mathematics

Article

A Bayesian Variable Selection Method for Spatial Autoregressive Quantile Models

Yuanying Zhao [1,*] and Dengke Xu [2]

[1] College of Mathematics and Information Science, Guiyang University, Guiyang 550005, China
[2] School of Economics, Hangzhou Dianzi University, Hangzhou 310018, China
* Correspondence: zhaoyuanying_@126.com

Abstract: In this paper, a Bayesian variable selection method for spatial autoregressive (SAR) quantile models is proposed on the basis of spike and slab prior for regression parameters. The SAR quantile models, which are more generalized than SAR models and quantile regression models, are specified by adopting the asymmetric Laplace distribution for the error term in the classical SAR models. The proposed approach could perform simultaneously robust parametric estimation and variable selection in the context of SAR quantile models. Bayesian statistical inferences are implemented by a detailed Markov chain Monte Carlo (MCMC) procedure that combines Gibbs samplers with a probability integral transformation (PIT) algorithm. In the end, empirical numerical examples including several simulation studies and a Boston housing price data analysis are employed to demonstrate the newly developed methodologies.

Keywords: Bayesian variable selection; quantile regression; spatial autoregressive models; Gibbs sampling; PIT algorithm

MSC: 62F15

check for updates

Citation: Zhao, Y.; Xu, D. A Bayesian Variable Selection Method for Spatial Autoregressive Quantile Models. *Mathematics* **2023**, *11*, 987. https://doi.org/10.3390/math11040987

Academic Editors: Min Wang, Haijun Gong, Liucang Wu and Songfeng Zheng

Received: 20 December 2022
Revised: 4 February 2023
Accepted: 13 February 2023
Published: 15 February 2023

1. Introduction

Spatial regression models play a critical part in analyzing and tackling spatial data that is broadly available in spatial statistics, regional science, and spatial econometrics.In particularly, the spatial autoregressive (SAR) models proposed by Cliff and Ord [1] have received a lot of attention in recent years. For instance, LeSage and Pace [2] listed several Bayesian estimation and maximum likelihood estimation methods for SAR models in their monograph; based on a generalized method of moments estimator, Du et al. [3] made some statistical inferences for partially linear additive SAR models; under the condition of independent and identical distributed random errors terms in SAR models, Liu et al. [4] proposed a penalized quasi-maximum likelihood method which could perform parameter estimation and model selection simultaneously; Xie et al. [5] considered performing variable selection in SAR models with a diverging number of parameters; Jin and Lee [6] obtained a GEL estimation and investigated several typical test statistics of high-order SAR models; and recently, Ju et al. [7] developed Bayesian statistical diagnostics procedures in the framework of skew-normal SAR models. However, it is unfortunate that the methods proposed by all the cited works are based on the mean regression.

Quantile regression [8–10] studies the relationship among the quantile of the response with the explanatory variables, which provides a more robust and systematic path to investigate the dependence of the response on explanatory variables than mean regression. As a matter of fact, quantile regression considers how explanatory variables have impacts on the conditional quantiles of response variables instead of the conditional mean of the response variable, and presents a more comprehensive and complete picture of the relationship of the response variable with the explanatory variables. We refer the reader to Koenker [8] for an overview on quantile regression. It is noted that there has been a number of research papers on quantile regression issues in the Bayesian statistical framework. For example, Dunson and Taylor [11] presented a Bayesian method for quantile regression analysis based on the substitution likelihood idea [12]; Lancaster and Jun [13] conducted quantile regression analysis by utilizing Bayesian exponentially tilted empirical likelihood; Kottas and Krnjajic [14] developed a generalized Bayesian framework for quantile regression on the basis of Dirichlet processes; Yang and He [15] proposed a Bayesian quantile regression method which is equipped with empirical likelihood; and Rodrigues et al. [16] presented a Bayesian pyramid quantile regression approach that could make simultaneously statistical inferences at several different quantile levels. In particularly, it is a very natural, simple, and effective way to model Bayesian quantile regression by using asymmetric Laplace distribution, which has been studied by several authors, including Yu and Moyeed [17], Kozumi and Kobayashi [18], Hu et al. [19], and Wang and Tang [20], among others. Nevertheless, to the best of our knowledge, little work has been carried out on Bayesian quantile analysis for SAR models due to their complex spatially dependent structure.Therefore, based on references [17,18] and spike and slab prior [21–23], which is often regarded as the gold standard in Bayesian variable selection setting, a Bayesian procedure is proposed to perform simultaneously parameter estimation and variable selection in SAR models. The novel contributions of this paper are listed as follows: (i) we consider a more generalized model than the SAR model and the quantile model, whose Bayesian analysis has not been done, and investigate the sensitivity of the Bayesian estimates to different prior inputs; (ii) we adopt asymmetric Laplace distribution for the error term [17,18] in the SAR models to build the hierarchical SAR quantile models and a full MCMC algorithm combining the Gibbs sampler and the probability integral transformation (PIT) algorithm is developed simultaneously to perform robust parametric estimation, to identify significant explanatory variables, and to build accurate predictive models in the considered models based on spike and slab prior [21–23]; (iii) the required conditional posterior distributions, which are more tedious than those in analysis of the SAR model and the quantile model, are derived, and the implementation of the PIT algorithm for generating observations from the tedious conditional posterior distribution is presented. (iv) Results obtained from empirical numerical examples show that the estimate performance, predictive performance, and variable selection performance of our proposed approach are indeed quite satisfactory.

This rest of the paper is arranged as follows. Hierarchical SAR quantile models using the asymmetric Laplace distribution for the error term [17,18] in the SAR models are proposed in Section 2. In Section 3, we explicitly describe a Bayesian variable selection procedure based on the spike and slab prior [21–23] in the SAR quantile model setting, which combines Gibbs sampling [24] with the probability integral transformation (PIT) algorithm [25] to perform parameter estimation and variable selection simultaneously. Several simulation studies are conducted and a real Boston housing price data anaysis is used to demonstrate our proposed methodologies in Section 4. A discussion is presented in the final section. Some sampling technique details are described within the appendix.

2. Hierarchical Bayesian Quantile Modeling for SAR Models

A Bayesian quantile regression approach for SAR models is proposed in this paper. At a given quantile level $\tau \in (0,1)$, a SAR quantile model has the following form,

$$y_i = \rho_\tau \sum_{j=1}^{n} \omega_{ij} y_j + x_i^T \boldsymbol{\beta}_\tau + \varepsilon_i \tag{1}$$

for $i = 1, \cdots, n$, where y_i is the observation on the response variables and denotes $\boldsymbol{y} = \{y_1, \ldots, y_n\}$; ρ_τ is spatial parameter; ω_{ij} is the *ith* row and *jth* column element of an $n \times n$ spatial weight matrix $\boldsymbol{\omega}$, whose diagonal elements are zeros and other elements are known constants; x_i is a $p \times 1$ observation for explanatory variables of linear regressors $\boldsymbol{x} = (x_1, x_2, \cdots, x_n)^T$; the vector $\boldsymbol{\beta}_\tau = (\beta_1, \ldots, \beta_p)^T$ is the unknown regression coefficients; ε_i is a random error term with τth quantile equaling 0, i.e., $\int_{-\infty}^{0} p_\tau(\varepsilon_i) d\varepsilon_i = \tau$. Quantile regression is often implemented by dealing with a minimization problem based on the check function. Under Equation (1), the specification problem evolves to estimate ρ_τ and $\boldsymbol{\beta}_\tau$ by minimizing the following equation

$$L(\boldsymbol{y}, \boldsymbol{x}) = \sum_{i=1}^{n} \varphi_\tau(y_i - \rho_\tau \sum_{j=1}^{n} \omega_{ij} y_j - x_i^T \beta_\tau), \tag{2}$$

in which $\varphi_\tau(.)$ is the check function defined by $\varphi_\tau(u) = u\{\tau - I(u < 0)\}$ and $I(.)$ represents the common indicator function. In the Bayesian statistical framework, it is assumed that $\varepsilon_i, \cdots, \varepsilon_n$ are independently and identically distributed random variables and ε_i follows an asymmetric Laplace distribution with probability density function

$$p_\tau(\varepsilon_i) = \frac{\tau(1-\tau)}{\sigma} exp\{-\frac{1}{\sigma}\varphi_\tau(\varepsilon_i)\}),$$

in which σ is the scale parameter. Then the conditional distribution of $\boldsymbol{y}$ is specified by

$$p(\boldsymbol{y}|\boldsymbol{x}) = [\frac{\tau(1-\tau)}{\sigma}]^n exp\{-\frac{1}{\sigma}\sum_{i=1}^{n} -\varphi_\tau(y_i - \rho_\tau \sum_{j=1}^{n} \omega_{ij} y_j - x_i^T \beta_\tau)\}. \tag{3}$$

Therefore, minimizing the Equation (2) is equivalent to maximizing the Equation (3). Followed by the location–scale mixture expression of asymmetric Laplace distribution [14], Equation (1) could be rewritten as

$$\begin{cases} y_i = \rho_\tau \sum\limits_{j=1}^{n} \omega_{ij} y_j + x_i^T \boldsymbol{\beta}_\tau + k_1 e_i + \sqrt{k_2 \sigma e_i} z_i, \\ e_i \sim \exp\{\frac{1}{\sigma}\}, \\ z_i \sim N(0,1), \\ i = 1, \cdots, n, \end{cases} \tag{4}$$

in which $e_i \sim \exp\{\frac{1}{\sigma}\}$ denotes that e_i follows an exponential distribution with parameter σ, whose probability density function is $p(e_i|\sigma) = \frac{1}{\sigma}\exp\{-\frac{1}{\sigma}e_i\}I(e_i \geq 0)$; z_i is the standard normal random variable, e_i and z_i are mutually independent; $k_1 = \frac{1-2\tau}{\tau(1-\tau)}$ and $k_2 = \frac{2}{\tau(1-\tau)}$, respectively. The models defined in Equation (4) are referred to as the SAR quantile models in the paper. For ease of notation, we will delete τ in the representation in the following.

3. A Bayesian Variable Selection Procedure for SAR Quantile Models

3.1. Prior Specifications

To perform Bayesian statistical inferences, it is necessary to specify the prior distributions. The spike and slab prior for $\boldsymbol{\beta}$ following references [21–23] is chosen to implement parameter estimation and variable selection in the framework of SAR quantile models in this paper. Specifically,

$$\pi(\boldsymbol{\beta}|\boldsymbol{\gamma}, \boldsymbol{\delta}) \sim N(\boldsymbol{0}_p, \boldsymbol{H}_\beta), \tag{5}$$

in which $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_p)^T$ is the indicator variable vector, $\boldsymbol{\delta} = (\delta_1^2, \ldots, \delta_p^2)^T$, $\boldsymbol{0}_p$ denotes a $p \times 1$ vector whose each element is 0, $\boldsymbol{H}_\beta = diag(\gamma_1\delta_1^2, \ldots, \gamma_p\delta_p^2)$. Or equivalently, for $k = 1, \cdots, p$,

$$\pi(\beta_k|\gamma_k, \delta_k^2) \sim N(0, \gamma_k\delta_k^2), \tag{6}$$

i.e.,

$$\pi(\beta_k|\gamma_k = 1, \delta_k^2) \propto (\delta_k^2)^{-\frac{1}{2}}\exp\left\{-\frac{\beta_k^2}{2\delta_k^2}\right\}, \pi(\beta_k = 0|\gamma_k = 0, \delta_k^2) = 1.$$

Similar to Kozumi and Kobayashi (2011), the inverse Gamma prior is specified for σ, that is

$$\pi(\sigma) \sim InvGamma(a_\sigma, b_\sigma), \tag{7}$$

where $InvGamma(a_\sigma, b_\sigma)$ denotes an inverse Gamma distribution with shape parameter a_σ and scale parameter b_σ, i.e.,

$$\pi(\sigma) \propto \sigma^{-(a_\sigma+1)} \exp\left\{-\frac{b_\sigma}{\sigma}\right\}.$$

Suppose that the prior distribution for ρ is as follows:

$$\pi(\rho) \sim U(\rho_{min}, \rho_{max}); \tag{8}$$

in which $U(\rho_{min}, \rho_{max})$ represents the uniform distribution on the interval (ρ_{min}, ρ_{max}).

The following prior distribution for $\boldsymbol{\gamma}$ is considered in this paper:

$$\pi(\boldsymbol{\gamma}) = \prod_{k=1}^{p} \pi(\gamma_k) = \prod_{k=1}^{p} q^{\gamma_k}(1-q)^{1-\gamma_k}.$$

In other words, for $k = 1, \cdots, p$, the prior distribution of γ_k involved in $\boldsymbol{\gamma}$ is assumed to be a Bernoulli distribution, that is

$$\pi(\gamma_k) = q^{\gamma_k}(1-q)^{1-\gamma_k}. \tag{9}$$

We set $q = \frac{1}{2}$ that represents uniform prior in all numerical examples. Some other prior specifications for $\boldsymbol{\gamma}$ could be found by Cripps et al. [26].

For $k = 1, \cdots, p$, the prior distribution of δ_k^2 involved in $\boldsymbol{\delta}$ is taken to be a inverse Gamma distribution, which is given by

$$\pi(\delta_k^2) \sim InvGamma(a_\delta, b_\delta). \tag{10}$$

In the end, a_σ, b_σ, ρ_{min}, ρ_{max}, a_δ, b_δ in the above prior specifications are hyperparameters whose values are already given. If not specified,we choose $a_\sigma = b_\sigma = 0.001$ $\rho_{min} = -1, \rho_{max} = 1$, $a_\delta = b_\delta = 0.001$ within all numerical examples, which may stand for a case with noninformative prior.

3.2. Gibbs Sampling and Probability Integral Transformation (PIT) Algorithm

Denoting $\boldsymbol{e} = (e_1, \ldots, e_n)^T$, a sequence of random samples is generated from the joint posterior distribution $p(\boldsymbol{\beta}, \sigma, \rho, \boldsymbol{\delta}, \boldsymbol{e}, \boldsymbol{\gamma}|\boldsymbol{y}, \boldsymbol{x})$ via the Gibbs sampler algorithm [24], and then parameter estimation and variable selection are simultaneously implemented by the obtained sequence of random draws. In our proposed algorithm, samplers$\{\boldsymbol{\beta}, \sigma, \rho, \boldsymbol{\delta}, \boldsymbol{e}, \boldsymbol{\gamma}\}$ are drawn iteratively from the following conditional posterior distributions: $p(\boldsymbol{\beta}|\sigma, \rho, \boldsymbol{\delta}, \boldsymbol{e}, \boldsymbol{\gamma}, \boldsymbol{y}, \boldsymbol{x})$, $p(\sigma|\boldsymbol{\beta}, \rho, \boldsymbol{\delta}, \boldsymbol{e}, \boldsymbol{\gamma}, \boldsymbol{y}, \boldsymbol{x})$, $p(\rho|\boldsymbol{\beta}, \sigma, \boldsymbol{\delta}, \boldsymbol{e}, \boldsymbol{\gamma}, \boldsymbol{y}, \boldsymbol{x})$, $p(\boldsymbol{\delta}|\boldsymbol{\beta}, \sigma, \rho, \boldsymbol{e}, \boldsymbol{\gamma}, \boldsymbol{y}, \boldsymbol{x})$, $p(\boldsymbol{e}|\boldsymbol{\beta}, \sigma, \rho, \boldsymbol{\delta}, \boldsymbol{\gamma}, \boldsymbol{y}, \boldsymbol{x})$, $p(\boldsymbol{\gamma}|\boldsymbol{\beta}, \sigma, \rho, \boldsymbol{\delta}, \boldsymbol{e}, \boldsymbol{y}, \boldsymbol{x})$. The corresponding conditional posterior distributions in performing Gibbs sampler algorithm are listed in the following.

(1) Sample β_k from conditional distribution $p(\beta_k|\sigma, \rho, \delta_k, \boldsymbol{e}, \gamma_k, \boldsymbol{y}, \boldsymbol{x})$, $k = 1, \cdots, p$, noting that according to Equations (4) and (6), we have

$$p(\beta_k = 0|\sigma, \rho, \delta_k, \boldsymbol{e}, \gamma_k = 0, \boldsymbol{y}, \boldsymbol{x}) = 1,\ p(\beta_k|\sigma, \rho, \delta_k, \boldsymbol{e}, \gamma_k = 1, \boldsymbol{y}, \boldsymbol{x}) \sim N(\beta_k^*, \xi_k^2), \tag{11}$$

where $\beta_k^* = \xi_k^2 \left\{ \frac{1}{k_2\sigma} \sum_{i=1}^{n} \frac{y_i - \mu_i^* - k_1 e_i}{e_i} x_{ik} \right\}$, $\xi_k^2 = \left\{ \frac{1}{\delta_k^2} + \frac{1}{k_2\sigma} \sum_{i=1}^{n} \frac{x_{ik}^2}{e_i} \right\}^{-1}$, and $\mu_i^* = \rho \sum_{j=1}^{n} \omega_{ij} y_j + \sum_{j=1, j\neq k}^{p} x_{ij}\beta_j$.

(2) Sample σ from conditional distribution $p(\sigma|\boldsymbol{\beta}, \rho, \boldsymbol{\delta}, \boldsymbol{e}, \boldsymbol{\gamma}, \boldsymbol{y}, \boldsymbol{x})$, and it can be shown to be

$$p(\sigma|\boldsymbol{\beta}, \rho, \boldsymbol{\delta}, \boldsymbol{e}, \boldsymbol{\gamma}, \boldsymbol{y}, \boldsymbol{x}) \sim InvGamma(a_\sigma^*, b_\sigma^*), \tag{12}$$

in which $a_\sigma^* = a_\sigma + \frac{3n}{2}$ and $b_\sigma^* = b_\sigma + \sum_{i=1}^{n} [\frac{(y_i - \mu_i - k_1 e_i)^2}{2k_2 e_i} + e_i]$, $\mu_i = \rho \sum_{j=1}^{n} \omega_{ij} y_j + \boldsymbol{x}_i^T \boldsymbol{\beta}$.

(3) Sample ρ from conditional distribution $p(\rho|\boldsymbol{\beta}, \sigma, \boldsymbol{\delta}, \boldsymbol{e}, \boldsymbol{\gamma}, \boldsymbol{y}, \boldsymbol{x})$. Through a simple algebra calculation, we have

$$p(\rho|\boldsymbol{\beta}, \sigma, \boldsymbol{\delta}, \boldsymbol{e}, \boldsymbol{\gamma}, \boldsymbol{y}, \boldsymbol{x}) \propto det(\boldsymbol{A})\exp\left\{-\frac{1}{2k_2\sigma}\sum_{i=1}^{n}\frac{(y_i - \mu_i - k_1 e_i)^2}{e_i}\right\} I\{\rho_{min} < \rho < \rho_{max}\}, \tag{13}$$

in which $\boldsymbol{A} = \boldsymbol{I}_n - \rho\boldsymbol{\omega}$, I_n denotes a $n \times n$ identity matrix, $det(\boldsymbol{A})$ denotes the determinant of $\boldsymbol{A}$, $\mu_i = \rho \sum_{j=1}^{n} \omega_{ij} y_j + \boldsymbol{x}_i^T \boldsymbol{\beta}$.

(4) Sample δ_k from conditional distribution $p(\delta_k^2|\beta_k, \gamma_k)$, $k = 1, \cdots, p$. Noting that

$$p(\delta_k^2|\beta_k, \gamma_k = 0) \sim InvGamma(a_\delta, b_\delta), \;\; p(\delta_k^2|\beta_k, \gamma_k = 1) \sim InvGamma(a_\delta^*, b_\delta^*), \tag{14}$$

where $a_\delta^* = a_\delta + \frac{1}{2}$, $b_\delta^* = b_\delta + \frac{\beta_k^2}{2}$.

(5) Sample e_i from conditional distribution $p(e_i|\boldsymbol{\beta}, \sigma, \rho, \boldsymbol{y}, \boldsymbol{x}_i)$, $i = 1, \cdots, n$. It can be shown that

$$p(e_i|\boldsymbol{\beta}, \sigma, \rho, \boldsymbol{y}, \boldsymbol{x}_i) \sim GIG(\frac{1}{2}, m_i^2, n_i^2), \tag{15}$$

where $m_i^2 = \frac{(y_i - \mu_i)^2}{k_2\sigma}$, $n_i^2 = \frac{k_1^2}{k_2\sigma} + \frac{2}{\sigma}$, $\mu_i = \rho \sum_{j=1}^{n} \omega_{ij} y_j + \boldsymbol{x}_i^T \boldsymbol{\beta}$, $GIG(\nu, m^2, n^2)$ denotes the generalized inverse Gauss distribution with parameters ν, m^2 and n^2, and $p(x) \sim GIG(\nu, m^2, n^2)$ if and only if

$$p(x) \propto x^{\nu-1}\exp\left\{-\frac{1}{2}(m^2 x^{-1} + n^2 x)\right\}.$$

(6) Sample γ_k from conditional distriution $p(\gamma_k|\boldsymbol{\beta}, \sigma, \rho, \boldsymbol{e}, \boldsymbol{y}, \boldsymbol{x})$, $k = 1, \cdots, p$. Noting that $p(\gamma_k|\boldsymbol{\beta}, \sigma, \rho, \boldsymbol{e}, \boldsymbol{y}, \boldsymbol{x})$ is a Bernoulli distribution with

$$p(\gamma_k = 1|\boldsymbol{\beta}, \sigma, \rho, \boldsymbol{e}, \boldsymbol{y}, \boldsymbol{x}) = \frac{1}{1+h}, p(\gamma_k = 0|\boldsymbol{\beta}, \sigma, \rho, \boldsymbol{e}, \boldsymbol{y}, \boldsymbol{x}) = \frac{h}{1+h}, \tag{16}$$

in which $h = h_1 \frac{\pi(\gamma_k=0)}{\pi(\gamma_k=1)}$ with $h_1 = \prod_{i=1}^{n} \frac{p(y_i|\boldsymbol{\beta},\sigma,\rho,e_i,\gamma_k=0,\boldsymbol{x}_i)}{p(y_i|\boldsymbol{\beta},\sigma,\rho,e_i,\gamma_k=1,\boldsymbol{x}_i)} \times \frac{\pi(\beta_k|\gamma_k=0,\delta_k^2)}{\pi(\beta_k|\gamma_k=1,\delta_k^2)}$

$$= \prod_{i=1}^{n} exp\{-\frac{1}{2}\frac{(y_i - \rho\sum_{j=1}^{n} W_{ij} y_j - \sum_{j\neq k} x_{ij}\beta_j - k_1 e_i)^2 - (y_i - \rho\sum_{j=1}^{n} W_{ij} y_j - \sum_{j=1}^{p} x_{ij}\beta_j - k_1 e_i)^2}{k_2\sigma e_i}\} \times \frac{1}{(2\pi\delta_k^2)^{-\frac{1}{2}} exp\{-\frac{\beta_k^2}{2\delta_k^2}\}}.$$

It is easily found that the conditional distributions (11), (12), (14). and (16) involved in the above Gibbs sampling method are lots of familiar distributions, such as the normal, inverse Gamma, and Bernoulli distribution, for example, whose sampling is fast and straightforward. In addition, there exist some efficient algorithms [27,28] to draw from the generalized inverse Gauss distribution (15). However, since conditional distribution (13) is nonstandard and unfamiliar distribution, it is rather difficult to draw directly observations for ρ. Hence, the probability integral transformation (PIT) algorithm [25], which is a sampling procedure that we recommend to use in applications (e.g., in our simulation study and Boston data analysis of Section 4), is used to draw observations from it, and the sampling detail is described within the Appendix A. Finally, the MCMC algorithm is summarized in the following Algorithm 1.

Algorithm 1: An MCMC-based sampling algorithm for the SAR quantile models

Input: setup initial values $\boldsymbol{\beta}^{(0)}, \sigma^{(0)}, \rho^{(0)}, \boldsymbol{e}^{(0)}, \boldsymbol{\delta}^{(0)}$, and $\boldsymbol{\gamma}^{(0)}$, and the number of iterations of the sampling algorithm T.

```
for t ← 1 : T do
    for k ← 1 : p do
        Sample β_k^(t) from Equation (11) ;
    end
    Sample σ^(t) from Equation (12) ;
    Sample ρ^(t) from Equation (13) according to PIT algorithm ;
    for k ← 1 : p do
        Sample δ_k^(t) from Equation (14) ;
    end
    for i ← 1 : n do
        Sample e_i^(t) from Equation (15) ;
    end
    for k ← 1 : p do
        Sample γ_k^(t) from Equation (16) ;
    end
end
```

Output: a sequence of samples $\{(\boldsymbol{\beta}^{(t)}, \sigma^{(t)}, \rho^{(t)}, \boldsymbol{\delta}^{(t)}, \boldsymbol{e}^{(t)}, \boldsymbol{\gamma}^{(t)}) : t = 1, \cdots, T\}$.

3.3. Bayesian Estimates and Standard Errors

Observations simulated from the above proposed MCMC algorihtm (Algorithm 1) could be employed to calculate the joint Bayesian estimates, standard errors of unknown parameters $\{\boldsymbol{\beta}, \sigma, \rho, \boldsymbol{\delta}\}$, and latent variables $\{\boldsymbol{e}, \boldsymbol{\gamma}\}$. Let $\{\boldsymbol{\beta}^{(t)}, \sigma^{(t)}, \rho^{(t)}, \boldsymbol{\delta}^{(t)}, \boldsymbol{e}^{(t)}, \boldsymbol{\gamma}^{(t)} : t = 1, 2, \cdots, \mathrm{T}\}$ be observations simulated from the joint posterior distribution $p(\boldsymbol{\beta}, \sigma, \rho, \boldsymbol{\delta}, \boldsymbol{e}, \boldsymbol{\gamma} | \boldsymbol{y}, \boldsymbol{x})$ via the proposed method after the algorithm converges. The joints Bayesian estimates (consistent estimates) of $\{\boldsymbol{\beta}, \sigma, \rho, \boldsymbol{\delta}, \boldsymbol{e}, \boldsymbol{\gamma}\}$ are caculated by their posterior sample mean [29],

$$\hat{\beta} = \frac{1}{T}\sum_{t=1}^{T}\beta^{(t)}, \hat{\sigma} = \frac{1}{T}\sum_{t=1}^{T}\sigma^{(t)}, \hat{\rho} = \frac{1}{T}\sum_{t=1}^{T}\rho^{(t)}, \hat{\delta} = \frac{1}{T}\sum_{t=1}^{T}\delta^{(t)}, \hat{e} = \frac{1}{T}\sum_{t=1}^{T}e^{(t)}, \hat{\gamma} = \frac{1}{T}\sum_{t=1}^{T}\gamma^{(t)}.$$

Similarly, the estimated standard errors of $\{\boldsymbol{\beta}, \sigma, \rho, \boldsymbol{\delta}, \boldsymbol{e}, \boldsymbol{\gamma}\}$ are obtained by their posterior sample standard errors.

4. Numerical Examples

In this section, several simulation studies and Boston housing price data analysis are employed to demonstrate the proposed Bayesian variable selection procedure.

4.1. Simulation Studies

The response y_i is generated from the following model

$$y_i = \rho\sum_{j=1}^{n}\omega_{ij}y_j + x_i^T\beta + \epsilon_i, \qquad i = 1, \cdots, n$$

with $n = 100$, where each p-dimensional x_i is independently generated from the normal distribution $N(0, \Sigma)$, in which the $(j,k)th$ element of Σ is $0.5^{|j-k|}$ for $j = 1, \cdots, p$ and $k = 1, \cdots, p$. The same as Chen et al. [30], we set $\omega_{1,2} = 1, \omega_{n,n-1} = 1$, and $\omega_{i,i-1} = \omega_{i,i+1} = 0.5$ for $i = 2, \cdots, n-1$, other elements of the spatial matrix ω are all zero. For spatial parameter, $\rho = -0.8, 0.0, 0.8$, which stand for three spatial dependences of the responses, are considered in each simulation setting, respectively. $\rho = -0.8$ represents negative and relatively strong spatial dependence, $\rho = 0.0$ represents independence, whereas

$\rho = 0.8$ represents positive and relatively strong spatial dependence. Similar to Alhamzawi et al. [31], the true values of β are taken into account in the following four scenarios:

Scenario 1: $\beta^{true} = (0.85, 0.85, 0.85, 0.85, 0.85, 0.85, 0.85, 0.85)^T$, which corresponds to the dense case;

Scenario 2: $\beta^{true} = (3, 1.5, 0, 0, 2, 0, 0, 0)^T$,which could be regarded as the sparse case;

Scenario 3: $\beta^{true} = (5, 0, 0, 0, 0, 0, 0, 0)^T$, which could be viewed as the very sparse case;

Scenario 4: $\beta^{true} = (5, 5, 5, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0)^T$ with $p = 18$, which may represent a moderate p case.

Within each scenario, we consider 4 different choices for the distribution of error term ϵ_i, such that the $\tau th(0 < \tau < 1)$ quantile is 0:

(a): The normal distribution $N(\mu, 1)$ (denoted as normal), with μ such that τth quantile of ϵ_i is 0;

(b): The Student's t distribution $t(3) + \mu$ (denoted as t), with μ such that τth quantile of ϵ_i is 0;

(c): The Laplace distribution $Laplace(\mu, 1)$ (denoted as Laplace),with μ such that τth quantile of ϵ_i is 0;

(d): The mixture of normal distribution $0.9N(\mu, 1) + 0.1N(\mu, 9)$ (denoted as mixed-normal), with μ such that τth quantile of ϵ_i is 0.

For each scenario of β and each choice of the error distribution, each spatial parameter ρ and with three different quantile levels $\tau = \{0.1, 0.3, 0.5\}$, we run 500 replications. In each replication, 5000 samples are collected to calculate Bayesian estimates of unknown parameters after 5000 burn-ins. Table 1 reports the mean and the root mean square error (denoted as RMSE in parentheses) of the Bayesian estimates based on 500 replications for scenario 1 setting. From Table 1, it is easy to see that (1) all the Bayesian estimates perform reasonably well, and the performance under normal error distributions is better than the performance under t, Laplace and mixed-normal error distributions in term of RMSEs; (2) the means of Bayesian estimates are quite close to their true values, which suggests that our proposed Bayesian quantile approach is very effective under different error distributions, different quantile levels, and different spatial parameters ρ. Tables 2–4 summarize the numerical results for Bayesian parametric estimations and variable selection simultaneously, based on 500 replications for scenarios 2–4, one for each scenario setting. Denoting MSE= $\frac{1}{p}\|\hat{\beta} - \beta^{true}\|_2^2$, MPE= $\frac{1}{n}\|X(\hat{\beta} - \beta^{true})\|_2^2$, and TP as the true positive number, which means the number of correctly identified zero coefficients.FP denotes the false positive number, which means the number of incorrectly identified non-zero coefficients. In our Bayesian statistical framework, the coefficient is identified as zero if the 95% credible interval of this parameter covers zero. Otherwise, it is identified as non-zero. Furthermore, FPR (false positive rate), TPR (true positive rate), and MCC (Matthews correlation coefficient) [32] are defined as follows: FPR= $\frac{\text{FP}}{\text{FP + TN}}$, TPR= $\frac{\text{TP}}{\text{TP + FN}}$, MCC= $\frac{\text{TP} \times \text{TN} - \text{FP} \times \text{FN}}{\sqrt{(\text{TP + FP})(\text{TP + FN})(\text{TN + FP})(\text{TN + FN})}}$, in which FN and TN are the number of incorrectly identified zero coefficients and the number of correctly identified non-zero coefficients, respectively. It is clear that MCC has a range from to -1 to 1 and models with MCC closer to 1 have higher selection accuracy. The reported TP values (denoted as TP), FP values (denoted as FP), FPR values (denoted as FPR), TPR values (denoted as FPR), and MCC values (denoted as MCC) in Tables 2–4 are averaged values over 500 replications. It is noted that the MSE is an estimation performance index, the MPE is a prediction performance index, and TP, FP, FPR, TPR and MCC are variable selection performance indexes. Examinations of Tables 2–4 indicate that (i) Most MSEs are not more than 0.02, which illustrates that the Bayesian estimation is robust and accurate; (ii) The vast majority of MPEs are not more than 0.25 and each MPE is smaller than 0.4, which demonstrates the predictive effect of our proposed approach is very good; (iii) Each TP is close to the number of true zero coefficients, each FP and FPR is close to 0, each TPR is close to 1. In particular, each MCC is bigger than 0.93 and most of them are clearly close to 1. All TPs, FPs, FPRs, TPRs and MCCs suggest that our proposed Bayesian variable selection procedure has extremely high selection accuracy. In short, the empirical performance (containing estimation performance, the predictive performance, and variable selection performance) of the Bayesian approach is quite satisfactory in our considered settings.

Table 1. Means and RMSEs (in parentheses) in the first simulation study under Scenario 1.

(ρ, τ)	Error Distribution	β_1	β_2	β_3	β_4	β_5	β_6	β_7	β_8	ρ
(−0.8, 0.1)	normal	0.8369	0.8319	0.8480	0.8314	0.8499	0.8226	0.8221	0.8462	−0.8000
		(0.2016)	(0.2547)	(0.2427)	(0.2642)	(0.2587)	(0.2487)	(0.2485)	(0.1981)	(0.0230)
	t	0.7582	0.8125	0.8311	0.8008	0.8400	0.7929	0.8454	0.7670	−0.8071
		(0.4365)	(0.5072)	(0.4791)	(0.4964)	(0.5178)	(0.4912)	(0.4828)	(0.4297)	(0.0294)
	Laplace	0.7904	0.8352	0.8287	0.8069	0.8333	0.8045	0.7962	0.8230	−0.8051
		(0.3827)	(0.4591)	(0.4604)	(0.4557)	(0.4597)	(0.4547)	(0.4665)	(0.3834)	(0.0307)
	mixed-normal	0.8201	0.8322	0.8116	0.8332	0.7974	0.8639	0.8432	0.8013	−0.8042
		(0.3098)	(0.3571)	(0.3724)	(0.3807)	(0.3927)	(0.3665)	(0.3621)	(0.3166)	(0.0280)
(−0.8, 0.3)	normal	0.8391	0.8369	0.8555	0.8398	0.8458	0.8372	0.8478	0.8359	−0.7983
		(0.1601)	(0.1967)	(0.2070)	(0.2012)	(0.1861)	(0.1916)	(0.1756)	(0.1662)	(0.0183)
	t	0.7751	0.8372	0.7963	0.8516	0.8113	0.8290	0.8070	0.7978	−0.7987
		(0.3294)	(0.3864)	(0.3953)	(0.3950)	(0.3874)	(0.3959)	(0.4080)	(0.3270)	(0.0223)
	Laplace	0.8111	0.8388	0.8363	0.8168	0.8194	0.8525	0.8250	0.8303	−0.7994
		(0.2532)	(0.3002)	(0.3208)	(0.3046)	(0.3186)	(0.3059)	(0.3077)	(0.2569)	(0.0214)
	mixed-normal	0.8257	0.8191	0.8297	0.8477	0.8438	0.8346	0.8199	0.8504	−0.7996
		(0.2108)	(0.2821)	(0.2728)	(0.2707)	(0.2657)	(0.2695)	(0.2572)	(0.2080)	(0.0196)
(−0.8, 0.5)	normal	0.8411	0.8372	0.8457	0.8412	0.8526	0.8319	0.8535	0.8339	−0.7987
		(0.1521)	(0.1883)	(0.1916)	(0.1914)	(0.1918)	(0.1828)	(0.1733)	(0.1532)	(0.0166)
	t	0.7882	0.8526	0.8044	0.8218	0.8176	0.8091	0.8350	0.8034	−0.7983
		(0.2980)	(0.3565)	(0.3654)	(0.3601)	(0.3619)	(0.3609)	(0.3566)	(0.2923)	(0.0209)
	Laplace	0.8306	0.8165	0.8422	0.8199	0.8530	0.8197	0.8458	0.8100	−0.7976
		(0.2471)	(0.2945)	(0.3039)	(0.3017)	(0.2992)	(0.3015)	(0.2924)	(0.2559)	(0.0199)
	mixed-normal	0.8302	0.8302	0.8558	0.8419	0.8248	0.8196	0.8561	0.8156	−0.7977
		(0.2138)	(0.2732)	(0.2636)	(0.2802)	(0.2766)	(0.2686)	(0.2531)	(0.2225)	(0.0186)
(0.0, 0.1)	normal	0.8412	0.8332	0.8415	0.8474	0.8255	0.8497	0.8438	0.8381	−0.0001
		(0.2005)	(0.2427)	(0.2399)	(0.2431)	(0.2305)	(0.2289)	(0.2497)	(0.2117)	(0.0460)
	t	0.7865	0.8152	0.8114	0.8114	0.7895	0.8298	0.8302	0.7691	−0.0451
		(0.4391)	(0.4947)	(0.5027)	(0.4993)	(0.5087)	(0.4930)	(0.5137)	(0.4457)	(0.0881)
	Laplace	0.7865	0.8339	0.8234	0.7975	0.8307	0.8015	0.8369	0.7908	−0.0281
		(0.3918)	(0.4583)	(0.4490)	(0.4734)	(0.4680)	(0.4726)	(0.4675)	(0.4105)	(0.0796)
	mixed-normal	0.8045	0.8504	0.8356	0.8246	0.8588	0.8054	0.8426	0.7967	−0.0207
		(0.3248)	(0.3640)	(0.3983)	(0.3839)	(0.3667)	(0.3859)	(0.3915)	(0.3202)	(0.0681)

Table 1. *Cont.*

(ρ, τ)	Error Distribution	β_1	β_2	β_3	β_4	β_5	β_6	β_7	β_8	ρ
(0.0, 0.3)	normal	0.8328	0.8357	0.8545	0.8346	0.8420	0.8493	0.8447	0.8362	−0.0044
		(0.1590)	(0.1831)	(0.1939)	(0.2020)	(0.1921)	(0.1922)	(0.1867)	(0.1636)	(0.0418)
	t	0.8013	0.8135	0.8215	0.8412	0.8047	0.8242	0.8170	0.8180	−0.0079
		(0.3155)	(0.3802)	(0.3714)	(0.3657)	(0.3805)	(0.3702)	(0.3655)	(0.2912)	(0.0507)
	Laplace	0.8177	0.8242	0.8243	0.8263	0.8300	0.8242	0.8466	0.8166	0.0065
		(0.2483)	(0.2955)	(0.2960)	(0.3086)	(0.3059)	(0.3073)	(0.3115)	(0.2566)	(0.0475)
	mixed-normal	0.8145	0.8346	0.8380	0.8334	0.8460	0.8004	0.8492	0.8425	−0.0025
		(0.2218)	(0.2573)	(0.2751)	(0.2698)	(0.2676)	(0.2945)	(0.2812)	(0.2050)	(0.0445)
(0.0, 0.5)	normal	0.8494	0.8303	0.8361	0.8553	0.8271	0.8478	0.8422	0.8343	−0.0008
		(0.1443)	(0.1794)	(0.1834)	(0.1794)	(0.1808)	(0.1740)	(0.1767)	(0.1439)	(0.0392)
	t	0.7926	0.8107	0.8244	0.8465	0.8161	0.8245	0.8140	0.8035	0.0005
		(0.3105)	(0.3761)	(0.3617)	(0.3550)	(0.3684)	(0.3702)	(0.3716)	(0.3053)	(0.0484)
	Laplace	0.8204	0.8404	0.8156	0.8341	0.8302	0.8443	0.8192	0.8125	0.0023
		(0.2315)	(0.2807)	(0.3050)	(0.2964)	(0.2959)	(0.2831)	(0.2835)	(0.2473)	(0.0431)
	mixed-normal	0.8163	0.8446	0.8002	0.8564	0.8187	0.8487	0.8237	0.8332	−0.0016
		(0.2289)	(0.2833)	(0.2940)	(0.2660)	(0.2601)	(0.2637)	(0.2646)	(0.2113)	(0.0459)
(0.8, 0.1)	normal	0.8383	0.8459	0.8479	0.8249	0.8435	0.8583	0.8352	0.8232	0.7948
		(0.2102)	(0.2520)	(0.2507)	(0.2570)	(0.2413)	(0.2541)	(0.2852)	(0.2145)	(0.0193)
	t	0.7648	0.8402	0.8513	0.8168	0.8302	0.8364	0.8152	0.7768	0.7709
		(0.4449)	(0.5331)	(0.5462)	(0.5576)	(0.5227)	(0.5219)	(0.5349)	(0.4874)	(0.0454)
	Laplace	0.7888	0.8551	0.8177	0.8296	0.8268	0.8154	0.8438	0.7945	0.7784
		(0.4253)	(0.5056)	(0.4861)	(0.5132)	(0.4958)	(0.4921)	(0.4873)	(0.4101)	(0.0372)
	mixed-normal	0.8366	0.7999	0.8658	0.8185	0.8627	0.8195	0.8190	0.8373	0.7799
		(0.3547)	(0.4346)	(0.4389)	(0.4130)	(0.4107)	(0.4155)	(0.4364)	(0.3501)	(0.0354)
(0.8, 0.3)	normal	0.8430	0.8464	0.8409	0.8402	0.8681	0.8325	0.8422	0.8449	0.7965
		(0.1668)	(0.1982)	(0.2092)	(0.1924)	(0.1869)	(0.1854)	(0.1997)	(0.1582)	(0.0177)
	t	0.7955	0.8303	0.8407	0.8037	0.8409	0.8242	0.8230	0.7939	0.7913
		(0.3270)	(0.3802)	(0.3792)	(0.3971)	(0.3943)	(0.3892)	(0.3948)	(0.3164)	(0.0256)
	Laplace	0.8401	0.8158	0.8391	0.8429	0.8335	0.8347	0.8201	0.8199	0.7923
		(0.2574)	(0.3410)	(0.3517)	(0.3331)	(0.3323)	(0.3267)	(0.3311)	(0.2650)	(0.0243)
	mixed-normal	0.8419	0.8274	0.8437	0.8304	0.8361	0.8340	0.8509	0.8132	0.7937
		(0.2277)	(0.2861)	(0.2992)	(0.2900)	(0.2976)	(0.2758)	(0.2732)	(0.2464)	(0.0203)
(0.8, 0.5)	normal	0.8412	0.8429	0.8500	0.8343	0.8501	0.8409	0.8505	0.8453	0.7976
		(0.1452)	(0.1804)	(0.1808)	(0.1924)	(0.1822)	(0.1771)	(0.1828)	(0.1544)	(0.0173)
	t	0.8072	0.8157	0.8202	0.8123	0.8473	0.8138	0.8550	0.7780	0.7969
		(0.2930)	(0.3710)	(0.3803)	(0.3685)	(0.3664)	(0.3736)	(0.3623)	(0.3187)	(0.0212)
	Laplace	0.8312	0.8152	0.8367	0.8316	0.8404	0.8206	0.8328	0.8206	0.7967
		(0.2235)	(0.2957)	(0.2857)	(0.2864)	(0.3060)	(0.2876)	(0.2965)	(0.2380)	(0.0189)
	mixed-normal	0.8239	0.8228	0.8493	0.8192	0.8457	0.8327	0.8359	0.8340	0.7985
		(0.2262)	(0.2553)	(0.2585)	(0.2716)	(0.2633)	(0.2686)	(0.2619)	(0.2157)	(0.0184)

Table 2. Numerical results of the simulation study under Scenario 2.

(ρ, τ)	Error Distribution	MSE	MPE	TP	FP	FPR	TPR	MCC
(−0.8, 0.1)	normal	0.0154	0.0962	4.9340	0.0060	0.0020	0.9868	0.9810
	t	0.0381	0.2326	4.9300	0.0080	0.0027	0.9860	0.9794
	Laplace	0.0388	0.2339	4.9020	0.0060	0.0020	0.9804	0.9727
	mixed-normal	0.0215	0.1301	4.9600	0.0060	0.0020	0.9920	0.9878
(−0.8, 0.3)	normal	0.0077	0.0482	4.9880	0.006	0.0020	0.9976	0.9952
	t	0.0111	0.0721	5.000	0.008	0.0027	1.0000	0.9979
	Laplace	0.0098	0.0619	4.9980	0.0040	0.0013	0.9996	0.9984
	mixed-normal	0.0095	0.0631	5.0000	0.0040	0.0013	1.0000	0.9989
(−0.8, 0.5)	normal	0.0072	0.0467	4.9980	0.004	0.0013	0.9996	0.9984
	t	0.0088	0.0560	5.0000	0.006	0.0020	1.0000	0.9984
	Laplace	0.0073	0.0470	5.0000	0.0020	0.0007	1.0000	0.9995
	mixed-normal	0.0085	0.0528	5.0000	0.0020	0.0007	1.0000	0.9995
(0.0, 0.1)	normal	0.0134	0.0806	4.9580	0.0000	0.0000	0.9916	0.9889
	t	0.0330	0.1985	4.9560	0.0060	0.0020	0.9912	0.9868
	Laplace	0.0359	0.2187	4.9140	0.0040	0.0013	0.9828	0.9764
	mixed-normal	0.0219	0.1306	4.9400	0.0020	0.0007	0.9880	0.9836
(0.0, 0.3)	normal	0.0073	0.0457	4.9900	0.0000	0.0000	0.9980	0.9973
	t	0.0099	0.0607	4.9980	0.0000	0.0000	0.9996	0.9995
	Laplace	0.0097	0.0591	4.9980	0.0000	0.0000	0.9996	0.9995
	mixed-normal	0.0085	0.0531	4.9960	0.0000	0.0000	0.9992	0.9989
(0.0, 0.5)	normal	0.0063	0.0399	4.9980	0.0000	0.0000	0.9996	0.9995
	t	0.0081	0.0505	4.9980	0.0000	0.0000	0.9996	0.9995
	Laplace	0.0070	0.0426	5.0000	0.0000	0.0000	1.0000	1.0000
	mixed-normal	0.0073	0.0452	4.9980	0.0000	0.0000	0.9996	0.9995
(0.8, 0.1)	normal	0.0143	0.0890	4.940	0.0080	0.0027	0.9880	0.9820
	t	0.0405	0.2325	4.9580	0.0100	0.0033	0.9916	0.9862
	Laplace	0.0410	0.2413	4.9120	0.0080	0.0027	0.9824	0.9748
	mixed-normal	0.0259	0.1494	4.9340	0.0080	0.0027	0.9868	0.9805
(0.8, 0.3)	normal	0.0084	0.0543	4.996	0.008	0.0027	0.9992	0.9968
	t	0.0112	0.0702	5.0000	0.006	0.0020	1.0000	0.9984
	Laplace	0.0097	0.0618	5.0000	0.0040	0.0013	1.0000	0.9989
	mixed-normal	0.0098	0.0602	4.9960	0.0080	0.0027	0.9992	0.9968
(0.8, 0.5)	normal	0.0067	0.0426	4.994	0.0020	0.0007	0.9988	0.9979
	t	0.0093	0.0585	1.0000	0.0060	0.0020	1.0000	0.9984
	Laplace	0.0073	0.0476	5.0000	0.0060	0.0020	1.0000	0.9984
	mixed-normal	0.0088	0.0547	5.0000	0.0040	0.0013	1.0000	0.9989

Table 3. Numerical results in the simulation study under Scenario 3.

(ρ, τ)	Error Distribution	MSE	MPE	TP	FP	FPR	TPR	MCC
(−0.8, 0.1)	normal	0.0078	0.0557	6.9060	0.0020	0.0020	0.9866	0.9484
	t	0.0183	0.1309	6.9060	0.0040	0.0040	0.9866	0.9472
	Laplace	0.0185	0.1342	6.8840	0.0020	0.0020	0.9834	0.9375
	mixed-normal	0.0111	0.0771	6.9040	0.0000	0.0000	0.9863	0.9486
(−0.8,0.3)	normal	0.0029	0.0224	6.9940	0.0000	0.0000	0.9991	0.9966
	t	0.0034	0.0265	7.0000	0.0020	0.0020	1.0000	0.9989
	Laplace	0.0035	0.0272	6.9960	0.0020	0.0020	0.9994	0.9966
	mixed-normal	0.0043	0.0310	6.9980	0.0040	0.0040	0.9997	0.9966
(−0.8, 0.5)	normal	0.0025	0.0189	6.9900	0.0002	0.0002	0.9986	0.9932
	t	0.0028	0.0219	6.9980	0.0020	0.0020	0.9997	0.9977
	Laplace	0.0026	0.0191	6.9980	0.0020	0.0020	0.9997	0.9977
	mixed-normal l	0.0032	0.0241	6.9960	0.0020	0.0020	0.9994	0.9966
(0.0, 0.1)	normal	0.0054	0.0405	6.9180	0.0000	0.0000	0.9883	0.9557
	t	0.0165	0.1197	6.8820	0.0000	0.0000	0.9831	0.9378
	Laplace	0.0176	0.1265	6.8700	0.0000	0.0000	0.9814	0.9319
	mixed-normal	0.0100	0.0717	6.9280	0.0000	0.0000	0.9897	0.9609
(0.0, 0.3)	normal	0.0021	0.0163	6.9980	0.0000	0.0000	0.9997	0.9989
	t	0.0031	0.0238	6.9980	0.0000	0.0000	0.9997	0.9989
	Laplace	0.0024	0.0188	7.0000	0.0000	0.0000	1.0000	1.0000
	mixed-normal	0.0030	0.0231	6.9860	0.0000	0.0000	0.9980	0.9921
(0.0, 0.5)	normal	0.0019	0.0150	6.9940	0.0000	0.0000	0.9991	0.9966
	t	0.0020	0.0155	7.0000	0.0000	0.0000	1.0000	1.0000
	Laplace	0.0020	0.0159	6.9960	0.0000	0.0000	0.9994	0.9977
	mixed-normal	0.0023	0.0179	6.9960	0.0000	0.0000	0.9994	0.9977
(0.8, 0.1)	normal	0.0071	0.0520	6.9320	0.0040	0.0040	0.9903	0.9606
	t	0.0186	0.1334	6.9240	0.0020	0.0020	0.9891	0.9576
	Laplace	0.0204	0.1467	6.8760	0.0020	0.0020	0.9823	0.9336
	nixed-normal	0.0129	0.0912	6.9060	0.0040	0.0040	0.9866	0.9472
(0.8, 0.3)	normal	0.0032	0.0246	7.0000	0.0020	0.0020	1.0000	0.9989
	t	0.0043	0.0328	6.9980	0.0000	0.0000	0.9997	0.9989
	Laplace	0.0035	0.0273	6.9960	0.0040	0.0040	0.9994	0.9954
	mixed-normal	0.0037	0.0286	6.9880	0.0040	0.0040	0.9983	0.9909
(0.8, 0.5)	normal	0.0024	0.0189	6.9960	0.0000	0.0000	0.9994	0.9977
	t	0.0028	0.0216	6.9980	0.0020	0.0020	0.9997	0.9977
	Laplace	0.0028	0.0207	7.0000	0.0020	0.0020	1.0000	0.9989
	mixed-normal	0.0034	0.0253	6.9940	0.0020	0.0020	0.9991	0.9954

Table 4. Numerical results in the simulation study under Scenario 4.

(ρ,τ)	Error Distribution	MSE	MPE	TP	FP	FPR	TPR	MCC
$(-0.8,0.1)$	normal	0.0203	0.3328	14.9280	0.0900	0.0300	0.9952	0.9675
	t	0.0260	0.3732	14.8820	0.0600	0.0200	0.9921	0.9647
	Laplace	0.0276	0.3905	14.8300	0.0540	0.0180	0.9887	0.9561
	mixed-normal	0.0218	0.3380	14.9300	0.0740	0.0247	0.9953	0.9712
$(-0.8, 0.3)$	normal	0.0117	0.1825	14.9900	0.0580	0.01933	0.9993	0.9864
	t	0.0132	0.2111	14.9840	0.0400	0.0133	0.9989	0.9888
	Laplace	0.0133	0.2131	15.0000	0.0480	0.0160	1.0000	0.9904
	mixed-normal	0.0121	0.1924	14.9960	0.0560	0.0187	0.9997	0.9880
$(-0.8, 0.5)$	normal	0.0103	0.1667	14.9980	0.0500	0.0167	0.9999	0.9896
	t	0.0114	0.1733	14.9980	0.0460	0.0153	0.9999	0.9904
	Laplace	0.0114	0.1797	14.9980	0.0460	0.0153	0.9999	0.9904
	mixed-normal	0.0114	0.1753	14.9980	0.0380	0.0127	0.9999	0.9920
$(0.0, 0.1)$	normal	0.0078	0.1000	14.8640	0.0000	0.0000	0.9909	0.9736
	t	0.0203	0.2629	14.8320	0.0040	0.0013	0.9888	0.9668
	Laplace	0.0199	0.2609	14.8420	0.0000	0.0000	0.9895	0.9695
	mixed-normal	0.0134	0.1735	14.8540	0.0000	0.0000	0.9903	0.9718
$(0.0, 0.3)$	normal	0.0043	0.0550	14.9640	0.0000	0.0000	0.9976	0.9929
	t	0.0054	0.0662	14.9920	0.0000	0.0000	0.9995	0.9984
	Laplace	0.0053	0.0638	14.9960	0.0000	0.0000	0.9997	0.9992
	mixed-normal	0.0053	0.0655	14.9820	0.0000	0.0000	0.9988	0.9964
$(0.0, 0.5)$	normal	0.0034	0.0427	14.9900	0.0000	0.0000	0.9993	0.9980
	t	0.0046	0.0551	15.0000	0.0000	0.0000	1.0000	1.0000
	Laplace	0.0037	0.0435	14.9980	0.0000	0.0000	0.9999	0.9996
	mixed-normal	0.0043	0.0524	14.9880	0.0000	0.0000	0.9992	0.9976
$(0.8, 0.1)$	normal	0.0183	0.2771	14.9040	0.0820	0.0273	0.9936	0.9645
	t	0.0272	0.3877	14.8720	0.0560	0.0187	0.9915	0.9636
	Laplace	0.0279	0.3719	14.8140	0.0500	0.0167	0.9876	0.9540
	mixed-normal	0.0223	0.3175	14.8980	0.0540	0.0180	0.9932	0.9690
$(0.8, 0.3)$	normal	0.0118	0.1848	14.9900	0.0480	0.0160	0.9993	0.9884
	t	0.0121	0.1948	14.9980	0.0480	0.0160	0.9999	0.9900
	Laplace	0.0120	0.1965	14.9880	0.0460	0.0153	0.9992	0.9884
	mixed-normal	0.0125	0.1950	14.9980	0.0540	0.0180	0.9999	0.9888
$(0.8, 0.5)$	normal	0.0106	0.1781	14.9860	0.0500	0.0167	0.9991	0.9872
	t	0.0116	0.1853	14.9940	0.0580	0.0193	0.9996	0.9872
	Laplace	0.0111	0.1622	14.9980	0.0420	0.0140	0.9999	0.9912
	mixed-normal	0.0120	0.1851	15.000	0.0560	0.0187	1.0000	0.9888

4.2. Boston Housing Price Data Analysis

In this subsection, a real example relating to Boston housing price data, which was analyzed by Harrison and Rubinfeld [33], Du et al. [3], Liu et al. [4], Xie et al. [5] and many other authors, is adopted to illustrate the proposed Bayesian methodologies. The Boston housing price dataset contains 14 variables with 506 individuals, which cab be downloaded from the link: http://lib.stat.cmu.edu/datasets/boston (accessed on 18 July

2022); a detailed description of all the variables is summarized in Table 5. Our scientific interest is to investigate the relationship between the house price and the other variables in this study, while accounting for choosing several important variables to explain home price under the SAR quantile models. Based on the work of Harrison and Rubinfeld [33] and Liu et al. [4], log(MEDV) is treated as the response variable, and the other variables are taken as explanatory variables, where DIS, RAD, LSTAT are processed with the logarithm, and RM and NOX are processed with the square. For convenient analysis, all the variables are standardized in the paper such that their sample means become zero. Finally, all 14 variables are taken into account by the following SAR quantile model

$$\begin{cases} y_i = \rho \sum\limits_{j=1}^{n} \omega_{ij} y_j + \sum\limits_{k=1}^{p} x_{ik}\beta_k + k_1 e_i + \sqrt{k_2 \sigma e_i} z_i, \\ p(e_i|\sigma) = \frac{1}{\sigma}\exp\{-\frac{1}{\sigma}e_i\} I(e_i \geq 0), \\ z_i \sim N(0,1), \\ i = 1, \cdots, n, \end{cases}$$

where $n = 506$, $p = 13$, $k_1 = \frac{1-2\tau}{\tau(1-\tau)}$ and $k_2 = \frac{2}{\tau(1-\tau)}$, respectively. The response variable y_i denotes MEDV and the other 13 explanatory variables are x_{i1} (CRIM), x_{i2} (ZN), x_{i3} (INDUS), x_{i4} (CHAS), x_{i5} (NOX), Vx_{i6} (RM), x_{i7} (AGE), x_{i8} (DIS), x_{i9} (RAD), x_{i10} (TAX), x_{i11} (PTRATIO), x_{i12} (B), and x_{i13} (LSTAT), respectively. Moreover, similar to Ertur and KochGrowth [34], the initial spatial weight matrix $\omega^* = (\omega^*_{ij})_{n\times n}$ is considered by the following

$$\omega^*_{ij} = \begin{cases} 0, & \text{i=j}, \\ e^{-2d_{ij}/1000}, & \text{i} \neq \text{j}, \end{cases}$$

d_{ij} denotes the great-circle distance between the latitude and longitude coordinates of any two houses. Subsequently, for $i = 1, \cdots, n$, we set $\omega_{ii} = 0$ and the initial weights ω^*_{ij} are normalized such that $\sum\limits_{j=1}^{n} \omega_{ij} = 1$. The prior distributions and hyperparameters are specified as presented in Section 3.

Table 5. Description of the variables in Boston housing price data.

Variables	Description
CRIM	per capita crime rate by town
ZN	proportion of residential land zones for lots over 25,000 sq.ft
INDUS	proportion of non-retail business acres per town
CHAS	Charles River dummy variable
NOX	nitric oxides concentration
RM	average number of rooms per dwelling
AGE	proportion of owner-occupied units built prior to 1940
DIS	weighted distances to five Boston employment centers
RAD	index of accessibility to radial highways
TAX	full-value property-tax rate per USD 10,000
PTRATIO	pupil–teacher ratio by town
B	1000(B-0.63)^2 where B is the proportion of Black people by town
LSTAT	percentage of lower status of the population
MEDV	the median value of owner-occupied homes in USD 1000s

A total of 5000 posterior samplers after 5000 burn-ins are collected in the posterior analysis, Bayesian estimates (ESTs), standard error estimates (SEs). and 95% credible intervals (CIs) of the unknown parameters under Gibbs sampling with the PIT algorithm and 3 different quantile levels ($\tau = 0.2, 0.5, 0.8$) are reported in Table 6. From these empirical results, it can be seen that the Bayesian variable selection method could simultaneously obtain robust parameter estimations and identify important explanatory variables under different quantile levels ($\tau = 0.2, 0.5, 0.8$). For example, under $\tau = 0.5$, x_{i1}, x_{i5}, x_{i8}, x_{i11}, and x_{i13} are identified to be important explanatory variables with a significantly negative effect on MEDV, and x_{i6} and x_{i12} are detected to be important explanatory variables with a significantly positive impact on MEDV, since their corresponding 95% CIs do not cover zero; while x_{i2}, x_{i3}, x_{i4}, x_{i7}, x_{i9}, and x_{i10} appear to be insignificant at significance level 0.05 because their 95% CIs cover zero.

Table 6. Bayesian estimation results based on SAR quantile models in the Boston housing price data analysis.

τ	Parameters	EST	SE	95% CI
0.2	β_1	−0.3247	0.0385	(−0.4002, −0.2491)
	β_2	−0.0007	0.0074	(−0.0153, 0.0138)
	β_3	0.0019	0.0099	(−0.0174, 0.0213)
	β_4	0.0057	0.0004	(0.0050, 0.0064)
	β_5	−0.1108	0.0251	(−0.1600, −0.0616)
	β_6	0.2518	0.0313	(0.1905, 0.3131)
	β_7	−0.0101	0.0230	(−0.0552, 0.0350)
	β_8	−0.1833	0.0333	(−0.2487, −0.1180)
	β_9	0.1001	0.0367	(0.0281, 0.1721)
	β_{10}	−0.0913	0.0533	(−0.1958, 0.0131)
	β_{11}	−0.0948	0.0175	(−0.1291, −0.0606)
	β_{12}	0.1072	0.0203	(0.0674, 0.1471)
	β_{13}	−0.3307	0.0387	(−0.4067, −0.2548)
	ρ	0.3510	0.0372	(0.2781, 0.4238)
0.5	β_1	−0.1570	0.0217	(−0.1996, −0.1144)
	β_2	−0.0002	0.0032	(−0.0065, 0.0060)
	β_3	−0.0004	0.0051	(−0.0104, 0.0097)
	β_4	−0.0000	0.0001	(−0.0001, 0.0001)
	β_5	−0.0954	0.0312	(−0.1567, −0.0342)
	β_6	0.2816	0.0223	(0.2378, 0.3254)
	β_7	−0.0061	0.0170	(−0.0394, 0.0273)
	β_8	−0.1764	0.0306	(−0.2364, −0.1164)
	β_9	0.0466	0.0512	(−0.0537, 0.1468)
	β_{10}	−0.0405	0.0510	(−0.1405, 0.0594)
	β_{11}	−0.0986	0.0184	(−0.1347, −0.0626)
	β_{12}	0.0998	0.0184	(0.0636, 0.1359)
	β_{13}	−0.3132	0.0284	(−0.3688, −0.2575)
	ρ	0.3901	0.0406	(0.3105, 0.4696)
0.8	β_1	−0.1267	0.0368	(−0.1989, −0.0546)
	β_2	0.0042	0.0132	(−0.0217, 0.0301)
	β_3	−0.0008	0.0095	(−0.0195, 0.0179)
	β_4	−0.0063	0.0004	(−0.0071, −0.0055)
	β_5	−0.1563	0.0306	(−0.2162, −0.0964)
	β_6	0.2351	0.0287	(0.1788, 0.2912)
	β_7	−0.0085	0.0233	(−0.0542, 0.0372)
	β_8	−0.2680	0.0323	(−0.3313, −0.2047)
	β_9	0.0651	0.0374	(−0.0082, 0.1384)
	β_{10}	−0.0016	0.0098	(−0.0209, 0.0176)
	β_{11}	−0.1245	0.0223	(−0.1682, −0.0807)
	β_{12}	0.1179	0.0234	(0.0720, 0.1638)
	β_{13}	−0.3722	0.0366	(−0.4439, −0.3006)
	ρ	0.3602	0.0480	(0.2662, 0.4542)

Furthermore, we obtain parameter estimates based on a Bayesian estimation procedure for SAR models proposed by LeSage and Pace [2], and calculate the predictive root mean square error (PRMSE), which is defined as PRMSE $= \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\boldsymbol{y}_i - \hat{\boldsymbol{y}}_i)^2}$,where $\hat{\boldsymbol{y}}_i$ is the mean of $\{y_i^{(j)}, j = 1, \ldots, 5000\}$, and $y_i^{(j)}$ is the predicted value of y_i in the jth iteration after 5000 burn-in iterations. The related computing results are reported in Table 7, and the PRMSE values corresponding to our proposed Bayesian method ($\tau = 0.5$, which is the case with median regression) and the method proposed by LeSage and Pace are given by 0.4204 and 0.5136, respectively. Comparing these results with Table 3, we may find that: (1) the regression parameter $\boldsymbol{\beta}$ and spatial parameter ρ vary with different quantile levels (e.g., $\tau = 0.2, 0.5$ and 0.8) in SAR quantile models, which implies the way that the covariates affect the MEDV (response variable) is different at different levels of the distribution of the MEDV, the same as to the spatial relationship (the latitude and longitude coordinates) of any two houses. In a word, compared with the Bayesian estimation procedure for SAR models proposed by LeSage and Pace (in general, ordinal mean regression methods), our proposed SAR quantile models and methodologies could provide a more comprehensive and complete description of the Boston housing price data structure; (2) in terms of estimation and prediction performance based on the SAR model of LeSage and Pace and our proposed SAR quantile mothod ($\tau = 0.5$), our median regression performs better than LeSage and Pace's mean regression, since our proposed method has smaller standard errors and smaller PRMSE than those obtained by the method proposed by LeSage and Pace.

Table 7. Bayesian estimation results based on SAR models [2] in the Boston housing price data analysis.

Parameters	EST	SE	95% CI
β_1	−0.2046	0.0244	(−0.2531, −0.1573)
β_2	0.0329	0.0262	(−0.0177, 0.0848)
β_3	0.0050	0.0361	(−0.0663, 0.0764)
β_4	−0.0001	0.0004	(−0.0014, 0.0002)
β_5	−0.1677	0.0353	(−0.2360, −0.0995)
β_6	0.1415	0.0257	(0.0897, 0.1914)
β_7	0.0171	0.0332	(−0.0477, 0.0826)
β_8	−0.2824	0.0398	(−0.3601, −0.2054)
β_9	0.1723	0.0376	(0.0987, 0.2456)
β_{10}	−0.1023	0.0467	(−0.1937, −0.0087)
β_{11}	−0.0991	0.0250	(−0.1487, −0.0513)
β_{12}	0.0710	0.0210	(0.0301, 0.1124)
β_{13}	−0.4473	0.0347	(−0.5147, −0.3797)
ρ	0.4532	0.0436	(0.3676, 0.5367)
σ^2	0.2530	0.0102	(0.2339, 0.2737)

5. Discussion

A Bayesian quantile regression method for spatial autoregressive (SAR) models is presented in this paper. Furthermore, an efficient MCMC algorithm is elaborately designed for posterior inferences. Note that our proposed approach could perform simultaneously to obtain robust parameter estimations, to identify significant explanatory variables, and to build accurate predictive models in the context of SAR quantile models, as shown by our empirical numerical studies. Therefore, we strongly recommend using our proposed Bayesian procedure in spatial data analysis.

In spite of the excellent performances of our proposed approach, it also suffers from some limitations in application: (1) the linear relationship of the response variable and explanatory variables in SAR quantile models might not hold; (2) the quantile curves related to different quantile levels are fitted separately in the current version, and they might cross (violating the definition of quantiles); (3) the stability of our proposed approach is not good enough when quantile levels are very close to 0 and 1 (e.g., $\tau < 10^{-6}, \tau > 1 - 10^{-6}$); (4) in high dimensional or ultrahigh dimensional spatial data analysis, the performances of parametric estimation and variable selection for our proposed method are rather poor. The main contribution of this paper is to adopt the spike and slab prior method and the

Bayesian quantile technique [17,18] to SAR models, then similar ideas could be further extended to other spatial regression models in future works. Furthermore, it is a potential future project to consider more robust Bayesian quantile methods (e.g., Bayesian composite quantile regression), and more advanced Bayesian variable selection techniques for SAR models in high dimensional or ultrahigh dimensional settings.

Author Contributions: Conceptualization, D.X.; methodology, Y.Z.; software, Y.Z. and D.X.; data curation, Y.Z. and D.X.; formal analysis, Y.Z. and D.X.; Writing—original draft, Y.Z. and D.X. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China, grant number 11761016 and 12161014; by Zhejiang Provincial Natural Science Foundation of China, grant number LY23A010013; by the National Statistical Science Research Project of China, grant number 2021LY011; by the Project of High Level Creative Talents in Guizhou Province of China; and by Guiyang University Multidisciplinary Team Construction Projects in 2021, grant number 2021-xk04.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The research data is available on the website http://lib.stat.cmu.edu/datasets/boston (accessed on 18 July 2022).

Acknowledgments: Sincere thanks to everyone who suggested revisions and improved this paper.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A. Sampling for the Parameter ρ

The MH algorithm, which is a popular sampling method, is adopted to draw observations from the conditional distribution (13). At the tth iteration with a current value $\rho^{(t)}$, a new candidate ρ^* is generated from the following equation:

$$\rho^* = \rho^{(t)} + c \cdot u^{(t)},$$

where $u^{(t)}$ is a random sample from the standard normal distribution, $\rho^{(t)}$ and $u^{(t)}$ are mutually independent; and c is a tuning parameter. ρ^* is accepted with probability

$$\min\left\{1, \frac{p(\rho^*|\boldsymbol{\beta}, \sigma, \boldsymbol{\delta}, \boldsymbol{e}, \boldsymbol{\gamma}, \boldsymbol{y}, \boldsymbol{x})}{p(\rho^{(t)}|\boldsymbol{\beta}, \sigma, \boldsymbol{\delta}, \boldsymbol{e}, \boldsymbol{\gamma}, \boldsymbol{y}, \boldsymbol{x})}\right\}.$$

Similar to LeSage and Pace [2], c could be chosen such that the average acceptance rate is about [0.4, 0.6].

The probability integral transformation (PIT) algorithm [25], another alternative sampling approach, could be used to draw observations from the the conditional distribution (13) by the following:

Step (i): denote $\rho_1 = \rho_{min}$, $\rho_d = \rho_{max}$, divide the interval $[\rho_{min}, \rho_{max}]$ into $d-1$ equally spaced points such that,

$$\rho_j = \rho_1 + (j-1)h \qquad for\ j = 2, \cdots, d-1,$$

where $h = \frac{2}{d-1}$. Obtain $F(\rho_1), F(\rho_2), \cdots, F(\rho_d)$ by quadrature numerical integration, where $F(\rho)$ is the posterior cumulative distribution function for ρ.

Step (ii): generate $u \sim U(0,1)$, for $j = 1, \cdots, d$, if $u = F(\rho_j)$, then ρ_j is a random observation of the conditional distribution (13); Denoted $f(\rho)$ as the posterior probability density function for ρ, i.e., $f(\rho) = \frac{p(\rho|\boldsymbol{\beta},\sigma,\boldsymbol{\delta},\boldsymbol{e},\boldsymbol{\gamma},\boldsymbol{y},\boldsymbol{x})}{\int p(\rho|\boldsymbol{\beta},\sigma,\boldsymbol{\delta},\boldsymbol{e},\boldsymbol{\gamma},\boldsymbol{y},\boldsymbol{x})d\rho}$. If $F(\rho_j) < u < F(\rho_{j+1})$, let ρ^* be $u = F(\rho^*)$, then using the Taylor expansion

$$u = F(\rho^*) \approx F(\rho_j) + f(\rho_j)(\rho^* - \rho_j),$$

thus, we obtain

$$\rho^* = \rho_j + \frac{u - F(\rho_j)}{f(\rho_j)}.$$

In the end, if $\rho^* \in [\rho_j, \rho_{j+1}]$, then ρ^* is a random observation of $f(\rho)$; otherwise, we set $\rho^* = \frac{\rho_j + \rho_{j+1}}{2}$, which is regarded as a random observation of $f(\rho)$.

It is noted that d can be selected empirically, or be chosen as a large value in simulation studies and data analysis (e.g., $d = 500$ in the paper).In addition, a simulation study (refer to Appendix B) is conducted to compute the Bayesian estimates of parameters based on the above proposed MH algorithm and PIT algorithm in our SAR quantile models and compare their performances.

Appendix B

To compare the empirical performance of our proposed sampling approach (denoted as Gibbs sampling with PIT algorithm) with the typical sampling approach (such as Gibbs sampling with MH algorithm), the following simulation study is considered. The dataset is generated from the model and parameter settings given in scenario 1 of the simulation study in Section 4.1. In each replication, 5000 samples are collected to calculate Bayesian estimates of unknown parameters after 5000 burn-ins by using Gibbs sampling with the MH algorithm. The related computing results based on 500 data sets are reported in Table A1; Table A1 only gives partial results to save space. From Tables A1 and 1, we have the following findings: these two MCMC procedures (Gibbs sampling with PIT algorithm and Gibbs sampling with MH algorithm) are quite effective.In general, they obtain similar and accurate estimate results about unknown parameters and their differences are very minor. Comparing the computing time of these two MCMC approaches, it roughly takes 6.1 s in a Thinkpad X240 server to run a data set for Gibbs sampling with the PIT algorithm, and it takes about 29.2 s to run a replication for Gibbs sampling with the MH algorithm. Note that Gibbs sampling with the MH algorithm depends heavily on the accept probability and the proposal distribution. Therefore, we recommend to use Gibbs sampling with the PIT algorithm (i.e., our proposed sampling approach) in applications.

Table A1. Means and RMSEs (in parentheses) in the simulation study of Appendix B.

(ρ, τ)	Error Distribution	β_1	β_2	β_3	β_4	β_5	β_6	β_7	β_8	ρ
(−0.8, 0.5)	normal	0.8461	0.8414	0.8329	0.8493	0.8334	0.8478	0.8441	0.8418	−0.7983
		(0.1643)	(0.1867)	(0.1942)	(0.1754)	(0.1909)	(0.1802)	(0.1910)	(0.1522)	(0.0173)
	t	0.7983	0.8286	0.8231	0.8084	0.8384	0.8201	0.8062	0.8201	−0.7969
		(0.3020)	(0.3432)	(0.3525)	(0.3701)	(0.3606)	(0.3670)	(0.3653)	(0.2924)	(0.0215)
	Laplace	0.8259	0.8284	0.8438	0.8406	0.8305	0.8189	0.8433	0.7973	−0.7978
		(0.2225)	(0.2890)	(0.2908)	(0.2665)	(0.2783)	(0.2830)	(0.2929)	(0.2421)	(0.0194)
	mixed-normal	0.8164	0.8340	0.8358	0.8457	0.8273	0.8205	0.8508	0.8194	−0.7988
		(0.2161)	(0.2635)	(0.2605)	(0.2803)	(0.2793)	(0.2821)	(0.2812)	(0.2326)	(0.0189)
(0.8, 0.5)	normal	0.8473	0.8410	0.8428	0.8528	0.8438	0.8345	0.8480	0.8422	0.7985
		(0.1594)	(0.1864)	(0.1857)	(0.1907)	(0.1912)	(0.1840)	(0.1852)	(0.1571)	(0.0183)
	t	0.8308	0.8287	0.8052	0.8410	0.8205	0.8158	0.8530	0.7742	0.7983
		(0.2829)	(0.3697)	(0.3524)	(0.3447)	(0.3771)	(0.3835)	(0.3882)	(0.3151)	(0.0197)
	Laplace	0.8058	0.8294	0.8394	0.8217	0.8238	0.8603	0.8088	0.8327	0.7978
		(0.2467)	(0.3066)	(0.2945)	(0.3027)	(0.3032)	(0.2893)	(0.2944)	(0.2380)	(0.0192)
	mixed-normal	0.8375	0.8436	0.8334	0.8396	0.8366	0.8266	0.8529	0.8147	0.7964
		(0.2164)	(0.2409)	(0.2514)	(0.2568)	(0.2649)	(0.2553)	(0.2451)	(0.2193)	(0.0191)

Appendix C

In order to survey the sensitivity of the Bayesian estimates to different prior inputs, we conduct the following simulation study. The dataset is generated from the model and parameter settings given in Scenario 1 of the simulation study in Section 4.1; the following two different prior inputs are considered:

Type (i): $a_\sigma = b_\sigma = 0.01$, $a_\delta = b_\delta = 0.01$, other prior inputs are set to be the same as those given in Section 4.1, which is regarded as another noninformative prior case, different from Section 4.1;

Type (ii): $a_\sigma = 0.5, b_\sigma = 0.0164$, $a_\delta = 0.5, b_\delta = 0.0164$, other prior inputs are taken to be the same as those given in Section 4.1, which is based on the suggestion of Fong et al. [35].

The corresponding results on the basis of 500 datasets are reported in Table A2, Table A2 only gives partial results to save space. Tables A2 and 1 imply that: (1) estimates with type (i) and type (ii) prior inputs are better than those obtained from prior inputs in Section 4.1, but their differences are minor; (2) all the "means" values based on the 3 different prior inputs are very close to the true values of unknown parameters, which illustrates that Bayesian estimates are very accurate and are not sensitive to the prior inputs in our considered cases.

Table A2. Means and RMSs (in parentheses) in the simulation study in Appendix C.

Type (i) Prior Inputs										
(ρ,τ)	Error distribution	β_1	β_2	β_3	β_4	β_5	β_6	β_7	β_8	ρ
(0.8, 0.5)	normal	0.8359	0.8587	0.8395	0.8519	0.8385	0.8506	0.8432	0.8360	0.7975
		(0.1478)	(0.1581)	(0.1592)	(0.1686)	(0.1743)	(0.1654)	(0.1732)	(0.1514)	(0.0182)
	t	0.8451	0.8299	0.8312	0.8288	0.8464	0.8390	0.8531	0.8240	0.7969
		(0.1987)	(0.2489)	(0.2349)	(0.2327)	(0.2233)	(0.2353)	(0.2344)	(0.2165)	(0.0212)
	Laplace	0.8521	0.8405	0.8281	0.8395	0.8535	0.8388	0.8277	0.8397	0.7967
		(0.1663)	(0.2012)	(0.2200)	(0.2000)	(0.1976)	(0.1966)	(0.2037)	(0.1724)	(0.0175)
	mixed-normal	0.8383	0.8361	0.8466	0.8474	0.8480	0.8326	0.8465	0.8217	0.7971
		(0.1649)	(0.1920)	(0.1856)	(0.2069)	(0.2046)	(0.1964)	(0.1860)	(0.1657)	(0.0198)
Type (ii) Prior Inputs										
(ρ,τ)	Error distribution	β_1	β_2	β_3	β_4	β_5	β_6	β_7	β_8	ρ
(0.8, 0.5)	normal	0.8273	0.8367	0.8505	0.8374	0.8375	0.8427	0.8355	0.8312	0.8001
		(0.1419)	(0.1654)	(0.1671)	(0.1659)	(0.1606)	(0.1619)	(0.1600)	(0.1408)	(0.0165)
	t	0.8133	0.8311	0.8253	0.8490	0.8332	0.8284	0.8373	0.8245	0.7990
		(0.1915)	(0.2125)	(0.2035)	(0.2081)	(0.1958)	(0.1865)	(0.1956)	(0.1791)	(0.0186)
	Laplace	0.8140	0.8415	0.8212	0.8329	0.8466	0.8236	0.8335	0.8209	0.7996
		(0.1647)	(0.1757)	(0.1877)	(0.1862)	(0.1914)	(0.1949)	(0.1943)	(0.1526)	(0.0176)
	mixed-normal	0.8199	0.8407	0.8257	0.8506	0.8254	0.8426	0.8342	0.8195	0.7999
		(0.1656)	(0.1772)	(0.1961)	(0.1858)	(0.1952)	(0.1864)	(0.1872)	(0.1702)	(0.0183)

References

1. Cliff, A.D.; Ord, J.K. *Spatial Autocorrelation*; Pion Ltd.: London, UK, 1973.
2. LeSage, J.; Pace, R.K. *Introduction to Spatial Econometrics*; Chapman and Hall: London, UK, 2009.
3. Du, J.; Sun, X.Q.; Cao, R.Y.; Zhang, Z.Z. Statistical inference for partially linear additive spatial autoregressive models. *Spat. Stat.* **2018**, *25*, 52–67. [CrossRef]
4. Liu, X.; Chen, J.B.; Cheng, S.L. A penalized quasi-maximum likelihood method for variable selection in the spatial autoregressive model. *Spat. Stat.* **2018**, *25*, 86–104. [CrossRef]
5. Xie, T.F.; Cao, R.Y.; Du, J. Variable selection for spatial autoregressive models with a diverging number of parameters. *Stat. Pap.* **2020**, *61*, 1125–1145. [CrossRef]
6. Jin, F.; Lee, L. GEL estimation and tests of spatial autoregressive models. *J. Econom.* **2019**, *208*, 585–612. [CrossRef]

7. Ju, Y.; Yang, Y.; Hu, M.; Dai, L.; Wu, L. Bayesian Influence Analysis of the Skew-Normal Spatial Autoregression Models. *Mathematics* **2022**, *10*, 1306. [CrossRef]
8. Koenker, R. *Quantile Regression*; Cambridge University Press: London, UK, 2005.
9. Kim, M.O.; Yang, Y. Semiparametric approach to a random effects quantile regression model. *J. Am. Stat. Assoc.* **2011**, *106*, 1405–1417. [CrossRef]
10. Lu, W.; Zhu, Z.; Lian, H. High-dimensional quantile tensor regression. *J. Mach. Learn. Res.* **2020**, *21*, 1–31.
11. Dunson, D.B.; Taylor, J. Approximate Bayesian inference for quantiles. *J. Nonparametric Stat.* **2005**, *17*, 385–400. [CrossRef]
12. Lavine, M. On an approximate likelihood for quantiles. *Biometrika* **1995**, *82*, 220–222. [CrossRef]
13. Lancaster, T.; Jun, S.J. Bayesian quantile regression methods. *J. Appl. Econom.* **2010**, *25*, 287–307. [CrossRef]
14. Kottas, A.; Krnjaji, M. Bayesian Semiparametric Modelling in Quantile Regression. *Scand. J. Stat.* **2009**, *36*, 297–319. [CrossRef]
15. Yang, Y.; He, X. Bayesian empirical likelihood for quantile regression. *Ann. Stat.* **2012**, *40*, 1102–1131. [CrossRef]
16. Rodrigues, T.; Dortet-Bernadet, J.L.; Fan, Y. Pyramid quantile regression. *J. Comput. Graph. Stat.* **2019**, *28*, 732–746. [CrossRef]
17. Yu, K.; Moyeed, R.A. Bayesian quantile regression. *Stat. Probab. Lett.* **2001**, *54*, 437–447. [CrossRef]
18. Kozumi, H.; Kobayashi, G. Gibbs sampling methods for Bayesian quantile regression. *J. Stat. Comput. Simul.* **2011**, *81*, 1565–1578. [CrossRef]
19. Hu, Y.; Zhao, K.; Lian, H. Bayesian quantile regression for partially linear additive models. *Stat. Comput.* **2015**, *25*, 651–668. [CrossRef]
20. Wang, Z.Q.; Tang, N.S. Bayesian quantile regression with mixed discrete and nonignorable missing Covariates. *Bayesian Anal.* **2020**, *15*, 579–604. [CrossRef]
21. George, E.; McCulloch, R. Variable selection via Gibbs sampling. *J. Am. Stat. Assoc.* **1993**, 88, 881–889. [CrossRef]
22. Ishwaran, H.; Rao, J.S. Spike and slab variable selection: Frequentist and Bayesian strategies. *Ann. Stat.* **2005**, *33*, 730–773. [CrossRef]
23. Panagiotelis, A.; Smith, M. Bayesian identification, selection and estimation of semiparametric functions in high-dimensional additive models. *J. Econom.* **2008**, *143*, 291–316. [CrossRef]
24. Geman, S.; Geman, D. Stochastic relaxation, Gibbs distribution, and the Bayesian restoration of images. *IEEE Trans. Pattern Anal. Mach. Intell.* **1984**, *6*, 721–741. [CrossRef] [PubMed]
25. Wichitaksorn, N.; Tsurumi, H. Comparison of MCMC algorithms for the estimation of Tobit model with non-normal error: The case of asymmetric Laplace distribution. *Comput. Stat. Data Anal.* **2013**, *67*, 226–235. [CrossRef]
26. Cripps, E.; Carter, C.; Kohn, R. Variable selection and covariance selection in multivariate regression models. *Handb. Stat.* **2005**, *25*, 519–552.
27. Dagpunar, J. An easily implemented generalized inverse Gaussian generator. *Commun. Stat.- Simul. Comput.* **1989**, *18*, 703–710. [CrossRef]
28. Jøgensen, B. *Statistical Properties of the Generalized Inverse GAUSSIAN Distribution*; Springer: New York, NY, USA; Berlin, Germany, 1982.
29. Geyer, C.J. Practical markov chain monte carlo. *Stat. Sci.* **1992**, *7*, 473–483. [CrossRef]
30. Chen, J.; Wang, R.; Huang, Y. Semiparametric Spatial Autoregressive Model: A Two-Step Bayesian Approach. *Ann. Public Health Res.* **2015**, *2*, 1012–1024.
31. Alhamzawi, R.; Yu, K.; Benoit, D.F. Bayesian adaptive Lasso quantile regression. *Stat. Model.* **2012**, *12*, 279–297. [CrossRef]
32. Matthews, B. Comparison of the predicted and observed secondary structure of t4 phage lysozyme. *Biochim. Biophys. Acta (BBA)-Protein Struct.* **1975**, *405*, 442–451. [CrossRef]
33. Harrison, D.H.; Rubinfeld, D.L. Hedonic housing prices and the demand for clean air. *J. Environ. Econ. Manag.* **1978**, *5*, 81–102. [CrossRef]
34. Ertur, C.; KochGrowth, W. Growth technological interdependence and spatial externalities: Theory and evidence. *J. Appl. Econom.* **2007**, *22*, 1033–1062. [CrossRef]
35. Fong, Y.; Rue, H.; Wakefield, J. Bayesian influence for generalized linear mixed models. *Biostatistics* **2010**, *11*, 397–412. [CrossRef] [PubMed]