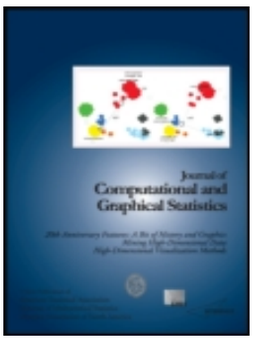

Journal of Computational and Graphical Statistics

ISSN: 1061-8600 (Print) 1537-2715 (Online) Journal homepage: http://www.tandfonline.com/loi/ucgs20

A New Algorithm for Simulating a Correlation Matrix Based on Parameter Expansion and Reparameterization

Xuefeng Liu & Michael J Daniels

To cite this article: Xuefeng Liu & Michael J Daniels (2006) A New Algorithm for Simulating a Correlation Matrix Based on Parameter Expansion and Reparameterization, Journal of Computational and Graphical Statistics, 15:4, 897-914, DOI: 10.1198/106186006X160681

To link to this article: https://doi.org/10.1198/106186006X160681

Published online: 01 Jan 2012.

A New Algorithm for Simulating a Correlation Matrix Based on Parameter Expansion and Reparameterization

Xuefeng LIU and Michael J. DANIELS

The correlation matrix (denoted by R) plays an important role in many statistical models. Unfortunately, sampling the correlation matrix in Markov chain Monte Carlo (MCMC) algorithms can be problematic. In addition to the positive definite constraint of covariance matrices, correlation matrices have diagonal elements fixed at one. In this article, we propose an efficient two-stage parameter expanded reparameterization and Metropolis-Hastings (PX-RPMH) algorithm for simulating R. Using this algorithm, we draw all elements of R simultaneously by first drawing a covariance matrix from an inverse Wishart distribution, and then translating it back to a correlation matrix through a reduction function and accepting it based on a Metropolis-Hastings acceptance probability. This algorithm is illustrated using multivariate probit (MVP) models and multivariate regression (MVR) models with a common correlation matrix across groups. Via both a simulation study and a real data example, the performance of the PX-RPMH algorithm is compared with those of other common algorithms. The results show that the PX-RPMH algorithm is more efficient than other methods for sampling a correlation matrix.

Key Words: Candidate prior; Data augmentation; PX-RPMH algorithm; Target prior.

1. INTRODUCTION

The correlation matrix (R) is directly involved in the specification of a variety of statistical models. For example, in multivariate probit models, it is common to restrict the covariance matrix to be a correlation matrix for identifiability (Chib and Greenberg 1998). In multivariate regression models, where the variances, but not correlations, are assumed to differ across some categorical covariates, common correlation models are often fit (Manly and Rayner 1987; Pourahmadi, Daniels, and Park 2006). Sampling the correlation matrix

Xuefeng Liu is Assistant Professor, Division of Translational Research and Clinical Epidemiology, School of Medicine at Wayne State University, Detroit, MI 48201 (E-mail: *xli@med.wayne.edu*). Michael J. Daniels is Associate Professor, Department of Epidemiology and Biostatistics, Department of Statistics, University of Florida, Gainesville, FL 32611 (E-mail: *mdaniels@stat.ufl.edu*).

Journal of Computational and Graphical Statistics, Volume 15, Number 4, Pages 897–914
DOI: 10.1198/106186006X160681

in these models is necessary and can be problematic due to the two constraints on a correlation matrix: positive definiteness and fixed diagonal elements. In addition, the number of unknown elements in the correlation matrix increases quadratically with the dimension T. These facts make simulating a correlation matrix difficult.

Barnard, McCulloch, and Meng (2000) suggested using the Griddy Gibbs sampler (Ritter and Tanner 1992) to draw the components of a correlation matrix one at a time in the context of a hierarchical shrinkage model for the marginal variances of a covariance matrix. Although the Griddy Gibbs sampler is simple to implement, it is not computationally efficient. Chib and Greenberg (1998) sampled the correlation matrix in the multivariate probit model using a random walk Metropolis-Hastings (RW-MH) algorithm. This algorithm is computationally more efficient than Griddy Gibbs sampler as it samples more than one component at a time. However, it has the disadvantage that the values sampled from the proposal density are not guaranteed to be positive definite. In addition, it has the problem of potentially slow mixing associated with most RW-MH algorithms. Liechty, Liechty, and Müller (2004) proposed another approach to simulate all the components of the correlation matrix one by one through introduction of a latent variable. Other approaches have been discussed by Nandram and Chen (1994), Nobile (1998), Daniels and Kass (1999), and Wong, Carter, and Kohn (2003).

Parameter expansion has been explored by several authors for simulating a correlation matrix. Liu (2001) considered simulating a correlation matrix through reparameterization in the multivariate probit regression model. However, this method is restricted to a particular prior for R. Zhang, Boscardin, and Belin (2006) proposed a parameter-extended Metropolis-Hastings algorithm (PX-MH) for sampling R in Bayesian models with correlated latent variables. The seminal idea in their method is that instead of a marginal prior for R, they specified a joint prior for R and D (unidentified marginal standard deviations) derived from an inverse Wishart distribution for $\Sigma = DRD$. Then sampling (R, D) jointly was accomplished through a Metropolis Hastings algorithm by first drawing Σ from a prespecified Wishart distribution with degrees of freedom being the sample size and the scale matrix being the current value of Σ. Using this method, all components of R are drawn at one time. However, this algorithm is similar to random walk algorithms since it samples Σ from a candidate distribution "centered" at the previous value. Though all the components are sampled at once, the random walk aspect can slow down mixing and convergence.

In this article, building on previous work, we propose a two-stage parameter expanded reparameterization and Metropolis-Hastings (PX-RPMH) algorithm for simulating a correlation matrix. In stage 1, using the idea of parameter expanded reparameterization in Liu (2001), R can be transformed into a less constrained covariance matrix, say $\Sigma = DRD$, through four steps such that the posterior distribution of Σ is an inverse Wishart distribution. Here D is the expansion parameter (more details later). In the second stage, simulating R in the original model is equivalent to first simulating Σ from the inverse Wishart distribution in the new model, and then translating it back to R through the reduction function ($R = D^{-1}\Sigma D^{-1}$) and accepting it based on a Metropolis-Hastings (M-H) acceptance probability. The major contributions of this algorithm over those previous proposed are as

follows. First, we expand on the idea in Liu (2001) to provide a general framework for deriving the parameter expanded candidate density of $\Sigma = DRD$ for correlation matrices and illustrate how these steps are related so that the algorithm can be easily extended to the more general model settings that involve sampling correlation matrices. Second, the M-H step allows the prior for R needed for the algorithm to be implemented to be different from the prior for R we use for inference, unlike in previous work (Liu 2001). Third, the family of inverse Wishart distributions used as proposal densities in the M-H step of our algorithm are derived based on the data model. The degrees of freedom and the scale matrix vary with different models. This differs from the approach in Zhang et al. (2006).

The rest of the article is organized as follows. In Section 2, we review multivariate probit (MVP) models and multivariate regression models (MVR) with a common correlation matrix across groups. In Section 3, we introduce a parameter expanded reparameterization and Metropolis-Hastings algorithm (PX-RPMH) for simulating a correlation matrix. This algorithm is then derived for MVP models and for MVR models with a common correlation matrix. Results of a simulation study are reported on in Section 4 and a real data example is analyzed in Section 5.

2. MULTIVARIATE PROBIT MODELS AND REGRESSION MODELS

In this section, we review two models in which simulating a correlation matrix is necessary: multivariate probit (MVP) models and multivariate regression (MVR) models with a common correlation matrix across groups.

2.1 MULTIVARIATE PROBIT MODELS

The MVP model was first proposed by Ashford and Sowden (1970). Later, Chib and Greenberg (1998) provided a tractable computational framework for overcoming the problems in fitting MVP models. Let $Q_i = (Q_{i1}, \ldots, Q_{iT})$ $(i = 1, \ldots, n)$ denote a T-dimensional longitudinal vector of binary responses and let $Y_i = (Y_{i1}, \ldots, Y_{iT})$ denote a longitudinal vector of latent variables with distribution $Y_i \sim N(X_i\beta, \Sigma)$, where X_i is a $T \times p$ design matrix, β is a $p \times 1$ vector of unknown regression coefficients, and Σ is the covariance matrix of Y_i. The MVP model has the following relationship between Y_i and Q_i: $Q_{ij} = \mathrm{I}(Y_{ij})$ where I is the indicator function such that $\mathrm{I}(Y) = 1$ if $Y > 0$ and $\mathrm{I}(Y) = 0$ otherwise $(i = 1, \ldots, n;\ j = 1, \ldots, T)$. Thus, the probability of $Q_i = q_i$, conditional on parameters β and Σ, is given by

$$P(Q_i = q_i|\beta, \Sigma) = \int_{A_{iT}} \cdots \int_{A_{i1}} \frac{1}{(2\pi)^{nT/2}|\Sigma|^{\frac{1}{2}}} \times \exp\left\{-\frac{1}{2}\sum_{i=1}^{n}(Y_i - X_i\beta)'\Sigma^{-1}(Y_i - X_i\beta)\right\} dY_i, \quad (2.1)$$

where A_{ij} is the interval $(0, \infty)$ if $Q_{ij} = 1$ and the interval $(-\infty, 0]$ if $Q_{ij} = 0$. As noted

by Chib and Greenberg (1998), the parameters (β, Σ) are not identifiable from the observed data model. To ensure identifiability of the model parameters, it is common to restrict the covariance matrix Σ to be a correlation matrix R.

2.2 MULTIVARIATE REGRESSION MODELS

We now consider modeling multivariate continuous responses from several groups. Let Y_{ij} be a T-dimensional response vector for the jth individual in group i ($i = 1, \ldots, c;\ j = 1, \ldots, n_i$). The following linear regression model is considered

$$Y_{ij} \sim N(X_{ij}\beta, \Sigma_i), \tag{2.2}$$

where X_{ij} is a $T \times p$ design matrix, β is a $p \times 1$ vector of regression coefficients, and Σ_i is the covariance matrix of Y_{ij}. This is a multivariate regression model with the same effect of covariates across responses.

When there are not enough data to estimate a completely separate covariance matrix Σ_i for each group in (2.2), we might consider the model with a common correlation matrix in which Σ_i can be decomposed as $V_i R V_i$, with $V_i = \text{diag}(\sigma_{i1}, \ldots, \sigma_{iT})$ and σ_{ik} the standard deviation of the kth response in group i ($i = 1, \ldots, c;\ k = 1, \ldots, T$). Thus, variances are allowed to differ across groups, but the correlation matrix is assumed constant.

3. THE PX-RPMH ALGORITHM

In this section, we propose a parameter expanded reparameterization and Metropolis-Hastings (PX-RPMH) algorithm to sample a correlation matrix. We also provide the specific details necessary to implement the algorithm for MVP and MVR models.

3.1 SUMMARY OF THE GENERAL ALGORITHM

The PX-RPMH algorithm can be viewed as a two-stage algorithm. We call the first stage parameter expanded reparameterization (PXRP) and the second stage parameter expanded Metropolis-Hastings (PXMH). The goal at the first stage is to derive a candidate density which is close to the target density and from which it is easy to simulate; at the second stage, we implement a parameter expanded Metropolis-Hastings step to ensure that the correlation matrix is sampled from the correct posterior distribution.

The PXRP stage has four steps. Step 1 is to find the expansion parameters. The scale parameters of a correlation matrix R are fixed to be one. This constraint makes drawing R difficult. This difficulty can be overcome by transforming R into a less constrained covariance matrix Σ, by borrowing the scale from some (matrix valued) expansion parameter (D). We refer to this step as "expansion parameter mining" or "XP mining." Step 2 is to find a candidate transformation $\mathbb{T}$, which maps the standard parameter space to the constrained expanded parameter space. Under this transformation, the joint posterior distribution indexed by the standard parameters can be converted into the one indexed by the expanded parameters. We refer to this step as "candidate transformations mining" or "CT mining." In the third step, we derive a *candidate prior* for R that after transformation, results in a

candidate density for Σ from which it is easy to sample. Typically, the *candidate prior* is different from the *target prior* of R which is used for posterior inference. This step is referred to as "*candidate prior* mining" or "CP mining." The final step is to derive the parameter expanded candidate density (PXCD) of Σ. This density will be used as the proposal density to sample R. It is usually an inverse Wishart distribution with degrees of freedom and scale matrices varying in different model settings. We refer to this step as parameter expanded candidate density deriving (PXCD deriving).

In the PXMH stage, we first simulate (R, D) from the PXCD of Σ and then according to a M-H acceptance probability, we decide whether to keep the candidate R.

3.2 THE PX-RPMH ALGORITHM FOR MVP MODELS

Recall the MVP models in Section 2.1. For Bayesian inference, we specify independent priors for β and R as

$$\pi(\beta) \propto \exp\{-\frac{1}{2}(\beta-\beta_0)'\Psi_\beta^{-1}(\beta-\beta_0)\} \tag{3.1}$$

$$\pi(R) \propto I\{R_{jk} : R_{jk} = 1(j=k), |R_{jk}| < 1(j \neq k) \text{ and } R \text{ is pos. def.}\} \tag{3.2}$$

where R_{kl} is the element in the kth row and lth column of R. We will discuss weakening this independence assumption at the end of Section 3.3. Note that (3.2) is the joint uniform prior defined by Barnard, McCulloch, and Meng (2000).

For this model, the PX-RPMH algorithm is derived as follows,

XP mining: The expanded parameter vector is $\theta^* = (\beta, \Sigma)$ where $\Sigma = DRD$ and the reduction function is $P(\theta^*) = (\beta, D^{-1}\Sigma D^{-1})$ with expansion parameter D satisfying $R = D^{-1}\Sigma D^{-1}$.

CT mining: Based on the reduction function, $P(\theta^*)$ from the XP mining step, we derive the following one-to-one mapping from $\{Y_i, R\}$ to $\{Y_i^*, \Sigma\}$ for creating draws of the correlation matrix,

$$\begin{cases} Y_i = X_i\beta + D^{-1}Y_i^* \\ R = D^{-1}\Sigma D^{-1} \end{cases} \qquad (i = 1, \ldots, n), \tag{3.3}$$

where Σ is a $T \times T$ positive definite matrix and $\sum_{i=1}^n Y_{it}^{*2} = 1$ for any $t = 1, \ldots, T$. Note that β in (3.3) is fixed and not involved in the transformation. Given β, the step that draws Y_i implicitly draws Y_i^* and D because $\sum_{i=1}^n (Y_{it} - x_{it}'\beta)^2 = D_{tt}^{-1}\sum_{i=1}^n Y_{it}^{*2} = D_{tt}^{-1}$, where D_{tt} is the tth element of D and x_{it}' is the tth row of X_i. The space for (Y_i^*, Σ) is higher dimensional than that for (Y_i, R) since R has fewer parameters than Σ. The constraints, $\sum_{i=1}^n Y_{it}^{*2} = 1$ for any $t = 1, \ldots, T$, are needed to make the candidate transformation a one-to-one mapping.

CP mining: Based on (2.1) and (3.3), to obtain the PXCD of Σ, we first need to find the *candidate prior* for R, given by

$$\pi(R) \propto |R|^{-\frac{T+1}{2}} I\{R_{jk} : R_{jk} = 1(j=k), |R_{jk}| < 1(j \neq k) \text{ and } R \text{ is pos. def.}\}, \tag{3.4}$$

where T is the number of rows in R. We note that (3.4) is different from (3.2).

PXCD deriving: The parameter expanded candidate density of Σ is given in the following proposition.

Proposition 1. *If we choose priors (3.1) and (3.2) for β and R, respectively, then from likelihood function (2.1), transformation (3.3) and candidate prior (3.4), we obtain*

$$\pi(\Sigma|Y^*, \beta) \propto |\Sigma|^{-\frac{\nu+T+1}{2}} \exp\left\{ -\frac{1}{2} tr(S\Sigma^{-1}) \right\}, \tag{3.5}$$

where $\nu = n$, $S = \sum_{i=1}^{n} Y_i^* Y_i^{*'}$, $Y^* = (Y_1^*, \ldots, Y_n^*)$ and $Y_i^* = D(Y_i - X_i\beta)$. That is, $\Sigma|Y^*, \beta$ has an inverse Wishart distribution with degrees of freedom ν and scale covariance matrix S.

Proof: The proof is in the Appendix (p. 909).

All the proofs of other propositions and theorems in the article are also placed in the Appendix without further mention.

Proposition 1 gives the PXCD of Σ to use as the proposal density in the PXMH stage. In this stage, we first simulate Σ from (3.5) and then obtain the correlation matrix R through the reduction function $P(\beta, \Sigma) = (\beta, D^{-1}\Sigma D^{-1})$. We keep the candidate R with probability α (the acceptance rate in the M-H algorithm). The entire PX-RPMH algorithm for MVP models is given in the following theorem.

Theorem 1. *Assume that β and R are, a priori, independent, that is, $\pi(\beta, R) = \pi(\beta)\pi(R)$. If we specify (3.1) and (3.2) as priors for β and R, respectively, then under transformation (3.3) and candidate prior (3.4), simulating R is equivalent to simulating Σ first from the inverse Wishart distribution (3.5), and then translating it back to R through $R = D^{-1}\Sigma D^{-1}$ in (3.3) and accepting the candidate R using a Metropolis-Hastings step with acceptance rate α, where $\alpha = min\left\{1, \exp\left(\frac{T+1}{2}(\log|R| - \log|R^{(k)}|)\right)\right\}$ at iteration $k+1$.*

Theorem 1 provides a simple way to simulate the correlation matrix in MVP models.

3.3 THE PX-RPMH ALGORITHM FOR MVR MODELS

Recall the MVR Models introduced in Section 2.2. We specify the priors for β and R as in (3.1) and (3.2) and the prior for V_i as

$$\pi(\text{diag}(V_i)) \propto \prod_{t=1}^{T} (\sigma_{it}^2)^{-(r_{it}+1)} \exp\left\{ -\frac{\lambda_{it}}{\sigma_{it}^2} \right\}, \quad (i = 1, \ldots, c), \tag{3.6}$$

where $\text{diag}(V_i)$ is the vector of the diagonal elements of V_i.

We derive the PX-RPMH algorithm as follows. In the PXRP stage, XP mining is similar to that in MVP models. In CT mining, we consider the following one-to-one mapping $\mathbb{T}_1 : \{Y_{ij}, R\} \to \{\epsilon_{ij}, \Sigma\}$

$$\begin{cases} Y_{ij} = X_{ij}\beta + D^{-1}\epsilon_{ij} \\ R = D^{-1}\Sigma D^{-1} \end{cases} \qquad (i = 1, \ldots, c;\ j = 1, \ldots, n_i), \tag{3.7}$$

where β is fixed and not involved in $\mathbb{T}_1$, and ϵ_{ij} is a $T \times 1$ vector satisfying $\sum_{i=1}^{c} \sum_{j=1}^{n_i} \epsilon_{ijt}^2 = 1$ $(t = 1, \ldots, T)$. Note that given β, the step that draws Y_{ij} implicitly draws ϵ_{ij} and

D because $\sum_{i=1}^{c}\sum_{j=1}^{n_i}(Y_{ijt} - x'_{ijt}\beta)^2 = D_{tt}^{-2}\sum_{i=1}^{c}\sum_{j=1}^{n_i}(\epsilon_{ijt}^2) = D_{tt}^{-2}$ for any $t = 1, \ldots, T$, where D_{tt} is the tth diagonal element of D and x_{ijt} is the vector of covariates associated with the tth response of subject j in group i. Based on (2.2) and (3.7), the *candidate prior* for R is the same as (3.4) given in MVP models. The following proposition gives the PXCD of Σ.

Proposition 2. *If we specify (3.1), (3.2), and (3.6) as priors for β, R and V_i, respectively, then under $\mathbb{T}_1$ (3.7) and candidate prior (3.4), the posterior distribution of Σ given ϵ_{ij} and V_i $(i = 1, \ldots, c;\ j = 1, \ldots, n_i)$ is an inverse Wishart with degrees of freedom ν_1 and scale covariance matrix S_1, that is,*

$$\pi(\Sigma|\epsilon, \beta, V_1, \ldots, V_c) \propto |\Sigma|^{-\frac{\nu_1+T+1}{2}} \exp\left\{-\frac{1}{2}tr(S_1\Sigma^{-1})\right\}, \tag{3.8}$$

where $\nu_1 = \sum_{i=1}^{c} n_i$, $S_1 = \sum_{i=1}^{c}\sum_{j=1}^{n_i} V_i^{-1}\epsilon_{ij}\epsilon'_{ij}V_i^{-1}$, $\epsilon = (\epsilon'_{11}, \ldots, \epsilon'_{cn_c})'$ and $\epsilon_{ij} = D(Y_{ij} - X_{ij}\beta)$.

For the PXMH stage, the M-H step will be implemented as in MVP models. The PX-RPMH algorithm for MVR models is given in the following theorem

Theorem 2. *Assume that β, R, and V_i are, a priori, independent, that is, $\pi(\beta, R, V_i) = \pi(\beta)\pi(R)\pi(diag(V_i))$. If we specify (3.1), (3.2), and (3.6) as priors for β, R and V_i, respectively, then under transformation (3.7) and candidate prior (3.4), simulating R is equivalent to simulating Σ first from the inverse Wishart distribution (3.8), and then translating it back to R through $R = D^{-1}\Sigma D^{-1}$ in (3.7) and accepting the candidate R through a Metropolis-Hastings step with acceptance rate α, where $\alpha = min\left\{1, \exp\left(\frac{T+1}{2}(\log|R| - \log|R^{(k)}|)\right)\right\}$ at iteration $k+1$.*

We conclude this section with several remarks.

Remark 1. We have described the MVP model in the context of longitudinal binary data. However, this model can also be applied to multiple (non-longitudinal) binary responses from the same subject. We also point out that in the longitudinal setting, the model is only appropriate when the responses are observed at the same set of time points for all subjects (i.e., temporally aligned longitudinal data).

Remark 2. Extending the idea in Liu (2001), we provide a general framework for deriving the PXCD of Σ through the four steps in the PXRP stage. We used the phrase "mining" in the first three steps, a term borrowed from the "data mining" and "statistical methods mining" literature (Berry and Linoff 2000).

Remark 3. The four steps in the PXRP stage are highly dependent. The first three steps in conjunction with the likelihood function and the *candidate/target* priors for the parameters (besides R) in the model determine the fourth step, PXCD deriving In second step, CT mining, the candidate transformation varies with the likelihood and the priors for $\theta_{(-R)}$, where $\theta_{(-R)}$ is the parameter vector excluding R.

Remark 4. In the algorithm, we discussed two different terms: the *target prior* and the *candidate prior*. They play different roles in our algorithm. The *target prior* is the prior we specify for posterior inference. The *candidate prior* is used to derive the PXCD which will be used as the proposal density in the PXMH stage of the algorithm. We also point out

we specified the *target prior* and the *candidate prior* for all the parameters except R to be the same. However, this is not necessary.

Remark 5. The parameters were a priori independent in MVP models and MVR models. This algorithm can also be applied to cases where these priors are not independent (Liu 2001; Liu and Sun 2000).

Remark 6. PXCD deriving is the bridge between the PXRP stage and the PXMH stage. Under some mild conditions, the PX-RPMH algorithm can be reduced to a simpler algorithm. If the *candidate prior* and the *target prior* for R are same, the acceptance probability α in the second stage of the PX-RPMH is equal to 1. We call this simpler algorithm parameter expanded reparameterization (PX-RP) algorithm.

Remark 7. In the PXMH stage, we draw the expansion parameter D and correlation matrix R simultaneously. The draws of latent data and other parameters in the model can be then adjusted based on the current draw of D using the transformation in CT mining. Thus, as mentioned in Liu (2001), by redrawing the scale parameter D, the MCMC sampling scheme contains a way to adjust the current draws of the latent data and the other parameters, and therefore has a faster convergence rate than the standard Gibbs sampler.

Remark 8. Since the likelihood function does not change and the *target prior*(s) and the *candidate prior*(s) for all the parameters except R are the same in our specification of MVP and MVR models, the acceptance probability in the PXMH stage depends only on the *target prior* (3.2) and the *candidate prior* (3.4) for R.

Remark 9. We can also use a more general *candidate prior* for R

$$\pi(R) \propto |R|^{-\frac{a+1}{2}} \exp\left\{-\frac{1}{2}\mathrm{tr}(WR^{-1})\right\} I\{R_{jk} : R_{jk} = 1(j = k), |R_{jk}| < 1(j \neq k) \text{ and } R \text{ is pos. def.}\}, \quad (3.9)$$

where a is a constant and W is a $T \times T$ positive semi-definite tuning parameter matrix. (3.9) is a more general version of (3.4). When $W = 0$, it reduces to the candidate prior used here, $\pi(R) \propto |R|^{-\frac{a+1}{2}}$. Note that when $a \leq -1$, the reduced prior is proper, but when $a > -1$, it is improper. The advantage of the more general specification in (3.9) is that we can tune W if necessary to increase the acceptance rate in the M-H step. This will be helpful if the *target prior* for R is informative (e.g., Zhang et al. 2006).

4. A SIMULATION STUDY

The Griddy-Gibbs sampler (Barnard et al. 2000) and the random walk Metropolis-Hastings (RW-MH) algorithm (Chib and Greenberg 1998) are two common methods for sampling from the distribution of a correlation matrix. In the following, we compare the performance of the PX-RPMH algorithm with these two algorithms.

We simulated data under MVP models with $n = 200$ subjects at $T = 7$ time points. The six-dimensional columns of the design matrix were set as: $X_{it} = (1, t, t^2, b_i, b_i * t, b_i * t^2)$, where b_i is a binary covariate and t is the time point ($i = 1, \ldots, n$; $t = 1, \ldots, T$). Without

loss of generality, we set β to be a vector of 1's and R to be

$$R = \begin{pmatrix} 1 & 0.8 & 0.6 & 0.4 & 0.2 & 0 & 0 \\ 0.8 & 1 & 0.8 & 0.6 & 0.4 & 0.2 & 0 \\ 0.6 & 0.8 & 1 & 0.8 & 0.6 & 0.4 & 0.2 \\ 0.4 & 0.6 & 0.8 & 1 & 0.8 & 0.6 & 0.4 \\ 0.2 & 0.4 & 0.6 & 0.8 & 1 & 0.8 & 0.6 \\ 0 & 0.2 & 0.4 & 0.6 & 0.8 & 1 & 0.8 \\ 0 & 0 & 0.2 & 0.4 & 0.6 & 0.8 & 1 \end{pmatrix}.$$

The *target priors* for β and R were specified as

$$\pi(\beta) \propto 1$$

and

$$\pi(R) \propto I\{R_{jk} : R_{jk} = 1(j = k), |R_{jk}| < 1(j \neq k) \text{ and } R \text{ is pos. def.}\}.$$

Using results from Chen and Shao (1999), it can be shown that the joint posterior distribution for this model is proper.

The observed data, Q_i, were specified as $Q_{ij} = 1$ if $Y_{ij} > 0$ and 0 otherwise. Define $Y = (Y_1, \ldots, Y_{200})$ and $Q = (Q_1, \ldots, Q_{200})$. We used a combination of Gibbs sampling (Gelfand and Smith 1990) and data augmentation (Tanner and Wong 1987) to sample from the posterior distribution. Details on the full conditional distributions are given next. $Y|Q, \beta, R$ follows a truncated multivariate normal distribution with density

$$f(Y|Q, \beta, R) \propto \frac{\exp\left\{-\frac{1}{2}\sum_{i=1}^{n}(Y_i - X_i\beta)'R^{-1}(Y_i - X_i\beta)\right\}}{|R|^{\frac{n}{2}}\int_C \exp\left\{-\frac{1}{2}\sum_{i=1}^{n}(Y_i - X_i\beta)'R^{-1}(Y_i - X_i\beta)\right\}dY}, \tag{4.1}$$

where C is the truncation region defined as $C = \{Y : Y_{ij} \geq 0 \text{ if } Q_{ij} = 1 \text{ and } Y_{ij} < 0 \text{ if } Q_{ij} = 0\}$. $\beta|Y, R$ has a normal distribution,

$$\pi(\beta|Y, R) \propto \frac{1}{|\Sigma_\beta|^{\frac{1}{2}}} \exp\left\{-\frac{1}{2}(\beta - \mu_\beta)'\Sigma_\beta^{-1}(\beta - \mu_\beta)\right\}, \tag{4.2}$$

where $\mu_\beta = \left(\sum_{i=1}^{n} X_i'R^{-1}X_i\right)^{-1}\left(\sum_{i=1}^{n} X_i'R^{-1}Y_i\right)$ and $\Sigma_\beta = \left(\sum_{i=1}^{n} X_i'R^{-1}X_i\right)^{-1}$. The full conditional distribution for R is

$$\pi(R|Y, \beta) \propto |R|^{-\frac{n}{2}} \exp\left\{-\frac{1}{2}\text{tr}\left[\sum_{i=1}^{n}(Y_i - X_i\beta)(Y_i - X_i\beta)'R^{-1}\right]\right\}$$
$$\times I\{R_{jk} : R_{jk} = 1(j = k), |R_{jk}| < 1(j \neq k) \text{ and } R \text{ is pos. def.}\}. \tag{4.3}$$

We sampled from the full conditional distribution of R using the PX-RPMH algorithm, the Griddy-Gibbs sampling algorithm and the RW-MH algorithm. For the RW-MH algorithm, we chose as the proposal density a normal distribution with mean $r^{(k)}$ and covariance τA, where $r^{(k)}$ is the vector of the components of $R^{(k)}$ from the current iteration k defined as $r^{(k)} = (R_{12}^{(k)}, \ldots, R_{(T-1)T}^{(k)})$, A is the inverse of the negative of the Hessian matrix of

$\log L(R|Y, \hat{\beta})$, Y are the true values from the simulated data (unknown in a real dataset), $\hat{\beta}$ is the MLE of β, and τ is a tuning constant. The RW-MH algorithm, at iteration $k+1$, is given by

Step 1. Generate a proposal value r from the normal density $q(r|r^{(k)}, Y^{(k)}, \beta^{(k)})$.

Step 2. Set

$$r^{(k+1)} = \begin{cases} r & \text{with probability } \alpha \\ r^{(k)} & \text{with probability} 1-\alpha, \end{cases}$$

where $\alpha = \min\left\{\frac{f(Y^{(k)}|\beta^{(k)},R)}{f(Y^{(k)}|\beta^{(k)},R^{(k)})}, 1\right\}$.

Since R has 21 elements, to improve mixing, the Metropolis-Hastings step was applied to $r = (r_1, r_2, r_3, r_4)$ in four blocks, where r_1, r_2, and r_3 each consist of six components and r_4, of three components in a row-wise expansion of R. We chose $\tau = 1.5$ for the first three blocks and $\tau = 2.0$ for the fourth block such that the acceptance rate (a percentage of times a move to a new point is made) was equal to 0.21 (Roberts, Gelman, and Gilks 1997; Roberts and Rosenthal 2001).

To compare the running time of the PX-RPMH algorithm, the Griddy-Gibbs sampler and the RW-MH algorithm, we generated 5,000 samples from (4.3). For the Griddy-Gibbs sampler, for simplicity, we discretized the domain of $\pi(r_{ij}|r_{(-ij)}, Y, \beta)$ uniformly using the $m = 20$ points $r_1, \ldots, r_m$. The sampling required 991 seconds for the Griddy-Gibbs sampler, 11 seconds for the RW-MH algorithm, and 6 seconds for the PX-RPMH algorithm. Thus, the PX-RPMH was more than 150 times faster than the Griddy-Gibbs sampler.

We also examined the mixing of the PX-RPMH algorithm and the random walk Metropolis-Hastings algorithm graphically. We used the DA algorithm to iteratively generate Y, β, and R from (4.1), (4.2), and (4.3), respectively. A chain of 50,000 iterations was generated and only the last 2,000 samples were retained. We randomly chose five components of R: R_{15}, R_{27}, R_{35}, R_{47}, and R_{56} and examined the trace plots (see Figure 1). The first column is from the PX-RPMH algorithm and the second is from the RW-MH algorithm. It is clear that the PX-RPMH algorithm is mixing better than the RW-MH algorithm (and at the cost of no additional computational complexity). In the next section, we apply the algorithm to a longitudinal dataset.

5. A REAL DATA EXAMPLE

We illustrate the performance of the algorithm using data from a longitudinal clinical trial of growth hormone for maintaining muscle strength in the elderly. Details of the trial can be found in Kiel et al. (1998) and Daniels and Hogan (2000). One hundred sixty subjects were recruited and randomized to one of four treatments: (1) Placebo; (2) Growth hormone only; (3) Exercise plus placebo; (4) Exercise plus growth hormone. The placebo and growth hormone treatments were administered daily via injections. Various muscle strength measures were recorded at baseline (0 months), 6 months and 12 months. We focus on the outcome mean quadriceps strength. The dropout rates in the four treatment

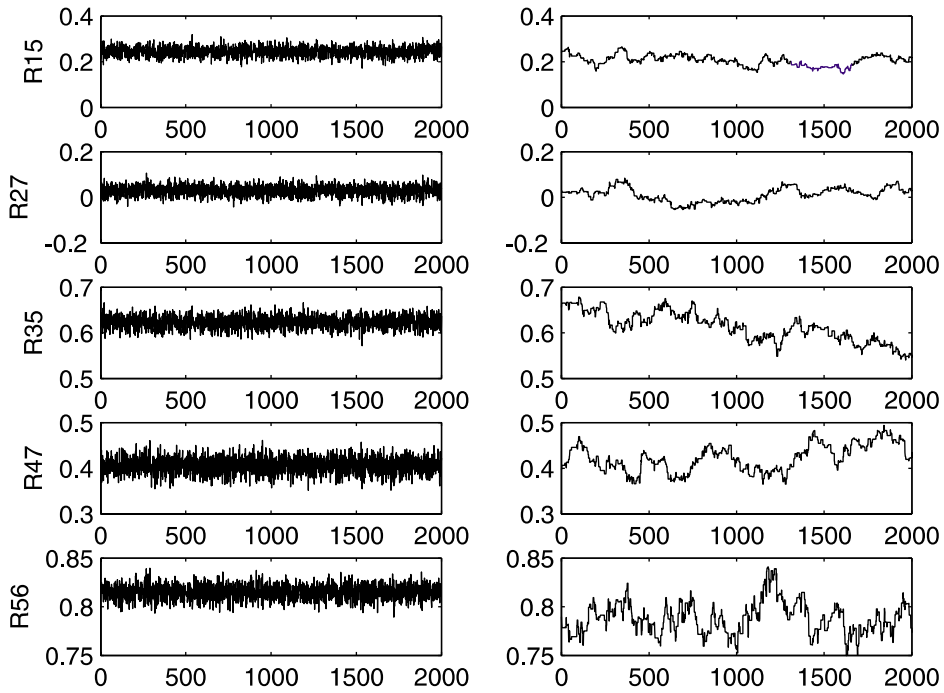

Figure 1. Trace plots of five components of R over the last 2,000 iterations of 50,000 iterations for the PX-RPMH algorithm and RW-MH algorithm in the simulation study.

groups were 11/41, 13/41, 9/40 and 16/38, respectively. We will conduct our analysis under an assumption of random dropout for illustration of our method (Rubin 1976).

Let $Y_{ij} = (Y_{ij1}, Y_{ij2}, Y_{ij3})$ denote the vector of longitudinal responses for subject j in treatment group i ($i = 1, \ldots, 4$; $j = 1, \ldots, n_i$). X_{ij} is specified such that each component of β corresponds to a time/treatment specific mean. We use the model described in Section 2.2, $Y_{ij} \sim N(X_{ij}\beta, \Sigma_i)$, with $\Sigma_i = V_i R V_i$. We specify priors for β, V_i and R as follows,

$$\pi(\beta) \propto \exp\{-\frac{1}{2}(\beta - \beta_0)'\Psi_\beta^{-1}(\beta - \beta_0)\}, \tag{5.1}$$

$$\pi(\text{diag}(V_i)) \propto \prod_{t=1}^{T}(\sigma_{it}^2)^{-(r_{it}+1)} \exp\left\{-\frac{\lambda_{it}}{\sigma_{it}^2}\right\} \quad (i = 1, \ldots, 4), \tag{5.2}$$

$$\pi(R) \propto I\{R_{jk} : R_{jk} = 1(j = k), |R_{jk}| < 1(j \neq k) \text{ and } R \text{ is pos. def.}\}. \tag{5.3}$$

In the following, Y_{mis} denotes the vector of missing responses. Details on sampling $(Y_{\text{mis}}, \beta, V_i)$ can be found in the Appendix. The full conditional distribution of R is given by

$$\pi(R|Y, \beta) \propto |R|^{-\frac{n_i}{2}} \exp\left\{-\frac{1}{2}\text{tr}\left[\sum_{i=1}^{4}\sum_{j=1}^{n_i} V_i \left(Y_{ij} - X_{ij}\beta\right)\left(Y_{ij} - X_{ij}\beta\right)' V_i R^{-1}\right]\right\}$$

$$\times I\{R_{jk} : R_{jk} = 1(j = k), |R_{jk}| < 1(j \neq k) \text{ and } R \text{ is pos. def.}\}. \tag{5.4}$$

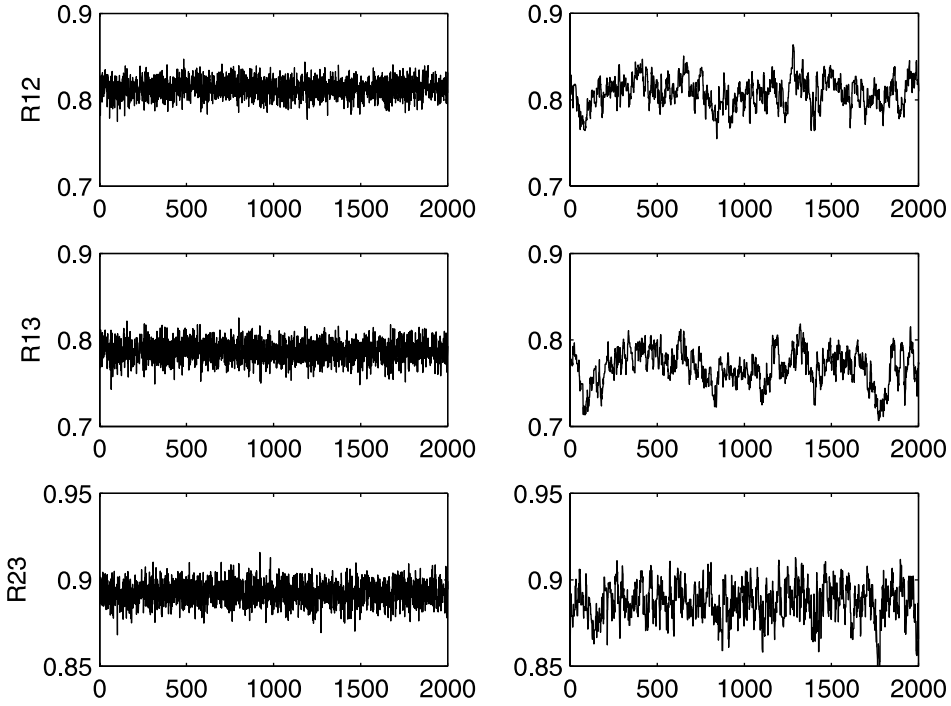

Figure 2. Trace plots of the three components of R over the last 2,000 iterations of 10,000 iterations for the PX-RPMH algorithm and RW-MH algorithm in the real data example in Section 5.

To sample R, we use the PX-RPMH algorithm as described in Section 3.3. For comparison, we also sample R using the RW-MH algorithm as described in Section 4.

We ran a chain of 10,000 iterations using the two different methods to sample R and examined trace plots of the last 2,000 samples (see Figure 2). It is clear that the PX-RPMH algorithm is mixing better than the RW-MH algorithm as was the case in the simulations in Section 4. The ease with which this model can be fit using the PX-RPMH algorithm should encourage its consideration. For completeness, Table 1 contains the posterior mean and 95% credible intervals for the elements of the common correlation matrix (R), the group-specific variances (V_i), and the mean parameters based on the chain run using the PX-RPMH algorithm.

6. DISCUSSION

We have proposed a PX-RPMH algorithm for sampling a correlation matrix, R. This approach places no restrictions on the choice of the prior for R unlike previous methods and is easy to implement. Our comparisons show that the PX-RPMH algorithm is superior to the RW-MH algorithm and the Griddy-Gibbs sampler in terms of computational time and mixing.

The PX-RPMH algorithm can also be applied to the other cases in which sampling a correlation matrix is necessary, such as shrinkage models for normal regression (Barnard,

Table 1. Posterior Means and 95% Credible Intervals for the Correlations, Variances, and Regression Parameters for the Growth Hormone Data.

Groups (i)	*R*	*diag(V_i)*	β
	0.81(0.75,0.86)	634(462,855)	68.9(60.8,76.7)
1	0.79(0.72,0.85)	937(662,1355)	80.5(69.9,90.6)
	0.89(0.83,0.91)	913(644,1273)	79.1(68.3,89.6)
		514(384,742)	68.4(61.4,75.2)
2	–	589(448,822)	66.0(58.6,73.7)
		492(357,672)	65.1(58.0,72.0)
		606 (465,775)	65.6(58.1,73.0)
3	–	640(507,818)	81.0(73.0,88.7)
		465(350,601)	72.4(65.6,79.2)
		689(517,932)	65.2(57.4,73.2)
4	-	628 (477,868)	62.1(54.5,69.8)
		554 (407,782)	62.6(55.0,70.4)

McCulloch, and Meng 2001), scale mixtures of multivariate normal link models (Chen and Dey 1998, 2000) and joint models for longitudinal binary and continuous responses (Gueorguieva and Agresti 2001); for the final case, the PX-RPMH algorithm has been implemented (Liu 2006). The same family of inverse Wishart distributions should be able to be used as the parameter expanded candidate density (PXCD) in the PXRP stage for these models, but the degrees of freedom and the scale matrix of the PXCD will vary and more expansion parameters, in addition to the scale parameters D, may be needed to implement the algorithm for certain models. This will make the PXRP stage more complex. As the associate editor and one referee pointed out, care must be taken when this algorithm is applied to models where the dimensions of the response vector and the correlation matrix are not the same. In these cases, deriving the parameter expanded candidate density can be difficult. We will address this issue in future work.

The expansion parameters play an important role in our algorithm. If the same expansion parameters also make the DA algorithm converge faster, then we may combine the DA algorithm and the PX-RPMH algorithm and make the MCMC algorithm more efficient. In general model settings, whether the expansion parameters used in our algorithm can also accelerate the DA algorithm needs further study.

APPENDIX

Proof of Proposition 1: Extending the idea in Liu (2001), based on transformation (3.3), we derive the Jacobian. The following five determinants are needed

$$
\begin{aligned}
\left|\frac{\partial(Y_1,\ldots,Y_n)'}{\partial(\sigma_{21},\ldots,\sigma_{T,T-1})}\right| &= 0\\
\left|\frac{\partial(r_{21},\ldots,r_{T,T-1})'}{\partial(\sigma_{21},\ldots,\sigma_{T,T-1})}\right| &= \prod_{j=1}^{T}(\sigma_{jj}^{-1})^{T-1} = |D^{-1}|^{T-1}\\
\left|\frac{\partial(\sigma_{11}^{-1},\ldots,\sigma_{T,T}^{-1})'}{\partial(\sigma_{11}^{2},\ldots,\sigma_{T,T}^{2})}\right| &= \left(-\frac{1}{2}\right)^{T}\prod_{j=1}^{T}(\sigma_{jj}^{-2})^{\frac{3}{2}} \propto |D^{-1}|^{3}\\
\left|\frac{\partial(Y_1,\ldots,Y_n)'}{\partial(Y_1^*,\ldots,Y_{n-1}^*)}\right| &= \left|\left(\mathrm{I}_{n-1}\bigotimes D^{-1}\right)P_1\right| \propto |D^{-1}|^{n-1}\\
\left|\frac{\partial(Y_1,\ldots,Y_n)'}{\partial(\sigma_{11}^{-1},\ldots,\sigma_{T,T}^{-1})}\right| &= |P_2|,
\end{aligned}
$$

where P_1 and P_2 are matrices, not depending on D. Then the Jacobian is

$$
\begin{aligned}
J &= \left|\frac{\partial(Y_1,\ldots,Y_n,R)}{\partial(Y_1^*,\ldots,Y_{n-1}^*,\Sigma)}\right|_+ = \left|\frac{\partial(Y_1,\ldots,Y_n,r_{21},\ldots,r_{T,T-1})'}{\partial(Y_1^*,\ldots,Y_{n-1}^*,\sigma_{11}^2,\ldots,\sigma_{T,T}^2,\sigma_{21},\ldots,\sigma_{T,T-1})}\right|_+\\
&= \left|\begin{array}{cc}
\frac{\partial(Y_1,\ldots,Y_n)'}{\partial(Y_1^*,\ldots,Y_{n-1}^*,\sigma_{11}^2,\ldots,\sigma_{T,T}^2)} & \frac{\partial(r_{21},\ldots,r_{T,T-1})'}{\partial(Y_1^*,\ldots,Y_{n-1}^*,\sigma_{11}^2,\ldots,\sigma_{T,T}^2)}\\
\frac{\partial(Y_1,\ldots,Y_n)'}{\partial(\sigma_{21},\ldots,\sigma_{T,T-1})} & \frac{\partial(r_{21},\ldots,r_{T,T-1})'}{\partial(\sigma_{21},\ldots,\sigma_{T,T-1})}
\end{array}\right|_+\\
&= \left|\frac{\partial(Y_1,\ldots,Y_n)'}{\partial(Y_1^*,\ldots,Y_{n-1}^*,\sigma_{11}^2,\ldots,\sigma_{T,T}^2)}\right|_+\left|\frac{\partial(r_{21},\ldots,r_{T,T-1})'}{\partial(\sigma_{21},\ldots,\sigma_{T,T-1})}\right|_+\\
&= \left|\frac{\partial(\sigma_{11}^{-1},\ldots,\sigma_{T,T}^{-1})'}{\partial(\sigma_{11}^2,\ldots,\sigma_{T,T}^2)}\right|_+\left|\frac{\partial(Y_1,\ldots,Y_n)'}{\partial(Y_1^*,\ldots,Y_{n-1}^*,\sigma_{11}^{-1},\ldots,\sigma_{T,T}^{-1})}\right|_+\left|\frac{\partial(r_{21},\ldots,r_{T,T-1})'}{\partial(\sigma_{21},\ldots,\sigma_{T,T-1})}\right|_+\\
&= \left|\frac{\partial(\sigma_{11}^{-1},\ldots,\sigma_{T,T}^{-1})'}{\partial(\sigma_{11}^2,\ldots,\sigma_{T,T}^2)}\right|_+\left|\begin{array}{c}
\frac{\partial(Y_1,\ldots,Y_n)'}{\partial(Y_1^*,\ldots,Y_{n-1}^*)}\\
\frac{\partial(Y_1,\ldots,Y_n)'}{\partial(\sigma_{11}^{-1},\ldots,\sigma_{T,T}^{-1})}
\end{array}\right|_+\left|\frac{\partial(r_{21},\ldots,r_{T,T-1})'}{\partial(\sigma_{21},\ldots,\sigma_{T,T-1})}\right|_+\\
&\propto |D^{-1}|^3\left|\begin{array}{c}
(\mathrm{I}_{n-1}\bigotimes D^{-1})P_1\\
P_2
\end{array}\right||D^{-1}|^{T-1}\\
&= |D^{-1}|^{T+2}\left|\begin{pmatrix}\mathrm{I}_{n-1}\bigotimes D^{-1} & 0\\ 0 & \mathrm{I}_3\end{pmatrix}\begin{pmatrix}P_1\\P_2\end{pmatrix}\right|_+\\
&= |D^{-1}|^{T+2}|D^{-1}|^{n-1} = |D^{-1}|^{n+T+1},
\end{aligned}
$$

where n is the sample size, T is the number of rows in R, and $|\ |_+$ is the absolute value of the corresponding determinant. Using prior (3.4), the joint distribution of Y and R given β is

$$
p(Y,R|\beta) \propto |R|^{-\frac{n+T+1}{2}}\exp\left\{-\frac{1}{2}\sum_{i=1}^{n}(Y_i - X_i\beta)'R^{-1}(Y_i - X_i\beta)\right\}. \tag{A.1}
$$

Through the one-to-one mapping $\mathbb{T}: (Y, R) \to (Y^*, \Sigma)$, we have

$$\begin{aligned} p(Y^*, \Sigma|\beta) &\propto |R|^{-\frac{n+T+1}{2}} \times J \times \exp\left\{-\frac{1}{2}\sum_{i=1}^{n} Y_i^{*'}\Sigma^{-1}Y_i^*\right\} \\ &= |R|^{-\frac{n+T+1}{2}}|D^{-1}|^{n+T+1}\exp\left\{-\frac{1}{2}\sum_{i=1}^{n} Y_i^{*'}\Sigma^{-1}Y_i^*\right\} \\ &= |\Sigma|^{-\frac{n+T+1}{2}}\exp\left\{-\frac{1}{2}\sum_{i=1}^{n} Y_i^{*'}\Sigma^{-1}Y_i^*\right\}. \end{aligned} \tag{A.2}$$

So,

$$\pi(\Sigma|Y^*, \beta) \propto p(Y^*, \Sigma|\beta) \propto |\Sigma|^{-\frac{\nu+T+1}{2}}\exp\left\{-\frac{1}{2}S\Sigma^{-1}\right\}, \tag{A.3}$$

where $\nu = n$, $S = \sum_{i=1}^{n} Y_i^* Y_i^{*'}$, $Y_i^* = D(Y_i - X_i\beta)$, and $Y^* = (Y_1^*, \ldots, Y_n^*)$.

Proof of Theorem 1: By the candidate transformation $\mathbb{T}$ given in (3.3), and (A.1) and (A.2) in the proof of Proposition 1, drawing Y and R given β is equivalent to drawing Y^* and Σ first, and then translating back to Y and R through $\mathbb{T}$. $R = D^{-1}\Sigma D^{-1}$ is the candidate value. It is accepted in the Metropolis-Hastings step with probability α, derived as follows. Let π_1 denote the prior or full conditional distribution of R and π_2 denote the corresponding *candidate prior* or proposal density. Then

$$\pi_1(R|Y, \beta) \propto \pi_1(R) f(Y|\beta, R),$$

and

$$\pi_2(R|Y, \beta) \propto \pi_2(R) f(Y|\beta, R),$$

where $\pi_1(R)$, $\pi_2(R)$, and $f(Y|\beta, R)$ are given by (3.2), (3.4), and (2.1), respectively. The probability of acceptance at iteration $k+1$ is

$$\begin{aligned} \alpha &= \min\left\{1, \frac{\pi_1(R|Y, \beta)\pi_2(R^{(k)}|Y, \beta)}{\pi_1(R^{(k)}|Y, \beta)\pi_2(R|Y, \beta)}\right\} = \min\left\{1, \frac{\pi_1(R)\pi_2(R^{(k)})}{\pi_1(R^{(k)})\pi_2(R)}\right\} \\ &= \min\left\{1, \frac{\pi_2(R^{(k)})}{\pi_2(R)}\right\} = \min\left\{1, \frac{|R^{(k)}|^{-\frac{T+1}{2}}}{|R|^{-\frac{T+1}{2}}}\right\} \\ &= \min\left\{1, \exp\left(\frac{T+1}{2}(\log|R| - \log|R^{(k)}|)\right)\right\}. \end{aligned} \tag{A.4}$$

Proof of Proposition 2: Similar to the proof of Proposition 1, the Jacobian can be shown to be $J = |D^{-1}|^{\sum_{i=1}^{c} n_i + T + 1}$, where n_i is the number of individuals in group i and T is the dimension of R. Using prior (3.4), the joint distribution of Y and R given β and V_i is

$$\begin{aligned} p(Y, R|\beta, V_1, \ldots, V_c) &\propto |R|^{-\frac{\sum_{i=1}^{c} n_i + T + 1}{2}} \\ &\times \exp\left\{-\frac{1}{2}\sum_{i=1}^{c}\sum_{j=1}^{n_i}\left(Y_{ij} - X_{ij}\beta\right)' V_i^{-1} R^{-1} V_i^{-1}\left(Y_{ij} - X_{ij}\beta\right)\right\}. \end{aligned} \tag{A.5}$$

Define ϵ to be $\epsilon = (\epsilon'_{11}, \ldots, \epsilon'_{cn_c})'$, where $\epsilon_{ij} = D(Y_{ij} - X_{ij}\beta)$. Through the transformation (3.7), we have

$$\begin{aligned} & p(\epsilon, \Sigma|\beta, V_1, \ldots, V_c) \\ \propto\ & |R|^{-\frac{\sum_{i=1}^c n_i+T+1}{2}} \times J \times \exp\left\{-\frac{1}{2}\sum_{i=1}^{c}\sum_{j=1}^{n_i}\epsilon'_{ij}V_i^{-1}\Sigma^{-1}V_i^{-1}\epsilon_{ij}\right\} \\ =\ & |R|^{-\frac{\sum_{i=1}^c n_i+T+1}{2}}|D^{-1}|^{\sum_{i=1}^c n_i+T+1}\exp\left\{-\frac{1}{2}\sum_{i=1}^{c}\sum_{j=1}^{n_i}\epsilon'_{ij}V_i^{-1}\Sigma^{-1}V_i^{-1}\epsilon_{ij}\right\} \\ =\ & |\Sigma|^{-\frac{\sum_{i=1}^c n_i+T+1}{2}}\exp\left\{-\frac{1}{2}\sum_{i=1}^{c}\sum_{j=1}^{n_i}\epsilon'_{ij}V_i^{-1}\Sigma^{-1}V_i^{-1}\epsilon_{ij}\right\}. \end{aligned} \tag{A.6}$$

Hence,

$$\pi(\Sigma|\epsilon, \beta, V_1, \ldots, V_c) \propto p(\epsilon, \Sigma|\beta, V_1, \ldots, V_c) \propto |\Sigma|^{-\frac{\nu_1+T+1}{2}}\exp\left\{-\frac{1}{2}S_1\Sigma^{-1}\right\}, \tag{A.7}$$

where $\nu_1 = \sum_{i=1}^c n_i$ and $S_1 = \sum_{i=1}^c \sum_{j=1}^{n_i} V_i^{-1}\epsilon_{ij}\epsilon'_{ij}V_i^{-1}$.

Proof of Theorem 2: By the candidate transformation $\mathbb{T}_1$ given in (3.7), and (A.5) and (A.6) in the proof of Proposition 2, drawing Y and R given β and V_i $(i = 1, \ldots, c)$ is equivalent to drawing ϵ and Σ first, and then translating back to Y and R through $\mathbb{T}_1$. $R = D^{-1}\Sigma D^{-1}$ is used as the candidate value. It is accepted in the Metropolis-Hastings step with probability α, which can be derived similarly to the acceptance probability in Theorem 1.

Sampling algorithm details for data example in Section 5. We provide some details on the data augmented Gibbs sampling algorithm for $(\beta, V_i, Y_{\text{mis}})$. We start with Y_{mis}. If we partition Y_{ij}, μ_{ij}, and Σ_i as

$$Y_{ij} = \begin{pmatrix} Y_{ij,\text{mis}} \\ Y_{ij,\text{obs}} \end{pmatrix},$$

$$\mu_{ij} = \begin{pmatrix} \mu_{ij,\text{mis}} \\ \mu_{ij,\text{obs}} \end{pmatrix},$$

and

$$\Sigma_i = \begin{pmatrix} \Sigma_{i11} & \Sigma_{i12} \\ \Sigma_{i21} & \Sigma_{i22} \end{pmatrix},$$

respectively, then the data augmentation step is implemented by sampling Y_{mis} from

$$\begin{aligned} & f(Y_{\text{mis}}|Y_{\text{obs}}, \beta, V_i, R) \\ & \propto \exp\left\{-\frac{1}{2}\sum_{i}^{4}\sum_{j=1}^{n_i}\left(Y_{ij,\text{mis}} - \mu^{(c)}_{ij,\text{mis}}\right)'\Sigma^{(c)^{-1}}_{i,\text{mis}}\left(Y_{ij,\text{mis}} - \mu^{(c)}_{ij,\text{mis}}\right)\right\}, \end{aligned} \tag{A.8}$$

where $\mu_{ij,\text{mis}}^{(c)}$ and $\Sigma_{i,\text{mis}}^{(c)}$ are the conditional mean and covariance of $Y_{ij,\text{mis}}$ given $Y_{ij,\text{obs}}$ for subject j in group i, given by

$$\mu_{ij,\text{mis}}^{(c)} = \mu_{ij,\text{mis}} + \Sigma_{i12}\Sigma_{i22}^{-1}(Y_{ij,\text{obs}} - \mu_{ij,\text{obs}}) \quad \text{and} \quad \Sigma_{i,\text{mis}}^{(c)} = \Sigma_{i11} - \Sigma_{i12}\Sigma_{i22}^{-1}\Sigma_{i12}'.$$

Once we sample Y_{mis} from (A.8), we have complete data Y, $Y = (Y_{\text{mis}}', Y_{\text{obs}}')'$. Thus, the full conditional distribution for β, $\beta|Y, V_i, R$ will follow a normal distribution,

$$\pi(\beta|Y, V_i, R) \propto \exp\left\{-\frac{1}{2}(\beta - \mu_\beta)'\Sigma_\beta^{-1}(\beta - \mu_\beta)\right\}, \tag{A.9}$$

where

$$\mu_\beta = \left(\sum_{i=1}^{4}\sum_{j=1}^{n_i} X_{ij}'(V_i R V_i)^{-1}X_{ij}\right)^{-1}\left(\sum_{i=1}^{4}\sum_{j=1}^{n_i} X_{ij}'(V_i R V_i)^{-1}Y_{ij}\right),$$

and

$$\Sigma_\beta = \left(\sum_{i=1}^{4}\sum_{j=1}^{n_i} X_{ij}'(V_i R V_i)^{-1}X_{ij}\right)^{-1}.$$

The full conditional distribution of each component of V_i is given by

$$\pi\left(\frac{1}{\sigma_{it}}|Y, \beta, R, \sigma_{i(-t)}\right) \propto \sigma_{it}^{-(n_i+2r_{it}-1)} \exp\left\{-\frac{a_{it}}{\sigma_{it}^2} - \frac{b_{it}}{\sigma_{it}}\right\}$$

$$(i = 1, \ldots, 4; t = 1, \ldots, 3), \tag{A.10}$$

where $a_{it} = \lambda_{it} + \frac{1}{2}C_{tt}\sum_{j=1}^{n_i}\epsilon_{ijt}^2$ and $b_{it} = \sum_{l=t+1}^{T} C_{tl}\sigma_{il}^{-1}(\sum_{j=1}^{n_i}\epsilon_{ijt}\epsilon_{ijl})$. Here C_{tl} is the (t, l)th element of R^{-1}. The components of V_i can be sampled one by one from (A.10), using the Metropolis-Hastings algorithm (for details, see Liu 2006).

ACKNOWLEDGMENTS

This project is supported by NIH grants CA85295 and HL079457.

[Received October 2005. Revised May 2006]

REFERENCES

Ashford, J. R., and Sowden, R. R. (1970), "Multivariate Probit Analysis," *Biometrics*, 26, 535–546.

Barnard, J., McCulloch, R., and Meng, X. L. (2000), "Modeling Covariance Matrices in Terms of Standard Deviations and Correlations with Application to Shrinkage," *Statistica Sinica*, 10, 1281–1311.

Berry, J. A. M., and Linoff, G. S. (2000), *Mastering Data Mining: The Art and Science of Customer Relationship Management*, New York: Wiley.

Chen, M.-H., and Dey, D. K. (1998), "Bayesian Modelling of Correlated Binary Responses via Scale Mixture of Multivariate Normal Link Functions," *Sankhya*, Series A, 60, 322–343.

——— (2000), "A Unified Bayesian Analysis for Correlated Ordinal Data Models," *Brazilian Journal of Probability and Statistics*, 14, 87–111.

Chen, M.-H., and Shao, Q.-M. (1999), "Properties of Prior and Posterior Distributions for Multivariate Categorical Response Data Models," *Journal of Multivariate Analysis*, 71, 277–296.

Chib, S., and Greenberg, E. (1998), "Bayesian Analysis of Multivariate Probit Models," *Biometrika*, 85, 347–361.

Daniels, M. J. (2006), "Bayesian Modelling of Several Covariance Matrices and Some Results on Propriety of the Posterior for Linear Regression with Correlated and/or Heterogeneous Errors," *Journal of Multivariate Analysis*, 97, 1185–1207.

Daniels, M. J., and Hogan, J. W. (2000), "Re-parameterizing the Pattern Mixture Model for Sensitivity Analyses Under Informative Dropout," *Biometrics*, 56, 1241–1248.

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.

Gelfand, A. E., and Smith, A. F. M. (1990), "Sampling-Based Approaches to Calculating Marginal Densities," *Journal of the American Statistical Association*, 85, 398–409.

Gueorguieva, R. V., and Agresti, A. (2001), "A Correlated Probit Model for Joint Modelling of Clustered Binary and Continuous Responses," *Journal of the American Statistical Association*, 96, 1102–1112.

Kiel, D. P., Puhl, J., Rosen, C. J., Berg, K., Murphy, J. B., and MacLean, D. B. (1998), "Lack of Association Between Insulin-like Growth Factor I and Body Composition, Muscle Strength, Physical Performance or Self Reported Mobility Among Older Persons with Functional Limitations," *Journal of the American Geriatrics Society*, 46, 822–828.

Liechty, J., Liechty, M., and Müller, P. (2004), "Bayesian Correlation Estimation," *Biometrika*, 91, 1–14.

Liu, C. (2001), Comment on "The Art of Data Augmentation," *Journal of Computational and Graphical Statistics*, 10, 75–81.

Liu, X. F. (2006), "Bayesian Methodology for Models with Multivariate (Longitudinal) Outcomes," unpublished Ph.D. thesis, Department of Statistics, University of Florida.

Liu, C., and Sun, D. X. (2000), "Analysis of Interval Censored Data from Fractionated Experiments Using Covariance Adjustments," *Technometrics*, 42, 353–365.

Manly, B. F. J., and Rayner, J. C. W. (1987), "The Comparison of Sample Covariance Matrices using Likelihood Ratio Tests," *Biometrika*, 74, 841–847.

Nandram, B., and Chen M.-H. (1994), "Accelerating Gibbs Sampler Convergence in the Generalized Linear Models via a Reparameterization," *Journal of Statistical Computation and Simulation*, 81, 27–40.

Nobile, A. (1998), "A Hybrid Markov Chain for the Bayesian Analysis of the Multinomial Probit Model," *Statistics and Computing*, 8, 229–242.

Pourahmadi, M., Daniels, M. J., and Park, T. (2006), "Simultaneously Modelling of the Cholesky Decomposition of Several Covariance Matrices," *Journal of Multivariate Analysis*, (in press).

Ritter, C., and Tanner, M. A. (1992), "Facilitating the Gibbs Sampler: The Gibbs Stopper and the Griddy-Gibbs Sampler," *Journal of the American Statistical Association*, 87, 861–868.

Roberts, G. O., Gelman, A., and Gilks, W. R. (1997), "Weak Convergence and Optimal Scaling of Random Walk Metropolis Hastings Algorithms," *The Annals of Applied Probability*, 7, 110–120.

Roberts, G. O., and Rosenthal, J. S. (2001), "Optimal Scaling for Various Metropolis-Hastings Algorithms," *Statistical Science*, 16, 351–367.

Rubin, D. B. (1976), "Inference and Missing Data," *Biometrika*, 63, 581–590.

Tanner, M. A., and Wong, W. H. (1987), "The Calculation of Posterior Distribution by Data Augmentation" (with discussion), *Journal of the American Statistical Association*, 82, 528–550.

Wong, F., Carter, C. K., and Kohn, R. (2003), "Efficient Estimation of Covariance Selection Models," *Biometrika*, 90, 809–830.

Zhang, X., Boscardin, W. J., and Belin, T. R. (2006), "Sampling Correlation Matrices in Bayesian Models with Correlated Latent Variables," *Journal of Computational and Graphical Statistics*, 15, TKK–TKK >>ERIC, FILL IN<<.