Journal of Statistical and Econometric Methods, vol.xx, no.xxx, 2018, xxx-xxx
ISSN: 2241-0384 (print), 2241-0376 (online)
International Scientific Press, 2018

Skew-normal generalized spatial panel data model

Mohadeseh Alsadat Farzammehr[1], Mohammad Reza Zadkarami[2] and Geoffrey J. McLachlan[3]

Abstract

This paper proposes a new generalized spatial panel data model that accounts for skewness in the data. Existing studies on spatial panel data models typically assume a normal distribution for the random error components. This assumption may not be appropriate in many applications. Here we consider a more flexible and powerful approach that generalizes the traditional model. We propose a skew-normal generalized spatial panel data model that adopts a multivariate skew normal distribution for the random error components. For parameter estimation, a Bayesian inference algorithm is developed. A simulation study is conducted to compare the proposed skew normal spatial model with the traditional (normal) spatial model. The proposed model is also illustrated using a real data set of cigarette demand. For this applied data, we analyze the elasticity of per pack cigarette price and per capita disposable income.

Mathematics Subject Classification : xxxxx
Keywords: Spatial panel data model; Spatial dependence; Random effects; Bayesian inference; MCMC; Gibbs sampler; Multivariate skew-normal distribution; Skewness; Spatial econometrics

---

[1]Department of Mathematical Sciences and Computer, Shahid Chamran University of Ahvaz, Ahvaz, Islamic Republic of Iran, E-mail: M-Farzammehr@phdstu.scu.ac.ir
[2]Department of Mathematical Sciences and Computer, Shahid Chamran University of Ahvaz, Ahvaz, Islamic Republic of Iran, E-mail: zadkarami m@scu.ac.ir
[3]Department of Mathematics University of Queensland St. Lucia, Brisbane, Australia 4072, E-mail: g.maclachlan@uq.edu.au

Article Info: *Revised* : September 8, 2018. *Published online* : , 2018

1 Introduction

Panel data models are now widely used in many different disciplines, including political science, sociology, finance, economics, and marketing. Analysis of panel data regression models with the structure of spatial correlation (or spatial dependence) has great importance and interest in empirical economic research. Specifically, we focus on the estimation of panel models with random-effects (RE) specification which takes into account the spatial dependence in the disturbance terms; including unit specific errors and remainder disturbances. Anselin (1988) provides a panel model with a spatial autoregressive (SAR) structure on the remainder disturbances; Kapoor *et al.* (2007) propose a different specification of the model with a SAR structure in the overall disturbance term; Anselin *et al.* (2008) and Fingleton (2008) adopt different spatial moving average (SMA) structure in the two disturbances. In an attempt to nest the models by Anselin (1988) and Kapoor *et al.* (2007), Baltagi *et al.* (2007) suggest an extended model. Subsequently, Lee and Yu (2012) pointed out that a RE model with serial correlation process in the remainder disturbances encompasses not only the previous structures, but also a general case of both SAR and SMA structures (SARMA). This generalized spatial panel data model is not necessarily more valuable nor easier to deal with than the parsimonious models in an empirical application. However, it is a more general approach that allows for both SAR and SMA structure. Hence, we consider this generalized model as the basis of our new generalized spatial panel data model in this paper. As such, the proposed model formally encompasses the aforementioned models as special cases.

In applied work in economics, one often uses the logarithmic transformation of the variables, rather than working with the raw variables directly. This is because the estimated regression coefficients can be directly interpreted as elasticities (the relative response of one variable to changes in another variable) and thus there is no need to take the partial derivatives afterwards. Although this approach gives reasonable empirical results, it has several drawbacks such as reduced information, difficulty in interpreting the results, no universal transformation, and no guarantee of joint normality. These issues motivate us to consider the generalized spatial panel data model that takes into account the skewness in the disturbance terms.

There are several multivariate versions of the skew-normal distribution in the literature; see Azzalini and Dalla Valle (1996), Azzalini and Capitanio (1999), Arellano-Valle and Azzalini (2006) for a discussion of the many existing variants, and a unified treatment of these; Lee and McLachlan (2013, 2014, 2016) also discuss some of these multivariate variants. All of these distributions share several properties similar to the multivariate normal distribution. Arellano-Valle *et al.* (2007) used the skew-normal distribution proposed by Sahu *et al.* (2003) for a linear mixed effects model, but did not consider spatial dependence. Indeed, the aforementioned model is really a skew-normal linear mixed model with spatial dependence. We are not aware of any linear regression model within a spatial econometric context that considers a skew-normal distribution for the error term, although there have been a few applications of the skew-normal distribution in the context of spatial linear regression-type models with continuous observations; see, for example, Molenaar *et al.* (2010), Meintanis and Hlavka (2010), Smith *et al.* (2012), Lin *et al.* (2016) and Bhat *et al.* (2017).

In this paper, we show how the skew-normal is particularly well suited for spatial panel model analysis because it requires only two additional parameters to be estimated relative to the traditional spatial panel models. This is because the structural error terms are marginally skew-normal and distributed with the same amount of skew parameters across random error components. We exploit this feature and impose this specific restrictive form on the multivariate skew-normal distribution that has not been used in the literature. Then, we show that how the generalized spatial panel model with random error components of the skew-normal variety can be estimated with relative ease using the Bayesian estimation approach described in Arellano-Valle *et al.* (2007).

Overall, this paper contributes to the spatial panel data analysis field by providing a more flexible spatial panel data model with a general and robust formulation for the disturbance terms using the skew-normal distribution of Sahu *et al.* (2003) given by Arellano-Valle and Genton (2005). Logarithmic specification for estimation is also adopted because we believe it is based on a structural basis which is certainly plausible in many empirical settings of spatial interaction. However, our methodology is immediately applicable to the spatial disturbance specification without the need to perform transformation of data and we have more general spatial specifications too. Although the coefficients in the skew spatial panel models do not directly provide elasticity values of each variable, we can derive them with a simple computation.

The remainder of the paper is organized as follows. In Section 2, the generalized model with various spatial panel models as special cases of the model and the intended multivariate skew-normal (MSN) distribution will be briefly reviewed. In that section, we assume that the random error components of the generalized model follow the MSN distribution. After that, we show that the log-likelihood function of the observed data is of high dimension and does not have a closed form expression, and so the method of Bayesian hierarchical estimation is used. In Section 3, we specify the prior distributions for the unknown model parameters and obtain the joint posterior distributions of the parameters using the Markov chain Monte Carlo (MCMC) algorithm. In Section 4, based on several information criteria, we determine which model is more appropriate for our practical application. In Section 5, two applications of the proposed model are presented using a simulated data set and the real data set of cigarette consumption. Finally, Section 6 is dedicated to the concluding remarks.

2 The Generalized Model

In this section we review the generalized spatial panel data model as pointed out by Lee and Yu (2012), but with no serial correlation, which nest existing RE spatial panels as special cases. Let y_{it} denote the observation on the ith space for the tth time point, $i = 1, ..., N$, $t = 1, ..., T$. The model with spatial dependence and RE using Baltagi and Liu (2016) notation is given as follows:

$$\begin{aligned} \boldsymbol{y}_t &= \boldsymbol{X}_t\boldsymbol{\beta} + \boldsymbol{u}_t, \quad t = 1, ..., T, \\ \boldsymbol{u}_t &= \boldsymbol{u}_1 + \boldsymbol{u}_{2t}, \end{aligned} \tag{1}$$

with

$$\begin{aligned} \boldsymbol{u}_1 &= \rho_1 \boldsymbol{W}_1 \boldsymbol{u}_1 + (\boldsymbol{I}_N + \varrho_1 \boldsymbol{M}_1)\boldsymbol{\mu}, \\ \boldsymbol{u}_{2t} &= \rho_2 \boldsymbol{W}_2 \boldsymbol{u}_{2t} + (\boldsymbol{I}_N + \varrho_2 \boldsymbol{M}_2)\boldsymbol{\nu}_t, \end{aligned} \tag{2}$$

where $\boldsymbol{y}_t = (y_{1t}, ..., y_{Nt})'$ is an $(N \times 1)$ column vector of continues response values at time t, with N unique space units, $\boldsymbol{X}_t = [\boldsymbol{x}_{1t}, ..., \boldsymbol{x}_{Nt}]'$ denotes the $(N \times K)$ matrix of observations on the non-stochastic explanatory variables in all spaces at time t for a set of K exogenous variables corresponding to the fixed effects, and $\boldsymbol{\beta}$ denotes the parameter vector of $(K \times 1)$ regression coefficients. The disturbance component $\boldsymbol{u}_t$ follows an error term model which involves the sum of two disturbances. The $(N \times 1)$ vector of random error variables $\boldsymbol{u}_1 = (u_1, ..., u_N)'$ is assumed to be a time-invariant space-specific vector with no time subscript, while the $(N \times 1)$ vector of the remainder disturbance $\boldsymbol{u}_{2t} = (u_{12t}, ..., u_{N2t})'$ varies with time.

In spatial panel data models, the correlation structure between each pair of spatial units is specified by means of spatial weights matrices. In the above model, error components $\boldsymbol{u}_1$ and $\boldsymbol{u}_{2t}$ are allowed to incorporate spatial correlation with both autoregressive and moving average features. The matrices $\boldsymbol{W}_r$ (and $\boldsymbol{M}_r$); $r = 1, 2$ are $(N \times N)$ non-stochastic spatial weights matrices that generate the spatial dependence with zeros in the main diagonals and ones corresponding to neighbouring. As we assume that $\boldsymbol{W}_r$ (and $\boldsymbol{M}_r$) is non-normalized, thus ρ_r (and ϱ_r) is the scalar spatial autoregressive (and spatial moving average) coefficients with the compact subset of $|\rho_r| < 1/\kappa_{rmax}$ (and $\varrho_r < 1/\vartheta_{rmax}$) where κ_{rmax} (and ϑ_{rmax}) are the absolute values of the largest eigenvalues of $\boldsymbol{W}_r$ (and $\boldsymbol{M}_r$) which are assumed to be bounded away from zero by some fixed positive constant. To limit dependence, $\boldsymbol{W}_r$ (and $\boldsymbol{M}_r$) can also be normalized so that each row sums to one and ρ_r (and ϱ_r) are uniformly bounded in absolute value. Following Baltagi *et al.* (2012), $\boldsymbol{B}_1 = \boldsymbol{I}_N - \rho_1 \boldsymbol{W}_1$, $\boldsymbol{B}_2 = \boldsymbol{I}_N - \rho_2 \boldsymbol{W}_2$, $\boldsymbol{D}_1 = \boldsymbol{I}_N + \varrho_1 \boldsymbol{M}_1$ and $\boldsymbol{D}_2 = \boldsymbol{I}_N + \varrho_2 \boldsymbol{M}_2$ is defined. It follows that $\boldsymbol{u}_1$ and $\boldsymbol{u}_{2t}$ in equation (2) can be rewritten as

$$\begin{aligned} \boldsymbol{u}_1 &= \boldsymbol{A}_1^{-1} \boldsymbol{\mu}, \\ \boldsymbol{u}_{2t} &= \boldsymbol{A}_2^{-1} \boldsymbol{\nu}_t, \end{aligned} \tag{3}$$

where $\boldsymbol{A}_1 = \boldsymbol{D}_1^{-1} \boldsymbol{B}_1$, $\boldsymbol{A}_2 = \boldsymbol{D}_2^{-1} \boldsymbol{B}_2$ and they are assumed to be non-singular.

In this study, we assume that all four weight matrices are equal to $\boldsymbol{W}$, but with different spatial coefficients. It is commonly assumed that in the $(N \times 1)$ space-specific vector $\boldsymbol{\mu} = (\mu_1, ..., \mu_N)'$, the μ_i's are the unobserved space-specific effects. The common assumption of distributions for the vector $\boldsymbol{\mu}$ and $\boldsymbol{\nu_t}$ are $N_N(\boldsymbol{0}, \sigma_\mu^2 \boldsymbol{I}_N)$ and $N_N(\boldsymbol{0}, \sigma_\nu^2 \boldsymbol{I}_N)$ respectively and they are taken to be independent.

In spatial panel models, the data are ordered by time periods rather than space units, because this ordering for modelling spatial dependence is more convenient. Consequently, the extended form of the model (1-3) at all spaces and at all times is given by,

$$\begin{aligned} \boldsymbol{y} &= \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{u}, \\ \boldsymbol{u} &= (\boldsymbol{\iota}_T \otimes \boldsymbol{A}_1^{-1})\boldsymbol{\mu} + (\boldsymbol{I}_T \otimes \boldsymbol{A}_2^{-1})\boldsymbol{\nu}, \end{aligned} \tag{4}$$

where the response $\boldsymbol{y} = [\boldsymbol{y_1'}, ..., \boldsymbol{y_T'}]'$ is of dimension $(NT \times 1)$ and the explanatory matrix $\boldsymbol{X} = [\boldsymbol{X}_1', ..., \boldsymbol{X}_T']'$ is $(NT \times K)$. In the disturbance term, $\boldsymbol{\iota}_T$ and $\boldsymbol{I}_T$ denote a vector of ones of dimension T and an identity matrix of dimension T, respectively, and $\otimes$ denotes the Kronecker product.

This model nests various RE spatial panel models in the spatial econometrics literature, including the following:
(1) $\rho_1 = 0$ and $\varrho_1 = \varrho_2 = 0$, Anselin (1988)-random effects spatial autoregressive (RE-SAR) model.
(2) $\rho_1 = \rho_2$ and $\varrho_1 = \varrho_2 = 0$, Kapoor *et al.* (2007)-spatial autoregressive random effects (SAR-RE) model.
(3) $\varrho_1 = \varrho_2 = 0$, Baltagi *et al.* (2007)-generalized random effects spatial autoregressive (GRE-SAR) model.
(4) $\rho_1 = \rho_2 = 0$ and $\varrho_1 = 0$, Anselin *et al.* (2008)-random effects spatial moving average (RE-SMA) model.
(5) $\rho_1 = \rho_2 = 0$ and $\varrho_1 = \varrho_2$, Fingleton (2008)-spatial moving average random effects (SMA-RE) model.
In addition, it also includes
(6) Lee and Yu (2012) spatial autoregressive moving average random effects (SARMA-RE) model with $\rho_1 = \rho_2$ and $\varrho_1 = \varrho_1$.

In this paper, we develop a new spatial panel data model by replacing the multivariate normality assumption with the multivariate skew-normal distribution for $\boldsymbol{u}$ in equation (4). We consider the MSN distribution proposed by Sahu *et al.* (2003) and adopt a parametrization that has the advantage of being closed under marginalization and conditioning. In the next section, we will examine the stochastic representation of this skew-normal random vector as given in Arellano-Valle and Genton (2007).

2.1 The MSN Distribution of Sahu *et al.* (2003)

An n-dimensional random vector $\boldsymbol{Y}$ follows a n-variate skew-normal distribution with $(n \times 1)$ location parameter vector $\boldsymbol{m}$, $(n \times n)$ positive-definite dispersion matrix $\boldsymbol{\Sigma}$ and $(n \times n)$ skewness diagonal matrix $\boldsymbol{\Delta}$ if its density function is given by

$$f(\boldsymbol{y}|\boldsymbol{m},\boldsymbol{\Sigma},\boldsymbol{\Delta}) = 2^n \phi_n(\boldsymbol{y}|\boldsymbol{m},\boldsymbol{\Sigma}+\boldsymbol{\Delta}\boldsymbol{\Delta}^{'}) \times \Phi_n(\boldsymbol{\Delta}^{'}(\boldsymbol{\Sigma}+\boldsymbol{\Delta}\boldsymbol{\Delta}^{'})^{-1}(\boldsymbol{y}-\boldsymbol{m})|\boldsymbol{0},(\boldsymbol{I}_n+\boldsymbol{\Delta}^{'}\boldsymbol{\Sigma}^{-1}\boldsymbol{\Delta})^{-1}) \tag{5}$$

where $\phi_n(\boldsymbol{x}|\boldsymbol{m},\boldsymbol{\Sigma})$ denotes the probability density function (pdf) and $\Phi_n(\boldsymbol{x}|\boldsymbol{m},\boldsymbol{\Sigma})$ denotes the cumulative distribution function (cdf) of the n-variate normal distribution $N_n(\boldsymbol{m},\boldsymbol{\Sigma})$ with mean vector $\boldsymbol{m}$ and covariate matrix $\boldsymbol{\Sigma}$ evaluated at $\boldsymbol{x}$. Note that, if $\boldsymbol{m} = \boldsymbol{0}$ and $\boldsymbol{\Sigma} = \boldsymbol{I}_n$; $\phi_n(\boldsymbol{x}|\boldsymbol{m},\boldsymbol{\Sigma})$ and $\Phi_n(\boldsymbol{x}|\boldsymbol{m},\boldsymbol{\Sigma})$ will be denoted by $\phi_n(\boldsymbol{x})$ and $\Phi_n(\boldsymbol{x})$, respectively.

For the multivariate skew-normal distribution defined in (5), we note that $(\boldsymbol{I}_n + \boldsymbol{\Delta}^{'}\boldsymbol{\Sigma}^{-1}\boldsymbol{\Delta})^{-1} = \boldsymbol{I}_n - \boldsymbol{\Delta}^{'}(\boldsymbol{\Sigma} + \boldsymbol{\Delta}\boldsymbol{\Delta}^{'})^{-1}\boldsymbol{\Delta}$ and the matrices $\boldsymbol{\Delta}$ and $\boldsymbol{\Sigma}$ are diagonal. To simplify implementation, we let $\boldsymbol{\Sigma} = \sigma^2\boldsymbol{I}_n$ where $\sigma^2 > 0$ is a variance component. Also, we have that $\boldsymbol{\Delta} = \text{diag}(\delta_1, \delta_2, ..., \delta_n)$ in which $\boldsymbol{\delta} = (\delta_1, \delta_2, ..., \delta_n)^{'}$ is a vector of skewness parameters. In the sequel we assume $\delta_1 = \delta_2 = ... = \delta_n$ to simplify the fitting, and hence $\boldsymbol{\Delta} = \delta I_n$, where $\delta \in \boldsymbol{R}$ is a scalar skewness parameter. In our empirical analysis, we expect δ to be positive because the log transformation is the most popular transformation for positively skewed distribution in linear regression.

If $\boldsymbol{Y}$ has the density (5) we write $\boldsymbol{Y} \sim SN_n(\boldsymbol{m},\boldsymbol{\Sigma},\boldsymbol{\Delta})$. When $\boldsymbol{\Delta} = \boldsymbol{0}$, this MSN distribution reduces to the usual symmetric multivariate $N_n(\boldsymbol{m},\boldsymbol{\Sigma})$ density. Furthermore, $\boldsymbol{Y}$ has the following convenient stochastic representation

$$\boldsymbol{Y}{=}\boldsymbol{\Delta}|\boldsymbol{X}_0| + \boldsymbol{X}_1, \tag{6}$$

where $\boldsymbol{X}_0 \sim N_n(\boldsymbol{0},\boldsymbol{I}_n)$ and $\boldsymbol{X}_1 \sim N_n(\boldsymbol{m},\boldsymbol{\Sigma})$ are two independent random vectors. Note that $|\boldsymbol{X}_0|$ is the vector where its ith element is equal to the absolute value of the ith element of $\boldsymbol{X}_0$. This stochastic representation of $\boldsymbol{Y}$ is useful for random number generation and for our theoretical purposes.

Let $\boldsymbol{w} = |\boldsymbol{X}_0|$ and let $I\{\boldsymbol{w} > \boldsymbol{0}\}$ be an indicator function which is 1 if every element of $\boldsymbol{w}$ is greater than zero, and zero otherwise. Then $\boldsymbol{w}$ follows a $n \times 1$ standard normal distribution $N_n(\mathbf{0}, \boldsymbol{I}_n)$ that is truncated to the space $\boldsymbol{w} > 0$. Thus it follows that the hierarchical representation of (6) is given by

$$\begin{aligned} \boldsymbol{Y}|\boldsymbol{w} &\sim N_n(\boldsymbol{m} + \boldsymbol{\Delta}\boldsymbol{w}, \boldsymbol{\Sigma}), \\ \boldsymbol{w} &\sim N_n(\mathbf{0}, \boldsymbol{I}_n) I\{\boldsymbol{w} > 0\}. \end{aligned} \tag{7}$$

2.2 The Generalized Model with the MSN Distribution

Considerable attention has been given to relaxing the normality assumption for inference about the unknown parameters in linear mixed effects models. Arellano-Valle *et al.* (2005) assumed that both random effects and error terms are distributed as skew-normal distributions. Here, we suppose that $\boldsymbol{\mu}$ has independently and identically (iid) $SN_N(\mathbf{0}, \sigma_\mu^2 \boldsymbol{I}_N, \delta_\mu \boldsymbol{I}_N)$ distribution and $\boldsymbol{\nu}_t$ has iid $SN_N(\mathbf{0}, \sigma_\nu^2 \boldsymbol{I}_N, \delta_\nu \boldsymbol{I}_N)$ distribution assumed to be independent of each other. The problem of departure from normality will be supported using the above hierarchical model considerations.

We also suppose that the non-stochastic matrices $\boldsymbol{A}_1$ and $\boldsymbol{A}_2$ are orthogonal matrices. Thus, under the present assumptions and, as a direct consequence of the location-scale extension of the skew-normal distribution, Corollary 2.3 in Arellano-Valle and Genton (2005), we have

$$\boldsymbol{u}_1 = \boldsymbol{A}_1^{-1}\boldsymbol{\mu}|\sigma_\mu^2, \delta_\mu, \rho_1, \varrho_1 \sim SN_N(\mathbf{0}, (\boldsymbol{A}_1^{\prime}\boldsymbol{A}_1)^{-1}\sigma_\mu^2, \boldsymbol{A}_1^{-1}\delta_\mu). \tag{8}$$

Hence the response distribution at time point t is given by,

$$\boldsymbol{y}_t|\boldsymbol{\beta}, \boldsymbol{\mu}, \sigma_\nu^2, \delta_\nu, \rho_1, \rho_2, \varrho_1, \varrho_2 \sim SN_N(\boldsymbol{X}_t\boldsymbol{\beta} + \boldsymbol{A}_1^{-1}\boldsymbol{\mu}, (\boldsymbol{A}_2^{\prime}\boldsymbol{A}_2)^{-1}\sigma_\nu^2, \boldsymbol{A}_2^{-1}\delta_\nu). \tag{9}$$

Similarly, under the assumption that

$$\boldsymbol{u}_2 = (\boldsymbol{I}_T \otimes \boldsymbol{A}_2^{-1})\boldsymbol{\nu}|\sigma_\nu^2, \delta_\nu, \rho_2, \varrho_2 \sim SN_{NT}(\mathbf{0}, (\boldsymbol{I}_T \otimes (\boldsymbol{A}_2^{\prime}\boldsymbol{A}_2)^{-1})\sigma_\nu^2, (\boldsymbol{I}_T \otimes \boldsymbol{A}_2^{-1})\delta_\nu), \tag{10}$$

the response distribution at all time points is given by,

$$\boldsymbol{y}|\boldsymbol{\beta},\boldsymbol{\mu},\sigma^2_\nu,\delta_\nu,\rho_1,\rho_2,\varrho_1,\varrho_2 \sim SN_{NT}(\boldsymbol{X\beta}+(\boldsymbol{\iota}_T\otimes\boldsymbol{A}_1^{-1})\boldsymbol{\mu}, \\ (\boldsymbol{I}_T\otimes(\boldsymbol{A}_2'\boldsymbol{A}_2)^{-1})\sigma^2_\nu,(\boldsymbol{I}_T\otimes\boldsymbol{A}_2^{-1})\delta_\nu). \tag{11}$$

Hence, the conditional pdf of the response $\boldsymbol{y}$ is

$$f_Y(\boldsymbol{y}|\boldsymbol{\mu},\boldsymbol{\beta},\sigma^2_\nu,\delta_\nu,\rho_1,\rho_2,\varrho_1,\varrho_2) = \\ 2^{NT}\phi_{NT}(\boldsymbol{y}|\boldsymbol{X\beta}+(\boldsymbol{\iota}_T\otimes\boldsymbol{A}_1^{-1})\boldsymbol{\mu},(\boldsymbol{I}_T\otimes(\boldsymbol{A}_2'\boldsymbol{A}_2)^{-1})(\sigma^2_\nu+\delta^2_\nu)) \tag{12} \\ \times\Phi_{NT}(\tfrac{\delta_\nu}{(\sigma^2_\nu+\delta^2_\nu)}(\boldsymbol{I}_T\otimes\boldsymbol{A}_2^{-1})^{-1}(\boldsymbol{y}-\boldsymbol{X\beta}-(\boldsymbol{\iota}_T\otimes\boldsymbol{A}_1^{-1})\boldsymbol{\mu}|\boldsymbol{0},(\tfrac{\sigma^2_\nu}{\sigma^2_\nu+\delta^2_\nu})\boldsymbol{I}_{NT}).$$

It is of particular interest to make inference about the unknown collection of parameters $\theta=\{\boldsymbol{\beta}',\sigma^2_\nu,\sigma^2_\mu,\delta_\nu,\delta_\mu,\rho_1,\rho_2,\varrho_1,\varrho_2\}$ that are assumed to be independent of each another. Inference with this type of model has to be based on the marginal distribution for the response $\boldsymbol{y}$ unless the data are analysed in a Bayesian framework (Molenberghs and Verbeke 2000). By computing the following integral, the marginal density of $\boldsymbol{y}$ can be obtained,

$$f_Y(\boldsymbol{y}|\theta)=\textstyle\int_{R^N} f(\boldsymbol{y}|\boldsymbol{\mu},\boldsymbol{\beta},\sigma^2_\nu,\delta_\nu,\rho_2,\varrho_2)f(\boldsymbol{\mu}|\sigma^2_\mu,\delta_\mu,\rho_1,\varrho_1)d\boldsymbol{\mu} \\ =\textstyle\int_{R^N} 2^{N(T+1)}\phi_{NT}(\boldsymbol{y}|\boldsymbol{X\beta}+(\boldsymbol{\iota}_T\otimes\boldsymbol{A}_1^{-1})\boldsymbol{\mu},(\sigma^2_\nu+\delta^2_\nu)(\boldsymbol{I}_T\otimes(\boldsymbol{A}_2'\boldsymbol{A}_2)^{-1})) \\ \times\Phi_{NT}(\tfrac{\delta_\nu}{\sqrt{\sigma^2_\nu(\sigma^2_\nu+\delta^2_\nu)}}(\boldsymbol{I}_T\otimes\boldsymbol{A}_2^{-1})^{-1}(\boldsymbol{y}-\boldsymbol{X\beta}-(\boldsymbol{\iota}_T\otimes\boldsymbol{A}_1^{-1})\boldsymbol{\mu})|\boldsymbol{0},\boldsymbol{I}_{NT}) \\ \times\phi_N(\boldsymbol{\mu}|\boldsymbol{0},(\sigma^2_\mu+\delta^2_\mu)\boldsymbol{I}_N)\times\Phi_N(\tfrac{\delta_\mu}{(\sigma^2_\mu+\delta^2_\mu)}\boldsymbol{\mu}|\boldsymbol{0},(\tfrac{\sigma^2_\mu}{\sigma^2_\mu+\delta^2_\mu})\boldsymbol{I}_N)d\boldsymbol{\mu}.$$

In the above, we let

$$\phi_{NT}(\boldsymbol{y}|\boldsymbol{X\beta}+(\boldsymbol{\iota}_T\otimes\boldsymbol{A}_1^{-1})\boldsymbol{\mu},(\sigma^2_\nu+\delta^2_\nu)(\boldsymbol{I}_T\otimes(\boldsymbol{A}_2'\boldsymbol{A}_2)^{-1}))\times\phi_N(\boldsymbol{\mu}|\boldsymbol{0},(\sigma^2_\mu+\delta^2_\mu)\boldsymbol{I}_N) \\ =\phi_{NT}(\boldsymbol{y}|\boldsymbol{X\beta},\boldsymbol{\Sigma})\phi_N(\boldsymbol{\mu}|\boldsymbol{\mu_1},\boldsymbol{\Lambda}), \tag{14}$$

where

$$\boldsymbol{\Sigma} = (\sigma_\nu^2 + \delta_\nu^2)(\boldsymbol{I}_T \otimes (\boldsymbol{A}_2^{'}\boldsymbol{A}_2)^{-1}) + (\sigma_\mu^2 + \delta_\mu^2)(\boldsymbol{J}_T \otimes (\boldsymbol{A}_1^{'}\boldsymbol{A}_1)^{-1}), \tag{15}$$

$$\boldsymbol{\mu}_1 = (\sigma_\nu^2 + \delta_\nu^2)^{-1}\boldsymbol{\Lambda}(\boldsymbol{A}_2^{'}\boldsymbol{A}_2)^{-1'}(\boldsymbol{I}_T \otimes (\boldsymbol{A}_2^{'}\boldsymbol{A}_2)^{-1})^{-1}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}), \tag{16}$$

and

$$\boldsymbol{\Lambda} = ((\sigma_\mu^2 + \delta_\mu^2)^{-1}\boldsymbol{I}_N + (\sigma_\nu^2 + \delta_\nu^2)^{-1}(\boldsymbol{\iota}_T \otimes \boldsymbol{A}_1^{-1})'(\boldsymbol{I}_T \otimes (\boldsymbol{A}_2^{'}\boldsymbol{A}_2)^{-1})^{-1}(\boldsymbol{\iota}_T \otimes \boldsymbol{A}_1^{-1}). \tag{17}$$

Also, we let

$$\begin{aligned}
&\Phi_{NT}(\tfrac{\delta_\nu}{\sqrt{\sigma_\nu^2(\sigma_\nu^2+\delta_\nu^2)}}(\boldsymbol{I}_T \otimes \boldsymbol{A}_2^{-1})^{-1}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta} - (\boldsymbol{\iota}_T \otimes \boldsymbol{A}_1^{-1})\boldsymbol{\mu})|\boldsymbol{0}, \boldsymbol{I}_{NT}) \\
&\times\Phi_N(\tfrac{\delta_\mu}{(\sigma_\mu^2+\delta_\mu^2)}\boldsymbol{\mu}|\boldsymbol{0}, \tfrac{\sigma_\mu^2}{(\sigma_\mu^2+\delta_\mu^2)}\boldsymbol{I}_N) \\
&= \Phi_{N(T+1)}(-\boldsymbol{\Gamma}\boldsymbol{\mu}| - \boldsymbol{\mu}_2, \boldsymbol{R}),
\end{aligned} \tag{18}$$

where

$$\boldsymbol{\mu}_2 = \begin{pmatrix} \frac{\delta_\nu}{\sqrt{\sigma_\nu^2(\sigma_\nu^2+\delta_\nu^2)}}(\boldsymbol{I}_T \otimes \boldsymbol{A}_2^{-1})^{-1}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}) \\ \boldsymbol{0}_N \end{pmatrix}, \tag{19}$$

$$\boldsymbol{\Gamma} = \begin{pmatrix} \frac{\delta_\nu}{\sqrt{\sigma_\nu^2(\sigma_\nu^2+\delta_\nu^2)}}(\boldsymbol{I}_T \otimes \boldsymbol{A}_2^{-1})^{-1}(\boldsymbol{\iota}_T \otimes \boldsymbol{A}_1^{-1}) \\ -\frac{\delta_\mu}{\sqrt{\sigma_\mu^2(\sigma_\mu^2+\delta_\mu^2)}}\boldsymbol{I}_N \end{pmatrix}, \tag{20}$$

$$\boldsymbol{R} = \begin{pmatrix} \boldsymbol{I}_{NT} & \mathbf{0} \\ \mathbf{0} & \frac{\sigma_\mu^2}{(\sigma_\mu^2+\delta_\mu^2)}\boldsymbol{I}_N \end{pmatrix}. \tag{21}$$

Lastly, to find the marginal density of $\boldsymbol{y}$, we make use of equations (14) and (18) to obtain

$$\begin{aligned} f_{\boldsymbol{Y}}(\boldsymbol{y}|\theta) &= \int_{R^N} 2^{N(T+1)}\phi_{NT}(\boldsymbol{y}|\boldsymbol{X\beta},\boldsymbol{\Sigma})\phi_N(\boldsymbol{\mu}|\boldsymbol{\mu}_1,\boldsymbol{\Lambda})\Phi_{N(T+1)}(-\boldsymbol{\Gamma\mu}|-\boldsymbol{\mu}_2,\boldsymbol{R})d\boldsymbol{\mu} \\ &= 2^{N(T+1)}\phi_{NT}(\boldsymbol{y}|\boldsymbol{X\beta},\boldsymbol{\Sigma})\int_{R^N}\phi_N(\boldsymbol{\mu}|\boldsymbol{\mu}_1,\boldsymbol{\Lambda})\Phi_{N(T+1)}(-\boldsymbol{\Gamma\mu}|-\boldsymbol{\mu}_2,\boldsymbol{R})d\boldsymbol{\mu} \\ &= 2^{N(T+1)}\phi_{NT}(\boldsymbol{y}|\boldsymbol{X\beta},\boldsymbol{\Sigma})\Phi_{N(T+1)}(\boldsymbol{\mu}_2|\boldsymbol{\Gamma\mu},\boldsymbol{R}+\boldsymbol{\Gamma\Lambda\Gamma}'). \end{aligned} \tag{22}$$

The proof of the above results is given in the Appendix of the Arellano-Valle *et al.* (2007). The complete log-likelihood function for θ given the response sample $\boldsymbol{y}$ can be written as

$$\iota(\theta|\boldsymbol{y}) \propto -\frac{1}{2}log|\boldsymbol{\Sigma}|-\frac{1}{2}(\boldsymbol{y}-\boldsymbol{X\beta})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{y}-\boldsymbol{X\beta})+log\Phi_{N(T+1)}(\boldsymbol{\mu}_2-\boldsymbol{\Gamma\mu}|\mathbf{0},\boldsymbol{R}+\boldsymbol{\Gamma\Lambda\Gamma}'). \tag{23}$$

As this likelihood function is of high dimension and has no closed form, the estimation of the parameters is complex and we cannot directly calculate the posterior distribution of θ based on the observed data. Therefore, as an alternative, we apply a Bayesian computational method via the MCMC procedure. Thus, from (23) we can sample using the Gibbs sampler and infer the marginal posterior distribution of interest. The Gibbs sampler works by drawing samples iteratively from the conditional posterior distributions.

3 Specifying prior and posterior distributions for Bayesian inference

In order to determine the posterior distributions of our model parameters through Bayesian inference implementation, we need to specify the prior distribution on all the unknown parameters in θ. As we do not have prior information about historical or previous experiments data, we proceed by assigning conjugate but weakly informative prior distributions to obtain well-defined and proper posteriors for our parameters. A popular choice to ensure posterior propriety in linear models with RE is to consider proper conditionally conjugate priors like non-informative distributions, that is, prior distributions with large variances. Here, we use normal priors for the fixed-effects, inverse gamma priors for variance parameters and truncated-normal for skewness parameters, as suggested in Arellano-Valle *et al.* (2007). The prior distributions were adopted with their density functions for the $\boldsymbol{\beta}$, σ_ν^2, σ_μ^2, δ_ν, δ_μ parameters given as:

- $\pi(\boldsymbol{\beta}|\boldsymbol{\beta}_0, \boldsymbol{\Gamma}_{\boldsymbol{\beta}_0}) \propto exp\{-\frac{1}{2}(\boldsymbol{\beta}-\boldsymbol{\beta}_0)'\boldsymbol{\Gamma}_{\boldsymbol{\beta}_0}^{-1}(\boldsymbol{\beta}-\boldsymbol{\beta}_0)\}$,
- $\pi(\sigma_\nu^2|\sigma_{\nu_0}^2, \gamma_{\nu 0}) = \frac{(\gamma_{\nu_0})^{\sigma_{\nu_0}^2}}{\Gamma(\sigma_{\nu_0}^2)}(\frac{1}{\sigma_\nu^2})^{\sigma_{\nu_0}^2+1} exp\{-\frac{\gamma_{\nu_0}}{\sigma_\nu^2}\}$,
- $\pi(\sigma_\mu^2|\sigma_{\mu_0}^2, \gamma_{\mu_0}) = \frac{(\gamma_{\mu_0})^{\sigma_{\mu_0}^2}}{\Gamma(\sigma_{\mu_0}^2)}(\frac{1}{\sigma_\mu^2})^{\sigma_{\mu_0}^2+1} exp\{-\frac{\gamma_{\mu_0}}{\sigma_\mu^2}\}$,
- $\pi(\delta_\nu|\mu_\nu, \gamma_\nu^2) \propto exp\{-\frac{1}{2\gamma_\nu^2}(\delta_\nu-\mu_\nu)^2\}I\{\delta_\nu > 0\}$,
- $\pi(\delta_\mu|\mu_\mu, \gamma_\mu^2) \propto exp\{-\frac{1}{2\gamma_\mu^2}(\delta_\mu-\mu_\mu)^2\}I\{\delta_\mu > 0\}$.

In addition to the specification of these parameters, the specification of a subjective prior to either the scalar spatial autoregressive coefficients ρ_1, ρ_2, and scalar spatial moving average coefficients ϱ_1 and ϱ_2 is difficult without previous data. It seems that a uniform prior which assigns equal weight to all values of the spatial parameters is not unreasonable, as most of the spatial dependence models are based on real data sets (Zheng *et al.* 2008 and Baltagi and Liu 2016). When $\boldsymbol{W}$ is non-normalized, we assume that λ_{max} is the absolute of the largest eigenvalues of $\boldsymbol{W}$ and define a prior distribution for the parameters space $\rho_r \sim U(-1/\lambda_{max}, 1/\lambda_{max})$ and $\varrho_r \sim U(-1/\lambda_{max}, 1/\lambda_{max})$; $r = 1, 2$. For the normalized case, we define $\rho_r \sim U(-1, 1)$ and $\varrho_r \sim U(-1, 1)$; $r = 1, 2$ with $U(a, b)$ denoting the uniform distribution between a and b. We can make statistical inference for the parameters with these specifications and equation (23), based on the posterior distributions of the parameters under the hierarchical Bayesian approach. To achieve this, the joint posterior distribution of all parameters in θ is proportional to the observed data and the prior distributions as follows:

$$\pi(\boldsymbol{\beta}, \boldsymbol{\mu}, \sigma_\nu^2, \sigma_\mu^2, \delta_\nu, \delta_\mu, \rho_1, \rho_2, \varrho_1, \varrho_2|\boldsymbol{y}) \propto$$

$$\begin{aligned}
&\phi_{NT}(\boldsymbol{y}|\boldsymbol{X\beta} + (\boldsymbol{\iota}_T \otimes \boldsymbol{A}_1^{-1})\boldsymbol{\mu}, (\sigma_\nu^2 + \delta_\nu^2)(\boldsymbol{I}_T \otimes (\boldsymbol{A}_2'\boldsymbol{A}_2)^{-1}))\\
&\times \Phi_{NT}(\tfrac{\delta_\nu}{\sqrt{\sigma_\nu^2(\sigma_\nu^2+\delta_\nu^2)}}(\boldsymbol{I}_T \otimes \boldsymbol{A}_2^{-1})^{-1}(\boldsymbol{y} - \boldsymbol{X\beta} - (\boldsymbol{\iota}_T \otimes \boldsymbol{A}_1^{-1})\boldsymbol{\mu})|\boldsymbol{0}, \boldsymbol{I}_{NT})\\
&\times \phi_N(\boldsymbol{\mu}|\boldsymbol{0}, (\sigma_\mu^2 + \delta_\mu^2)\boldsymbol{I}_N) \times \Phi_N(\tfrac{\delta_\mu}{(\sigma_\mu^2+\delta_\mu^2)}\boldsymbol{\mu}|\boldsymbol{0}, (\tfrac{\sigma_\mu^2}{\sigma_\mu^2+\delta_\mu^2})\boldsymbol{I}_N)\\
&\times exp\{-\tfrac{1}{2}(\boldsymbol{\beta} - \boldsymbol{\beta}_0)'\boldsymbol{\Gamma}_{\beta_0}^{-1}(\boldsymbol{\beta} - \boldsymbol{\beta}_0)\}\\
&\times exp\{-\tfrac{1}{2\gamma_\nu^2}(\delta_\nu - \mu_\nu)^2\}I\{\delta_\nu > 0\} \times exp\{-\tfrac{1}{2\gamma_\mu^2}(\delta_\mu - \mu_\mu)^2\}I\{\delta_\mu > 0\}\\
&\times \tfrac{(\gamma_{\mu_0})^{\sigma_{\mu_0}^2}}{\Gamma(\sigma_{\mu_0}^2)}(\tfrac{1}{\sigma_\mu^2})^{\sigma_{\mu_0}^2+1} exp\{-\tfrac{\gamma_{\mu_0}}{\sigma_\mu^2}\} \times \tfrac{(\gamma_{\nu_0})^{\sigma_{\nu_0}^2}}{\Gamma(\sigma_{\nu_0}^2)}(\tfrac{1}{\sigma_\nu^2})^{\sigma_{\nu_0}^2+1} exp\{-\tfrac{\gamma_{\nu_0}}{\sigma_\nu^2}\}.
\end{aligned} \tag{24}$$

The complicated form of distribution (24) and the difficulty to obtain analytically the marginal distributions of each unknown stochastic parameter in θ have motivated us to implement an MCMC algorithm. The MCMC algorithm draws samples from the posterior distributions of all unknown stochastic parameters using the Gibbs sampler. The posterior estimates of ρ_1, ρ_2, ϱ_1, and ϱ_2 are obvious. The posterior distributions of the other parameters can be shown after some algebra in the next section.

3.1 The MCMC Algorithm

In this section, the MCMC computations for the class of skew-normal linear mixed effects models that incorporate spatial autoregressive and moving average process will be considered by the following spatial hierarchical Bayes model. The hierarchical representation given in equation (7) is important to obtain the parameter estimates and will allow us to easily implement the methods using the available R under OpenBUGS (Lunn *et al.* 2009) software to write BUGS codes. When the values of ρ_r and ϱ_r, $r = 1, 2$ coefficients are treated as constants, and by introducing two $(N \times 1)$ random vector $\boldsymbol{w}_{\nu_t}$ and $\boldsymbol{w}_{\mu}$, the hierarchical representation of the model is formulated as:

$$\begin{gathered}\boldsymbol{y}_t|\boldsymbol{\mu},\boldsymbol{\beta},\sigma^2_{\nu},\delta_{\nu},\rho_1,\rho_2,\varrho_1,\varrho_2,\boldsymbol{w}_{\nu_t} \sim N_N(\boldsymbol{X}_t\boldsymbol{\beta}+\boldsymbol{A}_1^{-1}\boldsymbol{\mu}+\delta_{\nu}\boldsymbol{A}_2^{-1}\boldsymbol{w}_{\nu_t},\sigma^2_{\nu}(\boldsymbol{A}_2'\boldsymbol{A}_2)^{-1}),\\ \boldsymbol{w}_{\nu_t} \sim N_N(\boldsymbol{0},\boldsymbol{I}_N)I\{\boldsymbol{w}_{\nu_t}>\boldsymbol{0}\}, t=1,...,T\end{gathered} \tag{25}$$

$$\begin{gathered}\boldsymbol{\mu}|\sigma^2_{\mu},\delta_{\mu},\boldsymbol{w}_{\mu} \sim N_N(\delta_{\mu}\boldsymbol{w}_{\mu},\sigma^2_{\mu}\boldsymbol{I}_N),\\ \boldsymbol{w}_{\mu} \sim N_N(\boldsymbol{0},\boldsymbol{I}_N)I\{\boldsymbol{w}_{\mu}>\boldsymbol{0}\},\end{gathered} \tag{26}$$

where $\boldsymbol{w}_k = |\boldsymbol{\xi}|$ with $\boldsymbol{\xi} \sim N_k(\boldsymbol{0},\boldsymbol{I}_k)$.

The complete-conditional posterior distributions of parameters are required in order to apply the Gibbs sampling methodology to simulate samples from the distributions iteratively. Here, we will present the complete-conditional posterior distribution of the parameters $\boldsymbol{\beta}$, $\boldsymbol{\mu}$, $\boldsymbol{w}_{\nu_t}$, $\boldsymbol{w}_{\mu}$, σ^2_{ν}, σ^2_{μ}, δ_{ν}, δ_{μ}, while omitting the detailed derivation. For the regression coefficient vector $\boldsymbol{\beta}$, we obtain

$$\boldsymbol{\beta}|\boldsymbol{\mu},\sigma^2_{\nu},\delta_{\nu},\rho_1,\rho_2,\varrho_1,\varrho_2,\boldsymbol{w}_{\nu_t} \sim N_k(\boldsymbol{m}_{\beta},\boldsymbol{\Sigma}_{\beta}), \tag{27}$$

where the mean is given by $\boldsymbol{m}_{\beta} = \boldsymbol{\Sigma}_{\beta}(\boldsymbol{\Gamma}_{\beta_0}^{-1}\boldsymbol{\beta}_0 + \frac{1}{\sigma^2_{\nu}}\sum_{t=1}^{T}\boldsymbol{X}_t'(\boldsymbol{A}_2'\boldsymbol{A}_2)\boldsymbol{d}_{t\beta})$, $\boldsymbol{d}_{t\beta} = \boldsymbol{y}_t - \boldsymbol{A}_1^{-1}\boldsymbol{\mu} - \delta_{\nu}\boldsymbol{A}_2^{-1}\boldsymbol{w}_{\nu_t}$, and the covariance matrix is given by $\boldsymbol{\Sigma}_{\beta} = (\boldsymbol{\Gamma}_{\beta_0}^{-1} + \frac{1}{\sigma^2_{\nu}}\sum_{t=1}^{T}\boldsymbol{X}_t'(\boldsymbol{A}_2'\boldsymbol{A}_2)$
$\boldsymbol{X}_t)^{-1}$. For the vector $\boldsymbol{\mu}$, we have

$$\boldsymbol{\mu}|\boldsymbol{\beta},\sigma^2_{\nu},\sigma^2_{\mu},\delta_{\nu},\delta_{\mu},\boldsymbol{w}_{\nu_t},\boldsymbol{w}_{\mu},\rho_1,\rho_2,\varrho_1,\varrho_2, \sim N_N(\boldsymbol{m}_{\mu},\boldsymbol{\Sigma}_{\mu}), \tag{28}$$

where $\boldsymbol{m}_{\mu} = \boldsymbol{\Sigma}_{\mu}(\frac{1}{\sigma^2_{\nu}}\boldsymbol{A}_1'^{-1}(\boldsymbol{A}_2'\boldsymbol{A}_2)\sum_{t=1}^{T}\boldsymbol{d}_{t\mu} + \frac{1}{\sigma^2_{\mu}}\delta_{\mu}\boldsymbol{w}_{\mu})$, $\boldsymbol{d}_{t\mu} = \boldsymbol{y}_t - \boldsymbol{X}_t\boldsymbol{\beta} - \delta_{\nu}\boldsymbol{A}_2^{-1}\boldsymbol{w}_{\nu_t}$, and $\boldsymbol{\Sigma}_{\mu} = (\frac{T}{\sigma^2_{\epsilon}}\boldsymbol{A}_1'^{-1}(\boldsymbol{A}_2'\boldsymbol{A}_2)\boldsymbol{A}_1^{-1} + \frac{1}{\sigma^2_{\mu}}\boldsymbol{I}_N)^{-1}$. It can be shown that the complete-conditional distribution of $\boldsymbol{w}_{\nu_t}$ and $\boldsymbol{w}_{\mu}$ is given, respectively, by

$$\boldsymbol{w}_{\nu_t}|\boldsymbol{\beta},\boldsymbol{\mu},\sigma^2_{\nu},\delta_{\nu},\rho_1,\rho_2,\varrho_1,\varrho_2 \sim N_N(\boldsymbol{m}_{w_{\nu_t}},\boldsymbol{\Sigma}_{w_{\nu_t}})I\{\boldsymbol{w}_{\nu_t}>\boldsymbol{0}\} \tag{29}$$

and

$$\boldsymbol{w}_\mu|\boldsymbol{\mu},\sigma_\mu^2,\delta_\mu,\rho_2,\varrho_2 \sim N_N(\boldsymbol{m}_{w_\mu},\boldsymbol{\Sigma}_{w_\mu})I\{\boldsymbol{w}_\mu > \boldsymbol{0}\}, \tag{30}$$

where $\boldsymbol{\mu}_{w_{\nu_t}} = \boldsymbol{\Sigma}_{w_{\nu_t}}(\frac{\delta_\nu}{\sigma_\nu^2}\boldsymbol{A}_2\sum_{t=1}^T \boldsymbol{d}_{w_{\nu_t}})$, $\boldsymbol{d}_{w_{\nu_t}} = \boldsymbol{y}_t - \boldsymbol{X}_t\boldsymbol{\beta} - \boldsymbol{A}_1^{-1}\boldsymbol{\mu}$, $\boldsymbol{\Sigma}_{w_{\nu_t}} = \sigma_\nu^2(\sigma_\nu^2 + \delta_\nu^2)^{-1}\boldsymbol{I}_N$, $\boldsymbol{m}_{w_\mu} = \boldsymbol{\Sigma}_{w_\mu}(\frac{\delta_\mu}{\sigma_\mu^2}\boldsymbol{\mu})$ and $\boldsymbol{\Sigma}_{w_\mu} = \sigma_\mu^2(\delta_\mu^2 + \delta_\mu^2)^{-1}\boldsymbol{I}_N$. It follows that the complete-conditional distribution of σ_ν^2 and σ_μ^2 is given by

$$\sigma_\nu^2|\delta_\nu,\boldsymbol{w}_\nu,\rho_1,\rho_2,\varrho_1,\varrho_2 \sim IG(\tau_{\sigma_\nu^2},\gamma_{\sigma_\nu^2}) \tag{31}$$

and

$$\sigma_\mu^2|\boldsymbol{\mu},\delta_\mu,\boldsymbol{w}_\mu,\rho_1,\rho_2,\varrho_1,\varrho_2 \sim IG(\tau_{\sigma_\mu^2},\gamma_{\sigma_\mu^2}), \tag{32}$$

respectively, where $\tau_{\sigma_\nu^2} = \frac{T}{2} + \sigma_{\nu_0}^2$, $\gamma_{\sigma_\nu^2} = \frac{1}{2}\sum_{t=1}^T \boldsymbol{d}_{t\sigma_\nu^2}^{'}(\boldsymbol{A}_2^{'}\boldsymbol{A}_2)\boldsymbol{d}_{t\sigma_\nu^2} + \gamma_{\nu_0}^{-1}$, $\boldsymbol{d}_{t\sigma_\nu^2} = \boldsymbol{y}_t - \boldsymbol{X}_t\boldsymbol{\beta} - \boldsymbol{A}_1^{-1}\boldsymbol{\mu} - \delta_\nu\boldsymbol{A}_2^{-1}\boldsymbol{w}_{\nu_t}$, $\tau_{\sigma_\mu^2} = \frac{N}{2} + \sigma_{\mu_0}^2$, and $\gamma_{\sigma_\mu^2} = \frac{1}{2}(\boldsymbol{\mu} - \delta_\mu\boldsymbol{w}_\mu)^{'}(\boldsymbol{\mu} - \delta_\mu\boldsymbol{w}_\mu) + \gamma_{\mu_0}$. Also, for the skewness parameters δ_ν and δ_μ, we have

$$\delta_\nu|\boldsymbol{\beta},\boldsymbol{\mu},\sigma_\nu^2,\boldsymbol{w}_{\nu t},\rho_1,\rho_2,\varrho_1,\varrho_2 \sim N(m_{\delta_\nu},\eta_{\delta_\nu})I\{\delta_\nu > 0\} \tag{33}$$

and

$$\delta_\mu|\boldsymbol{\mu},\sigma_\mu^2,\boldsymbol{w}_\mu,\rho_1,\rho_2,\varrho_1,\varrho_2 \sim N(m_{\delta_\mu},\eta_{\delta_\mu})I\{\delta_\mu > 0\}, \tag{34}$$

where $m_{\delta_\nu} = \eta_{\delta_\nu}(\frac{\mu_\nu}{\gamma_\nu^2} + \frac{1}{\sigma_\nu^2}\sum_{t=1}^T \boldsymbol{w}_{\nu_t}^{'}\boldsymbol{A}_2\boldsymbol{d}_{w_{\nu_t}})$, $\eta_{\delta_\nu} = (\frac{1}{\gamma_\nu^2} + \frac{1}{\sigma_\nu^2}\sum_{t=1}^T \boldsymbol{w}_{\nu_t}^{'}\boldsymbol{w}_{\nu_t})^{-1}$, $m_{\delta_\mu} = \eta_{\delta_\mu}(\frac{\mu_\mu}{\gamma_\mu^2} + \frac{1}{\sigma_\mu^2}\boldsymbol{w}_\mu^{'}\boldsymbol{\mu})$, and $\eta_{\delta_\mu} = (\frac{1}{\gamma_\mu} + \frac{1}{\sigma_\mu^2}\boldsymbol{w}_\mu^{'}\boldsymbol{\mu})^{-1}$.

4 Model Selection

Progress in the computation of Bayesian posterior distributions of model parameters based on a MCMC algorithm has made it possible to fit increasingly complex statistical models to determine the best model fitting. In this paper, 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 skew-normal model and compare it to normal model. They are defined by AIC= $\widehat{D} + 2P$, BIC= $\widehat{D} + Plog(N)$ and HQC= $\widehat{D} + 2Plog(log(N))$, respectively, where $\widehat{D}$ is -2 times log-likelihood evaluated at the maximum likelihood estimate, P is the number of parameters and N is the sample size. Models with smaller AIC, BIC, and HQC are preferred when comparing different fitted results.

5 Application

In this section, we fit and compare the performances of each spatial panel data model in the generalized class with the MSN and normal distributions through simulation and analysis in our motivating application on cigarette consumption, to demonstrate the usefulness of our methodology in spatial econometrics field. The results were analyzed through the R and packages R2OpenBUGS and BRugs (Ligges *et al.* 2017 and Ligges 2017) with OpenBUGS software.

5.1 Analysis of simulated data

This section presents a simulation experiment to evaluate the performance of the skew-normal linear mixed model with spatial error dependence. The data generating process started with a simple panel data regression with random error components disturbances as follows:

$$y_{it} = 0.5x_{it1} + 7x_{it2} + u_{it} \quad i = 1, ..., N; t = 1, ..., T. \tag{35}$$

The explanatory variables x_{it1} and x_{it2} are generated as in Nerlove (1971) with $x_{ith} = 0.1t + 0.5x_{it-1h} + z_{ith}$, where z_{ith} is a random variable distributed on $U(-0.5, 0.5)$ and $x_{i0h} = 5 + 10z_{i0h}$ for $h = 1, 2$. Following Baltagi and Liu (2016), we select the same spatial weight matrix $\boldsymbol{W}$ for equation (2). The matrix $\boldsymbol{W}$ is created such that its ith row has non-zero elements in positions $i + 5$ and $i - 5$. Therefore, the ith element of $\boldsymbol{W}$ is directly related to the five elements immediately before it and the five elements immediately after it. This matrix is defined in a circular world so that the non-zero elements in rows 1 and N are in positions $(2, 3, 4, 5, 6, N-4, N-3, N-2, N-1, N)$ and $(1, 2, 3, 4, 5, N-5, N-4, N-3, N-2, N-1)$, respectively. This matrix is row-normalized so that all of its non-zero elements are equal to 1/10. The disturbance term $\boldsymbol{u}_t$ were generated as a spatially correlated process with the following data generating process:

1. SAR: $\varrho_1 = \varrho_2 = 0$, ρ_1 and ρ_2 take values from $U(-1, 1)$.
2. SMA: $\rho_1 = \rho_2 = 0$, ϱ_1 and ϱ_2 take values from $U(-1, 1)$.
3. SARMA: ϱ_1, ϱ_2, ρ_1 and ρ_2 take values from $U(-1, 1)$.

The data generating process is taken to be a simple error component regression model with spatial disturbances of the autoregressive, moving average or spatial autoregressive moving average type, where the space-specific effect μ_i is a random variable uniformly distributed as $\mu_i \sim SN(0, 4, 5)$ and the random variable ν_{it} uniformly distributed as $\nu_{it} \sim SN(0, 1, 10)$. Since the skew parameters are positive, there is a right skewness in the density of the disturbance terms. The difference in shape between the implied skew-normal distribution and the normal distribution may appear rather minimal, but the difference is enough to make substantial differences in the impacts of specific policy actions. In particular, using the normal distribution when the skew-normal is the appropriate distribution leads to an underestimation (Bhat *et al.* 2017).

The sample sizes were chosen as $N = 25$ and $T = 5$. For each experiment, we perform 100 replications. The model specified by equation (35) was fitted to this data set via the following prior specifications:

$$\beta_0 \sim N_1(0, 10^2),\ \ \beta_1 \sim N_1(0, 10^2),\ \sigma^2_\mu \sim IG(10^{-2}, 10^{-2}),\ \sigma^2_\nu \sim IG(10^{-2}, 10^{-2}),$$
$$\delta_\mu \sim N_1(0, 10^2) I\{\delta_\mu > 0\},\ \delta_\nu \sim N_1(0, 10^2) I\{\delta_\nu > 0\}$$

Note that these prior distributions are considered as close to non-informative prior distributions (with large variances). Thus, we expect the results to be somewhat robust with respect to prior distributions. We ran the algorithm described in section 3.1 with 10,000 iterations and discarded the 3000 initial iterates. Concerning the Gibbs sampling result, the skew models are preferred over the normal models for this data set according to the recorded criteria AIC, BIC, and HQC.

After collecting the final MCMC samples, summaries of the Monte Carlo (MC) results given in Tables 1 and 2. The posterior mean average (MC mean), standard deviation average (MC SD) and average of the 95% credible intervals (MC CIs) of MC results are estimated for β_1, β_2, and skewness parameters, where skew-normal and normality assumptions were considered. The estimated parameters can be considerably different between the models with these assumptions. It shows that the MC mean results over the 100 data sets of skew models are concentrated around the true values. The estimated 95% credible intervals of skewness parameters do not include zero, confirming that consideration of a departure from normality may improve the model fit. We focus on the regression coefficients, because inference about these parameters is important. The data set was generated by assuming $(\beta_1, \beta_2) = (0.5, 7)$. The parameters estimated have the smaller MC SD with skew models related that obtained for normal models.

The apparent conclusion is that the price to be paid for estimating the skew density is worthwhile. In other words, these findings suggest that it is important to assume a model with a skew distribution in order to achieve reasonable results when the data exhibit non-normal characteristics. As in the simulation here, it can lead to more accurate estimates for regression coefficients.

Table 1: mean (Mc SD) and MC CIs for the parameters of spatial panel data models: SAR

spatial dependence type	parameter	MC mean(MC SD)	MC CIs
RE-SAR : $\rho_1 = 0, \rho_2 \sim U(-1,1)$			
Skew-Normal			
	β_1	0.592 (0.593)	(-0.582,1.741)
	β_2	7.783 (0.591)	(6.633,8.946)
	δ_μ	6.488 (1.804)	(3.566,10.388)
	δ_ν	7.125 (0.973)	(5.272,9.034)
Normal			
	β_1	1.583 (0.721)	(0.186,3.018)
	β_2	8.780 (0.772)	(7.291,10.348)
SAR-RE : $\rho_1 = \rho_2 \sim U(-1,1)$			
Skew-Normal			
	β_1	1.508 (0.530)	(0.465,2.540)
	β_2	9.389 (0.523)	(8.375,10.40)
	δ_μ	11.16 (2.037)	(7.618,15.58)
	δ_ν	11.08 (1.014)	(9.049,13.01)
Normal			
	β_1	2.571 (0.844)	(0.961,4.315)
	β_2	10.48 (0.933)	(8.706,12.39)
GRE-SAR: $\rho_1 \sim U(-1,1), \rho_2 \sim U(-1,1)$			
Skew-Normal			
	β_1	0.818 (0.593)	(-0.348,1.959)
	β_2	7.891 (0.594)	(6.725,9.044)
	δ_μ	6.845 (1.667)	(3.951,10.39)
	δ_ν	7.872 (1.017)	(5.912,9.874)
Normal			
	β_1	1.753 (0.740)	(0.300,3.216)
	β_2	8.768 (0.763)	(7.298,10.30)

Table 2: MC mean (MC SD) and MC CIs for the parameters of spatial panel data models: SMA and SARMA

spatial dependence type	parameter	MC mean(MC SD)	MC CIs
RE-SMA : $\varrho_1 = 0, \varrho_2 \sim U(-1,1)$			
Skew-Normal			
	β_1	1.471 (0.806)	(-0.111,3.041)
	β_2	7.275 (0.833)	(6.638,15.89)
	δ_μ	9.321 (2.498)	(4.998,14.37)
	δ_ν	7.801 (2.057)	(4.159,11.86)
Normal			
	β_1	2.564 (0.860)	(0.880,4.295)
	β_2	9.465 (0.921)	(13.69,17.33)
SMA-RE : $\varrho_1 = \varrho_2 \sim U(-1,1)$			
Skew-Normal			
	β_1	1.678 (0.604)	(0.471,2.882)
	β_2	7.681 (0.692)	(6.207,9.036)
	δ_μ	6.838 (1.444)	(4.538,10.13)
	δ_ν	8.453 (0.975)	(6.447,10.18)
Normal			
	β_1	3.764 (0.768)	(2.235,5.255)
	β_2	9.216 (0.754)	(7.729,10.80)
SARMA-RE : $\varrho_1, \varrho_2, \rho_1, \rho_2 \sim U(-1,1)$			
Skew-Normal			
	β_1	1.273 (0.703)	(-0.101,2.667)
	β_2	10.70 (0.707)	(9.617,12.38)
	δ_μ	8.695 (2.300)	(4.479,13.45)
	δ_ν	7.052 (1.655)	(4.176,10.39)
Normal			
	β_1	2.131 (0.770)	(0.627,3.659)
	β_2	11.81 (0.808)	(10.26,13.45)

5.2 An Application

We now illustrate the usefulness of our approach using skew distributions by applying it to a real spatial panel data on cigarette consumption collected in the United States over the period 1963 to 1992. The file includes the per pack price of cigarettes (price), per capita disposable income (ndi) consumer price index (cpi), and cigarette sales in packs per capita (sales) over time and across space units, reported in Croissant (2017). Studies on this data are considered in Baltagi and Levin (1986), Baltagi and Levin (1992), Elhorst (2005), Baltagi (2008), Zheng *et al.* (2008), Elhorst (2010), Lee and Yu (2010), and Leorato and Mezzetti (2016), among others.

Cigarette consumption as a worldwide phenomenon has significantly destructive effects on the health consequences of cigarette consumers. The willingness to buy cigarette is strongly influenced by a consumer's sense of real price and income level variables. As a result, its necessary to estimate the price elasticity of cigarettes to understand how price influences smoking decisions. The price elasticity is measured by a individual's sensitivity to price changes. Moreover, income variable is often used to measure per capita cigarette consumption in demand analysis. Therefore, there has been considerable interest in the data that relates cigarette demand to cigarette price and per capita disposable income.

Under usual circumstances when cigarettes are more expensive, demand decreases and price elasticity is reported as a negative value in this case. The percent change in demand for cigarettes because of price change defines the concept of price elasticity. Income elasticity indicates the ability of the population to purchase cigarette as per capita income changes. The common relationship is in a positive direction with the capacity to demand to increase as per capita income increases; see Lee and Yu (2012) and Leorato and Mezzetti (2016). Basically, a positive income elasticity of demand is linked with necessity goods and services meaning a rise in income will lead to a rise in demand. While a negative income elasticity of demand is linked with inferior goods and services. In this case, rising incomes will lead to a drop in demand and may mean changes to luxury goods and services.

Here, the response variable of the model is sales which is measured as the number of cigarettes per person in aged 14 years and older. This is regressed on the real price of cigarettes (price/cpi) and real per capita disposable income (ndi/cpi).

As mentioned previously, logarithms are used in economics because the estimated coefficients in log-linear regressions can be directly interpreted as elasticities and traditionally it has been assumed that the error components follows a normal distribution. These data are expressed via their logarithms on the Elhorst website, www.regroningen.nl/elhorst.

Figure 1(a) suggests that cigarette demand decreases over time in most states but with substantial interspace variation. The histogram of the raw cigarette consumption data in Figure 1(b) shows its asymmetric nature and clearly indicate that the normality assumption is not satisfied. This also was confirmed by using the Shapiro-Wilk test where the test statistic is equal to 0.8351 and p-value is 2.2e-16. Fitting a MSN model to the dataset seems more appropriate. To investigate this we posit the 6 multivariate models (three different SAR structures, two different SMA structures and one SARMA structure) with normal and skew-normal distributions for disturbance terms. Hence, we conduct an empirical analysis of the cigarette demand to estimate price and income elasticities to check the validity of our results through non-normal disturbances with log-linear models.

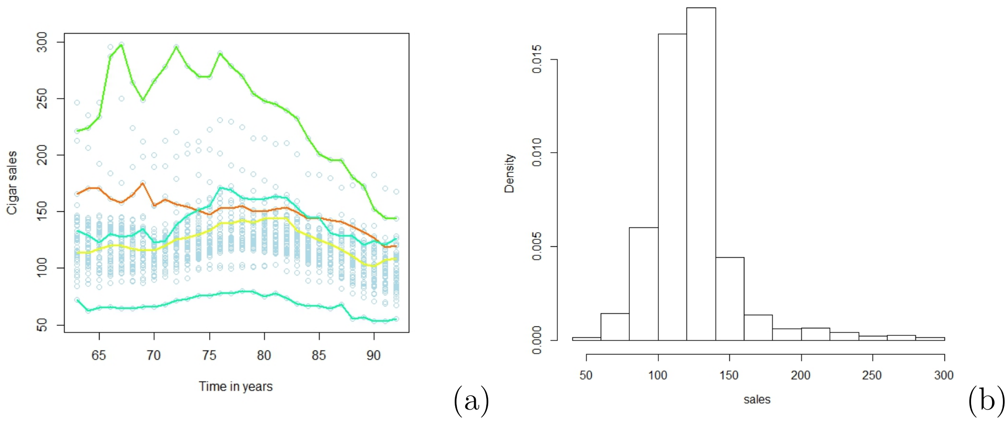


Figure 1: Cigarette demand data. (a) Cigarette sales for 46 states of the US, with trajectories highlighted for 5 random states. (b) Histogram of the raw data.

5.3 Prior distributions

To proceed with Bayesian inferences on the aforementioned data, we need to specify the priors at the population level. In the absence of historical data and experiments, we specify weakly informative prior distributions for all model parameters to obtain well-defined (proper) posteriors. More specifically, we specify the following priors: independent Normal(0, Precision = 0.01) distribution for the components of $\boldsymbol{\beta}$, non-informative inverse gamma distribution IG(0.01, 0.01) for the scale parameters $\sigma_\nu^2 > 0$ and $\sigma_\mu^2 > 0$ (so that the distribution has mean 1 and variance 100), and truncated normal(0, Precision = 0.01) for each of the skewness parameter δ_ν and δ_μ which is used to accommodate positive or negative skewness of the data. Furthermore, in our analysis, it is assumed that the spatial weight matrix $\boldsymbol{W}$ is non-normalized and the parameters space ρ_1, ρ_2, ϱ_1 and ϱ_2 has prior values that vary over the set $\{0; 0.1; 0.15\}$ due to $1/\lambda_{max} \approx 0.19$. To help the estimation process, we have fixed the space parameters at these values. We then estimated the entire spatial skew-normal generalized panel data model all simultaneously (including two skew parameters and all other parameters) to obtain the final results for the model.

5.3.1 Results of analysis

For all special cases in the generalized model, we ran one long chain of 300,000 iterations to estimate the posterior distribution of the parameters after an initial 20,000 burn-in iterations. In order to reduce autocorrelation among successive Markov draws, we used a thinning interval of 100. We formed estimates of the population posterior mean (PM), the corresponding standard deviation (SD) and 95% percentiles credible intervals (CIs) for the unknown parameters. The following items summarize the results obtained for each of the six spatial dependence models with normal and skew-normal assumptions. The estimates are reported along with some of the corresponding values of ρ_1, ρ_2, ϱ_1 and ϱ_2. Tables 3 to 8 present the estimation results. The regression coefficients of price and income are estimated by β_1 and β_2, respectively. This paper by considering the following formula provides price and income elasticity of demand which is important in spatial econometrics applications areas where either skewness should be considered or log-transformation can be avoided.

Price elasticity = Coefficient of price $(\beta_1) \times$ Average $\frac{price}{cpi}$/Average sales.

Income elasticity = Coefficient of income $(\beta_2) \times$ Average $\frac{ndi}{cpi}$/Average sales.

Case 1. When $\rho_1 = 0$ and $\varrho_1 = \varrho_2 = 0$, the model reduces to the RE-SAR model. The results of the model with (a) normal and (b) skew-normal assumption for two values of $\rho_2 = \{0.1, 0.15\}$ are shown in Table 3.

Table 3: **A1:RE-SAR**

A1.1: $\rho_1 = 0, \rho_2 = 0.1$

model	(a) Normal			(b) Skew-Normal		
parameter	PM	SD	95% CI	PM	SD	95% CI
β_1	-7.60e+01	4.07e+00	(-8.35e+01,-6.78e+01)	-8.30e+01	3.37e+00	(-8.97e+01,-7.65e+01)
β_2	-2.13e-01	4.08e-02	(-2.93e-01,-1.31e-01)	-1.25e-01	3.91e-02	(-1.90e-01,-3.96e-02)
σ_μ^2	4.92e+04	1.14e+04	(3.15e+04,7.48e+04)	3.05e+04	7.27e+03	(1.96e+04,4.84e+04)
σ_ν^2	1.61e+02	6.28e+00	(1.49e+02,1.74e+02)	3.53e+01	5.84e+00	(2.49e+01,4.86e+01)
δ_μ	-	-	-	2.11e+01	9.52e+00	(3.22e+00,4.04e+01)
δ_ν	-	-	-	1.87e+01	6.85e-01	(1.74e+01,2.01e+01)

A1.2 : $\rho_1 = 0, \rho_2 = 0.15$

model	(a) Normal			(b) Skew-Normal		
parameter	PM	SD	95% CI	PM	SD	95% CI
β_1	-7.42e+01	4.68e+00	(-8.22e+01,-6.49e+01)	-8.24e+01	4.33e+00	(-9.10e+01,-7.45e+01)
β_2	-2.77e-01	5.09e-02	(-3.77e-01,-1.75e-01)	-1.87e-01	4.87e-02	(-2.82e-01,-9.40e-02)
σ_μ^2	5.12e+04	1.19e+04	(3.22e+04,7.80e+04)	2.51e+04	6.35e+03	(1.54e+04,4.05e+04)
σ_ν^2	1.61e+02	6.31e+00	(1.49e+02,1.74e+02)	3.53e+01	5.95e+00	(2.45e+01,4.69e+01)
δ_μ	-	-	-	2.36e+01	1.01e+01	(4.85e+00,4.31e+01)
δ_ν	-	-	-	1.88e+01	7.33e-01	(1.73e+01,2.02e+01)

Case 2. When $\rho_1 = \rho_2$ and $\varrho_1 = \varrho_2 = 0$ the model reduces to the SAR-RE model. The results of the model with (a) normal and (b) skew-normal assumption for values of $\rho_1 = \rho_2 = 0.1$ and $\rho_1 = \rho_2 = 0.15$ are shown in Table 4.

Table 4: **A1:SAR-RE**

A1.3: $\rho_1 = 0.1, \rho_2 = 0.1$

model	(a) Normal			(b) Skew-Normal		
parameter	PM	SD	95% CI	PM	SD	95% CI
β_1	-7.59e+01	4.14e+00	(-8.33e+01,-6.75e+01)	-8.24e+01	3.75e+00	(-9.04e+01,-7.53e+01)
β_2	-2.14e-01	4.09e-02	(-2.93e-01,-1.28e-01)	-1.25e-01	3.60e-02	(-1.98e-01,-4.65e-02)
σ_μ^2	1.92e+04	4.42e+03	(1.23e+04,2.86e+04)	8.90e+03	2.75e+03	(4.02e+03,1.46e+04)
σ_ν^2	1.61e+02	6.20e+00	(1.49e+02,1.74e+02)	3.52e+01	5.78e+00	(2.60e+01,4.67e+01)
δ_μ	-	-	-	3.49e+01	1.07e+01	(1.30e+01,5.76e+01)
δ_ν	-	-	-	1.87e+01	7.21e-01	(1.72e+01,2.01e+01)

A1.4 : $\rho_1 = 0.15, \rho_2 = 0.15$

model	(a) Normal			(b) Skew-Normal		
parameter	PM	SD	95% CI	PM	SD	95% CI
β_1	-7.43e+01	4.73e+00	(-8.31e+01,-6.51e+01)	-8.29e+01	4.33e+00	(-9.10e+01,-7.40e+01)
β_2	-2.77e-01	5.16e-02	(-3.75e-01,-1.70e-01)	-1.81e-01	4.82e-02	(-2.73e-01,-8.54e-02)
σ_μ^2	1.16e+04	2.68e+03	(7.37e+03, 1.74e+04)	5.35e+03	1.71e+03	(2.78e+03,9.56e+03)
σ_ν^2	1.61e+02	6.23e+00	(1.49e+02,1.74e+02)	3.58e+01	6.12e+00	(2.56e+01,4.86e+01)
δ_μ	-	-	-	2.96e+01	9.62e+00	(9.96e+00,4.70e+01)
δ_ν	-	-	-	1.86e+01	7.41e-01	(1.72e+01,2.01e+01)

Case 3. When $\varrho_1 = \varrho_2 = 0$ the model reduces to the GRE-SAR model. The results of the model with (a) normal and (b) skew-normal assumption for values of $\rho_1 = 0.1$, $\rho_2 = 0.15$ and $\rho_1 = 0.15$, $\rho_2 = 0.1$ are shown in Table 5.

Table 5: **A1:GRE-SAR**

A1.5: $\rho_1 = 0.1, \rho_2 = 0.15$

model	(a) Normal			(b) Skew-Normal		
parameter	PM	SD	95% CI	PM	SD	95% CI
β_1	-7.42e+01	4.74e+00	(-8.26e+01,-6.50e+01)	-8.26e+01	4.30e+00	(-9.05e+01,-7.39e+01)
β_2	-2.77e-01	5.15e-02	(-3.74e-01,-1.71e-01)	-1.74e-01	4.62e-02	(-2.63e-01,-8.20e-02)
σ^2_μ	2.01e+04	4.64e+03	(1.28e+04,2.98e+04)	8.42e+03	2.55e+03	(4.12e+03,1.45e+04)
σ^2_ν	1.61e+02	6.28e+00	(1.49e+02,1.75e+02)	3.52e+01	5.66e+00	(2.50e+01,4.78e+01)
δ_μ	-	-	-	3.34e+01	1.11e+01	(1.17e+01,5.50e+01)
δ_ν	-	-	-	1.87e+01	6.93e-01	(1.73e+01,2.03e+01)

A1.6 : $\rho_1 = 0.15, \rho_2 = 0.1$

model	(a) Normal			(b) Skew-Normal		
parameter	PM	SD	95% CI	PM	SD	95% CI
β_1	-7.60e+01	4.14e+00	(-8.39e+01,-6.75e+01)	-8.28e+01	3.61e+00	(-9.01e+01,-7.53e+01)
β_2	-2.13e-01	4.09e-02	(-2.94e-01,-1.26e-01)	-1.21e-01	3.62e-02	(-1.83e-01,-4.23e-02)
σ^2_μ	1.11e+04	2.55e+03	(7.08e+03,1.67e+04)	4.90e+03	1.70e+03	(2.13e+03,8.80e+03)
σ^2_ν	1.61e+02	6.15e+00	(1.49e+02,1.73e+02)	3.55e+01	6.34e+00	(2.45e+01,4.86e+01)
δ_μ	-	-	-	3.36e+01	9.69e+00	(1.31e+01,5.32e+01)
δ_ν	-	-	-	1.87e+01	7.83e-01	(1.70e+01,2.03e+01)

Case 4. When $\rho_1 = \rho_2 = 0$ and $\varrho_2 = 0$ the model reduces to the RE-SMA model. The results of the model with (a) normal and (b) skew-normal assumption for two values of $\varrho_1 = \{0.1, 0.15\}$ are shown in Table 6.

Table 6: **RE-SMA**

A2.1: $\varrho_1 = 0.1, \varrho_2 = 0$

model	(a) Normal			(b) Skew-Normal		
parameter	PM	SD	95% CI	PM	SD	95% CI
β_1	-7.88e+01	3.13e+00	(-8.49e+01,-7.24e+01)	-8.27e+01	2.82e+00	(-8.77e+01,-7.67e+01)
β_2	-1.20e-01	2.79e-02	(-1.75e-01,-5.87e-02)	-6.40e-02	2.65e-02	(-1.19e-01,-1.51e-02)
σ_μ^2	2.52e+04	5.75e+03	(1.62e+04,3.82e+04)	1.45e+04	4.36e+03	(8.24e+03,2.51e+04)
σ_ν^2	1.74e+02	6.61e+00	(1.62e+02,1.88e+02)	4.09e+01	7.65e+00	(2.80e+01,5.68e+01)
δ_μ	-	-	-	2.98e+01	1.06e+01	(9.30e+00,5.15e+01)
δ_ν	-	-	-	1.94e+01	7.86e-01	(1.78e+01,2.09e+01)

A2.2 : $\varrho_1 = 0.15, \varrho_2 = 0$

model	(a) Normal			(b) Skew-Normal		
parameter	PM	SD	95% CI	PM	SD	95% CI
β_1	-7.88e+01	3.13e+00	(-8.50e+01,-7.25e+01)	-8.28e+01	2.49e+00	(-8.78e+01,-7.80e+01)
β_2	-1.20e-01	2.79e-02	(-1.75e-01,-5.95e-02)	-6.35e-02	2.23e-02	(-1.03e-01,-1.73e-02)
σ_μ^2	2.06e+04	4.70e+03	(1.32e+04,3.12e+04)	1.07e+04	2.96e+03	(5.85e+03,1.81e+04)
σ_ν^2	1.74e+02	6.58e+00	(1.62e+02,1.88e+02)	4.06e+01	6.32e+00	(2.93e+01,5.41e+01)
δ_μ	-	-	-	3.22e+01	1.02e+01	(1.17e+01, 5.12e+01)
δ_ν	-	-	-	1.94e+01	7.21e-01	(1.79e+01,2.08e+01)

Case 5. When $\rho_1 = \rho_2 = 0$ and $\varrho_1 = \varrho_2$ the model reduces to the SMA-RE model. The results of the model with (a) normal and (b) skew-normal assumption for values of $\varrho_1 = \varrho_2 = 0.1$ and $\varrho_1 = \varrho_2 = 0.15$ are shown in Table 7.

Table 7: **SMA-RE**

A2.3: $\varrho_1 = 0.1, \varrho_2 = 0.1$

model	(a) Normal			(b) Skew-Normal		
parameter	PM	SD	95% CI	PM	SD	95% CI
β_1	-7.71e+01	3.84e+00	(-8.39e+01,-6.92e+01)	-8.29e+01	3.45e+00	(-8.98e+01,-7.65e+01)
β_2	-1.64e-01	3.68e-02	(-2.35e-01,-8.70e-02)	-8.53e-02	3.31e-02	(-1.57e-01,-2.49e-02)
σ_μ^2	2.58e+04	5.92e+03	(1.66e+04,3.86e+04)	1.38e+04	4.20e+03	(6.43e+03,2.31e+04)
σ_ν^2	1.74e+02	6.71e+00	(1.62e+02,1.88e+02)	4.11e+01	6.80e+00	(2.83e+01,5.53e+01)
δ_μ	-	-	-	3.04e+01	1.10e+01	(8.51e+00,5.40e+01)
δ_ν	-	-	-	1.93e+01	7.91e-01	(1.77e+01,2.07e+01)

A2.4 : $\varrho_1 = 0.15, \varrho_2 = 0.15$

model	(a) Normal			(b) Skew-Normal		
parameter	PM	SD	95% CI	PM	SD	95% CI
β_1	-7.65e+01	4.19e+00	(-8.40e+01,-6.83e+01)	-8.29e+01	3.79e+00	(-9.04e+01,-7.52e+01)
β_2	-1.69e-01	4.20e-02	(-2.51e-01,-8.32e-02)	-7.73e-02	3.80e-02	(-1.55e-01,-3.39e-03)
σ_μ^2	2.11e+04	4.85e+03	(1.35e+04,3.13e+04)	9.85e+03	2.87e+03	(4.83e+03,1.63e+04)
σ_ν^2	1.90e+02	7.35e+00	(1.76e+02,2.06e+02)	4.64e+01	7.82e+00	(3.29e+01,6.30e+01)
δ_μ	-	-	-	3.26e+01	1.07e+01	(1.13e+01,5.28e+01)
δ_ν	-	-	-	2.01e+01	8.81e-01	(1.83e+01,2.16e+01)

Case 6. When $\rho_1 = \rho_2$ and $\varrho_1 = \varrho_2$ the model reduces to the SARMA model. The results of the model with (a) normal and (b) skew-normal assumption for values of $\rho_r = \varrho_r = 0.1$ and $\rho_r = \varrho_r = 0.15$; $r = 1, 2$ are shown in Table 8.

Table 8: **SARMA-RE**

A3.1: $\rho_1 = 0.1, \rho_2 = 0.1, \varrho_1 = 0.1, \varrho_2 = 0.1$

model	(a) Normal			(b) Skew-Normal		
parameter	PM	SD	95% CI	PM	SD	95% CI
β_1	-7.50e+01	4.75e+00	(-8.36e+01,-6.57e+01)	-8.30e+01	4.20e+00	(-9.04e+01,-7.51e+01)
β_2	-2.19e-01	5.25e-02	(-3.21e-01,-1.11e-01)	-1.09e-01	5.02e-02	(-2.08e-01,-1.09e-02)
σ^2_μ	1.28e+04	2.95e+03	(8.08e+03,1.90e+04)	4.82e+03	1.86e+03	(1.71e+03,9.23e+03)
σ^2_ν	1.88e+02	7.29e+00	(1.74e+02,2.04e+02)	4.49e+01	7.81e+00	(3.19e+01,6.25e+01)
δ_μ	-	-	-	3.54e+01	1.05e+01	(1.44e+01,5.38e+01)
δ_ν	-	-	-	2.00e+01	8.50e-01	(1.82e+01,2.17e+01)

A3.2 : $\rho_1 = 0.1, \rho_2 = 0.1, \varrho_1 = 0.15, \varrho_2 = 0.15$

model	(a) Normal			(b) Skew-Normal		
parameter	PM	SD	95% CI	PM	SD	95% CI
β_1	-7.51e+01	4.10e+00	(-8.41e+01,-6.60e+01)	-8.32e+01	4.70e+00	(-9.15e+01,-7.33e+01)
β_2	-1.92e-01	5.88e-02	(-3.07e-01,-7.43e-02)	-8.69e-02	5.43e-02	(-1.95e-01,-9.37e-03)
σ^2_μ	1.13e+04	2.62e+03	(7.09e+03,1.68e+04)	4.69e+03	1.52e+03	(2.09e+03,7.93e+03)
σ^2_ν	2.19e+02	8.50e+00	(2.02e+02,2.37e+02)	5.36e+01	9.05e+00	(3.81e+01,7.33e+01)
δ_μ	-	-	-	3.14e+01	9.70e+00	(1.34e+01,5.17e+01)
δ_ν	-	-	-	5.36e+01	9.05e+00	(3.81e+01,7.33e+01)

Our estimation of the parameters in the generalized class of spatial dependence models can be summarized as follows: (i) the most 95% CIs associated with each of fitted values from the normal models are wider than that from the corresponding skew-normal models; (ii) all of the parameter estimates are significant since the 95% CIs do not include zero; (iii) the estimate of the variance component of the random effects, σ^2_μ, and the variance component of the random errors, σ^2_ν, are smaller in the skew-normal cases as compared to the normal case primarily because of the interrelation between high variability and skewness; (iv) among all of the parameters estimated, mostly the related SD obtained from the skew-normal distribution is the smallest; (v) both skewness parameters δ_ϵ and δ_μ are significantly positive for all skewed versions of the fitted models and thus suggests there is moderate right skewness in the data. In particular, as the CIs of δ_ν and δ_μ do not include zero, this confirms the positive skewness of the responses. These estimates indicate that consideration of departures from normality by the adaption of a model that allows for skewness may improve the generalized model fit. The marginal posterior density plots of the parameters for our competing models are shown in Figures 2 and 3 which all support posterior results for the parameters under our proposed models.

5.3.2 Goodness-of-Fit

Along with the above results from Case (1-6) in Tables 3 to 8, we further report results for the goodness-of-fit of the normal and skew-normal models, using the previous model choice criteria (AIC, BIC and HQC). Table 9 presents the comparison among the competing models. The results indicate that the spatial panel models with the skew-normal distribution assumption provide better fits as indicated by their lower average values of the model choice criteria relative to the spatial panel models with normal distribution assumption. Thus, there is a potential gain of efficiency in estimating the certain parameters when the normality assumption does not apply to the data.

Table 9: Model comparison using AIC, BIC, HQC criteria

model	(a) Normal			(b) Skew-Normal		
Criteria	AIC	BIC	HQC	AIC	BIC	HQC
A1.1	10878	10876.65	10871.77	9809	9806.98	9799.65
A1.2	10888	10886.65	10881.77	9780	9777.98	9770.65
A1.3	10878	10876.65	10871.77	9807	9804.98	9797.65
A1.4	10878	10876.65	10871.77	9802	9799.98	9792.65
A1.5	10888	10886.65	10881.77	9783	9780.98	9773.65
A1.6	10878	10876.65	10871.77	9802	9799.98	9792.65
A2.1	10988	10986.65	10981.77	10012	10009.98	10002.65
A2.2	10988	10986.65	10981.77	10010	10007.98	10000.65
A2.3	10988	10986.65	10981.77	9988	9985.98	9978.65
A2.4	11118	11116.65	11111.77	1092	10189.98	10181.32
A3.1	11098	11096.65	11091.77	10122	10119.98	10112.65
A3.2	11308	11306.65	11301.77	10402	10399.98	10392.65

The estimated coefficient for the mean price (income) in Tables 3 to 8 are statistically significant. It can be interpreted as mean price (income) elasticity of cigarette demand easily. A comprehensive description of the estimation techniques for the price and income elasticity of demand for cigarettes can be found in Gallet and List (2003). In most previous studies, all three variables in the cigarette demand data (cigarette sales, $\frac{price}{cpi}$ and $\frac{ndi}{cpi}$) are expressed in logarithms to have the routine use in the estimation of the elasticities. In the empirical analysis were used sales as a dependent variable and real price and real income as independent variables. Indeed, elasticity values, rather than the coefficient estimates for price and income, are used as the dependent variables because elasticities are unit-free, easily interpreted, and comparable across studies. We now compare this approach with our proposed methodology. For the studies used here, price (income) elasticity value for the demand from raw cigarette data with the skewed assumption is approximately -0.6 (-0.1) and for the log-transformed model with the normal assumption is -0.8 (nearly zero). The results of the data fit comparisons are presented in Table 10.

Table 10: Estimated Elasticities by Model Characteristics

model	raw		transformed	
Criteria	Mean Price Elasticity	Mean Income Elasticity	Mean Price Elasticity	Mean Income Elasticity
A1.1	-0.609	-0.097	-0.773	0.021
A1.2	-0.605	-0.145	-0.828	0.056
A1.3	-0.605	-0.097	-0.773	0.028
A1.4	-0.608	-0.140	-0.828	0.056
A1.5	-0.606	-0.135	-0.850	0.076
A1.6	-0.608	-0.094	-0.773	0.021
A2.1	-0.607	-0.050	-0.707	0.003
A2.2	-0.607	-0.049	-0.707	0.003
A2.3	-0.608	-0.066	-0.748	0.020
A2.4	-0.608	-0.060	-0.789	0.058
A3.1	-0.609	-0.084	-0.822	0.077
A3.2	-0.610	-0.067	-0.865	0.128

Under the Bayesian approach based on the normal distribution for error terms of model A1.1, the analysis in Zheng *et al.* (2008) concluded that a significant negative effect of cigarette price is on cigarette consumption, while income has no effect on cigarette consumption with an insignificant income elasticitiy. In this case, by ignoring the amount of spatial dependence parameter, our estimated parameters are in line with the results of Zheng *et al.* (2008), which is amenable to popular software R. Further, the elasticity values for Case1 to 6 in Table 10 suggest that it is somewhat different between the amount of the estimated mean price (income) elasticity given by the raw and transformed models. In other words, our method does not noticeably influence the price elasticity and they are more stable within the different models. But, income elasticity estimates are in a different direction. More detailes, there is a significant negative effect of income on raw cigarette sales based on the posterior mean income elasticities. It means that as consumers' income rises, their cigarette consumption is less than before. Negative values of the income elasticity of demand usually depend on high education, such that higher income lead to less cigarette consumption; see Baltagi and Llevin (1986), Baltagi and Li (2004) and Zheng *et al.* (2008) for more detailes. Therefore, our results are similar to the main findings of Baltagi and Levin (1992), namely, there is a significant negative price elasticity and significant but small negative income elasticity.

It can be inferred from this study that the variation in these elasticity estimates arises because of differences between the assumptions inherent in the model and the estimation techniques. Therefore, we think that using our estimation technique to determine the econometric estimates is important. Hence, our method can be viewed as an objective method in econometric modelling as it is very easy to implement and yield different conclusions about the relationship between elasticity values for the provision of a reasonable model.

6 Conclusions

This paper has investigated the generalized panel data model proposed by Lee and Yu (2012) which encompasses various spatial panel regression models. It contributes to the spatial econometrics field by proposing a more flexible and powerful model that allows for a non-normal (through parametric) random error components specifications of the spatial panel model. To our knowledge, this is the first spatial non-normal generalized panel data model proposed in the economic literature. The skew-normal distribution that we use for the random error components is mathematically tractable, parsimonious in parameters, and includes the normal distribution as a special case. It is not only flexible because it allows a continuity of shapes from normality to non-normality, including positive skewness, but allows easy implementable Bayesian analysis to be performed. We have implemented a Bayesian inference approach for estimation via the generalized model, which is based on the method by Arellano-Valle *et al.* (2007). We have demonstrated the proposed methodology of using a skew distribution for the random error components of the generalized model is useful and important in spatial econometrics especially when the data exhibit skewness and the log transformation can be avoided.

ACKNOWLEDGEMENTS. The first auther would like to acknowledge the support of Shahid Chamran University of Ahvaz to enable her to visit the University of Queensland. We would like to thank Dr. Sharon Lee of the Department of Mathematics, University of Queensland for her valuable comments and corrections which enhanced the preparation of this paper.

Appendix

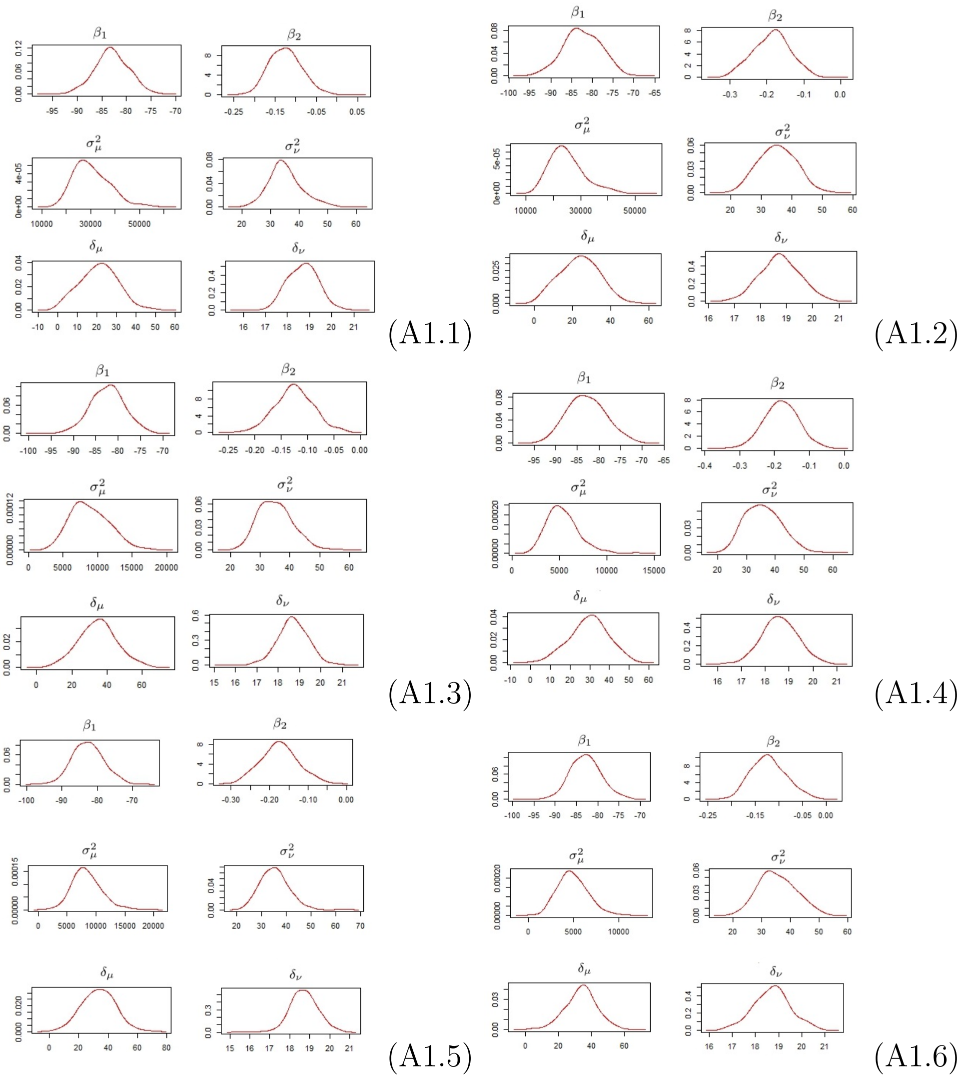
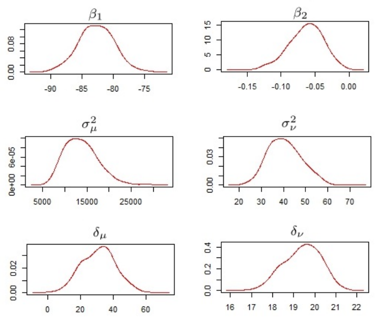
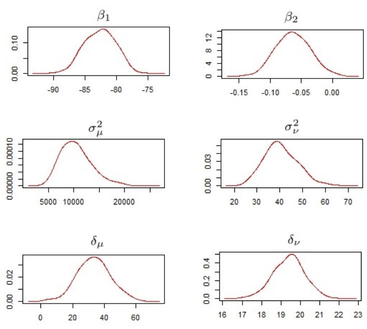
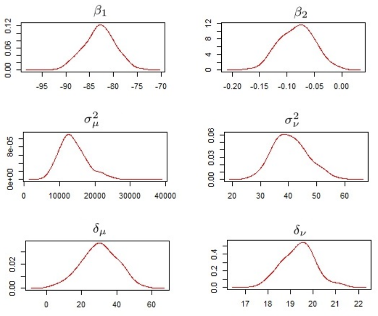
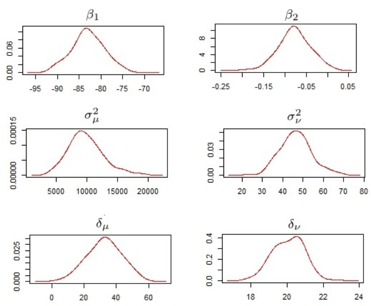


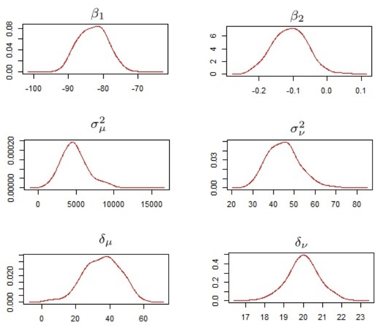
Figure 2: The marginal posterior density plots of the skew models parameters: SAR


(A2.1)


(A2.2)


(A2.3)


(A2.4)


(A3.1)

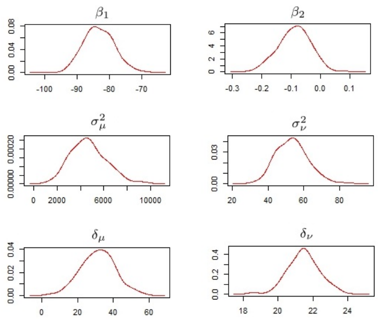

(A3.2)

Figure 3: The marginal posterior density plots of the skew models parameters: SMA and SARMA

References

[1] H. Akaike, A new look at the statistical model identification, *IEEE Transactions on Automatic Control*, **19**(6), (1974), 716 - 723.

[2] L. Anselin, *Spatial Econometrics: Methods and Models*, Kluwer Accademic: Dordrecht, 1988.

[3] L. Anselin, J. Le Gallo and H. Jayet, Spatial panel econometrics, *In the Econometrics of Panel Data*, (2008), Berlin: Springer, 625 - 660.

[4] R. B. Arellano-Valle, H. Bolfarine and V. H. Lachos, Bayesian inference for skew-normal linear mixed models, *Journal of Applied Statistics*, **34**(6), (2007), 663 - 682.

[5] R. B. Arellano-Valle and A. Azzalini, On the unification of families of skew-normal distributions, *Scandinavian Journal of Statistics*, **33**(3), (2006), 561 - 574.

[6] R. B. Arellano-Valle and M. G. Genton, On fundamental skew distributions, *Journal of Multivariate Analysis*, **96**(1), (2005), 93 - 116.

[7] A. Azzalini and A. Capitanio, Statistical applications of the multivariate skew normal distribution, *Journal of the Royal Statistical Society Series B*, **61**(3), (1999), 579 - 602.

[8] A. Azzalini and A. D. Valle, The multivariate skew-normal distribution, *Biometrika*, **83**(4), (1996), 715–726.

[9] B.H. Baltagi, S.H. Song, B.C. Jung and W. Koh, Testing for serial correlation, spatial autocorrelation and random effects using panel data, *Journal of Econometrics*, **140**(1), (2007), 5 - 51.

[10] B. H. Baltagi and L. Liu , Prediction in a generalized spatial panel data model with serial correlation, *Journal of Forecasting*, **35**(7), (2016), 573 - 591.

[11] B. H. Baltagi and D. Levin, Estimating dynamic demand for cigarettes using panel data: the effects of bootlegging, taxation and advertising reconsidered, *Review of Economics and Statistics*, (1986), 148 - 155.

[12] B. H. Baltagi, Econometric Analysis of Panel Data, *Proceedings of the conference name*, Chichester: John Wiley and Sons, Ltd, 2008.

[13] B. H. Baltagi and D. Levin, Cigarette taxation: Raising revenues and reducing consumption, *Structural Change and Economic Dynamics*, (1992), 321 - 335.

[14] B. H. Baltagi, G. Bresson and A. Pirotte, Forecasting with spatial panel data, *Computational Statistics and Data Analysis*, **56**(11), (2012), 3381 - 3397.

[15] C. R. Bhat, S. Astroza and A. S. Hamdi, A spatial generalized ordered-response model with skew normal kernel error terms with an application to bicycling frequency, *Transportation Research, Part B: Methodological*, **95**, (2017), 126 - 148.

[16] Y. Croissant , Package Ecdat, (2017).

[17] P. J. Elhorst, Unconditional maximum likelihood estimation of linear and log-linear dynamic models for spatial panels, *Geographical Analysis*, **37**(1), (2005), 62 - 83.

[18] P. J. Elhorst, Spatial Panel Data Models, In M. M. Fischer and A. Getis (Eds.), *Handbook of Applied Spatial Analysis*, Berlin: Springer, (2010), 377 - 407.

[19] B. Fingleton, A generalized method of moments estimators for a spatial panel model with an endogenous spatial lag and spatial moving average errors, *Spatial Economic Analysis*, **3**(1), (2008), 27 - 44.

[20] C. A. Gallet and J. A. List, Cigarette demand: a meta-analysis of elasticities, *Health Economics* , **12**(10), (2003), 821 - 835.

[21] E. J. Hannan and B. G. Quinn, The determination of the order of an autoregression, *Journal of the Royal Statistical Society Series B*, **41**, (1979), 190 - 195.

[22] M. Kapoor, H. H. Kelejian and I. R. Prucha , Panel data models with spatially correlated error components, *Journal of Econometrics*, **140**(1), (2007), 97 - 130.

[23] S. X. Lee and G. J. McLachlan, On mixtures of skew normal and skew t-distributions, *Advances in Data Analysis and Classification*, **7**(3), (2013), 241 - 266.

[24] S. X. Lee and G. J. McLachlan, Finite mixtures of multivariate skew t-distributions: some recent and new results, *Statistics and Computing*, **24**(2), (2014), 181 - 202.

[25] S. X. Lee and G. J. McLachlan, Finite mixtures of canonical fundamental skew t-distributions: the unification of the restricted and unrestricted

skew t-mixture models, *Statistics and Computing*, **26**(3), (2016), 573 - 589.

[26] L.F. Lee and J. Yu, Spatial panels: random components versus fixed effects, *International Economic Review*, **53**(4), (2012), 1369 - 1412.

[27] L. F. Lee and J. Yu, Some recent developments in spatial panel data models, *Regional Science and Urban Economics*, **40**(5), (2010), 255 - 271.

[28] S. Leorato and M. Mezzetti, Spatial panel data model with error dependence: a Bayesian separable covariance approach, *Bayesian Analysis*, **11**(4), (2016), 1035 - 1069.

[29] U. Ligges, J. Kerman, S. OpenBUGS and M. N. Thomas, Package R2OpenBUGS, (2017).

[30] M. U. Ligges, Package BRugs, (2017).

[31] T. I. Lin, G. J. McLachlan and S. X. Lee, Extending mixtures of factor models using the restricted multivariate skew-normal distribution, *Journal of Multivariate Analysis*, **143**, (2016), 398 - 413.

[32] D. Lunn, D. Spiegelhalter, A. Thomas and N. Best, The BUGS project: Evolution, critique and future directions, *Statistics in Medicine*, **28**(25), (2009), 3049 - 3067.

[33] S. G. Meintanis and Z. Hlavka, Goodness-of-Fit tests for bivariate and multivariate skew normal distributions., *Scandinavian Journal of Statistics*, **37**(4), (2010), 701 - 714.

[34] G. Molenberghs and G. Verbeke, *Linear Mixed Models for Longitudinal Data*, Berlin: Springer, 2000.

[35] D. Molenaar, C. V. Dolan and N. D. Verhelst, Testing and modelling non-normality within the one-factor model, *British Journal of Mathematical and Statistical Psychology*, **63**(2), (2010), 293 - 317.

[36] M. Nerlove, A note on error components models, *Econometrica* , **39**, (1971), 383 - 396.

[37] S. K. Sahu, D. K. Dey and M. D. Branco, A new class of multivariate skew distributions with applications to Bayesian regression models, *Canadian Journal of Statistics*, **31**(2), (2003), 129 - 150.

[38] G. Schwarz, Estimating the dimension of a model, *The Annals of Statistics*, **6**(2), (1978), 461 - 464.

[39] M. S. Smith, Q. Gan and R. J. Kohn, Modelling dependence using skew t copulas: Bayesian inference and applications, *Journal of Applied Econometrics*, **27**(3), (2012), 500 - 522.

[40] Y. Zheng, J. Zhu and D. Li, Analyzing spatial panel data of cigarette demand: A Bayesian hierarchical modelling approach, *Proceedings of the conference name*, **6**(4), (2008), 467 - 489.