Pak. J. Statist.
2015 Vol. 31(3), 307-325

STATISTICAL INFERENCE OF GEOGRAPHICALLY AND TEMPORALLY WEIGHTED REGRESSION MODEL

Haiyan Xuan[1§], Shuaifeng Li[2] and Muhammad Amin[3§]

[1] School of Economics and Management, Lanzhou University of Technology, Lanzhou, P.R. China.

[2] College of Applied Science and Technology, Hainan University, Hainan, P.R. China.

[3] Nuclear Institute for Food and Agriculture (NIFA) Peshawar, Pakistan.

[§] Corresponding authors Email: haiyanxuan@gmail.com
aminkanju@gmail.com

ABSTRACT

The fundamental issues of statistical inference related to geographically and temporally weighted regression (GTWR) model are studied. Initially, the test statistics for hypothesis testing problems of global stationarity, spatial nonstationarity and temporal nonstationarity are proposed by analysis of variance technique. The heteroscedasticity in GTWR model is detected and SCORE test statistic is provided. Finally, an approximation method is proposed to compute the *p*-values for aforementioned test statistics. A Simulation study is carried out to assess the performance of these test methods, and a real example of per capita GDP in Chinese 92 cities is given.

KEYWORDS

Geographically and temporally weighted regression; spatial non-stationarity; temporal non-stationarity; heteroscedasticityn.

1. INTRODUCTION

In many applied research fields such as geography, economics and epidemiology, the data are generally belongs to the geographical locations. This type of data is called spatial data. It is well known that the variation of the geographical location in the spatial data analysis and the relationship of exchange of variables lead to spatial non-stationarity.To study the characteristic of spatial non-stationarity, Foster and Gorr (1986), Gorr and Olligschlaeger (1994) proposed spatially-varying parameter regression model of the following form

$$y_i = \sum_{k=0}^{d} \beta_k\,(u_i, v_i) x_{ik} + \varepsilon_i \qquad (1)$$

Just assume that $x_{i0} \equiv 1$, can make the model (1) contain intercept term. Where $(y_i; x_{i1}, x_{i2}, \cdots, x_{id})$ are observations of the response variable Y and explanatory variables $X_1, X_2, \cdots, X_d$ at location (u_i, v_i) in the study region, $\varepsilon_i (i = 1,2, \cdots, n)$ is the error term with mean zero and common variance σ^2. $\beta_k(u_i, v_i)$ is $d + 1$ unknown function of geographical locations for the k^{th} independent variable $(k = 0,1,2, \cdots, d)$. Spatially-varying parameter regression model assumes that the regression coefficients are the functions of geographical locations, the spatial characteristics of data are involved in the model. This type of model creates condition for exploring spatial non-stationarity of the regression. Fotheringham and his colleagues proposed a well-known fitting method to estimate the unknown parameters in the model (1) based on the idea of local polynomial smoothing in 1996. This method describes the weight function depend on the geographical location of the data by locally weighted least squares method, and available to the estimated values of each parameter for each location. These estimated values in each geographical location can effectively explain the spatial variation characteristics of the observed data. The geographically weighted regression (GWR) technique has a great attraction in analysis of spatial data and has been successfully applied to such kind of practical problems. The main results regarding this topic are summarized in Fotheringham et al. (2002).

However, in many cases, the structure of data not only belongs to the geographical location but also is related to the time factor. These type of data sets have the following characteristics: at a given time surface, it belongs to the spatial data, related with the geographical location; in a specific geographical location, the observed data is time series, related with the time factor, such kind of observed data is called spatio-temporal data.

Recently, in order to embed the spatial and temporal characteristics of the data in the regression model, Huang et al. (2010) proposed geographically and temporally weighted regression model (GTWR) of the form

$$y_i = \sum_{k=0}^{d} \beta_k (u_i, v_i, t_i) x_{ik} + \varepsilon_i \tag{2}$$

Just assume that $x_{i0} \equiv 1$, can make the model (2) contain spatial and temporal variability intercept term. Where $(y_i; x_{i1}, x_{i2}, \cdots, x_{id})$ are observations of the response variable Y and explanatory variables $X_1, X_2, \cdots, X_d$ at location u_i, v_i, t_i in the study region, $\varepsilon_i (i = 1,2, \cdots, n)$ is the error term having zero mean and constant variance σ^2. $\beta_k(u_i, v_i, t_i)$ $(k = 0,1,2, \cdots, d)$ is $d + 1$ unknown function of geographical locations and times under observation.

The GTWR model is the extension of GWR model. This model assumes that the regression coefficients are the functions of geographical locations and observed times in varying-coefficient model. Spatial and temporal characteristics of data are involved in the GTWR model, which provide the base to explore the spatial non-stationarity and temporal non-stationarity. Huang et al. (2010) proposed a procedure to fit the GTWR model, and also given the related select principle of weight function and cross-validation for fixing bandwidth parameter. In order to test its improved performance, GTWR was compared with global ordinary least squares, temporal weighted regression (TWR), and GWR in terms of goodness-of-fit and other statistical measures by using a case study of residential housing sales. The results show that there were substantial benefits in modeling both spatial and temporal non-stationarity simultaneously (Huang et al. 2010).

This study mainly focuses on the development of formal statistical testing procedures related to the GTWR model in (2). In Section 2, we construct a statistic for examining the goodness-of-fit of geographically and temporally weighted regression model versus linear regression model. Section 3 suggests a procedure for testing a variation of the parameters over geographical locations, which is very important in exploring spatial non-stationarity. Section 4 describes a procedure for testing variation of the parameters over observation times, which is very important in exploring temporal non-stationarity. In Section 5, we detect heteroscedasticity in GTWR model. The SCORE test statistic is given. Finally, in Section 6, an approximation method is proposed for computing the p values of the earlier test statistics.

2. TESTING FOR GLOBAL STATIONARITY

Based on geographically and temporally weighted regression model (2), in order to check whether the regression relationship is global stationarity, we use the test problem for a given spatio-temporal data set $(y_i; x_{i1}, x_{i2}, \cdots, x_{id})$, $i = 1,2, \cdots, n$: whether the goodness-of-fit of geographically and temporally weighted regression model is significantly better than linear regression model. The form of linear regression model is as follows

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_d x_{id} + \varepsilon_i \tag{3}$$

Through this test, we can judge the stationarity in the relationship between the dependent variable Y and the independent variables $X_1, X_2, \cdots, X_d$.

By using the above analysis, the hypothesis can be written as:

$$\begin{cases} H_0^{(1)}: y_i = \beta_0 x_{i0} + \beta_1 x_{i1} + \cdots + \beta_d x_{id} + \varepsilon_i \\ \quad H_1 : \ y_i = \sum_{k=0}^{d} \beta_k(u_i, v_i, t_i)\, x_{ik} + \varepsilon_i \end{cases} \tag{4}$$

Just assume that $x_{i0} \equiv 1$. If $H_0^{(1)}$ is true, we consider that the regression relationship varies non-significantly over geographical locations and observed time.

The linear regression model is fitted by the ordinary least squares procedure under the null hypothesis $H_0^{(1)}$, thus we can obtain the vector of fitted values and the residual sum of squares. Their forms are as follows:

$$\hat{Y}_H = HY \tag{5}$$

$$RSS\left(H_0^{(1)}\right) = Y^T(I-H)^T(I-H)Y = Y^T R_H Y \tag{6}$$

where

$$H = X(X^T X)^{-1} X^T \tag{7}$$

$$R_H = (I-H)^T(I-H) \tag{8}$$

$$Y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}, X = \begin{pmatrix} 1 & x_{11} & \dots & x_{1d} \\ 1 & x_{21} & \dots & x_{2d} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & \dots & x_{nd} \end{pmatrix} \tag{9}$$

Under the alternative hypothesis H_1, the GTWR model is fitted using weighted least squares method, the estimation of $\beta(u_i, v_i, t_i)$ can be followed as (Huang et al. 2010)

$$\begin{aligned}\hat{\beta}(u_i,v_i,t_i) &= (\hat{\beta}_0(u_i,v_i,t_i)\ \hat{\beta}_1(u_i,v_i,t_i) \cdots\ \hat{\beta}_d(u_i,v_i,t_i))^T \\ &= [X^T W(u_i,v_i,t_i)X]^{-1} X^T W(u_i,v_i,t_i)Y \end{aligned} \tag{10}$$

$$W(u_i,v_i,t_i) = \mathrm{diag}\left(w_1(u_i,v_i,t_i),\ \ w_2(u_i,v_i,t_i),\ \ \ldots,\ \ w_n(u_i,v_i,t_i)\right)$$

where $w_i(u_0,v_0,t_0) = K_{h_0}(d_{0i}),\ i = 1,2,\ldots n$, (u_0,v_0,t_0) is any given point in the study area. $d_{0i}, i = 1,2,\ldots n$ are the distance between (u_0,v_0,t_0) and (u_i,v_i,t_i). Let $(u_0,v_0,t_0) = (u_i,v_i,t_i), i = 1,2,\ldots n$, then we get the weighting matrix $W(u_i,v_i,t_i)$ with kernal function and bandwidth h_0.

Let $\mathrm{X_i^T} = (1 \quad x_{i1} \quad x_{i2} \quad \cdots \quad x_{id})$ be the ith row of X. Then the fitted value of y_i is

$$\begin{aligned}\hat{y}_i &= X_i^T \hat{\beta}(u_i,v_i,t_i) \\ &= X_i^T\ [X^T W(u_i,v_i,t_i)X]^{-1} X^T W(u_i,v_i,t_i)Y \end{aligned} \tag{11}$$

Let $\hat{Y} = (\hat{y}_1 \quad \hat{y}_2 \quad \ldots \quad \hat{y}_n)^T$ is the vector of the fitted values, $\hat{\varepsilon} = (\hat{\varepsilon}_1 \quad \hat{\varepsilon}_2 \quad \ldots \quad \hat{\epsilon}_n)^T$ is the vector of the residuals. Then

$$\hat{Y}_S = SY \tag{12}$$

$$\varepsilon = Y - \hat{Y} = (I - S)Y \tag{13}$$

where

$$S = \begin{pmatrix} X_1^T [X^T W(u_1,v_1,t_1)X]^{-1} X^T W(u_1,v_1,t_1)Y \\ X_2^T [X^T W(u_2,v_2,t_2)X]^{-1} X^T W(u_2,v_2,t_2)Y \\ \vdots \\ X_n^T [X^T W(u_n,v_n,t_n)X]^{-1} X^T W(u_n,v_n,t_n)Y \end{pmatrix}$$

The residual sum of squares of the model can be expressed as follows:

$$RSS(H_1) = Y^T (I - S)^T (I - S)Y = Y^T R_S Y \tag{14}$$

where

$$R_S = (I - S)^T (I - S) \tag{15}$$

A test statistic is constructed as

$$F_1 = \frac{RSS\left(H_0^{(1)}\right) - RSS(H_1)}{RSS(H_1)} = \frac{Y^T (R_H - R_S)Y}{Y^T R_S Y} \tag{16}$$

If the alternative hypothesis H_1 is true, $RSS\left(H_0^{(1)}\right)$ will be sufficiently larger than $RSS(H_1)$, F_1 will have increasing trend, otherwise, F_1 a decreasing trend. Therefore, the p-value of the statistic for testing $H_0^{(1)}$ vs H_1 is

$$p_1 = P_{H_0^{(1)}}(F_1 > f_1) \tag{17}$$

where, the observation f_1 of F_1 is obtained by the formula (16). For a given significance level α, if $p_1 \geq \alpha$, accept $H_0^{(1)}$ that the regression relationship varies nonsignificantly over geographical locations and observation times. If $p_1 < \alpha$, accept H_1 that the regression relationship is significant over geographical locations and observation times.

3. TESTING FOR SPATIAL NON-STATIONARITY

Based on geographically and temporally weighted regression model (2), in order to know whether the regression relationship is spatial non-stationarity, we submit the test problem for a given spatio-temporal data set $(y_i; x_{i1}, x_{i2}, \cdots, x_{id})$, $i = 1,2,\cdots,n$: whether the goodness-of-fit of geographically and temporally weighted regression model is significantly better than temporally weighted regression model. The form of temporally weighted regression model is as follows

$$y_i = \sum_{k=0}^{d} \beta_k\,(t_i) x_{ik} + \varepsilon_i \quad (18)$$

This test can help us to decide the significance of spatial non-stationarity regarding the relationship between the dependent variable Y and the independent variables $X_1, X_2, \cdots, X_d$.

Based on the above analysis, the hypothesis is as follows:

$$\begin{cases} H_0^{(2)}: y_i = \sum_{k=0}^{d} \beta_k(t_i)\, x_{ik} + \varepsilon_i \\ H_1: y_i = \sum_{k=0}^{d} \beta_k(u_i, v_i, t_i)\, x_{ik} + \varepsilon_i \end{cases} \quad (19)$$

Just assume that $x_{i_0} \equiv 1$. In case of true hypothesis $H_0^{(2)}$, we believe that the regression relationship varies non-significantly over geographical locations.

The temporally weighted regression model is fitted by the local linear fitting method under the null hypothesis $H_0^{(2)}$ (Fan et al. 1999). Let all of $\beta_k(\cdot)$,$k = 0,1,2,\cdots,d$ have continuous second derivative, the linear function approximation of regression coefficients can be expressed by the Taylor formula at the neighborhood of a given point t_0. Their forms are as follows:

$$\beta_k(t) \approx a_k(t_0) + b_k(t_0)(t - t_0),\ k = 0,1,2,\cdots,d \quad (20)$$

As a result, the estimation problem of regression coefficient upon temporally weighted regression model become the problem of locally weighted least squares. For a given kernel function $K(\cdot)$ and smooth parameter h, $(a_k(t_0), b_k(t_0)(t - t_0))$, $k = 0,1,2,\cdots,d$ can be estimated by minimizing

$$\sum_{i=1}^{n} \left[y_i - a_0(t_0) + b_0(t_0)(t - t_0) - \sum_{k=1}^{d} \left(a_k(t_0) + b_k(t_0)(t - t_0)\right) x_{ik}\right]^2 \cdot K_h(t_i - t_0) \quad (21)$$

where

$$\mathrm{K_h}(\cdot) = \mathrm{K}(\cdot/\mathrm{h})/\mathrm{h} \quad (22)$$

Let the form of matrix X_0 is as follows:

$$\begin{pmatrix} 1 & t_1 - t_0 & x_{11} & x_{11}(t_1 - t_0) & \ldots & x_{1d} & x_{1d}(t_1 - t_0) \\ 1 & t_2 - t_0 & x_{21} & x_{21}(t_2 - t_0) & \ldots & x_{2d} & x_{2d}(t_2 - t_0) \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 1 & t_n - t_0 & x_{n1} & x_{n1}(t_n - t_0) & \ldots & x_{nd} & x_{nd}(t_n - t_0) \end{pmatrix}$$

and

$$W_0 = diag(K_h(t_1 - t_0) \quad K_h(t_2 - t_0) \quad \ldots \quad K_h(t_n - t_0))$$

Then the estimated values of $a_k(t_0)$ and $b_k(t_0)$ can be obtained at the point of $(k = 0,1,2,\cdots,d)$. The form of matrix can be expressed as

$$\begin{aligned}\hat{\beta}(t_0) &= (\hat{a}_0(t_0) \quad \hat{b}_0(t_0) \quad \hat{a}_1(t_0) \quad \hat{b}_1(t_0) \quad \cdots \quad \hat{a}_d(t_0) \quad \hat{b}_d(t_0))^T \\ &= [{X_0}^T W_0 X_0]^{-1} {X_0}^T W_0 Y\end{aligned} \tag{23}$$

$\hat{\beta}_k(t_0) = \hat{a}_k(t_0)$ can be seen from the formula (20). Then

$$\hat{\beta}_k(t_0) = e_{2k+1,2d+2}^T [{X_0}^T W_0 X_0]^{-1} {X_0}^T W_0 Y \tag{24}$$

where $e_{2k+1,2d+2}$ is $2d+2$ dimensional column vector where its $(2k+2)$th row element is 1, the remaining elements are 0.

Let $t_0 = t_i$ for each $k = 0,1,2,\cdots,d$, the estimated values of $\beta_k(t)$ can be expressed at each point t_i as follows $(i = 1,2,\cdots,n)$

$$\hat{\beta}_k(t_i) = e_{2k+1,2d+2}^T [{X_i}^T W_i X_i]^{-1} {X_i}^T W_i Y \tag{25}$$

where X_i and W_i are obtained by substituting $t_0 = t_i$ into the above formulas X_0 and W_0. Then, the fitted value of y_i can be expressed as

$$\begin{aligned}\hat{y}_i &= \hat{\beta}_0(t_i) + \hat{\beta}_1(t_i)x_{i1} + \hat{\beta}(t_i)x_{i2} + \cdots + \hat{\beta}(t_i)x_{id} \\ &= (1 \quad x_{i1} \quad x_{i2} \quad \cdots \quad x_{id})(\hat{\beta}_0(t_i) \quad \hat{\beta}_1(t_i) \quad \cdots \quad \hat{\beta}_d(t_i))^T \\ &= (1 \quad x_{i1} \quad x_{i2} \quad \cdots \quad x_{id}) \begin{pmatrix} 1 & 0 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 1 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 1 & 0 \end{pmatrix} [{X_i}^T W_i X_i]^{-1} {X_i}^T W_i Y \\ &= x_i^T N [{X_i}^T W_i X_i]^{-1} {X_i}^T W_i Y\end{aligned}$$

where N is $(d+1) \times (2d+2)$ order matrix that its $(k+1)$th row element is 1, the remaining elements are 0 $(k = 0,1,2,\cdots,d)$.

Thus, the vector of fitted values can be expressed as

$$\hat{Y}_M = (\hat{y}_0 \quad \hat{y}_1 \quad \cdots \quad \hat{y}_n)^T = MY \tag{26}$$

where

$$M = \begin{pmatrix} x_1^T N [{X_1}^T W_1 X_1]^{-1} {X_1}^T W_1 \\ x_2^T N [{X_2}^T W_2 X_2]^{-1} {X_2}^T W_2 \\ \vdots \\ x_n^T N [{X_n}^T W_n X_n]^{-1} {X_n}^T W_n \end{pmatrix} \tag{27}$$

The residual vector can be expressed as

$$\hat{\varepsilon} = (\hat{\varepsilon}_1 \quad \hat{\varepsilon}_2 \quad \ldots \quad \hat{\epsilon}_n)^T = Y - \hat{Y}_M = (I - M)Y \tag{28}$$

The residual sum of squares can be expressed as

$$RSS\left(H_0^{(2)}\right) = Y^T (I-M)^T (I-M) Y = Y^T R_M Y \tag{29}$$

where

$$R_M = (I-M)^T (I-M) \tag{30}$$

The vector of fitted values and the residual sum of squares for geographically and temporally weighted regression model under the alternative hypothesis H_1 can be expressed as (12) and (14).

A test statistic is constructed as

$$F_2 = \frac{RSS\left(H_0^{(2)}\right) - RSS(H_1)}{RSS(H_1)} = \frac{Y^T(R_M - R_S)Y}{Y^T R_S Y} \tag{31}$$

If the alternative hypothesis H_1 is true, $RSS\left(H_0^{(2)}\right)$ will be sufficiently larger than $RSS(H_1)$, F_2 has an increasing trend, otherwise, F_2 has an decreasing trend. Therefore, the p-value of the statistic for testing $H_0^{(2)}$ vs H_1 is

$$p_2 = P_{H_0^{(2)}}(F_2 > f_2) \tag{32}$$

where, the observation f_2 of F_2 is obtained by the formula (31). For a given significance level α, if $p_2 \geq \alpha$, accept $H_0^{(2)}$ that the regression relationship varies nonsignificantly over geographical locations. If $p_2 < \alpha$, accept H_1 that the regression relationship varies significantly over geographical locations.

4. TESTING FOR TEMPORAL NON-STATIONARITY

By considering the GTWR model (2), for the sake of exploring whether the regression relationship is temporal non-stationarity, we submit the test problem for a given spatio-temporal data set $(y_i; x_{i1}, x_{i2}, \cdots, x_{id})$, $i = 1,2,\cdots,n$: whether the goodness-of-fit of geographically and temporally weighted regression model is significantly better than geographically weighted regression model. The form of geographically weighted regression model is as follows

$$y_i = \sum_{k=0}^{d} \beta_k (u_i, v_i) x_{ik} + \varepsilon_i \tag{33}$$

$$\begin{cases} H_0^{(3)}: y_i = \sum_{k=0}^{d} \beta_k (u_i, v_i)\, x_{ik} + \varepsilon_i \\ H_1: y_i = \sum_{k=0}^{d} \beta_k (u_i, v_i, t_i)\, x_{ik} + \varepsilon_i \end{cases} \tag{34}$$

Just assume that $x_{i0} \equiv 1$. If $H_0^{(3)}$ is true, we consider that the regression relationship varies nonsignificantly over observation times.

The spatially-varying parameter regression model is fitted by the geographical weighted regression technique under the null hypothesis $H_0^{(3)}$. Then, a vector of fitted values of the dependent variable Y can be expressed as

$$\hat{Y}_L = (\hat{y}_1 \quad \hat{y}_2 \quad \dots \quad \hat{y}_n)^T = LY \tag{35}$$

where

$$L = \begin{pmatrix} x_1^T [X^T W(u_1, v_1) X]^{-1} X^T W(u_1, v_1) \\ x_2^T [X^T W(u_2, v_2) X]^{-1} X^T W(u_2, v_2) \\ \vdots \\ x_n^T [X^T W(u_n, v_n) X]^{-1} X^T W(u_n, v_n) \end{pmatrix} \tag{36}$$

$$W(u_i, v_i) = diag\left(w_1(u_i, v_i),\ \ w_2(u_i, v_i),\ \ \ldots,\ \ w_n(u_i, v_i)\right)$$

where $w_i(u_0, v_0) = K_{h_1}(d_{0i}), i = 1,2, \ldots n$, (u_0, v_0) is any given point in the study area, $d_{0i}, i = 1,2, \ldots n$ are the distance between (u_0, v_0) and (u_i, v_i). Let $(u_0, v_0) = (u_i, v_i), i = 1,2, \ldots n$, then we get the weighting matrix $W(u_i, v_i)$ with kernal function and bandwidth h_1.

The residual vector can be expressed as

$$\hat{\varepsilon} = (\hat{\varepsilon}_1 \quad \hat{\varepsilon}_2 \quad \ldots \quad \hat{\varepsilon}_n)^T = Y - \hat{Y}_L = (I - L)Y \tag{37}$$

The residual sum of squares can be expressed as

$$RSS\left(H_0^{(3)}\right) = Y^T(I - L)^T(I - L)Y = Y^T R_L Y \tag{38}$$

where

$$R_L(I - L)^T(I - L) \tag{39}$$

The vector of fitted values and the residual sum of squares for geographically and temporally weighted regression model under the alternative hypothesis H_1 can be written as (12) and (14).

A test statistic is constructed as

$$F_3 = \frac{RSS\left(H_0^{(3)}\right) - RSS(H_1)}{RSS(H_1)} = \frac{Y^T(R_L - R_S)Y}{Y^T R_S Y} \tag{40}$$

If the alternative hypothesis H_1 is true, $RSS\left(H_0^{(3)}\right)$ will be sufficiently larger than $RSS(H_1)$, F_3 has an increasing trend, otherwise, F_3 has an decreasing trend. Therefore, the p-value of the statistic for testing $H_0^{(3)}$ vs H_1 is

$$p_3 = P_{H_0^{(3)}}(F_3 > f_3) \tag{41}$$

where, the observation f_3 of F_3 is obtained by the formula (40). For a given significance level α, if $\mathrm{p}_3 \geq \alpha$, accept $H_0^{(3)}$ that the regression relationship varies nonsignificantly over observation times. If $\mathrm{p}_3 < \alpha$, accept H_1 that the regression relationship varies significantly over observation times.

5. STATISTICAL ANALYSIS OF HETEROSCEDASTICITY

For the GTWR model (2), it is assumed generally that $E(\varepsilon_i) = 0$, while variance is the function of geographical locations and observation times. For the similicity, the variance function may express as $Var(\varepsilon_i) = \sigma^2 G(\alpha, Z_i)$. Where $G(\alpha, Z_i)$ is relevant function of the variable $Z_i = (u_i, v_i, t_i)^T$ and unknown parameter α. $G(\cdot)$ is a known function that there is a unique α_0, which make $G(\alpha_0, Z_i) = 1$ can be established for all of i. Thus, the heteroscedasticity testing problem for model (2) is equivalent to the following hypothesis testing problem:

$$H_0: \alpha = \alpha_0 \Leftrightarrow H_1^{(1)}: \alpha \neq \alpha_0 \tag{42}$$

The model (2) is provided with heteroscedasticity under the alternative hypothesis H_1, $\varepsilon_i \sim N(0, \delta^2 G(\alpha, Z_i))$. The penalized log-likelihood function of model (2) can be expressed as (Sliverman et al. 1993)

$$L(\alpha, \beta_i, \sigma^2) = -\frac{n}{2} ln 2\pi - \frac{1}{2}\sum_{i=1}^{n} lnG(\alpha, Z_i) - \sum_{i=1}^{n} \frac{\varepsilon_i^2}{2\sigma^2} G(\alpha, Z_i) - \frac{\lambda}{2}\sum_{i=1}^{n} \beta^T(i)\Omega\beta(i) - \frac{n}{2} ln\sigma^2 \quad (43)$$

where $\beta(i) = (\beta_0(u_i, v_i, t_i), \beta_1(u_i, v_i, t_i), \cdots, \beta_d(u_i, v_i, t_i))^T$, Ω is a diagonal matrix, which is related to observed point (u_i, v_i, t_i).

Let $\theta = (\alpha, \beta(i), \sigma^2)$, and $\hat{\theta}$ is the estimated value of the parameter θ in model (2) under the null hypothesis H_0. By fitting the model (2) under the null hypothesis H_0, the estimated value of the parameter can be obtained. Then the residual at point u_i, v_i, t_i is defined as

$$\hat{\varepsilon}_i = y_i - \sum_{k=0}^{d} \hat{\beta}_k (u_i, v_i, t_i) x_{ik} \quad (44)$$

Just assume that $x_{i0} \equiv 1$, The estimated value of σ^2 is denoted as follows

$$\sigma^2 = \frac{1}{n}\sum_{i=1}^{n} \hat{\varepsilon}_i^2 \quad (45)$$

For the heteroscedasticity testing problem of the model (2), the construction method of test statistic is summarized into the following lemma:

Lemma 1:

If $M = \left(\frac{\partial G(\alpha, Z_1)}{\partial \alpha} \quad \frac{\partial G(\alpha, Z_2)}{\partial \alpha} \quad \dots \quad \frac{\partial G(\alpha, Z_n)}{\partial \alpha}\right)^T |_{\hat{\theta}}$, $\bar{M} = (I - 11^T/n) M$, $R = (\hat{\varepsilon}_1^2/\hat{\sigma}^2 \quad \hat{\varepsilon}_2^2/\hat{\sigma}^2 \quad \dots \quad \hat{\varepsilon}_n^2/\hat{\sigma}^2)_{n\times 1}^T$, where I is n order identity matrix, $1 = (1 \quad 1 \quad \dots \quad 1)_{n\times 1}^T$. Then, the SCORE test statistic of the problem (42) can be expressed as

$$SC = \frac{1}{2} R^T \bar{M} (\bar{M}^T \bar{M})^{-1} \bar{M}^T R \quad (46)$$

Proof:

For the penalized log-likelihood function $L(\theta)$ such as formula (43), the partial derivatives on the parameter α, $\beta(i)$ and σ^2 can be expressed as follows

$$\frac{\partial L(\theta)}{\partial \alpha} = -\frac{1}{2}\sum_{i=1}^{n} \frac{1}{G(\alpha, Z_i)} \frac{\partial G(\alpha, Z_i)}{\partial \alpha} + \frac{1}{2}\sum_{i=1}^{n} \frac{\varepsilon_i^2}{\sigma^2 G^2(\alpha, Z_i)} \frac{\partial G(\alpha, Z_i)}{\partial \alpha}$$

$$= \frac{1}{2}\sum_{i=1}^{n} \frac{1}{G(\alpha, Z_i)} \frac{\partial G(\alpha, Z_i)}{\partial \alpha} [\frac{\varepsilon_i^2}{\sigma^2 G(\alpha, Z_i)} - 1]$$

$$\frac{\partial L(\theta)}{\partial \beta(i)} = \sum_{i=1}^{n} \frac{y_i - x_i^T \beta(i)}{\sigma^2 G^2(\alpha, Z_i)} x_i^T - \lambda \sum_{i=1}^{n} \beta^T(i)\, \Omega$$

$$\frac{\partial L(\theta)}{\partial \sigma^2} = \frac{1}{2}\sum_{i=1}^{n} \frac{\varepsilon_i^2}{\sigma^4 G(\alpha, Z_i)} - \frac{n}{2\sigma^2}$$

$$\frac{\partial^2 L(\theta)}{\partial \alpha \partial \alpha^T} = \frac{1}{2}\sum_{i=1}^{n} \left\{ \left[\frac{\varepsilon_i^2}{\sigma^2 G(\alpha, Z_i)} - 1\right] \frac{\partial}{\partial \alpha^T} \left[\frac{1}{G(\alpha, Z_i)} \frac{\partial G(\alpha, Z_i)}{\partial \alpha}\right] - \frac{\varepsilon_i^2}{\sigma^2 G^3(\alpha, Z_i)} \frac{\partial G(\alpha, Z_i)}{\partial \alpha} \frac{\partial G(\alpha, Z_i)}{\partial \alpha^T} \right\}$$

$$\frac{\partial^2 L(\theta)}{\partial\alpha\partial\sigma^2} = -\frac{1}{2}\sum_{i=1}^{n}\frac{\partial G(\alpha,Z_i)}{\partial\alpha}\frac{\varepsilon_i^2}{\sigma^4 G^2(\alpha,Z_i)}$$

$$\frac{\partial^2 L(\theta)}{\partial\sigma^2\partial\alpha^T} = -\frac{1}{2}\sum_{i=1}^{n}\frac{\partial G(\alpha,Z_i)}{\partial\alpha^T}\frac{\varepsilon_i^2}{\sigma^4 G^2(\alpha,Z_i)}$$

$$\frac{\partial^2 L(\theta)}{\partial\sigma^2\partial\sigma^2} = \frac{n}{2\sigma^4} - \sum_{i=1}^{n}\frac{\varepsilon_i^2}{\sigma^6 G^2(\alpha,Z_i)}$$

$$\frac{\partial L(\theta)}{\partial\beta(i)\partial\alpha^T} = -\sum_{i=1}^{n}\frac{\varepsilon_i x_i^T}{\sigma^2 G^2(\alpha,Z_i)}\frac{\partial G(\alpha,Z_i)}{\partial\alpha^T}$$

$$\frac{\partial L(\theta)}{\partial\beta(i)\partial\sigma^2} = -\sum_{i=1}^{n}\frac{\varepsilon_i x_i^T}{\sigma^4 G(\alpha,Z_i)}$$

$$\frac{\partial L(\theta)}{\partial\beta(i)\partial\beta^T(i)} = -\sum_{i=1}^{n}\frac{x_i x_i^T}{\sigma^2 G(\alpha,Z_i)} - \lambda\sum_{i=1}^{n}\Omega$$

$$\frac{\partial^2 L(\theta)}{\partial\alpha\partial\beta^T(i)} = 0$$

$$\frac{\partial^2 L(\theta)}{\partial\sigma^2\partial\beta^T(i)} = 0$$

Let J is Fisher information array about the parameter θ, and $J^{\alpha\alpha}$ is block matrix corresponding to α. At present, the estimated value of parameter θ be $\hat{\theta}$ under the null hypothesis H_0, and $\partial G(\alpha_0, Z_i) = 1$. Thus

$$\frac{\partial L(\theta)}{\partial\alpha}|_{\hat{\theta}} = \frac{1}{2}M^T(R-1) \tag{47}$$

The form of Fisher information array J can be expressed as

$$J(\Theta) = \begin{pmatrix} A_{11} & A_{12} & A_{13} \\ A_{21} & A_{22} & A_{23} \\ A_{31} & A_{32} & A_{33} \end{pmatrix}$$

where, A_{11} is corresponding to parameter α. A_{22} is corresponding to parameter $\beta(i)$, and A_{33} is corresponding to parameter σ^2, the elements are computed at $\theta = \hat{\theta}$. It is noted that $E(\varepsilon_i^2) = \sigma^2 G(\alpha_0, Z_i) = \sigma^2$. Thus

$$A_{11} = E\left[-\frac{\partial^2 L(\theta)}{\partial\alpha\partial\alpha^T}\right] = \frac{1}{2}M^T M$$

$$A_{12} = E\left[\frac{\partial^2 L(\theta)}{\partial\alpha\partial\beta^T(i)}\right] = 0$$

$$A_{13} = E\left[-\frac{\partial^2 L(\theta)}{\partial\alpha\partial\sigma^2}\right] = \frac{1}{2\hat{\sigma}^2}M^T 1$$

$$A_{21} = E\left[-\frac{\partial^2 L(\theta)}{\partial\beta(i)\partial\alpha^T}\right] = 0$$

$$A_{22} = E\left[-\frac{\partial^2 L(\theta)}{\partial\beta(i)\partial\beta^T(i)}\right] = \frac{X^T X}{\hat{\sigma}^2} + n\lambda\Omega$$

$$A_{23} = E\left[-\frac{\partial^2 L(\theta)}{\partial\beta(i)\partial\sigma^2}\right] = 0$$

$$A_{31} = E\left[-\frac{\partial^2 L(\theta)}{\partial\sigma^2 \partial\alpha^T}\right] = \frac{1}{2\hat{\sigma}^2} 1^T$$

$$A_{32} = E\left[-\frac{\partial^2 L(\theta)}{\partial\sigma^2 \partial\beta^T(i)}\right] = 0$$

$$A_{33} = E\left[-\frac{\partial^2 L(\theta)}{\partial\sigma^2 \partial\sigma^2}\right] = \frac{n}{2\hat{\sigma}^4}$$

$J^{\alpha\alpha}$ can be obtained by inverse matrix algorithm in block matrix. We have

$$\begin{aligned}
J^{\alpha\alpha} &= \left[A_{11} - (A_{12} \quad A_{13}) \begin{pmatrix} A_{22} & A_{23} \\ A_{32} & A_{33} \end{pmatrix}^{-1} \begin{pmatrix} A_{21} \\ A_{31} \end{pmatrix}\right]^{-1} \\
&= \left[A_{11} - (0 \quad A_{13}) \begin{pmatrix} A_{22} & 0 \\ 0 & A_{33} \end{pmatrix}^{-1} \begin{pmatrix} A_{21} \\ A_{31} \end{pmatrix}\right]^{-1} \\
&= \left[A_{11} - (0 \quad A_{13}{A_{33}}^{-1}) \begin{pmatrix} A_{21} \\ A_{31} \end{pmatrix}\right]^{-1} \\
&= \left[A_{11} - A_{13}{A_{33}}^{-1}A_{31}\right]^{-1} \\
&= \left[\frac{1}{2}M^T M - \frac{1}{2\hat{\sigma}^2} M^T 1 \cdot \frac{2\hat{\sigma}^4}{n} \cdot \frac{1}{2\hat{\sigma}^2} 1^T M\right]^{-1} \\
&= \left[\frac{1}{2}M^T (I - 11^T/n)^T (I - 11^T/n) M\right]^{-1} \\
&= 2(\bar{M}^T \bar{M})^{-1}
\end{aligned}$$

For hypothesis testing problem (42), the SCORE test statistic (Eubank. 1988) can be expressed as

$$\begin{aligned}
SC &= \left[\left(\frac{\partial L(\theta)}{\partial\alpha}\right)^T J^{\alpha\alpha} \left(\frac{\partial L(\theta)}{\partial\alpha}\right)\right]_{\theta=\hat{\theta}} \\
&= \frac{1}{2}(R-1)^T M \cdot 2(\bar{M}^T \bar{M})^{-1} \cdot \frac{1}{2} M^T (R-1) \\
&= \frac{1}{2}\left[R^T \left(I - \frac{11^T}{n}\right) M (\bar{M}^T \bar{M})^{-1} M^T \left(I - \frac{11^T}{n}\right) R\right] \\
&= \frac{1}{2} R^T \bar{M} (\bar{M}^T \bar{M})^{-1} \bar{M}^T R
\end{aligned}$$

Thus, the formula (46) is established.

If the alternative hypothesis H_1 is true, the test statistic SC has an increasing trend, otherwise, SC has an decreasing trend. Therefore, the p-value of the statistic for testing H_0 vs H_1 is

$$p_0 = P_{H_0}(SC > sc_0) \tag{48}$$

where, the observation sc_0 of SC is obtained by the formula (46). For a given significance level α, if $p_0 \geq \alpha$, accept H_0; otherwise, do not accept H_0.

For simplicity, if the variance structure of ε_i is only related with one of u, v, t in the model (2), let

$$G(\alpha, t_i) = exp(\alpha t_i^2) \quad (49)$$

and $m_i = \frac{\partial G(\alpha, Z_i)}{\partial \alpha}|_{\hat{\theta}}$. At present, we should pay attention to m_i is a numerical value but not a matrix. We have

$$m_i = \frac{\partial G(\alpha, Z_i)}{\partial \alpha}|_{\hat{\theta}} = t_i^2 ex\,p(\alpha_0 t_i^2) = t_i^2 \quad (50)$$

Then

$$\begin{aligned} R^T \bar{M} &= (R-1)^T M \\ &= \begin{pmatrix} \hat{\varepsilon}_1^2/\hat{\sigma}^2 - 1 \\ \hat{\varepsilon}_2^2/\hat{\sigma}^2 - 1 \\ \vdots \\ \hat{\varepsilon}_n^2/\hat{\sigma}^2 - 1 \end{pmatrix}^T \begin{pmatrix} m_1 \\ m_2 \\ \vdots \\ m_n \end{pmatrix} \\ &= \sum_{i=1}^n m_i (\hat{\varepsilon}_i^2/\hat{\sigma}^2 - 1) \end{aligned}$$

and

$$\begin{aligned} &\bar{M}^T \bar{M} \\ &= \begin{pmatrix} m_1 \\ m_2 \\ \vdots \\ m_n \end{pmatrix}^T \begin{pmatrix} 1-\frac{1}{n} & -\frac{1}{n} & \cdots & -\frac{1}{n} \\ -\frac{1}{n} & 1-\frac{1}{n} & \cdots & -\frac{1}{n} \\ \vdots & \vdots & \ddots & \vdots \\ -\frac{1}{n} & -\frac{1}{n} & \cdots & 1-\frac{1}{n} \end{pmatrix} \begin{pmatrix} m_1 \\ m_2 \\ \vdots \\ m_n \end{pmatrix} \\ &= \sum_{i=1}^n m_i^2 - \left(\sum_{i=1}^n m_1\right)^2/n \end{aligned}$$

Thus, the SCORE test statistic can be expressed as

$$\begin{aligned} SC &= \frac{1}{2} R^T \bar{M} (\bar{M}^T \bar{M})^{-1} \bar{M}^T R \\ &= \frac{\left[\sum_{i=1}^n m_i (\hat{\varepsilon}_i^2/\hat{\sigma}^2 - 1)\right]^2}{2\left[\sum_{i=1}^n m_i^2 - (\sum_{i=1}^n m_1)^2/n\right]} \end{aligned} \quad (51)$$

6. COMPUTATION OF P-VALUE

In the case of the previous fitting methods, even if the error terms $\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_n$ in the model are assumed to be independent and identically distributed as normal distribution $N(0, \sigma^2)$, F_1 in (16) is generally not distributed as an F-distribution because R_S is not idempotent and the numerator and denominator are not independent. However, under the assumption that the fitted value $\hat{y}_i$ of the dependent variable is unbiased estimate of $E(y_i)$, that is, $E(\hat{y}_i) = E(y_i)$. F_1 can be expressed as a ratio of quadratic forms in error terms so that we can use some distributional results of quadratic forms of normal variates to compute the p-value if we further assume that the error terms are normally distributed.

For this purpose, the following lemma is given:

Lemma 1:

Under the assumption that the fitted value $\hat{y}_i$ of the dependent variable is the unbiased estimate of $E(y_i)$, F_1 can be expressed as

$$F_1 = \frac{RSS\left(H_0^{(1)}\right) - RSS(H_1)}{RSS(H_1)} = \frac{\varepsilon^T (R_H - R_S)\varepsilon}{\varepsilon^T R_S \varepsilon} \tag{52}$$

where $\varepsilon = (\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_n)^T$ is the error vector of the model.

Proof:

under the null hypothesis $H_0^{(1)}$, a linear regression model is valid to the data, and it is well known that

$$RSS\left(H_0^{(1)}\right) = \varepsilon^T R_H \varepsilon \tag{53}$$

On the other hand, under the assumption of $E(\hat{y}_i) = E(y_i)$, we have

$$\begin{aligned}(I - S)Y &= Y - E(Y) + E\left(\hat{Y}\right) - \hat{Y} \\ &= \varepsilon - [SY - SE(Y)] \\ &= \varepsilon - S\varepsilon \\ &= (I - S)\varepsilon\end{aligned} \tag{54}$$

Thus

$$\begin{aligned}RSS(H_1) &= Y^T (I - S)^T (I - S)Y \\ &= \varepsilon^T (I - S)^T (I - S)\varepsilon \\ &= \varepsilon^T R_S \varepsilon\end{aligned} \tag{55}$$

The lemma is then proved by substituting (53) and (55) into (16).

Similarly, we can prove that F_2 and F_3 can be expressed as

$$F_2 = \frac{\varepsilon^T (R_M - R_S)\varepsilon}{\varepsilon^T R_S \varepsilon} \tag{56}$$

$$F_3 = \frac{\varepsilon^T (R_L - R_S)\varepsilon}{\varepsilon^T R_S \varepsilon} \tag{57}$$

This section provides F distribution approximation method (Cleveland et al. 1988) to compute the previous p-values. The main idea of F distribution approximation method is to approximate normal variable quadratic distribution of the numerator and denominator in the test statistic by χ^2 distribution with appropriate multiples and degrees of freedom respectively, then the distribution of test statistic is approximated by F distribution with appropriate degrees of freedom. Below, we discuss a brief introduction about approximation method for F distribution.

Let $Z = \xi^T A \xi$, where $\xi \sim N(0, I)$ and A is n order semi-positive definite real symmetric matrix. If A is idempotent matrix, Z is a random variable distributed as χ^2 distribution with degrees of freedom $tr(A)$. The distribution of Z is generally approximated by $a\chi_b^2$. Through determining the constants a and b, make the expectation and variance of Z be equal to the expectation and variance of $a\chi_b^2$.

$$\mathrm{E}(a\chi_b^2) = ab,\ Var(a\chi_b^2) = 2a^2 b, \text{ and } E(Z) = tr(A),\ Var(Z) = 2tr(A^2).$$

Let

$$\begin{cases} ab = tr(A) \\ 2a^2 b = 2tr(A^2) \end{cases}$$

Then

$$a = \frac{tr(A^2)}{tr(A)}\ b = \frac{[tr(A)]^2}{tr(A^2)} \quad (58)$$

For quadratic form of normal variate $F = \frac{\xi^T A_1 \xi}{\xi^T A_2 \xi}$, where $\xi \sim N(0, I)$, A_1 and A_2 are semi-positive definite real symmetric matrices. If A_1 and A_1 are idempotent, and $A_1 A_2 = 0$, then F is distributed as F distribution with degrees of freedom $tr(A_1)$, and $tr(A_2)$. The distribution of $\xi^T A_i \xi$ is generally approximated by $a_i \chi^2_{b_i}$ for $i = 1,2$. Let

$$\tilde{F} = \frac{\left(\frac{1}{a_1}\xi^T A_1 \xi\right)/b_1}{\left(\frac{1}{a_2}\xi^T A_2 \xi\right)/b_2} = \frac{tr(A_1)}{tr(A_2)} F \quad (59)$$

thus, we apply $\mathrm{F}(\mathrm{b}_1, \mathrm{b}_2)$ to approximate the distribution of $\tilde{\mathrm{F}}$.

Under the assumption that the fitted values $\hat{\mathrm{y}}_{\mathrm{i}}$ of the dependent variable is the unbiased estimate of $\mathrm{E}(\mathrm{y}_{\mathrm{i}})$, we provide the approximate computing formula of p-value based on F distribution approximation method.

For the test statistic F_1, we have

$$\begin{aligned} p_1 &= P_{H_0^{(1)}}(F_1 > f_1) \\ &= P_{H_0^{(1)}}\left(\frac{\varepsilon^T (R_H - R_S)\varepsilon}{\varepsilon^T R_S \varepsilon} > f_1\right) \\ &= P_{H_0^{(1)}}\left(\frac{tr(R_S)}{tr(R_H - R_S)}\frac{\varepsilon^T (R_H - R_S)\varepsilon}{\varepsilon^T R_S \varepsilon} > \frac{tr(R_S)}{tr(R_H - R_S)} f_1\right) \\ &\approx P\left(F(r_{H-S}, r_S) > \frac{tr(R_S)}{tr(R_H - R_S)} f_1\right) \end{aligned}$$

where $r_{H-S} = \frac{[tr(R_H - R_S)]^2}{tr(R_H - R_S)^2}$, $r_S = \frac{[tr(R_S)]^2}{tr(R_S^2)}$. $F(r_{H-S}, r_S)$ is a random variable distributed as F distribution with degrees of freedom r_{H-S}, r_S.

Similarly, for the test statistics F_2 and F_3, we have

$$p_2 = P_{H_0^{(1)}}(F_2 > f_2) \approx P\left(F(r_{H-S}, r_S) > \frac{tr(R_S)}{tr(R_H - R_S)} f_2\right)$$

where $r_{M-S} = \frac{[tr(R_M - R_S)]^2}{tr(R_M - R_S)^2}$, $r_S = \frac{[tr(R_S)]^2}{tr(R_S^2)}$. $F(r_{M-S}, r_S)$ is a random variable distributed as F distribution with degrees of freedom r_{M-S}, r_S.

$$p_3 = P_{H_0^{(1)}}(F_3 > f_3) \approx P\left(F(r_{L-S}, r_S) > \frac{tr(R_S)}{tr(R_L - R_S)} f_3\right)$$

where $r_{L-S} = \frac{[tr(R_L - R_S)]^2}{tr(R_L - R_S)^2}$, $r_S = \frac{[tr(R_S)]^2}{tr(R_S^2)}$. $F(r_{L-S}, r_S)$ is a random variable distributed as F distribution with degrees of freedom r_{L-S}, r_S.

Here, we provide the approximate formula to compute p-value for the test statistic SC. Let

$$\widetilde{SC} = \sqrt{SC} \quad (60)$$

that is

$$\widetilde{SC} = \frac{1}{\left[2\sum_{i=1}^{n} m_i^2 - 2\left(\sum_{i=1}^{n} m_i\right)^2/n\right]^{\frac{1}{2}}}\left(\sum_{i=1}^{n} m_i \hat{\varepsilon}_i^2/\hat{\sigma}^2 - \sum_{i=1}^{n} m_i\right)$$

$$= \frac{1}{\varphi_2}\left[\frac{nY^T(I-S)^T D(I-S)Y}{Y^T(I-S)^T(I-S)Y} - \varphi_1\right]$$

where

$$D = diag(m_1 \quad m_2 \quad \cdots \quad m_n)$$

$$\varphi_1 = tr(D),\ \varphi_2 = [2\sum_{i=1}^{n} m_i^2 - 2\left(\sum_{i=1}^{n} m_i\right)^2/n]^{\frac{1}{2}}$$

If the estimation bias of regression function can be ignored, that is, $E\left(Y - \hat{Y}_S\right) = 0$. Then, we have

$$\widetilde{SC} = \frac{1}{\varphi_2}\left[\frac{n\varepsilon^T(I-S)^T D(I-S)\varepsilon}{\varepsilon^T(I-S)^T(I-S)\varepsilon} - \varphi_1\right]$$

$$= \frac{1}{\varphi_2}\left(\frac{n\varepsilon^T M_1 \varepsilon}{\varepsilon^T M_2 \varepsilon} - \varphi_1\right)$$

where

$$M_1 = (I-S)^T D(I-S)\ M_2 = (I-S)^T(I-S)$$

Then

$$P_{H_0^{(1)}}(SC > sc_0)$$

$$= P_{H_0^{(1)}}\left(\widetilde{SC} > \sqrt{sc_0}\right) + P_{H_0}\left(\widetilde{SC} < -\sqrt{sc_0}\right)$$

$$= 1 - P_{H_0^{(1)}}\left(-\sqrt{sc_0} < \widetilde{SC} < \sqrt{sc_0}\right)$$

$$= 1 - P_{H_0^{(1)}}\left(\frac{\varphi_1 - \varphi_2\sqrt{sc_0}}{n} < \frac{\varepsilon^T M_1 \varepsilon}{\varepsilon^T M_2 \varepsilon} < \frac{\varphi_1 + \varphi_2\sqrt{sc_0}}{n}\right)$$

$$= 1 - P_{H_0^{(1)}}\left(\frac{tr(M_2)}{tr(M_1)}\frac{\varphi_1 - \varphi_2\sqrt{sc_0}}{n} < \frac{tr(M_2)}{tr(M_1)}\frac{\varepsilon^T M_1 \varepsilon}{\varepsilon^T M_2 \varepsilon} < \frac{tr(M_2)}{tr(M_1)}\frac{\varphi_1 + \varphi_2\sqrt{sc_0}}{n}\right)$$

$$\approx 1 - \mathrm{P}\left(\frac{tr(M_2)}{tr(M_1)}\frac{\varphi_1 - \varphi_2\sqrt{sc_0}}{n} < F(c_1, c_2) < \frac{tr(M_2)}{tr(M_1)}\frac{\varphi_1 + \varphi_2\sqrt{sc_0}}{n}\right)$$

where $c_1 = \frac{[tr(M_1)]^2}{tr\left(M_1^2\right)}$, $c_2 = \frac{[tr(M_2)]^2}{tr\left(M_2^2\right)}$. $F(c_1, c_2)$ is a random variable distributed as F distribution with c_1 and c_2 degrees of freedom.

7. TESTING FOR SPATIAL NON-STATIONARITY

To examine the applicability of GTWR, a case study was implemented using per capita GDP observed between 2004 and 2013 in Chinese 92 cities. The main purpose of the analysis is to explore the underlying spatio-temporal patterns of per capita GDP in the mainland of China and to analyze of the factors which affect per capita GDP.

GDP is often used as indicators of the development of economics. It can not only reflect a country's economic performance, but also reflect the strength and wealth of a country. In general, the GDP has four different components, including consumption, investment, government spending and net exports.

Taking indexes of per capita GDP as dependent variable in this article and selecting macroeconomics indexes which can affect the GDP as independent variable. A set of 3680 observations were available. Some macroscopic factors which would influence the indexes of GDP were obtained. The data got from China city statistical yearbook.

Let Y represent per capita GDP, X_1 represent the per capita investment in fixed assets, X_2 represent the per capita public finance expenditure, X_3 represent the per capita total volume of retail sales of the social consumer goods. The model for analyzing the data is of the form

$$y_i = \beta_0(u_i, v_i, t_i) + \sum_{j=1}^{3} \beta_j\,(u_i, v_i, t_i)x_{ij} + \varepsilon_i \; i = 1,2,\cdots,92$$

where $\beta_0(u_i, v_i, t_i)$ represents the average rate of per capita GDP with spatio-temporal location. $\beta_1(u_i, v_i, t_i)$ represents the average rate of per capita GDP With the per capita investment in fixed assets in each cities. $\beta_2(u_i, v_i, t_i)$ represents the average rate of per capita GDP With the per capita public finance expenditure in each cities. $\beta_3(u_i, v_i, t_i)$ represents the average rate of per capita GDP With the per capita total volume of retail sales of the social consumer goods in each cities.

Table 1
Bandwidths and $\boldsymbol{p}$-values with the Local Linear Estimation

h_S	h_T	p	p_0	p_1	p_2	p_3
46km	0.014year	0	0	0.0038117	0.0012413	1.6665E-6

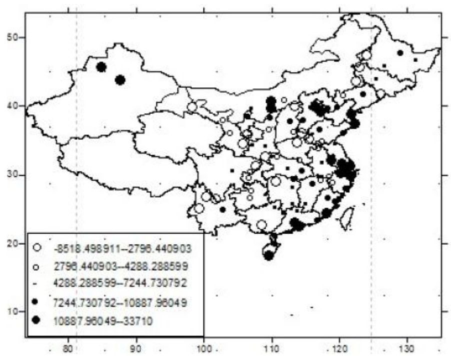

Fig. 1: The $\boldsymbol{\beta_0}$ Distribution of 92 Cities

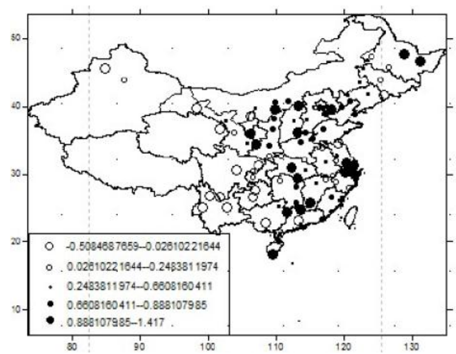

Fig. 2: The $\boldsymbol{\beta_1}$ Distribution of 92 Cities

Since GTWR, locally linear TWR and GWR method needs Cartesian coordinates of the spatial locations. We first transform the longitude and latitude of each meteorological station into the Cartesian coordinate. However, location and time are usually measured in different units (in our case, location in kilometers and time in years), thus they have different scale effects. We choose parameter of spatial and temporal bandwidths, respectively. We choose the Gaussian kernel function and use cross-validation method to determine the bandwidths.

According to section 3-6, we construct the test statistic and respectively obtain the global non-stationary test p-value and the p-values of temporal and spatial nonstationarity.

According to significance test method which was proposed by Mei and Wang (2012), we obtain respectively the significance test p-values p_0, p_1, p_2, p_3 of the regression coefficient functions $\beta_0(u_i, v_i, t_i)$, $\beta_1(u_i, v_i, t_i)$, $\beta_2(u_i, v_i, t_i)$, $\beta_3(u_i, v_i, t_i)$, which reflect the significant changes of the regression coefficients with spatio-temporal location by using the GTWR estimation method.

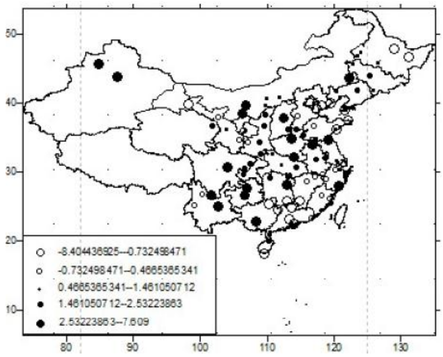

Fig. 3: The β_2 Distribution of 92 Cities

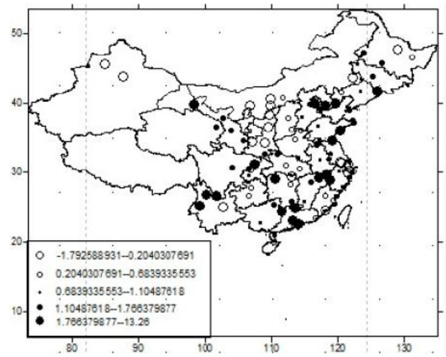

Fig. 4: The β_3 Distribution of 92 Cities

The results of global significance test in Table 1 shows that the global non-stationary test p-value of regression model and the significance test p-values of the regression coefficient functions are all very small (approximately 0). The results show that it has significant influence of the per capita investment in fixed assets, per capita public finance expenditure and per capita total volume of retail sales of the social consumer goods on per capita GDP. It also shows the coefficient functions vary with spatio-temporal location.

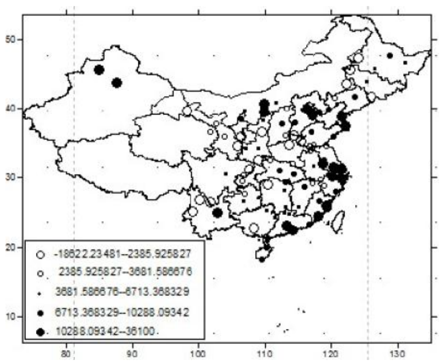

Fig. 5: The β_0 Distribution by GWR

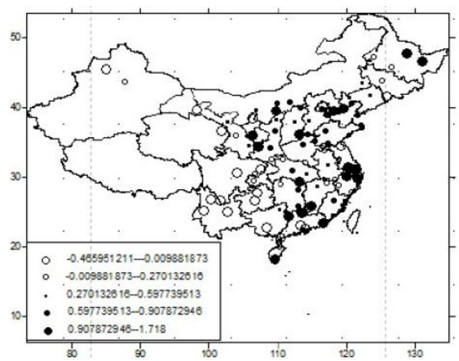

Fig. 6: The β_1 Distribution by GWR

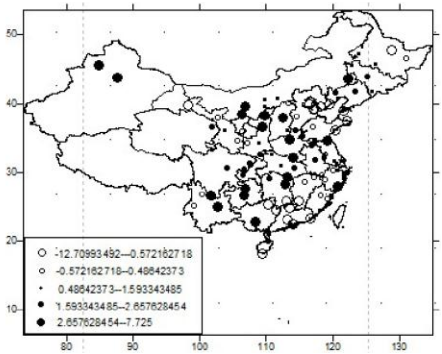

Fig. 7: The $\boldsymbol{\beta_2}$ Distribution by GWR

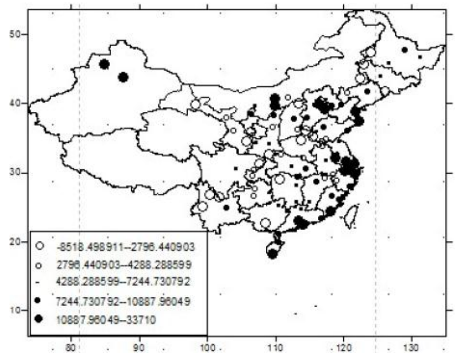

Fig. 8: The $\boldsymbol{\beta_3}$ Distribution by GWR

According to the spatio-temporal data set, based on the results calculated by SAS software, we use Surfer software to draw the distribution maps of $\beta_0(u_i, v_i, t_i)$, $\beta_1(u_i, v_i, t_i)$, $\beta_2(u_i, v_i, t_i)$, $\beta_3(u_i, v_i, t_i)$ of 92 cities in 2013. They are shown in Fig. 1-4. It can be seen from Fig. 1-4 that the distribution of estimated parameters shows an evident spatial heterogeneity over the mainland of China.

Thus, the spatial distribution of $\beta_0(u_i, v_i, t_i)$ is the highest of the coast, the Beijing Tianjin, Wulumuqi, Kelamayi, Eerduosi. And the other places are lower than these cities. The spatial distribution of $\beta_1(u_i, v_i, t_i)$ shows a increasing trend from western to east, which has the similar changes with the distribution of per capita GDP. The spatial distribution of $\beta_2(u_i, v_i, t_i)$ shows a increasing trend from the southeast to the inland. The spatial distribution of $\beta_3(u_i, v_i, t_i)$ is the lowest of inland of China, increasing from inside to outside.

The results show that there are significant spatial correlation and spatial nonstationarity of country in China. The conclusions are as follow. Investment factor has a positive effect on the level of the country economy, but the contribution to the economic development gradually decreased.

The p-values of temporal nonstationarity is 8.793E-11. The p-values of spatial nonstationarity is 0, so we can get that it has temporal and spatial non-stationary.

It can be seen from the result obtained by using the locally linear TWR method, there is no significant spatial variation over time. This is somewhat trite, because TWR only models temporal heterogeneity, which indicates that the spatial variation of this coefficient is not obvious. Moreover, the spatial variation in GWR and GTWR share analogous distributions (we can see from Fig. 5-8), except that the spatial variation in the GTWR model portrays heterogeneity in more detail. It can be inferred that the spatio-temporal nonstationarity of the GTWR model is dominated by the spatial effect for the test data set.

8. CONCLUSION

The current paper studies three cases of hypothesis testing such as global stationarity, spatial non-stationarity and temporal non-stationarity. Moreover, some significant issues regarding geographically and temporally weighted regression model are also investigated. By considering some mild assumptions and fitting the corresponding models, the respective test statistics are constructed. Furthermore, the Score test statistic applicable to geographically and temporally weighted regression model has been structured and it is applied to test the significance of the error term' heteroscedasticity. For the above test statistic, it is assumed that the error term follow a normal distribution, therefore, we use F distribution approximation approach for calculating the p-values. As demonstrated by the simulation results above, geographically and temporally weighted regression in the practical application of our research is a very important tool and has practical significance.

REFERENCES

1. Foster, S.A. and Gorr, W.L. (1986). An adaptive filter for estimating spatially-varying parameters: application to modeling police hours spent in response to calls for service. *Management Science*, 32(7), 878-889.
2. Gorr, W.L. and Olligschlaeger, A.M. (1994). Weighted spatial adaptive filtering: monte carlo studies and application to illicit drug market modeling. *Geographical Analysis*, 26(1), 67-87.
3. Fotheringham, A.S., Brunsdon, C. and Charlton, M. (2002). *Geographically Weighted Regression - the Analysis of Spatially Varying Relationships*. Wiley, Chichester.
4. 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.
5. Fan, J. and Zhang, W. (1999). Statistical estimation in varying coefficient models. *The Annals of Statistics*, 27(5), 1491-1518.
6. Green, P.J. and Sliverman, B.W. (1993). *Nonparametric Regression and Generalized Linear Models*. Chapman and Hall, London.
7. Eubank, R.L. (1988). *Spline Smoothing and Nonparametric Regression*. Marcel Dekker, Inc, New York.
8. Cleveland, W.S. and Devlin, S.J. (1988). Locally weighted regression: an approach to regression analysis by local fitting. *Journal of the American Statistical Association*, 83(403), 596-610.
9. Mei, C. and Wang N. (2012). *Recent Regression Analysis Methods*. Science Press, Beijing, China.