Statistics and Its Interface Volume 15 (2022) 105–113

Bayesian estimation for partially linear varying coefficient spatial autoregressive models

Ruiqin Tian, Dengke Xu, Jiang Du*, and Junfei Zhang

We propose a fully Bayesian estimation approach for partially linear varying coefficient spatial autoregressive models on the basis of B-spline approximations of nonparametric components. A computational efficient MCMC method that combines the Gibbs sampler with Metropolis-Hastings algorithm is implemented to simultaneously obtain the Bayesian estimates of unknown parameters, as well as their standard error estimates. Monte Carlo simulations are used to investigate the finite sample performance of the proposed method. Finally, a real data analysis of Boston housing data is used to illustrate the usefulness of the proposed methodology.

AMS 2000 subject classifications: Primary 62F15; secondary 62G05.
Keywords and phrases: Spatial autoregressive models, Partially linear varying coefficient models, Bayesian estimate, Gibbs sampler, B-spline.

1. INTRODUCTION

Spatial regression models are important tools in dealing with spatial dependent data which is widely available in many fields such as spatial econometrics, regional science economics, finance, biology and so on. Among them, the spatial autoregressive models (SARMs) proposed by Cliff and Ord (1973) are powerful and useful approaches for discovering structural dependence in data. Partially because that SARMs have prominent advantages such as high precision, the quite direct physical interpretation, the desiring prediction capability and the easy accessible inference procedures that are implemented in existing procedures in statistic software. Therefore, in recent years, SARMs have received much attention and have become an important research topic in econometrics and statistics. However, parametric spatial autoregressive models are highly sensitive to model misspecification, so they may not be adequate in many complex situations. To capture the underlying nonlinear relationships between the dependent variables and their associated covariates, some nonparametric and semiparametric spatial autoregressive models have recently been proposed and systematically analyzed in a frequentist setting. For example, Su and Jin (2010) developed a profile quasi-maximum likelihood estimation approach for partially linear spatial autoregressive model, and studied the asymptotic properties of the proposed estimators. Sun (2016) studied a spatial varying coefficient models with nonparametric spatial weights. Wei et al. (2017) proposed a semiparametric partially linear varying coefficient spatial autoregressive model, where a profile quasi-maximum likelihood approach based on the local-linear method was introduced. Du et al. (2018) developed the partially linear additive spatial autoregressive models, and proposed the estimation method via combining the spline approximations with instrumental variables estimation method. Under mild conditions, they established the asymptotic normality for the finite parametric vector and the optimal convergence rate for nonparametric functions. Sun et al. (2019) proposed a series-based two-stage least square estimating method and established the asymptotic theories of the proposed estimates.

In addition, with the fast development of Markov chain Monte Carlo (MCMC) methods and the Bayesian method has some good properties such that (i) it considers the use of prior information in obtaining better results; (ii) it can give more reliable results even though under small sample sizes; and (iii) it presents the conditional distributions of the parameters of interest, from which some quantities such as parameter of interest and its variance can be estimated, so that Bayesian inference for various statistical models has been receiving a lot of attention in recent years. For example, Tang and Duan (2012) developed a semiparametric Bayesian method for the generalized partial linear mixed models with longitudinal data. Combining the variance modelling technique with spline approximation, Xu and Zhang (2013) introduced Bayesian inference for the partially linear model with variance modelling. Tang et al. (2018) introduced a semiparametric Bayesian approach to make Bayesian inference on the transformation linear mixed models. Pfarrhofer and Piribauer (2019) proposed two shrinkage priors to make Bayesian variable selection for high-dimensional spatial autoregressive models. Hardouin (2019) proposed a variational bayesian method to estimate the logistic spatial regression. Wang and Tang (2020) considered Bayesian inference on a quantile regression model in the presence of nonignorable missing covariates. However, to the best of our knowledge, there is little work done for Bayesian analysis of partially linear varying coefficient spatial autoregressive models. To

*Corresponding author.

end this gap, we propose a hybrid algorithm to simultaneously estimate the parametric component and nonparametric coefficient functions via combining the Gibbs sampler with Metropolis-Hastings algorithm in this paper.

The outline of the paper is as follows. In Section 2, we introduce the partially linear varying coefficient spatial autoregressive models. A Bayesian procedure based on a data augmentation scheme for obtaining estimates is developed in Section 3. The full conditional distributions for implementing the sampling-based methods are also derived. To illustrate the proposed methodology, results obtained from a simulation study are presented in Section 4. As an application example, the Boston house price data are analyzed by the proposed method in Section 5. The paper is concluded with a brief discussion in Section 6.

2. PARTIALLY LINEAR VARYING COEFFICIENT SPATIAL AUTOREGRESSIVE MODELS

By accounting for spatial dependence among responses in partially linear varying coefficient models (Fan and Huang, 2005), we build the following partially linear varying coefficient spatial autoregressive models:

$$Y_i = \rho \sum_{j \neq i}^{n} w_{ij} Y_j + X_i^T \beta + Z_i^T \alpha(U_i) + \varepsilon_i, \quad i = 1, 2, \cdots, n, \quad (1)$$

where Y_i is a dependent variable corresponding to the ith observation, X_i is a vector of explanatory variables and U_i is a univariate observed covariate; $|\rho| \leq 1$ is an unknown parameter reflecting spatial autocorrelation between neighbors, $W = (w_{ij}), 1 \leq i, j \leq n$ is a specified $n \times n$ exogenously spatial weight matrix with zero diagonal elements; $\beta = (\beta_1, \cdots, \beta_p)^T$ is a $p \times 1$ vector of unknown regression coefficients and $\alpha(U_i) = (\alpha_1(U_i), \cdots, \alpha_d(U_i))^T$ is a $d \times 1$ vector of unknown smooth functions in the model; Z_i is the observation of explanatory variables associated with the nonparametric components. In addition, the superscript T denotes the transposed of a vector (or matrix). In this paper, we only consider univariate U_i, without loss of generality, it is assumed to be in the unit interval $[0, 1]$.

Let us work with the matrix notation. Denote $Y = (Y_1, Y_2, \cdots, Y_n)^T$, $X = (X_1, X_2, \cdots, X_n)^T$, $Z = (Z_1, Z_2, \cdots, Z_n)^T$, $\varepsilon = (\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_n)^T$, $F = (Z_1^T \alpha(U_1), Z_2^T \alpha(U_2), \cdots, Z_n^T \alpha(U_n))^T$. Then model (1) can be written as

$$Y = \rho W Y + X\beta + F + \varepsilon, \quad (2)$$

where we suppose the random errors $\varepsilon \sim N(0, \sigma^2 I_n)$, and I_n is a $n \times n$ identity matrix.

Based on model (2), we can obtain the following likelihood function

$$L(\beta, \alpha(\cdot), \sigma^2, \rho | Y, X, Z) \propto |A| (\sigma^2)^{-\frac{n}{2}} \exp\left\{ -\frac{1}{2\sigma^2} \|AY - X\beta - F\|^2 \right\}, \quad (3)$$

where $A = I_n - \rho W$. Due to $\alpha(\cdot)$ consists of nonparametric functions, (3) is not directly ready for optimization. Then, similar to He et al.(2005), we approximate the unknown smooth functions $\alpha(\cdot)$ in (3) by its B spline basis function. More specifically, let $B(u) = (B_1(u), \cdots, B_L(u))^T$ be B-spline basis functions with the order of M, where $L = k_n + M$, and k_n is the number of interior knots. We use the B-spline basis functions because they have bounded support and are numerically stable. Selection of knots is generally an important aspect of spline smoothing. In this paper, similar to He et al. (2005), the number of internal knots is taken to be the integer part of $n^{1/5}$. Then, $\alpha_k(u)$ can be approximated by

$$\alpha_k(u) \approx B(u)^T \gamma_k, \quad k = 1, \cdots, d.$$

Substituting this into (3), we can obtain

$$\ell(\beta, \gamma, \sigma^2, \rho | Y, X, Z) \propto |A| (\sigma^2)^{-\frac{n}{2}} \exp\left\{ -\frac{1}{2\sigma^2} \|AY - X\beta - \Omega\gamma\|^2 \right\}, \quad (4)$$

where $\gamma = (\gamma_1^T, \cdots, \gamma_d^T)^T$ and $\omega_i = I_d \otimes B(U_i) \cdot Z_i, \Omega = (\omega_1, \omega_2, \cdots, \omega_n)^T$.

3. BAYESIAN ESTIMATION OF PARTIALLY LINEAR VARYING COEFFICIENT SPATIAL AUTOREGRESSIVE MODELS

3.1 Prior density of parameters

To implement a Bayesian approach to estimate the parameters of the models, we need to specify a prior distribution for the parameters of the models. For simplicity, normal priors are chosen as $\beta \sim N(\beta_0, b_\beta)$, $\gamma \sim N(\gamma_0, \tau^2 I_K)(K = dL)$, where the hyperparameters β_0, γ_0 and b_β are assumed to be known vectors or matrix. In addition, τ^2 is taken to be Inverse Gamma $IG(a_\tau, b_\tau)$ with density function given by

$$p(\tau^2 | a_\tau, b_\tau) \propto (\tau^2)^{-a_\tau - 1} \exp\left(\frac{-b_\tau}{\tau^2} \right),$$

and similarly σ^2 is taken to be $IG(c_0, d_0)$, where a_τ, b_τ and c_0, d_0 are known positive numbers. For ρ we take a normal prior $\rho \sim N(0, c)$, where c may be chosen dependent on the prior belief of the researcher regarding the strength of spatial association in the data.

3.2 Gibbs sampling and full conditional distributions

Let $\theta = (\beta, \gamma, \sigma^2, \rho, \tau^2)$. Based on (4), we can sample from joint posterior distribution $p(\theta|Y, X, Z)$ by Gibbs sampling as following process.

Step 1. Setting initial values of parameters as $\theta^{(0)} = (\beta^{(0)}, \gamma^{(0)}, \sigma^{2(0)}, \rho^{(0)}, \tau^{2(0)})$.

Step 2. Based on $\theta^{(l)} = (\beta^{(l)}, \gamma^{(l)}, \sigma^{2(l)}, \rho^{(l)}, \tau^{2(l)})$, compute $A^{(l)} = I_n - \rho^{(l)} W$ and $\Sigma^{(l)} = \sigma^{2(l)} I_n$.

Step 3. Based on $\theta^{(l)} = (\beta^{(l)}, \gamma^{(l)}, \sigma^{2(l)}, \rho^{(l)}, \tau^{2(l)})$, sample $\theta^{(l+1)} = (\beta^{(l+1)}, \gamma^{(l+1)}, \sigma^{2(l+1)}, \rho^{(l+1)}, \tau^{2(l+1)})$ as follows:

• Sampling $\tau^{2(l+1)}|\cdot \sim$ Inverse Gamma $IG(c_\tau, d_\tau)$, where

$$c_\tau = \frac{K}{2} + a_\tau \text{ and } d_\tau = \frac{(\gamma^{(l)} - \gamma_0)^T (\gamma^{(l)} - \gamma_0)}{2} + b_\tau.$$

• Sampling $\gamma^{(l+1)}|\cdot \sim$ normal distribution $N(\tilde{\mu}_\gamma, \tilde{\Sigma}_\gamma)$, where

$$\tilde{\mu}_\gamma = \tilde{\Sigma}_\gamma (\tau^{-2(l+1)} I_K \gamma_0 + \Omega^T {\Sigma^{(l)}}^{-1} (A^{(l)} Y - X\beta^{(l)})),$$
$$\text{and } \tilde{\Sigma}_\gamma = (\tau^{-2(l+1)} I_K + \Omega^T {\Sigma^{(l)}}^{-1} \Omega)^{-1}.$$

• Sampling $\beta^{(l+1)}|\cdot \sim$ normal distribution $N(\tilde{\mu}_\beta, \tilde{\Sigma}_\beta)$, where

$$\tilde{\mu}_\beta = \tilde{\Sigma}_\beta (b_\beta^{-1} \beta_0 + X^T {\Sigma^{(l)}}^{-1} (A^{(l)} Y - \Omega\gamma^{(l+1)})),$$
$$\text{and } \tilde{\Sigma}_\beta = (b_\beta^{-1} + X^T {\Sigma^{(l)}}^{-1} X)^{-1}.$$

• Sampling $\sigma^{2(l+1)}|\cdot \sim$ Inverse Gamma $IG(c_\sigma, d_\sigma)$, where

$$c_\sigma = \frac{n}{2} + c_0, d_\sigma = e^T(\rho^{(l)}, l+1) e(\rho^{(l)}, l+1)/2 + d_0,$$
$$\text{and } e(\rho^{(l)}, l+1) = A^{(l)} Y - X\beta^{(l+1)} - \Omega\gamma^{(l+1)}.$$

• Sampling $\rho^{(l+1)}|\cdot \sim$: The full conditional distribution of ρ is

$$p(\rho|Y, X, Z, \beta^{(l+1)}, \gamma^{(l+1)}, \sigma^{2(l+1)}) \propto |A| \exp\{-\frac{1}{2} e^T(\rho, l+1)$$
$${\Sigma^{(l+1)}}^{-1} e(\rho, l+1) - \frac{\rho^2}{2c}\}.$$

Step 4. Repeating Steps 2 and 3.

Consequently, we can obtain sample series $(\beta^{(l)}, \gamma^{(l)}, \sigma^{2(l)}, \rho^{(l)}, \tau^{2(l)}), l = 1, 2, \cdots$ from the above the proposed algorithm. It is easy to show that the full conditional distributions of τ^2, γ, β and σ^2 are some familiar distributions, such as the Inverse Gamma and normal distributions. Therefore, it is fast and convenient to sample random variables from these standard distributions. Unfortunately, the full conditional distribution $p(\rho|Y, X, Z, \beta^{(l+1)}, \gamma^{(l+1)}, \sigma^{2(l+1)})$ is a non-standard distribution and rather complicated, thus drawing observations from this full conditional distribution is rather difficult. Hence, we use the well known Metropolis-within-Gibbs algorithm to sample observations from this full conditional distribution, for more details see Pfarrhofer and Piribauer (2019).

3.3 Bayesian estimation

We could generate the observations from the above proposed algorithm to obtain Bayesian estimators of the unknown parameters $(\beta, \gamma, \sigma^2, \rho)$ and their corresponding standard errors.

Let $\{(\beta^{(j)}, \gamma^{(j)}, \sigma^{2(j)}, \rho^{(j)}) : j = 1, 2, \cdots, J\}$ be the observations of $(\beta, \gamma, \sigma^2, \rho)$ generated from the joint full conditional distribution $p(\beta, \gamma, \sigma^2, \rho|Y, X, Z)$ via the proposed hybrid algorithm. The Bayesian estimates of $(\beta, \gamma, \sigma^2, \rho)$ are defined as:

$$\hat{\beta} = \frac{1}{J} \sum_{j=1}^{J} \beta^{(j)}, \hat{\gamma} = \frac{1}{J} \sum_{j=1}^{J} \gamma^{(j)}, \hat{\sigma}^2 = \frac{1}{J} \sum_{j=1}^{J} \sigma^{(2j)},$$
$$\hat{\rho} = \frac{1}{J} \sum_{j=1}^{J} \rho^{(j)}.$$

As is shown by Geyer (1992), $(\hat{\beta}, \hat{\gamma}, \hat{\sigma}^2, \hat{\rho})$ are consistent estimates of their corresponding posterior means. Similarly, the consistent estimates of the posterior covariance matrices $\mathrm{Var}(\beta|\theta_{\beta^-}, Y, X, Z), \mathrm{Var}(\gamma|\theta_{\gamma^-}, Y, X, Z), \mathrm{Var}(\sigma^2|\theta_{\sigma^{2-}}, Y, X, Z)$ and $\mathrm{Var}(\rho|\theta_{\rho^-}, Y, X, Z)$ can be obtained via the sample covariance matrices of the observations $\{(\beta^{(j)}, \gamma^{(j)}, \sigma^{2(j)}, \rho^{(j)}) : j = 1, 2, \cdots, J\}$, where θ_{β^-} denotes the parameter θ excluding β. For example,

$$\widehat{\mathrm{Var}}(\beta|\theta_{\beta^-}, Y, X, Z) = (J-1)^{-1} \sum_{j=1}^{J} (\beta^{(j)} - \hat{\beta})(\beta^{(j)} - \hat{\beta})^T.$$

Thus, standard errors can be obtained from the diagonal elements of these matrices.

4. SIMULATION STUDY

In this section, we investigate the finite sample properties of the proposed estimation method. The data are generated from the following model

$$Y_i = \rho \sum_{j \neq i}^{n} w_{ij} Y_j + X_i^T \beta + Z_i^T \alpha(U_i) + \varepsilon_i, \quad i = 1, 2, \cdots, n, \quad (5)$$

where X_i is generated from multivariate normal distribution with zero mean vector and covariance matrix being $\tilde{\Sigma} = 0.5^{|k-j|} (k = 1, 2, 3, j = 1, 2, 3)$, Z_i is also generated from multivariate normal distribution with zero mean vector and covariance matrix being $\tilde{\Sigma} = 0.5^{|k-j|} (k = 1, 2, j = 1, 2)$, U_i follows a uniform distribution $U(0, 1)$. We set $\beta = (1, 1, 1)^T$, and $\alpha(u_i) = (\alpha_1(u_i), \alpha_2(u_i))^T$, where $\alpha_1(u_i) = \sin(2\pi u_i)$ and $\alpha_2(u_i) = 8u_i(1-u_i^2)$. For comparison, three different values $\sigma^2 = 0.1, 0.5$ and 1, which represent strong and weak signal-to-noise ratios, are considered in this section. In addition, let the spatial parameter $\rho = -0.5, 0, 0.5$, which represents different spatial dependencies. Similar to Xie et al. (2020), the weight matrix is set to be $W = I_R \otimes H_m, H_m = (l_m l_m^T - I_m)/(m-1)$, where l_m is an m-dimensional vector with all elements being 1,"$\otimes$" means Kronecker product, $n = R \times m$. We choose R to be 25, 40 and m to be 4, 8 in this simulation.

In the simulation, we consider the noninformative prior information type of hyperparameter values for unknown parameters $\beta, \gamma, \sigma^2, \rho, \tau^2$: $\beta_0 = (0,0,0)^T, b_\beta = 100 \times I_3$, $\gamma_0 = (0,0,0)^T, c_0 = d_0 = 0.01, a_\tau = b_\tau = 1, c = 10$.

Based on the above settings and the generated data set, the preceding proposed hybrid algorithm is used to evaluate the Bayesian estimates of unknown parameters based on 100 replications. To investigate the convergence of the proposed hybrid algorithm, the estimated potential scale reduction (EPSR) values (Gelman, 1996) are computed for a number of test runs on the base of three parallel chains of observations via three different starting values. We observe that in all test runs, the EPSR values are close to 1 and less than 1.2 after 3000 iterations. Therefore, we can collect the observations after 3000 iterations with $J = 2000$ in producing the Bayesian estimates for each replication and posterior summary of the parameters is presented in Tables 1-3. In order to evaluate the performance of estimator $\hat{\alpha}_1(u)$ and $\hat{\alpha}_2(u)$, we use the square root of average square errors to assess, which is defined as follow: (RASE),

$$\text{RASE}(\hat{\alpha}_k(u)) = E\left\{\frac{1}{n}\sum_{i=1}^{n}\left[\hat{\alpha}_k(u_i) - \alpha_k(u_i)\right]^2\right\}^{\frac{1}{2}}, k = 1, 2.$$

The simulation results are reported in Table 4. Moreover, to see accuracy of estimate of functions $\alpha_1(u)$ and $\alpha_2(u)$ directly, we plot the true value of functions $\alpha_1(u)$ and $\alpha_2(u)$ against its estimate under different cases. To save space, we only list some nonparametric curve fitting results under different spatial parameters in Figures 1-6.

In Tables 1-3, "Bias" denotes the absolute difference between the true value and the average of the Bayesian estimates of the parameters based on 100 replications, "SD" denotes standard deviation of the Bayesian estimates and "RMS" is the root mean square between the estimates based on 100 replications and its true value. We can draw the following observations from Tables 1-3. (i) Bayesian estimates are reasonably accurate under all the considered settings in the sense that their Bias values and RMS values are reasonable small, where the RMS values are close to their corresponding SD values;(ii) Based on different spatial parameters, the results of Bayesian estimation are similar; (iii) As the variance σ^2 becomes smaller and smaller, the Bayesian estimation method behaves better and better. Examination

Table 1. Bayesian estimates of unknown parameters under different cases when $\rho = 0.5$

σ^2	R	Para.	$m=4$			$m=8$		
			Bias	SD	RMS	Bias	SD	RMS
1	25	β_1	0.0210	0.1077	0.1098	0.0045	0.0829	0.0831
		β_2	0.0211	0.1441	0.1456	0.0143	0.1000	0.1010
		β_3	0.0023	0.1315	0.1315	0.0170	0.0920	0.0936
		σ^2	0.0117	0.1540	0.1544	0.0187	0.0998	0.1015
		ρ	0.0023	0.0328	0.0329	0.0022	0.0304	0.0305
	40	β_1	0.0047	0.1024	0.1025	0.0082	0.0583	0.0589
		β_2	0.0027	0.1113	0.1114	0.0041	0.0634	0.0635
		β_3	0.0157	0.0970	0.0982	0.0009	0.0658	0.0658
		σ^2	0.0142	0.1281	0.1289	0.0026	0.0786	0.0786
		ρ	0.0064	0.0266	0.0274	0.0001	0.0252	0.0252
0.5	25	β_1	0.0007	0.0928	0.0928	0.0112	0.0674	0.0683
		β_2	0.0048	0.1039	0.1040	0.0087	0.0681	0.0686
		β_3	0.0075	0.0895	0.0898	0.0021	0.0628	0.0629
		σ^2	0.0152	0.0784	0.0799	0.0027	0.0524	0.0524
		ρ	0.0019	0.0216	0.0217	0.0039	0.0248	0.0251
	40	β_1	0.0034	0.0685	0.0685	0.0038	0.0467	0.0469
		β_2	0.0003	0.0858	0.0858	0.0026	0.0504	0.0505
		β_3	0.0041	0.0741	0.0742	0.0109	0.0421	0.0435
		σ^2	0.0191	0.0617	0.0646	0.0032	0.0384	0.0385
		ρ	0.0016	0.0166	0.0167	0.0003	0.0187	0.0187
0.1	25	β_1	0.0053	0.0411	0.0415	0.0034	0.0281	0.0283
		β_2	0.0052	0.0468	0.0471	0.0024	0.0258	0.0259
		β_3	0.0018	0.0371	0.0371	0.0047	0.0250	0.0254
		σ^2	0.0039	0.0145	0.0151	0.0004	0.0102	0.0103
		ρ	0.0017	0.0111	0.0112	0.0006	0.0103	0.0104
	40	β_1	0.0016	0.0304	0.0305	0.0006	0.0187	0.0187
		β_2	0.0027	0.0382	0.0383	0.0033	0.0224	0.0227
		β_3	0.0000	0.0342	0.0342	0.0028	0.0215	0.0217
		σ^2	0.0044	0.0125	0.0133	0.0004	0.0087	0.0087
		ρ	0.0009	0.0080	0.0081	0.0021	0.0073	0.0076

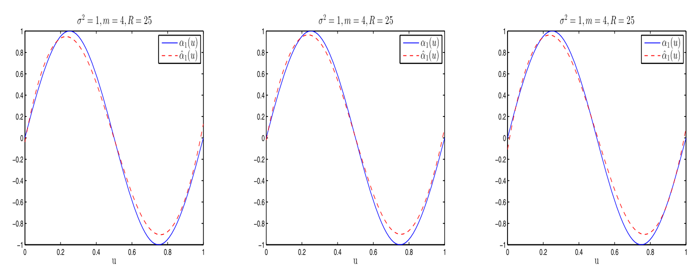

Figure 1. The estimated function versus true value of $\alpha_1(u)$ with different ρ's(from left to right responding to $\rho = -0.5, 0, 0.5$) when $m = 4, R = 25, \sigma^2 = 1$.

of Figures 1-6 shows that the shapes of the estimated nonparametric function are very close to the corresponding true line under all the considered settings, which agrees with what was discovered from Table 4. All in all, all the above findings show that the preceding proposed estimation procedure can well recover the true information in partially linear varying coefficient spatial autoregressive models.

Table 2. Bayesian estimates of unknown parameters under different cases when $\rho = 0$

σ^2	R	Para.	$m=4$ Bias	$m=4$ SD	$m=4$ RMS	$m=8$ Bias	$m=8$ SD	$m=8$ RMS
1	25	β_1	0.0123	0.1074	0.1081	0.0116	0.0863	0.0871
		β_2	0.0002	0.1349	0.1349	0.0048	0.0821	0.0822
		β_3	0.0057	0.1145	0.1146	0.0166	0.0772	0.0790
		σ^2	0.0014	0.1493	0.1493	0.0081	0.1009	0.1012
		ρ	0.0018	0.0512	0.0513	0.0073	0.0532	0.0537
	40	β_1	0.0039	0.1049	0.1050	0.0078	0.0728	0.0732
		β_2	0.0165	0.1188	0.1199	0.0005	0.0863	0.0863
		β_3	0.0057	0.1035	0.1036	0.0076	0.0779	0.0783
		σ^2	0.0055	0.1121	0.1122	0.0026	0.0770	0.0770
		ρ	0.0021	0.0421	0.0422	0.0039	0.0404	0.0406
0.5	25	β_1	0.0094	0.0762	0.0767	0.0079	0.0619	0.0624
		β_2	0.0002	0.0959	0.0959	0.0053	0.0573	0.0576
		β_3	0.0039	0.0811	0.0812	0.0108	0.0555	0.0565
		σ^2	0.0017	0.0747	0.0748	0.0037	0.0509	0.0510
		ρ	0.0000	0.0390	0.0390	0.0035	0.0408	0.0409
	40	β_1	0.0030	0.0653	0.0654	0.0080	0.0494	0.0500
		β_2	0.0033	0.0841	0.0841	0.0048	0.0581	0.0583
		β_3	0.0131	0.0759	0.0770	0.0013	0.0530	0.0530
		σ^2	0.0077	0.0540	0.0546	0.0023	0.0413	0.0414
		ρ	0.0081	0.0267	0.0279	0.0032	0.0290	0.0292
0.1	25	β_1	0.0024	0.0347	0.0348	0.0034	0.0280	0.0282
		β_2	0.0018	0.0437	0.0437	0.0023	0.0257	0.0258
		β_3	0.0025	0.0383	0.0384	0.0048	0.0249	0.0254
		σ^2	0.0015	0.0157	0.0158	0.0004	0.0102	0.0102
		ρ	0.0012	0.0179	0.0179	0.0009	0.0193	0.0193
	40	β_1	0.0029	0.0320	0.0321	0.0036	0.0221	0.0223
		β_2	0.0014	0.0354	0.0355	0.0023	0.0261	0.0262
		β_3	0.0040	0.0326	0.0329	0.0005	0.0237	0.0237
		σ^2	0.0015	0.0119	0.0120	0.0007	0.0083	0.0083
		ρ	0.0007	0.0136	0.0137	0.0012	0.0135	0.0136

Table 3. Bayesian estimates of unknown parameters under different cases when $\rho = -0.5$

σ^2	R	Para.	$m=4$ Bias	$m=4$ SD	$m=4$ RMS	$m=8$ Bias	$m=8$ SD	$m=8$ RMS
1	25	β_1	0.0123	0.1092	0.1099	0.0015	0.0783	0.0783
		β_2	0.0102	0.1373	0.1376	0.0022	0.0861	0.0861
		β_3	0.0070	0.1190	0.1192	0.0063	0.0882	0.0884
		σ^2	0.0058	0.1581	0.1582	0.0046	0.1101	0.1102
		ρ	0.0002	0.0637	0.0637	0.0024	0.0786	0.0786
	40	β_1	0.0021	0.0926	0.0927	0.0116	0.0706	0.0715
		β_2	0.0029	0.1176	0.1176	0.0063	0.0816	0.0818
		β_3	0.0198	0.1074	0.1093	0.0018	0.0748	0.0749
		σ^2	0.0110	0.1100	0.1105	0.0029	0.0830	0.0830
		ρ	0.0116	0.0446	0.0461	0.0052	0.0548	0.0550
0.5	25	β_1	0.0061	0.0759	0.0761	0.0081	0.0619	0.0624
		β_2	0.0016	0.0968	0.0968	0.0051	0.0572	0.0574
		β_3	0.0065	0.0816	0.0819	0.0110	0.0555	0.0566
		σ^2	0.0027	0.0760	0.0761	0.0040	0.0508	0.0510
		ρ	0.0016	0.0484	0.0484	0.0039	0.0568	0.0569
	40	β_1	0.0019	0.0657	0.0658	0.0082	0.0499	0.0506
		β_2	0.0020	0.0833	0.0833	0.0046	0.0579	0.0581
		β_3	0.0142	0.0763	0.0776	0.0012	0.0530	0.0530
		σ^2	0.0067	0.0549	0.0553	0.0021	0.0414	0.0415
		ρ	0.0087	0.0333	0.0344	0.0037	0.0402	0.0404
0.1	25	β_1	0.0043	0.0350	0.0352	0.0035	0.0280	0.0282
		β_2	0.0013	0.0450	0.0450	0.0023	0.0256	0.0257
		β_3	0.0025	0.0371	0.0372	0.0048	0.0249	0.0254
		σ^2	0.0015	0.0157	0.0157	0.0005	0.0102	0.0102
		ρ	0.0029	0.0235	0.0237	0.0011	0.0269	0.0270
	40	β_1	0.0025	0.0311	0.0312	0.0037	0.0223	0.0226
		β_2	0.0037	0.0343	0.0345	0.0022	0.0260	0.0261
		β_3	0.0087	0.0302	0.0315	0.0004	0.0237	0.0237
		σ^2	0.0038	0.0117	0.0123	0.0007	0.0083	0.0083
		ρ	0.0005	0.0168	0.0168	0.0016	0.0188	0.0189

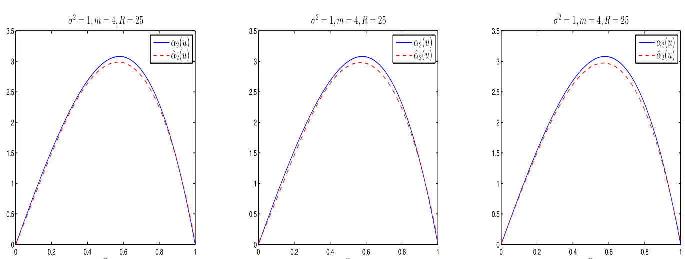

Figure 2. The estimated function versus true value of $\alpha_2(u)$ with different ρ's(from left to right responding to $\rho = -0.5, 0, 0.5$) when $m = 4, R = 25, \sigma^2 = 1$.

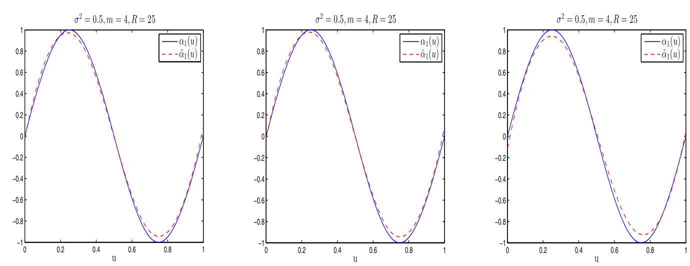

Figure 3. The estimated function versus true value of $\alpha_1(u)$ with different ρ's(from left to right responding to $\rho = -0.5, 0, 0.5$) when $m = 4, R = 25, \sigma^2 = 0.5$.

In addition, to investigate the sensitivity of the Bayesian estimate for $\alpha(u_i)$ to the selection of the number of internal knots, we consider the other two choices of K, i.e. $K_1 = \lfloor K_0/1.5 \rfloor$ and $K_2 = \lfloor 1.5K_0 \rfloor$, where K_0 is the optimal number of interior knots and $\lfloor s \rfloor$ denotes the largest integer not greater than s. Here, we only present the results of Bayesian estimates in Table 5 and list some curves of nonparametric estimation in Figures 7-8 when $\rho = 0.5$ and $\sigma^2 = 0.5$ under different choices of K.

By viewing Table 5 and comparing it with Table 1, we can see that the Bayesian estimates are reasonably accurate regardless of the values of K in the sense of their SD values and RMS values. The results of the simulation studies show the proposed estimation method performs well in K_1 and K_2. In

Table 4. The estimate for the nonparametric components($\hat{\alpha}_1(u)$ and $\hat{\alpha}_2(u)$) under different cases

ρ	σ^2	R	$m=4$ RASE($\hat{\alpha}_1(u)$)	$m=4$ RASE($\hat{\alpha}_2(u)$)	$m=8$ RASE($\hat{\alpha}_1(u)$)	$m=8$ RASE($\hat{\alpha}_2(u)$)
0.5	1	25	0.0635	0.0747	0.0335	0.0453
		40	0.0528	0.0455	0.0203	0.0263
	0.5	25	0.0527	0.0459	0.0287	0.0239
		40	0.0440	0.0350	0.0172	0.0187
	0.1	25	0.0188	0.0256	0.0111	0.0071
		40	0.0178	0.0121	0.0102	0.0071
0	1	25	0.0576	0.0716	0.0421	0.0392
		40	0.0496	0.0501	0.0351	0.0333
	0.5	25	0.0360	0.0408	0.0262	0.0246
		40	0.0314	0.0344	0.0210	0.0191
	0.1	25	0.0153	0.0087	0.0111	0.0072
		40	0.0126	0.0080	0.0107	0.0057
-0.5	1	25	0.0637	0.0593	0.0465	0.0375
		40	0.0514	0.0552	0.0322	0.0335
	0.5	25	0.0375	0.0399	0.0263	0.0249
		40	0.0319	0.0353	0.0210	0.0194
	0.1	25	0.0173	0.0101	0.0111	0.0072
		40	0.0160	0.0108	0.0107	0.0059

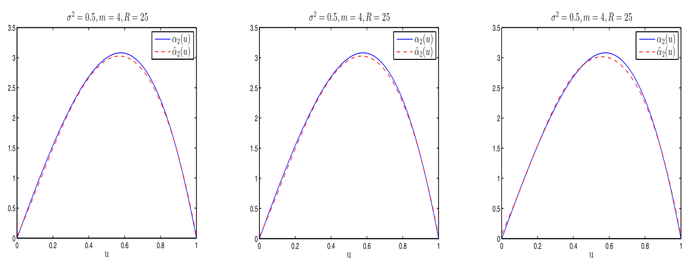

Figure 4. The estimated function versus true value of $\alpha_2(u)$ with different ρ's(from left to right responding to $\rho = -0.5, 0, 0.5$) when $m = 4, R = 25, \sigma^2 = 0.5$.

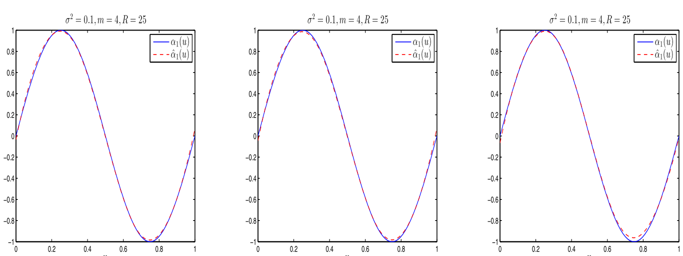

Figure 5. The estimated function versus true value of $\alpha_1(u)$ with different ρ's(from left to right responding to $\rho = -0.5, 0, 0.5$) when $m = 4, R = 25, \sigma^2 = 0.1$.

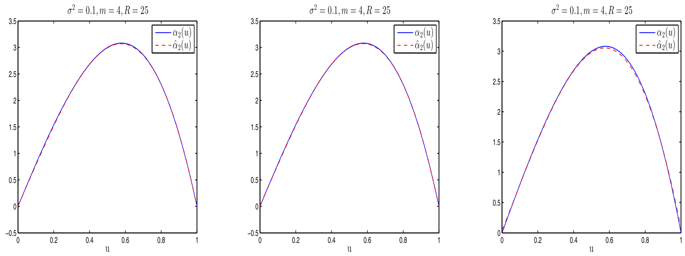

Figure 6. The estimated function versus true value of $\alpha_2(u)$ with different ρ's(from left to right responding to $\rho = -0.5, 0, 0.5$) when $m = 4, R = 25, \sigma^2 = 0.1$.

Table 5. Bayesian estimates of unknown parameters when $\rho = 0.5$ and $\sigma^2 = 0.5$ under different choices of K.

K	R	Para.	$m=4$ Bias	$m=4$ SD	$m=4$ RMS	$m=8$ Bias	$m=8$ SD	$m=8$ RMS
K_1	25	β_1	0.0041	0.0805	0.0806	0.0012	0.0606	0.0606
		β_2	0.0048	0.0961	0.0962	0.0002	0.0648	0.0648
		β_3	0.0011	0.0841	0.0842	0.0036	0.0540	0.0542
		σ^2	0.0166	0.0834	0.0851	0.0184	0.0534	0.0565
		ρ	0.0012	0.0188	0.0188	0.0000	0.0212	0.0212
	40	β_1	0.0063	0.0677	0.0680	0.0008	0.0496	0.0496
		β_2	0.0025	0.0760	0.0761	0.0010	0.0472	0.0472
		β_3	0.0079	0.0637	0.0642	0.0039	0.0452	0.0454
		σ^2	0.0077	0.0551	0.0556	0.0057	0.0400	0.0404
		ρ	0.0053	0.0186	0.0194	0.0004	0.0178	0.0178
K_2	25	β_1	0.0283	0.0934	0.0976	0.0163	0.0610	0.0632
		β_2	0.0172	0.1072	0.1086	0.0058	0.0672	0.0674
		β_3	0.0029	0.0898	0.0898	0.0019	0.0611	0.0612
		σ^2	0.0011	0.0690	0.0690	0.0042	0.0597	0.0599
		ρ	0.0029	0.0240	0.0242	0.0020	0.0225	0.0225
	40	β_1	0.0142	0.0697	0.0711	0.0048	0.0504	0.0506
		β_2	0.0019	0.0726	0.0726	0.0088	0.0572	0.0578
		β_3	0.0045	0.0686	0.0688	0.0015	0.0520	0.0520
		σ^2	0.0138	0.0602	0.0618	0.0032	0.0376	0.0378
		ρ	0.0019	0.0165	0.0166	0.0005	0.0159	0.0159

Table 6. The estimate for the nonparametric components($\hat{\alpha}_1(u)$ and $\hat{\alpha}_2(u)$) when $\rho = 0.5$ and $\sigma^2 = 0.5$ under different choices of K

K	R	$m=4$ RASE($\hat{\alpha}_1(u)$)	$m=4$ RASE($\hat{\alpha}_2(u)$)	$m=8$ RASE($\hat{\alpha}_1(u)$)	$m=8$ RASE($\hat{\alpha}_2(u)$)
K_1	25	0.0741	0.0520	0.0700	0.0251
	40	0.0701	0.0227	0.0277	0.0198
K_2	25	0.0541	0.0574	0.0247	0.0238
	40	0.0368	0.0400	0.0164	0.0159

addition, we found that the proposed method is insensitive to the tuning parameter K, which is chosen in the interval $[K_{1n}, K_{2n}]$, where $K_{1n} = \lfloor n^{1/5}/1.5 \rfloor$ and $K_{2n} = \lfloor 1.5n^{1/5} \rfloor$. From Figures 7-8, we can obtain that the shapes of the estimated nonparametric function are very similar to those in Figures 3-4, which show that the proposed method is consistent.

Lastly, in order to show that nonparametric modeling is very important if there is a nonlinear relationship between

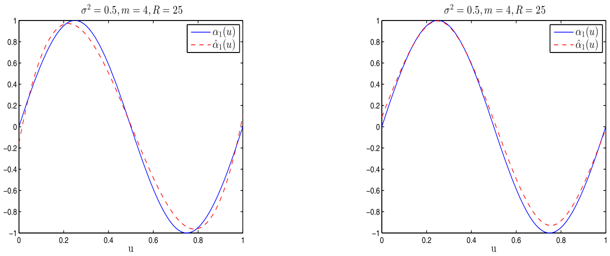

Figure 7. The estimated function versus true value of $\alpha_1(u)$ for different choices of K (from left to right responding to $K = K_1, K_2$) when $m = 4, R = 25, \sigma^2 = 0.5$.

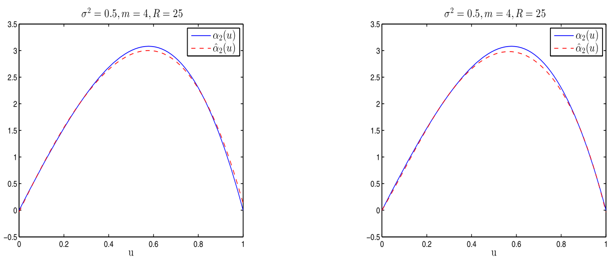

Figure 8. The estimated function versus true value of $\alpha_2(u)$ for different choices of K (from left to right responding to $K = K_1, K_2$) when $m = 4, R = 25, \sigma^2 = 0.5$.

the response variables and the covariates, we compare the proposed models (denoted as "M1") with the classical spatial autoregressive models (denoted as "M2") using average square error and biases, which are defined as follows:

$$\text{MSE}_y = \frac{1}{n}\sum_{i=1}^{n}(y_i - y_i^*)^2, \text{Bias}_y = \frac{1}{n}\sum_{i=1}^{n}(y_i - y_i^*),$$

where y_i^* is the fitting value. Here, we present the results when $\rho = 0.5$ and $\sigma^2 = 0.5$ in Table 7 under different variances σ^2. Examination of Table 7 indicates that (i) The proposed method (M1) is superior to the general spatial autoregressive models (M2) in the sense of their Bias values and MSE values. (ii) As the variance σ^2 becomes smaller and smaller, the proposed Bayesian estimation method behaves better and better.

5. REAL DATA ANALYSIS

In this section, we apply the proposed method to the Boston housing price data which contains 506 observations in Boston Standard Metropolitan Statistical Area in 1970. This data set has been analyzed by Pace and Gilley (1997), Zhang et al.(2011) and Fan and Huang (2005) without considering the spatial dependence of different observations. Similar to Fan and Huang (2005), we use partially linear varying coefficient models with the spatial effect to fit the

Table 7. Results of comparing the two methods when $\rho = 0.5$ and $\sigma^2 = 0.5$ under different variances σ^2

σ^2	R	Method	$m = 4$		$m = 8$	
			Bias_y	MSE_y	Bias_y	MSE_y
1	25	$M1$	0.0005	1.2844	0.0088	1.1315
		$M2$	−0.0205	9.4926	0.0260	7.7129
	40	$M1$	0.0133	1.3788	0.0033	1.2053
		$M2$	−0.0420	9.5792	0.0036	7.7111
0.5	25	$M1$	−0.0103	0.6486	−0.0116	0.5713
		$M2$	−0.0275	8.7363	0.0276	7.0942
	40	$M1$	0.0033	0.7106	0.0024	0.6020
		$M2$	0.0271	8.5937	0.0029	7.0653
0.1	25	$M1$	0.0007	0.1261	0.0016	0.1154
		$M2$	−0.0219	8.0652	0.0036	6.2635
	40	$M1$	0.0046	0.1370	0.0011	0.1204
		$M2$	−0.0481	8.2443	0.0021	6.5401

data, i.e. use the following model to fit the data

$$Y_i = \rho\sum_{j=1}^{n} w_{ij}Y_j + \sum_{k=1}^{3} Z_{ik}\alpha_k(U_i) + X_{i1}\beta_1 + X_{i2}\beta_2 + \varepsilon_i. \quad (6)$$

Similar to Fan and Huang (2005), we take MEDV (median value of owner-occupied homes in 1,000USD) as the response, and the following predictors as the Z-variables and X-variables: CRIM (per capita crime rate by town, denoted by Z_1), RM(average number of rooms per dwelling, denoted by Z_2), TAX (full-value property-tax rate per 10,000 USD, denoted by Z_3), PTRATIO(pupil teacher ratio by town, denoted by X_1), and AGE (proportion of owner-occupied units built prior to 1940, denoted by X_2). We set index variable as $U = \sqrt{LSTAT}$, square root transformation on LSTAT(the percentage of lower status of the population). In addition, similar to Ertur and KochGrowth (2007), we consider the following spatial weight matrix and let $W = (w_{ij})$, where

$$w_{ij} = \begin{cases} 0, i = j \\ e^{-2d_{ij}/1000}, \text{otherwise}, \end{cases}$$

d_{ij} is the great-circle distance between the longitude and latitude coordinates of any two houses.

The preceding proposed hybrid algorithm is used to obtain Bayesian estimates of β's, σ^2's and ρ's, where we use noninformative prior information for all unknown parameters. To test the convergence of the algorithm, plot of the EPSR values for all the unknown parameters against iterations is presented in Figure 9, which indicates that the algorithm converges about 3000 iterations because EPSR values of all unknown parameters are less than 1.05 about 3000 iterations. We calculate Bayesian estimates (EST), standard deviation estimates (SD), and 95% credible intervals (CI) of the Bayesian estimates of β's, σ^2's and ρ's. Results are given in Table 8. Although we obtained confidence interval with 95% confidence level for regression coefficient for

Table 8. *The real example: Bayesian estimates, standard deviation estimates and 95% credible intervals.*

Parameter	EST	SD	CI
β_1	-0.2982	0.0613	$(-0.4188, -0.1817)$
β_2	0.0121	0.0092	$(-0.0063, 0.0300)$
σ^2	16.3510	1.5197	(14.4372, 18.6093)
ρ	0.1328	0.0418	(0.0754, 0.1845)

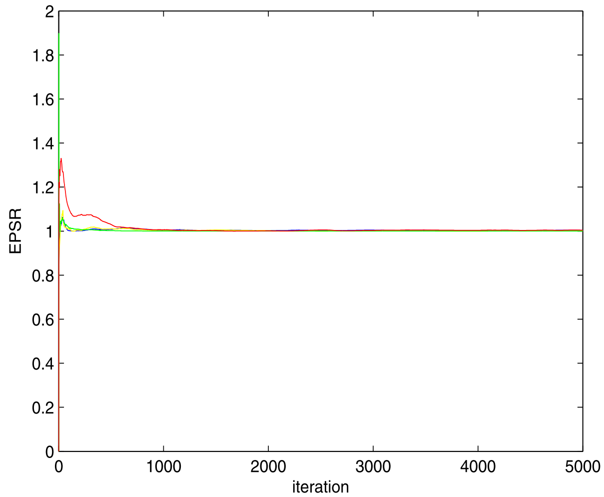

Figure 9. EPSR values of all parameters against iterations in the real example.

X_2 which include the zero, we find the P-value is 0.0541. Therefore, we slightly accept the null hypothesis that X_2 is insignificant. For the nonparametric components, the curves of the estimated functions $\alpha_1(u), \alpha_2(u), \alpha_3(u)$ and the corresponding 95% pointwise confidence intervals are shown in Figure 10. From the analysis results, we can obtain the following conclusions. The estimation of the spatial parameter is $\hat{\rho} = 0.1328$ with its confidence interval not concluding zero which indicates there exists a substantial spatial relationship among the responses. The regression coefficient of X_1 is negative, which reveals that the housing price would decrease as the pupil/teacher ratio increases. These conclusions are consistent with the analysis results of Du et al. (2018). The regression coefficient of X_2 is positive and small, which implies that the proportion of owner-occupied units built prior to 1940 has a relatively small positive effect on the housing prices. From Figure 10, the variables CRIM, RM, TAX demonstrate non-linear effects on the response, which is consistent with the findings of Du et al.(2018) and other literatures.

6. CONCLUSION AND DISCUSSION

In this paper, we propose a fully Bayesian estimation approach for partially linear varying coefficient spatial autoregressive models on the basis of B-spline approximations of nonparametric components. A computational efficient MCMC method which combines the Gibbs sampler with Metropolis-Hastings algorithm is implemented to simultaneously obtain the Bayesian estimates of unknown parameters. A simulation study and a real data analysis of Boston housing data are used to show the efficiency of the proposed Bayesian approach. The results show that the developed Bayesian method is highly efficient and computationally fast.

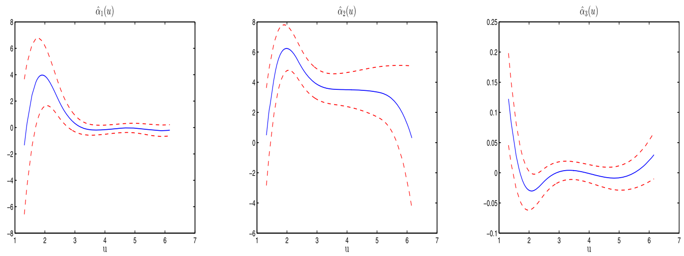

Figure 10. The averages of 95% confidence regions for $\alpha_1(u), \alpha_2(u)$ and $\alpha_3(u)$, where solid curves are the estimated curves of $\alpha_1(u), \alpha_2(u)$ and $\alpha_3(u)$.

Furthermore, there are several interesting issues that merit further research. Specific considerations are as follows:(i)Comparing to classical mean regression, quantile regression possesses much more flexibility in distribution of random error so that it is robust to non-normal errors and outliers. Therefore, how to combine quantile regression and spatial autoregressive models warrants a future investigation. Thus, the proposed method does not depend on the normality assumption of random errors; (ii) The covariate selection is an important issue in spatial analysis. In particular, in a Bayesian framework, Park and Casella (2008) proposed a BLasso by imposing the double exponential prior on the regression coefficients and the gamma distribution on the shrinkage parameter. It is interesting to extend the BLasso approach to spatial autoregressive models;(iii) A possible extension of the current model is considering when covariates are missing under different missingness mechanism.

ACKNOWLEDGEMENTS

The authors are grateful to the Editor, the Associate Editor and the referees for their helpful comments and suggestions, which lead to improvement of an earlier version of this paper. Tian's work was supported by the National Natural Science Foundation of China (11801514), and Startup Foundation for Talents in Hangzhou Normal University(2019QDL039); Xu's work was supported by the Key projects of statistical research in Zhejiang Province(20TJZZ13); Du's work was supported by the National Natural Science Foundation of China (11971045,11771032), the Natural Science Foundation of

Beijing Municipality (1202001) and the Science and Technology Project of Beijing Municipal Education Commission (KM201910005015,KM202010005026).

Received 9 July 2020

REFERENCES

[1] ANSELIN, L. (1988). *Spatial econometrics: methods and models*. Kluwer Academic Publishers, The Netherlands.
[2] CLIFF, A. and ORD, J. K. (1973). *Spatial autocorrelation*. Pion, London.
[3] DU, J., SUN, X. Q., CAO, R. Y. and ZHANG, Z. Z. (2018). Statistical inference for partially linear additive spatial autoregressive models. *Spatial Statistics* **25** 52–67. MR3809255
[4] ERTUR, C. and KOCHGROWTH, W. (2007). Growth, technological interdependence and spatial externalities:theory and evidence. *Journal of Applied Economic* **22** 1033–1062. MR2408972
[5] FAN, J. Q. and HUANG, T. (2005). Profile likelihood inferences on semiparametric varyingcoefficient partially linear models. *Bernoulli* **11(6)** 1031–1057. MR2189080
[6] HARDOUIN, C. (2019). A variational method for parameter estimation in a logistic spatial regression. *Spatial Statistics* **31** 100365–100365. MR3957509
[7] GELMAN, A. (1996). *Inference and Monitoring Convergence in Markov Chain Monte Carlo in Practice*. Chapman and Hall, London.
[8] GEYER, C. J.(1992). Practical Markov Chain Monte Carlo. *Statistical Science* **7** 473–483.
[9] HE, X., FUNG, W. K. and ZHU, Z. Y.(2005). Robust estimation in a generalized partial linear model for clustered data. *Journal of the American Statistical Association* **100** 1176–1184. MR2236433
[10] PACE, R. K. and GILLEY, O. W. (1997). Using the spatial configuration of the data to improve estimation. *Journal of Real Estate Finance & Economics* **14(3)**330–340.
[11] PARK, T. and CASELLA, G.(2008). The Bayesian lasso. *Journal of the American Statistical Association* **103(482)**681–686. MR2524001
[12] PFARRHOFER, M. and PIRIBAUER, P. (2019). Flexible shrinkage in high-dimensional Bayesian spatial autoregressive models. *Spatial Statistics* **29** 109–128. MR3880101
[13] SU, L. J. and JIN, S. N. (2010). Profile quasi-maximum likelihood estimation of partially linear spatial autoregressive models. *Journal of Econometrics* **157** 18–33. MR2652276
[14] SUN, Y., ZHANG, Y. and HUANG, J. Z. (2019). Estimation of a semiparametric varying-coefficient mixed regressive spatial autoregressive model. *Econometrics and Statistics* **9** 140–155. MR3907678
[15] SUN, Y. G. (2016). Functional-coefficient spatial autoregressive models with nonparametric spatial weights. *Journal of Econometrics* **195** 134–153. MR3545298
[16] TANG, N. S.and DUAN, X. D. (2012). A semiparametric Bayesian approach to generalized partial linear mixed models for longitudinal data. *Computational Statistics & Data Analysis* **56** 4348–4365. MR2957878
[17] TANG, N. S., WU, Y. and CHEN, D.(2018). Semiparametric Bayesian analysis of transformation linear mixed models. *Journal of Multivariate Analysis* **166** 225–240. MR3799645
[18] WANG, Z. Q. and TANG, N. S. (2020). Bayesian Quantile Regression with Mixed Discrete and Nonignorable Missing Covariates. *Bayesian Analysis* **15** 579–604. MR4078726
[19] WEI, C. H., GUO, S. and ZHAI, S. F. (2017). Statistical inference of partially linear varying coefficient spatial autoregressive models. *Economic Modelling* **64** 553–559. MR3548001
[20] XIE, T. F., CAO, R. Y. and DU, J. (2020). Variable selection for spatial autoregressive models with a diverging number of parameters. *Statistical Papers* **61** 1125–1145. MR4096410
[21] XU, D. K. and ZHANG, Z. Z.(2013). A semiparametric Bayesian approach to joint mean and variance models. *Statistics & Probability Letters* **83** 1624–1631. MR3062273
[22] ZHANG, H. H., CHENG, G. and LIU, Y. (2011). Linear or nonlinear? Automatic structure discovery for partially linear models. *Journal of the American Statistical Association* **106(495)** 1099–1112. MR2894767

Ruiqin Tian
School of Science, Hangzhou Normal University, Hangzhou
People's Republic of China
E-mail address: ruiqintian@163.com

Dengke Xu
Department of Statistics, Zhejiang Agriculture and Forestry University, Hangzhou
People's Republic of China
E-mail address: xudengke1983@163.com

Jiang Du
School of Statistics and Data Science, Faculty of Science, Beijing University of Technology, Beijing
People's Republic of China
E-mail address: dujiang84@163.com

Junfei Zhang
School of Statistics and Mathematics, Central University of Finance and Economics, Beijing
People's Republic of China
E-mail address: junfeizhang_cufe@163.com