Fixed and random effects selection in linear and logistic models

Satkartar K. Kinney

Institute of Statistics and Decision Sciences, Duke University,
Box 90251, Durham, North Carolina 27705, U.S.A.
email: saki@stat.duke.edu

and

David B. Dunson

Biostatistics Branch, MD A3-03, National Institute of Environmental Health Sciences,
P.O. Box 12233, Research Triangle Park, North Carolina 27709, U.S.A.

August 23, 2006

SUMMARY. We address the problem of selecting which variables should be included in the fixed and random components of logistic mixed effects models for correlated data. A fully Bayesian variable selection is implemented using a stochastic search Gibbs sampler to estimate the exact model-averaged posterior distribution. This approach automatically identifies subsets of predictors having non-zero fixed effect coefficients or non-zero random effects variance, while allowing uncertainty in the model selection process. Default priors are proposed for the variance components and an efficient parameter expansion Gibbs sampler is developed for posterior computation. The approach is illustrated using simulated data and an epidemiologic example.

KEY WORDS: Bayesian model selection; Logistic regression; Mixed effects model; Model averaging; Parameter expansion; Random effects; Variance components test; Variable selection

1. Introduction

Logistic mixed models are widely used, flexible models for unbalanced repeated measures data. Often in linear and nonlinear mixed effects models, random effects are chosen to control for specific factors which are expected to cause random variation in the coefficients, such as batch effects and within-subject variation in repeated measurements. A difficult question is how to decide which predictors have coefficients that vary among subjects. Standard model selection criteria and test procedures are not appropriate for comparing models with different numbers of random effects due to constraints on the parameter space of the variance components.

This problem has motivated a growing literature on frequentist tests for homogeneity in variance components. In the setting of linear mixed models with one variance component, Crainiceanu and Ruppert (2004) derived finite and asymptotic distributions of likelihood ratio test statistics. Such results are not yet available for logistic mixed models. Lin (1997) proposed a score test for whether all variance components in a generalized linear mixed model (GLMM) are zero (See also Verbeke and Molenberghs, 2003 and Hall and Praestgaard, 2001); however, such methods cannot be used for general comparisons of models with different numbers of random effects. Jiang et al. (2006) recently proposed an innovated "fence" method to select predictors with random effects in linear mixed models; however, this approach does not allow inferences on whether a given predictor has a random component, and uncertainty in the model selection process is not accounted for.

Motivated by practical difficulties that arise in implementing a frequentist approach to this problem, we focus on Bayesian methods. An advantage of a Bayesian approach implemented with MCMC is that one bases inferences on estimates of the exact posterior distribution, with the estimation accuracy improving with number of MCMC iterations. Moreover, one has greater flexibility in comparing models with differing numbers of variance components. Potential disadvantages include computational burden and sensitivity to the prior.

A great body of work exists on Bayesian model selection for fixed effects; however, very little exists on selection of random effects. Bayesian methods for variable selection in regression models are described in Kuo and Mallick (1998) and George and McCulloch (1997) and Bayesian model selection and averaging are reviewed in Clyde and George (2004). Pauler, Wakefield and Kass (1999) compare variance component models using Bayes factors and Sinharay and Stern (2001) consider the problem of comparing two GLMMs using the Bayes factor. Motivated by sensitivity to the choice of prior, Chung and Dey (2002) develop an intrinsic Bayes factor approach for balanced variance component models. Chen and Dunson (2003) developed a more general stochastic search variable selection (George and McCulloch, 1993; Geweke, 1996) approach to the linear mixed effects model. Relying on Taylor series approximators to intractable integrals, Cai and Dunson (2006) recently extended this approach to all GLMMs. Although their approximation should be accurate in most cases, it is appealing to consider exact alternatives.

In this article we propose a related approach to that considered by Chen and Dunson (2003), but for logistic mixed effects model for binary data. We also revisit linear mixed models to carefully consider prior specification and computational efficiency. A new default prior structure based on a multivariate generalization of Gelman's (2005) folded-t prior is proposed. A parameter-expanded stochastic search Gibbs sampler is then developed for efficient posterior computation. Section 2 describes the approach for linear mixed models and Section 3 extends the approach to logistic mixed models and Section 4 suggests some additional extensions. Section 5 evaluates the approach using simulation examples and Section 6 gives a data example. Section 7 discusses the results.

2. Linear mixed models

The approach of Chen and Dunson (2003) allowed for Bayesian selection of predictors with fixed and random effects in the linear mixed effects (LME) model (Laird and Ware, 1982); however, their approach had the disadvantages of (i) relying on subjective priors that are difficult to elicit, and (ii) computational inefficiency due to slow mixing of the Gibbs sampler. Our initial goal is to address these two problems using a generalization of the approach of Gelman (2005), which focused on priors and computation for a single ANOVA model.

If we have n subjects under study, each with n_i observations, $i = 1, \dots, n$, let y_{ij} denote the jth response for subject i, $\boldsymbol{X}_{ij}$ a $p \times 1$ vector of predictors, and $\boldsymbol{Z}_{ij}$ a $q \times 1$ vector of predictors. Then the LME model is denoted as

$$y_{ij} = \boldsymbol{X}'_{ij}\boldsymbol{\beta} + \boldsymbol{Z}'_{ij}\boldsymbol{a}_i + \epsilon_{ij}, \quad \epsilon_{ij} \sim N(0, \sigma^2) \tag{1}$$

where $\boldsymbol{a}_i \sim N(\boldsymbol{0}, \boldsymbol{\Omega})$. Here $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)'$ are the fixed effects and $\boldsymbol{a}_i = (a_{i1}, \dots, a_{iq})'$ are the random effects. In practice $\boldsymbol{Z}_{ij}$ is typically chosen to be a subset of the predictors in $\boldsymbol{X}_{ij}$ believed to have random effects, often only the intercept for simplicity. If we let $\boldsymbol{X}_{ij}$ and $\boldsymbol{Z}_{ij}$ include all candidate predictors, then the problem of interest is to locate a subset of these predictors to be included in the model.

2.1 *General Approach and Review*

Our approach is to set up the problem as a Bayesian stochastic search variable selection (SSVS) problem. SSVS, originally proposed for subset selection in linear regression, searches for models having high posterior probability by (i) starting with the full model containing all p candidate predictors; (ii) choosing mixture priors that allow predictors to drop out by zeroing their coefficients; and (iii) running a Gibbs sampler relying on conditional conjugacy to sample from the posterior distribution. The resulting draws will differ in the subset of predictors having non-zero coefficients and, after discarding initial burn-in draws, one can estimate posterior model probabilities using the proportion of MCMC draws spent in each model. In general, all 2^{2p} models will not be visited; hence, many or most of the candidate models will be estimated to have zero posterior probability. Although there is no guarantee that the model with highest posterior probability will be visited when p is large, SSVS tends to quickly locate good models.

The SSVS approach has been applied successfully in a wide variety of regression applications, including challenging gene selection problems. One challenge in developing SSVS approaches for random effects models is the constraint that the random effects covariance matrix $\mathbf{\Omega}$ be positive semi-definite. Chen and Dunson (2003) addressed this problem by using a modified Cholesky decomposition of $\mathbf{\Omega}$:

$$\mathbf{\Omega} = \mathbf{\Lambda}\mathbf{\Gamma}\mathbf{\Gamma}'\mathbf{\Lambda}, \tag{2}$$

where $\mathbf{\Lambda}$ is a positive diagonal matrix with diagonal elements $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_q)'$ proportional to the random effects standard deviations, so that setting $\lambda_l = 0$ is equivalent to dropping the lth random effect from the model. $\mathbf{\Gamma}$ is a lower triangular matrix with diagonal elements equal to 1 and free elements that describe the random effects correlations. In the case of independent random effects, $\mathbf{\Gamma}$ is simply the identity matrix $\boldsymbol{I}$ and the diagonal elements $\lambda_l, l = 1, \ldots, q$ of $\mathbf{\Lambda}$ equal the random effects standard deviations.

Applying the covariance decomposition to (1) we have:

$$y_{ij} = \boldsymbol{X}'_{ij}\boldsymbol{\beta} + \boldsymbol{Z}'_{ij}\mathbf{\Lambda}\mathbf{\Gamma}\boldsymbol{b}_i + \epsilon_{ij}, \quad \epsilon_{ij} \sim N(0, \sigma^2) \tag{3}$$

where $\boldsymbol{b}_i \sim N(\mathbf{0}, \boldsymbol{I})$. Chen and Dunson (2003) show that by rearranging terms, the diagonal elements, $\lambda_l, l = 1, \ldots, q$, of $\mathbf{\Lambda}$ can be expressed as linear regression coefficients, conditional on $\mathbf{\Gamma}$ and $\boldsymbol{b}_i$. Similarly, the free elements $\gamma_k, k = 1, \ldots, q(q-1)/2$, of $\mathbf{\Gamma}$ can be expressed as linear regression coefficients, conditional on $\mathbf{\Lambda}$ and $\boldsymbol{b}_i$. Hence the variance parameters $\boldsymbol{\lambda}$ and $\boldsymbol{\gamma}$ have desirable conditional conjugacy properties for constructing a Gibbs sampling algorithm for sampling the posterior distribution and we are able to use the SSVS approach.

2.2 *Priors*

Prior selection is a key step in any Bayesian analysis; however, in this context it is particularly important as problems can arise when default priors are applied without caution. In particular, flat or excessively diffuse priors are not recommended for hierarchical models given the potential for an improper posterior and the difficulty of verifying propriety due to the intractable nature of the density, even when the output from a Gibbs chain seems reasonable (Hobert and Casella, 1996). Proper distributions are also desired for Bayes factors to be well-defined (Pauler et al., 1999).

A mixture of a point mass at zero and a normal or heavier-tailed distribution is a common choice of prior for fixed effects coefficients, $\beta_l, l = 1, \ldots, p$, in Bayesian model selection problems. Smith and Kohn (1996) introduce a vector $\boldsymbol{J}$ of indicator variables, where $J_l = 1$, indicates that the lth variable is in the model, $l = 1, \ldots, p$, and assign a Zellner g prior (Zellner and Siow, 1980) to $\boldsymbol{\beta_J}$, the vector of coefficients in the current model. As a notational convention, we let $\boldsymbol{\beta}$ denote the $p \times 1$ vector $(\{\beta_l : J_l = 1\} = \boldsymbol{\beta_J}, \{\beta_l : J_l = 0\} = \mathbf{0})$. Hence, conditional on the model index $\boldsymbol{J}$, the prior for $\boldsymbol{\beta}$ is induced through the prior for $\boldsymbol{\beta_J}$. Consistency issues can arise when comparing models based on these priors; however, for linear models, placing a conjugate gamma prior on g induces a t prior on the coefficients. In the special case where the t has degrees of freedom equal 1, the Cauchy distribution is induced, which has been recommended for Bayesian robustness (Clyde and George, 2004).

For standard deviation parameters in hierarchical models, Gelman (2005) recommends a family of folded-t prior distributions over the commonly used inverse gamma family, due to their flexibility and behavior when random effects are very small. These priors are induced using a parameter-expansion approach which has the added benefit of improving computational efficiency by reducing dependence among the parameters (Liu, Rubin and Wu, 1998; Liu and Wu, 1999). This yields a Gibbs sampler less prone to slow mixing when the standard deviations are near zero.

Extending the parameter expansion approach proposed by Gelman (2005) for simple variance component models to the LME model, we replace (3) with:

$$y_{ij} = \boldsymbol{X}'_{ij}\boldsymbol{\beta} + \boldsymbol{Z}'_{ij}\boldsymbol{A}\boldsymbol{\Gamma}\boldsymbol{\xi}_i + \epsilon_{ij}, \quad \epsilon_{ij} \sim N(0, \sigma^2), \tag{4}$$

where $\boldsymbol{\xi}_i \sim N(\mathbf{0}, \boldsymbol{D})$ and $\boldsymbol{A} = \mathrm{diag}(\alpha_1, \ldots, \alpha_q)'$ and $\boldsymbol{D} = \mathrm{diag}(d_1, \ldots, d_q)'$ are diagonal matrices, $\alpha_l \sim N(0, 1), l = 1, \ldots, q$, and $d_l \sim IG(\frac{1}{2}, \frac{N}{2}), l = 1, \ldots, q$, IG denoting the inverse gamma distribution. Note that the latent random effects have been multiplied by a redundant multiplicative parameter. In this case the implied covariance decomposition is $\boldsymbol{\Omega} = \boldsymbol{A}\boldsymbol{\Gamma}\boldsymbol{D}\boldsymbol{\Gamma}'\boldsymbol{A}$.

The parameters $\alpha_l, l = 1, \ldots, q$, are proportional to λ_l and thus to the random effects standard deviations, so setting $\alpha_l = 0$ effectively drops out the random effects for the lth predictor. When random effects are assumed to be uncorrelated, i.e., $\Gamma = I$ and $\lambda_l, l = 1, \ldots, q$ equal the random effects standard deviations, a folded t prior on $\lambda_l = |\alpha_l|\sqrt{d_l}, l = 1, \ldots, q$ is induced, as described in Gelman (2005). Generalizing to the case of correlated random effects, a folded-t prior is not induced; however, improved computational efficiency is still achieved, as illustrated in Section 5.

In our proposed prior structure we use a Zellner-type prior for the fixed effects components. Specifically, we let $\boldsymbol{\beta_J} \sim N\left(\mathbf{0}, \sigma^2(\mathbf{X^{J}}'\mathbf{X^J})^{-1}/g\right)$, $g \sim G(\frac{1}{2}, \frac{N}{2})$, $\sigma^2 \propto \frac{1}{\sigma^2}$, and $J_l \sim Be(p_0), l = 1, \ldots, p$, with Be denoting the Bernoulli distribution and $G(a, b)$ denoting the Gamma distribution with mean a/b and variance a/b^2. We give $\alpha_l, l = 1, \ldots, q$, a zero-inflated half-normal prior, $ZI - N^+(0, 1, p_{l0})$, where p_{l0} is the prior probability that $\alpha_l = 0$. Lastly, the free elements of $\boldsymbol{\Gamma}$ are treated as a $q(q-1)/2$-vector with prior $p(\boldsymbol{\gamma}|\boldsymbol{\alpha}) = N(\boldsymbol{\gamma}_0, \boldsymbol{V}_\gamma) \cdot 1(\boldsymbol{\gamma} \in \boldsymbol{R_\alpha})$ where $\boldsymbol{R_\alpha}$ constrains elements of $\boldsymbol{\gamma}$ to be zero when the corresponding random effects are zero. For simplicity, we do not allow uncertainty in which random effects are correlated.

2.3 *Posterior computation*

The joint posterior distribution for $\boldsymbol{\theta} = (\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}, \sigma^2)$ is given by:

$$p(\boldsymbol{\theta}|y) \propto \prod_{i=1}^{n} N_p(\boldsymbol{\xi}_i; \mathbf{0}, \boldsymbol{D}) \prod_{j=1}^{n_i} \left\{N(y_{ij}; \boldsymbol{X}'_{ij}\boldsymbol{\beta} + \boldsymbol{Z}'_{ij}\boldsymbol{A}\boldsymbol{\Gamma}\boldsymbol{\xi}_i, \sigma^2)\right\} p(\sigma^2)p(\boldsymbol{\beta}, \boldsymbol{J}, g)p(\boldsymbol{\alpha}, \boldsymbol{\gamma})p(\boldsymbol{D}) \tag{5}$$

This distribution has a complex form which we cannot sample from directly; instead we employ a parameter-expanded Gibbs sampler (Liu et al., 1998; Liu and Wu, 1999). The Gibbs sampler proceeds by iteratively sampling from the full conditional distributions of all parameters $\boldsymbol{\alpha}, \boldsymbol{\gamma}, \boldsymbol{\beta}, \sigma^2$, hyperparameters g and $\boldsymbol{J}$, and the diagonal elements $d_l, l = 1, \ldots, q$ of $\boldsymbol{D}$.

The full conditional posterior distributions are given in the Appendix and follow from (5) using straightforward algebraic routes. After discarding draws from the burn-in period, posterior summaries of model parameters can be estimated the usual way from the Gibbs sampler output. Models with high posterior probability can be identified as those appearing most often in the output and considered for further evaluation. Marginal inclusion probabilities for a given coefficient may also be calculated using the proportion of draws in which the coefficient is nonzero.

3. Logistic Mixed Models

Our approach for logistic mixed models is to formulate the model in such a way that its coefficients are conditionally linear and the SSVS approach can again be applied. This entails the use of a data augmentation strategy and approximation of the logistic density, with approximation error corrected for using importance weights. The covariance decomposition in (2) and parameter expansion approach described in Section 2.2 are again used.

Defining terms as in (3), the logistic mixed model for a binary response variable y is written as:

$$\text{logit}(P(y_{ij} = 1|\boldsymbol{X}_{ij}, \boldsymbol{Z}_{ij}, \boldsymbol{\beta}, \boldsymbol{a}_i)) = \boldsymbol{X}'_{ij}\boldsymbol{\beta} + \boldsymbol{Z}'_{ij}\boldsymbol{a}_i, \quad \boldsymbol{a}_i \sim N(\boldsymbol{0}, \boldsymbol{\Omega}) \tag{6}$$

We would like to be able to apply the SSVS approach as in the normal case. If we apply the covariance decomposition in (2) to the logistic mixed model, we have:

$$\text{logit}(P(y_{ij} = 1|\boldsymbol{X}_{ij}, \boldsymbol{Z}_{ij}, \boldsymbol{\beta}, \boldsymbol{\lambda}, \boldsymbol{\gamma}, \boldsymbol{b}_i)) = \boldsymbol{X}'_{ij}\boldsymbol{\beta} + \boldsymbol{Z}'_{ij}\boldsymbol{\Lambda}\boldsymbol{\Gamma}\boldsymbol{b}_i, \quad \boldsymbol{b}_i \sim N(\boldsymbol{0}, \boldsymbol{I}) \tag{7}$$

In this case the model is nonlinear and we do not immediately have conditional linearity for the variance parameters $\boldsymbol{\lambda}$ and $\boldsymbol{\gamma}$ as in the normal case. In order to obtain conditional linearity for the model coefficients, we take advantage of the fact that the logistic distribution can be closely approximated by the t distribution (Albert and Chib, 1993; Holmes and Knorr-Held, 2003; O'Brien and Dunson, 2004), and that the t distribution can be expressed as a scale mixture of normals (West, 1987).

First, note that (7) is equivalent to the specification:

$$y_{ij} \quad = \quad \begin{cases} 1 & w_{ij} > 0 \\ 0 & w_{ij} \leq 0 \end{cases},$$

where w_{ij} is a logistically distributed random variable with location parameter $\boldsymbol{X}'_{ij}\boldsymbol{\beta}+\boldsymbol{Z}'_{ij}\boldsymbol{\Lambda}\boldsymbol{\Gamma}\boldsymbol{b}_i$ and density function

$$\mathcal{L}(w_{ij}|\boldsymbol{X}_{ij}, \boldsymbol{Z}_{ij}, \boldsymbol{\beta}, \boldsymbol{\lambda}, \boldsymbol{\gamma}) = \frac{\exp\{-(w_{ij} - \boldsymbol{X}'_{ij}\boldsymbol{\beta} - \boldsymbol{Z}'_{ij}\boldsymbol{\Lambda}\boldsymbol{\Gamma}\boldsymbol{b}_i)\}}{\{1 + \exp[-(w_{ij} - \boldsymbol{X}'_{ij}\boldsymbol{\beta} - \boldsymbol{Z}'_{ij}\boldsymbol{\Lambda}\boldsymbol{\Gamma}\boldsymbol{b}_i)]\}^2}.$$

Then, as w_{ij} is approximately distributed as a non-central t_ν with location parameter $\boldsymbol{X}'_{ij}\boldsymbol{\beta} + \boldsymbol{Z}'_{ij}\boldsymbol{\Lambda}\boldsymbol{\Gamma}\boldsymbol{b}_i$ and scale parameter $\tilde{\sigma}^2$, we can express it as a scale mixture of normals and write:

$$w_{ij} = \boldsymbol{X}'_{ij}\boldsymbol{\beta} + \boldsymbol{Z}'_{ij}\boldsymbol{\Lambda}\boldsymbol{\Gamma}\boldsymbol{b}_i + \epsilon_{ij}, \quad \epsilon_{ij} \sim N(0, \tilde{\sigma}^2/\phi_{ij}) \tag{8}$$

where $\phi_{ij} \sim G(\frac{\nu}{2}, \frac{\nu}{2})$. Setting $\nu = 7.3$ and $\tilde{\sigma}^2 = \pi^2(\nu - 2)/3\nu$ makes the approximation nearly exact. The approximation error, though negligible except in the extreme tails, may be corrected for by importance weighting when making inferences. Under this model formulation, we have a model in which all coefficients are conditionally normal, and we are able to apply SSVS to the problem. We also are able to take advantage of the improved computational efficiency of a parameter expanded model as in (4). Applying the parameter expansion to (8) we have:

$$w_{ij} = \boldsymbol{X}'_{ij}\boldsymbol{\beta} + \boldsymbol{Z}_{ij}\boldsymbol{A}\boldsymbol{\Gamma}\boldsymbol{\xi}_i + \epsilon_{ij}, \quad \epsilon_{ij} \sim N(0, \tilde{\sigma}^2/\phi_{ij})$$

where terms are defined as in (4) and (8). We will use this model formulation to propose a prior structure and compute posterior distributions.

3.1 *Priors and posterior computation*

We use the same priors for the random effects parameters as in the normal case, and similar priors for the fixed effects parameters. We specify $\boldsymbol{\beta_J} \sim N\left(0, (\mathbf{X^{J}}'\mathbf{X^{J}})^{-1}/g\right)$, $g \sim G(\frac{1}{2}, \frac{N}{2})$, and $J_l \sim Be(p_0), l = 1, \ldots, p$. Using the t-distribution to approximate the likelihood as previously described, the joint posterior distribution for $\theta = (\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\phi})$ is given by:

$$p(\boldsymbol{\theta}|\mathbf{y}) \propto p(\boldsymbol{\beta}, \boldsymbol{J}, g)p(\boldsymbol{\gamma}, \boldsymbol{\alpha})p(\boldsymbol{D}) \left(\prod_{i=1}^{n} N_q(\boldsymbol{\xi}_i; \mathbf{0}, \boldsymbol{D}) \right. \\ \left. \times \prod_{j=1}^{n_i} \left[N\left(w_{ij}; \boldsymbol{X}_{ij}\boldsymbol{\beta} + \boldsymbol{Z}_{ij}\boldsymbol{A\Gamma\xi}_i, \frac{\tilde{\sigma}^2}{\phi_{ij}}\right) \{1(w_{ij} > 0)y_{ij} + 1(w_{ij} \leq 0)(1 - y_{ij})\} p(\phi_{ij}) \right] \right) \quad (9)$$

Again we have a complex posterior from which we cannot directly sample and we employ a Gibbs sampler. In introducing a latent variable w_{ij} we have applied a data augmentation strategy related to Albert and Chib (1993) and used for multivariate logistic models by O'Brien and Dunson (2004). This auxiliary variable is updated in the Gibbs sampler and its full conditional posterior follows immediately from (9) as a normal distribution truncated above or below by 0 depending on y_{ij}:

$$p(w_{ij}|\boldsymbol{\theta}, y_{ij}) = \frac{N\left(w_{ij}; \boldsymbol{X}_{ij}\boldsymbol{\beta} + \boldsymbol{Z}'_{ij}\boldsymbol{A\Gamma\xi}_i, \frac{\tilde{\sigma}^2}{\phi_{ij}}\right) \cdot 1\left((-1)^{y_{ij}} w_{ij} < 0\right)}{\Phi\left(0; \boldsymbol{X}'_{ij}\boldsymbol{\beta} + \boldsymbol{Z}'_{ij}\boldsymbol{A\Gamma\xi}_i, \frac{\tilde{\sigma}^2}{\phi_{ij}}\right)^{1-y_{ij}} \left\{1 - \Phi\left(0; \boldsymbol{X}'_{ij}\boldsymbol{\beta} + \boldsymbol{Z}'_{ij}\boldsymbol{A\Gamma\xi}_i, \frac{\tilde{\sigma}^2}{\phi_{ij}}\right)\right\}^{y_{ij}}} \quad (10)$$

where $\Phi(\cdot)$ indicates the normal cumulative distribution function. The Gibbs sampler proceeds by iteratively sampling from the full conditional distributions of all parameters $\boldsymbol{\alpha}, \boldsymbol{\gamma}, \boldsymbol{\beta}, \boldsymbol{\phi}$, hyperparameters g and $\boldsymbol{J}$, as well as the latent variable $\boldsymbol{\xi}_i, i = 1, \ldots, n$ and the diagonal elements $d_l, l = 1, \ldots, q$ of $\boldsymbol{D}$. The full conditional posterior distributions follow from (9). They are similar in form to the normal case and are given in the Appendix.

This Gibbs sampler generates samples from an approximate posterior as we have approximated the logistic likelihood in (8). To correct for this, importance weights (Hastings, 1970) may be applied when computing posterior summaries to obtain exact inferences. If we have M iterations of our Gibbs sampler, excluding the burn-in interval, then our importance weights $r^{(t)}, t = 1, \ldots, M$ can be computed as:

$$r^{(t)} = \prod_{i=1}^{n} \prod_{j=1}^{n_i} \frac{\mathcal{L}(w_{ij}; \boldsymbol{X}'_{ij}\boldsymbol{\beta} + \boldsymbol{Z}'_{ij}\boldsymbol{A\Gamma\xi}_i)}{\mathcal{T}_\nu(w_{ij}; \boldsymbol{X}'_{ij}\boldsymbol{\beta} + \boldsymbol{Z}'_{ij}\boldsymbol{A\Gamma\xi}_i, \tilde{\sigma}^2)}$$

where $\mathcal{L}(\cdot)$ is the logistic density function and $\mathcal{T}_\nu(\cdot)$ is the t density function with degrees of freedom ν.

Posterior means, probabilities, and other summaries of the model parameters can be estimated from the Gibbs sampler output using an importance-weighted sample average. For example, the posterior probability for a given model m is the sum of the weights corresponding to each occurence of model m in the posterior sample, divided by the sum of all M weights. The approximation is very close and hence the weights are close to one. In our simulation and data examples we found very little difference between weighted and unweighted results.

In lieu of approximating the logistic distribution with the t distribution, we also considered the slice sampler for sampling from the exact posterior distribution as applied by Gerlach, Bird and Hall (2002) to variable selection for logistic models. In this approach, the model is considered linear with response variable $v_{ij} = \text{logit}(p(y_{ij} = 1))$, the vector of log odds. This vector is updated in a data augmented Gibbs sampler where an auxiliary variable $u_{ij} \sim U\left(0, \frac{1}{1+\exp(w_{ij})}\right)$ is introduced so that the full conditional posterior distribution for v_{ij} is simplified to a truncated normal distribution. While slice sampling in general has been noted to have appealing theoretical properties (Neal, 2000; Mira and Tierney, 2002), it demonstrated unsatisfactory convergence properties for our purposes due to asymmetries induced by the likelihood (Green, 1997). In simulations using the slice sampler approach, the correct models were quickly located; however, the Gibbs chains for nonzero model coefficients were extremely slow to converge.

4. Other models

4.1 *Logistic models for ordinal data*

This framework can also be adapted to accommodate logistic mixed models with ordinal response variables $y_{ij} \in \{1, \ldots, C\}$:

$$\text{logit}(P(y_{ij} \leq c | \boldsymbol{X}_{ij}, \boldsymbol{Z}_{ij}, \boldsymbol{\beta}, \boldsymbol{a}_i, \boldsymbol{\tau})) = \tau_c - \boldsymbol{X}'_{ij}\boldsymbol{\beta} - \boldsymbol{Z}'_{ij}\boldsymbol{a}_i, \quad c \in \{1, \ldots, C\} \tag{11}$$

where terms in the linear predictor are as defined in (3) and $\boldsymbol{\tau} = (\tau_1, \ldots, \tau_{C-1})'$ where $\tau_1 = 0$ for identifiability and $-\infty = \tau_0 < \tau_1 < \cdots < \tau_C = \infty$ are threshold parameters for the ordered categories. Our data augmentation stochastic search Gibbs sampler can be applied to (11) with modifications to truncate w_{ij} to $[\tau_{c-1}, \tau_c]$ for $y_{ij} = c$ and to update the threshold parameters $\boldsymbol{\tau}$. Although updating of $\boldsymbol{\tau}$ can potentially proceed after augmentation as described in Albert and Chib (1993), such an approach has a tendency to mix very slowly (Johnson and Albert, 1999). A modification in which the latent variables $\{w_{ij}\}$ are integrated out and a Metropolis-Hastings step is used yields better results. An alternative, which allows the baseline parameters $\boldsymbol{\tau}$ to be updated jointly from a multivariate normal posterior after augmentation, is to consider a continuation-ratio logit formulation instead of (11) (Agresti, 1990). Such formulations characterize the ordinal distribution in terms of the discrete hazard so are natural in time to event applications (Albert and Chib, 2001).

4.2 *Probit models*

Logistic models are often preferred over probit models due to the more intuitive interpretation of their regression coefficients in terms of odds ratios; however, it is worth noting that our approach for normal models is easily modified to accomplish model selection for probit mixed models by applying the well-known data augmentation Gibbs sampler described in Albert and Chib (1993). In our simulations this MCMC algorithm exhibited good mixing and convergence properties.

5. Simulation Examples

We evaluate the proposed approach using a simulation example for a binary response logistic model. We generate 3 covariates from $U(-2, 2)$ for 30 observations on each of 200 subjects, so $\boldsymbol{X}_{ij} = (1, X_{ij1}, X_{ij2}, X_{ij3})'$ and we let $\boldsymbol{Z}_{ij} = \boldsymbol{X}_{ij}$. We let $\boldsymbol{\beta} = (1, 0, 1, 1)'$ and $\alpha_i \sim N(\boldsymbol{0}, \boldsymbol{\Omega})$, choosing a range of realistic values for the random effects variances:

$$\boldsymbol{\Omega} = \begin{pmatrix} .90 & .48 & .06 & 0 \\ .48 & .40 & .10 & 0 \\ .06 & .10 & .10 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

We generate $\text{logit}(P(y_{ij} = 1))$ according to model (6) and then y_{ij} from $\text{Be}(p(y_{ij}))$. We follow the prior specification outlined in Section 3.1 and induce heavy-tailed priors on the fixed effects coefficients and random effects variances. These default priors do not require subjective choice of hyperparameter values, with the exception of the prior inclusion probabilities, which can be chosen as $p = 0.5$ to give equal probability to inclusion and exclusion, and the prior mean and variance of $\boldsymbol{\gamma}$. Our prior specification does include an informative normal prior for γ; however, γ is scaled in the parameter-expanded model and hence the informative prior can reasonably be chosen. A prior that modestly shrinks the correlations towards zero is desirable for stable estimation while still allowing the data to inform the relationships between the random effects. As a reasonable choice, we set the prior mean and variance for γ to be $\mathbf{0}$ and $0.5\boldsymbol{I}$, which can be used as a default in other applications.

We ran the Gibbs sampler for 20,000 iterations, after a burnin period of 5000 iterations. Three chains with dispersed starting values were run and found to converge after a few thousand iterations. The resulting MCMC chains for the random effects variances are shown in Figure 1 and the posterior means for the fixed effects coefficients and random effects variances are given in Table 1, along with the PQL estimates computed by glmmPQL in R.

We compare our results to the penalized quasi-likelihood (PQL) approach (Breslow and Clayton, 1993), as this approach is widely used for estimating GLMMs. Although our focus is on selection and inferences in the variance components allowing for model uncertainty, which is not addressed by current frequentist methods, we also obtain model-averaged coefficient estimates. Based on the limited number of simulations run, our estimates tend to be closer to the true values than the PQL estimates, which are known to be biased (Breslow, 2003; Jang and Lim, 2005). Our algorithm is too computationally intense to run a large enough simulation to definitively assess the frequentist operating characteristics of our approach.

We also compute credible intervals for the random effects variances; whereas, there is no good approach for estimating uncertainty for PQL variance estimates, or other frequentist variance estimates. In addition we are able to simultaneously compute marginal posterior inclusion probabilities for both the fixed effects and random effects and correctly locate the true model as the one with highest posterior probability.

To evaluate sensitivity to the prior inclusion probability, we also repeated the simulation with prior probabilities set to 0.2 and 0.8 and found very little difference in the posterior means shown in Table 1. Posterior model probabilities were slightly different when the prior inclusion probabilities were changed; however there was no difference in parameter estimates, inferences or model ranking. In each case the true model had the highest posterior probability.

To evaluate the effect of using the priors induced by the parameter expanded model, we compare simulation results between two Gibbs samplers, one including and one excluding the redundant multiplicative parameter in the random effects component. As expected, we do not see any real difference in the point estimates; however, as seen in Figure 2, the parameter expansion approach resulted in improved computational efficiency and MCMC chains for the random effects variances. Table 2 shows the reduction in autocorrelation in the Gibbs chains. Note we have not directly drawn from the posterior distribution of the variances, rather we have computed them from the MCMC draws for $\boldsymbol{\alpha}, \boldsymbol{\gamma}, \boldsymbol{\lambda}$ and $\mathbf{d}$. The overparameterization causes the Gibbs chains for these parameters to mix poorly, but in combination they produce well-behaved chains for the random effects variances.

6. Epidemiology Application

As a motivating example, we consider data from the Collaborative Perinatal Project (CPP) conducted between 1959 and 1966. We examine the effect of DDE, a metabolite of DDT, as measured in maternal serum, on pregnancy loss, a binary response variable. Potential confounding variables include mother's age, body mass index, smoking status, and serum levels of cholesterol and triglycerides. Data were collected across twelve different study centers and there is potential for heterogeneity across centers. We are interested in selecting a logistic mixed effects model relating DDE levels and pregnancy loss, accounting for heterogeneity among study centers in those factors that vary in their effects across centers. In addition, inferences on whether predictors such as DDE vary in their effect is of substantial interest.

We let our binary response variable $y_{ij} = 1$ indicate pregnancy loss for participant j in study center i, $i = 1, \dots, 12; j = 1, \dots, n_i$, for 5389 total participants. Our covariate vector is $\boldsymbol{X}_{ij} = (1, X_{ij1}, \dots, X_{ij5})'$ where X_{ij1} is the level of DDE, and $X_{ij2}, \dots, X_{ij5}$ are the potential confounding variables. All covariates are continuous and centered at their means, and we let $\boldsymbol{Z}_{ij} = \boldsymbol{X}_{ij}$, thus considering all coefficients, including the intercept, for possible heterogeneity among centers.

Priors were chosen as in the simulation example and the Gibbs sampler run for 30,000 iterations after a burnin period of 5,000. The Gibbs sampling results indicate that there is no heterogeneity present among study centers and that a fixed effects model is appropriate. The preferred model, as shown in Table 3, includes only the intercept, body mass index, and age, as predictors. The posterior means for all variances are close to zero. A few models with nonzero posterior probability do contain a random effect. The posterior means for the fixed effect are similar to the PQL results returned by glmmPQL in R for the full model, shown in Table 4. These results also show that DDE did not have an appreciable effect on pregancy loss in the CPP study. The PQL results indicate that DDE had a very small but statistically significant effect; however, this may be due to bias in the PQL approach. Applying the BIC criteria to select the best fixed effects model, we obtain the high posterior probability model shown in Table 3.

7. Discussion

This article has proposed a new method for Bayesian model selection of fixed and random effects in continuous and binary response models. Our approach relies on a generalization of the folded-t priors and parameter expansion algorithm of Gelman (2005) to the setting of variable selection of multivariate, correlated random effects. Clear advantages over an earlier approach proposed by Chen and Dunson (2003) include robustness using default priors, efficiency of posterior computation and use of data augmentation to allow binary or categorical responses.

Our proposed approach is advantageous in that fixed and random effects are selected simultaneously. Additionally, the Bayesian framework allows for marginal posterior inclusion probabilities to be computed for each predictor along with model-averaged coefficient estimates. Posterior model probabilities can be used to compare models; whereas frequentist testing for variance components is more limited. In addition to model selection and averaging, the proposed prior structure and computational algorithm should be useful for efficient Gibbs sampling for fitting single mixed effects models. In particular, the prior and computational algorithm represent a useful alternative to approaches that rely on inverse-Wishart priors for variance components (e.g.Gilks et al., 1993). There is an increasing realization that inverse-Wishart priors are a poor choice, particularly when limited prior information is available. Although we have focused on LMEs of the Laird and Ware (1982) type, it is straightforward to adapt our methods for a broader class of linear mixed models, accomodating varying coefficient models, spatially correlated data, and other applications (Zhao et al., 2006).

Gibbs sampling chains from random effects model parameters tend to exhibit slow mixing and convergence. Gelfand, Sahu and Carlin (1996) recommend hierarchical centering for improved convergence and posterior surface behavior. Vines, Gilks and Wild (1994) also propose a transformation of random effects to improve mixing. A challenge in implementing the hierarchically centered model is to efficiently update the correlation matrix in the context of random effects selection where we are interested in separating out the variances. One solution is proposed by Chib and Greenberg (1998); however, it is prohibitively slow for more than a couple random effects.

ACKNOWLEDGEMENTS

This work was supported in part by the Intramural Research Program of the NIH, National Institute of Environmental Health Sciences, and by the National Science Foundation under Agreement No. DMS-0112069 with the Statistical and Mathematical Sciences Institute (SAMSI). The authors are grateful to Merlise Clyde and Abel Rodriguez at Duke University and to the participants of the Model Uncertainty working group of the SAMSI Program on Latent Variable Modeling in the Social Sciences.

REFERENCES

Agresti, A. (1990). *Categorical Data Analysis*. New York: John Wiley & Sons.

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

Albert, J. H. and Chib, S. (2001). Sequential ordinal modeling with applications to survival data. *Biometrics* **57**, 829–836.

Breslow, N. (2003). Whither PQL? *UW Biostatistics Working Paper Series* Working Paper 192.

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

Cai, B. and Dunson, D. B. (2006). Bayesian covariance selection in generalized linear mixed models. *Biometrics* **62**, 446–457.

Chen, Z. and Dunson, D. B. (2003). Random effects selection in linear mixed models. *Biometrics* **59**, 762–769.

Chib, S. and Greenberg, E. (1998). Bayesian analysis of multivariate probit models. *Biometrika* **85**, 347–361.

Chung, Y. and Dey, D. (2002). Model determination for the variance component model using reference priors. *Journal of Statistical Planning and Inference* **100**, 49–65.

Clyde, M. and George, E. I. (2004). Model uncertainty. *Statistical Science* **19**, 81–94.

Crainiceanu, C. and Ruppert, D. (2004). Likelihood ratio tests in linear mixed models with one variance component. *Journal of the Royal Statistical Society B* **66**, 165–185.

Gelfand, A. E., Sahu, S. K. and Carlin, B. P. (1996). Efficient parameterizations for generalized linear mixed models. *Bayesian Statistics* **5**.

Gelman, A. (2005). Prior distributions for variance parameters in hierarchical models. *Bayesian Analysis* **1**, 1–19.

George, E. I. and McCulloch, R. E. (1993). Variable selection via Gibbs sampling. *Journal of the American Statistical Association* **88**, 881–889.

George, E. I. and McCulloch, R. E. (1997). Approaches for Bayesian variable selection. *Statistica Sinica* **7**, 339–374.

Gerlach, R., Bird, R. and Hall, A. (2002). Bayesian variable selection in logistic regression: Predicting company earnings direction. *Australian and New Zealand Journal of Statistics* **42**, 155–168.

Geweke, J. (1996). Variable selection and model comparison in regression. In *Bayesian Statistics 5 - Proceedings of the Fifth Valencia International Meeting*, pages 609–620.

Gilks, W., Wang, C., Yvonnet, B. and Coursaget, P. (1993). Random-effects models for longitudinal data using Gibbs sampling. *Biometrics* **49**, 441–453.

Green, P. J. (1997). Discussion of "The EM algorithm - an old folk song sung to a fast new tune," by Meng and van Dyk. *Journal of the Royal Statistical Society, Series B* **59**, 554–555.

Hall, D. and Praestgaard, J. (2001). Order-restricted score tests for homogeneity in generalised linear and nonlinear mixed models. *Biometrika* **88**, 739–751.

Hastings, W. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* **57**, 97–109.

Hobert, J. P. and Casella, G. (1996). The effect of improper priors on Gibbs sampling in hierarchical linear mixed models. *Journal of the American Statistical Association* **91**.

Holmes, C. and Knorr-Held, L. (2003). Efficient simulation of Bayesian logistic regression models. Technical report, Ludwig Maximilians University Munich.

Jang, W. and Lim, J. (2005). Estimation bias in generalized linear mixed models. Technical report, Institute for Statistics and Decision Sciences, Duke University.

Jiang, J., Rao, J., Gu, Z. and Nguyen, T. (2006). Fence methods for mixed model selection. *Submitted* .

Johnson, V. E. and Albert, J. H. (1999). *Ordinal Data Modeling*. New York: Springer-Verlag.

Kuo, L. and Mallick, B. (1998). Variable selection for regression models. *Sankhya: The Indian Journal of Statistics, Series B* **60**, 65–81.

Laird, N. and Ware, J. (1982). Random-effects models for longitudinal data. *Biometrics* **38**, 963–974.

Lin, X. (1997). Variance component testing in generalised linear models with random effects. *Biometrika* **84**, 309–326.

Liu, C., Rubin, D. B. and Wu, Y. N. (1998). Parameter expansion to accelerate EM: the PX-EM algorithm. *Biometrika* **85**, 755–770.

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

Mira, A. and Tierney, L. (2002). Efficiency and convergence properties of slice samplers. *Scandinavian Journal of Statistics* **29**, 1–12.

Neal, R. M. (2000). Slice sampling. Technical report, Department of Statistics, University of Toronto.

O'Brien, S. M. and Dunson, D. B. (2004). Bayesian multivariate logistic regression. *Biometrics* **60**, 739–746.

Pauler, D. K., Wakefield, J. C. and Kass, R. E. (1999). Bayes factors and approximations for variance component models. *Journal of the Americal Statistical Association* **94**.

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* pages 507–516.

Smith, M. and Kohn, R. (1996). Nonparametric regression using Bayesian variable selection. *Journal of Econometrics* **75**, 317–343.

Verbeke, G. and Molenberghs, G. (2003). The use of score tests for inference on variance components. *Biometrics* **59**, 254–262.

Vines, S., Gilks, W. and Wild, P. (1994). Fitting Bayesian multiple random effects models. Technical report, Biostatistics Unit, Medical Research Council, Cambridge.

West, M. (1987). On scale mixtures of normal distributions. *Biometrika* **74**, 646–648.

Zellner, A. and Siow, A. (1980). Posterior odds ratios for selected regression hypotheses. In *Bayesian Statistics: Proceedings of the First International Meeting held in Valencia (Spain).*

Zhao, Y., Staudenmayer, J., Coull, B. and Wand, M. (2006). General design Bayesian generalized linear mixed models. *Statistical Science* **21**, 35–51.

APPENDIX

Full conditional distributions

Distributions are presented for the logistic case. Except where noted, the corresponding distribution for the normal case is obtained by substituting y_{ij} for w_{ij}, and σ^2 for $\tilde{\sigma}^2/\phi_{ij}$. Additionally, σ^2 is sampled only in the normal case, and w_{ij} and ϕ_{ij} only in the logistic case. Let $\boldsymbol{X}^{\mathbf{J}}_{ij}$ denote the subvector of X_{ij}, $\{X_{ijl} : J_l = 1\}$ and $\boldsymbol{\psi}$ be the N-vector such that $\psi_{ij} = w_{ij} - \mathbf{X^J}'_{ij}\boldsymbol{\beta} - \boldsymbol{Z}'_{ij}\boldsymbol{A\Gamma\xi}_i$. R code for sampling from these distributions is available through the first author's website.

- The full conditional posterior $p(\boldsymbol{\beta_J}|\boldsymbol{J}, \boldsymbol{\alpha}, \boldsymbol{\gamma}, \boldsymbol{\phi}, \boldsymbol{\xi}, \boldsymbol{y}, \boldsymbol{X}, \boldsymbol{Z})$ is $N(\hat{\boldsymbol{\beta}}_{\mathbf{J}}, \boldsymbol{V}_{\mathbf{J}})$ where

$$\hat{\boldsymbol{\beta}}_{\mathbf{J}} = \left(\sum_{i=1}^{n} \sum_{j=1}^{n_i} \frac{\phi_{ij}}{\tilde{\sigma}^2} \psi_{ij} \mathbf{X^J}'_{ij} \right) \cdot \boldsymbol{V}_{\mathbf{J}} \text{ and } \boldsymbol{V}_{\mathbf{J}} = \left(\sum_{i=1}^{n} \sum_{j=1}^{n_i} \mathbf{X^J}_{ij} \mathbf{X^J}'_{ij} \left(\frac{\phi_{ij}}{\tilde{\sigma}^2} + g \right) \right)^{-1}$$

- To calculate the posterior for $\boldsymbol{J}$ we need to update each J_l individually. We calculate $p(J_l = 1|\boldsymbol{J}_{-l}, \boldsymbol{\alpha}, \boldsymbol{\gamma}, \boldsymbol{\phi}, \boldsymbol{\xi}, \boldsymbol{y}, \boldsymbol{X}, \boldsymbol{Z})$ for $l = 1, \ldots, p$, by integrating out $\boldsymbol{\beta}$ as in Smith and Kohn (1996) and obtaining $p(J_l = 1|\boldsymbol{J}_{-l}, \boldsymbol{\alpha}, \boldsymbol{\gamma}, \boldsymbol{\phi}, \boldsymbol{\xi}, \boldsymbol{y}, \boldsymbol{X}, \boldsymbol{Z}) = \frac{1}{1+h_l}$, where $J_{-l} = \{J_i : i \neq l\}$ and $h_l = \frac{1-p_{l0}}{p_{l0}} \cdot c^{1/2} \cdot \frac{S(J_l=0)}{S(J_l=1)}$. For the logistic case we set $c = \frac{1}{g}$ and define

$$S(\boldsymbol{J}) = |\mathbf{X}^{\mathbf{J}\prime}\mathbf{X}^{\mathbf{J}}|^{1/2} \cdot |\boldsymbol{V}_{\mathbf{J}}|^{1/2} \exp\left\{-\frac{1}{2}\left(\sum_{i=1}^{n}\sum_{j=1}^{n_i}\phi_{ij}\psi_{ij}^2 - \hat{\boldsymbol{\beta}}_{\mathbf{J}}'\boldsymbol{V}_{\mathbf{J}}^{-1}\hat{\boldsymbol{\beta}}_{\mathbf{J}}\right)\right\}$$

  and for the normal case we set $c = 1 + \frac{1}{g}$. We integrate out σ^2 and define:

$$S(\boldsymbol{J}) = \left(\boldsymbol{\psi}'\boldsymbol{\psi} - \hat{\boldsymbol{\beta}}_{\mathbf{J}}\boldsymbol{V}_{\mathbf{J}}^{-1}\hat{\boldsymbol{\beta}}_{\mathbf{J}}\right)^{-N/2}$$

  $S(J_l = 0)$ is equivalent to $S(\boldsymbol{J})$ but with the element J_l of $\boldsymbol{J}$ set to 0, so $\boldsymbol{\psi}$, $\mathbf{X}^{\mathbf{J}}$, $\hat{\boldsymbol{\beta}}_{\mathbf{J}}$ and $\boldsymbol{V}_{\mathbf{J}}$ may need to be recomputed to correspond to $J_l = 0$. Similarly for $S(J_l = 1)$.

- The gamma prior on g yields a conjugate gamma posterior:

$$\Gamma\left(\frac{p_{\mathbf{J}}+1}{2}, \frac{\boldsymbol{\beta}_{\boldsymbol{J}}'\mathbf{X}^{\mathbf{J}\prime}\mathbf{X}^{\mathbf{J}}\boldsymbol{\beta}_{\boldsymbol{J}}/\sigma^2 + N}{2}\right)$$

  where $p_{\mathbf{J}} = \sum_{l=1}^{p} 1(J_l = 1)$. Set σ^2 =1 for the logistic case.

- Similary each ϕ_{ij} has a conjugate gamma posterior:

$$G\left(\frac{\nu+1}{2}, \frac{(w_{ij} - \boldsymbol{Z}_{ij}\boldsymbol{A}\boldsymbol{\Gamma}\boldsymbol{\xi}_i - \boldsymbol{X}_{ij}'\boldsymbol{\beta})^2/\tilde{\sigma}^2 + \nu}{2}\right)$$

- The posterior for $p(\sigma^2|\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{J}, \gamma, \boldsymbol{\xi}, \boldsymbol{y}, \boldsymbol{X}, \boldsymbol{Z})$ is:

$$IG\left(\frac{N + p_{\mathbf{J}}}{2}, \frac{\boldsymbol{\psi}'\boldsymbol{\psi} + g\boldsymbol{\beta}_{\boldsymbol{J}}'\mathbf{X}^{\mathbf{J}\prime}\mathbf{X}^{\mathbf{J}}\boldsymbol{\beta}_{\boldsymbol{J}}}{2}\right)$$

- The full conditional posterior $p(\gamma|\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\lambda}, \boldsymbol{\xi}, \boldsymbol{\phi}, \boldsymbol{y}, \boldsymbol{X}, \boldsymbol{Z})$ is given by $N(\hat{\gamma}, \hat{\boldsymbol{V}}_{\gamma}) \cdot 1(\gamma \in \boldsymbol{R}_{\boldsymbol{\lambda}})$ where

$$\hat{\boldsymbol{V}}_{\gamma} = \left(\sum_{i=1}^{n}\sum_{j=1}^{n_i}\frac{\phi_{ij}}{\tilde{\sigma}^2}\boldsymbol{u}_{ij}\boldsymbol{u}_{ij}' + \boldsymbol{V}_{\boldsymbol{\gamma}}^{-1}\right)^{-1} \text{ and } \hat{\gamma} = \left(\sum_{i=1}^{n}\sum_{j=1}^{n_i}\frac{\phi_{ij}}{\tilde{\sigma}^2}(w_{ij} - \mathbf{X}^{\mathbf{J}}{}_{ij}\boldsymbol{\beta}_{\boldsymbol{J}})u_{ij}' + \boldsymbol{\gamma}_0\boldsymbol{V}_{\boldsymbol{\gamma}}^{-1}\right) \cdot \hat{\boldsymbol{V}}_{\boldsymbol{\gamma}}$$

  The $q(q-1)/2$ vector $\boldsymbol{u}_{ij}$ is defined as $(\xi_{il}\alpha_m Z_{ijm} : l = 1, \ldots, q, m = l+1, \ldots, q)'$ so that the random effects term $\boldsymbol{Z}_{ij}'\boldsymbol{A}\boldsymbol{\Gamma}\boldsymbol{\xi}_i$ can be written as $\boldsymbol{u}_{ij}'\boldsymbol{\gamma}$.

- The latent variables $\boldsymbol{\xi}_i$ have posterior $p(\boldsymbol{\xi}_i|\boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{\gamma}, \boldsymbol{\phi}, \boldsymbol{y}, \boldsymbol{X}, \boldsymbol{Z})$ given by $N(\hat{\boldsymbol{\xi}}_i, \boldsymbol{V}_\xi)$ where

$$\hat{\boldsymbol{\xi}}_i = \sum_{j=i}^{n_i} \phi_{ij}(w_{ij} - \mathbf{X^J}'_{ij}\boldsymbol{\beta_J})\boldsymbol{Z}'_{ij}\boldsymbol{A}\boldsymbol{\Gamma}\boldsymbol{V}_\xi\tilde{\sigma}^{-2} \text{ and } \boldsymbol{V}_\xi = \left(\sum_{j=1}^{n_i} \phi_{ij}\boldsymbol{\Gamma}'\boldsymbol{A}\boldsymbol{Z}_{ij}\boldsymbol{Z}'_{ij}\boldsymbol{A}\boldsymbol{\Gamma}\tilde{\sigma}^{-2} + \boldsymbol{D}^{-1}\right)^{-1}$$

- Each α_l must be updated individually. The zero-inflated truncated normal prior for α_l yields a conjugate posterior $p(\alpha_l|\boldsymbol{\alpha}_{-l}, \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\xi}, \boldsymbol{\phi}, \boldsymbol{y}, \boldsymbol{X}, \boldsymbol{Z}) = ZI - N^+(\hat{\alpha}, V_{\alpha l}, \hat{p}_l)$ where

$$\hat{\alpha} = \left(\frac{\sum_{i=1}^n \sum_{j=1}^{n_i} \phi_{ij} t_{ijl} T_{ij}}{\tilde{\sigma}^2}\right) V_{\alpha l}, \qquad V_{\alpha l} = \left(\sum_{i=1}^n \sum_{j=1}^{n_i} \frac{\phi_{ij} t_{ijl}^2}{\tilde{\sigma}^2} + 1\right)^{-1}$$

$$\hat{p}_l = \frac{p_{al}}{p_{al} + (1 - p_{al})\frac{N(0;0,1)}{N(0;\hat{\alpha},V_{\alpha l})} \cdot \frac{1-\Phi(0;\hat{\alpha},V_{\alpha l})}{1-\Phi(0;0,1)}}$$

  where $T_{ij} = w_{ij} - \mathbf{X^J}'_{ij}\boldsymbol{\beta_J} - \sum_{k\neq l} t_{ijk}\alpha_k$ and $N(0; m, v)$ denotes the normal density with mean m and variance v evaluated at 0 and $\Phi(0; m, v)$ is the normal cumulative distribution function with mean m and variance v evaluated at 0. The q vector

$$\boldsymbol{t}_{ij} = \left(Z_{ijl}\left(\xi_{il} + \sum_{m=1}^{l-1} \xi_{im}\gamma_{ml}\right) : l = 1, \ldots, q\right)^T$$

  is defined so that the random effects term $\boldsymbol{Z}'_{ij}\boldsymbol{A}\boldsymbol{\Gamma}\boldsymbol{\xi}_i$ can be written as $\boldsymbol{t}'_{ij}\boldsymbol{\alpha}$.

- The diagonal elements of $\boldsymbol{D}$ have inverse gamma priors $IG(\frac{1}{2}, \frac{N}{2})$; hence the posterior is given by $p(d_l|\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\xi}, \boldsymbol{\phi}, \boldsymbol{y}) = IG\left(\frac{1}{2} + \frac{n}{2}, \frac{N}{2} + \frac{\sum_{i=1}^n \xi_{il}^2}{2}\right)$

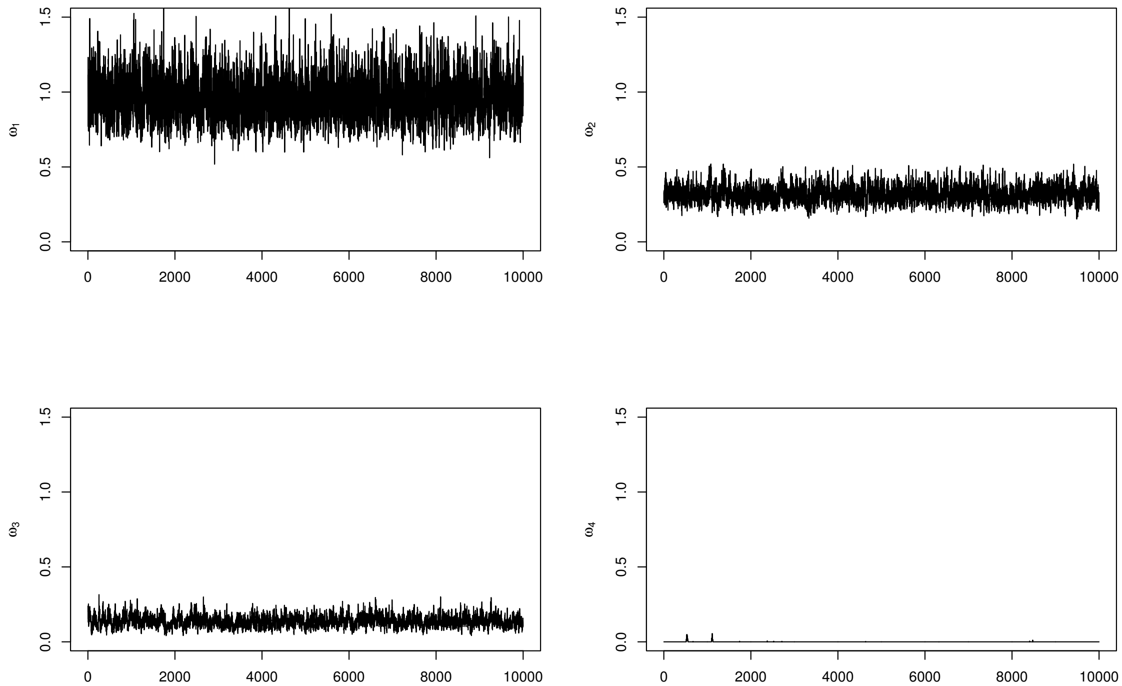

Figure 1. Gibbs chains for random effects variances

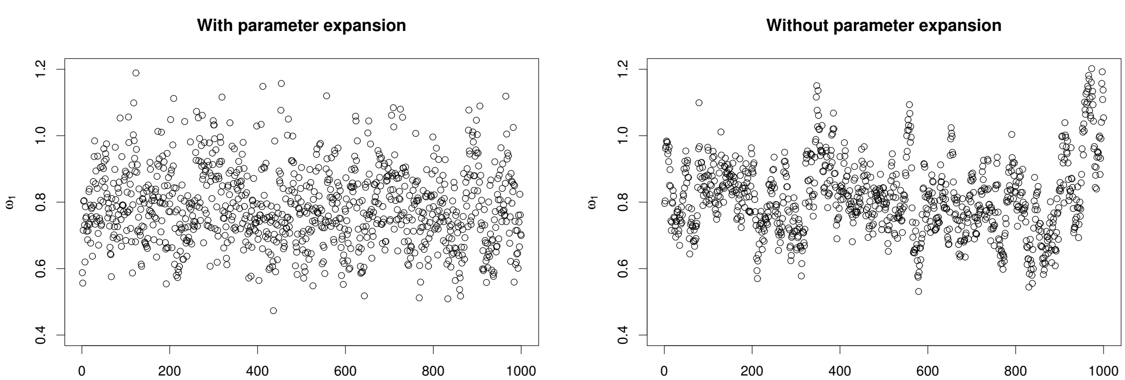

Figure 2. Illustration of parameter expansion effect on mixing of the Gibbs sampler

Table 1

Simulation results

Parameter	True value	PQL	95% CI	Post. Mean	95% CI	Pr(inclusion)
β_0	1.0	0.901	(0.753,1.048)	0.892	(0.759, 1.027)	1.000
β_1	0.0	0.031	(-0.062,0.125)	0.001	(0.000, 0.006)	0.044
β_2	1.0	0.900	(0.820,0.980)	0.929	(0.845, 1.016)	1.000
β_3	1.0	0.961	(0.896,1.025)	0.990	(0.920, 1.061)	1.000
ω_1	0.9	0.899		0.958	(0.721, 1.252)	1.000
ω_2	0.4	0.298		0.315	(0.221, 0.427)	1.000
ω_3	0.1	0.143		0.136	(0.072, 0.215)	1.000
ω_4	0.0	0.026		0.000	(0.000, 0.000)	0.008

Table 2

Autocorrelations in Gibbs chains, with and without parameter expansion

	Lag:	1	2	3	4	5	6	7	8	9	10
ω_1	w/o parameter exp	0.902	0.810	0.726	0.645	0.574	0.511	0.451	0.392	0.342	0.299
	w/parameter exp	0.422	0.350	0.288	0.252	0.208	0.177	0.154	0.142	0.132	0.117
ω_2	w/o parameter exp	0.783	0.653	0.558	0.484	0.422	0.369	0.324	0.286	0.251	0.225
	w/parameter exp	0.563	0.461	0.375	0.326	0.290	0.251	0.222	0.184	0.160	0.148
ω_3	w/o parameter exp	0.853	0.756	0.682	0.618	0.572	0.529	0.487	0.450	0.422	0.393
	w/parameter exp	0.811	0.711	0.639	0.574	0.520	0.477	0.441	0.417	0.388	0.362
ω_4	w/o parameter exp	0.808	0.629	0.439	0.335	0.228	0.162	0.087	0.038	0.008	-0.003
	w/parameter exp	0.595	0.399	0.358	0.295	0.198	-0.001	-0.001	-0.001	-0.001	-0.001

Table 3

Models with highest posterior probability

Prob	Model
.58	X_0, X_{bmi}, X_{age}
.16	X_0, X_{age}
.09	$X_0, X_{bmi}, X_{age}, X_{dde}$
.05	$X_0, X_{chol}, X_{bmi}, X_{age}$
.03	X_0, X_{age}, X_{dde}
.02	$X_0, X_{tg}, X_{bmi}, X_{age}$
.01	$X_0, X_{bmi}, X_{age}, Z_{chol}$
.01	X_0, X_{chol}, X_{age}
.01	$X_0, X_{chol}, X_{bmi}, X_{age}, X_{dde}$
.01	X_0, X_{age}, Z_{bmi}

Table 4

Posterior summary of fixed effects in CPP example

	PQL	95% CI	Mean	95% CI	$p(\beta_l = 0)$
β_0	-1.813	(-1.943, -1.700)	-1.793	(-1.871 ,-1.716)	0.000
β_{tg}	0.014	(-0.087, 0.101)	0.000	(0.000 , 0.000)	0.968
β_{chol}	-0.081	(-0.219, -0.001)	-0.002	(-0.034, 0.000)	0.932
β_{bmi}	-0.138	(-0.229, -0.055)	-0.096	(-0.210 , 0.000)	0.239
β_{age}	0.295	(0.211, 0.372)	0.279	(0.205 , 0.352)	0.000
β_{dde}	0.088	(0.009, 0.189)	0.005	(0.000 , 0.067)	0.876