

Statistical Inference on Semiparametric Spatial Additive Model

Chuanhua Wei¹, Ran Yan¹, Shuang Guo¹ & Tao Tao¹

¹ School of Science, Minzu University of China, Beijing 100081, P.R.China

Correspondence: Chuanhua Wei, School of Science, Minzu University of China, Beijing 100081, P.R.China.
 E-mail: chweisd@163.com

Received: January 15, 2020 Accepted: February 12, 2020 Online Published: February 19, 2020

doi:10.5539/jmr.v12n2p1 URL: <https://doi.org/10.5539/jmr.v12n2p1>

Abstract

There has been a growing interest on using nonparametric and semiparametric modelling techniques for the analysis of spatial data because of their powerfulness in extracting the underlying local patterns in the data. In this study, stimulated by the Boston house price data, we apply a semiparametric spatial additive model to incorporation of spatial effects in regression models. For this semiparametric model, we develop a linear hypothesis test of parametric coefficients as well as a test for the existence of the spatial effects. For the problem of variable selection, the adaptive Lasso method was applied. Monte Carlo simulation studies are conducted to illustrate the finite sample performance of the proposed inference procedures. Finally, an application in Boston housing data is studied.

Keywords: spatial additive model, profile least-squares approach, backfitting method, adaptive lasso

1. Introduction

The Boston house price data of Rubinfeld (1978) and Gilley and Pace (1996) is frequently used in literature to illustrate some new statistical methods. The data set consists of the median value of owner-occupied homes in 506 census tracts in the Boston Standard Metropolitan Statistical Area in 1970, and 13 accompanying sociodemographic and related variables, and is available through the R package **Sedep**.

For the Boston house price data, to capture the “large-scale” locational effects between response variable and the associated 13 covariates, Pace and Gilley (1997) proposed the following linear regression model

$$y_i = \sum_{j=1}^{13} \beta_j x_{ij} + \beta_{14} u_i v_i + \beta_{15} u_i + \beta_{16} v_i + \beta_{17} u_i^2 + \beta_{18} v_i^2 + \varepsilon_i, i = 1, 2, \dots, 506, \quad (1)$$

where (u_i, v_i) is the latitude(LAT) and longitude(LON) of the i th observation. However, sometimes, the quadratic expression involving latitude and longitude is not adequate for the real locational effects. To solve this problem, we apply the following semiparametric spatial additive model to fit the data set

$$y_i = f(u_i) + g(v_i) + \mathbf{x}_i^T \boldsymbol{\beta} + \varepsilon_i, \quad i = 1, 2, \dots, n, \quad (2)$$

where y_i and $\mathbf{x}_i = (x_{i1}, x_{i2}, \dots, x_{ip})^T$ are observations of the response and associated explanatory variables at location (u_i, v_i) in the studied geographical region, $\boldsymbol{\beta} = (\beta_1, \beta_2, \dots, \beta_p)^T$ is a vector of p -dimensional unknown parameters, and ε_i is model error with mean zero and variance σ^2 . $f(\cdot)$ and $g(\cdot)$ are unknown smooth functions of latitude and longitude respectively. For the identifiability of nonparametric functions, we assume that $E[f(u)] = 0$ and $E[g(v)] = 0$. Without loss of generality, we also assume that both the y_i and \mathbf{x}_i have been centered about their respective means. Model (2) can be regarded as an extension of the following semiparametric spatial model of McMillen (2012),

$$y_i = m(u_i, v_i) + \mathbf{x}_i^T \boldsymbol{\beta} + \varepsilon_i, \quad i = 1, 2, \dots, n.$$

where $m(u, v)$ is an unknown nonparametric function on the latitude and longitude. Compared with spatial autoregressive models, spatial error model and other standard spatial econometric models, model (2) does not require use of an $n \times n$ spatial weight matrix \mathbf{W} , and has a greater flexibility to deal with spatial effects.

It is remarked that model (2) is a special partially linear additive models in statistical literature, which have been studied by Li (2000), Jiang *et al.* (2007), Liang *et al.* (2008), Liu *et al.* (2011), Wei and Liu (2012), more details can be found in Li and Racine (2007). In this paper, we focus on the statistical inference of model (2), including testing and variable selection.

This rest of this paper is organized as follows. We introduce the profile least-square estimation approach in Section 2. Testing for the parametric and nonparametric components is discussed in Section 3. Section 4 studies the variable selection procedure. Simulations are studied in Section 5 to illustrate the performance of the proposed approaches. As an application example, the Boston house price data are analyzed by the proposed methods in Section 6. Discussion is given in Section 7.

2. Profile Least-Squares Estimation

As the basis for the problem of testing and variable selection of model (2), estimation of model (2) is first briefly described in this section. Many methods can be applied to estimate model (2), we apply the profile least-square approach of Liang *et al.* (2008) and Wei and Liu (2012) as its simplicity.

Let

$$\mathbf{Y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \mathbf{X} = \begin{bmatrix} \mathbf{x}_1^T \\ \mathbf{x}_2^T \\ \vdots \\ \mathbf{x}_n^T \end{bmatrix}, \mathbf{f} = \begin{bmatrix} f(u_1) \\ f(u_2) \\ \vdots \\ f(u_n) \end{bmatrix}, \mathbf{g} = \begin{bmatrix} g(v_1) \\ g(v_2) \\ \vdots \\ g(v_n) \end{bmatrix}, \boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix},$$

then, model (2) can be written as

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{f} + \mathbf{g} + \boldsymbol{\varepsilon}. \tag{3}$$

Suppose the parametric components $\boldsymbol{\beta}$ and the nonparametric function \mathbf{g} are known, then model (2) becomes the following standard nonparametric regression model

$$y_i^* = f(u_i) + \varepsilon_i, \quad i = 1, 2, \dots, n, \tag{4}$$

with $y_i^* = y_i - \mathbf{x}_i^T \boldsymbol{\beta} - g(v_i)$. We apply the local linear approach to estimate the nonparametric function $f(\cdot)$. Assume that $f(u_i)$ has a continuous second derivative for any fixed u , then by Taylor expansion, we have

$$f(u_i) \approx f(u) + f'(u)(u_i - u), \quad i = 1, 2, \dots, n.$$

Then, we can estimate $f(u)$ and $f'(u)$ by the following weighted local least-squares problems:

$$\sum_{i=1}^n \{ [y_i - \mathbf{X}_i^T \boldsymbol{\beta} - g(v_i)] - [f(u) + f'(u)(u_i - u)] \}^2 K_{h_1}(u_i - u), \tag{5}$$

with kernel function $K_{h_1}(\cdot) = K(\cdot/h_1)/h_1$, and bandwidth h_1 . By solving the problem (5), we obtain the initial estimator of $f(u)$ as

$$\hat{f}(u) = \mathbf{e}_1^T \{ \mathbf{D}_u^T \mathbf{K}_u \mathbf{D}_u \}^{-1} \mathbf{D}_u^T \mathbf{K}_u (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{g}), \tag{6}$$

with $\mathbf{e}_1 = (1, 0)^T$, $\mathbf{K}_u = \text{diag}\{K_{h_1}(u_1 - u), K_{h_1}(u_2 - u), \dots, K_{h_1}(u_n - u)\}$, and

$$\mathbf{D}_u = \begin{bmatrix} 1 & u_1 - u \\ 1 & u_2 - u \\ \vdots & \vdots \\ 1 & u_n - u \end{bmatrix}, \mathbf{S}_u = \begin{bmatrix} \mathbf{e}_1^T \{ \mathbf{D}_{u_1}^T \mathbf{K}_{u_1} \mathbf{D}_{u_1} \}^{-1} \mathbf{D}_{u_1}^T \mathbf{K}_{u_1} \\ \mathbf{e}_1^T \{ \mathbf{D}_{u_2}^T \mathbf{K}_{u_2} \mathbf{D}_{u_2} \}^{-1} \mathbf{D}_{u_2}^T \mathbf{K}_{u_2} \\ \vdots \\ \mathbf{e}_1^T \{ \mathbf{D}_{u_n}^T \mathbf{K}_{u_n} \mathbf{D}_{u_n} \}^{-1} \mathbf{D}_{u_n}^T \mathbf{K}_{u_n} \end{bmatrix}.$$

We take u to be each of u_1, u_2, \dots, u_n , together with the condition $\sum_{i=1}^n f(u_i) = 0$, we can obtain the estimator of \mathbf{f} as

$$\hat{\mathbf{f}} = [\hat{f}(u_1), \hat{f}(u_2), \dots, \hat{f}(u_n)]^T = \mathbf{S}_u^* (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{g}). \tag{7}$$

with $\mathbf{S}_u^* = (\mathbf{I}_n - \mathbf{1}\mathbf{1}^T) \mathbf{S}_u$,

Similarly, the estimator of \mathbf{g} is

$$\hat{\mathbf{g}} = [\hat{g}(v_1), \hat{g}(v_2), \dots, \hat{g}(v_n)]^T = \mathbf{S}_v^* (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{f}), \tag{8}$$

with $\mathbf{S}_v^* = (\mathbf{I}_n - \mathbf{1}\mathbf{1}^T) \mathbf{S}_v$,

$$\mathbf{S}_v = \begin{bmatrix} \mathbf{e}_1^T \{ \mathbf{D}_{v_1}^T \mathbf{K}_{v_1} \mathbf{D}_{v_1} \}^{-1} \mathbf{D}_{v_1}^T \mathbf{K}_{v_1} \\ \mathbf{e}_1^T \{ \mathbf{D}_{v_2}^T \mathbf{K}_{v_2} \mathbf{D}_{v_2} \}^{-1} \mathbf{D}_{v_2}^T \mathbf{K}_{v_2} \\ \vdots \\ \mathbf{e}_1^T \{ \mathbf{D}_{v_n}^T \mathbf{K}_{v_n} \mathbf{D}_{v_n} \}^{-1} \mathbf{D}_{v_n}^T \mathbf{K}_{v_n} \end{bmatrix}, \mathbf{D}_v = \begin{bmatrix} 1 & v_1 - v \\ 1 & v_2 - v \\ \vdots & \vdots \\ 1 & v_n - v \end{bmatrix}.$$

Combining (7) and (8), we have the following backfitting estimating equation system

$$\begin{bmatrix} \mathbf{I}_n & \mathbf{S}_1^* \\ \mathbf{S}_2^* & \mathbf{I}_n \end{bmatrix} \begin{bmatrix} \mathbf{f} \\ \mathbf{g} \end{bmatrix} = \begin{bmatrix} \mathbf{S}_1^* \\ \mathbf{S}_2^* \end{bmatrix} (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}). \tag{9}$$

By Hastie and Tibshirani (1990) and Opsomer and Ruppert (1997), the estimators of \mathbf{f} and \mathbf{g} can be obtained as

$$\hat{\mathbf{f}} = \mathbf{W}_1(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}), \quad \hat{\mathbf{g}} = \mathbf{W}_2(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}), \tag{10}$$

with $\mathbf{W}_1 = \mathbf{I}_n - (\mathbf{I}_n - \mathbf{S}_u^* \mathbf{S}_v^*)^{-1}(\mathbf{I}_n - \mathbf{S}_u^*)$, $\mathbf{W}_2 = \mathbf{I}_n - (\mathbf{I}_n - \mathbf{S}_v^* \mathbf{S}_u^*)^{-1}(\mathbf{I}_n - \mathbf{S}_v^*)$.

Replacing \mathbf{f} and \mathbf{g} of (3) by their estimators $\hat{\mathbf{f}}$ and $\hat{\mathbf{g}}$, respectively, we can obtain a synthetic linear regression model

$$(\mathbf{I}_n - \mathbf{S})\mathbf{Y} = (\mathbf{I}_n - \mathbf{S})\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \tag{11}$$

with $\mathbf{S} = \mathbf{W}_1 + \mathbf{W}_2$.

The profile least squares estimation of $\boldsymbol{\beta}$ can be obtained by the least-squares approach and model (11),

$$\hat{\boldsymbol{\beta}} = [\mathbf{X}^T(\mathbf{I}_n - \mathbf{S})^T(\mathbf{I}_n - \mathbf{S})\mathbf{X}]^{-1} \mathbf{X}^T(\mathbf{I}_n - \mathbf{S})^T(\mathbf{I}_n - \mathbf{S})\mathbf{Y}, \tag{12}$$

Furthermore, we define the final estimators of \mathbf{f} and \mathbf{g} as

$$\hat{\mathbf{f}} = \mathbf{W}_1(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}), \quad \hat{\mathbf{g}} = \mathbf{W}_2(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}). \tag{13}$$

By (12) and (13), we have

$$\hat{\mathbf{Y}} = \mathbf{X}\hat{\boldsymbol{\beta}} + \hat{\mathbf{f}} + \hat{\mathbf{g}} = \mathbf{X}\hat{\boldsymbol{\beta}} + \mathbf{S}(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}) = \mathbf{L}\mathbf{Y} \quad \text{and} \quad \hat{\boldsymbol{\varepsilon}} = \mathbf{Y} - \hat{\mathbf{Y}} = (\mathbf{I}_n - \mathbf{L})\mathbf{Y}, \tag{14}$$

with $\hat{\mathbf{Y}} = (\hat{y}_1, \hat{y}_2, \dots, \hat{y}_n)^T$, and $\hat{\boldsymbol{\varepsilon}} = (\hat{\varepsilon}_1, \hat{\varepsilon}_2, \dots, \hat{\varepsilon}_n)^T$, $\mathbf{L} = \mathbf{S} + \bar{\mathbf{X}}[\bar{\mathbf{X}}^T\bar{\mathbf{X}}]^{-1}\bar{\mathbf{X}}^T(\mathbf{I}_n - \mathbf{S})$, and $\bar{\mathbf{X}} = (\mathbf{I}_n - \mathbf{S})\mathbf{X}$.

According to the result in Hastie and Tibshirani (1990, Section 3.4.3), we can use the Cross-Validation (CV) technique to choose bandwidth. The bandwidth is chosen to minimize the expression

$$CV(\mathbf{h}) = \sum_{i=1}^n \left(\frac{\hat{\varepsilon}_i}{1 - l_{ii}} \right)^2,$$

where $\mathbf{h} = (h_1, h_2)^T$, $\hat{\varepsilon}_i$ and l_{ii} ($i = 1, 2, \dots, n$) are respectively the residuals and the diagonal elements of the matrix \mathbf{L} which are shown in (14).

3. Hypothesis Tests

Based on the estimation method described in the previous section, we propose two hypothesis tests. The first one is for general hypothesis testing of regression parameters, and the second one is for testing the nonparametric functions.

3.1 Testing for the Parametric Components

we consider the following linear hypothesis

$$H_0 : \mathbf{A}\boldsymbol{\beta} = \mathbf{0} \quad \text{VS} \quad H_1 : \mathbf{A}\boldsymbol{\beta} \neq \mathbf{0}, \tag{15}$$

where \mathbf{A} is a $k \times p$ matrix of known constants and $\mathbf{0}$ is a k -vector of zero. We shall also assume that $\text{rank}(\mathbf{A}) = k$.

Different to the generalized likelihood ratio test of Wei and Liu (2012), we will develop a generalized F -test statistic for the testing problem (15). To construct the test statistic, we first estimate model (2) under the null hypothesis $\mathbf{A}\boldsymbol{\beta} = \mathbf{0}$. Applying the Lagrange multiplier technique to linear model (11), the restricted profile least-squares estimator of $\boldsymbol{\beta}$ is obtained as

$$\hat{\boldsymbol{\beta}}_r = \hat{\boldsymbol{\beta}} - (\bar{\mathbf{X}}^T\bar{\mathbf{X}})^{-1}\mathbf{A}^T \left[\mathbf{A}(\bar{\mathbf{X}}^T\bar{\mathbf{X}})^{-1}\mathbf{A}^T \right]^{-1} \mathbf{A}\hat{\boldsymbol{\beta}}. \tag{16}$$

Moreover, the restricted estimator of \mathbf{f} and \mathbf{g} can be defined as

$$\hat{\mathbf{f}}_r = \mathbf{W}_1(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}_r), \quad \hat{\mathbf{g}}_r = \mathbf{W}_2(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}_r), \tag{17}$$

Therefore, we can get the residual sum of squares under H_0 as

$$RSS(H_0) = (\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}_r - \hat{\mathbf{f}}_r - \hat{\mathbf{g}}_r)^T(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}_r - \hat{\mathbf{f}}_r - \hat{\mathbf{g}}_r). \tag{18}$$

On the other hand, based on the equation (14), the residual sum of squares under the alternative hypothesis is

$$RSS(H_1) = (\mathbf{Y} - \hat{\mathbf{Y}})^T(\mathbf{Y} - \hat{\mathbf{Y}}) = \mathbf{Y}^T(\mathbf{I} - \mathbf{L})^T(\mathbf{I} - \mathbf{L})\mathbf{Y} = \mathbf{Y}^T\mathbf{M}_1\mathbf{Y}, \tag{19}$$

where $\mathbf{M}_1 = (\mathbf{I} - \mathbf{L})^T(\mathbf{I} - \mathbf{L})$.

By simple calculation, we have

$$RSS(H_0) - RSS(H_1) = \mathbf{Y}^T\mathbf{M}_0\mathbf{Y}, \tag{20}$$

where $\mathbf{M}_0 = (\mathbf{I} - \mathbf{S})^T\bar{\mathbf{X}}[\bar{\mathbf{X}}^T\bar{\mathbf{X}}]^{-1}\mathbf{A}^T[\mathbf{A}(\bar{\mathbf{X}}^T\bar{\mathbf{X}})^{-1}\mathbf{A}^T]^{-1}\mathbf{A}[\bar{\mathbf{X}}^T\bar{\mathbf{X}}]^{-1}\bar{\mathbf{X}}^T(\mathbf{I} - \mathbf{S})$.

It is obviously that, if H_0 is true, there will be little difference between $RSS(H_0)$ and $RSS(H_1)$. Under the alternative hypothesis, there should be significant difference between $RSS(H_0)$ and $RSS(H_1)$, and $RSS(H_1)$ should become systematically larger than $RSS(H_0)$. Using this fact, we can construct a generalized F test statistic as

$$F_1 = \frac{[RSS(H_0) - RSS(H_1)]/\nu_1}{RSS(H_1)/\delta_1} = \frac{\mathbf{Y}^T\mathbf{M}_0\mathbf{Y}/\nu_1}{\mathbf{Y}^T\mathbf{M}_1\mathbf{Y}/\delta_1}. \tag{21}$$

where $\nu_1 = \text{tr}(\mathbf{M}_0)$, $\delta_1 = \text{tr}(\mathbf{M}_1)$.

It is obviously that \mathbf{M}_0 and \mathbf{M}_1 are not symmetric and idempotent matrix, so the statistical distribution of the test statistic F_1 is more complicated. However, By the result of Cleveland and Devlin (1988), if the model error is normal distribution, the null distribution of the test statistic F_1 can be approximated by an F - distribution with degrees of freedom (r_1, r_2) , where $r_1 = \nu_1^2/\nu_2, r_2 = \delta_1^2/\delta_2, \nu_2 = \text{tr}(\mathbf{M}_0^2), \delta_2 = \text{tr}(\mathbf{M}_1^2)$. The application of this approximate procedure can be found in Leung *et al.* (2000a, 2000b) and Fotheringham *et al.* (2002).

3.2 Testing for the Existence of Spatial Effects

For the semiparametric spatial additive model (2), an important question is to test the existence of spatial effects. This leads to the following testing problem

$$H_0 : f(u_i) = g(v_i) = 0, i = 1, 2, \dots, n.$$

If the null hypothesis H_0 is true, model (2.1) is turned as the following standard linear regression model

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}. \tag{22}$$

Then, the residual sum of squares under H_0 based on the least-squares approach is

$$RSS(H_0) = \mathbf{Y}^T[\mathbf{I}_n - \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T]\mathbf{Y} = \mathbf{Y}^T\mathbf{N}_0\mathbf{Y}, \tag{23}$$

with $\mathbf{N}_0 = \mathbf{I}_n - \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T$. On the other hand, by (19), the residual sum of squares under the alternative hypothesis H_1 is $\mathbf{Y}^T\mathbf{M}_1\mathbf{Y}$. Similar to the construction of F_1 , we define the following test statistic as

$$F_2 = \frac{[RSS(H_0) - RSS(H_1)]/\tau_1}{RSS(H_1)/\delta_1} = \frac{\mathbf{Y}^T(\mathbf{N}_0 - \mathbf{M}_1)\mathbf{Y}/\tau_1}{\mathbf{Y}^T\mathbf{M}_1\mathbf{Y}/\delta_1}. \tag{24}$$

where $\tau_1 = \text{tr}(\mathbf{N}_0 - \mathbf{M}_1)$.

Then, according to Cleveland and Devlin (1988), we know that the null distribution of test statistic F_2 can be approximated by an F - distribution with degrees of freedom (r_1^*, r_2) , where $r_1^* = \tau_1^2/\tau_2, r_2 = \delta_1^2/\delta_2, \tau_2 = \text{tr}[(\mathbf{N}_0 - \mathbf{M}_1)^2], \delta_2 = \text{tr}(\mathbf{M}_1^2)$.

4. Variable Selection by ALASSO Method

As well as other type regression models, an important problem in using the model (2) is variable selection. In practice, many variables can be introduced to the initial analysis. Deciding which covariates to be kept in the final statistical model has always been a tricky task for data analysis. For the standard linear regression model, stepwise regression method can be applied directly by statistical software. However, traditional variable selection methods suffer from several drawbacks, the most severe one of which is the lack of stability as pointed out by Breiman (1996). To solve this problem, some penalized methods for significant variable selection have been proposed in the last decades. Examples include the bridge regression of Frank and Friedman (1993), the least absolute shrinkage and selection operator (LASSO) of Tibshirani (1996), the smoothly clipped absolute deviation (SCAD) of Fan and Li (2001), the elastic net penalty of Zou and Hastie (2006) and the adaptive Lasso (ALASSO) of Zou (2006). For model (2), Liu *et al.* (2011) developed a variable selection procedure to identify significant linear components using the SCAD based on the polynomial splines method. In the following, we consider the problem of simultaneous variable selection and estimation in model (2) based on the ALASSO

method. Firstly, according to Tibshirani (1996), the Lasso estimator of the coefficient β in model (2) is obtained by minimizing the residual sum of squares:

$$\sum_{i=1}^n [y_i - \mathbf{x}_i^T \beta - f(u_i) - g(v_i)]^2 \tag{25}$$

with respect to β subject to the constraint $\sum_{j=1}^p \beta_j \leq s$, where s is a tuning parameter. Equivalently, the Lasso estimator of β can be defined as

$$\beta_{\text{Lasso}} = \arg \min_{\beta} Q(\beta) = \sum_{i=1}^n [y_i - \mathbf{x}_i^T \beta - f(u_i) - g(v_i)]^2 + \lambda \sum_{j=1}^p |\beta_j|, \tag{26}$$

where λ is a tuning parameter depending on s . The purpose of penalty $\lambda \sum_{j=1}^p |\beta_j|$ is to shrink some of the coefficients to exactly zero. This makes the LASSO a simultaneous estimation and variable selection procedure. To eliminate the unknown coefficient functions $\hat{f}(\cdot)$ and $\hat{g}(\cdot)$ in $Q(\beta)$, we replace $f(u_i)$ and $g(v_i)$ in (26) by $\hat{f}(u_i)$ and $\hat{g}(v_i)$ which were defined in (2.7) respectively, then we can obtain

$$\beta_{\text{Lasso}} = \arg \min_{\beta} Q(\beta) = (\bar{Y} - \bar{X}\beta)^T (\bar{Y} - \bar{X}\beta) + \lambda \sum_{j=1}^p |\beta_j|. \tag{27}$$

As noted by Zou (2006), the Lasso estimator can not attain the oracle property. To solve this problem, following Zou (2006), the ALASSO estimator of β can be defined as

$$\hat{\beta}_{\text{ALasso}} = \arg \min_{\beta} \left\{ (\bar{Y} - \bar{X}\beta)^T (\bar{Y} - \bar{X}\beta) + \lambda \sum_{j=1}^p \frac{|\beta_j|}{|\tilde{\beta}_j|^\gamma} \right\} \tag{28}$$

where $\tilde{\beta} = (\tilde{\beta}_1, \tilde{\beta}_2, \dots, \tilde{\beta}_p)^T$ is a consistent estimator of β , $\gamma > 0$ is constant. One possible choice for $\tilde{\beta}$ is the profile least-squares estimator $\hat{\beta}$ in (12). The LARS algorithm based on R package can be used to obtain $\hat{\beta}_{\text{ALasso}}$.

5. Simulation Studies

In this section, some simulations are conducted to examine the performance of the proposed procedures.

5.1 Spatial Layout and Design of the Experiments

In spatial analysis, a regular lattice has quite a strong background in applications. In our simulations, the observations are collected from a uniform, two-dimensional grid consisting of $m \times m$ lattice points with unit distance between any two neighboring points along the horizontal and vertical axes. These m^2 points are arranged in an orthogonal coordinate system in such a way that the spatial coordinates (u_i, v_i) for observing the data are

$$(u_i, v_i) = \left(\text{mod}(i - 1, m), \left\lfloor \frac{i - 1}{m} \right\rfloor \right), i = 1, 2, \dots, m^2,$$

where $\text{mod}(i - 1, m)$ is the remainder of $i - 1$ divided by m and $\lfloor \frac{i-1}{m} \rfloor$ is the integer part of number $\frac{i-1}{m}$.

The data are generated from the following semiparametric spatial additive model

$$y_i = f(u_i) + g(v_i) + \beta_1 x_{i1} + \beta_2 x_{i2} + \varepsilon_i, \quad i = 1, 2, \dots, m^2,$$

with $x_{i1} \sim N(0, 1)$, $x_{i2} \sim U(-2, 2)$, the following three type models are considered. Model 1 and Model 2 are true spatial additive models, while the Model 3 is not a true additive structure.

- Model 1** : $f(u_i) = u_i, \quad g(v_i) = 1 + 2v_i;$
- Model 2** : $f(u_i) = \cos\left(\frac{2\pi u_i}{m-1}\right), \quad g(v_i) = 0;$
- Model 3** : $f(u_i) + g(v_i) = \frac{3}{m-1} \sqrt{\left(\frac{m-1}{2} - \left|\frac{m-1}{2} - u_i\right|\right) \left(\frac{m-1}{2} - \left|\frac{m-1}{2} - v_i\right|\right)}.$

To gain an idea of the effect of the distribution of the error on our results, we take the following four different types of the error distribution whose scales are so adjusted that they all have a common variance $\sigma^2 = 0.25$, (1) $\varepsilon_i \sim N(0, 0.5^2)$, (2) $\varepsilon_i \sim U(-\sqrt{3}/2, \sqrt{3}/2)$, (3) $\varepsilon_i \sim \frac{1}{8}\chi_8^2 - 1$, (4) $\varepsilon_i \sim \sqrt{3}/4t(8)$. The Epanechnikov kernel $K(x) = 0.75(1 - x^2)\mathbf{I}_{|x| \leq 1}$ is used in our simulation. Furthermore, we use the CV technique of Section 2 to choose the bandwidths.

5.2 The Finite Sample Performance of the Profile Least-Squares Estimation

Let $\beta_1 = 1, \beta_2 = 2$, sample size $m = 8, 10, 12$, for each setting, simulation were repeated $N = 1000$ times. By the profile least-squares approach in Section 2, we can obtain the estimators of β_1 and β_2 . we asses the performance of the estimators based on sample mean (Mean), sample standard deviation (SD) and sample mean squared error (MSE). The simulation results are presented in Table 1-3. We can find that all the estimators of the parametric components are close to the true value. As the sample size increases, the biases, sample standard deviation and sample mean squared error of estimators decrease.

Table 1. Finite Sample Performance of the Profile Least-squares Estimator of Model 1

Error	m	$\hat{\beta}_1$			$\hat{\beta}_2$		
		Mean	SD	MSE	Mean	SD	MSE
$N(0, 0.5^2)$	8	1.0014	0.0645	0.0042	2.0008	0.0554	0.0031
	10	0.9990	0.0498	0.0025	2.0028	0.0450	0.0020
	15	0.9997	0.0335	0.0011	2.0003	0.0289	0.0008
$U(-\frac{\sqrt{3}}{2}, \frac{\sqrt{3}}{2})$	8	1.0003	0.0663	0.0044	1.9992	0.0576	0.0033
	10	1.0018	0.0514	0.0026	1.9985	0.0439	0.0019
	15	1.0001	0.0333	0.0011	2.0012	0.0293	0.0009
$\frac{1}{8}\chi_8^2 - 1$	8	0.9949	0.0678	0.0046	2.0005	0.0557	0.0031
	10	0.9992	0.0514	0.0026	2.0017	0.0429	0.0018
	15	0.9994	0.0355	0.0013	2.0018	0.0297	0.0009
$\frac{\sqrt{3}}{4}t(8)$	8	0.9987	0.0657	0.0043	1.9963	0.0567	0.0032
	10	0.9975	0.0511	0.0026	1.9990	0.0445	0.0020
	15	0.9999	0.0342	0.0012	2.0002	0.0284	0.0008

Table 2. Finite Sample Performance of the Profile Least-squares Estimator of Model 2

Error	m	$\hat{\beta}_1$			$\hat{\beta}_2$		
		Mean	SD	MSE	Mean	SD	MSE
$N(0, 0.5^2)$	8	0.9997	0.0702	0.0049	1.9973	0.0583	0.0034
	10	1.0012	0.0547	0.0030	1.9982	0.0457	0.0021
	15	1.0018	0.0351	0.0012	2.0000	0.0298	0.0009
$U(-\frac{\sqrt{3}}{2}, \frac{\sqrt{3}}{2})$	8	0.9999	0.0685	0.0047	2.0010	0.0591	0.0035
	10	1.0034	0.0531	0.0028	2.0008	0.0462	0.0021
	15	1.0023	0.0351	0.0012	1.9994	0.0308	0.0010
$\frac{1}{8}\chi_8^2 - 1$	8	1.0026	0.0687	0.0047	2.0014	0.0616	0.0038
	10	0.9996	0.0530	0.0028	2.0007	0.0461	0.0021
	15	1.0001	0.0362	0.0013	2.0022	0.0304	0.0009
$\frac{\sqrt{3}}{4}t(8)$	8	1.0012	0.0697	0.0049	2.0008	0.0593	0.0035
	10	1.0003	0.0543	0.0029	1.9993	0.0472	0.0022
	15	0.9993	0.0342	0.0012	2.0002	0.0297	0.0009

Table 3. Finite Sample Performance of the Profile Least-squares Estimator of Model 3

Error	m	$\hat{\beta}_1$			$\hat{\beta}_2$		
		Mean	SD	MSE	Mean	SD	MSE
$N(0, 0.5^2)$	8	1.0016	0.0753	0.0057	2.0001	0.0643	0.0041
	10	0.9971	0.0575	0.0033	2.0029	0.0474	0.0022
	15	1.0020	0.0347	0.0012	1.9996	0.0310	0.0010
$U(-\frac{\sqrt{3}}{2}, \frac{\sqrt{3}}{2})$	8	1.0045	0.0740	0.0055	2.0009	0.0619	0.0038
	10	0.9994	0.0563	0.0032	2.0017	0.0473	0.0022
	15	1.0038	0.0353	0.0013	2.0004	0.0309	0.0010
$\frac{1}{8}\chi_8^2 - 1$	8	0.9950	0.0740	0.0055	1.9995	0.0616	0.0038
	10	0.9964	0.0545	0.0030	2.0040	0.0465	0.0022
	15	1.0002	0.0367	0.0013	2.0014	0.0319	0.0010
$\frac{\sqrt{3}}{4}t(8)$	8	1.0028	0.0719	0.0052	1.9998	0.0635	0.0040
	10	0.9962	0.0574	0.0033	1.9990	0.0493	0.0024
	15	0.9989	0.0369	0.0014	2.0019	0.0316	0.0010

5.3 Performance of the Proposed Test Statistic F_1

To study the finite sample performance of the test statistic F_1 in Section 3.1, for models 1-3, we consider the following testing problem

$$H_0 : \beta_1 - \beta_2 = 0 \quad \text{VS} \quad H_1 : \beta_1 - \beta_2 = c.$$

where $c = 0, \pm 0.2, \pm 0.4, \pm 0.6, \pm 0.8$ respectively, and $\beta_1 = 1, \beta_2 = 1 - c$. For each given value of c and each type of the error distribution, 1000 replications with $m = 10$ were run and the rejection rate under the significance level $\alpha = 0.05$ was computed as the simulated power of our proposed test procedure. The results are shown in Table 4-6. We can see that the rejection frequencies (estimated sizes) under H_0 are all reasonably close to the corresponding significance level 0.05. Under the alternative hypothesis, the rejection rate seems very robust to the variation of the type of error distribution, and increases rapidly as the alternative hypothesis deviates from the null hypothesis.

Table 4. The Rejection frequencies of the test statistic F_1 of Model 1

c Value	Error		Distribution	
	$N(0, 0.5^2)$	$U(-\sqrt{3}/2, \sqrt{3}/2)$	$\frac{1}{8}\chi_8^2 - 1$	$\frac{\sqrt{3}}{4}t(8)$
0.0	0.059	0.034	0.059	0.063
-0.2	0.291	0.331	0.319	0.323
0.2	0.340	0.311	0.312	0.329
-0.4	0.839	0.862	0.831	0.807
0.4	0.832	0.816	0.831	0.829
-0.6	0.993	0.991	0.994	0.982
0.6	0.985	0.994	0.988	0.990
-0.8	0.999	1.000	0.999	1.000
0.8	1.000	1.000	0.998	1.000

Table 5. The Rejection frequencies of the test statistic F_1 of Model 2

c Value	Error		Distribution	
	$N(0, 0.5^2)$	$U(-\sqrt{3}/2, \sqrt{3}/2)$	$\frac{1}{8}\chi_8^2 - 1$	$\frac{\sqrt{3}}{4}t(8)$
0.0	0.052	0.039	0.039	0.049
-0.2	0.297	0.314	0.259	0.304
0.2	0.275	0.286	0.285	0.290
-0.4	0.774	0.802	0.775	0.803
0.4	0.797	0.811	0.792	0.791
-0.6	0.985	0.993	0.982	0.985
0.6	0.982	0.993	0.988	0.984
-0.8	0.999	1.000	1.000	0.999
0.8	1.000	1.000	0.998	0.999

Table 6. The Rejection frequencies of the test statistic F_1 of Model 3

c Value	Error		Distribution	
	$N(0, 0.5^2)$	$U(-\sqrt{3}/2, \sqrt{3}/2)$	$\frac{1}{8}\chi_8^2 - 1$	$\frac{\sqrt{3}}{4}t(8)$
0.0	0.044	0.055	0.057	0.052
-0.2	0.287	0.290	0.267	0.292
0.2	0.279	0.272	0.270	0.288
-0.4	0.764	0.782	0.772	0.780
0.4	0.775	0.769	0.781	0.780
-0.6	0.983	0.981	0.979	0.973
0.6	0.982	0.976	0.978	0.975
-0.8	0.999	1.000	1.000	1.000
0.8	0.998	1.000	0.999	0.998

5.4 Performance of the Proposed Test Statistic F_2

To study the finite sample performance of the test statistic F_2 in Section 3.2, we consider the testing problem $H_0 : f(u_i) = g(v_i) = 0$ for the following models

- Model 4** : $f(u_i) = cu_i, \quad g(v_i) = dv_i;$
- Model 5** : $f(u_i) = c \cos\left(\frac{2\pi u_i}{m-1}\right), \quad g(v_i) = 0;$
- Model 6** : $f(u_i) + g(v_i) = c \frac{3}{m-1} \sqrt{\left(\frac{m-1}{2} - \left|\frac{m-1}{2} - u_i\right|\right)\left(\frac{m-1}{2} - \left|\frac{m-1}{2} - v_i\right|\right)}.$

Take $(c, d) = (0, 0), (0, 0.1), (-0.1, 0), (0.1, 0.1), (0, -0.2), (0.2, 0), (0.2, 0.2)$ in model 4, and $c = 0, \pm 0.1, \pm 0.2, \pm 0.3, \pm 0.5, \pm 0.6$ in model 5 and model 6. In addition, 1000 realizations were generated with $m = 10$ to calculate the size and power of F_2 , results are reported in Table 4-6. We can see that the rejection frequencies (estimated sizes) under H_0 are all close to the nominal level 0.05. On the other hand, the rejection rate increases rapidly as the alternative hypothesis deviates from the null hypothesis.

Table 7. The Rejection frequencies of the test statistic F_2 of Model 4

(c,d)	Error		Distribution	
Value	$N(0, 0.5^2)$	$U(-\sqrt{3}/2, \sqrt{3}/2)$	$\frac{1}{8}\chi_8^2 - 1$	$\frac{\sqrt{3}}{4}t(8)$
(0,0)	0.113	0.085	0.087	0.103
(0,0.1)	0.999	0.999	1.000	0.997
(-0.1,0)	0.998	1.000	1.000	0.999
(0.1,0.1)	1.000	1.000	1.000	1.000
(0,-0.2)	1.000	1.000	1.000	1.000
(0.2,0)	1.000	1.000	1.000	1.000
(0.2,0.2)	1.000	1.000	1.000	1.000

Table 8. The Rejection frequencies of the test statistic F_2 of Model 5

c	Error		Distribution	
Value	$N(0, 0.5^2)$	$U(-\sqrt{3}/2, \sqrt{3}/2)$	$\frac{1}{8}\chi_8^2 - 1$	$\frac{\sqrt{3}}{4}t(8)$
0	0.094	0.082	0.091	0.113
0.1	0.211	0.204	0.201	0.193
-0.1	0.216	0.201	0.199	0.212
0.2	0.581	0.577	0.586	0.577
-0.2	0.591	0.563	0.584	0.604
0.3	0.909	0.933	0.922	0.925
-0.3	0.915	0.929	0.907	0.914
0.5	0.999	1.000	1.000	1.000
-0.5	1.000	1.000	1.000	0.999

Table 9. The Rejection frequencies of the test statistic F_2 of Model 6

c	Error		Distribution	
Value	$N(0, 0.5^2)$	$U(-\sqrt{3}/2, \sqrt{3}/2)$	$\frac{1}{8}\chi_8^2 - 1$	$\frac{\sqrt{3}}{4}t(8)$
0	0.094	0.118	0.095	0.108
0.1	0.137	0.136	0.131	0.134
-0.1	0.123	0.117	0.117	0.118
0.2	0.238	0.196	0.238	0.271
-0.2	0.234	0.238	0.224	0.240
0.3	0.464	0.436	0.441	0.431
-0.3	0.460	0.440	0.447	0.452
0.5	0.889	0.878	0.871	0.888
-0.5	0.882	0.872	0.872	0.857
0.6	0.967	0.952	0.970	0.968
-0.6	0.966	0.962	0.965	0.968

5.5 Finite Sample Performance of Variable Selection Procedure

In this section, we conduct the following numeric simulation to illustrate the proposed variable selection procedures. The data are generated from the following semiparametric spatial additive regression model

$$y_i = f(u_i) + g(v_i) + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_8 x_{i8} + \varepsilon_i, i = 1, 2, \dots, m^2,$$

where $x_{ik} \sim N(0, 1), k = 1, 2, \dots, 8, \varepsilon_i \sim N(0, 0.25)$. Sample size $n = 8^2, 10^2$ and 15^2 , respectively. Let $(\beta_1, \beta_2, \dots, \beta_8) = (2, 1, 0, 0, 0, 0.5, 0, 0)$. Two cases for the spatial nonparametric functions are designed as

Model 7 : $f(u_i) = u_i, \quad g(v_i) = 1 + 2v_i;$

Model 8 : $f(u_i) + g(v_i) = \frac{3}{m-1} \sqrt{\left(\frac{m-1}{2} - \left|\frac{m-1}{2} - u_i\right|\right)\left(\frac{m-1}{2} - \left|\frac{m-1}{2} - v_i\right|\right)}, g(v_i) = 0.$

Model 7 is a true spatial additive model, the other model is not.

To assess the accuracy of the proposed estimators of the parametric components $\beta = (\beta_1, \beta_2, \dots, \beta_8)^T$, a criterion for measuring the “goodness” of an estimator is needed. For this purpose, the mean squared errors (MSE) criterion, defined as $MSE = E\|\hat{\beta} - \beta\|^2$ is used. For comparison with the ALASSO, we also evaluate the MSE of the LASSO, ORACLE and FULL model estimators. The ORACLE estimator is a profile least-squares estimator defined in (12) based on the true model that contains none of the covariates with zero coefficients, while the FULL model estimator is a profile least-squares estimator defined in (12) based on the full model that contains all of the covariates. The ORACLE estimator is expected to perform best since it is based on the true model which is unknown in practice, and thus serves as a benchmark for comparisons. All our simulations are based on 1000 replications. Table 10 reports the values of MSE of the various estimators.

From Table 10, we can see that, Both Lasso and Adaptive Lasso procedures significantly improve the MSE over the full model for both the model (7) and model (8). As expected, the ORACLE estimator performs best among all estimators. Finally, the increase of sample size certainly improves the estimation accuracy of all estimating procedures.

Table 11 presents the average number of “correct” and “incorrect” zero estimates for the ALASSO and LASSO based on 1000 replications. In Table 11, the columns labelled with “C” give the average number of the five zero coefficients correctly set to 0, the columns labelled with “I” give the average number of the three nonzero coefficients incorrectly set to 0. We observe from the table that in all cases the ALASSO provide more accurate number of correct zeros than does the LASSO.

Table 10. MSE of estimators for β

Method	Model 7			Model 8		
	$n = 8^2$	$n = 10^2$	$n = 15^2$	$n = 8^2$	$n = 10^2$	$n = 15^2$
ALASSO	0.0184	0.0112	0.0040	0.0224	0.0108	0.0048
LASSO	0.0320	0.0192	0.0080	0.0384	0.0224	0.0088
ORACLE	0.0135	0.0081	0.0036	0.0162	0.0096	0.0039
FULL	0.0397	0.0240	0.0096	0.0464	0.0272	0.0104

Table 11. Average numbers of correct(C)and incorrect(I) zeros

		Model 7			Model 8		
		$n = 8^2$	$n = 10^2$	$n = 15^2$	$n = 8^2$	$n = 10^2$	$n = 15^2$
ALASSO	C	4.152	4.261	4.491	4.171	4.283	4.478
	I	0	0	0	0	0	0
LASSO	C	2.354	2.294	2.38	2.348	2.338	2.369
	I	0	0	0	0	0	0

6. Real Data Analysis

We now illustrate the proposed testing procedure by analyzing the well-known Boston housing prices data. The dependent variable is the log of the median value (in 1000 USD) of the owner-occupied homes in each of the census tracts. Thirteen explanatory variables include levels of nitrogen oxides (NOX), average number of rooms (RM), proportion of structures built before 1940 (AGE), black population proportion (B), lower status population proportion (LSTAT), crime rate (CRIM), proportion of area zoned with large lots (ZN), proportion of nonretail business areas (INDUS), property tax

rate (TAX), pupilteacher ratio (PTRATIO), location contiguous to the Charles River (CHAS), weighted distances to the employment centers (DIS), and an index of accessibility (RAD).

To modelling the spatial effects more flexible, we use the following apply the semiparametric spatial additive model

$$y_i = f(u_i) + g(v_i) + \sum_{j=1}^{13} \beta_j x_{ij} + \beta_{14} u_i v_i + \varepsilon_i, i = 1, 2, \dots, 506, \tag{29}$$

We assume that both the response variable and covariates have been centered about their respective means. Applying the profile least-squares approach of Section 2, both the ALASSO and LASSO procedures of Section 4 to estimate the coefficients $\beta_j, j = 1, 2, \dots, 14$, results are presented in Table 12.

From Table 12, we can see that higher CRIM, NOX², DIS, TAX, PTRATIO, and LSTAT are all expected to lead to lower housing prices. On the other hand, larger ZN, RM², RAD and B are all expected to lead to higher housing prices. Overall, these estimates are consistent with our expectations. ZN is found to be insignificant in determining housing prices by the results of ALASSO.

Table 12. Estimates of the coefficients

	PLS	ALASSO	LASSO
CRIM	-0.0107	-0.0108	-0.0106
ZN	0.0005	0.0000	0.0005
INDUS	0.0015	-0.0009	0.0013
CHAS	0.0677	0.0715	0.0670
NOX ²	-0.58732	-0.6170	-0.5814
RM ²	0.0072	0.0071	0.0072
AGE	-0.0008	-0.0003	-0.0008
DIS	-0.08243	-0.0873	-0.0819
RAD	0.0161	0.01146	0.0158
TAX	-0.0007	-0.0339	-0.0007
PTRATIO	-0.0326	0.0002	-0.0324
B	0.0004	0.0003	0.0004
LSTAT	-0.0279	-0.0290	-0.0279
LAT*LON	-0.0337	-0.02843	-0.0317

7. Discussion

In this article, we propose a new semiparametric spatial additive model to incorporate spatial effects into regression models. The spatial effects of latitude and longitude were modelled as the nonparametric smoothing functions. However, their interaction effect was modelled as parametric structure. How to analyse the interaction effect by the nonparametric or semiparametric method is an important problem. On the other hand, our study was focused only on spatial cross sectional data. A natural extension is the following semiparameric spatio-temporal additive model by incorporating temporal effects into the model (2),

$$y_i = f(u_i) + g(v_i) + m(t_i) + \mathbf{x}_i^T \boldsymbol{\beta} + \varepsilon_i, \quad i = 1, 2, \dots, n,$$

where $m(\cdot)$ is the unknown smooth functions of time. The estimation, testing and variable selection procedures of this paper can be applied to this model, this is an important area of future research.

Conflicts of Interest

The authors declare no conflict of interest.

References

Breiman, L. (1996). Heuristics of Instability and Stabilization in Model Selection. *The Annals of Statistics*, 24, 2350-2383. <https://doi.org/10.1214/aos/1032181158>

Cleveland, W. S., & Devlin, S. J. (1988). Locally Weighted Regression Analysis by Local Fitting. *Journal of the American Statistical Association*, 83, 596-610. <https://doi.org/10.1080/01621459.1988.10478639>

Fan, J., & Hardle, W., & Mammen, E. (1998). Direct Estimation of Additive and Linear Components for High-Dimensional Data. *The Annals of Statistics*, 26(3), 943-71. <https://doi.org/10.1214/aos/1024691083>.

- Fan, J., & Li, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical Association*, 9, 1348-1360. <https://doi.org/10.1198/016214501753382273>
- Fotheringham, A. S., & Brunson, C., & Charlton, M. (2002). *Geographically Weighted Regression the Analysis of Spatial Varying Relationships*. Wiley, West Sussex. <https://doi.org/10.1111/j.1538-4632.2003.tb01114.x>
- Frank, I. E., & Friedman, J. H. (1993). A statistical view of some chemometrics tools. *Technometrics*, 35, 109-135. <https://doi.org/10.1080/00401706.1993.10485033>
- Gilley, O. W., & Pace, R. K. (1996). On the Harrison and Rubinfeld data. *Journal of Environmental Economics and Management*, 31, 403-405. <https://doi.org/10.1006/jjeem.1996.0052>
- Harrison, D., & Rubinfeld, D. L. (1978). Hedonic housing prices and the demand for clean air. *Journal of Environmental Economics and Management*, 5, 81-102. [https://doi.org/10.1016/0095-0696\(78\)90006-2](https://doi.org/10.1016/0095-0696(78)90006-2)
- Hastie, T. J., & Tibshirani, R. (1990). *Generalized Additive Models*. New York: Chapman and Hall.
- Jiang, J., Zhou, H., Jiang, X., & Peng, J. (2007). Generalized likelihood ratio tests for the structures of semiparametric additive models. *The Canadian Journal of Statistics*, 35, 381-398. <https://doi.org/10.2307/20445263>
- Leung, Y., & Mei, C. L., & Zhang, W. X. (2000a). Statistical tests for spatial nonstationarity based on the geographically weighted regression model. *Environment and Planning A*, 32, 9-32 <https://doi.org/10.1068/a3162>
- Leung, Y., & Mei, C. L., & Zhang, W. X. (2000b) Testing for spatial autocorrelation among the residuals of the geographically weighted regression. *Environment and Planning A*, 32, 871-890. <https://doi.org/10.1068/a32117>
- Li, Q. (2000). Efficient estimation of additive partially linear models. *International Economic Review*, 41, 1073-1092. <https://doi.org/10.1111/1468-2354.00096>
- Li, Q., & Racine, J. (2007). *Nonparametric Econometrics: Theory and Practice*. Princeton University, Princeton and Oxford.
- Liang, H., & Thurston, S., & Ruppert, D., & Apanasovich, T. (2008). Additive partial linear models with measurement errors. *Biometrika*, 95, 667-678. <https://doi.org/10.1093/biomet/asn024>
- Liu, X., & Wang, L., & Liang, H. (2011). Variable selection and estimation for semiparametric additive partial linear models. *Statistica Sinica*, 21, 1225-1248. <https://doi.org/10.2307/24309561>
- McMillen, D. P. (2012). Perspectives on Spatial Econometrics: Linear Smoothing with Structured Models. *Journal of Regional Science*, 52(2), 192-209. <https://doi.org/10.1111/j.1467-9787.2011.00746.x>
- Opsomer, J. D., & Ruppert, D. (1999). Fitting a bivariate additive model by local polynomial regression. *Annals of Statistics*, 25, 186-211. <https://doi.org/10.1214/aos/1034276626>
- Pace, R. K., & Gilley, O. W. (1997). Using the spatial configuration of the data to improve estimation. *Journal of Real Estate Finance and Economics*, 14, 333-340. <https://doi.org/10.1023/A:1007762613901>
- Tibshirani, R. (1996). Regression shrinkage and selection via the LASSO. *Journal of the Royal Statistical Society Series B*, 58, 267-288. <https://doi.org/10.1111/j.2517-6161.1996.tb02080.x>
- Wei, C. H., & Liu, C. L. (2012). Statistical inference on semiparametric partially linear additive models. *Journal of Nonparametric Statistics*, 24, 809-823. <https://doi.org/10.1080/10485252.2012.716155>
- Zou, H. (2006). The adaptive lasso and its oracle properties. *Journal of the American Statistical Association*, 110, 1418-1429 <https://doi.org/10.1198/016214506000000735>
- Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society, Series B*, 67, 301-320. <https://doi.org/10.2307/3647580>

Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/4.0/>).