Research Article

Bayesian Estimation of Partially Linear Additive Spatial Autoregressive Models with P-Splines

Zhiyong Chen [ID],[1] **Minghui Chen,**[1] **and Guodong Xing**[2]

[1]*School of Mathematics and Statistics, Fujian Normal University, Fuzhou 350117, China*
[2]*School of Mathematics and Statistics, Hefei Normal University, Hefei 230061, China*

Correspondence should be addressed to Zhiyong Chen; zychen1024@163.com

In this paper, we aim to develop a partially linear additive spatial autoregressive model (PLASARM), which is a generalization of the partially linear additive model and spatial autoregressive model. It can be used to simultaneously evaluate the linear and nonlinear effects of the covariates on the response for spatial data. To estimate the unknown parameters and approximate nonparametric functions by Bayesian P-splines, we develop a Bayesian Markov Chain Monte Carlo approach to estimate the PLASARM and design a Gibbs sampler to explore the joint posterior distributions of unknown parameters. Furthermore, we illustrate the performance of the proposed model and estimation method by a simulation study and analysis of Chinese housing price data.

1. Introduction

Spatial econometrics is a subfield of econometrics dealing with spatial interaction effects among geographical units Elhorst [1]; it has been widely applied in many research fields such as economics, geography, and environmental science. The traditional spatial parametric autoregressive models have received a lot of attention from theoretical econometricians and applied researchers, for example, Anselin [2], Case [4], Cressie [3], LeSage [4], LeSage [5], Anselin and Bera [6], Lee [7], Lee [8], Kakamu and Wago [9], LeSage and Pace [10], Kazar and Celik [11], and Piribauer and Crespo Cuaresma [12] among others. However, the spatial parametric autoregressive models are highly sensitive to model misspecification; they may not be adequate in many complex situations. Indeed, it is confirmed that some economic relationships in space exhibit highly nonlinear shapes [13–16]. Neglecting the underlying nonlinear relationship in spatial parametric autoregressive models frequently leads to inconsistent estimation of the parameters.

Although spatial nonparametric autoregressive models can be used to improve the performance of the spatial parametric autoregressive models, it is unavoidable to bear the risk of misspecifying the link function. However, when the dimension of the explanatory variables is high, a fully nonparametric spatial autoregressive model is too hard to be explored because of the so-called "curse of dimensionality" [17]. Several nonparametric dimension-reduction techniques have been developed to overcome the problem, for instance, single-index model [18, 19], additive model [20], and varying-coefficient model [21], among others. Thus, semiparametric spatial autoregressive models have been proposed for dealing with spatial data. To capture the underlying relationships between the response variables and their associated covariates, semiparametric spatial autoregressive models have gained much attention in the literature of statistics and econometrics. For example, Su and Jin [22] discussed the quasi-likelihood estimator for the semiparametric partially linear spatial autoregressive model; Su [23] proposed semiparametric generalized method of moment (GMM) estimation of the semiparametric spatial autoregressive model; Sun et al. [24] developed a profile likelihood estimator for the semiparametric spatial dynamic model; Chen et al. [25] presented a two-step Bayesian approach for the semiparametric spatial autoregressive model; Wei and Sun [26] considered the semiparametric GMM

estimation of a spatial model with space-varying coefficients; Hoshino [27] proposed a semiparametric series GMM estimator for the semiparametric spatial autoregressive model; Krisztin [28] studied a novel Bayesian semiparametric estimation for the penalized spline spatial autoregressive model; and Wei et al. [29] considered profile quasi-maximum likelihood method to estimate the partially linear varying-coefficient spatial autoregressive model. Sun and Wu [30] discussed the GMM to estimate the partially linear single-index spatial autoregressive model. Cheng et al. [31] developed the GMM estimation of the partially linear single-index spatial autoregressive model. Du et al. [32] investigated a GMM estimator for PLASARM. Hu et al. [33] presented a new partial functional linear spatial autoregressive model.

In recent years, semiparametric models have become an important research topic between statistics and econometrics due to the explanation of parametric modeling and the flexibility of nonparametric modeling. Among the various semiparametric models, the most popular ones are perhaps the partially linear additive models. They contain both linear and nonlinear additive components. As they can not only provide more flexible models than the stringent linear models but also mitigate the "curse of dimensionality" phenomenon encountered in nonparametric regression. Therefore, the partially linear additive models provide a good balance between the flexibility of the additive models and the interpretation of the partially linear models. A lot of authors have developed various methods to analyze such models: local linear method [34], spline estimation [35–37], variable selection [38–41], quantile regression [42–45], and so on. But most of this work has not been applied to analyze spatial data. It is a distinctive challenge facing analysts of spatial data to characterize certain flexible functional forms that try to account for the latent nonlinearity of PLASARM. Combining the partially linear additive models with the spatial autoregressive models, we develop a Bayesian P-splines method and design a Gibbs sampler to explore the joint posterior distributions, along with a Markov chain Monte Carlo algorithm, to estimate the unknown parameters and approximate the nonparametric functions of PLASARM.

The rest of the article proceeds as follows: In Section 2, we present PLASARM for spatial dependent data and discuss its identifiability conditions and then obtain the likelihood functions by approximating the link function with a Bayesian P-spline method. In Section 3, we specify the prior distributions, derive the full conditional posteriors of the unknown parameters, and describe the detailed sampling algorithms in order to provide a Bayesian P-splines method for analysis. The applicability and practicality of the proposed model and estimation method are demonstrated through a simulation study and a real dataset in Section 4. In Section 5, we conclude the paper with a summary.

2. Model and Likelihood

2.1. Model. The partially linear additive spatial autoregressive model is defined as follows:

$$y_i = \rho \sum_{l=1}^{n} w_{il} y_l + x_i^T \alpha + \sum_{j=1}^{p} g_j(z_{ij}) + \varepsilon_i, \quad i = 1, \ldots, n, \quad (1)$$

where y_i is the i-th observation of response variable, $x_i = (x_{i1}, \ldots, x_{iq})^T$ and $z_i = (z_{i1}, \ldots, z_{ip})^T$ are the i-th observations of q and p dimensional covariate vectors, respectively, w_{il} is a specified constant spatial weight, $g_j(\cdot)$ $(j = 1, \ldots, p)$ are unknown univariate nonparametric functions, ρ is an unknown spatial parameter reflecting spatial autocorrelation between neighbors with stability condition $|\rho| < 1$, $\alpha = (\alpha_1, \ldots, \alpha_q)^T$ is a $q \times 1$ vector of the unknown parameters, and ε_i's are mutually independent and identically distributed normal random variables with zero mean and variance σ^2. To ensure model identifiability of the nonparametric functions, it is often assumed that the condition $\sum_{i=1}^{n} g_j(z_{ij}) = 0$ for $j = 1, \ldots, p$.

2.2. Bayesian P-Splines and Likelihood. In analyzing the proposed PLASARMs defined by (1), modeling the smooth nonparametric function is an important issue. Because of the advantages of the Bayesian approach and the nice features of Bayesian P-splines [46], we will concentrate on applying the Bayesian P-splines approach. We intend to approximate unknown function $g_j(\cdot)$ in (1) by B-splines [47]. For $j = 1, \ldots, p$, the unknown function $g_j(\cdot)$ is a polynomial spline of degree m_j with k_j order interior knots $\xi_j = (\xi_{j1}, \ldots, \xi_{jk_j})^T$ with $a_j < \xi_{j1} < \cdots < \xi_{jk_j} < b_j$, that is,

$$g_j(z_j) = \sum_{l=1}^{K_j} B_{jl}(z_j)\beta_{jl} = B_j^T(z_j)\beta_j, \quad z_j \in [a_j, b_j], \quad (2)$$

where $K_j = 1 + m_j + k_j$ is the number of splines determined by the number of knots, $B_j(z_j) = (B_{j1}(z_j), \ldots, B_{jK_j}(z_j))^T$ is a $K_j \times 1$ vector of spline basis that is determined by the knot vector ξ_j, $\beta_j = (\beta_{j1}, \ldots, \beta_{jK_j})^T$ is a $K_j \times 1$ vector of spline coefficients, and

$$a_j = \min_{1 \leq i \leq n}\{z_{ij}\} \text{ and } b_j = \max_{1 \leq i \leq n}\{z_{ij}\}. \quad (3)$$

In practice, a choice of K_j in the range of 10 to 30 provides the flexibility of fitting. To prevent overfitting due to using a relatively large number of knots, Eilers and Marx [48] proposed applying a difference penalty on coefficients of adjacent B-splines. Let $1_n = (1, \ldots, 1)^T$ and $\mathbf{B}_j$ be an $n \times K_j$ matrix with $B_j^T(z_{ij})$ as its i-th row. To achieve identification, we set $\sum_{i=1}^{n} \sum_{l=1}^{K_j} B_{jl}(z_{ij})\beta_{jl} = 0$, which is equivalent to $1_n^T \mathbf{B}_j \beta_j = 0$. Denote $Q_j = 1_n^T \mathbf{B}_j$; then, the constraint becomes $Q_j \beta_j = 0$.

It follows from (2) that model (1) can be represented as

$$y_i = \rho \sum_{l=1}^{n} w_{il} y_l + x_i^T \alpha + \sum_{j=1}^{p} B_j^T(z_j)\beta_j + \varepsilon_i$$
$$= \rho \sum_{l=1}^{n} w_{il} y_l + x_i^T \alpha + B^T(z_i)\beta + \varepsilon_i, \quad i = 1, \ldots, n, \quad (4)$$

where $B^T(z_i) = (B_1^T(z_{i1}), \ldots, B_p^T(z_{ip}))^T$ and $\beta = (\beta_1^T, \ldots, \beta_p^T)^T$. Then, the matrix form of the model (1) can be written as

$$y = \rho W y + x^T \alpha + B^T(z)\beta + \varepsilon, \qquad (5)$$

$$
\begin{aligned}
&L\left(\alpha, \beta, \sigma^2, \rho \mid y, x, z\right) \\
&\propto \sigma^{-n} |I_n - \rho W| \exp\left\{-\frac{1}{2\sigma^2}\left[y - \rho W y - x^T\alpha - B^T(z)\beta\right]^T\left[y - \rho W y - x^T\alpha - B^T(z)\beta\right]\right\} \\
&= \sigma^{-n} |A(\rho)| \exp\left\{-\frac{1}{2\sigma^2}\left[A(\rho)y - x^T\alpha - B^T(z)\beta\right]^T\left[A(\rho)y - x^T\alpha - B^T(z)\beta\right]\right\} \\
&\doteq \sigma^{-n} |A(\rho)| \exp\left\{-\frac{1}{2\sigma^2}\left[A(\rho)y - \mathfrak{B}(x,z)\theta\right]^T\left[A(\rho)y - \mathfrak{B}(x,z)\theta\right]\right\},
\end{aligned}
\qquad (6)
$$

where $x = (x_1, \ldots, x_n)^T$, $z = (z_1, \ldots, z_n)^T$, $K = \sum_{j=1}^{p} K_j$, $\theta = (\alpha^T, \beta^T)^T$ is a $(K+q) \times 1$ vector of the regression coefficient, $\mathfrak{B}(x,z) = (x, B^T(z))$ is an $n \times (K+q)$ matrix, I_n is an identity matrix of order n, and $A(\rho) = I_n - \rho W$.

3. Bayesian Estimation

We develop a Bayesian P-splines method with the Gibbs sampler to estimate the proposed model in this section. We begin with the specification of the prior distributions, then the derivations of the full conditional posteriors of all of the unknown parameters, and the narration of the detailed sampling scheme.

3.1. Priors. In order to develop a Bayesian P-splines method to estimate the model and eschew the use of improper to prevent improper joint posterior, we specify appropriate prior distributions for all the unknown parameters which include regression coefficients $\theta = (\alpha^T, \beta^T)^T$, variance σ^2, and spatial parameter ρ.

Firstly, we set a hierarchical prior for α, which consists of a conjugate normal prior,

$$\pi\left(\alpha \mid \sigma^2, \tau_0\right) \propto \left(2\pi\tau_0\sigma^2\right)^{-q/2} \exp\left\{-\frac{\alpha^T\alpha}{2\tau_0\sigma^2}\right\}, \qquad (7)$$

and an inverse-gamma prior,

$$\pi(\tau_0) \propto \tau_0^{-r_{\tau_{\alpha0}}/2-1} \exp\left\{-\frac{s_{\tau_{\alpha0}}^2}{2\tau_0}\right\}, \qquad (8)$$

where $r_{\tau_{\alpha0}}$ and $s_{\tau_{\alpha0}}^2$ are prespecified hyperparameters.

Secondly, to identify the unspecified smooth function $g_j(\cdot)$ for $j = 1, \ldots, p$, the identifiability constraint $Q_j\beta_j = 0$ should be imposed on β_j. Under this constraint, we choose the random walk prior,

where $y = (y_1, \ldots, y_n)^T$, $x = (x_1, \ldots, x_n)^T$, $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_n)^T$, $W = (w_{il})$ is an $n \times n$ specified constant spatial weight matrix, and $B^T(z)$ is an $n \times (K_1 + \cdots + K_p)$ matrix with $B^T(z_i)$ as its i-th row.

The likelihood function corresponding to (5) is as follows:

$$\pi\left(\beta_j \mid \sigma^2, \tau_j\right) \propto \left(2\pi\tau_j\sigma^2\right)^{-(K_j-d)/2} \exp\left\{-\frac{\beta_j^T M_{\beta_j}\beta_j}{2\tau_j\sigma^2}\right\} I\{Q_j\beta_j = 0\}, \qquad (9)$$

for β_j, where $I\{\cdot\}$ is the indicator function, d is the order of the random walk, M_{β_j} is the penalty matrix that equals $(D_{d-1} \times \cdots \times D_0)^T(D_{d-1} \times \cdots \times D_0)$ for d-order random walk prior, D_l is a $(K_j - l - 1) \times (K_j - l)$ matrix of the form

$$D_l = \begin{pmatrix} -1 & 1 & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & -1 & 1 \end{pmatrix}, \quad l = 0, \ldots, d-1. \qquad (10)$$

To enhance the flexibility and robustness of the method, we further put an inverse-gamma prior,

$$\pi(\tau_j) \propto \tau_j^{-r_{\tau_{\beta_j0}}/2-1} \exp\left\{-\frac{s_{\tau_{\beta_j0}}^2}{2\tau_j}\right\}, \qquad (11)$$

on τ_j, where $r_{\tau_{\beta_j0}}$ and $s_{\tau_{\beta_j0}}^2$ are prespecified hyperparameters. In addition, the following conjugate prior distribution of σ^2 is assigned as

$$\pi(\sigma^2) \propto (\sigma^2)^{-r_0/2-1} \exp\left\{-\frac{s_0^2}{2\sigma^2}\right\}, \qquad (12)$$

where r_0 and s_0^2 are prespecified hyperparameters. Throughout this article, we set $r_0 = s_0^2 = 1$ to obtain a Cauchy distribution of σ^2 and use $r_{\tau_{\alpha0}} = r_{\tau_{\beta_j0}} = 1$ and $s_{\tau_{\alpha0}}^2 = s_{\tau_{\beta_j0}}^2 = 0.005$ to obtain a highly dispersed inverse-gamma prior for τ_j, $j = 0, 1, \ldots, p$.

Finally, the spatial autocorrelation coefficient ρ is given a uniform prior $\rho \sim U(\lambda_{\min^{-1}}, \lambda_{\max^{-1}})$, where $\lambda_{\min}$ and $\lambda_{\max}$ are the minimum and maximum eigenvalues of the standardized spatial weight matrix W, respectively. In other words,

$$\pi(\rho) \propto 1. \qquad (13)$$

The joint prior of all unknown parameters is given by

$$\pi\left(\rho, \alpha, \beta, \sigma^2, \tau\right) = \pi(\rho)\pi\left(\sigma^2\right)\pi\left(\tau_0\right)\pi\left(\alpha|\sigma^2, \tau_0\right)\prod_{j=1}^{p}\pi\left(\beta_j|\sigma^2, \tau_j\right)\pi\left(\tau_j\right), \tag{14}$$

where $\tau = (\tau_0, \tau_1, \ldots, \tau_p)$. Note that we have treated hyperparameter τ as a parameter vector for computational convenience.

3.2. The Full Conditional Posterior Distributions of the Parameters.
In this subsection, we present the full conditional posterior distributions of all parameters that are used in the Gibbs sampler algorithm and describe the detailed sampling method.

It follows from likelihood function (6) and priors (14) that, given the remaining unknown quantities, the conditional posterior distribution of spatial autocorrelation coefficient ρ is

$$p\left(\rho|y, x, z, \alpha, \beta, \sigma^2, \tau\right)$$

$$\propto |A(\rho)|\exp\left\{-\frac{1}{2\sigma^2}\left[A(\rho)y - x^T\alpha - B^T(z)\beta\right]^T\left[A(\rho)y - x^T\alpha - B^T(z)\beta\right]\right\}. \tag{15}$$

Because it is not a standard density function, we cannot directly do simulations from (15). We prefer the Metropolis–Hastings algorithm [49, 50] to overcome the difficulty: generate a candidate ρ^* from a truncated Cauchy distribution with location ρ and scale σ_ρ on interval $(\lambda_{\min}^{-1}, \lambda_{\max}^{-1})$, where σ_ρ acts as a tuning parameter, and accept ρ^* with probability

$$\min\left\{1, \frac{p\left(\rho^*|y, x, z, \alpha, \beta, \sigma^2, \tau\right)}{p\left(\rho|y, x, z, \alpha, \beta, \sigma^2, \tau\right)} \times C_\rho\right\}, \tag{16}$$

given the factor

$$C_\rho = \frac{\arctan\left[\left(\lambda_{\max}^{-1} - \rho\right)/\sigma_\rho\right] - \arctan\left[\left(\lambda_{\min}^{-1} - \rho\right)/\sigma_\rho\right]}{\arctan\left[\left(\lambda_{\max}^{-1} - \rho^*\right)/\sigma_\rho\right] - \arctan\left[\left(\lambda_{\min}^{-1} - \rho^*\right)/\sigma_\rho\right]}. \tag{17}$$

It is easy to see from likelihood function (6) and priors (14) that, given the spatial autocorrelation coefficient ρ and the hyperparameter τ, the joint posterior of all the parameters is proportional to

$$p\left(\theta, \sigma^2|y, x, z, \rho, \tau\right)$$

$$\propto \sigma^{-n}\exp\left\{-\frac{1}{2\sigma^2}\left[A(\rho)y - x^T\alpha - B^T(z)\beta\right]^T\left[A(\rho)y - x^T\alpha - B^T(z)\beta\right]\right\}$$

$$\times \sigma^{-r_0-2}\exp\left\{-\frac{s_0^2}{2\sigma^2}\right\} \times \sigma^{-q}\exp\left\{-\frac{\alpha^T\alpha}{2\tau_0\sigma^2}\right\} \times \prod_{j=1}^{p}\sigma^{-K_j+d}\exp\left\{-\frac{\beta_j^T M_{\beta_j}\beta_j}{2\tau_j\sigma^2}\right\}I\{Q_j\beta_j = 0\}$$

$$\propto \exp\left\{-\frac{1}{2\sigma^2}[A(\rho)y - \mathfrak{B}(x,z)\theta]^T[A(\rho)y - \mathfrak{B}(x,z)\theta]\right\} \tag{18}$$

$$\times \exp\left\{-\frac{1}{2\sigma^2}\theta^T\text{diag}\{\tau^{-1}\}\theta\right\}I\{Q\beta = 0\} \times \sigma^{-n-r_0-q-K+p\,d-2}\exp\left\{-\frac{s_0^2}{2\sigma^2}\right\}$$

$$\propto \left(\sigma^2\right)^{-(n+r_0-pd)/2-1}\exp\left\{-\frac{S^2 + s_0^2}{2\sigma^2}\right\}$$

$$\times\left(2\pi\sigma^2\right)^{-(K+q)/2}|\Xi|^{1/2}\exp\left\{-\frac{1}{2\sigma^2}(\theta - \widehat{\theta})^T\Xi(\theta - \widehat{\theta})\right\}I\{Q\beta = 0\},$$

where $Q = (Q_1, \ldots, Q_p)$, $\text{diag}\{\tau^{-1}\} = \text{diag}\{\tau_0^{-1} I_q, \tau_1^{-1} M_{\beta_1}, \ldots, \tau_p^{-1} M_{\beta_p}\}$, $\Xi = \text{diag}\{\tau^{-1}\} + \mathfrak{B}^T (x, z) \mathfrak{B} (x, z)$, $\hat{\theta} = \Xi^{-1} \mathfrak{B}^T (x, z) A(\rho) y$, and $S^2 = y^T A(\rho)^T A(\rho) y - \hat{\theta}^T \Xi \hat{\theta}$.

We can see from (18) that the method of composition [51] can be used to draw σ^2 from the conditional inverse-gamma posterior,

$$p(\sigma^2 | y, x, z, \theta, \rho, \tau) \propto (\sigma^2)^{-(n + r_0 - p\, d)/2 - 1} \exp\left\{-\frac{S^2 + s_0^2}{2\sigma^2}\right\}, \tag{19}$$

and to generate θ from the conditional normal posterior,

$$p(\theta | y, x, z, \rho, \sigma^2, \tau) \propto (2\pi\sigma^2)^{-(K+q)/2} |\Xi|^{1/2} \exp\left\{-\frac{1}{2\sigma^2}(\theta - \hat{\theta})^T \Xi (\theta - \hat{\theta})\right\} I\{Q\beta = 0\}. \tag{20}$$

In addition, to achieve identification, the constraint $Q_j \beta_j = 0$ should be imposed on β_j. According to Panagiotelis and Smith [52], sampling β_j from (20) is equivalent to sampling $\beta_j^{(\text{new})}$ from $N(\beta_j^*, \Sigma_j^*)$, where $\Sigma_j^* = (\tau_j^{-1} M_{\beta_j} + \mathbf{B}_j^T \mathbf{B}_j)^{-1}$, then $\beta_j^{(\text{new})}$ is transformed to β_j by

$$\beta_j = \beta_j^{(\text{new})} - \Sigma_j^* Q_j^T (Q_j \Sigma_j^* Q_j^T)^{-1} Q_j \beta_j^{(\text{new})}, \quad j = 1, \ldots, p. \tag{21}$$

It is obvious that the hyperparameter τ_j for $j = 0, 1, \ldots, p$ is mutually independent in the posterior. The full conditional posterior of τ_j is an inverse-gamma distribution with density

$$p(\tau_0 | \alpha, \sigma^2) \propto \tau_0^{-(q + r_{\tau_{\alpha 0}})/2 - 1} \exp\left\{-\frac{s_{\tau_{\alpha 0}}^2 + \alpha^T \alpha / \sigma^2}{2\tau_0}\right\}, \tag{22}$$

$$p(\tau_j | \beta, \sigma^2) \propto \tau_j^{-(K_j - d + r_{\tau_{\beta j 0}})/2 - 1} \exp\left\{-\frac{s_{\tau_{\beta j 0}}^2 + \beta_j^T M_{\beta_j} \beta_j / \sigma^2}{2\tau_j}\right\}, \quad j = 1, \ldots, p, \tag{23}$$

which can be simulated directly from (22) and (23), respectively.

3.3. Sampling.

Let the parameter vector $\Theta = \{\rho, \sigma^2, \theta, \tau\}$. The Bayesian estimate of Θ is obtained by observations drawn from the full conditional posterior of all parameters with running some Markov chain Monte Carlo (MCMC) tools such as the Gibbs sampler and the Metropolis–Hastings algorithm. The pseudocode of the MCMC algorithm in our method is implemented in Algorithm 1.

4. Numerical Illustration

A simulation study demonstrates that the proposed model and methodology perform satisfactorily in the finite sample. We apply the proposed model to analyze a real dataset.

4.1. Simulation.

Consider the following model:

$$y_i = \rho \sum_{l=1}^{n} w_{il} y_l + x_i^T \alpha + \sum_{j=1}^{2} g_j(z_{ij}) + \varepsilon_i, \quad i = 1, \ldots, n, \tag{24}$$

where the covariate vector $x_i = (x_{i1}, x_{i2})^T$ follows a bivariate normal distribution with mean 0 and covariance matrix $\Sigma = \begin{pmatrix} 1 & -0.5 \\ -0.5 & 1 \end{pmatrix}$, $z_i = (z_{i1}, z_{i2})^T$ is a bivariate vector, and both z_{i1} and z_{i2} are independent; they follow uniform distributions on $(-1, 1)$ and $(0, 1)$, respectively. The link functions $g_1(z) = \sin(\pi z)$ and $g_2(z) = 4z(1 - z^2) - 1$, $\varepsilon_i \sim N(0, \sigma^2)$. The true values of parameters are assumed as $\alpha = (1, -1)^T$ and $\sigma^2 = 0.25$, respectively. For comparison, we consider three different cases of spatial parameters $\rho_1 = 0.25$, $\rho_2 = 0.5$, and $\rho_3 = 0.75$, which represent from weak to strong spatial dependence of the response. Two kinds of spatial weight matrices W are used: one is Rook weight matrix [2] with n units, and the other is Case weight matrix [53] with r districts and m members in each district. The sample size of Rook weight matrix and Case weight matrix is $n = \{100, 169, 400, 625\}$ and $(r, m) = \{(20, 5), (20, 8), (80, 5), (80, 8)\}$, respectively.

According to the above design, we conduct each simulation with 500 replications. For $j = 1, \ldots, p$, we applied a natural cubic spline in which the knots with $K_j = 22$ are placed at an equally spaced interval of the predictor variables, used second-order random walk penalties, and assigned hyperparameters $(r_0, s_0^2, r_{\tau_{\alpha 0}}, s_{\tau_{\alpha 0}}^2, r_{\tau_{\beta j 0}}, s_{\tau_{\beta j 0}}^2) = (1, 1, 1, 0.005, 1, 0.005)$ in our computation. The second-order random walk penalties are used for the Bayesian

---

Input: Samples: $\{(y_i, x_i, z_i)\}_{i=1,\ldots,n}.$

Initialization: Initialize the MCMC algorithm in iteration $t = 0$ with $\Theta^{(0)}$, where all unknown quantities are drawn from the respective prior distributions.

MCMC iterations: For $t = 1, 2, 3, \ldots$, given the current state $\Theta^{(t-1)}$ successively, resample $\Theta^{(t)}$ from $p(\Theta|y, x, z)$. The detailed Gibbs sampler cycles are outlined as follows:

  (i) Draw $\rho^{(t)}$ from $p(\rho|y, x, z, \theta^{(t-1)}, \sigma^{2(t-1)}, \tau^{(t-1)})$;

  (ii) Draw $\sigma^{2(t)}$ from $p(\sigma^2|y, x, z, \theta^{(t-1)}, \rho^{(t-1)}, \tau^{(t-1)})$;

  (iii) Draw $\theta^{(t)}$ from $p(\theta|y, x, z, \rho^{(t-1)}, \sigma^{2(t-1)}, \tau^{(t-1)})$, and adjust $\beta^{(t)}$ according to (21);

  (iv) Draw $\tau^{(t)}$ from $p(\tau|\theta^{(t-1)}, \sigma^{2(t-1)})$, which is replaced by the following two steps:

  (v) Draw $\tau_0^{(t)}$ from $p(\tau_0|\alpha^{(t-1)}, \sigma^{2(t-1)})$;

  (vi) Draw $\tau_j^{(t)}$ from $p(\tau_j|\beta^{(t-1)}, \sigma^{2(t-1)})$ for $j = 1, \ldots, p$.

Output: An MCMC sample from the joint posterior distribution: $\left\{\Theta^{(t)}\right\}_{t=1,2,3,\ldots}.$

ALGORITHM 1: The pseudocode of the MCMC sampling scheme.

P-splines to approximate the unknown smooth functions. The initial states of the Markov chain are chosen as follows: the unknown quantities are drawn from the respective prior distributions. By incrementally increasing or decreasing, the tuning parameter σ_ρ is used such that the resultant acceptable rate for the parameter is around 25%.

We can generate 6000 sampled values and delete the first 2000 values as a burn-in period for each of the replications by the proposed Gibbs sampler. In order to check the convergence of the proposed MCMC algorithm, we run five times with different starting values by the proposed Gibbs sampler for an arbitrarily selected replication. The sampled traces of parts of the unknown quantities for replication with $(r, m) = (80, 5)$ and $\rho = 0.5$ are displayed in Figure 1, where the unknown quantities include model parameters and fitted function on grid points. It is evident that the five parallel sequences mix reasonably well. The other cases have similar performance. Based on the five parallel sequences, we further calculate the "potential scale reduction factor" $\sqrt{\widehat{R}}$ [54] for each model parameter and fitted function on 10 selected grid points. All the values of $\sqrt{\widehat{R}}$ against the iteration numbers are displayed in Figure 2 (the case of the spatial parameter $\rho = 0.5$). We observe that 2000 burn-in iterations have converged as all the values of $\sqrt{\widehat{R}}$ were less than 1.2. Based on 4000 sampled values, we then compute the means, the standard errors, and the 2.5[th] and 97.5[th] percentiles of the posterior distributions for all unknown quantities.

We evaluate the performance of the estimated functions by the mean absolute deviation errors (MADE):

$$\text{MADE}_j = \frac{1}{101} \sum_{i=1}^{101} \left|\widehat{g}_j(z_{ji}) - g_j(z_{ji})\right| \text{ and } \text{MADE} = \frac{1}{p} \sum_{j=1}^{p} \text{MADE}_j, \tag{25}$$

at 101 fixed grid points $\{z_{ji}\}_{i=1}^{101}$ that are equally spaced chosen from interval $[a_j, b_j]$. To assess the accuracy of the estimates for the remaining parameters, we calculate the corresponding means across 500 replications for the posterior mean (Mean), bias (Bias), standard error (SE), and 2.5[th] and 97.5[th] percentiles of the parameters (the 95% posterior credible intervals, 95% CI). We also compute the standard derivations (SD) of the estimated posterior means to compare them with the means of the estimated posterior standard errors.

Figure 3(a) displays the boxplots of the MADE values with sample size $n = 100$. Based on the Rook weight matrix, the medians are $\text{MADE}_1 = 0.1023$, $\text{MADE}_2 = 0.0898$, and $\text{MADE} = 0.0975$. Based on the Case weight matrix, the medians are $\text{MADE}_1 = 0.1019$, $\text{MADE}_2 = 0.0896$, and $\text{MADE} = 0.0974$. Figure 3(b) shows the boxplots of the MADE values with sample size $n = 400$. Based on the Rook weight matrix, the medians are $\text{MADE}_1 = 0.0507$, $\text{MADE}_2 = 0.0460$, and $\text{MADE} = 0.0488$. Based on the Case weight matrix, the medians are $\text{MADE}_1 = 0.0505$, $\text{MADE}_2 = 0.0461$, and $\text{MADE} = 0.0488$. It shows that both the Rook and Case weight matrices can obtain reasonable estimations in fitting nonparametric functions.

Table 1 summarizes the estimation results. It is observed that all of the means of the estimates are very close to the respective true values, and the average values of the standard errors are close to the corresponding standard derivations, which indicate that the parameter estimates and the standard errors are accurate. This indicates that the proposed estimators of the parametric model perform better with increasing sample size. By comparing the estimates for ρ under the same sample sizes, we find the estimates of ρ with the Case weight matrix are much better than those with the Rook weight matrix. The possible main reason lies in the fact that the Rook weight matrix is more complicated than the Case weight matrix. Finally, the bigger the sample size under the same spatial complexity, the more accurate the estimates are. We have also repeated the aforementioned experiences with different starting values, and the results are similar. This indicates that the proposed Gibbs sampler works well.

The median and SD of MADE values of the link functions are reported in Table 2. It shows that the median and SD of
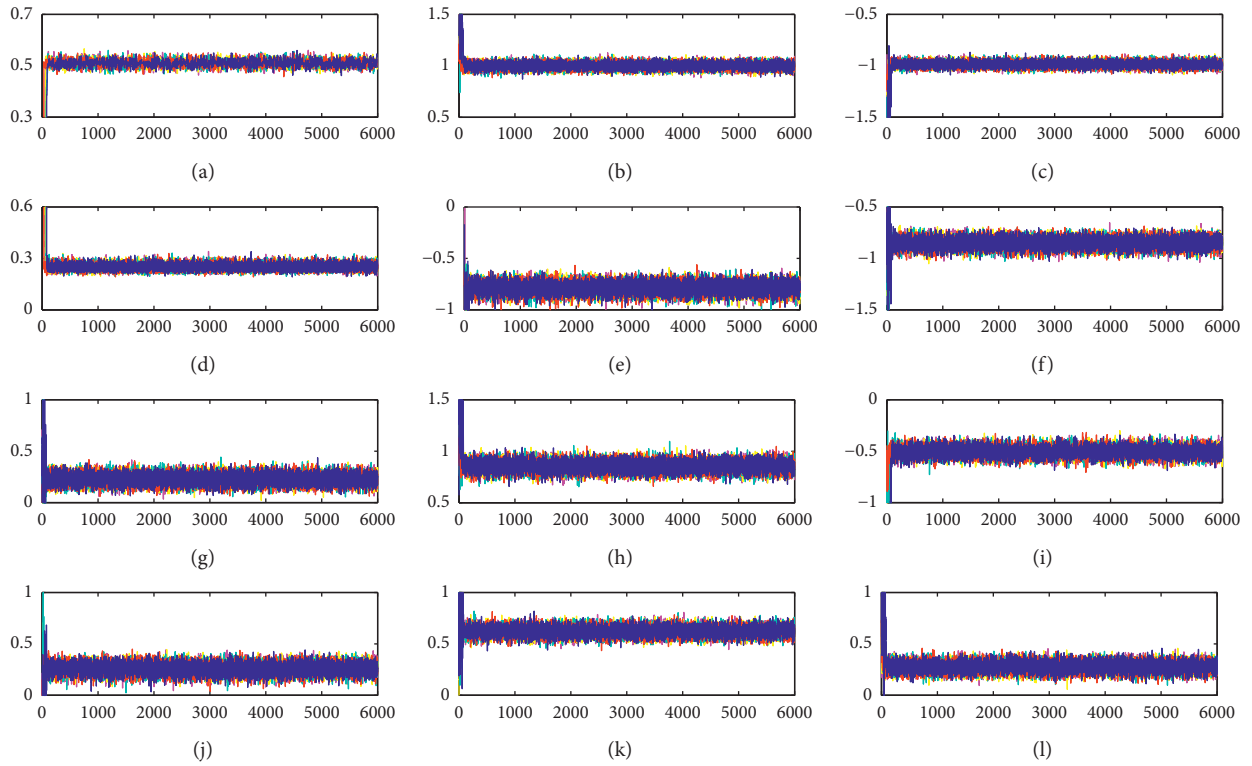
FIGURE 1: Trace plots of five parallel sequences corresponding to different starting values for parts of the unknown quantities. (a) ρ, (b) α_1, (c) α_2, (d) σ^2, (e) $g_1(z_{15})$, (f) $g_1(z_{35})$, (g) $g_1(z_{55})$, (h) $g_1(z_{75})$, (i) $g_2(z_{15})$, (j) $g_2(z_{35})$, (k) $g_2(z_{55})$, and (l) $g_2(z_{75})$.
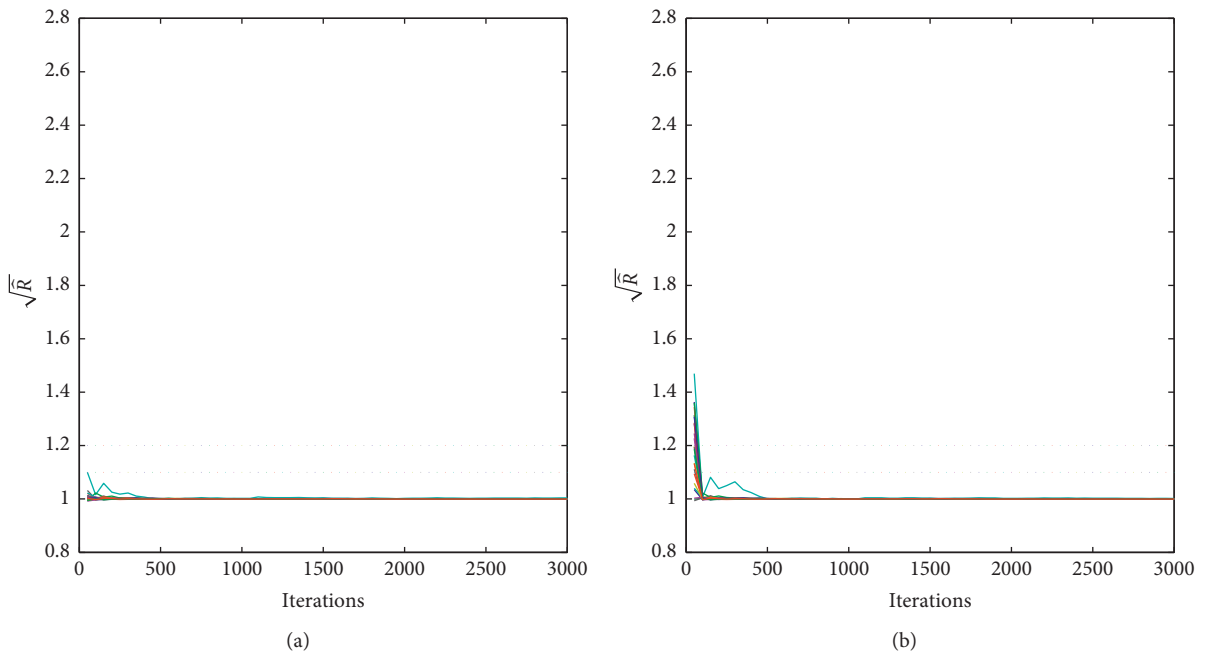

FIGURE 2: The "potential scale reduction factor" $\sqrt{\widehat{R}}$ for simulation results. (a) Rook weight matrix ($n = 400$). (b) Case weight matrix ($r = 80$, $m = 5$).
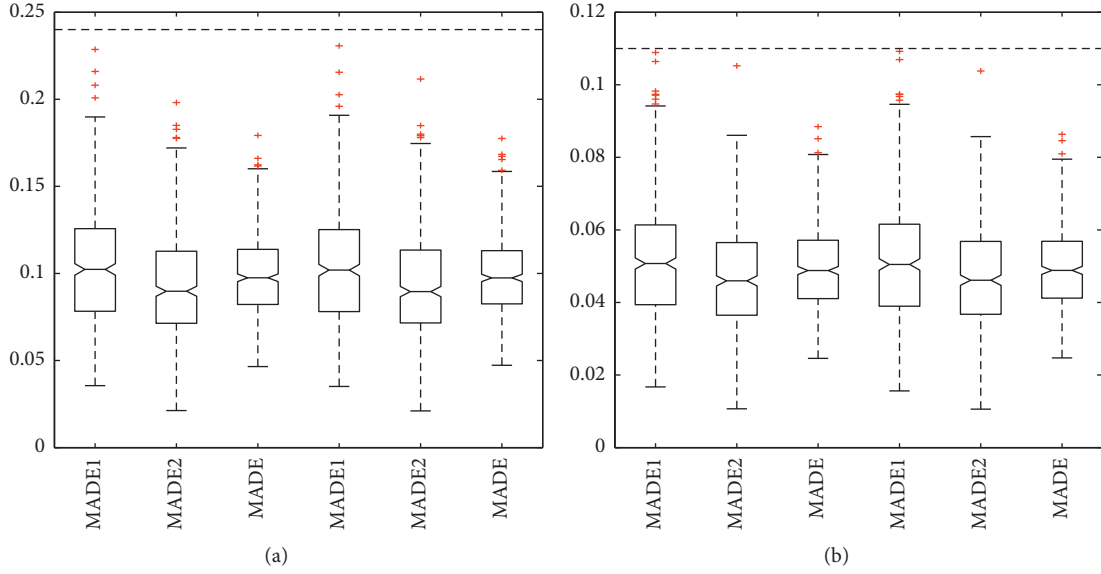
FIGURE 3: The boxplots (a) show the mean absolute deviation errors with sample size $n = 100$. The boxplots (b) show the mean absolute deviation errors with sample size $n = 400$ (the three panels on the left are based on the Rook weight matrix and the three panels on the right are based on the Case weight matrix). (a) MADE ($\rho = 0.5$, $n = 100$), (b) MADE ($\rho = 0.5$, $n = 400$).

MADE values of the link functions decrease along with increasing the sample size, which indicates that the estimates of the unknown functions are convergent. The spatial correlation coefficients also have little effect on the nonparametric estimates.

The estimated functions, together with the 95% pointwise posterior credible intervals, are depicted in Figure 4 from a typical sample under the spatial dependence $\rho = 0.5$ when the sample size is $n = 100$ and $n = 400$, respectively. The typical sample is selected in such a way that its MADE value is equal to the median in the 500 replications. This indicates that the proposed estimators of nonparametric functions are performed better with increasing sample size. It took about 30 seconds for $n = 100$ and 60 seconds for $n = 400$ to run each replication using a PC with Intel (R) Core (TM) i7-8750H 2.20 GHz, respectively. Computer code is available upon request from the authors.

4.2. A Real Data Example. As an illustration of our proposed model and method with a real example, we analyze housing prices data in Chinese cities. The data are obtained from the China City Statistical Yearbook 2013 [56] and the China Statistical Yearbook for Regional Economy 2013 [57]. They cover 286 cities at/above the prefecture level (except for the cities in Taiwan, Hong Kong, and Macau). In our real data analysis, we consider that the dependent variable y is each city's average selling price of residential houses (denoted by HP). There are three covariates: (1) population density (PD); (2) per capita disposable income of urban households (INC); and (3) loan-to-GDP ratio (MON), which indicates the degree of monetary policy whether ease or tightness.

This motivates us to consider the following partially linear additive spatial autoregressive model:

$$y_i = \rho \sum_{l=1}^{n} w_{il} y_l + \alpha x_i + \sum_{j=1}^{2} g_j(z_{ij}) + \varepsilon_i, \quad i = 1, \ldots, n,$$

$$(26)$$

where the response variable $y_i = \mathrm{HP}_i$, $x_{i1} = \mathrm{PD}_i$, $z_{i1} = \mathrm{INC}_i$, $z_{i2} = \mathrm{MON}_i$. With regard to the choice of the weight matrix, according to the practice in Sun et al. [24], we use the Euclidean distance in terms of any two houses to calculate the spatial weight matrix W. We use longitude and latitude to represent the location, which is denoted as $s_i = (\mathrm{Lon}_i, \mathrm{Lat}_i)$. The spatial weight w_{il} is

$$w_{il} = \exp\{-\|s_i - s_l\|\} / \sum_{k \neq i} \exp\{-\|s_i - s_k\|\}. \quad (27)$$

We set as

$$\mathrm{PD} = \frac{\left[\mathrm{PD}^{1/3} - \min\left(\mathrm{PD}^{1/3}\right)\right]}{\left[\max\left(\mathrm{PD}^{1/3}\right) - \min\left(\mathrm{PD}^{1/3}\right)\right]}, \quad (28)$$

where the transformation is used for transferring the asymmetric distribution of PD to nearly uniform distribution on $(0, 1)$. The other covariates variables of INC and MON are transformed so that the marginal distribution is approximately $N(0, 1)$. These do not alter the model but facilitate implementation. For this dataset, we adopted natural cubic P-splines, used second-order random walk penalties, and assigned $(\lambda, r_0, s_0^2, r_{\tau_{\alpha 0}}, s_{\tau_{\alpha 0}}^2, r_{\tau_{\beta j 0}}, s_{\tau_{\beta j 0}}^2)$ $= (2, 1, 1, 1, 0.005, 1, 0.005)$. The tuning parameter σ_ρ is used such that the acceptable rate for updating ρ is around 25%.

We run five times with different initial states by the proposed Gibbs sampler and generate 10000 sampled values following a burn-in of 20000 iterations in each

TABLE 1: Simulation results of parameter estimation.

Parameter	n	Rook weight matrix					(r, m)	Case weight matrix				
		Mean	Bias	SE	SD	95% CI		Mean	Bias	SE	SD	95% CI
$\rho = 0.250$		0.246	−0.004	0.048	0.047	(0.152, 0.340)		0.245	−0.005	0.043	0.044	(0.160, 0.329)
$\alpha_1 = 1.000$		0.992	−0.008	0.062	0.061	(0.870, 1.114)		0.992	−0.008	0.062	0.060	(0.870, 1.114)
$\alpha_2 = -1.000$		−0.999	0.001	0.062	0.063	(−1.122, −0.877)		−1.000	0.000	0.062	0.063	(−1.122, −0.877)
$\sigma^2 = 0.250$		0.251	0.001	0.038	0.035	(0.187, 0.335)		0.251	0.001	0.038	0.035	(0.187, 0.335)
$\rho = 0.500$		0.494	−0.006	0.041	0.041	(0.412, 0.574)		0.495	−0.005	0.031	0.031	(0.434, 0.554)
$\alpha_1 = 1.000$	100	0.993	−0.007	0.063	0.061	(0.870, 1.115)	(20, 5)	0.993	−0.007	0.063	0.060	(0.870, 1.115)
$\alpha_2 = -1.000$		−1.000	0.000	0.062	0.063	(−1.123, −0.877)		−1.000	0.000	0.062	0.063	(−1.123, −0.878)
$\sigma^2 = 0.250$		0.251	0.001	0.038	0.035	(0.187, 0.336)		0.251	0.001	0.038	0.035	(0.187, 0.336)
$\rho = 0.750$		0.744	−0.006	0.028	0.029	(0.687, 0.799)		0.747	−0.003	0.016	0.017	(0.715, 0.778)
$\alpha_1 = 1.000$		0.993	−0.007	0.063	0.062	(0.870, 1.117)		0.994	−0.006	0.063	0.060	(0.870, 1.117)
$\alpha_2 = -1.000$		−1.001	−0.001	0.063	0.063	(−1.124, −0.878)		−1.001	−0.001	0.063	0.064	(−1.125, −0.878)
$\sigma^2 = 0.250$		0.251	0.001	0.038	0.035	(0.187, 0.337)		0.251	0.001	0.039	0.035	(0.187, 0.337)
$\rho = 0.250$		0.250	0.000	0.036	0.036	(0.180, 0.320)		0.240	−0.010	0.043	0.041	(0.157, 0.323)
$\alpha_1 = 1.000$		0.994	−0.006	0.046	0.046	(0.903, 1.085)		1.001	0.001	0.048	0.049	(0.907, 1.095)
$\alpha_2 = -1.000$		−0.998	0.002	0.046	0.046	(−1.089, −0.907)		−0.995	0.005	0.048	0.049	(−1.088, −0.901)
$\sigma^2 = 0.250$		0.249	−0.001	0.028	0.027	(0.200, 0.310)		0.250	0.000	0.029	0.028	(0.199, 0.313)
$\rho = 0.500$		0.498	0.002	0.030	0.031	(0.438, 0.558)		0.493	0.007	0.029	0.028	(0.435, 0.550)
$\alpha_1 = 1.000$	169	0.994	−0.006	0.046	0.046	(0.903, 1.086)	(20, 8)	1.001	0.001	0.048	0.049	(0.908, 1.095)
$\alpha_2 = -1.000$		−0.998	0.002	0.046	0.046	(−1.089, −0.907)		−0.995	0.005	0.048	0.049	(−1.089, −0.901)
$\sigma^2 = 0.250$		0.249	−0.001	0.028	0.028	(0.200, 0.310)		0.250	0.000	0.029	0.028	(0.199, 0.313)
$\rho = 0.750$		0.748	−0.002	0.021	0.021	(0.707, 0.788)		0.746	−0.004	0.015	0.015	(0.716, 0.775)
$\alpha_1 = 1.000$		0.995	−0.005	0.047	0.046	(0.904, 1.087)		1.002	0.002	0.048	0.049	(0.908, 1.096)
$\alpha_2 = -1.000$		−0.999	0.001	0.047	0.046	(−1.090, −0.907)		−0.996	0.004	0.048	0.049	(−1.090, −0.902)
$\sigma^2 = 0.250$		0.249	−0.001	0.028	0.028	(0.200, 0.311)		0.250	0.000	0.029	0.029	(0.199, 0.314)
$\rho = 0.250$		0.249	−0.001	0.023	0.023	(0.204, 0.294)		0.248	−0.002	0.020	0.020	(0.209, 0.286)
$\alpha_1 = 1.000$		1.001	0.001	0.029	0.028	(0.943, 1.059)		1.001	0.001	0.029	0.028	(0.943, 1.059)
$\alpha_2 = -1.000$		−0.997	0.003	0.029	0.030	(−1.055, −0.940)		−0.997	0.003	0.029	0.030	(−1.055, −0.940)
$\sigma^2 = 0.250$		0.250	0.000	0.018	0.018	(0.217, 0.288)		0.250	0.000	0.018	0.18	(0.217, 0.288)
$\rho = 0.500$		0.499	−0.001	0.020	0.020	(0.460, 0.537)		0.498	−0.002	0.014	0.014	(0.470, 0.526)
$\alpha_1 = 1.000$	400	1.001	0.001	0.030	0.030	(0.943, 1.059)	(80, 5)	1.001	0.001	0.030	0.030	(0.943, 1.059)
$\alpha_2 = -1.000$		−0.997	0.003	0.030	0.030	(−1.055, −0.940)		−0.997	0.003	0.030	0.030	(−1.055, −0.940)
$\sigma^2 = 0.250$		0.250	0.000	0.018	0.018	(0.217, 0.288)		0.250	0.000	0.018	0.0018	(0.217, 0.288)
$\rho = 0.750$		0.749	−0.001	0.013	0.014	(0.722, 0.775)		0.749	−0.001	0.007	0.007	(0.734, 0.763)
$\alpha_1 = 1.000$		1.001	0.001	0.030	0.029	(0.943, 1.060)		1.002	0.002	0.030	0.029	(0.943, 1.060)
$\alpha_2 = -1.000$		−0.998	0.002	0.030	0.029	(−1.056, −0.939)		−0.998	0.002	0.030	0.030	(−1.056, −0.940)
$\sigma^2 = 0.250$		0.250	0.000	0.018	0.018	(0.217, 0.288)		0.250	0.000	0.018	0.018	(0.217, 0.289)
$\rho = 0.250$		0.247	−0.003	0.018	0.020	(0.211, 0.283)		0.247	0.003	0.020	0.019	(0.208, 0.287)
$\alpha_1 = 1.000$		1.000	0.000	0.023	0.022	(0.954, 1.046)		0.998	−0.002	0.023	0.024	(0.953, 1.044)
$\alpha_2 = -1.000$		−0.999	0.001	0.023	0.023	(−1.045, −0.953)		−1.002	−0.002	0.023	0.023	(−1.047, −0.957)
$\sigma^2 = 0.250$		0.251	0.001	0.014	0.014	(0.224, 0.280)		0.250	0.000	0.014	0.014	(0.224, 0.279)
$\rho = 0.500$		0.497	−0.003	0.016	0.017	(0.467, 0.528)		0.498	−0.002	0.014	0.013	(0.471, 0.525)
$\alpha_1 = 1.000$	625	1.001	0.001	0.024	0.023	(0.955, 1.047)	(80, 8)	0.999	−0.001	0.023	0.024	(0.953, 1.044)
$\alpha_2 = -1.000$		−0.999	0.001	0.024	0.023	(−1.045, −0.953)		−1.002	−0.002	0.023	0.023	(−1.047, −0.957)
$\sigma^2 = 0.250$		0.251	0.001	0.014	0.014	(0.224, 0.280)		0.250	0.000	0.014	0.014	(0.224, 0.279)
$\rho = 0.750$		0.748	−0.002	0.011	0.011	(0.727, 0.769)		0.749	−0.001	0.007	0.007	(0.735, 0.763)
$\alpha_1 = 1.000$		1.001	0.001	0.024	0.023	(0.955, 1.047)		0.999	−0.001	0.023	0.024	(0.953, 1.044)
$\alpha_2 = -1.000$		−0.999	0.001	0.024	0.023	(−1.046, −0.953)		−1.002	−0.002	0.023	0.023	(−1.048, −0.957)
$\sigma^2 = 0.250$		0.251	0.001	0.014	0.014	(0.224, 0.281)		0.250	0.000	0.014	0.014	(0.224, 0.280)

replication. Figure 5 plots the traces of parts of the unknown quantities. It can be seen that the five parallel sequences mix very well. Based on the five parallel sequences, the "potential scale reduction factor" $\sqrt{\widehat{R}}$ is calculated and plotted in Figure 6, from which we observe that the Markov chain has converged within 20000 burn-in iterations.

The estimated parameters together with their standard errors and 95% posterior credible intervals are reported in Table 3. The estimate of the spatial parameter is $\widehat{\rho} = 0.615$ with the 0.034

TABLE 2: The median and SD of MADE values for simulation results.

| Para.	Rook	$g_1(z)$		$g_2(z)$		Case	$g_1(z)$		$g_2(z)$	
ρ	n	Median	SD	Median	SD	(r, m)	Median	SD	Median	SD
0.250		0.102	0.034	0.090	0.031		0.102	0.034	0.090	0.030
0.500	100	0.102	0.034	0.090	0.031	(20, 5)	0.102	0.034	0.090	0.031
0.750		0.102	0.034	0.089	0.031		0.102	0.034	0.089	0.031
0.250		0.076	0.026	0.070	0.023		0.079	0.027	0.073	0.023
0.500	169	0.076	0.026	0.070	0.023	(20, 8)	0.079	0.027	0.073	0.023
0.750		0.076	0.026	0.070	0.023		0.079	0.027	0.073	0.023
0.250		0.051	0.016	0.046	0.014		0.051	0.016	0.046	0.014
0.500	400	0.051	0.016	0.046	0.014	(80, 5)	0.051	0.016	0.046	0.014
0.750		0.051	0.016	0.046	0.014		0.051	0.016	0.046	0.014
0.250		0.042	0.013	0.036	0.012		0.040	0.014	0.037	0.011
0.500	625	0.042	0.013	0.037	0.013	(80, 8)	0.040	0.014	0.037	0.011
0.750		0.042	0.013	0.036	0.013		0.040	0.014	0.037	0.011
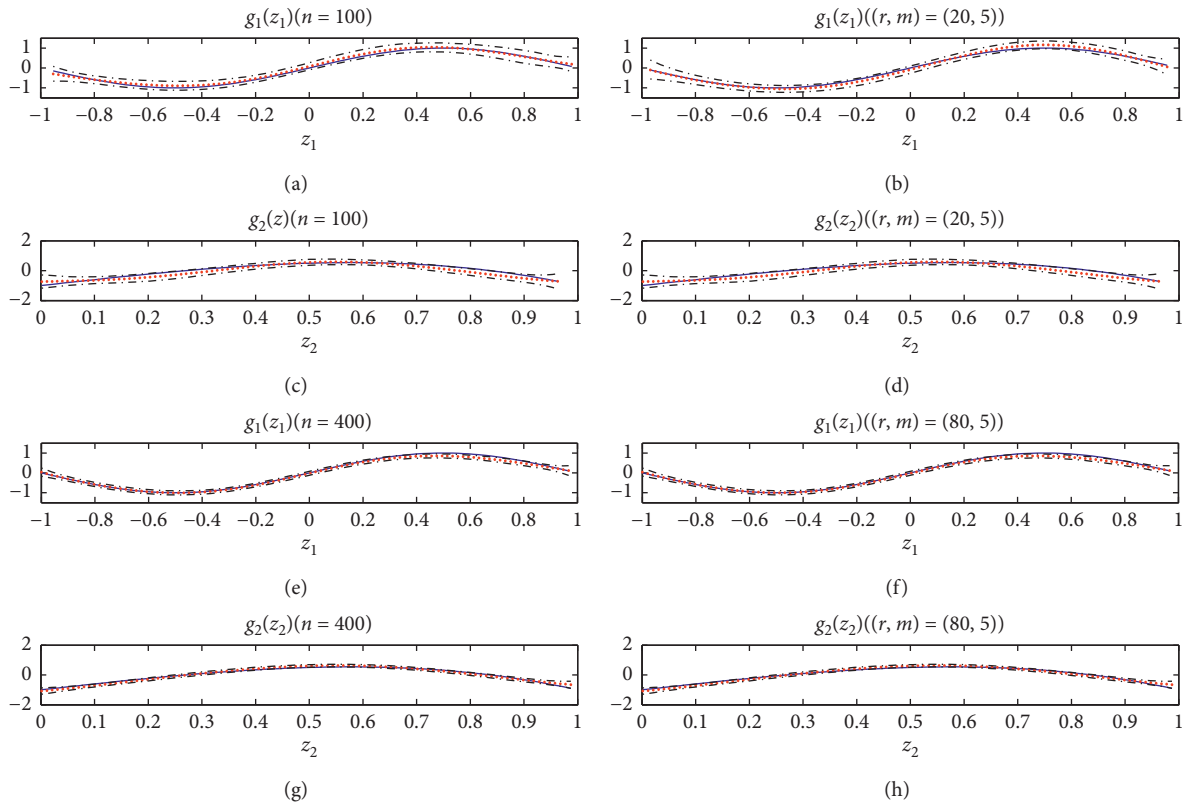
FIGURE 4: The estimated functions (dotted lines) and their 95% pointwise posterior credible intervals (dot-dashed lines) for a typical sample. The solid lines represent the true functions (the two panels on the left are based on the Rook weight matrix and the two panels on the right are based on the Case weight matrix).

standard deviation which indicates there exists a strong and positive spatial relationship among the response. We observe that the covariate PD has significant effects on the response. The regression coefficient of PD is $\hat{\alpha} = 0.258 > 0$, in which the population density has a positive effect on the housing price.

The estimated functions, together with the 95% pointwise posterior credible intervals, are depicted in Figure 7, which look like two nonlinear functions with an upward trend. The curves show that $g_1(z_1)$ and $g_2(z_2)$ have a local maximum of 0.764 at around $z_1 = 3.335$ and 0.272 at around
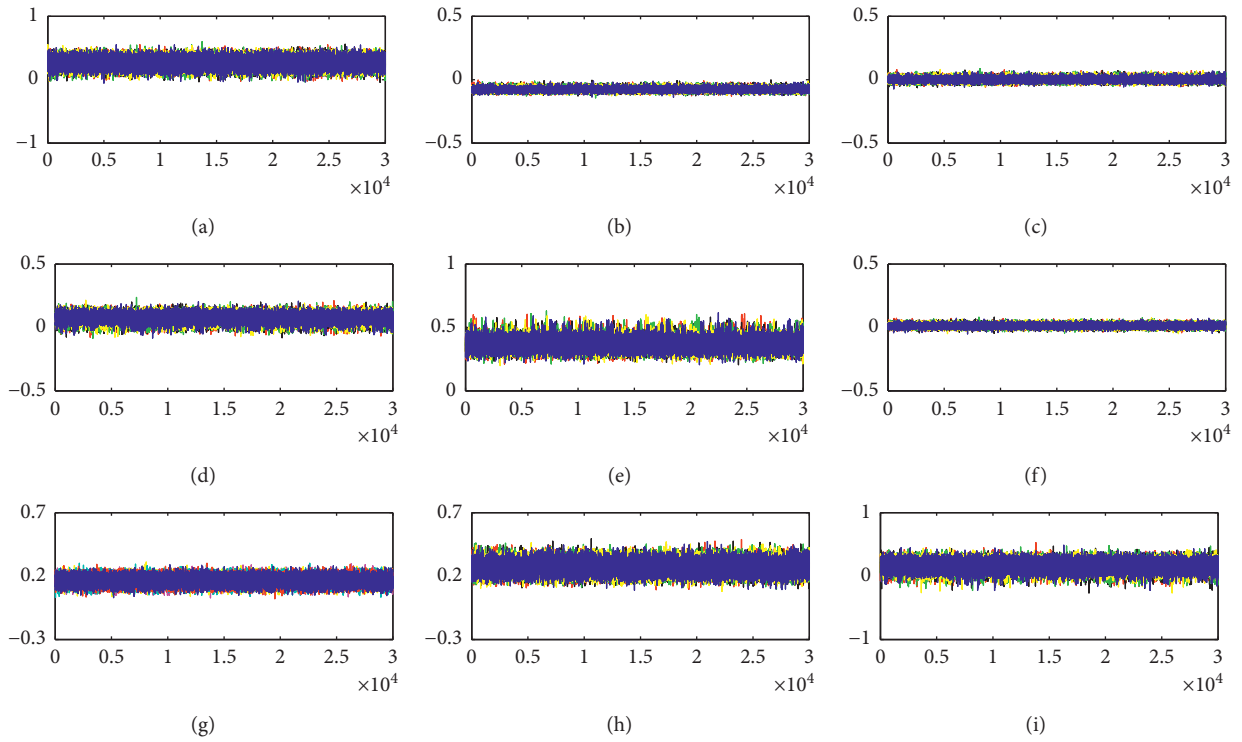
FIGURE 5: Trace plots of five parallel sequences corresponding to different starting values for parts of the unknown quantities. (a) α, (b) $g_1(z_{15})$, (c) $g_1(z_{35})$, (d) $g_1(z_{55})$, (e) $g_1(z_{75})$, (f) $g_2(z_{25})$, (g) $g_2(z_{45})$, (h) $g_2(z_{65})$, and (i) $g_2(z_{85})$.
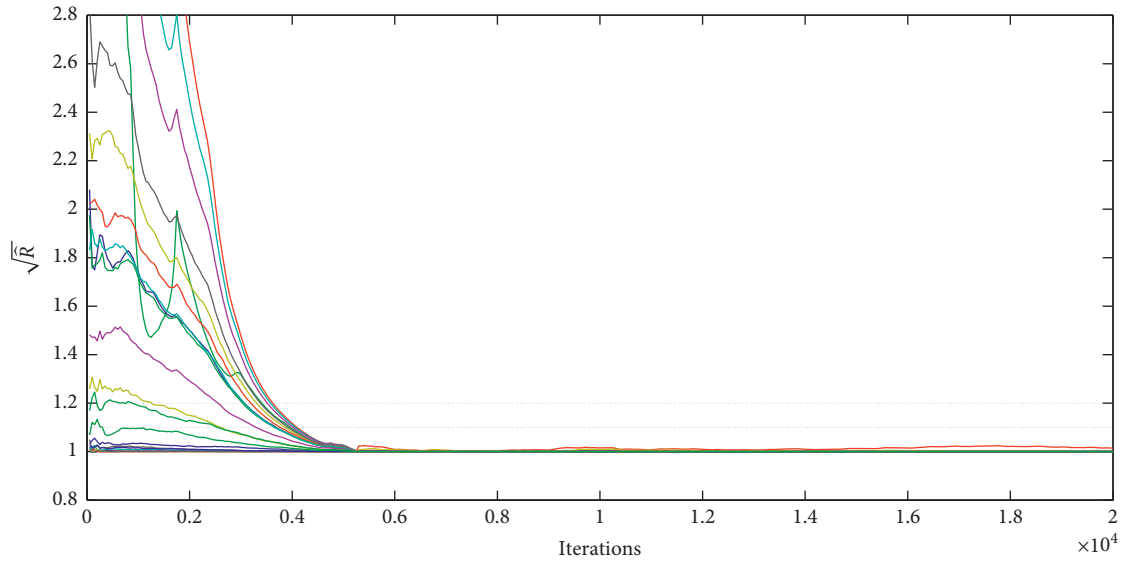

FIGURE 6: The "potential scale reduction factor" $\sqrt{\widehat{R}}$ for the housing prices of Chinese cities.

TABLE 3: The parameter estimation for the housing prices of Chinese cities.

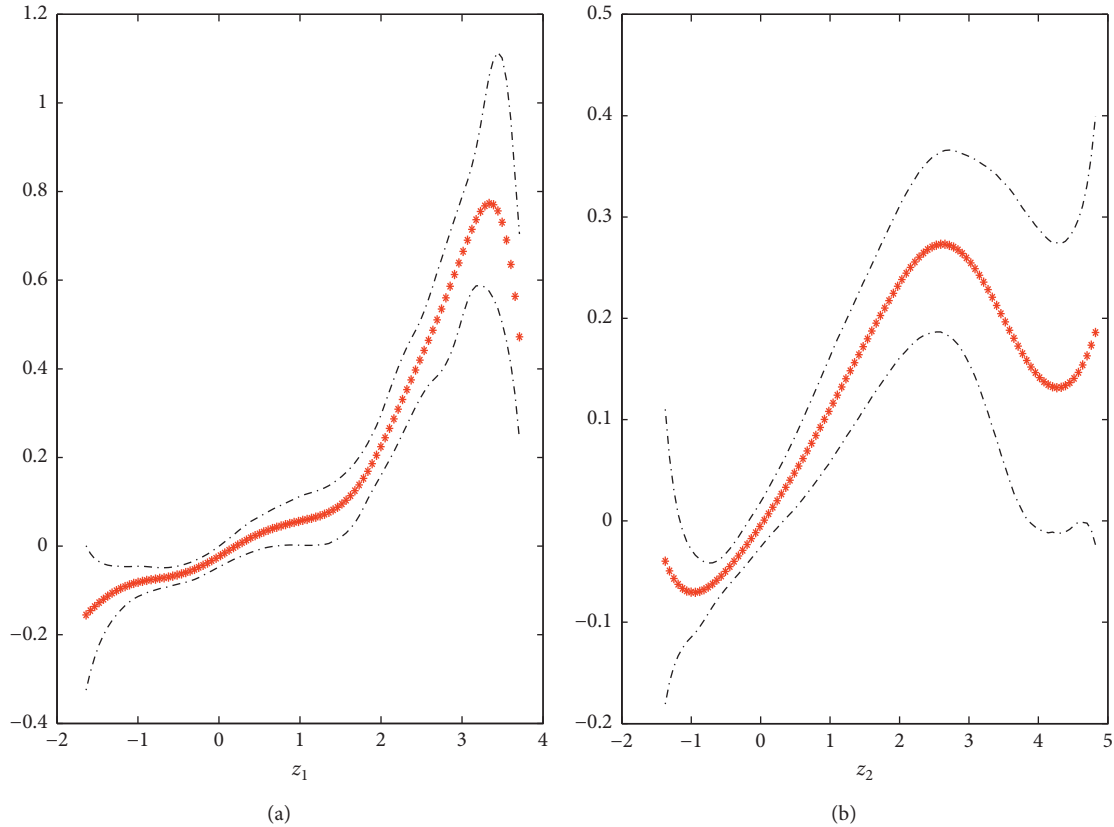| Parameter	Mean	SE	95% CI
α	0.258	0.073	(0.112, 0.401)
ρ	0.615	0.034	(0.550, 0.680)
σ^2	0.0168	0.0015	(0.0141, 0.0199)

FIGURE 7: The estimated functions (dot line) and their 95% pointwise posterior credible intervals (dot-dashed lines) for the housing prices of Chinese cities in model (26). (a) $g_1(z_1)$ and (b) $g_2(z_2)$.

$z_2 = 2.592$, respectively. It provides evidence that the effects of the covariates INC and MON on the response are significant and nonlinear.

5. Summary

Spatial data are often encountered in economics, geography, and environmental science and can be analyzed by the spatial autoregressive models. To reduce the high risk of misspecification of the traditional spatial autoregressive models and avoid some serious drawbacks of fully nonparametric models, we have developed a PLASARM for spatial data, which combines the partially linear additive model and spatial autoregressive model. In this paper, we have considered a Bayesian MCMC approach with P-splines to analyze the proposed model and designed a Gibbs sampler to explore the joint posterior distributions based on the Bayesian P-splines technique. A simulation study demonstrates that the proposed model and methodology perform satisfactorily in the finite sample. The results have shown that they are reliable for recovering the true parameters and nonparametric components. We have further applied the proposed model and methodology to analyze a real dataset.

Various extensions can be considered in the future. We can pursue the case where the covariates are high dimension in the model. The proposed models can be generalized in several ways. Although we focus on a PLASARM to assess

the effects of the covariates on the response, the other semiparametric models such as single-index spatial autoregressive models and varying-coefficient spatial autoregressive models can also be considered. In addition, it is straightforward to generalize the model and methodology to handle more complicated data types, for example, mixed discrete, continuous, and ordinal data [55]. We leave these topics for future research.

Data Availability

The data are obtained from the China City Statistical Yearbook 2013 [56] and the China Statistical Yearbook for Regional Economy 2013 [57]. The data can be made available from the corresponding author or obtained from the above Yearbook.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

References

[1] J. P. Elhorst, *Spatial Econometrics: From Cross-Sectional Data to Spatial Panels*, Springer, Heidelberg, Germany, 2014.

[2] L. Anselin, *Spatial Econometrics: Methods and Models*, Kluwer Academic Publishers, Amsterdam, The Netherlands, 1988.

[3] N. Cressie, "Statistics for spatial data," *Terra Nova*, vol. 4, no. 5, pp. 613–617, 1992.

[4] J. P. LeSage, "Bayesian estimation of spatial autoregressive models," *International Regional Science Review*, vol. 20, no. 1-2, pp. 113–129, 1997.

[5] J. P. LeSage, "Bayesian estimation of limited dependent variable spatial autoregressive models," *Geographical Analysis*, vol. 32, no. 1, pp. 19–35, 2000.

[6] L. Anselin and A. K. Bera, "Spatial dependence in linear regression models with an introduction to spatial econometrics," in *Handbook of Applied Economics Statistics* Marcel Dekker, New York, NY, USA, 1998.

[7] L.-F. Lee, "Asymptotic distributions of quasi-maximum likelihood estimators for spatial autoregressive models," *Econometrica*, vol. 72, no. 6, pp. 1899–1925, 2004.

[8] L.-f. Lee, "GMM and 2SLS estimation of mixed regressive, spatial autoregressive models," *Journal of Econometrics*, vol. 137, no. 2, pp. 489–514, 2007.

[9] K. Kakamu and H. Wago, "Small-sample properties of panel spatial autoregressive models: comparison of the Bayesian and maximum likelihood methods," *Spatial Economic Analysis*, vol. 3, no. 3, pp. 305–319, 2008.

[10] P. J. LeSage and R. K. Pace, *Introduction to Spatial Econometrics*, CRC Press, Boca Raton, FL, USA, 2009.

[11] B. M. Kazar and M. Celik, *Spatial Autoregressive Model*, Springer Press, New York, NY, USA, 2012.

[12] P. Piribauer and J. Crespo Cuaresma, "Bayesian variable selection in spatial autoregressive models," *Spatial Economic Analysis*, vol. 11, no. 4, pp. 457–479, 2016.

[13] J. H. P. Paelinck and L. H. Klaassen, *Spatial Econometrics*, Gower Press, Aldershot, England, 1979.

[14] R. Basile, "Regional economic growth in Europe: a semiparametric spatial dependence approach," *Papers in Regional Science*, vol. 87, no. 4, pp. 527–544, 2008.

[15] R. Basile, "Productivity polarization across regions in europe," *International Regional Science Review*, vol. 32, no. 1, pp. 92–115, 2009.

[16] R. Basile and B. Gress, "Semi-parametric spatial auto-covariance models of regional growth behaviour in Europe," *Région et Développement*, vol. 21, pp. 93–118, 2005.

[17] R. E. Bellman, *Adaptive Control Processes*, Princeton University Press, Princeton, NJ, USA, 1961.

[18] J. H. Friedman and W. Stuetzle, "Projection pursuit regression," *Journal of the American Statistical Association*, vol. 76, no. 376, pp. 817–823, 1981.

[19] H.-B. Wang, "A Bayesian multivariate partially linear single-index probit model for ordinal responses," *Journal of Statistical Computation and Simulation*, vol. 88, no. 8, pp. 1616–1636, 2018.

[20] T. J. Hastie and R. J. Tibshirani, *Generalized Additive Models*, Chapman and Hall, London, UK, 1990.

[21] T. Hastie and R. Tibshirani, "Varying-coefficient models," *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 55, no. 4, pp. 757–779, 1993.

[22] L. Su and S. Jin, "Profile quasi-maximum likelihood estimation of partially linear spatial autoregressive models," *Journal of Econometrics*, vol. 157, no. 1, pp. 18–33, 2010.

[23] L. Su, "Semiparametric GMM estimation of spatial autoregressive models," *Journal of Econometrics*, vol. 167, no. 2, pp. 543–560, 2012.

[24] Y. Sun, H. J. Yan, W. Y. Zhang, and Z. Lu, "A Semiparametric spatial dynamic model," *Annals of Statistics*, vol. 42, no. 2, pp. 700–727, 2014.

[25] J. Q. Chen, R. F. Wang, and Y. X. Huang, "Semiparametric spatial autoregressive model: a two-step Bayesian approach," *Annals of Public Health and Research*, vol. 2, no. 1, p. 1012, 2015.

[26] H. Wei and Y. Sun, "Heteroskedasticity-robust semi-parametric GMM estimation of a spatial model with space-varying coefficients," *Spatial Economic Analysis*, vol. 12, no. 1, pp. 113–128, 2016.

[27] T. Hoshino, "Semiparametric spatial autoregressive models with endogenous regressors: with an application to crime data," *Journal of Business & Economic Statistics*, vol. 36, no. 1, pp. 160–172, 2017.

[28] T. Krisztin, "The determinants of regional freight transport: a spatial, semiparametric approach," *Geographical Analysis*, vol. 49, no. 3, pp. 268–308, 2017.

[29] C. Wei, S. Guo, and S. Zhai, "Statistical inference of partially linear varying coefficient spatial autoregressive models," *Economic Modelling*, vol. 64, pp. 553–559, 2017.

[30] Y. Sun and Y. Wu, "Estimation and testing for a partially linear single-index spatial regression model," *Spatial Economic Analysis*, vol. 13, no. 4, pp. 473–489, 2018.

[31] S. Cheng, J. Chen, and X. Liu, "GMM Estimation of partially linear single-index spatial autoregressive model," *Spatial Statistics*, vol. 31, p. 100354, 2019.

[32] J. Du, X. Sun, R. Cao, and Z. Zhang, "Statistical inference for partially linear additive spatial autoregressive models," *Spatial Statistics*, vol. 25, pp. 52–67, 2018.

[33] Y. Hu, S. Wu, S. Feng, and J. Jin, "Estimation in partial functional linear spatial autoregressive model," *Mathematics*, vol. 8, no. 10, p. 1680, 2020.

[34] H. Liang, S. W. Thurston, D. Ruppert, T. Apanasovich, and R. Hauser, "Additive partial linear models with measurement errors," *Biometrika*, vol. 95, no. 3, pp. 667–678, 2008.

[35] L. Wang and L. J. Yang, "Spline-backfitted kernel smoothing of nonlinear additive autoregression model," *Annals of Statistics*, vol. 35, pp. 2474–2503, 2007.

[36] J. Zhang and H. Lian, "Partially linear additive models with unknown link functions," *Scandinavian Journal of Statistics*, vol. 45, no. 2, pp. 255–282, 2018.

[37] G. Deng and H. Liang, "Model averaging for semiparametric additive partial linear models," *Science China Mathematics*, vol. 53, no. 5, pp. 1363–1376, 2010.

[38] P. Du, G. Cheng, and H. Liang, "Semiparametric regression models with additive nonparametric components and high dimensional parametric components," *Computational Statistics & Data Analysis*, vol. 56, no. 6, pp. 2006–2017, 2012.

[39] L. Wang, X. Liu, H. Liang, and R. J. Carroll, "Estimation and variable selection for generalized additive partial linear models," *Annals of Statistics*, vol. 39, no. 4, pp. 1827–1851, 2011.

[40] X. Liu, L. Wang, and H. Liang, "Estimation and variable selection for semiparametric additive partial linear models," *Statistica Sinica*, vol. 21, no. 3, pp. 1225–1248, 2011.

[41] J. Guo, M. Tang, M. Tian, and K. Zhu, "Variable selection in high-dimensional partially linear additive models for composite quantile regression," *Computational Statistics & Data Analysis*, vol. 65, pp. 56–67, 2013.

[42] Y. Hu, K. Zhao, and H. Lian, "Bayesian quantile regression for partially linear additive models," *Statistics and Computing*, vol. 25, no. 3, pp. 651–668, 2015.

[43] H. Lian, "Semiparametric estimation of additive quantile regression models by two-fold penalty," *Journal of Business & Economic Statistics*, vol. 30, no. 3, pp. 337–350, 2012.

[44] B. Sherwood and L. Wang, "Partially linear additive quantile regression in ultra-high dimension," *Annals of Statistics*, vol. 44, pp. 288–317, 2016.

[45] K. Yu and Z. Lu, "Local linear additive quantile regression," *Scandinavian Journal of Statistics*, vol. 31, no. 3, pp. 333–346, 2004.

[46] S. Lang and A. Brezger, "Bayesian P-splines," *Journal of Computational and Graphical Statistics*, vol. 13, no. 1, pp. 183–212, 2004.

[47] C. De Boor, *A Practical Guide to Splines*, Springer, New York, NY, USA, 1978.

[48] P. H. Eilers and B. D. Marx, "Flexible smoothing with B-splines and penalties," *Statistical Science*, vol. 11, pp. 89–102, 1996.

[49] W. K. Hastings, "Monte Carlo sampling methods using Markov chains and their applications," *Biometrika*, vol. 57, no. 1, pp. 97–109, 1970.

[50] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, "Equation of state calculations by fast computing machines," *The Journal of Chemical Physics*, vol. 21, no. 6, pp. 1087–1092, 1953.

[51] M. A. Tanner, *Tools for Statistical Inference: Methods for the Exploration of Posterior Distributions and Likelihood Functions*, Springer-Verlag, New York, NY, USA, 2nd edition, 1993.

[52] A. Panagiotelis and M. Smith, "Bayesian identification, selection and estimation of semiparametric functions in high-dimensional additive models," *Journal of Econometrics*, vol. 143, no. 2, pp. 291–316, 2008.

[53] A. C. Case, "Spatial patterns in household demand," *Econometrica*, vol. 59, no. 4, pp. 953–965, 1991.

[54] A. Gelman and D. B. Rubin, "Inference from iterative simulation using multiple sequences," *Statistical Science*, vol. 7, pp. 457–511, 1992.

[55] Z.-Y. Chen and H.-B. Wang, "Latent single-index models for ordinal data," *Statistics and Computing*, vol. 28, no. 3, pp. 699–711, 2018.

[56] National Bureau of Statistics of China, *China City Statistical Yearbook*, China Statistics Press, Beijing, China, 2013.

[57] National Bureau of Statistics of China, *China Statistical Yearbook for Regional Economy*, China Statistics Press, Beijing, China, 2013.