

A New Algorithm for Simulating a Correlation Matrix Based on Parameter Expansion and Re-parameterization *

Xuefeng Liu[†]

Michael Daniels[‡]

Department of Statistics

Department of Statistics

University of Florida

University of Florida

*This project is supported by NIH grants CA85295 and HL079457.

[†]Email: xliu@stat.ufl.edu

[‡]Email: mdaniels@stat.ufl.edu

Abstract

The correlation matrix (denoted by R) plays an important role in many statistical models. Unfortunately, sampling the correlation matrix in MCMC algorithms can be problematic. In addition to the positive definite constraint of covariance matrices, it has diagonal elements fixed at 1. In this paper, we propose an efficient two-stage parameter expanded re-parameterization and Metropolis-Hastings (PX-RPMH) algorithm for simulating R . The theory of the PX-RPMH algorithm and sufficient conditions to implement it are shown in detail. 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. A simulation study and a real data example are done to compare the performance of the PX-RPMH algorithm with those of other common algorithms. The results show that our algorithm is more efficient than other methods for sampling a correlation matrix.

Key Words: Data Augmentation, Parameter Expansion, Correlation Matrix, Candidate Prior, Target Prior, PX-RPMH Algorithm.

1 Introduction

The correlation matrix is directly involved in 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, 2005). Sampling the correlation matrix 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 p . 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, a disadvantage is 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 Muller (2004) provide 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 in

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 re-parameterization in the multivariate probit regression model. This method is restricted to a particular prior for R . Zhang, Boscardin and Belin (2004) 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 variances) derived from some inverse Wishart distribution of $\Sigma = DRD$ in model estimation. Then sampling (R, D) jointly was accomplished through a Metropolis Hastings algorithm by first drawing Σ from a pre-specified 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 they are sampling Σ from a candidate distribution 'centered' at the previous value. Though they sample all the components at once, the random walk aspect can slow down mixing and convergence.

In this paper, building on previous work, we propose a two-stage parameter expanded re-parameterization and Metropolis-Hastings (PX-RPMH) algorithm for simulating a correlation matrix from its conditional distribution. In stage 1, using the idea of parameter expanded re-parameterization 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 paper over previous work 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 emphasize the intuition behind 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 Liu (2001), there is no restriction on the prior for R for Bayesian inference. 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* (2004).

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 and techniques to sample from their respective posterior distributions. In Section 3, we propose a parameter expanded re-parameterization and Metropolis-Hastings algorithm (PX-RPMH) for simulating a correlation matrix. This algorithm is derived both for MVP models and for MVR models with a common correlation matrix. Its implementation and properties are explained in detail. A simulation study is reported on in Section 4 and a real data example 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 unified simulation-based inferential methodology for overcoming the problems in fitting MVP models. Let $Q_i = (Q_{i1}, \dots, Q_{iT})$ ($i = 1, \dots, n$) denote a T -dimensional longitudinal vector of binary responses. Let $Y_i = (Y_{i1}, \dots, Y_{iT})$ 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 : $Q_{ij} = I(Y_{ij})$ where I is the indicator function such that $I(Y) = 1$ if $Y > 0$ and $I(Y) = 0$ otherwise ($i = 1, \dots, n; j = 1, \dots, T$). Thus the probability of $Q_i = q_i$, conditional on parameters β and Σ , is given by

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

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 .

To sample from the posterior distribution of (β, R) , a combination of Gibbs sampling (Gelfand and Smith, 1990) and data augmentation (Tanner and Wong, 1987) is typically used. This can be conceptualized as two steps: an imputation (I) step and a posterior (P) step as follows

I step Impute the latent values $Y = \{Y_i : i = 1, \dots, n\}$ by drawing Y_{ij} from the full conditional $f(Y_{ij}|Y_{i(-j)}, \beta, R)$, where Y_{ij} is the j th element of Y_i and $Y_{i(-j)}$ is the vector containing all the elements of Y_i except Y_{ij} ($i = 1, \dots, n; j = 1, \dots, T$).

P step Generate (β, R) from $\pi(\beta, R|Y)$ using a Gibbs sampler by first drawing β from $\pi(\beta|Y, R)$ and then drawing R from $\pi(R|Y, \beta)$.

2.2 Multivariate regression models

We now consider modelling multivariate continuous responses from several groups. Let Y_{ij} be a T -dimensional response vector for the j th individual in group i ($i = 1, \dots, c; j = 1, \dots, n_i$). 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 separate correlation matrix R_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 $V_i = \text{diag}(\sigma_{i1}, \dots, \sigma_{iT})$ and σ_{ik} the standard deviation of the k -th response in group i

($i = 1, \dots, c$; $k = 1, \dots, T$). Thus, variances are allowed to differ across groups, but the correlation matrix is assumed constant.

The parameters β in MVP models and $(\beta(\gamma), V_i)$ in MVR Models can be sampled from their full conditionals using the straightforward techniques as those in other standard MCMC algorithms. We will not provide details in this paper. In what follows, we will show that the PX-RPMH algorithm provides a simple way to sample a correlation matrix in general and specifically in these models.

3 The PX-RPMH algorithm

We propose a parameter expanded re-parameterization and Metropolis-Hastings (PX-RPMH) algorithm to sample a correlation matrix. We derive the algorithm for both MVP models and MVR Model I.

3.1 Overview of algorithm

The PX-RPMH algorithm can be viewed as a two-stage algorithm. We call the first stage parameter expanded re-parameterization (PXR) and the second stage parameter expanded Metropolis-Hastings (PXM). 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 modified parameter expanded Metropolis-Hastings step to ensure that the correlation matrix is sampled from the correct posterior.

The PXR 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 and the off-diagonal elements are restricted to a convex

region $V \subset [-1, 1]^{-\frac{T(T-1)}{2}}$. The first constraint makes drawing R more difficult than drawing a covariance matrix Σ . 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). The correlation matrix R can be converted into a covariance matrix Σ such that the conditional distribution of Σ has a standard distribution (e.g. inverse Wishart distribution). The expansion of the parameter space with the scale parameters, however, must be done to preserve the observed-data model in the following sense: there is a many-to-one mapping, called the reduction function, $\theta = P(\theta^*)$, such that $f(Y_{obs}|\theta) = f(Y_{obs}|\theta^*)$, where Y_{obs} is the observed data, θ is the parameter vector in the original model and θ^* is the parameter vector in the expanded model. For example, in MVP models, $Y_{obs} = Q$, $\theta = (\beta, R)$, $\theta^* = (\beta, \Sigma)$ and $P(\theta^*) = (\beta, D^{-1}\Sigma D^{-1})$. We refer to this step as ‘expansion parameter mining’ or ‘XP mining’.

Once we find the appropriate set of parameters for parameter expansion, we need 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. The standard parameter space has lower dimension than the expanded parameter space because R here has fewer parameters than Σ . Constraints on the expanded space are needed since our candidate transformation needs to be a one-to-one mapping. These constraints vary with the candidate transformation and depend on the specific models or designs. We refer to this step as ‘candidate transformations mining’ or ‘CT mining’. CT mining comprises the second step of the PXR stage.

In the third step, we derive a *candidate prior* for R . We introduce this prior, based on the candidate transformation in step 2 so that the conditional distribution of R can be transformed

into the parameter expanded conditional distribution of Σ from which Σ is easy to sample. This prior for R is not pre-specified or fixed. It is derived based on the CT in step 2, the likelihood function and the priors for parameters, except R , to obtain the candidate density of Σ derived below, from which it is easy to sample. Typically, the *candidate prior* chosen in this step is different from the prior for R we pre-specify for posterior inference. To distinguish them, we call the latter the *target prior* for R . We refer to this step as ‘*candidate prior mining*’ or ‘CP mining’.

The final step is to derive the parameter expanded candidate density (PXCD) of Σ based on the CT in step 2 and the CP in step 3. 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).

Once we obtain the PXCD of Σ , we go to the second stage, i.e, the PXMH stage. In this 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

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

$$\pi(R) \propto \mathbf{I}\{R_{kl} : |R_{kl}| \leq 1 \text{ and } R \text{ is pos. def., } k, l = 1, \dots, T\}, \quad (3.2)$$

where R_{kl} is the element in the k th row and l th column of R . We will weaken this independence assumption in Section 4. Note that (3.2) is the joint uniform prior defined in 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 $P(\theta^*)$ from the XP mining, we find the following one-to-one mapping from $\{Y_i, R\}$ to $\{Y_i^*, \Sigma\}$ for creating draws of the correlation matrix,

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

where Σ is a $T \times T$ positive definite matrix and $\sum_{i=1}^n Y_{it}^{*2} = 1$ for any $t = 1, \dots, T$. 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 t th element of D and x'_{it} is the t th 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 $t = 1, \dots, T$, 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

$$\pi(R) \propto |R|^{-\frac{T+1}{2}} \mathbf{I}\{R_{kl} : |R_{kl}| \leq 1 \text{ and } R \text{ is pos. def.}\}, \quad (3.4)$$

where T is the dimension of 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*

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

where $\nu = n$, $S = \sum_{i=1}^n Y_i^* Y_i^{*'}$, $Y^* = (Y_1^*, \dots, Y_n^*)$ 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 parameter S .

Proof: In Appendix I.

All the proofs of other Propositions and Theorems in the paper are also placed in Appendix I 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 $P(\beta, \Sigma) = (\beta, D^{-1}\Sigma D^{-1})$. Second, 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, i.e., $\pi(\beta, R) = \pi(\beta)\pi(R)$. 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 some 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 MVR Models 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

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

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 PXR stage, XP mining is similar to that in MVP models. In CT mining, we consider the following one-to-one mapping $\mathbb{T}_1 : \{Y_{ij}, R\} \rightarrow \{\epsilon_{ij}, \Sigma\}$

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

where β is fixed and not involved in \mathbb{T}_1 , and ϵ_{ij} is a $T \times 1$ vector satisfying $\sum_{i=1}^c \sum_{j=1}^{n_i} \epsilon_{ijt}^2 = 1$ ($t = 1, \dots, T$). 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 $t = 1, \dots, T$, where D_{tt} is the t -th diagonal element of D and x_{ijt} is the vector of covariates associated with the t th 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) 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 ($i = 1, \dots, c; j =$*

$1, \dots, n_i$) is an inverse Wishart with degrees of freedom ν_1 and scale covariance matrix S_1 , i.e.

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

where $\nu_1 = \sum_{i=1}^c n_i$, $S_1 = \sum_{i=1}^c \sum_{j=1}^{n_i} V_i^{-1} \epsilon_{ij} \epsilon_{ij}' V_i^{-1}$, $\epsilon = (\epsilon'_{11}, \dots, \epsilon'_{cn_c})'$ 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 Model I is given in the following theorem

Theorem 2. Assume that β , R and V_i are, a priori, independent, i.e., $\pi(\beta, R, V_i) = \pi(\beta)\pi(R)\pi(\text{diag}(V_i))$.

If we specify (3.1), (3.2), and (3.6) 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. Extending the idea in Liu (2001), we provide a general framework for deriving the PXCD of Σ through the four steps in the PXR 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 2. XP mining, CT mining, CP mining and PXCD deriving are highly dependent. The first three steps, the likelihood function, and the *candidate/target priors* for the parameters (besides R) in the model determine the last step. The PXCD derived in the fourth step varies with the candidate transformation found in the second step and the *candidate prior* in the third step. In CT

mining, the candidate transformation may vary with the likelihood and the priors for $\theta_{(-R)}$, where $\theta_{(-R)}$ is the parameter vector except R .

Remark 3. 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 that the *target prior* and the *candidate prior* for the parameters except R are the same. However, this condition is not necessary.

Remark 4. The parameters were *a priori* independent in MVP models and MVR Model I. This algorithm can also be applied to cases where these priors are not independent (Liu, 2001; Liu and Sun, 2000). In addition, the target and candidate priors for $\theta_{(-R)}$ can be different.

Remark 5. PXCD deriving is the bridge between the PXRPM 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 re-parameterization (PX-RP) algorithm.

Remark 6. 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 re-drawing 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 7. Since the likelihood function does not change and the *target prior(s)* and the *candidate prior(s)*

for the parameters except R are the same, the acceptance probability in the PXMH stage depends only on the *target prior* (3.2) and the *candidate prior* (3.4) for R .

Remark 8. We can also use a more general *candidate prior* for R

$$\pi(R) \propto |R|^{-\frac{a+1}{2}} \exp \left\{ -\frac{1}{2} \text{tr}(WR^{-1}) \right\} \mathbb{I}\{R_{kl} : |R_{kl}| \leq 1 \text{ and } R \text{ is pos. def.}\} \quad (3.9)$$

where a is a constant and $W = P\Lambda P'$ is a $T \times T$ positive semi-definite tuning parameter matrix, where Λ is a diagonal matrix containing the eigenvalues of W and P is an orthogonal matrix with eigenvectors as columns. (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* (2004)).

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. 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 6-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 ($i = 1, \dots, n; t = 1, \dots, T$). 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

$$\pi(\beta) \propto 1 \quad \text{and} \quad \pi(R) \propto \mathbf{I}\{R_{jk} : |R_{jk}| \leq 1 \text{ and } R \text{ is pos. def.}\}.$$

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 $Y = (Y_1, \dots, Y_{200})$ and $Q = (Q_1, \dots, Q_{200})$. We used the DA algorithm from Section 2.1 to sample from the posterior. Some more details of the full conditional distributions are given next. $Y|Q, \beta, R$ is a truncated multivariate normal with density

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

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

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

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

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

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 τW , where $r^{(k)}$ is the vector of the components of $R^{(k)}$ from the current iteration k defined as $r^{(k)} = (R_{12}^{(k)}, \dots, R_{(T-1)T}^{(k)})$ ($T = 7$), W is the inverse of the negative of the Hessian matrix of $\log L(R|Y, \hat{\beta})$, where Y is 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. Let r be the candidate value from the proposal density $q(r|r^{(k)}, Y, \beta)$. 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 $\alpha = \min \left\{ \frac{f(Y^{(k)}|\beta^{(k)}, R)}{f(Y^{(k)}|\beta^{(k)}, R^{(k)})}, 1 \right\}$.

The Metropolis-Hastings step was applied to $r = (r_1, r_2, r_3, r_4)$ in four blocks since R has 21 elements, 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 5000 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 points r_1, \dots, r_m (set $m = 20$). The sampling required 991 seconds for the Griddy-Gibbs sampler; for the RW-MH algorithm, the running time was 11 seconds; for the PX-RPMH, it was 6 seconds, which is more than 150 times faster than the Griddy-Gibbs sampler.

Second, we examined graphically the mixing of the PX-RPMH algorithm and the random walk Metropolis-Hastings algorithm. 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 2000 samples were retained. We randomly chose 5 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.

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 are interested in the outcome of mean quadriceps strength. The dropout rates in the four treatment groups were 11/41, 13/41, 9/40 and 16/38, respectively. We will conduct our analysis under an assumption of random dropout (MAR) 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 group i ($i = 1, \dots, 4; j = 1, \dots, n_i$). 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,

$$\pi(\beta) \propto \exp\left\{-\frac{1}{2}(\beta - \beta_0)' \Psi_\beta^{-1}(\beta - \beta_0)\right\} \quad (5.1)$$

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

$$\pi(R) \propto \mathbb{I}\{R_{kl} : |R_{kl}| \leq 1 \text{ and } R \text{ is pos. def.}, k, l = 1, \dots, T\}. \quad (5.3)$$

Details on sampling (Y_{mis}, β, V_i) can be found in the Appendix. The full conditional distribution of R is given by

$$\pi(R|Y, \beta) \propto |R|^{-\frac{n_i}{2}} \exp\left\{-\frac{1}{2}\text{tr}\left[\sum_{i=1}^4 \sum_{j=1}^{n_i} V_i (Y_{ij} - X_{ij}\beta)(Y_{ij} - X_{ij}\beta)' V_i R^{-1}\right]\right\} \times \quad (5.4)$$

$$\mathbb{I}\{R_{jk} : |R_{jk}| \leq 1 \text{ and } R \text{ is pos. def.}\}.$$

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 2000 samples (see Figure 2). It is clear that the PX-RPMH algorithm is mixing better than the RW-MH algorithm as was the case with the simulations in Section 4. 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. The ease with which this model can be fit using the PX-RPMH algorithm should encourage its consideration.

6 Discussion

We have proposed a PX-RPMH algorithm for sampling R . This method does not restrict the choice of the prior for R as previous methods and is easy to implement. Our comparisons show that the PX-RPMH algorithm is superior to a 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, 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 and Daniels, 2005). The same family of inverse Wishart distributions should be able to be derived for these models as the parameter expanded candidate density (PXCD) in the PXR stage, 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 PXRП stage more complex. As one referee pointed out, we need to be careful when we apply this algorithm to models in which the dimension of the responses and the correlation matrix are not the same. In this case, deriving the parameter expanded candidate density can be difficult. We will address this issue in the future research.

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 unify 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.

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: John Wiley and Sons, Inc..
- Chen, M.-H. and Dey, D.K. (1998). Bayesian modelling of correlated binary responses via scale mixture of multivariate normal link functions. *Sankhya, Serial A* 60, 322-343.
- Chen, M.-H. and Dey, D.K. (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* (in press).
- 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 Muller, P. (2004) Bayesian Correlation Estimation. *Biometrika* 91, 1-14.

- Liu, X. (2006). Joint models for binary and continuous longitudinal processes. *PhD 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.
- Liu, X.F. and Daniels, M.J. (2005). Joint Models for the association of longitudinal binary and continuous processes (manuscript).
- 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.
- van Dyk, D.A. and Meng, X.L. (2001). The art of data augmentation (with discussion). *Journal of Computational and Graphical Statistics* 10, 1-111.
- 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. (2004). Sampling correlation matrices in Bayesian models with correlated latent variables (manuscript).

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

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

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

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

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

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

Through the one-to-one mapping $\mathbb{T} : (Y, R) \rightarrow (Y^*, \Sigma)$, 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 $\nu = n$, $S = \sum_{i=1}^n Y_i^* Y_i^{*'}$, $Y^* = (Y_1^*, \dots, Y_n^*)$ and $Y_i^* = D(Y_i - X_i \beta)$.

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} . $R = D^{-1} \Sigma D^{-1}$ will be used as the candidate value. It is accepted in the Metropolis-Hastings step with probability α . The α can be 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

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

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

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

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

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

Hence,

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

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 ($i = 1, \dots, c$) is equivalent to drawing ϵ and Σ first, and then translating back to Y and R through \mathbb{T}_1 . $R = D^{-1}\Sigma D^{-1}$ will be used as the candidate value. It is accepted in the Metropolis-Hastings step with probability α . The α can be derived similarly to that in the Proof of Theorem 1.

Sampling algorithm details for data example in Section 5

We provide some details on the data augmented Gibbs sampling algorithm for the (β, V_i, Y_{mis}) .

We start with Y_{mis} . If we partition Y_{ij} , μ_{ij} and Σ_i as $Y_{ij} = \begin{pmatrix} Y_{ij,mis} \\ Y_{ij,obs} \end{pmatrix}$, $\mu_{ij} = \begin{pmatrix} \mu_{ij,mis} \\ \mu_{ij,obs} \end{pmatrix}$ and

$\Sigma_i = \begin{pmatrix} \Sigma_{i11} & \Sigma_{i12} \\ \Sigma_{i21} & \Sigma_{i22} \end{pmatrix}$, then the data augmentation step is implemented by sampling Y_{mis} from

$$f(Y_{mis}|Y_{obs}, \beta, V_i, R) \propto \exp \left\{ -\frac{1}{2} \sum_i \sum_{j=1}^{n_i} (Y_{ij,mis} - \mu_{ij,mis}^{(c)})' \Sigma_{i,mis}^{(c)-1} (Y_{ij,mis} - \mu_{ij,mis}^{(c)}) \right\}, \quad (6.1)$$

where $\mu_{ij,mis}^{(c)}$ and $\Sigma_{i,mis}^{(c)}$ are the conditional mean and covariance of $Y_{ij,mis}$ given $Y_{ij,obs}$ for subject j in group i . They can be formulated as

$$\mu_{ij,mis}^{(c)} = \mu_{ij,mis} + \Sigma_{i12} \Sigma_{i22}^{-1} (Y_{ij,obs} - \mu_{ij,obs}) \quad \text{and} \quad \Sigma_{i,mis}^{(c)} = \Sigma_{i11} - \Sigma_{i12} \Sigma_{i22}^{-1} \Sigma_{i21}'.$$

Once we sample Y_{mis} from (6.1), we obtain complete data Y , $Y = (Y_{mis}', Y_{obs}')'$. The full conditional

distribution for β , $\beta|Y, V_i, R$ is normally distributed as

$$\pi(\beta|Y, V_i, R) \propto \exp \left\{ -\frac{1}{2}(\beta - \mu_\beta)' \Sigma_\beta^{-1} (\beta - \mu_\beta) \right\}, \quad (6.2)$$

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

$$\pi\left(\frac{1}{\sigma_{it}} | Y, \beta, R, \sigma_{i(-t)}\right) \propto \sigma_{it}^{-(n_i+2r_{it}-1)} \exp \left\{ -\frac{a_{it}}{\sigma_{it}^2} - \frac{b_{it}}{\sigma_{it}} \right\} \quad (i = 1, \dots, 4; t = 1, \dots, 3), \quad (6.3)$$

where $a_{it} = \lambda_{it} + \frac{1}{2} C_{tt} \sum_{j=1}^{n_i} \epsilon_{ijt}^2$ and $b_{it} = \sum_{l=t+1}^T C_{tl} \sigma_{il}^{-1} (\sum_{j=1}^{n_i} \epsilon_{ijl} \epsilon_{ijl})$. Here C_{tl} is the (t, l) th element of R^{-1} . The components of V_i can be sampled one by one from (6.3), using the Metropolis-Hastings algorithm (for details, see Liu (2006)).

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
1	0.81(0.75,0.86)	634(462,855)	68.9(60.8,76.7)
	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)
2	-	514.03(384,742)	68.4(61.4,75.2)
	-	589.(448,822)	66.0(58.6,73.7)
	-	492.(357,672)	65.1(58.0,72.0)
3	-	606 (465,775)	65.6(58.1,73.0)
	-	640(507,818)	81.0(73.0,88.7)
	-	465(350,601)	72.4(65.6,79.2)
4	-	689(517,932)	65.2(57.4,73.2)
	-	628 (477,868)	62.1(54.5,69.8)
	-	554 (407,782)	62.6(55.0,70.4)

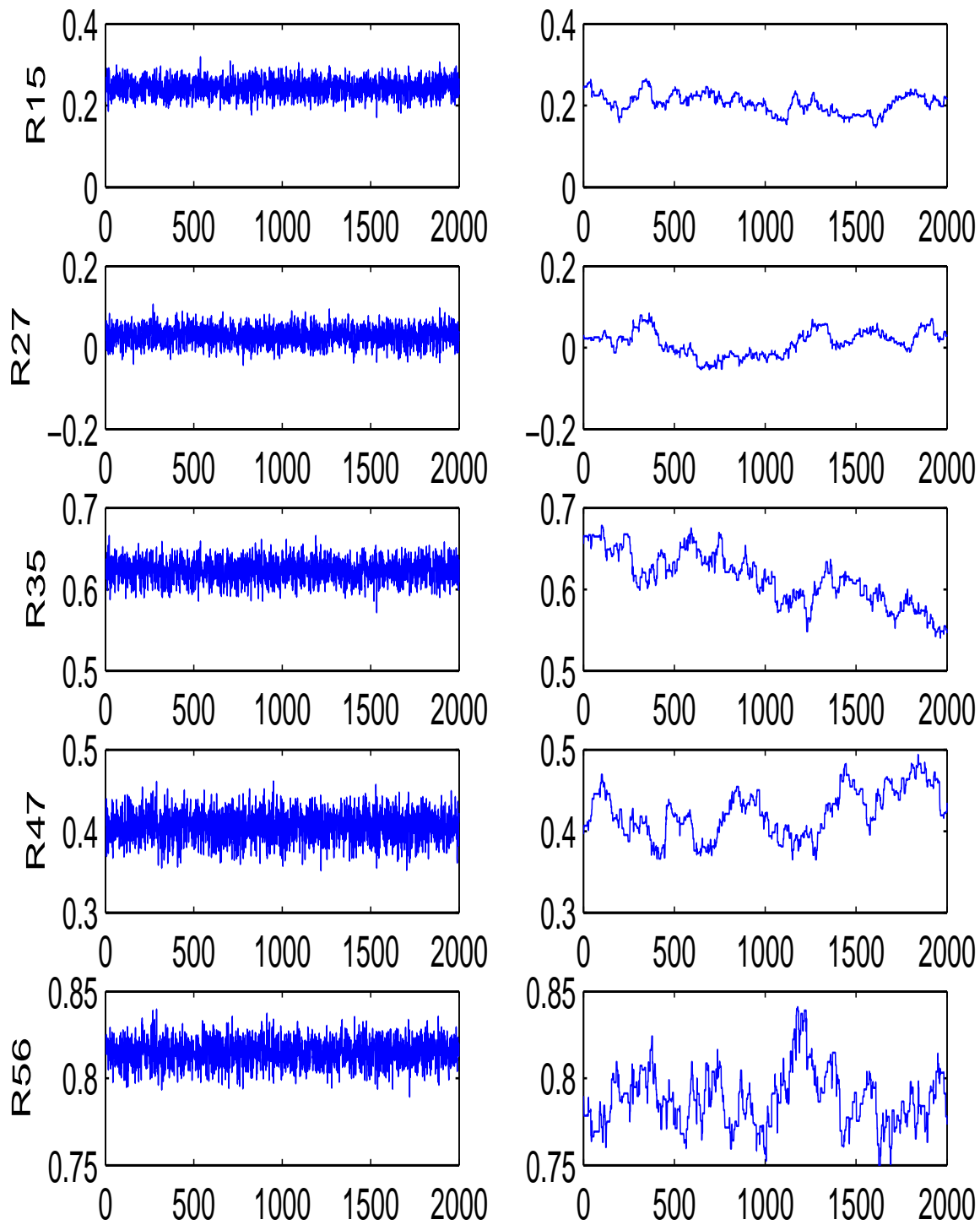


Figure 1: Trace plots of 5 components of R over the last 2000 iterations of 50,000 iterations for the PX-RPMH algorithm and RW-MH algorithm in the simulation study.

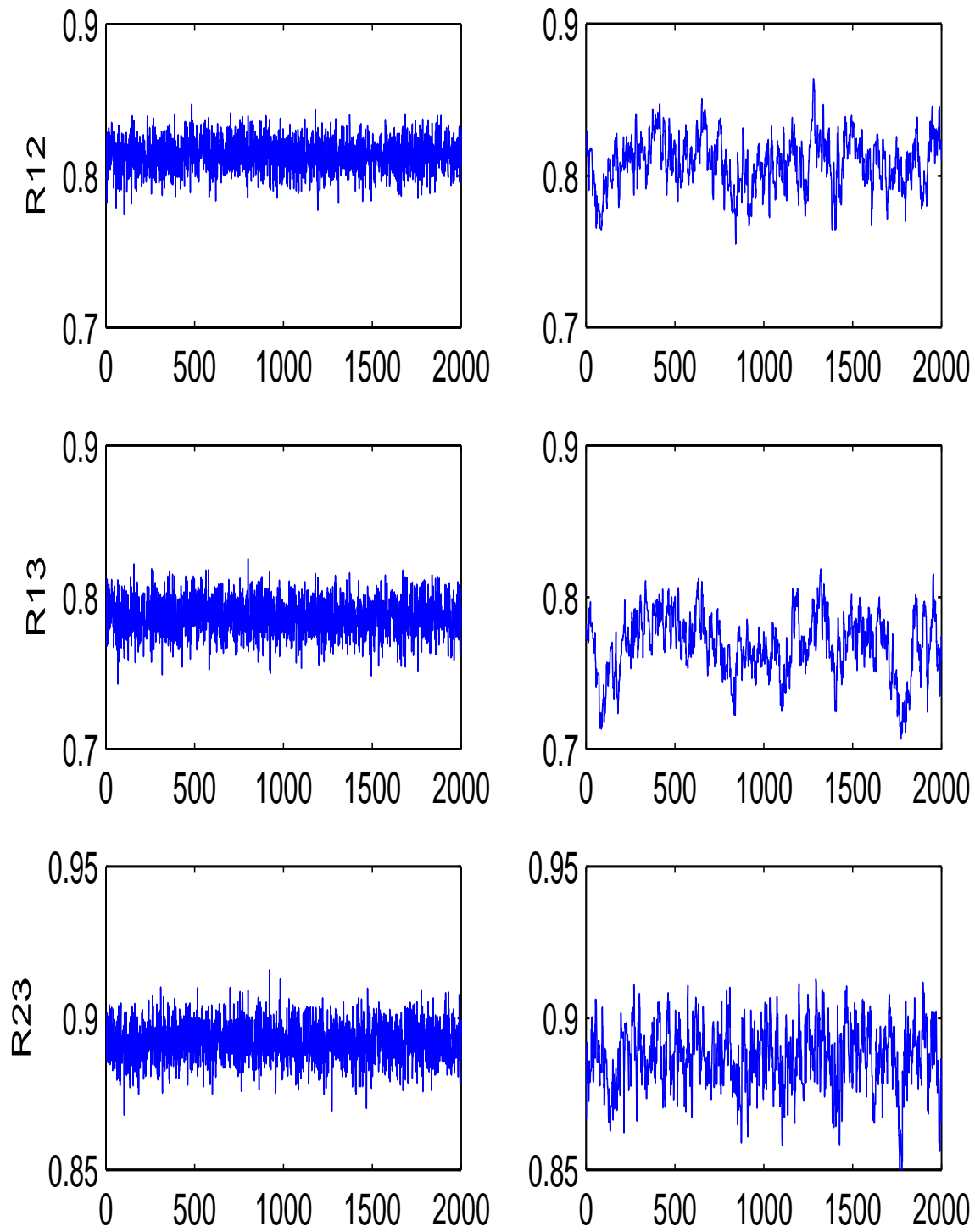


Figure 2: Trace plots of the three components of \mathcal{A} over the last 2000 iterations of 10,000 iterations for the PX-RPMH algorithm and RW-MH algorithm in the real data example in Section 5.