Online Supplement for:

Streamlined Variational Inference for Linear Mixed Models with Crossed Random Effects

By Marianne Menictas[1], Gioia Di Credico[2] and Matt P. Wand[3]

Harvard University[1], University of Trieste[2] and University of Technology Sydney[3]

S.1 The Inverse G-Wishart and Inverse χ^2 Distributions

The Inverse G-Wishart corresponds to the matrix inverses of random matrices that have a *G-Wishart* distribution (e.g. Atay-Kayis & Massam, 2005). For any positive integer d, let G be an undirected graph with d nodes labeled $1,\ldots,d$ and set E consisting of sets of pairs of nodes that are connected by an edge. We say that the symmetric $d\times d$ matrix $\boldsymbol{M}$ *respects* G if

$$\boldsymbol{M}_{ij}=0 \quad \text{for all} \quad \{i,j\}\notin E.$$

A $d\times d$ random matrix $\boldsymbol{X}$ has an Inverse G-Wishart distribution with graph G and parameters $\xi>0$ and symmetric $d\times d$ matrix $\boldsymbol{\Lambda}$, written

$$\boldsymbol{X}\sim \text{Inverse-G-Wishart}(G,\xi,\boldsymbol{\Lambda})$$

if and only if the density function of $\boldsymbol{X}$ satisfies

$$\mathfrak{p}(\boldsymbol{X})\propto |\boldsymbol{X}|^{-(\xi+2)/2}\exp\{-\tfrac{1}{2}\text{tr}(\boldsymbol{\Lambda}\,\boldsymbol{X}^{-1})\}$$

over arguments $\boldsymbol{X}$ such that $\boldsymbol{X}$ is symmetric and positive definite and $\boldsymbol{X}^{-1}$ respects G. Two important special cases are

$$G=G_{\text{full}}\equiv \text{totally connected } d\text{-node graph},$$

for which the Inverse G-Wishart distribution coincides with the ordinary Inverse Wishart distribution, and

$$G=G_{\text{diag}}\equiv \text{totally disconnected } d\text{-node graph},$$

for which the Inverse G-Wishart distribution coincides with a product of independent Inverse Chi-Squared random variables. The subscripts of G_{full} and G_{diag} reflect the fact that $\boldsymbol{X}^{-1}$ is a full matrix and $\boldsymbol{X}^{-1}$ is a diagonal matrix in each special case.

The $G=G_{\text{full}}$ case corresponds to the ordinary Inverse Wishart distribution. However, with modularity in mind, we will work with the more general Inverse G-Wishart family throughout this article.

In the $d=1$ special case the graph $G=G_{\text{full}}=G_{\text{diag}}$ and the Inverse G-Wishart distribution reduces to the Inverse Chi-Squared distributions. We write

$$x\sim \text{Inverse-}\chi^2(\xi,\lambda)$$

for this Inverse-G-Wishart$(G_{\text{diag}},\xi,\lambda)$ special case with $d=1$ and $\lambda>0$ scalar.

S.2 The Generalized Blockdiag Operator

If $\boldsymbol{M}_1$, $\boldsymbol{M}_2$ and $\boldsymbol{M}_3$ are each 2×2 matrices then a well-established notation is

$$\underset{1\le i\le 3}{\text{blockdiag}}(\boldsymbol{M}_i)=\begin{bmatrix}\boldsymbol{M}_1 & \boldsymbol{O}_{2\times 2} & \boldsymbol{O}_{2\times 2}\\ \boldsymbol{O}_{2\times 2} & \boldsymbol{M}_2 & \boldsymbol{O}_{2\times 2}\\ \boldsymbol{O}_{2\times 2} & \boldsymbol{O}_{2\times 2} & \boldsymbol{M}_3\end{bmatrix}$$

where $\boldsymbol{O}_{2\times 2}$ denotes the 2×2 matrix of zeroes.

Suppose instead that $\boldsymbol{M}_2$ is $n \times 2$ where $n = 0$. Then $\boldsymbol{M}_2$ is null but here we allow its column dimension to be a positive integer. The *generalized* blockdiag operator is such that

$$\underset{1\le i\le 3}{\text{blockdiag}}(\boldsymbol{M}_i) = \begin{bmatrix} \boldsymbol{M}_1 & \boldsymbol{O}_{2\times 2} & \boldsymbol{O}_{2\times 2} \\ \boldsymbol{O}_{2\times 2} & \boldsymbol{O}_{2\times 2} & \boldsymbol{M}_3 \end{bmatrix}.$$

The key aspect is that, after positioning $\boldsymbol{M}_1$, the column index is incremented by 2 before positioning $\boldsymbol{M}_3$. This is due to $\boldsymbol{M}_2$ being a "matrix" having *generalized dimension* 0×2.

The generalized blockdiag operator is useful for describing the design matrices that arise from model (1). Suppose that $m = 5$, $m' = 3$, $q' = 2$ and that the $n_{ii'}$ values are as given by Table S.1.

	$i' = 1$	$i' = 2$	$i' = 3$
$i = 1$	2	4	0
$i = 2$	0	3	1
$i = 3$	0	0	6
$i = 4$	7	2	9
$i = 5$	5	0	8

Table S.1: *The $n_{ii'}$ values for an illustrative example concerning the use of the generalized blockdiag operator to describe cross random effects design matrices.*

Then, according to the definitions given in Section 2.1 and the rules of the generalized blockdiag operator:

$$\overset{\blacksquare}{\boldsymbol{Z}}{}'_1 = \underset{1\le i'\le 3}{\text{blockdiag}}(\boldsymbol{Z}'_{1i'}) = \begin{bmatrix} \boldsymbol{Z}'_{11} & \boldsymbol{O}_{2\times 2} & \boldsymbol{O}_{2\times 2} \\ \boldsymbol{O}_{4\times 2} & \boldsymbol{Z}'_{12} & \boldsymbol{O}_{4\times 2} \end{bmatrix},$$

$$\overset{\blacksquare}{\boldsymbol{Z}}{}'_2 = \underset{1\le i'\le 3}{\text{blockdiag}}(\boldsymbol{Z}'_{2i'}) = \begin{bmatrix} \boldsymbol{O}_{3\times 2} & \boldsymbol{Z}'_{22} & \boldsymbol{O}_{3\times 2} \\ \boldsymbol{O}_{1\times 2} & \boldsymbol{O}_{1\times 2} & \boldsymbol{Z}'_{23} \end{bmatrix},$$

$$\overset{\blacksquare}{\boldsymbol{Z}}{}'_3 = \underset{1\le i'\le 3}{\text{blockdiag}}(\boldsymbol{Z}'_{3i'}) = \begin{bmatrix} \boldsymbol{O}_{6\times 2} & \boldsymbol{O}_{6\times 2} & \boldsymbol{Z}'_{33} \end{bmatrix},$$

$$\overset{\blacksquare}{\boldsymbol{Z}}{}'_4 = \underset{1\le i'\le 3}{\text{blockdiag}}(\boldsymbol{Z}'_{4i'}) = \begin{bmatrix} \boldsymbol{Z}'_{41} & \boldsymbol{O}_{7\times 2} & \boldsymbol{O}_{7\times 2} \\ \boldsymbol{O}_{2\times 2} & \boldsymbol{Z}'_{42} & \boldsymbol{O}_{2\times 2} \\ \boldsymbol{O}_{9\times 2} & \boldsymbol{O}_{9\times 2} & \boldsymbol{Z}'_{43} \end{bmatrix}$$

$$\text{and}\quad \overset{\blacksquare}{\boldsymbol{Z}}{}'_5 = \underset{1\le i'\le 3}{\text{blockdiag}}(\boldsymbol{Z}'_{5i'}) = \begin{bmatrix} \boldsymbol{Z}'_{51} & \boldsymbol{O}_{5\times 2} & \boldsymbol{O}_{5\times 2} \\ \boldsymbol{O}_{8\times 2} & \boldsymbol{O}_{8\times 2} & \boldsymbol{Z}'_{53} \end{bmatrix}.$$

Algorithm S.1 provides full details of the generalized blockdiag operator for general input matrices, with some possibly having generalized dimension for which 0 is allowed.

S.3 Derivation of the $\mathfrak{q}(\boldsymbol{\beta}, \boldsymbol{u}_{\text{all}})$ Parameters Updates Under Product Restriction I

The full conditional distribution of $\boldsymbol{\beta}$ is

$$\mathfrak{p}(\boldsymbol{\beta}|\text{rest}) \propto \mathfrak{p}(\boldsymbol{y}|\boldsymbol{\beta}, \boldsymbol{u}, \boldsymbol{u}', \sigma^2)\mathfrak{p}(\boldsymbol{\beta}).$$

Algorithm S.1 *The generalized blockdiag algorithm.*

Inputs: $\{\boldsymbol{M}_i : 1 \leq i \leq d\}$ where $\boldsymbol{M}_i$ is an $r_i \times c_i$ generalized dimensioned matrix

for integers $r_i \geq 0$ and $c_i \geq 0$.

$\boldsymbol{A} \longleftarrow$ the $\left(\sum_{i=1}^{d} r_i\right) \times \left(\sum_{i=1}^{d} c_i\right)$ matrix of zeroes

$r_{\text{stt}} \longleftarrow 1 \quad ; \quad c_{\text{stt}} \longleftarrow 1$

for $i = 1, \ldots, m$

$r_{\text{end}} \longleftarrow r_{\text{stt}} + r_i - 1 \quad ; \quad c_{\text{end}} \longleftarrow c_{\text{stt}} + c_i - 1$

if $r_i > 0$ and $c_i > 0$

(rows r_{stt} to r_{end} and columns c_{stt} to c_{end} of $\boldsymbol{A}$) $\longleftarrow \boldsymbol{M}_i$

$r_{\text{stt}} \longleftarrow r_{\text{end}} + 1 \quad ; \quad c_{\text{stt}} \longleftarrow c_{\text{end}} + 1$

Output: $\boldsymbol{A}$

Note that $\mathfrak{p}(\boldsymbol{y}|\boldsymbol{\beta}, \boldsymbol{u}, \boldsymbol{u}', \sigma^2)$ can be expressed as the $N\left(\boldsymbol{X\beta}, \sigma^2 \boldsymbol{I}\right)$ density function in the vector

$$\boldsymbol{y} - \underset{1\leq i\leq m}{\text{stack}}\left\{\underset{1\leq i'\leq m'}{\text{stack}}\left(\boldsymbol{Z}_{ii'}\boldsymbol{u}_i + \boldsymbol{Z}'_{ii'}\boldsymbol{u}'_{i'}\right)\right\}$$

Also, $\mathfrak{p}(\boldsymbol{\beta})$ is the $N\left(\boldsymbol{\mu_\beta}, \boldsymbol{\Sigma_\beta}\right)$ density function in the vector $\boldsymbol{\beta}$. Then, under product restriction I, standard quadratic form manipulations lead to the optimal $\mathfrak{q}$-density function of $\boldsymbol{\beta}$ being that of the $N(\boldsymbol{\mu}_{\mathfrak{q}(\boldsymbol{\beta})}, \boldsymbol{\Sigma}_{\mathfrak{q}(\boldsymbol{\beta})})$ distribution with updates

$$\boldsymbol{\mu}_{\mathfrak{q}(\boldsymbol{\beta})} \longleftarrow (\mu_{\mathfrak{q}(\sigma^2)}\boldsymbol{X}^T\boldsymbol{X}+\boldsymbol{\Sigma}_{\boldsymbol{\beta}}^{-1})^{-1}(\mu_{\mathfrak{q}(\sigma^2)}\boldsymbol{X}^T\boldsymbol{r}+\boldsymbol{\Sigma}_{\boldsymbol{\beta}}^{-1}\boldsymbol{\mu}_{\boldsymbol{\beta}}^{-1}) \text{ and } \boldsymbol{\Sigma}_{\mathfrak{q}(\boldsymbol{\beta})} \longleftarrow (\mu_{\mathfrak{q}(\sigma^2)}\boldsymbol{X}^T\boldsymbol{X}+\boldsymbol{\Sigma}_{\boldsymbol{\beta}}^{-1})^{-1}.$$

where

$$\boldsymbol{r} \equiv \boldsymbol{y} - \underset{1\leq i\leq m}{\text{stack}}\left\{\underset{1\leq i'\leq m'}{\text{stack}}\left(\boldsymbol{Z}_{ii'}\boldsymbol{\mu}_{\mathfrak{q}(\boldsymbol{u}_i)} + \boldsymbol{Z}'_{ii'}\boldsymbol{\mu}_{\mathfrak{q}(\boldsymbol{u}'_{i'})}\right)\right\}.$$

If $\boldsymbol{b}$ and $\boldsymbol{B}$ are defined according to the updates in (9) then simple algebra shows that

$$\boldsymbol{B}^T\boldsymbol{b} = \mu_{\mathfrak{q}(\sigma^2)}\boldsymbol{X}^T\boldsymbol{r} + \boldsymbol{\Sigma}_{\boldsymbol{\beta}}^{-1}\boldsymbol{\mu}_{\boldsymbol{\beta}}^{-1} \quad \text{and} \quad \boldsymbol{B}^T\boldsymbol{B} = \mu_{\mathfrak{q}(\sigma^2)}\boldsymbol{X}^T\boldsymbol{X} + \boldsymbol{\Sigma}_{\boldsymbol{\beta}}^{-1}.$$

Therefore, the $\boldsymbol{\mu}_{\mathfrak{q}(\boldsymbol{\beta})}$ update corresponds to the least squares solution $\boldsymbol{x} = (\boldsymbol{B}^T\boldsymbol{B})^{-1}\boldsymbol{B}^T\boldsymbol{b}$ and the update of $\boldsymbol{\Sigma}_{\mathfrak{q}(\boldsymbol{\beta})}$ corresponds to $(\boldsymbol{B}^T\boldsymbol{B})^{-1}$.

Analogous arguments can be used to justify the updates for the parameters of $\mathfrak{q}(\boldsymbol{u}_i)$, $1 \leq i \leq m$, and $\mathfrak{q}(\boldsymbol{u}'_{i'})$, $1 \leq i' \leq m'$.

S.4 The SolveLeastSquares Algorithm

The SolveLeastSquares is concerned with solving the least squares problem

$$\min_{\boldsymbol{x}} \|\boldsymbol{b} - \boldsymbol{Bx}\|^2.$$

which has solution $\boldsymbol{x} = (\boldsymbol{B}^T\boldsymbol{B})^{-1}\boldsymbol{B}^T\boldsymbol{b}$. The matrix $(\boldsymbol{B}^T\boldsymbol{B})^{-1}$ is also of intrinsic interest. In next subsection a version of this problem is solved for the situation where $\boldsymbol{B}$ has two-level sparse structure. In this subsection there is no sparseness structure imposed on $\boldsymbol{B}$.

S.5 The SolveTwoLevelSparseLeastSquares Algorithm

The SolveTwoLevelSparseLeastSquares algorithm solves a sparse version of the the least squares problem:

$$\min_{\boldsymbol{x}} \|\boldsymbol{b} - \boldsymbol{Bx}\|^2$$

Algorithm S.2 SOLVELEASTSQUARES *for solving the least squares problem: minimise* $\|\boldsymbol{b} - \boldsymbol{B}\,\boldsymbol{x}\|^2$ *in* $\boldsymbol{x}$ *and obtaining* $(\boldsymbol{B}^T\boldsymbol{B})^{-1}$.

Inputs: $\left\{\boldsymbol{b}(\tilde{n}\times 1),\ \boldsymbol{B}(\tilde{n}\times p)\right\}$

Decompose $\boldsymbol{B} = \boldsymbol{Q}\begin{bmatrix}\boldsymbol{R}\\ \boldsymbol{0}\end{bmatrix}$ such that $\boldsymbol{Q}^{-1} = \boldsymbol{Q}^T$ and $\boldsymbol{R}$ is upper-triangular.

$\boldsymbol{c} \longleftarrow \boldsymbol{Q}^T\boldsymbol{b}$; $\boldsymbol{c}_1 \longleftarrow$ first p rows of $\boldsymbol{c}$

$\boldsymbol{x} \longleftarrow \boldsymbol{R}^{-1}\boldsymbol{c}_1$; $(\boldsymbol{B}^T\boldsymbol{B})^{-1} \longleftarrow \boldsymbol{R}^{-1}\boldsymbol{R}^{-T}$

Output: $\left(\boldsymbol{x}, (\boldsymbol{B}^T\boldsymbol{B})^{-1}\right)$

which has solution $\boldsymbol{x} = \boldsymbol{A}^{-1}\boldsymbol{B}^T\boldsymbol{b}$ where $\boldsymbol{A} = \boldsymbol{B}^T\boldsymbol{B}$ where $\boldsymbol{B}$ and $\boldsymbol{b}$ have the following structure:

$$\boldsymbol{B} \equiv \left[\begin{array}{c|c|c|c|c} \boldsymbol{B}_1 & \dot{\boldsymbol{B}}_1 & \boldsymbol{O} & \cdots & \boldsymbol{O} \\ \hline \boldsymbol{B}_2 & \boldsymbol{O} & \dot{\boldsymbol{B}}_2 & \cdots & \boldsymbol{O} \\ \hline \vdots & \vdots & \vdots & \ddots & \vdots \\ \hline \boldsymbol{B}_m & \boldsymbol{O} & \boldsymbol{O} & \cdots & \dot{\boldsymbol{B}}_m \end{array}\right] \quad \text{and} \quad \boldsymbol{b} = \left[\begin{array}{c} \boldsymbol{b}_1 \\ \hline \boldsymbol{b}_2 \\ \hline \vdots \\ \hline \boldsymbol{b}_m \end{array}\right]. \tag{S.1}$$

The sub-vectors of $\boldsymbol{x}$ and the sub-matrices of $\boldsymbol{A}$ corresponding to its non-zero blocks of are labeled as follows:

$$\boldsymbol{x} = \left[\begin{array}{c} \boldsymbol{x}_1 \\ \hline \boldsymbol{x}_{2,1} \\ \hline \boldsymbol{x}_{2,2} \\ \hline \vdots \\ \hline \boldsymbol{x}_{2,m} \end{array}\right] \quad \text{and} \quad \boldsymbol{A}^{-1} = \left[\begin{array}{c|c|c|c|c} \boldsymbol{A}^{11} & \boldsymbol{A}^{12,1} & \boldsymbol{A}^{12,2} & \cdots & \boldsymbol{A}^{12,m} \\ \hline \boldsymbol{A}^{12,1\,T} & \boldsymbol{A}^{22,1} & \times & \cdots & \times \\ \hline \boldsymbol{A}^{12,2\,T} & \times & \boldsymbol{A}^{22,2} & \cdots & \times \\ \hline \vdots & \vdots & \vdots & \ddots & \vdots \\ \hline \boldsymbol{A}^{12,m\,T} & \times & \times & \cdots & \boldsymbol{A}^{22,m} \end{array}\right] \tag{S.2}$$

with $\times$ denoting sub-blocks that are not of interest. The SOLVETWOLEVELSPARSELEASTSQUARES algorithm is given in Algorithm S.3.

S.6 Derivation of Result 1

The full conditional density function of $(\boldsymbol{\beta}, \boldsymbol{u})$ satisfies

$$\mathfrak{p}(\boldsymbol{\beta}, \boldsymbol{u}|\text{rest}) \propto \mathfrak{p}(\boldsymbol{y}|\boldsymbol{\beta}, \boldsymbol{u}, \boldsymbol{u}', \sigma^2)\mathfrak{p}(\boldsymbol{\beta}, \boldsymbol{u}|\boldsymbol{\Sigma}).$$

Note that $\mathfrak{p}(\boldsymbol{y}|\boldsymbol{\beta}, \boldsymbol{u}, \boldsymbol{u}', \sigma^2)$ can be expressed as the

$$N\left(\overset{\blacktriangle}{\boldsymbol{C}}\begin{bmatrix}\boldsymbol{\beta}\\ \boldsymbol{u}\end{bmatrix}, \sigma^2\boldsymbol{I}\right)$$

density function in the vector $\overset{\blacktriangle}{\boldsymbol{r}}'$, where

$$\overset{\blacktriangle}{\boldsymbol{C}} \equiv \left[\underset{1\le i\le m}{\text{stack}}(\overset{\blacktriangle}{\boldsymbol{X}}_i) \quad \underset{1\le i\le m}{\text{blockdiag}}(\overset{\blacktriangle}{\boldsymbol{Z}}_i)\right] \quad \text{and} \quad \overset{\blacktriangle}{\boldsymbol{r}}' \equiv \underset{1\le i\le m}{\text{stack}}\left\{\overset{\blacktriangle}{\boldsymbol{y}}_i - \underset{1\le i'\le m'}{\text{stack}}(\boldsymbol{Z}'_{ii'}\boldsymbol{u}'_{i'})\right\}.$$

Algorithm S.3 SolveTwoLevelSparseLeastSquares *for solving the two-level sparse matrix least squares problem: minimise* $\|\boldsymbol{b}-\boldsymbol{B}\,\boldsymbol{x}\|^2$ *in* $\boldsymbol{x}$ *and sub-blocks of* $\boldsymbol{A}^{-1}$ *corresponding to the non-zero sub-blocks of* $\boldsymbol{A}=\boldsymbol{B}^T\boldsymbol{B}$*. The sub-block notation is given by (S.1) and (S.2).*

Inputs: $\big\{\big(\boldsymbol{b}_i(\tilde{n}_i\times 1),\ \boldsymbol{B}_i(\tilde{n}_i\times p),\ \overset{\bullet}{\boldsymbol{B}}_i(\tilde{n}_i\times q)\big):\ 1\le i\le m\big\}$

$\boldsymbol{\omega}_3 \longleftarrow$ NULL ; $\boldsymbol{\Omega}_4 \longleftarrow$ NULL

For $i=1,\ldots,m$:

Decompose $\overset{\bullet}{\boldsymbol{B}}_i=\boldsymbol{Q}_i\begin{bmatrix}\boldsymbol{R}_i\\ \boldsymbol{0}\end{bmatrix}$ such that $\boldsymbol{Q}_i^{-1}=\boldsymbol{Q}_i^T$ and $\boldsymbol{R}_i$ is upper-triangular.

$\boldsymbol{c}_{0i}\longleftarrow \boldsymbol{Q}_i^T\boldsymbol{b}_i$; $\boldsymbol{C}_{0i}\longleftarrow \boldsymbol{Q}_i^T\boldsymbol{B}_i$

$\boldsymbol{c}_{1i}\longleftarrow$ first q rows of $\boldsymbol{c}_{0i}$; $\boldsymbol{c}_{2i}\longleftarrow$ remaining rows of $\boldsymbol{c}_{0i}$; $\boldsymbol{\omega}_3\longleftarrow\begin{bmatrix}\boldsymbol{\omega}_3\\ \boldsymbol{c}_{2i}\end{bmatrix}$

$\boldsymbol{C}_{1i}\longleftarrow$ first q rows of $\boldsymbol{C}_{0i}$; $\boldsymbol{C}_{2i}\longleftarrow$ remaining rows of $\boldsymbol{C}_{0i}$; $\boldsymbol{\Omega}_4\longleftarrow\begin{bmatrix}\boldsymbol{\Omega}_4\\ \boldsymbol{C}_{2i}\end{bmatrix}$

Decompose $\boldsymbol{\Omega}_4=\boldsymbol{Q}\begin{bmatrix}\boldsymbol{R}\\ \boldsymbol{0}\end{bmatrix}$ such that $\boldsymbol{Q}^{-1}=\boldsymbol{Q}^T$ and $\boldsymbol{R}$ is upper-triangular.

$\boldsymbol{c}\longleftarrow$ first p rows of $\boldsymbol{Q}^T\boldsymbol{\omega}_3$; $\boldsymbol{x}_1\longleftarrow\boldsymbol{R}^{-1}\boldsymbol{c}$; $\boldsymbol{A}^{11}\longleftarrow\boldsymbol{R}^{-1}\boldsymbol{R}^{-T}$

For $i=1,\ldots,m$:

$\boldsymbol{x}_{2,i}\longleftarrow\boldsymbol{R}_i^{-1}(\boldsymbol{c}_{1i}-\boldsymbol{C}_{1i}\boldsymbol{x}_1)$; $\boldsymbol{A}^{12,i}\longleftarrow -\boldsymbol{A}^{11}(\boldsymbol{R}_i^{-1}\boldsymbol{C}_{1i})^T$

$\boldsymbol{A}^{22,i}\longleftarrow\boldsymbol{R}_i^{-1}(\boldsymbol{R}_i^{-T}-\boldsymbol{C}_{1i}\boldsymbol{A}^{12,i})$

Output: $\Big(\boldsymbol{x}_1,\boldsymbol{A}^{11},\big\{\big(\boldsymbol{x}_{2,i},\boldsymbol{A}^{22,i},\boldsymbol{A}^{12,i}\big):\ 1\le i\le m\big\}\Big)$

Also,

$$\mathfrak{p}(\boldsymbol{\beta},\boldsymbol{u}|\boldsymbol{\Sigma})\quad\text{is the}\quad N\left(\begin{bmatrix}\boldsymbol{\mu}_{\boldsymbol{\beta}}\\ \boldsymbol{0}\end{bmatrix},\begin{bmatrix}\boldsymbol{\Sigma}_{\boldsymbol{\beta}} & \boldsymbol{O}\\ \boldsymbol{O} & \boldsymbol{I}_m\otimes\boldsymbol{\Sigma}\end{bmatrix}\right)$$

density function in the vector $(\boldsymbol{\beta},\boldsymbol{u})$. Then, under product restriction II, standard quadratic form manipulations lead to the optimal $\mathfrak{q}$-density function of $(\boldsymbol{\beta},\boldsymbol{u})$ being that of the $N(\boldsymbol{\mu}_{\mathfrak{q}(\boldsymbol{\beta},\boldsymbol{u})},\boldsymbol{\Sigma}_{\mathfrak{q}(\boldsymbol{\beta},\boldsymbol{u})})$ distribution with updates

$$\boldsymbol{\mu}_{\mathfrak{q}(\boldsymbol{\beta},\boldsymbol{u})}\longleftarrow(\overset{\blacktriangle}{\boldsymbol{C}}{}^T\overset{\blacktriangle}{\boldsymbol{R}}{}^{-1}_{\text{MFVB}}\overset{\blacktriangle}{\boldsymbol{C}}+\overset{\blacktriangle}{\boldsymbol{D}}_{\text{MFVB}})^{-1}(\overset{\blacktriangle}{\boldsymbol{C}}{}^T\overset{\blacktriangle}{\boldsymbol{R}}{}^{-1}_{\text{MFVB}}\overset{\blacktriangle}{\boldsymbol{r}}{}'_{\text{MFVB}}+\overset{\blacktriangle}{\boldsymbol{o}}_{\text{MFVB}})\quad\text{and}\quad\boldsymbol{\Sigma}_{\mathfrak{q}(\boldsymbol{\beta},\boldsymbol{u})}\longleftarrow(\overset{\blacktriangle}{\boldsymbol{C}}{}^T\boldsymbol{R}^{-1}_{\text{MFVB}}\overset{\blacktriangle}{\boldsymbol{C}}+\overset{\blacktriangle}{\boldsymbol{D}}_{\text{MFVB}})^{-1}.$$

Here $\overset{\blacktriangle}{\boldsymbol{R}}_{\text{MFVB}}\equiv\mu^{-1}_{\mathfrak{q}(1/\sigma^2)}\boldsymbol{I}$,

$$\overset{\blacktriangle}{\boldsymbol{D}}_{\text{MFVB}}\equiv\begin{bmatrix}\boldsymbol{\Sigma}_{\boldsymbol{\beta}}^{-1} & \boldsymbol{O}\\ \boldsymbol{O} & \boldsymbol{I}_m\otimes\boldsymbol{M}_{\mathfrak{q}(\boldsymbol{\Sigma}^{-1})}\end{bmatrix},\quad\overset{\blacktriangle}{\boldsymbol{o}}_{\text{MFVB}}\equiv\begin{bmatrix}\boldsymbol{\Sigma}_{\boldsymbol{\beta}}^{-1}\boldsymbol{\mu}_{\boldsymbol{\beta}}\\ \boldsymbol{0}\end{bmatrix}$$

and

$$\overset{\blacktriangle}{\boldsymbol{r}}{}'_{\text{MFVB}}\equiv\underset{1\le i\le m}{\text{stack}}\Big\{\overset{\blacktriangle}{\boldsymbol{y}}_i-\underset{1\le i'\le m'}{\text{stack}}\big(\boldsymbol{Z}'_{ii'}\boldsymbol{\mu}_{\mathfrak{q}(\boldsymbol{u}'_{i'})}\big)\Big\}.$$

If $\boldsymbol{b}$ and $\boldsymbol{B}$ are defined according to (S.1) and the matrices $\boldsymbol{b}_i$, $\boldsymbol{B}_i$ and $\overset{\bullet}{\boldsymbol{B}}_i$ are defined as in Result 1 then

$$\boldsymbol{B}^T\boldsymbol{b}=\overset{\blacktriangle}{\boldsymbol{C}}{}^T\overset{\blacktriangle}{\boldsymbol{R}}{}^{-1}_{\text{MFVB}}\overset{\blacktriangle}{\boldsymbol{r}}{}'_{\text{MFVB}}+\overset{\blacktriangle}{\boldsymbol{o}}_{\text{MFVB}}\quad\text{and}\quad\boldsymbol{B}^T\boldsymbol{B}=\overset{\blacktriangle}{\boldsymbol{C}}{}^T\overset{\blacktriangle}{\boldsymbol{R}}{}^{-1}_{\text{MFVB}}\overset{\blacktriangle}{\boldsymbol{C}}+\overset{\blacktriangle}{\boldsymbol{D}}_{\text{MFVB}}.$$

Therefore, with this assignment of $\boldsymbol{b}_i$, $\boldsymbol{B}_i$ and $\overset{\bullet}{\boldsymbol{B}}_i$, the $\boldsymbol{\mu}_{\mathfrak{q}(\boldsymbol{\beta},\boldsymbol{u}_{\text{all}})}$ update corresponds to the least squares solution $\boldsymbol{x}=(\boldsymbol{B}^T\boldsymbol{B})^{-1}\boldsymbol{B}^T\boldsymbol{b}$ and the updates of the sub-blocks of $\boldsymbol{\Sigma}_{\mathfrak{q}(\boldsymbol{\beta},\boldsymbol{u}_{\text{all}})}$ listed in the first two rows of Table 1 correspond to the sub-blocks of $(\boldsymbol{B}^T\boldsymbol{B})^{-1}$ in the positions where $\boldsymbol{B}^T\boldsymbol{B}$ has non-zero sub-blocks.

S.7 Derivation of Result 2

Result 2 uses the following re-ordering of the overall design matrix:

$$\widetilde{\boldsymbol{C}} \equiv \left[\boldsymbol{X} \ \underset{1\le i\le m}{\text{stack}}(\overset{\blacksquare}{\boldsymbol{Z}}_i') \ \ \underset{1\le i\le m}{\text{blockdiag}}(\overset{\blacktriangle}{\boldsymbol{Z}}_i) \right].$$

rather than $\boldsymbol{C} \equiv [\boldsymbol{X} \ \boldsymbol{Z}]$ in the generalized ridge regression expressions of Section 3. This re-ordering involves the q-density parameters of $\boldsymbol{u}'$ preceding those of $\boldsymbol{u}$ and is brought about by our $m \ge m'$ convention throughout this article and the requirement that the potentially very large

$$\underset{1\le i\le m}{\text{blockdiag}}(\overset{\blacktriangle}{\boldsymbol{Z}}_i)$$

appears on the right for embedding within the two-level sparse least squares infrastructure of Nolan & Wand (2020) and Nolan *et al.* (2019). The re-ordering means that the updates for $\boldsymbol{\mu}_{\mathsf{q}(\boldsymbol{\beta},\boldsymbol{u}',\boldsymbol{u})}$ and $\boldsymbol{\Sigma}_{\mathsf{q}(\boldsymbol{\beta},\boldsymbol{u}',\boldsymbol{u})}$ are

$$\begin{aligned}
\boldsymbol{\mu}_{\mathsf{q}(\boldsymbol{\beta},\boldsymbol{u}',\boldsymbol{u})} &\longleftarrow (\widetilde{\boldsymbol{C}}^T \boldsymbol{R}_{\text{MFVB}}^{-1}\widetilde{\boldsymbol{C}} + \widetilde{\boldsymbol{D}}_{\text{MFVB}})^{-1}(\widetilde{\boldsymbol{C}}^T \boldsymbol{R}_{\text{MFVB}}^{-1}\boldsymbol{y} + \boldsymbol{o}_{\text{MFVB}}) \quad \text{and}\\
\boldsymbol{\Sigma}_{\mathsf{q}(\boldsymbol{\beta},\boldsymbol{u}',\boldsymbol{u})} &\longleftarrow (\widetilde{\boldsymbol{C}}^T \boldsymbol{R}_{\text{MFVB}}^{-1}\widetilde{\boldsymbol{C}} + \widetilde{\boldsymbol{D}}_{\text{MFVB}})^{-1}
\end{aligned}$$

where

$$\widetilde{\boldsymbol{D}}_{\text{MFVB}} \equiv \begin{bmatrix} \boldsymbol{\Sigma}_{\boldsymbol{\beta}}^{-1} & \boldsymbol{O} & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{I}_{m'} \otimes \boldsymbol{M}_{\mathsf{q}((\boldsymbol{\Sigma}')^{-1})} & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{O} & \boldsymbol{I}_m \otimes \boldsymbol{M}_{\mathsf{q}(\boldsymbol{\Sigma}^{-1})} \end{bmatrix}$$

has the $\boldsymbol{M}_{\mathsf{q}((\boldsymbol{\Sigma}')^{-1})}$ matrices appearing before the $\boldsymbol{M}_{\mathsf{q}(\boldsymbol{\Sigma}^{-1})}$ matrices due to the switch in the ordering of the random effects vectors.

If $\boldsymbol{b}$ and $\boldsymbol{B}$ are defined according to (S.1) with the matrices $\boldsymbol{b}_i$, $\boldsymbol{B}_i$ and $\overset{\bullet}{\boldsymbol{B}}_i$ defined as in Result 2 then straightforward matrix algebra can be used to show that

$$\boldsymbol{B}^T\boldsymbol{b} = \widetilde{\boldsymbol{C}}^T \boldsymbol{R}_{\text{MFVB}}^{-1}\boldsymbol{y} + \boldsymbol{o}_{\text{MFVB}} \quad \text{and} \quad \boldsymbol{B}^T\boldsymbol{B} = \widetilde{\boldsymbol{C}}^T \boldsymbol{R}_{\text{MFVB}}^{-1}\widetilde{\boldsymbol{C}} + \widetilde{\boldsymbol{D}}_{\text{MFVB}}.$$

Therefore, with this assignment of $\boldsymbol{b}_i$, $\boldsymbol{B}_i$ and $\overset{\bullet}{\boldsymbol{B}}_i$, the $\boldsymbol{\mu}_{\mathsf{q}(\boldsymbol{\beta},\boldsymbol{u}',\boldsymbol{u})}$ update corresponds to the least squares solution $\boldsymbol{x} = (\boldsymbol{B}^T\boldsymbol{B})^{-1}\boldsymbol{B}^T\boldsymbol{b}$ and the updates of the sub-blocks of $\boldsymbol{\Sigma}_{\mathsf{q}(\boldsymbol{\beta},\boldsymbol{u}',\boldsymbol{u})}$ listed in the first four rows of Table 1 correspond to the sub-blocks of $(\boldsymbol{B}^T\boldsymbol{B})^{-1}$ in the positions where $\boldsymbol{B}^T\boldsymbol{B}$ has non-zero sub-blocks.

S.8 Marginal Log-Likelihood Lower Bound and Derivation

The logarithmic form of the variational lower bound on the marginal log-likelihood, corresponding to model (1) with prior specification (B) and product restriction III is

$$\begin{aligned}
\log \underline{\mathfrak{p}}(\boldsymbol{y};\mathsf{q}) = E_{\mathsf{q}}\big\{&\log \mathfrak{p}\left(\boldsymbol{y},\boldsymbol{\beta},\boldsymbol{u},\boldsymbol{u}',a_{\sigma^2},\boldsymbol{A}_{\boldsymbol{\Sigma}},\boldsymbol{A}_{\boldsymbol{\Sigma}'},\sigma^2,\boldsymbol{\Sigma},\boldsymbol{\Sigma}'\right)\\
&-\log \mathsf{q}^*\left(\boldsymbol{\beta},\boldsymbol{u},\boldsymbol{u}',a_{\sigma^2},\boldsymbol{A}_{\boldsymbol{\Sigma}},\boldsymbol{A}_{\boldsymbol{\Sigma}'},\sigma^2,\boldsymbol{\Sigma},\boldsymbol{\Sigma}'\right)\big\}\\
= E_{\mathsf{q}}&\left\{\mathfrak{p}\left(\boldsymbol{y}|\boldsymbol{\beta},\boldsymbol{u},\boldsymbol{u}',\sigma^2\right)\right\} + E_{\mathsf{q}}\left\{\log \mathfrak{p}\left(\boldsymbol{\beta},\boldsymbol{u},\boldsymbol{u}'|\boldsymbol{\Sigma},\boldsymbol{\Sigma}'\right)\right\}\\
&-E_{\mathsf{q}}\left\{\log \mathsf{q}^*\left(\boldsymbol{\beta},\boldsymbol{u},\boldsymbol{u}'\right)\right\} + E_{\mathsf{q}}\left\{\log \mathfrak{p}\left(\sigma^2|a_{\sigma^2}\right)\right\} - E_{\mathsf{q}}\left\{\log \mathsf{q}^*\left(\sigma^2\right)\right\}\\
&+E_{\mathsf{q}}\left\{\log \mathfrak{p}\left(a_{\sigma^2}\right)\right\} - E_{\mathsf{q}}\left\{\log \mathsf{q}^*\left(a_{\sigma^2}\right)\right\} + E_{\mathsf{q}}\left\{\log \mathfrak{p}\left(\boldsymbol{\Sigma}|\boldsymbol{A}_{\boldsymbol{\Sigma}}\right)\right\}\\
&-E_{\mathsf{q}}\left\{\log \mathsf{q}^*\left(\boldsymbol{\Sigma}\right)\right\} + E_{\mathsf{q}}\left\{\log \mathfrak{p}\left(\boldsymbol{A}_{\boldsymbol{\Sigma}}\right)\right\} - E_{\mathsf{q}}\left\{\log \mathsf{q}^*\left(\boldsymbol{A}_{\boldsymbol{\Sigma}}\right)\right\}\\
&+E_{\mathsf{q}}\left\{\log \mathfrak{p}\left(\boldsymbol{\Sigma}'|\boldsymbol{A}_{\boldsymbol{\Sigma}'}\right)\right\} - E_{\mathsf{q}}\left\{\log \mathsf{q}^*\left(\boldsymbol{\Sigma}'\right)\right\} + E_{\mathsf{q}}\left\{\log \mathfrak{p}\left(\boldsymbol{A}_{\boldsymbol{\Sigma}'}\right)\right\}\\
&-E_{\mathsf{q}}\left\{\log \mathsf{q}^*\left(\boldsymbol{A}_{\boldsymbol{\Sigma}'}\right)\right\}.
\end{aligned}$$

The first of the $\log \underline{\mathfrak{p}}(\boldsymbol{y};\mathfrak{q})$ terms is

$$
\begin{aligned}
E_{\mathfrak{q}}\left\{\mathfrak{p}\left(\boldsymbol{y}|\boldsymbol{\beta},\boldsymbol{u},\boldsymbol{u}',\sigma^2\right)\right\} = &-\tfrac{1}{2}n_{\bullet\bullet}\log(2\pi) - \tfrac{1}{2}n_{\bullet\bullet}E_{\mathfrak{q}}\left\{\log(\sigma^2)\right\} \\
&-\tfrac{1}{2}\mu_{\mathfrak{q}(1/\sigma^2)}\sum_{i=1}^{m}\sum_{i'=1}^{m'}\left\{\left|\left|\boldsymbol{y}_{ii'} - \boldsymbol{X}_{ii'}\boldsymbol{\mu}_{\mathfrak{q}(\boldsymbol{\beta})} - \boldsymbol{Z}_{ii'}\boldsymbol{\mu}_{\mathfrak{q}(\boldsymbol{u}_i)} - \boldsymbol{Z}'_{ii'}\boldsymbol{\mu}_{\mathfrak{q}(\boldsymbol{u}'_{i'})}\right|\right|^2\right. \\
&\quad +\text{tr}\left(\boldsymbol{X}_{ii'}^T\boldsymbol{X}_{ii'}\boldsymbol{\Sigma}_{\mathfrak{q}(\boldsymbol{\beta})}\right) + \text{tr}\left(\boldsymbol{Z}_{ii'}^T\boldsymbol{Z}_{ii'}\boldsymbol{\Sigma}_{\mathfrak{q}(\boldsymbol{u}_i)}\right) \\
&\quad +\text{tr}\left(\left(\boldsymbol{Z}'_{ii'}\right)^T\boldsymbol{Z}'_{ii'}\boldsymbol{\Sigma}_{\mathfrak{q}(\boldsymbol{u}'_{i'})}\right) \\
&\quad +2\text{tr}\left[\boldsymbol{Z}_{ii'}^T\boldsymbol{X}_{ii'}E_{\mathfrak{q}}\{(\boldsymbol{\beta}-\boldsymbol{\mu}_{\mathfrak{q}(\boldsymbol{\beta})})(\boldsymbol{u}_i-\boldsymbol{\mu}_{\mathfrak{q}(\boldsymbol{u}_i)})^T\}\right] \\
&\quad +2\text{tr}\left[\left(\boldsymbol{Z}'_{ii'}\right)^T\boldsymbol{X}_{ii'}E_{\mathfrak{q}}\{(\boldsymbol{\beta}-\boldsymbol{\mu}_{\mathfrak{q}(\boldsymbol{\beta})})(\boldsymbol{u}'_{i'}-\boldsymbol{\mu}_{\mathfrak{q}(\boldsymbol{u}'_{i'})})^T\}\right] \\
&\quad \left. +2\text{tr}\left[\boldsymbol{Z}_{ii'}^T\boldsymbol{Z}'_{ii'}E_{\mathfrak{q}}\{(\boldsymbol{u}_i-\boldsymbol{\mu}_{\mathfrak{q}(\boldsymbol{u}_i)})(\boldsymbol{u}'_{i'}-\boldsymbol{\mu}_{\mathfrak{q}(\boldsymbol{u}'_{i'})})^T\}\right]\right\}.
\end{aligned}
$$

Under product restrictions I and II, $E_{\mathfrak{q}}\left\{\mathfrak{p}\left(\boldsymbol{y}|\boldsymbol{\beta},\boldsymbol{u},\boldsymbol{u}',\sigma^2\right)\right\}$ simplify further as we have

$$E_{\mathfrak{q}}\{(\boldsymbol{\beta}-\boldsymbol{\mu}_{\mathfrak{q}(\boldsymbol{\beta})})(\boldsymbol{u}'_{i'}-\boldsymbol{\mu}_{\mathfrak{q}(\boldsymbol{u}'_{i'})})^T\} = \boldsymbol{O}, \quad 1 \le i' \le m',$$

and

$$E_{\mathfrak{q}}\{(\boldsymbol{u}_i-\boldsymbol{\mu}_{\mathfrak{q}(\boldsymbol{u}_i)})(\boldsymbol{u}'_{i'}-\boldsymbol{\mu}_{\mathfrak{q}(\boldsymbol{u}'_{i'})})^T\} = \boldsymbol{O}, \quad 1 \le i \le m, 1 \le i' \le m'.$$

Under product restriction I we also have

$$E_{\mathfrak{q}}\{(\boldsymbol{\beta}-\boldsymbol{\mu}_{\mathfrak{q}(\boldsymbol{\beta})})(\boldsymbol{u}_i-\boldsymbol{\mu}_{\mathfrak{q}(\boldsymbol{u}_i)})^T\} = \boldsymbol{O}, \quad 1 \le i \le m.$$

The second of the $\log \underline{\mathfrak{p}}(\boldsymbol{y};\mathfrak{q})$ terms is

$$
\begin{aligned}
E_{\mathfrak{q}}[\log\{\mathfrak{p}(\boldsymbol{\beta},\boldsymbol{u},\boldsymbol{u}' \mid \boldsymbol{\Sigma},\boldsymbol{\Sigma}')\}] &= E_{\mathfrak{q}}[\log\{\mathfrak{p}(\boldsymbol{\beta})\} + \log\{\mathfrak{p}(\boldsymbol{u}|\boldsymbol{\Sigma})\} + \log\{\mathfrak{p}(\boldsymbol{u}'|\boldsymbol{\Sigma}')\}] \\
&= -\tfrac{1}{2}(p+mq+m'q')\log(2\pi) - \tfrac{1}{2}\log|\boldsymbol{\Sigma}_{\boldsymbol{\beta}}| - \frac{m}{2}E_{\mathfrak{q}}\left\{\log|\boldsymbol{\Sigma}|\right\} \\
&\quad -\frac{m'}{2}E_{\mathfrak{q}}\left\{\log|\boldsymbol{\Sigma}'|\right\} \\
&\quad -\tfrac{1}{2}\text{tr}\left(\boldsymbol{\Sigma}_{\boldsymbol{\beta}}^{-1}\left\{\left(\boldsymbol{\mu}_{\mathfrak{q}(\boldsymbol{\beta})}-\boldsymbol{\mu}_{\boldsymbol{\beta}}\right)\left(\boldsymbol{\mu}_{\mathfrak{q}(\boldsymbol{\beta})}-\boldsymbol{\mu}_{\boldsymbol{\beta}}\right)^T + \boldsymbol{\Sigma}_{\mathfrak{q}(\boldsymbol{\beta})}\right\}\right) \\
&\quad -\tfrac{1}{2}\text{tr}\left(\boldsymbol{M}_{\mathfrak{q}(\boldsymbol{\Sigma}^{-1})}\left\{\sum_{i=1}^{m}\left(\boldsymbol{\mu}_{\mathfrak{q}(\boldsymbol{u}_i)}\boldsymbol{\mu}_{\mathfrak{q}(\boldsymbol{u}_i)}^T + \boldsymbol{\Sigma}_{\mathfrak{q}(\boldsymbol{u}_i)}\right)\right\}\right) \\
&\quad -\tfrac{1}{2}\text{tr}\left(\boldsymbol{M}_{\mathfrak{q}\left((\boldsymbol{\Sigma}')^{-1}\right)}\left\{\sum_{i'=1}^{m'}\left(\boldsymbol{\mu}_{\mathfrak{q}(\boldsymbol{u}'_{i'})}\boldsymbol{\mu}_{\mathfrak{q}(\boldsymbol{u}'_{i'})}^T + \boldsymbol{\Sigma}_{\mathfrak{q}(\boldsymbol{u}'_{i'})}\right)\right\}\right).
\end{aligned}
$$

The third of the $\log \underline{\mathfrak{p}}(\boldsymbol{y};\mathfrak{q})$ terms is the negative of

$$E_{\mathfrak{q}}[\log\{\mathfrak{q}(\boldsymbol{\beta},\boldsymbol{u},\boldsymbol{u}')\}] \;=\; -\tfrac{1}{2}(p+mq+m'q') - \tfrac{1}{2}(p+mq+m'q')\log(2\pi) - \tfrac{1}{2}\log|\boldsymbol{\Sigma}_{\mathfrak{q}(\boldsymbol{\beta},\boldsymbol{u},\boldsymbol{u}')}|.$$

The fourth of the $\log \underline{\mathfrak{p}}(\boldsymbol{y};\mathfrak{q})$ terms is

$$
\begin{aligned}
E_{\mathfrak{q}}[\log\{\mathfrak{p}(\sigma^2|a_{\sigma^2})\}] \;&=\; E_{\mathfrak{q}}\left(\log\left[\frac{\{1/(2a_{\sigma^2})\}^{\nu_{\sigma^2}/2}}{\Gamma(\nu_{\sigma^2}/2)}(\sigma^2)^{-(\nu_{\sigma^2}/2)-1}\exp\{-1/(2a_{\sigma^2}\sigma^2)\}\right]\right) \\
&=\; -\tfrac{1}{2}\nu_{\sigma^2}\,E_{\mathfrak{q}}\{\log(2a_{\sigma^2})\} - \log\{\Gamma(\tfrac{1}{2}\nu_{\sigma^2})\} - (\tfrac{1}{2}\nu_{\sigma^2}+1)E_{\mathfrak{q}}\{\log(\sigma^2)\} \\
&\quad -\tfrac{1}{2}\mu_{\mathfrak{q}(1/a_{\sigma^2})}\mu_{\mathfrak{q}(1/\sigma^2)}.
\end{aligned}
$$

The fifth of the $\log \underline{\mathfrak{p}}(\boldsymbol{y};\mathfrak{q})$ terms is the negative of

$$\begin{aligned}
E_{\mathfrak{q}}[\log\{\mathfrak{q}(\sigma^2)\}] &= E_{\mathfrak{q}}\left(\log\left[\frac{\{\lambda_{\mathfrak{q}(\sigma^2)}/2\}^{\xi_{\mathfrak{q}(\sigma^2)}/2}}{\Gamma(\xi_{\mathfrak{q}(\sigma^2)}/2)}(\sigma^2)^{-(\xi_{\mathfrak{q}(\sigma^2)}/2)-1}\exp\{-\lambda_{\mathfrak{q}(\sigma^2)}/(2\sigma^2)\}\right]\right)\\
&= \tfrac{1}{2}\xi_{\mathfrak{q}(\sigma^2)}\,\log(\lambda_{\mathfrak{q}(\sigma^2)}/2)-\log\{\Gamma(\tfrac{1}{2}\xi_{\mathfrak{q}(\sigma^2)})\}-(\tfrac{1}{2}\xi_{\mathfrak{q}(\sigma^2)}+1)E_{\mathfrak{q}}\{\log(\sigma^2)\}\\
&\quad -\tfrac{1}{2}\lambda_{\mathfrak{q}(\sigma^2)}\mu_{\mathfrak{q}(1/\sigma^2)}.
\end{aligned}$$

The sixth of the $\log \underline{\mathfrak{p}}(\boldsymbol{y};\mathfrak{q})$ terms is

$$\begin{aligned}
E_{\mathfrak{q}}[\log\{\mathfrak{p}(a_{\sigma^2})\}] &= E_{\mathfrak{q}}\left(\log\left[\frac{\{1/(2\nu_{\sigma^2}s_{\sigma^2}^2)\}^{1/2}}{\Gamma(1/2)}a_{\sigma^2}^{-(1/2)-1}\exp\{-1/(2\nu_{\sigma^2}s_{\sigma^2}^2a_{\sigma^2})\}\right]\right)\\
&= -\tfrac{1}{2}\,\log(2\nu_{\sigma^2}s_{\sigma^2}^2)-\log\{\Gamma(\tfrac{1}{2})\}-(\tfrac{1}{2}+1)E_{\mathfrak{q}}\{\log(a_{\sigma^2})\}-\{1/(2\nu_{\sigma^2}s_{\sigma^2}^2)\}\mu_{\mathfrak{q}(1/a_{\sigma^2})}.
\end{aligned}$$

The seventh of the $\log \underline{\mathfrak{p}}(\boldsymbol{y};\mathfrak{q})$ terms is the negative of

$$\begin{aligned}
E_{\mathfrak{q}}[\log\{\mathfrak{q}(a_{\sigma^2})\}] &= E_{\mathfrak{q}}\left(\log\left[\frac{\{\lambda_{\mathfrak{q}(a_{\sigma^2})}/2\}^{\xi_{\mathfrak{q}(a_{\sigma^2})}/2}}{\Gamma(\xi_{\mathfrak{q}(a_{\sigma^2})}/2)}(a_{\sigma^2})^{-(\xi_{\mathfrak{q}(a_{\sigma^2})}/2)-1}\exp\{-\lambda_{\mathfrak{q}(a_{\sigma^2})}/(2a_{\sigma^2})\}\right]\right)\\
&= \tfrac{1}{2}\xi_{\mathfrak{q}(a_{\sigma^2})}\,\log(\lambda_{\mathfrak{q}(a_{\sigma^2})}/2)-\log\{\Gamma(\tfrac{1}{2}\xi_{\mathfrak{q}(a_{\sigma^2})})\}-(\tfrac{1}{2}\xi_{\mathfrak{q}(a_{\sigma^2})}+1)E_{\mathfrak{q}}\{\log(a_{\sigma^2})\}\\
&\quad -\tfrac{1}{2}\lambda_{\mathfrak{q}(a_{\sigma^2})}\mu_{\mathfrak{q}(1/a_{\sigma^2})}.
\end{aligned}$$

The eighth of the $\log \underline{\mathfrak{p}}(\boldsymbol{y};\mathfrak{q})$ terms is

$$\begin{aligned}
E_{\mathfrak{q}}[\log\{\mathfrak{p}(\boldsymbol{\Sigma}|\boldsymbol{A}_{\boldsymbol{\Sigma}})\}] &= E_{\mathfrak{q}}\left(\frac{|\boldsymbol{A}_{\boldsymbol{\Sigma}}|^{-\frac{1}{2}(\nu_{\boldsymbol{\Sigma}}+q-1)}|\boldsymbol{\Sigma}|^{-\frac{1}{2}(\nu_{\boldsymbol{\Sigma}}+2q)}}{2^{\frac{q}{2}(\nu_{\boldsymbol{\Sigma}}+2q-1)}\pi^{\frac{q}{4}(q-1)}\prod_{j=1}^{q}\Gamma(\frac{1}{2}(\nu_{\boldsymbol{\Sigma}}+2q-j))}\exp\{-\tfrac{1}{2}\text{tr}(\boldsymbol{A}_{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\Sigma}^{-1})\}\right)\\
&= -\tfrac{1}{2}(\nu_{\boldsymbol{\Sigma}}+q-1)E_{\mathfrak{q}}\{\log|\boldsymbol{A}_{\boldsymbol{\Sigma}}|\}-\tfrac{1}{2}(\nu_{\boldsymbol{\Sigma}}+2q)E_{\mathfrak{q}}\{\log|\boldsymbol{\Sigma}|\}\\
&\quad -\tfrac{1}{2}\text{tr}(\boldsymbol{M}_{\mathfrak{q}(\boldsymbol{A}_{\boldsymbol{\Sigma}}^{-1})}\boldsymbol{M}_{\mathfrak{q}(\boldsymbol{\Sigma}^{-1})})-\frac{q}{2}(\nu_{\boldsymbol{\Sigma}}+2q-1)\log(2)-\frac{q}{4}(q-1)\log(\pi)\\
&\quad -\sum_{j=1}^{q}\log\Gamma(\tfrac{1}{2}(\nu_{\boldsymbol{\Sigma}}+2q-j)).
\end{aligned}$$

The ninth of the $\log \underline{\mathfrak{p}}(\boldsymbol{y};\mathfrak{q})$ terms is the negative of

$$\begin{aligned}
E_{\mathfrak{q}}[\log\{\mathfrak{q}(\boldsymbol{\Sigma})\}] &= E_{\mathfrak{q}}\left(\frac{|\boldsymbol{\Lambda}_{\mathfrak{q}(\boldsymbol{\Sigma})}|^{\frac{1}{2}(\xi_{\mathfrak{q}(\boldsymbol{\Sigma})}-q+1)}|\boldsymbol{\Sigma}|^{-\frac{1}{2}(\xi_{\mathfrak{q}(\boldsymbol{\Sigma})}+2)}}{2^{\frac{q}{2}(\xi_{\mathfrak{q}(\boldsymbol{\Sigma})}+1)}\pi^{\frac{q}{4}(q-1)}\prod_{j=1}^{q}\Gamma(\frac{1}{2}(\xi_{\mathfrak{q}(\boldsymbol{\Sigma})}+2-j))}\exp\{-\tfrac{1}{2}\text{tr}(\boldsymbol{\Lambda}_{\mathfrak{q}(\boldsymbol{\Sigma})}\boldsymbol{\Sigma}^{-1})\}\right)\\
&= \tfrac{1}{2}(\xi_{\mathfrak{q}(\boldsymbol{\Sigma})}-q+1)\log|\boldsymbol{\Lambda}_{\mathfrak{q}(\boldsymbol{\Sigma})}|-\tfrac{1}{2}(\xi_{\mathfrak{q}(\boldsymbol{\Sigma})}+2)E_{\mathfrak{q}}\{\log|\boldsymbol{\Sigma}|\}-\tfrac{1}{2}\text{tr}(\boldsymbol{\Lambda}_{\mathfrak{q}(\boldsymbol{\Sigma})}\boldsymbol{M}_{\mathfrak{q}(\boldsymbol{\Sigma}^{-1})})\\
&\quad -\frac{q}{2}(\xi_{\mathfrak{q}(\boldsymbol{\Sigma})}+1)\log(2)-\frac{q}{4}(q-1)\log(\pi)-\sum_{j=1}^{q}\log\Gamma(\tfrac{1}{2}(\xi_{\mathfrak{q}(\boldsymbol{\Sigma})}+2-j)).
\end{aligned}$$

The tenth of the $\log \underline{\mathfrak{p}}(\boldsymbol{y};\mathfrak{q})$ terms is

$$\begin{aligned}
E_{\mathfrak{q}}[\log\{\mathfrak{p}(\boldsymbol{A}_{\boldsymbol{\Sigma}})\}] &= E_{\mathfrak{q}}\left(\frac{|\boldsymbol{\Lambda}_{\boldsymbol{A}_{\boldsymbol{\Sigma}}}|^{\frac{1}{2}(2-q)}|\boldsymbol{A}_{\boldsymbol{\Sigma}}|^{-3/2}}{2^{q}\pi^{\frac{q}{4}(q-1)}\prod_{j=1}^{q}\Gamma(\frac{1}{2}(3-j))}\exp\{-\tfrac{1}{2}\text{tr}(\boldsymbol{\Lambda}_{\boldsymbol{A}_{\boldsymbol{\Sigma}}}\boldsymbol{A}_{\boldsymbol{\Sigma}}^{-1})\}\right)\\
&= -\tfrac{1}{2}q(2-q)\log(\nu_{\boldsymbol{\Sigma}})-\tfrac{1}{2}(2-q)\sum_{j=1}^{q}\log(s_{\boldsymbol{\Sigma},j}^2)-\frac{3}{2}E_{\mathfrak{q}}\{\log|\boldsymbol{A}_{\boldsymbol{\Sigma}}|\}\\
&\quad -\tfrac{1}{2}\sum_{j=1}^{q}1/(\nu_{\boldsymbol{\Sigma}}s_{\boldsymbol{\Sigma},j}^2)\left(\boldsymbol{M}_{\mathfrak{q}(\boldsymbol{A}_{\boldsymbol{\Sigma}}^{-1})}\right)_{jj}-q\log(2)-\frac{q}{4}(q-1)\log(\pi)\\
&\quad -\sum_{j=1}^{q}\log\Gamma(\tfrac{1}{2}(3-j)).
\end{aligned}$$

The eleventh of the $\log \underline{\mathfrak{p}}(\boldsymbol{y};\mathfrak{q})$ terms is the negative of

$$
\begin{aligned}
E_{\mathfrak{q}}[\log\{\mathfrak{q}(\boldsymbol{A}_{\boldsymbol{\Sigma}})\}] &= E_{\mathfrak{q}}\left(\frac{|\boldsymbol{\Lambda}_{\mathfrak{q}(\boldsymbol{A}_{\boldsymbol{\Sigma}})}|^{\frac{1}{2}(\xi_{\mathfrak{q}(\boldsymbol{A}_{\boldsymbol{\Sigma}})}-q+1)}|\boldsymbol{A}_{\boldsymbol{\Sigma}}|^{-\frac{1}{2}(\xi_{\mathfrak{q}(\boldsymbol{A}_{\boldsymbol{\Sigma}})}+2)}}{2^{\frac{q}{2}(\xi_{\mathfrak{q}(\boldsymbol{A}_{\boldsymbol{\Sigma}})}+1)}\pi^{\frac{q}{4}(q-1)}\prod_{j=1}^{q}\Gamma(\frac{1}{2}(\xi_{\mathfrak{q}(\boldsymbol{A}_{\boldsymbol{\Sigma}})}+2-j))}\exp\{-\tfrac{1}{2}\mathrm{tr}(\boldsymbol{\Lambda}_{\mathfrak{q}(\boldsymbol{A}_{\boldsymbol{\Sigma}})}^{-1}\boldsymbol{A}_{\boldsymbol{\Sigma}}^{-1})\}\right)\\
&= \tfrac{1}{2}(\xi_{\mathfrak{q}(\boldsymbol{A}_{\boldsymbol{\Sigma}})}-q+1)\log|\boldsymbol{\Lambda}_{\mathfrak{q}(\boldsymbol{A}_{\boldsymbol{\Sigma}})}|-\tfrac{1}{2}(\xi_{\mathfrak{q}(\boldsymbol{A}_{\boldsymbol{\Sigma}})}+2)E_{\mathfrak{q}}\{\log|\boldsymbol{A}_{\boldsymbol{\Sigma}}|\}\\
&\quad -\tfrac{1}{2}\mathrm{tr}(\boldsymbol{\Lambda}_{\mathfrak{q}(\boldsymbol{A}_{\boldsymbol{\Sigma}})}\boldsymbol{M}_{\mathfrak{q}(\boldsymbol{A}_{\boldsymbol{\Sigma}}^{-1})})-\frac{q}{2}(\xi_{\mathfrak{q}(\boldsymbol{A}_{\boldsymbol{\Sigma}})}+1)\log(2)-\frac{q}{4}(q-1)\log(\pi)\\
&\quad -\sum_{j=1}^{q}\log\Gamma(\tfrac{1}{2}(\xi_{\mathfrak{q}(\boldsymbol{A}_{\boldsymbol{\Sigma}})}+2-j)).
\end{aligned}
$$

The remaining four terms of $\log \underline{\mathfrak{p}}(\boldsymbol{y};\mathfrak{q})$ are

$$
\begin{aligned}
E_{\mathfrak{q}}[\log\{\mathfrak{p}(\boldsymbol{\Sigma}'|\boldsymbol{A}_{\boldsymbol{\Sigma}'})\}] &= -\tfrac{1}{2}(\nu_{\boldsymbol{\Sigma}'}+q'-1)E_{\mathfrak{q}}\{\log|\boldsymbol{A}_{\boldsymbol{\Sigma}'}|\}-\tfrac{1}{2}(\nu_{\boldsymbol{\Sigma}'}+2q')E_{\mathfrak{q}}\{\log|\boldsymbol{\Sigma}'|\}\\
&\quad -\tfrac{1}{2}\mathrm{tr}(\boldsymbol{M}_{\mathfrak{q}((\boldsymbol{A}_{\boldsymbol{\Sigma}'})^{-1})}\boldsymbol{M}_{\mathfrak{q}((\boldsymbol{\Sigma})^{-1})})-\frac{q'}{2}(\nu_{\boldsymbol{\Sigma}'}+2q'-1)\log(2)-\frac{q'}{4}(q'-1)\log(\pi)\\
&\quad -\sum_{j=1}^{q'}\log\Gamma(\tfrac{1}{2}(\nu_{\boldsymbol{\Sigma}'}+2q'-j)),
\end{aligned}
$$

the negative of

$$
\begin{aligned}
E_{\mathfrak{q}}[\log\{\mathfrak{q}(\boldsymbol{\Sigma}')\}] = {} & \tfrac{1}{2}(\xi_{\mathfrak{q}(\boldsymbol{\Sigma}')}-q'+1)\log|\boldsymbol{\Lambda}_{\mathfrak{q}(\boldsymbol{\Sigma}')}|-\tfrac{1}{2}(\xi_{\mathfrak{q}(\boldsymbol{\Sigma}')}+2)E_{\mathfrak{q}}\{\log|\boldsymbol{\Sigma}'|\}-\tfrac{1}{2}\mathrm{tr}(\boldsymbol{\Lambda}_{\mathfrak{q}(\boldsymbol{\Sigma}')}\boldsymbol{M}_{\mathfrak{q}((\boldsymbol{\Sigma}')^{-1})})\\
& -\frac{q'}{2}(\xi_{\mathfrak{q}(\boldsymbol{\Sigma}')}+1)\log(2)-\frac{q'}{4}(q'-1)\log(\pi)-\sum_{j=1}^{q'}\log\Gamma(\tfrac{1}{2}(\xi_{\mathfrak{q}(\boldsymbol{\Sigma}')}+2-j)),
\end{aligned}
$$

$$
\begin{aligned}
E_{\mathfrak{q}}[\log\{\mathfrak{p}(\boldsymbol{A}_{\boldsymbol{\Sigma}'})\}] = {} & -\tfrac{1}{2}q'(2-q')\log(\nu_{\boldsymbol{\Sigma}'})-\tfrac{1}{2}(2-q')\sum_{j'=1}^{q'}\log(s_{\boldsymbol{\Sigma},j}^2)-\frac{3}{2}E_{\mathfrak{q}}\{\log|\boldsymbol{A}_{\boldsymbol{\Sigma}}|\}\\
& -\tfrac{1}{2}\sum_{j=1}^{q'}1/(\nu_{\boldsymbol{\Sigma}'}s_{\boldsymbol{\Sigma},j}^2)\left(\boldsymbol{M}_{\mathfrak{q}((\boldsymbol{A}_{\boldsymbol{\Sigma}'})^{-1})}\right)_{jj}-q'\log(2)-\frac{q'}{4}(q'-1)\log(\pi)\\
& -\sum_{j=1}^{q'}\log\Gamma(\tfrac{1}{2}(3-j)),
\end{aligned}
$$

and the negative of

$$
\begin{aligned}
E_{\mathfrak{q}}[\log\{\mathfrak{q}(\boldsymbol{A}_{\boldsymbol{\Sigma}'})\}] = {} & \tfrac{1}{2}(\xi_{\mathfrak{q}(\boldsymbol{A}_{\boldsymbol{\Sigma}'})}-q'+1)\log|\boldsymbol{\Lambda}_{\mathfrak{q}(\boldsymbol{A}_{\boldsymbol{\Sigma}'})}|-\tfrac{1}{2}(\xi_{\mathfrak{q}(\boldsymbol{A}_{\boldsymbol{\Sigma}'})}+2)E_{\mathfrak{q}}\{\log|\boldsymbol{A}_{\boldsymbol{\Sigma}'}|\}\\
& -\tfrac{1}{2}\mathrm{tr}(\boldsymbol{\Lambda}_{\mathfrak{q}(\boldsymbol{A}_{\boldsymbol{\Sigma}'})}\boldsymbol{M}_{\mathfrak{q}((\boldsymbol{A}_{\boldsymbol{\Sigma}'})^{-1})})-\frac{q'}{2}(\xi_{\mathfrak{q}(\boldsymbol{A}_{\boldsymbol{\Sigma}'})}+1)\log(2)-\frac{q'}{4}(q'-1)\log(\pi)\\
& -\sum_{j=1}^{q'}\log\Gamma(\tfrac{1}{2}(\xi_{\mathfrak{q}(\boldsymbol{A}_{\boldsymbol{\Sigma}'})}+2-j)).
\end{aligned}
$$

In the summation of each of these $\log \underline{\mathfrak{p}}(\boldsymbol{x};q)$ terms, note that the coefficient of $E_{\mathfrak{q}}\{\log(\sigma^2)\}$ is

$$-\tfrac{1}{2}\,n_{\bullet\bullet}-\tfrac{1}{2}\nu_{\sigma^2}-1+\tfrac{1}{2}\xi_{\mathfrak{q}(\sigma^2)}+1=-\tfrac{1}{2}\,n_{\bullet\bullet}-\tfrac{1}{2}\nu_{\sigma^2}-1+\tfrac{1}{2}(\nu_{\sigma^2}+n_{\bullet\bullet})+1=0.$$

The coefficient of $E_{\mathfrak{q}}\{\log(a_{\sigma^2})\}$ is

$$-\tfrac{1}{2}\nu_{\sigma^2}-(\tfrac{1}{2}+1)+\tfrac{1}{2}\xi_{\mathfrak{q}(a_{\sigma^2})}+1=-\tfrac{1}{2}\nu_{\sigma^2}-(\tfrac{1}{2}+1)+\tfrac{1}{2}(\nu_{\sigma^2}+1)+1=0.$$

The coefficient of $E_{\mathfrak{q}}\{\log|\mathbf{\Sigma}|\}$ is

$$-\frac{m}{2} - \tfrac{1}{2}(\nu_{\mathbf{\Sigma}} + 2q) + \tfrac{1}{2}(\xi_{\mathfrak{q}(\mathbf{\Sigma})} + 2) = -\tfrac{1}{2}(m + \nu_{\mathbf{\Sigma}} + 2q) + \tfrac{1}{2}(m + \nu_{\mathbf{\Sigma}} + 2q) = 0.$$

The coefficient of $E_{\mathfrak{q}}\{\log|\boldsymbol{A}_{\mathbf{\Sigma}}|\}$ is

$$-\tfrac{1}{2}(\nu_{\mathbf{\Sigma}} + q - 1) - \frac{3}{2} + \tfrac{1}{2}(\xi_{\mathfrak{q}(\boldsymbol{A}_{\mathbf{\Sigma}})} + 2) = -\tfrac{1}{2}(\nu_{\mathbf{\Sigma}} + q + 2) + \tfrac{1}{2}(\nu_{\mathbf{\Sigma}} + q + 2) = 0.$$

The coefficient of $E_{\mathfrak{q}}\{\log|\mathbf{\Sigma}'|\}$ is

$$-\frac{m'}{2} - \tfrac{1}{2}(\nu_{\mathbf{\Sigma}'} + 2q') + \tfrac{1}{2}(\xi_{\mathfrak{q}(\mathbf{\Sigma}')} + 2) = -\tfrac{1}{2}(m' + \nu_{\mathbf{\Sigma}'} + 2q') + \tfrac{1}{2}(m' + \nu_{\mathbf{\Sigma}'} + 2q') = 0.$$

The coefficient of $E_{\mathfrak{q}}\{\log|\boldsymbol{A}_{\mathbf{\Sigma}'}|\}$ is

$$-\tfrac{1}{2}(\nu_{\mathbf{\Sigma}'} + q' - 1) - \frac{3}{2} + \tfrac{1}{2}(\xi_{\mathfrak{q}(\boldsymbol{A}_{\mathbf{\Sigma}'})} + 2) = -\tfrac{1}{2}(\nu_{\mathbf{\Sigma}'} + q' + 2) + \tfrac{1}{2}(\nu_{\mathbf{\Sigma}'} + q' + 2) = 0.$$

Therefore, the terms in $E_{\mathfrak{q}}\{\log(\sigma^2)\}$, $E_{\mathfrak{q}}\{\log(a)\}$, $E_{\mathfrak{q}}\{\log|\mathbf{\Sigma}|\}$ and $E_{\mathfrak{q}}\{\log|\boldsymbol{A}_{\mathbf{\Sigma}}|\}$ can be dropped and we then have

$$\log \underline{\mathfrak{p}}(\boldsymbol{y};\mathfrak{q}) = \sum_{i=1}^{15} T_i$$

where

$$\begin{aligned}
T_1 = &-\tfrac{1}{2}n_{\bullet\bullet}\log(2\pi)\\
&-\tfrac{1}{2}\mu_{\mathfrak{q}(1/\sigma^2)}\sum_{i=1}^{m}\sum_{i'=1}^{m'}\Bigg\{\left|\left|\boldsymbol{y}_{ii'} - \boldsymbol{X}_{ii'}\boldsymbol{\mu}_{\mathfrak{q}(\boldsymbol{\beta})} - \boldsymbol{Z}_{ii'}\boldsymbol{\mu}_{\mathfrak{q}(\boldsymbol{u}_i)} - \boldsymbol{Z}'_{ii'}\boldsymbol{\mu}_{\mathfrak{q}(\boldsymbol{u}'_{i'})}\right|\right|^2\\
&\quad +\text{tr}\left(\boldsymbol{X}_{ii'}^T\boldsymbol{X}_{ii'}\mathbf{\Sigma}_{\mathfrak{q}(\boldsymbol{\beta})}\right) + \text{tr}\left(\boldsymbol{Z}_{ii'}^T\boldsymbol{Z}_{ii'}\mathbf{\Sigma}_{\mathfrak{q}(\boldsymbol{u}_i)}\right) + \text{tr}\left(\left(\boldsymbol{Z}'_{ii'}\right)^T\boldsymbol{Z}'_{ii'}\mathbf{\Sigma}_{\mathfrak{q}(\boldsymbol{u}'_{i'})}\right)\\
&\quad +2\text{tr}\left[\boldsymbol{Z}_{ii'}^T\boldsymbol{X}_{ii'}E_{\mathfrak{q}}\left\{\left(\boldsymbol{\beta} - \boldsymbol{\mu}_{\mathfrak{q}(\boldsymbol{\beta})}\right)\left(\boldsymbol{u}_i - \boldsymbol{\mu}_{\mathfrak{q}(\boldsymbol{u}_i)}\right)^T\right\}\right]\\
&\quad +2\text{tr}\left[\left(\boldsymbol{Z}'_{ii'}\right)^T\boldsymbol{X}_{ii'}E_{\mathfrak{q}}\left\{\left(\boldsymbol{\beta} - \boldsymbol{\mu}_{\mathfrak{q}(\boldsymbol{\beta})}\right)\left(\boldsymbol{u}'_{i'} - \boldsymbol{\mu}_{\mathfrak{q}(\boldsymbol{u}'_{i'})}\right)^T\right\}\right]\\
&\quad +2\text{tr}\left[\boldsymbol{Z}_{ii'}^T\boldsymbol{Z}'_{ii'}E_{\mathfrak{q}}\left\{\left(\boldsymbol{u}_i - \boldsymbol{\mu}_{\mathfrak{q}(\boldsymbol{u}_i)}\right)\left(\boldsymbol{u}'_{i'} - \boldsymbol{\mu}_{\mathfrak{q}(\boldsymbol{u}'_{i'})}\right)^T\right\}\right]\Bigg\},\\
T_2 = &-\tfrac{1}{2}(p + mq + m'q')\log(2\pi) - \tfrac{1}{2}\log|\mathbf{\Sigma}_{\boldsymbol{\beta}}|\\
&-\tfrac{1}{2}\text{tr}\left(\mathbf{\Sigma}_{\boldsymbol{\beta}}^{-1}\left\{\left(\boldsymbol{\mu}_{\mathfrak{q}(\boldsymbol{\beta})} - \boldsymbol{\mu}_{\boldsymbol{\beta}}\right)\left(\boldsymbol{\mu}_{\mathfrak{q}(\boldsymbol{\beta})} - \boldsymbol{\mu}_{\boldsymbol{\beta}}\right)^T + \mathbf{\Sigma}_{\mathfrak{q}(\boldsymbol{\beta})}\right\}\right)\\
&-\tfrac{1}{2}\text{tr}\left(\boldsymbol{M}_{\mathfrak{q}(\mathbf{\Sigma}^{-1})}\left\{\sum_{i=1}^{m}\left(\boldsymbol{\mu}_{\mathfrak{q}(\boldsymbol{u}_i)}\boldsymbol{\mu}_{\mathfrak{q}(\boldsymbol{u}_i)}^T + \mathbf{\Sigma}_{\mathfrak{q}(\boldsymbol{u}_i)}\right)\right\}\right)\\
&-\tfrac{1}{2}\text{tr}\left(\boldsymbol{M}_{\mathfrak{q}\left((\mathbf{\Sigma}')^{-1}\right)}\left\{\sum_{i'=1}^{m'}\left(\boldsymbol{\mu}_{\mathfrak{q}(\boldsymbol{u}'_{i'})}\boldsymbol{\mu}_{\mathfrak{q}(\boldsymbol{u}'_{i'})}^T + \mathbf{\Sigma}_{\mathfrak{q}(\boldsymbol{u}'_{i'})}\right)\right\}\right),\\
T_3 = &\ \tfrac{1}{2}(p + mq + m'q') + \tfrac{1}{2}(p + mq + m'q')\log(2\pi) + \tfrac{1}{2}\log\left|\mathbf{\Sigma}_{\mathfrak{q}(\boldsymbol{\beta},\boldsymbol{u},\boldsymbol{u}')}\right|,\\
T_4 = &\ \tfrac{1}{2}\nu_{\sigma^2}\log(2) - \log\{\Gamma(\tfrac{1}{2}\nu_{\sigma^2})\} - \tfrac{1}{2}\mu_{\mathfrak{q}(1/a_{\sigma^2})}\mu_{\mathfrak{q}(1/\sigma^2)},\\
T_5 = &-\tfrac{1}{2}\xi_{\mathfrak{q}(\sigma^2)}\log(\lambda_{\mathfrak{q}(\sigma^2)}/2) + \log\{\Gamma(\tfrac{1}{2}\xi_{\mathfrak{q}(\sigma^2)})\} + \tfrac{1}{2}\lambda_{\mathfrak{q}(\sigma^2)}\mu_{\mathfrak{q}(1/\sigma^2)},\\
T_6 = &-\tfrac{1}{2}\log(2\nu_{\sigma^2}s_{\sigma^2}^2) - \log\{\Gamma(\tfrac{1}{2})\} - \{1/(2\nu_{\sigma^2}s_{\sigma^2}^2)\}\mu_{\mathfrak{q}(1/a_{\sigma^2})}
\end{aligned}$$

$$
\begin{aligned}
T_7 &= -\tfrac{1}{2}\xi_{\mathfrak{q}(a_{\sigma^2})}\ \log(\lambda_{\mathfrak{q}(a_{\sigma^2})}/2) + \log\{\Gamma(\tfrac{1}{2}\xi_{\mathfrak{q}(a_{\sigma^2})})\} + \tfrac{1}{2}\lambda_{\mathfrak{q}(a_{\sigma^2})}\mu_{\mathfrak{q}(1/a_{\sigma^2})},\\
T_8 &= -\tfrac{1}{2}\text{tr}(\boldsymbol{M}_{\mathfrak{q}(\boldsymbol{A}_{\boldsymbol{\Sigma}}^{-1})}\boldsymbol{M}_{\mathfrak{q}(\boldsymbol{\Sigma}^{-1})}) - \frac{q}{2}(\nu_{\boldsymbol{\Sigma}} + 2q - 1)\log(2) - \frac{q}{4}(q-1)\log(\pi)\\
&\quad -\sum_{j=1}^{q}\log\Gamma(\tfrac{1}{2}(\nu_{\boldsymbol{\Sigma}} + 2q - j)),\\
T_9 &= -\tfrac{1}{2}(\xi_{\mathfrak{q}(\boldsymbol{\Sigma})} - q + 1)\log|\boldsymbol{\Lambda}_{\mathfrak{q}(\boldsymbol{\Sigma})}| + \tfrac{1}{2}\text{tr}(\boldsymbol{\Lambda}_{\mathfrak{q}(\boldsymbol{\Sigma})}\boldsymbol{M}_{\mathfrak{q}(\boldsymbol{\Sigma}^{-1})}) + \frac{q}{2}(\xi_{\mathfrak{q}(\boldsymbol{\Sigma})} + 1)\log(2),\\
&\quad +\frac{q}{4}(q-1)\log(\pi) + \sum_{j=1}^{q}\log\Gamma(\tfrac{1}{2}(\xi_{\mathfrak{q}(\boldsymbol{\Sigma})} + 2 - j)),\\
T_{10} &= -\tfrac{1}{2}q(2-q)\log(\nu_{\boldsymbol{\Sigma}}) - \tfrac{1}{2}(2-q)\sum_{j=1}^{q}\log(s^2_{\boldsymbol{\Sigma},j}) - \tfrac{1}{2}\sum_{j=1}^{q}1/(\nu_{\boldsymbol{\Sigma}}s^2_{\boldsymbol{\Sigma},j})\left(\boldsymbol{M}_{\mathfrak{q}(\boldsymbol{A}_{\boldsymbol{\Sigma}}^{-1})}\right)_{jj}\\
&\quad -q\log(2) - \frac{q}{4}(q-1)\log(\pi) - \sum_{j=1}^{q}\log\Gamma(\tfrac{1}{2}(3-j)),\\
T_{11} &= -\tfrac{1}{2}(\xi_{\mathfrak{q}(\boldsymbol{A}_{\boldsymbol{\Sigma}})} - q + 1)\log|\boldsymbol{\Lambda}_{\mathfrak{q}(\boldsymbol{A}_{\boldsymbol{\Sigma}})}| + \tfrac{1}{2}\text{tr}(\boldsymbol{\Lambda}_{\mathfrak{q}(\boldsymbol{A}_{\boldsymbol{\Sigma}})}\boldsymbol{M}_{\mathfrak{q}(\boldsymbol{A}_{\boldsymbol{\Sigma}}^{-1})})\\
&\quad +\frac{q}{2}(\xi_{\mathfrak{q}(\boldsymbol{A}_{\boldsymbol{\Sigma}})} + 1)\log(2) + \frac{q}{4}(q-1)\log(\pi) + \sum_{j=1}^{q}\log\Gamma(\tfrac{1}{2}(\xi_{\mathfrak{q}(\boldsymbol{A}_{\boldsymbol{\Sigma}})} + 2 - j)),
\end{aligned}
$$

$$
\begin{aligned}
T_{12} &= -\tfrac{1}{2}\text{tr}(\boldsymbol{M}_{\mathfrak{q}(\boldsymbol{A}_{\boldsymbol{\Sigma}'}^{-1})}\boldsymbol{M}_{\mathfrak{q}((\boldsymbol{\Sigma}')^{-1})}) - \frac{q'}{2}(\nu_{\boldsymbol{\Sigma}'} + 2q' - 1)\log(2) - \frac{q'}{4}(q'-1)\log(\pi)\\
&\quad -\sum_{j=1}^{q'}\log\Gamma(\tfrac{1}{2}(\nu_{\boldsymbol{\Sigma}'} + 2q' - j)),\\
T_{13} &= -\tfrac{1}{2}(\xi_{\mathfrak{q}(\boldsymbol{\Sigma}')} - q' + 1)\log|\boldsymbol{\Lambda}_{\mathfrak{q}(\boldsymbol{\Sigma}')}| + \tfrac{1}{2}\text{tr}(\boldsymbol{\Lambda}_{\mathfrak{q}(\boldsymbol{\Sigma}')}\boldsymbol{M}_{\mathfrak{q}((\boldsymbol{\Sigma}')^{-1})}) + \frac{q'}{2}(\xi_{\mathfrak{q}(\boldsymbol{\Sigma}')} + 1)\log(2),\\
&\quad +\frac{q'}{4}(q'-1)\log(\pi) + \sum_{j=1}^{q'}\log\Gamma(\tfrac{1}{2}(\xi_{\mathfrak{q}(\boldsymbol{\Sigma}')} + 2 - j)),\\
T_{14} &= -\tfrac{1}{2}q'(2-q')\log(\nu_{\boldsymbol{\Sigma}'}) - \tfrac{1}{2}(2-q')\sum_{j=1}^{q'}\log(s^2_{\boldsymbol{\Sigma},j}) - \tfrac{1}{2}\sum_{j=1}^{q'}1/(\nu_{\boldsymbol{\Sigma}'}s^2_{\boldsymbol{\Sigma},j})\left(\boldsymbol{M}_{\mathfrak{q}((\boldsymbol{A}_{\boldsymbol{\Sigma}'})^{-1})}\right)_{jj}\\
&\quad -q'\log(2) - \frac{q'}{4}(q'-1)\log(\pi) - \sum_{j=1}^{q'}\log\Gamma(\tfrac{1}{2}(3-j))
\end{aligned}
$$

and

$$
\begin{aligned}
T_{15} &= -\tfrac{1}{2}(\xi_{\mathfrak{q}(\boldsymbol{A}_{\boldsymbol{\Sigma}'})} - q' + 1)\log|\boldsymbol{\Lambda}_{\mathfrak{q}(\boldsymbol{A}_{\boldsymbol{\Sigma}'})}| + \tfrac{1}{2}\text{tr}(\boldsymbol{\Lambda}_{\mathfrak{q}(\boldsymbol{A}_{\boldsymbol{\Sigma}'})}\boldsymbol{M}_{\mathfrak{q}((\boldsymbol{A}_{\boldsymbol{\Sigma}'})^{-1})})\\
&\quad +\frac{q'}{2}(\xi_{\mathfrak{q}(\boldsymbol{A}_{\boldsymbol{\Sigma}'})} + 1)\log(2) + \frac{q'}{4}(q'-1)\log(\pi) + \sum_{j=1}^{q'}\log\Gamma(\tfrac{1}{2}(\xi_{\mathfrak{q}(\boldsymbol{A}_{\boldsymbol{\Sigma}'})} + 2 - j)).
\end{aligned}
$$

Note that the component of $\log \underline{\mathfrak{p}}(\boldsymbol{y};\mathfrak{q})$ which does not get updated during the coordinate ascent iterations, except for the irreducible $\log\Gamma$ terms, and which we will call 'const' is:

$$\begin{aligned}
\text{const} \equiv & -\tfrac{1}{2}n_{\bullet\bullet}\log(2\pi) - \tfrac{1}{2}(p+mq+m'q')\log(2\pi) - \tfrac{1}{2}\log|\boldsymbol{\Sigma}_{\boldsymbol{\beta}}| + \tfrac{1}{2}(p+mq+m'q') \\
& + \tfrac{1}{2}(p+mq+m'q')\log(2\pi) - \tfrac{1}{2}\nu_{\sigma^2}\log(2) + \tfrac{1}{2}(\xi_{\mathfrak{q}(\sigma^2)})\log(2) - \tfrac{1}{2}\log(2\nu_{\sigma^2}s_{\sigma^2}^2) \\
& - \tfrac{1}{2}q(\nu_{\boldsymbol{\Sigma}}+2q-1)\log(2) - \frac{q}{2}(q-1)\log(\pi) + \tfrac{1}{2}q(\xi_{\mathfrak{q}(\boldsymbol{\Sigma})}+1)\log(2) + \frac{q}{2}(q-1)\log(\pi) \\
& - \tfrac{1}{2}q(2-q)\log(\nu_{\boldsymbol{\Sigma}}) - \tfrac{1}{2}(2-q)\sum_{j=1}^{q}\log(s_{\boldsymbol{\Sigma},j}^2) - q\log(2) + \tfrac{1}{2}q(\xi_{\mathfrak{q}(\boldsymbol{A}_{\boldsymbol{\Sigma}})}+1)\log(2) \\
& - \tfrac{1}{2}q'(\nu_{\boldsymbol{\Sigma}'}+2q'-1)\log(2) - \frac{q'}{2}(q'-1)\log(\pi) + \tfrac{1}{2}q'(\xi_{\mathfrak{q}(\boldsymbol{\Sigma}')}+1)\log(2) \\
& + \frac{q'}{2}(q'-1)\log(\pi) - \tfrac{1}{2}q'(2-q')\log(\nu_{\boldsymbol{\Sigma}'}) - \tfrac{1}{2}(2-q')\sum_{j=1}^{q'}\log(s_{\boldsymbol{\Sigma},j}^2) - q'\log(2) \\
& + \tfrac{1}{2}q'(\xi_{\mathfrak{q}(\boldsymbol{A}_{\boldsymbol{\Sigma}'})}+1)\log(2) - \log\Gamma(\tfrac{1}{2}) \\
= & -\tfrac{1}{2}\left(n_{\bullet\bullet}+1\right)\log(\pi) - \tfrac{1}{2}\log|\Sigma_{\boldsymbol{\beta}}| + \tfrac{1}{2}(p+mq+m'q') - \tfrac{1}{2}\log(\nu_{\sigma^2}) - \tfrac{1}{2}\log(s_{\sigma^2}^2) \\
& + \tfrac{1}{2}\left\{q(\nu_{\boldsymbol{\Sigma}}+q+m-1) + q'(\nu_{\boldsymbol{\Sigma}'}+q'+m'-1) - 1\right\}\log(2) \\
& - \tfrac{1}{2}q(2-q)\log(\nu_{\boldsymbol{\Sigma}}) - \tfrac{1}{2}(2-q)\sum_{j=1}^{q}\log(s_{\sigma^2}^2) - \tfrac{1}{2}q'(2-q')\log(\nu_{\boldsymbol{\Sigma}'}) - \tfrac{1}{2}(2-q')\sum_{j=1}^{q'}\log(s_{\sigma^2}^2)
\end{aligned}$$

Our final $\log \underline{\mathfrak{p}}(\boldsymbol{y};\mathfrak{q})$ expression is then

$$\begin{aligned}
\log \underline{\mathfrak{p}}(\boldsymbol{y};\mathfrak{q}) = & -\tfrac{1}{2}\left(n_{\bullet\bullet}+1\right)\log(\pi) - \tfrac{1}{2}\log|\Sigma_{\boldsymbol{\beta}}| + \tfrac{1}{2}(p+mq+m'q') - \tfrac{1}{2}\log(\nu_{\sigma^2}) - \tfrac{1}{2}\log(s_{\sigma^2}^2) \\
& + \tfrac{1}{2}\left\{q(\nu_{\boldsymbol{\Sigma}}+q+m-1) + q'(\nu_{\boldsymbol{\Sigma}'}+q'+m'-1) - 1\right\}\log(2) - \tfrac{1}{2}q(2-q)\log(\nu_{\boldsymbol{\Sigma}}) \\
& - \tfrac{1}{2}(2-q)\sum_{j=1}^{q}\log(s_{\sigma^2}^2) - \tfrac{1}{2}q'(2-q')\log(\nu_{\boldsymbol{\Sigma}'}) - \tfrac{1}{2}(2-q')\sum_{j=1}^{q'}\log(s_{\sigma^2}^2) - \log\{\Gamma(\tfrac{1}{2}\nu_{\sigma^2})\} \\
& - \tfrac{1}{2}\text{tr}\left(\boldsymbol{\Sigma}_{\boldsymbol{\beta}}^{-1}\left\{\left(\boldsymbol{\mu}_{\mathfrak{q}(\boldsymbol{\beta})}-\boldsymbol{\mu}_{\boldsymbol{\beta}}\right)\left(\boldsymbol{\mu}_{\mathfrak{q}(\boldsymbol{\beta})}-\boldsymbol{\mu}_{\boldsymbol{\beta}}\right)^T + \boldsymbol{\Sigma}_{\mathfrak{q}(\boldsymbol{\beta})}\right\}\right) + \tfrac{1}{2}\log|\boldsymbol{\Sigma}_{\mathfrak{q}(\boldsymbol{\beta},\boldsymbol{u},\boldsymbol{u}')}| \\
& - \tfrac{1}{2}\text{tr}\left(\boldsymbol{M}_{\mathfrak{q}(\boldsymbol{\Sigma}^{-1})}\left\{\sum_{i=1}^{m}\left(\boldsymbol{\mu}_{\mathfrak{q}(\boldsymbol{u}_i)}\boldsymbol{\mu}_{\mathfrak{q}(\boldsymbol{u}_i)}^T + \boldsymbol{\Sigma}_{\mathfrak{q}(\boldsymbol{u}_i)}\right)\right\}\right) - \tfrac{1}{2}\mu_{\mathfrak{q}(1/a_{\sigma^2})}\mu_{\mathfrak{q}(1/\sigma^2)} \\
& - \tfrac{1}{2}\text{tr}\left(\boldsymbol{M}_{\mathfrak{q}((\boldsymbol{\Sigma}')^{-1})}\left\{\sum_{i'=1}^{m'}\left(\boldsymbol{\mu}_{\mathfrak{q}(\boldsymbol{u}'_{i'})}\boldsymbol{\mu}_{\mathfrak{q}(\boldsymbol{u}'_{i'})}^T + \boldsymbol{\Sigma}_{\mathfrak{q}(\boldsymbol{u}'_{i'})}\right)\right\}\right) - \{1/(2\nu_{\sigma^2}s_{\sigma^2}^2)\}\mu_{\mathfrak{q}(1/a_{\sigma^2})} \\
& - \tfrac{1}{2}\xi_{\mathfrak{q}(\sigma^2)}\,\log(\lambda_{\mathfrak{q}(\sigma^2)}/2) + \log\{\Gamma(\tfrac{1}{2}\xi_{\mathfrak{q}(\sigma^2)})\} + \tfrac{1}{2}\lambda_{\mathfrak{q}(\sigma^2)}\mu_{\mathfrak{q}(1/\sigma^2)} \\
& - \tfrac{1}{2}\text{tr}(\boldsymbol{M}_{\mathfrak{q}(\boldsymbol{A}_{\boldsymbol{\Sigma}}^{-1})}\boldsymbol{M}_{\mathfrak{q}(\boldsymbol{\Sigma}^{-1})}) - \sum_{j=1}^{q}\log\Gamma(\tfrac{1}{2}(\nu_{\boldsymbol{\Sigma}}+2q-j)) - \tfrac{1}{2}\text{tr}(\boldsymbol{\Lambda}_{\mathfrak{q}(\boldsymbol{\Sigma})}\boldsymbol{M}_{\mathfrak{q}(\boldsymbol{\Sigma}^{-1})}) \\
& - \tfrac{1}{2}\text{tr}(\boldsymbol{M}_{\mathfrak{q}(\boldsymbol{A}_{\boldsymbol{\Sigma}'}^{-1})}\boldsymbol{M}_{\mathfrak{q}((\boldsymbol{\Sigma}')^{-1})}) - \sum_{j=1}^{q'}\log\Gamma(\tfrac{1}{2}(\nu_{\boldsymbol{\Sigma}'}+2q'-j)) - \tfrac{1}{2}\text{tr}(\boldsymbol{\Lambda}_{\mathfrak{q}(\boldsymbol{\Sigma}')}\boldsymbol{M}_{\mathfrak{q}((\boldsymbol{\Sigma}')^{-1})}) \\
& - \tfrac{1}{2}\xi_{\mathfrak{q}(a_{\sigma^2})}\,\log(\lambda_{\mathfrak{q}(a_{\sigma^2})}/2) + \log\{\Gamma(\tfrac{1}{2}\xi_{\mathfrak{q}(a_{\sigma^2})})\} + \tfrac{1}{2}\lambda_{\mathfrak{q}(a_{\sigma^2})}\mu_{\mathfrak{q}(1/a_{\sigma^2})} \\
& - \sum_{j=1}^{q}\log\Gamma(\tfrac{1}{2}(\xi_{\mathfrak{q}(\boldsymbol{\Sigma})}+2-j)) + \tfrac{1}{2}(\xi_{\mathfrak{q}(\boldsymbol{\Sigma})}-q+1)\log|\boldsymbol{\Lambda}_{\mathfrak{q}(\boldsymbol{\Sigma})}|
\end{aligned}$$

$$
-\sum_{j=1}^{q'} \log\Gamma(\tfrac{1}{2}(\xi_{\mathfrak{q}(\boldsymbol{\Sigma}')}+2-j)) + \tfrac{1}{2}(\xi_{\mathfrak{q}(\boldsymbol{\Sigma}')}-q'+1)\log|\boldsymbol{\Lambda}_{\mathfrak{q}(\boldsymbol{\Sigma}')}|
$$

$$
\begin{aligned}
-\tfrac{1}{2}\mu_{\mathfrak{q}(1/\sigma^2)}\sum_{i=1}^{m}\sum_{i'=1}^{m'}\Bigg\{&\Big|\Big|\boldsymbol{y}_{ii'}-\boldsymbol{X}_{ii'}\boldsymbol{\mu}_{\mathfrak{q}(\boldsymbol{\beta})}-\boldsymbol{Z}_{ii'}\boldsymbol{\mu}_{\mathfrak{q}(\boldsymbol{u}_i)}-\boldsymbol{Z}'_{ii'}\boldsymbol{\mu}_{\mathfrak{q}(\boldsymbol{u}'_{i'})}\Big|\Big|^2\\
&+\text{tr}\left(\boldsymbol{X}_{ii'}^T\boldsymbol{X}_{ii'}\boldsymbol{\Sigma}_{\mathfrak{q}(\boldsymbol{\beta})}\right)+\text{tr}\left(\boldsymbol{Z}_{ii'}^T\boldsymbol{Z}_{ii'}\boldsymbol{\Sigma}_{\mathfrak{q}(\boldsymbol{u}_i)}\right)+\text{tr}((\boldsymbol{Z}'_{ii'})^T\boldsymbol{Z}'_{ii'}\boldsymbol{\Sigma}_{\mathfrak{q}(\boldsymbol{u}'_{i'})})\\
&+2\text{tr}\left[\boldsymbol{Z}_{ii'}^T\boldsymbol{X}_{ii'}E_{\mathfrak{q}}\{(\boldsymbol{\beta}-\boldsymbol{\mu}_{\mathfrak{q}(\boldsymbol{\beta})})(\boldsymbol{u}_i-\boldsymbol{\mu}_{\mathfrak{q}(\boldsymbol{u}_i)})^T\}\right]\\
&+2\text{tr}\left[\left(\boldsymbol{Z}'_{ii'}\right)^T\boldsymbol{X}_{ii'}E_{\mathfrak{q}}\{(\boldsymbol{\beta}-\boldsymbol{\mu}_{\mathfrak{q}(\boldsymbol{\beta})})(\boldsymbol{u}'_{i'}-\boldsymbol{\mu}_{\mathfrak{q}(\boldsymbol{u}'_{i'})})^T\}\right]\\
&+2\text{tr}\left[\boldsymbol{Z}_{ii'}^T\boldsymbol{Z}'_{ii'}E_{\mathfrak{q}}\{(\boldsymbol{u}_i-\boldsymbol{\mu}_{\mathfrak{q}(\boldsymbol{u}_i)})(\boldsymbol{u}'_{i'}-\boldsymbol{\mu}_{\mathfrak{q}(\boldsymbol{u}'_{i'})})^T\}\right]\Bigg\}.
\end{aligned}
$$

The $\log \underline{\mathfrak{p}}(\boldsymbol{y};\mathfrak{q})$ expression simplifies under product restrictions I and II, since we have

$$
E_{\mathfrak{q}}\{(\boldsymbol{\beta}-\boldsymbol{\mu}_{\mathfrak{q}(\boldsymbol{\beta})})(\boldsymbol{u}'_{i'}-\boldsymbol{\mu}_{\mathfrak{q}(\boldsymbol{u}'_{i'})})^T\}=\boldsymbol{O},\quad 1\le i'\le m',
$$

and

$$
E_{\mathfrak{q}}\{(\boldsymbol{u}_i-\boldsymbol{\mu}_{\mathfrak{q}(\boldsymbol{u}_i)})(\boldsymbol{u}'_{i'}-\boldsymbol{\mu}_{\mathfrak{q}(\boldsymbol{u}'_{i'})})^T\}=\boldsymbol{O},\quad 1\le i\le m, 1\le i'\le m'.
$$

Under product restriction I we also have

$$
E_{\mathfrak{q}}\{(\boldsymbol{\beta}-\boldsymbol{\mu}_{\mathfrak{q}(\boldsymbol{\beta})})(\boldsymbol{u}_i-\boldsymbol{\mu}_{\mathfrak{q}(\boldsymbol{u}_i)})^T\}=\boldsymbol{O},\quad 1\le i\le m.
$$

From Theorem 1 of Nolan & Wand (2020), the $\log|\boldsymbol{\Sigma}_{\mathfrak{q}(\boldsymbol{\beta},\boldsymbol{u},\boldsymbol{u}')}|$ term has the following streamlined form:

$$
\begin{aligned}
\log|\boldsymbol{\Sigma}_{\mathfrak{q}(\boldsymbol{\beta},\boldsymbol{u},\boldsymbol{u}')}| = &\log\left|\boldsymbol{A}^{11}\text{ component of }\mathcal{S}\text{ from Algorithm 2}\right|\\
&-\sum_{i=1}^{m}\log\left|\mu_{\mathfrak{q}(1/\sigma^2)}\overset{\blacktriangle}{\boldsymbol{Z}}{}_i^T\overset{\blacktriangle}{\boldsymbol{Z}}_i+\boldsymbol{M}_{\mathfrak{q}(\boldsymbol{\Sigma}^{-1})}\right|,
\end{aligned}
$$

under product restriction I, and

$$
\log|\boldsymbol{\Sigma}_{\mathfrak{q}(\boldsymbol{\beta},\boldsymbol{u},\boldsymbol{u}')}| = \log\left|\boldsymbol{\Sigma}_{\mathfrak{q}(\boldsymbol{\beta})}\right| + \sum_{i=1}^{m}\log\left|\boldsymbol{\Sigma}_{\mathfrak{q}(\boldsymbol{u}_i)}\right| - \sum_{i=1}^{m}\log\left|\mu_{\mathfrak{q}(1/\sigma^2)}\overset{\blacktriangle}{\boldsymbol{Z}}{}_i^T\overset{\blacktriangle}{\boldsymbol{Z}}_i+\boldsymbol{M}_{\mathfrak{q}(\boldsymbol{\Sigma}^{-1})}\right|
$$

under product restrictions II and III.

S.9 Streamlined Computing for Frequentist Inference

As an aside we point out that the approach used by Algorithm 8 for product restriction III, in which the $\mathfrak{q}$-density updates for the $(\boldsymbol{\beta},\boldsymbol{u}_{\text{all}})$ parameters are embedded within the SolveTwoLevelSparseLeastSquares infrastructure, can also be used for streamlined *frequentist* inference when m' is moderate in size. To the best of our knowledge, the results given here for efficient computation of the important sub-blocks of the relevant covariance matrix are novel.

The frequentist Gaussian response two-level linear mixed model with crossed random effects is

$$
\begin{aligned}
&\boldsymbol{y}_{ii'}|\boldsymbol{\beta},\boldsymbol{u}_i,\boldsymbol{u}'_{i'}\overset{\text{ind.}}{\sim}N(\boldsymbol{X}_{ii'}\boldsymbol{\beta}+\boldsymbol{Z}_{ii'}\,\boldsymbol{u}_i+\boldsymbol{Z}'_{ii'}\,\boldsymbol{u}'_{i'},\sigma^2\,\boldsymbol{I}),\\
&\boldsymbol{u}_i\overset{\text{ind.}}{\sim}N(\boldsymbol{0},\boldsymbol{\Sigma}),\ \boldsymbol{u}'_{i'}\overset{\text{ind.}}{\sim}N(\boldsymbol{0},\boldsymbol{\Sigma}'),\ 1\le i\le m,\ 1\le i\le m'.
\end{aligned}
\tag{S.3}
$$

The best linear unbiased predictor of $[\boldsymbol{\beta}^T\ \boldsymbol{u}^T]^T$ and its corresponding covariance matrix are

$$\begin{bmatrix}\widehat{\boldsymbol{\beta}}\\ \widehat{\boldsymbol{u}}\end{bmatrix} = (\boldsymbol{C}^T\boldsymbol{R}_{\text{BLUP}}^{-1}\boldsymbol{C} + \boldsymbol{D}_{\text{BLUP}})^{-1}\boldsymbol{C}^T\boldsymbol{R}_{\text{BLUP}}^{-1}\boldsymbol{y}$$
$$\text{and}\quad \text{Cov}\left(\begin{bmatrix}\widehat{\boldsymbol{\beta}}\\ \widehat{\boldsymbol{u}}-\boldsymbol{u}\end{bmatrix}\right) = (\boldsymbol{C}^T\boldsymbol{R}_{\text{BLUP}}^{-1}\boldsymbol{C} + \boldsymbol{D}_{\text{BLUP}})^{-1} \tag{S.4}$$

where $\boldsymbol{C} \equiv [\boldsymbol{X}\ \boldsymbol{Z}]$, with $\boldsymbol{X}$ and $\boldsymbol{Z}$ as defined in Section 2.1.

$$\boldsymbol{D}_{\text{BLUP}} \equiv \begin{bmatrix}\boldsymbol{O} & \boldsymbol{O}\\ \boldsymbol{O} & \begin{bmatrix}\boldsymbol{I}_m\otimes\boldsymbol{\Sigma}^{-1} & \boldsymbol{O}\\ \boldsymbol{O} & \boldsymbol{I}_{m'}\otimes(\boldsymbol{\Sigma}')^{-1}\end{bmatrix}\end{bmatrix} \quad\text{and}\quad \boldsymbol{R}_{\text{BLUP}} \equiv \sigma^2\boldsymbol{I}.$$

Note that the following sub-blocks are required for adding pointwise confidence intervals to mean estimates:

$$\text{Cov}(\widehat{\boldsymbol{\beta}}),\quad \text{Cov}(\widehat{\boldsymbol{u}}_i-\boldsymbol{u}_i),\quad \text{Cov}(\widehat{\boldsymbol{u}}'_{i'}-\boldsymbol{u}'_{i'}),$$
$$E\{\widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{u}}_i-\boldsymbol{u}_i)^T\},\quad E\{\widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{u}}'_{i'}-\boldsymbol{u}'_{i'})^T\}\quad\text{and}\quad E\{(\widehat{\boldsymbol{u}}_i-\boldsymbol{u}_i)(\widehat{\boldsymbol{u}}'_{i'}-\boldsymbol{u}'_{i'})^T\} \tag{S.5}$$

for $1\le i\le m$ and $1\le i'\le m'$.

Result S.1. *Computation of* $[\widehat{\boldsymbol{\beta}}^T\ \ \widehat{\boldsymbol{u}}^T]^T$ *and each of the sub-blocks of* $Cov([\widehat{\boldsymbol{\beta}}\ \ \widehat{\boldsymbol{u}}-\boldsymbol{u}]^T)$ *listed in (S.5) are expressible as the two-level sparse matrix least squares form:*

$$\left\|\boldsymbol{b}-\boldsymbol{B}\begin{bmatrix}\boldsymbol{\beta}\\ \boldsymbol{u}\end{bmatrix}\right\|^2$$

where $\boldsymbol{b}$ *and the non-zero sub-blocks of* $\boldsymbol{B}$*, according to the notation in (S.1), are, for* $1\le i\le m$*,*

$$\boldsymbol{b}_i \equiv \begin{bmatrix}\sigma^{-1}\boldsymbol{y}_i\\ \boldsymbol{0}\\ \boldsymbol{0}\end{bmatrix},\quad \boldsymbol{B}_i \equiv \begin{bmatrix}\sigma^{-1}\boldsymbol{X}_i & \sigma^{-1}\boldsymbol{Z}'_i\\ \boldsymbol{O} & m^{-1/2}(\boldsymbol{I}_{m'}\otimes(\boldsymbol{\Sigma}')^{-1/2})\\ \boldsymbol{O} & \boldsymbol{O}\end{bmatrix}\quad \textit{and}\quad \overset{\bullet}{\boldsymbol{B}}_i \equiv \begin{bmatrix}\sigma^{-1}\boldsymbol{Z}_i\\ \boldsymbol{O}\\ \boldsymbol{\Sigma}^{-1/2}\end{bmatrix}.$$

Each of these matrices has $m'(n_{ii'}+q')+q$ *rows. The* $\boldsymbol{B}_i$ *matrices each have* $p+m'q'$ *columns and the* $\overset{\bullet}{\boldsymbol{B}}_i$ *each have* q *columns. The solutions are*

$$\widehat{\boldsymbol{\beta}} = \textit{first } p \textit{ rows of } \boldsymbol{x}_1,\ \ \text{Cov}(\widehat{\boldsymbol{\beta}}) = \textit{top left } p\times p \textit{ sub-block of } \boldsymbol{A}^{11},$$

$$\underset{1\le i'\le m'}{\text{stack}}\left(\widehat{\boldsymbol{u}}'_{i'}\right) = \textit{subsequent } (m'q')\times 1 \textit{ entries of } \boldsymbol{x}_1 \textit{ following } \widehat{\boldsymbol{\beta}},$$

$$E\{\widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{u}}'_{i'}-\boldsymbol{u}'_{i'})^T\} = \textit{subsequent } p\times q' \textit{ sub-blocks of } \boldsymbol{A}^{11} \textit{ to the right of } \text{Cov}(\widehat{\boldsymbol{\beta}}),$$

$$\text{Cov}(\widehat{\boldsymbol{u}}'_{i'}-\boldsymbol{u}'_{i'}) = \textit{subsequent } q'\times q' \textit{ diagonal sub-blocks of } \boldsymbol{A}^{11} \textit{ following } \text{Cov}(\widehat{\boldsymbol{\beta}}),\ \ 1\le i'\le m',$$

$$\widehat{\boldsymbol{u}}_i = \boldsymbol{x}_{2,i},\ \ \text{Cov}(\widehat{\boldsymbol{u}}_i-\boldsymbol{u}_i) = \boldsymbol{A}^{22,i},\ \ E\{\widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{u}}_i-\boldsymbol{u}_i)^T\} = \textit{first } p \textit{ rows of } \boldsymbol{A}^{12,i}$$

$$\underset{1\le i'\le m'}{\text{stack}}\left[E\{(\widehat{\boldsymbol{u}}_i-\boldsymbol{u}_i)(\widehat{\boldsymbol{u}}'_{i'}-\boldsymbol{u}'_{i'})^T\}\right] = \textit{remaining } m'q' \textit{ rows of } \boldsymbol{A}^{12,i},\ \ 1\le i\le m,$$

where the $\boldsymbol{x}_1$*,* $\boldsymbol{x}_{2,i}$*,* $\boldsymbol{A}^{11}$*,* $\boldsymbol{A}^{22,i}$ *and* $\boldsymbol{A}^{12,i}$ *notation is given by (S.2).*

Algorithm S.4 proceduralizes Result S.1 to facilitate computation of best linear unbiased predictors for the fixed and random effects parameters in (S.3) for fixed values of the covariance parameters. In practice, the covariance parameters would need to be replaced by estimates obtained using an approach such as restricted maximum likelihood. Algorithm S.4 also delivers the matrices in (S.5). In the case where m' is moderate but m is potentially very large Algorithm S.4 performs efficient streamlined computing.

Algorithm S.4 *Streamlined algorithm for obtaining best linear unbiased predictions and corresponding covariance matrix components for the linear mixed model with crossed random effects.*

Data Inputs: $\left\{\left(\overset{\blacktriangle}{\boldsymbol{y}}_i, \overset{\blacktriangle}{\boldsymbol{X}}_i, \overset{\blacktriangle}{\boldsymbol{Z}}_i, \overset{\blacksquare}{\boldsymbol{Z}}'_i\right):\ 1 \le i \le m\right\}$

Covariance Matrix Inputs: $\sigma^2 > 0,\ \ \boldsymbol{\Sigma}'(q' \times q'),\ \ \boldsymbol{\Sigma}(q \times q),$ symmetric and positive definite.

For $i = 1, \ldots, m$:

$$\boldsymbol{b}_i \longleftarrow \begin{bmatrix} \sigma^{-1}\boldsymbol{y}_i \\ \boldsymbol{0} \\ \boldsymbol{0} \end{bmatrix},\ \boldsymbol{B}_i \longleftarrow \begin{bmatrix} \sigma^{-1}\boldsymbol{X}_i & \sigma^{-1}\boldsymbol{Z}'_i \\ \boldsymbol{O} & m^{-1/2}(\boldsymbol{I}_{m'} \otimes (\boldsymbol{\Sigma}')^{-1/2}) \\ \boldsymbol{O} & \boldsymbol{O} \end{bmatrix},$$

$$\overset{\bullet}{\boldsymbol{B}}_i \longleftarrow \begin{bmatrix} \sigma^{-1}\boldsymbol{Z}_i \\ \boldsymbol{O} \\ \boldsymbol{\Sigma}^{-1/2} \end{bmatrix}$$

$\mathcal{S} \longleftarrow$ SolveTwoLevelSparseLeastSquares$\Big(\big\{(\boldsymbol{b}_i, \boldsymbol{B}_i, \overset{\bullet}{\boldsymbol{B}}_i) : 1 \le i \le m\big\}\Big)$

$\widehat{\boldsymbol{\beta}} \longleftarrow$ first p-rows of $\boldsymbol{x}_1$ component of $\mathcal{S}$

Cov$(\widehat{\boldsymbol{\beta}}) \longleftarrow$ top left $p \times p$ sub-block of $\boldsymbol{A}^{11}$ component of $\mathcal{S}$

$i_{\text{stt}} \longleftarrow p + 1$

For $i' = 1, \ldots, m'$:

- $i_{\text{end}} \longleftarrow i_{\text{stt}} + q' - 1$
- $\widehat{\boldsymbol{u}}_{i'} \longleftarrow$ sub-vector of $\boldsymbol{x}_1$ component of $\mathcal{S}$ with entries i_{stt} to i_{end}
- Cov$(\widehat{\boldsymbol{u}}'_{i'} - \boldsymbol{u}'_{i'}) \longleftarrow$ diagonal sub-block of $\boldsymbol{A}^{11}$ component of $\mathcal{S}$ with rows i_{stt} to i_{end} and columns i_{stt} to i_{end}
- $E\{\widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{u}}'_{i'} - \boldsymbol{u}'_{i'})^T\} \longleftarrow$ sub-block of $\boldsymbol{A}^{11}$ component of $\mathcal{S}$ with rows 1 to p and columns i_{stt} to i_{end}
- $i_{\text{stt}} \longleftarrow i_{\text{end}} + 1$

For $i = 1, \ldots, m$:

- $\widehat{\boldsymbol{u}}_i \longleftarrow \boldsymbol{x}_{2,i}$ component of $\mathcal{S}$; Cov$(\widehat{\boldsymbol{u}}_i - \boldsymbol{u}_i) \longleftarrow \boldsymbol{A}^{22,i}$ component of $\mathcal{S}$
- $E\{\widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{u}}_i - \boldsymbol{u}_i)^T\} \longleftarrow$ sub-matrix of $\boldsymbol{A}^{12,i}$ component of $\mathcal{S}$ with rows 1 to p
- $i_{\text{stt}} \longleftarrow p + 1$
- For $i' = 1, \ldots, m'$:
  - $i_{\text{end}} \longleftarrow i_{\text{stt}} + q' - 1$; $\boldsymbol{\Omega} \longleftarrow \boldsymbol{A}^{12,i}$ component of $\mathcal{S}$
  - $E\{(\widehat{\boldsymbol{u}}_i - \boldsymbol{u}_i)(\widehat{\boldsymbol{u}}'_{i'} - \boldsymbol{u}'_{i'})^T\} \longleftarrow$ sub-matrix of $\boldsymbol{\Omega}^T$ with columns i_{stt} to i_{end}
  - $i_{\text{stt}} \longleftarrow i_{\text{end}} + 1$

Outputs: $\widehat{\boldsymbol{\beta}}$, Cov$(\widehat{\boldsymbol{\beta}})$, $\Big\{(\widehat{\boldsymbol{u}}'_{i'},\ E\{\widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{u}}'_{i'} - \boldsymbol{u}'_{i'})^T\},\ \text{Cov}(\widehat{\boldsymbol{u}}'_{i'} - \boldsymbol{u}'_{i'})) :\ 1 \le i' \le m',$

$\Big(\widehat{\boldsymbol{u}}_i, E\{\widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{u}}_i - \boldsymbol{u}_i)^T\}, E\{(\widehat{\boldsymbol{u}}_i - \boldsymbol{u}_i)(\widehat{\boldsymbol{u}}'_{i'} - \boldsymbol{u}'_{i'})^T\}, \text{Cov}(\widehat{\boldsymbol{u}}_i - \boldsymbol{u}_i)\Big) : 1 \le i' \le m',$

$1 \le i \le m\}$

S.10 Full List of Items in the National Education Longitudinal Study

Table S.2 lists each of the 24 items within the National Education Longitudinal Study data set used in Section 6. Several of the measurements involve item response theory, which is abbreviated as IRT.

item	description
1	reading IRT-estimated number right
2	mathematics IRT-estimated number right
3	science IRT-estimated number right
4	history/citizenship/geography IRT-estimated number right
5	reading standardized score
6	mathematics standardized score
7	science standardized score
8	history/citizenship/geography standardized score
9	reading IRT estimate of ability
10	mathematics IRT estimate of ability
11	science IRT estimate of ability
12	history/citizenship/geography IRT estimate of ability
13	standardized test composite (reading, mathematics)
14	reading level 1: probability of proficiency
15	reading level 2: probability of proficiency
16	reading level 3: probability of proficiency
17	mathematics level 1: probability of proficiency
18	mathematics level 2: probability of proficiency
19	mathematics level 3: probability of proficiency
20	mathematics level 4: probability of proficiency
21	science level 1: probability of proficiency
22	science level 2: probability of proficiency
23	science level 3: probability of proficiency
24	science level 4: probability of proficiency

Table S.2: *Descriptions of each of the 24 items in the National Education Longitudinal Study data used in Section 6. The abbreviation IRT stands for item response theory. Fuller details are provided by Thurgood et al. (2003).*

Reference

Atay-Kayis, A. & Massam, H. (2005). A Monte Carlo method for computing marginal likelihood in nondecomposable Gaussian graphical models. *Biometrika*, **92**, 317–335.