## Research article

# Statistical inference in functional semiparametric spatial autoregressive model 

Gaosheng Liu ${ }^{1, *}$ and Yang Bai ${ }^{2}$<br>${ }^{1}$ School of Sciences, Tianjin University of Commerce, Tianjin, 300134, China<br>${ }^{2}$ School of Statistics and Management, Shanghai University of Finance and Economics, Shanghai 200433, China

* Correspondence: Email: 772936104@qq.com; Tel: +8613122900767.


#### Abstract

Semiparametric spatial autoregressive model has drawn great attention since it allows mutual dependence in spatial form and nonlinear effects of covariates. However, with development of scientific technology, there exist functional covariates with high dimensions and frequencies containing rich information. Based on high-dimensional covariates, we propose an interesting and novel functional semiparametric spatial autoregressive model. We use B-spline basis function to approximate the slope function and nonparametric function and propose generalized method of moments to estimate parameters. Under certain regularity conditions, the asymptotic properties of the proposed estimators are obtained. The estimators are computationally convenient with closed-form expression. For slope function and nonparametric function estimators, we propose the residual-based approach to derive its pointwise confidence interval. Simulation studies show that the proposed method performs well.


Keywords: semiparametric spatial autoregressive model; functional data analysis; B-spline approximation; generalized method of moments
Mathematics Subject Classification: 62G05, 62J05, 62M30

## 1. Introduction

Spatial autoregressive model (SAR) and its derivatives have been widely used in many areas such as economics, political science, public health and so on. There are lots of literatures concerning about spatial autoregressive models such as Anselin [1], LeSage [19], Anselin and Bera [2], Lee and Yu [20], LeSage and Pace [21], Lee [22], Dai, Li and Tian [9]. In particular, Lee [22] utilised generalized method of moments to make inference about spatial autoregressive model. Xu and Li [25] investigated the instrumental variable (IV) and maximum likelihood estimators for spatial autoregressive model using a nonlinear transformation of dependent variable. However, spatial autoregressive model may
not be flexible enough to capture nonlinear impact of some covariates since its parametric structure. In order to enrich model adaptability and flexibility, some semiparametric spatial autoregressive models have been proposed. For example, Su [31] studied a semiparametric SAR, which includes nonparametric covariates. Su and Jin [32] proposed partially linear SAR with both linear covariates and nonparametric explanatory variables. Sun et al. [33] studied a semiparametric spatial dynamic model with a profile likelihood approach. Wei and Sun [36] derived semiparametric generalized method of moments estimator. Hoshino [35] proposed a semiparametric series generalized method of moments estimator and established consistency and asymptotic normality of the proposed estimator.

However, with development of economic and scientific technology, huge amounts of data can be easily collected and stored. In particular, some types of data are observed in high dimensions and frequencies containing rich information. We usually call them functional data. When those types data are included in a model as covariates, it is common to use functional linear model (FLM). There exist vast literatures on estimation and prediction for FLM (See, for example, Reiss et al. [26], Ramsay and Dalzell [27], Delaigle and Hall [11], Aneiros-Pérez and Vieu [3]). Many methods were proposed to estimate the slop function such as Cardot et al. [5], Hall and Horowitz [14], Crambes et al. [8], Shin [28]. In particular, Hall and Horowitz [14] established minimax convergence rates of estimation. Cai and Hall [6] proposed functional principle components method and a reproducing kernel Hilbert space approach was used in Yuan and Cai [7].

In many applications of spatial data, there are often covariates with nonlinear effect and functional explanatory variables. This motivates us to propose an interesting and novel functional semiparametric spatial autoregressive model. The model is relatively flexible because it utilises functional linear model to deal with functional covariate and semiparametric SAR model to allow spatial dependence and nonlinear effect of scalar covariate. Recently, some models consider both functional covariates and spatial dependence. For instance, Pineda-Ríos [24] proposed functional SAR model and used least squares and maximum likelihood methods to estimate parameters. The functional SAR model considers spatial effect for error term instead of spatial effect for response variable. Huang et al. [12] considered spatial functional linear model and they developed an estimation method based on maximum method and functional principle component analysis. Hu et al. [13] developed generalized methods of moments to estimate parameters in spatial functional linear model. In the paper, we proposed a generalized method of moments estimator which is heteroskedasticity robust and takes a closed-form written explicitly.

The rest of the paper is organized as follows. Section 2 introduces the proposed model and the estimation procedure. The asymptotic properties of the proposed estimators are established in Section 3. Section 4 conducts simulation studies to evaluate the empirical performance of the proposed estimators. Section 5 gives some discussions about the model. All technical proofs are provided in appendix.

## 2. Model and estimation method

### 2.1. Model and notations

Consider the following novel functional semiparametric spatial autoregressive model,

$$
\begin{equation*}
\boldsymbol{Y}=\rho \boldsymbol{W} \boldsymbol{Y}+\boldsymbol{Z}^{\prime} \boldsymbol{\theta}+\int_{0}^{1} X(t) \beta(t) d t+g(\boldsymbol{U})+\boldsymbol{\varepsilon} \tag{2.1}
\end{equation*}
$$

where $\boldsymbol{Y}$ is a response variable, $\rho$ is an unknown coefficient of the spatial neighboring effect, $\boldsymbol{W}$ is the constant spatial weight matrix with a zero diagonal, $\boldsymbol{Z}=\left(Z_{1}, \ldots, Z_{p}\right)^{\prime}$ is a $p$-dimensional covariate and $\boldsymbol{\theta}$ is its coefficient. $\{X(t): t \in[0,1]\}$ is a zero-mean and second-order (i.e. $\left.E|X(t)|^{2}<\infty, \forall t \in[0,1]\right)$ stochastic process defined on $(\Omega, \mathcal{B}, P)$ with sample paths in $L^{2}[0,1]$, the Hilbert space containing integrable functions with inner product $\langle x, y\rangle=\int_{0}^{1} x(t) y(t) d t, \forall x, y \in L^{2}[0,1]$ with norm $\|x\|=\langle x, x\rangle^{1 / 2}$. The slope function $\beta(t)$ is a square integrable function on $[0,1], \boldsymbol{U}$ is a random variable, $g(\cdot)$ is an unknown function on its support $[0,1]$ without loss of generality. We assume $E[g(\cdot)]=0$ to ensure the identifiability of the nonparametric function. $\boldsymbol{\varepsilon}$ is a random error with zero mean and finite variance $\sigma^{2}$, independent of $\boldsymbol{Z}, \boldsymbol{U}$ and $X(t)$.

Remark 1. The model (2.1) is more flexible to take different models. It generalizes both semiparametric spatial autoregressive model [32] and functional partial linear model [28] which correspond to the cases $\beta(t)=0$ and $\rho=0$, respectively. The model can be represented by $\boldsymbol{Y}=(\boldsymbol{I}-\rho \boldsymbol{W})^{-1} \int_{0}^{1} \boldsymbol{X}(t) \beta(t) d t+(\boldsymbol{I}-\rho \boldsymbol{W})^{-1} \boldsymbol{Z}^{\prime} \boldsymbol{\theta}+(\boldsymbol{I}-\rho \boldsymbol{W})^{-1} g(\boldsymbol{U})+(\boldsymbol{I}-\rho \boldsymbol{W})^{-1} \boldsymbol{\varepsilon}$. We assume $\boldsymbol{I}-\rho \boldsymbol{W}$ could be inverted to make the presentation valid. Thus $Y_{i}$ is also influenced by its neighbours' covariates $X_{j}(t)$ as $j \neq i$. Parameter $\rho$ indicates the basic impact of the neighbours. Greater absolute value of $\rho$ means that the response variable is more likely to be affected by its neighbours.

### 2.2. Estimation method

In the section, we give a method to estimate unknown parameter $\rho$ and $\boldsymbol{\theta}$, slope function $\beta(\cdot)$ and nonparametric function $g(\cdot)$. We use B-spline basis function to approximate $g(\cdot)$ and $\beta(\cdot)$. Let $0=u_{0}<$ $u_{1} \ldots<u_{k_{1}+1}=1$ be a partition of interval [0,1]. Using $u_{i}$ as knots, we have $N_{1}=k_{1}+l_{1}+1$ normalized B-spline basis functions of order $l_{1}+1$ that from a basis function for the linear spline space. Put the basis function as a vector $\boldsymbol{B}_{1}(t)=\left(B_{11}(t), \ldots, B_{1 N_{1}}(t)\right)^{\prime}$ and then the slope function $\beta(\cdot)$ is approximated by $\boldsymbol{B}_{1}^{\prime}(\cdot) \gamma$. Similarly, let $\boldsymbol{B}_{2}(u)=\left(B_{21}(u), \ldots, B_{2 N_{2}}(u)\right)^{\prime}$ be normalized B-spline basis function vector determined by $k_{2}$ interior knots in $[0,1]$ and the order $l_{2}+1$ to approximate $g(\cdot)$, where $N_{2}=k_{2}+l_{2}+1$. Then it follows that

$$
\beta(t) \approx \boldsymbol{B}_{1}^{\prime}(t) \gamma, \quad g(u) \approx \boldsymbol{B}_{2}^{\prime}(u) \boldsymbol{\zeta}
$$

where $\boldsymbol{\gamma}=\left(\gamma_{1}, \ldots, \gamma_{N_{1}}\right)^{\prime}$ and $\zeta=\left(\zeta_{1}, \ldots, \zeta_{N_{2}}\right)^{\prime}$.
Let $\mathbf{D}=\left\langle X(t), \boldsymbol{B}_{1}(t)\right\rangle=\left(\int_{0}^{1} X(t) B_{11}(t) d t, \ldots, \int_{0}^{1} X(t) B_{1 N_{1}}(t) d t\right)^{\prime}, \mathbf{D}_{i}=\left\langle X_{i}(t), \boldsymbol{B}_{1}(t)\right\rangle$. Then the model can be rewritten as

$$
\boldsymbol{Y} \approx \rho \boldsymbol{W} \boldsymbol{Y}+\boldsymbol{Z}^{\prime} \boldsymbol{\theta}+\mathbf{D}^{\prime} \boldsymbol{\gamma}+\boldsymbol{B}_{2}^{\prime}(U) \zeta+\boldsymbol{\varepsilon}
$$

Let $\boldsymbol{P}=\boldsymbol{\Pi}\left(\boldsymbol{\Pi} \boldsymbol{\Pi}^{\prime} \boldsymbol{\Pi}\right)^{-1} \boldsymbol{\Pi}^{\prime}$ denote the projection matrix onto the space by $\boldsymbol{\Pi}$, where $\boldsymbol{\Pi}=\left(\mathbf{D}^{\prime}, \boldsymbol{B}_{2}^{\prime}(U)\right)^{\prime}$. Similar to Zhang and Shen [39], profiling out the functional approximation, we obtain

$$
(\boldsymbol{I}-\boldsymbol{P}) \boldsymbol{Y} \approx \rho(\boldsymbol{I}-\boldsymbol{P}) \boldsymbol{W} \boldsymbol{Y}+(\boldsymbol{I}-\boldsymbol{P}) \boldsymbol{Z}^{\prime} \boldsymbol{\theta}+(\boldsymbol{I}-\boldsymbol{P}) \boldsymbol{\varepsilon}
$$

Let $\mathbf{Q}=(\boldsymbol{W} \boldsymbol{Y}, \boldsymbol{Z})$ and $\boldsymbol{\eta}=\left(\rho, \boldsymbol{\theta}^{\prime}\right)^{\prime}$. Applying the two stage least squares procedure proposed by Kelejian and Prucha [17], we propose the following estimator

$$
\hat{\boldsymbol{\eta}}=\left(\mathbf{Q}^{\prime}(\boldsymbol{I}-\boldsymbol{P}) \boldsymbol{M}(\boldsymbol{I}-\boldsymbol{P}) \mathbf{Q}\right)^{-1} \mathbf{Q}^{\prime}(\boldsymbol{I}-\boldsymbol{P}) \boldsymbol{M}(\boldsymbol{I}-\boldsymbol{P}) \boldsymbol{Y}
$$

where $\boldsymbol{M}=\boldsymbol{H}\left(\boldsymbol{H}^{\prime} \boldsymbol{H}\right)^{-1} \boldsymbol{H}^{\prime}$ and $\boldsymbol{H}$ is matrices of instrumental variables. Moreover,

$$
(\hat{\gamma}, \hat{\zeta})=\left(\boldsymbol{\Pi}^{\prime} \boldsymbol{\Pi}\right)^{-1} \boldsymbol{\Pi}^{\prime}(\boldsymbol{Y}-\mathbf{Q} \hat{\boldsymbol{\eta}}) .
$$

Consequently, we use $\hat{\beta}(t)=\boldsymbol{B}_{1}^{\prime}(t) \hat{\gamma}, \hat{g}(u)=\boldsymbol{B}_{2}^{\prime}(u) \hat{\zeta}$ as the estimator of $\beta(t)$ and $g(u)$.
For statistical inference based on $\hat{\boldsymbol{\eta}}$, consistent estimators of the asymptotic covariance matrices are needed. Define the following estimator

$$
\hat{\sigma}^{2}=\frac{1}{n}\left\|\boldsymbol{Y}-\hat{\rho} \boldsymbol{W} \boldsymbol{Y}-\boldsymbol{Z}^{\prime} \hat{\boldsymbol{\theta}}-\mathbf{D}^{\prime} \hat{\boldsymbol{\gamma}}-\boldsymbol{B}_{2}^{\prime}(U) \hat{\zeta}\right\|^{2},
$$

and

$$
\hat{\boldsymbol{\Sigma}}=\frac{1}{n} \mathbf{Q}^{\prime}(\boldsymbol{I}-\boldsymbol{P}) \boldsymbol{M}(\boldsymbol{I}-\boldsymbol{P}) \mathbf{Q}
$$

where $\|\cdot\|$ is the $L^{2}$ norm for a function or the Euclidean norm for a vector. In order to make statistical inference about $\sigma^{2}$, it need to get the value $\omega=E\left[\left(\varepsilon_{1}^{2}-\sigma^{2}\right)^{2}\right]$. Therefore, we use the following estimator $\hat{\omega}$ to estimate $\omega$

$$
\hat{\omega}=\frac{1}{n} \sum_{i=1}^{n}\left(\hat{\varepsilon}_{i}^{2}-\hat{\sigma}^{2}\right)^{2} .
$$

Similar to Zhang and Shen [39], we use an analogous idea for the construction of instrument variables. In the first step, the following instrumental variables are obtained $\tilde{\boldsymbol{H}}=(\boldsymbol{W}(\boldsymbol{I}-$ $\left.\tilde{\rho} \boldsymbol{W})^{-1}\left(\boldsymbol{Z}, \mathbf{D}^{\prime} \tilde{\gamma}, \boldsymbol{B}_{2}^{\prime}(U) \tilde{\zeta}\right), \boldsymbol{Z}, \boldsymbol{\Pi}\right)$, where $\tilde{\rho}, \tilde{\gamma}$ and $\tilde{\zeta}$ are obtained by simply regressing $\boldsymbol{Y}$ on pseudo regressor variables $\boldsymbol{W} \boldsymbol{Y}, \boldsymbol{\Pi}$. In the second step, instrumental variables $\tilde{\boldsymbol{H}}$ are used to obtain the estimators $\overline{\boldsymbol{\eta}}, \bar{\gamma}$ and $\bar{\zeta}$, which are necessary to construct the following instrumental variables $\boldsymbol{H}=$ $\left(\boldsymbol{W}(\boldsymbol{I}-\bar{\rho} \boldsymbol{W})^{-1}\left(\boldsymbol{Z}^{\prime} \overline{\boldsymbol{\theta}}+\mathbf{D}^{\prime} \overline{\boldsymbol{\gamma}}+\boldsymbol{B}_{2}^{\prime}(U) \overline{\boldsymbol{\zeta}}\right), \boldsymbol{Z}\right)$. Finally, we use the instrumental variables $\boldsymbol{H}$ to obtain the final estimators $\hat{\rho}$ and $\hat{\boldsymbol{\theta}}$.

## 3. Asymptotic theory

In this section, we derive the asymptotic normality and rates of convergence for the estimators defined in previous section. Firstly, we introduce some notations. For convenience and simplicity, $c$ denote a generic positive constant, which may take different values at different places. Let $\beta_{0}(\cdot)$ and $g_{o}(\cdot)$ be the true value of function $\beta(\cdot)$ and $g(\cdot)$ respectively. $K(t, s)=\operatorname{Cov}(X(t), X(s))$ denotes the covariance function of $X(\cdot) . a_{n} \sim b_{n}$ means that $a_{n} / b_{n}$ is bounded away from zero and infinity as $n \rightarrow \infty$. In the paper, we make the following assumptions.

C1 The matrix $\boldsymbol{I}-\rho \boldsymbol{W}$ is nonsingular with $|\rho|<1$.
C2 The row and column sums of the matrices $\boldsymbol{W}$ and $(\boldsymbol{I}-\rho \boldsymbol{W})^{-1}$ are bounded uniformly in absolute value for any $|\rho|<1$.
C3 For matrix $\boldsymbol{S}=\boldsymbol{W}(\boldsymbol{I}-\rho \boldsymbol{W})^{-1}$, there exists a constant $\lambda_{c}$ such that $\lambda_{c} \boldsymbol{I}-\boldsymbol{S} \boldsymbol{S}^{\prime}$ is positive semidefinite for all $n$.
C4 Matrix $\frac{1}{n} \tilde{\mathbf{Q}}^{\prime}(\boldsymbol{I}-\boldsymbol{P}) \boldsymbol{M}(\boldsymbol{I}-\boldsymbol{P}) \tilde{\mathbf{Q}} \rightarrow \boldsymbol{\Sigma}$ in probability for some positive definite matrix, where $\tilde{\mathbf{Q}}=$ $\left(\boldsymbol{W}(\boldsymbol{I}-\rho \boldsymbol{W})^{-1}\left(\boldsymbol{Z}^{\prime} \boldsymbol{\theta}+\int_{0}^{1} X(t) \beta(t) d t+g(\boldsymbol{U})\right), \boldsymbol{Z}\right)$.
C5 For matrix $\tilde{\mathbf{Q}}$, there exsits a constant $\lambda_{c}^{*}$ such that $\lambda_{c}^{*} \boldsymbol{I}-\tilde{Q} \tilde{\mathbf{Q}}^{\prime}$ is positive semidefinite for all $n$.
C6 $X(t)$ has a finite fourth moment, that is, $E\|X(t)\|^{4} \leq c$.

C7 $K(t, s)$ is positive definite.
C8 The nonparametric function $g(\cdot)$ has bounded and continuous derivatives up to order $r(\geq 2)$ and the slope function $\beta(t) \in C^{r}[0,1]$.
C9 The density of $\boldsymbol{U}, f_{U}(u)$, is bounded away from 0 and $\infty$ on $[0,1]$. Furthermore, we assume that $f_{U}(u)$ is continuously differentiable on $[0,1]$.
C10 For the knots number $k_{j},(j=1,2)$, it is assumed $k_{1} \sim k_{2} \sim k$.
Assumptions C1-C3 are required in the setting of spatial autoregressive model (see, for example, Lee [23], Kelejian and Prucha [18], Zhang and Shen [39]). They concern the restriction of spatial weight matrix and SAR parameter. Assumption C4 (see Du et al. [10]) is used to represent the asymptotic covariance matrix of $\hat{\boldsymbol{\eta}}$. Moreover, assumption C4 requires implicitly that the generated regressors $\boldsymbol{C Z}$ and $\boldsymbol{Z}$, deviated from their functional part of projection onto $\boldsymbol{\Pi}$, are not asymptotically multicollinear. Assumption C5 is required to ensure the identifiability of parameter $\boldsymbol{\eta}$. Assumptions C6-C7 are commonly assumed in functional linear model [14]. Assumption C6 is a mild restriction to prove the convergence of our estimator. Assumption C7 guarantees the identifiability of $\beta(t)$. Assumption C 8 ensures that $\beta(\cdot)$ and $g(\cdot)$ are sufficiently smoothed and can be approximated by basis functions in the spline space. Assumption C9 requires a bounded condition on the covariates. It is often assumed in asymptotic analysis of nonparametric regression problems (see, for example [15,37]). Assumption C10 is required to achieve the optimal convergence rate of $\hat{\beta}(\cdot)$ and $\hat{g}(\cdot)$.

Let

$$
\begin{aligned}
\Delta_{n}= & E\left(\mathbf{D D}^{\prime}\right)-E\left\{E\left(\mathbf{D} \boldsymbol{B}_{2}^{\prime}(u) \mid \boldsymbol{U}\right)\left[E\left(\boldsymbol{B}_{2}(u) \boldsymbol{B}_{2}^{\prime}(u) \mid \boldsymbol{U}\right)\right]^{-1} E\left(\boldsymbol{B}_{2}(u) \mathbf{D}^{\prime} \mid \boldsymbol{U}\right)\right\}, \\
& \Omega_{n}=E\left(\boldsymbol{B}_{2} \boldsymbol{B}_{2}^{\prime}\right)-E\left\{E\left(\boldsymbol{B}_{2}(u) \mathbf{D}^{\prime} \mid \mathbf{V}\right)\left[E\left(\mathbf{D} \mathbf{D}^{\prime} \mid \mathbf{V}\right)\right]^{-1} E\left(\mathbf{D} \boldsymbol{B}_{2}^{\prime}(u) \mid \mathbf{V}\right)\right\},
\end{aligned}
$$

where $\mathbf{V}=\left\langle X(t), \beta_{0}(t)\right\rangle$. The following theorems state the asymptotic properties of the estimators for parameter and nonparametric function.

Theorem 1. Suppose assumptions C1-C10 hold, then

$$
\sqrt{n}(\hat{\boldsymbol{\eta}}-\boldsymbol{\eta}) \xrightarrow{d} N\left(0, \sigma^{2} \boldsymbol{\Sigma}^{-1}\right) .
$$

Theorem 2. Assumptions C1-C10 hold and $k \sim n^{\frac{1}{2 r+1}}$, then

$$
\begin{aligned}
& \left\|\hat{\beta}(\cdot)-\beta_{0}(\cdot)\right\|^{2}=O_{p}\left(n^{-\frac{2 r}{2 r+1}}\right) . \\
& \left\|\hat{g}(\cdot)-g_{0}(\cdot)\right\|^{2}=O_{p}\left(n^{-\frac{2 r}{2 r+1}}\right) .
\end{aligned}
$$

Remark 2. Theorem 2 gives the consistency of function estimators. The slope function estimator $\hat{\beta}(\cdot)$ and nonparametric function estimator $\hat{g}(\cdot)$ have the same optimal global convergence rate established by Stone [29].

Theorem 3. Suppose assumptions C1-C10 hold, and $E\left(\left|\varepsilon_{1}\right|^{4+r}\right)<\infty$ for some $r>0$, then

$$
\hat{\sigma}^{2} \xrightarrow{p} \sigma^{2}, \quad \hat{\omega} \xrightarrow{p} \omega, \quad \text { and } \quad \hat{\boldsymbol{\Sigma}} \xrightarrow{p} \boldsymbol{\Sigma} .
$$

Remark 3. From the proof of Theorem 3, if $\operatorname{trace}(\boldsymbol{S}) / n=o(1)$, it can be shown that

$$
\sqrt{n}\left(\hat{\sigma}^{2}-\sigma^{2}\right) \xrightarrow{d} N(0, \omega) .
$$

Theorem 4. Suppose assumptions C1-C10 hold and $n /\left(k_{1}^{2 r+1}\right)=n /\left(k_{2}^{2 r+1}\right)=o(1)$, for any fixed points $t, u \in(0,1)$, as $n \rightarrow \infty$,

$$
\begin{aligned}
& \sqrt{\frac{n}{k_{1}}}\left(\hat{\beta}(t)-\beta^{*}(t)\right) \xrightarrow{d} N(0, \Xi(t)), \\
& \sqrt{\frac{n}{k_{2}}}\left(\hat{g}(u)-g^{*}(u)\right) \xrightarrow{d} N(0, \Lambda(u)) .
\end{aligned}
$$

where $\beta^{*}(t)=\boldsymbol{B}_{1}^{\prime}(t) \boldsymbol{\gamma}_{0}, \quad g^{*}(u)=\boldsymbol{B}_{2}^{\prime}(u) \zeta_{0}, \quad \Xi(t)=\lim _{n \rightarrow \infty} \frac{\sigma^{2}}{k_{1}} \boldsymbol{B}_{1}^{\prime}(t) \Delta_{n} \boldsymbol{B}_{1}(t), \quad \Lambda(u) \quad=$ $\lim _{n \rightarrow \infty} \frac{\sigma^{2}}{k_{2}} \boldsymbol{B}_{2}^{\prime}(u) \Omega_{n} \boldsymbol{B}_{2}(u), \boldsymbol{\gamma}_{0}$ and $\zeta_{0}$ are defined in Lemma 1 of appendix.

Remark 4. The above conclusions is similar to those of Yu et al. [38], which gave the asymptotic normality for spline estimators in single-index partial functional linear regression model. Note that $\hat{\beta}(t)-\beta_{0}(t)=\left(\beta^{*}(t)-\beta_{0}(t)\right)+\left(\hat{\beta}(t)-\beta^{*}(t)\right)$. We obtain that $\beta^{*}(t)-\beta_{0}(t)=O\left(k_{1}^{-r}\right)$ by Lemma 1 on Appendix and $\hat{\beta}(t)-\beta^{*}(t)$ dominates $\beta^{*}(t)-\beta_{0}(t)$. Therefore we can use the asymptotic behaviors of $\hat{\beta}(t)-\beta^{*}(t)$ to describe the asymptotic behaviors of $\hat{\beta}(t)-\beta_{0}(t)$.

The variance $\Xi(t)$ and $\Lambda(u)$ are involved in basic function and knots. Different basis functions and knots can get different variance estimators. Moreover, the variance expression contains unknown quantities. Replacing them by consistent estimators can lead to approximation errors. What's more, there may exist heteroscedasticity in error term and then the estimator $\hat{\sigma}^{2}$ is not consistent. Consequently, we propose the following residual-based method to construct piecewise confidence interval in practice.

It is crucial that spatial structure must be preserved during data resampling in models with spatial dependence [1]. Therefore, we employ the residual-based bootstrap procedure to derive the empirical pointwise standard error of $\hat{\beta}(t)$ and $\hat{g}(\cdot)$. The procedure can be described as follows:
(1) Based on the data sets $\{\boldsymbol{Y}, \boldsymbol{Z}, X(t), \boldsymbol{U}\}$ and spatial matrix $\boldsymbol{W}$, one fits the proposed model and obtain the residual vector $\hat{\boldsymbol{\varepsilon}}_{1}=\left(\hat{\varepsilon}_{11}, \ldots, \hat{\varepsilon}_{n 1}\right)^{\prime}$. Then, we derive the centralized residual vector $\hat{\boldsymbol{\varepsilon}}$.
(2) Draw a bootstrap sample $\hat{\boldsymbol{\varepsilon}}^{*}$ with replacement from the empirical distribution function of $\hat{\boldsymbol{\varepsilon}}$ and generate $\boldsymbol{Y}^{*}=(\boldsymbol{I}-\hat{\rho} \boldsymbol{W})^{-1}\left(\boldsymbol{Z}^{\prime} \boldsymbol{\theta}+\mathbf{D}^{\prime} \hat{\gamma}+\boldsymbol{B}_{2}^{\prime}(U) \hat{\boldsymbol{\zeta}}+\hat{\boldsymbol{\varepsilon}}^{*}\right)$.
(3) Based on the new data sets $\left\{\boldsymbol{Y}^{*}, \boldsymbol{Z}, X(t), \boldsymbol{U}\right\}$ and spatial matrix $\boldsymbol{W}$, we fit the proposed model again to derive the estimator $\hat{\beta}^{*}(t)$ and $\hat{g}^{*}(u)$. Repeat the process many times. Thus, for given $t$ and $u$, calculate empirical variance of $\hat{\beta}^{*}(t)$ and $\hat{g}^{*}(u)$ respectively. Consequently, we use the empirical variance to construct its confidence interval.

## 4. Simulation studies

In this section, we use simulation examples to study the properties of the proposed estimators. The data is generated from the following model:

$$
Y_{i}=\rho \sum_{i=1}^{n} w_{i j} Y_{j}+Z_{1 i} \theta_{1}+Z_{2 i} \theta_{2}+\int_{0}^{1} \beta(t) X_{i}(t) d t+g\left(U_{i}\right)+\varepsilon_{i}, i=1, \ldots, n
$$

where $\rho=0.5, \beta(t)=\sqrt{2} \sin (\pi t / 2)+3 \sqrt{2} \sin (3 \pi t / 2)$ and $X(t)=\sum_{j=1}^{50} \gamma_{j} \phi_{j}(t)$, where $\gamma_{j}$ is distributed as independent normal with mean 0 and variance $\lambda_{j}=((j-0.5) \pi)^{-2}$ and $\phi_{j}(t)=\sqrt{2} \sin ((j-0.5) \pi t)$.
$Z_{i 1}$ and $Z_{i 2}$ are independent and follow standard normal distribution, $\theta_{1}=\theta_{2}=1, U_{i} \sim U(0,1)$, $g(u)=\sin \left(\frac{\pi(u-A)}{C-A}\right), A=\frac{\sqrt{3}}{2}-\frac{1.654}{\sqrt{12}}, C=\frac{\sqrt{3}}{2}+\frac{1.654}{\sqrt{12}}$. The spatial weight matrix $\boldsymbol{W}=\left(w_{i j}\right)_{n \times n}$ is generated based on mechanism that $w_{i j}=0.3^{|i-j|} I(i \neq j), 1 \leq i, j \leq n$ with $w_{i i}=0, i=1, \ldots, n$. A standardized transformation is used to convert the matrix $\boldsymbol{W}$ to have row-sums of unit. We set the following three kinds of error term: (1) $\varepsilon_{i} \sim N\left(0, \sigma^{2}\right)$; (2) $\varepsilon_{i} \sim 0.75 t(3)$; (3) $\varepsilon_{i} \sim\left(1+0.5 U_{i}\right) N\left(0, \sigma^{2}\right)$, where $\sigma^{2}=1$. In order to compare the different situations for magnitudes of $\rho$, we set $\rho=\{0.2,0.7\}$ with error term $N\left(0, \sigma^{2}\right)$. Simulation results are derived based on 1000 replications.

To achieve good numerical performances, the order $l_{1}$ and $l_{2}$ of splines and the number of interior knots $k_{1}$ and $k_{2}$ should be chosen. To reduce the burden of computation, we use the cubic B-spline with four evenly distributed knots (i.e., $k_{1}=k_{2}=2$ ) for slope function $\beta(\cdot)$ and nonparametric function $g(\cdot)$ respectively. These choices of $k_{1}$ and $k_{2}$ are small enough to avoid overfitting in typical problem with sample size not too small and big enough to flexibly approximate many smooth function. We use the square root of average square errors (RASE) to assess the performance of estimators $\hat{\beta}(\cdot)$ and $\hat{g}(\cdot)$ respectively

$$
\begin{aligned}
& \operatorname{RASE}_{1}=\left\{\sum_{i=1}^{n_{1}} \frac{\left(\hat{\beta}\left(t_{i}\right)-\beta\left(t_{i}\right)\right)^{2}}{n_{1}}\right\}^{1 / 2}, \\
& \operatorname{RASE}_{2}=\left\{\sum_{i=1}^{n_{2}} \frac{\left(\hat{g}\left(u_{i}\right)-g\left(u_{i}\right)\right)^{2}}{n_{2}}\right\}^{1 / 2},
\end{aligned}
$$

where $\left\{t_{i}, i=1, \ldots, n_{1}\right\},\left\{u_{i}, i=1, \ldots, n_{2}\right\}$ and $n_{1}=n_{2}=200$ are grid points chosen equally spaced in the domain of $\beta(\cdot)$ and $g(\cdot)$ respectively.

Tables 1-3 show simulation results with different kinds of error terms. Table 4 presents different magnitudes of $\rho$ with error term $N(0,1)$. They show the bias (Bias), standard deviation (SD), standard error (SE) and coverage probability (CP) with nominal level of $95 \%$ for estimator and the mean and standard deviation (SD) of $\operatorname{RASE}_{j}(j=1,2)$ for $\hat{\beta}(\cdot)$ and $\hat{g}(\cdot)$. The simulation results can be summarized as follows:
(1) The estimators $\hat{\rho}, \hat{\theta}_{1}, \hat{\theta}_{2}, \hat{\sigma}^{2}$ are approximately unbiased and the estimated standard errors are close to sample standard deviations in normal error distribution. The empirical coverage probabilities approximate the nominal level of $95 \%$ well.
(2) Figure 1 gives an example of the estimated function curve $\hat{\beta}(\cdot)$ and $\hat{g}(\cdot)$ and its empirical $95 \%$ confidence interval with sample size $n=300$ for error term $N(0,1)$. From the mean and standard deviation (SD) of $\operatorname{RASE}_{j}(j=1,2)$, combined with Figure 1, we conclude that the proposed function estimators $\hat{\beta}(\cdot)$ and $\hat{g}(\cdot)$ perform well.
(3) For error term $0.75 t(3)$ and $\left(1+0.5 U_{i}\right) N(0,1)$, the estimators $\hat{\rho}, \hat{\theta}_{1}, \hat{\theta}_{2}$ are approximately unbiased and the estimated standard errors are close to sample standard deviations. In addition, the mean and standard deviation for RASE of estimated coefficient function $\hat{\beta}(\cdot)$ and $\hat{g}(\cdot)$ are decreasing. It indicates that parametric and non parametric estimators perform well in non-normal error term.
(4) From Table 1 and Table 4, as basic spatial effect $\rho$ increases, the SE and SD of $\hat{\rho}$ decrease. For the different magnitudes of $\rho$, the Bias and SD of parametric estimators for $\hat{\theta}_{1}$ and $\hat{\theta}_{2}$, and the mean of RASE for $\hat{\beta}(\cdot)$ and $\hat{g}(\cdot)$ remain stable. It means that the magnitudes of $\rho$ do not affect the other parametric and nonparametric estimators.

Table 1. Simulation results for $\rho=0.5$ with error term $N(0,1)$.

| $n$ | Est | Bias | SD | SE | CP |
| :--- | :--- | :--- | :--- | :--- | :--- |
| 100 | $\hat{\rho}$ | -0.0091 | 0.0789 | 0.0799 | 0.9430 |
|  | $\hat{\theta}_{1}$ | -0.0065 | 0.1058 | 0.1030 | 0.9500 |
|  | $\hat{\theta}_{2}$ | -0.0012 | 0.1052 | 0.1078 | 0.9430 |
|  | $\hat{\sigma}^{2}$ | -0.1100 | 0.1379 | 0.1232 | 0.7900 |
|  | RASE $_{1}$ | 1.4361 | 0.7046 |  |  |
|  | RASE $_{2}$ | 0.1950 | 0.0683 |  |  |
| 300 | $\hat{\rho}$ | -0.0031 | 0.0441 | 0.0444 | 0.9520 |
|  | $\hat{\theta}_{1}$ | -0.0011 | 0.0594 | 0.0594 | 0.9580 |
|  | $\hat{\theta}_{2}$ | 0.0027 | 0.0595 | 0.0595 | 0.9480 |
|  | $\hat{\sigma}^{2}$ | -0.0291 | 0.0834 | 0.0785 | 0.8970 |
|  | RASE $_{1}$ | 0.7932 | 0.3728 |  |  |
|  | RASE $_{2}$ | 0.1108 | 0.0392 |  |  |
| 500 | $\hat{\rho}$ | -0.0019 | 0.0339 | 0.0332 | 0.9600 |
|  | $\hat{\theta}_{1}$ | -0.0039 | 0.0456 | 0.0442 | 0.9610 |
|  | $\hat{\theta}_{2}$ | -0.0004 | 0.0455 | 0.0461 | 0.9410 |
|  | $\hat{\sigma}^{2}$ | -0.0212 | 0.0653 | 0.0616 | 0.9100 |
|  | RASE $_{1}$ | 0.6253 | 0.2847 |  |  |
|  | RASE $_{2}$ | 0.0838 | 0.0303 |  |  |

Note: It shows that the bias (Bias), standard deviation (SD), standard error (SE) and coverage probability (CP) with nominal level of $95 \%$ for estimator and the mean and standard deviation (SD) of RASE for $\hat{\beta}(\cdot)$ and $\hat{g}(\cdot)$ from 1000 repetitions.

Table 2. Simulation results for $\rho=0.5$ with error term $0.75 t(3)$.

| $n$ | Est | Bias | SD | SE | CP |
| :--- | :--- | :--- | :--- | :--- | :--- |
| 00 | $\hat{\rho}$ | -0.0152 | 0.1010 | 0.1092 | 0.9620 |
|  | $\hat{\theta}_{1}$ | -0.0059 | 0.1332 | 0.1297 | 0.9540 |
|  | $\hat{\theta}_{2}$ | 0.0028 | 0.1332 | 0.1376 | 0.9350 |
|  | RASE $_{1}$ | 1.8018 | 1.0485 |  |  |
|  | RASE $_{2}$ | 0.2440 | 0.1053 |  |  |
| 300 | $\hat{\rho}$ | -0.0093 | 0.0567 | 0.0600 | 0.9530 |
|  | $\hat{\theta}_{1}$ | 0.0062 | 0.0761 | 0.0805 | 0.9460 |
|  | $\hat{\theta}_{2}$ | -0.0002 | 0.0759 | 0.0793 | 0.9400 |
|  | RASE $_{1}$ | 0.9992 | 0.5281 |  |  |
|  | RASE $_{2}$ | 0.1393 | 0.0591 |  |  |
| 500 | $\hat{\rho}$ | -0.0017 | 0.0429 | 0.0431 | 0.9500 |
|  | $\hat{\theta}_{1}$ | -0.0008 | 0.0577 | 0.0571 | 0.9540 |
|  | $\hat{\theta}_{2}$ | -0.0008 | 0.0578 | 0.0596 | 0.9410 |
|  | RASE $_{1}$ | 0.8034 | 0.4053 |  |  |
|  | RASE $_{2}$ | 0.1073 | 0.0427 |  |  |

Note: It shows that the bias (Bias), standard deviation (SD), standard error (SE) and coverage probability (CP) with nominal level of $95 \%$ for estimator and the mean and standard deviation (SD) of RASE for $\hat{\beta}(\cdot)$ and $\hat{g}(\cdot)$ from 1000 repetitions.

Table 3. Simulation results for $\rho=0.5$ with error term $\left(1+0.5 U_{i}\right) N(0,1)$.

| $n$ | Est | Bias | SD | SE | CP |
| :--- | :--- | :--- | :--- | :--- | :--- |
| 100 | $\hat{\rho}$ | -0.0143 | 0.0979 | 0.0984 | 0.9490 |
|  | $\hat{\theta}_{1}$ | -0.0065 | 0.1331 | 0.1465 | 0.9500 |
|  | $\hat{\theta}_{2}$ | 0.0048 | 0.1328 | 0.1339 | 0.9540 |
|  | RASE $_{1}$ | 1.8170 | 0.8823 |  |  |
|  | RASE $_{2}$ | 0.2431 | 0.0922 |  |  |
| 300 | $\hat{\rho}$ | -0.0037 | 0.0556 | 0.0540 | 0.9530 |
|  | $\hat{\theta}_{1}$ | 0.0012 | 0.0750 | 0.0735 | 0.9500 |
|  | $\hat{\theta}_{2}$ | -0.0032 | 0.0752 | 0.0755 | 0.9470 |
|  | RASE $_{1}$ | 1.0316 | 0.4720 |  |  |
|  | RASE $_{2}$ | 0.1411 | 0.0536 |  |  |
| 500 | $\hat{\rho}$ | -0.0017 | 0.0431 | 0.0430 | 0.9470 |
|  | $\hat{\theta}_{1}$ | 0.0026 | 0.0577 | 0.0555 | 0.9570 |
|  | $\hat{\theta}_{2}$ | -0.0026 | 0.0577 | 0.0574 | 0.9440 |
|  | RASE $_{1}$ | 0.8035 | 0.3615 |  |  |
|  | RASE $_{2}$ | 0.1058 | 0.0385 |  |  |

Note: It shows that the bias (Bias), standard deviation (SD), standard error (SE) and coverage probability (CP) with nominal level of $95 \%$ for estimator and the mean and standard deviation (SD) of RASE for $\hat{\beta}(\cdot)$ and $\hat{g}(\cdot)$ from 1000 repetitions.

Table 4. Simulation results for different magnitudes of $\rho$ with error term $N(0,1)$.

| $\rho$ | $n$ | Est | Bias | SE | SD | CP |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| 0.2 | 100 | $\hat{\rho}$ | -0.0114 | 0.1036 | 0.1046 | 0.9410 |
|  |  | $\hat{\theta}_{1}$ | -0.0036 | 0.1046 | 0.1064 | 0.9390 |
|  |  | $\hat{\theta}_{2}$ | -0.0019 | 0.1052 | 0.1009 | 0.9550 |
|  |  | $\hat{\sigma}^{2}$ | -0.1141 | 0.1334 | 0.1228 | 0.7500 |
|  |  | $\mathrm{RASE}_{1}$ | 1.4329 | 0.6579 |  |  |
|  |  | $\mathrm{RASE}_{2}$ | 0.1922 | 0.0705 |  |  |
|  | 300 | $\hat{\rho}$ | 0.0013 | 0.0565 | 0.0532 | 0.9620 |
|  |  | $\hat{\theta}_{1}$ | -0.0029 | 0.0585 | 0.0594 | 0.9500 |
|  |  | $\hat{\theta}_{2}$ | -0.0011 | 0.0586 | 0.0586 | 0.9480 |
|  |  | $\hat{\sigma}^{2}$ | -0.0389 | 0.0817 | 0.0779 | 0.8820 |
|  |  | $\mathrm{RASE}_{1}$ | 0.8313 | 0.3826 |  |  |
|  |  | $\mathrm{RASE}_{2}$ | 0.1115 | 0.0397 |  |  |
|  | 500 | $\hat{\rho}$ | -0.0012 | 0.0434 | 0.0435 | 0.9540 |
|  |  | $\hat{\theta}_{1}$ | -0.0006 | 0.0452 | 0.0457 | 0.9460 |
|  |  | $\hat{\theta}_{2}$ | -0.0021 | 0.0452 | 0.0439 | 0.9520 |
|  |  | $\hat{\sigma}^{2}$ | -0.0210 | 0.0658 | 0.0617 | 0.9090 |
|  |  | $\mathrm{RASE}_{1}$ | 0.6203 | 0.2789 |  |  |
|  |  | $\mathrm{RASE}_{2}$ | 0.0863 | 0.0307 |  |  |
| 0.7 | 100 | $\hat{\rho}$ | -0.0059 | 0.0569 | 0.0553 | 0.9600 |
|  |  | $\hat{\theta}_{1}$ | 0.0028 | 0.1075 | 0.1101 | 0.9400 |
|  |  | $\hat{\theta}_{2}$ | -0.0011 | 0.1068 | 0.1125 | 0.9340 |
|  |  | $\hat{\sigma}^{2}$ | -0.0990 | 0.1396 | 0.1239 | 0.7740 |
|  |  | $\mathrm{RASE}_{1}$ | 1.4460 | 0.6813 |  |  |
|  |  | $\mathrm{RASE}_{2}$ | 0.1935 | 0.0679 |  |  |
|  | 300 | $\hat{\rho}$ | -0.0012 | 0.0317 | 0.0319 | 0.9500 |
|  |  | $\hat{\theta}_{1}$ | 0.0008 | 0.0597 | 0.0606 | 0.9480 |
|  |  | $\hat{\theta}_{2}$ | 0.0010 | 0.0597 | 0.0578 | 0.9520 |
|  |  | $\hat{\sigma}^{2}$ | -0.0324 | 0.0840 | 0.0782 | 0.8860 |
|  |  | $\mathrm{RASE}_{1}$ | 0.7987 | 0.3837 |  |  |
|  |  | $\mathrm{RASE}_{2}$ | 0.1109 | 0.0415 |  |  |
|  | 500 | $\hat{\rho}$ | -0.0017 | 0.0242 | 0.2382 | 0.9560 |
|  |  | $\hat{\theta}_{1}$ | -0.0035 | 0.0459 | 0.0443 | 0.9580 |
|  |  | $\hat{\theta}_{2}$ | 0.0006 | 0.0459 | 0.0488 | 0.9410 |
|  |  | $\hat{\sigma}^{2}$ | -0.0199 | 0.0651 | 0.0617 | 0.9030 |
|  |  | $\mathrm{RASE}_{1}$ | 0.6124 | 0.2717 |  |  |
|  |  | $\mathrm{RASE}_{2}$ | 0.0853 | 0.0311 |  |  |

[^0] $95 \%$ for estimator and the mean and standard deviation (SD) of RASE for $\hat{\beta}(\cdot)$ from 1000 repetitions.


Figure 1. It displays the true curve $\beta(t)$ and $g(u)$ (red solid line), the estimated curve $\hat{\beta}(t)$ and $\hat{g}(u)$ (green dotted line) and ponitwise 2.5 and 97.5 percentile of the estimated function $\hat{\beta}(\cdot)$ and $\hat{g}(u)$ (light green line) in 500 replications with sample size $n=300$ respectively. In the firgure, the left one shows estimator $\hat{\beta}(t)$ and the right one shows estimator $\hat{g}(u)$ with error term $\mathrm{N}(0,1)$.

## 5. Discussion

In this paper, an interesting and novel functional semiparametric spatial autoregressive model is proposed. The model considers functional covariates based on semiparametric spatial autoregressive model. The slope function and nonparametric function are approximated by B-spline basis function.

Then generalized method of moments is proposed to estimate parameters. Under mild conditions, we establish the asymptotic properties for proposed estimators.

In order to use our model in practical applications, firstly, response variable needs spatial dependence. Secondly, there are covariates with nonlinear effect and functional variables. A problem of practical interest is to extend our model to take into account functional covariates and single index function simultaneously. What's more, making a test about spatial dependence and nonlinear effect of covariates is an important issue. Those topics are left for future work.

## Acknowledgments

We would like to thank the referees for their helpfull suggestions and comments which lead to the improvement of this article. Bai's work was supported by the National Natural Science Foundation of China (No.11771268).

## Conflict of interest

No potential conflict of interest was reported by the authors.

## References

1. L. Anselin, Spatial econometrics: methods and models, The Netherlands: Kluwer Academic Publishers, 1988.
2. L. Anselin, A. K. Bera, Spatial dependence in linear regression models with an introduction to spatial econometrics, In: Handbook of Applied Economic Statistics, New York: Marcel Dekker, 1998.
3. G. Aneiros-Pérez, P. Vieu, Semi-functional partial linear regression, Stat. Probabil. Lett., 76 (2006), 1102-1110.
4. C. de Boor, A practical guide to splines, New York: Springer-Verlag, 2001.
5. H. Cardot, F. Ferraty, P. Sarda, Spline estimators for the functional linear model, Stat. Sinica, 13 (2003), 571-592.
6. T. Cai, P. Hall, Prediction in functional linear regression, Ann. Statist., 34 (2006), 2159-2179.
7. T. Cai, M. Yuan, Minimax and adaptive prediction for functional linear regression, J. Am. Stat. Assoc., 107 (2012), 1201-1216.
8. C. Crambes, A. Kneip, P. Sarda, Smoothing splines estimators for functional linear regression, Ann. Statist., 37 (2009), 35-72.
9. X. Dai, S. Li, M. Tian, Quantile regression for partially linear varying coefficient spatial autoregressive models, 2016, arXiv:1608.01739.
10. J. Du, X. Sun, R. Cao, Z. Zhang, Statistical inference for partially linear additive spatial autoregressive models, Spat. Stat., 25 (2018), 52-67.
11. A. Delaigle, P. Hall, Methodology and theory for patial least squares applied to functional data, Ann. Statist., 40 (2012), 322-352.
12. T. Huang, S. Gilbert, H. Wang, S. Wang, Spatial functional linear model and its estimation method, 2018, arXiv:1811.00314.
13. Y. Hu, S. Wu, S. Feng, J. Jin, Estimation in partial functional linear spatial autoregressive model, Mathematics, 8 (2020), 1-12.
14. P. Hall, J. L. Horowitz, Methodology and convergence rates for functional linear regression, Ann. Statist., 35 (2007), 70-91.
15. J. Huang, Efficient estimation of the partly linear additive Cox model, Ann. Statist., 27 (1999), 1536-1563.
16. J. Z. Huang, Local asymptotics for polynomial spline regression, Ann. Statist., 31 (2003), 16001635.
17. 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 Finan. Econ., 17 (1998), 99-121.
18. H. H. Kelejian, I. R. Prucha, A generalized moments estimator for the autoregressive parameter in a spatial model, International Economic Review, 40 (1999), 509-533.
19. J. P. LeSage, The theory and practice of spatial econometrics, Ohio: University of Toledo, 1999.
20. L. F. Lee, J. H. Yu, Estimation of spatial autoregressive panel data models with fixed effects, J. Econometrics, 154 (2010), 165-185.
21. J. LeSage, R. K. Pace, Introduction to spatial econometrics, Boca Raton: Chapman and Hall/CRC, 2009.
22. L. F. Lee, GMM and 2SLS estimation of mixed regressive, spatial autoregressive models, $J$. Econometrics, 137 (2007), 489-514.
23. L. F. Lee, Best spatial two-stage least squares estimators for a spatial autoregressive model with autoregressive disturbances, Econometric Rev., 22 (2003), 307-335.
24. W. Pineda-Rios, R. Giraldo, E. Porcu, Functional SAR models: With application to spatial econometrics, Spat. Stat., 29 (2019), 145-159.
25. X. Qu, L. F. Lee, A spatial autoregressive model with a nonlinear transformation of the dependent variable, J. Econometrics, 184 (2015), 209-232.
26. P. T. Reiss, J. Goldsmith , H. L. Shang, R. T. Ogden, Methods for scalar-on-function regression, Int. Stat. Rev., 85 (2017), 228-249.
27. J. O. Ramsay, C. J. Dalzell, Some tools for functinal data analysis (with discussion), J. R. Stat. Soc. B, 53 (1991), 539-572.
28. H. Shin, Partial functional linear regression, J. Stat. Plan. Infer., 139 (2009), 3405-3418.
29. C. J. Stone, Optimal rates of convergence for nonparametric estimators, Ann. Statist., 8 (1980), 1348-1360.
30. C. J. Stone, Additive regression and other nonparametric models, Ann. Statist., 13 (1985), 689-705.
31. L. J. Su, Semiparametric GMM estimation of spatial autoregressive models, J. Econometrics, 167 (2012), 543-560.
32. L. J. Su, S. N. Jin, Profile quasi-maximum likelihood estimation of partially linear spatial autoregressive models, J. Econometrics, 157 (2010), 18-33.
33. Y. Sun, H. Yan, W. Zhang, Z. Lu, A semiparametric spatial dynamic model, Ann. Statist., 42 (2014), 700-727.
34. L. Schumaker, Spline functions: basic theory, Cambridge University Press, 2007.
35. H. Tadao, Semiparametric spatial autoregressive models with endogenous regressors: with an application to crime data, J. Bus. Econ. Stat., 36 (2018), 160-172.
36. H. Wei, Y. Sun, Heteroskedasticity-robust semi-parametric GMM estimation of a spatial model with space-varying coefficients, Spatial Economic Analysis, 12 (2017), 113-128.
37. L. Wang, X. Liu, H. Liang, R. Carroll, Estimation and variable selection for generalized additive partial linear models, Ann. Statist., 39 (2011), 1827-1851.
38. P. Yu, J. Du, Z. Zhang, Single-index partial functional linear regression model, Stat. Papers, 11 (2018), 1-17.
39. Y. Q. Zhang, D. M. Shen, Eseimation of semi-parametric varying-coefficient spatial panel data models with random effects, J. Statist. Plann. Infer., 159 (2015), 64-80.

## A. Appendix

Lemma 1. Assume condition C 8 holds for $g_{0}(u)$ and $\beta_{0}(t)$, there exits $\boldsymbol{\gamma}_{0}$ and $\zeta_{0}$ such that

$$
\sup _{t \in(0,1)}\left\|\beta_{0}(t)-\boldsymbol{B}_{1}^{\prime}(t) \boldsymbol{\gamma}_{0}\right\| \leq c_{1} k_{1}^{-r}, \quad \sup _{u \in(0,1)}\left\|g_{0}(u)-\boldsymbol{B}_{2}^{\prime}(u) \zeta_{0}\right\| \leq c_{2} k_{2}^{-r},
$$

where $\boldsymbol{\gamma}=\left(\gamma_{01}, \ldots, \gamma_{0 N_{1}}\right)^{\prime}, \zeta=\left(\zeta_{01}, \ldots, \zeta_{0 N_{2}}\right)^{\prime}$ and $c_{1}>0, c_{2}>0$ depend only on $l_{1}$ and $l_{2}$, respectively.
Proof of Lemma 1. It can be followed by spline's approximation property ( $[4,16,34]$ ).
Proof of Theorem 1. The proof is similar to Theorem 1 in [10] and we omit here.
Proof of Theorem 2. Let $\delta=n^{-\frac{r}{2 r+1}}, \boldsymbol{T}_{1}=\delta^{-1}\left(\boldsymbol{\gamma}-\boldsymbol{\gamma}_{0}\right), \boldsymbol{T}_{2}=\delta^{-1}\left(\boldsymbol{\zeta}-\zeta_{0}\right)$ and $\boldsymbol{T}=\left(\boldsymbol{T}_{1}^{\prime}, \boldsymbol{T}_{2}^{\prime}\right)^{\prime}$. We then prove that for any given $\epsilon>0$, there exits a sufficient large constant $L=L_{\epsilon}$ such that

$$
p\left\{\inf _{\|T\|=L} l\left(\boldsymbol{\phi}_{0}+\delta \boldsymbol{T}\right)>l\left(\boldsymbol{\phi}_{0}\right)\right\} \geq 1-\epsilon,
$$

where $\boldsymbol{\phi}_{0}=\left(\gamma_{0}^{\prime}, \boldsymbol{\zeta}_{0}^{\prime}\right)^{\prime}, l(\gamma, \zeta)=\sum_{i=1}^{n}\left(Y_{i}-\mathbf{Q}_{i}^{\prime} \hat{\boldsymbol{\eta}}-\mathbf{D}_{i}^{\prime} \boldsymbol{\gamma}-\boldsymbol{B}_{2}^{\prime}\left(u_{i}\right) \zeta\right)^{2}$. This implies that with the probability at least $1-\epsilon$ that there exits a local minimizer in the ball $\left\{\boldsymbol{\phi}_{0}+\delta \boldsymbol{T}:\|\boldsymbol{T}\| \leq L\right\}$. By Taylor expansion and simple calculation, it holds that

$$
\begin{aligned}
\left\{l\left(\boldsymbol{\phi}_{0}+\delta \boldsymbol{T}\right)-l\left(\boldsymbol{\phi}_{0}\right)\right\} \geq & -2 \delta \sum_{i=1}^{n}\left(\varepsilon_{i}+R_{1 i}+R_{2 i}+\mathbf{Q}_{i}^{\prime}(\boldsymbol{\eta}-\hat{\boldsymbol{\eta}})\right) V_{i} \\
& +\delta^{2} \sum_{i=1}^{n} V_{i}^{2}+o_{p}(1) \\
= & A_{1}+A_{2}+o_{p}(1),
\end{aligned}
$$

where $R_{1 i}=\left\langle X_{i}(t), \beta_{0}(t)-\boldsymbol{B}_{1}^{\prime}(t) \boldsymbol{\gamma}_{0}\right\rangle, R_{2 i}=g_{0}\left(u_{i}\right)-\boldsymbol{B}_{2}^{\prime}\left(u_{i}\right) \zeta_{0}, V_{i}=\mathbf{D}_{i}^{\prime} \boldsymbol{T}_{1}+\boldsymbol{B}_{2}^{\prime}\left(u_{i}\right) \boldsymbol{T}_{2}$. By assumption C6, Lemmas 1 and 8 of Stone [30], we derive that $\left\|R_{1 i}\right\| \leq c k_{1}^{-r},\left\|R_{2 i}\right\|=O_{p}\left(k_{2}^{-r}\right)$. Then by simple calculation, we obtain that

$$
\sum_{i=1}^{n} R_{1 i} V_{i}=\sum_{i=1}^{n} R_{1 i}\left(\mathbf{D}_{i}^{\prime} \boldsymbol{T}_{1}+\boldsymbol{B}_{2}^{\prime}\left(u_{i}\right) \boldsymbol{T}_{2}\right)=O_{p}\left(n k^{-r}\right)\|\boldsymbol{T}\|
$$

Similarly, it hold that $\sum_{i=1}^{n} \varepsilon_{i} V_{i}=O_{p}(\sqrt{n})\|\boldsymbol{T}\|, \sum_{i=1}^{n} R_{2 i} V_{i}=O_{p}\left(n k^{-r}\right)\|\boldsymbol{T}\|, \sum_{i=1}^{n} V_{i}^{2}=O_{p}(n)\|\boldsymbol{T}\|^{2}$. Similar to the proof of Theorem 2 in Du et al. [10], we get $(\boldsymbol{\eta}-\hat{\boldsymbol{\eta}})^{\prime} \mathbf{Q}^{\prime} \mathbf{Q}(\boldsymbol{\eta}-\hat{\boldsymbol{\eta}})=O_{p}(1)$. Then it holds that $\sum_{i=1}^{n} \mathbf{Q}_{i}(\boldsymbol{\eta}-\hat{\boldsymbol{\eta}}) V_{i}=O_{p}(\sqrt{n})\|\boldsymbol{T}\|$. Consequently, we show that $A_{1}=O_{p}\left(n \delta^{2}\right)\|\boldsymbol{T}\|, A_{2}=O_{p}\left(n \delta^{2}\right)\|\boldsymbol{T}\|^{2}$. Then through choosing a sufficiently large $L, A_{2}$ dominates $A_{1}$ uniformly in $\|\boldsymbol{T}\|=L$. Thus, there exits local minimizers $\hat{\gamma}, \hat{\zeta}$ such that $\|\gamma-\hat{\gamma}\|=O_{p}(\delta),\|\zeta-\hat{\zeta}\|=O_{p}(\delta)$.

Let $R_{1 k_{1}}(t)=\beta_{0}(t)-\boldsymbol{B}_{1}^{\prime}(t) \boldsymbol{\gamma}_{0}$. Then we get

$$
\begin{aligned}
\left\|\hat{\boldsymbol{\beta}}(t)-\beta_{0}(t)\right\|^{2} & =\int_{0}^{1}\left(\boldsymbol{B}_{1}^{\prime}(t) \hat{\gamma}-\beta_{0}(t)\right)^{2} d t \\
& =\int_{0}^{1}\left(\boldsymbol{B}_{1}^{\prime}(t) \hat{\gamma}-\boldsymbol{B}_{1}^{\prime}(t) \boldsymbol{\gamma}_{0}+R_{1 k_{1}}(t)\right)^{2} d t \\
& \leq 2 \int_{0}^{1}\left\{\boldsymbol{B}_{1}^{\prime}(t)\left(\hat{\gamma}-\boldsymbol{\gamma}_{0}\right)\right\}^{2} d t+2 \int_{0}^{1} R_{1 k_{1}}^{2}(t) d t \\
& =2\left(\hat{\gamma}-\boldsymbol{\gamma}_{0}\right)^{\prime} \int_{0}^{1} \boldsymbol{B}_{1}(t) \boldsymbol{B}_{1}^{\prime}(t) d t\left(\hat{\gamma}-\boldsymbol{\gamma}_{0}\right)+2 \int_{0}^{1} R_{1 k_{1}}^{2}(t) d t .
\end{aligned}
$$

Since $\|\boldsymbol{\gamma}-\hat{\boldsymbol{\gamma}}\|=O_{p}(\delta)$ and $\left\|\int_{0}^{1} \boldsymbol{B}_{1}(t) \boldsymbol{B}_{1}^{\prime}(t) d t\right\|=O(1)$, then we have

$$
\left(\hat{\gamma}-\boldsymbol{\gamma}_{0}\right)^{\prime} \int_{0}^{1} \boldsymbol{B}_{1}(t) \boldsymbol{B}_{1}^{\prime}(t) d t\left(\hat{\gamma}-\boldsymbol{\gamma}_{0}\right)=O_{p}\left(\delta^{2}\right)
$$

In addition, by Lemma1, it holds that $\int_{0}^{1} R_{1 k_{1}}^{2}(t) d t=O_{p}\left(\delta^{2}\right)$. Thus, we obtain $\left\|\hat{\beta}(t)-\beta_{0}(t)\right\|^{2}=O_{p}\left(\delta^{2}\right)$. Similarly, we get $\left\|\hat{g}(u)-g_{0}(u)\right\|^{2}=O_{p}\left(\delta^{2}\right)$.

Proof of Theorem 3. The proof is similar to Theorem 3 in [10] and we omit here.
Proof of Theorem 4. By the definition of $l(\gamma, \zeta)$ in the proof of Theorem 2, we have

$$
\begin{align*}
-\frac{1}{2 n} \frac{\partial l(\hat{\gamma} \hat{\boldsymbol{\zeta}})}{\partial \gamma} & =\frac{1}{n} \sum_{i=1}^{n}\left[Y_{i}-\mathbf{Q}_{i}^{\prime} \hat{\boldsymbol{\eta}}-\mathbf{D}_{i}^{\prime} \boldsymbol{\gamma}-\boldsymbol{B}_{2}^{\prime}\left(u_{i}\right) \zeta\right] \mathbf{D}_{i}^{\prime}  \tag{A.1}\\
& =\frac{1}{n} \sum_{i=1}^{n}\left[\tilde{e}_{i}-\mathbf{D}_{i}^{\prime}\left(\hat{\gamma}-\boldsymbol{\gamma}_{0}\right)-\boldsymbol{B}_{2}^{\prime}\left(u_{i}\right)\left(\hat{\boldsymbol{\zeta}}-\zeta_{0}\right)\right] \mathbf{D}_{i}^{\prime}+o_{p}(1)=0,
\end{align*}
$$

where $\tilde{e}_{i}=\varepsilon_{i}+R_{1 i}+R_{2 i}, R_{1 i}=\left\langle X_{i}(t), \beta_{0}(t)-\boldsymbol{B}_{1}^{\prime}(t) \gamma_{0}\right\rangle, R_{2 i}=g_{0}\left(u_{i}\right)-\boldsymbol{B}_{2}^{\prime}\left(u_{i}\right) \zeta_{0}$. The remainder is $o_{p}(1)$ because $\frac{1}{n} \sum_{i=1}^{n} \mathbf{Q}_{i}(\hat{\boldsymbol{\eta}}-\boldsymbol{\eta})=o_{p}(1)$ by Theorem 1. In addition, we have

$$
\begin{equation*}
-\frac{1}{2 n} \frac{\partial l(\hat{\gamma}, \hat{\zeta})}{\partial \zeta}=\frac{1}{n} \sum_{i=1}^{n}\left[\tilde{e}_{i}-\mathbf{D}_{i}^{\prime}\left(\hat{\gamma}-\boldsymbol{\gamma}_{0}\right)-\boldsymbol{B}_{2}^{\prime}\left(u_{i}\right)\left(\hat{\zeta}-\zeta_{0}\right)\right] \boldsymbol{B}_{2}^{\prime}\left(u_{i}\right)+o_{p}(1)=0 . \tag{A.2}
\end{equation*}
$$

It follows from (A.2) that

$$
\begin{equation*}
\hat{\zeta}-\zeta_{0}=\left[\frac{1}{n} \sum_{i=1}^{n} \boldsymbol{B}_{2}^{\prime}\left(u_{i}\right) \boldsymbol{B}_{2}\left(u_{i}\right)\right]^{-1}\left\{\frac{1}{n} \sum_{i=1}^{n} \tilde{e}_{i} \boldsymbol{B}_{2}\left(u_{i}\right)-\frac{1}{n} \sum_{i=1}^{n} \boldsymbol{B}_{2}\left(u_{i}\right) \mathbf{D}_{i}^{\prime}\left(\hat{\gamma}-\gamma_{0}\right)+o_{p}(1)\right\} . \tag{A.3}
\end{equation*}
$$

Let

$$
\bar{\Lambda}_{n}=\frac{1}{n} \sum_{i=1}^{n} \mathbf{D}_{i} \mathbf{D}_{i}^{\prime}-\frac{1}{n} \sum_{i=1}^{n} \mathbf{D}_{i} \boldsymbol{B}_{2}^{\prime}\left(u_{i}\right)\left[\frac{1}{n} \sum_{i=1}^{n} \boldsymbol{B}_{2}\left(u_{i}\right) \boldsymbol{B}_{2}^{\prime}\left(u_{i}\right)\right]^{-1} \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{B}_{2}\left(u_{i}\right) \mathbf{D}_{i}^{\prime} .
$$

By substituting (A.3) into (A.1), we obtain

$$
\hat{\gamma}-\boldsymbol{\gamma}_{0}=\left(\bar{\Lambda}_{n}\right)^{-1}\left\{\frac{1}{n} \sum_{i=1}^{n} \tilde{e}_{i}\left[\mathbf{D}_{i}-\left(\frac{1}{n} \mathbf{D}_{i} \boldsymbol{B}_{2}^{\prime}\left(u_{i}\right)\right) \times\left[\frac{1}{n} \sum_{i=1}^{n} \boldsymbol{B}_{2}\left(u_{i}\right) \boldsymbol{B}_{2}^{\prime}\left(u_{i}\right)\right]^{-1} \boldsymbol{B}_{2}\left(u_{i}\right)\right]+o_{p}(1)\right\} .
$$

Since $\hat{\beta}(t)-\beta^{*}(t)=\boldsymbol{B}_{1}^{\prime}(t)\left(\hat{\gamma}-\gamma_{0}\right)$ and for any $t \in(0,1)$, as $n \rightarrow \infty$, by the law of large numbers, the slutsky's theorem and the property of multivariate normal distribution, we obtain that

$$
\sqrt{\frac{n}{k_{1}}}\left(\hat{\beta}(t)-\beta^{*}(t)\right) \xrightarrow{d} N(0, \Xi(t)),
$$

where $\Xi(t)=\lim _{n \rightarrow \infty} \frac{\sigma^{2}}{k_{1}} \boldsymbol{B}_{1}^{\prime}(t) \Delta_{n} \boldsymbol{B}_{1}(t)$. Similar arguments hold for $\hat{g}(u)$.
(c) 2021 the Author(s), licensee AIMS Press. This is an open access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0)


[^0]:    Note: It shows that the bias (Bias), standard error (SE), standard deviation (SD) and coverage probability (CP) with nominal level of

