Biometrics 66, 70–78
March 2010
DOI: 10.1111/j.1541-0420.2009.01227.x

Bayesian Inference in Semiparametric Mixed Models for Longitudinal Data

Yisheng Li,[1,*] Xihong Lin,[2] and Peter Müller[1]

[1]Department of Biostatistics, Division of Quantitative Sciences, University of Texas M. D. Anderson Cancer Center, Houston, Texas 77030, U.S.A.

[2]Department of Biostatistics, Harvard School of Public Health, Boston, Massachusetts 02115, U.S.A.

[*]*email:* ysli@mdanderson.org

Summary. We consider Bayesian inference in semiparametric mixed models (SPMMs) for longitudinal data. SPMMs are a class of models that use a nonparametric function to model a time effect, a parametric function to model other covariate effects, and parametric or nonparametric random effects to account for the within-subject correlation. We model the nonparametric function using a Bayesian formulation of a cubic smoothing spline, and the random effect distribution using a normal distribution and alternatively a nonparametric Dirichlet process (DP) prior. When the random effect distribution is assumed to be normal, we propose a uniform shrinkage prior (USP) for the variance components and the smoothing parameter. When the random effect distribution is modeled nonparametrically, we use a DP prior with a normal base measure and propose a USP for the hyperparameters of the DP base measure. We argue that the commonly assumed DP prior implies a nonzero mean of the random effect distribution, even when a base measure with mean zero is specified. This implies weak identifiability for the fixed effects, and can therefore lead to biased estimators and poor inference for the regression coefficients and the spline estimator of the nonparametric function. We propose an adjustment using a postprocessing technique. We show that under mild conditions the posterior is proper under the proposed USP, a flat prior for the fixed effect parameters, and an improper prior for the residual variance. We illustrate the proposed approach using a longitudinal hormone dataset, and carry out extensive simulation studies to compare its finite sample performance with existing methods.

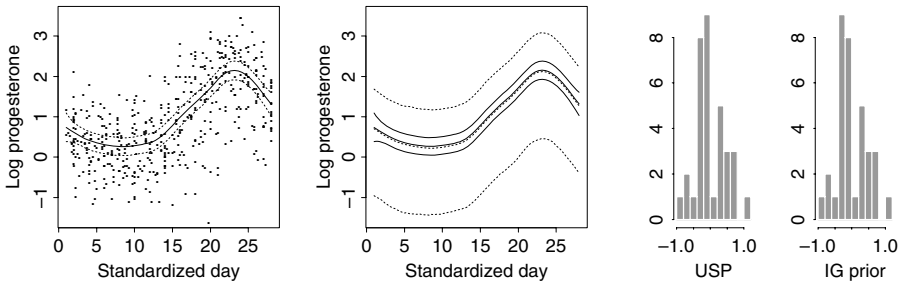
Key words: Dirichlet process prior; Identifiability; Postprocessing; Random effects; Smoothing spline; Uniform shrinkage prior; Variance components.

1. Introduction

Longitudinal data arise in many fields, such as epidemiology, clinical trials, and survey research. Linear mixed models (LMMs) (Laird and Ware, 1982) are commonly used for longitudinal data analysis, where covariate effects are modeled parametrically and within-subject correlation is modeled using random effects. Semiparametric mixed models (SPMMs) extend LMMs by modeling a covariate effect, e.g., time effect, using a nonparametric function (Zeger and Diggle, 1994; Zhang et al., 1998), and other covariate effects parametrically. Inference in SPMMs has been mainly developed using frequentist methods, such as kernel and profile methods (Zeger and Diggle, 1994; Lin and Carroll, 2001; Fan and Li, 2004; Wang, Carroll, and Lin, 2005) and smoothing spline and P-spline methods using a joint maximum penalized likelihood (Zhang et al., 1998; Verbela et al., 1999; Ruppert, Wand, and Carroll, 2003). All these approaches assume a parametric normal distribution for the random effects. In this article, we develop Bayesian methods for inference in SPMMs. We model the time effect nonparametrically using a Bayesian formulation of splines and model the random effect distribution parametrically or nonparametrically using a Dirichlet process (DP) prior.

This work is motivated by a longitudinal study on the reproductive hormone progesterone (Sowers et al., 1998). Scientific interest included estimation of the time course of the progesterone level in a menstrual cycle as well as the effects of age and body mass index (BMI). Figure 1a plots the log-transformed progesterone level as a function of day in a standardized menstrual cycle, suggesting that progesterone level varies over time in a complicated manner, and that it is difficult to model its time trend with a simple parametric function. Figure 1c shows histograms of the posterior estimates (means) of the random intercepts assuming a normal random effect distribution. It suggests substantial departure from normality in the random effect distribution (Verbeke and Lesaffre, 1996). These observations motivate us to develop a Bayesian method with a nonparametric model for the progesterone profile and a nonparametric random effect distribution.

Bayesian methods for parametric LMMs have been extensively developed. Assuming normal random effects, posterior inference can be easily implemented using standard Gibbs sampling. Kleinman and Ibrahim (1998) relaxed the normality assumption by modeling the random effect distribution nonparametrically using a DP prior assuming a normal base measure with a zero mean. Similar approaches have been considered by many other authors, including, e.g., Müller and Rosner (1997), Mukhopadhyay and Gelfand (1997), and Pennell and Dunson (2007).



(a) (b) (c)

Figure 1. The posterior means and 95% pointwise CIs of $f(t)$ for the progesterone data. (a) Assuming normal random intercepts and USPs for the random intercept variance and smoothing parameter: —— posterior mean ····· 95% CIs. Average $\widehat{f(t)}$, denoted as $\widehat{\widehat{f(t)}} = 0.98$ and average CI length = 0.46. The unconnected dots represent the raw data. (b) ——: Posterior means and 95% CIs assuming a DP prior with center adjustment and a gamma prior for $M(\sim G(0.5,\ 1))$ for the random intercept distribution and USPs for the random intercept variance and smoothing parameter ($\widehat{\widehat{f(t)}} = 0.99$, Average CI length = 0.45). ·····: Posterior means and 95% CIs assuming the traditional DP prior without center adjustment under fixed $M(=0.75)$ and IG priors for the random intercept variance and smoothing parameter ($\widehat{\widehat{f(t)}} = 0.94$, Average CI length = 2.62). (c) Histograms of the 34 posterior means of the random intercepts when their distribution is assumed normal for the progesterone data. The left and right panels correspond to the models assuming USPs and IG priors for the variance of the random intercepts and smoothing parameter, respectively.

In this article, we discuss extensions to SPMMs. Spline estimation in the SPMM relies heavily on estimation of the smoothing parameter, which is treated as an extra variance component in the Bayesian model. Thus the choice of appropriate prior distributions for variance components is critical. We develop a uniform shrinkage prior (USP) for variance components in SPMMs and study its theoretical properties. Conjugate inverse gamma (IG) and inverse Wishart (IW) priors have been commonly used for variance components, mainly for computational convenience. Several authors have noted serious difficulties in using diffuse IG priors for random effect variances (Natarajan and Kass, 1998, 2000; Gelman, 2006), and showed that the resulting posterior is near improper and the variance component estimators could be excessively biased, especially when the number of subjects is small. Similar problems arise in using diffuse IW priors. To address this issue, Daniels (1999) and Natarajan and Kass (2000) proposed USPs for variance components in two-level parametric (generalized) LMMs. We show that their proposed USPs are restricted to problems where random effects are independent and the associated design matrix is block diagonal. A Bayesian smoothing spline in SPMMs induces additional random effects that result in a nonblock-diagonal design matrix for combined random effect vectors. Therefore, the methods from Daniels (1999) and Natarajan and Kass (2000) are not applicable. We propose modified USPs that can be applied. We show that under mild conditions, the proposed USPs combined with a flat prior for the fixed effect parameters and an improper prior for the residual variance lead to a proper posterior distribution.

The second contribution of this article is to model random effects using a center-adjusted DP prior in SPMMs. We show that a standard DP prior can lead to biased estimators and poor inference on the nonparametric mean function. We develop a postprocessing procedure to adjust inference for the nonparametric function. Extensive simulation studies are conducted to evaluate finite sample performance of the proposed methods and to compare with the existing methods.

Section 2 presents the Bayesian SPMM. Section 3 develops the USP for the variance components when the random effect distribution is assumed normal. Section 4 discusses posterior adjustments when using nonparametric random effect distribution with a nonzero mean. Section 5 reports an application. Section 6 summarizes extensive simulation studies.

2. A Bayesian SPMM

Let $Y_{ij}(i = 1, \ldots, m, j = 1, \ldots, n_i)$ be the outcome for the ith subject at time t_{ij}. We assume Y_{ij} follows the SPMM (Zeger and Diggle, 1994; Zhang et al., 1998)

$$Y_{ij} = \mathbf{X}_{ij}^T \boldsymbol{\beta} + f(t_{ij}) + \mathbf{Z}_{ij}^T \boldsymbol{b}_i + \epsilon_{ij}, \tag{1}$$

where $\boldsymbol{\beta}$ is a $p \times 1$ vector of fixed effects associated with covariates $\mathbf{X}_{ij}$, $f(t)$ is an unknown twice-differentiable smooth function of time, and $\boldsymbol{b}_i$ is a $q \times 1$ vector of random effects associated with covariates $\mathbf{Z}_{ij}$ and follows some distribution $F(\cdot)$. We consider both parametric and nonparametric models for the random effect distribution $F(\boldsymbol{b}_i)$. We assume residuals $\epsilon_{ij} \sim N(0,\ \sigma^2)$ are independent and are independent of $\boldsymbol{b}_i$.

We estimate the nonparametric function $f(t)$ using a smoothing spline by assuming $f(t)$ follows the integrated Wiener prior (Wahba, 1978)

$$f(t) = \delta_0 + \delta_1 t + \lambda^{-1/2} \int_0^t (t-u) dW(u),\ t \in [T_1, T_2], \tag{2}$$

where $\boldsymbol{\delta} = (\delta_0, \delta_1)^T$ is an unknown 2×1 vector of parameters, λ is a tuning parameter, T_1 and T_2 specify the range of t, and $\int_0^t (t-u)dW(u)$ is a two-fold integrated Wiener process with $W(t)$ being a standard Wiener process. Because $f(t)$ already includes an intercept and a linear term in t, we assume that the design vector $\mathbf{X}_{ij}$ in equation (1) does not have rows corresponding to 1 or t.

Following Zhang et al. (1998), denote by $\mathbf{t}^0$ a vector of ordered distinct time points of the $t_{ij} (i = 1, \ldots, m, j = 1, \ldots, n_i)$ and $\mathbf{T} = (\mathbf{1}, \mathbf{t}^0)$, where $\mathbf{1}$ is an $r \times 1$ vector of 1 with r being the number of distinct time points. Let $\mathbf{Y}_i = (Y_{i1}, \ldots, Y_{in_i})^T$, and define $\mathbf{X}_i, \mathbf{Z}_i, \boldsymbol{\epsilon}_i (i = 1, \ldots, m)$ similarly. Let $\mathbf{N}_i$ be an incidence matrix for the ith subject mapping $\mathbf{t}_i = (t_{i1}, \ldots, t_{in_i})^T$ with $\mathbf{t}^0$ such that the (j, l)th element of $\mathbf{N}_i$ is 1 if $t_{ij} = t_l^0$ and 0 otherwise ($j = 1, \ldots, n_i$, $l = 1, \ldots, r$). Then model (1) can be written as

$$\mathbf{Y}_i = \mathbf{X}_i \boldsymbol{\beta} + \mathbf{N}_i \mathbf{f} + \mathbf{Z}_i \boldsymbol{b}_i + \boldsymbol{\epsilon}_i, \tag{3}$$

where $\mathbf{f} = \{f(t_1^0), \ldots, f(t_r^0)\}^T$. Further denote $\mathbf{Y} = (\mathbf{Y}_1^T, \ldots, \mathbf{Y}_m^T)^T$, and similarly define $\mathbf{X}$, $\mathbf{N}$ and $\boldsymbol{\epsilon}$. Let $\mathbf{Z} = diag(\mathbf{Z}_1, \ldots, \mathbf{Z}_m)$. We have

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{N}\mathbf{f} + \mathbf{Z}\boldsymbol{b} + \boldsymbol{\epsilon}, \tag{4}$$

where $\boldsymbol{b} = (\boldsymbol{b}_1^T, \ldots, \boldsymbol{b}_m^T)^T$ and $\boldsymbol{\epsilon} = (\boldsymbol{\epsilon}_1^T, \ldots, \boldsymbol{\epsilon}_m^T)^T \sim N(\mathbf{0}, \sigma^2 \mathbf{I})$ with $\mathbf{I}$ denoting an identity matrix of dimension $n = \sum_{i=1}^n n_i$.

The integrated Wiener prior (2) for the smoothing spline $f(t)$ is equivalent to a finite-dimensional smoothing spline prior for $\mathbf{f}$. Let $\mathbf{K}$ denote the integrated Wiener covariance at the design points $\mathbf{t}^0$ with the (j, k)th element $\{t_j^0\}^2 (3t_j^0 - t_k^0)/6$ for $t_j^0 \geq t_k^0$. Let $\mathbf{L}$ be an $r \times (r-2)$ full rank matrix satisfying $\mathbf{K} = \mathbf{L}\mathbf{L}^T$ and $\mathbf{L}^T\mathbf{T} = \mathbf{0}$. Finally let $\mathbf{B} = \mathbf{L}(\mathbf{L}^T\mathbf{L})^{-1}$. We assume

$$\mathbf{f} = \mathbf{T}\boldsymbol{\delta} + \mathbf{B}\boldsymbol{a}, \tag{5}$$

where $\boldsymbol{\delta}$ is a 2×1 vector, $\boldsymbol{a}$ is an $r \times 1$ vector. Equation (5) can be interpreted as specifying that a priori $f(\cdot)$ is centered around a linear function with nonlinearities (or roughness) characterized by the random variation term $\mathbf{B}\boldsymbol{a}$. Let $\mathbf{0}$ and $\mathbf{I}$ generically denote a vector of all zeroes and the identity matrix of suitable dimension. The finite-dimensional Bayesian smoothing spline assumes a nearly flat prior $\boldsymbol{\delta} \sim N(\mathbf{0}, d\mathbf{I})$ and a normal prior $\boldsymbol{a} \sim N(\mathbf{0}, \tau\mathbf{I})$, where d is a large constant, and τ is the smoothing parameter. The spline (5) is equivalent to equation (2) in the sense that the posterior mean of $f(t^0)$ is identical under both priors (5) and (2). This vague prior specification for the smoothing spline is applicable when little prior information is available regarding the feature of the underlying $f(t)$. When features of $f(t)$ are known a priori, such information should be incorporated in the prior, leading to alternative approaches of modeling $f(t)$. It follows that the SPMM (4) has the following LMM representation

$$\mathbf{Y} = \mathbf{X}_\star \boldsymbol{\beta}_\star + \mathbf{Z}^{(1)}\boldsymbol{b} + \mathbf{Z}^{(2)}\boldsymbol{a} + \boldsymbol{\epsilon}, \tag{6}$$

where $\mathbf{Z}^{(1)} = \mathbf{Z}, \mathbf{X}_\star = (\mathbf{X}, \mathbf{N}\mathbf{T}), \boldsymbol{\beta}_\star = (\boldsymbol{\beta}^T, \boldsymbol{\delta}^T)^T, \mathbf{Z}^{(2)} = \mathbf{N}\mathbf{B}$, and $\boldsymbol{b}$ and $\boldsymbol{a}$ are random effects with $\boldsymbol{b} \sim F(\cdot)$ that might depend on variance components $\boldsymbol{D}$, and $\boldsymbol{a} \sim N(0, \tau\mathbf{I})$ with the smoothing parameter τ being treated as an extra variance component.

3. A Uniform Shrinkage Prior for the SPMM

In this section, we consider SPMM (6) with a normal random effect distribution, $\boldsymbol{b} \sim N(\mathbf{0}, \mathcal{D})$, where $\mathcal{D} = diag(\boldsymbol{D}, \ldots, \boldsymbol{D})$. We propose a USP for the covariance matrix $\boldsymbol{D}$ and the smoothing parameter τ. The idea of the USP is easiest explained in a simple normal sampling model, $x_i \overset{i.i.d.}{\sim} N(\mu, \sigma^2)$, $i = 1, \ldots, n$, with prior $\mu \sim N(0, \sigma_0^2)$. The posterior mean, $E(\mu \,|\, \sigma^2, x_1, \ldots, x_n) = (\sigma^2 \cdot 0 + n\sigma_0^2 \bar{x}) / (\sigma^2 + n\sigma_0^2)$ is a shrinkage estimator, i.e., a weighted average of the prior mean 0 and the sample average $\bar{x} = (1/n)\sum_{i=1}^n x_i$. Assuming a uniform prior for the shrinkage coefficient $\sigma^2/(\sigma^2 + n\sigma_0^2)$ implicitly defines a prior for σ^2 known as the USP. Daniels (1999) and Natarajan and Kass (2000) extended this idea to two-stage hierarchical models. Assuming independence across clusters, a shrinkage matrix for each individual random effect vector is defined. This independence is equivalent to assuming a block-diagonal design matrix for the random effects. An average shrinkage matrix for all random effect vectors can then be sensibly defined, e.g., through the use of a harmonic mean in a simple normal–normal model (Daniels, 1999).

However, in the SPMM-induced LMM (6), a shrinkage matrix for each individual random effect vector cannot be readily defined due to two reasons. First, the model contains two sets of random effects, $\boldsymbol{b}$ and $\boldsymbol{a}$. The responses $\mathbf{Y}_i$ and $\mathbf{Y}_j$ are correlated by sharing the same set of random effects $\boldsymbol{a}$. As a result, there is not a shrinkage matrix associated with each individual $\boldsymbol{b}_i$ estimate alone. Second, even without the random effects $\boldsymbol{b}$ in the model, because the design matrix $\mathbf{Z}^{(2)}$ is not block diagonal, the shrinkage matrix for $\boldsymbol{a}$ is still not diagonal. In other words, there is no shrinkage coefficient that is naturally associated with each a_i either. Therefore, in model (6) a new definition of the joint USP for $(\boldsymbol{D}, \tau)$ (conditional on σ^2) is needed. Below we propose a stepwise procedure to define a prior for $(\sigma^2, \boldsymbol{D}, \tau)$ by deriving conditional USPs for $\boldsymbol{D}$ given σ^2, and τ given $(\sigma^2, \boldsymbol{D})$.

We first factor the prior on $(\sigma^2, \boldsymbol{D}, \tau)$ as $\pi(\sigma^2)$ $\pi(\boldsymbol{D} \,|\, \sigma^2)\pi(\tau \,|\, \sigma^2, \boldsymbol{D})$. We assume $\pi(\sigma^2) \sim IG(\alpha, \nu) \propto \exp(-\nu/\sigma^2)/(\sigma^2)^{\alpha+1}$, using, e.g., $\alpha = \nu = 0.01$, to represent a vague prior.

To derive $\pi(\boldsymbol{D} \,|\, \sigma^2)$, we first consider the following simplified LMM:

$$\mathbf{Y} = \mathbf{X}_\star \boldsymbol{\beta}_\star + \mathbf{Z}^{(1)}\boldsymbol{b} + \boldsymbol{\epsilon}, \tag{7}$$

i.e., model (6) with random effects $\boldsymbol{a}$ removed. Let $V = 1/m \sum_{i=1}^m \mathbf{Z}_i^{(1)^T} \mathbf{Z}_i^{(1)}$. Following Daniels (1999) and Natarajan and Kass (2000), we define an average shrinkage matrix for the posterior mean of $\boldsymbol{b}_i$ conditional on $(\boldsymbol{\beta}_\star, \sigma^2)$ as $\left(\boldsymbol{D}^{-1} + 1/\sigma^2 V\right)^{-1} 1/\sigma^2 V$. By placing a uniform prior on the above shrinkage matrix, we obtain the conditional prior

$$\pi(\boldsymbol{D} \,|\, \sigma^2) \propto \left|\mathbf{I} + \boldsymbol{D}\left\{1/\sigma^2\, V\right\}\right|^{-q-1}.$$

To derive $\pi(\tau \,|\, \sigma^2, \boldsymbol{D})$, we rewrite the SPMM-induced LMM (6) as

$$\mathbf{Y} = \mathbf{X}_\star \boldsymbol{\beta}_\star + \mathbf{Z}^{(2)}\boldsymbol{a} + \boldsymbol{\epsilon}_\star, \tag{8}$$

where $\boldsymbol{\epsilon}_\star = \mathbf{Z}^{(1)}\ \boldsymbol{b} + \boldsymbol{\epsilon}$. Let $\boldsymbol{R} = \mathbf{Z}^{(1)}\mathcal{D}\mathbf{Z}^{(1)^T} + \sigma^2\mathbf{I}$.

The lack of a block-diagonal structure of $\mathbf{Z}^{(2)}$ implies that the posterior mean of each a_i conditional on $(\boldsymbol{\beta}_\star, \boldsymbol{D}, \sigma^2)$

cannot be represented as a shrinkage estimator toward its prior mean 0. In other words, there is no natural definition of a shrinkage coefficient associated with the estimate of each a_i. This hinders the application of the USP proposed by Daniels (1999) or Natarajan and Kass (2000). We propose to work with the shrinkage matrix in the posterior mean of the vector $\boldsymbol{a}$, which can be shown to be equal to $\mathbf{S} = (\tau^{-1}\mathbf{I} + \mathbf{Z}^{(2)^T}\boldsymbol{R}^{-1}\mathbf{Z}^{(2)})^{-1}1/\tau\,\mathbf{I}$.

Let $\mathbf{Q}$ be the diagonalizing orthogonal matrix that satisfies $\mathbf{Z}^{(2)^T}\boldsymbol{R}^{-1}\mathbf{Z}^{(2)} = \mathbf{Q}\boldsymbol{\Lambda}_{\sigma^2,\boldsymbol{D}}\mathbf{Q}^T$, where $\boldsymbol{\Lambda}_{\sigma^2,\boldsymbol{D}}$ is diagonal. Denote $\lambda_i \geq 0$ the diagonal elements of $\boldsymbol{\Lambda}_{\sigma^2,\boldsymbol{D}}$. Both $\boldsymbol{Q}$ and λ_i are functions of (σ^2, $\boldsymbol{D}$), but not of τ. The shrinkage matrix $\mathbf{S}$ hence can be rewritten as $\mathbf{S} = \mathbf{Q}\mathbf{S}_\star\mathbf{Q}^T$, where $\mathbf{S}_\star = (\tau^{-1}\mathbf{I} + \boldsymbol{\Lambda}_{\sigma^2,\boldsymbol{D}})^{-1}\tau^{-1}\mathbf{I}$. Because $\mathbf{Q}$ does not depend on τ, $\mathbf{S}$ can be regarded as a linear transformation of $\mathbf{S}_\star$. Therefore, a uniform prior on $\mathbf{S}$ is equivalent to a uniform prior on $\mathbf{S}_\star$. The diagonalized shrinkage matrix $\mathbf{S}_\star$ contains $r-2$ shrinkage coefficients $(\tau^{-1} + \lambda_i)^{-1}\,\tau^{-1}$ along the diagonal. Conditional on (σ^2, $\boldsymbol{D}$), in the spirit of Daniels (1999) and Natarajan and Kass (2000), we define a USP for τ by placing a uniform prior on $(\tau^{-1} + \bar{\lambda})^{-1}\tau^{-1}$ with $\bar{\lambda} = (r-2)^{-1}\sum_{i=1}^{r-2}\lambda_i$. This leads to a USP for τ as

$$\pi(\tau \mid \sigma^2, \boldsymbol{D}) \propto 1/(1+\bar{\lambda}\tau)^2. \tag{9}$$

THEOREM 1: *In the normal SPMM* (6)*, assume* ($\mathbf{X}_\star$, $\mathbf{Z}^{(1)}$, $\mathbf{Z}^{(2)}$) *is of full column rank* $p + mq + r$ *and* $n > p + mq + r$. *Then* (1) *the joint conditional USP* $\pi(\boldsymbol{D}, \tau \mid \sigma^2)$ *is proper; and* (2) *under the improper prior* $\pi(\boldsymbol{\beta}_\star, \boldsymbol{D}, \tau, \sigma^2) \propto 1/\sigma^2 \cdot \pi(\boldsymbol{D}, \tau \mid \sigma^2)$*, the posterior is proper.*

The proof is given in Web Appendix A.1. Statement (1) is one of the desirable properties for a noninformative prior as described in Daniels (1999). Statement (2) allows us to combine the proposed USP with commonly used improper priors for the fixed effects and residual variance.

The proposed USP is applicable to a general class of LMMs. Specifically, consider the following general LMM:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathcal{Z}\boldsymbol{c} + \boldsymbol{\epsilon}, \tag{10}$$

where the random effects $\boldsymbol{c} = (c_1, \ldots, c_m) \sim N(\mathbf{0}, \tau\mathbf{I})$ and $\boldsymbol{\epsilon} \sim N(\mathbf{0}, \boldsymbol{R})$. The SPMM-induced LMM (8) is a special case, with $\boldsymbol{c} = \boldsymbol{a}$, $\mathcal{Z} = \boldsymbol{Z}^{(2)}$. The posterior mean of $\boldsymbol{c}$ involves a shrinkage matrix $\mathbf{S} = \left(\tau^{-1}\mathbf{I} + \mathcal{Z}^T\boldsymbol{R}^{-1}\mathcal{Z}\right)^{-1}\tau^{-1}\mathbf{I}$ toward $\mathbf{0}$. Denote λ_i the eigenvalues of $\mathcal{Z}^T\boldsymbol{R}^{-1}\mathcal{Z}$ and let $\bar{\lambda} = 1/m \cdot \sum_{i=1}^{m}\lambda_i$. The USP for τ in equation (10) can be defined by placing a uniform prior on $\left(1/\tau + \bar{\lambda}\right)^{-1}1/\tau$. When the matrix $\mathcal{Z}^T\boldsymbol{R}^{-1}\mathcal{Z}$ is block diagonal, e.g., two-stage clustered LMMs, the proposed USP reduces to the corresponding USP defined by Natarajan and Kass (2000) in the normal outcome case. For independent data, it reduces to that in Daniels (1999). The proofs are given in Web Appendix A.2.

One can also show a more general posterior propriety result. That is, Theorem 1 holds for general LMM (10) as long as the covariance matrix has a conditionally (on σ^2) proper prior, assuming the above improper priors for $\boldsymbol{\beta}$ and σ^2. These results extend the results of Chen, Shao, and Xu (2002) for two-stage clustered random effect models with a normal outcome to general LMMs.

Finally, we note one important computational detail. To compute the prior (9), one can write $\bar{\lambda} = \mathrm{tr}(\mathbf{Z}^{(2)^T}\boldsymbol{R}^{-1}\mathbf{Z}^{(2)})/(r-2) = \mathrm{tr}(\mathbf{Z}^{(2)}\mathbf{Z}^{(2)^T}\boldsymbol{R}^{-1})/(r-2)$, where $tr(\cdot)$ represents the trace of a matrix. This way the calculation of $\bar{\lambda}$ does not involve finding the orthogonal transformation of $\mathbf{Z}^{(2)T}\ \boldsymbol{R}^{-1}\ \mathbf{Z}^{(2)}$, and the computational burden is greatly reduced. In addition, $\boldsymbol{R}^{-1}$ can be computed using the Schur complement formula (Searle, 1982, p. 261). This approach converts calculation of the inverse of an $n \times n$ matrix to inverting an $(m \times q) \times (m \times q)$ matrix.

Posterior Markov chain Monte Carlo (MCMC) simulation is implemented using the Gibbs sampler with Metropolis steps for sampling the variance components. See Web Appendix A.5 for details.

4. Bayesian SPMMs with Nonparametric Random Effects

4.1 *SPMM with a Hierarchically Centered DP for Random Effects*

In this section, we discuss inference in the SPMM (1) when the random effect distribution $F(\boldsymbol{b})$ is assigned a nonparametric DP prior. Specifically, we assume

$$\boldsymbol{b}_i \mid G \overset{\text{i.i.d.}}{\sim} G,\ G \sim DP(M, G_0), \tag{11}$$

where $DP(M, G_0)$ denotes a DP with a total mass parameter M and base measure G_0.

In LMMs when a DP prior is assumed for the random effect distribution, the base measure G_0 is typically assumed to have a zero mean, as in $G_0 = N(\mathbf{0}, \boldsymbol{D})$ (Christopher and MacEachern, 1996; Kleinman and Ibrahim, 1998; Pennell and Dunson, 2007, among others). Despite the centering of $\boldsymbol{b}_i$ at zero *a priori*, i.e., $E(\boldsymbol{b}_i \mid \boldsymbol{D}) = \mathbf{0}$, the random distribution G has a nonzero mean almost surely. This will lead to biased inference for the fixed effects that are paired with the random effects $\boldsymbol{b}_i$, i.e., effects associated with common columns in the design matrices $\mathbf{X}$ and $\mathbf{Z}$, e.g., an intercept and a random intercept (Li, Müller, and Lin, 2007). The problem arises in SPMMs because the spline estimation of the nonparametric function $f(t)$ involves the fixed effect intercept and slope of time, and the subject-specific random effects $\boldsymbol{b}_i$ often contain random intercepts. We propose to adjust with the random moments of a hierarchically centered DP to address this issue. We refer to the resulting inference as center-adjusted inference.

The center-adjusted DP approach involves two modifications, namely rewriting model (11) as a centered hierarchical model and a postprocessing adjustment. First, following Li et al. (2007), we remove the paired fixed effects from $\boldsymbol{X}_\star$ in equation (6) and absorb them in the paired random effects by defining $G_0 = N(\boldsymbol{\beta}_b, \boldsymbol{D})$ with $\boldsymbol{\beta}_b$ being an unknown parameter vector. Therefore, the SPMM (6) is rewritten as

$$\mathbf{Y}_i = \tilde{\mathbf{X}}_i\tilde{\boldsymbol{\beta}} + \mathbf{Z}_i^{(2)}\boldsymbol{a} + \mathbf{Z}_i^{(1)}\boldsymbol{b}_i + \boldsymbol{\epsilon}_i,\ i = 1, \ldots, m, \tag{12}$$

where $\tilde{\boldsymbol{\beta}}$ corresponds to $\boldsymbol{\beta}_\star$ with the paired fixed effects removed, and similarly for $\tilde{\mathbf{X}}$.

The matrices $\mathbf{Z}_i^{(2)} = \mathbf{N}_i\mathbf{B}$, and $\tilde{\mathbf{X}}_i$ and $\mathbf{Z}_i^{(1)}$, are the matrix blocks in $\tilde{\mathbf{X}}$ and $\mathbf{Z}^{(1)}$ that correspond to the ith subject; and $\boldsymbol{b}_i \sim DP(M, G_0)$, with $G_0 = N(\boldsymbol{\beta}_b, \boldsymbol{D})$. We complete the DP SPMM with hyperpriors on $\boldsymbol{D}$, $\boldsymbol{\beta}_\mathrm{b}$, τ, σ^2. Specifically, we use the USP defined in Section 3 for ($\boldsymbol{D}$, τ). For $\boldsymbol{\beta}_\mathrm{b}$, we assume

$\boldsymbol{\beta}_{\mathbf{b}} \sim N(\mathbf{0}, \boldsymbol{\Sigma}_0)$ with $\boldsymbol{\Sigma}_0 = d' \cdot \mathbf{I}$, where d' is a large constant. We continue using an IG prior for σ^2.

The second and novel modification is an adjustment in the reported posterior inference. Inference about $\boldsymbol{\beta}_{\mathrm{b}}$ is replaced by inference on the (random) moments of the random probability measure G. Closed-form expressions for the conditional moments are given in Li et al. (2007). See also Section 4.2.

Similar to the earlier discussion of posterior propriety for the normal SPMMs, we are interested in posterior propriety under the DP SPMM with an improper prior on $(\tilde{\boldsymbol{\beta}}, \boldsymbol{\beta}_{\mathbf{b}}, \sigma^2)$. This seems to have been largely ignored in the literature due to the common use of conjugate vague yet proper priors. Li et al. (2007) show posterior propriety in a LMM similar to equation (12), except with the $\mathbf{Z}_i^{(2)}\, \boldsymbol{a}$ term excluded. The inclusion of the random effects $\mathbf{Z}_i^{(2)}\, \boldsymbol{a}$ for modeling the smoothing spline leads to dependence across subjects, thus preventing a direct application of their results. Theorem 2 shows that a similar result still holds.

THEOREM 2: *In the DP SPMM* (12)*, assuming the centered DP prior and the conditions in Theorem* 1 *hold, under the improper prior* $\pi(\tilde{\boldsymbol{\beta}}, \boldsymbol{\beta}_{\mathbf{b}}, \boldsymbol{D}, \tau, \sigma^2) \propto 1/\sigma^2 \cdot \pi(\boldsymbol{D}, \tau \,|\, \sigma^2)$*, the posterior is proper.*

The proof of Theorem 2 is given in Web Appendix A.3. The conclusion remains valid when a DP prior is assumed for only a subvector of the random effects, with a multivariate normal distribution for the remaining random effects and an arbitrary design matrix for all random effects.

4.2 *Center-Adjusted Inference for Fixed Effects Paired with Nonparametric Random Effects*

Posterior simulation in model (12) is carried out by MCMC simulation. See Web Appendix A.6 for details. We now describe a postprocessing technique to adjust for the random moments of G. Li et al. (2007, Propositions 2 and 3) give explicit formulas for the posterior conditional mean and covariance matrix of the (random) first two moments, $\boldsymbol{\mu}_G = \int \boldsymbol{b} dG(\boldsymbol{b})$ and $\mathbf{Cov}_G = \int (\boldsymbol{b} - \boldsymbol{\mu}_G)(\boldsymbol{b} - \boldsymbol{\mu}_G)^T dG(\boldsymbol{b})$ of the random probability measure G. We use these closed-form expressions to draw inference on the fixed effects paired with $\boldsymbol{b}_i$ and the variance components of $\boldsymbol{b}_i$. Let $\mathbf{y}$ generically denote the observed data. We report $E(\boldsymbol{\mu}_G \,|\, \mathbf{y})$ and Cov $(\boldsymbol{\mu}_G \,|\, \mathbf{y})$ as posterior inference for the fixed effects paired with $\boldsymbol{b}_i$, and $E(\mathbf{Cov}_G \,|\, \mathbf{y})$ and Cov $(\mathbf{Cov}_G \,|\, \mathbf{y})$ as posterior moments for the variance components of $\boldsymbol{b}_i$. We report posterior credible intervals (CIs) for the components of the mean and covariance matrix of G using either a normal approximation or by matching their first and second moments to those of a log-normal distribution, the latter being used for the variances. We select the log-normal distribution because of its positive support.

4.3 *Center-Adjusted Inference for $f(t)$*

The estimation of $f(t)$ in equation (12) involves estimation of the fixed effect intercept, slope of time, and induced random effects $\boldsymbol{a}$. If the nonparametric random effects $\boldsymbol{b}_i$ contain a random intercept, slope, or both, then we need to adjust inference on $f(t)$, similar to the adjusted inference for fixed effects. When no random intercept or slope term is present, no adjustment is required for $f(t)$.

In equation (12), let $\boldsymbol{\beta}_{\mathrm{c}}$ and $\boldsymbol{c}_i$ be the subvectors of $\boldsymbol{\beta}_{\mathrm{b}}$ and $\boldsymbol{b}_i$ that correspond to the fixed and random intercept/slope/both, as appropriate. Let $r = M/(m + M)$, as before. Further let $\boldsymbol{\mu}_{G_\star,\mathbf{c}} = r\boldsymbol{\beta}_{\mathbf{c}} + \sum_{i=1}^m (1-r)\boldsymbol{c}_i$ and $\mathbf{Cov}_{G_\star,\mathbf{c}} = \left\{ r\left(\boldsymbol{\beta}_{\mathbf{c}}\boldsymbol{\beta}_{\mathbf{c}}{}^T + \boldsymbol{D}\right) + (1-r)\sum_{i=1}^m \boldsymbol{c}_i \boldsymbol{c}_i{}^T \right\} - \boldsymbol{\mu}_{G_\star,\mathbf{c}}\boldsymbol{\mu}_{G_\star,\mathbf{c}}{}^T$. If $\boldsymbol{b}_i$ includes a random intercept and/or slope, then we report posterior inference on $\mathbf{f} \equiv \mathbf{T}\boldsymbol{\mu}_G + \mathbf{B}\boldsymbol{a}$.

THEOREM 3:

$$\text{(i)} \quad E(\mathbf{f} \,|\, \mathbf{Y}) = \mathbf{T} \cdot \left\{ E(r\boldsymbol{\beta}_{\mathbf{c}} \,|\, \mathbf{Y}) + E\left((1-r) \sum_{i=1}^m \boldsymbol{c}_i \,|\, \mathbf{Y} \right) \right\} + \mathbf{B} \cdot E(\boldsymbol{a} \,|\, \mathbf{Y}); \tag{13}$$

$$\text{(ii)} \quad \mathrm{Cov}(\mathbf{f} \,|\, \mathbf{Y}) = (\mathbf{T}\ \mathbf{B})\, \mathrm{Cov}\left(\begin{pmatrix} \boldsymbol{\mu}_{G_\star,\mathbf{c}} \\ \boldsymbol{a} \end{pmatrix} \Big| \mathbf{Y} \right) \begin{pmatrix} \mathbf{T}^T \\ \mathbf{B}^T \end{pmatrix} + \mathbf{T} \cdot E\left(\frac{\mathbf{Cov}_{G_\star,\mathbf{c}}}{m + M + 1} \Big| \mathbf{Y} \right) \cdot \mathbf{T}^T. \tag{14}$$

The proof of Theorem 3 is given in Web Appendix A.4. These results show that the posterior mean and covariance matrix of $\mathbf{f}$ can be evaluated using the posterior samples of $(\boldsymbol{c}_i, \boldsymbol{\beta}_{\mathrm{c}}, \boldsymbol{D}, \boldsymbol{a}, M)$. The posterior CI for each component of $\mathbf{f}$ is obtained using a normal approximation.

5. An Application to the Progesterone Data

The progesterone dataset contained a total of 492 observations from 34 healthy women with each woman contributing from 11 to 28 observations over time. The menstrual cycle lengths of these women ranged from 23 to 56 days, with an average of 29.6 days. Each woman's menstrual cycle length was standardized uniformly to a reference 28-day cycle (Sowers et al., 1998). There were 98 distinct standardized time points. Figure 1a shows the data.

Denote by Y_{ij} the jth log-transformed progesterone value measured at standardized day t_{ij} since menstruation for the ith woman, and by AGE_i and BMI_i her age and BMI, respectively. We centered the t_{ij}'s at the average of all distinct time points and scaled by 10, i.e., scaled to a range of 2.8 units. We consider the following SPMM:

$$Y_{ij} = \beta_1 AGE_i + \beta_2 BMI_i + f(t_{ij}) + b_i + \epsilon_{ij}, \tag{15}$$

where $f(t)$ is a smooth function, b_i are i.i.d. random intercepts with mean zero, and ϵ_{ij} are independent residuals following a $N(0, \sigma^2)$ distribution. We centered age and BMI at the medians 36 years and 26 kg/m^2 and divided by 100. The curve $f(t)$ hence represents the progesterone profile for the population of women with age 36 years and BMI 26 kg/m^2.

We first fit the SPMM (15) by assuming a $N(0, \theta)$ distribution for b_i. We used independent $N(0, 10^4)$ priors for all fixed effect parameters in the LMM representation of equation (15) and an $IG(10^{-2}, 10^{-2})$ prior for σ^2. A USP was used for (θ, τ), where τ is the smoothing parameter for $f(t)$. The specification reflects a lack of prior information from the investigators on the model parameters, in particular, features of the progesterone profile. Some implementation detail can be found in Web Appendix A.6. Figure 1a plots the posterior mean of $f(t)$ and the pointwise 95% CIs. For comparison, we also evaluated the corresponding estimates using independent

Table 1

Estimates of the regression coefficients and variance components for the progesterone data. DPP-ADJ – DP prior with center adjustment; DPP-UN – traditional DP prior without adjustment. IGP: IG prior.

		Normal prior		DPP-ADJ $M = 0.75$		DPP-UN $M = 0.75$		DPP-ADJ $M \sim G(0.5, 1)$		DPP-UN $M \sim G(0.5, 1)$	
Par	Prior of var	Post mean	95% CI	Post mean	95% CI	Post mean	95% CI	Post mean	95% CI	Post mean	95% CI
β_1	USP	1.66	(−2.04, 5.34)	2.57	(0.57, 4.39)	2.62	(0.60, 4.43)	2.58	(0.32, 4.57)	2.59	(0.31, 4.59)
	IGP	1.618	(−2.24, 5.33)	2.78	(0.68, 4.60)	2.86	(0.72, 4.61)	2.70	(0.44, 4.69)	2.73	(0.41, 4.73)
β_2	USP	−2.67	(−7.21, 1.79)	−5.16	(−9.75, 0.75)	−5.19	(−9.72, 0.60)	−4.74	(−9.43, 1.33)	−4.77	(−9.56, 1.24)
	IGP	−2.75	(−7.49, 2.09)	−4.43	(−9.65, 1.15)	−4.36	(−9.67, 1.40)	−4.45	(−9.50, 1.86)	−4.37	(−9.43, 1.76)
σ^2	USP	0.34	(0.30, 0.38)	0.34	(0.30, 0.39)	0.34	(0.30, 0.39)	0.34	(0.29, 0.38)	0.34	(0.29, 0.38)
	IGP	0.34	(0.29, 0.38)	0.34	(0.30, 0.39)	0.34	(0.30, 0.39)	0.34	(0.29, 0.38)	0.34	(0.30, 0.38)
θ	USP	0.25	(0.14, 0.42)	0.25	(0.14, 0.39)		NA	0.25	(0.15, 0.41)		NA
	IGP	0.27	(0.15, 0.47)	0.27	(0.02, 1.16)		NA	0.27	(0.11, 0.56)		NA

$IG(10^{-2}, 10^{-2})$ priors for θ and τ. The fitted curve and the 95% CIs remain virtually unchanged (not shown). The estimates of the regression coefficients and variance components are presented in Table 1.

The posterior means of the random intercepts under both the USP and IG priors are clearly bimodal with peaks around −0.35 and 0.5 (Figure 1c), suggesting nonnormality of the random intercept distribution. We therefore fit the DP SPMM (15) assuming a nonparametric distribution for the random intercepts b_i with a hierarchically centered $DP(M, N(\beta_b, \theta))$ prior. Accordingly, $f(t)$ in equation (15) is replaced by $f_c(t)$, a centered version, i.e., with μ_G subtracted. We considered two choices for the total mass parameter, $M = 0.75$ and $M \sim G(0.5, 1)$, a gamma prior with $E(M) = Var(M) = 0.5$. We assign a $N(0, 10^4)$ prior for β_b and a USP for (θ, τ). For comparison, we also carried out inference under $\pi(\theta, \tau) = \pi(\theta)\pi(\tau)$ with $\pi(\theta) = \pi(\tau) = IG(10^{-2}, 10^{-2})$.

To compare with inference assuming normal random effects, we report inference on $f(t) \equiv \mu_G + f_c(t)$. The estimate of $f(t)$ and corresponding 95% pointwise CIs are plotted in Figure 1b (solid lines) with an $IG(0.5, 1)$ prior for M and a USP for (θ, τ). The plots under the IG priors for (θ, τ), and for fixed $M(=0.75)$ and either a USP or IG priors for (θ, τ) (none shown) are practically indistinguishable. Table 1 reports fixed effect and variance component estimates under these models.

These results suggest that the posterior mean of $f(t)$ is not greatly affected by the distributional assumption on the random intercepts, as expected. The posterior means of $f(t)$ are similar and the CIs using the center-adjusted DP prior are slightly shorter than those using a normal prior. The progesterone level remains relatively low and stable in the first half of a menstrual cycle and increases markedly after ovulation. It reaches a peak around reference day 23 and then decreases. Except for β_1 and β_2, inference for other parameters are essentially the same across different models. Remarkable changes were observed in the estimate of β_1. In contrast to the parametric model, the results under the DP model suggest a significant age effect on the log progesterone.

To compare our results with those using Zhang et al.'s (1998) methods, we implemented their (frequentist) approach without including the stochastic process term in their model for a fair comparison. Their results (not shown) are very similar to those using our Bayesian approach assuming normal random intercepts and a USP. Compared to their results, ours using the center-adjusted DP prior and the USP are distinct in two aspects. First, we found a significant age effect on log progesterone. The biological meaning of this finding is yet to be investigated. We noticed that the 95% CIs for the fixed effects are generally shorter than the frequentist confidence intervals. Second, although not large, there is an appreciable difference in inference for the random intercept variance and smoothing parameter. The 95% CI lengths are smaller for these parameters than their frequentist counterparts. As a result, the average 95% CI lengths for the nonparametric function estimates are slightly smaller. Although these differences are small in our data, when the number of distinct time points is small or the number of subjects is small, larger differences are expected in estimating the nonparametric function, random intercept variance, and smoothing parameter. We note that the comparison of posterior credible versus frequentist confidence interval estimates should not be overinterpreted. However, a reasonably constructed Bayesian CI may be expected to possess desirable frequentist properties, such as those considered in our simulation studies. A thorough discussion of this comparison is beyond the scope of this article.

For comparison, we also present in Table 1 the parameter estimates obtained under the DP prior without center adjustments. Inference for the regression coefficients and residual variance is similar across all DP models. Figure 1b (dashed lines) plots the pointwise posterior means of $f(t)$ and their 95% pointwise CIs for fixed $M(=0.75)$ and IG priors for θ and τ using the conventional DP prior without center adjustments. Compared to the solid lines, the posterior means are shifted slightly downward and the CIs are dramatically wide. For other prior combinations, i.e., fixed versus random M and USP versus IG priors, the posterior means were shifted upward and the CIs were similarly wide. Specifically, for fixed M + USP, the average $\widehat{f(t)} = 1.10$ and the average CI length = 1.49; for random M + USP, the average $\widehat{f(t)} = 1.06$ and average CI length = 0.99; and for random M + IG priors, the

Table 2
Comparison between inference assuming normal prior and DP prior with center adjustment (DPP-ADJ) for the random intercept distribution: simulation results based on 200 *replicates under the model with age and BMI included with* $m = 34$

Par	Prior for var	Sim. truth: R.e. dist.:	$b_i \sim N(0, 0.2303)$ Normal				$b_i \sim 11/18\ N(-0.35, 0.03) + 7/18\ N(0.55, 0.05)$ Normal				
		RB	MSE	SE	CIL	CP	RB	MSE	SE	CIL	CP
β_1	USP	0.01	3.61	0.16	7.02	0.93	−0.05	3.09	0.29	7.02	0.95
	IGP	−0.01	3.52	0.17	7.41	0.96	−0.0007	2.87	0.33	7.35	0.96
β_2	USP	−0.02	4.97	0.22	8.64	0.95	0.03	5.42	0.55	8.64	0.92
	IGP	−0.002	4.86	0.23	9.13	0.95	0.05	5.04	0.45	9.06	0.97
σ^2	USP	0.01	0.0006	0.00003	0.09	0.94	0.001	0.0005	0.00005	0.09	0.94
	IGP	0.002	0.0005	0.00002	0.09	0.95	0.001	0.0005	0.00005	0.09	0.96
θ	USP	−0.02	0.004	0.0002	0.25	0.92	−0.03	0.002	0.0002	0.25	0.96
	IGP	0.09	0.01	0.0003	0.29	0.95	0.07	0.003	0.0004	0.29	0.98
	Prior for var	$b_i \sim 11/18\ N(-0.35, 0.03) + 7/18\ N(0.55, 0.05)$ DPP-ADJ ($M = 0.75$)					$b_i \sim 11/18\ N(-0.35, 0.03) + 7/18\ N(0.55, 0.05)$ DPP-ADJ ($M \sim G(0.5, 1)$)				
β_1	USP	−0.02	1.97	0.25	4.51	0.88	0.03	1.98	0.21	4.85	0.90
	IGP	0.01	1.86	0.21	4.43	0.89	0.03	2.10	0.24	5.08	0.91
β_2	USP	0.02	3.08	0.32	5.75	0.90	0.01	3.13	0.30	6.20	0.92
	IGP	−0.01	3.04	0.36	5.59	0.91	0.04	20.91	0.36	6.53	0.95
σ^2	USP	0.02	0.0007	0.00007	0.09	0.93	0.01	0.0005	0.00004	0.09	0.96
	IGP	0.02	0.0006	0.00006	0.09	0.93	0.004	0.0004	0.00004	0.09	0.97
θ	USP	−0.07	0.002	0.0002	0.20	0.95	−0.05	0.002	0.0003	0.21	0.97
	IGP	0.37	0.08	0.03	1.45	1.00	0.28	0.01	0.0007	0.43	0.98

average $\widehat{f(t)} = 1.07$ and average CI length = 1.12. In comparison, the average posterior mean was 0.99 and the average CI length was 0.45 under the centered DP prior with adjustments and a gamma prior for M.

6. Simulation Studies

We carried out extensive simulation studies to investigate the (frequentist) performance of the proposed inference approach. In all simulations we set up a simulation truth as in model (1) (or model (1) with both age and BMI excluded) to mimic the design in the progesterone study. Unless otherwise stated we used a moderate sample size of $m = 34$ in the simulations. More details can be found in Web Appendix A.7. Some results are summarized in Tables 2 and 3. We report the relative biases (RBs), mean squared errors (MSEs), standard errors of the MSEs (SEs), 95% CI lengths (CILs), and the 95% coverage probabilities (CPs). Here RB is defined as the ratio of bias and the absolute value of the nonzero true parameter. The simulations with 200 replicates, 21,000 iterations with 1,000 burn-in in each replicate, took on average about 15 hours to run.

First we used a simulation truth model as in equation (1) with a normal random effect distribution, $b_i \sim N(0, \theta)$, two fixed effects, (β_1, β_2) and $f(\cdot)$ like in the progesterone study. The simulation results indicate that the proposed inference performs well assuming normal random effects and under both the USPs and IG priors for the variance components and the smoothing parameter. Some summaries are shown in Table 2.

Next we considered a scenario with a bimodal normal mixture as the true random effect distribution. The results show unbiased estimates of all model parameters under both the USP and IG priors for the variance components. However, the MSEs for the estimated β_1 and β_2 assuming normal random effects are significantly larger than those using the center-adjusted inference under the DP priors. The 95% CI lengths assuming normal priors are significantly larger, although their CPs are closer to the nominal 95%. Some summaries are shown in Table 3.

For comparison with a traditional DP prior, we carried out inference under the same model assuming a DP prior without center adjustments. We find markedly biased estimates, wide CIs, and overly high CPs for $f(t)$. See Figure 2 in the Web Appendix. The estimated random effect distribution, $E(G \mid Y)$, which included the corresponding fixed effect as its mean, is plotted in Figure 2 in the article. Details on how the random effect distribution was estimated are given in Web Appendix A.8. Also shown in Figure 2 is $E(G \mid Y)$ under a normality assumption for the random effect distribution. It is symmetric, as expected. In comparison, the estimates using the centered-adjusted DP prior capture the bimodality of the true distribution. The estimates are reasonable considering the latent nature of the random effects, the relatively small sample size ($m = 34$) and the discrete nature of the DP prior.

To investigate performance under smaller sample sizes, we reduced the number of subjects to $m = 10$. The simulation results are shown in Table 3. We find that the IG prior leads to considerably biased inference for the random effect variance θ. Similar results were obtained when we increased the simulation truth of the variance of the random intercepts. These results suggest that the USP provides a robust alternative to the IG priors for inference on variance components, especially when the number of subjects is small.

Finally, we performed more simulations with a bimodal random effect distribution and conducted center-adjusted

Table 3

Comparison between the performance of USP and IG prior (IGP): simulation results based on 200 *replicates under the model with age and BMI excluded. DPP-ADJ – DP prior with center adjustment. All assessments regarding $f(t)$ are averages across all distinct time points.*

		$b_i \sim N(0,\ 0.2303)$ Normal prior $m = 10$ $\theta = 0.2303$					$b_i \sim N(0,\ 11.2847)$ Normal prior $m = 10$ $\theta = 11.2847$					$b_i \sim \frac{11}{18}N(-0.35, 0.03) + \frac{7}{18}N(0.55, 0.05)$ DPP-ADJ ($M = 0.75$) $m = 34$ $\theta = 0.2303$				
Par	Prior for var	RB	MSE	SE	CIL	CP	RB	MSE	SE	CIL	CP	RB	MSE	SE	CIL	CP
σ^2	USP	0.01	0.002	0.0002	0.17	0.94	0.01	0.002	0.0002	0.16	0.96	0.02	0.0005	0.00005	0.09	0.95
	IGP	0.02	0.002	0.0002	0.17	0.92	0.02	0.002	0.0002	0.17	0.92	0.02	0.0006	0.00006	0.09	0.95
θ	USP	−0.01	0.02	0.002	0.49	0.89	−0.02	28.05	2.72	21.42	0.89	−0.08	0.002	0.0002	0.19	0.91
	IGP	0.28	0.02	0.003	0.72	0.97	0.27	45.74	4.94	31.65	0.95	0.61	0.62	0.49	1.55	1.00
$f(\cdot)$	USP	0.004	0.03	0.003	0.72	0.93	−0.03	1.22	0.11	3.75	0.88	−0.02	0.01	0.001	0.43	0.96
	IGP	0.01	0.03	0.003	0.79	0.96	0.02	1.36	0.12	4.22	0.88	−0.002	0.01	0.001	0.49	0.97

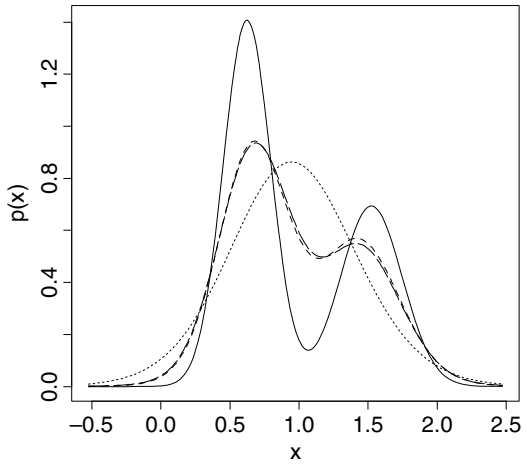


Figure 2. The densities of the posterior means of the random intercept distributions in the model with age and BMI included and the true random intercept distribution being $11/18 \times N(-0.35,\ 0.03) + 7/18 \times N(0.55,\ 0.05)$. ·····: normal prior for the random intercept distribution – – –: DP prior with center adjustment with fixed $M(= 0.75)$ — —: DP prior with center adjustment with a gamma prior for M ($\sim G(0.5,\ 1)$). USPs are used for the random intercept variance and smoothing parameter. The solid curve (——) is the true density of the random intercept shifted 0.9737 unit to the right. This shift corresponds to the true average values of $f(t)$ across all distinct time points. Results are based on 200 replicates. The corresponding estimated densities under the inverse gamma priors for θ and τ are practically indistinguishable (not shown).

inference with a DP prior. We compared the IG prior versus the USP for the variance of the base measure. The results are summarized in Table 3. The IG prior leads to seriously biased estimate of the random effect variance θ. The problem persists even when the sample size m was doubled to 68 (results not shown), suggesting difficulties with the use of the diffuse IG priors for variance components. More simulation results can be found in the Web Appendix.

7. Discussion

We proposed a framework for Bayesian inference in SPMMs. We addressed two important gaps in the literature, a default prior for the variance components in such models, and adjustment of inference for fixed effects that are linked with nonparametrically modeled random effects. The latter includes inference for a nonparametric mean function over time.

The proposed approach is fully model based, and thus enjoys all the advantages of such methods. In particular, posterior inference on any well defined event of interest is meaningful and possible. Uncertainties can be characterized by appropriate summaries of the posterior distribution. The proposed methods allow unequally spaced time points in estimating the nonparametric function. When data are missing completely at random or missing at random, due to the model-based nature, the proposed inference procedures are still valid.

Some limitations remain. For example, in the application example, we assumed a SPMM with random intercept for the log progesterone level, with the assumption that the individual progesterone profiles differ by a vertical shift. A possible extension of the model would be to model individual departure from the mean curve as another spline function. On the other hand, an informal inspection of the residual curves in the progesterone example does not show important violations of the current assumptions. The periodicity of the curve estimates, when desired, could be incorporated in our model. Our current model does not impose this constraint. This does not seem to be a problem, as the posterior log progesterone profile appears to already satisfy (approximately) the periodicity constraint.

We focus on estimation of a single nonparametric function in this article. But the methods can be generalized to generalized additive mixed models in the presence of multiple nonparametric additive covariate effects and non-Gaussian outcomes (Lin and Zhang, 1999).

8. Supplementary Materials

The Web Appendices are downloadable from `http://www.biometrics.tibs.org`. The `C++` code that was used for our implementation is available for interested readers upon request.

Acknowledgements

Drs YL and XL's research is supported by the NCI grant R37 CA76404-12. We thank the editor, associate editor, and two reviewers for their constructive comments that led to an improved manuscript.

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.

Christopher, A. B. and MacEachern, S. N. (1996). A semiparametric Bayesian model for randomised block designs. *Biometrika* **83,** 275–285.

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

Fan, J. and Li, R. (2004). New estimation and model selection procedures for semiparametric modeling in longitudinal data analysis. *Journal of the American Statistical Association* **99,** 710–723.

Gelman, A. (2006). Prior distributions for variance parameters in hierarchical models. *Bayesian Analysis* **1,** 515–533.

Kleinman, K. P. and Ibrahim, J. G. (1998). A semiparametric Bayesian approach to the random effects model. *Biometrics* **54,** 921–938.

Laird, N. M. and Ware, J. H. (1982). Random effects models for longitudinal data. *Biometrics* **38,** 963–974.

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, Texas.

Lin, X. and Carroll, R. J. (2001). Semiparametric regression for clustered data. *Biometrika* **88,** 1179–1185.

Mukhopadhyay, S. and Gelfand, A. (1997). Dirichlet process mixed generalized linear models. *Journal of the American Statistical Association* **92,** 633–639.

Müller, P. and Rosner, G. (1997). A Bayesian population model with hierarchical mixture priors applied to blood count data. *Journal of the American Statistical Association* **92,** 1279–1292.

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

Natarajan, R. and McCulloch, C. E. (1998). Gibbs sampling with diffuse proper priors: A valid approach to data-driven inference? *Journal of Computational and Graphical Statistics* **7,** 267–277.

Pennell, M. L. and Dunson, D. B. (2007). Fitting semiparametric random effects models to large data sets. *Biostatistics* **8,** 821–834.

Ruppert, D., Wand, M. P., and Carroll, R. J. (2003). *Semiparametric Regression*. New York: Cambridge University Press.

Searle, S. R. (1982). *Matrix Algebra Useful for Statistics*. New York: Wiley.

Sowers, M. F., Crutchfield, M., Shapiro, B., Zhang, B., La Pietra, M., Randolph, J. F., and Schork, M. A. (1998). Urinary ovarian and gonadotrophin hormone levels in premenopausal women with low bone mass. *Journal of Bone and Mineral Research* **13,** 1191–1202.

Verbeke, G. and Lesaffre, E. (1996). A linear mixed-effects model with heterogeneity in the random-effects population. *Journal of the American Statistical Association* **433,** 217–221.

Verbela, A. P., Cullis, B. R., Kenward, M. G., and Welham, S. J. (1999). The analysis of designed experiments and longitudinal data by using smoothing splines. *Journal of the Royal Statistical Society, Series C* **48,** 269–300.

Wahba, G. (1978). Improper priors, spline smoothing and the problem of guarding against model errors in regression. *Journal of the Royal Statistical Society, Series B* **40,** 364–372.

Wang, N., Carroll, R. J., and Lin, X. (2005). Efficient semiparametric marginal estimation for longitudinal/clustered data. *Journal of the American Statistical Association* **100,** 147–157.

Zeger, S. L. and Diggle, P. J. (1994). Semi-parametric models for longitudinal data with application to CD4 cell numbers in HIV seroconverters. *Biometrics* **8,** 81–89.

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

Received March 2008. *Revised November* 2008.
Accepted November 2008.