A bootstrap test for constant coefficients in geographically weighted regression models

Chang-Lin Mei, Min Xu and Ning Wang

Department of Statistics, School of Mathematics and Statistics, Xi’an Jiaotong University, Xi’an, PR China

ABSTRACT
Statistical tests for whether some coefficients really vary over space play an important role in using the geographically weighted regression (GWR) to explore spatial non-stationarity of the regression relationship. In view of some shortcomings of the existing inferential methods, we propose a residual-based bootstrap test to detect the constant coefficients in a GWR model. The proposed test is free of the assumption that the model error term is normally distributed and admits some useful extensions for identifying more complicated spatial patterns of the coefficients. Some simulation with comparison to the existing test methods is conducted to assess the test performance, including the accuracy of the bootstrap approximation to the null distribution of the test statistic, the power in identifying spatially varying coefficients and the robustness to collinearity among the explanatory variables. The simulation results demonstrate that the bootstrap test works quite well. Furthermore, a real-world data set is analyzed to illustrate the application of the proposed test.

1. Introduction

Due to its simple implementation and ease of interpretation, the geographically weighted regression (GWR) technique, originally proposed by Brunsdon et al. (1996), has become a popular tool for exploring spatial non-stationarity of a regression relationship in geo-referenced data analysis and has been applied to a variety of fields including geography, ecology, econometrics, epidemiology, and environmental science. In this methodology, a spatially varying coefficient model

\[ Y = \sum_{j=1}^{q} \beta_j(u,v)X_j + \varepsilon \quad (1) \]

is locally calibrated by the weighted least-squares procedure based on the observations \( \{y_i; x_{i1}, \ldots, x_{iq}\}_{i=1}^{n} \) of the response \( Y \) and the explanatory variables \( X_1, X_2, \ldots, X_q \) collected at the geographical locations \( (u_i, v_i) \) \( (i = 1, 2, \ldots, n) \). In the model, \( \beta_j(u,v) \) \( (j = 1, 2, \ldots, q) \) are unknown functions of the location co-ordinate \( (u,v) \), and \( \varepsilon \) is the error term with mean

CONTACT Chang-Lin Mei clmei@mail.xjtu.edu.cn

© 2016 Taylor & Francis
zero and variance $\sigma^2$. The model can include a spatially varying intercept by setting $X_1=1$.

Model (1) allows the regression coefficients to vary over space and therefore can be used to explore spatial non-stationarity of the regression relationship via the spatial variation patterns of the estimated coefficients. It is a common practice in the GWR literature that the estimated coefficients are mapped on the geographic region to show the degree of spatial non-stationarity in the regression relationship. In this sense, the coefficient estimators play an important role in reflecting spatial non-stationarity of the regression relationship and the analysis results depend largely upon the estimators of the underlying coefficients.

As a local fitting technique, however, the GWR estimator of each coefficient varies with the focal location no matter whether the underlying coefficient is constant or varies over the space. The excessive flexibility of this model calibration method may lead to spurious variation and even sign reversals of the estimated coefficients (Jetz et al. 2005, Páez et al. 2011). Moreover, collinearity among the explanatory variables may also result in sign changes and strong correlation in the coefficient estimators (Wheeler and Tiefelsdorf 2005, Wheeler and Calder 2007). On the other hand, if the coefficients are of very different degrees of smoothness with, for example, some coefficients being constant and the others varying over space, a global bandwidth selected by some commonly used data-driven methods, such as CV, GCV, and AIC, cannot yield accurate estimators for all of the coefficients, which has been theoretically proved and empirically confirmed in the statistical literature of varying coefficient models (Fan and Zhang 1999) and has also been noted in the GWR literature (Farber and Páez 2007, Yang et al. 2011, 2012).

In summary, there are several causes of inaccurate coefficient estimators and consequently misleading interpretation on the spatial variation patterns of the regression relationship. To alleviate the adverse effect of the collinearity and obtain more accurate estimators of the coefficient, Wheeler (2007) proposed the ridge regression and Bárcena et al. (2014) used the regularization methods for calibrating the GWR models. Furthermore, Wheeler (2009) presented the geographically weighted lasso for coefficient penalization and model selection. Recently, Brunsdon et al. (2012) and Gollini et al. (2015) developed a new ridge estimation approach, called locally compensated ridge GWR, to locally fit the ridge regression at the locations where the collinearity is a problem. Moreover, Wang et al. (2008) suggested the local-linear GWR to reducing the bias and boundary effect of the coefficient estimators. Yang et al. (2011, 2012) proposed the flexible bandwidth GWR method to estimate the coefficients with different degrees of smoothness. In addition to the above improved estimation methods, statistical tests for judging whether some of the coefficients vary over the space are fundamental in achieving a valid interpretation of spatial non-stationarity of the regression relationship. In particular, if the test indicates that some coefficients are constant over space, then we can, on one hand, conclude that the impact of the corresponding explanatory variables on the response is spatially stationary and, on the other hand, identify a more appropriate model, i.e. the mixed GWR model (Brunsdon et al. 1999) to fit the given data.

In fact, since the inception of GWR, much attention has been paid to such statistical inferences. In the pioneering work by Brunsdon et al. (1996, 1998), two kinds of permutation test were proposed respectively based on the selected bandwidth size and the sample variances of the estimated coefficients at all of the observation locations. However, these
two tests can only detect global stationarity of the regression relationship because it is under the null hypothesis that all of the coefficients are constant such that any permutations of the observations \( \{y_i; x_{i1}, \ldots, x_{ip}\}_{i=1}^n \) among the sampling locations are equally likely to occur and then the \( p \)-values of the tests could be derived. Shortly afterwards, Brunsdon et al. (1999) developed a set of analytically derived tests for the null hypothesis of no spatial stationarity by comparing the residual sum of squares from the GWR estimation of Model (1) (called GWR model henceforth) with that from the ordinary least-squares estimation of the corresponding global linear model. Furthermore, they proposed so-called mixed GWR models with the back-fitting calibration procedure on which the aforementioned tests can be used to identify one or more constant coefficients in the full GWR model in Equation (1). Leung et al. (2000) also introduced residual sum of squares based statistics to test for global stationarity of the regression relationship and another statistic that is based on the sample variance of each coefficient estimator at the sampling locations to test whether each coefficient varies spatially. In both papers, the null distributions of the test statistics were approximated by an \( F \) distribution with proper degrees of freedom. This kind of approximation roots in the distributional theory of the quadratic forms in normal variables and is closely related to the assumption that the model error term follows a normal distribution. Furthermore, when formulating the numerator and (or) denominator of each statistic as the quadratic forms of the model errors, the assumption that the fitted values of the response variable are unbiased estimators of the expectations of the corresponding observations has to be made. This assumption is in general untrue because it has been proved in the statistical nonparametric literature that the local-constant kernel estimation like GWR always produces biased estimators of the regression function except for the case of ordinary linear models (e.g. Fan and Zhang 1999, 2008). Thirdly, when using an \( F \) distribution to approximate the null distributions of the test statistics, the dependence between the numerator and the denominator of each statistic is neglected although they are generally co-dependent in the framework of nonparametric estimation. Mei et al. (2004) employed the sample variance based statistic to identify a mixed GWR model and suggested a more accurate approach, called three-moment \( \chi^2 \) approximation, to derive the \( p \)-value of the test. Although the requirement of independence between the numerator and denominator of the test statistic is avoided in this approximation, the assumptions of the normality of the model error term and the unbiasedness of the coefficient estimators are still retained. Páez et al. (2002a, 2002b) developed a maximum likelihood estimation and inferential framework for the GWR model. However, the assumption that the model error term follows a heteroscedastic normal distribution is still needed.

The above discussions suggest that there is still much room for improvement in the existing inferential methods of the GWR technique. One of the possible ways is to derive the asymptotic null distributions of the test statistics. However, it has been recognized that the limiting distributions may not necessarily provide a valid approximation to the null distributions in the case of finite sample size (e.g. Fan and Jiang 2007). Driven by many sophisticated statistical inferential problems in a variety of areas and fueled by the power of modern computers, Monte Carlo methods are a popular method to simulate the distribution of a complex statistic. The most attractive one is perhaps the bootstrap method originally developed by Efron (1979). Because of its solid theoretical basis and fewer restrictions on the underlying distribution of the error term, the bootstrap method has
been widely applied to the literature of parametric and nonparametric regression analysis for statistical inferences. This method has also been used in the GWR literature to test for zero coefficients in a mixed GWR model (Mei et al. 2006) and to investigate coefficient non-stationarity via comparing the basic GWR with some standard spatially autocorrelated regressions (Harris et al. 2015). In this paper, we focus on the residual sum of squares based statistic in Brunsdon et al. (1999) and propose a bootstrap procedure to approximate its null distribution to achieve the task of testing for the constant coefficients in GWR models. In view of the model calibration methods, although the local-linear GWR in Wang et al. (2008) performs better than the conventional GWR in reducing the bias and the boundary effect of the coefficient estimators, it is very computationally expensive because both the coefficients and their partial derivatives should be simultaneously estimated. Here, we still use the conventional GWR technique proposed by Brunsdon et al. (1996) to calibrate Model (1) and the two-step procedure derived by Fotheringham et al. (2002, Chapter 3) to fit a mixed GWR model because of their computational efficiency.

The rest of this paper is organized as follows. In Section 2, the estimation procedures for a GWR model and a mixed GWR model are briefly described. In Section 3, the residual sum of squares based statistic is introduced to test the null hypothesis that some of the coefficients in the GWR model are constant and a residual-based bootstrap procedure is proposed to compute the p-value of the test. Furthermore, some issues including the computational strategy and some possible extensions of the test are discussed. A simulation study with comparison to the existing test methods is conducted in Section 4 to evaluate the performance of the test under different error distributions and different degrees of collinearity among the explanatory variables. The Boston housing price data are analyzed in Section 5 to demonstrate the application of the proposed test. The paper is then concluded with some remarks.

2. A brief review of the conventional estimation approaches of GWR and mixed GWR models

2.1. Estimation of a GWR model and the resulting residual sum of squares

Let \( \{y_i; x_{i1}, \ldots, x_{iq}\}_{i=1}^n \) be the observations independently collected at the locations \((u_i, v_i) (i = 1, 2, \ldots, n)\) and \(d_{0i}\) be the Euclidean distance between a given focal location \((u_0, v_0)\) and \((u_i, v_i)\). The GWR estimation of Model (1) is to minimize the objective function

\[
\sum_{i=1}^n \left[ y_i - \sum_{j=1}^q \beta_j(u_0, v_0)x_{ij} \right]^2 K_h(d_{0i})
\]

with respect to \(\beta_j(u_0, v_0) (j = 1, 2, \ldots, q)\), where \(K_h() = K()h\) with \(K()\) being a kernel function and \(h\) being a bandwidth. Denote

\[
X = \begin{pmatrix}
x_{11} & x_{12} & \ldots & x_{1q} \\
x_{21} & x_{22} & \ldots & x_{2q} \\
\vdots & \vdots & \ddots & \vdots \\
x_{n1} & x_{n2} & \ldots & x_{nq}
\end{pmatrix},
\]

\[
Y = \begin{pmatrix}
y_1 \\
y_2 \\
\vdots \\
y_n
\end{pmatrix}.
\]
\( W_h(u_0, v_0) = \text{Diag}(K_h(d_{01}), K_h(d_{02}), \ldots, K_h(d_{0n})) \),

and

\[
\beta(u_0, v_0) = (\beta_1(u_0, v_0), \beta_2(u_0, v_0), \ldots, \beta_q(u_0, v_0))^T.
\]

Then, the GWR estimator of the coefficient vector \( \beta(u_0, v_0) \) is

\[
\hat{\beta}(u_0, v_0) = (\hat{\beta}_1(u_0, v_0), \hat{\beta}_2(u_0, v_0), \ldots, \hat{\beta}_q(u_0, v_0))^T
= [X^T W_h(u_0, v_0) X]^{-1} X^T W_h(u_0, v_0) Y.
\]

Especially, taking the focal location \((u_0, v_0) = (u_i, v_i)\) \(i = 1, 2, \ldots, n\), respectively, the fitted values of the response variable \( Y \) at the \( n \) sampling locations are computed by

\[
y_i = x_i^T \hat{\beta}(u_i, v_i) = x_i^T [X^T W_h(u_i, v_i) X]^{-1} X^T W_h(u_i, v_i) Y, \quad i = 1, 2, \ldots, n,
\]

where \( x_i = (x_{i1}, x_{i2}, \ldots, x_{in}) \) is the \( i \)th row of the matrix \( X \) in Equation (2). The fitted vector of the response variable \( Y \) at the sampling locations can be denoted by

\[
\hat{Y} = (\hat{y}_1, \hat{y}_2, \ldots, \hat{y}_n)^T = S(h) Y,
\]

where

\[
S(h) = \begin{pmatrix}
    x_1^T [X^T W_h(u_1, v_1) X]^{-1} X^T W_h(u_1, v_1) \\
    x_2^T [X^T W_h(u_2, v_2) X]^{-1} X^T W_h(u_2, v_2) \\
    \vdots \\
    x_n^T [X^T W_h(u_n, v_n) X]^{-1} X^T W_h(u_n, v_n)
\end{pmatrix}.
\]

The residual vector is then

\( \hat{\varepsilon}(h) = (\hat{\varepsilon}_1, \hat{\varepsilon}_2, \ldots, \hat{\varepsilon}_n)^T = Y - \hat{Y} = [I - S(h)] Y \)

and the residual sum of squares is

\( \text{RSS}(h) = (\hat{\varepsilon}(h))^T \hat{\varepsilon}(h) = Y^T [I - S(h)]^T [I - S(h)] Y, \)

where \( I \) is the identity matrix of order \( n \).

The size of the bandwidth \( h \) can be determined by the data-driven methods, such as CV, GCV, and AIC\(_c\). Here, the AIC\(_c\) procedure is used to select the bandwidth. That is, let

\[
\text{AIC}_c(h) = \log \left[ \frac{1}{n} \text{RSS}(h) \right] + \frac{n + \text{tr}(S(h))}{n - 2 - \text{tr}(S(h))},
\]

where \( \text{tr}(\cdot) \) stands for the trace of a matrix. The optimal size of \( h \) is

\[
h_1 = \arg \min_{h > 0} \text{AIC}_c(h).
\]

2.2. Estimation of a mixed GWR model and the resulting residual sum of squares

After properly re-ordering the explanatory variables, a mixed GWR model in its sample form (Brunsdon et al. 1999) stands as
\[
y_i = \sum_{j=1}^{r} \beta_j(u_j, v_j)x_{ij} + \sum_{j=r+1}^{q} \beta_jx_{ij} + \epsilon_i, \quad i = 1, 2, \ldots, n,
\]

where \( \beta_j(u,v) \) \((j = 1, 2, \ldots, r)\) are spatially varying coefficients and \( \beta_j \) \((j = r + 1, r + 2, \ldots, q)\) are constant coefficients. Here, the errors \( \epsilon_i \) \((i = 1, 2, \ldots, n)\) are also assumed to be independent and identically distributed random variables with mean zero and variance \( \sigma^2 \).

Fotheringham et al. (2002, chapter 3) proposed a two-step calibration method for the mixed GWR model. This method is less computationally intensive than the back-fitting approach suggested in Brunsdon et al. (1999) and admits the explicit expressions of the estimators for both the constant and the spatially varying coefficients. What follows is the description of the two-step method with the resulting residual sum of squares.

For the mixed GWR model in Equation (9), let

\[
X_c = \begin{pmatrix}
X_{11} & X_{12} & \cdots & X_{1r} \\
X_{21} & X_{22} & \cdots & X_{2r} \\
\vdots & \vdots & \ddots & \vdots \\
X_{n1} & X_{n2} & \cdots & X_{nr}
\end{pmatrix}
\]

be the design matrices corresponding to the spatially varying coefficient part and the constant coefficient part, respectively, and let

\[
\beta_c = \begin{pmatrix}
\beta_1(u,v) \\
\beta_2(u,v) \\
\vdots \\
\beta_q(u,v)
\end{pmatrix}
\]

be the vectors of the spatially varying coefficients and the constant coefficients, respectively. According to the two-step estimation, the estimator of \( \hat{\beta}_c \) is

\[
\hat{\beta}_c = \left( \hat{\beta}_{r+1}, \hat{\beta}_{r+2}, \ldots, \hat{\beta}_q \right)^T
\]

\[
= \left( X_c^T (I - L(h))^T (I - L(h)) X_c \right)^{-1} X_c^T (I - L(h))^T (I - L(h)) Y
\]

and the estimator of \( \hat{\beta}_v(u,v) \) at each sampling location \((u_i, v_i)\) is

\[
\hat{\beta}_v(u_i, v_i) = \left( \hat{\beta}_1(u_i, v_i), \hat{\beta}_2(u_i, v_i), \ldots, \hat{\beta}_q(u_i, v_i) \right)^T
\]

\[
= \left( X_v^T W_h(u_i, v_i) X_v \right)^{-1} X_v^T W_h(u_i, v_i) (Y - X_c \hat{\beta}_c)
\]

where

\[
L(h) = \begin{pmatrix}
x_{u1}^T [X_v^T W_h(u_1, v_1) X_v]^{-1} X_v^T W_h(u_1, v_1) \\
x_{u2}^T [X_v^T W_h(u_2, v_2) X_v]^{-1} X_v^T W_h(u_2, v_2) \\
\vdots \\
x_{wn}^T [X_v^T W_h(u_n, v_n) X_v]^{-1} X_v^T W_h(u_n, v_n)
\end{pmatrix},
\]

\( x_{ui}^T = (x_{i1}, x_{i2}, \ldots, x_{ir}) \) is the \( i \)th row of \( X_v \), \( W_h(u_v, v) \) is the same diagonal weight matrix in Equation (3) with \((u_0, v_0)\) replaced by \((u,v)\) \((i = 1, 2, \ldots, n)\) and \( Y \) is the observation vector.
in Equation (2). With the estimators $\hat{\beta}_c$ and $\hat{\beta}_v(u_i, v_i)$ ($i = 1, 2, \ldots, n$), the fitted vector of the response variable $Y$ in the mixed GWR model is then

$$\hat{Y}_M = (\hat{y}_{1M}, \hat{y}_{2M}, \ldots, \hat{y}_{nM})^T = S_M(h)Y,$$  \hspace{1cm} (11)

where

$$S_M(h) = L(h) + (I - L(h))X_c \left[ X_c^T (I - L(h)) (I - L(h))X_c \right]^{-1} \cdot X_c^T (I - L(h)) (I - L(h)).$$  \hspace{1cm} (12)

The residual vector is therefore

$$\hat{\varepsilon}_M(h) = (\hat{\varepsilon}_{1M}, \hat{\varepsilon}_{2M}, \ldots, \hat{\varepsilon}_{nM})^T = Y - \hat{Y}_M = [I - S_M(h)]Y,$$  \hspace{1cm} (13)

and the residual sum of squares is

$$RSS_M(h) = (\hat{\varepsilon}_M(h))^T \hat{\varepsilon}_M(h) = Y^T [I - S_M(h)]^T [I - S_M(h)] Y.$$  \hspace{1cm} (14)

In the above estimation of the mixed GWR model, the optimal size of the bandwidth $h$ can also be selected by the AICc criterion. That is, let

$$AIC_{cM}(h) = \log \left[ \frac{1}{n} RSS_M(h) \right] + \frac{n + \text{tr}(S_M(h))}{n - 2 - \text{tr}(S_M(h))}.$$  \hspace{1cm} (16)

The optimal size of $h$ is

$$h_M = \arg \min_{h > 0} AIC_{cM}(h).$$  \hspace{1cm} (15)

**3. A residual-based bootstrap test for constant coefficients in the GWR model**

In this section, we first focus on testing the null hypothesis

$H_0$: some coefficients in Model (1) are constant

versus the alternative hypothesis

$H_1$: all of the coefficients in Model (1) vary over the space

and formulate a residual-based bootstrap test for the hypotheses. Then, some related issues including the computational strategy of the $p$-value and some possible extensions of the test are discussed.

**3.1. Construction of the test statistic**

Based on the data set $\{y_i; x_{i1}, x_{i2}, \ldots, x_{iq}; (u_i, v_i)\}_{i=1}^n$, fit the GWR model in Equation (1) (the alternative model) according to the estimation procedure described in Subsection 2.1 and compute the residual sum of squares
\[
RSS(h_1) = Y^T [I - S(h_1)]^T [I - S(h_1)] Y = Y^T R(h_1) Y, \tag{16}
\]

where \( R(h_1) = [I - S(h_1)]^T [I - S(h_1)] \) and \( h_1 \) is the optimal bandwidth size selected by the AICc criterion in Equation (8). Specifically, it is assumed that one wants to test whether the last \( q - r \) coefficients in Model (1) are constant. Then the null hypothesis is

\[
H_0 : \beta_j(u, v) = \beta_j \text{ for } j = r + 1, r + 2, \ldots, q, \tag{17}
\]

which leads to the null model of the mixed GWR model in Equation (9). Calibrate the null model with the bandwidth size \( h_1 \) selected under the alternative model and compute the residual sum of squares

\[
RSS_M(h_1) = Y^T [I - S_M(h_1)]^T [I - S_M(h_1)] Y = Y^T R(H_0) Y, \tag{18}
\]

where \( R(H_0) = [I - S_M(h_1)]^T [I - S_M(h_1)] \). The test statistic is then constructed as

\[
T = \frac{RSS_M(h_1) - RSS(h_1)}{RSS(h_1)} = \frac{Y^T [R(H_0) - R(H_1)] Y}{Y^T R(H_1) Y}. \tag{19}
\]

The above statistic has been used in the GWR literature (Brunsdon et al. 1999, Leung et al. 2000, Mei et al. 2006) for comparing the goodness-of-fit of two related models. In the nonparametric regression literature, this statistic is called a generalized likelihood ratio statistic originally proposed by Fan et al. (2001) and has been widely used to perform inference for various nonparametric or semiparametric models (Fan and Jiang 2007). It has been proved in many situations that this statistic holds a useful property, called Wilks’ phenomenon, meaning that its limiting distribution is free of the nonparametric smoothing methods used for calibrating the model. As pointed out by Fan and Huang (2005), when the generalized likelihood ratio statistic is used for hypothesis testing, it is essential to calibrate both the null and the alternative models with the same bandwidth size unless the null model is a full parametric model which does not need a bandwidth in its estimation. Otherwise, the adjustment of the bandwidth may make the residual sums of squares from the null and alternative models incomparable, leading to a loss in the power of the test. Furthermore, as noted in Fan and Jiang (2007) and Cai (2007), for a given data set, since one is not certain whether the null model is correct, the fit from the alternative model should be used to generate the residuals for bootstrap sampling because this fit results in consistent residuals under both the null and alternative models. For these reasons, we use in this paper the bandwidth size \( h_1 \) selected in the GWR model (alternative model) to fit the mixed GWR model (null model) in the construction of the test statistic in Equation (19), which meets the above two requirements simultaneously.

Intuitively, if the null hypothesis \( H_0 \) in Equation (17) is indeed true, the difference between \( RSS_M(h_1) \) and \( RSS(h_1) \) should be small. Otherwise, \( RSS_M(h_1) \) tends to be significantly larger than \( RSS(h_1) \). Hence, a large value of the statistic \( T \) supports rejection of the null hypothesis and the \( p \)-value of the test is then

\[
p = P_{H_0}(T \geq t), \tag{20}
\]
where \( t \) is the observed value of \( T \) obtained by Equation (19) and \( P_{H_0} \) means that the probability is computed according to the null distribution of \( T \). Given a significance level \( a \), if \( p < a \), then reject the null hypothesis \( H_0 \); otherwise, do not reject \( H_0 \).

3.2. A bootstrap procedure for computing the p-value of the test

To overcome the shortcomings of the existing inferential methods in approximating the null distribution of the test statistic \( T \), we suggest a bootstrap procedure to compute the p-value of the test. Here, it is only assumed that the model errors \( \varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n \) are independent and identically distributed random variables with mean zero and variance \( \sigma^2 \). The bootstrap procedure in our setting is as follows.

**Step 1**: Based on the original data \( \{y_i; x_{i1}, x_{i2}, \ldots, x_{iq}; (u_i, v_i)\}_{i=1}^n \), the alternative model (i.e., the full GWR model in Equation (1)) is calibrated and the optimal bandwidth size \( h_1 \) is selected. Then, compute the residual vector \( \hat{\varepsilon}(h_1) = (\hat{\varepsilon}_1, \hat{\varepsilon}_2, \ldots, \hat{\varepsilon}_n)^T = [I - S(h_1)]Y \) in Equation (6) and the residual sum of squares \( RSS(h_1) = \hat{Y}^T R(H_1) \hat{Y} \) in Equation (16).

**Step 2**: Fit the null model (i.e., the mixed GWR model in Equation (9)) using the two-step calibration method and the same bandwidth size \( h_1 \) in Step 1. Then, compute the fitted vector of the response variable \( \hat{\varepsilon}(h_1) = (\hat{\varepsilon}_1, \hat{\varepsilon}_2, \ldots, \hat{\varepsilon}_n)^T = [I - S(h_1)]Y \) in Equation (6) and the residual sum of squares \( RSS(h_1) = \hat{Y}^T R(H_1) \hat{Y} \) in Equation (18).

**Step 3**: Compute the observed value of the statistic \( T \) by \( t = \frac{RSS_{H_1}(h_1) - RSS(h_1)}{RSS(h_1)} \) and the centered residual vector \( \hat{\varepsilon}_c = (\hat{\varepsilon}_1c, \hat{\varepsilon}_2c, \ldots, \hat{\varepsilon}_{nc})^T \) with \( \hat{\varepsilon}_ic = \hat{\varepsilon}_i - \frac{1}{n} \sum_{j=1}^n \hat{\varepsilon}_j (i = 1, 2, \ldots, n) \). where \( \hat{\varepsilon}_1, \hat{\varepsilon}_2, \ldots, \hat{\varepsilon}_n \) are the residuals obtained in Step 1.

**Step 4**: Draw a bootstrap sample \( \varepsilon^* = (\varepsilon_1^*, \varepsilon_2^*, \ldots, \varepsilon_n^*)^T \) with replacement from \( \hat{\varepsilon}_c = (\hat{\varepsilon}_1c, \hat{\varepsilon}_2c, \ldots, \hat{\varepsilon}_{nc})^T \) and compute the bootstrap sample of the response variable by \( Y^* = (y_1^*, y_2^*, \ldots, y_n^*)^T = S(h_1)Y + \varepsilon^* \), yielding the bootstrap data \( \{y_i; x_{i1}, x_{i2}, \ldots, x_{iq}; (u_i, v_i)\}_{i=1}^n \).

**Step 5**: Based on the bootstrap data, re-calibrate both the alternative and null models by their respective estimation approaches with the same bandwidth size \( h_1 \). Compute the residual sums of the squares \( RSS_{H_1}^*(h_1) \) and \( RSS_{H_0}^*(h_1) \) of the alternative and the null models, respectively. Then a bootstrap value of the statistic \( T \) is obtained by \( t^* = \frac{RSS_{H_1}^*(h_1) - RSS_{H_0}^*(h_1)}{RSS(h_1)} \).

**Step 6**: Repeat Steps 4 and 5 for \( b \) times and obtain \( b \) bootstrap values of the statistic \( T \), which we denote by \( t_1^*, t_2^*, \ldots, t_b^* \). The p-value in Equation (20) is then estimated by

\[
\hat{p} = \frac{1}{b} \sum_{i=1}^b I(t_i^* \geq t),
\]

where \( I(\cdot) \) is the indicator function.

3.3. Some comments on the test

(1) Testing a globally stationary regression relationship. As one of the most important inferences in the GWR literature, the test for a globally stationary regression relationship
can provide the information that a spatially varying coefficient model is really necessary for the given data. The proposed bootstrap test is applicable to this case. In fact, the null hypothesis in this case is

\[ H_0: \text{all of the coefficients in Model (1) are constant,} \]

which corresponds to the ordinary linear regression model

\[ y_i = \sum_{j=1}^{q} \beta_j x_{ij} + \epsilon_i, \quad i = 1, 2, \ldots, n. \]

In this case, the above null model is simply calibrated by the least-squares procedure. As a result, the matrix \( S_M(h_1) \) and the residual sum of squares \( RSS_M(h_1) \) in Step 2 do not depend on the bandwidth size \( h_1 \) and have the forms of

\[ S_M(h_1) = X(X^TX)^{-1}X^T \quad \text{and} \quad RSS_M(h_1) = Y^T[I - X(X^TX)^{-1}X^T]Y, \]

respectively, where \( X \) is the design matrix in Equation (2).

(2) A simplified algorithm to compute the bootstrap value of the test statistic. Step 5 says that both the alternative and the null models must be recalibrated based on the bootstrap data \( \{ y_i^*: x_{i1}, x_{i2}, \ldots, x_{iq}; (u_i, v_i) \}_{i=1}^{n} \}. \) However, the calibrations of the alternative and null models do not need to be performed repeatedly. In fact, it is noted that both the original data and the bootstrap data have the same observations of the explanatory valuables and the corresponding sampling locations. Furthermore, the bandwidth size used for re-calibrating the two models is unchanged in each repeat of Steps 4 and 5. Therefore, re-calibrating the alternative and the null models will result in the same matrices \( S_M(h_1) \) and \( S_M^*(h_1) \) and therefore the same matrices \( R(H_0) \) and \( R(H_1) \). Based on this fact, the residual sums of squares \( RSS^*(h_1) \) and \( RSS_M^*(h_1) \) can be simply computed by

\[ RSS^*(h_1) = (Y^*)^T[I - S_M(h_1)]^T[I - S_M(h_1)]Y^* = (Y^*)^T R(H_1) Y^* \]

and

\[ RSS_M^*(h_1) = (Y^*)^T[I - S_M^*(h_1)]^T[I - S_M^*(h_1)]Y^* = (Y^*)^T R(H_0) Y^* \]

respectively, where \( Y^* = S_M(h_1) Y + \epsilon^* \). Furthermore, the bootstrap value of the test statistic \( T \) is

\[ t^* = \frac{(Y^*)^T[R(H_0) - R(H_1)]Y^*}{(Y^*)^T R(H_1) Y^*}. \]

Therefore, in each of the \( b \) repeats of computing the bootstrap value in Step 6, it is only needed to replace \( Y^* \) with the updated one, which can largely reduce the computational complexity of the bootstrap test.

(3) Some possible extensions of the test. Although the test is proposed mainly for identifying the constant coefficients in a GWR model, it admits some extensions for testing more complex structures of the coefficients.

One extension is to test the null hypothesis that some coefficients, for example, the last \( q - r \) coefficients in Model (1), are linear combinations of some known functions of the spatial coordinate \((u, v)\). Specifically, the null hypothesis can be expressed as
\[
H_0 : \beta_j(u,v) = \sum_{k=1}^{l_j} \theta_{jk} g_{jk}(u,v) \quad \text{for } j = r + 1, r + 2, \ldots, q,
\]

where for each \( j = r + 1, r + 2, \ldots, q \), \( \theta_{jk} (k = 1, 2, \ldots, l_j) \) are unknown constant parameters and \( g_{jk}(u,v) \) \((k = 1, 2, \ldots, l_j)\) are known linearly independent functions of the spatial coordinate \((u,v)\). Under the null hypothesis, the null model is of the form

\[
y_i = \sum_{j=1}^{r} \beta_j(u_i, v_i) x_{ij} + \sum_{j=r+1}^{q} \left[ \sum_{k=1}^{l_j} \theta_{jk} g_{jk}(u_i, v_i) x_{ij} \right], \quad i = 1, 2, \ldots, n.
\]

Let

\[
z_{ijk} = g_{jk}(u_i, v_i)x_{ij}, \quad k = 1, 2, \ldots, l_j; \quad j = r + 1, r + 2, \ldots, q; \quad i = 1, 2, \ldots, n.
\]

The null model becomes

\[
y_i = \sum_{j=1}^{r} \beta_j(u_i, v_i) x_{ij} + \sum_{j=r+1}^{l_j} \sum_{k=1}^{l_{r+1,j}} \theta_{r+1,k} z_{i,r+1,k} + \sum_{j=r+2}^{l_j} \sum_{k=1}^{l_{r+2,k}} \theta_{r+2,k} z_{i,r+2,k}
\]

\[
\ldots + \sum_{j=1}^{l_q} \sum_{k=1}^{l_{q,k}} \theta_{q,k} z_{iqk} + \epsilon_i, \quad i = 1, 2, \ldots, n,
\]

which is a mixed GWR model because \( z_{ijk} \) \((k = 1, 2, \ldots, l_j; j = 1, 2, \ldots, q; i = 1, 2, \ldots, n)\) are all known. Therefore, the proposed bootstrap test is applicable to the null hypothesis in Equation (21).

When each \( g_{jk}(u,v) \) in Equation (21) is a power function of \( u \) and \( v \), the coefficients \( \beta_j(u,v) \) \((j = r + 1, r + 2, \ldots, q)\) are polynomial functions of the spatial coordinate. Especially, when the null hypothesis includes all the coefficients (i.e. \( r = 0 \)), the resulting null model is in fact a geographical expansion model which has been comprehensively studied in Cassetti (1972, 1997) and Jones and Cassetti (1992). Therefore, as a special case of the null hypothesis in Equation (21), the bootstrap test can be used to identify whether an expansion model is suitable for a given data set.

Another extension is to test the hypothesis of one mixed GWR model against another mixed GWR model. In this case, the alternative model is a mixed GWR model. The null model may be diverse. For example, in order to evaluate whether some explanatory variables in the constant-coefficient part of the alternative model do influence the response variable, one can set such a null model that the coefficients corresponding to these explanatory variables are all zero, resulting in the same test method in Mei et al. (2006).

4. A simulation study

In this section, some simulation with comparison to the existing test methods in the GWR literature is conducted to assess the performance of the proposed bootstrap test. The simulation focuses on the validity of the bootstrap approximation to the null distribution of the test statistic and the power of the test in detecting the spatially varying coefficients under normal and some typical non-normal error distributions. Furthermore, in view of the fact that collinearity among the explanatory variables may have an adverse impact on the coefficient estimators (Wheeler and Tiefelsdorf 2005,
Wheeler and Calder (2007), it is therefore of interest to evaluate the influence of the collinearity on the performance of the test.

4.1. Design of the simulation

(1) The spatial layout. The spatial region for the simulation is taken to be a square with the length of each side being 10 units and a Cartesian coordinate system is installed in such a way that its origin is located at the bottom-left corner of the square and the two axes coincide with the mutually orthogonal two sides of the square. The sampling locations are $m \times m$ lattice points with their spatial coordinates under the Cartesian system being

$$(u_i, v_i) = \left(\frac{10}{m-1}\mod(i-1,m), \frac{10}{m-1}\text{int}(i-1,m)\right), \quad i = 1, 2, \ldots, m^2,$$

where $\text{mod}(a,b)$ and $\text{int}(a,b)$ stand for the remainder and the integer part of $a$ divided by $b$, respectively. In the simulation, $m = 16$ and 21 are designated, resulting in the sample sizes $n = 256$ and 441, respectively.

(2) The model for generating data. Let $Y$ be the response variable, $X_1, X_2, X_3$, and $X_4$ be the explanatory variables, and $\{y_i; x_{i1}, x_{i2}, x_{i3}, x_{i4}; (u_i, v_i)\}_{i=1}^n$ be the data. The following GWR model is considered throughout the simulation:

$$y_i = \beta_1(u_i, v_i)x_{i1} + \beta_2(u_i, v_i)x_{i2} + \beta_3(u_i, v_i)x_{i3} + \beta_4(u_i, v_i)x_{i4} + \varepsilon_i, \quad i = 1, 2, \ldots, n,$$

where

$$\begin{align*}
\beta_1(u, v) &= 4 \sin\left(\frac{\pi}{10}u\right); \\
\beta_2(u, v) &= \frac{4}{625}uv(10-u)(10-v); \\
\beta_3(u, v) &= 1 + 4c \left[1 + e^{-\left(u-5\right)}\right]^{-1}; \\
\beta_4(u, v) &= 0.5 + 2c \left[1 + \frac{1}{125} \left(v-5\right)^3\right],
\end{align*}$$

with $c$ being a constant which will be designated different values to evaluate the power of the test. The model errors $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n$ are independent and identically distributed random variables with mean zero and variance $\sigma^2$.

(3) The error distribution. In order to examine the effect of the error distribution on the performance of the test, we consider the following three different types of the error distribution whose scales are so adjusted that they all have a common variance $\sigma^2 = 1$:

(i) standard normal distribution $N(0,1)$;
(ii) uniform distribution $U(-\sqrt{3}, \sqrt{3})$;
(iii) transformed and centered $\chi^2$ distribution $\frac{1}{2}\chi^2(2) - 1$.

(4) The way of generating the observations of the explanatory variables. Let $Z_1, Z_2, Z_3$, and $Z_4$ be independent random variables with the common distribution $U(0,1)$. The explanatory variables are taken to be
\[
\begin{pmatrix}
X_1 \\
X_2 \\
X_3 \\
X_4
\end{pmatrix} = \begin{pmatrix}
1 & \gamma & 0 & 0 \\
\gamma & 1 & \tau & 0 \\
0 & \tau & 1 & \delta \\
0 & 0 & \delta & 1
\end{pmatrix} \begin{pmatrix}
Z_1 \\
Z_2 \\
Z_3 \\
Z_4
\end{pmatrix},
\]

where \( 0 \leq \gamma, \tau, \rho < 1 \). The following four cases are considered:

(i) \( \gamma = \tau = \delta = 0 \). In this case, the explanatory variables are mutually independent.

(ii) \( \tau = \delta = 0, \) and \( \gamma = 0.27 \) and \( 0.5, \) respectively. In this case \((X_1, X_2, X_3, X_4)\) are mutually independent while \( X_1 \) and \( X_2 \) are dependent. The correlation coefficient between \( X_1 \) and \( X_2 \) is

\[
\rho_{X_1X_2} = \frac{\text{Cov}(X_1, X_2)}{\sqrt{\text{Var}(X_1) \cdot \text{Var}(X_2)}} = \frac{\text{Cov}(Z_1, \gamma Z_2, \gamma Z_1 + Z_2)}{\sqrt{\text{Var}(Z_1 + \gamma Z_2) \cdot \text{Var}(\gamma Z_1 + Z_2)}} = \frac{2\gamma}{1 + \gamma^2}.
\]

\( \gamma = 0.27 \) leads to \( \rho_{X_1X_2} = 0.5 \), a moderate correlation coefficient between \( X_1 \) and \( X_2 \); \( \gamma = 0.5 \) yields \( \rho_{X_1X_2} = 0.8 \), a high correlation coefficient between \( X_1 \) and \( X_2 \).

(iii) \( \gamma = \delta = 0, \) and \( \tau = 0.27 \) and \( 0.5, \) respectively. In this case \((X_1, X_2, X_3)\), and \( X_4 \) are mutually independent while the correlation coefficients between \( X_2 \) and \( X_3 \) are \( \rho_{X_2X_3} = 0.5 \) and \( 0.8, \) respectively.

(iv) \( \gamma = \delta = 0, \) and \( \delta = 0.27 \) and \( 0.5, \) respectively. Similarly, \( X_1, X_2, \) and \( (X_3, X_4) \) are mutually independent while the correlation coefficients between \( X_3 \) and \( X_4 \) are \( \rho_{X_3X_4} = 0.5 \) and \( 0.8, \) respectively.

Given each set of the values for \( \gamma, \tau, \) and \( \delta, \) the observations of the explanatory variables are generated as follows. Draw independently a sample \((z_1, z_2, z_3, z_4)^T\) from the distribution \( U(0,1) \) and the observation of the explanatory variables \((x_1, x_2, x_3, x_4)^T\) is computed according to Equation (23). Furthermore, the corresponding observation of the response variable is derived according to Model (22).

(5) The null hypothesis. The null hypothesis to be tested in the simulation is

\[
H_0 : \beta_3(u, v) = \beta_3 \text{ and } \beta_4(u, v) = \beta_4,
\]

where \( \beta_3 \) and \( \beta_4 \) are two constants. This null hypothesis corresponds to \( c = 0 \) in the coefficients \( \beta_3(u, v) \) and \( \beta_4(u, v) \).

(6) Selection of the kernel function and bandwidth. Throughout the simulation, the kernel function is chosen as the Gaussian kernel, i.e. \( K(t) = \frac{1}{\sqrt{2\pi}} e^{-t^2/2} \), and the optimal bandwidth size for the alternative model is selected by the \( \text{AIC}_c \) criterion in Equations (8).

### 4.2. Simulation results with analysis

For each of the experimental settings, we run \( N = 500 \) replications and, in each replication, \( b = 500 \) bootstrap samples were drawn to compute the \( p \)-value.

(1) Validity of the bootstrap approximation to the null distribution of the test statistic. Under the null hypothesis (i.e. setting \( c = 0 \) in the coefficients \( \beta_3(u, v) \) and \( \beta_4(u, v) \)) and for each of the experimental settings, we calculated the rate of rejecting the null hypothesis.
among $N = 500$ replications at the significance levels $\alpha = 0.01$, 0.05, and 0.10, respectively. The rate is an estimator of the type I error of the test under the given setting and significance level.

Moreover, the bootstrap test was compared with the test in Brunsdon et al. (1999) where the test statistic is the same as that in Equation (19) except for an additional scaling factor, but the null distribution of the statistic is approximated by an $F$ distribution. It should also be pointed out that, in the present simulation, the null model was calibrated by the two-step method described in Section 2.2 instead of the originally proposed back-fitting procedure in Brunsdon et al. (1999). Furthermore, although the degrees of freedom of the foregoing $F$ distribution are derived under the assumption of the normally distributed error terms, the formulas can be routinely used to compute the degrees of freedom for other error distributions.

For each of the experimental settings, the rejection rate was computed for both test methods and the results are reported in Table 1, where the letters 'B' and 'F' refer respectively to the bootstrap test and the test in Brunsdon et al. (1999) which we call $F$-test henceforth.

It is known from Table 1 that, for the bootstrap test, the rejection rates under the null hypothesis are reasonably close to the respective significance levels in all of the experimental settings even for the small sample size of $n = 256$, indicating that the bootstrap method yields an accurate approximation to the null distribution of the test statistic at least on the right tail of the null distribution on which the $p$-value of the test is computed. Furthermore, although the bootstrap approximation seemingly performs best for the normal error distribution, it can also account well for non-normal error distributions in the sense that the rejection rates do not show an evident deviation from the respective significance levels for both the short-tailed distribution of $U(-\sqrt{3}, \sqrt{3})$ and the skew distribution of $\frac{1}{2}\chi^2(2)-1$. More interestingly, although the collinearity among the explanatory variables may lead to spurious correlation between the GWR estimators of the spatially varying coefficients (Wheeler and Tiefelsdorf 2005), it does not bring about obvious influence on the rejection rates under the null hypothesis, which indicates that the bootstrap approximation to the null distribution of the test statistic is rather robust to the collinearity among the explanatory variables.

For the $F$-test, however, the rejection rates are much lower than their respective significance levels in all of the experimental settings, which indicates that the derived $F$ distribution may have a heavier tail on the right than the real null distribution of the test statistic, leading to an inaccurate approximation to the null distribution. Although the rejection rates under the larger significance levels tend to increase with the sample size increasing, they are still much smaller than the corresponding significance levels. One more finding is that, under the null hypothesis, the rejection rates of the $F$-test are little influenced not only by the collinearity between the explanatory variables but also by the non-normality of the error distribution.

(2) Power of the bootstrap test. When $c \neq 0$, the alternative hypothesis that all of the coefficients in Model (22) vary over the space is true. In this situation, given a significance level $\alpha$, the rate of rejecting the null hypothesis in $N = 500$ replications is an estimator of the test power. Here, we conduct some simulation to assess the power of the bootstrap test via a comparison with that of the $F$-test.
Set the significance level $\alpha = 0.05$ and the values of $c$ to be from 0.1 to 0.7 by an increment of 0.1. Here, we only focused on the cases where the explanatory variables $X_1$, $X_2$, $X_3$, and $X_4$ are mutually independent and two of them are highly correlated with $\rho_{(X_1,X_2)} = \rho_{(X_3,X_4)} = 0.8$, respectively. Furthermore, the power of the $F$-test was also computed in each experimental setting for the purpose of comparison. The power functions for both tests are depicted in Figure 1.

It can be observed from Figure 1 that the bootstrap test is of high power in identifying the varying coefficients in the GWR model no matter what the model error term follows a normal or a non-normal distribution. In contrast, the $F$-test shows evidently lower power in
all of the settings. Especially, when the difference between the null and alternative hypotheses is small, the F-test is much more conservative than the bootstrap test. For other aspects, both tests show similar characteristics. Specifically, the power increases rapidly with either the sample size increasing or the alternative hypothesis deviating from the null hypothesis gradually. Compared to the results in the case of mutually independent explanatory variables, the power of both tests is somewhat influenced by the collinearity among the explanatory variables in complicated ways. Generally, there is a loss of power especially in the case of $\rho(X_2; X_3) = 0.8$, where the collinearity exists between the explanatory variables in the null and alternative hypotheses, respectively; in the case of $\rho(X_3; X_4) = 0.8$, where the correlated explanatory variables are all in the alternative hypothesis, however, the power increases slightly for both tests.

To further understand the impact of the collinearity on the power of the tests, we also run the simulation for negative correlation between the explanatory variables. That is, the parameters $\gamma$, $\tau$, and $\delta$ take the corresponding negative values, respectively. The results for both tests are in general similar to those mentioned above, but for the case of $\rho(X_2; X_3) = -0.8$, the power of both tests shows an evident increase.
compared to the case of $\rho(X_2, X_3) = 0.8$. The results for the case of $\rho(X_2, X_3) = -0.8$ are illustrated in the last row of Figure 1.

In addition, the bootstrap test was compared with the test suggested in Leung et al. (2000) (called $L$-test henceforth) in the simulation study. In the $L$-test, the test statistic is constructed based on the sample variance of the GWR estimators of each coefficient at all of the sampling locations (see Equation (65) in Leung et al. 2000) and its null distribution is also approximated by an $F$ distribution. It should be noted that since each coefficient is separately tested for a constant in the $L$-test, testing the null hypothesis in Equation (24) with the $L$-test amounts to simultaneously testing the following two null hypotheses

$$H_{03}: \beta_3(u, v) = \beta_3$$ and $$H_{04}: \beta_4(u, v) = \beta_4.$$

In all of the experimental settings, the rejection rates under the null hypotheses $H_{03}$ and $H_{04}$ were also computed with the $L$-test. Because of the limited space, the results are omitted here, but those for $n = 441$ are attached as a supplementary material. It is found from the results that the $L$-test yields an excessively high type I error. That is, when the null hypotheses are indeed true, the rejection rates in all settings are much larger than the corresponding significance levels, indicating that an $F$ distribution cannot provide an accurate approximation to the null distributions of the test statistics. Although the results improve with the sample size increasing, the rejection rates are still generally two or three times higher than the corresponding significance levels even for the larger sample size of $n = 441$ (see Table A in the supplementary material). In terms of the test power, the $L$-test yields a more complex pattern than the bootstrap test (see Figure A in the supplementary material). Generally speaking, the power of $L$-test is higher than that of the bootstrap test especially for small value of $c$. With the value of $c$ increasing, however, the power of $L$-test for $H_{04}$ becomes lower than that of the bootstrap test. Moreover, the impact of the collinearity on the power seems different for these two tests with the most serious case being $\rho(X_2, X_3) = 0.8$ for the $L$-test, but $\rho(X_2, X_3) = 0.8$ for the bootstrap test.

In summary, the simulation study demonstrates that the bootstrap test outperforms the existing $F$-test and $L$-test in terms of type I error and test power. The bootstrap test can well approximate the null distribution of the test statistic even for small sample size and is free of the normality assumption on the error distribution. Furthermore, it is powerful in identifying the spatially varying coefficients. However, collinearity among the explanatory variables has an adverse impact on the power of the test, although the influence tends to be alleviated with the sample size increasing. Similar conclusions have been reached by Páez et al. (2011) in the empirical study of estimation of GWR models.

### 5. Application to Boston housing data

The proposed bootstrap test is illustrated by an application to the Boston housing data set given in Harrison and Rubinfeld (1978) and corrected for a few minor errors by Gilley and Pace (1996). The data set consists of the median value ($MEDV$ in $1000$) of owner-occupied homes in 506 US census tracts in the Boston area in 1970 as well as the following explanatory variables:
- **CRIM**, the per capita crime rate by town;
- **ZN**, the proportion of a town’s residential land zoned for lots greater than 25,000 square feet;
- **INDVS**, the proportion of non-retail business acres per town;
- **CHAS**, the Charles River dummy variable (**CHAS** = 1 if the tract bounds the Charles River; **CHAS** = 0 otherwise);
- **NOX**, the nitrogen oxide concentration in parts per 10 million;
- **RM**, the average number of rooms in owner-occupied homes per dwelling;
- **AGE**, the proportion of owner-occupied homes built prior to 1940;
- **DIS**, the weighted distances to five employment centers in the Boston region;
- **RAD**, the index of accessibility to radial highways;
- **TAX**, the full value property tax rate per $10000;
- **PTRATIO**, the pupil–teacher ratio by town school district;
- **LSTAT**, the proportion of population that is in the lower status;
- **BK**, the black proportion of population by town.

In addition, the Cartesian coordinate \((u,v)\) for each tract is attached to the data set.

The variables **ZN** and **CHAS** were excluded in the subsequent analysis because more than two-thirds of the observations of **ZN** are zero and the location information that **CHAS** reflects has been included in the spatially varying coefficients of the GWR model. Firstly, we considered the following GWR model between **MEDV** and the remaining eleven explanatory variables:

\[
MEDV = \beta_0(u,v) + \beta_1(u,v)CRIM + \beta_2(u,v)INDVS + \beta_3(u,v)NOX + \beta_4(u,v)RM + \beta_5(u,v)AGE + \beta_6(u,v)DIS + \beta_7(u,v)RAD + \beta_8(u,v)TAX + \beta_9(u,v)PTRATIO + \beta_{10}(u,v)LSTAT + \beta_{11}(u,v)BK + \varepsilon. \tag{25}
\]

Based on the data set, the bootstrap test was used to identify the constant coefficients in the alternative model in Equation (25). Here, the Gaussian kernel was used once again for fitting the alternative model and the bandwidth size selected by the AICc criterion is \(h_1 = 1.44\). We conducted a series of the tests for different null hypotheses that some specific coefficients in Model (25) are assumed to be constant. In each of the tests, 1000 bootstrap replications were run to compute the \(p\)-value. Some typical tests with their \(p\)-values are reported in Table 2.

From the last row of Table 2 it is concluded that, under the significance level \(\alpha = 0.05\), we cannot reject the null hypothesis that the coefficients of the explanatory variables **CRIM**, **INDVS**, **NOX**, **AGE**, **RAD**, **TAX**, **PTRATIO**, and **BK** are constants. Furthermore, the additional tests were performed by adding one or more remaining coefficients \(\beta_6(u,v)\), \(\beta_4(u,v)\), and \(\beta_{10}(u,v)\), into the null hypothesis in the last row of Table 2 and the resulting \(p\)-values are all between 0.000 and 0.017, indicating that the coefficients \(\beta_6(u,v)\), \(\beta_4(u,v)\), \(\beta_6(u,v)\), and \(\beta_{10}(u,v)\) vary over the space.

According to the above testing results, the following mixed GWR model might be appropriate to the data set. After rearranging the additive terms, this mixed GWR model stands as
Table 2. Some typical null hypotheses with their p-values of the tests under 1000 bootstrap replications.

| Null hypothesis | Corresponding variables | p-Value |
|-----------------|-------------------------|---------|
| $\beta_0(u,v) = \beta_0$ | Intercept | 0.000 |
| $\beta_1(u,v) = \beta_1$ | CRIM | 0.568 |
| $\beta_2(u,v) = \beta_2$ | INDVS | 0.667 |
| $\beta_3(u,v) = \beta_3$ | NOX | 0.095 |
| $\beta_4(u,v) = \beta_4$ | RM | 0.000 |
| $\beta_5(u,v) = \beta_5$ | AGE | 0.678 |
| $\beta_6(u,v) = \beta_6$ | DIS | 0.021 |
| $\beta_7(u,v) = \beta_7$ | RAD | 0.398 |
| $\beta_8(u,v) = \beta_8$ | TAX | 0.596 |
| $\beta_9(u,v) = \beta_9$ | PTRATIO | 0.115 |
| $\beta_{10}(u,v) = \beta_{10}$ | LSTAT | 0.000 |
| $\beta_{11}(u,v) = \beta_{11}$ | BK | 0.258 |
| $\beta_j(u,v) = \beta_j$ | | |
| $j = 1, 2, 3, 5, 7, 8, 9, 11$ | | |

For the mixed GWR model in Equation (26), of further interest is to test whether some explanatory variables with constant coefficients influence the response MEDV significantly, which amounts to the tests for identifying which constant coefficients are zero. As mentioned in Section 3.3, the bootstrap test is applicable to such kind of tests. With the Gaussian kernel and the bandwidth selection criterion in Equation (15), the optimal size of the bandwidth for the mixed GWR model in Equation (26) is $h_M = 0.68$. We first performed the tests with 1000 bootstrap replications for the null hypothesis that each single constant coefficient in Model (26) is zero. The resulting p-values are listed in Table 3.

It is known from Table 3 that, for each single explanatory variable with constant coefficient, INDVS has not significant impact on the response MEDV. Furthermore, we added one or more other constant coefficients into the null hypothesis $\beta_2 = 0$, the resulting p-values are all less than or equal to 0.005, indicating that INDVS is the only variable that does not influence MEDV significantly.

By removing the insignificant explanatory variable INDVS from Model (26) and re-ordering both the spatially varying coefficients and the constant coefficients, we finally specified a mixed GWR model for the Boston housing data as

\[
MEDV = \beta_0(u,v) + \beta_4(u,v)RM + \beta_6(u,v)DIS + \beta_{10}(u,v)LSTAT \\
+ \beta_1 CRIM + \beta_2 INDVS + \beta_3 NOX + \beta_5 AGE + \beta_7 RAD \\
+ \beta_8 TAX + \beta_9 PTRATIO + \beta_{11} BK + \varepsilon. \tag{26}
\]
\[ MEDV = \beta_0(u, v) + \beta_1(u, v)RM + \beta_2(u, v)DIS + \beta_3(u, v)LSTAT \\
+ \beta_4CRIM + \beta_5NOX + \beta_6AGE + \beta_7RAD \\
+ \beta_8TAX + \beta_9PTRATIO + \beta_{10}BK + \epsilon. \] (27)

The above model can be calibrated by the two-step method in Fotheringham et al. (2002, chapter 3) and can be used to explain the spatial characteristics of the regression relationship between the median value of owner-occupied homes and the related explanatory variables. This content is beyond the scope of this paper and is therefore omitted.

6. Final remarks

Understanding whether some of the coefficients in a GWR model really vary over the space is essential to validly explain spatial non-stationarity of the regression relationship. To achieve this task, we propose in this paper a residual-based bootstrap test to identify constant coefficients in a GWR model. The simulation study shows that the bootstrap test outperforms the existing test methods in terms of type I error and test power. Moreover, the bootstrap test is applicable to diverse hypotheses and provides a useful way of building a possible mixed GWR model for a georeferenced data set.

In terms of future research, it should be noted that only the effect of global collinearity between the explanatory variables is evaluated in the simulation study whereas the impact of local collinearity on the bootstrap and other tests remains to be investigated. Furthermore, a more important issue is how to alleviate the adverse impact of global or local collinearity on the tests and deserves to be further studied. Moreover, the geographically and temporally weighted regression (GTWR), a promising extension of the GWR model, has recently been developed to account for local effects in both space and time (Huang et al. 2010, Fotheringham et al. 2015) and has attracted a growing interest on its applications. However, the similar inferential problems to those in the GWR model are far from being solved. The bootstrap test might be a potential candidate for achieving the task, but the details still need to be clearly addressed in view of the complexity of GTWR.

Acknowledgements

The authors thank the reviewers for their helpful comments and suggestions, which lead to a great deal of improvement on the manuscript.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported by the National Natural Science Foundation of China [No. 11271296].
References

Bárcena, M.J., et al., 2014. Alleviating the effect of collinearity in geographically weighted regression. *Journal of Geographical Systems*, 16 (4), 441–466. doi:10.1007/s10109-014-0199-6

Brunsdon, C., Charlton, M., and 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., and Charlton, M., 1996. Geographically weighted regression: a method for exploring spatial non-stationarity. *Geographical Analysis*, 28 (4), 281–298. doi:10.1111/j.1538-4632.1996.tb00936.x

Brunsdon, C., Fotheringham, A.S., and Charlton, M., 1998. Geographically weighted regression: modelling spatial non-stationarity. *The Statistician*, 47 (3), 431–443.

Brunsdon, C., Fotheringham, A.S., and Charlton, M., 1999. Some notes on parametric significance tests for geographically weighted regression. *Journal of Regional Science*, 39 (3), 497–524. doi:10.1111/0022-4146.00146

Cai, Z., 2007. Trending time-varying coefficient time series models with serially correlated errors. *Journal of Econometrics*, 136 (1), 163–188. doi:10.1016/j.jeconom.2005.08.004

Casetti, E., 1972. Generating models by the expansion method: applications to geographical research. *Geographical Analysis*, 4 (1), 81–91. doi:10.1111/gean.1972.4.issue-1

Casetti, E., 1997. The expansion method, mathematical modeling, and spatial econometrics. *International Regional Science Review*, 20 (1), 9–32.

Efron, B., 1979. Bootstrap methods: another look at the jackknife. *The Annals of Statistics*, 7 (1), 1–26. doi:10.1214/aos/1176344552

Fan, J. and Huang, T., 2005. Profile likelihood inference on semiparametric varying-coefficient partial linear models. *Bernoulli*, 11 (6), 1031–1057.

Fan, J. and Jiang, J., 2007. Nonparametric inference with generalized likelihood ratio tests. *Test*, 16 (3), 409–444.

Fan, J., Zhang, C., and Zhang, J., 2001. Generalized likelihood ratio statistics and Wilks phenomenon. *The Annals of Statistics*, 29 (1), 153–193.

Fan, J. and Zhang, W., 1999. Statistical estimation in varying coefficient models. *The Annals of Statistics*, 27 (5), 1491–1518.

Fan, J. and Zhang, W., 2008. Statistical methods with varying coefficient models. *Statistics and Its Interface*, 1 (1), 179–195.

Farber, S. and Páez, A., 2007. A systematic investigation of cross-validation in GWR model estimation: empirical analysis and Monte Carlo simulations. *Journal of Geographical Systems*, 9 (4), 371–396.

Fotheringham, A.S., Brunsdon, C., and Charlton, M., 2002. *Geographically weighted regression: the analysis of spatially varying relationships*. Chichester, UK: John Wiley and Sons.

Fotheringham, A.S., Crespo, R., and Yao, J., 2015. *Geographically and temporally weighted regression* (GTWR). *Geographical Analysis*, 47 (4), 431–452.

Gilley, O.W. and Pace, R.K., 1996. On the Harrison and Rubinfeld data. *Journal of Environmental Economics and Management*, 31 (3), 403–405.

Gollini, I., et al., 2015. GW model: an R package for exploring spatial heterogeneity using geographically weighted models. *Journal of Statistical Software*, 63 (17), 1–50.

Harris, P., et al., 2015. Using bootstrap methods to investigate coefficient non-stationarity in regression model: an empirical case study. *Procedia Environmental Sciences*, 27, 112–115.

Huang, B., Wu, B., and Barry, M., 2010. Geographically and temporally weighted regression for modeling spatio-temporal variation in house prices. *International Journal of Geographical Information Science*, 24 (3), 383–401.

Hurrison, D. and Rubinfeld, D.L., 1978. Hedonic housing prices and the demand for clean air. *Journal of Environmental Economics and Management*, 5 (1), 81–102.

Jetz, W., Rahbek, C., and Lichstein, J.W., 2005. Local and global approaches to spatial data analysis in ecology. *Global Ecology and Biogeography*, 14 (1), 97–98. doi:10.1111/geb.2005.14.issue-1

Jones III, J.P. and Casetti, E., 1992. *Applications of the Expansion Method*. London: Routledge.
Leung, Y., Mei, C.-L., and Zhang, W.-X., 2000. Statistical tests for spatial nonstationarity based on the geographically weighted regression model. Environment and Planning A, 32 (1), 9–32. doi:10.1068/a3162

Mei, C.-L., He, S.-Y., and Fang, K.-T., 2004. A Note on the Mixed Geographically Weighted Regression Model. Journal of Regional Science, 44 (1), 143–157. doi:10.1111/j.1085-9489.2004.00331.x

Mei, C.-L., Wang, N., and Zhang, W.-X., 2006. Testing the importance of the explanatory variables in a mixed geographically weighted regression model. Environment and Planning A, 38 (3), 587–598. doi:10.1068/a3768

Páez, A., Farber, S., and Wheeler, D.C., 2011. A simulation-based study of geographically weighted regression as a method for investigating spatially varying relationships. Environment and Planning A, 43 (12), 2992–3010. doi:10.1068/a44111

Páez, A., Uchida, T., and Miyamoto, K., 2002a. A general framework for estimation and inference of geographically weighted regression models: 1. Location-specific kernel bandwidths and a test for locational heterogeneity. Environment and Planning A, 34 (4), 733–754. doi:10.1068/a34110

Páez, A., Uchida, T., and Miyamoto, K., 2002b. A general framework for estimation and inference of geographically weighted regression models: 2. Spatial association and model specification tests. Environment and Planning A, 34 (5), 883–904. doi:10.1068/a34133

Wang, N., Mei, C.-L., and Yan, X.-D., 2008. Local linear estimation of spatially varying coefficient models: an improvement on the geographically weighted regression technique. Environment and Planning A, 40 (4), 986–1005. doi:10.1068/a3941

Wheeler, D.C., 2007. Diagnostic tools and a remedial method for collinearity in geographically weighted regression. Environment and Planning A, 39 (10), 2464–2481. doi:10.1068/a38325

Wheeler, D.C., 2009. Simultaneous coefficient penalization and model selection in geographically weighted regression: the geographically weighted lasso. Environment and Planning A, 41 (3), 722–742. doi:10.1068/a40256

Wheeler, D.C. and Calder, C., 2007. An assessment of coefficient accuracy in linear regression models with spatially varying coefficients. Journal of Geographical Systems, 9 (2), 145–166. doi:10.1007/s10109-006-0040-y

Wheeler, D.C. and Tiefelsdorf, M., 2005. Multicollinearity and correlation among local regression coefficients in geographically weighted regression. Journal of Geographical Systems, 7 (2), 161–187. doi:10.1007/s10109-005-0155-6

Yang, W., Fotheringham, A.S., and Harris, P., 2011. Model selection in GWR: the development of a flexible bandwidth GWR. London: Geocomputation 2011, UK.

Yang, W., Fotheringham, A.S., and Harris, P., 2012. An extension of geographically weighted regression with flexible bandwidths. Lancaster: GISRUK, UK.