國立政治大學統計學系

碩士論文

A Conditional Modification to
Geographically Weighted Regression

指導老師：余清祥 博士

研究生：梁穎誼

中華民國一百年六月

謝誌

當初在 Malaysia 什麼都不懂，來到了這塊地方。或許來之前就已經找好人生方向了，來的這裡就一直讀書，貪婪的吸取統計的知識。結果一直讀書啊，一直讀書。不知不覺讀了八年。歲月不留人，人老的特別快，以前還能蹦蹦跳跳，現在動久了就會 sakit 到鬼這樣。驚覺回頭，才發現要獲得知識並不是沒代價的，除了口袋有形的 RM(或 NTD)，還必須付出無形的健康跟時間。錢能找的回來，其它是很難的。給所有看到這篇的人，研究之餘請注意健康。

台灣其實是一個學術風氣不錯的地方。我必須感謝台灣這塊土地。很幸運的，我在台灣遇到很多貴人。其中，我想最應該感謝的是余清祥老師，老師不但在學術上給予了我良好的指導，而且也好幾次雪中送炭，讓我能夠無後顧之憂完成了這份研究。我想這篇論文有一半應屬於老師的。另外，我也必須感謝江老師，雖然在碩士階段的交集不多，不過每次我找老師幫忙，即使老師在百忙之中還幫助我。

此外，還有其他的人，包括一起學習的貝姐、BRK、燴飯、OOO。一起 drink drink 的損友。還有 mee gia, gai gai, old papa 等。Terima Kasih。

中文摘要

在二十世紀九十年代，學者提出地理加權迴歸（**Geographically Weighted Regression;**簡稱GWR)。GWR是一個企圖解決空間非穩定性的方法。此方法最大的特性，是模型中的迴歸係數可以依空間的不同而改變，這也意味著不同的地理位置可以有不同的迴歸係數。在係數的估計上，每個觀察值都擁有一個固定環寬，而估計值可以由環寬範圍內的觀察值取得。然而，若變數之間的特性不同，固定環寬的設定可能會產生不可靠的估計值。

為了解決這個問題，本文章提出CGWR(Conditional-based GWR)的方法嘗試修正估計值，允許各迴歸變數有不同的環寬。在估計的程序中，CGWR運用疊代法與交叉驗證法得出最終的估計值。本文驗證了CGWR的收斂性，也同時透過電腦模擬比較GWR, CGWR與local linear法(Wang and Mei, 2008)的表現。研究發現，當迴歸係數之間存有正相關時，CGWR比其他兩個方法來的優異。最後，本文使用CGWR分析台灣高齡老人失能資料，驗證CGWR的效果。

關鍵字：地理加權迴歸、廣義加法模型、交叉驗證法、Jacobi疊代法、電腦模擬、MAUP問題

Abstract

Geographically weighted regression (GWR), first proposed in the 1990s, is a modelling technique used to deal with spatial non-stationarity. The main characteristic of GWR is that it allows regression coefficients to vary across space, and so the values of the parameters can vary depending on locations. The parameters for each location can be estimated by observations within a fixed range (or bandwidth). However, if the parameters differ considerably, the fixed bandwidth may produce unreliable or even unstable estimates.

To deal with the estimation of greatly varying parameter values, we propose Conditional-based GWR (CGWR), where a different bandwidth is selected for each independent variable. The bandwidths for the independent variables are derived via an iteration algorithm using cross-validation. In addition to showing the convergence of the algorithm, we also use computer simulation to compare the proposed method with the basic GWR and a local linear method (Wang and Mei, 2008). We found that the CGWR outperforms the other two methods if the parameters are positively correlated. In addition, we use elderly disability data from Taiwan to demonstrate the proposed method.

Keywords: Geographically weighted regression, Generalized additive model, Cross validation, Jacobi iteration, Computer simulation, Modifiable areal unit problem

Index

Table Index

Figure Index

Chapter 1.

Introduction

In recent years, spatial regression has become an important tool for analyzing spatial data. In applying the analysis, first order stationarity is a common assumption and the expected values at different locations are assumed to be fixed. Thus, similar to the development of time series analysis, the focus of spatial regression is on the covariance structure. Some well-known spatial covariance models such as Simultaneous autoregressive (SAR) and moving average (MA) can be treated as spatial versions of the AR and MA models used in the time series analysis. However, first order stationarity is a questionable assumption in practice. For example, the term Modifiable areal unit problem (MAUP), used in the geography, particularly on the scaling effect can be treated as a spatial version of Simpson Paradox, is often encountered when dealing with spatial data. Other than the problem that the parameters' values are not identical in all locations, including the data with different attributes for estimation may also produce biased results.

In other words, estimating the parameters using all of the data and the ordinary least squares (OLS) method would distort the local distinctness. One possible solution is only considering the data from areas with similar attributes. However, it is difficult to define the locations of homogeneity. Moreover, the mean of a non-stationary process is usually continuous across the space and does not follow a step-function (Fotheringham et al., 2002), making it difficult to find the exact boundary of the appropriate data points.

The other possibility is the varying coefficient model (Cleveland et al., 1991). The model allows the coefficient terms being a functional form of a scalar,

which is related to locally linear models (Fan and Zhang, 2008). Note that, the varying coefficient model has been plugged into different kinds of application such as survival analysis, time series model, etc, and is a popular tool to explore the dynamic feature. Based on the concept of varying coefficient model, geographically weighted regression (GWR) was proposed to solve the problem of MAUP (Brunsdon et al., 1996).

The GWR allows the regression coefficients to vary across the space or map. It is used to obtain estimates from a moving data window and is analogous to kernel regression for obtaining a smoothing estimate. The OLS can be treated as a special case of GWR, i.e., a window with infinite length, but the local distinctness is likely to be lost by averaging all of the observations. The optimal length (or bandwidth) of the moving windows in GWR is usually determined by cross-validation (CV) or Akaike's information criterion (AIC), and is fixed for all variables.

Nevertheless, there is still room for improvement, mainly related to the selection and testing of the bandwidth. For example, using CV and AIC to select the optimal bandwidth is a data-driven method, and like in kernel regression, the estimates are sensitive to outliers (Farber and Páez, 2007). In addition, the data variations are not necessarily the same and the bandwidth should not be fixed as well. On the other hand, the hypothesis testing of the parameters depends on the bandwidth. For example, Leung et al. (2000) proposed some goodness-of-fit tests and found that the degree of freedom of GWR residuals is a function of the bandwidth. Since selecting the bandwidth can be somewhat subjective, it is possible to come up with totally different results by using different tests.

Several GWR modifications were proposed since its introduction, most of which are related to determining the bandwidth. Brunsdon et al. (1999) introduced

a mixed GWR model that allows the inclusion of both stationary and non-stationary coefficients. They also proposed vector bandwidths, allowing coefficients with different bandwidths via a backfitting algorithm and bandwidths that were functions of data density. Shi et al. (2006) proposed a weighting function that allows the GWR spatial weight to be influenced by the attribute, rather than the distance. Furthermore, Farber and Páez (2007) found that it is possible to reduce the bias by modifying the CV procedure. Recently, Wang et al. (2008) introduced a local linear estimation, or a polynomial fitting technique, to reduce the bias.

However, in spite of these GWR modifications, the issue that the variations of each coefficient can be very different has not yet been discussed in details. Intuitively, it is reasonable to provide separate bandwidths if the coefficients behave differently. Under the classic GWR setting, the bandwidth is the same for all of the coefficients and the estimates are likely to be over-smoothed or under-smoothed.

Meanwhile, enormous data analysis based on GWR showed that the correlations often exist between GWR coefficients. Despite most of them didn't mention the exact index, the correlations are apparent through the outcome mapping. Bivand and Brunstad (2005) even calculated the correlation index, which is nearly high correlated. Note that, the correlation can be additional information to improve the estimation, as the Control variant technique on Variance reduction. Therefore, the information should be sufficiently use rather than ignore it. Although the problem we are dealing with is not entire identical in variance reduction, the idea can be borrowed to modify the estimator.

To complete the puzzle, in this paper, we propose a method to provide separate bandwidths for different coefficients via an iteration algorithm. We also

evaluate the validity of the proposed method by comparing it with the basic GWR and local linear method by Wang et al. (2008). The next section will first introduce the GWR and its extension. Next, Section 3 will introduce proposed method, namely Conditional-based GWR (CGWR), and its theoretical results. Section 4 discusses simulation studies of comparing the performance of the proposed method with that of the basic GWR and local linear method. In addition to simulations, we also apply the proposed method to a real data set, as shown in Section 5. Section 6 provides a discussion of the proposed method that includes its limitations and scope for possible future developments.

Chapter 2.

GWR and its extension

2.1 GWR

We shall first provide a brief introduction of the GWR, followed by an introduction of the proposed method. The GWR uses a dependent variable y modelled as a linear function of a set of p independent variables, $x_1, x_2, \ldots, x_p$, and can be expressed as

$$Y_i = \beta_{i0} + \sum_{k=1}^{p} \beta_{ik} x_{ik} + \varepsilon_i, \tag{1}$$

where β_{ik} and x_{ik} are the parameter and observed value of variable k ($k = 1, \ldots, p$) for observation i, respectively, and ε_i is the error term for observation i, which is generally assumed to be from a normal distribution with zero mean and constant variance σ^2 (i.e., $\varepsilon_i \sim N(0, \sigma^2)$). The subscript i represents the spatial location of observation i ($i = 1, \ldots, n$). In other words, in the GWR model, each location can be thought of as having its own regression model for the dependent variable and independent variables. Thus, the essence of GWR is to obtain local regression with nearby data for each location, and choosing an appropriate range for the data becomes a key factor. We will use the term "bandwidth" to represent this data range for the rest of this paper.

The parameter set $\boldsymbol{\beta}_i$ of observation i can be estimated by

$$\hat{\boldsymbol{\beta}}_i = (\mathbf{X^T W_i X})^{-1} \mathbf{X^T W_i} Y, \tag{2}$$

where $\hat{\boldsymbol{\beta}}_i = (\hat{\beta}_{i0}, \hat{\beta}_{i1}, \ldots, \hat{\beta}_{ip})^T, \mathbf{X} = (\boldsymbol{I}, \boldsymbol{x_1}, \ldots, \boldsymbol{x_p})^T, Y = (Y_1, \ldots, Y_n)^T$, and $\mathbf{W_i}$ is a diagonal weight matrix. The weight matrix contains the weights of the neighbourhoods for the ith observation, and can be expressed as:

$$\mathbf{W}_i = \begin{pmatrix} w_{i1} & & \cdots & & 0 \\ & \ddots & & & \\ \vdots & & w_{ij} & & \vdots \\ & & & \ddots & \\ 0 & & \cdots & & w_{in} \end{pmatrix} \tag{3}$$

where w_{ij} ($j = 1, ..., n$) are obtained by kernel functions. The kernel functions chosen are usually distance-weighted and Gaussian type,

$$w_{ij} = \exp[-\frac{1}{2}(d_{ij}/h)^2] \tag{4}$$

However, if data locations are sparse at somewhere on the space, distance-weighted kernel might not appropriate due to lack of information. Brunsdon et al. (2002) introduced rank-based and k nearest neighbourhood method as the solutions. In addition, as we mentioned before, bandwidth selection is generally calibrated via minimizing Cross-validation score or AIC.

2.2 Local linear method

Wang et al. (2008) proposed local linear-fitting-based GWR, which is Taylor expansion version of GWR. The main essence of it is to reduce bias by using second continuous partial derivatives with respect to geographical location coordinates. For any given location, the regression parameter can be locally approximated as

$$\begin{aligned} \beta_j &\approx \beta_{ij} + \beta_{ij}^{(u)}(u - u_0) + \beta_{ij}^{v}(v - v_0) \\ j &= 1, 2, ..., p, \end{aligned} \tag{5}$$

where β_j is any of regression parameter and β_{ij} is the parameter of the given location. $\beta_{ij}^{(u)}$ and $\beta_{ij}^{(v)}$ are the partial derivatives of β_j at location i, that is, two-dimensional coordinates on the location, with respect to each axis.

6

To estimate the parameters, the matrix of independent variables can be simply rewritten as

$$\mathbf{X}_i^* = \begin{pmatrix} x_{11} & x_{11}(u_1-u_i) & x_{11}(v_1-v_i) & \cdots & {}_px_1 & {}_px_1(u_1-u_{1i}) & {}_px_1(v_1-v_i) \\ x_{21} & x_{21}(u_2-u_i) & x_{21}(v_2-v_i) & \cdots & {}_px_2 & {}_px_2(u_2-u_{2i}) & {}_px_2(v_2-v_i) \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ x_{n1} & x_{n1}(u_n-u_i) & x_{n1}(v_n-v_i) & \cdots & x_{np} & x_{np}(u_{np}-u_n)u_i & (x_{np})v \end{pmatrix} \quad (6)$$

where u and v are the coordinates of the locations. By adding intercept into the matrix, (i.e. $\mathbf{X}_i^{**} = (\mathbf{1}, \mathbf{u}\text{-}\mathbf{u}_i, \mathbf{v}\text{-}\mathbf{v}_i, \mathbf{X}_i^*)$), the parameters can be estimated by

$$\hat{\beta}_i = (\mathbf{X}_i^{*T}\mathbf{W}_i\mathbf{X}_i^*)^{-1}\mathbf{X}_i^{*T}\mathbf{Y} \quad (7)$$

which is similar to the original GWR estimator.

2.3 Drawbacks

Using a single bandwidth in the GWR is likely to create unsatisfactory estimates if the corresponding surfaces of independent variables have different attributes (such as different shapes or variations). Allowing different bandwidths seems to be a natural modification to the GWR. Unfortunately, the weighted least squares and equation (2) cannot handle the situation of varying bandwidths for different variables. Brunsdon et al. (1999) proposed an approach similar to a stepwise regression and backfitting algorithm for selecting different bandwidths. However, they did not provide suggestions for choosing bandwidths. Of course, choosing the optimal bandwidths is similar to the problem of minimization, and thus the methods such as grid search can be used. However, these methods usually require intensive computation and might not be feasible.

Local linear method solves the aforementioned problems when all the coefficient surfaces are linear, since the remainder of the Taylor expansion would be zero in this case. In other words, all bandwidths are infinite and same due to

unbiasness. But, the method suffers from the bandwidth problem again when it is not the case.

On the other hand, there is another problem that we should concern on local linear estimation. In such complex surface, the slope could different on different location, and we might expect this method would mislead the surface at somewhere slopes change rapidly, which mostly happen on non-linear surface. Finally, we would like to point that adding additional dependent variable could increase the variance too.

In this study, we introduce an approach to determine different bandwidths using iteration. This approach is inspired by the vector bandwidth of Brunsdon et al. (1999) and by the kernel smoothing used in the varying-coefficient model of Wu and Chiang (2000). We combine these ideas and propose a method (and an algorithm to establish the estimation) for selecting each variable's bandwidth. The proposed method uses some existing techniques such as the generalized additive model (GAM) and the Jacobi iteration method, which will be mentioned first.

Chapter 3.

Methodology

3.1 Generalized Additive Model

Hastie and Tishibirani (1990) proposed the GAM, replacing the linear function x_k by a polynomial function of x_k, or $m_k(x_k)$, in equation (1). This model can be written as:

$$Y_i = m_0 + m_1(x_1) + m_2(x_2) + \cdots m_{p-1}(x_{p-1}) + \varepsilon_i \qquad (8)$$

where m_0 is a constant and $m_k(x_k)$ is a function of x_k. Then, the following linear system can be used to estimate these $m_k(\cdot)$ functions:

$$\begin{bmatrix} \mathbf{I} & \mathbf{S_1} & & & \mathbf{S_1} \\ \mathbf{S_2} & \mathbf{I} & . & . & \mathbf{S_2} \\ & & . & . & \\ & & . & . & \\ \mathbf{S_p} & \mathbf{S_p} & & & \mathbf{I} \end{bmatrix} \begin{bmatrix} f_1 \\ f_2 \\ . \\ . \\ f_p \end{bmatrix} = \begin{bmatrix} \mathbf{S_1}Y \\ \mathbf{S_2}Y \\ . \\ . \\ \mathbf{S_p}Y \end{bmatrix} \qquad (9)$$

where $\mathbf{I}$ is an identity matrix, $\mathbf{S_1},\ldots,\mathbf{S_p}$ are a set of $n \times n$ smoothing matrices, and $f_1,\ldots,f_p$ are a set of $n \times 1$ vectors. Some typical form of S matrix are Moving average, Spline, B-Spline, etc. Hastie and Tisibirani (1990) applied the block Gauss-Seidel iteration procedure to estimate the $m_k(\cdot)$ function, which is also the essence of our approach. Under some regular conditions, the Gauss-Siedel iteration will converge. The reader can refer Hastie and Tisibirani (1990) for more details. In the following, we discuss the details of the Jacobi and Gauss-Seidel iterations.

3.2 Jacobi Iteration

Jacobi iteration is a two-century old numerical method to solve a linear system, $\mathbf{A}x = b$, iteratively. It is a widely-known method and has many successors such as the Gauss-Seidel iteration and the successive over-relaxation method. The

major difference between these methods involves the convergence rate, with the Jacobi iteration converging slower in most cases. However, the Jacobi iteration is easy to implement and can be written as the following steps.

Consider a linear system: $\mathbf{A}x = b$, where $\mathbf{A}$ is an $n \times n$ matrix, x and b are $n \times 1$ vectors, and x is the solution to be determined. The Jacobi iteration procedure is as follows: (Let $x_i\{l\}$ denote the lth iteration value of x_i.)

1. Set the initial solution of x_i to be 0, i.e., $x_i\{0\} = 0$.

2. For each element x_i, $x_i\{l\} = \dfrac{b_i - \sum\limits_{i \neq j} a_{ij} x_j\{l-1\}}{a_{ii}}$

3. Repeat (2) until the stopping criteria is reached.

3.3 Proposed CGWR

The proposed method has the following main characteristics: it allows a different bandwidth for each coefficient and an iteration algorithm is incorporated into the approach. We adapt the ideas of the GAM and Jacobi iteration and use them to determine appropriate bandwidths. By applying the concept behind the GAM, the GWR model can be rewritten as:

$$Y_i = f_{i1} + ... + f_{ik} + ... + f_{ip} + \varepsilon_i \tag{10}$$

where $f_{ik} = \beta_{ik} \times x_{ik}$ and β_{ik} is the parameter coefficient of the variable k at location i. If f_{ik} is the intercept, then x_{ik} is set to be 1. Then, we can use the Jacobi iteration to iteratively solve for parameter f_{ik} one-by-one. Again, we let $f_k\{l\}$ denote the lth iteration vector of f_k, and let $f_k\{l\}$ be a n×1 vector composed by f_k. The proposed method can be summarized as follows: (*CGWR procedure*)

1. Set the initial solution of f_k to be 0, or $f_k\{0\} = \mathbf{0}$, $k = 1, ..., p$, and let $l = 1$.

10

2. For each element $f_k\{l\}$, apply the basic GWR model with only one independent variable, x_k. The value of the dependent variable used is

$$y^* = y - \sum_{\substack{j=1 \\ j \neq k}}^{p} f_j\{l-1\}, \text{ (i.e., we regress } \left(y - \sum_{\substack{j=1 \\ j \neq k}}^{p} f_j\{l-1\} \right) \text{ on variable } x_k$$

without a fitting intercept). The bandwidth is obtained by minimizing the CVSS (or AIC).

3. Repeat (2) until the stopping criteria is reached.

The bandwidths are modified iteratively according to the preceding process. Note that the vector bandwidth and the mixed GWR can be treated as special cases of CGWR. The bandwidth is usually large at first, and it is possible to reach the global estimate first, before the local effect enters. We anticipate that the proposed procedure can increase the estimation accuracy because it finds the individual optimal bandwidth for each component.

There are at least two reasons for finding optimal bandwidth solutions individually using the Jacobi iteration. First, although more complex numerical methods (such as the quasi-Newton method) could be used, it can reduce the computation time and cost. Second, as mentioned previously, there are algorithms that converge faster than the Jacobi iteration, but they are likely to produce biased estimates. For example, in the Gauss-Siedel iteration process, the estimate of one component is updated based on the simultaneous estimates of other components. If the estimates of some components have severe biases, it is possible to contaminate the estimates of other components. In other words, although the proposed method is conservative, we are willing to sacrifice some of the efficiency in exchange for accuracy and stability. Therefore, unlike the vector

bandwidth and mixed GWR, the Jacobi iteration is used in the proposed method, instead of the Gauss-Siedel iteration.

The rest of this section summarizes some properties related to the CGWR. First, although Brunsdon et al. (1999) proposed an estimation method for solving different bandwidths, they did not prove the convergence of this method. We will first use the case of two coefficients as an example and give the proof of convergence in the following. Note that the method of Brunsdon et al. can be treated as a special case of CGWR (i.e., the bandwidths are never updated) and the next result might shed light on the complete proof of CGWR.

Proposition 1: *Suppose that the bandwidths are fixed. Assume that there is only an independent variable, i.e., there are two coefficients, an intercept and an explanatory variable. In addition, the bandwidths are fixed during the iteration process. If the largest absolute eigenvalue of the product of two smoothing matrices, i.e., $\mathbf{S}_1 \times \mathbf{S}_2$, is smaller than 1, then the CGWR estimation would converge.*

The outline of the proof is given in Appendix A. Note that the convergence can be extended to a case with more than two coefficients (or two or more independent variables), with a similar proof.

Unfortunately, showing the convergence of the CGWR estimation (i.e., changing bandwidths) is more complicated, although it is similar to that in Proposition 1. Nonetheless, we consider several scenarios and the estimation process seems to produce stable and reliable estimates. In the next section, we will evaluate the stability of the CGWR via computer simulation, and compare it with the basic GWR and a local linear method.

For the remainder of this section, we introduce some results of the CGWR, and also provide possible explanations. First, the estimation variances with

12

CGWR are usually smaller than those for the basic GWR. We will provide a heuristic proof, since the smoothing matrix of CGWR cannot be written in a closed form. For simplification, we take only the case of two components as an example. If there are only two components, then the variance of each estimator ($V\left(\hat{f}_1^{GWR}\right)$, $V\left(\hat{f}_1^{CGWR}\right)$) can be written as the following equation (11), where $\mathbf{S}_i\,(i=1,2)$ is the corresponding CGWR smoothing matrix. Obviously, the denominator of the basic GWR is smaller than that of CGWR, and thus the variance of CGWR is smaller.

$$\hat{f}_1^{GWR} = \mathbf{S}\mathbf{Y} \Rightarrow V\left(\hat{f}_1^{GWR}\right) = \sigma^2 \mathbf{S}\mathbf{S}^{\mathbf{T}}$$

$$\hat{f}_1^{CGWR} = \mathbf{S}^*\mathbf{Y} \Rightarrow V\left(\hat{f}_1^{CGWR}\right) = \sigma^2 \mathbf{S}^*\mathbf{S}^{*\mathbf{T}}$$

$$\left(\mathbf{S}^* = \mathbf{S}_1 - \mathbf{S}_1\mathbf{S}_2 + \mathbf{S}_1\mathbf{S}_2\mathbf{S}_1 - \mathbf{S}_1\mathbf{S}_2\mathbf{S}_1\mathbf{S}_2 + ... + \mathbf{R}_1\right), \text{ where } \mathbf{R}_1 \text{ denote the remainder term.}$$

$$\mathbf{S} = \begin{bmatrix} \dfrac{w_{11}\left[\left(\sum_i w_{1i}x_i^2\right)-\left(\sum_i w_{1i}x_i\right)x_1\right]}{\left(\sum_i w_{1i}\right)\left(\sum_i w_{1i}x_i^2\right)-\left(\sum_i w_{1i}x_i\right)^2} & \cdots & \dfrac{w_{1n}\left[\left(\sum_i w_{1i}x_i^2\right)-\left(\sum_i w_{1i}x_i\right)x_n\right]}{\left(\sum_i w_{1i}\right)\left(\sum_i w_{1i}x_i^2\right)-\left(\sum_i w_{1i}x_i\right)^2} \\ \vdots & \ddots & \vdots \\ \dfrac{w_{n1}\left[\left(\sum_i w_{ni}x_i^2\right)-\left(\sum_i w_{ni}x_i\right)x_1\right]}{\left(\sum_i w_{ni}\right)\left(\sum_i w_{ni}x_i^2\right)-\left(\sum_i w_{ni}x_i\right)^2} & \cdots & \dfrac{w_{1n}\left[\left(\sum_i w_{ni}x_i^2\right)-\left(\sum_i w_{ni}x_i\right)x_n\right]}{\left(\sum_i w_{ni}\right)\left(\sum_i w_{ni}x_i^2\right)-\left(\sum_i w_{ni}x_i\right)^2} \end{bmatrix}_{n\times n}$$

$$\mathbf{S}_1 = \begin{bmatrix} \dfrac{w1_{11}}{\sum_i w1_{1i}} & \cdots & \dfrac{w1_{1n}}{\sum_i w1_{1i}} \\ \vdots & \ddots & \vdots \\ \dfrac{w1_{n1}}{\sum_i w1_{ni}} & \cdots & \dfrac{w1_{nn}}{\sum_i w1_{ni}} \end{bmatrix}_{n\times n} \qquad \mathbf{S}_2 = \begin{bmatrix} \dfrac{w2_{11}x_1x_1}{\sum_i w2_{1i}x_i^2} & \cdots & \dfrac{w2_{1n}x_1x_n}{\sum_i w2_{1i}x_i^2} \\ \vdots & \ddots & \vdots \\ \dfrac{w2_{n1}x_nx_1}{\sum_i w2_{ni}x_i^2} & \cdots & \dfrac{w2_{n1}x_nx_n}{\sum_i w2_{ni}x_i^2} \end{bmatrix}_{n\times n} \qquad (11)$$

Moreover, there is no guarantee that the proposed method works when two or more coefficients are negatively correlated with each other. A possible explanation is that the dissimilarity of the coefficients would inflate the mean-bias, which will be zero if the coefficients are perfectly positively correlated. This is similar to factor analysis, since if two or more variables are highly correlated, a single factor (i.e., variable) can provide a good summary of these variables. However, if two variables are negatively correlated, different signs would complicate the convergence process.

Another explanation is the unstable of negative correlation matrix. If all the element on the correlation matrix greater than zero, there are some perfect properties such as all eigenvalue will be the real number that smaller or equal to 1 . However, it doesn't hold when some of them are negative as results as bad estimation.

Finally, there are two intrinsic meanings on the term 'conditional'. To begin with, CGWR update each variable conditionally by fixing others. In fact, the concept is more like the Hasting algorithm in Monte Carlo Markov Chain method. The constraint on CGWR is another aspect, since CGWR only works well when inter-correlation between variables is positive.

Chapter 4.

Simulation Study

We use simulations to evaluate the performance of the proposed CGWR, comparing it to the basic GWR and a local linear estimation. These simulation studies are separated into two parts: fixed effect model and random effect model. For the fixed effect model, we examine both single type and mixed type surfaces. The difference between these two scenarios is whether the coefficients' surfaces are similar or different. Moreover, we assume that the coefficients at different locations follow one of the following four surfaces functions: linear, quadratic, ridge and hillside. To simplify the discussion, we assume that there were two coefficients, i.e., one intercept and one independent variable,

$$Y_i = \beta_{i0} + \beta_{i1}x_i + \varepsilon_i \qquad i = 1, 2, ..., n \tag{12}$$

where i indicates the location of the observation.

We should define the term S/N ratio, i.e., signal vs. noise. In a sense, the signal represents the variations in the coefficients' surface, and the noise is the random fluctuation from observations. Hence, the S/N ratio can be interpreted as the ratio of standard deviations or variation levels, with larger S/N ratios associated with larger variations in the coefficient surface (i.e., the trend of the coefficients is more obvious). In particular, we assume that:

1. Signal $= 3 \times \left(\dfrac{\sum\limits_i (\beta_{ik} - \bar{\beta}_k)^2}{n-1} \right)^{\frac{1}{2}}$

2. Noise = Standard deviation of the error term. (Here it is 0.5.)

Intuitively, the coefficient estimates will be close to the true values if the noise is small (or the S/N ratio is large). The bias and variance of the estimates can be used together to evaluate the accuracy of the proposed CGWR. We can define:

15

$$\text{Average Discount rate:} \quad \frac{n^{-1}\sum_i\left(MSE_{method}\right)}{n^{-1}\sum_i\left(MSE_{OLS}\right)} \tag{13}$$

where MSE refers to the mean squares error, or the squared root of (bias)[2] plus variance. Note that (13) uses the MSE of the estimate from OLS as a bench mark. Furthermore, (13) can be computed for all locations and we can use the weighted mean of the discount rate for comparisons,

$$\frac{\sum_i\left[(MSE_{OLS})_{location\ i}\times\left(MSE_{method}/MSE_{OLS}\right)_{location\ i}\right]}{\sum_i(MSE_{OLS})_{location\ i}}. \tag{14}$$

Moreover, we can also define other evaluation indices:

$$\text{Average bandwidth:} \quad n^{-1}\sum_i\left(bandwidth_{method}\right) \tag{15}$$

$$\text{Average Variance:} \quad n^{-1}\sum_i\left(Var\left(\hat{\beta}_{ik}\right)\right) \tag{16}$$

$$\text{Average Bias:} \quad n^{-1}\sum_i\left(\left(\hat{\beta}_{ik}-\beta_{ik}\right)^2\right) \tag{17}$$

The above evaluation indices allow us to check the accuracy and precision of the estimators individually. In general, a good estimator should have small variance, bias and MSE.

4.1 Simulation Setting:

As mentioned above, there were four types of surfaces for the coefficients (Figure 1). Cases 1 and 2 were polynomial functions (close to linear functions) of the independent variables, and cases 3 and 4 were non-linear functions. Similar settings also appeared in previous studies on GWR, and they have practical meanings. For example, the quadratic surface (Case 2) often occurs in situations involving housing prices, where prices are significantly higher for locations near a town centre or transportation centre. For example, the relationship between the

16

size of a house and its price is more obvious in downtown areas than in the countryside. Non-linear surfaces can be used to determine the relationship between floor space and price for both industrial and commercial areas. For every non-stationary surface, we assume that there are 10 by 10 regular lattice points (100 locations).

For the fixed effect model, two major scenarios are tested under different S/N ratio. The first scenario is single type surfaces. In this scenario, both intercept and independent variables are composed by same surface, and all four types of aforementioned surfaces will be considered. The second scenario, i.e., mixed type surfaces, is more realistic and assumes that intercept and independent variables follow different surfaces. We examine the second scenario with two cases, linear-quadratic (polynomial surface) and ridge-hillside (non-polynomial surface).

To ensure the randomness of the experiment and prevent the nuisance effect caused by specific pattern, the independent variable at each location is drawn from a uniform distribution on the interval (0, 1) and is fixed for all 100 simulation runs. We consider various values for the independent variable and the results are very consistent. To simply the discussion, we will choose only one of the results as a demonstration. For both models, the errors are drawn from a normal distribution with a mean of 0 and a standard deviation of 0.5. Also, the reason for choosing the standard deviation 0.5 is to incorporate the values of the S/N ratio.

However, in the real scenario, the dependent variables are often random due to measurement errors. Moreover, performing the simulation under different dependent variable sets can make the results carry more conviction. Hence, we redo the single-type surfaces simulations under the random effect model, which

the dependent variable set is drawn repeatedly from uniform distribution on the interval $(0, 1)$ at each sample.
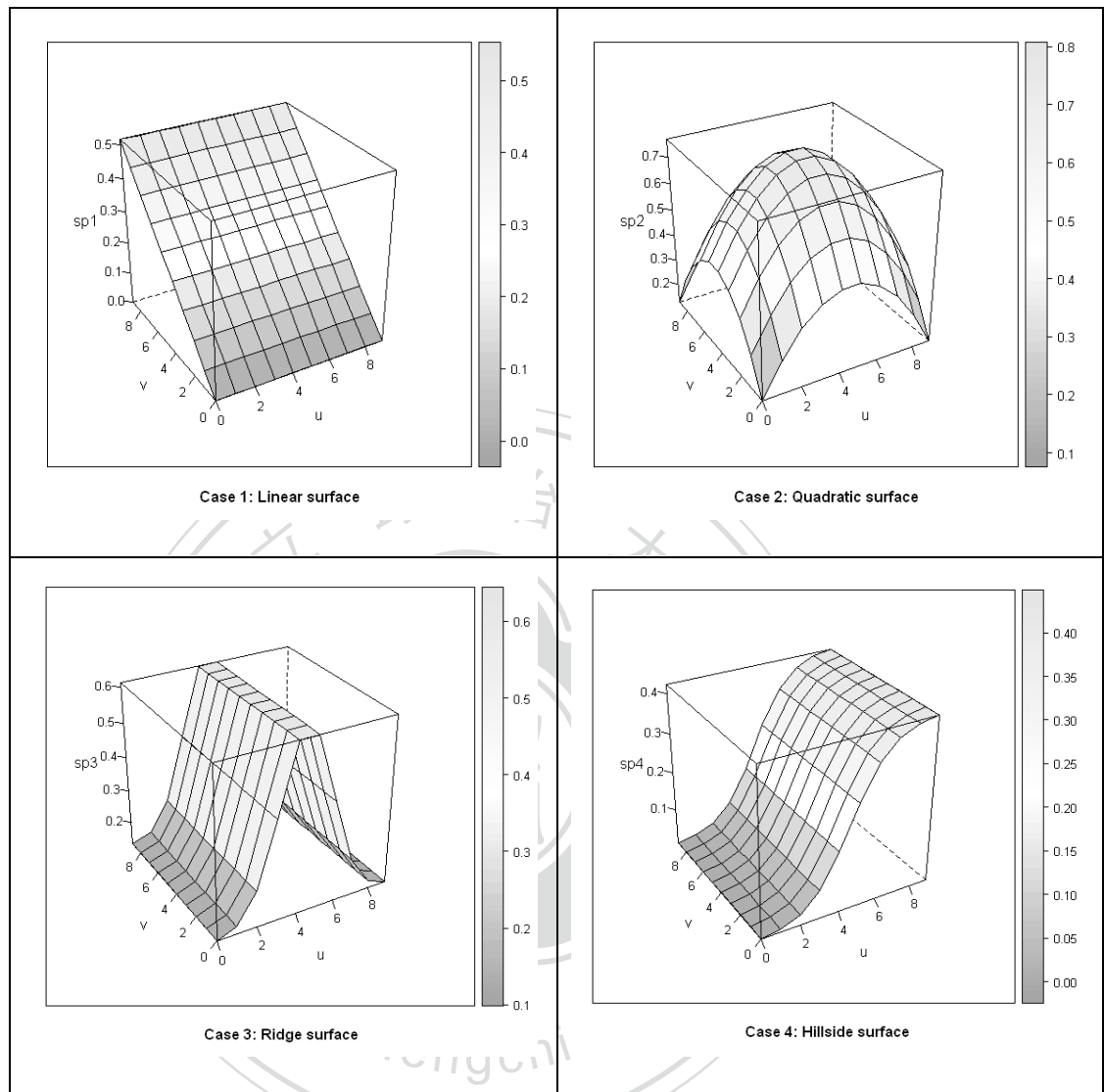
Figure 1: Four coefficient surfaces. Cases 1 & 2 are polynomial surfaces and Cases 3 & 4 are non-polynomial surfaces.

For the CGWR, the Gaussian kernel is chosen and the optimal bandwidth is the one with the minimum cross-validated sum of squares (CVSS). Furthermore, we require that the bandwidth is within a reasonable range, to prevent the estimation from being too localized or globalized, for extremely small or large bandwidths, respectively. The upper bound of the range is the maximum length on the map and the lower bound is to have at least 5 data points (each with 1/5 of the

weight). The preceding setting is also used in the "spgwr" package (Bivard and Yu, 2010, ver0.5-4) in R, a free statistical software.

The rest of this section discusses the simulation results for the single type surfaces, followed by those for the mixed type surfaces, and finally the random effect models. Specifically, we will compare the performances of three GWR estimations based on the average discount rate, the average bandwidth and the average variance and average bias of the estimates. In addition, the stopping criteria in following simulations are average absolute change rate (β_0 & β_1) less than 0.005, which is 0.01% of the noise. The settings is different from the usual settings because the convergence times to 10^{-3} is much greater than 0.005. Due to the mass simulations considered in this paper, we sacrifice a little precision in exchange of the simulation time and the results would be similar for smaller stopping criteria.

To simplify the notation, we will use β_0 and β_1 to denote the coefficients of the intercept and slope of independent variable x in the following tables, respectively.

4.2 Single type surfaces:

Table 1-1 shows the results of the discount rates in the case where B0 and B1 satisfy a linear surface. We can see that both the proposed CGWR and the local linear method have significant improvements over the basic CGWR. Interestingly, the local linear method is better (with respect to smaller discount rates) than the GWR when the S/N ratio is large, but the basic GWR is better when the S/N is small. The reason might be that larger noises produce larger fluctuations, and thus the average tangent line in the local linear method is not accurate or stable. Similar patterns also appear for the other three surfaces (Tables

1-2, 1-3, 1-4). This suggests that the local linear method is not very stable if the S/N ratio is small.

The CGWR and the local linear method again outperform the basic GWR in the case of a quadratic surface. However, the CGWR appears to be the best and the edge is more obvious when the S/N increased. For the nonlinear surfaces, the CGWR continues to work satisfactorily, but the local linear model does not. It is possible to create worse results than the basic GWR. The CGWR is still reliable in the cases of non-linear surfaces, performing much better than the other two methods.

Table 1-1. The average discount rates on a linear surface (single-type)

		B0			B1		
B1	B0	1	3	5	1	3	5
	1	1.064	0.312	0.158	1.815	1.589	1.328
Basic GWR	3	1.348	0.321	0.173	0.598	0.663	0.755
	5	1.775	0.382	0.199	0.329	0.397	0.459
	1	1.317	0.192	0.078	2.111	1.033	0.686
Local-Linear	3	1.107	0.164	0.069	0.403	0.347	0.284
	5	1.299	0.188	0.059	0.197	0.179	0.126
	1	0.664	0.195	0.082	1.076	0.775	0.412
CGWR	3	0.880	0.202	0.086	0.400	0.366	0.286
	5	0.899	0.227	0.092	0.198	0.207	0.178

Table 1-2. The average discount rates on a quadratic surface (single-type)

		B0			B1		
B1	B0	1	3	5	1	3	5
	1	1.955	0.791	0.510	2.877	7.729	7.542
Basic GWR	3	3.217	0.896	0.570	1.418	2.332	3.650
	5	4.334	1.108	0.639	0.844	1.316	1.819
	1	3.279	0.914	0.535	5.159	8.900	7.941
Local-Linear	3	5.177	1.119	0.599	2.039	2.634	3.481
	5	6.147	1.196	0.602	1.001	1.169	1.488
	1	1.168	0.324	0.146	1.529	1.612	0.928

| CGWR | 3 | 1.491 | 0.380 | 0.166 | 0.675 | 0.713 | 0.730 |
| | 5 | 1.709 | 0.429 | 0.183 | 0.374 | 0.451 | 0.430 |

Table 1-3. The average discount rates on a ridge surface (single-type)

		B0			**B1**		
	B0 B1	1	3	5	1	3	5
Basic GWR	1	1.740	0.571	0.383	2.642	2.325	2.249
	3	2.623	0.728	0.401	1.150	1.375	1.321
	5	2.892	0.824	0.433	0.591	0.810	0.894
Local-Linear	1	2.890	1.023	0.621	4.745	4.374	3.672
	3	4.340	1.206	0.618	1.960	2.255	2.023
	5	5.385	1.386	0.677	1.081	1.326	1.344
CGWR	1	0.884	0.203	0.111	1.029	0.428	0.258
	3	1.048	0.239	0.116	0.529	0.385	0.250
	5	0.854	0.221	0.107	0.253	0.208	0.159

Table 1-4. The average discount rates on a hillside surface (single-type)

		B0			**B1**		
	B0 B1	1	3	5	1	3	5
Basic GWR	1	1.075	0.315	0.153	1.697	1.905	1.359
	3	1.353	0.340	0.173	0.583	0.603	0.668
	5	1.152	0.343	0.185	0.234	0.278	0.360
Local-Linear	1	1.339	0.243	0.149	2.187	1.549	1.239
	3	1.403	0.289	0.188	0.617	0.529	0.722
	5	1.113	0.339	0.209	0.266	0.307	0.421
CGWR	1	0.726	0.170	0.077	1.098	0.761	0.473
	3	0.926	0.199	0.079	0.384	0.348	0.239
	5	0.578	0.192	0.079	0.159	0.158	0.140

To further evaluate the performances of the three different GWR methods, we can consider the hillside surface (Table 1-4) and the S/N ratio of B0 = B1 = 5 as a demonstration. We can see that the CGWR produces the best fit and the mean

surface looks almost identical to the true surface (Figures 1 and 2). The edge of the mean surface from the basic GWR is not smooth, which may be due to the fewer observations involved in the estimation, i.e., the edge effect. The mean surface of the local linear method does not look like the shape of a hillside at all, but looks like a linear surface.

| Method 1: Basic GWR | Method 2: Local liner estimation | Method 3: CGWR |
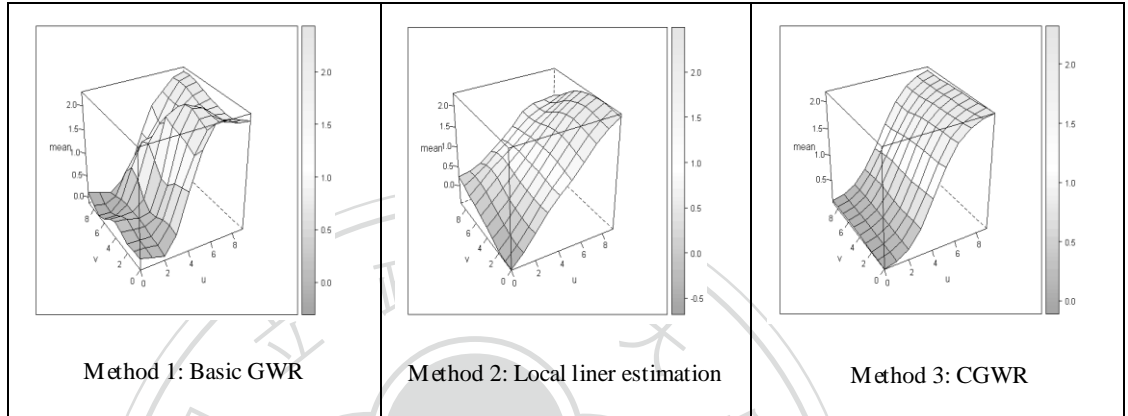
Figure 2: The B1 mean surface of each estimation method from 100 simulations. The simulation scenario is a hillside, and the S/N ratio of B0 = B1 = 5.

Table 2. The average bandwidths for linear and hillside surfaces (single-type) For CGWR, the first and second values are the average bandwidths of B0 and B1.

Method S/N	Linear surface			Hillside surface		
	Basic GWR	Local Linear	CGWR	Basic GWR	Local Linear	CGWR
1	2.35	9.73	4.16 ; 7.24	2.25	9.78	4.72 ; 7.45
3	1.46	10.8	1.65 ; 2.31	1.47	9.01	1.57 ; 2.80
5	1.2	10.34	1.25 ; 1.47	1.24	6.04	1.25 ; 1.66

Intuitively, we expect that the bandwidth is small if the S/N is large, since distant observations can be very different and cause biased estimation. Basically, all three GWR methods have significant drops in bandwidth if we increase the S/N ratio from 1 to 3. Moreover, the bandwidths in the case of a linear surface should be larger than those of a nonlinear surface under the same S/N ratio, since the surface change is quite homogenous in any direction.

The bandwidth results can also be used to explain why the CGWR outperforms the other two methods. We will choose two surfaces (linear and hillside) to discuss these results (Table 2). The local linear method often yields large optimal bandwidths. If the true surface is close to linear, we can rely on observations within a larger bandwidth and thus have smaller variances than those in the case of a nonlinear surface. Since the shape of a hillside is close to linear, the bandwidths of the hillside case are very similar to those in the linear case (Table 2) and are much larger than those of the quadratic and ridge cases (Appendix B).

Furthermore, the values of the CGWR bandwidth match the expected pattern. For example, if the S/N ratio is small, the bandwidth is expected to be large in order to provide a stable estimate. If we fix the S/N ratio of B1, the B0 bandwidth of CGWR decreases as the S/N ratio of B0 increases, for all four surfaces (Appendix B). Similar results hold for the B1 bandwidths if we fix the S/N ratio of B0. This is exactly why we want to consider a different bandwidth for each coefficient, since the numbers of observations needed can be different.

Table 3-1. The average variances and biases of B0 and B1 on a linear surface (single-type)

(i) B0

	OLS		Basic GWR		Local linear		CGWR	
S/N of B0:	1	5	1	5	1	5	1	5
Conditional on S/N of B1 = 1								
Variance	0.009	0.011	0.039	0.067	0.051	0.052	0.024	0.036
Bias	0.029	0.028	0.002	0.004	0.0003	0.0002	0.001	0.0002
Conditional on S/N of B1 = 5								
Variance	0.010	0.009	0.090	0.107	0.057	0.043	0.048	0.057
Bias	0.723	0.719	0.026	0.038	0.0005	0.0002	0.012	0.009

(ii) B1

	OLS		Basic GWR		Local linear		CGWR	
S/N of B0:	**1**	**5**	**1**	**5**	**1**	**5**	**1**	**5**
Conditional on S/N of B1 = 1								
Variance	0.031	0.036	0.113	0.190	0.146	0.145	0.071	0.117
Bias	0.038	0.695	0.013	0.051	0.001	0.0005	0.004	0.028
Conditional on S/N of B1 = 5								
Variance	0.030	0.027	0.248	0.294	0.152	0.114	0.088	0.152
Bias	0.191	0.880	0.048	0.124	0.001	0.0002	0.004	0.010

Table 3-2. The average variances and biases of B0 and B1 on a ridge surface (single-type)

(i) B0

	OLS		Basic GWR		Local linear		CGWR	
S/N of B0:	**1**	**5**	**1**	**5**	**1**	**5**	**1**	**5**
Conditional on S/N of B1 = 1								
Variance	0.011	0.010	0.068	0.152	0.118	0.304	0.038	0.044
Bias	0.035	0.049	0.013	0.020	0.011	0.017	0.003	0.006
Conditional on S/N of B1 = 5								
Variance	0.010	0.010	0.241	0.265	0.487	0.530	0.089	0.084
Bias	0.914	0.947	0.114	0.152	0.091	0.122	0.014	0.018

(ii) B1

	OLS		Basic GWR		Local linear		CGWR	
S/N of B0:	**1**	**5**	**1**	**5**	**1**	**5**	**1**	**5**
Conditional on S/N of B1 = 1								
Variance	0.026	0.033	0.180	0.433	0.333	0.843	0.067	0.188
Bias	0.049	0.923	0.022	0.136	0.030	0.198	0.011	0.055
Conditional on S/N of B1 = 5								
Variance	0.029	0.029	0.650	0.717	1.302	1.436	0.097	0.178
Bias	0.377	1.370	0.271	0.541	0.203	0.459	0.008	0.046

The variances and biases of the estimates from the three GWR methods can also be used for comparisons. Again, we will only use the cases of linear and

ridge surfaces. Because there are many combinations for the S/N ratios of B0 and B1, we will only show the results of S/N = 1 and 5 (Tables 3-1 to 3-2). Unlike in the previous comparisons, we will also provide the variances and biases of the OLS estimates. In general, a larger S/N ratio tends to produce a larger bias. Moreover, the OLS estimates fail to capture the spatial trend causing the largest bias, but it uses all of the observations in the estimation (i.e., infinity bandwidth) and thus has the smallest variance. As for the three GWR estimations, the variances of the estimates are generally larger than the biases of the estimates.

The results for a linear surface are shown in Table 3-1. As mentioned earlier, the average bandwidths of the local linear method were the largest, possibly indicating the smallest variances. In addition, the local linear method has the smallest bias, and also the smallest discount rates in the case of a linear surface (Table 1-1). Although the CGWR has a larger bias than the local linear method in the case of a linear surface, it dominates the basic GWR with respect to both variance and bias. The CGWR has the best performance in the case of a ridge surface, and it also outperforms both the basic GWR and the local linear method for both the variance and bias.

4.3 Mixed type surfaces:

We also repeat the same comparisons of the three GWR estimation methods with the mixed type surfaces. The results are similar to those in the single type surfaces, and thus we will only show the results of the discount rates. As we mentioned before, there are two cases in this scenario: linear-quadratic (polynomial surface) and ridge-hillside (non-polynomial surface). For the first case, the underlying surface of intercept is linear, and the slope is quadratic. For the second case, all surfaces are of non-polynomial type and it is more complex than the first one.

Table 4-1. The average discount rates of a linear-quadratic surface (mixed-type)

		B0			B1		
	B0 B1	1	3	5	1	3	5
	1	1.062	0.279	0.155	1.934	1.643	1.420
Basic GWR	3	2.110	0.407	0.181	1.048	1.067	0.935
	5	3.119	0.537	0.231	0.688	0.687	0.711
	1	1.492	0.225	0.080	2.888	1.514	0.905
Local-Linear	3	3.100	0.534	0.169	1.405	1.314	0.954
	5	5.357	0.775	0.268	0.928	0.871	0.753
	1	0.764	0.183	0.100	1.301	0.802	0.538
CGWR	3	1.132	0.345	0.179	0.708	0.849	0.785
	5	1.257	0.388	0.217	0.404	0.523	0.584

Table 4-2: The average discount rates of a ridge-hillside surface (mixed-type)

		B0			B1		
	B0 B1	1	3	5	1	3	5
	1	1.457	0.545	0.337	2.160	3.268	2.868
Basic GWR	3	1.623	0.582	0.349	0.622	1.373	1.786
	5	1.511	0.604	0.344	0.251	0.586	0.943
	1	2.105	0.853	0.536	3.312	5.602	4.755
Local-Linear	3	2.236	0.903	0.497	0.896	2.348	2.595
	5	1.999	0.980	0.540	0.333	1.076	1.522
	1	0.956	0.245	0.116	1.166	0.750	0.351
CGWR	3	1.279	0.361	0.151	0.544	0.677	0.443
	5	1.338	0.485	0.204	0.245	0.434	0.415

Basically, The CGWR also has smaller discount rates than the basic GWR for the mixed type surfaces (Tables 4-1 to 4-2). We will focus on the results that are different than those of the single type surfaces. Although the local linear estimation is better than the GWR in the linear-quadratic surfaces, it has adverse

26

performance in the ridge-hillside surfaces. The results might suggest the instability of local-linear approach. On the contrary, for both cases, the CGWR dominates other two methods, similar to that in the single type scenario.

From the computer simulations, we found that the proposed CGWR makes a significant improvement over the basic GWR. If the coefficients surfaces are non-linear, the CGWR also outperforms the local linear method. In the following section, we will use an empirical example to compare the CGWR and basic GWR, providing further evidence for supporting the CGWR.

4.4 Random effect model

We also repeated the same comparisons of the three GWR estimation methods for the random effect model. The results were similar to those for the fixed effect model, and thus we will only show the results of the discount rates of single-type surfaces.

Table 5-1. The average discount rates on a linear surface (random effect).

	B0 / B1	B0			B1		
		1	**3**	**5**	**1**	**3**	**5**
	1	0.903	0.264	0.134	1.583	1.693	1.493
Basic GWR	**3**	1.214	0.312	0.159	0.590	0.755	0.794
	5	1.497	0.353	0.169	0.310	0.379	0.426
	1	0.912	0.146	0.049	1.642	1.017	0.601
Local-Linear	**3**	0.914	0.143	0.053	0.384	0.338	0.271
	5	0.857	0.131	0.052	0.142	0.125	0.111
	1	0.608	0.154	0.067	0.989	0.747	0.355
CGWR	**3**	0.762	0.180	0.085	0.365	0.364	0.283
	5	0.770	0.168	0.079	0.180	0.164	0.164

Table5-2: The average discount rates on a quadratic surface (random effect).

		B0			B1		
	B0 B1	1	3	5	1	3	5
Basic GWR	1	1.500	0.618	0.375	2.316	4.047	1.745
	3	2.410	0.682	0.456	1.264	1.767	2.376
	5	3.056	0.847	0.519	0.745	1.020	1.379
Local-Linear	1	1.959	0.569	0.317	3.243	3.654	3.522
	3	2.706	0.621	0.338	1.265	1.497	1.562
	5	2.800	0.776	0.350	0.641	0.794	0.856
CGWR	1	1.021	0.300	0.136	1.336	0.969	0.644
	3	1.071	0.329	0.167	0.638	0.642	0.624
	5	1.230	0.355	0.159	0.371	0.383	0.363

Table 5-3. The average discount rates on a ridge surface (random effect).

		B0			B1		
	B0 B1	1	3	5	1	3	5
Basic GWR	1	1.508	0.506	0.305	2.862	3.497	3.807
	3	2.295	0.594	0.340	1.073	1.566	1.931
	5	2.481	0.666	0.426	0.586	0.824	1.186
Local-Linear	1	2.020	0.681	0.382	3.857	4.313	4.286
	3	2.664	0.778	0.407	1.291	1.861	1.968
	5	3.030	0.835	0.548	0.746	0.937	1.319
CGWR	1	0.820	0.227	0.123	1.273	0.883	0.667
	3	1.100	0.244	0.131	0.552	0.471	0.484
	5	1.047	0.254	0.124	0.291	0.275	0.244

Table 5-4: The average discount rates on a hillside surface (random effect).

		B0			B1		
	B1 \ B0	1	3	5	1	3	5
	1	0.969	0.250	0.134	1.624	1.862	1.526
Basic GWR	3	1.238	0.269	0.140	0.545	0.645	0.752
	5	1.356	0.328	0.152	0.307	0.373	0.411
	1	1.021	0.179	0.121	1.797	1.191	0.991
Local-Linear	3	1.202	0.227	0.123	0.535	0.468	0.489
	5	1.100	0.250	0.122	0.245	0.277	0.259
	1	0.620	0.149	0.081	0.913	0.761	0.551
CGWR	3	0.769	0.165	0.069	0.359	0.337	0.243
	5	0.747	0.130	0.070	0.189	0.147	0.153

The CGWR and local linear estimation also had smaller discount rates than the basic GWR for the random effect model (Tables 5-1 to 5-4). We will focus on the results that were different than those of the fixed effect model. Although the CGWR still dominated in the cases of ridge and hillside surfaces, the local linear estimation had better performances when the S/N ratio was larger. In general, for the random effect model, the local linear estimation was the best in the case of a linear surface, and the CGWR was the best for the other three cases, just as with the fixed effect model.

This section showed how computer simulations were used to demonstrate that the proposed CGWR is a significant improvement over the basic GWR. Except for the case of a linear surface, the CGWR also outperformed the local linear method. In the following section, we will show how an empirical study was used to compare the CGWR and basic GWR to provide further evidence to support the CGWR.

Chapter 5.

Empirical Study

We now apply the CGWR to a real data set. The data considered are from the 2000 Taiwan Population Census and the goal is to study the relationship between elderly disability and related social factors. Like in many developed countries, population aging is a serious problem and the elderly population in Taiwan has increased rapidly. For example, the proportion of people 65 years of age and over was 10.5% in the beginning of 2009, compared to 7% in 1994. However, most medical resources are concentrated in metropolitan areas or northern Taiwan, and this may not match the needs of the elderly. Hu and Yue (2002) studied this issue and found that the distribution of elderly disability is spatially autocorrelated and a classic spatial regression model can be applied. However, Brunsdon (1999) claimed that the phenomenon of spatial autocorrelation is likely to be caused by correlated random errors or spatial non-stationarity, (i.e., identifiability). His claim motivates us to re-examine the data using the GWR-based model.

The Taiwan census data are at the township level and there are 350 townships involved. The dependent variable is the proportion of disabled elderly people in each township. Since this variable appears to be right skewed, a log transformation (i.e., $y_i^* = \log(y_i + 1)$) is applied. Four independent variables are selected: population density (POP), proportion of elderly (ELD), higher-age mortality rate (HMR) and education level (EDU). These independent variables are standardized into the interval [0, 1] before being plugging into the GWR model. Before applying the GWR, we first test the spatial non-stationarity (Leung et al., 2000). The F-test of Leung et al. suggests that the model is spatial non-stationary

(p-value < 0.001). This confirms the conjecture of Brunsdon (1999) and we shall continue to proceed with the GWR-type analysis.

Table 6. The correlations of coefficients from the disability data. The variables INT and POP were in group 1, and ELD, HMR, & EDU were in group 2.

	INT	POP	ELD	HMR	EDU
INT	1.000	0.463	-0.949	-0.951	-0.971
POP	0.463	1.000	-0.281	-0.281	-0.653
ELD	-0.949	-0.281	1.000	0.914	0.875
HMR	-0.951	-0.281	0.914	1.000	0.868
EDU	-0.971	-0.653	0.875	0.868	1.000

Since the correlation of intercept and the variable POP is 0.463 (Table 6), they are put in the same group. Similarly, the variables ELD, HMR, and EDU are in the other group. Recall that the CGWR does not work well in the case of negative correlation. Thus, we focus on the variables with positive correlation and we choose to work on the group of intercept and variable POP. First, we treat variable ELD, HMR, EDU as constant, after getting estimates for them from the basic GWR. Then, we apply the CGWR to the intercept and the variable POP, as follows:

$$(y_i^* - \hat{\beta}_{i2}^{GWR} ELD_i - \hat{\beta}_{i3}^{GWR} HMR_i - \hat{\beta}_{i4}^{GWR} EDU_i) = \hat{\beta}_{i0}^{CGWR} + \hat{\beta}_{i1}^{CGWR} POP_i + r_i \qquad (18)$$

After fitting the CGWR, the calibrated bandwidths vary across the variables. Here, we let the lower and upper bounds of the bandwidth be (1 km, 400 km).

Through the surface after fitting with each method, we find that there is a dramatically different between CGWR and others (Figure 3). For the intercept term, the other methods review the North-South pattern. However, CGWR shows that high disability occurred on inland or mountain area. On the other hand, local

linear method seems to fit with a linear expansion surface, which is a strong assumption.

We use the pseudo R-square and residual plot to evaluate the performance of CGWR (Figure 4). The pseudo R-square is the Pearson product moment correlation coefficient of the fitted value and the observed value, and a larger value usually indicates a better fit. In this example, the pseudo R-square values of basic GWR and CGWR are 0.514 and 0.894, respectively. Moreover, the residual plots are also in favour of the CGWR, since there are fewer outliers and the outliers appear to have smaller variance.

It should be noted that either one of the variable sets can be chosen as constant. If we apply the CGWR procedure on the other variable set (i.e., ELD, HMR, and EDU) then the CGWR still has a better fitting result, although the pseudo R-square is 0.874, which is slightly lower.
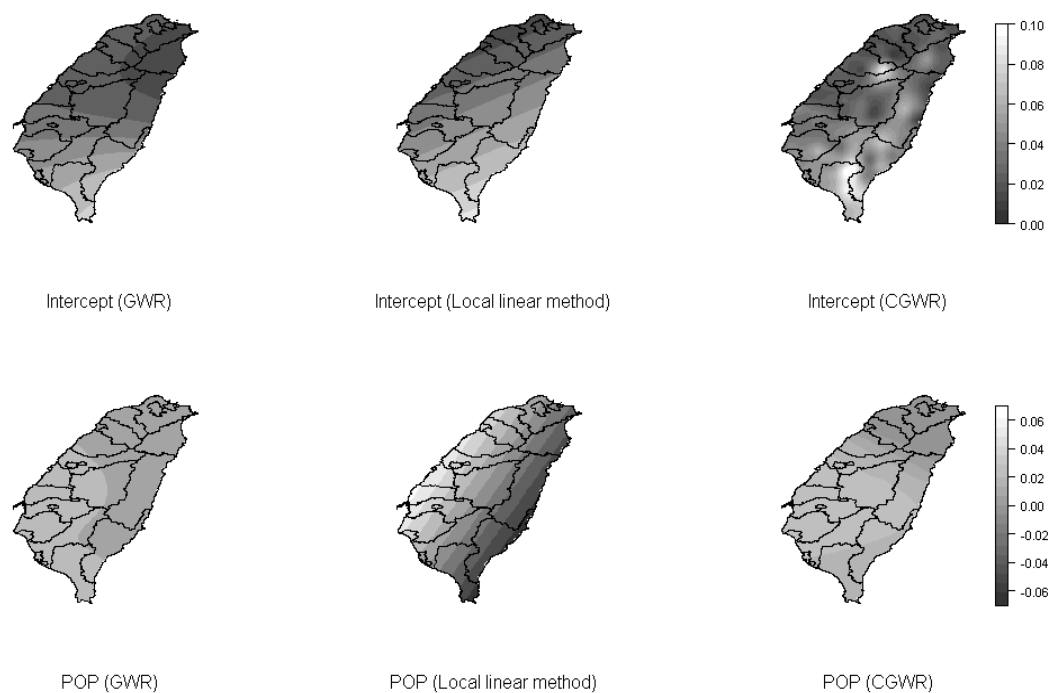

Figure 3: The surface of Intercept and variable POP from different fitting method.
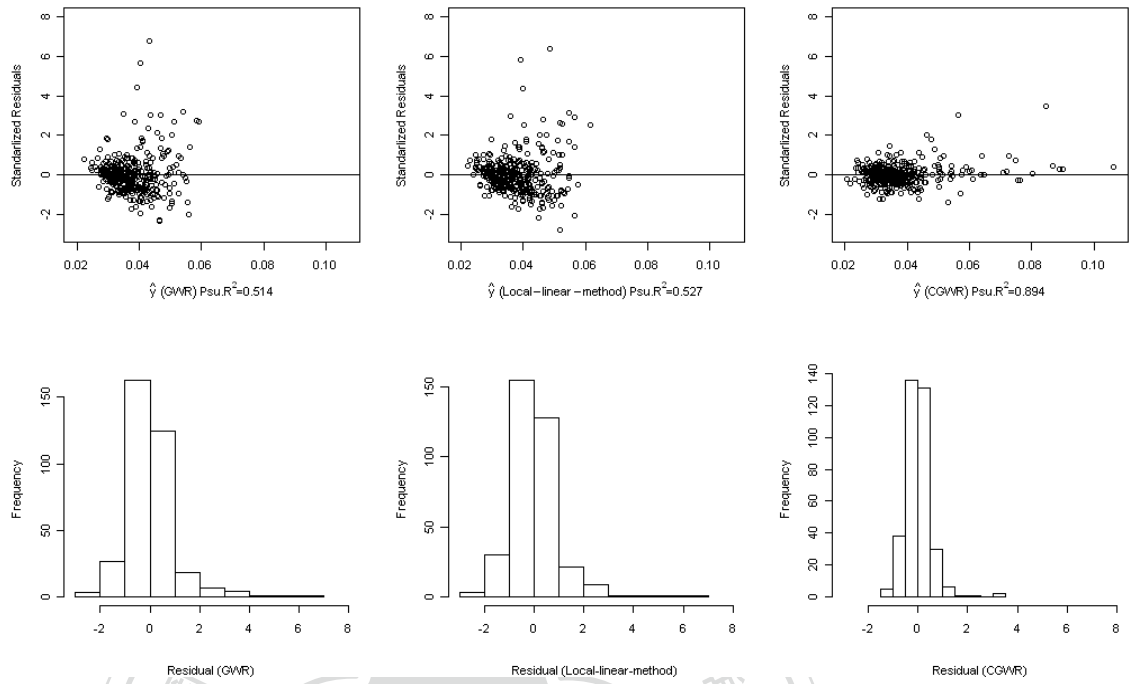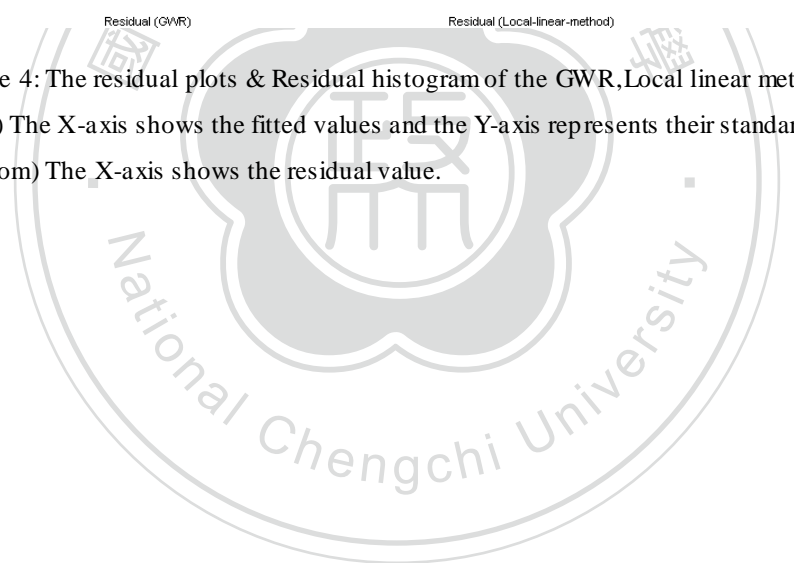
Figure 4: The residual plots & Residual histogram of the GWR, Local linear method and CGWR.
(Top) The X-axis shows the fitted values and the Y-axis represents their standardized residuals.
(Bottom) The X-axis shows the residual value.

Chapter 6.

Discussion and Concluding Remarks

Beginning from past decade, the GWR has acted as a new modelling technique used to deal with spatial non-stationarity. This technique allows regression coefficients to vary across space and obtains their estimates from a fixed bandwidth of observations. However, using a fixed bandwidth might not be appropriate since the independent variables would behave differently. In this paper, we proposed a method (CGWR) for modifying the GWR by relaxing the restriction of a fixed bandwidth. We compared the proposed method to the GWR and a local linear estimation, which was shown to be effective at reducing the bias. Based on simulation results, we found that the CGWR has the best performances given that the regression coefficients are positively correlated, and this advantage is especially noticeable in the cases of non-linear surfaces.

Of course, the improvements in the CGWR are due to the fact that it allows the bandwidth to vary for each coefficient. If the coefficients are positively correlated, the CGWR can reduce the bias, as well as the variance. In the empirical study, if the coefficients are not always positively correlated, the CGWR can still be modified and applied to a set of variables that are positively correlated for any pairs of variables.

However, the proposed method also has limitations. First, the most critical one is probably that the current settings for CGWR do not work well in the case of negative correlation, allowing the basic GWR to outperform the proposed CGWR. As a solution, we suggest calculating the correlation coefficients of the variables before applying the CGWR. Nevertheless, we can perform a two-stage fitting as we did at previous empirical study. In order to verify the feasibility of this approach, we conduct a experiment in two variables. That is, there are three

components in this model, which is intercept and two independent variables. Here, we redo the simulation by fixing second variable at first stage. By ratio of average discount rate, the outcome is similar to previous result. CGWR seems to fit well by this approach. Second, the CGWR is a computer intensive method, and would become extremely time-consuming if there are more variables. To increase the iteration speed, a moving average method could be used to increase the convergence speed. Third, we have not shown that the CGWR will converge if there are many variables, although we found that it would for cases up to four variables. A possible modification to a case with more variables would be to separate the variables into two groups and use double iteration. Then, CGWR could be used on each group of variables (inner loop), and the process re-iterated between the two groups (outer loop) until both groups of variables converge. To show that this idea is feasible, we conduct an experiment with six variables. The variables are separated into two groups of three variables each, and the estimation does converge.
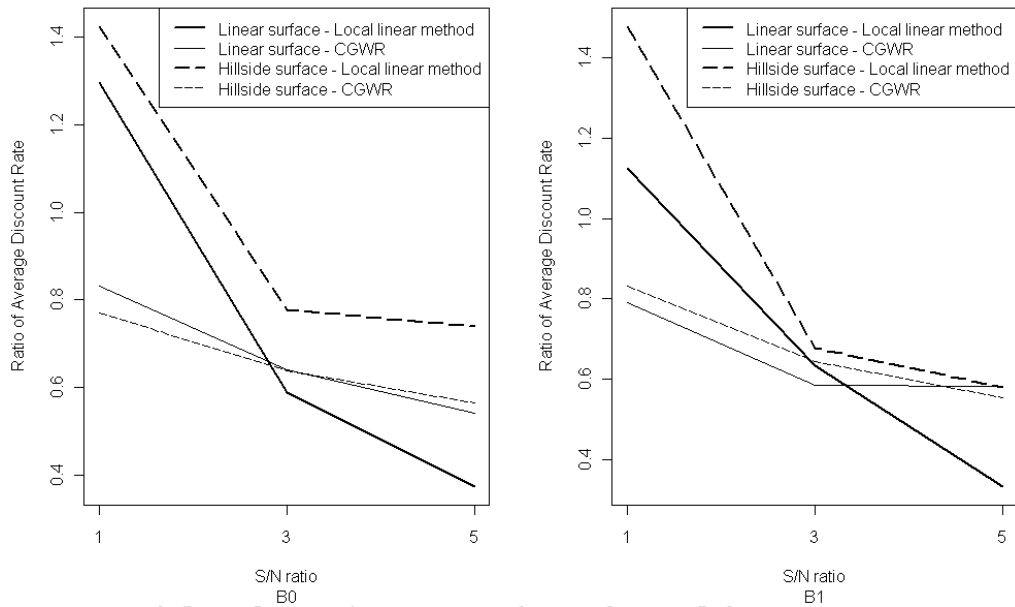
Figure 5: Average discount rate of each method by fixing variable x2, The baseline of the ratio of average discount rate is the GWR.

In this study, we also found some potential problems in applying the GWR. The GWR still has room for further improvement, especially when the S/N ratio is small, the surfaces of the coefficients are non-linear and the coefficients are very different. In addition, the variance reduction of CGWR over GWR is much more obvious than that for the bias reduction. This indicates that the GWR estimates have large variances. In other words, if the variances of the GWR estimates could be reduced, the bias could also be further reduced, and thus it is likely that the estimates would be more stable.

References

Bivand R, Danlin Y, 2010, R manual for package "spgwr",

    http://cran.r-project.org/web/packages/spgwr/spgwr.pdf (2010/08/01)

Bivand R, R. Brunstad R, 2005, Further Explorations of Interactions between

    Agricultural Policy and Regional Growth in Western Europe. Approaches to

    Nonstationarity in Spatial Econometrics,

    http://www.feweb.vu.nl/ersa2005/final_papers/671.pdf (2010/09/30)

Brunsdon C, Fotheringham A S, Charlton M, 1998, "Geographically weighted

    regression: modelling spatial nonstationarity" The Statistician 47, 431 – 443

Brunsdon C, Fotheringham A S, Charlton M, 1999, "Some notes on parametric

    significance tests for geographically weighted regression" Journal of Regional

    Science 39, 497 – 524

Brunsdon C, Fotheringham A S, Charlton M, 2002, Geographically Weighted

    Regression: the analysis of spatially varying relationships (John Wiley &

    Sons)

Cleveland W S, Grosse E, Shyu W M, 1991, " Local regression models"

    Statistical Models in S (Chambers, J. M. and Hastie,T. J., eds), 309–376.

    (Wadsworth & Brooks, Pacific Grove.)

Cressie N, 1993, Statistics for spatial data (John Wiley & Sons)

Fan J Q, Zhang W Y, 2008, " Statistical methods with varying coefficient models"

    Statistics and its interface 1, 179 – 195

Farber S, Páez A, 2007, "A systematic investigation of cross-validation in GWR

    model estimation: empirical analysis and Monte Carlo simulations" Journal of

    Geographical Systems 9, 371 – 396

Haining R, 2003, Spatial data analysis, Theory and practice (Cambridge

    University Press)

Hastie T J, Tibshirani R J, 1990, Generalized additive models, (Chapman and Hall, London, New York)

Hu Y W, Yue J C, 2002, "Spatial Analysis of Taiwan Elderly Disability" Technical Report, Department of Statistics, National Chengchi University, Taipei, Taiwan R.O.C.

Leung Y, Mei C L, Zhang W X, 2000, "Statistical tests for spatial nonstationarity based on the geographically weighted regression model" Environment and Planning A 32, 9 – 32

Manoranjan V S, Olmos G M, 1997, "A Two-Step Jacobi-Type Iterative Method" Computers Math. Application 34, 1, 1 – 9

Shi H J, Zhang L J, Liu J G, 2006, "A new spatial-attribute weighting function for geographically weighted regression" Canadian Journal of Forest Research 36, 4 996 – 1005

Wang N, Mei C L, Yan X D, 2008, "Local linear estimation of spatially varying coefficient models: an improvement on the geographically weighted regression technique" Environment and Planning A 40, 986 – 1005

Wu C O, Chiang C T, 2000, "Kernel smoothing on varying coefficient models with longitudinal dependent variable" Statistica Sinica 10, 433 – 456

Appendix A. The outline proof of the convergence for CGWR with fixed bandwidths

Consider a non–stationary model with only the intercept and a single explanatory variable (i.e., $y_i = f_{i1} + f_{i2} + \varepsilon_i$), where ε_i is random error.

To show the convergence of the CGWR algorithm, we first let the initial values be $\hat{f}_1(0) = \hat{f}_2(0) = \boldsymbol{0}$. Then the estimation of $\hat{f}_1 \, \& \, \hat{f}_2$ can be written as:

$$\hat{f}_1(k) = \mathbf{S}_1 \boldsymbol{Y} - \mathbf{S}_1 \mathbf{S}_2 \boldsymbol{Y} + \mathbf{S}_1 \mathbf{S}_2 \mathbf{S}_1 \boldsymbol{Y} - \mathbf{S}_1 \mathbf{S}_2 \mathbf{S}_1 \mathbf{S}_2 \boldsymbol{Y} + ... + \boldsymbol{R}_1$$
$$\hat{f}_2(k) = \mathbf{S}_2 \boldsymbol{Y} - \mathbf{S}_2 \mathbf{S}_1 \boldsymbol{Y} + \mathbf{S}_2 \mathbf{S}_1 \mathbf{S}_2 \boldsymbol{Y} - \mathbf{S}_2 \mathbf{S}_1 \mathbf{S}_2 \mathbf{S}_1 \boldsymbol{Y} + ... + \boldsymbol{R}_2$$

where $\boldsymbol{R}_1 \, \& \, \boldsymbol{R}_2$ are the remainder terms of the estimators:

To prove the convergence of the CGWR algorithm, we need to show that both R_1 and R_2 will converge to 0. Since the discussions of R_1 and R_2 are similar, we will only show the case for R_1. The case for $\boldsymbol{R}_1$ can be separated into two parts:

Let $(\mathbf{S}_1 \mathbf{S}_2)^{k^*} = \mathbf{P}^{-1} \mathbf{B}^{k^*} \mathbf{P}$, where $\mathbf{P}$ is a matrix of eigenvectors for $\mathbf{S}_1 \mathbf{S}_2$ at its columns, $\mathbf{B}^{k^*}$ is a diagonal matrix with eigenvalues for $\mathbf{S}_1 \mathbf{S}_2$ on the diagonal, and k^* is an arbitrary natural number.

If the largest eigenvalue in the absolute values of $\mathbf{S}_1 \mathbf{S}_2$, $|\lambda|$, is smaller than 1, then the remainder $\boldsymbol{R}_1$ converges to 0 as $k \to \infty$. □

The smoothing matrices in the proof are similar to those used in the GAM. According to Hastie and Tishibirani (1990), the smoother matrices should satisfy the bounded condition (p. 121). Moreover, this assumption seems to be true in practice and it is always the case in our simulation and empirical studies.

Appendix B. Detailed results for bandwidths (fixed effect models, single-type)

Note that the CGWR has two bandwidths for two coefficients, B0 and B1, respectively.

	B0 / B1	Linear surface			Quadratic surface		
		1	**3**	**5**	**1**	**3**	**5**
GWR	**1**	2.35	1.57	1.35	4.03	1.13	0.98
	3	1.75	1.46	1.26	1.26	1.03	0.94
	5	1.56	1.31	1.20	1.09	0.98	0.92
Local-Linear	**1**	9.73	9.56	10.26	2.63	1.58	1.22
	3	10.56	10.8	10.06	1.71	1.33	1.13
	5	10.01	9.82	10.34	1.50	1.21	1.06
CGWR	**1**	4.16 ; 7.24	1.60 ; 4.82	1.23 ; 5.19	6.58 ; 7.77	0.87 ; 8.27	0.66 ; 6.28
	3	4.23 ; 2.70	1.65 ; 2.31	1.18 ; 2.28	6.02 ; 3.69	0.82 ; 3.82	0.61 ; 2.10
	5	3.50 ; 1.70	1.60 ; 1.65	1.25 ; 1.47	4.77 ; 1.44	0.89 ; 1.35	0.61 ; 0.88
		Ridge surface			Hillside surface		
GWR	**1**	4.2	1.01	0.83	2.25	1.67	1.41
	3	1.16	0.89	0.77	1.79	1.47	1.31
	5	1.01	0.80	0.73	1.63	1.43	1.24
Local-Linear	**1**	2.61	1.33	1.03	9.78	9.78	8.2
	3	1.68	1.12	0.95	9.75	9.01	7.4
	5	1.28	1.01	0.91	8.67	8.93	6.04
CGWR	**1**	4.79 ; 8.79	0.85 ; 8.69	0.65 ; 5.45	4.72 ; 7.45	1.63 ; 5.83	1.25 ; 5.34
	3	5.24 ; 4.07	0.83 ; 2.15	0.66 ; 1.12	4.98 ; 3.11	1.57 ; 2.80	1.24 ; 2.70
	5	3.64 ; 1.16	0.80 ; 0.99	0.65 ; 0.79	5.09 ; 1.67	1.68 ; 1.64	1.25 ; 1.66