Brazilian Journal of Probability and Statistics (2000), 14, pp. 87–111.

A UNIFIED BAYESIAN APPROACH FOR ANALYZING CORRELATED ORDINAL RESPONSE DATA

Ming-Hui Chen
Department of Mathematical Sciences
Worcester Polytechnic Institute
100 Institute Road
Worcester - MA 01609, USA

Dipak K. Dey
Department of Statistics
University of Connecticut
BOX U-120
Storrs, CT 06269, USA

Summary
Generalized linear models with scale mixture of multivariate normal (SMMVN) link functions are considered to model correlated ordinal response data. A unified Bayesian approach, which includes prior elicitation and model comparison, is proposed. In order to incorporate available prior information, we propose a class of informative prior distributions on model parameters. The propriety of the proposed informative prior is also examined in detail. Due to the complexity of SMMVN models, Markov chain Monte Carlo sampling is used to carry out all posterior computations. Finally, a real data example from prostate cancer studies is used to illustrate the proposed methodologies.

Key Words: Bayesian computation; hierarchical model; latent variables; Markov chain Monte Carlo; multivariate generalized linear models; scale mixture of multivariate normal links.

1 Introduction

It is the current interest to model and analyze correlated or repeated categorical response data as these types of data often arise in biometrics, pharmaceutical industries, and social sciences. The responses are naturally correlated when they are taken from the same individual or subject at one time or over time. For example, data obtained from surveys are inherently categorical, items in a questionnaire usually consist of two-to-five options (e.g., "disagree", "neutral" and "agree"), and the responses of the questionnaire are correlated since each respondent answers all questions in the survey.

Prentice (1988) provided a comprehensive review of various modeling 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 a Bayesian framework, Chib and Greenberg (1998) used the multivariate probit (MVP) model for correlated binary data, while Chen and Dey (1998) considered general scale mixture of multivariate normal (SMMVN) link functions for longitudinal binary responses.

Unlike the correlated binary response data, the literature is still sparse in modeling and analyzing correlated or repeated ordinal data. A correlated ordinal data problem is not a simple generalization of the one for the correlated binary data. In general, an ordinal problem is harder than the binary one. To see this, we consider generalized linear models for independent categorical data models. It is well known that when the categorical response is binary, a generalized linear model reduces to a binomial model, which belongs to the exponential family. However, an appropriate model for ordinal response data is multinomial while a multinomial model is not a member of the exponential family. This may be one of the reasons why analyzing ordinal data problems is typically more difficult than the binary ones.

In the context of Bayesian analysis, Albert and Chib (1993) introduced latent variables into the generalized linear model to perform Bayesian computation using the Gibbs sampler. The primary goal for introducing the strategy involving latent variables is to simplify Bayesian computations. This goal was successfully achieved for independent binary probit models; see Albert and Chib (1993). However, even for a simple independent ordinal probit model, the Gibbs sampler may present challenging problems in achieving its convergence. To present a clear explanation, we let $Y = (Y_1, Y_2, \ldots, Y_n)'$ denote an $n \times 1$ vector of independent ordinal (1 through L) response random variables. Assume

$$P(Y_i \leq l) = \Phi(\gamma_l - x_i\beta), \text{ for } \ l = 1, 2, \cdots, L, \tag{1.1}$$

where Φ is the standard normal cumulative distribution function (cdf), $\beta = (\beta_1, \ldots, \beta_p)'$ is a $p \times 1$ vector of unknown regression coefficients, and $x_i = (x_{i1}, x_{i2}, \cdots, x_{ip})$ is a $1 \times p$ vector of the corresponding covariates for the i^{th} observation for $i = 1, 2, \ldots, n$. In (1.1) $-\infty = \gamma_0 \leq \gamma_1 = 0 \leq \gamma_2 \leq \cdots \leq \gamma_{L-1} \leq \gamma_L = \infty$ are cutpoints dividing the real line into L intervals. Then, we take an n-dimensional latent random vector $w = (w_1, w_i, \ldots, w_n)'$ such that

$$Y_i = l, \quad \text{if} \ \ \gamma_{l-1} \leq w_i < \gamma_l, \tag{1.2}$$

for $i = 1, 2, \ldots, n$. If a diffuse prior is specified for β and $\gamma = (\gamma_2, \cdots, \gamma_{L-1})'$, letting $y = (y_1, y_2, \ldots, y_n)$ be the observed data, the posterior distribution for β and γ is of the form

$$p(\beta, \gamma|y) \propto \prod_{i=1}^{n} \left[\Phi(\gamma_{y_i} - x_i\beta) - \Phi(\gamma_{y_i-1} - x_i\beta)\right].$$

Instead of directly sampling β and γ from $p(\beta, \gamma|y)$, Albert and Chib (1993) incorporated the unknown latent variables w as additional parameters to run the Gibbs sampler. It can be easily observed that the full conditional distributions for w and β are

$$w_i|\beta, \gamma, y_i \ \sim \ N(x_i\beta, 1) \tag{1.3}$$

conditional on $\gamma_{y_i-1} \le w_i \le \gamma_{y_i}$ and

$$\beta|w \ \sim \ N\left\{(X'X)^{-1}X'w, (X'X)^{-1}\right\}, \tag{1.4}$$

where $X = (x_1', \ldots, x_n')'$. Letting $\gamma_{(-l)} = (\gamma_2, \ldots, \gamma_{l-1}, \gamma_{l+1}, \ldots, \gamma_{L-1})'$, the conditional distribution of γ_l is

$$\gamma_l \mid \gamma_{(-l)}, w, y \ \sim \ U(a_l, b_l), \tag{1.5}$$

where $a_l = \max\{\max\{w_i : y_i = l\}, \gamma_{l-1}\}$, $b_l = \min\{\min\{w_i : y_i = l+1\}, \gamma_{l+1}\}$ and $l = 2, \cdots, L-1$. Albert and Chib (1993) implemented the Gibbs sampler by drawing w, β, and γ from (1.3), (1.4) and (1.5), which is straightforward. However, Cowles (1996) pointed out that the Gibbs sampler converges very slowly when each γ_l is generated from its full conditional (1.5). She reasoned that the interval (a_l, b_l) within which each γ_l must be generated from its full conditional can be very narrow, and, therefore, the cutpoint values can change very little between successive iterations. This is problematic as it makes the iterates highly correlated. Of course, the slower convergence of the γ_l is associated with the fact that there is little information on the latent variable distributions. Her empirical study also shows that the slow convergence of the cutpoints seriously affects the convergence of β. Then, instead of directly drawing γ from (1.5) she proposed a Metropolis-Hastings algorithm to generate γ from its marginal posterior distribution using a multivariate normal proposal density, which improves convergence of the Gibbs sampler used by Albert and Chib (1993). In the same spirit of Cowles (1996), Nandram and Chen (1996) proposed an alternative algorithm using a Dirichlet proposal density along with a reparameterization technique which improves convergence even further.

From the above discussions, we can see that to obtain a fully Bayesian analysis for correlated ordinal response data, we must develop novel modeling strategies as well as innovative computational algorithms. Cowles, Carlin and Connett (1996) used multivariate tobit models for analyzing longitudinal ordinal data which include correlations among the latent variables. However, they considered only the three-levels ordinal responses. Here we develop a new procedure for exact small sample Bayesian analysis of general correlated ordinal data models. In Section 2, we propose a novel modeling approach by considering latent variables, which is based on multivariate link functions using a very rich class of scale mixture of normals. A semi-automatic prior elicitation scheme by incorporating available historical information is developed, and the sufficient conditions for the propriety of the resulting informative prior distribution are established in Section 3. The distribution theory involved in the posterior calculations as well as various efficient computational algorithms for this complex simulation problem are presented in Section 4. In Section 5, we develop efficient Monte Carlo methods using the marginal likelihoods for model comparisons. A real data example from the two prostate cancer studies is used for illustrating the proposed methods in Section 6. Finally, Section 7 gives brief concluding remarks.

2 Models

We first introduce some notation which will be used throughout the remaining sections. Suppose that we observe an ordinal (1 through L_j) response Y_{ij} on the ith observations and the jth variable and let $x_{ij} = (x_{ij1}, x_{ij2}, \ldots, x_{ijp_j})$ be the corresponding p_j-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 whereas Y_1, Y_2, ..., Y_n are independent. Let $y_i = (y_{i1}, y_{i2}, \ldots, y_{iJ})'$, $y = (y_1, y_2, \ldots, y_n)$ be the observed data, and $D = (n, y, x)$ denote the data from the current study. Also let $\beta_j = (\beta_{j1}, \beta_{j2}, \ldots, \beta_{jp_j})'$ be a p_j-dimensional column vector of regression coefficients and $\beta = (\beta_1', \beta_2', \ldots, \beta_J')'$.

In order to set up the scale mixture of multivariate normal (SMMVN) link models for the correlated ordinal response data, we introduce a J-dimensional (latent) random vector $w_i^* = (w_{i1}^*, w_{i2}^*, \ldots, w_{iJ}^*)'$ such that

$$Y_{ij} = l, \quad \text{if} \ \ \gamma_{j,l-1}^* \leq w_{ij}^* < \gamma_{jl}^*, \tag{2.1}$$

where $-\infty = \gamma_{j0}^* \leq \gamma_{j1}^* \leq \gamma_{j2}^* \leq \gamma_{j,L_j-1}^* \leq \gamma_{jL_j}^* = \infty$ are cutpoints for the jth ordinal response, which divide the real line into L_j intervals. As explained by Nandram and Chen (1996), we specify $\gamma_{j1}^* = 0$ to ensure the identifiability of the cutpoint parameters. Here, we introduce different sets of cutpoints for different ordinal responses since in many practical problems, each ordinal response may behave quite differently. We further assume that

$$w_i^* \quad \sim \quad N(x_i\beta^*, \kappa(\lambda)\Sigma^*), \tag{2.2}$$

and

$$\lambda \sim \pi(\lambda), \tag{2.3}$$

where $\kappa(\lambda)$ is a positive function of one-dimensional positive-valued scale mixing variable λ, $\pi(\lambda)$ is a mixing distribution which is either discrete or continuous, $x_i = diag(x_{i1}, x_{i2}, \ldots, x_{iJ})$, and $\beta^* = (\beta_1^{*\prime}, \beta_2^{*\prime}, \ldots, \beta_J^{*\prime})'$ is a $p = \sum_{j=1}^J p_j$ dimensional column vector of regression coefficients corresponding to the cutpoints $\gamma_j^* = (\gamma_{j2}^*, \gamma_{j3}^*, \ldots, \gamma_{j,L_j-1}^*)'$ for $j = 1, 2, \ldots, J$. In (2.2) we further take $\Sigma^* = \left(\rho_{jj^*}^*\right)_{J\times J}$ to be a correlation matrix such that $\rho_{jj}^* = 1$ to ensure the identifiability of the parameters. 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.

For a special case where $\Sigma^* = I_J$, the $J \times J$ identity matrix, and $\kappa(\lambda) = 1$, the model defined by (2.1) and (2.2) reduces to an independent probit model given in (1.1). As discussed in the introduction section, the Gibbs sampler has already presented challenging problems in achieving convergence even for an independent probit model. Since the SMMVN-link models are much more complicated than the one for independent ordinal response data, it is expected that the computation be more challenging. This can be observed from the fact that for the SMMVN-link models, we need to deal with two difficult sampling problems: (i) generating cutpoints and (ii) generating correlation matrix.

To ease the computational burden, we extend the reparameterization technique of Nandram and Chen (1996) to the SMMVN-link models by considering the following transformation:

$$\delta_j = 1/\gamma_{j,L_j-1}^*, \ \gamma_{jl} = \delta_j \gamma_{jl}^*, \ \beta_j = \delta_j \beta_j^*, \ \text{and } w_{ij} = \delta_j w_{ij}^* \tag{2.4}$$

for $j = 1, 2, \ldots, J$ and $i = 1, 2, \ldots, n$. With transformation (2.4), the SMMVN-link models given by (2.1) and (2.2) become

$$Y_{ij} = l, \quad \text{if} \quad \gamma_{j,l-1} \le w_{ij} < \gamma_{jl}, \tag{2.5}$$

and

$$w_i \sim N(x_i\beta, \kappa(\lambda)\Sigma), \tag{2.6}$$

where the reparameterized cutpoints are $-\infty = \gamma_{j0} \le \gamma_{j1} = 0 \le \gamma_{j2} \le \cdots \le \gamma_{j,L_j-1} = 1 \le \gamma_{jL_j} = \infty$, $\Sigma = (\sigma_{jj^*})$, $\sigma_{jj} = \delta_j^2$, and $\sigma_{jj^*} = \delta_j \delta_{j^*} \rho_{jj^*}^*$ for $j \ne j^*$. The models given by (2.5) and (2.6) are thus called the reparameterized SMMVN-link models.

The reparameterized SMMVN-link models have several attractive features. First, the number of unknown cutpoints is reduced by J since in (2.5), we have only $L_j - 3$ unknown cutpoints for each j. Second, all unknown cutpoints γ_{jl} are between 0 and 1, i.e., $0 \le \gamma_{jl} \le 1$ for $l = 2, 3, \ldots, L_j - 2$ and $j = 1, 2, \ldots, J$. Third, the variance-covariance matrix Σ for w_i is unrestricted, which has a great advantage in implementation of MCMC sampling. Fourth, when $L_j = 3$, there are no unknown cutpoints. Due to these nice features, we use the reparameterized SMMVN-link models throughout the remaining sections.

Finally, we note that reparameterization (2.4) does not affect the distribution of the scale mixing variable λ. That is, we still have the same mixing distribution $\pi(\lambda)$ for the mixing variable λ. We also note the distribution of w_i determines the joint distribution of Y_i through (2.5) and the variance-covariance matrix Σ captures the correlations among the Y_{ij}'s. More specifically, we have the joint distribution of the correlated ordinal responses given by

$$\begin{aligned} & P(Y_{i1} = y_{i1}, Y_{i2} = y_{i2}, \ldots, Y_{iJ} = y_{iJ} | \beta, \Sigma^{-1}, \gamma, \lambda, x_i) \\ & = \int_{A_{i1}} \int_{A_{i2}} \cdots \int_{A_{iJ}} \frac{1}{(2\pi\kappa(\lambda))^{J/2} |\Sigma|^{\frac{1}{2}}} \\ & \times \exp\left\{ -\frac{\kappa^{-1}(\lambda)}{2} (w_i - x_i\beta)' \Sigma^{-1} (w_i - x_i\beta) \right\} dw_i, \end{aligned} \tag{2.7}$$

where $\gamma = (\gamma_1', \gamma_2', \ldots, \gamma_J')'$, $\gamma_j = (\gamma_{j2}, \gamma_{j3}, \ldots, \gamma_{j,L_j-2})'$, and

$$A_{ij} = (\gamma_{j,l-1}, \quad \gamma_{jl}] \text{ if } y_{ij} = l, \text{ for } j = 1, 2, \ldots, J. \tag{2.8}$$

The class of SMMVN links is quite rich, which includes multivariate probit (MVP), t-link (MVT), logit (MVL), symmetric stable distribution family links (MVS), symmetric exponential and power distribution family links (MVEP) models. To preserve space, we will give a brief explanation for MVP, MVT, and MVL as follows. The detailed discussions for the other links can be found in Chen and Dey (1998).

Taking $\kappa(\lambda) = 1$ and the mixing distribution $\pi(\{1\}) = 1$, the SMMVN-link reduces to the multivariate probit, i.e., MVP. Similar to the MVP, when we take $\kappa(\lambda) = 1/\lambda$ and $\lambda \sim \mathcal{G}\left(\frac{\nu}{2}, \frac{\nu}{2}\right)$, i.e., $\pi(\lambda) \propto \lambda^{\frac{\nu}{2}-1} \exp\left\{-\frac{\nu}{2}\lambda\right\}$, the SMMVN-link gives a multivariate t-link (MVT) with ν degrees of freedom. Note that the special case of MVT-link with $\nu = 1$ is termed as a multivariate Cauchy (MVC) link, and another special case of MVT-link with $\nu \to \infty$ is the MVP. Logistic regression is widely used to fit binary response data (e.g., see Prentice, 1988). The multivariate logit is a special SMMVN-link by taking $\kappa(\lambda) = 4\lambda^2$, where λ follows an asymptotic Kolmogorov distribution with density

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

The MVL models are attractive since the exchangeability on the correlation structure is not required, which is advantageous compared to the random effects type of logistic regression models, for example, stratified and mixture models as given in Prentice (1988).

It should be noticed that in the class of SMMVN links, the MVP and MVC links serve as the two extremes in light of the tail behavior, that is, the MVP has the lightest tail and the MVC link has the heaviest tail, while the others such as MVEP-link, MVL, and MVS-link have heavier tails than the MVP and lighter tails than the MVC.

3 Prior Distribution

3.1 Prior Elicitation

In this subsection, we will present a prior elicitation scheme from the historical studies for correlated ordinal response models.

Our prior construction is based on the notion of the existence of a previous study that measures the same response variable and covariates as the current study. For ease of exposition, we assume only one previous study, as the extension to multiple previous studies is straightforward. To this end, let $D_0 = (n_0, y_0, x_0)$ be the data from the historical study, where $y_0 = (y_{01}, y_{02}, \ldots, y_{0n_0})$ and $y_{0i} = (y_{0i1}, y_{0i2}, \ldots, y_{0iJ})$. Denote $w_{0i} = (w_{0i1}, \ldots, w_{0iJ})'$ to be the latent variable vector associated with the historical study. We propose a prior distribution for β given Σ^{-1}, γ, and a_0 of the form

$$\begin{aligned}
& \pi(\beta|\Sigma^{-1}, \gamma, a_0, D_0) \\
\propto \quad & \pi^*(\beta|\Sigma^{-1}, \gamma, a_0, D_0) = \prod_{i=1}^{n_0} \int_0^\infty \int_{A_{0i1}} \cdots \int_{A_{0iJ}} \frac{a_0^{J/2}\, |\Sigma|^{-1/2}}{(2\pi\kappa(\lambda_{0i}))^{J/2}} \\
& \times \exp\left\{ -\frac{a_0 \kappa^{-1}(\lambda_{0i})}{2} (w_{0i} - x_{0i}\beta)' \Sigma^{-1} (w_{0i} - x_{0i}\beta) \right\} \pi(\lambda_{0i}) dw_{0i} d\lambda_{0i},
\end{aligned} \tag{3.1}$$

where $\pi^*(\beta|\Sigma^{-1}, \gamma, a_0, D_0)$ is an unnormalized prior distribution, the scale mixing distribution $\pi(\lambda_{0i})$ is given in (2.3), and

$$A_{0ij} = (\gamma_{j,l-1}, \gamma_{jl}], \quad \text{if } y_{0ij} = l \text{ for } j = 1, \ldots, J. \tag{3.2}$$

In (3.1), a_0 can be interpreted as a scalar prior parameter that weights the prior data relative to the likelihood of the current study. It is reasonable to restrict the range of a_0 to be between 0 and 1, and thus we take $0 \le a_0 \le 1$. Notice that (3.1) has several appealing interpretations. Small values of a_0 give little prior weight to the historical control data relative to the likelihood of the current study whereas values of a_0 close to 1, for example, give roughly equal weight to the prior and the likelihood of the current study. When $a_0 \to 0$, (3.1) reduces to an improper uniform prior on

β, resulting in no incorporation of the historical data. The parameter a_0 allows the investigator to control the influence of the historical data on the current study. Such control is important in cases where there is heterogeneity between the previous and current study, or when the sample sizes of the two studies are quite different.

The prior specification is completed by specifying prior distributions for Σ^{-1}, γ, and a_0. We propose a joint prior distribution for $(\beta, \Sigma^{-1}, \gamma, a_0)$ of the form

$$\pi(\beta, \Sigma^{-1}, \gamma, a_0|D_0) \propto \pi^*(\beta|\Sigma^{-1}, \gamma, a_0, D_0)\pi_1(\Sigma^{-1}|K_0, Q_0)\pi_2(\gamma)\pi_3(a_0|\delta_0, \zeta_0), \tag{3.3}$$

where $\pi_1(\Sigma^{-1}|K_0, Q_0)$ is the probability density function of a Wishart distribution $W_J(K_0, Q_0)$ with degrees of freedom K_0 and mean matrix $K_0 Q_0$, and Q_0 is $J \times J$ symmetric and positive definite matrix. In (3.3), we also take independent uniform priors on γ, i.e., $\pi_2(\gamma) \propto 1$, and

$$\pi_3(a_0|\delta_0, \zeta_0) \ \propto \ a_0^{\delta_0 - 1}(1 - a_0)^{\zeta_0 - 1}, \quad 0 < a_0 < 1. \tag{3.4}$$

Note that in (3.3), hyperparameters K_0, Q_0, δ_0, and ζ_0 are prespecified *a priori*. Also note that when $\delta_0 = 1$ and $\zeta_0 = 1$, (3.4) reduces to a uniform prior on a_0 and that the prior specification (3.1) for β is indeed the generalization of *power prior distributions* proposed by Ibrahim, Ryan, and Chen (1998) and Chen, Ibrahim, and Yiannoutsos (1999) for analyzing univariate binary response data by using logistic regression, and Chen, Manatunga and Williams (1998) for human twin data models. However, the notion of power prior distributions cannot be directly adopted here due to the complexity of our SMMVN-link models.

3.2 Propriety of the Informative Prior Distribution

In this subsection, we investigate the propriety of the prior distribution $\pi(\beta, \Sigma^{-1}, \gamma, a_0|D_0)$ given in (3.3). We note that it is not immediately clear whether the prior distribution given in (3.3) is proper, since we use improper uniform priors for the regression coefficients β and the cutpoints γ.

Let $1\{S\}$ denote the indicator function such that $1\{S\} = 1$ if S is true, and $1\{S\} = 0$ if S is not true. Also, let

$$\begin{aligned}
&\tilde{x}_{0ij} = -x_{0ij}1\{2 \le y_{0ij} \le L_j\}, \quad \hat{x}_{0ij} = x_{0ij}1\{1 \le y_{0ij} \le L_j - 1\}, \\
&\tilde{c}_{ij} = (1\{3 \le y_{0ij}\}, \cdots, 1\{L_j \le y_{0ij}\}), \\
&\hat{c}_{ij} = -(1\{2 \le y_{0ij}\}, \cdots, 1\{L_j - 1 \le y_{0ij}\})1\{1 \le y_{0ij} \le L_j - 1\}, \\
&g_{ij} = \begin{pmatrix} \tilde{x}_{0ij} \\ \hat{x}_{0ij} \end{pmatrix}, \quad h_{ij} = \begin{pmatrix} \tilde{c}_{ij} \\ \hat{c}_{ij} \end{pmatrix}, \\
&g_i = \mathrm{diag}(g_{i,1}, \cdots, g_{i,J}), \quad h_i = \mathrm{diag}(h_{i,1}, \cdots, h_{i,J}),
\end{aligned}$$

$$G = \begin{pmatrix} g_1 \\ \vdots \\ g_{n_0} \end{pmatrix} \quad \text{and} \quad H = \begin{pmatrix} h_1 \\ \vdots \\ h_{n_0} \end{pmatrix}.$$

The following lemma plays a key role in the proof of our theorem concerning the propriety of the prior distribution.

Lemma 3.1 *Let $\theta = (\theta_1, ..., \theta_k)'$ and $\tau = (\tau_1, ..., \tau_l)'$, M be an $n_0 \times k$ matrix and N be an $n_0 \times l$ matrix, where $n_0 > k + l$. Assume that (M, N) is of full rank and that there exists a positive vector a such that*

$$a'M = 0 \quad \textit{and} \quad a'N \geq 0. \tag{3.5}$$

Then there exists a constant K depending only on (M, N) such that

$$||\eta|| \leq K||u|| \tag{3.6}$$

whenever

$$(M, N)\eta \leq u \quad \textit{and} \quad \tau \geq 0, \tag{3.7}$$

where $\eta = (\theta', \tau')'$ and $||\cdot||$ denotes the Euclidean norm.

Proof: Let $\mathcal{E} = \{\varepsilon = (\varepsilon_1, ..., \varepsilon_{k+l})' \in R^{k+l} : \quad \varepsilon_i = \pm 1\}$. Since (M, N) is of full rank, for every $\varepsilon \in \mathcal{E}$, there is a $b_\varepsilon \in R^n$ such that

$$b'_\varepsilon(M, N) = \varepsilon'. \tag{3.8}$$

Let $a = (a_1, ..., a_n)' \in R^n$ be the positive vector satisfying (3.5). Put

$$\delta = \frac{\min_{1 \leq i \leq n}(a_i)}{2 \max_{\varepsilon \in \mathcal{E}} ||b_\varepsilon||}.$$

For $\varepsilon = \varepsilon_\eta = \text{sign}(\eta') = (\text{sign}(\eta_1), ..., \text{sign}(\eta_{k+l}))'$, we have $\delta > 0$ and $a + \delta b_\varepsilon > 0$. Hence, it follows from (3.5) and (3.7) that

$$\begin{aligned} (a + \delta b_\varepsilon)'u &\geq (a + \delta b_\varepsilon)'(M, N)\eta \\ &= a'M\theta + a'N\tau + \delta b'_\varepsilon(M, N)\eta \\ &\geq \delta\, b'_\varepsilon(M, N)\eta = \delta\, \text{sign}(\eta')\eta \\ &\geq (\delta/(k + l))||\eta||, \end{aligned}$$

as desired. □

Now, we are led to the following theorem concerning the propriety of the prior distribution with improper uniform priors for β and γ.

Theorem 3.1 *Assume that the following conditions are satisfied:*

$(C1)$ (G, H) *is of full rank,*

$(C2)$ *There exists a positive vector* a *such that*

$$a'G = 0 \quad \text{and} \quad a'H \geq 0,$$

$(C3)$ $\int_0^\infty \kappa^{q/2}(\lambda)\pi(\lambda)d\lambda < \infty$, *where* $q = \sum_{j=1}^J p_j + \sum_{j=1}^J (L_j - 2)$,

$(C4)$ $\delta_0 > q/2$, *and* $\zeta_0 > 0$.

Then, the joint prior given in (3.3) is proper, that is

$$\int \pi^*(\beta|\Sigma, \gamma, a_0, D_0)\pi_1(\Sigma^{-1}|K_0, Q_0)\ a_0^{\delta_0 - 1}(1 - a_0)^{\zeta_0 - 1} d\beta d\Sigma d\gamma da_0 < \infty. \tag{3.9}$$

Proof: Let $\lambda_{01}, \cdots, \lambda_{0n_0}$ be independent random variables with the common probability density function π. Let $\tilde{w}_{0i} = (\tilde{w}_{0i1}, \cdots, \tilde{w}_{0iJ})'$ denote independent random variables such that

$$\tilde{w}_{0i}|\lambda_{0i} \sim N(0, (1/a_0)\kappa(\lambda_{0i})\Sigma)$$

and

$$A_{0i} = A_{0i1} \times A_{0i2} \times \cdots \times A_{0iJ}.$$

Thus, we can rewrite $\pi^*(\beta|\Sigma^{-1}, \gamma, a_0, D_0)$ given in (3.1) as

$$\begin{aligned} \pi^*(\beta|\Sigma^{-1}, \gamma, a_0, D_0) &= E\left[1\{(\tilde{w}_{0i} + x_{0i}\beta) \in A_{0i}, 1 \leq i \leq n_0\}\right] \\ &= E\Big[1\{\tilde{w}_{0ij} + x_{0ij}\beta_j \in A_{0ij}, 1 \leq j \leq J, 1 \leq i \leq n_0\}\Big]. \end{aligned}$$

Let

$$\gamma_j = \begin{pmatrix} \gamma_{j2} \\ \gamma_{j3} - \gamma_{j2} \\ \vdots \\ \gamma_{j,L_J-1} - \gamma_{j,L_J-2} \end{pmatrix}, \quad \eta = \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_J \\ \gamma_1 \\ \vdots \\ \gamma_J \end{pmatrix}, \ \tilde{w}^* = \begin{pmatrix} \tilde{w}_1^* \\ \vdots \\ \tilde{w}_{n_0}^* \end{pmatrix},$$

$$\tilde{w}_i^* = \Big(-\tilde{w}_{0ij}1\{1 \leq y_{0ij} \leq L_j - 1\}, \tilde{w}_{0ij}1\{2 \leq y_{0ij} \leq L_j\}, 1 \leq j \leq J\Big)'.$$

Noting that $\gamma_{j0} = -\infty, \gamma_{j1} = 0$ and $\gamma_{jL_j} = \infty$, we have

$$\begin{aligned}
&\{\tilde{w}_{0ij} + x_{0ij}\beta_j \in A_{0ij}, 1 \le j \le J, 1 \le i \le n_0\} \\
&\quad = \{\tilde{w}_{0ij} + x_{0ij}\beta_j < \gamma_{j,y_{0ij}}, 1 \le y_{0ij} \le L_j - 1, 1 \le i \le n_0\} \\
&\qquad \cap \{\tilde{w}_{0ij} + x_{0ij}\beta_j \ge \gamma_{j,y_{0ij}-1}, 2 \le y_{0ij} \le L_j, 1 \le i \le n_0\} \\
&\quad \subset \Big\{-\sum_{l=2}^{L_j-1}(\gamma_{jl} - \gamma_{j,l-1})1\{l \le y_{0ij}\} + x_{0ij}\beta_j \le -\tilde{w}_{0ij}, 1 \le y_{0ij} \le L_j - 1, \\
&\qquad\qquad 1 \le i \le n\Big\} \\
&\qquad \cap \Big\{\sum_{l=2}^{L_j-1}(\gamma_{jl} - \gamma_{j,l-1})1\{l+1 \le y_{0ij}\} - x_{0ij}\beta_j < \tilde{w}_{0ij}, 2 \le y_{0ij} \le L_j, \\
&\qquad\qquad 1 \le i \le n_0\Big\} \\
&\quad = \Big\{\hat{c}_{ij}\gamma_j + \hat{x}_{0ij}\beta_j \le -\tilde{w}_{0ij}1\{1 \le y_{0ij} \le L_j - 1\}, 1 \le i \le n_0\Big\} \\
&\qquad \cap \Big\{\tilde{c}_{ij}\gamma_j + \tilde{x}_{0ij}\beta_j \le \tilde{w}_{0ij}1\{2 \le y_{0ij} \le L_j\}, 1 \le i \le n_0\Big\} \\
&\quad = \{(G,H)\eta \le w^*\}.
\end{aligned}$$

Thus,

$$\pi^*(\beta|\Sigma^{-1}, \gamma, a_0, D_0) \le E\Big(1\{(G,H)\eta \le w^*\}\Big),$$

and by (C1), (C2), and Lemma 3.1,

$$\begin{aligned}
&\int \pi^*(\beta|\Sigma^{-1}, \gamma, a_0, D_0)\pi_1(\Sigma^{-1}|K_0, Q_0)\ a_0^{\delta_0-1}(1-a_0)^{\zeta_0-1} d\beta d\Sigma d\gamma da_0 \\
&\quad \le \int E\Big(1\{(G,H)\eta \le w^*\}\Big)\pi_1(\Sigma^{-1}|K_0, Q_0)\ a_0^{\delta_0-1}(1-a_0)^{\zeta_0-1} d\eta d\Sigma da_0 \\
&\quad \le \int E\Big(1\{||\beta|| \le K||w^*||\}\Big)\pi_1(\Sigma^{-1}|K_0, Q_0)\ a_0^{\delta_0-1}(1-a_0)^{\zeta_0-1} d\eta d\Sigma da_0) \\
&\quad \le K\int E\Big(\max_{1\le i\le n_0} ||\tilde{w}_i^*||\Big)^q \pi_1(\Sigma^{-1}|K_0, Q_0)\ a_0^{\delta_0-1}(1-a_0)^{\zeta_0-1} d\Sigma da_0 \\
&\quad \le K\int \sum_{i=1}^{n_0}\sum_{j=1}^{J} E|\tilde{w}_{0ij}|^q \pi_1(\Sigma^{-1}|K_0, Q_0)\ a_0^{\delta_0-1}(1-a_0)^{\zeta_0-1} d\Sigma da_0 \\
&\quad \le K\int \sum_{j=1}^{J} E\left[|\kappa(\lambda)/a_0|^{q/2}\right]\pi_1(\Sigma^{-1}|K_0, Q_0)\ a_0^{\delta_0-1}(1-a_0)^{\zeta_0-1} d\Sigma da_0 \\
&\quad = KJ\int E\left[|\kappa(\lambda)|^{q/2}\right]\pi_1(\Sigma^{-1}|K_0, Q_0) a_0^{\delta_0-q/2-1}(1-a_0)^{\zeta_0-1} d\Sigma da_0 \\
&\quad < \infty
\end{aligned}$$

by (C3) and (C4). □

We note that the condition $(C2)$ is easy to check via some standard optimization software. More specifically, an equivalent condition to $(C2)$ is

$$\inf_{a'H\geq 0,\ a_i\geq 1,\ i=1,2,\ldots,n} a'GG'a = 0, \tag{3.10}$$

which is a standard restricted quadratic optimization problem. For example, we may use the IMSL subroutine DNCONG to find the minimum value of the right hand side of (3.10). Then, the condition $(C2)$ is satisfied if the minimum value is 0.

4 Posterior Computation

We use a Markov chain Monte Carlo (MCMC) sampling scheme, namely, a Metropolis-Hastings algorithm (e.g., Metropolis *et al.*, 1953; Hastings, 1970; Tierney, 1994; and Chib and Greenberg, 1995), to perform the posterior computation. The posterior distribution for our correlated ordinal data model is of the form:

$$p(\beta, \Sigma^{-1}, \gamma, a_0|D, D_0) \ \propto\ L(\beta, \Sigma^{-1}, \gamma, D)\pi(\beta, \Sigma^{-1}, \gamma, a_0|D_0), \tag{4.1}$$

where the prior distribution $\pi(\beta, \Sigma^{-1}, \gamma, a_0|D_0)$ is given by (3.3) and the likelihood is

$$\begin{aligned} L(\beta, \Sigma^{-1}, \gamma, D) = \prod_{i=1}^{n} \int_0^\infty \int_{A_{i1}} \int_{A_{i2}} \cdots \int_{A_{iJ}} \frac{|\Sigma|^{-1/2}}{(2\pi\kappa(\lambda_i))^{J/2}} \\ \cdot \exp\left\{ -\frac{\kappa^{-1}(\lambda_i)}{2}(w_i - x_i\beta)'\Sigma^{-1}(w_i - x_i\beta) \right\} \pi(\lambda_i) dw_i d\lambda_i. \end{aligned} \tag{4.2}$$

To sample β, Σ^{-1}, γ, a_0 from (4.1), we introduce several auxiliary variables. These include the latent variables $w = (w_1', w_2', \ldots, w_n')'$ and the mixing variables $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_n)'$ for the current study and $w_0 = (w_{01}', \ldots, w_{0n_0}')'$ and $\lambda_0 = (\lambda_{01}, \ldots, \lambda_{0n_0})'$ for the historical study. To run the Metropolis-Hastings sampling algorithm, we need to generate β, Σ^{-1}, γ, w, w_0, λ, λ_0, and a_0 from their respective conditional distributions. Necessary steps of the Metropolis-Hastings algorithm are briefly described as follows.

Let $B = a_0 \sum_{i=1}^{n_0} \kappa^{-1}(\lambda_{0i}) x_{0i}' \Sigma^{-1} x_{0i} + \sum_{i=1}^{n} \kappa^{-1}(\lambda_i) x_i' \Sigma^{-1} x_i$ and

$$\hat{\beta} = B^{-1} \left(a_0 \sum_{i=1}^{n_0} \kappa^{-1}(\lambda_{0i}) x_{0i}' \Sigma^{-1} w_{0i} + \sum_{i=1}^{n} \kappa^{-1}(\lambda_i) x_i' \Sigma^{-1} w_i \right).$$

Then, given Σ^{-1}, w, w_0, λ, λ_0, and a_0, we have

$$\beta \mid \Sigma^{-1}, w, \lambda, w_0, \lambda_0, a_0, D, D_0 \quad \sim \quad N(\hat{\beta}, B^{-1}). \tag{4.3}$$

The conditional distribution of Σ^{-1} given β, w, w_0, λ, and λ_0 is a Wishart distribution, that is,

$$\Sigma^{-1} \mid \beta, w, w_0, \lambda, \lambda_0, D, D_0 \quad \sim \quad W_J(n + n_0 + K_0, Q^*), \tag{4.4}$$

where

$$\begin{aligned} Q^{*-1} &= Q_0^{-1} + \sum_{i=1}^{n} \kappa^{-1}(\lambda_i)(w_i - x_i\beta)(w_i - x_i\beta)' \\ &+ a_0 \sum_{i=1}^{n_0} \kappa^{-1}(\lambda_{0i})(w_{0i} - x_{0i}\beta)(w_{0i} - x_{0i}\beta)'. \end{aligned}$$

Therefore, generating β and Σ^{-1} from (4.3) and (4.4) is straightforward. We notice that without transformation (2.4), we must draw the correlation matrix Σ^* from its conditional posterior distribution based on (2.2). From Chen and Dey (1998) or Chib and Greenberg (1998), it can be seen that generating a correlation matrix is much more difficult than drawing a variance-covariance matrix from a Wishart distribution.

Next, we consider sampling a_0. From (3.1) and (3.4), the conditional posterior distribution of a_0 is of the form

$$\begin{aligned} & p(a_0|\beta, \Sigma^{-1}, w_0, \lambda_0, D_0) \\ \propto \; & a_0^{\frac{Jn_0}{2}} \exp\left\{ -\frac{a_0}{2} \sum_{i=1}^{n_0} \kappa^{-1}(\lambda_{0i})(w_{0i} - x_{0i}\beta)'\Sigma^{-1}(w_{0i} - x_{0i}\beta) \right\} \\ & \times a_0^{\delta_0 - 1}(1 - a_0)^{\zeta_0 - 1}, \end{aligned} \tag{4.5}$$

where $0 \le a_0 \le 1$. Since $p(a_0|\beta, \Sigma^{-1}, w_0, \lambda_0, D_0)$ is not log-concave in general, we use a Metropolis algorithm to sample a_0. Consider the transformation:

$$a_0 = \frac{\exp(\xi_i)}{1 + \exp(\xi_i)}. \tag{4.6}$$

Then, the conditional posterior distribution $[\xi_i|\beta, \Sigma^{-1}, w_0, \lambda_0, D_0]$ is

$$p(\xi_i|\beta, \Sigma^{-1}, w_0, \lambda_0, D_0) \;\propto\; p(a_0|\beta, \Sigma^{-1}, w_0, \lambda_0, D_0)\frac{\exp(\xi_i)}{(1 + \exp(\xi_i))^2}, \tag{4.7}$$

where $p(a_0|\beta, \Sigma^{-1}, w_0, \lambda_0, D_0)$ is given by (4.5) and a_0 is evaluated at $a_0 = \exp(\xi_i)/(1 + \exp(\xi_i))$. Instead of directly generating a_0 from (4.5), we first generate ξ_i from (4.7) and then use (4.6) to obtain a_0. To generate ξ_i, we use a normal proposal $N(\hat{\xi}_i, \hat{\tau}^2_{\hat{\xi}_i})$ where $\hat{\xi}_i$ is a maximizer of the logarithm of the right side of (4.7), which can be obtained by using, for example, the Nelder-Mead algorithm implemented by O'Neill (1971), and $\hat{\tau}^2_{\hat{\xi}_i}$ is the minus of the inverse of the second derivative of $\ln p(\xi_i|\beta, \Sigma^{-1}, w_0, \lambda_0, D_0)$ evaluated at $\xi_i = \hat{\xi}_i$, that is

$$\hat{\tau}^{-2}_{\hat{\xi}_i} = - \left. \frac{d^2 \ln p(\xi_i|\beta, \Sigma^{-1}, w_0, \lambda_0, D_0)}{d\xi_i^2} \right|_{\xi_i = \hat{\xi}_i} .$$

The algorithm to generate ξ_i operates as follows: (i) let ξ_i be the current value; (ii) generate a proposal value ξ_i^* from $N(\hat{\xi}_i, \hat{\tau}^2_{\hat{\xi}_i})$; and (iii) a move from ξ_i to ξ_i^* is made with probability

$$\min \left\{ \frac{p(\xi_i^*|\beta, \Sigma^{-1}, w_0, \lambda_0, D_0)\phi\left(\frac{\xi_i - \hat{\xi}_i}{\hat{\tau}_{\hat{\xi}_i}}\right)}{p(\xi_i|\beta, \Sigma^{-1}, w_0, \lambda_0, D_0)\phi\left(\frac{\xi_i^* - \hat{\xi}_i}{\hat{\tau}_{\hat{\xi}_i}}\right)}, 1 \right\},$$

where ϕ is the standard normal probability density function. After we generate ξ_i, we compute a_0 by using (4.6).

To generate the γ_j, w, and w_0 from their conditional distributions, the Gibbs sampler may present challenging problems in achieving convergence as discussed in the introduction section. Therefore, we propose an alternative efficient MCMC sampling scheme as follows. Let $w_{(j)} = (w_{1j}, w_{2j}, \ldots, w_{nj})'$ and denote $w_{(-j)}$ to be w with $w_{(j)}$ deleted for $j = 1, 2, \ldots, J$. Also let $w_{0(j)} = (w_{01j}, w_{02j}, \ldots, w_{0n_0j})'$ and denote $w_{0(-j)}$ to be w_0 with $w_{0(j)}$ deleted for $j = 1, 2, \ldots, J$. Then, we use a cycle of J Gibbs steps to generate γ_j, $w_{(j)}$, and $w_{0(j)}$ jointly from their conditional distributions for $j = 1, 2, \ldots, J$ in turn. For $j = 1, 2, \ldots, J$, we first draw γ_j from $[\gamma_j|\beta, \Sigma^{-1}, w_{(-j)}, w_{0(-j)}, \lambda, \lambda_0, a_0, D]$, then draw $w_{(j)}$ from $[w_{(j)}|\gamma_j, \beta, \Sigma^{-1}, w_{(-j)}, \lambda, D]$ and draw $w_{0(j)}$ from $[w_{0(j)}|\gamma_j, \beta, \Sigma^{-1}, w_{0(-j)}, \lambda_0, a_0, D]$. We use the algorithm of Geweke (1991) to generate $w_{(j)}$ and $w_{0(j)}$, since their respective conditional posterior distributions are truncated multivariate normals over intervals defined by (2.8) or (3.2). It can be easily observed that given γ_j, β, Σ^{-1}, λ, $\lambda^{(0)}$, and D, w_{1j}, w_{2j}, ..., w_{nj}, w_{01j}, w_{02j}, ..., and w_{0n_0j}, are independent. Therefore, the conditional posterior distribution $[\gamma_j|\beta, \Sigma^{-1}, w_{(-j)}, w_{0(-j)}, \lambda, \lambda_0, a_0, D, D_0]$ is

$$p(\gamma_j|\beta, \Sigma^{-1}, w_{(-j)}, w_{0(-j)}, \lambda, \lambda_0, a_0, D, D_0) \quad \propto$$

$$\prod_{i:\ y_{ij}=2}\Big\{\Phi\Big(\frac{\gamma_{j2}-\tilde{\mu}_{ij}}{\tilde{\sigma}_{ij}}\Big)-\Phi\Big(-\frac{\tilde{\mu}_{ij}}{\tilde{\sigma}_{ij}}\Big)\Big\}$$
$$\times\prod_{i:\ y_{ij}=3}\Big\{\Phi\Big(\frac{\gamma_{j3}-\tilde{\mu}_{ij}}{\tilde{\sigma}_{ij}}\Big)-\Phi\Big(\frac{\gamma_{j2}-\tilde{\mu}_{ij}}{\tilde{\sigma}_{ij}}\Big)\Big\}\cdots$$
$$\times\prod_{i:\ y_{ij}=L_j-1}\Big\{\Phi\Big(\frac{1-\tilde{\mu}_{ij}}{\tilde{\sigma}_{ij}}\Big)-\Phi\Big(\frac{\gamma_{j,L_j-2}-\tilde{\mu}_{ij}}{\tilde{\sigma}_{ij}}\Big)\Big\}$$
$$\times\prod_{i:\ y_{0ij}=2}\Big\{\Phi\Big(\frac{\gamma_{j2}-\tilde{\mu}_{0ij}}{\tilde{\sigma}_{0ij}}\Big)-\Phi\Big(-\frac{\tilde{\mu}_{0ij}}{\tilde{\sigma}_{0ij}}\Big)\Big\}$$
$$\times\prod_{i:\ y_{0ij}=3}\Big\{\Phi\Big(\frac{\gamma_{j3}-\tilde{\mu}_{0ij}}{\tilde{\sigma}_{0ij}}\Big)-\Phi\Big(\frac{\gamma_{j2}-\tilde{\mu}_{0ij}}{\tilde{\sigma}_{0ij}}\Big)\Big\}\cdots$$
$$\times\prod_{i:\ y_{0ij}=L_j-1}\Big\{\Phi\Big(\frac{1-\tilde{\mu}_{0ij}}{\tilde{\sigma}_{0ij}}\Big)-\Phi\Big(\frac{\gamma_{j,L_j-2}-\tilde{\mu}_{0ij}}{\tilde{\sigma}_{0ij}}\Big)\Big\}, \tag{4.8}$$

where

$$\tilde{\mu}_{ij}=x_{ij}\beta_j+\Sigma_{jj}\Sigma_{(-j)}^{-1}(w_{i(-j)}-x_{i(-j)}\beta_{(-j)}) \tag{4.9}$$

and

$$\tilde{\sigma}_{ij}^2=\kappa(\lambda_i)\left(\sigma_{jj}-\Sigma_{jj}\Sigma_{(-j)}^{-1}\Sigma_{jj}'\right). \tag{4.10}$$

In (4.9) and (4.10), $x_{i(-j)}$ is x_i with the ith row deleted, $\beta_{(-j)}$ is β with β_j deleted, $\Sigma_{(-j)}$ is Σ with the jth row and jth column deleted, and $\Sigma_{jj}=(\sigma_{j1},\ldots,\sigma_{j,j-1},\sigma_{j,j+1},\ldots,\ \sigma_{jJ})$. In (4.8), the definition of $\tilde{\mu}_{0ij}$ or $\tilde{\sigma}_{0ij}^2$ is similar to (4.9) or (4.10) with replacing $w_{i(-j)}$ and $x_{i(-j)}$ by $w_{0i(-j)}$ and $x_{0i(-j)}$.

Generating γ_j from (4.8) is a challenging problem. Cowles (1996) proposed a Hastings scheme using a multivariate truncated normal proposal distribution and Nandram and Chen (1996) developed an improved algorithm using Dirichlet proposal distribution to draw γ_j simultaneously from its conditional distribution (4.8). Chen and Schmeiser (1998) suggested to use a nearly automatic algorithm, a random-direction interior-point (RDIP) approach, to generate γ_j. The RDIP requires the minimum input from a user, but it may not be very efficient due to the nature of black-box algorithms. Here, we propose a simple Metropolis-Hastings algorithm using a transformation technique. Let

$$\gamma_{jl}=\frac{\gamma_{j,l-1}+e^{\zeta_{jl}}}{1+e^{\zeta_{jl}}},\quad l=2,\ldots,L_j-2 \tag{4.11}$$

and $\zeta_j=(\zeta_{j2},\ldots,\zeta_{j,L_j-2})'$. Then, the conditional distribution $[\zeta_j|\beta,\ \Sigma^{-1},$

$w_{(-j)}$, $w_{0(-j)}$, λ, λ_0, a_0, D, D_0] is

$$\begin{aligned} &p(\zeta_j|\beta, \Sigma^{-1}, w_{(-j)}, w_{0(-j)}, \lambda, \lambda_0, a_0, D, D_0) \\ &\quad \propto \quad p(\gamma_j|\beta, \Sigma^{-1}, w_{(-j)}, w_{0(-j)}, \lambda, \lambda_0, a_0, D, D_0) \prod_{l=2}^{L_j-2} \frac{(1-\gamma_{j,l-1})e^{\zeta_{jl}}}{(1+e^{\zeta_{jl}})^2}, \end{aligned} \tag{4.12}$$

where $p(\gamma_j|\beta, \Sigma^{-1}, w_{(-j)}, w_{0(-j)}, \lambda, \lambda_0, a_0, D, D_0)$ is given by (4.8) and γ_j is evaluated at $\gamma_{jl} = (\gamma_{j,l-1} + e^{\zeta_{jl}})/(1 + e^{\zeta_{jl}})$ for $l = 2, 3, \ldots, L_j - 2$. Now we follow the remaining steps of the Metropolis-Hastings algorithm for generating a_0. Compared to the algorithm of Nandram and Chen (1996), our Metropolis-Hastings algorithm does not require the cell counts $m_{jl} = \sum_{i=1}^{n} 1_{\{y_{ij}=l\}} + \sum_{i=1}^{n_0} 1_{\{y_{0ij}=l\}}$ to be balanced, where the indicator function $1_{\{y_{ij}=l\}} = 1$ if $y_{ij} = l$ and 0 otherwise.

Finally, we briefly discuss how to generate mixing variables λ_i and λ_{0i}. The random generation for λ_i or λ_{0i} requires the known form of the mixing distribution $\pi(\lambda)$. For a MVP model, it does not require to generate λ_i or λ_{0i}, since both $\pi(\{\lambda_i = 1\}) = 1$ and $\pi(\{\lambda_{0i} = 1\}) = 1$. For a MVT model, $[\lambda_i|\beta, \Sigma^{-1}, w_i, D]$ and $[\lambda_{0i}|\beta, \Sigma^{-1}, w_{0i}, D_0]$ are gamma distributions, which are easy to sample. For the MVL, MVS and MVEP link models, Chen and Dey (1998) developed various efficient Metropolis algorithms. Their algorithms can be directly applied to our correlated ordinal data models, and thus we omit the details.

5 Model Comparisons

In this section we consider the problem of accounting for uncertainty about model form. Here we are faced with many models within a class of SMMVN-link models. Although we may wish to summarize our findings with a single model, there are usually many choices to be made. In this context, we consider marginal likelihood approach (Chib and Greenberg, 1998) for model comparisons since this approach is particularly suitable for the correlated ordinal data models.

To compare different SMMVN-link models, we calculate the marginal likelihoods for each of the models and choose the model which yields the largest marginal likelihood. As discussed in Chen and Dey (1998) and Chib and Greenberg (1998), the marginal likelihood approach is essentially equivalent to the Bayes factor approach of Kass and Raftery (1995).

Let $m(D)$ be the marginal likelihood. Then, $m(D)$ is the normalizing constant of the posterior density $p(\beta, \Sigma^{-1}, \gamma, a_0|D, D_0)$ given in (4.1). That is,

$$p(\beta, \Sigma^{-1}, \gamma, a_0|D, D_0) = \frac{1}{m(D)} L(\beta, \Sigma^{-1}, \gamma, D)\pi(\beta, \Sigma^{-1}, \gamma, a_0|D_0). \tag{5.1}$$

To compute the marginal likelihood, we adopt a data-augmentation-based method of Chib (1995). Given some point $(\beta^*, \Sigma^*, \gamma^*, a_0^*)$ (typically the posterior means of β, Σ, γ, and a_0), using (5.1) we have an identity for the marginal likelihood on the natural log scale as

$$\begin{aligned} \ln m(D) &= \ln L(\beta^*, (\Sigma^*)^{-1}, \gamma^*, D) + \ln \pi(\beta^*, (\Sigma^*)^{-1}, \gamma^*, a_0^* | D_0) \\ &- \ln p(\beta^*, (\Sigma^*)^{-1}, \gamma^*, a_0^* | D, D_0). \end{aligned} \tag{5.2}$$

Following Chib and Greenberg (1998), we write

$$\begin{aligned} \ln p(\beta^*, (\Sigma^*)^{-1}, \gamma^*, a_0^* | D, D_0) &= \ln p(\beta^* | D, D_0) + \ln p((\Sigma^*)^{-1} | \beta^*, D, D_0) \\ &+ \ln p(\gamma^* | \beta^*, (\Sigma^*)^{-1}, D, D_0) \\ &+ \ln p(a_0^* | \beta^*, (\Sigma^*)^{-1}, \gamma^*, D, D_0) \end{aligned}$$

and

$$\begin{aligned} \ln \pi(\beta^*, (\Sigma^*)^{-1}, \gamma^*, a_0^* | D_0) &= \ln \pi(\beta^* | D_0) + \ln \pi((\Sigma^*)^{-1} | \beta^*, D_0) \\ &+ \ln \pi(\gamma^* | \beta^*, (\Sigma^*)^{-1}, D_0) \\ &+ \ln \pi(a_0^* | \beta^*, (\Sigma^*)^{-1}, \gamma^*, D_0). \end{aligned}$$

It can be shown that given β^*, $(\Sigma^*)^{-1}$, and γ^*,

$$p(a_0^* | \beta^*, (\Sigma^*)^{-1}, \gamma^*, D, D_0) = \pi(a_0^* | \beta^*, (\Sigma^*)^{-1}, \gamma^*, D_0).$$

Thus, (5.2) reduces to

$$\begin{aligned} \ln m(D) &= \ln L(\beta^*, (\Sigma^*)^{-1}, \gamma^*, D) \\ &+ \ln \pi(\beta^* | D_0) + \ln \pi((\Sigma^*)^{-1} | \beta^*, D_0) \\ &+ \ln \pi(\gamma^* | \beta^*, (\Sigma^*)^{-1}, D_0) \\ &- \Big[\ln p(\beta^* | D, D_0) + \ln p((\Sigma^*)^{-1} | \beta^*, D, D_0) \\ &+ \ln p(\gamma^* | \beta^*, (\Sigma^*)^{-1}, D, D_0) \Big]. \end{aligned} \tag{5.3}$$

Except for $L(\beta^*, (\Sigma^*)^{-1}, \gamma^*, D)$, all the other six terms in the right side of (5.3) are the marginal or conditional posterior or prior densities of β, Σ^{-1}, and γ evaluated at β^*, $(\Sigma^*)^{-1}$, and γ^*. Interestingly, the most efficient conditional marginal density estimation (CMDE) method of Gelfand, Smith and Lee (1992), can be adopted for computing these quantities. We give a brief explanation as follows. Let $\Big\{ \Big(\beta^{(1r)}, (\Sigma^{(1r)})^{-1}, \gamma^{(1r)}, a_0^{(1r)}\ w^{(1r)},\ \lambda^{(1r)},$

$w_0^{(1r)},\ \lambda_0^{(1r)}\Big),\ r=1,2,\ldots,R\Big\}$ be the first MCMC sample from the posterior distribution $p(\beta,\Sigma^{-1},\ \gamma,\ a_0,\ w,\ \lambda, w_0,\lambda_0|D,D_0)$ using the Metropolis-Hastings algorithm described in Section 4. Then, a simulation-consistent estimator of $p(\beta^*|D,D_0)$ is given by

$$\begin{aligned}\hat{p}(\beta^*|D,D_0) &= \frac{1}{R}\sum_{r=1}^{R}\left(\frac{1}{2\pi}\right)^{p/2}|B^{(r)}|^{1/2} \\ &\times \exp\left\{-\frac{1}{2}(\beta^*-\hat{\beta}^{(r)})'B^{(r)}(\beta^*-\hat{\beta}^{(r)})\right\}, \end{aligned} \tag{5.4}$$

where

$$\begin{aligned}\hat{\beta}^{(r)} &= (B^{(r)})^{-1}\Bigg(a_0^{(1r)}\sum_{i=1}^{n_0}\kappa^{-1}(\lambda_{0i}^{(1r)})x_{0i}'(\Sigma^{(1r)})^{-1}w_{0i}^{(1r)} \\ &+ \sum_{i=1}^{n}\kappa^{-1}(\lambda_i^{(1r)})x_i'(\Sigma^{(1r)})^{-1}w_i^{(1r)}\Bigg),\end{aligned}$$

$B^{(r)}=a_0^{(1r)}\sum_{i=1}^{n_0}\kappa^{-1}(\lambda_{0i}^{(1r)})x_{0i}'(\Sigma^{(1r)})^{-1}x_{0i}+\sum_{i=1}^{n}\kappa^{-1}(\lambda_i^{(1r)})x_i'(\Sigma^{(1r)})^{-1}x_i$, and $p=\sum_{j=1}^{J}p_j$. Assume that $\Big\{\Big((\Sigma^{(2r)})^{-1},\gamma^{(2r)},a_0^{(2r)}\ \ w^{(2r)},\ \lambda^{(2r)},\ w_0^{(2r)},\ \lambda_0^{(2r)}\Big),\ r=1,2,\ldots,R\Big\}$ is the second MCMC sample from the conditional posterior distribution $p(\Sigma^{-1},\ \gamma,\ a_0,\ w,\ \lambda,w_0,\lambda_0|\beta^*,D,D_0)$, which is independent of the first MCMC sample. Then, a simulation-consistent estimator of $p((\Sigma^*)^{-1}|\beta^*,D,D_0)$ is given by

$$\begin{aligned}&\hat{p}(\Sigma^{*-1}|\beta^*,D,D_0)=\frac{1}{R}\sum_{r=1}^{R}|Q^{*(r)}|^{-(n+n_0+K_0)/2} \\ &\times\frac{\exp\left\{-\frac{1}{2}tr\left((Q^{*(r)})^{-1}(\Sigma^*)^{-1}\right)\right\}|\Sigma^*|^{-\frac{n+n_0+K_0-J-1}{2}}}{2^{(n+n_0+K_0)J/2}\pi^{J(J-1)/4}\prod_{j=1}^{J}\Gamma\left(\frac{n+n_0+K_0-j+1}{2}\right)},\end{aligned} \tag{5.5}$$

where $tr\left((Q^{*(r)})^{-1}(\Sigma^*)^{-1}\right)$ denotes the trace of $(Q^{*(r)})^{-1}(\Sigma^*)^{-1}$, and

$$\begin{aligned}(Q^{*(r)})^{-1} &= Q_0^{-1}+\sum_{i=1}^{n_0}\kappa^{-1}(\lambda_{0i}^{(2r)})(w_{0i}^{(2r)}-x_{0i}\beta^*)(w_{0i}^{(2r)}-x_{0i}\beta^*)' \\ &+ \sum_{i=1}^{n}\kappa^{-1}(\lambda_i^{(2r)})(w_i^{(2r)}-x_i\beta^*)(w_i^{(2r)}-x_i\beta^*)'.\end{aligned}$$

Now, let $\left\{\left(\gamma^{(3r)}, a_0^{(3r)}\ w^{(3r)},\ \lambda^{(3r)},\ w_0^{(3r)},\ \lambda_0^{(3r)}\right),\ r = 1, 2, \ldots, R\right\}$ be the third MCMC sample from the conditional posterior distribution $p(\gamma, a_0, w, \lambda, w_0, \lambda_0|\beta^*, (\Sigma^*)^{-1}, D, D_0)$, which is independent of the first two MCMC samples. Let

$$S_{lj} = \{i : \quad y_{ij} = l,\ i = 1, 2, \ldots, n\} \cup \{i : \quad y_{0ij} = l,\ i = 1, 2, \ldots, n_0\}$$

for $l = 2, 3, \ldots, L_j - 1$ and $j = 1, 2, \ldots, J$. If all sets S_{lj}'s are not empty, a simulation-consistent estimator of $p(\gamma^*|\beta^*, (\Sigma^*)^{-1}, D, D_0)$ is given by

$$\hat{p}(\gamma^*|\beta^*, (\Sigma^*)^{-1}, D, D_0) = \frac{1}{R}\sum_{r=1}^{R}\prod_{j=1}^{J}\prod_{l=2}^{L_j-2}\frac{1}{T_{2lj}^{(r)} - T_{1lj}^{(r)}}1_{\{T_{1lj}^{(r)} < \gamma_{jl}^* \le T_{2lj}^{(r)}\}}, \quad (5.6)$$

where $T_{2lj}^{(r)} = \min\left\{\min\{w_{ij}^{(3r)} : y_{ij} = l + 1\}, \min\{w_{0ij}^{(3r)} : y_{0ij} = l + 1\}\right\}$ and $T_{1lj}^{(r)} = \max\left\{\max\{w_{ij}^{(3r)} : y_{ij} = l\}, \max\{w_{0ij}^{(3r)} : y_{0ij} = l\}\right\}$. When, J is large, the estimator given in (5.6) may not be efficient, a better Monte Carlo estimator can be obtained using a sequence of $J - 3$ dimensional conditional marginal distributions for γ. By the chain rule decomposition, we can write

$$\begin{aligned} p(\gamma^*|\beta^*, (\Sigma^*)^{-1}, D, D_0) &= p(\gamma_1^*|\beta^*, (\Sigma^*)^{-1}, D, D_0) \\ &\times p(\gamma_2^*|\gamma_1^*, \beta^*, (\Sigma^*)^{-1}, D, D_0)\cdots \\ &\times p(\gamma_J^*|\gamma_1^*, \ldots, \gamma_{J-1}^*, \beta^*, (\Sigma^*)^{-1}, D, D_0). \end{aligned}$$

Similar to (5.6), we obtain an estimate for each $p(\gamma_j^*|\gamma_1^*,\ \ldots,\ \gamma_{j-1}^*,\ \beta^*,\ (\Sigma^*)^{-1},\ D, D_0)$ for $j = 1, 2, \ldots, J$.

Analogous to (5.4), (5.5), and (5.6), we can obtain efficient Monte Carlo estimators for $\pi(\beta^*|D_0)$, $\pi((\Sigma^*)^{-1}|\beta^*, D_0)$ and $\pi(\gamma^*|\beta^*, (\Sigma^*)^{-1}, D_0)$ using three MCMC samples from the corresponding prior or conditional prior distributions. Finally, we consider how to compute $L(\beta^*, (\Sigma^*)^{-1}, \gamma^*, D)$. Let $\Sigma_d^* = diag(\Sigma^*)$, $c_L = L(\beta^*, (\Sigma^*)^{-1}, \gamma^*, D)$, and $c_L^* = L(\beta^*, (\Sigma_d^*)^{-1}, \gamma^*, D)$. Then, c_L^* can be evaluated numerically, since

$$c_L^* = \prod_{i=1}^{n}\int_0^\infty \prod_{j=1}^{J}\left[\Phi\left(\frac{\gamma_{jy_{ij}}^* - x_{ij}\beta_j^*}{\sqrt{\kappa(\lambda_i)\sigma_{jj}^*}}\right) - \Phi\left(\frac{\gamma_{j,y_{ij}-1}^* - x_{ij}\beta_j^*}{\sqrt{\kappa(\lambda_i)\sigma_{jj}^*}}\right)\right]\pi(\lambda_i)d\lambda_i.$$

Letting

$$\begin{aligned} p(w, \lambda \mid \beta^*, \Sigma^*, D) &= \prod_{i=1}^{n}\frac{|\Sigma^*|^{-\frac{1}{2}}}{(2\pi\kappa(\lambda_i))^{J/2}} \\ &\times \exp\left\{-\frac{\kappa^{-1}(\lambda_i)}{2}(w_i - x_i\beta^*)'(\Sigma^*)^{-1}(w_i - x_i\beta^*)\right\}\pi(\lambda_i), \end{aligned}$$

we have

$$\frac{c_L}{c_L^*} = E\left[\frac{p(w, \lambda \mid \beta^*, \Sigma^*, D)}{p(w, \lambda \mid \beta^*, \Sigma_d^*, D)}\right],$$

where the expectation is taken with respect to $p(w, \lambda \mid \beta^*, \Sigma_d^*, D)$. Then, $\frac{c_L}{c_L^*}$ can be easily computed using a MCMC sample from $p(w, \lambda \mid \beta^*, \Sigma_d^*, D)$ (see Chen and Shao, 1997a).

6 An Illustrative Example

To illustrate the proposed methodologies, we consider a real data example from prostate cancer studies. Adenocarcinoma of the prostate is the second-leading cause of cancer mortality in men. In order to examine the relation between several prostate cancer response variables and the important preoperative staging system predictors, two similar studies were conducted in the Hospital of the University of Pennsylvania at Philadelphia and Brigham and Women's Hospital at Boston, respectively. All patients involved in these two studies had undergone surgery. Two data sets, called the PENN data and the MASS data, were collected from these two studies. The PENN data contain 713 patients and the same pathologist was involved for all patients from 1989 to 1995. The MASS data contain the information for a prospective study of 104 patients with prostate cancer and the treatment took place between August of 1995 and April of 1996. For illustrative purposes, we consider two clinical categorical response variables (i.e., Pathological Extracapsular Extension (PECE) and Pathological Positive Surgical Margins (PPSM)) and three most important predictors, which are Prostate Specific Antigen (PSA), Clinical Gleason Score (GLEAS), and Clinical Stage (CLINS). PECE is an ordinal response that takes the values of "0", "1" or "2" and defines whether or not cancer has penetrated the prostatic capsule, where a "0" indicates a negative value which means that there is no cancer present in or near the capsule at all, a "1" indicates that the disease extends into but not through the capsule, and a "2" means that the disease has penetrated through the capsule. PPSM is another ordinal response having values of "0", "1" or "2" which distinguishes whether the cancer has been completely removed or not, where a value of "0" indicates a negative outcome and "2" indicates a positive outcome while "1" gives an outcome between "negative" and "positive". Notice that here "0", "1", and "2" denote levels 1, 2, and 3 in our general notation in Section 2. In the prostate cancer study, it is important to predict the outcomes of PECE, PPSM, and PSVI in order to determine whether a prostate cancer patient needs to undergo the surgery. See Desjardin (1997) for more detailed descriptions and discussions.

For patient i, we let Y_{i1} and Y_{i2} denote PECE and PPSM and let x_{i1}, x_{i2}, x_{i3}, and x_{i4} be an intercept, PSA, GLEAS, and CLINS. Then, both Y_{i1} and Y_{i2} are ordinal and each of them has three levels. Therefore, $J = 2$ and $L_1 = L_2 = 3$, which implies that there are no unknown cutpoints.

Since Y_{i1} and Y_{i2} are observed from the same patient, they are naturally correlated. Furthermore, the MASS study was conducted recently while the PENN study was done earlier. Therefore, the MASS study naturally serves as a current study while the PENN study is a historical study. The sample size of the MASS data is 104, i.e., $n = 104$, which is relatively small. In order to perform a more accurate statistical analysis, it is important to include the available historical information, i.e., the PENN data, into the analysis.

For illustrative purposes, we consider four SMMVN-link models to fit the prostate cancer data. These models are the MVP, MVT10 (i.e., MVT with $\nu = 10$), MVL, and MVC (i.e., MVT with $\nu = 1$). 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. The MVL model is roughly in the "halfway" between the MVP and MVC models, while MVT10 model is between the MVP and MVL models. We implement the MCMC algorithms proposed in Section 4. We take $K_0 = 3$ and $Q_0^{-1} = 0.001 I_2$, where I_2 is a 2×2 identity matrix. To ease computational burden, we standardize all three covariates. To incorporate the historical information from the PENN study, we choose $\delta_0 = 6$ and $\zeta_0 = 1$ in (3.4) to ensure the propriety of the informative prior distribution. We check the convergence of the MCMC algorithm using several diagnostic procedures recommended by Cowles and Carlin (1996) and after convergence, we find that the autocorrelations among the MCMC iterations are negligible with respect to their standard deviations at lag 15.

First, we compute the marginal likelihoods for all four models. To obtain simulation-consistent estimates of the marginal likelihoods, the Monte Carlo sample sizes were taken to be $R = 10,000$ in all calculations. Furthermore, we use a procedure provided by Chib (1995) to compute the simulation standard errors for marginal likelihood estimates. The estimated $\ln m(D)$'s and the corresponding simulation standard errors in parentheses are -3097.76 (0.86), -2573.19 (0.76), -2111.34 (0.63), and -2393.63 (0.71) for the MVC, MVL, MVT10, and MVP models, respectively. Surprisingly, the MVT10 model is the best based on the marginal likelihoods. This finding is interesting as the best model is not one of the most commonly used models such as MVP and MVL.

Second, using 50,000 MCMC iterates after convergence, we compute the posterior estimates and 95% highest posterior density (HPD) intervals for the MVT10 model. To obtain the HPD intervals, we use the newly developed Monte Carlo methods of Chen and Shao (1999). The results for the MVT10 model are presented in Table 1. From Table 1, it can be observed that all 95% HPD intervals except for the one predictor, i.e., CLINS, associated with ordinal response PPSM do not contain 0, which implies that both PSA and GLEAS are statistically significant predictors for both ordinal responses while CLINS is statistically significant only for PECE. In addition, the posterior mean, the posterior standard deviation, and the 95% HPD interval for $\rho_{12} = \sigma_{12}/\sqrt{\sigma_{11}\sigma_{22}}$ are 0.741, 0.031, and (0.678, 0.780) respectively. These results indicate that there is a strong correlation between two ordinal responses PECE and PPSM.

Table 1 *Bayesian Estimates of the Regression Coefficients.*

		Posterior		95% HPD
Response Variable	Covariates	Mean	Std. Dev.	Intervals
PECE	PSA	0.225	0.036	(0.153, 0.295)
	GLEAS	0.148	0.039	(0.072, 0.227)
	CLINS	0.247	0.059	(0.128, 0.360)
PPSM	PSA	0.880	0.210	(0.504, 1.309)
	GLEAS	0.432	0.221	(0.004, 0.873)
	CLINS	0.380	0.316	(-0.231, 1.014)

Finally, we compute Bayesian estimates without using any historical information, i.e., $a_0 = 0$ with probability 1. To preserve space, the detailed outputs will not be reported here. However, all the posterior standard deviations for all parameters with $a_0 = 0$ are larger than those in Table 1. Therefore, when the sample size of the current study is small, it is important to incorporate the available historical information into analysis.

7 Concluding Remarks

As discussed in the introduction section, an independent ordinal response data problem is more difficult to deal with than the independent binary one. Therefore, intuitively a correlated ordinal response problem should be even harder. However, when we use a reparameterized SMMVN-link model, the resulting posterior computation actually becomes easier. More specifically, the most challenging computational problem for generating a correlation matrix from its conditional posterior distribution in the Bayesian correlated binary data model disappears. Instead, we require only generating a variance-covariance matrix from a Wishart distribution for the correlated ordinal model. In addition, the computational algorithm developed for computing marginal likelihoods for the correlated ordinal data model is also more efficient than the one for the correlated binary data model. In the binary case, a product kernel density estimation is required for estimating the marginal posterior density for the correlation matrix. As shown in Chen and Shao (1997b), the conditional marginal density estimation is much more efficient than the kernel estimation in general. Therefore, the use of the reparameterized SMMVN-link models is advantageous not only for the posterior estimation but also for the model comparisons. From Section 3, it is clear that the introduction of latent variables (w and w_0) and mixing variables (λ and λ_0) makes the posterior computation for a SMMVN-link model possible.

Acknowledgements

The authors thank Dr. Anthony V. D'Amico of the Joint Center for Radiation Therapy at Harvard Medical School for providing the prostate cancer data sets. Dr. Chen's research was supported by the National Science Foundation under Grant No. DMS-9702172.

(Received February, 2000. Revised June, 2000.)

References

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

Chen, M.-H. and Dey, D.K. (1998). Bayesian analysis of longitudinal binary data models using scale mixture of multivariate normals link functions. *Sankhyā*, Series A, **60**, 322–343.

Chen, M.-H., Ibrahim, J.G., and Yiannoutsos, C. (1999). Prior elicitation, variable selection, and Bayesian computation for logistic regression models. *Journal of the Royal Statistical Society*, Series B, **61**, 223–242.

Chen, M.-H., Manatunga, A.K. and Williams, C.J. (1998). Heritability estimates from human twin data by incorporating historical prior information. *Biometrics*, **54**, 1348–1362.

Chen, M.-H. and Schmeiser, B.W. (1998). Towards black-box sampling: a random-direction interior-point Markov chain approach. *Journal of Computational and Graphical Statistics*, **7**, 1–22.

Chen, M.-H. and Shao, Q.-M. (1997a). On Monte Carlo methods for estimating ratios of normalizing constants, *Annals of Statistics*, **25**, 1563–1594.

Chen, M.-H. and Shao, Q.-M. (1997b). Performance study of marginal posterior density estimation via Kullback-Leibler divergence. *Test, A Journal of the Spanish Society of Statistics and O.R.*, **6**, 321–350.

Chen, M.-H. and Shao, Q.-M. (1999). Monte Carlo estimation of Bayesian credible and HPD intervals. *Journal of Computational and Graphical Statistics*, **9**, 69–92.

Chib, S. (1995). Marginal likelihood from the Gibbs output. *Journal of the American Statistical Association*, **90**, 1313–1321.

Chib, S. and Greenberg, E. (1995). Understanding the Metropolis-Hastings algorithm. *The American Statistician*, **49**, 327–335.

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

Cowles, M.K. (1996). Accelerating Monte Carlo Markov chain convergence for cumulative-link generalized linear models. *Statistics and Computing*, **6**, 101–111.

Cowles, M.K. and Carlin, B.P. (1996). Markov Chain Monte Carlo convergence diagnostics: a comparative review. *Journal of the American Statistical Association*, **91**, 883–904.

Cowles, M.K., Carlin, B.P., and Connett, J.E. (1996). Bayesian tobit modeling of longitudinal ordinal clinical trial compliance data with nonignorable missingness. *Journal of the American Statistical Association*, **91**, 86–98.

Desjardin, A.M. (1997). Statistical inference on studies of adenocarcinoma of the prostate. *Unpublished Masters Thesis, Department of Mathematical Sciences, Worcester Polytechnic Institute, Worcester, MA*.

Gelfand, A.E., Smith, A.F.M. and Lee, T.M. (1992). Bayesian analysis of constrained parameter and truncated data problems using Gibbs sampling. *Journal of the American Statistical Association*, **87**, 523–532.

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*, (E. M. Keramidas, Eds.) Fairfax Station, VA: Interface Foundation of North America Inc., 571–578.

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

Ibrahim, J.G., Ryan, L.M., and Chen, M.-H. (1998). Use of historical controls to adjust for covariates in trend tests for binary data. *Journal of the American Statistical Association*, **93**, 1282–1293.

Kass, R.E. and Raftery, A.E. (1995). Bayes Factor. *Journal of the American Statistical Association*, **90**, 773–795.

Liang, K.-Y. and Zeger, S.L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika*, **73**, 13–22.

Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H. and Teller, E. (1953). Equations of state calculations by fast computing machines. *Journal of Chemical Physics*, **21**, 1087–1092.

Nandram, B. and Chen, M.-H. (1996). Accelerating Gibbs sampler convergence in the generalized linear models via a reparameterization. *Journal of Statistical Computation and Simulation*, **54**, 129–144.

O'Neill, R. (1971). Algorithm AS47-function minimization using a simplex procedure. *Applied Statistics*, **20**, 338–345.

Prentice, R.L. (1988). Correlated binary regression with covariates specific to each binary observation. *Biometrics*, **44**, 1033–1048.

Tierney, L. (1994). Markov chains for exploring posterior distributions (with discussion). *The Annals of Statistics*, **22**, 1701–1762.

Zeger, S.L. and Liang, K.-Y. (1986). Longitudinal data analysis for discrete and continuous outcomes. *Biometrics*, **42**, 121–130.