



Estimation for partially linear additive regression with spatial data

Tang Qingguo¹ · Chen Wenyu¹

Received: 3 September 2021 / Revised: 4 May 2022 / Accepted: 10 May 2022 /

Published online: 17 June 2022

© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2022

Abstract

This paper studies a partially linear additive regression with spatial data. A new estimation procedure is developed for estimating the unknown parameters and additive components in regression. The proposed method is suitable for high dimensional data, there is no need to solve the restricted minimization problem and no iterative algorithms are needed. Under mild regularity assumptions, the asymptotic distribution of the estimator of the unknown parameter vector is established, the asymptotic distributions of the estimators of the unknown functions are also derived. Finite sample properties of our procedures are studied through Monte Carlo simulations. A real data example about spatial soil data is used to illustrate our proposed methodology.

Keywords Spatial data · Partially linear additive regression · Asymptotic distribution

1 Introduction

Spatial data, which are collected at different sites on the surface of the earth, arise in various areas of research, including econometrics, epidemiology, environmental science, image analysis, oceanography and many others. Many authors (e.g. Ripley 1981; Cressie 1991) have studied the parametric methods for statistical inference in spatial models. However, one often encounters situations where a particular parametric model cannot be adopted with confidence and thus a nonparametric method is used as an alternative. In the last two decades, efforts have been made in the literature to explore nonlinear relationship in spatial data. Among them, Tran (1990), Hallin et al. (2001) and Lee et al. (2004) discussed kinds of asymptotic properties of density estimation for spatial processes. Hallin et al. (2004) established weak consistency and asymptotic normality for local linear regression estimation. Gao et al. (2006) developed the prop-

✉ Tang Qingguo
tangguo@163.com

¹ School of Economics and Management, Nanjing University of Science and Technology, Nanjing 210094, China

erties of the local linear kernel estimator for semiparametric spatial regression model. Al-Sulami et al. (2017) developed a two-step estimation procedure for semiparametric spatial time-series autoregressive model and established the asymptotic properties of the estimators. Nandy et al. (2017) developed a regularized variable selection technique for building a spatial additive model. The focus of our work is on the estimation and inference of partially linear additive regression with spatial data.

Let $\{(Y_{ij}, \mathbf{Z}_{ij}, \mathbf{X}_{ij})\}$ be a random field indexed by $(i, j) \in \mathbf{E}^2$, where $\mathbf{E} = \{0, \pm 1, \pm 2, \dots\}$ and Y_{ij} , with values in R , $\mathbf{Z}_{ij} = (Z_{ij1}, \dots, Z_{ijd_1})^T$, with values in R^{d_1} , $\mathbf{X}_{ij} = (X_{ij1}, \dots, U_{ijd_2})^T$, with values in R^{d_2} , are defined over some probability space (Ω, \mathcal{F}, P) . A point $(i, j) \in \mathbf{E}^2$ is referred to as a site. Let S and S' be two sets of sites. The Borel fields $\mathcal{B}(S) = \mathcal{B}((Y_{ij}, \mathbf{Z}_{ij}, \mathbf{X}_{ij}) : (i, j) \in S)$ and $\mathcal{B}(S') = \mathcal{B}((Y_{ij}, \mathbf{Z}_{ij}, \mathbf{X}_{ij}) : (i, j) \in S')$ are the σ -fields generated by the random vectors $(Y_{ij}, \mathbf{Z}_{ij}, \mathbf{X}_{ij})$ with (i, j) being the elements of S and S' respectively. Let $d(S, S')$ be the Euclidean distance between S and S' . We will assume that $\{(Y_{ij}, \mathbf{Z}_{ij}, \mathbf{X}_{ij})\}$ satisfies the following mixing condition as defined in the literature (cf., Tran 1990; Hallin et al. 2004): there exists a function $\varphi(t) \downarrow 0$ as $t \rightarrow \infty$, such that whenever $S, S' \subset \mathbf{E}^2$,

$$\alpha(\mathcal{B}(S), \mathcal{B}(S')) = \sup\{|P(AB) - P(A)P(B)|, A \in \mathcal{B}(S), B \in \mathcal{B}(S')\} \leq \psi(\text{Card}(S), \text{Card}(S'))\varphi(d(S, S')), \tag{1.1}$$

where $\text{Card}(S)$ denotes the cardinality of S , and ψ is a symmetric positive function nondecreasing in each variable. In particular, if $\psi(\infty, \infty) \leq C_0$ for some positive constant C_0 , $\{(Y_{ij}, \mathbf{Z}_{ij}, \mathbf{X}_{ij})\}$ is α -mixing (or strongly mixing). Introduced by Rosenblatt (1956), α -mixing dependence is a property shared by many time series models such as the autoregressive linear processes (see Athreya and Pantala 1986).

A crucial problem for a number of applications is the problem of spatial regression, where the influence of the vectors \mathbf{X}_{ij} and \mathbf{Z}_{ij} on some response variable Y_{ij} is to be studied in a context of complex spatial dependence. In many practical situations, it is desirable to model multiple covariates nonparametrically. However, it is well known that multivariate nonparametric estimation is subject to the curse of dimensionality. A widely used approach for dimensionality reduction is to consider an additive model for the nonparametric part of the regression function in the partly linear model, which in turn results in the partially linear additive model. In this paper, we try to approximate the conditional mean function $E(Y_{ij}|\mathbf{Z}_{ij}, \mathbf{X}_{ij})$ by a partially linear additive function of the form

$$\Psi(\mathbf{Z}_{ij}, \mathbf{X}_{ij}) = \mu + \mathbf{Z}_{ij}^T \boldsymbol{\beta}_0 + \sum_{r=1}^{d_2} f_r(X_{ijr}), \tag{1.2}$$

such that $E(Y_{ij} - \Psi(\mathbf{Z}_{ij}, \mathbf{X}_{ij}))^2$ or, equivalently, $E(E(Y_{ij}|\mathbf{Z}_{ij}, \mathbf{X}_{ij}) - \Psi(\mathbf{Z}_{ij}, \mathbf{X}_{ij}))^2$ is minimized over a class of partially linear additive functions of the form $\Psi(\mathbf{Z}_{ij}, \mathbf{X}_{ij})$, where μ is an unknown constant, $\boldsymbol{\beta}_0 = (\beta_{01}, \dots, \beta_{0d_1})^T$ is a d_1 -dimensional vector of unknown parameters, $f_r(x_r)$, $x_r \in [c_{r1}, c_{r2}]$, $r = 1, \dots, d_2$ are unknown functions. For identifiability, we assume that $E[f_r(X_r)] = 0$ for all $1 \leq r \leq d_2$.

Additive models are a special case of the more general projection pursuit regression model of Friedman and Stuetzle (1981). Additive models are useful for multivariate data analysis. They enable us to avoid the curse of dimensionality while retaining great flexibility in the regression function. However, in many empirical situations the mean of the outcome is assumed to depend on some covariates parametrically and some other covariates nonparametrically. Partially linear additive regression (PLAR), which contains both linear and nonlinear additive components, are more flexible than stringent linear models and are more parsimonious than general nonparametric regression models. PLAR have been studied by many authors, see, for example, Li (2000), Fan and Li (2003), Liang et al. (2008), You and Zhou (2013), Cheng et al. (2014), Lian et al. (2015), Sherwood and Wang (2016), Yang et al. (2019) and among others.

Gao et al. (2006) used marginal integration technique to estimate the additive components in PLAR with spatial data. However, this method is not suitable for high-dimensional vector \mathbf{X}_{ij} . In this paper, we develop a new estimation method to estimate the additive components $f_r(x_r)$ in (1.1). Our method has the following advantages comparing with the existing methods in the literature: (1) different from spline approach (Stone 1985; Huang et al. 2010) which needs to solve the constrained minimization problem and backfitting procedures (Brieman and Friedman 1985; Buja et al. 1989) which needs iterative algorithms, our estimators of additive functions have definite expressions, there is no need to solve the restricted minimization problem and no iterative algorithms are needed, resulting in fast and efficient estimation; (2) since there is only a smoothing parameter in B-spline procedure, when additive functions have different degrees of smoothness, the estimators of some functions become inefficient. In our proposed method, we first obtain the estimators of β_0 and the piecewise polynomial estimators of $f_r(x_r)$, $r = 1, \dots, d_2$, the local linear estimator of $f_r(u_r)$ is obtained by solving the univariate regression problem and the bandwidth h_r is easily chosen; (3) different from marginal integration approach of Gao et al. (2006), when \mathbf{X}_{ij} is high-dimensional vector and the sample size is not big, our estimators behavior well, this can be seen from the following simulation and application; (4) under mild regularity conditions, the asymptotic properties of the proposed estimators are established.

In our spatial PLAR, spatial information is added into a regression modeling by the spatial dependence of the data $\{(Y_{ij}, \mathbf{Z}_{ij}, \mathbf{X}_{ij}), (i, j) \in \mathbf{E}^2\}$, where \mathbf{Z}_{ij} and \mathbf{X}_{ij} may contain both exogenous and endogenous variables, that is, neighboring values of Y_{ij} . Moreover, a component Z_{ijr} of \mathbf{Z}_{ij} or a component X_{ijr} of \mathbf{X}_{ij} may itself be a linear combination of neighboring values of Y_{ij} . In the literature, there is another stream of spatial regression modeling by using a bivariate function to model the spatial effect into the model, for example, Ramsay (2002) use a deterministic smooth surface function to describe the variations and connections among values at different locations. Yu et al. (2020) introduce a class of generalized geoaddditive models, a synthesis of geostatistics and generalized additive models, where the effect of explanatory variables are modeled with additive univariate functions and the spatial effect is modeled via a bivariate function.

The paper is organized as follows. Section 2 describes the estimation method. Section 3 presents the asymptotic theory of our estimators. In Sect. 4, we conduct simulation studies to examine the finite-sample performance of the proposed proce-

dures. In Sect. 5, the proposed methods are illustrated by analyzing a real estate data set. Technical proofs are relegated to the Appendix.

2 Estimation method

As stated in Sect. 1, the conditional mean function $E(Y_{ij}|\mathbf{Z}_{ij}, \mathbf{X}_{ij})$ is approximated by a partially linear additive function of the form $\Psi(\mathbf{Z}_{ij}, \mathbf{X}_{ij}) = \mu + \mathbf{Z}_{ij}^T \boldsymbol{\beta}_0 + \sum_{r=1}^{d_2} f_r(X_{ijr})$ such that

$$E(Y_{ij} - \Psi(\mathbf{Z}_{ij}, \mathbf{X}_{ij}))^2 = E[Y_{ij} - (\mu + \mathbf{Z}_{ij}^T \boldsymbol{\beta}_0 + \sum_{r=1}^{d_2} f_r(X_{ijr}))]^2$$

is minimized over a class of partially linear additive functions of the form $\Psi(\mathbf{Z}_{ij}, \mathbf{X}_{ij})$ with $E(f_r(X_{ijr})) = 0$, which implies that $\mu = E(Y_{ij}) - E(\mathbf{Z}_{ij})^T \boldsymbol{\beta}_0$. Since

$$\begin{aligned} & E[Y_{ij} - (\mu + \mathbf{Z}_{ij}^T \boldsymbol{\beta}_0 + \sum_{r=1}^{d_2} f_r(X_{ijr}))]^2 \\ &= E(E([Y_{ij} - E(Y_{ij}|\mathbf{X}_{ij}) - (\mathbf{Z}_{ij} - E(\mathbf{Z}_{ij}|\mathbf{X}_{ij}))^T \boldsymbol{\beta}_0]^2|\mathbf{X}_{ij})) \\ &+ E[E(Y_{ij}|\mathbf{X}_{ij}) - \mu - E(\mathbf{Z}_{ij}|\mathbf{X}_{ij})^T \boldsymbol{\beta}_0 - \sum_{r=1}^{d_2} f_r(X_{ijr})]^2, \end{aligned}$$

we then have

$$\boldsymbol{\beta}_0 = (E[(\mathbf{Z}_{ij} - E(\mathbf{Z}_{ij}|\mathbf{X}_{ij}))(\mathbf{Z}_{ij} - E(\mathbf{Z}_{ij}|\mathbf{X}_{ij}))^T])^{-1} E[(\mathbf{Z}_{ij} - E(\mathbf{Z}_{ij}|\mathbf{X}_{ij}))(Y_{ij} - E(Y_{ij}|\mathbf{X}_{ij}))]$$

provided that the inverse exists, and $\sum_{r=1}^{d_2} f_r(X_{ijr})$ is a projection of $E(Y_{ij}|\mathbf{X}_{ij}) - \mu - E(\mathbf{Z}_{ij}|\mathbf{X}_{ij})^T \boldsymbol{\beta}_0$ on the space of additive functions of $X_{ijr}, r = 1, \dots, d_2$.

Assume that we have observations $(Y_{ij}, \mathbf{Z}_{ij}, \mathbf{X}_{ij})$ for $1 \leq i \leq m$ and $1 \leq j \leq n$. The total sample size is thus $N = m \times n$. We first estimate $\boldsymbol{\beta}_0$. In order to approximate $f_r(x_r)$ for $x_r \in [c_{r1}, c_{r2}]$, we construct piecewise polynomial estimators of $f_r(x_r)$ of degree p_r . We split equally $[c_{r1}, c_{r2}]$ into M_{rN} subintervals. Then the length of every subinterval is $2h_{r0} = (c_{r2} - c_{r1})/M_{rN}$. Let $I_{rv} = [c_{r1} + 2(v - 1)h_{r0}, c_{r1} + 2vh_{r0})$ for $1 \leq v \leq M_{rN} - 1$ and $I_{rM_{rN}} = [c_{r2} - 2h_{r0}, c_{r2}]$. Let x_{rv} denote the center of the interval I_{rv} and χ_{rv} denote the indicator function of I_{rv} , so that $\chi_{rv}(x_r) = 1$ or 0 according to $x_r \in I_{rv}$ or $x_r \notin I_{rv}$.

Denote $\check{A}_{rvk}(x_r) = \chi_{rv}(x_r)[(x_r - x_{rv})/h_{r0}]^k$ for $k = 0, 1, \dots, p_r; v = 1, \dots, M_{rN}; r = 1, \dots, d_2$. We use $\check{f}_r(x_r) = \sum_{v=1}^{M_{rN}} \sum_{k=0}^{p_r} a_{rvk} \check{A}_{rvk}(x_r)$ to approximate $f_r(x_r)$. Since $E[f_r(X_r)] = 0$, we then set $E[\check{f}_r(X_r)] = 0$, this can be done by using $\check{A}_{rvk}(x_r) - \chi_{rv}(x_r)E[\check{A}_{rvk}(X_r)]/E[\chi_{rv}(X_r)]$ to replace $\check{A}_{rvk}(x_r)$ for $k = 1, \dots, p_r$ and setting $\sum_{v=1}^{M_{rN}} a_{rv0} E[\chi_{rv}(X_r)] = 0$. Let $a_{rM_{rN}0} = \sum_{v=1}^{M_{rN}-1} a_{rv0} E[\chi_{rv}(X_r)]/E[\chi_{rM_{rN}}(X_r)]$. Since $E[\check{A}_{rvk}(X_r)]$ and $E[\chi_{rv}(X_r)]$ are unknown, we then use

$$\bar{A}_{rvk} = \sum_{i=1}^m \sum_{j=1}^n \chi_{rv}(X_{ijr}) [(X_{ijr} - x_{rv})/h_{r0}]^k / \left[\sum_{i=1}^m \sum_{j=1}^n \chi_{rv}(X_{ijr}) \right]$$

to replace $E[\check{A}_{rvk}(X_r)]/E[\chi_{rv}(X_r)]$. Denote

$$\begin{aligned} A_{rv0}(x_r) &= \chi_{rv}(x_r) - \frac{\sum_{i=1}^m \sum_{j=1}^n \chi_{rv}(X_{ijr})}{\sum_{i=1}^m \sum_{j=1}^n \chi_{rM_{rN}}(X_{ijr})} \chi_{rM_{rN}}(x_r), \quad v = 1, \dots, M_{rN} - 1, \\ A_{rvk}(x_r) &= \chi_{rv}(x_r) \{ [(x_r - x_{rv})/h_{r0}]^k - \bar{A}_{rvk} \}, \quad k = 1, \dots, p_r, \\ \mathbf{A}_{rv}(x_r) &= (A_{rv0}(x_r), A_{rv1}(x_r), \dots, A_{rvq_r}(x_r))^T, \quad v = 1, \dots, M_{rN} - 1, \\ \mathbf{A}_{rM_{rN}}(x_r) &= (A_{rM_{rN}1}(x_r), \dots, A_{rM_{rN}q_r}(x_r))^T, \\ \mathbf{A}_r(x_r) &= (\mathbf{A}_{r1}(x_r)^T, \dots, \mathbf{A}_{rM_{rN}}(x_r)^T)^T, \quad r = 1, \dots, d_2. \end{aligned}$$

Let $\mathbf{a}_{rv} = (a_{rv0}, a_{rv1}, \dots, a_{rvp_r})^T$ for $v = 1, \dots, M_{rN} - 1$, $\mathbf{a}_{rM_{rN}} = (a_{rM_{rN}1}, \dots, a_{rM_{rN}p_r})^T$ and $\mathbf{a}_r = (\mathbf{a}_{r1}^T, \dots, \mathbf{a}_{rM_{rN}}^T)^T$. We use $\tilde{f}_r(x_r) = \mathbf{A}_r^T(x_r)\mathbf{a}_r$ to approximate $f_r(x_r)$. Note that $\tilde{f}_r(x_r)$ is a piecewise polynomial of degree p_r and satisfies that $\frac{1}{N} \sum_{i=1}^m \sum_{j=1}^n \tilde{f}_r(X_{ijr}) = 0$. Let $\bar{Y} = \frac{1}{N} \sum_{i=1}^m \sum_{j=1}^n Y_{ij}$ and $\bar{\mathbf{Z}} = \frac{1}{N} \sum_{i=1}^m \sum_{j=1}^n \mathbf{Z}_{ij}$ be the estimator of $E(Y)$ and $E(\mathbf{Z})$ respectively. Based on spatial observations $\{(Y_{ij}, \mathbf{Z}_{ij}, \mathbf{X}_{ij}) : 1 \leq i \leq m; 1 \leq j \leq n\}$, we use $\bar{Y} - \bar{\mathbf{Z}}^T \boldsymbol{\beta}$ to approximate μ and solve the following minimization problem

$$\min \sum_{i=1}^m \sum_{j=1}^n \{ Y_{ij} - \bar{Y} - (\mathbf{Z}_{ij} - \bar{\mathbf{Z}})^T \boldsymbol{\beta} - \sum_{r=1}^{d_2} \mathbf{A}_r^T(X_{ijr})\mathbf{a}_r \}^2 \tag{2.1}$$

with respect to the $\boldsymbol{\beta}, \mathbf{a}_r$. Denote $\bar{Y}_{ij} = Y_{ij} - \bar{Y}$, $\bar{\mathbf{Z}}_{ij} = \mathbf{Z}_{ij} - \bar{\mathbf{Z}}$, $\mathbf{A}(\mathbf{X}_{ij}) = (\mathbf{A}_1^T(X_{ij1}), \dots, \mathbf{A}_{d_2}^T(X_{ijd_2}))^T$ and $\mathbf{a} = (\mathbf{a}_1^T, \dots, \mathbf{a}_{d_2}^T)^T$. Then the estimator $\hat{\boldsymbol{\beta}}$ of $\boldsymbol{\beta}_0$ is given by

$$\hat{\boldsymbol{\beta}} = \boldsymbol{\Gamma}_N^{-1} \left[\sum_{i=1}^m \sum_{j=1}^n \bar{\mathbf{Z}}_{ij} \bar{Y}_{ij} - \mathbf{W}_N \mathbf{A}_N^{-1} \left(\sum_{i=1}^m \sum_{j=1}^n \mathbf{A}(\mathbf{X}_{ij}) \bar{Y}_{ij} \right) \right], \tag{2.2}$$

where

$$\boldsymbol{\Gamma}_N = \sum_{i=1}^m \sum_{j=1}^n \bar{\mathbf{Z}}_{ij} \bar{\mathbf{Z}}_{ij}^T - \mathbf{W}_N \mathbf{A}_N^{-1} \mathbf{W}_N^T, \tag{2.3}$$

$$\mathbf{W}_N = \sum_{i=1}^m \sum_{j=1}^n \bar{\mathbf{Z}}_{ij} \mathbf{A}(\mathbf{X}_{ij})^T, \quad \mathbf{A}_N = \sum_{i=1}^m \sum_{j=1}^n \mathbf{A}(\mathbf{X}_{ij}) \mathbf{A}(\mathbf{X}_{ij})^T. \tag{2.4}$$

The estimator $\tilde{\mathbf{a}} = (\tilde{\mathbf{a}}_1^T, \dots, \tilde{\mathbf{a}}_{d_2}^T)^T$ of \mathbf{a} is given by

$$\tilde{\mathbf{a}} = \mathbf{A}_N^{-1} \left(\sum_{i=1}^m \sum_{j=1}^n \mathbf{A}(\mathbf{X}_{ij}) (\bar{Y}_{ij} - \bar{\mathbf{Z}}_{ij}^T \hat{\boldsymbol{\beta}}) \right) \tag{2.5}$$

and piecewise polynomial estimator $\tilde{f}_r(x_r)$ of $f_r(x_r)$ is given by

$$\tilde{f}_r(x_r) = \mathbf{A}_r^T(x_r)\tilde{\mathbf{a}}_r, \quad x_r \in [c_{r1}, c_{r2}].$$

The estimator of μ is given by

$$\hat{\mu} = \bar{Y} - \bar{\mathbf{Z}}^T \hat{\boldsymbol{\beta}}. \tag{2.6}$$

Since $\tilde{f}_r(x_r)$ is only piecewise smooth, it may not be appealing to use this piecewise polynomial directly as the estimate of the function $f_r(x_r)$. After we estimated $\boldsymbol{\beta}_0$, for a given $x_{0r} \in [c_{r1}, c_{r2}]$, for x_r in the neighborhood of x_{0r} , we use $b_{r0} + b_{r1}(x_r - x_{0r})$ to approximate the unknown coefficient function $f_r(x_r)$. We then solve the following minimization problem

$$\min \sum_{i=1}^m \sum_{j=1}^n \left\{ Y_{ij} - \bar{Y} - (\mathbf{Z}_{ij} - \bar{\mathbf{Z}})^T \hat{\boldsymbol{\beta}} - \sum_{k \neq r} \tilde{f}_k(X_{ijk}) - [b_{r0} + b_{r1}(X_{ijr} - x_{0r})] \right\}^2 K\left(\frac{X_{ijr} - x_{0r}}{h_r}\right) \tag{2.7}$$

with respect to b_{r0}, b_{r1} , where $K(\cdot)$ is a given kernel function and h_r is a chosen bandwidth. Let $\mathbf{b}_r = (b_{r0}, b_{r1})^T$ and $\mathbf{B}_{ijr} = (1, (X_{ijr} - x_{0r})/h_r)^T$. Then the estimator of \mathbf{b}_r is given by

$$\hat{\mathbf{b}}_r = \left(\sum_{i=1}^m \sum_{j=1}^n K\left(\frac{X_{ijr} - x_{0r}}{h_r}\right) \mathbf{B}_{ijr} \mathbf{B}_{ijr}^T \right)^{-1} \left(\sum_{i=1}^m \sum_{j=1}^n K\left(\frac{X_{ijr} - x_{0r}}{h_r}\right) \mathbf{B}_{ijr} \tilde{Y}_{ijr} \right) \tag{2.8}$$

where

$$\tilde{Y}_{ijr} = Y_{ij} - \bar{Y} - (\mathbf{Z}_{ij} - \bar{\mathbf{Z}})^T \hat{\boldsymbol{\beta}} - \sum_{r' \neq r} \tilde{f}_{r'}(X_{ijr'}). \tag{2.9}$$

The estimator of $f_r(x_{0r})$ is given by $\hat{f}_r(x_{0r}) = \hat{b}_{r0} = (1, 0)\hat{\mathbf{b}}_r$ for $r = 1, \dots, d_2$.

To implement our estimation method, appropriate values M_{rN} and h_r are necessary. Here, M_{rN} are mainly used to estimate the parameters $\boldsymbol{\beta}_0$, for simplicity, we set $M_{1N} = \dots = M_{dN} = M_N$. The value for M_N can be selected by the following BIC information criterion:

$$\begin{aligned} BIC(M_N) = \log\left\{ \frac{1}{N} \sum_{i=1}^m \sum_{j=1}^n [Y_{ij} - \bar{Y} - (\mathbf{Z}_{ij} - \bar{\mathbf{Z}})^T \hat{\boldsymbol{\beta}} \right. \\ \left. - \sum_{r=1}^{d_2} \mathbf{A}_r^T(X_{ijr})\hat{\mathbf{a}}_r]^2 \right\} + M_N \log N / (2N). \end{aligned} \tag{2.10}$$

Large values of BIC indicate poor fits.

The bandwidth h_r can be selected by the following cross-validation (CV) score:

$$CV(h_r) = \frac{1}{N} \sum_{i=1}^m \sum_{j=1}^n \left(Y_{ij} - \hat{\mu}^{-ij} - \mathbf{Z}_{ij}^T \hat{\boldsymbol{\beta}}^{-ij} - \sum_{r' \neq r} \tilde{f}_{r'}^{-ij}(X_{ijr'}) - \hat{f}_r^{-ij}(X_{ijr}) \right)^2, \tag{2.11}$$

where $\hat{\mu}^{-ij}$, $\hat{\boldsymbol{\beta}}^{-ij}$, $\tilde{f}_{r'}^{-ij}(X_{ijr'})$ and $\hat{f}_r^{-ij}(X_{ijr})$ are estimated by removing the (i,j)th observation.

3 Asymptotic properties of the estimators

In this section, we shall describe the asymptotic properties of the estimators $\hat{\boldsymbol{\beta}}$ and $\hat{f}_r(x_r)$, $r = 1, \dots, d_2$. We first list following assumptions.

1. The random field $\{(Y_{ij}, \mathbf{Z}_{ij}, \mathbf{X}_{ij}) : (i, j) \in \mathbf{E}^2\}$ is strictly stationary. The marginal density $g_r(x_r)$ of X_{ijr} is continuous and bounded away from 0 uniformly over $[c_{r1}, c_{r2}]$ for $r = 1, \dots, d_2$.
2. $f_r(x_r) \in C^{p_r+1}[c_{r1}, c_{r2}]$ for $r = 1, \dots, d_2$.
3. $E\|\mathbf{Z}_{ij}\|^{4+2\delta} < \infty$ and $E|\varepsilon_{ij}|^{4+2\delta} < \infty$ for some $\delta > 0$, where $\varepsilon_{ij} = Y_{ij} - (\mu + \mathbf{Z}_{ij}^T \boldsymbol{\beta}_0 + \sum_{r=1}^{d_2} f_r(X_{ijr}))$.
4. The function $\psi(\cdot, \cdot)$ and φ satisfy that $\psi(m, n) \leq \min(m, n)$ and

$$\lim_{k \rightarrow \infty} k^\gamma \sum_{t=k}^{\infty} t[\varphi(t)]^{\delta/(2+\delta)} = 0$$

for some constant $\gamma > (4 + \delta)/(2 + \delta)$.

5. $M_N^{4+2\delta/(4+\delta)}/N \rightarrow 0$ and $NM_N^{-4(p_r+1)} \rightarrow 0$, where $M_N = \max_{1 \leq r \leq d_2} M_{rN}$.
6. $\min\{m, n\} \rightarrow \infty$ and there exist two sequences of positive integer vectors, $\mathbf{p}_{m,n} = (p_1, p_2) \in \mathbf{E}^2$ and $\mathbf{q}_{m,n} = (q, q) \in \mathbf{E}^2$, with $q \rightarrow \infty$ such that $q/p_1 \rightarrow 0$, $q/p_2 \rightarrow 0$ and $m/p_1 \rightarrow \infty$, $n/p_2 \rightarrow \infty$, and $p_N = p_1 p_2 = o((Nh_r)^{1/2})$, $N\varphi(q) \rightarrow 0$.
7. The kernel function $K(\cdot) \geq 0$ is a bounded symmetric function with a compact support.

Assumption 1 is standard in this context; it has been used, for instance, by Gao et al. (2006) in the spatial context. The constants γ and δ in assumption 4 may be interpreted as two indices of the mixing dependence. Generally, larger γ and smaller δ indicate less dependence. If $\varphi(t) = O(t^{-\kappa})$ for some $\kappa > 4(3 + \delta)/\delta$ or $\varphi(t) = O(e^{-\iota t})$ for some $\iota > 0$, then assumption 4 holds. Assumptions 6 is used to derive the asymptotic normality of the estimators.

Let \mathcal{F} denote the class of additive functions such that $F(\mathbf{x}) \in \mathcal{F}$ if $F(\mathbf{x}) = \mu + \sum_{r=1}^{d_2} F_r(x_r)$ and $F_r(x_r)$ satisfies assumption 2 and $E(F_r(X_r)) = 0$ for $r = 1, \dots, d_2$. To obtain the asymptotic distribution of $\hat{\boldsymbol{\beta}}$, we first need to adjust for the dependence of \mathbf{Z}_{ij} and \mathbf{X}_{ij} , which is a common complication in semiparametric models. Let

$$f_k(\mathbf{X}_{ij}) = \mu_k + \sum_{r=1}^{d_2} f_{kr}(X_{ijr}) = \operatorname{arg\,inf}_{F(\mathbf{x}) \in \mathcal{F}} E[Z_{ijk} - F(\mathbf{X}_{ij})]^2$$

and $\theta_k(\mathbf{X}_{ij}) = E(Z_{ijk}|\mathbf{X}_{ij})$. Since

$$E[Z_{ijk} - f_k(\mathbf{X}_{ij})]^2 = E[Z_{ijk} - \theta_k(\mathbf{X}_{ij})]^2 + E[\theta_k(\mathbf{X}_{ij}) - f_k(\mathbf{X}_{ij})]^2,$$

therefore, $f_k(\mathbf{X}_{ij})$ are the projections of $\theta_k(\mathbf{X}_{ij})$ onto the additive functional space \mathcal{F} (under the L_2 -norm). In other words, $f_k(\mathbf{X}_{ij})$ is an element that belongs to \mathcal{F} and it is the closest function to $\theta_k(\mathbf{X}_{ij})$ among all the functions in \mathcal{F} , for any $k = 1, \dots, d_1$. Let $V_{ijk} = Z_{ijk} - f_k(\mathbf{X}_{ij})$ and $\mathbf{V}_{ij} = (V_{ij1}, \dots, V_{ijd_1})^T$. Set $\mathbf{\Gamma} = E(\mathbf{V}_{ij}\mathbf{V}_{ij}^T)$ and $\mathbf{\Omega} = \sum_{i=-\infty}^{+\infty} \sum_{j=-\infty}^{+\infty} E[\mathbf{V}_{00}\mathbf{V}_{ij}^T \varepsilon_{00}\varepsilon_{ij}]$. The following theorem gives the asymptotic distribution of the estimator $\hat{\boldsymbol{\beta}}$.

Theorem 3.1 *Suppose that assumptions 1–6 hold. Then*

$$\sqrt{N}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \xrightarrow{d} N(\mathbf{0}, \mathbf{\Gamma}^{-1}\mathbf{\Omega}\mathbf{\Gamma}^{-1}). \tag{3.1}$$

Let $\mu_k = \int x^k K(x)dx$, $v_k = \int x^k K^2(x)dx$ for $k = 0, 1, \dots$, and $\sigma_r^2(x_r) = E(\varepsilon_{ij}^2|X_{ijr} = x_r)$. The following Theorem 3.2 gives the asymptotic distribution of the estimators $\hat{f}_r(x_r)$ for $r = 1, \dots, d_2$.

Theorem 3.2 *Suppose that assumptions 1–7 hold and $\sigma_r^2(x_r)$ is continuous in some neighborhood of x_{0r} and $M_N^{1+\delta/(4+\delta)}h_r \leq C_2$ for some positive constant C_2 . If x_{0r} is an interior point of $[c_{r1}, c_{r2}]$, then,*

$$\sqrt{Nh_r}(\hat{f}_r(x_{0r}) - f_r(x_{0r}) - \frac{\mu_2 h_r^2}{2\mu_0} f_r''(x_{0r})) \xrightarrow{d} N(0, \frac{v_0 \sigma_r^2(x_{0r})}{\mu_0^2 g_r(x_{0r})}). \tag{3.2}$$

The result of Theorem 3.2 indicates that the estimator of $f_r(x_{0r})$ in partially linear additive model has the same asymptotic distribution as the estimator in spatial univariate nonparametric regression.

4 Simulations

We now study the finite sample performance of our proposed method in Sect. 2 using a simulation spatial dataset. We shall use the following spatial partially linear additive model. Let $\{\varepsilon_{ij}^{(k)} : (i, j) \in \mathbf{E}^2\}$, $k = 1, 2, 3$ be three mutually independent i.i.d. $N(0, 1)$ white-noise processes and $\{\varepsilon_{ij}^{(4)} = (\varepsilon_{ij1}^{(4)}, \varepsilon_{ij2}^{(4)}, \varepsilon_{ij3}^{(4)})^T : (i, j) \in \mathbf{E}^2\}$ be an i.i.d. $N(0, \Theta)$ process with $\Theta = (\theta_{kk'})_{3 \times 3}$ and $\theta_{kk'} = \exp(-|k - k'|/3)$. Let $\{\varepsilon_{ij}^{(5)} : (i, j) \in \mathbf{E}^2\}$ be an i.i.d. $N(0, 0.1)$ white-noise process. Let

$$Y_{ij} = \mu + Z_{ij1}\beta_1 + Z_{ij2}\beta_2 + Z_{ij3}\beta_3 + \sum_{r=1}^3 f_r(X_{ijr}) + \varepsilon_{ij} \tag{4.1}$$

with $\mu = 2.4, \beta_1 = 1.6, \beta_2 = -2.5, \beta_3 = 3.1$ and $f_r(X_r) = f_{0r}(X_r) - E(f_{0r}(X_r))$, where

$$\begin{aligned} f_{01}(X_1) &= 1.2(X_1 - 0.9)^2, f_{02}(X_2) = -2.1 \sin(X_2 + 1.5), f_{03}(X_3) \\ &= 1.8 \cos(3X_3 - 0.6). \end{aligned}$$

We set $Z_{ij1} = X_{ij1}^2 - 2X_{ij2} + \varepsilon_{ij}^{(1)}, Z_{ij2} = X_{ij1} + X_{ij2}^2 - 0.6X_{ij3}^2 + \varepsilon_{ij}^{(2)}, Z_{ij3} = 0.9 + \sin(2X_{ij2} - X_{ij3}) + \varepsilon_{ij}^{(3)}, X_{ij1} = (U_{ij1} + U_{ij2})/2, X_{ij2} = (U_{ij2} + U_{ij3})/2, X_{ij3} = (U_{ij1} + U_{ij3})/2. \{U_{ijr} : (i, j) \in \mathbf{E}^2\}, r = 1, 2, 3$ are generated by the spatial autoregression

$$U_{ijr} = \sin(U_{(i-1)jr} + U_{i(j-1)r} + U_{(i+1)jr} + U_{i(j+1)r}) + \varepsilon_{ijr}^{(4)},$$

and $\{\varepsilon_{ij} : (i, j) \in \mathbf{E}^2\}$ is generated by the spatial autoregression $\varepsilon_{ij} = (\varepsilon_{(i-1)j} + \varepsilon_{i(j-1)} + \varepsilon_{(i+1)j} + \varepsilon_{i(j+1)})/5 + \varepsilon_{ij}^{(5)}$.

Data were simulated from this model over a rectangular domain of $m \times n$ sites—more precisely, over a grid of the form $\{(i, j) : 76 \leq i \leq 75+m, 76 \leq j \leq 75+n\}$, for various values of m and n . Each replication was obtained iteratively by the steps given in Hallin et al. (2004). 500 simulated spatial data sets are independently generated. For each simulated data set, the estimators of $\beta_r, r = 1, 2, 3$ and μ were computed by (2.2) with $p_r = 1$ and (2.6). The number of subinterval $M_{1N} = M_{2N} = M_{3N} = M_N$ were determined by BIC criterion given by (2.10). The estimators of $f_r(x_r), r = 1, 2, 3$ were computed by (2.8) with Epanechnikov kernel. Bandwidth h_r was select by “leave-one-out” cross-validation procedure given by (2.10).

Table 1 summaries the mean square errors (MSE) of the estimators $\hat{\mu}$ and $\hat{\beta}_r$ for $r = 1, 2, 3$ and the mean integrated squared errors (MISE) of the estimators $\hat{f}_r(x_r)$ for $r = 1, 2, 3$ based on 500 simulations for different m and n . Figure 1 depicts the actual functions $f_r(x_r), r = 1, 2, 3$ and the mean estimated curves over 500 simulations with $m = 20, n = 20$ and their 95% pointwise confidence bands. We see from Table 1 and Fig. 1 that our estimation procedure gives satisfactory results ever if for small m and n .

We now compare the proposed method (PM) with the B-spline method (BM) given in Huang et al. (2010). In B-spline method, cubic splines with equally spaced knots are used to approximate the additive functions $f_r(u_r), r = 1, 2, 3$, the smoothing parameter K_n were determined by BIC criterion. For different m and n , Table 2 reports the MSEs of parametric estimators $\hat{\mu}$ and $\hat{\beta}_r, r = 1, 2, 3$ and the mean of the weighted average squared error $WASE_r$ of the estimators $\hat{f}_r(x_r), r = 1, 2, 3$ based on 500 simulations, where $WASE_r$ is defined as

$$WASE_r = \frac{1}{mn + 1 - 2[0.025mn]} \sum_{i=[0.025mn]}^{mn-[0.025mn]} \frac{[\hat{f}_r(x_{(ir)}) - f_r(x_{(ir)})]^2}{[range(f_r)]^2},$$

Table 1 MSEs and MISEs of the estimators based on regular grid data

| | MSE | | | | MISE | | |
|-----------|-------------|-----------------|-----------------|-----------------|------------------|------------------|------------------|
| | $\hat{\mu}$ | $\hat{\beta}_1$ | $\hat{\beta}_2$ | $\hat{\beta}_3$ | $\hat{f}_1(u_1)$ | $\hat{f}_2(u_2)$ | $\hat{f}_3(u_3)$ |
| m=10,n=10 | 0.0358 | 0.0099 | 0.0099 | 0.0038 | 0.7340 | 0.3659 | 0.4870 |
| m=20,n=10 | 0.0315 | 0.0077 | 0.0062 | 0.0022 | 0.4964 | 0.2803 | 0.4355 |
| m=15,n=20 | 0.0236 | 0.0069 | 0.0041 | 0.0013 | 0.3736 | 0.2101 | 0.4108 |
| m=20,n=20 | 0.0219 | 0.0062 | 0.0036 | 0.0011 | 0.3403 | 0.1833 | 0.3950 |
| m=30,n=20 | 0.0197 | 0.0058 | 0.0026 | 0.0009 | 0.2890 | 0.1561 | 0.3871 |

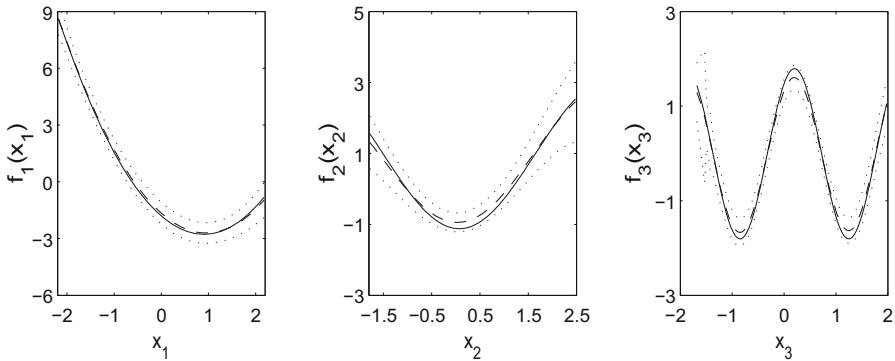


Fig. 1 The actual and the mean estimated curves for $f_r(x_r)$, $r = 1, 2, 3$ in model (4.1) with $m = 20$, $n = 20$ the 95% pointwise confidence bands. —, true curves; - - -, mean estimated curves;, 95% pointwise confidence bands

$x_{(1r)} \leq x_{(2r)} \leq \dots \leq x_{(mnr)}$ is a permutation of $\{x_{ijr} : 76 \leq i \leq 75 + m, 76 \leq j \leq 75 + n\}$ and $\text{range}(f_r)$ is the range of the function $f_r(x_r)$. Since the data near the boundary points are sparse, the estimators near the boundary points become poor. So we removed these points from $WASE_r$. Based on 500 simulations with $m = 20$, $n = 10$, Fig. 2 displays the boxplots for squared errors (SE) of parametric estimators $\hat{\mu}$ and $\hat{\beta}_r$, $r = 1, 2, 3$ and $WASE_r$ of the estimators $\hat{f}_r(x_r)$, $r = 1, 2, 3$. We see from Table 2 and Fig. 2 that there are no obvious differences between the parametric estimators of two methods, but the estimators of additive components $f_r(x_r)$, $r = 1, 2, 3$ obtained by the proposed method outperform that obtained by B-spline method. Because the number of knots of all the B-spline functions approximating the additive components $f_r(u_r)$, $r = 1, 2, 3$ is the same, when additive functions have different degrees of smoothness, some estimators will become inefficient. In the proposed method, we first obtain the estimators of β_0 and the piecewise polynomial estimators of $f_r(x_r)$, $r = 1, 2, 3$, the local linear estimator of $f_r(u_r)$ is obtained by solving the univariate regression problem and the bandwidth h_r is easily chosen.

To investigate the performances of the estimators under different M_n , Table 3 displays the simulation results based on 500 simulations with $m = 20$, $n = 10$. Table 3 shows that the MSEs for parametric estimators and the WASEs for nonparametric estimators are not sensitive to the choice of M_n and these estimators are efficient

Table 2 Comparison of the proposed method with B-spline method

| | Method | MSE | | | | WASE | | |
|-----------|--------|-------------|-----------------|-----------------|-----------------|------------------|------------------|------------------|
| | | $\hat{\mu}$ | $\hat{\beta}_1$ | $\hat{\beta}_2$ | $\hat{\beta}_3$ | $\hat{f}_1(u_1)$ | $\hat{f}_2(u_2)$ | $\hat{f}_3(u_3)$ |
| m=10,n=10 | PM | 0.0215 | 0.0063 | 0.0067 | 0.0029 | 0.0028 | 0.0418 | 0.0364 |
| | BM | 0.0318 | 0.0070 | 0.0085 | 0.0036 | 0.1615 | 0.3057 | 0.5053 |
| m=20,n=10 | PM | 0.0180 | 0.0054 | 0.0044 | 0.0013 | 0.0015 | 0.0296 | 0.0267 |
| | BM | 0.0179 | 0.0040 | 0.0053 | 0.0024 | 0.0796 | 0.2164 | 0.2062 |

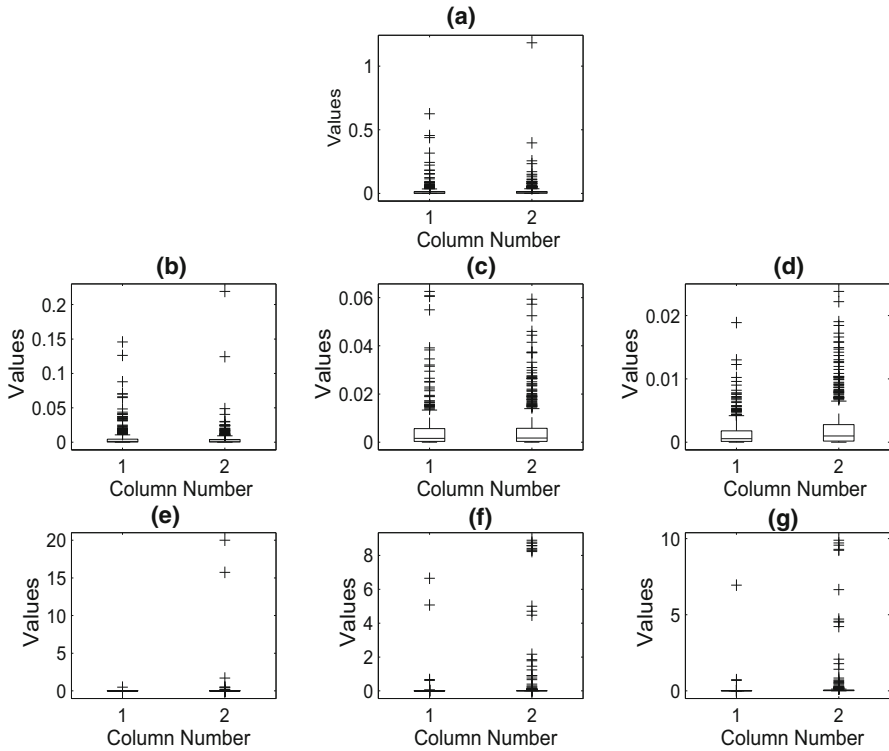


Fig. 2 Box plots for SE and WASE. Label 1 is boxplot for the proposed method and Label 2 is boxplot for B-spline method. **a** is boxplot for $\hat{\mu}$, **b** is boxplot for $\hat{\beta}_1$, **c** is boxplot for $\hat{\beta}_2$, **d** is boxplot for $\hat{\beta}_3$, **e** is boxplot for $\hat{f}_1(x_1)$, **f** is boxplot for $\hat{f}_2(x_2)$ and **g** is boxplot for $\hat{f}_3(x_3)$

under broad ranges of M_n . These results also illustrate that the above *BIC* smoothing parameter selection procedure generally gives the satisfactory results.

In order to investigate the influence of degree of the piecewise polynomial on the estimators, based on 500 simulations, Table 4 reports the simulation results under $q_r = 1, 2, 3$, which correspond to piecewise linear function, piecewise quadratic polynomial and piecewise cubic polynomial respectively. We see from Table 4 that there is no essential difference between the three estimates. Of course, in comparison, piecewise linear function is easier to operate.

Table 3 The MSEs and WASEs for the estimators with different M_N

| M_N | MSE | | | | WASE | | |
|-------|-------------|-----------------|-----------------|-----------------|------------------|------------------|------------------|
| | $\hat{\mu}$ | $\hat{\beta}_1$ | $\hat{\beta}_2$ | $\hat{\beta}_3$ | $\hat{f}_1(x_1)$ | $\hat{f}_2(x_2)$ | $\hat{f}_3(x_3)$ |
| 2 | 0.2777 | 0.1291 | 0.0147 | 0.0097 | 0.0023 | 0.0414 | 0.0040 |
| 3 | 0.1289 | 0.0251 | 0.0321 | 0.0062 | 0.0012 | 0.0145 | 0.0046 |
| 4 | 0.0389 | 0.0108 | 0.0065 | 0.0026 | 0.0006 | 0.0056 | 0.0022 |
| 5 | 0.0303 | 0.0086 | 0.0067 | 0.0020 | 0.0006 | 0.0067 | 0.0042 |
| 6 | 0.0197 | 0.0061 | 0.0047 | 0.0013 | 0.0009 | 0.0080 | 0.0035 |
| 7 | 0.0165 | 0.0050 | 0.0036 | 0.0011 | 0.0009 | 0.0104 | 0.0066 |
| 8 | 0.0144 | 0.0041 | 0.0035 | 0.0010 | 0.0027 | 0.0290 | 0.0226 |
| 9 | 0.0144 | 0.0051 | 0.0029 | 0.0010 | 0.0047 | 0.0879 | 0.0727 |
| 10 | 0.0204 | 0.0062 | 0.0028 | 0.0010 | 0.0439 | 0.0493 | 0.0636 |

Table 4 The MSEs and WAESs for the estimators with different p_r

| | p_r | MSE | | | | WASE | | |
|------------|-------|-------------|-----------------|-----------------|-----------------|------------------|------------------|------------------|
| | | $\hat{\mu}$ | $\hat{\beta}_1$ | $\hat{\beta}_2$ | $\hat{\beta}_3$ | $\hat{f}_1(u_1)$ | $\hat{f}_2(u_2)$ | $\hat{f}_3(u_3)$ |
| m=20, n=10 | 1 | 0.0235 | 0.0031 | 0.0023 | 0.0007 | 0.0014 | 0.0128 | 0.0123 |
| | 2 | 0.0195 | 0.0029 | 0.0024 | 0.0008 | 0.0027 | 0.0135 | 0.0234 |
| | 3 | 0.0206 | 0.0064 | 0.0037 | 0.0008 | 0.0005 | 0.0126 | 0.0150 |
| m=20, n=20 | 1 | 0.0153 | 0.0026 | 0.0015 | 0.0003 | 0.0007 | 0.0016 | 0.0067 |
| | 2 | 0.0067 | 0.0021 | 0.0012 | 0.0004 | 0.0011 | 0.0022 | 0.0123 |
| | 3 | 0.0128 | 0.0045 | 0.0020 | 0.0005 | 0.0004 | 0.0018 | 0.0149 |

In order to investigate the performances of our estimators based on irregular grid data, we first generate data over the rectangular grid of $m \times n$ sites. We then randomly selected half of the data to fit the model, that is, each site has a probability 0.5 of being selected. The final sample size is about $mn/2$. For different m and n , Table 5 summaries the MSEs of the parametric estimators and the MISEs of the nonparametric estimators $\hat{f}_r(x_r)$ for $r = 1, 2, 3$ based on 500 simulations. Comparing Table 5 with Table 1, we see that there is no clear difference between two estimators of μ and $\beta_r, r = 1, 2, 3$. When the sample size is small ($N=100,200$), the estimators of $f_r(x_r)$ for $r = 1, 2, 3$ based on regular grid data is slightly better than that based on irregular grid data, but the sample size is large ($N=300,400,600$), there is no clear difference between two estimators of $f_r(x_r)$ for $r = 1, 2, 3$. All these show that our estimation method is also efficient for irregular grid data.

5 A real data example

In this section, we apply our methods to analyze a real estate data set which includes the real estate data for 203 second-, third-and fourth-tier cities in China in 2016. Our

Table 5 MSEs and MISEs of the estimators based on irregular grid data

| | MSE | | | | MISE | | |
|-----------|-------------|-----------------|-----------------|-----------------|------------------|------------------|------------------|
| | $\hat{\mu}$ | $\hat{\beta}_1$ | $\hat{\beta}_2$ | $\hat{\beta}_3$ | $\hat{f}_1(u_1)$ | $\hat{f}_2(u_2)$ | $\hat{f}_3(u_3)$ |
| m=20,n=10 | 0.0784 | 0.0070 | 0.0081 | 0.0030 | 0.7366 | 0.4266 | 0.5191 |
| m=20,n=20 | 0.0513 | 0.0062 | 0.0047 | 0.0013 | 0.4999 | 0.3119 | 0.4523 |
| m=30,n=20 | 0.0406 | 0.0041 | 0.0030 | 0.0008 | 0.3702 | 0.2130 | 0.4154 |
| m=40,n=20 | 0.0334 | 0.0061 | 0.0027 | 0.0006 | 0.3437 | 0.1707 | 0.3962 |
| m=30,n=40 | 0.0250 | 0.0037 | 0.0016 | 0.0004 | 0.2760 | 0.1606 | 0.3842 |

purpose is to study the relationship between urban housing prices and their influencing factors. The response variable is urban housing price (Price). The covariates of primary interests include the average annual income of urban residents (Income), urban category, urban population, urban GDP, the ratio of no house and one house (RNO), bank interest rate (BIR), urban livability index (ULI), urban comprehensive competitiveness (UCC) and urban development index (UDI). We note that among these variables the data of some variables such as population and GDP are very large, whereas that of some variables such as ULI and UCC are small. For this purpose, for each data of these variables, we first make the following modification: each observation was divided by the maximum of all observations for this variable, so that the maximum of modified data of the variable is 1.

We first consider the following partially linear additive regression

$$\log(\text{Price}) = \mu + \beta_1 Z_1 + \beta_2 Z_2 + f_1(\text{Income}) + f_2(\text{Population}) + f_3(\text{GDP}) + f_4(\text{RNO}) + f_5(\text{BIR}) + f_6(\text{ULI}) + f_7(\text{UCC}) + f_8(\text{UDI}) + \varepsilon, \tag{5.1}$$

where $Z_1 = 1, Z_2 = 0$ stands for second-tier city, $Z_1 = 0$ and $Z_2 = 1$ stands for third-tier city and $Z_1 = 0$ and $Z_2 = 0$ stands for fourth-tier city.

The estimators of β_1 and β_2 are computed by (2.2) with $p_r = 1$, the estimators of μ is computed by (2.6). The number of subinterval M_N are determined by BIC criterion given in Sect. 2. The estimators of $f_r(\cdot), r = 1, \dots, 8$ are computed by (2.8) with Epanechnikov kernel. Bandwidth h_r are selected by “leave-one-out” cross-validation procedure. Table 6 exhibits the parametric estimators, and Fig. 3 shows the estimated curves of the additive components and their 95% pointwise confidence intervals. The fact that $\hat{\beta}_1 > \hat{\beta}_2 > 0$ in Table 6 indicates that the housing price for a second-tier city is larger than that for a third-tier city and the housing price for a third-tier city is larger than that for a fourth-tier city. We see from Fig. 3 that the estimated curves $\hat{f}_1(\text{Income}), \hat{f}_4(\text{RNO})$ and $\hat{f}_8(\text{UDI})$ show rising trends as Income, RNO and UDI increase, while $\hat{f}_5(\text{BIR})$ show decreasing trends as BIR increase. Figure 3 shows the subtle impact of the urban population and urban livability index (ULI) on housing prices, which overall declines as the urban population increases, while housing prices overall rise along with the urban livability index. This can be explained that some central and western cities, such as Fuyang, Shangqiu, Nanyang and Handan, have

Table 6 The parametric estimators for model (5.1). The values in parenthesis are standard error

| $\hat{\mu}$ | $\hat{\beta}_1$ | $\hat{\beta}_2$ |
|-------------|-----------------|-----------------|
| 8.4092 | 0.6110 | 0.3325 |
| (0.0110) | (0.0385) | (0.0135) |

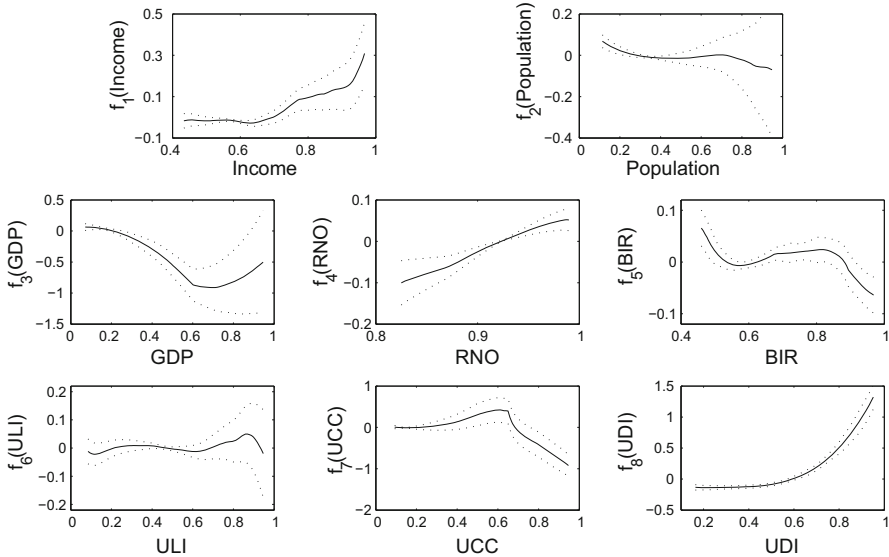


Fig. 3 The estimated curves (solid line) and their 95% pointwise confidence intervals (dotted lines)

large populations, while housing prices in these cities are low. Figure 3 shows that $\hat{f}_3(GDP)$ appears the changing trend of decreasing first and then rising, this is due to lower house prices in some central and western cities, and higher prices in some eastern cities with high GDP. Figure 3 also shows that $\hat{f}_7(UCC)$ appears the changing trend of rising first and then decreasing.

We find from Fig. 3 that the estimated curves of $f_2(Population)$, $f_4(RNO)$ and $f_6(ULI)$ approximate straight line. We then further construct the following partially linear additive regression

$$\log(Price) = \mu + \beta_1 Z_1 + \beta_2 Z_2 + \beta_3 Population + \beta_4 RNO + \beta_5 ULI + f_1(Income) + f_2(GDP) + f_3(BIR) + f_4(UCC) + f_5(UDI) + \varepsilon, \tag{5.2}$$

Table 7 displays the parametric estimators and Fig. 4 shows the estimated curves of the additive components and their 95% pointwise confidence intervals in model (5.2). We see from Table 7 that $\log(Price)$ is negatively associated with urban population and positively associated with the ratio of no house and one house and urban livability index. Since the absolute values of $\hat{\beta}_3$ and $\hat{\beta}_5$ are small, it is shown to have a small effect on $\log(Price)$. Figure 4 shows that the estimated curves in model (5.2) are similar to that in model (5.1).

Table 7 The parametric estimators for model (5.2). The values in parenthesis are standard error

| $\hat{\mu}$ | $\hat{\beta}_1$ | $\hat{\beta}_2$ | $\hat{\beta}_3$ | $\hat{\beta}_4$ | $\hat{\beta}_5$ |
|--------------------|--------------------|--------------------|---------------------|--------------------|--------------------|
| 7.9804 (0.1475) | 0.4515 (0.0413) | 0.2860 (0.0178) | -0.1311 (0.0694) | 0.5394 (0.1651) | 0.0519 (0.0870) |

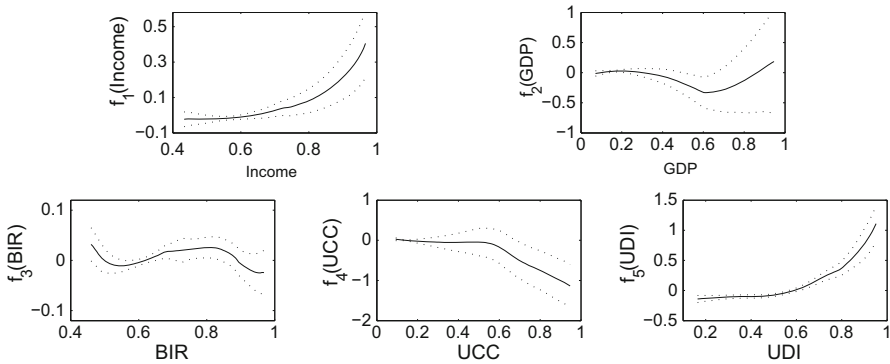


Fig. 4 The estimated curves (solid line) and their 95% pointwise confidence intervals (dotted lines)

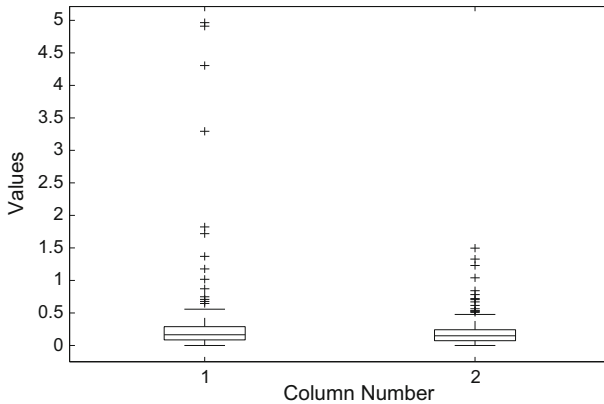


Fig. 5 Boxplot for the absolute prediction error $|\log(\widehat{Price}_i) - \log(Price_i)|$, $i = 1, \dots, 203$ for two models. Here 1 is the boxplot for model (5.1) and 2 is the boxplot for models (5.2)

To evaluate the prediction performance of model (5.1) and model (5.2), we applied leave-one-out cross-validation to the data; i.e., when predicting the housing price for the i th city, we omit the data for this city when fitting the models. Figure 5 displays the boxplots for the absolute prediction errors $|\log(\widehat{Price}_i) - \log(Price_i)|$, $i = 1, \dots, 203$, for models (5.1) and (5.2), where $Price_i$ denotes the housing price for the i th city. The mean values of these errors for the two models are 0.3050 and 0.2031 respectively. We see from Fig. 5 and the mean absolute prediction errors that model (5.2) behaviors better than model (5.1) in prediction performance.

6 Conclusion and future studies

We have proposed a new estimation procedure for partially linear additive regressions with spatial data using piecewise polynomials combined with local linear methods. The proposed method is suitable for high dimensional data, there is no need to solve the restricted minimization problem and no iterative algorithms are needed. Under mild regularity assumptions, the asymptotic distribution of the estimator of the unknown parameter vector is established, the asymptotic distributions of the estimators of the unknown functions are also derived.

We method focuses on data on a regular grid. Simulation studies in Sect. 4 demonstrate the proposed procedure is also efficient for data irregularly positioned, but the relevant theoretical properties will need to be developed, which are left for future research. As pointed out by a referee, geoadditive models in which spatial effect is modeled via a bivariate function have been proposed and investigated in literature. Our method can be extended to geoadditive models. We leave such extensions to future work.

Appendix: Proofs

In this section, let $C > 0$ denote a generic constant of which the value may change from line to line. For a matrix $B = (b_{ij})$, set $\|B\|_\infty = \max_i \sum_j |b_{ij}|$ and $|B|_\infty = \max_{i,j} |b_{ij}|$. For a vector $v = (v_1, \dots, v_k)^T$, set $\|v\|_\infty = \sum_{j=1}^k |v_j|$ and $|v|_\infty = \max_{1 \leq j \leq k} |v_j|$.

Let $f_{rv}(x_r) = \chi_{rv}(x_r) f_r(x_r)$, $\bar{f}_{rv} = (\sum_{i=1}^m \sum_{j=1}^n f_{rv}(X_{ijr})) / (\sum_{i=1}^m \sum_{j=1}^n \chi_{rv}(X_{ijr}))$,

$$\check{f}_{rv}(x_r) = f_r(x_{rv}) \check{A}_{rv0}(x_r) + h_{r0} f'_r(x_{rv}) \check{A}_{rv1}(x_r) + \dots + h_{r0}^{pr} f_r^{(pr)}(x_{rv}) \check{A}_{rvqr}(x_r) / pr!,$$

$\check{\check{f}}_{rv} = (\sum_{i=1}^m \sum_{j=1}^n \check{f}_{rv}(X_{ijr})) / (\sum_{i=1}^m \sum_{j=1}^n \chi_{rv}(X_{ijr}))$ and $f_{rv}^*(X_{ijr}) = [f_{rv}(X_{ijr}) - \chi_{rv}(X_{ijr}) \bar{f}_{rv}] - [\check{f}_{rv}(X_{ijr}) - \chi_{rv}(X_{ijr}) \check{\check{f}}_{rv}]$. Noting that $f_r(x_r) = \sum_{v=1}^{M_{rN}} f_{rv}(x_r)$ and $f_{rv}(X_{ijr}) = \chi_{rv}(X_{ijr}) \bar{f}_{rv} + [\check{f}_{rv}(X_{ijr}) - \chi_{rv}(X_{ijr}) \check{\check{f}}_{rv}] + f_{rv}^*(X_{ijr})$, we get that

$$f_r(X_{ijr}) = \mathbf{A}_r^T(X_{ijr}) \mathbf{a}_{0r} + \bar{F}_{rM_{rN}} \chi_{rM_{rN}}(X_{ijr}) + \sum_{v=1}^{M_{rN}} f_{rv}^*(X_{ijr}), \tag{A.1}$$

where $\mathbf{a}_{0r} = (\mathbf{a}_{0r1}^T, \dots, \mathbf{a}_{0rM_{rN}}^T)^T$ with $\mathbf{a}_{0rv} = (\bar{f}_{rv}, h_{r0} f'_r(x_{rv}), \dots, h_{r0}^{pr} f_r^{(pr)}(x_{rv}) / pr!)^T$ for $v = 1, \dots, M_{rN} - 1$ and $\mathbf{a}_{0rM_{rN}} = (h_{r0} f'_r(x_{rM_{rN}}), \dots, h_{r0}^{pr} f_r^{(pr)}(x_{rM_{rN}}) / pr!)^T$, $\bar{F}_{rM_{rN}} = (\sum_{i=1}^m \sum_{j=1}^n f_r(X_{ijr})) / (\sum_{i=1}^m \sum_{j=1}^n \chi_{rM_{rN}}(X_{ijr}))$. Let $f_{krv}(x_r) = \chi_{krv}(x_r) f_{kr}(x_r)$, $f_{krv}^*(X_{ijr})$, $k = 1, \dots, d_1$ and $\bar{F}_{krM_{rN}}$ are defined similarly as $\bar{f}_{rv}^*(X_{ijr})$ and $\bar{F}_{rM_{rN}}$. Denote $\bar{Y} = \bar{Y} - E(Y)$, $\bar{Z} = \bar{Z} - E(Z)$,

$\check{Z}_{ij} = (\check{Z}_{ij1}, \dots, \check{Z}_{ijd_1})^T$ with

$$\check{Z}_{ijk} = \sum_{r=1}^{d_2} (\bar{F}_{krM_{rN}} \chi_{rM_{rN}}(X_{ijr}) + \sum_{v=1}^{M_{rN}} f_{krv}^*(X_{ijr})) - \bar{Z}_k + V_{ijk}, \quad k = 1, \dots, d_1. \tag{A.2}$$

Let

$$\check{Y}_{ij} = \sum_{r=1}^{d_2} (\bar{F}_{rM_{rN}} \chi_{rM_{rN}}(X_{ijr}) + \sum_{v=1}^{M_{rN}} f_{rv}^*(X_{ijr})) + \check{\mathbf{Z}}^T \boldsymbol{\beta}_0 - \bar{Y} + \varepsilon_{ij}. \tag{A.3}$$

Then, we have

$$\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0 = \check{\boldsymbol{\Gamma}}_N^{-1} \left[\sum_{i=1}^m \sum_{j=1}^n \check{\mathbf{Z}}_{ij} \check{Y}_{ij} - \check{\mathbf{W}}_N \mathbf{A}_N^{-1} (\sum_{i=1}^m \sum_{j=1}^n \mathbf{A}(X_{ij}) \check{Y}_{ij}) \right], \tag{A.4}$$

with

$$\check{\boldsymbol{\Gamma}}_N = \sum_{i=1}^m \sum_{j=1}^n \check{\mathbf{Z}}_{ij} \check{\mathbf{Z}}_{ij}^T - \check{\mathbf{W}}_N \mathbf{A}_N^{-1} \check{\mathbf{W}}_N^T, \quad \check{\mathbf{W}}_N = \sum_{i=1}^m \sum_{j=1}^n \check{\mathbf{Z}}_{ij} \mathbf{A}(X_{ij})^T. \tag{A.5}$$

Lemma A.1 *Under assumptions 1–5, it holds that*

$$\max_{1 \leq r \leq d, 1 \leq v \leq M_{rN}, 1 \leq k \leq p_r} \left| \frac{1}{N} \sum_{i=1}^m \sum_{j=1}^n \check{A}_{rvk}(X_{ijr}) - E(\check{A}_{rvk}(X_r)) \right| = o_p(M_N^{-2}).$$

Proof Let $D_N = \{(i, j) : 1 \leq i \leq m, 1 \leq j \leq n\}$, $\check{A}_{rvk}(X_{ijr}) = \check{A}_{rvk}(X_{ijr}) - E(\check{A}_{rvk}(X_{ijr}))$ and $A^* = \max_{1 \leq r \leq d, 1 \leq v \leq M_{rN}, 1 \leq k \leq p_r} \left| \frac{1}{N} \sum_{i=1}^m \sum_{j=1}^n \check{A}_{rvk}(X_{ijr}) \right|$. For any sufficiently small positive constant ε , we have

$$P(A^* > \varepsilon M_N^{-2}) \leq \frac{M_N^4}{\varepsilon^2 N^2} \sum_{r=1}^{d_2} \sum_{v=1}^{M_{rN}} \sum_k \left[\sum_{(i,j) \in D_N} E(\check{A}_{rvk}^2(X_{ijr})) + \sum_{(i',j') \neq (i,j) \in D_N} |E(\check{A}_{rvk}(X_{ijr}) \check{A}_{rvk}(X_{i'j'r}))| \right]. \tag{A.6}$$

Let $c_{Nk} = [M_N^{\delta/(2+\delta)\tau}]$ for $k = 1, 2$, where $\tau > 2(4 + \delta)/(2 + \delta)$ is a constant. Let the set $\{(i, j) \neq (i', j') \in D_N\}$ be split into the following two parts

$$\begin{aligned} \mathbf{S}_1 &= \{(i, j) \neq (i', j') \in D_N : |i - i'| \leq c_{N1}, |j - j'| \leq c_{N2}\}, \\ \mathbf{S}_2 &= \{(i, j) \neq (i', j') \in D_N : |i - i'| > c_{N1} \text{ or } |j - j'| > c_{N2}\}. \end{aligned}$$

By assumption 5, we have

$$\frac{M_N^4}{\varepsilon^2 N^2} \sum_{r=1}^{d_2} \sum_{v=1}^{M_{rN}} \sum_k \sum_{(i',j'),(i,j)} \sum_{\in \mathbf{S}_1} |E(\check{A}_{rvk}(X_{ijr}) \check{A}_{rvk}(X_{i'j'r}))| \leq C M_N^{4+2\delta/(2+\delta)\tau} / N = o(1). \tag{A.7}$$

Turning to S_2 , using Lemma 5.1 of Hallin et al. (2004b), we obtain that

$$|E(\tilde{A}_{rvk}(X_{ijr})\tilde{A}_{rvk}(X_{i'j'r}))| \leq C(E|\tilde{A}_{rvk}(X_{ijr})|^{2+\delta})^{2/(2+\delta)}(\varphi(\|(i', j') - (i, j)\|))^{\delta/(2+\delta)} \leq CM_N^{-2/(2+\delta)}(\varphi(\|(i', j') - (i, j)\|))^{\delta/(2+\delta)}.$$

Therefore, by assumptions 4 and 5, we get

$$\frac{M_N^4}{\varepsilon^2 N^2} \sum_{r=1}^{d_2} \sum_{v=1}^{M_{rN}} \sum_k \sum_{(i', j'), (i, j)} \sum_{\in S_2} |E(\tilde{A}_{rvk}(X_{ijr})\tilde{A}_{rvk}(X_{i'j'r}))| \leq C \frac{M_N^4}{\varepsilon^2 N^2} N \sum_{l=1}^2 c_{Nl}^\tau \sum_{t=c_{Nl}}^N t(\varphi(t))^{\delta/(2+\delta)} = o(1). \tag{A.8}$$

Now Lemma A.1 follows from (A.6)–(A.8) and the fact that $\sum_{(i, j) \in D_N} E(\tilde{A}_{rvk}^2(X_{ijr})) \leq NM_N^{-1}$. □

Lemma A.2 *Under assumptions 1-5, it holds that*

$$\|\check{W}_N/N\|_\infty = o_p(M_N^{1/2+\delta/(2+\delta)\tau} / N^{1/2}).$$

Proof We first prove

$$\left\| \frac{1}{N} \sum_{i=1}^m \sum_{j=1}^n \mathbf{A}(X_{ij}) V_{ijl} \right\|_\infty = o_p(M_N^{1/2+\delta/(2+\delta)\tau} / N^{1/2}), \quad l = 1, \dots, d_1. \tag{A.9}$$

Let $\xi_{ijrv1} = \chi_{rv}(X_{ijr}) - E(\chi_{rv}(X_r))\chi_{rM_{rN}}(X_{ijr})/E(\chi_{rM_{rN}}(X_r))$,

$$\begin{aligned} \xi_{rv2} &= \frac{\sum_{i=1}^m \sum_{j=1}^n \chi_{rv}(X_{ijr})}{\sum_{i=1}^m \sum_{j=1}^n \chi_{rM_{rN}}(X_{ijr})} - \frac{E(\chi_{rv}(X_r))}{E(\chi_{rM_{rN}}(X_r))}, \\ \eta_{ijrvk1} &= \left(\frac{X_{ijr} - x_{rv}}{h_{r0}} \right)^k - \frac{E(((X_r - x_{rv})/h_{r0})^k \chi_{rv}(X_r))}{E(\chi_{rv}(X_r))}, \\ \eta_{rvk2} &= \frac{\sum_{i=1}^m \sum_{j=1}^n \chi_{rv}(X_{ijr})((X_{ijr} - x_{rv})/h_{r0})^k}{\sum_{i=1}^m \sum_{j=1}^n \chi_{rv}(X_{ijr})} - \frac{E(\chi_{rv}(X_r)((X_r - x_{rv})/h_{r0})^k)}{E(\chi_{rv}(X_r))}. \end{aligned}$$

Then

$$\begin{aligned} & \left\| \frac{1}{N} \sum_{i=1}^m \sum_{j=1}^n \mathbf{A}(X_{ij}) V_{ijl} \right\|_\infty \\ & \leq \sum_{r=1}^{d_2} \sum_{v=1}^{M_{rN}-1} \frac{1}{N} \left| \sum_{i=1}^m \sum_{j=1}^n (\xi_{ijrv1} + \chi_{rM_{rN}}(X_{ijr})\xi_{rv2}) V_{ijl} \right| \\ & \quad + \sum_{r=1}^{d_2} \sum_{v=1}^{M_{rN}} \sum_{k=1}^{p_r} \frac{1}{N} \left| \sum_{i=1}^m \sum_{j=1}^n \chi_{rv}(X_{ijr})(\eta_{ijrvk1} + \eta_{rvk2}) V_{ijl} \right|. \end{aligned} \tag{A.10}$$

Similar to the proof of Lemma A.1, we obtain that $E[\sum_{i=1}^m \sum_{j=1}^n \xi_{ijrv1} V_{ijl}]^2 \leq CNM_N^{2\delta/((2+\delta)\tau)-1}$ and $E[\sum_{i=1}^m \sum_{j=1}^n \chi_{rv}(X_{ijr})\eta_{ijrvk1} V_{ijl}]^2 \leq CNM_N^{2\delta/((2+\delta)\tau)-1}$. Hence,

$$\sum_{r=1}^{d_2} \sum_{v=1}^{M_{rN}-1} \frac{1}{N} \left| \sum_{i=1}^m \sum_{j=1}^n \xi_{ijrv1} V_{ijl} \right| = o_p(M_N^{1/2+\delta/(2+\delta)\tau} / N^{1/2}), \tag{A.11}$$

$$\sum_{r=1}^{d_2} \sum_{v=1}^{M_{rN}} \sum_{k=1}^{p_r} \frac{1}{N} \left| \sum_{i=1}^m \sum_{j=1}^n \chi_{rv}(X_{ijr}) \eta_{ijrvk1} V_{ijl} \right| = o_p(M_N^{1/2+\delta/(2+\delta)\tau} / N^{1/2}). \tag{A.12}$$

Lemma A.1 implies $\max_{1 \leq r \leq d_2, 1 \leq v \leq M_{rN}-1} |\xi_{rv2}| = o_p(M_N^{-1})$, $\max_{1 \leq r \leq d_2, 1 \leq v \leq M_{rN}} |\eta_{rvk2}| = o_p(M_N^{-1})$. Since $E[\chi_{rv}(X_{ijr})V_{ijl}] = E[(\chi_{rv}(X_{ijr}) - E\chi_{rv}(X_{ijr}))V_{ijl}] + E(\chi_{rv}(X_{ijr}))E(V_{ijl}) = 0$, then by arguments similar to those used in the proof of Lemma A.1, we have

$$\frac{1}{N} \left| \sum_{i=1}^m \sum_{j=1}^n \chi_{rM_{rN}}(X_{ijr}) V_{ijl} \right| = o_p(N^{-1/2}), \tag{A.13}$$

Therefore,

$$\sum_{r=1}^{d_2} \sum_{v=1}^{M_{rN}-1} \frac{|\xi_{rv2}|}{N} \left| \sum_{i=1}^m \sum_{j=1}^n \chi_{rM_{rN}}(X_{ijr}) V_{ijl} \right| = o_p(N^{-1/2}), \tag{A.14}$$

$$\sum_{r=1}^{d_2} \sum_{v=1}^{M_{rN}} \sum_{k=1}^{p_r} \frac{|\eta_{rvk2}|}{N} \left| \sum_{i=1}^m \sum_{j=1}^n \chi_{rv}(X_{ijr}) V_{ijl} \right| = o_p(N^{-1/2}). \tag{A.15}$$

Now (A.9) follows from (A.10)–(A.12), (A.14) and (A.15). Similar to the proof of (A.9), we deduce that

$$\left\| \frac{1}{N} \sum_{i=1}^m \sum_{j=1}^n \chi_{rM_{rN}}(X_{ijr}) \mathbf{A}(\mathbf{X}_{ij}) \right\|_{\infty} = o_p(M_N^{-1}). \tag{A.16}$$

Using the fact that $E(f_r(X_r)) = 0$, we get that $\bar{F}_{rM_{rN}} = O_p(M_N^{3/2} / N^{1/2})$. Hence,

$$\left\| \frac{1}{N} \sum_{i=1}^m \sum_{j=1}^n \bar{F}_{rM_{rN}} \chi_{rM_{rN}}(X_{ijr}) \mathbf{A}(\mathbf{X}_{ij}) \right\|_{\infty} = o_p(M_N^{1/2} / N^{1/2}). \tag{A.17}$$

Similar to the proof of (A.9) and using Assumptions , we have

$$\left\| \frac{1}{N} \sum_{i=1}^m \sum_{j=1}^n \left[\sum_{r=1}^{d_2} \sum_{v=1}^{M_{rN}} f_{krv}^*(X_{ijr}) \right] \mathbf{A}(\mathbf{X}_{ij}) \right\|_{\infty} = o_p(M_N^{-p_r} N^{-1/2}). \tag{A.18}$$

Similar to the proof of (A.17), we get that

$$\left\| \frac{1}{N} \sum_{i=1}^m \sum_{j=1}^n \bar{Z}_k \mathbf{A}(\mathbf{X}_{ij}) \right\|_{\infty} = o_p(M_N^{-1/2} N^{-1/2}). \tag{A.19}$$

Now Lemma A.2 follows from (A.9) and (A.17)–(A.19) and Assumption 5. □

Lemma A.3 *Under Assumptions 1-5, it holds that*

$$\check{\Gamma}_N/N = \Gamma + o_p(1).$$

Proof We first prove that $M_N \mathbf{A}_N/N$ is invertible. Let λ_{min} be the minimum eigenvalue of $M_N \mathbf{A}_N/N$. By Lemma 3 of Stone (1985) and Lemma A.1 and using Assumption 4 and the fact that $\chi_{rv}(X_{ijr})\chi_{rv'}(X_{ijr}) = 0$ for $v \neq v'$, we have that

$$\begin{aligned} \lambda_{min} &= \inf_{\|\mathbf{a}\|=1} \frac{M_N}{N} \mathbf{a}^T \mathbf{A}_N \mathbf{a} \\ &= \inf_{\|\mathbf{a}\|=1} \frac{M_N}{N} \sum_{i=1}^m \sum_{j=1}^n \left(\sum_{r=1}^{d_2} [\sum_{v=1}^{M_{rN}-1} \sum_{k=0}^{p_r} a_{rvk} A_{rvk}(X_{ijr}) \right. \\ &\quad \left. + \sum_{k=1}^{p_r} a_{rM_{rN}k} A_{rM_{rN}k}(X_{ijr})] \right)^2 \\ &\geq C \inf_{\|\mathbf{a}\|=1} \sum_{r=1}^{d_2} \frac{M_N}{N} \sum_{i=1}^m \sum_{j=1}^n \left[\sum_{v=1}^{M_{rN}-1} (a_{rv0} \chi_{rv}(X_{ijr}) \right. \\ &\quad \left. + \sum_{k=1}^{p_r} a_{rvk} A_{rvk}(X_{ijr}))^2 \right. \\ &\quad \left. + (-\chi_{rM_{rN}}(X_{ijr}) \sum_{v=1}^{M_{rN}-1} a_{rv0} E_{rM_{rN}v} + \sum_{k=1}^{p_r} a_{rM_{rN}k} A_{rM_{rN}k}(X_{ijr}))^2 \right] \\ &\geq C \inf_{\|\mathbf{a}\|=1} \sum_{r=1}^{d_2} \left(\sum_{v=1}^{M_{rN}-1} \mathbf{a}_{rv}^T G_r \mathbf{a}_{rv} + \mathbf{a}_{rM_{rN}}^T G_r^* \mathbf{a}_{rM_{rN}} \right) + o_p(1), \end{aligned}$$

where $E_{rM_{rN}v} = \sum_{i=1}^m \sum_{j=1}^n \chi_{rv}(X_{ijr}) / \sum_{i=1}^m \sum_{j=1}^n \chi_{rM_{rN}}(X_{ijr})$, $G_r = (g_{rij})_{(p_r+1) \times (p_r+1)}$ with $g_{r11} = 2$ and $g_{rij} = \int_{-1}^1 x_r^{i+j-2} dx_r - \int_{-1}^1 x_r^{i-1} dx_r \int_{-1}^1 x_r^{j-1} dx_r / 2$ for $i > 1$ or $j > 1$, $G_r^* = (g_{rij}^*)_{p_r \times p_r}$ with $g_{rij}^* = \int_{-1}^1 x_r^{i+j} dx_r - \int_{-1}^1 x_r^i dx_r \int_{-1}^1 x_r^j dx_r / 2$ and $\mathbf{a}_{rv} = (a_{rv0}, a_{rv1}, \dots, a_{rvp_r})^T$ for $v = 1, \dots, M_{rN} - 1$ and $\mathbf{a}_{rM_{rN}} = (a_{rM_{rN}1}, \dots, a_{rM_{rN}p_r})^T$. For fixed p_r , it is easy to prove that G_r and G_r^* are positive definite. Hence, there exists a positive constant C_1^* such that $\lambda_{min} \geq C_1^* + o_p(1)$ and consequently $M_N \mathbf{A}_N/N$ is invertible. By arguments similar to those used to prove Lemma A.1 and using the fact that $E(f_{kr}(U_r)) = 0$ for $k = 1, \dots, d_1$, we get that

$$\frac{1}{N} \sum_{i=1}^m \sum_{j=1}^n \check{\mathbf{Z}}_{ij} \check{\mathbf{Z}}_{ij}^T = E(\mathbf{V}\mathbf{V}^T) + o_p(1). \tag{A.20}$$

Using Lemma A.2 and assumption 5, we obtain that

$$|\check{\mathbf{W}}_N \mathbf{A}_N^{-1} \check{\mathbf{W}}_N^T / N|_\infty \leq M_N \|\check{\mathbf{W}}_N / N\|_\infty^2 \cdot |(M_N \mathbf{A}_N / N)^{-1}|_\infty = o_p(1). \tag{A.21}$$

Now Lemma A.3 follows from (A.5), (A.20) and (A.21). □

Proof of Theorem 3.1 Using (A.13), we obtain that

$$\sum_{r=1}^{d_2} \frac{1}{N} \sum_{i=1}^m \sum_{j=1}^n \bar{F}_{rM_{rN}} \chi_{rM_{rN}}(X_{ijr}) V_{ijk} = o_p(M_N^{3/2} N^{-1}). \tag{A.22}$$

Similar to the proof of (A.18), we have

$$\frac{1}{N} \sum_{i=1}^m \sum_{j=1}^n \left[\sum_{r=1}^{d_2} \sum_{v=1}^{M_{rN}} f_{krv}^*(X_{ijr}) \right] V_{ijk} = o_p(M_N^{-Pr} N^{-1/2}). \tag{A.23}$$

Since

$$\frac{1}{N} \sum_{i=1}^m \sum_{j=1}^n (\bar{\mathbf{Z}}^T \boldsymbol{\beta}_0 + \bar{Y}) V_{ijk} = (\bar{\mathbf{Z}}^T \boldsymbol{\beta}_0 + \bar{Y}) \frac{1}{N} \sum_{i=1}^m \sum_{j=1}^n V_{ijk} = o_p(M_N N^{-1}), \tag{A.24}$$

then by (A.3), we have

$$N^{-1/2} \sum_{i=1}^m \sum_{j=1}^n \check{Y}_{ij} V_{ijk} = N^{-1/2} \sum_{i=1}^m \sum_{j=1}^n \varepsilon_{ij} V_{ijk} + o_p(1). \tag{A.25}$$

Similarly, $N^{-1/2} \sum_{i=1}^m \sum_{j=1}^n (\check{Z}_{ijk} - V_{ijk}) \varepsilon_{ij} = o_p(1)$. Under the assumptions of Theorem 3.1, it is easy to prove that

$$N^{-1/2} \sum_{i=1}^m \sum_{j=1}^n (\check{Z}_{ijk} - V_{ijk})(\check{Y}_{ij} - \varepsilon_{ij}) = o_p(1).$$

Hence,

$$N^{-1/2} \sum_{i=1}^m \sum_{j=1}^n \check{\mathbf{Z}}_{ij} \check{Y}_{ij} = N^{-1/2} \sum_{i=1}^m \sum_{j=1}^n \mathbf{V}_{ij} \varepsilon_{ij} + o_p(1). \tag{A.26}$$

Similar to the proof of Lemma A.2, we deduce that

$$\left\| \frac{1}{N} \sum_{i=1}^m \sum_{j=1}^n \mathbf{A}(\mathbf{X}_{ij}) \check{Y}_{ij} \right\|_\infty = o_p(M_N^{1/2+\delta/(2+\delta)\tau} / N^{1/2}). \tag{A.27}$$

Therefore, Using Lemma A.2 and (A.27), we conclude that

$$\begin{aligned} & N^{-1/2} \|\check{\mathbf{W}}_N \mathbf{A}_N^{-1} \left(\sum_{i=1}^m \sum_{j=1}^n \mathbf{A}(\mathbf{X}_{ij}) \check{Y}_{ij} \right)\|_\infty \\ & \leq N^{1/2} M_N \|\check{\mathbf{W}}_N / N\|_\infty \cdot \|(M_N \mathbf{A}_N / N)^{-1}\|_\infty \cdot \left\| \frac{1}{N} \sum_{i=1}^m \sum_{j=1}^n \mathbf{A}(\mathbf{X}_{ij}) \check{Y}_{ij} \right\|_\infty \\ & = N^{1/2} M_N o_p(M_N^{1+2\delta/(2+\delta)\tau} / N) = o_p(1). \end{aligned} \tag{A.28}$$

By arguments similar to those used in the proof of Lemma 6 of Tang and Cheng (2009), we can prove that $N^{-1/2} \sum_{i=1}^m \sum_{j=1}^n \mathbf{V}_{ij} \varepsilon_{ij}$ is asymptotically normal. Therefore, (3.2) follows from (A.4), Lemma (A.3), (A.25) and (A.28). The proof of Theorem 3.1 is completed. \square

Proof of Theorem 3.2 Let $\mathbf{a}_0 = (\mathbf{a}_{01}^T, \dots, \mathbf{a}_{0d}^T)^T$. By arguments similar to those used to prove Lemma A.2, we deduce that

$$\begin{aligned} \|\tilde{\mathbf{a}} - \mathbf{a}_0\|^2 &\leq CM_N^2 \left\| \sum_{i=1}^m \sum_{j=1}^n \mathbf{A}(\mathbf{X}_{ij}) (\bar{Y}_{ij} - \bar{\mathbf{Z}}_{ij}^T \hat{\boldsymbol{\beta}} - \mathbf{A}^T(\mathbf{X}_{ij}) \mathbf{a}_0) \right\|^2 \\ &= o_p(M_N^{2+2\delta/(2+\delta)\tau} / N). \end{aligned}$$

Hence, under the assumptions of Theorem 3.2, by arguments similar to those used to prove Lemma A.2, we obtain that

$$\begin{aligned} &\left(\sum_{i=1}^m \sum_{j=1}^n K\left(\frac{X_{ijr} - x_{0r}}{h_r}\right) \left(\sum_{r' \neq r} \tilde{f}_{r'}(X_{ijr'}) - \sum_{r' \neq r} f_{r'}(X_{ijr'}) \right) \right)^2 \\ &\leq \|\tilde{\mathbf{a}} - \mathbf{a}_0\|^2 \left\| \sum_{i=1}^m \sum_{j=1}^n K\left(\frac{X_{ijr} - x_{0r}}{h_r}\right) \mathbf{A}_{-r}(\mathbf{X}_{ij}) \right\|^2 \\ &+ \left[\sum_{i=1}^m \sum_{j=1}^n K\left(\frac{X_{ijr} - x_{0r}}{h_r}\right) \sum_{r' \neq r} (\bar{F}_{r'M_{r'N}} \chi_{r'M_{r'N}}(X_{ijr'}) + \sum_{v=1}^{M_{r'N}} f_{r'v}^*(X_{ijr'})) \right]^2 \\ &= o_p(NM_N^{1+2\delta/(2+\delta)\tau} h_r^2) = o_p(Nh_r), \end{aligned}$$

where $\mathbf{A}_{-r}(\mathbf{X}_{ij}) = (\mathbf{A}_1(X_{ij1}), \dots, \mathbf{A}_{r-1}(X_{ij(r-1)}), \mathbf{A}_{r+1}(X_{ij(r+1)}), \dots, \mathbf{A}_d(X_{ijd}))^T$. Therefore,

$$(Nh_r)^{-1/2} \sum_{i=1}^m \sum_{j=1}^n K\left(\frac{X_{ijr} - x_{0r}}{h_r}\right) \mathbf{B}_{ijr} \left(\sum_{r' \neq r} \tilde{f}_{r'}(X_{ijr'}) - \sum_{r' \neq r} f_{r'}(X_{ijr'}) \right) = o_p(1). \tag{A.29}$$

Now by arguments similar to those used in the proof of Theorem 3.1 of Hallin et al. (2004) and using (A.29), we can easily complete the proof of Theorem 3.2. \square

References

Al-Sulami D, Jiang Z, Lu Z, Zhu J (2017) Estimation for semiparametric nonlinear regression of irregularly located spatial time-series data. *Econom Stat* 2:22–35

Athreya KB, Pantala SG (1986) A note on strong mixing of ARMA processes. *Stat Probab Lett* 4:187–190

Brieman L, Friedman J (1985) Estimating optimal transformations for multiple regression and correlation (with discussion). *J Am Stat Assoc* 80:580–619

Buja A, Hastie T, Tibshirani R (1989) Linear smoothers and additive models (with discussion). *Ann Stat* 17:453–555

Cheng G, Zhou L, Huang J (2014) Efficient semiparametric estimation in generalized partially linear additive models for longitudinal/clustered data. *Bernoulli* 20:141–163

Cressie NAC (1991) *Statistics for spatial data*. Wiley, New York

Fan Y, Li Q (2003) A kernel-based method for estimating additive partially linear models. *Stat Sin* 13:739–62

Friedman J, Stuetzle W (1981) Projection pursuit regression. *J Am Stat Assoc* 76:817–823

- Gao J, Lu Z, Tjøstheim D (2006) Estimation in semiparametric spatial regression. *Ann Stat* 34:1395–1435
- Hallin M, Lu Z, Tran LT (2001) Density estimation for spatial linear processes. *Bernoulli* 7:657–668
- Hallin M, Lu Z, Tran LT (2004) Local linear spatial regression. *Ann Stat* 32:2469–2500
- Huang J, Horowitz J, Wei F (2010) Variable selection in nonparametric additive models. *Ann Stat* 38:2282–2313
- Lee YK, Choi H, Park BU, Yu KS (2004) Local likelihood density estimation on random fields. *Stat Probab Lett* 68:347–57
- Li Q (2000) Efficient estimation of additive partially linear models. *Int Econ Rev* 41:1073–1092
- Lian H, Liang H, Ruppert D (2015) Separation of covariates into nonparametric and parametric parts in high-dimensional partially linear additive models. *Stat Sin* 25:591–607
- Liang H, Thurston S, Ruppert D, Apanasovich T, Hauser R (2008) Additive partial linear models with measurement errors. *Biometrika* 95:667–678
- Nandy S, Lim C, Maiti T (2017) Additive model building for spatial regression. *J R Stat Soc B* 79:779–800
- Ramsay T (2002) Spline smoothing over difficult regions. *J R Stat Soc B* 64:307–319
- Ripley B (1981) *Spatial statistics*. Wiley, New York
- Rosenblatt M (1956) A central limit theorem and a strong mixing condition. *Proc Natl Acad Sci USA* 42:43–47
- Sherwood B, Wang L (2016) Partially linear additive quantile regression in ultra-high dimension. *Ann Stat* 44:288–317
- Stone C (1985) Additive regression and other nonparametric models. *Ann Stat* 13:689–705
- Tang Q, Cheng L (2009) B-spline estimation for varying coefficient regression with spatial data. *Sci Chin Ser A* 52(11):2321–2340
- Tran LT (1990) Kernel density estimation on random field. *J Multivar Anal* 34:37–53
- Yang J, Yang H, Lu F (2019) Rank-based shrinkage estimation for identification in semiparametric additive models. *Stat Pap* 60:1255–1281
- You J, Zhou X (2013) Efficient estimation in panel data partially additive linear model with serially correlated errors. *Stat Sin* 23:271–303
- Yu S, Wang G, Wang L, Liu C, Yang L (2020) Estimation and Inference for Generalized Geoadditive Models. *J Am Stat Assoc* 115:761–774

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.