Sankhyā: The Indian Journal of Statistics
Special Issue on Bayesian Analysis
1998, Volume 60, Series A, Pt. 3, pp. 322-343

BAYESIAN MODELING OF CORRELATED BINARY RESPONSES VIA SCALE MIXTURE OF MULTIVARIATE NORMAL LINK FUNCTIONS

By MING-HUI CHEN*
Worcester Polytechnic Institute, Worcester
and
DIPAK K. DEY
University of Connecticut, Storrs

SUMMARY. In this article, we consider using scale mixture of multivariate normal links (SMMVN) to model binary responses when binary observations are taken from the same individuals or are taken over time in a longitudinal fashion. SMMVN-links are quite rich, which include multivariate probit, Student's t links, logit, symmetric stable link, and exponential power link. Fully parametric classical approaches to these are intractable and thus Bayesian methods are pursued using a Markov chain Monte Carlo (MCMC) sampling based approach. Necessary theory involved in Bayesian modeling and computation is provided. In particular, we produce a new look at the multivariate logit model, the most popular model in this context. Further, we develop various efficient computational algorithms for this complex simulation problem. Finally, a real data example from the Indonesian Children's Health Study is used to illustrate the proposed methodology.

1. Introduction

There is growing interest in the statistical literature concerning modeling and analysis of correlated binary data. This type of data often arises when two or more binary responses are taken at one time for the same subjects or when repeated measurements are taken over time such as in longitudinal studies. Generalized linear regression methods are considered for such correlated binary data to study the relation between various covariates and the dichotomous outcome measure.

Prentice (1988) has provided a comprehensive review of various modeling

AMS (1991) *subject classifications.* 62F15, 62J12.
Key words and phrases. Bayesian computation; Bayesian hierarchical model; exponential power link; Gibbs sampler; Markov chain Monte Carlo; multivariate generalized linear models; multivariate logit; multivariate probit; multivariate Student's t link; positive stable link.
* Research supported by the National Science Foundation under Grant No. DMS-9702172.

strategies using generalized linear regression analysis of correlated binary data with covariates associated at each binary response. Following Liang and Zeger (1986) and Zeger and Liang (1986), Prentice used the generalized estimating equation (GEE) approach to obtain consistent and asymptotically normal estimators of regression coefficients.

In this paper we develop an exact small sample Bayesian analysis of such correlated binary data models. We propose a very novel modeling approach which is based on multivariate link functions using a very rich class of scale mixtures of normals. Such models are very flexible and include, as a special case, multivariate probit (MVP), t-link (MVT), logit (MVL), stable distribution family links (MVS), and exponential power distribution family links (MVEP). The MVP model was introduced by Ashford and Sowden (1970) and studied further by Amemiya (1985). Recently Chib and Greenberg (1998) used the MVP model, and Dey and Chen (1996) used the MVP and MVT models along with models proposed by Prentice (1988) in a Bayesian framework.

The objective of this paper is to explore different modeling strategies for the analysis of correlated or longitudinal binary responses from a Bayesian perspective by incorporating scale mixtures of normals as link functions. By considering a rich class of link functions in such models (within a parametric framework) this approach unifies all the previous methods in a systematic manner. This paper also produces a new look at the MVL, the most popular model in this context. We adopt the Markov chain Monte Carlo (MCMC) framework (e.g., Gelfand and Smith, 1990 and Tierney, 1994) to simulate the posterior distribution for proposed models. Various efficient computational algorithms for this complex simulation problem are developed. For example, we develop efficient proposal densities in several Metropolis's steps.

The rest of the paper is organized as follows. In Section 2, we propose the general structure of the scale mixture of multivariate normals (SMMVN) link models and discuss several special cases of such models. Section 3 is devoted to the development of the prior distributions and on the development of the distribution theory involved in the posterior calculations. In particular, we develop efficient Metropolis algorithms for the MVL, MVS and MVEP link models. In Section 4 we apply our proposed methodology to the Indonesian children data. Finally, Section 5 gives brief concluding remarks.

2. The Scale Mixture of Multivariate Normals Link Models

We first introduce some notation which will be used throughout the paper. Suppose that we observe a binary (0-1) response Y_{ij} on the i-th observations and j-th variable and let $x_{ij} = (x_{ij1}, x_{ij2}, \ldots, x_{ijp})$ be the corresponding p-dimensional row regression vector for $i = 1, 2, \ldots, n$ and $j = 1, 2, \ldots, J$. (Note that x_{ij1} may be 1, which corresponds to an intercept.) Denote $Y_i =$

$(Y_{i1}, Y_{i2}, \ldots, Y_{iJ})'$ and assume that Y_{i1}, Y_{i2}, ..., Y_{iJ} are dependent and Y_1, Y_2, ..., Y_n are independent. Let $y_i = (y_{i1}, y_{i2}, \ldots, y_{iJ})'$ and $y = (y_1, y_2, \ldots, y_n)$ be the observed data. Also let $\beta = (\beta_1, \beta_2, \ldots, \beta_p)'$ be a p-dimensional column vector of regression coefficients.

2.1. *SMMVN-link models.* In spirit of the models for correlated binary data given in Prentice (1988), Dey and Chen (1996) considered stratified and mixture models and conditional models. Further Dey and Chen (1996) used multivariate probit (MVP) model (see also Chib and Greenberg 1998) and multivariate t-link models for the correlated binary response Y_{i1}, Y_{i2}, ..., Y_{iJ}. Here, we consider more general scale mixture of multivariate normal link functions.

In order to set up our general SMMVN-link models, we introduce a J-dimensional (latent) random vector $w_i = (w_{i1}, w_{i2}, \ldots, w_{iJ})'$ such that

$$Y_{ij} = \begin{cases} 1 & \text{if } w_{ij} > 0 \\ 0 & \text{if } w_{ij} \le 0 \end{cases} \qquad \ldots (2.1)$$

and assume that

$$w_i \sim N(x_i\beta, \kappa(\lambda)\Sigma), \qquad \ldots (2.2)$$

and

$$\lambda \sim \pi(\lambda), \qquad \ldots (2.3)$$

where $\kappa(\lambda)$ is a positive function of one-dimensional positive-valued scale mixture variable λ, $\pi(\lambda)$ is a mixture distribution which is either discrete or continuous, and

$$x_i = \begin{pmatrix} x_{i1} \\ x_{i2} \\ \vdots \\ x_{iJ} \end{pmatrix}.$$

In (2.2), we take $\Sigma = (\rho_{jj^*})_{J\times J}$ to be a correlation matrix such that $\rho_{jj} = 1$ to ensure the identifiability of the parameters. See Chib and Greenberg (1998) or Dey and Chen (1996) for the detailed discussions. Such a w_i is sometimes called a tolerance variable since in a bioassay setting w_i can be a lethal dose of a drug. It follows that the distribution of w_i determines the joint distribution of Y_i through (2.1) and the correlation matrix Σ captures the correlations among the Y_{ij}'s. More specifically, we have the joint distribution of the responses as

$$P(Y_{i1} = y_{i1}, Y_{i2} = y_{i2}, \ldots, Y_{iJ} = y_{iJ} | \beta, \Sigma, \lambda, x_i)$$
$$= \int_{A_{i1}} \int_{A_{i2}} \cdots \int_{A_{iJ}} \frac{1}{(2\pi\kappa(\lambda))^{J/2} |\Sigma|^{\frac{1}{2}}} \exp\left\{ -\frac{\kappa^{-1}(\lambda)}{2} (w_i - x_i\beta)'\Sigma^{-1}(w_i - x_i\beta) \right\} dw_i, \qquad \ldots (2.4)$$

where

$$A_{ij} = \begin{cases} (-\infty, 0] & \text{if } y_{ij} = 0 \\ (0, \infty) & \text{if } y_{ij} = 1 \end{cases}. \qquad \ldots (2.5)$$

REMARK 1. Alternative to our SMMVN-link models, random effects models are commonly used in this context. One version of a random effects model is given as follows. Assume that

$$P(Y_{ij} = y_{ij}|\beta, \epsilon_i) = F^{y_{ij}}(x_{ij}\beta + \epsilon_i)\,[1 - F(x_{ij}\beta + \epsilon_i)]^{y_{ij}} \ ,$$

for $j = 1, 2, \ldots, J$, where F is a cumulative distribution function, which serves as a link function, and random effects, ϵ_i, are assumed to be independent and identically distributed with $E(\epsilon_i) = 0$ and $Var(\epsilon_i) = \sigma^2$ for $i = 1, 2, \ldots, n$. In the above setting, random effects, ϵ_i's, are automatically taken to be exchangeable. Furthermore, it is well known that the variance σ^2 is nearly not identifiable. The SMMVN-link models overcome the difficulty that arises in random effects models. In the SMMVN-link models, Σ is not a general variance and covariance matrix and indeed, it is a correlation matrix. Therefore, the identifiability problem does not exist any more. Furthermore, the SMMVN-link models do not assume the exchangeability on the correlation structure.

Next we present various special cases of our SMMVN-link models.

2.2. *MVP and MVT models.* It is easy to observe that MVP and MVT models are two special cases of the SMMVN-link models. Note that MVP models were considered by Chib and Greenberg (1998) and further elaborated by Dey and Chen (1996) while MVT models were discussed only in Dey and Chen (1996). In fact, when we take $\kappa(\lambda) = 1$ and the mixture distribution $\pi(\{1\}) = 1$, then the SMMVN models lead to the MVP models. Similar to the MVP models, when we take $\kappa(\lambda) = 1/\lambda$ and $\lambda \sim \mathcal{G}(\nu/2, \nu/2)$, i.e.,

$$\pi(\lambda) = \frac{1}{\Gamma\left(\frac{\nu}{2}\right)} \left(\frac{\nu}{2}\right)^{\nu/2} \lambda^{\nu/2-1} \exp\left\{-\frac{\nu}{2}\lambda\right\},$$

the SMMVN-link models give the MVT models. The latter result can be simply seen by the following identity:

$$\int_0^\infty \left(\frac{\lambda}{2\pi}\right)^{\frac{J}{2}} |\Sigma|^{-\frac{1}{2}} \exp\left\{-\frac{\lambda}{2}(w_i - x_i\beta)'\Sigma^{-1}(w_i - x_i\beta)\right\} \frac{1}{\Gamma\left(\frac{\nu}{2}\right)} \left(\frac{\nu}{2}\right)^{\frac{\nu}{2}} \lambda^{\frac{\nu}{2}-1} \exp\left\{-\frac{\nu}{2}\lambda\right\} d\lambda$$
$$= \frac{\Gamma\left(\frac{1}{2}(\nu + J)\right)}{(\pi\nu)^{(1/2)J}\Gamma\left(\frac{1}{2}\nu\right)|\Sigma|^{1/2}} \left(1 + \nu^{-1}(w_i - x_i\beta)'\Sigma^{-1}(w_i - x_i\beta)\right)^{-\frac{1}{2}(\nu+J)}. \qquad \ldots(2.6)$$

The right hand side of (2.6) is exactly the probability density function of $t_\nu(x_i\beta, \Sigma)$. Note that the special case of $t_\nu(x_i\beta, \Sigma)$ where $\nu = 1$ is termed a multivariate Cauchy (MVC) distribution, and another special case of $t_\nu(x_i\beta, \Sigma)$ with $\nu \to \infty$ is the multivariate normal distribution $N(x_i\beta, \Sigma)$. To limit model complexity, we consider only fixed ν so that we can investigate different MVT-link models. A uniform prior on $1/\nu$ $(0 < 1/\nu \leq 1)$ can be considered. However, this will bring additional computational burden.

2.3. *Multivariate logit models.* Logit models are widely used to fit binary data (e.g., see Prentice 1988). Here, we propose multivariate logit (MVL) models to fit the longitudinal binary data. It is interesting to note that a MVL model is also a special case of the SMMVN-link model. As pointed out by Choy (1995), the SMMVN-link model leads to the MVL model when $\kappa(\lambda) = 4\lambda^2$ and λ follows an asymptotic Kolmogorov distribution with density

$$\pi(\lambda) = \pi_K(\lambda) = 8\sum_{k=1}^{\infty}(-1)^{k+1}k^2\lambda\exp\{-2k^2\lambda^2\}. \qquad \ldots(2.7)$$

Choy (1995) did not pursue the logistic distribution for any data analysis due to the complexity of this mixture density. However, when we use a trick of the t approximation to the logistic distribution, the MVL models become quite computationally attractive. More importantly, the MVL models do not assume exchangeability on the correlation structure, which is advantageous compared to the random effects type of logistic regression models, for example, stratified and mixture models given in Prentice (1988).

In order to obtain an approximation to the Kolmogorov distribution $\pi_K(\lambda)$, we use a connection between a univariate t_ν distribution having ν degrees of freedom, and logistic distribution, both with location parameter 0 and scale parameter 1. Recall that for the univariate logistic distribution, the probability density function (pdf) is given by $\pi_L(w) = e^{-w}\left(1+e^{-w}\right)^{-2}$ and for the t_ν, the pdf is

$$\pi_{t_\nu}(w) = \frac{\Gamma\left(\frac{1}{2}(\nu+1)\right)}{\sqrt{\pi\nu}\,\Gamma\left(\frac{1}{2}\nu\right)}\left(1+\frac{w^2}{\nu}\right)^{-(\nu+1)/2}.$$

Then, we have the following two identities:

$$\pi_L(w) = \int_0^\infty \frac{1}{\sqrt{2\pi}}\left(\frac{1}{4\lambda^2}\right)^{1/2}\exp\left\{-\left(\frac{1}{4\lambda^2}\right)\frac{w^2}{2}\right\}\pi_K(\lambda)d\lambda \qquad \ldots(2.8)$$

$$\pi_{t_\nu}(w) = \int_0^\infty \frac{\lambda^{1/2}}{\sqrt{2\pi}}\exp\left\{-\lambda\frac{w^2}{2}\right\}\frac{1}{\Gamma\left(\frac{\nu}{2}\right)}\left(\frac{\nu}{2}\right)^{\nu/2}\lambda^{\frac{\nu}{2}-1}\exp\left\{-\frac{\nu}{2}\lambda\right\}d\lambda. \quad \ldots(2.9)$$

Albert and Chib (1993) observed empirically that a t_ν random variable is approximately b times a logistic random variable with appropriate choices of positive-valued ν and b. Therefore, the logistic pdf $\pi_L(w)$ can be approximated by

$$\pi_L(w) \approx b\pi_{t_\nu}(bw) = \int_0^\infty \frac{b\lambda^{1/2}}{\sqrt{2\pi}}\exp\left\{-b^2\lambda\frac{w^2}{2}\right\}\frac{1}{\Gamma\left(\frac{\nu}{2}\right)}\left(\frac{\nu}{2}\right)^{\nu/2}\lambda^{\frac{\nu}{2}-1}\exp\left\{-\frac{\nu}{2}\lambda\right\}d\lambda. \qquad \ldots(2.10)$$

In (2.10), transforming $b^2\lambda$ to $1/(4\lambda^2)$ leads to

$$\pi_L(w) \approx \int_0^\infty \frac{1}{\sqrt{2\pi}} \left(\frac{1}{4\lambda^2}\right)^{1/2} \exp\left\{-\left(\frac{1}{4\lambda^2}\right)\frac{w^2}{2}\right\} \frac{\left(\frac{\nu}{8b^2}\right)^{\nu/2}}{\Gamma\left(\frac{\nu}{2}\right)(\lambda^2)^{\nu/2+1}} \exp\left\{-\left(\frac{\nu}{8b^2}\right)\frac{1}{\lambda^2}\right\} 2\lambda d\lambda. \quad \ldots(2.11)$$

We denote that

$$g_L(\lambda|\nu, b) = \frac{\left(\frac{\nu}{8b^2}\right)^{\nu/2}}{\Gamma\left(\frac{\nu}{2}\right)(\lambda^2)^{\nu/2+1}} \exp\left\{-\left(\frac{\nu}{8b^2}\right)\frac{1}{\lambda^2}\right\} 2\lambda. \quad \ldots(2.12)$$

By comparing the right hand sides of (2.8) and (2.11), it is expected that $g_L(\lambda|\nu, b)$ would serve as a good proposal density of $\pi_K(\lambda)$ with appropriate choices of ν and b. It is interesting to mention that if $\lambda^2 \sim \mathcal{IG}\left(\nu/2,\ \nu/8b^2\right)$, where $\mathcal{IG}(u, v)$ is an inverse gamma distribution with a probability density function $\pi_{\mathcal{IG}}(\lambda|u, v) = \frac{v^u}{\Gamma(u)\lambda^{u+1}} e^{-v/\lambda}$, $\lambda > 0$, then $\lambda \sim g_L(\lambda|\nu, b)$.

In Appendix A, we show that the best choices of ν and b are $\nu = 5$ and $b = .712$ and we also provide an efficient way to evaluate the infinite series of $\pi_K(\lambda)$. Finally, we comment that use of $g_L(\lambda|\nu, b)$ will simplify the implementation of MCMC sampling, for example, Metropolis steps. We will further elaborate this point in Section 3 below.

2.4. *Symmetric multivariate stable link models.* Univariate stable distributions, which are a class of limiting distributions for sums of *i.i.d.* random variables, are characterized by four parameters, the index of stability α, $[0 < \alpha \leq 2]$, the scale parameter σ, $[\sigma > 0]$, the skewness parameter γ, $[-1 \leq \gamma \leq 1]$ and the shift parameter μ, $[-\infty < \mu < \infty]$. The parameter α measures the degree of peakedness and the heaviness of the tails of the stable distribution. Setting the skewness parameter $\gamma = 0$, we get a symmetric stable distribution. When $\alpha = 2$, the stable distribution corresponds to a normal distribution, while when $\alpha < 2$, the variance of the distribution becomes infinite. This characteristic makes the stable distribution useful in modeling data that admit observations of very large magnitude.

A multivariate stable distribution can be obtained as a scale mixture of multivariate normals with $\kappa(\lambda) = 2\lambda$ and the mixture distribution $\pi(\lambda) = S^P(\alpha, 1)$, where the pdf of the positive stable distribution $S^P(\alpha, 1)$ in the polar form is given by

$$\pi_{S^P}(\lambda|\alpha, 1) = \frac{\alpha}{1-\alpha}\lambda^{-\left(\frac{\alpha}{1-\alpha}+1\right)} \int_0^1 s(u) \exp\left\{-\frac{s(u)}{\lambda^{\frac{\alpha}{1-\alpha}}}\right\} du, \quad \text{for } 0 < \alpha < 1 \quad \ldots(2.13)$$

with

$$s(u) = \left(\frac{\sin(\alpha\pi u)}{\sin(\pi u)}\right)^{\frac{\alpha}{1-\alpha}} \left(\frac{\sin[(1-\alpha)\pi u]}{\sin(\pi u)}\right).$$

(See Appendix B for more details of the positive stable distribution $S^P(\alpha, 1)$.) It is easy to see that if λ and U follow the bivariate distribution with its pdf

$$\pi(\lambda, u|\alpha) = \frac{\alpha}{1-\alpha} \lambda^{-\left(\frac{\alpha}{1-\alpha}+1\right)} s(u) \exp\left\{-\frac{s(u)}{\lambda^{\frac{\alpha}{1-\alpha}}}\right\}, \quad \text{for } \lambda > 0, \quad 0 < u < 1, \qquad \ldots(2.14)$$

then the marginal distribution of λ is the positive stable distribution $S^P(\alpha, 1)$. In our scenario, to obtain a robust multivariate link model, we consider a symmetric multivariate stable (MVS) distribution $S_J(2\alpha, 0, x_i\beta, \Sigma)$ for w_i where the log characteristic function of $S_J(2\alpha, 0, x_i\beta, \Sigma)$ is given by

$$\ln \psi(t) = \mathbf{i}\ (x_i\beta)' t - (t'\Sigma t)^{\alpha}, \text{ for } \alpha \in [1/2, 1),$$

with $t = (t_1, \ldots, t_J)'$ and $\mathbf{i}^2 = -1$. Note that when $\alpha = 1/2$, $S_J(1, 0, x_i\beta, \Sigma)$ is the multivariate Cauchy distribution, while $S_J(2, 0, x_i\beta, \Sigma) = \lim_{\alpha \to 1} S_J(2\alpha, 0, x_i\beta, \Sigma)$ is a multivariate normal distribution. Therefore, MVC is a special case of MVT as well as a special case of MVS, and MVP is a limiting case of MVS. For ease of model complexity, we consider only fixed α so that we can investigate different MVS-link models. A uniform prior on α can be considered and it brings additional computational burden. However, a method proposed by Buckle (1996) can be adopted in implementing the Gibbs sampler, to circumvent this problem.

2.5. *Symmetric multivariate exponential power link models.* Exponential power family distributions play an important role in Bayesian modeling, as indicated in Box and Tiao (1992), where they used a univariate exponential power family to model random effects in linear and nonlinear models. In our context, we develop a symmetric multivariate exponential power family to obtain a robust link function. Formally, the pdf of the multivariate exponential power family (MVEP) distribution has the form

$$\pi_{EP}(w_i|x_i\beta, \Sigma, \alpha) = c_J\, |\Sigma|^{-1/2} \exp\left\{-\left[c_0 (w_i - x_i\beta)' \Sigma^{-1} (w_i - x_i\beta)\right]^{\alpha}\right\}, \quad \text{for } 1/2 \le \alpha \le 1, \qquad \ldots(2.15)$$

where α is called the kurtosis parameter and constants c_0 and c_J are defined as

$$c_0 = \frac{\Gamma\left(\frac{3}{2\alpha}\right)}{\Gamma\left(\frac{1}{2\alpha}\right)} \quad \text{and} \quad c_J = \frac{\alpha c_0^{J/2} \Gamma\left(\frac{J}{2}\right)}{\Gamma\left(\frac{J}{2\alpha}\right) \pi^{J/2}}. \qquad \ldots(2.16)$$

Again, it follows from Andrews and Mallows (1974), West (1987), and Choy (1995) that a MVEP-link model is also a special case of the SMMVN-link model with $\kappa(\lambda) = 1/(2c_0\lambda)$ and $\pi(\lambda) = \left(\frac{1}{\lambda}\right)^{J/2} \pi_{S^P}(\lambda|\alpha, 1)$, where $\pi_{S^P}(\lambda|\alpha, 1)$ is defined in (2.13). There are two interesting special cases of the MVEP distributions, that is, the multivariate normal ($\alpha = 1$) and the multivariate double exponential distribution ($\alpha = 1/2$). Similar to the MVS-link models, we consider only fixed α in our example.

3. The Prior Distributions and Posterior Computations

In this section we present prior distributions for various SMMVN-link models and develop algorithms to perform posterior computations for such models.

3.1. *Prior distributions.* First, we choose the same prior distribution for the regression coefficient vector β for all SMMVN-link models presented in Section 2. That is,

$$\pi(\beta|\beta_0, B_0) \;\propto\; \exp\left\{-\frac{1}{2}(\beta-\beta_0)' B_0 (\beta-\beta_0)\right\}, \qquad \ldots (3.1)$$

where B_0 is a precision matrix, β_0 is a location parameter vector, and both β_0 and B_0 are prespecified. Typically, we choose $\beta_0 = 0$ and $B_0 = diag\,(B_1, B_2, \ldots, B_p)$ where B_j is chosen to be small (e.g., $B_j = 0.01$) so that a vague prior distribution for β is obtained, which ensures that the posterior is driven by the data.

Second, we denote $vec^*(\Sigma) = (\rho_{12}, \rho_{13}, \ldots, \rho_{J-1,J})'$. Then, analogous to Chib and Greenburg (1998), we choose

$$\pi(vec^*(\Sigma)|\Sigma_0, G_0) \;\propto\; \exp\left\{-\frac{1}{2}(vec^*(\Sigma) - vec^*(\Sigma_0))' G_0 (vec^*(\Sigma) - vec^*(\Sigma_0))\right\} \qquad \ldots (3.2)$$

for $vec^*(\Sigma) \in V$ where Σ_0 is a $J \times J$ correlation matrix with all diagonal elements equal to one, G_0 is a $(J(J-1)/2) \times (J(J-1)/2)$ precision matrix, and the region V is a subset of the region $[-1,1]^{J(J-1)/2}$ that leads to a proper correlation matrix. As mentioned by Chib and Greenburg (1998) and also shown by Rousseeuw and Molenberghs (1994), the region V forms a convex solid body in the hypercube $[-1,1]^{J(J-1)/2}$. Note that both hyper parameters Σ_0 and G_0 are to be specified. The simplest choices of Σ_0 and G_0 are $\Sigma_0 = I_J$ and $G_0 = I_{J(J-1)/2}$, which are the J and $J(J-1)/2$ dimensional identity matrices.

3.2. *Posterior computations.* We use Gibbs sampling (e.g, Geman and Geman, 1984 and Gelfand and Smith, 1990) to perform the posterior computation. We present the steps needed to perform the Gibbs sampling algorithms for all SMMVN-link models considered in Section 2 in turn.

3.2.1. *Multivariate probit models.* To run the Gibbs sampler for the multivariate models, we need to sample β, w_i, and Σ from their respective conditional distributions. Let $\hat{\beta} = B^{-1}\left(B_0\beta_0 + \sum_{i=1}^n x_i'\Sigma^{-1}w_i\right)$ and $B = B_0 + \sum_{i=1}^n x_i'\Sigma^{-1}x_i$. Then, given the w_i and Σ, we have

$$\beta \mid w_1, w_2, \ldots, w_n, \Sigma, y \;\sim\; N\left(\hat{\beta}, B^{-1}\right).$$

From (2.4), it can be seen that the full conditional distribution of w_i is multivariate normal truncated to a region determined by y_i. More specifically,

$$\pi(w_i \mid \beta, \Sigma, y_i) \propto$$
$$\prod_{j=1}^{J} \left[1_{\{w_{ij}>0\}}1_{\{y_{ij}=1\}} + 1_{\{w_{ij}\leq 0\}}1_{\{y_{ij}=0\}}\right] \exp\left\{-\frac{1}{2}(w_i - x_i\beta)'\Sigma^{-1}(w_i - x_i\beta)\right\}.$$

As suggested by Geweke (1991), we generate this truncated multivariate normal variate w_i to use a cycle of J Gibbs steps through the components of w_i so that w_{ij} is sampled from a truncated normal over the interval A_{ij} given in (2.5).

Finally, we sample Σ from its conditional distribution. The conditional likelihood function $L(\Sigma|\beta, w, y)$, ignoring the normalizing constant, is

$$|\Sigma|^{-\frac{n}{2}} \exp\left\{-\frac{1}{2}\sum_{i=1}^{n}(w_i - x_i\beta)'\Sigma^{-1}(w_i - x_i\beta)\right\},$$

where $w = (w_1, w_2, \ldots, w_n)$ and $vec^*(\Sigma) \in V$. The full conditional density is proportional to $L(\Sigma|\beta, w, y)\pi(vec^*(\Sigma)|\Sigma_0, G_0)$. Because of the complexity of this conditional distribution, we use a Hastings algorithm (e.g., Metropolis *et al.*, 1951, Hastings, 1970, and Tierney, 1994) to generate Σ. Let Σ be the current value. Using an algorithm analogous to Chib and Greenburg (1998), we generate candidate values Σ^* by specifying a random walk chain $\Sigma^* = \Sigma + H$, where $H = (h_{ij})$ is an increment matrix with zeros on the diagonals and with means $E(h_{ij}) = 0$. Let ξ be the least eigenvalue of Σ. Alternatively to Chib and Greenburg's algorithm, we use a Metropolized hit-and-run algorithm (Chen and Schmeiser, 1993) to simulate H. This algorithm operates as follows:

(i) generate an *i.i.d* $N(0,1)$ random variate sequence of $z_{12}, z_{13}, \ldots, z_{J-1,J}$;

(ii) generate a signed distance d from $N(0, \sigma_d^2)$ truncated to $(-\frac{\xi}{\sqrt{2}}, \frac{\xi}{\sqrt{2}})$, where σ_d is a prespecified tuning parameter;

(iii) calculate

$$h_{ij} = \frac{dz_{ij}}{\left(\sum_{j=1}^{J-1}\sum_{l=j}^{J} z_{jl}^2\right)^{1/2}}$$

for $i < j$, $h_{ii} = 0$, and $h_{ij} = h_{ji}$ for $i > j$.

Note that in (ii), σ_d^2 is appropriately chosen so as to avoid excessive rejections in the Hastings algorithm, and in practice it is sufficient to specify $0.3 \leq \sigma_d \leq 1$. As discussed in Marsaglia and Olkin (1994), H generated in this manner guarantees that Σ^* is positive definite.

Let ξ^* be the least eigenvalue of Σ^*. Following Chen and Schmeiser (1993), given the proposal value, a move to the point Σ^* is made with probability

$$\min\left\{\frac{L(\Sigma^*|\beta, w, y)\pi(vec^*(\Sigma^*)|\Sigma_0, G_0)\left(\Phi\left(\frac{\xi^*}{\sqrt{2}\sigma_d}\right) - \Phi\left(-\frac{\xi^*}{\sqrt{2}\sigma_d}\right)\right)}{L(\Sigma|\beta, w, y)\pi(vec^*(\Sigma)|\Sigma_0, G_0)\left(\Phi\left(\frac{\xi}{\sqrt{2}\sigma_d}\right) - \Phi\left(-\frac{\xi}{\sqrt{2}\sigma_d}\right)\right)}, 1\right\},$$

where $\Phi(\cdot)$ is the standard normal cumulative distribution function. Compared to the algorithm of Chib and Greenburg (1998), our Metropolized hit-and-run algorithm is more preferable since the use of the hit-and-run algorithm ensures the candidate correlation matrix Σ^* to be non-negative no matter which σ_d^2 is chosen.

3.2.2. *Multivariate t-Link models.* Using (2.6), we have

$$\pi(w_i|\nu, x_i\beta, \Sigma) = \int_0^\infty \left[\left(\tfrac{\lambda_i}{2\pi}\right)^{J/2} |\Sigma|^{-1/2} \exp\left\{-\tfrac{\lambda_i}{2}(w_i - x_i\beta)'\Sigma^{-1}(w_i - x_i\beta)\right\}\right] \cdot \tfrac{1}{\Gamma\left(\frac{\nu}{2}\right)} \left(\tfrac{\nu}{2}\right)^{\nu/2} \lambda_i^{\frac{\nu}{2}-1} e^{-\frac{\nu}{2}\lambda_i} d\lambda_i. \quad \ldots (3.3)$$

To run the Gibbs sampler, we sample λ_i, w_i, β, and Σ from their respective conditional distributions. Using (3.3), we independently generate

$$\lambda_i \mid w_i, \beta, \Sigma \sim \mathcal{G}\left(\frac{\nu + J}{2}, \frac{1}{2}\left[\nu + (w_i - x_i\beta)'\Sigma^{-1}(w_i - x_i\beta)\right]\right),$$

where $\mathcal{G}(u, v)$ denotes a gamma distribution with density $\pi_{\mathcal{G}}(\lambda|u, v) \propto \lambda^{u-1}e^{-v\xi}$. In a manner similar to the MVP models with obvious adjustments, we can generate w_i, β, and Σ given the λ_i for the MVT models. For example, given λ_i, w_i, and Σ, we have

$$\beta \mid \lambda_i, w_i, i = 1, 2, \ldots, n, \Sigma, y \sim N\left(\hat{\beta}_t, B_t^{-1}\right),$$

where $\hat{\beta}_t = B_t^{-1}(B_0\beta_0 + \sum_{i=1}^n \lambda_i x_i'\Sigma^{-1}w_i)$ and $B_t = B_0 + \sum_{i=1}^n \lambda_i x_i'\Sigma^{-1}x_i$.

3.2.3. *Multivariate logit models.* Similar to (2.8), we have

$$\pi(w_i|x_i\beta, \Sigma) = \int_0^\infty \left(\frac{1}{8\pi\lambda_i^2}\right)^{J/2} |\Sigma|^{-1/2} \exp\left\{-\frac{1}{8\lambda_i^2}(w_i - x_i\beta)'\Sigma^{-1}(w_i - x_i\beta)\right\} \pi_K(\lambda_i)d\lambda_i, \quad \ldots (3.4)$$

where $\pi_K(\lambda_i)$ is given in (2.7).

To implement the Gibbs sampler, we draw λ_i, w_i, β, and Σ from their respective conditional distributions. Similar to the MVT-link models, with obvious adjustments we can draw w_i, β, and Σ given the λ_i for the MVL models. For example, given λ_i, w_i, and Σ, we take

$$\beta \mid \lambda_i, w_i, i = 1, 2, \ldots, n, \Sigma, y \sim N\left(\hat{\beta}_t, B_t^{-1}\right),$$

where $\hat{\beta}_t = B_t^{-1}(B_0\beta_0 + \sum_{i=1}^n \left(4\lambda_i^2\right)^{-1} x_i'\Sigma^{-1}w_i)$ and $B_t = B_0 + \sum_{i=1}^n \left(4\lambda_i^2\right)^{-1} x_i'\Sigma^{-1}x_i$. However, to draw λ_i given w_i, β, and Σ, we use the Metropolis sampling scheme.

Let λ_i be the current value. Using (2.11) and (3.4), we independently generate

$$\lambda^{*2}_i \sim \mathcal{IG}\left(\frac{J+\nu}{2}, \frac{1}{8}\left[(w_i - x_i\beta)'\Sigma^{-1}(w_i - x_i\beta) + \frac{\nu}{b^2}\right]\right). \qquad \ldots (3.5)$$

Then, a move to the proposal point λ_i^* is made with probability

$$\min\left\{\frac{\pi_K(\lambda_i^*)/g_L(\lambda_i^*|\nu, b)}{\pi_K(\lambda_i)/g_L(\lambda_i|\nu, b)}, 1\right\}, \qquad \ldots (3.6)$$

where $\pi_K(\lambda)$ and $g_L(\lambda_i|\nu, b)$ are given in (2.7) and (2.12) respectively and ν and b are chosen in Section 2.3.

REMARK 2. In our Metropolis step, the full proposal density is

$$c_L^*\left(\frac{1}{8\pi\lambda_i^2}\right)^{J/2}|\Sigma|^{-1/2}\exp\left\{-\frac{1}{8\lambda_i^2}(w_i - x_i\beta)'\Sigma^{-1}(w_i - x_i\beta)\right\}g_L(\lambda_i|\nu, b),$$

where c_L^* is a normalizing constant and in (3.6) the common term

$$c_L^*\left(\frac{1}{8\pi\lambda_i^2}\right)^{J/2}|\Sigma|^{-1/2}\exp\left\{-\frac{1}{8\lambda_i^2}(w_i - x_i\beta)'\Sigma^{-1}(w_i - x_i\beta)\right\}$$

was canceled out in calculation. We also note that due to the introduction of the proposal density $g_L(\lambda_i|\nu, b)$, the Metropolis step is straightforward.

3.2.4. *Symmetric multivariate stable link models.* For the MVS-link models, we have

$$\pi(w_i|x_i\beta, \Sigma, \alpha)$$

$$= \int_0^\infty \left(\frac{1}{4\pi\lambda_i}\right)^{J/2}|\Sigma|^{-1/2}\exp\left\{-\frac{1}{4\lambda_i}(w_i - x_i\beta)'\Sigma^{-1}(w_i - x_i\beta)\right\}\pi_{S^P}(\lambda_i|\alpha, 1)d\lambda_i, \qquad \ldots (3.7)$$

where $\pi_{S^P}(\lambda_i|\alpha, 1)$ is defined in (2.13) and $1/2 \le \alpha < 1$. In an almost similar manner with an obvious adjustment, we can draw w_i, β, and Σ from their respective conditionals distributions in the implementation of the Gibbs sampler. To draw λ_i from its conditional distribution $[\lambda_i \mid w_i, x_i\beta, \Sigma, \alpha]$, Choy (1995) proposed the Sampling-Resampling (SR) method and the generalized Ratio-of-Uniform (ROU) algorithm, which requires use of (2.14). Denote $N_J(w_i|x_i\beta, 2\lambda_i\Sigma)$ to be the pdf of the multivariate normal distribution with $x_i\beta$ and $2\lambda_i\Sigma$ as its mean and variance-covariance matrix. Then, the SR method can be implemented as follows:

Step 1. Generate m independently proposed points $\lambda_{i(1)}$, $\lambda_{i(2)}$, $\ldots$, $\lambda_{i(m)}$ from $\pi_{S^P}(\lambda_i|\alpha, 1)$.

Step 2. Calculate the standardized weights

$$\omega_k^* = \frac{N_J(w_i|x_i\beta, 2\lambda_{i(k)}\Sigma)}{\sum_{j=1}^m N_J(w_i|x_i\beta, 2\lambda_{i(j)}\Sigma)}, \quad k = 1, 2, \ldots, m. \qquad \ldots(3.8)$$

Step 3. Generate a discrete approximate realization λ_i^* from $(\lambda_{i(1)}, \lambda_{i(2)}, \ldots, \lambda_{i(m)})$ with probability $(\omega_1^*, \omega_2^*, \ldots, \omega_m^*)$.

As Choy (1995) pointed out, under mild regularity conditions,

$$\lambda_i^* \quad \xrightarrow{\mathcal{D}} \quad [\lambda_i \mid w_i, x_i\beta, \Sigma, \alpha] \ \text{ as } \ m \to \infty. \qquad \ldots(3.9)$$

Note that to draw λ from $\pi_{S^P}(\lambda|\alpha, 1)$ $(0 < \alpha < 1)$, we need the following steps:

Step 1. Generate $U \sim U(0,1)$ and $E \sim \mathcal{E}(1)$ where $\mathcal{E}(1)$ denotes an exponential distribution with mean 1.

Step 2. Calculate

$$\lambda = \frac{\sin(\alpha\pi U)}{\sin^{1/\alpha}(\pi U)} \left(\frac{\sin((1-\alpha)\pi U)}{E}\right)^{\frac{1-\alpha}{\alpha}}.$$

(See, e.g., Chambers, Mallows, and Stuck, 1976 or Choy, 1995 for details.) The SR method is straightforward to implement. However, as pointed by Choy (1995), there is a hidden danger in using the SR method since the full conditional distribution of λ_i might be very different from the stable $\pi_{S^P}(\lambda|\alpha, 1)$ distribution which generates the proposed sample. For such cases, the convergence of (3.9) may be slow and therefore, a large value of m is needed, which makes the SR algorithm inefficient. To obtain a more efficient sampling scheme, we can use the following generalized version of the ROU algorithm of Wakefield, Gelfand and Smith (1991). For our MVS-link models, using (2.14), the conditional distribution of λ_i given u_i, w_i, $x_i\beta$, Σ and α is

$$[\lambda_i \mid u_i, w_i, x_i\beta, \Sigma, \alpha] \quad \propto \quad N_J(w_i|x_i\beta, 2\lambda_i\Sigma)\lambda_i^{-(a+1)} \exp\left\{-s(u_i)\lambda_i^{-a}\right\}, \ \lambda_i > 0, \qquad \ldots(3.10)$$

where $a = \frac{\alpha}{1-\alpha}$ and the conditional distribution of u_i given λ_i and α is

$$[u_i \mid \lambda_i, \alpha] \quad \propto \quad s(u_i) \exp\left\{-s(u_i)\lambda_i^{-a}\right\}, \ 0 < u_i < 1. \qquad \ldots(3.11)$$

To generate λ_i from (3.10), we use the following two steps:

Step 1. Draw ψ_i from

$$[\psi_i \mid u_i, w_i, x_i\beta, \Sigma, \alpha]$$

$$\propto \exp\left\{-\left[(J/2 + a)\psi_i + \frac{(w_i - x_i\beta)'\Sigma^{-1}(w_i - x_i\beta)}{4} e^{-\psi_i} + s(u_i)e^{-a\psi_i}\right]\right\} \qquad \ldots(3.12)$$

by using the fast adaptive rejection sampling algorithm of Gilks and Wild (1992) since the conditional distribution of ψ_i is log-concave.

Step 2. Calculate $\lambda_i = \exp(\psi_i)$.

The generation of u_i from (3.11) can be implemented by the following steps:

Step 1. Draw ϕ_i from the conditional distribution

$$[\phi_i \mid \lambda_i, \alpha] \quad \propto \quad \frac{u_i s\left(e^{\phi_i}/(1+e^{\phi_i})\right)}{1+e^{\phi_i}} \exp\left\{-s\left(e^{\phi_i}/(1+e^{\phi_i})\right)\lambda_i^{-a}\right\} \qquad \ldots (3.13)$$

by using the ROU algorithm.

Step 2. Calculate $u_i = \frac{e^{\phi_i}}{1+e^{\phi_i}}$.

The ROU algorithm is much more efficient than the SR method. However, the ROU algorithm requires the numerical maximizations, making the implementation very expensive. Alternatively to the ROU algorithm, we propose a Metropolis method with an inverse gamma proposal distribution. Letting λ_i be the current value, we draw

$$\lambda_i^* \quad \sim \quad \mathcal{IG}\left(\frac{J+1}{2}, \frac{1}{4}\left[(w_i - x_i\beta)'\Sigma^{-1}(w_i - x_i\beta) + 1\right]\right). \qquad \ldots (3.14)$$

Then, a move to the proposal point λ_i^* is made with probability

$$\min\left\{\frac{\pi_{S^P}(\lambda_i^*|\alpha,1)/\pi_{\mathcal{IG}}(\lambda_i^*|1/2,1/4)}{\pi_{S^P}(\lambda_i|\alpha,1)/\pi_{\mathcal{IG}}(\lambda_i|1/2,1/4)}, 1\right\}. \qquad \ldots (3.15)$$

Remark 3. In the above Metropolis scheme, the full proposal density is proportional to

$$\left(\frac{1}{4\pi\lambda_i}\right)^{J/2} |\Sigma|^{-1/2} \exp\left\{-\frac{1}{4\lambda_i}(w_i - x_i\beta)'\Sigma^{-1}(w_i - x_i\beta)\right\} \pi_{\mathcal{IG}}(\lambda_i|1/2,1/4) \qquad \ldots (3.16)$$

and using Appendix B it is easy to show that the proposal distribution given in (3.16) has heavier tails than the conditional distribution of λ_i. This proposal density works well when α is not far away from 1/2. To monitor how this proposal density performs, one may check the acceptance probability. Our empirical study indicates that the acceptance probabilities are greater than 0.70 for $\alpha \leq .75$. (See Section 4 for an illustrative example.) Note that (3.15) requires evaluating one-dimensional integrals. Since one-dimensional numerical integration method is well-developed and computationally faster, one can use, for example, IMSL subroutine QDAG or QDAGI to achieve this. This Metropolis scheme makes the implementation of the Gibbs sampler for the MVS-link models much simpler than the ROU does and it is expected to perform well when α is not too close to 1.

3.2.5. *Symmetric multivariate exponential power link models.* For the MVEP-link models, we have

$$\pi(w_i|x_i\beta,\Sigma,\alpha)=\int_0^\infty\left(\frac{c_0\lambda_i}{\pi}\right)^{J/2}|\Sigma|^{-1/2}\exp\left\{-c_0\lambda_i(w_i-x_i\beta)'\Sigma^{-1}(w_i-x_i\beta)\right\}$$
$$\left(\frac{1}{\lambda_i}\right)^{J/2}\pi_{S^P}(\lambda_i|\alpha,1)d\lambda_i, \qquad \ldots(3.17)$$

where c_0 is defined in (2.16). To run the Gibbs sampler, similar to the MVS-link models, we take λ_i, w_i, β, and Σ from their respective conditional distributions for the MVEP-link models. For example, to implement the SR method for generating λ_i from its conditional distribution, instead of (3.8), we calculate the standardized weights as follows:

$$\omega_k^*=\frac{\lambda_{i(k)}^{-J/2}N_J(w_i|x_i\beta,(2c_0\lambda_{i(k)})^{-1}\Sigma)}{\sum_{j=1}^m\lambda_{i(j)}^{-J/2}N_J(w_i|x_i\beta,(2c_0\lambda_{i(j)})^{-1}\Sigma)},\quad k=1,2,\ldots,m.$$

For the generalized version of the ROU algorithm, instead of (3.12), we generate ψ_i from the conditional distribution

$$[\psi_i \mid u_i,w_i,x_i\beta,\Sigma,\alpha]$$
$$\propto\exp\left\{-\left[a\psi_i+c_0(w_i-x_i\beta)'\Sigma^{-1}(w_i-x_i\beta)e^{\psi_i}+s(u_i)e^{-a\psi_i}\right]\right\},$$

where $a=\frac{\alpha}{1-\alpha}$ and c_0 is given in (2.16). Then we take $\lambda_i=\exp(\psi_i)$.

Similar to the MVS-link models, we propose a Metropolis algorithm. Instead of using an inverse gamma proposal, we use an inverse Gaussian proposal. Letting λ_i be the current value, we draw

$$\lambda_i^* \sim \mathcal{IN}(\mu_{EP}^*,\sigma_{EP}^*), \qquad \ldots(3.18)$$

where

$$\mu_{EP}^*=\left(4c_0(w_i-x_i\beta)'\Sigma^{-1}(w_i-x_i\beta)\right)^{-1/2}\ \text{and}\ \sigma_{EP}^*=\frac{1}{2},$$

and the density of $\mathcal{IN}(\mu^*,\sigma^*)$ is

$$\pi_{\mathcal{IN}}(\lambda|\mu^*,\sigma^*)=\sqrt{\frac{\sigma^*}{2\pi\lambda^3}}\exp\left\{-\frac{\sigma^*(\lambda-\mu^*)^2}{2\mu^{*2}\lambda}\right\},\quad \text{for}\quad \lambda>0,$$

with parameters $\mu^*>0$ and $\sigma^*>0$. Then, a move to the proposal point λ_i^* is made with probability

$$\min\left\{\frac{\pi_{S^P}(\lambda_i^*|\alpha,1)/\pi_{\mathcal{IG}}(\lambda_i^*|1/2,1/4)}{\pi_{S^P}(\lambda_i|\alpha,1)/\pi_{\mathcal{IG}}(\lambda_i|1/2,1/4)},1\right\}. \qquad \ldots(3.19)$$

Remark 4. In the Metropolis algorithm, the full proposal density is proportional to

$$\left(\frac{c_0\lambda_i}{\pi}\right)^{J/2} |\Sigma|^{-1/2} \exp\left\{-c_0\lambda_i(w_i - x_i\beta)'\Sigma^{-1}(w_i - x_i\beta)\right\} \left(\frac{1}{\lambda_i}\right)^{J/2} \pi_{\mathcal{IG}}(\lambda_i|1/2, 1/4),$$

which exactly matches with an inverse Gaussian distribution.

Remark 5. To generate $\lambda \sim \mathcal{IN}(\mu^*, \sigma^*)$, we simply use the following steps:

Step 1. Generate $Z \sim N(0,1)$, set $\xi = Z^2$, and calculate $\lambda^* = \mu^* + \frac{\mu^{*2}\xi}{2\sigma^*} - \frac{\mu^*}{2\sigma^*}\sqrt{4\mu^*\sigma^*\xi + \mu^{*2}\xi^2}$.

Step 2. Generate $U \sim U(0,1)$ and let $\lambda = \lambda^*$ if $U \le \frac{\mu^*}{\mu^*+\lambda^*}$ and $\lambda = \frac{\mu^{*2}}{\lambda^*}$ if otherwise.

See Devroye (1986, p 148) for the details.

Finally, it is interesting to mention that calculation of the acceptance probability given in (3.19) for the MVEP-Link models is exactly the same as the one for the MVS-Link models.

4. An Illustrative Example: Respiratory Infection of Indonesian Children

To apply and illustrate our methodology, we consider the data on respiratory infection in Indonesian preschool children. One hundred and twenty two preschool children in Indonesia were examined for upto six consecutive quarters for the respiratory infection. We examine a subset of a cohort study by Sommer (1982) ignoring the missing data and covariates to form our subset. The data is also described along with covariates in Diggle, Liang, and Zeger (1994). In our analysis we consider gender, height for age, seasonal cosine and sine, presence of Xerophthalmia (vitamin A deficiency), age and age square. In order to help the numerical stability in the implementation of the Gibbs sampler, we standardize all of the covariates.

We consider six SMMVN-link models to fit the data on respiratory infection in Indonesian preschool children. These models are the MVP, MVEP with $\alpha = .75$, MVL, MVT with $\nu = 8$, MVS with $\alpha = .75$ and MVC (i.e., MVT with $\nu = 1$). For ease of notation, we call MVEP with $\alpha = .75$, MVT with $\nu = 8$, and MVS with $\alpha = .75$ simply as MVEP, MVT and MVS thereafter. These models capture different aspects and features of the SMMVN-link models. For example, the MVP and the MVC correspond to the lightest and the heaviest tails, respectively. In the implementation of the Gibbs sampler, we use $g_L(\lambda|5, .712)$ for the MVL model to implement the Metropolis step, which results in an acceptance probability of approximately 80%. We also use the Metropolis algorithms proposed in Section 3.2 for both the MVEP and MVS models and the acceptance

Table 1. Bayesian posterior estimates

Model	Covariates	Mean	Std Dev	Median	95% HPD Interval
MVP	gender	-.049	.155	-.049	(-.350, .253)
	height for age	-.071	.080	-.069	(-.226, .082)
	seasonal cosine	-.393	.103	-.391	(-.597, -.190)
	seasonal sine	-.029	.109	-.027	(-.247, .180)
	Xerophtalmia	.312	.366	.322	(-.383, 1.04)
	age	-.218	.094	-.217	(-.400, -.035)
	age^2	-.195	.089	-.194	(-.374, -.029)
MVEP	gender	-.070	.194	-.070	(-.459, .306)
	height for age	-.089	.101	-.088	(-.293, .100)
	seasonal cosine	-.521	.135	-.520	(-.777, -.254)
	seasonal sine	-.032	.141	-.032	(-.298, .254)
	Xerophtalmia	.406	.461	.422	(-.476, 1.30)
	age	-.286	.120	-.285	(-.520, -.052)
	age^2	-.266	.120	-.260	(-.513, -.040)
MVL	gender	-.153	.314	-.150	(-.808, .422)
	height for age	-.136	.164	-.130	(-.458, .186)
	seasonal cosine	-.822	.223	-.818	(-1.25, -.381)
	seasonal sine	-.065	.227	-.059	(-.485, .398)
	Xerophtalmia	.614	.728	.659	(-.881, 1.97)
	age	-.486	.208	-.480	(-.904, -.079)
	age^2	-.473	.201	-.467	(-.895, -.106)
MVT	gender	-.096	.195	-.093	(-.504, .276)
	height for age	-.078	.099	-.076	(-.274, .116)
	seasonal cosine	-.513	.139	-.510	(-.792, -.248)
	seasonal sine	-.042	.138	-.039	(-.314, .226)
	Xerophtalmia	.379	.463	.412	(-.555, 1.25)
	age	-.302	.136	-.298	(-.582, -.048)
	age^2	-.298	.134	-.291	(-.566, -.047)
MVS	gender	-.305	.368	-.289	(-1.06, .385)
	height for age	-.114	.183	-.111	(-.462, .247)
	seasonal cosine	-1.05	.314	-1.02	(-1.72, -.495)
	seasonal sine	-.064	.306	-.046	(-.712, .520)
	Xerophtalmia	.590	.956	.713	(-1.40, 2.22)
	age	-.895	.369	-.869	(-1.63, -.177)
	age^2	-.906	.349	-.883	(-1.58, -.237)
MVC	gender	-.838	.651	-.779	(-2.17, .478)
	height for age	-.368	.330	-.355	(-1.01, .261)
	seasonal cosine	-2.22	.750	-2.08	(-3.85, -.983)
	seasonal sine	-.050	.623	-.042	(-1.27, 1.18)
	Xerophtalmia	.149	2.20	.572	(-4.49, 3.76)
	age	-2.29	.739	-2.26	(-3.69, -.799)
	age^2	-2.25	.731	-2.22	(-3.67, -.937)

probability, for example, for the MVS model is approximately 73%. We check the convergence of the Gibbs sampler using several diagnostic procedures as recommended by Cowles and Carlin (1996) and after convergence, we generate a large number of Gibbs iterates for further various Bayesian calculations.

Using 100,000 Gibbs iterates after convergence, we compute the posterior estimates and 95% highest posterior density (HPD) intervals for all six models. Note that an easily used Monte Carlo method developed by Chen and Shao (1998) is used for computing HPD intervals. The results are presented in Table 1. Although we use a subset of the data and the different scales of the covariates than Diggle, Liang and Zeger (1994), the signs and significance of the coefficients for all covariates are consistent with those obtained by Diggle *et, al.* (1994). We also mention that the posterior estimate of the correlation matrix indicates that the equicorrelation assumption on the correlation structure is questionable. For example, the 95% HPD intervals for ρ_{41} are (.090, .816) and (.016, .786) for the MVP and MVL models, respectively, while the 95% HPD intervals for all other ρ_{ij}'s contain zero. These results imply that there is a significant positive correlation between the fourth quarter visit and the first quarter visit.

5. **Concluding Remarks**

Correlated binary data often arise in experiments when two or more measurements are taken at one time for the same subjects or when repeated measurements are taken over time. If such correlation is ignored in the model, overstatement of the precision of parameter estimates results. We have considered a unified approach in this paper to incorporate the correlation structure, using the notion of multivariate generalized linear models.

Our proposed modeling approach is based on multivariate link functions using a very rich class of scale mixtures of normals. Such models are very flexible and include all the standard link functions in a generalized linear model scenario. In addition our modeling approach gives more insight of logit, stable distribution and exponential power distribution family links. Our illustrative example empirically demonstrates that our proposed modeling approach and various computational algorithms work well.

There is another advantage of considering our approach over the usual random effects model which is based on the assumption of exchangeability. This is clearly reflected in our data analysis which shows that the equicorrelation assumption is not valid. In addition other advantages of Bayesian modeling over classical approach prevail in our studies. This includes more precise influence, exact small sample analysis, incorporation of the prior information, and inclusion of a large number of covariates.

Appendix A

Choices of ν and b for $g_L(\nu, b)$ and evaluation of $\pi_K(\lambda)$ for the MVL models. First, we discuss how to choose ν and b. By matching the quantiles of a t_ν distribution and the logistic distribution π_L, Albert and Chib (1993) obtained

$\nu = 8$ and $b = .634$. We determine the values of ν and b by matching the moments of two distributions. We obtain $\nu = 9$ and $b = .625$ by matching the first four moments of a t_ν distribution (assuming $\nu > 4$) and the logistic distributions, we also obtain various values of b with fixed ν by matching the second moments of these two distributions and these b's are .637, .652, .712 for $\nu = 8, 7, 5$ respectively. Further, we obtain $\nu = 7.581$ and $b = .643$ by matching the first two moments of the distribution $g_L(\lambda|\nu, b)$ and the asymptotic Kolmogorov distribution $\pi_K(\lambda)$ and surprisingly, these values of ν and b are comparable to those given by Albert and Chib (1993).

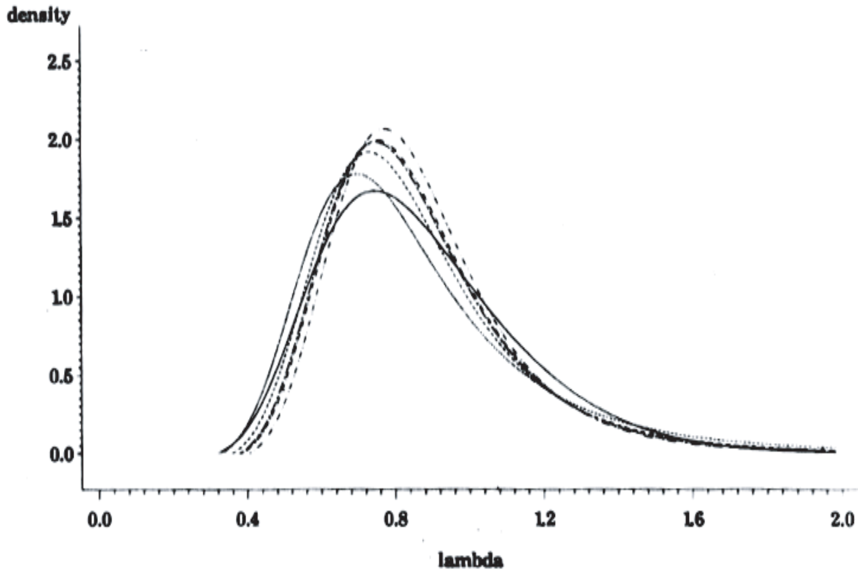

Figure 1. Density Curves of $\pi_K(\lambda)$ and $g_L(\lambda|\nu, b)$. The solid curve is $\pi_K(\lambda)$; the dotted curve is $g_L(\lambda|5, .712)$; the dashed curve is $g_L(\lambda|7, .652)$; the long-dashed curve is $g_L(\lambda|7.581, .643)$; the dot-dashed curve is $g_L(\lambda|8, .637)$; the dot-dot-dashed curve is $g_L(\lambda|9, .625)$.

The density function curves of $\pi_K(\lambda)$ and $g_L(\lambda|\nu, b)$ with these choices of ν and b are displayed in Figure 1. Note that we used 100,000 terms to calculate π_K. From Figure 1, $g_L(\lambda|7.581, .643)$ and $g_L(\lambda|8, .637)$ are very close to each other, the mode of $g_L(\lambda|5, .712)$ (the mode of $g_L(\lambda|9, .625)$) is a little bit smaller (greater) than that of $\pi_K(\lambda)$, which implies that any other g_L having degrees of freedom smaller than 5 or greater than 9 will be farther apart away from $\pi_K(\lambda)$, and all five $g_L(\lambda|\nu, b)$'s matches $\pi_K(\lambda)$ fairly well. However, In Section 3, we use $g_L(\lambda|\nu, b)$ to construct a proposal density for the Metropolis steps in the Gibbs sampler. For such purposes, an ideal $g_L(\lambda|\nu, b)$ ought to have a shape similar to $\pi_K(\lambda)$ as well as heavier tails than those of $\pi_K(\lambda)$. Thus, $g_L(\lambda|5, .712)$ will be the best choice, which can be immediately observed from the following theorem.

THEOREM A.1 *Let $w \sim \pi_{t_\nu}$ and $w^* \sim \pi_L$. Then, we have*

$$\sup_\nu \left\{ \nu \geq 3: \quad Var(w) = Var(bw^*), \ \lim_{\lambda \to 0} \frac{\pi_K(\lambda)}{g_L(\lambda|\nu, b)} < \infty, \right.$$

and

$$\left. \lim_{\lambda \to \infty} \frac{\pi_K(\lambda)}{g_L(\lambda|\nu, b)} < \infty \right\} = 5. \qquad \ldots (A.1)$$

With $\nu = 5$, $Var(w) = Var(bw^*)$ leads to $b^2 = 5/\pi^2$ and henceforth, $b = .712$. Furthermore, we have

$$\lim_{\lambda \to \infty} \frac{\pi_K(\lambda)}{g_L(\lambda|5, .712)} = 0 \qquad \ldots (A.2)$$

and

$$\lim_{\lambda \to 0} \frac{\pi_K(\lambda)}{g_L(\lambda|5, .712)} = 0. \qquad \ldots (A.3)$$

PROOF. From (2.7), it can be easily seen that for $\nu \geq 3$, (A.2) is always true and therefore, $\lim_{\lambda \to \infty} \frac{\pi_K(\lambda)}{g_L(\lambda|\nu,b)} < \infty$ automatically holds. Note that if $\nu \geq 3$, $Var(w) = Var(bw^*)$ leads to $\frac{\nu}{b^2} = \frac{\nu-2}{3}\pi^2$. Using an alternative expression of $\pi_L(\lambda)$ by the theory of theta functions (see, for example, Whittaker and Watson 1927), which is given by

$$\pi_L(\lambda) = \frac{\sqrt{2\pi}}{\lambda} \sum_{k=1}^{\infty} \left[\frac{(2k-1)^2\pi^2}{4\lambda^3} - \frac{1}{\lambda} \right] \exp\left\{ -\frac{(2k-1)^2\pi^2}{8\lambda^2} \right\}, \lambda > 0, \quad \ldots (A.4)$$

we have that the necessary and sufficient conditions for $\lim_{\lambda \to 0} \frac{\pi_K(\lambda)}{g_L(\lambda|\nu,b)} < \infty$ is $3 \leq \nu \leq 5$. Furthermore, with $\nu = 5$ it is easy to show that (A.3) is true. Therefore, we complete the proof of the theorem. □

Second, we discuss how to efficiently evaluate the infinite series of $\pi_L(\lambda)$. We call $\pi_L(\lambda)$ defined by (2.7) and (A.4) as the first series and second series respectively. As discussed by Devroye (1986, pp 161-162), both can be written as alternative series of the form

$$\pi_L(\lambda) = ch(\lambda) \sum_{k=0}^{\infty} (-1)^k a_k(\lambda). \qquad \ldots (A.5)$$

In (A.5), for the first series we take

$$\begin{aligned} ch(\lambda) &= 8\lambda \exp\left\{-2\lambda^2\right\}, \ \lambda > 0, \\ a_k(\lambda) &= (k+1)^2 \exp\left\{-2\lambda^2((k+1)^2 - 1)\right\}, k \geq 0 \text{ and } \lambda > 0, \end{aligned} \qquad \ldots (A.6)$$

and for the second series we take

$$ch(\lambda) = \frac{\sqrt{2\pi}\pi^2}{4\lambda^4}\exp\left\{-\frac{\pi^2}{8\lambda^2}\right\},$$

$$a_k(\lambda) = \begin{cases} \frac{4\lambda^2}{\pi^2}\exp\left\{-\frac{(k^2-1)\pi^2}{8\lambda^2}\right\}, & k \text{ odd}, \lambda > 0 \\ (k+1)^2\exp\left\{-\frac{((k+1)^2-1)\pi^2}{8\lambda^2}\right\}, & k \text{ even}, \lambda > 0 \end{cases} . \quad \ldots (A.7)$$

Then, Devroye (1986, p 162) showed that for the first series the terms $a_k(\lambda)$ given in (A.6) are monotonically decreasing for $\lambda > \sqrt{1/3}$ and for the second series the terms $a_k(\lambda)$ given in (A.7) are monotonically decreasing for $\lambda < \pi/2$. As suggested by Devroye (1986), we use the first series expansion to evaluate $\pi_K(\lambda)$ when $\lambda > .75$ and the second series when $\lambda \leq .75$. Furthermore, for a given numerical precision $0 < \delta < 1$ the number of terms ($K^*(\lambda)$) required for evaluating $\pi_K(\lambda)$ can be determined by

$$K^*(\lambda) = \inf_k \{k : \ ch(\lambda)a_k(\lambda) \leq \delta\}.$$

Appendix B

Positive stable distribution $S^P(\alpha, 1)$. Choy (1995) pointed out that when $\alpha = 1/2$, $S^P(\frac{1}{2}, 1)$ is the $\mathcal{IG}(1/2, 1/4)$, that is,

$$\pi_{S^P}(\lambda|1/2, 1) = \frac{(1/4)^{1/2}}{\Gamma(1/2)\lambda^{1/2+1}}\exp\left\{-\frac{1}{4\lambda}\right\}.$$

Figure 2 displays the density curves for various values of α for $1/2 \leq \alpha < 1$.

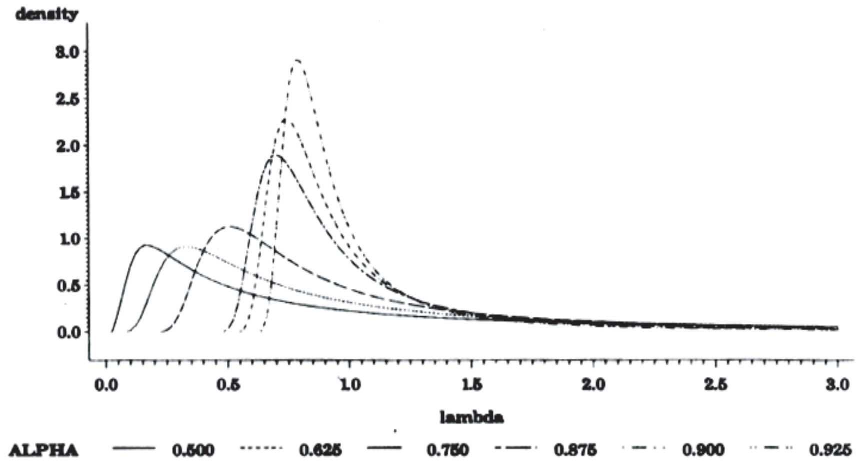

Figure 2: Density Curves of $S^P(\alpha, 1)$ for Various Values of α

From Figure 2, we can observe that (i) the tails of $S^P(\alpha, 1)$ are getting heavier when α gets smaller and (ii) the density curves of $S^P(\alpha, 1)$ are shifted to the right when α becomes larger. Let $\pi_{\mathcal{IG}}(\lambda|1/2, 1/4)$ denote the pdf of $\mathcal{IG}(1/2, 1/4)$. Then, using (2.13) it is easy to show that

$$\lim_{\lambda\to 0} \frac{\pi_{SP}(\lambda|1/2,1)}{\pi_{\mathcal{IG}}(\lambda|1/2,1/4)} = \lim_{\lambda\to\infty} \frac{\pi_{SP}(\lambda|1/2,1)}{\pi_{\mathcal{IG}}(\lambda|1/2,1/4)} = 0, \quad \text{for } 1/2 \le \alpha < 1. \qquad \ldots(B.1)$$

Acknowledgement. The authors wish to thank Charles B. Hall for providing the data set.

References

Albert, J.H. and Chib, S. (1993). Bayesian analysis of Binary and polychotomous response data, *Jour. Amer. Statist. Assoc.,* **88**, 669-679.

Amemiya, T. (1985). *Advanced Econometrics*, Harvard University Press, Boston.

Andrews, D.F. and Mallows, C.L. (1974). Scale mixtures of normal distributions, *Jour. Royal Statist. Soc.,* Series B, **36**, 99-102.

Ashford, J.R. and Sowden, R.R. (1970). Multivariate probit analysis, *Biometrics,* **26**, 535-546.

Box, G.E.P. and Tiao, G.C. (1992). *Bayesian Inference in Statistical Analysis*, Wiley: New York.

Buckle, D.J. (1995). Bayesian inference for stable distributions, *Jour. Amer. Statist. Assoc.,* **90**, 605-613.

Chambers, J.M., Mallows, C.L., and Stuck, B.W. (1976). A method for simulating stable random variables, *Jour. Amer. Statist. Assoc.* **71**, 340-344.

Chen, M.-H. and Schmeiser, B.W. (1993). Performance of the Gibbs, hit-and-run, and metropolis samplers, *The Journal of Computational and Graphical Statistics,* **2**, 251-272.

Chen, M.-H. and Shao, Q.-M. (1998). Monte Carlo estimation of Bayesian credible and HPD intervals, *The Journal of Computational and Graphical Statistics,* **7**, to appear.

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

Cowles, M.K. and Carlin, B.P. (1996). Markov Chain Monte Carlo convergence diagnostics: A comparative review, *Jour. Amer. Statist. Assoc.,* **91**, 883-904.

Choy, S.T.B. (1995). Robust Bayesian analysis using scale mixture of normals distributions, *Ph.D. Dissertation*, Department of Mathematics, Imperial College, London.

Dey, D.K. and Chen, M.-H. (1996). Bayesian analysis of correlated binary data models, *Technical Report 96-02*, Department of Statistics, University of Connecticut.

Devroye, L. (1986). *Non-Uniform Random Variate Generation*, Springer-Verlag: New York.

Diggle, P.J., Liang, K.-Y., and Zeger, S.L. (1994). *Analysis of Longitudinal Data*, Clearendon Press: Oxford.

Gelfand, A. E. and Smith, A.F.M. (1990). Sampling based approaches to calculating marginal densities, *Jour. Amer. Statist. Assoc.,* **85**, 398-409.

Geman, S. and Geman, D. (1984). Stochastic Relaxation, Gibbs Distributions and the Bayesian Restoration of Images, *IEEE Transactions on Pattern Analysis and Machine Intelligence,* **6,** 721-741.

Geweke, J. (1991). Efficient simulation from the multivariate normal and student-t distributions subject to linear constraints, *Computing Science and Statistics: Proceedings of the Twenty-Third Symposium on the Interface*, 571-578.

Gilks, W.R. and Wild, P. (1992). Adaptive rejection sampling for Gibbs sampling, *Applied Statistics,* **41**, 337-348.

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

LIANG, K.-Y. AND ZEGER, S.L. (1986). Longitudinal data analysis using generalized linear Models, *Biometrika,* **73**, 13-22.

MARSAGLIA, G. AND OLKIN, I. (1984). Generating correlation matrices, *SIAM Jour. Scientific and Statistical Computations,* **5**, 470-475.

METROPOLIS, N., ROSENBLUTH, A.W., ROSENBLUTH, M.N., TELLER, A.H. AND TELLER, E. (1953). Equations of state calculations by fast computing machines, *Jour. chemical physics,* **21**, 1087-1092.

PRENTICE, R.L. (1988). Correlated Binary regression with covariate specific to each binary observation, *Biometrics,* **44**, 1033-1048.

ROUSSEEUW, P. AND MOLENBERGHS, G. (1994). The shape of correlation matrices, *American Statistician,* **48**, 276-279.

SOMMER, A. (1982). *Nutritional Blindness*, Oxford University Press: New York.

TIERNEY, L. (1994). Markov Chains for exploring posterior distributions (with discussions), *Ann. Statist.,* **22**, 1701-1762.

WAKEFIELD, J.C., GELFAND, A.E. AND SMITH, A.F.M. (1991). Efficient generation of random variates via the ratio-of-uniforms method, *Statistics and Computing,* **1**, 129-133.

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

WHITTAKER, E.T. AND WATSON, G.N. (1927). *A Course of Modern Analysis*, University Press: Cambridge.

ZEGER, S.L. AND LIANG, K.-Y. (1986). Longitudinal data analysis for discrete and continuous outcomes, *Biometrics,* **42**, 121-130.

MING-HUI CHEN
DEPARTMENT OF MATHEMATICAL SCIENCES
WORCESTER POLYTECHNIC INSTITUTE
100 INSTITUTE ROAD
WORCESTER, MA 01609
USA
e-mail : mhchen@wpi.edu

DIPAK K. DEY
DEPARTMENT OF STATISTICS
UNIVERSITY OF CONNECTICUT
BOX U-120, STORRS, CT 06269
USA
e-mail : dey@stat.uconn.edu