsymbol{X}_{1:m,1})^{-1}\big\}^{-1}$$

$$\times(\boldsymbol{X}^T_{m+1,1}\boldsymbol{X}_{m+1,1})^{-1}\boldsymbol{X}^T_{m+1,1}\overset{\star}{\boldsymbol{X}}_{m+1,1},$$

$$\boldsymbol{\Gamma}_1 \equiv \boldsymbol{X}_{1:m,1}\boldsymbol{M}'\boldsymbol{\Gamma}_2\boldsymbol{M}'\boldsymbol{X}^T_{1:m,1} \quad \text{and} \quad \boldsymbol{\Gamma}_2 \equiv -\big\{\boldsymbol{M}+\boldsymbol{M}' + \lambda(\boldsymbol{X}^T_{m+1,1}\boldsymbol{X}_{m+1,1})^{-1}\big\}^{-1}.$$

Application of Woodbury's matrix identity (S.33) to $\big(\boldsymbol{Q}_{m1}+\boldsymbol{\Gamma}_1\big)^{-1}$ with $\boldsymbol{S} = \boldsymbol{Q}_{m1}$, $\boldsymbol{U} = \boldsymbol{X}_{1:m,1}\boldsymbol{M}'$, $\boldsymbol{V} = \boldsymbol{M}'\boldsymbol{X}^T_{1:m,1}$ and $\boldsymbol{T} = \boldsymbol{\Gamma}_2$ leads to

$$\mathfrak{T}_1 = \overset{\star}{\boldsymbol{X}}{}^T_{1:m,1}\boldsymbol{Q}^{-1}_{m1}\overset{\star}{\boldsymbol{X}}_{1:m,1}$$

$$-\overset{\star}{\boldsymbol{X}}{}^T_{1:m,1}\boldsymbol{Q}^{-1}_{m1}\boldsymbol{X}_{1:m,1}\boldsymbol{M}'\big\{\boldsymbol{\Gamma}_2 + \boldsymbol{M}'\boldsymbol{X}^T_{1:m,1}\boldsymbol{Q}^{-1}_{m1}\boldsymbol{X}_{1:m,1}\boldsymbol{M}'\big\}^{-1}\boldsymbol{M}'\boldsymbol{X}^T_{1:m,1}\boldsymbol{Q}^{-1}_{m1}\overset{\star}{\boldsymbol{X}}_{1:m,1}$$

$$= \frac{n_{11}+\ldots+n_{m1}}{\lambda}\Big[\text{lower right } \overset{\star}{d}\times\overset{\star}{d} \text{ block of } \big\{E\big([\boldsymbol{X}_\circ\,\overset{\star}{\boldsymbol{X}}_\circ{}^T]^{\otimes 2}\big)\big\}^{-1}\Big]^{-1}\{1+o_P(1)\}$$

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

$$\overset{\star}{\boldsymbol{X}}{}^T_{1:m,1}\big(\boldsymbol{Q}_{m1} + \boldsymbol{\Gamma}_1\big)^{-1}\boldsymbol{X}_{1:m,1} = \overset{\star}{\boldsymbol{X}}{}^T_{1:m,1}\boldsymbol{Q}^{-1}_{m1}\boldsymbol{X}_{1:m,1} - \overset{\star}{\boldsymbol{X}}{}^T_{1:m,1}\boldsymbol{Q}^{-1}_{m1}\boldsymbol{X}_{1:m,1}$$

$$\times\boldsymbol{M}'\big\{\boldsymbol{\Gamma}_2 + \boldsymbol{M}'\boldsymbol{X}^T_{1:m,1}\boldsymbol{Q}^{-1}_{m1}\boldsymbol{X}_{1:m,1}\boldsymbol{M}'\big\}^{-1}\boldsymbol{M}'\boldsymbol{X}^T_{1:m,1}\boldsymbol{Q}^{-1}_{m1}\boldsymbol{X}_{1:m,1}$$

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

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

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

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

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

$$\boldsymbol{Q}_{12} = \begin{bmatrix} \boldsymbol{X}_{11} & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{X}_{12} \end{bmatrix} \begin{bmatrix} \boldsymbol{M}+\boldsymbol{M}' & \boldsymbol{M} \\ \boldsymbol{M} & \boldsymbol{M}+\boldsymbol{M}' \end{bmatrix} \begin{bmatrix} \boldsymbol{X}_{11} & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{X}_{12} \end{bmatrix}^T + \lambda \boldsymbol{I}.$$

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

$$\begin{aligned}
\overset{\star}{\boldsymbol{X}}{}^T\boldsymbol{Q}^{-1}_{12}\overset{\star}{\boldsymbol{X}} = {} & \begin{bmatrix} \boldsymbol{I}_{\overset{\star}{d}} \\ \boldsymbol{I}_{\overset{\star}{d}} \end{bmatrix}^T \begin{bmatrix} \overset{\star}{\boldsymbol{X}}_{11} & \boldsymbol{O} \\ \boldsymbol{O} & \overset{\star}{\boldsymbol{X}}_{12} \end{bmatrix}^T \left\{ \begin{bmatrix} \boldsymbol{X}_{11} & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{X}_{12} \end{bmatrix} \begin{bmatrix} \boldsymbol{M}+\boldsymbol{M}' & \boldsymbol{M} \\ \boldsymbol{M} & \boldsymbol{M}+\boldsymbol{M}' \end{bmatrix} \begin{bmatrix} \boldsymbol{X}_{11} & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{X}_{12} \end{bmatrix}^T + \lambda\boldsymbol{I} \right\}^{-1} \\
& \times \begin{bmatrix} \overset{\star}{\boldsymbol{X}}_{11} & \boldsymbol{O} \\ \boldsymbol{O} & \overset{\star}{\boldsymbol{X}}_{12} \end{bmatrix} \begin{bmatrix} \boldsymbol{I}_{\overset{\star}{d}} \\ \boldsymbol{I}_{\overset{\star}{d}} \end{bmatrix} \\
= {} & \mathfrak{T}_5 + \mathfrak{T}_6
\end{aligned}$$

where

$$\begin{aligned}
\mathfrak{T}_5 = {} & (1/\lambda) \begin{bmatrix} \boldsymbol{I}_{\overset{\star}{d}} \\ \boldsymbol{I}_{\overset{\star}{d}} \end{bmatrix}^T \begin{bmatrix} \overset{\star}{\boldsymbol{X}}_{11} & \boldsymbol{O} \\ \boldsymbol{O} & \overset{\star}{\boldsymbol{X}}_{12} \end{bmatrix}^T \\
& \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}^T \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}^T \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}}{}^T_{11}\overset{\star}{\boldsymbol{X}}_{11} \right) - \left(\frac{1}{n_{11}} \overset{\star}{\boldsymbol{X}}{}^T_{11}\boldsymbol{X}_{11} \right) \left(\frac{1}{n_{11}} \boldsymbol{X}^T_{11}\boldsymbol{X}_{11} \right)^{-1} \left(\frac{1}{n_{11}} \boldsymbol{X}^T_{11}\overset{\star}{\boldsymbol{X}}_{11} \right) \right\} \\
& + \frac{n_{12}}{\lambda} \left\{ \left(\frac{1}{n_{12}} \overset{\star}{\boldsymbol{X}}{}^T_{12}\overset{\star}{\boldsymbol{X}}_{12} \right) - \left(\frac{1}{n_{12}} \overset{\star}{\boldsymbol{X}}{}^T_{11}\boldsymbol{X}_{12} \right) \left(\frac{1}{n_{12}} \boldsymbol{X}^T_{12}\boldsymbol{X}_{12} \right)^{-1} \left(\frac{1}{n_{12}} \boldsymbol{X}^T_{12}\overset{\star}{\boldsymbol{X}}_{12} \right) \right\} \\
= {} & \frac{n_{11}+n_{12}}{\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{}^T]^{\otimes 2}\big)\big\}^{-1}\Big]^{-1} \{1+o_P(1)\}.
\end{aligned}$$

and

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

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

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

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

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}^T \boldsymbol{Q}_{12}^{-1} \overset{\star}{\boldsymbol{X}} &= \begin{bmatrix} \boldsymbol{I}_d \\ \boldsymbol{I}_d \end{bmatrix}^T \begin{bmatrix} \boldsymbol{X}_{11} & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{X}_{12} \end{bmatrix}^T \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}^T + \lambda \boldsymbol{I} \right\}^{-1} \\ &\qquad \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} \\ &= \begin{bmatrix} \boldsymbol{I}_d \\ \boldsymbol{I}_d \end{bmatrix}^T \begin{bmatrix} \boldsymbol{M}+\boldsymbol{M}'+\lambda(\boldsymbol{X}_{11}^T\boldsymbol{X}_{11})^{-1} & \boldsymbol{M} \\ \boldsymbol{M} & \boldsymbol{M}+\boldsymbol{M}'+\lambda(\boldsymbol{X}_{12}^T\boldsymbol{X}_{12})^{-1} \end{bmatrix}^{-1} \\ &\qquad \times \begin{bmatrix} \left(\frac{1}{n_{11}}\boldsymbol{X}_{11}^T\boldsymbol{X}_{11}\right)^{-1} \left(\frac{1}{n_{11}}\boldsymbol{X}_{11}^T\overset{\star}{\boldsymbol{X}}_{11}\right) \\ \left(\frac{1}{n_{12}}\boldsymbol{X}_{11}^T\boldsymbol{X}_{12}\right)^{-1} \left(\frac{1}{n_{12}}\boldsymbol{X}_{12}^T\overset{\star}{\boldsymbol{X}}_{12}\right) \end{bmatrix} \\ &\xrightarrow{P} \begin{bmatrix} \boldsymbol{I}_d \\ \boldsymbol{I}_d \end{bmatrix}^T \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{\star}{\boldsymbol{X}}{}_\circ^T) \\ &= (\boldsymbol{M} + \tfrac{1}{2}\boldsymbol{M}')^{-1} \left\{ E(\boldsymbol{X}_\circ^{\otimes 2}) \right\}^{-1} E(\boldsymbol{X}_\circ \overset{\star}{\boldsymbol{X}}{}_\circ^T) \end{aligned}$$

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

For general m and m', note that the behavior of $\boldsymbol{X}^T \boldsymbol{Q}_{mm'}^{-1} \overset{\star}{\boldsymbol{X}}$ mimics that of the $\boldsymbol{X}^T \boldsymbol{Q}_{mm'}^{-1} \boldsymbol{X}$ special case, with the $\left\{ E(\boldsymbol{X}_\circ^{\otimes 2}) \right\}^{-1} E(\boldsymbol{X}_\circ \overset{\star}{\boldsymbol{X}}{}_\circ^T)$ 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.