The Annals of Statistics
2012, Vol. 40, No. 6, 2823–2849
DOI: 10.1214/12-AOS1052

CONVERGENCE ANALYSIS OF THE GIBBS SAMPLER FOR BAYESIAN GENERAL LINEAR MIXED MODELS WITH IMPROPER PRIORS

BY JORGE CARLOS ROMÁN AND JAMES P. HOBERT[1]

Vanderbilt University and University of Florida

Bayesian analysis of data from the general linear mixed model is challenging because any nontrivial prior leads to an intractable posterior density. However, if a conditionally conjugate prior density is adopted, then there is a simple Gibbs sampler that can be employed to explore the posterior density. A popular default among the conditionally conjugate priors is an improper prior that takes a product form with a flat prior on the regression parameter, and so-called power priors on each of the variance components. In this paper, a convergence rate analysis of the corresponding Gibbs sampler is undertaken. The main result is a simple, easily-checked sufficient condition for geometric ergodicity of the Gibbs–Markov chain. This result is close to the best possible result in the sense that the sufficient condition is only slightly stronger than what is required to ensure posterior propriety. The theory developed in this paper is extremely important from a practical standpoint because it guarantees the existence of central limit theorems that allow for the computation of valid asymptotic standard errors for the estimates computed using the Gibbs sampler.

1. Introduction. The general linear mixed model (GLMM) takes the form

$$Y = X\beta + Zu + e, \tag{1}$$

where Y is an $N \times 1$ data vector, X and Z are known matrices with dimensions $N \times p$ and $N \times q$, respectively, β is an unknown $p \times 1$ vector of regression coefficients, u is a random vector whose elements represent the various levels of the random factors in the model and $e \sim \mathrm{N}_N(0, \sigma_e^2 I)$. The random vectors e and u are assumed to be independent. Suppose there are r random factors in the model. Then u and Z are partitioned accordingly as $u = (u_1^T \ \ u_2^T \ \ \cdots \ \ u_r^T)^T$ and $Z = (Z_1 \ \ Z_2 \ \ \cdots \ \ Z_r)$, where u_i is $q_i \times 1$, Z_i is $N \times q_i$ and $q_1 + \cdots + q_r = q$. Then

$$Zu = \sum_{i=1}^{r} Z_i u_i,$$

Received April 2012; revised September 2012.

[1]Supported by NSF Grants DMS-08-05860 and DMS-11-06395.

MSC2010 subject classifications. Primary 60J27; secondary 62F15.

Key words and phrases. Convergence rate, geometric drift condition, geometric ergodicity, Markov chain, Monte Carlo, posterior propriety.

and it is assumed that $u \sim \mathrm{N}_q(0, D)$, where $D = \bigoplus_{i=1}^r \sigma_{u_i}^2 I_{q_i}$. Let σ^2 denote the vector of variance components, that is, $\sigma^2 = (\sigma_e^2 \;\; \sigma_{u_1}^2 \;\; \cdots \;\; \sigma_{u_r}^2)^T$. For background on this model, which is sometimes called the *variance components model*, see Searle, Casella and McCulloch (1992).

A Bayesian version of the GLMM can be assembled by specifying a prior distribution for the unknown parameters, β and σ^2. A popular choice is the proper (conditionally) conjugate prior that takes β to be multivariate normal, and takes each of the variance components to be inverted gamma. One obvious reason for using such a prior is that the resulting posterior has conditional densities with standard forms, and this facilitates the use of the Gibbs sampler.

In situations where there is little prior information, the hyperparameters of this proper prior are often set to extreme values as this is thought to yield a "noninformative" prior. Unfortunately, these extreme proper priors approximate improper priors that correspond to improper posteriors, and this results in various forms of instability. This problem has led several authors, including Daniels (1999) and Gelman (2006), to discourage the use of such extreme proper priors, and to recommend alternative default priors that are improper, but lead to proper posteriors. Consider, for example, the one-way random effects model given by

$$Y_{ij} = \beta + \alpha_i + e_{ij}, \tag{2}$$

where $i = 1, \dots, c$, $j = 1, \dots, n_i$, the α_i's are i.i.d. $\mathrm{N}(0, \sigma_\alpha^2)$, and the e_{ij}'s, which are independent of the α_i's, are i.i.d. $\mathrm{N}(0, \sigma_e^2)$. This is an important special case of model (1). (See Section 5 for a detailed explanation of how the GLMM reduces to the one-way model.) The standard diffuse prior for this model, which is among those recommended by Gelman (2006), has density $1/(\sigma_e^2 \sqrt{\sigma_\alpha^2})$. This prior, like many of the improper priors for the GLMM that have been suggested and studied in the literature, is called a "power prior" because it is a product of terms, each a variance component brought to a (possibly negative) power. Of course, like the proper conjugate priors mentioned above, power priors also lead to posteriors whose conditional densities have standard forms.

In this paper, we consider the following parametric family of priors for (β, σ^2):

$$p(\beta, \sigma^2; a, b) = (\sigma_e^2)^{-(a_e+1)} e^{-b_e/\sigma_e^2} \left[\prod_{i=1}^r (\sigma_{u_i}^2)^{-(a_i+1)} e^{-b_i/\sigma_{u_i}^2} \right] I_{\mathbb{R}_+^{r+1}}(\sigma^2), \tag{3}$$

where $a = (a_e, a_1, \dots, a_r)$ and $b = (b_e, b_1, \dots, b_r)$ are fixed hyperparameters, and $\mathbb{R}_+ := (0, \infty)$. By taking b to be the vector of 0's, we can recover the power priors described above. Note that β does not appear on the right-hand side of (3); that is, we are using a so-called flat prior for β. Consequently, even if all the elements of a and b are strictly positive, so that every variance component gets a proper prior, the overall prior remains improper. There have been several studies concerning posterior propriety in this context, but it is still not known exactly which values of a and b yield proper posteriors. The best known result is due to Sun, Tsutakawa and He (2001), and we state it below so that it can be used in a comparison later in this section.

Define $\theta = (\beta^T \ \ u^T)^T$ and $W = (X \ \ Z)$, so that $W\theta = X\beta + Zu$. Let y denote the observed value of Y, and let $\phi_d(x; \mu, \Sigma)$ denote the $\mathrm{N}_d(\mu, \Sigma)$ density evaluated at the vector x. By definition, the posterior density is proper if

$$m(y) := \int_{\mathbb{R}_+^{r+1}} \int_{\mathbb{R}^{p+q}} \pi^*(\theta, \sigma^2|y)\, d\theta\, d\sigma^2 < \infty,$$

where

$$\pi^*(\theta, \sigma^2|y) = \phi_N(y; W\theta, \sigma_e^2 I)\phi_q(u; 0, D)p(\beta, \sigma^2; a, b). \tag{4}$$

A routine calculation shows that the posterior is improper if $\operatorname{rank}(X) < p$. The following result provides sufficient (and nearly necessary) conditions for propriety. (Throughout the paper, the symbol P subscripted with a matrix will denote the projection onto the column space of that matrix.)

THEOREM 1 [Sun, Tsutakawa and He (2001)]. *Assume that* $\operatorname{rank}(X) = p$, *and let* $t = \operatorname{rank}(Z^T(I - P_X)Z)$ *and* $\mathrm{SSE} = \|(I - P_W)y\|^2$. *If the following four conditions hold, then* $m(y) < \infty$:

(A) *For each* $i \in \{1, 2, \ldots, r\}$, *one of the following holds*:

$$\text{(A1)} \quad a_i < b_i = 0; \qquad \text{(A2)} \quad b_i > 0;$$

(B) *for each* $i \in \{1, 2, \ldots, r\}$, $q_i + 2a_i > q - t$;
(C) $N + 2a_e > p - 2\sum_{i=1}^r a_i I_{(-\infty,0)}(a_i)$;
(D) $2b_e + \mathrm{SSE} > 0$.

If $m(y) < \infty$, then the posterior density is well defined (i.e., proper) and is given by $\pi(\theta, \sigma^2|y) = \pi^*(\theta, \sigma^2|y)/m(y)$, but it is intractable in the sense that posterior expectations cannot be computed in closed form, nor even by classical Monte Carlo methods. However, there is a simple two-step Gibbs sampler that can be used to approximate the intractable posterior expectations. This Gibbs sampler simulates a Markov chain, $\{(\theta_n, \sigma_n^2)\}_{n=0}^\infty$, that lives on $\mathsf{X} = \mathbb{R}^{p+q} \times \mathbb{R}_+^{r+1}$, and has invariant density $\pi(\theta, \sigma^2|y)$. If the current state of the chain is (θ_n, σ_n^2), then the next state, $(\theta_{n+1}, \sigma_{n+1}^2)$, is simulated using the usual two steps. Indeed, we draw θ_{n+1} from $\pi(\theta|\sigma_n^2, y)$, which is a $(p+q)$-dimensional multivariate normal density, and then we draw σ_{n+1}^2 from $\pi(\sigma^2|\theta_{n+1}, y)$, which is a product of $r+1$ univariate inverted gamma densities. The exact forms of these conditional densities are given in Section 2.

Because the Gibbs–Markov chain is Harris ergodic (see Section 2), we can use it to construct consistent estimates of intractable posterior expectations. For $k > 0$, let $L_k(\pi)$ denote the set of functions $g : \mathbb{R}^{p+q} \times \mathbb{R}_+^{r+1} \to \mathbb{R}$ such that

$$E_\pi|g|^k := \int_{\mathbb{R}_+^{r+1}} \int_{\mathbb{R}^{p+q}} |g(\theta, \sigma^2)|^k \pi(\theta, \sigma^2|y)\, d\theta\, d\sigma^2 < \infty.$$

If $g \in L_1(\pi)$, then the ergodic theorem implies that the average

$$\overline{g}_m := \frac{1}{m} \sum_{i=0}^{m-1} g(\theta_i, \sigma_i^2)$$

is a strongly consistent estimator of $E_\pi g$, no matter how the chain is started. Of course, in practice, an estimator is only useful if it is possible to compute an associated (probabilistic) bound on the difference between the estimate and the truth. Typically, this bound is based on a standard error. All available methods of computing a valid asymptotic standard error for $\overline{g}_m$ are based on the existence of a central limit theorem (CLT) for $\overline{g}_m$ [see, e.g., Bednorz and Łatuszyński (2007), Flegal, Haran and Jones (2008), Flegal and Jones (2010), Jones et al. (2006)]. Unfortunately, even if $g \in L_k(\pi)$ for all $k > 0$, Harris ergodicity is not enough to guarantee the existence of a CLT for $\overline{g}_m$ [see, e.g., Roberts and Rosenthal (1998, 2004)]. The standard method of establishing the existence of CLTs is to prove that the underlying Markov chain converges at a geometric rate.

Let $\mathcal{B}(\mathsf{X})$ denote the Borel sets in X, and let $P^n : \mathsf{X} \times \mathcal{B}(\mathsf{X}) \to [0, 1]$ denote the n-step Markov transition function of the Gibbs–Markov chain. That is, $P^n((\theta, \sigma^2), A)$ is the probability that $(\theta_n, \sigma_n^2) \in A$, given that the chain is started at $(\theta_0, \sigma_0^2) = (\theta, \sigma^2)$. Also, let $\Pi(\cdot)$ denote the posterior distribution. The chain is called geometrically ergodic if there exist a function $M : \mathsf{X} \to [0, \infty)$ and a constant $\varrho \in [0, 1)$ such that, for all $(\theta, \sigma^2) \in \mathsf{X}$ and all $n = 0, 1, \ldots,$ we have

$$\| P^n((\theta, \sigma^2), \cdot) - \Pi(\cdot) \|_{\mathrm{TV}} \le M(\theta, \sigma^2) \varrho^n,$$

where $\| \cdot \|_{\mathrm{TV}}$ denotes the total variation norm. The relationship between geometric convergence and CLTs is simple: if the chain is geometrically ergodic and $E_\pi |g|^{2+\delta} < \infty$ for some $\delta > 0$, then there is a CLT for $\overline{g}_m$. Our main result (Theorem 2 in Section 3) provides conditions under which the Gibbs–Markov chain is geometrically ergodic. The conditions of Theorem 2 are not easy to interpret, and checking them may require some nontrivial numerical work. On the other hand, the following corollary to Theorem 2 is a slightly weaker result whose conditions are very easy to check and understand.

COROLLARY 1. *Assume that* rank$(X) = p$. *If the following four conditions hold, then the Gibbs–Markov chain is geometrically ergodic.*

(A) *For each* $i \in \{1, 2, \ldots, r\}$, *one of the following holds*:

$$\text{(A1)} \quad a_i < b_i = 0; \qquad \text{(A2)} \quad b_i > 0;$$

(B′) *for each* $i \in \{1, 2, \ldots, r\}$, $q_i + 2a_i > q - t + 2$;
(C′) $N + 2a_e > p + t + 2$;
(D) $2b_e + \mathrm{SSE} > 0$.

As we explain in Section 2, the best result we could possibly hope to obtain is that the Gibbs–Markov chain is geometrically ergodic whenever the posterior is proper. With this in mind, note that the conditions of Corollary 1 are very close to the conditions for propriety given in Theorem 1. In fact, the former imply the latter. To see this, assume that (A), (B′), (C′) and (D) all hold. Then, obviously, (B) holds, and all that remains is to show that (C) holds. This would follow immediately if we could establish that

$$t \geq -2 \sum_{i=1}^{r} a_i I_{(-\infty,0)}(a_i). \tag{5}$$

We consider two cases. First, if $\sum_{i=1}^{r} I_{(-\infty,0)}(a_i) = 0$, then it follows that $-2\sum_{i=1}^{r} a_i I_{(-\infty,0)}(a_i) = 0$, and (5) holds (since t is nonnegative). On the other hand, if $\sum_{i=1}^{r} I_{(-\infty,0)}(a_i) > 0$, then there is at least one negative a_i, and (B′) implies that

$$\sum_{i=1}^{r} (q_i + 2a_i) I_{(-\infty,0)}(a_i) > q - t.$$

This inequality combined with the fact that $q = q_1 + \cdots + q_r$ yields

$$t > q - \sum_{i=1}^{r} (q_i + 2a_i) I_{(-\infty,0)}(a_i) \geq -2 \sum_{i=1}^{r} a_i I_{(-\infty,0)}(a_i),$$

so (5) holds, and this completes the argument.

The strong similarity between the conditions of Corollary 1 and those of Theorem 1 might lead the reader to believe that the proofs of our results rely somehow on Theorem 1. This is not the case, however. In fact, we do not even assume posterior propriety before embarking on our convergence rate analysis; see Section 3.

The only other existing result on geometric convergence of Gibbs samplers for linear mixed models with *improper* priors is that of Tan and Hobert (2009), who considered (a slightly reparameterized version of) the one-way random effects model (2) and priors with $b = (b_e, b_1) = (0, 0)$. We show in Section 5 that our Theorem 2 (specialized to the one-way model) improves upon the result of Tan and Hobert (2009) in the sense that our sufficient conditions for geometric convergence are weaker. Moreover, it is known in this case exactly which priors lead to proper posteriors (when SSE > 0), and we use this fact to show that our results can be very close to the best possible. For example, if the standard diffuse prior is used, then the posterior is proper if and only if $c \geq 3$. On the other hand, our results imply that the Gibbs–Markov chain is geometrically ergodic as long as $c \geq 3$, and the total sample size, $N = n_1 + n_2 + \cdots + n_c$, is at least $c + 2$. The extra condition that $N \geq c + 2$ is extremely weak. Indeed, SSE > 0 implies that $N \geq c + 1$, so, for fixed $c \geq 3$, our condition for geometric ergodicity fails only in the single case where $N = c + 1$.

An analogue of Corollary 1 for the GLMM with *proper* priors can be found in Johnson and Jones (2010). In contrast with our results, one of their sufficient conditions for geometric convergence is that $X^T Z = 0$, which rarely holds in practice. Overall, the proper and improper cases are similar, in the sense that geometric ergodicity is established via geometric drift conditions in both cases. However, the drift conditions are quite disparate, and the analysis required in the improper case is substantially more demanding. Finally, we note that the linear models considered by Papaspiliopoulos and Roberts (2008) are substantively different from ours because these authors assume that the variance components are known.

The remainder of this paper is organized as follows. Section 2 contains a formal definition of the Gibbs–Markov chain. The main convergence result is stated and proven in Section 3, and an application involving the two-way random effects model is given in Section 4. In Section 5, we consider the one-way random effects model and compare our conditions for geometric convergence with those of Tan and Hobert (2009). Finally, Section 6 concerns an interesting technical issue related to the use of improper priors.

2. The Gibbs sampler. In this section, we formally define the Markov chain underlying the Gibbs sampler, and state some of its properties. Recall that $\theta = (\beta^T \;\; u^T)^T$, $\sigma^2 = (\sigma_e^2 \;\; \sigma_{u_1}^2 \;\; \cdots \;\; \sigma_{u_r}^2)^T$ and $\pi^*(\theta, \sigma^2|y)$ is the potentially improper, unnormalized posterior density defined at (4). Suppose that

$$\int_{\mathbb{R}^{p+q}} \pi^*(\theta, \sigma^2|y)\, d\theta < \infty \tag{6}$$

for all σ^2 outside a set of measure zero in $\mathbb{R}_+^{r+1}$, and that

$$\int_{\mathbb{R}_+^{r+1}} \pi^*(\theta, \sigma^2|y)\, d\sigma^2 < \infty \tag{7}$$

for all θ outside a set of measure zero in $\mathbb{R}^{p+q}$. These two integrability conditions are necessary, but not sufficient, for posterior propriety. (Keep in mind that it is not known exactly which priors yield proper posteriors.) When (6) and (7) hold, we can define conditional densities as follows:

$$\pi(\theta|\sigma^2, y) = \frac{\pi^*(\theta, \sigma^2|y)}{\int_{\mathbb{R}^{p+q}} \pi^*(\theta, \sigma^2|y)\, d\theta} \quad \text{and} \quad \pi(\sigma^2|\theta, y) = \frac{\pi^*(\theta, \sigma^2|y)}{\int_{\mathbb{R}_+^{r+1}} \pi^*(\theta, \sigma^2|y)\, d\sigma^2}.$$

Clearly, when the posterior is proper, these conditionals are the usual ones based on $\pi(\theta, \sigma^2|y)$. When the posterior is improper, they are incompatible conditional densities; that is, there is no (proper) joint density that generates them. In either case, we can run the Gibbs sampler as usual by drawing alternately from the two conditionals. However, as we explain below, if the posterior is improper, then the resulting Markov chain cannot be geometrically ergodic. Despite this fact, we do not restrict attention to the cases where the sufficient conditions for propriety in

Theorem 1 are satisfied. Indeed, we hope to close the gap that currently exists between the necessary and sufficient conditions for propriety by finding weaker conditions than those in Theorem 1 that imply geometric ergodicity (and hence posterior propriety).

We now provide a set of conditions that guarantee that the integrability conditions are satisfied. Define

$$\tilde{s} = \min\{q_1 + 2a_1, q_2 + 2a_2, \dots, q_r + 2a_r, N + 2a_e\}.$$

The proof of the following result is straightforward and is left to the reader.

PROPOSITION 1. *The following four conditions are sufficient for* (6) *and* (7) *to hold*:

(S1) $\operatorname{rank}(X) = p$;
(S2) $\min\{b_1, b_2, \dots, b_r\} \geq 0$;
(S3) $2b_e + \mathrm{SSE} > 0$;
(S4) $\tilde{s} > 0$.

Note that $\mathrm{SSE} = \mathrm{SSE}(X, Z, y) = \|y - W\hat{\theta}\|^2$, where $W = (X \;\; Z)$ and $\hat{\theta} = (W^T W)^- W^T y$. Therefore, if condition (S3) holds, then for all $\theta \in \mathbb{R}^{p+q}$,

$$2b_e + \|y - W\theta\|^2 = 2b_e + \|y - W\hat{\theta}\|^2 + \|W\theta - W\hat{\theta}\|^2 \geq 2b_e + \mathrm{SSE} > 0.$$

Note also that if $N > p + q$, then SSE is strictly positive with probability one under the data generating model.

Assume now that (S1)–(S4) hold so that the conditional densities are well defined. Routine manipulation of $\pi^*(\theta, \sigma^2|y)$ shows that $\pi(\theta|\sigma^2, y)$ is a multivariate normal density with mean vector

$$m = \begin{bmatrix} (X^T X)^{-1} X^T (I - (\sigma_e^2)^{-1} Z Q^{-1} Z^T (I - P_X)) y \\ (\sigma_e^2)^{-1} Q^{-1} Z^T (I - P_X) y \end{bmatrix}$$

and covariance matrix

$$V = \begin{bmatrix} \sigma_e^2 (X^T X)^{-1} + R Q^{-1} R^T & -R Q^{-1} \\ -Q^{-1} R^T & Q^{-1} \end{bmatrix},$$

where $Q = (\sigma_e^2)^{-1} Z^T (I - P_X) Z + D^{-1}$ and $R = (X^T X)^{-1} X^T Z$.

Things are a bit more complicated for $\pi(\sigma^2|\theta, y)$ due to the possible existence of a bothersome set of measure zero. Define $A = \{i \in \{1, 2, \dots, r\} : b_i = 0\}$. If A is empty, then $\pi(\sigma^2|\theta, y)$ is well defined for every $\theta \in \mathbb{R}^{p+q}$, and it is the following product of $r + 1$ inverted gamma densities:

$$\pi(\sigma^2|\theta, y) = f_{\mathrm{IG}}\left(\sigma_e^2; \frac{N}{2} + a_e, b_e + \frac{\|y - W\theta\|^2}{2}\right) \times \prod_{i=1}^{r} f_{\mathrm{IG}}\left(\sigma_{u_i}^2; \frac{q_i}{2} + a_i, b_i + \frac{\|u_i\|^2}{2}\right),$$

where

$$f_{\mathrm{IG}}(v;c,d)=\begin{cases}\dfrac{d^c}{\Gamma(c)v^{c+1}}e^{-d/v}, & v>0,\\ 0, & v\le 0,\end{cases}$$

for $c,d>0$. On the other hand, if A is nonempty, then

$$\int_{\mathbb{R}_+^{r+1}}\pi^*(\theta,\sigma^2|y)\,d\sigma^2=\infty,$$

whenever $\theta\in\mathcal{N}:=\{\theta\in\mathbb{R}^{p+q}:\prod_{i\in A}\|u_i\|=0\}$. The fact that $\pi(\sigma^2|\theta,y)$ is not defined when $\theta\in\mathcal{N}$ is irrelevant from a simulation standpoint because the probability of observing a θ in $\mathcal{N}$ is zero. However, in order to perform a theoretical analysis, the Markov transition density (Mtd) of the Gibbs Markov chain must be defined for *every* $\theta\in\mathbb{R}^{p+q}$. Obviously, the Mtd can be defined arbitrarily on a set of measure zero. Thus, for $\theta\notin\mathcal{N}$, we define $\pi(\sigma^2|\theta,y)$ as in the case where A is empty, while if $\theta\in\mathcal{N}$, we define it to be $f_{\mathrm{IG}}(\sigma_e^2;1,1)\prod_{i=1}^r f_{\mathrm{IG}}(\sigma_{u_i}^2;1,1)$. Note that this definition can also be used when A is empty if we simply define $\mathcal{N}$ to be $\varnothing$ in that case.

The Mtd of the Gibbs–Markov chain, $\{(\theta_n,\sigma_n^2)\}_{n=0}^\infty$, is defined as

$$k(\theta,\sigma^2|\tilde\theta,\tilde\sigma^2)=\pi(\sigma^2|\theta,y)\pi(\theta|\tilde\sigma^2,y).$$

It is easy to see that the chain is ψ-irreducible, and that $\pi^*(\theta,\sigma^2|y)$ is an invariant density. It follows that the chain is positive recurrent if and only if the posterior is proper [Meyn and Tweedie (1993), Chapter 10]. Since a geometrically ergodic chain is necessarily positive recurrent, the Gibbs–Markov chain cannot be geometrically ergodic when the posterior is improper. The point here is that conditions implying geometric ergodicity also imply posterior propriety.

The marginal sequences, $\{\theta_n\}_{n=0}^\infty$ and $\{\sigma_n^2\}_{n=0}^\infty$, are themselves Markov chains; see, for example, Liu, Wong and Kong (1994). The σ^2-chain lives on $\mathbb{R}_+^{r+1}$ and has Mtd given by

$$k_1(\sigma^2|\tilde\sigma^2)=\int_{\mathbb{R}^{p+q}}\pi(\sigma^2|\theta,y)\pi(\theta|\tilde\sigma^2,y)\,d\theta$$

and invariant density $\int_{\mathbb{R}^{p+q}}\pi^*(\theta,\sigma^2|y)\,d\theta$. Similarly, the θ-chain lives on $\mathbb{R}^{p+q}$ and has Mtd

$$k_2(\theta|\tilde\theta)=\int_{\mathbb{R}_+^{r+1}}\pi(\theta|\sigma^2,y)\pi(\sigma^2|\tilde\theta,y)\,d\sigma^2$$

and invariant density $\int_{\mathbb{R}_+^{r+1}}\pi^*(\theta,\sigma^2|y)\,d\sigma^2$. Since the two marginal chains are also ψ-irreducible, they are positive recurrent if and only if the posterior is proper. Moreover, when the posterior is proper, routine calculations show that all three chains are Harris ergodic; that is, ψ-irreducible, aperiodic and positive Harris recurrent; see Román (2012) for details. An important fact that we will exploit is that geometric ergodicity is a solidarity property for the three chains $\{(\theta_n, \sigma_n^2)\}_{n=0}^{\infty}$, $\{\theta_n\}_{n=0}^{\infty}$ and $\{\sigma_n^2\}_{n=0}^{\infty}$; that is, either all three are geometric or none of them is [Diaconis, Khare and Saloff-Coste (2008), Liu, Wong and Kong (1994), Roberts and Rosenthal (2001)]. In the next section, we prove that the Gibbs–Markov chain converges at a geometric rate by proving that one of the marginal chains does.

3. The main result. In order to state the main result, we need a bit more notation. For $i \in \{1, \dots, r\}$, define R_i to be the $q_i \times q$ matrix of 0's and 1's such that $R_i u = u_i$. In other words, R_i is the matrix that *extracts* u_i from u. Here is our main result.

THEOREM 2. *Assume that* (S1)–(S4) *hold so that the Gibbs sampler is well defined. If the following two conditions hold, then the Gibbs–Markov chain is geometrically ergodic*:

(1) *For each* $i \in \{1, 2, \dots, r\}$, *one of the following holds*:

$$\text{(i)} \quad a_i < b_i = 0; \qquad \text{(ii)} \quad b_i > 0.$$

(2) *There exists an* $s \in (0, 1] \cap (0, \tilde{s}/2)$ *such that*

$$2^{-s}(p+t)^s \frac{\Gamma(N/2 + a_e - s)}{\Gamma(N/2 + a_e)} < 1 \tag{8}$$

and

$$2^{-s} \sum_{i=1}^{r} \left\{ \frac{\Gamma(q_i/2 + a_i - s)}{\Gamma(q_i/2 + a_i)} \right\} \big(\operatorname{tr}\big(R_i (I - P_{Z^T(I-P_X)Z}) R_i^T\big)\big)^s < 1, \tag{9}$$

where $t = \operatorname{rank}(Z^T(I - P_X)Z)$ *and* $P_{Z^T(I-P_X)Z}$ *is the projection onto the column space of* $Z^T(I - P_X)Z$.

REMARK 1. It is important to reiterate that, by themselves, (S1)–(S4) *do not* imply that the posterior density is proper. Of course, if conditions (1) and (2) in Theorem 2 hold as well, then the chain is geometric, so the posterior is necessarily proper.

REMARK 2. A numerical search could be employed to check the second condition of Theorem 2. Indeed, one could evaluate the left-hand sides of (8) and (9) at all values of s on a fine grid in the interval $(0, 1] \cap (0, \tilde{s}/2)$. The goal, of course, would be to find a single value of s at which both (8) and (9) are satisfied. It can be shown that, if there does exist an $s \in (0, 1] \cap (0, \tilde{s}/2)$ such that (8) and (9) hold, then $N + 2a_e > p + t$ and, for each $i = 1, 2, \dots, r$, $q_i + 2a_i > \operatorname{tr}(R_i(I - P_{Z^T(I-P_X)Z})R_i^T)$. Thus, it would behoove the user to verify these simple conditions before engaging in any numerical work.

REMARK 3. When evaluating (9), it may be helpful to write $P_{Z^T(I-P_X)Z}$ as $U^T P_\Lambda U$, where U and Λ are the orthogonal and diagonal matrices, respectively, in the spectral decomposition of $Z^T(I-P_X)Z$. That is, U is a q-dimensional orthogonal matrix and Λ is a diagonal matrix containing the eigenvalues of $Z^T(I-P_X)Z$. Of course, the projection P_Λ is a $q \times q$ binary diagonal matrix whose ith diagonal element is 1 if and only if the ith diagonal element of Λ is positive.

REMARK 4. Note that

$$\begin{aligned}\sum_{i=1}^r \operatorname{tr}\big(R_i(I-P_{Z^T(I-P_X)Z})R_i^T\big) &= \operatorname{tr}\left[(I-P_{Z^T(I-P_X)Z})\left(\sum_{i=1}^r R_i^T R_i\right)\right]\\ &= \operatorname{tr}(I-P_{Z^T(I-P_X)Z})\\ &= \operatorname{rank}(I-P_{Z^T(I-P_X)Z})\\ &= q-t.\end{aligned}$$

Moreover, when $r>1$, the matrix $I-P_{Z^T(I-P_X)Z}$ has $q=q_1+q_2+\cdots+q_r$ diagonal elements, and the (nonnegative) term $\operatorname{tr}(R_i(I-P_{Z^T(I-P_X)Z})R_i^T)$ is simply the sum of the q_i diagonal elements that correspond to the ith random factor.

REMARK 5. Recall from the Introduction that Corollary 1 provides an alternative set of sufficient conditions for geometric ergodicity that are harder to satisfy, but easier to check. A proof of Corollary 1 is given at the end of this section.

We will prove Theorem 2 indirectly by proving that the σ^2-chain is geometrically ergodic (when the conditions of Theorem 2 hold). This is accomplished by establishing a *geometric drift condition* for the σ^2-chain.

PROPOSITION 2. *Assume that* (S1)–(S4) *hold so that the Gibbs sampler is well defined. Under the two conditions of Theorem* 2, *there exist a* $\rho \in [0,1)$ *and a finite constant* L *such that, for every* $\tilde{\sigma}^2 \in \mathbb{R}_+^{r+1}$,

$$E\big(v(\sigma^2)|\tilde{\sigma}^2\big) \le \rho v(\tilde{\sigma}^2) + L, \tag{10}$$

where the drift function is defined as

$$v(\sigma^2) = \alpha(\sigma_e^2)^s + \sum_{i=1}^r (\sigma_{u_i}^2)^s + \alpha(\sigma_e^2)^{-c} + \sum_{i=1}^r (\sigma_{u_i}^2)^{-c},$$

and α *and* c *are positive constants. Hence, under the two conditions of Theorem* 2, *the* σ^2*-chain is geometrically ergodic.*

REMARK 6. The formulas for $\rho = \rho(\alpha, s, c)$ and $L = L(\alpha, s, c)$ are provided in the proof, as is a set of acceptable values for the pair (α, c). Recall that the value of s is given to us in the hypothesis of Theorem 2.

PROOF OF PROPOSITION 2. The proof has two parts. In part I, we establish the validity of the geometric drift condition, (10). In part II, we use results from Meyn and Tweedie (1993) to show that (10) implies geometric ergodicity of the σ^2-chain.

Part I. By conditioning on θ and iterating, we can express $E(v(\sigma^2)|\tilde{\sigma}^2)$ as

$$E\left[\alpha E\big((\sigma_e^2)^s|\theta\big) + E\left(\sum_{i=1}^r (\sigma_{u_i}^2)^s|\theta\right) + \alpha E\big((\sigma_e^2)^{-c}|\theta\big) + E\left(\sum_{i=1}^r (\sigma_{u_i}^2)^{-c}|\theta\right)\middle|\tilde{\sigma}^2\right].$$

We now develop upper bounds for each of the four terms inside the square brackets. Fix $s \in S := (0, 1] \cap (0, \tilde{s}/2)$, and define

$$G_0(s) = 2^{-s}\frac{\Gamma(N/2 + a_e - s)}{\Gamma(N/2 + a_e)}$$

and, for each $i \in \{1, 2, \dots, r\}$, define

$$G_i(s) = 2^{-s}\frac{\Gamma(q_i/2 + a_i - s)}{\Gamma(q_i/2 + a_i)}.$$

Note that, since $s \in (0, 1]$, $(x_1 + x_2)^s \le x_1^s + x_2^s$ whenever $x_1, x_2 \ge 0$. Thus,

$$\begin{aligned}
E\big((\sigma_e^2)^s|\theta\big) &= 2^s G_0(s)\left(b_e + \frac{\|y - W\theta\|^2}{2}\right)^s \\
&\le 2^s G_0(s)\left[b_e^s + \left(\frac{\|y - W\theta\|^2}{2}\right)^s\right] \\
&= G_0(s)\big(\|y - W\theta\|^2\big)^s + 2^s G_0(s) b_e^s.
\end{aligned}$$

Similarly,

$$E\big((\sigma_{u_i}^2)^s|\theta\big) = 2^s G_i(s)\left(b_i + \frac{\|u_i\|^2}{2}\right)^s \le G_i(s)\big(\|u_i\|^2\big)^s + 2^s G_i(s) b_i^s.$$

Now, for any $c > 0$, we have

$$\begin{aligned}
E\big((\sigma_e^2)^{-c}|\theta\big) &= 2^{-c} G_0(-c)\left(b_e + \frac{\|y - W\theta\|^2}{2}\right)^{-c} \\
&\le 2^{-c} G_0(-c)\left(b_e + \frac{\mathrm{SSE}}{2}\right)^{-c}
\end{aligned}$$

and, for each $i \in \{1, 2, \dots, r\}$,

$$\begin{aligned}
E\big((\sigma_{u_i}^2)^{-c}|\theta\big) &= 2^{-c} G_i(-c)\left(b_i + \frac{\|u_i\|^2}{2}\right)^{-c} \\
&\le G_i(-c)\big[\big(\|u_i\|^2\big)^{-c} I_{\{0\}}(b_i) + (2b_i)^{-c} I_{(0,\infty)}(b_i)\big].
\end{aligned}$$

Recall that $A = \{i : b_i = 0\}$, and note that $E(\sum_{i=1}^r (\sigma_{u_i}^2)^{-c}|\theta)$ can be bounded above by a constant if A is empty. Thus, we consider the case in which A is empty separately from the case where $A \neq \varnothing$. We begin with the latter, which is the more difficult case.

Case I: A is nonempty. Combining the four bounds above (and applying Jensen's inequality twice), we have

$$\begin{aligned} E\big(v(\sigma^2)|\tilde{\sigma}^2\big) &\le \alpha G_0(s)\big[E\big(\|y - W\theta\|^2|\tilde{\sigma}^2\big)\big]^s + \sum_{i=1}^r G_i(s)\big[E\big(\|u_i\|^2|\tilde{\sigma}^2\big)\big]^s \\ &\quad + \sum_{i \in A} G_i(-c)E\big[\|u_i\|^{-2c}|\tilde{\sigma}^2\big] + \kappa(\alpha, s, c), \end{aligned} \tag{11}$$

where

$$\kappa(\alpha, s, c) = \alpha 2^s G_0(s) b_e^s + 2^s \sum_{i=1}^r G_i(s) b_i^s + \alpha 2^{-c} G_0(-c)\left(b_e + \frac{\text{SSE}}{2}\right)^{-c} + \sum_{i : b_i > 0} G_i(-c)(2b_i)^{-c}.$$

Appendix A.2 contains a proof of the following inequality:

$$E\big[\|y - W\theta\|^2|\tilde{\sigma}^2\big] \le (p + t)\tilde{\sigma}_e^2 + \big(\|(I - P_X)y\| + \|(I - P_X)Z\|K\big)^2, \tag{12}$$

where $\|\cdot\|$ with a matrix argument denotes the Frobenius norm, and the constant $K = K(X, Z, y)$ is defined and shown to be finite in Appendix A.1. It follows immediately that

$$\big[E\big(\|y - W\theta\|^2|\tilde{\sigma}^2\big)\big]^s \le (p + t)^s(\tilde{\sigma}_e^2)^s + \big(\|(I - P_X)y\| + \|(I - P_X)Z\|K\big)^{2s}.$$

In Appendix A.3, it is shown that, for each $i \in \{1, 2, \ldots, r\}$, we have

$$E\big[\|u_i\|^2|\tilde{\sigma}^2\big] \le \xi_i\tilde{\sigma}_e^2 + \zeta_i \sum_{j=1}^r \tilde{\sigma}_{u_j}^2 + \big(\|R_i\|K\big)^2,$$

where $\xi_i = \text{tr}(R_i(Z^T(I - P_X)Z)^+ R_i^T)$, $\zeta_i = \text{tr}(R_i(I - P_{Z^T(I-P_X)Z})R_i^T)$ and A^+ denotes the Moore–Penrose inverse of the matrix A. It follows that

$$\big[E\big(\|u_i\|^2|\tilde{\sigma}^2\big)\big]^s \le \xi_i^s(\tilde{\sigma}_e^2)^s + \zeta_i^s \sum_{j=1}^r (\tilde{\sigma}_{u_j}^2)^s + \big(\|R_i\|K\big)^{2s}. \tag{13}$$

In Appendix A.4, it is established that, for any $c \in (0, 1/2)$, and for each $i \in \{1, 2, \ldots, r\}$, we have

$$E\big[\|u_i\|^{-2c}|\tilde{\sigma}^2\big] \le 2^{-c}\frac{\Gamma(q_i/2 - c)}{\Gamma(q_i/2)}\big[\lambda_{\max}^c(\tilde{\sigma}_e^2)^{-c} + (\tilde{\sigma}_{u_i}^2)^{-c}\big], \tag{14}$$

where $\lambda_{\max}$ denotes the largest eigenvalue of $Z^T(I-P_X)Z$. Using (12)–(14) in (11), we have

$$
\begin{aligned}
E\big(v(\sigma^2)|\tilde{\sigma}^2\big) \le{}& \alpha\left(\delta_1(s)+\frac{\delta_2(s)}{\alpha}\right)\big(\tilde{\sigma}_e^2\big)^s+\delta_3(s)\sum_{j=1}^r\big(\tilde{\sigma}_{u_j}^2\big)^s \\
&+\alpha\frac{\delta_4(c)}{\alpha}\big(\tilde{\sigma}_e^2\big)^{-c}+\delta_5(c)\sum_{j\in A}\big(\tilde{\sigma}_{u_j}^2\big)^{-c}+L(\alpha,s,c),
\end{aligned}
\tag{15}
$$

where

$$
\delta_1(s):=G_0(s)(p+t)^s,\qquad \delta_2(s):=\sum_{i=1}^r\xi_i^sG_i(s),\qquad \delta_3(s):=\sum_{i=1}^r\zeta_i^sG_i(s),
$$

$$
\delta_4(c):=2^{-c}\lambda_{\max}^c\sum_{i\in A}G_i(-c)\frac{\Gamma(q_i/2-c)}{\Gamma(q_i/2)},
$$

$$
\delta_5(c):=2^{-c}\max_{i\in A}\left[G_i(-c)\frac{\Gamma(q_i/2-c)}{\Gamma(q_i/2)}\right]
$$

and

$$
\begin{aligned}
L(\alpha,s,c)={}&\kappa(\alpha,s,c)+\alpha G_0(s)\big(\|(I-P_X)y\|+\|(I-P_X)Z\|K\big)^{2s} \\
&+\sum_{i=1}^rG_i(s)\big(\|R_i\|K\big)^{2s}.
\end{aligned}
$$

Hence,

$$
E\big(v(\sigma^2)|\tilde{\sigma}^2\big)\le\rho(\alpha,s,c)v\big(\tilde{\sigma}^2\big)+L(\alpha,s,c),
$$

where

$$
\rho(\alpha,s,c)=\max\left\{\delta_1(s)+\frac{\delta_2(s)}{\alpha},\delta_3(s),\frac{\delta_4(c)}{\alpha},\delta_5(c)\right\}.
$$

We must now show that there exists a triple $(\alpha,s,c)\in\mathbb{R}_+\times S\times(0,1/2)$ such that $\rho(\alpha,s,c)<1$. We begin by demonstrating that, if c is small enough, then $\delta_5(c)<1$. Define $\tilde{a}=-\max_{i\in A}a_i$. Also, set $C=(0,1/2)\cap(0,\tilde{a})$. Fix $c\in C$ and note that

$$
\delta_5(c)=\max_{i\in A}\left[\frac{\Gamma(q_i/2+a_i+c)}{\Gamma(q_i/2+a_i)}\frac{\Gamma(q_i/2-c)}{\Gamma(q_i/2)}\right].
$$

For any $i\in A$, $c+a_i<0$, and since $\tilde{s}>0$, it follows that

$$
0<\frac{q_i}{2}+a_i<\frac{q_i}{2}+a_i+c<\frac{q_i}{2}.
$$

But, $\Gamma(x-z)/\Gamma(x)$ is decreasing in x for $x>z>0$, so we have

$$
\frac{\Gamma(q_i/2+a_i)}{\Gamma(q_i/2+a_i+c)}=\frac{\Gamma(q_i/2+a_i+c-c)}{\Gamma(q_i/2+a_i+c)}>\frac{\Gamma(q_i/2-c)}{\Gamma(q_i/2)},
$$

and it follows immediately that $\delta_5(c) < 1$ whenever $c \in C$. The two conditions of Theorem 2 imply that there exists an $s^\star \in S$ such that $\delta_1(s^\star) < 1$ and $\delta_3(s^\star) < 1$. Let $c^\star$ be any point in C, and choose $\alpha^\star$ to be any number larger than

$$\max\left\{\frac{\delta_2(s^\star)}{1-\delta_1(s^\star)}, \delta_4(c^\star)\right\}.$$

A simple calculation shows that $\rho(\alpha^\star, s^\star, c^\star) < 1$, and this completes the argument for case I.

Case II: $A = \varnothing$. Since we no longer have to deal with $E(\sum_{i=1}^r (\sigma_{u_i}^2)^{-c}|\theta)$, bound (15) becomes

$$E\big(v(\sigma^2)|\tilde{\sigma}^2\big) \le \alpha\left(\delta_1(s) + \frac{\delta_2(s)}{\alpha}\right)\big(\tilde{\sigma}_e^2\big)^s + \delta_3(s)\sum_{j=1}^r \big(\tilde{\sigma}_{u_j}^2\big)^s + L(\alpha, s, c),$$

and there is no restriction on c other than $c > 0$. [Note that the constant term $L(\alpha, s, c)$ requires no alteration when we move from case I to case II.] Hence,

$$E\big(v(\sigma^2)|\tilde{\sigma}^2\big) \le \rho(\alpha, s)v\big(\tilde{\sigma}^2\big) + L(\alpha, s, c),$$

where

$$\rho(\alpha, s) = \max\left\{\delta_1(s) + \frac{\delta_2(s)}{\alpha}, \delta_3(s)\right\}.$$

We must now show that there exists a $(\alpha, s) \in \mathbb{R}_+ \times S$ such that $\rho(\alpha, s) < 1$. As in case I, the two conditions of Theorem 2 imply that there exists an $s^\star \in S$ such that $\delta_1(s^\star) < 1$ and $\delta_3(s^\star) < 1$. Let $\alpha^\star$ be any number larger than

$$\frac{\delta_2(s^\star)}{1-\delta_1(s^\star)}.$$

A simple calculation shows that $\rho(\alpha^\star, s^\star) < 1$, and this completes the argument for case II. This completes part I of the proof.

Part II. We begin by establishing that the σ^2-chain satisfies certain properties. Recall that its Mtd is given by

$$k_1(\sigma^2|\tilde{\sigma}^2) = \int_{\mathbb{R}^{p+q}} \pi(\sigma^2|\theta, y)\pi(\theta|\tilde{\sigma}^2, y)\, d\theta.$$

Note that k_1 is strictly positive on $\mathbb{R}_+^{r+1} \times \mathbb{R}_+^{r+1}$. It follows that the σ^2-chain is ψ-irreducible and aperiodic, and that its maximal irreducibility measure is equivalent to Lebesgue measure on $\mathbb{R}_+^{r+1}$; for definitions, see Meyn and Tweedie (1993), Chapters 4 and 5. Let P_1 denote the Markov transition function of the σ^2-chain; that is, for any $\tilde{\sigma}^2 \in \mathbb{R}_+^{r+1}$ and any Borel set A,

$$P_1(\tilde{\sigma}^2, A) = \int_A k_1(\sigma^2|\tilde{\sigma}^2)\, d\sigma^2.$$

We now demonstrate that the σ^2-chain is a Feller chain; that is, for each fixed open set O, $P_1(\cdot, O)$ is a lower semi-continuous function on $\mathbb{R}_+^{r+1}$. Indeed, let $\{\tilde{\sigma}_m^2\}_{m=1}^{\infty}$ be a sequence in $\mathbb{R}_+^{r+1}$ that converges to $\tilde{\sigma}^2 \in \mathbb{R}_+^{r+1}$. Then

$$
\begin{aligned}
\liminf_{m\to\infty} P_1(\tilde{\sigma}_m^2, O) &= \liminf_{m\to\infty} \int_O k_1(\sigma^2|\tilde{\sigma}_m^2)\, d\sigma^2 \\
&= \liminf_{m\to\infty} \int_O \left[\int_{\mathbb{R}^{p+q}} \pi(\sigma^2|\theta, y)\pi(\theta|\tilde{\sigma}_m^2, y)\, d\theta \right] d\sigma^2 \\
&\ge \int_O \int_{\mathbb{R}^{p+q}} \pi(\sigma^2|\theta, y)\Big[\liminf_{m\to\infty} \pi(\theta|\tilde{\sigma}_m^2, y)\Big]\, d\theta\, d\sigma^2 \\
&= \int_O \left[\int_{\mathbb{R}^{p+q}} \pi(\sigma^2|\theta, y)\pi(\theta|\tilde{\sigma}^2, y)\, d\theta \right] d\sigma^2 \\
&= P_1(\tilde{\sigma}^2, O),
\end{aligned}
$$

where the inequality follows from Fatou's lemma, and the third equality follows from the fact that $\pi(\theta|\sigma^2, y)$ is continuous in σ^2; for a proof of continuity, see Román (2012). We conclude that $P_1(\cdot, O)$ is lower semi-continuous, so the σ^2-chain is Feller.

The last thing we must do before we can appeal to the results in Meyn and Tweedie (1993) is to show that the drift function, $v(\cdot)$, is unbounded off compact sets; that is, we must show that, for every $d \in \mathbb{R}$, the set

$$S_d = \{\sigma^2 \in \mathbb{R}_+^{r+1} : v(\sigma^2) \le d\}$$

is compact. Let d be such that S_d is nonempty (otherwise S_d is trivially compact), which means that d and d/α must be larger than 1. Since $v(\sigma^2)$ is a continuous function, S_d is closed in $\mathbb{R}_+^{r+1}$. Now consider the following set:

$$T_d = \big[(d/\alpha)^{-1/c}, (d/\alpha)^{1/s}\big] \times \big[d^{-1/c}, d^{1/s}\big] \times \cdots \times \big[d^{-1/c}, d^{1/s}\big].$$

The set T_d is compact in $\mathbb{R}_+^{r+1}$. Since $S_d \subset T_d$, S_d is a closed subset of a compact set in $\mathbb{R}_+^{r+1}$, so it is compact in $\mathbb{R}_+^{r+1}$. Hence, the drift function is unbounded off compact sets.

Since the σ^2-chain is Feller and its maximal irreducibility measure is equivalent to Lebesgue measure on $\mathbb{R}_+^{r+1}$, Meyn and Tweedie's (1993) Theorem 6.0.1 shows that every compact set in $\mathbb{R}_+^{r+1}$ is petite. Hence, for each $d \in \mathbb{R}$, the set S_d is petite, so $v(\cdot)$ is unbounded off petite sets. It now follows from the drift condition (10) and an application of Meyn and Tweedie's (1993) Lemma 15.2.8 that condition (iii) of Meyn and Tweedie's (1993) Theorem 15.0.1 is satisfied, so the σ^2-chain is geometrically ergodic. This completes part II of the proof. $\square$

We end this section with a proof of Corollary 1.

PROOF OF COROLLARY 1. It suffices to show that, together, conditions (B′) and (C′) of Corollary 1 imply the second condition of Theorem 2. Clearly, (B′) and (C′) imply that $\tilde{s}/2 > 1$, so $(0, 1] \cap (0, \tilde{s}/2) = (0, 1]$. Take $s^\star = 1$. Condition (C′) implies

$$2^{-s^\star}(p+t)^{s^\star}\frac{\Gamma(N/2+a_e-s^\star)}{\Gamma(N/2+a_e)} = \frac{p+t}{N+2a_e-2} < 1.$$

Now, we know from Remark 4 that $\sum_{i=1}^r \operatorname{tr}(R_i(I - P_{Z^T(I-P_X)Z})R_i^T) = q - t$. Hence,

$$\begin{aligned}
2^{-s^\star}&\sum_{i=1}^r\left\{\frac{\Gamma(q_i/2+a_i-s^\star)}{\Gamma(q_i/2+a_i)}\right\}\bigl(\operatorname{tr}\bigl(R_i(I-P_{Z^T(I-P_X)Z})R_i^T\bigr)\bigr)^{s^\star}\\
&=\sum_{i=1}^r\frac{\operatorname{tr}(R_i(I-P_{Z^T(I-P_X)Z})R_i^T)}{q_i+2a_i-2}\\
&\le\frac{\sum_{i=1}^r\operatorname{tr}(R_i(I-P_{Z^T(I-P_X)Z})R_i^T)}{\min_{j\in\{1,2,\ldots,r\}}\{q_j+2a_j-2\}}\\
&=\frac{q-t}{\min_{j\in\{1,2,\ldots,r\}}\{q_j+2a_j-2\}} < 1,
\end{aligned}$$

where the last inequality follows from condition (B′). □

4. An application of the main result. In this section, we illustrate the application of Theorem 2 using the two-way random effects model with one observation per cell. The model equation is

$$Y_{ij} = \beta + \alpha_i + \gamma_j + \varepsilon_{ij},$$

where $i = 1, 2, \ldots, m$, $j = 1, 2, \ldots, n$, the α_i's are i.i.d. $N(0, \sigma_\alpha^2)$, the γ_j's are i.i.d. $N(0, \sigma_\gamma^2)$ and the ε_{ij}'s are i.i.d. $N(0, \sigma_e^2)$. The α_i's, γ_j's and ε_{ij}'s are all independent.

We begin by explaining how to put this model in GLMM (matrix) form. There are a total of $N = m \times n$ observations and we arrange them using the usual (lexicographical) ordering

$$Y = (Y_{11}\cdots Y_{1n}\quad Y_{21}\cdots Y_{2n}\quad \cdots\quad Y_{m1}\cdots Y_{mn})^T.$$

Since β is a univariate parameter common to all of the observations, $p = 1$ and X is an $N \times 1$ column vector of ones, which we denote by 1_N. There are $r = 2$ random factors with $q_1 = m$ and $q_2 = n$, so $q = m + n$. Letting $\otimes$ denote the Kronecker product, we can write the Z matrix as $(Z_1\ \ Z_2)$, where $Z_1 = I_m \otimes 1_n$ and $Z_2 = 1_m \otimes I_n$. We assume throughout this section that SSE > 0.

We now examine conditions (8) and (9) of Theorem 2 for this particular model. A routine calculation shows that $Z^T(I - P_X)Z$ is a block diagonal matrix given by

$$n\left(I_m - \frac{1}{m}J_m\right) \oplus m\left(I_n - \frac{1}{n}J_n\right),$$

where J_d is a $d \times d$ matrix of ones (and $\oplus$ is the direct sum operator). It follows immediately that

$$t = \operatorname{rank}(Z^T(I - P_X)Z) = \operatorname{rank}\left(I_m - \frac{1}{m}J_m\right) + \operatorname{rank}\left(I_n - \frac{1}{n}J_n\right) = m + n - 2.$$

Hence, (8) becomes

$$2^{-s}(m + n - 1)^s \frac{\Gamma(mn/2 + a_e - s)}{\Gamma(mn/2 + a_e)} < 1.$$

Now, it can be shown that

$$I - P_{Z^T(I-P_X)Z} = \left(\frac{1}{m}J_m\right) \oplus \left(\frac{1}{n}J_n\right).$$

Hence, we have

$$\operatorname{tr}(R_1(I - P_{Z^T(I-P_X)Z})R_1^T) = \operatorname{tr}\left(\frac{1}{m}J_m\right) = 1$$

and

$$\operatorname{tr}(R_2(I - P_{Z^T(I-P_X)Z})R_2^T) = \operatorname{tr}\left(\frac{1}{n}J_n\right) = 1.$$

Therefore, (9) reduces to

$$2^{-s}\left\{\frac{\Gamma(m/2 + a_1 - s)}{\Gamma(m/2 + a_1)} + \frac{\Gamma(n/2 + a_2 - s)}{\Gamma(n/2 + a_2)}\right\} < 1.$$

Now consider a concrete example in which $m = 5$, $n = 6$, and the prior is

$$\frac{I_{\mathbb{R}_+}(\sigma_e^2)I_{\mathbb{R}_+}(\sigma_\alpha^2)I_{\mathbb{R}_+}(\sigma_\gamma^2)}{\sigma_e^2\sqrt{\sigma_\alpha^2\sigma_\gamma^2}}.$$

So, we are taking $b_e = b_1 = b_2 = a_e = 0$ and $a_1 = a_2 = -1/2$. Corollary 1 implies that the Gibbs–Markov chain is geometrically ergodic whenever $m, n \geq 6$, but this result is not applicable when $m = 5$ and $n = 6$. Hence, we turn to Theorem 2. In this case, $\tilde{s} = 4$, so we need to find an $s \in (0, 1]$ such that

$$2^{-s}\max\left\{(10)^s \frac{\Gamma(30/2 + 0 - s)}{\Gamma(30/2 + 0)},\right.$$
$$\left.\frac{\Gamma(5/2 + (-1/2) - s)}{\Gamma(5/2 + (-1/2))} + \frac{\Gamma(6/2 + (-1/2) - s)}{\Gamma(6/2 + (-1/2))}\right\} < 1.$$

The reader can check that, when $s = 0.9$, the left-hand side is approximately 0.87. Therefore, Theorem 2 implies that the Gibbs–Markov chain is geometrically ergodic in this case.

5. Specializing to the one-way random effects model. The only other existing results on geometric convergence of Gibbs samplers for linear mixed models with *improper* priors are those of Tan and Hobert (2009) (hereafter, T&H). These authors considered the one-way random effects model, which is a simple, but important special case of the GLMM given in (1). In this section, we show that our results improve upon those of T&H.

Recall that the one-way model is given by

$$Y_{ij} = \beta + \alpha_i + e_{ij}, \tag{16}$$

where $i = 1, \dots, c$, $j = 1, \dots, n_i$, the α_i's are i.i.d. $\mathrm{N}(0, \sigma_\alpha^2)$, and the e_{ij}'s, which are independent of the α_i's, are i.i.d. $\mathrm{N}(0, \sigma_e^2)$. It is easy to see that (16) is a special case of the GLMM. Obviously, there are a total of $N = n_1 + \cdots + n_c$ observations, and we arrange them in a column vector with the usual ordering as follows:

$$Y = (\,Y_{11} \cdots Y_{1n_1} \quad Y_{21} \cdots Y_{2n_2} \quad \cdots \quad Y_{c1} \cdots Y_{cn_c}\,)^T .$$

As in the two-way model of Section 4, β is a univariate parameter common to all of the observations, so $p = 1$ and $X = 1_N$. Here there is only one random factor (with c levels), so $r = 1$, $q = q_1 = c$ and $Z = \bigoplus_{i=1}^c 1_{n_i}$. Of course, in this case, $\mathrm{SSE} = \sum_{i=1}^c \sum_{j=1}^{n_i} (y_{ij} - \overline{y}_i)^2$, where $\overline{y}_i = n_i^{-1} \sum_{j=1}^{n_i} y_{ij}$. We assume throughout this section that $\mathrm{SSE} > 0$.

We note that T&H actually considered a slightly different parameterization of the one-way model. In their version, β does not appear in the model equation (16), but rather as the mean of the u_i's. In other words, T&H used the "centered" parameterization, whereas here we are using the "noncentered" parameterization. Román (2012) shows that, because β and $(\alpha_1 \cdots \alpha_c)^T$ are part of a single "block" in the Gibbs sampler, the centered and noncentered versions of the Gibbs sampler converge at exactly the same rate.

T&H considered improper priors for $(\beta, \sigma_e^2, \sigma_\alpha^2)$ that take the form

$$(\sigma_e^2)^{-(a_e+1)} I_{\mathbb{R}_+}(\sigma_e^2) (\sigma_\alpha^2)^{-(a_1+1)} I_{\mathbb{R}_+}(\sigma_\alpha^2),$$

and they showed that the Gibbs sampler for the one-way model is geometrically ergodic if $a_1 < 0$ and

$$\begin{gathered} N + 2a_e \geq c + 3 \quad \text{and} \\ c \min\left\{ \left(\sum_{i=1}^c \frac{n_i}{n_i + 1} \right)^{-1}, \frac{n^*}{N} \right\} < 2 \exp\left\{ \Psi\left(\frac{c}{2} + a_1 \right) \right\}, \end{gathered} \tag{17}$$

where $n^* = \max\{n_1, n_2, \dots, n_c\}$ and $\Psi(x) = \frac{d}{dx} \log(\Gamma(x))$ is the digamma function.

We now consider the implications of Theorem 2 in the case of the one-way model. First, $t = \mathrm{rank}(Z^T (I - P_X) Z) = c - 1$. Combining this fact with Remark 4, it follows that the two conditions of Theorem 2 will hold if $a_1 < 0$, and there exists an $s \in (0,1) \cap (0, a_1 + \frac{c}{2}) \cap (0, a_e + \frac{N}{2})$ such that

$$2^{-s} \max\left\{ c^s \frac{\Gamma(N/2 + a_e - s)}{\Gamma(N/2 + a_e)}, \frac{\Gamma(c/2 + a_1 - s)}{\Gamma(c/2 + a_1)} \right\} < 1.$$

Román (2012) shows that such an s does indeed exist (so the Gibbs chain is geometrically ergodic) when

$$N + 2a_e \geq c + 2 \quad \text{and} \quad 1 < 2\exp\left\{ \Psi\left(\frac{c}{2} + a_1\right)\right\}. \tag{18}$$

Now, it's easy to show that

$$1 \leq c \min\left\{ \left(\sum_{i=1}^{c} \frac{n_i}{n_i + 1} \right)^{-1}, \frac{n^*}{N} \right\}.$$

Consequently, if (17) holds, then so does (18). In other words, our sufficient conditions are weaker than those of T&H, so our result improves upon theirs. Moreover, in contrast with the conditions of T&H, our conditions do not directly involve the group sample sizes, $n_1, n_2, \dots, n_c$.

Of course, the best result possible would be that the Gibbs–Markov chain is geometrically ergodic whenever the posterior is proper. Our result is very close to the best possible in the important case where the standard diffuse prior is used; that is, when $a_1 = -1/2$ and $a_e = 0$. The posterior is proper in this case if and only if $c \geq 3$ [Sun, Tsutakawa and He (2001)]. It follows from (18) that the Gibbs–Markov chain is geometrically ergodic as long as $c \geq 3$, and the total sample size, $N = n_1 + n_2 + \cdots + n_c$, is at least $c + 2$. This additional sample size condition is *extremely weak*. Indeed, the positivity of SSE implies that $N \geq c + 1$, so, for fixed $c \geq 3$, our condition for geometric ergodicity fails only in the single case where $N = c + 1$. Interestingly, in this case, the conditions of Corollary 1 reduce to $c \geq 5$ and $N \geq c + 3$.

6. Discussion. Our decision to work with the σ^2-chain rather than the θ-chain was based on an important technical difference between the two chains that stems from the fact that $\pi(\sigma^2|\theta, y)$ is not continuous in θ for each fixed σ^2 (when the set A is nonempty). Indeed, recall that, for $\theta \notin \mathcal{N}$,

$$\pi(\sigma^2|\theta, y) = f_{\mathrm{IG}}\left(\sigma_e^2; \frac{N}{2} + a_e, b_e + \frac{\|y - W\theta\|^2}{2}\right) \times \prod_{i=1}^{r} f_{\mathrm{IG}}\left(\sigma_{u_i}^2; \frac{q_i}{2} + a_i, b_i + \frac{\|u_i\|^2}{2}\right),$$

but for $\theta \in \mathcal{N}$,

$$\pi(\sigma^2|\theta, y) = f_{\mathrm{IG}}(\sigma_e^2; 1, 1) \prod_{i=1}^{r} f_{\mathrm{IG}}(\sigma_{u_i}^2; 1, 1).$$

Also, recall that the Mtd of the σ^2-chain is given by

$$k_1(\sigma^2|\tilde{\sigma}^2) = \int_{\mathbb{R}^{p+q}} \pi(\sigma^2|\theta, y)\pi(\theta|\tilde{\sigma}^2, y)\, d\theta.$$

Since the set $\mathcal{N}$ has measure zero, the "arbitrary part" of $\pi(\sigma^2|\theta, y)$ washes out of k_1. However, the same cannot be said for the θ-chain, whose Mtd is given by

$$k_2(\theta|\tilde{\theta}) = \int_{\mathbb{R}_+^{r+1}} \pi(\theta|\sigma^2, y)\pi(\sigma^2|\tilde{\theta}, y)\, d\sigma^2.$$

This difference between k_1 and k_2 comes into play when we attempt to apply certain "topological" results from Markov chain theory, such as those in Chapter 6 of Meyn and Tweedie (1993). In particular, in our proof that the σ^2-chain is a Feller chain (which was part of the proof of Proposition 2), we used the fact that $\pi(\theta|\sigma^2, y)$ is continuous in σ^2 for each fixed θ. Since $\pi(\sigma^2|\theta, y)$ is not continuous, we cannot use the same argument to prove that the θ-chain is Feller. In fact, we suspect that the θ-chain is not Feller, and if this is true, it means that our method of proof will not work for the θ-chain.

It is possible to circumvent the problem described above by removing the set $\mathcal{N}$ from the state space of the θ-chain. In this case, we are no longer required to define $\pi(\sigma^2|\theta, y)$ for $\theta \in \mathcal{N}$, and since $\pi(\sigma^2|\theta, y)$ is continuous (for fixed σ^2) on $\mathbb{R}^{p+q} \setminus \mathcal{N}$, the Feller argument for the θ-chain will go through. On the other hand, the new state space has "holes" in it, and this could complicate the search for a drift function that is unbounded off compact sets. For example, consider a toy drift function given by $v(x) = x^2$. This function is clearly unbounded off compact sets when the state space is $\mathbb{R}$, but not when the state space is $\mathbb{R} \setminus \{0\}$. The modified drift function $v^*(x) = x^2 + 1/x^2$ is unbounded off compact sets for the "holey" state space.

T&H overlooked a set of measure zero (similar to our $\mathcal{N}$), and this oversight led to an error in the proof of their main result (Proposition 3). However, Román (2012) shows that T&H's proof can be repaired and that their result is correct as stated. The fix involves deleting the offending null set from the state space, and adding a term to the drift function.

APPENDIX: UPPER BOUNDS

A.1. Preliminary results. Here is our first result.

LEMMA 1. *The following inequalities hold for all $\sigma^2 \in \mathbb{R}_+^{r+1}$ and all $i \in \{1, 2, \ldots, r\}$:*

(1) $Q^{-1} \preceq (Z^T(I - P_X)Z)^+\sigma_e^2 + (I - P_{Z^T(I-P_X)Z})(\sum_{j=1}^r \sigma_{u_j}^2)$;
(2) $\operatorname{tr}((I - P_X)ZQ^{-1}Z^T(I - P_X)) \leq \operatorname{rank}(Z^T(I - P_X)Z)\sigma_e^2$;
(3) $(R_iQ^{-1}R_i^T)^{-1} \preceq ((\sigma_e^2)^{-1}\lambda_{\max} + (\sigma_{u_i}^2)^{-1})I_{q_i}$.

PROOF. Recall from Section 3 that $U^T \Lambda U$ is the spectral decomposition of $Z^T(I - P_X)Z$, and that P_Λ is a binary diagonal matrix whose ith diagonal element is 1 if and only if the ith diagonal element of Λ is positive. Let $\sigma_\bullet^2 = \sum_{j=1}^r \sigma_{u_j}^2$. Since $(\sigma_\bullet^2)^{-1} I_q \preceq D^{-1}$, we have

$$(\sigma_e^2)^{-1} Z^T(I - P_X)Z + (\sigma_\bullet^2)^{-1} I_q \preceq (\sigma_e^2)^{-1} Z^T(I - P_X)Z + D^{-1},$$

and this yields

$$\begin{aligned} Q^{-1} &= \big((\sigma_e^2)^{-1} Z^T(I - P_X)Z + D^{-1}\big)^{-1} \\ &\preceq \big((\sigma_e^2)^{-1} Z^T(I - P_X)Z + (\sigma_\bullet^2)^{-1} I_q\big)^{-1} \\ &= U^T\big(\Lambda(\sigma_e^2)^{-1} + I_q(\sigma_\bullet^2)^{-1}\big)^{-1} U. \end{aligned} \tag{19}$$

Now let Λ^+ be a diagonal matrix whose diagonal elements, $\{\lambda_i^+\}_{i=1}^q$, are given by

$$\lambda_i^+ = \begin{cases} \lambda_i^{-1}, & \lambda_i \neq 0, \\ 0, & \lambda_i = 0. \end{cases}$$

Note that, for each $i \in \{1, 2, \dots, r\}$, we have

$$\frac{1}{\lambda_i(\sigma_e^2)^{-1} + (\sigma_\bullet^2)^{-1}} \le \lambda_i^+ \sigma_e^2 + I_{\{0\}}(\lambda_i)\sigma_\bullet^2.$$

This shows that

$$\big(\Lambda(\sigma_e^2)^{-1} + I_q(\sigma_\bullet^2)^{-1}\big)^{-1} \preceq \Lambda^+ \sigma_e^2 + (I - P_\Lambda)\sigma_\bullet^2.$$

Together with (19), this leads to

$$\begin{aligned} Q^{-1} &\preceq U^T\big(\Lambda(\sigma_e^2)^{-1} + I_q(\sigma_\bullet^2)^{-1}\big)^{-1} U \preceq U^T\big(\Lambda^+\sigma_e^2 + (I - P_\Lambda)\sigma_\bullet^2\big)U \\ &= \big(Z^T(I - P_X)Z\big)^+ \sigma_e^2 + U^T(I - P_\Lambda)U\sigma_\bullet^2. \end{aligned}$$

So to prove the first statement, it remains to show that $U^T(I - P_\Lambda)U = I - P_{Z^T(I-P_X)Z}$. But notice that letting $A = Z^T(I - P_X)Z$ and using its spectral decomposition, we have

$$\begin{aligned} A(A^T A)^+ A^T &= U^T \Lambda U\big(U^T \Lambda^T \Lambda U\big)^+ U^T \Lambda U \\ &= U^T \Lambda\big(\Lambda^T \Lambda\big)^+ \Lambda U = U^T P_\Lambda U, \end{aligned}$$

which implies that

$$I - P_{Z^T(I-P_X)Z} = I - A\big(A^T A\big)^+ A^T = I - U^T P_\Lambda U = U^T(I - P_\Lambda)U.$$

The proof of the first statement is now complete. Now let $\tilde{Z} = (I - P_X)Z$. Multiplying the first statement on the left and the right by $\tilde{Z}$ and $\tilde{Z}^T$, respectively, and then taking traces yields

$$\operatorname{tr}(\tilde{Z}Q^{-1}\tilde{Z}^T) \le \operatorname{tr}(\tilde{Z}(\tilde{Z}^T\tilde{Z})^+\tilde{Z}^T)\sigma_e^2 + \operatorname{tr}(\tilde{Z}U^T(I-P_\Lambda)U\tilde{Z}^T)\sigma_\bullet^2. \tag{20}$$

Since $(\tilde{Z}^T\tilde{Z})(\tilde{Z}^T\tilde{Z})^+$ is idempotent, we have

$$\begin{aligned}\operatorname{tr}(\tilde{Z}(\tilde{Z}^T\tilde{Z})^+\tilde{Z}^T) &= \operatorname{tr}(\tilde{Z}^T\tilde{Z}(\tilde{Z}^T\tilde{Z})^+) = \operatorname{rank}(\tilde{Z}^T\tilde{Z}(\tilde{Z}^T\tilde{Z})^+)\\ &= \operatorname{rank}(\tilde{Z}^T\tilde{Z}).\end{aligned}$$

Furthermore,

$$\begin{aligned}\operatorname{tr}(\tilde{Z}U^T(I-P_\Lambda)U\tilde{Z}^T) &= \operatorname{tr}(U^T(I-P_\Lambda)UZ^T(I-P_X)Z)\\ &= \operatorname{tr}(U^T(I-P_\Lambda)UU^T\Lambda U)\\ &= \operatorname{tr}(U^T(I-P_\Lambda)\Lambda U) = 0,\end{aligned}$$

where the last line follows from the fact that $(I-P_\Lambda)\Lambda = 0$. It follows from (20) that

$$\operatorname{tr}((I-P_X)ZQ^{-1}Z^T(I-P_X)) \le \operatorname{rank}(Z^T(I-P_X)Z)\sigma_e^2,$$

and the second statement has been established. Recall from Section 3 that $\lambda_{\max}$ is the largest eigenvalue of $Z^T(I-P_X)Z$, and that R_i is the $q_i \times q$ matrix of 0's and 1's such that $R_i u = u_i$. Now, fix $i \in \{1,2,\dots,r\}$ and note that

$$Q = (\sigma_e^2)^{-1}Z^T(I-P_X)Z + D^{-1} \preceq (\sigma_e^2)^{-1}\lambda_{\max}I_q + D^{-1}.$$

It follows that

$$R_i((\sigma_e^2)^{-1}\lambda_{\max}I_q + D^{-1})^{-1}R_i^T \preceq R_iQ^{-1}R_i^T,$$

and since these two matrices are both positive definite, we have

$$\begin{aligned}(R_iQ^{-1}R_i^T)^{-1} &\preceq (R_i((\sigma_e^2)^{-1}\lambda_{\max}I_q + D^{-1})^{-1}R_i^T)^{-1}\\ &= ((\sigma_e^2)^{-1}\lambda_{\max} + (\sigma_{u_i}^2)^{-1})I_{q_i},\end{aligned}$$

and this proves that the third statement is true. □

Let $\tilde{z}_i$ and y_i denote the ith column of $\tilde{Z}^T = ((I-P_X)Z)^T$ and the ith component of y, respectively. Also, define K to be

$$\sum_{i=1}^N |y_i| \sqrt{\sup_{w\in\mathbb{R}_+^{N+q}} t_i^T\left(t_it_i^T + \sum_{j\in\{1,2,\dots,N\}\setminus\{i\}} w_jt_jt_j^T + \sum_{j=N+1}^{N+q} w_jt_jt_j^T + w_iI_q\right)^{-2} t_i},$$

where, for $j = 1,2,\dots,N$, $t_j = \tilde{z}_j$, and for $j \in \{N+1,\dots,N+q\}$, the t_j are the standard orthonormal basis vectors in $\mathbb{R}^q$; that is, t_{N+l} has a one in the lth position and zeros everywhere else.

LEMMA 2. *For any $\sigma^2 \in \mathbb{R}_+^{r+1}$, we have*

$$h(\sigma^2) := \|(\sigma_e^2)^{-1} Q^{-1} Z^T (I - P_X) y\| \le K < \infty.$$

The following result from Khare and Hobert (2011) will be used in the proof of Lemma 2.

LEMMA 3. *Fix $n \in \{2, 3, \ldots\}$ and $m \in \mathbb{N}$, and let $x_1, \ldots, x_n$ be vectors in $\mathbb{R}^m$. Then*

$$C_{m,n}(x_1; x_2, \ldots, x_n) := \sup_{w \in \mathbb{R}_+^n} x_1^T \left(x_1 x_1^T + \sum_{i=2}^n w_i x_i x_i^T + w_1 I \right)^{-2} x_1$$

is finite.

PROOF OF LEMMA 2. Recall that we defined $\tilde{z}_i$ and y_i to be the ith column of $\tilde{Z}^T = ((I - P_X)Z)^T$ and the ith component of y, respectively. Now,

$$\begin{aligned} h(\sigma^2) &= \|(Z^T (I - P_X) Z + \sigma_e^2 D^{-1})^{-1} Z^T (I - P_X) y\| \\ &= \left\| \sum_{i=1}^N (\tilde{Z}^T \tilde{Z} + \sigma_e^2 D^{-1})^{-1} \tilde{z}_i y_i \right\| \\ &\le \sum_{i=1}^N \|(\tilde{Z}^T \tilde{Z} + \sigma_e^2 D^{-1})^{-1} \tilde{z}_i y_i\| \\ &= \sum_{i=1}^N \left\| \left(\sum_{j=1}^N \tilde{z}_j \tilde{z}_j^T + \sigma_e^2 D^{-1} \right)^{-1} \tilde{z}_i y_i \right\| \\ &= \sum_{i=1}^N |y_i| K_i(\sigma^2), \end{aligned}$$

where

$$K_i(\sigma^2) := \left\| \left(\tilde{z}_i \tilde{z}_i^T + \sum_{j \in \{1,2,\ldots,N\} \setminus \{i\}} \tilde{z}_j \tilde{z}_j^T + \sigma_e^2 D^{-1} \right)^{-1} \tilde{z}_i \right\|.$$

For each $i \in \{1, 2, \ldots, N\}$, define

$$\hat{K}_i = \sqrt{\sup_{w \in \mathbb{R}_+^{N+q}} t_i^T \left(t_i t_i^T + \sum_{j \in \{1,2,\ldots,N\} \setminus \{i\}} w_j t_j t_j^T + \sum_{j=N+1}^{N+q} w_j t_j t_j^T + w_i I_q \right)^{-2} t_i},$$

and notice that K can be written as $K = \sum_{i=1}^{N} |y_i| \hat{K}_i$. Therefore, it is enough to show that, for each $i \in \{1, 2, \dots, N\}$, $K_i(\sigma^2) \le \hat{K}_i < \infty$. Now,

$$\begin{aligned}
K_i^2(\sigma^2) &= \tilde{z}_i^T \Biggl(\tilde{z}_i \tilde{z}_i^T + \sum_{j \in \{1,2,\dots,N\} \setminus \{i\}} \tilde{z}_j \tilde{z}_j^T + \sigma_e^2 D^{-1}\Biggr)^{-2} \tilde{z}_i \\
&= \tilde{z}_i^T \Biggl(\tilde{z}_i \tilde{z}_i^T + \sum_{j \in \{1,2,\dots,N\} \setminus \{i\}} \tilde{z}_j \tilde{z}_j^T + \sigma_e^2 \Bigl(D^{-1} - \frac{1}{\sigma_\bullet^2} I_q\Bigr) + \frac{\sigma_e^2}{\sigma_\bullet^2} I_q\Biggr)^{-2} \tilde{z}_i \\
&\le \sup_{w \in \mathbb{R}_+^{N+q}} t_i^T \Biggl(t_i t_i^T + \sum_{j \in \{1,2,\dots,N\} \setminus \{i\}} w_j t_j t_j^T + \sum_{j=N+1}^{N+q} w_j t_j t_j^T + w_i I_q\Biggr)^{-2} t_i \\
&= \hat{K}_i^2.
\end{aligned}$$

Finally, an application of Lemma 3 shows that $\hat{K}_i^2$ is finite, and the proof is complete. □

Let $\chi_k^2(\mu)$ denote the noncentral chi-square distribution with k degrees of freedom and noncentrality parameter μ.

LEMMA 4. *If $J \sim \chi_k^2(\mu)$ and $\gamma \in (0, k/2)$, then*

$$E[J^{-\gamma}] \le \frac{2^{-\gamma}\Gamma(k/2 - \gamma)}{\Gamma(k/2)}.$$

PROOF. Since $\Gamma(x - \gamma)/\Gamma(x)$ is decreasing for $x > \gamma > 0$, we have

$$\begin{aligned}
E[J^{-\gamma}] &= \sum_{i=0}^{\infty} \frac{\mu^i e^{-\mu}}{i!} \int_{\mathbb{R}_+} x^{-\gamma} \biggl[\frac{1}{\Gamma(k/2 + i) 2^{k/2+i}} x^{k/2+i-1} e^{-x/2}\biggr] dx \\
&= 2^{-\gamma} \sum_{i=0}^{\infty} \frac{\mu^i e^{-\mu}}{i!} \frac{\Gamma(k/2 + i - \gamma)}{\Gamma(k/2 + i)} \\
&\le 2^{-\gamma} \frac{\Gamma(k/2 - \gamma)}{\Gamma(k/2)}.
\end{aligned}$$

□

A.2. An upper bound on $E[\|y - W\theta\|^2|\sigma^2]$. We remind the reader that $\theta = (\beta^T \ \ u^T)^T$, $W = (X \ \ Z)$, and that $\pi(\theta|\sigma^2, y)$ is a multivariate normal density with mean m and covariance matrix V. Thus,

$$E[\|y - W\theta\|^2|\sigma^2] = \mathrm{tr}(W V W^T) + \|y - Wm\|^2, \tag{21}$$

and we have

$$
\begin{aligned}
&\operatorname{tr}(WVW^T) \\
&\quad = \sigma_e^2 \operatorname{tr}(P_X) + \operatorname{tr}(P_X Z Q^{-1} Z^T P_X) - 2\operatorname{tr}(Z Q^{-1} Z^T P_X) + \operatorname{tr}(Z Q^{-1} Z^T) \\
&\quad = p\sigma_e^2 - \operatorname{tr}(Z Q^{-1} Z^T P_X) + \operatorname{tr}(Z Q^{-1} Z^T) \\
&\quad = p\sigma_e^2 + \operatorname{tr}(Z Q^{-1} Z^T (I - P_X)) \qquad (22)\\
&\quad = p\sigma_e^2 + \operatorname{tr}((I - P_X) Z Q^{-1} Z^T (I - P_X)) \\
&\quad \le p\sigma_e^2 + \operatorname{rank}(Z^T (I - P_X) Z)\sigma_e^2 \\
&\quad = (p + t)\sigma_e^2,
\end{aligned}
$$

where the inequality is an application of Lemma 1. Finally, a simple calculation shows that

$$y - Wm = (I - P_X)[I - (\sigma_e^2)^{-1} Z Q^{-1} Z^T (I - P_X)]y.$$

Hence,

$$
\begin{aligned}
\|y - Wm\| &= \|(I - P_X)y - (\sigma_e^2)^{-1}(I - P_X) Z Q^{-1} Z^T (I - P_X) y\| \\
&\le \|(I - P_X)y\| + \|(\sigma_e^2)^{-1}(I - P_X) Z Q^{-1} Z^T (I - P_X) y\| \qquad (23)\\
&\le \|(I - P_X)y\| + \|(I - P_X)Z\| \|(\sigma_e^2)^{-1} Q^{-1} Z^T (I - P_X) y\| \\
&\le \|(I - P_X)y\| + \|(I - P_X)Z\| K,
\end{aligned}
$$

where $\|\cdot\|$ denotes the Frobenius norm and the last inequality uses Lemma 2. Finally, combining (21), (22) and (23) yields

$$E[\|y - W\theta\|^2 | \sigma^2] \le (p + t)\sigma_e^2 + (\|(I - P_X)y\| + \|(I - P_X)Z\| K)^2.$$

A.3. An upper bound on $E[\|u_i\|^2|\sigma^2]$. Note that

$$E[\|u_i\|^2|\sigma^2] = E[\|R_i u\|^2|\sigma^2] = \operatorname{tr}(R_i Q^{-1} R_i^T) + \|E[R_i u|\sigma^2]\|^2. \qquad (24)$$

By Lemma 1, we have

$$
\begin{aligned}
\operatorname{tr}(R_i Q^{-1} R_i^T) &\le \operatorname{tr}(R_i (Z^T (I - P_X) Z)^+ R_i^T)\sigma_e^2 \\
&\qquad + \operatorname{tr}(R_i (I - P_{Z^T(I-P_X)Z}) R_i^T) \sum_{j=1}^r \sigma_{u_j}^2 \qquad (25)\\
&= \xi_i \sigma_e^2 + \zeta_i \sum_{j=1}^r \sigma_{u_j}^2.
\end{aligned}
$$

Now, by Lemma 2,

$$\|E[R_i u|\sigma^2]\| \le \|R_i\| \|E[u|\sigma^2]\| = \|R_i\| h(\sigma^2) \le \|R_i\| K. \qquad (26)$$

Combining (24), (25) and (26) yields

$$E\big[\|u_i\|^2|\sigma^2\big] \le \xi_i \sigma_e^2 + \zeta_i \sum_{j=1}^{r} \sigma_{u_j}^2 + \big(\|R_i\| K\big)^2.$$

A.4. An upper bound on $E[(\|u_i\|^2)^{-c}|\sigma^2]$. Fix $i \in \{1, 2, \dots, r\}$. Given σ^2, $(R_i Q^{-1} R_i^T)^{-1/2} u_i$ has a multivariate normal distribution with identity covariance matrix. It follows that, conditional on σ^2, the distribution of $u_i^T (R_i Q^{-1} R_i^T)^{-1} u_i$ is $\chi^2_{q_i}(w)$. It follows from Lemma 4 that, as long as $c \in (0, 1/2)$, we have

$$E\big[\big[u_i^T (R_i Q^{-1} R_i^T)^{-1} u_i\big]^{-c} |\sigma^2\big] \le 2^{-c} \frac{\Gamma(q_i/2 - c)}{\Gamma(q_i/2)}.$$

Now, by Lemma 1,

$$\begin{aligned}
&E\big[\big(\|u_i\|^2\big)^{-c}|\sigma^2\big] \\
&\quad = \big((\sigma_e^2)^{-1}\lambda_{\max} + (\sigma_{u_i}^2)^{-1}\big)^c E\big[\big[u_i^T \big((\sigma_e^2)^{-1}\lambda_{\max} + (\sigma_{u_i}^2)^{-1}\big) I_{q_i} u_i\big]^{-c} |\sigma^2\big] \\
&\quad \le \big((\sigma_e^2)^{-1}\lambda_{\max} + (\sigma_{u_i}^2)^{-1}\big)^c E\big[\big[u_i^T (R_i Q^{-1} R_i^T)^{-1} u_i\big]^{-c} |\sigma^2\big] \\
&\quad \le \big((\sigma_e^2)^{-1}\lambda_{\max} + (\sigma_{u_i}^2)^{-1}\big)^c 2^{-c} \frac{\Gamma(q_i/2 - c)}{\Gamma(q_i/2)} \\
&\quad \le 2^{-c} \frac{\Gamma(q_i/2 - c)}{\Gamma(q_i/2)} \big[\lambda_{\max}^c (\sigma_e^2)^{-c} + (\sigma_{u_i}^2)^{-c}\big].
\end{aligned}$$

Acknowledgments. The authors thank three anonymous reviewers for helpful comments and suggestions.

REFERENCES

BEDNORZ, W. and ŁATUSZYŃSKI, K. (2007). A few remarks on "Fixed-width output analysis for Markov chain Monte Carlo" by Jones et al. [MR2279478]. *J. Amer. Statist. Assoc.* **102** 1485–1486. MR2412582

DANIELS, M. J. (1999). A prior for the variance in hierarchical models. *Canad. J. Statist.* **27** 567–578. MR1745822

DIACONIS, P., KHARE, K. and SALOFF-COSTE, L. (2008). Gibbs sampling, exponential families and orthogonal polynomials (with discussion). *Statist. Sci.* **23** 151–178. MR2446500

FLEGAL, J. M., HARAN, M. and JONES, G. L. (2008). Markov chain Monte Carlo: Can we trust the third significant figure? *Statist. Sci.* **23** 250–260. MR2516823

FLEGAL, J. M. and JONES, G. L. (2010). Batch means and spectral variance estimators in Markov chain Monte Carlo. *Ann. Statist.* **38** 1034–1070. MR2604704

GELMAN, A. (2006). Prior distributions for variance parameters in hierarchical models. *Bayesian Anal.* **1** 515–534.

JOHNSON, A. A. and JONES, G. L. (2010). Gibbs sampling for a Bayesian hierarchical general linear model. *Electron. J. Stat.* **4** 313–333. MR2645487

JONES, G. L., HARAN, M., CAFFO, B. S. and NEATH, R. (2006). Fixed-width output analysis for Markov chain Monte Carlo. *J. Amer. Statist. Assoc.* **101** 1537–1547. MR2279478

KHARE, K. and HOBERT, J. P. (2011). A spectral analytic comparison of trace-class data augmentation algorithms and their sandwich variants. *Ann. Statist.* **39** 2585–2606. MR2906879

LIU, J. S., WONG, W. H. and KONG, A. (1994). Covariance structure of the Gibbs sampler with applications to the comparisons of estimators and augmentation schemes. *Biometrika* **81** 27–40. MR1279653

MEYN, S. P. and TWEEDIE, R. L. (1993). *Markov Chains and Stochastic Stability*. Springer, London. MR1287609

PAPASPILIOPOULOS, O. and ROBERTS, G. (2008). Stability of the Gibbs sampler for Bayesian hierarchical models. *Ann. Statist.* **36** 95–117. MR2387965

ROBERTS, G. O. and ROSENTHAL, J. S. (1998). Markov-chain Monte Carlo: Some practical implications of theoretical results (with discussion). *Canad. J. Statist.* **26** 5–31. MR1624414

ROBERTS, G. O. and ROSENTHAL, J. S. (2001). Markov chains and de-initializing processes. *Scand. J. Stat.* **28** 489–504. MR1858413

ROBERTS, G. O. and ROSENTHAL, J. S. (2004). General state space Markov chains and MCMC algorithms. *Probab. Surv.* **1** 20–71. MR2095565

ROMÁN, J. C. (2012). Convergence analysis of block Gibbs samplers for Bayesian general linear mixed models. Ph.D. thesis, Dept. Statistics, Univ. Florida, Gainesville, FL.

SEARLE, S. R., CASELLA, G. and MCCULLOCH, C. E. (1992). *Variance Components*. Wiley, New York. MR1190470

SUN, D., TSUTAKAWA, R. K. and HE, Z. (2001). Propriety of posteriors with improper priors in hierarchical linear mixed models. *Statist. Sinica* **11** 77–95. MR1820002

TAN, A. and HOBERT, J. P. (2009). Block Gibbs sampling for Bayesian random effects models with improper priors: Convergence and regeneration. *J. Comput. Graph. Statist.* **18** 861–878. MR2598033

DEPARTMENT OF MATHEMATICS
VANDERBILT UNIVERSITY
NASHVILLE, TENNESSEE 37240
USA
E-MAIL: jc.roman@vanderbilt.edu

DEPARTMENT OF STATISTICS
UNIVERSITY OF FLORIDA
GAINESVILLE, FLORIDA 32611
USA
E-MAIL: jhobert@stat.ufl.edu