Ann Inst Stat Math (2008) 60:499–517
DOI 10.1007/s10463-007-0127-3

Bayesian hierarchical linear mixed models for additive smoothing splines

Dongchu Sun · Paul L. Speckman

Received: 12 September 2005 / Revised: 8 November 2006 / Published online: 19 April 2007

Abstract Bayesian hierarchical models have been used for smoothing splines, thin-plate splines, and L-splines. In analyzing high dimensional data sets, additive models and backfitting methods are often used. A full Bayesian analysis for such models may include a large number of random effects, many of which are not intuitive, so researchers typically use noninformative improper or nearly improper priors. We investigate propriety of the posterior for these cases. Our findings extend known results for normal linear mixed models to certain cases with Bayesian additive smoothing spline models.

Keywords Generalized linear mixed models · Gibbs sampling · Linear mixed models · Markov chain Monte Carlo · Multivariate normal · Variance components

1 Introduction

There is a large literature on the use of improper priors with Gaussian linear mixed models. Beginning with Hobert and Casella (1996), a number of authors have shown how the use of improper priors can lead to improper posterior distributions, with consequent disastrous performance in Markov chain Monte Carlo simulations. The implication is that nearly improper priors can also give misleading results. Speckman and Sun (2003) considered a special case motivated by nonparametric function estimation

Supported by National Science Foundation grant SES-0351523 and by National Institutes of Health grants R01-CA100760 and R01-MH071418.

D. Sun (✉) · P. L. Speckman
Department of Statistics, University of Missouri-Columbia, Columbia, MO 65211, USA
e-mail: sund@missouri.edu

P. L. Speckman
e-mail: speckmanp@missouri.edu

using Bayesian smoothing splines. Among all inverse gamma type prior distributions on the variance components, they characterized the ones with proper posterior distributions. The purpose of this paper is to extend the results to additive regression models.

Consider regression on a function of r variables, $y = f(t_1, \ldots, t_r) + e$. In high dimensions, the problem of estimating f without assuming a parametric form suffers from the "curse of dimensionality," a practical problem that has led to a large number of methods. One common assumption, introduced by Stone (1985) and implemented by Hastie and Tibshirani (1990) among others, is that the regression function is additive, e.g.,

$$f(t_1, \ldots, t_r) = \beta_0 + \sum_{k=1}^{r} f_k(t_k), \tag{1}$$

where for identifiability f_k is orthogonal to the space of constants for all k. In fact, there is no need to restrict the variables t_k to be univariate; $f_k(t_k)$ itself may be a function of several variables. Variations on this model include tensor product splines (Wahba 1990).

A convenient way to put priors on the functions in model (1) is via Wahba's (1983) Gaussian prior equivalent to spline smoothing. Consider the single component nonparametric regression model

$$y_i = f(t_i) + e_i, \quad i = 1, \ldots, n.$$

One popular estimate of f is the smoothing spline, defined as the solution to the optimization problem

$$f = \arg\min_g \left[\sum_{i=1}^{n} \{y_i - g(t_i)\}^2 + \eta \int \{g^{(m)}(t)\}^2 \, \mathrm{d}t \right],$$

for an appropriate smoothing parameter η and order m (see, e.g., Wahba 1990; Green and Silverman 1994; Eubank 1999). The smoothing spline is intuitively appealing because it balances fidelity to the data as measured by squared error with roughness of the fit in terms of the L_2 norm on $g^{(m)}$. The solution is a natural polynomial spline of order $2m$. In particular, the case $m = 2$ yields a cubic smoothing spline. Wahba showed that the cubic smoothing spline arises as the limit of posterior means for suitable Gaussian process priors on the function $f(t)$. Fully Bayesian versions have been implemented by Carter and Kohn (1994, 1996), and Nychka (2000) has a nice account of the connection between spatial smoothing and nonparametric regression.

Wahba's limiting prior on $\boldsymbol{f} = (f(t_1), \ldots, f(t_n))'$ can be shown to have the form

$$[\boldsymbol{f} \mid \delta_1] \propto \delta_1^{-(n-m)/2} \mathrm{e}^{-\boldsymbol{f}'\boldsymbol{A}\boldsymbol{f}/(2\delta_1)}, \tag{2}$$

where $\boldsymbol{A}$ is a positive semidefinite matrix of rank $n - m$ depending on $t_1, \ldots, t_n$ and m, and δ_1 is a variance component that must be specified, estimated or given a prior distribution. A derivation of representation (2) is given in Speckman and Sun (2003), for example. The null space of $\boldsymbol{A}$ is exactly the linear subspace generated by polynomials of degree at most $m-1$, i.e., the subspace spanned by the vectors $(t_1^j, \dots, t_n^j)'$, $j = 0, \dots, m-1$. This prior is "partially informative" in the sense that the prior is improper (and noninformative) on the null space of $\boldsymbol{A}$ and a proper Gaussian prior on the range space of $\boldsymbol{A}$. Specifically, suppose $\boldsymbol{A}$ has spectral decomposition $\boldsymbol{A} = \boldsymbol{\Gamma}\boldsymbol{\Lambda}\boldsymbol{\Gamma}'$ for $\boldsymbol{\Lambda} = \operatorname{diag}(0, \dots, 0, \lambda_{m+1}, \dots, \lambda_n)$ and $\boldsymbol{\Gamma}'\boldsymbol{\Gamma} = \boldsymbol{I}_n$. Let $\boldsymbol{\Gamma} = (\boldsymbol{T}, \boldsymbol{X})$, where $\boldsymbol{T}$ is an $n \times m$ matrix spanning the space of polynomials of degree $m-1$, and $\boldsymbol{X}$ is an $n \times (n-m)$ matrix orthogonal to $\boldsymbol{T}$. Let $\boldsymbol{\theta} = (\theta_1, \dots, \theta_m)'$ have constant prior, and let $\boldsymbol{u} \sim \mathrm{N}_{n-m}(\mathbf{0}, \delta_1 \operatorname{diag}(\lambda_{m+1}, \dots, \lambda_n)^{-1})$. Then prior (2) is equivalent to $\boldsymbol{f} = \boldsymbol{T}\boldsymbol{\theta} + \boldsymbol{X}\boldsymbol{u}$. Speckman and Sun (2003) termed this a partially improper normal distribution with precision matrix $\delta_1^{-1}\boldsymbol{A}$, and denoted it by $\boldsymbol{f} \sim \mathrm{PIN}(\mathbf{0}, \delta_1^{-1}\boldsymbol{A})$. If the error terms e_i are independent of $\boldsymbol{f}$ and iid $\mathrm{N}(0, \delta_0)$, the posterior of $\boldsymbol{f}$ is easily seen to be $\mathrm{N}_n(\boldsymbol{S}_\eta \boldsymbol{y}, \delta_0 \boldsymbol{S}_\eta)$, where $\eta = \delta_0/\delta_1$ and $\boldsymbol{S}_\eta = (\boldsymbol{I} + \eta\boldsymbol{A})^{-1}$. (In the smoothing literature, $\boldsymbol{S}_\eta$ is the smoother matrix.)

The complete Bayesian specification requires prior distributions on δ_0 and δ_1. While it may be possible to elicit prior information on δ_0, the second variance component δ_1 appears to be more difficult, and consideration of noninformative priors seems natural. Within the class of inverse gamma-type priors

$$[\delta_k \mid a_k, b_k] \propto \frac{1}{\delta_k^{a_k+1}} \mathrm{e}^{-b_k/\delta_k}, \tag{3}$$

$k = 0, 1$, Speckman and Sun (2003) were able to characterize the conditions under which the posterior is proper. (Here and in the rest of the paper, we follow the Bayesian convention where $[\delta_k \mid a_k, b_k]$ denotes the density of δ_k given a_k and b_k.) This model is closely related to the mixed models of Hobert and Casella (1996) and Sun et al. (2001), but the fact that $\boldsymbol{A}$ has rank nearly equal to n necessitates new proofs. The results of Speckman and Sun also pertain to CAR (Besag 1974) and IAR (Besag and Kooperberg 1995) priors commonly used for discrete spatial models as well as multidimensional thin-plate smoothing spline priors. One nice aspect of the priors associated with spline smoothing is that they extend naturally to $f(t)$ for arbitrary unobserved t (Nychka 2000), so posterior inference involving $f(t)$ at points not in the data set is easy to obtain. In this respect, the smoothing spline priors are comparable to some of the priors used in kriging unlike the CAR prior.

This class of Gaussian priors on the function f is easily extended to the additive model

$$y_i = \beta_0 + \sum_{k=1}^{r} f_k(t_{ik}) + e_i, \quad i = 1, \dots, n, \tag{4}$$

with iid normal errors e_i, adding the identifiability assumption $\sum_{i=1}^n f_k(t_{ik}) = 0$. Let $\boldsymbol{f}_k = (f(t_{1k}), \dots, f(t_{nk}))'$, $k = 1, \dots, r$. Since $\boldsymbol{f}_k \perp \mathbf{1}$ with $\mathbf{1} = (1, \dots, 1)'$, we assume independent singular priors with improper densities of the form

$$[\boldsymbol{f}_k \mid \delta_k] \propto \begin{cases} \delta_k^{-(n-m_k)/2} \mathrm{e}^{-\boldsymbol{f}_k' \boldsymbol{A}_k \boldsymbol{f}_k/(2\delta_k)}, & \boldsymbol{f}_k'\mathbf{1} = 0, \\ 0, & \boldsymbol{f}_k'\mathbf{1} \neq 0, \end{cases} \tag{5}$$

$k = 1, \ldots, r$, where $\boldsymbol{A}_k$ is positive semidefinite with rank $q_k = n - m_k$. The restriction $\sum_{i=1}^n f_k(t_{ik}) = 0$, $k = 1, \ldots, r$, can be thought of as a projection onto the space orthogonal to constants of the PIN$(\mathbf{0}, \delta_k^{-1}\boldsymbol{A}_k)$ distribution. An explicit representation is given in (9) below. By analogy, if the predictor variables $t_1, \ldots, t_r$ are all one dimensional, this specification of priors corresponds to estimating model (1) by solving the variational problem

$$\min_{\beta_0, g_1, \ldots, g_r} \left[\sum_{i=1}^n \left\{ y_i - \beta_0 - g_1(t_{i1}) - \cdots - g_r(t_{ir}) \right\}^2 + \sum_{k=1}^r \eta_k \int \left\{ g_k^{(m_k)}(t_k) \right\}^2 \mathrm{d}t_k \right]$$

with $\eta_k = \delta_0/\delta_k$ for $k = 1, \ldots, r$. This model is commonly fit by the backfitting algorithm (Hastie and Tibshirani 1990). The first Bayesian analysis of the additive case of spline smoothing was given by Wong and Kohn (1996). Subsequent treatments include Shively et al. (1999), Hastie and Tibshirani (2000), Fahrmeir and Lang (2001), and Wood et al. (2002). These and other authors commonly use diffuse priors on the variance components δ_k, $k = 0, \ldots, r$. However, the question of propriety of the posterior under improper priors remains.

To precisely specify the prior (5) corresponding to additive model (4), it is necessary to have an identifiability assumption. Clearly the model is not identifiable if there is collinearity among the explanatory variables t_{ij}. The model is also not identifiable if there is a polynomial dependency among the explanatory variables ("concurvity"). These notions can be made precise with the following model representation. Let $\boldsymbol{T}_k$ be an $n \times (m_k - 1)$ matrix such that $\boldsymbol{T}_k'\mathbf{1} = \mathbf{0}$ and $(\mathbf{1}, \boldsymbol{T}_k)$ spans the null space of $\boldsymbol{A}_k$. We assume that the model is identifiable in the sense that the matrix $\boldsymbol{X}_0 = (\mathbf{1}, \boldsymbol{T}_1, \ldots, \boldsymbol{T}_r)$ has full rank

$$p = r(\boldsymbol{X}_0) = 1 + \sum_{k=1}^r (m_k - 1). \tag{6}$$

In addition, throughout the rest of the paper, we assume that the data are not fit perfectly by the implicit linear model of rank p, i.e., suppose

$$\boldsymbol{y}'(\boldsymbol{I}_n - \boldsymbol{X}_0(\boldsymbol{X}_0'\boldsymbol{X}_0)^{-1}\boldsymbol{X}_0')\boldsymbol{y} > 0.$$

Finally, we assume the constant prior on β_0,

$$[\beta_0] \equiv 1. \tag{7}$$

Combined with the implicit constant prior on the range of $\boldsymbol{T}_k$ for all k, this implies a constant prior on the range of $\boldsymbol{X}_0$.

While the mixed effects models here are closely related to models considered by Hobert and Casella (1996), Sun et al. (2001), and others, there are several important differences that make new proofs necessary. First, as customary in applications of additive models, we do not include a separate term for an "error sum of squares." In the applications we have in mind, additive models are not used for designed experiments and there are no repeated measurements. Secondly, the dimension of the random effect terms is large, and there typically are more than n random components. For example, with a simple additive model with two smooth terms and no fixed effects, the prior corresponding to cubic spline smoothing has $m_1 = m_2 = 2$. The random effect vectors for $\boldsymbol{f}_1$ and $\boldsymbol{f}_2$ both have $n-2$ components, so previous results designed for fixed length random effects do not apply. Note that the priors on $\boldsymbol{f}_1, \ldots, \boldsymbol{f}_r$ are always partially improper since a flat prior is used for the range of $\boldsymbol{X}_0$. The first two results treat propriety in terms of the prior on the smoothing parameters $\eta_k = \delta_0/\delta_k, k = 1, \ldots, r$.

Theorem 1 *Under the priors given by* (5) *on the* $\boldsymbol{f}_k$ *assuming the full rank condition* (6) *and the constant prior* (7) *on* β_0*, suppose the prior on* $\boldsymbol{\eta} = (\eta_1, \ldots, \eta_r)'$ *is proper. If the prior on* δ_0 *is proper and satisfies* $E\delta_0^{-(n-p)/2} < \infty$*, then the joint posterior distribution is proper.*

Remark 1 This result covers many interesting priors. For example, if any proper prior is used on $\boldsymbol{\eta}$, a lognormal or (proper) gamma or inverse gamma prior may be taken for δ_0.

This theorem is much weaker than necessary. The prior moment condition on δ_0 is sufficient for propriety but far from necessary. In fact, the prior on δ_0 need not even be proper. An important special case is the invariance prior, where we have a complete characterization.

Theorem 2 *With the priors given by* (5) *assuming* (6) *holds, the constant prior* (7) *on* β_0*, and the invariance prior* $[\delta_0] \propto \delta_0^{-1}$*, the joint posterior distribution is proper if and only if the prior on* $\boldsymbol{\eta} = (\eta_1, \ldots, \eta_r)'$ *is proper.*

Following Hobert and Casella (1996), our next results concern independent improper inverse gamma-type priors on the δ_k of the form

$$[\delta_0, \delta_1, \ldots, \delta_r \mid a_k, b_k, k = 0, \ldots, r] = \prod_{k=0}^{r} \frac{\mathrm{e}^{-b_k/\delta_k}}{\delta_k^{a_k+1}}, \tag{8}$$

where $a_k \in \mathbb{R}$ and $b_k \geq 0, k = 1, \ldots, r$. In the general case under priors on the δ_k of the form (8) with $b_k = 0$ for all k, we have the following characterization.

Theorem 3 *With priors given by* (5) *assuming* (6), (7)*, and* (8) *with* $b_0 = \cdots = b_r = 0$*, the joint posterior distribution is proper if and only if*

(a) $a_k < 0, k = 1, \ldots, r$.
(b) $n - p + 2\sum_{k=0}^{r} a_k > 0$.

The final result gives a sufficient condition for propriety when $b_k > 0$ for all k.

Theorem 4 *With priors given by* (5) *assuming* (6)*,* (7)*, and* (8) *with* $b_k > 0$ *for* $k = 0, \ldots, r$*, the joint posterior distribution is proper if*

$$n - p + 2\sum_{k=0}^{r} \min(0, a_k) > 0.$$

The paper is organized as follows. A small simulated example is presented in Sect. 2 illustrating the potential effects of running an MCMC algorithm with an improper posterior distribution. The model is developed further and proofs of the main results are presented in Sect. 3.

2 Simulation results

From the theory suggested in Theorem 1 and the proofs in the next section, the problem with improper priors of the form (3) with $b_k = 0$ is that there is a spurious mode in the posterior at 0 for the variance components δ_k. To illustrate the potential harmful effects of the wrong chioce of noninformative priors, we conducted a small simulation with $n = 30$ observations and two components. We sampled (t_1, t_2) from a bivariate normal distribution with means zero, variances set to one, and correlation 0.7. We chose $f_1(t_1) = \sin(2t_1)$ and $f_2(t_2) = t_2^2$. The observations were taken to be $y_i = f_1(t_{i1}) + f_2(t_{i2}) + e_i$ with independent, normally distributed error terms with mean zero and standard deviation 0.4. For the first MCMC chain, the priors for the variances were $[\delta_k] \propto \delta_k^{-1}$, i.e. $a_k = b_k = 0$, for $k = 0, 1, 2$. Part of the output from an MCMC run of 90,000 cycles following 10,000 burnin cycles is shown in the left panels of Fig. 1. Trace plots are shown for simulated draws from the posterior distribution of δ_k sampled once every 10 iterations, $k = 0, 1, 2$. Note that the simulation for δ_2 is trapped in the spurious mode at zero for roughly the first 50,000 cycles. (This behavior depends on the initial seed for the pseudo random number generator.) Trace plots from a second run of 100,000 cycles with $a_k = -0.5$, $k = 0, 1, 2$ are shown in the right panels of Fig. 1. These plots show no such instability.

We should point out that this simulation study agrees with the results in Lambert et al. (2005), who demonstrate the instability of the posterior estimates in linear mixed models with inverse gamma $(\varepsilon, \varepsilon)$ prior for variance components when ε is a small value such as 0.01, 0.001, etc.

3 Bayesian additive models

Consider the additive model for $\boldsymbol{y} = (y_1, \ldots, y_n)'$,

$$\boldsymbol{y} = \sum_{k=1}^{r} \boldsymbol{f}_k + \boldsymbol{e}, \quad \boldsymbol{e} \sim N_n(\boldsymbol{0}, \delta_0 \boldsymbol{W}^{-1}),$$

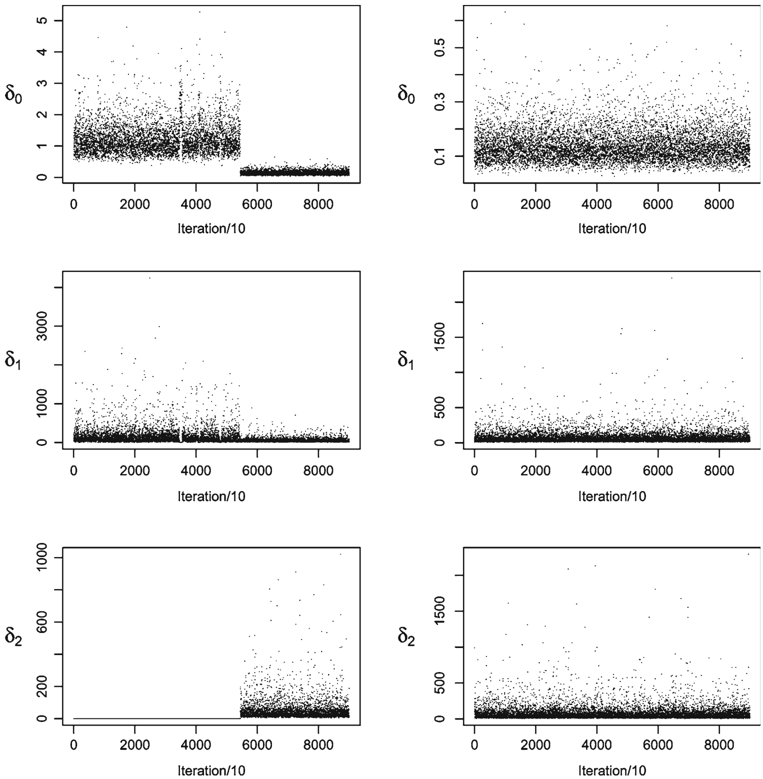

Fig. 1 *Left panels*: trace plots of δ_k from MCMC run of 90,000 cycles sampled every 20 iterations for simulated data described in text and improper priors $[\delta_k] \propto \delta_k^{-1}$, $k = 0, 1, 2$. *Right panels*: similar trace plots with priors $[\delta_k] \propto \delta_k^{-0.5}$, $k = 0, 1, 2$

where $\boldsymbol{W}$ is a known positive definite matrix and $\boldsymbol{f}_k \sim \mathrm{PIN}(\boldsymbol{0}, \delta_k^{-1}\boldsymbol{A}_k)$. As stated above, the model is typically not identifiable since each component $\boldsymbol{f}_k$ includes a constant term. However, with $\eta_k = \delta_0/\delta_k$, the full conditional distributions of the $\boldsymbol{f}_k$ are proper normal:

$$\begin{aligned}&(\boldsymbol{f}_k \mid \boldsymbol{y}, \delta_0, \boldsymbol{\eta}, \boldsymbol{f}_j, j \neq k)\\ &\quad \sim N\left\{(\boldsymbol{W} + \eta_k \boldsymbol{A}_k)^{-1}\boldsymbol{W}\left(\boldsymbol{y} - \sum_{j \neq k} \boldsymbol{f}_j\right), \delta_0(\boldsymbol{W} + \eta_k \boldsymbol{A}_k)^{-1}\right\}.\end{aligned}$$

Here $\boldsymbol{\eta} = (\eta_1, \ldots, \eta_r)$. To compensate for the lack of identifiability, Hastie and Tibshirani (2000) suggested projecting $\boldsymbol{f}_k$ onto the space orthogonal to constants after each MCMC iteration. This is standard practice and justifiable in the Gaussian case (e.g., Besag et al. 1995; Gelfand and Sahu 1999). Shively et al. (1999) and Hastie and Tibshirani (2000) have efficient algorithms for sampling from the full conditionals.

For our purposes, we need an alternative approach with an explicit representation for each of the partially informative terms, conditioning on the subspace orthogonal to constants. For each k, let $\boldsymbol{A}_k = \boldsymbol{X}_k \boldsymbol{\Lambda}_k \boldsymbol{X}_k'$, where $\boldsymbol{\Lambda}_k$ is diagonal, $\boldsymbol{\Lambda}_k > 0$, and $\boldsymbol{X}_k' \boldsymbol{X}_k = \boldsymbol{I}_{n-m_k}$. In addition, let $\boldsymbol{T}_k$ be an $n \times (m_k - 1)$ matrix that spans the null space of $\boldsymbol{A}_k$ and is orthogonal to the space of constants. Then

$$\boldsymbol{f}_k = \boldsymbol{1}_n \beta_k + \boldsymbol{T}_k \boldsymbol{\theta}_k + \boldsymbol{X}_k \boldsymbol{u}_k$$

is a unique representation. Assuming independence, the PIN prior implies that $\boldsymbol{\theta}_k$ has constant prior and

$$\boldsymbol{u}_k \overset{\text{indep}}{\sim} N_{n-m_k}(\boldsymbol{0}, \delta_k \boldsymbol{\Lambda}_k^{-1}). \tag{9}$$

Collecting constant terms with $\beta_0 = \sum_{k=1}^r \beta_k$, we have the model

$$\boldsymbol{y} = \boldsymbol{1}_n \beta_0 + \sum_{k=1}^r (\boldsymbol{T}_k \boldsymbol{\theta}_k + \boldsymbol{X}_k \boldsymbol{u}_k) + \boldsymbol{e}.$$

Without loss of generality, we assume that $\boldsymbol{W}$ equals $\boldsymbol{I}_n$, the n-dimensional identity matrix. The linear terms in the equation can be collected in a single full rank component of the form $\boldsymbol{1}_n \beta_0 + \sum_{k=1}^r \boldsymbol{T}_k \boldsymbol{\theta}_k = \boldsymbol{X}_0 \boldsymbol{\theta}$, resulting in the reduced model

$$\boldsymbol{y} = \boldsymbol{X}_0 \boldsymbol{\theta} + \sum_{k=1}^r \boldsymbol{X}_k \boldsymbol{u}_k + \boldsymbol{e}, \quad \boldsymbol{e} \sim N_n(\boldsymbol{0}, \delta_0 \boldsymbol{I}_n). \tag{10}$$

An interesting and broad class of models related to model (10) is Bayesian P-splines (see Lang and Brezger 2004). One essential difference between the models treated here and P-splines is that the rank of $\boldsymbol{X}_k$ is fixed at some dimension $r_k \ll n$ for P-splines, so results obtained here do not apply.

It is not hard to show that the full conditionals of $\boldsymbol{\theta}$ and the $\boldsymbol{u}_k$ are multivariate normal. For a fully Bayesian model, assume inverse-gamma-type priors (possibly improper) (3) on the variance components. Then the full conditional distributions of the δ_k are again inverse-gamma. These distributions form the basis of a straightforward MCMC algorithm. In our notation, Hastie and Tibshirani (2000) suggested priors of the form (3) with $a_0 = b_0 = 0$, $a_k = 1$ for $k > 1$, and b_k very small.

From a practical standpoint, the most difficult part of the implementation is choosing parameters of the hyperpriors for the δ_j. In his discussion of Hastie and Tibshirani's article, Gelfand questioned the uncritical use of nearly improper priors for the variance components. The results of Speckman and Sun (2003) for the one component case show that the limiting posterior under Hastie and Tibshirani's prior when $b_1 = 0$ is in fact not proper.

Recall that $\boldsymbol{X}_k$ has rank $q_k = n - m_k$ and $\eta_k = \delta_0/\delta_k$, $k = 1, \ldots, r$. In addition, define

$$\boldsymbol{u} = (\boldsymbol{u}_1', \ldots, \boldsymbol{u}_r')', \quad \boldsymbol{B}_{\boldsymbol{\eta}} = \operatorname{diag}(\eta_1^{-1} \boldsymbol{\Lambda}_1, \ldots, \eta_r^{-1} \boldsymbol{\Lambda}_r),$$

and $q = q_1 + \cdots + q_r$. Then (9) is equivalent to

$$(\boldsymbol{u} \mid \delta_0, \boldsymbol{\eta}) \sim N_q(\mathbf{0}, \delta_0 \boldsymbol{B}_{\boldsymbol{\eta}}). \tag{11}$$

Without loss of generality, we assume that $\boldsymbol{X}_k \boldsymbol{X}_k'$ is idempotent, that is, $(\boldsymbol{X}_k \boldsymbol{X}_k')^2 = \boldsymbol{X}_k \boldsymbol{X}_k'$. If not, take the singular value decomposition $\boldsymbol{X}_k \boldsymbol{\Lambda}_k^{1/2} = \boldsymbol{P}_k \boldsymbol{D}_k \boldsymbol{Q}_k'$, where $\boldsymbol{P}_k$ and $\boldsymbol{Q}_k$ are suborthogonal matrices and $\boldsymbol{D}_k$ is a positive definite diagonal matrix. Then (10) and (11) hold with $\boldsymbol{X}_k$ replaced by $\boldsymbol{P}_k$, $\boldsymbol{u}_k$ replaced by $\boldsymbol{D}_k \boldsymbol{Q}_k' \boldsymbol{\Lambda}_k^{-1/2} \boldsymbol{u}_k$, and $\boldsymbol{\Lambda}_k$ replaced by $\boldsymbol{D}_k^2$. Let $\boldsymbol{X} = (\boldsymbol{X}_1, \ldots, \boldsymbol{X}_r)$. We can rewrite (10) and (11) as follows:

$$\begin{aligned}
[\boldsymbol{y} \mid \boldsymbol{\theta}, \boldsymbol{u}, \delta_0] &= \frac{1}{(2\pi\delta_0)^{n/2}} \exp\left\{-\frac{1}{2\delta_0}(\boldsymbol{y} - \boldsymbol{X}_0\boldsymbol{\theta} - \boldsymbol{X}\boldsymbol{u})'(\boldsymbol{y} - \boldsymbol{X}_0\boldsymbol{\theta} - \boldsymbol{X}\boldsymbol{u})\right\}, \\
[\boldsymbol{u} \mid \delta_0, \boldsymbol{\eta}] &= \frac{1}{(2\pi\delta_0)^{q/2}|\boldsymbol{B}_{\boldsymbol{\eta}}|^{1/2}} \exp\left\{-\frac{1}{2\delta_0}\boldsymbol{u}'\boldsymbol{B}_{\boldsymbol{\eta}}^{-1}\boldsymbol{u}\right\}.
\end{aligned}$$

To be consistent with the prior specification in the introduction, we assume independent priors for $\boldsymbol{\theta}$ and δ_0 with the form

$$[\boldsymbol{\theta}] \propto 1, \tag{12}$$

$$[\delta_0] \propto \frac{1}{\delta_0^{a_0+1}} \mathrm{e}^{-b_0/\delta_0}. \tag{13}$$

A popular prior for δ_0 is the so-called "invariance prior" corresponding to $a_0 = b_0 = 0$. If there are no random effects $\boldsymbol{u}_k$, the likelihood is a scale and location family, and the invariance prior is a commonly used objective prior. See, for example, Berger (1985).

By assumption, $\boldsymbol{X}_0$ has rank p (full rank). Define

$$\begin{aligned}
\boldsymbol{H}_0 &= \boldsymbol{X}_0(\boldsymbol{X}_0'\boldsymbol{X}_0)^{-1}\boldsymbol{X}_0', \\
\boldsymbol{G}_{\boldsymbol{\eta}} &= \boldsymbol{B}_{\boldsymbol{\eta}}^{-1} + \boldsymbol{X}'(\boldsymbol{I}_n - \boldsymbol{H}_0)\boldsymbol{X}, \\
\boldsymbol{R}_{\boldsymbol{\eta}} &= \boldsymbol{I}_n - \boldsymbol{H}_0 - (\boldsymbol{I}_n - \boldsymbol{H}_0)\boldsymbol{X}\boldsymbol{G}_{\boldsymbol{\eta}}^{-1}\boldsymbol{X}'(\boldsymbol{I} - \boldsymbol{H}_0).
\end{aligned} \tag{14}$$

Lemma 1 *Consider the linear mixed model* (10) *and* (11).

(a) *Under prior* (12)*, the marginal likehood function of* $(\delta_0, \boldsymbol{\eta})$ *is given by*

$$\begin{aligned}
L_1(\delta_0, \boldsymbol{\eta}) &\equiv \int_{\mathbb{R}^q}\int_{\mathbb{R}^p} [\boldsymbol{y} \mid \boldsymbol{\theta}, \boldsymbol{u}, \delta_0][\boldsymbol{u} \mid \delta_0, \boldsymbol{\eta}]\, \mathrm{d}\boldsymbol{\theta}\, \mathrm{d}\boldsymbol{u} \\
&= \frac{1}{(2\pi\delta_0)^{\frac{n-p}{2}} |\boldsymbol{X}_0'\boldsymbol{X}_0|^{\frac{1}{2}} |\boldsymbol{B}_{\boldsymbol{\eta}}|^{\frac{1}{2}} |\boldsymbol{G}_{\boldsymbol{\eta}}|^{\frac{1}{2}}} \exp\left\{-\frac{\boldsymbol{y}'\boldsymbol{R}_{\boldsymbol{\eta}}\boldsymbol{y}}{2\delta_0}\right\}.
\end{aligned} \tag{15}$$

(b) *Under prior* (12) *and* (13) *with* $n - p + 2a_0 > 0$, *the marginal likehood function of* $\boldsymbol{\eta}$ *is*

$$
\begin{aligned}
L_2(\boldsymbol{\eta}) &\equiv \int_0^\infty L_1(\delta_0, \boldsymbol{\eta}) \frac{\mathrm{e}^{-\beta_0/\delta_0}}{\delta_0^{a_0+1}} \mathrm{d}\delta_0 \\
&= \frac{2^{a_0} \Gamma\left(\frac{n-p}{2} + a_0\right)}{\pi^{\frac{n-p}{2}} \left|\boldsymbol{X}_0' \boldsymbol{X}_0\right|^{\frac{1}{2}} \left|\boldsymbol{B}_{\boldsymbol{\eta}}\right|^{\frac{1}{2}} \left|\boldsymbol{G}_{\boldsymbol{\eta}}\right|^{\frac{1}{2}} \left(\boldsymbol{y}' \boldsymbol{R}_{\boldsymbol{\eta}} \boldsymbol{y} + 2b_0\right)^{\frac{n-p}{2}+a_0}}.
\end{aligned} \tag{16}
$$

Proof Note that $(\boldsymbol{y} - \boldsymbol{X}_0\boldsymbol{\theta} - \boldsymbol{X}\boldsymbol{u})'(\boldsymbol{y} - \boldsymbol{X}_0\boldsymbol{\theta} - \boldsymbol{X}\boldsymbol{u})$ can be decomposed as

$$
(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{u})'(\boldsymbol{I}_n - \boldsymbol{H}_0)(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{u}) + (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})\boldsymbol{X}_0'\boldsymbol{X}_0(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}),
$$

where $\hat{\boldsymbol{\theta}} = (\boldsymbol{X}_0'\boldsymbol{X}_0)^{-1}\boldsymbol{X}_0'\,\boldsymbol{y}$. Then

$$
\int_{\mathbb{R}^p} [\boldsymbol{y} \mid \boldsymbol{\theta}, \boldsymbol{u}, \delta_0]\, \mathrm{d}\boldsymbol{\theta} = \frac{|\boldsymbol{X}_0'\boldsymbol{X}_0|^{-\frac{1}{2}}}{(2\pi\delta_0)^{\frac{n-p}{2}}} \exp\left\{-\frac{1}{2\delta_0}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{u})'(\boldsymbol{I}_n - \boldsymbol{H}_0)(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{u})\right\}.
$$

It is easy to verify that

$$
\begin{aligned}
&(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{u})'(\boldsymbol{I}_n - \boldsymbol{H}_0)(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{u}) + \boldsymbol{u}'\boldsymbol{B}_{\boldsymbol{\eta}}^{-1}\boldsymbol{u} \\
&\quad = (\boldsymbol{u} - \boldsymbol{c})'\boldsymbol{G}_{\boldsymbol{\eta}}(\boldsymbol{u} - \boldsymbol{c}) + \boldsymbol{y}'(\boldsymbol{I}_n - \boldsymbol{H}_0)\boldsymbol{y} - \boldsymbol{c}'\boldsymbol{G}_{\boldsymbol{\eta}}\boldsymbol{c},
\end{aligned}
$$

where $\boldsymbol{c} = \boldsymbol{G}_{\boldsymbol{\eta}}^{-1}\boldsymbol{X}'(\boldsymbol{I}_n - \boldsymbol{H}_0)\boldsymbol{y}$. Thus

$$
\begin{aligned}
L_1(\delta_0, \boldsymbol{\eta}) &= \int_{\mathbb{R}^q} \frac{\exp\left\{-\frac{1}{2\delta_0}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{u})'(\boldsymbol{I}_n - \boldsymbol{H}_0)(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{u})\right\}}{(2\pi\delta_0)^{\frac{n-p}{2}}|\boldsymbol{X}_0'\boldsymbol{X}_0|^{\frac{1}{2}}} [\boldsymbol{u} \mid \delta_0, \boldsymbol{\eta}]\, \mathrm{d}\boldsymbol{u} \\
&= \frac{|\boldsymbol{X}_0'\boldsymbol{X}_0|^{-\frac{1}{2}}}{(2\pi\delta_0)^{\frac{n-p}{2}}\left|\boldsymbol{B}_{\boldsymbol{\eta}}\right|^{\frac{1}{2}}\left|\boldsymbol{G}_{\boldsymbol{\eta}}\right|^{\frac{1}{2}}} \exp\left\{-\frac{1}{2\delta_0}[\boldsymbol{y}'(\boldsymbol{I}_n - \boldsymbol{H}_0)\boldsymbol{y} - \boldsymbol{c}'\boldsymbol{G}_{\boldsymbol{\eta}}\boldsymbol{c}]\right\}.
\end{aligned}
$$

Because

$$
\boldsymbol{c}'\boldsymbol{G}_{\boldsymbol{\eta}}\boldsymbol{c} = \boldsymbol{y}'(\boldsymbol{I}_n - \boldsymbol{H}_0)\boldsymbol{X}\boldsymbol{G}_{\boldsymbol{\eta}}^{-1}\boldsymbol{X}'(\boldsymbol{I}_n - \boldsymbol{H}_0)\boldsymbol{y},
$$

we have $\boldsymbol{y}'(\boldsymbol{I}_n - \boldsymbol{H}_0)\boldsymbol{y} - \boldsymbol{c}'\boldsymbol{G}_{\boldsymbol{\eta}}\boldsymbol{c} = \boldsymbol{y}'\boldsymbol{R}_{\boldsymbol{\eta}}\boldsymbol{y}$, and (15) holds. Part (b) follows from part (a) immediately.

In the following, we use the partial ordering on $m \times m$ matrices $\boldsymbol{A}_1 \le \boldsymbol{A}_2$ if and only if $\boldsymbol{A}_2 - \boldsymbol{A}_1$ is nonnegative definite. We need the following condition.

Assumption A $(\boldsymbol{I}_n - \boldsymbol{H}_0)\boldsymbol{X}_k\boldsymbol{X}_k' = \boldsymbol{I}_n - \boldsymbol{H}_0$ for $k = 1, \ldots, r$.

Note that Assumption A is satisfied automatically for the additive models considered here since, for each $k = 1, \ldots, r$, $\boldsymbol{H}_k = \boldsymbol{X}_k \boldsymbol{X}_k'$ is a projection matrix with $\text{null}(\boldsymbol{H}_k) \subset \text{range}(\boldsymbol{H}_0)$.

Remark 2 This assumption does not hold for typical linear mixed models and seems characteristic of the mixed models associated with smoothing treated here. In particular, the assumption does not hold for P-splines.

We will write $\boldsymbol{\Lambda}_k = \text{diag}(\lambda_{k1}, \ldots, \lambda_{k\,q_k})$, $k = 1, \ldots, r$, and define

$$\lambda_{\min} = \min(\lambda_{kj}, k = 1, \ldots, r, j = 1, \ldots, q_k),$$
$$\lambda_{\max} = \max(\lambda_{kj}, k = 1, \ldots, r, j = 1, \ldots, q_k).$$

Lemma 2 *Under Assumption A,*

$$\left(\prod_{k=1}^{r}\prod_{j=1}^{q_k}\frac{\lambda_{kj}}{\lambda_{\max}}\right)\left(1+\lambda_{\max}\sum_{k=1}^{r}\frac{1}{\eta_k}\right)^{n-p}$$
$$\leq |\boldsymbol{G}_{\boldsymbol{\eta}}|\,|\boldsymbol{B}_{\boldsymbol{\eta}}| \leq \left(\prod_{k=1}^{r}\prod_{j=1}^{q_k}\frac{\lambda_{kj}}{\lambda_{\min}}\right)\left(1+\lambda_{\min}\sum_{k=1}^{r}\frac{1}{\eta_k}\right)^{n-p}, \tag{17}$$
$$\frac{1}{1+\lambda_{\max}\sum_{k=1}^{r}\frac{1}{\eta_k}}(\boldsymbol{I}_n - \boldsymbol{H}_0) \leq \boldsymbol{R}_{\boldsymbol{\eta}} \leq \frac{1}{1+\lambda_{\min}\sum_{k=1}^{r}\frac{1}{\eta_k}}(\boldsymbol{I}_n - \boldsymbol{H}_0). \tag{18}$$

Proof Write $\widetilde{\boldsymbol{X}} = (\boldsymbol{I}_n - \boldsymbol{H}_0)\boldsymbol{X}$. Then

$$\boldsymbol{G}_{\boldsymbol{\eta}} = \boldsymbol{B}_{\boldsymbol{\eta}}^{-1} + \widetilde{\boldsymbol{X}}'\widetilde{\boldsymbol{X}}.$$

Let $\widetilde{\boldsymbol{B}}_\eta = diag(\frac{1}{\eta_1}\boldsymbol{I}_{q_1}, \ldots, \frac{1}{\eta_r}\boldsymbol{I}_{q_r})$. We know that

$$\lambda_{\max}^{-1}\widetilde{\boldsymbol{B}}_\eta^{-1} + \widetilde{\boldsymbol{X}}'\widetilde{\boldsymbol{X}} \;\leq\; \boldsymbol{G}_{\boldsymbol{\eta}} \;\leq\; \lambda_{\min}^{-1}\widetilde{\boldsymbol{B}}_\eta^{-1} + \widetilde{\boldsymbol{X}}'\widetilde{\boldsymbol{X}}. \tag{19}$$

Using a well-known formula (e.g., Christensen 2002, p. 416), for any $c > 0$,

$$(c^{-1}\widetilde{\boldsymbol{B}}_\eta^{-1} + \widetilde{\boldsymbol{X}}'\widetilde{\boldsymbol{X}})^{-1} = c\widetilde{\boldsymbol{B}}_\eta - c^2\widetilde{\boldsymbol{B}}_\eta\widetilde{\boldsymbol{X}}'(\boldsymbol{I}_n + c\widetilde{\boldsymbol{X}}\widetilde{\boldsymbol{B}}_\eta\widetilde{\boldsymbol{X}}')^{-1}\widetilde{\boldsymbol{X}}\widetilde{\boldsymbol{B}}_\eta. \tag{20}$$

Noting that

$$\widetilde{\boldsymbol{X}}\widetilde{\boldsymbol{B}}_\eta = (\boldsymbol{I}_n - \boldsymbol{H}_0)\boldsymbol{X}\widetilde{\boldsymbol{B}}_\eta = (\boldsymbol{I}_n - \boldsymbol{H}_0)\Big(\frac{1}{\eta_1}\boldsymbol{X}_1, \ldots, \frac{1}{\eta_r}\boldsymbol{X}_r\Big),$$

we have

$$\widetilde{\boldsymbol{X}}\widetilde{\boldsymbol{B}}_\eta\widetilde{\boldsymbol{X}}' = (\boldsymbol{I}_n - \boldsymbol{H}_0)\left(\sum_{k=1}^{r}\frac{1}{\eta_k}\boldsymbol{X}_k\boldsymbol{X}_k'\right)(\boldsymbol{I}_n - \boldsymbol{H}_0)$$

$$= \left(\sum_{k=1}^{r}\frac{1}{\eta_k}\right)(\boldsymbol{I}_n - \boldsymbol{H}_0). \tag{21}$$

The last equality follows from Assumption A. Thus for any $c > 0$,

$$\boldsymbol{I}_n + c\widetilde{\boldsymbol{X}}\boldsymbol{B}_{\boldsymbol{\eta}}\widetilde{\boldsymbol{X}}' = \boldsymbol{H}_0 + \left(1 + c\sum_{k=1}^{r}\frac{1}{\eta_k}\right)(\boldsymbol{I}_n - \boldsymbol{H}_0),$$

$$\left(\boldsymbol{I}_n + c\widetilde{\boldsymbol{X}}\boldsymbol{B}_{\boldsymbol{\eta}}\widetilde{\boldsymbol{X}}'\right)^{-1} = \boldsymbol{H}_0 + \left(1 + c\sum_{k=1}^{r}\frac{1}{\eta_k}\right)^{-1}(\boldsymbol{I}_n - \boldsymbol{H}_0), \tag{22}$$

and

$$|\boldsymbol{I}_n + c\widetilde{\boldsymbol{X}}\boldsymbol{B}_{\boldsymbol{\eta}}\widetilde{\boldsymbol{X}}'| = \left(1 + c\sum_{k=1}^{r}\frac{1}{\eta_k}\right)^{n-p}.$$

The last two equalities hold because $\boldsymbol{H}_0$ is a projection matrix. If two positive definite matrices satisfy $\boldsymbol{A}_1 \le \boldsymbol{A}_2$, then $|\boldsymbol{A}_1| \le |\boldsymbol{A}_2|$. (To see this, note that $\boldsymbol{A}_1^{-1/2}\boldsymbol{A}_2\boldsymbol{A}_1^{-1/2} \ge \boldsymbol{I}$.) Moreover, for any invertible $q \times q$ matrix $\boldsymbol{A}$ and $n \times q$ matrix $\boldsymbol{X}$, $|\boldsymbol{A} + \boldsymbol{X}'\boldsymbol{X}| = |\boldsymbol{A}||\boldsymbol{I}_n + \boldsymbol{X}\boldsymbol{A}^{-1}\boldsymbol{X}'|$ (e.g., Harville 1997, p. 188). Thus, from (19),

$$|\boldsymbol{G}_{\boldsymbol{\eta}}| \ge |\lambda_{\max}^{-1}\widetilde{\boldsymbol{B}}_\eta^{-1} + \widetilde{\boldsymbol{X}}'\widetilde{\boldsymbol{X}}| = \frac{1}{\lambda_{\max}^{q}|\widetilde{\boldsymbol{B}}_\eta|}\left|\boldsymbol{I}_n + \lambda_{\max}\widetilde{\boldsymbol{X}}\widetilde{\boldsymbol{B}}_\eta\widetilde{\boldsymbol{X}}'\right|;$$

$$|\boldsymbol{G}_{\boldsymbol{\eta}}| \le |\lambda_{\min}^{-1}\widetilde{\boldsymbol{B}}_\eta^{-1} + \widetilde{\boldsymbol{X}}'\widetilde{\boldsymbol{X}}| = \frac{1}{\lambda_{\min}^{q}|\widetilde{\boldsymbol{B}}_\eta|}\left|\boldsymbol{I}_n + \lambda_{\min}\widetilde{\boldsymbol{X}}\widetilde{\boldsymbol{B}}_\eta\widetilde{\boldsymbol{X}}'\right|.$$

Since $|\boldsymbol{B}_{\boldsymbol{\eta}}|/|\widetilde{\boldsymbol{B}}_\eta| = \prod_{k=1}^{r}\prod_{j=1}^{q_k}\lambda_{kj}$, inequality (17) holds.

From its definition (14) and the first inequality of (19),

$$\boldsymbol{R}_{\boldsymbol{\eta}} = \boldsymbol{I}_n - \boldsymbol{H}_0 - \widetilde{\boldsymbol{X}}\boldsymbol{G}_{\boldsymbol{\eta}}^{-1}\widetilde{\boldsymbol{X}}'$$

$$\le \boldsymbol{I}_n - \boldsymbol{H}_0 - \widetilde{\boldsymbol{X}}\left(\lambda_{\min}^{-1}\widetilde{\boldsymbol{B}}_\eta^{-1} + \widetilde{\boldsymbol{X}}'\widetilde{\boldsymbol{X}}\right)^{-1}\widetilde{\boldsymbol{X}}'.$$

Using (20) and (22),

$$
\begin{aligned}
\boldsymbol{R}_{\eta} &\leq \boldsymbol{I}_n - \boldsymbol{H}_0 - \widetilde{\boldsymbol{X}}\left[\lambda_{\min}\widetilde{\boldsymbol{B}}_{\eta} - \lambda_{\min}^2\widetilde{\boldsymbol{B}}_{\eta}\widetilde{\boldsymbol{X}}'(\boldsymbol{I}_n + \lambda_{\min}\widetilde{\boldsymbol{X}}\widetilde{\boldsymbol{B}}_{\eta}\widetilde{\boldsymbol{X}}')^{-1}\widetilde{\boldsymbol{X}}\widetilde{\boldsymbol{B}}_{\eta}\right]\widetilde{\boldsymbol{X}}' \\
&= \boldsymbol{I}_n - \boldsymbol{H}_0 - \lambda_{\min}\widetilde{\boldsymbol{X}}\widetilde{\boldsymbol{B}}_{\eta}\widetilde{\boldsymbol{X}}' \\
&\quad + \lambda_{\min}^2\widetilde{\boldsymbol{X}}\widetilde{\boldsymbol{B}}_{\eta}\widetilde{\boldsymbol{X}}'\left[\boldsymbol{H}_0 + \left(1 + \lambda_{\min}\sum_{k=1}^{r}\frac{1}{\eta_k}\right)^{-1}(\boldsymbol{I}_n - \boldsymbol{H}_0)\right]\widetilde{\boldsymbol{X}}\widetilde{\boldsymbol{B}}_{\eta}\widetilde{\boldsymbol{X}}'.
\end{aligned}
$$

Finally, because $\boldsymbol{H}_0\widetilde{\boldsymbol{X}} = \boldsymbol{0}$ and $(\boldsymbol{I}_n - \boldsymbol{H}_0)\widetilde{\boldsymbol{X}} = \widetilde{\boldsymbol{X}}$, equality (21) implies

$$
\begin{aligned}
\boldsymbol{R}_{\boldsymbol{\eta}} &\leq \boldsymbol{I}_n - \boldsymbol{H}_0 - \lambda_{\min}\sum_{k=1}^{r}\frac{1}{\eta_k}(\boldsymbol{I}_n - \boldsymbol{H}_0) \\
&\quad + \lambda_{\min}^2\left(\sum_{k=1}^{r}\frac{1}{\eta_k}\right)^2\left(1 + \lambda_{\min}\sum_{k=1}^{r}\frac{1}{\eta_k}\right)^{-1}(\boldsymbol{I}_n - \boldsymbol{H}_0) \\
&= \frac{\left(1 + \lambda_{\min}\sum_{k=1}^{r}\frac{1}{\eta_k}\right)\left(1 - \lambda_{\min}\sum_{k=1}^{r}\frac{1}{\eta_k}\right) + \lambda_{\min}^2\left(\sum_{k=1}^{r}\frac{1}{\eta_k}\right)^2}{1 + \lambda_{\min}\sum_{k=1}^{r}\frac{1}{\eta_k}}(\boldsymbol{I}_n - \boldsymbol{H}_0) \\
&= \frac{1}{1 + \lambda_{\min}\sum_{k=1}^{r}\frac{1}{\eta_k}}(\boldsymbol{I}_n - \boldsymbol{H}_0).
\end{aligned}
$$

This proves the righthand inequality of (18). The other inequality can be shown similarly.

Proof of Theorem 1. Applying the last lemma, the term $|\boldsymbol{B}_{\boldsymbol{\eta}}\boldsymbol{G}_{\boldsymbol{\eta}}|$ is bounded below by a positive constant.

Lemma 3 *Consider the linear mixed model* (10) *and* (11) *under priors* (12) *and* (13) *with* $n - p + 2a_0 > 0$*, and suppose Assumption A is satisfied. Then for all* $\boldsymbol{\eta} = (\eta_1, \ldots, \eta_r)$ *with* $\eta_k > 0$*,*

$$
\begin{aligned}
&\frac{K(a_0, \lambda_{\min})\left(1 + \lambda_{\min}\sum_{k=1}^{r}\frac{1}{\eta_k}\right)^{-\frac{n-p}{2}}}{\left[\frac{\boldsymbol{y}'(\boldsymbol{I}_n - \boldsymbol{H}_0)\boldsymbol{y}}{1+\lambda_{\min}\sum_{k=1}^{r}\frac{1}{\eta_k}} + 2b_0\right]^{\frac{n-p}{2}+a_0}} \\
&\quad \leq L_2(\boldsymbol{\eta}) \leq \frac{K(a_0, \lambda_{\max})\left(1 + \lambda_{\max}\sum_{k=1}^{r}\frac{1}{\eta_k}\right)^{-\frac{n-p}{2}}}{\left[\frac{\boldsymbol{y}'(\boldsymbol{I}_n - \boldsymbol{H}_0)\boldsymbol{y}}{1+\lambda_{\max}\sum_{k=1}^{r}\frac{1}{\eta_k}} + 2b_0\right]^{\frac{n-p}{2}+a_0}},
\end{aligned}
\tag{23}
$$

where $K(\cdot,\cdot)$ *is the positive function*

$$
K(a, \lambda) = \frac{2^a\Gamma\left(\frac{n-p}{2} + a\right)}{\pi^{\frac{n-p}{2}}|\boldsymbol{X}_0'\boldsymbol{X}_0|^{\frac{1}{2}}}\left(\prod_{k=1}^{r}\prod_{j=1}^{q_k}\frac{\lambda}{\lambda_{kj}}\right)^{\frac{1}{2}}, \quad a, \lambda \in \mathbb{R}. \tag{24}
$$

Proof Inequality (18) implies that

$$\frac{\boldsymbol{y}'(\boldsymbol{I}_n - \boldsymbol{H}_0)\boldsymbol{y}}{1 + \lambda_{\max}\sum_{k=1}^{r}\frac{1}{\eta_k}} \le \boldsymbol{y}'\boldsymbol{R}_{\boldsymbol{\eta}}\boldsymbol{y} \le \frac{\boldsymbol{y}'(\boldsymbol{I}_n - \boldsymbol{H}_0)\boldsymbol{y}}{1 + \lambda_{\min}\sum_{k=1}^{r}\frac{1}{\eta_k}}. \tag{25}$$

Applying (17) and (25) to (16), we have

$$L_2(\boldsymbol{\eta}) \ge \frac{\Gamma\left(\frac{n-p}{2} + a_0\right)\left(\prod_{k=1}^{r}\prod_{j=1}^{q_k}\frac{\lambda_{\min}}{\lambda_{kj}}\right)^{1/2}}{\pi^{\frac{n-p}{2}}|\boldsymbol{X}_0'\boldsymbol{X}_0|^{\frac{1}{2}}\left[\frac{\boldsymbol{y}'(\boldsymbol{I}_n-\boldsymbol{H}_0)\boldsymbol{y}}{1+\lambda_{\min}\sum_{k=1}^{r}\frac{1}{\eta_k}} + 2b_0\right]^{\frac{n-p}{2}+a_0}}\left(1 + \lambda_{\min}\sum_{k=1}^{r}\frac{1}{\eta_k}\right)^{-\frac{n-p}{2}},$$

$$L_2(\boldsymbol{\eta}) \le \frac{\Gamma\left(\frac{n-p}{2} + a_0\right)\left(\prod_{k=1}^{r}\prod_{j=1}^{q_k}\frac{\lambda_{\max}}{\lambda_{kj}}\right)^{1/2}}{\pi^{\frac{n-p}{2}}|\boldsymbol{X}_0'\boldsymbol{X}_0|^{\frac{1}{2}}\left[\frac{\boldsymbol{y}'(\boldsymbol{I}_n-\boldsymbol{H}_0)\boldsymbol{y}}{1+\lambda_{\max}\sum_{k=1}^{r}\frac{1}{\eta_k}} + 2b_0\right]^{\frac{n-p}{2}+a_0}}\left(1 + \lambda_{\max}\sum_{k=1}^{r}\frac{1}{\eta_k}\right)^{-\frac{n-p}{2}}.$$

This proves the lemma.

As an immediate consequence of the lemma, we have the following characterization.

Theorem 5 *For the linear mixed model* (10) *and* (11) *together with priors* (12) *and* (13) *and Assumption A, if* $a_0 = b_0 = 0$*, the joint posterior of* $(\boldsymbol{\theta}, \boldsymbol{u}, \delta_0, \boldsymbol{\eta})$ *is proper if and only if the prior of* $\boldsymbol{\eta}$ *is proper.*

Proof Noting that if $a_0 = b_0 = 0$, (23) becomes

$$\frac{K(a_0, \lambda_{\min})}{\left[\boldsymbol{y}'(\boldsymbol{I}_n - \boldsymbol{H}_0)\boldsymbol{y}\right]^{(n-p)/2}} \le L_2(\boldsymbol{\eta}) \le \frac{K(a_0, \lambda_{\max})}{\left[\boldsymbol{y}'(\boldsymbol{I}_n - \boldsymbol{H}_0)\boldsymbol{y}\right]^{(n-p)/2}}.$$

Because the upper and lower bounds do not depend on $\boldsymbol{\eta}$, the result holds.

Theorem 2 follows from Theorem 5 as a special case.

We now restrict consideration to independent inverse gamma-type priors for the δ_k of the form (8). Note that (8) is equivalent to $(\delta_0, \boldsymbol{\eta})$ having joint prior

$$[\delta_0, \boldsymbol{\eta}] = \frac{e^{-b_0/\delta_0}}{\delta_0^{1+a_+}}\prod_{k=1}^{r}\eta_k^{a_k-1}e^{-b_k\eta_k/\delta_0}, \tag{26}$$

where $a_+ = \sum_{k=0}^{r} a_k$. The following conditions will be needed.

Condition C1. $n - p + 2a_+ > 0$.
Condition C2. $a_k < 0$ for $k = 0, 1, \ldots, r$.
Condition C3. $n - p + 2\sum_{k=0}^{r}\min(0, a_k) > 0$.

Lemma 4 *(a) Under Condition C1, we have*

$$\int_0^\infty L_1(\delta_0, \boldsymbol{\eta})[\delta_0, \boldsymbol{\eta}]\, d\delta_0$$
$$= \frac{2^{a_+}\Gamma\left(\frac{n-p}{2}+a_+\right)\prod_{k=1}^r \eta_k^{a_k-1}}{\pi^{\frac{n-p}{2}}\left|\boldsymbol{X}_0'\boldsymbol{X}_0\right|^{\frac{1}{2}}\left|\boldsymbol{B}_{\boldsymbol{\eta}}\right|^{\frac{1}{2}}\left|\boldsymbol{G}_{\boldsymbol{\eta}}\right|^{\frac{1}{2}}\left(\boldsymbol{y}'\boldsymbol{R}_{\boldsymbol{\eta}}\boldsymbol{y}+2b_0+2\sum_{k=1}^r b_k\eta_k\right)^{\frac{n-p}{2}+a_+}}.$$

(b) If Assumption A and Condition C1 hold and $b_0 = \cdots = b_r = 0$, we have

$$\frac{K(a_+, \lambda_{\min})g(\lambda_{\min}, \boldsymbol{\eta})}{\left[\boldsymbol{y}'(\boldsymbol{I}_n - \boldsymbol{H}_0)\boldsymbol{y}\right]^{\frac{n-p}{2}+a_+}} \le \int_0^\infty L_1(\delta_0, \boldsymbol{\eta})[\delta_0, \boldsymbol{\eta}]\, \mathrm{d}\delta_0 \le \frac{K(a_+, \lambda_{\max})g(\lambda_{\max}, \boldsymbol{\eta})}{\left[\boldsymbol{y}'(\boldsymbol{I}_n - \boldsymbol{H}_0)\boldsymbol{y}\right]^{\frac{n-p}{2}+a_+}},$$

where $K(\cdot,\cdot)$ is given by (24) and

$$g(c, \boldsymbol{\eta}) = \left(\prod_{k=1}^r \eta_k^{a_k-1}\right)\left(1 + c\sum_{k=1}^r \frac{1}{\eta_k}\right)^{a_+}, \quad c > 0,\ \ \eta_k > 0.$$

(c) If Assumption A and Condition C1 hold, $b_k > 0$ for $k = 0, \ldots, r$, and $\boldsymbol{y}'(\boldsymbol{I}_n - \boldsymbol{H}_0)\boldsymbol{y} > 0$, there exist positive constants $0 < K_1, K_2 < \infty$ such that

$$K_1 h(\boldsymbol{\eta}) \le \int_0^\infty L_1(\delta_0, \boldsymbol{\eta})[\delta_0, \boldsymbol{\eta}]\, \mathrm{d}\delta_0 \le K_2 h(\boldsymbol{\eta}),$$

where

$$h(\boldsymbol{\eta}) = \frac{g(1, \boldsymbol{\eta})}{\left[1 + \left(1 + \sum_{k=1}^r \frac{1}{\eta_k}\right)(1 + \sum_{k=1}^r \eta_k)\right]^{(n-p)/2+a_+}}, \quad \eta_k > 0.$$

Proof Integrating out the product of $L_1(\delta_0, \boldsymbol{\eta})$ in (15) and the prior (26) with respect to δ_0, we get part (a). Parts (b) and (c) follow from (25) as in the proof of Lemma 3.

The following result contains the assertions of Theorem 3.

Theorem 6 *Consider the linear mixed model* (10) *and* (11)*. Assume that* $[\boldsymbol{\theta}] \equiv 1$*, and let the prior for* $(\delta_0, \delta_1, \ldots, \delta_k)$ *be specified by* (8)*.*

(a) *If $b_0 = \cdots = b_r = 0$, then Condition C1 is necessary for the existence of the posterior of* $(\boldsymbol{\theta}, \boldsymbol{u}, \delta_0, \boldsymbol{\eta})$.
(b) *Under Assumption A, if $b_0 = \cdots = b_r = 0$, then Conditions C1 and C2 are necessary and sufficient for the existence of the posterior of* $(\boldsymbol{\theta}, \boldsymbol{u}, \delta_0, \boldsymbol{\eta})$.
(c) *Under Assumption A, if $b_k > 0$, $k = 0, \ldots, r$, then Condition C3 is sufficient for the existence of the posterior of* $(\boldsymbol{\theta}, \boldsymbol{u}, \delta_0, \boldsymbol{\eta})$.

Proof To prove (a), we note that in part (a) of Lemma 4, $L_1(\delta_0, \boldsymbol{\eta})[\delta_0, \boldsymbol{\eta}]$ is integrable with respect to δ_0 only if Condition C1 holds. To prove (b), we need only show that

$$J_c \equiv \int_0^\infty \cdots \int_0^\infty g(c, \boldsymbol{\eta})\, \mathrm{d}\eta_1 \cdots \mathrm{d}\eta_r < \infty, \quad c > 0,$$

if and only if Condition C2 holds. In fact, by the transformation $t_k = c/\eta_k$, we have

$$J_c = c^{a_1+\cdots+a_r} \int_0^\infty \cdots \int_0^\infty \frac{(1 + t_1 + \cdots + t_r)^{a_0+a_1+\cdots+a_r}}{t_1^{a_1+1} \cdots t_r^{a_r+1}}\, \mathrm{d}t_1 \cdots \mathrm{d}t_r.$$

Make the transformation $s_1 = t_1/(1+t_1)$, $s_2 = t_2/(1+t_1+t_2), \ldots, s_r = t_r/(1+t_1+\cdots+t_r)$. It is easy to see that $s_k \in (0, 1)$ and $(1+t_1 \cdots +t_k) = 1/[(1-s_1)\cdots(1-s_k)]$, for all $k = 1, \ldots, r$, hence $t_1 = s_1/(1-s_1)$, $t_2 = s_2/[(1-s_1)(1-s_2)], \ldots, t_r = s_r/[(1-s_1)\cdots(1-s_r)]$. The Jacobian of the transformation is

$$\begin{aligned} \left|\frac{\partial(t_1, \ldots, t_r)}{\partial(s_1, \ldots, s_r)}\right| &= \frac{1}{(1-s_1)^2} \prod_{k=2}^{r} \frac{1}{(1-s_1)\cdots(1-s_{k-1})(1-s_k)^2} \\ &= \prod_{k=2}^{r} \frac{1}{(1-s_k)^{r-k+2}}. \end{aligned}$$

After some simplification,

$$J_c = c^{a_1+\cdots+a_r} \prod_{k=1}^{r} \int_0^1 s_k^{-a_k-1} (1-s_k)^{-(a_0+a_1+\cdots+a_{k-1})-1}\, \mathrm{d}s_k.$$

But J_c clearly is finite if and only if Condition C2 holds, i.e., if and only if $a_k < 0$ for all k.

To prove (c), define

$$V \equiv \int_0^\infty \cdots \int_0^\infty h(\boldsymbol{\eta})\, \mathrm{d}\boldsymbol{\eta}.$$

Using Lemma 4 and Condition C1,

$$V < \int_0^\infty \cdots \int_0^\infty \frac{\prod_{k=1}^r \eta_k^{a_k-1}}{\left(1 + \sum_{k=1}^r \frac{1}{\eta_k}\right)^{(n-p)/2} \left(1 + \sum_{k=1}^r \eta_k\right)^{(n-p)/2+a_+}} \mathrm{d}\boldsymbol{\eta}.$$

For an ordered subset $J \subset \{1, \ldots, r\}$, define

$$V_0(J) = \int_{0<\eta_k<1, k\in J} \frac{\prod_{k\in J} \eta_k^{a_k-1}}{\left(1 + \sum_{k\in J} \frac{1}{\eta_k}\right)^{(n-p)/2}} \mathrm{d}\boldsymbol{\eta}_J$$

$$V_\infty(J) = \int_{1\le\eta_k<\infty, k\in J} \frac{\prod_{k\in J} \eta_k^{a_k-1}}{\left(1 + \sum_{k\in J} \eta_k\right)^{(n-p)/2+a_+}} \mathrm{d}\boldsymbol{\eta}_J.$$

(The notation $\mathrm{d}\boldsymbol{\eta}_J$ means $\mathrm{d}\eta_{k_1} \cdots \mathrm{d}\eta_{k_j}$ when $J = \{k_1, \ldots, k_j\}$.) Then

$$V < \sum_{\mathrm{all} J} V_0(J) V_\infty(J^c),$$

where the sum is over all ordered subsets J of $\{1, \cdots, r\}$ including the empty set. Without loss of generality, consider $J = \{1, \ldots, r\}$. Using the transformations from the proof for part (b) above,

$$\begin{aligned} V_0(\{1, \ldots, r\}) &= \int_1^\infty \cdots \int_1^\infty \prod_{k=1}^r t_k^{-a_k-1} (1 + t_1 + \cdots + t_r)^{-\frac{n-p}{2}} \mathrm{d}t_1 \cdots \mathrm{d}t_r \\ &= \int_{\frac{1}{2}}^1 \int_{\frac{1}{3}}^1 \cdots \int_{\frac{1}{r+1}}^1 \prod_{k=1}^r s_k^{-a_k-1} \prod_{k=1}^r (1 - s_k)^{\frac{n-p}{2} + \sum_{j=k}^r a_j - 1} \mathrm{d}s_1 \cdots \mathrm{d}s_r. \end{aligned}$$

Clearly, $V_0(J)$ is finite if $(n-p)/2 + \sum_{j=k}^r a_j > 0$ for $k = 1, \ldots, r$. Considering all subsets J, if

$$(n-p)/2 + \sum_{k\in J} a_k > 0 \quad \text{for all } J \subset \{1, \ldots, r\}, \tag{27}$$

then $V_0(J) < \infty$ for all J.

Similarly, using the transformation $s_1 = \eta_1/(1+\eta_1)$, $s_2 = \eta_2/(1+\eta_1+\eta_2)$, ..., $s_r = \eta_r/(1 + \eta_1 + \cdots + \eta_r)$,

$$V_\infty(\{1, \ldots, r\}) = \int_{\frac{1}{2}}^1 \int_{\frac{1}{3}}^1 \cdots \int_{\frac{1}{r+1}}^1 \prod_{k=1}^r s_k^{a_k-1} \prod_{k=1}^r (1 - s_k)^{\frac{n-p}{2} + \sum_{j=0}^k a_j - 1} \mathrm{d}s_1 \cdots \mathrm{d}s_r,$$

which is finite if $(n-p)/2 + \sum_{j=1}^k a_j > 0$ for $k = 1, \ldots, r$. Thus $V_\infty(J) < \infty$ for all subsets J if

$$(n-p)/2 + a_0 + \sum_{k\in J} a_k > 0 \quad \text{for all } J \subset \{1, \ldots, r\}. \tag{28}$$

Taken together, conditions (27) and (28) are equivalent to Condition C3.

Remark 3 Consider the special case of independent priors of the form $[\delta_k] \propto \delta_k^{-a_k-1}$. If $(\boldsymbol{\theta}, \delta_0)$ has joint prior $[\boldsymbol{\theta}, \delta_0] = 1/\delta_0^{a_0+1}$ and Assumption A holds, Theorems 5 and 6 imply that there are exactly two choices for a_0:

(a) If $a_0 = 0$, one must use a proper prior for $\boldsymbol{\eta}$. Consequently, the joint posterior of $(\boldsymbol{\theta}, \boldsymbol{u}, \delta_0, \delta_1, \ldots, \delta_r)$ (or $(\boldsymbol{\theta}, \boldsymbol{u}, \delta_0, \boldsymbol{\eta})$) is improper for any prior of $(\delta_1, \ldots, \delta_r)$ of the form

$$[\delta_1, \ldots, \delta_r] = \prod_{k=1}^{r} \frac{1}{\delta_k^{a_k+1}}. \tag{29}$$

(b) If $a_0 < 0$, then one can choose priors for $(\delta_1, \ldots, \delta_r)$ given by (29) as long as $a_k < 0$ for every $k = 1, \ldots, r$ and $(n - p)/2 + a_0 + a_1 + \cdots + a_r > 0$.

Remark 4 As one of the referees pointed out, the prior $[\delta_0] = 1/\delta_0$ is also the limiting case when δ_0 follows a log normal (μ_0, σ_0^2) prior as $\sigma_0 \to \infty$. Thus, under a limiting log normal prior for δ_0, the prior for $\boldsymbol{\eta}$ must be proper for the existence of the posterior.

Acknowledgments The authors would like to thank the editors and referees for constructive comments to an early version of the paper.

References

Berger, J. (1985). *Statistical decision theory and Bayesian analysis*. New York: Springer.

Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems (with discussion). *Journal of the Royal Statistical Society, Series B, Methodological, 36*, 192–236.

Besag, J., Green, P., Higdon, D., Mengersen, K. (1995). Bayesian computation and stochastic systems (Disc: P41-66). *Statistical Science, 10*, 3–41.

Besag, J., Kooperberg, C. (1995). On conditional and intrinsic autoregressions. *Biometrika, 82*, 733–746.

Carter, C. K., Kohn, R. (1994). On Gibbs sampling for state space models. *Biometrika, 81*, 541–553.

Carter, C. K., Kohn, R. (1996). Markov chain Monte Carlo in conditionally Gaussian state space models. *Biometrika, 83*, 589–601.

Christensen, R. (2002). *Plane answers to complex questions: the theory of linear models*. New York: Springer.

Eubank, R. L. (1999). *Nonparametric regression and spline smoothing*. Marcel Dekker.

Fahrmeir, L., Lang, S. (2001). Bayesian inference for generalized additive mixed models based on Markov random field priors. *Applied Statistics, 50(2)*, 201–220.

Gelfand, A. E., Sahu, S. K. (1999). Identifiability, improper priors, and Gibbs sampling for generalized linear models. *Journal of the American Statistical Association, 94*, 247–253.

Green, P. J., Silverman, B. W. (1994). *Nonparametric regression and generalized linear models: a roughness penalty approach*. London; New York: Chapman & Hall.

Harville, D. A. (1997). *Matrix algebra from a statistician's perspective*. New York: Springer.

Hastie, T. J., Tibshirani, R. J. (1990). *Generalized additive models*. London; New York: Chapman & Hall.

Hastie, T., Tibshirani, R. (2000). Bayesian backfitting (with comments and a rejoinder by the authors. *Statistical Science, 15(3)*, 196–223.

Hobert, J. P., 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.

Lambert, P. C., Sutton, A. J., Burton, P. R., Abrams, K. R., Jones, D. R. (2005). How vague is vague? a simulation study of the impact of the use of vague prior distributions in mcmc using winbugs. *Statistics in Medicine, 24*, 2401–2428.

Lang, S., Brezger, A. (2004). Bayesian p-splines. *Journal of Computational and Graphical Statistics, 13*, 183–212.

Nychka, D. W. (2000). Spatial-process estimates as smoothers. In M. G. a. Schimek (ed.), *Smoothing and regression: approaches, computation, and application* (pp. 393–424). New York: Wiley

Shively, T. S., Kohn, R., Wood, S. (1999). Variable selection and function estimation in additive nonparametric regression using a data-based prior (C/R: p794-806). *Journal of the American Statistical Association, 94*, 777–794.

Speckman, P. L., Sun, D. (2003). Fully Bayesian spline smoothing and intrinsic autoregressive priors. *Biometrika, 90(2)*, 289–302.

Stone, C. J. (1985). Additive regression and other nonparametric models. *The Annals of Statistics, 13*, 689–705.

Sun, D., Tsutakawa, R. K., He, Z. (2001). Propriety of posteriors with improper priors in hierarchical linear mixed models. *Statistica Sinica, 11(1)*, 77–95.

Wahba, G. (1983). Bayesian "confidence intervals" for the cross-validated smoothing spline. *Journal of the Royal Statistical Society, Series B, Methodological, 45*, 133–150.

Wahba, G. (1990). *Spline models for observational data*. Philadelphia: Society for Industrial and Applied Mathematics.

Wong, C.-M., Kohn, R. (1996). A Bayesian approach to additive semiparametric regression. *Journal of Econometrics, 74*, 209–235.

Wood, S., Kohn, R., Shively, T., Jiang, W. (2002). Model selection in spline nonparametric regression. *Journal of the Royal Statistical Society, Series B, Methodological, 64(1)*, 119–139.