Economic Modelling xxx (xxxx) xxx

Contents lists available at ScienceDirect

Economic Modelling

journal homepage: www.journals.elsevier.com/economic-modelling

Scale-adaptive estimation of mixed geographically weighted regression models

Feng Chen [a], Chang-Lin Mei [b,*]

[a] Department of Statistics, School of Mathematics and Statistics, Xi'an Jiaotong University, Xi'an, China
[b] School of Science, Xi'an Polytechnic University, Xi'an, China

ARTICLE INFO

JEL classification:
C13
C14
C21

Keywords:
Geographically weighted regression
Backfitting
Spatial scale
Bandwidth

ABSTRACT

Mixed geographically weighted regression (GWR) models, a combination of linear and spatially varying coefficient models, have found their applications in a variety of disciplines including economic modelling for geo-referenced data analysis. Generally, different explanatory variables may operate at different spatial scales, leading to different levels of spatial heterogeneity of the varying coefficients. To deal with such a multiscale problem, we propose a scale-adaptive method to calibrate mixed GWR models, in which a different bandwidth is separately assumed for each spatially varying coefficient and is selected based on the backfitting procedure. Extensive simulations with different spatial layouts and a real-world example based on the Dublin voter turnout data demonstrate that the scale-adaptive method can not only significantly improve the estimation accuracy of the spatially varying coefficients, but also provide valuable information on the scale at which each explanatory variable operates.

1. Introduction

Regression models are one of the most important tools in economic modelling. Because of the complexity of economic phenomena, spatial heterogeneity in a regression relationship is very common in many real-world problems such as regional housing market, policy-making and cost-benefit analysis (Holly et al., 2010; Oikarinen et al., 2018; Mauricio, 2019). Various kinds of regression models including multilevel models (Goldstein, 1986), random coefficient models (Longford, 1995) and geographically weighted regression (GWR) models (Brunsdon et al., 1996) have been developed to deal with spatial heterogeneity. Among these models, GWR models have been a very popular tool in exploring spatial heterogeneity of a regression relationship due to their flexibility and interpretability. A GWR model is of the form

$$y_i = \beta_0(u_i, v_i) + \sum_{j=1}^{r} \beta_j(u_i, v_i)x_{ij} + \varepsilon_i,\ i = 1, 2, \ldots, n,$$

where $\{y_i; x_{i1}, x_{i2}, \ldots, x_{ir}\}_{i=1}^{n}$ are the observations of the response variable Y and the explanatory variables $X_1, X_2, \ldots, X_r$ at n spatial sampling locations $\{(u_i, v_i)\}_{i=1}^{n}$, $\{\beta_j(u, v)\}_{j=0}^{r}$ are $r+1$ unknown spatially varying coefficients to be estimated, and $\{\varepsilon_i\}_{i=1}^{n}$ are independent and identically distributed errors with $\mathrm{E}(\varepsilon_i) = 0$ and $\mathrm{Var}(\varepsilon_i) = \sigma^2$. The spatially varying coefficients are estimated by the locally weighted least square procedure and the estimated coefficients provide the information about spatial heterogeneity of the regression relationship.

The GWR model assumes that all of the coefficients vary over space. In practice, however, the influence of some explanatory variables on the response variable may be spatially global, whereas that of the others is spatially local. To model such a regression relationship, Brunsdon et al. (1999) proposed a semi-parametric counterpart of the GWR model, called the mixed GWR model, in which some coefficients are assumed to be constant and the others vary over space. That is, the mixed GWR model can be represented as

$$y_i = \sum_{l=1}^{q} \alpha_l z_{il} + \sum_{j=1}^{p} \beta_j(u_i, v_i)x_{ij} + \varepsilon_i,\quad i = 1, 2, \ldots, n, \tag{1}$$

where we denote the constant coefficients by $\{\alpha_l\}_{l=1}^{q}$ and the corresponding explanatory variables by $Z_1, Z_2, \ldots, Z_q$. Generally, one takes $Z_1 = 1$ or $X_1 = 1$ to make the model contain a constant or spatially varying intercept.

As aforementioned, the traditional GWR model is calibrated by the locally weighted least square or kernel smoothing method in which the weights are generated by a kernel function with a single bandwidth size for all of the spatially varying coefficients (Brunsdon et al., 1996). For

* Corresponding author.
E-mail address: clmei@xpu.edu.cn (C.-L. Mei).

https://doi.org/10.1016/j.econmod.2020.02.015
Received 28 September 2019; Received in revised form 6 February 2020; Accepted 7 February 2020
Available online XXX
0264-9993/

the mixed GWR model in Equation (1), Brunsdon et al. (1999) treated the constant coefficient part and the spatially varying coefficient part separately to calibrate the model. Specifically, all of the spatially varying coefficients are simultaneously estimated by the traditional GWR procedure and the constant coefficients are estimated by the ordinary least square procedure. The two procedures are alternatively performed until a given convergence criterion is met. Afterwards, Fotheringham et al. (2002) proposed a two-step method to calibrate the mixed GWR model without iteration. The two-step method can not only reduce the computation time, but also yield closed forms of both constant and varying coefficient estimators. Nevertheless, a single bandwidth size is still used to estimate all of the spatially varying coefficients. In view of the identification of the constant coefficients in a mixed GWR model, Mei et al. (2004) formulated a test based on the residual sums of squares from the fits of the GWR model and its mixed counterpart; Lu et al. (2014b) derived a Monte Carlo test using the AIC criterion; and Mei et al. (2016) developed a residual-based bootstrap test without the assumption that the model errors are normally distributed. Nowadays, in addition to their applications in many other fields, both GWR and mixed GWR models have been widely applied to economic modelling (recently for example, Cho et al., 2010; Öcal and Yildirim, 2010; Heckert and Mennis, 2012; Lu et al., 2014a; Cao et al., 2019; Cortés and Iturra, 2019; Li et al., 2019; Salvati et al., 2019).

Spatial phenomena are intrinsically subject to scale effect (Gao and Bian, 2016). In general, different spatial processes may operate at different spatial scales (Fotheringham et al., 2017). And specific explanatory variables with different scales may have different effects on the response variable (Wu et al., 2019; Murakami et al., 2019). For example, the characteristics of houses are usually measured with different scales. Some variables are measured at the dwelling scale such as the average number of rooms and housing units; some are measured at the real estate level like the property fee; and some are measured at the census or town level such as the population density and crime rate (Wu et al., 2019). In the GWR model, each coefficient reflects the intensity of the corresponding explanatory variable on the response variable and the explanatory variables with different spatial scales may make the coefficients admit different levels of spatial heterogeneity. As well known in statistical varying coefficient models with non-parametric kernel smoothing estimation, a coefficient with strong variation needs a small bandwidth size while a smoother coefficient requires a relatively large bandwidth size in order to obtain efficient estimators for all of the coefficients (Fan and Zhang, 1999). This means that the bandwidth size can be used to quantify the heterogeneity level of each coefficient and different bandwidth sizes should be selected for the coefficients with different levels of heterogeneity. As aforementioned, however, both traditional calibration methods for the GWR and the mixed GWR models assume a single bandwidth size for all of the spatially varying coefficients. The selected optimal bandwidth size is therefore a balance among the different levels of spatial heterogeneity, leading to efficiency or accuracy loss in the estimators of the varying coefficients.

The above issue, termed as the multiscale problem, has been noted and well handled for the GWR model. Specifically, Fotheringham et al. (2017) and Leong and Yue (2017) employed the backfitting procedure (Hastie and Tibshirani, 1986) to calibrate the GWR model, which makes it possible to adaptively select an appropriate bandwidth size for each spatially varying coefficient. Meanwhile, the simulation studies which they conducted have demonstrated that, in contrast to the traditional GWR estimation, the backfitting-based method can significantly improve the accuracy of the coefficient estimators especially when the coefficients are of different levels of spatially heterogeneity. More importantly, this estimation method can provide valuable information on the scale at which each explanatory variable operates.

Although the mixed GWR model has partially considered the multiscale problem by allowing a subset of the coefficients to be constant and the others to vary over space, its existing calibration methods still have the limitation that all of the varying coefficients are assumed to be operated at a same spatial scale (Fotheringham et al., 2017; Wu et al., 2019). Given the wide application backgrounds of the mixed GWR model, the development of scale-adaptive calibration methods is therefore of great importance in obtaining efficient estimators for the coefficients and providing useful scale information for the explanatory variables. Motivated by the multiscale estimation of the GWR model in Fotheringham et al. (2017) and Leong and Yue (2017), a backfitting-based scale-adaptive estimation method is proposed in this paper for the mixed GWR model. Specifically, the two-step estimation method (Fotheringham et al., 2002) is used to obtain the initial estimators of both constant and varying coefficients. Then, the constant and varying coefficients are re-estimated by the backfitting procedure in which the optimal bandwidth size for each varying coefficient is separately selected by the commonly used AIC_c criterion.

The rest of this paper is organized as follows. In Section 2, the scale-adaptive calibration method for the mixed GWR model is introduced in details. In Section 3, extensive simulation studies with a comparison to the existing two-step estimation are conducted to assess the performance of the proposed estimation approach. A real-world example based on the Dublin voter turnout data is given in Section 4 to demonstrate the application of the proposed method. The paper is then concluded with a brief summary.

2. Scale-adaptive estimation for mixed GWR models

The mixed GWR model in Equation (1) can be re-expressed in matrix notation as

$$\boldsymbol{y} = \boldsymbol{Z}\boldsymbol{\alpha} + \sum_{j=1}^{p} \boldsymbol{\beta}_j(u,v) * \boldsymbol{x}_j + \boldsymbol{\varepsilon}, \tag{2}$$

where $\boldsymbol{y} = (y_1, y_2, \ldots, y_n)^{\mathrm{T}}$; $\boldsymbol{Z} = (\boldsymbol{z}_1, \boldsymbol{z}_2, \ldots, \boldsymbol{z}_q)$ with $\boldsymbol{z}_l = (z_{1l}, z_{2l}, \ldots, z_{nl})^{\mathrm{T}}$ $(l = 1, 2, \ldots, q)$; $\boldsymbol{x}_j = (x_{1j}, x_{2j}, \ldots, x_{nj})^{\mathrm{T}}$ $(j = 1, 2, \ldots, p)$; $\boldsymbol{\beta}_j(u,v) = \left(\beta_j(u_1,v_1), \beta_j(u_2,v_2), \ldots, \beta_j(u_n,v_n)\right)^{\mathrm{T}}$ $(j = 1, 2, \ldots, p)$; $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \ldots, \alpha_q)^{\mathrm{T}}$; $\boldsymbol{\varepsilon} = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n)^{\mathrm{T}}$; and "$*$" stands for the element-wise multiplication of two vectors such that, for each $j = 1, 2, \ldots, p$, $\boldsymbol{\beta}_j(u,v) * \boldsymbol{x}_j = \left(\beta_j(u_1,v_1)x_{1j}, \beta_j(u_2,v_2)x_{2j}, \ldots, \beta_j(u_n,v_n)x_{nj}\right)^{\mathrm{T}}$.

Let $\boldsymbol{X} = (\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_p)$, $\widetilde{\boldsymbol{x}}_i^{\mathrm{T}} = (x_{i1}, x_{i2}, \ldots, x_{ip})$, and

$$\boldsymbol{L}_v(h) = \begin{pmatrix} \widetilde{\boldsymbol{x}}_1^{\mathrm{T}}\left(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{W}_h(u_1,v_1)\boldsymbol{X}\right)^{-1}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{W}_h(u_1,v_1) \\ \widetilde{\boldsymbol{x}}_2^{\mathrm{T}}\left(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{W}_h(u_2,v_2)\boldsymbol{X}\right)^{-1}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{W}_h(u_2,v_2) \\ \vdots \\ \widetilde{\boldsymbol{x}}_n^{\mathrm{T}}\left(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{W}_h(u_n,v_n)\boldsymbol{X}\right)^{-1}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{W}_h(u_n,v_n) \end{pmatrix},$$

where $\boldsymbol{W}_h(u_i,v_i) = \mathrm{Diag}\left(K(d_{i1}/h), K(d_{i2}/h), \ldots, K(d_{in}/h)\right)$ with $K(\cdot)$ being a kernel function, d_{ij} being the Euclidean distance between the locations (u_i,v_i) and (u_j,v_j), and h being the bandwidth. According to the two-step estimation in Fotheringham et al. (2002), the estimators of the constant coefficients are obtained by

$$\begin{aligned} \widehat{\boldsymbol{\alpha}} &= \left(\widehat{\alpha}_1, \widehat{\alpha}_2, \ldots, \widehat{\alpha}_q\right)^{\mathrm{T}} \\ &= \left(\boldsymbol{Z}^{\mathrm{T}}(\boldsymbol{I}_n - \boldsymbol{L}_v(h))^{\mathrm{T}}(\boldsymbol{I}_n - \boldsymbol{L}_v(h))\boldsymbol{Z}\right)^{-1}\boldsymbol{Z}^{\mathrm{T}}(\boldsymbol{I}_n - \boldsymbol{L}_v(h))^{\mathrm{T}}(\boldsymbol{I}_n - \boldsymbol{L}_v(h))\boldsymbol{y}, \end{aligned} \tag{3}$$

where $\boldsymbol{I}_n$ is the identity matrix of order n.

The estimators of the varying coefficients at the location (u_i,v_i) are derived by

$$\begin{aligned} \widehat{\boldsymbol{\beta}}(u_i,v_i) &= \left(\widehat{\beta}_1(u_i,v_i), \widehat{\beta}_2(u_i,v_i), \ldots, \widehat{\beta}_p(u_i,v_i)\right)^{\mathrm{T}} \\ &= \left(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{W}_h(u_i,v_i)\boldsymbol{X}\right)^{-1}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{W}_h(u_i,v_i)\left(\boldsymbol{y} - \boldsymbol{Z}\widehat{\boldsymbol{\alpha}}\right), \ i = 1, 2, \ldots, n. \end{aligned} \tag{4}$$

The optimal bandwidth size, denoted by h^*, is selected by the AIC_c criterion, that is,

$$h^* = \underset{h>0}{\mathrm{argmin}}\, \mathrm{AIC_c}(h), \tag{5}$$

where

$$\mathrm{AIC_c}(h) = \log(\hat{\sigma}^2(h)) + \frac{n + \mathrm{tr}(\boldsymbol{H}(h))}{n - 2 - \mathrm{tr}(\boldsymbol{H}(h))}, \tag{6}$$

$$\hat{\sigma}^2(h) = \frac{1}{n}\boldsymbol{y}^{\mathrm{T}}(\boldsymbol{I}_n - \boldsymbol{H}(h))^{\mathrm{T}}(\boldsymbol{I}_n - \boldsymbol{H}(h))\boldsymbol{y},$$

$$\boldsymbol{H}(h) = (\boldsymbol{I}_n - \boldsymbol{L}_v(h))\boldsymbol{Z}\left(\boldsymbol{Z}^{\mathrm{T}}(\boldsymbol{I}_n - \boldsymbol{L}_v(h))^{\mathrm{T}}(\boldsymbol{I}_n - \boldsymbol{L}_v(h))\boldsymbol{Z}\right)^{-1}\boldsymbol{Z}^{\mathrm{T}}(\boldsymbol{I}_n - \boldsymbol{L}_v(h))^{\mathrm{T}} + \boldsymbol{L}_v(h),$$

and $\mathrm{tr}(\cdot)$ is the trace of a matrix.

Taking the above estimators of the constant and varying coefficients as their respective initial estimators, we re-estimate the constant and varying coefficients using the following backfitting procedure, in which all of the constant coefficients are simultaneously estimated in each iteration.

(i) Set $\hat{\boldsymbol{\alpha}}$ in Equation (3) to be an initial estimator of the constant coefficient vector $\boldsymbol{\alpha}$, which we denote by $\boldsymbol{\alpha}^{(0)}$, and let $\tilde{\boldsymbol{y}}^{(0)} = \boldsymbol{y} - \boldsymbol{Z}\boldsymbol{\alpha}^{(0)}$. Construct the GWR model as

$$\tilde{\boldsymbol{y}}^{(0)} = \sum_{j=1}^{p}\boldsymbol{\beta}_j(u,v) * \boldsymbol{x}_j + \boldsymbol{\varepsilon}. \tag{7}$$

(ii) Set the initial values $\boldsymbol{\beta}_j^{(0)}(u,v)\ (j = 1,2,\ldots,p)$ of $\boldsymbol{\beta}_j(u,v)(j = 1,2,\ldots,p)$ to be their respective estimators from the two-step method. That is, let

$$\begin{aligned}\boldsymbol{\beta}_j^{(0)}(u,v) &= \left(\beta_j^{(0)}(u_1,v_1), \beta_j^{(0)}(u_2,v_2), \ldots, \beta_j^{(0)}(u_n,v_n)\right)^{\mathrm{T}} \\ &= \left(\hat{\beta}_j(u_1,v_1), \hat{\beta}_j(u_2,v_2), \ldots, \hat{\beta}_j(u_n,v_n)\right)^{\mathrm{T}},\end{aligned} \tag{8}$$

where $\hat{\beta}_j(u_i,v_i)\ (j = 1,2,\ldots,p; i = 1,2,\ldots,n)$, obtained by Equation (4), are the two-step estimators of the varying coefficients at n sampling locations.

(iii) Given each $k = 1,2,\ldots,p$, rewrite the model in Equation (7) as

$$\tilde{\boldsymbol{y}}^{(0)} - \sum_{j=1,j\neq k}^{p}\boldsymbol{\beta}_j(u,v) * \boldsymbol{x}_j = \boldsymbol{\beta}_k(u,v) * \boldsymbol{x}_k + \boldsymbol{\varepsilon}. \tag{9}$$

Substituting the initial values of $\boldsymbol{\beta}_j(u,v)(j = 1,\ldots,k-1,k+1,\ldots,p)$ into Equation (9) yields the following GWR model with the single explanatory variable X_k:

$$\boldsymbol{y}_k^{(0)} \triangleq \tilde{\boldsymbol{y}}^{(0)} - \sum_{j=1,j\neq k}^{p}\boldsymbol{\beta}_j^{(0)}(u,v) * \boldsymbol{x}_j = \boldsymbol{\beta}_k(u,v) * \boldsymbol{x}_k + \boldsymbol{\varepsilon}. \tag{10}$$

According to the traditional GWR estimation (Brunsdon et al., 1996), the ith element $\beta_k(u_i,v_i)$ in $\boldsymbol{\beta}_k(u,v)$ is estimated by

$$\hat{\beta}_k(u_i,v_i) = \left(\boldsymbol{x}_k^{\mathrm{T}}\boldsymbol{W}_{h_k}(u_i,v_i)\boldsymbol{x}_k\right)^{-1}\boldsymbol{x}_k^{\mathrm{T}}\boldsymbol{W}_{h_k}(u_i,v_i)\boldsymbol{y}_k^{(0)}, \tag{11}$$

where $\boldsymbol{W}_{h_k}(u_i,v_i)$ is same as $\boldsymbol{W}_h(u_i,v_i)$ except that h_k here is the bandwidth size for the kth varying coefficient. Denote

$$\boldsymbol{\beta}_k^{(1)}(u,v) = \left(\hat{\beta}_k(u_1,v_1), \hat{\beta}_k(u_2,v_2), \ldots, \hat{\beta}_k(u_n,v_n)\right)^{\mathrm{T}} = \boldsymbol{P}_k(h_k)\boldsymbol{y}_k^{(0)}, \tag{12}$$

where

$$\boldsymbol{P}_k(h_k) = \begin{pmatrix}\left(\boldsymbol{x}_k^{\mathrm{T}}\boldsymbol{W}_{h_k}(u_1,v_1)\boldsymbol{x}_k\right)^{-1}\boldsymbol{x}_k^{\mathrm{T}}\boldsymbol{W}_{h_k}(u_1,v_1) \\ \left(\boldsymbol{x}_k^{\mathrm{T}}\boldsymbol{W}_{h_k}(u_2,v_2)\boldsymbol{x}_k\right)^{-1}\boldsymbol{x}_k^{\mathrm{T}}\boldsymbol{W}_{h_k}(u_2,v_2) \\ \vdots \\ \left(\boldsymbol{x}_k^{\mathrm{T}}\boldsymbol{W}_{h_k}(u_n,v_n)\boldsymbol{x}_k\right)^{-1}\boldsymbol{x}_k^{\mathrm{T}}\boldsymbol{W}_{h_k}(u_n,v_n)\end{pmatrix}.$$

The hat matrix for the GWR model in Equation (10) is then

$$\boldsymbol{H}_k(h_k) = \mathrm{Diag}\,(x_{1k}, x_{2k}, \ldots, x_{nk})\,\boldsymbol{P}_k(h_k). \tag{13}$$

With the hat matrix in Equation (13), the optimal bandwidth size of h_k, which we denote by $h_k^{(1)}$, is determined by the $\mathrm{AIC_c}$ criterion, that is,

$$h_k^{(1)} = \arg\min_{h_k>0}\mathrm{AIC_c}(h_k), \tag{14}$$

where

$$\mathrm{AIC_c}(h_k) = \log(\hat{\sigma}^2(h_k)) + \frac{n + \mathrm{tr}(\boldsymbol{H}_k(h_k))}{n - 2 - \mathrm{tr}(\boldsymbol{H}_k(h_k))},$$

and

$$\hat{\sigma}^2(h_k) = \frac{1}{n}\left(\boldsymbol{y}_k^{(0)}\right)^{\mathrm{T}}(\boldsymbol{I}_n - \boldsymbol{H}_k(h_k))^{\mathrm{T}}(\boldsymbol{I}_n - \boldsymbol{H}_k(h_k))\boldsymbol{y}_k^{(0)}.$$

(iv) After Step (iii) is completed for each k, we obtain the first iterative value $\boldsymbol{\beta}_k^{(1)}(u,v)$ of $\boldsymbol{\beta}_k(u,v)$ and the optimal bandwidth size $h_k^{(1)}$ for each varying coefficient. Then, the estimator of $\boldsymbol{\alpha} = (\alpha_1,\alpha_2,\ldots,\alpha_q)^{\mathrm{T}}$ is updated by

$$\boldsymbol{\alpha}^{(1)} = \left(\boldsymbol{Z}^{\mathrm{T}}\boldsymbol{Z}\right)^{-1}\boldsymbol{Z}^{\mathrm{T}}\left(\boldsymbol{y} - \sum_{k=1}^{p}\boldsymbol{\beta}_k^{(1)}(u,v) * \boldsymbol{x}_k\right). \tag{15}$$

(v) Replace $\tilde{\boldsymbol{y}}^{(0)}$ and $\{\boldsymbol{\beta}_k^{(0)}(u,v)\}_{k=1}^{p}$ in Step (iii) by $\tilde{\boldsymbol{y}}^{(1)} = \boldsymbol{y} - \boldsymbol{Z}\boldsymbol{\alpha}^{(1)}$ and $\{\boldsymbol{\beta}_k^{(1)}(u,v)\}_{k=1}^{p}$ in Equation (12) respectively, and repeat Steps (iii) and (iv) until a given convergence threshold, say δ, is reached. Specifically, let $\boldsymbol{\alpha}^{(t-1)}$, $\boldsymbol{\beta}_k^{(t-1)}(u,v)$, $\boldsymbol{\alpha}^{(t)}$ and $\boldsymbol{\beta}_k^{(t)}(u,v)$ be two successive iterative values of $\boldsymbol{\alpha}$ and $\{\boldsymbol{\beta}_k(u,v)\}_{k=1}^{p}$. If

$$\left(\left(\boldsymbol{\alpha}^{(t)} - \boldsymbol{\alpha}^{(t-1)}\right)^{\mathrm{T}}\left(\boldsymbol{\alpha}^{(t)} - \boldsymbol{\alpha}^{(t-1)}\right) + \frac{1}{n}\sum_{k=1}^{p}\left(\boldsymbol{\beta}_k^{(t)}(u,v) - \boldsymbol{\beta}_k^{(t-1)}(u,v)\right)^{\mathrm{T}}\left(\boldsymbol{\beta}_k^{(t)}(u,v) - \boldsymbol{\beta}_k^{(t-1)}(u,v)\right)\right)^{\frac{1}{2}} \leq \delta, \tag{16}$$

then we take $\boldsymbol{\alpha}^{(t)}$ and $\{\boldsymbol{\beta}_k^{(t)}(u,v)\}_{k=1}^{p}$ as the final estimators of $\boldsymbol{\alpha}$ and $\{\boldsymbol{\beta}_k(u,v)\}_{k=1}^{p}$, and the corresponding optimal bandwidth sizes $\{h_k^{(t)}\}_{k=1}^{p}$, which we denote by $\{h_k^{*}\}_{k=1}^{p}$, as the final bandwidth sizes of the varying coefficients.

3. Simulation studies

In this section, simulation studies with different spatial layouts and different kinds of spatial weights are conducted to assess the performance of the scale-adaptive estimation for the mixed GWR model. In addition, to examine the impact of collinearity between the explanatory variables on the estimation accuracy of the spatially varying coefficients, different levels of collinearity between the explanatory variables are also considered.

3.1. Simulation 1: regular lattice sampling points with fixed bandwidth weights

3.1.1. The spatial layout and spatial weights

We took a unit square as the studied region and fit the origin of a Cartesian coordinate system at the bottom-left corner of the unit square with the two axes being along with its mutually orthogonal sides. The sampling points were designated as the lattice points with the coordinates being

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

where $mod\,(a,b)$ and $\mathrm{int}\,(a,b)$ are the remainder and the integer part of a divided by b, respectively, leading to the sample size $n = m^2$. In

this simulation, the spatial weights at each sampling point were generated by the Gaussian function with fixed bandwidth. That is, the spatial weights at each (u_i, v_i) are

$$w_{ij} = K\left(d_{ij}/h\right) = \exp\left(-\frac{1}{2}\left(\frac{d_{ij}}{h}\right)^2\right),\ j = 1, 2, \ldots, n, \tag{17}$$

where d_{ij} is the Euclidean distance between (u_i, v_i) and (u_j, v_j) and h is the bandwidth. The value of m was set to be 21 and 25, respectively, meaning that two sample sizes $n = 441$ and 625 were considered.

3.1.2. The models for generating data

The mixed GWR models for generating data were designed as

$$y_i = \alpha_1 z_{i1} + \alpha_2 z_{i2} + \beta_1(u_i, v_i) + \beta_2(u_i, v_i)x_{i2} + \beta_3(u_i, v_i)x_{i3} + \varepsilon_i,$$
$$i = 1, 2, \ldots, n. \tag{18}$$

The constant coefficients α_1 and α_2 were set to be 4 and 5, respectively, and the following three groups of the spatially varying coefficients were considered.

Group 1:

$$\begin{cases} \beta_1(u,v) &= 2 + 5(4u-2)exp(-(2-4u)^2 - (2-4v)^2), \\ \beta_2(u,v) &= 4(\sin(\pi(u-0.5))\sin(\pi v))^2, \\ \beta_3(u,v) &= 2 + 2(u-v). \end{cases}$$

Group 2:

$$\begin{cases} \beta_1(u,v) = 2 + \frac{1}{7}\exp(u+v), \\ \beta_2(u,v) = 2 + \frac{1}{2}(u-v), \\ \beta_3(u,v) = 2 + 5(4u-2)\exp\left(-(2-4u)^2 - (2-4v)^2\right). \end{cases}$$

Group 3:

$$\begin{cases} \beta_1(u,v) = 2 + \frac{1}{6}(u-v), \\ \beta_2(u,v) = 2 + \frac{1}{2}\sin(\pi u), \\ \beta_3(u,v) = 2 + 5(4u-2)\exp\left(-(2-4u)^2 - (2-4v)^2\right). \end{cases}$$

The true surfaces of the three groups of the spatially varying coefficients are depicted in Fig. 1 to visually understand their respective variation patterns.

The observations $\{z_{i1}\}_{i=1}^n$ and $\{z_{i2}\}_{i=1}^n$ of the explanatory variables Z_1 and Z_2 were independently drawn from the standard normal distribution $N(0,1)$. The observations of X_2 and X_3 were generated in the following way. Let

$$\begin{pmatrix} X_2 \\ X_3 \end{pmatrix} = \begin{pmatrix} 1 & \gamma \\ \gamma & 1 \end{pmatrix}\begin{pmatrix} D_1 \\ D_2 \end{pmatrix}, \tag{19}$$

where $0 \le \gamma < 1$, and D_1 and D_2 are independent random variables with the common distribution $N(0,1)$. It is easy to derive that the correlation coefficient between X_2 and X_3 is

$$\rho(X_2, X_3) = \frac{\mathrm{Cov}(X_2, X_3)}{\sqrt{\mathrm{Var}(X_2)\,\mathrm{Var}(X_3)}} = \frac{2\gamma}{1+\gamma^2}.$$

In the simulation, three levels of collinearity between X_2 and X_3 were considered:

(i) $\gamma = 0$, meaning that X_2 and X_3 are mutually independent.
(ii) $\gamma = 0.27$, yielding $\rho(X_2, X_3) \approx 0.5$, a moderate correlation between X_2 and X_3.
(iii) $\gamma = 0.63$, leading to $\rho(X_2, X_3) \approx 0.9$, a strong correlation between X_2 and X_3.

Given each value of γ, we drew independently the observations of D_1 and D_2 from $N(0,1)$ and generated the observations $\{x_{i2}\}_{i=1}^n$ and $\{x_{i3}\}_{i=1}^n$ of X_2 and X_3 according to Equation (19). Each experimental setting was repeated for N times, in which the observations of the explanatory variables were fixed and $\{\varepsilon_i\}_{i=1}^n$ were independently drawn from $N(0,1)$ in each replication.

3.1.3. The indices for measuring accuracy of the coefficient estimators

For the constant coefficients, we used the usual indices including mean, standard deviation and root mean square error to measure the estimation accuracy. Let $\widehat{\alpha}_{l(r)}$ be the estimator of the lth constant coefficient in the rth replication. The mean value of the estimators of α_l in the N replications is defined by

$$\mathrm{Mean}\left(\widehat{\alpha}_l\right) = \frac{1}{N}\sum_{r=1}^{N}\widehat{\alpha}_{l(r)},\ l = 1, 2, \ldots, q,$$

the standard deviation is

$$\mathrm{SD}\left(\widehat{\alpha}_l\right) = \left(\frac{1}{N}\sum_{r=1}^{N}\left(\widehat{\alpha}_{l(r)} - \mathrm{Mean}\left(\widehat{\alpha}_l\right)\right)^2\right)^{\frac{1}{2}},\ l = 1, 2, \ldots, q,$$

and the root mean square error is

$$\mathrm{RMSE}(\widehat{\alpha}_l) = \left(\frac{1}{N}\sum_{r=1}^{N}\left(\widehat{\alpha}_{l(r)} - \alpha_l\right)^2\right)^{\frac{1}{2}},\ l = 1, 2, \ldots, q.$$

For each spatially varying coefficient, the root mean square error at each sampling point and the averaged value of the root mean square errors over all of the n sampling points were taken as the estimation accuracy measures. Let $\widehat{\beta}_{k(r)}(u_i, v_i)$ be the estimator of the kth varying coefficient at location (u_i, v_i) in the rth replication. The root mean square error of the estimators of $\beta_k(u_i, v_i)$ at location (u_i, v_i) in the N replications is

$$\mathrm{RMSE}\left(\widehat{\beta}_k(u_i, v_i)\right) = \left(\frac{1}{N}\sum_{r=1}^{N}\left(\widehat{\beta}_{k(r)}(u_i, v_i) - \beta_k(u_i, v_i)\right)^2\right)^{\frac{1}{2}},\ i = 1, 2, \ldots, n,$$

and the averaged value of the root mean squared errors for the kth spatially varying coefficient is defined by

$$\mathrm{ARMSE}_k = \frac{1}{n}\sum_{i=1}^{n}\mathrm{RMSE}\left(\widehat{\beta}_k(u_i, v_i)\right),\ k = 1, 2, \ldots, p.$$

3.1.4. The designation of the experimental parameter values

For each experimental setting, the replication number was set to be $N = 500$ and the convergence threshold in the scale-adaptive method was taken as $\delta = 0.001$. In this simulation, both the proposed and the two-step methods were performed for the purpose of comparison.

3.1.5. The simulation results with analysis

Estimation of the constant coefficients. The values of the indices for the constant coefficients from both estimation methods are reported in Table 1, where "Scale-adaptive" refers to the proposed method and "Two-step" indicates the two-step method.

It can be observed from Table 1 that the two methods perform almost equally well in estimating the constant coefficients no matter whether the explanatory variables in the spatially varying coefficient part are mutually independent or correlated. It is reasonable that collinearity between the explanatory variables in the varying coefficient part has little influence on the constant coefficient estimators. Moreover, it is interesting that no obvious improvement is achieved by iteratively estimating the constant coefficients.

Bandwidth selection. Figs. 2–4 show the boxplots of the optimal bandwidth sizes selected in the two-step and scale-adaptive methods in $N = 500$ replications for the three groups of the varying coefficients, in

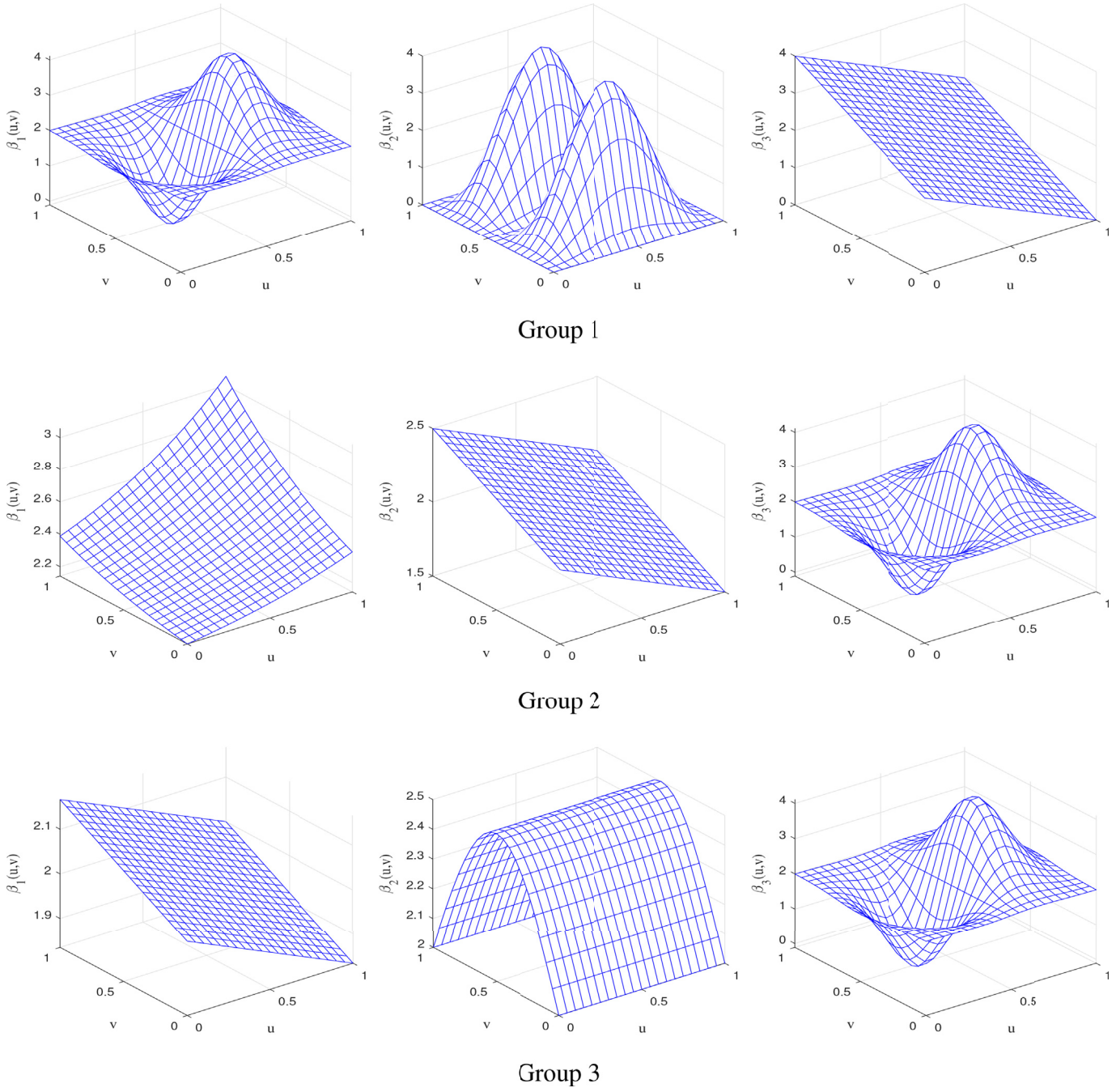


Fig. 1. True surfaces of the spatially varying coefficients in the three groups.

which h^* represents the optimal bandwidth size in the two-step method, and h_1^*, h_2^* and h_3^* denote the respective optimal bandwidth sizes of $\beta_1(u,v)$, $\beta_2(u,v)$ and $\beta_3(u,v)$ in the scale-adaptive method when the convergence criterion is met in each replication.

For the varying coefficients in Group 1, it can be observed from Fig. 1 that $\beta_1(u,v)$ and $\beta_2(u,v)$ are of a similar level of spatial heterogeneity, while $\beta_3(u,v)$ is much smoother than $\beta_1(u,v)$ and $\beta_2(u,v)$. As shown in Fig. 2, smaller bandwidth sizes are generally selected in the scale-adaptive method for $\beta_1(u,v)$ and $\beta_2(u,v)$ and much larger bandwidth sizes are chosen for $\beta_3(u,v)$ in the 500 replications. This result accords with the common finding in non-parametric smoothing techniques that a smoother function needs a larger bandwidth size and a function with a high level of heterogeneity needs a smaller bandwidth size in order to obtain their efficient estimators (Fan and Zhang, 1999).

For the two-step method, however, the optimal bandwidth sizes are generally between the smallest and largest ones in the scale-adaptive method, implying that the optimal bandwidth size in the two-step method is a trade-off of those for different levels of spatial heterogeneity of the varying coefficients. Furthermore, it seems that the optimal bandwidth size in the two-step method is mainly controlled by the coefficients with a high level of spatial heterogeneity. This finding can be observed from Fig. 2 where the values of h^* in the two-step method are more similar to those of h_1^* and h_2^* in the scale-adaptive method. For the other two groups of the varying coefficients, the similar finding can be observed from the optimal bandwidth sizes in the two estimation methods. In conclusion, the scale-adaptive method is capable of selecting proper bandwidth sizes for the coefficients with different levels of spatial heterogeneity.

Table 1
Means, standard deviations and root mean square errors of the constant coefficient estimators in the 500 replications for Simulation 1.

n	Varying coefficient	ρ	Method	$\alpha_1 = 4$			$\alpha_2 = 5$		
				Mean	SD	RMSE	Mean	SD	RMSE
441	Group 1	0	Scale-adaptive	3.9881	0.0687	0.0697	5.0190	0.0630	0.0657
			Two-step	3.9931	0.0700	0.0703	5.0050	0.0640	0.0641
		0.5	Scale-adaptive	3.9914	0.0684	0.0688	5.0212	0.0629	0.0663
			Two-step	3.9941	0.0700	0.0702	5.0048	0.0640	0.0642
		0.9	Scale-adaptive	3.9996	0.0680	0.0679	5.0272	0.0630	0.0686
			Two-step	3.9943	0.0703	0.0704	5.0044	0.0643	0.0644
	Group 2	0	Scale-adaptive	4.0097	0.0669	0.0675	5.0005	0.0617	0.0616
			Two-step	4.0125	0.0683	0.0694	5.0009	0.0628	0.0628
		0.5	Scale-adaptive	4.0115	0.0670	0.0680	4.9993	0.0618	0.0618
			Two-step	4.0138	0.0684	0.0697	4.9973	0.0630	0.0629
		0.9	Scale-adaptive	4.0127	0.0670	0.0681	4.9978	0.0622	0.0622
			Two-step	4.0150	0.0686	0.0702	4.9930	0.0632	0.0635
	Group 3	0	Scale-adaptive	4.0218	0.0665	0.0700	5.0024	0.0616	0.0616
			Two-step	4.0127	0.0683	0.0694	5.0024	0.0628	0.0628
		0.5	Scale-adaptive	4.0222	0.0666	0.0701	5.0019	0.0618	0.0618
			Two-step	4.0147	0.0683	0.0698	4.9989	0.0629	0.0629
		0.9	Scale-adaptive	4.0220	0.0669	0.0704	5.0007	0.0621	0.0621
			Two-step	4.0165	0.0686	0.0705	4.9948	0.0632	0.0633
625	Group 1	0	Scale-adaptive	4.0193	0.0599	0.0629	4.9559	0.0640	0.0777
			Two-step	4.0120	0.0606	0.0617	4.9838	0.0648	0.0667
		0.5	Scale-adaptive	4.0181	0.0600	0.0626	4.9549	0.0638	0.0781
			Two-step	4.0127	0.0606	0.0618	4.9817	0.0649	0.0674
		0.9	Scale-adaptive	4.0132	0.0601	0.0614	4.9551	0.0634	0.0777
			Two-step	4.0137	0.0605	0.0619	4.9790	0.0650	0.0682
	Group 2	0	Scale-adaptive	4.0183	0.0588	0.0615	5.0186	0.0612	0.0639
			Two-step	4.0151	0.0596	0.0614	5.0098	0.0629	0.0636
		0.5	Scale-adaptive	4.0173	0.0589	0.0613	5.0196	0.0611	0.0641
			Two-step	4.0152	0.0597	0.0615	5.0085	0.0630	0.0635
		0.9	Scale-adaptive	4.0161	0.0590	0.0612	5.0225	0.0609	0.0649
			Two-step	4.0154	0.0598	0.0617	5.0065	0.0632	0.0634
	Group 3	0	Scale-adaptive	4.0092	0.0589	0.0596	5.0100	0.0615	0.0623
			Two-step	4.0135	0.0597	0.0611	5.0087	0.0629	0.0635
		0.5	Scale-adaptive	4.0074	0.0590	0.0594	5.0104	0.0614	0.0622
			Two-step	4.0139	0.0597	0.0612	5.0076	0.0629	0.0633
		0.9	Scale-adaptive	4.0064	0.0593	0.0596	5.0121	0.0612	0.0623
			Two-step	4.0145	0.0598	0.0615	5.0056	0.0631	0.0633

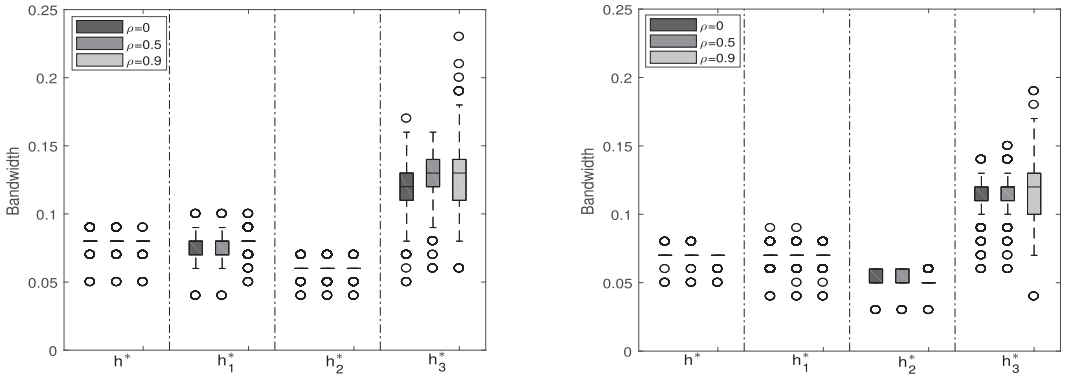
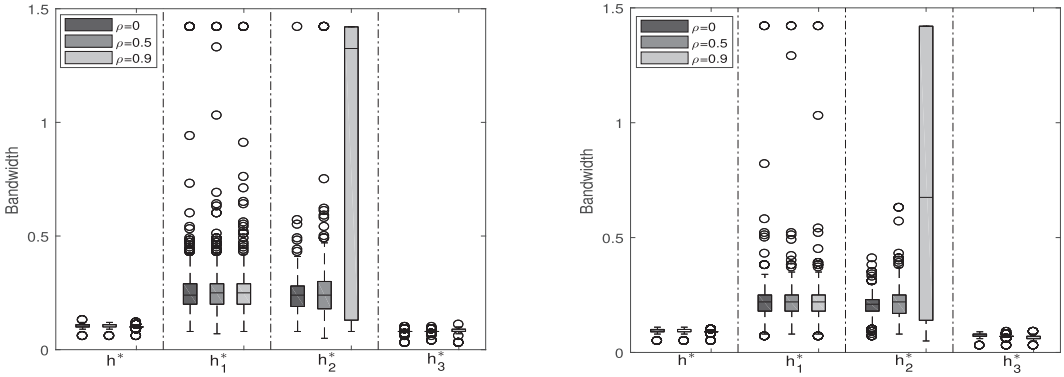


Fig. 2. Boxplots of the selected bandwidth sizes in the 500 replications for the coefficients of Group 1 in Simulation 1.

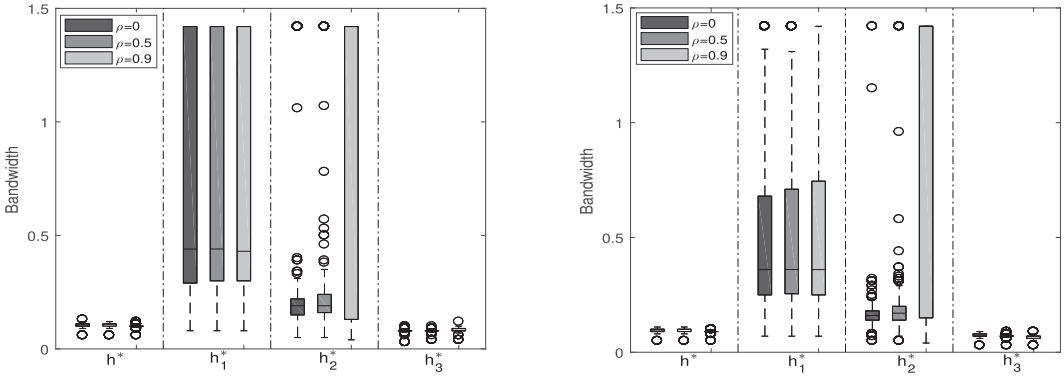
For the impact of collinearity between the explanatory variables on the optimal bandwidth sizes in the two methods, it seems that the optimal bandwidth sizes of the spatially varying coefficients with a lower level of spatial heterogeneity are more affected than those of the coefficients with a high level of spatial heterogeneity. As shown in Fig. 2, for example, when the value of $\rho(X_2, X_3)$ increases from 0 to 0.9, the uncertainty of the values of h_3^* increases obviously, while that of the values of h_1^* and h_2^* varies slightly. The similar finding is also observed from Figs. 3 and 4. When the sample size increases, however, the uncertainty of the optimal bandwidth sizes seems decreasing.

Estimation of the spatially varying coefficients. Table 2 reports the averaged values of the root mean squared errors of spatially varying coefficient estimators for the scale-adaptive and two-step methods. It is observed that the ARMSE values of the scale-adaptive estimators are consistently smaller than those of the two-step estimators in all of the experimental settings, indicating that the proposed method can significantly improve the estimation accuracy of the spatially varying coefficients. The most significant improvement is achieved for such coefficients that their optimal bandwidth sizes selected in the two methods are of large difference. Specifically, for the coefficients in Group 1, the values of h_3^* in the scale-adaptive method are in general much larger than those of h^* in the two-step method. Then, the accuracy improve-



(a) $n = 441$ (b) $n = 625$

Fig. 3. Boxplots of the selected bandwidth sizes in the 500 replications for the coefficients of Group 2 in Simulation 1.



(a) $n = 441$ (b) $n = 625$

Fig. 4. Boxplots of the selected bandwidth sizes in the 500 replications for the coefficients of Group 3 in Simulation 1.

ment for the estimator of $\beta_3(u,v)$ is most notable. For the coefficients in Groups 2 and 3, the values of h_1^* and h_2^* are generally much larger than those of h^* and the accuracy improvement for the estimators of $\beta_1(u,v)$ and $\beta_2(u,v)$ is obviously illustrated. For both methods, as expected, the estimation accuracy for all of the varying coefficients increases with the sample size increasing.

For the impact of collinearity between the explanatory variables on the estimation accuracy of the varying coefficients, it is known from Table 2 that, for both methods, strong collinearity leads to a decrease on the accuracy of the coefficient estimators. However, it seems that the influence is generally weaker for the scale-adaptive method. For example, in the case of $n = 441$, when the value of $\rho(X_2, X_3)$ increases from 0 to 0.9, the ARMSE values of $\beta_2(u,v)$ and $\beta_3(u,v)$ in Group 1 for the two-step method increase by 0.6712–0.4244 = 0.2468 and 0.5518–0.2496 = 0.3022, respectively, while those for the scale-adaptive method increase by 0.5401–0.3893 = 0.1508 and 0.4154–0.1968 = 0.2186, respectively. Nevertheless, this difference seems decreasing with the sample size increasing. The results demonstrate that the scale-adaptive method is more robust to the collinearity between the explanatory variables.

Table 2
Values of ARMSE for the spatially varying coefficient estimators in the 500 replications for Simulation 1.

n	Varying coefficient	ρ	Scale-adaptive			Two-step		
			$\beta_1(u,v)$	$\beta_2(u,v)$	$\beta_3(u,v)$	$\beta_1(u,v)$	$\beta_2(u,v)$	$\beta_3(u,v)$
441	Group 1	0	0.2584	0.3893	0.1968	0.2632	0.4244	0.2496
		0.5	0.2605	0.4100	0.2195	0.2666	0.4511	0.2854
		0.9	0.2642	0.5401	0.4154	0.2764	0.6712	0.5518
	Group 2	0	0.1104	0.1264	0.2515	0.1654	0.1772	0.2578
		0.5	0.1108	0.1449	0.2528	0.1682	0.2068	0.2739
		0.9	0.1113	0.2499	0.3174	0.1752	0.3876	0.4133
	Group 3	0	0.0877	0.1494	0.2524	0.1645	0.1812	0.2567
		0.5	0.0870	0.1599	0.2550	0.1668	0.2081	0.2757
		0.9	0.0862	0.2321	0.3003	0.1738	0.3871	0.4187
625	Group 1	0	0.2233	0.3441	0.1619	0.2285	0.3607	0.2248
		0.5	0.2238	0.3555	0.1842	0.2309	0.3776	0.2410
		0.9	0.2251	0.4667	0.3633	0.2384	0.5408	0.4336
	Group 2	0	0.0966	0.1120	0.2339	0.1464	0.1617	0.2419
		0.5	0.0967	0.1238	0.2377	0.1497	0.1833	0.2543
		0.9	0.0986	0.2214	0.2962	0.1602	0.3325	0.3646
	Group 3	0	0.0787	0.1292	0.2331	0.1466	0.1630	0.2419
		0.5	0.0793	0.1409	0.2406	0.1501	0.1844	0.2542
		0.9	0.0816	0.2062	0.2818	0.1613	0.3333	0.3643

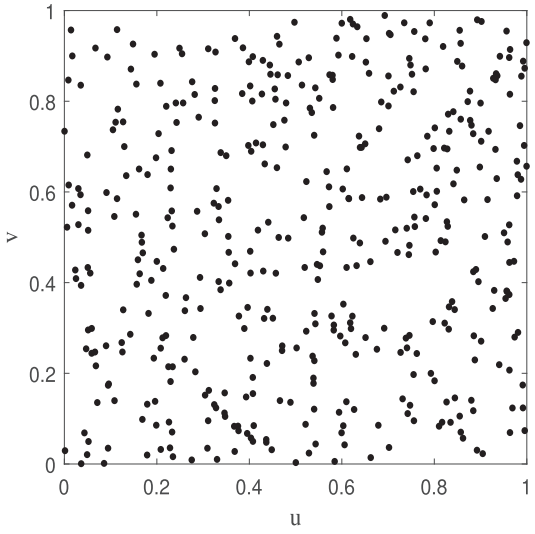


Fig. 5. Sampling points in Simulation 2.

Furthermore, the values of RMSE $\left(\widehat{\beta}_k(u_i, v_i)\right)$ $(k = 1, 2, 3)$ for each experimental setting can be depicted on the spatial region to make more comprehensive understanding on the root mean square errors for the two estimation methods. As an example, we drew the surfaces of RMSE $\left(\widehat{\beta}_k(u_i, v_i)\right)$ $(k = 1, 2, 3)$ for the sample size $n = 441$ and the coefficients in Group 1 and provided the figures as supplementary materials (see Section S1 in the supplementary materials) because of the limited space for the main body of the paper. The figures illustrate that larger estimation errors for both methods generally appear at the boundary area and the neighbourhoods of the locations where the coefficients reach their extreme values.

3.2. Simulation 2: irregular sampling points with variable bandwidth weights

3.2.1. The spatial layout and spatial weights

In this simulation, the same unit square as that in Simulation 1 was taken as the studied region, but each coordinate (u_i, v_i) of sampling points was determined by independently drawing a pair of random numbers from the uniform distribution $U(0, 1)$. Only the sample size $n = 441$ was considered in this simulation and the sampling points are depicted in Fig. 5.

As pointed out by Gollini et al. (2013), for irregular sampling points, the spatial weights with a variable or adapted bandwidth for the coefficient estimation perform better than the weights with a fixed bandwidth. Therefore, we generated the spatial weights in this simulation by a variable bandwidth. In order to make the simulation focus on examining the performance of the scale-adaptive method for irregular sampling points, the Gaussian function was once again used to generate the weights. In this case, however, h in Equation (17) is changed as d_{ik}, the Euclidean distance from the focal location (u_i, v_i) to its kth nearest sampling point, and the parameter k is treated as a "real" bandwidth to be selected in the calibration process.

The other experimental designs including the number of replications and convergence threshold were all kept to be the same as those in Simulation 1. For the purpose of comparison, the results of the two-step method were also computed.

3.2.2. The simulation results with analysis

Table 3 reports the results of the constant coefficient estimation for both methods. It is known from the table that the two methods estimate the constant coefficients equally well and all yield accurate estimators of the constant coefficients when the sampling points are irregularly distributed over the space.

Fig. 6 shows the boxplots of the optimal values of the parameter k, i.e., the optimal bandwidth sizes in the case of variable bandwidth weights. Compared with the corresponding results in Figs. 2–4 for $n = 441$, the similar variation patterns for the values of the parameter k can be observed, indicating that the scale-adaptive method can also provide valid information for the levels of spatial heterogeneity of the varying coefficients when the sampling points are irregularly distributed over space.

Table 4 shows the ARMSE values of the three groups of varying coefficients obtained by both methods. In contrast to the corresponding results in Table 2 for $n = 441$, although the ARMSE values for both methods show a slight increase in most cases, the ARMSE values from the scale-adaptive estimation are still consistently smaller than those from the two-step estimation.

In addition to the foregoing simulation studies, the other two simulation studies were also conducted according to one reviewer's suggestions. One is to use a real-world spatial layout with an irregular boundary of the region and irregularly distributed sampling points. Here, the Boston area with 506 US census tracts (Gilley and Pace, 1996) was taken as the spatial layout for this simulation study. The simulation results demonstrate that the proposed scale-adaptive method also performs very well under such a spatial layout. The details on the exper-

Table 3
Means, standard deviations and root mean square errors of the constant coefficient estimators in the 500 replications for Simulation 2.

Varying coefficient	ρ	Method	$\alpha_1 = 4$			$\alpha_2 = 5$		
			Mean	SD	RMSE	Mean	SD	RMSE
Group 1	0	Scale-adaptive	3.9906	0.0677	0.0683	5.0156	0.0628	0.0647
		Two-step	3.9838	0.0697	0.0715	5.0219	0.0641	0.0677
	0.5	Scale-adaptive	3.9868	0.0677	0.0689	5.0194	0.0625	0.0654
		Two-step	3.9804	0.0698	0.0724	5.0270	0.0642	0.0696
	0.9	Scale-adaptive	3.9804	0.0676	0.0703	5.0268	0.0623	0.0677
		Two-step	3.9769	0.0702	0.0738	5.0328	0.0643	0.0721
Group 2	0	Scale-adaptive	4.0024	0.0668	0.0668	5.0101	0.0619	0.0626
		Two-step	4.0037	0.0673	0.0673	5.0095	0.0624	0.0631
	0.5	Scale-adaptive	3.9952	0.0668	0.0669	5.0099	0.0619	0.0626
		Two-step	4.0004	0.0675	0.0675	5.0073	0.0624	0.0628
	0.9	Scale-adaptive	3.9907	0.0667	0.0673	5.0067	0.0617	0.0620
		Two-step	3.9957	0.0678	0.0679	5.0042	0.0627	0.0628
Group 3	0	Scale-adaptive	3.9920	0.0671	0.0675	5.0123	0.0620	0.0632
		Two-step	4.0015	0.0674	0.0674	5.0132	0.0624	0.0637
	0.5	Scale-adaptive	3.9850	0.0669	0.0685	5.0135	0.0619	0.0633
		Two-step	3.9990	0.0676	0.0675	5.0109	0.0623	0.0632
	0.9	Scale-adaptive	3.9840	0.0668	0.0686	5.0096	0.0617	0.0624
		Two-step	3.9950	0.0678	0.0680	5.0078	0.0626	0.0630

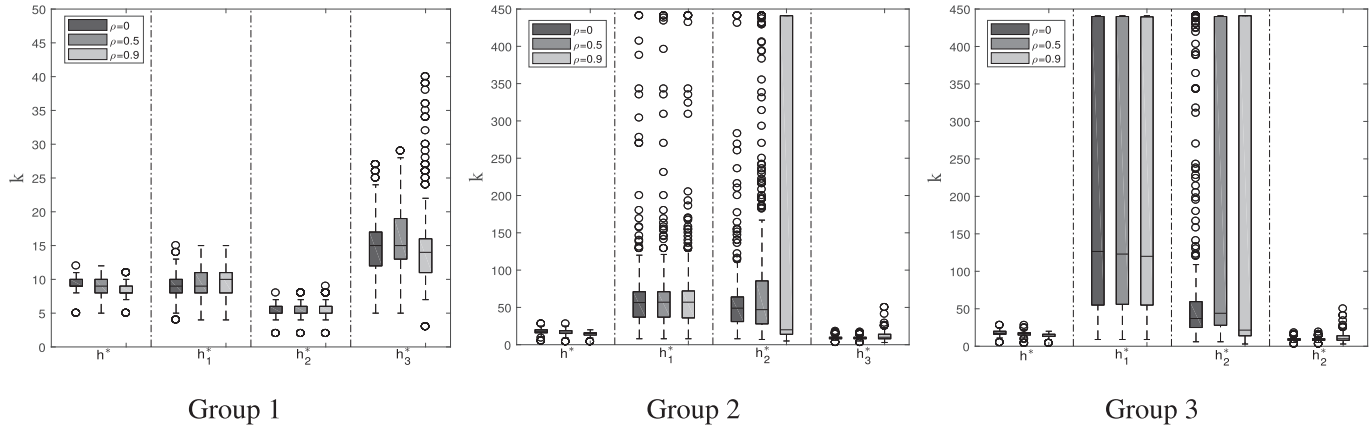


Fig. 6. Boxplots of the selected values of k in the 500 replications for the three groups of varying coefficients in Simulation 2.

Table 4
Values of ARMSE for the spatially varying coefficient estimators in the 500 replications for Simulation 2.

Varying coefficient	ρ	Scale-adaptive			Two-step		
		$\beta_1(u,v)$	$\beta_2(u,v)$	$\beta_3(u,v)$	$\beta_1(u,v)$	$\beta_2(u,v)$	$\beta_3(u,v)$
Group 1	0	0.2686	0.4167	0.2002	0.2908	0.4499	0.2395
	0.5	0.2697	0.4349	0.2221	0.2938	0.4708	0.2714
	0.9	0.2710	0.5563	0.3846	0.3007	0.6419	0.4933
Group 2	0	0.1170	0.1361	0.2910	0.1489	0.1654	0.3115
	0.5	0.1183	0.1575	0.2948	0.1541	0.1863	0.3258
	0.9	0.1202	0.2931	0.3690	0.1660	0.3649	0.4648
Group 3	0	0.0929	0.1521	0.2889	0.1466	0.1680	0.3064
	0.5	0.0932	0.1637	0.2956	0.1512	0.1875	0.3215
	0.9	0.0943	0.2705	0.3545	0.1627	0.3602	0.4591

imental design and the related results can be found in Section S2 of the supplementary materials. The other simulation study is to empirically investigate the performance of the scale-adaptive method when the traditional GWR estimation is replaced in the algorithm by the local-linear GWR estimation (Wang et al., 2008). The general findings are in what follows. The original GWR-based scale-adaptive and the local-linear-based one perform equally well on the constant coefficient estimation and on the selection of proper bandwidth sizes for varying coefficients. For the estimation of the varying coefficients, the local-linear-based method does reduce bias of the estimators, but increases their variance because more parameters including the coefficients themselves and their partial derivatives should be estimated at each location. As a result, the corresponding ARMSE values obtained from these two kinds of scale-adaptive methods are comparable. Because of the limited space, part of the simulation results with a brief description of the local-linear GWR estimation is provided as Section S3 in the supplementary materials.

4. Real-world example

The Dublin voter turnout data set, taken from Gollini et al. (2013), is used to demonstrate the application of the scale-adaptive method. This data set consists of nine percentage variables including voter turnout from the Irish 2004 Dáil elections and eight characteristics of the social structures in 322 electoral divisions of Greater Dublin. The related variables are explained in what follows.

- GenE12004: percentage of people who voted in the election in each electoral division;
- DiffAdd: percentage of one-year migrants in each electoral division;
- LARent: percentage of local authority renters in each electoral division;
- SC1: percentage of people with social class one (high social class) in each electoral division;
- Unempl: percentage of unemployed people in each electoral division;
- LowEduc: percentage of people without any formal education in each electoral division;
- Age1: percentage of people aged from 18 to 24 in each electoral division;
- Age2: percentage of people aged from 25 to 44 in each electoral division;
- Age3: percentage of people aged from 45 to 64 in each electoral division.

Lu et al. (2014b) built a GWR model with GenE12004 being the response variable and the other eight variables being explanatory variables to analyze spatial heterogeneity of the regression relationship, and a Monte Carlo test suggested that the coefficients of DiffAdd, LARent, LowEduc, Age2, Age3 and the Intercept are constant, while the coefficients of SC1, Unempl and Age1 vary over space. This test result specifies the following mixed GWR model:

$$\begin{aligned}\text{GenE12004}_i = {} & \alpha_1 + \alpha_2\text{DiffAdd}_i + \alpha_3\text{LARent}_i + \alpha_4\text{LowEduc}_i + \alpha_5\text{Age2}_i \\ & + \alpha_6\text{Age3}_i + \beta_1(u_i,v_i)\text{SC1}_i + \beta_2(u_i,v_i)\text{Unempl}_i \\ & + \beta_3(u_i,v_i)\text{Age1}_i + \varepsilon_i,\ i = 1,2,\ldots,322. \end{aligned} \tag{20}$$

We then employed both two-step and scale-adaptive methods to calibrate this model, in which the Gaussian kernel function with a variable bandwidth was used to generate the spatial weights, and the convergence threshold δ was set to be 0.001 in the scale-adaptive method.

The estimated constant coefficients by the two methods are listed in Table 5. It is known from the table that both methods yield the simi-

Table 5
Estimated constant coefficients in the model in Equation (20).

Variable Constant coefficient	Intercept α_1	DiffAdd α_2	LARent α_3	LowEduc α_4	Age2 α_5	Age3 α_6
Scale-adaptive	78.9289	−0.1262	−0.1313	−0.1573	−0.4237	−0.1104
Two-step	84.7056	−0.1371	−0.1124	0.0928	−0.5133	−0.2390

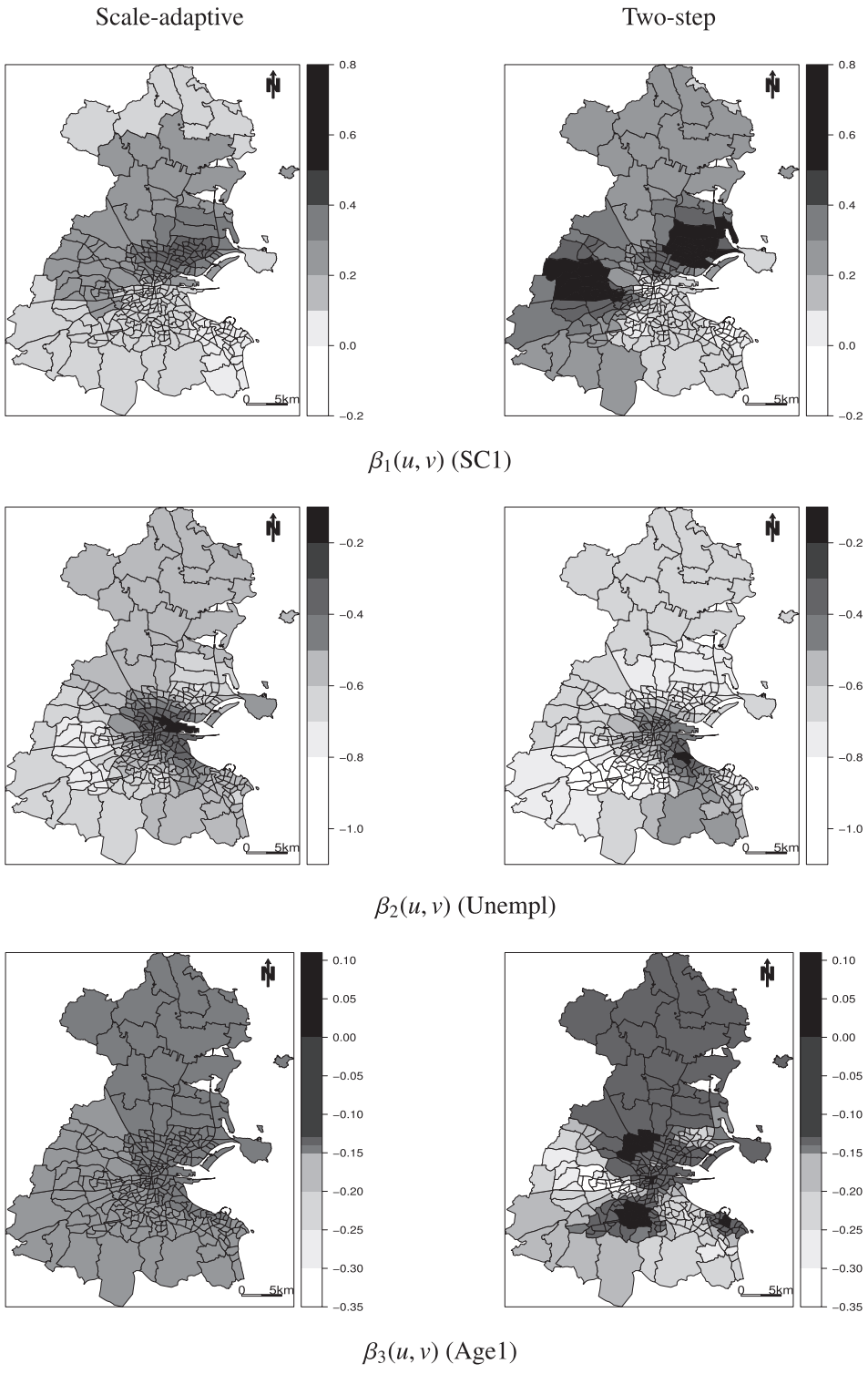


Fig. 7. Maps of the varying coefficient estimators obtained by the two estimation methods. The left-hand column is for the scale-adaptive method and the right-hand column is for the two-step method.

lar estimated values of the constant coefficients except for that of the coefficient of LowEduc. For the scale-adaptive method, the estimators of the constant coefficients are all negative, implying that the corresponding variables negatively influence the voting percentage consistently over the 322 electoral divisions. The results seem reasonable according to the practical meanings of these explanatory variables. For the two-step method, however, the positive estimator of α_4 apparently conflicts with the meaning of the corresponding explanatory variable LowEduc, although the significance of LowEduc influencing the response variable remains to be statistically tested. In this sense, the scale-adaptive method yields more reasonable estimators for the constant coefficients.

For the spatially varying coefficients of the variables SC1, Unempl and Age1, the optimal values of the parameter k in the weights are respectively 26, 13 and 218 for the scale-adaptive method, and 16 for the two-step method. The optimal values of k from the scale-adaptive method provide the information that the coefficient $\beta_3(u,v)$ of Age1 is much smoother than the other coefficients, implying that, in contrast to the impact of Age1 on the voting percentage, that of SC1 and Unempl varies dramatically over the 322 electoral divisions.

The maps of the estimators of the spatially varying coefficients are shown in Fig. 7. It is observed that the corresponding estimators of each varying coefficient by the two-step and scale-adaptive methods show quite different spatial patterns. In view of the fact that the scale-adaptive method is capable of selecting an appropriate bandwidth size for each varying coefficient, the estimation result from the scale-adaptive method is more convincible. From the maps in the left-hand column of Fig. 7, we known that SC1 has positive impact on the voting percentage at all electoral divisions, but the influence intensity varies over the divisions, showing a pattern of strong influence in the center part and weak influence in the north and south parts. Unempl negatively influences the voting percentage with the larger intensity appearing on the surrounding area of the city. Age1 has negative impact on the voting percentage with the intensity showing slight difference between the northeast and southwest parts.

5. Conclusions

In this paper, a scale-adaptive method is proposed to deal with the multiscale problem in mixed GWR models. The extensive simulations demonstrate that the proposed method not only achieves significant improvement on the estimation accuracy of the spatially varying coefficients, but also provides valuable information about the level of spatial heterogeneity of each varying coefficient or spatial scale that each explanatory variable operates. Moreover, the scale-adaptive method is quite robust to the spatial layout and collinearity between the explanatory variables. The real-world example based on the Dublin voter turnout data demonstrates the application potentials of the scale-adaptive estimation for mixed GWR models.

Declaration of competing interest

The authors declare that there are no conflicts of interest.

Acknowledgment

This work was supported by the National Natural Science Foundation of China [No. 11871056]. The authors would like to thank the Co-Editor Angus C. Chu and the reviewers for their valuable comments and suggestions which led to significant improvement on the manuscript.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.econmod.2020.02.015.

References

Brunsdon, C., Fotheringham, A.S., Charlton, M., 1996. Geographically weighted regression: a method for exploring spatial nonstationarity. Geogr. Anal. 28 (4), 281–298.
Brunsdon, C., Fotheringham, A.S., Charlton, M., 1999. Some notes on parametric significance test for geographically weighted regression. J. Reg. Sci. 39 (3), 497–524.
Cao, K., Diao, M., Wu, B., 2019. A big data–based geographically weighted regression model for public housing prices: a case study in Singapore. Ann. Am. Assoc. Geogr. 109 (1), 173–186.
Cho, S.H., Lambert, D.M., Roberts, R.K., Kim, S.G., 2010. Moderating urban sprawl: is there a balance between shared open space and housing parcel size? J. Econ. Geogr. 10 (5), 763–783.
Cortés, Y., Iturra, V., 2019. Market versus public provision of local goods: an analysis of amenity capitalization within the metropolitan region of Santiago de Chile. Cities 89, 92–104.
Fan, J., Zhang, W., 1999. Statistical estimation in varying coefficient models. Ann. Stat. 27 (5), 1491–1518.
Fotheringham, A.S., Brunsdon, C., Charlton, M., 2002. Geographically Weighted Regression: the Analysis of Spatially Varying Relationships. John Wiley and Sons, Chichester, UK.
Fotheringham, A.S., Yang, W., Kang, W., 2017. Multiscale geographically weighted regression (MGWR). Ann. Am. Assoc. Geogr. 107 (6), 1247–1265.
Gao, P., Bian, L., 2016. Scale effects on spatially embedded contact networks. Comput. Environ. Urban Syst. 59, 142–151.
Gilley, O.W., Pace, R.K., 1996. On the Harrison and Rubinfeld data. J. Environ. Econ. Manag. 31 (3), 403–405.
Goldstein, H., 1986. Multilevel mixed linear model analysis using iterative generalized least squares. Biometrika 73 (1), 43–56.
Gollini, I., Lu, B., Charlton, M., Brunsdon, C., Harris, P., 2013. GWmodel: an R package for exploring spatial heterogeneity using geographically weighted models. J. Stat. Software 63 (17), 1–52.
Hastie, T., Tibshirani, R., 1986. Generalized additive models. Stat. Sci. 1 (3), 297–310.
Heckert, M., Mennis, J., 2012. The economic impact of greening urban vacant land: a spatial difference-in-differences analysis. Environ. Plann. A: Econ. Space 44, 3010–3027.
Holly, S., Pesaran, M.H., Yamagata, T., 2010. A spatio-temporal model of house prices in the USA. J. Econom. 158, 160–173.
Leong, Y.Y., Yue, J.C., 2017. A modification to geographically weighted regression. Int. J. Health Geogr. 16 (11), 1–18.
Li, H., Wei, Y.D., Wu, Y., Tian, G., 2019. Analyzing housing prices in Shanghai with open data: amenity, accessibility and urban structure. Cities 91, 165–179.
Longford, N.T., 1995. Random coefficient models. In: Arminger, G., Clogg, C.C., Sobel, M.E. (Eds.), Handbook of Statistical Modeling for the Social and Behavioral Sciences. Springer, Boston.
Lu, B., Charlton, M., Harris, P., Fotheringham, A.S., 2014a. 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.
Lu, B., Harris, P., Charlton, M., Brunsdon, C., 2014b. The GWmodel R package: further topics for exploring spatial heterogeneity using geographically weighted models. Geo Spatial Inf. Sci. 17 (2), 85–101.
Mauricio, S., 2019. Do monetary subjective well-being evaluations vary across space? comparing continuous and discrete spatial heterogeneity. Spatial Econ. Anal. 14 (1), 53–87.
Mei, C.L., He, S.Y., Fang, K.T., 2004. A note on the mixed geographically weighted regression model. J. Reg. Sci. 44 (1), 143–157.
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.
Murakami, D., Lu, B., Harris, P., Brunsdon, C., Charlton, M., Nakaya, T., Griffith, D.A., 2019. The importance of scale in spatially varying coefficient modeling. Ann. Am. Assoc. Geogr. 109 (1), 50–70.
Öcal, N., Yildirim, J., 2010. Regional effects of terrorism on economic growth in Turkey: a geographically weighted regression approach. J. Peace Res. 47 (4), 477–489.
Oikarinen, E., Bourassa, S.C., Hoesli, M., Engblom, J., 2018. U.S. metropolitan house price dynamics. J. Urban Econ. 105, 54–69.
Salvati, L., Ciommi, M.T., Serra, P., Chelli, F.M., 2019. Exploring the spatial structure of housing prices under economic expansion and stagnation: the role of socio-demographic factors in metropolitan Rome, Italy. Land Use Pol. 81, 143–152.
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: Econ. Space 40 (4), 986–1005.
Wu, C., Ren, F., Hu, W., Du, Q., 2019. Multiscale geographically and temporally weighted regression: exploring the spatiotemporal determinants of housing prices. Int. J. Geogr. Inf. Sci. 33 (3), 489–511.