

SUPPLEMENT TO “CLUSTERING FOR MULTIVARIATE CONTINUOUS AND DISCRETE LONGITUDINAL DATA”

BY ARNOŠT KOMÁREK AND LENKA KOMÁRKOVÁ

Charles University in Prague and the University of Economics in Prague

This is a supplement to the paper published by *The Annals of Applied Statistics*. It contains (A) more detailed description of the assumed prior distribution for model parameters giving also some guidelines for the selection of the hyperparameters to achieve a weakly informative prior distribution; (B) more details on the posterior distribution and the sampling MCMC algorithm; (C) additional information to the analysis of the Mayo Clinic PBC data; (D) more detailed results of the simulation study.

This is a supplement to the paper “Clustering for multivariate continuous and discrete longitudinal data” (Komárek and Komárková, 2012). Sections in the supplement are considered as appendices to the main paper and are numbered using capital letters. Cross-references in this supplement that do not contain a capital letter refer to corresponding parts of the main paper.

Appendix A provides a detailed description of the prior distribution in the Bayesian specification of the mixture multivariate generalized linear mixed model as well as guidelines for selection of the fixed hyperparameters to achieve weakly informative prior distribution. Appendix B gives more details on the posterior distribution and the sampling MCMC algorithm. Additional information to the analysis of the Mayo Clinic PBC data is given in Appendix C. Finally, Appendix D shows additional results of the simulation study.

Keywords and phrases: Classification, functional data, generalized linear mixed model, multivariate longitudinal data, repeated observations

A. Prior distribution. In this appendix, we provide a detailed overview of the prior distribution assumed in the Bayesian specification of the mixture multivariate generalized linear mixed model. We also give some guidelines for the selection of the hyperparameters to achieve a weakly informative prior distribution.

A.1. *Notation.* First, we review and slightly extend the notation:

- $\boldsymbol{\alpha} = (\boldsymbol{\alpha}_1^\top, \dots, \boldsymbol{\alpha}_R^\top)^\top$: the fixed effects, where $\boldsymbol{\alpha}_r$ is a $c_r \times 1$ vector ($r = 1, \dots, R$) and $\boldsymbol{\alpha}$ is a $c \times 1$ vector ($c = \sum_{r=1}^R c_r$);
- $\boldsymbol{\phi} = (\phi_1, \dots, \phi_R)^\top$: GLMM dispersion parameters of which some of them might be constants by definition, e.g., for Bernoulli response with a logit link (logistic regression), corresponding ϕ_r is equal to 1, and the same is true for Poisson response with a log link. Further, we denote by $\boldsymbol{\phi}^{-1}$ a vector of inverted dispersion parameters, i.e., $\boldsymbol{\phi}^{-1} = (\phi_1^{-1}, \dots, \phi_R^{-1})^\top$;
- $\mathbf{B} = (\mathbf{b}_1, \dots, \mathbf{b}_N)$: individual (latent) values of random effects, where $\mathbf{b}_i = (\mathbf{b}_{i,1}^\top, \dots, \mathbf{b}_{i,R}^\top)^\top$, $\mathbf{b}_{i,r}$ ($r = 1, \dots, R$) are $d_r \times 1$ vectors and \mathbf{b}_i ($i = 1, \dots, N$) are $d \times 1$ vectors ($d = \sum_{r=1}^R d_r$);
- $\mathbf{w} = (w_1, \dots, w_K)^\top$: mixture weights from the distribution of random effects;
- $\boldsymbol{\mu}_1, \dots, \boldsymbol{\mu}_K$: $d \times 1$ vectors of mixture means;
- $\mathbf{D}_1, \dots, \mathbf{D}_K$: $d \times d$ mixture covariance matrices;
- \mathbf{s} : $d \times 1$ shift vector used to calculate the shifted-scaled random effects (see below);
- \mathbf{S} : $d \times d$ diagonal scale matrix used to calculate the shifted-scaled random effects;
- $\mathbf{B}^* = (\mathbf{b}_1^*, \dots, \mathbf{b}_N^*)$: individual (latent) values of shifted-scaled random effects, where $\mathbf{b}_i^* = (\mathbf{b}_{i,1}^{*\top}, \dots, \mathbf{b}_{i,R}^{*\top})^\top$, $\mathbf{b}_{i,r}^{*\top}$ ($r = 1, \dots, R$) are $d_r \times 1$ vectors and $\mathbf{b}_i^{*\top}$ ($i = 1, \dots, N$) are $d \times 1$ vectors;
- $\boldsymbol{\mu}_1^*, \dots, \boldsymbol{\mu}_K^*$: $d \times 1$ vectors of mixture means from the distribution of shifted-scaled random effects;
- $\mathbf{D}_1^*, \dots, \mathbf{D}_K^*$: $d \times d$ mixture covariance matrices from the distribution of shifted-scaled random effects;
- $\mathbf{u} = (u_1, \dots, u_N)^\top$: (latent) component allocations;
- $\boldsymbol{\gamma}_b = (\gamma_{b,1}, \dots, \gamma_{b,d})^\top$: random hyperparameter introduced in an additional hierarchy level when specifying the prior distribution for the mixture covariance matrices (see below). Further, let $\boldsymbol{\gamma}_b^{-1} = (\gamma_{b,1}^{-1}, \dots, \gamma_{b,d}^{-1})^\top$;
- $\boldsymbol{\gamma}_\phi = (\gamma_{\phi,1}, \dots, \gamma_{\phi,R})^\top$: random hyperparameter introduced in an ad-

ditional hierarchy level when specifying the prior distribution for the GLMM dispersion parameters (see below). Further, let $\boldsymbol{\gamma}_\phi^{-1} = (\gamma_{\phi,1}^{-1}, \dots, \gamma_{\phi,R}^{-1})^\top$.

The set of GLMM related parameters further composes a vector $\boldsymbol{\psi}$, and the set of mixture related parameters further composes a vector $\boldsymbol{\theta}$, i.e.,

$$\boldsymbol{\psi} = (\boldsymbol{\alpha}^\top, \boldsymbol{\phi}^\top)^\top, \quad \boldsymbol{\theta} = (\boldsymbol{w}^\top, \boldsymbol{\mu}_1^\top, \dots, \boldsymbol{\mu}_K^\top, \text{vec}(\boldsymbol{D}_1), \dots, \text{vec}(\boldsymbol{D}_K))^\top.$$

In the following, $\boldsymbol{\psi}$ and $\boldsymbol{\theta}$ are sometimes referenced as *frequentist* parameters as they are the same as unknown parameters of the corresponding frequentist model. Finally, let $p(\cdot)$ and $p(\cdot|\cdot)$ be generic symbols for (conditional) distributions and let $\varphi(\cdot|\boldsymbol{\mu}, \boldsymbol{D})$ be a density of the (multivariate) normal distribution with a mean $\boldsymbol{\mu}$ and a covariance matrix \boldsymbol{D}

A.2. *Shifted-scaled random effects.* Due to possibility of improving the mixing and numerical stability of the MCMC algorithm described in Section B, we introduce shifted-scaled random effects:

$$\boldsymbol{b}_i^* = \boldsymbol{S}^{-1}(\boldsymbol{b}_i - \boldsymbol{s}), \quad i = 1, \dots, N,$$

where \boldsymbol{s} is a $d \times 1$ fixed shift vector and \boldsymbol{S} is a $d \times d$ fixed diagonal scale matrix. It follows from Eq. (4) that

$$\boldsymbol{b}_i^* | \boldsymbol{\theta}^* \stackrel{\text{i.i.d.}}{\sim} \sum_{k=1}^K w_k \mathcal{MVN}(\boldsymbol{\mu}_k^*, \boldsymbol{D}_k^*), \quad i = 1, \dots, N,$$

where $\boldsymbol{\theta}^* = (\boldsymbol{w}^\top, \boldsymbol{\mu}_1^{*\top}, \dots, \boldsymbol{\mu}_K^{*\top}, \text{vec}(\boldsymbol{D}_1^*), \dots, \text{vec}(\boldsymbol{D}_K^*))^\top$:

$$(A.1) \quad \begin{cases} \boldsymbol{\mu}_k^* &= \boldsymbol{S}^{-1}(\boldsymbol{\mu}_k - \boldsymbol{s}), \\ \boldsymbol{D}_k^* &= \boldsymbol{S}^{-1} \boldsymbol{D}_k \boldsymbol{S}^{-1}, \end{cases} \quad k = 1, \dots, K.$$

Without the loss of generality, we may set $\boldsymbol{s} = \mathbf{0}$ (zero vector) and $\boldsymbol{S} = \boldsymbol{I}_d$ ($d \times d$ identity matrix). However for numerical stability of the MCMC which is subsequently used to obtain a sample from the corresponding posterior distribution, it is useful if the shifted-scaled random effects have approximately zero means and unit variances. We return to this point in Sec. A.5.

A.3. *Factorization of the prior distribution.* With the MCMC algorithm, we primarily obtain a sample from the posterior distribution of $\boldsymbol{\theta}^*$ (parameters of the mixture distribution of shifted-scaled random effects) and then use the transformation (A.1) to get a sample from the posterior distribution

of $\boldsymbol{\theta}$ (parameters of the mixture distribution of the original random effects). For this reason, all expression explaining the assumed prior distribution will be expressed in terms of $\boldsymbol{\theta}^*$ rather than $\boldsymbol{\theta}$.

As it is usual for hierarchically specified models, the joint prior distribution which involves also latent random effects \mathbf{B} and component allocations \mathbf{u} exploits conditional independence between sets of unrelated parameters. Namely, it is factorized as

$$(A.2) \quad p(\boldsymbol{\psi}, \boldsymbol{\theta}^*, \mathbf{B}, \mathbf{u}) = p(\mathbf{B} | \boldsymbol{\theta}^*, \mathbf{u}) \times p(\mathbf{u} | \boldsymbol{\theta}^*) \times p(\boldsymbol{\theta}^*) \times p(\boldsymbol{\psi}),$$

where

$$(A.3) \quad \left\{ \begin{array}{l} p(\mathbf{B} | \boldsymbol{\theta}^*, \mathbf{u}) = \prod_{i=1}^N p(\mathbf{b}_i | \boldsymbol{\theta}^*, u_i), \\ p(\mathbf{u} | \boldsymbol{\theta}^*) = \prod_{i=1}^N p(u_i | \boldsymbol{w}). \end{array} \right.$$

The products on the right-hand side of Eq. (A.3) follow from assumed independence between subjects/patients. For the prior distribution $p(\boldsymbol{\theta}^*)$ of the mixture related parameters, we shall use a multivariate version of the classical proposal of [Richardson and Green \(1997\)](#) which exploits a factorization:

$$(A.4) \quad p(\boldsymbol{\theta}^*) = p(\boldsymbol{w}) \times \prod_{k=1}^K \left\{ p(\boldsymbol{\mu}_k^*) \times p(\mathbf{D}_k^{*-1}) \right\}.$$

For $p(\boldsymbol{\psi})$, the last part of the prior expression related to the GLMM parameters, we shall adopt classically used priors in this context (see, e.g., [Fong, Rue and Wakefield, 2010](#)) with factorization:

$$(A.5) \quad p(\boldsymbol{\psi}) = p(\boldsymbol{\phi}^{-1}) \times p(\boldsymbol{\alpha}) = \prod_{r=1}^R p(\phi_r^{-1}) \times p(\boldsymbol{\alpha}).$$

In the remained of this Section, we describe our particular choices for the factors in Eqs. (A.3), (A.4) and (A.5). Further, we explain a data motivated guidelines for selection of the fixed hyperparameters which may be followed to obtain a weakly informative prior distribution. These guidelines are motivated by the classical proposal of [Richardson and Green \(1997\)](#) in the case of mixture related parameters and the classical proposal for GLMM related parameters (see, e.g., [Fong, Rue and Wakefield, 2010](#)).

A.4. *Weakly informative prior distribution.* To be able to follow the Richardson and Green’s guidelines for the selection of the fixed hyperparameters leading to the weakly informative prior distribution, we have to have some initial idea concerning the location, scale and individual values of the random effects in our multivariate GLMM. This can be obtained by using standard software and calculating separate maximum-likelihood estimates in each of R GLMM models (3) under the classical assumption of normally distributed random effects. In the following, let for each $r = 1, \dots, R$, $\beta_{ML,r}^0$ be so obtained initial estimate for the overall mean vector of random effects, and $\mathbf{d}_{ML,r}^0$ analogously obtained initial estimates for the vector of overall standard deviations of random effects. Further, let for each $r = 1, \dots, R$, $\mathbf{b}_{EB,i,r}^0$, $i = 1, \dots, N$ be empirical Bayes estimates of individual values of random effects in the r th ML estimated GLMM under the assumption of normality of random effects. Furthermore, let

$$\begin{aligned}\beta_{ML}^0 &= (\beta_{ML,1}^0{}^\top, \dots, \beta_{ML,R}^0{}^\top)^\top, \\ \mathbf{d}_{ML}^0 &= (\mathbf{d}_{ML,1}^0{}^\top, \dots, \mathbf{d}_{ML,R}^0{}^\top)^\top, \\ \mathbf{b}_{EB,i}^0 &= (\mathbf{b}_{EB,i,1}^0{}^\top, \dots, \mathbf{b}_{EB,i,R}^0{}^\top)^\top,\end{aligned}$$

$i = 1, \dots, N$. The right-hand side of the Expression (A.2) is then composed of the factors explained in Sec. A.5–A.11.

A.5. *Prior for random effects.* $p(\mathbf{b}_i | \boldsymbol{\theta}^*, u_i)$ is for each $i = 1, \dots, N$ a (multivariate) normal distribution $\mathcal{N}(\mathbf{s} + \mathbf{S}\boldsymbol{\mu}_{u_i}^*, \mathbf{SD}_{u_i}^* \mathbf{S}^\top)$, that is,

$$\begin{aligned}\text{(A.6)} \quad & p(\mathbf{b}_i | \boldsymbol{\theta}^*, u_i) \\ & \propto |\mathbf{SD}_{u_i}^* \mathbf{S}^\top|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}(\mathbf{b}_i - \mathbf{s} - \mathbf{S}\boldsymbol{\mu}_{u_i}^*)^\top (\mathbf{SD}_{u_i}^* \mathbf{S}^\top)^{-1} (\mathbf{b}_i - \mathbf{s} - \mathbf{S}\boldsymbol{\mu}_{u_i}^*)\right\}.\end{aligned}$$

This follows from assumed mixture model (4) For numerical stability of the MCMC which is subsequently used to obtain a sample from the posterior distribution, it is useful if the shifted-scaled random effects

$$\text{(A.7)} \quad \mathbf{b}_i^* = \mathbf{S}^{-1}(\mathbf{b}_i - \mathbf{s}), \quad i = 1, \dots, N$$

have approximately zero means and unit variances. This can be achieved by setting $\mathbf{s} = \beta_{ML}^0$ and $\mathbf{S} = \text{diag}(\mathbf{d}_{ML}^0)$. In a sequel, let further $\mathbf{b}_{EB,i}^{0*} = \mathbf{S}^{-1}(\mathbf{b}_{EB,i}^0 - \mathbf{s})$, $i = 1, \dots, N$ be vectors of shifted and scaled values of the initial ML based empirical Bayes estimates of random effects, and let $\mathbf{R}_{EB}^{0*} = (R_{EB,1}^{0*}, \dots, R_{EB,d}^{0*})$ be a vector of sample standard deviations based on elements of $\mathbf{b}_{EB,i}^{0*}$ multiplied by some factor c , where $c > 1$. In our applications we use $c = 6$. Finally, let $(\mathbf{R}_{EB}^{0*})^2 = ((R_{EB,1}^{0*})^2, \dots, (R_{EB,d}^{0*})^2)^\top$.

A.6. *Prior for component allocations.* $p(u_i | \mathbf{w})$, $i = 1, \dots, N$, also follows from the assumed mixture model for the random effects and coincides with the discrete distribution with

$$P(u_i = k | \mathbf{w}) = w_k, \quad k = 1, \dots, K.$$

A.7. *Prior for mixture weights.* $p(\mathbf{w})$ is a Dirichlet distribution $\mathcal{D}(\delta, \dots, \delta)$, where δ is a fixed hyperparameter (prior sample size). That is,

$$p(\mathbf{w}) = \frac{\Gamma(K\delta)}{\Gamma^K(\delta)} \prod_{k=1}^K w_k^{\delta-1},$$

Weakly informative prior is obtained by setting δ to a small positive value, e.g., $\delta = 1$.

A.8. *Prior for mixture means.* $p(\boldsymbol{\mu}_k^*)$, $k = 1, \dots, K$, is a (multivariate) normal distribution $\mathcal{N}(\boldsymbol{\xi}_b, \mathbf{C}_b)$, where $\boldsymbol{\xi}_b$ and \mathbf{C}_b are fixed prior mean and covariance matrix, respectively, that is,

$$p(\boldsymbol{\mu}_k^*) \propto |\mathbf{C}_b|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}(\boldsymbol{\mu}_k^* - \boldsymbol{\xi}_b)^\top \mathbf{C}_b^{-1}(\boldsymbol{\mu}_k^* - \boldsymbol{\xi}_b)\right\}, \quad k = 1, \dots, K.$$

Following the suggestions of [Richardson and Green \(1997\)](#) adapted to the multivariate GLMM, weakly informative prior is obtained by taking $\boldsymbol{\xi}_b$ to be a sample mean of $\mathbf{b}_{EB,i}^{0*}$, $i = 1, \dots, N$ and $\mathbf{C}_b = \text{diag}((\mathbf{R}_{EB}^{0*})^2)$.

A.9. *Prior for mixture covariance matrices.* $p(\mathbf{D}_k^{*-1} | \boldsymbol{\gamma}_b)$, $k = 1, \dots, K$, is a Wishart distribution $\mathcal{W}(\zeta_b, \boldsymbol{\Xi}_b)$, where ζ_b are fixed prior degrees of freedom and $\boldsymbol{\Xi}_b = \text{diag}(\boldsymbol{\gamma}_b)$ is a diagonal scale matrix with random diagonal elements. That is,

$$(A.8) \quad p(\mathbf{D}_k^{*-1} | \boldsymbol{\gamma}_b) \propto |\boldsymbol{\Xi}_b|^{-\frac{\zeta_b}{2}} |\mathbf{D}_k^{*-1}|^{\frac{\zeta_b-d-1}{2}} \exp\left\{-\frac{1}{2}\text{tr}(\boldsymbol{\Xi}_b^{-1} \mathbf{D}_k^{*-1})\right\},$$

$$k = 1, \dots, K.$$

To get a weakly informative prior, we set ζ_b to a small positive number, e.g., $\zeta_b = d + 1$. Further observe that a priori, $\mathbf{E}(\mathbf{D}_k^{*-1} | \boldsymbol{\gamma}_b) = \zeta_b \boldsymbol{\Xi}_b$. As noted by [Richardson and Green \(1997\)](#) in the univariate i.i.d. context, it seems restrictive to suppose that a knowledge of the variability of the data, represented by \mathbf{R}_{EB}^{0*} in our case, implies much about the size of matrices \mathbf{D}_k^{*-1} , $k = 1, \dots, K$ and subsequently, it is not advisable to take a fixed scale matrix $\boldsymbol{\Xi}_b$ based on \mathbf{R}_{EB}^{0*} . For this reason, they suggest to introduce additional hierarchical level in the prior distribution. Their proposal adapted to

our multivariate setting then takes $\Xi_b = \text{diag}(\gamma_b)$ with independent gamma $\mathcal{G}(g_{b,1}, h_{b,1}), \dots, \mathcal{G}(g_{b,d}, h_{b,d})$ priors for $\gamma_{b,1}^{-1}, \dots, \gamma_{b,d}^{-1}$, the inverted diagonal elements of Ξ_b . That is,

$$p(\gamma_b^{-1}) = \prod_{l=1}^d p(\gamma_{b,l}^{-1}) = \prod_{l=1}^d \left\{ \frac{h_{b,l}^{g_{b,l}}}{\Gamma(g_{b,l})} (\gamma_{b,l}^{-1})^{g_{b,l}-1} \exp(-h_{b,l} \gamma_{b,l}^{-1}) \right\},$$

where $(g_{b,1}, h_{b,1}), \dots, (g_{b,d}, h_{b,d})$ are fixed hyperparameters. Weakly informative prior is obtained by setting $g_{b,l}$ to a small positive value, e.g., $g_{b,l} = 0.2$, $l = 1, \dots, d$, and setting $h_{b,l}$ to moderate multiple of $1/(R_{EB,l}^{0*})^2$, e.g., $h_{b,l} = 10/(R_{EB,l}^{0*})^2$, $l = 1, \dots, d$.

A.10. *Prior for GLMM dispersion parameters.* $p(\phi_r^{-1} | \gamma_{\phi,r})$, $r = 1, \dots, R$, is nondegenerate distribution only for such $r = 1, \dots, R$ for which the dispersion parameter ϕ_r of the corresponding GLMM is not constant by definition. With respect to the analysis of PBC910 data where we consider three types of GLMM's: (i) Gaussian with identity link, (ii) Bernoulli with logit link, (iii) Poisson with log link, this only applies to the Gaussian GLMM with identity link (\equiv LMM) for which ϕ_r is the residual variance σ_r^2 . In such situations, $p(\phi_r^{-1} | \gamma_{\phi,r})$ is a gamma distribution $\mathcal{G}(\zeta_{\phi,r}/2, \gamma_{\phi,r}^{-1}/2)$, that is,

$$p(\phi_r^{-1} | \gamma_{\phi,r}) = \frac{2^{-\frac{\zeta_{\phi,r}}{2}}}{\Gamma(\frac{\zeta_{\phi,r}}{2})} \gamma_{\phi,r}^{-\frac{\zeta_{\phi,r}}{2}} (\phi_r^{-1})^{\frac{\zeta_{\phi,r}}{2}-2} \exp\left(-\frac{1}{2} \gamma_{\phi,r}^{-1} \phi_r^{-1}\right).$$

Note that with this parameterization of the gamma distribution we have a direct correspondence to the univariate Wishart distribution with the same meaning for ζ and γ parameters as in the prior distribution for the inverted mixture covariance matrices, see (A.8). Weakly informative prior distribution is obtained by setting $\zeta_{\phi,r}$ to a small positive number, e.g., $\zeta_{\phi,r} = 2$. Further, similarly to the prior distribution for the inverted mixture covariance matrices, we will assume that $\gamma_{\phi,r}^{-1}$ is random and assign it a gamma $\mathcal{G}(g_{\phi,r}, h_{\phi,r})$ prior distribution, i.e.,

$$p(\gamma_{\phi,r}^{-1}) = \left\{ \frac{h_{\phi,r}^{g_{\phi,r}}}{\Gamma(g_{\phi,r})} (\gamma_{\phi,r}^{-1})^{g_{\phi,r}-1} \exp(-h_{\phi,r} \gamma_{\phi,r}^{-1}) \right\},$$

where $(g_{\phi,r}, h_{\phi,r})$ are fixed hyperparameters. This allows to specify weakly informative prior distribution and at the same time keeps the posterior distribution proper. Weakly informative prior is obtained by setting $g_{\phi,r}$ to a small positive value, e.g., $g_{\phi,r} = 0.2$, and setting $h_{\phi,r}$ to moderate multiple of $1/(R_{res,r}^{0*})^2$, e.g., $h_{\phi,r} = 10/(R_{res,r}^{0*})^2$, where $R_{res,r}^{0*}$ is a range of residuals from the initial ML fit to the r th (G)LMM.

A.11. *Prior for fixed effects.* $p(\boldsymbol{\alpha})$ is a (multivariate) normal distribution $\mathcal{N}(\boldsymbol{\xi}_\alpha, \mathbf{C}_\alpha)$, where $\boldsymbol{\xi}_\alpha$ and \mathbf{C}_α are fixed prior hyperparameters, i.e.,

$$(A.9) \quad p(\boldsymbol{\alpha}) \propto |\mathbf{C}_\alpha|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}(\boldsymbol{\alpha} - \boldsymbol{\xi}_\alpha)^\top \mathbf{C}_\alpha^{-1}(\boldsymbol{\alpha} - \boldsymbol{\xi}_\alpha)\right\}.$$

As usual in the context of hierarchical regression models, weakly informative prior is obtained by setting $\boldsymbol{\xi}_\alpha = \mathbf{0}$ (zero vector) and \mathbf{C}_α to diagonal matrix with large positive numbers, e.g., 10 000 on the diagonal. It is always useful to check that the variance of the (marginal) posterior distribution of $\boldsymbol{\alpha}$ is indeed considerably smaller.

A.12. *Joint prior distribution including random hyperparameters.* When we incorporate also random hyperparameters γ_b and γ_ϕ into the expression of the joint prior distribution, we obtain:

$$(A.10) \quad \left\{ \begin{aligned} & p(\boldsymbol{\psi}, \boldsymbol{\theta}^*, \mathbf{B}, \mathbf{u}, \gamma_b, \gamma_\phi) \\ &= p(\mathbf{B} | \boldsymbol{\theta}^*, \mathbf{u}) \times p(\mathbf{u} | \boldsymbol{\theta}^*) \\ &\quad \times p(\boldsymbol{\theta}^* | \gamma_b) \times p(\gamma_b^{-1}) \times p(\boldsymbol{\psi} | \gamma_\phi) \times p(\gamma_\phi^{-1}) \\ &= \prod_{i=1}^N \left\{ p(\mathbf{b}_i | \boldsymbol{\theta}^*, u_i) \times p(u_i | \mathbf{w}) \right\} \times p(\mathbf{w}) \\ &\quad \times \prod_{k=1}^K \left\{ p(\boldsymbol{\mu}_k^*) \times p(\mathbf{D}_k^{*-1} | \gamma_b) \right\} \times \prod_{l=1}^d p(\gamma_{b,l}^{-1}) \\ &\quad \times \prod_{r=1}^R \left\{ p(\phi_r^{-1} | \gamma_{\phi,r}) \times p(\gamma_{\phi,r}^{-1}) \right\} \times p(\boldsymbol{\alpha}), \end{aligned} \right.$$

where all factors appearing in (A.10) have been explained in Sec. A.5–A.11.

B. Posterior distribution and the sampling algorithm. In this appendix, we provide more detailed explanation of the Monte Carlo Markov chain (MCMC) algorithm mentioned in Komárek and Komárková (2012, Section 2.2) which is used to obtain the sample from the posterior distribution. As it is usual with the MCMC applications, we generate a sample from the joint posterior distribution of all model parameters:

$$(B.1) \quad \begin{cases} p(\boldsymbol{\psi}, \boldsymbol{\theta}^*, \mathbf{B}, \mathbf{u}, \boldsymbol{\gamma}_b, \boldsymbol{\gamma}_\phi | \mathbf{y}) \\ \propto p(\boldsymbol{\psi}, \boldsymbol{\theta}^*, \mathbf{B}, \mathbf{u}, \boldsymbol{\gamma}_b, \boldsymbol{\gamma}_\phi) \times L^{Bayes}(\boldsymbol{\psi}, \boldsymbol{\theta}^*, \mathbf{B}, \mathbf{u}, \boldsymbol{\gamma}_b, \boldsymbol{\gamma}_\phi), \end{cases}$$

where

$$L^{Bayes}(\boldsymbol{\psi}, \boldsymbol{\theta}^*, \mathbf{B}, \mathbf{u}, \boldsymbol{\gamma}_b, \boldsymbol{\gamma}_\phi) = p(\mathbf{y} | \boldsymbol{\psi}, \boldsymbol{\theta}^*, \mathbf{B}, \mathbf{u}, \boldsymbol{\gamma}_b, \boldsymbol{\gamma}_\phi)$$

is the likelihood of the Bayesian model.

B.1. *Likelihood for the Bayesian model.* Given the (Bayesian) model parameters, the likelihood is given by

$$(B.2) \quad \begin{aligned} L^{Bayes}(\boldsymbol{\psi}, \boldsymbol{\theta}^*, \mathbf{B}, \mathbf{u}, \boldsymbol{\gamma}_b, \boldsymbol{\gamma}_\phi) &= p(\mathbf{y} | \boldsymbol{\psi}, \boldsymbol{\theta}^*, \mathbf{B}, \mathbf{u}, \boldsymbol{\gamma}_b, \boldsymbol{\gamma}_\phi) \\ &= p(\mathbf{y} | \boldsymbol{\psi}, \mathbf{B}) = \prod_{i=1}^N \prod_{r=1}^R \prod_{j=1}^{n_{i,r}} p(y_{i,r,j} | \phi_r, \boldsymbol{\alpha}_r, \mathbf{b}_{i,r}), \end{aligned}$$

where the form of $p(y_{i,r,j} | \phi_r, \boldsymbol{\alpha}_r, \mathbf{b}_{i,r})$ follows from a particular GLMM, in which

$$(B.3) \quad \begin{aligned} h_r^{-1}\{E(Y_{i,r,j} | \boldsymbol{\alpha}_r, \mathbf{b}_{i,r})\} &= \eta_{i,r,j} = \mathbf{x}_{i,r,j}^\top \boldsymbol{\alpha}_r + \mathbf{z}_{i,r,j}^\top \mathbf{b}_{i,r}, \\ &i = 1, \dots, N, r = 1, \dots, R, j = 1, \dots, n_{i,r}, \end{aligned}$$

where h_r^{-1} is the link function for the r th response, $\eta_{i,r,j}$ is the linear predictor, and $\mathbf{x}_{i,r,j}$, $\mathbf{z}_{i,r,j}$ are $c_r \times 1$ and $d_r \times 1$, respectively, vectors of known covariates. It is assumed that $p(y_{i,r,j} | \phi_r, \boldsymbol{\alpha}_r, \mathbf{b}_{i,r})$ belongs to an exponential family. For notational convenience, we will assume that h_r^{-1} is a canonical link function which implies that

$$(B.4) \quad \begin{aligned} p(y_{i,r,j} | \phi_r, \boldsymbol{\alpha}_r, \mathbf{b}_{i,r}) &= p(y_{i,r,j} | \phi_r, \eta_{i,r,j}) \\ &= \exp\left\{ \frac{y_{i,r,j} \eta_{i,r,j} - q_r(\eta_{i,r,j})}{\phi_r} + k_r(y_{i,r,j}, \phi_r) \right\}, \end{aligned}$$

where k_r and q_r are appropriate distribution specific functions, see Table B.1 for their form in three most common situations of the Gaussian, Bernoulli and Poisson distribution.

In a sequel, we shall exploit additional notation. Let

TABLE B.1

Components of three most common generalized linear models. All distributions are parametrized such that the mean is λ .

	Distribution		
	Gaussian $\mathcal{N}(\lambda, \sigma^2)$	Bernoulli $\mathcal{A}(\lambda)$	Poisson $\mathcal{P}o(\lambda)$
Canonical link $h^{-1}(\lambda)$	λ	$\text{logit}(\lambda) = \log\left(\frac{\lambda}{1-\lambda}\right)$	$\log(\lambda)$
Inverted canonical link $h(\eta)$	η	$\text{ilogit}(\eta) = \frac{\exp(\eta)}{1 + \exp(\eta)}$	$\exp(\eta)$
Dispersion ϕ	σ^2	1	1
Variance $v(\phi, \eta)$	$\phi = \sigma^2$	$h(\eta) \{1 - h(\eta)\} = \lambda(1 - \lambda)$	$h(\eta) = \lambda$
$k(y, \phi)$	$\frac{\log(2\pi\phi)}{2} - \frac{y^2}{2\phi}$	0	$-\log(y!)$
$q(\eta)$	$\frac{\eta^2}{2}$	$\log\{1 + \exp(\eta)\}$	$\exp(\eta)$

- $n_{\bullet r} = \sum_{i=1}^N n_{i,r}$ be the number of all observations of the r th marker;
- $\mathbf{y}_{\bullet r} = (\mathbf{y}_{1,r}^\top, \dots, \mathbf{y}_{N,r}^\top)^\top$ be the vector of all observations of the r th marker;
- $\boldsymbol{\eta}_{i,r} = (\eta_{i,r,1}, \dots, \eta_{i,r,n_{i,r}})^\top$ be the vector of linear predictors for the observations of the r th marker of the i th subject;
- $\boldsymbol{\eta}_{\bullet r} = (\boldsymbol{\eta}_{1,r}^\top, \dots, \boldsymbol{\eta}_{N,r}^\top)^\top$ be the vector of linear predictors for all observations of the r th marker;
- $\lambda_{i,r,j} = \mathbb{E}(Y_{i,r,j} | \boldsymbol{\alpha}_r, \mathbf{b}_{i,r}) = \mathbb{E}(Y_{i,r,j} | \eta_{i,r,j})$, $i = 1, \dots, N$, $r = 1, \dots, R$, $j = 1, \dots, n_{i,r}$, be the mean of the (i, r, j) th response obtained from the GLMM (B.3) with the response density (B.4) which in general is given by

$$\lambda_{i,r,j} = \frac{\partial q_r}{\partial \eta}(\eta_{i,r,j}), \quad i = 1, \dots, N, \quad r = 1, \dots, R, \quad j = 1, \dots, n_{i,r};$$

- $\boldsymbol{\lambda}_{i,r} = (\lambda_{i,r,1}, \dots, \lambda_{i,r,n_{i,r}})^\top$, $i = 1, \dots, N$, $r = 1, \dots, R$, be the vector of means of the r th marker for the i th subject;
- $\boldsymbol{\lambda}_{\bullet r} = (\boldsymbol{\lambda}_{1,r}^\top, \dots, \boldsymbol{\lambda}_{N,r}^\top)^\top$, $r = 1, \dots, R$, be the vector of the means for the r th marker;
- $v_{i,r,j} = \text{var}(Y_{i,r,j} | \phi_r, \boldsymbol{\alpha}_r, \mathbf{b}_{i,r}) = \text{var}(Y_{i,r,j} | \phi_r, \eta_{i,r,j})$, $i = 1, \dots, N$, $r = 1, \dots, R$, $j = 1, \dots, n_{i,r}$, be the variance of the (i, r, j) th response obtained from the GLMM (B.3) with the response density (B.4). which in general is given by

$$v_{i,r,j} = \phi_r \frac{\partial^2 q_r}{\partial \eta^2}(\eta_{i,r,j}), \quad i = 1, \dots, N, \quad r = 1, \dots, R, \quad j = 1, \dots, n_{i,r},$$

see Table B.1 for explicit formulas in three most common situations;

- $\mathbf{v}_{i,r} = (v_{i,r,1}, \dots, v_{i,r,n_{i,r}})^\top$, $i = 1, \dots, N$, $r = 1, \dots, R$, be the vector of variances of the r th marker for the i th subject;
- $\mathbf{V}_{i,r} = \text{diag}(\mathbf{v}_{i,r})$, $i = 1, \dots, N$, $r = 1, \dots, R$, be the diagonal matrix with variances of the r th marker for the i th subject on a diagonal;
- $\mathbf{v}_{\bullet r} = (\mathbf{v}_{1,r}^\top, \dots, \mathbf{v}_{N,r}^\top)^\top$, $r = 1, \dots, R$, be the vector of the variances for the r th marker;
- $\mathbf{V}_{\bullet r} = \text{diag}(\mathbf{v}_{\bullet r})$, $r = 1, \dots, R$, be the diagonal matrix with variances for the r th marker on a diagonal;
- $\mathbf{X}_{i,r}$, $i = 1, \dots, N$, $r = 1, \dots, R$, be the matrix of covariates for the fixed effects from the GLMM (B.3) for the r th response of the i th subject, i.e.,

$$\mathbf{X}_{i,r} = \begin{pmatrix} \mathbf{x}_{i,r,1}^\top \\ \vdots \\ \mathbf{x}_{i,r,n_{i,r}}^\top \end{pmatrix};$$

- $\mathbf{X}_{\bullet r}$, $r = 1, \dots, R$ be the matrix of covariates for the fixed effects from the GLMM (B.3) for the r th response, i.e.,

$$\mathbf{X}_{\bullet r} = \begin{pmatrix} \mathbf{X}_{1,r} \\ \vdots \\ \mathbf{X}_{N,r} \end{pmatrix};$$

- $\mathbf{Z}_{i,r}$, $i = 1, \dots, N$, $r = 1, \dots, R$ be the matrix of covariates for the random effects from the GLMM (B.3) for the r th response of the i th subject, i.e.,

$$\mathbf{Z}_{i,r} = \begin{pmatrix} \mathbf{z}_{i,r,1}^\top \\ \vdots \\ \mathbf{z}_{i,r,n_{i,r}}^\top \end{pmatrix};$$

- $q_{i,r,j} = q_r(\eta_{i,r,1})$, $i = 1, \dots, N$, $r = 1, \dots, R$, $j = 1, \dots, n_{i,r}$;
- $\mathbf{q}_{i,r} = (q_{i,r,1}, \dots, q_{i,r,n_{i,r}})^\top$, $i = 1, \dots, N$, $r = 1, \dots, R$;
- $\mathbf{q}_{\bullet r} = (\mathbf{q}_{1,r}^\top, \dots, \mathbf{q}_{N,r}^\top)^\top$, $r = 1, \dots, R$;
- $k_{i,r,j} = k_r(y_{i,r,j}, \phi_r)$, $i = 1, \dots, N$, $r = 1, \dots, R$, $j = 1, \dots, n_{i,r}$;
- $\mathbf{k}_{i,r} = (k_{i,r,1}, \dots, k_{i,r,n_{i,r}})^\top$, $i = 1, \dots, N$, $r = 1, \dots, R$;
- $\mathbf{k}_{\bullet r} = (\mathbf{k}_{1,r}^\top, \dots, \mathbf{k}_{N,r}^\top)^\top$, $r = 1, \dots, R$;
- \mathbf{s}_r , $r = 1, \dots, R$, be the part of the shift vector \mathbf{s} , see Expression (A.7), which corresponds to the r th marker, i.e., $\mathbf{s} = (\mathbf{s}_1^\top, \dots, \mathbf{s}_R^\top)^\top$;
- \mathbf{S}_r , $r = 1, \dots, R$ be the diagonal block of the scale matrix \mathbf{S} , see Expression (A.7), which corresponds to the r th marker, i.e., $\mathbf{S} = \text{block-diag}(\mathbf{S}_1, \dots, \mathbf{S}_R)$.

Consequently, the likelihood (B.2) can be written as

$$\begin{aligned}
 \text{(B.5)} \quad L^{Bayes}(\boldsymbol{\psi}, \boldsymbol{\theta}^*, \mathbf{B}, \mathbf{u}, \boldsymbol{\gamma}_b, \boldsymbol{\gamma}_\phi) &= L^{Bayes}(\boldsymbol{\psi}, \mathbf{B}) \\
 &= \exp \left[\sum_{i=1}^N \sum_{r=1}^R \phi_r^{-1} \{ \mathbf{y}_{i,r}^\top \boldsymbol{\eta}_{i,r} - \mathbf{1}^\top \mathbf{q}_{i,r} \} + \mathbf{1}^\top \mathbf{k}_{i,r} \right] \\
 &= \exp \left[\sum_{r=1}^R \phi_r^{-1} \{ \mathbf{y}_{\bullet r}^\top \boldsymbol{\eta}_{\bullet r} - \mathbf{1}^\top \mathbf{q}_{\bullet r} \} + \mathbf{1}^\top \mathbf{k}_{\bullet r} \right],
 \end{aligned}$$

where $\mathbf{1}$ is a vector of ones of appropriate dimension.

B.2. Monte Carlo Markov chain. In this Section, we introduce the MCMC procedure to obtain a sample

$$\mathcal{S}_M = \left\{ (\boldsymbol{\psi}^{(m)}, \boldsymbol{\theta}^{*(m)}, \mathbf{B}^{(m)}, \mathbf{u}^{(m)}, \boldsymbol{\gamma}_b^{(m)}, \boldsymbol{\gamma}_\phi^{(m)}) : m = 1, \dots, M \right\}$$

from the joint posterior distribution (B.1). We exploit the Gibbs algorithm and update the parameters in blocks of related parameters by sampling from the appropriate full conditional distributions, further denoted as $p(\cdot | \dots)$. In the remainder of this Section, we derive these full conditional distributions and explain how to sample from them in situations when they do not have a form of any classical distribution. We exploit the following additional notation:

- $N_k = \sum_{i=1}^N \mathbb{I}(u_i = k)$, $k = 1, \dots, K$, will denote the number of subjects who are assigned into the k th mixture component according to the values of component allocations. Note that $\sum_{k=1}^K N_k = N$, the number of subjects;
- $\overline{\mathbf{b}}_{[k]}^*$, $k = 1, \dots, K$, will be the sample mean of shifted-scaled values of random effects of those subjects assigned into the k th mixture component according to the values of component allocations \mathbf{u} , i.e.,

$$\overline{\mathbf{b}}_{[k]}^* = N_k^{-1} \sum_{i:u_i=k} \mathbf{b}_i^*, \quad k = 1, \dots, K.$$

B.2.1. Mixture weights. Mixture weights form one block of the Gibbs algorithm. The corresponding full conditional distribution is the Dirichlet distribution $\mathcal{D}(\delta + N_1, \dots, \delta + N_K)$. That is,

$$p(\mathbf{w} | \dots) = \frac{\Gamma(K\delta + N)}{\prod_{k=1}^K \Gamma(\delta + N_k)} \prod_{k=1}^K w_k^{\delta + N_k - 1},$$

which is the distribution we can directly sample from.

B.2.2. *Mixture means.* The second block of the Gibbs algorithm is composed of mixture means. Their full conditional distribution is a product of K independent normal distributions, i.e.,

$$p(\boldsymbol{\mu}_1^*, \dots, \boldsymbol{\mu}_k^* | \dots) = \prod_{k=1}^K p(\boldsymbol{\mu}_k^* | \dots) \sim \prod_{k=1}^K \mathcal{N}(\mathbf{E}(\boldsymbol{\mu}_k^* | \dots), \text{var}(\boldsymbol{\mu}_k^* | \dots)),$$

where

$$\begin{aligned} \text{var}(\boldsymbol{\mu}_k^* | \dots) &= (N_k \mathbf{D}_k^{*-1} + \mathbf{C}_b^{-1})^{-1}, \\ \mathbf{E}(\boldsymbol{\mu}_k^* | \dots) &= \text{var}(\boldsymbol{\mu}_k^* | \dots) (N_k \mathbf{D}_k^{*-1} \bar{\mathbf{b}}_{[k]}^* + \mathbf{C}_b^{-1} \boldsymbol{\xi}_b), \quad k = 1, \dots, K. \end{aligned}$$

B.2.3. *Mixture covariance matrices.* The third block of the Gibbs algorithm is formed of the inversions of the mixture covariance matrices. Their full conditional distribution is a product of K independent Wishart distributions, i.e.,

$$p(\mathbf{D}_1^{*-1}, \dots, \mathbf{D}_k^{*-1} | \dots) = \prod_{k=1}^K p(\mathbf{D}_k^{*-1} | \dots) \sim \prod_{k=1}^K \mathcal{W}(\tilde{\zeta}_{b,k}, \tilde{\boldsymbol{\Xi}}_{b,k}),$$

where

$$\begin{aligned} \tilde{\zeta}_{b,k} &= \zeta_b + N_k, \\ \tilde{\boldsymbol{\Xi}}_{b,k} &= \left\{ \boldsymbol{\Xi}_b^{-1} + \sum_{i:u_i=k} (\mathbf{b}_i^* - \boldsymbol{\mu}_k^*)(\mathbf{b}_i^* - \boldsymbol{\mu}_k^*)^\top \right\}^{-1}, \quad k = 1, \dots, K. \end{aligned}$$

B.2.4. *Variance hyperparameters for mixture covariance matrices.* In the fourth block of the Gibbs algorithm, we update the variance hyperparameters γ_b^{-1} whose inversions form a random diagonal of the scale matrix $\boldsymbol{\Xi}_b$ in the Wishart prior for the inversions of the mixture covariance matrices. Their full conditional distribution is a product of independent gamma distributions, i.e.,

$$p(\gamma_b^{-1} | \dots) = \prod_{l=1}^d p(\gamma_{b,l}^{-1} | \dots) \sim \prod_{l=1}^d \mathcal{G}(\tilde{g}_{b,l}, \tilde{h}_{b,l}),$$

where

$$\tilde{g}_{b,l} = g_{b,l} + \frac{K \zeta_b}{2}, \quad \tilde{h}_{b,l} = h_{b,l} + \frac{1}{2} \sum_{k=1}^K Q_k(l, l), \quad l = 1, \dots, d,$$

where $Q_k(l, l)$ is the l th diagonal element of the matrix \mathbf{D}_k^{*-1} .

B.2.5. *Component allocations.* In the fifth block of the Gibbs algorithm, we update the latent component allocations \mathbf{u} . Their full conditional distribution is a product over subjects ($i = 1, \dots, N$) of the following discrete distributions:

$$P(u_i = k \mid \dots) = \frac{w_k \varphi(\mathbf{b}_i^* \mid \boldsymbol{\mu}_k^*, \mathbf{D}_k^*)}{\sum_{l=1}^K w_l \varphi(\mathbf{b}_i^* \mid \boldsymbol{\mu}_l^*, \mathbf{D}_l^*)}, \quad k = 1, \dots, K.$$

B.2.6. *Fixed effects.* The sixth block of the Gibbs algorithm consists in updating the vector of fixed effects $\boldsymbol{\alpha}$. The full conditional distribution factorizes with respect to the R longitudinal markers involved in a model, i.e.,

$$p(\boldsymbol{\alpha} \mid \dots) = \prod_{r=1}^R p(\boldsymbol{\alpha}_r \mid \dots).$$

To proceed, we introduce some additional notation. Let for each $r = 1, \dots, R$,

$$L_{\alpha,r}^{Bayes}(\boldsymbol{\psi}, \mathbf{B}) = \exp\left\{\phi_r^{-1}(\mathbf{y}_{\bullet r}^\top \boldsymbol{\eta}_{\bullet r} - \mathbf{1}^\top \mathbf{q}_{\bullet r}) + \mathbf{1}^\top \mathbf{k}_{\bullet r}\right\},$$

be the part of the likelihood (B.5) pertaining to the r th marker. Combining this with the prior distribution (A.9) leads to the following full conditional distribution for the vector of the fixed effects from the r th GLMM:

$$(B.6) \quad p(\boldsymbol{\alpha}_r \mid \dots)$$

$$\propto \exp\left\{\phi_r^{-1}(\mathbf{y}_{\bullet r}^\top \boldsymbol{\eta}_{\bullet r} - \mathbf{1}^\top \mathbf{q}_{\bullet r}) - \frac{1}{2}(\boldsymbol{\alpha}_r - \boldsymbol{\xi}_{\alpha,r})^\top \mathbf{C}_{\alpha,r}^{-1}(\boldsymbol{\alpha}_r - \boldsymbol{\xi}_{\alpha,r})\right\},$$

$r = 1, \dots, R$, where $\boldsymbol{\xi}_{\alpha,r}$ and $\mathbf{C}_{\alpha,r}$ are appropriate subvector and submatrix, respectively, of the prior mean $\boldsymbol{\xi}_\alpha$ and prior covariance matrix \mathbf{C}_α , respectively. In a situation when the r th GLMM distribution (B.4) is normal $\mathcal{N}(\eta_{i,r,j}, \phi_r)$, it can easily be shown that (B.6) is a (multivariate) normal distribution with the variance and mean given by

$$(B.7) \quad \text{var}(\boldsymbol{\alpha}_r \mid \dots) = (\phi_r^{-1} \mathbf{X}_{\bullet r}^\top \mathbf{X}_{\bullet r} + \mathbf{C}_{\alpha,r}^{-1})^{-1},$$

$$(B.8) \quad \mathbf{E}(\boldsymbol{\alpha}_r \mid \dots) = \text{var}(\boldsymbol{\alpha}_r \mid \dots) (\phi_r^{-1} \mathbf{X}_{\bullet r}^\top \mathbf{e}_{\alpha,r} + \mathbf{C}_{\alpha,r}^{-1} \boldsymbol{\xi}_{\alpha,r}),$$

where

$$\mathbf{e}_{\alpha,r} = \begin{pmatrix} \mathbf{e}_{\alpha,1,r} \\ \vdots \\ \mathbf{e}_{\alpha,N,r} \end{pmatrix} = \begin{pmatrix} \mathbf{y}_{1,r} - \mathbf{Z}_{1,r} \mathbf{b}_{1,r} \\ \vdots \\ \mathbf{y}_{N,r} - \mathbf{Z}_{N,r} \mathbf{b}_{N,r} \end{pmatrix},$$

from which we can easily sample from. Nevertheless, in general, (B.6) does not have a form of any classically used distribution. In that case, we follow

a proposal of [Gamerman \(1997\)](#) and update $\boldsymbol{\alpha}_r$ using a Metropolis-Hastings step with a normal proposal based on appropriately tuned second-order approximation to (B.6) created using one Newton-Raphson step. Let in the m th MCMC iteration $\boldsymbol{\alpha}_r^{(m-1)}$ denote the previously sampled value of $\boldsymbol{\alpha}_r$, and $\boldsymbol{\lambda}_{\bullet r}^{(m-1)}$ the vector of residuals $\boldsymbol{\lambda}_{\bullet r}$ calculated using $\boldsymbol{\alpha}_r = \boldsymbol{\alpha}_r^{(m-1)}$. Further, let

$$\begin{aligned}\mathcal{U}_{\alpha,r}(\boldsymbol{\alpha}_r) &= \frac{\partial \log\{L_{\alpha,r}^{Bayes}(\boldsymbol{\psi}, \mathbf{B})\}}{\partial \boldsymbol{\alpha}_r} = \phi_r^{-1} \mathbf{X}_{\bullet r}^\top (\mathbf{y}_{\bullet r} - \boldsymbol{\lambda}_{\bullet r}), \\ \mathcal{J}_{\alpha,r}(\boldsymbol{\alpha}_r) &= -\frac{\partial^2 \log\{L_{\alpha,r}^{Bayes}(\boldsymbol{\psi}, \mathbf{B})\}}{\partial \boldsymbol{\alpha}_r \partial \boldsymbol{\alpha}_r^\top} = \phi_r^{-2} \mathbf{X}_{\bullet r}^\top \mathbf{V}_{\bullet r} \mathbf{X}_{\bullet r},\end{aligned}$$

be the score and observed information matrix from the r th GLMM with respect to the vector of fixed effects $\boldsymbol{\alpha}_r$. A new value of $\boldsymbol{\alpha}_r$ is proposed by sampling from a (multivariate) normal distribution with the variance and mean given by

$$(B.9) \quad \widehat{\text{var}}(\boldsymbol{\alpha}_r | \dots) = \sigma_{\alpha,r} \left\{ \mathcal{J}_{\alpha,r}(\boldsymbol{\alpha}_r^{(m-1)}) + \mathbf{C}_{\alpha,r}^{-1} \right\}^{-1},$$

$$(B.10) \quad \widehat{\mathbf{E}}(\boldsymbol{\alpha}_r | \dots) = \boldsymbol{\alpha}_r^{(m-1)} + \Delta_{\alpha,r}^{(m)},$$

where

$$(B.11) \quad \begin{aligned}\Delta_{\alpha,r}^{(m)} &= \left\{ \mathcal{J}_{\alpha,r}(\boldsymbol{\alpha}_r^{(m-1)}) + \mathbf{C}_{\alpha,r}^{-1} \right\}^{-1} \\ &\quad \left\{ \mathcal{U}_{\alpha,r}(\boldsymbol{\alpha}_r^{(m-1)}) - \mathbf{C}_{\alpha,r}^{-1} (\boldsymbol{\alpha}_r^{(m-1)} - \boldsymbol{\xi}_{\alpha,r}) \right\} \\ &= \left(\phi_r^{-2} \mathbf{X}_{\bullet r}^\top \mathbf{V}_{\bullet r} \mathbf{X}_{\bullet r} + \mathbf{C}_{\alpha,r}^{-1} \right)^{-1} \\ &\quad \left\{ \phi_r^{-1} \mathbf{X}_{\bullet r}^\top (\mathbf{y}_{\bullet r} - \boldsymbol{\lambda}_{\bullet r}^{(m-1)}) - \mathbf{C}_{\alpha,r}^{-1} (\boldsymbol{\alpha}_r^{(m-1)} - \boldsymbol{\xi}_{\alpha,r}) \right\},\end{aligned}$$

and $\sigma_{\alpha,r} > 0$ is a tuning parameter. In practical analyzes, we recommend to start with $\sigma_{\alpha,r} = 1$ which often leads to reasonable mixing. Note that in the case of normally distributed response, setting $\sigma_{\alpha,r} = 1$ leads to (B.9) which is identical to (B.7) and (B.10) coincides with (B.8).

Finally, we remark that both the proposal covariance matrix (B.9) (with $\sigma_{\alpha,r} = 1$) and the Newton-Raphson step (B.11) can efficiently be obtained as a least squares solution and QR decomposition to a particular problem. To show that, let $\mathbf{L}_{\alpha,r}$ be the Cholesky decomposition of the inversion of the prior covariance matrix $\mathbf{C}_{\alpha,r}$, i.e., $\mathbf{C}_{\alpha,r}^{-1} = \mathbf{L}_{\alpha,r} \mathbf{L}_{\alpha,r}^\top$. Further, let

$$\widetilde{\mathbf{X}}_{\bullet r} = \begin{pmatrix} \phi_r^{-1} \mathbf{V}_{\bullet r}^{1/2} \mathbf{X}_{\bullet r} \\ \mathbf{L}_{\alpha,r}^\top \end{pmatrix}, \quad \widetilde{\mathbf{e}}_{\bullet r} = \begin{pmatrix} \mathbf{V}_{\bullet r}^{-1/2} (\mathbf{y}_{\bullet r} - \boldsymbol{\lambda}_{\bullet r}^{(m-1)}) \\ -\mathbf{L}_{\alpha,r}^\top (\boldsymbol{\alpha}_r^{(m-1)} - \boldsymbol{\xi}_{\alpha,r}) \end{pmatrix}.$$

It follows from (B.11) that

$$\Delta_{\alpha,r}^{(m)} = \left(\widetilde{\mathbf{X}}_{\bullet,r}^\top \widetilde{\mathbf{X}}_{\bullet,r} \right)^{-1} \widetilde{\mathbf{X}}_{\bullet,r}^\top \widetilde{\mathbf{e}}_r,$$

i.e., $\Delta_{\alpha,r}^{(m)}$ is the least squares solution to $\widetilde{\mathbf{X}}_{\bullet,r} \delta = \widetilde{\mathbf{e}}_r$. That is, if $\widetilde{\mathbf{X}}_{\bullet,r} = \widetilde{\mathbf{Q}}_{\bullet,r} \widetilde{\mathbf{R}}_{\bullet,r}$ is the QR decomposition of the matrix $\widetilde{\mathbf{X}}_{\bullet,r}$, we have

$$\Delta_{\alpha,r}^{(m)} = \widetilde{\mathbf{R}}_{\bullet,r}^{-1} \widetilde{\mathbf{Q}}_{\bullet,r}^\top \widetilde{\mathbf{e}}_r,$$

and also

$$\{\widehat{\text{var}}(\boldsymbol{\alpha}_r | \dots)\}^{-1} = \sigma_{\alpha,r}^{-1} \widetilde{\mathbf{R}}_{\bullet,r}^\top \widetilde{\mathbf{R}}_{\bullet,r}.$$

B.2.7. Random effects. The seventh block of the Gibbs algorithm consists in updating the vectors of random effects $\mathbf{b}_1, \dots, \mathbf{b}_N$. The full conditional distribution factorizes with respect to N subjects, i.e.,

$$p(\mathbf{B} | \dots) = \prod_{i=1}^N p(\mathbf{b}_i | \dots).$$

The remainder of this part is similar to Paragraph B.2.6. First, we introduce additional notation. Let for each $i = 1, \dots, N$,

$$(B.12) \quad L_{b,i}^{Bayes}(\boldsymbol{\psi}, \mathbf{B}) = \exp \left\{ \sum_{r=1}^R \phi_r^{-1} (\mathbf{y}_{i,r}^\top \boldsymbol{\eta}_{i,r} - \mathbf{1}^\top \mathbf{q}_{i,r}) + \mathbf{1}^\top \mathbf{k}_{i,r} \right\},$$

be the part of the likelihood (B.5) pertaining to the i th subject. If we combine this with the model (A.6) for random effects, we obtain that the full conditional distribution for the i th random effect vector \mathbf{b}_i is

$$\begin{aligned} & p(\mathbf{b}_i | \dots) \\ & \propto \exp \left\{ \sum_{r=1}^R \phi_r^{-1} (\mathbf{y}_{i,r}^\top \boldsymbol{\eta}_{i,r} - \mathbf{1}^\top \mathbf{q}_{i,r}) - \frac{1}{2} (\mathbf{b}_i - \mathbf{s} - \mathbf{S} \boldsymbol{\mu}_{u_i}^*)^\top (\mathbf{S} \mathbf{D}_{u_i}^* \mathbf{S}^\top)^{-1} (\mathbf{b}_i - \mathbf{s} - \mathbf{S} \boldsymbol{\mu}_{u_i}^*) \right\}, \end{aligned}$$

When working with shifted-scaled random effects, see Expression (A.7), we obtain the following full conditional distribution

$$(B.13) \quad \begin{aligned} & p(\mathbf{b}_i | \dots) \\ & \propto \exp \left\{ \sum_{r=1}^R \phi_r^{-1} (\mathbf{y}_{i,r}^\top \boldsymbol{\eta}_{i,r} - \mathbf{1}^\top \mathbf{q}_{i,r}) - \frac{1}{2} (\mathbf{b}_i^* - \boldsymbol{\mu}_{u_i}^*)^\top \mathbf{D}_{u_i}^{*-1} (\mathbf{b}_i^* - \boldsymbol{\mu}_{u_i}^*) \right\}, \end{aligned}$$

from which we have to sample from during this block of the Gibbs algorithm. In a situation when all R GLMM distributions (B.4) are normal $\mathcal{N}(\eta_{i,r,j}, \phi_r)$, $r = 1, \dots, R$, it can easily be shown that (B.13) is a (multivariate) normal distribution with the variance and mean given by

$$(B.14) \quad \text{var}(\mathbf{b}_i^* | \dots) = (\mathbf{J}_{b,i} + \mathbf{D}_{u_i}^{*-1})^{-1}$$

$$(B.15) \quad \mathbf{E}(\mathbf{b}_i^* | \dots) = \text{var}(\mathbf{b}_i^* | \dots) (\mathbf{U}_{b,i} + \mathbf{D}_{u_i}^{*-1} \boldsymbol{\mu}_{u_i}^*),$$

where

$$\mathbf{U}_{b,i} = \begin{pmatrix} \phi_1^{-1} \mathbf{S}_1^\top \mathbf{Z}_{i,1}^\top (\mathbf{y}_{i,1} - \mathbf{X}_{i,1} \boldsymbol{\alpha}_1 - \mathbf{Z}_{i,1} \mathbf{s}_1) \\ \vdots \\ \phi_R^{-1} \mathbf{S}_R^\top \mathbf{Z}_{i,R}^\top (\mathbf{y}_{i,R} - \mathbf{X}_{i,R} \boldsymbol{\alpha}_R - \mathbf{Z}_{i,R} \mathbf{s}_R) \end{pmatrix},$$

$$\mathbf{J}_{b,i} = \text{block-diag}(\phi_r^{-1} \mathbf{S}_r^\top \mathbf{Z}_{i,r}^\top \mathbf{Z}_{i,r} \mathbf{S}_r : r = 1, \dots, R),$$

$i = 1, \dots, N$, and we can directly update the values of random effects by sampling from this multivariate normal distribution. Analogously to Paragraph B.2.6, the full conditional distribution (B.13) does not have a form of any classically used distribution as soon as at least one of R GLMM distributions (B.4) is not normal. In that case, we update each \mathbf{b}_i^* (and \mathbf{b}_i) using a Metropolis-Hastings step with a normal proposal based on appropriately tuned second-order approximation to (B.13) created using one Newton-Raphson step. Let in the m th MCMC iteration, $\mathbf{b}_i^{*(m-1)}$ denote the previously sampled value of \mathbf{b}_i^* . Further, let

$$\mathcal{U}_{b,i}(\mathbf{b}_i^*) = \frac{\partial \log \{L_{b,i}^{Bayes}(\boldsymbol{\psi}, \mathbf{B})\}}{\partial \mathbf{b}_i^*} = \begin{pmatrix} \phi_1^{-1} \mathbf{S}_1^\top \mathbf{Z}_{i,1}^\top (\mathbf{y}_{i,1} - \boldsymbol{\lambda}_{i,1}^{(m-1)}) \\ \vdots \\ \phi_R^{-1} \mathbf{S}_R^\top \mathbf{Z}_{i,R}^\top (\mathbf{y}_{i,R} - \boldsymbol{\lambda}_{i,R}^{(m-1)}) \end{pmatrix},$$

$$\begin{aligned} \mathcal{J}_{b,i}(\mathbf{b}_i^*) &= - \frac{\partial^2 \log \{L_{b,i}^{Bayes}(\boldsymbol{\psi}, \mathbf{B})\}}{\partial \mathbf{b}_i^* \partial \mathbf{b}_i^{*\top}} \\ &= \text{block-diag}(\phi_r^{-2} \mathbf{S}_r^\top \mathbf{Z}_{i,r}^\top \mathbf{V}_{i,r} \mathbf{Z}_{i,r} \mathbf{S}_r : r = 1, \dots, R), \end{aligned}$$

be the score and observed information matrix with respect to the shifted-scaled vector of random effects corresponding to the likelihood contribution (B.12). A new value of \mathbf{b}_i^* is proposed by sampling from a (multivariate) normal distribution with the variance and mean given by

$$(B.16) \quad \widehat{\text{var}}(\mathbf{b}_i^* | \dots) = \sigma_b \left\{ \mathcal{J}_{b,i}(\mathbf{b}_i^{*(m-1)}) + \mathbf{D}_{u_i}^{*-1} \right\}^{-1},$$

$$(B.17) \quad \widehat{\mathbf{E}}(\mathbf{b}_i^* | \dots) = \mathbf{b}_i^{*(m-1)} + \Delta_{b,i}^{(m)},$$

where

$$(B.18) \quad \Delta_{b,i}^{(m)} = \left\{ \mathcal{J}_{b,i}(\mathbf{b}_i^{*(m-1)}) + \mathbf{D}_{u_i}^{*-1} \right\}^{-1} \left\{ \mathcal{U}_{b,i}(\mathbf{b}_i^{*(m-1)}) - \mathbf{D}_{u_i}^{*-1}(\mathbf{b}_i^{*(m-1)} - \boldsymbol{\mu}_{u_i}^*) \right\},$$

and $\sigma_b > 0$ is a tuning parameter. Also in this case, we recommend to start with $\sigma_b = 1$ which often leads to reasonable mixing. Note that in the case of normally distributed response, setting $\sigma_b = 1$ leads to (B.16) which is identical to (B.14) and (B.17) coincides with (B.15). Analogously to Paragraph B.2.6, both the proposal covariance matrix (B.16) (with $\sigma_b = 1$) and the Newton-Raphson step (B.18) can efficiently be obtained as a least squares solution and QR decomposition to a particular problem.

B.2.8. GLMM dispersion parameters. The eight block of the Gibbs algorithm involves sampling the GLMM dispersion parameters ϕ_1, \dots, ϕ_R for those $r = 1, \dots, R$ for which ϕ_r of the corresponding GLMM is not constant by definition. In these situations, the full conditional distribution of the inverted dispersion parameter is given by

$$(B.19) \quad p(\phi_r^{-1} | \dots) \propto (\phi_r^{-1})^{\frac{\zeta_{\phi,r}-2}{2}} \exp \left\{ -\frac{1}{2} \gamma_{\phi,r}^{-1} \phi_r^{-1} + \phi_r^{-1} (\mathbf{y}_{\bullet r}^\top \boldsymbol{\eta}_{\bullet r} - \mathbf{1}^\top \mathbf{q}_{\bullet r}) + \mathbf{1}^\top \mathbf{k}_{\bullet r} \right\}.$$

If we concentrate on three most common situations described in Table B.1, the only case when it is necessary to sample the dispersion parameter ϕ_r^{-1} is normally distributed response with $\phi_r = \sigma_r^2$ being the error variance of the corresponding linear mixed model. In this case the full conditional distribution (B.19) is the gamma distribution, i.e.,

$$p(\phi_r^{-1} | \dots) \sim \mathcal{G}(\tilde{\zeta}_{\phi,r}/2, \tilde{\gamma}_{\phi,r}^{-1}/2),$$

where

$$\tilde{\zeta}_{\phi,r} = \zeta_{\phi,r} + n_{\bullet r}, \quad \tilde{\gamma}_{\phi,r} = \left\{ \gamma_{\phi,r}^{-1} + (\mathbf{y}_{\bullet r} - \boldsymbol{\eta}_{\bullet r})^\top (\mathbf{y}_{\bullet r} - \boldsymbol{\eta}_{\bullet r}) \right\}^{-1}.$$

B.2.9. Variance hyperparameter for dispersions. The last, ninth block of the Gibbs algorithm consists of sampling the variance hyperparameters $\gamma_{\phi,r}^{-1}$ for those $r = 1, \dots, R$ for which ϕ_r of the corresponding GLMM is not constant by definition. In these situations, the full conditional distribution of $\gamma_{\phi,r}^{-1}$ is the gamma distribution, i.e.,

$$p(\gamma_{\phi,r}^{-1} | \dots) \sim \mathcal{G}(\tilde{g}_{\phi,r}, \tilde{h}_{\phi,r}),$$

where

$$\tilde{g}_{\phi,r} = g_{\phi,r} + \frac{\zeta_{\phi,r}}{2}, \quad \tilde{h}_{\phi,r} = h_{\phi,r} + \frac{\phi_r^{-1}}{2}.$$

C. Analysis of the Mayo Clinic PBC Data. This section supplements the main paper by some details on the analysis of the PBC910 data. The MMGLMM for the clustering of patients included in the PBC910 data is based on longitudinal measurements of

- (1) logarithmic serum bilirubin (`lbili`, $Y_{i,1,j}$),
- (2) platelet counts (`platelet`, $Y_{i,2,j}$),
- (3) dichotomous presence of blood vessel malformations (`spiders`, $Y_{i,3,j}$),

with assumed (1) Gaussian, (2) Poisson, and (3) Bernoulli distribution, respectively and the mean structure (3) given by

$$(C.1) \quad \begin{cases} \mathbb{E}(Y_{i,1,j} | \mathbf{b}_{i,1}) = b_{i,1,1} + b_{i,1,2} t_{i,1,j}, \\ \log\{\mathbb{E}(Y_{i,2,j} | \mathbf{b}_{i,2})\} = b_{i,2,1} + b_{i,2,2} t_{i,2,j}, \\ \text{logit}\{\mathbb{E}(Y_{i,3,j} | b_{i,3}, \alpha_3)\} = b_{i,3} + \alpha_3 t_{i,3,j}, \end{cases}$$

$i = 1, \dots, N$, $j = 1, \dots, n_{i,r}$, $r = 1, 2, 3$, where $1 \leq n_{i,r} \leq 5$. In model (C.1), $t_{i,r,j}$ is the time in months from the start of follow-up when the value of $Y_{i,r,j}$ was obtained.

C.1. *Summary of model parameters.* In the main analysis, we shall try to classify patients in two groups and hence a two component mixture ($K = 2$) will be considered in the distribution of random effects. In summary, the considered model involves the following parameters of the main interest.

- Random effect vectors $\mathbf{b}_i = (b_{i,1,1}, b_{i,1,2}, b_{i,2,1}, b_{i,2,2}, b_{i,3})^\top$, $i = 1, \dots, N$, where $b_{i,1,1}$ and $b_{i,1,2}$ are the random intercept and random slope from the linear mixed model for logarithmic bilirubin, $b_{i,2,1}$ and $b_{i,2,2}$ are the random intercept and random slope from the log-Poisson model for platelet counts and $b_{i,3}$ is the random intercept from the logit model for presence of blood vessel malformations.
- Mixture parameters $w_k, \boldsymbol{\mu}_k^*, \mathbf{D}_k^*$, $k = 1, 2$ and their counterparts $\boldsymbol{\mu}_k = \mathbf{s} + \mathbf{S}\boldsymbol{\mu}_k^*$, $\mathbf{D}_k = \mathbf{S}\mathbf{D}_k^*\mathbf{S}^\top$, $k = 1, 2$ which represent the mean and the covariance matrix for each cluster.
- Dispersion parameter ϕ_1 from the linear mixed model for logarithmic bilirubin. In this case, ϕ_1 is the residual variance and we will denote it also as $\sigma_1^2 = \phi_1$.
- Fixed effects $\boldsymbol{\alpha} = \alpha_3$, where α_3 is the slope from the logit model for presence of blood vessel malformations.

For the purpose of estimation, we have

$$\boldsymbol{\psi} = (\sigma_1^2, \alpha_3)^\top, \quad \boldsymbol{\theta}^* = (\mathbf{w}^\top, \boldsymbol{\mu}_1^{*\top}, \boldsymbol{\mu}_2^{*\top}, \text{vec}(\mathbf{D}_1^*), \text{vec}(\mathbf{D}_2^*))^\top.$$

C.2. *Fixed hyperparameters, shift vector, scale matrix.* The values of fixed hyperparameters, the shift vector and the scale matrix chosen according to the guidelines given in Section A.4 were as follows.

$$\begin{aligned} \mathbf{s} &= (0.315, 0.00765, 5.53, -0.00663, -2.75)^\top, \\ \mathbf{S} &= \text{diag}(0.864, 0.02008, 0.35, 0.01565, 3.23), \\ \delta &= 1, \\ \boldsymbol{\xi}_b &= (0, 0, 0, 0, 0)^\top, \quad \mathbf{C}_b = \text{diag}(36, 36, 36, 36, 36)^\top, \\ \zeta_b &= 6, \quad g_{b,l} = 0.2, \quad h_{b,l} = 0.278, \quad l = 1, \dots, 5, \\ \zeta_{\phi,1} &= 2, \quad g_{\phi,1} = 0.2, \quad h_{\phi,1} = 2.76, \\ \boldsymbol{\xi}_\alpha &= 0, \quad \mathbf{C}_\alpha = 10\,000. \end{aligned}$$

C.3. *Monte Carlo Markov chain.* The reported results are based on 10 000 iterations of 1:100 thinned MCMC obtained after a burn-in period of 1 000 iterations. This took 75 minutes on Intel Core 2 Duo 3 GHz CPU with 3.25 GB RAM running on Linux OS. Convergence of the MCMC was evaluated using the R package `coda` (Plummer et al., 2006). We illustrate the performance of the MCMC on traceplots of last 1 000 sampled values (1/10 of the sample used for the inference) of the model deviance $D(\boldsymbol{\psi}, \boldsymbol{\theta}) = -2 \log\{L(\boldsymbol{\psi}, \boldsymbol{\theta})\}$ on the left panel of Figure C.1. Estimated autocorrelations for this quantity are shown on the right panel of Figure C.1.

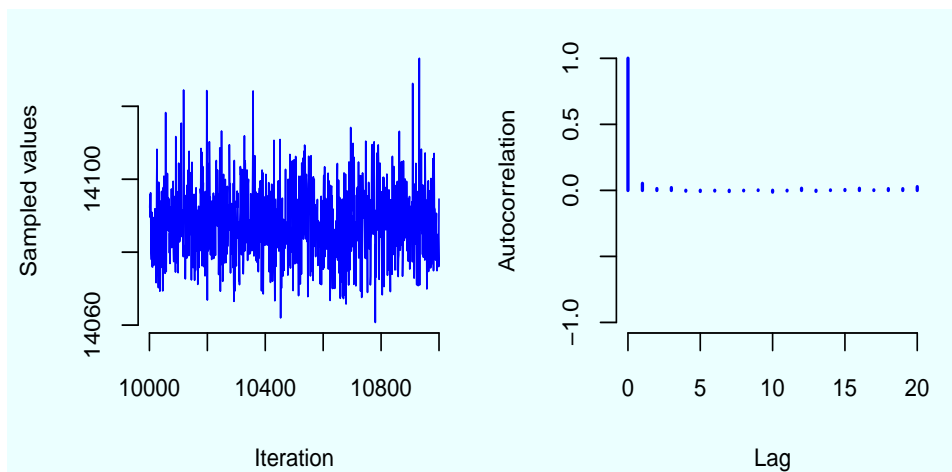


FIG C.1. *PBC910 Data.* Traceplot of last 1 000 sampled values of the observed data deviance $D(\boldsymbol{\psi}, \boldsymbol{\theta})$ and estimated autocorrelation based on 10 000 sampled values of the observed data deviance $D(\boldsymbol{\psi}, \boldsymbol{\theta})$.

D. Simulation study. In this section, we provide more details on the simulation study. Namely, Table D.1 shows the true values of the mixture covariance matrices in the distributions of random effects used to generate the data. Further, Figures D.1 and D.2 show observed longitudinal profiles for one simulated dataset in settings with $K = 2$, and $K = 3$, respectively, normally distributed random effects in each cluster and in total 200 subjects. Plots are also supplemented by the true cluster-specific marginal mean evolutions of each outcome. Supplementary results of the simulation study follow in Tables D.2–D.6.

TABLE D.1
Simulation study. Standard deviations (on a diagonal) and correlations (off-diagonal elements) for each mixture component derived from the true covariance matrices \mathbf{D}_1 , \mathbf{D}_2 , \mathbf{D}_3 .

	Intercept (Gaus.)	Slope (Gaus.)	Intercept (Pois.)	Slope (Pois.)	Intercept (Bernoulli)
$k = 1$					
Intercept (Gaus.)	0.5	0.0	-0.3	0.0	0.3
Slope (Gaus.)		0.01	0.0	-0.2	0.1
Intercept (Pois.)			0.3	0.0	0.0
Slope (Pois.)				0.003	0.0
Intercept (Bernoulli)					2.0
$k = 2$					
Intercept (Gaus.)	0.3	-0.2	0.1	-0.1	0.2
Slope (Gaus.)		0.01	0.0	0.3	0.1
Intercept (Pois.)			0.5	0.0	-0.2
Slope (Pois.)				0.005	0.0
Intercept (Bernoulli)					3.0
$k = 3$					
Intercept (Gaus.)	0.7	0.0	0.0	0.0	0.0
Slope (Gaus.)		0.03	0.0	0.0	0.0
Intercept (Pois.)			0.7	0.0	0.0
Slope (Pois.)				0.010	0.0
Intercept (Bernoulli)					0.5

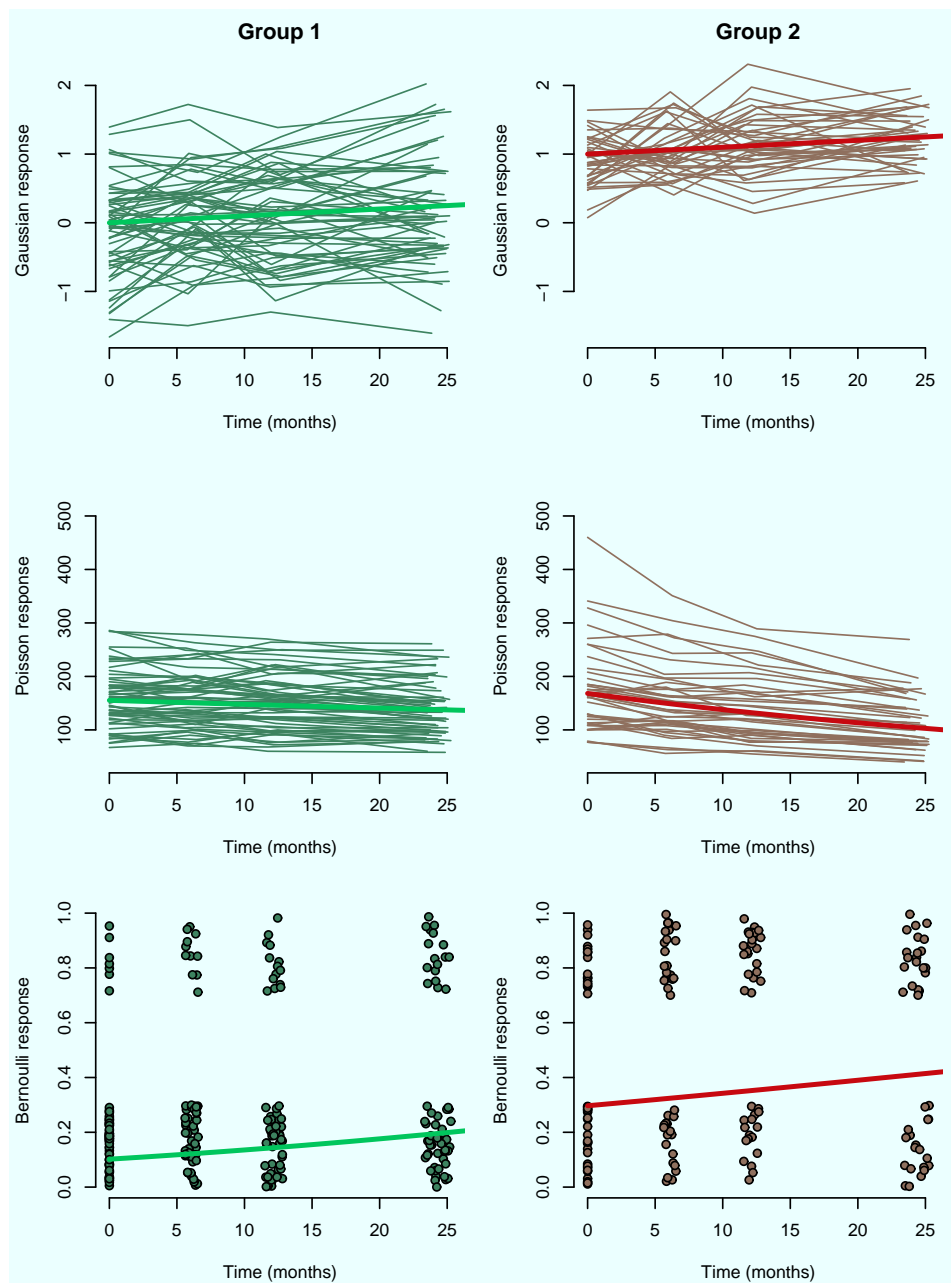


FIG D.1. *Simulation study.* One simulated dataset for a model with $K = 2$ ($N = (60, 40)$) and normally distributed random effects in each cluster. Thick lines show true cluster-specific marginal mean evolution. The values of a dichotomous Bernoulli response are vertically jittered.

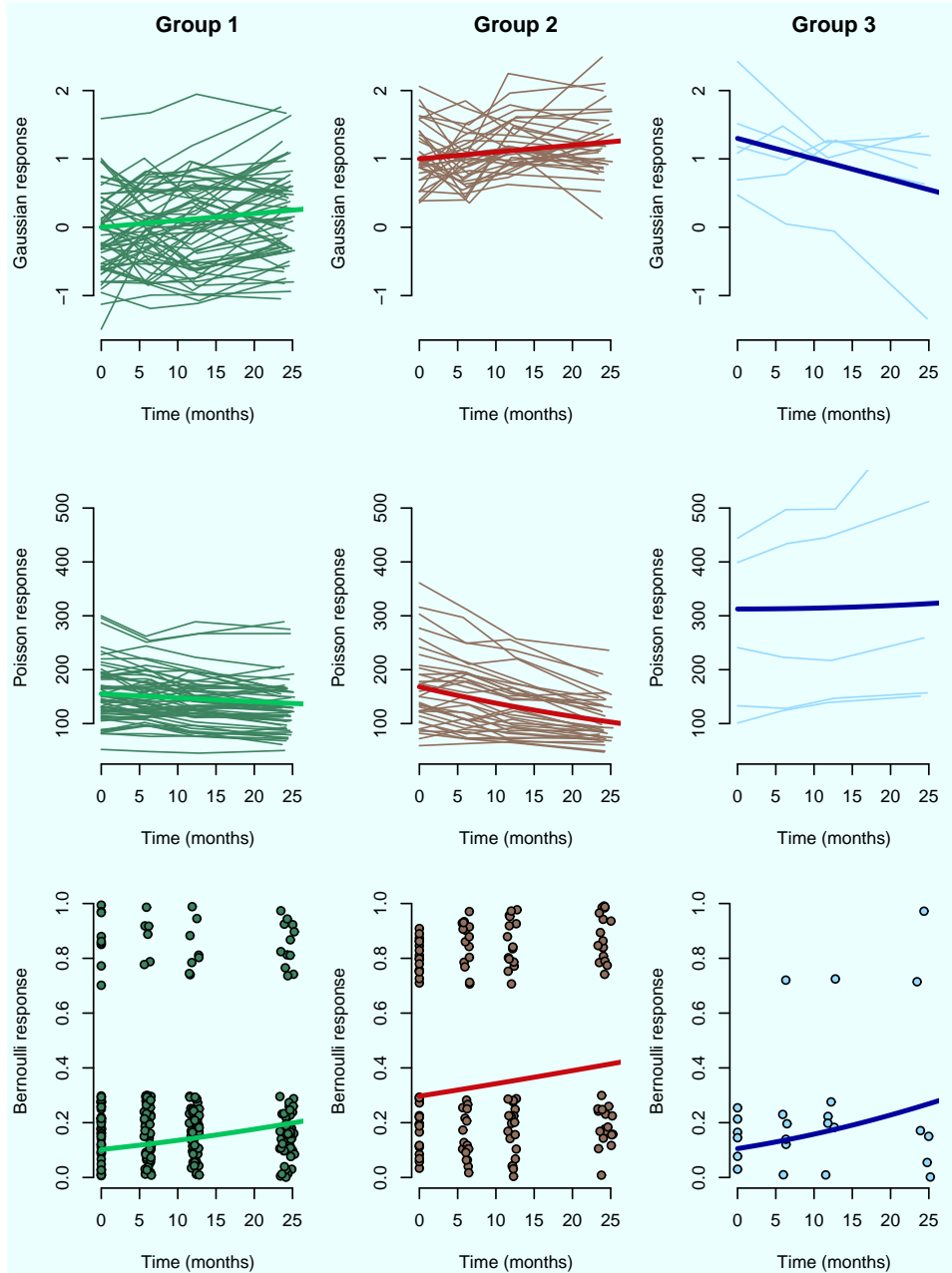


FIG D.2. *Simulation study. One simulated dataset for a model with $K = 3$ ($N = (60, 34, 6)$) and normally distributed random effects in each cluster. Thick lines show true cluster-specific marginal mean evolution. The values of a dichotomous Bernoulli response are vertically jittered.*

TABLE D.2

Simulation study: error rates of classification obtained from a model with correctly specified number of clusters.

Setting	Overall error rate (%)	Conditional error rate (%) given component		
		1	2	3
Normal ($K = 2$)				
$N = (30, 20)$	15.8	7.9	27.6	N.A.
(60, 40)	7.8	4.0	13.4	N.A.
(120, 80)	5.8	4.0	8.4	N.A.
MVT ₅ ($K = 2$)				
$N = (30, 20)$	20.6	10.9	35.2	N.A.
(60, 40)	10.8	3.5	21.9	N.A.
(120, 80)	8.7	4.3	15.4	N.A.
Normal ($K = 3$)				
$N = (30, 17, 3)$	26.5	13.9	41.9	66.0
(60, 34, 6)	17.4	11.9	20.0	56.8
(120, 68, 12)	10.1	6.4	10.8	42.3
MVT ₅ ($K = 3$)				
$N = (30, 17, 3)$	23.8	11.2	39.3	62.3
(60, 34, 6)	18.8	14.8	19.3	55.2
(120, 68, 12)	9.7	6.4	11.9	30.4

TABLE D.3

Simulation study: mixture weights. Bias, standard deviation and mean squared error (MSE) are based on posterior means as parameter estimates, the reported MSE is the average MSE over the K mixture components.

Setting	Bias	Std. Dev.	$\sqrt{\text{MSE}}$
$w = (0.600, 0.400)$			
Normal ($K = 2$)			
$N = (30, 20)$	(0.053, -0.053)	(0.161, 0.161)	0.169
(60, 40)	(0.030, -0.030)	(0.057, 0.057)	0.064
(120, 80)	(0.008, -0.008)	(0.037, 0.037)	0.037
MVT ₅ ($K = 2$)			
$N = (30, 20)$	(0.071, -0.071)	(0.198, 0.198)	0.209
(60, 40)	(0.064, -0.064)	(0.111, 0.111)	0.128
(120, 80)	(0.033, -0.033)	(0.097, 0.097)	0.102
$w = (0.600, 0.340, 0.060)$			
Normal ($K = 3$)			
$N = (30, 17, 3)$	(0.023, -0.051, 0.028)	(0.190, 0.127, 0.127)	0.154
(60, 34, 6)	(0.003, 0.003, -0.006)	(0.113, 0.090, 0.051)	0.088
(120, 68, 12)	(0.005, 0.003, -0.008)	(0.054, 0.056, 0.029)	0.048
MVT ₅ ($K = 3$)			
$N = (30, 17, 3)$	(0.045, -0.057, 0.012)	(0.133, 0.104, 0.070)	0.113
(60, 34, 6)	(-0.029, 0.008, 0.021)	(0.158, 0.106, 0.086)	0.122
(120, 68, 12)	(-0.007, -0.011, 0.017)	(0.065, 0.060, 0.034)	0.056

TABLE D.4

Simulation study. Mixture means of the random effects in the model for Gaussian response. Bias, standard deviation and mean squared error (MSE) are based on posterior means as parameter estimates, the reported MSE is the average MSE over the K mixture components.

Setting	Bias	Std. Dev.	$\sqrt{\text{MSE}}$
$\mu_{*,1} = (0.000, 1.000)$ (Intercept, Gaussian resp.)			
Normal ($K = 2$)			
$N = (30, 20)$	(0.140, -0.102)	(0.194, 0.213)	0.237
(60, 40)	(0.053, -0.020)	(0.128, 0.162)	0.150
(120, 80)	(0.014, 0.003)	(0.074, 0.055)	0.066
MVT ₅ ($K = 2$)			
$N = (30, 20)$	(0.174, -0.140)	(0.215, 0.738)	0.564
(60, 40)	(0.075, -0.041)	(0.145, 0.592)	0.433
(120, 80)	(0.038, -0.085)	(0.114, 0.516)	0.378
$\mu_{*,1} = (0.000, 1.000, 1.300)$ (Intercept, Gaussian resp.)			
Normal ($K = 3$)			
$N = (30, 17, 3)$	(0.184, -0.117, -0.427)	(0.181, 0.295, 0.495)	0.444
(60, 34, 6)	(0.124, -0.086, -0.124)	(0.161, 0.209, 0.533)	0.360
(120, 68, 12)	(0.030, -0.014, -0.040)	(0.084, 0.107, 0.428)	0.260
MVT ₅ ($K = 3$)			
$N = (30, 17, 3)$	(0.181, -0.107, -0.282)	(0.196, 0.336, 0.827)	0.563
(60, 34, 6)	(0.095, -0.100, -0.362)	(0.151, 0.201, 0.574)	0.424
(120, 68, 12)	(0.013, 0.017, -0.163)	(0.091, 0.086, 0.579)	0.353
$\mu_{*,2} = (0.0100, 0.0100)$ (Slope, Gaussian resp.)			
Normal ($K = 2$)			
$N = (30, 20)$	(0.0006, -0.0001)	(0.0036, 0.0058)	0.0048
(60, 40)	(-0.0001, -0.0001)	(0.0029, 0.0039)	0.0034
(120, 80)	(-0.0003, -0.0002)	(0.0018, 0.0026)	0.0022
MVT ₅ ($K = 2$)			
$N = (30, 20)$	(0.0000, 0.0023)	(0.0038, 0.0185)	0.0134
(60, 40)	(0.0006, 0.0000)	(0.0025, 0.0144)	0.0103
(120, 80)	(0.0002, -0.0006)	(0.0018, 0.0067)	0.0049
$\mu_{*,2} = (0.0100, 0.0100, -0.0300)$ (Slope, Gaussian resp.)			
Normal ($K = 3$)			
$N = (30, 17, 3)$	(-0.0005, -0.0056, 0.0200)	(0.0039, 0.0138, 0.0248)	0.0204
(60, 34, 6)	(-0.0008, -0.0015, 0.0034)	(0.0031, 0.0049, 0.0279)	0.0165
(120, 68, 12)	(-0.0001, -0.0008, -0.0028)	(0.0022, 0.0029, 0.0214)	0.0126
MVT ₅ ($K = 3$)			
$N = (30, 17, 3)$	(-0.0008, -0.0042, 0.0223)	(0.0041, 0.0173, 0.0294)	0.0237
(60, 34, 6)	(-0.0007, -0.0024, 0.0093)	(0.0029, 0.0068, 0.0347)	0.0211
(120, 68, 12)	(0.0002, -0.0016, 0.0051)	(0.0020, 0.0079, 0.0327)	0.0196

TABLE D.5

Simulation study: mixture means of the random effects in the model for Poisson response. Bias, standard deviation and mean squared error (MSE) are based on posterior means as parameter estimates, the reported MSE is the average MSE over the K mixture components.

Setting	Bias	Std. Dev.	$\sqrt{\text{MSE}}$
$\mu_{*,3} = (5.00, 5.00)$ (Intercept, Poisson resp.)			
Normal ($K = 2$)			
$N = (30, 20)$	(0.00, -0.03)	(0.06, 0.16)	0.12
(60, 40)	(-0.01, 0.00)	(0.05, 0.11)	0.09
(120, 80)	(0.00, 0.00)	(0.03, 0.06)	0.05
MVT ₅ ($K = 2$)			
$N = (30, 20)$	(0.00, -0.04)	(0.05, 0.51)	0.36
(60, 40)	(0.00, -0.01)	(0.04, 0.67)	0.47
(120, 80)	(0.00, -0.02)	(0.03, 0.22)	0.16
$\mu_{*,3} = (5.00, 5.00, 5.50)$ (Intercept, Poisson resp.)			
Normal ($K = 3$)			
$N = (30, 17, 3)$	(0.02, 0.10, -0.19)	(0.10, 0.32, 0.40)	0.33
(60, 34, 6)	(0.02, 0.03, -0.08)	(0.08, 0.15, 0.42)	0.26
(120, 68, 12)	(0.01, 0.01, 0.02)	(0.06, 0.09, 0.31)	0.19
MVT ₅ ($K = 3$)			
$N = (30, 17, 3)$	(0.01, 0.16, -0.21)	(0.08, 0.52, 0.70)	0.52
(60, 34, 6)	(0.02, 0.02, -0.12)	(0.06, 0.14, 0.51)	0.31
(120, 68, 12)	(0.00, 0.01, -0.14)	(0.03, 0.10, 0.35)	0.22
$\mu_{*,4} = (-0.0050, -0.0200)$ (Slope, Poisson resp.)			
Normal ($K = 2$)			
$N = (30, 20)$	(-0.0017, 0.0003)	(0.0025, 0.0046)	0.0039
(60, 40)	(-0.0006, -0.0005)	(0.0017, 0.0018)	0.0018
(120, 80)	(-0.0002, -0.0002)	(0.0006, 0.0012)	0.0010
MVT ₅ ($K = 2$)			
$N = (30, 20)$	(-0.0026, 0.0019)	(0.0029, 0.0058)	0.0051
(60, 40)	(-0.0011, 0.0007)	(0.0020, 0.0044)	0.0035
(120, 80)	(-0.0006, 0.0001)	(0.0016, 0.0022)	0.0020
$\mu_{*,4} = (-0.0050, -0.0200, 0.0000)$ (Slope, Poisson resp.)			
Normal ($K = 3$)			
$N = (30, 17, 3)$	(-0.0023, 0.0049, -0.0042)	(0.0021, 0.0057, 0.0089)	0.0073
(60, 34, 6)	(-0.0011, 0.0025, 0.0001)	(0.0020, 0.0042, 0.0078)	0.0054
(120, 68, 12)	(-0.0002, 0.0005, 0.0004)	(0.0007, 0.0022, 0.0055)	0.0034
MVT ₅ ($K = 3$)			
$N = (30, 17, 3)$	(-0.0023, 0.0049, -0.0039)	(0.0027, 0.0062, 0.0097)	0.0078
(60, 34, 6)	(-0.0010, 0.0027, -0.0038)	(0.0019, 0.0043, 0.0075)	0.0058
(120, 68, 12)	(-0.0002, 0.0010, -0.0021)	(0.0011, 0.0036, 0.0071)	0.0048

TABLE D.6

Simulation study: mixture means of the random effects in the model for Bernoulli response. Bias, standard deviation and mean squared error (MSE) are based on posterior means as parameter estimates, the reported MSE is the average MSE over the K mixture components.

Setting	Bias	Std. Dev.	$\sqrt{\text{MSE}}$
$\mu_{*,5} = (-3.00, -1.00)$ (Intercept, Bernoulli resp.)			
Normal ($K = 2$)			
$N = (30, 20)$	(-0.06, -0.27)	(1.15, 2.33)	1.84
(60, 40)	(-0.08, 0.02)	(0.61, 1.05)	0.86
(120, 80)	(-0.08, 0.00)	(0.42, 0.50)	0.46
MVT ₅ ($K = 2$)			
$N = (30, 20)$	(0.09, 0.35)	(0.87, 4.63)	3.33
(60, 40)	(-0.03, -0.18)	(0.63, 3.02)	2.17
(120, 80)	(-0.11, 0.26)	(0.45, 2.02)	1.47
$\mu_{*,5} = (-3.00, -1.00, -2.00)$ (Intercept, Bernoulli resp.)			
Normal ($K = 3$)			
$N = (30, 17, 3)$	(-0.01, -0.54, -3.91)	(1.03, 2.69, 7.28)	5.04
(60, 34, 6)	(0.09, -0.46, -2.74)	(0.75, 1.95, 4.06)	3.07
(120, 68, 12)	(-0.02, 0.10, -1.28)	(0.47, 0.55, 2.32)	1.58
MVT ₅ ($K = 3$)			
$N = (30, 17, 3)$	(0.08, -1.10, -2.05)	(0.86, 3.98, 5.12)	3.99
(60, 34, 6)	(0.02, -0.33, -1.83)	(0.83, 1.01, 4.04)	2.66
(120, 68, 12)	(0.01, 0.07, -0.76)	(0.36, 0.52, 2.53)	1.56

References.

- FONG, Y., RUE, H. and WAKEFIELD, J. (2010). Bayesian inference for generalized linear mixed models. *Biostatistics* **11** 397–412.
- GAMERMAN, D. (1997). Sampling from the posterior distribution in generalized linear mixed models. *Statistics and Computing* **7** 57–68.
- KOMÁREK, A. and KOMÁRKOVÁ, L. (2012). Clustering for multivariate continuous and discrete longitudinal data. *The Annals of Applied Statistics*.
- PLUMMER, M., BEST, N., COWLES, K. and VINES, K. (2006). CODA: Convergence Diagnosis and Output Analysis for MCMC. *R News* **6** 7–11.
- RICHARDSON, S. and GREEN, P. J. (1997). On Bayesian analysis of mixtures with unknown number of components (with Discussion). *Journal of the Royal Statistical Society, Series B* **59** 731–792.

A. KOMÁREK
FACULTY OF MATHEMATICS AND PHYSICS
CHARLES UNIVERSITY IN PRAGUE
SOKOLOVSKÁ 83
CZ–186 75, PRAHA 8
CZECH REPUBLIC
E-MAIL: arnost.komarek@mff.cuni.cz
URL: <http://www.karlin.mff.cuni.cz/~komarek>

L. KOMÁRKOVÁ
FACULTY OF MANAGEMENT
THE UNIVERSITY OF ECONOMICS IN PRAGUE
JAROŠOVSKÁ 1117
CZ–377 01, JINDŘICHŮV HRADEC
CZECH REPUBLIC
E-MAIL: komarkol@fm.vse.cz