Research article

Estimation of partially linear single-index spatial autoregressive model using B-splines

Lei Liu[1,*]**and Jun Dai**[2]

[1] Yunnan Key Laboratory of Statistical Modeling and Data Analysis, Yunnan University, Kunming 650091, China

[2] School of Mathematics, Statistics and Mechanics, Beijing University of Technology, Beijing 100124, China

* **Correspondence:** Email: liulei_yndx@163.com.

Abstract: Possible dependence across spatial units is a relevant issue in many areas in practice. In this paper, we consider the partially linear single-index spatial autoregressive model to analyze the dependence of the spatial units and suggest an estimation method. An algorithm procedure is proposed to estimate the link function for the single index and the parameters in the single index, as well as the parameters in the linear component and the spatial parameter of the model. The nonparametric function is estimated based on B-spline approximation. The Nelder-Mead iteration algorithm is adopted to calculate the parametric and nonparametric parts simultaneously in the optimization. The asymptotic properties of parameter and function estimates are established. Monte Carlo simulation studies are conducted to investigate the performance of the proposed estimation methodology and calculation procedure. Furthermore, we use the proposed method to analyze air quality data and rural household income data in China.

Keywords: spatial autoregressive model; partially linear single-index; B-splines

1. Introduction

In empirical practice of economics and statistics, possible dependence across spatial units is a relevant issue in urban, real estate, regional, public, agricultural, environmental economics, and industrial organization [1]. Many studies have proposed approaches and techniques to deal with these problems by imposing some structures on the model in order to capture spatial dependence or spatial autocorrelation in cross-sectional and/or panel data. In recent years, spatial autoregressive (SAR) models have attracted great interest due to their usefulness in practice. For instance, SAR models have the property of using autocorrelation structures according to spatial patterns on data. These

models allow for the analysis of spatial dependencies and interactions among observations, making them particularly useful in fields such as geography, environmental science, and urban planning where spatial relationships play crucial roles.

Lee [2] explained that ordinary least squares (OLS) estimators can be inconsistent for estimating an SAR model due to the fact that the spatial lagged variables can be correlated with the disturbances in a spatial model. In order to obtain a consistent estimator, Lee [1] proposed the quasi-maximum likelihood estimator (QMLE) for the SAR model, investigated the asymptotic properties of QMLE, and found out that the rates of convergence of QMLE may depend on some general features of the spatial weight matrix of the model. However, the quasi-maximum likelihood procedures often pose computational challenges when dealing with large sample sizes, as analytical solutions are not available and numerical methods must be employed to obtain numerical solutions. Kelejian and Prucha [3] proposed the generalized spatial two-stage least squares procedure for estimating the SAR model in which they demonstrated consistency and asymptotic normality of the resulting estimator. However, the estimator is not asymptotically effective. Lee [4] proposed best spatial two stage least squares estimators for the SAR model to improve the efficiency of the estimator. Su [5] extended the SAR to a semi-parametric setting and proposed the GMM estimation method, and showed that the resulting estimator is consistent and asymptotically normal. Xu and Lee [6] proposed the maximum likelihood estimator (MLE) for the SAR Tobit model and studied the asymptotic properties.

Recently, there has been a growing literature on semiparametric spatial autoregressive models. Invoking the flexibility of partial linear model, Su and Jin [7] developed asymptotic theories for the QMLE of partial linear SAR models based on the profile quasi-maximum likelihood theme. In Su and Jin [7], the rate of convergence of the spatial parameter estimator depends on some general features of the spatial weight matrix of the model. Du et al. [8] proposed the partial linear additive SAR model. Obviously, the model proposed by Du et al. [8] did not impose any interaction within nonparametric components in spatial data. Therefore, the proposed method by Du et al. [8] cannot solve the problem of interaction effects on response variables in spatial data. Moreover, the nonparametric components are subject to the curse of dimensionality and can only accommodate low-dimensional covariates; see Wang [9]. The single-index model is an important tool in multivariate nonparametric regression. By reducing the dimensionality from multivariate predictors to a univariate index, single-index models avoid the so-called "curse of dimensionality" while still capturing important features in high-dimensional data. Cheng and Chen [10] introduced the partially linear single-index spatial autoregressive model (PLSISARM), employing local linear approximation to estimate unknown link functions and utilizing MLE. In contrast to the partial linear SAR model and the partial linear additive SAR model, PLSISARM considers covariate interactions within its nonparametric component. This paper proposes a new estimation procedure for PLSISARM. Specifically, we utilize B-splines to approximate the unknown link function, enhancing smoothness, and adopt QMLE as the estimation method. We also introduce a feasible Nelder-Mead iteration algorithm tailored for this new estimation procedure. Furthermore, we establish the asymptotic properties for the QMLE. Finally, we demonstrate the application of PLSISARM on two distinct datasets from meteorology and economics, thereby extending the practical utility of the model. Compared to Cheng and Chen's work [10], B-spline approximation is computationally more efficient and smoother than local linear approximation, facilitating easier interpretation in practical applications. Additionally, this paper employs QMLE instead of MLE, resulting in differences in asymptotic distributions. Notably, when

the error term does not satisfy the normal distribution, QMLE can still estimate the asymptotic variance, while MLE fails to do so.

This paper is organized as follows. In Section 2, we describe the technical setting and propose the estimation procedure for PLSISARM via QMLE. In Section 3, we establish the asymptotic properties for the proposed procedure. In Section 4, we present the simulation studies to assess the finite sample performance of the proposed procedure. In Section 5, we analyze two real data sets to illustrate the practical usefulness of the proposed model. In Section 6, we conclude the paper. The appendix provides the proofs of the main results.

2. The model and estimation procedures

We consider the partially linear single-index spatial autoregressive model (PLSISARM)

$$Y_N = \lambda W_N Y_N + X_N \beta + g(Z_N \alpha) + V_N, \tag{2.1}$$

where N is the total number of spatial units; Y_N is the N-dimensional vector of observations of the dependent variable; W_N is an $N \times N$ specified constant spatial weight matrix with zero diagonal elements, the weights obtained based on physical distance, social networks, or "economic" distance; $X_N = (x_1, x_2, \ldots, x_N)^T$ and $Z_N = (z_1, z_2, \ldots, z_N)^T$ are the $N \times p$ and $N \times d$ matrices of regressors, respectively; g is an unknown differentiable function, the nonparametric components assume that the influence of the covariates z can be collapsed to a single index, $z^T \alpha$, via a nonparametric link function g; $\alpha \in R^d, \|\alpha\| = 1$, and the first nonzero element of α is positive for identifiability; and $V_N = (\epsilon_1, \epsilon_2, \ldots, \epsilon_N)^T$ is an N-dimensional vector of i.i.d. disturbances with zero mean and finite variance σ^2.

Remark 1: PLSISARM is flexible enough to cover a variety of situations. When $\lambda = 0, \beta = 0$, and $d = 1$, the proposed model is reduced to the partial linear single-index model, the single-index spatial autoregressive model, and the partial linear spatial autoregressive model, respectively.

In order to obtain the estimator of the parameters $\theta = (\sigma^2, \lambda, \alpha^T, \beta^T)^T$ and the function $g(\cdot)$, similar to Lee [1] and Su and Jin [7], we maximize the following log quasi-likelihood function to obtain the estimator

$$L_N(\theta, g(\cdot)) = -\frac{N}{2} \log(2\pi) - \frac{N}{2} \log \sigma^2 + \log |M_N(\lambda)| - \frac{1}{2\sigma^2} V_N^T(\theta) V_N(\theta),$$

where $S_N(\lambda) = I_N - \lambda W_N$, and $V_N(\theta) = Y_N - \lambda W_N Y_N - X_N \beta - g(Z_N \alpha)$. Since the infinite parameter $g(\cdot)$ in the objective function $L_N(\theta)$ is unknown, we are unable to maximize $L_N(\theta)$ to obtain the estimator of θ directly. The B-spline method has been widely used in estimation for nonparametric models and semi-parametric models [11]. B-splines also have numerous precedences for their application in semi-parametric spatial autoregressive models, as demonstrated in Zhang et al. [12], Du et al. [8], and Tian et al. [13]. In this paper, we use B-splines to approximate the nonparametric function $g(t)$ due to the computational advantages. Let $B(t) = (B_1(t), \ldots, B_{K_N+l+1}(t))^T$ be a set of B-spline basis functions of order $l+1$ with knots $a = t_0 < t_1 < \ldots < t_{k_N} < t_{K_N+1} = b$ for $a < b \in \mathbb{R}$. Therefore, approximating $g(t)$ by a linear combination of normalized B-spline basis functions, it has

$$g(t) \approx \sum_{j=1}^{K_N+l+1} B_j(t) \gamma_j = B(t)^T \gamma,$$

where $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_{(K_N+l+1)})^T$ is the unknown spline coefficient vector. Based on the above B-spline approximation, model (2.1) can be rewritten as

$$Y_N \approx \lambda W_N Y_N + X_N \boldsymbol{\beta} + \boldsymbol{\Pi}(Z_N \boldsymbol{\alpha})\boldsymbol{\gamma} + V_N, \tag{2.2}$$

where $\boldsymbol{\Pi}(Z_N \boldsymbol{\alpha}) = (B(z_1^T \boldsymbol{\alpha}), \ldots, B(z_N^T \boldsymbol{\alpha}))^T$.

Given $\boldsymbol{\alpha}$, apply the QMLE method [1] to model (2.2) to obtain the estimator of $\boldsymbol{\delta} = (\sigma^2, \lambda, \boldsymbol{\beta}^T, \boldsymbol{\gamma}^T)^T$, denoted as $\hat{\boldsymbol{\delta}}(\boldsymbol{\alpha}) = (\sigma^2(\boldsymbol{\alpha}), \hat{\lambda}(\boldsymbol{\alpha}), \hat{\boldsymbol{\beta}}(\boldsymbol{\alpha})^T, \hat{\boldsymbol{\gamma}}(\boldsymbol{\alpha})^T)^T$. Next, we consider the estimation of the parameter $\boldsymbol{\alpha}$. Substitute $\hat{\boldsymbol{\delta}}(\boldsymbol{\alpha})$ back into model (2.2). Then we have

$$Y_N \approx \hat{\lambda}(\boldsymbol{\alpha})W_N Y_N + X_N \hat{\boldsymbol{\beta}}(\boldsymbol{\alpha}) + \boldsymbol{\Pi}(Z_N \boldsymbol{\alpha})\hat{\boldsymbol{\gamma}}(\boldsymbol{\alpha}) + V_N. \tag{2.3}$$

Minimize the following objective function to achieve the estimator of $\boldsymbol{\alpha}$:

$$Q_N(\boldsymbol{\alpha}) = \|Y_N - \hat{\lambda}(\boldsymbol{\alpha})W_N Y_N - X_N \hat{\boldsymbol{\beta}}(\boldsymbol{\alpha}) - \boldsymbol{\Pi}(Z_N \boldsymbol{\alpha})\hat{\boldsymbol{\gamma}}(\boldsymbol{\alpha})\|^2.$$

Once we get the estimator $\hat{\boldsymbol{\alpha}}$, we can further obtain $\hat{\boldsymbol{\delta}}(\hat{\boldsymbol{\alpha}}) = (\hat{\sigma}^2(\hat{\boldsymbol{\alpha}}), \hat{\lambda}(\hat{\boldsymbol{\alpha}}), \hat{\boldsymbol{\beta}}(\hat{\boldsymbol{\alpha}})^T, \hat{\boldsymbol{\gamma}}(\hat{\boldsymbol{\alpha}})^T)^T$. Then, we can achieve the estimator of $g(\cdot)$ as $\hat{g} = B(z^T \hat{\boldsymbol{\alpha}})^T \hat{\boldsymbol{\gamma}}$.

To elaborate the ideas of the estimation methodology, we detail the algorithm as follows.

Algorithm for stage one: Calculate the initial value of the parameters $\boldsymbol{\theta} = (\sigma^2, \lambda, \boldsymbol{\beta}^T, \boldsymbol{\alpha}^T)^T$ and $\boldsymbol{\gamma}$.

Step 1. Calculate the initial value of $\boldsymbol{\alpha}$ for iteration.

Consider the following model:

$$Y_N = \lambda W_N Y_N + X_N \boldsymbol{\beta} + Z_N \boldsymbol{\alpha} + V_N. \tag{2.4}$$

Apply the QMLE of Lee (2004) to model (2.4) to get the estimator of the parameters. Denote the initial value of $\boldsymbol{\alpha}$ as $\hat{\boldsymbol{\alpha}}^{(0)}$. Normalize $\hat{\boldsymbol{\alpha}}^{(0)}$ so that $\|\hat{\boldsymbol{\alpha}}^{(0)}\| = 1$ and the first element of $\hat{\boldsymbol{\alpha}}^{(0)}$, $\hat{\alpha}_i^{(0)} > 0$.

Step 2. Substitute $\hat{\boldsymbol{\alpha}}^{(0)}$ into model (2.2) and calculate the QMLE for the parameters in the following model:

$$Y_N = \lambda W_N Y_N + X_N \boldsymbol{\beta} + \boldsymbol{\Pi}(Z_N \hat{\boldsymbol{\alpha}}^{(0)})\boldsymbol{\gamma} + V_N.$$

Let the resulting estimators $\hat{\lambda}^{(0)}$, $\hat{\boldsymbol{\beta}}^{(0)}$ and $\hat{\boldsymbol{\gamma}}^{(0)}$ be the initial values of λ, $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$, respectively.

Algorithm for stage two: The iteration procedure.

Step 3. Update the estimation for $\boldsymbol{\alpha}$.

Substitute the estimators $\hat{\lambda}^{(k)}$, $\hat{\boldsymbol{\beta}}^{(k)}$, and $\hat{\boldsymbol{\gamma}}^{(k)}$ into model (2.2) to obtain the following model:

$$Y_N = \hat{\lambda}^{(k)} W_N Y_N + X_N \hat{\boldsymbol{\beta}}^{(k)} + \boldsymbol{\Pi}(Z_N \boldsymbol{\alpha})\hat{\boldsymbol{\gamma}}^{(k)} + V_N.$$

Then, minimize the following objective function to get the update estimation of $\boldsymbol{\alpha}$, $\hat{\boldsymbol{\alpha}}^{(k+1)}$,

$$Q(\boldsymbol{\alpha}) = \|Y_N - \hat{\lambda}^{(k)} W_N Y_N - X_N \hat{\boldsymbol{\beta}}^{(k)} - \boldsymbol{\Pi}(Z_N \boldsymbol{\alpha})\hat{\boldsymbol{\gamma}}^{(k)}\|^2.$$

Normalize $\hat{\boldsymbol{\alpha}}^{(k+1)}$ so that $\|\hat{\boldsymbol{\alpha}}^{(k+1)}\| = 1$ and the first element of $\hat{\boldsymbol{\alpha}}^{(k+1)}$, $\hat{\alpha}_i^{(k+1)} > 0$. The optimization for $Q(\boldsymbol{\alpha})$ is completed by the Nelder-Mead method of the "optim" function in R.

Step 4. Update the estimation for $\sigma^2, \lambda, \boldsymbol{\beta}$, and $\boldsymbol{\gamma}$.

Substitute $\hat{\boldsymbol{\alpha}}^{(k+1)}$ into model (2.2) and calculate the QMLE for the parameters in the following model to get the update estimators $\hat{\sigma}^{2(k+1)} \hat{\lambda}^{(k+1)}, \hat{\boldsymbol{\beta}}^{(k+1)}$, and $\hat{\boldsymbol{\gamma}}^{(k+1)}$:

$$\boldsymbol{Y}_N = \lambda \boldsymbol{W}_N \boldsymbol{Y}_N + \boldsymbol{X}_N \boldsymbol{\beta} + \boldsymbol{\Pi}(\boldsymbol{Z}_N \hat{\boldsymbol{\alpha}}^{(k+1)}) \boldsymbol{\gamma} + \boldsymbol{V}_N.$$

Repeat Step 3 and Step 4 until the estimator converges, which is

$$|\hat{\lambda}^{(k+1)} - \hat{\lambda}^{(k)}|^2 + \|\hat{\boldsymbol{\alpha}}^{(k+1)} - \hat{\boldsymbol{\alpha}}^{(k)}\|^2 + \|\hat{\boldsymbol{\beta}}^{(k+1)} - \hat{\boldsymbol{\beta}}^{(k)}\|^2 + \|\hat{\boldsymbol{\gamma}}^{(k+1)} - \hat{\boldsymbol{\gamma}}^{(k)}\|^2 < C,$$

where C is a predetermined constant which is $C = 10^{-4}$ in the simulation studies. Thus, the estimator for $g(\cdot)$ is $\boldsymbol{B}(\boldsymbol{z}^T \hat{\boldsymbol{\alpha}}^{(k+1)})^T \hat{\boldsymbol{\gamma}}^{(k+1)}$.

Selection of knots: It is very important to select the knots (including the number of knots and the position of knots for B-spline estimation) in this paper. We use the BIC criterion to select the number of knots. Given the number of knots, it is suggested that the position of the knots could be equally spaced or equally spaced sample quantiles of the predictor variables. The order of splines is not necessarily better when it is higher. To avoid overfitting, lower-order splines, such as quadratic and cubic splines, are typically chosen. However, the approximation accuracy also needs to be considered, so cubic splines are often selected in practice, which is also a common choice in literature [8, 13]. In this paper, we use cubic splines with equally spaced sample quantiles of the predictor variable knots to estimate the B-spline basis.

3. Asymptotic properties

For the purpose of statistical inference, the large-sample properties of the estimates need to be established. Let $\boldsymbol{M}(\lambda) = \boldsymbol{I} - \lambda \boldsymbol{W}_N$, and for any matrix $\boldsymbol{M}$, denote the projection matrix $\boldsymbol{P}_M = \boldsymbol{I} - \boldsymbol{M}(\boldsymbol{M}^T \boldsymbol{M})^{-1} \boldsymbol{M}^T$. The log likelihood function of model (2.2) is

$$L_N(\boldsymbol{\theta}, \boldsymbol{\gamma}) = -\frac{N}{2} \log(2\pi) - \frac{N}{2} \log \sigma^2 + \log |\boldsymbol{M}(\lambda)| - \frac{1}{2\sigma^2} \boldsymbol{V}_N^T(\boldsymbol{\theta}, \boldsymbol{\gamma}) \boldsymbol{V}_N(\boldsymbol{\theta}, \boldsymbol{\gamma}), \quad (3.1)$$

where $\boldsymbol{V}_N(\boldsymbol{\theta}, \boldsymbol{\gamma}) = \boldsymbol{M}_N(\lambda) \boldsymbol{Y}_N - \boldsymbol{X}_N \boldsymbol{\beta} - \boldsymbol{\Pi}(\boldsymbol{Z}_N \boldsymbol{\alpha}) \boldsymbol{\gamma}$. Then, we concentrate out $\boldsymbol{\gamma}$, and, when it does not cause ambiguity, denote $\boldsymbol{\Pi}(\boldsymbol{Z}_N \boldsymbol{\alpha})$ simply as $\boldsymbol{\Pi}$. The concentrated log likelihood function of $\boldsymbol{\theta}$ is

$$\tilde{L}_N(\boldsymbol{\theta}) \propto -\frac{N}{2} \log \sigma^2 + \log |\boldsymbol{M}(\lambda)| - \frac{1}{2\sigma^2} (\boldsymbol{M}(\lambda) \boldsymbol{Y}_N - \boldsymbol{X}_N \boldsymbol{\beta})^T \boldsymbol{P}_{\boldsymbol{\Pi}} (\boldsymbol{M}(\lambda) \boldsymbol{Y}_N - \boldsymbol{X}_N \boldsymbol{\beta}). \quad (3.2)$$

Moreover, let $\boldsymbol{\eta} = (\lambda, \boldsymbol{\alpha})$. Given $\boldsymbol{\eta}$, one can derive the solutions for $\boldsymbol{\beta}$ and σ^2 from (3.2), which are

$$\hat{\boldsymbol{\beta}}(\boldsymbol{\eta}) = (\boldsymbol{X}_N^T \boldsymbol{P}_{\boldsymbol{\Pi}} \boldsymbol{X}_N)^{-1} \boldsymbol{X}_N^T \boldsymbol{P}_{\boldsymbol{\Pi}} \boldsymbol{M}(\lambda) \boldsymbol{Y}_N, \quad (3.3)$$

and

$$\hat{\sigma}^2(\boldsymbol{\eta}) = \frac{1}{N} [\boldsymbol{M}(\lambda) \boldsymbol{Y}_N]^T \boldsymbol{P}_{\boldsymbol{\Pi}}^T \boldsymbol{P}_{\boldsymbol{P}_{\boldsymbol{\Pi}} \boldsymbol{X}_N} \boldsymbol{P}_{\boldsymbol{\Pi}} \boldsymbol{M}(\lambda) \boldsymbol{Y}_N. \quad (3.4)$$

By concentrating $\boldsymbol{\beta}$ and σ^2 from (3.2), we derive the concentrated likelihood

$$\tilde{L}_N^c(\boldsymbol{\eta}) = \frac{N}{2} \log 2\pi - \frac{N}{2} + \log |\boldsymbol{M}(\lambda)| - \frac{N}{2} \log\{\hat{\sigma}^2(\boldsymbol{\eta})\}. \quad (3.5)$$

Let $\boldsymbol{\theta}_0 = (\sigma_0^2, \boldsymbol{\eta}_0^T, \boldsymbol{\beta}_0^T)^T$ with $\boldsymbol{\eta}_0 = (\lambda_0, \boldsymbol{\alpha}_0^T)^T$ be the true parameters. Define $G(\lambda) = \boldsymbol{W}_N \boldsymbol{M}^{-1}(\lambda)$, $\boldsymbol{M}_0 = \boldsymbol{M}(\lambda_0)$, $\boldsymbol{G}_0 = \boldsymbol{W}_N \boldsymbol{M}_0^{-1}$, and $\boldsymbol{A}_0 = \left(g(z_1^T \boldsymbol{\alpha}_0), \ldots, g(z_N^T \boldsymbol{\alpha}_0)\right)^T$. We then rewrite the model as

$$\boldsymbol{Y}_N = \boldsymbol{X}_N \boldsymbol{\beta}_0 + \lambda_0 \boldsymbol{G}^{-1}(\boldsymbol{X}_N \boldsymbol{\beta}_0 + \boldsymbol{A}_0) + \boldsymbol{M}_0^{-1} \boldsymbol{V}_N, \tag{3.6}$$

which is derived from the equation $\boldsymbol{M}_0^{-1} = \boldsymbol{I} + \lambda_0 \boldsymbol{G}_0$. Denote $Q(\boldsymbol{\eta}) = \max_{\beta, \sigma} E(\tilde{L}_N(\boldsymbol{\theta}))$. Then, the solution of this problem is

$$\boldsymbol{\beta}^*(\boldsymbol{\eta}) = (\boldsymbol{X}_N^T \boldsymbol{P}_\Pi \boldsymbol{X}_N)^{-1} \boldsymbol{X}_N^T \boldsymbol{P}_\Pi \boldsymbol{M}(\lambda) \boldsymbol{M}_0^{-1}(\boldsymbol{X}_N \boldsymbol{\beta} + \boldsymbol{A}_0) \tag{3.7}$$

and

$$\begin{aligned}
\sigma^{2*}(\boldsymbol{\eta}) =& \frac{1}{N} \left(\boldsymbol{A}_0 + (\lambda_0 - \lambda)\boldsymbol{R}\right)^T \boldsymbol{P}_\Pi^T \boldsymbol{P}_{\boldsymbol{P}_\Pi \boldsymbol{X}_N} \boldsymbol{P}_\Pi \left(\boldsymbol{A}_0 + (\lambda_0 - \lambda)\boldsymbol{R}\right) \\
&+ \frac{\sigma_0^2}{N} tr\left[\left(\boldsymbol{P}_\Pi \boldsymbol{M}(\lambda)\boldsymbol{M}_0^{-1}\right)^T \boldsymbol{P}_{\boldsymbol{P}_\Pi \boldsymbol{X}_N} \boldsymbol{P}_\Pi \boldsymbol{M}(\lambda)\boldsymbol{M}_0^{-1}\right].
\end{aligned} \tag{3.8}$$

where $\boldsymbol{R} = \boldsymbol{G}_0(\boldsymbol{X}_N \boldsymbol{\beta}_0 + \boldsymbol{A}_0)$. Consequently,

$$Q(\boldsymbol{\eta}) = \frac{N}{2} \log 2\pi - \frac{N}{2} + \log |\boldsymbol{M}(\lambda)| - \frac{N}{2} \log\{\sigma^{2*}(\boldsymbol{\eta})\}. \tag{3.9}$$

Let $\boldsymbol{\Pi}_0 = \boldsymbol{\Pi}(\boldsymbol{Z}_N \boldsymbol{\alpha}_0), \mu_3^* = E(|\epsilon_i|^3)$, and $\mu_4^* = E(|\epsilon_i|^4)$. Before establishing the asymptotic properties of the estimator, we need to make the following assumptions:

A1. (i) The disturbance ϵ_i is i.i.d. with 0 mean and σ_0^2 variance, and the moment $E(|\epsilon_i|^{4+\nu})$ exists for some $\nu > 0$.
(ii) The regressors x_i and z_i are i.i.d. non-stochastic sequences and have bounded support sets on $\mathbb{R}^p$ and $\mathbb{R}^q$, respectively.
(iii) The distribution of $z_i^T \boldsymbol{\alpha}$ is absolutely continuous and its density is bounded away from zero and infinity on $\mathbb{U} = \{u = z_i^T \boldsymbol{\alpha}, z_i \in \mathbb{R}^q, \|\boldsymbol{\alpha}\| = 1\}$.
(iv) The real-valued function $g(\cdot) \in C^r(\mathbb{U})$, where $r \le 2$.

A2. (i) The matrix $\boldsymbol{M}(\lambda)$ is nonsingular for all $\lambda \in (-1, 1)$.
(ii) The entries w_{ij} of $\boldsymbol{W}_N$ satisfies that $w_{ii} = 0$, $\sum_{j=1}^N w_{ij} = 1$, and $w_{ij} = O_p(1/h_N)$ for all $i, j = 1, 2, \ldots, N$ with $\lim_{N \to \infty} h_N/N = 0$.
(iii) The matrices $\boldsymbol{W}_N$ and $\boldsymbol{M}_0^{-1}$ are uniformly bounded in both row and column sums. The matrix $\boldsymbol{M}^{-1}(\lambda)$ is uniformly bounded in both row and column sums, uniformly in λ in a compact parameter space $\Lambda \subset (-1, 1)$. The true λ_0 is an interior point of Λ.

A3. (i) Let $t_1 < \ldots < t_{k_N}$ be the interior knots of $\mathbb{U}$ and t_0, t_{k_N+1} the bound points of $\mathbb{U}$. Letting $\varrho_j = t_j - t_{j-1}$, there exists a constant C such that $\max\{\varrho_j\}/\min\{\varrho_j\} \le C$ and $\max\{\varrho_j\} = o(K_N^{-1})$.
(ii) The knots number K_N is assumed to satisfy $K_N = O(N^{\frac{1}{2r+1}})$.

A4. The limit $\lim_{n \to \infty} \left(\boldsymbol{X}_N, -\dot{A}(\boldsymbol{Z}_N \boldsymbol{\alpha}_0)\boldsymbol{Z}_N, \boldsymbol{R}_N\right)^T \boldsymbol{P}_{\Pi_0} \left(\boldsymbol{X}_N, -\dot{A}(\boldsymbol{Z}_N \boldsymbol{\alpha}_0)\boldsymbol{Z}_N, \boldsymbol{R}_N\right)$ exists and is nonsingular, where $\dot{A}(\boldsymbol{Z}_N \boldsymbol{\alpha}_0) = diag\left\{\dot{g}(z_1^T \boldsymbol{\alpha}_0), \ldots, \dot{g}(z_N^T \boldsymbol{\alpha}_0)\right\}$ and $\dot{g}$ is the first derivative of g.

Assumption A1 outlines the essential features of the disturbances, regressors, and function g for the model. Assumption A2 is parallel to Assumption 2, Assumption 5, and Assumption 7 of Lee [1], which characterize the spatial weight matrix $\boldsymbol{W}_N$. Assumption A3 constrains the characteristics of the knots in B-splines, in parallel with the work of Du et al. [8] and Tian et al. [13]. Assumption A4 ensures the existence of the limiting distribution of the parameters. Subsequently, we present the following main results.

Theorem 1. *Suppose that Assumptions A1–A4 hold. Then, $\boldsymbol{\theta}_0$ is globally identifiable and $\hat{\boldsymbol{\theta}}$ is consistent for $\boldsymbol{\theta}_0$.*

Theorem 2. *Suppose that Assumptions A1–A4 hold. Then, $\hat{g}(\boldsymbol{z}_i^T \hat{\alpha}) \overset{P}{\longrightarrow} g(\boldsymbol{z}_i^T \alpha_0)$, where "$\overset{P}{\longrightarrow}$" denotes convergence in probability.*

Theorem 3. *Suppose that Assumptions A1–A4 hold. Then, we have $\sqrt{N}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \overset{D}{\longrightarrow} N(0, \boldsymbol{\Sigma}_{\theta_0}^{-1} \boldsymbol{\Omega}_{\theta_0} \boldsymbol{\Sigma}_{\theta_0}^{-1})$, where "$\overset{D}{\longrightarrow}$" denotes convergence in distribution, and*

$$\boldsymbol{\Sigma}_{\theta_0} = \lim_{N \to \infty} E\left\{ \frac{1}{N} \frac{\partial^2 \tilde{L}_N(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} \right\}, \quad \boldsymbol{\Omega}_{\theta_0} = \lim_{N \to \infty} E\left\{ \frac{1}{N} \frac{\partial \tilde{L}_N(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} \frac{\partial \tilde{L}_N(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}^T} \right\}, \tag{3.10}$$

which are defined in the appendix. Additionally, if ϵ_i follows $N(0, \sigma_0^2)$, then $\sqrt{N}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \overset{D}{\longrightarrow} N(0, \boldsymbol{\Sigma}_{\theta_0}^{-1})$.

The proofs of all theorems can be found in the Appendix. Theorem 1 and Theorem 2 guarantee the consistency of the parameter $\hat{\boldsymbol{\theta}}$ and the link function $\hat{g}$, respectively, while Theorem 3 provides the asymptotic distribution of $\hat{\boldsymbol{\theta}}$, which can be used for statistical inference. By substituting the corresponding parts in $\boldsymbol{\Omega}_{\theta_0}$ and $\boldsymbol{\Sigma}_{\theta_0}$ with estimators, we can obtain the sample analog of the asymptotic variance.

4. Simulation studies

In this section, we investigate the finite sample performance of the proposed estimation procedures via Monte Carlo simulation studies. The data is generated from the following model:

$$\boldsymbol{Y}_N = \lambda_0 \boldsymbol{W}_N \boldsymbol{Y}_N + \boldsymbol{X}_N \boldsymbol{\beta}_0 + g(\boldsymbol{Z}_N \alpha_0) + \boldsymbol{V}_N, \tag{4.1}$$

where $\boldsymbol{X}_N = (\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_N)^T$ and $\boldsymbol{x}_i = (x_{i1}, x_{i2})^T$, with x_{i1} and x_{i2} values generated from $N(0, 1)$, $i = 1, 2, \ldots, N$; $\boldsymbol{Z}_N = (\boldsymbol{z}_1, \boldsymbol{z}_2, \ldots, \boldsymbol{z}_N)^T$ and $\boldsymbol{z}_i = (z_{i1}, z_{i2}, z_{i3})^T$, with z_{i1}, z_{i2}, and z_{i3} values generated from $U(0, 1)$, $i = 1, 2, \ldots, N$; $\boldsymbol{\beta}_0 = (1, -1)^T$; $\alpha_0 = (1, -1, 1)^T$; and $\boldsymbol{V}_N \sim N(0, \sigma^2 \boldsymbol{I}_N)$. In the simulation, similar to Carroll et al. [14] and Yu et al. [15], suppose $g(t) = \sin\left(\frac{\pi t - A}{C - A}\right)$, where $A = \sqrt{3}/2 - 1.645/\sqrt{12}$, $C = \sqrt{3}/2 + 1.645/\sqrt{12}$. We focus on the spatial scenario in Case [16] where there are R districts and each district contains m members. The sample size is $N = mR$. The weight matrix is $\boldsymbol{W}_N = \boldsymbol{I}_R \otimes \boldsymbol{B}_m$, where $\boldsymbol{B}_m = (\boldsymbol{l}_m \boldsymbol{l}_m^T - \boldsymbol{I}_m)/(m-1)$, $\boldsymbol{l}_m$ is the m-dimensional column vector of ones, and $\otimes$ is the Kronecker product. The numerical study examined different values of R from 10, 30, and 50, and m from 20, 40, and 60. For comparison, three different values are considered for λ_0, that is, $\lambda_0 = 0.1$, 0.4, 0.7, which represents from weak to strong spatial dependence of the responses. $\lambda_0 = 0.1$ represents weak spatial dependence, $\lambda_0 = 0.4$ represents mild spatial dependence, and $\lambda_0 = 0.7$ represents relatively strong spatial dependence.

We use the empirical bias (denoted as Bias) and the empirical standard deviation (denoted as SD) to measure the effectiveness of the parameter estimates. Additionally, we have calculated the asymptotic standard deviation (denoted as AsySD) of the parameter estimates to verify the validity of the asymptotic theory. To assess the accuracy of the proposed nonparametric estimator, we choose the integrated mean squared error as the criterion for judgment, and its empirical version is defined as

$$\widehat{\text{IMSE}}(\hat{g}) = \frac{1}{K} \sum_{i=1}^{K} [\hat{g}(t_i) - g(t_i)]^2, \tag{4.2}$$

where $K = 500$ is the number of grid points on $[0, 1]$.

Tables 1–9 show the simulation results based on 1000 replications. We can observe the following: (i) As the sample size $N = mR$ increases, the SD of the parameter estimators (except $\hat{\lambda}$) and the $\widehat{IMSE}$ of $\hat{g}$ decreases, with the bias remaining near 0, suggesting the consistency of the estimators. (ii) The performance of $\hat{\lambda}$ improves as R increases, but it does not necessarily improve as m increases. This phenomenon is expected, as only increasing m can violate Assumption A2, leading to the unavailability of asymptotic properties. (iii) The performance of $\hat{\lambda}$ improves as λ_0 increases. This can be understood through the knowledge obtained from the asymptotic variance, as the asymptotic variance of $\hat{\lambda}$ strongly depends on λ_0. (iv) The SD of the estimator for finite-dimensional parameters is positively correlated with σ^2. (v) The empirical standard deviation of the parameter estimates is very close to the asymptotic standard deviation, validating the asymptotic theory presented in Theorem 3. The performance of the nonparametric estimates for g is demonstrated in Figure 1. The true function g and the mean of each estimated $\hat{g}$ over the 1000 replicates are plotted. We only display the figure for $\lambda_0 = 0.7$, $R = 30$, $m = 40$ and $\sigma^2 = 1.0$, as the plots for other occasions are similar. To save spaces, we omit them.

Table 1. Simulation results for $(\lambda_0, \sigma^2) = (0.1, 0.1)$.

σ^2	m	Est	R = 10			R = 20			R = 30		
			Bias ($\times 100$)	SD/$\widehat{IMSE}$ ($\times 100$)	AysSD ($\times 100$)	Bias ($\times 100$)	SD/$\widehat{IMSE}$ ($\times 100$)	AysSD ($\times 100$)	Bias ($\times 100$)	SD/$\widehat{IMSE}$ ($\times 100$)	AysSD ($\times 100$)
		$\hat{\lambda}$	−2.068	7.815	7.424	−0.561	4.134	4.340	−0.357	2.985	3.272
		$\hat{\alpha}_1$	0.081	2.465	2.713	0.040	1.841	1.910	0.007	1.521	1.593
		$\hat{\alpha}_2$	0.077	2.452	3.077	−0.005	1.825	2.428	0.164	1.510	2.128
	20	$\hat{\alpha}_3$	−0.168	2.611	2.652	−0.131	1.799	1.970	0.095	1.564	1.604
		$\hat{\beta}_1$	−0.031	2.467	2.245	−0.021	1.666	1.559	−0.071	1.226	1.276
		$\hat{\beta}_2$	−0.089	2.151	2.140	0.120	1.668	1.620	−0.082	1.392	1.329
		$\hat{g}$		1.740			0.646			0.640	
		$\hat{\lambda}$	−0.964	5.106	5.864	−0.544	4.920	5.494	−0.070	2.663	2.858
		$\hat{\alpha}_1$	−0.408	2.134	2.121	0.145	1.335	1.503	0.058	1.169	1.202
		$\hat{\alpha}_2$	−0.166	1.893	2.824	−0.031	1.327	1.879	−0.123	1.294	1.549
0.1	30	$\hat{\alpha}_3$	0.133	2.074	2.206	−0.225	1.436	1.486	−0.221	1.209	1.234
		$\hat{\beta}_1$	−0.100	1.872	1.865	−0.002	1.329	1.311	−0.083	1.039	1.056
		$\hat{\beta}_2$	−0.032	1.795	1.835	−0.015	1.269	1.246	0.017	1.150	1.074
		$\hat{g}$		1.115			0.308			0.266	
		$\hat{\lambda}$	−0.992	5.732	5.806	−0.595	3.698	4.065	−0.407	3.423	3.369
		$\hat{\alpha}_1$	−0.037	1.800	1.744	−0.057	1.206	1.328	0.019	1.049	1.110
		$\hat{\alpha}_2$	−0.153	1.799	2.235	−0.039	1.222	1.674	0.015	1.076	1.409
	40	$\hat{\alpha}_3$	−0.201	1.815	1.917	−0.021	1.230	1.318	−0.033	1.086	1.064
		$\hat{\beta}_1$	−0.143	1.590	1.577	−0.102	1.162	1.127	−0.030	0.896	0.919
		$\hat{\beta}_2$	0.068	1.506	1.489	0.057	1.164	1.115	−0.010	0.849	0.914
		$\hat{g}$		0.935			0.272			0.194	

Table 2. Simulation results for $(\lambda_0, \sigma^2) = (0.1, 0.5)$.

σ^2	m	Est	R = 10			R = 20			R = 30		
			Bias ($\times100$)	SD/$\widehat{IMSE}$ ($\times100$)	AysSD ($\times100$)	Bias ($\times100$)	SD/$\widehat{IMSE}$ ($\times100$)	AysSD ($\times100$)	Bias ($\times100$)	SD/$\widehat{IMSE}$ ($\times100$)	AysSD ($\times100$)
		$\hat{\lambda}$	−6.699	15.934	14.490	−1.845	8.318	8.602	−1.086	5.995	6.552
		$\hat{\alpha}_1$	0.212	5.748	5.793	0.024	4.215	4.089	0.053	3.354	3.398
		$\hat{\alpha}_2$	0.388	5.727	6.257	0.312	4.187	4.911	0.482	3.456	4.340
	20	$\hat{\alpha}_3$	−0.735	6.232	5.636	−0.163	4.094	4.202	0.124	3.428	3.450
		$\hat{\beta}_1$	−0.142	5.544	5.017	−0.013	3.727	3.485	−0.179	2.742	2.852
		$\hat{\beta}_2$	−0.107	4.823	4.769	0.309	3.730	3.620	−0.162	3.111	2.970
		$\hat{g}$	0.000	13.118	0.000	0.000	3.909	0.000	0.000	3.874	0.000
		$\hat{\lambda}$	−3.985	11.105	11.979	−2.138	9.421	10.243	−0.577	5.524	5.866
		$\hat{\alpha}_1$	−0.527	4.884	4.463	0.214	3.026	3.272	0.039	2.545	2.629
		$\hat{\alpha}_2$	−0.060	4.362	5.489	0.107	3.044	3.953	0.018	2.664	3.316
0.5	30	$\hat{\alpha}_3$	−0.106	4.816	4.547	−0.362	3.297	3.235	−0.196	2.594	2.698
		$\hat{\beta}_1$	−0.203	4.197	4.165	−0.033	2.971	2.929	−0.111	2.318	2.360
		$\hat{\beta}_2$	−0.104	4.009	4.098	−0.041	2.830	2.785	0.009	2.569	2.400
		$\hat{g}$	0.000	6.997	0.000	0.000	1.630	0.000	0.000	1.305	0.000
		$\hat{\lambda}$	−3.703	12.135	11.881	−1.955	7.685	8.209	−1.419	6.876	6.748
		$\hat{\alpha}_1$	−0.216	4.136	3.750	−0.162	2.749	2.901	−0.058	2.351	2.420
		$\hat{\alpha}_2$	−0.161	4.038	4.569	−0.017	2.754	3.549	−0.003	2.286	2.999
	40	$\hat{\alpha}_3$	−0.384	4.132	4.049	−0.053	2.791	2.879	−0.089	2.435	2.328
		$\hat{\beta}_1$	−0.339	3.555	3.524	−0.255	2.599	2.518	−0.081	2.003	2.055
		$\hat{\beta}_2$	0.137	3.378	3.329	0.136	2.605	2.493	−0.020	1.901	2.043
		$\hat{g}$	0.000	5.859	0.000	0.000	1.434	0.000	0.000	0.969	0.000

Table 3. Simulation results for $(\lambda_0, \sigma^2) = (0.1, 1.0)$.

σ^2	m	Est	R=10			R=20			R=30		
			Bias ($\times100$)	SD/$\widehat{IMSE}$ ($\times100$)	AysSD ($\times100$)	Bias ($\times100$)	SD/$\widehat{IMSE}$ ($\times100$)	AysSD ($\times100$)	Bias ($\times100$)	SD/$\widehat{IMSE}$ ($\times100$)	AysSD ($\times100$)
		$\hat{\lambda}$	−9.291	19.214	16.714	−2.673	10.363	10.183	−1.577	7.620	7.865
		$\hat{\alpha}_1$	0.128	8.474	7.703	−0.143	6.207	5.502	0.060	4.861	4.582
		$\hat{\alpha}_2$	0.755	8.554	7.853	0.651	6.177	6.215	0.773	5.033	5.582
	20	$\hat{\alpha}_3$	−1.407	9.367	7.479	−0.182	5.976	5.639	0.075	4.936	4.685
		$\hat{\beta}_1$	−0.242	7.864	7.089	−0.003	5.274	4.929	−0.263	3.882	4.032
		$\hat{\beta}_2$	−0.152	6.828	6.718	0.463	5.283	5.117	−0.222	4.399	4.197
		$\hat{g}$	0.000	49.203	0.000	0.000	11.433	0.000	0.000	9.957	0.000
		$\hat{\lambda}$	−6.432	14.909	14.607	−3.069	11.441	11.455	−1.045	7.182	7.253
		$\hat{\alpha}_1$	−0.812	7.361	5.913	0.201	4.372	4.480	−0.020	3.610	3.628
		$\hat{\alpha}_2$	0.089	6.425	6.731	0.228	4.404	5.217	0.125	3.782	4.457
1.0	30	$\hat{\alpha}_3$	−0.379	7.204	5.935	−0.506	4.767	4.437	−0.211	3.714	3.723
		$\hat{\beta}_1$	−0.306	5.954	5.885	−0.052	4.200	4.140	−0.132	3.277	3.337
		$\hat{\beta}_2$	−0.146	5.670	5.789	−0.066	3.998	3.937	0.003	3.632	3.392
		$\hat{g}$	0.000	20.222	0.000	0.000	3.543	0.000	0.000	2.762	0.000
		$\hat{\lambda}$	−5.679	15.696	14.471	−2.917	10.009	9.900	−2.105	8.628	8.077
		$\hat{\alpha}_1$	−0.460	6.139	5.062	−0.301	3.975	3.990	−0.124	3.331	3.323
		$\hat{\alpha}_2$	−0.308	5.910	5.829	0.010	3.974	4.718	0.039	3.296	4.002
	40	$\hat{\alpha}_3$	−0.811	6.150	5.395	−0.101	4.001	3.962	−0.133	3.497	3.206
		$\hat{\beta}_1$	−0.489	5.027	4.983	−0.370	3.678	3.560	−0.117	2.834	2.905
		$\hat{\beta}_2$	0.168	4.794	4.707	0.194	3.685	3.525	−0.026	2.689	2.888
		$\hat{g}$	0.000	17.027	0.000	0.000	3.131	0.000	0.000	2.042	0.000

Table 4. Simulation results for $(\lambda_0, \sigma^2) = (0.4, 0.1)$.

σ^2	m	Est	R = 10			R = 20			R = 30		
			Bias ($\times 100$)	SD/$\widehat{IMSE}$ ($\times 100$)	AysSD ($\times 100$)	Bias ($\times 100$)	SD/$\widehat{IMSE}$ ($\times 100$)	AysSD ($\times 100$)	Bias ($\times 100$)	SD/$\widehat{IMSE}$ ($\times 100$)	AysSD ($\times 100$)
	20	$\hat{\lambda}$	−0.800	2.980	3.354	−0.368	2.712	3.104	−0.343	2.522	2.746
		$\hat{\alpha}_1$	−0.351	2.597	2.622	−0.026	1.763	1.924	0.041	1.461	1.529
		$\hat{\alpha}_2$	0.018	2.478	3.395	−0.027	1.718	2.302	0.066	1.497	1.866
		$\hat{\alpha}_3$	0.204	2.453	2.675	−0.082	1.812	1.906	−0.029	1.401	1.544
		$\hat{\beta}_1$	−0.037	2.398	2.291	0.019	1.558	1.516	0.012	1.248	1.291
		$\hat{\beta}_2$	0.083	2.486	2.320	0.008	1.684	1.663	−0.008	1.347	1.304
		$\hat{g}$	0.000	1.363	0.000	0.000	0.470	0.000	0.000	0.398	0.000
0.1	30	$\hat{\lambda}$	−0.455	4.187	4.588	−0.859	3.775	3.735	−0.240	2.161	2.486
		$\hat{\alpha}_1$	−0.001	2.042	2.119	−0.190	1.448	1.547	−0.098	1.110	1.249
		$\hat{\alpha}_2$	0.006	2.059	2.674	−0.055	1.355	1.973	0.036	1.130	1.622
		$\hat{\alpha}_3$	−0.100	1.955	2.095	0.083	1.450	1.542	0.100	1.154	1.269
		$\hat{\beta}_1$	0.069	2.136	1.978	−0.044	1.291	1.280	−0.021	1.050	1.094
		$\hat{\beta}_2$	0.036	1.846	1.876	0.040	1.314	1.301	0.070	1.112	1.110
		$\hat{g}$	0.000	1.022	0.000	0.000	0.372	0.000	0.000	0.261	0.000
	40	$\hat{\lambda}$	−1.456	6.342	6.789	−0.285	2.684	2.982	−0.258	1.863	2.066
		$\hat{\alpha}_1$	0.003	1.635	1.765	−0.041	1.220	1.357	0.001	0.989	1.060
		$\hat{\alpha}_2$	0.130	1.687	2.224	−0.047	1.219	1.666	−0.002	0.980	1.441
		$\hat{\alpha}_3$	0.056	1.592	1.756	−0.047	1.323	1.274	−0.028	0.996	1.108
		$\hat{\beta}_1$	−0.190	1.544	1.563	−0.031	1.181	1.154	−0.094	0.901	0.883
		$\hat{\beta}_2$	0.047	1.622	1.612	0.031	1.123	1.127	0.094	0.867	0.881
		$\hat{g}$	0.000	0.544	0.000	0.000	0.224	0.000	0.000	0.204	0.000

Table 5. Simulation results for $(\lambda_0, \sigma^2) = (0.4, 0.5)$.

σ^2	m	Est	R = 10			R = 20			R = 30		
			Bias ($\times 100$)	SD/$\widehat{IMSE}$ ($\times 100$)	AysSD ($\times 100$)	Bias ($\times 100$)	SD/$\widehat{IMSE}$ ($\times 100$)	AysSD ($\times 100$)	Bias ($\times 100$)	SD/$\widehat{IMSE}$ ($\times 100$)	AysSD ($\times 100$)
	20	$\hat{\lambda}$	−2.551	6.428	6.921	−1.347	5.261	5.408	−1.159	5.111	5.512
		$\hat{\alpha}_1$	−1.011	6.023	5.449	−0.153	4.006	4.109	−0.034	3.360	3.341
		$\hat{\alpha}_2$	0.246	5.782	6.641	0.082	3.936	4.687	0.135	3.424	3.954
		$\hat{\alpha}_3$	0.372	5.617	5.599	−0.184	4.094	4.088	−0.119	3.199	3.370
		$\hat{\beta}_1$	−0.018	5.346	5.100	0.071	3.490	3.373	0.011	2.795	2.887
		$\hat{\beta}_2$	0.141	5.562	5.184	0.021	3.764	3.717	−0.023	3.006	2.915
		$\hat{g}$	0.000	9.385	0.000	0.000	2.622	0.000	0.000	2.171	0.000
0.5	30	$\hat{\lambda}$	−2.258	8.623	8.838	−2.215	7.157	6.847	−0.951	4.330	4.794
		$\hat{\alpha}_1$	−0.065	4.623	4.430	−0.370	3.192	3.354	−0.261	2.457	2.736
		$\hat{\alpha}_2$	0.185	4.719	5.178	−0.084	2.953	4.100	0.096	2.542	3.466
		$\hat{\alpha}_3$	−0.312	4.609	4.362	0.028	3.290	3.333	0.190	2.604	2.780
		$\hat{\beta}_1$	0.109	4.793	4.418	−0.113	2.886	2.863	−0.013	2.347	2.444
		$\hat{\beta}_2$	0.074	4.123	4.190	0.116	2.940	2.908	0.157	2.482	2.482
		$\hat{g}$	0.000	6.673	0.000	0.000	1.950	0.000	0.000	1.387	0.000
	40	$\hat{\lambda}$	−4.679	12.349	11.693	−1.092	5.388	5.820	−0.792	3.788	4.170
		$\hat{\alpha}_1$	−0.301	3.719	3.777	−0.129	2.753	2.971	−0.037	2.256	2.313
		$\hat{\alpha}_2$	0.104	3.869	4.524	−0.089	2.741	3.580	0.015	2.238	3.053
		$\hat{\alpha}_3$	0.028	3.842	3.765	−0.166	2.950	2.797	−0.079	2.270	2.409
		$\hat{\beta}_1$	−0.431	3.446	3.493	−0.086	2.642	2.579	−0.208	2.015	1.972
		$\hat{\beta}_2$	0.134	3.638	3.603	0.062	2.513	2.519	0.200	1.938	1.969
		$\hat{g}$	0.000	2.994	0.000	0.000	1.179	0.000	0.000	1.087	0.000

Table 6. Simulation results for $(\lambda_0, \sigma^2) = (0.4, 1.0)$.

σ^2	m	Est	R = 10			R = 20			R = 30		
			Bias (×100)	SD/$\widehat{IMSE}$ (×100)	AysSD (×100)	Bias (×100)	SD/$\widehat{IMSE}$ (×100)	AysSD (×100)	Bias (×100)	SD/$\widehat{IMSE}$ (×100)	AysSD (×100)
1.0	20	$\hat{\lambda}$	−3.973	8.684	8.745	−2.112	6.863	6.549	−1.631	6.150	6.165
		$\hat{\alpha}_1$	−2.137	9.396	7.106	−0.358	5.916	5.517	−0.081	4.870	4.585
		$\hat{\alpha}_2$	0.343	8.941	8.030	0.249	5.800	5.972	0.198	4.913	5.202
		$\hat{\alpha}_3$	0.349	8.572	7.308	−0.300	6.003	5.508	−0.322	4.639	4.610
		$\hat{\beta}_1$	−0.012	7.559	7.181	0.107	4.939	4.750	0.012	3.955	4.083
		$\hat{\beta}_2$	0.166	7.874	7.326	0.032	5.322	5.256	−0.035	4.251	4.123
		$\hat{g}$	0.000	43.795	0.000	0.000	6.492	0.000	0.000	4.961	0.000
	30	$\hat{\lambda}$	−3.646	11.145	10.405	−2.870	8.577	7.700	−1.459	5.481	5.636
		$\hat{\alpha}_1$	−0.124	6.847	5.811	−0.485	4.573	4.584	−0.422	3.511	3.769
		$\hat{\alpha}_2$	0.484	7.000	6.294	−0.068	4.257	5.353	0.143	3.638	4.612
		$\hat{\alpha}_3$	−0.649	6.991	5.713	−0.114	4.712	4.538	0.222	3.732	3.827
		$\hat{\beta}_1$	0.140	6.805	6.238	−0.159	4.085	4.050	−0.009	3.318	3.456
		$\hat{\beta}_2$	0.112	5.826	5.920	0.173	4.159	4.112	0.223	3.506	3.509
		$\hat{g}$	0.000	21.085	0.000	0.000	4.331	0.000	0.000	3.008	0.000
	40	$\hat{\lambda}$	−6.211	14.879	12.560	−1.642	6.765	6.878	−1.179	4.884	5.097
		$\hat{\alpha}_1$	−0.538	5.429	5.084	−0.218	3.955	4.090	−0.075	3.241	3.180
		$\hat{\alpha}_2$	0.190	5.685	5.754	−0.067	3.928	4.791	0.073	3.203	4.054
		$\hat{\alpha}_3$	−0.082	5.625	5.072	−0.273	4.225	3.862	−0.123	3.257	3.299
		$\hat{\beta}_1$	−0.598	4.874	4.937	−0.126	3.738	3.647	−0.295	2.849	2.788
		$\hat{\beta}_2$	0.187	5.147	5.094	0.086	3.556	3.562	0.281	2.741	2.783
		$\hat{g}$	0.000	7.083	0.000	0.000	2.533	0.000	0.000	2.357	0.000

**Table 7.** Simulation results for $(\lambda_0, \sigma^2) = (0.7, 0.1)$.

σ^2	m	Est	R = 10			R = 20			R = 30		
			Bias (×100)	SD/$\widehat{IMSE}$ (×100)	AysSD (×100)	Bias (×100)	SD/$\widehat{IMSE}$ (×100)	AysSD (×100)	Bias (×100)	SD/$\widehat{IMSE}$ (×100)	AysSD (×100)
0.1	20	$\hat{\lambda}$	−0.434	2.291	3.358	−0.129	1.438	1.794	−0.175	1.278	1.570
		$\hat{\alpha}_1$	0.089	2.754	2.642	−0.053	1.828	1.936	0.040	1.540	1.564
		$\hat{\alpha}_2$	0.058	2.484	3.117	−0.066	1.782	2.307	−0.031	1.426	1.877
		$\hat{\alpha}_3$	−0.213	2.694	2.615	−0.098	1.824	1.846	−0.128	1.476	1.583
		$\hat{\beta}_1$	−0.019	2.029	2.065	−0.005	1.676	1.621	0.052	1.291	1.310
		$\hat{\beta}_2$	−0.013	2.145	2.090	−0.057	1.613	1.608	0.040	1.258	1.281
		$\hat{g}$	0.000	1.354	0.000	0.000	0.490	0.000	0.000	0.361	0.000
	30	$\hat{\lambda}$	−0.388	2.051	2.027	−0.264	1.361	1.537	−0.058	1.050	1.155
		$\hat{\alpha}_1$	−0.152	1.990	2.034	−0.091	1.422	1.488	−0.021	1.058	1.231
		$\hat{\alpha}_2$	−0.040	2.004	2.697	0.038	1.453	1.920	−0.069	1.078	1.566
		$\hat{\alpha}_3$	0.014	1.828	2.119	0.075	1.438	1.555	−0.079	1.108	1.187
		$\hat{\beta}_1$	0.078	1.829	1.762	−0.045	1.255	1.294	0.061	1.087	1.081
		$\hat{\beta}_2$	0.008	1.670	1.650	0.085	1.300	1.279	0.000	1.043	1.067
		$\hat{g}$	0.000	0.778	0.000	0.000	0.414	0.000	0.000	0.260	0.000
	40	$\hat{\lambda}$	−0.475	2.033	2.286	−0.179	1.438	1.539	−0.107	1.066	1.206
		$\hat{\alpha}_1$	−0.092	1.655	1.718	−0.118	1.136	1.324	0.014	0.965	1.148
		$\hat{\alpha}_2$	−0.081	1.661	2.348	0.035	1.185	1.757	−0.116	0.933	1.397
		$\hat{\alpha}_3$	−0.061	1.659	1.857	0.117	1.186	1.375	−0.154	0.999	1.078
		$\hat{\beta}_1$	0.120	1.611	1.557	0.011	1.119	1.120	0.030	0.906	0.903
		$\hat{\beta}_2$	−0.067	1.624	1.544	0.006	1.083	1.102	−0.053	0.929	0.894
		$\hat{g}$	0.000	0.479	0.000	0.000	0.297	0.000	0.000	0.169	0.000

Table 8. Simulation results for $(\lambda_0, \sigma^2) = (0.7, 0.5)$.

σ^2	m	Est	R = 10			R = 20			R = 30		
			Bias ($\times100$)	SD/$\widehat{IMSE}$ ($\times100$)	AysSD ($\times100$)	Bias ($\times100$)	SD/$\widehat{IMSE}$ ($\times100$)	AysSD ($\times100$)	Bias ($\times100$)	SD/$\widehat{IMSE}$ ($\times100$)	AysSD ($\times100$)
	20	$\hat{\lambda}$	−1.631	4.793	5.106	−0.544	2.864	3.065	−0.671	2.486	2.616
		$\hat{\alpha}_1$	0.119	6.104	5.610	−0.201	4.160	4.143	−0.096	3.417	3.394
		$\hat{\alpha}_2$	0.226	5.444	6.099	−0.074	4.091	4.709	0.023	3.216	3.951
		$\hat{\alpha}_3$	−0.788	5.997	5.498	−0.323	4.221	3.958	−0.167	3.318	3.427
		$\hat{\beta}_1$	−0.011	4.555	4.618	0.009	3.744	3.614	0.120	2.886	2.929
		$\hat{\beta}_2$	−0.085	4.792	4.653	−0.136	3.601	3.583	0.086	2.803	2.847
		$\hat{g}$	0.000	9.417	0.000	0.000	2.820	0.000	0.000	1.949	0.000
0.5	30	$\hat{\lambda}$	−1.376	4.181	3.938	−0.776	2.758	2.882	−0.283	2.090	2.195
		$\hat{\alpha}_1$	−0.413	4.540	4.326	−0.131	3.248	3.224	−0.110	2.393	2.689
		$\hat{\alpha}_2$	−0.016	4.607	5.387	0.140	3.232	3.990	−0.132	2.436	3.335
		$\hat{\alpha}_3$	−0.122	4.260	4.469	−0.003	3.268	3.362	−0.178	2.522	2.594
		$\hat{\beta}_1$	0.127	4.098	3.940	−0.077	2.805	2.891	0.136	2.432	2.415
		$\hat{\beta}_2$	−0.013	3.724	3.678	0.184	2.910	2.856	0.000	2.325	2.376
		$\hat{g}$	0.000	4.647	0.000	0.000	2.271	0.000	0.000	1.345	0.000
	40	$\hat{\lambda}$	−1.617	4.363	4.359	−0.743	2.908	2.922	−0.383	2.153	2.295
		$\hat{\alpha}_1$	−0.275	3.756	3.704	−0.142	2.699	2.889	−0.003	2.211	2.510
		$\hat{\alpha}_2$	−0.120	3.799	4.818	−0.019	2.799	3.721	−0.208	2.133	2.994
		$\hat{\alpha}_3$	−0.215	3.758	3.983	−0.074	2.762	2.988	−0.331	2.260	2.363
		$\hat{\beta}_1$	0.286	3.601	3.480	0.026	2.500	2.504	0.071	2.024	2.015
		$\hat{\beta}_2$	−0.156	3.634	3.452	0.004	2.422	2.460	−0.122	2.077	1.999
		$\hat{g}$	0.000	2.669	0.000	0.000	1.610	0.000	0.000	0.868	0.000

Table 9. Simulation results for $(\lambda_0, \sigma^2) = (0.7, 1.0)$.

σ^2	m	Est	R = 10			R = 20			R = 30		
			Bias ($\times100$)	SD/$\widehat{IMSE}$ ($\times100$)	AysSD ($\times100$)	Bias ($\times100$)	SD/$\widehat{IMSE}$ ($\times100$)	AysSD ($\times100$)	Bias ($\times100$)	SD/$\widehat{IMSE}$ ($\times100$)	AysSD ($\times100$)
	20	$\hat{\lambda}$	−2.416	6.098	5.732	−0.840	3.611	3.548	−0.989	3.058	2.986
		$\hat{\alpha}_1$	−0.292	9.656	7.409	−0.471	6.112	5.587	−0.218	4.930	4.645
		$\hat{\alpha}_2$	0.100	8.394	7.384	−0.067	6.080	6.011	0.115	4.611	5.217
		$\hat{\alpha}_3$	−1.822	9.440	7.208	−0.578	6.223	5.339	−0.262	4.807	4.675
		$\hat{\beta}_1$	0.028	6.444	6.527	0.011	5.294	5.104	0.163	4.081	4.141
		$\hat{\beta}_2$	−0.226	6.763	6.569	−0.198	5.082	5.060	0.122	3.962	4.014
		$\hat{g}$	0.000	42.319	0.000	0.000	7.594	0.000	0.000	4.378	0.000
1.0	30	$\hat{\lambda}$	−2.066	5.395	4.810	−1.102	3.510	3.423	−0.453	2.651	2.634
		$\hat{\alpha}_1$	−0.715	6.799	5.759	−0.207	4.721	4.401	−0.201	3.424	3.704
		$\hat{\alpha}_2$	0.082	7.043	6.681	0.265	4.655	5.213	−0.179	3.481	4.460
		$\hat{\alpha}_3$	−0.400	6.480	5.907	−0.101	4.718	4.583	−0.299	3.626	3.574
		$\hat{\beta}_1$	0.159	5.811	5.570	−0.106	3.966	4.086	0.191	3.440	3.413
		$\hat{\beta}_2$	−0.020	5.264	5.190	0.264	4.118	4.034	0.001	3.283	3.351
		$\hat{g}$	0.000	14.225	0.000	0.000	5.203	0.000	0.000	2.863	0.000
	40	$\hat{\lambda}$	−2.448	5.830	5.197	−1.135	3.682	3.464	−0.568	2.735	2.725
		$\hat{\alpha}_1$	−0.488	5.445	5.025	−0.175	3.918	3.966	−0.035	3.183	3.449
		$\hat{\alpha}_2$	−0.086	5.558	6.191	0.044	4.085	4.933	−0.262	3.105	4.008
		$\hat{\alpha}_3$	−0.376	5.397	5.378	−0.196	3.996	4.086	−0.495	3.292	3.255
		$\hat{\beta}_1$	0.409	5.096	4.919	0.036	3.533	3.540	0.103	2.860	2.846
		$\hat{\beta}_2$	−0.217	5.140	4.880	0.002	3.425	3.474	−0.172	2.937	2.825
		$\hat{g}$	0.000	6.278	0.000	0.000	3.606	0.000	0.000	1.831	0.000
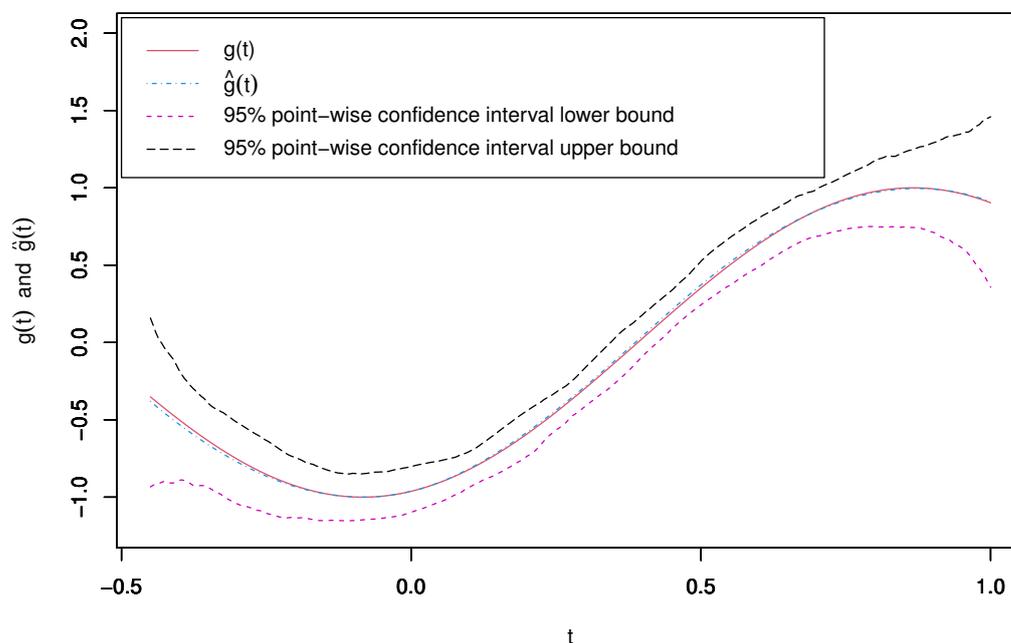
Figure 1. Curve estimate for a single replication of the model simulation study. The true cure g (solid curve), the mean of $\hat{g}$ with B-splines (*degree* = 3) over 1000 simulations ($\lambda_0 = 0.7, R = 30, m = 40, \sigma^2 = 1.0$), and the 95% point-wise confidence interval for g are shown.

5. Real data analysis

In this section, we illustrate the usefulness of the proposed estimation method via analysis of real data, especially for the dependence across spatial units.

5.1. China air quality data

Air quality is gaining more and more attention in China and is closely related to people's daily life. Wu and Kuo [17] utilized statistical models to analyze air quality in Taiwan, China. Chen et al. [18] employed machine learning methods to predict air quality in Hangzhou, China, highlighting the need to consider the impact of neighboring areas on air quality. The factors affecting air quality are complex and can be divided into several categories, such as urban characteristics, meteorological factors, emission factors, and so on. According to the fluidity of the air, the air quality of the core city is related to the surrounding cities and the degree of influence is related to the spatial distance. Therefore, we select the air quality data of cities in China to illustrate the usefulness of the proposed method in Section 2.

The data set contains information of 28 different capital cities in China including air quality index, urban characteristics, meteorological factors, and emission factors from January 1st to April 30th, 2014. There are 13 variables (see Table 10).

Table 10. Variables in China air quality data.

Variables	Explanation (data sources)
TEMP	Temperature [1]
WIND	Wind [1]
RAIN	Rainfall [1]
PEOPLE	Urban resident population/(10,000 people) [2]
CAR	Civil vehicle ownership/vehicle [2]
BUILDING	Building construction area/(10,000 m^2) [2]
SO2	Industrial sulfur dioxide emissions/t [3]
YANCHEN	Industrial soot emissions/t [3]
GREEN	Green area/hm^2 [3]
GN	Dichotomous variable, the heating period in northern area in China(GN=1)
LONG	Longitude coordinates [4]
LATT	Latitude coordinates [4]
AQI	Air quality index [1]

*data sources: (1) China Weather Network http://www.weather.com.cn/, (2) China Regional Economic Statistics Yearbook, (3) China City Statistical Yearbook, (4) http://www.gpsspg.com/maps.htm.

Air quality index (AQI) is the dependent variable. Urban permanent population (PEOPLE), green area (GREEN), civil vehicle ownership (CAR), industrial sulfur dioxide emissions (SO2), industrial soot emissions (YANCHEN), temperature (TEMP), wind (WIND), rainfall (RAIN), building construction area (BUILDING), and the heating problem in northern China (GN) are the independent variables. The model considers the influence of the spatial structure factors of the city on the air quality. According to the research background, we consider the variables log(RAIN), log(PEOPLE), log(CAR), log(BUILDING), GN as the linear independent variables and the variables TEMP, WIND, log(SO2), log(YANCHEN), log(GREEN) as the non-parametric independent variables.

Therefore, we consider fitting the data via the following model:

$$Y_N = \lambda W_N Y_N + \sum_{j=1}^{5} \beta_j X_j + g\left(\sum_{j=1}^{5} \alpha_j Z_j\right) + V_N, \tag{5.1}$$

where Y_N is AQI; X_1, X_2, X_3, X_4, X_5 are log(RAIN), log(PEOPLE), log(CAR), log(BUILDING), GN, respectively; and Z_1, Z_2, Z_3, Z_4, Z_5 are TEMP, WIND, log(SO2), log(YANCHEN), log(GREEN), respectively. Logarithmic transformation for variables is taken to ease off the trouble caused by big gaps in the domain [19].

According to the practice in Du et al. [8], Su and Yang [20], and Xie et al. [21], let the weight matrix $W_N = (w_{ij})$ be

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

where $d_{ij} \geqslant 0$ represents the spatial distance of the ith and jth units, the Euclidean distance is calculated in terms of the longitude and latitude coordinates of any two cities, and d_0 is the threshold distance chosen to median of d_{ij}. We choose the weight matrix in this form as the dependent variable AQI is spatially correlated. That is, when the distance between two cities is far enough, there is tiny spatial

relation between them, and a threshold distance is chosen to set the tiny weight to be zero to alleviate the potential effect of tiny weights on the data analysis. Furthermore, we normalize the weight matrix so that the sum of rows of $\boldsymbol{W}_N$ is equal to 1, i.e., $\sum_{j=1}^{N} w_{ij} = 1$. Obviously, the spatial weight matrix $\boldsymbol{W}_N$ is exogenous. We use the proposed method in Section 2 to analyze the processed data. The analysis results for the estimation of unknown parametric coefficients are listed in Table 11. For the nonparametric component, the univariate function estimates are displayed in Figure 2.

Table 11. Analysis results for China air quality data.

Parameter	λ	log(RAIN) (X_1)	log(PEOPLE) (X_2)	log(CAR) (X_3)	log(BUILDING) (X_4)	GN (X_5)
Estimation	0.261	0.437	0.351	0.120	0.075	1.232
SD	0.134	0.259	0.314	0.247	0.166	0.434
Parameter		TEMP (Z_1)	WIND (Z_2)	log(SO2) (Z_3)	log(YANCHEN) (Z_4)	log(GREEN) (Z_5)
Estimation		0.023	−0.188	0.981	0.027	0.020
SD		0.296	0.329	0.510	0.617	0.324

From the analysis results, the estimation of the spatial parameter is $\hat{\lambda} = 0.261$, which indicates a substantial spatial relationship exists among the AQI of the cities in China. The regression coefficients of $X_1 - X_5$ are positive, implying that the factors of rainfall, urban resident population, civil vehicle ownership/vehicle, building construction area, and the heating period have positive effects on AQI of the cites. The results are generally consistent with our judgment and expectations. As population grows, resource consumption increases, resulting in air pollution. Increased car ownership and construction area lead to increased emissions of pollutants. Increased and large amounts of pollutants during the heating period in the northern area lead to a drop of air quality.

From Figure 2, in the nonparametric component of the model, we observe that the single-index has an "S" shape nonlinear effect on the AQI of the city. It indicates that there exists interaction among the covariates in the nonparametric components. From the results of the single-index coefficients of $Z_1 - Z_5$, the variables temperature, wind, industrial sulfur dioxide emissions, industrial soot emissions, and green area have non-linear effects on the AQI. The single-index coefficients of temperature, industrial sulfur dioxide emissions, industrial soot emissions, green area are positive, and the single-index coefficient of wind are negative. The results are generally consistent with our judgment and expectations. Increased wind power helps the diffusion of pollutants in the air, which improves air quality and leads to a decline of AQI. Increased temperature is not conducive to the decomposition of air pollutants. Increased industrial sulfur dioxide and industrial soot emissions lead to an increase of air pollutants.

However, more rainfall is conducive to reducing the presence of pollutants in the air and increased green plants are conducive to absorption and degradation of air pollutants, which contradicts our data analysis results ($X_1 = 0.437, Z_5 = 0.020$). The model has not undergone variable selection, therefore significance cannot be solely determined by coefficient values. The abnormality of these two coefficients may indicate that they are statistically insignificant.
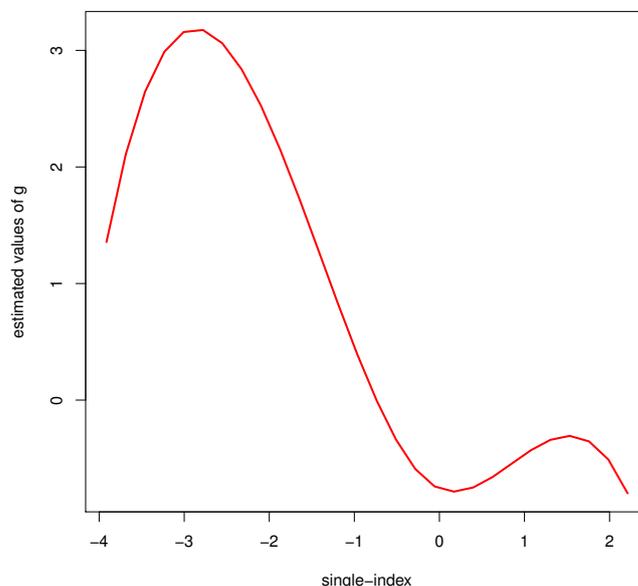
Figure 2. Curve estimate for China air quality data, with $\boldsymbol{Z}^T\hat{\boldsymbol{\alpha}}$ on the x-axis and $\hat{g}(t)$ on the y-axis.

5.2. Rural household income data

The development of the rural economy is significant to the overall economic development in China. Recently, it is facing many problems, such as large populations, low incomes and increasing difficulty in employment. Among them, rural household income is affected by multiple factors. We take the rural household income data of Fuping County in China as an example, including 209 villages, to analyze the factors of rural household income by the proposed method in Section 1. In this model, the dependent variable is the per capita net income of farmers, and the independent variables are 7 factors, including rural natural conditions, village industries, and labor skills, which have great impact on rural household income (see Table 12).

Table 12. Variables in rural household income data.

Variables	Explanation
income	the per capita net income of farmers
agNum	the number of labors engaging in agriculture
urNum	the number of labors left the farm for urban
farm	Farmland
agFa	Facility agriculture
area	Effective irrigation area
indusNum	Number of rural industries
workNum	Number of skilled workers

We consider the variables agNum and farm as the linear independent variables because they represent agriculture-related productivity and means of production in economics, which signify the scale of agriculture. We anticipate that 'income' will exhibit a linear relationship with respect to these two factors. The other variables urNum, agFa, area, indusNum, workNum are the non-parametric independent variables. Next, we fit the data via the following model:

$$Y_N = \lambda W_N Y + X_1 \beta_1 + X_2 \beta_2 + g(Z_1 \alpha_1 + Z_2 \alpha_2 + Z_3 \alpha_3 + Z_4 \alpha_4 + Z_5 \alpha_5) + V_N, \tag{5.2}$$

where Y_N is income; X_1, X_2 are agNum and farm, respectively; and Z_1, Z_2, Z_3, Z_4, Z_5 are urNum, agFa, area, indusNum, and workNum, respectively.

Obtained from spatial geographic information, we can get the neighboring situation of the villages. Therefore,

$$w_{ij} = \begin{cases} 1 & Adjacent \\ 0 & Otherwise \end{cases}, \tag{5.3}$$

where $w_{ij} = 1$ represents adjacent of the ith and jth villages; when two villages are not adjacent, there is tiny spatial relation between them, then $w_{ij} = 0$. As the response variable could be spatially correlated, we choose the weight matrix in this form. Furthermore, we normalize the weight matrix so that the sum of rows of W_N is equal to 1, i.e., $\sum_{j=1}^{N} w_{ij} = 1$. Obviously, the spatial weight matrix W_N is exogenous. The analysis results for the estimation of unknown parametric coefficients are listed in Table 13. For the nonparametric components, the univariate function estimate is displayed in Figure 3.

Table 13. Analysis results for rural household income data.

Parameter	λ	agNum (X_1)	farm (X_2)	urNum (Z_1)	agFa (Z_2)	area (Z_3)	indusNum (Z_4)	workNum (Z_5)
Estimation	0.242	−0.044	−0.228	0.309	−0.349	0.826	0.310	−0.072
SD	0.072	0.073	0.072	0.285	0.189	0.283	0.359	0.367

According to the result, the estimation of the spatial parameter is $\hat{\lambda} = 0.260$ indicates a substantial spatial relationship exists in the model. The regression coefficients of X_1 and X_2 are negative, which implies that the amount of labor engaging in agriculture and farmland would not increase rural household income.

From Figure 3, in the nonparametric parts of the model, we observe that single-index has a nonlinear effect on rural household income. There exists interaction among the covariates in the nonparametric components. The single-index coefficients of $Z_1 - Z_5$ (variables urNum, agFa, area, indusNum, workNum) have significant non-linear effects on rural household income.

Further research is needed to develop a model structure selection method for partially linear single-index spatial autoregressive models. Variable selection decides which covariates have linear effects, and which have nonlinear effects for data analysis.
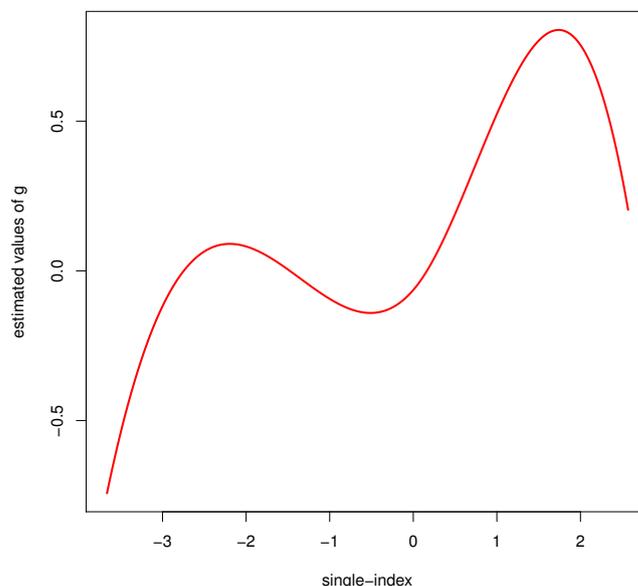
Figure 3. Curve estimate for rural household income data, with $\mathbf{Z}^T\hat{\alpha}$ on the x-axis and $\hat{g}(t)$ on the y-axis.

6. Conclusions

In this paper, B-spline approximation is employed for the link function in the partially linear single-index spatial autoregressive model, offering a smoother and more interpretable alternative compared to local linear approximation. A feasible Nelder-Mead iteration algorithm tailored for this model is proposed to solve the QMLE. The parameter estimation and function estimation have been proven to be consistent, and the asymptotic distribution of parameter estimates for non-normal, general error terms is provided. A series of Monte Carlo simulations are performed to verify the effectiveness of the proposed model and estimation method. Finally, the model and method are applied to air quality datasets and rural income datasets, thereby extending the application scope of the model.

Use of AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article.

Acknowledgments

Conflict of interest

The authors declare there is no conflicts of interest.

References

1. L. F. Lee, Asymptotic distributions of quasi-maximum likelihood estimators for spatial autoregressive models, *Econometrica*, **72** (2004), 1899–1925. https://doi.org/10.1111/j.1468-0262.2004.00558.x

2. L. F. Lee, Consistency and efficiency of least squares estimation for mixed regressive, spatial autoregressive models, *Econometric Theory*, **18** (2002), 252–277. https://doi.org/10.1017/S0266466602182028

3. H. H. Kelejian, I. R. Prucha, A generalized spatial two-stage least squares procedure for estimating a spatial autoregressive model with autoregressive disturbances, *J. Real Estate Finance Econ.*, **17** (1998), 99–121. https://doi.org/10.1023/A:1007707430416

4. L. F. Lee, Best spatial two stage least squares estimators for a spatial autoregressive model with autoregressive disturbances, *Econometric Rev.*, **22** (2003), 307–335. https://doi.org/10.1081/ETC-120025891

5. L. J. Su, Semiparametric GMM estimation of spatial autoregressive models, *J. Econometrics*, **167** (2012), 543–560. https://doi.org/10.1016/j.jeconom.2011.09.034

6. X. Xu, L. F. Lee, Maximum likelihood estimation of a spatial autoregressive tobit model, *J. Econometrics*, **188** (2015), 264–280. https://doi.org/10.1016/j.jeconom.2015.05.004

7. L. J. Su, S. N. Jin, Profile quasi-maximum likelihood estimation of partially linear spatial autoregressive models, *J. Econometrics*, **157** (2010), 18–33. https://doi.org/10.1016/j.jeconom.2009.10.033

8. J. Du, X. Q. Sun, R. Y. Cao, Z. Z. Zhang, Statistical inference for partially linear additive spatial autoregressive models, *Spat. Stat.*, **25** (2018), 52–67. https://doi.org/10.1016/j.spasta.2018.04.008

9. J. L. Wang, L. G. Xue, L. X. Zhu, Y. S. Chong, Estimation for a partial-linear single-index model, *Ann. Stat.*, **38** (2010), 246–274. https://doi.org/10.1214/09-AOS712

10. S. L. Cheng, J. B. Chen, Estimation of partially linear single-index spatial autoregressive model, *Stat. Pap.*, **62** (2021), 495–531. https://doi.org/10.1007/s00362-019-01105-y

11. C. De Boor, *A Practical Guide to Splines*, Springer, 1978.

12. Y. Zhang, C. X. Xia, L. R. Zeng, B-spline estimation for varying-coefficient single-index model, *Chin. J. Appl. Prob. Stat.*, **29** (2013), 433–442. https://doi.org/10.3969/j.issn.1001-4268.2013.09.008

13. R. Q. Tian, M. J. Xia, D. K. Xu, Profile quasi-maximum likelihood estimation for semiparametric varying-coefficient spatial autoregressive panel models with fixed effects, *Stat. Pap.*, **65** (2024), 5109–5143. https://doi.org/10.1007/s00362-024-01586-6

14. R. J. Carroll, J. Q. Fan, I. Gijbels, M. P. Wand, Generalized partially linear single-index models, *J. Am. Stat. Assoc.*, **92** (1997), 477–489. https://doi.org/10.1080/01621459.1997.10474001

15. P. Yu, J. Du, Z. Z. Zhang, Single-index partially functional linear regression model, *Stat. Pap.*, **61** (2020), 1107–1123. https://doi.org/10.1007/s00362-018-0980-6

16. A. C. Case, Spatial patterns in household demand, *Econometrica*, **59** (1991), 953–965. https://doi.org/10.2307/2938168

17. E. Y. Wu, S. L. Kuo, A study on the use of a statistical analysis model to monitor air pollution status in an air quality total quantity control district, *Atmosphere*, **4** (2013), 349–364. https://doi.org/10.3390/atmos4040349

18. L. Chen, H. Y. Long, J. H. Xu, B. Q. Wu, Z. Hang, T. Xing, et al. Deep citywide multisource data fusion-based air quality estimation, *IEEE Trans. Cybern.*, **54** (2024), 111–122. https://doi.org/10.1109/TCYB.2023.3245618

19. J. Wang, L. J. Yang, Efficient and fast spline-backfitted kernel smoothing of additive models, *Ann. Inst. Stat. Math.*, **61** (2009), 663–690. https://doi.org/10.1007/s10463-007-0157-x

20. L. J. Su, Z. L. Yang, *Instrumental Variable Quantile Estimation of Spatial Autoregressive Models*, 2011. Available from: https://ink.library.smu.edu.sg/soe_research/1074

21. T. F. Xie, R. Y. Cao, J. Du, Variable selection for spatial autoregressive models with a diverging number of parameters, *Stat. Pap.*, **61** (2020), 1125–1145. https://doi.org/10.1007/s00362-018-0984-2

22. L. L. Schumaker, *Spline Functions*, Wiley, 1981.

23. H. H. Kelejian, I. R. Prucha, Specification and estimation of spatial autoregressive models with autoregressive and heteroscedastic disturbances, *J. Econometrics*, **157** (2010), 53–67. https://doi.org/10.1016/j.jeconom.2009.10.025

Appendix

Lemma 1. *Suppose that assumptions A1-A4 hold. Then, $\frac{1}{\sqrt{N}}H^T P_{\Pi_0}A_0 = o_p(1)$, for $H = V_N, A_0, X_N, WY_N, \dot{A}(Z_N\alpha)Z_N$.*

Proof. According to Corollary 6.21 in Schumaker [22], we find that $\|A_0 - \Pi_0\gamma_0\| = O_p(\sqrt{N}K^{-r})$. Using this, similar to Lemma 1 of Tian et al. [13], it suffices to show $\|\frac{1}{\sqrt{N}}H^T P_{\Pi_0}A_0\| = o_p(1)$. □

Lemma 2. *Suppose that assumptions A1-A4 hold. Then, $\frac{tr(P_{\Pi}^T P P_{\Pi} x_N P_{\Pi})}{N} = 1 + o(1)$, $\frac{tr(P_{\Pi})}{N} = 1 + o(1)$, $\frac{tr(G_0 P_{\Pi})}{N} = \frac{tr(G_0)}{N} + o(1)$, $\frac{tr(G_0^T P_{\Pi} G_0)}{N} = \frac{tr(G_0^T G_0)}{N} + o(1)$.*

Proof. Given that the eigenvalues of a projection matrix consist solely of 0s and 1s, with the number of 1s equaling the rank of the matrix, and recognizing that the trace of a matrix is equivalent to the sum of its eigenvalues, we can readily prove the lemma. □

Proof of Theorem 1. First, we prove that $\eta \overset{p}{\longrightarrow} \eta_0$. Similar to Su and Jin [7] and Cheng and Chen [10], it suffices to show that

$$\frac{1}{N}(\tilde{L}^c(\eta) - Q(\eta)) = o_p(1), \tag{A1}$$

and

$$\limsup_{N\to\infty} \max_{\eta\in N_\epsilon^c(\eta)} \frac{1}{N}[Q(\eta) - Q(\eta_0)] < 0, \text{ for any } \epsilon > 0. \tag{A2}$$

where $N_\epsilon^c(\eta_0)$ is the complement of an open neighborhood of η with diameter ϵ.

To prove (A1), it is sufficient to show that

$$\hat{\sigma}^2(\eta) - \sigma^{2*}(\eta) = o_p(1). \tag{A3}$$

From (3.4), we have

$$\begin{aligned}
\hat{\sigma}^2(\eta) &= \frac{1}{N}[M(\lambda)Y_N]^T P_\Pi^T P_{P_\Pi X_N} P_\Pi M(\lambda)Y_N. \\
&= \frac{1}{N}\left[(\lambda_0-\lambda)R + M(\lambda)M_0^{-1}V_N + A_0\right]^T P_\Pi^T P_{P_\Pi X_N} P_\Pi \left[(\lambda_0-\lambda)R + M(\lambda)M_0^{-1}V_N + A_0\right].
\end{aligned} \tag{A4}$$

Then,

$$\begin{aligned}
\hat{\sigma}^2(\eta) - \sigma^{2*}(\eta) &= 2(\lambda_0-\lambda)\frac{1}{N}R^T P_\Pi^T P_{P_\Pi X_N} P_\Pi M(\lambda)M_0^{-1}V_N + 2\frac{1}{N}A_0^T P_\Pi^T P_{P_\Pi X_N} P_\Pi M(\lambda)M_0^{-1}V_N. \\
&+ [M(\lambda)M_0^{-1}V_N]^T P_\Pi^T P_{P_\Pi X_N} P_\Pi M(\lambda)M_0^{-1}V_N \\
&- \frac{\sigma_0^2}{N}tr\left[\left(P_\Pi M(\lambda)M_0^{-1}\right)^T P_{P_\Pi X_N} P_\Pi M(\lambda)M_0^{-1}\right] \\
&\equiv 2(\lambda_0-\lambda)H_1(\eta) + 2H_2(\eta) + H_3(\eta) - \sigma_0^2(\eta).
\end{aligned} \tag{A5}$$

Similar to the proof of Theorem 1 in Tian et al. [13], by using Lemma 1, it is easy to prove that $H_2(\eta) = o_p(1), H_1(\eta) = o_p(1)$, and $H_3(\eta) = \sigma_0^2(\eta) + o_p(1)$ such that (A1) holds. Furthermore, referring to Tian et al. [13], based on Lemma 1, we can also prove the validity of (A2). The primary difference lies in the use of an unknown single-index function in our cross-sectional data context, whereas Tian et al. [13] employed a q-dimensional unknown function in their panel data setting.

Next, we prove that $\hat{\sigma}^2 \xrightarrow{P} \sigma_0^2$. Given that we now have $\hat{\lambda} = \lambda_0 + o_p(1)$ and $\hat{\alpha} = \alpha_0 + o_p(1)$, due to the continuity of $z_i^T\alpha$ in terms of α, it follows that $\Pi(Z_N\hat{\alpha}) = \Pi(Z_N\alpha_0)(1 + o_p(1)) \equiv \Pi_0(1 + o_p(1))$. Recall that

$$\begin{aligned}
\hat{\sigma}^2(\eta) &= \frac{1}{N}[M(\lambda)Y_N]^T P_\Pi^T P_{P_\Pi X_N} P_\Pi M(\lambda)Y_N \\
&= \frac{1}{N}[A_0 + V_N + (\lambda_0-\lambda)(R + G_0V_N)]^T P_\Pi^T P_{P_\Pi X_N} P_\Pi [A_0 + V_N + (\lambda_0-\lambda)(R + G_0V_N)],
\end{aligned} \tag{A6}$$

The second equality in the above equation relies on the fact that $M_0^{-1} = I + \lambda_0 G$, and the projection matrix $P_{P_\Pi X_N}$ eliminates the term $X_N\beta_0$. Then, by Lemma 1 and Lemma 2, we have

$$\begin{aligned}
E[\hat{\sigma}^2(\hat{\eta})] &= \frac{1}{N}E\left\{[A_0 + V_N + (\lambda_0-\hat{\lambda})(R + G_0V_N)]^T P_\Pi^T P_{P_\Pi X_N} P_\Pi [A_0 + V_N + (\lambda_0-\hat{\lambda})(R + G_0V_N)]\right\} + o(1) \\
&= E\left\{\frac{1}{N}V_N^T P_\Pi^T P_{P_\Pi X_N} P_\Pi V_N + (\lambda_0-\hat{\lambda})^2\left(\frac{R_N^T P_\Pi^T P_{P_\Pi X_N} P_\Pi R_N}{N} + \frac{V_N^T G_0^T P_\Pi^T P_{P_\Pi X_N} P_\Pi G_0 V_N}{N}\right)\right\} + o(1) \\
&= \sigma_0^2 + o(1).
\end{aligned}$$

and

$$Var[\hat{\sigma}^2(\hat{\boldsymbol{\eta}})] = \frac{1}{N^2} E\left\{[\boldsymbol{A}_0 + \boldsymbol{V}_N]^T \boldsymbol{P}_{\boldsymbol{\Pi}}^T \boldsymbol{P}_{\boldsymbol{P}_{\boldsymbol{\Pi}} X_N} \boldsymbol{P}_{\boldsymbol{\Pi}} [\boldsymbol{A}_0 + \boldsymbol{V}_N]\right\} - \sigma_0^4 + o(1)$$

$$= \frac{1}{N^2}\left[(\mu_4^* - 3\sigma_0^4)\sum_{i=1}^{N} p_{ii} + \sigma_0^4\{tr(\boldsymbol{P}_{\boldsymbol{\Pi}}^T \boldsymbol{P}_{\boldsymbol{P}_{\boldsymbol{\Pi}} X_N} \boldsymbol{P}_{\boldsymbol{\Pi}})\}^2 + 2tr(\boldsymbol{P}_{\boldsymbol{\Pi}}^T \boldsymbol{P}_{\boldsymbol{P}_{\boldsymbol{\Pi}} X_N} \boldsymbol{P}_{\boldsymbol{\Pi}})\right] - \sigma_0^4 + o(1)$$

$$= \frac{\sigma_0^4\{tr(\boldsymbol{P}_{\boldsymbol{\Pi}}^T \boldsymbol{P}_{\boldsymbol{P}_{\boldsymbol{\Pi}} X_N} \boldsymbol{P}_{\boldsymbol{\Pi}})\}^2}{N^2} - \sigma_0^4 + o(1) = o(1).$$

where p_{ii} is the diagonal element of $\boldsymbol{P}_{\boldsymbol{\Pi}}^T \boldsymbol{P}_{\boldsymbol{P}_{\boldsymbol{\Pi}} X_N} \boldsymbol{P}_{\boldsymbol{\Pi}}$. Then, we have $\hat{\sigma}^2(\hat{\boldsymbol{\eta}}) = \sigma_0^2 + o_p(1)$. The proof of consistency for $\hat{\beta}^2(\hat{\boldsymbol{\eta}})$ is analogous to that of $\hat{\sigma}^2(\hat{\boldsymbol{\eta}})$: it suffices to demonstrate that its expectation converges to β_0 and its variance converges to 0. Further elaboration on this point is omitted here for brevity. $\square$

Proof of Theorem 2. According to Theorem 1, we have $\boldsymbol{\Pi}(\boldsymbol{Z}_N\hat{\boldsymbol{\alpha}}) = \boldsymbol{\Pi}(\boldsymbol{Z}_N\boldsymbol{\alpha}_0)(1+o_p(1)) \equiv \boldsymbol{\Pi}_0(1+o_p(1))$. Then,

$$\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0 = [\boldsymbol{\Pi}^T(\boldsymbol{Z}_N\hat{\boldsymbol{\alpha}})\boldsymbol{\Pi}(\boldsymbol{Z}_N\hat{\boldsymbol{\alpha}}))]^{-1}\boldsymbol{\Pi}^T(\boldsymbol{Z}_N\hat{\boldsymbol{\alpha}})[\boldsymbol{M}^{-1}(\hat{\lambda})\boldsymbol{Y}_N - \boldsymbol{X}_N\hat{\boldsymbol{\beta}}] - \boldsymbol{\gamma}_0$$

$$= [\boldsymbol{\Pi}_0^T\boldsymbol{\Pi}_0]^{-1}\boldsymbol{\Pi}_0^T[\boldsymbol{M}^{-1}(\hat{\lambda})\boldsymbol{Y}_N - \boldsymbol{X}_N\hat{\boldsymbol{\beta}}](1 + o_p(1)) - \boldsymbol{\gamma}_0$$

$$= \{[\boldsymbol{\Pi}_0^T\boldsymbol{\Pi}_0]^{-1}\boldsymbol{\Pi}_0^T\boldsymbol{V}_N + [\boldsymbol{\Pi}_0^T\boldsymbol{\Pi}_0]^{-1}\boldsymbol{\Pi}_0^T\boldsymbol{\Delta} + [\boldsymbol{\Pi}_0^T\boldsymbol{\Pi}_0]^{-1}\boldsymbol{\Pi}_0^T\boldsymbol{X}_N(\boldsymbol{\beta}_0 - \hat{\boldsymbol{\beta}})$$

$$+ (\lambda_0 - \hat{\lambda})[\boldsymbol{\Pi}_0^T\boldsymbol{\Pi}_0]^{-1}\boldsymbol{\Pi}_0^T(\boldsymbol{R} + \boldsymbol{G}_0\boldsymbol{V}_N)\}(1 + o_p(1))$$

where $\boldsymbol{\Delta} = \boldsymbol{A}_0 - \boldsymbol{\Pi}_0\boldsymbol{\gamma}_0$. Similar to Theorem 3 of Tian et al. [13], we can prove all the components above are $o_p(1)$, which indicates $\|\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0\| = o_p(1)$.

According to de Boor [11], similar to Theorem 3 in Tian et al. [13], B-splines have the following properties: $B_j(u) \geq 0$, $\sum_{j=1}^{K_N} B_j(u) = 1$ for $u \in \mathbb{U}$, and

$$\frac{C_1}{K_N + l + 1}\sum_{j=1}^{K_N+l+1}\kappa_j^2 \leq \int_{\mathbb{U}}\left(\sum_{j=1}^{K_N+l+1}\kappa_j B_j(u)\right)^2 \mathrm{d}u \leq \frac{C_2}{K_N + l + 1}\sum_{j=1}^{K_N+l+1}\kappa_j^2.$$

where C_1 and C_2 are positive constants. With Cororllary 6.21 of Schumaker [22], we have $\|\boldsymbol{B}(u)^T\boldsymbol{\gamma}_0 - g(u)\| = O(K_N^{-r})$. Then,

$$\|\hat{g} - g\|^2 = \int_{\mathbb{U}}[\hat{g}(u) - g(u)]^2\mathrm{d}u$$

$$= \int_{\mathbb{U}}[\boldsymbol{B}(u)^T\hat{\boldsymbol{\gamma}} - \boldsymbol{B}(u)^T\boldsymbol{\gamma}_0 + \boldsymbol{B}(u)^T\boldsymbol{\gamma}_0 - g(u)]^2\mathrm{d}u$$

$$\leq 2\int_{\mathbb{U}}[\boldsymbol{B}(u)^T(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0)]^2\mathrm{d}u + 2\int_{\mathbb{U}}[\boldsymbol{B}(u)^T\boldsymbol{\gamma}_0 - g(u)]^2\mathrm{d}u$$

$$\leq \frac{2C}{K_N + l + 1}\|\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0\|^2 + 2\|\boldsymbol{B}(u)^T\boldsymbol{\gamma}_0 - g(u)\|^2\mathrm{d}u$$

$$= o_p(1).$$

so $\hat{g}(\boldsymbol{Z}_N^T\hat{\boldsymbol{\alpha}}) \xrightarrow{P} g(\boldsymbol{Z}_N^T\boldsymbol{\alpha_0})$. $\square$

Proof of Theorem 3. By Theorem 3.2 of Lee [1] and Theorem 3 of Tian et al. [13], the asymptotic normality of the parameter $\boldsymbol{\theta}$ can be derived from the Taylor expansion of $\partial\tilde{L}_N(\boldsymbol{\theta})/\partial\boldsymbol{\theta} = 0$. The first-order Taylor expansion at $\boldsymbol{\theta}_0$ is

$$\frac{\partial \tilde{L}_N(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} + \frac{\partial^2 \tilde{L}_N(\boldsymbol{\theta}^*)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) = 0,$$

where $\boldsymbol{\theta}^*$ lies between $\boldsymbol{\theta}_0$ and $\hat{\boldsymbol{\theta}}$. Then,

$$\sqrt{N}(\boldsymbol{\theta} - \boldsymbol{\theta}_0) = -\left(\frac{1}{N} \frac{\partial^2 \tilde{L}_N(\boldsymbol{\theta}^*)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T}\right)^{-1} \frac{1}{\sqrt{N}} \frac{\partial \tilde{L}_N(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} \tag{A7}$$

The proof of the theorem can be accomplished using Slusky's theorem, if it holds that

$$\frac{1}{\sqrt{N}} \frac{\partial \tilde{L}_N(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} \xrightarrow{D} N(0, \boldsymbol{\Omega}_{\boldsymbol{\theta}_0}) \tag{A8}$$

and

$$\frac{1}{N} \frac{\partial^2 \tilde{L}_N(\boldsymbol{\theta}^*)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} \xrightarrow{P} \boldsymbol{\Sigma}_{\boldsymbol{\theta}_0}. \tag{A9}$$

Recall $\boldsymbol{\Pi}_0 = \boldsymbol{\Pi}(\boldsymbol{Z}_N \boldsymbol{\alpha}_0)$ and $\boldsymbol{P}_{\boldsymbol{\Pi}_0} = \boldsymbol{I} - \boldsymbol{\Pi}_0(\boldsymbol{\Pi}_0^T \boldsymbol{\Pi}_0)^{-1} \boldsymbol{\Pi}_0^T$. Comining Lemma 1 and Assumption A1, we have

$$\frac{1}{\sqrt{N}} \frac{\partial \tilde{L}_N(\boldsymbol{\theta}_0)}{\partial \sigma^2} = -\frac{\sqrt{N}}{2\sigma_0^2} + \frac{1}{2\sigma_0^4 \sqrt{N}} \boldsymbol{V}_N^T \boldsymbol{P}_{\boldsymbol{\Pi}_0} \boldsymbol{V}_N + o_p(1),$$

$$\frac{1}{\sqrt{N}} \frac{\partial \tilde{L}_N(\boldsymbol{\theta}_0)}{\partial \lambda} = -\frac{1}{\sqrt{N}} tr\boldsymbol{G}_0 + \frac{1}{\sigma_0^2 \sqrt{N}} (\boldsymbol{R}^T \boldsymbol{P}_{\boldsymbol{\Pi}_0} \boldsymbol{V}_N + \boldsymbol{V}_N^T \boldsymbol{G}_0^T \boldsymbol{P}_{\boldsymbol{\Pi}_0} \boldsymbol{V}_N) + o_p(1),$$

$$\frac{1}{\sqrt{N}} \frac{\partial \tilde{L}_N(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\alpha}} = -\frac{1}{\sigma_0^2 \sqrt{N}} \boldsymbol{Z}_N^T \dot{\boldsymbol{A}}(\boldsymbol{Z}_N \boldsymbol{\alpha}_0) \boldsymbol{P}_{\boldsymbol{\Pi}_0} \boldsymbol{V}_N + o_p(1),$$

$$\frac{1}{\sqrt{N}} \frac{\partial \tilde{L}_N(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\beta}} = \frac{1}{\sigma_0^2 \sqrt{N}} \boldsymbol{X}_N^T \boldsymbol{P}_{\boldsymbol{\Pi}_0} \boldsymbol{V}_N + o_p(1), \tag{A10}$$

and

$$\frac{\partial^2 \tilde{L}_N(\boldsymbol{\theta})}{\partial \sigma^2 \partial \sigma^2} = \frac{N}{2\sigma^4} - \frac{1}{\sigma^6}(\boldsymbol{M}(\lambda)\boldsymbol{Y}_N - \boldsymbol{X}_N \boldsymbol{\beta})^T \boldsymbol{P}_{\boldsymbol{\Pi}}(\boldsymbol{M}(\lambda)\boldsymbol{Y}_N - \boldsymbol{X}_N \boldsymbol{\beta}),$$

$$\frac{\partial^2 \tilde{L}_N(\boldsymbol{\theta})}{\partial \sigma^2 \partial \lambda} = -\frac{1}{\sigma^4}(\boldsymbol{W}_N \boldsymbol{Y}_N)^T \boldsymbol{P}_{\boldsymbol{\Pi}}(\boldsymbol{M}(\lambda)\boldsymbol{Y}_N - \boldsymbol{X}_N \boldsymbol{\beta}),$$

$$\frac{\partial^2 \tilde{L}_N(\boldsymbol{\theta})}{\partial \sigma^2 \partial \boldsymbol{\alpha}} = \frac{1}{\sigma^4} \boldsymbol{Z}_N^T \dot{\boldsymbol{A}}(\boldsymbol{Z}_N \boldsymbol{\alpha}) \boldsymbol{P}_{\boldsymbol{\Pi}}(\boldsymbol{M}(\lambda)\boldsymbol{Y}_N - \boldsymbol{X}_N \boldsymbol{\beta}),$$

$$\frac{\partial^2 \tilde{L}_N(\boldsymbol{\theta})}{\partial \sigma^2 \partial \boldsymbol{\beta}} = -\frac{1}{\sigma^4} \boldsymbol{X}_N^T \boldsymbol{P}_{\boldsymbol{\Pi}}(\boldsymbol{M}(\lambda)\boldsymbol{Y}_N - \boldsymbol{X}_N \boldsymbol{\beta}),$$

$$\frac{\partial^2 \tilde{L}_N(\boldsymbol{\theta})}{\partial \lambda^2} = -\frac{1}{\sigma^2}(\boldsymbol{W}_N \boldsymbol{Y}_N)^T \boldsymbol{P}_{\boldsymbol{\Pi}} \boldsymbol{W}_N \boldsymbol{Y}_N - tr\boldsymbol{G}^2(\lambda), \tag{A11}$$

$$\frac{\partial^2 \tilde{L}_N(\boldsymbol{\theta})}{\partial \lambda \partial \boldsymbol{\alpha}} = \frac{1}{\sigma^2} \boldsymbol{Z}_N^T \dot{\boldsymbol{A}}(\boldsymbol{Z}_N \boldsymbol{\alpha}) \boldsymbol{P}_{\boldsymbol{\Pi}} \boldsymbol{W}_N \boldsymbol{Y}_N, \quad \frac{\partial^2 \tilde{L}_N(\boldsymbol{\theta})}{\partial \lambda \partial \boldsymbol{\beta}} = -\frac{1}{\sigma^2} \boldsymbol{X}_N^T \boldsymbol{P}_{\boldsymbol{\Pi}} \boldsymbol{W}_N \boldsymbol{Y}_N,$$

$$\frac{\partial^2 \tilde{L}_N(\boldsymbol{\theta})}{\partial \boldsymbol{\alpha} \partial \boldsymbol{\alpha}^T} = -\frac{1}{\sigma^2} \boldsymbol{Z}_N^T \dot{\boldsymbol{A}}(\boldsymbol{Z}_N \boldsymbol{\alpha}) \dot{\boldsymbol{A}}(\boldsymbol{Z}_N \boldsymbol{\alpha}) \boldsymbol{Z}_N, \quad \frac{\partial^2 \tilde{L}_N(\boldsymbol{\theta})}{\partial \boldsymbol{\alpha} \partial \boldsymbol{\beta}^T} = \frac{1}{\sigma^2} \boldsymbol{Z}_N^T \dot{\boldsymbol{A}}(\boldsymbol{Z}_N \boldsymbol{\alpha}) \boldsymbol{P}_{\boldsymbol{\Pi}} \boldsymbol{X}_N,$$

$$\frac{\partial^2 \tilde{L}_N(\boldsymbol{\theta})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^T} = -\frac{1}{\sigma^2} \boldsymbol{X}_N^T \boldsymbol{P}_{\boldsymbol{\Pi}} \boldsymbol{X}_N.$$

Notice that the above equation is linear, quadratic, or cubic in terms of $\beta, \alpha, 1/\sigma^2$, except for terms involving λ. Combining $\|\theta^* - \theta_0\| = o_p(1)$, by continuity we can obtain

$$\frac{1}{N} \frac{\partial^2 \tilde{L}_N(\theta^*)}{\partial \theta \partial \theta^T} \xrightarrow{P} \frac{1}{N} \frac{\partial^2 \tilde{L}_N(\theta_0)}{\partial \theta \partial \theta^T}, \tag{A12}$$

if $\frac{1}{N} \frac{\partial^2 \tilde{L}_N(\theta^*)}{\partial \lambda^2} \xrightarrow{P} \frac{1}{N} \frac{\partial^2 \tilde{L}_N(\theta_0)}{\partial \lambda^2}$. To show this, by the mean value theorem we have

$$\frac{1}{N} \frac{\partial^2 \tilde{L}_N(\theta^*)}{\partial \lambda^2} - \frac{1}{N} \frac{\partial^2 \tilde{L}_N(\theta_0)}{\partial \lambda^2} = \left(\frac{1}{\sigma_0^2} - \frac{1}{\sigma^{2*}} \right) \frac{1}{N} (W_N Y_N)^T P_\Pi W_N Y_N - \frac{2}{N} (\lambda^* - \lambda_0) tr[G^3(\tilde{\lambda})]. \tag{A13}$$

where $\tilde{\lambda}$ lies between λ^* and λ_0. According to Assumption A2, it easy to know that $tr[G^3(\tilde{\lambda})]/n = O_p(h_N^{-1}) = o_p(1)$, so (A12) is proved. To show

$$\frac{1}{N} \frac{\partial^2 \tilde{L}_N(\theta_0)}{\partial \theta \partial \theta^T} \xrightarrow{P} E \left\{ \frac{1}{N} \frac{\partial^2 \tilde{L}_N(\theta_0)}{\partial \theta \partial \theta^T} \right\}, \tag{A14}$$

we have

$$-E \left\{ \frac{1}{N} \frac{\partial^2 \tilde{L}_N(\theta_0)}{\partial \theta \partial \theta^T} \right\}$$

$$= \begin{bmatrix} \frac{tr P_{\Pi_0}}{N\sigma_0^4} + \frac{A_0^T P_{\Pi_0} A_0}{N\sigma_0^6} - \frac{1}{2\sigma_0^4} & * & * & * \\ \frac{R^T P_{\Pi_0} A_0}{N\sigma_0^4} + \frac{tr[G_0^T P_{\Pi_0}]}{n\sigma_0^2} & \frac{R^T P_{\Pi_0} R}{N\sigma_0^2} + \frac{tr G_0^2}{N} + \frac{\sigma_0^2 tr[G_0^T P_{\Pi_0} G_0]}{N} & * & * \\ -\frac{Z_N^T \dot{A}(Z_N \alpha) P_{\Pi_0} A_0}{N\sigma_0^4} & -\frac{Z_N^T \dot{A}(Z_N \alpha) P_{\Pi_0} R}{N\sigma_0^2} & \frac{Z_N^T \dot{A}(Z_N \alpha) \dot{A}(Z_N \alpha) Z_N}{N\sigma_0^2} & * \\ \frac{X_N^T P_{\Pi_0} A_0}{N\sigma_0^2} & \frac{X_N^T P_{\Pi_0} R}{N\sigma_0^2} & -\frac{X_N^T P_{\Pi_0} \dot{A}(Z_N \alpha) Z_N}{N\sigma_0^2} & \frac{X_N^T P_{\Pi_0} X_N}{N\sigma_0^2} \end{bmatrix} \tag{A15}$$

$$= \begin{bmatrix} \frac{1}{2\sigma_0^4} & * & * & * \\ \frac{tr G_0}{n\sigma_0^2} & \frac{R^T P_{\Pi_0} R}{N\sigma_0^2} + \frac{tr G_0^2}{N} + \frac{\sigma_0^2 tr[G_0^T G_0]}{N} & * & * \\ 0 & -\frac{Z_N^T \dot{A}(Z_N \alpha) P_{\Pi_0} R}{N\sigma_0^2} & \frac{Z_N^T \dot{A}(Z_N \alpha) \dot{A}(Z_N \alpha) Z_N}{N\sigma_0^2} & * \\ 0 & \frac{X_N^T P_{\Pi_0} R}{N\sigma_0^2} & -\frac{X_N^T P_{\Pi_0} \dot{A}(Z_N \alpha) Z_N}{N\sigma_0^2} & \frac{X_N^T P_{\Pi_0} X_N}{N\sigma_0^2} \end{bmatrix} + o(1)$$

$$\rightarrow \Sigma_{\theta_0},$$

which is the result of Lemma 2. The existence of Σ_{θ_0} is guaranteed by Assumptions A2 and A4. According to the law of large numbers, (A14) holds, and thus (A9) holds.

It is easy to see that the terms of (A10) are linear or quadratic functions of V_N, and therefore their means are all $o_p(1)$. By the central limit theorem of Kelejian and Prucha [23], we have

$$\frac{1}{\sqrt{N}} \frac{\partial \tilde{L}_N(\theta_0)}{\partial \theta} \xrightarrow{D} N(0, \Omega_{\theta_0}), \tag{A16}$$

where

$$E \left\{ \frac{1}{N} \frac{\partial \tilde{L}_N(\theta_0)}{\partial \theta} \frac{\partial \tilde{L}_N(\theta_0)}{\partial \theta^T} \right\} = \begin{bmatrix} \Omega_{11} & * & * & * \\ \Omega_{21} & \Omega_{22} & * & * \\ \Omega_{31} & \Omega_{32} & \Omega_{33} & * \\ \Omega_{41} & \Omega_{42} & \Omega_{43} & \Omega_{44} \end{bmatrix} + o(1) \rightarrow \Omega_{\theta_0}, \tag{A17}$$

with

$$\Omega_{11} = \frac{1}{2\sigma_0^4} + \frac{\mu_4^* - 3\sigma_0^4}{4\sigma_0^8}, \quad \Omega_{21} = \frac{tr\boldsymbol{G}_0}{n\sigma_0^2} + \frac{(\mu_4^* - 3\sigma_0^4)tr\boldsymbol{G}_0}{2N\sigma_0^6} + \frac{\mu_3^* \boldsymbol{R}^T \boldsymbol{P}_{\boldsymbol{\Pi}_0} diag(\boldsymbol{P}_{\boldsymbol{\Pi}_0})}{2N\sigma_0^6},$$

$$\Omega_{22} = \frac{\boldsymbol{R}^T \boldsymbol{P}_{\boldsymbol{\Pi}_0} \boldsymbol{R}}{N\sigma_0^2} + \frac{tr\boldsymbol{G}_0^2 + tr\boldsymbol{G}_0^T \boldsymbol{G}_0}{N} + \frac{(\mu_4^* - 3\sigma_0^4)\sum_{i=1}^N G_{ii}^2}{N\sigma_0^4} + \frac{2\mu_3^* \boldsymbol{R}^T \boldsymbol{P}_{\boldsymbol{\Pi}_0} diag(\boldsymbol{G}_0^T \boldsymbol{P}_{\boldsymbol{\Pi}_0})}{N\sigma_0^4},$$

$$\boldsymbol{\Omega}_{31} = \frac{\mu_3^* \boldsymbol{Z}_N^T \dot{A}(\boldsymbol{Z}_N \boldsymbol{\alpha}_0) \boldsymbol{P}_{\boldsymbol{\Pi}_0} diag(\boldsymbol{P}_{\boldsymbol{\Pi}_0})}{2N\sigma_0^6}, \quad \boldsymbol{\Omega}_{32} = -\frac{\boldsymbol{Z}_N^T \dot{A}(\boldsymbol{Z}_N \boldsymbol{\alpha}_0) \boldsymbol{P}_{\boldsymbol{\Pi}_0} \boldsymbol{R}}{N\sigma_0^2} - \frac{\mu_3^* \boldsymbol{Z}_N^T \dot{A}(\boldsymbol{Z}_N \boldsymbol{\alpha}_0) \boldsymbol{P}_{\boldsymbol{\Pi}_0} diag(\boldsymbol{G}_0 \boldsymbol{P}_{\boldsymbol{\Pi}_0})}{N\sigma_0^4}$$

$$\boldsymbol{\Omega}_{33} = \frac{\boldsymbol{Z}_N^T \dot{A}(\boldsymbol{Z}_N \boldsymbol{\alpha}_0) \boldsymbol{P}_{\boldsymbol{\Pi}_0} \dot{A}(\boldsymbol{Z}_N \boldsymbol{\alpha}_0) \boldsymbol{Z}_N}{N\sigma_0^2}, \quad \boldsymbol{\Omega}_{41} = \frac{\mu_3^* \boldsymbol{X}_N^T \boldsymbol{P}_{\boldsymbol{\Pi}_0} diag(\boldsymbol{P}_{\boldsymbol{\Pi}_0})}{2N\sigma_0^6},$$

$$\boldsymbol{\Omega}_{42} = \frac{\boldsymbol{X}_N^T \boldsymbol{P}_{\boldsymbol{\Pi}_0} \boldsymbol{R}}{N\sigma_0^2} + \frac{\mu_3^* \boldsymbol{X}_N^T \boldsymbol{P}_{\boldsymbol{\Pi}_0} diag(\boldsymbol{G}_0 \boldsymbol{P}_{\boldsymbol{\Pi}_0})}{N\sigma_0^4}, \quad \boldsymbol{\Omega}_{43} = -\frac{\boldsymbol{X}_N^T \boldsymbol{P}_{\boldsymbol{\Pi}_0} \dot{A}(\boldsymbol{Z}_N \boldsymbol{\alpha}_0) \boldsymbol{Z}_N}{N\sigma_0^2}, \quad \boldsymbol{\Omega}_{44} = \frac{\boldsymbol{X}_N^T \boldsymbol{P}_{\boldsymbol{\Pi}_0} \boldsymbol{X}_N}{N\sigma_0^2}.$$

It is not difficult to observe that when ϵ_i follows $N(0, \sigma_0^2)$, we have $\mu_3^* = 0, \mu_4^* = 3\sigma_0^4$, i.e., $\boldsymbol{\Omega}_{\boldsymbol{\theta}_0} = \boldsymbol{\Sigma}_{\boldsymbol{\theta}_0}$. $\quad\square$