BIOMETRIC METHODOLOGY



Tractable Bayes of skew-elliptical link models for correlated binary data

Zhongwei Zhang¹ Raphaël Huser¹

Zhongwei Zhang¹ Reinaldo B. Arellano-Valle² Marc G. Genton¹

¹Statistics Program, Computer, Electrical and Mathematical Sciences and Engineering Division, King Abdullah University of Science and Technology (KAUST), Thuwal, Saudi Arabia

²Departamento of Estadística, Pontificia Universidad Católica de Chile, Santiago, Chile

Correspondence

Raphaël Huser, Statistics Program, Computer, Electrical and Mathematical Sciences and Engineering Division, King Abdullah University of Science and Technology (KAUST), Thuwal 23955-6900, Saudi Arabia.

Email: raphael.huser@kaust.edu.sa

Abstract

Correlated binary response data with covariates are ubiquitous in longitudinal or spatial studies. Among the existing statistical models, the most well-known one for this type of data is the multivariate probit model, which uses a Gaussian link to model dependence at the latent level. However, a symmetric link may not be appropriate if the data are highly imbalanced. Here, we propose a multivariate skew-elliptical link model for correlated binary responses, which includes the multivariate probit model as a special case. Furthermore, we perform Bayesian inference for this new model and prove that the regression coefficients have a closed-form unified skew-elliptical posterior with an elliptical prior. The new methodology is illustrated by an application to COVID-19 data from three different counties of the state of California, USA. By jointly modeling extreme spikes in weekly new cases, our results show that the spatial dependence cannot be neglected. Furthermore, the results also show that the skewed latent structure of our proposed model improves the flexibility of the multivariate probit model and provides a better fit to our highly imbalanced dataset.

KEYWORDS

asymmetric link model, correlated binary data, COVID-19 pandemic, Markov chain Monte Carlo, tractable Bayes, unified skew-elliptical distribution

1 | INTRODUCTION

Correlated binary response data with covariates frequently arise in longitudinal or spatial studies. For instance, in longitudinal studies, the disease status (i.e., diseased or not diseased) is measured over time on the same person. Similarly, in a panel study of income dynamics, the employment status information may be collected over time from the same survey participant. The multivariate probit model (Ashford & Sowden, 1970; Chib & Greenberg, 1998) is well known for this type of data, as it describes the dependence between binary variables by a latent Gaussian link, which allows for flexible modeling of dependence, has a straightforward interpretation of the parameters, and is easily amenable to Bayesian inference.

A symmetric link, however, does not always provide the best fit to a given dataset; see Chen et al. (1999) and Kim et al. (2008) for some examples. In this case, the link might be misspecified, yielding substantial bias in the mean response estimates (Czado & Santner, 1992). Chen et al. (1999) used the rate at which the probability of a given binary response variable approaches 0 and 1 to guide the selection of a symmetric or asymmetric link. In other words, if the binary response data are highly imbalanced, the rate of the probability of the random variable approaching 0 is typically very different from the one approaching 1, so that an asymmetric link might be preferred over a symmetric link. Motivated by this observation, a variety of flexible asymmetric link models have been proposed for univariate binary response data, such as the skew-t link model (Kim, 2002), skew-probit link model (Bazán et al., 2010), and generalized extreme value link model (Wang & Dey, 2010). The last one was recently extended to model correlated imbalanced binary data by using a multivariate normal distribution to capture the dependence (Zhao et al., 2021). However, to the best of our knowledge, no other multivariate asymmetric link models have previously been proposed in the literature for correlated binary responses. The purpose of this paper is to propose a flexible multivariate skew-elliptical (Branco & Dey, 2001) link model for correlated binary responses, which includes the multivariate probit model as a special case and allows for tractable Bayesian inference; see Section 2.2 for details on the multivariate skew-elliptical distribution, used in our model as a key building block.

Durante (2019) has proved that for the univariate probit model with Gaussian priors the posterior of the regression coefficients belongs to the class of unified skewnormal distributions (Arellano-Valle & Azzalini, 2006). In this paper, we also consider Bayesian inference for our new multivariate model and prove that the posterior of the regression coefficients belongs to the unified skew-elliptical family (Arellano-Valle & Genton, 2010) for an elliptical prior. The closed-form and tractable posterior for the regression coefficients facilitates inference by using an algorithm that does not rely on data augmentation, and thus avoids the convergence and mixing issues of the classical data-augmentation algorithms for probit models; see Johndrow et al. (2019) for a discussion of this issue.

We illustrate the new methodology by an application to COVID-19 pandemic data from three different counties of the state of California, USA. By jointly modeling the occurrences of extreme spikes in weekly new infected cases using our new model, we can estimate the underlying spatial dependence structure, which might provide helpful quantitative insights into the transmission modes of the virus and help authorities mitigate its spread. Furthermore, our model has additional skewness parameters compared to the multivariate probit model, which improves its flexibility and makes it more appropriate for modeling our highly imbalanced dataset.

This paper is organized as follows. Section 2 describes the preliminaries of the skew-elliptical and unified skewelliptical distributions. Section 3 details our proposed methodology. We first introduce the new skew-elliptical link model and prove that the regression coefficients of this model have a unified skew-elliptical posterior, and then we focus on two important special cases, that is, the skew-normal and skew-*t* link models. Section 4 concerns a simulation study and an application to COVID-19 pandemic data. Section 5 concludes with a discussion and perspectives on future research.

2 | PRELIMINARIES: SKEW-ELLIPTICAL AND UNIFIED SKEW-ELLIPTICAL DISTRIBUTIONS

2.1 | The skew-elliptical distribution

The skew-elliptical distribution, originally proposed by Azzalini and Capitanio (1999), was formulated by multiplying an elliptical density with a skewing function. Branco and Dey (2001) proposed a new formulation of the skew-elliptical distribution by means of a conditioning mechanism. The close relationship between these two formulations is established in Azzalini and Capitanio (2003). Thanks to the construction in terms of a conditioning mechanism, the formulation in Branco and Dev (2001) has led to many attractive properties of this class of distribution, such as the existence of stochastic representation and closeness under marginalization and affine transformation. Fang (2003) considered a slightly wider class of distributions than Branco and Dey (2001) by adding an extra truncation parameter, which was later called the extended skew-elliptical distribution in Arellano-Valle and Genton (2010), and showed that this new distribution is closed under marginalization, affine transformation, and also conditioning.

Here we adopt a slightly different parameterization than Fang (2003) with the truncation parameter taken as 0 and consider only skew-elliptical random vectors which possess densities. Let $g^{(d+1)}$ be a density generator for a (d + 1)-dimensional elliptical random vector that satisfies $\int_0^\infty r^{(d+1)/2-1}g^{(d+1)}(r)dr = \Gamma\{(d + 1)/2\}\pi^{-(d+1)/2}$. Then a *d*-dimensional random vector X has a skew-elliptical distribution with location parameter vector $\xi \in \mathbb{R}^d$, positivedefinite scale matrix $\Sigma \in \mathbb{R}^{d \times d}$, skewness parameter vector $\alpha \in \mathbb{R}^d$, and density generator $g^{(d+1)}$, if its density function is

$$f_{\boldsymbol{X}}(\boldsymbol{x}) = 2|\boldsymbol{\Sigma}|^{-1/2} g^{(d)} \{ (\boldsymbol{x} - \boldsymbol{\xi})^{\top} \boldsymbol{\Sigma}^{-1} (\boldsymbol{x} - \boldsymbol{\xi}) \}$$
$$\times G\{ \boldsymbol{\alpha}^{\top} \sigma^{-1} (\boldsymbol{x} - \boldsymbol{\xi}); g_{a(\boldsymbol{x})} \}, \quad \boldsymbol{x} \in \mathbb{R}^{d}, \quad (1)$$

where $\sigma = \operatorname{diag}(\Sigma)^{1/2} \in \mathbb{R}^{d \times d}$, $q(\mathbf{x}) = (\mathbf{x} - \boldsymbol{\xi})^{\top} \Sigma^{-1} (\mathbf{x} - \boldsymbol{\xi})$, $g^{(d)}$ is the *d*-dimensional marginal density generator induced by $g^{(d+1)}$, and $G(\cdot; g_{q(\mathbf{x})})$ is the cumulative distribution function of the univariate elliptical distribution with mean 0, scale 1, and conditional density generator $g_{q(\mathbf{x})}(s) = g^{(d+1)} \{s + q(\mathbf{x})\} / g^{(d)} \{q(\mathbf{x})\}$. We write 1790 WILEY Biometrics

 $X \sim S\mathcal{E}_d(\boldsymbol{\xi}, \boldsymbol{\Sigma}, \boldsymbol{\alpha}, g^{(d+1)})$. When $\boldsymbol{\alpha} = \mathbf{0}$, the skew-elliptical distribution reduces to elliptical distribution.

The skew-elliptical distribution has two stochastic representations, that is, a convolution-type representation and a conditioning-type representation; see Equations (10) and (19) in Fang (2003). The former is useful for random sampling, and the latter allows us to express its cumulative distribution function in the following simple form:

$$F_{X}(\boldsymbol{x}) = 2G_{d+1}(\boldsymbol{x}_{*} - \boldsymbol{\xi}_{*}; \boldsymbol{\Sigma}_{*}, \boldsymbol{g}^{(d+1)}), \qquad (2)$$

with $\boldsymbol{x}_* = (0, \boldsymbol{x}^\top)^\top, \boldsymbol{\xi}_* = (0, \boldsymbol{\xi}^\top)^\top$ and

$$\Sigma_* = \begin{pmatrix} 1 & -\boldsymbol{\delta}^\top \boldsymbol{\sigma} \\ -\boldsymbol{\sigma} \boldsymbol{\delta} & \boldsymbol{\Sigma} \end{pmatrix},$$

where $\sigma = \operatorname{diag}(\Sigma)^{1/2} \in \mathbb{R}^{d \times d}$, $\delta = (1 + \alpha^{\top} \overline{\Sigma} \alpha)^{-1/2} \overline{\Sigma} \alpha$ with $\overline{\Sigma}$ being the correlation matrix corresponding to Σ , that is, $\Sigma = \sigma \overline{\Sigma} \sigma$, and $G_{d+1}(\boldsymbol{x}_* - \boldsymbol{\xi}_*; \Sigma_*, g^{(d+1)})$ denotes the cumulative distribution function of the (d + 1)-variate elliptical distribution with location vector $\boldsymbol{\xi}_* \in \mathbb{R}^{d+1}$, positive-definite covariance matrix $\Sigma_* \in \mathbb{R}^{(d+1) \times (d+1)}$, and density generator $g^{(d+1)}$. The positive definiteness of Σ_* implies that the admissible parameters of (Σ, α) are such that the matrix $\overline{\Sigma} - \delta \delta^{\top}$ is positive definite.

A prominent subclass of the skew-elliptical distribution is the skew-normal distribution (Azzalini & Dalla Valle, 1996). Specifically, when $g^{(d+1)}$ is the (d + 1)-variate normal density generator, the density function of X is

$$f_{\boldsymbol{X}}(\boldsymbol{x}) = 2\phi_d(\boldsymbol{x} - \boldsymbol{\xi}; \boldsymbol{\Sigma})\Phi\{\boldsymbol{\alpha}^{\top}\sigma^{-1}(\boldsymbol{x} - \boldsymbol{\xi})\}, \quad \boldsymbol{x} \in \mathbb{R}^d,$$

where $\phi_d(\mathbf{x} - \boldsymbol{\xi}; \Sigma)$ denotes the probability density function of the *d*-variate Gaussian distribution with mean vector $\boldsymbol{\xi}$ and covariance matrix Σ , and $\Phi(\cdot)$ is the cumulative distribution function of the standard normal distribution. We denote this distribution as $\mathbf{X} \sim S \mathcal{N}_d(\boldsymbol{\xi}, \Sigma, \boldsymbol{\alpha})$. When $\boldsymbol{\alpha} = \mathbf{0}$, it reduces to the *d*-dimensional normal distribution, $\mathcal{N}_d(\boldsymbol{\xi}, \Sigma)$.

When $g^{(d+1)}$ is the (d + 1)-variate *t* density generator with ν degrees of freedom, we get another important subclass of the skew-elliptical distribution, that is, the skew-*t* distribution (Branco & Dey, 2001; Azzalini & Capitanio, 2003). Its density has the following form:

$$f_X(\boldsymbol{x}) = 2t_d(\boldsymbol{x} - \boldsymbol{\xi}; \boldsymbol{\Sigma}, \boldsymbol{\nu})T\left[\boldsymbol{\alpha}^{\top} \boldsymbol{\sigma}^{-1}(\boldsymbol{x} - \boldsymbol{\xi}) \left\{ \frac{\boldsymbol{\nu} + \boldsymbol{p}}{q(\boldsymbol{x}) + \boldsymbol{\nu}} \right\}^{1/2}; \boldsymbol{\nu} + d \right],$$
$$\boldsymbol{x} \in \mathbb{R}^d,$$

where $t_d(\boldsymbol{x} - \boldsymbol{\xi}; \boldsymbol{\Sigma}, \boldsymbol{\nu})$ denotes the probability density function of the *d*-variate *t* distribution with location vector $\boldsymbol{\xi}$, scale matrix $\boldsymbol{\Sigma}$, and degrees of freedom $\boldsymbol{\nu}$, $T(\cdot; \boldsymbol{\nu} + d)$

denotes the univariate *t* distribution function with degrees of freedom v + d. We write $\mathbf{X} \sim S\mathcal{T}_d(\boldsymbol{\xi}, \Sigma, \boldsymbol{\alpha}, v)$. When $\boldsymbol{\alpha} = \mathbf{0}$, it reduces to the *d*-dimensional *t* distribution, and when $v \to \infty$, it tends to the *d*-dimensional skew-normal distribution.

2.2 | The unified skew-elliptical distribution

An extension of the skew-elliptical distribution is the unified skew-elliptical distribution (Arellano-Valle & Genton, 2010), which aims to gain more flexibility by unifying various skew-elliptical families under the same model. Specifically, a *d*-dimensional random vector X has a unified skew-elliptical distribution, denoted by $X \sim SU\mathcal{E}_{d,m}(\boldsymbol{\xi}, \Sigma, \Lambda, \tau, \Gamma, g^{(d+m)})$, if its density function is

$$\begin{split} f_X(\boldsymbol{x}) &= |\boldsymbol{\Sigma}|^{-1/2} g^{(d)} \{ (\boldsymbol{x} - \boldsymbol{\xi})^\top \boldsymbol{\Sigma}^{-1} (\boldsymbol{x} - \boldsymbol{\xi}) \} \\ &\times \frac{G_m \{ \boldsymbol{\tau} + \boldsymbol{\Lambda} \boldsymbol{\sigma}^{-1} (\boldsymbol{x} - \boldsymbol{\xi}); \boldsymbol{\Gamma}, \boldsymbol{g}_{q(\boldsymbol{x})}^{(m)} \}}{G_m(\boldsymbol{\tau}; \boldsymbol{\Gamma} + \boldsymbol{\Lambda} \bar{\boldsymbol{\Sigma}} \boldsymbol{\Lambda}^\top, \boldsymbol{g}^{(m)})}, \quad \boldsymbol{x} \in \mathbb{R}^d, \end{split}$$

where $q(\mathbf{x}) = (\mathbf{x} - \boldsymbol{\xi})^{\top} \Sigma^{-1} (\mathbf{x} - \boldsymbol{\xi})$, $g^{(d+m)}$ is a (d + m)variate elliptical density generator, $g^{(d)}$ and $g^{(m)}$ are its *d*variate and *m*-variate marginal density generators, respectively, $g_{q(\mathbf{x})}^{(m)}(s) = g^{(d+m)} \{s + q(\mathbf{x})\} / g^{(d)} \{q(\mathbf{x})\}$, $\boldsymbol{\xi} \in \mathbb{R}^d$ is a location parameter vector, $\tau \in \mathbb{R}^d$ introduces additional flexibility to capture skewness, $\Gamma \in \mathbb{R}^{m \times m}$ is a correlation matrix, and $\Lambda \in \mathbb{R}^{m \times d}$ encompasses the main effect on the skewness. When m = 1, it reduces to the extended skewelliptical distribution (Fang, 2003), and if we further have $\tau = \mathbf{0}$, it reduces to the skew-elliptical distribution (1).

Similar to the skew-elliptical distribution, the unified skew-elliptical distribution also has two special subclasses, that is, the unified skew-normal distribution (Arellano-Valle & Azzalini, 2006) and the unified skew-*t* distribution. When $g^{(d+m)}$ is the (d + m)-variate normal density generator, we get the unified skew-normal distribution with density

$$f_{X}(\boldsymbol{x}) = \phi_{d}(\boldsymbol{x} - \boldsymbol{\xi}; \boldsymbol{\Sigma}) \frac{\Phi_{m}\{\boldsymbol{\tau} + \Lambda \sigma^{-1}(\boldsymbol{x} - \boldsymbol{\xi}); \boldsymbol{\Gamma}\}}{\Phi_{m}(\boldsymbol{\tau}; \boldsymbol{\Gamma} + \Lambda \bar{\boldsymbol{\Sigma}} \Lambda^{\top})}, \quad \boldsymbol{x} \in \mathbb{R}^{d},$$
(3)

where $\Phi_m(\cdot; \Gamma)$ denotes the centered *m*-dimensional normal distribution function with covariance matrix Γ . We write $\mathbf{X} \sim SUN_{d,m}(\boldsymbol{\xi}, \Sigma, \Lambda, \tau, \Gamma)$. The definition (3) is equivalent to the one in Arellano-Valle and Azzalini (2006) with a slightly different parameterization but is consistent with Arellano-Valle and Genton (2010). When $g^{(d+m)}$ is the (d + m)-variate *t* density generator with ν degrees of freedom, we get the unified skew-*t* distribution with density

$$f_{\boldsymbol{X}}(\boldsymbol{x}) = t_d(\boldsymbol{x} - \boldsymbol{\xi}; \boldsymbol{\Sigma}, \boldsymbol{\nu}) \frac{T_m \left[\{ \boldsymbol{\tau} + \boldsymbol{\Lambda} \boldsymbol{\sigma}^{-1} (\boldsymbol{x} - \boldsymbol{\xi}) \} \left\{ \frac{\boldsymbol{\nu} + p}{q(\boldsymbol{x}) + \boldsymbol{\nu}} \right\}^{1/2}; \boldsymbol{\Gamma}, \boldsymbol{\nu} + d \right]}{T_m(\boldsymbol{\tau}; \boldsymbol{\Gamma} + \boldsymbol{\Lambda} \bar{\boldsymbol{\Sigma}} \boldsymbol{\Lambda}^{\top}, \boldsymbol{\nu})} \\ \boldsymbol{x} \in \mathbb{R}^d,$$

where $T_m(\cdot; \Gamma, \nu + d)$ denotes the centered *m*dimensional *t* distribution function with dispersion matrix Γ and degrees of freedom $\nu + d$. We write $X \sim SUT_{d,m}(\xi, \Sigma, \Lambda, \nu, \tau, \Gamma)$.

3 | POSTERIOR INFERENCE FOR THE SKEW-ELLIPTICAL LINK MODEL

3.1 | The skew-elliptical link model

As discussed in Section 1, when modeling correlated binary data, the multivariate probit model uses a Gaussian link to capture dependence at the "latent level." A symmetric link, however, does not always provide the best fit to a given dataset, in particular for binary response data that are highly imbalanced.

In this section, we extend the Gaussian link to the multivariate skew-elliptical link, which includes the skewnormal and skew-*t* links as special cases. Specifically, let Y_{ij} denote a binary {0, 1} response on the *i*th observation of the *j*th variable and denote by $\mathbf{Y}_i = (Y_{i1}, \dots, Y_{iM})^{\mathsf{T}}$ the collection of the *i*th observation on all *M* variables for $i = 1, \dots, n$. Let $\mathbf{Y}_i^* = (Y_{i1}^*, \dots, Y_{iM}^*)^{\mathsf{T}}$ be a vector of latent variables capturing dependence among the components of $\mathbf{Y}_i, \boldsymbol{\beta} \in \mathbb{R}^p$ be a vector of regression coefficients, $X_i = (\mathbf{x}_{i1}, \dots, \mathbf{x}_{iM})^{\mathsf{T}} \in \mathbb{R}^{M \times p}$ be the data matrix for the *i*th observation, and let $X = (X_1^{\mathsf{T}}, \dots, X_n^{\mathsf{T}})^{\mathsf{T}} \in \mathbb{R}^{nM \times p}$. Then the multivariate skew-elliptical link model can be expressed as

$$Y_{ij} = \begin{cases} 1, & \text{if } Y_{ij}^* > 0, \\ 0, & \text{otherwise,} \end{cases}$$
$$Y^* = (Y_1^{*\top}, \dots, Y_n^{*\top})^{\top} = X\beta + \varepsilon, \qquad (4)$$

$$\begin{pmatrix} \boldsymbol{\beta} \\ \boldsymbol{\varepsilon} \end{pmatrix} \mid \boldsymbol{\Sigma}, \boldsymbol{\alpha}, \boldsymbol{g}^{(p+nM+1)} \sim S\mathcal{E}_{p+nM} \begin{pmatrix} \boldsymbol{\mu} \\ \boldsymbol{0} \end{pmatrix}, \begin{pmatrix} \boldsymbol{\Omega} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I}_n \otimes \boldsymbol{\Sigma} \end{pmatrix}, \begin{pmatrix} \boldsymbol{0} \\ \boldsymbol{\alpha} \end{pmatrix}, \boldsymbol{g}^{(p+nM+1)} \end{pmatrix},$$

where $\boldsymbol{\mu} \in \mathbb{R}^p$ is a location parameter vector, $\Omega \in \mathbb{R}^{p \times p}$ is a positive-definite covariance matrix, $\mathbf{I}_n \in \mathbb{R}^{n \times n}$ is the identity matrix, \otimes denotes the Kronecker product, $\Sigma \in \mathbb{R}^{M \times M}$ is a positive-definite covariance matrix, $\boldsymbol{\alpha} \in \mathbb{R}^{nM}$ is a skewness parameter vector, and $g^{(p+nM+1)}$ is a (p + nM + 1)-variate elliptical density generator.

To better understand the assumption on the joint distribution of β and ε in the model (4), we express it in a different way. Using Proposition 2 in Fang (2003), an equivalent formulation is

$$\boldsymbol{\beta} \mid \boldsymbol{g}^{(p+nM+1)} \sim \mathcal{SE}_p(\boldsymbol{\mu}, \boldsymbol{\Omega}, \boldsymbol{0}, \boldsymbol{g}^{(p+1)}),$$
 (5)

$$\boldsymbol{\varepsilon} \mid \boldsymbol{\beta}, \boldsymbol{\Sigma}, \boldsymbol{\alpha}, \boldsymbol{g}^{(p+nM+1)} \sim \mathcal{SE}_{nM}(\mathbf{0}, \mathbf{I}_n \otimes \boldsymbol{\Sigma}, \boldsymbol{\alpha}, \boldsymbol{g}_{q(\boldsymbol{\beta})}^{(nM+1)}),$$
(6)

where $q(\boldsymbol{\beta}) = (\boldsymbol{\beta} - \boldsymbol{\mu})^{\mathsf{T}} \Omega^{-1} (\boldsymbol{\beta} - \boldsymbol{\mu}), g_{q(\boldsymbol{\beta})}^{(nM+1)}(s) = g^{(p+nM+1)}$ $\{s + q(\beta)\}/g^{(p)}\{q(\beta)\}$, and $g^{(p)}, g^{(p+1)}$ are the p- and $(p + q^{(p)})$ 1)-variate marginal density generators induced by the same generator $g^{(p+nM+1)}$, respectively. Assumption (5) may be understood as the elliptical prior, as the skewness parameter is zero, for β , while (6) is the distributional assumption for the error in the latent data vector Y^* . From (6), we observe that β and ε are dependent, but they are conditionally independent given $q(\beta)$. This weak dependence between them is broken when $g^{(p+nM+1)}$ is the normal density generator. Specifically, when $g^{(p+nM+1)}$ is the (p + nM + 1)-variate normal density generator, (5) becomes the typical Gaussian prior, $\mathcal{N}_{p}(\boldsymbol{\mu}, \Omega)$, and (6) becomes $\boldsymbol{\varepsilon} \mid \boldsymbol{\Sigma}, \boldsymbol{\alpha} \sim S \mathcal{N}_{nM}(\mathbf{0}, \mathbf{I}_n \otimes \boldsymbol{\Sigma}, \boldsymbol{\alpha})$, which is independent of β conditional on Σ and α . If we further have $\boldsymbol{\alpha} = \mathbf{0}$, then (6) becomes $\boldsymbol{\varepsilon} \mid \boldsymbol{\Sigma} \sim \mathcal{N}_{nM}(\mathbf{0}, \mathbf{I}_n \otimes \boldsymbol{\Sigma})$ and model (4) reduces to the well-known multivariate probit model (Ashford & Sowden, 1970; Chib & Greenberg, 1998) with a typical Gaussian prior for β . By assuming a joint distribution for β and ε , we can gain two major advantages. The first is that we are able to account not only for the dependence between β and ε but also for the dependence between the different observations Y_i , i =1, ..., n. The second is that this assumption allows us to get a tractable posterior for β ; see Section 3.2 for more details.

The multivariate probit model (Chib & Greenberg, 1998) assumes that the covariates are not shared by the *M* variables $Y_{i1}, ..., Y_{iM}$. In that case, $\boldsymbol{\beta}$ can be understood as $\boldsymbol{\beta} = (\boldsymbol{\beta}_1^\top, ..., \boldsymbol{\beta}_M^\top)^\top$, where $\boldsymbol{\beta}_j \in \mathbb{R}^{p_j}$ with $\sum_{j=1}^M p_j = p$ is the regression coefficients for the *j*th variables $Y_{1j}, ..., Y_{nj}$, and \boldsymbol{x}_{ij} is understood as the vector $\boldsymbol{x}_{ij} = (\boldsymbol{x}_{ij1}^\top, ..., \boldsymbol{x}_{ijM}^\top)^\top$ with $\boldsymbol{x}_{ijk} = \mathbf{0}$ for $k \neq j$, so that $\boldsymbol{x}_{ij}^\top \boldsymbol{\beta} = \boldsymbol{x}_{ijj}^\top \boldsymbol{\beta}_j$. This notation of expanded vector $\boldsymbol{\beta}$ and \boldsymbol{x}_{ij} simplifies the expression of our model (4).

From (6), we know that the admissible parameters of (Σ, α) are those such that the matrix $I_n \otimes \overline{\Sigma} - \delta \delta^{\top}$ is positive definite, where $\overline{\Sigma}$ is the correlation matrix corresponding to Σ and $\delta = \{1 + \alpha^{\top}(I_n \otimes \overline{\Sigma})\alpha\}^{-1/2}(I_n \otimes \overline{\Sigma})\alpha$. From (4), we obtain the joint probability mass function of $Y = (Y_1^{\top}, ..., Y_n^{\top})^{\top} = y$, given all the parameters and the data matrix X, as

$$p(\mathbf{y} \mid \boldsymbol{\beta}, \boldsymbol{\Sigma}, \boldsymbol{\alpha}, g^{(p+nM+1)}) = \int_{A_{nM}} \cdots \int_{A_{11}} 2|\mathbf{I}_n \otimes \boldsymbol{\Sigma}|^{-1/2}$$
$$\times g^{q(\boldsymbol{\beta}), nM} \{q(\mathbf{t})\} G(\boldsymbol{\alpha}^{\top} \mathbf{t}; g^{q(\boldsymbol{\beta})}_{a(t)}) d$$

where $q(t) = t^{\top}(I_n \otimes \Sigma^{-1})t$, $g_{q(t)}^{q(\beta)}(s) = g_{q(\beta)}^{(nM+1)}\{s + q(t)\}/g^{q(\beta),nM}\{q(t)\}$, $g^{q(\beta),nM}$ is the *nM*-variate marginal density generator induced by $g_{q(\beta)}^{(nM+1)}$, and $A_{ij}, i = 1, ..., n, j = 1, ..., n$ is the interval $(-\mathbf{x}_{ij}^{\top}\boldsymbol{\beta}, \infty)$ if $y_{ij} = 1$, and $(-\infty, \mathbf{x}_{ij}^{\top}\boldsymbol{\beta}]$ if $y_{ij} = 0$. Although the above joint probability involves multidimensional integration over a constrained space, we show in the following section that it can be substantially simplified.

3.2 Unified skew-elliptical posterior for the regression coefficients

In this section, we prove that for the multivariate skewelliptical link model (4), the regression coefficients parameter $\boldsymbol{\beta}$ has a unified skew-elliptical posterior for an elliptical prior. To prove this result, we first simplify the joint probability mass function $p(\boldsymbol{y} | \boldsymbol{\beta}, \boldsymbol{\Sigma}, \boldsymbol{\alpha}, g^{(p+nM+1)})$ of the observed data in the following lemma. All proofs are deferred to the Appendix.

Lemma 1. The joint probability mass function $p(\mathbf{y} | \boldsymbol{\beta}, \boldsymbol{\Sigma}, \boldsymbol{\alpha}, \mathbf{g}^{(p+nM+1)})$ based on (4) can be simplified to

$$p(\boldsymbol{y} \mid \boldsymbol{\beta}, \boldsymbol{\Sigma}, \boldsymbol{\alpha}, \boldsymbol{g}^{(p+nM+1)}) = 2G_{nM+1}(D_*\boldsymbol{\beta}; \boldsymbol{\Sigma}_*, \boldsymbol{g}_{q(\boldsymbol{\beta})}^{(nM+1)}),$$

where $D = \text{diag}(2\mathbf{y} - \mathbf{1}_{nM}) \in \mathbb{R}^{nM \times nM}$ with $\mathbf{1}_{nM} \in \mathbb{R}^{nM}$ being the vector of 1s, $D_* = (\mathbf{0}_p, (DX)^{\mathsf{T}})^{\mathsf{T}} \in \mathbb{R}^{(nM+1) \times p}$, $\mathbf{0}_p \in \mathbb{R}^p$ is a vector of 0s, and

$$\Sigma_* = \begin{pmatrix} 1 & -\boldsymbol{\delta}^\top D(\mathbf{I}_n \otimes \sigma) \\ -(\mathbf{I}_n \otimes \sigma) D\boldsymbol{\delta} & D(\mathbf{I}_n \otimes \Sigma) D \end{pmatrix} \in \mathbb{R}^{(nM+1) \times (nM+1)},$$

with $\delta \in \mathbb{R}^{nM}$, $\delta = \{1 + \alpha^{\top}(I_n \otimes \overline{\Sigma})\alpha\}^{-1/2}(I_n \otimes \overline{\Sigma})\alpha$, $\sigma = \text{diag}(\Sigma)^{1/2} \in \mathbb{R}^{d \times d}$ and $\overline{\Sigma}$ being the correlation matrix corresponding to Σ , that is, $\Sigma = \sigma \overline{\Sigma} \sigma$.

Now we are ready to present our main result that the posterior distribution of β coincides with a unified skew-elliptical distribution.

Theorem 1. Let $\mathbf{y} = (\mathbf{y}_1^{\top}, ..., \mathbf{y}_n^{\top})^{\top}$ be observations from the multivariate skew-elliptical link model (4) and $X = (X_1^{\top}, ..., X_n^{\top})^{\top}$ be the corresponding data matrix. Then

$$\boldsymbol{\beta} \mid \boldsymbol{y}, \boldsymbol{\Sigma}, \boldsymbol{\alpha}, \boldsymbol{g}^{(p+nM+1)} \sim$$

$$SU\mathcal{E}_{p,nM+1}(\boldsymbol{\mu}_{post}, \boldsymbol{\Omega}_{post}, \boldsymbol{\Lambda}_{post}, \boldsymbol{\tau}_{post}, \boldsymbol{\Gamma}_{post}, \boldsymbol{g}^{(p+nM+1)})$$

with posterior parameters

$$\mu_{post} = \mu, \ \Omega_{post} = \Omega, \ \Lambda_{post} = \sigma_*^{-1} D_* \omega,$$
$$\tau_{post} = \sigma_*^{-1} D_* \mu, \ \Gamma_{post} = \bar{\Sigma}_*,$$

where $D_* \in \mathbb{R}^{(nM+1) \times p}$ and $\Sigma_* \in \mathbb{R}^{(nM+1) \times (nM+1)}$ are the matrices defined in Lemma 1, $\sigma_* = \operatorname{diag}(\Sigma_*)^{1/2} \in \mathbb{R}^{(nM+1) \times (nM+1)}$, $\overline{\Sigma}_*$ is the correlation matrix corresponding to Σ_* , that is, $\Sigma_* = \sigma_* \overline{\Sigma}_* \sigma_*$, and $\omega = \operatorname{diag}(\Omega)^{1/2} \in \mathbb{R}^{p \times p}$.

In Bayesian regression, we are mostly interested in the posterior marginals, their moments, and more complex functionals such as measures of dependence and credible intervals. Thanks to the fundamental property of the unified skew-elliptical distribution that it is closed under marginalization, conditioning, and affine transformations, this type of inference is simplified. We refer to Arellano-Valle and Genton (2010) for details on how to obtain the parameters of the marginal distribution, conditional distribution, and the distribution after affine transformations. As for the calculation of the posterior moments and credible intervals, numerical integration of the marginal posterior densities can be used. When interest is in the posterior moments, another approach is to use the moment-generating function. We refer to Section 5 of Arellano-Valle and Genton (2010) for derivations of the moment generating function and moments of the unified skew-elliptical distribution.

3.3 | Special case: The skew-*t* link model

In this section, we consider a special case, that is the skew-*t* link model obtained when the density generator $g^{(p+nM+1)}$ in model (4) is the (p + nM + 1)-variate *t* density generator with ν degrees of freedom. The skew-normal link model is a limiting model of the skew-*t* link model as ν tends to infinity, and more details about this model are provided in the Supporting Information. Specifically, for the skew-*t* link model, the joint distributional assumption of β and ε is

$$\binom{\boldsymbol{\beta}}{\boldsymbol{\varepsilon}} \left| \boldsymbol{\Sigma}, \boldsymbol{\alpha}, \boldsymbol{\nu} \sim \mathcal{ST}_{p+nM} \left(\begin{pmatrix} \boldsymbol{\mu} \\ \boldsymbol{0} \end{pmatrix}, \begin{pmatrix} \boldsymbol{\Omega} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I}_n \otimes \boldsymbol{\Sigma} \end{pmatrix}, \begin{pmatrix} \boldsymbol{0} \\ \boldsymbol{\alpha} \end{pmatrix}, \boldsymbol{\nu} \right),$$

which is equivalent to assuming

$$\begin{split} \boldsymbol{\beta} \mid \boldsymbol{\nu} \sim \mathcal{T}_{p}(\boldsymbol{\mu}, \Omega, \boldsymbol{\nu}), \\ \boldsymbol{\varepsilon} \mid \boldsymbol{\beta}, \boldsymbol{\Sigma}, \boldsymbol{\alpha}, \boldsymbol{\nu} \sim \mathcal{ST}_{nM} \bigg(\mathbf{0}, \frac{\boldsymbol{\nu} + (\boldsymbol{\beta} - \boldsymbol{\mu})^{\mathsf{T}} \Omega^{-1} (\boldsymbol{\beta} - \boldsymbol{\mu})}{\boldsymbol{\nu} + p} (\mathbf{I}_{n} \otimes \boldsymbol{\Sigma}), \boldsymbol{\alpha}, \boldsymbol{\nu} + p \bigg) \end{split}$$

where $\mathcal{T}_p(\boldsymbol{\mu}, \Omega, \nu)$ denotes the *t* distribution with location parameter vector $\boldsymbol{\mu}$, dispersion matrix Ω , and degrees of freedom ν . The nonnegative parameter ν can be

considered as a hyper-parameter which controls the dependence between β and ε . As ν increases the dependence decreases, and when $\nu \rightarrow \infty$ the skew-*t* link model tends to the skew-normal link model and the dependence between them vanishes.

By taking $g^{(nM+1)}$ in Lemma 1 as the *t* density generator with ν degrees of freedom, we get the following explicit expression of the joint probability:

$$p(\boldsymbol{y} \mid \boldsymbol{\beta}, \boldsymbol{\Sigma}, \boldsymbol{\alpha}, \boldsymbol{\nu}) = 2T_{nM+1} \left[\left\{ \frac{\boldsymbol{\nu} + p}{\boldsymbol{\nu} + (\boldsymbol{\beta} - \boldsymbol{\mu})^{\mathsf{T}} \boldsymbol{\Omega}^{-1} (\boldsymbol{\beta} - \boldsymbol{\mu})} \right\}^{1/2} D_* \boldsymbol{\beta}; \boldsymbol{\Sigma}_*, \boldsymbol{\nu} + p \right]. (7)$$

In practice, we typically assume a weakly informative prior for $\boldsymbol{\beta}$, which means ν is often large and Ω is often taken as a diagonal matrix with large diagonal entries. This implies that $(\boldsymbol{\beta} - \boldsymbol{\mu})^{\top} \Omega^{-1} (\boldsymbol{\beta} - \boldsymbol{\mu})$ is often very small compared to ν and $\nu \approx \nu + (\boldsymbol{\beta} - \boldsymbol{\mu})^{\top} \Omega^{-1} (\boldsymbol{\beta} - \boldsymbol{\mu})$. Hence, if we assume that the diagonal entries of Σ are all equal, then Σ needs to be a correlation matrix because $p(\boldsymbol{y} \mid \boldsymbol{\beta}, \Sigma, \boldsymbol{\alpha}, \nu) \approx p(\boldsymbol{y} \mid \boldsymbol{\beta}, \boldsymbol{\beta}, b^2 \Sigma, \boldsymbol{\alpha}, \nu)$ for any positive number *b*. We now state the result that for the skew-*t* link model the posterior of $\boldsymbol{\beta}$ coincides with a unified skew-*t* distribution, which directly follows from Theorem 1 by taking $g^{(nM+1)}$ as the (nM + 1)-variate *t* density generator with ν degrees of freedom.

Corollary 1. Let $\mathbf{y} = (\mathbf{y}_1^{\top}, ..., \mathbf{y}_n^{\top})^{\top}$ be observations from the multivariate skew-t link model and $X = (X_1^{\top}, ..., X_n^{\top})^{\top}$ be the corresponding data matrix. Then

$$\boldsymbol{\beta} \mid \boldsymbol{y}, \boldsymbol{\Sigma}, \boldsymbol{\alpha}, \boldsymbol{\nu} \sim \mathcal{SUT}_{p, nM+1}(\boldsymbol{\mu}_{post}, \boldsymbol{\Omega}_{post}, \boldsymbol{\Lambda}_{post}, \boldsymbol{\nu}, \boldsymbol{\tau}_{post}, \boldsymbol{\Gamma}_{post})$$

where μ_{post} , Ω_{post} , Λ_{post} , and τ_{post} , Γ_{post} are defined in Theorem 1.

Similarly to the unified skew-normal distribution, the unified skew-*t* distribution is also closed under marginalization, conditioning, and affine transformations (Arellano-Valle & Genton, 2010), which simplifies the inference of the posterior marginals, their moments, and functionals such as measures of dependence and credible intervals. Thanks to the stochastic representation of the unified skew-*t* distribution, exact sampling from the distribution of $\beta \mid y, \Sigma, \alpha, \nu$ is also feasible. Specifically, using Equation (9) in Arellano-Valle and Genton (2010), $\beta \mid y, \Sigma, \alpha, \nu$ has the stochastic representation

$$\boldsymbol{\beta} \mid \boldsymbol{y}, \boldsymbol{\Sigma}, \boldsymbol{\alpha}, \boldsymbol{\nu} \stackrel{\mathrm{d}}{=} \boldsymbol{\mu} + \left(\frac{\boldsymbol{\nu} + \boldsymbol{U}_{1}^{\mathsf{T}} \boldsymbol{s} (\boldsymbol{D}_{*} \boldsymbol{\Omega} \boldsymbol{D}_{*}^{\mathsf{T}} + \boldsymbol{\Sigma}_{*})^{-1} \boldsymbol{s} \boldsymbol{U}_{1}}{\boldsymbol{\nu} + \boldsymbol{n} \boldsymbol{M} + 1} \right)^{1/2} \boldsymbol{U}_{0} + \boldsymbol{\Omega} \boldsymbol{D}_{*}^{\mathsf{T}} (\boldsymbol{D}_{*} \boldsymbol{\Omega} \boldsymbol{D}_{*}^{\mathsf{T}} + \boldsymbol{\Sigma}_{*})^{-1} \boldsymbol{s} \boldsymbol{U}_{1},$$
(8)

where $s = \text{diag}(D_*\Omega D_*^\top + \Sigma_*)^{1/2} \in \mathbb{R}^{(nM+1)\times(nM+1)}$, $U_0 \sim \mathcal{T}_p(\mathbf{0}, \Omega - \Omega D_*^\top (D_*\Omega D_*^\top + \Sigma_*)^{-1} D_*\Omega, \nu + nM + 1)$ is independent of U_1 , which follows a (nM + 1)-variate truncated

t distribution with location parameter vector **0**, dispersion matrix $s^{-1}(D_*\Omega D_*^{\top} + \Sigma_*)s^{-1}$, degrees of freedom ν , and truncated below the level $-s^{-1}D_*\mu$.

3.4 | Identifiability of α and ν

In this subsection, we investigate the identifiability of the skewness parameter α and the degree of freedom parameter ν in the skew-t link model. Lemma 1 implies that both α and ν play a role in determining the joint probability mass function $p(\mathbf{y} | \boldsymbol{\beta}, \boldsymbol{\Sigma}, \boldsymbol{\alpha}, \mathbf{g}^{(p+nM+1)})$, and hence they should be identifiable theoretically. Here we conduct a simulation study to verify that. We fix M = 1, $\Sigma = 1$, $\nu = 1$, $\beta =$ $(-1, -0.5, 0.5)^T$, $\mu = 0$, and $\Omega = 25 \times I_3$, and simulate n =100 data from the skew-*t* link model with skewness α = $\alpha_c \times \mathbf{1}_n$, $\alpha_c = -2$ or 2. The first column of the data matrix X is set to 1, and the other two columns are generated from a standard normal distribution. Then we compute the log-likelihood log $p(\mathbf{y} | \boldsymbol{\beta}, \boldsymbol{\Sigma}, \boldsymbol{\alpha}, g^{(p+nM+1)})$ with respect to different values of α_c . Specifically, all the other parameters except α are treated as nuisance parameters and we use their true values to compute $\log p(\mathbf{y} | \boldsymbol{\beta}, \boldsymbol{\Sigma}, \boldsymbol{\alpha}, g^{(p+nM+1)})$. Similarly, we fix $\alpha_c = -2$ and simulate n = 100 data from the skew-*t* link model with degrees of freedom $\nu = 1, 2$, or 5. Then the log-likelihood with respect to various values of ν is computed.

Figure 1 displays the log-likelihood with respect to varying α_c and ν . It shows that the skewness parameter α is weakly identifiable, which might be due to the fact that α appears only through the matrix Σ_* , and if α_c is large, then the vector δ in Lemma 1 would be approximately $(I_n \otimes \overline{\Sigma}) \mathbf{1}_n \operatorname{sign}(\alpha_c)$ and hence only the information about the sign of α_c is contained in the likelihood. Another observation is that small values of ν ($\nu = 1, 2$) are identifiable, while larger values of ν (e.g., $\nu = 5$) are difficult to identify. One explanation is that as ν increases, the skew-*t* link model tends to the skew-normal link model and thus larger values of ν would be more difficult to estimate.

4 | EMPIRICAL STUDIES

4.1 | Prior and posterior for α and Σ

As the skew-normal link model is a limiting case of the skew-*t* link model when the degree of freedom ν tends to ∞ , in this section we focus on the skew-*t* link model and consider a real-data application. A simulation study is provided in the Supporting Information. To make the model parsimonious, in both the simulation study and data application we assume that the skewness parameters are the same across different observations, that is, $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_M, \dots, \alpha_1, \dots, \alpha_M)^{\mathsf{T}} \in \mathbb{R}^{nM}$, and Σ is a



FIGURE 1 Log-likelihood with respect to α_c and ν . Left panel: true value of α_c is $\alpha_T = -2$ (black line) and $\alpha_T = 2$ (red line); right panel: true value of ν is $\nu_T = 1$ (black line), $\nu_T = 2$ (red line), and $\nu_T = 5$ (green line). This figure appears in color in the electronic version of this article, and any mention of color refers to that version

correlation matrix,that is, $\Sigma = \overline{\Sigma}$. The assumption of a correlation matrix for Σ is not very restrictive because it is approximately equivalent to assuming that all the diagonal entries in Σ are equal, as we discussed in Section 3.3. Now we specify the prior and posterior for the skewness parameter $\boldsymbol{\alpha}_s = (\alpha_1, ..., \alpha_M)^{\top}$ and the correlation matrix $\overline{\Sigma}$.

Bayesian modeling of unstructured covariance or correlation matrices is a fundamental and difficult task because of the constraint of positive definiteness and the quadratic increase of the number of parameters with respect to the number of correlated variables. More importantly, it is difficult to specify a prior for them (Gelman et al., 2014). Typical priors for correlation matrices include the marginally uniform prior, the jointly uniform prior (Barnard et al., 2000), and the so-called Lewandowski–Kurowicka–Joe (LKJ) prior (Lewandowski et al., 2009).

The marginally uniform prior means that each nondiagonal element in the correlation matrix has a uniform marginal distribution over [-1, 1], whereas the jointly uniform prior means that the correlation matrix has a joint uniform distribution over the compact space of valid correlation matrices. The LKJ prior is recommended in the R library rstan (Stan Development Team, 2022) and has the form $\pi(\bar{\Sigma}) \propto |\bar{\Sigma}|^{\eta-1}$, where $|\bar{\Sigma}|$ is the determinant of $\bar{\Sigma}$ and $\eta > 0$ is the shape parameter of the LKJ distribution. The jointly uniform prior is a special case of the LKJ prior when $\eta = 1$.

In this work, we adopt the jointly uniform prior for $\overline{\Sigma}$ by setting $\eta = 1$ in the LKJ prior and specify an independent weakly informative Gaussian prior π_{α_s} for α_s . Then, using Equation (7), the joint posterior of $(\overline{\Sigma}, \alpha_s)$ given the data

and the regression coefficients is

$$2T_{nM+1}\left[\left\{\frac{\nu+p}{\nu+(\beta-\mu)^{\top}\Omega^{-1}(\beta-\mu)}\right\}^{1/2}D_{*}\beta;\Sigma_{*},\nu+p\right]\pi_{\alpha_{s}}(\alpha_{s}).$$
 (9)

We evaluate the multivariate t probability on the righthand side of (9) using the R library tlrmvnmvt (Cao et al., 2022), which implements the classic Genz algorithm (Genz & Bretz, 1999) and exploits a tile-low-rank algorithm (Cao et al., 2021) to speed up the computation of the multivariate normal and t probabilities. To avoid sampling the correlation matrix from a constrained space, we consider the reparameterization adopted in Smith (2013), which reexpresses a correlation matrix in terms of the Cholesky factor of a positive definite matrix $\bar{\Sigma} = \Lambda_{\bar{\Sigma}}^{-1/2} L_{\bar{\Sigma}} L_{\bar{\Sigma}}^{\top} \Lambda_{\bar{\Sigma}}^{-1/2}$, where $L_{\bar{\Sigma}}$ is a lower triangular matrix and $\Lambda_{\bar{\Sigma}} = \text{diag}(L_{\bar{\Sigma}}L_{\bar{\Sigma}}^{\top})$. Here the diagonal entries of $L_{\bar{\Sigma}}$ are set to 1 such that the correspondence between $L_{\bar{\Sigma}}$ and $\bar{\Sigma}$ is one-to-one. We denote the collection of the M(M-1)/2 unconstrained parameters in $L_{\Sigma} = (l_{ij})$ by θ , that is, $\theta = \{l_{ij} : i > j, i, j =$ 1,..., M}, and the M(M-1)/2 constrained parameters in $\overline{\Sigma}$ by $vec(\bar{\Sigma})$. Then, using a change of variables we get the posterior of $(\boldsymbol{\theta}, \boldsymbol{\alpha}_s)$ as $p(\boldsymbol{\theta}, \boldsymbol{\alpha}_s \mid \boldsymbol{y}, \boldsymbol{\beta}, \boldsymbol{\nu}) = p(\bar{\Sigma}, \boldsymbol{\alpha}_s \mid \boldsymbol{y}, \boldsymbol{\beta}, \boldsymbol{\nu})|J| =$ $p(\bar{\Sigma}, \alpha_s \mid y, \beta, \nu) \prod_{i=1}^{M} (1 + \sum_{j < i} l_{ij}^2)^{-(M+1)/2}$, where |J| = $|\partial \operatorname{vec}(\bar{\Sigma})/\partial \theta|$ is the determinant of the Jacobian matrix of this transformation.

As direct sampling from the conditional distribution of θ , $\alpha_s \mid \mathbf{y}, \boldsymbol{\beta}, \nu$ is unknown, we propose to use a random walk Metropolis–Hastings algorithm to generate samples from it. Specifically, we first sample α'_s from a proposal

ALGORITHM 1 MCMC sampling scheme for the multivariate ST link model

Initialization: Set $\boldsymbol{\beta}^{(0)}, \bar{\Sigma}^{(0)}, \boldsymbol{\alpha}^{(0)}$; for iteration k from 1 to K do [1] Sample $\boldsymbol{U}_{0}^{(k)}$ from $\mathcal{T}_{p}(\mathbf{0}, \Omega - \Omega D_{*}^{\top}(D_{*}\Omega D_{*}^{\top} + \Sigma_{*})^{-1}D_{*}\Omega, \nu + nM + 1)$ (in R use rmvt); [2] Sample $\boldsymbol{U}_{1}^{(k)}$ from a (nM + 1)-variate truncated t distribution with location parameter vector **0**, dispersion matrix $s^{-1}(D_{*}\Omega D_{*}^{\top} + \Sigma_{*})s^{-1}$, degrees of freedom ν , and truncated below the level $-s^{-1}D_{*}\boldsymbol{\mu}$, using the accept-reject algorithm of Botev (2017) (in R use mvrandt); [3] Compute $\boldsymbol{\beta}^{(k)}$ via $\boldsymbol{\beta}^{(k)} = \boldsymbol{\mu} + \left\{ \frac{\nu + (\boldsymbol{U}_{1}^{(k)})^{\top}s(D_{*}\Omega D_{*}^{\top} + \Sigma_{*})^{-1}s\boldsymbol{U}_{1}^{(k)}}{\nu + nM + 1} \right\}^{1/2}\boldsymbol{U}_{0}^{(k)} + \Omega D_{*}^{\top}(D_{*}\Omega D_{*}^{\top} + \Sigma_{*})^{-1}s\boldsymbol{U}_{1}^{(k)};$ [4] Use the Metropolis-Hastings algorithm described in Section 4.1 to sample $(\boldsymbol{\theta}^{(k)}, \boldsymbol{\alpha}_{*}^{(k)})$ from the distribution of $(\boldsymbol{\theta}, \boldsymbol{\alpha}_{s} \mid \boldsymbol{y}, X, \boldsymbol{\beta}^{(k)})$, then return the resulting $\bar{\Sigma}^{(k)}$ and $\boldsymbol{\alpha}^{(k)}$. Output: $(\boldsymbol{\beta}^{(1)}, \bar{\Sigma}^{(1)}, \boldsymbol{\alpha}^{(1)}), \dots, (\boldsymbol{\beta}^{(K)}, \bar{\Sigma}^{(K)}, \boldsymbol{\alpha}^{(K)})$

distribution with density $q(\cdot | \boldsymbol{\alpha}_s)$ and $\boldsymbol{\theta}'$ from a proposal distribution with density $r(\cdot | \boldsymbol{\theta})$. Here we take both proposal densities q and r as symmetric normal densities, that is, $\boldsymbol{\alpha}'_s | \boldsymbol{\alpha}_s \sim \mathcal{N}_M(\boldsymbol{\alpha}_s, h_1 \mathbf{I}_M)$ and $\boldsymbol{\theta}' | \boldsymbol{\theta} \sim \mathcal{N}_K(\boldsymbol{\theta}, h_2 \mathbf{I}_K), K = M(M-1)/2$. Then the acceptance probability is $\alpha((\boldsymbol{\alpha}_s, \boldsymbol{\theta}), (\boldsymbol{\alpha}'_s, \boldsymbol{\theta}')) = \min \left\{ \frac{p(\boldsymbol{\theta}', \boldsymbol{\alpha}'_s | \boldsymbol{y}, \boldsymbol{X}, \boldsymbol{\beta}) \mathbf{1}((\boldsymbol{\theta}', \boldsymbol{\alpha}'_s) \in C)}{p(\boldsymbol{\theta}, \boldsymbol{\alpha}_s | \boldsymbol{y}, \boldsymbol{X}, \boldsymbol{\beta}) \mathbf{1}((\boldsymbol{\theta}, \boldsymbol{\alpha}_s) \in C)}, 1 \right\}$, where $1(\cdot)$ is the indicator function and C is the space of all $(\boldsymbol{\theta}, \boldsymbol{\alpha}_s)$ such that the resulting matrix $\bar{\boldsymbol{\Sigma}} - \boldsymbol{\delta} \boldsymbol{\delta}^{\top}$ is positive definite with $\boldsymbol{\delta} = (1 + \boldsymbol{\alpha}^{\top} \bar{\boldsymbol{\Sigma}} \boldsymbol{\alpha})^{-1/2} \bar{\boldsymbol{\Sigma}} \boldsymbol{\alpha}$.

4.2 | MCMC sampling scheme

As sampling from the distribution of $\beta \mid y, X, \overline{\Sigma}, \alpha$ is feasible using (8) and sampling from the distribution of $\overline{\Sigma}, \alpha \mid y, X, \beta$ has been described in Section 4.1, we now combine them to construct a Markov chain Monte Carlo (MCMC) sampler for the multivariate skew-*t* link model; see Algorithm 1.

4.3 | Application to COVID-19 pandemic data

Biometrics WILEY 1795

In this subsection, we illustrate our methodology on COVID-19 pandemic data from different counties of the state of California, USA, freely downloaded from the California Open Data Portal (2022). The dataset contains the number of daily new confirmed cases and deaths from March 18, 2020, to March 10, 2021, in 58 counties of California. There is a clear weekly cyclic pattern in this dataset, that is, the numbers of new confirmed cases on weekdays are often much larger than those during the weekends. To avoid modeling this artificial cyclic pattern, we aggregate the data and consider the weekly new confirmed cases, resulting in n = 51 weekly observations. As n is relatively small, we here only focus on the three most populous counties in California, that is, Los Angeles, San Diego, and Orange.

To remove the obvious trend, we apply cubic smoothing splines with six knots to the logarithm of each of the three time series, where the logarithm is used because most



FIGURE 2 Upper panel: smoothing splines for the time series of weekly new confirmed cases at Los Angeles (left), Orange (middle), and San Diego (right). Lower panel: the residuals obtained as the difference between the original data and the fitted splines, with red points considered as extreme spikes and green points as nonextreme values. This figure appears in color in the electronic version of this article, and any mention of color refers to that version

epidemics grow approximately exponentially during the initial phase (Ma, 2020). Figure 2 displays the observed data for the three counties, the smoothing splines for each time series, and the resulting residuals. We then consider a residual point as an extreme spike if it exceeds the empirical 90% quantile of the corresponding time series, and we denote it as 1; otherwise, we denote it as 0. In this way, we get three imbalanced binary time series and we aim to model the dependence among them.

We consider three covariates in total, that is, an intercept, one covariate as time, and another one as the square of time. Following the recommendation of Gelman et al. (2008), we standardize the two temporal predictors in a preliminary step to make them have a mean of 0 and a standard deviation of 1. To assess the performance of the multivariate skew-elliptical link model, we consider six models $\mathcal{M}_1, \mathcal{M}_2, \mathcal{M}_3, \mathcal{M}_4, \mathcal{M}_5$, and \mathcal{M}_6 of different complexity. $\mathcal{M}_1, \mathcal{M}_2$, and \mathcal{M}_3 are the multivariate skew-*t* link model with $\nu = 1, 2, 5$, respectively, \mathcal{M}_4 is the multivariate skew-normal model (i.e., obtained as $\nu \to \infty$), \mathcal{M}_5 is the multivariate probit model (obtained with $\nu \to \infty$ and $\alpha = 0$), and \mathcal{M}_6 is the independent probit model (obtained with $\nu \to \infty$, $\bar{\Sigma} = I_3$, $\alpha = 0$).

For each of these models, we run Algorithm 1 for 5000 iterations and remove the first 2000 samples as burn-in. The prior for the regression parameters $\boldsymbol{\beta}$ is specified as $\mathcal{T}_3(\mathbf{0}, 25I_3, \nu)$ for the models $\mathcal{M}_1, \mathcal{M}_2, \mathcal{M}_3$, and $\mathcal{N}_3(\mathbf{0}, 25I_3)$ for the models $\mathcal{M}_4, \mathcal{M}_5$, and \mathcal{M}_6 . The prior for the skewness parameters is taken as $\mathcal{N}_3(\mathbf{0}, 16I_3)$. The variances of the proposal densities in the Metropolis–Hastings algorithm are taken the same as in the simulation study, that is, $h_1 = 9$ and $h_2 = 0.36$.

Table 1 summarizes the estimation results for all the models. The results show that the estimates of the intercept for all the models are significantly negative. This is expected as 90% of the observations are 0, and only 10% are 1. We also observe that the credible intervals for the correlation and skewness parameters are generally quite large (as in the simulation study), implying that they are hard to estimate with only n = 51 observations. However, the

	\mathcal{M}_1 (numb	her of parameters $= 9$,	DIC = 86.2)	\mathcal{M} , (numb	er of parameters $= 9$,	DIC = 88.5)	\mathcal{M}_3 (numb	er of parameters = 9, I	DIC = 88.4
	Est	90% CI	Prob	Est	90% CI	Prob	Est	90% CI	Prob
Intercept	-9.61	(-23.46, -0.74)	100%	-2.92	(-6.70, -0.84)	100%	-1.79	(-3.28, -0.93)	100%
Time	9.52	(-0.08, 25.15)	94%	3.25	(-0.01, 9.23)	95%	1.98	(-0.04, 4.83)	95%
Time ²	-6.54	(-18.13, 0.63)	88%	-2.34	(-7.15, 0.39)	80%	-1.43	(-3.76, 0.36)	89%
$ar{\Sigma}_{12}$	0.40	(-0.03, 0.77)	93%	0.42	(-0.04, 0.80)	93%	0.41	(-0.05, 0.76)	91%
$ar{\Sigma}_{13}$	0.63	(0.26, 0.89)	%66	0.67	(0.30, 0.93)	%66	0.62	(0.16, 0.89)	98%
$ar{\Sigma}_{23}$	0.64	(0.28, 0.88)	100%	0.64	(0.24, 0.93)	100%	0.64	(0.33, 0.90)	100%
$lpha_1$	-0.04	(-6.14, 6.20)	49%	-0.70	(-7.25, 5.73)	57%	0.34	(-6.36, 7.14)	56%
α_2	-0.64	(-7.25, 5.89)	45%	-0.05	(-6.37, 7.71)	47%	-0.21	(-6.69, 6.77)	48%
α_3	0.12	(-6.67, 7.17)	49%	0.17	(-7.15, 7.64)	52%	0.23	(-5.47, 6.24)	49%
	\mathcal{M}_4 (numb	er of parameters =9,	DIC = 86.3)	\mathcal{M}_5 (numb	oer of parameters $= 9$,	DIC = 86.7	\mathcal{M}_6 (numb	er of parameters = 9, l	DIC = 98.1)
	Est	90% CI	Prob	Est	90% CI	Prob	Est	90% CI	Prob
Intercept	-1.48	(-1.89, -1.13)	100%	-1.50	(-1.90, -1.15)	100%	-1.49	(-1.82, -1.21)	100%
Time	1.64	(-0.05, 3.51)	94%	1.60	(-0.06, 3.54)	94%	1.47	(0.10, 3.00)	97%
$Time^{2}$	-1.18	(-2.80, 0.37)	89%	-1.15	(-2.82, 0.37)	88%	-0.99	(-2.34, 0.22)	91%
$ar{\Sigma}_{12}$	0.42	(-0.05, 0.77)	94%	0.41	(-0.04, 0.78)	94%			
$ar{\Sigma}_{13}$	0.63	(0.27, 0.89)	%66	0.63	(0.29, 0.89)	%66			
$ar{\Sigma}_{23}$	0.64	(0.27, 0.89)	100%	0.66	(0.32, 0.90)	100%			
$lpha_1$	0.88	(-6.04, 7.22)	59%						
α_2	0.43	(-6.12, 6.64)	56%						
α_3	0.05	(-6.22, 5.77)	51%						
<i>Note</i> : The first column is Abbreviations: DIC, devi	the posterior manual and a statements of the posterior of the second statements of the second st	ean, the second column is t n criterion credible interval.	he 90% credible interval.	(CI), and the th	nird column is the empiric	al probability that the po	sterior samples	have the same sign as the po	osterior mean.

Posterior estimates, number of parameters, and estimated DIC for different models fitted in our COVID-19 data application in Section 4.3 TABLE 1

1424024 (2012) 3. Dwnladed from https://onlinelibrary.wiky.com/doi/10.1111/bom.13731 by Cochrane Saudi Arabia, Wiley Online Library on [1209/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons Licenses

1798 WILEY Biometrics

correlation between the counties of Orange and San Diego, and, $\bar{\Sigma}_{23}$, seems to be quite strong, as its posterior mean for models $\mathcal{M}_1, \mathcal{M}_2, \mathcal{M}_3, \mathcal{M}_4$, and \mathcal{M}_5 is consistently far from zero (with an estimate close to 0.64) and its 90% credible interval always excludes zero. This indicates that these two counties are more connected together in terms of extreme COVID-19 cases than the other pairs of counties considered, which sheds some light on the spread of the epidemic. The extreme occurrences observed in the counties of Los Angeles and San Diego also seem fairly strongly interconnected since the estimate of Σ_{13} is also quite high, yet to a slightly milder degree.

To compare the different fitted models, we use the deviance information criterion (DIC) proposed by Spiegelhalter et al. (2002). The DIC is the Bayesian analogue of the Akaike information criterion and is defined as DIC = $D(\bar{\kappa}) + 2p_D$, where κ denotes the collection of all the parameters, $\bar{\boldsymbol{\kappa}} = \mathbf{E}[\boldsymbol{\kappa} \mid \boldsymbol{y}]$ is its posterior mean, $D(\cdot)$ is a deviance function and $p_D = E[D(\boldsymbol{\kappa}) | \boldsymbol{y}] - D(\bar{\boldsymbol{\kappa}})$ is the effective number of model parameters. Here we take the deviance function $D(\boldsymbol{\kappa})$ as $-2\log p(\boldsymbol{y} \mid \boldsymbol{\beta}, \bar{\Sigma}, \boldsymbol{\alpha}, \boldsymbol{\nu})$ when the model is the skew-t link model, or $-2\log p(\mathbf{y} \mid \boldsymbol{\beta}, \bar{\boldsymbol{\Sigma}}, \boldsymbol{\alpha})$ when the model is the skew-normal link model and estimate $E[D(\kappa) | y]$ by Monte Carlo using the samples generated from Algorithm 1. The smaller the DIC value, the better the model's goodness-of-fit and predictive performance. We refer to Spiegelhalter et al. (2002) for other properties about the DIC measure.

Table 1 reports the estimated DIC values for the six different models. The results show that the multivariate skew-*t* link model with a degree of freedom $\nu = 1$ provides the best fit to the data despite its high complexity, the multivariate skew-normal link model \mathcal{M}_4 has the second best performance and the independent symmetric probit model \mathcal{M}_6 is the worst. This has three major implications. The first is that spatial dependence plays an important role in the spread of the epidemic and ignoring the correlation would lead to a poor fit of the extreme spikes. The second is that adding the skewness parameter indeed improves the model's flexibility and can provide a better fit to our highly imbalanced dataset. The third is that for this data application, very heavy-tailed link (skew-t with $\nu = 1$) and very light-tailed link (skew-normal) seem to describe the data better than mild heavy-tailed links (skew-*t* with $\nu = 2, 5$), and the tail heaviness appears to have a larger effect than the skewness.

5 | DISCUSSION

Although we here focus on the skew-elliptical link model, the result of a closed-form posterior for the regression coefficients could also be obtained if we consider a more flexible class of distributions for the assumption (6). In fact, if $\varepsilon \mid \beta, \Sigma, \alpha$ has a distribution which is closed under affine transformation, following the proof of Lemma 1 and Theorem 1, one can show that the posterior of β coincides with a fundamental skew distribution (Arellano-Valle & Genton, 2005). This novel result opens up new avenues for the development of skewed link models for correlated binary data.

In this paper, we only considered the normal and t density generators because they are the most commonly used ones, but our results hold for any elliptical density generators. Moreover, since the skew-normal link model is a limiting model of the skew-t link model as the degree of freedom parameter ν tends to infinity, in practice, one can simply choose the skew-t link model with a series of different degrees of freedom for convenience, and then select the best-performing model based on certain measures such as DIC. Alternatively, if one has a more efficient algorithm to sample from high-dimensional truncated t distributions, one might try to include the estimation of the parameter ν in the Metropolis–Hastings algorithm described in Section 4.1.

There are various directions for future research. As the number of observations in our dataset is relatively small, we chose not to consider too many covariates and restricted the number of counties. An interesting extension of our real data application would be to consider a larger dataset with more informative covariates, such as daily weather information or population migration between different counties. Adding such extra covariates could potentially fit the data better and provide a more detailed and informed explanation of the spread of the epidemic. Another interesting methodological extension is to improve Algorithm 1. As we used the accept-reject algorithm of Botev (2017) within Algorithm 1 to sample from a multivariate truncated t distribution, its lack of scalability to higher dimensions is inevitably inherited. Therefore, more efficient and scalable algorithms to sample from high-dimensional truncated normal and t distributions would significantly improve the performance of Algorithm 1. Finally, although the skew-elliptical link models offer extra flexibility in modeling correlated binary data, there might be identifiability issues, as shown in Section 3.4, and the skewness parameters are especially difficult to estimate. So one research question is how to solve or avoid these issues, possibly by carefully designing informative prior distributions.

DATA AVAILABILITY STATEMENT

The data can be freely downloaded from the California Open Data Portal (https://data.ca.gov).

ORCID

Zhongwei Zhang https://orcid.org/0000-0002-2070-1993

Marc G. Genton https://orcid.org/0000-0001-6467-2998 *Raphaël Huser* https://orcid.org/0000-0002-1228-2071

REFERENCES

- Arellano-Valle, R.B. & Azzalini, A. (2006) On the unification of families of skew-normal distributions. *Scandinavian Journal of Statistics*, 33, 561–574.
- Arellano-Valle, R.B. & Genton, M.G. (2005) On fundamental skew distributions. *Journal of Multivariate Analysis*, 96(1), 93–116.
- Arellano-Valle, R.B. & Genton, M.G. (2010) Multivariate unified skew-elliptical distributions. *Chilean Journal of Statistics*, 1(1), 17–33.
- Ashford, J.R. & Sowden, R.R. (1970) Multivariate probit analysis. *Biometrics*, 26(3), 535–546.
- Azzalini, A. & Capitanio, A. (1999) Statistical applications of the multivariate skew normal distribution. *Journal of the Royal Statistics Society (Series B)*, 61(3), 579–602.
- Azzalini, A. & Capitanio, A. (2003) Distributions generated by perturbation of symmetry with emphasis on a multivariate skew *t*-distribution. *Journal of the Royal Statistics Society (Series B)*, 65(2), 367–389.
- Azzalini, A. & Dalla Valle, A. (1996) The multivariate skew-normal distribution. *Biometrika*, 83(4), 715–726.
- Barnard, J., McCullloch, R. & Meng, X. (2000) Modeling covariance matrices in terms of standard deviations and correlations, with application to shrinkage. *Statistica Sinica*, 10, 1281–1311.
- Bazán, J.L., Bolfarine, H. & Branco, M.D. (2010) A framework for skew-probit links in binary regression. *Communications in Statistics - Theory and Methods*, 39, 678–697.
- Botev, Z.I. (2017) The normal law under linear restrictions: simulation and estimation via minimax tilting. *Journal of the Royal Statistical Society (Series B)*, 79, 125–148.
- Branco, M.D. & Dey, D.K. (2001) A general class of multivariate skewelliptical distributions. *Journal of Multivariate Analysis*, 79, 99–113.
- California Open Data Portal (2022) https://data.ca.gov/ (accessed 5 April. 2022).
- Cao, J., Genton, M.G., Keyes, D.E. & Turkiyyah, G.M. (2021) Exploiting low rank covariance structures for computing highdimensional normal and Student-*t* probabilities. *Statistics and Computing*, 31(2), 2.
- Cao, J., Genton, M.G., Keyes, D.E. & Turkiyyah, G.M. (2022) tlrmvnmvt: computing high-dimensional multivariate normal and Student-t probabilities with low-rank methods in R. Journal of Statistical Software, 101, 4.
- Chen, M.H., Dey, D.K. & Shao, Q.M. (1999) A new skewed link model for dichotomous quantal response data. *Journal of the American Statistical Association*, 94(448), 1172–1186.
- Chib, S. & Greenberg, E. (1998) Analysis of multivariate probit models. *Biometrika*, 85(2), 347–361.
- Czado, C. & Santner, T.J. (1992) The effect of link misspecification on binary regression inference. *Journal of Statistical Planning and Inference*, 33, 213–231.
- Durante, D. (2019) Conjugate Bayes for probit regression via unified skew-normal distributions. *Biometrika*, 106(4), 765–779.
- Fang, B.Q. (2003) The skew elliptical distributions and their quadratic forms. *Journal of Multivariate Analysis*, 87, 298–314.

- Gelman, A., Carlin, J.B., Stern, H.S., Dunson, D.B., Vehtari, A. & Rubin, D.B. (2014) *Bayesian data analysis*, 3rd edition. Boca Raton, FL: CRC Press.
- Gelman, A., Jakulin, A., Pittau, M.G. & Su, Y.S. (2008) A weakly informative default prior distribution for logistic and other regression models. *The Annals of Applied Statistics*, 2(4), 1360–1383.
- Genz, A. & Bretz, F. (1999) Numerical computation of multivariate *t*-probabilities with application to power calculation of multiple contrasts. *Journal of Statistical Computation and Simulation*, 63, 361–378.
- Johndrow, J.E., Smith, A., Pillai, N. & Dunson, D.B. (2019) MCMC for imbalanced categorical data. *Journal of the American Statistical Association*, 14(527), 1394–1403.
- Kim, H. (2002) Binary regression with a class of skewed *t* link models. *Communications in Statistics - Theory and Methods*, 31(10), 1863–1886.
- Kim, S., Chen, M.H. & Dey, D.K. (2008) Flexible generalized *t*-link models for binary response data. *Biometrika*, 95(1), 93–106.
- Lewandowski, D., Kurowicka, D. & Joe, H. (2009) Generating random correlation matrices based on vines and extended onion method. *Journal of Multivariate Analysis*, 100, 1989–2001.
- Ma, J. (2020) Estimating epidemic exponential growth rate and basic reproduction number. *Infectious Disease Modelling*, 5, 129–141.
- Smith, M.S. (2013) Bayesian approaches to copula modelling. In P. Damien, P. Dellaportas, N. G. Polson, & D. A. Stephens (Eds.), *Bayesian theory and applications*. Oxford, UK: Oxford University Press, chapter 17, pp. 336–358.
- Spiegelhalter, D.J., Best, N.G., Carlin, B.P. & van der Linde, A. (2002) Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society (Series B)*, 64, 583–639.
- Stan Development Team (2022) RStan: the R interface to Stan. R package version 2.21.5.
- Wang, X. & Dey, D.K. (2010) Generalized extreme value regression for binary response data: an application to B2B electronic payments system adoption. *The Annals of Applied Statistics*, 4(4), 2000–2023.
- Zhao, X., Zhang, L. & Bandyopadhyay, D. (2021) A shared spatial model for multivariate extreme-valued binary data with non-random missingness. *Sankhya B*, 83, 374–396.

SUPPORTING INFORMATION Data S1

Tables and Figures referenced in Sections 3.4 and 4.3, and more details about the skew-normal link model and a simulation study are available with this paper at the Biometrics website on Wiley Online Library. For reproducibility, we also provide the R code that we developed, with a small example showing how to replicate our data analysis.

How to cite this article: Zhang, Z.,

Arellano-Valle, R.B., Genton, M.G., & Huser, R. (2023) Tractable Bayes of skew-elliptical link models for correlated binary data. *Biometrics*, 79, 1788–1800. https://doi.org/10.1111/biom.13731

Biometrics WILEY 1799

APPENDIX A

Proof of Lemma 1. Since a diagonal matrix diag(x) with $x \in \{-1, 1\}^{nM}$ has the property diag(x) $x = \mathbf{1}_{nM}$ and $(\text{diag}(x))^{-1} = \text{diag}(x)$, we have

$$p(\mathbf{y} \mid \boldsymbol{\beta}, \boldsymbol{\Sigma}, \boldsymbol{\alpha}, g^{(p+nM+1)}) = \Pr(\mathbf{Y} = \mathbf{y} \mid \boldsymbol{\beta}, \boldsymbol{\Sigma}, \boldsymbol{\alpha}, g^{(p+nM+1)})$$
$$= \Pr(2\mathbf{Y} - \mathbf{1}_{nM} = 2\mathbf{y} - \mathbf{1}_{nM} \mid \boldsymbol{\beta}, \boldsymbol{\Sigma}, \boldsymbol{\alpha}, g^{(p+nM+1)})$$
$$= \Pr\{D(2\mathbf{Y} - \mathbf{1}_{nM}) = \mathbf{1}_{nM} \mid \boldsymbol{\beta}, \boldsymbol{\Sigma}, \boldsymbol{\alpha}, g^{(p+nM+1)}\}$$
$$= \Pr(D\mathbf{Y}^* > \mathbf{0} \mid \boldsymbol{\beta}, \boldsymbol{\Sigma}, \boldsymbol{\alpha}, g^{(p+nM+1)})$$
$$= \Pr(-D\varepsilon - DX\boldsymbol{\beta} < \mathbf{0} \mid \boldsymbol{\beta}, \boldsymbol{\Sigma}, \boldsymbol{\alpha}, g^{(p+nM+1)}).$$

By (6), $\boldsymbol{\varepsilon} \mid \boldsymbol{\beta}, \boldsymbol{\Sigma} \boldsymbol{\alpha}, g^{(p+nM+1)} \sim \mathcal{SE}_{nM}(\mathbf{0}, \mathbf{I}_n \otimes \boldsymbol{\Sigma}, \boldsymbol{\alpha}, g^{(nM+1)}_{q(\boldsymbol{\beta})})$. Using Proposition 1 in Fang (2003), we know that

$$\begin{split} (-D\boldsymbol{\varepsilon} - D\boldsymbol{X}\boldsymbol{\beta}) \mid \boldsymbol{\beta}, \boldsymbol{\Sigma}, \boldsymbol{\alpha}, \boldsymbol{g}^{(p+nM+1)} \\ &\sim \boldsymbol{S} \boldsymbol{\mathcal{E}}_{nM} (-D\boldsymbol{X}\boldsymbol{\beta}, \boldsymbol{D}(\boldsymbol{I}_n \otimes \boldsymbol{\Sigma})\boldsymbol{D}, \boldsymbol{D}\boldsymbol{\alpha}, \boldsymbol{g}_{q(\boldsymbol{\beta})}^{(nM+1)}) \end{split}$$

Using (2), we finally get

$$p(\mathbf{y} \mid \boldsymbol{\beta}, \boldsymbol{\Sigma}, \boldsymbol{\alpha}, \mathbf{g}^{(p+nM+1)}) = 2G_{nM+1}(D_*\boldsymbol{\beta}; \boldsymbol{\Sigma}_*, \mathbf{g}_{q(\boldsymbol{\beta})}^{(nM+1)}).$$

Proof of Theorem 1. The posterior density of β is

$$p(\boldsymbol{\beta} \mid \boldsymbol{y}, \boldsymbol{\Sigma}, \boldsymbol{\alpha}, g^{(p+nM+1)}) \propto p(\boldsymbol{y} \mid \boldsymbol{\beta}, \boldsymbol{\Sigma}, \boldsymbol{\alpha}, g^{(p+nM+1)})$$
$$\times p(\boldsymbol{\beta} \mid g^{(p+nM+1)}).$$

Using Lemma 1 and the assumption (5), we have

$$p(\boldsymbol{\beta} \mid \boldsymbol{y}, \boldsymbol{\Sigma}, \boldsymbol{\alpha}, g^{(p+nM+1)})$$

$$\propto G_{nM+1}(D_*\boldsymbol{\beta}; \boldsymbol{\Sigma}_*, g_{q(\boldsymbol{\beta})}^{(nM+1)})g^{(p)}\{(\boldsymbol{\beta} - \boldsymbol{\mu})^{\top} \boldsymbol{\Omega}^{-1}(\boldsymbol{\beta} - \boldsymbol{\mu})\}$$

$$= G_{nM+1}(\sigma_*^{-1}D_*\boldsymbol{\beta}; \boldsymbol{\bar{\Sigma}}_*, g_{q(\boldsymbol{\beta})}^{(nM+1)})g^{(p)}\{(\boldsymbol{\beta} - \boldsymbol{\mu})^{\top} \boldsymbol{\Omega}^{-1}(\boldsymbol{\beta} - \boldsymbol{\mu})\}$$

$$= G_{nM+1}\{\sigma_*^{-1}D_*\boldsymbol{\mu} + \sigma_*^{-1}D_*(\boldsymbol{\beta} - \boldsymbol{\mu}); \boldsymbol{\bar{\Sigma}}_*, g_{q(\boldsymbol{\beta})}^{(nM+1)}\}$$

$$\times g^{(p)}\{(\boldsymbol{\beta} - \boldsymbol{\mu})^{\top} \boldsymbol{\Omega}^{-1}(\boldsymbol{\beta} - \boldsymbol{\mu})\}$$

$$= G_{nM+1}\{\sigma_*^{-1}D_*\boldsymbol{\mu} + \sigma_*^{-1}D_*\boldsymbol{\omega}\boldsymbol{\omega}^{-1}(\boldsymbol{\beta} - \boldsymbol{\mu}); \boldsymbol{\bar{\Sigma}}_*, g_{q(\boldsymbol{\beta})}^{(nM+1)}\}$$

$$\times g^{(p)}\{(\boldsymbol{\beta} - \boldsymbol{\mu})^{\top} \boldsymbol{\Omega}^{-1}(\boldsymbol{\beta} - \boldsymbol{\mu})\}$$

$$= G_{nM+1}\{\tau_{\text{post}} + \Lambda_{\text{post}}\boldsymbol{\omega}^{-1}(\boldsymbol{\beta} - \boldsymbol{\mu}); \boldsymbol{\bar{\Sigma}}_*, g_{q(\boldsymbol{\beta})}^{(nM+1)}\}$$

$$\times g^{(p)}\{(\boldsymbol{\beta} - \boldsymbol{\mu})^{\top} \boldsymbol{\Omega}^{-1}(\boldsymbol{\beta} - \boldsymbol{\mu})\}.$$

Hence, the required posterior is $\boldsymbol{\beta} \mid \boldsymbol{y}, \boldsymbol{\Sigma}, \boldsymbol{\alpha}, g^{(p+nM+1)} \sim SU\mathcal{E}_{p,nM+1}(\boldsymbol{\mu}_{post}, \boldsymbol{\Omega}_{post}, \boldsymbol{\Lambda}_{post}, \boldsymbol{\tau}_{post}, \boldsymbol{\Gamma}_{post}, g^{(p+nM+1)}).$