Journal of the Korean Statistical Society (2001), **30**: *4, pp 613–635*

Bayesian Analysis of a New Skewed Multivariate Probit Model for Correlated Binary Response Data†

Hea-Jung Kim[1]

ABSTRACT

This paper proposes a skewed multivariate probit model for analyzing a correlated binary response data with covariates. The proposed model is formulated by introducing an asymmetric link based upon a skewed multivariate normal distribution. The model connected to the asymmetric multivariate link, allows for flexible modeling of the correlation structure among binary responses and straightforward interpretation of the parameters. However, complex likelihood function of the model prevents us from fitting and analyzing the model analytically. Simulation-based Bayesian inference methodologies are provided to overcome the problem. We examine the suggested methods through two data sets in order to demonstrate their performances.

Keywords: Correlated binary data; MCMC method; latent variables approach; Monte Carlo accept-reject procedure; partial Bayes factor; skewed multivariate normal distribution.

1. Introduction

In many applications, one is confronted with multivariate binary response data: Response vectors of correlated binary variables, along with covariates, observed for each unit in a sample. The response vector may include repeated measurements of units on the same variable, as in longitudinal studies or in subsampling primary units. An example for the latter situation is common in genetic studies where a family is the cluster but responses are given by the members of the family. The response vector also arises in settings of multivariate measurements on a random cross-section of subjects where the response vector consists of different variables, e.g., different questions in an interview.

†This research was supported by KRF-2000-041-D00053

[1]Department of Statistics, Dongguk University, Seoul, 100-715, Korea

Models for the analysis of such data are available. Carey, Zeger, and Diggle (1993) and Glonek and McCullagh (1995) suggested the generalization of the binary logistic model to multivariate outcomes in conjunction with a particular reparameterized representation for the correlations among correlated binary data. Ashford and Sowden (1970) and Amemiya (1985) extended the binary probit model to get the multivariate probit model. Ochi and Prentice (1984) and Chib and Greenberg (1996) also considered estimation of the extended model.

Foregoing models are called as multivariate symmetric link models in the sense that they are obtained from generalizing the most common symmetric links, the probit and logit links. Although they are commonly used for regressing multivariate binary response on a set of covariates, sometimes they do not provide the best fit available for a given data set. In this case the link could be misspecified, which can yield substantial bias in the mean response estimates (see Czado and Santner 1992). In particular, rare event cases where respective marginal probabilities of one or more elements of the binary response vector approach 0 at different rate than they approach 1, the symmetric link models are known to be inappropriate (see Chen, Dey and Chao, 1999). The most intuitive approach to prevent such a misspecification is to construct a multivariate asymmetric link model whose marginal models embed a symmetric link into a wide parametric class of links. Many such parametric classes for univariate binary data have been proposed in the literature. For the related works, we refer Basu and Mukhopadhyay (2000), Chen, Dey and Chao (1999), Czado (1994) and references therein. However, unlike the univariate case, an asymmetric link model for the multivariate binary response data has not been seen yet.

The purpose of this paper is to propose a multivariate asymmetric link model. It is a variant of multivariate probit model considered by Chib and Greenberg (1996) and is motivated by using a skewed multivariate normal distribution of Azzalini and Valle (1996), where the marginal densities are scalar skew-normal. The format of the paper is thus as follows. Section 2 introduces a multivariate skewed probit model along with theoretical results necessary for constructing the model. In Section 3 we show that the suggested model is computationally attractive. In particular, by using a latent variable approach of Albert and Chib (1993), Markov chain Monte Carlo (MCMC) algorithms can be easily implemented to sample from the posterior distribution of the parameters of the model. Section 4 considers a Bayesian approach that enables us to do a model comparison between symmetric and asymmetric link models. In Section 5, two illustrative examples are given to examine and demonstrate the performances of the Bayesian inference methods considered in the previous sections. We finish this paper with brief concluding remarks in Section 6.

2. The Skewed Multivariate Probit Model

2.1. The Model

Let Y_{ij} denote a binary 1-0 response on the ith observation unit on jth variable and let $Y_i = (Y_{i1}, \ldots, Y_{ip})'$ denote the collection of correlated responses on all p binary response variables and $Y_1, \ldots, Y_n$ are independent. Suppose $x_{ij} = (x_{ij1}, \ldots, x_{ijp})'$ be the corresponding p-dimensional regression vector for $i = 1, \ldots, n$ and $j = 1, \ldots, p$. (Note that x_{ij1} may be 1, which correspond to an intercept.) Also let $\beta_j \in R^{k_j}$ be a k_j column vector of regression coefficients, and $\beta = (\beta_1', \ldots, \beta_p')' \in R^k$, $k = \sum_{j=1}^{p} k_j$.

In order to set up our skewed multivariate probit model, we introduce a p-dimensional independent latent random vectors $Z_i = (Z_{i1}, \ldots, Z_{ip})'$, $i = 1, \ldots, n$, such that

$$Y_{ij} = \begin{cases} 1 & \text{if} \quad Z_{ij} > 0 \\ 0 & \text{if} \quad Z_{ij} \leq 0, \end{cases} \tag{1}$$

and assume that

$$Z_i \sim N_p(X_i'\beta + \delta w_i, \Sigma), \tag{2}$$

$$w_i \overset{iid}{\sim} TN(0, 1), \tag{3}$$

a truncated standard normal with its density

$$g(w) = (2/\pi)^{1/2} \exp\{-w^2/2\}, \quad w > 0,$$

where $\delta = (\delta_1, \ldots, \delta_p)'$, a parameter vector and $X_i = diag(x_{i1}', \ldots, x_{ip}')$ is a $p \times k$ covariate matrix. In (2) we take $\Sigma = \{\rho_{ij}\}$ to be a correlation matrix to ensure the identifiability of the parameters. See Chib and Greenberg (1998) and Dey and Chen (1996) for detailed discussions.

The model defined by (1) through (3) has several attractive properties. First, when $\delta = 0$, it reduces to the standard multivariate probit model. Second, since the distribution of w_i is the truncated standard normal the conditional distribution of Z_i given w_i is a p-variate normal with mean $X_i\beta + \delta w_i$ and correlation matrix Σ, while the marginal distribution of Z_i is a skewed p-variate normal. The marginal probability density function of Z_i is given by the following Theorem.

Theorem 1. Under the skewed multivariate probit model, the distribution of Z_i is a skewed p-variate normal with its density

$$h(Z_i|\delta,\beta,\Sigma) = 2\phi_p(Z_i; X_i\beta, \Theta)\Phi(\alpha'(Z_i - X_i\beta)), \tag{4}$$

where

$$\alpha' = \delta'\Sigma^{-1}(1+\delta'\Sigma^{-1}\delta)^{-1/2}, \quad \Theta = \Sigma + \delta\delta' > 0,$$

and $\phi_p(Z_i; X_i\beta, \Theta)$ and $\Phi(\cdot)$ denote the probability density of p-variate normal with mean $X_i\beta$ and covariance matrix Θ and distribution function of the standard normal, respectively.

Proof. Let $U_i = Z_i - X_i\beta$, then using a standard method for transformations of random variables, the density of U_i at point $u_i \in R^p$ is

$$\begin{aligned}
&h(u_i|\delta,\Sigma)\\
&= \frac{2}{(2\pi)^{p/2}|\Sigma|^{1/2}}\int_0^\infty \phi(w_i)\exp\left\{-\frac{(u_i-\delta w_i)'\Sigma^{-1}(u_i-\delta w_i)}{2}\right\}dw_i\\
&= \frac{2\exp\left\{-u_i'\Sigma^{-1}u_i/2 + (u_i'\Sigma\delta)^2/[2(1+\delta'\Sigma^{-1}\delta)]\right\}}{(2\pi)^{p/2}|\Sigma|^{1/2}(1+\delta'\Sigma^{-1}\delta)^{1/2}}\\
&\times \int_0^\infty (1+\delta'\Sigma^{-1}\delta)^{1/2}\phi\left((1+\delta'\Sigma^{-1}\delta)^{1/2}(w_i - \delta'\Sigma^{-1}u_i/(1+\delta'\Sigma^{-1}\delta)\right)dw_i\\
&= \frac{2\exp\left\{-[u_i'\Sigma^{-1}u_i - (u_i'\Sigma^{-1}\delta)^2/(1+\delta'\Sigma^{-1}\delta)]/2\right\}}{(2\pi)^{p/2}|\Sigma|^{1/2}(1+\delta'\Sigma^{-1}\delta)^{1/2}}\Phi\left(\frac{\delta'\Sigma^{-1}u_i}{(1+\delta'\Sigma^{-1}\delta)^{1/2}}\right),
\end{aligned}$$

where $\phi(\cdot)$ is the probability density function of the standard normal. Since

$$(\Sigma+\delta\delta')^{-1} = \Sigma^{-1} - \frac{\Sigma^{-1}\delta\delta'\Sigma^{-1}}{1+\delta'\Sigma^{-1}\delta},$$

we see that

$$u_i'\Sigma^{-1}u_i - \frac{(u_i'\Sigma^{-1}\delta)^2}{1+\delta'\Sigma^{-1}\delta} = u_i'\Theta^{-1}u_i, \tag{5}$$

where

$$\Theta = \Sigma + \delta\delta' \text{ and } |\Theta| = |\Sigma + \delta\delta'| = |\Sigma|(1+\delta'\Sigma^{-1}\delta). \tag{6}$$

Moreover, $\Theta > 0$ because $\eta'\Theta\eta = \eta'\Sigma\eta + (\eta'\delta)^2 > 0$ for all $\eta \neq 0$ and $\Sigma > 0$. Replacing (5) and (6) in the above joint density of U_i and transforming U_i to Z_i via the relation $U_i = Z_i - X_i\beta$, we have the result.

Corollary 1. Let $Z_i = (Z_{i1}, \ldots, Z_{ip})'$, and let σ_w^2 and $\sigma_{Z_j}^2$ be the variances of w_i and Z_{ij}. If $\mu_w^{(3)}$ denote the standardized third moment of w_i; that is $\mu_w^{(3)} = E\{[w_i - E(w_i)]/\sigma_w\}^3$. Then marginal density of Z_{ij} is

$$f(z_{ij}|\delta_j, \beta_j) = 2\phi(z_{ij}; x_{ij}'\beta_j, 1+\delta_j^2)\Phi\left(\frac{\delta_j(z_{ij} - x_{ij}'\beta_j)}{(1+\delta_j^2)^{1/2}}\right), \quad j = 1, \ldots, p, \tag{7}$$

and its standardized third moment $\mu_{Z_j}^3$ is given by

$$\mu_{Z_j}^3 = E\left(\frac{Z_{ij} - EZ_{ij}}{\sigma_{Z_j}}\right)^3 = \frac{\delta_j^3 \sigma_w^3 \mu_w^{(3)}}{\sigma_{Z_j}^3}, \tag{8}$$

where $\phi(z_{ij}; x_{ij}'\beta_j, 1+\delta_j^2)$ is the pdf of $N(x_{ij}'\beta_j, 1+\delta_j^2)$.

Proof. Let $U_i = Z_i - X_i\beta$, then applying the result by Azzalini and Valle (1996), we have the moment generating function of U_i;

$$M_{U_i}(t) = 2\exp\{t'\Theta t/2\}\Phi\{t'\delta\}.$$

Therefore, after substantial reduction, the moment generating function of Z_i is given by

$$M_{Z_i}(t) = 2\exp\{t'X_i\beta + t'\Theta t/2\}\Phi\{t'\delta\}.$$

This yields the marginal moment generating function of Z_{ij}

$$M_{Z_{ij}}(t_j) = 2\exp\{t_j x_{ij}'\beta_j + (1+\delta_j^2)t_j^2/2\}\Phi\{t_j\delta_j\}, \quad j = 1, \ldots, p$$

which is the same moment generating function as that of Z_{ij} having the pdf (7). Also note that (7) is equivalent to the univariate skewed normal density given by Chen, Dey and Shao (1999). Thus (8) is immediate from the result of Chen, Dey and Shao (1999).

Remark 1. Under the skewed multivariate probit model, the probability $p_{ij} = Pr(Y_{ij} = 1) = Pr(Z_{ij} > 0)$ approaches 0 at the same rate as it approaches 1 for $\delta_j = 0$. When $\delta_j > 0$, the probability p_{ij} approaches 1 at a faster rate than it approaches 0. The opposite result is obtained when $\delta_j < 0$, where $1 - p_{ij} = Pr(Y_{ij} = 0)$.

Proof. Since $\mu_w^{(3)} > 0$ (a half-normal distribution defined by (3)), the sign of (8) mainly depends on that of δ_j, i.e. the distribution of Z_{ij} is skewed to the right (or the left) when $\delta_j > 0$ (or $\delta_j < 0$).

From Remark 1, we see that the skewed multivariate probit model (implying the univariate skewed probit model for each Z_{ij} marginally) accounts for different approach rates of p_{ij}'s by differing the values of δ_j's (including some of zero δ_j's) and takes care of correlation between Z_{ij}'s. Therefore, if multivariate binary response data marginally take a skewed link model as a true model, the symmetric multivariate probit model, will be either underfitted or overfitted.

Corollary 2. Suppose A_j is an interval defined by the value of y_{ij} so that $A_j = (-\infty, 0)$ if $y_{ij} = 0$ and $A_j = (0, \infty)$ if $y_{ij} = 1$ for $i = 1, \ldots, n$ and $j = 1, \ldots, p$. Then respective joint and marginal probabilities that $Y_i = y_i$ and $Y_{ij} = y_{ij}$, $j = 1, \ldots, p$, conditioned on parameters, β, Σ, δ and a set of covariates x_{ij} are

$$\begin{aligned} Pr(Y_i = y_i|\beta, \Sigma, \delta) &= \int_{A_p} \cdots \int_{A_1} h(Z_i|\beta, \Sigma, \delta) dZ_i \\ &= 2\int_{\Omega_p} \cdots \int_{\Omega_1} \phi_p(t; 0, \Theta)\Phi(\alpha' t) dt \quad \text{and} \\ Pr(Y_{ij} = y_{ij}|\beta_j, \delta_j) &= \int_{A_j} f(z_{ij}|\beta_j, \delta_j) dz_{ij} \end{aligned}$$

$$= \begin{cases} \int_0^\infty \Phi(x_{ij}'\beta_j + \delta_j w_i) g(w_i) dw_i, & \text{if} \quad y_{ij} = 1, \\ \int_0^\infty [1 - \Phi(x_{ij}'\beta_j + \delta_j w_i)] g(w_i) dw_i, & \text{if} \quad y_{ij} = 0, \end{cases}$$

where

$$\Omega_j = \begin{cases} (-x_{ij}'\beta_j, \infty) & \text{if} \quad y_{ij} = 1 \\ (-\infty, -x_{ij}'\beta_j) & \text{if} \quad y_{ij} = 0, \end{cases}$$

Proof. We see that the statements are immediate from Theorem 1 and Corollary 1.

Analytic evaluation of the probability $Pr(Y_i = y_i|\beta, \Sigma, \delta)$ is not available. Instead, in Section 3, we will give a Monte Carlo method for the evaluation.

2.2. The Likelihood Function

Let $\mathbf{Y} = (Y_1, \ldots, Y_n)$ and $\mathbf{X} = (X_1, \ldots, X_n)$, and let $D_{obs} = (n, \mathbf{Y}, \mathbf{X})$ denote the observed data then, from Corollary 2, the likelihood function for the skewed multivariate probit model is given by

$$L(\beta, \delta, \Sigma|D_{obs}) = \prod_{i=1}^{n} \int_{A_p} \cdots \int_{A_1} h(Z_i|\beta, \Sigma, \delta) dZ_i \; I(\varrho \in \mathbf{A}), \tag{9}$$

where $\varrho \equiv (\rho_{12}, \rho_{13}, \ldots, \rho_{p-1,p})'$, the $s = p(p-1)/2$ free parameter vector in the correlation matrix Σ, so that $\mathbf{A}$ denotes a convex solid body in the hypercub $[-1,1]^s$ that leads to a proper correlation matrix (see Rousseeuw and Molenberghs (1994) for more on the shape of correlation matrices). The likelihood function shows that the skewed multivariate normal distribution, which allows for flexible modeling of the correlation structure and rates for p_{ij} approaching to 1, induces the problem of evaluating the likelihood function.

Recently, developments in Markov chain Monte Carlo method has given rise to reasonably effective method for estimation the model (cf. Gelfand and Smith (1990), Chib and Greenberg (1996) and Albert and Chip (1993)).

Let $\mathbf{Z} = (Z_1, \ldots, Z_n)$ and $W = (w_1, \ldots, w_n)$, and let $D = (n, \mathbf{Y}, \mathbf{X}, \mathbf{Z}, W)$ denote complete data. Then complete data likelihood function of the parameters (β, Σ, δ) can be written as

$$
\begin{aligned}
&L(\beta, \delta, \Sigma | D) \\
&\propto \prod_{i=1}^{n} \Bigg[|\Sigma|^{-1/2} \exp\left\{ -\frac{1}{2}(Z_i - X_i\beta - \delta w_i)'\Sigma^{-1}(Z_i - X_i\beta - \delta w_i) \right\} \qquad (10) \\
&\times \prod_{j=1}^{p} \{ I(z_{ij} > 0) I(y_{ij} = 1) + I(z_{ij} \le 0) I(y_{ij} = 0) \} g(w_i) \Bigg] I(\varrho \in \mathbf{A}),
\end{aligned}
$$

The representation in (10) will ease computation. We demonstrate this idea and the role of the auxiliary variables Z_i and w_i in the MCMC algorithms in Section 3.

3. Markov Chain Monte Carlo Method

3.1. Posterior Simulation

Suppose that we consider a prior density $p(\beta, \delta, \varrho)$ on the parameters of a given multivariate skewed probit model and assume that β, δ, and ϱ are independent in priori, so that

$$\pi(\beta, \delta, \varrho) \propto \phi_k(\beta; \beta_0, B_0^{-1}) \phi_p(\delta; \delta_0, D_0^{-1}) \phi_s(\varrho; \varrho_0, G_0^{-1}), \ \varrho \in \mathbf{A},$$

where $s = p(p-1)/2$ and ϕ_s denotes the density of a s-variate normal distribution which is truncated to $\mathbf{A}$. The hyperparameters $(\beta_0, B_0, \delta_0, D_0, \varrho_0, G_0)$ are chosen to reflect the available prior information. The location of the prior information is controlled by the vectors β_0, δ_0 and ϱ_0 and strength by the precision matrices B_0, D_0 and G_0.

Our basic approach for fitting the skewed multivariate probit model by MCMC method is due to Albert and Chip (1993). In this approach, the parameter space is augmented by latent data $\mathbf{Z}$ and W. To sample from the posterior distribution $p(\beta, \Sigma, \delta | D_{obs})$, we introduce the latent variables $\mathbf{Z} = (Z_1, \dots, Z_n$ and $W = (w_1, \dots, w_n)$. Then the joint posterior distribution for $\Delta = (\beta, \delta, \varrho, \mathbf{Z}, W)$ is given by

$$
\begin{aligned}
& p(\Delta | \mathbf{X}, \mathbf{Y}) \\
\propto \; & \pi(\beta, \delta, \varrho) \left[\prod_{i=1}^{n} |\Sigma|^{-1/2} \exp\left\{ -\frac{1}{2}(Z_i - X_i\beta - \delta w_i)'\Sigma^{-1}(Z_i - X_i\beta - \delta w_i) \right\} \right. \\
\times \; & \left. \prod_{j=1}^{p} \{ I(z_{ij} > 0) I(y_{ij} = 1) + I(z_{ij} \le 0) I(y_{ij} = 0) \} g(w_i) \right] I(\varrho \in \mathbf{A}). \qquad (11)
\end{aligned}
$$

Given a set full conditional posterior distributions obtained from (11), to sample the identified parameters (β, δ, ϱ) in a MCMC method with augmentation, we need to iterate sampling of each element of Δ on the following steps a large number of times.

From (11), we see that the distributions in the first step of the MCMC sampling are the univariate normal distributions

$$
\begin{aligned}
& [w_i | Y_i, Z_i, \beta, \delta, \varrho] \sim \\
& N\left(\frac{\delta'\Sigma^{-1}(Z_i - X_i\beta)}{1 + \delta'\Sigma^{-1}\delta}, (1 + \delta'\Sigma^{-1}\delta)^{-1} \right) I(w_i > 0), \;\; i = 1, \dots, n, \qquad (12)
\end{aligned}
$$

truncated to the region $R_i = \{w_i; w_i > 0\}$, $i = 1, \dots, n$.

A possible complication of the sampling could be that from truncated normal distribution. This can be easily resolved by the algorithm of Devroye (1986). The full conditional distribution of Z_i is a truncated multivariate normal

$$
\begin{aligned}
& [Z_i | w_i, Y_i, \beta, \delta, \varrho] \sim \\
& N_p(X_i\beta + \delta w_i, \Sigma) \prod_{j=1}^{p} \{ I(z_{ij} > 0) I(y_{ij} = 1) + I(z_{ij} < 0) I(y_{ij} = 0) \}, \;\; i = 1, \dots, n. \qquad (13)
\end{aligned}
$$

This distribution can be simulated by the method of Geweke (1991), composing a cycle of p Gibbs steps through the components z_{ij} of Z_i. Instead of sampling z_{ij} in this manner, the entire vector Z_i can be sampled from $[Z_i | w_i, Y_i, X_i, \beta, \delta, \varrho]$ by the accept-reject method of Albet and Chip (1993).

The next two distributions for the MCMC sampling are

$$[\beta|\mathbf{Z}, W, \delta, \varrho] \sim N_k(\hat{\beta}, B^{-1}), \tag{14}$$

$$[\delta|\mathbf{Z}, W, \beta, \varrho] \sim N_p(\hat{\delta}, D^{-1}), \tag{15}$$

where

$$\hat{\beta} = B^{-1}(B_0\beta_0 + \sum_{i=1}^{n} X_i'\Sigma^{-1}(Z_i - \delta w_i)),\ \ B = B_0 + \sum_{i=1}^{n} X_i'\Sigma^{-1}X_i,$$

$$\hat{\delta} = D^{-1}(D_0\delta_0 + \sum_{i=1}^{n} w_i\Sigma^{-1}(Z_i - X_i\beta)),\ \ D = D_0 + \sum_{i=1}^{n} w_i^2\Sigma^{-1},$$

and $k = \sum_j^p k_j$.

Finally, the full conditional density of the unique elements of Σ is

$$p(\varrho|\mathbf{Z}, \mathbf{X}, W, \beta, \delta) \propto \phi_s(\varrho; \varrho_0, G_0^{-1}) f(\mathbf{Z}|\mathbf{X}, W, \beta, \delta, \varrho) I(\varrho \in \mathbf{A}). \tag{16}$$

where

$$\begin{aligned} f(\mathbf{Z}|\mathbf{X}, W, \beta, \delta, \varrho) &= (2\pi)^{-np/2}|\Sigma|^{-n/2} \\ &\times \ \exp\left\{-\frac{1}{2}\sum_{i=1}^{n}(Z_i - X_i\beta - \delta w_i)'\Sigma^{-1}(Z_i - X_i\beta - \delta w_i)\right\}. \end{aligned}$$

The analysis of this density and search for suitable bounds and dominating functions is difficult. Nevertheless this posterior density can be sampled by use of Metropolis-Hastings(MH) algorithm with a proposal density described by the random walk chain. $\varrho' = \varrho + h$ where ϱ' is the candidate value, ϱ is the current value and h is a zero mean increment vector. It is convenient to assume that h follows a symmetric distribution, such as multivariate normal, so that

$$\alpha(\varrho, \varrho^t) = \min\left\{1, \frac{f(\varrho^t)}{f(\varrho)}\right\},$$

where $f(\varrho) = \phi_s(\varrho; \varrho_0, G_0^{-1}) f(\mathbf{Z}|\mathbf{X}, W, \beta, \delta, \varrho)$. Then move to ϱ^t with probability $\alpha(\varrho, \varrho^t)$ and stay at ϱ with probability $1 - \alpha(\varrho, \varrho^t)$. Note that the proposal density need not enforce the positive definiteness constraint, because that constraint is part of $f(\varrho)$. In other words, when Σ^t is not positive definite or ϱ' is not element of $\mathbf{A}$, the conditional posterior is zero, and hence proposal value is rejected with certainty. See Chib and Greenberg (1995) for turning of the covariance matrix of the proposal density that guarantees proper acceptance rate. In case the dimension of Σ is large, it is best to partition ϱ into blocks and to apply the Metropolis-Hastings algorithm in sequence, cycling through the various blocks (cf. Chib and Greenberg 1996).

3.2. Posterior Probabilities

We need to calculate the posterior predicted probability of the observed choice of each individual:

$$Pr(Y_i = y_i|\beta, \Sigma, \delta) = \int_{A_p} \cdots \int_{A_1} h(Z_i|\beta, \Sigma, \delta) dZ_i, \ i = 1, \ldots, n, \tag{17}$$

where A_j, $j = 1, \ldots, p$, is the interval $A_j = (-\infty, 0)$ if $y_{ij} = 0$ and $A_j = (0, \infty)$ if $y_{ij} = 1$.

This integral can be accurately estimated by drawing a large number of Z_i values from a Monte Carlo accept-reject procedure by iterating on the following steps for $k = 1, \ldots, M$.

(Algorithm 1): Given the Bayes estimates β^*, δ^*, and Σ^*

- `Step 1:` `Simulate` $w_i^{(k)}$ `from` $TN(0,1)I(w_i > 0)$`, the truncated standard normal.`
- `Step 2:` `Simulate` $Z_i^{(k)}$ `from` $N_p(X_i\beta^* + \delta^* w_i^{(k)}, \Sigma^*)$`;`
- `Step 3:` `Calculate` $Pr(Y_i = y_i|Z_i^{(k)}, w_i^{(k)}, \beta^*, \delta^*, \Sigma^*)$.

The probability in Step 3 is 1 or 0 depending on whether $Z_i^{(k)}$ corresponds to the constraints (in terms of A_j's) imposed by y_i. Then from the law of large numbers,

$$M^{-1} \sum_{k=1}^{M} Pr(Y_i = y_i|Z_i^{(k)}, w_i^{(k)}, \beta^*, \delta^*, \Sigma^*) \rightarrow Pr(Y_i = y_i|\beta^*, \delta^*, \Sigma^*). \tag{18}$$

For this method to be effective, M must be large, but ensuring this is relatively simple because the computation is done at only one point $(\beta^*, \delta^*, \Sigma^*)$. Moreover, Step 1 and Step 2 require only the generations of Gaussian samples. As a by-product, estimation of the marginal predicted posterior probability of each component of Y_i, i.e.

$$Pr(Y_{ij} = y_{ij}|\beta_j, \delta_j) = \int_{A_j} f(z_{ij}|\beta_j, \delta_j) dz_{ij}, \ \ i = 1, \ldots, n; \ j = 1, \ldots, p, \tag{19}$$

can be obtained from the same Monte Carlo accept-reject procedure if we modify Step 2 and Step 3 to the marginal distributions of Z_{ij}'s and Y_{ij}'s. For this estimation, a numerical calculation (using a computer package such as *Mathematica*) that directly calculates $\int_{A_j} f(z_{ij}|\beta_j^*, \delta_j^*) dz_{ij}$ is also available.

4. Model Comparison

In Section 2 we proposed a multivariate skewed probit link model for multivariate binary response data, in which asymmetry of the link is determined by δ in (2). Therefore it is of practical interest to compare models formulated by different choices of δ in (2), symmetric (usual) multivariate probit link model with $\delta = 0$ and an asymmetric one with $\delta \neq 0$. To this end, we propose a conditional Bayes factor approach (see, e.g. Geweke 1996) to perform the model comparison. The approach can be made by modifying the MCMC sampling step of δ in Subsection 3.1.

The modification is as follows. Under the assumption that investigator's prior distributions for δ_j's, β and ϱ are mutually independent, we change the prior distribution of δ_j, $j = 1, \ldots, p$. With prior probability $q_j = P(\delta_j = 0)$

$$d\Pi(\delta_j) = q_j dH(\delta_j) + (1 - q_j)\phi(\delta_j; \delta_{0j}, d_{0j}^2), \;\; j = 1, \ldots, p, \tag{20}$$

where $\Pi(\cdot)$ denotes the prior c.d.f. of δ_j; $H(\delta_j) = 0$ if $\delta_j < 0$ and $H(\delta_j) = 1$ if $\delta_j \geq 0$. Here d_{0j}^2 denotes jth diagonal element of D_0^{-1}. We set the prior distributions of β and ϱ to be the same as those used in Subsection 3.1.

To apply the conditional Bayes factor approach, we need following modification in the fourth step of the MCMC sampling in Subsection 3.1(i.e. posterior sampling of δ_j) : Given $\delta_\ell (\ell \neq j)$, $\mathbf{Z}$, W, β and ϱ, define $U_i = Z_i - X_i\beta$ so that

$$U_i = \delta w_i + \varepsilon_i, \;\; \varepsilon_i \overset{iid}{\sim} N_p(0, \Sigma), \tag{21}$$

$i = 1, \ldots, n$. Then the conditional distribution of δ_j follows from the simplified model,

$$U_{ij} | U_{i(\backslash j)} \overset{iid}{\sim} N(w_i \delta_j + \theta_{ij}, \sigma_{(j)}^2),$$

where $\theta_{ij} = \Sigma_{j(\backslash j)} \Sigma_{(\backslash j)}^{-1} (U_{i(\backslash j)} - w_i \delta_{(\backslash j)})$, and $\sigma_{(j)}^2 = 1 - \Sigma_{j(\backslash j)} \Sigma_{(\backslash j)(\backslash j)}^{-1} \Sigma_{(\backslash j)j}$ denote respective conditional mean and variance of U_{ij} obtained from (21).

The likelihood function kernel is

$$\exp\{-\frac{1}{2\sigma_{(j)}^2} \sum_{i=1}^{n} (U_{ij} - \delta_j w_i - \theta_{ij})^2\} \psi_{(\backslash j)},$$

where

$$\psi_{(\backslash j)} = \exp\{-\frac{1}{2}(U_{i(\backslash j)} - \delta_{(\backslash j)} w_i)' \Sigma_{(\backslash j)(\backslash j)}^{-1} (U_{i(\backslash j)} - \delta_{(\backslash j)} w_i)\}.$$

Conditional on $\delta_j = 0$ the value of the kernel is

$$\exp\{-\frac{1}{2\sigma_{(j)}^2}\sum_{i=1}^{n}(U_{ij}-\theta_{ij})^2\}\psi_{(\backslash j)}. \tag{22}$$

Conditional on $\delta_j \neq 0$ the corresponding kernel density for δ_j is

$$\begin{aligned}
& (2\pi)^{-p/2}d_{0j}^{-1}\exp\{-\frac{1}{2\sigma_{(j)}^2}\sum_{i=1}^{n}(U_{ij}-w_i\delta_j-\theta_{ij})^2-\frac{1}{2d_{0j}^2}(\delta_j-\delta_{0j})^2\}\psi_{(\backslash j)} \\
= \; & (2\pi)^{-p/2}d_{0j}^{-1}\exp\{-\frac{1}{2}[d_j(\delta_j-\hat{\delta}_j)^2+\frac{\sum_{i=1}^{n}(U_{ij}-\theta_{ij})^2}{\sigma_{(j)}^2} \\
+ \; & \frac{\delta_{0j}^2}{d_{0j}^2}-d_j\hat{\delta}_j^2]\}\psi_{(\backslash j)},
\end{aligned} \tag{23}$$

where

$$\hat{\delta}_j = d_j^{-1}[\delta_{0j}/d_{0j}^2+\sum_{i=1}^{n}w_i(U_{ij}-\theta_{ij})/\sigma_{(j)}^2],\;\; d_j = 1/d_{0j}^2+\sum_{i=1}^{n}w_i^2/\sigma_{(j)}^2.$$

Thus the conditional posterior distribution for $\delta_j \neq 0$ is

$$[\delta_j|\mathbf{Z},W,\beta,\sigma] \sim N(\hat{\delta}_j, d_j^{-1}). \tag{24}$$

To calculate the conditional Bayes factor, it is necessary to integrate (23) over δ_j which yields conditional marginal likelihood

$$d_j^{-1/2}d_{0j}^{-1}\psi_{(\backslash j)}\exp\{-1/2[\sum_{i=1}^{n}(U_{ij}-\theta_{ij})^2/\sigma_{(j)}^2+\delta_{0j}^2/d_{0j}^2-d_j\hat{\delta}_j^2]\}. \tag{25}$$

Comparing this marginal likelihood to (22), we have the conditional Bayes factor in favor of $\delta \neq 0$, versus $\delta = 0$, that is

$$BF_j^c = d_j^{-1/2}d_{0j}^{-1}\exp\{-1/2[\delta_{0j}^2/d_{0j}^2-d_j\hat{\delta}_j^2]\}\;\; j=1,\ldots,p. \tag{26}$$

To draw δ_j from its conditional distribution, the conditional posterior probability that $\delta_j = 0$ is computed from the conditional Bayes factor (26):

$$q^c = q_j/\{q_j+(1-q_j)BF_j^c\}. \tag{27}$$

Based on a comparison of this probability with a drawing from the uniform distribution on [0, 1], the choice $\delta_j = 0$ or $\delta_j \neq 0$ is made. Therefore modifying the fourth step of the basic algorithm, we have the following algorithm for the model comparison.

(Algorithm 2): Algorithm for model comparison

- Sample w_i, $i = 1, \ldots, n$, from the conditional posterior (12);
- Sample z_{ij}, $j = 1, \ldots, p$ and $i = 1, \ldots, n$, from the conditional posterior (13);
- Sample β from the conditional posterior (14);
- The parameters $\delta_1, \ldots, \delta_p$ are drawn in succession so that, for δ_j, compute q_j^c from BF_j^c and generate u from $U(0,1)$, if $u \leq q_j^c$, set $\delta_j = 0$. Else, sample δ_j from $N(\hat{\delta}_j, d_j^{-1})$;
- Sample ϱ from the Metropolis-Hastings algorithm in Subsection 3.1.

The model comparison could be done in the obvious way, by recording the indicator variables for the model corresponding to the nonzero δ_j's at the end of each iteration.

5. Illustrative Examples

We now take up two examples of generalized multivariate regression with correlated binary responses data. The objectives in these examples are to (a) compare the proposed skewed model with the multivariate probit model by Amemiya (1985); (b) illustrate the numerical accuracy of the algorithms of the previous sections; and (c) study the relation between prior and posterior distributions in the proposed model.

5.1. Example 1: Artificial Data

In this example we consider a simulation with the skewed multivariate probit model. Our primary aims here are to examine the numerical accuracy of the algorithms and to study the relation between prior and posterior distributions in the model. We generate multiple simulated datasets from the following 4 dimensional binary response skewed probit model :

$$Z_i \sim N_4(X_i\beta + \delta w_i, \Sigma), \quad Y_{ij} = I(Z_{ij} > 0), \tag{28}$$

where $\Sigma = \{\rho_{jj'}\}$ with $\rho_{jj'} = 0.5$, for $j \neq j'$, a intraclass correlation matrix, $w_i \sim TN(0,1)I(w_i > 0)$, and $X_i = diag\{x_{i1}, x_{i2}, x_{i3}, x_{i4}\}$.

First, for each $j(j = 1,\ldots,4)$, we independently generate $x_{ij} \sim N((-1)^{j+1}, 3), j = 1,\ldots,4$, to obtain X_i $i = 1,\ldots,n$, and take $\beta = (1,-1,2,-2)'$. Then using $Z_i = (Z_{i1},\ldots,Z_{i4})'$, we generate independent binary response variables, Y_i, $i = 1,\ldots,n$ from (28) with $\delta = (5,-5,2,-2)'$. In the analysis, we considered various sample size n and different set of values of the hyperparameters, $B_0 = b_1 I_4$, $D_0 = b_2 I_4$ and $G_0 = b_3$ (for Σ is restricted to a intraclass covariance structure). We set the other hyperparameters, β_0, δ_0 and ϱ_0, to be zero vectors.

The posterior distributions of the parameters are obtained by applying the posterior simulation (in Subsection 3.1) for 10,000 cycles beyond 1000 burn-in iterations. Many standard diagnostic measures (see. e.g., Cowles and Carlin, 1996) have been computed to monitor convergence by using " CODA Output Analysis Menu" by Best et al. (1996). Those indicated rapid convergence within 1000 burn-in iterations. For each parameter, trace of the Markov chains obtained from twelve different starting points, appeared to settle to the same (or similar) distribution within 1000 iterations. Gelman and Rubin shrinkage factor also converged to 1 within 1000 iterations. Furthermore, the autocorrelations of each parameter from the MCMC algorithm disappeared at lag less than 3. In the Metropolis-Hasting step of the algorithm, we let the random walk proposal dens-

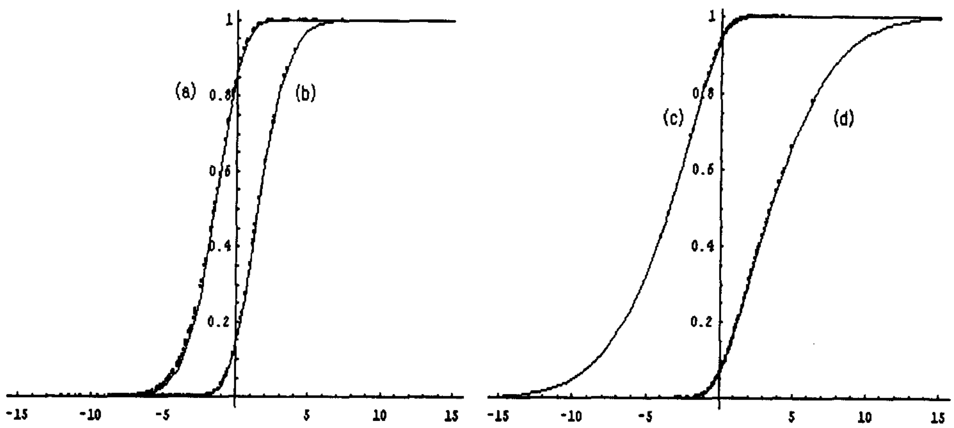

Figure 1. Probability Plots (overlapped plots of $(x'_{ij}\beta_j, p_{ij})$ and $(x'_{ij}\beta_j, \hat{p}_{ij})$): Respective solid lines in (a), (b), (c), and (d) correspond to the true probabilities, p_{i1}, p_{i2}, p_{i3}, and p_{i4}. The dotted values in (a), (b), (c), and (d) are estimated values of p_{i1}, p_{i2}, p_{i3}, and p_{i4}, respectively.

ity be a normal density with 0 mean and covariance matrix τ, where τ plays role of a tuning parameter for ensuring the acceptance rate being around .5 as recommended by Robert, Gelman and, Gilks (1994). The value of τ for the sample size $n = 20$ was 1.5 and that of sample size $n = 100$ was 2.0.

The posterior distributions are summarized in Table 1, where we report the posterior mean (the average of simulated values), numerical standard error of the posterior mean (computed by batch means method), the standard deviation (the standard deviation of the simulated values), the posterior median, the lower 2.5 and upper 97.5 percentiles of the simulated values.

From Table 1 it is clear that, regardless of the particular prior distributions, the posterior simulation has accurately produced a posterior distribution concentrated on the values that generated data. Three systematic effects of the prior distributions to the posterior distributions are evident. First, each posterior distribution of component of δ is spread out, and its 95% credible interval dose not include 0, which is evident for the skewed multivariate probit model.

Second, for each n, increase in values of b_ℓ, $\ell = 1, 2$, slightly decreases 95% credible intervals of the parameters. This is due to the fact that the values of the hyperparameters effect fairly small in the estimation compared to the mass of the likelihood function. Finally, for each set of values b_ℓ, $\ell = 1, 2$, increase in sample size n produces more accurate marginal posterior distribution of each parameter in the sense that, for $n = 100$, each marginal distribution has more concentrated mass around the true value than it does for $n = 20$.

To check the fit of the skewed multivariate probit model, we compare the posterior estimate of marginal probabilities (predicted posterior means), $\hat{p}_{ij} = \int_{-\infty}^{\infty} \Phi(x_{ij}'\beta_j^* + \delta_j^* w_i) g(w_i) dw_i$ to true probabilities $p_{ij} = \int_{-\infty}^{\infty} \Phi(x_{ij}'\beta_j + \delta_j w_i) g(w_i)$ dw_i, $j = 1, \ldots, 4$, for the model (28). Here $x_{ij}'\beta_j^*$ and δ_j^* denote Bayesian estimates. A part of the results (with $n = 100$ and $(b_1, b_2, b_3) = (.01, .01, 0.3)$) is shown in Figure 1. From Figure 1, it can be seen that the estimated probabilities are fairly close to the true ones. The figures also highlight the performance of the skew multivariate probit model; they show that, in case of fitting the symmetric multivariate probit model to the dataset, the marginal probability plots symmetric about 0.5 at $x_{ij}'\beta_j = 0$ would overestimate (or underestimate) the true probabilities.

Table 1. Summaries of the Posterior Distributions for Simulated Model

n	Para.	Mean	Mum. SE	SD	Median	Lower	Upper
			(b_1, b_2, b_3)=(.01, .01, 0.3)				
20	β_1	.9588	.0007	.0725	.9584	.8412	1.0783
	β_2	-1.0604	.0009	.0976	-1.0599	-1.2234	-.9013
	β_3	1.9630	.0006	.0642	1.9635	1.8554	2.0677
	β_4	-2.0857	.0007	.0758	-2.0851	-2.2111	-1.9606
	δ_1	5.2713	.0021	.2166	5.2706	4.9189	5.6262
	δ_2	-4.3462	.0025	.2519	-4.3474	-4.7576	-3.9312
	δ_3	2.3958	.0027	.2728	2.3956	1.9516	2.8463
	δ_4	-1.5088	.0032	.3277	-1.5117	-2.0533	-.9614
	$\rho_{jj'}$	.5202	.0009	.0929	.5322	.3501	.6544
			(b_1, b_2, b_3)=(.1, .1, 0.3)				
20	β_1	1.1475	.0011	.1141	1.1473	.9620	1.3389
	β_2	-1.0670	.0008	.0821	-1.0677	-1.2020	-.9328
	β_3	1.9707	.0007	.0726	1.9702	1.8525	2.0919
	β_4	-2.1140	.0006	.0609	-2.1152	-2.2124	-2.0126
	δ_1	4.8340	.0025	.2505	4.8336	4.4198	5.2529
	δ_2	-4.9436	.0030	.2954	-4.9457	-5.4265	-4.4515
	δ_3	1.8013	.0027	.2739	1.8003	1.3467	2.2494
	δ_4	-1.8554	.0030	.2994	-1.8536	-2.3476	-1.3664
	$\rho_{jj'}$	.4419	.0009	.0890	.4444	.2928	.5818
			(b_1, b_2, b_3)=(.01, .01, 0.3)				
100	β_1	.9619	.0004	.0417	.9615	.8936	1.0320
	β_2	-.9956	.0004	.0429	-.9951	-1.0670	-.9250
	β_3	1.9261	.0003	.0415	1.9267	1.8944	2.0603
	β_4	-2.0210	.0002	.0276	-2.0208	-2.0667	-1.9744
	δ_1	4.9458	.0010	.1046	4.9457	4.7735	5.1170
	δ_2	-4.9979	.0012	.1185	-4.9991	-5.1923	-4.8020
	δ_3	2.1165	.0013	.1249	2.1154	1.9109	2.3271
	δ_4	-1.9892	.0014	.1376	-1.9886	-2.2162	-1.7652
	$\rho_{jj'}$	.5227	.0004	.0395	.5252	.4558	.5830
			(b_1, b_2, b_3)=(.1, .1, 0.3)				
100	β_1	1.0055	.0001	.0135	1.0054	.9831	1.0276
	β_2	-.9786	.0001	.0116	-.9785	-.9978	-.9594
	β_3	2.0073	.0001	.0098	2.0073	1.9911	2.0235
	β_4	-2.0028	.0001	.0082	-2.0028	-2.0162	-1.9892
	δ_1	4.9849	.0003	.0332	4.9852	4.9290	5.0386
	δ_2	-5.0011	.0003	.0369	-5.0014	-5.0619	-4.9397
	δ_3	1.9582	.0003	.0392	1.9579	1.8931	2.0227
	δ_4	-1.9571	.0004	.0414	-1.9567	-2.0258	-1.8891
	$\rho_{jj'}$	.5130	.0003	.03701	.5113	.4511	.5746

5.2. Example 2: Voter Behavior Data

The data is a survey data of voting behavior collected from 95 residents of Troy, Michigan. This example is also considered in Green (1993). The objective of the study is to model two quantal responses as a function of covariates, allowing for correlation in responses. The two quantal responses were recorded: Y_{i1} = the first decision, measured by 1 or 0 with 1 being a state of sending at least one child to public school; Y_{i2} = the second, recorded on the binary (1-0) scale depending on whether to vote in favor of a school budget.

Let the covariates in x_{i11} be a constant, the natural logarithm of annual household income in dollars (INC), and the natural logarithm of property taxes paid per year in dollars (TAX); and those in x_{i21} be a constant, INC, TAX, and the number of years (YRS) the resident has been living in Troy. So that $x_{i1} = (x_{i11}, INC_i, TAX_i)'$ and $x_{i1} = (x_{i21}, INC_i, TAX_i, YRS_i)'$. (See Green 1993 for a detailed discussion of this data set). The summary statistics for the data set is given in Table 3.

Table 3. Summary of the Dataset

(Y_{i1}, Y_{i2})	Count	Decision		INC	TAX	YRS
(0, 0)	8	First	Mean	0.109	0.302	0.352
(0, 1)	7		S.D.	0.381	0.535	0.388
(1, 0)	28	Second	Mean	0.164	0.029	0.157
(1, 1)	52		S.D.	0.346	0.134	0.103

We want to fit the proposed model to this data set. The bivariate skewed probit model in which the marginal probabilities for the ith subject are given by

$$Pr(Y_{ij} = 1|\beta_j, \varrho, \delta_j) = \int_{-\infty}^{\infty} \Phi(x'_{ij}\beta_j + \delta_j w_i) g(w_i) dw_i,$$

and the joint probabilities are given through the cdf of the bivariate normal with correlation matrix equal to

$$\Sigma = \begin{pmatrix} 1 & \varrho \\ \varrho & 1 \end{pmatrix}.$$

Thus the model contains 9 unknown regression parameters (including δ_j's) and 1 unknown correlation parameter.

First Algorithm 2 is applied to the data set for testing the skewness of the bivariate probit model. For illustrative purpose we take $Pr(\delta_j = 0) = q_j =$

0.5, $j = 1, 2$, as a base prior probability that each δ_j is excluded from the model. To study the relation between the prior and the posterior distribution of δ_j's in conjunction with Algorithm 2, we also consider $q_j = 0.3$ and $q_j = 0.7$. In order to reflect the vagueness of the prior information about β, δ and ϱ, we represent our prior distribution through the hyperparameters $B_0 = .01I_7$, $D_0 = 0.01I_2$, $G_0 = 2$, $\beta_0 = 0$, $\delta_0 = 0$, and $\varrho_0 = 0$.

Posterior probabilities of alternative states of skewness parameters are presented in Table 4. These results are obtained from the method described in Section 4, with $m = 10^4$ iterations of Algorithm 2 beyond 10^3 burn-in iterations (decided based on the same convergence checkings as in Example 1). A systematic effect of the prior distributions of q_j's on the posterior probabilities is evident. Increases in q_j, the prior probability that $\delta_j = 0$, tends to favor symmetric probit model and vice versa. Thus giving more informative priors to δ_j's have the potential to effect our posterior inference about δ_j's. From Table 4, it can be also observed that, regardless of the particular prior distribution, the partial Bayes factor method yields the largest posterior probability for the state $(\delta_1 = 0, \delta_2 \neq 0)$. This non-zero value of the skewness parameter δ_2 suggested that the following bivariate skewed probit model may fit the data.

$$Z_i \sim N_2(X_i\beta + \delta w_i, \Sigma), \quad Y_{ij} = I(Z_{ij} > 0), \tag{29}$$

where $\delta = (0, \delta_2)'$. We proceed to estimate the model (29) by use of the MCMC sampling in Subsection 3.1. In the sampling process, we ignore the first m_0= 1000 draws and collect the next $m = 10^4$. These are used to approximate the posterior distributions of 9 parameters in the bivariate skewed probit model (Skewed). It is worth mentioning that the entire sampling process took less than 20 minutes on 600MHz PC.

Table 4. Posterior Probabilities of States of δ_1 and δ_2.

States	q_j	0.3	0.5	0.7
$(\delta_1 = 0, \delta_2 = 0)$		0.109	0.302	0.352
$(\delta_1 = 0, \delta_2 \neq 0)$		0.381	0.535	0.388
$(\delta_1 \neq 0, \delta_2 = 0)$		0.164	0.029	0.157
$(\delta_1 \neq 0, \delta_2 \neq 0)$		0.346	0.134	0.103

In this example, proposal values are also generated by the random walk chain, but with $\tau = 1$. Summaries of the posterior distributions are contrasted with those obtained from facilitating the bivariate probit model (Symm) and they are provided in Table 5. We observe that the posterior estimate of δ_2 is positive and that its 95% credible interval does not include 0. This coincides with the result in Table 4. The positive value of the skewness parameter suggests that usual probit model may not fit the responses Y_{2i}'s obtained from the second decision. We also note from Table 5 that when we fit Symm instead of Skewed all the posterior means of Symm are different from those of Skewed.

Table 5. Summaries of the Posterior Distributions

Para.	Model	Mean	Mum. SE	SD	Median	Lower	Upper
β_{11}	Symm	-4.187	0.007	3.564	-4.202	-11.356	2.846
	Skewed	-4.247	0.007	3.696	-4.217	-10.323	1.786
β_{12}	Symm	0.068	0.012	0.435	0.079	-0.782	0.907
	Skewed	0.101	0.013	0.445	0.107	-0.644	0.820
β_{13}	Symm	0.652	0.017	0.561	0.659	-0.469	1.743
	Skewed	0.616	0.016	0.563	0.614	-0.299	1.550
β_{21}	Symm	-0.469	0.080	3.789	-0.425	-7.889	6.846
	Skewed	-1.815	0.095	7.964	-1.682	-14.981	11.183
β_{22}	Symm	1.053	0.014	0.437	1.039	0.256	1.947
	Skewed	4.087	0.079	2.025	3.912	1.154	7.628
β_{23}	Symm	-1.382	0.016	0.579	-1.356	-2.671	-0.391
	Skewed	-6.147	0.094	2.905	-5.845	-11.324	-1.946
β_{24}	Symm	-0.018	0.002	0.016	-0.018	-0.043	0.012
	Skewed	-0.141	0.001	0.009	-0.134	-0.321	-0.001
δ_2	Skewed	9.708	0.085	2.503	9.832	5.421	13.541
ϱ	Symm	0.259	0.009	0.149	0.275	-0.130	0.639
	Skewed	0.023	0.002	0.150	0.025	-0.226	0.272

The difference is highlighted by mean and standard deviation of β_{24}. That is, the 95% credible interval for β_{24} include 0 if Symm is adopted for fitting the data, while Skewed model does not include 0 in its interval. Therefore, we say that the properties of the posterior estimates are not robust to the choice of models.

Finally, We compare the proposed model with the (symmetric) probit model in terms of predicted bivariate binary responses. The predicted responses, $\hat{Y}_i$, $i = 1, \ldots, n$, for each model are obtained in the following procedure. First, given the Bayes estimates in Table 5, we can calculate the posterior probabilities of four different responses of Y_i $((0,0),(0,1),(1,0),(1,1))$ using Algorithm 1. Then we take the highest probability response among them as predicted value of Y_i. An intuitively appealing way to summarize the predicted values from each fitted model (Skewed or Symm) is via a classification table. See, Efron (1975) and

Hosmer and Lemeshow (1989), for the use of the table as a criterion for goodness-of-fit. The results of classifying the predicted Y_i's using each fitted model given in Table 5 are noted in Table 6.

Table 6 notes following evidence. In comparison with the result of Skewed model, Symm model does not performs well in predicting the second binary response Y_{i2}. This coincides with an implication of Table 5 that fitted model involves significant skewness parameter in Y_{i2}.

Table 6. Classification Table Based on the Fitted Skewed Model (Result of the Fitted Symm Model is Listed in Parentheses).

		Predicted				
Observed	(Y_{i1}, Y_{i2})	(0, 0)	(0, 1)	(1, 0)	(1, 1)	Total
(Y_{i1}, Y_{i2})	Model					
(0, 0)	Skew	7	1	0	0	8
	Symm	(5)	(2)	(0)	(1)	(8)
(0, 1)	Skew	1	6	0	0	7
	Symm	(2)	(4)	(0)	(1)	(7)
(1, 0)	Skew	1	0	26	1	28
	Symm	(1)	(0)	(22)	(5)	(28)
(1, 1)	Skew	0	2	1	49	52
	Symm	(1)	(3)	(6)	(42)	(52)
Total	Skew	9	9	27	50	95
	Symm	(9)	(9)	(28)	(49)	(95)

6. Concluding Remarks

This paper has presented a skewed multivariate probit model for analyzing a correlated binary data with covariates. The model is described in terms of a skewed multivariate normal distribution for underlying variables that are manifested as discrete variables through a threshold specification. In addition, the paper has established Bayesian techniques for analyzing the skewed multivariate probit model from the output of posterior simulation via MCMC. Our two illustrative examples suggest that (i) the techniques can be applied to various binary response data sets and to high dimensional models that are intractable by using a frequentist method; (ii) the skewed multivariate probit model may be more appropriate than the multivariate probit model when the number of 1's is much different from the number of 0's in each component of the vectors in a correlated binary response data. The implication (ii) can be easily seen from Figure 1 and Table 6.

The advantages of the proposed model can be enumerated: (i) It allows flexibility to model skewness(in the sense that the skewness of the model is determined by the data). (ii) It is analytically tractable and easily implementable from a computational perspective. There are, however, a few aspects that warrant further study. In practice, there may be other plausible choice of the distributions of Z_i and w_i to get a more general skewed multivariate link model in the presence of binary correlated data with covariates. A study pertaining to suggesting a general skewed multivariate link model is an interesting research topic and it is left as a future study of interest.

REFERENCES

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

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

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

Azzalini, A. and Valle, A. D. (1996). "The multivariate skew-normal distribution," *Biometrika,* **83**, 715-726.

Basu, S. and Mukhopadhyay, S. (2000). Binary response regression with normal scale mixture links, in *Generalized Linear Models: A Bayesian Perspective,* eds. D. K. Dey, S. K. Ghosh, and B. K. Mallick, Marcel Dekker, New York.

Best, N., Cowles, M. K., and Vines, K. (1996). *CODA; Convergence Diagnosis and Output Analysis Software for Gibbs Sampling Output, Version 3.0,* MRC Biostatistics Unit, Cambridge.

Carey, V., Zeger, S. L., and Diggle, P. (1993). "Modelling multivariate binary data with alternating logistic regressions," *Biometrika,* **80**, 517-526.

Chen, M. H., Dey, D. K., and Shao, Q. M. (1999). "A new skewed link model for dichotomous quantal response model," *Journal of the American Statistical Association,* **94**, 1172-1186.

Chib, S. and Greenberg, E. (1996). Bayesian analysis of multivariate probit models in MCMC Preprint Service.

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

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.

Czado, C. (1994). "Parametric link modification of both tails in binary regression," *Statistical Papers,* **35**, 189-201.

Czado, C. and Santner, T. J. (1992). "The effect of link misspecification on binary regression inference," *Journal of Statistical Planning and Inference,* **33**, 213-231.

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

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

Geweke, J. (1996). "Variable Selection and Model Comparison in Regression," *Bayesian Statistics 5.* eds. J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith, Oxford University Press, 609-620.

Gelfand, A. E. and Smith, A. F. M. (1990). "Sampling-based approaches to calculating marginal densities," *Journal of the American Statistical Association,* **85**, 398-409.

Glonek, G. F. V. and McCullagh, P. (1995). "Multivariate logistic models," *Journal of the Royal Statistical Society,* B, **57**, 533-546.

Green, W. (1993). *Economic Analysis.* 2nd. ed. Macmillan, New York.

Ochi, Y. and Prentice, R. L. (1984). "Likelihood inference in a correlated probit regression model," *Biometrika,* **71**, 531-543.

Roberts, G. O., Gelman, A., and Gilks, W. R. (1994). Weak convergence and optimal scaling of random walk Metropolis algorithms, Technical Report, University of Cambridge.