Statistical Modelling xxxx; **xx(x)**: 1–41

Streamlined variational inference for higher level group-specific curve models

M. Menictas[1], T.H. Nolan[1,3], D.G. Simpson[2] and M.P. Wand[1,3]

[1]School of Mathematical and Physical Sciences, University of Technology Sydney, Ultimo, Australia.
[2]Department of Statistics, University of Illinois at Urbana-Champaign, Champaign, Illinois, United States of America.
[3]Australian Research Council Centre of Excellence for Mathematical and Statistical Frontiers, The University of Melbourne, Parkville, Australia.

Abstract: A two-level group-specific curve model is such that the mean response of each member of a group is a separate smooth function of a predictor of interest. The three-level extension is such that one grouping variable is nested within another one, and higher level extensions are analogous. Streamlined variational inference for higher level group-specific curve models is a challenging problem. We confront it by systematically working through two-level and then three-level cases and making use of the higher level sparse matrix infrastructure laid down in (Nolan and Wand, 2020, *ANZIAM Journal*, doi: 10.1017/S1446181120000061). A motivation is analysis of data from ultrasound technology for which three-level group-specific curve models are appropriate. Whilst extension to the number of levels exceeding three is not covered explicitly, the pattern established by our systematic approach sheds light on what is required for even higher level group-specific curve models.

Key words: approximate Bayesian inference, longitudinal data analysis, multilevel models, panel data, mean field variational Bayes

Received March 2019; **revised** March 2020; **accepted** April 2020

1 Introduction

We provide explicit algorithms for fitting and approximate Bayesian inference for multilevel models involving, potentially, thousands of noisy curves. The algorithms include covariance parameter estimation and allow for pointwise credible intervals around the fitted curves. Contrast function fitting and inference is also supported by our approach. Both two-level and three-level situations are covered, and a template for even higher level situations is laid down.

Models and methodology for statistical analyses of grouped data for which the basic unit is a noisy curve continues to be an important area of research. A

Address for correspondence: M.P. Wand, School of Mathematical and Physical Sciences, University of Technology Sydney, P.O. Box 123, Broadway 2007, Australia.
E-mail: matt.wand@uts.edu.au

 10.1177/1471082X20930894

driving force is rapid technological change which is resulting in the generation of curve-type data at fine resolution levels. Examples of such technology include accelerometers (e.g., Goldsmith et al., 2015) personal digital assistants (e.g., Trail et al., 2014) and quantitative ultrasound (e.g., Wirtzfeld et al., 2015). In some applications, curve-type data have higher levels of grouping, with groups at one level nested inside other groups. Our focus here is streamlined variational inference for such circumstances.

Some motivating data is shown in Figure 1 from an experiment involving quantitative ultrasound technology (Wirtzfeld et al., 2015). Each curve corresponds to a logarithmically transformed backscatter coefficient, with the backscatter coefficient measurements in units centimeters^{-1}steradian^{-1}, over a fine grid of frequency values for tumours in laboratory mice, with exactly 1 tumour per mouse. Each backscatter/frequency curve corresponds to one of five slices of the same tumour, where slices correspond to scan lines at different probe locations. The slices are grouped according to being from one of 10 tumours. We refer to such data as three-level data with frequency measurements at level 1, slices being the level 2 groups and tumours constituting the level 3 groups. The gist of this article is efficient and flexible variational fitting and inference for such data that scales well to much larger multilevel datasets. Indeed, our algorithms are linear in the number of groups at both levels 2 and 3. Simulation study results given later in this article show that curve-type data with thousands of groups can be analysed quickly using our new methodology. Depending on sample sizes and implementation language, fitting times range from a few seconds to several minutes. In contrast, naïve implementations become infeasible when the number of groups are in the several hundreds due to storage and computational demands.

We work with a variant of group-specific curve models that at least go back to Donnelly et al. (2015). Other contributions of this type include Brumback and Rice (1998), Verbyla et al. (1999), Wang (1998) and Zhang et al. (1998). The specific formulation that we use is that given by Durban et al. (2005) which involves an embedding within the class of linear mixed models (Robinson, 1991, e.g) with low-rank smoothing splines used for flexible function modelling and fitting. More recent contributions in this area such as Djeundje and Currie (2010), Heckman et al. (2013) and Djeundje (2016) point out potential pitfalls and provide remedies for group-specific curve models and other models involving mixed model-based spline penalization. For example, Heckman et al. (2013) advocate subject-area modelling of covariance structures and the use of sandwich-type standard errors, whereas Djeundje (2016) advocates the use of centring constraints. Whilst our streamlined algorithms for higher level group-specific curve models and variability bands assume the basic form corresponding to the first set of references in this paragraph, embellishments of the type proposed in the more recent literature could be considered depending on the nature of the data to which the methodology is applied.

We consider both frequentist and Bayesian group-specific curve models. The use of Bayesian versions, with variational approximations, is in keeping with the emerging variational Bayesian inference paradigm in statistical and machine learning contexts

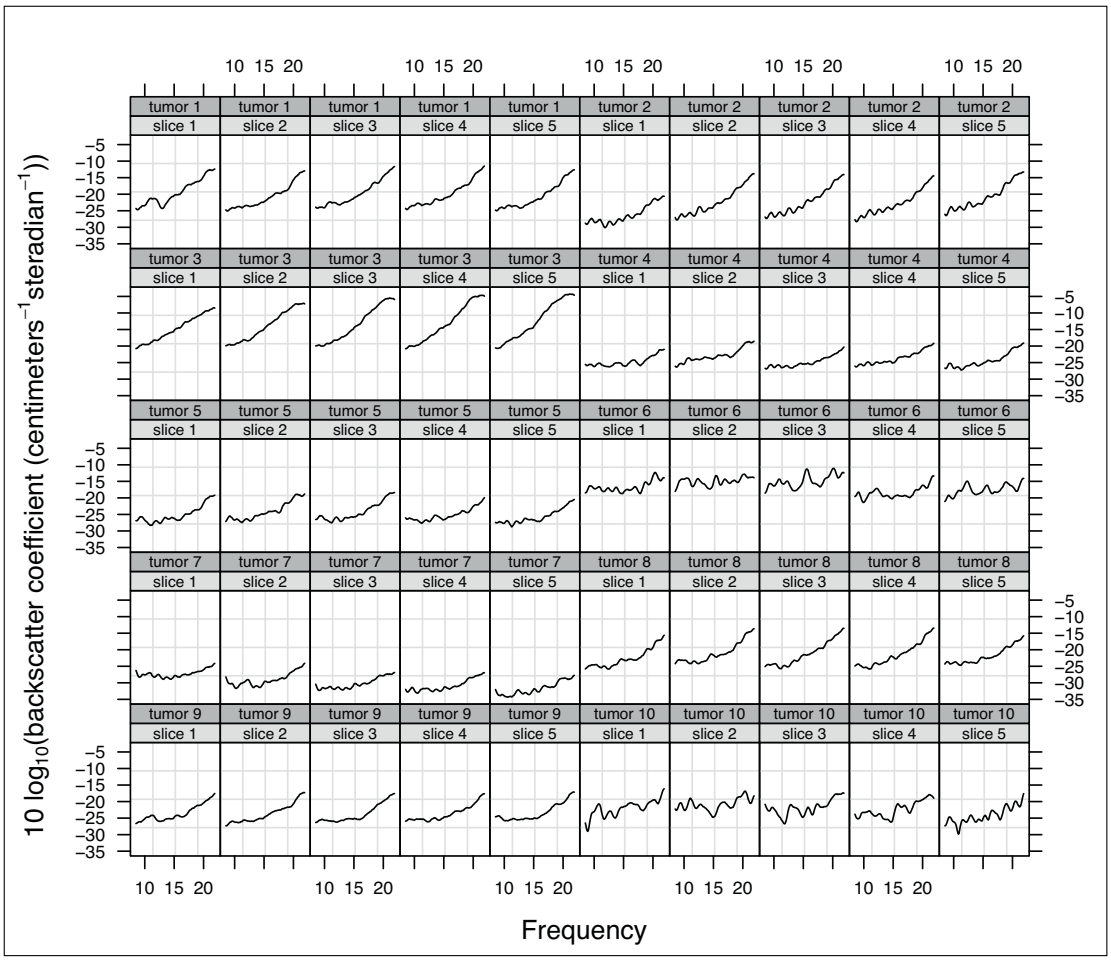

Figure 1 Illustrative three-level curve-type data. The response variable is $10\log_{10}$(backscatter coefficient) according to ultrasound technology. The units of backscatter coefficient are centimeters^{-1}steradian^{-1}. Level 1 corresponds to different ultrasound frequencies and matches the horizontal axes in each panel. Level 2 corresponds to different Slices of a Tumour due to differing probe locations. Level 3 corresponds to different Tumours with one Tumour for each of 10 laboratory mice

in which scalability to very large problems is paramount (e.g., Blei, Kucukelbir and McAuliffe, 2017).

Even though approximate Bayesian variational inference is our overarching goal, we also provide an important parallelism involving classical frequentist inference. Contemporary mixed model software such as `nlme()` (Pinheiro et al., 2018) and `lme4()` (Bates et al., 2015) in the R language provide streamlined algorithms for obtaining the best linear unbiased predictions of fixed and random effects in multilevel mixed models with details given in, for example, Pinheiro and Bayes

(2000). However, the sub-blocks of the covariance matrices required for construction of pointwise confidence interval bands around the estimated curves are *not* provided by such software. In the variational Bayesian analog, these sub-blocks are required for covariance parameter fitting and inference which, in turn, are needed for curve estimation. A significant contribution of this article is streamlined computation for both the best linear unbiased predictors and its corresponding covariance computation. Similar mathematical results lead to the mean field variational Bayesian inference equivalent. We present explicit ready-to-code algorithms for both two-level and three-level group-specific curve models. Extensions to higher level models could be derived using the blueprint that we establish here. Nevertheless, the algebraic overhead is increasingly burdensome with each increment in the number of levels. It is prudent to treat each multilevel case separately and here we already require several pages to cover two-level and three-level group-specific curve models. To our knowledge, this is the first article to provide streamlined algorithms for fitting three-level group-specific curve models.

Another important aspect of our group-specific curve fitting algorithms is the fact that they make use of the SolveTwoLevelSparseLeastSquares and SolveThreeLevelSparseLeastSquares algorithms developed for ordinary linear mixed models in Nolan et al. (2020). This realization means that the algorithms listed in Sections 2 and 3 are more concise and code-efficient: There is no need to repeat the implementation of these two fundamental algorithms for stable QR-based solving of higher level sparse linear systems. Sections S.12–S.13 of the supplementary materials provide details on the SolveTwoLevelSparseLeastSquares and SolveThreeLevelSparseLeastSquares algorithms.

Section 2 deals with the two-level case, and the three-level case is covered in Section 3. In Section 4, we provide some assessments concerning the accuracy and speed of the new variational inference algorithms. Concluding remarks are given in Section 5.

2 Two-level models

The simplest version of group-specific curve models involves the pairs (x_{ij}, y_{ij}) where x_{ij} is the jth value of the predictor variable within the ith group and y_{ij} is the corresponding value of the response variable. We let m denote the number of groups and n_i denote the number of predictor/response pairs within the ith group. The Gaussian response two-level group-specific curve model is

$$y_{ij} = f(x_{ij}) + g_i(x_{ij}) + \varepsilon_{ij}, \quad \varepsilon_{ij} \overset{\text{ind.}}{\sim} N(0, \sigma_\varepsilon^2), \quad 1 \le i \le m,\ 1 \le j \le n_i, \tag{2.1}$$

where the smooth function f is the global regression mean function and the smooth functions g_i, $1 \le i \le m$, allow for flexible group-specific deviations from f. As in Durban et al. (2005), we use mixed model-based penalized basis functions to model

f and the g_i. Specifically,

$$f(x) = \beta_0 + \beta_1 x + \sum_{k=1}^{K_{\text{gbl}}} u_{\text{gbl},k}\, z_{\text{gbl},k}(x), \quad u_{\text{gbl},k} \overset{\text{ind.}}{\sim} N(0, \sigma^2_{\text{gbl}}) \text{ and}$$

$$g_i(x) = u_{\text{lin},i0} + u_{\text{lin},i1}\, x + \sum_{k=1}^{K_{\text{grp}}} u_{\text{grp},ik}\, z_{\text{grp,k}}(x), \quad \begin{bmatrix} u_{\text{lin},i0} \\ u_{\text{lin},i1} \end{bmatrix} \overset{\text{ind.}}{\sim} N(\mathbf{0}, \boldsymbol{\Sigma}), \quad u_{\text{grp},ik} \overset{\text{ind.}}{\sim} N(0, \sigma^2_{\text{grp}}),$$

where $\{z_{\text{gbl},k}(\cdot) : 1 \le k \le K_{\text{gbl}}\}$ and $\{z_{\text{grp,k}}(\cdot) : 1 \le k \le K_{\text{grp}}\}$ are suitable sets of basis functions. Splines and wavelet families are the most common choices for the $z_{\text{gbl},k}(\cdot)$ and $z_{\text{grp,k}}(\cdot)$. In our illustrations and simulation studies, we use the canonical cubic O'Sullivan spline basis as described in Section 4 of Wand and Ormerod (2008), which corresponds to a low-rank version of classical smoothing splines (e.g., Wahba, 1990). The variance parameters σ^2_{gbl} and σ^2_{grp} control the effective degrees of freedom used for the global mean and group-specific deviation functions, respectively. Lastly, $\boldsymbol{\Sigma}$ is a 2×2 unstructured covariance matrix for the coefficients of the group-specific linear deviations.

We also use the notation:

$$\boldsymbol{x}_i \equiv \begin{bmatrix} x_{i1} \\ \vdots \\ x_{in_i} \end{bmatrix} \quad \text{and} \quad \boldsymbol{y}_i \equiv \begin{bmatrix} y_{i1} \\ \vdots \\ y_{in_i} \end{bmatrix}$$

for the vectors of predictors and responses corresponding to the ith group. Notation such as $z_{\text{gbl},1}(\boldsymbol{x}_i)$ denotes the $n_i \times 1$ vector containing $z_{\text{gbl},1}(x_{ij})$ values, $1 \le j \le n_i$.

The remainder of this section is concerned with streamlined methodologies for estimation and approximate inference for the model parameters. The basis dimension variables K_{gbl} and K_{grp} are not estimated via these methodologies and are additional tuning parameters that need to be chosen another way. Section 2.4 of Harezlak et al. (2018) summarizes theory, methodology and software concerned with this choice for ordinary nonparametric regression using penalized splines. Theory such as that presented in Kauermann et al. (2009) reveals that the basis dimension has a relatively minor effect after a particular threshold is reached. This suggests taking K_{gbl} and K_{grp} to be large relative to the predictor sample sizes, but this needs to be mitigated against computational cost. In practice, we recommend starting with 10–15 basis functions for each level of the hierarchy and assessing sensitivity to these choices using fits based on larger basis sizes.

2.1 Best linear unbiased prediction

Model (2.1) is expressible as a Gaussian response linear mixed model as follows:

$$\boldsymbol{y}|\boldsymbol{u} \sim N(\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}\boldsymbol{u}, \sigma^2_\varepsilon \boldsymbol{I}), \quad \boldsymbol{u} \sim N(\mathbf{0}, \boldsymbol{G}), \tag{2.2}$$

where

$$\boldsymbol{X} \equiv \begin{bmatrix} \boldsymbol{X}_1 \\ \vdots \\ \boldsymbol{X}_m \end{bmatrix} \quad \text{with} \quad \boldsymbol{X}_i \equiv [\boldsymbol{1}\ \boldsymbol{x}_i] \quad \text{and} \quad \boldsymbol{\beta} \equiv \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}$$

are the fixed effects design matrix and coefficients corresponding to the linear component of f. The random effects design matrix $\boldsymbol{Z}$ and corresponding random effects vector $\boldsymbol{u}$ are partitioned according to

$$\boldsymbol{Z} = \left[\boldsymbol{Z}_{\text{gbl}} \quad \underset{1\le i\le m}{\text{blockdiag}}([\boldsymbol{X}_i\ \boldsymbol{Z}_{\text{grp},i}])\right] \quad \text{and} \quad \boldsymbol{u} = \begin{bmatrix} \boldsymbol{u}_{\text{gbl}} \\ \begin{bmatrix} \boldsymbol{u}_{\text{lin},i} \\ \boldsymbol{u}_{\text{grp},i} \end{bmatrix}_{1\le i\le m} \end{bmatrix}, \tag{2.3}$$

where $\boldsymbol{u}_{\text{gbl}} = [u_{\text{gbl},1}\ \cdots\ u_{\text{gbl},K_{\text{gbl}}}]^T$ are the coefficients corresponding to the non-linear component of f, $\boldsymbol{u}_{\text{lin},i} = [u_{\text{lin},i0}\ u_{\text{lin},i1}]^T$ are the coefficients corresponding to the linear component of g_i and $\boldsymbol{u}_{\text{grp},i} = [u_{\text{grp},i1}\ \cdots\ u_{\text{grp},iK_{\text{grp}}}]^T$ are the coefficients corresponding to the non-linear component of g_i, $1 \le i \le m$. In (2.3), $\boldsymbol{Z}_{\text{gbl}} \equiv \text{stack}_{1\le i\le m}(\boldsymbol{Z}_{\text{gbl},i})$ and the matrices $\boldsymbol{Z}_{\text{gbl},i}$ and $\boldsymbol{Z}_{\text{grp},i}$, $1 \le i \le m$, contain, respectively, spline basis functions for the global mean function f and the ith group deviation functions g_i. Specifically,

$$\boldsymbol{Z}_{\text{gbl},i} \equiv [\, z_{\text{gbl},1}(\boldsymbol{x}_i) \cdots z_{\text{gbl},K_{\text{gbl}}}(\boldsymbol{x}_i)\,] \quad \text{and} \quad \boldsymbol{Z}_{\text{grp},i} = [\, z_{\text{grp},1}(\boldsymbol{x}_i) \cdots z_{\text{grp},K_{\text{grp}}}(\boldsymbol{x}_i)\,]$$

for $1 \le i \le m$. The corresponding fixed and random effects vectors are

$$\boldsymbol{u}_{\text{gbl}} \sim N(\boldsymbol{0}, \sigma^2_{\text{gbl}}\boldsymbol{I}_{K_{\text{gbl}}}) \quad \text{and} \quad \begin{bmatrix} \boldsymbol{u}_{\text{lin},i} \\ \boldsymbol{u}_{\text{grp},i} \end{bmatrix} \overset{\text{ind.}}{\sim} N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Sigma} & \boldsymbol{O} \\ \boldsymbol{O} & \sigma^2_{\text{grp}}\boldsymbol{I}_{K_{\text{grp}}} \end{bmatrix}\right), \quad 1 \le i \le m.$$

Hence, the full random effects covariance matrix is

$$\boldsymbol{G} = \text{Cov}(\boldsymbol{u}) = \begin{bmatrix} \sigma^2_{\text{gbl}}\boldsymbol{I}_{K_{\text{gbl}}} & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{I}_m \otimes \begin{bmatrix} \boldsymbol{\Sigma} & \boldsymbol{O} \\ \boldsymbol{O} & \sigma^2_{\text{grp}}\boldsymbol{I}_{K_{\text{grp}}} \end{bmatrix} \end{bmatrix}. \tag{2.4}$$

Next define the matrices

$$\boldsymbol{C} \equiv [\boldsymbol{X}\ \boldsymbol{Z}], \quad \boldsymbol{D}_{\text{BLUP}} \equiv \begin{bmatrix} \boldsymbol{O} & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{G}^{-1} \end{bmatrix} \quad \text{and} \quad \boldsymbol{R}_{\text{BLUP}} \equiv \sigma^2_\varepsilon \boldsymbol{I}. \tag{2.5}$$

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

$$\begin{aligned} \begin{bmatrix} \widehat{\boldsymbol{\beta}} \\ \widehat{\boldsymbol{u}} \end{bmatrix} &= (\boldsymbol{C}^T \boldsymbol{R}_{\text{BLUP}}^{-1} \boldsymbol{C} + \boldsymbol{D}_{\text{BLUP}})^{-1} \boldsymbol{C}^T \boldsymbol{R}_{\text{BLUP}}^{-1} \boldsymbol{y} \\ \text{and} \quad \text{Cov}\left(\begin{bmatrix} \widehat{\boldsymbol{\beta}} \\ \widehat{\boldsymbol{u}} - \boldsymbol{u} \end{bmatrix} \right) &= (\boldsymbol{C}^T \boldsymbol{R}_{\text{BLUP}}^{-1} \boldsymbol{C} + \boldsymbol{D}_{\text{BLUP}})^{-1}. \end{aligned} \tag{2.6}$$

This covariance matrix grows quadratically in m, so its storage becomes infeasible for large numbers of groups. However, only the following sub-blocks are required for adding pointwise confidence intervals to curve estimates:

$$\begin{aligned} \text{Cov}\left(\begin{bmatrix} \widehat{\boldsymbol{\beta}} \\ \widehat{\boldsymbol{u}}_{\text{gbl}} - \boldsymbol{u}_{\text{gbl}} \end{bmatrix} \right) &= \text{top left-hand } (2 + K_{\text{gbl}}) \times (2 + K_{\text{gbl}}) \\ &\quad \text{sub-block of } (\boldsymbol{C}^T \boldsymbol{R}_{\text{BLUP}}^{-1} \boldsymbol{C} + \boldsymbol{D}_{\text{BLUP}})^{-1}, \\ \text{Cov}\left(\begin{bmatrix} \widehat{\boldsymbol{u}}_{\text{lin},i} - \boldsymbol{u}_{\text{lin},i} \\ \widehat{\boldsymbol{u}}_{\text{grp},i} - \boldsymbol{u}_{\text{grp},i} \end{bmatrix} \right) &= \text{subsequent } (2 + K_{\text{grp}}) \times (2 + K_{\text{grp}}) \text{ diagonal} \\ &\quad \text{sub-blocks of } (\boldsymbol{C}^T \boldsymbol{R}_{\text{BLUP}}^{-1} \boldsymbol{C} + \boldsymbol{D}_{\text{BLUP}})^{-1} \\ &\quad \text{below Cov}\left(\begin{bmatrix} \widehat{\boldsymbol{\beta}} \\ \widehat{\boldsymbol{u}}_{\text{gbl}} - \boldsymbol{u}_{\text{gbl}} \end{bmatrix} \right), \; 1 \le i \le m, \text{ and} \\ E\left\{ \begin{bmatrix} \widehat{\boldsymbol{\beta}} \\ \widehat{\boldsymbol{u}}_{\text{gbl}} - \boldsymbol{u}_{\text{gbl}} \end{bmatrix} \begin{bmatrix} \widehat{\boldsymbol{u}}_{\text{lin},i} - \boldsymbol{u}_{\text{lin},i} \\ \widehat{\boldsymbol{u}}_{\text{grp},i} - \boldsymbol{u}_{\text{grp},i} \end{bmatrix}^T \right\} &= \text{subsequent } (2 + K_{\text{gbl}}) \times (2 + K_{\text{grp}}) \text{ sub-blocks} \\ &\quad \text{of } (\boldsymbol{C}^T \boldsymbol{R}_{\text{BLUP}}^{-1} \boldsymbol{C} + \boldsymbol{D}_{\text{BLUP}})^{-1} \text{ to the right of} \\ &\quad \text{Cov}\left(\begin{bmatrix} \widehat{\boldsymbol{\beta}} \\ \widehat{\boldsymbol{u}}_{\text{gbl}} - \boldsymbol{u}_{\text{gbl}} \end{bmatrix} \right), \; 1 \le i \le m. \end{aligned} \tag{2.7}$$

As in Nolan et al. (2020), we define the generic two-level sparse matrix problem to be determination of the vector $\boldsymbol{x}$ which minimizes the least squares criterion

$$\|\boldsymbol{b} - \boldsymbol{B}\boldsymbol{x}\|^2 \quad \text{where } \|\boldsymbol{v}\|^2 \equiv \boldsymbol{v}^T \boldsymbol{v} \text{ for any column vector } \boldsymbol{v}, \tag{2.8}$$

with $\boldsymbol{B}$ having the two-level sparse form

$$\boldsymbol{B} \equiv \left[\begin{array}{c|c|c|c|c} \boldsymbol{B}_1 & \dot{\boldsymbol{B}}_1 & \boldsymbol{O} & \cdots & \boldsymbol{O} \\ \hline \boldsymbol{B}_2 & \boldsymbol{O} & \dot{\boldsymbol{B}}_2 & \cdots & \boldsymbol{O} \\ \hline \vdots & \vdots & \vdots & \ddots & \vdots \\ \hline \boldsymbol{B}_m & \boldsymbol{O} & \boldsymbol{O} & \cdots & \dot{\boldsymbol{B}}_m \end{array}\right] \quad \text{and } \boldsymbol{b} \text{ partitioned according to } \quad \boldsymbol{b} \equiv \left[\begin{array}{c} \boldsymbol{b}_1 \\ \hline \boldsymbol{b}_2 \\ \hline \vdots \\ \hline \boldsymbol{b}_m \end{array}\right]. \tag{2.9}$$

In (2.9), for any $1 \leq i \leq m$, the matrices $\boldsymbol{B}_i$, $\dot{\boldsymbol{B}}_i$ and $\boldsymbol{b}_i$ each have the same number of rows. The numbers of columns in $\boldsymbol{B}_i$ and $\dot{\boldsymbol{B}}_i$ are arbitrary, whereas the $\boldsymbol{b}_i$ are column vectors. In addition to solving for $\boldsymbol{x}$, the sub-blocks of $(\boldsymbol{B}^T\boldsymbol{B})^{-1}$ corresponding to the non-sparse regions of $\boldsymbol{B}^T\boldsymbol{B}$ are included in our definition of a two-level sparse matrix least squares problem. Algorithm A.2 of Nolan et al. (2020) provides a stable and efficient solution to this problem and labels it the SolveTwoLevelSparseLeastSquares algorithm. Section S.13 of the supplementary materials contains details regarding this algorithm. It uses the the following notation for sub-blocks of $\boldsymbol{x} \equiv (\boldsymbol{B}^T\boldsymbol{B})^{-1}\boldsymbol{B}^T\boldsymbol{b}$ and $\boldsymbol{A}^{-1}$, where $\boldsymbol{A} \equiv \boldsymbol{B}^T\boldsymbol{B}$:

$$\boldsymbol{x} \equiv \left[\begin{array}{c} \boldsymbol{x}_1 \\ \hline \boldsymbol{x}_{2,1} \\ \hline \boldsymbol{x}_{2,2} \\ \hline \vdots \\ \hline \boldsymbol{x}_{2,m} \end{array}\right] \quad \text{and} \quad \boldsymbol{A}^{-1} \equiv \left[\begin{array}{c|c|c|c|c} \boldsymbol{A}^{11} & \boldsymbol{A}^{12,1} & \boldsymbol{A}^{12,2} & \cdots & \boldsymbol{A}^{12,m} \\ \hline \boldsymbol{A}^{12,1\,T} & \boldsymbol{A}^{22,1} & \times & \cdots & \times \\ \hline \boldsymbol{A}^{12,2\,T} & \times & \boldsymbol{A}^{22,2} & \cdots & \times \\ \hline \vdots & \vdots & \vdots & \ddots & \vdots \\ \hline \boldsymbol{A}^{12,m\,T} & \times & \times & \cdots & \boldsymbol{A}^{22,m} \end{array}\right]. \tag{2.10}$$

In Nolan et al. (2020), we used SolveTwoLevelSparseLeastSquares for fitting two-level linear mixed models. However, precisely the same algorithm can be used for fitting two-level group-specific curve models because of the following.

Result 1 *Computation of* $[\widehat{\boldsymbol{\beta}}^T \ \widehat{\boldsymbol{u}}^T]^T$ *and each of the sub-blocks of* $\text{Cov}([\widehat{\boldsymbol{\beta}}^T \ (\widehat{\boldsymbol{u}} - \boldsymbol{u})^T]^T)$ *listed in (2.7) are expressible as solutions to the two-level sparse matrix least squares problem:*

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

where the non-zero sub-blocks $\boldsymbol{B}$ *and* $\boldsymbol{b}$*, according to the notation in (2.9), are for* $1 \leq i \leq m$*:*

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

with each of these matrices having $\tilde{n}_i = n_i + K_{\text{gbl}} + 2 + K_{\text{grp}}$ *rows and with* $\boldsymbol{B}_i$ *having* $p = 2 + K_{\text{gbl}}$ *columns and* $\dot{\boldsymbol{B}}_i$ *having* $q = 2 + K_{\text{grp}}$ *columns. The solutions are, with sub-block labelling according to (2.10),*

$$\begin{bmatrix} \widehat{\boldsymbol{\beta}} \\ \widehat{\boldsymbol{u}}_{\text{gbl}} \end{bmatrix} = \boldsymbol{x}_1, \quad \text{Cov}\left(\begin{bmatrix} \widehat{\boldsymbol{\beta}} \\ \widehat{\boldsymbol{u}}_{\text{gbl}} - \boldsymbol{u}_{\text{gbl}} \end{bmatrix}\right) = \boldsymbol{A}^{11}$$

and

$$\begin{bmatrix} \widehat{\boldsymbol{u}}_{\text{lin},i} \\ \widehat{\boldsymbol{u}}_{\text{grp},i} \end{bmatrix} = \boldsymbol{x}_{2,i}, \quad E\left\{\begin{bmatrix} \widehat{\boldsymbol{\beta}} \\ \widehat{\boldsymbol{u}}_{\text{gbl}} - \boldsymbol{u}_{\text{gbl}} \end{bmatrix}\begin{bmatrix} \widehat{\boldsymbol{u}}_{\text{lin},i} - \boldsymbol{u}_{\text{lin},i} \\ \widehat{\boldsymbol{u}}_{\text{grp},i} - \boldsymbol{u}_{\text{grp},i} \end{bmatrix}^T\right\} = \boldsymbol{A}^{12,i},$$

$$\text{Cov}\left(\begin{bmatrix} \widehat{\boldsymbol{u}}_{\text{lin},i} - \boldsymbol{u}_{\text{lin},i} \\ \widehat{\boldsymbol{u}}_{\text{grp},i} - \boldsymbol{u}_{\text{grp},i} \end{bmatrix}\right) = \boldsymbol{A}^{22,i}, \quad 1 \leq i \leq m.$$

A derivation of Result 1 is given in Section S.1 of the supplementary materials. Algorithm 1 encapsulates streamlined best linear unbiased prediction computation together with coefficient covariance matrix sub-blocks of interest.

Algorithm 1 Streamlined algorithm for obtaining best linear unbiased predictions and corresponding covariance matrix components for the two-level group-specific curves model

Inputs: $\boldsymbol{y}_i(n_i \times 1),\ \boldsymbol{X}_i(n_i \times 2),\ \boldsymbol{Z}_{\text{gbl},i}(n_i \times K_{\text{gbl}}),\ \boldsymbol{Z}_{\text{grp},i}(n_i \times K_{\text{grp}}),\ 1 \le i \le m;$
$\sigma_\varepsilon^2, \sigma_{\text{gbl}}^2, \sigma_{\text{grp}}^2 > 0,\ \boldsymbol{\Sigma}(q \times q)$, symmetric and positive definite.

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

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

$$\dot{\boldsymbol{B}}_i \longleftarrow \begin{bmatrix} \sigma_\varepsilon^{-1}\boldsymbol{X}_i & \sigma_\varepsilon^{-1}\boldsymbol{Z}_{\text{grp},i} \\ \boldsymbol{O} & \boldsymbol{O} \\ \boldsymbol{\Sigma}^{-1/2} & \boldsymbol{O} \\ \boldsymbol{O} & \sigma_{\text{grp}}^{-1}\boldsymbol{I}_{K_{\text{grp}}} \end{bmatrix}$$

$$\mathcal{S}_1 \longleftarrow \mathit{SolveTwoLevelSparseLeastSquares}\left(\left\{(\boldsymbol{b}_i, \boldsymbol{B}_i, \dot{\boldsymbol{B}}_i) : 1 \le i \le m\right\}\right)$$

$$\begin{bmatrix} \widehat{\boldsymbol{\beta}} \\ \widehat{\boldsymbol{u}}_{\text{gbl}} \end{bmatrix} \longleftarrow \boldsymbol{x}_1 \text{ component of } \mathcal{S}_1 \text{ ; } \text{Cov}\left(\begin{bmatrix} \widehat{\boldsymbol{\beta}} \\ \widehat{\boldsymbol{u}}_{\text{gbl}} - \boldsymbol{u}_{\text{gbl}} \end{bmatrix}\right) \longleftarrow \boldsymbol{A}^{11} \text{ component of } \mathcal{S}_1$$

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

$$\begin{bmatrix} \widehat{\boldsymbol{u}}_{\text{lin},i} \\ \widehat{\boldsymbol{u}}_{\text{grp},i} \end{bmatrix} \longleftarrow \boldsymbol{x}_{2,i} \text{ component of } \mathcal{S}_1$$

$$\text{Cov}\left(\begin{bmatrix} \widehat{\boldsymbol{u}}_{\text{lin},i} - \boldsymbol{u}_{\text{lin},i} \\ \widehat{\boldsymbol{u}}_{\text{grp},i} - \boldsymbol{u}_{\text{grp},i} \end{bmatrix}\right) \longleftarrow \boldsymbol{A}^{22,i} \text{ component of } \mathcal{S}_1$$

$$E\left\{\begin{bmatrix} \widehat{\boldsymbol{\beta}} \\ \widehat{\boldsymbol{u}}_{\text{gbl}} - \boldsymbol{u}_{\text{gbl}} \end{bmatrix}\begin{bmatrix} \widehat{\boldsymbol{u}}_{\text{lin},i} - \boldsymbol{u}_{\text{lin},i} \\ \widehat{\boldsymbol{u}}_{\text{grp},i} - \boldsymbol{u}_{\text{grp},i} \end{bmatrix}^T\right\} \longleftarrow \boldsymbol{A}^{12,i} \text{ component of } \mathcal{S}_1$$

Output:

$$\left(\begin{bmatrix} \widehat{\boldsymbol{\beta}} \\ \widehat{\boldsymbol{u}}_{\text{gbl}} \end{bmatrix},\ \text{Cov}\left(\begin{bmatrix} \widehat{\boldsymbol{\beta}} \\ \widehat{\boldsymbol{u}}_{\text{gbl}} - \boldsymbol{u}_{\text{gbl}} \end{bmatrix}\right), \left\{\left(\begin{bmatrix} \widehat{\boldsymbol{u}}_{\text{lin},i} \\ \widehat{\boldsymbol{u}}_{\text{grp},i} \end{bmatrix},\ \text{Cov}\left(\begin{bmatrix} \widehat{\boldsymbol{u}}_{\text{lin},i} - \boldsymbol{u}_{\text{lin},i} \\ \widehat{\boldsymbol{u}}_{\text{grp},i} - \boldsymbol{u}_{\text{grp},i} \end{bmatrix}\right),\ E\left\{\begin{bmatrix} \widehat{\boldsymbol{\beta}} \\ \widehat{\boldsymbol{u}}_{\text{gbl}} - \boldsymbol{u}_{\text{gbl}} \end{bmatrix}\begin{bmatrix} \widehat{\boldsymbol{u}}_{\text{lin},i} - \boldsymbol{u}_{\text{lin},i} \\ \widehat{\boldsymbol{u}}_{\text{grp},i} - \boldsymbol{u}_{\text{grp},i} \end{bmatrix}^T\right\}\right) :\ 1 \le i \le m\right\}\right)$$

2.2 Mean field variational Bayes

We now consider the following Bayesian extension of (2.2) and (2.4):

$$\begin{gathered}
\boldsymbol{y}|\boldsymbol{\beta}, \boldsymbol{u}, \sigma_\varepsilon^2 \sim N(\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}\boldsymbol{u}, \sigma_\varepsilon^2 \boldsymbol{I}), \quad \boldsymbol{u}|\sigma_{\text{gbl}}^2, \sigma_{\text{grp}}^2, \boldsymbol{\Sigma} \sim N(\boldsymbol{0}, \boldsymbol{G}), \quad \boldsymbol{G} \text{ as defined in (2.4)}, \\
\boldsymbol{\beta} \sim N(\boldsymbol{\mu}_{\boldsymbol{\beta}}, \boldsymbol{\Sigma}_{\boldsymbol{\beta}}), \quad \sigma_\varepsilon^2|a_\varepsilon \sim \text{Inverse-}\chi^2(\nu_\varepsilon, 1/a_\varepsilon), \quad a_\varepsilon \sim \text{Inverse-}\chi^2(1, 1/(\nu_\varepsilon s_\varepsilon^2)), \\
\sigma_{\text{gbl}}^2|a_{\text{gbl}} \sim \text{Inverse-}\chi^2(\nu_{\text{gbl}}, 1/a_{\text{gbl}}), \quad a_{\text{gbl}} \sim \text{Inverse-}\chi^2(1, 1/(\nu_{\text{gbl}} s_{\text{gbl}}^2)), \\
\sigma_{\text{grp}}^2|a_{\text{grp}} \sim \text{Inverse-}\chi^2(\nu_{\text{grp}}, 1/a_{\text{grp}}), \quad a_{\text{grp}} \sim \text{Inverse-}\chi^2(1, 1/(\nu_{\text{grp}} s_{\text{grp}}^2)), \\
\boldsymbol{\Sigma}|\boldsymbol{A}_{\boldsymbol{\Sigma}} \sim \text{Inverse-G-Wishart}\big(G_{\text{full}}, \nu_{\boldsymbol{\Sigma}} + 2, \boldsymbol{A}_{\boldsymbol{\Sigma}}^{-1}\big), \\
\boldsymbol{A}_{\boldsymbol{\Sigma}} \sim \text{Inverse-G-Wishart}(G_{\text{diag}}, 1, \boldsymbol{\Lambda}_{\boldsymbol{A}_{\boldsymbol{\Sigma}}}), \quad \boldsymbol{\Lambda}_{\boldsymbol{A}_{\boldsymbol{\Sigma}}} \equiv \{\nu_{\boldsymbol{\Sigma}} \text{diag}(s_{\boldsymbol{\Sigma},1}^2, s_{\boldsymbol{\Sigma},2}^2)\}^{-1}.
\end{gathered} \tag{2.11}$$

Here the 2×1 vector $\boldsymbol{\mu}_{\boldsymbol{\beta}}$ and 2×2 symmetric positive definite matrix $\boldsymbol{\Sigma}_{\boldsymbol{\beta}}$ are hyperparameters corresponding to the prior distribution on $\boldsymbol{\beta}$ and

$$\nu_\varepsilon, s_\varepsilon, \nu_{\text{gbl}}, s_{\text{gbl}}, \nu_{\text{grp}}, s_{\text{grp}}, \nu_{\boldsymbol{\Sigma}}, s_{\boldsymbol{\Sigma},1}, s_{\boldsymbol{\Sigma},2} > 0$$

are hyperparameters for the variance and covariance matrix parameters. Details on the Inverse G-Wishart distribution, and the Inverse-χ^2 special case, are given in Section S.3 of the supplementary materials. The auxiliary variable a_ε is defined so that σ_ε has a Half-t distribution with degrees of freedom parameter ν_ε and scale parameter s_ε, with larger values of s_ε corresponding to greater noninformativity. Analogous comments apply to the other standard deviation parameters. Setting $\nu_{\boldsymbol{\Sigma}} = 2$ leads to the correlation parameter in $\boldsymbol{\Sigma}$ having a Uniform distribution on $(-1, 1)$ (Huang and Wand, 2013).

Throughout this article, we use $\mathfrak{p}$ generically to denote a density function corresponding to random quantities in Bayesian models such as (2.11). For example, $\mathfrak{p}(\boldsymbol{\beta})$ denotes the prior density function of $\boldsymbol{\beta}$ and $\mathfrak{p}(\boldsymbol{u}|\sigma_{\text{gbl}}^2, \sigma_{\text{grp}}^2, \boldsymbol{\Sigma})$ denotes the density function of $\boldsymbol{u}$ conditional on $(\sigma_{\text{gbl}}^2, \sigma_{\text{grp}}^2, \boldsymbol{\Sigma})$. Now consider the following mean field restriction on the joint posterior density function of all parameters in (2.11):

$$\mathfrak{p}(\boldsymbol{\beta}, \boldsymbol{u}, a_\varepsilon, a_{\text{gbl}}, a_{\text{grp}}, \boldsymbol{A}_{\boldsymbol{\Sigma}}, \sigma_\varepsilon^2, \sigma_{\text{gbl}}^2, \sigma_{\text{grp}}^2, \boldsymbol{\Sigma}|\boldsymbol{y}) \approx \mathfrak{q}(\boldsymbol{\beta}, \boldsymbol{u}, a_\varepsilon, a_{\text{gbl}}, a_{\text{grp}}, \boldsymbol{A}_{\boldsymbol{\Sigma}})\, \mathfrak{q}(\sigma_\varepsilon^2, \sigma_{\text{gbl}}^2, \sigma_{\text{grp}}^2, \boldsymbol{\Sigma}). \tag{2.12}$$

Here, generically, each $\mathfrak{q}$ denotes an approximate posterior density function of the random vector indicated by its argument according to the mean field restriction (2.12). Then application of the minimum Kullback–Leibler divergence equations (e.g., equation (10.9) of Bishop, 2006) leads to the optimal $\mathfrak{q}$-density functions for the parameters of interest being as follows:

$$\begin{aligned}
&\mathfrak{q}^*(\boldsymbol{\beta}, \boldsymbol{u}) \text{ has a } N\big(\boldsymbol{\mu}_{\mathfrak{q}(\boldsymbol{\beta},\boldsymbol{u})}, \boldsymbol{\Sigma}_{\mathfrak{q}(\boldsymbol{\beta},\boldsymbol{u})}\big) \text{ distribution,}\\
&\mathfrak{q}^*(\sigma^2_\varepsilon) \text{ has an Inverse-}\chi^2\big(\xi_{\mathfrak{q}(\sigma^2_\varepsilon)}, \lambda_{\mathfrak{q}(\sigma^2_\varepsilon)}\big) \text{ distribution,}\\
&\mathfrak{q}^*(\sigma^2_{\text{gbl}}) \text{ has an Inverse-}\chi^2\big(\xi_{\mathfrak{q}(\sigma^2_{\text{gbl}})}, \lambda_{\mathfrak{q}(\sigma^2_{\text{gbl}})}\big) \text{ distribution,}\\
&\mathfrak{q}^*(\sigma^2_{\text{grp}}) \text{ has an Inverse-}\chi^2\big(\xi_{\mathfrak{q}(\sigma^2_{\text{grp}})}, \lambda_{\mathfrak{q}(\sigma^2_{\text{grp}})}\big) \text{ distribution}\\
\text{and } &\mathfrak{q}^*(\boldsymbol{\Sigma}) \text{ has an Inverse-G-Wishart}\big(G_{\text{full}}, \xi_{\mathfrak{q}(\boldsymbol{\Sigma})}, \boldsymbol{\Lambda}_{\mathfrak{q}(\boldsymbol{\Sigma})}\big) \text{ distribution.}
\end{aligned}$$

The optimal $\mathfrak{q}$-density parameters are determined via an iterative coordinate ascent algorithm, with details given in Section S.5 of this article's supplementary materials. The stopping criterion is based on the variational lower bound on the marginal likelihood (e.g., Bishop, 2006, Section 10.2.2) and denoted $\underline{\mathfrak{p}}(\boldsymbol{y}; \mathfrak{q})$. Its logarithmic form and derivation are given in Section S.6 of the supplementary materials.

Note that updates for $\boldsymbol{\mu}_{\mathfrak{q}(\boldsymbol{\beta},\boldsymbol{u})}$ and $\boldsymbol{\Sigma}_{\mathfrak{q}(\boldsymbol{\beta},\boldsymbol{u})}$ may be written

$$\boldsymbol{\mu}_{\mathfrak{q}(\boldsymbol{\beta},\boldsymbol{u})} \leftarrow (\boldsymbol{C}^T\boldsymbol{R}^{-1}_{\text{MFVB}}\boldsymbol{C} + \boldsymbol{D}_{\text{MFVB}})^{-1}(\boldsymbol{C}^T\boldsymbol{R}^{-1}_{\text{MFVB}}\boldsymbol{y} + \boldsymbol{o}_{\text{MFVB}}) \quad \text{and} \quad \boldsymbol{\Sigma}_{\mathfrak{q}(\boldsymbol{\beta},\boldsymbol{u})} \leftarrow (\boldsymbol{C}^T\boldsymbol{R}^{-1}_{\text{MFVB}}\boldsymbol{C} + \boldsymbol{D}_{\text{MFVB}})^{-1}, \tag{2.13}$$

where

$$\boldsymbol{R}_{\text{MFVB}} \equiv \mu^{-1}_{\mathfrak{q}(1/\sigma^2_\varepsilon)}\boldsymbol{I}, \quad \boldsymbol{D}_{\text{MFVB}} \equiv \begin{bmatrix} \boldsymbol{\Sigma}^{-1}_{\boldsymbol{\beta}} & \boldsymbol{O} & \boldsymbol{O} \\ \boldsymbol{O} & \mu_{\mathfrak{q}(1/\sigma^2_{\text{gbl}})}\boldsymbol{I} & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{O} & \underset{1\le i\le m}{\text{blockdiag}}\begin{bmatrix} \boldsymbol{M}_{\mathfrak{q}(\boldsymbol{\Sigma}^{-1})} & \boldsymbol{O} \\ \boldsymbol{O} & \mu_{\mathfrak{q}(1/\sigma^2_{\text{grp}})}\boldsymbol{I} \end{bmatrix} \end{bmatrix} \tag{2.14}$$

$$\text{and} \quad \boldsymbol{o}_{\text{MFVB}} \equiv \begin{bmatrix} \boldsymbol{\Sigma}^{-1}_{\boldsymbol{\beta}}\boldsymbol{\mu}_{\boldsymbol{\beta}} \\ 0 \end{bmatrix}.$$

For increasingly large numbers of groups, the matrix $\boldsymbol{\Sigma}_{\mathfrak{q}(\boldsymbol{\beta},\boldsymbol{u})}$ approaches a size that is untenable for random access memory storage on standard 2020s workplace computers. However, only the following relatively small sub-blocks of $\boldsymbol{\Sigma}_{\mathfrak{q}(\boldsymbol{\beta},\boldsymbol{u})}$ are required for variational inference concerning the variance and covariance matrix parameters:

$$\begin{aligned}
&\boldsymbol{\Sigma}_{\mathfrak{q}(\boldsymbol{\beta},\boldsymbol{u}_{\text{gbl}})} = \text{top left-hand } (2 + K_{\text{gbl}}) \times (2 + K_{\text{gbl}}) \text{ sub-block of } (\boldsymbol{C}^T\boldsymbol{R}^{-1}_{\text{MFVB}}\boldsymbol{C} + \boldsymbol{D}_{\text{MFVB}})^{-1},\\
&\boldsymbol{\Sigma}_{\mathfrak{q}(\boldsymbol{u}_{\text{lin},i},\boldsymbol{u}_{\text{grp},i})} = \text{subsequent } (2 + K_{\text{grp}}) \times (2 + K_{\text{grp}}) \text{ diagonal sub-blocks of}\\
&\qquad\qquad (\boldsymbol{C}^T\boldsymbol{R}^{-1}_{\text{MFVB}}\boldsymbol{C} + \boldsymbol{D}_{\text{MFVB}})^{-1} \text{ below } \boldsymbol{\Sigma}_{\mathfrak{q}(\boldsymbol{\beta},\boldsymbol{u}_{\text{gbl}})},\ 1 \le i \le m, \text{ and}
\end{aligned}$$

$$E_{\mathfrak{q}}\left\{\left(\begin{bmatrix}\boldsymbol{\beta}\\ \boldsymbol{u}_{\text{gbl}}\end{bmatrix}-\boldsymbol{\mu}_{\mathfrak{q}(\boldsymbol{\beta},\boldsymbol{u}_{\text{gbl}})}\right)\left(\begin{bmatrix}\boldsymbol{u}_{\text{lin},i}\\ \boldsymbol{u}_{\text{grp},i}\end{bmatrix}-\boldsymbol{\mu}_{\mathfrak{q}(\boldsymbol{u}_{\text{lin},i},\boldsymbol{u}_{\text{grp},i}))}\right)^T\right\} = \text{subsequent}$$
$$(2+K_{\text{gbl}})\times(2+K_{\text{grp}}) \text{ sub-blocks of } (\boldsymbol{C}^T\boldsymbol{R}_{\text{MFVB}}^{-1}\boldsymbol{C}+\boldsymbol{D}_{\text{MFVB}})^{-1} \tag{2.15}$$
$$\text{to the right of } \boldsymbol{\Sigma}_{\mathfrak{q}(\boldsymbol{\beta},\boldsymbol{u}_{\text{gbl}})},\ 1\le i\le m.$$

Our streamlined mean field variational Bayes algorithm depends on Result 2.

Result 2 *The mean field variational Bayes updates of* $\boldsymbol{\mu}_{\mathfrak{q}(\boldsymbol{\beta},\boldsymbol{u})}$ *and each of the sub-blocks of* $\boldsymbol{\Sigma}_{\mathfrak{q}(\boldsymbol{\beta},\boldsymbol{u})}$ *in (2.15) are expressible as a two-level sparse matrix least squares problem of the form:*

$$\left\|\boldsymbol{b}-\boldsymbol{B}\boldsymbol{\mu}_{\mathfrak{q}(\boldsymbol{\beta},\boldsymbol{u})}\right\|^2,$$

where the non-zero sub-blocks $\boldsymbol{B}$ *and* $\boldsymbol{b}$*, according to the notation in (2.9), are, for* $1\le i\le m$*,*

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

and

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

with each of these matrices having $\tilde{n}_i = n_i + 2 + K_{\text{gbl}} + 2 + K_{\text{grp}}$ *rows and with* $\boldsymbol{B}_i$ *having* $p = 2 + K_{\text{gbl}}$ *columns and* $\dot{\boldsymbol{B}}_i$ *having* $q = 2 + K_{\text{grp}}$ *columns. The solutions are, with*

sub-block labelling according to (2.10),

$$\boldsymbol{\mu}_{\mathrm{q}(\boldsymbol{\beta},\boldsymbol{u}_{\mathrm{gbl}})} = \boldsymbol{x}_1, \quad \boldsymbol{\Sigma}_{\mathrm{q}(\boldsymbol{\beta},\boldsymbol{u}_{\mathrm{gbl}})} = \boldsymbol{A}^{11},$$

$$\boldsymbol{\mu}_{\mathrm{q}(\boldsymbol{u}_{\mathrm{lin},i},\boldsymbol{u}_{\mathrm{grp},i})} = \boldsymbol{x}_{2,i}, \quad \boldsymbol{\Sigma}_{\mathrm{q}(\boldsymbol{u}_{\mathrm{lin},i},\boldsymbol{u}_{\mathrm{grp},i})} = \boldsymbol{A}^{22,i},$$

and

$$E_q\left\{\begin{bmatrix}\boldsymbol{\beta}-\boldsymbol{\mu}_{\mathrm{q}(\boldsymbol{\beta})}\\ \boldsymbol{u}_{\mathrm{gbl}}-\boldsymbol{\mu}_{\mathrm{q}(\boldsymbol{u}_{\mathrm{gbl}})}\end{bmatrix}\begin{bmatrix}\boldsymbol{u}_{\mathrm{lin},i}-\boldsymbol{\mu}_{\mathrm{q}(\boldsymbol{u}_{\mathrm{lin},i})}\\ \boldsymbol{u}_{\mathrm{grp},i}-\boldsymbol{\mu}_{\mathrm{q}(\boldsymbol{u}_{\mathrm{grp},i})}\end{bmatrix}^T\right\} = \boldsymbol{A}^{12,i},\ 1 \le i \le m.$$

Algorithm 2 utilizes Result 2 to facilitate streamlined computation of the variational parameters.

Lastly, we note that Algorithm 2 is loosely related to Algorithm 2 of Lee and Wand (2016). One difference is that we are treating the Gaussian, rather than Bernoulli, response situation here. In addition, we are using the recent sparse multilevel matrix results of Nolan and Wand (2020) which are amenable to higher level extensions, such as the three-level group-specific curve model treated in Section 3.

2.3 Contrast function extension

In many curve-type data applications, the data can be categorized as being from two or more types. Of particular interest in such circumstances are contrast function estimates and accompanying standard errors. The streamlined approaches used in Algorithms 1 and 2 still apply for the contrast function extension regardless of the number of categories. The two category situation, where there is a single contrast function, is described here. The extension to higher numbers of categories is straightforward.

Suppose that the (x_{ij}, y_{ij}) pairs are from one of two categories, labelled A and B, and introduce the indicator variable data:

$$\iota_{ij}^{A} \equiv \begin{cases} 1 \text{ if } (x_{ij}, y_{ij}) \text{ is from category } A, \\ 0 \text{ if } (x_{ij}, y_{ij}) \text{ is from category } B. \end{cases} \tag{2.16}$$

Then penalized spline models for the global mean and deviation functions for each category are

$$\left.\begin{aligned} f^A(x) &= \beta_0^{\mathrm{A}} + \beta_1^{\mathrm{A}}\, x + \sum_{k=1}^{K_{\mathrm{gbl}}} u_{\mathrm{gbl},k}^{A} z_{\mathrm{gbl},k}(x) \\ g_i^A(x) &= u_{\mathrm{lin},i0}^{A} + u_{\mathrm{lin},i1}^{A}\, x + \sum_{k=1}^{K_{\mathrm{grp}}} u_{\mathrm{grp},ik}^{A} z_{\mathrm{grp},k}(x) \end{aligned}\right\} \text{ for category A}$$

Algorithm 2 QR-decomposition-based streamlined algorithm for obtaining mean field variational Bayes approximate posterior density functions for the parameters in the Bayesian two-level group-specific curves model (2.11) with product density restriction (2.12)

Data Inputs: $\boldsymbol{y}_i(n_i \times 1),\ \boldsymbol{X}_i(n_i \times 2),\ \boldsymbol{Z}_{\text{gbl},i}(n_i \times K_{\text{gbl}}),\ \boldsymbol{Z}_{\text{grp},i}(n_i \times K_{\text{grp}}),\ 1 \le i \le m$;

Hyperparameter Inputs: $\boldsymbol{\mu}_{\boldsymbol{\beta}}(2 \times 1)$, $\boldsymbol{\Sigma}_{\boldsymbol{\beta}}(2 \times 2)$ symmetric and positive definite,

$s_\varepsilon, \nu_\varepsilon, s_{\text{gbl}}, \nu_{\text{gbl}}, s_{\boldsymbol{\Sigma},1}, s_{\boldsymbol{\Sigma},2}, \nu_{\boldsymbol{\Sigma}}, s_{\text{grp}}, \nu_{\text{grp}} > 0$.

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

$\boldsymbol{C}_{\text{gbl},i} \longleftarrow [\boldsymbol{X}_i\ \boldsymbol{Z}_{\text{gbl},i}]$; $\boldsymbol{C}_{\text{grp},i} \longleftarrow [\boldsymbol{X}_i\ \boldsymbol{Z}_{\text{grp},i}]$

Initialize: $\mu_{\text{q}(1/\sigma_\varepsilon^2)}, \mu_{\text{q}(1/\sigma_{\text{gbl}}^2)}, \mu_{\text{q}(1/\sigma_{\text{grp}}^2)}, \mu_{\text{q}(1/a_\varepsilon)}, \mu_{\text{q}(1/a_{\text{gbl}})}, \mu_{\text{q}(1/a_{\text{grp}})} > 0$,

$\boldsymbol{M}_{\text{q}(\boldsymbol{\Sigma}^{-1})}(2 \times 2), \boldsymbol{M}_{\text{q}(A_{\boldsymbol{\Sigma}}^{-1})}(2 \times 2)$ both symmetric and positive definite.

$\xi_{\text{q}(\sigma_\varepsilon^2)} \longleftarrow \nu_\varepsilon + \sum_{i=1}^m n_i$; $\xi_{\text{q}(\sigma_{\text{gbl}}^2)} \longleftarrow \nu_{\text{gbl}} + K_{\text{gbl}}$; $\xi_{\text{q}(\boldsymbol{\Sigma})} \longleftarrow \nu_{\boldsymbol{\Sigma}} + 2 + m$

$\xi_{\text{q}(\sigma_{\text{grp}}^2)} \longleftarrow \nu_{\text{grp}} + mK_{\text{grp}}$; $\xi_{\text{q}(a_\varepsilon)} \longleftarrow \nu_\varepsilon + 1$; $\xi_{\text{q}(a_{\text{gbl}})} \longleftarrow \nu_{\text{gbl}} + 1$; $\xi_{\text{q}(a_{\text{grp}})} \longleftarrow \nu_{\text{grp}} + 1$

$\xi_{\text{q}(A_{\boldsymbol{\Sigma}})} \longleftarrow \nu_{\boldsymbol{\Sigma}} + 2$

Cycle:

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

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

$$\dot{\boldsymbol{B}}_i \longleftarrow \begin{bmatrix} \mu^{1/2}_{\text{q}(1/\sigma_\varepsilon^2)}\boldsymbol{X}_i & \mu^{1/2}_{\text{q}(1/\sigma_\varepsilon^2)}\boldsymbol{Z}_{\text{grp},i} \\ \boldsymbol{O} & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{O} \\ \boldsymbol{M}^{1/2}_{\text{q}(\boldsymbol{\Sigma}^{-1})} & \boldsymbol{O} \\ \boldsymbol{O} & \mu^{1/2}_{\text{q}(1/\sigma_{\text{grp}}^2)}\boldsymbol{I}_{K_{\text{grp}}} \end{bmatrix}$$

$\mathcal{S}_2 \longleftarrow \mathit{SolveTwoLevelSparseLeastSquares}\left(\left\{(\boldsymbol{b}_i, \boldsymbol{B}_i, \dot{\boldsymbol{B}}_i) : 1 \le i \le m\right\}\right)$

$\boldsymbol{\mu}_{\text{q}(\boldsymbol{\beta},\boldsymbol{u}_{\text{gbl}})} \longleftarrow \boldsymbol{x}_1$ component of $\mathcal{S}_2$; $\boldsymbol{\Sigma}_{\text{q}(\boldsymbol{\beta},\boldsymbol{u}_{\text{gbl}})} \longleftarrow \boldsymbol{A}^{11}$ component of $\mathcal{S}_2$

$\boldsymbol{\mu}_{\text{q}(\boldsymbol{u}_{\text{gbl}})} \longleftarrow$ last K_{gbl} rows of $\boldsymbol{\mu}_{\text{q}(\boldsymbol{\beta},\boldsymbol{u}_{\text{gbl}})}$

$\boldsymbol{\Sigma}_{\text{q}(\boldsymbol{u}_{\text{gbl}})} \longleftarrow$ bottom-right $K_{\text{gbl}} \times K_{\text{gbl}}$ sub-block of $\boldsymbol{\Sigma}_{\text{q}(\boldsymbol{\beta},\boldsymbol{u}_{\text{gbl}})}$

continued on a subsequent page ...

Algorithm 2 continued. This is a continuation of the description of this algorithm that commences on a preceding page

$\lambda_{\mathrm{q}(\sigma_\varepsilon^2)} \longleftarrow \mu_{\mathrm{q}(1/a_\varepsilon)}$; $\Lambda_{\mathrm{q}(\boldsymbol{\Sigma})} \longleftarrow \boldsymbol{M}_{\mathrm{q}(A_{\boldsymbol{\Sigma}}^{-1})}$; $\lambda_{\mathrm{q}(\sigma_{\mathrm{grp}}^2)} \longleftarrow \mu_{\mathrm{q}(1/a_{\mathrm{grp}})}$

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

$\boldsymbol{\mu}_{\mathrm{q}(\boldsymbol{u}_{\mathrm{lin},i}, \boldsymbol{u}_{\mathrm{grp},i})} \longleftarrow \boldsymbol{x}_{2,i}$ component of $\mathcal{S}_2$

$\boldsymbol{\Sigma}_{\mathrm{q}(\boldsymbol{u}_{\mathrm{lin},i}, \boldsymbol{u}_{\mathrm{grp},i})} \longleftarrow \boldsymbol{A}^{22,i}$ component of $\mathcal{S}_2$

$\boldsymbol{\mu}_{\mathrm{q}(\boldsymbol{u}_{\mathrm{lin},i})} \longleftarrow$ first 2 rows of $\boldsymbol{\mu}_{\mathrm{q}(\boldsymbol{u}_{\mathrm{lin},i}, \boldsymbol{u}_{\mathrm{grp},i})}$

$\boldsymbol{\Sigma}_{\mathrm{q}(\boldsymbol{u}_{\mathrm{lin},i})} \longleftarrow$ top left 2×2 sub-block of $\boldsymbol{\Sigma}_{\mathrm{q}(\boldsymbol{u}_{\mathrm{lin},i}, \boldsymbol{u}_{\mathrm{grp},i})}$

$\boldsymbol{\mu}_{\mathrm{q}(\boldsymbol{u}_{\mathrm{grp},i})} \longleftarrow$ last K_{grp} rows of $\boldsymbol{\mu}_{\mathrm{q}(\boldsymbol{u}_{\mathrm{lin},i}, \boldsymbol{u}_{\mathrm{grp},i})}$

$\boldsymbol{\Sigma}_{\mathrm{q}(\boldsymbol{u}_{\mathrm{grp},i})} \longleftarrow$ bottom right $K_{\mathrm{grp}} \times K_{\mathrm{grp}}$ sub-block of $\boldsymbol{\Sigma}_{\mathrm{q}(\boldsymbol{u}_{\mathrm{lin},i}, \boldsymbol{u}_{\mathrm{grp},i})}$

$$E_{\mathrm{q}}\left\{\left(\begin{bmatrix}\boldsymbol{\beta}\\ \boldsymbol{u}_{\mathrm{gbl}}\end{bmatrix} - \boldsymbol{\mu}_{\mathrm{q}(\boldsymbol{\beta}, \boldsymbol{u}_{\mathrm{gbl}})}\right)\left(\begin{bmatrix}\boldsymbol{u}_{\mathrm{lin},i}\\ \boldsymbol{u}_{\mathrm{grp},i}\end{bmatrix} - \boldsymbol{\mu}_{\mathrm{q}(\boldsymbol{u}_{\mathrm{lin},i}, \boldsymbol{u}_{\mathrm{grp},i}))}\right)^T\right\}$$

$$\longleftarrow \boldsymbol{A}^{12,i} \text{ component of } \mathcal{S}_2$$

$$\lambda_{\mathrm{q}(\sigma_\varepsilon^2)} \longleftarrow \lambda_{\mathrm{q}(\sigma_\varepsilon^2)} + \left\|\boldsymbol{y}_i - \mathbf{C}_{\mathrm{gbl},i}\boldsymbol{\mu}_{\mathrm{q}(\boldsymbol{\beta}, \boldsymbol{u}_{\mathrm{gbl}})} - \mathbf{C}_{\mathrm{grp},i}\boldsymbol{\mu}_{\mathrm{q}(\boldsymbol{u}_{\mathrm{lin},i}, \boldsymbol{u}_{\mathrm{grp},i})}\right\|^2$$

$$\lambda_{\mathrm{q}(\sigma_\varepsilon^2)} \longleftarrow \lambda_{\mathrm{q}(\sigma_\varepsilon^2)} + \mathrm{tr}\left(\mathbf{C}_{\mathrm{gbl},i}^T\mathbf{C}_{\mathrm{gbl},i}\boldsymbol{\Sigma}_{\mathrm{q}(\boldsymbol{\beta}, \boldsymbol{u}_{\mathrm{gbl}})}\right) + \mathrm{tr}\left(\mathbf{C}_{\mathrm{grp},i}^T\mathbf{C}_{\mathrm{grp},i}\boldsymbol{\Sigma}_{\mathrm{q}(\boldsymbol{u}_{\mathrm{lin},i}, \boldsymbol{u}_{\mathrm{grp},i})}\right)$$

$$\lambda_{\mathrm{q}(\sigma_\varepsilon^2)} \longleftarrow \lambda_{\mathrm{q}(\sigma_\varepsilon^2)}$$

$$+2\,\mathrm{tr}\left[\mathbf{C}_{\mathrm{grp},i}^T\mathbf{C}_{\mathrm{gbl},i}\,E_{\mathrm{q}}\left\{\left(\begin{bmatrix}\boldsymbol{\beta}\\ \boldsymbol{u}_{\mathrm{gbl}}\end{bmatrix} - \boldsymbol{\mu}_{\mathrm{q}(\boldsymbol{\beta}, \boldsymbol{u}_{\mathrm{gbl}})}\right)\left(\begin{bmatrix}\boldsymbol{u}_{\mathrm{lin},i}\\ \boldsymbol{u}_{\mathrm{grp},i}\end{bmatrix} - \boldsymbol{\mu}_{\mathrm{q}(\boldsymbol{u}_{\mathrm{lin},i}, \boldsymbol{u}_{\mathrm{grp},i})}\right)^T\right\}\right]$$

$$\boldsymbol{\Lambda}_{\mathrm{q}(\boldsymbol{\Sigma})} \longleftarrow \boldsymbol{\Lambda}_{\mathrm{q}(\boldsymbol{\Sigma})} + \boldsymbol{\mu}_{\mathrm{q}(\boldsymbol{u}_{\mathrm{lin},i})}\boldsymbol{\mu}_{\mathrm{q}(\boldsymbol{u}_{\mathrm{lin},i})}^T + \boldsymbol{\Sigma}_{\mathrm{q}(\boldsymbol{u}_{\mathrm{lin},i})}$$

$$\lambda_{\mathrm{q}(\sigma_{\mathrm{grp}}^2)} \longleftarrow \lambda_{\mathrm{q}(\sigma_{\mathrm{grp}}^2)} + \|\boldsymbol{\mu}_{\mathrm{q}(\boldsymbol{u}_{\mathrm{grp,i}})}\|^2 + \mathrm{tr}\left(\boldsymbol{\Sigma}_{\mathrm{q}(\boldsymbol{u}_{\mathrm{grp,i}})}\right)$$

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

$$\mu_{\mathrm{q}(1/\sigma_\varepsilon^2)} \leftarrow \xi_{\mathrm{q}(\sigma_\varepsilon)}/\lambda_{\mathrm{q}(\sigma_\varepsilon^2)} \quad ; \quad \mu_{\mathrm{q}(1/\sigma_{\mathrm{gbl}}^2)} \leftarrow \xi_{\mathrm{q}(\sigma_{\mathrm{gbl}}^2)}/\lambda_{\mathrm{q}(\sigma_{\mathrm{gbl}}^2)}$$

$$\boldsymbol{M}_{\mathrm{q}(\boldsymbol{\Sigma}^{-1})} \leftarrow (\xi_{\mathrm{q}(\boldsymbol{\Sigma})} - 1)\boldsymbol{\Lambda}_{\mathrm{q}(\boldsymbol{\Sigma})}^{-1} \quad ; \quad \mu_{\mathrm{q}(1/\sigma_{\mathrm{grp}}^2)} \leftarrow \xi_{\mathrm{q}(\sigma_{\mathrm{grp}}^2)}/\lambda_{\mathrm{q}(\sigma_{\mathrm{grp}}^2)}$$

$$\lambda_{\mathrm{q}(a_\varepsilon)} \longleftarrow \mu_{\mathrm{q}(1/\sigma_\varepsilon^2)} + 1/(\nu_\varepsilon s_\varepsilon^2) \quad ; \quad \mu_{\mathrm{q}(1/a_\varepsilon)} \longleftarrow \xi_{\mathrm{q}(a_\varepsilon)}/\lambda_{\mathrm{q}(a_\varepsilon)}$$

$$\boldsymbol{\Lambda}_{\mathrm{q}(A_{\boldsymbol{\Sigma}})} \longleftarrow \mathrm{diag}\left\{\mathrm{diagonal}\left(\boldsymbol{M}_{\mathrm{q}(\boldsymbol{\Sigma}^{-1})}\right)\right\} + \{\nu_{\boldsymbol{\Sigma}}\mathrm{diag}(s_{\boldsymbol{\Sigma},1}^2, s_{\boldsymbol{\Sigma},2}^2)\}^{-1}$$

$$\boldsymbol{M}_{\mathrm{q}(A_{\boldsymbol{\Sigma}}^{-1})} \longleftarrow \xi_{\mathrm{q}(A_{\boldsymbol{\Sigma}})}\boldsymbol{\Lambda}_{\mathrm{q}(A_{\boldsymbol{\Sigma}})}^{-1}$$

$$\lambda_{\mathrm{q}(a_{\mathrm{gbl}})} \longleftarrow \mu_{\mathrm{q}(1/\sigma_{\mathrm{gbl}}^2)} + 1/(\nu_{\mathrm{gbl}}s_{\mathrm{gbl}}^2) \quad ; \quad \mu_{\mathrm{q}(1/a_{\mathrm{gbl}})} \longleftarrow \xi_{\mathrm{q}(a_{\mathrm{gbl}})}/\lambda_{\mathrm{q}(a_{\mathrm{gbl}})}$$

$$\lambda_{\mathrm{q}(a_{\mathrm{grp}})} \longleftarrow \mu_{\mathrm{q}(1/\sigma_{\mathrm{grp}}^2)} + 1/(\nu_{\mathrm{grp}}s_{\mathrm{grp}}^2) \quad ; \quad \mu_{\mathrm{q}(1/a_{\mathrm{grp}})} \longleftarrow \xi_{\mathrm{q}(a_{\mathrm{grp}})}/\lambda_{\mathrm{q}(a_{\mathrm{grp}})}$$

until the increase in $\underline{\mathfrak{p}}(\boldsymbol{y}; \mathrm{q})$ is negligible.

Outputs: $\boldsymbol{\mu}_{\mathrm{q}(\boldsymbol{\beta}, \boldsymbol{u}_{\mathrm{gbl}})}$, $\boldsymbol{\Sigma}_{\mathrm{q}(\boldsymbol{\beta}, \boldsymbol{u}_{\mathrm{gbl}})}$, $\left\{\boldsymbol{\mu}_{\mathrm{q}(\boldsymbol{u}_{\mathrm{lin},i}, \boldsymbol{u}_{\mathrm{grp},i})}, \boldsymbol{\Sigma}_{\mathrm{q}(\boldsymbol{u}_{\mathrm{lin},i}, \boldsymbol{u}_{\mathrm{grp},i})},\right.$

$$\left. E_{\mathrm{q}}\left\{\left(\begin{bmatrix}\boldsymbol{\beta}\\ \boldsymbol{u}_{\mathrm{gbl}}\end{bmatrix} - \boldsymbol{\mu}_{\mathrm{q}(\boldsymbol{\beta}, \boldsymbol{u}_{\mathrm{gbl}})}\right)\left(\begin{bmatrix}\boldsymbol{u}_{\mathrm{lin},i}\\ \boldsymbol{u}_{\mathrm{grp},i}\end{bmatrix} - \boldsymbol{\mu}_{\mathrm{q}(\boldsymbol{u}_{\mathrm{lin},i}, \boldsymbol{u}_{\mathrm{grp},i})}\right)^T\right\} : 1 \le i \le m\right\},$$

$$\xi_{\mathrm{q}(\sigma_\varepsilon)}, \lambda_{\mathrm{q}(\sigma_\varepsilon^2)}, \xi_{\mathrm{q}(\sigma_{\mathrm{gbl}}^2)}, \lambda_{\mathrm{q}(\sigma_{\mathrm{gbl}}^2)}, \xi_{\mathrm{q}(\boldsymbol{\Sigma})}, \boldsymbol{\Lambda}_{\mathrm{q}(\boldsymbol{\Sigma})}^{-1}, \xi_{\mathrm{q}(\sigma_{\mathrm{grp}}^2)}, \lambda_{\mathrm{q}(\sigma_{\mathrm{grp}}^2)}.$$

and

$$\left.\begin{aligned} f^B(x) &= \beta_0^{\text{A}} + \beta_0^{\text{BvsA}} + (\beta_1^{\text{A}} + \beta_1^{\text{BvsA}})\, x + \sum_{k=1}^{K_{\text{gbl}}} u^B_{\text{gbl},k} z_{\text{gbl},k}(x) \\ g_i^B(x) &= u^B_{\text{lin},i0} + u^B_{\text{lin},i1}\, x + \sum_{k=1}^{K_{\text{grp}}} u^B_{\text{grp},ik} z_{\text{grp,k}}(x) \end{aligned}\right\} \text{ for category B.}$$

This allows us to estimate the global contrast function

$$c(x) \equiv f^B(x) - f^A(x) = \beta_0^{\text{BvsA}} + \beta_1^{\text{BvsA}}\, x + \sum_{k=1}^{K_{\text{gbl}}} (u^B_{\text{gbl},k} - u^A_{\text{gbl},k}) z_{\text{gbl},k}(x). \tag{2.17}$$

The distributions on the random coefficients are

$$[u^A_{\text{lin},i0}\ u^A_{\text{lin},i1}\ u^B_{\text{lin},i0}\ u^B_{\text{lin},i1}]^T \overset{\text{ind.}}{\sim} N(\mathbf{0}, \boldsymbol{\Sigma})$$

and

$$u^A_{\text{gbl},k} \overset{\text{ind.}}{\sim} N\big(0, (\sigma^A_{\text{gbl}})^2\big),\ \ u^B_{\text{gbl},k} \overset{\text{ind.}}{\sim} N\big(0, (\sigma^B_{\text{gbl}})^2\big),\ \ u^A_{\text{grp},ik} \overset{\text{ind.}}{\sim} N\big(0, \sigma^2_{\text{grp}}\big) \text{ and } u^B_{\text{grp},ik} \overset{\text{ind.}}{\sim} N\big(0, \sigma^2_{\text{grp}}\big)$$

independently of each other. In this two-category extension, the matrix $\boldsymbol{\Sigma}$ is an unstructured 4×4 covariance matrix.

Algorithms 1 and 2 can be used to achieve streamlined fitting and inference for the contrast curve extension, but with key matrices having new definitions. First, the $\boldsymbol{X}_i$, $\boldsymbol{Z}_{\text{gbl},i}$ and $\boldsymbol{Z}_{\text{grp},i}$ matrices need to be changed to:

$$\boldsymbol{X}_i = [\,\mathbf{1}\ \boldsymbol{x}_i\ \mathbf{1} - \boldsymbol{\iota}_i^A\ (\mathbf{1} - \boldsymbol{\iota}_i^A) \odot \boldsymbol{x}_i\,],$$

$$\boldsymbol{Z}_{\text{gbl},i} = [\,\boldsymbol{\iota}_i^A \odot z_{\text{gbl},1}(\boldsymbol{x}_i) \cdots \boldsymbol{\iota}_i^A \odot z_{\text{gbl},K_{\text{gbl}}}(\boldsymbol{x}_i)\ (\mathbf{1} - \boldsymbol{\iota}_i^A) \odot z_{\text{gbl},1}(\boldsymbol{x}_i) \cdots (\mathbf{1} - \boldsymbol{\iota}_i^A) \odot z_{\text{gbl},K_{\text{gbl}}}(\boldsymbol{x}_i)\,]$$

and

$$\boldsymbol{Z}_{\text{grp},i} = [\,\boldsymbol{\iota}_i^A \odot z_{\text{grp},1}(\boldsymbol{x}_i) \cdots \boldsymbol{\iota}_i^A \odot z_{\text{grp},K_{\text{grp}}}(\boldsymbol{x}_i)\ (\mathbf{1} - \boldsymbol{\iota}_i^A) \odot z_{\text{grp},1}(\boldsymbol{x}_i) \cdots (\mathbf{1} - \boldsymbol{\iota}_i^A) \odot z_{\text{grp},K_{\text{grp}}}(\boldsymbol{x}_i)\,].$$

Here $\boldsymbol{\iota}_i^A$ is the $n_i \times 1$ vector of ι_{ij}^A values, defined by (2.16), and $\boldsymbol{a} \odot \boldsymbol{b}$ denotes the element-wise product of equal-sized vectors $\boldsymbol{a}$ and $\boldsymbol{b}$.

In the case of best linear unbiased prediction, the updates for the $\boldsymbol{B}_i$ matrices in Algorithm 1 need to be replaced by:

$$\boldsymbol{B}_i \longleftarrow \begin{bmatrix} \sigma_\varepsilon^{-1}\boldsymbol{X}_i & \sigma_\varepsilon^{-1}\boldsymbol{Z}_{\text{gbl},i} \\ \boldsymbol{O} & m^{-1/2}\begin{bmatrix} (\sigma^A_{\text{gbl}})^{-1}\boldsymbol{I}_{K_{\text{gbl}}} & 0 \\ 0 & (\sigma^B_{\text{gbl}})^{-1}\boldsymbol{I}_{K_{\text{gbl}}} \end{bmatrix} \\ \boldsymbol{O} & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{O} \end{bmatrix}$$

and the output coefficient vectors change to

$$\begin{bmatrix} \widehat{\boldsymbol{\beta}} \\ \widehat{\boldsymbol{u}}^A_{\text{gbl}} \\ \widehat{\boldsymbol{u}}^B_{\text{gbl}} \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} \widehat{\boldsymbol{u}}^A_{\text{lin},i} \\ \widehat{\boldsymbol{u}}^B_{\text{lin},i} \\ \widehat{\boldsymbol{u}}^A_{\text{grp},i} \\ \widehat{\boldsymbol{u}}^B_{\text{grp},i} \end{bmatrix}.$$

In the case of mean field variational Bayes, the updates of the $\boldsymbol{B}_i$ matrices in Algorithm 2 need to be replaced by:

$$\boldsymbol{B}_i \longleftarrow \begin{bmatrix} \mu^{1/2}_{\mathfrak{q}(1/\sigma^2_\varepsilon)}\boldsymbol{X}_i & \mu^{1/2}_{\mathfrak{q}(1/\sigma^2_\varepsilon)}\boldsymbol{Z}_{\text{gbl},i} \\ m^{-1/2}\boldsymbol{\Sigma}_{\boldsymbol{\beta}}^{-1/2} & \boldsymbol{O} \\ \boldsymbol{O} & m^{-1/2}\begin{bmatrix} \mu^{1/2}_{\mathfrak{q}(1/(\sigma^A_{\text{gbl}})^2)}\boldsymbol{I}_{K_{\text{gbl}}} & 0 \\ 0 & \mu^{1/2}_{\mathfrak{q}(1/(\sigma^B_{\text{gbl}})^2)}\boldsymbol{I}_{K_{\text{gbl}}} \end{bmatrix} \\ \boldsymbol{O} & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{O} \end{bmatrix}.$$

A contrast curves adjustment to the mean field variational Bayes updates is also required for some of the covariance matrix parameters. However, these calculations are comparatively simple and analogous to those given in Section S.5 of the supplementary materials.

We now demonstrate the use of the contrast curves extension for data from a longitudinal study on adolescent somatic growth. The study was concerned with the mechanisms of human hypertension development and conducted at the Indiana University School of Medicine, Indianapolis, Indiana, USA. More details on these data can be found in Pratt et al. (1989). The data used is a subset for which only adolescents having 9 or more height measurements are included. The resulting number of subjects is 216, for which the number of height measurements ranges from 9 to 27 with an average of about 19 height measurements per subject. The variables

of interest are

y_{ij} = jth height measurement (centimetres) of subject i, and
x_{ij} = age (years) of subject i when y_{ij} is recorded,

for $1 \le i \le m$ and $1 \le j \le n_i$. The subjects are categorized into black ethnicity and white ethnicity and comparison of mean height between the two populations is of interest. Our group-specific curve model analyses involved spline bases of dimensions $K_{\text{gbl}} = 22$ and $K_{\text{grp}} = 12$. For the Bayesian fitting and inference, the data were transformed to have zero mean and unit standard deviation and hyperparameters were set to be

$$\boldsymbol{\mu}_{\boldsymbol{\beta}} = 0, \quad \boldsymbol{\Sigma}_{\boldsymbol{\beta}} = 10^{10}\boldsymbol{I}, \quad \nu_{\varepsilon} = \nu_{\text{gbl}} = \nu_{\text{grp}} = 1, \quad \nu_{\boldsymbol{\Sigma}} = 2, \quad s_{\varepsilon} = s^2_{\boldsymbol{\Sigma},1} = s^2_{\boldsymbol{\Sigma},2} = 10^5. \tag{2.18}$$

The fits were then back-transformed to correspond to the original units. These choices impose approximate noninformativity on the fixed effects vector $\boldsymbol{\beta}$ as well as all standard deviation and correlation parameters (Huang and Wand, 2013). In the common situation where there is no useful prior knowledge of model parameters, or when automation is important, we recommend such noninformative hyperparameter choices. The results are insensitive to large values of the hyperparameters appearing in (2.18).

The fits from Algorithm 2 are seen to have good agreement with the data in each sub-panel of the top two plots in Figure 2. The bottom panels of Figure 2 show the estimated height gap between black and white adolescents as a function of age. For the females, there is a significant height difference, with white adolescents taller than black adolescents, at 16–17 years old. Between 5 and 15 years, there is no obvious height difference. For the males, the difference is at its highest and statistically significant up to about 14 years of age, peaking at 13 years of age with black adolescents taller than white adolescents. Between 17 and 20 years old, there is no discernible height difference between the two populations.

3 Three-level models

The three-level version of group-specific curve models corresponds to curve-type data having two nested groupings. For example, the data in each panel of Figure 1 are first grouped according to slice, which is the level 2 group, and the slices are grouped according to tumour which is the level 3 group. We denote predictor/response pairs as (x_{ijk}, y_{ijk}) where x_{ijk} is the kth value of the predictor variable in the ith level 3 group and (i, j)th level 2 group and y_{ijk} is the corresponding value of the response variable. We let m denote the number of level 3 groups, n_i denote that number of level 2 groups in the ith level 3 group and o_{ij} denote the the number of units within the (i, j)th level 2 group. The Figure 1 data, which happen to be balanced, are such

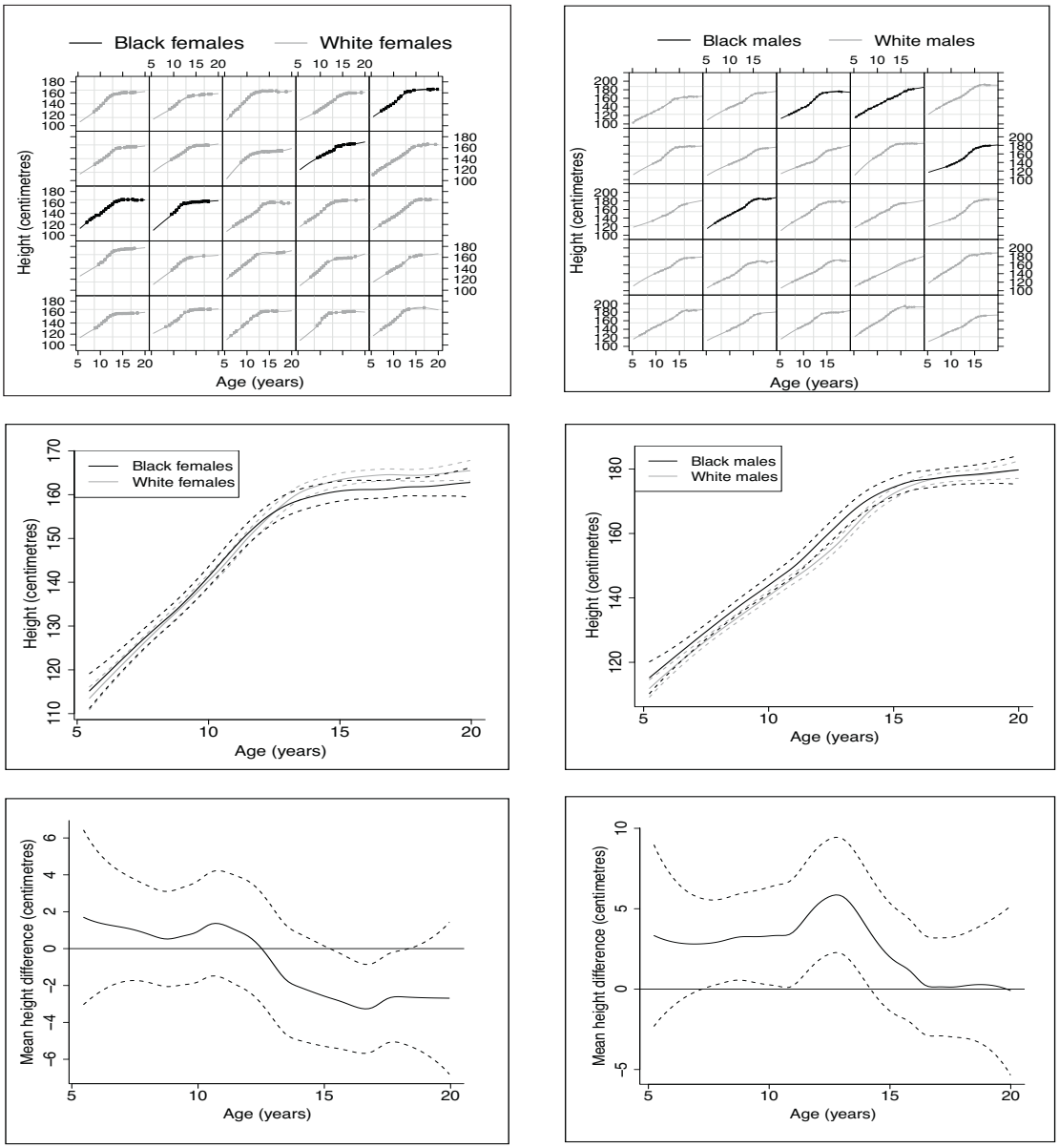

Figure 2 Top panels: Fitted group-specific growth curves for random subsets of 25 female subjects (left) and 25 male subjects (right) from the data on adolescent somatic growth (Pratt et al., 1989). The shading corresponds to approximate pointwise 99% credible intervals. Middle panels: Estimated mean growth curves with approximate 95% credible intervals. Bottom panels: Similar to the top panels but for the estimated contrast curve. Each contrast curve corresponds to mean height for black adolescents minus mean height for white adolescents. The left panel contrasts female adolescents and the right panel contrasts male adolescents. The shaded regions correspond to approximate pointwise 95% credible intervals

that

m = number of tumours = 10,
n_i = number of slices for the ith tumour = 5
and o_{ij} = number of predictor/response pairs for the ith tumour and jth slice = 128.

The Gaussian response three-level group-specific curve model for such data is

$$\begin{aligned} y_{ijk} &= f(x_{ijk}) + g_i(x_{ijk}) + h_{ij}(x_{ijk}) + \varepsilon_{ijk}, \quad \varepsilon_{ijk} \overset{\text{ind.}}{\sim} N(0, \sigma_\varepsilon^2), \\ & 1 \le i \le m,\ 1 \le j \le n_i,\ 1 \le k \le o_{ij}, \end{aligned} \tag{3.1}$$

where the smooth function f is the global mean function, the g_i functions, $1 \le i \le m$, allow for group-specific deviations according to membership of the ith level 3 group and the h_{ij}, $1 \le i \le m$ and $1 \le j \le n_i$ allow for an additional level of group-specific deviations according to membership of the jth level 2 group within the ith level 3 group. The mixed model-based penalized spline models for these functions are

$$f(x) = \beta_0 + \beta_1 x + \sum_{k=1}^{K_{\text{gbl}}} u_{\text{gbl},k}\, z_{\text{gbl},k}(x), \quad u_{\text{gbl},k} \overset{\text{ind.}}{\sim} N(0, \sigma_{\text{gbl}}^2),$$

$$g_i(x) = u^g_{\text{lin},i0} + u^g_{\text{lin},i1}\, x + \sum_{k=1}^{K^g_{\text{grp}}} u^g_{\text{grp},ik}\, z^g_{\text{grp,k}}(x), \quad \begin{bmatrix} u^g_{\text{lin},i0} \\ u^g_{\text{lin},i1} \end{bmatrix} \overset{\text{ind.}}{\sim} N(\mathbf{0}, \mathbf{\Sigma}_g), \quad u^g_{\text{grp},ik} \overset{\text{ind.}}{\sim} N\big(0, \sigma^2_{\text{grp},g}\big)$$

and

$$h_{ij}(x) = u^h_{\text{lin},ij0} + u^h_{\text{lin},ij1}\, x + \sum_{k=1}^{K^h_{\text{grp}}} u^h_{\text{grp},ijk}\, z^h_{\text{grp,k}}(x), \quad \begin{bmatrix} u^h_{\text{lin},ij0} \\ u^h_{\text{lin},ij1} \end{bmatrix} \overset{\text{ind.}}{\sim} N(\mathbf{0}, \mathbf{\Sigma}_h), \quad u^h_{\text{grp},ijk} \overset{\text{ind.}}{\sim} N\big(0, \sigma^2_{\text{grp},h}\big),$$

with all random effect distributions independent of each other. For this three-level case, we have three bases:

$$\{z_{\text{gbl},k}(\cdot) : 1 \le k \le K_{\text{gbl}}\}, \quad \{z^g_{\text{grp,k}}(\cdot) : 1 \le k \le K^g_{\text{grp}}\} \quad \text{and} \quad \{z^h_{\text{grp,k}}(\cdot) : 1 \le k \le K^h_{\text{grp}}\}.$$

The variance and covariance matrix parameters are analogous to the two-level model. For example, $\mathbf{\Sigma}_g$ and $\mathbf{\Sigma}_h$ are both unstructured 2×2 matrices corresponding to the linear components of the g_i and h_{ij}, respectively.

The following notation is useful for setting up the required design matrices: If $\boldsymbol{M}_1, \ldots, \boldsymbol{M}_d$ is a set of matrices each having the same number of columns then

$$\operatorname*{stack}_{1 \le i \le d}(\boldsymbol{M}_i) \equiv \begin{bmatrix} \boldsymbol{M}_1 \\ \vdots \\ \boldsymbol{M}_d \end{bmatrix}.$$

We then define, for $1 \leq i \leq m$ and $1 \leq j \leq n_i$,

$$\boldsymbol{x}_i \equiv \underset{1\leq j\leq n_i}{\text{stack}}\left(\boldsymbol{x}_{ij}\right) \quad \text{and} \quad \boldsymbol{x}_{ij} \equiv \underset{1\leq k\leq o_{ij}}{\text{stack}}\left(x_{ijk}\right).$$

3.1 Best linear unbiased prediction

Model (3.1) is expressible as a Gaussian response linear mixed model as follows:

$$\boldsymbol{y}|\boldsymbol{u} \sim N(\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}\boldsymbol{u}, \sigma_\varepsilon^2 \boldsymbol{I}), \quad \boldsymbol{u} \sim N(\boldsymbol{0}, \boldsymbol{G}), \tag{3.2}$$

where the design matrices are

$$\boldsymbol{X} = \underset{1\leq i\leq m}{\text{stack}}\left(\boldsymbol{X}_i\right) \quad \text{with} \quad \boldsymbol{X}_i = \underset{1\leq j\leq n_i}{\text{stack}}\left(\boldsymbol{X}_{ij}\right) \quad \text{and} \quad \boldsymbol{X}_{ij} \equiv [\boldsymbol{1}\ \boldsymbol{x}_{ij}]$$

and

$$\boldsymbol{Z} \equiv \left[\boldsymbol{Z}_{\text{gbl}}\ \underset{1\leq i\leq m}{\text{blockdiag}}\left[\underset{1\leq j\leq n_i}{\text{stack}}([\boldsymbol{X}_{ij}\ \boldsymbol{Z}^g_{\text{grp},ij}])\ \underset{1\leq j\leq n_i}{\text{blockdiag}}([\boldsymbol{X}_{ij}\ \boldsymbol{Z}^h_{\text{grp},ij}])\right]\right],$$

where

$$\boldsymbol{Z}_{\text{gbl}} \equiv \underset{1\leq i\leq m}{\text{stack}}\left(\underset{1\leq j\leq n_i}{\text{stack}}(\boldsymbol{Z}_{\text{gbl},ij})\right)$$

and the matrices $\boldsymbol{Z}_{\text{gbl},ij}$, $\boldsymbol{Z}^g_{\text{grp},ij}$ and $\boldsymbol{Z}^h_{\text{grp},ij}$, $1 \leq i \leq m$, $1 \leq j \leq n_i$, contain, respectively, spline basis functions for the global mean function f, the ith level one group deviation functions g_i and (i, j)th level two group deviation functions h_{ij}. Specifically,

$$\boldsymbol{Z}_{\text{gbl},ij} \equiv \left[\, z_{\text{gbl},1}(\boldsymbol{x}_{ij}) \cdots z_{\text{gbl},K_{\text{gbl}}}(\boldsymbol{x}_{ij}) \,\right], \quad \boldsymbol{Z}^g_{\text{grp},ij} = \left[z^g_{\text{grp},1}(\boldsymbol{x}_{ij}) \cdots z^g_{\text{grp},K^g_{\text{grp}}}(\boldsymbol{x}_{ij})\right]$$

$$\text{and } \boldsymbol{Z}^h_{\text{grp},ij} \equiv \left[z^h_{\text{grp},1}(\boldsymbol{x}_{ij}) \cdots z^h_{\text{grp},K^h_{\text{grp}}}(\boldsymbol{x}_{ij})\right] \quad \text{for } 1 \leq i \leq m \text{ and } 1 \leq j \leq n_i.$$

The fixed and random effects vectors are

$$\boldsymbol{\beta} \equiv \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} \quad \text{and} \quad \boldsymbol{u} \equiv \begin{bmatrix} \boldsymbol{u}_{\text{gbl}} \\ \underset{1\leq i\leq m}{\text{stack}}\left(\begin{bmatrix} \begin{bmatrix} \boldsymbol{u}^g_{\text{lin},i} \\ \boldsymbol{u}^g_{\text{grp},i} \end{bmatrix} \\ \begin{bmatrix} \boldsymbol{u}^h_{\text{lin},i1} \\ \boldsymbol{u}^h_{\text{grp},i1} \end{bmatrix} \\ \vdots \\ \begin{bmatrix} \boldsymbol{u}^h_{\text{lin},in_i} \\ \boldsymbol{u}^h_{\text{grp},in_i} \end{bmatrix} \end{bmatrix}\right) \end{bmatrix} \quad \text{where} \quad \boldsymbol{u}^g_{\text{lin},i} \equiv \begin{bmatrix} u^g_{\text{lin},i0} \\ u^g_{\text{lin},i1} \end{bmatrix}$$

with $\boldsymbol{u}^g_{\text{grp},i}$, $\boldsymbol{u}^b_{\text{lin},ij}$ and $\boldsymbol{u}^b_{\text{grp},ij}$ defined similarly and the covariance matrix of $\boldsymbol{u}$ is

$$\boldsymbol{G} = \text{Cov}(\boldsymbol{u}) = \begin{bmatrix} \sigma^2_{\text{gbl}}\boldsymbol{I} & & \boldsymbol{O} \\ & & \\ \boldsymbol{O} & \underset{1\le i\le m}{\text{blockdiag}} & \begin{bmatrix} \boldsymbol{\Sigma}_g & \boldsymbol{O} & \boldsymbol{O} \\ \boldsymbol{O} & \sigma^2_{\text{grp},g}\boldsymbol{I} & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{O} & \boldsymbol{I}_{n_i} \otimes \begin{bmatrix} \boldsymbol{\Sigma}_b & \boldsymbol{O} \\ \boldsymbol{O} & \sigma^2_{\text{grp},b}\boldsymbol{I} \end{bmatrix} \end{bmatrix} \end{bmatrix}. \tag{3.3}$$

We define matrices in a similar way to what is given in (2.5). The best linear unbiased predictor of $[\boldsymbol{\beta}\ \boldsymbol{u}]$ and corresponding covariance matrix are as shown in (2.6), but with entries as described in this section. This covariance matrix grows quadratically in both m and the n_is, and so storage becomes impractical for large numbers of level 2 and level 3 groups. However, only certain sub-blocks are required for the addition of pointwise confidence intervals to curve estimates. In particular, we only require the sub-matrices of $(\boldsymbol{C}^T\boldsymbol{R}^{-1}_{\text{BLUP}}\boldsymbol{C} + \boldsymbol{D}_{\text{BLUP}})^{-1}$ that correspond to the non-zero sub-blocks of the general three-level sparse matrix given in Section 3 of Nolan and Wand (2020).

In the case of the three-level Gaussian response linear mixed model, Nolan and Wand's

$\boldsymbol{A}_{11}$ sub-block corresponds to a $(2 + K_{\text{gbl}}) \times (2 + K_{\text{gbl}})$ matrix $\text{Cov}\left(\begin{bmatrix} \widehat{\boldsymbol{\beta}} \\ \widehat{\boldsymbol{u}}_{\text{gbl}} - \boldsymbol{u}_{\text{gbl}} \end{bmatrix}\right)$;

$\boldsymbol{A}_{22,i}$ sub-block corresponds to a $(2 + K^g_{\text{grp}}) \times (2 + K^g_{\text{grp}})$ matrix $\text{Cov}\left(\begin{bmatrix} \widehat{\boldsymbol{u}}^g_{\text{lin},i} - \boldsymbol{u}^g_{\text{lin},i} \\ \widehat{\boldsymbol{u}}^g_{\text{grp},i} - \boldsymbol{u}^g_{\text{grp},i} \end{bmatrix}\right)$;

$\boldsymbol{A}_{12,i}$ sub-block corresponds to a $(2 + K_{\text{gbl}}) \times (2 + K^g_{\text{grp}})$ matrix

$$E\left\{\begin{bmatrix} \widehat{\boldsymbol{\beta}} \\ \widehat{\boldsymbol{u}}_{\text{gbl}} - \boldsymbol{u}_{\text{gbl}} \end{bmatrix}\begin{bmatrix} \widehat{\boldsymbol{u}}^g_{\text{lin},i} - \boldsymbol{u}^g_{\text{lin},i} \\ \widehat{\boldsymbol{u}}^g_{\text{grp},i} - \boldsymbol{u}^g_{\text{grp},i} \end{bmatrix}^T\right\},\ 1 \le i \le m;$$

$\boldsymbol{A}_{22,ij}$ sub-block corresponds to a $(2 + K^b_{\text{grp}}) \times (2 + K^b_{\text{grp}})$ matrix $\text{Cov}\left(\begin{bmatrix} \widehat{\boldsymbol{u}}^b_{\text{lin},ij} - \boldsymbol{u}^b_{\text{lin},ij} \\ \widehat{\boldsymbol{u}}^b_{\text{grp},ij} - \boldsymbol{u}^b_{\text{grp},ij} \end{bmatrix}\right)$;

$\boldsymbol{A}_{12,ij}$ sub-block corresponds to a $(2 + K_{\text{gbl}}) \times (2 + K^b_{\text{grp}})$ matrix

$$E\left\{\begin{bmatrix} \widehat{\boldsymbol{\beta}} \\ \widehat{\boldsymbol{u}}_{\text{gbl}} - \boldsymbol{u}_{\text{gbl}} \end{bmatrix}\begin{bmatrix} \widehat{\boldsymbol{u}}^b_{\text{lin},ij} - \boldsymbol{u}^b_{\text{lin},ij} \\ \widehat{\boldsymbol{u}}^b_{\text{grp},ij} - \boldsymbol{u}^b_{\text{grp},ij} \end{bmatrix}^T\right\};$$

$\boldsymbol{A}_{12,i,j}$ sub-block corresponds to a $(2 + K^g_{\text{grp}}) \times (2 + K^b_{\text{grp}})$ matrix

$$E\left\{\begin{bmatrix} \widehat{\boldsymbol{u}}^g_{\text{lin},i} - \boldsymbol{u}^g_{\text{lin},i} \\ \widehat{\boldsymbol{u}}^g_{\text{grp},i} - \boldsymbol{u}^g_{\text{grp},i} \end{bmatrix}\begin{bmatrix} \widehat{\boldsymbol{u}}^b_{\text{lin},ij} - \boldsymbol{u}^b_{\text{lin},ij} \\ \widehat{\boldsymbol{u}}^b_{\text{grp},ij} - \boldsymbol{u}^b_{\text{grp},ij} \end{bmatrix}^T\right\},\ 1 \le i \le m,\ 1 \le j \le n_i.$$

As described in Nolan et al. (2020), the *SolveThreeLevelSparseLeastSquares* algorithm arises in the special case where $\boldsymbol{x}$ is the minimizer of the least squares problem given in equation (2.8), where $\boldsymbol{B}$ has the three-level sparse form and $\boldsymbol{b}$

is partitioned according to that given in Section 3 of Nolan and Wand (2020). This algorithm can be used for fitting three-level group-specific curve models by making use of Result 3. Note that Result 3 relies on notation for sub-blocks of $\boldsymbol{x} = (\boldsymbol{B}^T\boldsymbol{B})^{-1}\boldsymbol{B}^T\boldsymbol{b}$ and $\boldsymbol{A}^{-1}$ where $\boldsymbol{A} = \boldsymbol{B}^T\boldsymbol{B}$. This notation is too verbose to list here, but is given by (S.8) of the supplementary materials, as well as in Section 3 of Nolan and Wand (2020).

Result 3 *Computation of* $[\widehat{\boldsymbol{\beta}}^T \ \widehat{\boldsymbol{u}}^T]^T$ *and each of the sub-blocks of* $\text{Cov}([\widehat{\boldsymbol{\beta}}^T \ (\widehat{\boldsymbol{u}} - \boldsymbol{u})^T]^T)$ *listed in (2.7) are expressible as the three-level sparse matrix least squares form:*

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

where according to the notation in Section 3.1 of Nolan and Wand (2020), the non-zero sub-blocks of $\boldsymbol{B}$ *and* $\boldsymbol{b}$ *for* $1 \le i \le m$ *and* $1 \le j \le n_i$ *are as follows:*

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

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

with each of these matrices having $\tilde{o}_{ij} = o_{ij} + K_{\text{gbl}} + 2 + K^g_{\text{grp}} + 2 + K^h_{\text{grp}}$ *rows and with* $\boldsymbol{B}_i$ *having* $p = 2 + K_{\text{gbl}}$ *columns,* $\dot{\boldsymbol{B}}_i$ *having* $q_1 = 2 + K^g_{\text{grp}}$ *columns and* $\ddot{\boldsymbol{B}}_{ij}$ *having* $q_2 = 2 + K^h_{\text{grp}}$ *columns. The solutions are, with sub-block labelling according to (S.8) of the supplementary materials and Section 3 of Nolan and Wand (2020),*

$$\begin{bmatrix} \widehat{\boldsymbol{\beta}} \\ \widehat{\boldsymbol{u}}_{\text{gbl}} \end{bmatrix} = \boldsymbol{x}_1, \quad \text{Cov}\left(\begin{bmatrix} \widehat{\boldsymbol{\beta}} \\ \widehat{\boldsymbol{u}}_{\text{gbl}} - \boldsymbol{u}_{\text{gbl}} \end{bmatrix}\right) = \boldsymbol{A}^{11},$$

$$\begin{bmatrix} \widehat{\boldsymbol{u}}^g_{\text{lin},i} \\ \widehat{\boldsymbol{u}}^g_{\text{grp},i} \end{bmatrix} = \boldsymbol{x}_{2,i}, \quad E\left\{ \begin{bmatrix} \widehat{\boldsymbol{\beta}} \\ \widehat{\boldsymbol{u}}_{\text{gbl}} - \boldsymbol{u}_{\text{gbl}} \end{bmatrix} \begin{bmatrix} \widehat{\boldsymbol{u}}^g_{\text{lin},i} - \boldsymbol{u}^g_{\text{lin},i} \\ \widehat{\boldsymbol{u}}^g_{\text{grp},i} - \boldsymbol{u}^g_{\text{grp},i} \end{bmatrix}^T \right\} = \boldsymbol{A}^{12,i},$$

$$\text{Cov}\left(\begin{bmatrix} \widehat{\boldsymbol{u}}^g_{\text{lin},i} - \boldsymbol{u}^g_{\text{lin},i} \\ \widehat{\boldsymbol{u}}^g_{\text{grp},i} - \boldsymbol{u}^g_{\text{grp},i} \end{bmatrix} \right) = \boldsymbol{A}^{22,i}, \quad 1 \le i \le m,$$

$$\begin{bmatrix} \widehat{\boldsymbol{u}}^b_{\text{lin},ij} \\ \widehat{\boldsymbol{u}}^b_{\text{grp},ij} \end{bmatrix} = \boldsymbol{x}_{2,ij}, \quad E\left\{ \begin{bmatrix} \widehat{\boldsymbol{\beta}} \\ \widehat{\boldsymbol{u}}_{\text{gbl}} - \boldsymbol{u}_{\text{gbl}} \end{bmatrix} \begin{bmatrix} \widehat{\boldsymbol{u}}^b_{\text{lin},ij} - \boldsymbol{u}^b_{\text{lin},ij} \\ \widehat{\boldsymbol{u}}^b_{\text{grp},ij} - \boldsymbol{u}^b_{\text{grp},ij} \end{bmatrix}^T \right\} = \boldsymbol{A}^{12,ij},$$

$$E\left\{ \begin{bmatrix} \widehat{\boldsymbol{u}}^g_{\text{lin},i} - \boldsymbol{u}^g_{\text{lin},i} \\ \widehat{\boldsymbol{u}}^g_{\text{grp},i} - \boldsymbol{u}^g_{\text{grp},i} \end{bmatrix} \begin{bmatrix} \widehat{\boldsymbol{u}}^b_{\text{lin},ij} - \boldsymbol{u}^b_{\text{lin},ij} \\ \widehat{\boldsymbol{u}}^b_{\text{grp},ij} - \boldsymbol{u}^b_{\text{grp},ij} \end{bmatrix}^T \right\} = \boldsymbol{A}^{12,i,j}$$

and

$$\text{Cov}\left(\begin{bmatrix} \widehat{\boldsymbol{u}}^b_{\text{lin},ij} - \boldsymbol{u}^b_{\text{lin},ij} \\ \widehat{\boldsymbol{u}}^b_{\text{grp},ij} - \boldsymbol{u}^b_{\text{grp},ij} \end{bmatrix} \right) = \boldsymbol{A}^{22,ij}, \quad 1 \le i \le m, \ 1 \le j \le n_i.$$

A derivation of Result 3 is given in Section S.7 of the supplementary materials. Result 3 combined with Theorem 3.3 of Nolan and Wand (2020) leads to Algorithm 3. The SolveThreeLevelSparseLeastSquares algorithm is given in Section S.13 of the supplementary materials.

3.2 Mean field variational bayes

A Bayesian extension of (3.2) and (3.3) is as follows:

$$\boldsymbol{y}|\boldsymbol{\beta}, \boldsymbol{u}, \sigma^2_\varepsilon \sim N(\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}\boldsymbol{u}, \sigma^2_\varepsilon \boldsymbol{I}), \quad \boldsymbol{u}|\sigma^2_{\text{gbl}}, \sigma^2_{\text{grp},g}, \boldsymbol{\Sigma}_g, \sigma^2_{\text{grp},h}, \boldsymbol{\Sigma}_h \sim N(\boldsymbol{0}, \boldsymbol{G}), \quad \boldsymbol{G} \text{ as defined in (3.3)},$$

$$\boldsymbol{\beta} \sim N(\boldsymbol{\mu}_{\boldsymbol{\beta}}, \boldsymbol{\Sigma}_{\boldsymbol{\beta}}), \quad \sigma^2_\varepsilon|a_\varepsilon \sim \text{Inverse-}\chi^2(\nu_\varepsilon, 1/a_\varepsilon), \quad a_\varepsilon \sim \text{Inverse-}\chi^2(1, 1/(\nu_\varepsilon s^2_\varepsilon)),$$

$$\sigma^2_{\text{gbl}}|a_{\text{gbl}} \sim \text{Inverse-}\chi^2(\nu_{\text{gbl}}, 1/a_{\text{gbl}}), \quad a_{\text{gbl}} \sim \text{Inverse-}\chi^2(1, 1/(\nu_{\text{gbl}} s^2_{\text{gbl}})),$$

$$\sigma^2_{\text{grp},g}|a_{\text{grp},g} \sim \text{Inverse-}\chi^2(\nu_{\text{grp},g}, 1/a_{\text{grp},g}), \quad a_{\text{grp},g} \sim \text{Inverse-}\chi^2(1, 1/(\nu_{\text{grp},g} s^2_{\text{grp},g})),$$

$$\sigma^2_{\text{grp},h}|a_{\text{grp},h} \sim \text{Inverse-}\chi^2(\nu_{\text{grp},h}, 1/a_{\text{grp},h}), \quad a_{\text{grp},h} \sim \text{Inverse-}\chi^2(1, 1/(\nu_{\text{grp},h} s^2_{\text{grp},h})),$$

$$
\begin{aligned}
\boldsymbol{\Sigma}_g|\boldsymbol{A}_{\boldsymbol{\Sigma}_g} &\sim \text{Inverse-G-Wishart}\Big(G_{\text{full}},\, \nu_{\boldsymbol{\Sigma}_g}+2, \boldsymbol{A}_{\boldsymbol{\Sigma}_g}^{-1}\Big),\\
\boldsymbol{A}_{\boldsymbol{\Sigma}_g} &\sim \text{Inverse-G-Wishart}(G_{\text{diag}},\, 1,\, \boldsymbol{\Lambda}_{\boldsymbol{A}_{\boldsymbol{\Sigma}_g}}), \quad \boldsymbol{\Lambda}_{\boldsymbol{A}_{\boldsymbol{\Sigma}_g}} \equiv \big\{\nu_{\boldsymbol{\Sigma}_g}\text{diag}(s^2_{\boldsymbol{\Sigma}_g,1}, s^2_{\boldsymbol{\Sigma}_g,2})\big\}^{-1},\\
\boldsymbol{\Sigma}_h|\boldsymbol{A}_{\boldsymbol{\Sigma}_h} &\sim \text{Inverse-G-Wishart}\Big(G_{\text{full}},\, \nu_{\boldsymbol{\Sigma}_h}+2, \boldsymbol{A}_{\boldsymbol{\Sigma}_h}^{-1}\Big),\\
\boldsymbol{A}_{\boldsymbol{\Sigma}_h} &\sim \text{Inverse-G-Wishart}(G_{\text{diag}},\, 1,\, \boldsymbol{\Lambda}_{\boldsymbol{A}_{\boldsymbol{\Sigma}_h}}), \quad \boldsymbol{\Lambda}_{\boldsymbol{A}_{\boldsymbol{\Sigma}_h}} \equiv \big\{\nu_{\boldsymbol{\Sigma}_h}\text{diag}(s^2_{\boldsymbol{\Sigma}_h,1}, s^2_{\boldsymbol{\Sigma}_h,2})\big\}^{-1}.
\end{aligned}
\tag{3.4}
$$

Algorithm 3 Streamlined algorithm for obtaining best linear unbiased predictions and corresponding covariance matrix components for the three-level group specific curves model

Inputs: $\boldsymbol{y}_{ij}(o_{ij}\times 1)$, $\boldsymbol{X}_{ij}(o_{ij}\times 2)$, $\boldsymbol{Z}_{\text{gbl},ij}(o_{ij}\times K_{\text{gbl}})$, $\boldsymbol{Z}^g_{\text{grp},ij}(o_{ij}\times K^g_{\text{grp}})$,
$\boldsymbol{Z}^h_{\text{grp},ij}(o_{ij}\times K^h_{\text{grp}})$, $1\le i\le m$, $1\le j\le n_i$; $\sigma^2_\varepsilon, \sigma^2_{\text{gbl}}, \sigma^2_{\text{grp},g}, \sigma^2_{\text{grp},h} > 0$,
$\boldsymbol{\Sigma}_g(2\times 2)$, $\boldsymbol{\Sigma}_h(2\times 2)$, symmetric and positive definite.

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

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

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

$$
\dot{\boldsymbol{B}}_{ij} \longleftarrow \begin{bmatrix} \sigma_\varepsilon^{-1}\boldsymbol{X}_{ij} & \sigma_\varepsilon^{-1}\boldsymbol{Z}^g_{\text{grp},ij} \\ \boldsymbol{O} & \boldsymbol{O} \\ n_i^{-1/2}\boldsymbol{\Sigma}_g^{-1/2} & \boldsymbol{O} \\ \boldsymbol{O} & n_i^{-1/2}\sigma^{-1}_{\text{grp},g}\boldsymbol{I}_{K^g_{\text{grp}}} \\ \boldsymbol{O} & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{O} \end{bmatrix};\ \ddot{\boldsymbol{B}}_{ij} \longleftarrow \begin{bmatrix} \sigma_\varepsilon^{-1}\boldsymbol{X}_{ij} & \sigma_\varepsilon^{-1}\boldsymbol{Z}^h_{\text{grp},ij} \\ \boldsymbol{O} & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{O} \\ \boldsymbol{\Sigma}_h^{-1/2} & \boldsymbol{O} \\ \boldsymbol{O} & \sigma^{-1}_{\text{grp},h}\boldsymbol{I}_{K^h_{\text{grp}}} \end{bmatrix}
$$

$$
\mathcal{S}_3 \longleftarrow \textit{SolveThreeLevelSparseLeastSquares}\Big(\big\{(\boldsymbol{b}_{ij}, \boldsymbol{B}_{ij}, \dot{\boldsymbol{B}}_{ij}, \ddot{\boldsymbol{B}}_{ij}) : 1\le i\le m,\ 1\le j\le n_i\big\}\Big)
$$

$$
\begin{bmatrix} \widehat{\boldsymbol{\beta}} \\ \widehat{\boldsymbol{u}}_{\text{gbl}} \end{bmatrix} \longleftarrow \boldsymbol{x}_1 \text{ component of } \mathcal{S}_3\ ;\ \text{Cov}\left(\begin{bmatrix} \widehat{\boldsymbol{\beta}} \\ \widehat{\boldsymbol{u}}_{\text{gbl}} - \boldsymbol{u}_{\text{gbl}} \end{bmatrix}\right) \longleftarrow \boldsymbol{A}^{11} \text{ component of } \mathcal{S}_3
$$

continued on a subsequent page ...

Algorithm 3 continued. This is a continuation of the description of this algorithm that commences on a preceding page

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

$$\begin{bmatrix} \widehat{\boldsymbol{u}}^g_{\text{lin},i} \\ \widehat{\boldsymbol{u}}^g_{\text{grp},i} \end{bmatrix} \longleftarrow \boldsymbol{x}_{2,i} \text{ component of } \mathcal{S}_3$$

$$\text{Cov}\left(\begin{bmatrix} \widehat{\boldsymbol{u}}^g_{\text{lin},i} - \boldsymbol{u}^g_{\text{lin},i} \\ \widehat{\boldsymbol{u}}^g_{\text{grp},i} - \boldsymbol{u}^g_{\text{grp},i} \end{bmatrix}\right) \longleftarrow A^{22,i} \text{ component of } \mathcal{S}_3$$

$$E\left\{\begin{bmatrix} \widehat{\boldsymbol{\beta}} \\ \widehat{\boldsymbol{u}}_{\text{gbl}} - \boldsymbol{u}_{\text{gbl}} \end{bmatrix}\begin{bmatrix} \widehat{\boldsymbol{u}}^g_{\text{lin},i} - \boldsymbol{u}^g_{\text{lin},i} \\ \widehat{\boldsymbol{u}}^g_{\text{grp},i} - \boldsymbol{u}^g_{\text{grp},i} \end{bmatrix}^T\right\} \longleftarrow A^{12,i} \text{ component of } \mathcal{S}_3$$

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

$$\begin{bmatrix} \widehat{\boldsymbol{u}}^h_{\text{lin},ij} \\ \widehat{\boldsymbol{u}}^h_{\text{grp},ij} \end{bmatrix} \longleftarrow \boldsymbol{x}_{2,ij} \text{ component of } \mathcal{S}_3$$

$$E\left\{\begin{bmatrix} \widehat{\boldsymbol{\beta}} \\ \widehat{\boldsymbol{u}}_{\text{gbl}} - \boldsymbol{u}_{\text{gbl}} \end{bmatrix}\begin{bmatrix} \widehat{\boldsymbol{u}}^h_{\text{lin},ij} - \boldsymbol{u}^h_{\text{lin},ij} \\ \widehat{\boldsymbol{u}}^h_{\text{grp},ij} - \boldsymbol{u}^h_{\text{grp},ij} \end{bmatrix}^T\right\} \longleftarrow A^{12,ij} \text{ component of } \mathcal{S}_3$$

$$E\left\{\begin{bmatrix} \widehat{\boldsymbol{u}}^g_{\text{lin},i} - \boldsymbol{u}^g_{\text{lin},i} \\ \widehat{\boldsymbol{u}}^g_{\text{grp},i} - \boldsymbol{u}^g_{\text{grp},i} \end{bmatrix}\begin{bmatrix} \widehat{\boldsymbol{u}}^h_{\text{lin},ij} - \boldsymbol{u}^h_{\text{lin},ij} \\ \widehat{\boldsymbol{u}}^h_{\text{grp},ij} - \boldsymbol{u}^h_{\text{grp},ij} \end{bmatrix}^T\right\} \longleftarrow A^{12,\,i,\,j} \text{ component of } \mathcal{S}_3$$

$$\text{Cov}\left(\begin{bmatrix} \widehat{\boldsymbol{u}}^h_{\text{lin},ij} - \boldsymbol{u}^h_{\text{lin},ij} \\ \widehat{\boldsymbol{u}}^h_{\text{grp},ij} - \boldsymbol{u}^h_{\text{grp},ij} \end{bmatrix}\right) \longleftarrow A^{22,ij} \text{ component of } \mathcal{S}_3$$

Output:

$$\left(\begin{bmatrix} \widehat{\boldsymbol{\beta}} \\ \widehat{\boldsymbol{u}}_{\text{gbl}} \end{bmatrix},\ \text{Cov}\left(\begin{bmatrix} \widehat{\boldsymbol{\beta}} \\ \widehat{\boldsymbol{u}}_{\text{gbl}} - \boldsymbol{u}_{\text{gbl}} \end{bmatrix}\right), \left\{\left(\begin{bmatrix} \widehat{\boldsymbol{u}}_{\text{lin},i} \\ \widehat{\boldsymbol{u}}_{\text{grp},i} \end{bmatrix},\ \text{Cov}\left(\begin{bmatrix} \widehat{\boldsymbol{u}}_{\text{lin},i} - \boldsymbol{u}_{\text{lin},i} \\ \widehat{\boldsymbol{u}}_{\text{grp},i} - \boldsymbol{u}_{\text{grp},i} \end{bmatrix}\right),\right.\right.$$
$$\left.E\left\{\begin{bmatrix} \widehat{\boldsymbol{\beta}} \\ \widehat{\boldsymbol{u}}_{\text{gbl}} - \boldsymbol{u}_{\text{gbl}} \end{bmatrix}\begin{bmatrix} \widehat{\boldsymbol{u}}_{\text{lin},i} - \boldsymbol{u}_{\text{lin},i} \\ \widehat{\boldsymbol{u}}_{\text{grp},i} - \boldsymbol{u}_{\text{grp},i} \end{bmatrix}^T\right\}\right) :\ 1 \le i \le m,$$
$$\left(\begin{bmatrix} \widehat{\boldsymbol{u}}^h_{\text{lin},ij} \\ \widehat{\boldsymbol{u}}^h_{\text{grp},ij} \end{bmatrix},\ E\left\{\begin{bmatrix} \widehat{\boldsymbol{\beta}} \\ \widehat{\boldsymbol{u}}_{\text{gbl}} - \boldsymbol{u}_{\text{gbl}} \end{bmatrix}\begin{bmatrix} \widehat{\boldsymbol{u}}^h_{\text{lin},ij} - \boldsymbol{u}^h_{\text{lin},ij} \\ \widehat{\boldsymbol{u}}^h_{\text{grp},ij} - \boldsymbol{u}^h_{\text{grp},ij} \end{bmatrix}^T\right\},\right.$$
$$\left.\left.E\left\{\begin{bmatrix} \widehat{\boldsymbol{u}}^g_{\text{lin},i} - \boldsymbol{u}^g_{\text{lin},i} \\ \widehat{\boldsymbol{u}}^g_{\text{grp},i} - \boldsymbol{u}^g_{\text{grp},i} \end{bmatrix}\begin{bmatrix} \widehat{\boldsymbol{u}}^h_{\text{lin},ij} - \boldsymbol{u}^h_{\text{lin},ij} \\ \widehat{\boldsymbol{u}}^h_{\text{grp},ij} - \boldsymbol{u}^h_{\text{grp},ij} \end{bmatrix}^T\right\},\ \text{Cov}\left(\begin{bmatrix} \widehat{\boldsymbol{u}}^h_{\text{lin},ij} - \boldsymbol{u}^h_{\text{lin},ij} \\ \widehat{\boldsymbol{u}}^h_{\text{grp},ij} - \boldsymbol{u}^h_{\text{grp},ij} \end{bmatrix}\right)\right) :$$
$$\left.1 \le i \le m,\ 1 \le j \le n_i\right\}\Bigg)$$

The following mean field restriction is imposed on the joint posterior density function of all parameters in (3.4):

$$\begin{aligned} &\mathfrak{p}(\boldsymbol{\beta}, \boldsymbol{u}, a_\varepsilon, a_{\text{gbl}}, a_{\text{grp},\,g}, \boldsymbol{A}_{\boldsymbol{\Sigma}_g}, a_{\text{grp},\,h}, \boldsymbol{A}_{\boldsymbol{\Sigma}_h}, \sigma^2_\varepsilon, \sigma^2_{\text{gbl}}, \sigma^2_{\text{grp},g}, \boldsymbol{\Sigma}_g, \sigma^2_{\text{grp},h}, \boldsymbol{\Sigma}_h | \boldsymbol{y}) \\ &\qquad \approx \mathfrak{q}(\boldsymbol{\beta}, \boldsymbol{u}, a_\varepsilon, a_{\text{gbl}}, a_{\text{grp},\,g}, \boldsymbol{A}_{\boldsymbol{\Sigma}_g}, a_{\text{grp},\,h}, \boldsymbol{A}_{\boldsymbol{\Sigma}_h})\, \mathfrak{q}(\sigma^2_\varepsilon, \sigma^2_{\text{gbl}}, \sigma^2_{\text{grp},g}, \boldsymbol{\Sigma}_g, \sigma^2_{\text{grp},h}, \boldsymbol{\Sigma}_h). \end{aligned} \tag{3.5}$$

The optimal q-density functions for the parameters of interest are

$$\begin{aligned}
&\mathfrak{q}^*(\boldsymbol{\beta}, \boldsymbol{u}) \text{ has a } N\big(\boldsymbol{\mu}_{\mathfrak{q}(\boldsymbol{\beta},\boldsymbol{u})}, \boldsymbol{\Sigma}_{\mathfrak{q}(\boldsymbol{\beta},\boldsymbol{u})}\big) \text{ distribution,}\\
&\mathfrak{q}^*(\sigma^2_\varepsilon) \text{ has an Inverse-}\chi^2\big(\xi_{\mathfrak{q}(\sigma^2_\varepsilon)}, \lambda_{\mathfrak{q}(\sigma^2_\varepsilon)}\big) \text{ distribution,}\\
&\mathfrak{q}^*(\sigma^2_{\text{gbl}}) \text{ has an Inverse-}\chi^2\big(\xi_{\mathfrak{q}(\sigma^2_{\text{gbl}})}, \lambda_{\mathfrak{q}(\sigma^2_{\text{gbl}})}\big) \text{ distribution,}\\
&\mathfrak{q}^*(\sigma^2_{\text{grp},g}) \text{ has an Inverse-}\chi^2\big(\xi_{\mathfrak{q}(\sigma^2_{\text{grp},g})}, \lambda_{\mathfrak{q}(\sigma^2_{\text{grp},g})}\big) \text{ distribution}\\
&\mathfrak{q}^*(\sigma^2_{\text{grp},h}) \text{ has an Inverse-}\chi^2\big(\xi_{\mathfrak{q}(\sigma^2_{\text{grp},h})}, \lambda_{\mathfrak{q}(\sigma^2_{\text{grp},h})}\big) \text{ distribution}\\
&\mathfrak{q}^*(\boldsymbol{\Sigma}_g) \text{ has an Inverse-G-Wishart}(G_{\text{full}}, \xi_{\mathfrak{q}(\boldsymbol{\Sigma}_g)}, \boldsymbol{\Lambda}_{\mathfrak{q}(\boldsymbol{\Sigma}_g)}) \text{ distribution}\\
\text{and } &\mathfrak{q}^*(\boldsymbol{\Sigma}_h) \text{ has an Inverse-G-Wishart}(G_{\text{full}}, \xi_{\mathfrak{q}(\boldsymbol{\Sigma}_h)}, \boldsymbol{\Lambda}_{\mathfrak{q}(\boldsymbol{\Sigma}_h)}) \text{ distribution.}
\end{aligned}$$

The optimal q-density parameters are determined through an iterative coordinate ascent algorithm, details of which are given in Section S.10 of the supplementary materials. As in the two-level case, the updates for $\boldsymbol{\mu}_{\mathfrak{q}(\boldsymbol{\beta},\boldsymbol{u})}$ and $\boldsymbol{\Sigma}_{\mathfrak{q}(\boldsymbol{\beta},\boldsymbol{u})}$ may be written in the same form as (2.13), but with a three-level version of the $\boldsymbol{C}$ matrix and

$$\boldsymbol{D}_{\text{MFVB}} \equiv \begin{bmatrix} \boldsymbol{\Sigma}_{\boldsymbol{\beta}}^{-1} & \boldsymbol{O} & & \boldsymbol{O} & \\ \boldsymbol{O} & \boldsymbol{\mu}_{\mathfrak{q}(1/\sigma^2_{\text{gbl}})}\boldsymbol{I} & & \boldsymbol{O} & \\ & & \boldsymbol{M}_{\mathfrak{q}(\boldsymbol{\Sigma}_g^{-1})} & \boldsymbol{O} & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{O} & \boldsymbol{I}_m \otimes & \boldsymbol{O} & \boldsymbol{\mu}_{\mathfrak{q}(1/\sigma^2_{\text{grp},g})}\boldsymbol{I} \\ & & & \boldsymbol{O} & \boldsymbol{O} \\ & & & \boldsymbol{O} & \boldsymbol{I}_{n_i} \otimes \begin{bmatrix} \boldsymbol{M}_{\mathfrak{q}(\boldsymbol{\Sigma}_h^{-1})} & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{\mu}_{\mathfrak{q}(1/\sigma^2_{\text{grp},h})}\boldsymbol{I} \end{bmatrix} \end{bmatrix}. \tag{3.6}$$

For large numbers of level 2 and level 3 groups, $\boldsymbol{\Sigma}_{\mathfrak{q}(\boldsymbol{\beta},\boldsymbol{u})}$'s size becomes infeasible to deal with. However, only relatively small sub-blocks of $\boldsymbol{\Sigma}_{\mathfrak{q}(\boldsymbol{\beta},\boldsymbol{u})}$ are needed for variational inference regarding the variance and covariance parameters. These sub-block positions correspond to the non-zero sub-block positions of a general three-level sparse matrix defined in Section 3 of Nolan and Wand (2020). Here,

Nolan and Wand's

$\boldsymbol{A}_{11}$ sub-block corresponds to a $(2+K_{\text{gbl}})\times(2+K_{\text{gbl}})$ matrix $\boldsymbol{\Sigma}_{\text{q}(\boldsymbol{\beta},\boldsymbol{u}_{\text{gbl}})}$;

$\boldsymbol{A}_{22,i}$ sub-block corresponds to a $(2+K^g_{\text{grp}})\times(2+K^g_{\text{grp}})$ matrix $\boldsymbol{\Sigma}_{\text{q}(\boldsymbol{u}^g_{\text{lin},i},\boldsymbol{u}^g_{\text{grp},i})}$;
$\boldsymbol{A}_{12,i}$ sub-block corresponds to a $(2+K_{\text{gbl}})\times(2+K^g_{\text{grp}})$ matrix

$$E\left\{\left(\begin{bmatrix}\boldsymbol{\beta}\\ \boldsymbol{u}_{\text{gbl}}\end{bmatrix}-\boldsymbol{\mu}_{\text{q}(\boldsymbol{\beta},\boldsymbol{u}_{\text{gbl}})}\right)\left(\begin{bmatrix}\boldsymbol{u}^g_{\text{lin},i}\\ \boldsymbol{u}^g_{\text{grp},i}\end{bmatrix}-\boldsymbol{\mu}_{\text{q}(\boldsymbol{u}^g_{\text{lin},i},\boldsymbol{u}^g_{\text{grp},i})}\right)^T\right\},\ 1\le i\le m;$$

$\boldsymbol{A}_{22,ij}$ sub-block corresponds to a $(2+K^h_{\text{grp}})\times(2+K^h_{\text{grp}})$ matrix $\boldsymbol{\Sigma}_{\text{q}(\boldsymbol{u}^h_{\text{lin},ij},\boldsymbol{u}^h_{\text{grp},ij})}$;
$\boldsymbol{A}_{12,ij}$ sub-block corresponds to a $(2+K_{\text{gbl}})\times(2+K^h_{\text{grp}})$ matrix (3.7)

$$E\left\{\left(\begin{bmatrix}\boldsymbol{\beta}\\ \boldsymbol{u}_{\text{gbl}}\end{bmatrix}-\boldsymbol{\mu}_{\text{q}(\boldsymbol{\beta},\boldsymbol{u}_{\text{gbl}})}\right)\left(\begin{bmatrix}\boldsymbol{u}^h_{\text{lin},ij}\\ \boldsymbol{u}^h_{\text{grp},ij}\end{bmatrix}-\boldsymbol{\mu}_{\text{q}(\boldsymbol{u}^h_{\text{lin},ij},\boldsymbol{u}^h_{\text{grp},ij})}\right)^T\right\};$$

$\boldsymbol{A}_{12,\,i,\,j}$ sub-block corresponds to a $(2+K^g_{\text{grp}})\times(2+K^h_{\text{grp}})$ matrix

$$E\left\{\left(\begin{bmatrix}\boldsymbol{u}^g_{\text{lin},i}\\ \boldsymbol{u}^g_{\text{grp},i}\end{bmatrix}-\boldsymbol{\mu}_{\text{q}(\boldsymbol{u}^g_{\text{lin},i},\boldsymbol{u}^g_{\text{grp},i})}\right)\left(\begin{bmatrix}\boldsymbol{u}^h_{\text{lin},ij}\\ \boldsymbol{u}^h_{\text{grp},ij}\end{bmatrix}-\boldsymbol{\mu}_{\text{q}(\boldsymbol{u}^h_{\text{lin},ij},\boldsymbol{u}^h_{\text{grp},ij})}\right)^T\right\},$$

$1\le i\le m,\ 1\le j\le n_i$.

We appeal to Result 4 for a streamlined mean field variational Bayes algorithm.

Result 4*The mean field variational Bayes updates of* $\boldsymbol{\mu}_{\text{q}(\boldsymbol{\beta},\boldsymbol{u})}$ *and each of the sub-blocks of* $\boldsymbol{\Sigma}_{\text{q}(\boldsymbol{\beta},\boldsymbol{u})}$ *in (3.7) are expressible as a three-level sparse matrix least squares problem of the form:*

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

where the non-zero sub-blocks $\boldsymbol{B}$ *and* $\boldsymbol{b}$*, according to the notation in Section 3.1 of Nolan and Wand (2020), are for* $1\le i\le m$ *and* $1\le j\le n_i$*.*

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

$$\dot{\boldsymbol{B}}_{ij} \equiv \begin{bmatrix} \mu^{1/2}_{\mathrm{q}(1/\sigma^2_\varepsilon)}\boldsymbol{X}_{ij} & \mu^{1/2}_{\mathrm{q}(1/\sigma^2_\varepsilon)}\boldsymbol{Z}^g_{\mathrm{grp},ij} \\ \boldsymbol{O} & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{O} \\ n_i^{-1/2}\boldsymbol{M}^{1/2}_{\mathrm{q}(\boldsymbol{\Sigma}_g^{-1})} & \boldsymbol{O} \\ \boldsymbol{O} & n_i^{-1/2}\mu^{1/2}_{\mathrm{q}(1/\sigma^2_{\mathrm{grp},g})}\boldsymbol{I}_{K^g_{\mathrm{grp}}} \\ \boldsymbol{O} & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{O} \end{bmatrix} \quad \textit{and} \quad \ddot{\boldsymbol{B}}_{ij} \equiv \begin{bmatrix} \mu^{1/2}_{\mathrm{q}(1/\sigma^2_\varepsilon)}\boldsymbol{X}_{ij} & \mu^{1/2}_{\mathrm{q}(1/\sigma^2_\varepsilon)}\boldsymbol{Z}^h_{\mathrm{grp},ij} \\ \boldsymbol{O} & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{O} \\ \boldsymbol{M}^{1/2}_{\mathrm{q}(\boldsymbol{\Sigma}_h^{-1})} & \boldsymbol{O} \\ \boldsymbol{O} & \mu^{1/2}_{\mathrm{q}(1/\sigma^2_{\mathrm{grp},h})}\boldsymbol{I}_{K^h_{\mathrm{grp}}} \end{bmatrix}$$

with each of these matrices having $\tilde{o}_{ij} = o_{ij} + 2 + K_{\mathrm{gbl}} + 2 + K^g_{\mathrm{grp}} + 2 + K^h_{\mathrm{grp}}$ *rows and with* $\boldsymbol{B}_i$ *having* $p = 2 + K_{\mathrm{gbl}}$ *columns,* $\dot{\boldsymbol{B}}_i$ *having* $q_1 = 2 + K^g_{\mathrm{grp}}$ *columns and* $\ddot{\boldsymbol{B}}_{ij}$ *having* $q_2 = 2 + K^h_{\mathrm{grp}}$ *columns. The solutions are, with sub-block labelling according to (S.8) of the supplementary materials and Section 3 of Nolan and Wand (2020),*

$$\boldsymbol{\mu}_{\mathrm{q}(\boldsymbol{\beta},\boldsymbol{u}_{\mathrm{gbl}})} = \boldsymbol{x}_1, \quad \boldsymbol{\Sigma}_{\mathrm{q}(\boldsymbol{\beta},\boldsymbol{u}_{\mathrm{gbl}})} = \boldsymbol{A}^{11},$$

$$\boldsymbol{\mu}_{\mathrm{q}(\boldsymbol{u}^g_{\mathrm{lin},i},\boldsymbol{u}^g_{\mathrm{grp},i})} = \boldsymbol{x}_{2,i}, \quad E_q\left\{\begin{bmatrix} \boldsymbol{\beta} - \boldsymbol{\mu}_{\mathrm{q}(\boldsymbol{\beta})} \\ \boldsymbol{u}_{\mathrm{gbl}} - \boldsymbol{\mu}_{\mathrm{q}(\boldsymbol{u}_{\mathrm{gbl}})} \end{bmatrix}\begin{bmatrix} \boldsymbol{u}^g_{\mathrm{lin},i} - \boldsymbol{\mu}_{\mathrm{q}(\boldsymbol{u}^g_{\mathrm{lin},i})} \\ \boldsymbol{u}^g_{\mathrm{grp},i} - \boldsymbol{\mu}_{\mathrm{q}(\boldsymbol{u}^g_{\mathrm{grp},i})} \end{bmatrix}^T\right\} = \boldsymbol{A}^{12,i},$$

$$\boldsymbol{\Sigma}_{\mathrm{q}(\boldsymbol{u}^g_{\mathrm{lin},i},\boldsymbol{u}^g_{\mathrm{grp},i})} = \boldsymbol{A}^{22,i}, \quad 1 \le i \le m,$$

$$\boldsymbol{\mu}_{\mathrm{q}(\boldsymbol{u}^h_{\mathrm{lin},ij},\boldsymbol{u}^h_{\mathrm{grp},ij})} = \boldsymbol{x}_{2,ij}, \quad E_q\left\{\begin{bmatrix} \boldsymbol{\beta} - \boldsymbol{\mu}_{\mathrm{q}(\boldsymbol{\beta})} \\ \boldsymbol{u}_{\mathrm{gbl}} - \boldsymbol{\mu}_{\mathrm{q}(\boldsymbol{u}_{\mathrm{gbl}})} \end{bmatrix}\begin{bmatrix} \boldsymbol{u}^h_{\mathrm{lin},ij} - \boldsymbol{\mu}_{\mathrm{q}(\boldsymbol{u}^h_{\mathrm{lin},ij})} \\ \boldsymbol{u}^h_{\mathrm{grp},ij} - \boldsymbol{\mu}_{\mathrm{q}(\boldsymbol{u}^h_{\mathrm{grp},ij})} \end{bmatrix}^T\right\} = \boldsymbol{A}^{12,ij},$$

$$E_q\left\{\begin{bmatrix} \boldsymbol{u}^g_{\mathrm{lin},i} - \boldsymbol{\mu}_{\mathrm{q}(\boldsymbol{u}^g_{\mathrm{lin},i})} \\ \boldsymbol{u}^g_{\mathrm{grp},i} - \boldsymbol{\mu}_{\mathrm{q}(\boldsymbol{u}^g_{\mathrm{grp},i})} \end{bmatrix}\begin{bmatrix} \boldsymbol{u}^h_{\mathrm{lin},ij} - \boldsymbol{\mu}_{\mathrm{q}(\boldsymbol{u}^h_{\mathrm{lin},ij})} \\ \boldsymbol{u}^h_{\mathrm{grp},ij} - \boldsymbol{\mu}_{\mathrm{q}(\boldsymbol{u}^h_{\mathrm{grp},ij})} \end{bmatrix}^T\right\} = \boldsymbol{A}^{12,\,i,\,j}$$

and

$$\boldsymbol{\Sigma}_{\mathrm{q}(\boldsymbol{u}^h_{\mathrm{lin},ij},\boldsymbol{u}^h_{\mathrm{grp},ij})} = \boldsymbol{A}^{22,ij}, \quad 1 \le i \le m,\ 1 \le j \le n_i.$$

Algorithm 4 makes use of Result 4 to facilitate streamlined computation of all variational parameters in the three-level group-specific curves model. Section S.11 of the supplementary materials derives and provides an explicit expression for the approximate marginal log-likelihood.

Algorithm 4 QR-decomposition-based streamlined algorithm for obtaining mean field variational Bayes approximate posterior density functions for the parameters in the Bayesian three-level group-specific curves model (3.4) with product density restriction (3.5)

Data Inputs: $\boldsymbol{y}_{ij}(o_{ij} \times 1)$, $\boldsymbol{X}_{ij}(o_{ij} \times 2)$, $\boldsymbol{Z}_{\text{gbl},ij}(o_{ij} \times K_{\text{gbl}})$, $\boldsymbol{Z}^{g}_{\text{grp},ij}(o_{ij} \times K^{g}_{\text{grp}})$,

$\boldsymbol{Z}^{h}_{\text{grp},ij}(o_{ij} \times K^{h}_{\text{grp}})$ $1 \le i \le m,\ 1 \le j \le n_i$.

Hyperparameter Inputs: $\boldsymbol{\mu}_{\boldsymbol{\beta}}(2 \times 1)$, $\boldsymbol{\Sigma}_{\boldsymbol{\beta}}(2 \times 2)$ symmetric and positive definite,

$s_{\varepsilon}, \nu_{\varepsilon}, s_{\text{gbl}}, \nu_{\text{gbl}}, s_{\boldsymbol{\Sigma}_g,1}, s_{\boldsymbol{\Sigma}_g,2}, \nu_{\boldsymbol{\Sigma}_g}, s_{\text{grp},g}, \nu_{\text{grp},g}, s_{\boldsymbol{\Sigma}_h,1}, s_{\boldsymbol{\Sigma}_h,2}, \nu_{\boldsymbol{\Sigma}_h}, s_{\text{grp},h}, \nu_{\text{grp},h} > 0.$

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

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

$$\boldsymbol{C}_{\text{gbl},ij} \longleftarrow [\boldsymbol{X}_{ij}\ \boldsymbol{Z}_{\text{gbl},ij}] \quad ; \quad \boldsymbol{C}^{g}_{\text{grp},ij} \longleftarrow [\boldsymbol{X}_{ij}\ \boldsymbol{Z}^{g}_{\text{grp},ij}] \quad ; \quad \boldsymbol{C}^{h}_{\text{grp},ij} \longleftarrow [\boldsymbol{X}_{ij}\ \boldsymbol{Z}^{h}_{\text{grp},ij}]$$

Initialize: $\mu_{\text{q}(1/\sigma^2_{\varepsilon})}, \mu_{\text{q}(1/\sigma^2_{\text{gbl}})}, \mu_{\text{q}(1/\sigma^2_{\text{grp,g}})}, \mu_{\text{q}(1/\sigma^2_{\text{grp,h}})}, \mu_{\text{q}(1/a_{\varepsilon})}, \mu_{\text{q}(1/a_{\text{gbl}})}$,

$\mu_{\text{q}(1/a_{\text{grp,g}})}, \mu_{\text{q}(1/a_{\text{grp,h}})} > 0, \boldsymbol{M}_{\text{q}(\boldsymbol{\Sigma}_{\text{g}}^{-1})}(2 \times 2), \boldsymbol{M}_{\text{q}(\boldsymbol{\Sigma}_{\text{h}}^{-1})}(2 \times 2)$,

$\boldsymbol{M}_{\text{q}(A_{\text{g}}^{-1})}(2 \times 2), \boldsymbol{M}_{\text{q}(A_{\text{h}}^{-1})}(2 \times 2)$ symmetric and positive definite.

$$\xi_{\text{q}(\sigma^2_{\varepsilon})} \longleftarrow \nu_{\varepsilon} + \textstyle\sum_{i=1}^{m}\sum_{j=1}^{n_i} o_{ij} \quad ; \quad \xi_{\text{q}(\sigma^2_{\text{gbl}})} \longleftarrow \nu_{\text{gbl}} + K_{\text{gbl}} \quad ; \quad \xi_{\text{q}(\boldsymbol{\Sigma}_{\text{g}})} \longleftarrow \nu_{\boldsymbol{\Sigma}_g} + 2 + m$$

$$\xi_{\text{q}(\boldsymbol{\Sigma}_{\text{h}})} \longleftarrow \nu_{\boldsymbol{\Sigma}_h} + 2 + \textstyle\sum_{i=1}^{m} n_i \quad ; \quad \xi_{\text{q}(\sigma^2_{\text{grp},g})} \longleftarrow \nu_{\text{grp,g}} + mK^{g}_{\text{grp}}$$

$$\xi_{\text{q}(\sigma^2_{\text{grp},h})} \longleftarrow \nu_{\text{grp,h}} + K^{h}_{\text{grp}} \textstyle\sum_{i=1}^{m} n_i \quad ; \quad \xi_{\text{q}(a_{\varepsilon})} \longleftarrow \nu_{\varepsilon} + 1 \quad ; \quad \xi_{\text{q}(a_{\text{gbl}})} \longleftarrow \nu_{\text{gbl}} + 1$$

$$\xi_{\text{q}(a_{\text{grp,g}})} \longleftarrow \nu_{\text{grp,g}} + 1 \quad ; \quad \xi_{\text{q}(a_{\text{grp,h}})} \longleftarrow \nu_{\text{grp},h} + 1 \quad ; \quad \xi_{\text{q}(\boldsymbol{A}_{\boldsymbol{\Sigma}_g})} \longleftarrow \nu_{\boldsymbol{\Sigma}_g} + 2 \quad ; \quad \xi_{\text{q}(\boldsymbol{A}_{\boldsymbol{\Sigma}_h})} \longleftarrow \nu_{\boldsymbol{\Sigma}_h} + 2$$

Cycle:

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

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

$$\boldsymbol{b}_{ij} \longleftarrow \begin{bmatrix} \mu^{1/2}_{\text{q}(1/\sigma^2_{\varepsilon})}\boldsymbol{y}_{ij} \\ \frac{1}{\sqrt{\sum_{i=1}^{m} n_i}}\boldsymbol{\Sigma}_{\boldsymbol{\beta}}^{-1/2}\boldsymbol{\mu}_{\boldsymbol{\beta}} \\ \boldsymbol{0} \\ \boldsymbol{0} \\ \boldsymbol{0} \\ \boldsymbol{0} \\ \boldsymbol{0} \end{bmatrix}; \boldsymbol{B}_{ij} \longleftarrow \begin{bmatrix} \mu^{1/2}_{\text{q}(1/\sigma^2_{\varepsilon})}\boldsymbol{X}_{ij} & \mu^{1/2}_{\text{q}(1/\sigma^2_{\varepsilon})}\boldsymbol{Z}_{\text{gbl},ij} \\ \frac{1}{\sqrt{\sum_{i=1}^{m} n_i}}\boldsymbol{\Sigma}_{\boldsymbol{\beta}}^{-1/2} & \boldsymbol{O} \\ \boldsymbol{O} & \frac{1}{\sqrt{\sum_{i=1}^{m} n_i}}\mu^{1/2}_{\text{q}(1/\sigma^2_{\text{gbl}})}\boldsymbol{I}_{K_{\text{gbl}}} \\ \boldsymbol{O} & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{O} \end{bmatrix};$$

$$\dot{\boldsymbol{B}}_{ij} \longleftarrow \begin{bmatrix} \mu^{1/2}_{\text{q}(1/\sigma^2_{\varepsilon})}\boldsymbol{X}_{ij} & \mu^{1/2}_{\text{q}(1/\sigma^2_{\varepsilon})}\boldsymbol{Z}^{g}_{\text{grp},ij} \\ \boldsymbol{O} & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{O} \\ n_i^{-1/2}\boldsymbol{M}^{1/2}_{\text{q}(\boldsymbol{\Sigma}_g^{-1})} & \boldsymbol{O} \\ \boldsymbol{O} & n_i^{-1/2}\mu^{1/2}_{\text{q}(1/\sigma^2_{\text{grp},g})}\boldsymbol{I}_{K^{g}_{\text{grp}}} \\ \boldsymbol{O} & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{O} \end{bmatrix}$$

continued on a subsequent page ...

Figure 3 provides illustration of Algorithm 4 by showing the fits to the Figure 1 ultrasound data. The number of spline basis functions used at each level are $K_{\text{gbl}} = K^g_{\text{grp}} = K^h_{\text{grp}} = 15$. For the Bayesian fitting and inference, the data were transformed to have zero mean and unit standard deviation and hyperparameters values were analogous to those given in (2.18). The fits were then back-transformed to correspond to the original units. Posterior mean curves and (narrow) 99% pointwise credible

Algorithm 4 continued. This is a continuation of the description of this algorithm that commences on a preceding page

$$\ddot{\boldsymbol{B}}_{ij} \longleftarrow \begin{bmatrix} \mu^{1/2}_{\text{q}(1/\sigma^2_\varepsilon)}\boldsymbol{X}_i & \mu^{1/2}_{\text{q}(1/\sigma^2_\varepsilon)}\boldsymbol{Z}^h_{\text{grp},ij} \\ \boldsymbol{O} & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{O} \\ \boldsymbol{M}^{1/2}_{\text{q}(\boldsymbol{\Sigma}_h^{-1})} & \boldsymbol{O} \\ \boldsymbol{O} & \mu^{1/2}_{\text{q}(1/\sigma^2_{\text{grp},h})}\boldsymbol{I}_{K^h_{\text{grp}}} \end{bmatrix}$$

$\mathcal{S}_4 \longleftarrow \mathit{SolveThreeLevelSparseLeastSquares}\Big(\Big\{(\boldsymbol{b}_{ij}, \boldsymbol{B}_{ij}, \dot{\boldsymbol{B}}_{ij}, \ddot{\boldsymbol{B}}_{ij}) : 1 \le i \le m,$

$1 \le j \le n_i\Big\}\Big)$

$\boldsymbol{\mu}_{\text{q}(\boldsymbol{\beta},\boldsymbol{u}_{\text{gbl}})} \longleftarrow \boldsymbol{x}_1$ component of $\mathcal{S}_4$; $\boldsymbol{\Sigma}_{\text{q}(\boldsymbol{\beta},\boldsymbol{u}_{\text{gbl}})} \longleftarrow \boldsymbol{A}^{11}$ component of $\mathcal{S}_4$

$\boldsymbol{\mu}_{\text{q}(\boldsymbol{u}_{\text{gbl}})} \longleftarrow$ last K_{gbl} rows of $\boldsymbol{\mu}_{\text{q}(\boldsymbol{\beta},\boldsymbol{u}_{\text{gbl}})}$

$\boldsymbol{\Sigma}_{\text{q}(\boldsymbol{u}_{\text{gbl}})} \longleftarrow$ bottom-right $K_{\text{gbl}} \times K_{\text{gbl}}$ sub-block of $\boldsymbol{\Sigma}_{\text{q}(\boldsymbol{\beta},\boldsymbol{u}_{\text{gbl}})}$

$\lambda_{\text{q}(\sigma^2_\varepsilon)} \longleftarrow \mu_{\text{q}(1/a_\varepsilon)}$; $\Lambda_{\text{q}(\boldsymbol{\Sigma}_{\text{g}})} \longleftarrow \boldsymbol{M}_{\text{q}(A^{-1}_{\boldsymbol{\Sigma}_g})}$; $\Lambda_{\text{q}(\boldsymbol{\Sigma}_{\text{h}})} \longleftarrow \boldsymbol{M}_{\text{q}(A^{-1}_{\boldsymbol{\Sigma}_h})}$

$\lambda_{\text{q}(\sigma^2_{\text{grp, g}})} \longleftarrow \mu_{\text{q}(1/a_{\text{grp, g}})}$; $\lambda_{\text{q}(\sigma^2_{\text{grp, g}})} \longleftarrow \mu_{\text{q}(1/a_{\text{grp, g}})}$

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

$\boldsymbol{\mu}_{\text{q}(\boldsymbol{u}^g_{\text{lin},i}, \boldsymbol{u}^g_{\text{grp},i})} \longleftarrow \boldsymbol{x}_{2,i}$ component of $\mathcal{S}_4$

$\boldsymbol{\Sigma}_{\text{q}(\boldsymbol{u}^g_{\text{lin},i}, \boldsymbol{u}^g_{\text{grp},i})} \longleftarrow \boldsymbol{A}^{22,i}$ component of $\mathcal{S}_4$

$\boldsymbol{\mu}_{\text{q}(\boldsymbol{u}^g_{\text{lin},i})} \longleftarrow$ first 2 rows of $\boldsymbol{\mu}_{\text{q}(\boldsymbol{u}^g_{\text{lin},i}, \boldsymbol{u}^g_{\text{grp},i})}$

$\boldsymbol{\Sigma}_{\text{q}(\boldsymbol{u}^g_{\text{lin},i})} \longleftarrow$ top left 2×2 sub-block of $\boldsymbol{\Sigma}_{\text{q}(\boldsymbol{u}^g_{\text{lin},i}, \boldsymbol{u}^g_{\text{grp},i})}$

$\boldsymbol{\mu}_{\text{q}(\boldsymbol{u}^g_{\text{grp},i})} \longleftarrow$ last K^g_{grp} rows of $\boldsymbol{\mu}_{\text{q}(\boldsymbol{u}^g_{\text{lin},i}, \boldsymbol{u}^g_{\text{grp},i})}$

$\boldsymbol{\Sigma}_{\text{q}(\boldsymbol{u}^g_{\text{grp},i})} \longleftarrow$ bottom right $K^g_{\text{grp}} \times K^g_{\text{grp}}$ sub-block of $\boldsymbol{\Sigma}_{\text{q}(\boldsymbol{u}^g_{\text{lin},i}, \boldsymbol{u}^g_{\text{grp},i})}$

$$E_{\text{q}}\left\{\left(\begin{bmatrix}\boldsymbol{\beta}\\ \boldsymbol{u}_{\text{gbl}}\end{bmatrix} - \boldsymbol{\mu}_{\text{q}(\boldsymbol{\beta},\boldsymbol{u}_{\text{gbl}})}\right)\left(\begin{bmatrix}\boldsymbol{u}^g_{\text{lin},i}\\ \boldsymbol{u}^g_{\text{grp},i}\end{bmatrix} - \boldsymbol{\mu}_{\text{q}(\boldsymbol{u}^g_{\text{lin},i}, \boldsymbol{u}^g_{\text{grp},i})}\right)^T\right\}$$

$\longleftarrow \boldsymbol{A}^{12,i}$ component of $\mathcal{S}_4$

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

continued on a subsequent page ...

Algorithm 4 continued. This is a continuation of the description of this algorithm that commences on a preceding page

$\boldsymbol{\mu}_{\mathrm{q}(\boldsymbol{u}^b_{\mathrm{lin},\vec{ij}},\boldsymbol{u}^b_{\mathrm{grp},\vec{ij}})} \longleftarrow \boldsymbol{x}_{2,\vec{ij}}$ component of $\mathcal{S}_4$

$\boldsymbol{\Sigma}_{\mathrm{q}(\boldsymbol{u}^b_{\mathrm{lin},\vec{ij}},\boldsymbol{u}^b_{\mathrm{grp},\vec{ij}})} \longleftarrow \boldsymbol{A}^{22,\vec{ij}}$ component of $\mathcal{S}_4$

$\boldsymbol{\mu}_{\mathrm{q}(\boldsymbol{u}^b_{\mathrm{lin},\vec{ij}})} \longleftarrow$ first 2 rows of $\boldsymbol{\mu}_{\mathrm{q}(\boldsymbol{u}^b_{\mathrm{lin},\vec{ij}},\boldsymbol{u}^b_{\mathrm{grp},\vec{ij}})}$

$\boldsymbol{\Sigma}_{\mathrm{q}(\boldsymbol{u}^b_{\mathrm{lin},\vec{ij}})} \longleftarrow$ top left 2×2 sub-block of $\boldsymbol{\Sigma}_{\mathrm{q}(\boldsymbol{u}^b_{\mathrm{lin},\vec{ij}},\boldsymbol{u}^b_{\mathrm{grp},\vec{ij}})}$

$\boldsymbol{\mu}_{\mathrm{q}(\boldsymbol{u}^b_{\mathrm{grp},\vec{ij}})} \longleftarrow$ last K^b_{grp} rows of $\boldsymbol{\mu}_{\mathrm{q}(\boldsymbol{u}^b_{\mathrm{lin},\vec{ij}},\boldsymbol{u}^b_{\mathrm{grp},\vec{ij}})}$

$\boldsymbol{\Sigma}_{\mathrm{q}(\boldsymbol{u}^b_{\mathrm{grp},\vec{ij}})} \longleftarrow$ bottom right $K^b_{\mathrm{grp}} \times K^b_{\mathrm{grp}}$ sub-block of $\boldsymbol{\Sigma}_{\mathrm{q}(\boldsymbol{u}^b_{\mathrm{lin},\vec{ij}},\boldsymbol{u}^b_{\mathrm{grp},\vec{ij}})}$

$$E_{\mathrm{q}}\left\{\left(\begin{bmatrix}\boldsymbol{\beta}\\ \boldsymbol{u}_{\mathrm{gbl}}\end{bmatrix} - \boldsymbol{\mu}_{\mathrm{q}(\boldsymbol{\beta},\boldsymbol{u}_{\mathrm{gbl}})}\right)\left(\begin{bmatrix}\boldsymbol{u}^b_{\mathrm{lin},\vec{ij}}\\ \boldsymbol{u}^b_{\mathrm{grp},\vec{ij}}\end{bmatrix} - \boldsymbol{\mu}_{\mathrm{q}(\boldsymbol{u}^g_{\mathrm{lin},i},\boldsymbol{u}^g_{\mathrm{grp},i}))}\right)^T\right\}$$

$\longleftarrow \boldsymbol{A}^{12,\vec{ij}}$ component of $\mathcal{S}_4$

$$E_{\mathrm{q}}\left\{\left(\begin{bmatrix}\boldsymbol{u}^g_{\mathrm{lin},i}\\ \boldsymbol{u}^g_{\mathrm{grp},i}\end{bmatrix} - \boldsymbol{\mu}_{\mathrm{q}(\boldsymbol{u}^g_{\mathrm{lin},i},\boldsymbol{u}^g_{\mathrm{grp},i}))}\right)\left(\begin{bmatrix}\boldsymbol{u}^b_{\mathrm{lin},\vec{ij}}\\ \boldsymbol{u}^b_{\mathrm{grp},\vec{ij}}\end{bmatrix} - \boldsymbol{\mu}_{\mathrm{q}(\boldsymbol{u}^b_{\mathrm{lin},\vec{ij}},\boldsymbol{u}^b_{\mathrm{grp},\vec{ij}}))}\right)^T\right\}$$

$\longleftarrow \boldsymbol{A}^{12,i,j}$ component of $\mathcal{S}_4$

$$\lambda_{\mathrm{q}(\sigma^2_\varepsilon)} \longleftarrow \lambda_{\mathrm{q}(\sigma^2_\varepsilon)} + \Big\|\boldsymbol{y}_{\vec{ij}} - \boldsymbol{C}_{\mathrm{gbl},\vec{ij}}\boldsymbol{\mu}_{\mathrm{q}(\boldsymbol{\beta},\boldsymbol{u}_{\mathrm{gbl}})} - \boldsymbol{C}^g_{\mathrm{grp},\vec{ij}}\boldsymbol{\mu}_{\mathrm{q}(\boldsymbol{u}^g_{\mathrm{lin},i},\boldsymbol{u}^g_{\mathrm{grp},i})} - \boldsymbol{C}^b_{\mathrm{grp},\vec{ij}}\boldsymbol{\mu}_{\mathrm{q}(\boldsymbol{u}^b_{\mathrm{lin},\vec{ij}},\boldsymbol{u}^b_{\mathrm{grp},\vec{ij}})}\Big\|^2$$

$$\lambda_{\mathrm{q}(\sigma^2_\varepsilon)} \longleftarrow \lambda_{\mathrm{q}(\sigma^2_\varepsilon)} + \mathrm{tr}\big(\boldsymbol{C}^T_{\mathrm{gbl},\vec{ij}}\boldsymbol{C}_{\mathrm{gbl},\vec{ij}}\boldsymbol{\Sigma}_{\mathrm{q}(\boldsymbol{\beta},\boldsymbol{u}_{\mathrm{gbl}})}\big) + \mathrm{tr}\big((\boldsymbol{C}^g_{\mathrm{grp},\vec{ij}})^T\boldsymbol{C}^g_{\mathrm{grp},\vec{ij}}\boldsymbol{\Sigma}_{\mathrm{q}(\boldsymbol{u}^g_{\mathrm{lin},i},\boldsymbol{u}^g_{\mathrm{grp},i})}\big)$$

$$\lambda_{\mathrm{q}(\sigma^2_\varepsilon)} \longleftarrow \lambda_{\mathrm{q}(\sigma^2_\varepsilon)} + \mathrm{tr}\big((\boldsymbol{C}^b_{\mathrm{grp},\vec{ij}})^T\boldsymbol{C}^b_{\mathrm{grp},\vec{ij}}\boldsymbol{\Sigma}_{\mathrm{q}(\boldsymbol{u}^b_{\mathrm{lin},\vec{ij}},\boldsymbol{u}^b_{\mathrm{grp},\vec{ij}})}\big)$$

$$\lambda_{\mathrm{q}(\sigma^2_\varepsilon)} \longleftarrow \lambda_{\mathrm{q}(\sigma^2_\varepsilon)} + 2\,\mathrm{tr}\left[\boldsymbol{C}^T_{\mathrm{grp},i}\boldsymbol{C}_{\mathrm{gbl},i}\,E_{\mathrm{q}}\left\{\left(\begin{bmatrix}\boldsymbol{\beta}\\ \boldsymbol{u}_{\mathrm{gbl}}\end{bmatrix} - \boldsymbol{\mu}_{\mathrm{q}(\boldsymbol{\beta},\boldsymbol{u}_{\mathrm{gbl}})}\right)\times\left(\begin{bmatrix}\boldsymbol{u}^g_{\mathrm{lin},i}\\ \boldsymbol{u}^g_{\mathrm{grp},i}\end{bmatrix} - \boldsymbol{\mu}_{\mathrm{q}(\boldsymbol{u}^g_{\mathrm{lin},i},\boldsymbol{u}^g_{\mathrm{grp},i}))}\right)^T\right\}\right]$$

$$\lambda_{\mathrm{q}(\sigma^2_\varepsilon)} \longleftarrow \lambda_{\mathrm{q}(\sigma^2_\varepsilon)} + 2\,\mathrm{tr}\left[(\boldsymbol{C}^g_{\mathrm{grp},\vec{ij}})^T\boldsymbol{C}_{\mathrm{gbl},\vec{ij}}\,E_{\mathrm{q}}\left\{\left(\begin{bmatrix}\boldsymbol{\beta}\\ \boldsymbol{u}_{\mathrm{gbl}}\end{bmatrix} - \boldsymbol{\mu}_{\mathrm{q}(\boldsymbol{\beta},\boldsymbol{u}_{\mathrm{gbl}})}\right)\times\left(\begin{bmatrix}\boldsymbol{u}^b_{\mathrm{lin},\vec{ij}}\\ \boldsymbol{u}^b_{\mathrm{grp},\vec{ij}}\end{bmatrix} - \boldsymbol{\mu}_{\mathrm{q}(\boldsymbol{u}^b_{\mathrm{lin},\vec{ij}},\boldsymbol{u}^b_{\mathrm{grp},\vec{ij}}))}\right)^T\right\}\right]$$

$$\lambda_{\mathrm{q}(\sigma^2_\varepsilon)} \longleftarrow \lambda_{\mathrm{q}(\sigma^2_\varepsilon)} + 2\,\mathrm{tr}\left[(\boldsymbol{C}^g_{\mathrm{grp},\vec{ij}})^T\boldsymbol{C}^b_{\mathrm{grp},\vec{ij}}\,E_{\mathrm{q}}\left\{\left(\begin{bmatrix}\boldsymbol{u}^g_{\mathrm{lin},i}\\ \boldsymbol{u}^g_{\mathrm{grp},i}\end{bmatrix} - \boldsymbol{\mu}_{\mathrm{q}(\boldsymbol{u}^g_{\mathrm{lin},i},\boldsymbol{u}^g_{\mathrm{grp},i})}\right)\times\left(\begin{bmatrix}\boldsymbol{u}^b_{\mathrm{lin},\vec{ij}}\\ \boldsymbol{u}^b_{\mathrm{grp},\vec{ij}}\end{bmatrix} - \boldsymbol{\mu}_{\mathrm{q}(\boldsymbol{u}^b_{\mathrm{lin},\vec{ij}},\boldsymbol{u}^b_{\mathrm{grp},\vec{ij}})}\right)^T\right\}\right]$$

$$\boldsymbol{\Lambda}_{\mathrm{q}(\boldsymbol{\Sigma}_b)} \longleftarrow \boldsymbol{\Lambda}_{\mathrm{q}(\boldsymbol{\Sigma}_b)} + \boldsymbol{\mu}_{\mathrm{q}(\boldsymbol{u}^b_{\mathrm{lin},\vec{ij}})}\boldsymbol{\mu}^T_{\mathrm{q}(\boldsymbol{u}^b_{\mathrm{lin},\vec{ij}})} + \boldsymbol{\Sigma}_{\mathrm{q}(\boldsymbol{u}^b_{\mathrm{lin},\vec{ij}})}$$

$$\lambda_{\mathrm{q}(\sigma^2_{\mathrm{grp},b})} \longleftarrow \lambda_{\mathrm{q}(\sigma^2_{\mathrm{grp},b})} + \|\boldsymbol{\mu}_{\mathrm{q}(\boldsymbol{u}^b_{\mathrm{grp},\vec{ij}})}\|^2 + \mathrm{tr}\left(\boldsymbol{\Sigma}_{\mathrm{q}(\boldsymbol{u}^b_{\mathrm{grp},\vec{ij}})}\right)$$

until the increase in $\underline{\mathrm{p}}(\boldsymbol{y};\mathrm{q})$ is negligible.

continued on a subsequent page ...

Algorithm 4 continued. This is a continuation of the description of this algorithm that commences on a preceding page

$$\boldsymbol{\Lambda}_{\mathrm{q}(\boldsymbol{\Sigma}_g)} \longleftarrow \boldsymbol{\Lambda}_{\mathrm{q}(\boldsymbol{\Sigma}_g)} + \boldsymbol{\mu}_{\mathrm{q}(\boldsymbol{u}^g_{\mathrm{lin},i})}\boldsymbol{\mu}^T_{\mathrm{q}(\boldsymbol{u}^g_{\mathrm{lin},i})} + \boldsymbol{\Sigma}_{\mathrm{q}(\boldsymbol{u}^g_{\mathrm{lin},i})}$$

$$\lambda_{\mathrm{q}(\sigma^2_{\mathrm{grp},g})} \longleftarrow \lambda_{\mathrm{q}(\sigma^2_{\mathrm{grp},g})} + \|\boldsymbol{\mu}_{\mathrm{q}(\boldsymbol{u}^g_{\mathrm{grp},i})}\|^2 + \mathrm{tr}\left(\boldsymbol{\Sigma}_{\mathrm{q}(\boldsymbol{u}^g_{\mathrm{grp},i})}\right)$$

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

$$\mu_{\mathrm{q}(1/\sigma^2_\varepsilon)} \leftarrow \xi_{\mathrm{q}(\sigma_\varepsilon)}/\lambda_{\mathrm{q}(\sigma^2_\varepsilon)} \quad ; \quad \mu_{\mathrm{q}(1/\sigma^2_{\mathrm{gbl}})} \leftarrow \xi_{\mathrm{q}(\sigma^2_{\mathrm{gbl}})}/\lambda_{\mathrm{q}(\sigma^2_{\mathrm{gbl}})}$$

$$\boldsymbol{M}_{\mathrm{q}((\boldsymbol{\Sigma}_g)^{-1})} \leftarrow (\xi_{\mathrm{q}(\boldsymbol{\Sigma}_g)} - 2 + 1)\boldsymbol{\Lambda}^{-1}_{\mathrm{q}(\boldsymbol{\Sigma}_g)} \quad ; \quad \boldsymbol{M}_{\mathrm{q}((\boldsymbol{\Sigma}_h)^{-1})} \leftarrow (\xi_{\mathrm{q}(\boldsymbol{\Sigma}_h)} - 2 + 1)\boldsymbol{\Lambda}^{-1}_{\mathrm{q}(\boldsymbol{\Sigma}_h)}$$

$$\mu_{\mathrm{q}(1/\sigma^2_{\mathrm{grp},g})} \leftarrow \xi_{\mathrm{q}(\sigma^2_{\mathrm{grp},g})}/\lambda_{\mathrm{q}(\sigma^2_{\mathrm{grp},g})} \quad ; \quad \mu_{\mathrm{q}(1/\sigma^2_{\mathrm{grp},h})} \leftarrow \xi_{\mathrm{q}(\sigma^2_{\mathrm{grp},h})}/\lambda_{\mathrm{q}(\sigma^2_{\mathrm{grp},h})}$$

$$\lambda_{\mathrm{q}(a_\varepsilon)} \longleftarrow \mu_{\mathrm{q}(1/\sigma^2_\varepsilon)} + 1/(\nu_\varepsilon s^2_\varepsilon) \quad ; \quad \mu_{\mathrm{q}(1/a_\varepsilon)} \longleftarrow \xi_{\mathrm{q}(a_\varepsilon)}/\lambda_{\mathrm{q}(a_\varepsilon)}$$

$$\boldsymbol{M}_{\mathrm{q}(A^{-1}_{\boldsymbol{\Sigma}_g})} \longleftarrow \xi_{\mathrm{q}(A_{\boldsymbol{\Sigma}_g})}\boldsymbol{\Lambda}^{-1}_{\mathrm{q}(A_{\boldsymbol{\Sigma}_g})} \quad ; \; \boldsymbol{M}_{\mathrm{q}(A^{-1}_{\boldsymbol{\Sigma}_h})} \longleftarrow \xi_{\mathrm{q}(A_{\boldsymbol{\Sigma}_h})}\boldsymbol{\Lambda}^{-1}_{\mathrm{q}(A_{\boldsymbol{\Sigma}_h})}$$

$$\boldsymbol{\Lambda}_{\mathrm{q}(A_{\boldsymbol{\Sigma}_g})} \longleftarrow \mathrm{diag}\left\{\mathrm{diagonal}\left(\boldsymbol{M}_{\mathrm{q}(\boldsymbol{\Sigma}^{-1}_g)}\right)\right\} + \{\nu_{\boldsymbol{\Sigma}_g}\mathrm{diag}(s^2_{\boldsymbol{\Sigma}_g,1}, s^2_{\boldsymbol{\Sigma}_g,2})\}^{-1}$$

$$\boldsymbol{\Lambda}_{\mathrm{q}(A_{\boldsymbol{\Sigma}_h})} \longleftarrow \mathrm{diag}\left\{\mathrm{diagonal}\left(\boldsymbol{M}_{\mathrm{q}(\boldsymbol{\Sigma}^{-1}_h)}\right)\right\} + \{\nu_{\boldsymbol{\Sigma}_h}\mathrm{diag}(s^2_{\boldsymbol{\Sigma}_h,1}, s^2_{\boldsymbol{\Sigma}_h,2})\}^{-1}$$

$$\lambda_{\mathrm{q}(a_{\mathrm{gbl}})} \longleftarrow \mu_{\mathrm{q}(1/\sigma^2_{\mathrm{gbl}})} + 1/(\nu_{\mathrm{gbl}}s^2_{\mathrm{gbl}}) \quad ; \quad \mu_{\mathrm{q}(1/a_{\mathrm{gbl}})} \longleftarrow \xi_{\mathrm{q}(a_{\mathrm{gbl}})}/\lambda_{\mathrm{q}(a_{\mathrm{gbl}})}$$

$$\lambda_{\mathrm{q}(a_{\mathrm{grp},g})} \longleftarrow \mu_{\mathrm{q}(1/\sigma^2_{\mathrm{grp},g})} + 1/(\nu_{\mathrm{grp},g}s^2_{\mathrm{grp},g}) \quad ; \quad \mu_{\mathrm{q}(1/a_{\mathrm{grp},g})} \longleftarrow \xi_{\mathrm{q}(a_{\mathrm{grp},g})}/\lambda_{\mathrm{q}(a_{\mathrm{grp},g})}$$

$$\lambda_{\mathrm{q}(a_{\mathrm{grp},h})} \longleftarrow \mu_{\mathrm{q}(1/\sigma^2_{\mathrm{grp},h})} + 1/(\nu_{\mathrm{grp},h}s^2_{\mathrm{grp},h}) \quad ; \quad \mu_{\mathrm{q}(1/a_{\mathrm{grp},h})} \longleftarrow \xi_{\mathrm{q}(a_{\mathrm{grp},h})}/\lambda_{\mathrm{q}(a_{\mathrm{grp},h})}$$

Outputs: $\boldsymbol{\mu}_{\mathrm{q}(\boldsymbol{\beta},\boldsymbol{u}_{\mathrm{gbl}})}$, $\boldsymbol{\Sigma}_{\mathrm{q}(\boldsymbol{\beta},\boldsymbol{u}_{\mathrm{gbl}})}$, $\left\{\boldsymbol{\mu}_{\mathrm{q}(\boldsymbol{u}^g_{\mathrm{lin},i},\boldsymbol{u}^g_{\mathrm{grp},i})}, \boldsymbol{\Sigma}_{\mathrm{q}(\boldsymbol{u}^g_{\mathrm{lin},i},\boldsymbol{u}^g_{\mathrm{grp},i})}\right.$,

$$\left. E_{\mathrm{q}}\left\{\left(\begin{bmatrix}\boldsymbol{\beta}\\ \boldsymbol{u}_{\mathrm{gbl}}\end{bmatrix} - \boldsymbol{\mu}_{\mathrm{q}(\boldsymbol{\beta},\boldsymbol{u}_{\mathrm{gbl}})}\right)\left(\begin{bmatrix}\boldsymbol{u}^g_{\mathrm{lin},i}\\ \boldsymbol{u}^g_{\mathrm{grp},i}\end{bmatrix} - \boldsymbol{\mu}_{\mathrm{q}(\boldsymbol{u}^g_{\mathrm{lin},i},\boldsymbol{u}^g_{\mathrm{grp},i}))}\right)^T\right\} : 1 \le i \le m,\right.$$

$$E_{\mathrm{q}}\left\{\left(\begin{bmatrix}\boldsymbol{\beta}\\ \boldsymbol{u}_{\mathrm{gbl}}\end{bmatrix} - \boldsymbol{\mu}_{\mathrm{q}(\boldsymbol{\beta},\boldsymbol{u}_{\mathrm{gbl}})}\right)\left(\begin{bmatrix}\boldsymbol{u}^h_{\mathrm{lin},ij}\\ \boldsymbol{u}^h_{\mathrm{grp},ij}\end{bmatrix} - \boldsymbol{\mu}_{\mathrm{q}(\boldsymbol{u}^h_{\mathrm{lin},ij},\boldsymbol{u}^h_{\mathrm{grp},ij}))}\right)^T\right\},$$

$$E_{\mathrm{q}}\left\{\left(\begin{bmatrix}\boldsymbol{u}^g_{\mathrm{lin},i}\\ \boldsymbol{u}^g_{\mathrm{grp},i}\end{bmatrix} - \boldsymbol{\mu}_{\mathrm{q}(\boldsymbol{u}^g_{\mathrm{lin},i},\boldsymbol{u}^g_{\mathrm{grp},i})}\right)\left(\begin{bmatrix}\boldsymbol{u}^h_{\mathrm{lin},ij}\\ \boldsymbol{u}^h_{\mathrm{grp},ij}\end{bmatrix} - \boldsymbol{\mu}_{\mathrm{q}(\boldsymbol{u}^h_{\mathrm{lin},ij},\boldsymbol{u}^h_{\mathrm{grp},ij})}\right)^T\right\},$$

$$\left.\boldsymbol{\mu}_{\mathrm{q}(\boldsymbol{u}^h_{\mathrm{lin},ij},\boldsymbol{u}^h_{\mathrm{grp},ij})}, \boldsymbol{\Sigma}_{\mathrm{q}(\boldsymbol{u}^h_{\mathrm{lin},ij},\boldsymbol{u}^h_{\mathrm{grp},ij})} : 1 \le i \le m, 1 \le j \le n_i\right\}, \xi_{\mathrm{q}(\sigma_\varepsilon)},\ \lambda_{\mathrm{q}(\sigma^2_\varepsilon)},\ \xi_{\mathrm{q}(\sigma^2_{\mathrm{gbl}})},$$

$$\lambda_{\mathrm{q}(\sigma^2_{\mathrm{gbl}})}, \xi_{\mathrm{q}(\boldsymbol{\Sigma}_g)},\ \boldsymbol{\Lambda}^{-1}_{\mathrm{q}(\boldsymbol{\Sigma}_g)},\ \xi_{\mathrm{q}(\boldsymbol{\Sigma}_h)},\ \boldsymbol{\Lambda}^{-1}_{\mathrm{q}(\boldsymbol{\Sigma}_h)},\ \xi_{\mathrm{q}(\sigma^2_{\mathrm{grp},g})},\ \lambda_{\mathrm{q}(\sigma^2_{\mathrm{grp},g})},\ \xi_{\mathrm{q}(\sigma^2_{\mathrm{grp},h})},\ \lambda_{\mathrm{q}(\sigma^2_{\mathrm{grp},h})}.$$

intervals are shown. For this example, the fits for model (3.4) using our streamlined variational inference approach are shown to be very good. A deeper analysis shows that whilst the estimated $f + g_i + h_{ij}$ curves have good concordance with the slice within tumour trajectories, the estimated $f + g_i$ curves do not align well with the

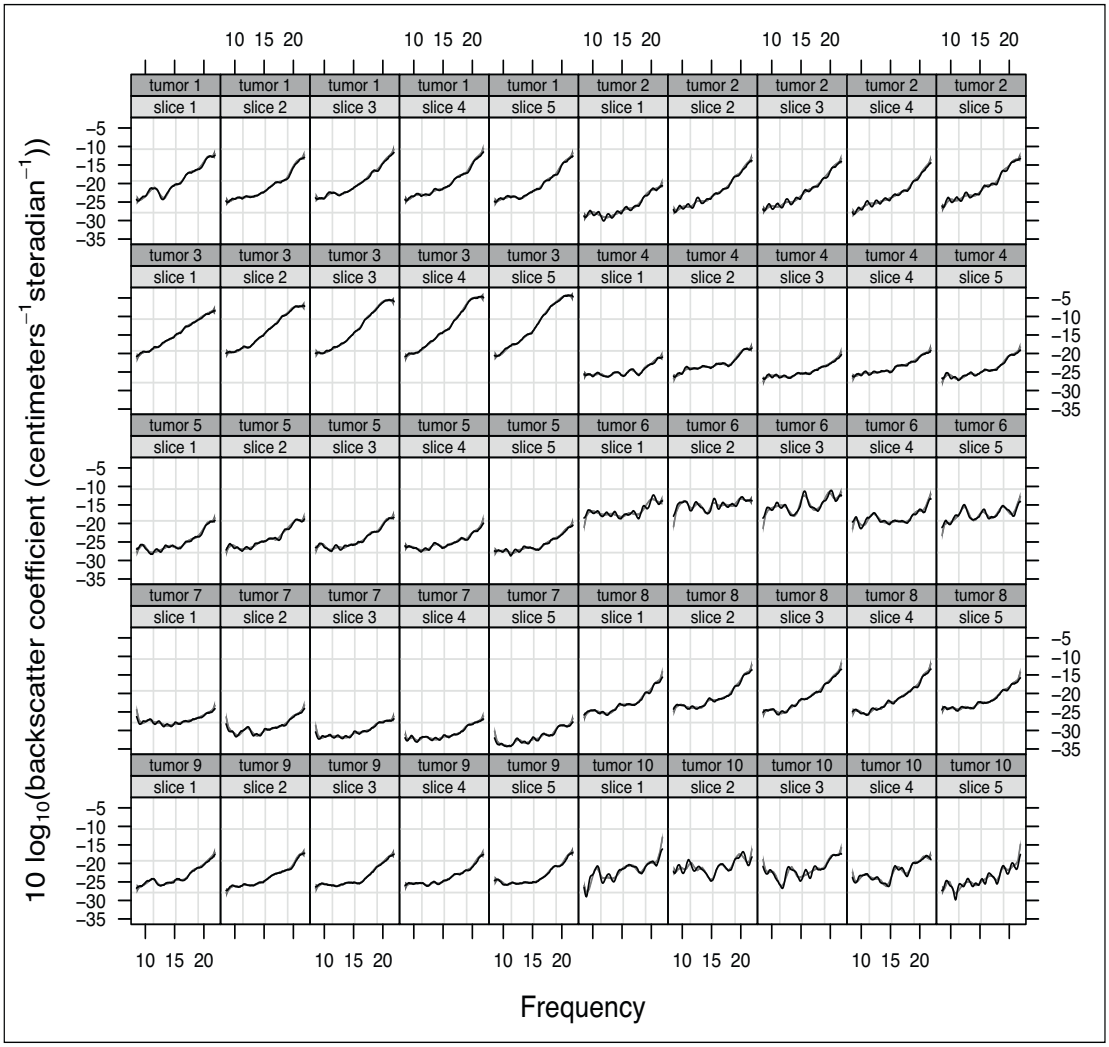

Figure 3 Illustrative three-level curve-type data with approximate fitted group specific curves and corresponding 99% credible sets based on mean field variational Bayes via Algorithm 4. The response variable is $10 \log_{10}$(backscatter coefficient) according to ultrasound technology. Level 1 corresponds to different ultrasound frequencies and matches the horizontal axes in each panel. Level 2 corresponds to different slices of a tumour due to differing probe locations. Level 3 corresponds to different tumours with one tumour for each of 10 laboratory mice

combined trajectories for the ith tumour when $i \in \{3, 7\}$. We suspect that this phenomenon is driven by identifiability issues and may be worthy of further study.

As discussed in the next section, fits such as those shown in Figure 3 can be obtained rapidly and accurately and Algorithm 4 is scalable to much larger datasets of the type illustrated by Figures 1 and 3.

4 Accuracy and speed assessment

In this section, we provide some assessment of the accuracy and speed of the inference delivered by streamlined variational inference for group-specific curves models.

4.1 Accuracy assessment

Mean field restrictions such as (2.12) and (3.5) imply that there is some loss of accuracy in inference produced by Algorithms 2 and 4. However, at least for the Gaussian response case treated here, approximate parameter orthogonality between the coefficient parameters and covariance parameters from likelihood theory implies that such restrictions are mild and mean field accuracy is high. Figure 4 corroborates this claim by assessing accuracy of the mean function estimates and 95% credible intervals at the median values of frequency for each panel in Figure 3. As a benchmark, we use Markov chain Monte Carlo-based inference via the rstan package (Guo et al., 2018). After a warm-up of size 1000, we retained 5000 Markov chain Monte Carlo samples from the mean function and median frequency posterior distributions and used kernel density estimation to approximate the corresponding posterior density function. For a generic univariate parameter θ, the accuracy of an approximation $\mathfrak{q}(\theta)$ to $\mathfrak{p}(\theta|\boldsymbol{y})$ is defined to be

$$\text{accuracy} \equiv 100\left\{1 - \tfrac{1}{2}\int_{-\infty}^{\infty}\left|\mathfrak{q}(\theta) - \mathfrak{p}(\theta|\boldsymbol{y})\right| d\theta\right\}\%. \tag{4.1}$$

The percentages in the top right-hand panel of Figure 4 correspond to (4.1) with replacement of $\mathfrak{p}(\theta|\boldsymbol{y})$ by the aforementioned kernel density estimate. In this case, accuracy is seen to be excellent, with accuracy percentages between 97% and 99% for all 40 curves. As explained in Section 3.1 of Menictas and Wand (2013) for a similar setting, such high accuracy is expected due to parameter orthogonality results that largely justify the product restrictions (2.12) and (3.5).

4.2 Speed assessment

We also conducted some simulation studies to assess the speed of streamlined variational higher level group-specific curve models, in terms of both comparative advantage over naïve implementation and absolute performance. The studies were run for both the two-level and three-level situations.

4.2.1 Two-level study

The focus of the two-level study was variational inference in the two-level case and to probe maximal speed potential. Algorithm 2 was implemented in the low-level computer language `Fortran 77`. An implementation of the naïve counterpart of Algorithm 2, involving storage and direct calculations concerning the full $\boldsymbol{\Sigma}_{\mathfrak{q}(\boldsymbol{\beta},\boldsymbol{u})}$ matrix, was also carried out. We then simulated data according to model (2.1) with $\sigma_\varepsilon = 0.2$,

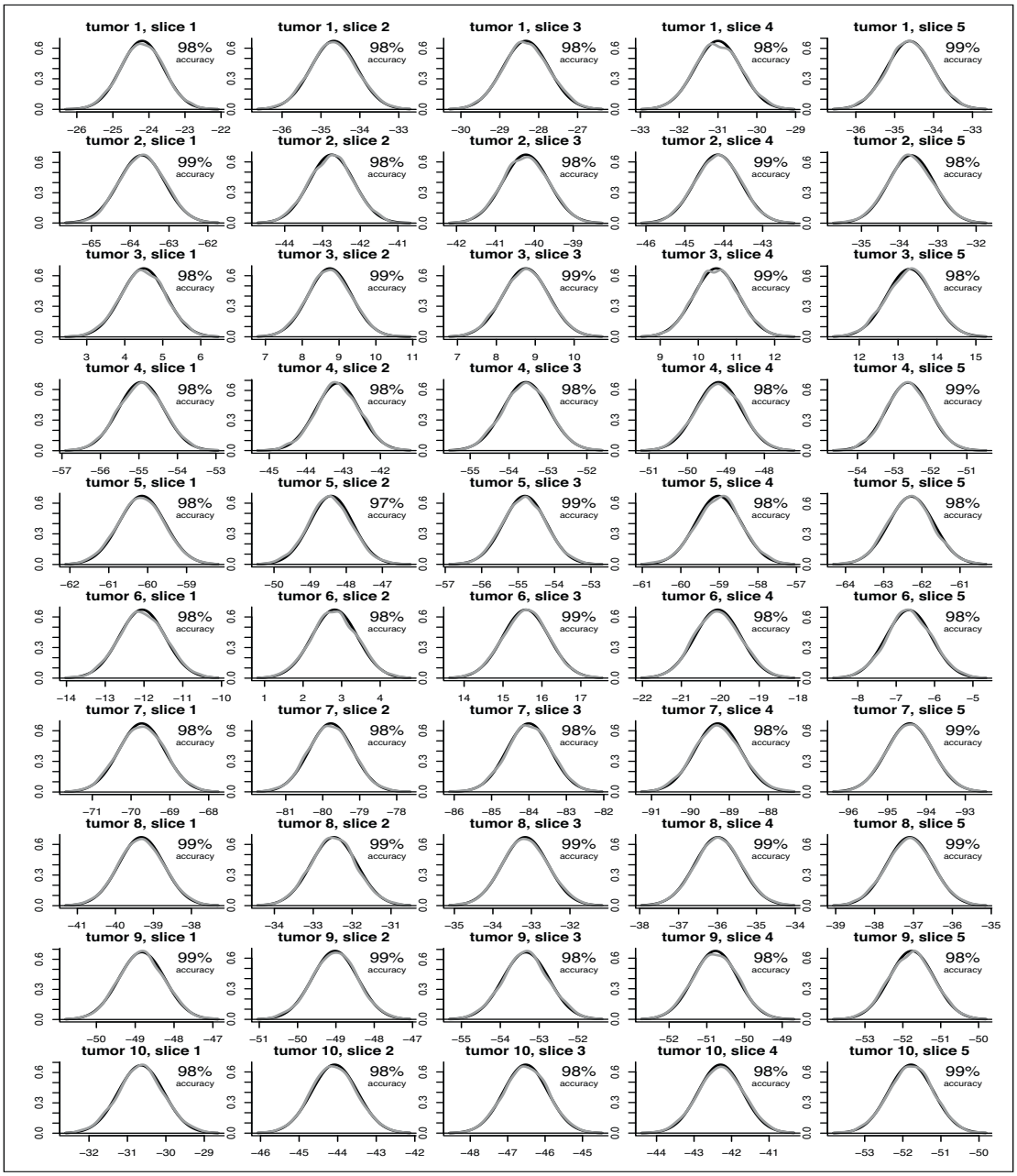

Figure 4 Accuracy assessment of Algorithm 4. Each panel displays approximate posterior density functions corresponding to mean function estimates according to the three-level group specific curve model (3.1). In each case the estimate is at the median frequency value. The grey density functions are based on Markov chain Monte Carlo and the black density functions are based on mean field variational Bayes. The accuracy percentage scores are defined by (4.1)

Table 1 Average (standard deviation) of elapsed computing times in seconds for fitting model (2.1) naïvely versus with streamlining via Algorithm 2. The NA entries indicate non applicability due to the naïve computations not being feasible

m	Naïve	Streamlined	Naïve/streamlined
100	75 (1.21)	0.748 (0.0334)	100
200	660 (7.72)	1.490 (0.0491)	442
300	2 210 (22.00)	2.260 (0.0567)	974
400	5 180 (92.20)	3.040 (0.0718)	1 700
500	NA	3.780 (0.0593)	NA

$$f(x) = 3\sqrt{x(1.3-x)}\,\Phi(6x-3) \quad \text{and} \quad g_i(x) = \alpha_1\alpha_2 \sin(2\pi x^{\alpha_3}),$$

where for each i, α_1, α_2 and α_3 are, respectively, random draws from the $N(\frac{1}{4}, \frac{1}{4})$ distribution and the sets $\{-1, 1\}$ and $\{1, 2, 3\}$. The level 2 sample sizes n_i were generated randomly from the set $\{30, 31, \ldots, 60\}$ and the level 1 sample sizes m ranging over the set $\{100, 200, 300, 400, 500\}$. All x_{ij} data were generated from a Uniform distribution over the unit interval. Table 1 summarizes the timings based on 100 replications with the number of mean field variational Bayes iterations fixed at 50. The study was run on a `MacBook Air` laptop with a 2.2 GHz processor and 8 GB of random access memory.

For m ranging from 100 to 400, we see that the naïve to streamlined ratios increase from about 100 to 1700. When $m = 500$, the naïve implementation fails to run due to its excessive storage demands. In contrast, the streamlined fits are produced in about 3 seconds. It is clear that streamlined variational inference is to be preferred and is the only option for large numbers of groups.

We then obtained timings for the streamlined algorithm for m becoming much larger, taking on values 100, 500, 2500 and 12500. The iterations in Algorithm 2 were stopped when the relative increase in the marginal log-likelihood fell below 10^{-5}. The average and standard deviation times in seconds over 100 replications are shown in Table 2. We see that the computational times are approximately linear in m. Even with 12 and a half thousand groups, Algorithm 2 is able to deliver fitting and inference on a contemporary laptop computer in about one and a half minutes.

For the contrast curves extension with two categories, the computing times are higher since the number of columns in the spline basis design matrices is doubled due to the presence of indicator variables. A running of the Figure 2 examples with and without categorization and contrast curves, and with the number of iterations

Table 2 Average (standard deviation) of elapsed computing times in seconds for fitting model (2.1) with streamlining via Algorithm 2 with implementation in the `Fortran 77` language

$m = 100$	$m = 500$	$m = 2,500$	$m = 12500$
0.635	2.900	16.90	95.00
(0.183)	(0.391)	(1.92)	(4.92)

Table 3 Average (standard deviation) of elapsed computing times in seconds for fitting a Bayesian three-level group-specific curves model with streamlining via Algorithm 4 with implementation in the `R` language

$m = 25$	$m = 50$	$m = 100$	$m = 200$	$m = 400$
82.4	174	394	978	2 650
(0.741)	(0.740)	(1.48)	(2.84)	(4.01)

kept constant, revealed an approximate fourfold increase in computing time for the contrast extension.

4.2.2 Three-level study

Our three-level study involved an implementation of Algorithm 4 in the slower, but friendlier, `R` computing language rather than `Fortran 77`. The data were simulated in an analogous way to the two-level study with sample sizes $m \in \{25, 50, 100, 200, 400\}$, $n_i = 50$ and $o_{ij} = 25$. The spline basis sizes were $K_{\text{gbl}} = 15$, $K^g_{\text{grp}} = 10$ and $K^b_{\text{grp}} = 7$. The number of mean field variational Bayes iterations was fixed at 50 and the number of replications was 100. The computer used for the two-level study was also used for this study.

The results of the three-level timing study are given in Table 3. They show that Algorithm 4 is approximately linear in m. The highest average computing time is about 44 minutes. The slowness of nested loops in the R language is partly to blame. Based on rough comparisons for our two-level R and Fortran 77 code, the highest computing times in Table 3 could be reduced to minutes rather than tens of minutes with low-level implementation.

5 Concluding remarks

We have used advanced matrix algebraic results to produce algorithms that allow the fitting of group-specific curve models with several thousand curves within minutes on contemporary personal hardware. In the Gaussian response setting considered here, the inferential accuracy is excellent due to parameter orthogonality results that imply the mildness of our mean field product restrictions. Our algorithms cover both two-level and three-level group-specific curve models and provide templates for higher level extensions.

6 Supplementary materials

Supplementary materials for this article containing

- derivations of all results,
- approximate marginal log-likelihood expressions

- and listings of the SolveTwoLevelSparseLeastSquares and SolveThreeLevelSparseLeastSquares algorithms

are available from `http://www.statmod.org/smij/archive.html`.

Acknowledgements

The ultrasound data were provided by the Bioacoustics Research Laboratory, Department of Electrical and Computer Engineering, University of Illinois at Urbana-Champaign, Illinois, USA.

Declaration of conflicting interests

The authors declared no potential conflicts of interest with respect to the research, authorship and/or publication of this article.

Funding

This research was partially supported by the Australian Research Council Discovery Project DP140100441 and U.S. National Institutes of Health grant R01CA226528-01A1.

References

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

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

Bishop CM (2006) *Pattern Recognition and Machine Learning*. New York, NY: Springer.

Blei DM, Kucukelbir A and McAuliffe JD (2017) Variational inference: A review for statisticians. *Journal of the American Statistical Association*, **112**, 859–877.

Brumback BA and Rice JA (1998) Smoothing spline models for the analysis of nested and crossed samples of curves (with discussion). *Journal of the American Statistical Association*, **93**, 961–994.

Djeundje VAB (2016) Systematic deviation in smooth mixed models for multi-level longitudinal data. *Statistical Methodology*, **32**, 203–217.

Djeundje VAB and Currie ID (2010) Appropriate covariance-specification via penalties for penalized splines in mixed models for longitudinal data. *Electronic Journal of Statistics*, **4**, 1202–1224.

Donnelly CA, Laird NM and Ware JH (1995) Prediction and creation of smooth curves for temporally correlated longitudinal data. *Journal of the American Statistical Association*, **90**, 984–989.

Durban M, Harezlak J, Wand MP and Carroll RJ (2005) Simple fitting of subject-specific curves for longitudinal data. *Statistics in Medicine*, **24**, 1153–1167.

Goldsmith J, Zipunnikov V and Schrack J (2015) Generalized multilevel function on-scalar regression and principal component analysis. *Biometrics*, 71, 344–353.

Guo J, Gabry J and Goodrich B (2018) *rstan: R interface to Stan*. R package version 2.18.2. URL http://mc-stan.org (last accessed 17 June 2020).

Harezlak J, Ruppert D and Wand MP (2018) *Semiparametric Regression with R*. New York, NY: Springer.

Heckman N, Lockhart R and Nielsen JD (2013) Penalized regression, mixed effects models and appropriate modelling. *Electronic Journal of Statistics*, **7**, 1517–1552.

Huang A and Wand MP (2013) Simple marginally noninformative prior distributions for covariance matrices. *Bayesian Analysis*, **8**, 439–452.

Kauermann G, Krivobokova T and Fahrmeir L (2009) Some asymptotic results on generalized penalized spline smoothing. *Journal of the Royal Statistical Society,Series B*, **71**, 487–503.

Lee CYY and Wand MP (2016) Variational inference for fitting complex Bayesian mixed effects models to health data. *Statistics in Medicine*, **35**, 165–188.

Menictas M and Wand MP (2013) Variational inference for marginal longitudinal semiparametric regression. *Stat*, **2**, 61–71.

Nolan TH, Menictas M and Wand MP (2020) *Streamlined computing for variational inference with higher level random effects*. Unpublished manuscript. URL https://arxiv.org/abs/1903.06616 (last accessed 17 June 2020).

Nolan TH and Wand MP (2020) Streamlined solutions to multilevel sparse matrix problems. *ANZIAM Journal*. doi: 10.1017/S1446181120000061

Pinheiro JC and Bates DM (2000) *Mixed-effects Models in S and S-PLUS*. New York, NY: Springer.

Pinheiro J, Bates D, DebRoy S, Sarkar D and R Core Team (2018) *nlme: Linear and nonlinear mixed effects models*. R package version 3.1. URL http://cran.r-project.org/package=nlme (last accessed 17 June 2020).

Pratt JH, Jones JJ, Miller JZ, Wagner MA and Fineberg NS (1989) Racial differences in aldosterone excretion and plasma aldosterone concentrations in children. *New England Journal of Medicine*, **321**, 1152–1157.

Robinson GK (1991) That BLUP is a good thing: The estimation of random effects. *Statistical Science*, **6**, 15–51.

Trail JB, Collins LM, Rivera DE, Li R, Piper ME and Baker TB (2014) Functional data analysis for dynamical system identification of behavioral processes. *Psychological Methods*, **19**, 175–187.

Verbyla AP, Cullis BR, Kenward MG and Welham SJ (1999) The analysis of designed experiments and longitudinal data by using smoothing splines (with discussion) *Applied Statistics*, **48**, 269–312.

Wahba G (1990) *Spline Models for Observational Data*. Philadelphia: Society for Industrial and Applied Mathematics.

Wand MP and Ormerod JT (2008) On semiparametric regression with O–Sullivan penalized splines. *Australian and New Zealand Journal of Statistics*, **50**, 179–198.

——— (2011) Penalized wavelets: Embedding wavelets into semiparametric regression. *Electronic Journal of Statistics*, **5**, 1654–1717.

Wang Y (1998) Mixed effects smoothing spline analysis of variance. *Journal of the Royal Statistical Society, Series B*, **60**, 159–174.

Wirtzfeld LA, Ghoshal G, Rosado-Mendez IM, Nam K, Park Y, Pawlicki AD, Miller RJ, Simpson DG, Zagzebski JA, Oelze MI, Hall TJ and O–Brien WD (2015) Quantitative ultrasound comparison of MAT and 4T1 mammary tumours in mice and rates across multiple imaging systems. *Journal of Ultrasound Medicine*, **34**, 1373–1383.

Zhang D, Lin X, Raz J and Sowers M (1998) Semi-parametric stochastic mixed models for longitudinal data. *Journal of the American Statistical Association*, **93**, 710–719.