

# Bayesian Covariance Selection in Generalized Linear Mixed Models

Bo Cai and David B. Dunson <sup>1</sup>

SUMMARY. The generalized linear mixed model (GLMM), which extends the generalized linear model (GLM) to incorporate random effects characterizing heterogeneity among subjects, is widely used in analyzing correlated and longitudinal data. Although there is often interest in identifying the subset of predictors that have random effects, random effects selection can be challenging, particularly when outcome distributions are non-normal. This article proposes a fully Bayesian approach to the problem of simultaneous selection of fixed and random effects in GLMMs. Integrating out the random effects induces a covariance structure on the multivariate outcome data, and an important problem which we also consider is that of covariance selection. Our approach relies on variable selection-type mixture priors for the components in a special LDU decomposition of the random effects covariance. A stochastic search MCMC algorithm is developed, which relies on Gibbs sampling, with Taylor series expansions used to approximate intractable integrals. Simulated data examples are presented for different exponential family distributions, and the approach is applied to discrete survival data from a time-to-pregnancy study.

KEY WORDS: Bayes factor; Latent variables; Marginal likelihood; MCMC algorithm; Random effects; Stochastic search; Taylor series; Variable selection.

---

<sup>1</sup>Bo Cai is Research Fellow, Biostatistics Branch, NIEHS, P.O.BOX 12233 MD A3-03, Research Triangle Park, NC 27709, USA (E-mail: cai@niehs.nih.gov); and David B. Dunson is senior investigator, Biostatistics Branch, NIEHS, P.O.BOX 12233 MD A3-03, Research Triangle Park, NC 27709, USA (E-mail: dunson1@niehs.nih.gov). The authors thank Beth Gladen and Grace Kissling for helpful comments and suggestions.

## 1. Introduction

With improvements in computation permitting routine implementation, the generalized linear mixed model (GLMM) has become very widely used in analyses of correlated and longitudinal data (McCulloch and Searle, 2001). Analogous to the mixed model (Laird and Ware, 1982) extension of the linear model, the GLMM extends the generalized linear model (GLM) to incorporate random effects characterizing heterogeneity among subjects or clusters. By integrating out the random effects, one can induce a dependency structure on the multiple responses, and hence GLMMs provide a convenient framework for modeling of multivariate non-Gaussian data. However, many complications arise in the non-Gaussian case, since integrals involved in marginalizing out the random effects do not have simple closed forms. This leads to some difficulties in model fitting and inferences on the fixed effects regression coefficients, problems addressed by Schall (1991), Zeger and Karim (1991), Breslow and Clayton (1993), McGilchrist (1994), and McCulloch (1997) among others. Greater challenges arise when interest instead focuses on selection of predictors to be included in the fixed and random effects components of the model, and when covariance structure modeling is the focus.

As motivation, we consider data from an epidemiologic study of time to pregnancy (Rowland et al., 1992). In this study, dental assistants completed a demographic and exposure history questionnaire, while also providing information on the number of menstrual cycles during which the woman was having noncontracepting sexual intercourse before the most recent pregnancy. Time to pregnancy (TTP) is a discrete event time, which can be analyzed using the following GLMM:

$$\Pr(T_i = t | T_i \geq t, \mathbf{x}_{it}, \mathbf{z}_{it}) = \frac{\exp(\mathbf{x}_{it}'\boldsymbol{\beta} + \mathbf{z}_{it}'\boldsymbol{\zeta}_i)}{1 + \exp(\mathbf{x}_{it}'\boldsymbol{\beta} + \mathbf{z}_{it}'\boldsymbol{\zeta}_i)}, \quad (1)$$

where  $T_i$  is the TTP for woman  $i$ ,  $\mathbf{x}_{it}$  and  $\mathbf{z}_{it}$  are vectors of predictors that may vary from cycle to cycle,  $\boldsymbol{\beta}$  are fixed effect regression coefficients,  $\boldsymbol{\zeta}_i \sim N(\mathbf{0}, \boldsymbol{\Sigma})$  are random effects for woman  $i$ , and  $\boldsymbol{\Sigma}$  is a covariance matrix. If the predictors to be included in the fixed ( $\mathbf{x}_{it}$ ) and random ( $\mathbf{z}_{it}$ ) effects components are known, then one can fit (1) in standard software packages (e.g., SAS or WinBUGS) to obtain estimates of  $\boldsymbol{\beta}$  and  $\boldsymbol{\Sigma}$  and to assess hypotheses on the fixed effects, such as  $H_{0l} : \beta_l = 0$  (the  $l$ th predictor has no effect on fecundability). However, it is also of interest to

determine which factors vary in their effects across women. For example, do the effects of aging, recent oral contraceptive use, and smoking vary? Addressing such hypotheses is equivalent to assessing whether random effects can be excluded from the model by effectively setting certain elements of  $\Sigma$  to 0, a difficult problem given the constraints on  $\Sigma$ .

Potentially, one can select a preferred GLMM by repeatedly fitting the model for all possible choices of  $\mathbf{x}_{it}$  and  $\mathbf{z}_{it}$  and then applying a standard criterion, such as the AIC or BIC. Such an approach is not feasible unless the number of candidate predictors is modest, and there is no general consensus on what the penalty for model complexity should be in a model with random effects. In addition, it is often not enough to say which model is preferred, one wants to report a weight of evidence (e.g., that the effect of smoking on fecundability varies among women). To assess whether one or more random effects should be included in the model, several authors have proposed frequentist score tests (Commenges and Jacqmin-Gadda, 1997; Lin, 1997; Hall and Praestgaard, 2001). In the Bayesian literature, Albert and Chib (1997) proposed an approach for testing whether a random intercept should be included, Sinharay and Stern (2001) developed a more general approach for calculating Bayes factors for comparing GLMMs, and Chen et al. (2003) proposed a class of informative priors for model selection in GLMMs. These methods focus on comparing two models at a time, and do not provide a general approach for searching for promising subsets of candidate predictors.

In the setting of linear mixed models for normal data, Chen and Dunson (2003) proposed a Bayesian approach for random effects selection based on using variable selection priors for the components in a special decomposition of the random effects covariance. Related approaches have been used in graphical (or covariance structure) modeling for multivariate normal data (refer to Wong, Carter and Kohn, 2003; Liechty, Liechty and Muller, 2004 for recent references). Bayesian variable selection in conventional GLMs has also received a lot of interest in the literature. Raftery (1996) proposed an approximate Bayes factor approach, Meyer and Laud (2002) considered predictive variable selection, Nott and Leonte (2004) developed an innovative sampling algorithm and Ntzoufras, Dellaportas, and Foster (2003) developed methods for joint variable and link selection.

In order to implement Bayesian selection of fixed and random effects while also considering covariance structure modeling, we first choose variable selection-type mixture priors for the fixed effects regression coefficients and the parameters in a special LDU decomposition of the random effects covariance proposed by Chen and Dunson (2003). These priors allow fixed effects to drop out of the model by placing probability mass on  $\beta_l = 0$ . In addition, following a related approach to Albert and Chib (1997) and Chen and Dunson (2003), we assign positive probability to random effects having 0 variance to effectively move between the full model with random effects for every predictor and submodels excluding one or more random effects. This prior specification has convenient computational properties, which is important given the potentially large number of models under consideration. In particular, given the conditional model probabilities, which can be approximated using Taylor series expansions, Gibbs sampling can be implemented as in typical GLMs using adaptive rejection sampling.

Outside of the realm of normal linear models, it is typically the case that Bayesian model selection requires the calculation of normalizing constants, which do not have closed form expressions. For this reason, many approaches have relied on approximations to intractable integrals, commonly using Laplace and other Taylor series approaches. The method proposed in this paper is based on the idea of using limited analytic approximations together with Markov chain Monte Carlo (MCMC) sampling. Similar ideas were implemented previously by Raftery, Madigan and Volinsky (1996) in the context of model averaging in survival analysis, and Chipman, George and McCulloch (2002, 2003) in implementing analyses of treed GLMs.

Section 2 describes the model and prior specification. Section 3 outlines the algorithm for posterior computation. Section 4 illustrates the approach using several simulated data examples. Section 5 applies the method to the time to pregnancy application, and Section 6 discusses the results.

## 2. The Model

### 2.1 Generalized Linear Mixed Models

For observation  $j$  ( $j = 1, \dots, n_i$ ) from subject  $i$  ( $i = 1, \dots, n$ ), let  $y_{ij}$  denote the response variable,

let  $\mathbf{x}_{ij}$  denote a  $p \times 1$  vector of candidate predictors, and let  $\mathbf{z}_{ij}$  denote a  $q \times 1$  vector of candidate predictors. The elements of  $\mathbf{y}_i = (y_{i1}, \dots, y_{i,n_i})'$  are assumed to be conditionally-independent random variables from a simple exponential family, so that the density function of  $y_{ij}$  given  $\mathbf{x}_{ij}$ ,  $\mathbf{z}_{ij}$  and random-effects  $\boldsymbol{\zeta}_i = (\zeta_{i1}, \dots, \zeta_{iq})'$  is expressed as:

$$\pi(y_{ij} | \mathbf{x}_{ij}, \mathbf{z}_{ij}, \boldsymbol{\zeta}_i) = \exp \left\{ \frac{y_{ij}\theta_{ij} - b(\theta_{ij})}{a_{ij}(\phi)} + c(y_{ij}, \phi) \right\}, \quad (2)$$

where  $\theta_{ij}$  is a location parameter,  $\phi$  is a scalar dispersion parameter, and  $a_{ij}(\cdot), b(\cdot), c(\cdot)$  are known functions, with  $a_{ij}(\phi)$  typically expressed as  $\phi/w_{ij}$ , where  $w_{ij}$  is a known weight. The canonical parameter  $\theta_{ij}$  is related to the linear predictor  $\eta_{ij} = \mathbf{x}_{ij}'\boldsymbol{\beta} + \mathbf{z}_{ij}'\boldsymbol{\zeta}_i$  through a monotone differentiable link function  $h(\cdot)$ , so that  $\theta_{ij} = h(\eta_{ij})$ , where  $\boldsymbol{\beta}$  is a  $p \times 1$  vector of fixed effect regression coefficients (referred to as fixed since the coefficients are constant for all subjects), and  $\boldsymbol{\zeta}_i \sim N_q(\mathbf{0}, \boldsymbol{\Sigma})$  is a  $q \times 1$  vector of subject-specific random effects with covariance matrix  $\boldsymbol{\Sigma}$ . In this initial specification, we assume that all the candidate predictors are included to define a full model. The resulting conditional mean and variance of  $y_{ij}$  are as follows:

$$\begin{aligned} \mu_{ij} &= E(y_{ij} | \mathbf{x}_{ij}, \mathbf{z}_{ij}, \boldsymbol{\zeta}_i) = b'(\theta_{ij}), \\ V_{ij} &= V(y_{ij} | \mathbf{x}_{ij}, \mathbf{z}_{ij}, \boldsymbol{\zeta}_i) = b''(\theta_{ij})\phi/w_{ij}, \end{aligned}$$

where we focus on the case in which  $a_{ij}(\phi) = \phi/w_{ij}$ .

Heterogeneity among subjects is accommodated by allowing the regression coefficients, and hence the linear predictor conditional on the covariates, to vary. When  $\mathbf{z}_{ij}$  is a subvector of  $\mathbf{x}_{ij}$ , the full model allows the regression coefficients for the covariates included in  $\mathbf{z}_{ij}$  to vary among subjects, while assuming that the remaining coefficients are fixed for all subjects. Integrating out the random effects  $\boldsymbol{\zeta}_i$ , this specification induces a correlation structure in the multiple observations from a subject. In particular, integrating out  $\boldsymbol{\zeta}_i$ , we have

$$\rho(y_{ij}, y_{ij'} | \mathbf{x}_{ij}, \mathbf{z}_{ij}) = \frac{\int y_{ij} y_{ij'} \pi(y_{ij} | \mathbf{x}_{ij}, \mathbf{z}_{ij}, \boldsymbol{\zeta}_i) \pi(y_{ij'} | \mathbf{x}_{ij'}, \mathbf{z}_{ij'}, \boldsymbol{\zeta}_i) \pi(\boldsymbol{\zeta}_i) dy_{ij} dy_{ij'} d\boldsymbol{\zeta}_i - \mu_{ij} \mu_{ij'}}{\sqrt{V_{ij} V_{ij'}}}, \quad (3)$$

which is not available in closed form except in the normal linear case. However, it is easy to see that the shared dependency on the random effects  $\boldsymbol{\zeta}_i$  in the models for  $y_{ij}$  and  $y_{ij'}$  induces correlation in these outcomes.

Let  $\mathbf{y} = (\mathbf{y}'_1, \dots, \mathbf{y}'_n)'$ ,  $\mathbf{y}_i = (y_{i1}, \dots, y_{in_i})'$ ,  $\mathbf{X} = (\mathbf{X}'_1, \dots, \mathbf{X}'_n)'$ ,  $\mathbf{X}_i = (\mathbf{x}_{i1}, \dots, \mathbf{x}_{in_i})'$ ,  $\mathbf{Z} = \text{diag}(\mathbf{Z}_1, \dots, \mathbf{Z}_n)$ ,  $\mathbf{Z}_i = (\mathbf{z}_{i1}, \dots, \mathbf{z}_{in_i})'$ , and  $\boldsymbol{\zeta} = (\boldsymbol{\zeta}'_1, \dots, \boldsymbol{\zeta}'_n)'$ . The joint distribution of responses  $\mathbf{y}$  and random effects  $\boldsymbol{\zeta}$  conditionally on the predictors  $\mathbf{X}$  and  $\mathbf{Z}$  is of the form

$$\pi(\mathbf{y}, \boldsymbol{\zeta} | \boldsymbol{\beta}, \phi, \boldsymbol{\Sigma}, \mathbf{X}, \mathbf{Z}) = \prod_{i=1}^n \prod_{j=1}^{n_i} \pi(y_{ij} | \eta_{ij}, \phi) \pi(\boldsymbol{\zeta}_i | \boldsymbol{\Sigma}), \quad (4)$$

where  $\pi(\boldsymbol{\zeta}_i | \boldsymbol{\Sigma}) = (2\pi)^{-\frac{q}{2}} |\boldsymbol{\Sigma}|^{-\frac{1}{2}} \exp\{-\frac{1}{2} \boldsymbol{\zeta}'_i \boldsymbol{\Sigma}^{-1} \boldsymbol{\zeta}_i\}$ . To facilitate our further development, expression (4) can also be written as

$$\pi(\mathbf{y}, \boldsymbol{\zeta} | \boldsymbol{\beta}, \phi, \boldsymbol{\Sigma}, \mathbf{X}, \mathbf{Z}) = \exp\{[\mathbf{y}' h(\boldsymbol{\eta}) - \mathbf{b}'(h(\boldsymbol{\eta})) \mathbf{1}_N] / \mathbf{a}' + \mathbf{c}'(\mathbf{y}, \phi) \mathbf{1}_N\} \pi(\boldsymbol{\zeta} | \boldsymbol{\Sigma}), \quad (5)$$

where  $\boldsymbol{\eta} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\zeta}$  and  $\mathbf{1}_N$  is an  $N \times 1$  vector of ones, where  $N = \sum_{i=1}^n n_i$ .

The specification of the model is completed by choosing a particular exponential family distribution for the conditional distribution of  $y_{ij}$  (e.g., Bernoulli) with a given link function (e.g., logistic). We will assume that the distribution and link function are known, but will account for uncertainty in the elements of  $\mathbf{x}_{ij}$  and  $\mathbf{z}_{ij}$  to be included in the model, as well as the covariance structure in the  $\boldsymbol{\zeta}_i$ 's. In particular, let  $\mathbf{x}_{ij}^{(M)}$ ,  $\mathbf{z}_{ij}^{(M)}$ ,  $\boldsymbol{\beta}^{(M)}$ ,  $\boldsymbol{\zeta}_i^{(M)}$ , and  $\boldsymbol{\Sigma}^{(M)}$  denote the fixed effect predictors, random effect predictors, fixed effect coefficients, random effects, and random effects covariance, respectively, for model  $M$ , which is specified as  $\eta_{ij}^{(M)} = \mathbf{x}_{ij}^{(M)'} \boldsymbol{\beta}^{(M)} + \mathbf{z}_{ij}^{(M)'} \boldsymbol{\zeta}_i^{(M)}$ , with the dispersion parameter, link function, and distributional form assumed common to the different models  $M \in \mathcal{M}$ . The predictors  $\mathbf{x}_{ij}^{(M)}$  consist of a  $p_M \leq p$  subset of  $\mathbf{x}_{ij}$ , while  $\mathbf{z}_{ij}^{(M)}$  is a  $q_M \leq q$  subset of  $\mathbf{z}_{ij}$ . In addition,  $\boldsymbol{\zeta}_i^{(M)} \sim N(\mathbf{0}, \boldsymbol{\Sigma}^{(M)})$  is a  $q_M \times 1$  vector of random effects, with  $q_M \times q_M$  covariance matrix  $\boldsymbol{\Sigma}^{(M)}$ , which can have zero off-diagonal elements corresponding to conditional independence relationships in the random effects included. The model space,  $\mathcal{M}$ , includes all possible combinations of subsets of  $\mathbf{x}_{ij}$  and  $\mathbf{z}_{ij}$  and zero off-diagonal elements of the random effects covariance matrices corresponding to these subsets. Hence, the total number of models is  $2^p \sum_{k=0}^q \binom{q}{k} 2^{\frac{1}{2}(q-k)(q-k-1)}$ .

## 2.2 Motivation

Our goal is to select good models from among the different possibilities for  $M$ . Our approach will rely on a stochastic search-type algorithm in which we embed the different candidate models within

the full model, and then drop out terms. Fixed effect predictors will be dropped out by using the common strategy of setting their regression coefficients to 0, while random effect selection involves a slightly more involved approach based on setting parameters in a decomposition of the random effects covariance equal to 0. To implement this strategy, we choose mixture priors with positive probability mass at zero for the fixed effects coefficients and the parameters in a decomposition of the random effects covariance.

In Bayesian literature, it is common practice to use decompositions in specifying priors for covariance matrices. For example, Daniels and Zhao (2003) used a special Cholesky decomposition to model changes in the random effects covariance over time. A related decomposition approach was considered by Daniels and Pourahmadi (2002). Daniels and Kass (1999) instead considered spectral decomposition. Following Chen and Dunson (2003), we use a special LDU decomposition, which has certain advantages over more commonly used Cholesky and spectral decompositions in model selection settings due to the conditionally linear structure. The decomposition contains a diagonal matrix with elements proportional to the standard deviations of the random effects and a lower triangular matrix related to the correlations among the random effects. Since 0 values for the standard deviations effectively correspond to random effects being excluded from the model, we can utilize mixture priors with mass at 0 for random effects selection. In addition, unlike Chen and Dunson (2003), we also allow zero off-diagonal elements through mixture priors for elements of the lower triangular matrix of the decomposition. Effectively, the prior allows movement between models of different dimension, with the covariance matrix of the random effects in each of these models being positive semi-definite. The details are given in the next section.

### 2.3 Reparameterization and Mixture Prior Specification

Let  $\sigma_{kk}$ ,  $k = 1, \dots, q$ , denote the diagonal entries of the symmetric random effects covariance  $\Sigma$ , and let  $\sigma_{mk} = \sigma_{km}$  denote the off-diagonal entries for  $m = k + 1, \dots, q, k = 1, \dots, q - 1$ . The covariance  $\Sigma$  may be factorized as

$$\Sigma = \Lambda \Gamma \Gamma' \Lambda,$$

where  $\mathbf{\Lambda} = \text{diag}(\lambda_1, \dots, \lambda_q)$ , with  $\lambda_k \geq 0$  for  $k = 1, \dots, q$ , and  $\mathbf{\Gamma}$  denotes lower triangular matrix,

$$\begin{pmatrix} 1 & & & \\ \gamma_{21} & 1 & & \\ \vdots & \vdots & \ddots & \\ \gamma_{q,1} & \gamma_{q,2} & \cdots & 1 \end{pmatrix}.$$

After some algebra, the elements of  $\mathbf{\Sigma}$  can be expressed as

$$\sigma_{kl} = \lambda_k \lambda_l \left( \gamma_{r_2, r_1} + \sum_{s=1}^{r_1-1} \gamma_{ks} \gamma_{ls} \right), \quad \text{for } k, l = 1, \dots, q, \quad (6)$$

where  $r_1 = \min(k, l)$ ,  $r_2 = \max(k, l)$ . This expression assures the positive semi-definite constraint on  $\mathbf{\Sigma}$  when  $\lambda_k > 0$  for all  $k$ . When  $\lambda_k = 0$ , all the elements of row  $k$  and column  $k$  of  $\mathbf{\Sigma}$  are zeros so that the submatrix of  $\mathbf{\Sigma}$  excluding row  $k$  and column  $k$  is still positive semi-definite. Such submatrices correspond to different choices of  $M$ , with each  $\mathbf{\Sigma}$  positive semi-definite. In this way, random effects are allowed to effectively drop out of the model.

Let  $\mathbf{\lambda} = (\lambda_1, \dots, \lambda_q)'$  and  $\boldsymbol{\gamma} = (\gamma_{mk} : m = k+1, \dots, q; k = 1, \dots, q-1)'$ . Prior distributions are placed on  $\mathbf{\lambda}$  and  $\boldsymbol{\gamma}$  in the full model, and submodels  $M$  are indexed by the 0 elements of  $\mathbf{\lambda}$  and  $\boldsymbol{\gamma}$ . To drop out the off-diagonal elements in the covariance matrix when a random effect is excluded, the support of the prior for  $\boldsymbol{\gamma}$  is defined as  $\mathcal{R}_{\mathbf{\lambda}} = \{\boldsymbol{\gamma} : \gamma_{mk} = \gamma_{kl} = 0 \text{ if } \lambda_k = 0, \text{ for } k = 1, \dots, q, 1 \leq l < k < m \leq q, l, m \in \mathcal{N}\}$ . Since the covariance matrix  $\mathbf{\Sigma}$  is a function of  $\mathbf{\lambda}$  and  $\boldsymbol{\gamma}$ , the prior density of  $\mathbf{\Sigma}$  is induced through the priors for  $\mathbf{\lambda}$  and  $\boldsymbol{\gamma}$ ,  $\pi(\mathbf{\lambda}, \boldsymbol{\gamma}) = \pi(\boldsymbol{\gamma}|\mathbf{\lambda})\pi(\mathbf{\lambda})$ . The prior for  $\mathbf{\lambda}$  is  $\prod_{k=1}^q \pi(\lambda_k)$ , where  $\pi(\lambda_k)$  is chosen as mixtures of point masses at zero and a truncated normal density:

$$\pi(\lambda_k) = \pi_{1,k0} 1(\lambda_k = 0) + (1 - \pi_{1,k0}) 1(\lambda_k > 0) \frac{N(\lambda_k; \mu_{1,k0}, \sigma_{1,k0}^2)}{F(0; -\mu_{1,k0}, \sigma_{1,k0}^2)}, \quad (7)$$

where  $\pi_{1,k0}$ ,  $\mu_{1,k0}$  and  $\sigma_{1,k0}^2$  are hyperparameters specified by investigators, and  $F(\cdot)$  is the normal distribution function. We refer to prior (7) as a zero-inflated positive normal density, ZI- $N^+(\lambda_k; \pi_{1,k0}, \mu_{1,k0}, \sigma_{1,k0}^2)$ . The prior probability of the  $k$ th random effect being excluded is  $\pi_{1,k0} = \Pr(H_{0k} : \lambda_k = 0)$ . The prior probability of the global null hypothesis of homogeneity is  $\Pr(H_0 : \lambda_1 = \dots = \lambda_q = 0) = \prod_{k=1}^q \pi_{1,k0}$ , which implies that all random effects are excluded from the model.

To allow fixed effect predictors to effectively drop out of the model, we also choose a zero-inflated normal density, ZI- $N(\beta_v | \pi_{2,v0}, \mu_{2,v0}, \sigma_{2,v0}^2)$ , as the prior for  $\beta_v$ , for  $v = 1, \dots, p$ . The prior

probability of the  $v$ th predictor being excluded is then  $\pi_{2,v0} = \Pr(\beta_v = 0)$ . Similar mixture priors have been widely used in the Bayesian variable selection literature (cf. Geweke, 1996).

We also allow zero off-diagonal elements in the covariance matrix by choosing mixture priors with masses at 0 for the  $\gamma$ 's. We choose a zero-inflated normal density,  $\text{ZI-N}(\gamma_{mk}; \pi_{3,mk,0}, \mu_{3,mk,0}, \sigma_{3,mk,0}^2)$ , with the constraint related to  $\boldsymbol{\lambda}$ , as the prior for  $\gamma_{mk}$ , for  $m = k + 1, \dots, q$  and  $k = 1, \dots, q - 1$ . This mixture prior fixes the prior probability of  $\gamma_{mk} = 0$  to be  $\pi_{3,mk,0}$ . In this way, the correlations between the random effects can be zero or non-zero. Explicitly, from (6), the correlation coefficient between the  $m$ th and the  $k$ th random effects is

$$\rho(\zeta_{im}, \zeta_{ik}; \boldsymbol{\gamma}) = \frac{\gamma_{mk} + \sum_{s=1}^{k-1} \gamma_{ks} \gamma_{ms}}{\sqrt{(1 + \sum_{s=1}^{m-1} \gamma_{ms}^2)(1 + \sum_{s=1}^{k-1} \gamma_{ks}^2)}}. \quad (8)$$

So the prior probability that the two random effects are uncorrelated is

$$\begin{aligned} \Pr\{\rho(\zeta_{im}, \zeta_{ik}) = 0\} &= \Pr\{\gamma_{m1}\gamma_{k1} = \dots = \gamma_{m,k-1}\gamma_{k,k-1} = \gamma_{mk} = 0\} \\ &= \pi_{3,mk,0} \prod_{s=1}^{k-1} [\pi_{3,ms,0}(1 - \pi_{3,ks,0}) + \pi_{3,ks,0}]. \end{aligned}$$

Note that the expression for the correlation coefficients,  $\rho(\zeta_{im}, \zeta_{ik})$ , for any two random effects that have non-zero variance ( $\lambda_m > 0, \lambda_k > 0$ ) does not involve  $\boldsymbol{\lambda}$ .

## 2.4 An Approximation

In implementing Bayesian model selection, it is necessary to calculate the marginal likelihood of  $\mathbf{y}$  conditional on the parameters by integrating out the random effects:

$$L(\boldsymbol{\beta}, \phi, \boldsymbol{\Sigma}; \mathbf{y}, \mathbf{X}, \mathbf{Z}) = \int_{\mathbb{R}^q} \pi(\mathbf{y}|\boldsymbol{\beta}, \phi, \boldsymbol{\zeta}, \mathbf{X}, \mathbf{Z}) \pi(\boldsymbol{\zeta}|\boldsymbol{\Sigma}) d\boldsymbol{\zeta}. \quad (9)$$

Let  $l(\boldsymbol{\beta}, \phi, \boldsymbol{\Sigma}; \mathbf{y}) = \log L(\boldsymbol{\beta}, \phi, \boldsymbol{\Sigma}; \mathbf{y})$ , suppressing the conditioning on  $\mathbf{X}$  and  $\mathbf{Z}$  as shorthand. To approximate (9), the classical way is Laplace's approximation (Solomon and Cox, 1992; Breslow and Clayton, 1993; Lin, 1997; Chipman et al., 2003, among others). From (6), we note that  $\boldsymbol{\Sigma}$  depends on the standard deviation of random effects which is proportional to  $\boldsymbol{\lambda}$ . When  $\boldsymbol{\lambda} = \mathbf{0}$ , the likelihood (9) reduces to ordinary GLM likelihood with  $\boldsymbol{\Sigma} = \mathbf{0}$ . We assume that the moments of the random effects  $\boldsymbol{\zeta}$  with order higher than two have order  $o(\boldsymbol{\lambda})$ . This is reasonable for Maclaurin

series, the special case of Taylor series. We begin by approximating the first integrand of (9) by taking a second order Taylor series expansion at  $E(\boldsymbol{\zeta}) = \mathbf{0}$ , the mean of the random effects:

$$\begin{aligned} L(\boldsymbol{\beta}, \phi, \boldsymbol{\zeta}; \mathbf{y}) &\approx L(\boldsymbol{\beta}, \phi, \boldsymbol{\zeta} = \mathbf{0}; \mathbf{y}) + \frac{\partial L(\boldsymbol{\beta}, \phi, \boldsymbol{\zeta}; \mathbf{y})}{\partial \boldsymbol{\zeta}} \Big|_{\boldsymbol{\zeta}=\mathbf{0}} \boldsymbol{\zeta} + \frac{1}{2} \boldsymbol{\zeta}' \frac{\partial^2 L(\boldsymbol{\beta}, \phi, \boldsymbol{\zeta}; \mathbf{y})}{\partial \boldsymbol{\zeta} \partial \boldsymbol{\zeta}'} \Big|_{\boldsymbol{\zeta}=\mathbf{0}} \boldsymbol{\zeta} \\ &= L(\boldsymbol{\beta}, \phi, \boldsymbol{\zeta} = \mathbf{0}; \mathbf{y}) \left\{ 1 + \frac{\partial l(\boldsymbol{\beta}, \phi, \boldsymbol{\zeta}; \mathbf{y})}{\partial \boldsymbol{\eta}} \Big|_{\boldsymbol{\zeta}=\mathbf{0}} \mathbf{Z} \boldsymbol{\zeta} + \frac{1}{2} (\mathbf{Z} \boldsymbol{\zeta})' \left[ \left( \frac{\partial l(\boldsymbol{\beta}, \phi, \boldsymbol{\zeta}; \mathbf{y})}{\partial \boldsymbol{\eta}} \frac{\partial l(\boldsymbol{\beta}, \phi, \boldsymbol{\zeta}; \mathbf{y})}{\partial \boldsymbol{\eta}'} \right) \right. \right. \\ &\quad \left. \left. + \text{DG} \left[ \frac{\partial^2 l(\boldsymbol{\beta}, \phi, \boldsymbol{\zeta}; \mathbf{y})}{\partial \boldsymbol{\eta} \partial \boldsymbol{\eta}'} \right] \right] \Big|_{\boldsymbol{\zeta}=\mathbf{0}} \mathbf{Z} \boldsymbol{\zeta} \right\}, \end{aligned}$$

where  $\boldsymbol{\eta} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\zeta}$ , and  $\text{DG}(A)$  denotes a diagonal matrix with diagonal entries of  $A$ . We note that (9) is actually the expectation of  $L(\boldsymbol{\beta}, \phi, \boldsymbol{\zeta}; \mathbf{y})$  with respect to  $\boldsymbol{\zeta}$ . Thus the approximation  $\tilde{L}(\boldsymbol{\beta}, \phi, \boldsymbol{\Sigma}; \mathbf{y})$  can be expressed as

$$\tilde{L}(\boldsymbol{\beta}, \phi, \boldsymbol{\Sigma}; \mathbf{y}) = L_0 \left\{ 1 + \frac{1}{2} \text{tr} \left[ \mathbf{Z}' \left( \frac{\partial l(\boldsymbol{\beta}, \phi, \boldsymbol{\zeta}; \mathbf{y})}{\partial \boldsymbol{\eta}} \frac{\partial l(\boldsymbol{\beta}, \phi, \boldsymbol{\zeta}; \mathbf{y})}{\partial \boldsymbol{\eta}'} + \text{DG} \left[ \frac{\partial^2 l(\boldsymbol{\beta}, \phi, \boldsymbol{\zeta}; \mathbf{y})}{\partial \boldsymbol{\eta} \partial \boldsymbol{\eta}'} \right] \right) \Big|_{\boldsymbol{\zeta}=\mathbf{0}} \mathbf{Z} \boldsymbol{\Sigma}^* \right] \right\}, \quad (10)$$

where  $L_0 = L(\boldsymbol{\beta}, \phi, \boldsymbol{\zeta} = \mathbf{0}; \mathbf{y})$ , which denotes the likelihood for the ordinary GLM,  $\text{tr}(A)$  denotes the trace of matrix  $A$ , and  $\boldsymbol{\Sigma}^* = I_n \otimes \boldsymbol{\Sigma}$ , the Kronecker product of  $I_n$  and  $\boldsymbol{\Sigma}$ . The second term in (10) involves the first and second derivative calculations. Thus, the approximation (10) is tractable, since the first and second derivatives of  $l(\boldsymbol{\beta}, \phi, \boldsymbol{\zeta}|\mathbf{y})$  are easily obtained as follows

$$\begin{aligned} \frac{\partial l(\boldsymbol{\beta}, \phi, \boldsymbol{\zeta}|\mathbf{y})}{\partial \boldsymbol{\eta}} &= \left( \mathbf{y} - \frac{\partial \psi(h(\boldsymbol{\eta}))}{\partial h(\boldsymbol{\eta})} \right) \frac{\partial h(\boldsymbol{\eta})}{\phi \partial \boldsymbol{\eta}} \\ \frac{\partial^2 l(\boldsymbol{\beta}, \phi, \boldsymbol{\zeta}|\mathbf{y})}{\partial \boldsymbol{\eta} \partial \boldsymbol{\eta}'} &= \left( \mathbf{y} - \frac{\partial \psi(h(\boldsymbol{\eta}))}{\partial h(\boldsymbol{\eta})} \right) \frac{\partial^2 h(\boldsymbol{\eta})}{\phi \partial \boldsymbol{\eta} \partial \boldsymbol{\eta}'} - \frac{\partial^2 \psi(h(\boldsymbol{\eta}))}{\phi \partial h(\boldsymbol{\eta}) \partial h(\boldsymbol{\eta}')} \frac{\partial h(\boldsymbol{\eta})}{\partial \boldsymbol{\eta}} \frac{\partial h(\boldsymbol{\eta})}{\partial \boldsymbol{\eta}'}. \end{aligned}$$

Then, in general, the approximation  $\tilde{L}(\boldsymbol{\beta}, \phi, \boldsymbol{\Sigma}; \mathbf{y})$  may be expressed as

$$L_0 \left\{ 1 + \frac{1}{2\phi} \left\{ \sum_{k=1}^q \sigma_{kk} \sum_{i=1}^n B_{i,k}^{(1)} + 2 \sum_{k=1}^{q-1} \sum_{m=k+1}^q \sigma_{mk} \sum_{i=1}^n B_{i,m,k}^{(2)} \right\} \right\}, \quad (11)$$

where  $B_{i,k}^{(1)}$  and  $B_{i,m,k}^{(2)}$  are functions of  $\boldsymbol{\beta}$  related to response variable  $\mathbf{y}$ , fixed effect predictors  $\mathbf{X}$ , and the random effect predictors  $\mathbf{Z}$ , and vary for particular GLMMs. In detail, the approximation (11) may be shown as

$$L_0 \left\{ 1 + \frac{1}{2\phi} \left\{ \sum_{k=1}^q \lambda_k^2 \left( 1 + \sum_{s=1}^{k-1} \gamma_{ks}^2 \right) \sum_{i=1}^n B_{i,k}^{(1)} + 2 \sum_{k=1}^{q-1} \sum_{m=k+1}^q \lambda_k \lambda_m \left( \gamma_{mk} + \sum_{s=1}^{k-1} \gamma_{ks} \gamma_{ms} \right) \sum_{i=1}^n B_{i,m,k}^{(2)} \right\} \right\}. \quad (12)$$

This form gives a general analytically tractable form for GLMMs which simplifies the subsequent computation. The general result can be applied in a straightforward manner to any particular

special case (e.g. logistic regression, Poisson, log linear models, etc). The detailed marginal distributions for normal linear, logistic regression and Poisson models are provided in the Appendix.

If covariance matrix components for random effects are in some sense relatively small, the approximation (11) may be rewritten as

$$L_0 \exp \left\{ \frac{1}{2\phi} \left\{ \sum_{k=1}^q \sigma_{kk} \sum_{i=1}^n B_{i,k}^{(1)} + 2 \sum_{k=1}^{q-1} \sum_{m=k+1}^q \sigma_{mk} \sum_{i=1}^n B_{i,m,k}^{(2)} \right\} \right\}. \quad (13)$$

### 3. Posterior Computation

We choose priors for  $\beta$ ,  $\lambda$  and  $\gamma$  as described in Section 2.3. For binomial and Poisson likelihoods, the scale or dispersion parameter is  $\phi = 1$ . For normal linear models,  $\phi$  is  $\sigma^2$ , and we follow common practice in choosing a gamma prior,  $\mathcal{G}(c_0, d_0)$ , for  $\sigma^{-2}$ . Posterior computation relies on a stochastic search variable selection (SSVS) Gibbs sampling algorithm (George and McCulloch, 1993) in which we iteratively sample from the full conditional distributions of each of the parameters. For  $\beta$ ,  $\lambda$  and  $\gamma$ , these posteriors will have a mixture structure consisting of point mass at 0 and non-conjugate distributions. In calculating the point mass probabilities, we rely on the approximation described in Section 2.4. To sample from the non-conjugate distribution, we use adaptive rejection Metropolis sampling (Gilks et al., 1995). One can use an alternative non-rejection-based sampling algorithm via latent variables proposed by Damien, Wakefield and Walker (1999).

Let  $\delta_{1,k}$  denote an indicator variable which is one if  $H_{0k}$  holds and zero if the alternative hypothesis holds. The prior distribution for  $\delta_{1,k}$  is clearly Bernoulli( $\pi_{1,k0}$ ). Let latent variable  $\lambda_k$  be  $(1 - \delta_{1,k})\tilde{\lambda}_k$  for  $k = 1, \dots, q$ . The prior distribution (7) may result from the following prior:

$$\pi(\tilde{\lambda}, \delta_1) = \prod_{k=1}^q \left[ N(\tilde{\lambda}_k | \mu_{1,k0}, \sigma_{1,k0}^2) \pi_{1,k0}^{\delta_{1,k}} (1 - \pi_{1,k0})^{1-\delta_{1,k}} \right].$$

Similar settings can be applied to  $\gamma$  and  $\beta$ . After specifying initial values for parameters, the algorithm iterates through the following steps a large number of times:

**Step 1:** Sample  $\tilde{\lambda}_k$  from its full conditional posterior distribution, which is

$$1(\tilde{\lambda}_k > 0) \tilde{L}(\beta, \tilde{\lambda}_k, \lambda_{(-k)}, \gamma, \phi; \mathbf{y}) N(\tilde{\lambda}_k; \mu_{1,k0}, \sigma_{1,k0}^2), \quad (14)$$

where the conditional probability of  $\tilde{\lambda}_k = 0$  is approximated by

$$\tilde{\pi}_{1,k} = \frac{\pi_{1,k0}}{\pi_{1,k0} + (1 - \pi_{1,k0})C_{1,k}}, \quad (15)$$

with  $C_{1,k} = \tilde{L}(\boldsymbol{\beta}, \lambda_k = \tilde{\lambda}_k, \boldsymbol{\lambda}_{(-k)}, \boldsymbol{\gamma}, \phi; \mathbf{y}) / \tilde{L}(\boldsymbol{\beta}, \lambda_k = 0, \boldsymbol{\lambda}_{(-k)}, \boldsymbol{\gamma}, \phi; \mathbf{y})$ , where  $\boldsymbol{\lambda}_{(-k)} = (\lambda_1, \dots, \lambda_{k-1}, \lambda_{k+1}, \dots, \lambda_q)'$ . Sampling from (15) can proceed by first sampling  $\delta_{1,k}$  from a Bernoulli  $(\tilde{\pi}_{1,k})$ . If  $\delta_{1,k}$  equals one, then let  $\lambda_k = 0$  and exclude the  $k$ th random effect. Otherwise sample  $\tilde{\lambda}_k$  for  $\lambda_k$  from  $\tilde{L}(\boldsymbol{\beta}, \tilde{\lambda}_k, \boldsymbol{\lambda}_{(-k)}, \boldsymbol{\gamma}, \phi; \mathbf{y})N(\tilde{\lambda}_k; \mu_{1,k0}, \sigma_{1,k0}^2)$ .

**Step 2:** Sample  $\tilde{\gamma}_{mk}$  ( $m > k$ ) from its full conditional posterior distribution, which is

$$1(\tilde{\gamma}_{mk} \neq 0)\tilde{L}(\boldsymbol{\beta}, \boldsymbol{\lambda}, \tilde{\gamma}_{mk}, \boldsymbol{\gamma}_{(-mk)}, \phi; \mathbf{y})N(\tilde{\gamma}_{mk}; \mu_{3,mk,0}, \sigma_{3,mk,0}^2), \quad (16)$$

where the conditional probability of  $\tilde{\gamma}_{mk} = 0$  is approximated by

$$\tilde{\pi}_{3,mk} = \frac{\pi_{3,mk,0}}{\pi_{3,mk,0} + (1 - \pi_{3,mk,0})C_{3,mk}}, \quad (17)$$

with  $C_{3,mk} = \tilde{L}(\boldsymbol{\beta}, \boldsymbol{\lambda}, \tilde{\gamma}_{mk}, \boldsymbol{\gamma}_{(-mk)}, \phi; \mathbf{y}) / \tilde{L}(\boldsymbol{\beta}, \boldsymbol{\lambda}, \tilde{\gamma}_{mk} = 0, \boldsymbol{\gamma}_{(-mk)}, \phi; \mathbf{y})$ , where  $\boldsymbol{\gamma}_{(-mk)} = (\gamma_{m'k'} : m' = k' + 1, \dots, q; k' = 1, \dots, k - 1, k + 1, \dots, q - 1)'$ . Sampling from (17) can proceed again by first sampling from a Bernoulli  $(\tilde{\pi}_{3,mk})$ . If this sample equals one, then let  $\gamma_{mk} = 0$ . Otherwise sample  $\tilde{\gamma}_{mk}$  for  $\gamma_{mk}$  from  $\tilde{L}(\boldsymbol{\beta}, \boldsymbol{\lambda}, \tilde{\gamma}_{mk}, \boldsymbol{\gamma}_{(-mk)}, \phi; \mathbf{y})N(\tilde{\gamma}_{mk}; \mu_{3,mk,0}, \sigma_{3,mk,0}^2)$ . However, if  $\lambda_m = 0$  or  $\lambda_k = 0$ ,  $\gamma_{mk} = 0$  according to its constraint related to  $\boldsymbol{\lambda}$ .

**Step 3:** Sample  $\tilde{\beta}_v$  from its full conditional posterior distribution, which is

$$1(\tilde{\beta}_v \neq 0)\tilde{L}(\tilde{\beta}_v, \boldsymbol{\beta}_{(-v)}, \boldsymbol{\lambda}, \boldsymbol{\gamma}, \phi; \mathbf{y})N(\tilde{\beta}_v; \mu_{2,v0}, \sigma_{2,v0}^2), \quad (18)$$

where the conditional probability of  $\tilde{\beta}_v = 0$ ,  $v = 1, \dots, p$ , is approximated by

$$\tilde{\pi}_{2,v} = \frac{\pi_{2,v0}}{\pi_{2,v0} + (1 - \pi_{2,v0})C_{2,v}}, \quad (19)$$

with  $C_{2,v} = \tilde{L}(\tilde{\beta}_v, \boldsymbol{\beta}_{(-v)}, \boldsymbol{\lambda}, \boldsymbol{\gamma}, \phi; \mathbf{y}) / \tilde{L}(\tilde{\beta}_v = 0, \boldsymbol{\beta}_{(-v)}, \boldsymbol{\lambda}, \boldsymbol{\gamma}, \phi; \mathbf{y})$ , where  $\boldsymbol{\beta}_{(-v)} = (\beta_1, \dots, \beta_{v-1}, \beta_{v+1}, \dots, \beta_p)'$ . Sampling from (19) can proceed again by first sampling from a Bernoulli  $(\tilde{\pi}_{2,v})$ . If this sample equals one, then let  $\beta_v = 0$ . Otherwise sample  $\tilde{\beta}_v$  for  $\beta_v$  from  $\tilde{L}(\tilde{\beta}_v, \boldsymbol{\beta}_{(-v)}, \boldsymbol{\lambda}, \boldsymbol{\gamma}, \phi; \mathbf{y})N(\tilde{\beta}_v; \mu_{2,v0}, \sigma_{2,v0}^2)$ .

**Step 4:** Sample  $\sigma^{-2}$ , if  $\phi = \sigma^2$ , from its full conditional distribution:  $\mathcal{G}(\sigma^{-2}; c_0, d_0) \tilde{L}(\boldsymbol{\beta}, \boldsymbol{\lambda}, \boldsymbol{\gamma}, \sigma^{-2}; \mathbf{y})$ .

Samples from the joint posterior distribution of the parameters are generated by repeating these steps for a large number of iterations after apparent convergence.

By varying the elements of  $\boldsymbol{\lambda}$ ,  $\boldsymbol{\beta}$  and  $\boldsymbol{\gamma}$  that are assigned 0 values, the algorithm effectively generates samples from the posterior distribution of  $M$ . As in SSVS algorithms for linear regression, we do not visit all the possible models in  $\mathcal{M}$ , since this number is typically enormous. Instead, by stochastically making local changes to the model based on (approximated) conditional model probabilities, we tend to visit models with relatively high posterior probability. However, for very large model spaces, there is no guarantee that we will visit the best model in  $\mathcal{M}$ . In addition, there may be a large number of models which have similar posterior probability. Hence, inferences are often based on marginal posterior probabilities of excluding a particular predictor from the fixed and/or random effects components.

Posterior model probabilities can be estimated by averaging indicator variables across iterations collected after apparent convergence. For example, to estimate the posterior probability of the  $k$ th random effect being excluded, one can simply add up the number of iterations for which  $\lambda_k = 0$  and divide by the total number of iterations. An alternative method is to use a Rao-Blackwell estimator  $\hat{\Pr}(\lambda_k = 0 | \text{data}) = \frac{1}{S} \sum_{s=1}^S \tilde{\pi}_{1,k}^{(s)}$ , where  $\tilde{\pi}_{1,k}^{(s)}$  is the value of  $\tilde{\pi}_{1,k}$  at iteration  $s$ , for  $s = 1, \dots, S$ . This estimator is potentially more efficient. The same approach can be used to calculate posterior probabilities of excluding predictors from the fixed effect component. To estimate the posterior probability that two random effects are uncorrelated given that they are both in the model (e.g.  $\sigma_{mk} = 0$ ), one can use the following estimator:

$$\hat{\Pr}(\sigma_{mk} = 0 | \lambda_m > 0, \lambda_k > 0, \text{data}) = \frac{\sum_{s: \lambda_m > 0, \lambda_k > 0} 1 \{ \rho(\zeta_{im}, \zeta_{ik}; \boldsymbol{\gamma}^{(s)}) = 0 \}}{\sum_{s=1}^S 1(\lambda_m > 0, \lambda_k > 0)},$$

so that we calculate the proportion of samples for which the random effects are uncorrelated from among the samples for which both random effects are in the model.

When (13) holds, the more explicit full conditional posterior distributions for some parameters and latent variables can be derived from the joint posterior distribution (refer to Appendix B).

#### 4. Simulation Studies

We carried out a simulation study to evaluate the behavior of the procedure. Simulated data based on different random effects covariance structures are generated from the GLMM with identity link, logistic link, and log link. We consider 100 subjects, each of which has 6 observations. The numbers of candidate predictors in the two components,  $p$  and  $q$ , are chosen as  $p = q = 3, 5$  or  $8$ . The covariates are  $\mathbf{x}_{ij} = (x_{ij1}, \dots, x_{ijp})'$ , where  $x_{ij1} = 1$  and  $x_{ijk} \sim \text{Bernoulli}(0.5)$ , for  $i = 1, \dots, 100$ ,  $j = 1, \dots, 6$ ,  $k = 2, \dots, p$ . Let  $\mathbf{z}_{ij} = \mathbf{x}_{ij}$ ,  $\boldsymbol{\beta}_{(-2)} \sim N(0, \mathbf{I})$ ,  $\beta_2 = 0$ , and  $\boldsymbol{\zeta}_i = (\zeta_{i1}, \dots, \zeta_{iq})' \sim N(0, \boldsymbol{\Sigma})$ , where  $\boldsymbol{\Sigma} = \boldsymbol{\Lambda} \boldsymbol{\Gamma} \boldsymbol{\Gamma}' \boldsymbol{\Lambda}$  with different structures which are (1)  $\boldsymbol{\lambda} = (1.2, 0.4, 0.6)'$  and  $\boldsymbol{\gamma} = (0.4, 0.5, 0.3)'$ ; (2)  $\boldsymbol{\lambda} = (0.2, 0, 0.7, 0, 0.5)'$  and  $\boldsymbol{\gamma} = (0, 0.4, 0, 0, 0, 0, 0.8, 0, 0.1, 0)'$ ; (3)  $\boldsymbol{\lambda} = (0.5, 0.8, 0.9, 0.2, 0.1, 0.1, 0.6, 0)'$  and  $\boldsymbol{\gamma} = (0.3, 0.6, 0.5, 0.4, 0.2, 0.1, 0.2, 0.3, 0.4, 0.3, 0.6, 0.1, 0.2, 0.1, 0.8, 0.3, 0.4, 0.8, 0.6, 0.3, 0.2, 0, 0, 0, 0, 0, 0, 0, 0)'$ . The corresponding covariance matrices for random effects are shown in the first row in Figure 1. For the GLMM with identity link,  $y_{ij} \sim N(\mathbf{x}_{ij}'\boldsymbol{\beta} + \mathbf{z}_{ij}'\boldsymbol{\zeta}_i, \sigma^{-2})$  with  $\sigma^{-2} = 2$ . For the GLMM with logistic link,  $y_{ij} \sim \text{Bernoulli}(\pi_{ij})$  with  $\text{logit}(\pi_{ij}) = \mathbf{x}_{ij}'\boldsymbol{\beta} + \mathbf{z}_{ij}'\boldsymbol{\zeta}_i$ . For the GLMM with log link,  $y_{ij} \sim \text{Poisson}(\lambda_{ij})$  with  $\log(\lambda_{ij}) = \mathbf{x}_{ij}'\boldsymbol{\beta} + \mathbf{z}_{ij}'\boldsymbol{\zeta}_i$ .

We chose the prior distribution for  $\lambda_k$  as  $\text{ZI-N}^+(\lambda_k; \pi_{1,k0}, 0, 10)$ . The prior distributions for the elements of  $\boldsymbol{\gamma}$  are chosen to be mixture priors,  $\text{ZI-N}(\gamma_{mk}; \pi_{3,u0}, 0, 1)$ , with the constraint related to  $\boldsymbol{\lambda}$ . A mixture prior distribution for  $\beta_v$  is chosen as  $\text{ZI-N}(\beta_v; \pi_{2,k0}, 0, 100)$ . A diffuse prior for parameter  $\sigma^{-2}$  is chosen to be  $\mathcal{G}(0.08, 0.08)$ . To study the effect of the prior probabilities of  $\lambda_k = 0$ ,  $\beta_v = 0$  and  $\gamma_{mk} = 0$  on the estimated posterior probabilities, we consider 0.2, 0.5, 0.8 for these prior probabilities.

For each simulated data set and choice of prior, we ran the Gibbs sampling algorithm described in Section 3 for 20,000 iterations after a 2000 burn-in. The diagnostic tests were carried out by using Geweke (1992) and Raftery and Lewis (1992), which showed rapid convergence and efficient mixing. A sample of size 4000 was obtained by thinning the MCMC chain by a factor of 5. For each simulated data set, we calculated (i) the posterior probabilities for the possible submodels under each link; and (ii) the estimated posterior means and the 95% credible intervals for each of the parameters.

Figure 1 displays image plots of the true covariance matrices for random effects corresponding to the simulated data and the estimated covariance matrices under each link. Table 2 shows the estimated posterior probabilities for the preferred subsets of fixed and random effect predictors for different simulation studies with different prior probabilities for  $\lambda_k = 0$ ,  $\beta_v = 0$  and  $\gamma_{mk} = 0$  under logistic link. We note that under the logistic link, although the estimated posterior probabilities change slightly according to the prior probabilities of  $\lambda_k = 0$ ,  $\beta_v = 0$  and  $\gamma_{mk} = 0$ , it is evident that the preferred model agrees with the true model specification regardless of the choices of  $\pi_{1,k0}$ ,  $\pi_{2,v0}$  and  $\pi_{3,u0}$ . The results under the identity and log links are similar to those under logistic link, though we do not show them here. Figure 2 presents boxplots of the samples of parameters for the second simulation under each link. The true values of all parameters fall in the 95% credible intervals.

To study the accuracy of the approximation proposed in Section 2.4, we consider a simple GLMM with identity link. Similarly, we consider 100 subjects, each of which has 6 observations, and  $p$  and  $q$  are chosen as 3. We choose different covariance structures with standard deviation components proportional to  $\boldsymbol{\lambda}$  from small to large, which are 1)  $\boldsymbol{\lambda} = (0.01, 0.02, 0.005)'$ ; 2)  $\boldsymbol{\lambda} = (1.2, 0.4, 0.6)'$ ; 3)  $\boldsymbol{\lambda} = (2.8, 4.3, 3.5)'$ ; 4)  $\boldsymbol{\lambda} = (27.5, 20.6, 35.1)'$ ; 5)  $\boldsymbol{\lambda} = (50.6, 30.8, 60.3)'$ .  $\boldsymbol{\gamma}$  is kept fixed at  $(0.4, 0.5, 0.3)'$ . We calculate the marginal likelihoods of the simulated data by using different methods, including exact calculation, Laplace approximation, importance sampling, and Chib's marginal likelihood method. For more details, see Sinharay and Stern (2001) and the references therein. Table 1 shows the results which show that all methods work basically the same.

Sensitivity of the results to the prior specification was assessed by repeating the analyses with the following different hyperparameters: (a) priors with variance  $/2$ ; (b) priors with variance  $\times 2$ ; (c) priors with moderately different means. Although we do not show details, inferences for all models are robust to the prior specification. The ranges in Table 2 illustrate this robustness.

## 5. Time-to-pregnancy Application

We illustrate the methodology through application to the time-to-pregnancy (TTP) study mentioned in Section 1. Female dental assistants, aged 19 to 39, were randomly selected from the

dental-assistants registry of the California Department of Consumer Affairs. 427 women provided detailed data on reproductive and contraceptive history, occupational exposures, and other factors related to fertility. Rowland et al. (1992) found an increased number of hours of exposure to unscavenged nitrous oxide was associated with decreased fecundability, the probability of conception within a single menstrual cycle with noncontracepted intercourse. Reanalyses were presented by Weinberg et al. (1993) and Dunson and Neelon (2003). These articles allow baseline heterogeneity in fecundability by allowing the hazard to change as the number of cycles attempting increases, but do not allow heterogeneity to vary with predictors.

Our goal is not only to assess covariate effects on fecundability, but also to identify factors that vary in their effects across women. To this end, we modeled the discrete-time hazard of conception using the logistic GLMM of expression (1), with candidate predictors including category indicators for age (19-24, 25-29, >30), intercourse frequency per week ( $\leq 1$ , 1-3, 3-4, >4), cigarettes smoked per day (nonsmoker, 1-5, 6-10, 11-15, >15), and the use of oral contraceptives in the cycle prior to beginning the pregnancy attempt (no, yes).

Let  $y_{ij}$  denote the pregnancy status (0 = no, 1 = yes) for woman  $i$  at the  $j$ th menstrual cycle,  $j = 1, \dots, T_i$ ,  $T_i \in \{1, \dots, 13\}$  (women not conceiving by cycle 13 are censored), and  $\mathbf{x}_{ij} = (x_{ij1}, \dots, x_{ij14})'$ , where  $x_{ij1} = 1$ , and  $x_{ij2}, \dots, x_{ij14}$  are categories of the predictors in the above order. To allow each regression coefficient to possibly vary across women, we let  $\mathbf{z}_{ij} = \mathbf{x}_{ij}$ . The prior distribution for  $\lambda_k$  is chosen as  $\text{ZI-N}^+(\lambda_k | \pi_{k0}, 0, 20)$ . The prior distribution for  $\beta_v$  is chosen as  $\text{ZI-N}(\beta_v | \pi_{v0}, 0, 100)$ . To reflect a somewhat diffuse prior on the correlations, we choose the prior distributions for the elements of  $\gamma$  as  $\text{ZI-N}(\gamma_{mk} | \pi_{3,u0}, 0, 1)$  with the constraint on  $\lambda$ . A diffuse prior for parameters  $\sigma^{-2}$  is chosen to be  $\mathcal{G}(0.08, 0.08)$ . We ran 80,000 iterations after a 10,000 burn-in. The chains passed Geweke (1992) and Raftery and Lewis (1992) convergence diagnostic tests. We retained every 20th sample for inferences of interest.

Table 3 presents the marginal posterior probabilities of including each predictor in the fixed and random effects components under different choices of  $\pi_{1,k0}$  and  $\pi_{2,v0}$ . The overall posterior probability of including age in the fixed effect component can be calculated as the posterior prob-

ability that any of the category indicators for age are included. Such overall posterior probabilities are calculated separately for the fixed and random effect components for each of the factors under consideration, including age, intercourse frequency, cigarettes smoked, and recent pill use. The results are shown in Table 3. The posterior probability of including age in the fixed effect component ranges from 0.95-0.97 (average = 0.96) depending on the prior. The corresponding ranges for intercourse frequency, cigarettes smoked, and recent pill use are 0.96-0.97 (average = 0.96), 0.92-0.97 (average = 0.94), and 0.87-0.98 (average = 0.92), respectively. Hence, as expected, there is some evidence that age, intercourse frequency, cigarette smoking, and recent pill use are predictive of fecundability on average, with the most evidence for age and intercourse frequency. The age effect is most apparent in women over 30+. In addition, the indicators for the highest categories of intercourse frequency (4+ acts/week) and cigarette smoking (15+/day) had the highest posterior probabilities of inclusion.

For the random effect component, the results were somewhat different. The posterior probability of inclusion for recent pill use ranged between 0.40 and 0.53 (average = 0.47), so there is no evidence of heterogeneity in the effect of recent pill use. However, there was some evidence of heterogeneity among women in the effects of each of the other factors. In particular, the posterior probability of including age in the random effect component ranged between 0.87 and 0.92 (average = 0.90), which is suggestive but not clear evidence. There was slightly more evidence of heterogeneity in the effects of intercourse frequency and cigarette smoking with the posterior probabilities of inclusion for these two factors ranging between 0.90-0.93 (average = 0.92) and 0.91-0.95 (average = 0.93), respectively.

Variability among women in the effect of age may be due to differences in the rate of decline of viable ova, or possibly due to environmental exposures. Selection by pregnancy history may also play a role, since some highly fertile couples may achieve their desired family size by younger ages. Heterogeneity in the aging effect has important implications for couples planning pregnancy, particularly given the growing concern that infertility may result if women delay attempting until their mid or late 30s. Heterogeneity in the impact of intercourse frequency and exposures, such as

smoking and recent pill use, are also extremely interesting. We plan to verify these preliminary results using larger data sets.

Table 4 provides the overall posterior summaries of the regression coefficients from our approach compared with the standard GLM with the logistic link. It is clear that there are no systematic differences between our model-averaged Bayesian point and interval estimates for the regression coefficients and the maximum likelihood estimates. This should be reassuring to frequentist statisticians that our prior choice has minimal impact on estimation, while facilitating inferences.

We ran extensive sensitivity analyses to evaluate the robustness of the results to the prior specifications by repeating the analyses with the following different hyperparameters: (a) priors with variance  $/2$ ; (b) priors with variance  $\times 2$ ; (c) priors with moderately different means within the range of the prior expectation. The ranges in Table 3 show the results for all of the different priors.

## 6. Discussion

In this article, we propose a Bayesian approach to the problem of random effects covariance selection in GLMMs. By using variable selection-type mixture priors for the fixed effect coefficients and the components in a special LDU decomposition of the random effects covariance, we develop an SSVS Gibbs sampling algorithm to avoid the overwhelming problem of calculating the posterior probabilities of all  $2^p \sum_{k=0}^q \binom{q}{k} 2^{\frac{1}{2}(q-k)(q-k-1)}$  submodels.

Since GLMMs extend GLMs to accommodate both non-normal response distributions and non-linear transformations of linear models containing random effects, there are a number of members included in this family. It is too often the case in the analysis of longitudinal and clustered data that investigators focus overly-much on the fixed effects, while considering the dependency structure as a nuisance. In a broad variety of applications, it is important and necessary to expand inferences under GLMMs to include the random effects covariance. This is certainly the case not only in fecundability studies, as we have illustrated, but also in general epidemiologic studies. In fact, in certain cases, such as family studies of heritability (Guo and Wang, 2002), inferences on the covariance structure are of primary interest. Since current methods for such inferences are limited

to criterion-based methods and simple score and likelihood ratio tests, our proposed Bayesian stochastic search approach should prove useful, both for model selection and for averaging across the (often extremely-high dimensional) set of possible models.

Although we have focused on settings in which the same type of outcome is measured repeatedly, the methods can also be used for covariance structure (or graphical) modeling of mixed scale and non-Gaussian variables. It is increasingly the case that interest focuses on very high-dimensional variables, such as are collected in genomic applications. Unfortunately, our approach is limited in the dimension that can feasibly be considered.

In our analysis, the execution time increases with  $p = q$  at a rate proportional to roughly the cube of the number of predictors. An important area for future research is the development of methods for large and high-dimensional data. When sample sizes are modest compared with the number of models under consideration, there can be sensitivity to the choice of priors. Although we have described subjective priors, it would be useful to consider reference and default-type priors in this setting.

## APPENDIX A

The normal linear mixed model of Laird and Ware (1982) is a special case of a GLMM having  $g(\mu_{ij}) = \mu_{ij} = \eta_{ij} = \mathbf{x}'_{ij}\boldsymbol{\beta} + \mathbf{z}'_{ij}\boldsymbol{\zeta}_i$ ,  $\phi = \sigma^2$  and  $\psi(\theta_{ij}) = \eta_{ij}^2/2$ . In this case,  $\frac{\partial l(\boldsymbol{\beta}, \phi, \boldsymbol{\zeta} | \mathbf{y})}{\partial \boldsymbol{\eta}} \frac{\partial l(\boldsymbol{\beta}, \phi, \boldsymbol{\zeta} | \mathbf{y})}{\partial \boldsymbol{\eta}'} = (\mathbf{y} - \boldsymbol{\eta})(\mathbf{y} - \boldsymbol{\eta})' / \sigma^2$  and  $\frac{\partial^2 l(\boldsymbol{\beta}, \phi, \boldsymbol{\zeta} | \mathbf{y})}{\partial \boldsymbol{\eta} \partial \boldsymbol{\eta}'} = -\mathbf{1}_N \mathbf{1}_N' / \sigma^2$ . Therefore we have

$$\begin{aligned} B_{i,k}^{(1)} &= ((\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta})' Z_{ik})^2 - Z_{ik}' Z_{ik} \\ B_{i,m,k}^{(2)} &= (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta})' Z_{im} Z_{ik}' (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta}) - Z_{ik}' Z_{im}, \end{aligned}$$

where  $Z_{ik}$  denotes the  $k$ th column of  $\mathbf{Z}_i$ , and  $L_0 = \exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^n \sum_{j=1}^{n_i} (y_{ij} - \mathbf{x}'_{ij}\boldsymbol{\beta})^2 \right\}$ .

When  $y_{ij}$  are 0-1 random variables, the logistic regression model can be obtained by the canonical link function  $g(\pi_{ij}) = \log \frac{\pi_{ij}}{1-\pi_{ij}} = \eta_{ij} = \mathbf{x}'_{ij}\boldsymbol{\beta} + \mathbf{z}'_{ij}\boldsymbol{\zeta}_i$ ,  $\phi = 1$ ,  $\psi(\theta_{ij}) = \log(1 + e^{\eta_{ij}}) = -\log(1 - \pi_{ij})$ , hence  $\frac{\partial l(\boldsymbol{\beta}, \phi, \boldsymbol{\zeta} | \mathbf{y})}{\partial \boldsymbol{\eta}} \frac{\partial l(\boldsymbol{\beta}, \phi, \boldsymbol{\zeta} | \mathbf{y})}{\partial \boldsymbol{\eta}'} = (\mathbf{y} - \boldsymbol{\pi})(\mathbf{y} - \boldsymbol{\pi})'$  and  $\frac{\partial^2 l(\boldsymbol{\beta}, \phi, \boldsymbol{\zeta} | \mathbf{y})}{\partial \boldsymbol{\eta} \partial \boldsymbol{\eta}'} = -\boldsymbol{\pi}(\mathbf{1}_N - \boldsymbol{\pi})'$ . Then

$$\begin{aligned} B_{i,k}^{(1)} &= ((\mathbf{y}_i - \boldsymbol{\pi}_i)' Z_{ik})^2 - \boldsymbol{\pi}_i' \text{DG}(Z_{ik} Z_{ik}') (\mathbf{1}_{n_i} - \boldsymbol{\pi}_i) \\ B_{i,m,k}^{(2)} &= (\mathbf{y}_i - \boldsymbol{\pi}_i)' Z_{im} Z_{ik}' (\mathbf{y}_i - \boldsymbol{\pi}_i) - \boldsymbol{\pi}_i' \text{DG}(Z_{im} Z_{ik}') (\mathbf{1}_{n_i} - \boldsymbol{\pi}_i), \end{aligned}$$

where  $\boldsymbol{\pi}_i = (\pi_{i1}, \dots, \pi_{in_i})'$  with  $\pi_{ij} = \exp(\mathbf{x}'_{ij}\boldsymbol{\beta}) / (1 + \exp(\mathbf{x}'_{ij}\boldsymbol{\beta}))$ , and  $L_0 = \exp\{y_{ij} \log \frac{\pi_{ij}}{1-\pi_{ij}} + \log(1 - \pi_{ij})\}$ .

Similarly, when  $y_{ij}$  are counts with mean  $\lambda_{ij}$ , the Poisson regression model can be obtained by the canonical link function  $g(\lambda_{ij}) = \log \lambda_{ij} = \eta_{ij} = \mathbf{x}'_{ij}\boldsymbol{\beta} + \mathbf{z}'_{ij}\boldsymbol{\zeta}_i$ ,  $\phi = 1$ ,  $\psi(\theta_{ij}) = e^{\eta_{ij}} = \lambda_{ij}$ ,  $\frac{\partial l(\boldsymbol{\beta}, \phi, \boldsymbol{\zeta}|\mathbf{y})}{\partial \boldsymbol{\eta}} \frac{\partial l(\boldsymbol{\beta}, \phi, \boldsymbol{\zeta}|\mathbf{y})}{\partial \boldsymbol{\eta}'} = (\mathbf{y} - \boldsymbol{\lambda})(\mathbf{y} - \boldsymbol{\lambda})'$  and  $\frac{\partial^2 l(\boldsymbol{\beta}, \phi, \boldsymbol{\zeta}|\mathbf{y})}{\partial \boldsymbol{\eta} \partial \boldsymbol{\eta}'} = -\boldsymbol{\lambda} \mathbf{1}'_N$ . Then we obtain that

$$\begin{aligned} B_{i,k}^{(1)} &= ((\mathbf{y}_i - \boldsymbol{\lambda}_i)' Z_{ik})^2 - \boldsymbol{\lambda}'_i \text{DG}(Z_{ik} Z'_{ik}) \mathbf{1}_{n_i} \\ B_{i,m,k}^{(2)} &= (\mathbf{y}_i - \boldsymbol{\lambda}_i)' Z_{im} Z'_{ik} (\mathbf{y}_i - \boldsymbol{\lambda}_i) - \boldsymbol{\lambda}'_i \text{DG}(Z_{im} Z'_{ik}) \mathbf{1}_{n_i}, \end{aligned}$$

where  $\boldsymbol{\lambda}_i = (\lambda_{i1}, \dots, \lambda_{in_i})'$  with  $\lambda_{ij} = \exp(\mathbf{x}'_{ij}\boldsymbol{\beta})$ , and  $L_0 = \exp\{y_{ij} \log \lambda_{ij} - \lambda_{ij} - \log y_{ij}!\}$ .

## APPENDIX B

The full conditional posteriors for the parameters in Section 2.4 when (13) holds:

**Step 1:** Sample  $\tilde{\lambda}_k$ ,  $k = 1, \dots, q$  from its full conditional distribution,

$$1(\tilde{\lambda}_k > 0) C_{1,k} \text{N}(\tilde{\lambda}_k; \mu_{1,k0}, \sigma_{1,k0}^2),$$

where

$$C_{1,k} = \exp \left\{ \frac{\tilde{\lambda}_k}{2\phi} \sum_{i=1}^n \left\{ \tilde{\lambda}_k \left( 1 + \sum_{s=1}^{k-1} \gamma_{ks}^2 \right) B_{i,k}^{(1)} + 2 \sum_{t=1, t \neq k}^q \lambda_t \left( \gamma_{w(t,k)} + \sum_{s=1}^{r-1} \gamma_{ks} \gamma_{ts} \right) B_{i,w(t,k)}^{(2)} \right\} \right\},$$

with  $r = \min(t, k)$ ,  $\gamma_{w(t,k)}$  equals  $\gamma_{kt}$  if  $t < k$  and  $\gamma_{tk}$  otherwise, and  $B_{i,w(t,k)}^{(2)}$  denotes  $B_{i,k,t}^{(2)}$  if  $t < k$  and  $B_{i,t,k}^{(2)}$  otherwise.

**Step 2:** Sample  $\tilde{\gamma}_{mk}$  for  $m > k$  from its full conditional distribution which is proportional to

$$\exp \left\{ \frac{\lambda_m \tilde{\gamma}_{mk}}{\phi} \sum_{i=1}^n \left\{ \frac{1}{2} \lambda_m \tilde{\gamma}_{mk} B_{i,m}^{(1)} + \sum_{t=k, t \neq m}^q \lambda_t \gamma_{tk} B_{i,w(t,m)}^{(2)} \right\} \right\} \text{N}(\tilde{\gamma}_{mk}; \mu_{3,mk,0}, \sigma_{3,mk,0}^2).$$

**Step 4:** Sample  $\sigma^{-2}$ , if  $\phi = \sigma^2$ , from the full conditional distribution:  $\mathcal{G}(c_0, d_0 - \log(\tilde{L}(\boldsymbol{\beta}, \boldsymbol{\lambda}, \boldsymbol{\gamma}, \phi; \mathbf{y})/L_0))$ .

## REFERENCES

Albert, J.H., and Chib, S. (1997). Bayesian Test and Model Diagnostics in Conditionally Independent Hierarchical Models. *Journal of the American Statistical Association*, 92, 916-925.

- Breslow, N.E. and Clayton, D.G. (1993). Approximate Inference in Generalized Linear Mixed Models. *Journal of the American Statistical Association*, 88, 9-25.
- Chen, M., Ibrahim, J.G., Shao, Q. and Weiss, R.E. (2003). Prior Elicitation for Model Selection and Estimation in Generalized Linear Mixed Models. *Journal of Statistical Planning and Inference*, 111, 57-76.
- Chen, Z. and Dunson, D.B. (2003). Random Effects Selection in Linear Mixed Models. *Biometrics*, 59, 762-769.
- Chipman, H., George, E.I. and McCulloch, R.E. (2002). Bayesian Treed Models. *Machine Learning*, 48, 299-320.
- Chipman, H., George, E.I. and McCulloch, R.E. (2003). Bayesian Treed Generalized Linear Models. *Bayesian Statistics 7* (J.M. Bernardo, M.J. Bayarri, J.O. Berger, A.P. Dawid, D. Heckerman, A.F.M. Smith and M. West, eds), Oxford: University Press, 323-349.
- Commenges, D. and Jacqmin-Gadda, H. (1997). Generalized Score Test of Homogeneity Based on Correlated Random Effects Models. *Journal of the Royal Statistical Society B*, 59, 157-171.
- Damien, P., Wakefield, J. and Walker, S. (1999). Gibbs Sampling for Bayesian Non-conjugate and Hierarchical Models by Using Auxillary Variables. *Journal of the Royal Statistical Society: Series B*, 61, 331-344.
- Daniels, M.J. and Kass, R.E. (1999). Nonconjugate Bayesian Estimation of Covariance Matrices and Its Use in Hierarchical Models. *Journal of the American Statistical Association*, 94, 1254-1263.
- Daniels, M.J. and Pourahmadi, M. (2002). Bayesian Analysis of Covariance Matrices and Dynamic Models for Longitudinal Data. *Biometrika*, 89, 553-566.
- Daniels, M.J. and Zhao, Y.D. (2003). Modelling the Random Effects Covariance Matrix in Longitudinal Data. *Statistics in Medicine*, 22, 1631-1647.

- Dunson, D.B. and Neelon, B. (2003). Bayesian Inference on Order-Constrained Parameters in Generalized Linear Models. *Biometrics*, 59, 286-295.
- George, E.I. and McCulloch, R.E. (1993). Variable Selection via Gibbs Sampling. *Journal of the American Statistical Association*, 88, 881-889.
- Geweke, J. (1992). Evaluating the Accuracy of Sampling-based Approaches to the Calculation of Posterior Moments. *Bayesian Statistics 4*, (J.M. Bernardo, J.O. Berger, A.P. Dawid, and A.F.M. Smith, eds), Oxford University Press, Oxford. 169-193.
- Geweke, J. (1996). Variable Selection and Model Comparison in Regression. *Bayesian Statistics 5*, (J.O. Berger, J.M. Bernardo, A.P. Dawid, and A.F.M. Smith, eds), Oxford University Press, Oxford. 609-620.
- Gilks, W.R., Best, N.G. and Tan, K.K.C. (1995). Adaptive Rejection Metropolis Sampling within Gibbs Sampling. *Journal of Applied Statistics*, 44, 455-472.
- Gilks, W.R., Neal, R.M., Best, N.G. and Tan, K.K.C. (1997). Corrigendum: Adaptive Rejection Metropolis Sampling. *Applied Statistics*, 46, 541-542.
- Guo, G. and Wang, J. (2002). The Mixed or Multilevel Model for Behavior Genetic Analysis. *Behavior Genetics*, 32, 37-49.
- Hall, D.B. and Praestgaard, J.T. (2001). Order-Restricted Score Tests for Homogeneity in Generalized Linear and Non-linear Mixed Models. *Biometrika*, 88, 739-751.
- Laird, N.M. and Ware, J.H. (1982). Random Effects Models for Longitudinal Data. *Biometrics*, 38, 963-974.
- Liechty, J.C., Liechty, M.W. and Muller, P. (2004). Bayesian Correlation Estimation. *Biometrika*, 91, 1-14.
- Lin, X. (1997). Variance Component Testing in Generalized Linear Models with Random Effects. *Biometrika*, 84(2), 309-326.

- McCulloch, C.E. (1997). Maximum Likelihood Algorithms for Generalized Linear Mixed Models. *Journal of the American Statistical Association*, 92, 162-170.
- McCulloch, C.E. and Searle, S. (2001). Generalized Linear and Mixed Models. New York: Wiley.
- McGilchrist, C.A. (1994). Estimation in Generalized Mixed Models. *Journal of the Royal Statistical Society B*, 56, 61-69.
- Meyer, M.C. and Laud, P.W. (2002). Predictive Variable Selection in Generalized Linear Models. *Journal of the American Statistical Association*, 97, 859-871.
- Nott, D.J. and Leone, D. (2004). Sampling Schemes for Bayesian Variable Selection in Generalized Linear Models. *Journal of Computational and Graphical Statistics*, 13, 362-382.
- Ntzoufras, I., Dellaportas, P., and Forster, J.J. (2003). Bayesian Variable Selection and Link Determination for Generalised Linear Models. *Journal of Statistical Planning and Inference*, 111, 165-180.
- Raftery, A. (1996). Approximate Bayes Factors and Accounting for Model Uncertainty in Generalized Linear Models. *Biometrika*, 83, 251-266.
- Raftery, A.E. and Lewis, S. (1992). How Many Iterations in the Gibbs Sampler? *Bayesian Statistics 4* (Bernardo, J. M., Berger, J. O., Dawid, A. P. and Smith, A. F. M., eds), Oxford: Oxford University Press, 763-773.
- Raftery, A.E., Madigan, D. and Volinsky, C.T. (1996). Accounting for Model Uncertainty in Survival Analysis Improves Predictive Performance. *Bayesian Statistics 5* (J.M. Bernardo, J.O. Berger, A.P. Dawid and A.F.M. Smith, eds), Oxford: University Press, 323-349.
- Rowland, A.S., Baird, D.D., Weinberg, C.R., Shore, D.L., Shy, C.M. and Wilcox, A.J. (1992). Reduced Fertility Among Women Employed as Dental Assistants Exposed to High Levels of Nitrous Oxide. *The New England Journal of Medicine*, 327(14), 993-997.

- Schall, R. (1991). Estimation in Generalized Linear Mixed Models with Random Effects. *Biometrika*, 78, 719-727.
- Sinharay, S. and Stern, H.S. (2001). Bayes Factors for Variance Component Testing in Generalized Linear Mixed Models. *Bayesian Methods with Applications to Science, Policy and Official Statistics (ISBA 2000 Proceedings)*, 507-516.
- Solomon, P.J. and Cox, D.R. (1992). Nonlinear Component of Variance Models. *Biometrika*, 79(1), 1-11.
- Weinberg, C.R., Baird, D.D. and Rowland, A.S. (1993). Pitfalls Inherent in Retrospective Time-to-event Studies: the Example of Time to Pregnancy. *Statistics in Medicine*, 12(9), 867-879.
- Wong, F., Carter, C.K., and Kohn, R. (2003). Efficient Estimation of Covariance Selection Models. *Biometrika*, 90, 809-830.
- Zeger, S.L. and Karim, M.R. (1991). Generalized Linear Models with Random Effects: A Gibbs Sampling Approach. *Journal of the American Statistical Association*, 86, 79-86.

Table 1: Comparison of approximated log marginal likelihoods for the GLMM with identity link

$\lambda$	Chib's	Exact	I.Sampling	Laplace	Proposed
(0.01, 0.02, 0.005)	-621.6	-621.6	-621.6	-621.6	-621.6
(1.2, 0.4, 0.6)	-539.1	-539.0	-539.0	-539.1	-539.1
(2.8, 4.3, 3.5)	-497.9	-497.9	-498.0	-498.0	-498.2
(27.5, 20.6, 35.1)	-426.3	-426.2	-426.2	-426.3	-426.4
(50.6, 30.8, 60.3)	-459.8	-459.7	-459.8	-459.9	-460.1

Table 2: Estimated model posterior probabilities in simulation studies under logistic link. Sub-models with posterior probability less than 0.02 are not displayed.

Model	$\pi_{1,k0}$		
	0.2	0.5	0.8
<i><u>Simulation 1</u></i>			
$x_1, x_3, z_1, z_2, z_3^a$	0.833 <sup>b</sup> <sub>(0.814,0.865)<sup>c</sup></sub>	0.796 <sub>(0.771,0.828)</sub>	0.748 <sub>(0.719,0.782)</sub>
$x_1, x_3, z_1, z_3$	0.085 <sub>(0.054,0.116)</sub>	0.098 <sub>(0.083,0.115)</sub>	0.116 <sub>(0.098,0.141)</sub>
$x_1, x_3, z_1, z_2$	0.066 <sub>(0.045,0.092)</sub>	0.070 <sub>(0.046,0.095)</sub>	0.082 <sub>(0.059,0.106)</sub>
<i><u>Simulation 2</u></i>			
$x_1, x_3, x_4, x_5, z_1, z_3, z_5^a$	0.437 <sub>(0.421,0.544)</sub>	0.519 <sub>(0.483,0.558)</sub>	0.568 <sub>(0.543,0.591)</sub>
$x_3, x_4, x_5, z_1, \dots, z_5$	0.106 <sub>(0.075,0.140)</sub>	0.095 <sub>(0.068,0.131)</sub>	0.084 <sub>(0.048,0.112)</sub>
$x_1, x_3, x_4, x_5, z_3, z_5$	0.103 <sub>(0.076,0.135)</sub>	0.139 <sub>(0.110,0.180)</sub>	0.177 <sub>(0.138,0.218)</sub>
$x_3, x_4, x_5, z_1, z_3, z_4, z_5$	0.024 <sub>(0.013,0.037)</sub>	0.039 <sub>(0.026,0.054)</sub>	0.052 <sub>(0.037,0.078)</sub>
$x_2, \dots, x_5, z_1, \dots, z_4$	0.022 <sub>(0.010,0.035)</sub>	0.037 <sub>(0.020,0.053)</sub>	0.045 <sub>(0.024,0.066)</sub>
<i><u>Simulation 3</u></i>			
$x_1, x_3, \dots, x_8, z_1, \dots, z_7^a$	0.547 <sub>(0.529,0.577)</sub>	0.581 <sub>(0.550,0.617)</sub>	0.633 <sub>(0.602,0.658)</sub>
$x_1, x_3, \dots, x_8, z_1, \dots, z_8$	0.090 <sub>(0.079,0.106)</sub>	0.075 <sub>(0.065,0.089)</sub>	0.064 <sub>(0.053,0.075)</sub>
$x_1, x_3, \dots, x_7, z_1, \dots, z_4, z_6, z_7$	0.051 <sub>(0.043,0.059)</sub>	0.074 <sub>(0.066,0.085)</sub>	0.078 <sub>(0.070,0.089)</sub>
$x_1, x_3, \dots, x_7, z_1, \dots, z_5, z_7$	0.032 <sub>(0.027,0.041)</sub>	0.034 <sub>(0.024,0.042)</sub>	0.037 <sub>(0.030,0.041)</sub>
$x_1, x_3, \dots, x_7, z_1, \dots, z_4, z_7$	0.025 <sub>(0.013,0.039)</sub>	0.027 <sub>(0.018,0.037)</sub>	0.032 <sub>(0.023,0.044)</sub>
$x_1, \dots, x_7, z_1, \dots, z_3, z_7$	0.023 <sub>(0.011,0.033)</sub>	0.025 <sub>(0.014,0.035)</sub>	0.028 <sub>(0.016,0.040)</sub>

<sup>a</sup> True model

<sup>b</sup> Posterior probability

<sup>c</sup> Range

Table 3: Estimated marginal posterior probabilities of including predictors in the fixed and random effects components under different prior probabilities of being zero in the time-to-pregnancy application. Probabilities over 0.9 are written in bold.

Predictor	Posterior Probability of Inclusion					
	Fixed Effect			Random Effect		
	0.2	0.5	0.8	0.2	0.5	0.8
<u>Intercept</u>	<b>0.90</b> <sub>(0.89,0.94)<sup>a</sup></sub>	0.87 <sub>(0.85,0.90)</sub>	0.83 <sub>(0.81,0.87)</sub>	<b>0.94</b> <sub>(0.90,0.96)</sub>	<b>0.90</b> <sub>(0.86,0.92)</sub>	0.85 <sub>(0.82,0.87)</sub>
<u>Age</u>						
25-29	0.83 <sub>(0.80,0.85)</sub>	0.75 <sub>(0.73,0.78)</sub>	0.72 <sub>(0.69,0.75)</sub>	0.55 <sub>(0.52,0.58)</sub>	0.50 <sub>(0.46,0.52)</sub>	0.43 <sub>(0.39,0.46)</sub>
30+	<b>0.93</b> <sub>(0.90,0.96)</sub>	0.89 <sub>(0.86,0.92)</sub>	0.87 <sub>(0.84,0.90)</sub>	0.88 <sub>(0.84,0.91)</sub>	0.86 <sub>(0.83,0.90)</sub>	0.81 <sub>(0.76,0.84)</sub>
Overall	<b>0.97</b> <sub>(0.94,0.99)</sub>	<b>0.96</b> <sub>(0.94,0.98)</sub>	<b>0.95</b> <sub>(0.93,0.97)</sub>	<b>0.92</b> <sub>(0.89,0.95)</sub>	<b>0.90</b> <sub>(0.87,0.92)</sub>	0.87 <sub>(0.84,0.91)</sub>
<u>Intercourse frequency</u>						
1-3	0.63 <sub>(0.61,0.66)</sub>	0.56 <sub>(0.54,0.59)</sub>	0.53 <sub>(0.49,0.56)</sub>	0.61 <sub>(0.56,0.64)</sub>	0.56 <sub>(0.52,0.59)</sub>	0.52 <sub>(0.49,0.56)</sub>
3-4	0.76 <sub>(0.73,0.79)</sub>	0.73 <sub>(0.71,0.76)</sub>	0.68 <sub>(0.65,0.71)</sub>	0.83 <sub>(0.78,0.87)</sub>	0.78 <sub>(0.75,0.82)</sub>	0.74 <sub>(0.70,0.78)</sub>
4+	<b>0.97</b> <sub>(0.93,0.98)</sub>	<b>0.94</b> <sub>(0.91,0.96)</sub>	0.88 <sub>(0.85,0.91)</sub>	0.54 <sub>(0.49,0.57)</sub>	0.50 <sub>(0.45,0.53)</sub>	0.44 <sub>(0.40,0.48)</sub>
Overall	<b>0.97</b> <sub>(0.94,0.99)</sub>	<b>0.96</b> <sub>(0.94,0.98)</sub>	<b>0.96</b> <sub>(0.93,0.98)</sub>	<b>0.93</b> <sub>(0.90,0.96)</sub>	<b>0.92</b> <sub>(0.89,0.94)</sub>	<b>0.90</b> <sub>(0.86,0.93)</sub>
<u>Cigarettes smoked</u>						
1-5	0.70 <sub>(0.68,0.73)</sub>	0.65 <sub>(0.62,0.67)</sub>	0.56 <sub>(0.52,0.59)</sub>	0.61 <sub>(0.59,0.65)</sub>	0.55 <sub>(0.52,0.59)</sub>	0.51 <sub>(0.47,0.53)</sub>
6-10	0.85 <sub>(0.81,0.88)</sub>	0.81 <sub>(0.77,0.83)</sub>	0.74 <sub>(0.72,0.78)</sub>	0.89 <sub>(0.85,0.92)</sub>	0.85 <sub>(0.82,0.88)</sub>	0.82 <sub>(0.77,0.85)</sub>
11-15	0.86 <sub>(0.82,0.88)</sub>	0.79 <sub>(0.76,0.82)</sub>	0.70 <sub>(0.67,0.74)</sub>	0.57 <sub>(0.53,0.61)</sub>	0.52 <sub>(0.49,0.55)</sub>	0.49 <sub>(0.46,0.54)</sub>
15+	<b>0.95</b> <sub>(0.92,0.97)</sub>	<b>0.92</b> <sub>(0.89,0.94)</sub>	0.88 <sub>(0.84,0.91)</sub>	0.69 <sub>(0.65,0.75)</sub>	0.66 <sub>(0.62,0.70)</sub>	0.61 <sub>(0.58,0.65)</sub>
Overall	<b>0.97</b> <sub>(0.94,0.99)</sub>	<b>0.94</b> <sub>(0.91,0.98)</sub>	<b>0.92</b> <sub>(0.89,0.95)</sub>	<b>0.95</b> <sub>(0.93,0.98)</sub>	<b>0.92</b> <sub>(0.89,0.95)</sub>	<b>0.91</b> <sub>(0.87,0.94)</sub>
<u>Recent pill use</u>						
	<b>0.98</b> <sub>(0.96,0.99)</sub>	<b>0.91</b> <sub>(0.87,0.95)</sub>	0.87 <sub>(0.85,0.91)</sub>	0.53 <sub>(0.48,0.59)</sub>	0.48 <sub>(0.43,0.51)</sub>	0.40 <sub>(0.37,0.45)</sub>

<sup>a</sup> Range

Table 4: Overall posterior means and 95% credible intervals of regression coefficients in the time-to-pregnancy application compared with the results from the GLM with the logistic link.

Effects	Proposed approach	Standard GLM
<u>Age</u>		
25-29	0.173 <sub>(-0.043,0.396)</sub>	0.177 <sub>(-0.068,0.422)</sub>
30+	0.358 <sub>(0.003,0.727)</sub>	0.354 <sub>(-0.029,0.719)</sub>
<u>Intercourse frequency per week</u>		
1-3	0.095 <sub>(-0.144,0.335)</sub>	0.088 <sub>(-0.151,0.327)</sub>
3-4	0.258 <sub>(-0.120,0.643)</sub>	0.265 <sub>(-0.129,0.659)</sub>
4+	0.877 <sub>(0.402,1.306)</sub>	0.885 <sub>(0.408,1.361)</sub>
<u>Cigarettes smoked per day</u>		
1-5	-0.119 <sub>(-0.570,0.313)</sub>	-0.126 <sub>(-0.736,0.482)</sub>
6-10	-0.278 <sub>(-0.694,0.241)</sub>	-0.285 <sub>(-0.873,0.303)</sub>
11-15	-0.425 <sub>(-1.239,0.337)</sub>	-0.433 <sub>(-1.413,0.547)</sub>
15+	-0.686 <sub>(-1.640,-0.164)</sub>	-0.681 <sub>(-1.622,0.260)</sub>
<u>Use of oral contraceptives</u>		
	-0.923 <sub>(-1.544,-0.268)</sub>	-0.931 <sub>(-1.619,-0.243)</sub>

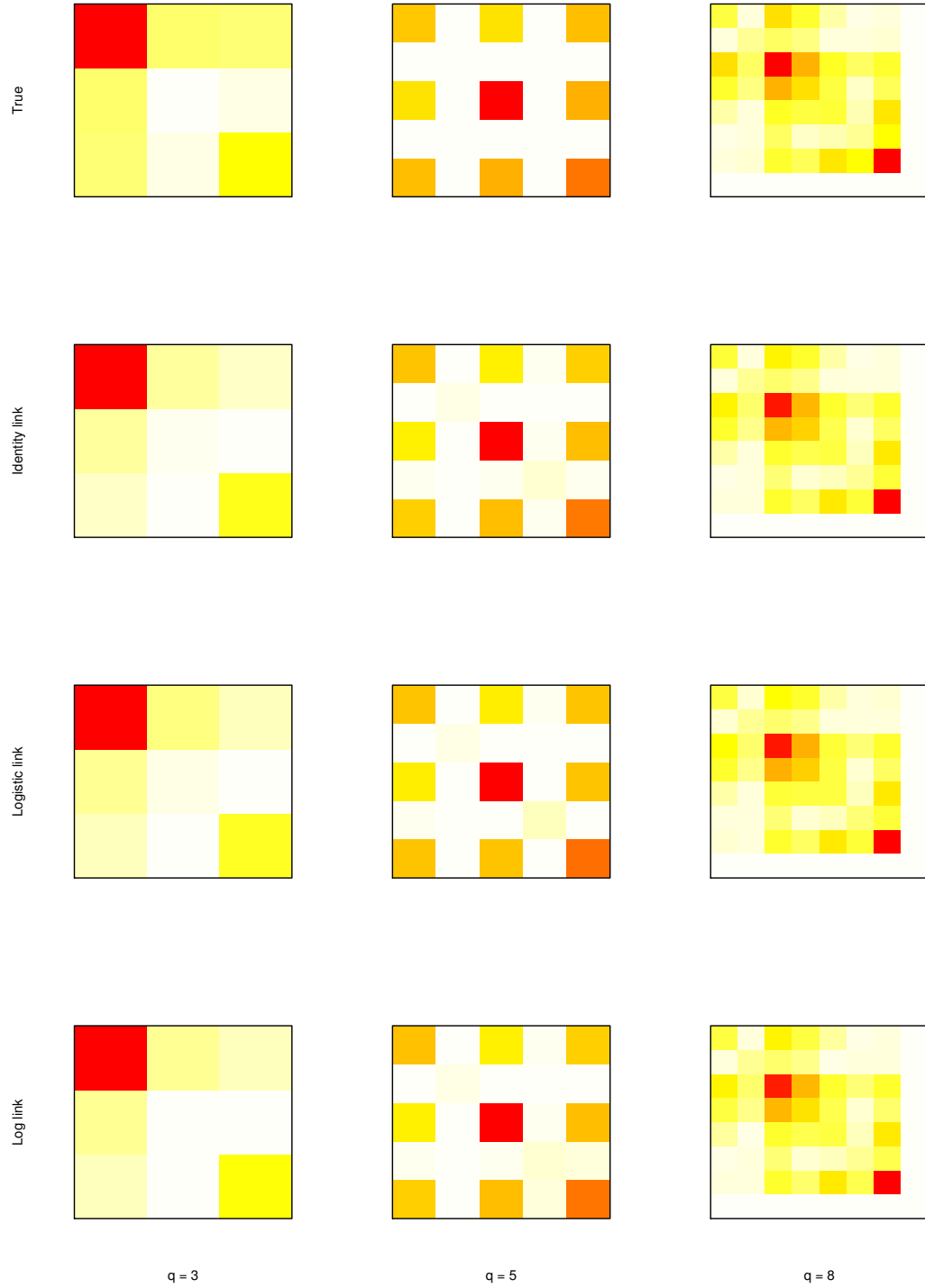


Figure 1: Image plots of the true and estimated random effects covariance matrices for simulated data under identity link, logistic link and log link with the number of candidate random effect predictors being 3, 5 and 8. The darker the color appears, the larger the value of the element is, with the white color corresponding to zero.

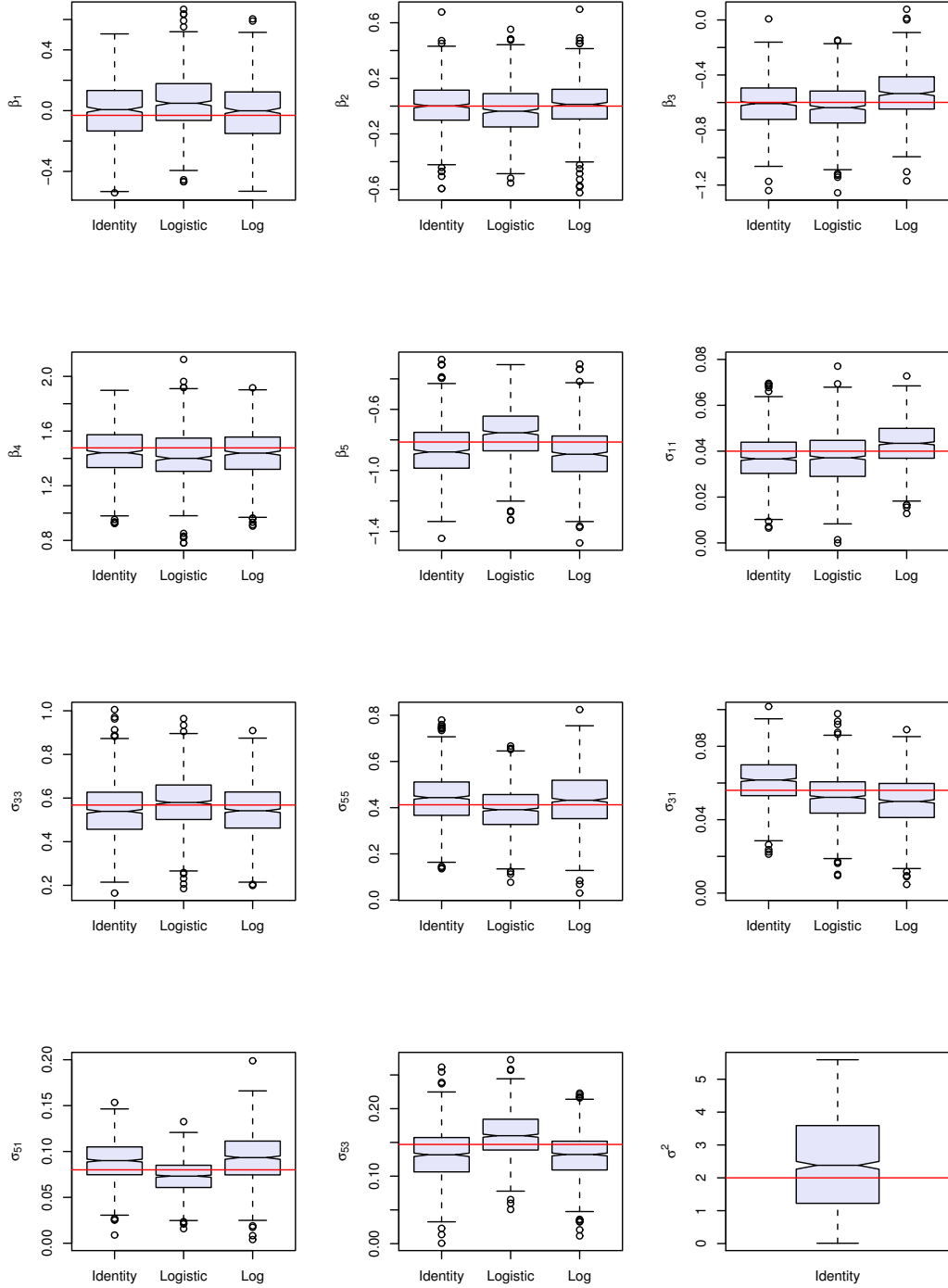


Figure 2: Boxplots of the samples of parameters for the second simulation under each link. The solid horizontal lines indicate the true values.