Statistica Sinica **11**(2001), 77-95

PROPRIETY OF POSTERIORS WITH IMPROPER PRIORS IN HIERARCHICAL LINEAR MIXED MODELS

Dongchu Sun, Robert K. Tsutakawa and Zhuoqiong He

University of Missouri-Columbia

Abstract: This paper examines necessary and sufficient conditions for the propriety of the posterior distribution in hierarchical linear mixed effects models for a collection of improper prior distributions. In addition to the flat prior for the fixed effects, the collection includes various limiting forms of the invariant gamma distribution for the variance components, including cases considered previously by Datta and Ghosh (1991), and Hobert and Casella (1996). Previous work is extended by considering a family of correlated random effects which include as special cases the intrinsic autoregressive models of Besag, York and Mollié (1991), the Autoregressive (AR) Model of Ord (1975), and the Conditional Autoregressive (CAR) Models of Clayton and Kaldor (1987), which have been found useful in the analysis of spatial effects. Conditions are then presented for the propriety of the posterior distribution for a generalized linear mixed model, where the first stage distribution belongs to an exponential family.

Key words and phrases: Generalized linear mixed models, Gibbs sampling, linear mixed models, Markov chain Monte Carlo, multivariate normal, variance components.

1. Introduction

Bayesian analysis, especially of hierarchical linear mixed models, has received much attention recently. There is often not enough information on hyperparameters for subjective Bayesian analysis and one often resorts to noninformative or default priors. For a recent review of noninformative priors, see Kass and Wasserman (1996).

There are several important reasons to consider noninformative priors for the generalized linear model. First, in the generalized linear model, it is difficult to derive the "standard" noninformative priors such as Jeffreys' prior, Berger and Bernardo's (1992) reference priors, or matching priors. Next, even when we can derive these "standard" noninformative priors, posterior distributions are usually computationally intractable because these priors often depend on sample sizes. Finally, simple noninformative priors are easy to implement. However without proper precaution, simple noninformative priors can be misused, sometimes unknowingly, and lead to other difficulties, such as the nonconvergence of the Gibbs sampler (cf. Hobert and Casella (1996)).

One commonly used noninformative prior is the constant prior density for some of the parameters, but this might lead to nonintegrable posteriors (cf. Ibrahim and Laud (1991)). Alternatively, scale-invariant priors are used in the literature. Consider the one-way random effects model (cf. Morris (1983)), where $(y_j|\theta_j, \delta_0)$ are i.i.d. $N(\theta_j, \delta_0)$, and $(\theta_j|\mu, \delta_1)$ are i.i.d. $N(\mu, \delta_1)$. Here for given (θ_j, δ_0), y_j and (μ, δ_1) are independent. Under the scale-invariant prior $p(\mu, \delta_0, \delta_1) \propto 1/(\delta_0\delta_1)$, the posterior of θ_j is improper. One solution is to use a dependent prior, $p(\mu, \delta_0, \delta_1) \propto (\delta_0 + \delta_1)^{-1}$ (cf. Kahn (1990)). However, such an approach to hierarchical models often involves the sample size and the design matrix. Consider also the balanced one-way ANOVA model:

$$Y_{ij} = \mu + u_i + e_{ij}, \quad i = 1, \ldots, I, \; j = 1, \ldots, n,$$

where the u_i are i.i.d. $N(0, \delta_1)$ and, independently, e_{ij} are i.i.d. $N(0, \delta_0)$. The Jeffreys prior for $(\mu, \delta_0, \delta_1)$ is of the form $p(\mu, \delta_0, \delta_1) \propto [\delta_0(\delta_0 + n\delta_1)^{3/2}]^{-1}$. Berger and Bernardo (1992) suggest the reference prior $p(\mu, \delta_0, \delta_1) \propto [\delta_0(\delta_0 + n\delta_1)]^{-1}$. Ghosh and Mukerjee (1993) and Datta (1996) showed that this reference prior is a matching prior when each or all three parameters are of interest. Such a prior depends on the sample size and use of this prior can be computationally impractical in practice, especially for an unbalanced design. When n gets large this prior is close to $1/(\delta_0\delta_1)$, the scale-invariant prior. A natural question is whether it will yield a proper posterior.

Datta and Ghosh (1991) considered a noninformative prior for a linear mixed model by assuming a constant prior for fixed effects and a prior density of the variance component δ_i of the form

$$\frac{1}{\delta_i^{a_i+1}} \exp(-b_i/\delta_i), \tag{1}$$

where $a_i \in \mathbb{R}$ and $b_i > 0$ (If $a_i > 0$, this is inverse gamma IG (a_i, b_i)). Ghosh, Natarajan, Stroud and Carlin (1998) studied a generalized linear mixed model with a constant prior for fixed effects and priors for random effects precision components that are IG (a_i, b_i) with a small but positive b_i. This is reasonable if one has information on δ_i (cf. Sun and Berger (1998)), but it is clearly suboptimal otherwise. Note that for the inverse gamma when both a_i and b_i tend to 0, the proper priors tend to the scale-invariant prior. It is of interest to know whether we can use the scale-invariant prior rather than the inverse gamma distribution with very small positive parameters. The problem is whether the posterior is proper in using flat prior for the fixed effects, and independent scale-invariant priors for the variance components of the random effects.

The scale-invariant prior for the one-way ANOVA model is a special case of (1) when $a_i = b_i = 0$. Recently, Hobert and Casella (1996) found some conditions on a_i for the propriety of posterior distribution when $b_i = 0$. They assumed the random effects are *a priori* independent. This would not be true in many situations, for example in image processes or in disease mapping over neighboring regions. See, for example, Bernardinelli, Clayton and Montomoli (1995) and Sun, Tsutakawa, Kim and He (2000).

It is important to explore simple noninformative priors for general hierarchical models systematically. In this paper, we study noninformative priors for hierarchical models. Our results include Datta and Ghosh (1991), Hobert and Casella (1996), and Ghosh, Natarajan, Stroud and Carlin (1998) as special cases. Some related work under different sampling plans was considered in Ghosh, Natarajan, Waller and Kim (1999).

The paper is arranged as follows. In Section 2, we consider a general linear mixed model whose random effects may be independent or correlated. The correlations are introduced through a family of distributions which includes as special cases the intrinsic AR model of Besag, York and Mollié (1991), the AR model of Ord (1975), and CAR models of Clayton and Kaldor (1987). We first prove the propriety of the posterior distribution when the fixed effects have a flat prior and the random effects have a proper prior. We then consider the propriety of the posterior distribution when (a1) $b_i > 0$ or (a2) $b_i = 0$ and $a_i < 0$. We give examples to show that the posterior distribution may or may not be proper in the case when $a_i = 0$ and $b_i \geq 0$. We also discuss special cases which may be applicable in practice. In Section 3, we present results which are useful in the implementation of the Gibbs sampler to our model. In Section 4, we first extend the results to the generalized hierarchical model where, conditionally on the parameters $\mathbf{v} = (v_1, \ldots, v_N)'$, the distribution of the observed random variable y_i depends only on v_i and $\mathbf{v}$ satisfies the general linear mixed model. This extension includes cases where the data are Poisson or binomial. We then extend our result to the generalized linear mixed model, where the first stage distribution belongs to an exponential family which may include shape and scale parameters in addition to $\mathbf{v}$.

2. Main Results

2.1. Preliminary

Consider a general linear mixed model

$$v_i = \mathbf{x}'_{1i}\boldsymbol{\theta} + \mathbf{x}'_{2i}\mathbf{u} + e_i, \tag{2}$$

where $\mathbf{v} = (v_1, \ldots, v_n)'$ is an $n \times 1$ vector of data, $\mathbf{X}_1 = (\mathbf{x}_{11}, \ldots, \mathbf{x}_{1n})'$ and $\mathbf{X}_2 = (\mathbf{x}_{21}, \ldots, \mathbf{x}_{2n})'$ are known design matrices, with dimensions $n \times p$ and $n \times q$, $\boldsymbol{\theta}$ is a $p \times 1$ vector of fixed effects, $\mathbf{u}$ is a $q \times 1$ vector of random effects, and

$\mathbf{e} = (e_1, \ldots, e_n)'$ is an $n \times 1$ vector of random errors. A Bayesian hierarchical linear mixed model begins with the assumptions

$$\begin{cases} \mathbf{u}|(\delta_1, \delta_2, \ldots, \delta_r) \sim N_q(\mathbf{0}, \mathbf{A}), \\ \mathbf{e}|\delta_0 \sim N_n(\mathbf{0}, \delta_0 \mathbf{\Sigma}), \end{cases} \tag{3}$$

where $\mathbf{u} = (\mathbf{u}_1', \ldots, \mathbf{u}_r')'$, $\mathbf{u}_i$ is $q_i \times 1$, $\sum_{i=1}^r q_i = q$, and $\mathbf{A} = \oplus_{i=1}^r \delta_i \mathbf{B}_i^{-1}$, where $\oplus_{i=1}^r \delta_i \mathbf{B}_i^{-1} = diag(\delta_1 \mathbf{B}_1^{-1}, \delta_2 \mathbf{B}_2^{-1}, \ldots, \delta_r \mathbf{B}_r^{-1})$. Here r subvectors of $\mathbf{u}$ correspond to the r different random vectors in the experiment, $\mathbf{0}$ is a vector of zeros, $\mathbf{\Sigma}$ is a known positive definite (pd) $n \times n$ matrix, and $\mathbf{B}_i$ is a pd $q_i \times q_i$ matrix. Without loss of generality, we assume that $\mathbf{\Sigma}$ equals $\mathbf{I}_n$, the n-dimensional identity matrix.

It is often the case that $\mathbf{B}_i$ is known, for example, $\mathbf{B}_i = \mathbf{I}_{q_i}$. In other cases, $\mathbf{B}_i$ contains some unknown parameters. In our case, we assume that $\mathbf{B}_i$ has the form,

$$\mathbf{B}_i = (\mathbf{I}_{q_i} - \rho_i \mathbf{C}_i)^{\xi_i}, \tag{4}$$

where $\mathbf{C}_i$ is a known $q_i \times q_i$ symmetric matrix, ξ_i is a known nonnegative integer and ρ_i is an unknown parameter such that $\mathbf{B}_i$ is pd. Let $\lambda_{i1} \le \lambda_{i2} \le \ldots \le \lambda_{iq_i}$ be the ordered eigenvalues of $\mathbf{C}_i$ satisfying $\lambda_{i1} < 0 < \lambda_{iq_i}$. When $1/\lambda_{i1} < \rho_i < 1/\lambda_{iq_i}$, $\mathbf{B}_i$ is positive definite. Note that $\mathbf{B}_i = \mathbf{I}_{q_i}$ if $\rho_i = 0$ so the model considered here is a generalization of independent random effects. We also note that if ρ_i is known, $\mathbf{B}_i$ would also be known and, by a suitable linear transformation on δ_i, our problem reduces to the case of independence. A common choice for $\mathbf{C}_i$ is the adjacency matrix used in spatial analysis, with element $c_{jl} = 1$ if regions j and l are adjacent and $c_{jl} = 0$ otherwise, including the case $j = l$. The common choices for ξ_i are 1 or 2. The former occurs in a CAR(1) model of Clayton and Kaldor (1987), and the latter in the AR model of Ord (1975).

Alternatively, the matrix $\mathbf{B}_i$ can be of the form

$$\mathbf{B}_i = (\mathbf{D}_i - \rho_i \mathbf{C}_i)^{\xi_i}, \tag{5}$$

where $\mathbf{D}_i$ is a known positive definite matrix. A limiting case with $\xi_i = 1$, $\rho = 1$, and $\mathbf{D}_i = diag(d_1, \ldots, d_{q_i})$, used by Besag *et al.* (1991), produces an improper prior. Here d_j is the sum of the jth column of the matrix $\mathbf{C}_i$. For this case, we could redefine $\mathbf{D}_i^{-1/2} = diag(d_1^{-1/2}, \ldots, d_{q_i}^{-1/2})$, $\mathbf{C}_i^* = \mathbf{D}_i^{-1/2} \mathbf{C}_i \mathbf{D}_i^{-1/2}$, $\mathbf{B}_i^* = (\mathbf{I}_{q_i} - \rho_i \mathbf{C}_i^*)^{\xi_i}$. Also, write $\mathbf{X}_2 = (\mathbf{X}_{21}, \ldots, \mathbf{X}_{2r})$, where $\mathbf{X}_{2i}$ is an $n \times q_i$ matrix and define $\mathbf{X}_{2i}^* = \mathbf{X}_{2i} \mathbf{D}_i^{1/2}$, $\mathbf{X}_2^* = (\mathbf{X}_{21}^*, \ldots, \mathbf{X}_{2r}^*)$. The linear mixed model with $\mathbf{B}_i$ defined by (5) is then equivalent to model (4) when $\mathbf{X}_2$ and $\mathbf{B}_i$ are replaced by $\mathbf{X}_2^*$ and $\mathbf{B}_i^*$, respectively.

The following conditional independence assumptions are also used:

(i) given $(\boldsymbol{\theta}, \mathbf{u})$, $\mathbf{v}$ is conditionally independent of $\mathbf{\Delta} \equiv (\delta_1, \ldots, \delta_r)$ and $\boldsymbol{\rho} = (\rho_1, \ldots, \rho_r)$.

(ii) given $(\boldsymbol{\Delta}, \boldsymbol{\rho})$, $\mathbf{u}$ is independent of $\boldsymbol{\theta}$ and δ_0;
(iii) $\boldsymbol{\theta}, \delta_1, \delta_2, \ldots, \delta_r, \rho_1, \rho_2, \ldots, \rho_r$ and δ_0 are mutually independent.
To fully specify the hierarchical model, we must specify the prior distributions of $\boldsymbol{\theta}$, $\boldsymbol{\Delta}, \boldsymbol{\rho}$ and δ_0. The most commonly used priors are multivariate normal for $\boldsymbol{\theta}$ and inverse gammas for the variance components δ_i. Finally, ρ_i is assumed to have an arbitrary distribution function $F_i(\cdot)$ on the interval $(\lambda_{i1}^{-1}, \lambda_{iq_i}^{-1})$. A degenerate prior at 0 for ρ_i has the components of $\mathbf{u}_i$ independent. A uniform distribution on $(\lambda_{i1}^{-1}, \lambda_{iq_i}^{-1})$ might be used if there is vague information on ρ_i.

Assume that $\mathbf{X}_1$ is of full rank so that $\mathbf{X}_1'\mathbf{X}_1$ is invertible. Let $\mathbf{X} = (\mathbf{X}_1, \mathbf{X}_2)$ and $\boldsymbol{\beta} = (\boldsymbol{\theta}', \mathbf{u}')'$. The usual least square estimator of $\boldsymbol{\beta}$ is given by $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{v}$, where $(\mathbf{X}'\mathbf{X})^{-}$ is a generalized inverse of $\mathbf{X}'\mathbf{X}$. If $rank(\mathbf{X}) = p + q$, then $(\mathbf{X}'\mathbf{X})^{-1}$ exists and equals $(\mathbf{X}'\mathbf{X})^{-}$. The usual sum of square errors and the regression sum of squares are

$$\mathrm{SSE} = (\mathbf{v} - \mathbf{X}\hat{\boldsymbol{\beta}})'(\mathbf{v} - \mathbf{X}\hat{\boldsymbol{\beta}}) = \mathbf{v}'\{\mathbf{I}_n - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\}\mathbf{v}, \tag{6}$$

$$\mathrm{SSR} = (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})'\mathbf{X}'\mathbf{X}(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}), \tag{7}$$

respectively. It is well known that both SSE and SSR are invariant for any choice of $(\mathbf{X}'\mathbf{X})^{-}$. In fact, $\mathbf{X}(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'$ is invariant for any choice of $(\mathbf{X}'\mathbf{X})^{-}$. Since $rank(\mathbf{X}_1) = p$, we know that $(\mathbf{X}_1'\mathbf{X}_1)^{-1}$ exists and we may define

$$\mathbf{R}_1 = \mathbf{I}_n - \mathbf{X}_1(\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1', \tag{8}$$

$$\mathbf{R}_2 \equiv \mathbf{R}_2(\delta_0, \ldots, \delta_r) = \frac{1}{\delta_0}\mathbf{X}_2'\mathbf{R}_1\mathbf{X}_2 + \mathbf{A}^{-1}, \tag{9}$$

$$\mathbf{R}_3 \equiv \mathbf{R}_3(\delta_0, \ldots, \delta_r) = \mathbf{X}_2'\mathbf{R}_1\mathbf{X}_2 - \frac{1}{\delta_0}\mathbf{X}_2'\mathbf{R}_1\mathbf{X}_2\mathbf{R}_2^{-1}\mathbf{X}_2'\mathbf{R}_1\mathbf{X}_2. \tag{10}$$

$$\mathbf{L}_1 \equiv \mathbf{L}_1(\mathbf{u}) = (\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{X}_2(\hat{\mathbf{u}} - \mathbf{u}), \tag{11}$$

$$\mathbf{L}_2 \equiv \mathbf{L}_2(\delta_0, \ldots, \delta_r) = \frac{1}{\delta_0}\mathbf{R}_2^{-1}\mathbf{X}_2'\mathbf{R}_1\mathbf{X}_2\hat{\mathbf{u}}. \tag{12}$$

Since $\mathbf{X}_2'\mathbf{R}_1\mathbf{X}_2$ is nonnegative definite and $\mathbf{A}$ is pd, $\mathbf{R}_2$ is pd and $\mathbf{R}_2^{-1}$ exists. Also let $t = rank(\mathbf{R}_1\mathbf{X}_2) = rank(\mathbf{X}_2'\mathbf{R}_1\mathbf{X}_2)$. Then we have $t \le q$.

2.2. Vague priors for fixed effects

Our first result is on the propriety of the joint posterior when the noninformative prior is only on the fixed effects $\boldsymbol{\theta}$. The proof is similar to that in Searle, Casella and McCulloch (1992, Section 9.2), which did not include $\boldsymbol{\rho}$ and $\boldsymbol{\Delta}$. We give the proof in the appendix for completeness.

Theorem 1. *Assume that the density of the fixed effect $\boldsymbol{\theta}$ is proportional to a constant and that the priors of $\boldsymbol{\Delta}, \boldsymbol{\rho}$ and δ_0 are proper. Let*

$$G \equiv G(\boldsymbol{\theta}, \boldsymbol{u}, \boldsymbol{\Delta}, \boldsymbol{\rho}; \boldsymbol{v}) = \frac{1}{\delta_0^{\frac{n}{2}}|\boldsymbol{A}|^{\frac{1}{2}}}\exp\Big\{-\frac{(\boldsymbol{v} - \boldsymbol{X}\boldsymbol{\beta})'(\boldsymbol{v} - \boldsymbol{X}\boldsymbol{\beta})}{2\delta_0} - \frac{\boldsymbol{u}'\boldsymbol{A}^{-1}\boldsymbol{u}}{2}\Big\}\prod_{i=0}^{r} g_i(\delta_i). \tag{13}$$

Then, we have

$$\int_{\mathbb{R}^{q+p}} G d\boldsymbol{\theta} d\boldsymbol{u} = \frac{(2\pi)^{\frac{1}{2}(p+q)}|\boldsymbol{X}_1'\boldsymbol{X}_1|^{-\frac{1}{2}}}{\delta_0^{\frac{1}{2}(n-p)}|\boldsymbol{R}_2|^{\frac{1}{2}}|\boldsymbol{A}|^{\frac{1}{2}}} \exp\Big\{-\frac{SSE}{2\delta_0} - \frac{\hat{\boldsymbol{u}}'\boldsymbol{R}_3\hat{\boldsymbol{u}}}{2\delta_0}\Big\}\prod_{i=0}^{r} g_i(\delta_i), \qquad (14)$$

which implies that the joint posterior distribution of $(\boldsymbol{\theta}, \boldsymbol{u}, \boldsymbol{\Delta}, \boldsymbol{\rho}, \delta_0)$ *is proper.*

2.3. Improper priors for both fixed effects and variance components

In addition to a flat prior on $\boldsymbol{\theta}$, we want to have noninformative priors on $(\delta_0, \boldsymbol{\Delta})$. We unify the results of Datta and Ghosh (1991) and Hobert and Casella (1996), while extending their results to the case $\rho_i \neq 0$. We consider the following conditions.

(a) For $i = 1, \ldots, r$, either of the two conditions holds:
  (a1) $a_i < b_i = 0$; (a2) $b_i > 0$;
(b1) $q_i + 2a_i > 0$.
(b2) $q_i + 2a_i > q - t$, for all $i = 1, \ldots, r$;
(c1) $n - p + 2a_0 + 2a_+ > 0$, where $a_+ = \sum_{i=1}^r a_i$;
(c2) $n - p + 2a_0 + 2\sum_{i=1}^r a_i^- > 0$, where $a_i^- = \min(0, a_i)$.

Theorem 2. *Suppose that the prior density of* $\boldsymbol{\theta}$ *is constant, that the prior density of* δ_i *is given at* (1)*, and that* ρ_i *has an arbitrary distribution* F_i*. Assume that* $2b_0 + SSE > 0$.
Case 1. *If* $t = q$ *or if* $r = 1$ *the conditions* (*a*)*,* (*b*2)*, and* (*c*1) *are necessary, and conditions* (*a*)*,* (*b*2) *and* (*c*2) *are sufficient for the propriety of the posterior distribution of* $(\boldsymbol{\theta}, \boldsymbol{u}, \boldsymbol{\Delta}, \boldsymbol{\rho}, \delta_0)$.
Case 2. *If* $t < q$ *and* $r > 1$*, conditions* (*a*)*,* (*b*1) *and* (*c*1) *are necessary, and conditions* (*a*)*,* (*b*2) *and* (*c*2) *are sufficient for the propriety of the joint posterior.*

The proof of this theorem is in the Appendix. There are two simple cases for the values of b_0. Since SSE ≥ 0, when $b_0 > 0$, $2b_0+$ SSE > 0. Note that SSE > 0 with probability one. When $b_0 = 0$, we have $2b_0 + \text{SSE} > 0$ with probability one so that the results will be true for almost all observations. Furthermore, if all $a_i \leq 0, i = 1, \ldots, r$, then conditions (c1) and (c2) are identical.

In Theorems 1 and 2, we assume a constant prior for the fixed effects $\boldsymbol{\theta}$. In practice, one could have a proper or partially informative prior (cf. Sun, Tsutakawa and Speckman (1999)), proper in some subspaces and constant in others. In these cases, prior densities are all bounded. Our results assuming a constant prior for $\boldsymbol{\theta}$ will remain true under such generalization.

2.4. Special cases

Case 1. Datta and Ghosh (1991) considered the case $b_i > 0$ for all $i = 1, \ldots, r$. From Theorem 2 we know that when all $b_i > 0$, (b1) and (c1) are necessary, while (b2) and (c2) are sufficient for a proper posterior. Clearly, (b2) is stronger than (b1) and (c2) is stronger than (c1). Our conditions for proper posterior for the special case where $\mathbf{B}_i = \mathbf{I}_{q_i}$ are slightly different from those of Datta and Ghosh (1991). Here are examples showing that the constraint $b_i > 0$ and $a_i = 0$ may or may not result in a proper posterior distribution.

Example 1. Consider the linear mixed model (2), with $n = 6, p = 2, r = 3, q_1 = q_2 = q_3 = 1, q = q_1+q_2+q_3 = 3$, $\boldsymbol{\theta} = (\theta_1, \theta_2)'$, and $\mathbf{u} = (u_1, u_2, u_3)'$. The matrices $\mathbf{X}_1$, $\mathbf{X}_2$, and $\mathbf{R}_1$ are given by

$$\mathbf{X}_1 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 0 \\ 0 & 1 \\ 1 & 0 \\ 0 & 1 \end{pmatrix}, \ \mathbf{X}_2 = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{pmatrix}, \ \text{ and } \mathbf{R}_1 = \frac{1}{3} \begin{pmatrix} 2 & 0 & -1 & 0 & -1 & 0 \\ 0 & 2 & 0 & -1 & 0 & -1 \\ -1 & 0 & 2 & 0 & -1 & 0 \\ 0 & -1 & 0 & 2 & 0 & -1 \\ -1 & 0 & -1 & 0 & 2 & 0 \\ 0 & -1 & 0 & -1 & 0 & 2 \end{pmatrix}.$$

Then $\mathbf{X}_2'\mathbf{R}_1\mathbf{X}_2$ is 3×3 matrix, whose three rows are given by $\frac{1}{3}(4, -2, -2)$, $\frac{1}{3}(-2, 4, -2)$ and $\frac{1}{3}(-2, -2, 4)$. Thus $t = rank(\mathbf{X}_2'\mathbf{R}_1\mathbf{X}_2) = 2$. We assume that the prior for (θ_1, θ_2) is constant, $b_0, b_1, b_2, b_3 > 0$, and $a_0 = a_1 = a_2 = a_3 = 0$. Condition (a) in Case 2 of Theorem 2. holds, and so does condition (c2) since $n - p - 2a_0 = 6 - 2 > 2 > 0$. However, condition (b2) fails. In this case, $\mathbf{A} = diag(\delta_1, \delta_2, \delta_3)$ and $|\mathbf{R}_2||\mathbf{A}| = |\delta_0^{-1}\mathbf{A}\mathbf{X}_2'\mathbf{R}_1\mathbf{X}_2 + \mathbf{I}_3|$, which equals

$$1 + \frac{4}{3}\left(\frac{\delta_1}{\delta_0} + \frac{\delta_2}{\delta_0} + \frac{\delta_3}{\delta_0} + \frac{\delta_1\delta_2}{\delta_0^2} + \frac{\delta_1\delta_3}{\delta_0^2} + \frac{\delta_2\delta_3}{\delta_0^2}\right) \geq \frac{4}{3}\frac{(\delta_1+\delta_2)\delta_3}{\delta_0^2} \ \geq \ \frac{8}{3\delta_0^2}\sqrt{\delta_1\delta_2}\,\delta_3.$$

Then for G defined by (13), we have

$$\int_0^\infty \cdots \int_0^\infty \int_{\mathbb{R}^{p+q}} G d\boldsymbol{\theta} d\mathbf{u} \prod_{i=0}^3 d\delta_i$$
$$\leq \int_0^\infty \cdots \int_0^\infty \frac{|\mathbf{X}_1'\mathbf{X}_1|^{-\frac{1}{2}} \exp\left\{-\frac{\text{SSE}}{2\delta_0} - \sum_{i=0}^3 \frac{b_i}{\delta_i}\right\}}{(2\pi)^{-\frac{1}{2}(p+q)} \delta_0^{\frac{1}{2}(n-p-2)} \delta_1^{\frac{1}{4}+1} \delta_2^{\frac{1}{4}+1} \delta_3^{\frac{1}{2}+1}} \prod_{i=0}^3 d\delta_i,$$

which is finite. In this case, necessary conditions (a), (b1) and (c2) in Case 2 of Theorem 2 hold, but a sufficient condition (b2) fails. However, the posterior distribution of $(\boldsymbol{\theta}, \mathbf{u}, \delta_0, \delta_1, \delta_2, \delta_3)$ is proper.

Example 2. Suppose that we have the same $\mathbf{X}_1$, but $\mathbf{X}_2 = (\mathbf{x}_{21}, \mathbf{x}_{22}, \mathbf{x}_{23})$, where $\mathbf{x}_{21} = (-1, 2, -1, 2, -1, 2)$, $\mathbf{x}_{22} = (-1, 0, -1, 0, 2, 0)$, $\mathbf{x}_{23} = (0, 2, 0, -1, 0, -1)$. The matrix $\mathbf{R}_1$ remains the same, but $\mathbf{X}_2'\mathbf{R}_1\mathbf{X}_2 = diag(0, 6, 6)$. So $|\mathbf{R}_2||\mathbf{A}| = (1 + \delta_2/\delta_0)(1 + \delta_3/\delta_0)$ and

$$G_1 = \left\{\delta_0\delta_1\delta_2\delta_3\sqrt{\left(\frac{\delta_2}{\delta_0}+1\right)\left(\frac{\delta_3}{\delta_0}+1\right)}\right\}^{-1} \exp\Big(-\sum_{i-0}^3 \frac{b_i}{\delta_i}\Big).$$

Since $\int_0^\infty \delta_1^{-1} \exp\{-b_1/\delta_1\} d\delta_1 = \infty$, $\int_0^\infty \int_0^\infty \int_0^\infty G_1 d\delta_1 d\delta_2 d\delta_3 = \infty$. So necessary conditions (a), (b1) and (c2) in Case 2 of Theorem 2 hold, but a sufficient condition (b2) fails. However, the posterior distribution of $(\boldsymbol{\theta}, \mathbf{u}, \delta_0, \delta_1, \delta_2, \delta_3)$ is improper.

Case 2. Hobert and Casella (1996) consider the case where $b_i \equiv 0$ and $\rho_i \equiv 0$ for all i, a special case of Theorem 2. Examine again the balanced one-way ANOVA discussed in Section 1. The prior $p(\mu, \delta_0, \delta_1) \propto (\delta_0\delta_1)^{-1}$ yields an improper posterior. The limiting case of the Jeffreys prior, as $n \to \infty$, $p(\mu, \delta_0, \delta_1) \propto \delta_0^{-1}\delta_1^{-3/2}$, will also end up with an improper posterior.

Case 3. The constant prior has been used in the literature. For example, Yang and Chen (1995) used constant priors for a random effects model. In general we could have

$$p(\boldsymbol{\theta}, \mathbf{u}, \boldsymbol{\Delta}, \boldsymbol{\rho}, \delta_0) \propto 1. \tag{15}$$

This is a special case here with $a_i = -1$ and $b_i = 0$ in (1). Note that under this constant prior, the restricted MLE of $(\boldsymbol{\Delta}, \boldsymbol{\rho}, \delta_0)$ is the marginal posterior mode. In this case, conditions (c1) and (c2) are identical. The following corollary is immediate.

Corollary 1. *Assume the prior defined by* (15) *is used. In addition, if* $q_i - 2 > q - t$ *and* $n - p - 2(r+1) > 0$*, the joint posterior distribution of* $(\boldsymbol{\theta}, \boldsymbol{u}, \boldsymbol{\Delta}, \boldsymbol{\rho}, \delta_0)$ *exists.*

Case 4. Another interesting improper prior for δ_i is of the form, $g_i(\delta_i) \propto \delta_i^{-1/2}$. This is a special case of (1) with $a_i = -1/2$ and $b_i = 0$.

Corollary 2. *Assume that* $g_i(\delta_i) \propto \delta_i^{-1/2}$*. Under the assumptions of Theorem* 2*, if* $q_i - 1 > q - t$ *and* $n - p - (r+1) > 0$*, the joint posterior distribution of* $(\boldsymbol{\theta}, \boldsymbol{u}, \boldsymbol{\Delta}, \boldsymbol{\rho}, \delta_0)$ *exists.*

The propriety of the posterior distribution in terms of the hyperparameters (a_i, b_i) $(i = 1, \ldots, r)$, may now be summarized according to the following table.

Table 1. Propriety of the posterior distribution.

	$a_i < 0$	$a_i = 0$	$a_i > 0$
$b_i = 0$	proper	improper	improper
$b_i > 0$	proper	proper or improper	proper

3. Computation of Posterior Distributions

We may use the Gibbs sampling procedure (cf. Gelfand and Smith (1990)) to get the posterior distribution. Here are the full conditional distributions for this purpose.

Proposition 1. *Suppose that the prior distributions of ρ_i is uniform on the interval $(\lambda_{i1}, \lambda_{iq_i})$. The full conditional distributions are given as follows:*

1. $\boldsymbol{\theta}|(\boldsymbol{u}, \delta_0, \boldsymbol{\Delta}, \boldsymbol{\rho}, \boldsymbol{v}) \sim MVN_p(\hat{\boldsymbol{\theta}} + \boldsymbol{L}_1, \delta_0(\boldsymbol{X}_1'\boldsymbol{X}_1)^{-1})$, *where $\boldsymbol{L}_1$ is given by* (11).
2. $\boldsymbol{u}|(\boldsymbol{\theta}, \delta_0, \boldsymbol{\Delta}, \boldsymbol{\rho}, \boldsymbol{v}) \sim MVN_q(\delta_0^{-1}\boldsymbol{M}_1\boldsymbol{X}_2'(\boldsymbol{v} - \boldsymbol{X}_1\boldsymbol{\theta}), \boldsymbol{M}_1)$, *where* $\boldsymbol{M}_1 = (\delta_0^{-1}\boldsymbol{X}_2'\boldsymbol{X}_2 + \boldsymbol{A}^{-1})^{-1}$.
3. $\delta_0|(\boldsymbol{\theta}, \boldsymbol{u}, \boldsymbol{\Delta}, \boldsymbol{\rho}, \boldsymbol{v}) \sim IG\ (a_0 + \frac{n}{2},\ b_0 + \frac{1}{2}(\boldsymbol{v} - \boldsymbol{X}_1\boldsymbol{\theta} - \boldsymbol{X}_2\boldsymbol{u})'(\boldsymbol{v} - \boldsymbol{X}_1\boldsymbol{\theta} - \boldsymbol{X}_2\boldsymbol{u}))$.
4. $\delta_i|(\boldsymbol{\theta}, \boldsymbol{u}, \boldsymbol{\Delta}_{(-i)}, \boldsymbol{\rho}, \boldsymbol{v}) \sim IG\ (a_i + \frac{q_i}{2},\ b_i + \frac{1}{2}\boldsymbol{u}_i'\boldsymbol{B}_i\boldsymbol{u}_i)$, *for* $i = 1, \ldots, r$. *Here* $\boldsymbol{\Delta}_{(-i)} = (\delta_1, \ldots, \delta_{i-1}, \delta_{i+1}, \ldots, \delta_r)$.
5. *let* $\boldsymbol{\rho}_{(-i)} = (\rho_1, \ldots, \rho_{i-1}, \rho_{i+1}, \ldots, \rho_r)$. *The conditional density of ρ_i given* $(\boldsymbol{\theta}, \boldsymbol{u}, \boldsymbol{\Delta}, \boldsymbol{\rho}_{(-i)}, \boldsymbol{v})$ *is of the form*

$$h_i(\rho_i) = |\boldsymbol{B}_i|^{q_i/2} \exp\Big\{ -\frac{1}{2\delta_i}\boldsymbol{u}_i'\boldsymbol{B}_i\boldsymbol{u}_i\Big\}.$$

Sampling from the first three conditional distributions is straightforward. Note that if $\xi_i = 0$, the parameter ρ_i does not effect other parameters. We will consider only the case where $q_i > 1$ and $\xi_i \geq 1$. Here is a property of logconcavity for the conditional density of ρ_i. From this, the adaptive rejection sampling method by Gilks and Wild (1992) can be used.

Lemma 1. *If $\xi_i \geq 1$, the conditional density of ρ_i is log-concave.*

Proof. Define $\mathbf{\Lambda}_i = diag(\lambda_{i1}, \ldots, \lambda_{iq_i})$. Let $\mathbf{\Gamma}_i$ be an $q_i \times q_i$ orthogonal matrix such that $\mathbf{C}_i = \mathbf{\Gamma}_i\mathbf{\Lambda}_i\mathbf{\Gamma}_i'$. Then $\mathbf{B}_i = \mathbf{\Gamma}_i(\mathbf{I}_{q_i} - \rho_j\mathbf{\Lambda}_i)^{\xi_i}\mathbf{\Gamma}_i'$ and

$$h_i(\rho_i) = \prod_{j=1}^{q_i}(1 - \rho_i\lambda_{ij})^{\xi_i/2} \exp\Big\{ -\frac{1}{2\delta_i}\mathbf{u}_i'\mathbf{\Gamma}_i(\mathbf{I}_{q_i} - \rho_i\mathbf{\Lambda}_i)^{\xi_i}\mathbf{\Gamma}_i'\mathbf{u}_i\Big\}.$$

Let $\mathbf{w} = (w_1, \ldots, w_{q_i})' = \mathbf{\Gamma}_i'\mathbf{u}_i$. Then

$$\frac{\partial^2}{\partial\rho_i^2}\log[h(\rho_i)] = -\frac{\xi_i}{2}\sum_{j=1}^{q_i}\lambda_{ij}^2\Big\{(1 - \rho_i\lambda_{ij})^{-2} + \frac{\xi_i - 1}{\delta_i}(1 - \rho_i\lambda_{ij})^{\xi_i - 2}w_j^2\Big\},$$

which is negative in the range $(\lambda_{i1}^{-1}, \lambda_{iq_i}^{-1})$. The result follows.

4. Extension to Generalized Hierarchical Models

4.1. One-parameter families

Consider the hierarchical model where $y_1, \ldots, y_N$ are conditionally independent given parameters $\mathbf{v} = (v_1, \ldots, v_N)$, and v_i follows the hierarchical prior defined by (2)–(4). Let $f_i(y_i|v_i)$ be the density of y_i given v_i.

Theorem 3. *Suppose the following conditions hold.*

(A) *v_i follows the linear mixed model* (2), *where the fixed effect $\boldsymbol{\theta}$ has a constant prior, the random effects $\boldsymbol{u}$ and $\boldsymbol{e}$ follow the distribution* (3), *the variance components δ_i have the prior* (1), *and the prior for ρ_i is arbitrary.*

(B) *There exists a subset of $\{1, \ldots, N\}$, say $\mathcal{J}_n = (i_1, \ldots, i_n)$, so that $p + q \leq n \leq N$ and the following conditions hold.*

(B1) *If $\boldsymbol{X}_j = (\boldsymbol{x}_{j1}, \ldots, \boldsymbol{x}_{j,N})'$ and $\boldsymbol{X}_j^* = (\boldsymbol{x}_{ji_1}, \ldots, \boldsymbol{x}_{j,i_n})'$. $rank(\boldsymbol{X}_1) = rank(\boldsymbol{X}_1^*) = p$ and $rank(\boldsymbol{X}_2) = rank(\boldsymbol{X}_2^*)$,*

(B2) *There is a constant $M > 0$ such that*

$$\begin{cases} \int f_j(y_j|v_j)dv_j < \infty, & j \in \mathcal{J}_n; \\ f_j(y_j|v_j) \leq M, & j \notin \mathcal{J}_n. \end{cases} \tag{16}$$

If $b_0 > 0$ and the conditions (*a*), (*b*2) *and* (*c*2) *in Theorem* 2 *hold, the posterior distribution of $(\boldsymbol{v}, \boldsymbol{\theta}, \boldsymbol{u}, \boldsymbol{\Delta}, \boldsymbol{\rho}, \delta_0)$ given $\boldsymbol{y} = (y_1, \ldots, y_N)$ is proper.*

The proof can be found in the appendix.

Remark 1. We now discuss the relationship between Theorem 3 and the sufficient part of Theorem 2. The relationship between Theorem 3 and Theorem 1 can be stated similarly. Note that the product of the likelihood of $\mathbf{v}$ for the linear mixed model (2) and the prior of $(\boldsymbol{\theta}, \mathbf{u}, \boldsymbol{\Delta}, \boldsymbol{\rho}, \delta_0)$, defined by (3)-(4), is the joint prior of $(\mathbf{v}, \boldsymbol{\theta}, \mathbf{u}, \boldsymbol{\Delta}, \boldsymbol{\rho}, \delta_0)$. From Theorem 2, if $b_0 + 0.5\text{SSE} > 0$, the sufficient conditions for the propriety of the posterior of $(\boldsymbol{\theta}, \mathbf{u}, \boldsymbol{\Delta}, \boldsymbol{\rho}, \delta_0)$ given $\mathbf{v}$ are (a), (b2) and (c2). Here since SSE is defined by (6), depending on $\mathbf{v}$, is often positive, we can choose $b_0 = 0$. Since n is often large enough, we might be able to choose $a_0 = 0$ so that (c2) holds. In such a case, we could use the improper prior $1/\delta_0$ for δ_0. Now for any one-parameter family, $f_i(y_i|v_i)$, in addition to the rank assumption (B1) on the design matrices and integrability assumptions on f_i, we need $b_0 > 0$, and conditions (a), (b2) and (c2) to insure the propriety of $(\mathbf{v}, \boldsymbol{\theta}, \mathbf{u}, \boldsymbol{\Delta}, \boldsymbol{\rho}, \delta_0)$. It is still an interesting open question whether we can relax the constraint $b_0 > 0$.

4.2. Generalized linear mixed models

We now consider generalized linear mixed models where $y_1, \ldots, y_N$ are independent random observations with the probability density

$$f_i(y_i|\eta_i, \phi) = \exp[s_i(\phi)^{-1}\{y_i\eta_i - \psi_i(\eta_i)\} + \gamma_i(y_i; \phi)]. \tag{17}$$

The function $s_i(\phi)$ is commonly of the form $s_i(\phi) = \phi w_i^{-1}$, where the w_i are prespecified weights. We wish to model the variability in η_i to account for various fixed covariates and random effects via (2), where $v_i = h_i(\eta_i)$. Note that for $\psi_i(\cdot)$

defined in (17), the first derivative $d\psi_i(\eta_i)/d\eta_i$ is a strictly increasing function. Let K_i be the inverse function of $d\psi_i(\eta_i)/d\eta_i$. Note that for any fixed scale parameter ϕ, the likelihood function $f_i(y_i|\eta_i,\phi)$ is bounded by

$$M_i(\phi) \equiv \sup_{\eta_i} f_i(y_i|\eta_i,\phi) \;=\; \exp[s_i(\phi)^{-1}\{y_iK_i(y_i) - \psi_i(K_i(y_i))\} + \gamma_i(y_i;\phi)]. \tag{18}$$

Using arguments similar to those for Theorem 3, we have the following result.

Theorem 4. *Assume that the conditions* (A) *and* (B1) *of Theorem* 3 *hold, and that for any prior density* p_0 *(proper or improper) of* ϕ*,*

$$\int \prod_{j\notin\mathcal{J}_n} M_j(\phi)\bigg\{\prod_{j\in\mathcal{J}_n}\int f_j(y_j|\eta_j,\phi)\frac{d}{d\eta_j}h_j(\eta_j)d\eta_j\bigg\}p_0(\phi)d\phi < \infty. \tag{19}$$

If $b_0 > 0$ *and the conditions* (a), (b2) *and* (c2) *in Theorem* 2 *hold, the posterior distribution of* $(\phi, \boldsymbol{v}, \boldsymbol{\theta}, \boldsymbol{u}, \boldsymbol{\Delta}, \boldsymbol{\rho}, \delta_0)$ *given* $\boldsymbol{y} = (y_1, \ldots, y_N)$ *is proper.*

Remark 2. As in Remark 1, we discuss the relationship between Theorem 4 and the sufficiency part of Theorem 2. For a generalized linear mixed model (17), we need rank assumption (B1) on the design matrices, the integrability assumptions (19) on the likelihood function and the marginal prior of the scale parameter ϕ, the assumption $b_0 > 0$, and the sufficient conditions (a), (b2) and (c2) in Theorem 2 to insure the propriety of $(\phi, \mathbf{v}, \boldsymbol{\theta}, \mathbf{u}, \boldsymbol{\Delta}, \boldsymbol{\rho}, \delta_0)$.

One important example has $y_i \sim N(\mu_i, \sigma^2)$. In this case, $\eta_i = \mu_i$, $\phi = \sigma^2$ and $A_i(\phi) = \phi$. It is easy to see that $M_i(\phi) = 1/\sqrt{2\pi\phi}$ and $\int f_i(y_i|\eta_i,\phi)d\eta_i = 1$. Condition (19) becomes $\int_0^\infty \phi^{-\frac{1}{2}(N-n)}p_0(\phi)d\phi < \infty$, which always holds when $N = n$ and F is a proper prior for ϕ.

Another example has $y_i \sim \text{gamma}(\alpha, \alpha/\mu_i)$. In this case, α is the shape parameter and μ_i is the mean of y_i for given (μ_i, α). So $\phi = \alpha$, $\eta_i = 1/\mu_i$, $A_i(\phi) = -1/\phi$, and $B(\eta_i) = \log(\eta_i)$. Choose $h_i(\eta_i) = \log(\eta_i) = -\log(\mu_i)$. Then $M_i(\phi) = y_i^{-1}\alpha^\alpha e^{-\alpha}/\Gamma(\alpha)$ and $\int_0^\infty f(y_i|\eta_i,\phi)\eta_i^{-1}d\eta_i = y_i^{-1}$. If $N = n$ and ϕ has a proper prior, condition (19) holds.

Acknowledgement

Sun's work was partially supported by the National Science Foundation grant DMS-9972598. He's research was partially supported by Federal Aid in Wildlife Restoration Project W-13-R. The authors would like to thank Hoon Kim and Roger Woodard for carely reading the earlier version of this paper. The authors wish to thank two referees for several helpful comments.

Appendix. Proofs and Lemmas

Proof of Theorem 1. We use the familiar decomposition $(\mathbf{v}-\mathbf{X}\boldsymbol{\beta})'(\mathbf{v}-\mathbf{X}\boldsymbol{\beta}) =$ SSE+SSR, where SSE and SSR are defined by (6) and (7), respectively. In addition, SSR has the decomposition SSR $= (\boldsymbol{\theta}-\hat{\boldsymbol{\theta}}-\mathbf{L}_1)'\mathbf{X}_1'\mathbf{X}_1(\boldsymbol{\theta}-\hat{\boldsymbol{\theta}}-\mathbf{L}_1) + (\mathbf{u}-\hat{\mathbf{u}})'\mathbf{X}_2'\mathbf{R}_1\mathbf{X}_2(\mathbf{u}-\hat{\mathbf{u}})$. Then,

$$\int_{\mathbb{R}^p} G d\boldsymbol{\theta} = \frac{(2\pi)^{\frac{p}{2}}|\mathbf{X}_1'\mathbf{X}_1|^{-\frac{1}{2}}}{\delta_0^{\frac{1}{2}(n-p)}|\mathbf{A}|^{\frac{1}{2}}} \cdot \exp\Big\{-\frac{\text{SSE}}{2\delta_0} - \frac{(\mathbf{u}-\hat{\mathbf{u}})'\mathbf{X}_2'\mathbf{R}_1\mathbf{X}_2(\mathbf{u}-\hat{\mathbf{u}})}{2\delta_0} - \frac{\mathbf{u}'\mathbf{A}^{-1}\mathbf{u}}{2}\Big\}\prod_{i=0}^{r} g_i(\delta_i).$$

Since $\mathbf{X}_2'\mathbf{R}_1\mathbf{X}_2$ is nonnegative definite and $\mathbf{A}$ is pd, $\mathbf{R}_2$ is pd and $\mathbf{R}_2^{-1}$ exists. Since

$$\delta_0^{-1}(\mathbf{u}-\hat{\mathbf{u}})\mathbf{X}_2'\mathbf{R}_1\mathbf{X}_2(\mathbf{u}-\hat{\mathbf{u}}) + \mathbf{u}'\mathbf{A}^{-1}\mathbf{u} = (\mathbf{u}-\mathbf{L}_2)'\mathbf{R}_2(\mathbf{u}-\mathbf{L}_2) + \delta_0^{-1}\hat{\mathbf{u}}'\mathbf{R}_3\hat{\mathbf{u}},$$

(14) follows immediately. Since $\mathbf{R}_3$ is nonnegative definite, the lower bound of $\hat{\mathbf{u}}'\mathbf{R}_3\hat{\mathbf{u}}$ is 0, or $\exp\{-\hat{\mathbf{u}}'\mathbf{R}_3\hat{\mathbf{u}}/(2\delta_0)\} \le 1$. Note that $\mathbf{X}_2'\mathbf{R}_1\mathbf{X}_2$ is nonnegative definite, and $|\mathbf{R}_2|\,|\mathbf{A}| \ge 1$. Therefore

$$\int_{\mathbb{R}^q}\int_{\mathbb{R}^p} G d\boldsymbol{\theta} d\mathbf{u} \le \frac{(2\pi)^{\frac{1}{2}(p+q)}|\mathbf{X}_1'\mathbf{X}_1|^{-\frac{1}{2}}}{\delta_0^{\frac{1}{2}(n-p)}}\exp\Big\{-\frac{\text{SSE}}{2\delta_0}\Big\}\prod_{i=0}^{r} g_i(\delta_i). \tag{A.1}$$

This proves the theorem.

Lemma 2. (Marshall and Olkin (1979)). *Assume that two $\nu\times\nu$ symmetric matrices $\boldsymbol{S}_1$ and $\boldsymbol{S}_2$ are both nonnegative definite. Let $\lambda_1(\boldsymbol{S}_i) \le \lambda_2(\boldsymbol{S}_i) \le \ldots \lambda_\nu(\boldsymbol{S}_i)$ be the eigenvalues of $\boldsymbol{S}_i$. Then*

$$\prod_{j=1}^{\nu}[\lambda_j(\boldsymbol{S}_1)+\lambda_j(\boldsymbol{S}_2)] \le |\boldsymbol{S}_1+\boldsymbol{S}_2| \le \prod_{j=1}^{\nu}[\lambda_j(\boldsymbol{S}_1)+\lambda_{\nu-j+1}(\boldsymbol{S}_2)].$$

Lemma 3. *Let $\lambda_1 \le \lambda_2 \le \cdots \le \lambda_\nu$ be real valued constants satisfying $\lambda_1 < 0 < \lambda_\nu$. Let ρ be a random variable on $(\lambda_1^{-1}, \lambda_\nu^{-1})$ with cumulative distribution function $F(\cdot)$.*

(a) *If integers t_1 and t_2 satisfy $1 \le t_1 \le t_2 \le \nu$, let $\beta = \frac{1}{2}(t_2 - t_1 + 1)$. Then for any real b,*

$$J \equiv \int_{(\lambda_1^{-1},\lambda_\nu^{-1})}\int_0^\infty \frac{1}{s^{\alpha+1}}\left\{\prod_{j=t_1}^{t_2}\frac{(1-\rho\lambda_j)^\xi}{s+(1-\rho\lambda_j)^\xi}\right\}^{\frac{1}{2}} e^{-b/s} ds\, F(d\rho) < \infty \tag{A.2}$$

for any nonnegative integer ξ *if and only if one of the following two sets of conditions holds:*

$$\alpha < b = 0 \text{ and } \alpha + \beta > 0; \tag{A.3}$$

$$b > 0 \text{ and } \alpha + \beta > 0. \tag{A.4}$$

(b) *Under condition* (A.3) *we have the following boundaries for* J.

$$\kappa_1(\lambda_1, \lambda_\nu, -\alpha\xi) \leq \frac{J}{Beta(-\alpha, \ \alpha+\beta)} \leq \kappa_u(\lambda_1, \lambda_\nu, -\alpha\xi), \tag{A.5}$$

where for $\tau > 0$,

$$\kappa_1(\lambda_1, \lambda_\nu, \tau) = \int_{(\lambda_1^{-1}, 0]} (1-\rho\lambda_1)^\tau F(d\rho) + \int_{(0, \lambda_\nu^{-1})} (1-\rho\lambda_\nu)^\tau F(d\rho); \tag{A.6}$$

$$\kappa_u(\lambda_1, \lambda_\nu, \tau) = \Big(1 - \frac{\lambda_\nu}{\lambda_1}\Big)^\tau + \Big(1 - \frac{\lambda_1}{\lambda_\nu}\Big)^\tau. \tag{A.7}$$

Here $Beta(\cdot, \cdot)$ *is the beta function.*

(c) *Under condition* (A.4)*we have the following boundaries for* J.

$$\frac{\kappa_2(\lambda_1, \lambda_\nu, \xi, \beta)}{e^b(\alpha+\beta)} \leq J \leq \begin{cases} \kappa_u(\lambda_1, \lambda_\nu, -\alpha\xi) Beta(-\alpha, \ \alpha+\beta), & \text{if } b > 0 \text{ and } \alpha \in (-\beta, 0); \\ \Gamma(\alpha) b^{-\alpha}, & \text{if } b > 0 \text{ and } \alpha > 0; \\ \kappa_u(\lambda_1, \lambda_\nu, \eta\xi)\Gamma(\eta) b^{-\eta}, & \text{if } b > 0 \text{ and } \alpha = 0, \end{cases} \tag{A.8}$$

where η *is an arbitrary constant such that* $0 < \eta < \beta$, *and*

$$\kappa_2(\lambda_1, \lambda_\nu, \xi, \beta) = \int_{(\lambda_1^{-1}, 0]} \left[\frac{(1-\rho\lambda_1)^\xi}{1 + (1-\rho\lambda_1)^\xi}\right]^\beta F(d\rho) + \int_{(0, \lambda_\nu^{-1})} \left[\frac{(1-\rho\lambda_\nu)^\xi}{1 + (1-\rho\lambda_\nu)^\xi}\right]^\beta F(d\rho). \tag{A.9}$$

Proof of Lemma 3. We have

$$0 < 1 - \rho\lambda_1 \leq 1 - \rho\lambda_2 \leq \ldots \leq 1 - \rho\lambda_\nu, \quad \text{for any} \quad \rho \in (\lambda_1^{-1}, 0];$$
$$1 - \rho\lambda_1 \geq 1 - \rho\lambda_2 \geq \ldots \geq 1 - \rho\lambda_\nu > 0, \quad \text{for any} \quad \rho \in (0, \lambda_\nu^{-1}).$$

Note that for any $s > 0$, $f(x) = x/(s+x)$ is increasing in $x > 0$. For $c > 0$, define

$$K_b(\alpha, \beta, c) \equiv \int_0^\infty \frac{e^{-b/s}}{s^{\alpha+1}} \Big(\frac{c}{s+c}\Big)^\beta ds. \tag{A.10}$$

We then have

$$J \geq \int_{(\lambda_1^{-1},0]} K_b(\alpha,\beta,(1-\rho\lambda_1)^{\xi})F(d\rho) + \int_{(0,\lambda_\nu^{-1})} K_b(\alpha,\beta,(1-\rho\lambda_\nu)^{\xi})F(d\rho); \quad \text{(A.11)}$$

$$J \leq \int_{(\lambda_1^{-1},0]} K_b(\alpha,\beta,(1-\rho\lambda_\nu)^{\xi})F(d\rho) + \int_{(0,\lambda_\nu^{-1})} K_b(\alpha,\beta,(1-\rho\lambda_1)^{\xi})F(d\rho). \quad \text{(A.12)}$$

Clearly K_b is finite if and only if (A.3) or (A.4) holds. Under (A.3) we use the transformation $u = s/(s+c)$ and get $K_0 = c^{-\alpha}\text{Beta}(-\alpha,\, \alpha+\beta)$. So the lower bound of (A.5) holds. Also

$$\begin{aligned}\frac{J}{\text{Beta}(-\alpha,\,\alpha+\beta)} &\leq \Big(1-\frac{\lambda_\nu}{\lambda_1}\Big)^{-\alpha\xi}\int_{(\lambda_1^{-1},0]} F(d\rho) + \Big(1-\frac{\lambda_1}{\lambda_\nu}\Big)^{-\alpha\xi}\int_{(0,\lambda_\nu^{-1})} F(d\rho)\\ &\leq \kappa_u(\lambda_1,\lambda_\nu,-\alpha\xi).\end{aligned}$$

Part (b) thus holds. For Part (c), under (A.4), by making the transformation $u = 1/s$,

$$K_b \;\geq\; \int_0^1 u^{\alpha+\beta-1}\Big(\frac{c}{c+1}\Big)^{\beta} e^{-b} du \;=\; \Big(\frac{c}{c+1}\Big)^{\beta}\,\frac{e^{-b}}{\alpha+\beta}. \quad \text{(A.13)}$$

Applying (A.13) to the two terms of the right hand side of (A.11) with c being $(1-\rho\lambda_\nu)^{\xi}$ and $(1-\rho\lambda_1)^{\xi}$, respectively, we get the lower bound in (27). For the upper bound,

$$K_b \leq \begin{cases} K_0 = c^{-\alpha}Beta(-\alpha,\ \alpha+\beta), & \text{if } \ b>0 \text{ and } \alpha\in(-\beta,0);\\ \dfrac{\Gamma(\alpha)}{b^{\alpha}}, & \text{if } \ b>0 \text{ and } \alpha>0. \end{cases} \quad \text{(A.14)}$$

The first case of the upper bound for J given in (A.8) holds since $\Gamma(\alpha)/b^{\alpha}$ does not depend on ρ. The second case holds from the same argument used for the lower bound in Part (b). For the third case, i.e., when $b>0$ and $\alpha=0$, we have $K_b \;\leq\; c^{\eta}\int_0^{\infty} u^{\eta-1}e^{-bu}du \;=\; c^{\eta}\Gamma(\eta)/b^{\eta}$. Here η is any small positive number satisfying $\eta<\beta$. Applying this formula of K_b to the two terms in (A.12) with c being $(1-\rho\lambda_1)^{\xi}$ and $(1-\rho\lambda_\nu)^{\xi}$, respectively, we get

$$\begin{aligned} J &\leq \Big[\int_{(\lambda_1^{-1},0]}(1-\rho\lambda_v)^{\eta\xi}F(d\rho) + \int_{(0,\lambda_v^{-1})}(1-\rho\lambda_1)^{\eta\xi}F(d\rho)\Big]\frac{\Gamma(\eta)}{b^{\eta}}\\ &\leq \kappa_u(\lambda_1,\lambda_\nu,\eta\xi)\frac{\Gamma(\eta)}{b^{\eta}}.\end{aligned}$$

Part (c) follows immediately.

Proof of Theorem 2. From the proof of Theorem 1 we know that, under the constant prior for $\boldsymbol{\theta}$, we have the inequality (A.1) where G is defined by (13). Let

$$G_1 \equiv G_1(\boldsymbol{\Delta}, \boldsymbol{\rho}; \mathbf{v}) = \frac{1}{|\mathbf{R}_2|^{\frac{1}{2}}|\mathbf{A}|^{\frac{1}{2}}} \prod_{i=1}^{r} \delta_i^{-(a_i+1)} e^{-b_i/\delta_i}, \tag{A.15}$$

$$G_2 = \int_{\lambda_{r1}^{-1}}^{\lambda_{rq_r}^{-1}} \int_0^{\infty} \cdots \int_{\lambda_{11}^{-1}}^{\lambda_{1q_1}^{-1}} \int_0^{\infty} G_1 \, d\delta_1 \, F_1(d\rho_1) \ldots d\delta_r \, F_r(d\rho_r). \tag{A.16}$$

It follows from (14) that

$$\begin{aligned} G_3 &\equiv \int_{\lambda_{r1}^{-1}}^{\lambda_{rq_r}^{-1}} \int_0^{\infty} \cdots \int_{\lambda_{11}^{-1}}^{\lambda_{1q_1}^{-1}} \int_0^{\infty} \int_{\mathbb{R}^q} \int_{\mathbb{R}^p} G d\boldsymbol{\theta} d\mathbf{u} d\delta_1 \, F_1(d\rho_1) \ldots d\delta_r \, F_r(d\rho_r) \\ &= \frac{(2\pi)^{\frac{1}{2}(p+q)}|\mathbf{X}_1'\mathbf{X}_1|^{-\frac{1}{2}}}{\delta_0^{\frac{1}{2}(n-p)+a_0+1}} \exp\Big\{-\frac{\text{SSE}+2b_0}{2\delta_0} - \frac{\hat{\mathbf{u}}'\mathbf{R}_3\hat{\mathbf{u}}}{2\delta_0}\Big\} G_2. \end{aligned} \tag{A.17}$$

Thus the joint posterior is proper if and only if $\int_0^{\infty} G_3 d\delta_0$ is finite. There are two situations to consider under Case 1.

Case 1.1. $t = q$. Since $|\mathbf{A}|^{-1} = \prod_{i=1}^{r} \delta_i^{-q_i}|\mathbf{B}_i|$ and $\mathbf{A}^{-1} = \oplus_{i=1}^{r}\delta_i^{-1}\mathbf{B}_i$, we get

$$\frac{1}{|\mathbf{R}_2|^{\frac{1}{2}}|\mathbf{A}|^{\frac{1}{2}}} = \left(\prod_{i=1}^{r} \frac{|\mathbf{B}_i|^{\frac{1}{2}}}{\delta_i^{\frac{1}{2}q_i}}\right) \frac{1}{|\delta_0^{-1}\mathbf{X}_2'\mathbf{R}_1\mathbf{X}_2 + \oplus_{j=1}^{r}\delta_j^{-1}\mathbf{B}_j|^{\frac{1}{2}}}. \tag{A.18}$$

Let λ_{min} (λ_{max}) be the smallest (largest) eigenvalue of $\mathbf{X}_2'\mathbf{R}_1\mathbf{X}_2$. Since $t = q$, $\mathbf{X}_2'\mathbf{R}_1\mathbf{X}_2$ is positive definite and $\lambda_{min} > 0$. Using the upper bound of Lemma 2,

$$|\mathbf{R}_2| = |\delta_0^{-1}\mathbf{X}_2'\mathbf{R}_1\mathbf{X}_2 + \oplus_{i=1}^{r}\delta_i^{-1}\mathbf{B}_i| \le \prod_{i=1}^{r} \delta_i^{-q_i} \prod_{j=1}^{q_i} \Big[\delta_i \frac{\lambda_{max}}{\delta_0} + \lambda_j(\mathbf{B}_i)\Big]. \tag{A.19}$$

Substituting (A.19) into (A.18), using the fact that $|\mathbf{B}_i| = \prod_{j=1}^{q_i} \lambda_j(\mathbf{B}_i)$, and integrating G_1 with respect to the $(\delta_i, \rho_i)'s$, we get

$$\begin{aligned} G_2 &\ge \prod_{i=1}^{r} \int_{\lambda_{i1}^{-1}}^{\lambda_{iq_i}^{-1}} \int_0^{\infty} \frac{e^{-b_i/\delta_i}}{\delta_i^{a_i+1}} \prod_{j=1}^{q_i} \left(\frac{\lambda_j(\mathbf{B}_i)}{\delta_i\lambda_{max}/\delta_0 + \lambda_j(\mathbf{B}_i)}\right)^{\frac{1}{2}} d\delta_i \, F_i(d\rho_i) \\ &= \Big\{\frac{\lambda_{max}}{\delta_0}\Big\}^{a_+} \prod_{i=1}^{r} \int_{\lambda_{i1}^{-1}}^{\lambda_{iq_i}^{-1}} \int_0^{\infty} \frac{e^{-b_i^*/s}}{s^{a_i+1}} \prod_{j=1}^{q_i} \left(\frac{[1-\rho_i\lambda_j(\mathbf{C}_i)]^{\xi_i}}{s + [1-\rho_i\lambda_j(\mathbf{C}_i)]^{\xi_i}}\right)^{\frac{1}{2}} ds \, F_i(d\rho_i). \end{aligned} \tag{A.20}$$

The last equality holds by making the transformation $s = \delta_i\lambda_{max}/\delta_0$. Here $b_i^* = b_i\lambda_{max}/\delta_0$, $a_+ = \sum_{i=1}^{r} a_i$, defined in condition (c1). Apply Lemma 3 with $t_1 = 1$ and $t_2 = q_i$ so that $t_2 - t_1 + 1 = q_i$.

Thus the right hand side of (A.20) is finite if and only if conditions (a) and (b2) hold. Using the lower bound in Lemma 2 and defining $\tilde{b}_i = b_i\lambda_{min}/\delta_0$,

$$G_2 \le \Big\{\frac{\lambda_{min}}{\delta_0}\Big\}^{a_+}\prod_{i=1}^{r}\int_{\lambda_{i1}^{-1}}^{\lambda_{iq_i}^{-1}}\int_0^\infty \frac{e^{-\tilde{b}_i/s}}{s^{a_i+1}}\prod_{j=1}^{q_i}\left(\frac{[1-\rho_i\lambda_j(\mathbf{C}_i)]^{\xi_i}}{s+[1-\rho_i\lambda_j(\mathbf{C}_i)]^{\xi_i}}\right)^{\frac{1}{2}} ds\, F_i(d\rho_i), \tag{A.21}$$

which is finite only if conditions (a) and (b2) hold. Using the lower bounds of Lemma 3 (b) and (c) for the integrals on the right hand side of (A.20),

$$G_2 \ \ge\ \delta_0^{-a_+}\exp\Big\{-\frac{\lambda_{max}\sum_{i=1}^r b_i}{\delta_0}\Big\}\lambda_{max}^{a_+}\prod_{i=1}^r H_{1i}, \tag{A.22}$$

where

$$H_{1i} = \begin{cases} \kappa_1(\lambda_{i1},\lambda_{iq_i},\xi_i,\frac{q_i}{2})/(a_i+\frac{q_i}{2}), & \text{if conditions (a1) and (b2) hold,}\\ \kappa_2(\lambda_{i1},\lambda_{iq_i},\xi_i,\frac{q_i}{2})\ \text{Beta}(-a_i,\, a_i+\frac{q_i}{2}), & \text{if conditions (a2) and (b2) hold,}\end{cases} \tag{A.23}$$

and κ_1 and κ_2 are defined by (A.6) and (A.9), respectively.

Using the upper bounds of Lemma 3 (b) and (c) for the integrals on the right hand side of (40), $G_2 \le \delta_0^{-\tilde{a}}\lambda_{min}^{\tilde{a}}\prod_{i=1}^r H_{2i}$, where $\tilde{a} = \sum_{i=1}^r a_i^- + \eta\sum_{i=1}^r 1(a_i=0)$ for some small $\eta > 0$, and

$$H_{2i} = \begin{cases} \kappa_u(\lambda_{i1},\lambda_{iq_i},-\frac{a_iq_i}{2})Beta(-a_i,\, a_i+\frac{q_i}{2}), & \text{if } \ b_i \ge 0 \text{ and } a_i\in(-\frac{q_i}{2},\,0);\\ \Gamma(a_i), & \text{if } \ b_i>0 \text{ and } a_i>0;\\ \kappa_u(\lambda_{i1},\lambda_{iq_i},\eta\xi)\Gamma(\eta), & \text{if } \ b_i>0 \text{ and } a_i=0.\end{cases} \tag{A.24}$$

Combining (A.17), (A.16) and (A.22), we have

$$\frac{(2\pi)^{\frac{1}{2}(p+q)}}{|\mathbf{X}_1'\mathbf{X}_1|^{\frac{1}{2}}}\lambda_{max}^{a_+}\left[\prod_{i=1}^r H_{1i}\right]\frac{\exp\Big\{-\frac{1}{2\delta_0}(\text{SSE}+2b_0+\lambda_{max}\sum_{i=1}^r b_i+\hat{\mathbf{u}}'\mathbf{X}_2'\mathbf{R}_1\mathbf{X}_2\hat{\mathbf{u}})\Big\}}{\delta_0^{\frac{1}{2}(n-p)+a_++a_0+1}}$$

$$\le G_3 \ \le\ \frac{(2\pi)^{\frac{1}{2}(p+q)}}{|\mathbf{X}_1'\mathbf{X}_1|^{\frac{1}{2}}}\lambda_{min}^{\tilde{a}}\left[\prod_{i=1}^r H_{2i}\right]\frac{1}{\delta_0^{\frac{1}{2}(n-p)+\tilde{a}+a_0+1}}\exp\Big\{-\frac{\text{SSE}+2b_0}{2\delta_0}\Big\}. \tag{A.25}$$

Clearly, the integral on the left hand side with respect to δ_0 is finite only if condition (c1) holds. From condition (c2), there is a small positive number η such that $\eta < \min\{q_1/2,\ldots,q_r\}$ and $n-p+2a_0+2\sum_{i=1}^r a_i^- + \eta\sum_{i=1}^r 1(a_i=0) > 0$, where $1(a_i=0)$ is the indicator. Thus $\frac{1}{2}(n-p)+\tilde{a}+a_0 > 0$, for some small $\eta > 0$. Consequently, the integral $\int G_3 d\delta_0$ is finite if conditions (a), (b2) and (c2) hold. This proves Case 1.1.

Case 1.2. $t < q$ and $r = 1$. In this case, we have $q_1 = q$ and

$$\frac{1}{|\mathbf{R}_2|^{\frac{1}{2}}|\mathbf{A}|^{\frac{1}{2}}} = \frac{|\mathbf{B}_1|^{\frac{1}{2}}}{\delta_1^{\frac{q}{2}}}\left|\frac{\mathbf{X}_2'\mathbf{R}_1\mathbf{X}_2}{\delta_0} + \frac{\mathbf{B}_1}{\delta_1}\right|^{-\frac{1}{2}}$$
$$= \prod_{j=1}^{q}\Big[\lambda_j(\mathbf{B}_1)\Big]^{\frac{1}{2}}\left|\frac{\delta_1}{\delta_0}\mathbf{X}_2'\mathbf{R}_1\mathbf{X}_2 + \mathbf{B}_1\right|^{-\frac{1}{2}}.$$

Since $\mathrm{rank}(\mathbf{X}_2'\mathbf{R}_1\mathbf{X}_2) = t < q$, $\mathbf{X}_2'\mathbf{R}_1\mathbf{X}_2$ has t positive eigenvalues. Let λ_{sp} and λ_{max} be the smallest positive eigenvalue and the largest eigenvalue of $(\mathbf{X}_2'\mathbf{R}_1\mathbf{X}_2)$, respectively. From Lemma 2,

$$\frac{e^{-b_1/\delta_1}}{\delta_1^{a_1+1}}\prod_{j=1}^{t}\left[\frac{\lambda_j(\mathbf{B}_1)}{\frac{\delta_1}{\delta_0}\lambda_{max} + \lambda_j(\mathbf{B}_1)}\right]^{\frac{1}{2}} \le G_1 \le \frac{e^{-b_1/\delta_1}}{\delta_1^{a_1+1}}\prod_{j=1}^{t}\left[\frac{\lambda_{q+1-j}(\mathbf{B}_1)}{\frac{\delta_1}{\delta_0}\lambda_{sp} + \lambda_{q+1-j}(\mathbf{B}_1)}\right]^{\frac{1}{2}}.$$

Using Lemma 3 with $(t_1, t_2) = (1, t)$ so that $t_2 - t_1 + 1 = t$, the lower bound and upper bound of G_2 with $r = 1$ and $q_1 = t$ hold iff conditions (a) and (b2) are satisfied.

Case 2. $t < q$ and $r > 1$. Since the inequality (A.19) still holds, the proof of necessity for Case 1 can be applied here. For the sufficiency note that, as in Case 2, the matrix $\mathbf{X}_2'\mathbf{R}_1\mathbf{X}_2$ has $q - t$ zero eigenvalues. From Lemma 2 we know that

$$\Big|\frac{1}{\delta_0}\mathbf{X}_2'\mathbf{R}_1\mathbf{X}_2 + \oplus_{i=1}^{r}\frac{1}{\delta_i}\mathbf{B}_i\Big|$$
$$\ge \Big\{\prod_{j=1}^{q-t}\lambda_j\Big(\oplus_{i=1}^{r}\frac{1}{\delta_i}\mathbf{B}_i\Big)\Big\}\Big\{\prod_{j=q-t+1}^{q}\Big|\frac{1}{\delta_0}\lambda_{sp} + \lambda_j\Big(\oplus_{i=1}^{r}\frac{1}{\delta_i}\mathbf{B}_i\Big)\Big|\Big\},$$

where λ_{sp} is defined in Case 1.1 above. Under conditions (a) and (b2), $q_i > q - t > 0$. Thus

$$\Big|\frac{1}{\delta_0}\mathbf{X}_2'\mathbf{R}_1\mathbf{X}_2 + \oplus_{i=1}^{r}\frac{1}{\delta_i}\mathbf{B}_i\Big| \ge \prod_{i=1}^{r}\frac{1}{\delta_i^{q_i}}\left(\prod_{j=1}^{q-t}\lambda_j(\mathbf{B}_i)\right)\prod_{j=q-t+1}^{q_i}\Big[\delta_i\frac{\lambda_{sp}}{\delta_0} + \lambda_j(\mathbf{B}_i)\Big].$$

Substituting this inequality into G_1, defined by (A.15), we have

$$G_2 \le \left(\frac{\lambda_{sp}}{\delta_0}\right)^{a_+}\prod_{i=1}^{r}\int_{\lambda_{i1}^{-1}}^{\lambda_{iq_i}^{-1}}\int_0^\infty \frac{e^{-b_i^*/\delta_i}}{s^{a_i+1}}\prod_{j=q-t+1}^{q_i}\left(\frac{[1-\rho_i\lambda_j(\mathbf{C}_i)]^{\xi_i}}{s + [1-\rho_i\lambda_j(\mathbf{C}_i)]^{\xi_i}}\right)^{\frac{1}{2}} ds\, F_i(d\rho_i), \tag{A.26}$$

where $b_i^* = b_i\lambda_{sp}/\delta_0$. It follows from Lemma 3, with (t_1, t_2) being $(q - t + 1, q_i)$, that the right side of (A.26) is finite if conditions (a) and (b2) hold. In this case,

$G_2 \le (\delta_0)^{-\tilde{a}}(\lambda_{sp})^{\tilde{a}} \prod_{i=1}^r H_{2i}$, where $\tilde{a}$ is defined in the proof of Case 1 and H_{2i} is given by (A.24) with q_i replaced by $q_i - q + t$. It follows From (A.17) and (A.16) that

$$G_3 \le \frac{(2\pi)^{\frac{1}{2}(p+q)}(\lambda_{sp})^{\tilde{a}}}{|\mathbf{X}_1'\mathbf{X}_1|^{\frac{1}{2}}} \left[\prod_{i=1}^r H_{2i}\right] \frac{1}{\delta_0^{\frac{1}{2}(n-p)+\tilde{a}+a_0+1}} \exp\Big\{-\frac{\mathrm{SSE}+2b_0}{2\delta_0}\Big\}. \quad \text{(A.27)}$$

The integral of the right hand side with respect to δ_0 is finite if condition (c2) holds. This proves the case when $t < q$ and $r > 1$.

Proof of Theorem 3. Without loss of generality, assume that $i_j = j$, $j = 1, \ldots, n$, so that $\mathbf{v}^* \equiv (v_1, \ldots, v_n)'$. Write $\mathbf{y} = (\mathbf{y}^{*\prime}, \widetilde{\mathbf{y}}')'$, $\mathbf{v} = (\mathbf{v}^{*\prime}, \widetilde{\mathbf{v}}')'$, and $\widetilde{\mathbf{v}} \equiv (v_{n+1}, \ldots, v_N)' = \widetilde{\mathbf{X}}_1\boldsymbol{\theta} + \widetilde{\mathbf{X}}_2\mathbf{u} + \widetilde{\mathbf{e}}$. The posterior density of $(\mathbf{v}, \boldsymbol{\theta}, \mathbf{u}, \boldsymbol{\Delta}, \boldsymbol{\rho}, \delta_0)$ given $\mathbf{y}$ is

$$\begin{aligned} & p(\mathbf{v}, \boldsymbol{\theta}, \mathbf{u}, \boldsymbol{\Delta}, \boldsymbol{\rho}, \delta_0|\mathbf{y}) \\ \propto & \frac{\prod_{i=1}^N f_i(y_i|v_i)}{\delta_0^{\frac{1}{2}(N-n)}} \prod_{i=n+1}^N \exp\Big[-\frac{1}{2\delta_0}(\widetilde{\mathbf{v}} - \widetilde{\mathbf{X}}_1\boldsymbol{\theta} - \widetilde{\mathbf{X}}_2\mathbf{u})'(\widetilde{\mathbf{v}} - \widetilde{\mathbf{X}}_1\boldsymbol{\theta} - \widetilde{\mathbf{X}}_2\mathbf{u})\Big]\, G^*. \end{aligned}$$

Here G^* is defined by (13) with $\mathbf{X}_1$ and $\mathbf{X}_2$ replaced by $\mathbf{X}_1^*$ and $\mathbf{X}_2^*$, respectively. Using the second inequality in (16) and integrating with respect to $(v_{n+1}, \ldots, v_N)$, $p(\mathbf{v}^*, \boldsymbol{\theta}, \mathbf{u}, \boldsymbol{\Delta}, \boldsymbol{\rho}, \delta_0|\mathbf{y})$ is proportional to $\prod_{i=1}^n f_i(y_i|v_i)\, G^*$. From the inequality (A.25) for Case 1. and (A.27) for Case 3. in the proof of Theorem 2, $p(\mathbf{v}^*, \delta_0|\mathbf{y}) \le \widetilde{M}\delta_0^{-\frac{1}{2}(n-p)-\tilde{a}-a_0-1}\exp\{-b_0/(2\delta_0)\}$, where $\tilde{a}$ is the same as in the proof of Case 1 and $\widetilde{M}$ is a generic constant depending only on $(\mathbf{X}_1^*, \mathbf{X}_2^*)$ and the eigenvalues of $\mathbf{C}$. From the assumption (c2) in Theorem 2, we know that $p(\mathbf{v}^*|\mathbf{y}) \le \prod_{i=1}^n f_i(y_i|v_i)$, which is proper by the first inequality in (16). This completes the proof.

References

Bernardinelli, L., Clayton, D. and Montomoli, C. (1995). Bayesian estimates of disease maps: how important are priors? *Statist. Medicine* **14**, 2411-2431.

Berger, J. O. and Bernardo, J. M. (1992). Reference priors in variance components problems. In *Proc. Indo-USA Workshop on Bayesian Analysis Statist. Econometrics* (Edited by P. K. Goel and N. S. Iyenger), 177-194. Springer-Verlag, New York.

Besag, J., York, J. and Mollié, A. (1991). Bayesian image restoration, with two applications in spatial statistics (with discussion). *Ann. Inst. Statist. Math.* **43**, 1-59.

Clayton, D. and Kaldor, J. (1987). Empirical Bayes estimates of age-standardized relative risks for use in disease mapping. *Biometrics* **43**, 671-681.

Datta, G. S. (1996). On priors providing frequentist validity for Bayesian inference for multiple parametric functions. *Biometrika* **83**, 287-298.

Datta, Gauri S. and Ghosh, M. (1991). Bayesian prediction in linear models: applications to small area estimation. *Ann. Statist.* **19**, 1748-1770.

Gelfand, A. E. and Smith, A. F. M. (1990). Sampling based approaches to calculating marginal densities. *J. Amer. Statist. Assoc.* **85**, 398-409.

Ghosh, J. K. and Mukerjee, R. (1993). On priors that match posterior and frequentist distribution functions. *Canad. J. Statist.* **21**, 89-96.

Ghosh, M., Natarajan, K., Stroud, T. W. F. and Carlin, B. P. (1998). Generalized linear models for small area estimation. *J. Amer. Statist. Assoc.* **93**, 273-282.

Ghosh, M., Natarajan, K., Waller, L. A. and Kim, D. (1999). Hierarchical Bayes GLMs for the analysis of spatial data: an application to disease mapping. *J. Statist. Plann. Inference* **75**, 305-318.

Gilks, W. R. and Wild, P. (1992). Adaptive rejection sampling for Gibbs sampling. *Appl. Statist.* **41**, 337-348.

Hobert, J. P. and Casella, G. (1996). The effect of improper priors on Gibbs sampling in hierarchical linear mixed models. *J. Amer. Statist. Assoc.* **91**, 1461-1473.

Ibrahim, J. G. and Laud, P. W. (1991). On Bayesian analysis of generalized linear models using Jeffreys's prior. *J. Amer. Statist. Assoc.* **86**, 981-986.

Kahn, M. J. (1990). Bayes empirical Bayes beta-binomial modeling with covariates, applied to health care policy. Ph.D dissertation, Department of Statistics, University of Washington.

Kass, R. E. and Wasserman, L. (1996). The selection of prior distributions by formal rules. *J. Amer. Statist. Assoc.* **91**, 1343-1370.

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

Morris, C. N. (1983). Parametric empirical Bayes inference: theory and applications (with discussion). *J. Amer. Statist. Assoc.* **78**, 47-65.

Ord, K. (1975). Estimation methods for models with spatial interaction. *J. Amer. Statist. Assoc.* **70**, 120-126.

Searle, S. R., Casella, G. and McCulloch, C. E. (1992). *Variance Components.* Wiley, New York.

Sun, D. and Berger, J. O. (1998). Reference priors with partial information. *Biometrika* **85**, 55-71.

Sun, D., Tsutakawa, R. K. and Speckman, P. L. (1999). Posterior distribution of hierarchical models using CAR (1) distributions. *Biometrika* **86**, 341-350.

Sun, D., Tsutakawa, R. K., Kim, H. and He, Z. (2000). Spatio-temporal interaction with disease mapping. *Statist. Medicine* **19**, 2015-2035.

Yang, R. and Chen, M. (1995). Bayesian analysis for random coefficient models using noninformative priors. *J. Multivariate Anal.* **55**, 283-311.

Department of Statistics, 222 Math Science Building, University of Missouri-Columbia, Columbia, Missouri 65211, U.S.A.

E-mail: dsun@stat.missouri.edu

(Received April 1998; accepted May 2000)