Journal of Econometrics and Statistics
Vol. 1, Issue 1, 2021, pp. 17-41

URL : www.arfjournals.com/jes

ESTIMATION AND SPECIFICATION TEST OF PARTIALLY LINEAR SINGLE-INDEX SPATIAL AUTOREGRESSIVE MODEL

Tizheng Li[1]
[1]*Department of Mathematics, School of Science, Xi'an University of Architecture and Technology, Xi'an 710055, China. Email: tizhengli@xauat.edu.cn*

Article History

Received : 10 March 2021
Revised : 15 March 2021
Accepted : 31 March 2021
Published : 2 September 2021

To cite this paper

Li, T. (2021). Estimation and Specification test of Partially Linear Single-Index Spatial Autoregressive Model. *Journal of Econometrics and Statistics*. 1(1), 17-41.

Abstract: The partially linear single-index spatial autoregressive model is a new class of semiparametric spatial autoregressive models, which achieves both dimension reduction and nice model interpretation. In this paper, we propose a new estimation method for the partially linear single-index spatial autoregressive model by combining local linear smoothing approach and quasi-maximum likelihood method. Compared to existing estimation method, the proposed method does not need to select instrumental variables. Furthermore, we propose a generalized likelihood ratio test to check the parametric form of the nonparametric component, in which a residual-based bootstrap procedure is suggested to calculate *p*-value of the proposed test. Some simulation studies are conducted to assess the performance of the proposed estimation and test methods and simulation results show that both methods perform well in finite samples. A real data example is provided to illustrate the proposed estimation and test methods.

Keywords: Spatial dependence; Single-index modeling; Quasi-maximum likelihood method; Local linear smoothing method; Bootstrap.

1. Introduction

In recent years, spatial dependence among cross-sectional units has become a standard notion of economic research activities in relation to crime rates, social interaction, economic growth, spillover effects, peer effects, price competition, tax competition, house prices, land prices, etc., and has received an increasing attention by theoretical econometricians and applied researchers. Among various models characterizing spatial dependence, the most popular one is perhaps spatial autoregressive models, in which outcome of a spatial unit is allowed to depend on a weighted average of outcomes of its neighboring units and the values of the explanatory variables. Linear and nonparametric spatial autoregressive models are two important classes of spatial autoregressive models and both have their unique advantages. The linear spatial autoregressive model is simple,

easy to estimate and interpret, and can afford most efficient statistical inference if the linear assumption is valid. The nonparametric spatial autoregressive model makes no assumption on the form of the regression function and lets the data determine a functional form tailored to the data, hence it carries no risk of model misspecification and can afford maximal flexibility and adaptability. Partially linear spatial autoregressive model, a class of models between the linear and nonparametric spatial autoregressive models, inherits advantages from both sides by allowing the response variable to depend on its spatial lag and some of the explanatory variables in a linear way and nonlinearly relate to the remaining explanatory variables. Since the introduction in Su and Jin (2010), the partially linear spatial autoregressive model has gained considerable attention in recent years. For example, Su and Jin (2010) developed a profile quasi-maximum likelihood method for partially linear spatial autoregressive model, and studied the asymptotic properties of the resulting estimators. However, the estimation method proposed by Su and Jin (2010) requires the error terms to be homoscedastic, which is rather restrictive in some empirical applications. To take the heteroscedasticity of the error term into account, Zhang (2013) and Zhang and Yang (2015a) proposed the pairwise difference estimation method and the instrumental variable estimation method for the partially linear spatial autoregressive model, respectively. Li and Mei (2013,2016) studied related test problems in the partially linear spatial autoregressive model such as whether the nonparametric component poses some interesting parametric forms and whether the parameters in the parametric component are significant or more generally satisfy certain linear constraint conditions. Recently, some researchers (Zhang and Sun, 2015; Zhang and Yang, 2015b; Ai and Zhang, 2017) extended the partially linear spatial autoregressive model from cross-section data to panel data and studied related estimation problems. More recently, Li and Guo (2020) considered the problem of variable selection in the partially linear spatial autoregressive model. They proposed a class of penalized likelihood method to simultaneously select significant explanatory variables in the parametric component and estimate the corresponding nonzero parameters, and studied asymptotic properties of the resulting penalized estimator.

However, as far as the model structure is concerned, the partially linear spatial autoregressive model still has the following two drawbacks. First, when the number of the explanatory variables in its nonparametric component is large, the partially linear spatial autoregressive model still suffers from the same drawbacks as the nonparametric spatial autoregressive model such as the "curse of dimensionality", the difficulty of interpretation and the lack of extrapolation capability. Second, the partially linear spatial autoregressive model requires the explanatory variables in its nonparametric component are all continuous, which is rather stringent in practical applications.

To avoid the above mentioned two drawbacks of the partially linear spatial autoregressive model, Sun and Wu (2018) and Cheng *et al.* (2019) independently proposed partially linear single-index spatial autoregressive model, in which the response variable linearly depends on its spatially lagged term and some of the explanatory variables but nonlinearly depends on a linear combination of the remaining explanatory variables. Specifically, let (X_i, Z_i, Y_i) be the observation collected from the i^{th} spatial unit ($i = 1, ..., n$), where X_i and Z_i are, respectively, the $p \times 1$ and $q \times 1$ vector of

exogenous explanatory variables, and Y_i is the response variable of interest. Then the sample form of the proposed partially linear single-index spatial autoregressive model is

$$Y_i = \rho \sum_{j \neq i} w_{ij} Y_j + X_i^T \beta + \eta(Z_i^T \alpha) + \varepsilon_i, \ i = 1, ..., n, \tag{1}$$

where $w_{ij}(i, j = 1, ..., n; i \neq j)$ are pre-specified exogenous spatial weights that determine the structure of neighborhood among spatial units, ρ is the spatial autoregressive parameter that measures the intensity of spatial correlation among the observations of the response variable, β is the vector of regression coefficients, α is the vector of index parameters, $\eta(\cdot)$ is the unknown link function, and $\varepsilon_i(i = 1, ..., n)$ are the independent and identically distributed error terms with mean zero and finite variance σ^2. To make model (1) identifiable, we assume that both X and Z do not contain constant term and at least one component of Z is continuous, the link function $\eta(\cdot)$ is differentiable and not constant on the support of $Z^T\alpha$, and the index parameter vector α satisfies $\|\alpha\| = 1$ and its first element is positive, where $\|\cdot\|$ denotes the Euclidean norm.

From the viewpoint of statistical modeling, the partially linear single-index spatial autoregressive model is of the following three attractive advantages. First, by introducing index term $Z^T\alpha$, partially linear single-index spatial autoregressive model not only avoids the "curse of dimensionality" since only one-dimensional nonparametric smoothing is involved regardless of the dimension of Z, but also has a nice interpretation with the impact of Z on Y being described by the finite-dimensional parameter vector and the univariate function $\eta(\cdot)$. Second, different from the partially linear spatial autoregressive model, the partially linear single-index spatial autoregressive model allows the discrete explanatory variables to appear in the nonparametric component. Third, through conducting a hypothesis test or variable selection procedure on the index parameter vector α, one can identify which explanatory variables in the nonparametric component have significant effect on the response variable. However, it is quite difficult to determine which explanatory variables in the nonparametric component are significant in the partially linear spatial autoregressive model.

To estimate parameter vector $(\alpha^T, \beta^T, \rho)^T$ and link function $\eta(\cdot)$ in model (1), Sun and Wu (2018) and Cheng *et al.* (2019) independently developed semiparametric generalized method of moment (GMM) estimation method based on local linear smoothing method and generalized method of moment, and studied asymptotic properties of the resulting estimators. The only difference between Sun and Wu (2018) and Cheng *et al.* (2019) lies in the treatment of the error terms of model (1). In Sun and Wu (2018), the error terms of model (1) are allowed to be heteroscedastic, while they are assumed to be independent and identically distributed in Cheng *et al.* (2019). The semiparametric GMM estimation method needs to select the instrumental variables and the choice of instrumental variables may affect the finite sample performance of the method. More importantly, the optimal choice of the instrumental variables is a very difficult problem. Furthermore, both Sun and Wu (2018) and Cheng *et al.* (2019) did not consider the estimation of the error variance σ^2, which is a vital parameter in model (1) because it measures the intensity of influence of random factors or some missing explanatory variables on the response variable.

In this paper, we develop a new estimation method for model (1) by combining local linear smoothing method and quasi-maximum likelihood method. To be specific, we first treat the spatial autoregressive parameter, index parameter vector and regression coefficient vector as if they were known, and use the local linear smoothing method to estimate the link function $\eta(\cdot)$. Then the quasi-maximum likelihood method is used to estimate the parameter vector $\theta = (\alpha^T, \beta^T, \rho, \sigma^2)$. Given the estimate of θ, the final estimate of $\eta(\cdot)$ can be obtained. Compared to the semiparametric GMM estimation method of Sun and Wu (2018) and Cheng *et al.* (2019), the great advantage of our method is that there is no need to select the instrumental variables. Thus, our estimation method may have better finite sample performance than the semiparametric GMM estimation method of Sun and Wu (2018) and Cheng *et al.* (2019), which is empirically verified by a simulation study in Section 5. Furthermore, our method can estimate the error variance σ^2. However, our estimation method also has the following two drawbacks. First, although our method does not require the error terms to follow normal distribution, it require the error terms to be homoscedastic, which is rather restrictive in some empirical applications. Second, it may be quite difficult to extend our method to model (1) with multiple spatial weight matrices. In principle, we can formulate a profile quasi log-likelihood function for model (1) with multiple spatial weight matrices. Nevertheless, during the search of the profile quasi-maximum likelihood estimator, we need to focus on the parameter space and evaluation of the determinant of the Jacobian transformation. For model (1), the parameter space in many circumstances can be taken to be $(-1, 1)$. But, the parameter space becomes rather complicated for model (1) with multiple spatial weight matrices. Even if the error terms of model (1) with multiple spatial weight matrices are normally distributed, the profile quasi-maximum likelihood method would be hard to implement as the determinant of the Jacobian transformation becomes more complicated than that of model (1). However, the semiparametric GMM estimation method of Sun and Wu (2018) and Cheng *et al.* (2019) can be easily extended to model (1) with multiple spatial weight matrices.

After fitting the partially linear single-index spatial autoregressive model (1), one of the important inferential problems is to check whether some interesting parametric forms are appropriate to the nonparametric component (namely, if a parametric spatial autoregressive model is adequate). This problem is vital important since a parametric spatial autoregressive model, which is powerful in explanation and easy to be fitted, is more preferred unless a partially linear single-index spatial autoregressive model is necessary for a given spatial data-set. To address this issue, we construct a test statistic based on the difference of the maximal profile quasi log-likelihood under the alternative model and the maximal quasi log-likelihood under the null model. Some simulation studies are conducted to assess the performance of the proposed estimation and test methods and the simulation results show that both methods perform well in finite samples. The Boston housing price data are analyzed to illustrate the application of the proposed estimation and test methods.

The rest of this paper is organized as follows. In Section 2, we introduce the proposed estimation method in detail. In Section 3, we discuss some issues related to the practical implementation of the proposed estimation method. A generalized likelihood ratio test statistic is constructed in Section 4

to check the parametric form of the link function $\eta(\cdot)$, in which a residual-based bootstrap procedure is provided to approximate the null distribution of the resulting test statistic. Some simulation studies are conducted to evaluate the finite sample performance of the proposed estimation and test methods in Section 5. In Section 6, a real data example is given to demonstrate the application of the proposed estimation and test methods. The paper is then concluded with some remarks in Section 7.

2. Estimation Method

Let $w_{ii} = 0 (i = 1, ..., n)$, $w = (w_{ij})$, $Y = (Y_1, ..., Y_n)^T$, $X = (X_1, ..., X_n)^T$, $Z = (Z_1, ..., Z_n)^T$, $\eta(Z\alpha) = (\eta(Z_1^T\alpha),...\eta(Z_n^T\alpha))^T$, and $\varepsilon = (\varepsilon_1, ..., \varepsilon_n)^T$. Then model (1) can be expressed in matrix form as

$$Y = \rho WY + X\beta + \eta(Z\alpha) + \varepsilon. \tag{2}$$

Let $T(\rho) = I - \rho W$ and $\varepsilon(\delta) = Y - \rho WY - X\beta - \eta(Z\alpha)$, where I be an identity matrix of order n and $\delta = (\alpha^T, \beta^T, \rho)^T$. Then, the Gaussian quasi log-likelihood function of model (2) is

$$\log L(\theta, \eta(\cdot)) = -\frac{n}{2}\log(2\pi\sigma^2) + \log(|T(\rho)|) - \frac{1}{2\sigma^2}\varepsilon(\delta)^T\varepsilon(\delta). \tag{3}$$

Since the unknown function $\eta(\cdot)$ is present in Equation (3), we propose estimating the finite-dimensional parameter vector θ by the following two-stage procedure:

(i) Estimate $\eta(\cdot)$ for fixed θ and denote the resulting estimator as $\eta_\theta(\cdot)$;

(ii) Plug in $\eta_\theta(\cdot)$ into $\varepsilon(\delta)$ in (3) and obtain the estimator $\hat{\theta}$ of by using the quasi-maximum likelihood method, and finally obtain the estimator $\eta_{\hat{\theta}}(\cdot)$ of $\eta(\cdot)$.

To estimate $\eta(\cdot)$ for fixed θ in the first stage, we employ the local linear smoothing method although other nonparametric smoothing methods such as the Nadaraya-Watson kernel method and the spline methods are applicable. The main reason for preferring the local linear smoothing method is because it possesses many attractive properties such as high statistical efficiency in an asymptotic minimax sense, design adaptation, and automatic boundary corrections (for details see Fan and Gijbels, 1996).

Assume that the link function $\eta(\cdot)$ has continuous second order derivative. Then for any given u in the domain of the index term $U = Z^T\alpha$, it follows from the Taylor's expansion that

$$\eta(v) \approx \eta(u) + \eta(u)(v - u)$$

for any v in a neighborhood of u. The local linear smoothing method finds $\eta(u)$ and $\eta'(u)$ by minimizing the following locally weighted least squares function

$$\sum_{i=1}^{n}\left[Y_i - \rho\sum_{j=1}^{n} w_{ij}Y_j - X_i^{\mathrm{T}}\beta - \eta(u) - \eta'(u)(U_i - u)\right]^2 K_h(U_i - u), \tag{4}$$

where $U_i = Z_i^T\alpha$, and $K_h(\cdot) = K(\cdot/h)/h$ with $K(\cdot)$, being a kernel function defined on $\Re$ and h being a bandwidth.

Let $\Psi(u) = (\eta(u), \eta'(u))^T$, $W(u, \alpha) = \text{diag}(K_h(U_1 - u), ..., K_h(U_n - u))^T$, and $Z(u, \alpha) = \begin{pmatrix} 1 & \cdots & 1 \\ U_1 - u & \cdots & U_n - u \end{pmatrix}^T$.

Given θ, the solution of the weighted least squares problem (4), that is, the local linear estimator of $\Psi(u)$, can be expressed as

$$\Psi_\theta(u) = S(u, \alpha)\,[T(\rho)Y - X\beta],$$

where $S(u, \alpha) = [Z(u, \alpha)^T W(u, \alpha) Z(u, \alpha)]^{-1} Z(u, \alpha)^T W(u, \alpha)$.

In particular, the local linear estimator of the link function $\eta(u)$ is given by

$$\eta_\theta(u) = s(u, \alpha)[T(\rho)Y - X\beta], \tag{5}$$

where $s(u, \alpha) = e^T S(u, \alpha)$ with $e = (1, 0)^T$.

With $\eta(\cdot)$ in (3) being replaced by $\eta_\theta(u)$, we obtain the following profile quasi log-likelihood function

$$\log L(\theta) = -\frac{n}{2}\log(2\pi\sigma^2) + \log(|T(\rho)|) - \frac{1}{2\sigma^2} \times$$

$$[T(\rho)Y - X\beta - \eta_\theta(Z\alpha)]^T [T(\rho)Y - X\beta - \eta_\theta(Z\alpha)], \tag{6}$$

where $\eta_\theta(Z\alpha) = (\eta_\theta(Z_1^T\alpha), ..., \eta_\theta(Z_n^T\alpha))^T$.

Maximizing $\log L(\theta)$ under the constraint conditions $\alpha^T\alpha = 1$, $\sigma^2 > 0$ and $-1 < \rho < 1$ yields the profile quasi-maximum likelihood estimator of θ as $\hat{\theta}$. Then the final local linear estimator $\hat{\eta}(u)$ of $\eta(u)$ is taken as $\eta_\theta(u)$ with θ being replaced by $\hat{\theta}$. As a result, the residual vector is

$$\hat{\varepsilon} = (\hat{\varepsilon}_1, \cdots, \hat{\varepsilon}_n)^T = Y - \hat{\rho}WY - X\hat{\beta} - \eta_{\hat{\theta}}(Z\hat{\alpha}). \tag{7}$$

3. Implementation of Estimation Method

3.1. An Iterative Algorithm

Since it is difficult to directly maximize the profile quasi log-likelihood function $\log L(\theta)$, we propose an iterative algorithm to obtain the profile quasi-maximum likelihood estimator $\hat{\theta}$ of θ.

Step 1. Initialize $\theta^{(0)} = (\alpha^{(0)}, \beta^{(0)}, \rho^{(0)}, \sigma^{2(0)})$.

Step 2. Update $\sigma^{2(m+1)} = \arg\max_{\sigma^2 \in (0, +\infty)} \log L(\alpha^{(m)}, \beta^{(m)}, \rho^{(m)}, \sigma^2)$.

Step 3. Update $\rho^{(m+1)} = \arg\max_{\rho \in (-1, 1)} \log L(\alpha^{(m)}, \beta^{(m)}, \rho, \sigma^{2(m+1)})$.

Step 4. Update $(\alpha^{(m+1)}, \beta^{(m+1)}) = \arg\max_{(\alpha, \beta) \in \mathbb{R}^{p+q}} \log L(\alpha, \beta, \rho^{(m+1)}, \sigma^{2(m+1)})$.

Step 5. Update Steps 2-4 until convergence and denote the final estimator of $(\alpha, \beta, \rho, \sigma^2)$ as $(\hat{\alpha}, \hat{\beta}, \hat{\rho}, \hat{\sigma}^2)$, then $\hat{\theta} = (\hat{\alpha}^T, \hat{\beta}^T, \hat{\rho}, \hat{\sigma}^2)^T$.

Remark 1. In Step 1, two methods can be used to obtain the initial estimator $\theta^{(0)}$ of θ. For example, the initial estimator $\theta^{(0)}$ can be obtained by fixing $\rho = 0$ and fitting a partially linear single-index model $Y = X\beta + \eta(Z\alpha) + \varepsilon$. Alternatively, the initial estimator $\theta^{(0)}$ can also be obtained by fitting a linear spatial autoregressive model $Y = \rho WY + X\beta + Z\alpha + \varepsilon$ by the quasi-maximum likelihood method. In Steps 2 and 3, both are one-dimensional nonlinear optimization problems which can be solved by using, for example, the function **fminbnd** in the toolbox **optimization** of the computer software **Matlab**. In Step 4, updating $(\alpha^{(m+1)}, \beta^{(m+1)})$ is equivalent to fitting the following partially linear single-index model

$$Y^* = X\beta + \eta(Z\alpha) + \varepsilon, \tag{8}$$

where $Y^* = Y - \rho^{(m+1)}WY$. There are several estimation methods available in the literature to fit model (8) such as the back-fitting method of Carroll *et al.* (1997), the penalized spline estimation method of Yu and Ruppert (2002) the minimum average variance estimation method of Xia and Härdle (2006), and the profile least squares method of Liang *et al.* (2010). Here we employ the profile least squares method to estimate (α, β) in model (8). The reason for such a choice is that the estimator of (α, β) obtained by the profile least squares method is semiparametrically efficient (for details, see Liang *et al.*, 2010).

3.2. Bandwidth Selection

With estimated $\hat{\theta}$, we obtain an approximated nonparametric regression model

$$Y_i - \hat{\rho}\sum_{j=1}^{n} w_{ij}Y_j - X_i^{\mathrm{T}}\hat{\beta} \approx \eta\left(\hat{U}_i\right) + \varepsilon_i, \quad i = 1, \cdots, n, \tag{9}$$

where $\hat{U}_i = Z_i^{T}\hat{\alpha}$. Hence, the value of the bandwidth h can be determined by some data-driven criteria such as the rule of thumb (ROT), the cross validation (CV), the generalized cross validation (GCV) and the corrected Akaika information criterion (AIC_c). To reduce the heavy computational burden, we employ the computationally simple rule of thumb (ROT) method to determine the value of h, that is, $h = s_U n^{-1/5}$, where s_U is the sample standard deviation of $\hat{U}_1, \cdots, \hat{U}_n$.

4. Testing for Parametric form of Link Function

4.1. The Hypotheses

The nonparametric estimate of the link function $\eta(\cdot)$ can provide us with descriptive and graphical information for exploratory data analysis. Using this information, it is possible to formulate a parametric spatial autoregressive model that takes into account the features that emerged from the preliminary analysis. To this end, we introduce a goodness-of-fit test to assess appropriateness of a parametric spatial autoregressive model. Without loss of generality, we consider a simple linear spatial autoregressive model under the null hypothesis. Accordingly, the null and alternative hypotheses can be described as follows:

$$H_0 : \eta(u) = \gamma_0 + \gamma_1 u \text{ for all } u \leftrightarrow H_1 : \eta(u) \neq \gamma_0 + \gamma_1 u \text{ for some } u, \tag{10}$$

where γ_0 and γ_1 are two unknown constant parameters.

4.2. Construction of Test Statistic

Firstly, under the alternative hypothesis H_1, we fit the partially linear single-index spatial autoregressive model (1) by the estimation method proposed in Section 2 and obtain the maximal profile quasi log-likelihood as

$$l(H_1) = -\frac{n}{2}\left[\log(2\pi)+1\right] - \frac{n}{2}\log\left(n^{-1}\mathrm{RSS}_1\right) + \log\left(\left|T(\hat{\rho})\right|\right), \tag{11}$$

where $\mathrm{RSS}_1 = \hat{\varepsilon}^T\hat{\varepsilon}$.

Secondly, under the null hypothesis H_0, model (1) becomes

$$Y_i = \rho\sum_{j\neq i} w_{ij}Y_j + X_i^T\beta + \gamma_0 + Z_i^T(\gamma_1\alpha) + \varepsilon_i, \quad i = 1,\cdots,n. \tag{12}$$

Let $\bar{\beta} = (\beta^T, \gamma_0, \gamma_1\alpha^T)^T$ and $\bar{X} = (X, 1, Z)$ with 1 is an $n \times 1$ vector with all of its elements being 1. Then, model (12) can be further written as

$$Y = \rho WY + \bar{X}\bar{\beta} + \varepsilon. \tag{13}$$

Model (13) is a standard linear spatial autoregressive model, and the quasi-maximum likelihood method can be used to fit this model. The Gaussian quasi log-likelihood function of model (13) is

$$\log L(\bar{\theta}) = -\frac{n}{2}\log(2\pi\sigma^2) + \log\left(\left|T(\rho)\right|\right) - \frac{1}{2\sigma^2}\times\left[T(\rho)Y - \bar{X}\bar{\beta}\right]^T\left[T(\rho)Y - \bar{X}\bar{\beta}\right], \tag{14}$$

where $\bar{\theta} = \left(\bar{\beta}^T, \rho, \sigma^2\right)^T$.

Given ρ, $\log L(\bar{\theta})$ can be partially maximized, which yields quasi-maximum likelihood estimators of $\bar{\beta}$ and σ^2, respectively, as

$$\tilde{\bar{\beta}}(\rho) = (\bar{X}^T\bar{X})^{-1}\bar{X}^T T(\rho)Y \tag{15}$$

and

$$\tilde{\sigma}^2(\rho) = Y^T T(\rho)^T M_0 T(\rho)Y, \tag{16}$$

where $M_0 = I - \bar{X}\left(\bar{X}^T\bar{X}\right)^{-1}\bar{X}^T$. Substituting $\tilde{\bar{\beta}}(\rho)$ and $\tilde{\sigma}^2(\rho)$ into (14) leads to the concentrated quasi log-likelihood function of ρ as

$$\log\tilde{L}(\rho) = -\frac{n}{2}\left[\log(2\pi)+1\right] - \frac{n}{2}\log\left(\tilde{\sigma}^2(\rho)\right) + \log\left(\left|T(\rho)\right|\right). \tag{17}$$

Maximizing $\log \tilde{L}(\rho)$ subject to the constraint condition $-1 < \rho < 1$ gives the quasi-maximum likelihood estimator $\tilde{\rho}$ of ρ. Substituting $\tilde{\rho}$ into $\tilde{\tilde{\beta}}(\rho)$ and $\tilde{\sigma}^2(\rho)$ yields the final estimator $\tilde{\tilde{\beta}} \equiv \tilde{\tilde{\beta}}(\tilde{\rho})$ of $\bar{\beta}$, the estimator $\tilde{\sigma}^2 \equiv \tilde{\sigma}^2(\tilde{\rho})$ of σ^2 and, consequently, the estimator $\tilde{\bar{\theta}} = (\tilde{\tilde{\beta}}^T, \tilde{\rho}, \tilde{\sigma}^2)^T$ of $\bar{\theta}$. Therefore, the maximal quasi log-likelihood under H_0 can be expressed as

$$l(H_0) = -\frac{n}{2}\left[\log(2\pi)+1\right] - \frac{n}{2}\log\left(n^{-1}\mathrm{RSS}_0\right) + \log\left(\left|T(\tilde{\rho})\right|\right), \tag{18}$$

where $\mathrm{RSS}_0 = \left[T(\tilde{\rho})Y - \bar{X}\tilde{\tilde{\beta}}\right]^{\mathrm{T}}\left[T(\tilde{\rho})Y - \bar{X}\tilde{\tilde{\beta}}\right]$ is the residual sum of squares under H_0 and $\tilde{\sigma}^2 = n^{-1}\mathrm{RSS}_0$.

Based on $l(H_1)$ and $l(H_0)$, a generalized likelihood ratio statistic is constructed as

$$T = l(H_1) - l(H_0) = \frac{n}{2}\log\left(\frac{\mathrm{RSS}_0}{\mathrm{RSS}_1}\right) + \log\left(\frac{\left|T(\hat{\rho})\right|}{\left|T(\tilde{\rho})\right|}\right). \tag{19}$$

Intuitively, the null hypothesis H_0 should be rejected if the value of T is large enough. Therefore, the p-value of the test is

$$p_0 = P_{H_0}(T \geq t), \tag{20}$$

where $P_{H_0}(\cdot)$ refers to the probability computed under the null hypothesis H_0 and t is the observation of T. For a given significance level α, if $p_0 < \alpha$, reject H_0; otherwise not reject H_0.

Remark 2. Although the test statistic T is derived for linear form of null hypothesis, our test method can be done for other more general forms of null hypothesis like $\eta(u) = f(u, \gamma)$, where $f(u, \gamma)$ is a function whose form is completely known but with an unknown parameter vector γ. In this case, under the null hypothesis, model (1) becomes a nonlinear spatial autoregressive model

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

It is very difficult to fit this model by the quasi-maximum likelihood method because of its complexity. To overcome this difficulty, under the null hypothesis H_0, we use the same parametric estimators $\hat{\alpha}$, $\hat{\beta}$, and $\hat{\rho}$ as those obtained under the alternative hypothesis H_1, and obtain an estimator of γ by solving the following nonlinear least squares function

$$S(\gamma) = \sum_{i=1}^{n}\left[Y_i - \hat{\rho}\sum_{j=1}^{n} w_{ij}Y_j - X_i^T\hat{\beta} - f\left(Z_i^T\hat{\alpha}, \gamma\right)\right]^2$$

Then, the resulting residual sum of squares under the null and alternative hypotheses are

$$\mathrm{RSS}_0 = \sum_{i=1}^{n}\left[Y_i - \hat{\rho}\sum_{j=1}^{n} w_{ij}Y_j - X_i^T\hat{\beta} - f\left(Z_i^T\hat{\alpha}, \hat{\gamma}\right)\right]^2$$

and

$$\mathrm{RSS}_1=\sum_{i=1}^{n}\left[Y_i-\hat{\rho}\sum_{j=1}^{n}w_{ij}Y_j-X_i^T\hat{\beta}-\eta_{\hat{\theta}}\left(Z_i^T\hat{\alpha}\right)\right]^2$$

Thus, for null hypothesis like $\eta(u) = f(u, \gamma)$, the test statistic T becomes

$$T=\frac{n}{2}\frac{\mathrm{RSS}_0-\mathrm{RSS}_1}{\mathrm{RSS}_1}.$$

4.3. Calculation of the p-value

To calculate the *p*-value of the proposed test, one of the commonly used methods is to derive the asymptotic null distribution of the test statistic T. However, the presence of the spatially lagged term of the response variable in the model makes the derivation of the asymptotic null distribution of the test statistic T very difficult. On the other hand, even if one can derive the asymptotic null distribution of T, as pointed out by many researchers (Hall and Hart, 1990; Härdle and Mammen, 1993; Fan and Jiang, 2007), *p*-value computed by the asymptotic null distribution of the test statistic may be invalid under the situation of finite sample sizes. Therefore, we propose a bootstrap procedure to approximate the null distribution of the test statistic T.

Among the existing bootstrap sampling schemes, the residual-based bootstrap procedure has been extensively used to approximate the null distribution of related test statistics in the literature of the nonparametric and semi-parametric regression (Stute *et al.*, 1998; Cai *et al.*, 2000; Fan and Huang, 2005; Fan and Jiang, 2005). Moreover, as pointed out by Anselin (1988), it is crucial that spatial structure must be preserved during data resampling in models with spatial dependence, and particularly with a spatially lagged term of the response variable. Thus, we employ the residual-based bootstrap procedure to approximate the null distribution of the test statistic T. In our case, the procedure can be described as follows.

Step 1. Based on the data set $\{Y, X, Z\}$ and a predetermined value of the bandwidth h, compute under H_1 the residual vector $\hat{\varepsilon}$ shown in (7) and centralize it to obtain $\hat{\varepsilon}_c=\left(\hat{\varepsilon}_1-\bar{\hat{\varepsilon}},\cdots,\hat{\varepsilon}_n-\bar{\hat{\varepsilon}}\right)^T$ in which $\bar{\hat{\varepsilon}}=\frac{1}{n}\sum_{i=1}^{n}\hat{\varepsilon}_i$. Furthermore, compute under H_0 the estimators $\tilde{\tilde{\beta}}$ and $\hat{\rho}$. With the estimation results under H_0 and H_1, compute the observed value t of the test statistic T by (19).

Step 2. Draw a bootstrap sample $\varepsilon^* = (\varepsilon_1^*, ..., \varepsilon_n^*)^T$ with replacement from the empirical distribution function of $\hat{\varepsilon}_c$.

Step 3. Generate $Y^*=(I-\tilde{\rho}W)^{-1}(\bar{X}\tilde{\tilde{\beta}}+\varepsilon^*)$ and calculate the bootstrap version T^* of the test statistic T by

$$T^* = \frac{n}{2}\log\left(\frac{\mathrm{RSS}_0^*}{\mathrm{RSS}_1^*}\right) + \log\left(\frac{\left|T\left(\hat{\rho}^*\right)\right|}{\left|T\left(\tilde{\rho}^*\right)\right|}\right), \tag{21}$$

where RSS_0^* and RSS_1^* are, respectively, the residual sum of squares obtained under H_0 and H_1 based on the data set $\{Y^*, X, Z\}$, and $\tilde{\rho}^*$ and $\hat{\rho}^*$ are the estimators of ρ based on the data set $\{Y^*, X, Z\}$ under H_0 and H_1, respectively.

Step 4. Repeat steps 2 and 3 m times and obtain a bootstrap sample of the test statistic T as $T_1^*, \ldots T_m^*$. The p-value is then estimated by

$$\hat{p}_0 = \frac{\#\left\{T_i^* \mid T_i^* \geq t\right\}}{m}, \tag{22}$$

where #A denotes the number of the elements in a set A.

5. Simulation Studies

In this section, we investigate the finite sample performance of the proposed estimation and test methods through simulation studies. In both simulation studies and real data analysis in Section 6, we employ the Gaussian kernel function $K(u) = \frac{1}{\sqrt{2\pi}}\exp\left(-u^2/2\right)$ and the bandwidth selection procedure proposed in Section 3.2.

5.1. Spatial Layout and Design of Experiment

The spatial layout for simulation studies is taken as a square region with the length of each side being l units. This type of spatial layout is of wide application backgrounds in the field of remote sensing. The $l \times l$ lattice squares in the region, which leads to a sample size of $n = l^2$, are designed as the spatial units at which the observations of the response variable and the explanatory variables are made. These n spatial units are labeled by 1 to n with the order from left to right and from bottom to top.

Given the above spatial layout, the spatial weight matrix W is constructed based on the Rook contiguity and the exponential function of the distance between spatial units, respectively. For the Rook contiguity, the standardized spatial weight matrix W is generated as follows:

(i) Let $w_{ij} = 1$ if spatial unit j shares a common edge with spatial unit i and let $w_{ij} = 0$ otherwise;

(ii) divide each element w_{ij} by the corresponding row sum to form the standardized spatial weight matrix W. For the latter way, the element w_{ij} of the spatial weight matrix W is taken as $w_{ij} = \exp(-d_{ij}) \Big/ \sum_{k=1}^{n} \exp(-d_{ik})$, where d_{ij} is the Euclidean distance between spatial units i and j.

We generate 500 data sets, each consisting of $n = 49$ and $n = 100$ random observations, from the following partially linear single-index spatial autoregressive model

$$Y_i = \rho \sum_{j \neq i} w_{ij} Y_j + X_i^T \beta + \eta\left(Z_i^T \alpha\right) + \varepsilon_i, \quad i = 1, \cdots, n, \tag{23}$$

where $X_i = (X_{i1}, X_{i2}, X_{i3})^T$ (i = 1, ..., n) were randomly drawn from the normal distribution with zero mean vector and the covariance matrix $\begin{pmatrix} 1 & 0.5 & 0.5 \\ 0.5 & 1 & 0.5 \\ 0.5 & 0.5 & 1 \end{pmatrix}$, $Z_i = (Z_{i1}, Z_{i2}, Z_{i3})^T$ in which Z_{ij}(j = 1, 2, 3) are independent and uniformly distributed on interval (0, 1), α = (0.5774, 0,5774, 0.5774)^T, β = (0.5, 1.0, 1.5)^T, and $\eta(u) = \sin(2\pi u)$. The value of the spatial autoregressive parameter ρ was taken to be 0.2, 0.5 and 0.8, respectively, to see the impact of the intensity of the spatial dependence among the observations of the response variable on the performance of the proposed estimation and test methods.

In order to investigate the influence of the error distribution on the performance of the proposed estimation and test methods, we consider the following three types of error distribution whose scales are adjusted such that they all have mean zero and variance 0.25:

(I) Normal distribution $N(0, 0.25)$;

(II) Uniform distribution $U(-\sqrt{3}/2, \sqrt{3}/2)$;

(III) Transformed and centralized chi-square distribution $\frac{1}{8}\chi^2(8) - 1$, where $\chi^2(8)$ denotes the random variable of a chi-square distribution with 8 degrees of freedom.

5.2. Performance of the Proposed Estimation Method

Simulation results for the proposed estimator $\hat{\theta}$ of θ under the given three types of error distribution are, respectively, summarized in Tables 1-3, in which “Mean” and “SD” stand for the mean and the standard deviation of the 500 estimates of θ, respectively.

We summarize some empirical findings from Tables 1-3. First, we can see that the bias and SD for the estimator $\hat{\theta}$ of θ are fairly small for almost all cases and they decrease dramatically when the sample size n increases, which demonstrates that the proposed estimation method gives very accurate estimate of θ. Second, as the degree of complexity of the spatial weight matrix W increases, that is the proportion of nonzero elements in W, both the bias and SD for the estimator $\hat{\rho}$ of ρ increase significantly, whereas the bias and SD of other estimators α, β and σ^2 are little affected, similar phenomenon is also observed in linear spatial autoregressive models (Lee, 2004), partially linear spatial autoregressive models (Su and Jin, 2010) and varying coefficient spatial autoregressive models (Li and Chen, 2013). This is reasonable because the higher the proportion of nonzero elements in W is, the stronger the spatial dependence of the response variable will be, which makes it more difficult to estimate the spatial autoregressive parameter ρ. Third, it can be observed from Tables 1-3 that the simulation results for $\hat{\theta}$ under the three types of error distribution have no

Table 1: Simulation results for θ under the error distribution N(0, 0.25).

W	*n*	ρ	*Index*	α_1	α_2	α_3	β_1	β_2	β_3	ρ	σ^2
Rook	49	0.2	Mean	0.5743	0.576	0.5737	0.4998	0.9897	1.5031	0.1960	0.2138
			SD	0.0546	0.058	0.0542	0.1052	0.1050	0.1060	0.0569	0.0439
		0.5	Mean	0.5739	0.5674	0.5681	0.4928	1.0007	1.5068	0.4905	0.2154
			SD	0.0674	0.1207	0.0833	0.1125	0.1085	0.1086	0.0519	0.0471
		0.8	Mean	0.5673	0.5599	0.5543	0.4984	1.0090	1.4959	0.7903	0.2187
			SD	0.0761	0.1372	0.1816	0.1036	0.1192	0.1079	0.0360	0.0591
	100	0.2	Mean	0.5728	0.5795	0.5769	0.5014	1.0017	1.4987	0.2014	0.2337
			SD	0.0336	0.0327	0.0340	0.0676	0.0642	0.0708	0.0364	0.0330
		0.5	Mean	0.5761	0.5739	0.5789	0.4983	1.0012	1.4988	0.4959	0.2320
			SD	0.0356	0.0340	0.0355	0.0699	0.0694	0.0659	0.0330	0.0320
		0.8	Mean	0.5781	0.5722	0.5717	0.4950	1.0048	1.4975	0.7972	0.2364
			SD	0.0416	0.0619 0	.0776	0.0677	0.0655	0.0670	0.0199	0.0399
EXP	49	0.2	Mean	0.5691	0.5754	0.5793	0.4893	1.0065	1.4935	0.1485	0.2141
			SD	0.0575	0.0562	0.0553	0.1050	0.1044	0.0996	0.1557	0.0455
		0.5	Mean	0.5711	0.5756	0.5774	0.4997	0.9987	1.5008	0.4594	0.2128
			SD	0.0553	0.0549	0.0550	0.1107	0.1023	0.1025	0.1369	0.0452
		0.8	Mean	0.5749	0.5733	0.5665	0.5002	0.9963	1.4916	0.7346	0.2166
			SD	0.0650	0.0719	0.1026	0.1053	0.1007	0.1014	0.1117	0.0507
	100	0.2	Mean	0.5776	0.5751	0.5763	0.4974	1.0083	1.4982	0.1822	0.2345
			SD	0.0339	0.0349	0.0334	0.0656	0.0694	0.0687	0.0869	0.0346
		0.5	Mean	0.5737	0.5752	0.5797	0.5013	0.9973	1.4980	0.4832	0.2330
			SD	0.0355	0.0367	0.0364	0.0651	0.0640	0.0632	0.0742	0.0333
		0.8	Mean	0.5760	0.5775	0.5752	0.5050	1.0037	1.4954	0.7802	0.2343
			SD	0.0349	0.0355	0.0378	0.0660	0.0683	0.0667	0.0535	0.0324

evident difference, which shows that the performance of the proposed estimator $\hat{\theta}$ of θ is quite robust to the variation of the error distribution.

The finite sample performance of the estimator $\hat{\eta}(\cdot)$ of $\eta(\cdot)$ is evaluated by using the mean square error (MSE) which is defined as

$$\mathrm{MSE}\left(\hat{\eta}(\cdot)\right) = \frac{1}{n_0}\sum_{k=1}^{n_0}\left[\hat{\eta}\left(u_k\right) - \eta\left(u_k\right)\right]^2,$$

where u_k (k = 1, ..., n_0) are some grid points that lie between the minimum value and maximum value of $\left\{U_i = Z_i^T\hat{\alpha},\ i = 1,\cdots,n\right\}$. In our simulation, we took n_0 = 100. Simulation results for the proposed estimator under the given three types of error distribution are reported in Table 4.

Table 2: Simulation results for under the error distribution $U(-\sqrt{3}/2, \sqrt{3}/2)$.

W	n	ρ	*Index*	α_1	α_2	α_3	β_1	β_2	β_3	ρ	σ^2
Rook	49	0.2	Mean	0.5697	0.5775	0.5768	0.5019	0.9979	1.4987	0.1970	0.2125
			SD	0.0588	0.0548	0.0524	0.1025	0.1081	0.1005	0.0613	0.0349
		0.5	Mean	0.5770	0.5722	0.5740	0.4902	1.0016	1.5055	0.4938	0.2130
			SD	0.0567	0.0610	0.0575	0.1078	0.1073	0.1075	0.0494	0.0333
		0.8	Mean	0.5713	0.5574	0.5522	0.5042	1.0037	1.4935	0.7926	0.2213
			SD	0.0652	0.1500	0.1772	0.1085	0.1052	0.1090	0.0338	0.0516
	100	0.2	Mean	0.5774	0.5781	0.5734	0.4981	0.9989	1.5013	0.1990	0.2323
			SD	0.0340	0.0359	0.0347	0.0658	0.0688	0.0703	0.0387	0.0230
		0.5	Mean	0.5767	0.5752	0.5768	0.5074	0.9931	1.5046	0.4944	0.2329
			SD	0.0356	0.0361	0.0347	0.0687	0.0670	0.0692	0.0315	0.0232
		0.8	Mean	0.5776	0.5688	0.5712	0.4981	0.9999	1.5030	0.7973	0.2369
			SD	0.0419	0.0936	0.0784	0.0705	0.0709	0.0737	0.0202	0.0311
EXP	49	0.2	Mean	0.5730	0.5776	0.5737	0.5020	0.9909	1.5007	0.1541	0.2141
			SD	0.0546	0.0534	0.0556	0.1012	0.1016	0.1041	0.1599	0.0321
		0.5	Mean	0.5747	0.5761	0.5728	0.5025	0.9953	1.4962	0.4616	0.2136
			SD	0.0587	0.0558	0.0570	0.1072	0.1073	0.1083	0.1257	0.0357
		0.8	Mean	0.5726	0.5751	0.5718	0.4968	1.0053	1.4978	0.7505	0.2143
			SD	0.0614	0.0852	0.0580	0.1092	0.1095	0.1132	0.1009	0.0358
	100	0.2	Mean	0.5789	0.5769	0.5730	0.5003	0.9974	1.4996	0.1865	0.2329
			SD	0.0366	0.0330	0.0368	0.0674	0.0693	0.0675	0.0909	0.0231
		0.5	Mean	0.5792	0.5755	0.5744	0.4962	1.0009	1.5015	0.4858	0.2348
			SD	0.0343	0.0327	0.0342	0.0704	0.0665	0.0646	0.0769	0.0240
		0.8	Mean	0.5771	0.5774	0.5742	0.4954	1.0011	1.4995	0.7797	0.2331
			SD	0.0341	0.0358	0.0369	0.0686	0.0747	0.0720	0.0519	0.0230

We can see from Table 4 that the MSE of the estimator $\hat{\eta}(\cdot)$ seems quite robust with respect to the variation of the error distribution, the spatial weight matrix and the spatial autoregressive parameter, and decreases remarkably as the sample size n increases.

5.3. Performance of the Proposed Test Method

In this subsection, we evaluate the finite sample performance of the proposed test method, including the validity of the bootstrap approximation to the null distribution of the test statistic and the power of the test. To this end, we took the link function $\eta(u)$ in (23) to be $\eta(u) = u + c\sin(2\pi u)$, where c

Table 3: Simulation results for θ under the error distribution $\frac{1}{8}\chi^2(8)-1$.

W	*n*	ρ	*Index*	α_1	α_2	α_3	β_1	β_2	β_3	ρ	σ^2
Rook	49	0.2	Mean	0.5739	0.5762	0.5732	0.5049	0.9974	1.4932	0.1889	0.2125
			SD	0.0597	0.0547	0.0585	0.1048	0.1031	0.1038	0.0568	0.0536
		0.5	Mean	0.5767	0.5654	0.5731	0.5019	0.9966	1.5009	0.4932	0.2116
			SD	0.0551	0.0977	0.0820	0.1037	0.1068	0.1039	0.0507	0.0546
		0.8	Mean	0.5723	0.5516	0.5439	0.5012	1.0040	1.4996	0.7914	0.2221
			SD	0.0772	0.1761	0.1885	0.1088	0.1063	0.1097	0.0348	0.0696
	100	0.2	Mean	0.5755	0.5768	0.5766	0.5035	1.0000	1.4969	0.1971	0.2355
			SD	0.0343	0.0354	0.0350	0.0769	0.0654	0.0638	0.0376	0.0429
		0.5	Mean	0.5744	0.5783	0.5761	0.5039	1.0008	1.4968	0.4970	0.2339
			SD	0.0351	0.0369	0.0344	0.0680	0.0676	0.0689	0.0294	0.0416
		0.8	Mean	0.5776	0.5777	0.5736	0.4975	1.0048	1.4992	0.7979	0.2341
			SD	0.0333	0.0352	0.0364	0.0694	0.0647	0.0674	0.0199	0.0415
EXP	49	0.2	Mean	0.5752	0.5720	0.5765	0.5005	1.0000	1.4935	0.1665	0.2103
			SD	0.0562	0.0560	0.0572	0.1048	0.1082	0.1110	0.1503	0.0548
		0.5	Mean	0.5685	0.5719	0.5819	0.4995	0.9971	1.4936	0.4555	0.2156
			SD	0.0607	0.0605	0.0619	0.1038	0.1100	0.1077	0.1386	0.0608
		0.8	Mean	0.5719	0.5747	0.5770	0.4985	1.0024	1.5027	0.7464	0.2155
			SD	0.0570	0.0565	0.0572	0.1068	0.1059	0.1028	0.0988	0.0587
	100	0.2	Mean	0.5782	0.5747	0.5757	0.5026	1.0004	1.4996	0.1835	0.2300
			SD	0.0346	0.0374	0.0361	0.0681	0.0648	0.0703	0.0932	0.0404
		0.5	Mean	0.5784	0.5782	0.5720	0.4992	1.0013	1.4977	0.4801	0.2317
			SD	0.0349	0.0355	0.0375	0.0704	0.0658	0.0708	0.0750	0.0430
		0.8	Mean	0.5781	0.5740	0.5768	0.5044	0.9980	1.4983	0.7795	0.2318
			SD	0.0350	0.0366	0.0341	0.0652	0.0693	0.0620	0.0548	0.0419

is such a constant that will take different values for different purposes. Note that the null hypothesis H_0 is true when $c = 0$ while the alternative hypothesis H_1 holds with $c \neq 0$.

In the simulation study performed here, we took the value of c in the link function $\eta(u)$ to be 0, 0.15, 0.30 and 0.45, respectively, to examine the validity of the bootstrap approximation to the null distribution of the test statistic T and the power of the test. The remainder of the experimental design was kept to be the same as that in Subsections 5.1 and 5.2 except that the spatial weight matrix was taken as the Rook and the spatial autoregressive parameter was taken to be 0.5. The reason why we only considered the case of Rook spatial weight matrix and $\rho = 0.5$ is that the involved computation is heavily huge and the simulation results under other cases are rather similar.

Table 4: MSE index for η(·) in model (23).

W	n	ρ	$N(0, 0.25)$	$U(-\sqrt{3}/2, \sqrt{3}/2)$	$\frac{1}{8}\chi^2(8)-1$
Rook	49	0.2	0.0645	0.0668	0.0697
		0.5	0.0740	0.0703	0.0685
		0.8	0.0898	0.0847	0.0891
	100	0.2	0.0418	0.0423	0.0427
		0.5	0.0427	0.0410	0.0427
		0.8	0.0441	0.0452	0.0413
Exp	49	0.2	0.0708	0.0737	0.0748
		0.5	0.0814	0.0757	0.0790
		0.8	0.1553	0.1174	0.1199
	100	0.2	0.0447	0.0432	0.0413
		0.5	0.0420	0.0466	0.0458
		0.8	0.0514	0.0466	0.0477

For each given value of c and each type of the error distribution, we run 200 replications of the test method and recorded the frequency of rejecting the null hypothesis under a given significance level α (0.01,0.05 and 0.10) as the empirical size of the test under H_0 (that is, $c = 0$) and the empirical power of the test under H_1 (that is, $c \neq 0$). And for each replication, the p-value in (22) was computed based on $m = 500$ bootstrap samples. The simulation results for the given three types of the error distribution are reported in Table 5.

We conclude some empirical findings from Table 5. First, in all of the experimental settings, the empirical sizes are all reasonably close to the corresponding significance levels α even for the very small sample size of $n = 49$. This demonstrates that the proposed bootstrap procedure yields an accurate approximation to the null distribution of the test statistic T at least on the right tail of the null distribution on which the p-value of the test is computed. Second, we can observe from the results that the empirical sizes have not evident difference for the three error distributions considered here, which shows that the bootstrap approximation to the null distribution of the test statistic T is quite robust to the variation of the error distribution. Third, the empirical power increases rapidly as the alternative hypothesis deviates away from the null hypothesis or the sample size n increases, which indicates that the proposed test method is powerful in identifying the linear form of the link function. Fourth, the empirical power is also quite robust with respect to the variation of the error distribution.

5.4. An Additional Simulation Study

According to comment 1 of the reviewer, we add a simulation study to compare the finite sample performance of the proposed estimation method with that of the semiparametric GMM estimation

Table 5: The rejection frequencies of testing for the linearity of link function η(·) in model (23).

		49			100		
Error distribution	*c*	*0.01*	*0.05*	*0.10*	*0.01*	*0.05*	*0.10*
$N(0, 0.25)$	0	0	0.045	0.080	0.010	0.065	0.120
	0.15	0.070	0.200	0.335	0.165	0.330	0.445
	0.30	0.240	0.515	0.670	0.720	0.880	0.935
	0.45	0.675	0.900	0.945	0.975	0.995	1.000
$U(-\sqrt{3}/2, \sqrt{3}/2)$	0	0.025	0.055	0.095	0.010	0.045	0.115
	0.15	0.050	0.185	0.240	0.105	0.275	0.400
	0.30	0.240	0.495	0.625	0.750	0.900	0.935
	0.45	0.630	0.880	0.915	0.985	0.995	0.995
$\frac{1}{8}\chi^2(8)-1$	0	0.015	0.065	0.140	0.012	0.035	0.125
	0.15	0.080	0.205	0.295	0.090	0.290	0.440
	0.30	0.375	0.555	0.660	0.710	0.875	0.925
	0.45	0.710	0.885	0.920	0.975	1.000	1.000

method which independently developed by Sun and Wu (2018) and Cheng *et al.* (2019). Following Sun and Wu (2018), we consider the following data generating process:

$$Y_i = 0.5\sum_{j\neq i} w_{ij}Y_j + 0.3X_i + \sin\left(\frac{\pi\left[\left(Z_{i1}+Z_{i2}+Z_{i3}\right)/\sqrt{3}-a\right]}{b-a}\right) + 0.3e_i, \quad i = 1,\cdots,n, \tag{24}$$

where $a = \sqrt{3}/2 - 1.645/\sqrt{12}$, $b = \sqrt{3}/2 + 1.645/\sqrt{12}$, $X_i (i = 1, ..., n)$ are drawn independently from binomial distribution $B(1, 0.5)$, $Z_{ij} (j = 1, 2, 3)$ are independently generated from uniform distribution $U(0, 1)$, $e_i (i = 1, ..., n)$ are drawn independently from standard normal distribution $N(0, 1)$. The spatial weights $w_{ij} (i, j = 1, ..., n)$ are specified based on the spatial scenario in Case (1991). To be specific, suppose there are R districts and each district has m members. Hence, the sample size is $n = Rm$. Moreover, each neighbour of a member in a district is given equal weight. In this case, the spatial weight matrix is $W = I_R \otimes B_m$, where $\otimes$ is the Kronecher product and $B_m = (1/(m-1))(1_m 1_m^T - I_m)$ with 1_m being a $m \times 1$ vector whose elements are all 1. Like that in Sun and Wu (2018), we consider three values of (R, m) : (10, 10), (20, 10) and (20, 20), which corresponds to the sample size 100, 200 and 400, respectively. Following Sun and Wu (2018), for each case, by repeating both estimation procedures 400 times, we get the means, biases, standard deviations (SDs) and mean squared errors (MSEs) for all parameter estimators. To evaluate the accuracy of the estimate of the link function, we consider mean integrated squared error (MISE) of $\hat{\eta}(\cdot)$ which is defined as:

$$\text{MISE} = E\left(\int\left[\hat{\eta}(u) - \eta(u)\right]^2 du\right).$$

We can see from Table 6 that, for the index parameters α_1, α_2 and α_3, the spatial autoregressive parameter ρ, and the link function $\eta(\cdot)$, the proposed estimation method gives more accurate estimate than the semiparametric GMM estimation method which independently developed by Sun and Wu (2018) and Cheng *et al.* (2019). One possible reason is that the instrumental variables use in Sun and Wu (2018) and Cheng *et al.* (2019) may be not optimal, which affect the finite sample performance of their method. However, for the regression coefficient β, the semiparametric GMM estimation method gives more accurate estimate than our method. Although this simulation study shows that our method slightly outperforms the semiparametric GMM estimation method of Sun and Wu (2018) and Cheng *et al.* (2019), overall comparison between two methods needs further investigation.

Table 6: Simulation results for profile quasi–maximum likelihood estimate (PQMLE) and semiparametric generalized method of moments estimate (SGMME)

		PQMLE				*SGMME*			
(R, m)	*Parameter*	*Mean*	*Bias*	*SD*	*MSE*	*Mean*	*Bias*	*SD*	*MSE*
(10, 10)	α_1	0.5743	–0.0030	0.0486	0.0024	0.5680	–0.0094	0.1344	0.0181
	α_2	0.5740	–0.0034	0.0491	0.0024	0.5533	–0.0241	0.1564	0.0250
	α_3	0.5776	0.0003	0.0482	0.0023	0.5556	–0.0218	0.1419	0.0205
	β	0.3853	0.0853	0.0620	0.0111	0.3018	0.0018	0.0768	0.0059
	ρ	0.4529	–0.0471	0.1039	0.0130	0.5634	0.0634	0.1327	0.0216
	$\eta(\cdot)$		MISE = 0.0347				MISE = 0.1088		
(20, 10)	α_1	0.5775	0.0002	0.0325	0.0011	0.5666	–0.0108	0.1111	0.0124
	α_2	0.5775	0.0002	0.0325	0.0011	0.5666	–0.0108	0.1111	0.0124
	α_3	0.5755	–0.0018	0.0330	0.0011	0.5666	–0.0108	0.1054	0.0112
	β	0.3636	0.0636	0.0424	0.0058	0.3075	0.0075	0.0547	0.0030
	ρ	0.4743	–0.0257	0.0633	0.0047	0.5887	0.0887	0.0742	0.0133
	$\eta(\cdot)$		MISE = 0.0143				MISE = 0.0681		
(20, 20)	α_1	0.5770	–0.0004	0.0236	0.0006	0.5662	–0.0112	0.0947	0.0090
	α_2	0.5774	0.0001	0.0233	0.0005	0.5664	–0.0110	0.0846	0.0072
	α_3	0.5761	–0.0012	0.0252	0.0006	0.5792	–0.0018	0.0839	0.0070
	β	0.3524	0.0524	0.0297	0.0036	0.3003	0.0003	0.0332	0.0011
	ρ	0.4774	–0.0226	0.0655	0.0048	0.5820	0.0820	0.0757	0.0124
	$\eta(\cdot)$		MISE = 0.0128				MISE = 0.0591		

6. An Illustration Example

In this section, we take the well-known Boston housing price data as a real example to illustrate the application of the proposed model and its estimation and test methods. The data consist of 506 observations of the median value (MV) of owner-occupied homes in 506 census tracts in the Boston Standard Metropolitan Statistical Area in 1970, together with 13 variables which might explain the variation of housing value (Harrison and Rubinfeld, 1978), and are now freely available through the GeoDa Center for Geospatial Analysis and Computation. The detailed description of these variables is summarized in Table 7.

In the literature of statistics, many researchers employed the semiparametric regression models with single-index term to analyze the Boston housing data. For example, Kong and Xia (2012) explored the relationship between the median value of owner-occupied homes and the remaining 13 variables by using a single-index quantile regression model. However, the single-index regression model may be inappropriate to fit the data because some variables may have linear influence on the housing price. To identify which variables have linear influence on the housing price, Zhang *et al.* (2011) first employed an additive model to fit the data and then applied a variable selection procedure to decide which variables have linear influence on the housing price. They concluded that variables RAD and PTRATIO have linear influence on the response variable, variables CRIM, NOX, RM, DIS, TAX and LSTAT have nonlinear influence on housing price, while the remaining five variables ZN, INDUS, CHAS, AGE and B were removed from the final model as insignificant variables. On

Table 7: Description of variables in Boston housing price data.

Variable	*Description*
MV	Median value of owner-occupied homes in $1,000 per tract
CRIM	Per capita crime rate per tract
ZN	Proportion of a town's residential land zoned for lots greater than 25,000 square feet
INDUS	Proportion of non-retail business acres per town
CHAS	Charles River dummy variable (= 1 if tract bounds the Charles River; 0 otherwise)
NOX	Nitrogen oxide concentration in pphm per town
RM	Average number of rooms per dwelling
AGE	Proportion of owner-occupied homes built prior to 1940 per tract
DIS	Weighted average of distances of a tract to five employment centers in the Boston region
RAD	Index of a town's accessibility to radial highways
TAX	Full-value property tax rate per $10,000 per town
PTRATIO	Pupil-teacher ratio by town school district
B	$1000(Bk - 0.63)^2$ in which Bk is the proportion of blacks per tract
LSTAT	Proportion of population that is in the lower status per tract

the other hand, according to the analysis results of many researchers (Pace and Gilley, 1997; Lesage and Pace, 2009; Li and Mei, 2013; Sun *et al.*, 2014; Du *et al.*, 2018; Liu *et al.*, 2018), the spatial dependence of the median value of owner-occupied homes is a non-ignorable factor that should be considered in the analysis of the Boston housing data.

To take the spatial effects into account and allow more flexible interpretation of the nonparametric component, we consider to fit the data via the following model

$$Y_i = \rho \sum_{j \neq i} w_{ij} Y_j + X_i^T \beta + \eta\left(Z_i^T \alpha\right) + \varepsilon_i, \quad i = 1, \cdots, n, \tag{25}$$

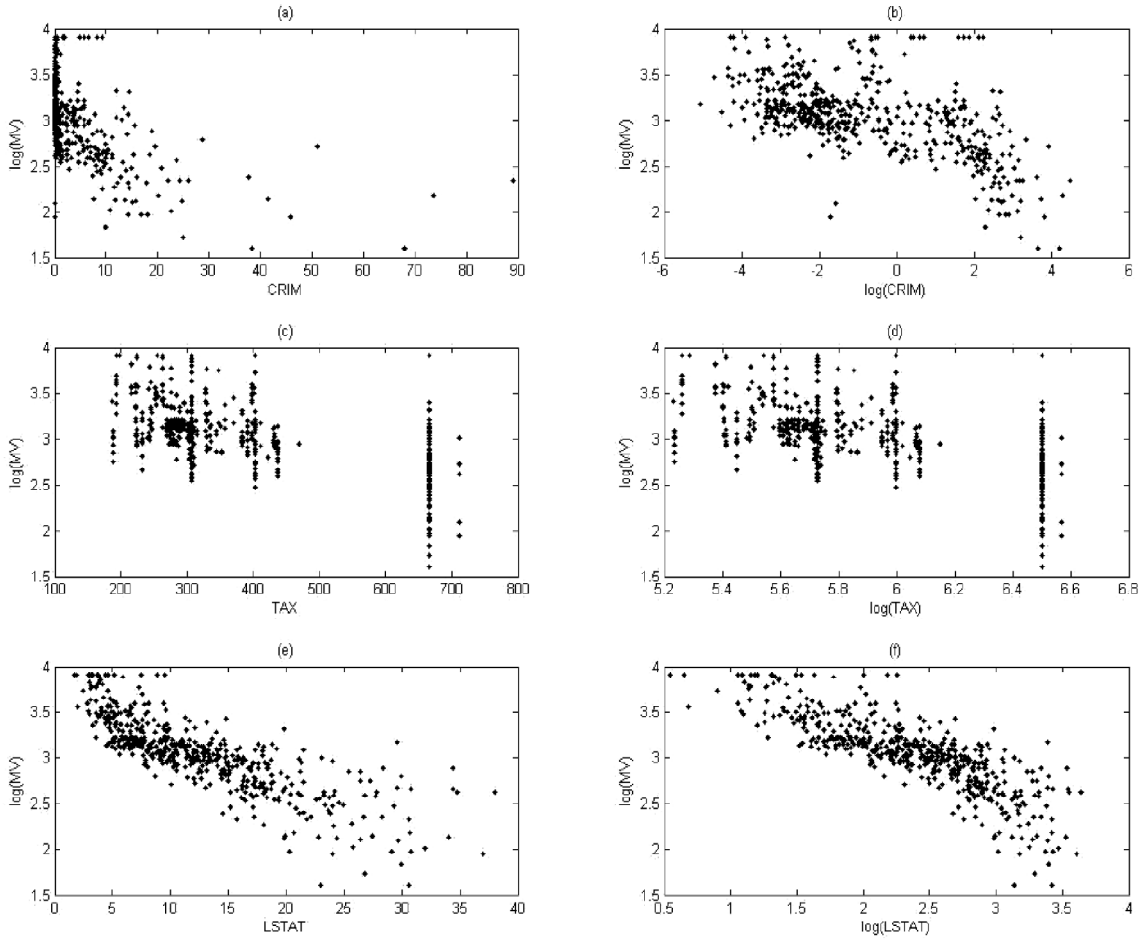
where n = 506, Z = log(log(CRIM), NOX, RM, DIS, log(TAX), log(LSTAT))^T, X = log(RAD, PTRATIO)^T and Y = log (MV). The reason for taking the logarithm of MV as the response variable, instead of MV itself, is that the correlation of log(MV) with the variables log(CRIM), DIS, RAD and log(TAX) is much stronger than that of MV with these variables. The correlation coefficient with log(MV) is –0.5719 for log(CRIM), 0.3425 for DIS, –0.4868 for RAD and –0.5619 for log(TAX), whereas the correlation coefficient with MV is –04573 for log(CRIM), 0.2493 for DIS, –0.3848 for RAD and –0.4783 for log(TAX). Another reason for using the logarithm of CRIM as the explanatory variable is that the observations of CRIM are uniformly distributed on the interval (–6, 6) with the aid of logarithmic transformation, which can be witnessed as well from Figures 1(a) and 1(b). By comparison, the logarithmic transformation for variables TAX and LSTAT is taken only to alleviate the trouble caused by big gaps in the domain, which can be seen from Figures 1(c)-1(f).

As for the choice of the spatial weight matrix $W = (w_{ij})$, following the practice in Pace and Gilley (1997), we take the element w_{ij} of W to be

$$w_{ij} = \max\left(1 - \frac{d_{ij}}{d_0}, 0\right), \tag{26}$$

where d_{ij} is the Euclidean distance calculated in terms of the longitude and latitude coordinates of census tract, and d_0 is a threshold distance which is used to control the degree of spatial dependence of the response variable. If the distance between two census tracts is far enough, the spatial dependence between them will probably attenuate. In order to alleviate the potential effect of tiny weights on the data analysis, a threshold distance is used to set the tiny weights to be zero. Furthermore, we normalize the spatial weight matrix such that the sum of each row of W is equal to 1. In our analysis, we take the value of d_0 to be 0.05, which yields a spatial weight matrix with 19.1% nonzero elements.

We first apply the proposed test method to check the linearity of the link function $\eta(\cdot)$ in model (25), in which m = 1000 bootstrap samples were drawn to compute the p-value of the test. The



Fugure 1: The scatter plots of log(MV) versus CRIM, log (CRIM), TAX, log(TAX), LSTAT and log(LSTAT).

resulting *p*-value is 0, which indicates that we should reject the null hypothesis of linearity. This provides strong evidence that model (25) is more appropriate than the traditional linear spatial autoregressive model for fitting the Boston housing price data.

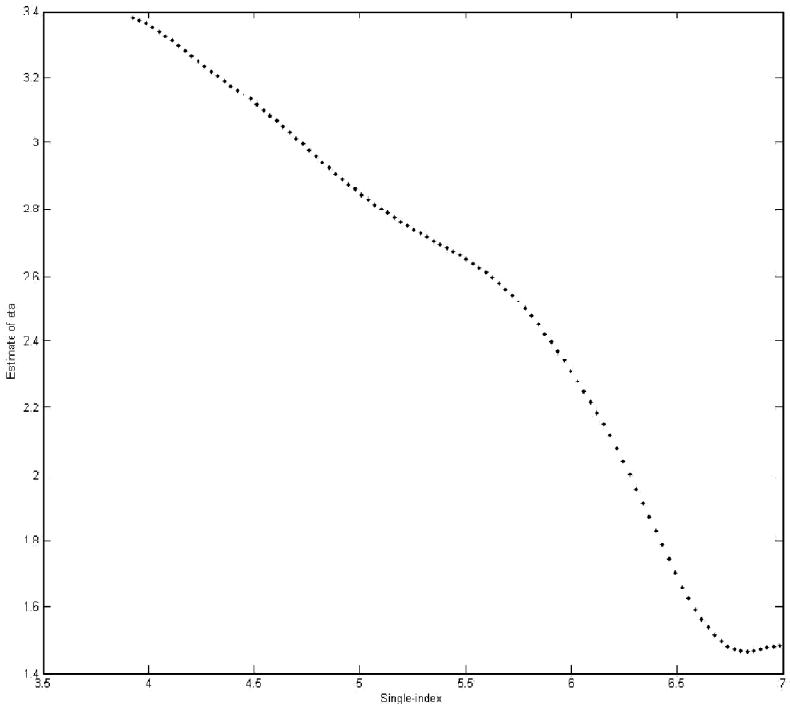
Then we use the proposed estimation method to estimate the unknown parameters and function in model (25). The estimators of the parameter vector θ and the link function $\eta(\cdot)$ are reported in Table 8 and Figure 2, respectively.

We summary some interesting empirical findings from the above analysis as follows. First, the estimated link function shows evident nonlinear decreasing trend as the value of $Z^T\hat{\alpha}$ increases, which together with the estimate of α indicate that the explanatory variables log(CRIM), NOX,

Table 8: Estimated parameter vector θ in model (25).

α_1	α_2	α_3	α_4	α_5	α_6	β_1	β_2	ρ	σ^2
0.0882	0.3854	–0.0643	0.0945	0.6441	0.6448	0.0259	–0.0274	0.2379	0.0310

DIS, log(TAX) and log(LSTAT) in the nonparametric component have negative impact on the housing price, while the influence of the explanatory variable RM is positive. Among them, NOX, and are three very influential adverse factors on the housing price. This is reasonable because people prefer for better air quality, lower TAT and higher educational status neighborhoods. Second, the estimated regression coefficient of X_1 is positive, which indicates that the accessibility to radial highways (RAD) has a positive impact on the housing price. This is reasonable because the larger the value of RAD is, the less time will be spent on commuting. The estimated regression coefficient of X_2 is negative, which reveals that the housing price would decrease as the pupil-teacher ratio increases. Although the relation between PTRATIO and school quality is not completely clear, a lower PTRATIO should imply more individual attention from the teacher. Third, the estimated spatial autoregressive parameter is 0.2379, which means that the housing prices in a neighborhood do affect each other. This is a true phenomenon in real world.



Fugure 2: The estimated link function in model (25)

7. Concluding Remarks

In this paper, we developed a new estimation method for the partially linear single-index spatial autoregressive model by combining the local linear smoothing method and the quasi-maximum likelihood method. The greatest advantage of our estimation method over the existing semiparametric GMM estimation method of Sun and Wu (2018) and Cheng *et al.* (2019) is that there is no need to select the instrumental variables. In comparison with the semiparametric GMM estimation method of Sun and Wu (2018) and Cheng *et al.* (2019), our estimation method has the following two drawbacks. First, our method requires the model error terms to be independent and identically distributed, which is rather restrictive in some empirical applications. Second, it is very difficult to extend our method to model (1) with multiple spatial weight matrices. Furthermore, a generalized likelihood ratio test was proposed to check the parametric form of the link function, in which a residual-based bootstrap procedure was suggested to calculate the *p*-value of the proposed test. The simulation studies show that the proposed estimation and test methods work well in finite samples. The Boston housing price data were analyzed to illustrate the application of the partially linear single-index spatial autoregressive model and its estimation and test methods, which led to some interesting empirical findings.

Two interesting future research topics about the partially linear single-index spatial autoregressive model should be mentioned. First, one fundamental issue in the partially linear single-index spatial autoregressive model is the structure identification or model selection, that is, how to determine which explanatory variables have linear impact on the response variable and which ones are of nonlinear impact on the response variable. In practice, data analysts usually assume a model structure according to their prior knowledge and then make estimation and inference based on the assumed model structure. However, such prior knowledge is rarely available, especially when the number of explanatory variables is large. As a consequence, we will face the dangers of model misspecification if we erroneously incorporate a explanatory variable which has nonlinear impact on the response variable into the linear part of the regression function and of loss of efficiency if we erroneously incorporate a explanatory variable which has linear impact on the response variable into the nonlinear part of the regression function. Thus, determining which explanatory variables have linear impact on the response variable is critical prior to the use of the partially linear single-index spatial autoregressive model. Moreover, during the initial stage of modeling, one often includes as many explanatory variables as possible to avoid misspecification due to the exclusion of the important explanatory variables from the model. However, including excessive unimportant explanatory variables in the model may decrease the efficiency of estimation and test and hence the precision of prediction. Hence, after correctly specifying the model structure of the partially linear single-index spatial autoregressive model, one needs to further select the important explanatory variables in both parametric and nonparametric component to increase the efficiency of estimation and test and finally the precision of prediction.

Second, the current studies about the partially linear single-index spatial autoregressive model all assumed that the spatial weight matrix is exogenous. This exogenous assumption is reasonable

if the spatial weight matrix is constructed based on contiguity or geographic distances among spatial units. However, in some practical applications especially in the field of economics, it is much better to construct the spatial weight matrix by using economic or socioeconomic distances. In this case, the spatial weight matrix is likely to be endogenous. Thus, it is appealing and necessary to study the estimation and related test problems of the partially linear single-index spatial autoregressive model with an endogenous spatial weight matrix.

Acknowledgements

This research was supported by the National Statistical Science Project [grant number 2019LY36] and the Natural Science Foundation of Shaanxi Province [grant number 2021JM349].

References

Ai, C.R., Zhang, Y.Q. (2017). Estimation of partially specified spatial panel data models with fixed-effects. *Econometric Reviews*, 36: 6-22.

Anselin, L. (1988). Spatial Econometrics: Methods and Models. Dordrecht: Kluwer Academic Publishers.

Cai, Z.W., Fan, J.Q., Yao, Q.W. (2000). Functional-coefficient regression models for nonlinear time series. *Journal of the American Statistical Association*, 95: 941-956.

Carroll, R.J., Fan, J.Q., Gijbels, I., Wand, M.P. (1997). Generalized partially linear single-index models. *Journal of the American Statistical Association*, 92: 477-489.

Case, A.C. (1991). Spatial patterns in household demand. *Econometrica*, 59: 953-965.

Cheng, S.L., Chen, J.B., Liu, X. (2019). GMM estimation of partially linear single-index spatial autoregressive model. *Spatial Statistics*, 31: 100354.

Du, J., Sun, X.Q., Cao, R.Y., Zhang, Z.Z. (2018). Statistical inference for partially linear additive spatial autoregressive models. *Spatial Statistics*, 25: 52-67.

Fan, J.Q., Gijbels, I. (1996). Local Polynomial Modelling and Its Applications. London: Chapman & Hall.

Fan, J.Q., Huang, T. (2005). Profile likelihood inferences on semiparametric varying-coefficient partially linear models. Bernoulli, 11: 1031-1057.

Fan, J.Q., Jiang, J.C. (2005). Nonparametric inferences for additive models. *Journal of the American Statistical Association*, 100: 890-907.

Fan, J.Q., Jiang, J.C. (2007). Nonparametric inference with generalized likelihood ratio tests. Test, 16: 409-444.

Hall, P., Hart, J.D. (1990). Bootstrap test for difference between means in nonparametric regression. *Journal of the American Statistical Association*, 412: 1039-1049.

Härdle, W., Mammen, E. (1993). Comparing nonparametric versus parametric regression fits. *The Annals of Statistics*, 21: 1926-1947.

Harrison, D., Rubinfeld, D.L. (1978). Hedonic housing prices and the demand for clean air. *Journal of Environmental Economics and Management*, 5: 81-102.

Kong, E., Xia, Y.C. (2012). A single-index quantile regression model and its estimation. Econometric Theory, 28: 730-768.

Lee, L.F. (2004). Asymptotic distributions of quasi-maximum likelihood estimators for spatial autoregressive models. *Econometrica*, 72: 1899-1925.

Lesage, J.P., Pace, R.K. (2009). Introduction to Spatial Econometrics. Boca Raton: CRC Press.

Li, K.M., Chen, J.B. (2013). Profile maximum likelihood estimation of semi-parametric varying coefficient spatial lag model. *The Journal of Quantitative & Technical Economics*, 30: 85-98.

Li, T.Z., Guo, Y. (2020). Penalized profile quasi-maximum likelihood method of partially linear spatial autoregressive model. *Journal of Statistical Computation and Simulation*, 90: 2681-2704.

Li, T.Z., Mei, C.L. (2013). Testing a polynomial relationship of the non-parametric component in partially linear spatial autoregressive models. *Papers in Regional Science*, 92: 633-649.

Li, T.Z., Mei, C.L. (2016). Statistical inference on the parametric component in partially linear spatial autoregressive models. Communications in Statistics-Simulation and Computation, 45: 1991-2006.

Liang, H., Liu, X., Li, R.Z., Tsai, C.-L. (2010). Estimation and testing for partially linear single-index models. The Annals of Statistics, 38: 3811-3836.

Liu, X., Chen, J.B., Cheng, S.L. (2018). A penalized quasi-maximum likelihood method for variable selection in the spatial autoregressive model. *Spatial Statistics*, 25: 86-104.

Pace, R.K., Gilley, O.W. (1997). Using the spatial configuration of the data to improve estimation. *Journal of Real Estate Finance and Economics*, 14: 333-340.

Stute, W., González-Manteiga, W., Presedo-Quindimil, M. (1998). Bootstrap approximation in model checks for regression. *Journal of the American Statistical Association*, 93: 141-149.

Su, L.J., Jin, S.N. (2010). Profile quasi-maximum likelihood estimation of partially linear spatial autoregressive models. *Journal of Econometrics*, 157: 18-33.

Sun, Y., Wu, Y.Q. (2018). Estimation and testing for a partially linear single—index spatial regression model. Spatial Economic Analysis, 13: 473-489.

Sun, Y., Yan, H.J., Zhang, W.Y., Lu, Z.D. (2014). A semiparametric spatial dynamic model. The Annals of Statistics, 42: 700-727.

Xia, Y.C., Härdle, W. (2006). Semi-parametric estimation of partially linear single-index models. *Journal of Multivariate Analysis*, 97: 1162-1184.

Yu, Y., Ruppert, D. (2002). Penalized spline estimation for partially linear single-index models. *Journal of the American Statistical Association*, 97: 1042-1054.

Zhang, H.H., Cheng, G., Liu, Y.F. (2011). Linear or nonlinear? automatic structure discovery for partially linear models. *Journal of the American Statistical Association*, 106: 1099-1112.

Zhang, Y.Q., Sun, Y.Q. (2015). Estimation of partially specified dynamic spatial panel data models with fixed-effects. Regional Science and Urban Economics, 51: 37-46.

Zhang, Y.Q., Yang, G.R. (2015a). Statistical inference of partially specified spatial autoregressive model. Acta Mathematicae Applicatae Sinica, English Series, 31: 1-16.

Zhang, Y.Q., Yang, G.R. (2015b). Estimation of partially specified spatial panel data models with random-effects. Acta Mathematica Sinica, English Series, 31: 456-478.

Zhang, Z.Y. (2013). A pairwise difference estimator for partially linear spatial autoregressive models. Spatial Economic Analysis, 8: 176-194.41