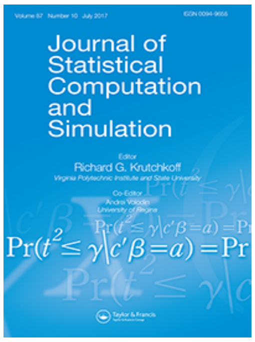

Journal of Statistical Computation and Simulation

Taylor & Francis
Taylor & Francis Group

ISSN: 0094-9655 (Print) 1563-5163 (Online) Journal homepage: https://www.tandfonline.com/loi/gscs20

Structure identification and variable selection in geographically weighted regression models

Wentao Wang & Dengkui Li

To cite this article: Wentao Wang & Dengkui Li (2017) Structure identification and variable selection in geographically weighted regression models, Journal of Statistical Computation and Simulation, 87:10, 2050-2068, DOI: 10.1080/00949655.2017.1311896

To link to this article: https://doi.org/10.1080/00949655.2017.1311896

JOURNAL OF STATISTICAL COMPUTATION AND SIMULATION, 2017
VOL. 87, NO. 10, 2050–2068
http://dx.doi.org/10.1080/00949655.2017.1311896

Taylor & Francis
Taylor & Francis Group

Structure identification and variable selection in geographically weighted regression models

Wentao Wang and Dengkui Li

Department of Statistics, School of Mathematics and Statistics, Xi'an Jiaotong University, Xi'an, People's Republic of China

ABSTRACT

Geographically weighted regression (GWR) is an important tool for exploring spatial non-stationarity of a regression relationship, in which whether a regression coefficient really varies over space is especially important in drawing valid conclusions on the spatial variation characteristics of the regression relationship. This paper proposes a so-called GWGlasso method for structure identification and variable selection in GWR models. This method penalizes the loss function of the local-linear estimation of the GWR model by the coefficients and their partial derivatives in the way of the adaptive group lasso and can simultaneously identify spatially varying coefficients, nonzero constant coefficients and zero coefficients. Simulation experiments are further conducted to assess the performance of the proposed method and the Dublin voter turnout data set is analysed to demonstrate its application.

ARTICLE HISTORY
Received 2 January 2017
Accepted 23 March 2017

KEYWORDS
Geographically weighted regression; spatial non-stationarity; adaptive group lasso; structure identification; variable selection

1. Introduction

Geographically weighted regression (GWR) [1,2] has been a very popular local modelling method that seeks to discover potential patterns of spatial non-stationarity in a regression relationship. In this methodology, the spatially varying coefficient model

$$y_i = \sum_{j=1}^{p} \beta_j(u_i, v_i)x_{ij} + \varepsilon_i, \quad i = 1, \dots, n, \tag{1}$$

is calibrated by the kernel smoothing technique, where y_i and $x_{i1}, x_{i2}, \dots, x_{ip}$ are respectively the observations of the response variable Y and those of the explanatory variables $X_1, X_2, \dots, X_p$ at the geographical location (u_i, v_i), ε_i is the random error term associated with (u_i, v_i) and $\beta_j(u, v)$ $(j = 1, \dots, p)$ are unknown regression coefficients to be estimated. Generally, $X_1 = 1$ is assumed to make the model include a spatially varying intercept. The model (1) with its calibration method is henceforth called GWR model as usual.

In the GWR literature, non-stationarity of the regression relationship is uncovered and explained by spatial variation patterns of the coefficient estimates. As a local modelling method, however, GWR always produces for each coefficient a set of local estimates which are generally different from one location to another no matter whether the coefficient is actually spatially varying or not. In some practical problems, however, it is possible that certain explanatory variables influencing the dependent variable are global in nature, whilst others are local [3], which leads to the consequence that

CONTACT Wentao Wang ✉ wtwang@hotmail.com

Supplemental data for this article can be accessed here http://dx.doi.org/10.1080/00949655.2017.1311896

some coefficients in the GWR model are constant and the others are spatially varying. Furthermore, it is also possible that constant coefficients are zero, indicating that the corresponding explanatory variables are irrelevant to the dependent variable. As pointed out by Wheeler [4], including some irrelevant explanatory variables in the model will degrade the estimation efficiency. In order to avoid a misleading explanation on non-stationarity of the underlying regression relationship, it is essential to know which coefficients in the GWR model really vary over the space and which coefficients are nonzero constant or zero, which belongs to structure identification and variable selection in a GWR model.

In fact, since the inception of GWR, much effort has been paid to the study of structure identification or model specification from the perspective of hypotheses testing. In the pioneering papers on GWR [1,5,6], a permutation test was proposed to check whether the coefficients in a GWR model are constant over the space. Furthermore, a residual sum of squares based test with the null distribution of the test statistics approximated by an F-distribution was also suggested by Brunsdon et al. [3] and Leung et al. [7] to check whether a GWR model or an ordinary linear model is appropriate for a given geo-referenced data set. However, it should be noted that the aforementioned kinds of tests focus on identifying global stationarity of the regression relationship. After introducing the mixed GWR model with its back-fitting calibration procedure, Brunsdon et al. [3] extended the residual sum of squares based test to identify constant coefficients in the GWR model. Furthermore, based on the same type of the test statistic and with the improved calibration method of a mixed GWR model in Fotheringham et al. [2] (Chapter 3), Mei et al. [8] proposed a bootstrap test to check whether some constant coefficients in a mixed GWR model are zero. Recently, considering some shortcomings of the residual sum of squares based test in the F-distribution approximation to the null distribution of the test statistic and normality assumption on the model error term, Mei et al. [9] suggested a bootstrap procedure to calculate the p-value of the test and the simulation study showed that the bootstrap method performs well. The above tests can be used to achieve the tasks of structure identification in a GWR model and variable selection in a mixed GWR model. In practice, however, there is generally not enough priori information for the analysts to know that which coefficients should be chosen to be tested for constants or zero. As a result, all possible combinations of the coefficients should be considered and a series of the tests should be performed, which is not an easy task especially when the number of the explanatory variables is large. Based on the sample variance of each estimated coefficient at all of the locations, Leung et al. [7] proposed an F-distribution to approximate the null distribution of the test statistic. Although this test can identify the constant coefficients in a GWR model one by one, as demonstrated in Mei et al. [9], it suffers from very high type I error. Therefore, it is worthwhile to develop some alternative methods for structure identification and variable selection in a GWR model in view of their importance in validly exploring spatial non-stationarity of the regression relationship.

The regularization or shrinkage methods developed in statistical learning have great potential to be used in regression models for model specification and variable selection. In the GWR literature, Wheeler [4] proposed a geographically weighted lasso method for simultaneous coefficient penalization and model selection mainly for alleviating the effect of collinearity among the explanatory variables on the coefficient estimates. This method performs local model selection by shrinking the coefficient estimates to zero at some locations and cannot be directly used for variable selection or constant coefficient identification. In the statistical literature on varying coefficient models, however, much attention has been paid to using the penalization methods for model structure identification and variable selection and many approaches with their own high spots have been developed [10–15]. Especially, based on the local-linear estimation and the smoothly clipped absolute deviation (SCAD, [16]) penalty, Ma and Zhang [17] recently proposed a new method for identifying varying coefficients, nonzero constant coefficients and zero coefficients in a varying coefficient model. In this method, the values of each coefficient and those of its derivative at all of the designed points are respectively grouped and taken as the penalty terms of the residual sum of squares of the local-linear estimation to make the coefficients and their derivatives shrink towards to zero. As a result, the varying, nonzero constant and zero coefficients can be simultaneously identified according to the penalized estimates of the coefficients.

Motivated by the methodology in Ma and Zhang [17] and based on the local-linear estimation of the GWR model [18] in which the estimates of coefficients and their partial derivatives can be obtained at any a focal location, we propose in this paper a so-called GWGlasso method to achieve the task of the structure identification and variable selection in GWR models. In the method, the values of each coefficient and those of its partial derivatives are respectively grouped and the adaptive group lasso method [19,20] is employed to penalize the residual sum of squares of the local-linear estimation of the GWR model. In this way, it is expected that, for the nonzero constant coefficients, their partial derivatives will shrink to zero, while for the zero coefficients, both the coefficients themselves and their partial derivatives will shrink to zero with which the nonzero constant and zero coefficients can be adaptively identified.

The remainder of this paper is organized as follows. In Section 2, the local-linear estimation method for GWR models is briefly described to facilitate the subsequent discussions. In Section 3, the GWGlasso methodology is introduced in detail and an iterative algorithm is formulated to compute the penalized estimates of the coefficients and their partial derivatives. A simulation experiment is conducted in Section 4 to assess the performance of the proposed method and a real-life data set is analysed in Section 5 to demonstrate its application. The paper is concluded with a brief summary and some possible extensions of the proposed method.

2. Local-linear estimation of the GWR model

As the basis of the forthcoming GWGlasso method, the local-linear estimation of the GWR model proposed by Wang et al. [18] is briefly described with the notations used in the current paper in order to facilitate the subsequent discussions.

Assume that the coefficients $\beta_j(u, v)$ $(j = 1, \dots, p)$ in model (1) have continuous partial derivatives with respect to u and v which we denote by $\beta_j^{(u)}(u, v)$ and $\beta_j^{(v)}(u, v)$, respectively. Given a designed location (u_k, v_k), let d_{ki} be the Euclidean distance between (u_k, v_k) and the ith observation location (u_i, v_i). According to Wang et al. [18], the local-linear estimates of the coefficients and their partial derivatives are the minimizer of the following objective function:

$$\mathcal{L}_h(u_k, v_k) = \sum_{i=1}^{n} \left\{ y_i - \sum_{j=1}^{p} [\beta_j(u_k, v_k) + \beta_j^{(u)}(u_k, v_k)(u_k - u_i) + \beta_j^{(v)}(u_k, v_k)(v_k - v_i)] x_{ij} \right\}^2 K_h(d_{ki}), \tag{2}$$

where $K_h(\cdot) = K(\cdot/h)$ with $K(\cdot)$ being a kernel function and h being a bandwidth. Let

$$\boldsymbol{X} = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{pmatrix}, \quad \boldsymbol{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}, \tag{3}$$

$$\begin{aligned} \boldsymbol{W}_h(u_k, v_k) &= \mathrm{Diag}[K_h(d_{k1}), K_h(d_{k2}), \dots, K_h(d_{kn})], \\ \boldsymbol{U}(u_k) &= \mathrm{Diag}[u_1 - u_k, u_2 - u_k, \dots, u_n - u_k], \\ \boldsymbol{V}(v_k) &= \mathrm{Diag}[v_1 - v_k, v_2 - v_k, \dots, v_n - v_k], \\ \boldsymbol{a}_r^{\mathrm{T}}(k) &= [\beta_1(u_k, v_k), \beta_2(u_k, v_k), \dots, \beta_p(u_k, v_k)], \\ \boldsymbol{b}_r^{\mathrm{T}}(k) &= [\beta_1^{(u)}(u_k, v_k), \beta_2^{(u)}(u_k, v_k), \dots, \beta_p^{(u)}(u_k, v_k)], \\ \boldsymbol{b}_r^{\mathrm{T}}(n+k) &= [\beta_1^{(v)}(u_k, v_k), \beta_2^{(v)}(u_k, v_k), \dots, \beta_p^{(v)}(u_k, v_k)]. \end{aligned} \tag{4}$$

The objective function in Equation (2) can be rewritten as

$$\begin{aligned}\mathcal{L}_h(u_k, v_k) &= (\boldsymbol{y} - \boldsymbol{X}\boldsymbol{a}_r(k) - \boldsymbol{U}(u_k)\boldsymbol{X}\boldsymbol{b}_r(k) - \boldsymbol{V}(v_k)\boldsymbol{X}\boldsymbol{b}_r(n+k))^{\mathrm{T}} \\ &\quad \cdot \boldsymbol{W}_h(u_k, v_k)(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{a}_r(k) - \boldsymbol{U}(u_k)\boldsymbol{X}\boldsymbol{b}_r(k) - \boldsymbol{V}(v_k)\boldsymbol{X}\boldsymbol{b}_r(n+k)). \qquad (5)\end{aligned}$$

Taking the potential derivatives of $\mathcal{L}_h(u_k, v_k)$ with respect to the vectors $\boldsymbol{a}_r(k)$, $\boldsymbol{b}_r(k)$ and $\boldsymbol{b}_r(n+k)$ respectively, letting each of them to be zero, and solving the equation, we can obtain the minimizer of $\mathcal{L}_h(u_k, v_k)$ as

$$[\hat{\boldsymbol{a}}_r^{\mathrm{T}}(k), \hat{\boldsymbol{b}}_r^{\mathrm{T}}(k), \hat{\boldsymbol{b}}_r^{\mathrm{T}}(n+k)]^{\mathrm{T}} = \hat{\boldsymbol{P}}_h(u_k, v_k)\boldsymbol{y}, \qquad (6)$$

where $\hat{\boldsymbol{P}}_h(u_k, v_k)$ is

$$\begin{pmatrix} \boldsymbol{X}^{\mathrm{T}}\boldsymbol{W}_h^{(0,0)}(k)\boldsymbol{X} & \boldsymbol{X}^{\mathrm{T}}\boldsymbol{W}_h^{(1,0)}(k)\boldsymbol{X} & \boldsymbol{X}^{\mathrm{T}}\boldsymbol{W}_h^{(0,1)}(k)\boldsymbol{X} \\ \boldsymbol{X}^{\mathrm{T}}\boldsymbol{W}_h^{(1,0)}(k)\boldsymbol{X} & \boldsymbol{X}^{\mathrm{T}}\boldsymbol{W}_h^{(2,0)}(k)\boldsymbol{X} & \boldsymbol{X}^{\mathrm{T}}\boldsymbol{W}_h^{(1,1)}(k)\boldsymbol{X} \\ \boldsymbol{X}^{\mathrm{T}}\boldsymbol{W}_h^{(0,1)}(k)\boldsymbol{X} & \boldsymbol{X}^{\mathrm{T}}\boldsymbol{W}_h^{(1,1)}(k)\boldsymbol{X} & \boldsymbol{X}^{\mathrm{T}}\boldsymbol{W}_h^{(0,2)}(k)\boldsymbol{X} \end{pmatrix}^{-1} \begin{pmatrix} \boldsymbol{X}^{\mathrm{T}}\boldsymbol{W}_h^{(0,0)}(k) \\ \boldsymbol{X}^{\mathrm{T}}\boldsymbol{W}_h^{(1,0)}(k) \\ \boldsymbol{X}^{\mathrm{T}}\boldsymbol{W}_h^{(0,1)}(k) \end{pmatrix} \qquad (7)$$

and

$$\begin{aligned}\boldsymbol{W}_h^{(\gamma,\psi)}(k) &= \mathrm{Diag}[(u_1 - u_k)^{\gamma}(v_1 - v_k)^{\psi}K_h(d_{1k}), (u_2 - u_k)^{\gamma}(v_2 - v_k)^{\psi}K_h(d_{2k}), \\ &\quad \ldots, (u_n - u_k)^{\gamma}(v_n - v_k)^{\psi}K_h(d_{nk})] \qquad (8)\end{aligned}$$

with $\gamma, \psi = 0, 1, 2$.

The optimal size of the bandwidth h can be selected by some data-driven procedures such as the CV and AIC_c criterions. Here, we give a detailed description of the CV criterion which will be used in the subsequent section.

Let $\boldsymbol{x}_i^{\mathrm{T}} = (x_{i1}, x_{i2}, \ldots, x_{ip})$ be the ith row of $\boldsymbol{X}$ defined in Equation (3). The fitted values of the response variable Y at the n locations can be computed by

$$\hat{\boldsymbol{y}} = (\hat{y}_1, \hat{y}_2, \ldots, \hat{y}_n)^{\mathrm{T}} = \boldsymbol{L}(h)\boldsymbol{y}, \qquad (9)$$

where

$$\boldsymbol{L}(h) = \begin{pmatrix} (\boldsymbol{x}_1^{\mathrm{T}}, \boldsymbol{0}_{1\times(2p)})\hat{\boldsymbol{P}}_h(u_1, v_1) \\ (\boldsymbol{x}_2^{\mathrm{T}}, \boldsymbol{0}_{1\times(2p)})\hat{\boldsymbol{P}}_h(u_2, v_2) \\ \vdots \\ (\boldsymbol{x}_n^{\mathrm{T}}, \boldsymbol{0}_{1\times(2p)})\hat{\boldsymbol{P}}_h(u_n, v_n) \end{pmatrix}, \qquad (10)$$

with $\boldsymbol{0}_{1\times(2p)}$ being a zero matrix of order $1 \times (2p)$.

It is observed from Equation (9) that the local-linear estimation method for the GWR model is a linear smoother. As well known, the CV score can be computed by

$$CV(h) = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{y_i - \hat{y}_i}{1 - l_{ii}(h)}\right)^2, \qquad (11)$$

where $l_{ii}(h)$ is the ith element in the diagonal of $\boldsymbol{L}(h)$. The optimal size, say h_o, of h is then

$$h_o = \arg\min_{h>0} CV(h). \qquad (12)$$

3. GWGlasso method for structure identification and variable selection in the GWR model

3.1. Penalized estimation of the coefficients and their partial derivatives via the adaptive group lasso

3.1.1. The penalized objective function

Let

$$\boldsymbol{A} = [\boldsymbol{a}_r^{\mathrm{T}}(1), \boldsymbol{a}_r^{\mathrm{T}}(2), \ldots, \boldsymbol{a}_r^{\mathrm{T}}(n)]^{\mathrm{T}} = [\boldsymbol{a}_c(1), \boldsymbol{a}_c(2), \ldots, \boldsymbol{a}_c(p)], \tag{13}$$

where the kth row $\boldsymbol{a}_r^{\mathrm{T}}(k)$ of $\boldsymbol{A}$, as shown in Equation (4), is the values of the p coefficients in the GWR model (1) at the location (u_k, v_k) and the jth column

$$\boldsymbol{a}_c(j) = [\beta_j(u_1, v_1), \beta_j(u_2, v_2), \ldots \beta_j(u_n, v_n)]^{\mathrm{T}}, \tag{14}$$

consists of the values of the jth coefficient at the n designated locations. Similarly, let

$$\boldsymbol{B} = [\boldsymbol{b}_r^{\mathrm{T}}(1), \ldots, \boldsymbol{b}_r^{\mathrm{T}}(n), \boldsymbol{b}_r^{\mathrm{T}}(n+1), \ldots, \boldsymbol{b}_r^{\mathrm{T}}(2n)]^{\mathrm{T}} = [\boldsymbol{b}_c(1), \boldsymbol{b}_c(2) \ldots, \boldsymbol{b}_c(p)], \tag{15}$$

where the kth row $\boldsymbol{b}_r^{\mathrm{T}}(k)$ of $\boldsymbol{B}$, as shown in Equation (4), is the values of the partial derivatives of the p coefficients at location (u_k, v_k) with respect to u $(1 \le k \le n)$ and v $(n+1 \le k \le 2n)$, respectively, and the jth column

$$\boldsymbol{b}_c(j) = [\beta_j^{(u)}(u_1, v_1), \ldots, \beta_j^{(u)}(u_n, v_n), \beta_j^{(v)}(u_1, v_1), \ldots, \beta_j^{(v)}(u_n, v_n)]^{\mathrm{T}}, \tag{16}$$

consists of the values of the two partial derivatives of the jth coefficient at the n designed locations.

In what follows, we denote by $\hat{\boldsymbol{a}}_r^{(0)}(k)$, $\hat{\boldsymbol{b}}_r^{(0)}(k)$ and $\hat{\boldsymbol{a}}_r^{(0)}(n+k)$ the local-linear estimates of $\boldsymbol{a}_r(k)$, $\boldsymbol{b}_r(k)$ and $\boldsymbol{a}_r(n+k)$ obtained by Equation (6) with the bandwidth h set to be its optimal value in Equation (11). Let $k = 1, 2, \ldots, n$, respectively, we can obtain the estimates of $\boldsymbol{A}$ and $\boldsymbol{B}$ which we denote by $\hat{\boldsymbol{A}}^{(0)}$ and $\hat{\boldsymbol{B}}^{(0)}$. The jth column of $\hat{\boldsymbol{A}}^{(0)}$ and $\hat{\boldsymbol{B}}^{(0)}$, that is, the estimates of $\boldsymbol{a}_c(j)$ and $\boldsymbol{b}_c(j)$ is denoted by $\boldsymbol{a}_c^{(0)}(j)$ and $\boldsymbol{b}_c^{(0)}(j)$, respectively.

According to the principle of the adaptive group lasso [19,20], we formulate the objective function of the penalized estimation of the GWR model in Equation (1) as

$$\mathcal{L}_{h,\lambda}(\boldsymbol{A}, \boldsymbol{B}) = \sum_{k=1}^{n} \mathcal{L}_h(u_k, v_k) + 2\lambda \left(\sum_{j=1}^{p} w_{1j} \|\boldsymbol{a}_c(j)\| + \sum_{j=1}^{p} w_{2j} \|\boldsymbol{b}_c(j)\| \right), \tag{17}$$

where $\|\cdot\|$ is the Euclidean norm of a vector, that is,

$$\|\boldsymbol{a}_c(j)\| = \left[\sum_{i=1}^{n} (\beta_j(u_i, v_i))^2 \right]^{1/2}, \|\boldsymbol{b}_c(j)\| = \left[\sum_{i=1}^{n} [(\beta_j^{(u)}(u_i, v_i))^2 + (\beta_j^{(v)}(u_i, v_i))^2] \right]^{1/2}, \tag{18}$$

w_{1j} and w_{2j} are the weights given by

$$w_{1j} = \frac{\sqrt{n}}{\hat{\boldsymbol{a}}_c^{(0)}(j)}, \quad w_{2j} = \frac{\sqrt{2n}}{\hat{\boldsymbol{b}}_c^{(0)}(j)}, \quad j = 1, 2, \ldots, p, \tag{19}$$

with $\hat{\boldsymbol{a}}_c^{(0)}(j)$ and $\hat{\boldsymbol{b}}_c^{(0)}(j)$ being the aforementioned estimates of $\boldsymbol{a}_c(j)$ and $\boldsymbol{b}_c(j)$, and λ is the penalization parameter. In the objective function $\mathcal{L}_{h,\lambda}(\boldsymbol{A}, \boldsymbol{B})$, the penalty terms $\sum_{j=1}^{p} w_{1j} \|\boldsymbol{a}_c(j)\|$ and $\sum_{j=1}^{p} w_{1j} \|\boldsymbol{b}_c(j)\|$ are used to cause shrinkage of the values of each coefficient and their two partial derivatives at all of the designed locations towards zero. Therefore, if the coefficient $\beta_j(u,v)$ is zero over space, the estimates of both $\boldsymbol{a}_c(j)$ and $\boldsymbol{b}_c(j)$ should be shrunk to zero; if $\beta_j(u,v)$ is nonzero constant, the estimates of only $\boldsymbol{b}_c(j)$ should be shrunk to zero. As a result, the shrunken estimates of $\boldsymbol{a}_c(j)$ and $\boldsymbol{b}_c(j)$ from $\mathcal{L}_{h,\lambda}(\boldsymbol{A},\boldsymbol{B})$, provide the evidence for identifying whether $\beta_j(u,v)$ is zero, nonzero constant, or varying over space.

3.1.2. Local quadratic approximation of the penalty terms and an iterative algorithm for the penalized estimation

It is noted that the penalty terms in the objective function $\mathcal{L}_{h,\lambda}(\boldsymbol{A},\boldsymbol{B})$ are not differentiable in the origin, which makes the common derivative-based algorithm unusable for obtaining the solutions of $\mathcal{L}_{h,\lambda}(\boldsymbol{A},\boldsymbol{B})$. However, as done in many papers (see, [11,15,17]), the local quadratic approximation proposed by Fan and Li [16] can be used to locally approximate the penalty terms. In our cases, this approximation shows, for each $j = 1,2,\ldots,p$, that

$$\begin{aligned} \|\boldsymbol{a}_c(j)\| &\approx \|\hat{\boldsymbol{a}}_c^{(m)}(j)\| + \frac{\|\boldsymbol{a}_c(j)\|^2 - \|\hat{\boldsymbol{a}}_c^{(m)}(j)\|^2}{2\|\hat{\boldsymbol{a}}_c^{(m)}(j)\|}; \\ \|\boldsymbol{b}_c(j)\| &\approx \|\hat{\boldsymbol{b}}_c^{(m)}(j)\| + \frac{\|\boldsymbol{b}_c(j)\|^2 - \|\hat{\boldsymbol{b}}_c^{(m)}(j)\|^2}{2\|\hat{\boldsymbol{b}}_c^{(m)}(j)\|}, \end{aligned} \tag{20}$$

where $\hat{\boldsymbol{a}}_c^{(m)}(j)$ and $\hat{\boldsymbol{b}}_c^{(m)}(j)$ are two known vectors with their norms being close to those of $\boldsymbol{a}_c(j)$ and $\boldsymbol{b}_c(j)$, respectively. In what follows, an iterative algorithm will be derived to compute the estimates of $\boldsymbol{a}_c(j)$ and $\boldsymbol{b}_c(j)$, in which $\hat{\boldsymbol{a}}_c^{(m)}(j)$ and $\hat{\boldsymbol{b}}_c^{(m)}(j)$ are set to be the latest estimates of $\boldsymbol{a}_c(j)$ and $\boldsymbol{b}_c(j)$.

3.1.3. An iterative algorithm for solving the objective function and identification of the types of the coefficients

With the quadratic approximation of the penalty terms in Equation (20), the objective function in Equation (17) can be approximated by

$$\mathcal{L}_{h,\lambda}(\boldsymbol{A},\boldsymbol{B}) \approx \sum_{k=1}^{n} \mathcal{L}_h(u_k,v_k) + 2\lambda\left(\sum_{j=1}^{p}\frac{w_{1j}\|\boldsymbol{a}_c(j)\|^2}{\|\hat{\boldsymbol{a}}_c^{(m)}(j)\|} + \sum_{j=1}^{p}\frac{w_{2j}\|\boldsymbol{b}_c(j)\|^2}{\|\hat{\boldsymbol{b}}_c^{(m)}(j)\|}\right) + c(m), \tag{21}$$

where $c(m) = -\lambda(\sum_{j=1}^{p} w_{1j}\|\hat{\boldsymbol{a}}_c^{(m)}(j)\| + \sum_{j=1}^{p} w_{2j}\|\hat{\boldsymbol{b}}_c^{(m)}(j)\|)$ is irrelevant to the elements of $\boldsymbol{A}$ and $\boldsymbol{B}$. From Equations (4) and (14), we obtain that

$$\begin{aligned} \sum_{j=1}^{p}\frac{w_{1j}}{\|\hat{\boldsymbol{a}}_c^{(m)}(j)\|}\|\boldsymbol{a}_c(j)\|^2 &= \sum_{j=1}^{p}\frac{w_{1j}}{\|\hat{\boldsymbol{a}}_c^{(m)}(j)\|}\left[\sum_{k=1}^{n}\beta_j^2(u_k,v_k)\right] \\ &= \sum_{k=1}^{n}\left[\sum_{j=1}^{p}\frac{w_{1j}}{\|\hat{\boldsymbol{a}}_c^{(m)}(j)\|}\beta_j^2(u_k,v_k)\right] = \sum_{k=1}^{n}\boldsymbol{a}_r^{\mathrm{T}}(k)\boldsymbol{D}_1^{(m)}\boldsymbol{a}_r(k), \end{aligned} \tag{22}$$

where

$$\boldsymbol{D}_1^{(m)} = \mathrm{Diag}\left(\frac{w_{11}}{\|\hat{\boldsymbol{a}}_c^{(m)}(1)\|}, \frac{w_{12}}{\|\hat{\boldsymbol{a}}_c^{(m)}(2)\|}, \ldots, \frac{w_{1p}}{\|\hat{\boldsymbol{a}}_c^{(m)}(p)\|}\right). \tag{23}$$

Similarly, we have

$$\sum_{j=1}^{p}\frac{w_{2j}}{\|\hat{\boldsymbol{b}}_c^{(m)}(j)\|}\|\boldsymbol{b}_c(j)\|^2 = \sum_{k=1}^{n}\boldsymbol{b}_r^{\mathrm{T}}(k)\boldsymbol{D}_2^{(m)}\boldsymbol{b}_r(k) + \sum_{k=1}^{n}\boldsymbol{b}_r^{\mathrm{T}}(n+k)\boldsymbol{D}_2^{(m)}\boldsymbol{b}_r(n+k), \tag{24}$$

where

$$\boldsymbol{D}_2^{(m)} = \mathrm{Diag}\left(\frac{w_{21}}{\|\hat{\boldsymbol{b}}_c^{(m)}(1)\|}, \frac{w_{22}}{\|\hat{\boldsymbol{b}}_c^{(m)}(2)\|}, \ldots, \frac{w_{2p}}{\|\hat{\boldsymbol{b}}_c^{(m)}(p)\|}\right). \tag{25}$$

Therefore, the objective function $\mathcal{L}_{h,\lambda}(\boldsymbol{A}, \boldsymbol{B})$ is of the form

$$\begin{aligned}\mathcal{L}_{h,\lambda}(\boldsymbol{A}, \boldsymbol{B}) = \sum_{k=1}^{n} \mathcal{L}_h(u_k, v_k) + \lambda \sum_{k=1}^{n} [\boldsymbol{a}_r^{\mathrm{T}}(k)\boldsymbol{D}_1^{(m)}\boldsymbol{a}_r(k) + \boldsymbol{b}_r^{\mathrm{T}}(k)\boldsymbol{D}_2^{(m)}\boldsymbol{b}_r(k) \\ + \boldsymbol{b}_r^{\mathrm{T}}(n+k)\boldsymbol{D}_2^{(m)}\boldsymbol{b}_r(n+k)] + c(m).\end{aligned} \tag{26}$$

Substituting $\mathcal{L}_h(u_k, v_k)$ in Equation (5) into the expression of $\mathcal{L}_{h,\lambda}(\boldsymbol{A}, \boldsymbol{B})$ and with the similar operation for deriving Equation (7), we obtain the following formula for iteratively computing the penalized estimates of $\boldsymbol{a}_r(k)$, $\boldsymbol{b}_r(k)$ and $\boldsymbol{b}_r(n+k)$ as

$$\begin{aligned}&[[\hat{\boldsymbol{a}}_r^{(m)}(k)]^{\mathrm{T}}, [\hat{\boldsymbol{b}}_r^{(m)}(k)]^{\mathrm{T}}, [\hat{\boldsymbol{b}}_r(n+k)]^{\mathrm{T}}]^{\mathrm{T}} \\ &= \begin{pmatrix} \boldsymbol{X}^{\mathrm{T}}\boldsymbol{W}_h^{(0,0)}(k)\boldsymbol{X} + \lambda\boldsymbol{D}_1^{(m)} & \boldsymbol{X}^{\mathrm{T}}\boldsymbol{W}_h^{(1,0)}(k)\boldsymbol{X} & \boldsymbol{X}^{\mathrm{T}}\boldsymbol{W}_h^{(0,1)}(k)\boldsymbol{X} \\ \boldsymbol{X}^{\mathrm{T}}\boldsymbol{W}_h^{(1,0)}(k)\boldsymbol{X} & \boldsymbol{X}^{\mathrm{T}}\boldsymbol{W}_h^{(2,0)}(k)\boldsymbol{X} + \lambda\boldsymbol{D}_2^{(m)} & \boldsymbol{X}^{\mathrm{T}}\boldsymbol{W}_h^{(1,1)}(k)\boldsymbol{X} \\ \boldsymbol{X}^{\mathrm{T}}\boldsymbol{W}_h^{(0,1)}(k)\boldsymbol{X} & \boldsymbol{X}^{\mathrm{T}}\boldsymbol{W}_h^{(1,1)}(k)\boldsymbol{X} & \boldsymbol{X}^{\mathrm{T}}\boldsymbol{W}_h^{(0,2)}(k)\boldsymbol{X} + \lambda\boldsymbol{D}_2^{(m)} \end{pmatrix}^{-1} \\ &\quad \cdot \begin{pmatrix} \boldsymbol{X}^{\mathrm{T}}\boldsymbol{W}_h^{(0,0)}(k) \\ \boldsymbol{X}^{\mathrm{T}}\boldsymbol{W}_h^{(1,0)}(k) \\ \boldsymbol{X}^{\mathrm{T}}\boldsymbol{W}_h^{(0,1)}(k) \end{pmatrix} \boldsymbol{y},\end{aligned} \tag{27}$$

where $\boldsymbol{W}_h^{(\gamma,\psi)}(k)$ $(\gamma, \psi = 0, 1, 2)$ are shown in Equation (8), and $\boldsymbol{D}_1^{(m)}$ and $\boldsymbol{D}_2^{(m)}$ are defined in Equations (23) and (25). In the above iterative algorithm, only $\boldsymbol{D}_1^{(m)}$ and $\boldsymbol{D}_2^{(m)}$ should be updated in each iteration. Once the bandwidth h, the penalization parameter λ and the initial values of $\boldsymbol{A}$ and $\boldsymbol{B}$ (or $\boldsymbol{a}_r(k)$, $\boldsymbol{b}_r(k)$ and $\boldsymbol{b}_r(n+k)$ for $k = 1, 2, \ldots, n$) are specified. Then penalized estimates of $\boldsymbol{a}_r(k)$, $\boldsymbol{b}_r(k)$ and $\boldsymbol{b}_r(n+k)$ for $k = 1, 2, \ldots, n$, and therefore those of $\boldsymbol{A}$ and $\boldsymbol{B}$ can be obtained by performing the iterations until convergence. Let $\hat{\boldsymbol{A}}^{(m)}$ and $\hat{\boldsymbol{B}}^{(m)}$ be the estimates of $\boldsymbol{A}$ and $\boldsymbol{B}$ in the mth iteration, the convergence criterion here is defined by

$$\left\| \begin{pmatrix} \hat{\boldsymbol{A}}^{(m)} \\ \hat{\boldsymbol{B}}^{(m)} \end{pmatrix} - \begin{pmatrix} \hat{\boldsymbol{A}}^{(m+1)} \\ \hat{\boldsymbol{B}}^{(m+1)} \end{pmatrix} \right\|_F < \tau, \tag{28}$$

where $\|\cdot\|_F$ indicates the Frobenius norm of a matrix which is equal to the squared root of sum of squares of all the elements of the matrix, and τ is a pre-specified threshold value. When the convergence criterion is reached, we take $\hat{\boldsymbol{A}}^{(m)}$ and $\hat{\boldsymbol{B}}^{(m)}$ as the final estimates of $\boldsymbol{A}$ and $\boldsymbol{B}$, which we denote in what follows as $\hat{\boldsymbol{A}}_{h,\lambda}$ and $\hat{\boldsymbol{B}}_{h,\lambda}$, respectively. Consequently, the final estimates of $\boldsymbol{a}_c(j)$ and $\boldsymbol{b}_c(j)$ for $j = 1, 2, \ldots, p$ are obtained, which we denote by $\hat{\boldsymbol{a}}_c(j)$ and $\hat{\boldsymbol{b}}_c(j)$ $(j = 1, 2, \ldots, p)$ hereafter.

Theoretically, if the coefficient $\beta_j(u, v)$ is zero, both the estimates $\hat{\boldsymbol{a}}_c(j)$ and $\hat{\boldsymbol{b}}_c(j)$ should be shrunk exactly to zero; if it is a nonzero constant, only $\hat{\boldsymbol{b}}_c(j)$ is shrunk exactly to zero. However, with the local quadratic approximation of the objective function $\mathcal{L}_{h,\lambda}(\boldsymbol{A}, \boldsymbol{B})$ and the iterative algorithm, $\hat{\boldsymbol{a}}_c(j)$ and $\hat{\boldsymbol{b}}_c(j)$ may not be exactly zero when the convergence criterion in Equation (28) is reached. As done in the literature on structure identification and variable selection of varying coefficient models (see, e.g.

[13,15,17]), a small threshold $\delta > 0$ should be designated to judge whether $\hat{\boldsymbol{a}}_c(j)$ and $\hat{\boldsymbol{b}}_c(j)$ have been shrunk to zero. Specifically, for $1 \le j \le p$, let

$$\hat{\boldsymbol{a}}_c(j) = [\hat{\beta}_j(u_1, v_1), \hat{\beta}_j(u_2, v_2), \dots, \hat{\beta}_j(u_n, v_n)]^{\mathrm{T}}, \tag{29}$$

$$\hat{\boldsymbol{b}}_c(j) = [\hat{\beta}_j^{(u)}(u_1, v_1), \dots, \hat{\beta}_j^{(u)}(u_n, v_n), \hat{\beta}_j^{(v)}(u_1, v_1), \dots, \hat{\beta}_j^{(v)}(u_n, v_n)]^{\mathrm{T}}. \tag{30}$$

Given each $j = 1, 2, \dots, p$, if $|\hat{\beta}_j(u_i, v_i)| < \delta$ for all $i = 1, 2, \dots, n$, we set $\hat{\boldsymbol{a}}_c(j) = \mathbf{0}$; if $|\hat{\beta}_j^{(u)}(u_i, v_i)| < \delta$ and $|\hat{\beta}_j^{(v)}(u_i, v_i)| < \delta$ for all $i = 1, 2, \dots, n$, we set $\hat{\boldsymbol{b}}_c(j) = \mathbf{0}$. With this reset estimates of $\boldsymbol{a}_c(j)$ and $\boldsymbol{b}_c(j)$, the corresponding coefficient $\beta_j(u, v)$ is identified to be zero, nonzero constant or spatially varying in the following way:

(1) If both $\hat{\boldsymbol{a}}_c(j) = \mathbf{0}$ and $\hat{\boldsymbol{b}}_c(j) = \mathbf{0}$, then $\beta_j(u, v) = 0$;
(2) if only $\hat{\boldsymbol{b}}_c(j) = \mathbf{0}$, then $\beta_j(u, v) = \beta_j$, where β_j is a nonzero constant;
(3) if otherwise, $\beta_j(u, v)$ is identified to be spatially varying.

3.1.4. Selection of the bandwidth h, the initial estimates of A and B, and the penalization parameter λ

In order to implement the aforementioned algorithm, the sizes of the bandwidth h and the penalization parameter λ should be properly selected and initial estimates $\boldsymbol{A}$ and $\boldsymbol{B}$ have to be designated. Ideally, the optimal sizes of h and λ should be simultaneously selected by some data-driven criterion. However, to do so is very computationally expensive. Following the way in Wang and Xia [11], Hu and Xia [14], and Ma and Zhang [17], we separately select the optimal sizes of h and λ in the following way.

The optimal size of the bandwidth h is taken to be h_o which is selected by the CV procedure in Equation (12). The aforementioned matrices $\hat{\boldsymbol{A}}^{(0)}$ and $\hat{\boldsymbol{B}}^{(0)}$ resulted from the local-linear estimates of $\hat{\boldsymbol{a}}_r(k)$, $\hat{\boldsymbol{b}}_r(k)$ and $\hat{\boldsymbol{b}}_r(n+k)$ $(k = 1, 2, \dots, n)$ are set to be the initial estimates of $\boldsymbol{A}$ and $\boldsymbol{B}$ for running the iterative algorithm in Equation (26).

For the penalization parameter λ, its optimal size is chosen by the BIC-type criterion proposed by Hu and Xia [14]. Specifically, given $\lambda > 0$ and the selected bandwidth h_o, run the iterative algorithm in Equation (26) and yields the estimates $\hat{\boldsymbol{A}}_{h_o,\lambda}$ and $\hat{\boldsymbol{B}}_{h_o,\lambda}$ of $\boldsymbol{A}$ and $\boldsymbol{B}$ with which the varying coefficients in the model is identified by the criterion described in the end of the last subsection. Let $df_{h_o,\lambda}$ be the number of the varying coefficients and

$$RSS_{h_o,\lambda} = \sum_{k=1}^{n} \sum_{i=1}^{n} \{y_i - \boldsymbol{x}_i[\hat{\boldsymbol{a}}_r(k) + \hat{\boldsymbol{b}}_r(k)(u_i - u_k) + \hat{\boldsymbol{b}}_r(n+k)(v_i - v_k)]\}^2 K_{h_o}(d_{ki}), \tag{31}$$

be the residual sum of squares computed by $\hat{\boldsymbol{A}}_{h_o,\lambda}$ and $\hat{\boldsymbol{B}}_{h_o,\lambda}$, where each row of $\hat{\boldsymbol{A}}_{h_o,\lambda}$ and $\hat{\boldsymbol{B}}_{h_o,\lambda}$ is still denoted by $\hat{\boldsymbol{a}}_r(k)$, $\hat{\boldsymbol{b}}_r(k)$ and $\hat{\boldsymbol{b}}_r(n+k)$ for notational simplicity. According to Hu and Xia [14], define the BIC-type criterion

$$\mathrm{BIC}(\lambda) = \log\left(\frac{1}{n^2} RSS_{h_o,\lambda}\right) + df_\lambda \frac{\log(nh)}{nh} + (p - df_\lambda)\frac{\log n}{n}. \tag{32}$$

The optimal size λ_o of the penalization parameter is selected by

$$\lambda_o = \arg\min_{\lambda > 0} \mathrm{BIC}(\lambda). \tag{33}$$

3.2. Implementation steps of the GWGlasso method

In order to facilitate the implementation of the GWGlasso method on computers, we summarize the basic steps in what follows.

Input: Spatial data set $\{y_i; x_{i1}, \ldots, x_{ip}; (u_i, v_i)\}_{i=1}^{n}$; the values of τ and δ; and the candidate sets of the bandwidth h and penalization parameter λ which are denoted by $\mathcal{H} = \{h_m\}_{m=1}^{M}$ and $\Lambda = \{\lambda_l\}_{l=1}^{L}$, respectively.

Output: The index sets of the varying coefficients, nonzero constant coefficients and zero coefficients which are denoted by $\mathcal{A}_V$, $\mathcal{A}_C$ and $\mathcal{A}_Z$, respectively.

Step 1: Compute the initial matrices $\hat{\boldsymbol{A}}^{(0)}$ and $\hat{\boldsymbol{B}}^{(0)}$ of $\boldsymbol{A}$ and $\boldsymbol{B}$ as well as the weights $\{w_{1j}, w_{2j}\}_{j=1}^{p}$.

Step 1.1: Given each $h_m \in \mathcal{H}$, compute the estimates $\hat{\boldsymbol{a}}_r^{(0)}(k)$, $\hat{\boldsymbol{b}}_r^{(0)}(k)$ and $\hat{\boldsymbol{a}}_r^{(0)}(n+k)$ of $\boldsymbol{a}_r^{(0)}(k)$, $\boldsymbol{b}_r^{(0)}(k)$ and $\boldsymbol{a}_r^{(0)}(n+k)$ according to Equation (7) for $k = 1, 2, \ldots, n$; then compute $CV(h_m)$ by Equation (11);

Step 1.2: Select $h_o = \arg\min_{1 \le m \le M} CV(h_m)$ and the estimates $\hat{\boldsymbol{a}}_r^{(0)}(k)$, $\hat{\boldsymbol{b}}_r^{(0)}(k)$ and $\hat{\boldsymbol{a}}_r^{(0)}(n+k)$ $(k = 1, 2, \ldots, n)$ corresponding to the bandwidth size h_o obtained in *Step* 1.1; and then formulate $\hat{\boldsymbol{A}}^{(0)}$ and $\hat{\boldsymbol{B}}^{(0)}$ according to Equations (13) and (15) as well as the weights $\{w_{1j}, w_{2j}\}_{j=1}^{p}$ according to Equation (19).

Step 2: Select the optimal size λ_o and corresponding estimates $\hat{\boldsymbol{A}}_{h_o,\lambda_o}$ and $\hat{\boldsymbol{B}}_{h_o,\lambda_o}$ of $\boldsymbol{A}$ and $\boldsymbol{B}$.

Step 2.1: For each $\lambda_l \in \Lambda$ and with $\hat{\boldsymbol{A}}^{(0)}$, $\hat{\boldsymbol{B}}^{(0)}$ and $\{w_{1j}, w_{2j}\}_{j=1}^{p}$ obtained in *Step* 1, run the iterative formula in Equation (27) for each $k = 1, 2, \ldots, n$ until the convergence criterion in Equation (28) is met and consequently obtain the estimates $\hat{\boldsymbol{A}}_{h_o,\lambda_l}$ and $\hat{\boldsymbol{B}}_{h_o,\lambda_l}$; then determine df_{h_o,λ_l} according to the identification criterion described in the end of Section 3.1.3 and compute $\mathrm{BIC}(\lambda_l)$ by Equation (32).

Step 2.2: Select $\lambda_o = \arg\min_{1 \le l \le L} \mathrm{BIC}(\lambda_l)$ and output the corresponding estimates $\hat{\boldsymbol{A}}_{h_o,\lambda_l}$ and $\hat{\boldsymbol{B}}_{h_o,\lambda_l}$ in Step 2.1.

Step 3: Reset each column of $\hat{\boldsymbol{A}}_{h_o,\lambda_l}$ and $\hat{\boldsymbol{B}}_{h_o,\lambda_l}$ according to the identification criterion in the end of Section 3.1.3 on which final results are obtained by

$\mathcal{A}_Z = \{j :$ the jth column of both $\hat{\boldsymbol{A}}_{h_o,\lambda_o}$ and $\hat{\boldsymbol{B}}_{h_o,\lambda_o}$ is reset to be a zero vector$\}$;

$\mathcal{A}_C = \{j :$ the jth column of only $\hat{\boldsymbol{B}}_{h_o,\lambda_o}$ is reset to be a zero vector$\}$;

$\mathcal{A}_V = \{j : j \notin \mathcal{A}_Z$ or $j \notin \mathcal{A}_C\}$.

Remark: In general, the GWGlasso method leads to a mixed GWR model when the irrelevant explanatory variables (if any) are removed. With the above notations, the mixed GWR model is of the form

$$y_i = \sum_{j \in \mathcal{A}_V} \beta_j(u_i, v_i) x_{ij} + \sum_{j \in \mathcal{A}_C} \beta_j x_{ij} + \varepsilon_i, \quad i = 1, 2, \ldots, n. \tag{34}$$

The GWGlasso method can also yield the shrunk estimates of the spatially varying coefficients and the nonzero constant coefficients. Specially, denote

$$\hat{\boldsymbol{A}}_{h_o,\lambda_o} = [\hat{\boldsymbol{a}}_c(1), \hat{\boldsymbol{a}}_c(2), \ldots \hat{\boldsymbol{a}}_c(p)], \tag{35}$$

with $\hat{\boldsymbol{a}}_c(j) = [\hat{\beta}_j(u_1, v_1), \hat{\beta}_j(u_2, v_2), \ldots, \hat{\beta}_j(u_n, v_n)]^{\mathrm{T}}$ for $j = 1, 2, \ldots, p$.

(i) If $j \in \mathcal{A}_V$, then the elements in $\hat{\boldsymbol{a}}_c(j)$ are the shrunk estimates of $\beta_j(u, v)$ at the n designed locations.

(ii) If $j \in \mathcal{A}_C$, the nonzero constant coefficient β_j can be estimated by averaging the shrunk estimates $\{\hat{\beta}_j(u_i, v_i)\}_{i=1}^{n}$ over the n locations. That is, $\beta_j = (1/n) \sum_{i=1}^{n} \{\hat{\beta}_j(u_i, v_i)\}_{i=1}^{n}$.

As well known, shrinkage estimation methodologies can well deal with the impact of collinearity among the explanatory variables. Therefore, if the collinearity is diagnosed to be serious, the GWGlasso method can simultaneously yield better estimates for both spatially varying coefficients and nonzero constant coefficients. However, a recent paper by Fotheringham and Oshan [21] have empirically demonstrated that the GWR method is not more seriously influenced by collinearity among the explanatory variables except in the extreme circumstances and the collinearity in a GWR model should be treated the same as in any regression framework. Therefore, when the collinearity is thought to be not so serious, the two-steps estimation method in Fotheringham et al. [2, Chapter 3] would better be used to calibrate the mixed GWR model because extra bias will be introduced to the shrunk estimates of the coefficients in order to lower their variance.

4. Simulation study

In this section, the simulation study is to show that the GWGlasso method is a better computational procedure than the residual-based bootstrap test of Mei et al. [9]. Concerning the bootstrap test, Mei et al. [9] have proposed two types of the residual-based bootstrap tests. The first bootstrap test is to detect the constant coefficients in a GWR model. Furthermore, the second bootstrap test is developed for judging whether some of constant coefficients are zero in a mixed GWR model. Therefore, we implement these two types of the bootstrap tests for the purpose of identifying zero, nonzero constant and varying coefficients in a GWR model, and denote the first and second bootstrap test by $\boldsymbol{T}_1$ and $\boldsymbol{T}_2$ respectively. Compared with the bootstrap tests, the GWGlasso method is a shrinkage method and able to achieve the purpose of identifying zero, nonzero constant and varying coefficients simultaneously. Therefore, we conduct the simulation study to assess the performance of the GWGlasso method in Section 4.2. Under the same design of experiment, the simulation results of the two type of the bootstrap tests are reported in Section 4.3. Furthermore, the deficiency and benefit of the above two methods are analysed and discussed in Section 4.4.

4.1. Design of the experiment

(i) The spatial layout. The spatial region for the experiment is taken to be a unit square. A Cartesian coordinate system is built in such a way that its origin locates at the bottom-left corner of the square and the two axes coincide with the mutually orthogonal two sides of the square. The sampling locations are $m \times m$ lattice points with their coordinates under the Cartesian coordinate system being

$$(u_i, v_i) = \left(\frac{1}{m-1} \mathrm{mod}\left(\frac{i-1}{m}\right), \frac{1}{m-1} \mathrm{int}\left(\frac{i-1}{m}\right) \right), \quad i = 1, \ldots, m^2, \tag{36}$$

where mod(a/b) and int(a/b) are the remainder and the integer part on a divided by b, respectively. In the experiment, we take $m = 21$ which results in the sample size $n = m^2 = 441$.

(ii) The models. The following three GWR models are considered in the experiment.

$$\begin{aligned}(\mathrm{I}): \quad y_i = {} & 3(u_i + v_i)x_{i1} + (1 + v_i^2)x_{i2} + 1.5x_{i3} + \alpha x_{i4} + 0x_{i5} \\ & + 0x_{i6} + 0x_{i7} + 0x_{i8} + 0.5\varepsilon_i,\end{aligned} \tag{37}$$

$$\begin{aligned}(\mathrm{II}): \quad y_i = {} & 3(u_i + v_i)x_{i1} + (1 + \alpha v_i^2)x_{i2} + 1.5x_{i3} + 1x_{i4} + 2x_{i5} \\ & + 0x_{i6} + 0x_{i7} + 0x_{i8} + 0.5\varepsilon_i,\end{aligned} \tag{38}$$

$$\begin{aligned}(\mathrm{III}): \quad y_i = {} & \exp(u_i + v_i)x_{i1} + 6u_i^2 x_{i2} + 2\alpha \sin(2\pi v_i)x_{i3} + 1.5x_{i4} + 1x_{i5} \\ & + 2x_{i6} + 0x_{i7} + 0x_{i8} + 0.5\varepsilon_i,\end{aligned} \tag{39}$$

where for each $i = 1, 2, \ldots, n$, $x_{i1} = 1$; $(x_{i2}, \ldots, x_{i8})^{\mathrm{T}}$ are drawn from the multivariate normal distribution $N(\mathbf{0}, \boldsymbol{\Sigma})$ with the covariance matrix $\boldsymbol{\Sigma} = (\rho^{|j-k|})_{2 \le j,\, k \le 8}$, and ε_i is independently drawn from the standard normal distribution $N(0, 1)$.

In the models, the parameter α controls the variation degree of the coefficient. By designating different values of α, we can evaluate the ability of the proposed GWGlasso method in solving three types of structure identification problems: (I) selection of nonzero constant and zero coefficients; (II) separation of varying and nonzero constant coefficients; (III) identification of varying and zero coefficients. The degree of collinearity among the explanatory variables is reflected by the parameter ρ in the covariance matrix which we use to assess the impact of the collinearity on the performance of the GWGlasso method. In the experiment, α was set to be 0.5 and 1, respectively; ρ was taken to be 0, 0.5, 0.9, respectively, indicating the cases that the explanatory variables are independent with each other, moderately correlated and highly correlated.

(iii) The kernel function and candidate sets of the bandwidth h and penalization parameter λ. Throughout the simulation, the Gaussian kernel $K(t) = (1/\sqrt{2\pi})\exp(-t^2/2)$ was used, the candidate sets of the bandwidth h and the penalization parameter λ were set to be $\mathcal{H} = \{h_m : h_m = 0.05 + 0.05 \times m\}_{m=1}^{20}$ and $\Lambda = \{\lambda_l : \lambda_l = 0.2 \times l\}_{l=1}^{20}$.

4.2. Simulation analysis of the GWGlasso method

In each experimental setting, 200 replications were conducted. Tables 1–3 report the frequencies of each coefficient identified to be spatially varying, nonzero constant or zero in the 200 replications. Throughout the simulation, we set $\tau = 10^{-4}$ and $\delta = 10^{-2}$ be the stopping criterion and the judging threshold, respectively.

From Table 1, it can be seen that for selecting the nonzero constant coefficients, the GWGlasso method works well in every case, with high frequency, it can correctly separate the varying, nonzero constant and zero coefficients. Since α is designed in model (I), the frequency of the GWGlasso method in cases $\alpha = 1$ is slightly higher than that of in cases $\alpha = 0.5$, which indicates that the

Table 1. Frequencies of each coefficient identified to be spatially varying (V), nonzero constant (C) or zero (Z) in Model (I).

	$\alpha = 0.5$			$\alpha = 1$		
Coefficient	V	C	Z	V	C	Z
$\rho = 0$						
$\beta_1(u,v)$	**200**	0	0	**200**	0	0
$\beta_2(u,v)$	**200**	0	0	**200**	0	0
$\beta_3 = 1.5$	2	**198**	0	2	**198**	0
$\beta_4 = \alpha$	4	**196**	0	1	**199**	0
$\beta_5 = 0$	2	0	**198**	0	3	**197**
$\beta_6 = 0$	0	0	**200**	0	1	**199**
$\beta_7 = 0$	0	2	**198**	0	1	**199**
$\beta_8 = 0$	1	1	**198**	0	1	**199**
$\rho = 0.5$						
$\beta_1(u,v)$	**200**	0	0	**200**	0	0
$\beta_2(u,v)$	**200**	0	0	**200**	0	0
$\beta_3 = 1.5$	3	**197**	0	3	**197**	0
$\beta_4 = \alpha$	2	**198**	0	3	**197**	0
$\beta_5 = 0$	0	0	**200**	1	3	**196**
$\beta_6 = 0$	1	4	**195**	1	0	**199**
$\beta_7 = 0$	0	3	**197**	0	2	**198**
$\beta_8 = 0$	0	2	**198**	0	0	**200**
$\rho = 0.9$						
$\beta_1(u,v)$	**200**	0	0	**200**	0	0
$\beta_2(u,v)$	**195**	5	0	**186**	14	0
$\beta_3 = 1.5$	21	**179**	0	21	**179**	0
$\beta_4 = \alpha$	8	**192**	0	5	**195**	0
$\beta_5 = 0$	5	3	**192**	1	2	**197**
$\beta_6 = 0$	3	3	**194**	0	4	**196**
$\beta_7 = 0$	0	1	**199**	1	2	**197**
$\beta_8 = 0$	0	2	**198**	0	2	**198**

Note: The significance of bold values presents the frequencies of the underlying coefficient correctly identified into the final model.

Table 2. Frequencies of each coefficient identified to be spatially varying (V), nonzero constant (C) or zero (Z) in Model (II).

	$\alpha = 0.5$			$\alpha = 1$		
Coefficient	V	C	Z	V	C	Z
$\rho = 0$						
$\boldsymbol{\beta}_1(u,v)$	**200**	0	0	**200**	0	0
$\boldsymbol{\beta}_2(u,v)$	**186**	14	0	**200**	0	0
$\boldsymbol{\beta}_3 = 1.5$	3	**197**	0	2	**198**	0
$\boldsymbol{\beta}_4 = 1$	7	**193**	0	2	**198**	0
$\boldsymbol{\beta}_5 = 2$	4	**196**	0	2	**198**	0
$\boldsymbol{\beta}_6 = 0$	1	1	**198**	0	1	**199**
$\boldsymbol{\beta}_7 = 0$	0	2	**198**	0	0	**200**
$\boldsymbol{\beta}_8 = 0$	0	1	**199**	0	0	**200**
$\rho = 0.5$						
$\boldsymbol{\beta}_1(u,v)$	**200**	0	0	**200**	0	0
$\boldsymbol{\beta}_2(u,v)$	**178**	22	0	**200**	0	0
$\boldsymbol{\beta}_3 = 1.5$	10	**190**	0	1	**199**	0
$\boldsymbol{\beta}_4 = 1$	9	**191**	0	2	**198**	0
$\boldsymbol{\beta}_5 = 2$	7	**193**	0	1	**199**	0
$\boldsymbol{\beta}_6 = 0$	1	1	**198**	1	2	**197**
$\boldsymbol{\beta}_7 = 0$	0	2	**198**	0	1	**199**
$\boldsymbol{\beta}_8 = 0$	0	2	**198**	1	0	**199**
$\rho = 0.9$						
$\boldsymbol{\beta}_1(u,v)$	**200**	0	0	**200**	0	0
$\boldsymbol{\beta}_2(u,v)$	**117**	83	0	**187**	13	0
$\boldsymbol{\beta}_3 = 1.5$	38	**162**	0	18	**182**	0
$\boldsymbol{\beta}_4 = 1$	15	**185**	0	1	**199**	0
$\boldsymbol{\beta}_5 = 2$	5	**195**	0	4	**196**	0
$\boldsymbol{\beta}_6 = 0$	2	2	**196**	1	3	**196**
$\boldsymbol{\beta}_7 = 0$	1	1	**198**	0	1	**199**
$\boldsymbol{\beta}_8 = 0$	0	0	**200**	0	3	**197**

Note: The significance of bold values presents the frequencies of the underlying coefficient correctly identified into the final model.

proposed method is stable to separate the nonzero constant and zero coefficients. As one can see from Table 2, the performance of the proposed method in cases $\alpha = 1$ remarkably outperforms than the cases $\alpha = 0.5$, which indicates that the GWGlasso method is sensitive to separate varying and nonzero constant coefficients and detect the changes in the amplitude of variation of the spatially varying coefficients. Specially, for the separation results of $\boldsymbol{\beta}_2(u,v)$ the separation of varying and nonzero constant coefficients of the proposed method in cases $\alpha = 0.5$ is getting gradually worse with the increase of ρ. To investigate the performance of the GWGlasso method for identifying of varying and zero coefficients, we have carried out the simulation in model (III). For Table 3, it can be observed that the identification results $\boldsymbol{\beta}_3(u,v)$ of the proposed method in cases $\alpha = 1$ are better than that of the cases $\alpha = 0.5$, which suggests that the proposed method is sensitive to identify the coefficients between varying and zero coefficients. Moreover, the proposed method is robust to the moderately correlated collinearity among the explanatory variables even when the highly correlated cases. Therefore, all the results reported in Tables 1–3 demonstrate the favourable performance of the proposed method, especially, the ability for variable selection and structure identification in GWR models.

4.3. Simulation analysis of the residual-based bootstrap tests

In this subsection, we briefly describe the two types of the residual-based bootstrap tests proposed by Mei et al. [9]. For $\boldsymbol{T}_1$, we focus on testing for constant coefficients in the GWR model

$$H_0: \quad \text{some coefficients in model (1) are constant}$$

versus

$$H_1: \quad \text{all the coefficients in model (1) vary over the space}$$

Table 3. Frequencies of each coefficient identified to be spatially varying (V), nonzero constant (C) or zero (Z) in Model (III).

	$\alpha = 0.5$			$\alpha = 1$		
Coefficient	V	C	Z	V	C	Z
$\rho = 0$						
$\beta_1(u, v)$	**192**	8	0	**192**	8	0
$\beta_2(u, v)$	**171**	2	27	**169**	1	30
$\beta_3(u, v)$	**102**	12	86	**141**	3	56
$\beta_4 = 1.5$	0	**200**	0	5	**195**	0
$\beta_5 = 1$	0	**200**	0	2	**198**	0
$\beta_6 = 2$	0	**200**	0	3	**197**	0
$\beta_7 = 0$	0	0	**200**	8	0	**192**
$\beta_8 = 0$	0	0	**200**	4	0	**196**
$\rho = 0.5$						
$\beta_1(u, v)$	**195**	5	0	**193**	7	0
$\beta_2(u, v)$	**170**	1	29	**169**	5	26
$\beta_3(u, v)$	**98**	4	98	**145**	6	49
$\beta_4 = 1.5$	6	**194**	0	12	**188**	0
$\beta_5 = 1$	4	**196**	0	5	**193**	2
$\beta_6 = 2$	5	**195**	0	6	**194**	0
$\beta_7 = 0$	9	0	**191**	9	0	**191**
$\beta_8 = 0$	1	1	**198**	4	0	**196**
$\rho = 0.9$						
$\beta_1(u, v)$	**196**	4	0	**190**	10	0
$\beta_2(u, v)$	**164**	5	31	**162**	7	31
$\beta_3(u, v)$	**104**	5	91	**160**	7	33
$\beta_4 = 1.5$	56	**134**	10	58	**138**	4
$\beta_5 = 1$	48	**121**	31	63	**116**	21
$\beta_6 = 2$	54	**145**	1	60	**139**	1
$\beta_7 = 0$	52	0	**148**	53	0	**147**
$\beta_8 = 0$	21	0	**179**	35	0	**165**

Note: The significance of bold values presents the frequencies of the underlying coefficient correctly identified into the final model.

and formulate a residual-based bootstrap test for the hypotheses. That is, $\boldsymbol{T}_1$ can be viewed as a structure identification procedure. And the detailed bootstrap procedure for calculating the p-value of the test is given in Section 3.2 in Mei et al. [9].

For $\boldsymbol{T}_2$, we aim at detecting for zero coefficients in the mixed GWR model

H_0 : some constant coefficients in the mixed GWR model are zero

versus

H_1 : all the coefficients in the mixed GWR model are nonzero.

As an extension of $\boldsymbol{T}_1$, $\boldsymbol{T}_2$ is to test the hypothesis of one mixed GWR model against another mixed GWR model. Therefore, $\boldsymbol{T}_2$ can be regarded as a variable selection procedure. Then, the related computational strategy of the p-value can be obtained from Mei et al. [8] and the Section 3.3 in Mei et al. [9].

Under the same experimental settings, 200 replications are conducted. For each replication, 1000 bootstrap samples are drawn to compute the p-value. Based on the spatial data set $\{y_i; x_{i1}, \ldots, x_{ip}; (u_i, v_i)\}_{i=1}^n$, the GWR model in Equation (1) is fitted according to the basic GWR technique and the alternative model of the mixed GWR model is calibrated by the two-step estimation in Fotheringham et al. [2]. As suggested by Mei et al. [9], the Gaussian kernel $K(t) = (1/\sqrt{2\pi}) \exp(-t^2/2)$ is used and the optimal bandwidth size is selected by AIC_c criterion throughout the simulation of test.

For each of the experimental settings, the null and the alternative hypotheses of $\boldsymbol{T}_1$ and $\boldsymbol{T}_2$ are known in advance, where $\boldsymbol{T}_1$ and $\boldsymbol{T}_2$ are designed for testing the constant coefficients in GWR models and the zero coefficients in mixed GWR models. All the cases of the experimental settings are listed in Table 4. Under the null hypotheses of $\boldsymbol{T}_1$ and $\boldsymbol{T}_2$ for each of the experimental settings, we compute

Table 4. The null and alternative hypothesis of the stages T_1 and T_2 in the Models (I), (II) and (III).

Model	Stage	H_0	H_1
(I)	T_1	$\beta_j(u,v)=\beta_j$, $j=3,4,5,6,7,8$.	All coefficients vary over the space.
	T_2	$\beta_j(u,v)=\beta_j$, $j=3,4,5,6,7,8$ and $\beta_j=0$, $j=5,6,7,8$.	$\beta_j(u,v)=\beta$, $j=3,4,5,6,7,8$.
(II)	T_1	$\beta_j(u,v)=\beta_j$, $j=3,4,5,6,7,8$.	All coefficients vary over the space.
	T_2	$\beta_j(u,v)=\beta_j$, $j=3,4,5,6,7,8$ and $\beta_j=0$, $j=6,7,8$.	$\beta_j(u,v)=\beta$, $j=3,4,5,6,7,8$.
(III)	T_1	$\beta_j(u,v)=\beta_j$, $j=4,5,6,7,8$.	All coefficients vary over the space.
	T_2	$\beta_j(u,v)=\beta_j$, $j=4,5,6,7,8$ and $\beta_j=0$, $j=7,8$.	$\beta_j(u,v)=\beta$, $j=4,5,6,7,8$.

Table 5. Rejection rates of $N=200$ replications of the testing procedure under the significance level of .05.

		$\alpha=1$			$\alpha=0.5$		
Model	Stage	$\rho=0$	$\rho=0.5$	$\rho=0.9$	$\rho=0$	$\rho=0.5$	$\rho=0.9$
(I)	T_1	0.000	0.015	0.020	0.040	0.015	0.020
	T_2	0.015	0.010	0.010	0.025	0.010	0.010
(II)	T_1	0.010	0.010	0.020	0.010	0.010	0.010
	T_2	0.030	0.005	0.020	0.010	0.005	0.020
(III)	T_1	0.000	0.015	0.000	0.010	0.015	0.005
	T_2	0.025	0.025	0.010	0.000	0.010	0.010

the rate of rejecting the null hypothesis among 200 replications at the significance level of .05. The rejection rate is an estimator of the type I error of the test under the given setting and significance level.

Then, the related rejection rate was calculated for both T_1 and T_2 and the results are reported in Table 5. It can be observed from Table 5 that the bootstrap test performs well to detect the constant coefficients in GWR models and identify the zero constant coefficients in mixed GWR models when the null hypotheses and the alternative of T_1 and T_2 can be accurately known in advance. The rejection rates under the null hypothesis are reasonable smaller than the pre-specific significance level in all of the experimental settings, indicating that the bootstrap test yields a small value of the type I error. Additionally, although the collinearity among the variables may lead to spurious correlation between the GWR estimator of the varying coefficients, which is discussed in Mei et al. [9] and further explored in Fotheringham and Oshan [21], it does not perform significant impact on the rejection rates, which suggests that the bootstrap test is rather robust to do structure identification and variable selection in the presence of the collinearity among the explanatory variables.

4.4. Comparison and discussion

In summary, the simulation study demonstrates that both the GWGlasso method and the residual-based bootstrap tests perform well for structure identification and variable selection in GWR models. As one of the important statistical inference method in GWR literature, the bootstrap tests are applicable to the case that the expressions of the null hypotheses of T_1 and T_2 are provided necessarily for the given data. In practice, however, there is generally not enough priori information for the analysts to know that which coefficients should be chosen to be tested for zero, constant or varying. Intuitively, all possible combinations of the coefficients should be considered and a series of the tests should be performed, which is not an easy task especially when the number of the explanatory variables is large. Although Mei et al. [9] (Section 3.3) have introduced a simplified algorithm of the bootstrap test to compute the p-value of the test statistic for reducing the computational complexity, it still involves an exhaustive search over 2^{p+1} candidate models to identify the varying, constant and zero coefficients consistently, which is quite demanding when p is large. Furthermore, how to create the selection criterion and choose the best model among all possible candidate models is still necessary to be further investigated. From a computational point of view, the GWGlasso method dominates the residual-based bootstrap test of Mei et al. [9]. Interestingly, the bootstrap test is much more robust than the GWGlasso method under the impact of the collinearity. Nevertheless, the GWGlasso method is the primary shrinkage method to tackle the structure identification and variable selection problem in GWR models, which has the notable improvement in computation and explanation.

5. Analysis of Dublin voter turnout data set

To further illustrate the usefulness of the GWGlasso algorithm, we apply the method to the Dublin voter turnout data set, which has been analysed by Kavanagh et al. [22] and Gollini et al. [23], and is publicly available in the R package called GWmodel. The data set includes the nine percentage variables which measures the voter turnout in the Irish 2004 Dáil elections and eight characteristics of social structures in 322 Electoral Divisions of Greater Dublin. Following Gollini et al. [23], we take GEI (the proportion of the electorate who turned out on voting night to cast their vote in the 2004 General Election in Ireland) as the response variable, the geographical locations (u, v) as the index variable, and the following variables as the explanatory variables:

- MDA: one year migrants, that is, moved to a different address one year ago;
- LAR: local authority renters;
- SCO: social class one;
- UEP: unemployed;
- LOE: without any formal educational;
- AGY: age group 18–24;
- AGM: age group 25–44;
- AGO: age group 45–64.

The eight explanatory variables reflect measures of migration, public housing, high social class, unemployment, educational attainment and three adult age groups. As pointed out in Gollini et al. [23], although the eight explanatory variables measured on the same scale, the variables are not of a similar magnitude. Therefore, before applying our method, all the explanatory variables are transformed so that their marginal distribution is approximately $N(0, 1)$. Moreover, the normalized procedure here is same to other variable selection methods, such as Wang and Xia [11]; Hu and Xia [14] and Ma and Zhang [17]. Using the Dublin voter turnout data set, we consider the following GWR model

$$\begin{aligned} GEI = {} & \beta_0(u,v) + \beta_1(u,v)\mathrm{MDA} + \beta_2(u,v)\mathrm{LAR} + \beta_3(u,v)\mathrm{SCO} \\ & + \beta_4(u,v)\mathrm{UEP} + \beta_5(u,v)\mathrm{LOE} + \beta_6(u,v)\mathrm{AGY} \\ & + \beta_7(u,v)\mathrm{AGM} + \beta_8(u,v)\mathrm{AGO} + \varepsilon. \end{aligned} \tag{40}$$

Based on the data set, the GWR model is calibrated by the local-linear estimation to capture the spatial variations of coefficients estimates. Here, the Gaussian kernel is used and the bandwidth size is selected by the cross-validation method without penalization defined by Equation (12). The optimal bandwidth size of local-linear estimation for the above GWR model is $h_o = 0.78$ of 10 km (A scale may indicate that 1 m equals 10 km), which is selected from $\mathcal{H} = \{h_m : h_m = 0.7 + 0.02m\}_{m=1}^{20}$. Then the local-linear estimation is used to provide the initial coefficients estimates and the adaptive weights in the global loss function in Equation (21). Similar preliminary processing has been extensively applied by many researchers, such as Wang and Xia [11]; Hu and Xia [14] and Ma and Zhang [17].

By the judging threshold defined in Section 3.1.3, it is evident that the judging threshold δ is vital in determining whether the coefficient estimates is zero, nonzero constant or varying over space. Therefore, it is necessary to investigate several levels of the judging thresholds in analysing the structure information in the above GWR model. With the stopping criterion $\tau = 10^{-4}$, we consider several levels of the judging thresholds to be $\delta = 0.1,\ 0.05,\ 0.01,\ 0.005,\ 0.001$. Then, the proposed algorithm is implemented, the optimal penalized parameter is selected by the BIC criterion in Equation (32). Concretely, we select the optimal penalized parameter from the following set $\Lambda = \{\lambda_l : \lambda_l = 0.2 \times l\}_{l=1}^{20}$. Although the optimal penalization parameter selected from Λ may be a local minima, the role played by the range of Λ is rather limited, which can be verified by results of the simulation study. Then the identification results are listed in Table 6.

In can be obtained from Table 6 that the large size δ leads to identify more zero coefficients, whereas the small value δ tends to choose more varying coefficients. Nevertheless, the identification results of $\delta = 0.01$ are likely to be persuasiveness. The resulting GWGlasso estimation with the threshold $\delta = 0.01$ suggests that INT, MDA, LAR, SCO, UEP, AGY, AGM and AGO are all relevant variables, whereas LOE is not. Furthermore, they all suggest that the coefficient functions of INT, MDA, LAR, SCO, UEP, LOE, AGM and AGO are spatially varying coefficients, and the coefficient of AGY is constant. By deleting the irrelevant variable and reordering the spatially varying coefficients and the constant coefficients, we obtained a mixed GWR model for Dublin voter turnout data

$$\begin{aligned}\mathrm{GEI} = {} & \beta_0(u,v) + \beta_1(u,v)\mathrm{MDA} + \beta_2(u,v)\mathrm{LAR} + \beta_3(u,v)\mathrm{SCO} \\ & + \beta_4(u,v)\mathrm{UEP} + \beta_5(u,v)\mathrm{AGM} + \beta_6(u,v)\mathrm{AGO} + \beta_7\mathrm{AGY} + \varepsilon. \end{aligned} \tag{41}$$

The Dublin voter turnout data set has been analysed for fitting GWR model and addressing the local collinearity problem in Gollini et al. [23, Section 7, 24–36]. For the local collinearity issues, penalized regression methods (such as geographically weighted ridge regression model [24] the geographically weighted lasso [4]) have been proposed to shrink regression coefficients to alleviate the effect of collinearity. As a penalized method, GWGlasso also has the ability to shrink the spatially varying coefficients to zero. To illustrate the shrinkage effect of GWGlasso, we present a comparison of all the coefficients estimates between the local-linear estimation and the GWGlasso estimation for the GWR model in (40). Moreover, the boxplots of the coefficients estimates of all the variables based on the unpenalized local-linear estimation and the GWGlasso estimation are illustrated in Figure 1. The constant coefficient estimates of AGY is −0.9750. Figure 1 shows that the GWGlasso method has a shrinkage effect on the coefficient estimates. Furthermore, the GWGlasso method shrinks the coefficient estimation of LOE to zero and identifies the coefficient estimation of AGY to be constant. It can also shrink the varying coefficients estimation to be a small scope.

To verify whether the coefficients are varying or not, we can provide some evidences from Table 6 and Figure 1. If the judging threshold is set to be large values $\delta = 0.1, 0.05$, MDA, LAR, LOE and AGY are irrelevant variables demonstrated in Table 6, which can be verified that the median of the coefficients estimates of these variables are much smaller than that of the other variables in Figure 1. When the judging threshold is set to be the value $\delta = 0.01, 0.005$, INT, MDA, LAR, SCO, UEP, AGY, AGM

Table 6. The structure identification and variable selection results with the GWR model of the Dublin voter turnout data.

δ	INT	MDA	LAR	SCO	UEP	LOE	AGY	AGM	AGO
0.1	V	Z	Z	V	C	Z	Z	V	V
0.05	V	Z	Z	V	C	Z	Z	V	V
0.01	V	V	V	V	V	Z	C	V	V
0.005	V	V	V	V	V	Z	C	V	V
0.001	V	V	V	V	V	V	C	V	V

Note: The coefficient of explanatory variable identified to be spatially varying (V), nonzero constant (C) or zero (Z). INT is the intercept.

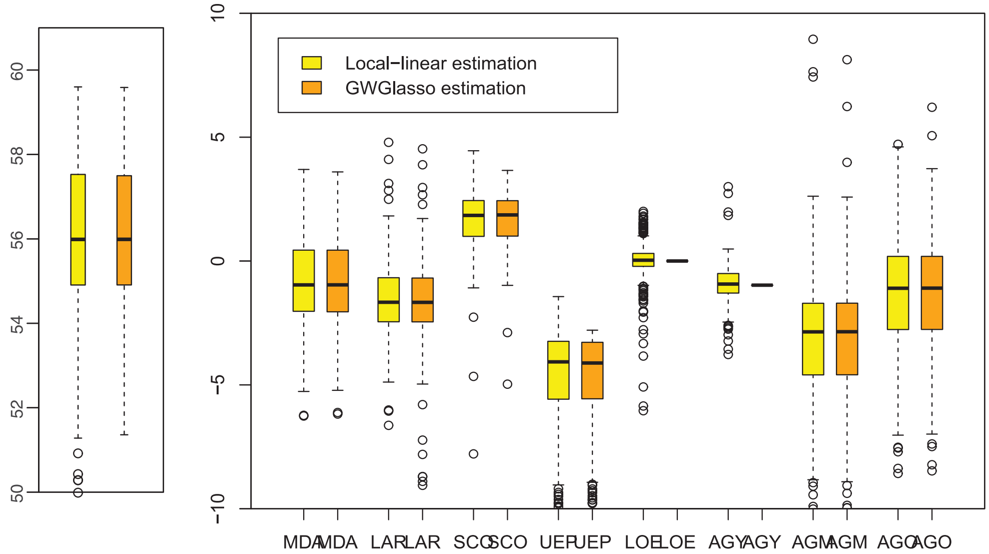

Figure 1. The boxplots of the spatially varying coefficients estimation based on the local-linear estimation and the GWGlasso estimation. The boxplots of intercept are illustrated left, and the boxplots of the coefficients of all the variables are depicted right.

and AGO are relevant explanatory variables. Figure 1 depicts that the medians of coefficients estimation of LOE is likely to be zero and the scope of LOE and AGY are relative small compared with that of other variables. It can be obtained from case $\delta = 0.001$ in Table 6, all the explanatory variables are relevant to the response and only AGY has constant effect, which indicates that the coefficients estimations of all the variables are likely to be spatially varying coefficients when the judging threshold is set to be small. These findings corroborate the identification results of the GWGlasso method very well.

In addition, to further understand the impact of the choice of the different kernels on the identification results, we have conducted the GWGlasso method with other kernels for the Dublin voter turnout data analysis. Among various kernels, the Bi-square kernel $K(t) = [(1 - t^2)_+]^2$ and the Epanechnikov kernel $K(t) = 0.75(1 - t^2)_+$ are frequently used in application. Respectively, the optimal bandwidth sizes of the local-linear estimation for the GWR model (40) with the Bi-square kernel and the Epanechnikov kernel are $h_o = 9$ and $h_o = 12$ of 10 km, which are selected from $\mathcal{H} = \{h_m : h_m = 5 + 0.5m\}_{m=1}^{20}$ by the CV criterion. With the stopping criterion $\tau = 10^{-4}$, several levels of the judging thresholds are set to be $\delta = 1, 0.5, 0.1, 0.05, 0.01$. For both the Bi-square kernel and the Epanechnikov kernel, the resulting GWGlasso estimation with these two kernels are similar. And the identification results of $\delta = 0.5$ are likely to be reasonable, which both suggest that the coefficient functions of INT, MDA, LAR, SCO, UEP, LOE, AGM and AGO are spatially varying coefficients, and the coefficients of LOE and AGY are zero constant.

Compared with the coefficient of AGY identified to be nonzero constant for the Gaussian kernel, the trivial difference of the resulting GWGlasso estimation with the Bi-square kernel and the Epanechnikov kernel is that the coefficient of AGY identified with these kernels is zero. Furthermore, although the different kernels may lead to the differences among the local-linear GWR estimators of varying coefficients with the different optimal bandwidth sizes, it does not bring about obvious influence on the identification results, which indicates that the GWGlasso method is rather robust to the choice of different kernels. Notice that the optimal judging threshold δ is intimate with the kernel function and should be selected carefully. Because of the limited space, the detailed results are omitted here, but those identification results are attached as a supplementary material. In summary, the Dublin voter turnout data analysis further confirms that the GWGlasso method is a promising shrinkage method for structure identification and variable selection in GWR models.

6. Conclusions

Variable selection and structure identification of the GWR model is important for exploring spatial non-stationarity in geo-referenced data analysis. As a local fitting technique, however, the GWR estimator of each coefficient varies with the local location regardless of whether the hidden coefficient is zero, nonzero constants or varies over the space. Therefore, identifying the spatially coefficients in a GWR model is vital to validly explain spatially non-stationarity of the regression relationship. To achieve the goal, we propose the GWGlasso algorithm to identify zero, nonzero constant or varying coefficients in a GWR model. Numerical experiments and real data analysis indicated that the proposed method is very effective. Moreover, GWGlasso provides a useful way of building a possible mixed GWR model for a geo-referenced data set. Finally, the R code of the GWGlasso algorithm and the residual-based bootstrap tests used in Sections 4 and 5 is provided in the supplemental material.

To conclude the paper, we would like to discuss some possible topics for future study. Firstly, our proposal is based on the group lasso method due to its simplicity. Similar ideas can be extend to other useful shrinkage methods, such as the group SCAD method, the group MCP method and other group selection methods [25]. Secondly, it will it can profitably to explore the proposed method in generalized GWR models such as the geographically weighted Poisson regression [26], the geographically weighted logistic regression [27] and so on. Furthermore, how to do variable selection and structure identification for generalized GWR model is an interesting topic for future research.

Acknowledgments

The authors thank the Associate Editor and the referee for their many insightful and inspiring comments and suggestions which greatly improved the paper, and also thank Professor Jingxiao Zhang for beneficial discussions.

Disclosure statement

No potential conflict of interest was reported by the authors.

References

[1] Brunsdon CE, Fotheringham AS, Charlton ME. Geographically weighted regression: a method for exploring spatial non-stationarity. Geograph Anal. 1996;28(4):281–298.

[2] Fotheringham AS, Brunsdon CE, Charlton ME. Geographically weighted regression – the analysis of spatial varying relationships. Chichester: John Wiley; 2002.

[3] Brunsdon CE, Fotheringham AS, Charlton ME. Some notes on parametric significance tests for geographically weighted regression. J Regional Sci. 1999;39(3):497–524.

[4] Wheeler DC. Simultaneous coefficient penalization and model selection in geographically weighted regression: the geographically weighted lasso. Environ Plan A. 2009;41(3):722–742.

[5] Brunsdon CE, Fotheringham AS, Charlton ME. Geographically weighted regression: modelling spatial non-stationarity. The Statistician. 1998;47(3):431–443.

[6] Fotheringham AS, Charlton ME, Brunsdon CE. Geographically weighted regression: a natural evolution of the expansion method for spatial data analysis. Environ Plan A. 1998;30(1):1905–1927.

[7] Leung Y, Mei C-L, Zhang W-X. Testing for spatial autotregression among the residuals of the geographically weighted regression. Environ Plan A. 2000;32(1):871–890.

[8] Mei C-L, Wang N, Zhang W-X. Testing the importance of the expalanatory variables in a mixed geographically weighted regression model. Environ Plan A. 2006;38(3):587–598.

[9] Mei C-L, Xu M, Wang N. A bootstrap test for constant coefficients in geographically weighted regression models. Int J Geogr Inf Sci. 2016;30(8):1622–1643.

[10] Wang L, Li H, Huang J-Z. Variable selection in nonparametric varying coefficient models for analysis of repeated measurements. J Am Stat Assoc. 2008;103(484):1556–1569.

[11] Wang H-S, Xia Y-C. Shrinkage estimation of the varying coefficient model. J Am Stat Assoc. 2009;104(486): 747–757.

[12] Kai B, Li R, Zou H. New efficient estimation and variable selection methods for semiparametric varying-coefficient partially linear models. Ann Stat. 2011;39(1):305–332.

[13] Tang Y-L, Wang HJ, Zhu Z-Y, et al. A unified variable selection approach for varying coefficient models. Statist Sinica. 2012;22:601–628.

[14] Hu T, Xia Y-C. Adapative semi-varying coefficient model selection. Statist Sinica. 2012;22:575–599.

[15] Wang K-N, Lin L. Robust structure identification and variable selection in partial linear varying coefficient models. J Stat Plan Inference. 2016;174(2016):153–168.

[16] Fan J, Li R. Variable selection via nonconcave penalized likelihood and its oracle properties. J Am Stat Assoc. 2001;96(456):1348–1360.

[17] Ma X-J, Zhang J-X. A new variable selection approach for varying coefficient models. Metrika. 2016;79:59–72.

[18] Wang N, Mei C-L, Yan X-D. Local linear estimation of spatially varying coefficient models: an improvement on the geographically weighted regression technique. Environ Plan A. 2008;40(4):986–1005.

[19] Yuan M, Lin Y. Model selection and estimation in regression with grouped variables. J R Stat Soc B. 2006;68(1):49–67.

[20] Wang H-S, Leng C-L. A note on adaptive group lasso. Comput Stat Data Anal. 2008;52:5277–5286.

[21] Fotheringham AS, Oshan TM. Geographically weighted regression and multicollinearity: dispelling the myth. J Geogr Syst. 2016;18:303–329.

[22] Kavanagh A, Fotheringham AS, Charlton ME. A geographically weighted regression analysis of the election specific turnout behaviour in the Republic of Ireland. Paper presented at: In elections, public opinion and parties conference, Nottingham ; 2006 September 8–10.

[23] Gollini I, Lu B, Charlton ME, et al. GW model: an R package for exploring spatial heterogeneity using geographically weighted models. J Stat Softw. 2015;63(17):1–50.

[24] Wheeler DC, Calder CA. An assessment of coefficient accuracy in linear regression models with spatially varying coefficients. J Geogr Syst. 2007;9(2):145–166.

[25] Huang J, Breheny P, Ma S. A selective review of group selection in high-dimensional models. Stat Sci. 2012;4:481–499.

[26] Nakaya T, Fotheringham AS, Brunsdon CE, et al. Geographically weighted Poisson regression for disease association mapping. Stat Med. 2005;24:2695–2717.

[27] Atkinson PM, German SE, Sear DA, et al. Exploring the relations between riverbank erosion and geomorphological controls using geographically weighted logistic regression. Geograph Anal. 2003;35(1):58–82.