Web-Supplement for:

Streamlined Variational Inference for Higher Level Group-Specific Curve Models

BY M. MENICTAS[1], T.H. NOLAN[1], D.G. SIMPSON[2] AND M.P. WAND[1]

University of Technology Sydney[1] *and University of Illinois*[2]

S.1 Derivation of Result 1

Straightforward algebra can be used to verify that

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

where $\boldsymbol{B}$ and $\boldsymbol{b}$ have sparse forms (2.9) with non-zero sub-blocks equal to

$$\boldsymbol{b}_i\equiv\begin{bmatrix}\sigma_\varepsilon^{-1}\boldsymbol{y}_i\\ \boldsymbol{0}\\ \boldsymbol{0}\\ \boldsymbol{0}\end{bmatrix},\quad \boldsymbol{B}_i\equiv\begin{bmatrix}\sigma_\varepsilon^{-1}\boldsymbol{X}_i & \sigma_\varepsilon^{-1}\boldsymbol{Z}_{\text{gbl},i}\\ \boldsymbol{O} & m^{-1/2}\sigma_{\text{gbl}}^{-1}\boldsymbol{I}_{K_{\text{gbl}}}\\ \boldsymbol{O} & \boldsymbol{O}\\ \boldsymbol{O} & \boldsymbol{O}\end{bmatrix}\quad\text{and}\quad \dot{\boldsymbol{B}}_i\equiv\begin{bmatrix}\sigma_\varepsilon^{-1}\boldsymbol{X}_i & \sigma_\varepsilon^{-1}\boldsymbol{Z}_{\text{grp},i}\\ \boldsymbol{O} & \boldsymbol{O}\\ \boldsymbol{\Sigma}^{-1/2} & \boldsymbol{O}\\ \boldsymbol{O} & \sigma_{\text{grp}}^{-1}\boldsymbol{I}_{K_{\text{grp}}}\end{bmatrix}.$$

Therefore, in view of (2.6),

$$\begin{bmatrix}\widehat{\boldsymbol{\beta}}\\ \widehat{\boldsymbol{u}}\end{bmatrix}=(\boldsymbol{B}^T\boldsymbol{B})^{-1}\boldsymbol{B}^T\boldsymbol{b}\quad\text{and}\quad\text{Cov}\left(\begin{bmatrix}\widehat{\boldsymbol{\beta}}\\ \widehat{\boldsymbol{u}}-\boldsymbol{u}\end{bmatrix}\right)=(\boldsymbol{B}^T\boldsymbol{B})^{-1}.$$

The sub-blocks of

$$\text{Cov}\left(\begin{bmatrix}\widehat{\boldsymbol{\beta}}\\ \widehat{\boldsymbol{u}}-\boldsymbol{u}\end{bmatrix}\right)$$

listed in (2.7) correspond to the non-$\times$ sub-blocks of $\boldsymbol{A}^{-1}=(\boldsymbol{B}^T\boldsymbol{B})^{-1}$ where $\boldsymbol{A}^{-1}$ is given by (2.10). The result follows immediately.

S.2 Derivation of Algorithm 1

Algorithm 1 is simply a proceduralization of Result 1.

S.3 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 message passing 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.4 Derivation of Result 2

It is straightforward to verify that the $\boldsymbol{\mu}_{\mathfrak{q}(\boldsymbol{\beta},\boldsymbol{u})}$ and $\boldsymbol{\Sigma}_{\mathfrak{q}(\boldsymbol{\beta},\boldsymbol{u})}$ updates, given at (2.12), may be written as

$$\boldsymbol{\mu}_{\mathfrak{q}(\boldsymbol{\beta},\boldsymbol{u})} \longleftarrow (\boldsymbol{B}^T\boldsymbol{B})^{-1}\boldsymbol{B}^T\boldsymbol{b} \quad \text{and} \quad \boldsymbol{\Sigma}_{\mathfrak{q}(\boldsymbol{\beta},\boldsymbol{u})} \longleftarrow (\boldsymbol{B}^T\boldsymbol{B})^{-1}$$

where $\boldsymbol{B}$ and $\boldsymbol{b}$ have the forms (2.9) with

$$\boldsymbol{b}_i \equiv \begin{bmatrix} \mu_{\mathfrak{q}(1/\sigma_\varepsilon^2)}^{1/2}\boldsymbol{y}_i \\ m^{-1/2}\boldsymbol{\Sigma}_{\boldsymbol{\beta}}^{-1/2}\boldsymbol{\mu}_{\boldsymbol{\beta}} \\ \boldsymbol{0} \\ \boldsymbol{0} \\ \boldsymbol{0} \end{bmatrix}, \quad \boldsymbol{B}_i \equiv \begin{bmatrix} \mu_{\mathfrak{q}(1/\sigma_\varepsilon^2)}^{1/2}\boldsymbol{X}_i & \mu_{\mathfrak{q}(1/\sigma_\varepsilon^2)}^{1/2}\boldsymbol{Z}_{\text{gbl},i} \\ m^{-1/2}\boldsymbol{\Sigma}_{\boldsymbol{\beta}}^{-1/2} & \boldsymbol{O} \\ \boldsymbol{O} & m^{-1/2}\mu_{\mathfrak{q}(1/\sigma_{\text{gbl}}^2)}^{1/2}\boldsymbol{I}_{K_{\text{gbl}}} \\ \boldsymbol{O} & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{O} \end{bmatrix}$$

and

$$\overset{\bullet}{\boldsymbol{B}}_i \equiv \begin{bmatrix} \mu_{\mathfrak{q}(1/\sigma_\varepsilon^2)}^{1/2}\boldsymbol{X}_i & \mu_{\mathfrak{q}(1/\sigma_\varepsilon^2)}^{1/2}\boldsymbol{Z}_{\text{grp},i} \\ \boldsymbol{O} & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{O} \\ \boldsymbol{M}_{\mathfrak{q}(\boldsymbol{\Sigma}^{-1})}^{1/2} & \boldsymbol{O} \\ \boldsymbol{O} & \mu_{\mathfrak{q}(1/\sigma_{\text{grp}}^2)}^{1/2}\boldsymbol{I}_{K_{\text{grp}}} \end{bmatrix}.$$

Result 2 immediately follows from Theorem 2 of Nolan & Wand (2018).

S.5 Derivation of Algorithm 2

We provide expressions for the q-densities for mean field variational Bayesian inference for the parameters in (2.10), with product density restriction (2.11). Arguments analogous to those given in, for example, Appendix C of Wand & Ormerod (2011) lead to:

$$\mathfrak{q}(\boldsymbol{\beta},\boldsymbol{u}) \text{ is a } N(\boldsymbol{\mu}_{\mathfrak{q}(\boldsymbol{\beta},\boldsymbol{u})},\boldsymbol{\Sigma}_{\mathfrak{q}(\boldsymbol{\beta},\boldsymbol{u})}) \text{ density function}$$

where

$$\boldsymbol{\Sigma}_{\mathfrak{q}(\boldsymbol{\beta},\boldsymbol{u})}=(\boldsymbol{C}^T\boldsymbol{R}_{\text{MFVB}}^{-1}\boldsymbol{C}+\boldsymbol{D}_{\text{MFVB}})^{-1}\quad\text{and}\quad\boldsymbol{\mu}_{\mathfrak{q}(\boldsymbol{\beta},\boldsymbol{u})}=\boldsymbol{\Sigma}_{\mathfrak{q}(\boldsymbol{\beta},\boldsymbol{u})}(\boldsymbol{C}^T\boldsymbol{R}_{\text{MFVB}}^{-1}\boldsymbol{y}+\boldsymbol{o}_{\text{MFVB}})$$

with $\boldsymbol{R}_{\text{MFVB}}$, $\boldsymbol{D}_{\text{MFVB}}$ and $\boldsymbol{o}_{\text{MFVB}}$ defined via (2.13),

$$\mathfrak{q}(\sigma_\varepsilon^2) \text{ is an Inverse-}\chi^2\left(\xi_{\mathfrak{q}(\sigma_\varepsilon^2)},\lambda_{\mathfrak{q}(\sigma_\varepsilon^2)}\right) \text{ density function}$$

where $\xi_{\mathfrak{q}(\sigma_\varepsilon^2)}=\nu_\varepsilon+\sum_{i=1}^m n_i$ and

$$\begin{aligned}
\lambda_{\mathfrak{q}(\sigma_\varepsilon^2)} &= \mu_{\mathfrak{q}(1/a_\varepsilon)}+\sum_{i=1}^m E_{\mathfrak{q}}\left\{\left\|\boldsymbol{y}_i-\boldsymbol{C}_{\text{gbl},i}\begin{bmatrix}\boldsymbol{\beta}\\ \boldsymbol{u}_{\text{gbl}}\end{bmatrix}-\boldsymbol{C}_{\text{grp},i}\begin{bmatrix}\boldsymbol{u}_{\text{lin},i}\\ \boldsymbol{u}_{\text{grp},i}\end{bmatrix}\right\|^2\right\}\\
&= \mu_{\mathfrak{q}(1/a_\varepsilon)}+\sum_{i=1}^m\left[\left\|E_{\mathfrak{q}}\left(\boldsymbol{y}_i-\boldsymbol{C}_{\text{gbl},i}\begin{bmatrix}\boldsymbol{\beta}\\ \boldsymbol{u}_{\text{gbl}}\end{bmatrix}-\boldsymbol{C}_{\text{grp},i}\begin{bmatrix}\boldsymbol{u}_{\text{lin},i}\\ \boldsymbol{u}_{\text{grp},i}\end{bmatrix}\right)\right\|^2\right.\\
&\qquad\left.+\text{tr}\left\{\text{Cov}_{\mathfrak{q}}\left(\boldsymbol{C}_{\text{gbl},i}\begin{bmatrix}\boldsymbol{\beta}\\ \boldsymbol{u}_{\text{gbl},i}\end{bmatrix}+\boldsymbol{C}_{\text{grp},i}\begin{bmatrix}\boldsymbol{u}_{\text{lin},i}\\ \boldsymbol{u}_{\text{grp},i}\end{bmatrix}\right)\right\}\right]\\
&= \mu_{\mathfrak{q}(1/a_\varepsilon)}+\sum_{i=1}^m\left\{\left\|E_{\mathfrak{q}}\left(\boldsymbol{y}_i-\boldsymbol{C}_{\text{gbl},i}\begin{bmatrix}\boldsymbol{\beta}\\ \boldsymbol{u}_{\text{gbl}}\end{bmatrix}-\boldsymbol{C}_{\text{grp},i}\begin{bmatrix}\boldsymbol{u}_{\text{lin},i}\\ \boldsymbol{u}_{\text{grp},i}\end{bmatrix}\right)\right\|^2\right.\\
&\qquad+\text{tr}(\boldsymbol{C}_{\text{gbl},i}^T\boldsymbol{C}_{\text{gbl},i}\boldsymbol{\Sigma}_{\mathfrak{q}(\boldsymbol{\beta},\boldsymbol{u}_{\text{gbl}})})+\text{tr}(\boldsymbol{C}_{\text{grp},i}^T\boldsymbol{C}_{\text{grp},i}\boldsymbol{\Sigma}_{\mathfrak{q}(\boldsymbol{u}_{\text{lin},i},\boldsymbol{u}_{\text{grp},i})})\\
&\qquad+2\,\text{tr}\left[\boldsymbol{C}_{\text{grp},i}^T\boldsymbol{C}_{\text{gbl},i}\,E_{\mathfrak{q}}\left\{\left(\begin{bmatrix}\boldsymbol{\beta}\\ \boldsymbol{u}_{\text{gbl}}\end{bmatrix}-\boldsymbol{\mu}_{\mathfrak{q}(\boldsymbol{\beta},\boldsymbol{u}_{\text{gbl}})}\right)\times\right.\right.\\
&\qquad\left.\left.\left.\left(\begin{bmatrix}\boldsymbol{u}_{\text{lin},i}\\ \boldsymbol{u}_{\text{grp},i}\end{bmatrix}-\boldsymbol{\mu}_{\mathfrak{q}(\boldsymbol{u}_{\text{lin},i},\boldsymbol{u}_{\text{grp},i})}\right)^T\right\}\right]\right\}
\end{aligned}$$

where $\boldsymbol{C}_{\text{gbl},i}\equiv[\boldsymbol{X}_i\,\boldsymbol{Z}_{\text{gbl},i}]$, $\boldsymbol{C}_{\text{grp},i}\equiv[\boldsymbol{X}_i\,\boldsymbol{Z}_{\text{grp},i}]$, and with reciprocal moment $\mu_{\mathfrak{q}(1/\sigma_\varepsilon^2)}=\xi_{\mathfrak{q}(\sigma_\varepsilon^2)}/\lambda_{\mathfrak{q}(\sigma_\varepsilon^2)}$,

$$\mathfrak{q}(\sigma_{\text{gbl}}^2) \text{ is an Inverse-}\chi^2\left(\xi_{\mathfrak{q}(\sigma_{\text{gbl}}^2)},\lambda_{\mathfrak{q}(\sigma_{\text{gbl}}^2)}\right) \text{ density function}$$

where $\xi_{\mathfrak{q}(\sigma_{\text{gbl}}^2)}=\nu_{\text{gbl}}+K_{\text{gbl}}$ and

$$\lambda_{\mathfrak{q}(\sigma_{\text{gbl}}^2)}=\mu_{\mathfrak{q}(1/a_{\text{gbl}})}+\|\boldsymbol{\mu}_{\mathfrak{q}(\boldsymbol{u}_{\text{gbl}})}\|^2+\text{tr}\left(\boldsymbol{\Sigma}_{\mathfrak{q}(\boldsymbol{u}_{\text{gbl}})}\right),$$

with reciprocal moment $\mu_{\mathfrak{q}(1/\sigma_{\text{gbl}}^2)}=\xi_{\mathfrak{q}(\sigma_{\text{gbl}}^2)}/\lambda_{\mathfrak{q}(\sigma_{\text{gbl}}^2)}$,

$$\mathfrak{q}(\sigma_{\text{grp}}^2) \text{ is an Inverse-}\chi^2\left(\xi_{\mathfrak{q}(\sigma_{\text{grp}}^2)},\lambda_{\mathfrak{q}(\sigma_{\text{grp}}^2)}\right) \text{ density function}$$

where $\xi_{\mathfrak{q}(\sigma_{\text{grp}}^2)}=\nu_{\text{grp}}+mK_{\text{grp}}$ and

$$\lambda_{\mathfrak{q}(\sigma_{\text{grp}}^2)}=\mu_{\mathfrak{q}(1/a_{\text{grp}})}+\sum_{i=1}^m\left\{\|\boldsymbol{\mu}_{\mathfrak{q}(\boldsymbol{u}_{\text{grp},i})}\|^2+\text{tr}\left(\boldsymbol{\Sigma}_{\mathfrak{q}(\boldsymbol{u}_{\text{grp},i})}\right)\right\},$$

with reciprocal moment $\mu_{\mathfrak{q}(1/\sigma^2_{\text{grp}})} = \xi_{\mathfrak{q}(\sigma^2_{\text{grp}})}/\lambda_{\mathfrak{q}(\sigma^2_{\text{grp}})}$,

$$\mathfrak{q}(\boldsymbol{\Sigma}) \text{ is an Inverse-G-Wishart } \big(G_{\text{full}}, \xi_{\mathfrak{q}(\boldsymbol{\Sigma})}, \boldsymbol{\Lambda}_{\mathfrak{q}(\boldsymbol{\Sigma})}\big) \text{ density function}$$

where $\xi_{\mathfrak{q}(\boldsymbol{\Sigma})} = \nu_{\boldsymbol{\Sigma}} + 2 + m$

$$\boldsymbol{\Lambda}_{\mathfrak{q}(\boldsymbol{\Sigma})} = \boldsymbol{M}_{\mathfrak{q}(\boldsymbol{A}_{\boldsymbol{\Sigma}}^{-1})} + \sum_{i=1}^{m} \Big(\boldsymbol{\mu}_{\mathfrak{q}(\boldsymbol{u}_{\text{lin},i})} \boldsymbol{\mu}_{\mathfrak{q}(\boldsymbol{u}_{\text{lin},i})}^T + \boldsymbol{\Sigma}_{\mathfrak{q}(\boldsymbol{u}_{\text{lin},i})} \Big),$$

with inverse moment $\boldsymbol{M}_{\mathfrak{q}(\boldsymbol{\Sigma}^{-1})} = (\xi_{\mathfrak{q}(\boldsymbol{\Sigma})} - 1)\boldsymbol{\Lambda}^{-1}_{\mathfrak{q}(\boldsymbol{\Sigma})}$,

$$\mathfrak{q}(a_\varepsilon) \text{ is an Inverse-}\chi^2(\xi_{\mathfrak{q}(a_\varepsilon)}, \lambda_{\mathfrak{q}(a_\varepsilon)}) \text{ density function}$$

where $\xi_{\mathfrak{q}(a_\varepsilon)} = \nu_\varepsilon + 1$,

$$\lambda_{\mathfrak{q}(a_\varepsilon)} = \mu_{\mathfrak{q}(1/\sigma^2_\varepsilon)} + 1/(\nu_\varepsilon s^2_\varepsilon)$$

with reciprocal moment $\mu_{\mathfrak{q}(1/a_\varepsilon)} = \xi_{\mathfrak{q}(a_\varepsilon)}/\lambda_{\mathfrak{q}(a_\varepsilon)}$,

$$\mathfrak{q}(a_{\text{gbl}}) \text{ is an Inverse-}\chi^2(\xi_{\mathfrak{q}(a_{\text{gbl}})}, \lambda_{\mathfrak{q}(a_{\text{gbl}})}) \text{ density function}$$

where $\xi_{\mathfrak{q}(a_{\text{gbl}})} = \nu_{\text{gbl}} + 1$,

$$\lambda_{\mathfrak{q}(a_{\text{gbl}})} = \mu_{\mathfrak{q}(1/\sigma^2_{\text{gbl}})} + 1/(\nu_{\text{gbl}} s^2_{\text{gbl}})$$

with reciprocal moment $\mu_{\mathfrak{q}(1/a_{\text{gbl}})} = \xi_{\mathfrak{q}(a_{\text{gbl}})}/\lambda_{\mathfrak{q}(a_{\text{gbl}})}$,

$$\mathfrak{q}(a_{\text{grp}}) \text{ is an Inverse-}\chi^2(\xi_{\mathfrak{q}(a_{\text{grp}})}, \lambda_{\mathfrak{q}(a_{\text{grp}})}) \text{ density function}$$

where $\xi_{\mathfrak{q}(a_{\text{grp}})} = \nu_{\text{grp}} + 1$,

$$\lambda_{\mathfrak{q}(a_{\text{grp}})} = \mu_{\mathfrak{q}(1/\sigma^2_{\text{grp}})} + 1/(\nu_{\text{grp}} s^2_{\text{grp}})$$

with reciprocal moment $\mu_{\mathfrak{q}(1/a_{\text{grp}})} = \xi_{\mathfrak{q}(a_{\text{grp}})}/\lambda_{\mathfrak{q}(a_{\text{grp}})}$ and

$$\mathfrak{q}(\boldsymbol{A}_{\boldsymbol{\Sigma}}) \text{ is an Inverse-G-Wishart } \big(G_{\text{diag}}, \xi_{\mathfrak{q}(\boldsymbol{A}_{\boldsymbol{\Sigma}})}, \boldsymbol{\Lambda}_{\mathfrak{q}(\boldsymbol{A}_{\boldsymbol{\Sigma}})}\big) \text{ density function}$$

where $\xi_{\mathfrak{q}(\boldsymbol{A}_{\boldsymbol{\Sigma}})} = \nu_{\boldsymbol{\Sigma}} + 2$,

$$\boldsymbol{\Lambda}_{\mathfrak{q}(\boldsymbol{A}_{\boldsymbol{\Sigma}})} = \text{diag}\big\{\text{diagonal}\big(\boldsymbol{M}_{\mathfrak{q}(\boldsymbol{\Sigma}^{-1})}\big)\big\} + \boldsymbol{\Lambda}_{\boldsymbol{A}_{\boldsymbol{\Sigma}}}$$

with inverse moment $\boldsymbol{M}_{\mathfrak{q}(\boldsymbol{A}_{\boldsymbol{\Sigma}}^{-1})} = \xi_{\mathfrak{q}(\boldsymbol{A}_{\boldsymbol{\Sigma}})} \boldsymbol{\Lambda}^{-1}_{\mathfrak{q}(\boldsymbol{A}_{\boldsymbol{\Sigma}})}$.

S.6 Approximate Marginal Log-Likelihood for Two-Level Models

The expression for the lower bound on the marginal log-likelihood for Algorithm 2 is

$$
\begin{aligned}
&\log \underline{\mathfrak{p}}(\boldsymbol{y};\mathfrak{q}) = \\
&-\tfrac{1}{2}\log(\pi)\sum_{i=1}^{m} n_i - \tfrac{1}{2}\log|\boldsymbol{\Sigma}_{\boldsymbol{\beta}}| - \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}\text{tr}\left(\boldsymbol{M}_{\mathfrak{q}(\boldsymbol{\Sigma}^{-1})}\left\{\sum_{i=1}^{m}\left(\boldsymbol{\mu}_{\mathfrak{q}(\boldsymbol{u}_{\text{lin},i}}\boldsymbol{\mu}_{\mathfrak{q}(\boldsymbol{u}_{\text{lin},i})}^T+\boldsymbol{\Sigma}_{\mathfrak{q}(\boldsymbol{u}_{\text{lin},i})}\right)\right\}\right)+\tfrac{1}{2}\{2+K_{\text{gbl}}+m(2+K_{\text{grp}})\} \\
&-\tfrac{1}{2}\mu_{\mathfrak{q}(1/\sigma^2_{\text{gbl}})}\left\{\|\boldsymbol{\mu}_{\mathfrak{q}(\boldsymbol{u}_{\text{gbl}})}\|^2+\text{tr}(\boldsymbol{\Sigma}_{\mathfrak{q}(\boldsymbol{u}_{\text{gbl}})})\right\}-\tfrac{1}{2}\mu_{\mathfrak{q}(1/\sigma^2_{\text{grp}})}\sum_{i=1}^{m}\left\{\|\boldsymbol{\mu}_{\mathfrak{q}(\boldsymbol{u}_{\text{grp},i}))}\|^2+\text{tr}(\boldsymbol{\Sigma}_{\mathfrak{q}(\boldsymbol{u}_{\text{grp},i})})\right\} \\
&+\tfrac{1}{2}\log|\boldsymbol{\Sigma}_{\boldsymbol{\beta}}|+\{\nu_{\boldsymbol{\Sigma}}+m+1+\tfrac{1}{2}(\nu_\varepsilon+\nu_{\text{gbl}}+K_{\text{gbl}}+\nu_{\text{grp}}+mK_{\text{grp}})\}\log(2)-\log\Gamma(\tfrac{\nu_\varepsilon}{2}) \\
&-\tfrac{1}{2}\mu_{\mathfrak{q}(1/a_\varepsilon)}\mu_{\mathfrak{q}(1/\sigma^2_\varepsilon)}-\tfrac{1}{2}\xi_{\mathfrak{q}(\sigma^2_\varepsilon)}\log(\lambda_{\mathfrak{q}(\sigma^2_\varepsilon)})+\log\{\Gamma(\tfrac{1}{2}\xi_{\mathfrak{q}(\sigma^2_\varepsilon)})\}+\tfrac{1}{2}\lambda_{\mathfrak{q}(\sigma^2_\varepsilon)}\mu_{\mathfrak{q}(1/\sigma^2_\varepsilon)}-\tfrac{1}{2}\log(\nu_\varepsilon s^2_\varepsilon) \\
&-3\log\{\Gamma(\tfrac{1}{2})\}-\tfrac{1}{2\nu_\varepsilon s^2_\varepsilon}\mu_{\mathfrak{q}(1/a_\varepsilon)}-\tfrac{1}{2}\xi_{\mathfrak{q}(a_\varepsilon)}\log(\lambda_{\mathfrak{q}(a_\varepsilon)})+\log\{\Gamma(\tfrac{1}{2}\xi_{\mathfrak{q}(a_\varepsilon)})\}+\tfrac{1}{2}\lambda_{\mathfrak{q}(a_\varepsilon)}\mu_{\mathfrak{q}(1/a_\varepsilon)} \\
&-\log\Gamma(\tfrac{\nu_{\text{gbl}}}{2})-\tfrac{1}{2}\mu_{\mathfrak{q}(1/a_{\text{gbl}})}\mu_{\mathfrak{q}(1/\sigma^2_{\text{gbl}})}-\tfrac{1}{2}\xi_{\mathfrak{q}(\sigma^2_{\text{gbl}})}\log(\lambda_{\mathfrak{q}(\sigma^2_{\text{gbl}})})+\log\{\Gamma(\tfrac{1}{2}\xi_{\mathfrak{q}(\sigma^2_{\text{gbl}})})\}-\tfrac{1}{2}\log(\nu_{\text{gbl}}s^2_{\text{gbl}}) \\
&+\tfrac{1}{2}\lambda_{\mathfrak{q}(\sigma^2_{\text{gbl}})}\mu_{\mathfrak{q}(1/\sigma^2_{\text{gbl}})}-\{1/(2\nu_{\text{gbl}}s^2_{\text{gbl}})\}\mu_{\mathfrak{q}(1/a_{\text{gbl}})}-\tfrac{1}{2}\xi_{\mathfrak{q}(a_{\text{gbl}})}\log(\lambda_{\mathfrak{q}(a_{\text{gbl}})})-\tfrac{1}{2}\mu_{\mathfrak{q}(1/a_{\text{grp}})}\mu_{\mathfrak{q}(1/\sigma^2_{\text{grp}})} \\
&+\log\{\Gamma(\tfrac{1}{2}\xi_{\mathfrak{q}(a_{\text{gbl}})})\}+\tfrac{1}{2}\lambda_{\mathfrak{q}(a_{\text{gbl}})}\mu_{\mathfrak{q}(1/a_{\text{gbl}})}-\log\Gamma(\tfrac{\nu_{\text{grp}}}{2})+\log\{\Gamma(\tfrac{1}{2}\xi_{\mathfrak{q}(\sigma^2_{\text{grp}})})\}-\tfrac{1}{2}\log(\nu_{\text{grp}}s^2_{\text{grp}}) \\
&-\tfrac{1}{2}\xi_{\mathfrak{q}(\sigma^2_{\text{grp}})}\log(\lambda_{\mathfrak{q}(\sigma^2_{\text{grp}})})+\tfrac{1}{2}\lambda_{\mathfrak{q}(\sigma^2_{\text{grp}})}\mu_{\mathfrak{q}(1/\sigma^2_{\text{grp}})}-\{1/(2\nu_{\text{grp}}s^2_{\text{grp}})\}\mu_{\mathfrak{q}(1/a_{\text{grp}})}-\tfrac{1}{2}\xi_{\mathfrak{q}(a_{\text{grp}})}\log(\lambda_{\mathfrak{q}(a_{\text{grp}})}) \\
&+\log\{\Gamma(\tfrac{1}{2}\xi_{\mathfrak{q}(a_{\text{grp}})})\}+\tfrac{1}{2}\lambda_{\mathfrak{q}(a_{\text{grp}})}\mu_{\mathfrak{q}(1/a_{\text{grp}})}-\tfrac{1}{2}\text{tr}(\boldsymbol{M}_{\mathfrak{q}(\boldsymbol{A}_{\boldsymbol{\Sigma}}^{-1})}\boldsymbol{M}_{\mathfrak{q}(\boldsymbol{\Sigma}^{-1})})+\tfrac{1}{2}\text{tr}(\boldsymbol{\Lambda}_{\mathfrak{q}(\boldsymbol{\Sigma})}\boldsymbol{M}_{\mathfrak{q}(\boldsymbol{\Sigma}^{-1})}) \\
&+\sum_{j=1}^{2}\log\Gamma(\tfrac{1}{2}(\xi_{\mathfrak{q}(\boldsymbol{A}_{\boldsymbol{\Sigma}})}+2-j))-\sum_{j=1}^{2}\log\Gamma(\tfrac{1}{2}(\nu_{\boldsymbol{\Sigma}}+4-j))-\tfrac{1}{2}(\xi_{\mathfrak{q}(\boldsymbol{\Sigma})}-1)\log|\boldsymbol{\Lambda}_{\mathfrak{q}(\boldsymbol{\Sigma})}| \\
&+\sum_{j=1}^{2}\log\Gamma(\tfrac{1}{2}(\xi_{\mathfrak{q}(\boldsymbol{\Sigma})}+2-j))-\tfrac{1}{2}\sum_{j=1}^{2}1/(\nu_{\boldsymbol{\Sigma}}s^2_{\boldsymbol{\Sigma},j})\left(\boldsymbol{M}_{\mathfrak{q}(\boldsymbol{A}_{\boldsymbol{\Sigma}}^{-1})}\right)_{jj}-\sum_{j=1}^{2}\log\Gamma(\tfrac{1}{2}(3-j)) \\
&-\tfrac{1}{2}(\xi_{\mathfrak{q}(\boldsymbol{A}_{\boldsymbol{\Sigma}})}-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})}) \\
&-\tfrac{1}{2}\mu_{\mathfrak{q}(1/\sigma^2_\varepsilon)}\sum_{i=1}^{m}\left\{\left\|E_{\mathfrak{q}}\left(\boldsymbol{y}_i-\boldsymbol{C}_{\text{gbl},i}\begin{bmatrix}\boldsymbol{\beta}\\ \boldsymbol{u}_{\text{gbl},i}\end{bmatrix}-\boldsymbol{C}_{\text{grp},i}\begin{bmatrix}\boldsymbol{u}_{\text{lin},i}\\ \boldsymbol{u}_{\text{grp},i}\end{bmatrix}\right)\right\|^2\right. \\
&\qquad\qquad+\text{tr}(\boldsymbol{C}_{\text{gbl},i}^T\boldsymbol{C}_{\text{gbl},i}\boldsymbol{\Sigma}_{\mathfrak{q}(\boldsymbol{\beta},\boldsymbol{u}_{\text{gbl}})})+\text{tr}(\boldsymbol{C}_{\text{grp},i}^T\boldsymbol{C}_{\text{grp},i}\boldsymbol{\Sigma}_{\mathfrak{q}(\boldsymbol{u}_{\text{lin},i},\boldsymbol{u}_{\text{grp},i})}) \\
&\qquad\qquad+2\,\text{tr}\left[\boldsymbol{C}_{\text{grp},i}^T\boldsymbol{C}_{\text{gbl},i}\,E_{\mathfrak{q}}\left\{\left(\begin{bmatrix}\boldsymbol{\beta}\\ \boldsymbol{u}_{\text{gbl}}\end{bmatrix}-\boldsymbol{\mu}_{\mathfrak{q}(\boldsymbol{\beta},\boldsymbol{u}_{\text{gbl}})}\right)\times\right.\right. \\
&\qquad\qquad\qquad\qquad\left.\left.\left.\left(\begin{bmatrix}\boldsymbol{u}_{\text{lin},i}\\ \boldsymbol{u}_{\text{grp},i}\end{bmatrix}-\boldsymbol{\mu}_{\mathfrak{q}(\boldsymbol{u}_{\text{lin},i},\boldsymbol{u}_{\text{grp},i}))}\right)^T\right\}\right]\right\}.
\end{aligned}
\tag{S.1}
$$

Derivation: The lower-bound on the marginal log-likelihood is achieved through the following expression:

$$
\begin{aligned}
\log \underline{\mathfrak{p}}(\boldsymbol{y};\mathfrak{q}) \;=\; & E_{\mathfrak{q}}\{\log \mathfrak{p}(\boldsymbol{y},\boldsymbol{\beta},\boldsymbol{u},\sigma^2_\varepsilon,a_\varepsilon,\sigma^2_{\text{gbl}},a_{\text{gbl}},\sigma^2_{\text{grp}},a_{\text{grp}},\boldsymbol{\Sigma},\boldsymbol{A}_{\boldsymbol{\Sigma}}) \\
& -\log \mathfrak{q}^*(\boldsymbol{\beta},\boldsymbol{u},\sigma^2_\varepsilon,a_\varepsilon,\sigma^2_{\text{gbl}},a_{\text{gbl}},\sigma^2_{\text{grp}},a_{\text{grp}},\boldsymbol{\Sigma},\boldsymbol{A}_{\boldsymbol{\Sigma}})\}
\end{aligned}
$$

$$
\begin{aligned}
&= E_{\mathfrak{q}}\{\log \mathfrak{p}(\boldsymbol{y}\,|\,\boldsymbol{\beta},\boldsymbol{u},\sigma_\varepsilon^2)\}\\
&\quad + E_{\mathfrak{q}}\{\log \mathfrak{p}(\boldsymbol{\beta},\boldsymbol{u}\,|\,\sigma_{\text{gbl}}^2,\sigma_{\text{grp}}^2,\boldsymbol{\Sigma})\} - E_{\mathfrak{q}}\{\log \mathfrak{q}^*(\boldsymbol{\beta},\boldsymbol{u})\}\\
&\quad + E_{\mathfrak{q}}\{\log \mathfrak{p}(\sigma_\varepsilon^2\,|\,a_\varepsilon)\} - E_{\mathfrak{q}}\{\log \mathfrak{q}^*(\sigma_\varepsilon^2)\} + E_{\mathfrak{q}}\{\log \mathfrak{p}(a_\varepsilon)\} - E_{\mathfrak{q}}\{\log \mathfrak{q}^*(a_\varepsilon)\}\\
&\quad + E_{\mathfrak{q}}\{\log \mathfrak{p}(\sigma_{\text{gbl}}^2\,|\,a_{\text{gbl}})\} - E_{\mathfrak{q}}\{\log \mathfrak{q}^*(\sigma_{\text{gbl}}^2)\} + E_{\mathfrak{q}}\{\log \mathfrak{p}(a_{\text{gbl}})\} - E_{\mathfrak{q}}\{\log \mathfrak{q}^*(a_{\text{gbl}})\}\\
&\quad + E_{\mathfrak{q}}\{\log \mathfrak{p}(\sigma_{\text{grp}}^2\,|\,a_{\text{grp}})\} - E_{\mathfrak{q}}\{\log \mathfrak{q}^*(\sigma_{\text{grp}}^2)\} + E_{\mathfrak{q}}\{\log \mathfrak{p}(a_{\text{grp}})\} - E_{\mathfrak{q}}\{\log \mathfrak{q}^*(a_{\text{grp}})\}\\
&\quad + E_{\mathfrak{q}}\{\log \mathfrak{p}(\boldsymbol{\Sigma}\,|\,\boldsymbol{A}_{\boldsymbol{\Sigma}})\} - E_{\mathfrak{q}}\{\log \mathfrak{q}^*(\boldsymbol{\Sigma})\} + E_{\mathfrak{q}}\{\log \mathfrak{p}(\boldsymbol{A}_{\boldsymbol{\Sigma}})\} - E_{\mathfrak{q}}\{\log \mathfrak{q}^*(\boldsymbol{A}_{\boldsymbol{\Sigma}})\}.
\end{aligned}
$$

First we note that

$$
\log \mathfrak{p}(\boldsymbol{y}\,|\,\boldsymbol{\beta},\boldsymbol{u},\sigma_\varepsilon^2) = -\frac{1}{2}\log(2\pi)\sum_{i=1}^m n_i - \frac{1}{2}\log(\sigma_\varepsilon^2)\sum_{i=1}^m n_i - \frac{1}{2\sigma_\varepsilon^2}\sum_{i=1}^m \|\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta}-\boldsymbol{Z}\boldsymbol{u}\|^2
$$

where

$$
\begin{aligned}
&\|\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta}-\boldsymbol{Z}\boldsymbol{u}\|^2\\
&= \left\| \begin{bmatrix}\boldsymbol{y}_1\\ \vdots\\ \boldsymbol{y}_m\end{bmatrix} - \begin{bmatrix}\boldsymbol{X}_1\\ \vdots\\ \boldsymbol{X}_m\end{bmatrix}\boldsymbol{\beta} - \begin{bmatrix}\boldsymbol{Z}_{\text{gbl},1}\\ \vdots\\ \boldsymbol{Z}_{\text{gbl},m}\end{bmatrix}\boldsymbol{u}_{\text{gbl}} - \underset{1\le i\le m}{\text{blockdiag}}([\boldsymbol{X}_i\ \boldsymbol{Z}_{\text{grp},i}]) \begin{bmatrix}\boldsymbol{u}_{\text{lin},i}\\ \boldsymbol{u}_{\text{grp},i}\end{bmatrix}_{1\le i\le m} \right\|^2\\
&= \sum_{i=1}^m \|\boldsymbol{y}_i - \boldsymbol{X}_i\boldsymbol{\beta} - \boldsymbol{Z}_{\text{gbl},i}\boldsymbol{u}_{\text{gbl},i} - \boldsymbol{X}_i\boldsymbol{u}_{\text{lin},i} - \boldsymbol{Z}_{\text{grp},i}\boldsymbol{u}_{\text{grp},i}\|^2\\
&= \sum_{i=1}^m \left\| \boldsymbol{y}_i - \boldsymbol{C}_{\text{gbl},i}\begin{bmatrix}\boldsymbol{\beta}\\ \boldsymbol{u}_{\text{gbl},i}\end{bmatrix} - \boldsymbol{C}_{\text{grp},i}\begin{bmatrix}\boldsymbol{u}_{\text{lin},i}\\ \boldsymbol{u}_{\text{grp},i}\end{bmatrix} \right\|^2
\end{aligned}
$$

and

$$
\boldsymbol{C}_{\text{gbl},i} \equiv [\boldsymbol{X}_i\ \boldsymbol{Z}_{\text{gbl},i}], \quad \boldsymbol{C}_{\text{grp},i} \equiv [\boldsymbol{X}_i\ \boldsymbol{Z}_{\text{grp},i}].
$$

Therefore,

$$
\begin{aligned}
&E_{\mathfrak{q}}\{\log \mathfrak{p}(\boldsymbol{y}\,|\,\boldsymbol{\beta},\boldsymbol{u},\sigma_\varepsilon^2)\}\\
&= -\tfrac{1}{2}\log(2\pi)\sum_{i=1}^m n_i - \frac{1}{2}E_{\mathfrak{q}}\{\log(\sigma_\varepsilon^2)\}\sum_{i=1}^m n_i\\
&\quad -\tfrac{1}{2}\mu_{\mathfrak{q}(1/\sigma_\varepsilon^2)}\sum_{i=1}^m \left\{ \left\| E_{\mathfrak{q}}\left(\boldsymbol{y}_i - \boldsymbol{C}_{\text{gbl},i}\begin{bmatrix}\boldsymbol{\beta}\\ \boldsymbol{u}_{\text{gbl},i}\end{bmatrix} - \boldsymbol{C}_{\text{grp},i}\begin{bmatrix}\boldsymbol{u}_{\text{lin},i}\\ \boldsymbol{u}_{\text{grp},i}\end{bmatrix}\right)\right\|^2 \right.\\
&\qquad + \text{tr}(\boldsymbol{C}_{\text{gbl},i}^T\boldsymbol{C}_{\text{gbl},i}\boldsymbol{\Sigma}_{\mathfrak{q}(\boldsymbol{\beta},\boldsymbol{u}_{\text{gbl}})}) + \text{tr}(\boldsymbol{C}_{\text{grp},i}^T\boldsymbol{C}_{\text{grp},i}\boldsymbol{\Sigma}_{\mathfrak{q}(\boldsymbol{u}_{\text{lin},i},\boldsymbol{u}_{\text{grp},i})})\\
&\qquad \left. + 2\,\text{tr}\left[\boldsymbol{C}_{\text{grp},i}^T\boldsymbol{C}_{\text{gbl},i}\,E_{\mathfrak{q}}\left\{\left(\begin{bmatrix}\boldsymbol{\beta}\\ \boldsymbol{u}_{\text{gbl}}\end{bmatrix} - \boldsymbol{\mu}_{\mathfrak{q}(\boldsymbol{\beta},\boldsymbol{u}_{\text{gbl}})}\right)\left(\begin{bmatrix}\boldsymbol{u}_{\text{lin},i}\\ \boldsymbol{u}_{\text{grp},i}\end{bmatrix} - \boldsymbol{\mu}_{\mathfrak{q}(\boldsymbol{u}_{\text{lin},i},\boldsymbol{u}_{\text{grp},i}))}\right)^T\right\}\right]\right\}
\end{aligned}
$$

The remainder of the expectations in (S.4) are expressed as:

$$
\begin{aligned}
E_{\mathfrak{q}}\{\log \mathfrak{p}(\boldsymbol{\beta},\boldsymbol{u}\,|\,\sigma_{\text{gbl}}^2,\sigma_{\text{grp}}^2,\boldsymbol{\Sigma})\} \quad = \quad & -\tfrac{1}{2}\{2+K_{\text{gbl}}+m(2+K_{\text{grp}})\}\log(2\pi) - \tfrac{1}{2}\log|\boldsymbol{\Sigma}_{\boldsymbol{\beta}}|\\
& -\tfrac{K_{\text{gbl}}}{2}E_{\mathfrak{q}}\{\log(\sigma_{\text{gbl}}^2)\} - \tfrac{m}{2}E_{\mathfrak{q}}\{\log|\boldsymbol{\Sigma}|\} - \tfrac{mK_{\text{grp}}}{2}E_{\mathfrak{q}}\{\log(\sigma_{\text{grp}}^2)\}\\
& -\tfrac{1}{2}\text{tr}\left(\boldsymbol{\Sigma}_{\boldsymbol{\beta}}^{-1}\left\{\left(\boldsymbol{\mu}_{\mathfrak{q}(\boldsymbol{\beta})}-\boldsymbol{\mu}_{\boldsymbol{\beta}}\right)\left(\boldsymbol{\mu}_{\mathfrak{q}(\boldsymbol{\beta})}-\boldsymbol{\mu}_{\boldsymbol{\beta}}\right)^T + \boldsymbol{\Sigma}_{\mathfrak{q}(\boldsymbol{\beta})}\right\}\right)\\
& -\tfrac{1}{2}\mu_{\mathfrak{q}(1/\sigma_{\text{gbl}}^2)}\left\{\|\boldsymbol{\mu}_{\mathfrak{q}(\boldsymbol{u}_{\text{gbl}})}\|^2 + \text{tr}(\boldsymbol{\Sigma}_{\mathfrak{q}(\boldsymbol{u}_{\text{gbl}})})\right\}\\
& -\tfrac{1}{2}\text{tr}\left(\boldsymbol{M}_{\mathfrak{q}(\boldsymbol{\Sigma}^{-1})}\left\{\sum_{i=1}^m\left(\boldsymbol{\mu}_{\mathfrak{q}(\boldsymbol{u}_{\text{lin},i}}\boldsymbol{\mu}_{\mathfrak{q}(\boldsymbol{u}_{\text{lin},i})}^T + \boldsymbol{\Sigma}_{\mathfrak{q}(\boldsymbol{u}_{\text{lin},i})}\right)\right\}\right)\\
& -\tfrac{1}{2}\mu_{\mathfrak{q}(1/\sigma_{\text{grp}}^2)}\sum_{i=1}^m\left\{\|\boldsymbol{\mu}_{\mathfrak{q}(\boldsymbol{u}_{\text{grp},i}))}\|^2 + \text{tr}(\boldsymbol{\Sigma}_{\mathfrak{q}(\boldsymbol{u}_{\text{grp},i})})\right\}
\end{aligned}
$$

$$
\begin{aligned}
E_{\mathfrak{q}}\{\log\mathfrak{q}^*(\boldsymbol{\beta},\boldsymbol{u})\} &= -\tfrac{1}{2}\{2+K_{\text{gbl}}+m(2+K_{\text{grp}})\}-\tfrac{1}{2}\{2+K_{\text{gbl}}+m(2+K_{\text{grp}})\}\log(2\pi)\\
&\quad -\tfrac{1}{2}\log|\boldsymbol{\Sigma}_{\boldsymbol{\beta}}|\\
E_{\mathfrak{q}}\{\log\mathfrak{p}(\sigma_\varepsilon^2\,|\,a_\varepsilon)\} &= -\tfrac{1}{2}\nu_\varepsilon E_{\mathfrak{q}}\{\log(2a_\varepsilon)\}-\log\Gamma(\nu_\varepsilon/2)-(\tfrac{1}{2}\nu_\varepsilon+1)E_{\mathfrak{q}}\{\log(\sigma_\varepsilon^2)\}\\
&\quad -\tfrac{1}{2}\mu_{\mathfrak{q}(1/a_\varepsilon)}\mu_{\mathfrak{q}(1/\sigma_\varepsilon^2)}\\
E_{\mathfrak{q}}\{\log\mathfrak{q}^*(\sigma_\varepsilon^2)\} &= \tfrac{1}{2}\xi_{\mathfrak{q}(\sigma_\varepsilon^2)}\log(\lambda_{\mathfrak{q}(\sigma_\varepsilon^2)}/2)-\log\{\Gamma(\tfrac{1}{2}\xi_{\mathfrak{q}(\sigma_\varepsilon^2)})\}-(\tfrac{1}{2}\xi_{\mathfrak{q}(\sigma_\varepsilon^2)}+1)E_{\mathfrak{q}}\{\log(\sigma_\varepsilon^2)\}\\
&\quad -\tfrac{1}{2}\lambda_{\mathfrak{q}(\sigma_\varepsilon^2)}\mu_{\mathfrak{q}(1/\sigma_\varepsilon^2)}\\
E_{\mathfrak{q}}\{\log\mathfrak{p}(a_\varepsilon)\} &= -\tfrac{1}{2}\log(2\nu_\varepsilon s_\varepsilon^2)-\log\{\Gamma(\tfrac{1}{2})\}-(\tfrac{1}{2}+1)E_{\mathfrak{q}}\{\log(a_\varepsilon)\}\\
&\quad -\{1/(2\nu_\varepsilon s_\varepsilon^2)\}\mu_{\mathfrak{q}(1/a_\varepsilon)}\\
E_{\mathfrak{q}}\{\log\mathfrak{q}^*(a_\varepsilon)\} &= \tfrac{1}{2}\xi_{\mathfrak{q}(a_\varepsilon)}\log(\lambda_{\mathfrak{q}(a_\varepsilon)}/2)-\log\{\Gamma(\tfrac{1}{2}\xi_{\mathfrak{q}(a_\varepsilon)})\}-(\tfrac{1}{2}\xi_{\mathfrak{q}(a_\varepsilon)}+1)E_{\mathfrak{q}}\{\log(a_\varepsilon)\}\\
&\quad -\tfrac{1}{2}\lambda_{\mathfrak{q}(a_\varepsilon)}\mu_{\mathfrak{q}(1/a_\varepsilon)}\\
E_{\mathfrak{q}}\{\log\mathfrak{p}(\sigma_{\text{gbl}}^2\,|\,a_{\text{gbl}})\} &= -\tfrac{1}{2}\nu_{\text{gbl}}E_{\mathfrak{q}}\{\log(2a_{\text{gbl}})\}-\log\Gamma(\nu_{\text{gbl}}/2)-(\tfrac{1}{2}\nu_{\text{gbl}}+1)E_{\mathfrak{q}}\{\log(\sigma_{\text{gbl}}^2)\}\\
&\quad -\tfrac{1}{2}\mu_{\mathfrak{q}(1/a_{\text{gbl}})}\mu_{\mathfrak{q}(1/\sigma_{\text{gbl}}^2)}\\
E_{\mathfrak{q}}\{\log\mathfrak{q}^*(\sigma_{\text{gbl}}^2)\} &= \tfrac{1}{2}\xi_{\mathfrak{q}(\sigma_{\text{gbl}}^2)}\log(\lambda_{\mathfrak{q}(\sigma_{\text{gbl}}^2)}/2)-\log\{\Gamma(\tfrac{1}{2}\xi_{\mathfrak{q}(\sigma_{\text{gbl}}^2)})\}-(\tfrac{1}{2}\xi_{\mathfrak{q}(\sigma_{\text{gbl}}^2)}+1)E_{\mathfrak{q}}\{\log(\sigma_{\text{gbl}}^2)\}\\
&\quad -\tfrac{1}{2}\lambda_{\mathfrak{q}(\sigma_{\text{gbl}}^2)}\mu_{\mathfrak{q}(1/\sigma_{\text{gbl}}^2)}\\
E_{\mathfrak{q}}\{\log\mathfrak{p}(a_{\text{gbl}})\} &= -\tfrac{1}{2}\log(2\nu_{\text{gbl}}s_{\text{gbl}}^2)-\log\{\Gamma(\tfrac{1}{2})\}-(\tfrac{1}{2}+1)E_{\mathfrak{q}}\{\log(a_{\text{gbl}})\}\\
&\quad -\{1/(2\nu_{\text{gbl}}s_{\text{gbl}}^2)\}\mu_{\mathfrak{q}(1/a_{\text{gbl}})}\\
E_{\mathfrak{q}}\{\log\mathfrak{q}^*(a_{\text{gbl}})\} &= \tfrac{1}{2}\xi_{\mathfrak{q}(a_{\text{gbl}})}\log(\lambda_{\mathfrak{q}(a_{\text{gbl}})}/2)-\log\{\Gamma(\tfrac{1}{2}\xi_{\mathfrak{q}(a_{\text{gbl}})})\}-(\tfrac{1}{2}\xi_{\mathfrak{q}(a_{\text{gbl}})}+1)E_{\mathfrak{q}}\{\log(a_{\text{gbl}})\}\\
&\quad -\tfrac{1}{2}\lambda_{\mathfrak{q}(a_{\text{gbl}})}\mu_{\mathfrak{q}(1/a_{\text{gbl}})}\\
E_{\mathfrak{q}}\{\log\mathfrak{p}(\sigma_{\text{grp}}^2\,|\,a_{\text{grp}})\} &= -\tfrac{1}{2}\nu_{\text{grp}}E_{\mathfrak{q}}\{\log(2a_{\text{grp}})\}-\log\Gamma(\nu_{\text{grp}}/2)-(\tfrac{1}{2}\nu_{\text{grp}}+1)E_{\mathfrak{q}}\{\log(\sigma_{\text{grp}}^2)\}\\
&\quad -\tfrac{1}{2}\mu_{\mathfrak{q}(1/a_{\text{grp}})}\mu_{\mathfrak{q}(1/\sigma_{\text{grp}}^2)}\\
E_{\mathfrak{q}}\{\log\mathfrak{q}^*(\sigma_{\text{grp}}^2)\} &= \tfrac{1}{2}\xi_{\mathfrak{q}(\sigma_{\text{grp}}^2)}\log(\lambda_{\mathfrak{q}(\sigma_{\text{grp}}^2)}/2)-\log\{\Gamma(\tfrac{1}{2}\xi_{\mathfrak{q}(\sigma_{\text{grp}}^2)})\}-(\tfrac{1}{2}\xi_{\mathfrak{q}(\sigma_{\text{grp}}^2)}+1)E_{\mathfrak{q}}\{\log(\sigma_{\text{grp}}^2)\}\\
&\quad -\tfrac{1}{2}\lambda_{\mathfrak{q}(\sigma_{\text{grp}}^2)}\mu_{\mathfrak{q}(1/\sigma_{\text{grp}}^2)}\\
E_{\mathfrak{q}}\{\log\mathfrak{p}(a_{\text{grp}})\} &= -\tfrac{1}{2}\log(2\nu_{\text{grp}}s_{\text{grp}}^2)-\log\{\Gamma(\tfrac{1}{2})\}-(\tfrac{1}{2}+1)E_{\mathfrak{q}}\{\log(a_{\text{grp}})\}\\
&\quad -\{1/(2\nu_{\text{grp}}s_{\text{grp}}^2)\}\mu_{\mathfrak{q}(1/a_{\text{grp}})}\\
E_{\mathfrak{q}}\{\log\mathfrak{q}^*(a_{\text{grp}})\} &= \tfrac{1}{2}\xi_{\mathfrak{q}(a_{\text{grp}})}\log(\lambda_{\mathfrak{q}(a_{\text{grp}})}/2)-\log\{\Gamma(\tfrac{1}{2}\xi_{\mathfrak{q}(a_{\text{grp}})})\}-(\tfrac{1}{2}\xi_{\mathfrak{q}(a_{\text{grp}})}+1)E_{\mathfrak{q}}\{\log(a_{\text{grp}})\}\\
&\quad -\tfrac{1}{2}\lambda_{\mathfrak{q}(a_{\text{grp}})}\mu_{\mathfrak{q}(1/a_{\text{grp}})}\\
E_{\mathfrak{q}}[\log\{\mathfrak{p}(\boldsymbol{\Sigma}|\boldsymbol{A}_{\boldsymbol{\Sigma}})\}] &= -\tfrac{1}{2}(\nu_{\boldsymbol{\Sigma}}+1)E_{\mathfrak{q}}\{\log|\boldsymbol{A}_{\boldsymbol{\Sigma}}|\}-\tfrac{1}{2}(\nu_{\boldsymbol{\Sigma}}+4)E_{\mathfrak{q}}\{\log|\boldsymbol{\Sigma}|\}-\tfrac{1}{2}\log(\pi)\\
&\quad -\tfrac{1}{2}\text{tr}(\boldsymbol{M}_{\mathfrak{q}(\boldsymbol{A}_{\boldsymbol{\Sigma}}^{-1})}\boldsymbol{M}_{\mathfrak{q}(\boldsymbol{\Sigma}^{-1})})-(\nu_{\boldsymbol{\Sigma}}+3)\log(2)-\textstyle\sum_{j=1}^2\log\Gamma(\tfrac{1}{2}(\nu_{\boldsymbol{\Sigma}}+4-j))\\
E_{\mathfrak{q}}[\log\{\mathfrak{q}(\boldsymbol{\Sigma})\}] &= \tfrac{1}{2}(\xi_{\mathfrak{q}(\boldsymbol{\Sigma})}-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 -(\xi_{\mathfrak{q}(\boldsymbol{\Sigma})}+1)\log(2)-\tfrac{1}{2}\log(\pi)-\textstyle\sum_{j=1}^2\log\Gamma(\tfrac{1}{2}(\xi_{\mathfrak{q}(\boldsymbol{\Sigma})}+2-j))\\
E_{\mathfrak{q}}[\log\{\mathfrak{p}(\boldsymbol{A}_{\boldsymbol{\Sigma}})\}] &= -\tfrac{3}{2}E_{\mathfrak{q}}\{\log|\boldsymbol{A}_{\boldsymbol{\Sigma}}|\}-\tfrac{1}{2}\textstyle\sum_{j=1}^2 1/(\nu_{\boldsymbol{\Sigma}}s_{\boldsymbol{\Sigma},j}^2)\left(\boldsymbol{M}_{\mathfrak{q}(\boldsymbol{A}_{\boldsymbol{\Sigma}}^{-1})}\right)_{jj}-2\log(2)-\tfrac{1}{2}\log(\pi)\\
&\quad -\textstyle\sum_{j=1}^2\log\Gamma(\tfrac{1}{2}(3-j))\\
E_{\mathfrak{q}}[\log\{\mathfrak{q}(\boldsymbol{A}_{\boldsymbol{\Sigma}})\}] &= \tfrac{1}{2}(\xi_{\mathfrak{q}(\boldsymbol{A}_{\boldsymbol{\Sigma}})}-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}\text{tr}(\boldsymbol{\Lambda}_{\mathfrak{q}(\boldsymbol{A}_{\boldsymbol{\Sigma}})}\boldsymbol{M}_{\mathfrak{q}(\boldsymbol{A}_{\boldsymbol{\Sigma}}^{-1})})\\
&\quad -(\xi_{\mathfrak{q}(\boldsymbol{A}_{\boldsymbol{\Sigma}})}+1)\log(2)-\tfrac{1}{2}\log(\pi)-\textstyle\sum_{j=1}^2\log\Gamma(\tfrac{1}{2}(\xi_{\mathfrak{q}(\boldsymbol{A}_{\boldsymbol{\Sigma}})}+2-j))
\end{aligned}
$$

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

$$-\tfrac{1}{2}\sum_{i=1}^m n_i-\tfrac{1}{2}\nu_\varepsilon-1+\tfrac{1}{2}\xi_{\mathfrak{q}(\sigma_\varepsilon^2)}+1=-\tfrac{1}{2}\sum_{i=1}^m n_i-\tfrac{1}{2}\nu_\varepsilon-1+\tfrac{1}{2}(\nu_\varepsilon+\sum_{i=1}^m n_i)+1=0.$$

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

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

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

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

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

$$-\frac{m}{2} - \tfrac{1}{2}(\nu_{\boldsymbol{\Sigma}} + 4) + \tfrac{1}{2}(\xi_{\mathfrak{q}(\boldsymbol{\Sigma})} + 2) = -\tfrac{1}{2}(m + \nu_{\boldsymbol{\Sigma}} + 4) + \tfrac{1}{2}(m + \nu_{\boldsymbol{\Sigma}} + 4) = 0.$$

The coefficient of $E_{\mathfrak{q}}\{\log(a_{\varepsilon})\}$ is

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

The coefficient of $E_{\mathfrak{q}}\{\log(a_{\text{gbl}})\}$ is

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

The coefficient of $E_{\mathfrak{q}}\{\log(a_{\text{grp}})\}$ is

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

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

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

Therefore, these terms can be dropped and the approximate marginal log-likelihood expression in (S.1) results.

S.7 Derivation of Result 3

If $\boldsymbol{B}$ and $\boldsymbol{b}$ have the same forms given by equation (7) in Nolan & Wand (2018) with

$$\boldsymbol{b}_{ij} \equiv \begin{bmatrix} \sigma_{\varepsilon}^{-1}\boldsymbol{y}_{ij} \\ \boldsymbol{0} \\ \boldsymbol{0} \\ \boldsymbol{0} \\ \boldsymbol{0} \\ \boldsymbol{0} \end{bmatrix}, \qquad \boldsymbol{B}_{ij} \equiv \begin{bmatrix} \sigma_{\varepsilon}^{-1}\boldsymbol{X}_{ij} & \sigma_{\varepsilon}^{-1}\boldsymbol{Z}_{\text{gbl},ij} \\ \boldsymbol{O} & \left(\sum_{i=1}^{m} n_i\right)^{-1/2}\sigma_{\text{gbl}}^{-1}\boldsymbol{I}_{K_{\text{gbl}}} \\ \boldsymbol{O} & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{O} \end{bmatrix},$$

$$\overset{\bullet}{\boldsymbol{B}}_{ij} \equiv \begin{bmatrix} \sigma_{\varepsilon}^{-1}\boldsymbol{X}_{ij} & \sigma_{\varepsilon}^{-1}\boldsymbol{Z}^{g}_{\text{grp},ij} \\ \boldsymbol{O} & \boldsymbol{O} \\ n_i^{-1/2}\boldsymbol{\Sigma}_g^{-1/2} & \boldsymbol{O} \\ \boldsymbol{O} & n_i^{-1/2}\sigma_{\text{grp},g}^{-1}\boldsymbol{I}_{K^g_{\text{grp}}} \\ \boldsymbol{O} & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{O} \end{bmatrix} \quad \textit{and} \quad \overset{\bullet\bullet}{\boldsymbol{B}}_{ij} \equiv \begin{bmatrix} \sigma_{\varepsilon}^{-1}\boldsymbol{X}_{ij} & \sigma_{\varepsilon}^{-1}\boldsymbol{Z}^{h}_{\text{grp},ij} \\ \boldsymbol{O} & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{O} \\ \boldsymbol{\Sigma}_h^{-1/2} & \boldsymbol{O} \\ \boldsymbol{O} & \sigma_{\text{grp},h}^{-1}\boldsymbol{I}_{K^h_{\text{grp}}} \end{bmatrix},$$

then straightforward algebra leads to

$$\boldsymbol{B}^T\boldsymbol{B} = \boldsymbol{C}^T\boldsymbol{R}_{\text{BLUP}}^{-1}\boldsymbol{C} + \boldsymbol{D}_{\text{BLUP}} \text{ and } \boldsymbol{B}^T\boldsymbol{b} = \boldsymbol{C}^T\boldsymbol{R}_{\text{BLUP}}^{-1}\boldsymbol{y}$$

where

$$\boldsymbol{C} \equiv [\boldsymbol{X}\ \boldsymbol{Z}], \quad \boldsymbol{D}_{\text{BLUP}} \equiv \begin{bmatrix} \boldsymbol{O} & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{G}^{-1} \end{bmatrix} \quad \text{and} \quad \boldsymbol{R}_{\text{BLUP}} \equiv \sigma_\varepsilon^2 \boldsymbol{I}, \tag{S.2}$$

and $\boldsymbol{G}$ as defined in (3.3). The remainder of the derivation of Result 3 is analogous to that of Result 1.

S.8 Derivation of Algorithm 3

Algorithm 3 is simply a proceduralization of Result 3.

S.9 Derivation of Result 4

It is straightforward to verify that the $\boldsymbol{\mu}_{\mathfrak{q}(\boldsymbol{\beta},\boldsymbol{u})}$ and $\boldsymbol{\Sigma}_{\mathfrak{q}(\boldsymbol{\beta},\boldsymbol{u})}$ updates, given at (2.12) but with $\boldsymbol{D}_{\text{MFVB}}$ as given in (3.6), may be written as

$$\boldsymbol{\mu}_{\mathfrak{q}(\boldsymbol{\beta},\boldsymbol{u})} \longleftarrow (\boldsymbol{B}^T\boldsymbol{B})^{-1}\boldsymbol{B}^T\boldsymbol{b} \quad \text{and} \quad \boldsymbol{\Sigma}_{\mathfrak{q}(\boldsymbol{\beta},\boldsymbol{u})} \longleftarrow (\boldsymbol{B}^T\boldsymbol{B})^{-1}$$

where $\boldsymbol{B}$ and $\boldsymbol{b}$ have the forms given by equation (7) in Nolan & Wand (2018) with

$$\boldsymbol{b}_{ij} \equiv \begin{bmatrix} \mu_{\mathfrak{q}(1/\sigma_\varepsilon^2)}^{1/2}\boldsymbol{y}_{ij} \\ \left(\sum_{i=1}^m n_i\right)^{-1/2}\boldsymbol{\Sigma}_{\boldsymbol{\beta}}^{-1/2}\boldsymbol{\mu}_{\boldsymbol{\beta}} \\ \boldsymbol{0} \\ \boldsymbol{0} \\ \boldsymbol{0} \\ \boldsymbol{0} \\ \boldsymbol{0} \end{bmatrix}, \quad \boldsymbol{B}_{ij} \equiv \begin{bmatrix} \mu_{\mathfrak{q}(1/\sigma_\varepsilon^2)}^{1/2}\boldsymbol{X}_{ij} & \mu_{\mathfrak{q}(1/\sigma_\varepsilon^2)}^{1/2}\boldsymbol{Z}_{\text{gbl},ij} \\ \left(\sum_{i=1}^m n_i\right)^{-1/2}\boldsymbol{\Sigma}_{\boldsymbol{\beta}}^{-1/2} & \boldsymbol{O} \\ \boldsymbol{O} & \left(\sum_{i=1}^m n_i\right)^{-1/2}\mu_{\mathfrak{q}(1/\sigma_{\text{gbl}}^2)}^{1/2}\boldsymbol{I}_{K_{\text{gbl}}} \\ \boldsymbol{O} & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{O} \end{bmatrix},$$

$$\dot{\boldsymbol{B}}_{ij} \equiv \begin{bmatrix} \mu_{\mathfrak{q}(1/\sigma_\varepsilon^2)}^{1/2}\boldsymbol{X}_{ij} & \mu_{\mathfrak{q}(1/\sigma_\varepsilon^2)}^{1/2}\boldsymbol{Z}_{\text{grp},ij}^g \\ \boldsymbol{O} & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{O} \\ n_i^{-1/2}\boldsymbol{M}_{\mathfrak{q}(\boldsymbol{\Sigma}_g^{-1})}^{1/2} & \boldsymbol{O} \\ \boldsymbol{O} & n_i^{-1/2}\mu_{\mathfrak{q}(1/\sigma_{\text{grp},g}^2)}^{1/2}\boldsymbol{I}_{K_{\text{grp}}^g} \\ \boldsymbol{O} & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{O} \end{bmatrix} \quad \textit{and} \quad \ddot{\boldsymbol{B}}_{ij} \equiv \begin{bmatrix} \mu_{\mathfrak{q}(1/\sigma_\varepsilon^2)}^{1/2}\boldsymbol{X}_{ij} & \mu_{\mathfrak{q}(1/\sigma_\varepsilon^2)}^{1/2}\boldsymbol{Z}_{\text{grp},ij}^h \\ \boldsymbol{O} & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{O} \\ \boldsymbol{M}_{\mathfrak{q}(\boldsymbol{\Sigma}_h^{-1})}^{1/2} & \boldsymbol{O} \\ \boldsymbol{O} & \mu_{\mathfrak{q}(1/\sigma_{\text{grp},h}^2)}^{1/2}\boldsymbol{I}_{K_{\text{grp}}^h} \end{bmatrix}.$$

Result 4 immediately follows from Theorem 4 of Nolan & Wand (2018).

S.10 Derivation of Algorithm 4

Algorithm 4 relies on expressions for the $\mathfrak{q}$-densities for mean field variational Bayesian inference for the parameters in (3.4) with product density restriction (3.5). We have

$$\mathfrak{q}(\boldsymbol{\beta},\boldsymbol{u}) \text{ is a } N(\boldsymbol{\mu}_{\mathfrak{q}(\boldsymbol{\beta},\boldsymbol{u})}, \boldsymbol{\Sigma}_{\mathfrak{q}(\boldsymbol{\beta},\boldsymbol{u})}) \text{ density function}$$

where

$$\boldsymbol{\Sigma}_{\mathfrak{q}(\boldsymbol{\beta},\boldsymbol{u})} = (\boldsymbol{C}^T\boldsymbol{R}_{\text{MFVB}}^{-1}\boldsymbol{C} + \boldsymbol{D}_{\text{MFVB}})^{-1} \quad \text{and} \quad \boldsymbol{\mu}_{\mathfrak{q}(\boldsymbol{\beta},\boldsymbol{u})} = \boldsymbol{\Sigma}_{\mathfrak{q}(\boldsymbol{\beta},\boldsymbol{u})}(\boldsymbol{C}^T\boldsymbol{R}_{\text{MFVB}}^{-1}\boldsymbol{y} + \boldsymbol{o}_{\text{MFVB}})$$

with $\boldsymbol{R}_{\text{MFVB}} \equiv \mu^{-1}_{\mathfrak{q}(1/\sigma^2_\varepsilon)}\boldsymbol{I}$, $\boldsymbol{o}_{\text{MFVB}} \equiv \left[\begin{array}{c}\boldsymbol{\Sigma}_{\boldsymbol{\beta}}^{-1}\boldsymbol{\mu}_{\boldsymbol{\beta}}\\ \boldsymbol{0}\end{array}\right]$ and $\boldsymbol{D}_{\text{MFVB}}$ as given in (3.3),

$$\mathfrak{q}(\sigma^2_\varepsilon) \text{ is an Inverse-}\chi^2\left(\xi_{\mathfrak{q}(\sigma^2_\varepsilon)},\lambda_{\mathfrak{q}(\sigma^2_\varepsilon)}\right) \text{ density function}$$

where $\xi_{\mathfrak{q}(\sigma^2_\varepsilon)} = \nu_\varepsilon + \sum_{i=1}^m\sum_{j=1}^{n_i} o_{ij}$ and

$$\begin{aligned}
\lambda_{\mathfrak{q}(\sigma^2_\varepsilon)} &= \mu_{\mathfrak{q}(1/a_\varepsilon)} + \sum_{i=1}^m\sum_{j=1}^{n_i} E_{\mathfrak{q}}\left\{\left\|\boldsymbol{y}_{ij} - \boldsymbol{C}_{\text{gbl},ij}\left[\begin{array}{c}\boldsymbol{\beta}\\ \boldsymbol{u}_{\text{gbl}}\end{array}\right] - \boldsymbol{C}^g_{\text{grp},ij}\left[\begin{array}{c}\boldsymbol{u}^g_{\text{lin},i}\\ \boldsymbol{u}^g_{\text{grp},i}\end{array}\right] - \boldsymbol{C}^h_{\text{grp},ij}\left[\begin{array}{c}\boldsymbol{u}^h_{\text{lin},ij}\\ \boldsymbol{u}^h_{\text{grp},ij}\end{array}\right]\right\|^2\right\}\\
&= \mu_{\mathfrak{q}(1/a_\varepsilon)} + \sum_{i=1}^m\sum_{j=1}^{n_i}\left[\left\|E_{\mathfrak{q}}\left(\boldsymbol{y}_{ij} - \boldsymbol{C}_{\text{gbl},ij}\left[\begin{array}{c}\boldsymbol{\beta}\\ \boldsymbol{u}_{\text{gbl}}\end{array}\right] - \boldsymbol{C}^g_{\text{grp},ij}\left[\begin{array}{c}\boldsymbol{u}^g_{\text{lin},i}\\ \boldsymbol{u}^g_{\text{grp},i}\end{array}\right] - \boldsymbol{C}^h_{\text{grp},ij}\left[\begin{array}{c}\boldsymbol{u}^h_{\text{lin},ij}\\ \boldsymbol{u}^h_{\text{grp},ij}\end{array}\right]\right)\right\|^2\right.\\
&\qquad\left.+\text{tr}\left\{\text{Cov}_{\mathfrak{q}}\left(\boldsymbol{C}_{\text{gbl},ij}\left[\begin{array}{c}\boldsymbol{\beta}\\ \boldsymbol{u}_{\text{gbl}}\end{array}\right] + \boldsymbol{C}^g_{\text{grp},ij}\left[\begin{array}{c}\boldsymbol{u}^g_{\text{lin},i}\\ \boldsymbol{u}^g_{\text{grp},i}\end{array}\right] + \boldsymbol{C}^h_{\text{grp},ij}\left[\begin{array}{c}\boldsymbol{u}^h_{\text{lin},ij}\\ \boldsymbol{u}^h_{\text{grp},ij}\end{array}\right]\right)\right\}\right]\\
&= \mu_{\mathfrak{q}(1/a_\varepsilon)} + \sum_{i=1}^m\sum_{j=1}^{n_i}\left\{\left\|E_{\mathfrak{q}}\left(\boldsymbol{y}_{ij} - \boldsymbol{C}_{\text{gbl},ij}\left[\begin{array}{c}\boldsymbol{\beta}\\ \boldsymbol{u}_{\text{gbl}}\end{array}\right] - \boldsymbol{C}^g_{\text{grp},ij}\left[\begin{array}{c}\boldsymbol{u}^g_{\text{lin},i}\\ \boldsymbol{u}^g_{\text{grp},i}\end{array}\right] - \boldsymbol{C}^h_{\text{grp},ij}\left[\begin{array}{c}\boldsymbol{u}^h_{\text{lin},ij}\\ \boldsymbol{u}^h_{\text{grp},ij}\end{array}\right]\right)\right\|^2\right.\\
&\qquad+\text{tr}(\boldsymbol{C}^T_{\text{gbl},ij}\boldsymbol{C}_{\text{gbl},ij}\boldsymbol{\Sigma}_{\mathfrak{q}(\boldsymbol{\beta},\boldsymbol{u}_{\text{gbl}})}) + \text{tr}((\boldsymbol{C}^g_{\text{grp},ij})^T\boldsymbol{C}^g_{\text{grp},ij}\boldsymbol{\Sigma}_{\mathfrak{q}(\boldsymbol{u}^g_{\text{lin},i},\boldsymbol{u}^g_{\text{grp},i})}) + \text{tr}((\boldsymbol{C}^h_{\text{grp},ij})^T\boldsymbol{C}^h_{\text{grp},ij}\boldsymbol{\Sigma}_{\mathfrak{q}(\boldsymbol{u}^h_{\text{lin},ij},\boldsymbol{u}^h_{\text{grp},ij})})\\
&\qquad+2\,\text{tr}\left[(\boldsymbol{C}^g_{\text{grp},ij})^T\boldsymbol{C}_{\text{gbl},ij}\,E_{\mathfrak{q}}\left\{\left(\left[\begin{array}{c}\boldsymbol{\beta}\\ \boldsymbol{u}_{\text{gbl}}\end{array}\right] - \boldsymbol{\mu}_{\mathfrak{q}(\boldsymbol{\beta},\boldsymbol{u}_{\text{gbl}})}\right)\left(\left[\begin{array}{c}\boldsymbol{u}^g_{\text{lin},i}\\ \boldsymbol{u}^g_{\text{grp},i}\end{array}\right] - \boldsymbol{\mu}_{\mathfrak{q}(\boldsymbol{u}^g_{\text{lin},i},\boldsymbol{u}^g_{\text{grp},i})}\right)^T\right\}\right]\\
&\qquad+2\,\text{tr}\left[(\boldsymbol{C}^h_{\text{grp},ij})^T\boldsymbol{C}_{\text{gbl},ij}\,E_{\mathfrak{q}}\left\{\left(\left[\begin{array}{c}\boldsymbol{\beta}\\ \boldsymbol{u}_{\text{gbl}}\end{array}\right] - \boldsymbol{\mu}_{\mathfrak{q}(\boldsymbol{\beta},\boldsymbol{u}_{\text{gbl}})}\right)\left(\left[\begin{array}{c}\boldsymbol{u}^h_{\text{lin},ij}\\ \boldsymbol{u}^h_{\text{grp},ij}\end{array}\right] - \boldsymbol{\mu}_{\mathfrak{q}(\boldsymbol{u}^h_{\text{lin},ij},\boldsymbol{u}^h_{\text{grp},ij})}\right)^T\right\}\right]\\
&\qquad\left.+2\,\text{tr}\left[(\boldsymbol{C}^g_{\text{grp},ij})^T\boldsymbol{C}^h_{\text{grp},ij}\,E_{\mathfrak{q}}\left\{\left(\left[\begin{array}{c}\boldsymbol{u}^g_{\text{lin},i}\\ \boldsymbol{u}^g_{\text{grp},i}\end{array}\right] - \boldsymbol{\mu}_{\mathfrak{q}(\boldsymbol{u}^g_{\text{lin},i},\boldsymbol{u}^g_{\text{grp},i})}\right)\left(\left[\begin{array}{c}\boldsymbol{u}^h_{\text{lin},ij}\\ \boldsymbol{u}^h_{\text{grp},ij}\end{array}\right] - \boldsymbol{\mu}_{\mathfrak{q}(\boldsymbol{u}^h_{\text{lin},ij},\boldsymbol{u}^h_{\text{grp},ij})}\right)^T\right\}\right]\right\}
\end{aligned}$$

where $\boldsymbol{C}_{\text{gbl},ij} \equiv [\boldsymbol{X}_{ij}\ \boldsymbol{Z}_{\text{gbl},ij}]$, $\boldsymbol{C}^g_{\text{grp},ij} \equiv [\boldsymbol{X}_{ij}\ \boldsymbol{Z}^g_{\text{grp},ij}]$, $\boldsymbol{C}^h_{\text{grp},ij} \equiv [\boldsymbol{X}_{ij}\ \boldsymbol{Z}^h_{\text{grp},ij}]$ and with reciprocal moment $\mu_{\mathfrak{q}(1/\sigma^2_\varepsilon)} = \xi_{\mathfrak{q}(\sigma^2_\varepsilon)}/\lambda_{\mathfrak{q}(\sigma^2_\varepsilon)}$,

$$\mathfrak{q}(\sigma^2_{\text{gbl}}) \text{ is an Inverse-}\chi^2\left(\xi_{\mathfrak{q}(\sigma^2_{\text{gbl}})},\lambda_{\mathfrak{q}(\sigma^2_{\text{gbl}})}\right) \text{ density function}$$

where $\xi_{\mathfrak{q}(\sigma^2_{\text{gbl}})} = \nu_{\text{gbl}} + K_{\text{gbl}}$ and

$$\lambda_{\mathfrak{q}(\sigma^2_{\text{gbl}})} = \mu_{\mathfrak{q}(1/a_{\text{gbl}})} + \|\boldsymbol{\mu}_{\mathfrak{q}(\boldsymbol{u}_{\text{gbl}})}\|^2 + \text{tr}\left(\boldsymbol{\Sigma}_{\mathfrak{q}(\boldsymbol{u}_{\text{gbl}})}\right),$$

with reciprocal moment $\mu_{\mathfrak{q}(1/\sigma^2_{\text{gbl}})} = \xi_{\mathfrak{q}(\sigma^2_{\text{gbl}})}/\lambda_{\mathfrak{q}(\sigma^2_{\text{gbl}})}$,

$$\mathfrak{q}(\sigma^2_{\text{grp},g}) \text{ is an Inverse-}\chi^2\left(\xi_{\mathfrak{q}(\sigma^2_{\text{grp},g})},\lambda_{\mathfrak{q}(\sigma^2_{\text{grp},g})}\right) \text{ density function}$$

where $\xi_{\mathfrak{q}(\sigma^2_{\text{grp},g})} = \nu_{\text{grp},\,g} + mK^g_{\text{grp}}$ and

$$\lambda_{\mathfrak{q}(\sigma^2_{\text{grp},g})} = \mu_{\mathfrak{q}(1/a_{\text{grp},\,g})} + \sum_{i=1}^m\left\{\|\boldsymbol{\mu}_{\mathfrak{q}(\boldsymbol{u}^g_{\text{grp},i})}\|^2 + \text{tr}\left(\boldsymbol{\Sigma}_{\mathfrak{q}(\boldsymbol{u}^g_{\text{grp},i})}\right)\right\},$$

with reciprocal moment $\mu_{\mathfrak{q}(1/\sigma^2_{\text{grp},g})} = \xi_{\mathfrak{q}(\sigma^2_{\text{grp},g})}/\lambda_{\mathfrak{q}(\sigma^2_{\text{grp},g})}$,

$$\mathfrak{q}(\boldsymbol{\Sigma}_g) \text{ is an Inverse-G-Wishart } \big(G_{\text{full}}, \xi_{\mathfrak{q}(\boldsymbol{\Sigma}_g)}, \boldsymbol{\Lambda}_{\mathfrak{q}(\boldsymbol{\Sigma}_g)}\big) \text{ density function}$$

where $\xi_{\mathfrak{q}(\boldsymbol{\Sigma}_g)} = \nu_{\boldsymbol{\Sigma}_g} + 2 + m$ and

$$\boldsymbol{\Lambda}_{\mathfrak{q}(\boldsymbol{\Sigma}_g)} = \boldsymbol{M}_{\mathfrak{q}(\boldsymbol{A}^{-1}_{\boldsymbol{\Sigma}_g})} + \sum_{i=1}^{m} \Big(\boldsymbol{\mu}_{\mathfrak{q}(\boldsymbol{u}^g_{\text{lin},i})} \boldsymbol{\mu}^T_{\mathfrak{q}(\boldsymbol{u}^g_{\text{lin},i})} + \boldsymbol{\Sigma}_{\mathfrak{q}(\boldsymbol{u}^g_{\text{lin},i})} \Big),$$

with inverse moment $\boldsymbol{M}_{\mathfrak{q}(\boldsymbol{\Sigma}_g^{-1})} = (\xi_{\mathfrak{q}(\boldsymbol{\Sigma}_g)} - 1)\boldsymbol{\Lambda}^{-1}_{\mathfrak{q}(\boldsymbol{\Sigma}_g)}$,

$$\mathfrak{q}(\sigma^2_{\text{grp},h}) \text{ is an Inverse-}\chi^2 \left(\xi_{\mathfrak{q}(\sigma^2_{\text{grp},h})}, \lambda_{\mathfrak{q}(\sigma^2_{\text{grp},h})}\right) \text{ density function}$$

where $\xi_{\mathfrak{q}(\sigma^2_{\text{grp},h})} = \nu_{\text{grp},\,h} + K^h_{\text{grp}} \sum_{i=1}^m n_i$ and

$$\lambda_{\mathfrak{q}(\sigma^2_{\text{grp},h})} = \mu_{\mathfrak{q}(1/a_{\text{grp},\,h})} + \sum_{i=1}^{m}\sum_{j=1}^{n_i} \left\{ \|\boldsymbol{\mu}_{\mathfrak{q}(\boldsymbol{u}^h_{\text{grp},ij})}\|^2 + \text{tr}\left(\boldsymbol{\Sigma}_{\mathfrak{q}(\boldsymbol{u}^h_{\text{grp},ij})}\right)\right\},$$

with reciprocal moment $\mu_{\mathfrak{q}(1/\sigma^2_{\text{grp},h})} = \xi_{\mathfrak{q}(\sigma^2_{\text{grp},h})}/\lambda_{\mathfrak{q}(\sigma^2_{\text{grp},h})}$,

$$\mathfrak{q}(\boldsymbol{\Sigma}_h) \text{ is an Inverse-G-Wishart } \big(G_{\text{full}}, \xi_{\mathfrak{q}(\boldsymbol{\Sigma}_h)}, \boldsymbol{\Lambda}_{\mathfrak{q}(\boldsymbol{\Sigma}_h)}\big) \text{ density function}$$

where $\xi_{\mathfrak{q}(\boldsymbol{\Sigma}_h)} = \nu_{\boldsymbol{\Sigma}_h} + 2 + \sum_{i=1}^m n_i$ and

$$\boldsymbol{\Lambda}_{\mathfrak{q}(\boldsymbol{\Sigma}_h)} = \boldsymbol{M}_{\mathfrak{q}(\boldsymbol{A}^{-1}_{\boldsymbol{\Sigma}_h})} + \sum_{i=1}^{m}\sum_{j=1}^{n_i} \Big(\boldsymbol{\mu}_{\mathfrak{q}(\boldsymbol{u}^h_{\text{lin},ij})} \boldsymbol{\mu}^T_{\mathfrak{q}(\boldsymbol{u}^h_{\text{lin},ij})} + \boldsymbol{\Sigma}_{\mathfrak{q}(\boldsymbol{u}^h_{\text{lin},ij})} \Big),$$

with inverse moment $\boldsymbol{M}_{\mathfrak{q}(\boldsymbol{\Sigma}_h^{-1})} = (\xi_{\mathfrak{q}(\boldsymbol{\Sigma}_h)} - 1)\boldsymbol{\Lambda}^{-1}_{\mathfrak{q}(\boldsymbol{\Sigma}_h)}$,

$$\mathfrak{q}(a_\varepsilon) \text{ is an Inverse-}\chi^2(\xi_{\mathfrak{q}(a_\varepsilon)}, \lambda_{\mathfrak{q}(a_\varepsilon)}) \text{ density function}$$

where $\xi_{\mathfrak{q}(a_\varepsilon)} = \nu_\varepsilon + 1$,

$$\lambda_{\mathfrak{q}(a_\varepsilon)} = \mu_{\mathfrak{q}(1/\sigma^2_\varepsilon)} + 1/(\nu_\varepsilon s^2_\varepsilon)$$

with reciprocal moment $\mu_{\mathfrak{q}(1/a_\varepsilon)} = \xi_{\mathfrak{q}(a_\varepsilon)}/\lambda_{\mathfrak{q}(a_\varepsilon)}$,

$$\mathfrak{q}(a_{\text{gbl}}) \text{ is an Inverse-}\chi^2(\xi_{\mathfrak{q}(a_{\text{gbl}})}, \lambda_{\mathfrak{q}(a_{\text{gbl}})}) \text{ density function}$$

where $\xi_{\mathfrak{q}(a_{\text{gbl}})} = \nu_{\text{gbl}} + 1$,

$$\lambda_{\mathfrak{q}(a_{\text{gbl}})} = \mu_{\mathfrak{q}(1/\sigma^2_{\text{gbl}})} + 1/(\nu_{\text{gbl}} s^2_{\text{gbl}})$$

with reciprocal moment $\mu_{\mathfrak{q}(1/a_{\text{gbl}})} = \xi_{\mathfrak{q}(a_{\text{gbl}})}/\lambda_{\mathfrak{q}(a_{\text{gbl}})}$,

$$\mathfrak{q}(a_{\text{grp},\,g}) \text{ is an Inverse-}\chi^2(\xi_{\mathfrak{q}(a_{\text{grp},\,g})}, \lambda_{\mathfrak{q}(a_{\text{grp},\,g})}) \text{ density function}$$

where $\xi_{\mathfrak{q}(a_{\text{grp},\,g})} = \nu_{\text{grp},\,g} + 1$,

$$\lambda_{\mathfrak{q}(a_{\text{grp},\,g})} = \mu_{\mathfrak{q}(1/\sigma^2_{\text{grp},g})} + 1/(\nu_{\text{grp},\,g} s^2_{\text{grp},\,g})$$

with reciprocal moment $\mu_{\mathfrak{q}(1/a_{\text{grp},\,g})} = \xi_{\mathfrak{q}(a_{\text{grp},\,g})}/\lambda_{\mathfrak{q}(a_{\text{grp},\,g})}$ and

$$\mathfrak{q}(\boldsymbol{A}_{\boldsymbol{\Sigma}_g}) \text{ is an Inverse-G-Wishart } \left(G_{\text{diag}}, \xi_{\mathfrak{q}(\boldsymbol{A}_{\boldsymbol{\Sigma}_g})}, \boldsymbol{\Lambda}_{\mathfrak{q}(\boldsymbol{A}_{\boldsymbol{\Sigma}_g})}\right) \text{ density function}$$

where $\xi_{\mathfrak{q}(\boldsymbol{A}_{\boldsymbol{\Sigma}_g})} = \nu_{\boldsymbol{\Sigma}_g} + 2$,

$$\boldsymbol{\Lambda}_{\mathfrak{q}(\boldsymbol{A}_{\boldsymbol{\Sigma}_g})} = \text{diag}\big\{\text{diagonal}\big(\boldsymbol{M}_{\mathfrak{q}(\boldsymbol{\Sigma}_g^{-1})}\big)\big\} + \boldsymbol{\Lambda}_{\boldsymbol{A}_{\boldsymbol{\Sigma}_g}}$$

with inverse moment $\boldsymbol{M}_{\mathfrak{q}(\boldsymbol{A}_{\boldsymbol{\Sigma}_g}^{-1})} = \xi_{\mathfrak{q}(\boldsymbol{A}_{\boldsymbol{\Sigma}_g})}\boldsymbol{\Lambda}_{\mathfrak{q}(\boldsymbol{A}_{\boldsymbol{\Sigma}_g})}^{-1}$,

$$\mathfrak{q}(a_{\text{grp},\,h}) \text{ is an Inverse-}\chi^2(\xi_{\mathfrak{q}(a_{\text{grp},\,h})}, \lambda_{\mathfrak{q}(a_{\text{grp},\,h})}) \text{ density function}$$

where $\xi_{\mathfrak{q}(a_{\text{grp},\,h})} = \nu_{\text{grp},\,h} + 1$,

$$\lambda_{\mathfrak{q}(a_{\text{grp},\,h})} = \mu_{\mathfrak{q}(1/\sigma^2_{\text{grp},h})} + 1/(\nu_{\text{grp},\,h}s^2_{\text{grp},\,h})$$

with reciprocal moment $\mu_{\mathfrak{q}(1/a_{\text{grp},\,h})} = \xi_{\mathfrak{q}(a_{\text{grp},\,h})}/\lambda_{\mathfrak{q}(a_{\text{grp},\,h})}$ and

$$\mathfrak{q}(\boldsymbol{A}_{\boldsymbol{\Sigma}_h}) \text{ is an Inverse-G-Wishart}\left(G_{\text{diag}}, \xi_{\mathfrak{q}(\boldsymbol{A}_{\boldsymbol{\Sigma}_h})}, \boldsymbol{\Lambda}_{\mathfrak{q}(\boldsymbol{A}_{\boldsymbol{\Sigma}_h})}\right) \text{ density function}$$

where $\xi_{\mathfrak{q}(\boldsymbol{A}_{\boldsymbol{\Sigma}_h})} = \nu_{\boldsymbol{\Sigma}_h} + 2$

$$\boldsymbol{\Lambda}_{\mathfrak{q}(\boldsymbol{A}_{\boldsymbol{\Sigma}_h})} = \text{diag}\big\{\text{diagonal}\big(\boldsymbol{M}_{\mathfrak{q}(\boldsymbol{\Sigma}_h^{-1})}\big)\big\} + \boldsymbol{\Lambda}_{\boldsymbol{A}_{\boldsymbol{\Sigma}_h}}$$

with inverse moment $\boldsymbol{M}_{\mathfrak{q}(\boldsymbol{A}_{\boldsymbol{\Sigma}_h}^{-1})} = \xi_{\mathfrak{q}(\boldsymbol{A}_{\boldsymbol{\Sigma}_h})}\boldsymbol{\Lambda}_{\mathfrak{q}(\boldsymbol{A}_{\boldsymbol{\Sigma}_h})}^{-1}$.

S.11 Approximate Marginal Log-Likelihood for Three-Level Models

$$
\begin{aligned}
&\log \underline{\mathfrak{p}}(\boldsymbol{y};\mathfrak{q}) = \\
&-\tfrac{1}{2}\left\{\sum_{i=1}^{m}\sum_{j=1}^{n_i} o_{ij}+4\right\}\log(\pi)+\tfrac{1}{2}\left\{2+K_{\text{gbl}}+m(2+K^g_{\text{grp}})+\sum_{i=1}^{m} n_i(2+K^h_{\text{grp}})\right\} \\
&-\tfrac{1}{2}\left(\sum_{i=1}^{m}\sum_{j=1}^{n_i} o_{ij}+\nu_\varepsilon+\nu_{\text{gbl}}+\nu_{\text{grp},\,g}+\nu_{\text{grp},\,h}\right)\log(2)+\left(2+m+\sum_{i=1}^{m} n_i+\nu_{\boldsymbol{\Sigma}_g}+\nu_{\boldsymbol{\Sigma}_h}\right)\log(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}_{\boldsymbol{\beta}}|+\tfrac{1}{2}\log|\boldsymbol{\Sigma}_{\mathfrak{q}(\boldsymbol{\beta},\boldsymbol{u})}| \\
&-\tfrac{1}{2}\text{tr}\left(\boldsymbol{M}_{\mathfrak{q}(\boldsymbol{\Sigma}_g^{-1})}\sum_{i=1}^{m}\left(\boldsymbol{\mu}_{\mathfrak{q}(\boldsymbol{u}^g_{\text{lin},i})}\boldsymbol{\mu}^T_{\mathfrak{q}(\boldsymbol{u}^g_{\text{lin},i})}+\boldsymbol{\Sigma}_{\mathfrak{q}(\boldsymbol{u}^g_{\text{lin},i})}\right)\right)-\tfrac{1}{2}\mu_{\mathfrak{q}(1/\sigma^2_{\text{grp},h})}\sum_{i=1}^{m}\sum_{j=1}^{n_i}\left\{\|\boldsymbol{\mu}_{\mathfrak{q}(\boldsymbol{u}^h_{\text{grp},ij}))}\|^2+\text{tr}(\boldsymbol{\Sigma}_{\mathfrak{q}(\boldsymbol{u}^h_{\text{grp},ij})})\right\} \\
&-\tfrac{1}{2}\text{tr}\left(\boldsymbol{M}_{\mathfrak{q}(\boldsymbol{\Sigma}_h^{-1})}\sum_{i=1}^{m}\sum_{j=1}^{n_i}\left(\boldsymbol{\mu}_{\mathfrak{q}(\boldsymbol{u}^h_{\text{lin},ij}}\boldsymbol{\mu}^T_{\mathfrak{q}(\boldsymbol{u}^h_{\text{lin},ij})}+\boldsymbol{\Sigma}_{\mathfrak{q}(\boldsymbol{u}^h_{\text{lin},ij})}\right)\right)-\tfrac{1}{2}\mu_{\mathfrak{q}(1/\sigma^2_{\text{gbl}})}\left\{\|\boldsymbol{\mu}_{\mathfrak{q}(\boldsymbol{u}_{\text{gbl}})}\|^2+\text{tr}(\boldsymbol{\Sigma}_{\mathfrak{q}(\boldsymbol{u}_{\text{gbl}})})\right\} \\
&-\tfrac{1}{2}\mu_{\mathfrak{q}(1/\sigma^2_{\text{grp},g})}\sum_{i=1}^{m}\left\{\|\boldsymbol{\mu}_{\mathfrak{q}(\boldsymbol{u}^g_{\text{grp},i}))}\|^2+\text{tr}(\boldsymbol{\Sigma}_{\mathfrak{q}(\boldsymbol{u}^g_{\text{grp},i})})\right\}-\log\Gamma(\frac{\nu_\varepsilon}{2})-\tfrac{1}{2}\mu_{\mathfrak{q}(1/a_\varepsilon)}\mu_{\mathfrak{q}(1/\sigma^2_\varepsilon)} \\
&-\tfrac{1}{2}\xi_{\mathfrak{q}(\sigma^2_\varepsilon)}\log(\lambda_{\mathfrak{q}(\sigma^2_\varepsilon)})+\log\{\Gamma(\tfrac{1}{2}\xi_{\mathfrak{q}(\sigma^2_\varepsilon)})\}+\tfrac{1}{2}\lambda_{\mathfrak{q}(\sigma^2_\varepsilon)}\mu_{\mathfrak{q}(1/\sigma^2_\varepsilon)}-\tfrac{1}{2}\log(\nu_\varepsilon s^2_\varepsilon)-\tfrac{1}{2\nu_\varepsilon s^2_\varepsilon}\mu_{\mathfrak{q}(1/a_\varepsilon)}-\tfrac{1}{2}\xi_{\mathfrak{q}(a_\varepsilon)}\log(\lambda_{\mathfrak{q}(a_\varepsilon)}) \\
&+\log\{\Gamma(\tfrac{1}{2}\xi_{\mathfrak{q}(a_\varepsilon)})\}+\tfrac{1}{2}\lambda_{\mathfrak{q}(a_\varepsilon)}\mu_{\mathfrak{q}(1/a_\varepsilon)}-\log\Gamma(\tfrac{\nu_{\text{gbl}}}{2})-\tfrac{1}{2}\mu_{\mathfrak{q}(1/a_{\text{gbl}})}\mu_{\mathfrak{q}(1/\sigma^2_{\text{gbl}})}-\tfrac{1}{2}\xi_{\mathfrak{q}(\sigma^2_{\text{gbl}})}\log(\lambda_{\mathfrak{q}(\sigma^2_{\text{gbl}})}) \\
&+\log\{\Gamma(\tfrac{1}{2}\xi_{\mathfrak{q}(\sigma^2_{\text{gbl}})})\}-\tfrac{1}{2}\log(\nu_{\text{gbl}}s^2_{\text{gbl}})+\tfrac{1}{2}\lambda_{\mathfrak{q}(\sigma^2_{\text{gbl}})}\mu_{\mathfrak{q}(1/\sigma^2_{\text{gbl}})}-\{1/(2\nu_{\text{gbl}}s^2_{\text{gbl}})\}\mu_{\mathfrak{q}(1/a_{\text{gbl}})}-\tfrac{1}{2}\xi_{\mathfrak{q}(a_{\text{gbl}})}\log(\lambda_{\mathfrak{q}(a_{\text{gbl}})}) \\
&+\log\{\Gamma(\tfrac{1}{2}\xi_{\mathfrak{q}(a_{\text{gbl}})})\}+\tfrac{1}{2}\lambda_{\mathfrak{q}(a_{\text{gbl}})}\mu_{q(1/a_{\text{gbl}})}-\tfrac{1}{2}\mu_{\mathfrak{q}(1/a_{\text{grp},\,g})}\mu_{\mathfrak{q}(1/\sigma^2_{\text{grp},g})}-\log\Gamma(\tfrac{\nu_{\text{grp},\,g}}{2})+\log\{\Gamma(\tfrac{1}{2}\xi_{\mathfrak{q}(\sigma^2_{\text{grp},g})})\} \\
&-\tfrac{1}{2}\log(\nu_{\text{grp},\,g}s^2_{\text{grp},\,g})-\tfrac{1}{2}\xi_{\mathfrak{q}(\sigma^2_{\text{grp},g})}\log(\lambda_{\mathfrak{q}(\sigma^2_{\text{grp},g})})+\tfrac{1}{2}\lambda_{\mathfrak{q}(\sigma^2_{\text{grp},g})}\mu_{\mathfrak{q}(1/\sigma^2_{\text{grp},g})}-\{1/(2\nu_{\text{grp},\,g}s^2_{\text{grp},\,g})\}\mu_{\mathfrak{q}(1/a_{\text{grp},\,g})} \\
&-\tfrac{1}{2}\xi_{\mathfrak{q}(a_{\text{grp},\,g})}\log(\lambda_{\mathfrak{q}(a_{\text{grp},\,g})})+\log\{\Gamma(\tfrac{1}{2}\xi_{\mathfrak{q}(a_{\text{grp},\,g})})\}+\tfrac{1}{2}\lambda_{\mathfrak{q}(a_{\text{grp},\,g})}\mu_{\mathfrak{q}(1/a_{\text{grp},\,g})}-\tfrac{1}{2}\mu_{\mathfrak{q}(1/a_{\text{grp},\,h})}\mu_{\mathfrak{q}(1/\sigma^2_{\text{grp},h})} \\
&-\log\Gamma(\tfrac{\nu_{\text{grp},\,h}}{2})+\log\{\Gamma(\tfrac{1}{2}\xi_{\mathfrak{q}(\sigma^2_{\text{grp},h})})\}-\tfrac{1}{2}\log(\nu_{\text{grp},\,h}s^2_{\text{grp},\,h})-\tfrac{1}{2}\xi_{\mathfrak{q}(\sigma^2_{\text{grp},h})}\log(\lambda_{\mathfrak{q}(\sigma^2_{\text{grp},h})})+\tfrac{1}{2}\lambda_{\mathfrak{q}(\sigma^2_{\text{grp},h})}\mu_{\mathfrak{q}(1/\sigma^2_{\text{grp},h})} \\
&-\{1/(2\nu_{\text{grp},\,h}s^2_{\text{grp},\,h})\}\mu_{\mathfrak{q}(1/a_{\text{grp},\,h})}-\tfrac{1}{2}\xi_{\mathfrak{q}(a_{\text{grp},\,h})}\log(\lambda_{\mathfrak{q}(a_{\text{grp},\,h})})+\log\{\Gamma(\tfrac{1}{2}\xi_{\mathfrak{q}(a_{\text{grp},\,h})})\}+\tfrac{1}{2}\lambda_{\mathfrak{q}(a_{\text{grp},\,h})}\mu_{\mathfrak{q}(1/a_{\text{grp},\,h})} \\
&-\tfrac{1}{2}\text{tr}(\boldsymbol{M}_{\mathfrak{q}(\boldsymbol{A}^{-1}_{\boldsymbol{\Sigma}_g})}\boldsymbol{M}_{\mathfrak{q}(\boldsymbol{\Sigma}_g^{-1})})+\tfrac{1}{2}\text{tr}(\boldsymbol{\Lambda}_{\mathfrak{q}(\boldsymbol{\Sigma}_g)}\boldsymbol{M}_{\mathfrak{q}(\boldsymbol{\Sigma}_g^{-1})})-\tfrac{1}{2}\text{tr}(\boldsymbol{M}_{\mathfrak{q}(\boldsymbol{A}^{-1}_{\boldsymbol{\Sigma}_h})}\boldsymbol{M}_{\mathfrak{q}(\boldsymbol{\Sigma}_h^{-1})})+\tfrac{1}{2}\text{tr}(\boldsymbol{\Lambda}_{\mathfrak{q}(\boldsymbol{\Sigma}_h)}\boldsymbol{M}_{\mathfrak{q}(\boldsymbol{\Sigma}_h^{-1})}) \\
&+\sum_{j=1}^{2}\log\Gamma(\tfrac{1}{2}(\xi_{\mathfrak{q}(\boldsymbol{A}_{\boldsymbol{\Sigma}_g})}+2-j))-\sum_{j=1}^{2}\log\Gamma(\tfrac{1}{2}(\nu_{\boldsymbol{\Sigma}_g}+4-j))-\tfrac{1}{2}(\xi_{\mathfrak{q}(\boldsymbol{\Sigma}_g)}-1)\log|\boldsymbol{\Lambda}_{\mathfrak{q}(\boldsymbol{\Sigma}_g)}| \\
&+\sum_{j=1}^{2}\log\Gamma(\tfrac{1}{2}(\xi_{\mathfrak{q}(\boldsymbol{\Sigma}_g)}+2-j))-\tfrac{1}{2}\sum_{j=1}^{2}1/(\nu_{\boldsymbol{\Sigma}_g}s^2_{\boldsymbol{\Sigma}_g,\,j})\left(\boldsymbol{M}_{\mathfrak{q}(\boldsymbol{A}^{-1}_{\boldsymbol{\Sigma}_g})}\right)_{jj}-\sum_{j=1}^{2}\log\Gamma(\tfrac{1}{2}(3-j)) \\
&-\tfrac{1}{2}(\xi_{\mathfrak{q}(\boldsymbol{A}_{\boldsymbol{\Sigma}_g})}-1)\log|\boldsymbol{\Lambda}_{\mathfrak{q}(\boldsymbol{A}_{\boldsymbol{\Sigma}_g})}|+\tfrac{1}{2}\text{tr}(\boldsymbol{\Lambda}_{\mathfrak{q}(\boldsymbol{A}_{\boldsymbol{\Sigma}_g})}\boldsymbol{M}_{\mathfrak{q}(\boldsymbol{A}^{-1}_{\boldsymbol{\Sigma}_g})}) \\
&+\sum_{j=1}^{2}\log\Gamma(\tfrac{1}{2}(\xi_{\mathfrak{q}(\boldsymbol{A}_{\boldsymbol{\Sigma}_h})}+2-j))-\sum_{j=1}^{2}\log\Gamma(\tfrac{1}{2}(\nu_{\boldsymbol{\Sigma}_h}+4-j))-\tfrac{1}{2}(\xi_{\mathfrak{q}(\boldsymbol{\Sigma}_h)}-1)\log|\boldsymbol{\Lambda}_{\mathfrak{q}(\boldsymbol{\Sigma}_h)}| \\
&+\sum_{j=1}^{2}\log\Gamma(\tfrac{1}{2}(\xi_{\mathfrak{q}(\boldsymbol{\Sigma}_h)}+2-j))-\tfrac{1}{2}\sum_{j=1}^{2}1/(\nu_{\boldsymbol{\Sigma}_h}s^2_{\boldsymbol{\Sigma}_h,\,j})\left(\boldsymbol{M}_{\mathfrak{q}(\boldsymbol{A}^{-1}_{\boldsymbol{\Sigma}_h})}\right)_{jj}-\sum_{j=1}^{2}\log\Gamma(\tfrac{1}{2}(3-j)
\end{aligned}
\tag{S.3}
$$

$$
\begin{aligned}
&-\tfrac{1}{2}(\xi_{\mathfrak{q}(\boldsymbol{A}_{\boldsymbol{\Sigma}_h})}-1)\log|\boldsymbol{\Lambda}_{\mathfrak{q}(\boldsymbol{A}_{\boldsymbol{\Sigma}_h})}|+\tfrac{1}{2}\mathrm{tr}(\boldsymbol{\Lambda}_{\mathfrak{q}(\boldsymbol{A}_{\boldsymbol{\Sigma}_h})}\boldsymbol{M}_{\mathfrak{q}(\boldsymbol{A}_{\boldsymbol{\Sigma}_h}^{-1})})\\
&-\tfrac{1}{2}\mu_{\mathfrak{q}(1/\sigma_\varepsilon^2)}\sum_{i=1}^{m}\sum_{j=1}^{n_i}\left\{\left\|E_{\mathfrak{q}}\left(\boldsymbol{y}_{ij}-\boldsymbol{C}_{\mathrm{gbl},ij}\left[\begin{array}{c}\boldsymbol{\beta}\\ \boldsymbol{u}_{\mathrm{gbl},i}\end{array}\right]-\boldsymbol{C}^g_{\mathrm{grp},ij}\left[\begin{array}{c}\boldsymbol{u}^g_{\mathrm{lin},i}\\ \boldsymbol{u}^g_{\mathrm{grp},i}\end{array}\right]-\boldsymbol{C}^h_{\mathrm{grp},ij}\left[\begin{array}{c}\boldsymbol{u}^h_{\mathrm{lin},ij}\\ \boldsymbol{u}^h_{\mathrm{grp},ij}\end{array}\right]\right)\right\|^2\right.\\
&\qquad+\mathrm{tr}\left(\boldsymbol{C}_{\mathrm{gbl},ij}^T\boldsymbol{C}_{\mathrm{gbl},ij}\boldsymbol{\Sigma}_{\mathfrak{q}(\boldsymbol{\beta},\boldsymbol{u}_{\mathrm{gbl}})}\right)+\mathrm{tr}\left(\boldsymbol{C}^g_{\mathrm{grp},ij}{}^T\boldsymbol{C}^g_{\mathrm{grp},ij}\boldsymbol{\Sigma}_{\mathfrak{q}(\boldsymbol{u}^g_{\mathrm{lin},i},\boldsymbol{u}^g_{\mathrm{grp},i})}\right)\\
&\qquad+\mathrm{tr}\left(\boldsymbol{C}^h_{\mathrm{grp},ij}{}^T\boldsymbol{C}^h_{\mathrm{grp},ij}\boldsymbol{\Sigma}_{\mathfrak{q}(\boldsymbol{u}^h_{\mathrm{lin},ij},\boldsymbol{u}^h_{\mathrm{grp},ij})}\right)\\
&\qquad+2\,\mathrm{tr}\left[\boldsymbol{C}^g_{\mathrm{grp},ij}{}^T\boldsymbol{C}_{\mathrm{gbl},ij}\,E_{\mathfrak{q}}\left\{\left(\left[\begin{array}{c}\boldsymbol{\beta}\\ \boldsymbol{u}_{\mathrm{gbl}}\end{array}\right]-\boldsymbol{\mu}_{\mathfrak{q}(\boldsymbol{\beta},\boldsymbol{u}_{\mathrm{gbl}})}\right)\left(\left[\begin{array}{c}\boldsymbol{u}^g_{\mathrm{lin},i}\\ \boldsymbol{u}^g_{\mathrm{grp}}\end{array}\right]-\boldsymbol{\mu}_{\mathfrak{q}(\boldsymbol{u}^g_{\mathrm{lin},i},\boldsymbol{u}^g_{\mathrm{grp},i}))}\right)^T\right\}\right]\\
&\qquad+2\,\mathrm{tr}\left[\boldsymbol{C}^h_{\mathrm{grp},ij}{}^T\boldsymbol{C}_{\mathrm{gbl},ij}\,E_{\mathfrak{q}}\left\{\left(\left[\begin{array}{c}\boldsymbol{\beta}\\ \boldsymbol{u}_{\mathrm{gbl}}\end{array}\right]-\boldsymbol{\mu}_{\mathfrak{q}(\boldsymbol{\beta},\boldsymbol{u}_{\mathrm{gbl}})}\right)\left(\left[\begin{array}{c}\boldsymbol{u}^h_{\mathrm{lin},ij}\\ \boldsymbol{u}^h_{\mathrm{grp}}\end{array}\right]-\boldsymbol{\mu}_{\mathfrak{q}(\boldsymbol{u}^h_{\mathrm{lin},ij},\boldsymbol{u}^h_{\mathrm{grp},ij}))}\right)^T\right\}\right]\\
&\qquad\left.+2\,\mathrm{tr}\left[\boldsymbol{C}^g_{\mathrm{grp},ij}{}^T\boldsymbol{C}^h_{\mathrm{grp},ij}\,E_{\mathfrak{q}}\left\{\left(\left[\begin{array}{c}\boldsymbol{u}^g_{\mathrm{lin},i}\\ \boldsymbol{u}^g_{\mathrm{grp}}\end{array}\right]-\boldsymbol{\mu}_{\mathfrak{q}(\boldsymbol{u}^g_{\mathrm{lin},i},\boldsymbol{u}^g_{\mathrm{grp},i})}\right)\left(\left[\begin{array}{c}\boldsymbol{u}^h_{\mathrm{lin},ij}\\ \boldsymbol{u}^h_{\mathrm{grp}}\end{array}\right]-\boldsymbol{\mu}_{\mathfrak{q}(\boldsymbol{u}^h_{\mathrm{lin},ij},\boldsymbol{u}^h_{\mathrm{grp},ij}))}\right)^T\right\}\right]\right\}.
\end{aligned}
$$

Derivation: The lower bound on the marginal log-likelihood is achieved through the following expression:

$$
\begin{aligned}
\log \underline{\mathfrak{p}}(\boldsymbol{y};\mathfrak{q}) \quad = \quad & E_{\mathfrak{q}}\{\log\mathfrak{p}(\boldsymbol{y},\boldsymbol{\beta},\boldsymbol{u},\sigma_\varepsilon^2,a_\varepsilon,\sigma^2_{\mathrm{gbl}},a_{\mathrm{gbl}},\sigma^2_{\mathrm{grp},g},\boldsymbol{\Sigma}_g,a_{\mathrm{grp},\,g},\boldsymbol{A}_{\boldsymbol{\Sigma}_g},\sigma^2_{\mathrm{grp},h},\boldsymbol{\Sigma}_h,a_{\mathrm{grp},\,h},\boldsymbol{A}_{\boldsymbol{\Sigma}_h})\\
& -\ \log\mathfrak{q}^*(\boldsymbol{\beta},\boldsymbol{u},\sigma_\varepsilon^2,a_\varepsilon,\sigma^2_{\mathrm{gbl}},a_{\mathrm{gbl}},\sigma^2_{\mathrm{grp},g},\boldsymbol{\Sigma}_g,a_{\mathrm{grp},\,g},\boldsymbol{A}_{\boldsymbol{\Sigma}_g},\sigma^2_{\mathrm{grp},h},\boldsymbol{\Sigma}_h,a_{\mathrm{grp},\,h},\boldsymbol{A}_{\boldsymbol{\Sigma}_h})\}\\
= \quad & E_{\mathfrak{q}}\{\log\mathfrak{p}(\boldsymbol{y}\,|\,\boldsymbol{\beta},\boldsymbol{u},\sigma_\varepsilon^2)\}\\
& +E_{\mathfrak{q}}\{\log\mathfrak{p}(\boldsymbol{\beta},\boldsymbol{u}\,|\,\sigma^2_{\mathrm{gbl}},\sigma^2_{\mathrm{grp},g},\boldsymbol{\Sigma}_g,\sigma^2_{\mathrm{grp},h},\boldsymbol{\Sigma}_h)\}-E_{\mathfrak{q}}\{\log\mathfrak{q}^*(\boldsymbol{\beta},\boldsymbol{u})\}\\
& +E_{\mathfrak{q}}\{\log\mathfrak{p}(\sigma_\varepsilon^2\,|\,a_\varepsilon)\}-E_{\mathfrak{q}}\{\log\mathfrak{q}^*(\sigma_\varepsilon^2)\}+E_{\mathfrak{q}}\{\log\mathfrak{p}(a_\varepsilon)\}-E_{\mathfrak{q}}\{\log\mathfrak{q}^*(a_\varepsilon)\}\\
& +E_{\mathfrak{q}}\{\log\mathfrak{p}(\sigma^2_{\mathrm{gbl}}\,|\,a_{\mathrm{gbl}})\}-E_{\mathfrak{q}}\{\log\mathfrak{q}^*(\sigma^2_{\mathrm{gbl}})\}+E_{\mathfrak{q}}\{\log\mathfrak{p}(a_{\mathrm{gbl}})\}-E_{\mathfrak{q}}\{\log\mathfrak{q}^*(a_{\mathrm{gbl}})\}\\
& +E_{\mathfrak{q}}\{\log\mathfrak{p}(\sigma^2_{\mathrm{grp},g}\,|\,a_{\mathrm{grp},\,g})\}-E_{\mathfrak{q}}\{\log\mathfrak{q}^*(\sigma^2_{\mathrm{grp},g})\}+E_{\mathfrak{q}}\{\log\mathfrak{p}(a_{\mathrm{grp},\,g})\}-E_{\mathfrak{q}}\{\log\mathfrak{q}^*(a_{\mathrm{grp},\,g})\}\\
& +E_{\mathfrak{q}}\{\log\mathfrak{p}(\boldsymbol{\Sigma}_g\,|\,\boldsymbol{A}_{\boldsymbol{\Sigma}_g})\}-E_{\mathfrak{q}}\{\log\mathfrak{q}^*(\boldsymbol{\Sigma}_g)\}+E_{\mathfrak{q}}\{\log\mathfrak{p}(\boldsymbol{A}_{\boldsymbol{\Sigma}_g})\}-E_{\mathfrak{q}}\{\log\mathfrak{q}^*(\boldsymbol{A}_{\boldsymbol{\Sigma}_g})\}\\
& +E_{\mathfrak{q}}\{\log\mathfrak{p}(\sigma^2_{\mathrm{grp},h}\,|\,a_{\mathrm{grp},\,g})\}-E_{\mathfrak{q}}\{\log\mathfrak{q}^*(\sigma^2_{\mathrm{grp},h})\}+E_{\mathfrak{q}}\{\log\mathfrak{p}(a_{\mathrm{grp},\,h})\}-E_{\mathfrak{q}}\{\log\mathfrak{q}^*(a_{\mathrm{grp},\,h})\}\\
& +E_{\mathfrak{q}}\{\log\mathfrak{p}(\boldsymbol{\Sigma}_h\,|\,\boldsymbol{A}_{\boldsymbol{\Sigma}_h})\}-E_{\mathfrak{q}}\{\log\mathfrak{q}^*(\boldsymbol{\Sigma}_h)\}+E_{\mathfrak{q}}\{\log\mathfrak{p}(\boldsymbol{A}_{\boldsymbol{\Sigma}_h})\}-E_{\mathfrak{q}}\{\log\mathfrak{q}^*(\boldsymbol{A}_{\boldsymbol{\Sigma}_h})\}.
\end{aligned}
\tag{S.4}
$$

First we note that

$$
\log\mathfrak{p}(\boldsymbol{y}\,|\,\boldsymbol{\beta},\boldsymbol{u},\sigma_\varepsilon^2)=-\frac{1}{2}\log(2\pi)\sum_{i=1}^{m}\sum_{j=1}^{n_i}o_{ij}-\frac{1}{2}\log(\sigma_\varepsilon^2)\sum_{i=1}^{m}\sum_{j=1}^{n_i}o_{ij}-\frac{1}{2\sigma_\varepsilon^2}\|\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta}-\boldsymbol{Z}\boldsymbol{u}\|^2
$$

where

$$
\begin{aligned}
&\|\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta}-\boldsymbol{Z}\boldsymbol{u}\|^2\\
&=\sum_{i=1}^{m}\sum_{j=1}^{n_i}\|\boldsymbol{y}_{ij}-\boldsymbol{X}_{ij}\boldsymbol{\beta}-\boldsymbol{Z}_{\mathrm{gbl},ij}\boldsymbol{u}_{\mathrm{gbl},i}-\boldsymbol{X}_{ij}\boldsymbol{u}^g_{\mathrm{lin},i}-\boldsymbol{Z}^g_{\mathrm{grp},ij}\boldsymbol{u}^g_{\mathrm{grp}}-\boldsymbol{X}_{ij}\boldsymbol{u}^h_{\mathrm{lin},ij}-\boldsymbol{Z}^h_{\mathrm{grp},ij}\boldsymbol{u}^h_{\mathrm{grp},ij}\|^2\\
&=\sum_{i=1}^{m}\sum_{j=1}^{n_i}\left\|\boldsymbol{y}_{ij}-\boldsymbol{C}_{\mathrm{gbl},ij}\left[\begin{array}{c}\boldsymbol{\beta}\\ \boldsymbol{u}_{\mathrm{gbl},i}\end{array}\right]-\boldsymbol{C}^g_{\mathrm{grp},ij}\left[\begin{array}{c}\boldsymbol{u}^g_{\mathrm{lin},i}\\ \boldsymbol{u}^g_{\mathrm{grp},i}\end{array}\right]-\boldsymbol{C}^h_{\mathrm{grp},ij}\left[\begin{array}{c}\boldsymbol{u}^h_{\mathrm{lin},ij}\\ \boldsymbol{u}^h_{\mathrm{grp},ij}\end{array}\right]\right\|^2
\end{aligned}
$$

and

$$
\boldsymbol{C}_{\mathrm{gbl},ij}\equiv[\boldsymbol{X}_{ij}\ \boldsymbol{Z}_{\mathrm{gbl},ij}],\quad \boldsymbol{C}^g_{\mathrm{grp},ij}\equiv[\boldsymbol{X}_{ij}\ \boldsymbol{Z}^g_{\mathrm{grp},ij}],\quad \boldsymbol{C}^h_{\mathrm{grp},ij}\equiv[\boldsymbol{X}_{ij}\ \boldsymbol{Z}^h_{\mathrm{grp},ij}].
$$

Therefore,

$$
\begin{aligned}
&E_{\mathfrak{q}}\{\log \mathfrak{p}(\boldsymbol{y}\,|\,\boldsymbol{\beta},\boldsymbol{u},\sigma^2_\varepsilon)\}\\
&\quad = -\tfrac{1}{2}\log(2\pi)\sum_{i=1}^m\sum_{j=1}^{n_i} o_{ij} - \frac{1}{2}E_{\mathfrak{q}}\{\log(\sigma^2_\varepsilon)\}\sum_{i=1}^m\sum_{j=1}^{n_i} o_{ij}\\
&\qquad -\tfrac{1}{2}\mu_{\mathfrak{q}(1/\sigma^2_\varepsilon)}\sum_{i=1}^m\sum_{j=1}^{n_i}\left\{\left\|E_{\mathfrak{q}}\left(\boldsymbol{y}_{ij} - \boldsymbol{C}_{\text{gbl},ij}\begin{bmatrix}\boldsymbol{\beta}\\ \boldsymbol{u}_{\text{gbl},i}\end{bmatrix} - \boldsymbol{C}^g_{\text{grp},ij}\begin{bmatrix}\boldsymbol{u}^g_{\text{lin},i}\\ \boldsymbol{u}^g_{\text{grp},i}\end{bmatrix} - \boldsymbol{C}^h_{\text{grp},ij}\begin{bmatrix}\boldsymbol{u}^h_{\text{lin},ij}\\ \boldsymbol{u}^h_{\text{grp},ij}\end{bmatrix}\right)\right\|^2\right.\\
&\qquad\quad +\text{tr}\left(\boldsymbol{C}^T_{\text{gbl},ij}\boldsymbol{C}_{\text{gbl},ij}\boldsymbol{\Sigma}_{\mathfrak{q}(\boldsymbol{\beta},\boldsymbol{u}_{\text{gbl}})}\right) + \text{tr}\left({\boldsymbol{C}^g_{\text{grp},ij}}^T\boldsymbol{C}^g_{\text{grp},ij}\boldsymbol{\Sigma}_{\mathfrak{q}(\boldsymbol{u}^g_{\text{lin},i},\boldsymbol{u}^g_{\text{grp},i})}\right)\\
&\qquad\quad +\text{tr}\left({\boldsymbol{C}^h_{\text{grp},ij}}^T\boldsymbol{C}^h_{\text{grp},ij}\boldsymbol{\Sigma}_{\mathfrak{q}(\boldsymbol{u}^h_{\text{lin},ij},\boldsymbol{u}^h_{\text{grp},ij})}\right)\\
&\qquad\quad +2\,\text{tr}\left[{\boldsymbol{C}^g_{\text{grp},ij}}^T\boldsymbol{C}_{\text{gbl},ij}\,E_{\mathfrak{q}}\left\{\left(\begin{bmatrix}\boldsymbol{\beta}\\ \boldsymbol{u}_{\text{gbl}}\end{bmatrix} - \boldsymbol{\mu}_{\mathfrak{q}(\boldsymbol{\beta},\boldsymbol{u}_{\text{gbl}})}\right)\left(\begin{bmatrix}\boldsymbol{u}^g_{\text{lin},i}\\ \boldsymbol{u}^g_{\text{grp}}\end{bmatrix} - \boldsymbol{\mu}_{\mathfrak{q}(\boldsymbol{u}^g_{\text{lin},i},\boldsymbol{u}^g_{\text{grp},i}))}\right)^T\right\}\right]\\
&\qquad\quad +2\,\text{tr}\left[{\boldsymbol{C}^h_{\text{grp},ij}}^T\boldsymbol{C}_{\text{gbl},ij}\,E_{\mathfrak{q}}\left\{\left(\begin{bmatrix}\boldsymbol{\beta}\\ \boldsymbol{u}_{\text{gbl}}\end{bmatrix} - \boldsymbol{\mu}_{\mathfrak{q}(\boldsymbol{\beta},\boldsymbol{u}_{\text{gbl}})}\right)\left(\begin{bmatrix}\boldsymbol{u}^h_{\text{lin},ij}\\ \boldsymbol{u}^h_{\text{grp}}\end{bmatrix} - \boldsymbol{\mu}_{\mathfrak{q}(\boldsymbol{u}^h_{\text{lin},ij},\boldsymbol{u}^h_{\text{grp},ij}))}\right)^T\right\}\right]\\
&\qquad\quad \left.+2\,\text{tr}\left[{\boldsymbol{C}^g_{\text{grp},ij}}^T\boldsymbol{C}^h_{\text{grp},ij}\,E_{\mathfrak{q}}\left\{\left(\begin{bmatrix}\boldsymbol{u}^g_{\text{lin},i}\\ \boldsymbol{u}^g_{\text{grp}}\end{bmatrix} - \boldsymbol{\mu}_{\mathfrak{q}(\boldsymbol{u}^g_{\text{lin},i},\boldsymbol{u}^g_{\text{grp},i})}\right)\left(\begin{bmatrix}\boldsymbol{u}^h_{\text{lin},ij}\\ \boldsymbol{u}^h_{\text{grp}}\end{bmatrix} - \boldsymbol{\mu}_{\mathfrak{q}(\boldsymbol{u}^h_{\text{lin},ij},\boldsymbol{u}^h_{\text{grp},ij}))}\right)^T\right\}\right]\right\}
\end{aligned}
$$

The remainder of the expectations in (S.4) are expressed as:

$$
\begin{aligned}
&E_{\mathfrak{q}}\{\log \mathfrak{p}(\boldsymbol{\beta},\boldsymbol{u}\,|\,\sigma^2_{\text{gbl}},\sigma^2_{\text{grp},g},\boldsymbol{\Sigma}_g,\sigma^2_{\text{grp},h},\boldsymbol{\Sigma}_h)\}\\
&= -\tfrac{1}{2}\{2 + K_{\text{gbl}} + m(2+K^g_{\text{grp}}) + (2+K^h_{\text{grp}})\sum_{i=1}^m n_i\}\log(2\pi) - \tfrac{1}{2}\log|\boldsymbol{\Sigma}_{\boldsymbol{\beta}}| - \tfrac{1}{2}K_{\text{gbl}}E_{\mathfrak{q}}\{\log(\sigma^2_{\text{gbl}})\}\\
&\quad -\tfrac{1}{2}m\left(E_{\mathfrak{q}}\{\log|\boldsymbol{\Sigma}_g|\} - K^g_{\text{grp}}E_{\mathfrak{q}}\{\log(\sigma^2_{\text{grp},g})\}\right) - \tfrac{1}{2}\left(E_{\mathfrak{q}}\{\log|\boldsymbol{\Sigma}_h|\} - K^h_{\text{grp}}E_{\mathfrak{q}}\{\log(\sigma^2_{\text{grp},h})\}\right)\sum_{i=1}^m n_i\\
&\quad -\tfrac{1}{2}\mu_{\mathfrak{q}(1/\sigma^2_{\text{gbl}})}\left\{\|\boldsymbol{\mu}_{\mathfrak{q}(\boldsymbol{u}_{\text{gbl}})}\|^2 + \text{tr}(\boldsymbol{\Sigma}_{\mathfrak{q}(\boldsymbol{u}_{\text{gbl}})})\right\} - \tfrac{1}{2}\mu_{\mathfrak{q}(1/\sigma^2_{\text{grp},g})}\sum_{i=1}^m\left\{\|\boldsymbol{\mu}_{\mathfrak{q}(\boldsymbol{u}^g_{\text{grp},i}))}\|^2 + \text{tr}(\boldsymbol{\Sigma}_{\mathfrak{q}(\boldsymbol{u}^g_{\text{grp},i})})\right\}\\
&\quad -\tfrac{1}{2}\mu_{\mathfrak{q}(1/\sigma^2_{\text{grp},h})}\sum_{i=1}^m\sum_{j=1}^{n_i}\left\{\|\boldsymbol{\mu}_{\mathfrak{q}(\boldsymbol{u}^h_{\text{grp},ij}))}\|^2 + \text{tr}(\boldsymbol{\Sigma}_{\mathfrak{q}(\boldsymbol{u}^h_{\text{grp},ij})})\right\}\\
&\quad -\tfrac{1}{2}\text{tr}\left(\boldsymbol{\Sigma}^{-1}_{\boldsymbol{\beta}}\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}_g)}\left\{\sum_{i=1}^m\left(\boldsymbol{\mu}_{\mathfrak{q}(\boldsymbol{u}^g_{\text{lin},i})}\boldsymbol{\mu}^T_{\mathfrak{q}(\boldsymbol{u}^g_{\text{lin},i})} + \boldsymbol{\Sigma}_{\mathfrak{q}(\boldsymbol{u}^g_{\text{lin},i})}\right)\right\}\right)\\
&\quad -\tfrac{1}{2}\text{tr}\left(\boldsymbol{M}_{\mathfrak{q}(\boldsymbol{\Sigma}^{-1}_h)}\left\{\sum_{i=1}^m\sum_{j=1}^{n_i}\left(\boldsymbol{\mu}_{\mathfrak{q}(\boldsymbol{u}^h_{\text{lin},ij})}\boldsymbol{\mu}^T_{\mathfrak{q}(\boldsymbol{u}^h_{\text{lin},ij})} + \boldsymbol{\Sigma}_{\mathfrak{q}(\boldsymbol{u}^h_{\text{lin},ij})}\right)\right\}\right),
\end{aligned}
$$

$$
\begin{aligned}
E_{\mathfrak{q}}\{\log \mathfrak{q}^*(\boldsymbol{\beta},\boldsymbol{u})\} \;=\; & -\tfrac{1}{2}\left\{2 + K_{\text{gbl}} + m(2+K^g_{\text{grp}}) + (2+K^h_{\text{grp}})\sum_{i=1}^m n_i\right\}\{1+\log(2\pi)\}\\
& -\tfrac{1}{2}\log|\boldsymbol{\Sigma}_{\mathfrak{q}(\boldsymbol{\beta},\boldsymbol{u})}|,
\end{aligned}
$$

Expressions for $E_{\mathfrak{q}}\{\log \mathfrak{p}(\sigma^2_\varepsilon\,|\,a_\varepsilon)\}$, $E_{\mathfrak{q}}\{\log \mathfrak{q}^*(\sigma^2_\varepsilon)\}$, $E_{\mathfrak{q}}\{\log \mathfrak{p}(a_\varepsilon)\}$, $E_{\mathfrak{q}}\{\log \mathfrak{q}^*(a_\varepsilon)\}$, $E_{\mathfrak{q}}\{\log \mathfrak{p}(\sigma^2_{\text{gbl}}\,|\,a_{\text{gbl}})\}$, $E_{\mathfrak{q}}\{\log \mathfrak{q}^*(\sigma^2_{\text{gbl}})\}$, $E_{\mathfrak{q}}\{\log \mathfrak{p}(a_{\text{gbl}})\}$ and $E_{\mathfrak{q}}\{\log \mathfrak{q}^*(a_{\text{gbl}})\}$ follow the same form as shown from the derivations in the two-level case.

Following on, we have

$$
\begin{aligned}
E_{\mathfrak{q}}\{\log \mathfrak{p}(\sigma^2_{\text{grp},g}\,|\,a_{\text{grp},\,g})\} \;=\; & -\tfrac{1}{2}\nu_{\text{grp},\,g}E_{\mathfrak{q}}\{\log(2a_{\text{grp},\,g})\} - \log\Gamma(\nu_{\text{grp},\,g}/2) - (\tfrac{1}{2}\nu_{\text{grp},\,g}+1)E_{\mathfrak{q}}\{\log(\sigma^2_{\text{grp},g})\}\\
& -\tfrac{1}{2}\mu_{\mathfrak{q}(1/a_{\text{grp},\,g})}\mu_{\mathfrak{q}(1/\sigma^2_{\text{grp},g})},
\end{aligned}
$$

$$
\begin{aligned}
E_{\mathfrak{q}}\{\log \mathfrak{q}^*(\sigma^2_{\text{grp},g})\} &= \tfrac{1}{2}\xi_{\mathfrak{q}(\sigma^2_{\text{grp},g})}\log(\lambda_{\mathfrak{q}(\sigma^2_{\text{grp},g})}/2) - \log\{\Gamma(\tfrac{1}{2}\xi_{\mathfrak{q}(\sigma^2_{\text{grp},g})})\} - (\tfrac{1}{2}\xi_{\mathfrak{q}(\sigma^2_{\text{grp},g})}+1)E_{\mathfrak{q}}\{\log(\sigma^2_{\text{grp},g})\}\\
&\quad -\tfrac{1}{2}\lambda_{\mathfrak{q}(\sigma^2_{\text{grp},g})}\mu_{\mathfrak{q}(1/\sigma^2_{\text{grp},g})},\\
E_{\mathfrak{q}}\{\log \mathfrak{p}(a_{\text{grp},\,g})\} &= -\tfrac{1}{2}\log(2\nu_{\text{grp},\,g}s^2_{\text{grp},\,g}) - \log\{\Gamma(\tfrac{1}{2})\} - (\tfrac{1}{2}+1)E_{\mathfrak{q}}\{\log(a_{\text{grp},\,g})\}\\
&\quad -\{1/(2\nu_{\text{grp},\,g}s^2_{\text{grp},\,g})\}\mu_{\mathfrak{q}(1/a_{\text{grp},\,g})},\\
E_{\mathfrak{q}}\{\log \mathfrak{q}^*(a_{\text{grp},\,g})\} &= \tfrac{1}{2}\xi_{\mathfrak{q}(a_{\text{grp},\,g})}\log(\lambda_{\mathfrak{q}(a_{\text{grp},\,g})}/2) - \log\{\Gamma(\tfrac{1}{2}\xi_{\mathfrak{q}(a_{\text{grp},\,g})})\} - (\tfrac{1}{2}\xi_{\mathfrak{q}(a_{\text{grp},\,g})}+1)E_{\mathfrak{q}}\{\log(a_{\text{grp},\,g})\}\\
&\quad -\tfrac{1}{2}\lambda_{\mathfrak{q}(a_{\text{grp},\,g})}\mu_{\mathfrak{q}(1/a_{\text{grp},\,g})},\\
E_{\mathfrak{q}}[\log\{\mathfrak{p}(\boldsymbol{\Sigma}_g|\boldsymbol{A}_{\boldsymbol{\Sigma}_g})\}] &= -\tfrac{1}{2}(\nu_{\boldsymbol{\Sigma}_g}+1)E_{\mathfrak{q}}\{\log|\boldsymbol{A}_{\boldsymbol{\Sigma}_g}|\} - \tfrac{1}{2}(\nu_{\boldsymbol{\Sigma}_g}+4)E_{\mathfrak{q}}\{\log|\boldsymbol{\Sigma}_g|\} - \tfrac{1}{2}\log(\pi)\\
&\quad -\tfrac{1}{2}\text{tr}(\boldsymbol{M}_{\mathfrak{q}(\boldsymbol{A}^{-1}_{\boldsymbol{\Sigma}_g})}\boldsymbol{M}_{\mathfrak{q}(\boldsymbol{\Sigma}_g^{-1})}) - (\nu_{\boldsymbol{\Sigma}_g}+3)\log(2) - \textstyle\sum_{j=1}^2\log\Gamma(\tfrac{1}{2}(\nu_{\boldsymbol{\Sigma}_g}+4-j)),\\
E_{\mathfrak{q}}[\log\{\mathfrak{q}^*(\boldsymbol{\Sigma}_g)\}] &= \tfrac{1}{2}(\xi_{\mathfrak{q}(\boldsymbol{\Sigma}_g)}-1)\log|\boldsymbol{\Lambda}_{\mathfrak{q}(\boldsymbol{\Sigma}_g)}| - \tfrac{1}{2}(\xi_{\mathfrak{q}(\boldsymbol{\Sigma}_g)}+2)E_{\mathfrak{q}}\{\log|\boldsymbol{\Sigma}_g|\} - \tfrac{1}{2}\text{tr}(\boldsymbol{\Lambda}_{\mathfrak{q}(\boldsymbol{\Sigma}_g)}\boldsymbol{M}_{\mathfrak{q}(\boldsymbol{\Sigma}_g^{-1})})\\
&\quad -(\xi_{\mathfrak{q}(\boldsymbol{\Sigma}_g)}+1)\log(2) - \tfrac{1}{2}\log(\pi) - \textstyle\sum_{j=1}^2\log\Gamma(\tfrac{1}{2}(\xi_{\mathfrak{q}(\boldsymbol{\Sigma}_g)}+2-j)),\\
E_{\mathfrak{q}}[\log\{\mathfrak{p}(\boldsymbol{A}_{\boldsymbol{\Sigma}_g})\}] &= -\tfrac{3}{2}E_{\mathfrak{q}}\{\log|\boldsymbol{A}_{\boldsymbol{\Sigma}_g}|\} - \tfrac{1}{2}\textstyle\sum_{j=1}^2 1/(\nu_{\boldsymbol{\Sigma}_g}s^2_{\boldsymbol{\Sigma}_g,\,j})\left(\boldsymbol{M}_{\mathfrak{q}(\boldsymbol{A}^{-1}_{\boldsymbol{\Sigma}_g})}\right)_{jj} - 2\log(2)\\
&\quad -\tfrac{1}{2}\log(\pi) - \textstyle\sum_{j=1}^2\log\Gamma(\tfrac{1}{2}(3-j)),
\end{aligned}
$$

and

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

In addition, the expressions for $E_{\mathfrak{q}}\{\log\mathfrak{p}(\sigma^2_{\text{grp},h}\,|\,a_{\text{grp},\,h})\}$, $E_{\mathfrak{q}}\{\log\mathfrak{q}^*(\sigma^2_{\text{grp},h})\}$, $E_{\mathfrak{q}}\{\log\mathfrak{p}(a_{\text{grp},\,h})\}$, $E_{\mathfrak{q}}\{\log\mathfrak{q}^*(a_{\text{grp},\,h})\}$, $E_{\mathfrak{q}}[\log\{\mathfrak{p}(\boldsymbol{\Sigma}|\boldsymbol{A}_{\boldsymbol{\Sigma}_h})\}]$, $E_{\mathfrak{q}}[\log\{\mathfrak{q}(\boldsymbol{\Sigma}_h)\}]$, $E_{\mathfrak{q}}[\log\{\mathfrak{p}(\boldsymbol{A}_{\boldsymbol{\Sigma}_h})\}]$, $E_{\mathfrak{q}}[\log\{\mathfrak{q}(\boldsymbol{A}_{\boldsymbol{\Sigma}_h})\}]$, are similar to the ones shown above.

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

$$
-\tfrac{1}{2}\sum_{i=1}^m\sum_{j=1}^{n_i}o_{ij} - \tfrac{1}{2}\nu_\varepsilon - 1 + \tfrac{1}{2}\xi_{\mathfrak{q}(\sigma^2_\varepsilon)} + 1 = -\tfrac{1}{2}\sum_{i=1}^m\sum_{j=1}^{n_i}o_{ij} - \tfrac{1}{2}\nu_\varepsilon - 1 + \tfrac{1}{2}(\nu_\varepsilon + \sum_{i=1}^m\sum_{j=1}^{n_i}o_{ij}) + 1 = 0.
$$

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

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

The coefficient of $E_{\mathfrak{q}}\{\log(\sigma^2_{\text{grp},g})\}$ is

$$
-\tfrac{1}{2}mK^g_{\text{grp}} - \tfrac{1}{2}\nu_{\text{grp},\,g} - 1 + \tfrac{1}{2}\xi_{\mathfrak{q}(\sigma^2_{\text{grp},g})} + 1 = -\tfrac{1}{2}mK^g_{\text{grp}} - \tfrac{1}{2}\nu_{\text{grp},\,g} - 1 + \tfrac{1}{2}(\nu_{\text{grp},\,g} + mK^g_{\text{grp}}) + 1 = 0.
$$

The coefficient of $E_{\mathfrak{q}}\{\log(\sigma^2_{\text{grp},h})\}$ is

$$
-\tfrac{1}{2}K^h_{\text{grp}}\sum_{i=1}^m n_i - \tfrac{1}{2}\nu_{\text{grp},\,h} - 1 + \tfrac{1}{2}\xi_{\mathfrak{q}(\sigma^2_{\text{grp},h})} + 1 = -\tfrac{1}{2}K^h_{\text{grp}}\sum_{i=1}^m n_i - \tfrac{1}{2}\nu_{\text{grp},\,h} - 1 + \tfrac{1}{2}\left(\nu_{\text{grp},\,h} + K^h_{\text{grp}}\sum_{i=1}^m n_i\right) + 1 = 0.
$$

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

$$
-\frac{m}{2} - \tfrac{1}{2}(\nu_{\boldsymbol{\Sigma}_g}+4) + \tfrac{1}{2}(\xi_{\mathfrak{q}(\boldsymbol{\Sigma}_g)}+2) = -\tfrac{1}{2}(m + \nu_{\boldsymbol{\Sigma}_g} + 4) + \tfrac{1}{2}(m + \nu_{\boldsymbol{\Sigma}_g} + 4) = 0.
$$

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

$$
-\tfrac{1}{2}\sum_{i=1}^m n_i - \tfrac{1}{2}(\nu_{\boldsymbol{\Sigma}_h}+4) + \tfrac{1}{2}(\xi_{\mathfrak{q}(\boldsymbol{\Sigma}_h)}+2) = -\tfrac{1}{2}\left(\sum_{i=1}^m n_i + \nu_{\boldsymbol{\Sigma}_h} + 4\right) + \tfrac{1}{2}\left(\sum_{i=1}^m n_i + \nu_{\boldsymbol{\Sigma}_h} + 4\right) = 0.
$$

The coefficient of $E_{\mathfrak{q}}\{\log(a_\varepsilon)\}$ is

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

The coefficient of $E_{\mathfrak{q}}\{\log(a_{\text{gbl}})\}$ is

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

The coefficient of $E_{\mathfrak{q}}\{\log(a_{\text{grp},\,g})\}$ is

$$-\tfrac{1}{2}\nu_{\text{grp},\,g} - \tfrac{1}{2} - 1 + \tfrac{1}{2}\xi_{\mathfrak{q}(a_{\text{grp},\,g})} + 1 = -\tfrac{1}{2}\nu_{\text{grp},\,g} - \tfrac{1}{2} - 1 + \tfrac{1}{2}(\nu_{\text{grp},\,g} + 1) + 1 = 0.$$

The coefficient of $E_{\mathfrak{q}}\{\log(a_{\text{grp},\,h})\}$ is

$$-\tfrac{1}{2}\nu_{\text{grp},\,h} - \tfrac{1}{2} - 1 + \tfrac{1}{2}\xi_{\mathfrak{q}(a_{\text{grp},\,h})} + 1 = -\tfrac{1}{2}\nu_{\text{grp},\,h} - \tfrac{1}{2} - 1 + \tfrac{1}{2}(\nu_{\text{grp},\,h} + 1) + 1 = 0.$$

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

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

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

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

Therefore these terms can be dropped and the cancellations led by the above expectations lead to the lower bound expression in (S.3).

S.12 The SolveTwoLevelSparseLeastSquares Algorithm

The SolveTwoLevelSparseLeastSquares is listed in Nolan *et al.* (2018) and based on Theorem 2 of Nolan & Wand (2018). Given its centrality to Algorithms 1 and 2 we list it again here. The algorithm solves a sparse version of the the least squares problem:

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

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:

$$\boldsymbol{B} \equiv \left[\begin{array}{c|c|c|c|c} \boldsymbol{B}_1 & \overset{\bullet}{\boldsymbol{B}}_1 & \boldsymbol{O} & \cdots & \boldsymbol{O} \\ \hline \boldsymbol{B}_2 & \boldsymbol{O} & \overset{\bullet}{\boldsymbol{B}}_2 & \cdots & \boldsymbol{O} \\ \hline \vdots & \vdots & \vdots & \ddots & \vdots \\ \hline \boldsymbol{B}_m & \boldsymbol{O} & \boldsymbol{O} & \cdots & \overset{\bullet}{\boldsymbol{B}}_m \end{array}\right] \quad \text{and} \quad \boldsymbol{b} = \left[\begin{array}{c} \boldsymbol{b}_1 \\ \hline \boldsymbol{b}_2 \\ \hline \vdots \\ \hline \boldsymbol{b}_m \end{array}\right]. \tag{S.5}$$

The sub-matrices corresponding to the non-zero blocks of $\boldsymbol{x}$ and $\boldsymbol{A}$ are labelled according to:

$$\boldsymbol{x} \equiv \left[\begin{array}{c} \boldsymbol{x}_1 \\ \hline \boldsymbol{x}_{2,1} \\ \hline \boldsymbol{x}_{2,2} \\ \hline \vdots \\ \hline \boldsymbol{x}_{2,m} \end{array}\right] \quad \text{and} \quad \boldsymbol{A}^{-1} = \left[\begin{array}{c|c|c|c|c} \boldsymbol{A}^{11} & \boldsymbol{A}^{12,1} & \boldsymbol{A}^{12,2} & \cdots & \boldsymbol{A}^{12,m} \\ \hline \boldsymbol{A}^{12,1\,T} & \boldsymbol{A}^{22,1} & \times & \cdots & \times \\ \hline \boldsymbol{A}^{12,2\,T} & \times & \boldsymbol{A}^{22,2} & \cdots & \times \\ \hline \vdots & \vdots & \vdots & \ddots & \vdots \\ \hline \boldsymbol{A}^{12,m\,T} & \times & \times & \cdots & \boldsymbol{A}^{22,m} \end{array}\right] \tag{S.6}$$

with $\times$ denoting sub-blocks that are not of interest. The SolveTwoLevelSparseLeastSquares algorithm is given in Algorithm S.1.

Algorithm S.1 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.5) and (S.6).*

Inputs: $\left\{\left(\boldsymbol{b}_i(\tilde{n}_i \times 1),\ \boldsymbol{B}_i(\tilde{n}_i \times p),\ \dot{\boldsymbol{B}}_i(\tilde{n}_i \times q)\right):\ 1 \le i \le m\right\}$

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

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

Decompose $\dot{\boldsymbol{B}}_i = \boldsymbol{Q}_i \begin{bmatrix} \boldsymbol{R}_i \\ \boldsymbol{0} \end{bmatrix}$ 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: $\left(\boldsymbol{x}_1, \boldsymbol{A}^{11}, \left\{\left(\boldsymbol{x}_{2,i}, \boldsymbol{A}^{22,i}, \boldsymbol{A}^{12,i}\right):\ 1 \le i \le m\right\}\right)$

S.13 The SolveThreeLevelSparseLeastSquares Algorithm

The SolveThreeLevelSparseLeastSquares, listed in Nolan *et al.* (2018) is a proceduralization of Theorem 4 of Nolan & Wand (2018). Since it is central to Algorithms 3 and 4 we list it here. The SolveThreeLevelSparseLeastSquares algorithm is concerned with solving the sparse three-level version of

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

with the 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:

$$\boldsymbol{B} \equiv \Big[\underset{1\le i\le m}{\text{stack}} \Big\{ \underset{1\le j\le n_i}{\text{stack}} (\boldsymbol{B}_{ij}) \Big\} \Big| \underset{1\le i\le m}{\text{blockdiag}} \Big\{ \big[\underset{1\le j\le n_i}{\text{stack}} (\dot{\boldsymbol{B}}_{ij}) \,\big|\, \underset{1\le j\le n_i}{\text{blockdiag}} (\ddot{\boldsymbol{B}}_{ij}) \big] \Big\} \Big] \tag{S.7}$$

and

$$\boldsymbol{b} \equiv \underset{1\le i\le m}{\text{stack}} \Big\{ \underset{1\le j\le n_i}{\text{stack}} (\boldsymbol{b}_{ij}) \Big\}. \tag{S.8}$$

The three-level sparse matrix inverse problem involves determination of all sub-blocks of $\boldsymbol{x}$ and the sub-blocks of $\boldsymbol{A}^{-1}$ corresponding to the non-zero sub-blocks of $\boldsymbol{A}$. Our notation for these sub-blocks is illustrated by

$$\boldsymbol{x} = \left[\begin{array}{c} \boldsymbol{x}_1 \\ \hline \boldsymbol{x}_{2,1} \\ \hline \boldsymbol{x}_{2,11} \\ \hline \boldsymbol{x}_{2,12} \\ \hline \boldsymbol{x}_{2,2} \\ \hline \boldsymbol{x}_{2,21} \\ \hline \boldsymbol{x}_{2,22} \\ \hline \boldsymbol{x}_{2,23} \end{array}\right]$$

$$\text{and } \boldsymbol{A}^{-1} = \left[\begin{array}{c|c|c|c|c|c|c|c}
\boldsymbol{A}^{11} & \boldsymbol{A}^{12,1} & \boldsymbol{A}^{12,11} & \boldsymbol{A}^{12,12} & \boldsymbol{A}^{12,2} & \boldsymbol{A}^{12,21} & \boldsymbol{A}^{12,22} & \boldsymbol{A}^{12,23} \\ \hline
\boldsymbol{A}^{12,1\,T} & \boldsymbol{A}^{22,1} & \boldsymbol{A}^{12,1,1} & \boldsymbol{A}^{12,1,2} & \times & \times & \times & \times \\ \hline
\boldsymbol{A}^{12,11\,T} & \boldsymbol{A}^{12,1,1\,T} & \boldsymbol{A}^{22,11} & \times & \times & \times & \times & \times \\ \hline
\boldsymbol{A}^{12,12\,T} & \boldsymbol{A}^{12,1,2\,T} & \times & \boldsymbol{A}^{22,12} & \times & \times & \times & \times \\ \hline
\boldsymbol{A}^{12,2\,T} & \times & \times & \times & \boldsymbol{A}^{22,2} & \boldsymbol{A}^{12,2,1} & \boldsymbol{A}^{12,2,2} & \boldsymbol{A}^{12,2,3} \\ \hline
\boldsymbol{A}^{12,21\,T} & \times & \times & \times & \boldsymbol{A}^{12,2,1\,T} & \boldsymbol{A}^{22,21} & \times & \times \\ \hline
\boldsymbol{A}^{12,22\,T} & \times & \times & \times & \boldsymbol{A}^{12,2,2\,T} & \times & \boldsymbol{A}^{22,22} & \times \\ \hline
\boldsymbol{A}^{12,23\,T} & \times & \times & \times & \boldsymbol{A}^{12,2,3\,T} & \times & \times & \boldsymbol{A}^{22,23}
\end{array}\right] \tag{S.9}$$

for the $m = 2$, $n_1 = 2$ and $n_2 = 3$ case. The $\times$ symbol denotes sub-blocks that are not of interest. The SolveThreeLevelSparseLeastSquares algorithm is given in Algorithm S.2.

Algorithm S.2 SolveThreeLevelSparseLeastSquares *for solving the three-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.7), (S.8) and (S.9).*

Inputs: $\left\{\left(\boldsymbol{b}_{ij}(\tilde{o}_{ij} \times 1),\ \boldsymbol{B}_{ij}(\tilde{o}_{ij} \times p),\ \dot{\boldsymbol{B}}_{ij}(\tilde{o}_{ij} \times q_1),\ \ddot{\boldsymbol{B}}_{ij}(\tilde{o}_{ij} \times q_2)\right) :\ 1 \le i \le m,\ 1 \le j \le n_i\right\}$

$\boldsymbol{\omega}_7 \longleftarrow \text{NULL}$; $\boldsymbol{\Omega}_8 \longleftarrow \text{NULL}$

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

$\boldsymbol{\omega}_9 \longleftarrow \text{NULL}$; $\boldsymbol{\Omega}_{10} \longleftarrow \text{NULL}$; $\boldsymbol{\Omega}_{11} \longleftarrow \text{NULL}$

For $j = 1, \ldots, n_i$:

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

$\boldsymbol{d}_{0ij} \longleftarrow \boldsymbol{Q}_{ij}^T\boldsymbol{b}_{ij}$; $\boldsymbol{D}_{0ij} \longleftarrow \boldsymbol{Q}_{ij}^T\boldsymbol{B}_{ij}$; $\dot{\boldsymbol{D}}_{0ij} \longleftarrow \boldsymbol{Q}_{ij}^T\dot{\boldsymbol{B}}_{ij}$

$\boldsymbol{d}_{1ij} \longleftarrow$ 1st q_2 rows of $\boldsymbol{d}_{0ij}$; $\boldsymbol{d}_{2ij} \longleftarrow$ remaining rows of $\boldsymbol{d}_{0ij}$; $\boldsymbol{\omega}_9 \longleftarrow \begin{bmatrix}\boldsymbol{\omega}_9\\ \boldsymbol{d}_{2ij}\end{bmatrix}$

$\boldsymbol{D}_{1ij} \longleftarrow$ 1st q_2 rows of $\boldsymbol{D}_{0ij}$; $\boldsymbol{D}_{2ij} \longleftarrow$ remaining rows of $\boldsymbol{D}_{0ij}$; $\boldsymbol{\Omega}_{10} \longleftarrow \begin{bmatrix}\boldsymbol{\Omega}_{10}\\ \boldsymbol{D}_{2ij}\end{bmatrix}$

$\dot{\boldsymbol{D}}_{1ij} \longleftarrow$ 1st q_2 rows of $\dot{\boldsymbol{D}}_{0ij}$; $\dot{\boldsymbol{D}}_{2ij} \longleftarrow$ remaining rows of $\dot{\boldsymbol{D}}_{0ij}$; $\boldsymbol{\Omega}_{11} \longleftarrow \begin{bmatrix}\boldsymbol{\Omega}_{11}\\ \dot{\boldsymbol{D}}_{2ij}\end{bmatrix}$

Decompose $\boldsymbol{\Omega}_{11} = \boldsymbol{Q}_i\begin{bmatrix}\boldsymbol{R}_i\\ \boldsymbol{0}\end{bmatrix}$ such that $\boldsymbol{Q}_i^{-1} = \boldsymbol{Q}_i^T$ and $\boldsymbol{R}_i$ is upper-triangular.

$\boldsymbol{c}_{0i} \longleftarrow \boldsymbol{Q}_i^T\boldsymbol{\omega}_9$; $\boldsymbol{C}_{0i} \longleftarrow \boldsymbol{Q}_i^T\boldsymbol{\Omega}_{10}$

$\boldsymbol{c}_{1i} \longleftarrow$ 1st q_1 rows of $\boldsymbol{c}_{0i}$; $\boldsymbol{c}_{2i} \longleftarrow$ remaining rows of $\boldsymbol{c}_{0i}$; $\boldsymbol{\omega}_7 \longleftarrow \begin{bmatrix}\boldsymbol{\omega}_7\\ \boldsymbol{c}_{2i}\end{bmatrix}$

$\boldsymbol{C}_{1i} \longleftarrow$ 1st q_1 rows of $\boldsymbol{C}_{0i}$; $\boldsymbol{C}_{2i} \longleftarrow$ remaining rows of $\boldsymbol{C}_{0i}$; $\boldsymbol{\Omega}_8 \longleftarrow \begin{bmatrix}\boldsymbol{\Omega}_8\\ \boldsymbol{C}_{2i}\end{bmatrix}$

Decompose $\boldsymbol{\Omega}_8 = \boldsymbol{Q}\begin{bmatrix}\boldsymbol{R}\\ \boldsymbol{0}\end{bmatrix}$ so that $\boldsymbol{Q}^{-1} = \boldsymbol{Q}^T$ and $\boldsymbol{R}$ is upper-triangular.

$\boldsymbol{c} \longleftarrow$ first p rows of $\boldsymbol{Q}^T\boldsymbol{\omega}_7$; $\boldsymbol{x}_1 \longleftarrow \boldsymbol{R}^{-1}\boldsymbol{c}$; $\boldsymbol{A}^{11} \longleftarrow \boldsymbol{R}^{-1}\boldsymbol{R}^{-T}$

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

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

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

For $j = 1, \ldots, n_i$:

$\boldsymbol{x}_{2,ij} \leftarrow \boldsymbol{R}_{ij}^{-1}(\boldsymbol{d}_{1ij} - \boldsymbol{D}_{1ij}\boldsymbol{x}_1 - \dot{\boldsymbol{D}}_{1ij}\boldsymbol{x}_{2,i})$

$\boldsymbol{A}^{12,ij} \leftarrow -\left\{\boldsymbol{R}_{ij}^{-1}(\boldsymbol{D}_{1ij}\boldsymbol{A}^{11} + \dot{\boldsymbol{D}}_{1ij}\boldsymbol{A}^{12,i\,T})\right\}^T$

$\boldsymbol{A}^{12,\,i,\,j} \leftarrow -\left\{\boldsymbol{R}_{ij}^{-1}(\boldsymbol{D}_{1ij}\boldsymbol{A}^{12,i} + \dot{\boldsymbol{D}}_{1ij}\boldsymbol{A}^{22,i})\right\}^T$

$\boldsymbol{A}^{22,ij} \leftarrow \boldsymbol{R}_{ij}^{-1}\left(\boldsymbol{R}_{ij}^{-T} - \boldsymbol{D}_{1ij}\boldsymbol{A}^{12,ij} - \dot{\boldsymbol{D}}_{1ij}\boldsymbol{A}^{12,\,i,\,j}\right)$

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