Electronic Journal of Statistics
Vol. 12 (2018) 4412–4439
ISSN: 1935-7524
https://doi.org/10.1214/18-EJS1506

Convergence analysis of the block Gibbs sampler for Bayesian probit linear mixed models with improper priors

Xin Wang

Department of Statistics, Miami Uiversity
Oxford, OH, 45056
e-mail: wangx172@miamioh.edu

and

Vivekananda Roy

Department of Statistics, Iowa State University
Ames, IA, 50011
e-mail: vroy@iastate.edu

Abstract: In this article, we consider Markov chain Monte Carlo (MCMC) algorithms for exploring the intractable posterior density associated with Bayesian probit linear mixed models under improper priors on the regression coefficients and variance components. In particular, we construct a two-block Gibbs sampler using the data augmentation (DA) techniques. Furthermore, we prove geometric ergodicity of the Gibbs sampler, which is the foundation for building central limit theorems for MCMC based estimators and subsequent inferences. The conditions for geometric convergence are similar to those guaranteeing posterior propriety. We also provide conditions for the propriety of posterior distributions with a general link function when the design matrices take commonly observed forms. In general, the Haar parameter expansion for DA (PX-DA) algorithm is an improvement of the DA algorithm and it has been shown that it is theoretically at least as good as the DA algorithm. Here we construct a Haar PX-DA algorithm, which has essentially the same computational cost as the two-block Gibbs sampler.

MSC 2010 subject classifications: Primary 60J05; secondary 62F15.
Keywords and phrases: Data augmentation, drift condition, geometric ergodicity, GLMM, Haar PX-DA algorithm, Markov chains, posterior propriety.

Received November 2017.

1. Introduction

Generalized linear mixed models (GLMMs) are generalized linear models with random terms in the linear predictor. The random effects in the GLMM can accommodate for overdispersion often present in non-Gaussian data, and dependence among correlated observations arising from longitudinal or repeated measures studies. GLMM is one of the most frequently used statistical models. Here, we consider a popular Bayesian GLMM for binary data, namely, the probit linear mixed model.

Let $(Y_1, Y_2, \ldots, Y_n)$ denote the vector of Bernoulli random variables. Let $\boldsymbol{x}_i$ and $\boldsymbol{z}_i$ be the $p \times 1$ and $q \times 1$ known covariates and random effect design vectors respectively associated with the ith observation for $i = 1, \ldots, n$. Let $\boldsymbol{\beta} \in \mathbb{R}^p$ be the unknown vector of regression coefficients and $\boldsymbol{u} \in \mathbb{R}^q$ be the random effects vector. A GLMM can be built (McCulloch, Searle and Neuhaus, 2011; Breslow and Clayton, 1993) with a link function that connects the expectation of Y_i with $\boldsymbol{x}_i$ and $\boldsymbol{z}_i$. One of the very popular link functions is the probit link function, Φ^{-1}, resulting in

$$P(Y_i = 1) = \Phi(\boldsymbol{x}_i^T \boldsymbol{\beta} + \boldsymbol{z}_i^T \boldsymbol{u}), \tag{1}$$

where $\Phi(\cdot)$ is the cumulative distribution function of the standard normal random variable. Assume that we have r random effects with $\boldsymbol{u} = (\boldsymbol{u}_1^T, \ldots, \boldsymbol{u}_r^T)^T$, where $\boldsymbol{u}_j$ is a $q_j \times 1$ vector with $q_j > 0$, $q_1 + \cdots + q_r = q$, and $\boldsymbol{u}_j \overset{\text{ind}}{\sim} N(0, \boldsymbol{I}_{q_j} 1/\tau_j)$, where $\tau_j \in \mathbb{R}_+ \equiv (0, \infty)$ is the precision parameter associated with $\boldsymbol{u}_j$ for $j = 1, \ldots, r$. Let $\boldsymbol{\tau} = (\tau_1, \ldots, \tau_r)$, thus, the data model for the probit GLMM is

$$\begin{aligned} Y_i | \boldsymbol{\beta}, \boldsymbol{u}, \boldsymbol{\tau} &\overset{\text{ind}}{\sim} \text{Bern}(\alpha_i) \text{ for } i = 1, \ldots, n \text{ with} \\ \alpha_i &= \Phi(\boldsymbol{x}_i^T \boldsymbol{\beta} + \boldsymbol{z}_i^T \boldsymbol{u}) \text{ for } i = 1, \ldots, n, \\ \boldsymbol{u}_j | \boldsymbol{\beta}, \boldsymbol{\tau} &\overset{\text{ind}}{\sim} N\Big(0, \frac{1}{\tau_j} \boldsymbol{I}_{q_j}\Big), \ j = 1, \ldots, r. \end{aligned} \tag{2}$$

Let $\boldsymbol{y} = (y_1, y_2, \ldots, y_n)^T$ be the observed Bernoulli response variables. Note that, the likelihood function for $(\boldsymbol{\beta}, \boldsymbol{\tau})$ is

$$\begin{aligned} L(\boldsymbol{\beta}, \boldsymbol{\tau} | \boldsymbol{y}) &= \int_{\mathbb{R}^q} \prod_{i=1}^n \left[\Phi(\boldsymbol{x}_i^T \boldsymbol{\beta} + \boldsymbol{z}_i^T \boldsymbol{u})\right]^{y_i} \left[1 - \Phi(\boldsymbol{x}_i^T \boldsymbol{\beta} + \boldsymbol{z}_i^T \boldsymbol{u})\right]^{1 - y_i} \\ &\times \ \phi_q(\boldsymbol{u}; \boldsymbol{0}, \boldsymbol{D}(\boldsymbol{\tau})^{-1}) d\boldsymbol{u}, \end{aligned} \tag{3}$$

which is not available in closed form. Here, $\phi_q(s; a, B)$ denotes the probability density function of the $q-$dimensional normal distribution with mean vector a, covariance matrix B and evaluated at s, and $\boldsymbol{D}(\boldsymbol{\tau}) = \oplus_{j=1}^r \tau_j \boldsymbol{I}_{q_j}$.

In Bayesian framework, one needs to specify the prior distributions of $\boldsymbol{\beta}$ and $\boldsymbol{\tau}$. Assume $\boldsymbol{\beta}$ and $\boldsymbol{\tau}$ are apriori independent. Let $\pi(\boldsymbol{\beta})$ and $\pi(\boldsymbol{\tau})$ be the prior densities of $\boldsymbol{\beta}$ and $\boldsymbol{\tau}$ respectively. Thus, the joint posterior density of $(\boldsymbol{\beta}, \boldsymbol{\tau})$ is

$$\pi(\boldsymbol{\beta}, \boldsymbol{\tau} | \boldsymbol{y}) = \frac{1}{c(\boldsymbol{y})} L(\boldsymbol{\beta}, \boldsymbol{\tau} | \boldsymbol{y}) \pi(\boldsymbol{\beta}) \pi(\boldsymbol{\tau}), \tag{4}$$

where

$$c(\boldsymbol{y}) = \int_{\mathbb{R}_+^r} \int_{\mathbb{R}^p} L(\boldsymbol{\beta}, \boldsymbol{\tau} | \boldsymbol{y}) \pi(\boldsymbol{\beta}) \pi(\boldsymbol{\tau}) d\boldsymbol{\beta} d\boldsymbol{\tau},$$

is the marginal density of $\boldsymbol{y}$. Since the likelihood function $L(\boldsymbol{\beta}, \boldsymbol{\tau} | \boldsymbol{y})$ is not available in closed form, the posterior density is intractable for any choice of the prior distributions of $\boldsymbol{\beta}$ and $\boldsymbol{\tau}$. In this article, we consider an improper flat prior for $\boldsymbol{\beta}$, that is, $\pi(\boldsymbol{\beta}) \propto 1$ and τ_j's, $j = 1, \ldots, r$, are apriori independent with

$$\pi(\tau_j) \propto e^{-b_j \tau_j} \tau_j^{a_j - 1}, \tag{5}$$

which can be proper or improper. In section 2, we discuss conditions under which the posterior density (4) is proper, that is $c(\boldsymbol{y}) < \infty$. Generally, Markov chain Monte Carlo (MCMC) algorithms are used for exploring the posterior density (4).

Even in the absence of random effects, for the probit regression model, the posterior distribution of $\boldsymbol{\beta}$ is difficult to sample from (Roy and Hobert, 2007). Albert and Chib's (1993) MCMC algorithm for sampling from the posterior distribution associated with the probit regression model is the most widely used data augmentation (DA) algorithm. The DA technique used in Albert and Chib (1993) can also be applied to the probit linear mixed model. Following Albert and Chib (1993), let $v_i \in \mathbb{R}$ be the continuous latent variable corresponding to the ith binary observation Y_i, such that $Y_i = I(v_i > 0)$, where $v_i|\boldsymbol{\beta}, \boldsymbol{u}, \boldsymbol{\tau} \overset{\text{ind}}{\sim} N(\boldsymbol{x}_i^T\boldsymbol{\beta} + \boldsymbol{z}_i^T\boldsymbol{u},\, 1)$ for $i = 1, \ldots, n$. Then

$$P(Y_i = 1) = P(v_i > 0) = \Phi(\boldsymbol{x}_i^T\boldsymbol{\beta} + \boldsymbol{z}_i^T\boldsymbol{u}), \tag{6}$$

that is, $Y_i|\boldsymbol{\beta}, \boldsymbol{u}, \boldsymbol{\tau} \overset{\text{ind}}{\sim} \text{Bern}(\alpha_i)$ as in (2). Note that $\boldsymbol{v}|\boldsymbol{\beta}, \boldsymbol{u}, \boldsymbol{\tau} \sim N(\boldsymbol{X\beta}+\boldsymbol{Zu}, \boldsymbol{I}_n)$, where $\boldsymbol{v} = (v_1, \ldots, v_n)^T$, $\boldsymbol{X}_{n\times p} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)^T$ and $\boldsymbol{Z}_{n\times q} = (\boldsymbol{z}_1, \ldots, \boldsymbol{z}_n)^T$.

Using the latent variables $\boldsymbol{v}$, we can introduce a joint density $\pi(\boldsymbol{\beta}, \boldsymbol{u}, \boldsymbol{v}, \boldsymbol{\tau}|\boldsymbol{y})$ (see section 3 for details) such that

$$\int_{\mathbb{R}^q}\int_{\mathbb{R}^n} \pi(\boldsymbol{\beta}, \boldsymbol{u}, \boldsymbol{v}, \boldsymbol{\tau}|\boldsymbol{y}) d\boldsymbol{v} d\boldsymbol{u} = \pi(\boldsymbol{\beta}, \boldsymbol{\tau}|\boldsymbol{y}), \tag{7}$$

where $\pi(\boldsymbol{\beta}, \boldsymbol{\tau}|\boldsymbol{y})$ is the posterior density defined in (4). If all the full conditionals of the joint density $\pi(\boldsymbol{\beta}, \boldsymbol{u}, \boldsymbol{v}, \boldsymbol{\tau}|\boldsymbol{y})$ are easy to sample from, then a Gibbs sampler can be run and it can be used to make inferences on the posterior density (4). Indeed this full Gibbs sampler is traditionally used in the analysis of Bayesian probit linear mixed models (Baragatti, 2011). In this article, instead of using full conditional distributions, we construct a two-block Gibbs sampler with $\boldsymbol{\eta} \equiv (\boldsymbol{\beta}^T, \boldsymbol{u}^T)^T$ as one block and $(\boldsymbol{v}^T, \boldsymbol{\tau}^T)^T$ as the other block — which is our first contribution. In general, block Gibbs samplers are known to be better than the Gibbs samplers based on full conditional distributions in terms of having smaller operator norm (Liu, Wong and Kong, 1994).

The above mentioned block Gibbs sampler has an everywhere strictly positive Markov transition density, implying that the underlying Markov chain is Harris ergodic (Asmussen and Glynn, 2011; Meyn and Tweedie, 1993). Thus, the time average estimators based on the block Gibbs sampler can be used to consistently estimate the (posterior) means with respect to the joint density $\pi(\boldsymbol{\beta}, \boldsymbol{u}, \boldsymbol{v}, \boldsymbol{\tau}|\boldsymbol{y})$. In practice, it is crucial to know whether the Monte Carlo errors associated with these estimates are sufficiently small. However, in order to provide valid standard errors, we need to establish a central limit theorem (CLT) for the time average estimators. Unlike for the ordinary Monte Carlo methods based on iid samples, mere existence of the finite second moment does not guarantee a CLT for MCMC estimators. One standard method of establishing a CLT for MCMC estimators is to prove that the underlying Markov chain is *geometrically ergodic* (Jones and

Hobert, 2001). Geometric ergodicity is also needed for consistently estimating the asymptotic variance in the Markov chain CLT (Flegal and Jones, 2010). Roy and Hobert (2007) and Chakraborty and Khare (2017) proved geometric ergodicity of Albert and Chib's (1993) DA algorithm for the Bayesian probit regression model under improper and proper priors on the regression coefficients. For linear models, Jones and Hobert (2004) and Tan and Hobert (2009) analyzed the Gibbs sampler for one-way random effects models under proper priors and improper priors respectively. Johnson and Jones (2010) analyzed the block Gibbs sampler for Bayesian linear mixed models under the assumption $\boldsymbol{X}^T\boldsymbol{Z} = \boldsymbol{0}$. Román and Hobert (2012) and Román and Hobert (2015) established geometric rate of convergence of the Gibbs samplers for Bayesian linear mixed models under improper and proper priors without the assumption of $\boldsymbol{X}^T\boldsymbol{Z} = \boldsymbol{0}$. Our second contribution, in this paper, is establishing geometric convergence rates for the block Gibbs sampler for Bayesian probit linear mixed models under improper priors.

DA algorithms are known to suffer from slow convergence (Meng and Van Dyk, 1999; Van Dyk and Meng, 2001). Liu and Wu (1999) proposed the *parameter expansion for data augmentation* (PX-DA) algorithm, which can converge faster than the DA algorithm without much extra computational effort (Van Dyk and Meng, 2001; Roy, 2014). Hobert and Marchev (2008) proved that the Haar PX-DA algorithm, that is based on a Haar measure, is better than any other PX-DA algorithm and the original DA algorithm in both the efficiency ordering and the operator norm ordering. For the probit regression model, Roy and Hobert (2007), through an example, showed that the Haar PX-DA algorithm can lead to huge gains in efficiency over the DA algorithm of Albert and Chib (1993). Our third contribution is to construct a Haar PX-DA algorithm improving the block Gibbs sampler mentioned before. Since geometric ergodicity of the Haar PX-DA algorithm follows from geometric ergodicity of the DA algorithm (Hobert and Marchev, 2008), we have CLTs for the Haar PX-DA algorithm based estimators as well.

The article is organized as follows. In section 2, we establish conditions for propriety of the posterior distribution (4) under improper priors, when $\boldsymbol{X}$ and $\boldsymbol{Z}$ take commonly observed forms. The results in section 2 hold for a general link function, not necessarily the probit link. In section 3, we construct the two-block Gibbs sampler for the Bayesian probit linear mixed model under improper priors. In section 4, we prove geometric ergodicity of the underlying Markov chain. In section 5, we present a corresponding Haar PX-DA algorithm. Section 6 contains some conclusions and discussions. Finally, the proofs of posterior propriety and geometric convergence of the Gibbs sampler appear in the appendices.

2. Propriety of posterior distributions

In this section, we discuss conditions under which the posterior density (4) is proper. The results in this section hold for GLMMs with a general link function. Let $F(\cdot)$ be a cumulative distribution function, and consider the link function

$F^{-1}(\cdot)$. Thus instead of the probit linear mixed model in (1), in this section we consider a GLMM with

$$P(Y_i = 1) = F(\boldsymbol{x}_i^T \boldsymbol{\beta} + \boldsymbol{z}_i^T \boldsymbol{u}). \tag{8}$$

Posterior propriety for Bayesian GLMMs under improper priors has been discussed in Chen, Shao and Xu (2002). We will first describe Chen, Shao and Xu's (2002) conditions. Then we will show, through examples, that these conditions often do not hold in practice. Finally, our conditions for posterior propriety will be presented.

Let $c_i = 1$ if $y_i = 0$ and $c_i = -1$ if $y_i = 1$ for $i = 1, \ldots, n$. Suppose $\boldsymbol{W}^*_{n\times(p+q)}$ is a matrix whose ith row is $c_i(\boldsymbol{x}_i^T, \boldsymbol{z}_i^T)$. In the special case when $b_j = 0$, that is, when τ_j has the power prior $\pi(\tau_j) \propto \tau_j^{a_j - 1}$ for $j = 1, \ldots, r$, a straightforward extension of Chen, Shao and Xu's (2002) Theorem 4.2 shows that the corresponding posterior distribution is proper if the following conditions hold:

(A1) $\boldsymbol{W} = (\boldsymbol{X}, \boldsymbol{Z})$ is a full rank matrix;
(A2) There exists an $n \times 1$ positive vector $\boldsymbol{e} > 0$ such that $\boldsymbol{e}^T \boldsymbol{W}^* = 0$;
(A3) $2a_j + q_j > 0$ for $j = 1, 2, \ldots, r$;
(A4) $a_j < 0$ for $j = 1, \ldots, r$;
(A5) $E|\delta|^{p-2\sum_{j=1}^r a_j} < \infty$, where $\delta \sim F$.

Roy and Hobert (2007) provided a simple method for checking the condition A2 using publicly available softwares.

The condition A1 assumes that $\boldsymbol{W}$ is a full rank matrix. Unfortunately, when $\boldsymbol{Z}$ is a design matrix with elements 1's and 0's, which is pretty common in practice, this assumption may not hold. For example, we consider the following important generalized two-way random effects model

$$F^{-1}(P(Y_{ij} = 1)) = \beta + \alpha_i + \gamma_j, \tag{9}$$

for $i = 1, 2, \ldots, n_1$, $j = 1, 2, \ldots, n_2$. Here, the α_i's are i.i.d $N(0, 1/\tau_1)$, and the γ_j's are i.i.d $N(0, 1/\tau_2)$. There are total $n = n_1 \times n_2$ observations, and we order them as $\boldsymbol{Y} = (Y_{11}, \ldots, Y_{1n_2}, \ldots, Y_{n_1 1}, \ldots, Y_{n_1 n_2})$. In this example, $p = 1$, and $\boldsymbol{X} = \mathbf{1}_n$ is an $n \times 1$ column vector of ones. Also, there are $r = 2$ random effects with $q_1 = n_1$, $q_2 = n_2$, $q = q_1 + q_2 = n_1 + n_2$, and $\boldsymbol{Z} = (\boldsymbol{Z}_1, \boldsymbol{Z}_2)$, where $\boldsymbol{Z}_1 = \boldsymbol{I}_{n_1} \otimes \mathbf{1}_{n_2}$ and $\boldsymbol{Z}_2 = \mathbf{1}_{n_1} \otimes \boldsymbol{I}_{n_2}$ with $\otimes$ denoting the Kronecker product. It can be checked that the rank of $\boldsymbol{W} = (\boldsymbol{X}, \boldsymbol{Z})$ is $n_1 + n_2 - 1$. Thus $\boldsymbol{W}$ is not a full rank matrix.

We now provide Theorem 1 showing the posterior propriety without the assumption A1. We also consider the more general prior $\pi(\tau_j)$ given in (5), that is, b_j may not be zero. We use certain transformations of the regression parameters $\boldsymbol{\beta}$ and random effects $\boldsymbol{u}$ to circumvent the problem with non-full rank matrix $\boldsymbol{W}$. Assume that the first column of $\boldsymbol{X}$ is a vector of 1's corresponding to an intercept term β_0 in $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_{p-1})^T$. Let $\boldsymbol{Z} = (\boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_r)$, where $\boldsymbol{Z}_j$ is an $n \times q_j$ matrix such that the (ik)th element is 1 if the observation i is observed at the kth level of the random effect $\boldsymbol{u}_j = \left(u_{j1}, \ldots, u_{jq_j}\right)^T$, 0 otherwise, for $i = 1, \ldots, n$, $k = 1, \ldots, q_j$ and $j = 1, \ldots, r$. Consider the following transformations,

$$\mu_0 = \beta_0 + \sum_{j=1}^{r} u_{j1}, \tag{10}$$

$$d_{jk} = u_{j,k+1} - u_{j1}, \text{for } k = 1, \ldots, q_j - 1, j = 1, \ldots, r. \tag{11}$$

Thus μ_0 is the sum of the intercept term and the first level effect of all r random effects. Also the (transformed) random effects d_{jk}'s denote the differences of the random effects compared to the first level effect.

Let $\tilde{\boldsymbol{\eta}} = (\mu_0, \beta_{1,}, \ldots, \beta_{p-1}, d_{11}, \ldots, d_{1,q_1-1}, \ldots d_{r1}, \ldots, d_{r,q_r-1})^T$. Define $\tilde{\boldsymbol{Z}} = (\tilde{\boldsymbol{Z}}_1, \ldots, \tilde{\boldsymbol{Z}}_r)$, where the $n \times (q_j - 1)$ matrix $\tilde{\boldsymbol{Z}}_j$ is $\boldsymbol{Z}_j$ without its first column. Thus, the vector $\boldsymbol{W}\boldsymbol{\eta}$ is the same as the vector $\tilde{\boldsymbol{W}}\tilde{\boldsymbol{\eta}}$, where $\tilde{\boldsymbol{W}} = \left(\boldsymbol{X}, \tilde{\boldsymbol{Z}}\right)$ with ith row $\tilde{\boldsymbol{w}}_i^T$. Let $\tilde{\boldsymbol{W}}^*$ be a matrix whose ith row is $\tilde{\boldsymbol{w}}_i^{*T} = c_i \tilde{\boldsymbol{w}}_i^T = c_i(\boldsymbol{x}_i^T, \tilde{\boldsymbol{z}}_i^T)$, where $\tilde{\boldsymbol{z}}_i^T$ is the ith row of $\tilde{\boldsymbol{Z}}$.

For the example (9), the transformed parameters μ_0 and d_{jk}'s become

$$\begin{aligned}
\mu_0 &= \beta + \alpha_1 + \gamma_1, \\
d_{1k} &= \alpha_{k+1} - \alpha_1 \text{ for } k = 1, \ldots, n_1 - 1, \\
d_{2k} &= \gamma_{k+1} - \gamma_1 \text{ for } k = 1, \ldots, n_2 - 1.
\end{aligned}$$

Thus in this example, we have $\tilde{\boldsymbol{\eta}} = (\mu_0, d_{11}, \ldots, d_{1,n_1-1}, d_{21}, \ldots, d_{2,n_2-1})^T$. Also note that $\tilde{\boldsymbol{W}}$ is a full rank matrix in this example, although $\boldsymbol{W}$ is not.

Theorem 1. *Assume the following conditions hold,*

(B1) $a_j < b_j = 0$, $q_j \geq 2$ *or* $b_j > 0$ *for* $j = 1, \ldots, r$;
(B2) $2a_j + q_j - 1 > 0$ *for* $j = 1, \ldots, r$;
(B3) $\tilde{\boldsymbol{W}}$ *is a full rank matrix;*
(B4) There exists an $n \times 1$ *positive vector* $\boldsymbol{e} > 0$ *such that* $\mathbf{e}^T \tilde{\boldsymbol{W}}^* = 0$.
(B5) $E|\delta|^{p+t} < \infty$, *where* $t = \sum_{j=1}^{r}[-2a_j I(b_j = 0) + (q_j - 1)I(b_j > 0)]$, *and* $\delta \sim F$.

Then the joint posterior density (4) *corresponding to the GLMM* (8) *is proper, i.e,*

$$\int_{\mathbb{R}_+^r} \int_{\mathbb{R}^q} \int_{\mathbb{R}^p} \prod_{i=1}^{n} \left[F\left(\boldsymbol{x}_i^T \boldsymbol{\beta} + \boldsymbol{z}_i^T \boldsymbol{u}\right)\right]^{y_i} \left[1 - F\left(\boldsymbol{x}_i^T \boldsymbol{\beta} + \boldsymbol{z}_i^T \boldsymbol{u}\right)\right]^{1-y_i}$$
$$\cdot \prod_{j=1}^{r} \tau_j^{\frac{q_j}{2} + a_j - 1} \exp\left[-\tau_j \left(b_j + \frac{1}{2}\boldsymbol{u}_j^T \boldsymbol{u}_j\right)\right] d\boldsymbol{\beta} d\boldsymbol{u} d\boldsymbol{\tau} < \infty. \tag{12}$$

A proof of Theorem 1 is given in Appendix A.

Remark 1. *When the probit link is considered, that is* $F(\cdot) = \Phi(\cdot)$, *the moment condition B5 holds automatically. Thus, for probit linear mixed models, the posterior density* (4) *is proper under* $B1 - B4$.

3. A two-block Gibbs sampler

We begin with deriving the joint density $\pi(\boldsymbol{\beta}, \boldsymbol{u}, \boldsymbol{v}, \boldsymbol{\tau}|\boldsymbol{y})$ mentioned in the introduction. Define the joint posterior density (up to a normalizing constant) of $\boldsymbol{\beta}, \boldsymbol{u}, \boldsymbol{v}, \boldsymbol{\tau}$, if it exists, as

$$\begin{aligned}\pi(\boldsymbol{\beta}, \boldsymbol{u}, \boldsymbol{v}, \boldsymbol{\tau}|\boldsymbol{y}) \propto & \prod_{i=1}^{n} \exp\left\{-\frac{1}{2}\left(v_i - \boldsymbol{x}_i^T\boldsymbol{\beta} - \boldsymbol{z}_i^T\boldsymbol{u}\right)^2\right\} \\ & \times \prod_{i=1}^{n} \left[1_{(0,\infty)}(v_i)\right]^{y_i} \left[1_{(-\infty,0]}(v_i)\right]^{1-y_i} \\ & \times \prod_{j=1}^{r} \tau_j^{\frac{q_j}{2}+a_j-1} \exp\left\{-\tau_j\left(b_j + \frac{\boldsymbol{u_j}^T\boldsymbol{u_j}}{2}\right)\right\}. \end{aligned} \tag{13}$$

From (3) and (6) it follows that (7) holds. In section 2, we discussed conditions under which the posterior density $\pi(\boldsymbol{\beta}, \boldsymbol{\tau}|\boldsymbol{y})$ given in (4) and hence the joint posterior density (13) is proper. Note that, these posterior densities are proper if and only if $c(\boldsymbol{y}) < \infty$.

Standard calculations show that the conditional density of $\boldsymbol{\eta}$ is

$$\pi(\boldsymbol{\eta}|\boldsymbol{v}, \boldsymbol{\tau}, \boldsymbol{y}) \propto \exp\left[-\frac{1}{2}(\boldsymbol{v} - \boldsymbol{W}\boldsymbol{\eta})^T(\boldsymbol{v} - \boldsymbol{W}\boldsymbol{\eta})\right] \cdot \exp\left[-\frac{1}{2}\boldsymbol{u}^T\boldsymbol{D}(\boldsymbol{\tau})\boldsymbol{u}\right], \tag{14}$$

Thus,

$$\boldsymbol{\eta}|\boldsymbol{v}, \boldsymbol{\tau}, \boldsymbol{y} \sim N_{p+q}\left(\boldsymbol{\Sigma}^{-1}\boldsymbol{W}^T\boldsymbol{v}, \boldsymbol{\Sigma}^{-1}\right), \tag{15}$$

where

$$\boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{X}^T\boldsymbol{X} & \boldsymbol{X}^T\boldsymbol{Z} \\ \boldsymbol{Z}^T\boldsymbol{X} & \boldsymbol{Z}^T\boldsymbol{Z} + \boldsymbol{D}(\boldsymbol{\tau}) \end{pmatrix}. \tag{16}$$

Similarly, the conditional density of $(\boldsymbol{v}, \boldsymbol{\tau})$ is

$$\begin{aligned}\pi(\boldsymbol{v}, \boldsymbol{\tau}|\boldsymbol{\eta}, \boldsymbol{y}) \quad \propto \quad & \prod_{i=1}^{n} \phi\left(v_i - \boldsymbol{w}_i^T\boldsymbol{\eta}; 0, 1\right)\left[1_{(0,\infty)}(v_i)\right]^{y_i}\left[1_{(-\infty,0]}(v_i)\right]^{1-y_i} \\ & \times \prod_{j=1}^{r} \tau_j^{\frac{q_j}{2}+a_j-1} \exp\left[-\tau_j\left(b_j + \frac{1}{2}\boldsymbol{u}_j^T\boldsymbol{u}_j\right)\right],\end{aligned}$$

where $\boldsymbol{w}_i^T$ is the ith row of $\boldsymbol{W}$ for $i = 1, \ldots, n$. Thus, conditional on $(\boldsymbol{\eta}, \boldsymbol{y})$, v_i, $i = 1, \ldots, n$ and $\boldsymbol{\tau}$ are independent. We have

$$v_i|\boldsymbol{\eta}, \boldsymbol{y} \overset{\text{ind}}{\sim} \text{TN}(\boldsymbol{w}_i^T\boldsymbol{\eta}, 1, y_i),\ i = 1, \ldots, n, \tag{17}$$

where $\text{TN}(\mu, \sigma^2, \omega)$ denotes the distribution of the normal random variable with mean μ and variance σ^2, that is truncated to have only positive values if $\omega = 1$, and only nonpositive values if $\omega = 0$. Also conditional on $\boldsymbol{\eta}, \boldsymbol{y}$, τ_j's are independent with $\tau_j \sim \text{Gamma}\left(a_j + q_j/2, b_j + \boldsymbol{u}_j^T\boldsymbol{u}_j/2\right)$ for $j = 1, \ldots, r$.

Thus, one single iteration of the block Gibbs sampler $\{\boldsymbol{\eta}^{(m)}, \boldsymbol{v}^{(m)}, \boldsymbol{\tau}^{(m)}\}_{m=0}^{\infty}$ has the following two steps:

Algorithm 1 The $(m+1)$st iteration of the two-block Gibbs sampler

1: Draw $\tau_j^{(m+1)} \overset{\text{ind}}{\sim} \text{Gamma}\left(a_j + q_j/2, b_j + \boldsymbol{u}_j^{(m)T}\boldsymbol{u}_j^{(m)}/2\right)$ for $j = 1, \ldots, r$, and independently draw $v_i^{(m+1)} \overset{\text{ind}}{\sim} \text{TN}(\boldsymbol{w}_i^T\boldsymbol{\eta}^{(m)}, 1, y_i)$, $i = 1, \ldots, n$.

2: Draw $\boldsymbol{\eta}^{(m+1)} \sim N_{p+q}\left(\left[\boldsymbol{\Sigma}^{(m+1)}\right]^{-1}\boldsymbol{W}^T\boldsymbol{v}^{(m+1)}, \left[\boldsymbol{\Sigma}^{(m+1)}\right]^{-1}\right)$, where $\boldsymbol{\Sigma}^{(m+1)}$ is evaluated at $\boldsymbol{\tau}^{(m+1)}$.

4. Geometric ergodicity of the block Gibbs sampler

In this section, we establish the geometric rate of convergence of the block Gibbs sampler $\{\boldsymbol{\eta}^{(m)}, \boldsymbol{v}^{(m)}, \boldsymbol{\tau}^{(m)}\}_{m=0}^{\infty}$. Since it is a two-block Gibbs sampler, it has the same rate of convergence as the $\boldsymbol{\eta}$-marginal Markov chain $\{\boldsymbol{\eta}^{(m)}\}_{m=0}^{\infty}$ (Roberts and Rosenthal, 2001). Below we analyze this $\boldsymbol{\Psi} \equiv \{\boldsymbol{\eta}^{(m)}\}_{m=0}^{\infty}$ chain.

Let $\boldsymbol{\eta}'$ be the current state and $\boldsymbol{\eta}$ be the next state of the Markov chain $\boldsymbol{\Psi}$, then the Markov transition density (Mtd) of $\boldsymbol{\Psi}$ is

$$k(\boldsymbol{\eta}|\boldsymbol{\eta}') = \int_{\mathbb{R}_+^r}\int_{\mathbb{R}^n} \pi(\boldsymbol{\eta}|\boldsymbol{v}, \boldsymbol{\tau}, \boldsymbol{y})\pi(\boldsymbol{v}, \boldsymbol{\tau}|\boldsymbol{\eta}', \boldsymbol{y})d\boldsymbol{v}d\boldsymbol{\tau}, \tag{18}$$

where $\pi(\cdot|\cdot, \boldsymbol{y})$'s are the conditional densities from section 3. Routine calculations show that $k(\boldsymbol{\eta}|\boldsymbol{\eta}')$ is reversible and thus is invariant with respect to the marginal density of $\boldsymbol{\eta}$ denoted as $\pi(\boldsymbol{\eta}|\boldsymbol{y}) \equiv \int_{\mathbb{R}_+^r}\int_{\mathbb{R}^n} \pi(\boldsymbol{\eta}, \boldsymbol{v}, \boldsymbol{\tau}|\boldsymbol{y})d\boldsymbol{v}d\boldsymbol{\tau}$. Let $h : \mathbb{R}^{p+q} \mapsto \mathbb{R}$ be a real valued function. Suppose our interest is to estimate the (posterior) mean $E(h(\boldsymbol{\eta})|\boldsymbol{y}) \equiv \int_{\mathbb{R}^{p+q}} h(\boldsymbol{\eta})\pi(\boldsymbol{\eta}|\boldsymbol{y})d\boldsymbol{\eta}$. Since $k(\boldsymbol{\eta}|\boldsymbol{\eta}')$ is strictly positive, the Markov chain $\boldsymbol{\Psi}$ is Harris ergodic (Meyn and Tweedie, 1993). Thus if $E(|h(\boldsymbol{\eta})||\boldsymbol{y}) < \infty$, then $E(h(\boldsymbol{\eta})|\boldsymbol{y})$ can be consistently estimated by

$$\bar{h}_m = \frac{1}{m}\sum_{i=0}^{m-1} h(\boldsymbol{\eta}^{(i)}).$$

As mentioned in the introduction, in order to provide an asymptotically valid confidence interval for $E(h(\boldsymbol{\eta})|\boldsymbol{y})$ based on $\bar{h}_m$, we need to establish a CLT for $\bar{h}_m$. We say a CLT exists for $\bar{h}_m$ if there exists a constant $\sigma_h^2 \in (0, \infty)$ such that,

$$\sqrt{m}\left(\bar{h}_m - E(h(\boldsymbol{\eta})|\boldsymbol{y})\right) \xrightarrow{d} N\left(0, \sigma_h^2\right) \text{ as } m \to \infty. \tag{19}$$

If (19) holds, and a consistent estimator $\hat{\sigma}_h^2$ of σ_h^2 is available, then the standard errors $\hat{\sigma}_h/\sqrt{m}$ can be used to provide an asymptotic confidence interval for $E(h(\boldsymbol{\eta})|\boldsymbol{y})$ (Roy and Hobert, 2007). Unfortunately, Harris ergodicity of $\boldsymbol{\Psi}$ does not guarantee (19), although it ensures consistency of $\bar{h}_m$. One method of proving (19) is to establish the geometric rate of convergence for the Markov chain $\mathbf{\Psi}$ (Jones and Hobert, 2001). Geometric ergodicity of $\mathbf{\Psi}$ also allows for consistent estimation of σ_h^2 using batch means or spectral variance methods (Flegal and Jones, 2010).

Let $\mathscr{B}$ denote the Borel σ-algebra of $\mathbb{R}^{p+q}$ and $K(\cdot,\cdot)$ be the Markov transition function corresponding to the Mtd $k(\cdot,\cdot)$ in (18), that is, for any set $O \in \mathscr{B}$, $\boldsymbol{\eta}' \in \mathbb{R}^{p+q}$ and any $j = 0, 1, \dots,$

$$K(\boldsymbol{\eta}', O) = \Pr(\boldsymbol{\eta}^{(j+1)} \in O | \boldsymbol{\eta}^{(j)} = \boldsymbol{\eta}') = \int_O k(\boldsymbol{\eta}|\boldsymbol{\eta}')d\boldsymbol{\eta}. \tag{20}$$

Then the m-step Markov transition function is $K^m(\boldsymbol{\eta}', O) = \Pr(\boldsymbol{\eta}^{(m+j)} \in O|\boldsymbol{\eta}^{(j)} = \boldsymbol{\eta}')$. Let $\Pi(\cdot|\boldsymbol{y})$ be the probability measure with density $\pi(\boldsymbol{\eta}|\boldsymbol{y})$. The Markov chain $\mathbf{\Psi}$ is geometrically ergodic if there exists a constant $0 < t < 1$ and a function $J : \mathbb{R}^{p+q} \mapsto \mathbb{R}^+$ such that for any $\boldsymbol{\eta} \in \mathbb{R}^{p+q}$,

$$||K^m(\boldsymbol{\eta}, \cdot) - \Pi(\cdot|\boldsymbol{y})||_{\mathrm{TV}} := \sup_{O \in \mathscr{B}} |K^m(\boldsymbol{\eta}, O) - \Pi(O|\boldsymbol{y})| \le J(\boldsymbol{\eta})t^m. \tag{21}$$

Harris ergodicity of $\mathbf{\Psi}$ implies that $||K^m(\boldsymbol{\eta}, \cdot) - \Pi(\cdot|\boldsymbol{y})||_{\mathrm{TV}} \downarrow 0$ as $m \to \infty$, while (21) guarantees its exponential rate of convergence. Roberts and Rosenthal (1997) showed that since $\mathbf{\Psi}$ is reversible, if (21) holds then there exists a CLT, that is (19) holds, for all h with $E(h^2(\boldsymbol{\eta})|\boldsymbol{y}) < \infty$.

In section 2, we provided two sets of conditions for posterior propriety. While the first set of conditions ($A1 - A5$) holds in the special case $b_j = 0$ for all $j = 1, \dots, r$, Theorem 1 holds for the general prior $\pi(\tau_j)$ given in (5). In Theorems 2 and 3, we provide conditions under which the Markov chain $\mathbf{\Psi}$ is geometrically ergodic, that is, (21) holds. Here we consider the general form of the prior distribution of τ_j as given in (5). Thus the parameters b_j's are not assumed to be zero. Since geometric ergodicity implies posterior propriety, Theorem 2 also provides conditions for posterior propriety for the probit linear mixed models in the general case when $b_j \neq 0$.

Theorem 2. *The Markov chain underlying the block Gibbs sampler is geometrically ergodic if the following conditions hold:*

(1) $a_j < b_j = 0$ *or* $b_j > 0$ *for* $j = 1, \dots, r$*;*
(2) $(A1) - (A3)$ *hold.*

A proof of Theorem 2 is given in the Appendix C. Theorem 3 shows geometric convergence of the Markov chain underlying the Gibbs sampler given in Algorithm 1 without the assumption A1.

Theorem 3. *The block Gibbs sampler is geometrically ergodic under the following conditions:*

(1) $(B1) - (B4)$ *hold;*
(2) There exists an $s \in (0, 1] \cap (0, \tilde{s})$ *such that*

$$2^{-s} \sum_{j=1}^{r} \frac{\Gamma(q_j/2 + a_j - s)}{\Gamma(q_j/2 + a_j)} \left[tr\left(R_j \left(\boldsymbol{I} - P_{\boldsymbol{Z}^T(\boldsymbol{I}-P_X)\boldsymbol{Z}} \right) R_j^T \right) \right]^s < 1, \tag{22}$$

where $\tilde{s} = \min\{a_1 + q_1/2, \ldots, a_r + q_r/2\}$, R_j *is a* $q_j \times q$ *matrix with 0's and 1's such that* $R_j \boldsymbol{u} = \boldsymbol{u}_j$ *and* $P_{\boldsymbol{Z}^T(\boldsymbol{I}-P_X)\boldsymbol{Z}}$ *is the projection matrix on the column space of* $\boldsymbol{Z}^T(\boldsymbol{I}-P_X)\boldsymbol{Z}$.

A proof of Theorem 3 is given in Appendix D.

Remark 2. *The extra condition (2) in Theorem 3 compared to Theorem 2 is due to the lack of the full rank assumption of* $\boldsymbol{W}$*, and the need to include an extra term in the drift function used to prove Theorem 3. This condition is also used in Román and Hobert (2012), who provide some discussions on this. The left-hand side of* (22) *can be evaluated at values of* s *on a fine grid in the interval* $(0,1] \cap (0,\tilde{s})$ *to numerically check the condition. Note that,* R_j *is the matrix that extracts* $\boldsymbol{u}_j$ *out of* $\boldsymbol{u}$*. Thus when* $r > 1$, *tr* $\left(R_j\left(\boldsymbol{I} - P_{\boldsymbol{Z}^T(\boldsymbol{I}-P_X)\boldsymbol{Z}}\right)R_j^T\right)$ *is the sum of the* q_j *diagonal elements of* $\boldsymbol{I} - P_{\boldsymbol{Z}^T(\boldsymbol{I}-P_X)\boldsymbol{Z}}$ *corresponding to the* j*th random effect.*

5. A Haar PX-DA algorithm

As mentioned in section 1, DA algorithms often suffer from slow convergence and high autocorrelations. Liu and Wu (1999) proposed parameter expansion for data augmentation (PX-DA) algorithms for speeding up the convergence of DA algorithms. Hobert and Marchev (2008) compared the performance of PX-DA algorithms based on a Haar measure (called Haar PX-DA algorithms) with PX-DA algorithms based on a probability measure and DA algorithms. In particular, they showed that, under some mild conditions, the Haar PX-DA algorithms are better than the general PX-DA algorithms and the DA algorithms in both the efficiency ordering and the operator norm ordering. As shown in Hobert and Marchev (2008), compared to the DA algorithm, in PX-DA, an extra step is added (sandwiched) between the two steps of the original DA algorithm. In order to construct this extra step, we derive the marginal density

$$\begin{aligned}\pi(\boldsymbol{v},\boldsymbol{\tau}|\boldsymbol{y}) &= \int_{\mathbb{R}^{p+q}} \pi(\boldsymbol{\eta},\boldsymbol{v},\boldsymbol{\tau}|\boldsymbol{y}) d\boldsymbol{\eta} \\ &\propto \prod_{i=1}^{n} \left[1_{(0,\infty)}(v_i)\right]^{y_i} \left[1_{(-\infty,0]}(v_i)\right]^{1-y_i} \prod_{j=1}^{r} \tau_j^{\frac{q_j}{2}+a_j} e^{-b_j\tau_j} \\ &\quad \cdot |\boldsymbol{\Sigma}|^{-1/2} \exp\left\{-\frac{1}{2}\boldsymbol{v}^T M_1 \boldsymbol{v}\right\},\end{aligned} \tag{23}$$

where $M_1 = \boldsymbol{I} - \boldsymbol{W}\boldsymbol{\Sigma}^{-1}\boldsymbol{W}^T$.

Let $\mathcal{Z}$ denote the subset of $\mathbb{R}^n$ where $\boldsymbol{v}$ lives, that is, $\mathcal{Z}$ is the Cartesian product of n half (positive or nonpositive) lines, where the ith component is $(0,\infty)$ (if $y_i = 1$) or $(-\infty, 0]$ (if $y_i = 0$). Let G be the unimodular multiplicative group on $\mathbb{R}_+$ with Haar measure $\nu(dg) = dg/g$, where dg is the Lebesgue measure on $\mathbb{R}_+$. For constructing an efficient extra step, as in Roy (2014), we let the group G act on $\mathcal{Z} \times \mathbb{R}_+^r$ through a group action $T(\boldsymbol{v},\boldsymbol{\tau}) = (g\boldsymbol{v},\boldsymbol{\tau}) = (gv_1, gv_2, \ldots, gv_n, \boldsymbol{\tau})$.

With the group action defined this way, it can be shown that the Lebesgue measure on $\mathcal{Z} \times \mathbb{R}_+^r$ is relatively left invariant with multiplier $\chi(g) = g^n$ (Roy, 2014; Hobert and Marchev, 2008). Following Hobert and Marchev (2008), consider a probability density function $\vartheta(g)$ on G where

$$\vartheta(g)\, dg \propto \pi(g\boldsymbol{v}, \boldsymbol{\tau}|\boldsymbol{y})\, \chi(g)\, \nu(dg) \propto g^{n-1} \exp\left\{-\frac{1}{2} g^2 \boldsymbol{v}^T M_1 \boldsymbol{v}\right\} dg. \tag{24}$$

Since propriety of the posterior density (13) implies that $\pi(\boldsymbol{v}, \boldsymbol{\tau}|\boldsymbol{y})$ is a valid density, $\boldsymbol{v}^T M_1 \boldsymbol{v}$ can be zero only on a set of measure 0 (in $\boldsymbol{v}$). Thus given $(\boldsymbol{v}, \boldsymbol{\tau})$, $\vartheta(g)$ is a valid density. From Hobert and Marchev (2008), it follows that the transition $(\boldsymbol{v}, \boldsymbol{\tau}) \to (\boldsymbol{v}', \boldsymbol{\tau}) \equiv T(\boldsymbol{v}, \boldsymbol{\tau}) = (g\boldsymbol{v}, \boldsymbol{\tau})$ where $g \sim \vartheta(g)$, is reversible with respect to $\pi(\boldsymbol{v}, \boldsymbol{\tau}|\boldsymbol{y})$ defined in (23). Given $\boldsymbol{\eta}^{(m)}$, below are the three steps involved in the $(m+1)$st iteration of the Haar PX-DA algorithm to move to the new state $\boldsymbol{\eta}^{(m+1)}$.

Algorithm 2 The $(m+1)$st iteration of the Haar PX-DA algorithm

1: $\tau_j \sim \text{Gamma}\left(a_j + q_j/2, b_j + \boldsymbol{u}_j^{(m)T} \boldsymbol{u}_j^{(m)}/2\right)$, for $j = 1, \ldots, r$ and independently draw $v_i | \boldsymbol{\eta}^{(m)}, \boldsymbol{y} \overset{\text{ind}}{\sim} \text{TN}(\boldsymbol{w}_i^T \boldsymbol{\eta}^{(m)}, 1, y_i)$ for $i = 1, \ldots, n$.
2: Draw g^2 from $\text{Gamma}(n/2, \boldsymbol{v}^T M_1 \boldsymbol{v}/2)$.
3: Set $v_i' = g v_i$ and let $\boldsymbol{v}' = (v_1', \ldots, v_n')^T$. Draw

$$\boldsymbol{\eta}^{(m+1)} \sim N_{p+q}\left(\boldsymbol{\Sigma}(\boldsymbol{\tau})^{-1} \boldsymbol{W}^T \boldsymbol{v}', \boldsymbol{\Sigma}(\boldsymbol{\tau})^{-1}\right).$$

The Mtd of the above Haar PX-DA algorithm can be written as

$$k^*(\boldsymbol{\eta}|\boldsymbol{\eta}') = \int_{\mathbb{R}^n} \int_{\mathbb{R}_+^r} \int_{\mathbb{R}^n} \pi(\boldsymbol{\eta}|\boldsymbol{v}', \boldsymbol{\tau}, \boldsymbol{y})\, Q(\boldsymbol{v}, d\boldsymbol{v}')\, \pi(\boldsymbol{v}, \boldsymbol{\tau}|\boldsymbol{\eta}', \boldsymbol{y})\, d\boldsymbol{v} d\boldsymbol{\tau} d\boldsymbol{v}', \tag{25}$$

where $Q(\cdot, \cdot)$ is the Markov transition function corresponding to the move $(\boldsymbol{v}, \boldsymbol{\tau}) \to (\boldsymbol{v}', \boldsymbol{\tau}) = T(\boldsymbol{v}, \boldsymbol{\tau})$. Let K^* and K be the Markov operators associated with the Mtds k^* and k defined in (25) and (18) respectively. From Hobert and Marchev (2008), we have $\|K^*\|_{\text{OP}} \le \|K\|_{\text{OP}}$, where $\|K\|_{\text{OP}}$ denotes the norm of the operator K (see also Roy, 2012a). Since the block Gibbs sampler is geometrically ergodic, we have $\|K^*\|_{\text{OP}} \le \|K\|_{\text{OP}} < 1$ (Roberts and Rosenthal, 1997). Thus we have the following corollary.

Corollary 1. *Under the conditions of Theorem 2 or Theorem 3, the Markov chain underlying the Haar PX-DA algorithm described in Algorithm 2 is geometrically ergodic.*

The extra step of Algorithm 2 is a single draw from the univariate density $\vartheta(g)$, which is easy to sample from. Thus, the computational burden, per iteration, for the Haar PX-DA algorithm is similar to that of the block Gibbs sampler described in section 3. Two other Haar PX-DA algorithms can be constructed by using group actions $T_1(\boldsymbol{v}, \boldsymbol{\tau}) = (\boldsymbol{v}, g\boldsymbol{\tau})$ and $T_2(\boldsymbol{v}, \boldsymbol{\tau}) = (g\boldsymbol{v}, g\boldsymbol{\tau})$. However, the corresponding $\vartheta(g)$'s are not easy to sample from, thus we do not consider them here.

6. Discussion

We develop a two-block Gibbs sampler for the Bayesian probit linear mixed models under improper priors. The block Gibbs algorithm samples the fixed effects and the random effects jointly. We prove the geometric ergodicity of the two-block Gibbs sampler, which guarantees the existence of central limit theorems for MCMC estimators under a finite second moment condition. We also propose a corresponding Haar PX-DA algorithm. The Haar PX-DA algorithm not only improves the efficiency of the Gibbs sampler, but also inherits its geometric convergence properties. We provide easily verifiable conditions for posterior propriety of Bayesian GLMMs.

Another popular link function is the logit link function. Polson, Scott and Windle (2013) proposed a DA algorithm for the logistic regression model. Choi and Hobert (2013) proved the uniform ergodicity of this DA algorithm under normal priors on the regression coefficients. Wang and Roy (2018) prove that the DA algorithm under the flat prior on the regression parameters is geometrically ergodic. As mentioned in Polson, Scott and Windle (2013), their DA algorithm can be extended to the logistic linear mixed model. However, the convergence properties of the corresponding Markov chain have not been studied, and can be a topic for future research. Another future project can be deriving similar extensions of the results in Roy (2012b) for proving geometric convergence of Gibbs samplers for robit linear mixed models.

Appendices

A. Proof of Theorem 1

Proof. Using the transformation $(\boldsymbol{\beta}^T, \boldsymbol{u}^T)^T \to (u_{11}, \dots, u_{r1}, \tilde{\boldsymbol{\eta}}^T)^T$, the integral in (12) can be written as,

$$\int_{\mathbb{R}^{p+q-r}} \int_{\mathbb{R}_+^r} \int_{\mathbb{R}^r} \prod_{i=1}^{n} \left[F\left(\tilde{\boldsymbol{w}}_i^T \tilde{\boldsymbol{\eta}}\right)\right]^{y_i} \left[1 - F\left(\tilde{\boldsymbol{w}}_i^T \tilde{\boldsymbol{\eta}}\right)\right]^{1-y_i} \tag{26}$$
$$\prod_{j=1}^{r} \tau_j^{\frac{q_j}{2}+a_j-1} \exp\left[-\frac{\tau_j}{2}\left(u_{j1}^2 + \sum_{k=1}^{q_j-1}(d_{jk}+u_{j1})^2 + 2b_j\right)\right] du_{11}\cdots du_{r1} d\boldsymbol{\tau} d\tilde{\boldsymbol{\eta}},$$

where $\tilde{\boldsymbol{w}}_i$ is defined in section 2. Let $\bar{d}_j = \sum_{k=1}^{q_j-1} d_{jk}$. Then (26) becomes,

$$\int_{\mathbb{R}^{p+q-r}} \int_{\mathbb{R}_+^r} \prod_{i=1}^{n} \left[F\left(\tilde{\boldsymbol{w}}_i^T \tilde{\boldsymbol{\eta}}\right)\right]^{y_i} \left[1 - F\left(\tilde{\boldsymbol{w}}_i^T \tilde{\boldsymbol{\eta}}\right)\right]^{1-y_i} (2\pi)^{\frac{r}{2}}$$
$$\cdot \prod_{j=1}^{r} q_j^{-1/2} \tau_j^{\frac{q_j}{2}+a_j-\frac{3}{2}} \exp\left[-\frac{\tau_j}{2}\left(\sum_{k=1}^{q_j-1}\left(d_{jk}-\bar{d}_j\right)^2 + \frac{q_j-1}{q_j}\bar{d}_j^2 + 2b_j\right)\right] d\boldsymbol{\tau} d\tilde{\boldsymbol{\eta}}$$
$$= \int_{\mathbb{R}^{p+q-r}} \prod_{i=1}^{n} \left[F\left(\tilde{\boldsymbol{w}}_i^T \tilde{\boldsymbol{\eta}}\right)\right]^{y_i} \left[1 - F\left(\tilde{\boldsymbol{w}}_i^T \tilde{\boldsymbol{\eta}}\right)\right]^{1-y_i} (2\pi)^{\frac{r}{2}}$$

$$\cdot \prod_{j=1}^{r} q_j^{-1/2} \Gamma\left(q_j/2 + a_j - 1/2\right) 2^{\frac{q_j}{2}+a_j-\frac{1}{2}}$$

$$\cdot \left(\sum_{k=1}^{q_j-1} \left(d_{jk} - \bar{d}_j\right)^2 + \frac{q_j-1}{q_j}\bar{d}_j^2 + 2b_j \right)^{-\frac{q_j}{2}-a_j+\frac{1}{2}} d\tilde{\boldsymbol{\eta}}$$

$$\leq \varphi_1 \int_{\mathbb{R}^{p+q-r}} \prod_{i=1}^{n} \left[F\left(\tilde{\boldsymbol{w}}_i^T \tilde{\boldsymbol{\eta}}\right)\right]^{y_i} \left[1 - F\left(\tilde{\boldsymbol{w}}_i^T \tilde{\boldsymbol{\eta}}\right)\right]^{1-y_i}$$

$$\cdot \prod_{j=1}^{r} \left(\sum_{k=1}^{q_j-1} \left(d_{jk} - \bar{d}_j\right)^2 + \frac{q_j-1}{q_j}\bar{d}_j^2 + 2b_j \right)^{-\frac{q_j}{2}-a_j+\frac{1}{2}} d\tilde{\boldsymbol{\eta}}, \quad (27)$$

where φ_1 is a constant depending on r, q_j and a_j, $j = 1, \ldots, r$.

Let $\delta_i, i = 1, \ldots, n$ be n i.i.d random variables with distribution function F. Let $\boldsymbol{\delta}^* = (c_1\delta_1, \ldots, c_n\delta_n)^T$, where $c_i = 1 - 2y_i$ as defined in section 2. We have $E\left[1\left\{c_i\tilde{\boldsymbol{w}}_i^T\tilde{\boldsymbol{\eta}} \leq c_i\delta_i\right\}\right] = \left[F\left(\tilde{\boldsymbol{w}}_i^T\tilde{\boldsymbol{\eta}}\right)\right]^{y_i}\left[1 - F\left(\tilde{\boldsymbol{w}}_i^T\tilde{\boldsymbol{\eta}}\right)\right]^{1-y_i}$, for $i = 1, \ldots, n$. Thus

$$\prod_{i=1}^{n} \left[F\left(\tilde{\boldsymbol{w}}_i^T \tilde{\boldsymbol{\eta}}\right)\right]^{y_i} \left[1 - F\left(\tilde{\boldsymbol{w}}_i^T \tilde{\boldsymbol{\eta}}\right)\right]^{1-y_i} = E\left[1\left\{\tilde{\boldsymbol{W}}^* \tilde{\boldsymbol{\eta}} \leq \boldsymbol{\delta}^*\right\}\right], \quad (28)$$

where $\tilde{\boldsymbol{W}}^*$ is the $n \times (p+q)$ matrix whose ith row is $c_i\tilde{\boldsymbol{w}}_i^T$.

Since conditions *B3* and *B4* are in force, according to Chen and Shao (2001) (Lemma 4.1), there exists a constant φ_0 depending on $\tilde{\boldsymbol{W}}$ and $\boldsymbol{y}$, such that $1\left\{\tilde{\boldsymbol{W}}^*\tilde{\boldsymbol{\eta}} \leq \boldsymbol{\delta}^*\right\} \leq 1\left\{\|\tilde{\boldsymbol{\eta}}\| \leq \varphi_0 \|\boldsymbol{\delta}^*\|\right\}$. Recall that $\tilde{\boldsymbol{\eta}} = (\mu_0, \beta_1, \ldots, \beta_{p-1}, d_{11}, \ldots, d_{1,q_1-1}, \ldots, d_{r1}, \ldots, d_{r,q_r-1})^T = (\mu_0, \beta_1, \ldots, \beta_{p-1}, \boldsymbol{d}_1^T, \ldots, \boldsymbol{d}_r^T)^T$, where $\boldsymbol{d}_j = (d_{j1}, \ldots, d_{j,q_j-1})^T$ for $j = 1, \ldots, r$. Thus from (27) and (28) it follows that (26) is bounded above by

$$\varphi_1 E\left[\int_{\mathbb{R}^{p+q-r}} 1\left\{\|\tilde{\boldsymbol{\eta}}\| \leq \varphi_0 \|\boldsymbol{\delta}^*\|\right\}\right.$$

$$\left. \times \prod_{j=1}^{r} \left(\sum_{k=1}^{q_j-1} \left(d_{jk} - \bar{d}_j\right)^2 + \frac{q_j-1}{q_j}\bar{d}_j^2 + 2b_j \right)^{-\frac{q_j}{2}-a_j+\frac{1}{2}} d\tilde{\boldsymbol{\eta}}\right]$$

$$\leq 2^p \varphi_0^p \varphi_1 E\left[\|\boldsymbol{\delta}^*\|^p\right.$$

$$\left. \times \int_{A_d} \prod_{j=1}^{r} \left(\sum_{k=1}^{q_j-1} \left(d_{jk} - \bar{d}_j\right)^2 + \frac{q_j-1}{q_j}\bar{d}_j^2 + 2b_j \right)^{-\frac{q_j}{2}-a_j+\frac{1}{2}} d\boldsymbol{d}_1 \cdots, d\boldsymbol{d}_r\right]$$

$$\leq 2^p \varphi_0^p \varphi_1 E \left[\|\boldsymbol{\delta}^*\|^p \int_{A_d} \prod_{j=1}^{r} \left(\frac{q_j - 1}{q_j} \bar{d}_j^2 + 2b_j \right)^{-\frac{q_j}{2} - a_j + \frac{1}{2}} d\boldsymbol{d}_1 \cdots, d\boldsymbol{d}_r \right], \quad (29)$$

where $A_d = \{|d_{jk}| \leq \varphi_0 \|\boldsymbol{\delta}^*\|, j = 1, \ldots, r,\ k = 1, \ldots q_j - 1\}$.

We consider two cases of condition *B1* separately.

Case 1: $a_j < b_j = 0$, $q_j \geq 2$. If $q_j = 2$, we have

$$\begin{aligned} \int_{|d_{j1}| \leq 1} \left[(d_{j1})^2 \right]^{-a_j - \frac{1}{2}} dd_{j1} &= -\frac{1}{2a_j} \left[(d_{j1})^2 \right]^{-a_j - \frac{1}{2}} d_{j1} \Big|_{-1}^{1} \\ &= -\frac{1}{2a_j} (1 + 1) = -\frac{1}{a_j} < \infty. \end{aligned}$$

For $q_j > 2$, note that,

$$\begin{aligned} &\int_{|d_{j1}| \leq 1} \left[\left(\sum_{k=1}^{q_j - 1} d_{jk} \right)^2 \right]^{-\frac{q_j}{2} - a_j + \frac{1}{2}} dd_{j1} \\ &= \frac{1}{2 - q_j - 2a_j} \left[\left(\sum_{k=1}^{q_j - 1} d_{jk} \right)^2 \right]^{-\frac{q_j}{2} - a_j + \frac{1}{2}} \left(\sum_{k=1}^{q_j - 1} d_{jk} \right) \Big|_{-1}^{1} \\ &= \frac{1}{2 - q_j - 2a_j} \left[\left(1 + \sum_{k=2}^{q_j - 1} d_{jk} \right)^2 \right]^{-\frac{q_j}{2} - a_j + \frac{1}{2}} \left(1 + \sum_{k=2}^{q_j - 1} d_{jk} \right) \\ &- \frac{1}{2 - q_j - 2a_j} \left[\left(-1 + \sum_{k=2}^{q_j - 1} d_{jk} \right)^2 \right]^{-\frac{q_j}{2} - a_j + \frac{1}{2}} \left(-1 + \sum_{k=2}^{q_j - 1} d_{jk} \right) \\ &\leq \frac{1}{2 - q_j - 2a_j} \left\{ \left[\left(1 + \sum_{k=2}^{q_j - 1} d_{jk} \right)^2 \right]^{-\frac{q_j}{2} - a_j + 1} \right. \\ &\quad \left. + \left[\left(-1 + \sum_{k=2}^{q_j - 1} d_{jk} \right)^2 \right]^{-\frac{q_j}{2} - a_j + 1} \right\}. \end{aligned} \quad (30)$$

Ignoring the constant multiple, continuing integrating (30) with respect to $d_{j2}, \ldots, d_{j,q_j - 1}$ consecutively, we arrive at some linear combinations of terms

$$\left[\left(\alpha_0 + d_{j,q_j - 1} \right)^2 \right]^{-\frac{q_j}{2} - a_j + \frac{q_j - 1}{2}} \left(\alpha_0 + d_{j,q_j - 1} \right) \Big|_{-1}^{1}, \quad (31)$$

where α_0's are constants. Since $a_j < 0$, each of these terms in (31) is finite. Then

$$\begin{aligned}
&\int_{A_{dj}} \left(\frac{q_j - 1}{q_j}\bar{d}_j^2 + b_j\right)^{-\frac{q_j}{2} - a_j + \frac{1}{2}} d\boldsymbol{d}_j \\
=&\left[q_j\left(q_j - 1\right)\right]^{\frac{q_j}{2} + a_j - \frac{1}{2}} \left(\varphi_0 \left\|\boldsymbol{\delta}^*\right\|\right)^{-2a_j} \\
&\cdot \int_{\{|d_{jk}| \le 1, k=1,\dots,q_j-1\}} \left[\left(\sum_{k=1}^{q_j-1} d_{jk}\right)^2\right]^{-\frac{q_j}{2} - a_j + \frac{1}{2}} d\boldsymbol{d}_j \\
\le&\varphi_{2j} \left\|\boldsymbol{\delta}^*\right\|^{-2a_j}, \qquad (32)
\end{aligned}$$

where $A_{dj} = \{|d_{jk}| \le \varphi_0 \|\boldsymbol{\delta}^*\|, k = 1, \dots q_j - 1\}$ and φ_{2j} is a finite positive constant.

Case 2: $b_j > 0$.

We have

$$\begin{aligned}
&\int_{A_{dj}} \left(\frac{q_j - 1}{q_j}\bar{d}_j^2 + 2b_j\right)^{-\frac{q_j}{2} - a_j + \frac{1}{2}} d\boldsymbol{d}_j \le \int_{A_{dj}} (2b_j)^{-\frac{q_j}{2} - a_j + \frac{1}{2}} d\boldsymbol{d}_j \\
\le&(2b_j)^{-\frac{q_j}{2} - a_j + \frac{1}{2}} 2^{q_j - 1} \varphi_0^{q_j - 1} \left\|\boldsymbol{\delta}^*\right\|^{q_j - 1} \le \varphi_{3j} \left\|\boldsymbol{\delta}^*\right\|^{q_j - 1}, \qquad (33)
\end{aligned}$$

where φ_{3j} is a finite positive constant.

Using (32), (33), and condition *B5*, it follows that (26) can be bounded above by

$$\begin{aligned}
&2^p \varphi_0^p \varphi_1 E\left[\left\|\boldsymbol{\delta}^*\right\|^p \prod_{j=1}^{r} \left\{\varphi_{2j} \left\|\boldsymbol{\delta}^*\right\|^{-2a_j} I(b_j = 0) + \varphi_{3j} \left\|\boldsymbol{\delta}^*\right\|^{q_j - 1} I(b_j > 0)\right\}\right] \\
=&2^p \varphi_0^p \varphi_1 \prod_{j: b_j = 0} \varphi_{2j} \prod_{j: b_j > 0} \varphi_{3j} \left[E \left\|\boldsymbol{\delta}^*\right\|^{p + \sum_{j=1}^{r}[-2a_j I(b_j = 0) + (q_j - 1) I(b_j > 0)]}\right] \\
<&\infty. \qquad \square
\end{aligned}$$

Remark 3. *If $q_j = 1$ and $b_j = 0$, (26) is ∞ since $\int_{\mathbb{R}_+} \tau_j^{q_j/2 + a_j - 3/2} d\tau_j = \infty$. If $b_j > 0$, the posterior density (4) can be proper even when $q_j = 1$.*

B. Two Lemmas

In this section, we list some technical results. For $\boldsymbol{\Sigma}$ defined in (16), note that

$$\boldsymbol{\Sigma}^{-1} = \begin{pmatrix} \left(\boldsymbol{X}^T \boldsymbol{X}\right)^{-1} + RS(\boldsymbol{\tau})^{-1} R^T & -RS(\boldsymbol{\tau})^{-1} \\ -S(\boldsymbol{\tau})^{-1} R^T & S(\boldsymbol{\tau})^{-1} \end{pmatrix}, \qquad (34)$$

with $S(\boldsymbol{\tau})$ and R defined as

$$S(\boldsymbol{\tau}) = \boldsymbol{Z}^T \left(\boldsymbol{I} - P_X\right) \boldsymbol{Z} + \boldsymbol{D}(\boldsymbol{\tau}), \text{ and } R = \left(\boldsymbol{X}^T \boldsymbol{X}\right)^{-1} \boldsymbol{X}^T \boldsymbol{Z} \qquad (35)$$

respectively, where

$$P_X = \boldsymbol{X}\left(\boldsymbol{X}^T\boldsymbol{X}\right)^{-1}\boldsymbol{X}^T. \tag{36}$$

Also the mean for the conditional distribution of $\boldsymbol{\eta}$ in (15) becomes

$$\boldsymbol{\Sigma}^{-1}\boldsymbol{W}^T\boldsymbol{v} = \begin{pmatrix} \left(\boldsymbol{X}^T\boldsymbol{X}\right)^{-1}\boldsymbol{X}^T\left[\boldsymbol{I} - \boldsymbol{Z}S(\boldsymbol{\tau})^{-1}\boldsymbol{Z}^T\left(\boldsymbol{I} - P_X\right)\right]\boldsymbol{v} \\ S(\boldsymbol{\tau})^{-1}\boldsymbol{Z}^T\left(\boldsymbol{I} - P_X\right)\boldsymbol{v} \end{pmatrix}.$$

Let $U^T\Lambda U$ be the spectral decomposition of $\boldsymbol{Z}^T\left(\boldsymbol{I} - P_X\right)\boldsymbol{Z}$ and let λ_j's be the diagonal elements of Λ. Then $\left(\boldsymbol{Z}^T\left(\boldsymbol{I} - P_X\right)\boldsymbol{Z}\right)^+ \equiv U^T\Lambda^+U$, where Λ^+ is a diagonal matrix whose jth diagonal element is $\lambda_j^+ = 1/\lambda_j$ if $\lambda_j \neq 0$, and 0 otherwise.

Lemma 1. *For the matrices $S(\boldsymbol{\tau})$ and P_X defined in (35) and (36), the following inequalities hold for all $\tau_j \in \mathbb{R}_+$, $j = 1, \ldots, r$:*

1. $S(\boldsymbol{\tau})^{-1} \preceq \left(\boldsymbol{Z}^T\left(\boldsymbol{I} - P_X\right)\boldsymbol{Z}\right)^+ + \sum_{j=1}^r 1/\tau_j\left(\boldsymbol{I} - P_{\boldsymbol{Z}^T(\boldsymbol{I}-P_X)\boldsymbol{Z}}\right)$.
2. $\left(R_jS(\boldsymbol{\tau})^{-1}R_j^T\right)^{-1} \preceq \left(\lambda_p + \tau_j\right)\boldsymbol{I}_{q_j}$, *where λ_p is the largest eigenvalue of $\boldsymbol{Z}^T\left(\boldsymbol{I} - P_X\right)\boldsymbol{Z}$ and R_j is a $q_j \times q$ matrix with 0's and 1's such that $R_j\boldsymbol{u} = \boldsymbol{u}_j$.*

The proof of the above result is similar to that of Lemma 1 in Román and Hobert (2012) and we omit it.

Lemma 2. *Let $S(\boldsymbol{\tau})$ and P_X be the two matrices as defined in (35) and (36). Let $\boldsymbol{l} = (l_1, \ldots, l_n)^T \in \mathbb{R}^n$. For any $\boldsymbol{\tau} \in \mathbb{R}_+^r$, we have,*

$$\left\|S(\boldsymbol{\tau})^{-1}\boldsymbol{Z}^T\left(\boldsymbol{I} - \boldsymbol{P}_X\right)\boldsymbol{l}\right\| \leq \hat{\varphi}\sum_{i=1}^n |l_i|,$$

where $\|\cdot\|$ denotes the Euclidean norm and $\hat{\varphi}$ is a finite number that depends on $\boldsymbol{W}$.

Proof. Let $\boldsymbol{Z}_P \equiv \left(\boldsymbol{I} - P_X\right)\boldsymbol{Z}$ and $\boldsymbol{z}_{Pi}^T$ be the ith row of $\boldsymbol{Z}_P$. Then

$$\begin{aligned}
&\left\|S(\boldsymbol{\tau})^{-1}\boldsymbol{Z}^T\left(\boldsymbol{I} - \boldsymbol{P}_X\right)\boldsymbol{l}\right\| = \left\|\left(\boldsymbol{Z}_P^T\boldsymbol{Z}_P + \boldsymbol{D}(\boldsymbol{\tau})\right)^{-1}\boldsymbol{Z}_P^T\boldsymbol{l}\right\| \\
&= \left\|\sum_{i=1}^n\left(\boldsymbol{Z}_P^T\boldsymbol{Z}_P + \boldsymbol{D}(\boldsymbol{\tau})\right)^{-1}\boldsymbol{z}_{Pi}l_i\right\| \leq \sum_{i=1}^n\left\|\left(\boldsymbol{Z}_P^T\boldsymbol{Z}_P + \boldsymbol{D}(\boldsymbol{\tau})\right)^{-1}\boldsymbol{z}_{Pi}l_i\right\| \\
&= \sum_{i=1}^n\left\|\left(\sum_{k=1}^n\boldsymbol{z}_{Pk}\boldsymbol{z}_{Pk}^T + \boldsymbol{D}(\boldsymbol{\tau})\right)^{-1}\boldsymbol{z}_{Pi}l_i\right\| \leq \sum_{i=1}^n |l_i|\,\varphi_i\left(\boldsymbol{\tau}\right),
\end{aligned}$$

where

$$\varphi_i^2\left(\boldsymbol{\tau}\right) = \boldsymbol{z}_{Pi}^T\left(\boldsymbol{z}_{Pi}\boldsymbol{z}_{Pi}^T + \sum_{k\in\{1,\ldots,n\}\setminus\{i\}}\boldsymbol{z}_{Pk}\boldsymbol{z}_{Pk}^T + \boldsymbol{D}(\boldsymbol{\tau})\right)^{-2}\boldsymbol{z}_{Pi}.$$

Note that for fixed $i \in \{1, 2 \ldots, n\}$,

$$\begin{aligned}\varphi_i^2(\boldsymbol{\tau}) &= \boldsymbol{z}_{Pi}^T \left(\boldsymbol{z}_{Pi}\boldsymbol{z}_{Pi}^T + \sum_{k\in\{1,\ldots,n\}\setminus\{i\}} \boldsymbol{z}_{Pk}\boldsymbol{z}_{Pk}^T + \boldsymbol{D}(\boldsymbol{\tau}) \right. \\ &\quad \left. - \frac{1}{\sum_{j=1}^r 1/\tau_j}\boldsymbol{I}_q + \frac{1}{\sum_{j=1}^r 1/\tau_j}\boldsymbol{I}_q \right)^{-2} \boldsymbol{z}_{Pi} \\ &\leq \sup_{\boldsymbol{\iota}\in\mathbb{R}_+^{n+q}} t_i^T \left(t_i t_i^T + \sum_{k\in\{1,\ldots n\}\setminus\{i\}} \iota_k t_k t_k^T + \sum_{k=n+1}^{n+q} \iota_k t_k t_k^T + \iota_1 \boldsymbol{I}_q \right)^{-2} t_i \\ &\equiv \hat{\varphi}_i^2,\end{aligned}$$

where $\boldsymbol{\iota} = (\iota_1, \iota_2, \ldots, \iota_{n+q})$, $t_k = \boldsymbol{z}_{Pk}$ for $k \in \{1, \ldots, n\}$ and for $k = n+1, \ldots, n+q$, define t_k to be a $q \times 1$ unit vector with 1 on the $(k-n)$th position, 0 elsewhere. The inequality follows from the fact that $\sum_{j=1}^r 1/\tau_j > 1/\tau_j$. By Lemma 3 in Román and Hobert (2012), we know that $\hat{\varphi}_i^2$ is finite. Let $\hat{\varphi} = \max_{1\leq i\leq n} \hat{\varphi}_i$, then

$$\left\| S(\boldsymbol{\tau})^{-1} \boldsymbol{Z}^T (\boldsymbol{I} - \boldsymbol{P}_X) \boldsymbol{l} \right\| \leq \sum_{i=1}^n |l_i| \hat{\varphi}_i \leq \hat{\varphi} \sum_{i=1}^n |l_i| . \qquad \square$$

C. Proof of Theorem 2

The two-block Gibbs sampler $\{\boldsymbol{\eta}^{(m)}, (\boldsymbol{v}^{(m)}, \boldsymbol{\tau}^{(m)})\}_{m=0}^\infty$ in Algorithm 1 has the same rate of convergence as its two marginal chains, namely, the $\boldsymbol{\eta}$-chain and the $(\boldsymbol{v}, \boldsymbol{\tau})$-chain. Here we work with the $\boldsymbol{\eta}$-chain, denoted as $\boldsymbol{\Psi} = \{\boldsymbol{\eta}^{(m)}\}_{m=0}^\infty$ and establish its geometric rate of convergence. Define $A \equiv \{j \in \{1, \ldots, r\} : b_j = 0\}$. Recall that given $\boldsymbol{\eta}$, the conditional distribution of $\boldsymbol{\tau}$ is given by independent $\text{Gamma}(a_j + q_j/2, b_j + \boldsymbol{u}_j^T \boldsymbol{u}_j/2)$, $j = 1, \ldots, r$, which is not defined when A is not empty and $\boldsymbol{\eta} \in \mathcal{N} = \left\{ \boldsymbol{\eta} \in \mathbb{R}^{p+q}; \prod_{j\in A} ||\boldsymbol{u}_j|| = 0 \right\}$. Since $\mathcal{N}$ is a set of measure zero, simulation of the Gibbs sampler is not affected by the fact that $\pi(\boldsymbol{\tau}|\boldsymbol{\eta}, \boldsymbol{y})$ is not defined on $\mathcal{N}$. But as mentioned in Román and Hobert (2012), for a theoretical analysis of the $\boldsymbol{\eta}$-chain, the Mtd of $\boldsymbol{\Psi}$ and hence $\pi(\boldsymbol{\tau}|\boldsymbol{\eta}, \boldsymbol{y})$ must be defined for *all* $\boldsymbol{\eta} \in \mathbb{R}^{p+q}$. Since $\mathcal{N}$ is a measure zero set, the Mtd of $\boldsymbol{\Psi}$ hence $\pi(\boldsymbol{\tau}|\boldsymbol{\eta}, \boldsymbol{y})$ can be defined arbitrarily on $\mathcal{N}$. If A is not empty for all $\boldsymbol{\eta} \in \mathbb{R}^{p+q}$, we define $\pi(\boldsymbol{\tau}|\boldsymbol{\eta}, \boldsymbol{y})$ as follows,

$$\pi(\boldsymbol{\tau}|\boldsymbol{\eta}, \boldsymbol{y}) = \begin{cases} \prod_{j=1}^r f_G\left(\tau_j, \frac{q_j}{2} + a_j, \frac{\boldsymbol{u}_j^T \boldsymbol{u}_j}{2} + b_j\right) & \text{if } \boldsymbol{\eta} \notin \mathcal{N} \\ \prod_{j=1}^r f_G(\tau_j, 1, 1) & \text{if } \boldsymbol{\eta} \in \mathcal{N} \end{cases},$$

where f_G stands for the density of a Gamma random variable.

We denote the $\{\boldsymbol{\eta}^{(m)}\}_{m=0}^\infty$ Markov chain defined on $\mathbb{R}^{p+q}\setminus\mathcal{N}$ as $\tilde{\boldsymbol{\Psi}}$. The chain $\tilde{\boldsymbol{\Psi}}$ is Harris ergodic on $\mathbb{R}^{p+q}\setminus\mathcal{N}$. Our proof of geometric ergodicity of $\boldsymbol{\Psi}$ is through that of $\tilde{\boldsymbol{\Psi}}$. The following proof establishes the geometric ergodicity of $\tilde{\boldsymbol{\Psi}}$.

Proof. We prove the geometric ergodicity of $\tilde{\boldsymbol{\Psi}}$ by establishing a drift function, which has the following form,

$$V(\boldsymbol{\eta}) = \sum_{i=1}^{n} \left(\boldsymbol{x}_i^T\boldsymbol{\beta} + \boldsymbol{z}_i^T\boldsymbol{u}\right)^2 + \sum_{j=1}^{r} \left(\boldsymbol{u}_j^T\boldsymbol{u}_j\right)^{-c}, \tag{37}$$

where $c \in (0, 1/2)$ is a positive constant determined later in the proof. Note that, since the condition A1 is in force, $V(\boldsymbol{\eta}) : \mathbb{R}^{p+q}\backslash\mathcal{N} \to [0, \infty)$ is unbounded off compact sets. We show that for any $\boldsymbol{\eta}, \boldsymbol{\eta}' \in \mathbb{R}^{p+q}\backslash\mathcal{N}$, there exists $\rho_1 \in [0, 1)$ and $L_1 > 0$ such that

$$E[V(\boldsymbol{\eta})\,|\boldsymbol{\eta}'] \leq \rho_1 V(\boldsymbol{\eta}') + L_1. \tag{38}$$

By Fubini's theorem, we have

$$\begin{aligned} E\left[V(\boldsymbol{\eta})\,|\boldsymbol{\eta}'\right] &= \int_{\mathbb{R}^{p+q}\backslash\mathcal{N}} V(\boldsymbol{\eta})\, k(\boldsymbol{\eta}|\boldsymbol{\eta}')\, d\boldsymbol{\eta} \\ &= \int_{\mathbb{R}^n}\int_{\mathbb{R}^r_+}\int_{\mathbb{R}^{p+q}\backslash\mathcal{N}} V(\boldsymbol{\eta})\, \pi(\boldsymbol{\eta}|\boldsymbol{v},\boldsymbol{\tau},\boldsymbol{y})\, \pi(\boldsymbol{v},\boldsymbol{\tau}|\boldsymbol{\eta}',\boldsymbol{y})\, d\boldsymbol{\eta}\, d\boldsymbol{\tau}\, d\boldsymbol{v}. \end{aligned}$$

Thus, the expectation on the left hand side of (38) can be evaluated using two steps. First, we calculate the expectation with respect to the conditional distribution of $\boldsymbol{\eta}$ given $\boldsymbol{v}$, $\boldsymbol{\tau}$ and $\boldsymbol{y}$, that is $E[V(\boldsymbol{\eta})|\boldsymbol{v},\boldsymbol{\tau},\boldsymbol{y}]$.

From (34) and (35), we have $\boldsymbol{W}\boldsymbol{\Sigma}^{-1}\boldsymbol{W}^T = P_X + (\boldsymbol{I} - P_X)\,\boldsymbol{Z}S(\boldsymbol{\tau})^{-1}\boldsymbol{Z}^T(\boldsymbol{I} - P_X)$. Also $(\boldsymbol{I} - P_X) = (\boldsymbol{I} - P_X)^2$. Let $\tilde{P} = (\boldsymbol{I} - P_X)\,\boldsymbol{Z}\boldsymbol{D}(\boldsymbol{\tau})^{-1/2}$, then

$$\begin{aligned} &(\boldsymbol{I} - P_X)\,\boldsymbol{Z}S(\boldsymbol{\tau})^{-1}\boldsymbol{Z}^T(\boldsymbol{I} - P_X) \\ =\,&(\boldsymbol{I} - P_X)^2\,\boldsymbol{Z}S(\boldsymbol{\tau})^{-1}\boldsymbol{Z}^T(\boldsymbol{I} - P_X)^2 \\ =\,&(\boldsymbol{I} - P_X)\,\tilde{P}\left(\tilde{P}^T\tilde{P} + \boldsymbol{I}\right)^{-1}\tilde{P}^T(\boldsymbol{I} - P_X) \preceq \boldsymbol{I} - P_X. \end{aligned}$$

Thus, $\boldsymbol{W}\boldsymbol{\Sigma}^{-1}\boldsymbol{W}^T \preceq P_X + \boldsymbol{I} - P_X = \boldsymbol{I}$. Here "$\boldsymbol{W}\boldsymbol{\Sigma}^{-1}\boldsymbol{W}^T \preceq \boldsymbol{I}$" means that $\boldsymbol{I} - \boldsymbol{W}\boldsymbol{\Sigma}^{-1}\boldsymbol{W}^T$ is a positive semidefinite matrix. From (15) and (16), it follows that

$$\begin{aligned} &E\left[\sum_{i=1}^{n}\left(\boldsymbol{x}_i^T\boldsymbol{\beta} + \boldsymbol{z}_i^T\boldsymbol{u}\right)^2|\boldsymbol{v},\boldsymbol{\tau},\boldsymbol{y}\right] \leq E\left[\boldsymbol{\eta}^T\boldsymbol{\Sigma}\boldsymbol{\eta}|\boldsymbol{\tau},\boldsymbol{v},\boldsymbol{y}\right] \\ &= p + q + \boldsymbol{v}\boldsymbol{W}\boldsymbol{\Sigma}^{-1}\boldsymbol{W}^T\boldsymbol{v} \leq p + q + \boldsymbol{v}^T\boldsymbol{v}. \end{aligned} \tag{39}$$

According to Román and Hobert (2012), for $c \in (0, 1/2)$ we have

$$E\left[\left(\boldsymbol{u}_j^T\boldsymbol{u}_j\right)^{-c}|\boldsymbol{v},\boldsymbol{\tau},\boldsymbol{y}\right] \leq 2^{-c}\frac{\Gamma(q_j/2 - c)}{\Gamma(q_j/2)}\left[\lambda_p^c + \tau_j^c\right], \tag{40}$$

where λ_p is the largest eigenvalue of $\boldsymbol{Z}^T(\boldsymbol{I} - P_{\boldsymbol{X}})\boldsymbol{Z}$. Using (39) and (40) from (37), we have

$$E\left[V(\boldsymbol{\eta})\,|\boldsymbol{\tau},\boldsymbol{v},\boldsymbol{y}\right] \leq \boldsymbol{v}^T\boldsymbol{v} + 2^{-c}\sum_{j=1}^{r}\frac{\Gamma(q_j/2 - c)}{\Gamma(q_j/2)}\tau_j^c + 2^{-c}\sum_{j=1}^{r}\frac{\Gamma(q_j/2 - c)}{\Gamma(q_j/2)}\lambda_p^c + p + q. \tag{41}$$

Now we consider the expectation corresponding to the conditional distribution of $\boldsymbol{v}$ and $\boldsymbol{\tau}$ given $\boldsymbol{\eta}'$ and $\boldsymbol{y}$. Using (10) from Roy and Hobert (2007), we have

$$E(v_i^2|\boldsymbol{\eta}',\boldsymbol{y}) = \begin{cases} 1+(\boldsymbol{x}_i^T\boldsymbol{\beta}'+\boldsymbol{z}_i^T\boldsymbol{u}')^2+\frac{(\boldsymbol{x}_i^T\boldsymbol{\beta}'+\boldsymbol{z}_i^T\boldsymbol{u}')\phi(\boldsymbol{x}_i^T\boldsymbol{\beta}'+\boldsymbol{z}_i^T\boldsymbol{u}')}{\Phi(\boldsymbol{x}_i^T\boldsymbol{\beta}'+\boldsymbol{z}_i^T\boldsymbol{u}')} & \text{if } y_i=1 \\ 1+(\boldsymbol{x}_i^T\boldsymbol{\beta}'+\boldsymbol{z}_i^T\boldsymbol{u}')^2-\frac{(\boldsymbol{x}_i^T\boldsymbol{\beta}'+\boldsymbol{z}_i^T\boldsymbol{u}')\phi(\boldsymbol{x}_i^T\boldsymbol{\beta}'+\boldsymbol{z}_i^T\boldsymbol{u}')}{1-\Phi(\boldsymbol{x}_i^T\boldsymbol{\beta}'+\boldsymbol{z}_i^T\boldsymbol{u}')} & \text{if } y_i=0 \end{cases}.$$

The above expectation can be written as,

$$E(v_i^2|\boldsymbol{\eta}',\boldsymbol{y}) = 1+\left(\boldsymbol{w}_i^{*T}\boldsymbol{\eta}'\right)^2-\frac{\left(\boldsymbol{w}_i^{*T}\boldsymbol{\eta}'\right)\phi\left(\boldsymbol{w}_i^{*T}\boldsymbol{\eta}'\right)}{1-\Phi\left(\boldsymbol{w}_i^{*T}\boldsymbol{\eta}'\right)}, \tag{42}$$

where $\boldsymbol{w}_i^* = c_i\boldsymbol{w}_i^T$ is the ith row of $\boldsymbol{W}^*$ defined in section 2. Also,

$$\begin{aligned} -\frac{\left(\boldsymbol{w}_i^{*T}\boldsymbol{\eta}'\right)\phi\left(\boldsymbol{w}_i^{*T}\boldsymbol{\eta}'\right)}{1-\Phi\left(\boldsymbol{w}_i^{*T}\boldsymbol{\eta}'\right)} &\le \begin{cases} \left|\frac{\left(\boldsymbol{w}_i^{*T}\boldsymbol{\eta}'\right)\phi\left(\boldsymbol{w}_i^{*T}\boldsymbol{\eta}'\right)}{1-\Phi\left(\boldsymbol{w}_i^{*T}\boldsymbol{\eta}'\right)}\right| & \text{if } \boldsymbol{w}_i^{*T}\boldsymbol{\eta}'\le 0 \\ 0 & \text{if } \boldsymbol{w}_i^{*T}\boldsymbol{\eta}'> 0 \end{cases} \\ &\le \sup_{u\in(-\infty,0]}\left|\frac{u\phi(u)}{1-\Phi(u)}\right| \equiv \Xi, \end{aligned} \tag{43}$$

where $\Xi \in (0,\infty)$.

We use $A_1,\ldots,A_{2^n}$ to denote all the subsets of $\mathbb{N}_n = \{1,2\ldots,n\}$. Following Roy and Hobert (2007), let

$$S_j = \left\{\boldsymbol{\eta}'\in\mathbb{R}^{p+q}\backslash\{\boldsymbol{0}\} : \boldsymbol{w}_i^T\boldsymbol{\eta}'\le 0 \text{ for all } i\in A_j \text{ and } \boldsymbol{w}_i^T\boldsymbol{\eta}'>0 \text{ for all } i\in\bar{A}_j\right\},$$

where $\bar{A}_j$ is the complement of A_j. As mentioned in Roy and Hobert (2007), the sets S_j's are disjoint, $\cup_{j=1}^{2^n}S_j = \mathbb{R}^{p+q}\backslash\{\boldsymbol{0}\}$ and some of the S_j's may be empty. For $j\in C\equiv\{i\in\mathbb{N}_{2^n} : S_i\neq\emptyset\}$, define

$$H_j(\boldsymbol{\eta}') = \frac{\sum_{i\in A_j}\left(\boldsymbol{w}_i^{*T}\boldsymbol{\eta}'\right)^2}{\sum_{i=1}^n\left(\boldsymbol{w}_i^{*T}\boldsymbol{\eta}'\right)^2} = \frac{\sum_{i\in A_j}\left(\boldsymbol{w}_i^{*T}\boldsymbol{\eta}'\right)^2}{\sum_{i\in A_j}\left(\boldsymbol{w}_i^{*T}\boldsymbol{\eta}'\right)^2+\sum_{i\in\bar{A}_j}\left(\boldsymbol{w}_i^{*T}\boldsymbol{\eta}'\right)^2}.$$

By (42), for $\boldsymbol{\eta}\in S_j$, $j\in C$, we have

$$\begin{aligned} &E\left[\sum_{i=1}^n v_i^2|\boldsymbol{\eta}',\boldsymbol{y}\right] \\ =&n+\sum_{i=1}^n\left(\boldsymbol{w}_i^{*T}\boldsymbol{\eta}'\right)^2-\sum_{i\in A_j}\frac{\left(\boldsymbol{w}_i^{*T}\boldsymbol{\eta}'\right)\phi\left(\boldsymbol{w}_i^{*T}\boldsymbol{\eta}'\right)}{1-\Phi\left(\boldsymbol{w}_i^{*T}\boldsymbol{\eta}'\right)}-\sum_{i\in\bar{A}_j}\frac{\left(\boldsymbol{w}_i^{*T}\boldsymbol{\eta}'\right)\phi\left(\boldsymbol{w}_i^{*T}\boldsymbol{\eta}'\right)}{1-\Phi\left(\boldsymbol{w}_i^{*T}\boldsymbol{\eta}'\right)} \\ =&n+\sum_{i=1}^n\left(\boldsymbol{w}_i^{*T}\boldsymbol{\eta}'\right)^2+\sum_{i\in A_j}\left|\frac{\left(\boldsymbol{w}_i^{*T}\boldsymbol{\eta}'\right)\phi\left(\boldsymbol{w}_i^{*T}\boldsymbol{\eta}'\right)}{1-\Phi\left(\boldsymbol{w}_i^{*T}\boldsymbol{\eta}'\right)}\right|-\sum_{i\in\bar{A}_j}\frac{\left(\boldsymbol{w}_i^{*T}\boldsymbol{\eta}'\right)\phi\left(\boldsymbol{w}_i^{*T}\boldsymbol{\eta}'\right)}{1-\Phi\left(\boldsymbol{w}_i^{*T}\boldsymbol{\eta}'\right)} \\ \le&n+\sum_{i=1}^n\left(\boldsymbol{w}_i^{*T}\boldsymbol{\eta}'\right)^2+n\Xi-\sum_{i\in\bar{A}_j}\left(\boldsymbol{w}_i^{*T}\boldsymbol{\eta}'\right)^2 \end{aligned}$$

$$=n\left(1+\Xi\right)+H_j\left(\boldsymbol{\eta}'\right)\sum_{i=1}^{n}\left(\boldsymbol{w}_i^{*T}\boldsymbol{\eta}'\right)^2,$$

where Ξ is defined in (43) and the inequality is due to the fact that $u\phi\left(u\right)/\left[1-\Phi\left(u\right)\right]\geq u^2$ for $u\geq 0$. Define $\lambda_j=\sup_{\boldsymbol{\eta}'\in S_j}\left\{H_j\left(\boldsymbol{\eta}'\right)\right\}\in[0,1]$ and

$$\lambda_0=\max_{j\in C}\lambda_j.$$

If $\boldsymbol{\eta}'=\mathbf{0}$, from (42), we have $E\left[\sum_{i=1}^{n}v_i^2|\boldsymbol{\eta}',\boldsymbol{y}\right]=n$. Thus, for all $\boldsymbol{\eta}'\in\mathbb{R}^{p+q}$,

$$E\left[\sum_{i=1}^{n}v_i^2|\boldsymbol{\eta}',\boldsymbol{y}\right]\leq\lambda_0\sum_{i=1}^{n}\left(\boldsymbol{x}_i^T\boldsymbol{\beta}'+\boldsymbol{z}_i^T\boldsymbol{u}'\right)^2+n\left(1+\Xi\right). \tag{44}$$

Since conditions A1 and A2 are in force, using the techniques in Roy and Hobert (2007), it can be shown that $\lambda_0<1$.

For $c\in(0,1/2)$, define

$$G_j(-c)=2^c\frac{\Gamma(q_j/2+a_j+c)}{\Gamma(q_j/2+a_j)}\text{ for }j=1,\ldots,r. \tag{45}$$

Since $\tau_j|\boldsymbol{\eta}',\boldsymbol{y}\sim\text{Gamma}(a_j+q_j/2,b_j+\boldsymbol{u}_j'^T\boldsymbol{u}_j'/2)$,

$$\begin{aligned}E\left[\tau_j^c|\boldsymbol{u}_j',\boldsymbol{y}\right]&=2^{-c}G_j\left(-c\right)\left[b_j+\frac{\boldsymbol{u}_j'^T\boldsymbol{u}_j'}{2}\right]^{-c}\\&\leq G_j\left(-c\right)\left[\left(2b_j\right)^{-c}I_{(0,\infty)}\left(b_j\right)+\left(\boldsymbol{u}_j'^T\boldsymbol{u}_j'\right)^{-c}I_{\{0\}}\left(b_j\right)\right].\end{aligned} \tag{46}$$

Recall that $A=\{j\in\{1,2,\ldots,r\}:b_j=0\}$. We consider two cases, namely, when A is empty and A is not empty.

Case 1: A is not empty.

Then using (44) and (46), from (41) we have

$$E\left[V\left(\boldsymbol{\eta}\right)|\boldsymbol{\eta}'\right]\leq\lambda_0\sum_{i=1}^{n}\left(\boldsymbol{x}_i^T\boldsymbol{\beta}'+\boldsymbol{z}_i^T\boldsymbol{u}'\right)^2+\delta_1\left(c\right)\sum_{j\in A}\left(\boldsymbol{u}_j'^T\boldsymbol{u}_j'\right)^{-c}+L_1\left(c\right),$$

where

$$\begin{aligned}\delta_1\left(c\right)&\equiv 2^{-c}\max_{j\in A}G_j\left(-c\right)\frac{\Gamma\left(q_j/2-c\right)}{\Gamma\left(q_j/2\right)},\\L_1\left(c\right)&\equiv n(1+\Xi)+p+q+2^{-c}\lambda_p^c\sum_{j=1}^{r}\frac{\Gamma\left(q_j/2-c\right)}{\Gamma\left(q_j/2\right)}\\&\quad+2^{-c}\sum_{j\notin A}G_j\left(-c\right)\frac{\Gamma\left(q_j/2-c\right)}{\Gamma\left(q_j/2\right)}\left(2b_j\right)^{-c}.\end{aligned} \tag{47}$$

By Román and Hobert (2012), there exists $c \in C_1 = (0, 1/2) \cap (0, -\max_{j \in A} a_j)$ such that $\delta_1(c) < 1$. Thus, taking $\rho_1 = \max(\lambda_0, \delta_1(c))$, and $L_1 = L_1(c)$, we have

$$E\left[V\left(\boldsymbol{\eta}\right)|\boldsymbol{\eta}'\right] \leq \rho_1 V\left(\boldsymbol{\eta}'\right) + L_1.$$

Case 2: A is empty.

In this case, the conditional expectation of τ_j^c can be bounded by a constant. Indeed from (46) we have

$$E\left[\tau_j^c|\boldsymbol{u}_j', \boldsymbol{y}\right] = 2^{-c} G_j\left(-c\right)\left[b_j + \frac{\boldsymbol{u}_j'^T \boldsymbol{u}_j'}{2}\right]^{-c} \leq G_j\left(-c\right)\left(2b_j\right)^{-c}.$$

Thus when A is empty, we have

$$E\left[V\left(\boldsymbol{\eta}\right)|\boldsymbol{\eta}'\right] \leq \lambda_0 \sum_{i=1}^n \left(\boldsymbol{x}_i^T \boldsymbol{\beta}' + \boldsymbol{z}_i^T \boldsymbol{u}'\right)^2 + L_1(c) \leq \lambda_0 V\left(\boldsymbol{\eta}'\right) + L_1(c).$$

Hence in both cases, (38) holds. We now show that $\boldsymbol{\eta}$-chain is a Feller chain on $\mathbb{R}^{p+q} \backslash \mathcal{N}$, which means that $K\left(\boldsymbol{\eta}, O\right)$ is a lower semi-continuous function on $\mathbb{R}^{p+q} \backslash \mathcal{N}$ for each fixed open set O. For a sequence $\{\boldsymbol{\eta}_m\}$ note that,

$$\begin{aligned}
\liminf_{m \to \infty} K\left(\boldsymbol{\eta}_m, O\right) &= \liminf_{m \to \infty} \int_O k\left(\boldsymbol{\eta}|\boldsymbol{\eta}_m\right) d\boldsymbol{\eta} \\
&= \liminf_{m \to \infty} \int_O \left[\int_{\mathbb{R}_+^r} \int_{\mathbb{R}^n} \pi(\boldsymbol{\eta}|\boldsymbol{v}, \boldsymbol{\tau}, \boldsymbol{y}) \pi(\boldsymbol{v}, \boldsymbol{\tau}|\boldsymbol{\eta}_m, \boldsymbol{y}) d\boldsymbol{v} d\boldsymbol{\tau}\right] d\boldsymbol{\eta} \\
&\geq \int_O \int_{\mathbb{R}_+^r} \int_{\mathbb{R}^n} \pi(\boldsymbol{\eta}|\boldsymbol{v}, \boldsymbol{\tau}, \boldsymbol{y}) \liminf_{m \to \infty} \pi(\boldsymbol{v}, \boldsymbol{\tau}|\boldsymbol{\eta}_m, \boldsymbol{y}) d\boldsymbol{v} d\boldsymbol{\tau} d\boldsymbol{\eta},
\end{aligned}$$

where the inequality follows from Fatou's lemma. Recall that $\pi(\boldsymbol{v}, \boldsymbol{\tau}|\boldsymbol{\eta}, \boldsymbol{y}) = \pi(\boldsymbol{v}|\boldsymbol{\eta}, \boldsymbol{y}) \pi(\boldsymbol{\tau}|\boldsymbol{\eta}, \boldsymbol{y})$. Note that, $\tau_j|\boldsymbol{\eta}', \boldsymbol{y} \sim \text{Gamma}(a_j + q_j/2, b_j + \boldsymbol{u}_j'^T \boldsymbol{u}_j'/2)$ and condition A3 holds. Thus, for all $\boldsymbol{\eta}' \in \mathbb{R}^{p+q} \backslash \mathcal{N}$ the conditional distribution of τ_j is a Gamma distribution with positive shape and scale parameters even if $b_j = 0$. Since both $\pi(\boldsymbol{v}|\boldsymbol{\eta}, \boldsymbol{y})$ and $\pi(\boldsymbol{\tau}|\boldsymbol{\eta}, \boldsymbol{y})$ are continuous functions in $\boldsymbol{\eta} \in \mathbb{R}^{p+q} \backslash \mathcal{N}$, if $\boldsymbol{\eta}_m \to \boldsymbol{\eta}$,

$$\begin{aligned}
\liminf_{m \to \infty} K\left(\boldsymbol{\eta}_m, O\right) &\geq \int_O \int_{\mathbb{R}_+^r} \int_{\mathbb{R}^n} \pi(\boldsymbol{\eta}|\boldsymbol{v}, \boldsymbol{\tau}, \boldsymbol{y}) \pi(\boldsymbol{v}, \boldsymbol{\tau}|\boldsymbol{\eta}, \boldsymbol{y}) d\boldsymbol{v} d\boldsymbol{\tau} d\boldsymbol{\eta} \\
&= K\left(\boldsymbol{\eta}, O\right).
\end{aligned}$$

Thus by Meyn and Tweedie (1993)(chap. 15), (38) implies the Markov chain $\tilde{\boldsymbol{\Psi}}$ is geometrically ergodic.

Next, we need to show that the original Markov chain $\boldsymbol{\Psi}$ is geometrically ergodic. The techniques of Lemma 12 in Román (2012) can be applied here for this purpose.

Let M and $\tilde{M}$ be the Mtfs of $\boldsymbol{\Psi}$ and $\tilde{\boldsymbol{\Psi}}$ respectively. Also, let M^m and $\tilde{M}^m$ be the corresponding m-step Mtfs, and $\mathsf{X} \equiv \mathbb{R}^{p+q}$, $\tilde{\mathsf{X}} \equiv \mathbb{R}^{p+q} \backslash \mathcal{N}$. Recall that $\mathscr{B}$

denotes the Borel σ-algebra of $\mathbb{R}^{p+q}$. Since the Lebesgue measure of $\mathcal{N}$ is 0, for any $\mathsf{x} \in \tilde{\mathsf{X}}$ and $\mathsf{B} \in \mathcal{B}_{\tilde{\mathsf{X}}} = \{\tilde{\mathsf{X}} \cap \mathsf{A} : \mathsf{A} \in \mathscr{B}\}$

$$\tilde{M}(\mathsf{x}, \mathsf{B}) = M(\mathsf{x}, \mathsf{B}).$$

Let μ and $\tilde{\mu}$ be the Lebesgue measures on X and $\tilde{\mathsf{X}}$ respectively. Then $\boldsymbol{\Psi}$ and $\tilde{\boldsymbol{\Psi}}$ are μ-irreducible and $\tilde{\mu}$-irreducible respectively. Also, μ and $\tilde{\mu}$ are the corresponding maximal irreducibility measures. These two Markov chains $\boldsymbol{\Psi}$ and $\tilde{\boldsymbol{\Psi}}$ are also aperiodic. According to Theorem 15.0.1 in Meyn and Tweedie (1993), there exists a ν-petite set $\mathsf{C} \in \mathcal{B}_{\tilde{\mathsf{X}}}$, $\rho_{\mathsf{C}} < 1$, $M_{\mathsf{C}} < \infty$, a number $\tilde{M}^{\infty}(\mathsf{C})$ such that $\tilde{\mu}(\mathsf{C}) > 0$ and

$$|\tilde{M}^m(\mathsf{x}, \mathsf{C}) - \tilde{M}^{\infty}(\mathsf{C})| < M_{\mathsf{C}} \rho_{\mathsf{C}}^m,$$

for all $\mathsf{x} \in \mathsf{C}$. Since the set C is a ν-petite set for $\tilde{\mathsf{X}}$, ν is a nontrivial measure on $\mathcal{B}_{\tilde{\mathsf{X}}}$ with,

$$\sum_{m=0}^{\infty} \tilde{M}^m(\mathsf{x}, \mathsf{B}) \tilde{a}(m) \geq \nu(\mathsf{B})$$

for all $\mathsf{x} \in \mathsf{C}$ and $\mathsf{B} \in \mathcal{B}_{\tilde{\mathsf{X}}}$, where $\tilde{a}(m)$ is a mass function on $\{0, 1, 2, \ldots, \}$.

Since $\tilde{M}^m(\mathsf{x}, \mathsf{B}) = M^m(\mathsf{x}, \mathsf{B})$ for any $\mathsf{x} \in \tilde{\mathsf{X}}$ and $\mathsf{B} \in \mathcal{B}_{\tilde{\mathsf{X}}}$, we have $M^m(\mathsf{x}, \mathsf{C}) = \tilde{M}^m(\mathsf{x}, \mathsf{C})$. So for all $x \in \mathsf{C}$

$$|M^m(\mathsf{x}, \mathsf{C}) - \tilde{M}^{\infty}(\mathsf{C})| < M_{\mathsf{C}} \rho_{\mathsf{C}}^m.$$

Also, since $\mu(\mathcal{N}) = 0$, we know that $\mu(\mathsf{C}) > 0$. It can be checked that C is also petite for the original Markov chain $\boldsymbol{\Psi}$. Thus from Theorem 15.0.1 of Meyn and Tweedie (1993), it follows that $\boldsymbol{\Psi}$ is geometrically ergodic. □

D. Proof of Theorem 3

Proof. As in Appendix C, we study the convergence properties of the $\boldsymbol{\eta}$-chain. Recall that $\mathcal{N} = \left\{\boldsymbol{\eta} \in \mathbb{R}^{p+q}; \prod_{j \in A} ||\boldsymbol{u}_j|| = 0\right\}$. When A is nonempty and $\boldsymbol{\eta} \in \mathcal{N}$, we define the conditional distribution of $\boldsymbol{\tau}$ given $\boldsymbol{\eta}, \boldsymbol{y}$ the same way as in Appendix C.

Consider the following drift function on $\mathbb{R}^{p+q} \backslash \mathcal{N}$,

$$V(\boldsymbol{\eta}) = \alpha \sum_{i=1}^{n} \left(\boldsymbol{x}_i^T \boldsymbol{\beta} + \boldsymbol{z}_i^T \boldsymbol{u}\right)^2 + \sum_{j=1}^{r} G_j(s) \left(\boldsymbol{u}_j^T \boldsymbol{u}_j\right)^s + \sum_{j=1}^{r} \left(\boldsymbol{u}_j^T \boldsymbol{u}_j\right)^{-c}.$$

where $G_j(\cdot)$ is defined in (45), α, $s \in \tilde{S} \equiv (0, 1] \cap (0, \tilde{s})$ for $\tilde{s}$ defined in Theorem 3, and $c \in C_1 = (0, 1/2) \cap (0, -\max_{j \in A} a_j)$ are positive constants to be chosen later. Under the assumption B3, $V(\boldsymbol{\eta}) : \mathbb{R}^{p+q} \backslash \mathcal{N} \to [0, \infty)$ is unbounded off compact sets (Since $\boldsymbol{W}$ is not a full rank matrix, the drift function considered in the proof of Theorem 2 is no more unbounded off compact sets.). We need to show that for any $\boldsymbol{\eta}, \boldsymbol{\eta}' \in \mathbb{R}^{p+q} \backslash \mathcal{N}$, there exists a constant $\rho_2 \in [0,1)$ and $L_2 > 0$ such that

$$E[V(\boldsymbol{\eta})|\boldsymbol{\eta}'] = E\{E[V(\boldsymbol{\eta}|\boldsymbol{v},\boldsymbol{\tau},\boldsymbol{y})]|\boldsymbol{\eta}',\boldsymbol{y}\} \le \rho_2 V(\boldsymbol{\eta}') + L_2. \tag{48}$$

First, we calculate the expectation of $V(\boldsymbol{\eta})$ with respect to the $\boldsymbol{\eta}$ conditional distribution given $\boldsymbol{v}, \boldsymbol{\tau}$ and $\boldsymbol{y}$. Same calculations as in the proof of Theorem 2 (see (39)) show that,

$$E\left[\sum_{i=1}^{n}\left(\boldsymbol{x}_i^T\boldsymbol{\beta} + \boldsymbol{z}_i^T\boldsymbol{u}\right)^2 |\boldsymbol{v},\boldsymbol{\tau},\boldsymbol{y}\right] \le p + q + \boldsymbol{v}^T\boldsymbol{v}. \tag{49}$$

For $s \in (0,1]$, by Jensen inequality,

$$E\left[\left(\boldsymbol{u}_j^T\boldsymbol{u}_j\right)^s |\boldsymbol{v},\boldsymbol{\tau},\boldsymbol{y}\right] \le \left[E\left(\boldsymbol{u}_j^T\boldsymbol{u}_j|\boldsymbol{v},\boldsymbol{\tau},\boldsymbol{y}\right)\right]^s. \tag{50}$$

Also, from (15) and (16) it follows that

$$E\left(\boldsymbol{u}_j^T\boldsymbol{u}_j|\boldsymbol{v},\boldsymbol{\tau},\boldsymbol{y}\right) = tr\left(R_j S(\boldsymbol{\tau})^{-1} R_j\right) + \left[E\left(R_j\boldsymbol{u}|\boldsymbol{v},\boldsymbol{\tau},\boldsymbol{y}\right)\right]^T \left[E\left(R_j\boldsymbol{u}|\boldsymbol{v},\boldsymbol{\tau},\boldsymbol{y}\right)\right], \tag{51}$$

where R_j is defined in Lemma 1. For the first part on the right hand side of (51), we have

$$\begin{aligned}
tr\left(R_j S(\boldsymbol{\tau})^{-1} R_j^T\right) &= tr\left[R_j\left(\boldsymbol{Z}^T\left(\boldsymbol{I} - P_X\right)\boldsymbol{Z}\right)^+ R_j^T\right] \\
&\quad + tr\left[R_j\left(\boldsymbol{I} - P_{\boldsymbol{Z}^T(\boldsymbol{I}-P_X)\boldsymbol{Z}}\right) R_j^T\right]\sum_{l=1}^{r}\tau_l^{-1} \\
&= \xi_j + \varsigma_j \sum_{l=1}^{r}\tau_l^{-1},
\end{aligned} \tag{52}$$

where $\xi_j = tr[R_j(\boldsymbol{Z}^T(\boldsymbol{I} - P_X)\boldsymbol{Z})^+ R_j^T]$ and $\varsigma_j = tr[R_j(\boldsymbol{I} - P_{\boldsymbol{Z}^T(\boldsymbol{I}-P_X)\boldsymbol{Z}})R_j^T]$. For the second part, we have

$$\begin{aligned}
&\left[E\left(R_j\boldsymbol{u}|\boldsymbol{v},\boldsymbol{\tau},\boldsymbol{y}\right)\right]^T\left[E\left(R_j\boldsymbol{u}|\boldsymbol{v},\boldsymbol{\tau},\boldsymbol{y}\right)\right] \\
&= \boldsymbol{v}^T\left(\boldsymbol{I} - P_X\right)\boldsymbol{Z}S(\boldsymbol{\tau})^{-1}R_j^T R_j S(\boldsymbol{\tau})^{-1}\boldsymbol{Z}^T\left(\boldsymbol{I} - P_X\right)\boldsymbol{v} \\
&\le \boldsymbol{v}^T\left(\boldsymbol{I} - P_X\right)\boldsymbol{Z}S(\boldsymbol{\tau})^{-1}S(\boldsymbol{\tau})^{-1}\boldsymbol{Z}^T\left(\boldsymbol{I} - P_X\right)\boldsymbol{v} \\
&= \left\|S(\boldsymbol{\tau})^{-1}\boldsymbol{Z}^T\left(\boldsymbol{I} - P_X\right)\boldsymbol{v}\right\|^2 \\
&\le \left(\hat{\varphi}\sum_{i=1}^{n}|v_i|\right)^2 \le \hat{\varphi}^2 n \sum_{i=1}^{n} v_i^2,
\end{aligned} \tag{53}$$

where the second inequality follows from Lemma 2 given in Appendix B. Combining (52) and (53), from (51) we have

$$\left[E\left(\boldsymbol{u}_j^T\boldsymbol{u}_j|\boldsymbol{v},\boldsymbol{\tau},\boldsymbol{y}\right)\right]^s \le \left[\xi_j + \varsigma_j\sum_{l=1}^{r}\tau_l^{-1} + \hat{\varphi}^2 n\sum_{i=1}^{n}v_i^2\right]^s$$

$$\leq \xi_j^s + \varsigma_j^s \sum_{l=1}^{r} \tau_l^{-s} + \hat{\varphi}^{2s} n^s \sum_{i=1}^{n} v_i^{2s}.$$

Note that, if $v_i^2 \leq 1$, then $v_i^{2s} \leq 1$, and if $v_i^{2s} > 1$, then $v_i^{2s} < v_i^2$. So $v_i^{2s} \leq 1 + v_i^2$. Thus,

$$\left[E\left(\boldsymbol{u}_j^T \boldsymbol{u}_j | \boldsymbol{v}, \boldsymbol{\tau}, \boldsymbol{y}\right)\right]^s \leq \varsigma_j^s \sum_{l=1}^{r} \tau_l^{-s} + \hat{\varphi}^{2s} n^s \sum_{i=1}^{n} v_i^2 + \hat{\varphi}^{2s} n^{1+s} + \xi_j^s. \tag{54}$$

Also recall from (40) that we also have,

$$E\left[\left(\boldsymbol{u}_j^T \boldsymbol{u}_j\right)^{-c} | \boldsymbol{v}, \boldsymbol{\tau}, \boldsymbol{y}\right] \leq 2^{-c} \frac{\Gamma\left(q_j/2 - c\right)}{\Gamma\left(q_j/2\right)} \left[\lambda_p^c + \tau_j^c\right].$$

Combining (40) , (49), (50) and (54) from (48) we have

$$\begin{aligned} E\left[V\left(\boldsymbol{\eta}\right) | \boldsymbol{v}, \boldsymbol{\tau}, \boldsymbol{y}\right] \leq &\left(\alpha + \delta_2\left(s\right)\right) \sum_{i=1}^{n} v_i^2 + \delta_3\left(s\right) \sum_{j=1}^{r} \tau_j^{-s} \\ &+ 2^{-c} \sum_{j=1}^{r} \frac{\Gamma\left(q_j/2 - c\right)}{\Gamma\left(q_j/2\right)} \tau_j^c + \kappa_1\left(\alpha, s, c\right), \end{aligned} \tag{55}$$

where

$$\delta_2\left(s\right) = \hat{\varphi}^{2s} n^s \sum_{j=1}^{r} G_j\left(s\right),$$

$$\delta_3\left(s\right) = \sum_{j=1}^{r} G_j\left(s\right) \varsigma_j^s, \text{ and}$$

$$\kappa_1\left(\alpha, s, c\right) = \alpha\left(p + q\right) + \sum_{j=1}^{r} G_j\left(s\right)\left(\hat{\varphi}^{2s} n^{1+s} + \xi_j^s\right) + 2^{-c} \lambda_p^c \sum_{j=1}^{r} \frac{\Gamma\left(q_j/2 - c\right)}{\Gamma\left(q_j/2\right)}.$$

Next we calculate the outer expectation in (48), that is, the expectation with respect to the conditional distribution of $\boldsymbol{v}$ and $\boldsymbol{\tau}$ given $\boldsymbol{\eta}'$ and $\boldsymbol{y}$.

When calculating the upper bound of $E(\sum_{i=1}^n v_i^2 | \boldsymbol{\eta}', \boldsymbol{y})$, we need to take into account the fact that $\boldsymbol{W}$ is not a full rank matrix in the current setting. But, $E(\sum_{i=1}^n v_i^2 | \boldsymbol{\eta}', \boldsymbol{y})$ can be written as,

$$E\left[\sum_{i=1}^{n} v_i^2 | \boldsymbol{\eta}', \boldsymbol{y}\right] = n + \sum_{i=1}^{n} \left(\tilde{\boldsymbol{w}}_i^{*T} \tilde{\boldsymbol{\eta}}'\right)^2 - \sum_{i=1}^{n} \frac{\left(\tilde{\boldsymbol{w}}_i^{*T} \tilde{\boldsymbol{\eta}}'\right) \phi\left(\tilde{\boldsymbol{w}}_i^{*T} \tilde{\boldsymbol{\eta}}'\right)}{1 - \Phi\left(\tilde{\boldsymbol{w}}_i^{*T} \tilde{\boldsymbol{\eta}}'\right)}.$$

where $\tilde{\boldsymbol{w}}_i^*$'s are defined in section 2.

Since the condition B3 is in force, we know that $\tilde{\boldsymbol{W}}$ is a full rank matrix. Then the same techniques (see (44)) as in the proof of Theorem 2 can be used to show that there exists $\lambda_0 \in [0, 1)$ such that

$$E\left[\sum_{i=1}^{n} v_i^2 | \boldsymbol{\eta}', \boldsymbol{y}\right] \leq \lambda_0 \sum_{i=1}^{n} \left(\tilde{\boldsymbol{w}}_i^{*T} \tilde{\boldsymbol{\eta}}'\right)^2 + n\left(1 + \Xi\right)$$

$$= \lambda_0 \sum_{i=1}^{n} \left(\boldsymbol{w}_i^T \boldsymbol{\eta}'\right)^2 + n\left(1+\Xi\right)$$
$$= \lambda_0 \sum_{i=1}^{n} \left(\boldsymbol{x}_i^T \boldsymbol{\beta}' + \boldsymbol{z}_i^T \boldsymbol{u}'\right)^2 + n\left(1+\Xi\right). \qquad (56)$$

For $s \in \tilde{S}$, we have

$$E\left[\tau_j^{-s} | \boldsymbol{\eta}', \boldsymbol{y}\right] = 2^s G_j(s) \left(b_j + \frac{\boldsymbol{u}_j'^T \boldsymbol{u}_j'}{2}\right)^s \le G_j(s) \left(\boldsymbol{u}_j'^T \boldsymbol{u}_j'\right)^s + 2^s G_j(s) b_j^s. \quad (57)$$

Also for $c \in C_1$, as in (46), we have

$$E\left[\tau_j^{c} | \boldsymbol{\eta}', \boldsymbol{y}\right] = 2^{-c} G_j(-c) \left[b_j + \frac{\boldsymbol{u}_j'^T \boldsymbol{u}_j'}{2}\right]^{-c}$$
$$\le G_j(-c) \left[(2b_j)^{-c} I_{(0,\infty)}(b_j) + \left(\boldsymbol{u}_j'^T \boldsymbol{u}_j'\right)^{-c} I_{\{0\}}(b_j)\right].$$

As in the proof of Theorem 2, we consider two cases, namely A is empty and A is not empty.

Case 1: A is not empty.

Using (46), (56) and (57) from (55), we have

$$E\left[V(\boldsymbol{\eta}) | \boldsymbol{\eta}'\right] = \alpha\lambda_0 \left(1 + \frac{\delta_2(s)}{\alpha}\right) \sum_{i=1}^{n} \left(\boldsymbol{x}_i^T \boldsymbol{\beta}' + \boldsymbol{z}_i^T \boldsymbol{u}'\right)^2$$
$$+ \delta_3(s) \sum_{j=1}^{r} G_j(s) \left(\boldsymbol{u}_j'^T \boldsymbol{u}_j'\right)^s + \delta_1(c) \sum_{j\in A} \left(\boldsymbol{u}_j'^T \boldsymbol{u}_j'\right)^{-c} + L_2(\alpha, s, c), \qquad (58)$$

where

$$L_2(\alpha, s, c) = \kappa_1(\alpha, s, c) + n(1+\Xi)(\alpha + \delta_2(s)) + \delta_3(s) 2^s \sum_{j=1}^{r} G_j(s) b_j^s$$
$$+ 2^{-c} \sum_{j\notin A} \frac{\Gamma(q_j/2 - c)}{\Gamma(q_j/2)} G_j(-c) (2b_j)^{-c},$$

and $\delta_1(c)$ is defined as in (47).

We know that for $c \in C_1$, $\delta_1(c) < 1$ as in Theorem 2. Since condition 2 of Theorem 3 holds, we have $\delta_3(s) < 1$. For a fixed s, $\lambda_0 (1 + \delta_2(s)/\alpha) < 1$ iff $\alpha > \lambda_0 \delta_2(s)/(1-\lambda_0)$. So there exists a ρ_2 such that

$$\rho_2 \equiv \rho_2(\alpha, s, c) = \max\left\{\lambda_0 (1 + \delta_2(s)/\alpha), \delta_3(s), \delta_1(c)\right\} < 1$$

and $L_2 \equiv L_2(\alpha, s, c) > 0$ such that (48) holds.

Case 2: A is empty.

In this case, the conditional expectation of τ_j^c can be bounded by a constant. Thus we have

$$E\left[V\left(\boldsymbol{\eta}\right)|\boldsymbol{\eta}'\right] = \alpha\lambda_0\left(1+\frac{\delta_2\left(s\right)}{\alpha}\right)\sum_{i=1}^{n}\left(\boldsymbol{x}_i^T\boldsymbol{\beta}'+\boldsymbol{z}_i^T\boldsymbol{u}'\right)^2$$
$$+\,\delta_3\left(s\right)\sum_{j=1}^{r}G_j\left(s\right)\left(\boldsymbol{u}_j'^T\boldsymbol{u}_j'\right)^s+L_2\left(\alpha,s,c\right).$$

As in case 1, it follows that (48) holds.

Since $\boldsymbol{\eta}$-chain is a Feller chain on $\mathbb{R}^{p+q}\backslash\mathcal{N}$, and $V(\boldsymbol{\eta})$ is unbounded off compact sets on $\mathbb{R}^{p+q}\backslash\mathcal{N}$, the $\boldsymbol{\eta}$-chain is geometrically ergodic on $\mathbb{R}^{p+q}\backslash\mathcal{N}$. Using the same techniques as in Appendix C, it can be shown that the original $\{\boldsymbol{\eta}^{(m)}\}_{m=0}^{\infty}$ Markov chain defined on $\mathbb{R}^{p+q}$ is also geometrically ergodic. □

Acknowledgment

The authors thank two anonymous reviewers and an anonymous associate editor for several helpful comments and suggestions that led to a substantially improved revision of the paper.

References

Albert, J. H. and Chib, S. (1993). Bayesian analysis of binary and polychotomous response data. *Journal of the American statistical Association* **88** 669–679. MR1224394

Asmussen, S. and Glynn, P. W. (2011). A new proof of convergence of MCMC via the ergodic theorem. *Statistics and Probability Letters* **81** 1482–1485. MR2818658

Baragatti, M. (2011). Bayesian variable selection for probit mixed models applied to gene selection. *Bayesian Analysis* **6** 209–229. MR2806242

Breslow, N. E. and Clayton, D. G. (1993). Approximate inference in generalized linear mixed models. *Journal of the American statistical Association* **88** 9–25. MR1397972

Chakraborty, S. and Khare, K. (2017). Convergence properties of Gibbs samplers for Bayesian probit regression with proper priors. *Electronic Journal of Statistics* **11** 177-210. MR3604022

Chen, M.-H. and Shao, Q.-M. (2001). Propriety of posterior distribution for dichotomous quantal response models. *Proceedings of the American Mathematical Society* **129** 293–302. MR1694452

Chen, M.-H., Shao, Q.-M. and Xu, D. (2002). Necessary and sufficient conditions on the properiety of posterior distributions for generalized linear mixed models. *Sankhyā: The Indian Journal of Statistics, Series A* **64** 57–85. MR1968375

Choi, H. M. and Hobert, J. P. (2013). The Polya-Gamma Gibbs sampler for Bayesian logistic regression is uniformly ergodic. *Electronic Journal of Statistics* **7** 2054–2064. MR3091616

Flegal, J. M. and Jones, G. L. (2010). Batch means and spectral variance estimators in Markov chain Monte Carlo. *The Annals of Statistics* **38** 1034–1070. MR2604704

Hobert, J. P. and Marchev, D. (2008). A theoretical comparison of the data augmentation, marginal augmentation and PX-DA algorithms. *The Annals of Statistics* **36** 532–554. MR2396806

Johnson, A. A. and Jones, G. L. (2010). Gibbs sampling for a Bayesian hierarchical general linear model. *Electronic Journal of Statistics* **4** 313–333. MR2645487

Jones, G. L. and Hobert, J. P. (2001). Honest exploration of intractable probability distributions via Markov chain Monte Carlo. *Statistical Science* **16** 312–334. MR1888447

Jones, G. L. and Hobert, J. P. (2004). Sufficient burn-in for Gibbs samplers for a hierarchical random effects model. *Annals of statistics* **32** 784–817. MR2060178

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

Liu, J. S. and Wu, Y. N. (1999). Parameter expansion for data augmentation. *Journal of the American Statistical Association* **94** 1264–1274. MR1731488

McCulloch, C. E., Searle, S. R. and Neuhaus, J. M. (2011). *Generalized, Linear, and Mixed Models*. John Wiley & Sons. MR2431553

Meng, X.-L. and Van Dyk, D. A. (1999). Seeking efficient data augmentation schemes via conditional and marginal augmentation. *Biometrika* **86** 301–320. MR1705351

Meyn, S. P. and Tweedie, R. L. (1993). *Markov chains and stochastic stability*. Springer. MR1287609

Polson, N. G., Scott, J. G. and Windle, J. (2013). Bayesian inference for logistic models using Pólya–Gamma latent variables. *Journal of the American statistical Association* **108** 1339–1349. MR3174712

Roberts, G. O., Rosenthal, J. S. (1997). Geometric ergodicity and hybrid Markov chains. *Electron. Comm. Probab* **2** 13–25. MR1448322

Roberts, G. O. and Rosenthal, J. S. (2001). Markov chains and de-initializing processes. *Scandinavian Journal of Statistics* **28** 489–504. MR1858413

Román, J. C. (2012). Convergence analysis of block Gibbs samplers for Bayesian general linear mixed models PhD thesis, University of Florida. MR3279056

Román, J. C. and Hobert, J. P. (2012). Convergence analysis of the Gibbs sampler for Bayesian general linear mixed models with improper priors. *The Annals of Statistics* **40** 2823–2849. MR3097961

Román, J. C. and Hobert, J. P. (2015). Geometric ergodicity of Gibbs samplers for Bayesian general linear mixed models with proper priors. *Linear*

Algebra and its Applications **473** 54–77. MR3338325

Roy, V. (2012a). Spectral analytic comparisons for data augmentation. *Stat. and Prob. Letters* **82** 103-108. MR2863030

Roy, V. (2012b). Convergence rates for MCMC algorithms for a robust Bayesian binary regression model. *Electronic Journal of Statistics* **6** 2463–2485. MR3020272

Roy, V. (2014). Efficient estimation of the link function parameter in a robust Bayesian binary regression model. *Computational Statistics & Data Analysis* **73** 87–102. MR3147976

Roy, V. and Hobert, J. P. (2007). Convergence rates and asymptotic standard errors for Markov chain Monte Carlo algorithms for Bayesian probit regression. *Journal of the Royal Statistical Society: Series B* **69** 607–623. MR2370071

Tan, A. and Hobert, J. P. (2009). Block Gibbs sampling for Bayesian random effects models with improper priors: Convergence and regeneration. *Journal of Computational and Graphical Statistics* **18** 861–878. MR2598033

Van Dyk, D. A. and Meng, X.-L. (2001). The Art of Data Augmentation (with discussion). *Journal of Computational and Graphical Statistics* **10** 1–50. MR1936358

Wang, X. and Roy, V. (2018). Geometric ergodicity of Pólya-Gamma Gibbs sampler for Bayesian logistic regression with a flat prior. *Electronic Journal of Statistics* **12** 3295–3311. MR3861283