A spatial heterogeneity mixed model with skew-elliptical distributions

Mohadeseh Alsadat Farzammehr[1,a], Geoffrey J. McLachlan[b]

[a]Iranian Judiciary Research Institute, Iran;
[b]Department of Mathematics, University of Queensland, Australia

Abstract

The distribution of observations in most econometric studies with spatial heterogeneity is skewed. Usually, a single transformation of the data is used to approximate normality and to model the transformed data with a normal assumption. This assumption is however not always appropriate due to the fact that panel data often exhibit non-normal characteristics. In this work, the normality assumption is relaxed in spatial mixed models, allowing for spatial heterogeneity. An inference procedure based on Bayesian mixed modeling is carried out with a multivariate skew-elliptical distribution, which includes the skew-t, skew-normal, student-t, and normal distributions as special cases. The methodology is illustrated through a simulation study and according to the empirical literature, we fit our models to non-life insurance consumption observed between 1998 and 2002 across a spatial panel of 103 Italian provinces in order to determine its determinants. Analyzing the posterior distribution of some parameters and comparing various model comparison criteria indicate the proposed model to be superior to conventional ones.

Keywords: panel data, linear mixed model, spatial heterogeneity, multivariate skew-elliptical distributions, Bayesian hierarchical approach, MCMC

1. Introduction

In panel data, observations are compiled over several time periods on a cross section. In spatial econometrics, spatial panels are cross-sections of spatial units (such as regions, provinces, or countries), in which panel data analysis has received much attention.

The spatial heterogeneity structure can be used to analyze economic data that explicitly incorporates spatial information. Differences in time or space might cause spatial heterogeneity. Fundamentally, spatial heterogeneity is introduced by enabling the model to have heterogeneous spatial and temporal effects. Chesson (1985) described three types of variations in these models. The distinction is between pure spatial variation, which is constant from one time period to another, pure temporal variation, which is constant across time and space, and pure spatiotemporal variation, for which both spatial and temporal variations can occur simultaneously. The first two types consider spatial and temporal perspectives separately, so they omit the possibility of spatio-temporal variations when studying spatial heterogeneity.

A spatial heterogeneity involving spatial-temporal variations is examined in this paper. This is achieved through incorporating spatial sampling units (space-specific effects) into model structures. In such a setting, four types of space-specific effects as presented by Zheng *et al.* Zheng *et al.* (2008)

---

are considered. Models for describing space-specific effects include (i) Homogeneous model (*homogeneous*), (ii) Heterogeneous model (*heterogeneous*), (iii) Fixed-effects model (*fixed*), and (iv) Random-effects model (*random*). These models are explained in Section 3.1.

Most spatial panel data exhibit characteristics like skewness in practice. A common technique in empirical econometric models is to transform the response and explanatory variables in order to apply classical models that are based on the normality assumption efficiently. Even though using a transformation to handle departures from normality may result in reasonable empirical results, it is in many cases inappropriate and restrictive. Frequently, an alternative theoretical model (that addresses skewness directly) performs better than data transformation (Farzammehr *et al.*, 2019). Furthermore, data transformation comes with several limitations, such as reduced information and no guarantee of normality of the transformed data. Additionally, the transformed variables can be difficult to interpret, and the transformations may not be applicable to other datasets. In light of this, we propose new spatial heterogeneity panel models that offer greater flexibility in the distributional assumption of error terms to reduce the impact of the unrealistic normality assumption.

It is becoming increasingly popular as an analytical tool for data sets that exhibit non-normal characteristics to use multivariate skew-elliptical distributions. These distributions have additional features such as asymmetry and heavy tails that are not available in traditional normal and student-t distributions. The connections between the various multivariate skew-t distributions are analogues to those between the skew-normal distributions discussed in many studies (see, for example, the papers by Azzalini (2005), Arellano-Valle and Genton (2005), Arellano-Valle and Azzalini (2006), Lee and Mchlachlan (2013) and Azzalini (2014)).

Several extensions of standard linear mixed models have been proposed in the statistical literature that replace the normality assumption for the errors. Using space-specific effects, mixed models also provide a convenient framework for modeling spatial heterogeneity. To the best of our knowledge, hardly any work has been performed to discuss the simultaneous effects induced by spatial heterogeneity and non-normality, which are inherent to spatial panel data.

The spatial heterogeneity perspective should always be considered when dealing with panel analyses based on sets of spaces, since additional methodological issues arise with panels of spaces, where the relevant markets within spaces are well defined by the heterogeneity structure affecting panels studies. As such, insurance market is widely believed to be important for the economic growth of a country. Based on premiums per capita, we analyze non-life insurance consumption in Italy's 103 provinces from 1998 to 2002. In order to accomplish this, we adopt an intermediate approach to the empirical study of insurance consumption, one that bridges existing panel studies and analyses of datasets on household income, wealth, and consumption. Note that the availability of data for Italian insurance premiums limits the analysis to the provincial level. Overall, the goal of this paper is to estimate the parameters of symmetric and asymmetric models. To do so, we follow the Bayesian hierarchical framework described in Jara *et al.* (2008), Lunn *et al.* (2009), which uses Markov chain Monte Carlo (MCMC) algorithm for estimating spatial heterogeneity panel model with multivariate skew-elliptical distributions. This alternative modeling approach has the advantage that models are easily fitted in free software R under OpenBUGS (Lunn *et al.*, 2009), and that the computational effort is almost the same as for the standard model.

The rest of the paper is organized as follow. Section 2 presents the spatial heterogeneity panel model and Section 3 explains multivariate skew-elliptical distributions. Further, we discuss model selection criteria that may be used to evaluate the performance of our models. Section 4 is devoted to a simulation study examining the appropriateness of the proposed skew models. Bayesian analysis of non-life insurance data set is presented in Section 5. The main goal of the paper is to investigate the

determinants of the development of the non-life insurance market for Italy. Finally, Section 6 includes concluding remarks.

2. The model

In order to begin, we define a regression model that is expressed as ,

$$y_{it} = \boldsymbol{x}_{it}\boldsymbol{\beta} + \lambda_{it} + \epsilon_{it}, \tag{2.1}$$

for $i = 1, \ldots, N$ space units, and $t = 1, \ldots, T$ time periods. In the above, we let y_{it} denote the response variable at space i and time t, $\boldsymbol{x}_{it}$ is a known $1 \times K$ vector of explanatory variables, and $\boldsymbol{\beta}$ is a $K \times 1$ vector of fixed but unknown regression parameters. The variable ϵ_{it} denotes an independently and identically distributed (iid) error term for space i and time t with zero mean and variance σ^2. For λ_{it}, there are various assumptions depending on the application. One of them is $\lambda_{it} = \alpha_i + \eta_t$, which is typically used for models that include both time-dependent and space-dependent effects simultaneously. Other commonly adopted assumptions include $\lambda_{it} = \alpha_i$ for models that have space-specific effects only. If only time-specific effects are required and expressed, the assumption $\lambda_{it} = \eta_t$ can also be used. The present study examines models with space-specific effects.

Using (2.1), the matrix form of the simple model on panel data is as follows,

$$\boldsymbol{y}_t = X_t\boldsymbol{\beta} + \boldsymbol{u}_t, \tag{2.2}$$

where $\boldsymbol{y}_t = (y_{1t}, \ldots, y_{Nt})'$ and $X_t = (\boldsymbol{x}_{1t}, \ldots, \boldsymbol{x}_{Nt})'$, respectively, denote the vector of the response variables and the matrix of the explanatory variables in all spaces at time t. Thus, $\boldsymbol{u}_t = (u_{1t}, \ldots, u_{Nt})'$ is the model error component involving the sum of two disturbances. One of these disturbances is the $N \times 1$ vector of $\boldsymbol{\alpha}_t = (\alpha_1, \ldots, \alpha_N)'$ for which, with appropriate specifications, we can obtain the four types of spatial panel models described in Zhang *et al.* (2008) (refered to as *homogeneous*, *heterogeneous*, *fixed* and *random*). This allows the model to account for differences among different spaces which is refered to as spatial heterogeneity structure. The other disturbance in $\boldsymbol{u}_t$ is the vector of the remainder disturbances $\boldsymbol{\epsilon}_t = (\epsilon_{1t}, \ldots, \epsilon_{Nt})'$ in each period.

Let $\boldsymbol{y} = (\boldsymbol{y}_1', \ldots, \boldsymbol{y}_T')'$ and $X = (X_1', \ldots, X_T')'$ be a concatenated form of $\boldsymbol{y}_t$ and X_t, respectively, such that they denote the response and explanatory variables, respectively, where $\boldsymbol{y}$ denotes an $NT \times 1$ vector of the response variable and X is a $NT \times K$ matrix of the non-stochastic exogenous regressors. In a similar manner, the (concatenated form of the) disturbance vector can be written as

$$\boldsymbol{u} = \boldsymbol{\alpha} + \boldsymbol{\epsilon}, \tag{2.3}$$

where $\boldsymbol{\alpha} = (\boldsymbol{I}_N \otimes \boldsymbol{\iota}_T)\boldsymbol{\alpha}_t$, ($\boldsymbol{\iota}_p$ is a $p \times 1$ vector of ones, $\boldsymbol{I}_p$ is a $p \times p$ identity matrix and the symbol $\otimes$ denotes the Kronecker product) and $\boldsymbol{\epsilon} = (\boldsymbol{\epsilon}_1', \ldots, \boldsymbol{\epsilon}_T')'$ denotes the $NT \times 1$ idiosyncratic error vector.

As previously mentioned, the common assumption for spatial panel models is that $\boldsymbol{u}$ follows a normal distribution. When the responses are skewed, the normality assumption may be unrealistic, so it would be useful to examine skew-elliptical distributions as an alternative to the normal distribution to account for skewness in the data. Here, for the *homogeneous*, *heterogeneous*, *fixed* and *random* space-specific effects, we consider the skew-elliptical distributions for the remainder error term.

Bayesian inference in the aforementioned models requires defining the prior distribution for all unknown parameters (2.2), from which a posterior distribution could be derived for model parameters. One solution for selecting prior distribution is to consider the conditionally non-informative conjugate priors proposed by Gelman and Hill (2006).

To achieve well-defined and proper posteriors, we assign conjugate weakly-informative priors, as in Jara *et al* (2008) methodology. It is important to note that a family of prior distributions $p(\varphi)$ is conditionally conjugate for φ if the conditional posterior distribution $p(\varphi|y)$ is also in that class. Besides these parameters, we also need to specify the prior distribution and derive the posterior distribution for different specification of unobserved effects α_t.

It is necessary to decide, in methodological papers and also when applying spatial models to real data, whether the unobserved space-specific effect α_i is to be treated as a random or a fixed effect. In conventional usage, α_i is called a random effect when treated as a random variable, and a fixed effect when treated as a parameter to be estimated for each cross-sectional observation i. When α_i refers to an unobserved random effect, it is frequently assumed that they are not correlated with the x_{it}; that is, it is assumed that $E(\alpha_i|x_{i1}, \ldots, x_{iT}) = 0$.

3. Multivariate skew-elliptical distributions

In the following, we give the formal definition of the unrestricted multivariate skew-normal (uMSN) and unrestricted multivariate skew-*t* (uMST) distributions according to Lee and McLachlan (2013) formulations as special cases of the class of skew-elliptical distributions by Arellano-Valle and Genton (2005). This type of multivariate skew-normal distribution was studied in Sahu *et al.* (2003). Suppose z_0 and z_1 are jointly normally distributed as

$$\begin{bmatrix} z_0 \\ z_1 \end{bmatrix} \sim N_{2n}\left(\begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} I_n & \mathbf{0} \\ \mathbf{0} & \Sigma \end{bmatrix} \right). \tag{3.1}$$

Let z denote a n-dimensional random vector. Then $z = \mu + \Delta|z_0| + z_1$ defines the convolution-type stochastic representation of the unrestricted multivariate skew-normal distribution, and its density is given by

$$f(z; \mu, \Sigma, \Delta) = 2^n \phi_n(z; \mu, \Sigma) \times \Phi_n\left(\Delta'\Omega^{-1}(z, \mu); \mathbf{0}, \Lambda \right), \tag{3.2}$$

where $\Omega = \Sigma + \Delta\Delta'$ and $\Lambda = I_n - \Delta'\Omega^{-1}\Delta = (I_n + \Delta'\Sigma^{-1}\Delta)^{-1}$. In the above, μ is a $n \times 1$ vector of location parameters, Σ is a $n \times n$ diagonal positive definite scale matrix and Δ is a $n \times n$ diagonal matrix of skewness parameters. Further, $\phi_n(\cdot; \mu, \Sigma)$ and $\Phi_n(\cdot; \mu, \Sigma)$, respectively, denote the probability density function (pdf) and the cumulative distribution function (cdf) of a $N_n(\mu, \Sigma)$ random variable. Note that when $\mu = \mathbf{0}$ and $\Sigma = I_n$; $\phi_n(\cdot; \mu, \Sigma)$ and $\Phi_n(\cdot; \mu, \Sigma)$ will be denoted by $\phi_n(\cdot)$ and $\Phi_n(\cdot)$ respectively.

We shall follow the terminology of Lee and McLachlan (2013) and adopt the notation $z \sim \text{uSN}_n(\mu, \Sigma, \Delta)$ if z has the uMSN density. It is noted that if $\Delta = \mathbf{0}$, the second part of (3.2) evaluates to 2^{-n}, and we again recover the multivariate normal density $\phi_n(z; \mu, \Sigma)$. Then the density of z reduces to the usual symmetric multivariate density, $N_n(\mu, \Sigma)$.

Following Sahu *et al.* (2003), the uMSN distribution admits a convenient hierarchial representation. Let $w = |z_0|$, the hierarchical representation of (3.2) is given by

$$\begin{aligned} z|w &\sim N_n(\mu + \Delta w, \Sigma), \\ w &\sim HN_n(\mathbf{0}, I_N), \end{aligned} \tag{3.3}$$

where $HN_n(\mathbf{0}, \Sigma)$ represents the n-dimensional half-normal distribution with mean $\mathbf{0}$ and scale matrix Σ.

Analogous to the skew-normal case, the uMST distribution has a similar stochastic representation to the uMSN distribution. Specifically, $z = \mu + \Delta|z_0| + z_1$ has the uMST distribution, where conditional

on the gamma variable r,

$$\begin{bmatrix} z_0 \\ z_1 \end{bmatrix} \sim N_{2n}\left(\begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \frac{1}{r}\begin{bmatrix} \mathbf{I}_n & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma} \end{bmatrix}\right). \tag{3.4}$$

It follows that the density of z is given by

$$f(z; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\Delta}, \tau) = 2^n t_n(z; \boldsymbol{\mu}, \boldsymbol{\Omega}, \tau) T_n\left(c(y)\sqrt{\frac{\tau + n}{\nu + d(z)}}, \mathbf{0}, \boldsymbol{\Lambda}, \tau + n\right), \tag{3.5}$$

where $c(z) = \boldsymbol{\Delta}'\boldsymbol{\Omega}^{-1}(z - \boldsymbol{\mu})$ and $d(z) = (z - \boldsymbol{\mu})^T\boldsymbol{\Omega}^{-1}(z - \boldsymbol{\mu})$. This density is expressed as the product of a multivariate student-t pdf and a multivariate student-t cdf. In fact, $T_n(\cdot\,; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \tau)$, is the cdf of $t_n(\cdot\,; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \tau)$. Here, the notation $z \sim \text{uST}_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\Delta}, \nu)$ will be used. In the latter case, ν corresponds to the degrees of freedom parameters. When $\nu \to \infty$, the multivariate skew-t distribution approaches the multivariate skew-normal distribution and reduces to the multivariate normal distribution when $\boldsymbol{\Delta} = \mathbf{0}$.

Similar to the uMSN version, the uMST admits a convenient hierarchial representation,

$$z|w \sim N_n\left(\boldsymbol{\mu} + \boldsymbol{\Delta}w, \frac{1}{r}\boldsymbol{\Sigma}\right),$$

$$w|r \sim HN_n\left(\mathbf{0}, \frac{1}{r}\boldsymbol{I}_N\right),$$

$$r \sim \gamma\left(\frac{\tau}{2}, \frac{\tau}{2}\right). \tag{3.6}$$

Here, for the *homogeneous*, *heterogeneous* and *fixed* region-specific effects, we only consider the skew-elliptical distributions (uMSN and uMST) for the remainder error component. For the *random* region-specific effect, we allow both ϵ_t and α_t to have a skew-t, skew-normal, student-t, and normal distributions.

The proposed panel model incorporates all types of spatial heterogenity mentioned above and is referred to as a skew-elliptical linear spatial panel model. The MCMC computations for the skew-elliptical panel model will be derived based on the following spatial hierarchical Bayes model. The hierarchical representation is important for obtaining the parameter estimates and will facilitate easy implementation in commonly used softwares such as R and OpenBUGS. In this case, from the stochastic representation, and by introducing the $N \times 1$ random vector $\boldsymbol{w}_k$, the hierarchical representation of the model can be formulated as follows. For $t = 1, \ldots, T$ we have,

$$\boldsymbol{y}_t|\boldsymbol{\beta}, \boldsymbol{\alpha}_t, \sigma_\epsilon^2, \boldsymbol{w}_{\epsilon_t}, \delta_\epsilon \sim N_N\left(X_t\boldsymbol{\beta} + \boldsymbol{\alpha}_t + \delta_\epsilon\boldsymbol{w}_{\epsilon_t}, \frac{1}{r}\sigma_\epsilon^2\boldsymbol{I}_N\right),$$

$$\boldsymbol{w}_{\epsilon_t} \sim N_N\left(\mathbf{0}, \frac{1}{r}\boldsymbol{I}_N\right)I(\boldsymbol{w}_{\epsilon_t} > \mathbf{0}),$$

$$r \sim \gamma\left(\frac{\tau}{2}, \frac{\tau}{2}\right). \tag{3.7}$$

where $I(\boldsymbol{w}_k > \mathbf{0})$ is an indicator function which is 1 if every element of $\boldsymbol{w}_k$ is greater than zero, and zero otherwise, and $\boldsymbol{w}_k = |\boldsymbol{\xi}|$ with $\boldsymbol{\xi} \sim N_k(\mathbf{0}, \boldsymbol{I}_k)$. Here, $|\boldsymbol{\xi}|$ denotes the vector with i^{th} element equal to the magnitude of the ith element of $\boldsymbol{\xi}$.

3.1. Distributions of priors and posteriors

In order to implement Bayesian inference for the aforementioned models, we need to specify the prior distribution for all unknown parameters in the models. A popular choice is to consider proper conditionally non-informative conjugate priors as suggested by Gelman and Hill (2006). To obtain well defined and appropriate posteriors we assign conjugate weakly informative priors. Note that, a family of prior distributions $p(\varphi)$ is conditionally conjugate for φ if the conditional posterior distribution $(\varphi|y)$ is also in that class.

The regression parameter $\boldsymbol{\beta}$ is customarily modelled using a multivariate normal distribution $N(\boldsymbol{\beta}_0, \boldsymbol{S}_{\beta_0})$. Usually vague priors are specified by taking $\boldsymbol{S}_{\beta_0}$ to be a diagonal matrix with large elements. The variance components σ_ϵ^2 and σ_α^2 are taken to have an inverse gamma distribution and the skewness parameters δ_ϵ and δ_α are taken to have independent truncated normal distributions. More specifically, we adopt $IG(\sigma_{\epsilon_0}^2, \gamma_{\epsilon_0})$ and $IG(\sigma_{\alpha_0}^2 \gamma_{\alpha_0})$ for σ_ϵ^2 and σ_α^2, respectively, and $N(\mu_\epsilon, \gamma_\epsilon)I(\delta_\epsilon > 0)$ and $N(\mu_\alpha, \gamma_\alpha)I(\delta_\alpha > 0)$ for δ_ϵ and δ_α respectively.

To implement Gibbs sampling methodology and, in particular, to simulate samples from the posterior distribution iteratively, it is critical to obtain the conditional posterior distributions of the parameters. The full conditional distributions of $\boldsymbol{\beta}$, $\boldsymbol{W}_{\epsilon_t}$, σ_ϵ^2, δ_ϵ, can be obtained after some algebraic manipulations and are given by

$$\boldsymbol{\beta}|\boldsymbol{\alpha}_t, \sigma_\epsilon^2, \delta_\epsilon, \boldsymbol{w}_{\epsilon_t} \sim N_k\left(A_\beta^{-1}\right)\boldsymbol{a}_\beta, A_\beta^{-1}, \tag{3.8}$$

where,

$$A_\beta = \boldsymbol{\Gamma}_{\beta_0}^{-1} + \sum_{t=1}^T \frac{X_t' X_t}{\sigma_\epsilon^2},$$

$$\boldsymbol{a}_\beta = \boldsymbol{\Gamma}_{\beta_0}^{-1}\boldsymbol{\beta}_0 + \sum_{t=1}^T \frac{X_t'\left(\boldsymbol{y}_t - \boldsymbol{\alpha}_t - \delta_\epsilon \boldsymbol{w}_{\epsilon_t}\right)}{\sigma_\epsilon^2},$$

$$\boldsymbol{w}_{\epsilon_t}|\boldsymbol{\beta}, \boldsymbol{\alpha}_t, \sigma_\epsilon^2, \delta_\epsilon \sim N_N\left(A_{w_{\epsilon_t}}^{-1}\boldsymbol{a}_{w_{\epsilon_t}}, A_{w_{\epsilon_t}}^{-1}\right)I\{\boldsymbol{w}_{\epsilon_t} > \boldsymbol{0}\}, \tag{3.9}$$

where,

$$A_{w_{\epsilon_t}} = \left(\frac{\delta_\epsilon^2}{\sigma_\epsilon^2} + 1\right)\boldsymbol{I}_N,$$

$$\boldsymbol{a}_{w_{\epsilon_t}} = \left(\frac{\delta_\epsilon}{\sigma_\epsilon^2}\right)\left(\boldsymbol{y}_t - X_t\boldsymbol{\beta} - \boldsymbol{\alpha}_t\right),$$

$$\sigma_\epsilon^2|\boldsymbol{\beta}, \boldsymbol{\alpha}_t, \delta_\epsilon, \boldsymbol{w}_{\epsilon_t} \sim IG\left(\sigma_{\epsilon_n}^2, \gamma_{\epsilon_n}\right), \tag{3.10}$$

$$\sigma_{\epsilon_n}^2 = \sigma_{\epsilon_0}^2 + \frac{NT}{2},$$

$$\gamma_{\epsilon_n} = \gamma_{\epsilon_0} + \sum_{t=1}^T \frac{\left(\boldsymbol{y}_t - X_t\boldsymbol{\beta} - \boldsymbol{\alpha}_t - \delta_\epsilon \boldsymbol{w}_{\epsilon_t}\right)'\left(\boldsymbol{y}_t - X_t\boldsymbol{\beta} - \boldsymbol{\alpha}_t - \delta_\epsilon \boldsymbol{w}_{\epsilon_t}\right)}{2},$$

$$\delta_\epsilon|\boldsymbol{\beta}, \boldsymbol{\alpha}_t, \sigma_\epsilon^2, \boldsymbol{w}_{\epsilon_t} \sim N\left(\mu_{\delta_\epsilon}, \sigma_{\delta_\epsilon}^2\right)I\{\delta_\epsilon > 0\}, \tag{3.11}$$

and where,

$$\mu_{\delta_\epsilon} = \sigma^2_{\delta_\epsilon} \left\{ \frac{\mu_\epsilon}{\gamma_\epsilon} + \sum_{t=1}^{T} \frac{w'_{\epsilon_t}(y_t - X_t\beta - \alpha_t)}{\sigma^2_\epsilon} \right\},$$

$$\sigma^2_{\delta_\epsilon} = \left(\frac{1}{\gamma_\epsilon} + \sum_{t=1}^{T} \frac{w'_{\epsilon_t}w_{\epsilon_t}}{\sigma^2_\epsilon} \right)^{-1}.$$

Additionally, we must specify the prior distribution and derive the posterior distribution for different specifications of unobserved effects α_t.

As previously mentioned, in this paper we assume that the α_i are unobserved effects and that they may vary with spatial and temporal variation depending on the specifications. We will consider all four spatial-temporal types of models specified in Zheng *et al.* (2008). The following sections provide technical details.

3.1.1. Homogeneous model

In the case of the homogeneous model, the unobserved region-specific effects at time t are assumed to be region-homogeneous and time-homogeneous with $\alpha_t = \alpha\iota_N$, where α is an overall mean. The prior distribution for α_t can be specified as follows,

$$\alpha_t \sim T_N(\alpha_0, \Gamma_{\alpha_0}, \nu), \tag{3.12}$$

where $\Gamma_{\alpha_0} = \gamma_{\alpha_0} I_N$. Thus the full conditional distribution of α_t is,

$$\alpha_t | y_t, \beta, \sigma^2_\epsilon, \delta_\epsilon, w_{\epsilon_t} \sim T_N\left(A_{\alpha_t}^{-1} a_t, A_{\alpha_t}^{-1}, \nu\right), \tag{3.13}$$

where $A_{\alpha_t} = (\mathrm{NT}/\sigma^2_\varepsilon + 1/\gamma_{\alpha_0})^{-1} I_N$ and $a_t = (y_t - X_t\beta - \delta_\epsilon w_{\epsilon_t})/\sigma^2_\epsilon + \alpha_0/\gamma_{\alpha_0}$.

3.1.2. Heterogeneous model

In the heterogeneous model, the unobserved region-specific effects are assumed to be region-homogeneous but time-inhomogeneous with $\alpha_t = \alpha_t\iota_N$, where α_t is a mean across regions at time t. Suppose that $\alpha = (\alpha_1, \ldots, \alpha_T)'$ denotes the vector of the means over time. Also, as the regression coefficients and the spatial autocorrelation coefficient are assumed to be region-homogeneous and time-inhomogeneous, we can write $\beta = (\beta'_1, \ldots, \beta'_T)'$, where $\beta_t = (\beta_{1t}, \beta_{2t}, \ldots, \beta_{kt})'$. By using the aforementioned stochastic representation, we have

$$y | \beta, \alpha, \sigma^2_\epsilon \sim uST_{\mathrm{NT}}\left(\tilde{X}\beta + \alpha, \sigma^2_\epsilon I_{\mathrm{NT}}, \delta_\epsilon I_{\mathrm{NT}}, \nu\right), \tag{3.14}$$

where $\tilde{X} = (\tilde{X}_1, \ldots, \tilde{X}_T)'$. Let $\tilde{X}_t = \iota_t \otimes X_t$, where ι_t denotes the tth row of I_T.

We obtain the full conditional distribution of β as follows,

$$\beta | \alpha, \sigma^2_\epsilon, \delta_\epsilon, w_\epsilon \sim N_p\left(A_\beta^{-1} a_\beta, A_\beta^{-1}\right), \tag{3.15}$$

where $A_\beta = \Gamma_{\beta_0}^{-1} + \sum_{t=1}^{T} \tilde{X}'_t \tilde{X}_t/\sigma^2_\epsilon$ and $a_\beta = \Gamma_{\beta_0}^{-1}\beta_0 + \sum_{t=1}^{T} \tilde{X}'_t(y_t - \alpha_t - \delta_\epsilon w_{\epsilon_t})/\sigma^2_\epsilon$, and the full conditional distribution of σ^2_ϵ is given by,

$$\sigma^2_\epsilon | \beta, \alpha_t, \delta_\epsilon, w_{\epsilon_t} \sim \mathrm{IG}\left(\sigma^2_{\epsilon_n}, \gamma_{\epsilon_n}\right), \tag{3.16}$$

where $\sigma^2_{\epsilon_n} = \sigma^2_{\epsilon_0} + \text{NT}/2$ and $\gamma_{\epsilon_n} = \gamma_{\epsilon_0} + \sum_{t=1}^T (\boldsymbol{y}_t - \tilde{X}_t \boldsymbol{\beta} - \boldsymbol{\alpha}_t - \delta_\epsilon \boldsymbol{w}_{\epsilon_t})'(\boldsymbol{y}_t - \tilde{X}_t \boldsymbol{\beta} - \boldsymbol{\alpha}_t - \delta_\epsilon \boldsymbol{w}_{\epsilon_t})/2$. In addition, we assume that

$$\boldsymbol{\alpha} \sim T_{NT}(\boldsymbol{\alpha}_0, \boldsymbol{\Gamma}_{\alpha_0}, \nu), \tag{3.17}$$

with $\boldsymbol{\Gamma}_{\alpha_0} = \gamma_{\alpha_0} \boldsymbol{I}_{\text{NT}}$, and thus the full conditional distribution of $\boldsymbol{\alpha}$ is given by,

$$\boldsymbol{\alpha}|\boldsymbol{y}, \boldsymbol{\beta}, \sigma^2_\epsilon \sim T_{\text{NT}}\left(A_\alpha^{-1} \boldsymbol{a}_\alpha, A_\alpha^{-1}, \nu\right), \tag{3.18}$$

where $A_\alpha = \boldsymbol{\Gamma}_{\alpha_0}^{-1} + \sigma^2_\epsilon A$ and $\boldsymbol{a}_\alpha = \boldsymbol{\Gamma}_{\alpha_0}^{-1} \boldsymbol{\alpha}_0 + \sigma^2_\epsilon B$, such that A is a $T \times T$ diagonal matrix with diagonal elements $A(t,t) = \boldsymbol{\iota}_N \boldsymbol{\iota}'_N$ and B is a T dimensional vector with the tth element $B(t) = \boldsymbol{\iota}'_N(\boldsymbol{y}_t - \tilde{X}_t \boldsymbol{\beta} - \delta_\epsilon \boldsymbol{w}_{\epsilon_t})$.

3.1.3. Fixed-effects model

In this model, the unobserved region-specific effects are assumed to be time invariant and thus we can set $\boldsymbol{\alpha}_t = \tilde{\boldsymbol{\alpha}}$ with no time subscript. Note that, however, this vector is assumed to be time-homogeneous but region-inhomogeneous, that is, $\boldsymbol{\alpha}_t = \tilde{\boldsymbol{\alpha}} = (\alpha_1, \dots, \alpha_N)'$ in which α_i is a mean over time for region i. We use the following prior distribution for $\tilde{\boldsymbol{\alpha}}$,

$$\tilde{\boldsymbol{\alpha}} \sim T_N(\boldsymbol{\alpha}_0, \boldsymbol{\Gamma}_{\alpha_0}, \nu), \tag{3.19}$$

with $\boldsymbol{\Gamma}_{\alpha_0} = \gamma_{\alpha_0} \boldsymbol{I}_N$. Thus the full conditional distribution of $\tilde{\boldsymbol{\alpha}}$ is,

$$\tilde{\boldsymbol{\alpha}}|\boldsymbol{y}, \boldsymbol{\beta}, \sigma^2_\epsilon \sim T_N\left(A_{\tilde{\alpha}}^{-1} \boldsymbol{a}_{\tilde{\alpha}}, A_{\tilde{\alpha}}^{-1}, \nu\right), \tag{3.20}$$

where $A_{\tilde{\alpha}} = \boldsymbol{\Gamma}_{\alpha_0}^{-1} + T/\sigma^2_\epsilon$ and $\boldsymbol{a}_{\tilde{\alpha}} = \boldsymbol{\Gamma}_{\alpha_0}^{-1} \boldsymbol{\alpha}_0 + \sum_{t=1}^T (\boldsymbol{y}_t - X_t \boldsymbol{\beta} - \delta_\epsilon \boldsymbol{w}_{\epsilon_t})$.

3.1.4. Random-effects model

In this case, the unobserved region-specific effects vector are assumed to be time-homogeneous but region-inhomogeneous with $\boldsymbol{\alpha}_t = \tilde{\boldsymbol{\alpha}} = (\alpha_1, \dots, \alpha_N)'$, where the α_i's are iid random variables with $N(\alpha^\star, \sigma^2_\alpha)$ distribution. Thus the full conditional distribution of $\tilde{\boldsymbol{\alpha}}$ is,

$$\tilde{\boldsymbol{\alpha}}|\boldsymbol{y}, \boldsymbol{\beta}, \sigma^2_\epsilon, \alpha^\star \sim T_N\left(A_{\tilde{\alpha}}^{-1} \boldsymbol{a}_{\tilde{\alpha}}, A_{\tilde{\alpha}}^{-1}, \nu\right), \tag{3.21}$$

where $A_{\tilde{\alpha}} = (1/\sigma^2_\alpha + T/\sigma^2_\epsilon)\boldsymbol{I}_N$ and $\boldsymbol{a}_{\tilde{\alpha}} = \alpha^\star \boldsymbol{\iota}_N/\sigma^2_\alpha + \sum_{t=1}^T (\boldsymbol{y}_t - X_t \boldsymbol{\beta} - \delta_\epsilon \boldsymbol{w}_{\epsilon_t})/\sigma^2_\epsilon$. We use the following prior distributions for $\alpha^\star$ and σ^2_α,

$$\alpha^\star \sim N\left(\alpha_0^\star, \gamma_{\alpha_0}^\star\right), \quad \sigma^2_\alpha \sim \text{IG}\left(\sigma^2_{\alpha_0}, \gamma_{\alpha_0}\right). \tag{3.22}$$

The full conditional distribution of $\alpha^\star$ is then,

$$\alpha^\star|\tilde{\boldsymbol{\alpha}}, \sigma^2_\alpha \sim N\left(\alpha_n^\star, \gamma_{\alpha_n}^\star\right), \tag{3.23}$$

where $\alpha_n^\star = \gamma_{\alpha_n}^\star(\alpha_0^\star/\gamma_{\alpha_0}^\star + \boldsymbol{\iota}'_N \tilde{\boldsymbol{\alpha}}/\sigma^2_\alpha)$ and $\gamma_{\alpha_n}^\star = (1/\gamma_{\alpha_0}^\star + N/\sigma^2_\alpha)^{-1}$. The full conditional distribution of σ^2_α is,

$$\sigma^2_\alpha|\tilde{\boldsymbol{\alpha}}, \alpha^\star \sim \text{IG}\left(\sigma^2_{\alpha_0} + \frac{N}{2}, \gamma_{\alpha_0} + \frac{(\tilde{\boldsymbol{\alpha}} - \alpha^\star \boldsymbol{\iota}_N)'(\tilde{\boldsymbol{\alpha}} - \alpha^\star \boldsymbol{\iota}_N)}{2}\right). \tag{3.24}$$

When $\boldsymbol{\alpha}_t$ has the skew-t distribution, we have that

$$\boldsymbol{\alpha}_t \big| \sigma_\alpha^2, \delta_\alpha, \boldsymbol{W}_{\alpha_t} = \boldsymbol{w}_{\alpha_t} \sim N_N\left(\delta_\alpha \boldsymbol{w}_{\alpha_t}, \frac{1}{r}\sigma_\alpha^2 \boldsymbol{I}_N\right),$$

$$\boldsymbol{w}_{\alpha_t} \sim N_N\left(\boldsymbol{0}, \frac{1}{r}\boldsymbol{I}_N\right) I(\boldsymbol{w}_{\alpha_t} > \boldsymbol{0}), \tag{3.25}$$

and so the full conditional distribution of $\boldsymbol{W}_{\alpha_t}$ given $\boldsymbol{\alpha}_t$, σ_α^2, and δ_α is given by

$$\boldsymbol{W}_{\alpha_t} | \boldsymbol{\alpha}_t, \sigma_\alpha^2, \delta_\alpha \sim N_N\left(\boldsymbol{A}_{\alpha_t}^{-1} \boldsymbol{a}_{w_{\alpha_t}}, \frac{1}{r}\boldsymbol{A}_{w_{\alpha_t}}^{-1}\right) I\{\boldsymbol{w}_{\alpha_t} > \boldsymbol{0}\}, \tag{3.26}$$

where $\boldsymbol{A}_{w_{\alpha_t}} = (\delta_\alpha^2/\sigma_\alpha^2 + 1)\boldsymbol{I}_N$ and $\boldsymbol{a}_{w_{\alpha_t}} = (\delta_\alpha/\sigma_\alpha^2)\boldsymbol{\alpha}_t$. Similarly, the full conditional distribution of σ_α^2 given $\boldsymbol{\alpha}_t$, δ_α, and $\boldsymbol{w}_{\alpha_t}$ is given by

$$\sigma_\alpha^2 \big| \boldsymbol{\alpha}_t, \delta_\alpha, \boldsymbol{w}_{\alpha_t} \sim \text{IG}\left(\sigma_{\alpha_n}^2, \gamma_{\alpha_n}\right), \tag{3.27}$$

where $\sigma_{\alpha_n}^2 = \sigma_{\alpha_0}^2 + NT/2$ and $\gamma_{\alpha_n} = \gamma_{\alpha_0} + \sum_{t=1}^T (\boldsymbol{\alpha}_t - \delta_\alpha \boldsymbol{w}_{\alpha_t})'(\boldsymbol{\alpha}_t - \delta_\alpha \boldsymbol{w}_{\alpha_t})/2$. Further, we have

$$\delta_\alpha \big| \boldsymbol{\alpha}_t, \sigma_\alpha^2, \boldsymbol{w}_{\alpha_t} \sim N\left(\mu_{\delta_\alpha}, \sigma_{\delta_\alpha}^2\right) I\{\delta_\alpha >\}, \tag{3.28}$$

where $\mu_{\delta_\alpha} = \sigma_{\delta_\alpha}^2 \{\mu_\alpha/\gamma_\alpha + \sum_{t=1}^T \boldsymbol{w}_{\alpha_t}'(\boldsymbol{\alpha}_t)/\sigma_\alpha^2\}$ and $\sigma_{\delta_\alpha}^2 = (1/\gamma_\alpha + \sum_{t=1}^T \boldsymbol{w}_{\alpha_t}' \boldsymbol{w}_{\alpha_t}/\sigma_\alpha^2)^{-1}$.

In order to determine whether the model fits the data adequately, key selection criteria must be applied. These criteria are based on the values of the penalized log likelihood, the Akaike information criterion (AIC) (Akaike, 1974), the Bayesian information criterion (BIC) (Schwarz, 1978) and the Hannan Quinn criterion (HQC) (Hannan and Quinn, 1979) are used to assess the performance of asymmetric models and compare it to the symmetric models. They are defined by AIC= $\hat{D} + 2P$, BIC= $\hat{D} + P\log(N)$, and HQC= $\hat{D} + 2P\log(\log(N))$, respectively, where $\hat{D}$ is -2 times log likelihood evaluated at the maximum likelihood estimate, P is the number of parameters, and N is the sample size. It should be noted that a model with a smaller value of AIC, BIC, and HQC is preferred when comparing different fitted results.

4. Simulation study

The following section presents simulated data to be used in assessing the performance of the panel models with spatial heterogeneity taken into account. We compare the results of skew models to those obtained by traditional models. Following is a representation of a simple panel data model that provides heterogeneity across units. To generate random samples for our simulations, we can use it as a simple method. For a given observation, an intercept varying over units results in the structure:

$$y_{it} = \sum_{r=1}^{3} \beta_r x_{rit} + \alpha_i + \epsilon_{it}, \quad i = 1, \ldots, N, \ t = 1, \ldots, T. \tag{4.1}$$

The true values of the model parameters are chosen such that all simulated data make sense and resemble a realistic scenario. Specifically, the true values of the model parameters are $(\beta_1, \beta_2, \beta_3) = (1.5, 6, 9)$ and for the model generating the $\epsilon_{it} \sim \text{SN}_1(0, 1, 10)$ and $\epsilon_{it} \sim \text{ST}_1(0, 1, 10)$. The x_{rit} is generated similar to Nervlove (1971) by taking $x_{rit} = 0.1t + 0.5x_{r,i,t-1} + z_{rit}$, where z_{rit} is uniformly

Table 1: The average mean (average standard deviation) and the equivalence 95% credible intervals for the estimated parameters of the simulation study are given in large parentheses

Model	Parameter	Region-specific effect type			
		homogeneous	*heterogeneous*	*fixed*	*random*
(a) Normal	β_1	1.97 (0.49)	2.00 (0.51)	2.42 (0.55)	1.75 (0.68)
		(0.79, 3.33)	(0.84, 2.99)	(0.96, 3.23)	(0.08, 3.01)
	β_2	5.62 (0.66)	5.68 (0.89)	5.67 (0.74)	5.89 (0.90)
		(5.01, 6.78)	(5.66, 6.75)	(5.22, 6.97)	(4.77, 6.22)
	β_3	8.53 (0.55)	8.25 (0.57)	8.94 (0.54)	8.57 (0.59)
		(7.66, 9.93)	(7.59, 9.21)	(7.93, 9.90)	(7.14, 9.60)
	$\alpha^\star$	-	-	-	10.62 (1.05)
		-	-	-	(8.43, 10.67)
	σ_ϵ^2	40.03 (4.55)	39.32 (5.01)	39.11 (5.65)	42.78 (5.88)
		(30.86, 48.74)	(26.38, 49.08)	(29.99, 49.62)	(32.14, 50.00)
	σ_α^2	-	-	-	1.11 (1.79)
		-	-	-	(0.11, 5.87)
(b) Skew-Normal	β_1	1.55 (0.50)	1.32 (0.49)	1.43 (0.47)	1.55 (0.31)
		(0.60, 1.93)	(0.49, 1.76)	(0.57, 1.64)	(0.55, 1.62)
	β_2	5.42 (0.62)	5.45 (0.60)	4.99 (0.64)	5.04 (0.81)
		(4.44, 5.93)	(4.77, 5.90)	(4.29, 5.98)	(4.33, 5.35)
	β_3	7.30 (0.32)	7.61 (0.41)	7.56 (0.35)	7.27 (0.55)
		(7.14, 8.99)	(7.17, 8.01)	(7.12, 8.79)	(7.20, 8.72)
	σ_ϵ^2	1.60 (1.29)	1.64 (1.38)	1.78 (1.34)	1.84 (1.10)
		(0.08, 4.23)	(0.13, 4.35)	(0.10, 4.49)	(0.37, 4.04)
	σ_α^2	-	-	-	4.06 (1.77)
		-	-	-	(1.56, 5.90)
	δ_ϵ	9.01 (0.88)	8.87 (0.76)	8.09 (0.69)	9.92 (0.99)
		(7.09, 9.64)	(7.02, 10.03)	(7.77, 9.89)	(8.59, 10.10)
	δ_α	-	-	-	4.08 (1.60)
		-	-	-	(0.65, 6.02)

distributed on the interval $(-0.5, 0.5)$ and $x_{ri0} = 5 + 10z_{ri0}$. There are four interpretations of α_i in this context: as a parameter to be estimated in the model (the so-called *homogeneous*, *heterogeneous* and *fixed* models) or alternatively, as a component of the disturbance process, giving rise to a composite error term $[\alpha_i + \epsilon_{it}]$ (the so-called *random* model), where $\alpha_i \sim SN_1(0, 4, 5)$ and $\alpha_i \sim ST_1(0, 4, 5, 3)$ in skew version of α_i. Under each interpretation, α_i is taken as a random variable. The sample sizes (N, T) in the different experiments were chosen as (25,5).

In this simulation, 100 Monte Carlo data sets were simulated from the model (4.1) according to the additional specifications described below. For the sake of comparison, prior distributions on parameters were the same for all type of models. Fixed effects are assigned a $N_3(\mathbf{0}, 10^2 \mathbf{I}_3)$ distribution. For the scale parameter of error distribution, σ_ϵ and σ_α, we use IG(0.01, 0.01) as the prior, and for the skewness parameters δ_ϵ and δ_α, independent truncated $N(0, 10^2)$ distributions are assigned. Note that all these prior distributions are considered close to being non-informative (with large variances). Thus, we expect robust results with respect to prior distributions. For the purpose of undertaking inference, we have used 50,000 iterates in a generated chain after discarding the initial 2,000 iterates. In Table 1 and Table 2, summaries of the Monte Carlo results for the posterior mean average, standard deviation average, and the equitailed 95% credible intervals average for the parameters for the four types of skew elliptical spatial panel models are recorded. These listed results for the equitailed 95% credible intervals are the averages of their lower and upper endpoints. We note that an another measure that could have been used for these intervals was their coverage probabilities.

Table 2: average mean (average standard deviation) and equitailed 95% credible intervals average in big parentheses for the estimated parameters of the simulation study

Model	Parameter	Region-specific effect type			
		homogeneous	*heterogeneous*	*fixed*	*random*
(a) Student-*T*	β_1	1.64 (0.45)	1.87 (0.48)	1.78 (0.51)	1.43 (0.65)
		(0.88, 2.57)	(0.85, 2.81)	(0.73, 3.31)	(0.90, 2.88)
	β_2	5.43 (0.61)	5.12 (0.62)	5.01 (0.49)	5.03 (0.52)
		(4.78, 6.22)	(4.12, 6.09)	(4.48, 6.66)	(4.90, 6.01)
	β_3	8.48 (0.57)	8.795 (0.59)	8.88 (0.52)	8.60 (0.55)
		(7.23, 9.87)	(7.60, 8.93)	(7.80, 9.10)	(7.76, 9.46)
	$\alpha^\star$	-	-	-	10.07 (1.11)
		-	-	-	(8.77, 10.46)
	σ_ϵ^2	39.39 (5.80)	38.81 (5.39)	39.06 (5.10)	37.95 (5.44)
		(29.60, 49.04)	(27.69, 50.01)	(29.80, 48.98)	(30.70, 49.57)
	σ_α^2	-	-	-	1.01 (1.67)
		-	-	-	(0.19, 5.59)
(b) Skew-*T*	β_1	1.47 (0.44)	1.53 (0.46)	1.52 (0.45)	1.52 (0.39)
		(0.67, 1.72)	(0.52, 1.70)	(0.58, 1.60)	(0.50, 1.59)
	β_2	5.22 (0.53)	5.28 (0.51)	5.02 (0.52)	5.05 (0.32)
		(4.69, 5.72)	(4.44, 5.70)	(4.11, 5.29)	(4.42, 5.28)
	β_3	7.72 (0.32)	7.46 (0.46)	7.57 (0.56)	7.85 (0.47)
		(6.12, 7.86)	(6.26, 7.99)	(7.11, 8.27)	(7.13, 8.22)
	σ_ϵ^2	1.49 (1.20)	1.55 (1.27)	1.98 (1.25)	1.02 (1.14)
		(0.39, 4.50)	(0.18, 4.32)	(0.36, 4.39)	(0.33, 3.96)
	σ_α^2	-	-	-	4.22 (3.13)
		-	-	-	(0.71, 4.93)
	δ_ϵ	9.23 (0.75)	9.03 (0.71)	8.90 (0.66)	9.97 (0.62)
		(7.40, 9.51)	(7.53, 9.80)	(7.62, 9.09)	(8.83, 10.06)
	δ_α	-	-	-	5.01 (1.23)
		-	-	-	(0.89, 5.72)

It is found that, for any of the 100 data sets for each model, no value of the AIC, BIC and HQC criteria leads to the selection of the normal model. This indicates an obvious departure from normality and suggests strong evidence of skewness. We also found that the parameter estimates among the four skew-elliptical models are different from those obtained with symmetric (normal and student-*t*) and asymmetric (skew-normal and skew-*t*) distributions. The parameters estimated for each of the four spatial panel models with skew-normal and skew-*t* error terms can be summarized as follows: (i) Equivalent 95% credible intervals associated with each parameter fitted value of symmetric models are larger than those of the corresponding asymmetric models; (ii) The parameters estimated for skew models have a smaller standard deviation compared to those for symmetric models; (iii) the estimate of the variance component of error σ_ϵ^2, and σ_α^2 for the *random* model is smaller in the class of skew models compared to the normal and student-*t* ones, primarily because of skewness; (iv) the estimated equitailed 95% credible intervals of δ_ϵ, and δ_α for the multivariate skew-normal and skew-*t* models do not include zero, confirming positive skewness of the responses.

In our empirical analysis, we expect the asymmetry parameters to be positive since the log transformation is the most popular transformation for positively skewed distributions in linear regression. The results also show that consideration of departures from normality can help in improving model fit.

The results shown in Table 1 and Table 2 indicate that we can recover the true values of the parameters used to generate the data with a reasonable level of accuracy. In fact, we analysed the simulated data generated from the model to check our ability in recovering the true values of the

parameters. In most of the different samples, the true values of each of these parameters lie within their respective posterior equilibrated 95% credible intervals, indicating that we are able to recover the true values of the parameters used to generate the datasets. As demonstrated by this simulation study, the skew-t model can provide a better fit for all types of models with unobserved spatial heterogeneity than the model that places the normality assumptions on error terms. Particularly, it seems that our evaluations of spatial heterogeneity models with the skew-t assumption are very useful in estimating the regression coefficients β.

5. Analysis of real demand data (Italian non-life insurance)

Using a dataset derived from records of insurance activities in Italy, we illustrate the proposed spatial panel models in this section. A country's economic development depends on the insurance industry, which is usually classified into life and non-life insurance. Life and non-life insurance consumption have different needs, so they're usually analyzed separately. Here we focus on non-life insurance consumption. Purchase of non-life insurance is an economic strategy used to provide protection for future losses by paying a premium today, thus transferring wealth from an indefinite to a definite state. This branch of the insurance market is believed to play an important role in promoting the welfare and growth of the industry. It does so by protecting families and companies against financial hardship caused by unexpected/unusual events such as fire, theft, disease, and accidents.

In our study, we examine non-life insurance consumption across 103 provinces in Italy spanning 1998–2002 in order to assess its determinants. The records include: real per-capita GDP (Gross Domestic Product) (rgdp), real per-capita bank deposits (bank), inhabitants density per square Km (den), real interest rate on lending to families and small enterprises (rirs), density of insurance agencies per 1000 inhabitants (agen), share of people with second grade schooling or more (school), share of value added, agriculture sector (vaagr), average number of family members (fam), judicial inefficiency index: average years to settle first degree of civil case (inef), survey result to the question "do you trust others?" (trust), real non-life insurance premiums per capita (ppcd), the province code (code) and the year of observation (year), reported in Millo and Carmeci (2010). These data are publicly available from splm package in R (Millo and Piras, 2012).

As outlined by Millo and Carmeci (2010), these data are challenging to analyze. In particular, it is difficult to estimate the determinants of the development of the nonlife insurance market for Italy and hence good proxies are needed. There is also a high spatial differentiation across the Italian provinces as well as a high degree of spatial correlation in insurance density. For this illustration, the explanatory variables are rgdp, bank, den, rirs, agen, school, vaagr, fam, inef and trust and the response variable is ppcd. For a discussion of the determinants of insurance, see the papers by Browne *et al.* (2000), Beck and Webb (2003), and Esho *et al.* (2004).

Let us consider the simple panel regression model (2.1) for the total number of non-life insurance premiums years in Italian provinces. The purpose of the study is to investigate the relationship between non-life insurance premiums and some determinants of insurance consumption in the model. In the context of non-life insurance premiums, some observable (income, wealth, etc) as well as unobservable effects (capital stock, litigation attitudes, etc) may affect the panel results.

There is a limit to the level of analysis for Italian insurance premiums. Due to this, we expect some heterogeneity within provinces. Another option would be to eliminate the differences between provinces that are associated with heterogeneity.

The unobserved space-specific effects are captured by the α_i terms. Therefore, following the arguments in Section 3.1, we consider different specifications for the unobserved vector α and then
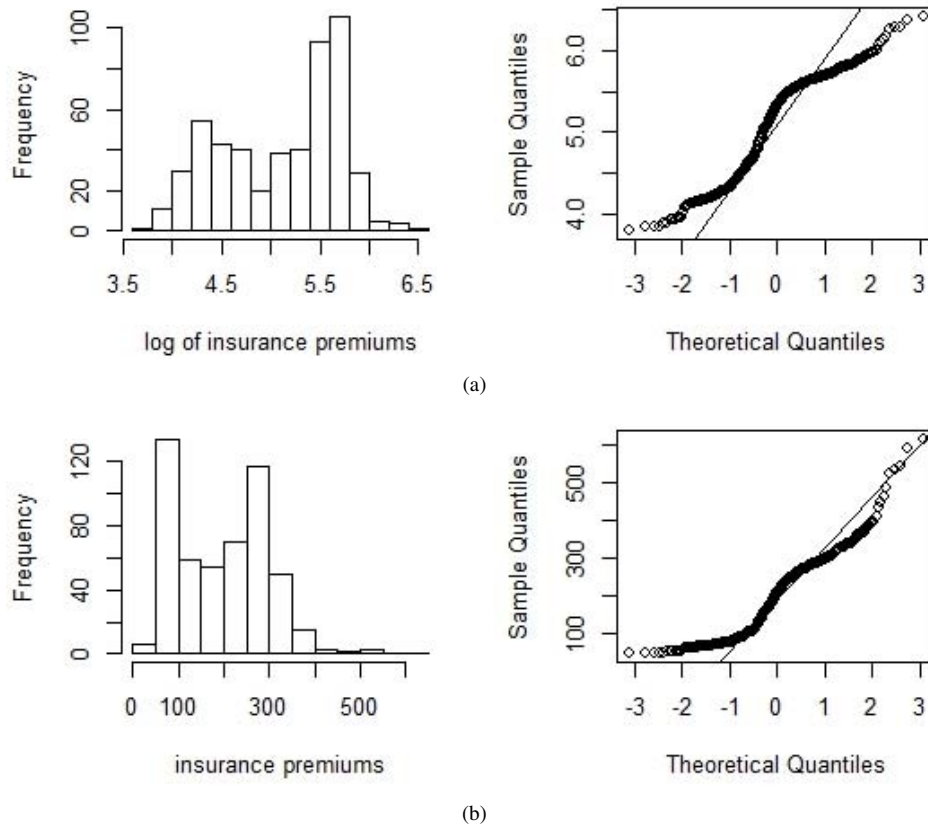
(a)

(b)

Figure 1: *(a) Histogram and Q-Q plot of Italian insurance premiums with log-transformed non-life insurance data. (b) Histogram and Q-Q plot of Italian insurance premiums with raw non-life insurance data.*

assess the spatial panel models based on the estimation results.

In the *homogeneous* model the provincial unobserved effects are the same across both time and provinces, such that α can be defined as

$$\alpha' = (\alpha \cdots \alpha \cdots\cdots \alpha \cdots \alpha), \tag{5.1}$$

which is a 515×1 vector. In the *heterogeneous* model the provincial unobserved effects are the same across the provinces but are different over time. In this case, α has the following structure,

$$\alpha' = (\alpha_1 \cdots \alpha_1 \cdots\cdots \alpha_5 \cdots \alpha_5). \tag{5.2}$$

On the other hand, in the *fixed* model and *random* model, the provincial unobserved effects are assumed to be the same over time but are different across provinces. Hence, in these cases, α is defined as follows,

$$\alpha' = (\alpha_1 \cdots \alpha_{103} \cdots\cdots \alpha_1 \cdots \alpha_{103}) = (\tilde{\alpha} \cdots\cdots \tilde{\alpha}), \tag{5.3}$$

where, in the *random* setting, α follows a specified distribution. For our model, the skew-elliptical distribution is used for the α_i terms. The above-mentioned structures provide four different ways
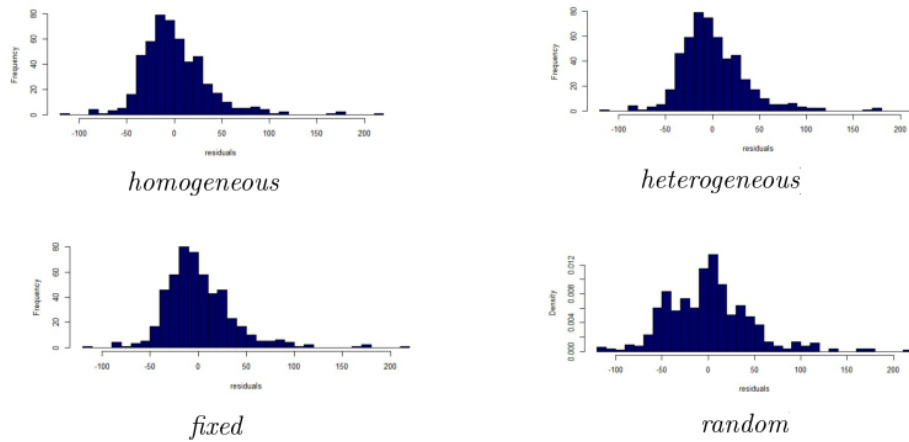
Figure 2: *Data on non-life insurance consumption in Italy. Histogram of residuals for each type of spatial effect after fitting the multivariate normal distribution, for each space.*

to introduce unobserved region-specific heterogeneity in the constant terms or random variables of the panel data model. We employ the approach described in Section 3 for the estimation of model parameters. These models allow us to account for the spatial heterogeneity among different regions. In fact, we follow the Bayesian methodology to perform the inference procedure. We assign independent prior distributions for β, σ_ϵ^2, σ_α^2, δ_ϵ and δ_α and obtain samples from the posterior distribution of the parameters using MCMC methods.

In Millo and Carmeci (2010) the log transformation of the data is used to approximate normality, and then the normal process is applied to model the transformed data. However, it can be observed from the histogram of log-transformed data in Figure 1(a) that the data do not appear normal even after log transformation. The histogram of the raw insurance consumption data in Figure 1(b) also shows that the data look asymmetric. This indicates that the normality assumption is not satisfied and that fitting a skew model to the data set seems more appropriate. Furthermore, we note that the skewness index for the raw non-life insurance premiums is 0.47.

As shown in Figure 2, the histograms display the distribution of residual values obtained after fitting the four types of spatial panel models to the normal distribution assumption. Evidently, the residual distributions are positively skewed and heavily tailed. We do not need to transform data since we use the proposed spatial heterogeneity panel models for the skewed data.

Analysis of the posterior distributions of some of the parameters and different model comparison criteria in Table 3 and Table 4 suggest that the proposed models outperform standard ones used in the literature. It motivates us to use skew models for skew data. Based on the model selection criteria values shown in the bottom of Table 3 and Table 4, the results indicate that skew-*t* model may be the best fitting model and skew-normal model is next, supporting the contention of a departure from normality (not shown). The estimates of other parameters of the skew-elliptical models show similar qualitative behaviour as seen in the analysis of the simulation data when compared with the normal models.

For the spatial heterogeneity panel models, details of implementation of the Gibbs sampler algorithm are as follows. For the skew-elliptical proposal distributions, to proceed with Bayesian inference as in Section 3, we ran the Gibbs sampler algorithm with 40,000 iterations, discarded the 20,000 initial iterates and stored every 20th iteration. Our findings in Table 3 and Table 4 are rather diverse, with

Table 3: The posterior mean (standard deviation) and equitailed 95% credible intervals for the parameters in the skew-normal spatial-temporal models fitted to the Italian non-life insurance data

Parameter	*homogeneous*	*heterogeneous*	*fixed*	*random*
rgdp	0.119 (0.007)	0.122 (0.008)	0.120 (0.007)	0.120 (0.007)
	(0.105, 0.133)	(0.108, 0.137)	(0.107, 0.134)	(0.108, 0.013)
bank	0.009 (0.001)	0.009 (0.001)	0.009 (0.001)	0.009 (0.001)
	(0.008, 0.105)	(0.007, 0.105)	(0.007, 0.0103)	(0.007, 0.010)
den	0.013 (0.004)	0.013 (0.005)	0.013 (0.004)	0.013 (0.004)
	(0.005, 0.021)	(0.004, 0.021)	(0.004, 0.021)	(0.005, 0.022)
rirs	−10.48 (1.722)	−10.17 (1.847)	−10.64 (1.684)	−10.05 (1.628)
	(−14.32, −7.298)	(−13.46, −6.624)	(−14.09, −7.497)	(−13.07, −6.854)
agen	−1.058 (8.997)	−1.341 (8.999)	−0.079 (8.897)	−0.100 (9.506)
	(-17.44, 16.46)	(-18.80, 16.62)	(-17.71, 17.23)	(-19.03, 18.31)
school	−0.882 (0.308)	−0.885 (0.402)	−1.015 (0.322)	−0.964 (0.360)
	(−1.515, −0.296)	(−1.710, −0.126)	(−1.692, −0.450)	(−1.756, −0.271)
vaagr	0.486 (0.713)	0.448 (0.681)	0.394 (0.734)	0.372 (0.670)
	(−0.881, 1.903)	(−0.947, 1.848)	(−1.064, 1.788)	(−0.996, 1.685)
fam	−26.85 (5.192)	−25.86 (6.149)	−27.49 (6.645)	−28.07 (6.277)
	(−36.92, −16.16)	(−37.37, −14.13)	(−39.88, −15.86)	(−40.11, −15.81)
inef	−4.360 (1.280)	−4.305 (1.244)	−4.453 (1.315)	−4.515 (1.321)
	(−6.911, −1.99)	(−6.678, −1.871)	(−7.088, −1.680)	(−6.987, −1.865)
trust	14.00 (4.970)	10.80 (5.834)	15.73 (7.128)	14.54 (6.204)
	(4.548, 24.07)	(0.847, 23.54)	(3.061, 31.43)	(0.689, 25.47)
σ_ϵ^2	307.4 (58.67)	302.7 (58.61)	284.5 (56.38)	294.5 (60.45)
	(209.3, 443.6)	(207.1, 432.6)	(190.0, 412.7)	(184.5, 422.2)
δ_ϵ	48.12 (2.507)	47.88 (2.473)	48.04 (2.584)	47.95 (2.647)
	(43.13, 52.86)	(43.07, 52.73)	(43.02, 52.99)	(42.49, 53.17)
−log likelihood	2383.5	2378	2359.5	2366
AIC	4797	4786	4749	4768
BIC	4797.19	4786.19	4749.19	4768.23
HQC	4766.11	4765.11	4728.11	4742.94

coefficients often changing sign and numerical values, which can be seen as evidence in favour of the need to investigate on skewness and unobserved heterogeneity in panel studies.

Most authors (e.g., Beenstock *et al.* (1988), Outreville (1990), Enz (2000), Esho *et al.* (2004), Millo and Carmeci (2010)) have commented on the elasticity of insurance consumption with respect to income and wealth. Elasticity estimates are frequently used as one of the basic indicators as they are unit-free, easily interpreted, and comparable across studies. In econometrics models with log-transformed data, elasticity values are estimated as the coefficients of the independent variables when the features of the dependent variable, estimation technique, and model structure are determined. However, there are only minor differences among the coefficient estimates and elasticites.

In a wider economic sense, the elasticity value of insurance consumption to income is an interesting subject for research. According to Millo and Carmeci (2010), model specifications and real GDP per capita account for both income and the general level of economic activity, and real bank deposits per capita, as an instrument for the stock of wealth. A negative value for income elasticity is considered consistent with the hypothesis of insurance as an inferior good (Millo and Carmeci, 2010). In empirical studies, it is discussed whether insurance is a superior or a normal good. If income elasticity is positive and smaller than one, then insurance is considered as a normal good. On the contrary, if elasticity is greater than one, insurance is a superior good.

Tables 3 and Table 4 provide estimates of the regression coefficients (posterior means) that appear reasonable based on the observable insurance determinantes. A study of the elasticity of insurance

Table 4: The posterior mean (standard deviation) and equitailed 95% credible intervals for the parameters in the skew-*t* spatial-temporal models fitted to the Italian non-life insurance data

Parameter	*homogeneous*	*heterogeneous*	*fixed*	*random*
rgdp	0.012 (0.001)	0.011 (0.001)	0.012 (0.001)	0.012 (0.001)
	(0.011, 0.013)	(0.010, 0.012)	(0.010, 0.013)	(0.011, 0.013)
bank	0.010 (0.001)	0.010 (0.001)	0.009 (0.001)	0.001 (0.000)
	(0.009, 0.011)	(0.008, 0.011)	(0.008, 0.011)	(0.009, 0.011)
den	0.015 (0.003)	0.012 (0.003.)	0.017 (0.004)	0.015 (0.003)
	(0.009, 0.022)	(0.006, 0.018)	(0.010, 0.024)	(0.008, 0.021)
rirs	−7.034 (1.522)	−9.835 (1.902)	−7.572 (1.593)	−7.577 (1.423)
	(−9.768, −3.906)	(−13.55, −6.491)	(−9.986, −3.383)	(−10.32, −4.819)
agen	−38.07 (1.977)	−20.93 (1.725)	−20.04 (1.747)	−31.48 (1.969)
	(−74.96, −1.522)	(−65.36, 6.566)	(−58.44, 4.77)	(−71.72, 1.646)
school	−0.5176 (0.242)	−0.401 (0.253)	−0.773 (0.286)	−0.362 (0.198)
	(−0.945, −0.031)	(−0.924, 0.052)	(−1.392, −0.241)	(−0.799, −0.026)
vaagr	0.6778 (0.555)	0.027 (0.545)	0.479 (0.595)	0.549 (0.558)
	(−0.497, 1.726)	(−1.018, 1.138)	(−0.681, 1.661)	(−0.523, 1.687)
fam	−46.81 (5.183)	−28.70 (7.510)	−51.36 (5.045)	−42.62 (4.768)
	(−56.67, −36.08)	(−44.41, −14.41)	(−62.19, −41.9)	(−50.65, −32.54)
inef	−3.022 (0.967)	−3.689 (1.08)	−3.007 (1.094)	−3.435 (1.159)
	(−4.708, −0.971)	(−5.598, −1.455)	(−5.147, −0.822)	(−5.615, −0.993)
trust	18.82 (5.752)	12.86 (5.426)	28.16 (4.893)	15.77 (4.967)
	(8.165, 28.43)	(4.427, 24.180)	(18.90, 37.03)	(6.391, 25.16)
σ_ϵ^2	73.04 (25.86)	75.36 (25.48)	72.84 (26.45)	63.71 (2.126)
	(34.45, 134.3)	(36.62, 139.3)	(30.74, 134.3)	(30.31, 107.9)
δ_ϵ	38.99 (3.273)	37.42 (3.005)	36.44 (3.26)	39.95 (3.012)
	(32.53, 45.34)	(31.14, 42.65)	(29.90, 42.49)	(34.09, 46.1)
−log likelihood	2244.5	2259	2244	2202
AIC	4519	4546	4516	4436
BIC	4519.19	4546.18	4516.18	4436.21
HQC	4498.11	4526.51	4496.51	4413.72

consumption with respect to income and wealth, which is relevant to spatial econometrics applications, is presented in this paper. Although there are some numerical differences between the posterior means for skew panel models with unobserved space-specific heterogeneity in the constant terms, they are mostly consistent. Focusing on the results of the *random* model with skew-*t* distribution in Table 4, the elasticity of insurance consumption with respect to income (1.049) turns out to be statistically positive and greater than one, thus asserting the view of non-life insurance as a superior good. The elasticity of wealth is positive but lower than one (0.420), suggesting that the tendency to insure is actually decreasing with wealth. The density of people in this area proves to be negative and significant, indicating that it is a poor agent for risk conditions. The effect of interest rates on insurance consumption is significantly negative. It is evident that the density of agencies is a positive incentive, consistent with the idea that insurance is a complex product with a substantial cost associated with locating an appropriate contract. In the non-life insurance market, family numerousness appears to play a weak role with a negative but significant coefficient, and so does education; the market appears not to be affected by person capital. Agriculture has a positive and significant share of the economy. Judicial system inefficiency has a meaningfully negative impact on insurance. The results confirm the argument that bad implementation of property laws negatively impacts people's willingness to insure. For probably similar reasons as in Guiso *et al.* (2004), trust is an important positive determinant of insurance. These findings are important for understanding Italy's non-life insurance market and its economic development.
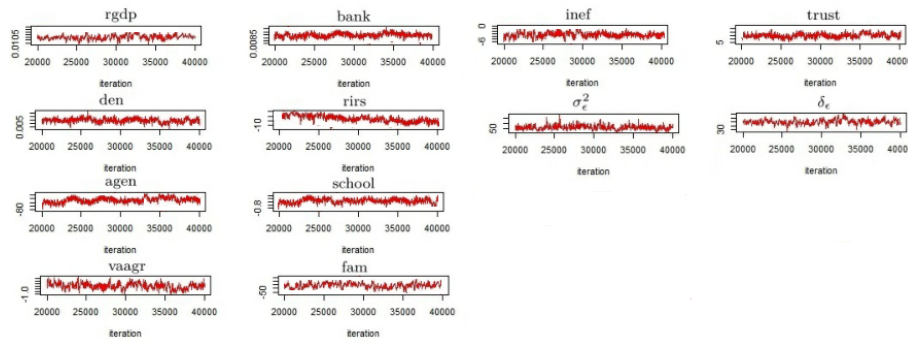
Figure 3: *Markov chain history for parameters under random model of the Italian non-life insurance data.*
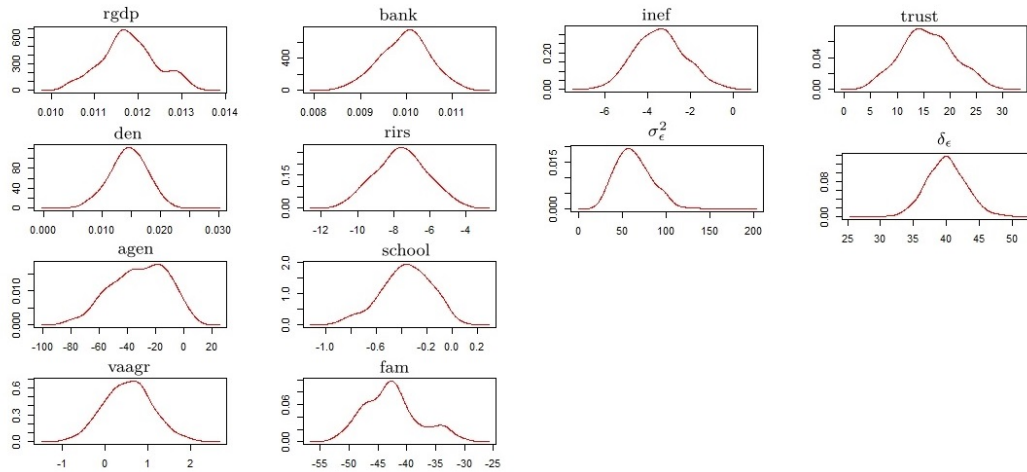

Figure 4: *Markov chain density for parameters under random model of the Italian non-life insurance data.*

The Markov chain history and density for some parameters under *random* skew-*t* spatial panel model is presented in Figures 3 and 4. The following plots confirm convergence of the MCMC.

6. Conclusion

Previous studies have rarely examined the behaviour of the error terms when using spatial heterogeneity, i.e. variation in relationships across spatially dispersed areas. In this paper, we examined multivariate skew-elliptical distributions for the remaining error components of the spatial panel data model. Specifically, our methodologies can be used to model data with skewness. The data can be modeled directly without the need for transformations. The standard approach of including skewness in the error terms of the econometrics is new and differs from the methods used by other studies to calculate other econometrics parameters estimates. Bayesian hierarchical parameter estimation was used for our model. Applying a number of criteria, we investigated the model fitting results for each sub model of our skew-elliptical spatial heterogeneity model. To illustrate the effectiveness of our proposed methodology, we conducted a simulation study and applied our models to real data sets, namely the insurance consumption data sets. The results from these analyses demonstrated the necessity of taking into account skewness in the data. By using a spatial panel data model with skew-elliptical pa-

rameters for the response models, more accurate estimates of econometric parameters can be obtained. In order to bring this about, we used R to implement the MCMC sampling scheme. Analysis of the non-life insurance consumption data set is reported to demonstrate the proposed methodologies. We have also discussed the elasticity of insurance consumption in relation to income and wealth. Panel models using non-normal remainder error components not only provide better results, but they also avoid the need for transformations and work with unrealistic normality assumptions.

References

Akaike H (1974). A new look at the statistical model identification, *IEEE Transactions on Automatic Control*, **19**, 716–723.

Arellano-Valle RB and Genton MG (2005). On fundamental skew distributions, *Journal of Multivariate Analysis*, **96**, 93–116.

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

Azzalini A (2005). The skew-normal distribution and related multivariate families, *Scandinavian Journal of Statistics* , **32**, 159–188.

Azzalini A (2014). *The Skew-Normal and Related Families*, Cambridge University Press.

Beck T and Webb I (2003). Economic, demographic and institutional determinants of life insurance consumption across countries, *World Bank Economic Review*, **17**, 51–88.

Beenstock M, Dickinson G, and Khajuria S (1988). The relationship between propertyliability insurance premiums and income: An international analysis, *The Journal of Risk and Insurance*, **55**, 259–272.

Browne MJ, Chung J, and Frees EW (2000). International property-liability insurance consumption, *The Journal of Risk and Insurance*, **67**, 73–90.

Chesson PL (1985). Coexistence of competitors in spatially and temporally varying environments: A look at the combined effects of different sorts of variability, *Theoretical Population Biology*, **28**, 263–287.

Enz R (2000). The *s*-curve relation between per-capita income and insurance penetration, *Geneva Papers on Risk and Insurance: Issues and Practice*, **25**, 396–406.

Esho NA, Kirievsky A, Ward D, and Zurbruegg R (2004). Law and the determinants of property-casualty insurance, *The Journal of Risk and Insurance*, **71**, 265–283.

Farzammehr MA, Zadkarami MR, and McLachlan GJ (2019). Skew-normal generalized spatial panel data model, *Communications in Statistics-Simulation and Computation*, **50**, 1–29.

Gelman A and Hill J (2006). *Data Analysis Using Regression and Multilevel/Hierarchical Models*, Cambridge University Press.

Guiso L, Sapienza P, and Zingales L (2004). The role of social capital in financial development, *American Economic Review*, **94**, 526–556.

Hannan EJ and Quinn BG (1979). The determination of the order of an autoregression, *Journal of the Royal Statistical Society Series B*, **41**, 190–195.

Jara A, Quintana F, and San Martin E (2008). Linear mixed models with skew-elliptical distributions: A Bayesian approach, *Computational Statistics and Data Analysis*, **52**, 5033–5045.

Lee SX and Mclachlan GJ (2013). On mixtures of skew normal and skew *t*-distributions, *Advances in Data Analysis and Classification*, **7**, 241–266.

Lunn D, Spiegelhalter D, Thomas A, and Best N (2009). The BUGS project: Evolution, critique and future directions, *Statistics in Medicine*, **28**, 3049–3067.

Millo G and Carmeci G (2010). Non-life insurtance consumption in Italy: A sub-regional panel data analysis, *Journal of Geographical Systems* , **13**, 273-–298.

Millo G and Piras G (2012). splm: Spatial panel data models in R, *Journal of Statistical Software*, **47**, 1–38.

Nerlove M (1971). Further evidence on the estimation of dynamic economic relations from a time series of cross sections, *Econometrica*, **39**, 359–382.

Outreville JF (1990). The economic significance of insurance markets in developing countries, *The Journal of Risk and Insurance*, **18**, 487–498.

Sahu SK, Dey DK, and Branco MD (2003). A new class of multivariate skew distributions with applications to Bayesian regression models, *Canadian Journal of Statistics*, **31**, 129–150.

Schwarz G (1978). Estimating the dimension of a model, *The Annals of Statistics*, **6**, 461–464.

Zheng Y, Zhu J, and Li D (2008). Analyzing spatial panel data of cigarette demand: A Bayesian hierarchical modeling approach, *Journal of Data Science*, **6**, 467–489.