APPLICATION OF GEOGRAPHICALLY WEIGHTED REGRESSION FOR ASSESSING SPATIAL NON-STATIONARITY

KHOKAN CHANDRA SIKDAR

Your file Votre référence
ISBN: 0-612-93059-9
Our file Notre référence
ISBN: 0-612-93059-9

Application of Geographically Weighted Regression for Assessing Spatial Non-stationarity

by

©Khokan Chandra Sikdar

A practicum submitted to the School of Graduate Studies
in partial fulfillment of the requirement for the Degree of
Master of Applied Statistics

Department of Mathematics and Statistics
Memorial University of Newfoundland

August, 2003

St. John's **Newfoundland and Labrador** **Canada**

Abstract

Linear regression is a commonly used method of statistical analysis. However, it is not able to capture any spatial variations that may exist in the relationship between explanatory and response variables. We will study geographically weighted regression, which is a local regression method that can account for spatial non-stationarity that may exist. We will describe the model, estimation and hypothesis testing, both in theory and in simulation studies. We will also apply the method to analyze data collected on housing prices in the Boston metropolitan area.

Acknowledgements

I owe my profound gratitude to my supervisor, Dr. Gary Sneddon. His insight, dedication and continuous guidance have made it possible for me to complete this endeavor. He has been very generous with this idea and given me a great privilege to work on this interesting problem of spatial non-stationarity in data analysis. I would like to thank him for his suggestion of working with this area.

I sincerely acknowledge the financial support provided by the School of Graduate Studies and Department of Mathematics and Statistics in the form of Graduate Fellowships and Teaching Assistantships. I also would like to thank Professors Herb Gaskill and Bruce Watson, the past and present Department Heads, for providing me with a friendly atmosphere and necessary facilities to complete the program.

I would also like to thank Dr. Veeresh Gadag and Dr. Paul Peng for reviewing my practicum and providing helpful comments and criticisms.

I am specially grateful to my parents for their eternal love and emotional support. Although they never really understood any of the statistical matters I studied, they always encouraged me to do my best.

Finally, I would like to express my sincere appreciation and thanks to Mr. Gopal Chowhan, Subrata K. Chakrabarty, Masud A. Khan, Tapon K. Bhandari, Brian Healey and Ms. Roxana Vernescu whose invaluable support and encouragement have sustained me during the tough times. Throughout the course of my study, they have helped me in several ways, for which I am grateful.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Relationship between Variables

Applied statistics is a discipline where learning from data is one of the most relevant and vital challenges. In many cases, the aim is to study the relationships among measurable variables, where one is interested in assessing if a change in one (or more) variable is associated with a change in another variable of interest. This relationship can be of two types; one is a functional relationship and the other is a statistical relationship (Neter et al 1985, p. 23-24). A functional relationship between two variables is expressed by a mathematical formula. If X is the independent variable and Y is the dependent variable, a functional relationship can be written in the form:

$$Y = f(X)$$

Given a particular value of X, the function f indicates the corresponding exact value of Y, which is the characteristic of all functional relations. On the other hand, a statistical relation, unlike a functional relation, is not perfect one. To study the relationship among variables one of the important statistical methods is regression analysis. In the underlying logic of regression analysis, one variable takes on the role of a response (or dependent) variable, while all others are viewed as explanatory, predictor or independent variables. By a statistical relationship, it is meant that the

observed values of the response variable in a regression model are generated by a probability distribution that is a function of other variables. To demonstrate this, suppose we have a set of observations $\{x_i\}$, $i = 1, 2, \ldots, n$, of an explanatory variable x, and $\{y_i\}$ of a dependent variable y. Then the usual simple regression model can be written as

$$y_i = b_0 + b_1 x_i + \epsilon_i \tag{1.1}$$

where ϵ_i is the error term.

In the study of the regression model (1.1), the explanatory variable x is used to explain how the response variable y varies if the values of the explanatory variable are changed. Regression analysis is used to estimate the quantitative functional relationships between dependent variables and one or more independent variables from the actual data, when the relationship among the variables is statistical in nature rather than exact. Regression analysis is widely used in many fields of research. The goal of regression analysis is to estimate the parameter values for a function that cause the function to best (in a least squares sense) fit a set of observations that are available.

1.2 Regression and Spatial Data

In practice, the study of regression models consists of more than one predictor, and hence the analysis is called the study of multiple regression. We are often interested in examining more than one predictor of our response variable, and to determine whether the inclusion of additional predictor variables leads to increase prediction of the outcome variable. A common feature of this procedure is that it is applied globally, that is, to the complete region under study. However, it is often desirable to examine the relationship at a more local scale. For example, in studying the relationship between house price and population density in a country, the relationship between the two variables may differ, depending on the geographical location within the country.

When data has been collected over a geographic region, there are often two issues for which we need to account. One is spatial dependency, which is when observations that are close in space exhibit spatial autocorrelation. This has been studied within a regression framework by Odland (1988) and Anselin (1993).The second is spatial non-stationarity, as discussed in detail by Bailey and Gatrell (1995). This indicates the variation in relationships over space. That is, the parameter values change from region to region, and hence the effect of the corresponding explanatory variable is not same over the whole area under study. It has been recognized that failure to take necessary steps to account for or ignore spatial autocorrelation can lead to serious errors in the model interpretation (Anselin and Griffith, 1988; Arbia, 1989). Therefore, in regression modeling, it is necessary to determine whether or not an identifiable spatial pattern exists in the data set. Getis and Ord (1992) suggests that spatial modeling should account for not only the dependence structure and spatial heteroskedasticity but also assess the effects of several predictors on a spatial scale.

There are several reasons why parameter estimates from a regression model might exhibit spatial variation. For instance, if a regression model is fitted to predict the price of houses, it might be usual that the value of an extra room may not be same in several towns. Similarly, if a particular type of illness (Fotheringham et al, 1996) is considered to be affected by the socio-economic or socio-cultural practices of the communities, the effect of certain predictor variables on the illness may vary from place to place.

1.3 Measuring Spatial Patterns

There are many ways to test for the existence of a spatial pattern in a data set. For example, we may test for such patterns by focusing on the locations of the sample points, by studying the values associated with these locations given the sampling pattern, or by combining these analyses. In many geographical analyses, the identification of spatial autocorrelation is performed through applying Moran's I statistic

(Besag and Newell, 1991; Getis and Ord, 1992). For the study of local patterns in spatial data, Getis and Ord (1992) introduced the G statistic and presented a comparative advantage between the G and I statistics with respect to the spatial pattern in a data set. These general tests are concerned with the overall pattern in a large study region, whereas a focused test concentrates upon one or more smaller regions selected because of some factors that have been previously hypothesized to be associated with the response variable. Besag and Newell (1991) discussed a focused test procedure, and pointed out some difficulties for the I statistic.

In linear regression analysis, the data may be drawn from geographical units and a single regression equation is estimated. In general, the ordinary least squares (OLS) technique is used to produce the global estimates of parameters which are considered to apply equally over the whole region. That is, the relationships being measured are assumed to be stationary over space. However, relationships which exhibit spatial non-stationarity create problems for the interpretation of the OLS estimates of parameters from the regression model. Naturally, it is of interest to combine these ideas: regression modeling that attempts to allow us to describe spatial non-stationarity in data. Along this line of thinking, we will discuss the technique of geographically weighted regression (GWR), in which the coefficients of a linear regression model are estimated by a weighted least squares procedure. The location in geographical space is used to produce the weight function and, therefore, GWR allows us to obtain the local estimates, rather than global, of the parameters in the regression model. Brunsdon et al (1998, 1999), for instance, suggested this method for analyzing a spatially autoregressive model.

It is our interest in this practicum to study the technique of GWR including its underlying theory, estimation and inference, and practical application. In Chapter 2, we discuss the theoretical aspects of GWR including several weight functions, bandwidth selection, estimation and testing procedures. The results of simulation studies are described in Chapter 3, where we concentrate on finding the GWR estimates of parameters, and to determine the power and size of the tests. With the purpose of

observing the performance of tests, we apply testing methods to data simulated using a single explanatory variable in the model, and with three explanatory variables. In Chapter 4, we choose a widely used socio-economic data set on Boston house prices for application of the GWR methods. We will compare these results to those found using the typical linear regression model. We will also use some model selection procedures to help determine a smaller number of explanatory variables to use in the GWR procedure. We will give our conclusions, and thoughts on possible future work, in Chapter 5.

Chapter 2

Theory of GWR

2.1 Introduction

As mentioned in Chapter 1, geographically weighted regression (GWR) is an alternate method of estimation that can incorporate the spatial non-stationarity in relationships over space. In this chapter we will study the theoretical aspects of GWR, focusing on parameter estimation and hypothesis testing, but the choices of weight function and bandwidth are also necessary parts of the methodological development. Since the spatial non-stationarity in relationships is the key issue, we will present in detail two statistical procedures to test for spatial variation in the parameter values of the GWR model.

2.2 Model for GWR

In spatial analysis the data are often assumed to be non-stationary over space. Geographically weighted regression is one of the statistical techniques through which the presence of spatial non-stationarity is examined. The statistical model of global regression can be written as

$$y_i = b_0 + \sum_{j=1}^{k} b_j x_{ij} + \epsilon_i \tag{2.1}$$

where y_i represents the i^{th} $(i = 1, 2, \ldots, n)$ response related to the j^{th} $(j = 1, 2, \ldots, k)$ predictor x_{ij}. The corresponding regression coefficient in (2.1) is b_j and an uncontrolled random error is ϵ_i.

GWR extends the usual regression framework of equation (2.1) that allows local rather than global parameters to be estimated (Fotheringham et al, 1998). Therefore, the model for GWR can be written as

$$y_i = b_0(u_i, v_i) + \sum_{j=1}^{k} b_j(u_i, v_i) x_{ij} + \epsilon_i \tag{2.2}$$

where (u_i, v_i) indicates the coordinates of the i^{th} point on the surface. If the entire study area is considered as a continuous surface of parameter values and the spatial variability of the surface is obtained through measurements of this surface at certain points, then $b_j(u_i, v_i)$ indicates the realization of the continuous function of $b_j(u, v)$ at point i. That is, in the analysis of spatial data, the parameters are assumed to be function of the locations at which the observations are obtained. Obviously, equation (2.1) is a special case of equation (2.2), where the parameter values are considered to be constant over space. Thus, the equation (2.2) can be approximated by the equation (2.1) considering the i^{th} region on the surface. When estimating a parameter for a given point i, an ordinary least squares (OLS) regression can be performed with a subset of the points in the data set that are close to i. Accordingly, an estimate of $b_j(u_i, v_i)$ is obtained for region i in the usual way, whereas for the next i, a new subset of nearby points is used, and so on. Thus, equation (2.2) is a recognition of the GWR expression through which one attempts to assess whether spatial variations in relationships exist (Fotheringham et al, 1998).

2.3 Estimation

The regression model in (2.2) leads to a probabilistic model for a given region, specified by i, on the surface. Specifying such a model for several points of the study area causes problems associated with estimating coefficients, and hence, model fitting. Unlike the OLS regression model in equation (2.1), this model (2.2) allows the parameters to vary in space. However, the model (2.2) consists of more unknown parameters than observations, and hence, being in unconstrained form it is not implementable directly. This is related to the notion of underdetermined regression models (Sneddon 1999). Hastie and Tibshirani (1990) have carried out work with these type of models. Also, the estimate of $b_j(u_i, v_i)$ for the i^{th} point involves some degree of bias since the coefficients of equation (2.2) recognize local behaviour rather than global. However, if the sample size is large enough for a specified location, the corresponding standard error of the parameter estimates will reduce. That is, the larger the local sample, the smaller the standard error of the estimates. Hence, the sample size of the local subset plays a key role in the estimation process of (2.2). The sample size works as a compromizing factor of increasing bias and decreasing standard error of the estimates.

2.3.1 Bias-Variance Compromise: Prediction Error Approach

The idea of a bias-variance compromise is discussed in many works where sampling is one of the vital platforms for research. Fotheringham et al (1998) present an extensive explanation using a diagram similar to Figure 2.1.

Considering the context of GWR, if $\mathbf{X}_i$ represents a set of predictors in location i on the surface, and $\hat{\mathbf{b}}$ is a set of coefficient estimators, then $\hat{\mathbf{y}}_i = \mathbf{X}_i^T\hat{\mathbf{b}}$ is an estimate of the response $\mathbf{y}$ at that location. Due to the random nature of $\mathbf{y}_i$, the estimator $\hat{\mathbf{b}}$, and hence, $\hat{\mathbf{y}}$ are random. Therefore, $\hat{\mathbf{y}}$ can be observed through its distributional pattern, which is characterized by its expected value $E(\hat{y})$ and standard deviation $SD(\hat{y})$. When for all $\mathbf{X}$, $E(\hat{y}) = E(y)$, the estimator is said to be unbiased. If for an estimator $\hat{\mathbf{y}}$, once unbiasedness holds, the lower the $SD(\hat{y})$ values the more efficient

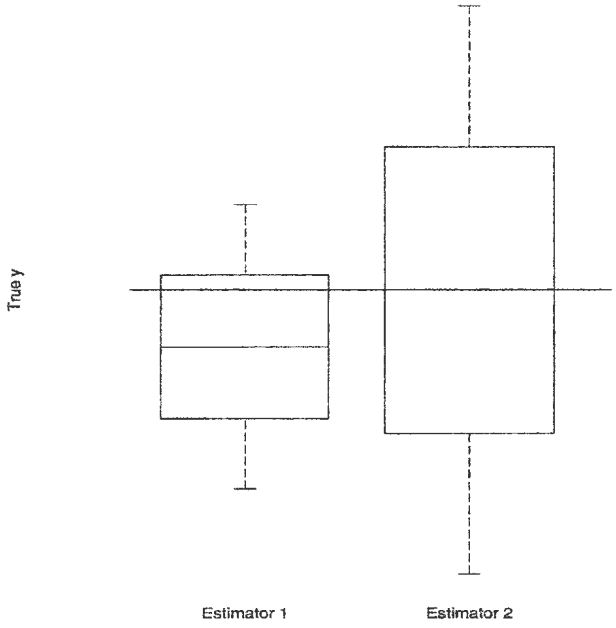

Figure 2.1: Prediction error distribution of y-estimators. Estimator 1 is $\hat{y}_1$, estimator 2 is $\hat{y}_2$.

$\hat{\mathbf{y}}$ is. Figure 2.1 presents an interpretation of two different estimators of y, say $\hat{\mathbf{y}}_1$ and $\hat{\mathbf{y}}_2$ respectively, in terms of bias and variance. The probability distributions of $\hat{\mathbf{y}}_1$ and $\hat{\mathbf{y}}_2$ are visualized by two boxplots. Considering the horizontal line for true y, it is obvious that $\hat{\mathbf{y}}_2$ is unbiased. Although $\hat{\mathbf{y}}_1$ is a biased estimator, its overall variability is less than that of $\hat{\mathbf{y}}_2$. Therefore, the estimator $\hat{\mathbf{y}}_1$ is presenting less prediction error of y even though $\hat{\mathbf{y}}_2$ is more advantageous as its distribution is centered at y. Due to the longer tails of the prediction squared error (PSE) of $\hat{\mathbf{y}}_2$, one may choose $\hat{\mathbf{y}}_1$ even if it is biased.

However, introducing more bias caused by a large sample approximation still seems to be a drawback of the estimation method. To reduce this effect, another adjustment is possible to consider. A weighted OLS estimation can be used so that it provides a means of computing localized regression estimates. This technique works well if the points further from region i are more likely to have coefficients differing from those closer to region i. If the estimation is performed through applying a monotone weighting function, then observations further from the point i, at which the parameter as well as the model is being estimated, receive less weight that observations closer to point i. Thus, estimation of equation (2.2) measures the relationship inherent in the model around each point i.

According to regression theory, the OLS estimate of coefficients in model (2.1), if written in matrix form, is given by

$$\hat{\mathbf{b}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}$$

where $\hat{\mathbf{b}}$ represents an estimate of $\mathbf{b}$, whereas $\mathbf{X}$ contains values of the predictor variable with $1's$ in the first column and $\mathbf{y}$ contains values of the response variable. However, if we do not want to place the same emphasis on each observation, a similar estimate for model (2.2) will be of the form

$$\hat{\mathbf{b}}(u_i, v_i) = [\mathbf{X}^T\mathbf{W}(u_i, v_i)\mathbf{X}]^{-1}\mathbf{X}^T\mathbf{W}(u_i, v_i)\mathbf{y}$$

where $\mathbf{W}(u_i, v_i)$ is an $n \times n$ weight matrix whose off-diagonal elements are zero. The diagonal elements of $\mathbf{W}(u_i, v_i)$ indicate weights for observations corresponding

to point i in the study area. The role of the weight matrix $\mathbf{W}(u_i, v_i)$ is to place different emphases on different observations to obtain parameter estimates. Hence, introduction of this geographical weight matrix $\mathbf{W}(u_i, v_i)$ leads to the estimation in such a way that the observed data near location i are weighted more than the observed data farther away. It is very important to choose an appropriate weight function to obtain a good estimate $\hat{\mathbf{b}}(u_i, v_i)$. The choice of weighting matrix is discussed in the next subsection.

2.3.2 Spatial Weighting Function

As noted above, even if a bias-variance balance is possible to meet through selecting a reasonably large sample, there is still an indication of increasing bias. This can be controlled when an appropriate weighting function is used for estimation. The choice of weighting function is one of the vital issues to estimate coefficients and later on to investigate spatial variability.

A simple but natural choice of weighting function at a specific location is to exclude those observations that are farther than some pre-specified distance. If we let w_{ik} be the $(i, j)^{th}$ element of W, this kind of weighting function is called a *binary weighting function*, and can be defined by

$$w_{ik} = \begin{cases} 1 & \text{if } d_{ik} < r \\ 0 & \text{otherwise} \end{cases} \tag{2.3}$$

where d_{ik} represents the distance between the i^{th} and k^{th} locations on the surface. To explain the weight function in equation (2.3): if i represents any point on the surface at which parameters are estimated, and k represents a specified point in space at which data are observed, then observations that are within some distance r from the locality i have a weight of unity, and observations whose distance exceeds this quantity r have weight zero. In the global model, where no spatial variation is considered, each observation has a weight of unity.

The binary weight function is a step function, which suffers from the problem of

discontinuity over the study area. This leads to a very sudden change of the spatial association between variables. One way to overcome this problem is to introduce a continuous weight function. One such function is the exponential weighting function given by

$$w_{ik} = \exp(-d_{ik}^2/2\beta^2) \tag{2.4}$$

The function (2.4) is called a Gaussian distance-decay-based weighting function. This is a continuous and monotone decreasing function of d_{ik}, because the larger the distance d_{ik}, the smaller the value of the weight. The weight would decay gradually with distance. More precisely, if i represents a point at which an observation was made, the weight assigned to that observation will be unity and the weights of the other points will decrease according to a Gaussian curve as d_{ij} increases. In the GWR estimation process, another weight function of the form

$$w_{ik} = \exp(-d_{ik}/\beta)$$

is also used. This is an alternative but very similar weighting function to equation (2.4).

Rather than considering an exponential form, another continuous function, which is known as a kernel function, is often used. The form of this function is

$$w_{ik} = \begin{cases} [1 - (d_{ik}/\beta)^2]^2 & \text{if } d_{ik} < r, \\ 0 & \text{otherwise} \end{cases} \tag{2.5}$$

The kernel function is denoted by K; that is,

$$w_{ik} = K(d_{ik})$$

The usual features of a kernel function K are (Brunsdon et al, 1998):

(i) $K(0) = 1$

(ii) $\lim_{d\to\infty} K(d) = 0$

(iii) K is a monotone decreasing function for positive real numbers.

The weighting function in equation (2.5) indicates setting the weights to zero outside a distance r and to decrease monotonically to zero with r as d_{ik} increases. Therefore, the kernel function (2.5) is a compromise between the weight functions of the binary (2.3) and exponential (2.4) forms.

2.3.3 Choice of β

Introducing a weighting function, the estimate of coefficients, when observations are corresponding to location i, can be written as

$$\hat{\mathbf{b}_i} = (\mathbf{X}^T\mathbf{W}_i\mathbf{X})^{-1}\mathbf{X}^T\mathbf{W}_i\mathbf{y} \tag{2.6}$$

where

$$\mathbf{W}_i = \begin{pmatrix} w_{i1} & 0 & \cdots & 0 \\ 0 & w_{i2} & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & w_{in} \end{pmatrix}$$

Here, the weight matrix $\mathbf{W}_i$ is a $n \times n$ diagonal matrix which consists of non-zero diagonal elements to indicate the weights for estimating $\mathbf{b}_i$ around region i in space. In fact, this matrix plays a key role in estimation in the GWR model. Once each w_{ik} has been computed, the $\hat{\mathbf{b}_i}$ vector can be computed through repeated application of expression (2.6) for each i. As noted previously, the weighting functions of continuous type are preferable for analyzing spatial data, since the degree of weighting changes with distance rather than suddenly dropping to zero. These functions include a constant β, which is often called the kernel bandwidth.

The bandwidth β is a non-negative constant depicting the way the Gaussian or kernel weights very with distance. For a given d_{ik}, the smaller the β, the less emphasis placed on the observation at location k. Accordingly, an important weighting note

is struck here—choice of an appropriate bandwidth value has more influence on estimation than the choice of weighting function (Simonoff 1996, p. 44). Our interest in this section is to emphasize how to choose a reasonable β value. In some cases, there is no theoretical basis of how to choose the value of β, although the properties of $\hat{\mathbf{b}}_i$ is greatly affected by the choice of β. Silverman (1986) suggests a subjective choice of β if no prior idea is available. The method of mean squared error and cross-validated sum of squared errors are used in this practicum to obtain the best possible β for every individual data set. The mean squared error of the estimate of coefficient b is defined by

$$\begin{aligned} MSE(\hat{b}) &= E(\hat{b} - b)^2 \\ &= Var(\hat{b}) + [Bias(\hat{b})]^2 \end{aligned} \tag{2.7}$$

Therefore, the MSE of an estimator can be decomposed into its variance and squared bias. To compare estimators by looking at their respective mean squared errors, naturally we would prefer one with smallest MSE. Hence, we are to choose a value of β for which the MSE of $\hat{b}$ attains its minimum. However, (2.7) cannot be found in practice, since the true b is unknown.

For a pre-specified weight function, let us consider the predicted value of y_i from GWR is denoted by $\hat{y}_i(\beta)$ (as a function of β). Then the *sum of squared errors* can be written as

$$SS(\beta) = \sum_i [y_i - \hat{y}_i(\beta)]^2 \tag{2.8}$$

A useful choice of β depends on a least square criterion. That is, we are to choose the value of β for which the quantity $SS(\beta)$ attains a minimum. In order to find the predicted value $\hat{y}_i(\beta)$, it is necessary to estimate the $b_j(u_i, v_i)$ at each of the sample points and then combine these with the x values at these points. However, a problem is encountered when minimizing sum of squared errors $SS(\beta)$. As $\beta \to 0$, $\hat{y}_i(\beta) \to y_i$; that is, $SS(\beta)$ in equation (2.8) is minimized when $\beta \to 0$. This is because, for all

kernel functions,

$$w_{ik} = \begin{cases} w_{ii} = 1 & \text{if } i = k, \\ w_{ik} = 0 & \text{as } \beta \to 0 \text{ if } i \neq k \end{cases}$$

To overcome this problem, a cross validation approach is suggested by Cleveland (1979) for local regression, and by Bowman (1984) for kernel density estimation; see also Golub, Heath and Wahba (1979), Li (1986) and the references therein for discussion of generalized cross validation in ridge regression. Let $\hat{y}_{(i)}(\beta)$ be the predicted value of y_i, obtained by omitting the i^{th} observation from the model, when the GWR estimation process is performed. Then the cross validated sum of squared errors is defined by

$$CVSS(\beta) = \sum_i [y_i - \hat{y}_{(i)}(\beta)]^2 \tag{2.9}$$

The value of β for which (2.9) attains its minimum is the logical choice that helps to overcome the problem obtained through equation (2.8).

2.4 Inference with the GWR Model

As described in the previous sections, the GWR estimation technique provides a means of computing localized regression estimates. It has been demonstrated to be a useful means for detecting spatial non-stationarity (Paez et al, 2002; Leung et al, 2000; Brunsdon et al, 1996). In GWR any spatial non-stationarity in the relationships being measured is accounted for by allowing the estimated model to vary spatially. From the statistical point of view, it is useful to assess the following two questions (Leung et al, 2000):

(1) On the whole, do the parameters in the GWR model vary significantly over the study region?

(2) Does each set of local parameters, $b_{ij} = b_j(u_i, v_i)$, $(i = 1, 2, \ldots, n)$ exhibit significant variation over the study region?

The first question can be modified as, "Does a GWR model describe the data significantly better than an OLS regression model?" This is, in fact, a goodness-of-fit test for a GWR model. The second question indicates that the variability of the local estimates could be thought of as a variance measure, and this is used to examine the plausibility of the stationarity assumption which is to be considered in classical regression. Furthermore, for any given j, the deviation of b_{ij} $(i = 1, 2, \ldots, n)$ can be used to evaluate the variation of the parameters associated with the j^{th} independent variable. However, it is very difficult to determine the null distribution of the estimated parameters. Therefore, a Monte-Carlo technique has been employed, called a permutation or randomization test.

2.4.1 F-Test Statistic

The work of Brunsdon et al (1999) provides a significance testing procedure for the GWR model. Following the conventional hypothesis testing framework, the notion of residual sum of squares is used to formulate the goodness-of-fit test. We assume that for calibrating the GWR model the weighting matrix is given. To find the distribution of the test statistic, the following two assumptions hold.

Assumption 1. The error terms ϵ_1, ϵ_2, ϵ_3, ..., ϵ_n are independent and identically distributed, following a normal distribution with zero mean and constant variance σ^2.

Assumption 2. Let $\hat{y}_i$ be the fitted value of y_i at location i. For all $i = 1, 2, \ldots, n$, $\hat{y}_i$ is an unbiased estimate of y_i. That is $E(\hat{y}_i) = y_i$ for all i.

The F-test is developed to test the null hypothesis that the coefficient $\mathbf{b}_j(u_i, v_i)$ is constant for all points (u, v) in the study area. No evidence of rejecting this hypothesis suggests that an ordinary, global regression model is adequate to describe the data set. Therefore, the hypotheses to be tested can be formulated as

$$H_0 : \frac{\partial b_j}{\partial u} = \frac{\partial b_j}{\partial v} = 0 \quad \forall j$$

versus

$$H_1 : \frac{\partial b_j}{\partial u} \neq 0 \text{ or } \frac{\partial b_j}{\partial v} \neq 0 \quad \forall j$$

The test statistic described in this section is produced by Brunsdon et al (1999) considering two models: the GWR model and the global regression model, where no variations are assumed for different localities. As in section 2.3, similar matrix notations are used here to provide a brief description of how to derive the appropriate test statistic.

In the GWR model, the coefficients $\mathbf{b}_j(u, v)$ vary across the study area. Following the OLS notation for the GWR model, $\mathbf{b}(u, v)$ can be treated as vector of coefficients in the global model, so that $\mathbf{b}(,)$ is a vector function mapping $\Re^2$, a two dimensional Euclidean plane, onto $\Re^m$, an m-dimensional Euclidean hyperplane. For the global model, an OLS estimate of the vector of parameters $\mathbf{b}(u, v)$ is given by

$$\hat{\mathbf{b}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}$$

Since no variation with respect to geographic space is assumed for global regression, the estimate $\hat{\mathbf{b}}$ is no longer a function of (u, v). Then the estimate of $\mathbf{y}$ can be written as

$$\hat{\mathbf{y}} = \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y} = \mathbf{S}_0\mathbf{y}$$

where $\mathbf{S}_0 = \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T$ is known as the hat matrix, or a smoothing operator since it transforms, or smooths, the observed $\mathbf{y}$ into $\hat{\mathbf{y}}$.

A weighted OLS estimate of $\mathbf{b}(u, v)$ is obtained when the estimation is performed with a weighting function $\mathbf{W}(u, v)$ such that the weighting changes as (u, v) varies. If the diagonal matrix $\mathbf{W}(u, v)$ consists of the diagonal element corresponding to the weighting for a particular (u, v), then

$$\hat{\mathbf{b}}(u, v) = [\mathbf{X}^T\mathbf{W}(u, v)\mathbf{X}]^{-1}\mathbf{X}^T\mathbf{W}(u, v)\mathbf{y}$$

For any given y_i, if the i^{th} row of $\mathbf{X}$ is $\mathbf{x}_i^T$ and the corresponding estimate is $\hat{\mathbf{b}}(u_i, v_i)$, then

$$\hat{y}_i = \mathbf{x}_i^T(\mathbf{X}^T\mathbf{W}(u,v)\mathbf{X})^{-1}\mathbf{X}^T\mathbf{W}(u,v)\mathbf{y}$$

The row vector $\mathbf{x}_i^T(\mathbf{X}^T\mathbf{W}(u,v)\mathbf{X})^{-1}\mathbf{X}^T\mathbf{W}(u,v)$ takes the observed $\mathbf{y}$ and smooths it to $\hat{y}_i$. Suppose $\mathbf{S}_1$ is the smoothing matrix for the GWR model so that its i^{th} row, $r_i = \mathbf{x}_i^T(\mathbf{X}^T\mathbf{W}(u,v)\mathbf{X})^{-1}\mathbf{X}^T\mathbf{W}(u,v)$. Then the estimate of $\mathbf{y}$ using the GWR model can be written as

$$\hat{\mathbf{y}} = \mathbf{S}_1\mathbf{y}$$

Obviously, both the hat matrices $\mathbf{S}_0$ and $\mathbf{S}_1$, computed from the global regression and GWR models, are independent of $\mathbf{y}$. In either model, the residuals may be expressed as

$$\hat{\epsilon} = (\mathbf{I} - \mathbf{S}_z)\mathbf{y}$$

where z is either 0 or 1. Therefore, the sum of squared residuals can be expressed as

$$\hat{\epsilon}^T\hat{\epsilon} = \mathbf{y}^T(\mathbf{I} - \mathbf{S}_z)^T(\mathbf{I} - \mathbf{S}_z)\mathbf{y} = \mathbf{y}^T\mathbf{R}_z\mathbf{y} \tag{2.10}$$

where $\mathbf{R}_z = (\mathbf{I} - \mathbf{S}_z)^T(\mathbf{I} - \mathbf{S}_z)$. The expression for the sum of squared residuals (2.8) is a quadratic form for both the GWR and classical regression models. If the assumptions about ϵ_i hold, then equation (2.10) is a quadratic form of normal variates. In this case, when both models can be expressed in the form of a hat matrix, Kendal and Stuart (1977) present the test statistic, for normally distributed $\mathbf{y}$, as

$$F = \left[\frac{\mathbf{y}^T\mathbf{R}_0\mathbf{y} - \mathbf{y}^T\mathbf{R}_1\mathbf{y}}{\nu}\right]\left[\frac{\mathbf{y}^T\mathbf{R}_1\mathbf{y}}{\delta}\right]^{-1} \tag{2.11}$$

where $\nu = Tr(\mathbf{R}_0 - \mathbf{R}_1)$ and $\delta = Tr(\mathbf{R}_1)$, and Tr is the trace of the matrix. The test statistic (2.11) has an approximate F distribution with degrees of freedom $(\nu^2/\nu', \delta^2/\delta')$, where $\nu' = Tr(\mathbf{R}_0 - \mathbf{R}_1)^2$, $\delta' = Tr(\mathbf{R}_1^2)$. These degrees of freedom are not necessarily integers; however, this does not affect the distributional assumption

provided $r > 0$ and $\delta > 0$. The approximation of the test statistic (2.11) to an F distribution depends on the fact that the numerator and denominator of (2.11) are quadratic forms of normal variates. These are well approximated by a χ^2 distribution with degrees of freedom chosen so that their first and second moments agree with those of the quadratic forms. Since there are hat matrices for both the GWR and classical regression models, it is possible to compare these two models using the test statistic (2.11). This is the extension of the conventional procedure of comparing classical regression models, where one consists of more explanatory variables than the other. In that case, the purpose is to fit the reduced model, where the F statistic follows an exact F distribution, because the degrees of freedom are an integer. Hence, an ANOVA table can be suggested for GWR-OLS comparisons, where the residual mean squared error for both is being compared.

2.4.2 Randomization Test

In the previous section, the F statistic aims to test whether the coefficients are constant over geographical space. Clearly, application of the F statistic (2.11) can produce a result of testing where the entire set of explanatory variables are used together for estimation. In this section, a different testing technique is illustrated, which aims to conduct similar inference but through inference on individual variables. Once a final model has been selected, we can further test whether or not each set of parameters in the model varies significantly across the study region. Brunsdon et al (1998) used the well established Monte Carlo techniques (Hope, 1968) to develop a method to test

$$H_0 : \quad b_j(u_i, v_i) = b_j, \forall i$$

versus

$$H_1 : \quad b_j(u_i, v_i) \text{ not all equal } \forall i$$

Testing of the above hypotheses actually measures the variability of $b_j(u_i, v_i)$ as i varies for a fixed j. Since the individual coefficient is to be tested, the test statistic is the variance of $b_j(u_i, v_i)$ across i :

$$v_j = \sum_i (\hat{b}_{ij} - \hat{b}_{.j})^2 / n \tag{2.12}$$

where $\hat{b}_{ij}$ is the GWR estimate of $b_j(u_i, v_i)$ and $\hat{b}_{.j}$ is obtained by averaging those over subscript i. The lower the value of v_j, the stronger the evidence that the coefficients corresponding to v_j is fixed. The null distribution of v_j is unknown, which leads us to apply a randomization testing technique to find its approximate distribution. Although Leung et al (2000) have used a transformation of v_j, and approximated as F distribution, we are not using this in our analysis. Under the null hypothesis, we assume that b_{ij} do not vary with i for a fixed predictor j. That is, little difference in the pattern of b_{ij} is suggested if the estimation of the GWR model were to be performed with locations of the observations randomly assigned to the predictor and response variables. More precisely, the spatial location should not greatly affect the parameter estimation if the b_{ij} are fixed over space. As explained by Brunsdon et al (1998), the randomization procedure, for given j, is as follows:

(a) Note the value of v_j for the correctly located observations.

(b) Randomly 'scramble' the locations p_i among the observations, and calculate v_j.

(c) Repeat the previous step P-1 times, noting v_j each time.

(d) Compute the rank of v_j for the correctly located case, R.

(e) The p-value for the randomization hypothesis is R/P.

Once the value of the bandwidth β is found by minimizing (2.9), the randomization test would be carried out following the steps as described above. In practice, a large number of random arrangements or scramblings is often required, so the overall computational requirements of this approach may be large.

2.5 Conclusion

In this chapter, we have demonstrated the GWR model and its methodology. The method of GWR can be used to produce localized parameter estimates, which appear to be a useful means to explore variation of parameters over space, and demonstrate complex spatial patterns. We have presented the details of two approaches of inference to assess spatial non-stationarity in relationships. In comparison to the linear regression model, GWR will provide less efficient estimates in the case when there is no spatial non-stationarity. However, it should be noted that when spatial non-stationarity is present, the classical regression model cannot provide a consistent estimate of the true model (Brunsdon et al, 2000).

Chapter 3

Simulation Studies

3.1 Introduction

The application of the estimation techniques and testing procedures to measure spatial non-stationarity is now described with simulated data. Our interest is to obtain estimates of the GWR coefficients and perform goodness-of-fit and randomization tests. The power and size of the tests will be studied empirically. This chapter presents results considering two different models: one with a single explanatory variable, and the other extended to three variables. Application of the weighting functions including choice of bandwidth, β, is also performed by using several methods.

3.2 Estimation in Single Predictor Model

The general form of the model with a single predictor, which is being simulated, is given by

$$y_i = b_{i0} + b_{i1}x_i + \epsilon_i \tag{3.1}$$

where $i = 1, 2, \ldots, n$; and $\epsilon_i \sim N(0, \sigma^2)$. The value $\sigma^2 = 1$ is chosen arbitrarily. The values of the independent variable x are drawn randomly from a uniform $(0, 1)$ distribution. The spatial region of interest consists of coordinates (u_i, v_i) taken from

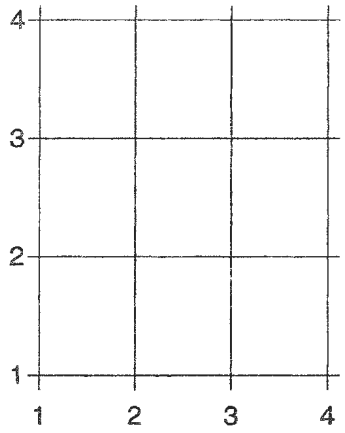

Figure 3.1: A grid with 4×4 Lattice Points

a square, two-dimensional grid. The simulation is performed in such a way that the grid consists of $m \times m$ lattice points with unit distance between any two of them along the horizontal and vertical axes. Figure 3.1 presents the lattice points in the study region to illustrate how the spatial region can be considered. Throughout this chapter, $i = 1$ refers to the location in the upper-left hand corner of the grid, $i = 2$ is the point to its right, and so on.

In the first case, we take $m = 4$, so we have $n = m^2 = 16$ observations in the study region. The spatial variation in the intercept and slope are chosen following a step changing approach. The b_{i0} and b_{i1} values for this case are considered as follows:

$$b_{i0} = \begin{cases} 1, & \text{for } i = 1, \ldots, 4 \\ 2, & \text{for } i = 5, \ldots, 8 \\ 3, & \text{for } i = 9, \ldots, 12 \\ 4, & \text{for } i = 13, \ldots, 16 \end{cases}$$

and

$$b_{i1} = \begin{cases} 1, & \text{for } i = 1, \ldots, 8 \\ -1, & \text{for } i = 9, \ldots, 16 \end{cases}$$

The value of the response variable $y_i, i = 1, \ldots, 16$, is generated by the model (3.1). Up to the stage of parameter estimation, we have used the weight functions of binary

Table 3.1: Summary statistics of parameter estimates: Binary weight function with $r = 2.0$.

	Intercept ($\hat{b}_0$)		Slope ($\hat{b}_1$)	
Points	Mean	St. dev.	Mean	St. dev.
1	1.723867	0.7838555	0.16604608	2.7338022
2	2.040854	0.6518693	-0.75787009	2.0100607
3	1.427912	0.6323068	0.90988045	1.0969543
4	1.231453	0.9695123	1.34819799	1.5703363
5	2.638822	0.6019597	-1.72037161	2.3612661
6	1.729607	0.4943086	0.97533019	0.8955057
7	1.841316	0.5422269	0.64377695	0.8858988
8	1.885211	0.7323107	0.62930668	1.1309353
9	2.438261	0.4896704	0.05075030	0.9974085
10	2.519601	0.4874250	-0.03684399	0.8666078
11	2.064949	0.6313220	0.79236817	0.9421332
12	2.482459	0.8763860	0.10675806	1.2172196
13	2.903235	0.6368353	0.05988004	1.1723625
14	3.014368	0.6207160	-0.28456138	1.0017228
15	3.254563	0.6785262	-0.69086805	1.0163031
16	4.014788	2.4692781	-1.40686089	2.8325846
Global statistic	2.168051	0.4349696	0.4388659	0.7490353

(2.3) and Gaussian (2.4) type. However, the latter one is used in the determination of power and size of the tests.

In the analysis, the Euclidian distance between points on the square grid (see Figure 3.1) are computed. Referring to the binary weight function in equation (2.3), the value of r is specified to 2 units. That is, to estimate the parameters of the GWR model corresponding to point i, a weight equal to 1 is considered for those points that are within 2 units of location i, and 0 for the points farther away. The estimates of b_{i0} and b_{i1} at 16 different points are computed from each of 500 simulated data sets. The mean and standard deviation of the estimates are presented in Table 3.1.

To interpret the result presented in Table 3.1, a close look at the parameter values and estimates of the corresponding points will help us to see the bias associated with

the GWR estimation method. As described previously, the true value of b_0 at each of the first four points is 1, whereas the means of the GWR estimates are 1.72, 2.04, 1.43 and 1.23 respectively. That is, for these points, the GWR estimates of the intercept term overestimate the true values, so there is positive bias with the estimates. For the next four points, the means of the b_0 estimates are 2.64, 1.73, 1.84 and 1.86 respectively, whereas the true value for each is 2. The estimate of b_0 at the fifth point has positive bias, while the other three have small negative bias. The true b_0 at the next four points is 3 and of the last four is 4. It is obvious that the GWR estimate of b_0 at the 16^{th} point shows slight positive bias, whereas the other 7 estimates are negatively biased. The standard deviations of the b_0 estimates are similar at each location except the 16^{th} point on the grid.

Unlike the b_0 estimates, the results obtained for $\hat{b}_1$ display greater departure from the true values. We know the coefficient b_1 equals to 1 at each of the first 8 locations and -1 at each of last 8 locations. Therefore, the two negative values of $\hat{b}_1$ averages corresponding to the second and fifth locations, whereas positive values corresponding to location 9 and 11-13 are not what we would expect. Also the standard deviations of the b_1 estimates are larger than those of $\hat{b}_0$.

The last row of Table 3.1 presents the statistics of $\hat{b}_0$ and $\hat{b}_1$ assuming the parameter values are unchanged over the study region. That is, these are simply the mean and standard deviation of the ordinary least squares (OLS) estimates of b_0 and b_1. Since no variation in parameter values are assumed for the OLS estimation, the model is called the global regression model and the statistics computed from estimates are known as global statistics. Obviously, many of the GWR estimates and corresponding parameter values are quite different from the respective global averages of b_0 and b_1.

Figure 3.2 displays the distribution of the $\hat{b}_0$ values at locations 1, 5, 9 and 13. Figure 3.3 presents the distribution of the $\hat{b}_1$ values at the corresponding locations. It is apparent that at each location the shape of the distributions of $\hat{b}_0$ and $\hat{b}_1$ are very close to normal. As well, they are centered close to the true b_0 values, and reasonably close to the true b_1 values.

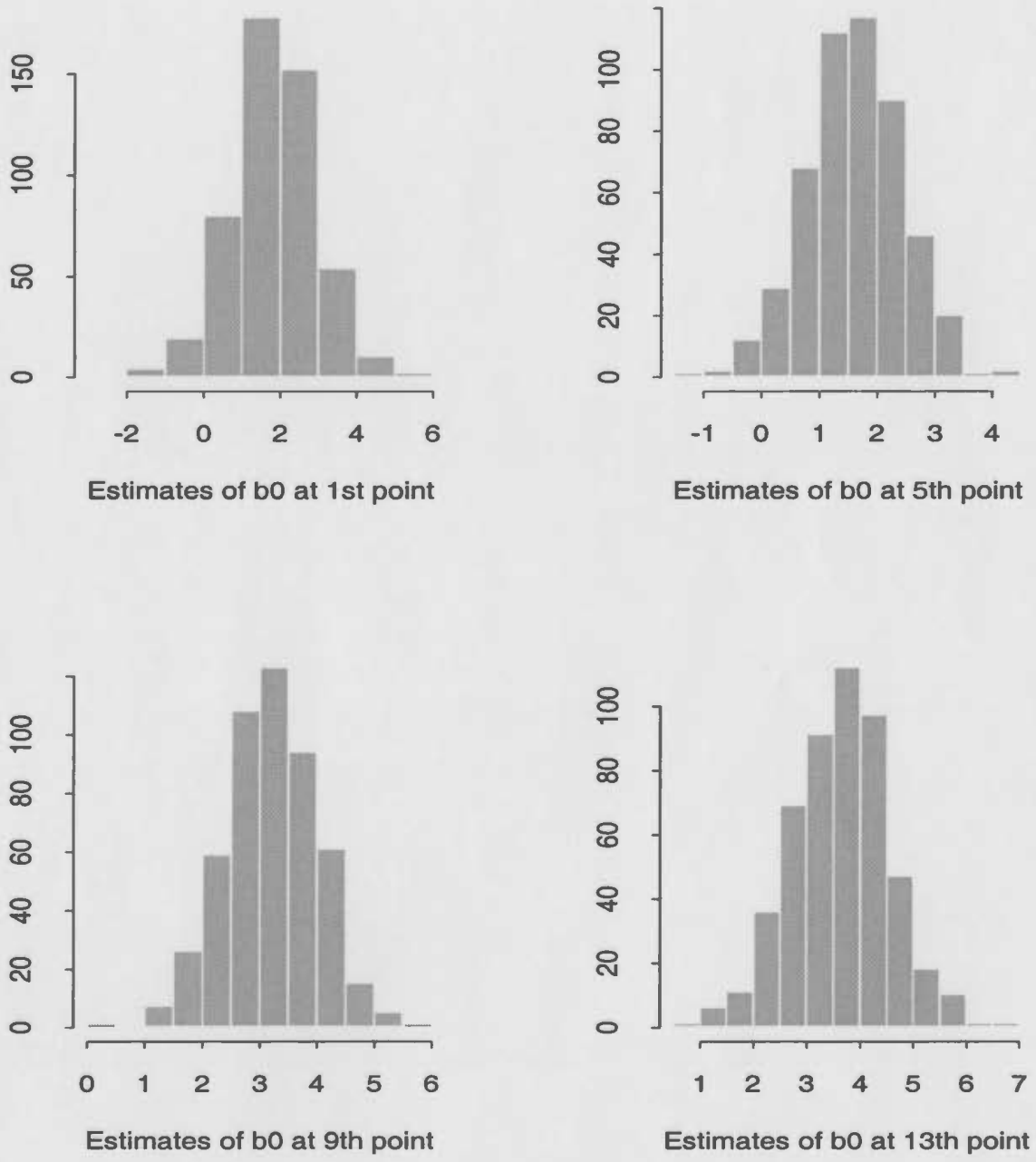

Figure 3.2: Distribution of the $\hat{b}_0$ values at several points on the grid

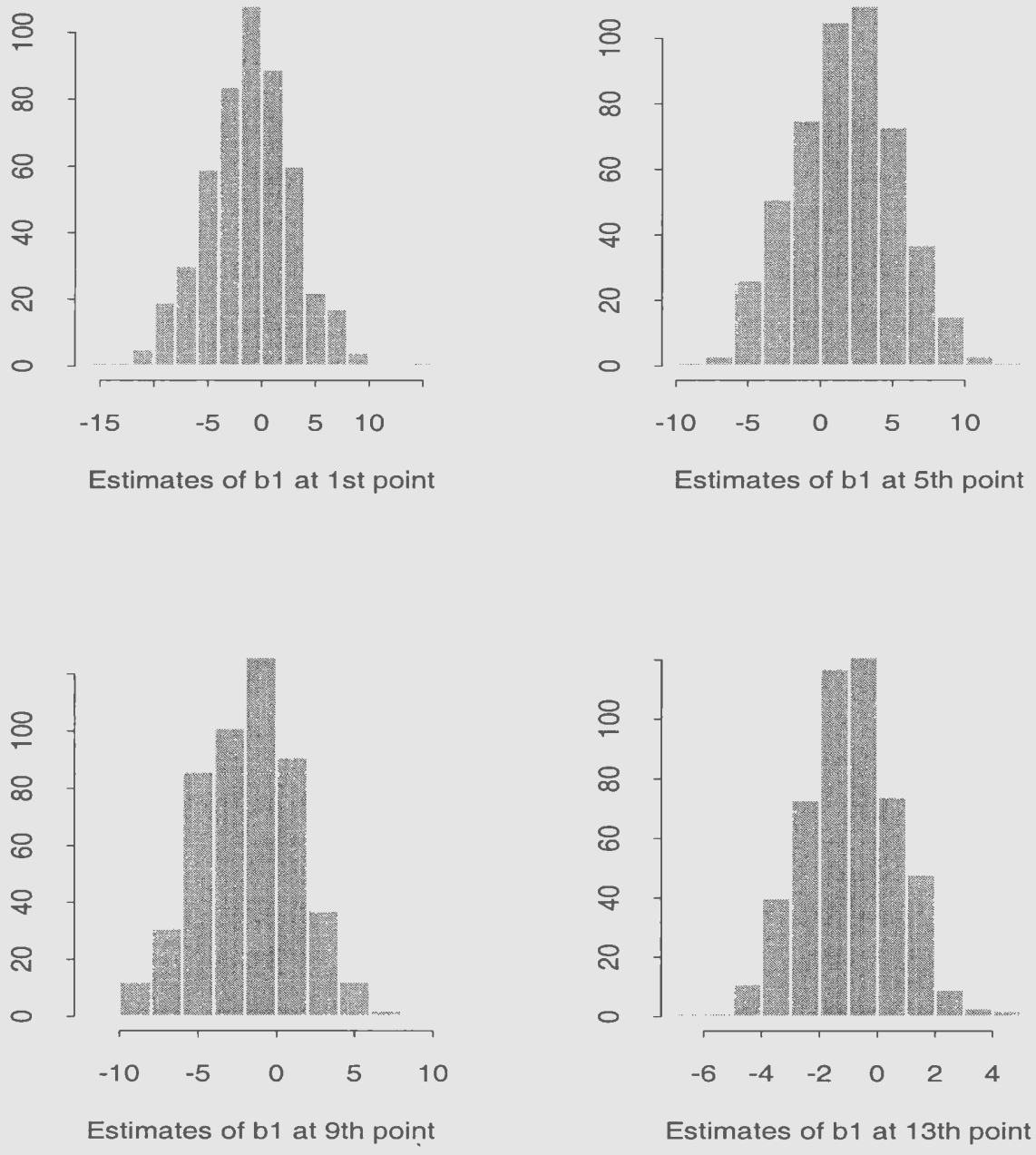

Figure 3.3: Distribution of the $\hat{b}_1$ values at several points on the grid

We have also used the Gaussian-distance-decay function (2.4) for analyzing this simulated data. Therefore, the bandwidth β had to be determined before parameter estimates were determined. The β value is first determined by the method of mean squared errors, and thereafter by the cross validation approach. Although the β values obtained from the latter approach are used in our analysis, choosing β based on the minimum attainable MSE in equation (2.7) gives an intuitive idea of cross checking the cross validation approach.

For empirical computation, the d_{ik}, distance between the i^{th} and k^{th} points ($i, k = 1, \ldots, 4$) on the square grid (Figure 3.1) is computed, and used to determine the Gaussian weight function (2.4). Initially, an arbitrary value of β is chosen to determine w_{ik}. Once the GWR estimates of b_0 and b_1 are obtained at each of 16 points on the grid by using the Gaussian weights, we determine the MSE and CVSS. Following the definition of MSE in (2.7), we can write its empirical formula as

$$MSE_i = \frac{1}{n}\sum_{j=1}^{n}(\hat{b}_{i,j} - b_i)^2$$

where b_i is the true value of b_i, and $\hat{b}_{i,j}$ is the corresponding GWR estimate of b. For each of 16 points on the grid, the MSE of $\hat{b}_0$ and $\hat{b}_1$ are estimated by using $n = 500$ simulated data. Trials over a range of β values help us to obtain the β for which the MSE attains its minimum. The results corresponding to three different choices of β are presented in Table 3.2. The MSE is minimized most often for $\beta = 2$. We determined this by examining a wide range of β values, but only three of those choices are presented in Table 3.2.

To find the empirical value of CVSS scores, we have used the formula in (2.9). Once the GWR estimates of b_0 and b_1 are found, the computation of $y_i - \hat{y}_{(i)}(\beta)$, i=1,2, ..., 16, is performed easily. Summing up the quantity $y_i - \hat{y}_{(i)}(\beta)$ over i gives the CVSS for a specific data set. Repetition of the same procedure on each data set gives 500 CVSS values, while the computational trials with several β values are performed to pick up the value of β which minimizes CVSS.

Table 3.3 presents the means of the CVSS scores and the number of times CVSS

Table 3.2: Mean squared errors of $\hat{b}_0$ and $\hat{b}_1$ corresponding to bandwidths $\beta = 1.0$, 1.5, 2.0

	β=1.0		β=1.5		β=2.0	
Points	mse $\hat{b}_0$	mse $\hat{b}_1$	mse $\hat{b}_0$	mse $\hat{b}_1$	mse $\hat{b}_0$	mse $\hat{b}_1$
1	0.9445774	2.093401	0.8344093	1.904298	0.7972115	1.4213389
2	0.8492792	2.137797	0.7469554	1.666718	0.8347536	1.2306280
3	4.5083120	8.812332	1.1692373	2.355303	0.9781568	1.2615225
4	19.1430513	50.836072	4.1119758	10.081826	1.3039886	2.0325877
5	1.3513548	2.137544	0.6492579	1.223976	0.4500557	1.1441368
6	0.8432945	1.671443	0.5027079	1.310045	0.5069929	1.5000168
7	2.2272753	4.495173	0.8262826	2.385633	0.7322085	2.2523445
8	13.3924956	28.378601	1.7542282	4.160823	1.0101686	3.0100945
9	1.1625503	2.258186	0.5655855	1.311947	0.4158165	1.2179893
10	1.3647600	2.096101	0.8608799	1.282856	0.5229588	0.9904287
11	2.2547807	4.659374	1.3445963	2.227735	0.7554245	1.2614566
12	5.4224494	9.697872	1.9648944	3.627675	1.0422258	1.8922757
13	1.1444721	4.903224	0.7484456	2.763793	0.6216082	1.5520608
14	1.0527807	4.973747	0.7432572	2.881863	0.6053905	1.8284279
15	1.2130773	4.476111	0.8607963	3.451822	0.6596120	2.3585991
16	3.6127100	10.766331	1.3867159	4.896241	0.8205366	3.0764862

Table 3.3: $CVSS$ scores for several values of bandwidth β

β values	Mean ($CVSS$)	Number of times $CVSS$ minimized
1	35.6428	78
1.5	26.9710	170
2.0	26.6153	163
5.0	31.9602	51
10.0	33.5571	5
20.0	33.9868	33
Total simulations		500

was minimized corresponding to the six different values of β. The means of the CVSS are obtained by averaging 500 CVSS scores under each value of β, where the least CVSS are computed separately for each data set over the six CVSS values corresponding to the values of β. Obviously, CVSS is minimized most often for β values equal to 1.5 and 2.0, which coincides with the least mean squared errors for β= 2.0. The smallest mean CVSS is 26.62, which is also found for β= 2.0. The next smallest mean CVSS is 26.97, which is observed when β= 1.5. However, the MSE quantity can only be calculated in simulation studies, when the true b_i values are known.

3.2.1 Binary versus Gaussian Weighting Function

As explained above, $\beta = 2.0$ is preferred based on minimum MSE and $CVSS$. We would like to compare the results when using the Gaussian weight function to what would happen if the binary weight function is used. Table 3.4 presents the average of the estimates of b_{i0} and b_{i1} using the Gaussian weight function with $\beta = 2.0$. These results can be compared to those in Table 3.1.

The average values obtained by using the Gaussian weight function seem to follow a downward trend within each of the four categories of true values of the intercept. For the slope, the first 8 are similar in value, as are the final 8 values. On the other hand, no such trend is observed for the binary weighting function. Rather, it shows large fluctuations among estimates in several points. Figures 3.4 and 3.5 display line diagrams separately for the b_0 and b_1 estimates and the corresponding true values at the 16 points on the grid.

Interestingly, the lines of $\hat{b}_0$ and $\hat{b}_1$ obtained by using the Gaussian weight function show a trend along the corresponding lines with the true parameter values and include a reasonable amount of positive and negative bias. However, the lines of the b_0 and b_1 estimates obtained using the binary weight function display large variability, in particular for $\hat{b}_1$.

Table 3.4: Summary statistics of parameter estimates: Gaussian weight functions with $\beta = 2.0$

	Intercept ($\hat{b}_0$))		Slope ($\hat{b}_1$)	
Points	Mean	St. dev.	Mean	St. dev.
1	1.761542	0.5367068	0.4559723999	1.0569348
2	1.718736	0.4994881	0.5884239226	0.8840573
3	1.640726	0.5473830	0.7695585314	0.9348922
4	1.525283	0.6557580	0.9785509492	1.0890110
5	2.059301	0.4722311	0.4013111474	0.8768544
6	1.995405	0.4606587	0.4990330045	0.7930556
7	1.912957	0.5066025	0.6263274860	0.8419422
8	1.819782	0.6003668	0.7726153581	0.9699079
9	2.487690	0.4662753	0.1029101711	0.8653345
10	2.381392	0.4553227	0.2435643254	0.7933064
11	2.281535	0.4880002	0.3643364632	0.8102761
12	2.225218	0.5748977	0.4424444235	0.8968705
13	2.965182	0.5547447	-0.3029614687	0.9833593
14	2.834534	0.5148213	-0.1260060126	0.8822922
15	2.727228	0.5191510	-0.0016574621	0.8533367
16	2.724838	0.6466617	-0.0001706193	0.9274530

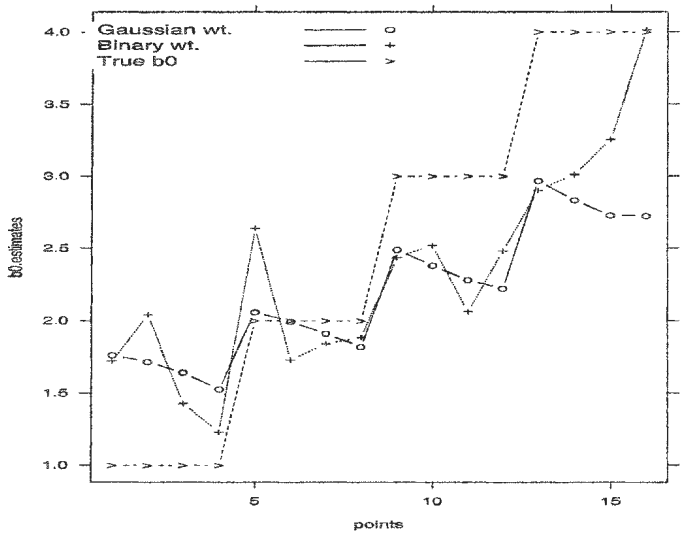

Figure 3.4: Pattern of b_0 estimates: Binary versus Gaussian weight function

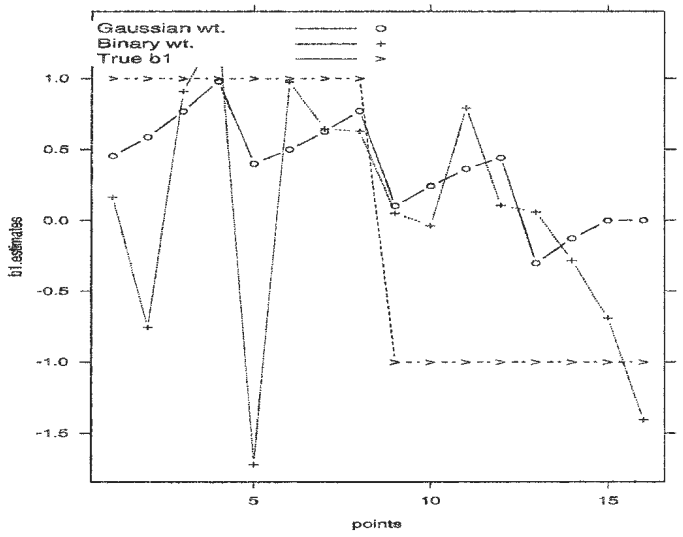

Figure 3.5: Pattern of b_1 estimates: Binary versus Gaussian weight function

3.3 Performance of Tests

In order to detect the presence of spatial variation among parameter values, two different statistical testing methods were described in chapter 2. We will now empirically study the performance of these tests. At this stage, a note of distinction between the two tests should be mentioned here. When one would like to apply goodness-of-fit test or the randomization test on local variation, the former one is used to make simultaneous inference on all parameters, whereas the latter assesses the contribution of an individual explanatory variable.

3.3.1 Power and Size of Tests: Single Predictor Model

To assess the performance of testing procedures of spatial non-stationarity, both the goodness-of-fit test and randomization test are applied to the data simulated by model (3.1). As described in section 2.4.1, the hypotheses to be tested under the goodness-of-fit testing procedure are

H_0 : The coefficients $\mathbf{b}(u_i, v_i)$ are constant for all points (u_i, v_i) in the study area versus

H_1 : The coefficients $\mathbf{b}(u_i, v_i)$ are not constant for some of the points (u_i, v_i).

The power and size of tests are two well-known statistical tools through which one can justify the level of acceptance of a test. Since for the simulated data the true state of nature of the parameter values are known, the power and size of tests are possible to determine. The power of a statistical test is the probability of rejecting the null hypothesis when in fact it is false and should be rejected. On the other hand, the size of a test is the probability that the test will lead to the rejection of the null hypothesis when the null hypothesis is true. To obtain empirically the power of the test, we have chosen b_0 and b_1 values at 16 different points in the grid in such a way that there exists spatial variation among the parameters. Two separate choices are considered: one with the purpose of observing the result with relatively small variation among the parameter values at different points, whereas the other with large variation among the parameter values. The size of the test will be studied empirically with data generated with no variation among the b_{i0} and b_{i1} values. The results are presented in Table 3.5.

Case 1: Simulated data with small variation in parameter values.

In this case, the same simulated data of section 3.2 are used. The Gaussian weighting function is used in the analysis, where the bandwidth β is chosen using the cross validation approach. The goodness-of-fit test produces the value of the F statistic of equation (2.11) for each of 500 data sets. Since the null distribution of the F statistic is approximated by an F distribution, and the analysis of these data gives 8.78 and 11.04 for the numerator and denominator degrees of freedom respectively, the corresponding p-values are obtained. At the 5% level of significance, it is found that the test would reject the null hypothesis 124 times out of 500 data sets which leads to the power of the test equal to 0.25 (see Table 3.5).

Table 3.5: Simulation results of single predictor model. NDF = the numerator degrees of freedom and DDF = the denominator degrees of freedom.

	Goodness-of-fit			Randomization	
Case	NDF	DDF	Power	Power v_0	Power v_1
1	8.78	11.04	0.25	0.06	0.09
2	12.70	9.10	1.00	0.57	1.00
			Size	Size v_0	Size v_1
3	2.99	14.00	0.01	0.07	0.05

Case 2: Simulated data with relatively large variation in parameter values.

In this case, simulations are performed considering the b_{i0} and b_{i1} values as follows:

$$b_{i0} = \begin{cases} 1, & \text{for } i = 1, \ldots, 4 \\ 10, & \text{for } i = 5, \ldots, 8 \\ 20, & \text{for } i = 9, \ldots, 12 \\ 50, & \text{for } i = 13, \ldots, 16 \end{cases}$$

and

$$b_{i1} = \begin{cases} 5, & \text{for } i = 1, \ldots, 8 \\ 20, & \text{for } i = 9, \ldots, 16 \end{cases}$$

Obviously, the spatial variation in the true parameter values is much greater in Case 2 than in Case 1. When the Gaussian weighting function is used for GWR analysis, all 500 data sets attain a minimum CV either for $\beta = .8$ or $\beta = 1.0$. As before, the analysis produces 500 F-statistic values along with the common degrees of freedom 12.70 and 9.10 for the numerator and denominator of the F distribution respectively. It is found that all of the 500 p-values are less than $\alpha = 0.05$. Therefore, the power of the test becomes 1.0 when the data has a relatively large amount of spatial variation in the parameters.

Case 3: Simulated data with no spatial variation in parameter values.

In this case, the interest is to apply the goodness-of-fit testing method and empirically find the size of the test. For these simulations we have chosen $b_{i0} = 3$ and $b_{i1} = 2$ for all i. No change of parameter values for the locations implies no spatial variation in the data set. To use the Gaussian weight function, over 50% of the minimum $CVSS$ scores are found for $\beta = 20$, and almost 25% for $\beta = 3$. Out of 500 data sets, only 5 tests found to be rejecting the null hypothesis at the 5% level of significance. Hence, the size of the test appears smaller than we would theoretically expect; see Table 3.5.

We now expand on what we observe on Table 3.5. In Case 1, when spatial non-stationarity exists but not with large variation in parameter values, some of the p-values of the corresponding test statistic are small and some are large. In Case 2, when spatial non-stationarity exists with relatively large variation in the parameter values, all of the p-values of the corresponding F-statistics becomes very small. In the third (stationary) case, we get very few small p-values. Hence, it can be concluded that when variations do exist, the p-values of the statistic becomes smaller, and accordingly the sensitivity of the statistics proposed to explore spatial variations in the parameters, and the power of the test becomes high. On the other hand, when no variation exists, the p-values of the statistic become large, and leads to the conclusion that the data do not support a model with spatial variability. Therefore, the goodness-of-fit test correctly reflects the properties of the true model in terms of simultaneous inference in the single predictor GWR model.

To assess spatial non-stationarity related to the individual contribution of each explanatory variable in the model, the randomization test of section 2.4.2 is applied to the data simulated in the above three cases. Since this procedure is performed to test whether or not each set of parameters in the model varies significantly across the study region, the hypothesis of interest would be to test whether the data support a model with constant intercept over location i and individually constant slope over location i. As described above, the context of data generation, number of simulations, weighting function and choice of bandwidth are the same. However, to obtain the value of the

test statistic v_j in equation (2.12) we use a Monte-Carlo approach such that in each of 500 simulated data sets, we randomly rearrange (or scramble) our observations among the 16 spatial locations. This scrambling is done 200 times, and v_j from (2.11) is calculated in each of these 200 rearrangements. These values, along with the v_j value of the original data arrangement, make up our randomization distribution for v_j. The p-value corresponding to v_j is equal to the number of scrambled v_j less than or equal to the unscrambled one divided by 200.

For Case 1, in which there is spatial non-stationarity with small variation in the true parameters, the power of the test corresponding to the statistics v_0 and v_1 are 0.062 and 0.09. For Case 2, in which there is spatial non-stationarity with larger variation in the parameter values, the power of the tests are 0.57 and 1.0 respectively. In Case 3, we find the size of the tests are 0.07 and 0.05 corresponding to the test statistics v_0 and v_1 at the 5% level of significance. As described in section 2.4.2, the statistic v_0 is for testing whether the intercept, b_0, varies over location i, and v_1 is for assessing whether the slope, b_1, changes over i.

Therefore, the conclusion for this test can be outlined as follows. If non-stationarity exists in the data with a single predictor, the power of the test tends to increase with increasing variability among the true parameter values. Although the result reflect the true state of parameter variations, such reflection may differ among coefficients. In this case, the test performs well to test variability of the slope over several locations in space; however the results are not as good for the intercept term.

3.3.2 Power and Size of Tests: Multi Predictor Model

The inferential analysis of a multi-predictor GWR model is simply an extension of section 3.3.1. The objective is to observe if there is any significant change in the performance of the testing procedures with data available on more than one predictor. We consider a GWR model with three predictors, which can be written in the form

$$y_i = b_{i0} + b_{i1}x_{i1} + b_{i2}x_{i2} + b_{i3}x_{i3} + \epsilon_i \tag{3.2}$$

where, as earlier, $\epsilon_i \sim N(0, \sigma^2)$ with $\sigma^2 = 1$. A similar arrangement in a grid is considered to represent the coordinates (u_i, v_i) of a spatial region of parameter variation. An extension of Figure 3.1 into 6×6 lattice points means we have $n = 36$ observations. The values of the independent variables x_1, x_2 and x_3 are taken randomly from a uniform distribution on interval $(0, 1)$, a normal distribution with mean 10, variance unity and a standard normal distribution respectively. The selection of true parameter values also follows a step changing approach in 36 points in the grid, and is described in each case below.

To determine the power of the tests we take the parameter values of model (3.2) so that it represents spatial non-stationarity among the points in the grid. To study the size of the tests we apply the same testing procedure to the model where no parameter variation exists. Considering these situations, the detail of data generation along with application of the tests described in Cases 4, 5 and 6, is as follows. The results are summarized in Table 3.6. Again we use 500 simulated data sets.

Case 4: Three predictor model data simulation: small variation in parameter values.

To generate data with spatial non-stationarity but relatively small variation in parameter values, the b_{i0} is chosen with the values:

$$b_{i0} = \begin{cases} 1, & \text{for } i = 1, \ldots, 9 \\ 2, & \text{for } i = 10, \ldots, 18 \\ 3, & \text{for } i = 19, \ldots, 27 \\ 4, & \text{for } i = 28, \ldots, 36 \end{cases}$$

The coefficients b_{i1}, b_{i2} and b_{i3} take 1, 2 and 1.5 for the first 18 points and -1, -2 and -1.5 for the last 18 points respectively. To apply GWR testing methods, the Gaussian weighting function is used with β =1.6 which minimizes the cross validated sum of squared errors for all 500 data sets. The F-statistic is computed from the data sets with common degrees of freedom 26.21 and 19.87 for numerator and denominator

respectively. All 500 p-values obtained from the goodness-of-fit test are less than $\alpha = 0.05$, and hence the power becomes 1.0.

Case 5: Three predictor model data simulation: large variation in parameter values.

Following a similar procedure, the data are simulated considering relatively large variation in parameter values. As before, to choose intercept values, the points are separated into four categories, each of which contains consecutive 9 points. For simulation, b_{i0} takes following values.

$$b_{i0} = \begin{cases} 1, & \text{for } i = 1, \ldots, 9 \\ 10, & \text{for } i = 10, \ldots, 18 \\ 20, & \text{for } i = 19, \ldots, 27 \\ 50, & \text{for } i = 28, \ldots, 36 \end{cases}$$

To choose the coefficients of the three explanatory variables, the points are divided into two separate categories so that each contains 18 points. The data sets are generated in such a way that b_{i1}, b_{i2} and b_{i3} take 5, 4 and 25 for the first 18 points, and 20, -10 and 5 for the last 18 points respectively. For 496 data sets, $\beta = 1.6$ and for the remaining four $\beta = 1.7$ minimizes the CV function. Table 3.5 indicates that the power of the test is found to be equal to 1.0.

Case 6: Three predictor model data simulation: no variation in parameter values.

Since our interest is to determine size of the test, we have chosen the parameter values to be equal at the different points on the grid. For the simulations, we take $b_{i0} = 3$, $b_{i1} = 2$, $b_{i2} = -1$ and $b_{i3} = 5$ for all i, following the model (3.2). In this case, $\beta = 5$ and 50 minimizes most of the CV scores when the Gaussian weight function is used for analysis. Interestingly, for the stationary case regardless of single or multi-predictor model, most of the minimum values of the cross validated sum of squared

Table 3.6: Simulation results for model with three predictors. NDF = the numerator degrees of freedom and DDF = the denominator degrees of freedom.

	Goodness-of-fit			Randomization			
Case	NDF	DDF	Power	Power v_0	Power v_1	Power v_2	Power v_3
4	26.21	19.87	1.00	0	1	0	1
5	26.21	19.87	1.00	0	1	0	0
			Size	Size v_0	Size v_1	Size v_2	Size v_3
6	1.67	31.11	0.02	0.04	0.04	0.05	0.06

errors are found at two quite different values of β. Out of 500 p-values corresponding to the F statistics, only 12 are found to be less than 0.05, and hence the size becomes 0.02.

For the randomization test, the steps described at the end of sub-section 2.4.2 are carried out. Application of this testing procedure to the model with three predictors means our interest is in assessing whether there is any variation for each individual set of parameters b_{ij} over the location $i = 1, 2, \ldots, n$ on the study region. As simulated in Cases 4, 5 and 6, exactly the same data sets are used to perform this test. A procedure that is similar to that done for the single predictor model case is applied to obtain the empirical values of the power and size for the model with three predictors. The number of permutations used in this case is also 200. The results corresponding to the coefficients of x_1, x_2 and x_3 are presented in the last four columns of Table 3.6. When the analysis of the data generated in Case 6 is done with the randomization test, the empirical values of the sizes corresponding to v_0, v_1, v_2 and v_3 become 4%, 4%, 5% and 6% respectivley . Since α is pre-considered at 0.05, the results of the empirical study seem reasonable. However, the power computed by using the data generated in Cases 4 and 5 are not what we would expect. For Case 4, the power of the randomization test corresponding to v_1 and v_3 are equal 1, whereas for v_0 and v_2 are 0. For case 5, the power corresponding to v_1 becomes 1; however for the other three it is 0. Therefore, it seems to be the test works well when the model

is of stationary over the study region. However, for spatial non-stationarity the test procedure for the multi-predictor model does not work as well. It is not clear why the randomization test gives the results that we observe. These test procedures will be applied to a socio-economic data set in Chapter 4 with possible explanations for the behaviour of the tests in Chapter 5.

3.4 Conclusion

The simulation studies presented in this chapter appears to validate the application of GWR. To strengthen the theoretical basis, we have used 500 data sets throughout the analysis, and 200 permutations for each to apply the Monte Carlo technique. The F-test results indicate its ability to identify the overall effect of spatial variation in relationships. However, as noted in our simulation results, the radomization test may have problems in detecting the presence of spatial non-stationarity for individual parameters, even when this non-stationarity is strong. Brunsdon et al (1998) has applied this test to a data set without verifying its performance. Although the performance of this test has been presented in some works with the simulation results of single predictor model (Leung et al, 2000), an extension of this to a multi-predictor model indicates we may need to be cautious in generalizing the technique.

Chapter 4

Analysis of a Socio-Economic Data and Spatial Non-stationarity

4.1 Introduction

The presence of spatial non-stationarity in many of data sets is not unusual, though analyses are often performed which ignore this issue. Socio-economic study is one of the important areas, where much of the data are obtained either from a census or large scale survey. Generally, the data from a relatively large study area include a wide variety in variables, which leads us to consider the context of spatial non-stationarity in analysis. To measure spatial variations in the relationships of parameter values, we will analyze house price data with the GWR methodology. The benefits of using the GWR inferential techniques is also discussed following the methods which assumes no spatial variation, i.e. ordinary least squares regression. The final issue will be to assess the goodness-of-fit and randomization testing procedures of the GWR model we have chosen. The stepwise regression and best subset methods are used initially for the purpose of selecting a suitable subset of the explanatory variables.

4.2 Data Description

Socio-economic data on housing prices obtained from the Boston Standard Metropolitan Statistical Area (SMSA) is used for illustrating the GWR inferential techniques of measuring spatial non-stationarity. Harrison and Rubinfeld (1978) used this data set to analyze various methodological issues of hedonic housing prices to estimate the demand for clean air. The hedonic price index is based on the fitted values of a regression of price on the various explanatory variables and is used to represent its qualitative determinants. The Census Bureau Publication in 1970 is the source of the majority of the data, whereas some of the explanatory variables were added from several other studies (see Harrison and Rubinfeld, 1978, Table IV, p 96-97). The basic data are a sample of 506 observations on 16 variables on census tracts (1 observation per census tract) in the Boston SMSA -1970. A brief description of the variables is presented in Table 4.1.

Harrison and Rubinfeld (1978) focused on the willingness to pay for air quality improvements using this data with several methods of analysis. To examine the effects of robust estimation, Belsley et al (1980) also used this data. Many authors, such as Krasker et al (1983), Subramanian and Carson (1988), Brieman and Friedman (1985), Lange and Ryan (1989), Breiman et al (1993), Pace (1993), have used the data to examine robust estimation, normality of residuals, and non-parametric and semi-parametric estimation.

Harrison and Rubinfeld (1978) described this data set as superior in comparison to others since it consists of a large number of neighborhood variables. The median value of the owner-occupied homes in the census tract is considered as the dependent variable for the regression models. It is expected that the house price is influenced by the structural aspects of houses, amenities in their neighbourhood, ease of access to employment, and how free the area is from air pollution. There are 15 explanatory variables available in the data set that are categorized into four different types with respect to the variables' contribution to the house value. These are structural, neighbourhood, accessibility/locality and air pollution (Table 4.1). As part of the

Table 4.1: Variables in the Boston house price data set

Variable	Definition
Dependent	
medv	Median value of owner-occupied homes in $1000s
Structural	
rm	Average number of rooms per dwelling
age	Proportion of owner occupied units built prior to 1940
Neighborhood	
b	$1000(Bk - 0.63)^2$ where Bk is the proportion of blacks by town
lstat	Proportion of lower status of the population
crim	Per capita crim rate by town
zn	Proportion of residential land zoned for lots over 25,000 square feet
indus	Proportion of non-retail business acres per town
tax	Full-value property-tax rate per $10,000
ptratio	Pupil-teacher ratio by town
chas	Charles River dummy variables (1 if tract bounds the river ; 0 otherwise)
Accessibility/locality	
dis	Weighted distances to five Boston employment centers
rad	Index of accessibility to radical highway
latt	Standardized Latitude coordinates
long	Standardized Longitude coordinates
Air pollution	
nox	Nitric oxides concentration (parts per 10 million)

structural aspect of houses, *rm* represents spaciousness and, in a certain sense, quantity of housing. The unit age is usually related to the structural quality. Out of 15 explanatory variables, eight are considered to be relating to the neighbourhood amenities. The context of the variable b helps to observe if there is any effect of Black-White neighbourhood on house price. Through an exploratory analysis Harrison and Rubinfeld (1978) observed that at a low to moderate proportion of Blacks in the Boston area, an increase in the Black population has a negative influence on housing, whereas the reverse trend is found when the proportion of Black people becomes very high. Based on this parabolic trend, b is created by shifting the proportion 0.63 towards the origin. The variable *lstat* is for indicating the proportion of lower status people, which is obtained by averaging the proportion of adults without some high school education and the proportion of male workers classified as laborers. It is assumed that the crime rates, defined by *crim* are generally proportional to people's perceptions of danger. Since crime gauges threat to the habitants' well-being and people of lower status might be undesirable neighbours, both are expected to have a negative effect on the housing price. The residential land zoned for lots greater than 25,000 square feet restricts construction of small lot houses, and accordingly welcomes the higher class people to be neighbours. This also maximizes the outdoor amenities to a community. Hence, the proportion *zn* should have a positive effect on housing price. The variable *indus* is considered to serve as proxy measure of externalities associated with industry, such as, noise, heavy traffic and unpleasant visual effects. The variable *tax* indicates the full value property taxes (\$/\$10,000), which measures the cost of public services in the communities. The local assessment ratio is used as a correction factor with the nominal tax rate and yields the full value tax rate which varies from town to town. The *ptratio* is treated as an indicator of measuring public sector benefits in towns. The lower the *ptratio*, the better opportunities for a child's education. Therefore, an increase in *indus*, *tax* and *ptratio* would not be expected by the residents and should have a negative influence on housing prices. The variable *chas* captures the amenities of a riverside location and thus the corresponding

regression coefficient would be expected to show a positive effect on house price.

4.3 Exploratory Analysis

The application of statistical methods usually demands some exploratory analysis of the variables included in the data set. Though our ultimate interest is to proceed to fit a GWR model with a limited number of explanatory variables of the house-price data, the exploratory look at several variables of the data set will help to make inference through analysis. Figure 4.1 presents the relationship of house price with changing the proportion of black people.

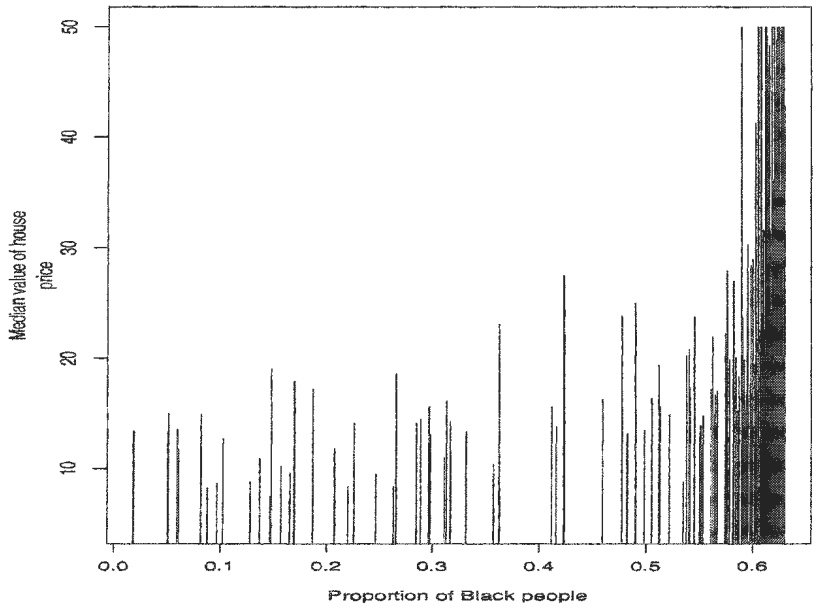

Figure 4.1: Pattern of house prices against the proportion of Black people

The house price increases when the proportion Black in the population becomes more than 60%. Due to this in Table 4.1 we observe that the variable b is created by shifting the proportion 0.63 to 0. The map in Figure 4.2 exhibits that the study area consists of eight different sub-regions: Inner Core, North Shore, North Suburban,

Table 4.2: Distribution of observations by sub-region

Sub-region	# of town	Observation	
		#	%
Inner Core	24	331	65.42
North Shore	11	36	7.11
North Suburban	9	33	6.52
Minuteman	4	12	2.37
Metro West	8	39	7.71
South West	4	4	0.79
Three Rivers	8	28	5.53
South Shore	11	23	4.55
Total	79	506	100.00

Minuteman, Metro West, South West, Three Rivers, South Shore.

For convenience of analysis, we attempt to observe sub-region summaries of several variables. Although the data set covers the Boston Standard Metropolitan Statistical Area (SMSA) in the 1970 census, most of the observations (65%) are selected from the sub-region of the Inner Core (Table 4.2). It was found from a further look that as a single city, Boston consists of the highest number of observations (132), which is also included in the Inner Core. Therefore, a major portion of the data tends to be concentrated in the city towns of the study area.

Since in the next sections of analysis our main attention will be to assess if there is any spatial variation of the effect of other variables on house price, we now explore more aspects of the data set. The results in Table 4.3 indicate that the houses with the lowest mean value are in the Inner Core area. One may find this surprising in the sense that this consists of the central towns of the Boston SMSA including the city of Boston, where the housing value may be assumed to be higher. However, large cities often have some poor, rundown areas. These are more likely in the city center. A large gap between the lowest ($20,692 in the Inner Core area) and the highest ($34,083 in Minuteman) means can be treated as an indication that the location may be a contributing factor to the house price. Observing the three largest standard

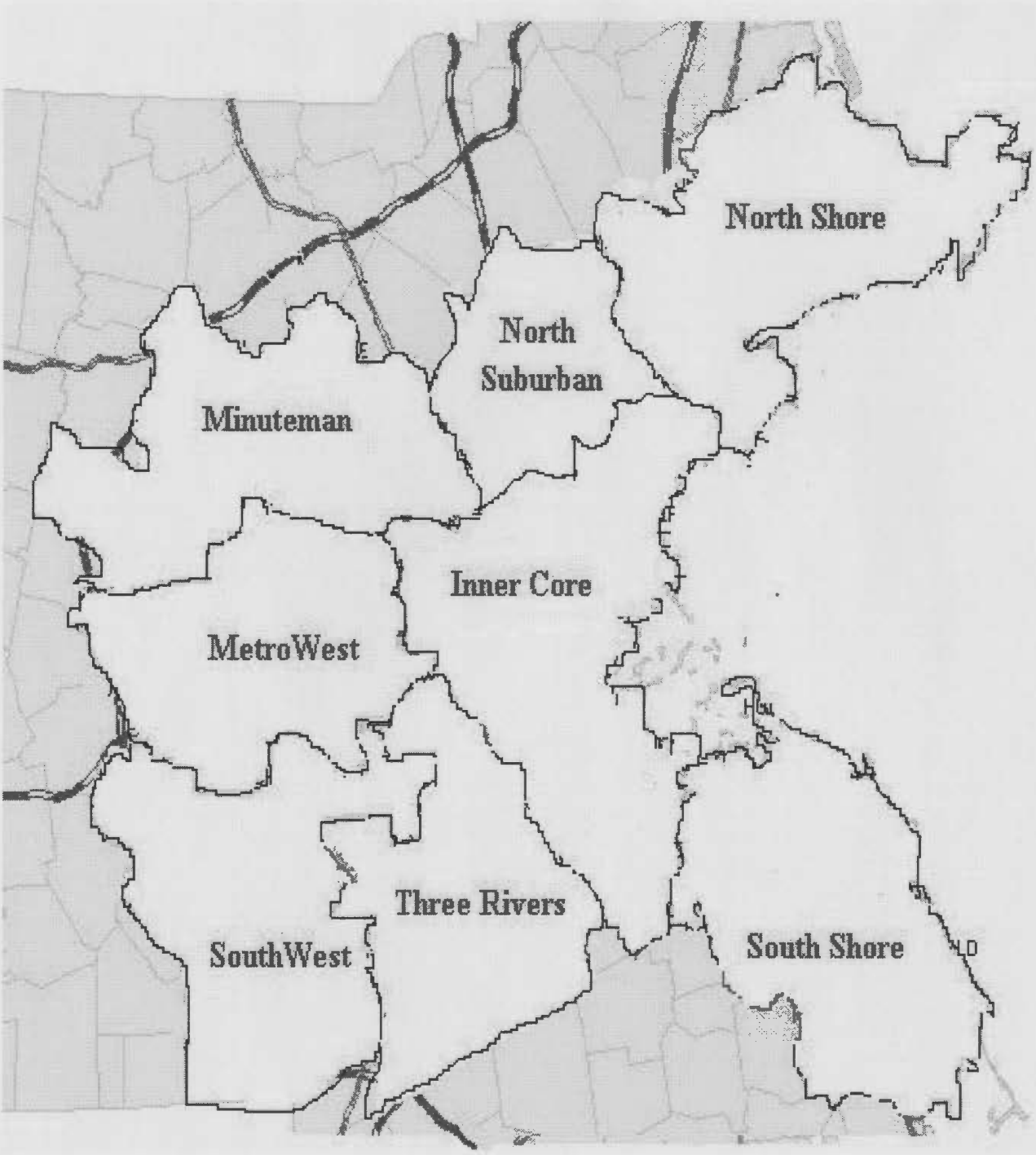

Figure 4.2: Map of Boston Standard Metropolitan Statistical Area

Table 4.3: Summary statistics of house price by sub-region, in $1000s.

Sub-region	Mean	Std. dev.	n
Inner Core	20.692	9.559	331
North Shore	23.336	5.941	36
North Suburban	24.776	5.723	33
Minuteman	34.083	5.660	12
Metro West	27.685	8.966	39
South West	34.025	15.201	4
Three Rivers	27.214	4.865	28
South Shore	22.087	4.504	23
Overall	22.533	9.197	506

Table 4.4: Summary statistics of house price by riverside residents, in $1000s

Do tracts bound the Charles River?	Mean	Std. dev.	n
Yes	28.440	11.817	35
No	22.094	8.831	471
Overall	22.533	9.197	506

deviations, it can be explained that the house prices in the South West, Inner Core and Metro West regions are more heterogeneous than the other regions.

As explained in section 4.2, the Charles River provides some additional amenities to the residents who are living by the riverside. From Table 4.4, we see that out of the total sample of 35 houses bounded by the Charles River, the mean price of those houses is $28,440, which is much higher than those are farther away from the river. All these issues produced from our exploratory analysis give us some indication that along with other variables in the data set, the location of census tracts or houses may have some influence on the housing values in the Boston SMSA. Therefore, combining the spatial aspect with the model fitting approach will be followed in the next stages of analysis to determine whether there is any spatial pattern in the relationships.

4.4 Choice of Model

In this section we will move through a sequential procedure of regression analysis with the aim of choosing a simplified model for house price. The model fitting approaches start with an ordinary least squares (OLS) regression, where all possible explanatory variables are included. The stepwise regression procedure is applied for the purpose of excluding less relevant variables from the model. A model using a subset of the predictors will be selected to apply the GWR method of estimation. An attempt of examining the traditional methods is also taken to observe the difference in the linear regression and GWR estimation techniques. We note that, to our knowledge, no procedures have been developed on variable selection in GWR.

4.4.1 OLS Regression with all Possible Regressors

As described in section 4.2, there are 506 observations in the data set with 15 explanatory variables available for model fitting. In the linear regression model, we begin by fitting a model with all possible regressors, and observe the contribution of individual variables on the house price. However, the ultimate purpose is to fit a GWR model and compare the results obtained from the OLS regression and GWR estimation methods. The spatial structure in the study area plays a key role in the GWR estimates, and leads to a difference with the estimates obtained from the OLS procedure. Therefore, to fit an OLS regression model, we use all the explanatory variables in Table 4.1 except *dis, latt* and *long*. Our starting linear regression model is as follows:

$$\begin{aligned} medv_i &= b_0 + b_1 crim + b_2 zn + b_3 indus + b_4 chas + b_5 nox + b_6 rm + b_7 age \\ &\quad + b_8 rad + b_9 tax + b_{10} ptratio + b_{11} b + b_{12} lstat + \epsilon_i \end{aligned} \tag{4.1}$$

The application of the OLS estimation method produces the F-statistic value equal to 101.5 with numerator and denominator degrees of freedom 12 and 493 respectively, and hence the $p-$value computed to zero. This indicates an overall significance that at

Table 4.5: Results of the OLS estimation with 12 predictors and 506 observations

Variable	**b-coefficient**	Std. Error(**b**)	t-value	P-value
Intercept	23.0244	5.0226	4.5842	0.0000
crim	-0.0796	0.0344	-2.3161	0.0210
zn	0.0055	0.0132	0.4176	0.6764
indus	0.1166	0.0633	1.8420	0.0661
chas	2.7872	0.9072	3.0723	0.0022
nox	-9.7996	3.8593	-2.5392	0.0114
rm	4.2228	0.4362	9.6816	0.0000
age	0.0292	0.0133	2.1912	0.0289
rad	0.3138	0.0699	4.4923	0.0000
tax	-0.0129	0.0040	-3.2520	0.0012
ptratio	-1.0434	0.1372	-7.6065	0.0000
b	0.0097	0.0028	3.4472	0.0006
lstat	-0.5392	0.0534	-10.1031	0.0000

least one of the explanatory variables included in (4.1) has a very strong contribution on the house price.

However, to observe the individual contribution of the variables we examine the estimated coefficients presented in Table 4.5. Clearly, out of 12 explanatory variables included in the model (4.1), only the *zn* shows a large $p-$value. The variable *indus* presents an evidence of positive relationship to the house price if we consider the level of significance $\alpha = 0.1$. However, the coefficients corresponding to the other explanatory variables exhibit a strong evidence that the house prices are affected by those variables at $\alpha = 0.05$, and in some cases at 0.01. The OLS estimates of the coefficients presented in the second column of Table 4.5 demand careful interpretation for the model. We see that a one unit increase of per capita crime rate implies a decrease of the median house price equal to $0.0796 \times \$1000 = \$79.60$. Similarly, an increase of the nitric oxide concentration by a single part per 10 million, full-value property tax rate by 1 per $\$10,000$, number of pupil per teacher by 1 and proportion of the lower status people by 1% implies a decrease in the house price by \$9799.60,

\$12.90, \$1043.40 and \$539.20 respectively. To explain the positive coefficients, an increase of the proportion of the non-retail business acres per town by 1%, average number of rooms per dwelling by 1, proportion of the owner occupied units built prior to 1940 by 1% and the transformed variable of proportion of Black people by 1 per 1000 lead to the increase of a house price by \$116.60, \$4222.80, \$29.20, and \$9.70 respectively. A house that bounds the Charles River is worth \$2787.20 more than a house farther away. Similarly, if a house is accessible to the radical highway, its price is \$313.80 more than a house that does not have access.

The R^2 value associated with this model indicates that 71.18% of the total variation in house price is explained by the regressors included in the model (4.1). In the OLS fit, it is essential to verify the assumptions about the distribution of the error term. It is usually assumed that the errors have a normal distribution with the zero mean and a constant variance. Figure 4.3 presents several plots of the residuals and fitted values of the house price in (4.1). The scatterplots of the residuals and the square root of absolute residuals against the fitted house price (Figure 4.3.a and 4.3.b) display a random scatter with no obvious pattern. This indicates independence and constant variance of the error term associated with the model. The normal probability plot is expected to be a straight line if the normality assumption is valid. However, Figure 4.3.d does not appear to satisfy this criterion. A similar trend at the right-end of Figure 4.3.c leads us to look for models with a transformation of the response variable. An attempt of several transformations, such as the natural log and square root of *medv* was made, and there was little difference in the results. Therefore, we move forward with our analysis considering the error term as approximately normally distributed and keep *medv* as the response variable. Figure 4.3.e is called the residual-fitted or $r - f$ plot. Ideally, we hope to see less variability in the plot labelled residuals, since the goal in regression is to explain the variability in the response y with our model. In this case the plot is not ideal since the variability is almost the same in both plots. The graph of Cook's distance against the observation numbers (Figure 4.3.f) along with other plots indicates influential observations. A preventive

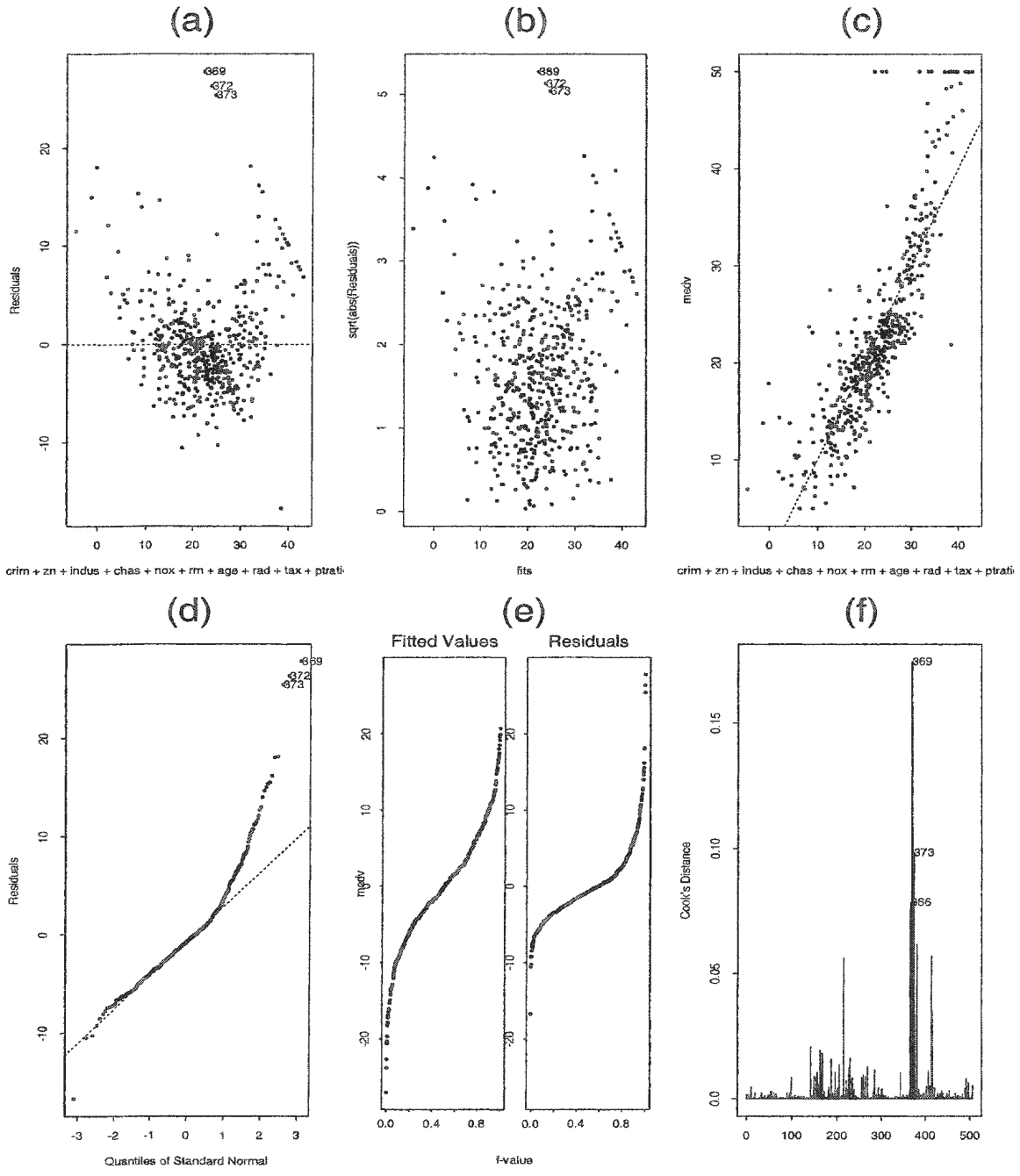

Figure 4.3: Diagnostic plots for the house price data with model (4.1)

measure of omitting those observations is taken prior to moving our analysis to the GWR model fitting and testing procedures.

4.4.2 Variable Selection in Linear Regression

With socio-economic data obtained from a census or large scale survey, consisting of a large number of variables, we often suspect that many of the explanatory variables might have some effect on the response variable. However, it is important to think about how many variables are included in the model, because too many explanatory variables make the interpretation of the model more difficult. In the analysis of the Boston house price data, the stepwise regression and the best subset (or leaps) method are applied for variable selection. An important variable selection note should be cited here — an independent variable is said to be important if the residual sum of squares is significantly reduced when it is added to the model (Leung et al, 2000). The value of the residual sum of squares is used to compute two different quantities: adjusted R^2 and Mallows' C_p. The stepwise regression procedure works with the testing of hypotheses to determine whether an explanatory variable should be included in or excluded from the model, whereas the leaps method applies the strategy of maximizing R^2 or adjusted R^2 and minimizing Mallows' C_p (with C_p close to p).

The stepwise regression combines the forward selection and backward elimination procedures, which to some extent work through the inclusion of the relevant variables and the exclusion of the irrelevant ones. To conduct the analysis, all 12 explanatory variables which have been used for the OLS fit in section 4.4.1 are initially entered. Considering the value of α equal to 0.05 for both entering and removing variables, the procedure is stopped at the fifth step (Table 4.6). The variables being suggested to include as explanatory are *lstat, rm, ptratio, b* and *chas.*

The leaps method produces all possible regression models and summarizes a subset containing the best of the many good models. A compromise between the adjusted R^2 and Mallows' C_p criteria is also considered here to select one of the simpler but reasonably good models (Weisberg, 1985). Table 4.7 presents the results of 12 different

Table 4.6: Stepwise regression results

α to enter = 0.05 α to remove = 0.05
Response is *medv* on 12 predictors, with n = 506.

Step	1	2	3	4	5
Constant	34.554	-1.358	18.567	12.055	11.854
lstat	-0.950	-0.642	-0.572	-0.513	-0.518
t value	-24.53	-14.69	-13.54	-11.53	-11.79
p value	0.000	0.000	0.000	0.000	0.000
rm		5.09	4.52	4.75	4.65
t value		11.46	10.60	11.18	11.08
p value		0.000	0.000	0.000	0.000
ptratio			-0.93	-0.90	-0.86
t value			-7.91	-7.72	-7.43
p value			0.000	0.000	0.000
b				0.0105	0.0101
t value				3.83	3.74
p value				0.000	0.000
chas					3.32
t value					3.68
p value					0.000
S	6.22	5.54	5.23	5.16	5.10
R^2	54.41	63.86	67.86	68.77	69.60
R^2-adjusted	54.32	63.71	67.67	68.53	69.30
C_p	247.9	94.6	30.7	17.7	6.1

Table 4.7: Twelve models with the smallest C_p and largest R^2

p	R^2	R^2_{adj}	C_p	Predictors in the Model
2	54.4	54.3	247.9	lstat
3	63.9	63.7	94.6	rm lstat
4	67.9	67.7	30.7	ptratio rm lstat
5	68.8	68.5	17.7	b ptratio rm lstat
6	69.6	69.3	6.1	chas b ptratio rm lstat
7	69.7	69.3	6.6	age chas b ptratio rm lstat
8	69.9	69.5	5.5	nox age chas b ptratio rm lstat
9	69.9	69.5	6.4	crim nox age chas b ptratio rm lstat
10	70.0	69.4	7.7	indus crim nox age chas b ptratio rm lstat
11	70.0	69.4	9.2	rad indus crim nox age chas b ptratio rm lstat
12	70.0	69.4	11.0	tax rad indus crim nox age chas b ptratio rm lstat
13	70.0	69.3	13.0	zn tax rad indus crim nox age chas b ptratio rm lstat

models obtained from the method of leaps. Every model consists of intercept term, and the values of p in the left column of Table 4.7 indicate the number of coefficients in a model, that is, the number of explanatory variables plus intercept term.

Table 4.7 presents 12 different models taking one from each of p=2, 3, ..., 13. Obviously, the C_p value closest to p is found for $p = 13$, that is, when all 12 variables included in the model, whereas the adjusted R^2 for any p equal to 6 and above is between 69.3 and 69.5. Clearly, the model with $p = 6$ is the same that is prescribed by the stepwise regression procedure, and in comparison to models in Table 4.5, its C_p value and adjusted R^2 are very reasonable. That is, with $p = 6$ we are getting a reasonable value of adjusted R^2, and C_p reasonably close to $p = 6$. Therefore, the model (4.2) with five regressors *lstat, rm, ptratio, b* and *chas* has been selected to proceed with our analysis using GWR.

Table 4.8: OLS regression with five regressors: with and without outliers

Variables	Model with $n = 506$	Model with $n = 502$ (excluding outliers)
Intercept	11.8536**	6.3162*
chas	3.3200**	3.0559**
rm	4.6523**	5.4458**
ptratio	-0.8583**	-0.9010**
b	0.0101**	0.0110**
lstat	-0.5181**	-0.4477**
F-statistic	228.9**	294.0**
R^2	69.6%	74.8%

**significant at 1% level and *significant at 5% level

4.4.3 OLS Regression with the Selected Regressors

The model (4.2) will be used for further analysis including the application of GWR:

$$medv_i = b_0 + b_1 chas + b_2 rm + b_3 ptratio + b_4 b + b_5 lstat + \epsilon_i \tag{4.2}$$

In section 4.1.1, we have described Figure 4.3 to check the assumptions on the error term associated with model (4.1). Similar plots are displayed in Figure 4.4 when OLS estimation is performed with the five regressors in model (4.2). A close look at the respective plots of Figure 4.3 and 4.4 help us to conclude that there is no significant difference in terms of the error assumption in the two models. However, there are four observations in the data set, indicated by the observation numbers 365, 369, 372 and 373, that are possible outliers. We note that all four of these values are from the Inner Core. Since outliers may be influential for parameter as well as model estimation, we have performed the analysis excluding those from the data set. Table 4.8 shows the results of the OLS estimation of the regression model with and without these outliers. The residual plots for the model which excludes the outliers are not included. However, we note that the QQ-plot in this case is similar to those seen in Figure 4.3 and 4.4.

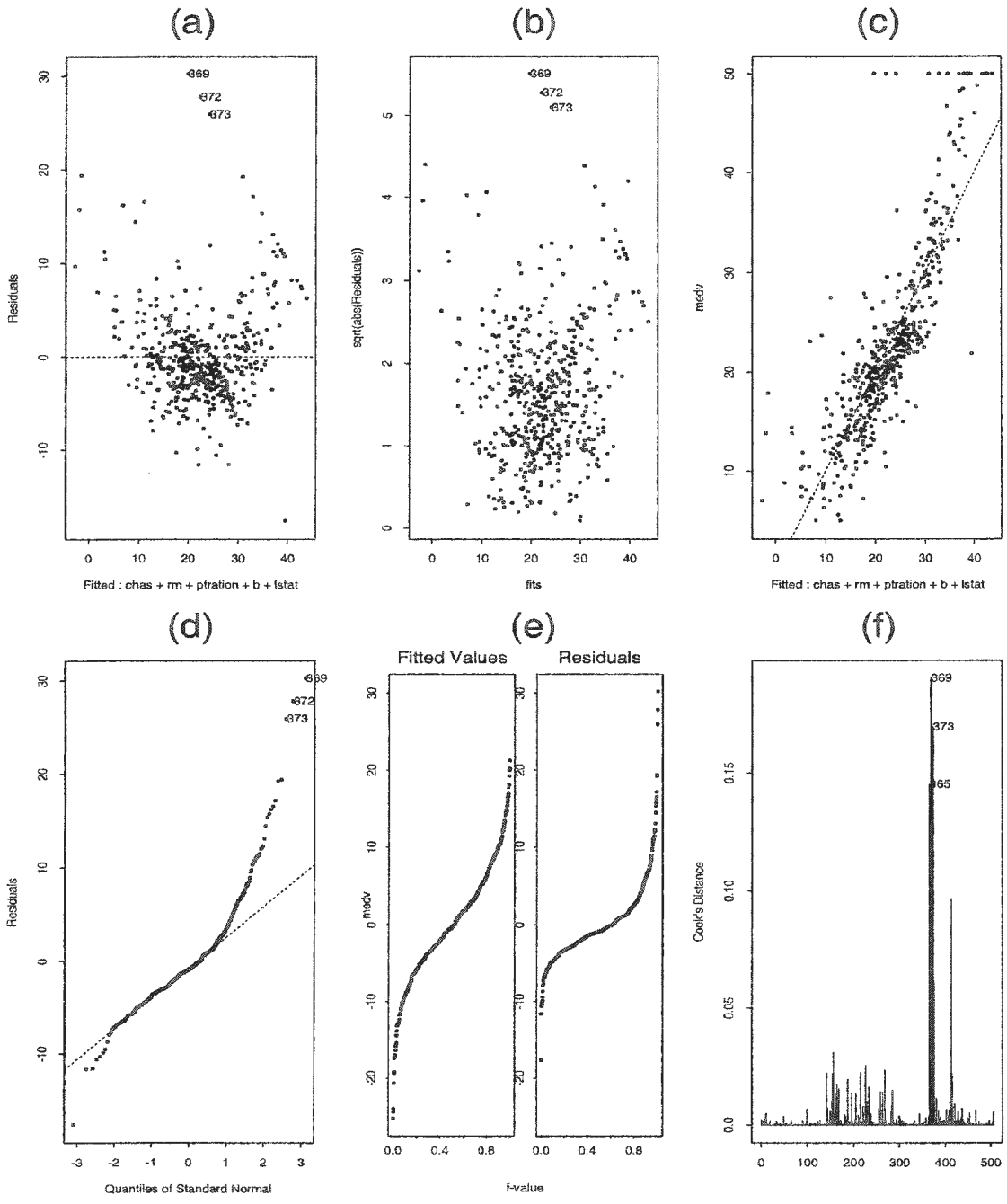

Figure 4.4: Diagnostic plots for the house price data with model (4.2)

In both models the individual effect of each of the five explanatory variables is highly significant on the house price. However, the exclusion of the outliers leads to an increase in R^2 from 69.6% to 74.8%. This is a significant achievement of having a higher percentage of the total variation of the house price explained by the model. Due to excluding those observations, there are some changes also found in the parameter estimates. The houses close to the Charles River exhibit a higher price than those further away. Also the house price tends to increase as the average number of rooms per dwelling increase and the proportion of Black people in towns increase. The values decrease with increasing pupil teacher ratio and the proportion of the lower status people. Since the exclusion of these observations appears reasonable, the analysis using the GWR estimation procedure will be performed excluding those outliers.

4.5 Proxy Variables for Measuring Spatial Variation

These house price data cover a relatively large study area of the metropolitan region of Boston. In the context of housing data, we usually assume that the greater the coverage area, the more variation in housing amenities. Similarly, we assume houses with more amenities would have a higher value than those with fewer amenities. Therefore, for model fitting it is important to consider location variation in the analysis. Along with the other explanatory variables, the effect of location variation will be determined through the application of the GWR estimation procedure and its associated hypothesis tests. However, an attempt of applying some traditional methods is also made to present a comparative advantage of GWR as a way of accounting for spatial non-stationarity.

Instead of determining the direct effect of some factors, we often see the use of proxy variables in model fitting. The utilization of such proxy variables is often found when the inclusion of an explanatory variable seems to be very relevant but not available directly in the data set. In the house price data, we have used one

Table 4.9: OLS regression with proxy variables as predictors

Variables	'weighted distance' as proxy	'sub-regions' as proxy
Intercept	10.051832	7.439438*
chas	2.672253**	2.847142**
rm	5.249001**	5.257867**
ptratio	-0.929174**	-0.865807**
b	0.011992**	0.011330**
lstat	-0.510961**	-0.483921**
dis	-0.390746**	—
sr.cat2: North Shore	—	-0.701148
sr.cat3: North Suburban	—	-0.867990
sr.cat4: Minuteman	—	0.367249
sr.cat5: Metro West	—	1.082042
sr.cat6: South West	—	1.644622
sr.cat7: Three Rivers	—	-1.629566*
sr.cat8: South Shore	—	-2.808603**
F-statistic	252.3**	125.8**
R^2	75.36%	75.54%

**Significant at 1% level, and *significant at 5% level

continuous variable termed as *dis* and another categorical variable called *sr* for two separate models.

The variable *dis* represents the weighted distance to five employment centers in Boston. Certainly, employment can play an important role as an amenity to the dwellers, and hence distance to the workplace may contribute to variation in the house prices. On the other hand, the categorical variable *sr* consists of eight sub-regions as described in Section 4.3. It is expected that the housing amenities in some of the sub-regions may differ from others; hence the variable *sr* may have some effect in exhibiting variation in housing price. Therefore, the analysis of the model (4.2) is extended to include these two variables separately in two different models. The results of the OLS estimates are presented in Table 4.9.

In both models, the OLS estimates of the parameters corresponding to each of

the five explanatory variables are highly significant. As observed in the two other OLS models in subsection 4.3.3, the effect of the Charles River, number of rooms per dwelling and proportion of Black people in towns is positive on the house price, whereas that of pupil teacher ratio and proportion of the lower status people is negative. Moreover, the inclusion of *dis* as proxy variable has a significant negative effect on the house price. That is, the houses that are close to employment centers have a higher price than those that are farther away. On the other hand, the inclusion of the categorical variable *sr* presents a comparative effect of seven other sub-regions with respect to the base category of the Inner Core sub-region. In the third column of Table 4.9, we see that the houses in the South Shore and Three Rivers have significantly lower price in the housing market than those in the Inner Core sub-region, given the other predictors in the regression model. We note that this differs from the summary statistics in Table 4.3, which indicate that the Inner Core has the lowest mean price. However, those sample means fail to account for the effect of the predictors that we include in our regression model. Although we observe the lower prices in the North Shore and North Suburban, and higher prices of the Minuteman, Metro West and South West areas as compared to the Inner Core, the $p-$values corresponding to these sub-regions present no evidence of statistical significance. In the use of proxy variables, we also see how the R^2 value changes. Comparing the R^2 values of both models in Table 4.9 with respect to the model without outliers in Table 4.8, we see that the inclusion of the two separate variables *dis* and *sr* as contributing location factors provides a model with slightly higher predictive skill. However, this small increase may be caused by the inclusion of an additional regressor in the models, rather than explaining it as evidence of significant contribution of the particular variable to house price.

Since the determination of spatial variability in the parameters will be the issue of the next section, we put special attention on the variables *dis* and *sr* as these are somehow representing spatial location. The smaller $p-$value corresponding to *dis*

Table 4.10: Partial F-test results. SSE = sum of squares residual, SSE_{df} = residual degrees of freedom, NDF = numerator degrees of freedom, DDF = denominator degrees of freedom, and for each model $n = 502$ used.

			Partial F-test			$p-value$
Models	SSE	SSE_{df}	F-value	NDF	DDF	
Reduced Model (4.2)	10201.1	496	–	–	–	–
Full Model (4.4)	9892.9	489	2.18	7	489	0.035

suggest to include this in the model. Therefore, the fitted model is of the form:

$$medv_i = b_0 + b_1 chas + b_2 rm + b_3 ptratio + b_4 b + b_5 lstat + b_6 dis + \epsilon_i \qquad (4.3)$$

However, the large $p-$values corresponding to the five sub-regions of sr (Table 4.9) leads us to perform a partial F-test which is also called full-versus-reduced model test (Neter et al 1985, p. 95). To test whether the variable sr should be included in the model, we consider the model:

$$\begin{aligned} medv_i &= b_0 + b_1 chas + b_2 rm + b_3 ptratio + b_4 b + b_5 lstat + b_6 sr{:}cat2 + b_7 sr{:}cat3 \\ &\quad + b_8 sr{:}cat4 + b_9 sr{:}cat5 + b_{10} sr{:}cat6 + b_{11} sr{:}cat7 + b_{12} sr{:}cat8 + \epsilon_i \end{aligned} \qquad (4.4)$$

as full or unrestricted, and the model (4.2) as the reduced model. Table 4.10 presents the value of the F-statistic = 2.18, and the results associated with the test. The $p-$value = 0.035 indicates sufficient evidence to include sr in the model.

4.6 Fitting GWR Model

Application of the GWR model fitting approach and testing techniques on the house price data is one of the key interests of this practicum. After a detail theoretical explanation (Chapter 2) and simulation results (Chapter 3), its application to a real data set helps to strengthen the methodological grounds, and to find further problems that may associated with some unknown characteristics in the data set. There are

two other variables available in the data set. These represent the measurement of the latitude and longitude of the sampled houses, and indicate the location of a house. To analyze the GWR methods, these two new variables are not used as explanatory variables, but are used to form the GWR weight function by computing distance between each location.

At the stage of analysis, the cross validation approach is used to determine the value of the bandwidth β. Using equation (2.9) this produces the minimum $CVSS$ score at β. To obtain the estimates of the GWR coefficients we have used the Gaussian weighting function (2.4) with $\beta = 0.40$. However, we did find some computational difficulties during the GWR estimation with the values of this weight function. In the previous section, we already omitted the four observations which had been identified as outliers. When the analysis determining $\hat{\mathbf{b}}_i$ in (2.6) was done with the remaining 502 observations, there were problems with inverting the $\mathbf{X}^T\mathbf{W}_i\mathbf{X}$ matrix. Some exploratory work identified ten more observations causing trouble such that the corresponding $\mathbf{W}_i$ matrix was found to be non-full rank. The exclusion of those observations finally set the house price data with sample size equal to 492 for the GWR analysis. We note that one of these observations was from the North Shore, and the remainder from the South Shore.

4.6.1 Results of F-Test and Randomization Test

The analysis for both testing procedures are performed with the five selected regressors included in the model. The interest of the F-test and randomization test is to test whether there is any evidence of spatial non-stationarity, where the overall significance of the five explanatory variables is to be tested by the former one and the significance of individual variables over the study area is to be inferred by the latter one. Through the F test we would like to test the null hypothesis that none of the five coefficients b_j varies over the study regions versus the alternative that at least one of the five b_j is not constant for all locations in the Boston SMSA. On the other hand, the randomization test examines whether the values of individual coefficients

Table 4.11: p-values of the Randomization test by several choice of scramblings

	Number of scramblings		
Coefficients	200	500	1000
b_0	1.000	0.996	0.997
b_1	0.890	0.896	0.911
b_2	1.000	0.994	0.997
b_3	0.880	0.902	0.912
b_4	0.950	0.960	0.969
b_5	0.980	0.964	0.979

change over the study region.

As stated above, we have excluded 14 observations from the data set, and the remaining 492 are considered for the GWR analysis. The value of F-statistic obtained from the analysis is 2.91 with the numerator and denominator degrees of freedom 173.86 and 389.76 respectively. The p-value of the test is approximately zero, and hence we conclude that there is a strong evidence of observing significant spatial non-stationarity in at least one of the five parameters over the Boston metropolitan area. In section 4.5, we have used *dis* and *sr* as proxy variables, and reached a conclusion that as other explanatory variables, location of houses has a significant effect on house price. The inference that we made from the GWR-based F-test supports this conclusion, and in addition it says the effect of at least one of the explanatory variables included in the model varies spatially over the study area.

For the randomization test, the number of explanatory variables included in the model, size of the sample considered for the analysis, the weight function computation procedure including the choice of β values are all the same as that of F test procedure. Initially, the test is conducted with 200 scramblings. Later on, it is increased to 500 and then 1000. The results are presented in Table 4.11.

The large $p-$values corresponding to the coefficients in Table 4.11 are not what we expect and do not match our findings obtained from the F-test. Increasing the number of scramblings from 200 to 500 and then 1000 produces no significant change

in the $p-$values. Similar results were also observed in Chapter 3 when its application to the simulation studies was performed for a multi-predictor model. To verify these results, we have attempted to visualize the GWR estimates of the parameters in this model. For every individual coefficient this has produced 492 estimates considering each household located at a point on the study region.

Figure 4.5 presents the image plots of the estimates on the map of the Boston metropolitan area. For an individual parameter estimates, the lighter the image, the higher effect of the corresponding explanatory variable on house price. Conversely, the darker the image, the lower its effect on house price. From Figure 4.5.a and 4.5.b, the effects of the Charles River and average number of rooms are lower in most areas of the Boston metropolitan city, but is quite strong in a few lightly shaded areas in the left of these plots. If we consider Figure 4.5.c and 4.5.e, the effects of pupil-teacher ratio and proportion of lower status people on house price are similar in most regions, except for those at the left of the plots. As an individual predictor, the effect of the proportion of Black population in the communities does not appear to vary over space. Therefore, as the F-test suggests, these image plots support that the relationship of some of the explanatory variables with house price varies over the location of the study region. All these findings strengthen our conclusions that the randomization test does not work as well for detecting spatial variation in relationship between an individual explanatory variable and response. Some possible explanations of such behaviour with this test are outlined in Chapter 5.

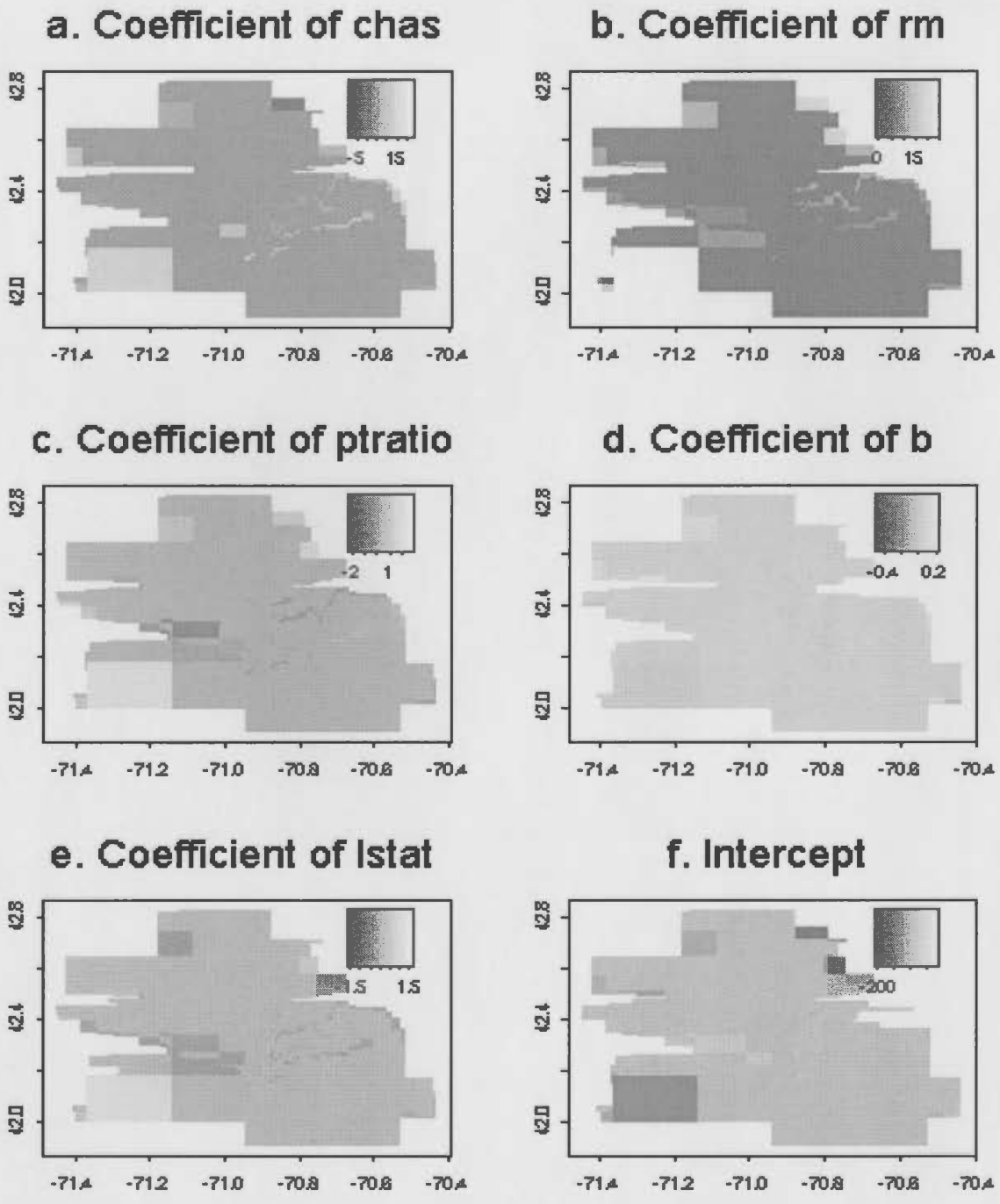

Figure 4.5: Images of parameter estimates on the Boston Metropolitan Area Map.

Chapter 5

Conclusions

In this practicum, we have considered the theoretical aspects of GWR, the performance of the estimation method and inference in simulation studies, and its use in analyzing a socio-economic data set. With the aim of assessing spatial non-stationarity in the parameter estimation, we combined the concept of regression analysis and spatial variation in our analysis. Though the F test and randomization test were chosen as the major area of concentration, we have performed several techniques of regression analysis throughout the study. The key findings are summarized below.

The performance of two different testing methods, presented in section 3.3, can be treated as one of the most significant outcomes of this study. For both the single and multi-predictor models, the simulation studies show that the F test can correctly reflect the spatial non-stationarity that may exist in the relationship among the explanatory and response variables. Also, in compare to the randomization procedure, it appears that the F test has advantages in terms of higher power and is less computationally intensive.

However, in determining the spatial variability of each individual parameter in the model, the application of the Monte Carlo technique exhibits some erratic behavior in our simulation results. When the true state of the data has no spatial variation in parameter values, the randomization test is of the appropriate size. Hence for the stationary case in both the single and multi-predictor models, this test correctly

identifies no spatial variation in the relationships. Also, if the data is exhibits spatial non-stationarity and we are aiming to fit a single predictor model, the test has adequate power. A contradictory performance is observed, when it is applied to a data set with spatial non-stationarity, and the purpose is to fit a multi-predictor model. One possible explanation of this problem is that the spatial variation in parameters could be removed by the addition of further explanatory variables (Brunsdon et al, 1998). According to Leung et al (2000), the validity of this randomization distribution is limited to the given data set. In this line of thinking, we conclude that the $p-$value associated with the randomization test may not work well to identify spatial non-stationarity. As explained by Brunsdon et al (1998), the confidence intervals might be more helpful than $p-$values as they convey an idea not only of the magnitude of parameters but also how precisely this has been determined. However, in our analysis only the Monte Carlo technique is applied to carry out this randomization test, rather than using a distribution of v_j in (2.12). Therefore, an attempt of cross-checking with a confidence interval may be considered for further work followed by a distributional form of v_j.

As shown in Chapter 4, several steps are recommended prior to applying GWR in a data set to assess the spatial non-stationarity in relationships. Considering the theoretical basis of GWR, it can be hoped that there are a number of research areas where the GWR method might be a valuable statistical technique in spatial data analysis.

As a future direction of studies, it may be useful to allow the GWR model to have a mix of variables with and without spatial variations that affect the response. Also, more simulation studies on the inference procedures may be needed. Leung et al (2000) present some results on a single predictor model with several patterns of spatial variability. However, their results were based on either a single, or very small number of simulations. This contrasts with our results with using a rectangular grid, as we used 500 simulated data sets in each of our studies.

Bibliography

[1] Anselin, L. & Griffith, D. A. (1988). Do spatial effects really matter in regression analysis? *Papers of the Regional Science Association*, **65**, 11-34.

[2] Anselin, L. (1993). SpaceStat: A program for the statistical analysis of spatial data. *Technical Report* 93106-4060, Department of Geography, University of California at Santa Barbara: NCGIA.

[3] Arbia, G. (1989). *Spatial data configuration in statistical analysis of regional economic and related problems.* Dordrecht: Kluwer.

[4] Bailey, T. C. & Gatrell, A. C. (1995). *Interactive Spatial Data Analysis.* Essex: Longman Scientific and Technical.

[5] Belsley, D. A., Kuh, E. & Welch, R. E. (1980). *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity*, Wiley, New York.

[6] Besag, J. & Newell, J. (1991). The detection of clusters in rare diseases. *Journal of the Royal Statistical Society A*, **154**, 143-155.

[7] Bowman, A. W. (1984). An alternative method of cross validation for the smoothing of density estimates. *Biometrika*, **71**, 353-360.

[8] Breiman, L., Friedman, J., Olshen, R. & Stone, C. J. (1993). *Classification and Regression Trees.* Chapman and Hall, New York.

[9] Breiman, L. & Friedman, J. (1985). Estimating optimal transformations for multiple regression and correlation. *Journal of the Americal Statistical Association*, **80**, 580-619.

[10] Brunsdon, C., Fotheringham, S. & Charlton, M. (1996). Geographically weighted regression: A method for exploring spatial non-stationarity. *Geographical analysis*, **28**, 281-289.

[11] Brunsdon, C., Fotheringham, S. & Charlton, M. (1998). Geographically weighted regression- modelling spatial non-stationarity. *The Statistician*, **47**, 431-443.

[12] Brunsdon, C., Fotheringham, S. & Charlton, M. (1999). Some notes on parametric significance tests for geographically weighted regression. *Journal of Regional Science*, **39**, 497-524.

[13] Brunsdon, C., Fotheringham, S. & Charlton, M. (2000). Geographically weighted regression as a statistical model. *Unpublished manuscript.*

[14] Cleveland, W. S. (1979). Robust locally weighted regression and smoothing scatterplots. *Journal of the Americal Statistical Association*, **83**, 596-610.

[15] Fotheringham, A. S., Brunsdon, C. & Charlton, M. (1996). The geography of parameter space: An investigation into spatial non-stationarity. *International Journal of Geographical Information Systems*, **10**, 605-627.

[16] Fotheringham, A. S., Charlton, M. E. & Brunsdon, C. (1998). Geographically weighted regression: a natural evaluation of the expansion method for spatial data analysis. *Environment and Planning A*, **30**, 1905-1927.

[17] Getis, A. & Ord, J. K. (1992). The analysis of spatial association by use of distance statistics. *Geographical Analysis*, **24**, 189-206.

[18] Golub, G., Heath, M. & Wahba, G. (1979). Generalized cross validation as a method for choosing a good ridge parameter. *Technometrics*, **21**, 215-223.

[19] Harrison, D. & Rubinfeld, D. L. (1978). Hedonic housing price and the demand for clean air. *Journal of Environmental Economics and Management*, **5**, 81-102.

[20] Hastie, T. & Tibshirani, R. (1990). *Generalized Additive Models*, Chapman and Hall, London.

[21] Hope, A. C. A. (1968). A simplified Monte-Carlo significance test procedure. *Journal of the Royal Statistical Society B*, **30**, 582-598.

[22] Krasker, W. S., Kuh, E. & Welsch, R. E. (1983). Estimation for dirty data and flawed models. *Handbook of Econometrics*, North Holland, Amsterdam, 1, 651-698.

[23] Lange, N. & Ryan, L. (1989). Assessing normality in random effect models. *Annals of Statistics*, **17**, 624-642.

[24] Leung, Y., Mei, C. & Zhang, W. A. (2000). Statistical tests for spatial nonstationarity based on the geographically weighted regression model. *Environment and Planning*, **32**, 9-32.

[25] Li, K. (1986). Asymptotic optimality of C_L and generalized cross-validation in ridge regression with application to spline smoothing. *Annals of Statistics*, **14**, 1101-1112.

[26] Neter, J., Wasserman, W. & Kutner, M. H. (1985). *Applied Linear Statistical Models*, Second edition, Irwin, Illinois, 2, 23-24.

[27] Odland, J. (1988). *Spatial Auto-correlation.* London Stage Publications.

[28] Pace, R. K. (1993). Nonparametric methods with application to hedonic models. *Journal of Real Estate Finance and Economics*, **7**, 185-204.

[29] Paez, A., Uchida, T. & Miyamoto, K. (2002). A general framework for estimation and inference of geographically weighted regression models: 2. Spatial association and model specification tests. *Environment and Planning A*, **34**, 883-904.

[30] Silverman, B. W. (1986). *Density Estimation for Statistics and Data Analysis.* Chapman and Hall, London.

[31] Simonoff, J. S. (1996). *Smoothing Methods in Statistics*, Springer, New York.

[32] Sneddon, G. (1999). Smoothing in an underdetermined linear model with random explanatory variables. *Canadian Journal of Statistics*, **27**, 63-80.

[33] Subramanian, S. & Carson, R. T. (1988). Robust regression in the presence of heteroskedasticity. *Advance Econometrics*, **7**, 85-138.

[34] Weisberg, S. (1985). *Applied Linear Regression.* Wiley Series, Second edition, 196-225.