

# Block Gibbs samplers for logistic mixed models: Convergence properties and a comparison with full Gibbs samplers

Yalin Rao

*Department of Mathematics and Statistics, University of Massachusetts Amherst*  
*e-mail:* [rao@math.umass.edu](mailto:rao@math.umass.edu)

and

Vivekananda Roy

*Department of Statistics, Iowa State University*  
*e-mail:* [vroy@iastate.edu](mailto:vroy@iastate.edu)

**Abstract:** The logistic linear mixed model (LLMM) is one of the most widely used statistical models. Generally, Markov chain Monte Carlo algorithms are used to explore the posterior densities associated with the Bayesian LLMMs. Polson, Scott and Windle's (2013) Pólya-Gamma data augmentation (DA) technique can be used to construct full Gibbs (FG) samplers for the LLMMs. Here, we develop efficient block Gibbs (BG) samplers for Bayesian LLMMs using the Pólya-Gamma DA method. We compare the FG and BG samplers in the context of a real data example, as the correlation between the fixed effects and the random effects changes as well as when the dimensions of the design matrices vary. These numerical examples demonstrate superior performance of the BG samplers over the FG samplers. We also derive conditions guaranteeing geometric ergodicity of the BG Markov chain when the popular improper uniform prior is assigned on the regression coefficients, and proper or improper priors are placed on the variance parameters of the random effects. This theoretical result has important practical implications as it justifies the use of asymptotically valid Monte Carlo standard errors for Markov chain based estimates of the posterior quantities.

**MSC2020 subject classifications:** Primary 60J05; secondary 62F15.

**Keywords and phrases:** Data augmentation, drift condition, geometric ergodicity, GLMM, Markov chain CLT, MCMC, standard errors.

Received February 2021.

## 1. Introduction

The logistic linear mixed model (LLMM) is an extensively used generalized linear mixed model for binary response data. Let  $(Y_1, Y_2, \dots, Y_n)$  denote the vector of Bernoulli responses. Let  $X$  and  $Z$  be the  $n \times p$  and  $n \times q$  known design matrices corresponding to the fixed effects and the random effects, respectively. Suppose  $x_i^\top$  and  $z_i^\top$  indicate the  $i^{\text{th}}$  row of  $X$  and  $Z$ , respectively, for  $i = 1, \dots, n$ . Let  $\beta \in \mathbb{R}^p$  be the regression coefficients vector and  $u \in \mathbb{R}^q$  be the random effects

vector. In general, a generalized linear mixed model (GLMM) can be built with a link function that connects the probability that the response variable  $Y$  equals 1 (that is, the expectation of  $Y$ ) with  $X$  and  $Z$ . For the LLMM,  $P(Y_i = 1) = F(x_i^\top \beta + z_i^\top u)$ , where  $F$  indicates the cumulative distribution function for the standard logistic random variable, that is,  $F(t) = e^t / (1 + e^t)$ ,  $t \in \mathbb{R}$ . Also, assume that we can divide  $u$  into  $r$  independent random effects. Let  $u = (u_1^\top, \dots, u_r^\top)^\top$ , where  $u_j$  is a  $q_j \times 1$  vector with  $q_j > 0$ ,  $j = 1, \dots, r$ , and  $\sum_{j=1}^r q_j = q$ . Assume,  $u_j \stackrel{\text{ind}}{\sim} N(0, (1/\tau_j)I_{q_j})$ , where  $\tau_j > 0$ , for  $j = 1, \dots, r$ . Let  $\tau = (\tau_1, \dots, \tau_r)$ . Thus the data model for the LLMM is

$$\begin{aligned} Y_i | \beta, u, \tau &\stackrel{\text{ind}}{\sim} \text{Ber}(F(x_i^\top \beta + z_i^\top u)) \quad \text{for } i = 1, \dots, n, \\ u_j | \tau_j &\stackrel{\text{ind}}{\sim} N(0, (1/\tau_j)I_{q_j}), \quad j = 1, \dots, r. \end{aligned} \quad (1)$$

Let  $y = (y_1, y_2, \dots, y_n)$  denote the vector of observed Bernoulli responses. Then, the likelihood function for  $(\beta, \tau)$  is

$$L(\beta, \tau | y) = \int_{\mathbb{R}^q} \prod_{i=1}^n [F(x_i^\top \beta + z_i^\top u)]^{y_i} [1 - F(x_i^\top \beta + z_i^\top u)]^{1-y_i} \phi_q(u; 0, D(\tau)^{-1}) du, \quad (2)$$

where  $D(\tau)^{-1} = \oplus_{j=1}^r \frac{1}{\tau_j} I_{q_j}$ , and  $\oplus$  indicates the direct sum. Here,  $\phi_q(u; 0, D(\tau)^{-1})$  denotes the probability density function of the  $q$ -dimensional normal distribution with mean vector 0, covariance matrix  $D(\tau)^{-1}$ , and evaluated at  $u$ .

When working in a Bayesian framework, one must specify priors for  $\beta$  and  $\tau$ . Here, we consider the prior for  $\beta$  as given by

$$\pi(\beta) \propto \exp \left[ -\frac{1}{2} (\beta - \mu_0)^\top Q (\beta - \mu_0) \right], \quad (3)$$

where  $\mu_0 \in \mathbb{R}^p$  and  $Q$  is a  $p \times p$  positive definite matrix (proper normal prior) or a zero matrix (improper uniform prior). Thus, if  $Q = 0$ , then  $\pi(\beta) \propto 1$ . The prior for  $\tau_j$  is

$$\pi(\tau_j) \propto \tau_j^{a_j-1} e^{-\tau_j b_j}, \quad j = 1, \dots, r, \quad (4)$$

which may be proper or improper depending on the values of  $a_j$  and  $b_j$ . Finally, we assume that  $\beta$  and  $\tau$  are apriori independent, and all the  $\tau_j$ s are also apriori independent. Hence, the joint posterior density for  $(\beta, \tau)$  is

$$\pi(\beta, \tau | y) = \frac{1}{c(y)} L(\beta, \tau | y) \pi(\beta) \pi(\tau), \quad (5)$$

where  $c(y) = \int_{\mathbb{R}_+^r} \int_{\mathbb{R}^p} L(\beta, \tau | y) \pi(\beta) \pi(\tau) d\beta d\tau$  is the marginal pmf of  $y$ . If  $c(y)$  is finite, then the posterior density is proper. Since we consider both proper and improper priors on  $(\beta, \tau)$ , if improper priors are used, then  $c(y)$  is not necessarily finite. Conditions for the posterior propriety of nonlinear mixed models with

general link functions are given in Chen, Shao and Xu (2002) and Wang and Roy (2018a). Theorem 1, given in Section 4 of this paper, provides easily verifiable sufficient conditions for propriety of (5) when  $\pi(\beta) \propto 1$ .

Since the likelihood function  $L(\beta, \tau | y)$  is not available in a closed form, the posterior density for  $(\beta, \tau)$  is not tractable for any choice of priors on these parameters. Markov chain Monte Carlo (MCMC) algorithms can be used to explore the posterior density  $\pi(\beta, \tau | y)$ . Even in the absence of the random effects, MCMC algorithms are generally used for exploring the posterior densities corresponding to the basic logistic model or other generalized linear models (GLMs). Using the data augmentation (DA) technique (van Dyk and Meng, 2001), in a highly cited paper, Albert and Chib (1993) constructed a Gibbs sampler for GLMs with the probit link. Since then there have been several attempts to construct such a DA Gibbs sampler for the logistic model (see e.g. Holmes and Held (2006) and Frühwirth-Schnatter and Frühwirth (2010)). Recently, using the Pólya-Gamma (PG) latent variables, Polson, Scott and Windle (2013) (denoted as PS&W hereafter) have proposed an efficient DA Gibbs sampler for the Bayesian logistic models. A random variable  $\omega$  follows a PG distribution with parameters  $a > 0$ ,  $b \geq 0$ , if  $\omega \stackrel{d}{=} (1/(2\pi^2)) \sum_{i=1}^{\infty} g_i / [(i - 1/2)^2 + b^2/(4\pi^2)]$ , where  $g_i \stackrel{iid}{\sim} \text{Gamma}(a, 1)$ . We denote this as  $\omega \sim \text{PG}(a, b)$ . PS&W's DA technique can be extended to construct a Gibbs sampler for the LLMMs. Indeed, with PG latent variables  $\omega = (\omega_1, \omega_2, \dots, \omega_n)$ , one can construct a joint posterior density  $\pi(\beta, u, \omega, \tau | y)$  (details are given in Section 2) such that

$$\int_{\mathbb{R}^q} \int_{\mathbb{R}_+^n} \pi(\beta, u, \omega, \tau | y) d\omega du = \pi(\beta, \tau | y), \quad (6)$$

where  $\mathbb{R}_+ = (0, \infty)$ , and  $\pi(\beta, \tau | y)$  is given in (5). Using the conditional distributions of the joint density  $\pi(\beta, u, \omega, \tau | y)$ , a full Gibbs sampler can be formed (The details for this Gibbs sampler are given in Section 2.1.). It is known that blocking parameters can improve the performance of the Gibbs sampler in terms of reducing its operator norm (Liu, Wong and Kong, 1994). In general, when one or more variables are correlated, sampling them jointly can improve the efficiency of MCMC algorithms (Roberts and Sahu, 1997; Chib and Ramamurthy, 2010; Turek et al., 2017). On the other hand, blocking may result in complex conditional distributions that are not easy to sample from. For the LLMMs, it turns out that an efficient two-block Gibbs sampler can be constructed by using the two blocks,  $\eta \equiv (\beta^\top, u^\top)^\top$  and  $(\omega, \tau)$ . We derive this block Gibbs sampler in Section 2.2. Using numerical examples, we show that blocking can lead to great gains in efficiency in Monte Carlo estimation for the LLMMs.

The block Gibbs Markov chain is Harris ergodic. Thus, the sample (time) averages are consistent estimators of the means with respect to the posterior density (5). On the other hand, in practice, it is important to ascertain the errors associated with the Monte Carlo estimates. A valid standard error for the Monte Carlo estimate can be formed if a central limit theorem (CLT) is available for the time average estimator (Jones and Hobert, 2001). Establishing geometric ergodicity (GE) of the underlying Markov chain is the most standard method for

guaranteeing CLTs hold for MCMC estimators, and is also used for consistently estimating the asymptotic variance in the CLT (Vats, Flegal and Jones (2018), Vats, Flegal and Jones (2019)). GE of Gibbs samplers for probit and logistic GLMs under different priors has been established in the literature (Roy and Hobert, 2007; Choi and Hobert, 2013; Chakraborty and Khare, 2017; Wang and Roy, 2018b). Also, GE of Gibbs samplers for probit mixed models and normal linear mixed models under improper priors on the regression coefficients and variance components is considered in Wang and Roy (2018a) and Román and Hobert (2012), respectively. From Roy (2012a) it follows that Wang and Roy's (2018a) GE result also holds for parameter expansion for DA algorithms for the probit mixed models. Wang and Roy (2018c) prove uniform ergodicity of a Gibbs sampler for the LLMMs under a proper normal prior on  $\beta$  and a truncated proper prior on  $\tau$ . Wang and Roy (2018c) establish a *minorization condition* to prove uniform ergodicity of the Gibbs sampler (see e.g. Roberts and Rosenthal, 2004, Theorem 8). Wang and Roy's (2018c) analysis of the Gibbs Markov chain puts a constraint on the support of the posterior density (5), in that the parameter  $\tau$  is bounded away from zero. Here, using a *drift condition*, we establish geometric convergence rates for the block Gibbs sampler, in the case when the popular improper uniform prior is assigned on  $\beta$  and the commonly used proper gamma priors or the improper power priors are assigned on  $\tau$ . In contrast to Wang and Roy (2018c), our result does not put any restriction on the support of the variance components, that is, the common support of  $\tau_j$ 's is the entire positive real line.

The rest of the article is organized as follows. In Section 2, we provide details on PG data augmentation and construct the full and the block Gibbs samplers. Section 3 contains numerical examples. These examples are used to compare the performance of the block and the full Gibbs samplers. In Section 4, we consider geometric convergence of the block Gibbs sampler under improper priors. Some concluding remarks are provided in Section 5. Finally, several theoretical results along with proofs of the results appear in the appendices.

## 2. Gibbs samplers

In this section, using the PG variables, we discuss DA for the LLMMs, and construct Gibbs samplers for (5). Following (2) and (5), the joint posterior density for  $(\beta, \tau)$  is

$$\pi(\beta, \tau | y) = \frac{\pi(\beta)\pi(\tau)}{c(y)} \int_{\mathbb{R}^q} \prod_{i=1}^n \frac{\exp\{y_i(x_i^\top \beta + z_i^\top u)\}}{1 + \exp(x_i^\top \beta + z_i^\top u)} \phi_q(u; 0, D(\tau)^{-1}) du.$$

By Theorem 1 in Polson, Scott and Windle (2013)

$$\begin{aligned} \pi(\beta, \tau | y) &= \int_{\mathbb{R}^q} \int_{\mathbb{R}_+^n} \left[ \prod_{i=1}^n \frac{\exp\{k_i(x_i^\top \beta + z_i^\top u) - \omega_i(x_i^\top \beta + z_i^\top u)^2/2\}}{2} p(\omega_i) \right] d\omega \\ &\quad \times \phi_q(u; 0, D(\tau)^{-1}) du \times \frac{\pi(\beta)\pi(\tau)}{c(y)}, \end{aligned} \quad (7)$$

where  $\omega = (\omega_1, \omega_2, \dots, \omega_n)$ ,  $k_i = y_i - 1/2$ ,  $i = 1, \dots, n$  and  $p(\omega_i)$  is the pdf of  $\omega_i \sim \text{PG}(1, 0)$  given by,

$$p(\omega_i) = \sum_{\ell=0}^{\infty} (-1)^\ell \frac{(2\ell+1)}{\sqrt{2\pi\omega_i^3}} \exp\left[-\frac{(2\ell+1)^2}{8\omega_i}\right], \omega_i > 0. \quad (8)$$

We now define the joint posterior density of  $\beta, u, \omega, \tau$  given  $y$  mentioned in (6) as

$$\begin{aligned} \pi(\beta, u, \omega, \tau \mid y) &\propto \left[ \prod_{i=1}^n \exp\{k_i(x_i^\top \beta + z_i^\top u) - \omega_i(x_i^\top \beta + z_i^\top u)^2/2\} p(\omega_i) \right] \pi(\beta) \pi(\tau) \\ &\quad \times \phi_q(u; 0, D(\tau)^{-1}) \\ &= \left[ \prod_{i=1}^n \exp\{k_i(x_i^\top \beta + z_i^\top u) - \omega_i(x_i^\top \beta + z_i^\top u)^2/2\} p(\omega_i) \right] \\ &\quad \times \phi_q(u; 0, D(\tau)^{-1}) \times \exp\left[-\frac{1}{2}(\beta - \mu_0)^\top Q(\beta - \mu_0)\right] \\ &\quad \times \prod_{j=1}^r \tau_j^{a_j-1} \exp(-b_j \tau_j), \end{aligned} \quad (9)$$

where (9) follows from the priors on  $\beta$  and  $\tau$  given in (3) and (4), respectively.

### 2.1. A full Gibbs sampler

Let  $\Omega(\omega)$  be the  $n \times n$  diagonal matrix with the  $i^{\text{th}}$  diagonal element  $\omega_i$ . Let  $M = [XZ]$ , and  $m_i^\top$  indicates the  $i^{\text{th}}$  row of  $M$  for  $i = 1, \dots, n$ . Let  $\kappa = (k_1, k_2, \dots, k_n)^\top$ . We begin with deriving the conditional densities required for the full Gibbs (FG) sampler. Based on (9), the conditional density of  $\beta$  given  $u, \omega, \tau, y$  is

$$\begin{aligned} \pi(\beta \mid u, \omega, \tau, y) &\propto \prod_{i=1}^n \exp\left[k_i x_i^\top \beta - \omega_i(x_i^\top \beta)^2/2 - \omega_i x_i^\top \beta z_i^\top u\right] \\ &\quad \times \exp\left[-\frac{1}{2}(\beta - \mu_0)^\top Q(\beta - \mu_0)\right] \\ &\propto \exp\left[-\frac{1}{2}\beta^\top (X^\top \Omega(\omega)X + Q)\beta\right. \\ &\quad \left.+ \beta^\top (X^\top \kappa + Q\mu_0 - X^\top \Omega(\omega)Zu)\right]. \end{aligned}$$

Hence,

$$\begin{aligned} \beta \mid u, \omega, \tau, y &\sim N((X^\top \Omega(\omega)X + Q)^{-1} \\ &\quad \times (X^\top \kappa + Q\mu_0 - X^\top \Omega(\omega)Zu), (X^\top \Omega(\omega)X + Q)^{-1}). \end{aligned} \quad (10)$$

Also from (9), the conditional density of  $u$  given  $\beta, \omega, \tau, y$  is

$$\begin{aligned}\pi(u \mid \beta, \omega, \tau, y) &\propto \prod_{i=1}^n \exp \left\{ k_i z_i^\top u - \frac{\omega_i}{2} \left[ (z_i^\top u)^2 + 2z_i^\top u x_i^\top \beta \right] \right\} \exp \left[ -\frac{1}{2} u^\top D(\tau) u \right] \\ &= \exp \left[ -\frac{1}{2} u^\top (Z^\top \Omega(\omega) Z + D(\tau)) u + u^\top (Z^\top \kappa - Z^\top \Omega(\omega) X \beta) \right].\end{aligned}$$

Thus, it follows that

$$\begin{aligned}u \mid \beta, \omega, \tau, y &\sim N((Z^\top \Omega(\omega) Z + D(\tau))^{-1} \\ &\quad \times (Z^\top \kappa - Z^\top \Omega(\omega) X \beta), (Z^\top \Omega(\omega) Z + D(\tau))^{-1}).\end{aligned}\quad (11)$$

Also from (9), the conditional density of  $\omega$  and  $\tau$  given  $\eta$  and  $y$  is as follows

$$\begin{aligned}\pi(\omega, \tau \mid \eta, y) &\propto \prod_{i=1}^n \exp(-\omega_i (m_i^\top \eta)^2 / 2) p(\omega_i) \mid D(\tau) \Big|^{\frac{1}{2}} \exp(-u^\top D(\tau) u / 2) \\ &\quad \times \prod_{j=1}^r \tau_j^{a_j-1} \exp(-b_j \tau_j),\end{aligned}\quad (12)$$

where  $|D(\tau)|$  is the determinant of  $D(\tau)$ . From the above, we see that  $\omega_i$ 's,  $i = 1, \dots, n$  are conditionally independent given  $(\eta, y)$ . Also,  $(\omega, \tau)$  are conditionally mutually independent given  $(\eta, y)$ . The conditional density for  $\omega_i$  is

$$\begin{aligned}\pi(\omega_i \mid \eta, y) &\propto \exp(-\omega_i (m_i^\top \eta)^2 / 2) p(\omega_i) \\ &= \sum_{\ell=0}^{\infty} (-1)^\ell \frac{(2\ell+1)}{\sqrt{2\pi\omega_i^3}} \exp \left( -\frac{(2\ell+1)^2}{8\omega_i} - \frac{\omega_i (m_i^\top \eta)^2}{2} \right),\end{aligned}\quad (13)$$

where the equality follows from (8). From Wang and Roy (2018b), the pdf for  $PG(a, b)$ ,  $a > 0$ ,  $b \in \mathbb{R}$  is

$$\begin{aligned}p(x \mid a, b) &= \left[ \cosh \left( \frac{b}{2} \right) \right]^a \frac{2^{a-1}}{\Gamma(a)} \sum_{\ell=0}^{\infty} (-1)^\ell \frac{\Gamma(\ell+a)}{\Gamma(\ell+1)} \frac{(2\ell+a)}{\sqrt{2\pi x^3}} \\ &\quad \times \exp \left( -\frac{(2\ell+a)^2}{8x} - \frac{xb^2}{2} \right),\end{aligned}$$

for  $x > 0$ , where the hyperbolic cosine function  $\cosh(t) = (e^t + e^{-t})/2$ . Hence, from (13) we have

$$\omega_i \mid \eta, y \stackrel{ind}{\sim} PG(1, |m_i^\top \eta|), \quad i = 1, \dots, n. \quad (14)$$

From (12), the conditional density for  $\tau_j$  is given by

$$\pi(\tau_j \mid \eta, y) \propto \tau_j^{q_j/2+a_j-1} \exp \left[ -\tau_j (b_j + u_j^\top u_j / 2) \right], \quad j = 1, \dots, r. \quad (15)$$

Thus, we have

$$\tau_j \mid \eta, y \stackrel{ind}{\sim} \text{Gamma}(a_j + q_j/2, b_j + u_j^\top u_j/2), j = 1, \dots, r.$$

We will allow  $a_j > -q_j/2$ , and  $b_j \geq 0$ . The case where  $b_j = 0$  is more complicated and will be discussed at length in Section 4.

Let  $(\beta^{(m)}, u^m, \omega^{(m)}, \tau^{(m)})$  denote the  $m^{th}$  element for  $(\beta, u, \omega, \tau)$  in the FG chain. Thus, a single iteration of the full Gibbs sampler  $\{\beta^{(m)}, u^m, \omega^{(m)}, \tau^{(m)}\}_{m=0}^\infty$  consists of the following four steps:

---

**Algorithm 1** The  $(m+1)$ st iteration of the full Gibbs sampler

---

1. Draw  $\tau_j^{(m+1)} \stackrel{ind}{\sim} \text{Gamma}(a_j + q_j/2, b_j + u_j^\top u_j/2)$ ,  $j = 1, \dots, r$  with  $u = u^{(m)}$ .
  2. Draw  $\omega_i^{(m+1)} \stackrel{ind}{\sim} PG(1, |m_i^\top \eta^{(m)}|)$ ,  $i = 1, \dots, n$ .
  3. Draw  $u^{(m+1)} \sim (11)$  with  $\tau = \tau^{(m+1)}$ ,  $\omega = \omega^{(m+1)}$ , and  $\beta = \beta^{(m)}$ .
  4. Draw  $\beta^{(m+1)} \sim (10)$  with  $\omega = \omega^{(m+1)}$ .
- 

## 2.2. A two-block Gibbs sampler

In this section, we construct a block Gibbs (BG) sampler for (5). As mentioned before,  $M = [XZ]$  with the  $i^{th}$  row  $m_i^\top$  for  $i = 1, \dots, n$ . Note that  $x_i^\top \beta + z_i^\top u = m_i^\top \eta$ . From (9), the conditional density of  $\eta$  given  $\omega, \tau, y$  is given by

$$\begin{aligned} \pi(\eta \mid \omega, \tau, y) &\propto \prod_{i=1}^n \exp[k_i m_i^\top \eta - \omega_i (m_i^\top \eta)^2 / 2] \exp\left[-\frac{1}{2} u^\top D(\tau) u\right] \\ &\quad \times \exp\left[-\frac{1}{2} (\beta - \mu_0)^\top Q (\beta - \mu_0)\right] \\ &\propto \exp\left[-\frac{1}{2} (\eta - \Sigma(M^\top \kappa + b))^\top \Sigma^{-1} (\eta - \Sigma(M^\top \kappa + b))\right], \end{aligned} \quad (16)$$

where  $\Sigma^{-1} = M^\top \Omega(\omega) M + A(\tau)$ ,

$$b_{(p+q) \times 1} = \begin{pmatrix} Q\mu_0 \\ 0_{q \times 1} \end{pmatrix} \text{ and } A(\tau)_{(p+q)(p+q)} = \begin{pmatrix} Q & 0 \\ 0 & D(\tau) \end{pmatrix}.$$

Hence,

$$\eta \mid \omega, \tau, y \sim N((M^\top \Omega(\omega) M + A(\tau))^{-1} (M^\top \kappa + b), (M^\top \Omega(\omega) M + A(\tau))^{-1}). \quad (17)$$

In the FG sampler in Section 2.1,  $\tau, \omega, u$  and  $\beta$  are drawn sequentially, whereas, in this section, we show that the conditional distribution of  $\eta$  given  $\omega, \tau, y$  is normal. From (12), we can see, conditional on  $(\eta, y)$ ,  $\omega$  and  $\tau$  are independent. Thus,  $\tau$  and  $\omega$  can be drawn jointly as a block and we have a two-block Gibbs sampler.

Let  $\eta^{(m)}, \omega^{(m)}$ , and  $\tau^{(m)}$  denote the values of  $\eta, \omega$ , and  $\tau$ , respectively, in the  $m^{th}$  iteration of the BG sampler. A single iteration of the block Gibbs sampler  $\{\eta^{(m)}, \omega^{(m)}, \tau^{(m)}\}_{m=0}^\infty$  consists of the following two steps:

---

**Algorithm 2** The  $(m+1)$ st iteration of the two-block Gibbs sampler

---

1. Draw  $\tau_j^{(m+1)} \stackrel{\text{ind}}{\sim} \text{Gamma}(a_j + q_j/2, b_j + u_j^\top u_j/2)$ ,  $j = 1, \dots, r$  with  $u = u^{(m)}$ , and independently draw  $\omega_i^{(m+1)} \stackrel{\text{ind}}{\sim} \text{PG}(1, |m_i^\top \eta^{(m)}|)$ ,  $i = 1, \dots, n$ .
  2. Draw  $\eta^{(m+1)} \sim (17)$  with  $\tau = \tau^{(m+1)}$  and  $\omega = \omega^{(m+1)}$ .
- 

The conditional distribution of  $\eta$  in the BG sampler and the conditional distributions of  $\beta$  and  $u$  in the FG sampler are all normal distributions of the form  $N(S^{-1}t, S^{-1})$  for some matrix  $S$  and a vector  $t$ . Note that, for the conditional distribution of  $\eta$ ,  $S$  is a  $(p+q) \times (p+q)$  matrix, whereas for  $\beta$  and  $u$  this is a  $p \times p$  and  $q \times q$  matrix, respectively. Thus, a naive method of drawing from  $N(S^{-1}t, S^{-1})$  is inefficient especially if  $p$  and/or  $q$  is large as it involves calculating inverse of the matrix  $S$ . Here, we use a known method of drawing from  $N(S^{-1}t, S^{-1})$  that does not require computing  $S^{-1}$ . The method is as follows:

---

Algorithm for drawing from  $N(S^{-1}t, S^{-1})$

---

1. Let  $S = LL^\top$  be the Cholesky decomposition of  $S$ .
  2. Solve  $Lw = t$ .
  3. Draw  $z \sim N(0, I_k)$  where  $k$  is the dimension of  $S$ .
  4. Solve  $L^\top x = w + z$ . Then  $x \sim N(S^{-1}t, S^{-1})$ .
- 

### 3. A real data example

We consider the student performance data set from Cortez and Silva (2008). This data set includes  $n = 649$  observations and 33 variables including several categorical variables. As in Cortez and Silva (2008), the binary response is defined as 1 if the final grade is greater than or equal to 10, otherwise, it is defined as 0. Recall that  $p$  denotes the number of columns for the design matrix  $X$ . Also, note that, categorical variables are incorporated into the LLMM as sets of dichotomous variables through what is known as dummy coding. We consider different subsets of variables while fitting the LLMM to compare the BG and FG samplers for different dimensions. In particular, we consider  $p = 3, 7, 23$ , including an intercept term. We also keep one random effect ‘school’ with 2 levels in the LLMM. As mentioned in Chib and Ramamurthy (2010), a key principle to block sampling in MCMC is that ‘parameters in different blocks are not strongly correlated whereas those within a block are’ (see also Roberts and Sahu, 1997; Turek et al., 2017). As we observe later in this Section, when  $p$  varies, the average absolute (posterior) correlations between  $\beta$  and  $u$  change greatly. The specific values of  $p$  we choose here are irrelevant for the general conclusions of this section. Indeed, for some other  $p$  values, close to the ones we choose here, the general pattern of the different empirical measures remains the same.

We analyze the data set by fitting the LLMM with a proper normal prior (3) on  $\beta$  with  $\mu_0 = 0$  and  $Q = 0.001I_3$ . Also, we use a proper Gamma prior (4) on

$\tau_1$  where the (prior) mean and variance of  $\tau_1$  are 1.2 and 100, respectively ( $a_1 = 0.0144$  and  $b_1 = 0.012$ ). We ran the BG sampler for  $m = 120,000$  iterations, starting at an initial value  $\eta^{(0)} = (\beta^{(0)}, u^{(0)})$  with a burn-in of  $B = 20,000$  iterations. Here,  $\beta^{(0)}$  is the estimate of  $\beta$  obtained by fitting a logistic linear model without any random effect. For  $p = 3, 7$ , the initial value  $u^{(0)}$  is a sample drawn from  $N(0, (1/\tau_1^{(0)})I_2)$ , where  $1/\tau_1^{(0)}$  is the estimate of the random effect variance component obtained from the R package lme4. For  $p = 23$ ,  $1/\tau_1^{(0)}$  is the estimate of the random effect variance component obtained from  $p = 7$  as lme4 did not run successfully in the case of  $p = 23$ . The FG sampler was also run for  $m = 120,000$  iterations with a burn-in of  $B = 20,000$  iterations.

The BG and FG samplers are compared using the lag  $k$  autocorrelation function (ACF) values  $k = 1, \dots, 5$ , the effective sample size (ESS) and the multivariate ESS (mESS) (See Roy (2020) for a simple introduction to these convergence diagnostic measures.). The ESS and mESS are calculated using the R package mcmcse. We also compute the mean squared jumps (MSJ) defined as  $\sum_{i=B+1}^m \|\beta^{(i+1)} - \beta^{(i)}\|^2 / (m - B)$  for the  $\beta$  variable, and similarly for the other variables. Here,  $\|\cdot\|$  denotes the Euclidean norm. Tables 1–3 provide the values of the ACF for  $\beta_0, \beta_1, \beta_2$  and  $\tau_1$  for the BG and FG samplers, as  $p$  varies. Better performance of the BG sampler compared to the FG sampler is observed from its mostly smaller ACF values. Table 4 provides the ESS values of the intercept parameter, first two regression coefficients and  $\tau_1$ . It also gives the mESS values for  $u$  and  $(\beta, \tau_1)$ , as  $p$  varies. Again, better efficiency of the BG sampler compared to the FG sampler is observed from its generally larger ESS and mESS values. We calculate the average of the absolute (posterior) correlations between the coordinates of the  $\beta$  vector and those of the  $u$  vector computed based on the BG samples mentioned before. For  $p = 3, 7$ , and  $23$ , these values are 0.2601, 0.1143, and 0.0206, respectively. From the mESS values of the parameters of interest, namely  $(\beta, \tau_1)$ , given in Table 4, we see that the increase in the efficiency of the BG sampler compared to the FG sampler is higher for smaller  $p$  values. From Table 5, it can be seen that the BG sampler leads to higher MSJ values than the FG sampler with the exception of  $\tau_1$  in some cases. Thus, Table 5 also corroborates better mixing of the BG sampler than the FG sampler. So, in practice, the BG sampler can provide significant gains compared to the FG sampler.

Finally, Table 4 also provides the time normalized efficiency (ESS and mESS values per second) for the two samplers, as  $p$  varies. From this Table, we see that the BG sampler has always resulted in larger ESS and mESS values per second than the FG sampler. Indeed, the BG sampler results in higher time normalized ESS values even in the cases when the FG sampler performs better in terms of the ESS (See e.g. ESS ( $\beta_1$ ) for  $p = 7$ ). Recall that in every iteration, the BG sampler makes a draw from a  $(p + q)$  dimensional normal distribution, whereas the FG sampler draws from a  $q$  dimensional normal distribution and then a  $p$  dimensional normal distribution. Other draws are the same for both the BG and FG samplers. Using the Cholesky decomposition method mentioned in Section 2.2, we observe that for all values of  $p$  considered here, the BG sampler takes

less time than the FG sampler to complete a certain number of iterations. On the other hand, when  $(p + q)$  takes much larger values, the BG sampler takes more running time than the FG sampler.

TABLE 1  
ACF for the BG and FG samplers for the student performance data with  $p = 3$

| Parameter | Sampler | lag 1 | lag 2 | lag 3 | lag 4 | lag 5 |
|-----------|---------|-------|-------|-------|-------|-------|
| $\beta_0$ | BG      | 0.434 | 0.385 | 0.359 | 0.333 | 0.319 |
|           | FG      | 0.985 | 0.974 | 0.964 | 0.956 | 0.948 |
| $\beta_1$ | BG      | 0.597 | 0.380 | 0.258 | 0.192 | 0.153 |
|           | FG      | 0.613 | 0.402 | 0.282 | 0.212 | 0.173 |
| $\beta_2$ | BG      | 0.836 | 0.734 | 0.667 | 0.621 | 0.586 |
|           | FG      | 0.838 | 0.739 | 0.671 | 0.623 | 0.587 |
| $\tau_1$  | BG      | 0.374 | 0.212 | 0.134 | 0.091 | 0.071 |
|           | FG      | 0.405 | 0.285 | 0.224 | 0.186 | 0.155 |

TABLE 2  
ACF for the BG and FG samplers for the student performance data with  $p = 7$

| Parameter | Sampler | lag 1 | lag 2 | lag 3 | lag 4 | lag 5 |
|-----------|---------|-------|-------|-------|-------|-------|
| $\beta_0$ | BG      | 0.463 | 0.409 | 0.365 | 0.340 | 0.320 |
|           | FG      | 0.924 | 0.867 | 0.821 | 0.781 | 0.747 |
| $\beta_1$ | BG      | 0.436 | 0.191 | 0.087 | 0.036 | 0.018 |
|           | FG      | 0.437 | 0.193 | 0.088 | 0.038 | 0.017 |
| $\beta_2$ | BG      | 0.419 | 0.185 | 0.085 | 0.047 | 0.031 |
|           | FG      | 0.420 | 0.187 | 0.090 | 0.053 | 0.032 |
| $\tau_1$  | BG      | 0.379 | 0.216 | 0.142 | 0.098 | 0.063 |
|           | FG      | 0.372 | 0.244 | 0.190 | 0.150 | 0.122 |

TABLE 3  
ACF for the BG and FG samplers for the student performance data with  $p = 23$

| Parameter | Sampler | lag 1 | lag 2 | lag 3 | lag 4 | lag 5 |
|-----------|---------|-------|-------|-------|-------|-------|
| $\beta_0$ | BG      | 0.870 | 0.811 | 0.761 | 0.714 | 0.669 |
|           | FG      | 0.933 | 0.874 | 0.820 | 0.770 | 0.723 |
| $\beta_1$ | BG      | 0.637 | 0.437 | 0.323 | 0.256 | 0.213 |
|           | FG      | 0.655 | 0.463 | 0.348 | 0.278 | 0.235 |
| $\beta_2$ | BG      | 0.885 | 0.807 | 0.753 | 0.713 | 0.681 |
|           | FG      | 0.880 | 0.801 | 0.744 | 0.702 | 0.669 |
| $\tau_1$  | BG      | 0.384 | 0.228 | 0.147 | 0.098 | 0.070 |
|           | FG      | 0.395 | 0.263 | 0.198 | 0.164 | 0.141 |

#### 4. Geometric ergodicity of the block Gibbs sampler

We begin this section with a discussion on the conditional density  $\pi(\tau \mid \eta, y)$ . Since we allow the prior rate parameter  $b_j$  for  $\tau_j$  to be zero, define  $A = \{j \in \{1, 2, \dots, r\} : b_j = 0\}$ . Recall from (15) that  $\tau_j \mid \eta, y \stackrel{\text{ind}}{\sim} \text{Gamma}(a_j + q_j/2, b_j + u_j^\top u_j/2)$ ,  $j = 1, \dots, r$  when  $a_j + q_j/2 > 0$  and  $b_j + u_j^\top u_j/2 > 0$ . The density  $\pi(\tau \mid \eta, y) = \prod_{j=1}^r \pi(\tau_j \mid \eta, y)$  is not defined when  $A$  is not empty and  $\|u_j\| = 0$

TABLE 4

Multivariate ESS or ESS for the BG and FG samplers for the student performance data with  $p = 3, 7$ , and  $23$ . The numbers inside the parentheses are the corresponding values per second. The column name MC stands for Markov chain.

| $p$ | MC | mESS ( $\beta$ $\tau_1$ ) | mESS ( $\beta$ ) | ESS ( $\beta_0$ ) | ESS ( $\beta_1$ ) | ESS ( $\beta_2$ ) | mESS( $u$ )     | ESS ( $\tau_1$ ) |
|-----|----|---------------------------|------------------|-------------------|-------------------|-------------------|-----------------|------------------|
| 3   | BG | 19,012<br>(321)           | 15,979<br>(270)  | 4,229<br>(71)     | 9,268<br>(156)    | 2,586<br>(44)     | 34,623<br>(585) | 31,688<br>(535)  |
|     | FG | 1,539<br>(22)             | 978<br>(14)      | 40<br>(1)         | 8,089<br>(115)    | 2,560<br>(36)     | 76<br>(1)       | 2,037<br>(29)    |
| 7   | BG | 27,474<br>(426)           | 26,702<br>(414)  | 4,494<br>(70)     | 36,015<br>(559)   | 31,294<br>(486)   | 34,844<br>(541) | 32,334<br>(502)  |
|     | FG | 13,533<br>(179)           | 12,793<br>(169)  | 1,894<br>(25)     | 38,572<br>(509)   | 29,826<br>(394)   | 1,931<br>(25)   | 11,401<br>(150)  |
| 23  | BG | 23,068<br>(295)           | 22,966<br>(294)  | 3,031<br>(39)     | 7,272<br>(93)     | 2,130<br>(27)     | 31,529<br>(404) | 28,561<br>(366)  |
|     | FG | 18,016<br>(187)           | 17,770<br>(184)  | 3,040<br>(32)     | 5,912<br>(61)     | 1,554<br>(16)     | 1,022<br>(11)   | 7,911<br>(82)    |

TABLE 5

Mean squared jumps for the BG and FG samplers for the student performance data with  $p = 3, 7$ , and  $23$

| p  | BG      |       |         | FG      |      |         |
|----|---------|-------|---------|---------|------|---------|
|    | $\beta$ | $u$   | $\tau$  | $\beta$ | $u$  | $\tau$  |
| 3  | 12.94   | 24.41 | 1319.13 | 0.73    | 0.10 | 1071.40 |
| 7  | 15.52   | 24.97 | 1192.74 | 3.03    | 0.10 | 1309.20 |
| 23 | 90.19   | 29.68 | 1240.48 | 74.84   | 0.11 | 1249.00 |

for  $j \in A$ . Let  $N = \{\eta \in \mathbb{R}^{p+q}, \prod_{j \in A} \|u_j\| = 0\}$ . The fact that  $\pi(\tau | \eta, y)$  is not defined on  $N$  is irrelevant for simulating the BG sampler as  $N$  is a measure zero set with respect to the Lebesgue measure on  $\mathbb{R}^{p+q}$ . But, for a theoretical analysis of the BG chain,  $\pi(\tau | \eta, y)$  needs to be defined for all  $\eta \in \mathbb{R}^{p+q}$ . For all  $\eta \in \mathbb{R}^{p+q}$ , we define

$$\pi(\tau | \eta, y) = \begin{cases} \prod_{j=1}^r f_G(\tau_j, a_j + \frac{q_j}{2}, b_j + \frac{1}{2}u_j^\top u_j) & \text{if } \eta \notin N \\ \prod_{j=1}^r f_G(\tau_j, 1, 1) & \text{if } \eta \in N \end{cases}. \quad (18)$$

Here,  $f_G(x, a, b)$  denotes the pdf of a gamma random variable with the shape parameter  $a$ , the rate parameter  $b$ , and evaluated at  $x$ . Thus,  $f_G(x, a, b) = (b^a/\Gamma(a))x^{a-1}\exp(-bx)$ . The Markov transition density (Mtd) of the BG chain  $\{\eta^{(m)}, \omega^{(m)}, \tau^{(m)}\}_{m=0}^\infty$  is

$$\begin{aligned} k(\eta, \omega, \tau | \eta', \omega', \tau') &= \pi(\eta | \omega, \tau, y) \pi(\omega, \tau | \eta', y) \\ &= \pi(\eta | \omega, \tau, y) \left[ \prod_{i=1}^n \pi(\omega_i | \eta', y) \right] \pi(\tau | \eta', y), \end{aligned} \quad (19)$$

where the conditional densities  $\pi(\eta | \omega, \tau, y)$ ,  $\pi(\omega_i | \eta', y)$ , and  $\pi(\tau | \eta', y)$  on the right side of (19) are given in (16), (13), and (18), respectively. It is easy to see that the joint density (9) is the invariant density of  $k$ , and  $k$  is  $\varphi$ -irreducible. Thus, if (9) is a proper density, that is, if  $c(y) < \infty$  in (5), then the BG chain  $\{\eta^{(m)}, \omega^{(m)}, \tau^{(m)}\}_{m=0}^\infty$  is Harris ergodic (Meyn and Tweedie, 1993, Chap 10), and hence, it can be used to consistently estimate means with respect to (9). Let  $S = \mathbb{R}^{p+q} \times \mathbb{R}_+^n \times \mathbb{R}_+^r$ . In fact, if  $g : S \rightarrow \mathbb{R}$  is integrable with respect to (9),

that is, if

$$\mathbb{E}_\pi |g(\eta, \omega, \tau)| := \int_S |g(\eta, \omega, \tau)| \pi(\beta, u, \omega, \tau | y) d\eta d\omega d\tau < \infty,$$

then  $\bar{g}_m := \sum_{i=0}^{m-1} g(\eta^{(i)}, \omega^{(i)}, \tau^{(i)})/m \rightarrow \mathbb{E}_\pi g$  almost surely as  $m \rightarrow \infty$ . On the other hand, even when  $\mathbb{E}_\pi g^2 < \infty$ , Harris ergodicity of  $k$  does not guarantee a CLT holds for  $\bar{g}_m$ , which is used to obtain valid standard errors of  $\bar{g}_m$ . We say a CLT for  $\bar{g}_m$  exists if  $\sqrt{m}(\bar{g}_m - \mathbb{E}_\pi g) \xrightarrow{d} N(0, \sigma_g^2)$  as  $m \rightarrow \infty$  for some  $\sigma_g^2 \in (0, \infty)$ . Certain convergence rates of the BG chain, as we explain next, ensure a CLT holds for  $\bar{g}_m$ .

Let  $\mathcal{B}(S)$  denotes the Borel  $\sigma$ -algebra of  $S$ . Let  $K^{(m)} : S \times \mathcal{B}(S) \rightarrow [0, 1]$  denotes the  $m$ -step Markov transition function (Mtf) corresponding to the Mtd (19), that is,

$$\begin{aligned} K^{(m)}((\eta', \omega', \tau'), B) &= P((\eta^{(m+j)}, \omega^{(m+j)}, \tau^{(m+j)}) \in B | (\eta^{(j)}, \omega^{(j)}, \tau^{(j)})) \\ &= (\eta', \omega', \tau'), \end{aligned}$$

for any  $j \in \{1, 2, \dots\}$  and for any measurable set  $B \in \mathcal{B}(S)$ . The BG chain is geometrically ergodic if there exist a function  $H : S \rightarrow [0, \infty)$  and a constant  $\rho \in (0, 1)$  such that for all  $m = 0, 1, 2, \dots$ ,

$$\begin{aligned} \|K^m((\eta', \omega', \tau'), \cdot) - \Pi(\cdot)\|_{\text{TV}} &:= \sup_{B \in \mathcal{B}(S)} \|K^m((\eta', \omega', \tau'), B) - \Pi(B)\| \\ &\leq H(\eta', \omega', \tau') \rho^m, \end{aligned} \quad (20)$$

where  $\Pi(\cdot)$  denotes the probability measure corresponding to the joint posterior density (9), and  $\|\cdot\|_{\text{TV}}$  denotes the total variation norm. Harris ergodicity of  $k$  implies the TV norm in (20)  $\downarrow 0$  as  $m \rightarrow \infty$ , but does not ascertain any rate at which this convergence takes place. On the other hand, (20) guarantees a CLT for  $\bar{g}_m$  if  $\mathbb{E}_\pi |g|^{2+\delta} < \infty$  for some  $\delta > 0$  (Roberts and Rosenthal, 2004). If (20) holds, it also implies that consistent batch means and a spectral variance estimator  $\hat{\sigma}_g^2$  of  $\sigma_g^2$  are available (Vats, Flegal and Jones (2018), Vats, Flegal and Jones (2019)), and thus a valid standard error (SE)  $\hat{\sigma}_g/\sqrt{m}$  for  $\bar{g}_m$  can be calculated. An advantage of being able to calculate a valid SE is that it can be used to decide ‘when to stop’ running the BG chain (Roy, 2020).

An important property that we are going to use in this article is that the marginal sequences  $\{\eta^{(m)}\}_{m=0}^\infty$ ,  $\{(\omega^{(m)}, \tau^{(m)})\}_{m=0}^\infty$  of the BG chain  $\{\eta^{(m)}, \omega^{(m)}, \tau^{(m)}\}_{m=0}^\infty$  are themselves Markov chains, and either all three chains are geometrically ergodic or none of them (Liu, Wong and Kong, 1994; Roberts and Rosenthal, 2001). Thus, we are free to analyze any of these chains to study their geometric convergence properties. Indeed, here we analyze the  $\{\eta^{(m)}\}_{m=0}^\infty$  marginal chain.

We denote the Markov chain  $\{\eta^{(m)}\}_{m=0}^\infty$  on  $\mathbb{R}^{p+q}$  by  $\Psi$  while the Markov chain  $\{\eta^{(m)}\}_{m=0}^\infty$  on  $\mathbb{R}^{p+q} \setminus N$  is denoted by  $\tilde{\Psi}$ . From (19) it follows that the Mtd of the  $\Psi$  chain is

$$\tilde{k}(\eta | \eta') = \int_{\mathbb{R}_+^p} \int_{\mathbb{R}_+^q} \pi(\eta | \omega, \tau, y) \pi(\omega, \tau | \eta', y) d\omega d\tau, \quad (21)$$

We can verify that  $\tilde{k}(\eta \mid \eta')\pi(\eta' \mid y) = \tilde{k}(\eta' \mid \eta)\pi(\eta \mid y)$  for all  $\eta, \eta' \in \mathbb{R}^{p+q}$  where  $\pi(\eta \mid y) = \int_{\mathbb{R}_+^r} \int_{\mathbb{R}_+^n} \pi(\eta, \omega, \tau \mid y) d\omega d\tau$  is the  $\eta$  marginal density of (9). Hence, (21) is reversible with respect to  $\pi(\eta \mid y)$ , and thus,  $\pi(\eta \mid y)$  is the invariant density for the Markov chain  $\{\eta^{(m)}\}_{m=0}^\infty$ . Also, since  $\{\eta^{(m)}\}_{m=0}^\infty$  is reversible, GE of the chain implies that CLTs hold for all square integrable functions with respect to  $\pi(\eta \mid y)$  (Roberts and Rosenthal, 1997). We first establish GE of the  $\tilde{\Psi}$  chain. As explained in the proof of Theorem 1, GE of  $\tilde{\Psi}$  implies that of  $\Psi$ .

**Theorem 1.** *If  $\pi(\beta) \propto 1$ , that is, if  $Q = 0$  in (3), 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.  $a_j + q_j/2 > 0$  for  $j = 1, \dots, r$ ;
3.  $M$  has full rank;
4. There exists a positive vector  $e > 0$  such that  $e'M^* = 0$  where  $M^*$  is an  $n \times (p + q)$  matrix with  $i$ th row  $c_i m_i^\top$ , where  $c_i = 1 - 2y_i$ ,  $i = 1, \dots, n$ .

The proof of Theorem 1 is given in the Appendix C. The condition 4 can be easily checked by an optimization method presented in Roy and Hobert (2007).

**Remark 1.** The conditions in Theorem 1 are the same as the conditions assumed in Wang and Roy's (2018a) Theorem 2 that establishes GE of Gibbs samplers for the probit linear mixed model.

**Remark 2.** As mentioned before, Wang and Roy (2018c) analyzed the PG sampler for LLMMs with proper normal priors on  $\beta$  and a truncated gamma prior on  $\tau$ . Wang and Roy's (2018c) proof established uniform ergodicity which is stronger than geometric ergodicity, but they assumed a stronger prior on  $\tau$ . Indeed, their proof involving a minorization condition requires that the support of  $\tau$  is bounded away from zero. Our analysis of the BG Markov chain does not entail any minorization condition, and does not put any restriction, other than being positive, on the support of the variance components.

## 5. Conclusion

In this article, we consider an efficient block Gibbs sampler based on the Pólya-Gamma DA (Polson, Scott and Windle, 2013) for one of the most widely used statistical models, namely the LLMMs. Through numerical examples, we observe that blocking can improve performance of the Pólya-Gamma Gibbs samplers. We hope that the article will encourage development and use of efficient blocking strategies for Monte Carlo estimation of other GLMMs, including spatial GLMMs where MCMC algorithms are known to suffer from slow mixing as noted in Evangelou and Roy (2019).

Undertaking a Foster-Lyapunov drift analysis, we establish CLTs for the BG sampler based MCMC estimators under the improper uniform prior on the regression coefficients and improper or proper priors on the variance components.

These theoretical results are crucial for obtaining standard errors for MCMC estimates of posterior means. In the process of our proof for demonstrating CLTs for the BG sampler, we also establish some general results on the Pólya-Gamma distribution. A potential future problem is to construct and study block Gibbs samplers for other GLMMs, including the mixed models with the robit link (Roy, 2012b).

## Appendices

### Appendix A: Some useful results

Recall from Section 2.2 that if  $Q = 0$ , that is, if  $\pi(\beta) \propto 1$ , then  $b = 0$  and  $A(\tau) = \begin{pmatrix} 0 & 0 \\ 0 & D(\tau) \end{pmatrix} = B(\tau)$ , say. In that case, (17) becomes

$$\eta \mid \omega, \tau, y \sim N((M^\top \Omega(\omega)M + B(\tau))^{-1} M^\top \kappa, (M^\top \Omega(\omega)M + B(\tau))^{-1}).$$

By using the method of calculating the inverse of a partitioned matrix, the covariance matrix is

$$\begin{aligned} (M^\top \Omega(\omega)M + B(\tau))^{-1} &= \begin{pmatrix} X^\top \Omega(\omega)X & X^\top \Omega(\omega)Z \\ Z^\top \Omega(\omega)X & Z^\top \Omega(\omega)Z + D(\tau) \end{pmatrix}^{-1} \\ &= \begin{pmatrix} (\tilde{X}^\top \tilde{X})^{-1} + \tilde{R}\tilde{S}^{-1}\tilde{R}^\top & -\tilde{R}\tilde{S}^{-1} \\ -\tilde{S}^{-1}\tilde{R}^\top & \tilde{S}^{-1} \end{pmatrix}, \end{aligned}$$

where  $\tilde{X} = \Omega(\omega)^{\frac{1}{2}}X$ ,  $\tilde{Z} = \Omega(\omega)^{\frac{1}{2}}Z$ ,  $P_{\tilde{X}} = \tilde{X}(\tilde{X}^\top \tilde{X})^{-1}\tilde{X}^\top$ ,  $\tilde{S} = \tilde{Z}^\top(I - P_{\tilde{X}})\tilde{Z} + D(\tau)$ , and  $\tilde{R} = (\tilde{X}^\top \tilde{X})^{-1}\tilde{X}^\top \tilde{Z}$ . For the mean vector, it follows that

$$\begin{aligned} (M^\top \Omega(\omega)M + B(\tau))^{-1} M^\top \kappa &= \begin{pmatrix} (\tilde{X}^\top \tilde{X})^{-1} + \tilde{R}\tilde{S}^{-1}\tilde{R}^\top & -\tilde{R}\tilde{S}^{-1} \\ -\tilde{S}^{-1}\tilde{R}^\top & \tilde{S}^{-1} \end{pmatrix} \begin{pmatrix} X^\top \kappa \\ Z^\top \kappa \end{pmatrix} \\ &= \begin{pmatrix} (\tilde{X}^\top \tilde{X})^{-1} X^\top \kappa + \tilde{R}\tilde{S}^{-1}\tilde{R}^\top X^\top \kappa - \tilde{R}\tilde{S}^{-1} Z^\top \kappa \\ -\tilde{S}^{-1}\tilde{R}^\top X^\top \kappa + \tilde{S}^{-1} Z^\top \kappa \end{pmatrix}. \end{aligned} \quad (22)$$

The first element in the right-hand side of (22) is the mean vector for  $\beta$ , while the second element in it is the mean vector for  $u$ . Thus,

$$u \mid \omega, \tau, y \sim N(-\tilde{S}^{-1}\tilde{R}^\top X^\top \kappa + \tilde{S}^{-1} Z^\top \kappa, \tilde{S}^{-1}). \quad (23)$$

**Lemma 1.** Let  $R_j$  be a  $q_j \times q$  matrix consisting of 0's and 1's such that  $R_j u = u_j$ . Then

$$(R_j \tilde{S}^{-1} R_j^\top)^{-1} \preceq \left( \sum_{i=1}^n \omega_i \text{tr}(Z^\top Z) + \tau_j \right) \mathbf{I}_{q_j}.$$

Here, for two matrices  $A$  and  $B$ ,  $A \preceq B$  means  $B - A$  is a positive semidefinite matrix.

*Proof.* Let  $\lambda_{\max}$  denote the largest eigenvalue for  $\tilde{Z}^\top(\mathbf{I} - P_{\tilde{X}})\tilde{Z}$ , then

$$\tilde{S} = \tilde{Z}^\top(\mathbf{I} - P_{\tilde{X}})\tilde{Z} + D(\tau) \preceq \lambda_{\max}\mathbf{I}_q + D(\tau) \preceq \text{tr}(\tilde{Z}^\top(\mathbf{I} - P_{\tilde{X}})\tilde{Z})\mathbf{I}_q + D(\tau),$$

where  $D(\tau) = \oplus_{j=1}^r \tau_j \mathbf{I}_{q_j}$  as defined before, and the second inequality follows from the fact that  $\tilde{Z}^\top(\mathbf{I} - P_{\tilde{X}})\tilde{Z}$  is a positive semidefinite matrix. Now

$$\begin{aligned} \text{tr}(\tilde{Z}^\top(\mathbf{I} - P_{\tilde{X}})\tilde{Z}) &\leq \text{tr}(\tilde{Z}^\top \tilde{Z}) = \text{tr}(Z^\top \Omega(\omega)Z) \\ &= \text{tr}\left(\sum_{i=1}^n \omega_i z_i z_i^\top\right) \\ &= \sum_{i=1}^n \text{tr}(\omega_i z_i z_i^\top) = \sum_{i=1}^n \omega_i \text{tr}(z_i z_i^\top) \leq \sum_{i=1}^n \omega_i \text{tr}(Z^\top Z), \end{aligned}$$

where  $z_i^\top$  denotes the  $i^{\text{th}}$  row of the  $Z$  matrix, and the first inequality is due to the fact that  $\tilde{Z}^\top P_{\tilde{X}} \tilde{Z}$  is a positive semidefinite matrix. Thus  $\tilde{S} \preceq \sum_{i=1}^n \omega_i \text{tr}(Z^\top Z) \mathbf{I}_q + D(\tau)$ . Hence,  $\tilde{S}^{-1} \succeq (\sum_{i=1}^n \omega_i \text{tr}(Z^\top Z) \mathbf{I}_q + D(\tau))^{-1}$ . Recall that  $R_j u = u_j$ . Extracting the result of the  $j^{\text{th}}$  random effect, we obtain:

$$\begin{aligned} R_j \tilde{S}^{-1} R_j^\top &\succeq R_j \left( \sum_{i=1}^n \omega_i \text{tr}(Z^\top Z) \mathbf{I}_q + D(\tau) \right)^{-1} R_j^\top \\ &= \left( \sum_{i=1}^n \omega_i \text{tr}(Z^\top Z) + \tau_j \right)^{-1} \mathbf{I}_{q_j}. \end{aligned}$$

Thus, we have  $(R_j \tilde{S}^{-1} R_j^\top)^{-1} \preceq (\sum_{i=1}^n \omega_i \text{tr}(Z^\top Z) + \tau_j) \mathbf{I}_{q_j}$ .  $\square$

## Appendix B: Some properties of the Pólya-Gamma distributions

**Lemma 2.** Suppose  $\omega \sim PG(a, b)$ .

1. If  $a \geq 1$ ,  $b \geq 0$ , then for  $0 < s \leq 1$ ,  $E(\omega^{-s}) \leq 2^s b^s + L(s)$ , where  $L(s)$  is a constant depending on  $s$ .
2. If  $a < 1$ ,  $b \geq 0$ , then for  $0 < s < a$ ,  $E(\omega^{-s}) \leq 2^{-s}(\pi^2 + b^2)^s \frac{\Gamma(a-s)}{\Gamma(a)}$ .

*Proof.* We first prove part 1 for  $a = 1$ . The probability density function of a  $PG(1, b)$  random variable is

$$f(x | 1, b) = \cosh(b/2) \sum_{\ell=0}^{\infty} (-1)^\ell \frac{(2\ell+1)}{\sqrt{2\pi x^3}} \exp\left[-\frac{(2\ell+1)^2}{8x} - \frac{b^2}{2}x\right], \quad x > 0.$$

We consider the two cases  $b = 0$  and  $b > 0$  separately.

**Case 1:**  $b = 0$ . Since  $0 < s \leq 1$ , for any  $x > 0$ , we have  $x^{-s} \leq x^{-1} + 1$ . Then,

$$E(\omega^{-s}) \leq \int_0^\infty (x^{-1} + 1) f(x | 1, 0) dx = \int_0^\infty x^{-1} f(x | 1, 0) dx + 1.$$

Now,

$$\begin{aligned}\int_0^\infty x^{-1} f(x | 1, 0) dx &= \int_0^\infty x^{-1} \sum_{\ell=0}^\infty (-1)^\ell \frac{(2\ell+1)}{\sqrt{2\pi x^3}} \exp\left[-\frac{(2\ell+1)^2}{8x}\right] dx \\ &= \int_0^\infty \sum_{\ell=0}^\infty (-1)^\ell x^{-\frac{5}{2}} \frac{(2\ell+1)}{\sqrt{2\pi}} \exp\left[-\frac{(2\ell+1)^2}{8x}\right] dx. \quad (24)\end{aligned}$$

Let  $h_1(x, \ell) = (-1)^\ell x^{-\frac{5}{2}} \frac{(2\ell+1)}{\sqrt{2\pi}} \exp\left[-\frac{(2\ell+1)^2}{8x}\right]$ , then

$$\begin{aligned}\sum_{\ell=0}^\infty \int_0^\infty |h_1(x, \ell)| dx &= \sum_{\ell=0}^\infty \frac{(2\ell+1)}{\sqrt{2\pi}} \int_0^\infty x^{-\frac{5}{2}} \exp\left[-\frac{(2\ell+1)^2}{8x}\right] dx \\ &= 8 \sum_{\ell=0}^\infty \frac{1}{(2\ell+1)^2} < \infty.\end{aligned}$$

Hence,  $|h_1|$  is integrable with respect to the product measure of the counting measure and the Lebesgue measure. By Fubini's Theorem, from (24) we have

$$\begin{aligned}\int_0^\infty x^{-1} f(x | 1, 0) dx &= \sum_{\ell=0}^\infty (-1)^\ell \frac{(2\ell+1)}{\sqrt{2\pi}} \int_0^\infty x^{-\frac{5}{2}} \exp\left[-\frac{(2\ell+1)^2}{8x}\right] dx \\ &= 8 \sum_{\ell=0}^\infty (-1)^\ell (2\ell+1)^{-2} = 8C, \quad (25)\end{aligned}$$

where  $C$  is Catalan's constant. Hence,  $E(\omega^{-s}) \leq 8C + 1$ .

**Case 2:**  $b > 0$ . Note that

$$\begin{aligned}E(\omega^{-s}) &= \int_0^\infty x^{-s} f(x | 1, b) dx \\ &= \int_0^\infty x^{-s-\frac{3}{2}} \cosh(b/2) \sum_{\ell=0}^\infty (-1)^\ell \frac{(2\ell+1)}{\sqrt{2\pi}} \exp\left[-\frac{(2\ell+1)^2}{8x} - \frac{b^2}{2}x\right] dx. \quad (26)\end{aligned}$$

According to 10.32.10 in Olver et al. (2010), we have

$$\int_0^\infty x^{-s-\frac{3}{2}} \exp\left[-\frac{(2\ell+1)^2}{8x} - \frac{b^2}{2}x\right] dx = 2K_{s+\frac{1}{2}}\left(\frac{b(2\ell+1)}{2}\right) \left(\frac{2b}{2\ell+1}\right)^{s+\frac{1}{2}}, \quad (27)$$

where  $K_v(\cdot)$  is the modified Bessel function of the second kind of order  $v$ . For  $x > 0$ , according to 10.32.8 in Olver et al. (2010),

$$\begin{aligned}K_{s+\frac{1}{2}}(x) &= \frac{\sqrt{\pi}(\frac{1}{2}x)^{s+\frac{1}{2}}}{\Gamma(s+1)} \int_1^\infty e^{-xt}(t^2-1)^s dt \\ &= \frac{\sqrt{\pi}(\frac{1}{2}x)^{s+\frac{1}{2}}}{\Gamma(s+1)} e^{-x} \int_0^\infty e^{-xt}(t^2+2t)^s dt \quad (28)\end{aligned}$$

$$\begin{aligned}
&\leq \frac{\sqrt{\pi}(\frac{1}{2}x)^{s+\frac{1}{2}}}{\Gamma(s+1)} e^{-x} \int_0^\infty e^{-xt}(t^{2s} + 2^s t^s) dt \\
&= \frac{\sqrt{\pi}(\frac{1}{2}x)^{s+\frac{1}{2}}}{\Gamma(s+1)} e^{-x} \left( \frac{\Gamma(2s+1)}{x^{2s+1}} + 2^s \frac{\Gamma(s+1)}{x^{s+1}} \right) \\
&= \sqrt{\pi} e^{-x} \left[ \frac{\Gamma(2s+1)}{\Gamma(s+1)} 2^{-s-1/2} x^{-s-1/2} + 2^{-1/2} x^{-1/2} \right]. \quad (29)
\end{aligned}$$

Also, from (28) we have

$$K_{s+\frac{1}{2}}(x) \geq \frac{\sqrt{\pi}(\frac{1}{2}x)^{s+\frac{1}{2}}}{\Gamma(s+1)} e^{-x} \int_0^\infty e^{-xt} 2^s t^s dt = \sqrt{\pi} e^{-x} 2^{-1/2} x^{-1/2}. \quad (30)$$

Let  $h_2(x, \ell) = x^{-s-3/2} \cosh(b/2) (-1)^\ell \frac{(2\ell+1)}{\sqrt{2\pi}} \exp \left[ -\frac{(2\ell+1)^2}{8x} - \frac{b^2}{2} x \right]$ , then

$$\begin{aligned}
&\sum_{\ell=0}^\infty \int_0^\infty |h_2(x, \ell)| dx \\
&= \sum_{\ell=0}^\infty \cosh(b/2) \frac{(2\ell+1)}{\sqrt{2\pi}} \int_0^\infty x^{-s-\frac{3}{2}} \exp \left[ -\frac{(2\ell+1)^2}{8x} - \frac{b^2}{2} x \right] dx \\
&= \cosh(b/2) \sum_{\ell=0}^\infty \frac{(2\ell+1)}{\sqrt{2\pi}} 2K_{s+\frac{1}{2}} \left( \frac{b(2\ell+1)}{2} \right) \left( \frac{2b}{2\ell+1} \right)^{s+\frac{1}{2}} \\
&\leq 2 \cosh(b/2) \sum_{\ell=0}^\infty \frac{(2\ell+1)}{\sqrt{2\pi}} \sqrt{\pi} e^{-\frac{b(2\ell+1)}{2}} \left[ \frac{\Gamma(2s+1)}{\Gamma(s+1)} 2^{-s-1/2} \right. \\
&\quad \times \left. \left( \frac{b(2\ell+1)}{2} \right)^{-s-1/2} + 2^{-1/2} \left( \frac{b(2\ell+1)}{2} \right)^{-1/2} \right] \left( \frac{2b}{2\ell+1} \right)^{s+\frac{1}{2}} \\
&= 2^s (1 + e^{-b}) \left[ \sum_{\ell=0}^\infty \frac{e^{-b\ell}}{(2\ell+1)^{2s}} \frac{\Gamma(2s+1)}{\Gamma(s+1)} + \sum_{\ell=0}^\infty \frac{e^{-b\ell}}{(2\ell+1)^s} b^s \right] < \infty.
\end{aligned}$$

The second equality follows by using (27). The inequality is based on (29). The convergence of the two series in the last step can be obtained by utilizing the ratio test. Hence,  $|h_2|$  is integrable with respect to the product measure of the counting measure and the Lebesgue measure. By Fubini's Theorem and (27), (26) becomes

$$E(\omega^{-s}) = \cosh(b/2) \sum_{\ell=0}^\infty (-1)^\ell \frac{(2\ell+1)}{\sqrt{2\pi}} 2K_{s+\frac{1}{2}} \left( \frac{b[2\ell+1]}{2} \right) \left( \frac{2b}{2\ell+1} \right)^{s+\frac{1}{2}}. \quad (31)$$

When  $\ell$  is even, applying (29) to (31), and when  $\ell$  is odd, applying (30) to (31), we obtain

$$E(\omega^{-s}) \leq 2 \cosh(b/2) \left\{ \sum_{\text{even } \ell} \frac{(2\ell+1)}{\sqrt{2\pi}} \sqrt{\pi} e^{-\frac{b(2\ell+1)}{2}} \left[ \frac{\Gamma(2s+1)}{\Gamma(s+1)} 2^{-s-1/2} \right. \right.$$

$$\begin{aligned}
& \times \left\{ \left( \frac{b(2\ell+1)}{2} \right)^{-s-1/2} + (b(2\ell+1))^{-1/2} \right\} \\
& - \sum_{\text{odd } \ell} \frac{(2\ell+1)}{\sqrt{2\pi}} \sqrt{\pi} e^{-\frac{b(2\ell+1)}{2}} 2^{-1/2} \left( \frac{b(2\ell+1)}{2} \right)^{-1/2} \Bigg\} \times \left( \frac{2b}{2\ell+1} \right)^{s+\frac{1}{2}} \\
& = e^{b/2} (1+e^{-b}) \left\{ \sum_{\text{even } \ell} (2\ell+1)^{-2s} 2^s e^{-b(\ell+1/2)} \frac{\Gamma(2s+1)}{\Gamma(s+1)} \right. \\
& \quad \left. + \sum_{\text{even } \ell} (2\ell+1)^{-s} (2b)^s e^{-b(\ell+1/2)} - \sum_{\text{odd } \ell} (2\ell+1)^{-s} (2b)^s e^{-b(\ell+1/2)} \right\} \\
& = (1+e^{-b}) b^s \sum_{\ell=0}^{\infty} (-e^{-b})^{\ell} (\ell+1/2)^{-s} + (1+e^{-b}) 2^{-s} \frac{\Gamma(2s+1)}{\Gamma(s+1)} \\
& \quad \times \sum_{\text{even } \ell} e^{-b\ell} (\ell+1/2)^{-2s} \\
& = (1+e^{-b}) b^s \Phi(-e^{-b}, s, 1/2) \\
& \quad + (1+e^{-b}) 2^{-s} \frac{\Gamma(2s+1)}{\Gamma(s+1)} \sum_{k=0}^{\infty} e^{-2bk} (2k+1/2)^{-2s} \\
& = (1+e^{-b}) \frac{b^s}{\Gamma(s)} \int_0^{\infty} \frac{t^{s-1} e^{-t/2}}{1+e^{-b-t}} dt + (1+e^{-b}) 2^{-s} \frac{\Gamma(2s+1)}{\Gamma(s+1)} \\
& \quad \times \sum_{k=0}^{\infty} e^{-2bk} (2k+1/2)^{-2s}, \tag{32}
\end{aligned}$$

where  $\Phi(\cdot)$  is the Lerch transcendent function.

For fixed  $s > 0$ , let

$$\begin{aligned}
f(b) &= (1+e^{-b}) \frac{b^s}{\Gamma(s)} \int_0^{\infty} \frac{t^{s-1} e^{-t/2}}{1+e^{-b-t}} dt - 2^s b^s \\
&= \frac{b^s}{\Gamma(s)} \left[ (1+e^{-b}) \int_0^{\infty} \frac{t^{s-1} e^{-t/2}}{1+e^{-b-t}} dt - \int_0^{\infty} t^{s-1} e^{-t/2} dt \right] \\
&= \frac{b^s e^{-b}}{\Gamma(s)} \int_0^{\infty} \frac{(1-e^{-t})}{1+e^{-b-t}} t^{s-1} e^{-t/2} dt.
\end{aligned}$$

Since  $(1-e^{-t})t^{s-1}e^{-t/2}/(1+e^{-b-t}) \leq t^{s-1}e^{-t/2}$ , which is integrable, by the Dominated Convergence Theorem (DCT), it follows that  $f(b)$  is a continuous function of  $b$ . Another application of DCT shows that

$$\begin{aligned}
\lim_{b \rightarrow \infty} \int_0^{\infty} \frac{1-e^{-t}}{1+e^{-b-t}} t^{s-1} e^{-t/2} dt &= \int_0^{\infty} (1-e^{-t}) t^{s-1} e^{-t/2} dt \\
&\leq \int_0^{\infty} t^{s-1} e^{-t/2} dt = 2^s \Gamma(s).
\end{aligned}$$

Hence,  $\lim_{b \rightarrow \infty} f(b) = 0$ . Since  $f(b)$  is a continuous function of  $b$ ,  $f(0) = 0$  and  $\lim_{b \rightarrow \infty} f(b) = 0$ , we can conclude that  $|f(b)|$  can be bounded by a positive

constant  $f_0$ , hence,

$$(1 + e^{-b}) \frac{b^s}{\Gamma(s)} \int_0^\infty \frac{t^{s-1} e^{-t/2}}{1 + e^{-b-t}} dt \leq 2^s b^s + f_0. \quad (33)$$

As for the second term in (32), we have

$$\begin{aligned} (1 + e^{-b}) 2^{-s} \frac{\Gamma(2s+1)}{\Gamma(s+1)} \sum_{k=0}^\infty e^{-2bk} (2k+1/2)^{-2s} \\ \leq (1 + e^{-b}) 2^{-s} \frac{\Gamma(2s+1)}{\Gamma(s+1)} [1/(e^{2b} - 1) + 4^s]. \end{aligned} \quad (34)$$

Here, the inequality is due to the fact  $(2k+1/2)^{-2s} \leq 1$  for  $k \geq 1$ . Note that for  $b \geq \varepsilon$ , where  $\varepsilon > 0$  is arbitrary, the upper bound of (34) becomes

$$(1 + e^{-b}) 2^{-s} \frac{\Gamma(2s+1)}{\Gamma(s+1)} [1/(e^{2b} - 1) + 4^s] \leq (1 + e^{-\varepsilon}) 2^{-s} \frac{\Gamma(2s+1)}{\Gamma(s+1)} [1/(e^{2\varepsilon} - 1) + 4^s].$$

Thus, combining (33) with the above result, from (32) we have for  $b \geq \varepsilon$ ,

$$E(\omega^{-s}) \leq 2^s b^s + f_0 + L(s, \varepsilon), \quad (35)$$

where  $L(s, \varepsilon) = (1 + e^{-\varepsilon}) 2^{-s} \frac{\Gamma(2s+1)}{\Gamma(s+1)} [1/(e^{2\varepsilon} - 1) + 4^s]$ .

Now, we consider  $0 < b < \varepsilon$ . Let  $k(b) \equiv E(\omega^{-s})$ , where  $\omega \sim PG(1, b)$ . Then,

$$\lim_{b \rightarrow 0} k(b) = \lim_{b \rightarrow 0} \cosh(b/2) \lim_{b \rightarrow 0} \int_0^\infty j(b, x) dx = \lim_{b \rightarrow 0} \int_0^\infty j(b, x) dx, \quad (36)$$

where

$$j(b, x) = x^{-s-\frac{3}{2}} \sum_{\ell=0}^\infty (-1)^\ell \frac{(2\ell+1)}{\sqrt{2\pi}} \exp \left[ -\frac{(2\ell+1)^2}{8x} - \frac{b^2}{2} x \right].$$

Note that

$$j(b, x) \leq (x^{-1} + 1) x^{-\frac{3}{2}} \sum_{\ell=0}^\infty (-1)^\ell \frac{(2\ell+1)}{\sqrt{2\pi}} \exp \left[ -\frac{(2\ell+1)^2}{8x} \right] = j(x), \text{ say.}$$

From (25), it follows that  $\int_0^\infty j(x) dx \leq 8C + 1$ . Then by the DCT, from (36), we have

$$\lim_{b \rightarrow 0} k(b) = \lim_{b \rightarrow 0} \int_0^\infty j(b, x) dx = k(0).$$

So,  $k(b) = E(\omega^{-s})$  is continuous at  $b = 0$ . Recall that  $E(\omega^{-s}) \leq 8C + 1$  for  $b = 0$  and  $0 < s \leq 1$ . Thus, there exists some small  $\varepsilon > 0$  such that  $E(\omega^{-s}) \leq 8C + 2$  for  $0 < b < \varepsilon$ . Combining this result with (35), we have  $E(\omega^{-s}) \leq 2^s b^s + L(s)$ , where  $L(s) = \max\{f_0 + L(s, \varepsilon), 8C + 2\}$ . Thus, part 1 is proved for  $a = 1$ .

Next, we prove the conclusion for  $a > 1$ . From Polson, Scott and Windle (2013), when  $\omega \sim PG(a, b)$ , we have

$$\omega \stackrel{d}{=} \frac{1}{2\pi^2} \sum_{\ell=1}^{\infty} \frac{g_{\ell}}{(\ell - 1/2)^2 + b^2/(4\pi^2)}, \quad (37)$$

where  $g_{\ell}$ 's are mutually independent  $\text{Gamma}(a, 1)$  random variables. Since  $a > 1$ ,  $g_{\ell} \stackrel{d}{=} \tilde{g}_{\ell} + g_{\ell}^*$ , where  $\tilde{g}_{\ell}$  and  $g_{\ell}^*$  are independent random variables following  $\text{Gamma}(a - 1, 1)$  and  $\text{Gamma}(1, 1)$ , respectively. Let  $x_1 = (1/[2\pi^2]) \sum_{\ell=1}^{\infty} g_{\ell}^* / [(\ell - 1/2)^2 + b^2/(4\pi^2)]$ . Then,  $x_1 \sim PG(1, b)$ . Thus, we have  $E(x_1^{-s}) \leq 2^s b^s + L(s)$ . Since for  $0 < s \leq 1$ ,  $E(\omega^{-s}) \leq E(x_1^{-s})$ , the same conclusion follows for  $\omega \sim PG(a, b)$ , where  $a > 1$ . Thus, the proof for part 1 is complete.

Next, we prove part 2. From (37), we have

$$\begin{aligned} E \omega^{-s} &= E \left[ \frac{1}{2\pi^2} \sum_{\ell=1}^{\infty} \frac{g_{\ell}}{(\ell - 1/2)^2 + b^2/(4\pi^2)} \right]^{-s} \\ &\leq E \left[ \frac{1}{2\pi^2} \frac{g_1}{(1 - 1/2)^2 + b^2/(4\pi^2)} \right]^{-s} \\ &= \left( \frac{\pi^2 + b^2}{2} \right)^s \int_0^{\infty} g_1^{-s} \frac{1}{\Gamma(a)} g_1^{a-1} \exp(-g_1) dg_1 \\ &= 2^{-s} (\pi^2 + b^2)^s \frac{\Gamma(a-s)}{\Gamma(a)}. \end{aligned}$$

Thus, the proof for part 2 is complete.  $\square$

**Lemma 3.** If  $\omega \sim PG(a, b)$ ,  $a > 0$ ,  $b \geq 0$ , then  $E\omega \leq a/4$ .

*Proof.* From (37), we have

$$\begin{aligned} E\omega &= E \left[ \frac{1}{2\pi^2} \sum_{\ell=1}^{\infty} \frac{g_{\ell}}{(\ell - 1/2)^2 + b^2/(4\pi^2)} \right] \\ &\leq E \left[ \frac{1}{2\pi^2} \sum_{\ell=1}^{\infty} \frac{g_{\ell}}{(\ell - 1/2)^2} \right] \\ &= \frac{1}{2\pi^2} \sum_{\ell=1}^{\infty} \frac{a}{(\ell - 1/2)^2} = aE\omega_2, \end{aligned} \quad (38)$$

where  $\omega_2 \sim PG(1, 0)$ . By Polson, Scott and Windle (2013),  $PG(1, 0) = J^*(1, 0)/4$ . From Devroye (2009), the density for  $J^*(1, 0)$  is

$$f^*(x) = \pi \sum_{\ell=0}^{\infty} (-1)^{\ell} (\ell + 1/2) \exp \left[ -\frac{(\ell + 1/2)^2 \pi^2 x}{2} \right],$$

then,

$$E J^*(1, 0) = \int_0^{\infty} x \pi \sum_{\ell=0}^{\infty} (-1)^{\ell} (\ell + 1/2) \exp \left[ -\frac{(\ell + 1/2)^2 \pi^2 x}{2} \right] dx$$

$$\begin{aligned}
&= \sum_{\ell=0}^{\infty} \int_0^{\infty} x \pi(-1)^{\ell} (\ell + 1/2) \exp \left[ -\frac{(\ell + 1/2)^2 \pi^2 x}{2} \right] dx \\
&= \sum_{\ell=0}^{\infty} \pi(-1)^{\ell} (\ell + 1/2) \Gamma(2) \left( \frac{1}{2} \left( \ell + \frac{1}{2} \right)^2 \pi^2 \right)^{-2} \\
&= \frac{4}{\pi^3} \sum_{\ell=0}^{\infty} \frac{(-1)^{\ell}}{(\ell + 1/2)^3} = \frac{4}{\pi^3} \Phi \left( -1, 3, \frac{1}{2} \right) = \frac{4}{\pi^3} \frac{\pi^3}{4} = 1,
\end{aligned}$$

where  $\Phi(\cdot)$  is the Lerch transcendent function and  $\Phi(-1, 3, \frac{1}{2}) = \pi^3/4$ . Note that, we achieve the second equality above by applying the following logic. Let  $h_3(x, \ell) = x \pi(-1)^{\ell} (\ell + \frac{1}{2}) \exp \left[ -\frac{(\ell + 1/2)^2 \pi^2 x}{2} \right]$ , then

$$\begin{aligned}
\sum_{\ell=0}^{\infty} \int_0^{\infty} |h_3(x, \ell)| dx &= \sum_{\ell=0}^{\infty} \int_0^{\infty} x \pi \left( \ell + \frac{1}{2} \right) \exp \left[ -\frac{(\ell + 1/2)^2 \pi^2 x}{2} \right] dx \\
&= \frac{4}{\pi^3} \sum_{\ell=0}^{\infty} \frac{1}{(\ell + 1/2)^3} < \infty,
\end{aligned}$$

Hence,  $h_3(x, \ell)$  is integrable with respect to the product measure of the counting measure and the Lebesgue measure. Thus, by Fubini's Theorem, the second equality follows. Consequently,  $E\omega_2 = E J^*(1, 0)/4 = 1/4$ . From (38), it follows  $E\omega \leq a/4$ .  $\square$

**Remark 3.** Wang and Roy (2018b) proved Lemma 2 in the special case when  $a = 1$ . Although their result is correct as stated, their proof has an error which can be repaired following the techniques used in the proof of Lemma 2 here. Lemma 3 for  $a = 1$  is also proved in Wang and Roy (2018b).

## Appendix C: Proof of Theorem 1

*Proof.* We first prove the geometric ergodicity of the  $\tilde{\Psi}$  chain by establishing a drift condition. We consider the following drift function

$$V(\eta) = \sum_{i=1}^n |m_i^{\top} \eta| + \sum_{j=1}^r (u_j^{\top} u_j)^{-c}, \quad (39)$$

where  $c \in (0, 1/2)$  to be determined later.

Since  $M$  has full rank,  $V(\eta) : \mathbb{R}^{p+q} \setminus N \rightarrow [0, \infty)$  is unbounded off compact sets. We prove that, for any  $\eta, \eta' \in \mathbb{R}^{p+q} \setminus N$ , there exist constants  $\rho \in [0, 1)$  and  $L > 0$  such that

$$E[V(\eta) \mid \eta'] = E\{E[V(\eta) \mid \omega, \tau, y] \mid \eta', y\} \leq \rho V(\eta') + L. \quad (40)$$

The first term in the drift function is  $\sum_{i=1}^n |m_i^{\top} \eta| = l^{\top} M \eta$ , where  $l = (l_1, l_2, \dots, l_n)$  is defined as  $l_i = 1$  if  $m_i^{\top} \eta \geq 0$ ,  $l_i = -1$  if  $m_i^{\top} \eta < 0$ . Since  $Q = 0$ , from (17)

we have

$$\begin{aligned}
& \mathbb{E} \left[ \sum_{i=1}^n |m_i^\top \eta| \middle| \omega, \tau, y \right] \\
&= l^\top M (M^\top \Omega(\omega) M + B(\tau))^{-1} M^\top \kappa \\
&\leq \sqrt{l^\top M (M^\top \Omega(\omega) M + B(\tau))^{-1} M^\top l} \sqrt{\kappa^\top M (M^\top \Omega(\omega) M + B(\tau))^{-1} M^\top \kappa} \\
&\leq \sqrt{l^\top M (M^\top \Omega(\omega) M)^{-1} M^\top l} \sqrt{\kappa^\top M (M^\top \Omega(\omega) M)^{-1} M^\top \kappa} \\
&= \sqrt{l^\top \Omega(\omega)^{-1/2} P_{\Omega(\omega)^{1/2} M} \Omega(\omega)^{-1/2} l} \sqrt{\kappa^\top M (M^\top \Omega(\omega) M)^{-1} M^\top \kappa} \\
&\leq \sqrt{\sum_{i=1}^n \frac{1}{\omega_i}} \sqrt{\kappa^\top M (M^\top \Omega(\omega) M)^{-1} M^\top \kappa}, \tag{41}
\end{aligned}$$

where the first inequality follows from the Cauchy-Schwarz inequality,  $P_{\Omega(\omega)^{1/2} M} \equiv \Omega(\omega)^{1/2} M (M^\top \Omega(\omega) M)^{-1} M^\top \Omega(\omega)^{1/2}$  is a projection matrix, and the third inequality follows from the fact that  $I \succeq P_{\Omega(\omega)^{1/2} M}$ . Recall that  $k_i = y_i - 1/2$ ,  $i = 1, \dots, n$ . Define,  $v_i = -2k_i m_i$  as the  $i^{\text{th}}$  row of an  $n \times (p+q)$  matrix  $V$ . Note that  $v_i v_i^\top = m_i m_i^\top$ ,  $i = 1, \dots, n$ . Since the conditions 3 and 4 in Theorem 1 are in force, by Lemma 3 in Wang and Roy (2018b), for the second part of (41), we have

$$\sqrt{\kappa^\top M (M^\top \Omega(\omega) M)^{-1} M^\top \kappa} = \sqrt{\frac{1}{4} \mathbf{1}^\top V (V^\top \Omega(\omega) V)^{-1} V^\top \mathbf{1}} \leq \sqrt{\frac{\rho_1}{4} \sum_{i=1}^n \frac{1}{\omega_i}}, \tag{42}$$

where  $\rho_1 \in [0, 1)$  is a constant. Applying (42) to (41), we have

$$\mathbb{E} \left[ \sum_{i=1}^n |m_i^\top \eta| \middle| \omega, \tau, y \right] \leq \frac{\sqrt{\rho_1}}{2} \sum_{i=1}^n \frac{1}{\omega_i}. \tag{43}$$

Next, we consider the inner expectation of the second term in the drift function (39). Note that, for  $c \in (0, 1/2)$ , we have

$$\begin{aligned}
& \mathbb{E} \left[ (u_j^\top u_j)^{-c} \middle| \omega, \tau, y \right] \\
&= \left( \sum_{i=1}^n \omega_i \text{tr}(Z^\top Z) + \tau_j \right)^c \mathbb{E} \left[ \left\{ u_j^\top \left( \sum_{i=1}^n \omega_i \text{tr}(Z^\top Z) + \tau_j \right) I_{q_j} u_j \right\}^{-c} \middle| \omega, \tau, y \right] \\
&\leq \left[ \left( \sum_{i=1}^n \omega_i \text{tr}(Z^\top Z) \right)^c + \tau_j^c \right] \mathbb{E} \left[ \left\{ u_j^\top (R_j \tilde{S}^{-1} R_j^\top)^{-1} u_j \right\}^{-c} \middle| \omega, \tau, y \right] \\
&\leq \frac{2^{-c} \Gamma(-c + q_j/2)}{\Gamma(q_j/2)} \left[ \left( \sum_{i=1}^n \omega_i \text{tr}(Z^\top Z) \right)^c + \tau_j^c \right], \tag{44}
\end{aligned}$$

where the first inequality follows from Lemma 1 and the fact that  $(a + b)^s \leq a^s + b^s$  for  $a > 0$ ,  $b > 0$ , and  $0 \leq s < 1$ . For the last inequality, note that, by (23), we have  $u_j \mid \omega, \tau, y \sim N(R_j(-\tilde{S}^{-1}\tilde{R}^\top X^\top \kappa + \tilde{S}^{-1}Z^\top \kappa), R_j\tilde{S}^{-1}R_j^\top)$  where  $R_j$  is a  $q_j \times q$  matrix consisting of 0's and 1's such that  $R_j u = u_j$ . Thus, given  $\omega, \tau, y$ ,  $(R_j\tilde{S}^{-1}R_j^\top)^{-\frac{1}{2}}u_j$  has a multivariate normal distribution with the identity covariance matrix. Hence, conditional on  $\omega, \tau, y$ , the distribution of  $u_j^\top (R_j\tilde{S}^{-1}R_j^\top)^{-1}u_j$  is  $\chi_{q_j}^2(w)$ , where  $w$  is some (unimportant) noncentrality parameter and  $q_j$  is the degrees of freedom for the relevant Chi-square distribution. Therefore, by Lemma 4 in Román and Hobert (2012), we have

$$\mathbb{E}\left[\left\{u_j^\top (R_j\tilde{S}^{-1}R_j^\top)^{-1}u_j\right\}^{-c} \mid \omega, \tau, y\right] \leq \frac{2^{-c}\Gamma(-c + q_j/2)}{\Gamma(q_j/2)}.$$

Applying the above result, the inequality in (44) is obtained.

Combining (43) and (44), from (39), we have

$$E[V(\eta) \mid \omega, \tau, y] \leq \frac{\sqrt{\rho_1}}{2} \sum_{i=1}^n \frac{1}{\omega_i} + \sum_{j=1}^r \frac{2^{-c}\Gamma(-c + q_j/2)}{\Gamma(q_j/2)} \left[ \left( \sum_{i=1}^n \omega_i \text{tr}(Z^\top Z) \right)^c + \tau_j^c \right]. \quad (45)$$

Next, we consider the outer expectation in (40). By Lemma 2, we have

$$\begin{aligned} \mathbb{E}\left[\frac{\sqrt{\rho_1}}{2} \sum_{i=1}^n \frac{1}{\omega_i} \mid \eta', y\right] &\leq \left[2 \sum_{i=1}^n |m_i^\top \eta'| + nL(1)\right] \frac{\sqrt{\rho_1}}{2} \\ &= \sqrt{\rho_1} \sum_{i=1}^n |m_i^\top \eta'| + \frac{nL(1)\sqrt{\rho_1}}{2}. \end{aligned} \quad (46)$$

For the outer expectation of the other terms on the right hand side of (45), we now consider the expectation for  $\tau_j^c$ . Recall from section 2.2 that  $\tau_j \mid \eta', y \stackrel{\text{ind}}{\sim} \text{Gamma}(a_j + q_j/2, b_j + u_j'^\top u_j'/2)$ ,  $j = 1, \dots, r$ . Then, it follows that

$$\mathbb{E}[\tau_j^c \mid \eta', y] = \frac{\Gamma(a_j + \frac{q_j}{2} + c)}{\Gamma(a_j + \frac{q_j}{2})} \left(b_j + \frac{1}{2}u_j'^\top u_j'\right)^{-c}.$$

Define,  $G_j(-c) = 2^c \Gamma(a_j + q_j/2 + c) / \Gamma(a_j + q_j/2)$ ,  $j = 1, 2, \dots, r$ . Hence,

$$\begin{aligned} \mathbb{E}[\tau_j^c \mid \eta', y] &= 2^{-c} G_j(-c) \left(b_j + \frac{u_j'^\top u_j'}{2}\right)^{-c} \\ &\leq G_j(-c) [(2b_j)^{-c} \mathbf{I}_{(0, \infty)}(b_j) + (u_j'^\top u_j')^{-c} \mathbf{I}_{\{0\}}(b_j)]. \end{aligned} \quad (47)$$

Also,

$$\sum_{j=1}^r \frac{2^{-c}\Gamma(q_j/2 - c)}{\Gamma(q_j/2)} G_j(-c) (u_j'^\top u_j')^{-c} \mathbf{I}_{\{0\}}(b_j) \leq \delta_1(c) \sum_{j=1}^r (u_j'^\top u_j')^{-c} \quad (48)$$

where  $\delta_1(c) = 2^{-c} \max_{j \in A} \frac{\Gamma(q_j/2-c)}{\Gamma(q_j/2)} G_j(-c) \geq 0$ . Recall that  $A = \{j \in \{1, 2, \dots, r\} : b_j = 0\}$ . From the condition 1 of Theorem 1, we have  $a_j < 0$  when  $b_j = 0$  (i.e. when  $j \in A$ ). From the proof of Proposition 2 in Román and Hobert (2012), it follows that, there exists  $c \in C \equiv (0, 1/2) \cap (0, \tilde{a})$ , where  $\tilde{a} = -\max_{j \in A} a_j$ , such that  $\delta_1(c) < 1$ .

Using (46), (47), and Jensen's inequality, from (45), we obtain

$$\begin{aligned} E[V(\eta) \mid \eta'] &= E\{E[V(\eta \mid \omega, \tau, y)] \mid \eta', y\} \\ &\leq \sqrt{\rho_1} \sum_{i=1}^n |m_i^\top \eta'| + \frac{nL(1)\sqrt{\rho_1}}{2} + \sum_{j=1}^r \frac{2^{-c}\Gamma(-c + q_j/2)}{\Gamma(q_j/2)} \\ &\quad \times \left\{ (\text{tr}(Z^\top Z) \sum_{i=1}^n E(\omega_i \mid \eta', y))^c + G_j(-c) \right. \\ &\quad \left. \times [(2b_j)^{-c} \mathbf{I}_{(0,\infty)}(b_j) + (u_j'^\top u_j')^{-c} \mathbf{I}_{\{0\}}(b_j)] \right\} \\ &\leq \sqrt{\rho_1} \sum_{i=1}^n |m_i^\top \eta'| + \delta_1(c) \sum_{j=1}^r (u_j'^\top u_j')^{-c} + L \leq \rho V(\eta') + L, \end{aligned} \quad (49)$$

where the second inequality is due to (48) and Lemma 3. Here,  $\rho = \max\{\sqrt{\rho_1}, \delta_1(c)\}$ , and

$$\begin{aligned} L &= \frac{nL(1)\sqrt{\rho_1}}{2} + \sum_{j=1}^r \frac{2^{-c}\Gamma(-c + q_j/2)}{\Gamma(q_j/2)} \left( \frac{\text{tr}(Z^\top Z)n}{4} \right)^c \\ &\quad + \sum_{j=1}^r \frac{2^{-c}\Gamma(-c + q_j/2)}{\Gamma(q_j/2)} G_j(-c) (2b_j)^{-c} \mathbf{I}_{(0,\infty)}(b_j). \end{aligned}$$

Recall that  $\rho_1, \delta_1(c) \in [0, 1)$ , thus  $\rho < 1$ . Consequently, (40) holds. In addition, we can show that  $\Psi$  chain is a Feller Markov chain by the following steps. Let  $K(\eta', \cdot)$  denote the Mtf corresponding to (21). To prove that the  $\Psi$  chain is a Feller Markov chain is to show that  $K(\eta', A)$  is a lower semi-continuous function on  $\mathbb{R}^{p+q} \setminus N$  for each fixed open set  $A$  in  $\mathbb{R}^{p+q} \setminus N$ . For a sequence  $\{\eta'_m\}$ , using (21), Fatou's Lemma and independence of the conditional distribution of  $\omega$  and  $\tau$  given  $(\eta', y)$ , we have

$$\begin{aligned} \liminf_{m \rightarrow \infty} K(\eta'_m, A) &= \liminf_{m \rightarrow \infty} \int_A \tilde{k}(\eta \mid \eta'_m) d\eta \\ &= \liminf_{m \rightarrow \infty} \int_A \int_{\mathbb{R}_+^r} \int_{\mathbb{R}_+^n} \pi(\eta \mid \omega, \tau, y) \pi(\omega, \tau \mid \eta'_m, y) d\omega d\tau d\eta \\ &\geq \int_A \int_{\mathbb{R}_+^r} \int_{\mathbb{R}_+^n} \pi(\eta \mid \omega, \tau, y) \liminf_{m \rightarrow \infty} \pi(\omega, \tau \mid \eta'_m, y) d\omega d\tau d\eta \\ &= \int_A \int_{\mathbb{R}_+^r} \int_{\mathbb{R}_+^n} \pi(\eta \mid \omega, \tau, y) \liminf_{m \rightarrow \infty} [\pi(\omega \mid \eta'_m, y) \pi(\tau \mid \eta'_m, y)] d\omega d\tau d\eta. \end{aligned}$$

Considering the conditions 1 and 2 in Theorem 1, for any fixed  $(\eta, \omega, y)$ , both  $\pi(\omega \mid \eta'_m, y)$  and  $\pi(\tau \mid \eta'_m, y)$  are continuous functions on  $\mathbb{R}^{p+q} \setminus N$ . Hence, if  $\eta'_m \rightarrow \eta'$ , then  $\liminf_{m \rightarrow \infty} K(\eta'_m, A) \geq K(\eta', A)$ , and we can conclude that the  $\Psi$  chain is a Feller Markov chain. Thus, by Lemma 15.2.8 in Meyn and Tweedie (1993), GE of the  $\tilde{\Psi}$  chain is proved.

Next, using the similar techniques as in Wang and Roy (2018a) and (Román, 2012, Lemma 12), we can establish that the GE of the original chain  $\Psi$  follows from that of  $\tilde{\Psi}$ . We include a proof here for completeness. Let  $X \equiv \mathbb{R}^{p+q}$ ,  $\tilde{X} \equiv \mathbb{R}^{p+q} \setminus N$ . Let  $K$  and  $\tilde{K}$  denote the Mtf's of  $\Psi$  and  $\tilde{\Psi}$  chains respectively. Also, since the Lebesgue measure of  $N$  is zero,  $\tilde{K}(x, B) = K(x, B)$ , for any  $x \in \tilde{X}$  and  $B \in \mathcal{B}_{\tilde{X}} = \{\tilde{X} \cap A : A \in \mathcal{B}_X\}$ , where  $\mathcal{B}_X$  denotes the Borel  $\sigma$ -algebra of  $\mathbb{R}^{p+q}$  and  $\mathcal{B}_{\tilde{X}}$  denotes the Borel  $\sigma$ -algebra of  $\mathbb{R}^{p+q} \setminus N$ , respectively.

Let  $\mu$  and  $\tilde{\mu}$  be the Lebesgue measures on  $X$  and  $\tilde{X}$ , respectively. As the Mtds are strictly positive for the two chains, the  $\Psi$  chain is  $\mu$ -irreducible and the  $\tilde{\Psi}$  chain is  $\tilde{\mu}$ -irreducible. Both chains are aperiodic. Note that,  $\mu$  and  $\tilde{\mu}$  are also the maximal irreducibility measures of  $\Psi$  and  $\tilde{\Psi}$  chains, respectively. By Theorem 15.0.1 in Meyn and Tweedie (1993), since we have established the GE of the  $\tilde{\Psi}$  chain, there exists a  $v$ -petite set  $C \in \mathcal{B}_{\tilde{X}}$ ,  $\rho_C < 1$ ,  $M_C < \infty$  and  $\tilde{K}^\infty(C) > 0$  such that  $\tilde{\mu}(C) > 0$  and

$$\left| \tilde{K}^m(x, C) - \tilde{K}^\infty(C) \right| < M_C \rho_C^m, \quad (50)$$

for all  $x \in C$ . Also, it can be shown that

$$K^m(x, B) = \tilde{K}^m(x, B \cap \tilde{X}), \quad (51)$$

for any  $x \in \tilde{X}$  and  $B \in \mathcal{B}_{\tilde{X}}$ . Note that,  $K^m$  and  $\tilde{K}^m$  indicate the corresponding  $m$ -step Mtf's. Thus, for  $x \in C$ ,  $K^m(x, C) = \tilde{K}^m(x, C)$ . Then, (50) becomes  $\left| K^m(x, C) - \tilde{K}^\infty(C) \right| < M_C \rho_C^m$ . Since  $\mu(N) = 0$ , we have  $\mu(C) = \tilde{\mu}(C)$ . Recall that  $\tilde{\mu}(C) > 0$ , thus  $\mu(C) > 0$ . Note that,  $C$  is  $v$ -petite for the  $\tilde{\Psi}$  chain, then for all  $x \in C$  and  $B \in \mathcal{B}_{\tilde{X}}$ ,

$$\sum_{m=0}^{\infty} \tilde{K}^m(x, B) a(m) \geq v(B), \quad (52)$$

where  $v$  is a nontrivial measure on  $\mathcal{B}_{\tilde{X}}$  and  $a(m)$  is a mass function on  $\{0, 1, 2, \dots\}$ . It can be shown that a nontrivial measure on  $\mathcal{B}_X$ , which is

$$v^*(\cdot) = v(\cdot \cap \tilde{X}), \quad (53)$$

is well defined. Then for any  $x \in C$  and any  $B \in \mathcal{B}_X$ , using (51), (52) and (53), we have

$$\sum_{m=0}^{\infty} K^m(x, B) a(m) = \sum_{m=0}^{\infty} \tilde{K}^m(x, B \cap \tilde{X}) a(m) \geq v(B \cap \tilde{X}) = v^*(B).$$

Hence,  $C$  is also a petite set for the  $\Psi$  chain. Applying Theorem 15.0.1 in Meyn and Tweedie (1993) again, GE of the  $\Psi$  chain is proved. Hence, we show that GE of  $\tilde{\Psi}$  implies that of the original chain  $\Psi$ .  $\square$

## Acknowledgments

The authors thank the editor and two anonymous reviewers for several helpful comments and suggestions that led to an 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](#)
- 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., SHAO, Q.-M. and XU, D. (2002). Necessary and sufficient conditions on the propriety of posterior distributions for generalized linear mixed models. *Sankhyā: The Indian Journal of Statistics, Series A* 57–85. [MR1968375](#)
- CHIB, S. and RAMAMURTHY, S. (2010). Tailored randomized block MCMC methods with application to DSGE models. *Journal of Econometrics* **155** 19–38. [MR2592847](#)
- 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](#)
- CORTEZ, P. and SILVA, A. M. G. (2008). Using data mining to predict secondary school student performance. In *Proceedings of 5th Annual Future Business Technology Conference* (A. BRITO and J. TEIXEIRA, eds.) 5–12. EUROSIS-ETI.
- DEVROYE, L. (2009). On exact simulation algorithms for some distributions related to Jacobi theta functions. *Statistics & Probability Letters* **79** 2251–2259. [MR2591982](#)
- EVANGELOU, E. and ROY, V. (2019). Estimation and prediction for spatial generalized linear mixed models with parametric links via reparameterized importance sampling. *Spatial statistics* **29** 289–315. [MR3903699](#)
- FRÜHWIRTH-SCHNATTER, S. and FRÜHWIRTH, R. (2010). Data augmentation and MCMC for binary and multinomial logit models. In *Statistical Modelling and Regression Structures* 111–132. Springer. [MR2664631](#)
- HOLMES, C. C. and HELD, L. (2006). Bayesian auxiliary variable models for binary and multinomial regression. *Bayesian Analysis* **1** 145–168. [MR2227368](#)
- 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](#)
- 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. [MR1287609](#)

- OLVER, F. W. J., LOZIER, D. W., F., B. R. and CLARK, C. W. (2010). *NIST Handbook of Mathematical Functions*. Cambridge University Press. [MR2655350](#)
- 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. and ROSENTHAL, J. S. (1997). Geometric ergodicity and hybrid Markov chains. *Electronic Communications in Probability* **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](#)
- ROBERTS, G. O. and ROSENTHAL, J. S. (2004). General state space Markov chains and MCMC algorithms. *Probability Surveys* **1** 20–71. [MR2095565](#)
- ROBERTS, G. O. and SAHU, S. K. (1997). Updating schemes, correlation structure, blocking and parameterization for the Gibbs sampler. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **59** 291–317. [MR1440584](#)
- 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](#)
- 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. (2020). Convergence Diagnostics for Markov Chain Monte Carlo. *Annual Review of Statistics and Its Application* **7** 387–412. [MR4104198](#)
- 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](#)
- TUREK, D., DE VALPINE, P., PACIOREK, C. J. and ANDERSON-BERGMAN, C. (2017). Automated parameter blocking for efficient Markov chain Monte Carlo sampling. *Bayesian Analysis* **12** 465–490. [MR3620741](#)
- 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](#)
- VATS, D., FLEGAL, J. M. and JONES, G. L. (2018). Strong consistency of multivariate spectral variance estimators in Markov chain Monte Carlo. *Bernoulli* **24** 1860–1909. [MR3757517](#)
- VATS, D., FLEGAL, J. M. and JONES, G. L. (2019). Multivariate output analysis for Markov chain Monte Carlo. *Biometrika* **106** 321–337. [MR3949306](#)

- WANG, X. and ROY, V. (2018a). Convergence analysis of the block Gibbs sampler for Bayesian probit linear mixed models with improper priors. *Electronic Journal of Statistics* **12** 4412–4439. [MR3892344](#)
- WANG, X. and ROY, V. (2018b). Geometric ergodicity of Pólya-Gamma Gibbs sampler for Bayesian logistic regression with a flat prior. *Electronic Journal of Statistics* **12** 3295–3311. [MR3861283](#)
- WANG, X. and ROY, V. (2018c). Analysis of the Pólya-Gamma block Gibbs sampler for Bayesian logistic linear mixed models. *Statistics & Probability Letters* **137** 251–256. [MR3776224](#)