Web Appendix to "Bayesian inference in semiparametric mixed models for longitudinal data"

by Yisheng Li, Xihong Lin and Peter Müller

A.1. Proof of Theorem 1

Lemma 1. The joint conditional USP $\pi(\boldsymbol{D}, \tau \mid \sigma^2)$ is proper.

Proof. By Theorem 2 of Natarajan and Kass (2000), $\pi(\boldsymbol{D} \mid \sigma^2)$ is proper. The conditional $\pi(\tau \mid \boldsymbol{D}, \sigma^2)$ is proper with a conditional density $\propto 1/(1+\bar{\lambda}\tau)^2$.

Proof of Theorem 1:

Let $m(\mathbf{Y}) = \int_{(0,\infty)} \int_P \int_{(0,\infty)} \int_{R^{\tilde{p}}} f(\mathbf{Y} \mid \boldsymbol{\beta}_\star, \boldsymbol{D}, \tau, \sigma^2)\pi(\boldsymbol{D}, \tau, \sigma^2) d\boldsymbol{\beta}_\star d\tau d\boldsymbol{D} d\sigma^2$, where $P = \{\boldsymbol{D} : \boldsymbol{D}$ is a $q \times q$ positive definite matrix$\}$. Without creating confusion we suppress the subscripts for the integration spaces below. Further let $\boldsymbol{\Sigma} = \mathbf{Z}\mathcal{C}\mathbf{Z}^T + \sigma^2\mathbf{I}$, where $\mathbf{Z} = (\mathbf{Z}^{(1)}, \mathbf{Z}^{(2)})$ and $\mathcal{C}_{\tilde{q}\times\tilde{q}} = \text{diag}(\boldsymbol{D}, \ldots, \boldsymbol{D}, \tau\mathbf{I}_{r-2})$ with $\tilde{q} = mq + r - 2$. Note this $\mathbf{Z}$ is not the one originally defined in Section 2.2. For convenience, we do not change the notation. Further let $\tilde{p} = p+2$.

To show $m(\mathbf{Y}) < \infty$, integrate out $\boldsymbol{\beta}_\star$ analytically. This yields

$$\int\int\int f(\mathbf{Y} \mid \boldsymbol{D}, \tau, \sigma^2)\pi(\boldsymbol{D}, \tau, \sigma^2) d\tau d\boldsymbol{D} d\sigma^2$$

$$\propto \int\int\int \frac{\exp\left[-\frac{1}{2}\left\{\mathbf{Y}^T\boldsymbol{\Sigma}^{-1}\mathbf{Y} - \mathbf{Y}^T\boldsymbol{\Sigma}^{-1}\mathbf{X}_\star(\mathbf{X}_\star^T\boldsymbol{\Sigma}^{-1}\mathbf{X}_\star)^{-1}\mathbf{X}_\star^T\boldsymbol{\Sigma}^{-1}\mathbf{Y}\right\}\right]}{(\sigma^2)^{(n-\tilde{q}-\tilde{p})/2}|\mathcal{C}|^{\frac{1}{2}}|\mathbf{X}_\star^T\mathbf{X}_\star|^{\frac{1}{2}}|\sigma^2\mathcal{C}^{-1} + \mathbf{Z}^T P_{X_\star}\mathbf{Z}|^{\frac{1}{2}}} \pi(\boldsymbol{D}, \tau, \sigma^2) d\tau d\boldsymbol{D} d\sigma^2$$

where $P_{X_\star} = \mathbf{I} - \mathbf{X}_\star(\mathbf{X}_\star^T\mathbf{X}_\star)^{-1}\mathbf{X}_\star^T$.

Denote $E(\mathcal{C}, \mathbf{Z}) \equiv \exp\left[-\frac{1}{2}\left\{\mathbf{Y}^T\boldsymbol{\Sigma}^{-1}\mathbf{Y} - \mathbf{Y}^T\boldsymbol{\Sigma}^{-1}\mathbf{X}_\star(\mathbf{X}_\star^T\boldsymbol{\Sigma}^{-1}\mathbf{X}_\star)^{-1}\mathbf{X}_\star^T\boldsymbol{\Sigma}^{-1}\mathbf{Y}\right\}\right]$. By the proof of Theorem 1 in Hobert and Casella (1996), if $\mathcal{C}$ is diagonal,

$$E(\mathcal{C}, \mathbf{Z}) \leq \exp\left[-\frac{1}{2\sigma^2}\mathbf{Y}^T\{P_Z - P_Z\mathbf{X}_\star(\mathbf{X}_\star^T P_Z\mathbf{X}_\star)^-\mathbf{X}_\star^T P_Z\}\mathbf{Y}\right], \tag{1}$$

where $P_Z = \mathbf{I} - \mathbf{Z}(\mathbf{Z}^T\mathbf{Z})^{-1}\mathbf{Z}^T$ and the notation $^-$ indicates the generalized inverse. Note since P_Z is idempotent, $r(P_Z) = tr(P_Z) = n - \tilde{q}$, where $r(\cdot)$ denotes the rank of a matrix.

We write $Q^T \mathcal{C} Q = \Lambda_{\mathcal{C}}$, where $Q = \mathrm{diag}(Q_0, \ldots, Q_0, \mathbf{I})$ with Q_0 being the orthogonal matrix that satisfies $Q_0^T \boldsymbol{D} Q_0 = \Lambda_D$, both $\Lambda_{\mathcal{C}}$ and Λ_D being diagonal. Replacing $\mathcal{C}$ by $\Lambda_{\mathcal{C}}$ and $\mathbf{Z}$ by $\tilde{\mathbf{Z}} = \mathbf{Z}Q$ in (1), we obtain

$$
\begin{aligned}
E(\mathcal{C}, \mathbf{Z}) = E(\Lambda_{\mathcal{C}}, \tilde{\mathbf{Z}}) &\leq \exp\left[-\frac{1}{2\sigma^2}\mathbf{Y}^T\{P_{\tilde{Z}} - P_{\tilde{Z}}\mathbf{X}_\star(\mathbf{X}_\star^T P_{\tilde{Z}}\mathbf{X}_\star)^{-}\mathbf{X}_\star^T P_{\tilde{Z}}\}\mathbf{Y}\right] \\
&= \exp\left[-\frac{1}{2\sigma^2}\mathbf{Y}^T\{P_Z - P_Z\mathbf{X}_\star(\mathbf{X}_\star^T P_Z\mathbf{X}_\star)^{-}\mathbf{X}_\star^T P_Z\}\mathbf{Y}\right].
\end{aligned}
$$

The last equality is based on the fact that $P_{\tilde{Z}} = \mathbf{I} - \tilde{\mathbf{Z}}(\tilde{\mathbf{Z}}^T\tilde{\mathbf{Z}})^{-1}\tilde{\mathbf{Z}}^T = \mathbf{I} - \mathbf{Z}QQ^T(\mathbf{Z}^T\mathbf{Z})^{-1}QQ^T\mathbf{Z}^T = P_Z$. In other words, (1) holds for non-diagonal $\mathcal{C}$ as well. Therefore, we have

$$
m(\mathbf{Y}) \leq \int\!\!\int\!\!\int \frac{\exp\left[-\frac{1}{2\sigma^2}\mathbf{Y}^T\{P_Z - P_Z\mathbf{X}_\star(\mathbf{X}_\star^T P_Z\mathbf{X}_\star)^{-}\mathbf{X}_\star^T P_Z\}\mathbf{Y}\right]}{(\sigma^2)^{(n-\tilde{q}-\tilde{p})/2}|\mathbf{X}_\star^T\mathbf{X}_\star|^{\frac{1}{2}}|\mathcal{C}|^{\frac{1}{2}}|\sigma^2\mathcal{C}^{-1} + \mathbf{Z}^T P_{X_\star}\mathbf{Z}|^{\frac{1}{2}}} \pi(\tau, \boldsymbol{D}, \sigma^2) d\tau d\boldsymbol{D} d\sigma^2, \tag{2}
$$

up to a constant.

Based on Lemma 2 of Hobert and Casella (1996), or Marshall and Olkin (1979), we have $|\sigma^2\mathcal{C}^{-1} + \mathbf{Z}^T P_{X_\star}\mathbf{Z}| \geq \prod_{i=1}^{\tilde{q}}\{\lambda_i(\sigma^2\mathcal{C}^{-1}) + \lambda_i(\mathbf{Z}^T P_{X_\star}\mathbf{Z})\} \geq (\sigma^2)^{\tilde{q}}/(\tau^{r-2}\prod_{i=1}^{mq}\delta_i)$. Therefore $|\mathcal{C}|^{\frac{1}{2}}|\sigma^2\mathcal{C}^{-1} + \mathbf{Z}^T P_{X_\star}\mathbf{Z}|^{\frac{1}{2}} \geq (\sigma^2)^{\tilde{q}/2}$. Since $\int \pi(\boldsymbol{D}, \tau \mid \sigma^2) d\tau d\boldsymbol{D} = 1$, $\pi(\sigma^2) \propto 1/\sigma^2$, and $n > p + mq + r \geq p + 2 = \tilde{p}$, it is easily seen that the integral with respect to τ and $\boldsymbol{D}$ in (2) is bounded by an inverse gamma denstiy of σ^2. Therefore the integral is finite. This is based on the fact that $\mathbf{Y}^T\{P_Z - P_Z\mathbf{X}_\star(\mathbf{X}_\star^T P_Z\mathbf{X}_\star)^{-}\mathbf{X}_\star^T P_Z\}\mathbf{Y} > 0$, a.s., since $P_Z - P_Z\mathbf{X}_\star(\mathbf{X}_\star^T P_Z\mathbf{X}_\star)^{-}\mathbf{X}_\star^T P_Z$ is idempotent with its rank $\geq n - \tilde{q} - \tilde{p} = n - (p + mq + r) \geq 1$.

We note that alternative proofs of the theorem are possible. For example, one can follow the derivation of posterior propriety in two-stage hierarchical LMMs in Chen, Shao and Xu (2002), where the prior for σ^2 is proportional to $1/\sigma^2$ and an inverse Wishart prior is assumed for the random effect covariance matrix. Another approach is to extend Theorem 1 in Sun, Tsutakawa and He (2001) by allowing unstructured random effect covariance matrix, conditionally (on σ^2) proper prior for the random effect covariance matrix and smoothing parameter, and a prior proportional to $1/\sigma^2$ for the residual variance.

A.2. Special cases of the proposed USP

Consider the following normal-normal model considered by Daniels (1999):

$$Y_i|\theta_i \sim N(\theta_i, \sigma_i),\ \ \theta_i|\tau \sim N(0, \tau),\ \ i = 1, \ldots, n. \tag{3}$$

He defined a USP $\pi(\tau) = \sigma_c/(\sigma_c + \tau)^2$, where σ_c is the harmonic mean of σ_i, i.e., $1/\sigma_c = 1/n \sum_{i=1}^{n} 1/\sigma_i$.

Result 1. Under model (3), a special case of $(10^{\star})$[1] in the article, our proposed USP for τ reduces to $\pi(\tau)$.

Proof. It is easily seen that $\mathbf{Z}^T \boldsymbol{R}^{-1}\mathbf{Z} = diag(1/\sigma_i)$. Therefore, $\left(1/\tau + \bar{\lambda}\right)^{-1} 1/\tau = (1/\tau + 1/\sigma_c)^{-1}$ $\cdot 1/\tau = \sigma_c/(\sigma_c + \tau)$. A uniform prior on this then leads to $\pi(\tau) = \sigma_c/(\sigma_c + \tau)^2$.

One can similarly show that when the data are repeated measures associated with a diagonal $\mathbf{Z}^T \boldsymbol{R}^{-1}\mathbf{Z}$, the proposed USP reduces to the corresponding USP defined by Natarajan and Kass (2000) for normal outcomes.

A.3. Proof of Theorem 2

Conditional on each configuration of ties in the Polya urn model the probability model on $\mathbf{Z}^{(1)}\boldsymbol{b} + \mathbf{Z}^{(2)}\boldsymbol{a}$ is a multivariate normal distribution with a non-block-diagonal variance-covariance matrix (i.e., dependent across clusters). By Theorem 1 the posterior conditional on the configuration is proper. The claim then follows following the argument in Proposition 1 of Li, Müller and Lin (2007).

[1] An equation number with a superscript $\star$ refers to the equation with the same number in the article body. One without a superscript $\star$ is instead defined within the scope of this web Appendix. Same below.

A.4. Proof of Theorem 3

(i) ($13^\star$) follows immediately from ($5^\star$) and Proposition 2 (i) of Li *et al.* (2007).

(ii) Let $\Psi = \begin{pmatrix} \boldsymbol{\delta} \\ \boldsymbol{a} \end{pmatrix}$. By the proof of Proposition 2 of Li *et al.* (2007), we have

$$E\left(\Psi \middle| \boldsymbol{c}, \boldsymbol{\beta}_{\mathbf{c}}, \boldsymbol{D}, \boldsymbol{a}, M\right) = \begin{pmatrix} \boldsymbol{\mu}_{G_\star,\mathbf{c}} \\ \boldsymbol{a} \end{pmatrix} \text{ and } Cov\left(\Psi \middle| \boldsymbol{c}, \boldsymbol{\beta}_{\mathbf{c}}, \boldsymbol{D}, \boldsymbol{a}, M\right) = \begin{pmatrix} \frac{Cov_{G_\star,\mathbf{c}}}{m+M+1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}.$$

($14^\star$) follows by writing

$$Cov(\mathbf{f} \mid \mathbf{Y}) = Cov\left((\mathbf{T}\ \mathbf{B})\,\Psi \middle| \mathbf{Y}\right) = (\mathbf{T}\ \mathbf{B})\, Cov\left(\Psi \middle| \mathbf{Y}\right) \begin{pmatrix} \mathbf{T}^T \\ \mathbf{B}^T \end{pmatrix},$$

and

$$Cov\left(\Psi \middle| \mathbf{Y}\right) = Cov\left(E\left(\Psi \middle| \boldsymbol{c}, \boldsymbol{\beta}_{\mathbf{c}}, \boldsymbol{D}, \boldsymbol{a}, M\right) \middle| \mathbf{Y}\right) + E\left(Cov\left(\Psi \middle| \boldsymbol{c}, \boldsymbol{\beta}_{\mathbf{c}}, \boldsymbol{D}, \boldsymbol{a}, M\right) \middle| \mathbf{Y}\right).$$

A.5 Gibbs sampling in the LMM ($6^\star$) *with normal random effects*

Under the priors specified in Sections 2 and 3 of the article, the full conditional distributions for $\boldsymbol{\beta}_\star$, $\boldsymbol{b}$, $\boldsymbol{a}$, σ^2, $\boldsymbol{D}$ and τ are as follows:

$$[\boldsymbol{\beta}_\star \mid \boldsymbol{b}, \boldsymbol{a}, \sigma^2, \mathbf{Y}] \sim N\left(\mathbf{T}_1 \mathbf{X}_\star^T (\mathbf{Y} - \mathbf{Z}^{(1)}\boldsymbol{b} - \mathbf{Z}^{(2)}\boldsymbol{a})/\sigma^2, \mathbf{T}_1\right),$$

$$[\boldsymbol{b} \mid \boldsymbol{\beta}_\star, \boldsymbol{a}, \boldsymbol{D}, \sigma^2, \mathbf{Y}] \sim N\left(\mathbf{T}_2 {\mathbf{Z}^{(1)}}^T (\mathbf{Y} - \mathbf{X}_\star\boldsymbol{\beta}_\star - \mathbf{Z}^{(2)}\boldsymbol{a})/\sigma^2, \mathbf{T}_2\right),$$

$$[\boldsymbol{a} \mid \boldsymbol{\beta}_\star, \boldsymbol{b}, \sigma^2, \tau, \mathbf{Y}] \sim N\left(\mathbf{T}_3 {\mathbf{Z}^{(2)}}^T (\mathbf{Y} - \mathbf{X}_\star\boldsymbol{\beta}_\star - \mathbf{Z}^{(1)}\boldsymbol{b})/\sigma^2, \mathbf{T}_3\right),$$

where $\mathbf{T}_1 = \left(\mathbf{X}_\star^T\mathbf{X}_\star/\sigma^2 + \mathbf{T}_0^{-1}\right)^{-1}$, $\mathbf{T}_2 = \left(\mathcal{D}^{-1} + \frac{{\mathbf{Z}^{(1)}}^T\mathbf{Z}^{(1)}}{\sigma^2}\right)^{-1}$, $\mathbf{T}_3 = \left(\frac{\mathbf{I}}{\tau} + \frac{{\mathbf{Z}^{(2)}}^T\mathbf{Z}^{(2)}}{\sigma^2}\right)^{-1}$; and

$$[\sigma^2 \mid \tau, \boldsymbol{D}, \boldsymbol{\beta}_\star, \boldsymbol{b}, \boldsymbol{a}, \mathbf{Y}] \propto \pi(\boldsymbol{D} \mid \sigma^2)\pi(\tau \mid \sigma^2, \mathbf{D}) IG\left(\alpha + n/2, \nu + \mathbf{V}^T\mathbf{V}/2\right),$$

$$[\boldsymbol{D} \mid \sigma^2, \tau, \boldsymbol{\beta}_\star, \boldsymbol{b}, \boldsymbol{a}, \mathbf{Y}] \propto \pi(\boldsymbol{D} \mid \sigma^2)\pi(\tau \mid \sigma^2, \boldsymbol{D}) IW\left(m, \left(\sum_{i=1}^{m} \boldsymbol{b}_i \boldsymbol{b}_i^T\right)^{-1}\right),$$

$$[\tau \mid \sigma^2, \boldsymbol{D}, \boldsymbol{\beta}_\star, \boldsymbol{b}, \boldsymbol{a}, \mathbf{Y}] \propto \pi(\tau \mid \sigma^2, \boldsymbol{D}) IG\left((r-2)/2 + 2, \boldsymbol{a}^T\boldsymbol{a}/2\right),$$

where $\mathbf{V} = \mathbf{Y} - \mathbf{X}\boldsymbol{\beta}_\star - \mathbf{Z}^{(1)}\boldsymbol{b} - \mathbf{Z}^{(2)}\boldsymbol{a}$. To sample from the last three full conditionals, we implement a Metropolis-Hastings transition probability using as proposal distributions an $IG\left(\alpha + n/2, \nu + \mathbf{V}^T\mathbf{V}/2\right)$, $IW\left(m, \left(\sum_{i=1}^{m} \boldsymbol{b}_i \boldsymbol{b}_i^T\right)^{-1}\right)$ and $IG\left((r-2)/2 + 2, \boldsymbol{a}^T\boldsymbol{a}/2\right)$ distribution, respectively. Similar proposal densities have been used by Natarajan and Kass (2000).

We briefly summarize the steps of the implementation of our inference procedure below:

1. Write the LMM formulation of the SPMM equation ($6^\star$), in particular, identify the matrices $\mathbf{X}_\star$, $\mathbf{Z}^{(1)}$ and $\mathbf{Z}^{(2)}$.

2. Choose starting values for all model parameters and start Gibbs sampling as described above.

3. Draw posterior inference based on the draws of posterior model parameters, in particular, draw inference for $f(t)$ based on Equation ($5^\star$).

A.6. Gibbs sampling in the LMM ($12^\star$) *with a hierarchically centered DP prior*

Denote by $\boldsymbol{b}_{-i} = (\boldsymbol{b}_j;\ j \neq i)$ the random effects vector without subject i. Let $\mathbf{V}_i = \mathbf{Y}_i - \tilde{\mathbf{X}}_i\tilde{\boldsymbol{\beta}} - \mathbf{Z}_i^{(2)}\boldsymbol{a}$. The full conditional distribution of $\boldsymbol{b}_i$ is

$$[\boldsymbol{b}_i \mid \boldsymbol{\beta}_{\mathbf{b}}, \tilde{\boldsymbol{\beta}}, \boldsymbol{b}_{-i}, \boldsymbol{a}, \sigma^2, \mathbf{Y}] \propto \sum_{j \neq i} q_j \delta_{\boldsymbol{b}_j} + M q_0 p(\boldsymbol{b}_i \mid \boldsymbol{\beta}_{\mathbf{b}}, \tilde{\boldsymbol{\beta}}, \boldsymbol{a}, \sigma^2, \boldsymbol{D}, \mathbf{Y}_i),$$

where $q_j = \exp\left\{-\frac{1}{2\sigma^2}\left(\mathbf{V}_i - \mathbf{Z}_i^{(1)}\boldsymbol{b}_j\right)^T\left(\mathbf{V}_i - \mathbf{Z}_i^{(1)}\boldsymbol{b}_j\right)\right\}$, $j \neq i$,

$$
\begin{aligned}
q_0 &= |\mathbf{Q}_i|^{1/2}|\boldsymbol{D}|^{-1/2} \\
&\times \exp\left[\frac{1}{2}\left\{\left(\boldsymbol{D}^{-1}\boldsymbol{\beta}_{\mathbf{b}} + \frac{\mathbf{Z}_i^{(1)^T}\mathbf{V}_i}{\sigma^2}\right)^T \mathbf{Q}_i\left(\boldsymbol{D}^{-1}\boldsymbol{\beta}_{\mathbf{b}} + \frac{\mathbf{Z}_i^{(1)^T}\mathbf{V}_i}{\sigma^2}\right) - \left(\boldsymbol{\beta}_{\mathbf{b}}^T\mathbf{D}^{-1}\boldsymbol{\beta}_{\mathbf{b}} + \frac{\mathbf{V}_i^T\mathbf{V}_i}{\sigma^2}\right)\right\}\right],
\end{aligned}
$$

and

$$p(\boldsymbol{b}_i \mid \boldsymbol{\beta}_{\mathbf{b}}, \tilde{\boldsymbol{\beta}}, \boldsymbol{a}, \sigma^2, \boldsymbol{D}, \mathbf{Y}_i) \propto \phi(\boldsymbol{b}_i \mid \boldsymbol{\beta}_{\mathbf{b}}, \boldsymbol{D}) p(\mathbf{Y}_i \mid \boldsymbol{b}_i, \tilde{\boldsymbol{\beta}}, \boldsymbol{a}, \sigma^2) \propto \phi(\boldsymbol{b}_i \mid \boldsymbol{\eta}_i, \mathbf{Q}_i)$$

with moments $\mathbf{Q}_i = \left(\boldsymbol{D}^{-1} + \frac{\mathbf{Z}_i^{(1)^T}\mathbf{Z}_i^{(1)}}{\sigma^2}\right)^{-1}$ and $\boldsymbol{\eta}_i = \mathbf{Q}_i\left\{\boldsymbol{D}^{-1}\boldsymbol{\beta}_{\mathbf{b}} + \frac{\mathbf{Z}_i^{(1)^T}\mathbf{V}_i}{\sigma^2}\right\}$. This full conditional distribution is a mixture of a normal distribution and $m-1$ point masses (some point masses may overlap).

After choosing random effects for each subject, we update the list of unique random effects, $\{\boldsymbol{\gamma}_j,\ j = 1, \ldots, k\}$ and k, the number of unique values.

To improve mixing over the entire parameter space we update the $\boldsymbol{\gamma}$'s conditional on the configuration s. Recall that $s_i = j$ if $\boldsymbol{\gamma}_j = \boldsymbol{b}_i$. The conditional density of $\boldsymbol{\gamma}_l$ is

$$[\boldsymbol{\gamma}_l \mid \boldsymbol{\beta}_{\mathbf{b}}, \tilde{\boldsymbol{\beta}}, \boldsymbol{b}, \boldsymbol{a}, \sigma^2, \boldsymbol{D}, \mathbf{Y}] \sim N\left(\mathbf{Q}_l\left\{\boldsymbol{D}^{-1}\boldsymbol{\beta}_{\mathbf{b}} + \sum_{i \in l}\mathbf{Z}_i^{(1)^T}\mathbf{V}_i/\sigma^2\right\}, \mathbf{Q}_l\right) \equiv \phi(\boldsymbol{\gamma}_l; \boldsymbol{\mu}_l, \boldsymbol{\Sigma}_l), \quad (4)$$

where $\mathbf{Q}_l = \left(\boldsymbol{D}^{-1} + \sum_{i \in l}\mathbf{Z}_i^{(1)^T}\mathbf{Z}_i^{(1)}/\sigma^2\right)^{-1}$.

Since the $\boldsymbol{\gamma}_l$'s are considered random samples from $N(\boldsymbol{\beta}_{\mathbf{b}}, \boldsymbol{D})$, $\boldsymbol{\beta}_{\mathbf{b}}$ and $\boldsymbol{D}$ can in turn be sampled as follows:

$$[\boldsymbol{\beta}_{\mathbf{b}} \mid \boldsymbol{\gamma}_1, \ldots, \boldsymbol{\gamma}_k, \boldsymbol{D}] \sim N\left(\left(k\boldsymbol{D}^{-1} + \boldsymbol{\Sigma}_0^{-1}\right)^{-1}\boldsymbol{D}^{-1}\sum_{l=1}^{k}\boldsymbol{\gamma}_l, \left(k\boldsymbol{D}^{-1} + \boldsymbol{\Sigma}_0^{-1}\right)^{-1}\right),$$

and

$$[\boldsymbol{D} \mid \boldsymbol{\gamma}_1, \ldots, \boldsymbol{\gamma}_k, \boldsymbol{\beta}_{\mathbf{b}}] \sim \pi(\boldsymbol{D} \mid \sigma^2)\pi(\tau \mid \sigma^2, \boldsymbol{D}) IW\left(k, \left\{\sum_{l=1}^{k}(\boldsymbol{\gamma}_l - \boldsymbol{\beta}_{\mathbf{b}})(\boldsymbol{\gamma}_l - \boldsymbol{\beta}_{\mathbf{b}})^T\right\}^{-1}\right). \quad (5)$$

Note that when $k < q$, an unlikely but not impossible case when M is small, an inverse Wishart distribution $IW\left(k, \left\{\sum_{l=1}^{k}(\boldsymbol{\gamma}_l - \boldsymbol{\beta}_{\mathbf{b}})(\boldsymbol{\gamma}_l - \boldsymbol{\beta}_{\mathbf{b}})^T\right\}^{-1}\right)$ is not well-defined because

$\sum_{l=1}^{k}(\boldsymbol{\gamma}_l - \boldsymbol{\beta}_{\mathbf{b}})(\boldsymbol{\gamma}_l - \boldsymbol{\beta}_{\mathbf{b}})^T$ is not positive-definite. In this case, the notation $IW\left(k, \left\{\sum_{l=1}^{k}(\boldsymbol{\gamma}_l - \boldsymbol{\beta}_{\mathbf{b}})(\boldsymbol{\gamma}_l - \boldsymbol{\beta}_{\mathbf{b}})^T\right\}^{-1}\right)$ in (5) reduces to a degenerate form, which is simply proportional to the joint normal density of the k independent $\boldsymbol{\gamma}_l$'s.

To sample $\boldsymbol{D}$ from (5), we again use the Metropolis-Hastings algorithm with a proposal density $IW\left(k+q, \left\{\mathbf{I}_q + \sum_{l=1}^{k}(\boldsymbol{\gamma}_l - \boldsymbol{\beta}_{\mathbf{b}})(\boldsymbol{\gamma}_l - \boldsymbol{\beta}_{\mathbf{b}})^T\right\}^{-1}\right)$ to ensure both the propriety of its inverse, a Wishart distribution (see, e.g., Gelman *et al.*, 2000, pp. 481), and the positive-definiteness of its scale matrix.

Sampling of the other parameters in the model remains the same as that given in Appendix A.5. If the priors for the variance components are taken to be inverse gamma or inverse Wishart, as used for comparison with the USP, the corresponding full conditionals can be derived straightforwardly using conjugacy.

Learning about M

West (1992) proposed a gamma hyperprior for modeling M. This can easily be included in the Gibbs sampling scheme. See West (1992) for details.

We again briefly summarize the steps of the implementation of our inference procedure below:

1. Write the LMM formulation of the DP SPMM equation ($12^\star$), in particular, identify the matrices $\tilde{\mathbf{X}}$, $\mathbf{Z}^{(1)}$ and $\mathbf{Z}^{(2)}$.

2. Choose starting values for all model parameters and start Gibbs sampling as described above.

3. Draw posterior inference based on the draws of posterior model parameters, in particular, draw inference for $f(t)$ based on Theorem 3 in the paper.

A.7. Some detail about the data example and simulation studies

Data example

For implementation, after a burn-in of 1,000 samples, the Gibbs sampler was run for 100,000 iterations with every 20th sample collected for posterior inference. We assessed convergence of the Markov chains by the method of Geweke (1992). For most model parameters, we evaluated Geweke statistics within ± 1.96, indicating no evidence against practical convergence of the Markov chains.

Simulation studies

Table 2 and Figure 2 in the paper include results based on the following simulation truth:

$$Y_{ij} = \beta_1 X_{1i} + \beta_2 X_{2i} + f(t_{ij}) + b_i + \epsilon_{ij},\ i = 1, \ldots, m,\ j = 1, \ldots, n_i, \tag{6}$$

where $m = 34$, $b_i \overset{i.i.d.}{\sim} N(0, 0.2303)$ and $f(t)$ is a smooth function of t, chosen to match the estimates under the progesterone data and X_1 and X_2 correspond to AGE and BMI in the progestrone data. Throughout, the design is identical to that of the progesterone study unless otherwise noted. Let $Var(\epsilon_{ij}) = \sigma^2$ and $Var(b_i) = \theta$. The true values of the parameters that are used throughout all simulations, when applicable, are: $\overline{f(t^0)} \equiv \sum_{i=1}^{r} f(t_i^0)/r = 0.9737$, $\beta_1 = 2.8743$, $\beta_2 = -4.7168$ and $\sigma^2 = 0.3396$, all chosen to match the estimates from the progesterone data. Alternatively, the following bimodal normal mixture distribution for the random intercepts was also considered:

$$b_i \sim \frac{11}{18} N(-0.35, 0.03) + \frac{7}{18} N(0.55, 0.05).$$

Table 3 in the paper includes simulation results based on a slightly different model as follows:

$$Y_{ij} = f(t_{ij}) + b_i + \epsilon_{ij},\ i = 1, \ldots, m,\ j = 1, \ldots, n_i,$$

i.e., model (6) with X_1 (age) and X_2 (BMI) excluded.

Some graphical results (Figures 1 and 2) on the inference for the nonparametric function are given at the end of this web Appendix.

A.8. Estimating the random effect distribution in the DP SPMM (12*)

The almost sure discrete nature of G implies a positive probability for ties among the random effects $\boldsymbol{b}_i$. Let $\boldsymbol{\gamma} = \left(\boldsymbol{\gamma}_1^T, \cdots, \boldsymbol{\gamma}_k^T\right)^T$ denote a vector of distinct random effects $\boldsymbol{b}_i$, and $\mathbf{s} = (s_1, \ldots, s_m)$ denote a vector of indicators defined by $s_i = j$ if $\boldsymbol{b}_i = \boldsymbol{\gamma}_j$. The pair $(\boldsymbol{\gamma}, \mathbf{s})$ is a reparametrization of $\boldsymbol{b}$. Let $g(\cdot)$ denote the density of $E(G \mid Y)$, the posterior mean of G. To evaluate g, we note that $E(G \mid Y) = p(\boldsymbol{b}_{m+1} \mid Y)$, i.e., the posterior mean of G is identical to the posterior predictive distribution of a future random effect. Let $g_c(b_{m+1}) = p(\boldsymbol{b}_{m+1} \mid \tilde{\boldsymbol{\beta}}, \boldsymbol{\beta}_{\mathbf{b}}, \mathbf{s}, \boldsymbol{D}, M)$ denote the conditional posterior predictive distribution for $\boldsymbol{b}_{m+1}$. We find

$$\begin{aligned} g_c(\boldsymbol{b}_{m+1}) = \; & \int \left\{ r\phi(\boldsymbol{b}_{m+1}; \boldsymbol{\beta}_{\mathbf{b}}, \boldsymbol{D}) + (1-r) \textstyle\sum_{l=1}^{k} h_l \delta_{\boldsymbol{\gamma}_l} \right\} \\ & d[\boldsymbol{\gamma}_1 \mid \mathbf{s}, \tilde{\boldsymbol{\beta}}, \boldsymbol{\beta}_{\mathbf{b}}, \boldsymbol{a}, \sigma^2, \boldsymbol{D}, \mathbf{Y}] \ldots d[\boldsymbol{\gamma}_k \mid \mathbf{s}, \tilde{\boldsymbol{\beta}}, \boldsymbol{\beta}_{\mathbf{b}}, \boldsymbol{a}, \sigma^2, \boldsymbol{D}, \mathbf{Y}], \end{aligned} \tag{7}$$

where $\phi(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma})$ denotes the density of $\mathbf{x} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ and h_l is the number of the $\boldsymbol{b}_i$'s that are equal to $\boldsymbol{\gamma}_l$ satisfying $\sum_{l=1}^{k} h_l = m$. Denote the density of each $[\boldsymbol{\gamma}_l \mid \mathbf{s}, \tilde{\boldsymbol{\beta}}, \boldsymbol{\beta}_{\mathbf{b}}, \boldsymbol{a}, \sigma^2, \boldsymbol{D}, \mathbf{Y}]$ as $\phi(\boldsymbol{\gamma}_l; \boldsymbol{\mu}_l, \boldsymbol{\Sigma}_l)$ (see (4)). From (7) we find

$$g_c(\boldsymbol{b}_{m+1}) = r\phi(\boldsymbol{b}_{m+1}; \boldsymbol{\beta}_{\mathbf{b}}, \boldsymbol{D}) + (1-r) \sum_{l=1}^{k} h_l \phi(\boldsymbol{b}_{m+1}; \boldsymbol{\mu}_l, \boldsymbol{\Sigma}_l).$$

Finally, the posterior predictive $g(\boldsymbol{b}_{m+1})$, and thus the posterior mean $E(G \mid Y)$, is evaluated as an average of the densities of the normal mixtures g_c across the posterior samples of $(\tilde{\boldsymbol{\beta}}, \boldsymbol{\beta}_{\mathbf{b}}, \mathbf{s}, \boldsymbol{D}, M)$.

References

Chen, M.-H., Shao, Q.-M. and Xu, D. (2002), Necessary and sufficient conditions on the propriety of posterior distributions for generalized linear mixed models, *Sankhyā A* **64**, 57–85.

Daniels, M. (1999), A prior for the variance in hierarchical models, *Canadian Journal of Statistics* **27**, 567–578.

Gelman, A., Carlin, J. B., Stern, H. S. and Rubin, D. B. (2000), *Bayesian Data Analysis*, Chapman and Hall, London.

Geweke, J. (1992), Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments, *in* J. M. Bernardo, J. Berger, A. P. Dawid, and A. F. M. Smith, eds, Bayesian Statistics 4, Oxford University Press, Oxford, U.K., pp. 169–193.

Hobert, J. P. and Casella, G. (1996), The effect of improper priors on Gibbs sampling in hierarchical linear mixed models, *Journal of the American Statistical Association* **91**, 1461–1473.

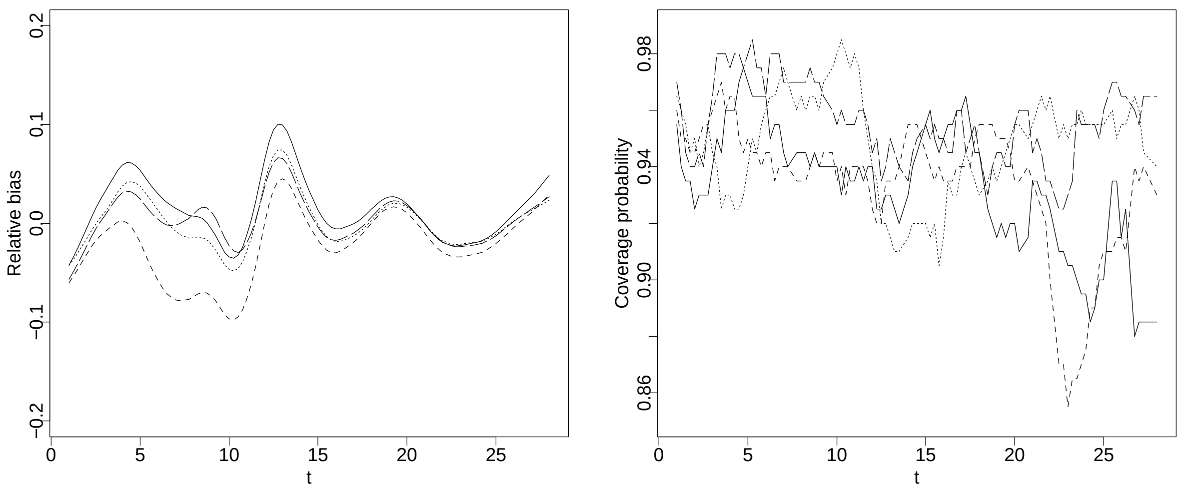
Li, Y., Müller, P. and Lin, X. (2007), Center-adjusted inference for a nonparametric Bayesian random effect distribution, Technical report, University of Texas M. D. Anderson Cancer Center.

Marshall, A. W. and Olkin, I. (1979), *Inequalities: Theories of Majorization and Its Applications*, Academic Press, New York.

Natarajan, R. and Kass, R. E. (2000), Reference Bayesian methods for generalized linear mixed models, *Journal of the American Statistical Association* **95**, 227–237.

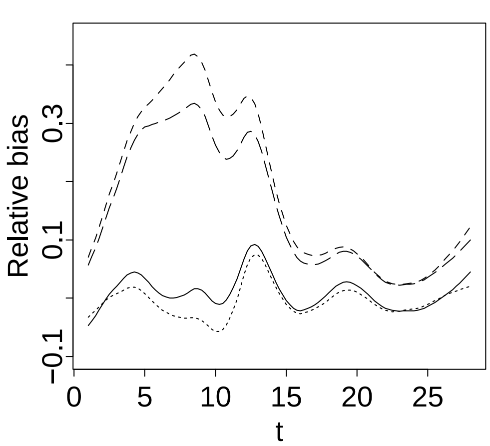
Sun, D., Tsutakawa, R. K. and He, Z. (2001), Properties of posteriors with improper priors in hierarchical linear mixed models, *Statistica Sinica* **11**, 77–95.

West, M. (1992), Hyperparameter estimation in Dirichlet process mixture models, Technical report, ISDS discussion paper 1992-03, Duke University. Available at http://www.stat.duke.edu/research/papers.



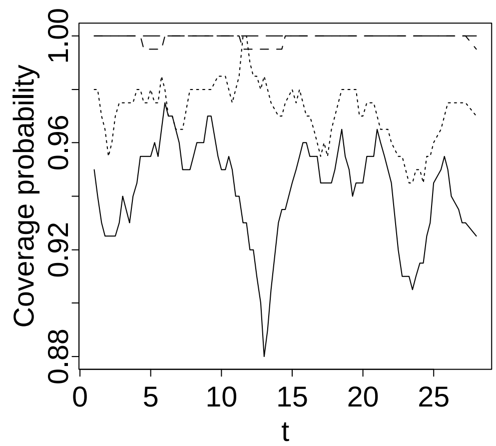
(a) (b)

Figure 1: Pointwise relative biases and 95% pointwise coverage probabilities for $f(t)$ under model (6) with the true random intercept distribution as either $N(0, .2303)$ or $11/18 \times N(-0.35, 0.03) + 7/18 \times N(0.55, 0.05)$. Normal prior is assumed for all random intercept distributions. In all panels, ——: Normal random intercept + uniform shrinkage prior (USP) (average relative bias = .01, average coverage probability = 93%, average credible interval (CI) length = .44) $\cdots\cdots$: Normal random intercept + inverse gamma (IG) prior (average relative bias = .002, average coverage probability = 95%, average CI length = .46) $- - -$: Bimodal random intercept + USP (average relative bias = -.02, average coverage probability = 93%, average CI length = .44) — —: Bimodal random intercept + IG prior (average relative bias = .003, average coverage probability = 96%, average CI length = .46). Results are based on 200 replicates.



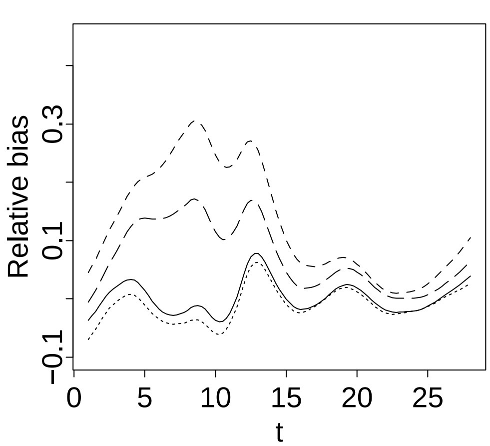
$(a)\, M = .75$

ARB: .01, -.01, .19, .16



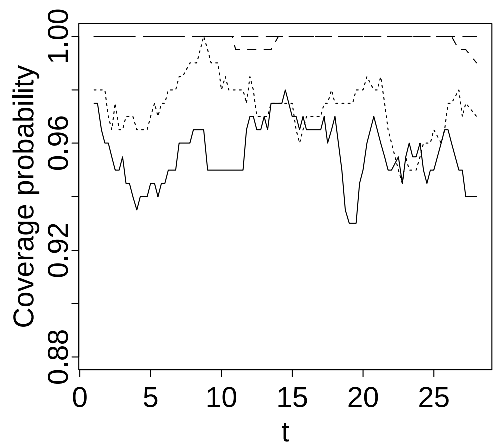
$(b)\, M = .75$

ACP (ACIL): 94 (.43), 97 (.49), 100 (1.18), 100 (1.69)



$(c)\, M \sim G(.5, 1)$

ARB: .001, -.01, .13, .07



$(d)\, M \sim G(.5, 1)$

ACP (ACIL): 96 (.44), 97(.49), 100 (.98), 100 (1.24)

Figure 2: Pointwise relative biases and 95% pointwise coverage probabilities for $f(t)$ under model (6) with a true random intercept distribution of $11/18 \times N(-0.35, 0.03) + 7/18 \times N(0.55, 0.05)$. The Dirichlet process (DP) prior with center adjustment (DPP-ADJ) and the traditional DP prior without center adjustment (DPP-UN) with a fixed $M\,(= 0.75)$ or a gamma prior for $M\,(\sim G(.5, 1))$ for the random intercept distribution. In all four panels, the line types correspond to the following procedures/priors used: —— ADJ/USP; $\cdots\cdots$ ADJ/IGP; - - - UN/USP; $- - -$ UN/IGP. The average relative biases (ARB), average coverage probabilities (ACP) and average credible interval lengths (ACIL) are reported in the above specified order. Results are based on 200 replicates.