Biometrics 000, 1–??
000 2021
DOI: 000

Tractable Bayes of Skew-Elliptical Link Models for Correlated Binary Data

Zhongwei Zhang[1], Reinaldo B. Arellano-Valle[2], Marc G. Genton[1], Raphaël Huser[1,*]

[1]Statistics Program, Computer, Electrical and Mathematical Sciences and Engineering (CEMSE) Division, King Abdullah University of Science and Technology (KAUST), Thuwal 23955-6900, Saudi Arabia

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

[*]*email*: raphael.huser@kaust.edu.sa

Summary: 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.

Key words: 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 and Sowden, 1970; Chib and 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 modelling of dependence, has 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), Kim et al. (2008) for some examples. In this case, the link might be misspecified, yielding substantial bias in the mean response estimates (Czado and 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 and 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 and 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 skew-normal distributions (Arellano-Valle and 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 and Genton, 2010) for an elliptical prior. The closed-form and tractable posterior for the regression coefficients facilitates inference by using an algorithm which 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 organised as follows. Section 2 describes the preliminaries about the skew-elliptical and unified skew-elliptical 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, i.e., 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 Dey (2001) has led to many attractive properties of this class of distribution, such as 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 parametrization 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 $\int_0^\infty r^{(d+1)/2-1} g^{(d+1)}(r)\mathrm{d}r = \Gamma\{(d+1)/2\}\pi^{-(d+1)/2}$. Then a d-dimensional random vector $\boldsymbol{X}$ has a skew-elliptical distribution with location parameter vector $\boldsymbol{\xi} \in \mathbb{R}^d$, positive-definite scale matrix $\Sigma \in \mathbb{R}^{d\times d}$, skewness parameter vector $\boldsymbol{\alpha} \in \mathbb{R}^d$, and density generator $g^{(d+1)}$, if its density function is

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

where $\sigma = \text{diag}(\Sigma)^{1/2} \in \mathbb{R}^{d\times d}$, $q(\boldsymbol{x}) = (\boldsymbol{x}-\boldsymbol{\xi})^\top\Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\xi})$, $g^{(d)}$ is the d-dimensional marginal density generator induced by $g^{(d+1)}$, and $G(\cdot; g_{q(\boldsymbol{x})})$ is the cumulative distribution function of the univariate elliptical distribution with mean 0, scale 1, and conditional density generator $g_{q(\boldsymbol{x})}(s) = g^{(d+1)}\{s+q(\boldsymbol{x})\}/g^{(d)}\{q(\boldsymbol{x})\}$. We write $\boldsymbol{X} \sim \mathcal{SE}_d(\boldsymbol{\xi}, \Sigma, \boldsymbol{\alpha}, g^{(d+1)})$. When $\boldsymbol{\alpha} = \mathbf{0}$, the skew-elliptical distribution reduces to an elliptical distribution.

The skew-elliptical distribution has two stochastic representations, i.e., 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

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

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

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

where $\sigma = \text{diag}(\Sigma)^{1/2} \in \mathbb{R}^{d\times d}$, $\boldsymbol{\delta} = (1 + \boldsymbol{\alpha}^\top\bar{\Sigma}\boldsymbol{\alpha})^{-1/2}\bar{\Sigma}\boldsymbol{\alpha}$ with $\bar{\Sigma}$ being the correlation matrix corresponding to Σ, i.e., $\Sigma = \sigma\bar{\Sigma}\sigma$, 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 $\bar{\Sigma} - \boldsymbol{\delta}\boldsymbol{\delta}^\top$ is positive definite.

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

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

where $\phi_d(\boldsymbol{x} - \boldsymbol{\xi}; \Sigma)$ 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 $\boldsymbol{X} \sim \mathcal{SN}_d(\boldsymbol{\xi}, \Sigma, \boldsymbol{\alpha})$. 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-variate Student's t density generator with ν degrees of freedom, we get another important subclass of the skew-elliptical distribution, i.e., the skew-t distribution (Branco and Dey, 2001; Azzalini and Capitanio, 2003). Its density has the following form

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

where $t_d(\boldsymbol{x} - \boldsymbol{\xi}; \Sigma, \nu)$ denotes the probability density function of the d-variate t distribution with location vector $\boldsymbol{\xi}$, scale matrix Σ, and degrees of freedom ν, $T(\cdot; \nu + d)$ denotes the univariate t distribution function with degrees of freedom $\nu + d$. We write $\boldsymbol{X} \sim \mathcal{ST}_d(\boldsymbol{\xi}, \Sigma, \boldsymbol{\alpha}, \nu)$. When $\boldsymbol{\alpha} = \mathbf{0}$, it reduces to the d-dimensional Student's t distribution, and when $\nu \to \infty$, 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 and Genton, 2010), which aims to gain more flexibility by unifying various skew-elliptical families under the same model. Specifically, a d-dimensional random vector $\boldsymbol{X}$ has a unified skew-elliptical distribution, denoted by $\boldsymbol{X} \sim \mathcal{SUE}_{d,m}(\boldsymbol{\xi}, \Sigma, \Lambda, \boldsymbol{\tau}, \Gamma, g^{(d+m)})$, if its density function is

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

where $q(\boldsymbol{x}) = (\boldsymbol{x} - \boldsymbol{\xi})^\top \Sigma^{-1}(\boldsymbol{x} - \boldsymbol{\xi})$, $g^{(d+m)}$ is a $(d+m)$-variate elliptical density generator, $g^{(d)}$ and $g^{(m)}$ are its d-variate and m-variate marginal density generators, respectively, $g^{(m)}_{q(\boldsymbol{x})}(s) = g^{(d+m)}\{s + q(\boldsymbol{x})\}/g^{(d)}\{q(\boldsymbol{x})\}$, $\boldsymbol{\xi} \in \mathbb{R}^d$ is a location parameter vector, $\boldsymbol{\tau} \in \mathbb{R}^d$ 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 skew-elliptical distribution (Fang, 2003), and if we further have $\boldsymbol{\tau} = \mathbf{0}$, it reduces to the skew-elliptical distribution (1).

Similar to the skew-elliptical distribution, the unified skew-elliptical distribution also has two special subclasses, i.e., the unified skew-normal distribution (Arellano-Valle and 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

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

where $\Phi_m(\cdot;\Gamma)$ denotes the centered m-dimensional normal distribution function with covariance matrix Γ. We write $\boldsymbol{X} \sim \mathcal{SUN}_{d,m}(\boldsymbol{\xi},\Sigma,\Lambda,\boldsymbol{\tau},\Gamma)$. The definition (3) is equivalent to the one in Arellano-Valle and Azzalini (2006) with a slightly different parametrization, but is consistent with Arellano-Valle and Genton (2010). When $g^{(d+m)}$ is the $(d+m)$-variate Student's t density generator with ν degrees of freedom, we get the unified skew-t distribution with density

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

where $T_m(\cdot;\Gamma,\nu+d)$ denotes the centered m-dimensional Student's t distribution function with dispersion matrix Γ and degrees of freedom $\nu+d$. We write $\boldsymbol{X} \sim \mathcal{SUT}_{d,m}(\boldsymbol{\xi},\Sigma,\Lambda,\nu,\boldsymbol{\tau},\Gamma)$.

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 skew-normal and skew-t links as special cases. Specifically, let Y_{ij} denote a binary 0/1 response on the ith observation of the jth variable and denote by $\boldsymbol{Y}_i = (Y_{i1},\ldots,Y_{iM})^\top$ the collection of the ith observation on all M variables for $i=1,\ldots,n$. Let $\boldsymbol{Y}_i^* = (Y_{i1}^*,\ldots,Y_{iM}^*)^\top$

be a vector of latent variables capturing dependence among the components of $\boldsymbol{Y}_i$, $\boldsymbol{\beta} \in \mathbb{R}^p$ be a vector of regression coefficients, $X_i = (\boldsymbol{x}_{i1}, \ldots, \boldsymbol{x}_{iM})^\top \in \mathbb{R}^{M \times p}$ be the data matrix for the ith observation, and let $X = (X_1^\top, \ldots, X_n^\top)^\top \in \mathbb{R}^{nM \times p}$. Then the multivariate skew-elliptical link model can be expressed as

$$
\begin{aligned}
Y_{ij} &= \begin{cases} 1, & \text{if } Y_{ij}^* > 0, \\ 0, & \text{otherwise,} \end{cases} \\
\boldsymbol{Y}^* &= (\boldsymbol{Y}_1^{*\top}, \ldots, \boldsymbol{Y}_n^{*\top})^\top = X\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \qquad (4)
\end{aligned}
$$

$$
\begin{pmatrix} \boldsymbol{\beta} \\ \boldsymbol{\varepsilon} \end{pmatrix} \Bigg| \Sigma, \boldsymbol{\alpha}, g^{(p+nM+1)} \sim \mathcal{SE}_{p+nM}\left(\begin{pmatrix} \boldsymbol{\mu} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \Omega & 0 \\ 0 & \mathrm{I}_n \otimes \Sigma \end{pmatrix}, \begin{pmatrix} \mathbf{0} \\ \boldsymbol{\alpha} \end{pmatrix}, g^{(p+nM+1)} \right),
$$

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, $\mathrm{I}_n \in \mathbb{R}^{n \times n}$ 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 $\boldsymbol{\beta}$ and $\boldsymbol{\varepsilon}$ in the model (4), we express it in a different way. Using Proposition 2 in Fang (2003), an equivalent formulation is

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

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

where $q(\boldsymbol{\beta}) = (\boldsymbol{\beta} - \boldsymbol{\mu})^\top \Omega^{-1} (\boldsymbol{\beta} - \boldsymbol{\mu})$, $g_{q(\boldsymbol{\beta})}^{(nM+1)}(s) = g^{(p+nM+1)}\{s + q(\boldsymbol{\beta})\}/g^{(p)}\{q(\boldsymbol{\beta})\}$, and $g^{(p)}$, $g^{(p+1)}$ are the p- and $(p+1)$-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 $\boldsymbol{\beta}$, while (6) is the distributional assumption for the latent data vector $\boldsymbol{Y}^*$. From (6) we observe that $\boldsymbol{\beta}$ and $\boldsymbol{\varepsilon}$ are dependent, but they are conditionally independent given $q(\boldsymbol{\beta})$. 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 $\boldsymbol{\varepsilon} \mid \Sigma, \boldsymbol{\alpha} \sim \mathcal{SN}_{nM}(\mathbf{0}, \mathrm{I}_n \otimes \Sigma, \boldsymbol{\alpha})$, which is independent of $\boldsymbol{\beta}$ conditional on Σ and $\boldsymbol{\alpha}$. If we further have $\boldsymbol{\alpha} = \mathbf{0}$, then (6) becomes $\boldsymbol{\varepsilon} \mid \Sigma \sim \mathcal{N}_{nM}(\mathbf{0}, \mathrm{I}_n \otimes \Sigma)$ and model (4) reduces to the well-known multivariate probit model (Ashford and Sowden, 1970; Chib and Greenberg, 1998) with a typical Gaussian prior for $\boldsymbol{\beta}$. By assuming a joint distribution for $\boldsymbol{\beta}$ and $\boldsymbol{\varepsilon}$, we can gain two major advantages. The first is that we are able to account not only for the dependence between $\boldsymbol{\beta}$ and $\boldsymbol{\varepsilon}$, but also for the dependence between the different observations $\boldsymbol{Y}_i, i = 1, \ldots, n$. The second is that this assumption allows us to get a tractable posterior for $\boldsymbol{\beta}$; see Section 3.2 for more details.

The multivariate probit model (Chib and Greenberg, 1998) assumes that the covariates are not shared by the M variables $Y_{i1}, \ldots, Y_{iM}$. In that case, $\boldsymbol{\beta}$ can be understood as $\boldsymbol{\beta} = (\boldsymbol{\beta}_1^\top, \ldots, \boldsymbol{\beta}_M^\top)^\top$, 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 variable $Y_{1j}, \ldots, Y_{nj}$, and $\boldsymbol{x}_{ij}$ is understood as the vector $\boldsymbol{x}_{ij} = (\boldsymbol{x}_{ij1}^\top, \ldots, \boldsymbol{x}_{ijM}^\top)^\top$ 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 $(\Sigma, \boldsymbol{\alpha})$ are those such that the matrix $\mathrm{I}_n \otimes \bar{\Sigma} - \boldsymbol{\delta}\boldsymbol{\delta}^\top$ is positive definite, where $\bar{\Sigma}$ is the correlation matrix corresponding to Σ and $\boldsymbol{\delta} = \{1 + \boldsymbol{\alpha}^\top (\mathrm{I}_n \otimes \bar{\Sigma}) \boldsymbol{\alpha}\}^{-1/2} (\mathrm{I}_n \otimes \bar{\Sigma}) \boldsymbol{\alpha}$. From (4) we obtain the joint probability mass function of $\boldsymbol{Y} = (\boldsymbol{Y}_1^\top, \ldots, \boldsymbol{Y}_n^\top)^\top = \boldsymbol{y}$, given all the parameters and the data matrix X, as

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

where $q(\boldsymbol{t}) = \boldsymbol{t}^\top (\mathrm{I}_n \otimes \Sigma^{-1}) \boldsymbol{t}$, $g_{q(\boldsymbol{t})}^{q(\boldsymbol{\beta})}(s) = g_{q(\boldsymbol{\beta})}^{(nM+1)}\{s + q(\boldsymbol{t})\} / g^{q(\boldsymbol{\beta}), nM}\{q(\boldsymbol{t})\}$, $g^{q(\boldsymbol{\beta}), nM}$ is the nM-variate marginal density generator induced by $g_{q(\boldsymbol{\beta})}^{(nM+1)}$, and $A_{ij}, i = 1, \ldots, n, j = 1, \ldots, M$ is the interval $(-\boldsymbol{x}_{ij}^\top \boldsymbol{\beta}, \infty)$ if $y_{ij} = 1$, and $(-\infty, \boldsymbol{x}_{ij}^\top \boldsymbol{\beta}]$ 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 skew-elliptical 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 $p(\boldsymbol{y} \mid \boldsymbol{\beta}, \Sigma, \boldsymbol{\alpha}, g^{(p+nM+1)})$ of the observed data in the following lemma. All proofs are deferred to the Appendix.

LEMMA 1: *The joint probability mass function $p(\boldsymbol{y} \mid \boldsymbol{\beta}, \Sigma, \boldsymbol{\alpha}, g^{(p+nM+1)})$ based on (4) can be simplified to*

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

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

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

with $\boldsymbol{\delta} \in \mathbb{R}^{nM}, \boldsymbol{\delta} = \{1 + \boldsymbol{\alpha}^\top(\mathrm{I}_n \otimes \bar{\Sigma})\boldsymbol{\alpha}\}^{-1/2}(\mathrm{I}_n \otimes \bar{\Sigma})\boldsymbol{\alpha}$, $\sigma = \operatorname{diag}(\Sigma)^{1/2} \in \mathbb{R}^{d\times d}$ and $\bar{\Sigma}$ being the correlation matrix corresponding to Σ, i.e., $\Sigma = \sigma\bar{\Sigma}\sigma$.

Now we are ready to present our main result that the posterior distribution of $\boldsymbol{\beta}$ coincides with a unified skew-elliptical distribution.

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

$$(\boldsymbol{\beta} \mid \boldsymbol{y}, \Sigma, \boldsymbol{\alpha}, g^{(p+nM+1)}) \sim \mathcal{SUE}_{p,nM+1}(\boldsymbol{\mu}_{post}, \Omega_{post}, \Lambda_{post}, \boldsymbol{\tau}_{post}, \Gamma_{post}, g^{(p+nM+1)}),$$

with posterior parameters

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

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, $\sigma_* = \operatorname{diag}(\Sigma_*)^{1/2} \in \mathbb{R}^{(nM+1)\times(nM+1)}$, $\bar{\Sigma}_*$ is the correlation matrix corresponding to Σ_*, i.e., $\Sigma_* = \sigma_*\bar{\Sigma}_*\sigma_*$, 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, i.e. the skew-t link model obtained when the density generator $g^{(p+nM+1)}$ in model (4) is the $(p+nM+1)$-variate Student's 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 is provided in the Supporting Information. Specifically, for the skew-t link model, the joint distributional assumption of $\boldsymbol{\beta}$ and $\boldsymbol{\varepsilon}$ is

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

which is equivalent to assuming

$$\boldsymbol{\beta} \mid \nu \sim \mathcal{T}_p(\boldsymbol{\mu}, \Omega, \nu),$$

$$\boldsymbol{\varepsilon} \mid \boldsymbol{\beta}, \Sigma, \boldsymbol{\alpha}, \nu \sim \mathcal{ST}_{nM}\left(\mathbf{0}, \frac{\nu + (\boldsymbol{\beta}-\boldsymbol{\mu})^\top \Omega^{-1} (\boldsymbol{\beta}-\boldsymbol{\mu})}{\nu + p} \left(\mathrm{I}_n \otimes \Sigma \right), \boldsymbol{\alpha}, \nu + p \right),$$

where $\mathcal{T}_p(\boldsymbol{\mu}, \Omega, \nu)$ denotes the Student's 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 $\boldsymbol{\beta}$ and $\boldsymbol{\varepsilon}$. As ν increases the dependence decreases, and when $\nu \to \infty$, 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 Student's t density generator with ν degrees of freedom, we get the following explicit expression of the joint probability

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

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 $p(\boldsymbol{y} \mid \boldsymbol{\beta}, \Sigma, \boldsymbol{\alpha}, \nu) \approx p(\boldsymbol{y} \mid b\boldsymbol{\beta}, b^2\Sigma, \boldsymbol{\alpha}, \nu)$ 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 Student's t density generator with ν degrees of freedom.

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

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

where $\boldsymbol{\mu}_{post}, \Omega_{post}, \Lambda_{post}, \boldsymbol{\tau}_{post}, \Gamma_{post}$ 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 and 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 $(\boldsymbol{\beta} \mid \boldsymbol{y}, \Sigma, \boldsymbol{\alpha}, \nu)$ is also feasible. Specifically, using Equation (9) in Arellano-Valle and Genton

(2010), $(\boldsymbol{\beta} \mid \boldsymbol{y}, \Sigma, \boldsymbol{\alpha}, \nu)$ has the stochastic representation

$$(\boldsymbol{\beta} \mid \boldsymbol{y}, \Sigma, \boldsymbol{\alpha}, \nu) \stackrel{\mathrm{d}}{=} \boldsymbol{\mu} + \left(\frac{\nu + \boldsymbol{U}_1^\top s (D_* \Omega D_*^\top + \Sigma_*)^{-1} s \boldsymbol{U}_1}{\nu + nM + 1} \right)^{1/2} \boldsymbol{U}_0 + \Omega D_*^\top (D_* \Omega D_*^\top + \Sigma_*)^{-1} s \boldsymbol{U}_1, \tag{8}$$

where $s = \mathrm{diag}(D_* \Omega D_*^\top + \Sigma_*)^{1/2} \in \mathbb{R}^{(nM+1) \times (nM+1)}$, $\boldsymbol{U}_0 \sim \mathcal{T}_p\big(\mathbf{0}, \Omega - \Omega D_*^\top (D_* \Omega D_*^\top + \Sigma_*)^{-1} D_* \Omega, \nu + nM + 1\big)$ is independent of $\boldsymbol{U}_1$, which follows a $(nM+1)$-variate truncated t distribution with location parameter vector $\mathbf{0}$, dispersion matrix $s^{-1}(D_* \Omega D_*^\top + \Sigma_*) s^{-1}$, degrees of freedom ν, and truncated below the level $-s^{-1} D_* \boldsymbol{\mu}$.

3.4 *Identifiability of* α *and* ν

In this subsection we investigate the identifiability of the skewness parameter $\boldsymbol{\alpha}$ and the degree of freedom parameter ν in the skew-t link model. Lemma 1 implies that both $\boldsymbol{\alpha}$ and ν play a role in determining the joint probability mass function $p(\boldsymbol{y} \mid \boldsymbol{\beta}, \Sigma, \boldsymbol{\alpha}, g^{(p+nM+1)})$, and hence they should be identifiable theoretically. Here we conduct a simulation study to verify that. We fix $M = 1$, $\Sigma = 1$, $\nu = 1$, $\boldsymbol{\beta} = (-1, -0.5, 0.5)^T$, $\boldsymbol{\mu} = \mathbf{0}$ and $\Omega = 25 \times \mathrm{I}_3$, and simulate $n = 100$ data from the skew-t link model with skewness $\boldsymbol{\alpha} = \alpha_c \times \mathbf{1}_n$, $\alpha_c = -2$ or 2. The first column of the data matrix X is set to $\mathbf{1}$ and the other two columns are generated from a standard normal distribution. Then we compute the log likelihood $\log p(\boldsymbol{y} \mid \boldsymbol{\beta}, \Sigma, \boldsymbol{\alpha}, g^{(p+nM+1)})$ with respect to different values of α_c. Specifically, all the other parameters except $\boldsymbol{\alpha}$ are treated as nuisance parameters and we use their true values to compute $\log p(\boldsymbol{y} \mid \boldsymbol{\beta}, \Sigma, \boldsymbol{\alpha}, g^{(p+nM+1)})$. 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 about here.]

Figure 1 displays the log likelihood with respect to varying α_c and ν. It shows that the skewness parameter $\boldsymbol{\alpha}$ is weakly identifiable, which might be due to the fact that $\boldsymbol{\alpha}$ appears only through the matrix Σ_*, and if α_c is large, then the vector $\boldsymbol{\delta}$ in Lemma 1 would be approximately $(\mathrm{I}_n \otimes \bar{\Sigma})\mathbf{1}_n \mathrm{sign}(\alpha_c)$ 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, whilst 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* $\boldsymbol{\alpha}$ *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, i.e., $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_M, \ldots, \alpha_1, \ldots, \alpha_M)^\top \in \mathbb{R}^{nM}$, and Σ is a correlation matrix, i.e., $\Sigma = \bar{\Sigma}$. 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 $\boldsymbol{\alpha}_s = (\alpha_1, \ldots, \alpha_M)^\top$ and the correlation matrix $\bar{\Sigma}$.

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 LKJ prior (Lewandowski et al., 2009).

The marginally uniform prior means that each non-diagonal 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 $\bar{\Sigma}$ by setting $\eta = 1$ in the LKJ prior and specify an independent weakly informative Gaussian prior for $\boldsymbol{\alpha}_s$. Then, using Equation (7), the joint posterior of $(\bar{\Sigma}, \boldsymbol{\alpha}_s)$ given the data and the regression coefficients is

$$p(\bar{\Sigma}, \boldsymbol{\alpha}_s \mid \boldsymbol{y}, \boldsymbol{\beta}, \nu) \propto 2T_{nM+1}\left[\left\{\frac{\nu + p}{\nu + (\boldsymbol{\beta} - \boldsymbol{\mu})^\top \Omega^{-1}(\boldsymbol{\beta} - \boldsymbol{\mu})}\right\}^{1/2} D_*\boldsymbol{\beta}; \Sigma_*, \nu + p\right]\pi(\boldsymbol{\alpha}_s). \quad (9)$$

We evaluate the multivariate Student's t probability on the right-hand side of (9) using the R library `tlrmvnmvt` (Cao et al., 2022), which implements the classic Genz algorithm (Genz and 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 reparametrization adopted in Smith (2013), which re-expresses 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 $L_{\bar{\Sigma}} = (l_{ij})$ by $\boldsymbol{\theta}$, i.e., $\boldsymbol{\theta} = \{l_{ij} : i > j, i, j = 1, \ldots, M\}$, and the $M(M-1)/2$ constrained parameters in $\bar{\Sigma}$ by $\text{vec}(\bar{\Sigma})$. Then, using a change of variables we get the posterior of $(\boldsymbol{\theta}, \boldsymbol{\alpha}_s)$ as $p(\boldsymbol{\theta}, \boldsymbol{\alpha}_s \mid \boldsymbol{y}, \boldsymbol{\beta}, \nu) = p(\bar{\Sigma}, \boldsymbol{\alpha}_s \mid \boldsymbol{y}, \boldsymbol{\beta}, \nu)|J| = p(\bar{\Sigma}, \boldsymbol{\alpha}_s \mid \boldsymbol{y}, \boldsymbol{\beta}, \nu) \prod_{i=1}^{M} \left(1 + \sum_{j<i} l_{ij}^2\right)^{-(M+1)/2}$, where $|J| = |\partial \text{vec}(\bar{\Sigma}) / \partial \boldsymbol{\theta}|$ is the determinant of the Jacobian matrix of this transformation.

As direct sampling from the distribution of $\boldsymbol{\theta}, \boldsymbol{\alpha}_s \mid \boldsymbol{y}, \boldsymbol{\beta}, \nu$ is unknown, we propose to use a random walk Metropolis-Hastings algorithm to generate samples from it. Specifically, we first sample $\boldsymbol{\alpha}_s'$ from a proposal distribution with density $q(\cdot \mid \boldsymbol{\alpha}_s)$ and $\boldsymbol{\theta}'$ from a proposal distribution with density $r(\cdot \mid \boldsymbol{\theta})$. Here we take both proposal densities q and r as symmetric normal densities, i.e., $\boldsymbol{\alpha}_s' \mid \boldsymbol{\alpha}_s \sim \mathcal{N}_M(\boldsymbol{\alpha}_s, h_1 \mathrm{I}_M)$ and $\boldsymbol{\theta}' \mid \boldsymbol{\theta} \sim \mathcal{N}_K(\boldsymbol{\theta}, h_2 \mathrm{I}_K), K = M(M-1)/2$. Then the acceptance probability is $\alpha((\boldsymbol{\alpha}_s, \boldsymbol{\theta}), (\boldsymbol{\alpha}_s', \boldsymbol{\theta}')) = \min\left\{\frac{p(\boldsymbol{\theta}', \boldsymbol{\alpha}_s' \mid \boldsymbol{y}, X, \boldsymbol{\beta}) 1((\boldsymbol{\theta}', \boldsymbol{\alpha}_s') \in C)}{p(\boldsymbol{\theta}, \boldsymbol{\alpha}_s \mid \boldsymbol{y}, X, \boldsymbol{\beta}) 1((\boldsymbol{\theta}, \boldsymbol{\alpha}_s) \in C)}, 1\right\}$, where $1(\cdot)$ is the indicator function and C is the space of all $(\boldsymbol{\theta}, \boldsymbol{\alpha}_s)$ such that the resulting matrix $\bar{\Sigma} - \boldsymbol{\delta}\boldsymbol{\delta}^\top$ is positive definite with $\boldsymbol{\delta} = (1 + \boldsymbol{\alpha}^\top \bar{\Sigma} \boldsymbol{\alpha})^{-1/2} \bar{\Sigma} \boldsymbol{\alpha}$.

4.2 *MCMC Sampling Scheme*

As sampling from the distribution of $(\boldsymbol{\beta} \mid \boldsymbol{y}, X, \bar{\Sigma}, \boldsymbol{\alpha})$ is feasible using (8) and sampling from the distribution of $(\bar{\Sigma}, \boldsymbol{\alpha} \mid \boldsymbol{y}, X, \boldsymbol{\beta})$ has been described in Section 4.1, we now combine them to construct an MCMC sampler for the multivariate skew-t link model.

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 $\mathcal{T}_p\big(\mathbf{0}, \Omega - \Omega D_*^\top (D_* \Omega D_*^\top + \Sigma_*)^{-1} D_* \Omega, \nu + nM + 1\big)$ (in R use *rmvt*);

[2] Sample $\boldsymbol{U}_1^{(k)}$ from a $(nM+1)$-variate truncated t distribution with location parameter vector $\mathbf{0}$, 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 $(\boldsymbol{\theta}^{(k)}, \boldsymbol{\alpha}_s^{(k)})$ 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: $(\boldsymbol{\beta}^{(1)}, \bar{\Sigma}^{(1)}, \boldsymbol{\alpha}^{(1)}), \ldots, (\boldsymbol{\beta}^{(K)}, \bar{\Sigma}^{(K)}, \boldsymbol{\alpha}^{(K)})$

4.3 *Application to COVID-19 Pandemic Data*

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, i.e., 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, i.e., 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 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.

[Figure 2 about here.]

We consider three covariates in total, i.e., 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 mean 0 and standard deviation 1. To assess the performance of the multivariate skew-elliptical link model, we consider six models $\mathcal{M}_1, \mathcal{M}_2, \mathcal{M}_3, \mathcal{M}_4, \mathcal{M}_5, \mathcal{M}_6$ of different complexity. $\mathcal{M}_1, \mathcal{M}_2, \mathcal{M}_3$ 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 $\boldsymbol{\alpha} = \mathbf{0}$), and $\mathcal{M}_6$ is the independent probit model (obtained with $\nu \to \infty$, $\bar{\Sigma} = \mathrm{I}_3, \boldsymbol{\alpha} = \mathbf{0}$).

For each of these models, we run the 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 $\mathcal{T}_3(\mathbf{0}, 25\mathrm{I}_3, \nu)$ for the model $\mathcal{M}_1, \mathcal{M}_2, \mathcal{M}_3$, and $\mathcal{N}_3(\mathbf{0}, 25\mathrm{I}_3)$ for the model $\mathcal{M}_4, \mathcal{M}_5, \mathcal{M}_6$. The prior for the skewness parameters is taken as $\mathcal{N}_3(\mathbf{0}, 16\mathrm{I}_3)$. The variances of the proposal densities in the Metropolis-Hastings algorithm are taken the same as in the simulation study, i.e. $h_1 = 9, h_2 = 0.36$.

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 correlation between the counties of Orange and San Diego, i.e., $\bar{\Sigma}_{23}$, seems to be quite strong, as its posterior mean for models $\mathcal{M}_1, \mathcal{M}_2, \mathcal{M}_3, \mathcal{M}_4, \mathcal{M}_5$ 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 into 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.

[Table 1 about here.]

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 (AIC) and is defined as DIC $= D(\bar{\boldsymbol{\kappa}}) + 2p_D$, where $\boldsymbol{\kappa}$ denotes the collection of all the parameters, $\bar{\boldsymbol{\kappa}} = \mathrm{E}[\boldsymbol{\kappa} \mid \boldsymbol{y}]$ is its posterior mean, $D(\cdot)$ is a deviance function and $p_D = \mathrm{E}[D(\boldsymbol{\kappa}) \mid \boldsymbol{y}] - D(\bar{\boldsymbol{\kappa}})$ is the effective number of model parameters. Here we take the deviance function $D(\boldsymbol{\kappa})$ as $-2\log p(\boldsymbol{y} \mid \boldsymbol{\beta}, \bar{\Sigma}, \boldsymbol{\alpha}, \nu)$ when the model is the skew-t link model, or $-2\log p(\boldsymbol{y} \mid \boldsymbol{\beta}, \bar{\Sigma}, \boldsymbol{\alpha})$ when the model is the skew-normal link model, and estimate $\mathrm{E}[D(\boldsymbol{\kappa}) \mid \boldsymbol{y}]$ 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 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 $\boldsymbol{\varepsilon} \mid \boldsymbol{\beta}, \Sigma, \boldsymbol{\alpha}$ 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 $\boldsymbol{\beta}$ coincides with a fundamental skew distribution (Arellano-Valle and

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 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 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).

References

Arellano-Valle, R. B. and Azzalini, A. (2006). On the unification of families of skew-normal distributions. *Scandinavian Journal of Statistics* **33,** 561–574.

Arellano-Valle, R. B. and Genton, M. G. (2005). On fundamental skew distributions. *Journal of Multivariate Analysis* **96,** 93–116.

Arellano-Valle, R. B. and Genton, M. G. (2010). Multivariate unified skew-elliptical distributions. *Chilean Journal of Statistics* **1,** 17–33.

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

Azzalini, A. and Capitanio, A. (1999). Statistical applications of the multivariate skew normal distribution. *Journal of the Royal Statistics Society (Series B)* **61,** 579–602.

Azzalini, A. and 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,** 367–389.

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

Barnard, J., McCullloch, R., and Meng, X. (2000). Modeling covariance matries in terms of standard deviations and correlations, with application to shrinkage. *Statistica Sinica* **10,** 1281–1311.

Bazán, J. L., Bolfarine, H., and 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. and Dey, D. K. (2001). A general class of multivariate skew-elliptical distributions. *Journal of Multivariate Analysis* **79,** 99–113.

California Open Data Portal (2022). https://data.ca.gov/ (accessed Apr. 5, 2022).

Cao, J., Genton, M. G., Keyes, D. E., and Turkiyyah, G. M. (2021). Exploiting low rank covariance structures for computing high-dimensional normal and Student-t probabilities. *Statistics and Computing* **31,** 2.

Cao, J., Genton, M. G., Keyes, D. E., and Turkiyyah, G. M. (2022). tlrmvnmvt: Computing high-dimensional multivariate normal and Student-t probabilities with low-rank methods in R. *Journal of Stiatistical Software* **101,** 4.

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

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

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.

Durante, D. (2019). Conjugate Bayes for probit regression via unified skew-normal distributions. *Biometrika* **106,** 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., and Rubin, D. B. (2014). *Bayesian Data Analysis*. CRC Press, third edition.

Gelman, A., Jakulin, A., Pittau, M. G., and Su, Y.-S. (2008). A weakly informative default prior distribution for logistic and other regression models. *The Annals of Applied Statistics* **2,** 1360–1383.

Genz, A. and 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., and Dunson, D. B. (2019). MCMC for imbalanced categorical data. *Journal of the American Statistical Association* **114,** 1394–1403.

Kim, H. (2002). Binary regression with a class of skewed t link models. *Communications in Statistics - Theory and Methods* **31,** 1863–1886.

Kim, S., Chen, M.-H., and Dey, D. K. (2008). Flexible generalized t-link models for binary response data. *Biometrika* **95,** 93–106.

Lewandowski, D., Kurowicka, D., and 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 Damien, P., Dellaportas, P., Polson, N. G., and Stephens, D. A., editors, *Bayesian Theory and Applications*, chapter 17, pages 336–358. Oxford University Press.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and 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. and 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,** 2000–2023.

Zhao, X., Zhang, L., and Bandyopadhyay, D. (2021). A shared spatial model for multivariate extreme-valued binary data with non-random missingness. *Sankhya B* **83,** 374–396.

Supporting Information

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.

Appendix

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

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

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

$$(-D\boldsymbol{\varepsilon} - DX\boldsymbol{\beta}) \mid \boldsymbol{\beta}, \Sigma, \boldsymbol{\alpha}, g^{(p+nM+1)} \sim \mathcal{SE}_{nM}(-DX\boldsymbol{\beta}, D(\mathrm{I}_n \otimes \Sigma)D, D\boldsymbol{\alpha}, g_{q(\boldsymbol{\beta})}^{(nM+1)}).$$

Using (2), we finally get

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

Proof of Theorem 1. The posterior density of the coefficients $\boldsymbol{\beta}$ is

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

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

$$\begin{aligned}
& p(\boldsymbol{\beta} \mid \boldsymbol{y}, \Sigma, \boldsymbol{\alpha}, g^{(p+nM+1)}) \\
\propto\; & G_{nM+1}(D_*\boldsymbol{\beta}; \Sigma_*, g^{(nM+1)}_{q(\boldsymbol{\beta})}) \cdot g^{(p)}\{(\boldsymbol{\beta}-\boldsymbol{\mu})^\top \Omega^{-1}(\boldsymbol{\beta}-\boldsymbol{\mu})\} \\
=\; & G_{nM+1}(\sigma_*^{-1}D_*\boldsymbol{\beta}; \bar{\Sigma}_*, g^{(nM+1)}_{q(\boldsymbol{\beta})}) \cdot g^{(p)}\{(\boldsymbol{\beta}-\boldsymbol{\mu})^\top \Omega^{-1}(\boldsymbol{\beta}-\boldsymbol{\mu})\} \\
=\; & G_{nM+1}\{\sigma_*^{-1}D_*\boldsymbol{\mu} + \sigma_*^{-1}D_*(\boldsymbol{\beta}-\boldsymbol{\mu}); \bar{\Sigma}_*, g^{(nM+1)}_{q(\boldsymbol{\beta})}\} \cdot g^{(p)}\{(\boldsymbol{\beta}-\boldsymbol{\mu})^\top \Omega^{-1}(\boldsymbol{\beta}-\boldsymbol{\mu})\} \\
=\; & G_{nM+1}\{\sigma_*^{-1}D_*\boldsymbol{\mu} + \sigma_*^{-1}D_*\omega\omega^{-1}(\boldsymbol{\beta}-\boldsymbol{\mu}); \bar{\Sigma}_*, g^{(nM+1)}_{q(\boldsymbol{\beta})}\} \cdot g^{(p)}\{(\boldsymbol{\beta}-\boldsymbol{\mu})^\top \Omega^{-1}(\boldsymbol{\beta}-\boldsymbol{\mu})\} \\
=\; & G_{nM+1}\{\boldsymbol{\tau}_{\mathrm{post}} + \Lambda_{\mathrm{post}}\omega^{-1}(\boldsymbol{\beta}-\boldsymbol{\mu}); \bar{\Sigma}_*, g^{(nM+1)}_{q(\boldsymbol{\beta})}\} \cdot g^{(p)}\{(\boldsymbol{\beta}-\boldsymbol{\mu})^\top \Omega^{-1}(\boldsymbol{\beta}-\boldsymbol{\mu})\}.
\end{aligned}$$

Hence, $(\boldsymbol{\beta} \mid \boldsymbol{y}, \Sigma, \boldsymbol{\alpha}, g^{(p+nM+1)}) \sim \mathcal{SUE}_{p,nM+1}(\boldsymbol{\mu}_{post}, \Omega_{post}, \Lambda_{post}, \boldsymbol{\tau}_{post}, \Gamma_{post}, g^{(p+nM+1)})$. □

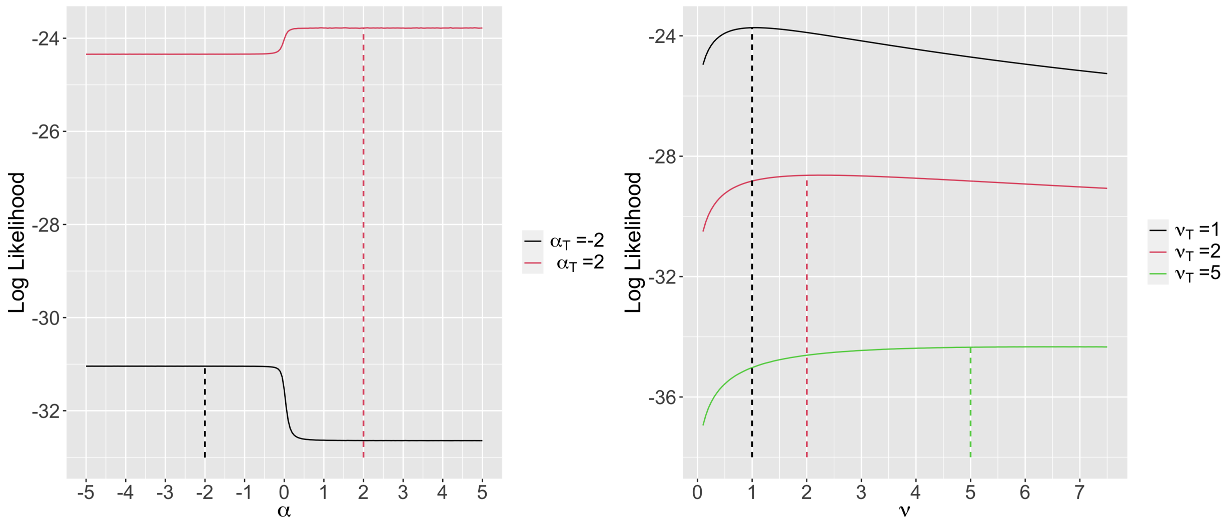

Figure 1. Log likelihood with respect to α_c and ν. Left panel: true value of α_c is $\alpha_T = -2$ (black line) and $\alpha_T = 2$ (red line); right panel: true value of ν is $\nu_T = 1$ (black line), $\nu_T = 2$ (red line), and $\nu_T = 5$ (green line). This figure appears in color in the electronic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

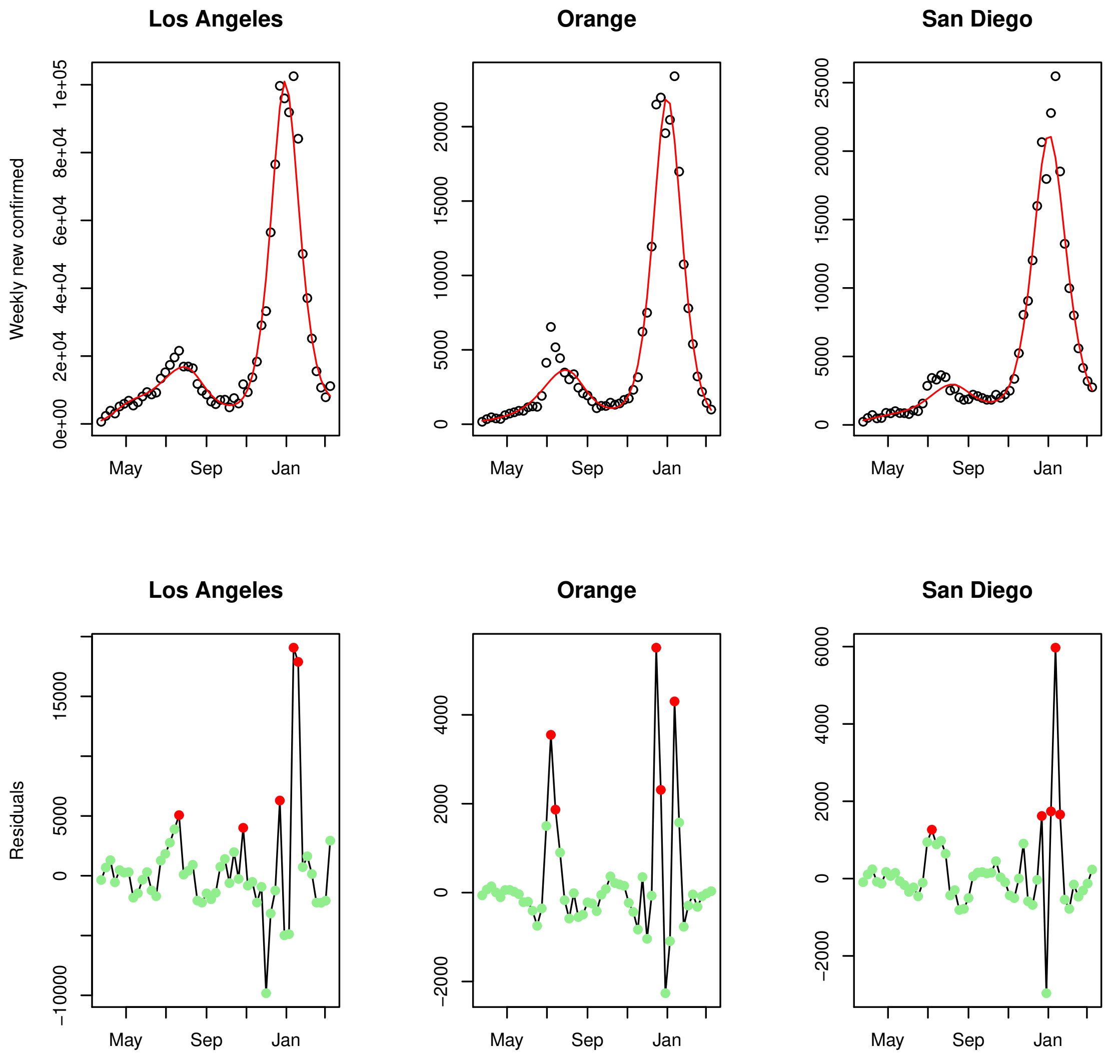

Figure 2. Upper panel: smoothing splines for the time series of weekly new confirmed cases at Los Angeles (left), Orange (middle), and San Diego (right). Lower panel: the residuals obtained as the difference between the original data and the fitted splines, with red points considered as extreme spikes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

Table 1

Posterior estimates, number of parameters, and estimated DIC for different models fitted in our COVID-19 data application in Section 4.3. The first column is the posterior mean, the second column is the 90% credible interval, and the third column is the empirical probability that the posterior samples have the same sign as the posterior mean

	$\mathcal{M}_1$ (# of par=9, DIC=86.2)			$\mathcal{M}_2$ (# of par=9, DIC=88.5)			$\mathcal{M}_3$ (# of par=9, 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)	90%	−1.43	(−3.76, 0.36)	89%
$\bar{\Sigma}_{12}$	0.40	(−0.03, 0.77)	93%	0.42	(−0.04, 0.80)	93%	0.41	(−0.05, 0.76)	91%
$\bar{\Sigma}_{13}$	0.63	(0.26, 0.89)	99%	0.67	(0.30, 0.93)	99%	0.62	(0.16, 0.89)	98%
$\bar{\Sigma}_{23}$	0.64	(0.28, 0.88)	100%	0.64	(0.24, 0.93)	100%	0.64	(0.33, 0.90)	100%
α_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%
	$\mathcal{M}_4$ (# of par=9, DIC=86.3)			$\mathcal{M}_5$ (# of par=9, DIC=86.7)			$\mathcal{M}_6$ (# of par=9, 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%
$\bar{\Sigma}_{12}$	0.42	(−0.05, 0.77)	94%	0.41	(−0.04, 0.78)	94%			
$\bar{\Sigma}_{13}$	0.63	(0.27, 0.89)	99%	0.63	(0.29, 0.89)	99%			
$\bar{\Sigma}_{23}$	0.64	(0.27, 0.89)	100%	0.66	(0.32, 0.90)	100%			
α_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%						