



Contents lists available at ScienceDirect

## Spatial Statistics

journal homepage: [www.elsevier.com/locate/spasta](http://www.elsevier.com/locate/spasta)



# Structure identification and model selection in geographically weighted quantile regression models



Wentao Wang<sup>a,\*</sup>, Shoufang Xu<sup>a,b</sup>, Tianshun Yan<sup>a</sup>

<sup>a</sup> Department of Statistics, School of Mathematics and Statistics, Xi'an Jiaotong University, Xi'an, Shaanxi, People's Republic of China

<sup>b</sup> College of Mathematics and Information Science, Xinxiang University, Xinxiang, Henan, People's Republic of China

## ARTICLE INFO

### Article history:

Received 28 December 2017

Accepted 20 May 2018

Available online 5 June 2018

### Keywords:

Geographical weighted quantile regression  
Spatial non-stationarity  
Variable selection  
Structure identification  
Adaptive group lasso  
Spatially varying coefficient models

## ABSTRACT

Geographical weighted quantile regression (GWQR) is an important tool for exploring the spatial non-stationarity of the regression relationship and providing the entire description of the response distribution, which is useful in practice because of its robustness against outliers and flexibility in dealing with non-normal distributions. This paper proposes the GWQlasso method for structure identification and variable selection in GWQR models. The proposed method combines the local-linear estimation of the GWQR model and the adaptive group lasso, which can simultaneously identify spatially varying coefficients, non-zero constant coefficients and zero coefficients. The selection consistency and the oracle property of the proposal are studied. Moreover, the derived algorithm for the GWQlasso method and the selection of the tuning parameter by the BIC criterion are established. Simulations and real examples are used to illustrate the method.

© 2018 Elsevier B.V. All rights reserved.

## 1. Introduction

Geographically weighted regression (GWR) (Brennan et al., 1996; Fotheringham et al., 2002) has become a very popular technique for locally modeling of spatial relationships by calibrating a spatially

\* Corresponding author.

E-mail address: [wttwang@hotmail.com](mailto:wttwang@hotmail.com) (W.T. Wang).

varying coefficient model of the form

$$Y_i = \sum_{j=1}^p \beta_j(u_i, v_i) X_{ij} + \epsilon_i, \quad i = 1, \dots, n \quad (1)$$

where  $(Y_i; X_{i1}, \dots, X_{ip})$  is the observation of the response variable  $Y$  and the explanatory variables  $X_1, \dots, X_p$  collected at the  $i$ th geographical location  $(u_i, v_i)$  in the study region,  $\beta_j(u, v)$  ( $j = 1, \dots, p$ ) are unknown coefficient functions of the spatial locations, and  $\epsilon_i$  is the random noise. The model (1) can include a spatially varying intercept by setting  $X_1 \equiv 1$ . Besides the GWR technique, the model (1) can also be calibrated by the Bayesian methods (Gamerman et al., 2003; Gelfand et al., 2003; Congdon, 2006, 2007). Except for many applications, most existing research about the GWR technique focuses on either parameter estimation (Brunsdon et al., 1996; Fotheringham et al., 2002; Wang et al., 2008; Shen et al., 2011; Yang et al., 2012; Lu et al., 2014, 2017; Fotheringham et al., 2017) or non-stationarity testing (Brunsdon et al., 1996, 1999; Leung et al., 2000; Páez et al., 2002; Mei et al., 2006, 2016; Harris et al., 2017). Methodologically, the GWR technique is inherently not robust to outliers and inefficient for many non-normal errors due to the least square-based mean regression in the process of parameter estimation (Zhang and Mei, 2011).

As an important extension to mean regression, quantile regression (Koenker and Bassett, 1978; Koenker, 2005) is widely used in application because of its robustness against outliers and flexibility in dealing with error distribution. Quantile regression for the spatially varying coefficient model (1) has been studied in the Bayesian method and the geographically weighted quantile regression (GWQR) method. In Bayesian approach, Reich et al. (2011) proposed Bayesian spatial quantile regression by specifying a semiparametric model for the entire quantile process. Yang and He (2015) considered Bayesian spatial quantile regression using empirical likelihood as a working likelihood for spatially correlated data. In GWQR approach, Hallin et al. (2009) proposed a local linear spatial quantile regression approach. Zhang and Mei (2011) considered geographical weighted median regression based on the least absolute deviation and further proposed two robust versions of the GWR approaches to handle outliers. Chen et al. (2012) proposed geographically weighted quantile regression (GWQR) for spatial quantile-based regression analysis and introduced a potential approach for testing the spatial non-stationarity. Following Chen et al. (2012), we refer to the spatially varying coefficient quantile regression model as the GWQR model.

However, much less has been done about the structure identification and variable selection for the GWQR model. As a local modeling technique, the GWQR estimator of the each coefficient varies with the focal location regardless of whether the underlying coefficient is zero, non-zero constant or varies over the space. In some practical problems, some coefficients in the GWQR model are assumed to be constant while the others are allowed to vary across the region (Chen et al., 2012). In addition, including irrelevant variables in the GWQR model will degrade the estimation efficiency. To avoid a misleading explanation on non-stationarity of the underlying regression relationship, it is essential to know which coefficients really vary over space and which coefficients are non-zero constants or zero.

In statistical learning literature, various shrinkage methods have been developed in quantile varying coefficient models (QVCM) for variable selection and model specification. Kai et al. (2011) considered the variable selection for the varying coefficient composite quantile regression models. Cai and Xiao (2012) proposed semiparametric quantile regression estimation and derived a constancy test in dynamic models. Hohsuk et al. (2012) used the one-step group SCAD method to conduct variable selection for QVCM. Tang et al. (2013) developed the variable selection procedure for QVCM with longitudinal data. Zhao et al. (2014) considered model selection for QVCM with categorical effect modifiers. Yang and He (2015) considered weighted composite quantile regression estimation and variable selection for QVCM. With local-linear quantile regression method and the adaptive group lasso method, Shen and Liang (2016) proposed variable selection for QVCM with censoring indicators missing at random. In this method, the values of each coefficient 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 shrink toward to zero. As a result, the nonzero and zero coefficients can be simultaneously identified according to the penalized estimates of the coefficients.

Motivated by the methodology in [Shen and Liang \(2016\)](#), the idea of shrinkage estimation is applied to the GWQR model in order to identify spatially varying, non-zero constant and zero coefficients. Based on the local-linear estimation of the GWQR model ([Chen et al., 2012](#)), we propose a shrinkage method so-called GWQlasso to simultaneously conduct variable selection and structure identification for the GWQR models. In the GWQlasso method, the values of the regression function and its partial derivatives are respectively grouped and the adaptive group lasso ([Yuan and Lin, 2006; Wang and Li, 2017](#)) is conducted to penalize the residual sum of squares of the local-linear estimation of the GWQR model. Additionally, the asymptotic analysis for the GWQlasso estimator is established to show that the GWQlasso method is able to identify the true model consistently and the estimated quantile varying coefficients retain the optimal convergence rate of nonparametric estimators. Moreover, the approximated iterative algorithm for the GWQlasso method has been developed to overcome the inherent difficulty brought by the nondifferentiability of the check loss function and the penalty function at the origin.

The rest of this paper is organized as follows. In Section 2, the local-linear estimation for the GWQR model is briefly described to facilitate the subsequent exposition. In Section 3, we propose the GWQlasso method and study its theoretical properties. The algorithm of the GWQlasso method and the selection of the tuning parameter by the BIC criterion are established in Section 4. Moreover, simulation experiments and Dublin voter turnout data analysis are presented in Section 5. Proofs of the theorems are provided in the supplemental materials.

## 2. Geographically weighted quantile regression model

### 2.1. Model and notations

Suppose that the observed spatial data set is

$$\mathbf{D}_n = \left\{ (Y_i, X_{1i}, \dots, X_{pi}; (u_i, v_i)), i = 1, \dots, n \right\},$$

where  $Y_i$  and  $\mathbf{X}_i^T = [X_{1i}, \dots, X_{pi}]$  are respectively the observations of the response variable  $Y$  and those of the explanatory variables  $X_1, \dots, X_p$  at the geographical location  $(u_i, v_i)$  ( $i = 1, \dots, n$ ). The geographically weighted quantile regression (GWQR) model with underlying true structure has the following form:

$$Y_i = \sum_{j \in \mathcal{A}_V} X_{ij} \beta_{\tau,j}(u_i, v_i) + \sum_{j \in \mathcal{A}_C} X_{ij} \beta_{\tau,j} + \sum_{j \in \mathcal{A}_Z} X_{ij} \theta_{\tau,j}(u_i, v_i) + \epsilon_{\tau,i}, \quad (2)$$

where  $\epsilon_{\tau,i}$  is the random error term with the conditional  $\tau$ th quantile being zero, and unknown sets  $\mathcal{A}_V, \mathcal{A}_C$  and  $\mathcal{A}_Z$  are the index sets for spatially varying effects, nonzero constant effects and zero effects, respectively. The index sets are mutually exclusive and satisfy  $\mathcal{A}_V \cup \mathcal{A}_C \cup \mathcal{A}_Z = \{1, \dots, p\}$ . Therefore, given data set  $\mathbf{D}_n$  and quantile level  $\tau$ , our main goal is to identify the index sets  $\mathcal{A}_V, \mathcal{A}_C$  and  $\mathcal{A}_Z$  as well as estimate the spatially varying coefficients  $\beta_{\tau,j}(u_i, v_i), j \in \mathcal{A}_V$  and non-zero constant coefficients  $\beta_{\tau,j}, j \in \mathcal{A}_C$  accurately.

### 2.2. Local-linear GWQR

The GWQR model (2) can be embedded into the following model

$$Q_\tau(\mathbf{X}_i; (u_i, v_i)) = X_{i1} \beta_{\tau,1}(u_i, v_i) + X_{i2} \beta_{\tau,2}(u_i, v_i) + \dots + X_{ip} \beta_{\tau,p}(u_i, v_i) = \mathbf{X}_i^T \boldsymbol{\beta}_\tau(u_i, v_i) \quad (3)$$

where  $Q_\tau(\mathbf{X}_i; (u_i, v_i)) \equiv \inf\{Y : F(Y | (\mathbf{X}_i; (u_i, v_i))) \geq \tau\}$  is the  $\tau$ th conditional quantile given  $\mathbf{X}_i$ . Assume that the coefficients vector  $\boldsymbol{\beta}_\tau(u, v) = [\beta_{\tau,1}(u, v), \dots, \beta_{\tau,p}(u, v)]^T \in \mathbb{R}^p$  has second continuous partial derivatives with respect to  $u$  and  $v$ . By Taylor's expansion,  $\boldsymbol{\beta}_\tau(u, v)$  can be approximated in the neighborhood of  $(u_0, v_0)$  by

$$\boldsymbol{\beta}_\tau(u, v) \approx \boldsymbol{\beta}_\tau(u_0, v_0) + \boldsymbol{\beta}_\tau^{(u)}(u_0, v_0)(u - u_0) + \boldsymbol{\beta}_\tau^{(v)}(u_0, v_0)(v - v_0),$$

where  $\beta_\tau^{(u)}(u_0, v_0)$  and  $\beta_\tau^{(v)}(u_0, v_0)$  are the partial derivatives vectors of  $\beta_\tau(u, v)$  at  $(u_0, v_0)$  with respect to  $u$  and  $v$ , respectively.

Let  $\rho_\tau(z) = z(\tau - I(z < 0))$  be the check loss function at quantile  $\tau \in (0, 1)$ , where  $I(\cdot)$  is the indicator function. For a focal location  $(u_t, v_t)$ , let  $d_{it} = \|(u_i, v_i) - (u_t, v_t)\|$ , where  $\|\cdot\|$  stands for the usual Euclidean norm. According to Chen et al. (2012), the local-linear GWQR estimates of the coefficients and their partial derivatives are the minimizers of the local weighted quantile loss function

$$\mathcal{L}_h(u_t, v_t) = \sum_{i=1}^n \rho_\tau \left\{ Y_i - \mathbf{X}_i^T [\beta_\tau(u_t, v_t) + \beta_\tau^{(u)}(u_t, v_t)(u_t - u_i) + \beta_\tau^{(v)}(u_t, v_t)(v_t - v_i)] \right\} K_h(d_{it}) \quad (4)$$

with respect to  $\beta_\tau(u_t, v_t)$ ,  $\beta_\tau^{(u)}(u_t, v_t)$  and  $\beta_\tau^{(v)}(u_t, v_t)$  for a given kernel function  $K_h(\cdot) = K(\cdot/h)/h^2$  and the bandwidth  $h$ . In what follows, we omit the subscript  $\tau$  in  $\beta_\tau(u_t, v_t)$ ,  $\beta_\tau^{(u)}(u_t, v_t)$  and  $\beta_\tau^{(v)}(u_t, v_t)$ , and they are denoted as  $\beta(u_t, v_t)$ ,  $\beta^{(u)}(u_t, v_t)$  and  $\beta^{(v)}(u_t, v_t)$ , respectively.

Following Fotheringham et al. (2002) and Chen et al. (2012), the cross-validation (CV) procedure is generally used to select the optimal bandwidth. The optimal value of the bandwidth  $h$  is selected by minimizing the following CV score

$$CV(h) = \frac{1}{n} \sum_{i=1}^n \rho_\tau \left\{ Y_i - Q_{\tau,(-i)}(\mathbf{X}_i; (u_i, v_i)) \right\}, \quad (5)$$

where  $Q_{\tau,(-i)}(\mathbf{X}_i; (u_i, v_i))$  is the leave-one-out version estimate of  $Q_\tau(\mathbf{X}_i; (u_i, v_i))$  obtained with the  $i$ th observation  $(Y_i, \mathbf{X}_i; (u_i, v_i))$  being omitted during estimation. Compared with the CV procedure in basic GWR model, the CV score in (5) replaces the quadratic loss function with the check loss function. The optimal bandwidth  $h$  is chosen as  $h_o = \arg \min_{h>0} CV(h)$ .

**Remark 1.** The estimates of the regression coefficients in the local-linear GWQR method can be solved by simplex method or interior point method, or linear programming method. An efficient way to solve the task is using **rq** function of the R package **quantreg** (Koenker, 2017). Specifically, for a given location  $(u_t, v_t)$ , let  $\mathbf{X} = [\mathbf{X}_1, \dots, \mathbf{X}_n]^T$ ,  $\mathbf{Y} = [Y_1, \dots, Y_n]^T$ ,

$$\mathbf{U}(t) = \text{diag}[u_1 - u_t, \dots, u_n - u_t], \quad \mathbf{V}(t) = \text{diag}[v_1 - v_t, \dots, v_n - v_t],$$

$$\mathbf{W}_t = [K_h(d_{t1}), \dots, K_h(d_{tn})]^T, \quad \mathbf{X}_t = [\mathbf{X}, \mathbf{U}(t)\mathbf{X}, \mathbf{V}(t)\mathbf{X}].$$

For the quantile level  $\tau$ , we recall the **quantreg** package by setting in the **rq** function with the format  $\text{rq}(\mathbf{Y} \sim \mathbf{X}_t, tau = \tau, weights = \mathbf{W}_t)$ . By taking  $(u_t, v_t)$  to be all of the locations  $\{(u_t, v_t)\}_{t=1}^n$ , we have  $\{\hat{\beta}(u_t, v_t), \hat{\beta}^{(u)}(u_t, v_t), \hat{\beta}^{(v)}(u_t, v_t)\}_{t=1}^n$ . Note that the intercept term is included in the default setting. It is necessary to delete the intercept term when the design matrix  $\mathbf{X}_t$  includes the intercept. In the statistical literature, the GWQR can be classified in the locally polynomial quantile regression (Koenker, 2005, pp. 222); in general implemented by the **lprq** function, where the key procedure of the **lprq** function is the **rq** function. Based on the **lprq** function, we recall the **rq** function to solve the estimation problem in Eq. (4) and obtain the estimation of the coefficients and their derivatives.

### 3. The GWQlasso method

#### 3.1. Penalized estimation

For ease of exposition, let

$$\mathbf{A} = [\beta(u_1, v_1), \dots, \beta(u_n, v_n)]^T = [\mathbf{a}_c(1), \dots, \mathbf{a}_c(p)], \quad (6)$$

where the  $i$ th row  $\beta^T(u_i, v_i)$  of  $\mathbf{A}$  is the values of the  $p$  coefficients in the GWQR model (3) and the  $j$ th column

$$\mathbf{a}_c(j) = [\beta_j(u_1, v_1), \dots, \beta_j(u_n, v_n)]^T$$

consists of the  $j$ th coefficient at the  $n$  designated locations. Let

$$\begin{aligned}\mathbf{B} &= \left[ \boldsymbol{\beta}^{(u)}(u_1, v_1), \dots, \boldsymbol{\beta}^{(u)}(u_n, v_n), \boldsymbol{\beta}^{(v)}(u_1, v_1), \dots, \boldsymbol{\beta}^{(v)}(u_n, v_n) \right]^T \\ &= [\mathbf{b}_c(1), \dots, \mathbf{b}_c(p)]\end{aligned}\quad (7)$$

where the  $i$ th row  $[\boldsymbol{\beta}^{(u)}(u_i, v_i)]^T$  and the  $(n+i)$ th row  $[\boldsymbol{\beta}^{(v)}(u_i, v_i)]^T$  of  $\mathbf{B}$  are the values of the partial derivatives of the  $p$  coefficients at location  $(u_i, v_i)$  with respect to  $u$  and  $v$ , respectively, and the  $j$ th column

$$\mathbf{b}_c(j) = [\beta_j^{(u)}(u_1, v_1), \dots, \beta_j^{(u)}(u_n, v_n), \beta_j^{(v)}(u_1, v_1), \dots, \beta_j^{(v)}(u_n, v_n)]^T$$

includes the two partial derivatives of  $j$ th coefficient at the  $n$  designated locations.

We define the unpenalized estimates of  $[\boldsymbol{\beta}(u_t, v_t)]^T$  and  $[[\boldsymbol{\beta}^{(u)}(u_t, v_t)]^T, [\boldsymbol{\beta}^{(v)}(u_t, v_t)]^T]$  as the  $t$ th row of  $\widehat{\mathbf{A}}^{[0]}$  and  $\widehat{\mathbf{B}}^{[0]}$ , which is obtained by the local-linear GWQR estimates with the optimal bandwidth  $h_0$ . By taking  $(u_t, v_t)$  to be all of the locations  $\{(u_t, v_t)\}_{t=1}^n$ , we have the initial estimates of  $\mathbf{A}$  and  $\mathbf{B}$  as  $\widehat{\mathbf{A}}^{[0]}$  and  $\widehat{\mathbf{B}}^{[0]}$ . Thus, the  $j$ th column of  $\widehat{\mathbf{A}}^{[0]}$  and  $\widehat{\mathbf{B}}^{[0]}$  is analogous to the estimates of  $\mathbf{a}_c(j)$  and  $\mathbf{b}_c(j)$ , i.e.,  $\widehat{\mathbf{a}}_c^{[0]}(j)$  and  $\widehat{\mathbf{b}}_c^{[0]}(j)$ .

Following the group lasso idea of Yuan and Lin (2006), we propose the penalized loss function

$$\mathcal{L}_{h,\lambda}(\mathbf{A}, \mathbf{B}) = \sum_{t=1}^n \mathcal{L}_h(u_t, v_t) + \sum_{j=1}^p (\lambda_{1j} \|\mathbf{a}_c(j)\| + \lambda_{2j} \|\mathbf{b}_c(j)\|), \quad (8)$$

where  $\lambda_1 = (\lambda_{11}, \dots, \lambda_{1p})^T \in \mathbb{R}^p$  and  $\lambda_2 = (\lambda_{21}, \dots, \lambda_{2p})^T \in \mathbb{R}^p$  are the tuning parameters, and  $\|\cdot\|$  is the usual Euclidean norm, i.e.,

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

In the objective function  $\mathcal{L}_{h,\lambda}(\mathbf{A}, \mathbf{B})$ , the penalty terms  $\sum_{j=1}^p \lambda_{1j} \|\mathbf{a}_c(j)\|$  and  $\sum_{j=1}^p \lambda_{2j} \|\mathbf{b}_c(j)\|$  are used to constrict the values of each coefficient and their two partial derivatives at all of the designed locations towards zero. Thus, if the coefficient  $\beta_j(u, v)$  is zero over space, the estimates of both  $\mathbf{a}_c(j)$  and  $\mathbf{b}_c(j)$  should be shrunk to zero; if  $\beta_j(u, v)$  is non-zero constant, the estimates of only  $\mathbf{b}_c(j)$  should be shrunk to zero. Furthermore, the estimated varying effects, nonzero constant effects and zero effects index sets are respectively,

$$\begin{aligned}\widehat{\mathcal{A}}_V &= \{j : \widehat{\mathbf{b}}_c(j) \neq \mathbf{0}, j = 1, \dots, p\}, \\ \widehat{\mathcal{A}}_C &= \{j : \widehat{\mathbf{a}}_c(j) \neq \mathbf{0}, \widehat{\mathbf{b}}_c(j) = \mathbf{0}, j = 1, \dots, p\}, \\ \widehat{\mathcal{A}}_Z &= \{j : \widehat{\mathbf{a}}_c(j) = \mathbf{0}, \widehat{\mathbf{b}}_c(j) = \mathbf{0}, j = 1, \dots, p\}.\end{aligned}\quad (9)$$

Grounded on this motivation, the shrinkage estimates of  $\mathbf{a}_c(j)$  and  $\mathbf{b}_c(j)$  from  $\mathcal{L}_{h,\lambda}(\mathbf{A}, \mathbf{B})$ , provide the evidence for identifying whether  $\beta_j(u, v)$  is zero, non-zero constant, or spatially varying coefficient. Because the estimator of coefficients and their derivatives are obtained based on a combination of GWQR technique and the lasso method, we refer to it as a geographically weighted quantile lasso (GWQlasso) estimate.

### 3.2. Theoretical properties

To study the selection consistency and the oracle property of the GWQlasso method, we define  $a_{1n} = \max\{\lambda_{1j} : 1 \leq j \leq p_1\}$ ,  $a_{2n} = \min\{\lambda_{1j} : p_1 < j \leq p\}$ ,  $b_{1n} = \max\{\lambda_{2j} : 1 \leq j \leq p_0\}$ , and  $b_{2n} = \min\{\lambda_{2j} : p_0 < j \leq p\}$ . Assume that  $\boldsymbol{\beta}_{\mathcal{A}_V}(u, v) = [\beta_1(u, v), \dots, \beta_{p_0}(u, v)]^T$ ,  $\boldsymbol{\beta}_{\mathcal{A}_C}(u, v) = [\beta_{p_0+1}, \dots, \beta_{p_1}]^T$ ,  $\boldsymbol{\beta}_{\mathcal{A}_Z}(u, v) = [\beta_{p_1+1}, \dots, \beta_p]^T$  are the spatially varying coefficients, the non-zero constant coefficients and the zero coefficients, respectively. We define  $\mathbf{X}_{i,\mathcal{A}_V} = [X_{i,1}, \dots, X_{i,p_0}]^T \in \mathbb{R}^{p_0}$ ,  $\mathbf{X}_{i,\mathcal{A}_C} = [X_{i,p_0+1}, \dots, X_{i,p_1}]^T \in \mathbb{R}^{p_1-p_0}$ ,  $\mathbf{X}_{i,\mathcal{A}_Z} = [X_{i,p_1+1}, \dots, X_{i,p}]^T \in \mathbb{R}^{p-p_1}$ . Let  $f_\epsilon(\cdot | u, v, \mathbf{X})$ ,  $F_\epsilon(\cdot | u, v, \mathbf{X})$  be the density function and cumulative distribution function of the error conditional on

$(U, V, X) = (u, v, \mathbf{X})$ , respectively. Moreover, we list the following regularity conditions which are needed to prove asymptotic properties.

(C1) For a  $\theta > 2$ ,  $E|Y_i|^{2\theta} < \infty$  and  $E\|X_{ij}\|^{2\theta} < \infty$ .

(C2) The kernel function  $K(\cdot)$  is a bounded positive, symmetric and Lipschitz continuous function with a compact support on  $\mathbb{R}$ .

(C3) The spatially varying coefficients  $\{\beta_j(\cdot), j = 1, \dots, p\}$  have continuous second partial derivatives.

(C4) The geographical location  $\{(u_i, v_i), i = 1, \dots, n\}$  is a sequence of fixed design points on a bounded compact support  $\mathcal{S}$ . For any vector  $\mathbf{s} = (\mu, v)^T \in \mathbb{R}^2$ , let  $d\mathbf{s} = d\mu dv$ , there exists a positive joint density function  $f(u, v)$  satisfying a Lipschitz condition such that

$$\sup_{(u, v) \in \mathcal{S}} \left| \frac{1}{n} \sum_{i=1}^n [\varphi(u_i, v_i) K_h(\|(u_i, v_i)^T - (u, v)^T\|)] - \int_{\mathbb{R}^2} \varphi(\mathbf{s}) K_h(\|\mathbf{s} - (u, v)^T\|) f(\mathbf{s}) d\mathbf{s} \right| = O(h)$$

for any bounded continuous function  $\varphi(\cdot)$  and  $K_h(\cdot) = K(\cdot/h)/h^2$ .  $f(u, v)$  is bounded away from zero on  $\mathcal{S}$ .

(C5) Let  $\mathbf{X}^T(*, i) = [\mathbf{X}_i^T, \mathbf{X}_i^T(u - u_i), \mathbf{X}_i^T(v - v_i)]$ . Matrix  $\Omega(u, v) = E\{\mathbf{X}(*, i)\mathbf{X}^T(*, i)| (u, v)\}$  is non-singular and has bounded second order derivatives on  $\mathcal{S}$ . Function  $E\{\|\mathbf{X}_i\|^4 | (u, v)\}$  is also bounded on  $\mathcal{S}$ .

(C6)  $F_\epsilon(\cdot | (u, v), \mathbf{X}) = \tau$  for all  $((u, v), \mathbf{X})$ , and  $f_\epsilon(\cdot | (u, v), \mathbf{X})$  is bounded away from zero and has a continuous and uniformly bounded derivative.

**Remark 2.** The conditions (C1)–(C3) are often assumed in asymptotic analysis of the nonparametric regression problems, see [Kai et al. \(2011\)](#), [Zhao et al. \(2014\)](#) and [Yang et al. \(2015\)](#). Since the sampling units can be regarded as given, the fixed bounded design condition (C4) is introduced by [Sun et al. \(2014\)](#) for technical convenience. (C4) does not preclude the condition that  $\{(u_i, v_i)\}_{i=1}^n$  are generated by some random mechanism. As pointed out in [Sun et al. \(2014\)](#), if  $(u_i, v_i)$  is i.i.d. with joint density  $f(u, v)$ , the condition (C4) holds with probability one which can be obtained in a similar way to [Hansen \(2008\)](#). (C5) is used in variable selection literature. (C6) is assumed in theoretical analysis of quantile regression.

**Theorem 1 (Consistent Selection).** Suppose that (C1)–(C6) hold. If  $h \propto n^{-1/5}$ ,  $nh^{-1/2}a_{1n} \rightarrow 0$ ,  $nh^{-1/2}b_{1n} \rightarrow 0$ ,  $nh^{-1/2}a_{2n} \rightarrow \infty$ ,  $nh^{-1/2}b_{2n} \rightarrow \infty$ , then we have

$$\begin{aligned} P\left\{\|\widehat{\mathbf{a}}_c(j)\| = 0\right\} &\rightarrow 1, \text{ for } p_0 < j \leq p; \\ P\left\{\|\widehat{\mathbf{b}}_c(j)\| = 0\right\} &\rightarrow 1, \text{ for } p_1 < j \leq p. \end{aligned}$$

**Theorem 1** describes the selection consistency of the GWQlasso method, which indicates that the proposed method can correctly identify spatially varying coefficients, constant coefficients and zero coefficients with probability tending to one over the entire geographical location support uniformly. According to **Theorem 1**, the GWQlasso method leads to a semi-parametric GWQR model when the irrelevant explanatory variables are removed. To study the oracle properties, we define the oracle estimator (i.e., the unpenalized estimator obtained under the true model) as

$$\begin{aligned} \widehat{\boldsymbol{\beta}}_{\text{oracle}}(u, v) = \arg \min_{\boldsymbol{\beta}_{\mathcal{A}_V}(u, v)} \sum_{i=1}^n \rho_\tau \left\{ Y_i - \mathbf{X}_{i, \mathcal{A}_C}^T \boldsymbol{\beta}_{\mathcal{A}_C} - (u_i - u) \mathbf{X}_{i, \mathcal{A}_V}^T \boldsymbol{\beta}_{\mathcal{A}_V}^{(u)}(u, v) \right. \\ \left. - (v_i - v) \mathbf{X}_{i, \mathcal{A}_V}^T \boldsymbol{\beta}_{\mathcal{A}_V}^{(v)}(u, v) \right\} K_h(\|(u_i, v_i) - (u, v)\|). \end{aligned}$$

**Theorem 2. (Oracle Property)** Suppose that (C1)–(C6) hold. If  $h \propto n^{-1/5}$ ,  $nh^{-1/2}a_{1n} \rightarrow 0$ ,  $nh^{-1/2}b_{1n} \rightarrow 0$ ,  $nh^{-1/2}a_{2n} \rightarrow \infty$ ,  $nh^{-1/2}b_{2n} \rightarrow \infty$ , then we obtain

$$\sup_{(u, v) \in \mathcal{S}} \|\widehat{\boldsymbol{\beta}}_{\mathcal{A}_V}(u, v) - \widehat{\boldsymbol{\beta}}_{\text{oracle}}(u, v)\| = o_p\{n^{-2/5}\}.$$

**Theorem 2** implies that the optimal point-wise convergence rate of the oracle estimator for the spatially varying coefficients is  $o_p\{n^{-2/5}\}$ . Note that the difference between the GWQlasso estimate and the oracle estimate for the spatially varying coefficients is negligible uniformly over the entire geographical location support when the true structure sets  $\mathcal{A}_V$ ,  $\mathcal{A}_C$  and  $\mathcal{A}_Z$  are identified correctly. The detailed proof of Theorem 1 and Theorem 2 is attached in supplementary materials.

## 4. Computational algorithm for GWQlasso

### 4.1. Derivation of the computational algorithm

Note that both the local weighted quantile loss function and the penalty function in (8) are non-differentiable at the origin, which leads to the common derivative-based algorithm unusable for obtaining the solution of  $\mathcal{L}_{h,\lambda}(\mathbf{A}, \mathbf{B})$ . In this section, we apply the quadratic approximation (Hunter and Lange, 2000) to approximate the local weighted quantile loss function, use the local quadratic approximation (Fan and Li, 2001) to approximate the penalty function and further establish the iterative algorithm of the GWQlasso method.

For the purpose of variable selection, we describe here an easy implementation based on the idea of the quadratic approximation (Hunter and Lange, 2000), the global weighted quantile loss function can be approximated by the following quadratic function

$$\sum_{t=1}^n \sum_{i=1}^n \rho_\tau(\mathbf{r}_{i,t}) K_h(d_{it}) \approx \frac{1}{4} \sum_{t=1}^n \sum_{i=1}^n \left[ \frac{\mathbf{r}_{i,t}^2}{\phi + |\mathbf{r}_{i,t}^{[m]}|} + (4\tau - 2)\mathbf{r}_{i,t} + C_1 \right] K_h(d_{it})$$

where  $\mathbf{r}_{i,t} = Y_i - \mathbf{X}_i^\top [\boldsymbol{\beta}(u_t, v_t) + \boldsymbol{\beta}^{(u)}(u_t, v_t)(u_t - u_i) + \boldsymbol{\beta}^{(v)}(u_t, v_t)(v_t - v_i)]$ ,  $\mathbf{r}_{i,t}^{[m]}$  is a given residual value at the  $m$ th iteration and  $C_1$  is a constant.  $\phi$  is the constant satisfying  $n\phi \ln \phi = \tau$ , where  $n$  is the sample size and  $\tau$  is the iterative convergence criterion postponed in (13). The detailed derivation process of the quadratic approximation can be referred in Hunter and Lange (2000, page 64–68).

For simplicity, the penalty function in (8) can be locally approximated by a quadratic function based on the idea of the local quadratic approximation (Fan and Li, 2001). For each  $j = 1, \dots, p$ , we have

$$\|\mathbf{a}_c(j)\| \approx \frac{\|\mathbf{a}_c(j)\|^2}{\|\widehat{\mathbf{a}}_c^{[m]}(j)\|} \quad \text{and} \quad \|\mathbf{b}_c(j)\| \approx \frac{\|\mathbf{b}_c(j)\|^2}{\|\widehat{\mathbf{b}}_c^{[m]}(j)\|},$$

where  $\widehat{\mathbf{a}}_c^{[m]}(j)$  and  $\widehat{\mathbf{b}}_c^{[m]}(j)$  are two known vectors with their norms being close to those of  $\mathbf{a}_c(j)$  and  $\mathbf{b}_c(j)$ , respectively. Next, an iterative algorithm will be derived to calculate the estimates of  $\mathbf{a}_c(j)$  and  $\mathbf{b}_c(j)$ , in which  $\widehat{\mathbf{a}}_c^{[m]}(j)$  and  $\widehat{\mathbf{b}}_c^{[m]}(j)$  are set to be the new updates of  $\mathbf{a}_c(j)$  and  $\mathbf{b}_c(j)$ .

Then the penalized function  $\mathcal{L}_{h,\lambda}(\mathbf{A}, \mathbf{B})$  can be approximated by

$$\frac{1}{4} \sum_{t=1}^n \sum_{i=1}^n \left[ \frac{\mathbf{r}_{i,t}^2}{\phi + |\mathbf{r}_{i,t}^{[m]}|} + (4\tau - 2)\mathbf{r}_{i,t} + C_1 \right] K_h(d_{it}) + \sum_{j=1}^p \left( \frac{\lambda_{1j} \|\mathbf{a}_c(j)\|^2}{\|\widehat{\mathbf{a}}_c^{[m]}(j)\|} + \frac{\lambda_{2j} \|\mathbf{b}_c(j)\|^2}{\|\widehat{\mathbf{b}}_c^{[m]}(j)\|} \right). \quad (10)$$

For ease of exposition, let

$$\begin{aligned} \mathbf{M}^{[m]}(t) &= \text{diag} \left[ \frac{K_h(d_{1t})}{\phi + |\mathbf{r}_{1,t}^{[m]}|}, \dots, \frac{K_h(d_{nt})}{\phi + |\mathbf{r}_{n,t}^{[m]}|} \right], \mathbf{N}(t) = \left[ (4\tau - 2)K_h(d_{1t}), \dots, (4\tau - 2)K_h(d_{nt}) \right]^\top, \\ \mathbf{D}_1^{[m]} &= \text{diag} \left[ \frac{\lambda_{11}}{\|\widehat{\mathbf{a}}_c^{[m]}(1)\|}, \dots, \frac{\lambda_{1p}}{\|\widehat{\mathbf{a}}_c^{[m]}(p)\|} \right], \mathbf{D}_2^{[m]} = \text{diag} \left[ \frac{\lambda_{21}}{\|\widehat{\mathbf{b}}_c^{[m]}(1)\|}, \dots, \frac{\lambda_{2p}}{\|\widehat{\mathbf{b}}_c^{[m]}(p)\|} \right], \\ \mathbf{r}_t &= \mathbf{Y} - \mathbf{X}\boldsymbol{\beta}(u_t, v_t) - \mathbf{U}(t)\mathbf{X}\boldsymbol{\beta}^{(u)}(u_t, v_t) - \mathbf{V}(t)\mathbf{X}\boldsymbol{\beta}^{(v)}(u_t, v_t). \end{aligned}$$

Following the idea of Wang and Li (2017), the penalized function (10) can be expressed as

$$\begin{aligned} \sum_{t=1}^n & \left\{ \frac{1}{4} \mathbf{r}_t^\top \mathbf{M}^{[m]}(t) \mathbf{r}_t + \mathbf{r}_t^\top \mathbf{N}(t) + \boldsymbol{\beta}^\top(u_t, v_t) \mathbf{D}_1^{[m]} \boldsymbol{\beta}(u_t, v_t) + [\boldsymbol{\beta}^{(u)}(u_t, v_t)]^\top \mathbf{D}_2^{[m]} \boldsymbol{\beta}^{(u)}(u_t, v_t) \right. \\ & \left. + [\boldsymbol{\beta}^{(v)}(u_t, v_t)]^\top \mathbf{D}_2^{[m]} \boldsymbol{\beta}^{(v)}(u_t, v_t) \right\} + C_2, \end{aligned} \quad (11)$$

where  $C_2 = \frac{C_1}{4} \sum_{t=1}^n \sum_{i=1}^n K_h(d_{it})$ . For  $t = 1, \dots, n$ , we obtain the formula for iteratively calculating the penalized estimates of  $\beta(u_t, v_t)$ ,  $\beta^{(u)}(u_t, v_t)$ ,  $\beta^{(v)}(u_t, v_t)$  as

$$\begin{aligned} & \left[ [\widehat{\beta}^{[m+1]}(u_t, v_t)]^\top, [\widehat{\beta}^{(u)[m+1]}(u_t, v_t)]^\top, [\widehat{\beta}^{(v)[m+1]}(u_t, v_t)]^\top \right]^\top = \\ & \left( \begin{array}{ccc} \mathbf{X}^\top \mathbf{M}^{[m]}(t) \mathbf{X} + 4\mathbf{D}_1^{[m]} & \mathbf{X}^\top \mathbf{M}^{[m]}(t) \mathbf{U}(t) \mathbf{X} & \mathbf{X}^\top \mathbf{M}^{[m]}(t) \mathbf{V}(t) \mathbf{X} \\ \mathbf{X}^\top \mathbf{M}^{[m]}(t) \mathbf{U}(t) \mathbf{X} & \mathbf{X}^\top \mathbf{M}^{[m]}(t) \mathbf{U}^2(t) \mathbf{X} + 4\mathbf{D}_2^{[m]} & \mathbf{X}^\top \mathbf{M}^{[m]}(t) \mathbf{U}(t) \mathbf{V}(t) \mathbf{X} \\ \mathbf{X}^\top \mathbf{M}^{[m]}(t) \mathbf{V}(t) \mathbf{X} & \mathbf{X}^\top \mathbf{M}^{[m]}(t) \mathbf{U}(t) \mathbf{V}(t) \mathbf{X} & \mathbf{X}^\top \mathbf{M}^{[m]}(t) \mathbf{V}^2(t) \mathbf{X} + 4\mathbf{D}_2^{[m]} \end{array} \right)^{-1} \\ & \times \begin{pmatrix} \mathbf{X}^\top \mathbf{M}^{[m]}(t) \mathbf{Y} + \mathbf{X}^\top \mathbf{N}(t) \\ \mathbf{X}^\top \mathbf{M}^{[m]}(t) \mathbf{U}(t) \mathbf{Y} + \mathbf{X}^\top \mathbf{U}(t) \mathbf{N}(t) \\ \mathbf{X}^\top \mathbf{M}^{[m]}(t) \mathbf{V}(t) \mathbf{Y} + \mathbf{X}^\top \mathbf{V}(t) \mathbf{N}(t) \end{pmatrix}. \end{aligned} \quad (12)$$

In the above iterative algorithm,  $\mathbf{D}_1^{[m]}$ ,  $\mathbf{D}_2^{[m]}$ ,  $\mathbf{M}^{[m]}(t)$  should be updated in each iteration. The derivation of the iterative formula (12) are given in the Supplemental materials.

In practice, once the bandwidth  $h$  and the penalization parameters  $\lambda_1$ ,  $\lambda_2$  and the initial values of  $\mathbf{A}$  and  $\mathbf{B}$  (or  $\beta(u_t, v_t)$ ,  $\beta^{(u)}(u_t, v_t)$ ,  $\beta^{(v)}(u_t, v_t)$  for  $t = 1, \dots, n$ ) are specified. The penalized estimates of  $\beta(u_t, v_t)$ ,  $\beta^{(u)}(u_t, v_t)$ ,  $\beta^{(v)}(u_t, v_t)$  for  $t = 1, \dots, n$ , and therefore those of  $\mathbf{A}$  and  $\mathbf{B}$  can be obtained by performing the iterations until convergence. Let  $\mathcal{L}_{h,\lambda}(\widehat{\mathbf{A}}^{[m]}, \widehat{\mathbf{B}}^{[m]})$  be the estimates of  $\mathcal{L}_{h,\lambda}(\mathbf{A}, \mathbf{B})$ , where  $\widehat{\mathbf{A}}^{[m]}$  and  $\widehat{\mathbf{B}}^{[m]}$  are the estimates of  $\mathbf{A}$  and  $\mathbf{B}$  in the  $m$ th iteration. Then we define the convergence criterion

$$tol = \left| \frac{\mathcal{L}_{h,\lambda}(\widehat{\mathbf{A}}^{[m]}, \widehat{\mathbf{B}}^{[m]}) - \mathcal{L}_{h,\lambda}(\widehat{\mathbf{A}}^{[m+1]}, \widehat{\mathbf{B}}^{[m+1]})}{\mathcal{L}_{h,\lambda}(\widehat{\mathbf{A}}^{[m]}, \widehat{\mathbf{B}}^{[m]})} \right| < \tau, \quad (13)$$

where  $\tau$  is a pre-specified threshold value.

#### 4.2. Choice of the bandwidth and the penalization parameters

By [Theorems 1](#) and [2](#), we know that as long as the following four conditions

$$nh^{-1/2}a_{1n} \rightarrow 0, \quad nh^{-1/2}b_{1n} \rightarrow 0, \quad nh^{-1/2}a_{2n} \rightarrow \infty, \quad nh^{-1/2}b_{2n} \rightarrow \infty$$

are satisfied, the optimal nonparametric convergence rate can be achieved and the true model can be consistently identified. However, it is difficult to simultaneously choose the sizes of the bandwidth and a total of  $2p$  penalization parameters. To achieve the task, we follow the idea of [Wang and Xia \(2009\)](#) and [Kai et al. \(2011\)](#), we separately select the optimal sizes of  $h$ ,  $\lambda_1$  and  $\lambda_2$  in the following way.

The optimal size of the bandwidth  $h$  is taken to be  $h_o$  which is selected by the CV criterion (5). The unpenalized estimates  $\widehat{\mathbf{A}}^{[0]}$  and  $\widehat{\mathbf{B}}^{[0]}$  resulted from the local-linear GWQR estimates of  $\beta(u_t, v_t)$ ,  $\beta^{(u)}(u_t, v_t)$ ,  $\beta^{(v)}(u_t, v_t)$  ( $t = 1, \dots, n$ ) are set to be the initial estimates of  $\mathbf{A}$  and  $\mathbf{B}$  for running the iterative formula in (12). For the penalization parameters  $\lambda_1$  and  $\lambda_2$ , we simplify the penalization parameters of  $\lambda_1$  and  $\lambda_2$  as

$$\lambda_{1j} = \frac{\lambda}{\left[ n^{-1/2} \|\widehat{\mathbf{a}}_c^{[0]}(j)\| \right]^{3/2}}, \quad \lambda_{2j} = \frac{\lambda}{\left[ (2n)^{-1/2} \|\widehat{\mathbf{b}}_c^{[0]}(j)\| \right]^{3/2}}, \quad (14)$$

where  $\widehat{\mathbf{a}}_c^{[0]}(j)$  and  $\widehat{\mathbf{b}}_c^{[0]}(j)$  are the  $j$ th column of the unpenalized estimate of  $\widehat{\mathbf{A}}^{[0]}$  and  $\widehat{\mathbf{B}}^{[0]}$ . One can verify that as long as  $\lambda nh^{-1/2} \rightarrow 0$  but  $\lambda n^{8/5}h^{-1/2} \rightarrow \infty$ , then the four conditions listed in [Theorem 2](#) are satisfied. Consequently, the original  $2p$ -dimensional problem about  $\lambda_1 \in \mathbb{R}^p$  and  $\lambda_2 \in \mathbb{R}^p$  becomes a univariate problem about  $\lambda \in \mathbb{R}^1$ . Given  $\lambda > 0$  and the selected bandwidth  $h_o$ , we run the iterative formula (12) to yield the penalized estimates  $\widehat{\mathbf{A}}_{h_o, \lambda}$  and  $\widehat{\mathbf{B}}_{h_o, \lambda}$  of  $\mathbf{A}$  and  $\mathbf{B}$ . The optimal penalization

parameter  $\lambda$  can be selected according to the following BIC-type criterion

$$BIC(\lambda) = \log\left(\frac{1}{n^2} RSS_{h_0, \lambda}\right) + df_\lambda \frac{\log(nh)}{nh} + (p - df_\lambda) \frac{\log n}{n}, \quad (15)$$

where  $df_{h_0, \lambda}$  is the number of the varying coefficients, and

$$\begin{aligned} RSS_{h_0, \lambda} = \sum_{t=1}^n \sum_{i=1}^n \rho_\tau & \left\{ Y_i - \mathbf{X}_i^\top [\widehat{\boldsymbol{\beta}}(u_t, v_t) \right. \\ & \left. + \widehat{\boldsymbol{\beta}}^{(u)}(u_t, v_t)(u_t - u_i) + \widehat{\boldsymbol{\beta}}^{(v)}(u_t, v_t)(v_t - v_i)] \right\} K_{h_0}(d_{it}). \end{aligned} \quad (16)$$

The optimal penalized parameter can be obtained as  $\lambda_o = \arg \min_{\lambda > 0} BIC(\lambda)$ .

#### 4.3. Implementation of the GWQlasso method as follows

In order to facilitate the implementation of the proposal, we present the operation steps as follows.

---

**Algorithm 1:** Geographically Weighted Quantile lasso (GWQlasso)

---

**Input:**  $\{Y_i; X_{i1}, \dots, X_{ip}; (u_i, v_i)\}_{i=1}^n, \tau, \tau, \delta, \mathcal{H} = \{h_k\}_{k=1}^K$  and  $\Lambda = \{\lambda_l\}_{l=1}^L$ .

**Output:**  $\widehat{\mathbf{A}}_V, \widehat{\mathcal{A}}_C, \widehat{\mathcal{A}}_Z, \widehat{\boldsymbol{\beta}}_{\widehat{\mathcal{A}}_V}, \widehat{\boldsymbol{\beta}}_{\widehat{\mathcal{A}}_C}$  and  $\widehat{\boldsymbol{\beta}}_{\widehat{\mathcal{A}}_Z}(u_t, v_t), (t = 1, \dots, n)$ .

**Initialization:** Compute  $\widehat{\mathbf{A}}^{[0]}, \widehat{\mathbf{B}}^{[0]}$  by the local-linear GWQR method by (4) and select the optimal bandwidth  $h_0$  by the CV score (5).

**Iterations:**

**for**  $l = 1$  to  $L$  **do**

    Compute the penalization parameters  $\lambda_1$  and  $\lambda_2$ .

**while**  $tol > \tau$  **do**

**for**  $t = 1$  to  $n$  **do**

            | Compute GWQlasso estimates  $\widehat{\boldsymbol{\beta}}(u_t, v_t), \widehat{\boldsymbol{\beta}}^{(u)}(u_t, v_t)$  and  $\widehat{\boldsymbol{\beta}}^{(v)}(u_t, v_t)$  by (12).

            | **end**

            | Compute the convergence criterion  $tol$  by (13).

        | **end**

        | By the identification criterion (See Remark 3), compute  $BIC(\lambda_l)$  values by (15).

**end**

Select the optimal size  $\lambda_o$  and obtain corresponding estimates  $\widehat{\mathbf{A}}_{h_0, \lambda_o}$  and  $\widehat{\mathbf{B}}_{h_0, \lambda_o}$ .

**Identification:** Obtain the index sets  $\widehat{\mathcal{A}}_V, \widehat{\mathcal{A}}_C$  and  $\widehat{\mathcal{A}}_Z$  by (9) and the corresponding estimates  $\widehat{\boldsymbol{\beta}}_{\widehat{\mathcal{A}}_V}, \widehat{\boldsymbol{\beta}}_{\widehat{\mathcal{A}}_C}$  and  $\widehat{\boldsymbol{\beta}}_{\widehat{\mathcal{A}}_Z}(u_t, v_t), (t = 1, \dots, n)$ .

---

**Remark 3.** Applying the GWQlasso method, we can obtain the limit values of the estimates  $\widehat{\mathbf{a}}_c(j)$  and  $\widehat{\mathbf{b}}_c(j)$  of  $\mathbf{a}_c(j)$  and  $\mathbf{b}_c(j)$  for  $j = 1, \dots, p$ . In practice, the elements in coefficient estimates in Eq. (12) are not penalized exactly to zero, thus thresholds have to be applied to set the small elements to zero. Similar idea can be found in statistical literature, such as Wang and Lin (2016) and Wang and Li (2017).

Specifically, for  $j = 1, \dots, p$ , if  $\|\widehat{\mathbf{a}}_c(j)\|/n < \delta$  for all  $i = 1, \dots, n$ , we set  $\widehat{\mathbf{a}}_c(j) = \mathbf{0}$ ; if  $\|\widehat{\mathbf{b}}_c(j)\|/(2n) < \delta$ , we set  $\widehat{\mathbf{b}}_c(j) = \mathbf{0}$ . According to (9), if  $j \in \widehat{\mathcal{A}}_V$ , then the elements in  $\widehat{\mathbf{a}}_c(j)$  are the shrunk estimates of  $\beta_j(u, v)$  at the  $n$  designed locations; if  $j \in \widehat{\mathcal{A}}_C$ , the nonzero constant coefficient  $\beta_j$  can be estimated by  $\beta_j = \frac{1}{n} \sum_{i=1}^n \widehat{\beta}_j(u_i, v_i)$ .

In the GWQlasso method, how to choose the appropriate value of the threshold is an important issue. In simulation studies, the threshold  $\delta$  is somewhat related to the convergence criterion  $\tau$  in Eq. (13). The smaller value  $\tau$  is set, the smaller value  $\delta$  can be selected, which indicates there is a closer approximation to set small elements to zero. However, the cost of the closer approximation is the more iterations in the inner loop and the more time used. Discovered by numerical simulations, the choice of the convergence criterion  $\tau = 0.01$  and the threshold  $\delta = 0.05$  seems useful and appropriate in simulation and application.

## 5. Numerical experiments

### 5.1. Simulation studies

To demonstrate the finite sample performance of the proposed GWQlasso method, we consider the following two GWQR models.

$$(I) Y_i = 10\left(\sqrt{u_i(1-u_i)} \sin\left(\frac{\pi(1+2^{-2.2})}{2(u_i+2^{-2.2})}\right) + \sqrt{v_i(1-v_i)} \sin\left(\frac{\pi(1+2^{-2.2})}{2(v_i+2^{-2.2})}\right)\right)X_{i1} \\ + 3 \sin(\pi u_i) \sin(2\pi v_i) X_{i2} + 4X_{i3} + 3X_{i4} + 0X_{i5} + 0X_{i6} + \epsilon_{\tau,i}, \\ (II) Y_i = 4 \exp(1 - 2((u_i - 0.5)^2 + (v_i - 0.5)^2)) X_{i1} + 5 \sin(2\pi u_i) \sin(2\pi v_i) X_{i2} \\ + 6(u_i^2 + v_i^2) X_{i3} + 4X_{i4} + 3X_{i5} + 0X_{i6} + \epsilon_{\tau,i},$$

where  $X_{i1} \equiv 1$ ,  $(X_{i2}, \dots, X_{i6})^\top$  are drawn from the multivariate normal distribution  $N(\mathbf{0}, \Sigma)$  with the covariance matrix  $\Sigma = (\rho^{|j-k|})_{1 \leq j, k \leq 6}$ . To assess the impact of the collinearity among the explanatory variables,  $\rho$  is taken to be 0, 0.5, 0.9, respectively, indicating the cases that the explanatory variables are independent of each other, moderately correlated and highly correlated. The random error  $\epsilon_{\tau,i}$  is given by  $\epsilon_{\tau,i} = \epsilon_i - \Phi^{-1}(\tau)$ , with  $\Phi(\cdot)$  being the common cumulative distribution function of  $\epsilon_i$ . And  $\Phi^{-1}(\tau)$  is subtracted from  $\epsilon_i$  to make the  $\tau$ th quantile of  $\epsilon_i$  zero for identifiability purpose. For  $\epsilon_i$ , we consider the following error distributions: (i) standard normal distribution  $N(0, 1)$ ; (ii)  $t$  distribution with 3 degrees of freedom  $t(3)$ ; (iii) mixture of normals  $0.9N(0, 1) + 0.1N(0, 10^2)$ ; (iv) logistic distribution  $L(0, 1)$ .

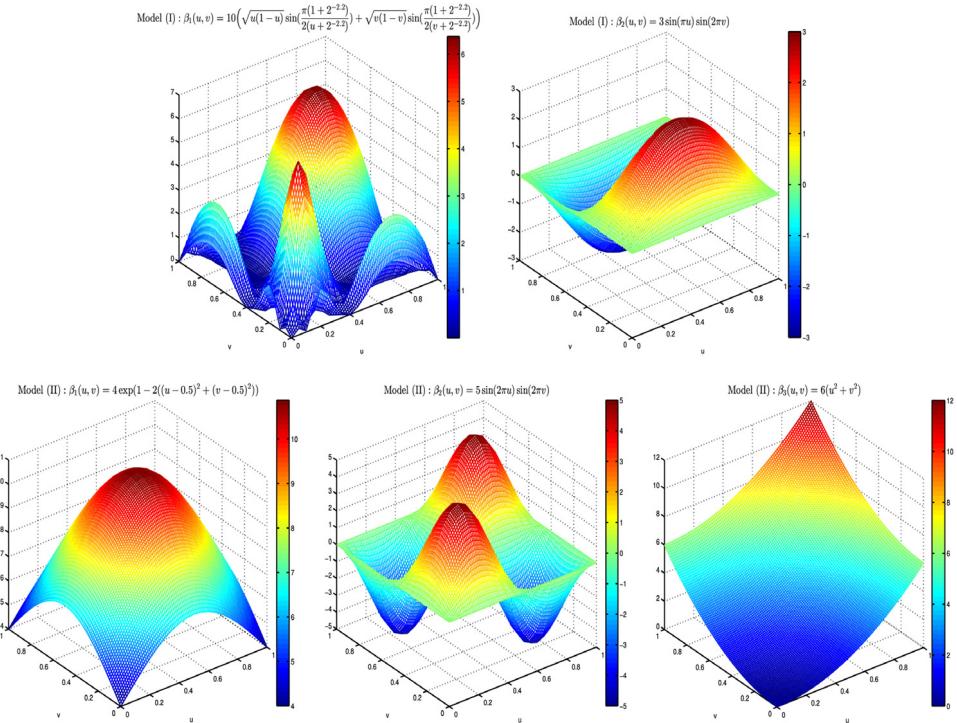
The spatial layout is a square region with the length of each side being  $m$  units. The observations of the response and the explanatory variables are located at equally spaced  $m \times m$  grid points in the square region. Here, we take  $m = 21$  which results in the sample size  $n = m^2 = 441$ . Let the origin of the Cartesian coordinate system be located at the bottom left corner of the square region. The spatial coordinates of the observations are

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

where  $\text{mod}(\frac{a}{b})$  and  $\text{int}(\frac{a}{b})$  are the remainder and the integer part on  $a$  divided by  $b$ , respectively. To visualize the variation patterns of the spatially varying coefficients in models (I) and (II), we depict the true surfaces of these coefficients in Fig. 1.

The kernel function  $K(\cdot)$  is chosen as the Epanechnikov kernel  $K(t) = 0.75(1-t^2)_+$  or the Gaussian kernel  $\frac{1}{\sqrt{2\pi}} \exp(-t^2/2)$ . The optimal bandwidth is selected by minimizing the leave-one-out CV score. The optimal bandwidth is used for GWQlasso method, where the optimal penalized parameter is determined by the BIC criterion in (15). The candidate sets of the bandwidth  $h$  and the penalization parameter  $\lambda$  are set to be  $\mathcal{H} = \{h_k : h_k = 0.2 + 0.05 \times k, k = 1, \dots, 20\}$  and  $\Lambda = \{\lambda_l : \lambda_l = 10 \times l, l = 1, \dots, 20\}$ .

For each setting, we take the number of replications  $N = 200$  and  $\tau = 0.25, 0.5, 0.75$ . Considering the limited space and the fact that the corresponding results for the quantile levels  $\tau = 0.25$  and  $0.75$ , we only present the identification results for  $\tau = 0.5$  and  $0.25$ . It is found from the results that the identification results of the GWQlasso method are a little influenced by the different quantile levels. Those results for  $\tau = 0.25$  are attached as a supplementary material. Tables 1–2 summarize the identification results at  $\tau = 0.5$ . In all of the simulation experiments, the GWQlasso method can correctly separate the varying, nonzero constant and zero coefficients with large frequency. The GWQlasso produces the most accurate identification performance when the model error follow  $N(0, 1)$  regardless of whether the kernel function is the Epanechnikov kernel or the Gaussian kernel. The identification results of the GWQlasso are little influenced by the non-normality of the error distribution. These findings indicate that the proposal is stable to non-normal random errors. Furthermore, the relative frequencies of the underlying coefficients correctly identified into the final model under the independent and the moderately correlated collinearity are reasonably close to one. And the relative frequencies which the varying coefficients are identified correctly are large under high correlated collinearity, which indicates that the GWQlasso is robust to the collinearity



**Fig. 1.** The plots depict the true surfaces of the spatially varying coefficients in models (I) and (II).

among the explanatory variables. Last, we find that the results of the Epanechnikov kernel is a little different to that of the Gaussian kernel. Thus, the role played by the choice of the kernel function on GWQlasso's performance is limited.

## 5.2. Analysis of Dublin voter turnout data

We now apply the GWQlasso method to the Dublin voter turnout data, which has been analyzed by [Gollini et al. \(2015\)](#) and is publicly available in the R package called **GWmodel**. Following [Gollini et al. \(2015\)](#), 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, the geographical locations ( $u, v$ ) as the index variable, and the following characteristic of social structures as the explanatory variables: MDA (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. In voter turnout studies, the researchers are more interested in how the characteristic of social structures explain the variation in low voter turnout (lower quantile) and high voter turnout (upper quantile) areas than in how they explain variation in average voter turnout regions. Moreover, it is also important for researchers to select relevant variables and identify constant variables under different quantile levels in the GWQR model. Based on the data set, we apply the GWQlasso method to identify the spatially varying, non-zero constant and zero coefficients in the GWQR model under different quantile levels.

Before applying the GWQlasso method, all the explanatory variables are transformed so that their marginal distribution is approximately  $N(0, 1)$ . The normalized procedure here is same to other

**Table 1**

Frequencies of each coefficient identified to be spatially varying (V), nonzero constant (C) or zero (Z) at  $\tau = 0.5$  in Model (I) over 200 simulation replications.

Coefficient	$\rho = 0$			$\rho = 0.5$						$\rho = 0.9$										
	Epanechnikov			Gaussian			Epanechnikov			Gaussian			Epanechnikov			Gaussian				
	V	C	Z	V	C	Z	V	C	Z	V	C	Z	V	C	Z	V	C	Z		
$N(0, 1)$																				
$\beta_1(u, v)$	<b>181</b>	19	0	<b>200</b>	0	0	<b>179</b>	21	0	<b>200</b>	0	0	<b>182</b>	18	0	<b>200</b>	0	0		
$\beta_2(u, v)$	<b>200</b>	0	0	<b>200</b>	0	0														
$\beta_3 = 3$	0	<b>200</b>	0	29	<b>171</b>	0	0	<b>200</b>	0	0										
$\beta_4 = 4$	0	<b>200</b>	0	9	<b>191</b>	0	0	<b>200</b>	0	0										
$\beta_5 = 0$	0	0	<b>200</b>	0	18	<b>182</b>	0	0	<b>200</b>	0	0									
$\beta_6 = 0$	0	0	<b>200</b>	0	3	<b>197</b>	0	0	<b>200</b>	0	0									
$t(3)$																				
$\beta_1(u, v)$	<b>198</b>	2	0	<b>200</b>	0	0	<b>196</b>	4	0	<b>200</b>	0	0	<b>187</b>	13	0	<b>200</b>	0	0		
$\beta_2(u, v)$	<b>200</b>	0	0	<b>200</b>	0	0														
$\beta_3 = 3$	2	<b>198</b>	0	0	<b>200</b>	0	8	<b>192</b>	0	0	<b>200</b>	0	80	<b>120</b>	0	25	<b>175</b>	0	0	
$\beta_4 = 4$	2	<b>198</b>	0	0	<b>200</b>	0	4	<b>196</b>	0	0	<b>200</b>	0	60	<b>140</b>	0	5	<b>195</b>	0	0	
$\beta_5 = 0$	5	0	<b>195</b>	0	0	<b>200</b>	10	0	<b>190</b>	0	0	<b>200</b>	70	0	<b>130</b>	9	0	<b>191</b>	0	0
$\beta_6 = 0$	7	0	<b>193</b>	0	0	<b>200</b>	0	0	<b>200</b>	0	0	<b>200</b>	33	0	<b>167</b>	1	0	<b>199</b>	0	0
$0.9N(0, 1) + 0.1N(0, 10^2)$																				
$\beta_1(u, v)$	<b>194</b>	6	0	<b>200</b>	0	0	<b>200</b>	0	0	<b>200</b>	0	0	<b>180</b>	20	0	<b>200</b>	0	0		
$\beta_2(u, v)$	<b>200</b>	0	0	<b>200</b>	0	0														
$\beta_3 = 3$	0	<b>200</b>	0	0	<b>200</b>	0	1	<b>199</b>	0	0	<b>200</b>	0	56	<b>144</b>	0	4	<b>196</b>	0	0	
$\beta_4 = 4$	0	<b>200</b>	0	45	<b>155</b>	0	0	<b>200</b>	0	0										
$\beta_5 = 0$	0	0	<b>200</b>	47	0	<b>153</b>	0	0	<b>200</b>	0	0									
$\beta_6 = 0$	1	0	<b>199</b>	0	0	<b>200</b>	0	0	<b>200</b>	0	0	<b>200</b>	26	0	<b>174</b>	0	0	<b>200</b>	0	0
$L(0, 1)$																				
$\beta_1(u, v)$	<b>196</b>	4	0	<b>200</b>	0	0	<b>189</b>	11	0	<b>200</b>	0	0	<b>193</b>	7	0	<b>200</b>	0	0		
$\beta_2(u, v)$	<b>200</b>	0	0	<b>200</b>	0	0														
$\beta_3 = 3$	11	<b>189</b>	0	0	<b>200</b>	0	16	<b>184</b>	0	0	<b>200</b>	0	90	<b>110</b>	0	19	<b>181</b>	0	0	
$\beta_4 = 4$	9	<b>191</b>	0	0	<b>200</b>	0	10	<b>190</b>	0	0	<b>200</b>	0	99	<b>101</b>	0	14	<b>186</b>	0	0	
$\beta_5 = 0$	14	0	<b>186</b>	0	0	<b>200</b>	20	0	<b>180</b>	0	0	<b>200</b>	98	0	<b>102</b>	8	0	<b>192</b>	0	0
$\beta_6 = 0$	15	0	<b>185</b>	0	0	<b>200</b>	13	0	<b>187</b>	0	0	<b>200</b>	55	0	<b>145</b>	0	0	<b>200</b>	0	0

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

quantile variable selection methods, such as Kai et al. (2011), Zhao et al. (2014) and Yang et al. (2015). For the geographical locations, the scale is used such that one decimeter equals one kilometer. Then we consider the following GWQR model

$$\begin{aligned} GEI = & \beta_0(u, v) + \beta_1(u, v)MDA + \beta_2(u, v)LAR + \beta_3(u, v)SCO + \beta_4(u, v)UEP \\ & + \beta_5(u, v)LOE + \beta_6(u, v)AGY + \beta_7(u, v)AGM + \beta_8(u, v)AGO + \epsilon_{\tau}. \end{aligned} \quad (17)$$

Based on the data set, we compute the GWQR estimates of the regression coefficients for the different quantile levels  $\tau = 0.25, 0.5, 0.75$ . Note that the judging threshold  $\delta$  in Remark 3 is vital in determining whether the coefficient estimates is zero, nonzero constant, or varying over space. Now we consider several levels of the judging thresholds to be  $\delta = 0.1, 0.05, 0.01$  with the stopping criterion  $\tau = 0.01$ . Here, the Epanechnikov kernel is used and the bandwidth size is selected by the cross-validation method. The candidate sets of the bandwidth  $h$  and the penalization parameter  $\lambda$  are set to be  $\mathcal{H} = \{h_k : h_k = 3 + 0.1 \times k, k = 1, \dots, 20\}$  and  $\Lambda = \{\lambda_l : \lambda_l = 2.5 \times l, l = 1, \dots, 20\}$ . As suggested by the reviewer, the comparisons between the GWQlasso with threshold and the GWQR with threshold need to be further discussed. Several levels of the thresholds  $\delta = 0.1, 0.05, 0.01$  are considered, which is convenient for comparison with the threshold in the GWQlasso method. For each of the threshold, the identification results were computed for two method and the results are reported in Table 3, where the letters 'L' and 'R' refer respectively to the GWQlasso with threshold and the GWQR with threshold.

**Table 2**

Frequencies of each coefficient identified to be spatially varying (V), nonzero constant (C) or zero (Z) at  $\tau = 0.5$  in Model (II) over 200 simulation replications.

Coefficient	$\rho = 0$			$\rho = 0.5$			$\rho = 0.9$											
	Epanechnikov			Gaussian			Epanechnikov			Gaussian			Epanechnikov			Gaussian		
	V	C	Z	V	C	Z	V	C	Z	V	C	Z	V	C	Z	V	C	Z
$N(0, 1)$																		
$\beta_1(u, v)$	<b>200</b>	0	0	<b>181</b>	19	0	<b>200</b>	0	0	<b>186</b>	14	0	<b>200</b>	0	0	<b>174</b>	26	0
$\beta_2(u, v)$	<b>200</b>	0	0	<b>181</b>	0	19	<b>200</b>	0	0	<b>191</b>	0	9	<b>200</b>	0	0	<b>173</b>	0	27
$\beta_3(u, v)$	<b>200</b>	0	0	<b>200</b>	0	0	<b>200</b>	0	0	<b>200</b>	0	0	<b>200</b>	0	0	<b>200</b>	0	0
$\beta_4 = 3$	0	<b>200</b>	0	1	<b>199</b>	0	0	<b>200</b>	0	2	<b>198</b>	0	25	<b>175</b>	0	23	<b>177</b>	0
$\beta_5 = 4$	0	<b>200</b>	0	0	<b>200</b>	0	0	<b>200</b>	0	1	<b>199</b>	0	101	<b>99</b>	0	116	<b>84</b>	0
$\beta_6 = 0$	0	0	<b>200</b>	0	0	<b>200</b>	0	0	<b>200</b>	0	0	<b>200</b>	8	0	<b>192</b>	64	0	<b>136</b>
$t(3)$																		
$\beta_1(u, v)$	<b>200</b>	0	0	<b>173</b>	27	0	<b>200</b>	0	0	<b>178</b>	22	0	<b>200</b>	0	0	<b>173</b>	27	0
$\beta_2(u, v)$	<b>200</b>	0	0	<b>171</b>	0	29	<b>200</b>	0	0	<b>186</b>	0	14	<b>200</b>	0	0	<b>182</b>	0	18
$\beta_3(u, v)$	<b>200</b>	0	0	<b>200</b>	0	0	<b>200</b>	0	0	<b>200</b>	0	0	<b>200</b>	0	0	<b>200</b>	0	0
$\beta_4 = 3$	4	<b>196</b>	0	14	<b>186</b>	0	8	<b>192</b>	0	20	<b>180</b>	0	103	<b>97</b>	0	68	<b>132</b>	0
$\beta_5 = 4$	2	<b>198</b>	0	15	<b>185</b>	0	4	<b>196</b>	0	8	<b>192</b>	0	62	<b>138</b>	0	147	<b>53</b>	0
$\beta_6 = 0$	1	0	<b>199</b>	5	0	<b>195</b>	3	0	<b>197</b>	6	0	<b>194</b>	71	0	<b>129</b>	118	0	<b>82</b>
$0.9N(0, 1) + 0.1N(0, 10^2)$																		
$\beta_1(u, v)$	<b>200</b>	0	0	<b>162</b>	38	0	<b>200</b>	0	0	<b>173</b>	27	0	<b>200</b>	0	0	<b>164</b>	36	0
$\beta_2(u, v)$	<b>200</b>	0	0	<b>177</b>	0	23	<b>200</b>	0	0	<b>181</b>	0	19	<b>200</b>	0	0	<b>170</b>	0	30
$\beta_3(u, v)$	<b>200</b>	0	0	<b>200</b>	0	0	<b>200</b>	0	0	<b>200</b>	0	0	<b>200</b>	0	0	<b>200</b>	0	0
$\beta_4 = 3$	0	<b>200</b>	0	5	<b>195</b>	0	1	<b>199</b>	0	1	<b>199</b>	0	110	<b>90</b>	0	56	<b>144</b>	0
$\beta_5 = 4$	1	<b>199</b>	0	4	<b>196</b>	0	4	<b>196</b>	0	3	<b>197</b>	0	150	<b>50</b>	0	127	<b>73</b>	0
$\beta_6 = 0$	4	0	<b>196</b>	1	0	<b>199</b>	0	0	<b>200</b>	0	0	<b>200</b>	60	0	<b>140</b>	86	0	<b>114</b>
$L(0, 1)$																		
$\beta_1(u, v)$	<b>200</b>	0	0	<b>160</b>	40	0	<b>200</b>	0	0	<b>166</b>	34	0	<b>200</b>	0	0	<b>163</b>	37	0
$\beta_2(u, v)$	<b>200</b>	0	0	<b>173</b>	0	27	<b>200</b>	0	0	<b>176</b>	0	24	<b>200</b>	0	0	<b>170</b>	0	30
$\beta_3(u, v)$	<b>200</b>	0	0	<b>200</b>	0	0	<b>200</b>	0	0	<b>200</b>	0	0	<b>200</b>	0	0	<b>200</b>	0	0
$\beta_4 = 3$	8	<b>192</b>	0	12	<b>188</b>	0	17	<b>183</b>	0	17	<b>183</b>	0	136	<b>64</b>	0	108	<b>92</b>	0
$\beta_5 = 4$	20	<b>180</b>	0	17	<b>183</b>	0	32	<b>168</b>	0	11	<b>189</b>	0	165	<b>35</b>	0	146	<b>54</b>	0
$\beta_6 = 0$	19	0	<b>181</b>	3	0	<b>197</b>	14	0	<b>186</b>	1	0	<b>199</b>	109	0	<b>91</b>	124	0	<b>76</b>

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

In can be obtained from Table 3 that the large value  $\delta$  leads to identify more zero coefficients, whereas the small value  $\delta$  tends to choose more varying coefficients. Nevertheless, the identification results of  $\delta = 0.05$  are likely to be persuasiveness for the three selected quantile levels. The resulting GWQlasso estimation with the threshold  $\delta = 0.05$  implies that MDA, LAR, SCO, UEP, AGM and AGO are all relevant variables, whereas LOE and AGY are not. These results suggest that the relationships between the six explanatory variables presented the characteristic of social structures and the voter turnout are non-stationary across the Dublin region.

From the identification results in Table 3, there are some differences between the identification results of the GWQR with the threshold and that of the GWQlasso method with the threshold. For the three selected quantile levels, the coefficient of AGY is non-zero constant in the GWQR technique with the threshold while the coefficient of AGY is zero in the GWQlasso method. For the quantile level  $\tau = 0.25$ , the coefficient of UEP varies spatially in the GWQR technique with threshold  $\delta = 0.1$  while the coefficient of UEP is non-zero constant in the GWQlasso method with threshold  $\delta = 0.1$ . Actually, the GWQR with threshold and the GWQlasso with threshold are essentially different. The principle of the threshold in the GWQlasso method is given in Remark 3 while the reason of the threshold in the GWQR technique is somewhat unknown. Although these two method give similar results, the GWQlasso method shows the shrinkage effect on the coefficient estimates.

To better understanding the spatial non-stationarity of the regression relationship for different quantile levels, we provide in Table 4 the maps of the GWQlasso estimates for the intercept and six explanatory variables at different quantile levels. As summarized in Table 4, among the maps of

**Table 3**

The identification results for GWQR with threshold and GWQlasso with threshold.

$\delta$	Method	$h_0$	$\lambda_0$	INT	MDA	LAR	SCO	UEP	LOE	AGY	AGM	AGO
$\tau = 0.25$												
0.01	L	3.7	22.5	V	V	V	V	V	V	V	V	V
	R			V	V	V	V	V	V	V	V	V
0.05	L	3.7	42.5	V	V	V	V	V	Z	Z	V	V
	R			V	V	V	V	V	Z	C	V	V
0.10	L	3.7	45	V	V	C	V	C	Z	Z	V	V
	R			V	V	C	V	V	Z	Z	V	V
$\tau = 0.5$												
0.01	L	3.7	20	V	V	V	V	V	V	V	V	V
	R			V	V	V	V	V	V	V	V	V
0.05	L	3.7	20	V	V	V	V	V	Z	Z	V	V
	R			V	V	V	V	V	Z	C	V	V
0.10	L	3.7	10	V	V	V	V	V	Z	Z	V	V
	R			V	V	V	V	V	Z	Z	V	V
$\tau = 0.75$												
0.01	L	3.8	15	V	V	V	V	V	V	V	V	V
	R			V	V	V	V	V	V	V	V	V
0.05	L	3.8	15	V	V	V	V	V	Z	Z	V	V
	R			V	V	V	V	V	Z	C	V	V
0.10	L	3.8	2.5	V	V	V	C	V	Z	Z	V	V
	R			V	V	V	V	V	Z	Z	V	V

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

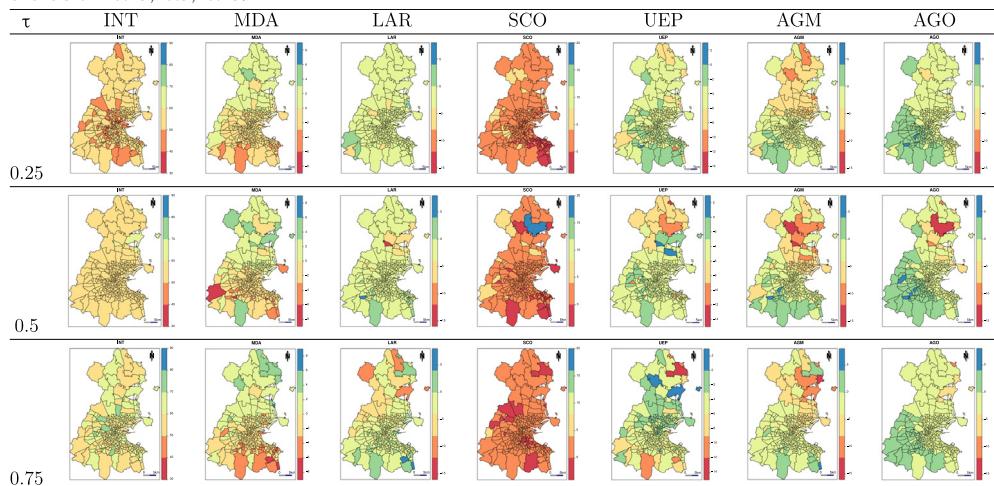
six relevant variables, the map of AGO is very similar at different quantile levels, which the positive association between AGO and GEI in the southwest region is stronger than that in the northeast. The difference is large for the maps of the other five variables and the intercept. Therefore, with multiple quantile levels being modeled, the GWQlasso method not only can provide the evidence to select relevant variables and analyze spatially non-stationarity relationship, but also can provide a thorough picture of how the characteristic of social structures are associated with the underlying conditional distribution of the voter turnout.

Additionally, the Dublin voter turnout data has been analyzed for fitting the GWR model and addressing the local collinearity problem in [Gollini et al. \(2015, section 7\)](#). To solve this issue, various penalized regression methods have been proposed to shrink regression coefficients to alleviate the effect of collinearity, such as the geographically weighted ridge regression ([Wheeler and Calder, 2007](#)), the geographically weighted lasso ([Wheeler, 2009](#)), the locally compensated ridge GWR ([Brunsdon et al., 2012](#)), the geographically weighted regularized regression ([Bárcena et al., 2014](#)), the GWQlasso method ([Wang and Li, 2017](#)) and so on. As a penalized method, we present a comparison of the coefficients estimates between the GWQlasso estimation and the GWQR estimation to illustrate the shrinkage effect in Figure 2 in supplemental materials. Actually, the GWQR model is different from the GWR model. To the best of our knowledge, there is no literature for addressing the local collinearity problem in using the GWQR model. Nevertheless, the simulation study indicates that the GWQlasso method is robust to the collinearity among the explanatory variables. Regardless of whether the local collinearity exists or not, the GWQlasso method can alleviate the effect of collinearity by shrinking the coefficients estimates of LOE and AGY to zero and the other coefficients estimates to be a small scope for the three different quantile levels. Thus Figure 2 further confirms that those variables eliminated by the GWQlasso method are unlikely to be relevant.

To investigate the necessity of using the GWQlasso method, the Dublin data set is further analyzed for fitting the identified model of the GWQlasso method and the full GWQR model under different quantile levels  $\tau = 0.25, 0.5, 0.75$ . The BIC-type criterion ([15](#)) is used to assess the fit of the two regression models. For  $\tau = 0.25$ , the BIC scores of the identified model and the GWQR model are 0.3513, 0.8635. For  $\tau = 0.5$ , the BIC scores of the two model are 0.3357, 0.7974. For  $\tau = 0.75$ , the BIC values of the models are 0.4106, 0.7061. The related results indicate that the proposed method is

**Table 4**

The GWQlasso estimates of the spatially varying coefficients for the intercept term and six relevant variables with three quantile levels  $\tau = 0.25, 0.5, 0.75$ .



able to provide a proper model specification information and the identified model gives the good fit for the Dublin data.

## 6. Conclusions and discussions

In the paper we propose a GWQlasso method for a shrinkage estimation of the geographically weighted quantile regression model. The proposed method is able to identify spatially varying coefficients, non-zero coefficients and zero coefficients simultaneously and estimate the corresponding varying and non-zero constant coefficients accurately. Simulation studies and real data analysis suggested that the proposed method is effective, which can provide a useful way to build a possible semi-parametric GWQR model for spatial data analysis. Finally, the R code of the local-linear GWQR technique and the GWQlasso method 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 developed for the GWQR model due to its simplicity. Similar ideas can be extended to other spatial models (McMillen, 2013; Shekhar et al., 2016). Secondly, shrinkage estimation of a semi-parametric GWQR model involves the quantile crossing problem. It is observed in Table 4 in real data analysis, the estimated quantile coefficients of the different quantile levels may cross. Furthermore, how to deal with the issue is another interesting topic open for discussion.

## Appendix A. Supplementary data

Supplementary material related to this article can be found online at <https://doi.org/10.1016/j.spa.2018.05.003>.

## References

- Bárcena, M.J., Menéndez, P., Palacios, M.B., Tusell, F., 2014. Alleviating the effect of collinearity in geographically weighted regression. *J. Geogr. Syst.* 16 (4), 441–466.
- Brunsdon, C., Charlton, M., Harris, P., 2012. Living with collinearity in local regression models. In: Proceedings of the 10th International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences. Brasil.
- Brunsdon, C., Fotheringham, A.S., Charlton, M., 1996. Geographically weighted regression: a method for exploring spatial non-stationarity. *Geogr. Anal.* 28 (4), 281–298.

- Brunsdon, C., Fotheringham, A.S., Charlton, M., 1999. Some notes on parametric significance tests for geographically weighted regression. *J. Reg. Sci.* 39 (3), 497–524.
- Cai, Z., Xiao, Z., 2012. Semiparametric quantile regression estimation in dynamic models with partially varying coefficients. *J. Econometrics* 167, 413–425.
- Chen, V.Y.J., Deng, W.S., Yang, T.C., Matthews, S.A., 2012. Geographically weighted quantile reffression (GWQR): An application to U.S. Mortality data. *Geogr. Anal.* 44, 134–150.
- Congdon, P., 2006. A model for non-parametric spatially varying regression effects. *Comput. Statist. Data Anal.* 50, 422–445.
- Congdon, P., 2007. Bayesian modelling strategies for spatially varying regression coefficients: A multivariate perspective for multiple outcomes. *Comput. Statist. Data Anal.* 51, 2586–2601.
- Fan, J., Li, R., 2001. Variable selection via nonconcave penalized likelihood and its oracle properties. *J. Amer. Statist. Assoc.* 96 (456), 1348–1360.
- Fotheringham, A.S., Brunsdon, C., Charlton, M., 2002. Geographically Weighted Regression: The Analysis of Spatial Varying Relationships. John Wiley, Chichester, Sussex.
- Fotheringham, A.S., Yang, W., Kang, W., 2017. Multiscale geographically weighted regression (MGWR). *Ann. Assoc. Am. Geogr.* 107 (6), 1247–1265.
- Gamerman, D., Moreira, A.R.B., Rue, H., 2003. Space-varying regression models: Specifications and simulation. *Comput. Statist. Data Anal.* 42, 513–533.
- Gelfand, A.E., Kim, H., Sirmans, C.F., Banerjee, S., 2003. Spatial modelling with spatially varying coefficient processes. *J. Amer. Statist. Assoc.* 98 (482), 387–396.
- Gollini, I., Lu, B., Charlton, M., Brunsdon, C., Harris, P., 2015. GWmodel: an R package for exploring spatial heterogeneity using geographically weighted models. *J. Stat. Softw.* 63 (17), 1–50.
- Hallin, M., Lu, Z., Yu, K., 2009. Local linear spatial quantile regression. *Bernoulli* 15, 659–686.
- Hansen, B.E., 2008. Uniform convergence rates for kernel estimation with dependent data. *Econom. Theor.* 24, 726–748.
- Harris, P., Brunsdon, C., Lu, B., Nakaya, T., Charlton, M., 2017. Introducing bootstrap methods to investigate coefficient non-stationarity in spatial regression models. *Spat. Stat.* 21, 241–261.
- Hohsuk, N., Kwanghun, C., Ingrid, V., 2012. Variable selection of varying coefficient models in quantile regression. *Electron. J. Stat.* 6, 1220–1238.
- Hunter, D.R., Lange, K., 2000. Quantile regression via an MM algorithm. *J. Comput. Graph. Statist.* 9 (1), 60–77.
- Kai, B., Li, R., Zou, H., 2011. New efficient estimation and variable selection method for semiparametric varying-coefficient partially linear models. *Ann. Statist.* 39 (1), 305–332.
- Koenker, R., 2005. Quantile Regression. Cambridge University Press, Cambridge.
- Koenker, R., 2017. quantreg: Quantile regression R package version 5.33.
- Koenker, R., Bassett, G., 1978. Regression quantiles. *Econometrica* 46 (1), 33–50.
- Leung, Y., Mei, C.-L., Zhang, W.-X., 2000. Statisticcal tests for spatial nonstationarity based on the geographically weighted regression model. *Environ. Plann. A* 32 (1), 9–32.
- Lu, B., Brunsdon, C., Charlton, M., Harris, P., 2017. Geographically weighted regression with parameter-specific distance metrics. *Int. J. Geogr. Inf. Sci.* 31 (5), 982–998.
- Lu, B., Charlton, M., Harris, P., Fotheringham, A.S., 2014. Geographically weighted regression with a non-Euclidean distance metric: A case study using hedonic house price data. *Int. J. Geogr. Inf. Sci.* 28 (4), 660–681.
- McMillen, D.P., 2013. Quantile Regression for Spatial Data. Springer, Heidelberg.
- Mei, C.-L., Wang, N., Zhang, W.-X., 2006. Testing the importance of the expalnatory variables in a mixed geographically weighted regression models. *Environ. Plann. A* 38 (3), 587–598.
- Mei, C.-L., Xu, M., Wang, N., 2016. A bootstrap test for constant coefficients in geographically weighted regression models. *Int. J. Geogr. Inf. Sci.* 30 (8), 1622–1643.
- Páez, A., Uchida, T., Miyamoto, K., 2002. A general framework for estimation and inference of geographically weighted regression models: Spatial association and model specification tests. *Environ. Plann. A* 34 (5), 883–904.
- Reich, B.J., Fuentes, M., Dunson, D.B., 2011. Bayesian spatial quantile regression. *J. Amer. Statist. Assoc.* 106 (483), 6–20.
- Shekhar, S., Feiner, S.K., Aref, W.G., 2016. Spatial computing. *Commun. ACM* 59 (1), 72–81.
- Shen, Y., Liang, H.Y., 2016. Quantile regression for partially linear varying-coefficient model with censoring indicators missing at random. *Comput. Statist. Data Anal.* 117, 1–18.
- Shen, S.-L., Mei, C.-L., Zhang, Y.-J., 2011. Spatially varying coefficient models: Testing for spatial heteroscedasticity and reweighting estimation of the coefficients. *Environ. Plann. A* 43 (7), 1723–1745.
- Sun, Y., Yan, H.J., Zhang, W.Y., Lu, Z., 2014. A semiparametric spatial dynamic models. *Ann. Statist.* 42 (2), 700–727.
- Tang, Y.L., Wang, H.X., Zhu, Z.Y., 2013. Variable selection in quantile varying coefficient models with longitudinal data. *Comput. Statist. Data Anal.* 57, 435–449.
- Wang, W.T., Li, D.K., 2017. Structure identification and variable selection in geographically weighted regression models. *J. Stat. Comput. Simul.* 87 (10), 2050–2068.
- Wang, K.N., Lin, L., 2016. Robust structure identification and variable selection in partial linear varying coefficient models. *J. Stat. Plan. Inference.* 174, 153–168.
- Wang, N., Mei, C.-L., Yan, X.-D., 2008. Local linear estimation of spatially varying coefficient models: An improvement on the geographically weighted regression technique. *Environ. Plann. A* 40 (4), 986–1005.
- Wang, H., Xia, Y., 2009. Shrinkage estimation of the varying coefficient model. *J. Amer. Statist. Assoc.* 104 (486), 747–757.
- Wheeler, D.C., 2009. Simultaneous coefficient penalization and model selection in geographically weighted regression: The geographically weighted lasso. *Environ. Plann. A* 41 (3), 722–742.

- Wheeler, D.C., Calder, C., 2007. An assessment of coefficient accuracy in linear regression models with spatially varying coefficients. *J. Geogr. Syst.* 9 (2), 145–166.
- Yang, W., Fotheringham, A.S., Harris, P., 2012. An Extension of Geographically Weighted Regression with Flexible Bandwidths. Lancaster, GISRUK, UK.
- Yang, Y., He, X., 2015. Quantile regression for spatially correlated data: An empirical likelihood approach. *Statist. Sinica* 25, 261–274.
- Yang, H., Lv, J., Guo, C.H., 2015. Weighted composite quantile regression estimation and variable selection for varying coefficient models with heteroscedasticity. *J. Korean Statist. Soc.* 44, 77–94.
- Yuan, M., Lin, Y., 2006. Model selection and estimation in regression with grouped variables. *J. R. Statist. Soc. Ser B.* 68 (1), 49–67.
- Zhang, H.-G., Mei, C.-L., 2011. Local least absolute deviation estimation of spatially varying coefficient models: Robust geographically weighted regression approaches. *Int. J. Geogr. Inf. Sci.* 25 (9), 1467–1489.
- Zhao, W.H., Zhang, R.Q., Liu, J.C., 2014. Regularization and model selection for quantile varying coefficient model with categorical effect modifiers. *Comput. Statist. Data Anal.* 79, 44–62.

## Further reading

- Knight, K., 1998. Limiting distributions for L1 regression estimators under general conditions. *Ann. Statist.* 26, 755–770.