



Article

Bayesian Analysis of Partially Linear Additive Spatial Autoregressive Models with Free-Knot Splines

Zhiyong Chen  and Jianbao Chen * 

School of Mathematics and Statistics, Fujian Normal University, Fuzhou 350117, China; chenzy@fjnu.edu.cn

* Correspondence: jbchen@fjnu.edu.cn

Abstract: This article deals with symmetrical data that can be modelled based on Gaussian distribution. We consider a class of partially linear additive spatial autoregressive (PLASAR) models for spatial data. We develop a Bayesian free-knot splines approach to approximate the nonparametric functions. It can be performed to facilitate efficient Markov chain Monte Carlo (MCMC) tools to design a Gibbs sampler to explore the full conditional posterior distributions and analyze the PLASAR models. In order to acquire a rapidly-convergent algorithm, a modified Bayesian free-knot splines approach incorporated with powerful MCMC techniques is employed. The Bayesian estimator (BE) method is more computationally efficient than the generalized method of moments estimator (GMME) and thus capable of handling large scales of spatial data. The performance of the PLASAR model and methodology is illustrated by a simulation, and the model is used to analyze a Sydney real estate dataset.

Keywords: spatial data; partially linear additive model spatial autoregressive; free-knot splines; MCMC tools; Gibbs sampler



Citation: Chen, Z.; Chen, J. Bayesian Analysis of Partially Linear Additive Spatial Autoregressive Models with Free-Knot Splines. *Symmetry* **2021**, *13*, 1635. <https://doi.org/10.3390/sym13091635>

Academic Editors: Célia Nunes, Miguel Fonseca, Elsa Moreira and Juan Luis García Guirao

Received: 5 July 2021

Accepted: 2 September 2021

Published: 6 September 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Spatial econometrics models are frequently proposed to analyze spatial data that arise in many disciplines such as urban, real estate, public, agricultural, environmental economics and industrial organizations. These models address relationships across geographic observations caused by spatial autocorrelation in cross-sectional or longitudinal data. Spatial econometrics models have a long history in both econometrics and statistics. Early developments and relevant surveys can be found in Cliff and Ord [1], Anselin [2], Case [3], Cressie [4], LeSage [5,6], Anselin and Bera [7].

Among spatial econometrics models, spatial autoregressive (SAR) models [1] have gained much attention by theoretical econometricians and applied researchers. Many approaches have been used to estimate the SAR models, which include the maximum likelihood estimator (MLE) [8], the generalized method of moment estimator (GMME) [9], and the quasi-maximum likelihood estimator (QMLE) [10]. However, these methods mainly focused on parametric SAR models, which are frequently assumed to be linear, few researchers have explicitly examined non-/semi-parametric SAR models. Indeed, it has been confirmed that a lot of economic variables exhibit highly nonlinear relationships on the dependent variables [11–13]. Neglecting the latent nonlinear functional forms often results in an inconsistent estimation of the parameters and misleading conclusions [14].

Although many empirical studies and econometric analyses applying the parametric SAR models ignore latent nonlinear relationships, several nonlinear forms [15–18] have been considered. Nevertheless, the nonlinear parametric SAR models can at most supply certain protection against some specific nonlinear functional forms. Since the nonlinear function is unknown, it is unavoidable to assume the risk of misspecifying the nonlinear function. As nonparametric techniques advance, the advantage of nonparametric SAR models are often used to model nonlinear economic relationships. However,

nonparametric components are only suitable for low dimensional covariates, otherwise the “curse of dimensionality” [19] problem is often encountered. Some nonparametric dimension-reduction tools have been considered to address this problem, for example, single-index model [20], partially linear model [21], the additive model [22], varying-coefficient model [23], among others. In recent years, many researchers have started using the advantage of semiparametric modeling in spatial econometrics. For example, Su and Jin [14] proposed the QMLE for semiparametric partially linear SAR models; Su [24] discussed GMME of semiparametric SAR models; Chen et al. [25] studied a Bayesian method for the semiparametric SAR models; Wei and Sun [26] considered GMME for the space-varying coefficients of a spatial model; Krisztin [27] investigated a novel Bayesian semiparametric estimation for the penalized spline SAR models; Krisztin [28] presented a genetic-algorithms for a nonlinear SAR model; Du et al. [29] established GMME of PLASAR models; Chen and Cheng [30] developed a GMME of a partially linear additive spatial error model.

Semiparametric models have received much attention from both econometrics and statistics owing to the explanatory of the parameters and the flexibility of nonparameters. The partially linear additive (PLA) model is probably one of the most popular among the various semiparametric models. As they can not only avoid the “curse of dimensionality” phenomenon encountered in nonparametric regression but also provide a more flexible structure than the generalized linear models. As a result, the PLA models provide good equilibrium between flexibility of the additive model and interpretation of the partially linear model. Many researchers have considered many approaches to analyze such models: local linear method [31], spline estimation [32–34], quantile regression [35–38], variable selection [39–42], etc. Characterizing the flexibility of nonparametric forms and attempting to explain the potential nonlinearity of PLASAR models are unique challenges faced by analysts of spatial data.

Combining PLA models with SAR models, we consider a class of PLASAR models for spatial data to capture the linear and nonlinear effects between the related variables in addition to spatial dependence between the neighbors in this article. We specify the prior of all unknown parameters, which led to a proper posterior distribution. The posterior summaries are obtained via MCMC tools. We develop an improved Bayesian method with free-knot splines [43–51], along with MCMC techniques to estimate unknown parameters, use a spline approach to approximate the nonparametric functions, and design a Gibbs sampler to explore the joint posterior distributions. Treating the number and the positions of knots as random variables can make the model spatially adaptive is an attractive feature of Bayesian free-knot splines [45,46]. In order to improve the rapidly-convergent performance of our algorithm, we further modify the movement step of Bayesian free-knots splines such that all knots can be repositioned in each iteration instead of only one knot moving. Finally, the performance of the PLASAR model and methodology is illustrated by a simulation, and they are used to analyze real data.

The rest of this paper is organized as follows. In Section 2, we propose the PLASAR model to analyze spatial data and discuss the proposed model’s identifiability condition, then acquire the likelihood function by fitting the nonparametric functions with a Bayesian free-knots splines approach. To provide a Bayesian framework, we specify the priors for the unknown parameters, derive the full conditional posterior distributions of the unknown parameters, modify the movement step of the Bayesian free-knots splines approach to accelerate the convergent performance of our algorithm, and describe the detailed sampling algorithm in Section 3. The applicability and practicality of the PLASA model and methodology for spatial data are evaluated by a simulation study, and the model is used to analyze a real dataset in Section 4. Section 5 concludes the paper with a summary.

2. Methodology

2.1. Model

We begin with the PLASAR model that is defined as

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

where $x_i = (x_{i1}, \dots, x_{iq})^T$ and $z_i = (z_{i1}, \dots, z_{ip})^T$ are covariate vectors, y_i is a response variable, w_{il} is a specified constant spatial weight, $g_j(\cdot)$ is unknown univariate nonparametric function for $j = 1, \dots, p$, $\alpha = (\alpha_1, \dots, \alpha_q)^T$ is $q \times 1$ vector of the unknown parameters, the unknown spatial parameter ρ reflects the spatial autocorrelation between the neighbors with stability condition $|\rho| < 1$, and ε_i 's are mutually independent and identically distributed normal with zero mean and variance σ^2 . In order to ensure model identifiability of the nonparametric function, it is often assumed that the condition $E[g_j(z_j)] = 0$ for $j = 1, \dots, p$.

2.2. Likelihood

We plan on approximating unknown functions $g_j(\cdot)$ in (1) by free-knot splines for $j = 1, \dots, p$. Assuming that $g_j(\cdot)$ has a polynomial spline of degree m_j with k_j order interior knots $\xi_j = (\xi_{j1}, \dots, \xi_{jk})^T$ with $a_j < \xi_{j1} < \dots < \xi_{jk} < b_j$, i.e.,

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

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

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

are boundary knots for $j = 1, \dots, p$. Let $\mathbf{B}_j = (B_1(z_{1j}), \dots, B_n(z_{nj}))^T$ and $\mathbf{1} = (1, \dots, 1)^T$. To achieve identification, we set $\sum_{i=1}^n \sum_{l=1}^{K_j} B_{jl}(z_{ij}) \beta_{jl} = 0$, which is written as $\mathbf{1}^T \mathbf{B}_j \beta_j = 0$. Denote $Q_j = \mathbf{1}^T \mathbf{B}_j$, then the constraint becomes $Q_j \beta_j = 0$.

It follows that the model (1) can be equivalent to

$$y_i = \rho \sum_{l=1}^n w_{il} y_l + x_i^T \alpha + \sum_{j=1}^p B_j^T(u_j) \beta_j + \varepsilon_i = \rho \sum_{l=1}^n w_{il} y_l + x_i^T \alpha + B_i^T(u) \beta + \varepsilon_i, \quad i = 1, \dots, n,$$

where $\beta = (\beta_1^T, \dots, \beta_p^T)^T$ and $B_i^T(u) = (B_1^T(u_1), \dots, B_p^T(u_p))^T$. Then the matrix form of the model (1) can be represented as

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

where $x = (x_1, \dots, x_n)^T$, $y = (y_1, \dots, y_n)^T$, $\varepsilon = (\varepsilon_1, \dots, \varepsilon_n)^T$, $W = (w_{il})$ is an $n \times n$ specified constant spatial weight matrix, $K = \sum_{j=1}^p K_j$, and $B^T(u)$ is an $n \times K$ matrix with $B_i^T(u)$ as its i th row.

The likelihood function for the PLASAR model is proportional to

$$\begin{aligned}
 & L(\alpha, \beta, k, \xi, \sigma^2, \rho | y, x, z) \\
 \propto & \sigma^{-n} |I_n - \rho W| \exp \left\{ -\frac{1}{2\sigma^2} [y - \rho W y - x^T \alpha - B^T(u) \beta]^T [y - \rho W y - x \alpha - B^T(u) \beta] \right\} \\
 = & \sigma^{-n} |A(\rho)| \exp \left\{ -\frac{1}{2\sigma^2} [A(\rho)y - x \alpha - B^T(u) \beta]^T [A(\rho)y - x^T \alpha - B^T(u) \beta] \right\} \tag{4} \\
 \doteq & \sigma^{-n} |A(\rho)| \exp \left\{ -\frac{1}{2\sigma^2} [A(\rho)y - \mathfrak{B}(x, u)\theta]^T [A(\rho)y - \mathfrak{B}(x, u)\theta] \right\},
 \end{aligned}$$

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

3. Bayesian Estimation

In this section, we consider a Bayesian free-knots splines approach with MCMC techniques to analyze the PLASAR model. We begin with the specification of the prior distributions of all unknown parameters, then the derivations of posterior distributions and the narration of the detailed sampling scheme for all of the unknown parameters. Meanwhile, we modify the movement step of Bayesian free-knots splines approach so that all the knots can be repositioned in each iteration.

3.1. Priors

As we will consider a Bayesian approach with free-knots splines to analyze the PLASAR models, all unknown parameters are assigned prior distributions. Note that besides regression coefficient vectors θ , the spatial autocorrelation coefficient ρ and the quantities σ^2 , the number of knots $k = (k_1^T, \dots, k_p^T)^T$ and location of knots $\xi = (\xi_1^T, \dots, \xi_p^T)^T$ also need prior distributions in the sense that they are random variables in the Bayesian approach with free-knot splines. We avoid the use of improper prior distributions to prevent improper joint posterior distributions.

For $j = 1, \dots, p$, we followed Poon and Wang [49] by putting a Poisson prior with mean λ_j for number k_j of the knots

$$\pi(k_j) = \frac{\lambda_j^{k_j}}{k_j!} e^{-\lambda_j}$$

and a conditional flat prior for knot location ξ_j

$$\pi(\xi_j | k_j) = \frac{k_j!}{(b_j - a_j)^{k_j}} \Delta_j,$$

where $\Delta_j = I\{a_j = \xi_{j0} < \xi_{j1} < \dots < \xi_{jk_j} < \xi_{j,k_j+1} = b_j\}$, a_j and b_j are defined in (3).

We set a conjugate normal inverse-gamma prior for the unknown parameters (θ, σ^2) , which is a composite of inverse-Gamma prior distributions for σ^2

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

where r_0 and s_0^2 are hyperparameters; a conditional normal prior distribution with mean vector $\mathbf{0}$ and covariance matrix $\tau_0 \sigma^2 I_q$ for α

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

and a conditional normal prior distribution with mean vector $\mathbf{0}$ and covariance matrix $\tau_j \sigma^2 I_{K_j}$ for β_j with the constraint $Q_j \beta_j = 0$ as follows:

$$\pi(\beta_j | k_j, \xi_j, \tau_j, \sigma^2) \propto (2\pi\tau_j\sigma^2)^{-\frac{K_j}{2}} \exp\left\{-\frac{\beta_j^T \beta_j}{2\tau_j\sigma^2}\right\} I\{Q_j \beta_j = 0\}$$

for $j = 1, \dots, p$. In order to improve the robustness of our method, we choose an inverse-gamma prior

$$\pi(\tau_0) \propto \tau_0^{-\frac{r_{\tau_0 0}}{2}-1} \exp\left\{-\frac{s_{\tau_0 0}^2}{2\tau_0}\right\} \text{ and } \pi(\tau_j) \propto \tau_j^{-\frac{r_{\tau_j 0}}{2}-1} \exp\left\{-\frac{s_{\tau_j 0}^2}{2\tau_j}\right\}$$

for $j = 0, 1, \dots, p$, where r_{τ_0} and $s_{\tau_0}^2$ are pre-specified hyper-parameters. Throughout this article we set $r_0 = s_0^2 = 1$ to obtain a Cauchy distribution of σ^2 and assign $r_{\tau_0} = r_{\tau_{\beta_j 0}} = 1$ and $s_{\tau_0}^2 = s_{\tau_{\beta_j 0}}^2 = 0.005$ to acquire a highly dispersed inverse gamma prior on τ_j for $j = 0, 1, \dots, p$.

In addition, we follow LeSage and Pace [52] by eliciting a uniform prior $U(\lambda_{\min}^{-1}, \lambda_{\max}^{-1})$ for the spatial autocorrelation coefficient ρ

$$\pi(\rho) \propto 1,$$

where λ_{\max} and λ_{\min} are the maximum and minimum eigenvalues of the standardized spatial weight matrix W , respectively.

Therefore, the joint priors of all of the quantities are defined as

$$\pi(\rho, \alpha, \beta, k, \xi, \sigma^2, \tau) = \pi(\rho)\pi(\sigma^2)\pi(\tau_0)\pi(\alpha|\sigma^2, \tau_0) \prod_{i=1}^p \pi(\beta_i | k_i, \xi_i, \tau_i, \sigma^2) \pi(k_i) \pi(\xi_i | k_i) \pi(\tau_i), \quad (5)$$

where $\tau = (\tau_0, \tau_1, \dots, \tau_p)$ is a hyperparameters vector. For computational convenience, we have treated the hyperparameter vector τ as a unknown parameter vector.

3.2. The Full Conditional Posterior Distributions of Unknown Quantities

Since the joint posterior distribution of the quantities is very complicated, it is difficult to generate samples directly. To solve this problem, we derive the full conditional posterior distributions of unknown quantities, modify the movement step of Bayesian free-knots splines to speed up the convergence, and describe the detailed sampling method in our algorithm.

It follows from the likelihood function (4) and the joint priors (5) that the conditional posterior distribution of ρ given the remaining unknown parameters is proportional to

$$p(\rho | y, x, z, \alpha, \beta, k, \xi, \sigma^2, \tau) \propto |A(\rho)| \exp\left\{-\frac{1}{2\sigma^2} [A(\rho)y - x^T \alpha - B^T(u)\beta]^T [A(\rho)y - x^T \alpha - B^T(u)\beta]\right\}. \quad (6)$$

It is not easy to directly simulate from (6), which does not have the form of any standard density function. Therefore, we prefer the Metropolis–Hastings algorithm [53,54] to solve this difficulty: draw ρ^* from a truncated Cauchy distribution with location ρ and scale σ_ρ on $(-1, 1)$, where σ_ρ is treated as a tuning parameter; and accept the candidate value ρ^* with probability

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

where

$$C_\rho = \frac{\arctan[(1-\rho)/\sigma_\rho] - \arctan[(-1-\rho)/\sigma_\rho]}{\arctan[(1-\rho^*)/\sigma_\rho] - \arctan[(-1-\rho^*)/\sigma_\rho]}.$$

From likelihood function (4) and priors (5), we can see that given (ρ, τ) , the joint posterior of $(\alpha, \beta, k, \xi, \sigma^2)$ is given by

$$\begin{aligned}
 & p(\alpha, \beta, k, \xi, \sigma^2 | x, y, z, \rho, \tau) \\
 \propto & \sigma^{-n} \exp \left\{ -\frac{1}{2\sigma^2} [A(\rho)y - x^T \alpha - B^T(u)\beta]^T [A(\rho)y - x^T \alpha - B^T(u)\beta] \right\} \\
 & \times \sigma^{-r_0 - q - 2} \exp \left\{ -\frac{s_0^2}{2\sigma^2} - \frac{\alpha^T \alpha}{2\tau_0 \sigma^2} \right\} \times \prod_{j=1}^p \left(\frac{\lambda_j}{b_j - a_j} \right)^{k_j} \Delta_j \\
 & \times \prod_{j=1}^p (2\pi\tau_j \sigma^2)^{-\frac{k_j}{2}} \exp \left\{ -\frac{\beta_j^T \beta_j}{2\tau_j \sigma^2} \right\} I\{Q_j \beta_j = 0\} \\
 \propto & \sigma^{-n - r_0 - K - q - 2} \exp \left\{ -\frac{1}{2\sigma^2} [A(\rho)y - \mathfrak{B}(x, u)\theta]^T [A(\rho)y - \mathfrak{B}(x, u)\theta] \right\} \quad (7) \\
 & \times \exp \left\{ -\frac{s_0^2}{2\sigma^2} - \frac{\theta^T \text{diag}\{\tau^{-1}\}\theta}{2\sigma^2} \right\} I\{Q\beta = \mathbf{0}\} \times \prod_{j=1}^p \left(\frac{\lambda_j \tau_j}{b_j - a_j} \right)^{k_j} \Delta_j \\
 \propto & |\Xi|^{-\frac{1}{2}} (S^2 + s_0^2)^{-\frac{n+r_0}{2}} \prod_{j=1}^p \left(\frac{\lambda_j \tau_j^{-\frac{1}{2}}}{b_j - a_j} \right)^{k_j} \Delta_j \\
 & \times (S^2 + s_0^2)^{\frac{n+r_0}{2}} (\sigma^2)^{-\frac{n+r_0}{2} - 1} \exp \left\{ -\frac{S^2 + s_0^2}{2\sigma^2} \right\} \\
 & \times (2\pi\sigma^2)^{-\frac{K+q}{2}} |\Xi|^{\frac{1}{2}} \exp \left\{ -\frac{1}{2\sigma^2} (\theta - \hat{\theta})^T \Xi (\theta - \hat{\theta}) \right\} I\{Q\beta = \mathbf{0}\},
 \end{aligned}$$

where $\theta = (\alpha^T, \beta^T)^T$, $\text{diag}\{\tau^{-1}\} = \text{diag}\{\tau_0^{-1}I_q, \tau_1^{-1}I_{K_1}, \dots, \tau_p^{-1}I_{K_p}\}$, $Q = (Q_1, \dots, Q_p)$, $\Xi = \text{diag}\{\tau^{-1}\} + \mathfrak{B}^T(x, u)\mathfrak{B}(x, u)$, $\hat{\theta} = \Xi^{-1}\mathfrak{B}^T(x, u)A(\rho)y$, and $S^2 = y^T A(\rho)^T A(\rho)y - \hat{\theta}^T \Xi \hat{\theta}$, which gives rise to a marginal posterior distribution

$$p(k, \xi | y, x, z, \rho, \tau) \propto |\Xi|^{-\frac{1}{2}} (S^2 + s_0^2)^{-\frac{n+r_0}{2}} \prod_{j=1}^p \left(\frac{\lambda_j \tau_j^{-\frac{1}{2}}}{b_j - a_j} \right)^{k_j} \Delta_j. \quad (8)$$

It is easy to see from (8) that

$$p(k_j, \xi_j | y, x, z, \rho, \alpha, k_{-j}, \xi_{-j}, \beta_{-j}, \tau) \propto |\Xi_j|^{-\frac{1}{2}} (S_j^2 + s_0^2)^{-\frac{n+r_0}{2}} \left(\frac{\lambda_j \tau_j^{-\frac{1}{2}}}{b_j - a_j} \right)^{k_j} \Delta_j, \quad j = 1, \dots, p. \quad (9)$$

where $\Xi_j = \tau_j^{-1}I_{K_j} + B_j(u_j)B_j^T(u_j)$, $y_j^* = A(\rho)y - x^T \alpha - B_{-j}^T(u_{-j})\beta_{-j}$, $\hat{\beta}_j = \Xi_j^{-1}B_j(u_j)y_j^*$, $S_j^2 = y_j^{*T}y_j^* - \hat{\beta}_j^T \Xi_j \hat{\beta}_j$ and $k_{-j}, \xi_{-j}, \beta_{-j}$ are k, ξ, β with k_j, ξ_j, β_j excluded, respectively.

It follows from (7) that the approach of composition [55] can be used to generate σ^2 from a conditional inverse-gamma posterior

$$p(\sigma^2 | y, x, z, \rho, \alpha, k, \xi, \beta, \tau) \propto (S^2 + s_0^2)^{\frac{n+r_0}{2}} (\sigma^2)^{-\frac{n+r_0}{2} - 1} \exp \left\{ -\frac{S^2 + s_0^2}{2\sigma^2} \right\} \quad (10)$$

and to sample θ from a conditional normal posterior

$$p(\theta | y, x, z, \rho, k, \xi, \sigma^2, \tau) \propto (2\pi\sigma^2)^{-\frac{K+q}{2}} |\Xi|^{\frac{1}{2}} \exp \left\{ -\frac{1}{2\sigma^2} (\theta - \hat{\theta})^T \Xi (\theta - \hat{\theta}) \right\} I\{Q\beta = \mathbf{0}\}. \quad (11)$$

It follows from (11) that

$$p(\alpha|y, x, z, \rho, k, \zeta, \beta, \sigma^2, \tau) \propto (2\pi\sigma^2)^{-\frac{q}{2}} |\Xi_0|^{\frac{1}{2}} \exp\left\{-\frac{1}{2\sigma^2}(\alpha - \hat{\alpha})^T \Xi_0(\alpha - \hat{\alpha})\right\}, \tag{12}$$

where $\Xi_0 = \tau_0^{-1}I_q + x^T x$, $y_0^* = A(\rho)y - B^T(u)\beta$, $\hat{\alpha} = \Xi_0^{-1}x^T y_0^*$, and

$$p(\beta_j|y, x, z, \rho, \alpha, k, \zeta, \beta_{-j}, \sigma^2, \tau) \propto (2\pi\sigma^2)^{-\frac{k_j}{2}} |\Xi_j|^{\frac{1}{2}} \exp\left\{-\frac{1}{2\sigma^2}(\beta_j - \hat{\beta}_j)^T \Xi_j(\beta_j - \hat{\beta}_j)\right\} I\{Q_j\beta_j = 0\}, \tag{13}$$

for $j = 1, \dots, p$. To achieve identification, we focus on the constraint $Q_j\beta_j = 0$, which should be imposed on β_j . According to Panagiotelis and Smith [56], drawing β_j from (13) is equivalent to drawing β_j^* from a normal distribution with mean vector $\hat{\beta}_j$ and covariance matrix Ξ_j , then β_j^* is transformed to β_j by

$$\beta_j = \beta_j^* - \Xi_j^* Q_j^T (Q_j \Xi_j^* Q_j^T)^{-1} Q_j \beta_j^*. \tag{14}$$

As it is convenient to sample $(\sigma^2, \alpha, \beta)$ from the conditional posterior (10), (12) and (13), we concentrate on sampling from (9). A sampling method is applied, in which the original Bayesian free-knots spline [43–51] is used as a reversible-jump sampler [57]. It includes three types of movement: the deletion, the addition and the movement of only one knot [48]. We keep the first two move-types unchanged but improve the movement step through the hit-and-run algorithm [58] so that all the knots can be repositioned in each iteration instead of only one knot: for $j = 1, \dots, p$ select a k_j dimension direction vector $c_j = (c_{j1}, \dots, c_{jk})^T$ randomly, and define

$$\Omega_j = \left\{ \omega_j : \zeta_j^* = \zeta_j + \omega c_j \quad \text{with } a_j < \zeta_{ji}^* < b_j, i = 1, \dots, k_j \right\} = (\omega_{j1}, \omega_{j2});$$

generate ω_j from a Cauchy distribution with location 0 and scale σ_{ζ_j} truncated on $(\omega_{j1}, \omega_{j2})$, where σ_{ζ_j} acts as a tuning parameter; assign $\zeta_j^* = \zeta_j + \omega c_j$ and reorder all of the knots. The proposed number and positions of knots are finally accepted with probability

$$\min \left\{ 1, A_j \left(\frac{|\Xi_j|}{|\Xi_j^*|} \right)^{\frac{1}{2}} \times \left(\frac{S_j^2 + s_0^2}{S_j^{*2} + s_0^2} \right)^{\frac{n+r_0}{2}} \right\},$$

where Ξ_j^* and S_j^{*2} correspond to Ξ_j and S_j^2 in the candidate posterior, respectively, and the factor

$$A_j = \frac{\arctan(\omega_{j2}/\sigma_{\zeta_j}) - \arctan(\omega_{j1}/\sigma_{\zeta_j})}{\arctan[(\omega_{j2} - \omega_j)/\sigma_{\zeta_j}] - \arctan[(\omega_{j1} - \omega_j)/\sigma_{\zeta_j}]}$$

It is evident that the posterior of hyperparameter τ_j is a conditional inverse-gamma distribution

$$p(\tau_0|\sigma^2, \alpha) \propto \tau^{-\frac{q+r\tau_{\alpha 0}}{2}-1} \exp\left\{-\frac{s_{\tau_{\alpha 0}}^2 + \alpha^T \alpha / \sigma^2}{2\tau_0}\right\} \tag{15}$$

and

$$p(\tau_j|\beta_j, k_j, \zeta_j, \sigma^2) \propto \tau_j^{-\frac{k_j+r\tau_{\beta_j 0}}{2}-1} \exp\left\{-\frac{s_{\tau_{\beta_j 0}}^2 + \beta_j^T \beta_j / \sigma^2}{2\tau_j}\right\}, \quad j = 1, \dots, p, \tag{16}$$

which can be simulated directly from (15) and (16).

3.3. Sampling Scheme

The Bayesian estimate of $\Theta = \{\rho, \alpha, k, \zeta, \beta, \sigma^2, \tau\}$ is obtained by observations generated from the posterior of all unknown quantities by running the Gibbs sampler. Moreover,

simulating β_j from (13) is challenging and nonstandard, and the parameter space on the constraint $Q_j\beta_j = 0$ for $j = 1, \dots, p$. According to Panagiotelis and Smith [56], it is equivalent that β_j^* is transformed to β_j by (14). The MCMC sampling algorithm (Algorithm 1) is described in the following manner.

Algorithm 1 The MCMC sampling algorithm.

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

Initialization: Initialize $\Theta^{(0)} = \{\rho^{(0)}, \alpha^{(0)}, k^{(0)}, \zeta^{(0)}, \beta^{(0)}, \sigma^{2(0)}, \tau^{(0)}\}$ in the MCMC algorithm, where the unknown parameters are generated from the priors, respectively.

MCMC iterations: Given the current state of $\Theta^{(t-1)}$ successively, draw $\Theta^{(t)}$ from $p(\Theta|x, y, z)$, for $t = 1, 2, 3, \dots$. The detailed MCMC sampling cycles are outlined in the following manner.

(a) Generate $\rho^{(t)}$ from $p(\rho|x, y, z, \theta^{(t-1)}, \alpha^{(t-1)}, k^{(t-1)}, \zeta^{(t-1)}, \sigma^{2(t-1)}, \tau^{(t-1)})$;

(b) Generate $\sigma^{2(t)}$ from $p(\sigma^2|x, y, z, \alpha^{(t)}, k^{(t)}, \zeta^{(t)}, \rho^{(t)}, \tau^{(t-1)})$;

(c) Generate $\alpha^{(t)}$ from $p(\alpha|x, y, z, \rho^{(t)}, k^{(t-1)}, \zeta^{(t-1)}, \beta^{(t-1)}, \tau^{(t-1)})$;

(d) Generate $(k_j^{(t)}, \zeta_j^{(t)})$ from $p(k_j^{(t)}, \zeta_j^{(t)}|x, y, z, \rho^{(t)}, k_{-j}^{(t-1)}, \zeta_{-j}^{(t-1)}, \beta_{-j}^{(t-1)}, \tau^{(t-1)})$ for $j = 1, \dots, p$;

(e) Generate $\beta_j^{(t)}$ from $p(\beta_j^{(t)}|x, y, z, \rho^{(t)}, \alpha^{(t)}, k^{(t)}, \zeta^{(t)}, \beta_{-j}^{(t-1)}, \sigma^{2(t)}, \tau^{(t-1)})$, and adjust $\beta_j^{(t)}$ according to (14) for $j = 1, \dots, p$;

(f) Generate $\tau_j^{(t)}$ from $p(\tau_j|k^{(t)}, \zeta^{(t)}, \beta_j^{(t)}, \sigma^{2(t)})$ for $j = 1, \dots, p$;

(g) Generate $\tau_0^{(t)}$ from $p(\tau_0|\alpha^{(t)}, \sigma^{2(t)})$.

Output: The MCMC sampling from the conditional posteriors of $\{\Theta^{(t)}\}_{t=1,2,3,\dots}$.

4. Empirical Illustrations

We demonstrate the performance of the PLSISAR model and methodology by a simulation and use them to analyze a real data. We set the Rook weight matrix [2] and the Case weight matrix [3] to examine the influence of the spatial weight matrix W . The Rook weight matrix is sampled from Rook contiguity in [59] by randomly allocating the n spatial units on a lattice of $m \times m$ ($\geq n$) squares, finding the neighbors for the unit, and then row normalizing. Meanwhile, we generated the Case weight matrix from the spatial scenario $W = I_r \otimes T_m$ in [3] with m members in a district and r districts, and each neighbor of a member in each district given equal weight [10], where \otimes is the Kronecker product, $T_m = (1/(m - 1))(\mathbf{1}_m \mathbf{1}_m^T - I_m)$ and $\mathbf{1}_m = (1, \dots, 1)^T$ is an m -dimensional vector.

4.1. Simulation

Consider the following PLSISAR models:

$$y_i = \rho \sum_{l=1}^n w_{il} y_l + x_i^T \alpha + g_1(z_{i1}) + g_2(z_{i2}) + \varepsilon_i, \quad i = 1, \dots, n,$$

where $x_i = (x_{i1}, x_{i2})^T$ follows a bivariate standard normal distribution, $z_i = (z_{i1}, z_{i2})^T$ is a bivariate vector, where z_{i1} and z_{i2} are mutually independent and follow uniform distributions on $(-1, 1)$ and $(0, 1)$, respectively. The nonparametric functions $g_1(z_1) = \sin(\pi z_1)$ and $g_2(z_2) = 4z_2(1 - z_2^2) - 1$, $\varepsilon_i \sim N(0, \sigma^2)$, the parameters are assumed as $\alpha = (1, -1)^T$ and two cases of variance $\sigma^2 = \{0.25, 0.75\}$, respectively. We consider three different cases of spatial parameters $\rho = \{0.2, 0.5, 0.7\}$, which represent the spatial dependence of the response from weak to strong. The sample size of the Case weight matrix and the Rook weight matrix is $(r, m) = \{(20, 5), (80, 5)\}$ and $n = \{100, 400\}$, respectively.

In our computation, we run each simulation with 1000 replications, adopt a quadratic B-spline and set hyper-parameters $(r_0, s_0^2, r_{\tau_{\alpha 0}}, s_{\tau_{\alpha 0}}^2) = (1, 1, 1, 0.005)$ and $(r_{\tau_{\beta_j 0}}, s_{\tau_{\beta_j 0}}^2) =$

$(1, 0.005)$ for $j = 1, \dots, p$. The initial state of the Markov chain of all unknown parameters is selected as follows. All unknown parameters are sampled from the respective priors by gradually decreasing or increasing the use of tuning parameters σ_ρ and σ_{ξ_j} for $j = 1, \dots, p$ so that the acceptable rates are about 25%. For each replication, we generate 6000 sampled values and then delete the first 2000 sampled values as a burn-in period by running our MCMC sampling. Based on the last 4000 sampled values, we compute the corresponding average of 1000 replications as the posterior mean (mean), the 95% posterior credible intervals (95% CI), and standard error (SE). In addition, the standard derivations (SD) of the estimated posterior mean are calculate to compare them with the mean of the estimated posterior SE.

We evaluate the performance of nonparametric estimators by the integrated squared bias (Bias), the root integrated mean squared errors (SSE), the mean absolute deviation errors (MADE)

$$\text{Bias}(\hat{g}_j) = \int [E\hat{g}_j(z) - g_j(z)]^2 dz, \quad \text{SSE}(\hat{g}_j) = \left\{ \int E[\hat{g}_j(z) - g_j(z)]^2 dz \right\}^{\frac{1}{2}},$$

$$\text{MADE}_j = \frac{1}{200} \sum_{i=1}^{200} |\hat{g}_j(z_{ji}) - g_j(z_{ji})| \quad \text{and} \quad \text{MADE} = \frac{1}{p} \sum_{j=1}^p \text{MADE}_j$$

for $j = 1, \dots, p$, where mathematical expectations are estimated by their corresponding empirical version, and the integrations are performed applying a Riemannian sum approximation at 200 fixed grid points $\{z_{ji}\}_{i=1}^{200}$ that are equally-spaced chosen from $[a_j, b_j]$. From the model (1), the marginal effects are given by $\frac{\partial y}{\partial x_j} = (I_n - \rho W)^{-1} I_n \alpha_j$ for $j = 1, \dots, q$. According to LeSage and Pace [52] suggestions, the mean of either the row sums or the column sums of the non-diagonal elements is used as the indirect effects, the mean of the diagonal elements is used as the direct effects, and the sum of the indirect and direct effects is taken as the total effects.

To check the convergence of our algorithm, we run five Markov chains corresponding to different starting values through the MCMC sampling algorithm to perform each replication. The sampled traces of some parameters and nonparametric functions on grid points are displayed in Figure 1. It is obvious that the five parallel sequences mix quite well. We compute the ‘‘potential scale reduction factor’’ $\sqrt{\hat{R}}$ [60] for all unknown parameters and nonparametric functions at 20 selected grid points. Figure 2 shows all the values of $\sqrt{\hat{R}}$ against the iteration numbers. According to the suggestion of Gelman and Rubin [60], it is easy to see that 2000 burn-in iterations are enough to make the MCMC algorithm converge as all the values of $\sqrt{\hat{R}}$ were less than 1.2.

The boxplots of the Bias values are displayed in Figure 3. Under the Rook weight matrix, the medians of which are $\text{Bias}_1 = 0.0147$ and $\text{Bias}_2 = 0.0104$ for $n = 100$, and $\text{Bias}_1 = 0.0039$ and $\text{Bias}_2 = 0.0030$ for $n = 400$, respectively. Under the Case weight matrix, the medians of which are $\text{Bias}_1 = 0.0148$ and $\text{Bias}_2 = 0.0104$ for $(r, m) = (20, 5)$, and $\text{Bias}_1 = 0.0038$ and $\text{Bias}_2 = 0.0030$ for $(r, m) = (80, 5)$, respectively. Figure 3 show the boxplots of the SSE values. Under the Rook weight matrix, the medians are $\text{SSE}_1 = 0.2490$ and $\text{SSE}_2 = 0.2271$ for $n = 100$, and $\text{SSE}_1 = 0.1253$ and $\text{SSE}_2 = 0.1151$ for $n = 400$, respectively. Under the Case weight matrix, the medians are $\text{SSE}_1 = 0.2495$ and $\text{SSE}_2 = 0.2267$ for $(r, m) = (20, 5)$, and $\text{SSE}_1 = 0.1257$ and $\text{SSE}_2 = 0.1154$ for $(r, m) = (80, 5)$, respectively. The results show that the Bias values and the SSE values of the nonparametric functions decrease with the increase in the sample size, indicating that the nonparametric estimation is convergent. It is evident that the weight matrix of Case and Rook can obtain a reasonable estimation effect.

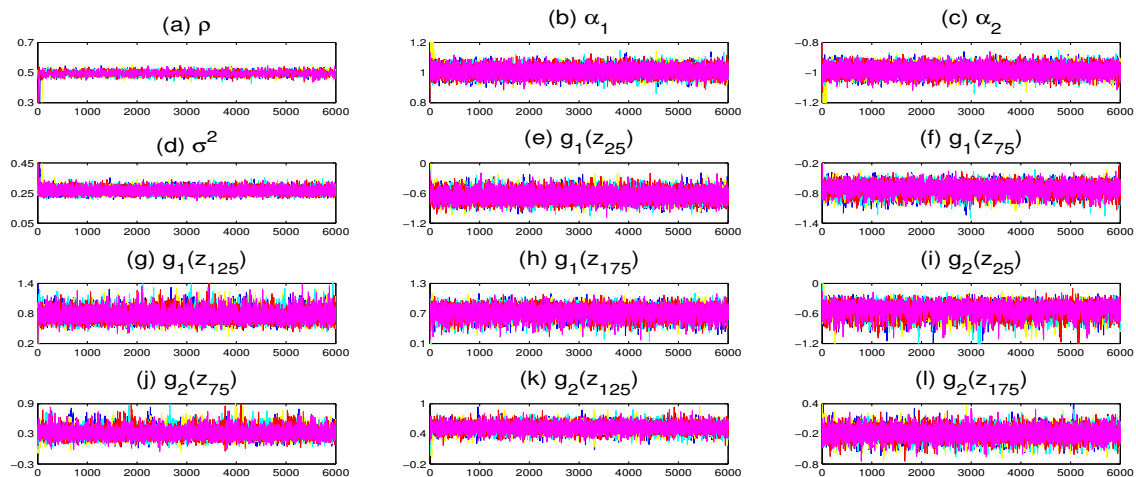


Figure 1. Trace plots of five parallel Markov chains with different starting values for some parameters and nonparametric functions (only a replication with $(r, m) = (80, 5)$ and $(\rho, \sigma^2) = (0.5, 0.25)$ is displayed).

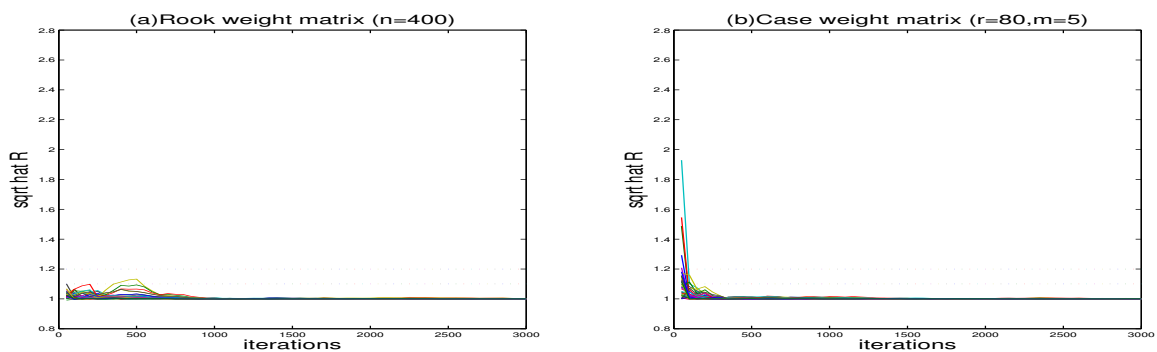


Figure 2. The “potential scale reduction factor” $\sqrt{\hat{R}}$ for simulation results (the case of $(\rho, \sigma^2) = (0.5, 0.25)$).

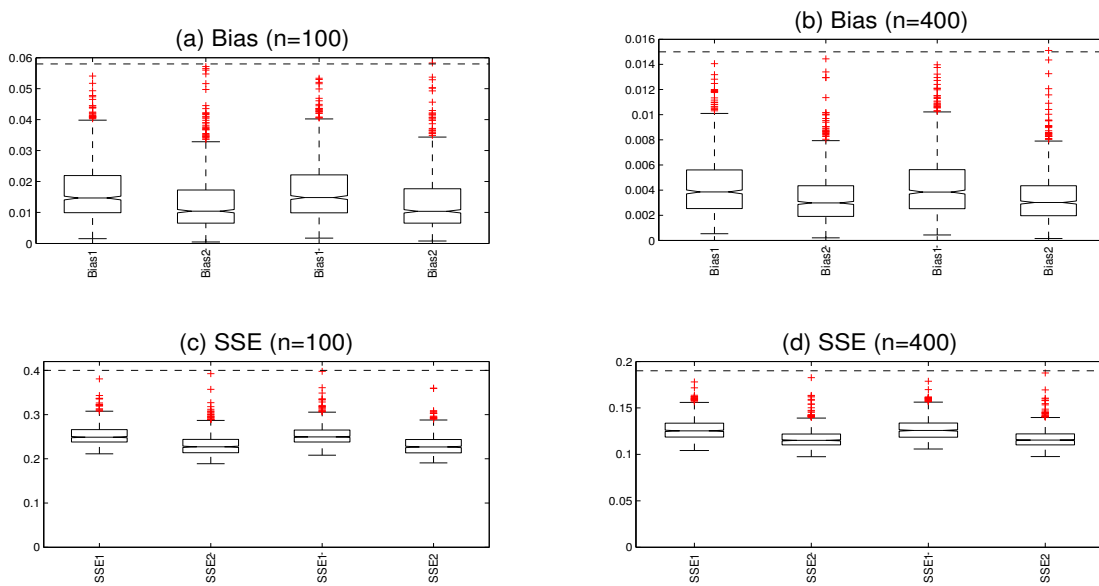


Figure 3. The boxplots (a,b) display the integrated squared bias, the boxplots (c,d) display the root integrated mean squared errors. (The two panels on the left are based on the Rook weight matrix, and the two panels on the right are based on the Case weight matrix with $(\rho, \sigma^2) = (0.5, 0.25)$).

The estimation results are reported in Table 1. We observe that the mean values of all estimators are very close to the corresponding true values, and the mean value of SE is close to the respective SD. The results show that the parameter estimation and SE are precise. Meanwhile, the larger the sample sizes under the same weight matrix, the more precise the estimates are. The above experiences corresponding to different starting values have been repeated, and the results are similar. It implies that the MCMC sampling works well. Moreover, we find that the estimation effect of ρ with the Case weight matrix is slightly better than that with Rook weight matrix under the same sample sizes. The possible main reason is that the performance of the Case weight matrix is superior to the Rook weight matrix under different variances σ^2 . In addition, the general pattern of estimates reported in Table 1 is that all estimators impose a relatively bigger bias on the total effect when the same sample sizes have a strong positive spatial dependence. Figure 4 depicts the fitted functions, together with its 95% CI from a typical sample with $\rho = 0.5$ and $\sigma^2 = 0.25$. The typical sample is selected in such a way that the SSE values are equal to the median in the 1000 replications. It is obvious that the fitted nonparametric functions are improving with increasing the sample size.

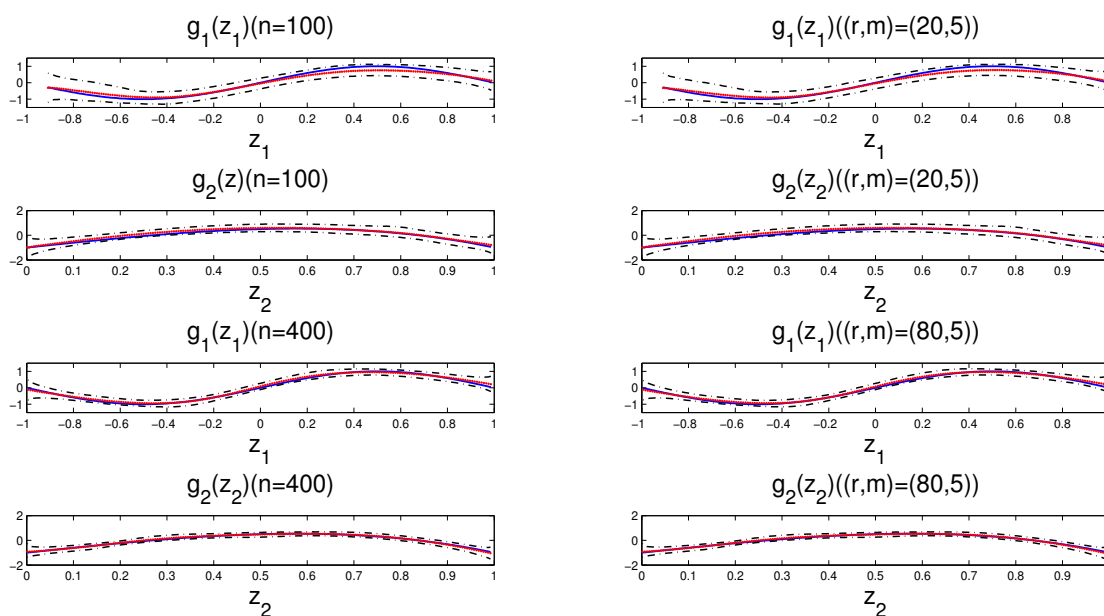


Figure 4. The true functions (solid lines), the fitted functions (dotted lines) and their 95% CI (dot-dashed lines) for a typical sample (the **left** panel based on the Rook weight matrix and the **right** panel based on the Case weight matrix with $(\rho, \sigma^2) = (0.5, 0.25)$).

Table 1. Simulation results of parametric estimation.

Parameter	<i>n</i>	Rook Weight Matrix				<i>(r, m)</i>	Case Weight Matrix				
		Mean	SE	SD	95% CI		Mean	SE	SD	95% CI	
$\rho = 0.2000$	100	0.2018	0.0552	0.0492	(0.0929, 0.3099)	(20, 5)	0.1987	0.0512	0.0473	(0.0973, 0.2990)	
$\alpha_1 = 1.0000$		0.9930	0.0649	0.0623	(0.8655, 1.1203)		0.9930	0.0648	0.0620	(0.8656, 1.1202)	
$\alpha_2 = -1.0000$		-0.9956	0.0650	0.0633	(-1.1231, -0.8680)		-0.9955	0.0650	0.0632	(-1.1202, -0.8679)	
$\sigma^2 = 0.2500$		0.2716	0.0447	0.0381	(0.1980, 0.3723)		0.2714	0.0447	0.0382	(0.1979, 0.3720)	
Total effect											
$x_1 = 1.2500$		1.2544	0.1188	0.1064	(1.0378, 1.5042)		1.2487	0.1134	0.1083	(1.0340, 1.4853)	
$x_2 = -1.2500$	-1.2581	0.1189	0.1115	(-1.5083, -1.0414)	-1.2516	0.1136	0.1057	(-1.4886, -1.0424)			
$\rho = 0.5000$	100	0.5000	0.0462	0.0414	(0.4081, 0.5896)	(20, 5)	0.4986	0.0344	0.0318	(0.4304, 0.5655)	
$\alpha_1 = 1.0000$		0.9930	0.0652	0.0627	(0.8650, 1.1210)		0.9931	0.0652	0.0620	(0.8651, 1.1211)	
$\alpha_2 = -1.0000$		-0.9957	0.0653	0.0633	(-1.1239, -0.8674)		-0.9957	0.0653	0.0635	(-1.1239, -0.8676)	
$\sigma^2 = 0.2500$		0.2719	0.0450	0.0383	(0.1980, 0.3731)		0.2717	0.0453	0.0384	(0.1978, 0.3728)	
Total effect											
$x_1 = 2.0000$		2.0153	0.2283	0.1971	(1.6229, 2.4969)		1.9972	0.1818	0.1734	(1.6633, 2.3758)	
$x_2 = -2.0000$	-2.0215	0.2311	0.2053	(-2.5037, -1.6281)	-2.0018	0.1821	0.1687	(-2.3817, -1.6673)			
$\rho = 0.7000$	100	0.6994	0.0348	0.0319	(0.6300, 0.7667)	(20, 5)	0.6988	0.0217	0.0200	(0.6557, 0.7409)	
$\alpha_1 = 1.0000$		0.9932	0.0655	0.0629	(0.8648, 1.1219)		0.9934	0.0656	0.0621	(0.8648, 1.1222)	
$\alpha_2 = -1.0000$		-0.9959	0.0655	0.0635	(-1.1245, -0.8673)		-0.9934	0.0656	0.0639	(-1.1249, -0.8671)	
$\sigma^2 = 0.2500$		0.2721	0.0451	0.0385	(0.1981, 0.3738)		0.2718	0.0451	0.0386	(0.1977, 0.3736)	
Total effect											
$x_1 = 3.3333$		3.3823	0.4428	0.3911	(2.6360, 4.3697)		3.3273	0.3034	0.2895	(2.7702, 3.9615)	
$x_2 = -3.3333$	-3.3924	0.4434	0.4034	(-4.3811, -2.6451)	-3.3346	0.3037	0.2810	(-3.9697, -2.7772)			
$\rho = 0.2000$	100	0.1982	0.0766	0.0750	(0.0469, 0.3471)	(20, 5)	0.1914	0.0713	0.0738	(0.0506, 0.3293)	
$\alpha_1 = 1.0000$		0.9839	0.1108	0.1075	(0.7663, 1.2011)		0.9840	0.1107	0.1071	(0.7665, 1.2013)	
$\alpha_2 = -1.0000$		-0.9882	0.1108	0.1095	(-1.2059, -0.7708)		-0.9878	0.1108	0.1093	(-1.2050, -0.7701)	
$\sigma^2 = 0.7500$		0.8234	0.1385	0.1386	(0.5957, 1.1359)		0.8230	0.1393	0.1391	(0.5955, 1.1353)	
Total effect											
$x_1 = 1.2500$		1.2487	0.1858	0.1743	(0.9164, 1.6442)		1.2369	0.1767	0.1771	(0.9169, 1.6083)	
$x_2 = -1.2500$	-1.2552	0.1865	0.1841	(-1.6525, -0.9221)	-1.2408	0.1771	0.1734	(-1.6130, -0.9189)			

Table 1. Cont.

Parameter	<i>n</i>	Rook Weight Matrix				<i>(r, m)</i>	Case Weight Matrix			
		Mean	SE	SD	95% CI		Mean	SE	SD	95% CI
$\rho = 0.5000$		0.4947	0.0637	0.0630	(0.3676, 0.6170)		0.4934	0.0480	0.0498	(0.3975, 0.5853)
$\alpha_1 = 1.0000$		0.9847	0.1112	0.1082	(0.7667, 1.2031)		0.9849	0.1112	0.1069	(0.7667, 1.2033)
$\alpha_2 = -1.0000$		-0.9890	0.1112	0.1095	(-1.2073, -0.7707)		-0.9887	0.1113	0.1097	(-1.2070, -0.7703)
$\sigma^2 = 0.7500$		0.8250	0.1398	0.1402	(0.5955, 1.1404)		0.8248	0.1404	0.1407	(0.5951, 1.1405)
Total effect										
$x_1 = 2.0000$		2.0079	0.3449	0.3154	(1.4275, 2.7545)		1.9782	0.2826	0.2831	(1.4639, 2.5733)
$x_2 = -2.0000$		-2.0187	0.3449	0.3321	(-2.7676, -1.4365)		-1.9844	0.2830	0.2778	(-2.5805, -1.4696)
$\rho = 0.7000$		0.6942	0.0479	0.0477	(0.5984, 0.7853)		0.6955	0.0301	0.0314	(0.6351, 0.7528)
$\alpha_1 = 1.0000$		0.9854	0.1115	0.1086	(0.7670, 1.2044)		0.9857	0.1116	0.1074	(0.7673, 1.2052)
$\alpha_2 = -1.0000$		-0.9896	0.1116	0.1099	(-1.2090, -0.7708)		-0.9857	0.1117	0.1104	(-1.2087, -0.7705)
$\sigma^2 = 0.7500$		0.8268	0.1417	0.1417	(0.5959, 1.1454)		0.8263	0.1420	0.1418	(0.5949, 1.1455)
Total effect										
$x_1 = 3.3333$		3.3768	0.6705	0.6079	(2.3087, 4.8873)		3.2961	0.4704	0.4722	(2.4414, 4.2873)
$x_2 = -3.3333$		-3.3950	0.6743	0.6351	(-4.9107, -2.3229)		-3.3061	0.4716	0.4637	(-4.2986, -2.4486)
$\rho = 0.2000$	400	0.1996	0.0245	0.0235	(0.1514, 0.2475)	(80, 5)	0.2006	0.0218	0.0208	(0.1577, 0.2431)
$\alpha_1 = 1.0000$		1.0005	0.0297	0.0286	(0.9422, 1.0587)		1.0005	0.0297	0.0287	(0.9423, 1.0587)
$\alpha_2 = -1.0000$		-0.9972	0.0297	0.0290	(-1.0554, -0.9389)		-0.9972	0.0297	0.0291	(-1.0554, -0.9390)
$\sigma^2 = 0.2500$		0.2549	0.0186	0.0185	(0.2210, 0.2939)		0.2549	0.0186	0.0185	(0.2210, 0.2939)
Total effect										
$x_1 = 1.2500$		1.2521	0.0524	0.0503	(1.1525, 1.3578)		1.2532	0.0496	0.0477	(1.1584, 1.3527)
$x_2 = -1.2500$		-1.2480	0.0524	0.0514	(-1.3539, -1.1484)		-1.2490	0.0494	0.0480	(-1.3484, -1.1546)
$\rho = 0.5000$		0.4995	0.0207	0.0198	(0.4587, 0.5399)		0.5003	0.0148	0.0140	(0.4712, 0.5290)
$\alpha_1 = 1.0000$		1.0006	0.0299	0.0287	(0.9420, 1.0590)		1.0005	0.0299	0.0288	(0.9419, 1.0590)
$\alpha_2 = -1.0000$		-0.9972	0.0299	0.0287	(-1.0558, -0.9387)		-0.9972	0.0299	0.0292	(-1.0557, -0.9387)
$\sigma^2 = 0.2500$		0.2550	0.0187	0.0186	(0.2210, 0.2941)		0.2549	0.0187	0.0186	(0.2209, 0.2941)
Total effect										
$x_1 = 2.0000$		2.0051	0.0974	0.0932	(1.8227, 2.2045)		2.0049	0.0797	0.0765	(1.8526, 2.1651)
$x_2 = -2.0000$		-1.9984	0.0973	0.0949	(-2.1977, -1.8162)		-1.9983	0.0795	0.0772	(-2.1580, -1.8465)

Table 1. Cont.

Parameter	<i>n</i>	Rook Weight Matrix				<i>(r, m)</i>	Case Weight Matrix			
		Mean	SE	SD	95% CI		Mean	SE	SD	95% CI
$\rho = 0.7000$		0.6996	0.0158	0.0151	(0.6684, 0.7304)		0.7001	0.0093	0.0087	(0.6817, 0.7183)
$\alpha_1 = 1.0000$		1.0006	0.0300	0.0288	(0.9419, 1.0594)		1.0004	0.0300	0.0289	(0.9416, 1.0593)
$\alpha_2 = -1.0000$		-0.9972	0.0300	0.0291	(-1.0560, -0.9385)		-0.9971	0.0300	0.0293	(-1.0560, -0.9384)
$\sigma^2 = 0.2500$		0.2550	0.0187	0.0183	(0.2210, 0.2942)		0.2549	0.0188	0.0187	(0.2208, 0.2942)
Total effect										
$x_1 = 3.3333$		3.3475	0.1920	0.1833	(2.9942, 3.7466)		3.3415	0.1335	0.1272	(3.0867, 3.6098)
$x_2 = -3.3333$		-3.3364	0.1917	0.1862	(-3.7347, -2.9837)		-3.3304	0.1332	0.1283	(-3.5980, -3.0764)
$\rho = 0.2000$	400	0.1986	0.0366	0.0365	(0.1266, 0.2701)	(80, 5)	0.1994	0.0326	0.0323	(0.1354, 0.2630)
$\alpha_1 = 1.0000$		0.9999	0.0512	0.0493	(0.8996, 1.1001)		0.9999	0.0512	0.0495	(0.8997, 1.1001)
$\alpha_2 = -1.0000$		-0.9942	0.0512	0.0501	(-1.0945, -0.8940)		-0.9943	0.0512	0.0501	(-1.0945, -0.8942)
$\sigma^2 = 0.7500$		0.7597	0.0555	0.0558	(0.6587, 0.8760)		0.7596	0.0555	0.0560	(0.6586, 0.8757)
Total effect										
$x_1 = 1.2500$		1.2527	0.0847	0.0827	(1.0936, 1.4257)		1.2530	0.0807	0.0786	(1.1000, 1.4165)
$x_2 = -1.2500$		-1.2457	0.0847	0.0843	(-1.4187, -1.0866)		-1.2459	0.0805	0.0800	(-1.4090, -1.0936)
$\rho = 0.5000$		0.4982	0.0306	0.0306	(0.4376, 0.5578)		0.4995	0.0219	0.0216	(0.4563, 0.5420)
$\alpha_1 = 1.0000$		1.0001	0.0514	0.0494	(0.8994, 1.1007)		1.0000	0.0514	0.0497	(0.8994, 1.1007)
$\alpha_2 = -1.0000$		-0.9944	0.0513	0.0501	(-1.0950, -0.8939)		-0.9943	0.0514	0.0502	(-1.0951, -0.8938)
$\sigma^2 = 0.7500$		0.7600	0.0559	0.0564	(0.6585, 0.8770)		0.7598	0.0559	0.0567	(0.6582, 0.8769)
Total effect										
$x_1 = 2.0000$		2.0068	0.1541	0.1509	(1.7231, 2.3272)		2.0048	0.1293	0.1256	(1.7601, 2.2668)
$x_2 = -2.0000$		-1.9956	0.1538	0.1538	(-2.3151, -1.7121)		-1.9935	0.1289	0.1280	(-2.2544, -1.7496)
$\rho = 0.7000$		0.6984	0.0232	0.0234	(0.6524, 0.7435)		0.6996	0.0137	0.0136	(0.6725, 0.7262)
$\alpha_1 = 1.0000$		1.0002	0.0515	0.0496	(0.8993, 1.1012)		1.0001	0.0516	0.0499	(0.8991, 1.1013)
$\alpha_2 = -1.0000$		-0.9945	0.0515	0.0502	(-1.0956, -0.8937)		-0.9944	0.0516	0.0504	(-1.0955, -0.8936)
$\sigma^2 = 0.7500$		0.7603	0.0563	0.0569	(0.6581, 0.8782)		0.7600	0.0564	0.0573	(0.6575, 0.8781)
Total effect										
$x_1 = 3.3333$		3.3537	0.2978	0.2935	(2.8195, 3.9858)		3.3415	0.2155	0.2091	(2.9334, 3.7779)
$x_2 = -3.3333$		-3.3351	0.2970	0.2967	(-3.9661, -2.8023)		-3.3226	0.2150	0.2134	(-3.7580, -2.9153)

For comparison purposes, we use the Bayesian P-splines approach to approximate the nonparametric functions [61], where we assign a second-order random walk prior to the spline coefficients. The boxplots of MADE values with the Case weight matrix in Figure 5. In our method, the medians of MADE are $MADE_1 = 0.0997$, $MADE_2 = 0.0804$ and $MADE = 0.0916$ for $(r, m) = (20, 5)$, and the medians of MADE are $MADE_1 = 0.0504$, $MADE_2 = 0.0433$ and $MADE = 0.0476$ for $(r, m) = (80, 5)$, respectively, which are slightly smaller than the Bayesian P-splines approach. The results show that the Bayesian free knots splines approach is superior to the Bayesian P-splines approach in terms of fitting unknown nonparametric functions and computing time. Furthermore, we also compare the performance between the generalized method of moment estimator (GMME) in Du et al. [29] and the Bayesian MCMC estimator (BE) in our method. In order to evaluate the estimation effect of the nonparametric functions, we calculate the integrated squared bias (Bias) and the root integrated mean squared errors (SSE). Table 2 reports the results of the nonparameter estimation for GMME and BE (only a replication with $(\rho, \sigma^2) = (0.5, 0.25)$ is displayed). It is evident that the estimates are improving with increasing the sample size, the Bias of the BE estimates are slightly smaller than the Bias of the GMME, and the SSE of the BE estimates are very smaller than Bias of the GMME under the same sample size, showing that BE is better than GMME, although the latter can also obtain a reasonable estimation.

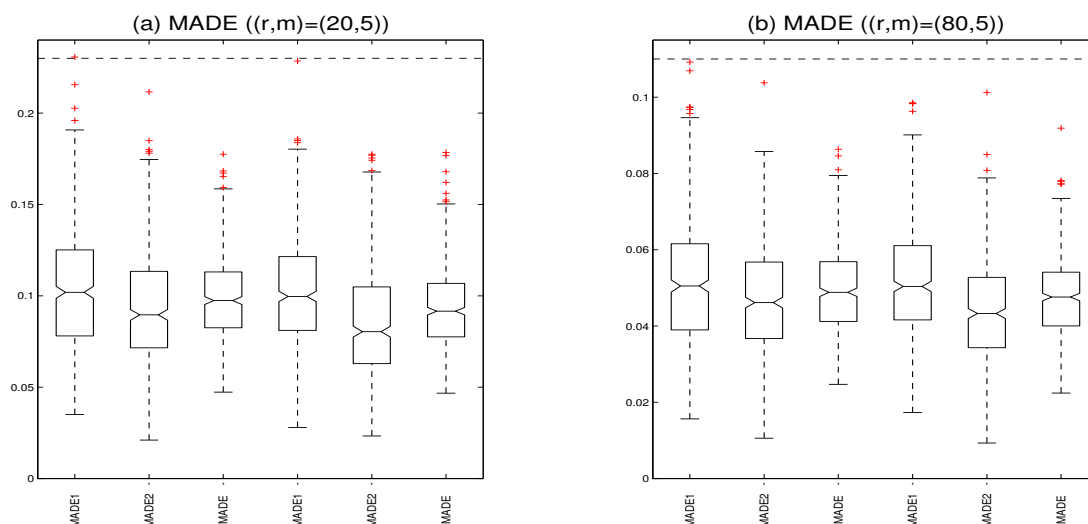


Figure 5. The boxplots (a) display the mean absolute deviation errors with the Case weight matrix $(r, m) = (20, 5)$ and (b) display the mean absolute deviation errors with the Case weight matrix $(r, m) = (80, 5)$ (the three panels on the left are based on Bayesian free knots splines and the three panels on the right are based on Bayesian P-splines).

Table 2. Simulation results of the nonparametric estimation.

Functions	(r, m)	With GMME		With BE	
		Bias	SSE	Bias	SSE
$g_1(\cdot)$	(60, 5)	0.014	7.947	0.006	0.146
$g_2(\cdot)$		0.009	7.246	0.007	0.141
$g_1(\cdot)$	(80, 5)	0.011	6.860	0.004	0.131
$g_2(\cdot)$		0.004	6.235	0.001	0.117

4.2. Application

We use the proposed model and estimation methods to analyze the well-known Sydney real estate data. A detailed description of the data set can be found in Harezlak et al. [62]. The data set contains a total of 37,676 properties sold in the Sydney Statistical Division in the calendar year of 2001, which is available from the HRW package in R.

We only focus on the last week of February to avoid the temporal issue, and there are 538 properties.

In this application, the house price (Price) is explained by four variables that include average weekly income (Income), levels of particulate matter with a diameter of less than a 10 micrometers level recorded at the air pollution monitoring station closest to the house (PM₁₀), lot size (LS), and distance from house to the nearest coastline location in kilometers (DC). On the one hand, Income and PM₁₀ have a linear effect on the response Price. On the other hand, LS and DC have a nonlinear effect on the response Price. Meanwhile, logarithmic transformation is performed on all variables to alleviate the trouble caused by large gaps in the domain. In addition, all variables are transformed such that the marginal distribution is approximately a standard normal distribution. This motivates us to consider the PLASAR model:

$$y_i = \rho \sum_{l=1}^n w_{il} y_l + x_i^T \alpha + g_1(z_{i1}) + g_2(z_{i2}) + \varepsilon_i, \quad i = 1, \dots, n, \tag{17}$$

where the response variable $y_i = \log(\text{Price}_i)$, $x_{i1} = \log(\text{Income}_i)$, $x_{i2} = \log(\text{PM}_{10i})$, $z_{i1} = \log(\text{LS}_i)$, $z_{i2} = \log(\text{DC}_i)$. Regarding the choice of the weight matrix, we use the Euclidean distance between any two houses to calculate the spatial weight matrix [63]. The spatial weight w_{il} is

$$w_{il} = \exp\{-\|s_i - s_l\|\} / \sum_{k \neq i} \exp\{-\|s_i - s_k\|\},$$

where $s_i = (\text{Lon}_i, \text{Lat}_i)$ is represented as the longitude and latitude of location. We apply a quadratic B-splines and assign hyperparameters $(\lambda, r_0, s_0^2, r_{\tau_{\alpha 0}}, s_{\tau_{\alpha 0}}^2, r_{\tau_{\beta 0}}, s_{\tau_{\beta 0}}^2) = (2, 1, 1, 1, 0.005, 1, 0.005)$ for $j = 1, \dots, p$ in our computation. We gradually decrease or increase the use of tuning parameters σ_ρ and σ_{ξ_j} such that the acceptable rates for updating ρ and (k_j, ξ_j) are around 25% for $j = 1, \dots, p$.

We generate 10,000 sampled values following a burn-in of 10,000 iterations and run the proposed Gibbs sampler five times with different initial states in each replication. Figure 6 plot the traces of some unknown parameters and nonparametric functions on grid points. It is obvious that the five parallel Markov chains mix well. We further calculate the ‘‘potential scale reduction factor’’ $\sqrt{\hat{R}}$ for each of the unknown parameters and nonparametric functions on 20 selected grid points, which are plotted in Figure 7. The result indicates that the Markov chains have converged within the first 10,000 burn-in iterations.

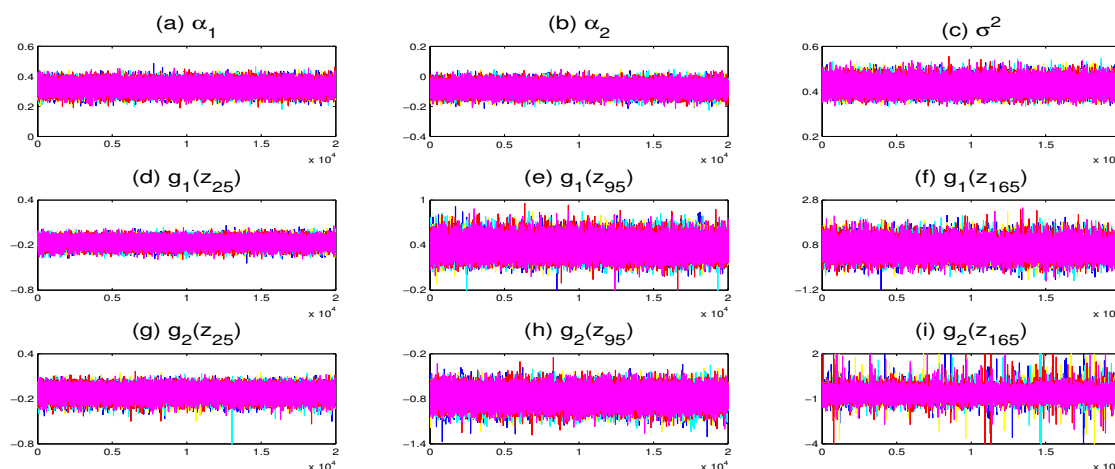


Figure 6. Trace plots of five parallel Markov chains with different starting values for some parameters and nonparametric functions.

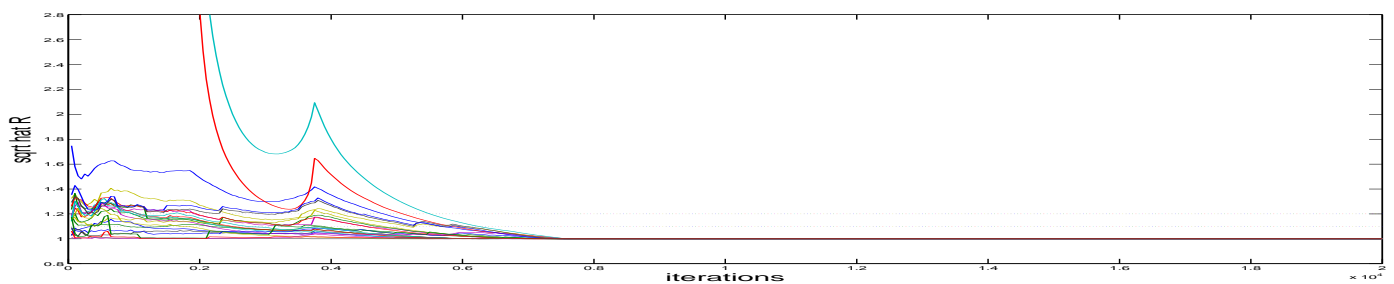


Figure 7. The “potential scale reduction factor” $\sqrt{\hat{R}}$ for Sydney real estate data.

Table 3 lists the estimated parameters together with their SE and 95% CI, which show that the estimation of $\hat{\rho}$ is 0.5548 with SE = 0.0307. It implies that there exists a significant and positive spatial relationship on the housing price. We find that two covaraites have significant effects on the housing price, and the effects of Income are positive, but PM₁₀ is negative. The regression coefficient of Income is $\hat{\alpha}_1 = 0.3269 > 0$, which indicates that the Income has a positive effect on the housing price. Moreover, the regression coefficient of PM₁₀ is $\hat{\alpha}_2 = -0.0810 < 0$, which reveals that the housing price would decrease as the PM₁₀ increases.

Table 3. Parametric estimation in the model (17) for Sydney real estate data.

Parameter	Mean	SE	95% CI
ρ	0.5548	0.0307	(0.4932, 0.6160)
α_1	0.3269	0.0326	(0.2630, 0.3908)
α_2	−0.0810	0.0318	(−0.1433, −0.0186)
σ^2	0.3269	0.0326	(0.2630, 0.3908)
Total effect			
x_1	0.7343	0.0875	(0.5573, 0.9073)
x_2	−0.1819	0.0866	(−0.3531, −0.0087)

Figure 8 depicts the fitted functions, together with its 95% CI, which look like two nonlinear functions. The curves show that $g_1(z_1)$ has a local maximum 0.6184 at around $z_1 = 3.9198$ and a local minimum 0.1557 at around $z_1 = -0.8237$, and $g_2(z_2)$ has a local minimum −0.8224 at around $z_2 = 2.1605$. The results provide evidence that the significant effects of LS and DC on the housing price have a nonlinear S-shape and U-shape, respectively.

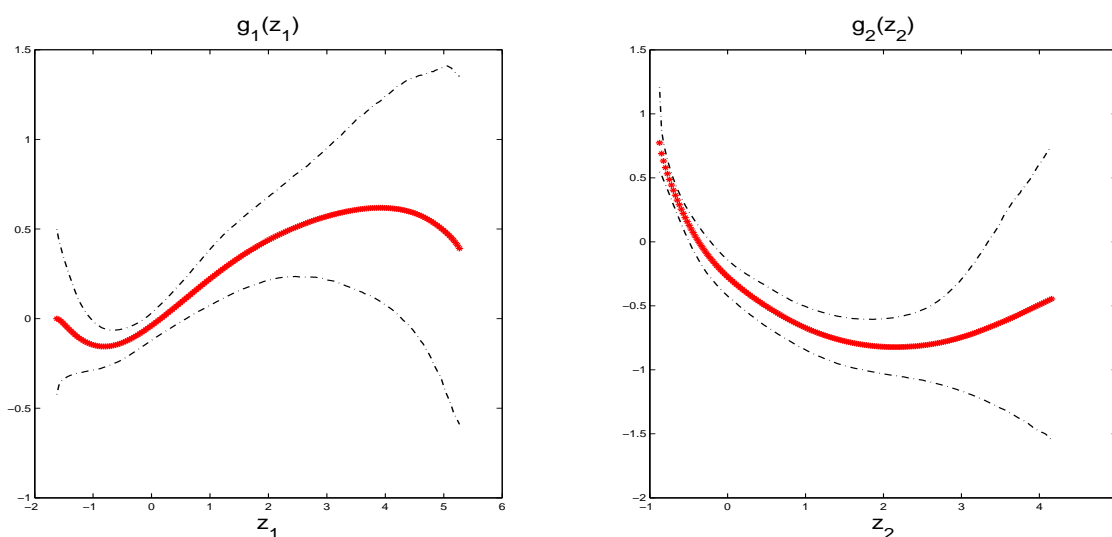


Figure 8. The fitted functions (dotted lines) and their 95% CI (dot-dashed lines) in the model (17) for Sydney real estate data.

5. Summary

Spatial data are frequently encountered in practical applications and can be analyzed through the SAR model. To avoid some serious shortcomings of fully nonparametric models and reduce the high risk of misspecification of the traditional SAR models, we have considered PLASAR models for spatial data, which combine the PLA model and SAR model. In addition to spatial dependence between neighbors, it captures the linear and nonlinear effects between the related variables. We specify the prior of all unknown parameters, which led to a proper posterior distribution. The posterior summaries are obtained via the MCMC technique, and we have considered a fully Bayesian approach with free-knot splines to analyze the PLASAR model and design a Gibbs sampler to explore the full conditional posterior distributions. To obtain a rapidly-convergent algorithm, a modified Bayesian free-knot splines approach incorporated with powerful MCMC techniques is employed. We have illustrated that the finite sample of the proposed model and estimation method perform satisfactorily through a simulation study. The results show that the Bayesian estimator is efficient relative to the GMME, although the latter can also obtain reasonable estimations. Finally, the proposed model and methodology are applied to analyze real data.

This article focuses only on symmetrical data and the homoscedasticity of independent errors. Since spatial data cannot easily meet the conditions, it is fairly straightforward to analyze the proposed model and methodology to deal with spatial errors and heteroscedasticity. While we use PLASAR models to assess the linear and nonlinear effects of the covariates on the spatial response, the other models, such as partially linear single-index SAR models and partially linear varying-coefficient SAR models, can also be considered. Moreover, it would be interesting to develop a model selection method in which covariates are linear or nonlinear. We leave these topics for future research.

Author Contributions: Supervision, Z.C. and J.C.; software, Z.C.; methodology, Z.C.; writing—original draft preparation, Z.C.; writing—review and editing, Z.C. and J.C. Both authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Natural Science Foundation of China (12001105), the Natural Science Foundation of Fujian Province (2020J01170), the Postdoctoral Science Foundation of China (2019M660156), the Program for Probability and Statistics: Theory and Application (No. IRTL1704), and the Program for Innovative Research Team in Science and Technology in Fujian Province University (IRTSTFJ).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are openly available in Reference [62].

Acknowledgments: The authors are most grateful to anonymous referees and the editors for their careful reading and insightful comments, who have helped to significantly improve this paper.

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

References

1. Cliff, A.D.; Ord, J.K. *Spatial Autocorrelation*; Pion Ltd.: London, UK, 1973.
2. Anselin, L. *Spatial Econometrics: Methods and Models*; Kluwer Academic Publisher: Dordrecht, The Netherlands, 1988.
3. Case, A.C. Spatial patterns in householed demand. *Econometrica* **1991**, *59*, 953–965. [[CrossRef](#)]
4. Cressie, N. *Statistics for Spatial Data*; John Wiley and Sons: New York, NY, USA, 1993.
5. LeSage, J.P. Bayesian estimation of spatial autoregressive models. *Int. Reg. Sci. Rev.* **1997**, *20*, 113–129. [[CrossRef](#)]
6. LeSage, J.P. Bayesian estimation of limited dependent variable spatial autoregressive models. *Geogr. Anal.* **2000**, *32*, 19–35. [[CrossRef](#)]
7. Anselin, L.; Bera, A.K. 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.
8. Ord, J. Estimation methods for models of spatial interaction. *J. Am. Stat. Assoc.* **1975**, *70*, 120–126. [[CrossRef](#)]

9. Kelejian, H.H.; Prucha, I.R. A generalized moments estimator for the autoregressive parameter in a spatial model. *Int. Econ. Rev.* **1999**, *40*, 509–533. [[CrossRef](#)]
10. Lee, L.F. Asymptotic distribution of quasi-maximum likelihood estimators for spatial autoregressive models. *Econometrica* **2004**, *72*, 1899–1925. [[CrossRef](#)]
11. Paelinck, J.H.P.; Klaassen, L.H.; Ancot, J.P.; Verster, A.C.P. *Spatial Econometrics*; Gower: Farnborough, UK, 1979.
12. Basile, R.; Gress, B. Semi-parametric spatial auto-covariance models of regional growth behaviour in Europe. *Rég. Dév.* **2004**, *21*, 93–118. [[CrossRef](#)]
13. Basile, R. Regional economic growth in Europe: A semiparametric spatial dependence approach. *Pape. Reg. Sci.* **2008**, *87*, 527–544. [[CrossRef](#)]
14. Su, L.J.; Jin, S.N. Profile quasi-maximum likelihood estimation of partially linear spatial autoregressive models. *J. Econom.* **2010**, *157*, 18–33. [[CrossRef](#)]
15. Baltagi, B.H.; Li, D. LM tests for functional form and spatial correlation. *Int. Reg. Sci. Rev.* **2001**, *24*, 194–225. [[CrossRef](#)]
16. Pace, P.K.; Barry, R.; Slawson, V.C., Jr.; Sirmans, C.F. Simultaneous spatial and functional form transformation. In *Advances in Spatial Econometrics*; Anselin, L., Florax, R., Rey, S.J., Eds.; Springer: Berlin, Germany, 2004; pp. 197–224.
17. van Gastel, R.A.J.J.; Paelinck, J.H.P. Computation of Box-cox transform parameters: A new method and its application to spatial econometrics. In *New Directions in Spatial Econometrics*; Anselin, L., Florax, R.J.G.M., Eds.; Springer: Berlin/Heidelberg, Germany, 1995; pp. 136–155.
18. Yang, Z.; Li, C.; Tse, Y.K. Functional form and spatial dependence in dynamic panels. *Econ. Lett.* **2006**, *91*, 138–145. [[CrossRef](#)]
19. Bellman, R.E. *Adaptive Control Processes*; Princeton University Press: Princeton, NJ, USA, 1961.
20. Friedman, J.H.; Stuetzle, W. Projection pursuit regression. *J. Am. Stat. Assoc.* **1981**, *76*, 817–823. [[CrossRef](#)]
21. Engle, R.F.; Granger, C.W.; Rice, J.; Weiss, A. Semiparametric Estimates of the Relation Between Weather and Electricity Sales. *J. Am. Stat. Assoc.* **1986**, *81*, 310–320. [[CrossRef](#)]
22. Hastie, T.J.; Tibshirani, R.J. *Generalized Additive Models*; Chapman and Hall: London, UK, 1990.
23. Hastie, T.J.; Tibshirani, R.J. Varying-coefficient models. *J. R. Stat. B* **1993**, *55*, 757–796. [[CrossRef](#)]
24. Su, L.J. Semiparametric GMM estimation of spatial autoregressive models. *J. Econom.* **2012**, *167*, 543–560. [[CrossRef](#)]
25. Chen, J.Q.; Wang, R.F.; Huang, Y.X. Semiparametric spatial autoregressive model: A two-step Bayesian approach. *Ann. Public Health Res.* **2015**, *2*, 1012.
26. Wei, H.J.; Sun, Y. Heteroskedasticity-robust semi-parametric GMM estimation of a spatial model with space-varying coefficients. *Spat. Econ. Anal.* **2016**, *12*, 113–128. [[CrossRef](#)]
27. Krisztin, T. The determinants of regional freight transport: A spatial, semiparametric approach. *Geogr. Anal.* **2017**, *49*, 268–308. [[CrossRef](#)]
28. Krisztin, T. Semi-parametric spatial autoregressive models in freight generation modeling. *Transp. Res. Part E Logist. Transp. Rev.* **2018**, *114*, 121–143. [[CrossRef](#)]
29. 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](#)]
30. Chen, J.B.; Cheng, S.L. GMM estimation of a partially linear additive spatial error model. *Mathematics* **2021**, *9*, 622. [[CrossRef](#)]
31. Liang, H.; Thurston, S.W.; Ruppert, D.; Apanasovich, T.; Hauser, R. Additive partial linear models with measurement errors. *Biometrika* **2008**, *95*, 667–678. [[CrossRef](#)] [[PubMed](#)]
32. Deng, G.H.; Liang, H. Model averaging for semiparametric additive partial linear models. *Sci. China Math.* **2010**, *53*, 1363–1376. [[CrossRef](#)]
33. Wang, L.; Yang, L.J. Spline-backfitted kernel smoothing of nonlinear additive autoregression model. *Ann. Stat.* **2007**, *35*, 2474–2503. [[CrossRef](#)]
34. Zhang, J.; Lian, H. Partially linear additive models with Unknown Link Functions. *Scand. J. Stat.* **2018**, *45*, 255–282. [[CrossRef](#)]
35. Hu, Y.A.; Zhao, K.F.; Lian, H. Bayesian Quantile Regression for Partially Linear Additive Models. *Stat. Comput.* **2015**, *25*, 651–668. [[CrossRef](#)]
36. Lian, H. Semiparametric estimation of additive quantile regression models by twofold penalty. *J. Bus. Econ. Stat.* **2012**, *30*, 337–350. [[CrossRef](#)]
37. Sherwood, B.; Wang, L. Partially linear additive quantile regression in ultra-high dimension. *Ann. Stat.* **2016**, *44*, 288–317. [[CrossRef](#)]
38. Yu, K.M.; Lu, Z.D. Local linear additive quantile regression. *Scand. J. Stat.* **2004**, *31*, 333–346. [[CrossRef](#)]
39. Du, P.; Cheng, G.; Liang, H. Semiparametric regression models with additive nonparametric components and high dimensional parametric components. *Comput. Stat. Data Anal.* **2012**, *56*, 2006–2017. [[CrossRef](#)]
40. Guo, J.; Tang, M.L.; Tian, M.Z.; Zhu, K. Variable selection in high-dimensional partially linear additive models for composite quantile regression. *Comput. Stat. Data Anal.* **2013**, *65*, 56–67. [[CrossRef](#)]
41. Liu, X.; Wang, L.; Liang, H. Estimation and variable selection for semiparametric additive partial linear models. *Stat. Sin.* **2011**, *21*, 1225–1248. [[CrossRef](#)]
42. Wang, L.; Liu, X.; Liang, H.; Carroll, R.J. Estimation and variable selection for generalized additive partial linear models. *Ann. Stat.* **2011**, *39*, 1827–1851. [[CrossRef](#)] [[PubMed](#)]
43. Denison, D.G.T.; Mallick, B.K.; Smith, A.F.M. Automatic Bayesian curving fitting. *J. R. Stat. B* **1998**, *60*, 333–350. [[CrossRef](#)]

44. Dimatteo, I.; Genovese, C.R.; Kass, R.E. Bayesian curve fitting with free-knot splines. *Biometrika* **2001**, *88*, 1055–1071. [[CrossRef](#)]
45. Holmes, C.C.; Mallick, B.K. Bayesian regression with multivariate linear splines. *J. R. Stat. B* **2001**, *63*, 3–17. [[CrossRef](#)]
46. Holmes, C.C.; Mallick, B.K. Generalized nonlinear modeling with multivariate free-knot regression splines. *J. Am. Stat. Assoc.* **2003**, *98*, 352–368. [[CrossRef](#)]
47. Lindstrom, M.J. Bayesian estimation of free-knot splines using reversible jump. *Comput. Stat. Data Anal.* **2002**, *41*, 255–269. [[CrossRef](#)]
48. Poon, W.-Y.; Wang, H.-B. Bayesian analysis of generalized partially linear single-index models. *Comput. Stat. Data Anal.* **2013**, *68*, 251–261. [[CrossRef](#)]
49. Poon, W.Y.; Wang, H.B. Multivariate partially linear single-index models: Bayesian analysis. *J. Nonparametr. Stat.* **2014**, *26*, 755–768. [[CrossRef](#)]
50. Chen, Z.Y.; Wang, H.B. Latent single-index models for ordinal data. *Stat. Comput.* **2018**, *28*, 699–711. [[CrossRef](#)]
51. Wang, H.B. A Bayesian multivariate partially linear single-index probit model for ordinal responses. *J. Stat. Comput. Sim.* **2018**, *88*, 1616–1636. [[CrossRef](#)]
52. LeSage, P.J.; Pace, R.K. *Introduction to Spatial Econometrics*; CRC Press: Boca Raton, FL, USA; London, UK; New York, NY, USA, 2009.
53. Metropolis, N.; Rosenbluth, A.W.; Rosenbluth, M.N.; Teller, A.H.; Teller, E. Equations of state calculations by fast computing machine. *J. Chem. Phys.* **1953**, *21*, 1087–1091. [[CrossRef](#)]
54. Hastings, W.K. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* **1970**, *57*, 97–109. [[CrossRef](#)]
55. Tanner, M.A. *Tools for Statistical Inference: Methods for the Exploration of Posterior Distributions and Likelihood Functions*, 2nd ed.; Springer: New York, NY, USA, 1993.
56. Panagiotelis, A.; Smith, M. Bayesian identification, selection and estimation of semiparametric functions in high-dimensional additive models. *J. Econom.* **2008**, *143*, 291–316. [[CrossRef](#)]
57. Green, P. Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika* **1995**, *82*, 711–732. [[CrossRef](#)]
58. Chen, M.-H.; Schmeiser, B.W. General hit-and-run Monte Carlo sampling for evaluating multidimensional integrals. *Oper. Res. Lett.* **1996**, *19*, 161–169. [[CrossRef](#)]
59. Su, L.J., Yang, Z. L. *Instrumental Variable Quantile Estimation of Spatial Autoregressive Models*; Working Paper; Singapore Management University: Singapore, 2009.
60. Gelman, A.; Rubin, D.B. Inference from iterative simulation using multiple sequences. *Stat. Sci.* **1992**, *7*, 457–511. [[CrossRef](#)]
61. Chen, Z.Y.; Chen, M.H.; Xing, G.D. Bayesian Estimation of Partially Linear Additive Spatial Autoregressive Models with P-Splines. *Math. Probl. Eng.* **2021**, *2021*, 1777469. [[CrossRef](#)]
62. Harezlak, J.; Ruppert, D.; Wand, M. *Semiparametric Regression with R*; Springer: New York, NY, USA, 2018.
63. Sun, Y.; Yan, H.J.; Zhang, W.Y.; Lu, Z. A Semiparametric spatial dynamic model. *Ann. Stat.* **2014**, *42*, 700–727. [[CrossRef](#)]