AN EXTENSION OF GEOGRAPHICALLY WEIGHTED REGRESSION WITH FLEXIBLE BANDWIDTHS

Wenbai Yang

A Thesis Submitted for the Degree of PhD
at the
University of St Andrews

2014

An Extension of Geographically Weighted Regression with Flexible Bandwidths

Wenbai Yang

Supervisor: Professor A. Stewart Fotheringham

This thesis is submitted in fulfilment for the degree of PhD
at the
University of St Andrews

May, 2014

1. Candidate's declarations:

I, Wenbai Yang, hereby certify that this thesis, which is approximately 37,370 words in length, has been written by me, that it is the record of work carried out by me and that it has not been submitted in any previous application for a higher degree.

I was admitted as a research student in October, 2011 and as a candidate for the degree of PhD in October, 2011; the higher study for which this is a record was carried out in the University of St Andrews between 2011 and 2014.

Date …… signature of candidate ……

2. Supervisor's declaration:

I hereby certify that the candidate has fulfilled the conditions of the Resolution and Regulations appropriate for the degree of PhD in the University of St Andrews and that the candidate is qualified to submit this thesis in application for that degree.

Date …… signature of supervisor ……

3. Permission for electronic publication:

In submitting this thesis to the University of St Andrews I understand that I am giving permission for it to be made available for use in accordance with the regulations of the University Library for the time being in force, subject to any copyright vested in the work not being affected thereby. I also understand that the title and the abstract will be published, and that a copy of the work may be made and supplied to any bona fide library or research worker, that my thesis will be electronically accessible for personal or research use unless exempt by award of an embargo as requested below, and that the library has the right to migrate my thesis into new electronic forms as required to ensure continued access to the thesis. I have obtained any third-party copyright permissions that may be required in order to allow such access and migration, or have requested the appropriate embargo below.

The following is an agreed request by candidate and supervisor regarding the electronic publication of this thesis:

Embargo on both all of printed copy and electronic copy for the same fixed period of two years on the following ground: publication would preclude future publication.

Date …… signature of candidate …… signature of supervisor ……

Abstract

Various statistical methods have been developed for local spatial analysis. Among them Geographically Weighted Regression (GWR) is a simple yet powerful method to explore spatially varying relationships between variables. This thesis examines how GWR can be extended to investigate spatially varying relationships at various geographical scales within one model.

GWR assumes that observations near to a regression location have more influence on the estimation of local regression coefficients than do observations farther away. A single bandwidth is employed in basic GWR to control the rate of distance-decay in this influence. The magnitude of the bandwidth affects the scale of variation in the estimated regression coefficients and thus usefully reflects the appropriate spatial scale at which the processes being modelled operate. A small bandwidth suggests the processes operate over a local spatial scale, whilst a large bandwidth indicates a more regional process.

In practice, a single bandwidth as in basic GWR may not be sufficient to reflect the potentially complex spatial variations in relationships between variables in a multivariate spatial model. Therefore, in order to estimate coefficient surfaces that may vary at different spatial scales for different variables, Flexible Bandwidth GWR (FBGWR) is proposed to allow different bandwidths to be individually specified for each independent variable in a regression framework. An algorithm based on back-fitting is developed to calibrate the FBGWR model.

The performance of FBGWR is investigated with simulated datasets where coefficients are predefined at various levels of non-stationarity across space. A case study is then carried out on data relating to the Irish Famine to demonstrate the application of FBGWR to real-world processes. The results suggest that FBGWR can distinguish various scales of non-stationarity in spatial processes and provide an improved model over basic GWR. FBGWR therefore represents a useful development in the modelling of spatially varying processes.

Dedicated to my parents for their never ending love and support.

谨以此文献给我敬爱的父母，感谢他们无尽的关爱与支持。

Acknowledgements

I have been extremely fortunate that my PhD spanned two universities: the National University of Ireland Maynooth (NUIM) and the University of St Andrews. I owe my deepest gratitude to so many people without whose guidance, advice, encouragement and support I could not have completed the work presented in this thesis.

First and foremost, I would like to express my sincere thanks to my supervisor, Professor A. Stewart Fotheringham for the knowledge, guidance and encouragement he provided throughout my PhD. It has been a great honour to have him as my supervisor. I am extremely grateful to Dr. Paul Harris, whose generous support, advice and encouragement has greatly contributed to my PhD experience. Thanks also to Professor Chris Brunsdon for his insightful suggestions during the early stages of my research.

I wish to thank Dr. Urška Demšar for her help during the transfer from Maynooth to St. Andrews and for her encouragement throughout my thesis. Thanks to Panagiota Spanou for taking care of everything I needed help with, my officemate Sila for all of the insightful discussions and encouragement that helped me through the hard times of my PhD and David in the next office for keeping me company through those late working nights.

I am grateful to everyone in the National Centre for Geocomputation, NUIM for making my time there a memorable experience. Special thanks to Dr. Jan Rigby for her advice on both a personal and academic level when acting as my PhD programme coordinator. Thanks to all of the administrative staff, especially Ann-Marie Burke and Melina Lawless for making my daily life much easier. Thanks to all of my fellow PhD students especially Conor, Binbin and Ishwari for their friendship. Thanks to all academic staff and post-docs for providing me with an inspiring work environment.

I also wish to extend my thanks to Dr. Chris Lloyd from University of Liverpool and Dr. Jed Long for examining my thesis.

Finally, I am forever indebted to my family, especially my parents and my husband Feng for their never ending love and support. I would particularly like to thank my baby boy Andi who kept me happy and energetic throughout my pregnancy. Writing this thesis with Andi in my tummy has been the most wonderful time in my life.

Funding

Research presented in this paper was funded by a Strategic Research Cluster grant (07/SRC/I1168) by Science Foundation Ireland under the National Development Plan. Additional funding was provided by the University of St Andrews as part of Professor A. Stewart Fotheringham's start up grant. Partial funding by the Chief Scientist Office in collocation with the University of Dundee through the award Geographical Differences in the Uptake of Colorectal and Breast Cancer Screening in Scotland is also gratefully acknowledged.

Contents

List of Figures

List of Tables

Chapter 1 Introduction

1.1 General overview

Constructing mathematical models of processes is a common theme in analytical research in a wide range of disciplines. “Researchers search for variables to identify various dimensions of phenomena and for relationships among the variables to interpret or change the real world.” (Casetti, 1972, P.82). Traditional analytical methods, however, tend to assume constant relationships among variables; that is, models with constant parameters are constructed to describe the relationships between variables. This is usually achievable and acceptable in physical sciences, where the investigated phenomena are determined by certain natural laws, for example, Newton’s law of universal gravitation involves a constant parameter, the gravitational constant. In social and environmental sciences, however, the relationships between variables may not be constant. For example, the relationship between elevation and precipitation may change according to complex geographical phenomena. Traditional global models can mask this non-stationarity in relationships (Fotheringham, 1997, Brunsdon et al., 1999a, Brunsdon et al., 1999b, Farber and Yeates, 2006). Models that allow parameters to vary across geographical space, over time, or according to other contexts are thus useful. Typical forms of such models are regression models that have spatially varying coefficients. This thesis focuses on a widely applied model of this type, namely Geographically Weighted Regression (GWR).

GWR (Brunsdon et al., 1996, Fotheringham et al., 1996, Fotheringham et al., 2002) is designed to model spatially varying relationships between variables. A local regression is performed at different geographical locations so that at each location a set of localised parameters is estimated, with each estimate representing a spatially local relationship. In order to achieve this, the basic assumption of GWR is that observations near to a regression point have more influence in the estimation of regression coefficients for that point than do observations farther away so that data are weighted according to their geographical proximity to the regression point with data from near observations weighted more than data from far observations. This geographical weighting is realized through a distance-decay weighting function with the bandwidth of the weighting function controlling the rate of distance-decay. The

larger a bandwidth is, the more slowly the weights decay and, as a result, the estimated parameters will exhibit more stationarity over space. Smaller bandwidths allow more rapid distance-decay in weighting and result in more localized parameter estimates. In other words, smaller bandwidths allow the analysis of spatially varying parameters at finer spatial scales. The magnitude of the bandwidth therefore indicates the spatial scale at which the local processes being examined operate.

A standard or classical GWR model defines a single bandwidth for the distance-decay function for all the relationships defined in the model. In a multivariate analysis, the relationships between the dependent variable and the various independent variables in the model are examined through this uniform bandwidth. In practice, however, it might well be the case that the processes being examined operate at different spatial scales. For example, consider the determinants of the price of a house. Some attributes, such as the presence of a nearby park, may have a similar effect on house prices throughout the city whereas others, such as the floor area of the house, might have an impact on prices that varies spatially. If a GWR model is constructed to analyse the relationships between house price and these two factors, floor area and proximity to parkland, the uniform bandwidth of the model can only reflect the average of the two spatial scales and detailed information about the complex nature of the spatial variations in relationships will be lost. A model that is capable of reflecting the different spatial scales at which different processes operate would thus be very useful. The purpose of this thesis is to develop such a modelling framework, termed here – Flexible Bandwidth GWR (FBGWR). To accomplish this, the weighting function for the coefficient of each independent variable in a GWR model needs to have its own bandwidth. This is not achievable in any current versions of GWR, although a special case can be considered in the existing mixed GWR (also called semi-parametric GWR, partial linear GWR) (Brunsdon et al., 1999b, Fotheringham et al., 2002, Mei et al., 2004, Nakaya et al., 2005a), where the relationship between the dependent variable and each independent variable is allowed to be either globally constant or locally varying. This thesis extends the mixed GWR formulation and concentrates on a full realization of GWR with predictor-specific bandwidths, namely FBGWR.

1.2 Research objectives

The main aim of the research is to develop FBGWR in which flexible bandwidths can be specified for different independent variables and for the intercept in a GWR model. Instead of a uniform bandwidth employed in basic GWR models, a vector of bandwidths will be defined in FBGWR. The key theoretical and practical issues of FBGWR are: (1) how to calibrate the model; (2) how to choose appropriate bandwidths to optimize model fit.

Since the FBGWR model may have a different weighting scheme for each independent variable, the Weighted Least Squares technique that is employed in basic GWR will no long apply and a new algorithm is required to calibrate a FBGWR model. This is the first objective of the research, to develop a model calibration strategy.

Given that the determination of a set of optimal bandwidths is the key point of FBGWR, only when appropriate bandwidths are specified in the model will the model be useful. The second objective of the research is then to develop a bandwidth selection strategy.

1.3 Thesis structure

The rest of this thesis is organized as follows: Chapter 2 provides a general overview of spatial analysis and spatial regression models. The developments of spatial analysis in various disciplines including geography, spatial econometrics and statistics are sketched out. Regression models in non-spatial circumstances as well as in spatial ones are briefly introduced. These techniques constitute the background and context for the development of GWR. Chapter 3 focuses on the development of GWR, describing the basic methodology of GWR, the applications of GWR and various extensions of GWR. Chapter 4 presents the methodology of FBGWR. The back-fitting algorithm is described, various strategies for bandwidth selection are discussed and an algorithm for calibrating FBGWR is proposed.

Chapter 5 investigates various aspects of the proposed FBGWR method through simulation experiments. Synthetic datasets are designed to test the algorithm under different configurations of spatial non-stationarity. The FBGWR algorithm is applied to several circumstances where various levels of heterogeneities are embedded in the

coefficients to be estimated and the results demonstrate the strength as well as the weakness of FBGWR.

Chapter 6 demonstrates the application of FBGWR through a case study on data on population loss during the Irish Famine. The procedure of how a FBGWR model can be constructed on the dataset is described and the results discussed. Chapter 7 concludes the thesis. It summarizes the research achievements, points out the limitations of current work and suggests possible directions for future work.

Chapter 2 Spatial analysis and spatial regression models

2.1 Spatial analysis in general

Spatial analysis, also called spatial data analysis, refers to a subset of methods for analysing spatial data, i.e. data that are referenced to a two or three dimensional space. To address the important role of space in the analysis, Goodchild and Longley (1999, P.569) further defined spatial analysis as "the subset of analytic techniques whose results depend on spatial frame, or will change if the frame changes, or if objects are repositioned within it".

According to Fischer and Getis (1997), modern spatial analysis originated from the development of quantitative geography and regional science in the late 1950s. Subsequently, spatial analysis has gained growing acceptance in the mainstream of social sciences (Goodchild et al., 2000). Examples can be seen in the new economic geography (Krugman, 1991, Martin, 1999), as well as in developments in sociology and political science (O'Loughlin et al., 1998). Goodchild et al. (2000) proposed the term "spatially integrated social science", referring to the application of spatial analysis tools in the social sciences to gain new scientific insights. Berry et al. (2008) also put forward a similar term "spatially integrated social environmental science".

Spatial analysis is not just non-spatial analysis with a space or distance component added into the model (Fotheringham, 1993). The demand for spatial analysis has been stimulated by theoretical questions raised in various fields of social sciences, as well as by technical developments in spatial data manipulation and the accessibility of increasing volumes of geo-referenced data. The development of geographic information systems (GIS) has helped to popularize spatial analytical practice through its expanding ability to manipulate and visualise spatial data, as well as the interaction capabilities it supplies to the user (for example, see Fotheringham and Wong (1991)).

One of the main approaches of spatial analysis is to use principles of statistics in a spatial context that are found in spatial statistics, geostatistics and spatial econometrics, allowing the exploration of data and processes from a spatial perspective to search for spatial patterns, correlations, outliers and residuals. Moreover, as stated by Fischer and Getis (1997, P.1), spatial analysis is more than

spatial statistics and data analysis: it "goes far beyond data sampling, data manipulation, exploratory and confirmatory spatial data analysis, into areas of spatial modelling".

Today, methods of spatial analysis have been developed and accumulated across a diverse range of disciplines, including not only geography and other earth sciences, but also economics, physics, engineering, mathematics and statistics. The following subsections will briefly outline the main disciplines that have played and continue to play important roles in the development of spatial analysis.

2.1.1 Quantitative geography

The importance of space and place has always been recognized in the fields of geography and regional science, while the practise of applying statistical methods to solve problems in geography has flourished since the quantitative revolution in geography (Berry and Marble, 1968) in the middle 1960s, which launched a long-term collaboration between geographers and spatial statisticians (Cliff and Ord, 2009). Due to the weakness of classical statistics in solving specific problems in geography, the demand for new statistical methods was put forward by quantitative geographers (Haining, 2009). To demonstrate this, classical statistics assumes that data are independent and identically distributed, but the assumption of independence is usually violated in geographical analysis because data values closer in geographical space to each other tend to be more similar than those farther apart (Tobler, 1970). Another assumption that is often violated in the analysis of spatial data is the stationarity of relationships between variables. It is quite conceivable, for example, that the processes generating spatial data vary over space. These violations of assumptions in spatial data analysis affect the use of classical statistics in geography (see Haining, 2009 for details). New statistical theories and techniques are thus required for use in spatial data analysis; these requirements have stimulated the development of spatial analysis in geography in both application and methodology.

Among other developments in spatial analysis, Fotheringham (1997) points out a movement away from global analysis, which assumes constant relationships between variables over space, to local analysis where the emphasis is to identify spatial variations in relationships. The output of local analysis, generally spatially varying parameters, can be mapped in a geographical context with GIS, providing more insight into the spatial processes of interest (Fotheringham et al., 2002).

2.1.2 Spatial econometrics

Spatial econometrics is a subfield of econometrics that deals with spatial interaction and spatial structure in spatial regression models for cross-sectional and panel data (Paelinck and Klaassen, 1979, Anselin, 1988) . As Anselin (2010) notes, the field of spatial econometrics stems not only from work towards spatial methods in regional science and regional and urban economics from the late 1960s, but also from the quantitative revolution in geography of the same time period. Spatial econometrics has eventually gained growing attention and wide application and developed into the mainstream of applied econometrics and social sciences methodology. The scope of data types has also been extended from cross-sectional data to the space-time domain. The field of spatial econometrics reached maturity in the early twenty-first century, with the general acceptance of both spatial statistics and spatial econometrics as mainstream methodologies (Anselin, 2010).

2.1.3 Statistics

In statistics, attention to spatial pattern and spatial dependency dates back to the 1950s (Whittle, 1954) and what would become standard methods were established in the following decades, for example, by Ripley (1981). Standard statistical analysis tools of linear and generalized linear regression models, which have been widely used in almost every scientific discipline, cannot be properly applied to geo-referenced data due to the violations of the standard assumptions of independence, identically distribution and stationarity, as described in 2.1.1.

Among the various disciplines that have branched from classic statistics, geostatistics is one that has direct relevance to spatial analysis. Based on the theory of regionalised variables, geostatistics considers a phenomenon at unknown locations as a set of correlated random variables to model the uncertainty associated with spatial estimation and simulation (Chilès and Delfiner, 1999). According to Haining (2009), the first use of the term geostatistics was in a geographical context by Hart (1954), while the current use of geostatistics stems from Matheron's (Matheron, 1963) theory for predicting properties in geographical space based on Krige's ideas for prediction using neighbouring samples (Krige, 1951). Geostatistics is now applied in diverse disciplines including geography, geology, geochemistry and soil science. Although developed in different disciplines, the aforementioned fields share some common

characteristics such as the spatial features of data, the focus on spatial patterns of residuals from regression models and the spatial aspects to be accounted for in spatial modelling.

2.2 Spatial issues in spatial modelling

The two key issues with spatial modelling, as pointed out by Anselin (1988), are spatial dependence and spatial heterogeneity, also expressed as spatial autocorrelation and spatial non-stationarity respectively.

2.2.1 Spatial dependence

Spatial dependence refers to the correlation among the values of a single variable measured at different locations in geographical space (Cliff and Ord, 1969), the main evidence of spatial dependence is spatial autocorrelation including positive and negative spatial autocorrelation. Positive spatial autocorrelation produces clustering of similar values, while negative spatial autocorrelation reflects dissimilar values at adjacent locations. In 1969, Cliff and Ord published "The Problem of Spatial Autocorrelation"(Cliff and Ord, 1969) and this article popularized the concept of spatial autocorrelation, raised the need to allow for spatial autocorrelation in model building and launched a revolution in the statistical analysis of spatial data (Griffith, 2009). Spatial autocorrelation may be intrinsic in real-world phenomena, as Tobler's first law of geography (Tobler, 1970) states: "everything is related to everything else, but near things are more related than distant things." Spatial autocorrelation can also result from missing exogenous factors or inappropriate spatial aggregation of nonhomogeneous units (Anselin, 1988, Tiefelsdorf and Griffith, 2007).

The occurrence of spatial autocorrelation violates the assumption of independence of observations in classical data analysis, making standard statistical analysis such as analysis of variance and ordinary least squares inappropriate. The effect of spatial autocorrelation is twofold: it can be a "problem" for geographers because if it needs to be detected and explained if found. For example, spatial autocorrelation in regression residuals indicates some unexplained variation in the model and actions may be taken to improve the model specification (Haining, 2009). On the other hand, spatial autocorrelation can bring opportunities for practitioners. For example, spatial autocorrelation is modelled in geostatistics for the purpose of spatial interpolation.

The description of spatial autocorrelation can be used to define spatial weighting functions which play important roles in spatial modelling (Fotheringham, 2009). An understanding of spatial dependence can also help in the design of spatial sampling (Haining, 2003).

In regression models, a typical method to deal with spatial autocorrelation is to include spatially lagged variables which are weighted averages of observations from a neighbourhood of a given location. Spatially lagged variables can be defined on the dependent variable (spatial lagged models, or autoregressive models), explanatory variables (spatial cross-regressive models), error terms (spatial error models) or combinations of them (Anselin, 2003).

2.2.2 Spatial heterogeneity

Spatial heterogeneity, also called spatial non-stationarity, refers to the variation in processes and relationships over space (Bailey and Gatrell, 1995). It results in spatially varying relationships, i.e., different responses may result from the same stimuli at different locations. The issue of spatial non-stationarity was first raised by Granger (1969). Fotheringham et al. (2002) summarized three reasons that may cause spatial non-stationarity in parameter estimates: random sampling variations, intrinsically different local behaviours across space, and model misspecification.

Similar to spatial autocorrelation, spatial non-stationarity brings not just a problem but also an opportunity to spatial analysis. Spatial non-stationarity is seen as the basis for local statistics (Fotheringham, 2009). The existence of spatial non-stationarity in data generating processes suggests that global models which assume a constant functional structure across space are not sufficient to describe the underlying mechanisms; local models are then preferred. In regression models, the existence of spatial non-stationarity requires the regression coefficients to vary spatially, either discretely between spatially distinct units, or continuously over space. In the latter case, regression coefficients can be defined as a function of locations as in the spatial expansion model (Casetti, 1997), or more directly, estimated from the data through a local regression process, for example, as in geographically weighted regression (Fotheringham et al., 2002).

2.2.3 Interaction between spatial autocorrelation and spatial non-stationarity

As the two main aspects in spatial modelling, spatial autocorrelation and spatial non-stationarity may coexist in spatial modelling, sometimes interacting with each other. A typical example of their coexistence is the spatial non-stationarity in spatial autocorrelation, which means spatial autocorrelation may vary over space. Anselin (1995) introduced Local Indicators of Spatial Association (LISA) as a means for measuring spatial variation in spatial autocorrelation across space. Fotheringham (2009) demonstrated the interaction between spatial autocorrelation and spatial non-stationarity and concluded that local statistical models make it possible to model spatial dependence and spatial non-stationarity within the same framework. Spatial autocorrelation, through the construction of spatial weighting functions, based on which local statistics and local statistical models can be built, can help to model spatial non-stationarity. On the other hand, when spatial autocorrelation in the residuals from a regression model is caused by spatial non-stationarity, a local regression model which accounts for spatial non-stationarity should be used instead of global models to account for the spatial autocorrelation.

2.3 Non-spatial regression models

As the basis of regression models, traditional non-spatial regression methods, including global methods and local methods that are localized in attribute space, are outlined in this section.

2.3.1 Standard regression

Regression analysis is used to study the dependence of a response (or dependent) variable on one or more predictor (or independent) variables. Suppose a dataset consisting of n observations, X_i $(i = 1, \dots n)$ denotes the vector of the independent variables measured for observation i, and y_i is the dependent variable measured for observation i, a regression model takes the form:

$$y_i = \mu(X_i) + \varepsilon_i, \tag{2-1}$$

where $\mu(X_i)$ is the mean function, and ε_i is the error term.

In traditional regression models, a parametric form is assumed for the mean function $\mu(X_i)$. The basic form is the normal linear form, where $\mu(X_i)$ is assumed to be linear in the parameters to be estimated and the error terms ε_i s are assumed to be independent normal random variables with mean zero and constant variance σ^2.

Suppose there are p independent variables, i.e., X_i is a p-dimensional vector $(x_{i1}, x_{i2}, \dots, x_{ip})$, a normal linear regression model has the form:

$$y_i = \sum_{j=1}^{p} \beta_j x_{ij} + \varepsilon_i, \qquad (2\text{-}2)$$

where β_j is the parameter associated with the jth independent variable. The β_js can be estimated by Ordinary Least Squares (OLS), which mininizes the sum of squared differences between the measured and predicted values of the dependent variable Y_i. In cases where the ε_is do not have constant variance, Weighted Least Squares (WLS) should be employed.

The traditional linear regression model benefits from its simplicity and well-developed statistical properties which enable validation tests and easy interpretation of the model's parameters and, consequantly it has been widely used in different disciplines to describe relationships between variables. Developments in regression analysis include nonlinear regression (Seber and Wild, 2003), nonparametric regression (Stone, 1977, Wegman and Wright, 1983), and local regression (Cleveland, 1979, Loader, 1999).

2.3.2 Local regression

Local regression techniques have long been developed in statistics to examine local relationships in non-spatial data, for example, smoothing splines (Reinsch, 1967, Henderson, 1924, Silverman, 1984, Wahba, 1978), locally weighted regression (Cleveland, 1979, Cleveland and Devlin, 1988) and other nonparametric regression models. Here, the term "local" is defined by the attribute space of the independent variable, referring to methodologies where only a subset of the observations, characterised by the independent variables having values within a certain range are used in the calibration of the model.

Locally weighted regression

The local fitting technique in locally weighted regression stems from the idea of smoothing in time series (Macaulay, 1931), where data are measured at equally

spaced time intervals. The idea is to allow the data to determine the appropriate functional relationship among variables instead of specifying a parametric model (Loader, 2012). Smoothing was extended by Watson (1964), Stone (1977) and Cleveland (1979) into the more general settings of regression analysis. One of the most typical methods is locally weighted regression, which was introduced by Cleveland (1979) as a univariate smoother, and then expanded by Cleveland and Devlin (1988) into a multivariate form. Also known as *lowess* or *loess*, locally weighted regression constructs a functional relationship between the dependent variable and the independent variables at any values of X_i in the space of the independent variables, using the observations whose values are closest to those of the X_i's. Each point in the space of the independent variable is weighted according to its distance from X_i. Points close to X_i have larger weight and points farther from X_i have smaller weights. A linear or a polynomial function of the independent variables is fitted to the dependent variable using WLS with these weights. The fitting is repeated for each X_i where an estimation of the dependent variable is needed. The implementation of *lowess* involves the choice of a neighbourhood size to determine the subsets of data for local fitting, the choice of a weighting function to calcuate the weight to be used in WLS, and the choice of the degree of local polynomials for the local functions.

Lowess provides the flexibilities of nonparametric regression and nonlinear regression by fitting simple polynomial models to localized subsets of the data. Being a local regression method, it does not require a global function to be specified for the whole dataset. These features have made *lowess* an attractive methodology for modelling a wide range of complex processes in data. It can be used for data exploration, or for diagnostic checking of parametric models (Cleveland and Devlin, 1988).

Lowess, however, has its restrictions. As in traditional regression, it requires the errors to be independently and normally distributed with constant variance. Although local functions can be defined on any number of independent variables, *lowess* becomes less useful for more than two independent variables, this is because: (i) the increase in the number of parameters in local models requires more expensive computation; (ii) as the number of independent variables increases, the dimension of space where data points are located in becomes high and data points become sparse in space, this is known as the "curse of dimensionality" (Friedman and Stuetzle, 1981) and brings

more variance to the regression. This is a common problem for local regression smoothers and dimension reduction procedures need to be introduced to adapt smoothers to higher dimensions; examples can be seen in Projection Pursuit Regression (Friedman and Stuetzle, 1981) and Additive Models (Hastie and Tibshirani, 1986).

2.3.3 Generalized additive model

An additive model (Hastie and Tibshirani, 1986, Hastie and Tibshirani, 1990) assumes the regression surface to be an addtive sum of smooth functions of the independent variables. The model can be written as:

$$Y = \sum_{j=1}^{p} s_j\left(X_j\right) + \varepsilon, \tag{2-3}$$

where Y is the dependent variable, $X_j (j = 1, \dots, p)$ is one of the p independent variables, ε is the error, and $s_j(X_j)$ is an unspecified smooth function to be determined by the data. Each function $s_j(X_j)$ is smoothed separately on a single independent variable X_j, with its own choice of smoother, including *lowess*, splines, among others (see Buja et al., 1989 for more smoothers).

The additive model extends a linear model by replacing the linear parametric functions with smooth nonparametric functions. It is also an extension of local regression smoothers in that it attempts to model a high-dimensional surface through low dimensional smoothers by assuming an additive structure among the independent variables.

The Generalized Additive Model (GAM) is a generalization of the additive model in the same way as generalized linear model (Nelder and Wedderburn, 1972) generalizes linear regression. In a generalized linear model (GLM), the dependent variable Y is assumed to have a distribution belonging to the exponential family, which includes the normal, bionomial, Poisson, gamma, logistic and many other common distributions. The mean of Y is related to a linear parametric combination of the independent variables $X_1, X_2, \dots, X_p$ via a prespecified link function g:

$$g\big(E(Y)\big) = \sum_{j=1}^{p} \beta_j X_j + \varepsilon. \tag{2-4}$$

The parameters $\beta_1, \beta_2, \dots, \beta_p$ are usually estimated with maximum likelihood or Bayesian approaches. GAM replaces the linear function in GLM with an additive function, acting as a nonparametric extension of GLM:

$$g\big(E(Y)\big) = \sum_{j=1}^{p} s_j\left(X_j\right) + \varepsilon, \tag{2-5}$$

where $s_j\left(X_j\right)$ is a nonparametric smooth function as in equation (2-3).

The smooth functions can be estimated by a local scoring algorithm (Hastie and Tibshirani, 1986), with the additive model estimated by an iterative procedure of back-fitting (Mosteller and Tukey, 1977, Friedman and Stuetzle, 1981, Breiman and Friedman, 1985). The back-fitting method will be demonstrated in more detail in Chapter 4.

GAM can be seen as a data-driven model, where the nature of the relationships between the dependent variable and independent variables is suggested by the data, rather than being prespecified by a given parametric form (Yee and Mitchell, 1991). It provides more flexibility than the linear parametric model and has been widely used in modelling complex nonlinear relationships. Applications can be seen in economics (Deaton and Muellbauer, 1980), ecology (Guisan et al., 2002), environmental sciences (Wood and Augustin, 2002) and many other areas.

However, because of its flexibily, caution needs to be exercised when applying GAM in order not to over-fit the data (Hastie and Tibshirani, 1990). In addition, although the additive structure of GAM provides more interpretability than a general high dimensional regression surface (Buja et al., 1989), allowing the effect of each independent variable on the dependent variable to be examined seperately, GAM is not as easily interpreted as a linear model and the results of GAM are not easily summarized and communicated (Hastie and Tibshirani, 1986).

As a further extension of GAM and of linear parametric regression models, Hastie and Tibshirani (1993) put forward a common framework called the varying coefficient model. It maintains the linear relationship between the dependent variable and the independent variables but allows the parameters to vary as smooth nonparametric functions of other variables which can be either a new set of independent variables or a specific variable such as time. The varying coefficient model helps to explore the dynamic pattern in datasets which may be ignored in traditional regression. It can also be used as a trial method to help in building new statistical models. Fan and Zhang (2008) give a review of the development of various types of varying coefficient models and demonstrate their broad applications in many scientific areas, especially in time series, longitudinal data analysis and survival analysis.

The above developments in non-spatial regression models can be employed in spatial regression but the spatial effects of spatial autocorrelation and spatial non-stationary which usually violate the assumptions of non-spatial regression models are not explicitly incorporated and new techniques are thus required to be developed for the use of spatial data. These are now introduced.

2.4 Spatial regression models

Spatial regression models refer to models that have been developed from traditional regression to make them applicable to spatial data, taking spatial autocorrelation and spatial non-stationarity into account. Methods of spatial regression were initially applied at the global levels; the results of the models being therefore constant across the study region forming a global statement of the spatial processes under study. The realization of spatial non-stationarity in relationships between explanatory variables and a response variable in a regression model has lead to the development of local methods of spatial regression that allow for spatially varying coefficients. Instead of assuming global statements, local methods of spatial regression focus on identifying spatial variations in spatial process; the results being local estimates of regression parameters. This section gives a brief review of both global methods and local methods of spatial regression.

2.4.1 Global methods

Two traditional spatial regression models are the spatial autoregressive model and the spatial error model (Ord, 1975, Anselin, 1988), which are both global methods.

The spatial autoregressive model (SAR) (Ord, 1975, Anselin, 1988), also referred to as a spatial lagged model, is one of the most commonly used spatial process specifications to account for spatial effects. The original form of SAR is:

$$Y = \rho WY + \beta \mathbf{X} + \varepsilon, \qquad (2\text{-}6)$$

where Y is the dependent variable, ρ is a spatial autoregressive parameter, $\mathbf{W}$ is a weighting matrix, $\mathbf{X}$ is a matrix of independent variables and a column of ones accounting for the intercept, $\boldsymbol{\beta}$ is the parameters to be estimated and $\boldsymbol{\varepsilon}$ is the independently and identically distributed error with variance σ^2.

The weighting matrix measures the spatial dependence between all pairs of observations associated with different spatial units, based on spatial contiguity, distance or other specifications of proximity. Suppose there are n spatial units, **W** will be an $n \times n$ matrix with zero on the diagonal suggesting that a spatial unit cannot be related to itself. The auroregressive parameter ρ is an overal measurement of spatial dependency. The model can be calibrated using maximum likelihood (Anselin, 1988) or the generalized method of moments (Kelejian and Prucha, 1998).

SAR produces a global model because it relates all spatial units in the system to each other, although for small values of ρ the covariance between observations associated with two spatial units may approach zero when the distance between the units is larger than a certain value.

SAR is a basic framework for spatial regression and it can be expanded to various forms. The autoregressive parameter can not only be defined on the dependent variable as in equation (2-6), but can also be defined on independent variables, the error term, or a combination of these terms. When it is present in the error term, the model turns into the so-called spatial error model (SEM). All these models have been widely applied. Various extensions have been developed on the SAR framework, for example, the biparametric model (Brandsma and Ketellapper, 1979), spatial Durbin model (Burridge, 1981) and SAR with various forms of weighting matrix (Hepple, 1995, Bavaud, 1998, Getis and Aldstadt, 2004, Aldstadt and Getis, 2006).

Although a useful tool for modelling spatial autocorrelation between spatial units and explaining whole-map processes, SAR has its limitations. One limitation is that the model is inflexible due to its high dependency on the pre-specified weighting matrix which affects the estimates and the interpretation of results (Anselin, 2002, Fingleton, 2003, Folmer and Oud, 2008). The biggest limitation, however, is that SAR only captures a single global measurement of spatial autocorrelation while ignoring any possible local variations. To compensate for this, Anselin (1995) disaggregates the weighting matrix into local decompositions and examines the influence of these local weighting matrices on the global model. Brunsdon et al. (1998a) introduces a spatially varying autoregressive model by applying geographically weighted regression to the framework of SAR and derives a set of localized values indicating local autocorrelation instead of an average one. The need for localized versions of spatial autoregressive models is revealed both in theory and through examples.

2.4.2 Local methods

The purpose of local regression models is to investigate the spatial heterogeneity in model behaviour and therefore to suggest a more accurate model specification. Several such regression models are introduced in the following sub-sections.

Spatial expansion model

The spatial expansion model serves as one of the initial treatments of spatial heterogeneity and is a special case of the expansion model (Casetti, 1972) which was initially designed as a method for the construction and modification of models. The expansion model starts from an initial model where the estimates of the parameters exhibit some instability over the study area. The expansion method then relates the parameters to some new variables, makes the parameters functions of them, and generates a new model which can remove the instability in the parameter estimates so as to increase predictability or to test specific theoretical hypotheses.

Assume an initial model describing a linear relationship between a dependent variable Y and p independent variables $X_1, X_2, \dots, X_p$:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p + \varepsilon \tag{2-7}$$

where $\beta_0,\ \beta_1,\ \beta_2, \dots,\ \beta_p$ are the parameters to be estimated, $\boldsymbol{\varepsilon}$ is the random error.

If the parameters vary over space, it is possible that their values might be related to other contextual variables across the space. The model may be expanded by assuming each parameter to be a function of contextual variables which can be chosen according to existing information or theoretical considerations about the real world phenomena. A variety of terminal models may be constructed depending on which contextual variables are chosen and what functional relationships are proposed. Particularly, the parameters can be assumed as functions of location. For example, linear expansion functions can be defined as following:

$$\begin{aligned} \beta_0 &= \beta_{00} + \beta_{01} u + \beta_{02} v \\ \beta_1 &= \beta_{10} + \beta_{11} u + \beta_{12} v \\ &\dots \\ \beta_p &= \beta_{p0} + \beta_{p1} u + \beta_{p2} v \end{aligned} \tag{2-8}$$

where u and v are coordinate variables representing locations. By replacing equations (2-8) into (2-7), a spatial expansion model is obtained:

$$Y = \beta_{00} + \beta_{01}u + \beta_{02}v + \beta_{10}X_1 + \beta_{11}uX_1 + \beta_{12}v X_1 + \cdots + \beta_{p0}X_p + \beta_{p1}uX_p + \beta_{p2}v X_p + \varepsilon \quad (2\text{-}9)$$

β_{01}, β_{02}, β_{11}, β_{12},…, β_{p1}, β_{p2} can be estimated by ordinary least squares and then compose parameter estimates for β_0, β_1, β_2, … , β_p, which are spatially varying. The testing of the expansion model can be carried out by testing the null hypotheses that parameters β_{01}, β_{02}, β_{11}, β_{12},…, β_{p1}, β_{p2} are not significantly different from zero. More complex, non-linear expansion functions can also be defined, although this will increase the complication in the estimation of the parameters as well as in the testing of their significance. The spatial expansion model provides a systematic approach to incorporating a spatial context into a non-spatial regression model. With the spatially varying parameters, the expansion model addresses the spatial variation in parameters while preserving the structure of the initial model.

The drawback of the spatial expansion model is that it relies on pre-defined parametric forms of expansion functions to describe the trends in relationships over space rather than revealing the spatial variations in relationships directly. To enable the estimation of model, the choice of expansion functions is limited; general smooth trends are assumed in relationships and therefore the spatial variations may be simplified and some important local variations may be missed (Fotheringham and Brunsdon, 1999).

Multilevel modelling

Multilevel modelling (Aitkin and Longford, 1986, Goldstein, 1987) treats a regression model as a hierarchical system composed of regression equations at various levels. The parameters in the regression model are assumed to be random variables to be estimated from certain probability distributions which are also estimated. Consider a two-level model specification which involves a low-level model based on individual data and a high-level model based on group data. The low-level model has the form:

$$y_{ij} = \beta_{0j} + \beta_{1j}x_{ij} + \varepsilon_{ij}, \quad (2\text{-}10)$$

where y_{ij} is the dependent variable measured at individual i in group j, x_{ij} is the independent variable measured at the same observation as y_{ij}, and ε_{ij} is the random error. The two parameters β_{0j} and β_{1j} are random variables and assumed to be dependent variables at the high level:

$$\beta_{0j} = \beta_0 + u_{0j},\ \beta_{1j} = \beta_1 + u_{1j}, \tag{2-11}$$

where β_0 and β_1 are the global parameters for the whole dataset, u_{0j} and u_{1j} are randomly varying variables for group j with mean zero and constant variance. Equation (2-10) can be rewritten as:

$$y_{ij} = \beta_0 + \beta_1 x_{ij} + (u_{0j} + u_{1j} x_{ij} + \varepsilon_{ij}), \tag{2-12}$$

which is the sum of a fixed part and a random part within the brackets. As the multilevel model contains more than one residual term, traditional OLS is not applicable for parameter estimation but an iterative generalised least-squares (IGLS) algorithm can be used (Goldstein, 1986). More alternative estimation procedures are summarized by Jones (1991). The multilevel model can be extended to include more levels, more independent variables as well as nonlinear forms (Goldstein, 2010). The hierarchical structure can also be extended to a more complex underlying data structure, for example, multiple membership multiple classification model (Browne et al., 2001) has been developed so that an observation can belong to more than one member of a classification.

Multilevel modelling allows the specification of models at several different scales simultaneously and the exploration of variations between groups at the same scale. It has gained wide application in regression analysis where data have a hierarchical structure, for example, in educational data (Goldstein, 1991), longitudinal data (Orford, 2000, Huang and Clark, 2002), and other multi-scale geographical data (Jones, 1991). Among the hierarchical structures incorporated in the model, space is an important structure, i.e., the multilevel model allows relationships to vary in different spaces, making the model a useful local spatial regression model. However, the limitation of the multilevel model is that it assumes no dependence or connection between neighbouring observations or groups and the continuous nature of space is not taken into account (Fotheringham, 1997). It also requires the nature of the spatial structure to be defined *a priori*; a task not always possible due to lack of knowledge about what this structure might be.

Spatial filtering model

Spatial filtering is another methodology designed to deal with spatial autocorrelation. The idea is to remove spatial autocorrelation in spatially autocorrelated variables by filtering the variables in order to separate spatial effects from non-spatial effects so

that the regression residuals remain uncorrelated and traditional linear regression techniques such as OLS can be applied to the filtered data.

Various spatial filtering procedures have been developed to achieve this aim. Tiefelsdorf and Griffith (2007) review the different methods from the early parametric models which aim to identify an underlying data generating spatial process (Griffith, 1979, Haining, 1991), to the more recent nonparametrical approaches based on data transformations, including the G_i-based approach developed by Getis (Getis, 1990, Getis, 1995), and the eigenvector approach by Griffith (1996).

Getis's filtering approach is based on local autocorrelation measures, namely the local Getis G_i-statistic (Getis and Ord, 1992). The approach helps to improve the specification of multiple regression relationships but it only works on a univariate spatial pattern and therefore each variable in a regression model needs to be treated individually (Tiefelsdorf and Griffith, 2007).

Griffith's approach is based on eigenvector decomposition using a subset of eigenvectors that are extracted from the underlying spatial structure to capture the dependencies in a regression model. This approach helps to explore the impact of spatial autocorrelation on correlation coefficients (Getis and Griffith, 2002). It can be applied to individual variables as well as to a regression system simultaneously. It can also be incorporated into generalized linear models (GLM), offering a wider range of applications. The drawback with this approach is that the calculation of eigenfunctions involves intensive computation.

Both approaches allow the use of traditional OLS for model calibration. Getis and Griffith (2002) compared the two approaches and concluded that they yield similar regression models.

Based on Griffith's eigenvector approach, Tiefelsdorf and Griffith (2007) further developed a semiparametric eigenvector spatial filtering approach which offers more feasibility, flexibility, and simplicity. More spatial filtering methods, including another eigenvector approach based on principal coordinates of neighboring matrices (Borcard and Legendre, 2002), are reviewed by Griffith (2010).

Whilst spatial filtering approaches might be useful in uncovering the impacts of spatial structures in spatial regression, the inconvenience is that the filtered variables are less interpretable in a regression model.

Bayesian spatially varying coefficient regression

Bayesian regression models with spatially varying coefficient processes (SVC) were introduced to model non-constant linear relationships between variables in the statistics literature (Gelfand et al., 2003). In the SVC framework, the parameters are treated as unknown random quantities from a multivariate spatial process. The SVC model is constructed by adding a continuous multivariate spatial process to account for spatially varying coefficients and to accommodate spatial dependence between variables. Suppose there are n observations in a geo-referenced space. A general form of SVC model is then:

$$y_i = x_i\boldsymbol{\beta} + \widetilde{\boldsymbol{x}}_i\boldsymbol{w}_i + \boldsymbol{\varepsilon}_i, \tag{2-13}$$

where $y_i(i = 1, \dots n)$ is the dependent variable measured for observation i, $\boldsymbol{x}_i$ is the vector of independent variables measured for i and $\boldsymbol{\beta}$ is the vector of regression coefficients which are assumed to be constant. $\widetilde{\boldsymbol{x}}_i$ is a sub-vector of $\boldsymbol{x}_i$ whose regression coefficients $\boldsymbol{w}_i$ are assumed to be spatially-varying. ε_i is the independent and identically distributed residual with zero mean and σ^2 variance. $\boldsymbol{w}_i$ follows a multivariate Gaussian process which is characterized by its mean and a cross-covariance function. $\widetilde{\boldsymbol{\beta}} = \boldsymbol{\beta} + \boldsymbol{w}_i$ forms the vector of spatially-varying coefficients. The calibration of the model involves the choice of prior distributions for the model parameters.

The advantages of SVC are its flexibility and capacity for inference. By using complex correlation structures for the coefficient processes, SVC provides a valid probability model from which an entire posterior inference for the spatially varying coefficient processes at both observed and unobserved locations can be obtained. It enables both hypotheses tests on model parameters and uncertainty assessment on predictions.

The disadvantage of SVC is the complexity, which not only means high computational costs, but also high demands to the practitioners who require training and experience in hierarchical Bayesian models. Gelman et al. (2004) and Banerjee et al. (2004) have provided guidance on general Bayesian modelling and hierarchical models for spatial data. The lack of available software to fit SVC models is another limitation of the widespread adoption of the models (Finley, 2011).

Moving window regression

Moving window regression (MWR) is a straightforward local regression approach to address spatial non-stationarity in relationships between variables. It employes a similar idea to locally weighted regression *lowess* in non-spatial cases, that only a subset of the dataset is used to fit a separate regression model at each point. The difference is that MWR defines the subsets in a geo-referenced space rather than the attribute space as used in *lowess*. The procedure of MWR can be described as: each location is visited in turn in a moving window fashion, a window centered at the location is imposed and the observations situated within the window are included to conduct a regression model with local regression coefficients estimated for this location. Once all locations are visited, each location then has its own set of regression coefficient estimates which can be mapped and compared.

By calibrating various regression models for each location, MWR can capture spatial variation in processes. A key decision, however, to be made when applying MWR is the choice of the window size, which reflects the geographical scale of the spatial processes being investigated. The smaller a window size is, the more local the regression will be, while the more variance will be in the model. The effects of various window sizes and the methods to choose an optimal size have been demonstrated by applications such as Lloyd and Shuttleworth (2005) in commuting analysis and Páez et al. (2008) in house price prediction. Empirical comparison with global regression and other local regression techniques including SAR indicates that MWR produces accurate estimates and helps to identify local anomalies (Farber and Yeates, 2006).

MWR is limited by the boundary effect, that estimates near to the boundaries of study area are less robust than those farther away. To demonstrate, suppose the window size is defined by the distance to the central location, a location near to boundaries will have fewer observations in its window than more central locations. Alternatively, if the window size is defined by the number of observations within it, i.e. a fixed number of observations are used for each local regression, then a location near to boundaries will have observations further away included in its window. Both situations will introduce bias in the estimation.

Another drawback of MWR is that it is discrete (Fotheringham et al., 2002); any information contained in the observations outside of a window is totally ignored in a local regression. This may not be reasonable for a real world spatial process.

Geographically weighted regression

Moving window regression can be seen as a special case of a more sophisticated local regression framework, geographically weighted regression (GWR) (Brunsdon et al., 1996, Fotheringham et al., 1996, Brunsdon et al., 1998b, Fotheringham et al., 1998). In MWR, all observations within the window can be considered as having an equal weight of one, while all observations outside of the window have a weight of zero. GWR defines the weights in a more general way, weighting the observations at various locations according to their proximity to the centre of the window. The closer an observation is located to the centre of the window, the more weight it receives, and *vice versa*. This is in accordance with the geographical principle of distance decay. A weighting scheme is employed to control the distance decay. Rather than a subset of data, all data can be included in the analysis but the weighting scheme will determine which observations have non-negligible weights.

GWR explicitly takes spatial location into account and it has been considered as one of the most important developments in the treatment of spatial non-stationarity (Anselin, 2010). More details about GWR will be presented in the next chapter.

2.5 Summary

This chapter gives a brief review of spatial analysis in general and spatial regression in particular. The importance of space has been recognized in various scientific disciplines involving theoretical domains and applied domains. The two main spatial effects that spatial data bring to data analysis are spatial autocorrelation and spatial non-stationarity. To address these two affects has thus been the main task of spatial data analysis. In terms of regression, traditional non-spatial regression methods are often insufficient in analysing spatial data while various fundamental methodologies including locally weighted regression, generalized additive modelling and so on have thrown some light on the development of local spatial regression.

Local spatial regression emphasises the exploration of spatial non-stationarity and constitutes one of the main trends in spatial regression. Local spatial regression

provides more information on spatial relationships between variables, improves the model specification and aids a better understanding of spatial processes. Among the many local spatial regression methods that have been developed, GWR is the focus of this thesis. Various aspects of GWR will be demonstrated in the next chapter.

Chapter 3 Geographically Weighted Regression

3.1 Basic methodology

Geographically Weighted Regression (GWR) (Brunsdon et al., 1996, Fotheringham et al., 1996, Brunsdon et al., 1998b, Fotheringham et al., 1998) is a useful technique to model spatial dependency and spatial heterogeneity. Based on the assumption that the relationship between the observed variables varies smoothly across a geographical space, it adopts the smoothing techniques used in Locally Weighted Regression models (Cleveland, 1979, Cleveland and Devlin, 1988) in attribute space and extends them into geographical space.

In GWR, the regression is repeated at every single location, taking into account the influence of data from surrounding locations. As a result, a series of local parameters is estimated. These parameters can be mapped as a surface, representing the variation in relationships over space. In this way, GWR describes not only the relationships between variables but also any spatial non-stationarity of these relationships.

Suppose a global linear regression model has the form:

$$y_i = \beta_0 + \sum_{k=1}^{m} \beta_k x_{ik} + \varepsilon_i \text{ for } i = 1, \ldots, n \qquad (3\text{-}1)$$

where y_i is the response or dependent variable measured at observation point i, x_{ik} $(k = 1,...,m)$ is the k th predictor or independent variable at observation point i, β_k is the parameter for x_{ik} describing the relationship between y_i and x_{ik}. β_0 is the intercept and ε_i is the error term. When the error terms meet the condition of being independently and identically drawn from a Normal Distribution with mean zero and common variance σ^2, the regression parameters β_0 and β_k can be estimated by Ordinary Least Squares (OLS), in which the value of $\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$ is minimised over the n observations in the dataset, where $\hat{y}_i$ stands for the predicted value.

The basic form of GWR according to Fotheringham et al. (2002) is:

$$y_i = \beta_0(u_i, v_i) + \sum_{k=1}^{m} \beta_k(u_i, v_i) x_{ik} + \varepsilon_i \qquad (3\text{-}2)$$

where (u_i, v_i) stands for the geographical coordinates of the i th observation. $\beta_0(u_i, v_i)$ and $\beta_k(u_i, v_i)$ are parameters describing the relationships around location (u_i, v_i). Other notations have the same meaning as those in the global regression model (3-1). By allowing the parameters to be location-specific, GWR incorporates into the model any potential spatial variation in relationships.

The fundamental mechanism of GWR is that an observation point is weighted according to its proximity to the regression point (the point at which the model is locally calibrated), so that data from observations close to a regression point are weighted higher than data from observations farther away. This accords with Tobler's first law of geography that "Everything is related to everything else, but near things are more related than distant things" (Tobler, 1970). The weighting scheme is centred on each regression location so that each regression point has its own weighting matrix to participate in the estimation of the locational specific parameter estimates through a Weighted Least Squares approach:

$$\hat{\beta}_i = (\mathbf{X}^{\mathbf{T}}\mathbf{W}_i\mathbf{X})^{-1}\mathbf{X}^{\mathbf{T}}\mathbf{W}_i y \tag{3-3}$$

where $\hat{\beta}_i$ is the vector of estimated parameters at regression point i, $\mathbf{X}$ is the design matrix which contains the values of the independent variables and a column of 1s, y is the vector of observed values of the dependent variable. $\mathbf{W}_i$ is a weight matrix conditioned on the location i with the form:

$$\mathbf{W_i} = \begin{matrix} w_{i1} & 0 & 0 & \dots & 0 \\ 0 & w_{i2} & 0 & \dots & 0 \\ 0 & & w_{i3} & \dots & 0 \\ . & . & . & \dots & . \\ 0 & 0 & 0 & \dots & w_{in} \end{matrix} \tag{3-4}$$

This is an n by n square matrix with all the off-diagonal elements zero and the diagonal elements ranging from zero to one representing the weights of observation points from 1 to n for the regression point i, e.g. w_{i1} stands for the weight of data from observation point 1, w_{i2} is the weight of data from observation point 2, and so on.

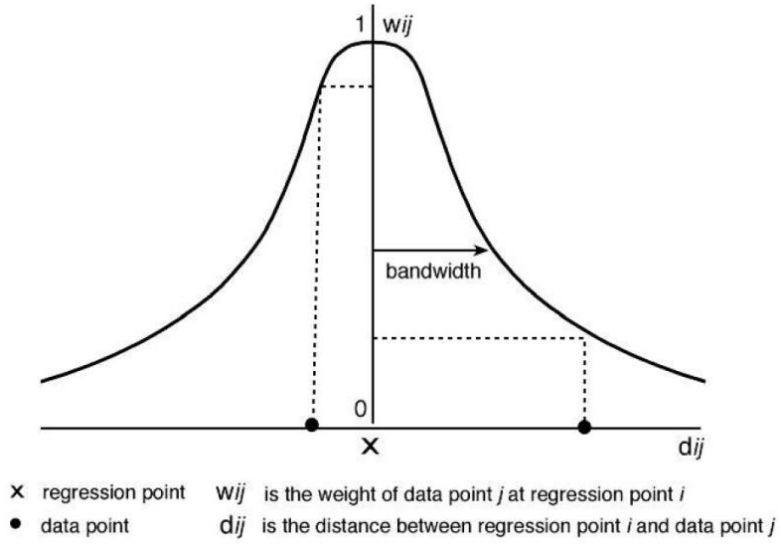
The weighting matrices are obtained through a function of distance W(*d*), named a weighting function or a kernel. The general requirements for a weighting function as stated by Brunsdon et al. (1998) are:

(1) a monotone decreasing function for positive real numbers, to guarantee that the weight decreases as the distance increases;

(2) W(0) = 1, so that the observed data at the regression point contributes fully to the regression;

(3) $\lim_{d\to\infty} W(d) = 0$, so that data from infinite distances hardly contribute to the regression.

A typical function is given in Figure 1.



(Fotheringham et al., 2002, P.44)

Figure 3-1 A typical weighting function

where × marks the regression point and the curve describes how the weight varies from one to close to zero according to the distance between the data point and the regression point. If a data point is far from ×, the weight will be small, while if a data point is close to ×, the weight will be large. If the data point is at the same location as the regression point, the weight will be the maximum value of one. It is worth noticing that the distance here is a general measurement of the proximity between spatial locations and it can be measured either in Euclidean distance or in some other meaningful type of distance according to the nature of the application.

There are many functions that can serve as a kernel. One common choice is a continuous exponential kernel, of which Gaussian function is probably the most widely applied in GWR:

$$w_{ij} = e^{-\frac{1}{2}(\frac{d_{ij}}{h})^2} \tag{3-5}$$

where w_{ij} is the weight of the observation point i for the regression point j, d_{ij} is the distance between point i and point j, h is known as a bandwidth and determines the size of the kernel and controls the rate of distance-decay in the weighting mechanism. The larger h is, the slower the weight will decay with d_{ij}. When h tends to infinity, the weight will tend to 1 for all values of d_{ij}; in this case, the model is approximate to a global regression model.

Another useful kernel is the bi-square function:

$$\begin{aligned} w_{ij} &= [1-(\frac{d_{ij}}{h})^2]^2 \text{ if } d_{ij} < h \\ &= 0 \qquad\qquad \text{otherwise} \end{aligned} \tag{3-6}$$

The difference between this kernel and a Gaussian kernel is that the bi-square excludes those observations beyond an influence distance of h by setting their weights to zero. This property reduces the computational cost of weights, making the model computationally more efficient; moreover, this property is in accordance with our understanding of many spatial processes.

The Gaussian and bi-square are the two most frequently used kernels although there are many other kernel functions that could be utilised depending on the spatial process being studied such as, for example, the tri-cube kernel function used by McMillen and McDonald (2004). A kernel function is usually conditioned by a general constant, the bandwidth. The bandwidth determines the rate at which the weights decay. A large bandwidth allows the weights to decrease slowly while a small bandwidth allows the weights to decay quickly.

While both the kernel function selected for the GWR model and the bandwidth selected for the kernel affects the estimated parameters of GWR, the effect of bandwidth selection is far greater than that of the kernel function (Brunsdon et al., 1996, Fotheringham et al., 1998). Bandwidth selection is a trade-off between bias and variance. When the bandwidth is large, more data points from far away will be involved in the regression so that the variance will be small while the bias will be large. If the bandwidth tends to infinity, the model will tend to a global regression model, where the surface of parameters will tend to be flat and local anomalies will be masked. If a small bandwidth is employed, the regression will be restricted to a local

area and the parameter estimates will depend on observations in close proximity to the regression point. Therefore the variance of the parameter estimates will increase but the bias in them will be small and more anomalies can be discovered.

Several data-driven approaches have been developed to select an optimal bandwidth for a given dataset, such as using cross-validation (CV) to minimize prediction error and using Akaike Information Criterion (AIC) to balance between goodness of fit and complexity of the model (Fotheringham et al. 2002). Whichever criterion is used, however, the result will be a constant optimal bandwidth for the model, all the relationships in the model will then be examined based on this bandwidth.

From the perspective of weighting, a global regression model can be considered as a special case of GWR where the weights are constant and equal to one for all regression points over all the dataset. Equally, moving window regression can be seen as having a boxcar function as the kernel:

$$\begin{aligned} w_{ij} &= 1 \quad \text{if } d_{ij} < h \\ &= 0 \quad \text{otherwise} \end{aligned} \tag{3-7}$$

where the bandwidth *h* equal to the window size. In this sense, GWR can be viewed as a general framework for a variety of regression models.

Although GWR was first introduced as an exploration tool to detect spatial non-stationarity in relationships between variables, the GWR estimates in the form of 3-3 provide a reasonable fit to the model in 3-2. It has been shown that, with the assumptions of independently normally distributed error terms having mean zero and common variances in formula (3-2), the least squares estimates of $\beta_k(u_i, v_i)$ are equivalent to estimates from a maximum likelihood approach (Fotheringham et al., 2002). Therefore GWR can be a statistical model with proper statistical inferences. Classical inference techniques, such as significance tests, confidence intervals, model comparison also apply to GWR.

3.2 Use of GWR

Since its introduction, GWR has proven to be a popular tool for the study of geo-referenced data and has been applied widely from social to natural sciences. The purpose of using GWR includes exploration, visualisation and prediction.

Initially, GWR was proposed as an exploratory tool to describe and map the spatial variations in the relationships between variables (Brunsdon et al., 1996, Fotheringham

et al., 1996). Fotheringham et al. (1998) suggested using it as a diagnostic to improve spatial understanding. Since the main outputs of GWR are the spatial variations in parameter estimates, these spatial variations can help to uncover interesting patterns in relationships, raise new research questions and guide further investigation into the processes being studied. Applications of GWR as an exploratory tool can been found in diverse disciplines, such as, ecology and environment (Foody, 2003, Pavel et al., 2008, Kupfer and Farris, 2007), climatology (Brunsdon et al., 2001), urban poverty (Longley and Tobón, 2004), regional science (Farber and Yeates, 2006, Huang and Leung, 2002), health and disease analysis (Nakaya et al., 2005b), environmental justice (Mennis and Jordan, 2005) and political science (Darmofal, 2008, Cho and Gimpel, 2010).

Uncovered variation in parameter estimates might be inherent to the geographical nature of the study area or might be caused by misspecification of the regression model. In the latter case, spatial variations could be removed by adding more explanatory variables into the model. As demonstrated by Brunsdon et al. (1998b), GWR can either be used to suggest the factors causing the spatial variation based on the geographical knowledge of the study area, thus helping to build a better model with more explanatory variables, or simply be used to visualize the spatial nature of relationships and the geographical effects which might cause spatial non-stationarity. While it is a common practice to map the parameter estimates and other outputs from GWR, research on further visualization techniques beyond conventional mapping methods has been carried out on GWR results. Mennis (2006) proposed adjustments to conventional data classification and colour schemes to improve the representation of the sign, magnitude, and significance of parameter estimates so as to improve the interpretation of GWR results. Demšar et al. (2008) treated GWR estimates as a multivariate dataset and explored them in a geovisual exploratory environment built on GeoVISTA Studio (Takatsuka and Gahegan, 2002), which is an integrated visual programming environment for geoscientific data analysis and visualisation. The geovisual exploratory post-analysis of GWR results has been used to identify spatial patterns of the parameter estimates and to compare the estimates of multiple variables. GWR is also applicable to prediction due to the property that the regression points at which local parameters are estimated are not limited to the points at which data are collected. Recently, the use of GWR as a predictor has gained increasing attention and considerable research has been undertaken on the comparison of GWR with other

approaches in producing accurate spatial predictions. For example, Gao et al. (2006) evaluate the prediction power of GWR together with a global linear regression model, a spatial dependency model, and a GWR mixed with spatial dependency model, through a house and land price dataset. Results show the superior performance of GWR on several empirical criteria. Kupfer and Farris (2007) apply GWR to predictive vegetation modelling and conclude that the prediction results from GWR had a closer fit to observed values with lower residuals than those from a global linear regression model. Bitter et al. (2007) compare GWR and the spatial expansion model in examining spatial heterogeneity in housing attribute prices within Tucson, Arizona and conclude that GWR outperforms the spatial expansion model in terms of explanatory power and predictive accuracy and that when explanatory power and predictive accuracy are the primary objectives of study, GWR is the superior approach. Lloyd (2010b) illustrates the benefits of local approaches including MWR and GWR against global regression for the characterization of spatial variation and for prediction. Simple kriging with local means (SKlm) conducted using local variograms estimated from GWR is also demonstrated. Harris et al. (2010b) compared GWR as a predictor with global multiple linear regression (MLR) and traditional geostatistical models of kriging, as well as a GWR based kriging (GWRK) where kriging models are specified with GWR as a mean component. Tests on simulated data with different levels of spatial heterogeneity and spatial autocorrelation show that both GWR and GWRK predictors performed better than MLR and ordinary kridging and that they could act as alternatives to the usually preferred kriging with external drift (KED) when predicting with non-stationary relationships.

Although the power of GWR in investigating local relationships and generating knowledge has been widely demonstrated, there has been a debate concerning the appropriate use of GWR in spatial inference (Wheeler and Tiefelsdorf, 2005, Wheeler and Calder, 2007). Páez et al. (2011) demonstrated through simulation experiments that GWR may introduce spurious inter-coefficient correlations in the resulting spatially varying coefficients when the sample size is small (160 in their experiments), which may lead to spurious interpretations of spatial patterns and therefore caution should be taken when using GWR as an inferential tool in these cases. For large sample sizes, the effect of spurious correlations is small and the estimated coefficient surfaces are reliable.

3.3 Software for GWR

Along with the wide application of GWR, various software and tools for GWR are available to spatial analysts. The earliest GWR software that is still in existence and known to the public is GWR 3.0 (Charlton et al., 2007), which is an update of the earlier version GWR 2.0 introduced by Fotheringham et al. (2002). Developed in the FORTRAN programming language, GWR 3.0 provides computational efficiency and a user-friendly interface which allows the users to configure and run a GWR model. Parameter estimation results together with essential diagnostics including standardised residuals, local r-squares etc. are produced. The parameter estimates can be later imported into spreadsheet or GIS platforms for visualization. Predictions at non-regression points can also be achieved. The kernel functions the user can select in GWR3.0 include a fixed Gaussian and an adaptive bi-square kernel. A bandwidth can be specified by the user or automatically selected through CV or AIC using a golden section search technique. GWR 3.0 also provides significance tests for local parameter estimates and for the spatial variation in a set of local parameter estimates.

A more recent and flexible software is GWR 4.0 developed by Nakaya et al. (2009) using the Microsoft .Net Framework. Compared to GWR 3.0, GWR 4.0 has an improved user interface and a wider range of options for kernel function specification and bandwidth selection. Flexible combinations of fixed/adaptive and Gaussian/bi-square kernel functions can be specified, AICc and BIC (Bayesian Information Criterion) criteria are added for model selection and an interval search method in which users can specify the lower/upper limits and a searching interval is added for bandwidth selection. The most significant feature of GWR 4.0 is the implementation of mixed GWR, where users can specify each coefficient as geographically fixed or varying either manually or through an automatic variable selection routine based on recursive model comparisons. A new geographical variability test of coefficients based on model comparison is also added.

There is also an R package, spgwr (Bivand and Yu, 2010), that has been widely used. Apart from geographically weighted regression, the spgwr package also provide functions for geographically weighted local statistics. Being open source, spgwr is attractive to users desiring more flexibility in model building and extending. One limitation of spgwr is that it does not provide spatial non-stationarity tests for coefficients.

Another R package implementing GWR is gwrr (Wheeler, 2013), which provides collinearity diagnostic and remedial tools for GWR models. Collinearity between independent variables has been raised as a drawback to the use of GWR (Wheeler and Tiefelsdorf, 2005), potentially producing unstable local regression coefficients and misleading the interpretation of the coefficients. The gwrr package calculates the variance-decomposition proportions and the condition indexes for the weighted design matrix of a GWR model to detect collinearity which, if indicated, can be relieved by geographically weighted ridge regression (Wheeler, 2007) and geographically weighted lasso (Wheeler, 2009) models provided in the package. The gwrr package does not supply as many options in the model specification as spgwr.

The most recently released R package GWmodel (Lu et al., 2013) provides a collection of geographically weighted techniques including GWR, GW principal components analysis, GW summary statistics and GW discriminant analysis. Various forms of GWR including basic, robust, generalised, mixed (semi-parametric GWR), heteroskedastic GWR are included in the package. Collinearity diagnostics are also provided and a remedy with locally compensated ridge supplied. Unlike GWR 4.0, the implementation of semi-parametric GWR in GWmodel does not allow an automatic selection of bandwidth for the locally varying terms.

GWR is also available in conventional GIS software such as ArcGIS (from ArcGIS 9.3) and spatial analysis software, for example SAM (Spatial Analysis in Macroecology) (Rangel et al., 2010) and Stata (Pearce, 1998). The implementations in them are not as flexible as in those dedicated tools mentioned above though.

This variety of software and tools help to facilitate the application and further development of GWR.

3.4 Alternative versions of GWR

Although the basic GWR framework specified by (Fotheringham et al., 2002) is the most popular version of GWR, there are other versions of GWR with various perspectives. Two typical ones are briefly described here.

Páez et al. (2002) changed the statistical perspective for GWR from modeling parametric non-stationarity to modeling spatial variance heterogeneity, arguing that GWR can be more fruitfully viewed as a local model of variance where non-stationarity is a consequence of spatial variance heterogeneity. In their GWR

framework based on maximum likelihood, variance is modeled as a smooth function of geographical distance, the parameter of kernel function is estimated on a local basis and the hypothesis of locational heterogeneity is tested. This alternative model makes GWR applicable to situations where the hypothesis of variance homogeneity is rejected by a statistical test.

Wang et al. (2008) proposed a locally linear approximation of GWR based on the Taylor's expansion of the spatially varying coefficient, expanding coefficients as linear functions of the spatial coordinates. This local linear-fitting-based GWR improves the overall model fit and reduces the coefficient bias compared to basic GWR.

3.5 Extensions of GWR

Following the introduction of GWR into geographical and statistical sciences, various extensions of GWR have been developed. Some arise from the model itself, for example mixed GWR, and some are proposed to remedy certain issues inherent in the basic GWR model, for example, penalized GWR. This section gives a brief introduction to various extensions.

3.5.1 Mixed GWR

One of the most straightforward extensions is mixed GWR. A mixed GWR model combines a global regression model together with a basic GWR model by incorporating geographically varying parameters as well as geographically constant parameters. The general form of mixed GWR is :

$$y_i = \sum_{k=1}^{l} \beta_k x_{ik} + \sum_{k=l+1}^{m} \beta_k (u_i, v_i) x_{ik} + \varepsilon_i \tag{3-8}$$

where x_{ik} $(k = 1,...,l)$ is the independent variable at observation point i with a fixed parameter β_k, x_{ik} $(k = l+1,...,m)$ is the independent variable at observation point i with a geographically varying parameter $\beta_k(u_i, v_i)$. y_i and ε_i have the same meanings as before.

In this way, the model mixes globally fixed terms and locally varying terms of explanatory variables simultaneously, accommodating the real-world datasets where

only some of the relationships to be examined between dependent variable and independent variables exhibit non-stationarity.

As global fixed terms can be considered as having a weighting function with an infinite bandwidth which yields constant weight equal to 1, mixed GWR can be seen as a two-bandwidth GWR: one bandwidth is infinity and the other is a local bandwidth specified for the locally varying terms. This is a special case or an early form of FBGWR proposed in this thesis.

3.5.2 Generalized GWR

Basic GWR is based on a conventional Gaussian regression framework which assumes a Gaussian distribution for the dependent variable. This assumption may not be realistic for all scientific analysis situations. For example, where the investigated dependent variable is a count, Poisson regression can be more appropriate (Lovett and Flowerdew, 1989).

Similar to the way in which Generalized Linear Models extend the basic linear regression model, the basic GWR model can also be extended to generalized GWR (GGWR), or geographically weighted generalised linear model (GWGLM) (Fotheringham et al., 2002). This allows a range of generalized linear modelling including Gaussian, Binomial, Poisson, Gamma and Logistic regression to be incorporated into GWR. Take Geographically Weighted Poisson Regression (GWPR, Nakaya et al. (2005b)) as an example, the model has the form:

$$y_i \sim \text{Poisson}[N_i exp(\beta_0(u_i, v_i) + \sum_{k=1}^{p} \beta_k(u_i, v_i) x_{ik})] \quad (3\text{-}9)$$

where Poisson[λ] is a Poisson distribution with mean λ, y_i is the dependent variable, which is a positive integer, measured at observation point i, N_i is the offset variable at the same location. As in equation (3-2), (u_i, v_i) represents the coordinates of the location, $\beta_0(u_i, v_i)$ and $\beta_k(u_i, v_i)$ are the locally varying parameters, and x_{ik} $(k = 1, \dots, p)$ is the kth independent variable at location (u_i, v_i). The model can be calibrated through local likelihood methods (Loader, 1999) and iteratively reweighted least squares with geographical weights from a kernel function. The outputs are the local estimated coefficients similar to those obtained in basic GWR.

3.5.3 Robust GWR

As in other regression models, outliers in the observations can reduce the accuracy of GWR parameter estimates. Robust GWR (Fotheringham et al., 2002, Harris et al., 2010c) is designed to account for outliers in GWR calibration. Fotheringham et al. (2002) demonstrate two approaches to robust GWR. One is the filtered data approach, which refits a GWR model on a filtered data set by removing observations that have large externally studentized residuals of an initial GWR fit. The other is an automatic approach where observations with large raw residuals are down-weighted after iterative GWR fits. The filtered data approach is more computationally intensive, while the drawback of the automatic approach is that it does not allow examination of unusual observations. Harris et al. (2010c) extended the filtered data approach by replacing the global outlier identification with a localized outlier identification where a mean smoother with the same spatial scale of the GWR model is applied on the residuals.

3.5.4 Penalized GWR

Penalized versions of GWR, including geographically weighted ridge regression (GWRR, Wheeler (2007)) and geographically weighted lasso (GWL, Wheeler (2009)), are designed to reduce the effect of local variable collinearity among GWR coefficient estimates. It is reported that in linear regression models, strong collinearity in the explanatory variables can increase the variance of the estimated regression coefficients (Wheeler and Calder, 2007). Wheeler and Tiefelsdorf (2005) investigate the collinearity effect in GWR, showing that collinearity in explanatory variables may be increased by the spatial weights and lead to potentially strong dependence in the local estimated coefficients and that this effect may decrease precision and power in parameter estimates in GWR and potentially mislead their interpretation.

GWRR and GWL are both penalization methods which place a constraint on the regression coefficients. GWRR uses a ridge regression parameter to constrain the estimated coefficients by minimizing the sum of the residual sum of squares and a penalty on the size of the squared coefficients. The ridge regression coefficients have the form:

$$\hat{\beta}^R = arg\,min_\beta \left\{ \sum_{i=1}^{n} \left(y_i - \beta_0 - \sum_{k=1}^{p} x_{ik}\beta_k \right)^2 + \lambda \sum_{k=1}^{p} {\beta_k}^2 \right\} \quad (3\text{-}10)$$

where λ is the ridge regression parameter that controls the amount of shrinkage. λ can be supplied by users or estimated by CV or other criteria. GWL takes the shrinkage of GWRR further by potentially shrinking the least significant variable coefficients to zero. Both penalized versions are said to be able to stabilize regression coefficients and thus improve the interpretation of GWR results (Wheeler, 2007, Wheeler, 2009).

3.5.5 Other Extensions

Heteroskedastic GWR (Fotheringham et al., 2002) relieves the assumption of constant variance error by basic GWR allowing the residual variance in the model to vary across space. The extension replaces the global estimate of residual variance in basic GWR with a local estimate. The extended model is calibrated in an iterative manner with re-weighted GWR fit in each step based on a mean smoothing over the observed squared residuals. This extension makes GWR more widely applicable in real-world studies. Harris et al. (2011) further extend heteroskedastic GWR into a predictive form which can increase prediction accuracy as well as prediction uncertainty accuracy over the basic GWR model.

Páez (2004) challenged the assumption of isotropy by incorporating directional variation in kernel functions of GWR to explore anisotropy in spatial processes. This extension allows the investigation of non-stationary parametric surfaces that may vary at different rates in different directions relative to a given location.

Lu (2012) extends the default Euclidean distance that is usually employed for the weighting scheme in GWR to a wide range of non-Euclidean distance metrics, particularly a network distance (Lu et al., 2011) and travel time metrics. The family of Minkowski distance matrices is also examined to yield an optimum distance metric for a given GWR model. This extension helps improve the understanding of the complex nature of geographical space.

FBGWR, as a new extension, challenges the assumption of equal scale of variability in the coefficients and attempts to incorporate multiple scales of analysis into the GWR model. The detail of FBGWR is now discussed.

Chapter 4 Flexible Bandwidth GWR

4.1 Why Flexible Bandwidths

As demonstrated in Chapter 3, bandwidth selection is an important decision in GWR reflecting a trade-off between bias and variance (Brunsdon et al., 1999b, Fotheringham et al., 2002). The bandwidth controls the rate at which the regression weights decay around a certain location. If the bandwidth is small, weights decline quickly as the distance between a data point and the regression point increases, thus the values of the regression coefficients change can rapidly over space. This may help to uncover local anomalies but can lead to instability in the parameter estimates. As a result, the model yields small bias but large variance in the parameter estimates. On the other hand, if the bandwidth is large, weights decline slowly as the distance increases and smoother coefficient surfaces are produced. The model estimates will have larger bias but smaller variance.

A lot of empirical studies have been undertaken on the impact of bandwidth on the descriptive and predictive power of GWR models. For example, Lloyd and Shuttleworth (2005) discuss the relationship between bandwidth size and observed spatial pattern; Propastin et al. (2008) explore the effect of scale on prediction uncertainty of GWR by varying the bandwidth; Guo et al. (2008) investigate the impacts of different bandwidths on model fitting, coefficient estimates and the residuals from GWR.

The bandwidth issue is essentially a scale issue, which is a focus of concern in geography (Haining, 2009, Lloyd, 2012). The size of bandwidth expresses the scale at which processes operate. A small bandwidth enables a local analysis and examination of effects at a small spatial scale; a large bandwidth suggests regional processes operating. When applying GWR to a given dataset to study spatially varying relationships between variables, an ideal bandwidth should be able to reflect the scale of variability in the relationships. In turn, *a priori* knowledge about the process to be examined could guide the bandwidth selection. In a multivariate spatial process however, variables may influence processes differently at different scales (Lam and Quattrochi, 1992, Poon, 2004), for example, Lloyd (2010a) demonstrates how population characteristics such as community background and employment level and

the relationships between these variables differ at different spatial scales as well as at different geographical locations. In these cases, the relationship between the dependent variable and each independent variable will have its own scale of spatial variation, thus a uniform bandwidth specified in basic GWR is insufficient to capture the spatial variations at various scales. A Flexible bandwidth GWR (FBGWR) model with potentially different bandwidths specified for each individual variable is thus needed. In short, FBGWR is required by the nature of spatial process, in that relationships may vary at different spatial scales.

4.2 Basic methodology

The formula of a basic GWR model as described in Chapter 3 is:

$$y_i = \beta_0(u_i, v_i) + \beta_1(u_i, v_i)x_{i1} + \beta_2(u_i, v_i)x_{i2} + ... + \beta_m(u_i, v_i)x_{im} + \varepsilon_i \quad (4\text{-}1)$$

where the notations are the same as in equation (3-2).

FBGWR extends this framework by replacing $\beta_k(u_i, v_i)$ with $\beta_{bwk}(u_i, v_i)$, and the formula is rewritten as:

$$y_i = \beta_{bw0}(u_i, v_i) + \beta_{bw1}(u_i, v_i)x_{i1} + \beta_{bw2}(u_i, v_i)x_{i2} + ... + \beta_{bwm}(u_i, v_i)x_{im} + \varepsilon_i \quad (4\text{-}2)$$

The new notation of $\beta_{bwk}(u_i, v_i)$ indicates that a different bandwidth is allowable for each local parameter estimate associated with a location. Thus not only do the regression coefficients vary at different locations, but the extension entails that they now also vary at different spatial scales for each independent variable. The scale of variability is controlled by the bandwidth of the coefficient. This property distinguishes FBGWR from standard GWR. It can be seen that standard GWR is a special case of FBGWR where the bandwidths are constant and where the coefficients vary at the same spatial scales for all independent variables. The intercept term $\beta_{bw0}(u_i, v_i)$, considered as a coefficient for a constant independent variable of all ones, will also have its own bandwidth.

To calibrate equation (4-2), two main issues need to be addressed:

1) Simultaneous calibration of all the terms. Standard GWR employs a Weighted Least Squares (WLS) approach to estimate the coefficients in the model (Fotheringham et al., 2002). The weighting changes as the location of the regression point *i* varies because a spatial weighting matrix is constructed so that each

observation is weighted according to its proximity to the location of point i. The choice of kernel function and bandwidth controls the weighting matrix and the coefficients in equation (4-1) will then be estimated with WLS at each regression point. In FBGWR, however, each term in the right-hand side of equation (4-2) has its own bandwidth to control the scale of variability of the coefficient as well as the weighting scheme; therefore, each term will have its own weighting matrix. Thus the simple WLS method of basic GWR will not apply to this situation and a new method has to be found to estimate all the parameters simultaneously.

2) Bandwidth selection. How to select the bandwidth for each independent variable or associated coefficient is the essential task of FBGWR. The selected bandwidths should be able to reflect the scale of variability in the processes described by each coefficient and can act as guidance for model selection with simpler regression models such as GWR, mixed GWR and MLR. A large bandwidth will suggest a global process over space, while a small bandwidth suggests a spatially local process. If roughly equal but local bandwidths are found for each term in FBGWR, a basic GWR will be enough to model the given dataset. If roughly equal but global bandwidths are found, then a MLR model suffices. If some bandwidths are roughly global whilst some are more local and equal then a mixed GWR model may suffice.

4.3 Back-fitting algorithm

One approach to calibrate FBGWR is to use the back-fitting method which has been similarly used by Brunsdon et al. (1999b) in calibrating mixed GWR. The idea is to calibrate each term in turn, assuming that all the other terms are known. The partial residual, defined as residual from the full model plus the estimate of the current term under consideration, is regressed on each individual variable, giving a new estimate for the term. The process moves on to the next variable with the new estimate for the previous term. The procedure continues in an iterative manner until the changes of all the terms on successive iterations are sufficiently small. In this way, all the calibrations are solved simultaneously.

The back-fitting algorithm itself is not new. It has long been used to solve equations which are in an additive form. Mosteller and Tukey (1977) used a similar algorithm to fit additive effects in an analysis of variance situation. Friedman and Stuetzle (1981) used back-fitting in their Projection Pursuit Regression, which is a nonparametric

multiple regression method that models the regression surface as a sum of general smooth functions of linear combinations of the predictor variables in an iterative manner. Back-fitting was defined as a procedure of readjustment of the smoothers along previously determined linear combinations when a new linear combination has been found. Breiman and Friedman (1985) used the back-fitting algorithm as the inner loop of their alternating conditional expectations (ACE) algorithm and demonstrated convergence for compact projection operators in a population setting. A more popular use of the back-fitting algorithm is in Generalized Additive Models (GAM) (Hastie and Tibshirani, 1986, Hastie and Tibshirani, 1987, Yee and Mitchell, 1991).

GAM relaxes the linear assumption for the covariates $X_1, X_2, \ldots X_p$ in normal linear regression model and linear logistic model and replaces the linear form of the covariates $\sum_{j=1}^{p} \beta_j X_j$ by a sum of smooth functions $\sum_{j=1}^{p} f_j(X_j)$, where the $f_j(\cdot)$'s are unspecified smooth functions to be estimated, usually by scatterplot smoothers. The smooth functions produced can then be used as a data description, for prediction, or to suggest covariate transformations.

The fit of a GAM to data requires two stages of convergence: the first is to fit an additive model, approximating the regression surface by a sum of smooth functions; the second is to fit the overall GAM. Back-fitting is used in the first stage which is then embedded in the second stage.

According to Hastie and Tibshirani (1986), the additive regression model is written as:

$$E(Y|X) = f_0 + \sum_{j=1}^{p} f_j(X_j) \qquad (4\text{-}3)$$

where the $f_j(\cdot)$'s are standardized so that $E\, f_j(X_j) = 0$.

The back-fitting algorithm to estimate f_0 and functions $f_j(\cdot)$ in (4-3) is described in the following steps:

Initialization: $f_0 = E(Y)$, $f_j^1(\cdot)(j = 1, 2, ..., p) = 0$, $m=0$;

Iterate: $m = m + 1$

For j=1 to p do:

$$R_j = Y - f_0 - \sum_{k=1}^{j-1} f_k^m(X_k) - \sum_{k=j+1}^{p} f_k^{m-1}(X_k)$$

$$f_j^m(X_j) = E(R_j \mid X_j)$$

Until: RSS = $E(Y - f_0 - \sum_{k=1}^{p} f_k^m(X_k))^2$ fails to decrease.

Here, $R_j = Y - f_0 - \sum_{\substack{k=1 \\ k \neq j}}^{p} f_k(X_k)$ is defined as the partial residual. $f_j^m(\cdot)$ denotes the estimate of $f_j(\cdot)$ at the *m*th iteration. $f_j^m(\cdot)$ can be estimated from any smoothers such as a running mean, running median, running least squares line, kernel estimate, or spline.

Buja et al. (1989) showed that the back-fitting algorithm is equivalent to the Gauss-Seidel method for solving an appropriate system of normal equations. They proved that convergence is assured if the smoothers used are linear, symmetric, and shrinking.

4.4 Bandwidth selection

4.4.1 Bandwidth selection for basic GWR

For Basic GWR, where a single bandwidth is used in the model, several criteria have been proposed to select an optimal bandwidth (Fotheringham et al., 2002), such as: cross-validation approach used for local regression by Cleveland (1979) and for kernel density estimation by Bowman (1984); generalised cross-validation as used in smoothing splines by Craven and Wahba (1979); the Akaike Information Criterion (Akaike, 1973) as suggested for nonparametric regression by Hurvich et al. (1998); and Bayesian Information Criterion (Schwarz, 1978) as used for GWR by Nakaya (2001).

Cross-validation (CV) is based on the least squares criterion. The classical CV score is defined as

$$CV(b) = \sum_i (y_i - \hat{y}_{\neq i}(b))^2 \tag{4-4}$$

where y_i is the observed value of the dependent variable at location *i*, $\hat{y}_{\neq i}(b)$ is the fitted value predicted by a model with bandwidth *b*, with the observation at point *i* removed from the model. The optimal bandwidth is the one that minimizes the CV score. The CV score originates from the sum of the predicted squared errors (PSEs) and is adopted in a "leave-one-out" manner so as to prevent the calibration from wrapping around the data points.

CV is the most popular approach for bandwidth selection in GWR but may be impacted disproportionately by a small selection of highly influential points (Farber and Páez, 2007). As a remedy, Farber and Páez (2007) developed several modified versions of the CV score based on row-standardization and row-normalization and simulation experiments showed that the modified CV scores tend to find larger bandwidths than traditional CV does and produces more accurate estimates of the regression coefficients, though less accurate estimates of the dependent variable.

Generalised cross-validation (GCV) is defined as (Loader, 1999):

$$GCV(b) = \frac{n\sum_i (y_i - \hat{y}_{\neq i}(b))^2}{(n - tr(S))^2} \tag{4-5}$$

where S is the hat matrix of the model and the trace of S represents the effective number of parameters in the model. n is the total number of observations.

Akaike Information Criterion (AIC) is a general model selection criterion that provides a trade-off between goodness-of-fit and degrees of freedom. The lower the AIC is, the better is the model. In GWR, a refinement to the AIC, namely the corrected AIC is usually used (Fotheringham et al., 2002):

$$AICc = 2n\log_e(\hat{\sigma}) + n\log_e(2\pi) + n\frac{n + tr(S)}{n - 2 - tr(S)} \tag{4-6}$$

where S and n are defined as before, $\hat{\sigma}$ is the estimated standard deviation of the error term. As a general criterion, AIC or AICc can also be used to compare GWR against a global model.

Bayesian Information Criterion (BIC) is defined as:

$$BIC = -2\log_e(L) + k\log_e(n) \tag{4-7}$$

where L is the model likelihood, k is the number of parameters. BIC is similar to AIC but tends to identify models with fewer parameters in larger samples than AIC does.

All these diagnostics are measures of the overall performance of a model with a particular bandwidth, reflecting a certain bias-variance combination. The bandwidth selection process is therefore an optimization process to minimize the chosen diagnostic on a certain searching interval. Usually, the searching range is defined as from the shortest distance to the longest distance between two data locations. One can search on the interval evenly by given step length or employ more efficient function minimisation techniques such as Golden Section Search (Greig, 1980) in numerical analysis.

4.4.2 Bandwidth selection for FBGWR

For FBGWR, a vector of bandwidths needs to be specified: one for each regression coefficient (intercept inclusive). As in basic GWR, bandwidths can be supplied from *a priori* knowledge or by automatic optimization. The simplest way is to specify the bandwidths directly from the analyst's experience and knowledge about the research application. If a process is known to act within a certain distance in space, this distance can be defined as the bandwidth for the coefficient associated with the process. Processes that vary at a local spatial scale should have small bandwidths and consequently the regression coefficients produced by the model will exhibit strong local variation. On the contrary, processes that are relatively stable across space should have large bandwidths which will allow the model to estimate spatially stable coefficients. To select bandwidths automatically from the data, a more complicated optimizing strategy needs to be developed.

A straightforward idea is to specify a series of bandwidth combinations, construct a FBGWR model with each combination, assess the performance of the model based on certain criterion, and find the best model among all possibilities. This brute-force approach is feasible for FBGWR models with only a few variables; however, as the number of variables increases, the time cost will increase rapidly and the method will be inefficient to apply.

An alternative approach is to select an optimal bandwidth for each coefficient (intercept inclusive), within each step of back-fitting, employing an existing criteria for a basic GWR. The optimal bandwidths might change during the back-fitting procedure but when the procedure terminates, the bandwidths optimized in the last iteration compose the final selection of bandwidths. As this procedure alters the traditional back-fitting by using potentially different bandwidths, and different estimators in each step, the properties of convergence will need to be further examined. Also, since the bandwidth optimizing procedure is carried out in each back-fitting step for each coefficient, the whole algorithm will be slowed down.

Furthermore, there are some prior exploratory analyses that may help bandwidth selection. For example, the correlation range from a variogram (Blanco‐moreno et al., 2008) of each independent variable can suggest a bandwidth for the coefficient associated with the given independent variable. Another possibility is to use the bandwidth optimized from a univariate GWR fit with the dependent variable and only

one independent variable as the bandwidth for this same independent variable in FBGWR.

Various bandwidth selection methods will be experimented with in Chapters 5 and 6. When applying FBGWR, practitioners can combine these methods and customize their own bandwidth selection strategy according to their research goals.

4.5 Algorithm development

Rewrite equation (4-2) into a vector form:

$$\boldsymbol{y} = \boldsymbol{\beta}_{bw0} + \boldsymbol{\beta}_{bw1} \otimes \boldsymbol{x}_1 + \boldsymbol{\beta}_{bw2} \otimes \boldsymbol{x}_2 + \ldots + \boldsymbol{\beta}_{bwm} \otimes \boldsymbol{x}_m + \boldsymbol{\varepsilon} \quad (4\text{-}8)$$

The bold type denotes a vector, all the vectors in the equation have the same length equal to the number of observations to be modelled. The symbol ⊗ denotes a multiplication operator between two vectors in the same length where each element of one vector is multiplied by the corresponding element of the other vector and the results construct a new vector.

A back-fitting algorithm for calibrating FBGWR in form of equation (4-8) is designed as following:

1) Give an initial guess for all the terms on the right hand side: $\boldsymbol{\beta}_{bw0}$, $\boldsymbol{\beta}_{bw1} \otimes \boldsymbol{x}_1$, $\boldsymbol{\beta}_{bw2} \otimes \boldsymbol{x}_2$, … , $\boldsymbol{\beta}_{bwm} \otimes \boldsymbol{x}_m$, name them as $\hat{\mathbf{f}}_0^{(0)}$, $\hat{\mathbf{f}}_1^{(0)}$, $\hat{\mathbf{f}}_2^{(0)}$, …, $\hat{\mathbf{f}}_m^{(0)}$;

   Set the initial residual $\boldsymbol{\varepsilon}^{(0)} = \boldsymbol{y} - \hat{\mathbf{f}}_0^{(0)} - \hat{\mathbf{f}}_1^{(0)} - \hat{\mathbf{f}}_2^{(0)} - \ldots - \hat{\mathbf{f}}_m^{(0)}$.

2) Set the iteration count $k = 0$;

   Set the maximum iterations (MI) as a pre-specified value.

3) If $k >$ MI, terminate;

   Otherwise, regress the partial residual against each independent variable in turn:

   Regress $\hat{\mathbf{f}}_0^{(k)} + \boldsymbol{\varepsilon}^{(k)}$ against $\mathbf{I}$ (a vector of 1s with the length equal to n), get a new fitted value $\hat{\mathbf{f}}_0^{(k+1)}$, and an updated residual $\boldsymbol{\varepsilon}_0^{(k)}$;

   Regress $\hat{\mathbf{f}}_1^{(k)} + \boldsymbol{\varepsilon}_0^{(k)}$ against $\boldsymbol{x}_1$, get a new fitted value $\hat{\mathbf{f}}_1^{(k+1)}$, and an updated residual $\boldsymbol{\varepsilon}_1^{(k)}$;

   Regress $\hat{\mathbf{f}}_2^{(k)} + \boldsymbol{\varepsilon}_1^{(k)}$ against $\boldsymbol{x}_2$, get a new fitted value $\hat{\mathbf{f}}_2^{(k+1)}$, and an updated residual $\boldsymbol{\varepsilon}_2^{(k)}$;

   …

Regress $\hat{\mathbf{f}}_m^{(k)} + \boldsymbol{\varepsilon}_{m-1}^{(k)}$ against $\boldsymbol{x}_m$, get a new fitted value $\hat{\mathbf{f}}_m^{(k+1)}$, and an updated residual $\boldsymbol{\varepsilon}_m^{(k)}$.

Assign $\boldsymbol{\varepsilon}^{(k+1)} = \boldsymbol{\varepsilon}_m^{(k)}$.

4) Calculate the score of change (SOC) after the k th iteration based on a certain criterion. If SOC is greater than a pre-specified termination threshold, set $k = k + 1$ and go to 3);

   Otherwise, if SOC is below the threshold, terminate.

The estimated coefficients from the last iteration comprise the final estimated coefficients.

$\hat{\boldsymbol{y}}^{(k+1)} = \hat{\mathbf{f}}_0^{(k+1)} + \hat{\mathbf{f}}_1^{(k+1)} + \hat{\mathbf{f}}_2^{(k+1)} + \ldots + \hat{\mathbf{f}}_m^{(k+1)}$ is the final fitted value for $\boldsymbol{y}$.

To implement the algorithm, a hierarchy of two modules is developed.

1) Basic univariate GWR: fit each individual term by regressing the partial residual against a single independent variable. The intercept term is considered as a parameter associated with an independent variable of all 1s. No intercept term will be added to other "real" independent variables, which allows the intercept in the fit to be absorbed in the residual.
2) Back-fitting algorithm: cycles through the individual terms in the model and updates each term using a basic univariate GWR fit. The cycles continue until the change after a whole iteration is small enough.

There are two main decisions need to be made for the back-fitting FBGWR algorithm: the choice of initial guesses, and the choice of termination criterion.

4.5.1 Choice of initial guesses

Various initial guesses can be supplied to $\hat{\mathbf{f}}_0^{(0)}$, $\hat{\mathbf{f}}_1^{(0)}$, $\hat{\mathbf{f}}_2^{(0)}$, …, $\hat{\mathbf{f}}_m^{(0)}$ to start the back-fitting. Some possibilities are:

1) Set $\hat{\mathbf{f}}_0^{(0)} = \hat{\mathbf{f}}_1^{(0)} = \hat{\mathbf{f}}_2^{(0)} = \ldots = \hat{\mathbf{f}}_m^{(0)} = 0$. These are the initial values used by Hastie and Tibshirani (1986) in their back-fitting to fit a generalized additive model (GAM).
2) Use the results from a basic GWR as initial guesses. Fit a basic GWR model as in equation (4-1) and rewrite the equation in a vector form:

$$\boldsymbol{y} = \boldsymbol{\beta}_0 + \boldsymbol{\beta}_1 \otimes \boldsymbol{x}_1 + \boldsymbol{\beta}_2 \otimes \boldsymbol{x}_2 + \ldots + \boldsymbol{\beta}_m \otimes \boldsymbol{x}_m + \boldsymbol{\varepsilon} \qquad (4\text{-}9)$$

The notations follow the same convention as in equation (4-8). Denote the estimated coefficients as $\widehat{\boldsymbol{\beta}}_0$, $\widehat{\boldsymbol{\beta}}_1$, $\widehat{\boldsymbol{\beta}}_2$, …, $\widehat{\boldsymbol{\beta}}_m$, set

$\hat{\mathbf{f}}_0^{(0)} = \widehat{\boldsymbol{\beta}}_0$

$\hat{\mathbf{f}}_1^{(0)} = \widehat{\boldsymbol{\beta}}_1 \otimes \boldsymbol{x}_1$

$\hat{\mathbf{f}}_2^{(0)} = \widehat{\boldsymbol{\beta}}_2 \otimes \boldsymbol{x}_2$

…

$\hat{\mathbf{f}}_m^{(0)} = \widehat{\boldsymbol{\beta}}_m \otimes \boldsymbol{x}_m$

3) Use the results from a multiple linear regression as initial guesses.

Different initial guesses may or may not yield different final results when the convergence is reached after sufficient iterations but will definitely affect the speed of convergence. Experiments with different initial guesses will be demonstrated in Chapter 5.

4.5.2 Choice of termination criterion

Hastie and Tibshirani (1986) used RSS to detect the convergence of back-fitting in their GAM fit. When RSS fails to decrease, the back-fitting procedure should terminate. In practice, the change between RSS from two successive iterations is monitored and once the change is below a pre-specified termination threshold, the procedure is deemed to converge. The score of change is defined as:

$$\text{SOC} = \text{RSS}new - \text{RSS}old \tag{4-10}$$

There are some other criteria that could be used to detect convergence. In the R package gam (Hastie, 2011) for fitting GAM models, the relative change in the estimates of smooth functions $f_j(\cdot)$ between two iterations is used. The score of change is calculated by the following formula:

$$\text{SOC} = \sqrt{\frac{\sum_{j=1}^{p} \frac{\sum_{i=1}^{n} (\hat{\mathbf{f}}_{ij}^{new} - \hat{\mathbf{f}}_{ij}^{old})^2}{n}}{\sum_{i=1}^{n} (\sum_{j=1}^{p} \hat{\mathbf{f}}_{ij}^{new})^2}} \tag{4-11}$$

where n is the number of observations, p is the number of terms to be fitted in the model, $\hat{\mathbf{f}}_{ij}^{new}$ is the fitted value for the jth term at observation i in current iteration

and $\hat{\mathbf{f}}_{ij}^{\,old}$ is the fitted value in the last iteration. The SOC is compared with the threshold at each iteration step to detect convergence.

In FBGWR, the change in estimated coefficients at each regression point can also be used with the SOC is calculated by a similar formula to (4-11) except replacing y_{nj} with the estimated coefficients. The approach is experimented through simulated datasets in the following chapter.

Chapter 5 Experiment on synthetic datasets

This chapter presents a series of experiments that were carried out on synthetic datasets to verify the algorithm proposed in Chapter 4 and to demonstrate the strengths and weaknesses of FBGWR.

5.1 Experimental design

5.1.1 Simulation study

Datasets used in this chapter are synthetic datasets that were generated from controlled simulations. Various simulation studies have been used in GWR-related research as demonstrated by Leung et al. (2000), Wheeler and Tiefelsdorf (2005), Mei et al. (2006), Farber and Paez (2007), Wang et al. (2008) and Harris et al. (2010). Simulation data are often preferred to empirical data in the study of a newly developed method because the properties of a simulated dataset can be well controlled and specific designs can be constructed for certain testing purposes.

For regression methods the true values of regression coefficients can be defined to generate a dataset; regression models can then be calibrated on the dataset to investigate the performance of the calibration method in coefficient estimation. In GWR the spatial non-stationarity of coefficients as well as the scale of non-stationarity can be controlled and the performance of GWR, FBGWR and other forms of GWR can then be evaluated and compared. While in an empirical study, although *a priori* knowledge and experience can help understand the spatial process under study, there are rarely accurate quantitative measurements of the relationships in the process and there are always possibilities of missing knowledge about the relationships. Estimated coefficients based on empirical datasets can also be obscured by artifacts (Wheeler, 2006) so that model evaluation can be sometimes arbitrary and misleading.

Therefore, a simulation study is carried out here for FBGWR and once the practicability and reliability of the model has been established, the study will move to a real-world dataset to demonstrate the utility of the method.

5.1.2 Dataset design

The model to generate the simulation datasets follows the experimental design by Wang et al. (2008) where the performance of a traditional GWR was compared with a modified GWR using local spatially expanded coefficients. Similar datasets have also been employed by Farber and Paez (2007) to test the goodness of fit and the ability of GWR to retrieve the coefficients of a spatially varying process and by Harris et al. (2010a) to evaluate and compare the performance of GWR in spatial prediction with other traditional geostatistical models.

The spatial region of the datasets is a two-dimensional grid consisting of 25 * 25 observation points, the distance between every two adjacent points along both horizontal and vertical axes is half of the unit distance. The grid is illustrated in Figure 5-1. The coordinate (u, v) of the point at the lower left corner is (0, 0), while the point at the upper right corner has coordinate (12, 12).

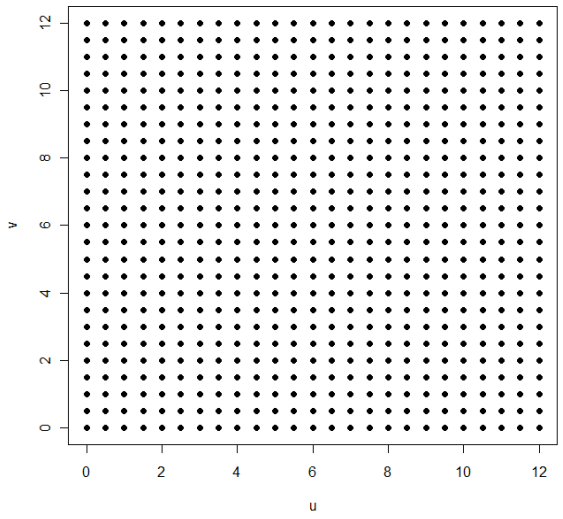


Figure 5-1 A grid for generating the simulation dataset

The data generating process is described as:

$$y = \beta_1 x_1 + \beta_2 x_2 + \varepsilon \tag{5-1}$$

where y is the dependent variable in a simple linear relationship with two independent variables x_1 and x_2 which are independently and randomly drawn from a uniform distribution with minimum 0 and maximum 1. ε is the error term randomly drawn from a normal distribution with zero mean and a variance equal to a certain proportion of the variance of the mean process, that is, the variance of $\beta_1 x_1 + \beta_2 x_2$. In the experiments, various levels of variance in error are employed for comparison.

The two coefficients β_1 and β_2 are defined as functions of the coordinate (u, v) of each observation point and a series of functions are designed to demonstrate various levels of spatial heterogeneity.

Note that there is no intercept in the data generating model so in the regression model to be calibrated there will be no intercept to be fitted either. The same idea has been used in Wheeler and Calder (2007) and Wheeler (2009).

Alternatively, x_1 can be a constant 1 so that β_1 represents an intercept term and equation (5-1) turns into a linear model with only one independent variable.

To better illustrate the data generating process, an example is presented here with x_1 and x_2 being random independent variables and β_1 is defined as a quadratic surface with β_2 fixed as a constant. The exact equations used to define β_1 and β_2 are:

$$\beta_1 = 1 + \left(\frac{1}{324}\right)[36 - (6-u)^2][36 - (6-v)^2] \quad (5\text{-}2)$$

$$\beta_2 = 3 \quad (5\text{-}3)$$

The variance of ε equals 1/3 of the variance of $\beta_1 x_1 + \beta_2 x_2$. The complete data generating process is visualized in two dimensional space in Figure 5-2 and in three dimensions in Figure 5-3. Note that in Figure 5-2 each plot is displayed in its own scale to present the variability or randomness of each element more clearly, while in Figure 5-3 all plots are re-scaled to the same level so that they can be compared more easily.

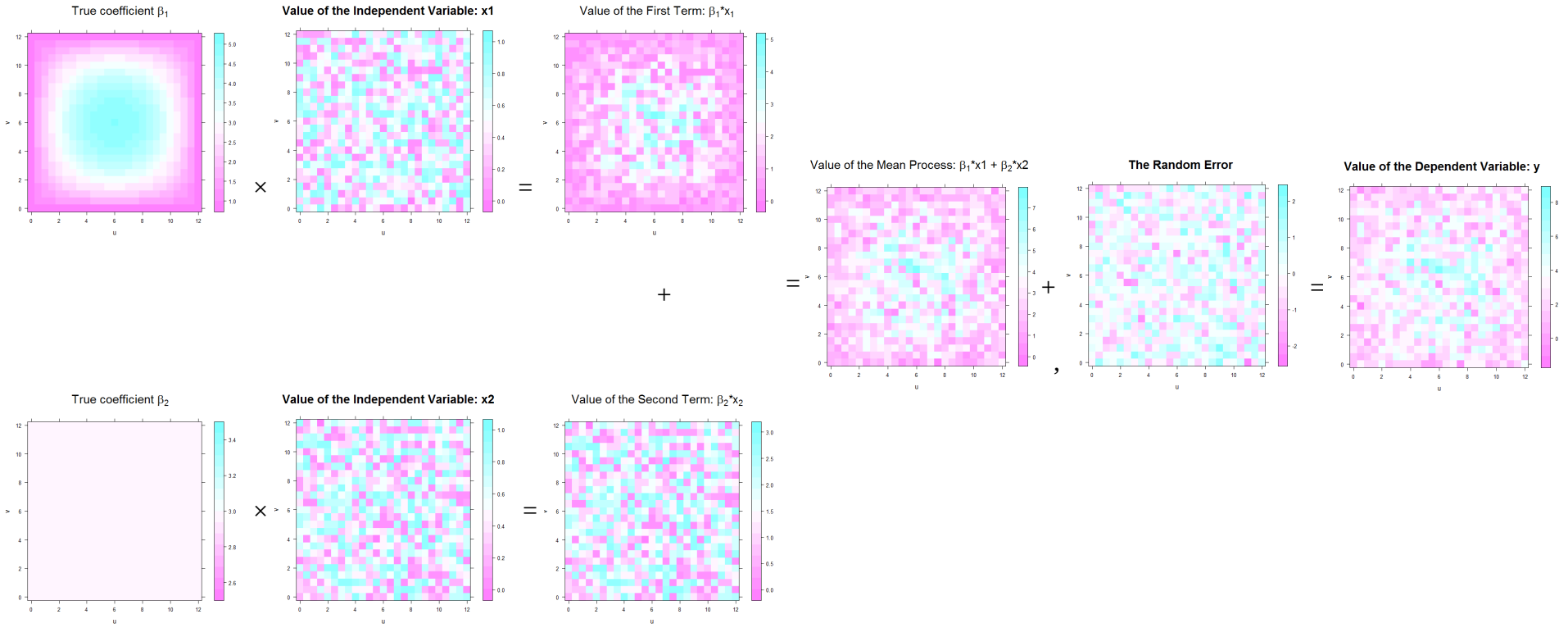


Figure 5-2 Illustration of the data generating process (2D)

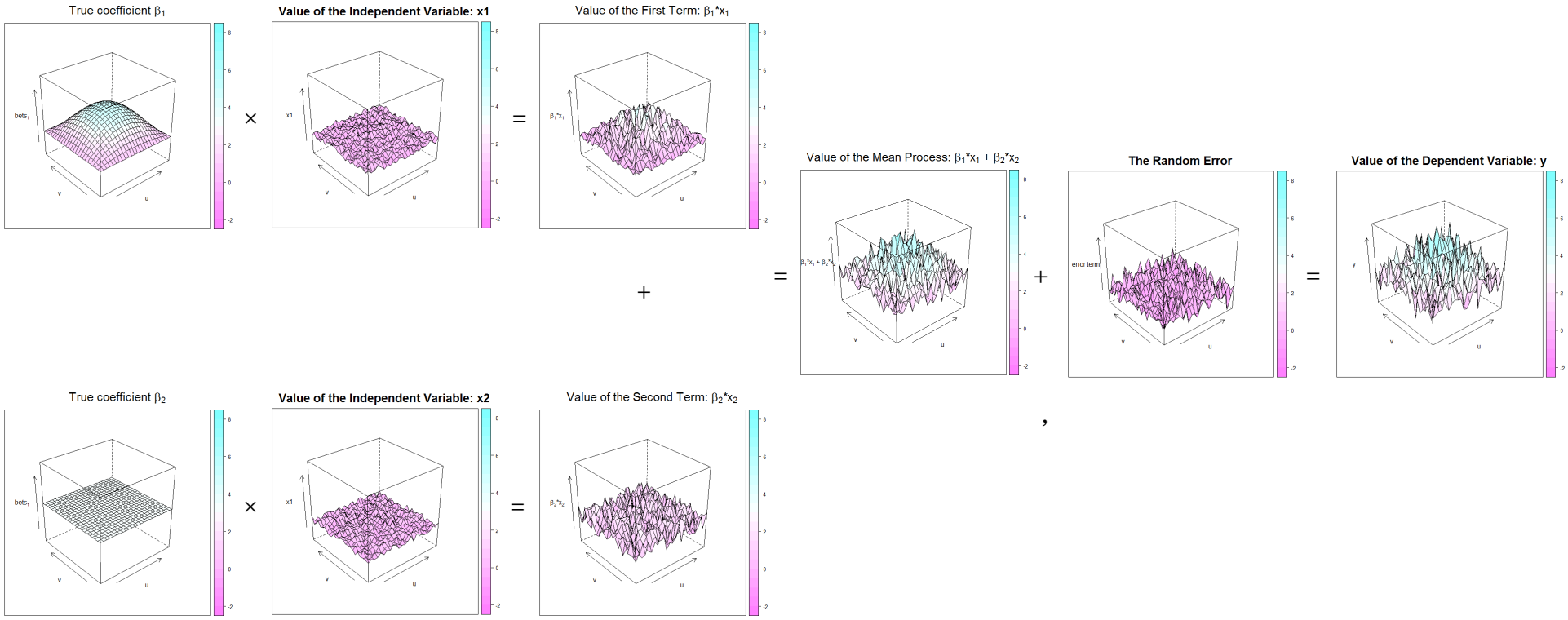


Figure 5-3 Illustration of the data generating process (3D)

5.1.3 Model evaluation

Throughout the simulation experiments various aspects of the regression models were examined to evaluate the model performance. The most important attention is paid to the bandwidths used in each model as the choice of bandwidth is the focus of FBGWR and affects the performance of GWR models considerably.

The second important concern is how the models compare in estimating the coefficients and reproducing the spatial variability of the coefficients. Since the true coefficients are known, the coefficients estimates are compared with the true coefficients and the deviance measured. Here the root mean square error (RMSE) is mainly employed to measure the accuracy of the coefficient estimates. The RMSE is the square root of the mean of the squared deviations of the estimates from the true values and it should be small for an accurate estimator. The calculation of RMSE is as below:

$$\mathrm{RMSE} = \sqrt{(1/n)\sum_{i=1}^{n}(z_i - \hat{z}_i)^2} \qquad (5\text{-}4)$$

where n is the number of observations, z_i is the true value at observation i, $\hat{z}_i$ is the estimate at the same observation. The deviance of the fitted value of dependent variable y from the true value is also measured and the RMSE of y compared. To facilitate an intuitive comparison, the surfaces of the coefficient estimates as well as of the true coefficients are also visualized. There are some cases where the surfaces estimated from different models are so similar that it is hard to tell simply from the maps which model provides a better estimate, in these cases, the quantitative comparison of RMSE has to be resorted to.

In terms of the goodness of fit, the residual sum of squares (RSS) is measured for each model. Although AIC presents a better criterion for the overall performance of model, this measurement cannot be calculated for FBGWR at the moment because the number of independent parameters cannot be defined for FBGWR. Instead, the AIC score of each fitted term composing an FBGWR model can be calculated separately and summed up as a total AIC value. The efficiency of the method is also reported in terms of the computation time required in the algorithm.

5.1.4 Experiment environment

All experiments in this chapter were implemented in the free software environment R (R Development Core Team, 2011), with the version R 2.13.0. Packages employed

include spgwr (Bivand and Yu, 2012) for basic GWR calibration and bandwidth optimization, gam (Hastie, 2011) for back-fitting, geoR (Diggle and Ribeiro, 2007) for data simulation, and lattice (Sarkar, 2008) for visualization.

The hardware where the experiments were conducted on is a Dell Precision T3500, with Processor: Intel(R) Xeon(R) CPU W3530 @2.80GHz 2.80GHz, RAM: 4.00 GB, OS: Windows 7 Ultimate, Service Pack 1, 64-bit.

5.2 Bandwidth selection

The core objective of FBGWR is to specify an appropriate bandwidth for each regression coefficient associated with each independent variable as well as the intercept. Although analysts can specify *a priori* bandwidths for FBGWR according to their experience and knowledge about the spatial processes under study, it will be helpful to let the algorithm automatically select optimal bandwidths from the data. In this section, various bandwidth selection strategies are experimented with using synthetic datasets and the results discussed.

5.2.1 Bandwidth optimization by brute force

The most straightforward method for bandwidth optimization is to configure FBGWR model with various combinations of possible bandwidths and to select the bandwidths that yield the best model performance. In this experiment, a series of values were assigned to each bandwidth in the FBGWR model in turn and the overall performance of the model with each different combination of bandwidths is evaluated in terms of the total AIC value of the model.

The dataset used here was generated from a model with two independent variables and no intercept:

$$y = \beta_1 x_1 + \beta_2 x_2 + \varepsilon \tag{5-5}$$

where x_1 and x_2 are randomly and independently drawn from a uniform distribution with minimum 0 and maximum 1. ε was randomly drawn from a normal distribution with zero mean and a variance equal to 1/3 of the variance of the mean process.

The two coefficients β_1 and β_2 are defined as:

$$\beta_1 = 1 + \left(\frac{1}{324}\right)[36 - (6-u)^2][36 - (6-v)^2] \tag{5-6}$$

$$\beta_2 = 3 \tag{5-7}$$

and the surfaces of β_1 and β_2 are visualized as following:

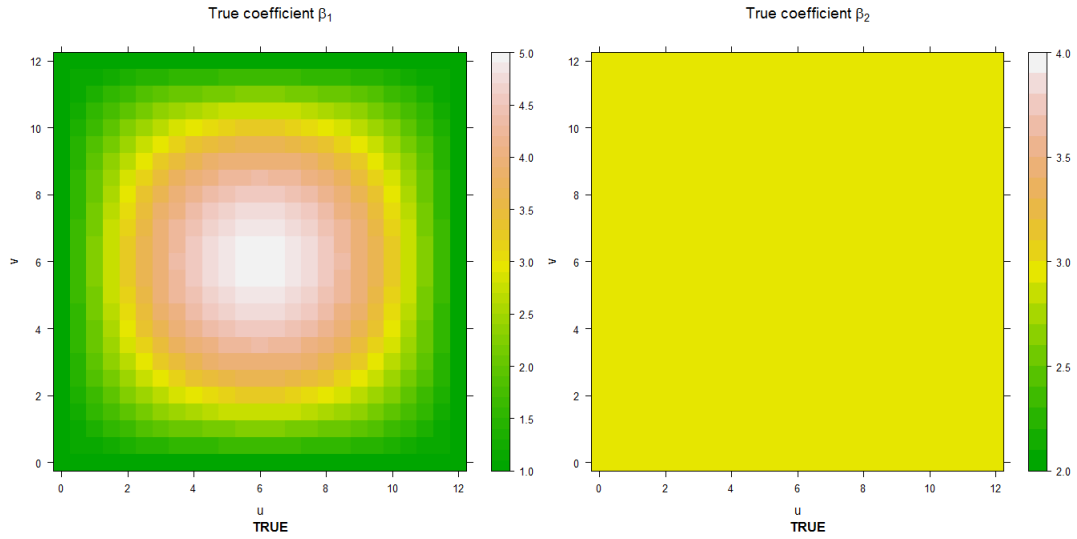


Figure 5-4 Surfaces of the true coefficients

The kernel function used here is a bi-square function. The bandwidth takes the form of an adaptive bandwidth, defined as the proportion of data points involved in the calibration of a local estimate. The same kernel function and adaptive bandwidth were used in the remaining experiments in this chapter. This adaptive kernel is chosen because although for datasets with observation points located on regular grids, fixed and adaptive bandwidths are virtually equivalent, the adaptive bandwidth in the form of a proportion is simpler to understand.

The bandwidths for x_1 and x_2, bw_1 and bw_2, were taken from a geometric progression of length 6: 1, 0.5, 0.25, 0.125, 0.0625, 0.03125, this means that there were 6*6=36 different combinations of bandwidths (bw_1, bw_2): (1, 1), (1, 0.5), (1, 0.25), (1, 0.125), (1, 0.0625), (1, 0.03125), …, (0.03125, 0.03125). Each combination was employed to construct a FBGWR model and the total AIC score and RSS are recorded in the following table:

Table 5-1 Total AIC and RSS from various bandwidth combinations

	bw_1	bw_2	Total AIC	RSS	Time Cost[1] (seconds)
1	0.03125	0.03125	3036	*279*	888
2	0.03125	0.0625	3025	312	658
3	0.03125	0.125	3018	322	652
4	0.03125	0.25	3013	327	633
5	0.03125	0.5	3011	330	641
6	0.03125	1	*3008*	331	480
7	0.0625	0.03125	3049	319	750
8	0.0625	0.0625	3079	369	623
9	0.0625	0.125	3079	383	579
10	0.0625	0.25	3071	388	608
11	0.0625	0.5	3067	391	555
12	0.0625	1	3065	392	649
13	0.125	0.03125	3068	336	730
14	0.125	0.0625	3097	389	609
15	0.125	0.125	3109	408	638
16	0.125	0.25	3104	414	615
17	0.125	0.5	3100	416	609
18	0.125	1	3099	418	613
19	0.25	0.03125	3149	366	704
20	0.25	0.0625	3169	420	659
21	0.25	0.125	3181	440	634
22	0.25	0.25	3181	449	709
23	0.25	0.5	3185	454	639
24	0.25	1	3195	461	694
25	0.5	0.03125	3270	406	717
26	0.5	0.0625	3289	467	658
27	0.5	0.125	3302	489	674
28	0.5	0.25	3310	502	566
29	0.5	0.5	3337	518	598
30	0.5	1	3388	543	769
31	1	0.03125	3352	436	737
32	1	0.0625	3379	504	637
33	1	0.125	3401	533	566
34	1	0.25	3430	556	615
35	1	0.5	3503	594	659
36	1	1	3616	654	651

[1] The time cost is measured by "user time" for the execution of user instructions of the process. The same measurement is used in other experiments in this thesis.

While smaller bandwidths usually result in smaller RSS values, this is not the case for AIC. The total AIC scores against bandwidths are visualized in a 3-D scatterplot as in Figure 5-5:

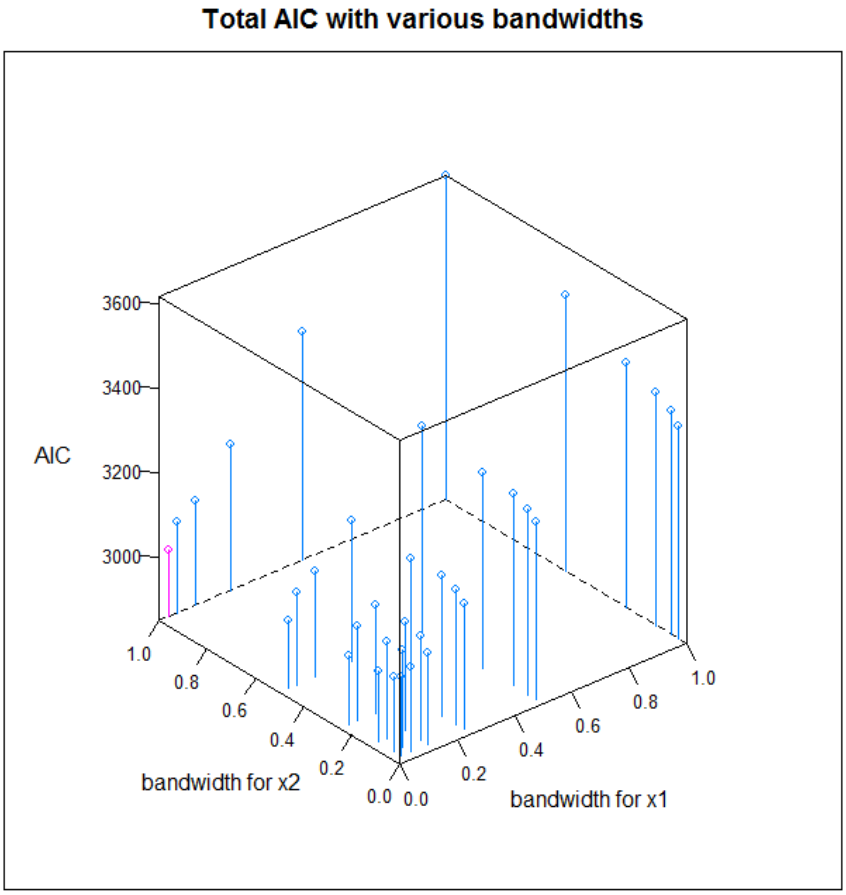


Figure 5-5 Scatterplot of total AIC score against various bandwidths

The minimum total AIC score was observed with bandwidth 0.03125 and 1, which suggested the best bandwidth combination for this dataset: a rather small bandwidth for x_1 and an extremely large one for x_2. This corresponds to the fact that β_1 varies at a local scale whilst β_2 is constant across the space. The FBGWR model fitted with these two bandwidths is further examined by plotting the residuals and a normal QQ plot, shown in Figure 5-6. The plots confirm that the residuals are normally distributed and that the model is a non-biased one.

The problem with this bandwidth optimization method is that it is rather time-consuming; the total time cost with this dataset was 23,416 seconds, or 6.5 hours. One FBGWR model needs to be fitted for every combination of bandwidths. Suppose there are *k* different bandwidths to be specified in the model. For each bandwidth,

there are m possible values to choose from and then there will be m^k different combinations of bandwidth, each in the form of a k dimensional bandwidth vector (bw_1, $bw_{2, \ldots,} bw_k$). The whole optimization procedure will involve k^m executions of back-fitting. When k or m is large, the method will be unacceptably slow. More efficient methods are required.

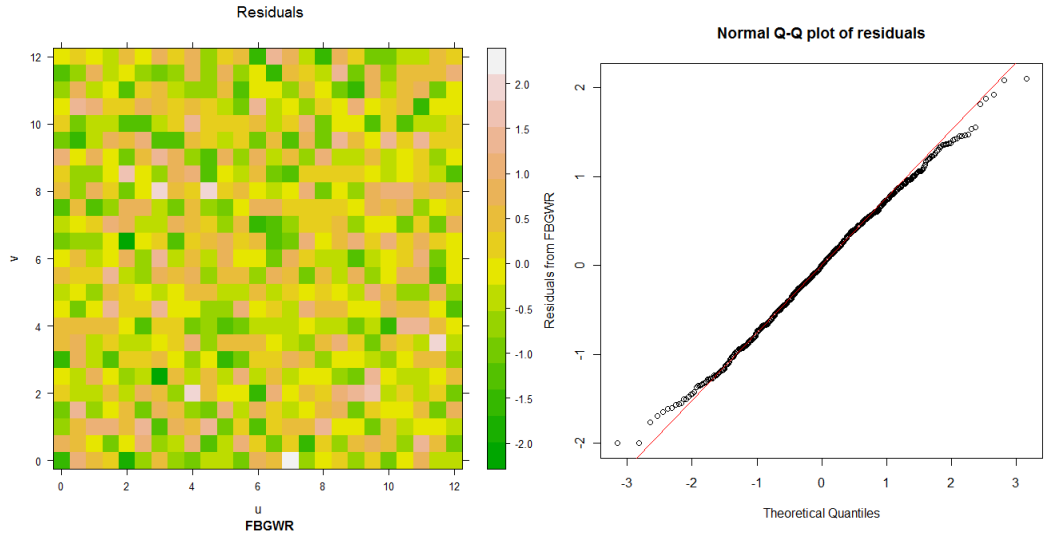


Figure 5-6 Surface of residuals (left) and the normal Q-Q plot (right)

5.2.2 Bandwidth optimization marginally

An alternative method for bandwidth optimization is to fix one bandwidth, bw_1, and let the other bandwidth, bw_2, be chosen automatically according to a certain criterion, such as AIC. As in the previous method, a series of values can be assigned to bw_1, each yielding a corresponding optimal bw_2 and a fitted FBGWR model. Then reverse the order, fix bw_2 and let bw_1 be chosen automatically. The resulting fitted models can be compared and the best one chosen.

An experiment using this technique was carried out on the same dataset as in 5.2.1 and with the same sequence of values (1, 0.5, 0.25, 0.125, 0.0625, 0.03125) assigned to each bandwidth respectively. Table 5-2 compares the total AIC score and RSS of the resulting 12 fitted models. The AIC scores against various bandwidths are also visualized in a 3-D plot in Figure 5-7. The lowest total AIC score was achieved by the model with bandwidth combination of 0.03125 and 0.9999, almost the same result as in 5.2.1. The lowest RSS was observed with the same bandwidth combination, while the second lowest RSS was gained by the bandwidth combination of 0.1455 and 0.03125, with the smallest possible bandwidth for bw_2.

Table 5-2 Total AIC and RSS from various bandwidth combinations (Case 1-6: bw_1 fixed, bw_2 automatically optimized; Case 7-12: bw_2 fixed, bw_1 automatically optimized)

	bw_1(fixed)	bw_2(optimized)	Total AIC	RSS	Time Cost (seconds)
1	0.03125	0.9999	*3008*	*331*	*5452*
2	0.0625	0.9999	3065	392	5557
3	0.125	0.9999	3099	418	6882
4	0.25	0.232	3181	448	17591
5	0.5	0.1632	3305	495	5124
6	1	0.1568	3408	540	4490
	bw_1(optimized)	bw_2(fixed)	Total AIC	RSS	Time Cost (seconds)
7	0.1455	0.03125	3080	341	7067
8	0.1455	0.0625	3107	394	5885
9	0.112	0.125	3104	405	5599
10	0.1168	0.25	3101	412	4460
11	0.1115	0.5	3094	414	3591
12	0.1114	1	3092	415	4916

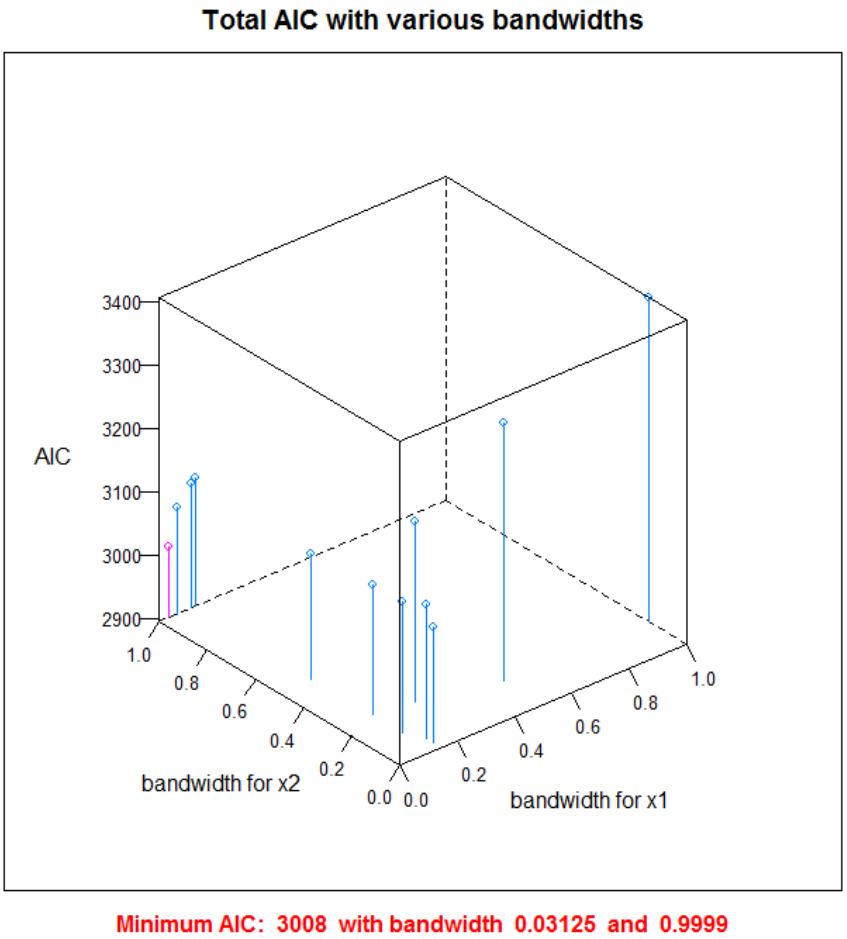


Figure 5-7 Scatterplot of total AIC score against various bandwidths

However, the time cost in this method is also very high with 76,614 seconds (21.3 hours) for the 12 trials. This is because each trial involves an automatic bandwidth optimization procedure through back-fitting which is computationally intensive. It would be more complicated and time-consuming if the method is applied to a model with more than two bandwidths.

5.2.3 Bandwidth optimization synchronously

A third method is to optimize the bandwidths synchronously, allowing the bandwidths to be updated iteratively through the back-fitting process. In the initial loop of back-fitting, an optimal bandwidth is selected for bw_1 and then bw_2 based on a certain criterion, such as AIC and CV. When the back-fitting process moves on to the next loop, as the estimates are updated the two bandwidths are also updated by two new optimal bandwidths. The updates continue until the whole back-fitting terminates. The bandwidths selected in the final loop are considered the optimal bandwidths for the dataset. To illustrate the procedure, assume the FBGWR model is written as:

$$y_i = \beta_{bw1}(u_i, v_i)x_{i1} + \beta_{bw2}(u_i, v_i)x_{i2} + \varepsilon_i . \tag{5-8}$$

Rewrite the model in a matrix form:

$$\mathrm{Y} = \boldsymbol{\beta}_1\mathrm{X}_1 + \boldsymbol{\beta}_2\mathrm{X}_2 + \boldsymbol{\varepsilon}, \tag{5-9}$$

denote the first term $\boldsymbol{\beta}_1\mathrm{X}_1$ with $\mathbf{f}_1(\mathrm{X}_1)$, and the second term $\boldsymbol{\beta}_2\mathrm{X}_2$ with $\mathbf{f}_2(\mathrm{X}_2)$. The back-fitting process embedded with bandwidth optimization can be depicted in Figure 5-8:

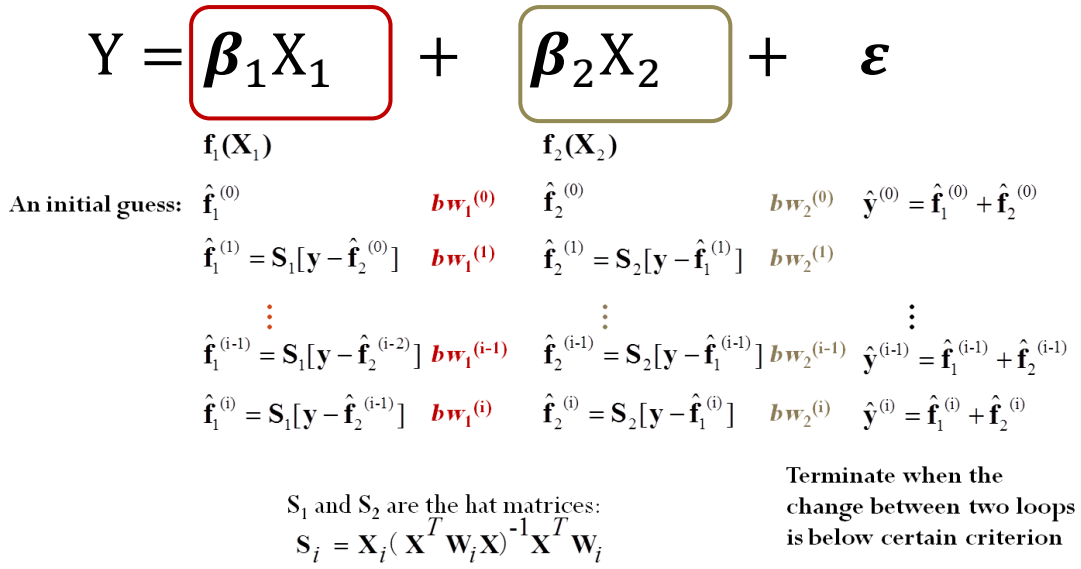
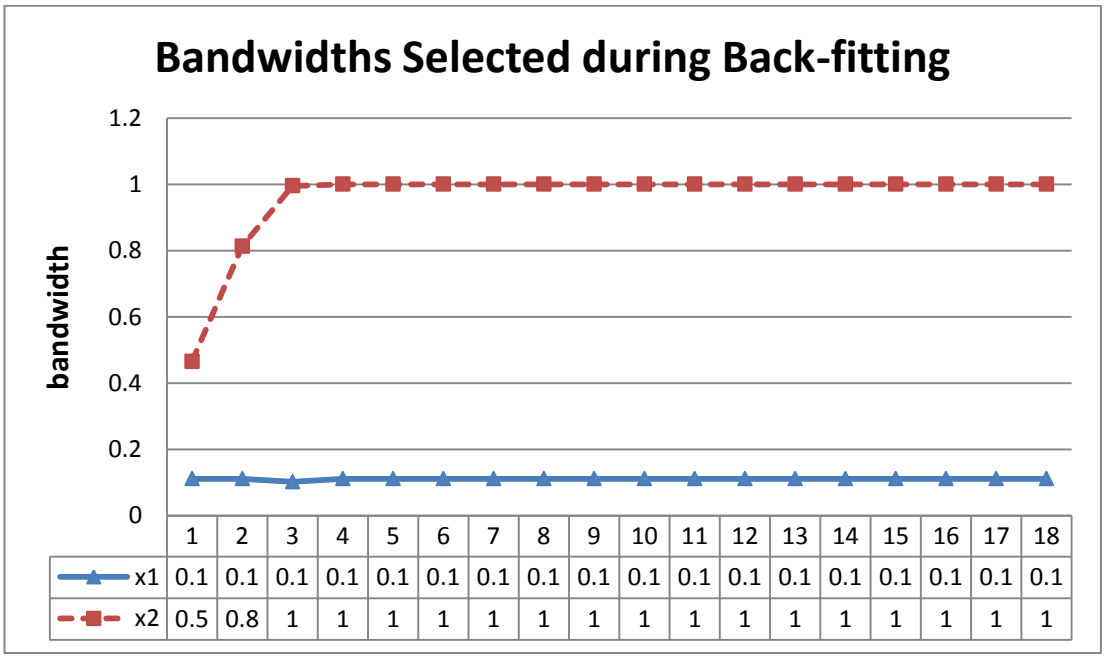


Figure 5-8 Procedure of iterative bandwidth selection through back-fitting

The whole procedure can be run automatically with no initial bandwidths required to be specified in advance: a reasonable interval that defines the lower and upper boundary for bandwidth searching will be enough to initiate the optimization.

Experiments were conducted on the same dataset as before. In each loop of back-fitting, the AIC score was employed as the criterion for bandwidth optimization and the optimization was carried out through the R function *optimize ()* (R Development Core Team, 2011) which combines golden section search and successive parabolic interpolation to find an optimal bandwidth. As an adaptive bandwidth was used, the intuitive interval (0,1) was defined for the bandwidth search.

Figure 5-9 shows the updates of each bandwidth during back-fitting. After the first four loops, both bandwidths stayed at the same values, with 0.11 for bw_1 and approximate 1 (the actual figure: 0.9999474) for bw_2. The back-fitting process terminated when the relative change in estimates between two successive loops fell below the predefined critical value of epsilon = 1e-7. The convergence process is shown in the left part of Figure 5-11. The total AIC value achieved in the final model is 3092, with 1568 from the estimation of x_1, and 1524 from the estimation of x_2. The RSS of the model is 415.

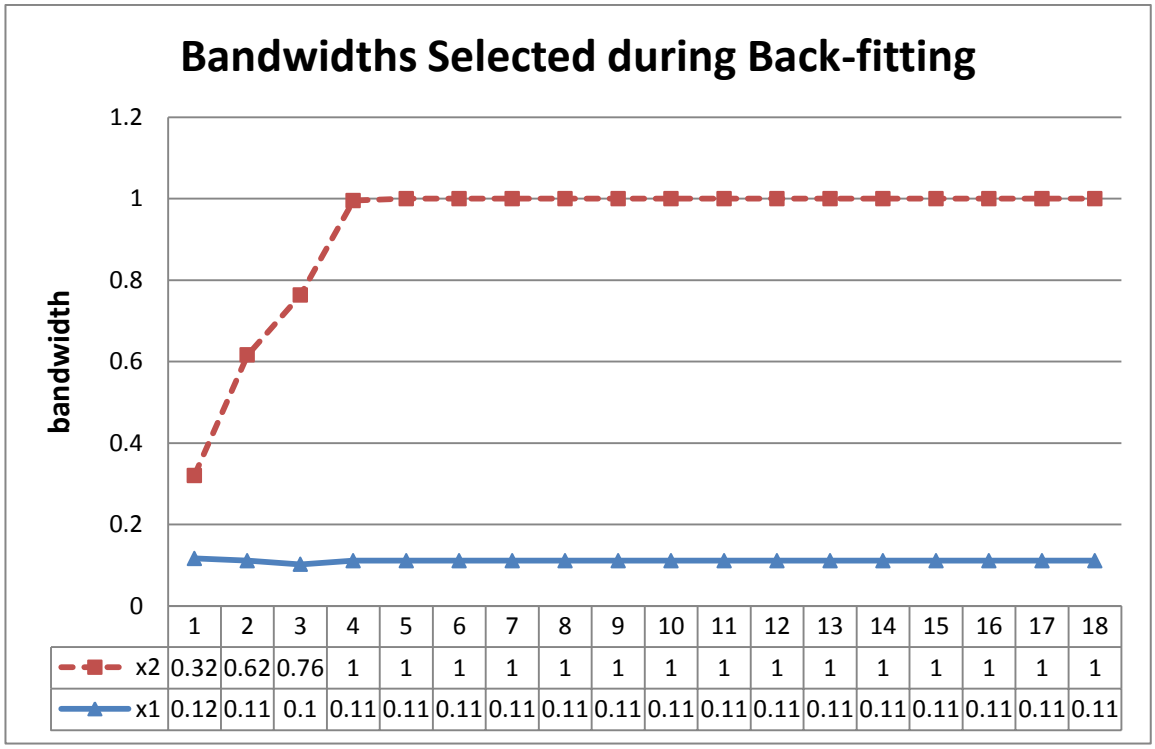


	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
x1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1
x2	0.5	0.8	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1

Figure 5-9 Updates of bandwidths selected during back-fitting

During back-fitting, the calibration of each sub-model interacts with the calibrations of other sub-models through the partial residual; therefore, the optimal bandwidth for each coefficient is also affected by the bandwidths chosen for other coefficients. To

examine the effect of this interaction, the order of x_1 and x_2 was exchanged in the model, that is, calibrate a model $Y = \boldsymbol{\beta}_2 X_2 + \boldsymbol{\beta}_1 X_1 + \boldsymbol{\varepsilon}$, instead of $Y = \boldsymbol{\beta}_1 X_1 + \boldsymbol{\beta}_2 X_2 + \boldsymbol{\varepsilon}$, the same back-fitting process was carried out, the updates of bandwidths are shown in Figure 5-10 and the convergence of the back-fitting is shown in the right part of Figure 5-11. Although different bandwidths were chosen in the first several loops, the updates stopped at the same values: 0.11 for bw_1 and approximately 1 (the same actual figure: 0.9999474) for bw_2, and the same AIC scores and RSS were achieved. The results therefore indicate some degree of stability in the algorithm.



	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
x2	0.32	0.62	0.76	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
x1	0.12	0.11	0.1	0.11	0.11	0.11	0.11	0.11	0.11	0.11	0.11	0.11	0.11	0.11	0.11	0.11	0.11	0.11

Figure 5-10 Updates of bandwidths selected during back-fitting (with X_1 and X_2 swapped in the model)

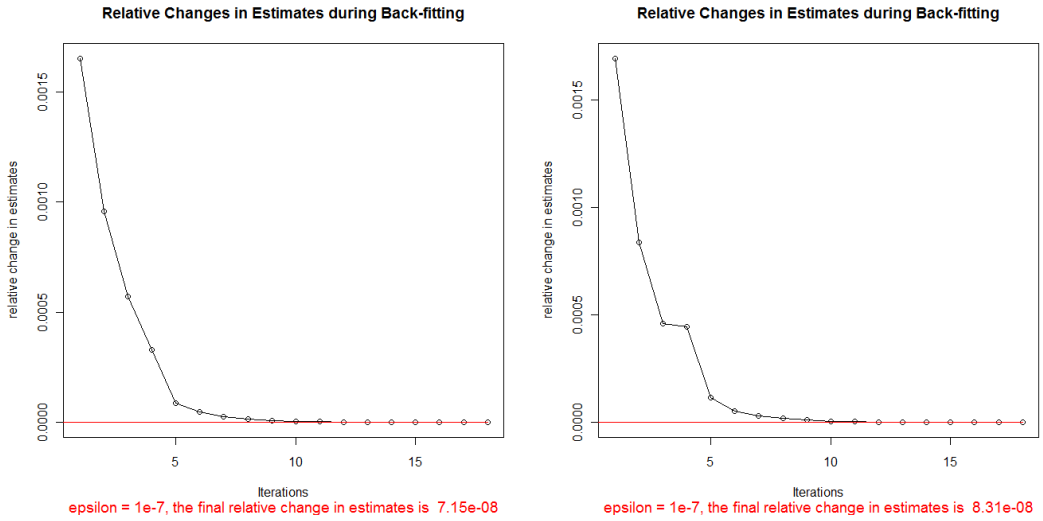


Figure 5-11 Convergence of the back-fitting processs

(Left: FBGWR model with X_1 before X_2; Right: FBGWR model with X_2 before X_1)

The time cost for this method was 9,870 seconds (2.7 hours) with the original model, and 9,381 seconds (2.6 hours) with the model with x_1 and x_2 were swapped. This method is much more efficient than the previous two and it is automatic and easy to implement without the requirement of specifying certain values for any bandwidth.

There is, however, one issue with this method that has been revealed through abundant replicated simulation experiments which is that the convergence of the algorithm cannot always be assured. For example, when the dataset as designed in equation (5-1) was generated separately 10 times and the same back-fitting algorithm with synchronous bandwidth optimization scheme embedded was executed on the 10 individual datasets, there were 4 datasets on which the algorithm failed to converge. That is, after the maximum number of iterations allowed by the algorithm (pre-defined as 30 here) had been reached, the relative change in estimates from two successive iterative steps was still above the pre-defined threshold (1e-7 here); the reason for choosing this value will be demonstrated in Section 5.4).

Figure 5-12 below displays an example where the convergence was not reached. The relative change in estimates during back-fitting failed to fall below the threshold of 1e-7 although the value was already quite small at 3.21e-05.

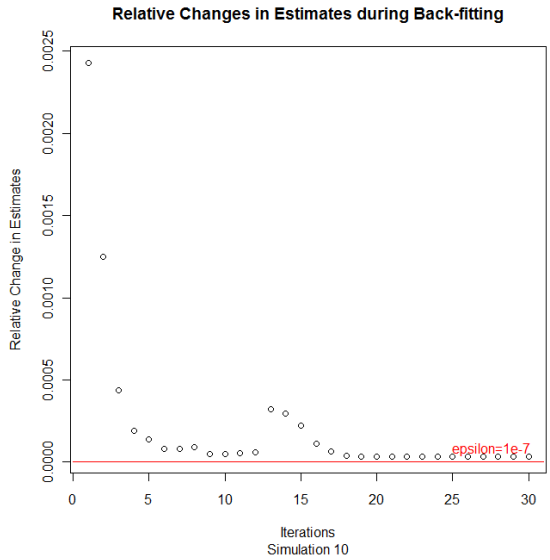


Figure 5-12 An example of nonconvergent back-fitting

The tracks of the optimal bandwidths updated during the back-fitting show that the two bandwidths were oscillating after reaching relatively steady values. The two graphics in Figure 5-13 show the updating procedure of the optimal bandwidths. The graphic on the left displays the entire procedure, while the ones on the right zoom in on the latter half to demonstrate the oscillation.

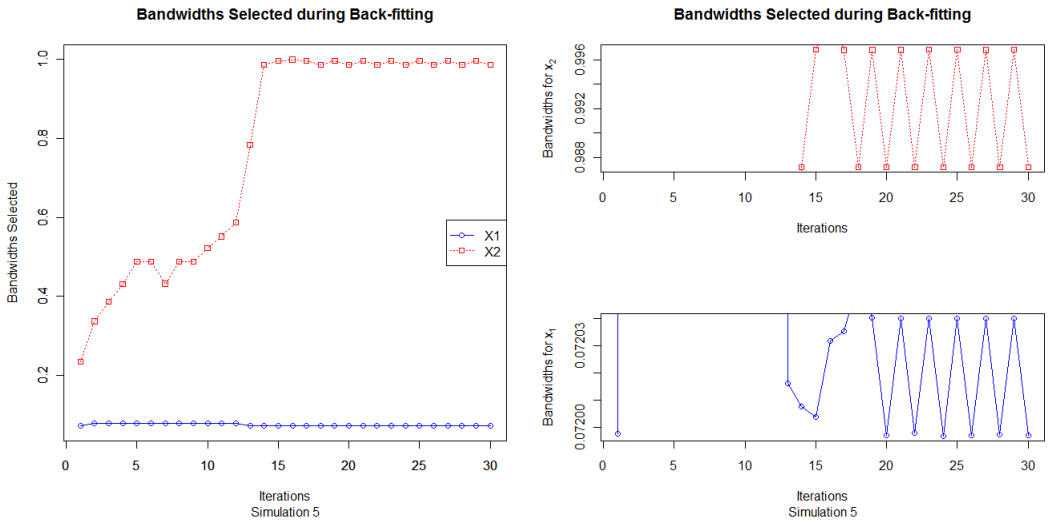


Figure 5-13 Updates of bandwidths during a nonconvergent back-fitting

It would appear that the oscillation has inhibited convergence and that a possible remedy is to stop the updating after both bandwidths have reached their relatively steady values. For the dataset above, this point is reached at step 20 where the bandwidths are 0.072 for x_1 and 0.99 for x_2. The FBGWR model was rerun with these two bandwidths and convergence was successfully achieved after 15 steps of

iteration, the relative change in estimates in the final step was 7.41e-08. The procedure of convergence is shown in Figure 5-14.

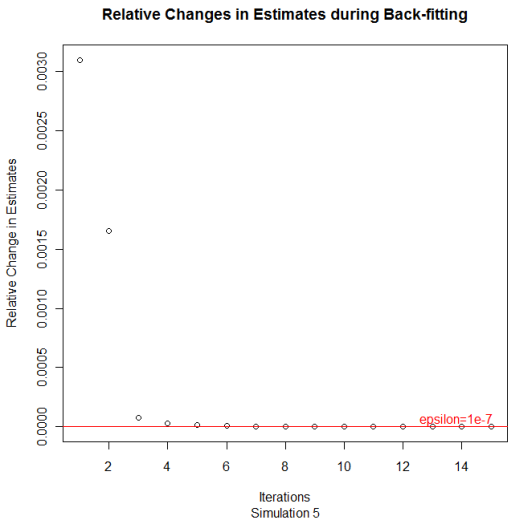
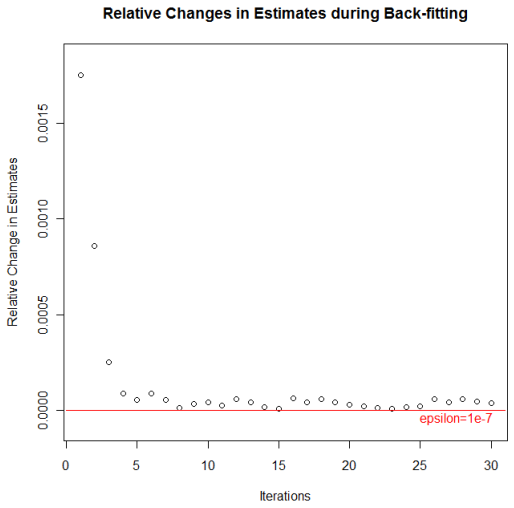
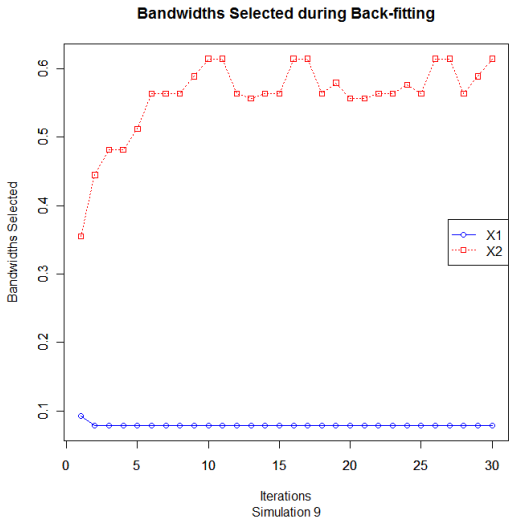


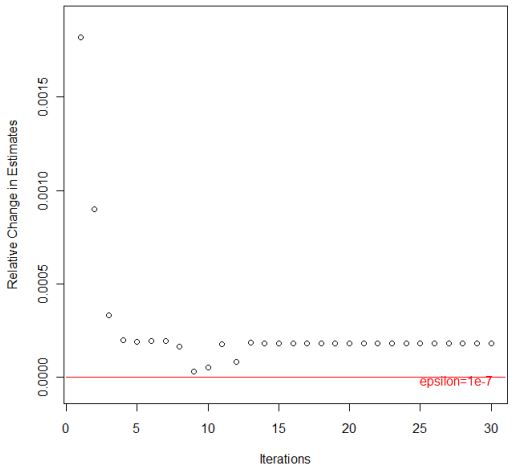
Figure 5-14 Convergence after remedy on a nonconvergent example

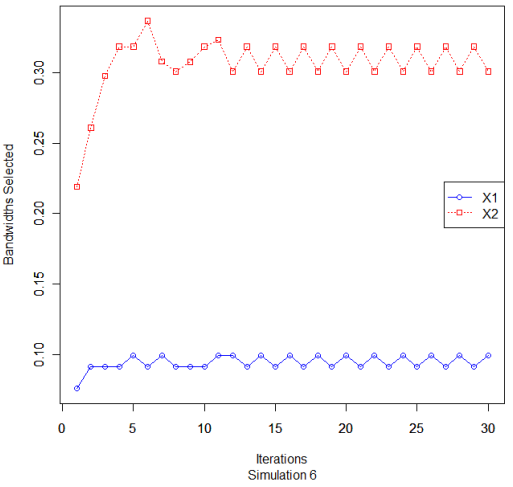
Comparison between the two models before and after this remedy shows that there are only negligible differences between the estimates of the coefficients and the fitted values of the dependent variables. This is easily understood as the two models were actually using almost the same bandwidths to calibrate the final models- to be precise, (0.07199693 0.9872001) in the non-convergent model and (0.07199703, 0.987185) in the convergent model. This suggests that, even though the convergence was not achieved under a strict criterion, the model is still reliable as far as the bandwidths reaching a relatively steady status.

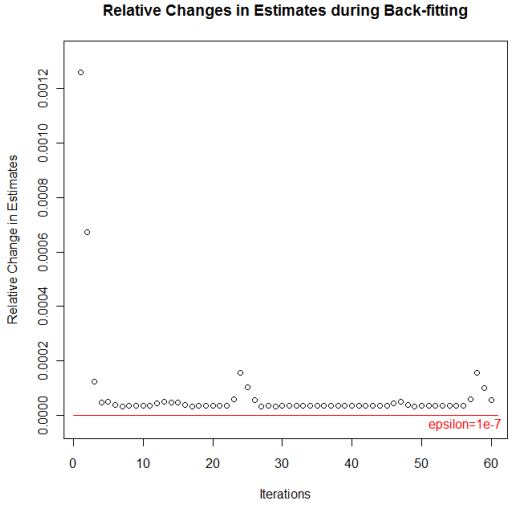
More experiments were carried out on 35 non-convergent cases where datasets were generated from various models with various parameters. Using the same remedy, the convergence was assured in all cases. Several typical cases among these non-convergent cases were selected and presented in Figure 5-15. In each case, the plots on the left display the relative changes in estimates during back-fitting; the ones on the right are the tracks of the bandwidth updates. The figures help to demonstrate how the oscillations in bandwidths prevented the back-fitting process from converging. The threshold used in these cases was 1e-7, while the maximum number of iterations was 30 for the first two examples, and 60 for the latter two.

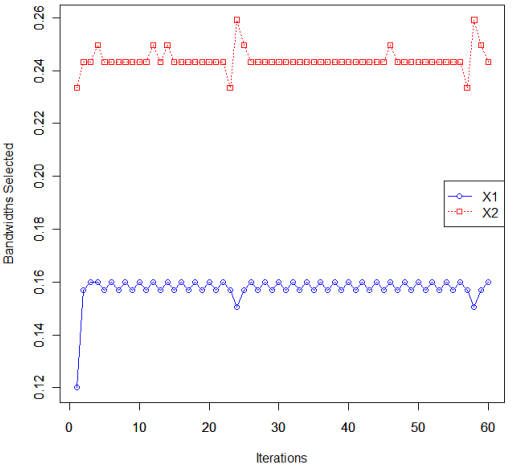

Relative Changes in Estimates during Back-fitting
Relative Change in Estimates
0.0000
0.0005
0.0010
0.0015
epsilon=1e-7
0
5
10
15
20
25
30
Iterations


Bandwidths Selected during Back-fitting
Bandwidths Selected
0.1
0.2
0.3
0.4
0.5
0.6
X1
X2
0
5
10
15
20
25
30
Iterations
Simulation 9


Relative Changes in Estimates during Back-fitting
Relative Change in Estimates
0.0000
0.0005
0.0010
0.0015
epsilon=1e-7
0
5
10
15
20
25
30
Iterations


Bandwidths Selected during Back-fitting
Bandwidths Selected
0.10
0.15
0.20
0.25
0.30
X1
X2
0
5
10
15
20
25
30
Iterations
Simulation 6


Relative Changes in Estimates during Back-fitting
Relative Change in Estimates
0.0000
0.0002
0.0004
0.0006
0.0008
0.0010
0.0012
epsilon=1e-7
0
10
20
30
40
50
60
Iterations


Bandwidths Selected during Back-fitting
Bandwidths Selected
0.12
0.14
0.16
0.18
0.20
0.22
0.24
0.26
X1
X2
0
10
20
30
40
50
60
Iterations

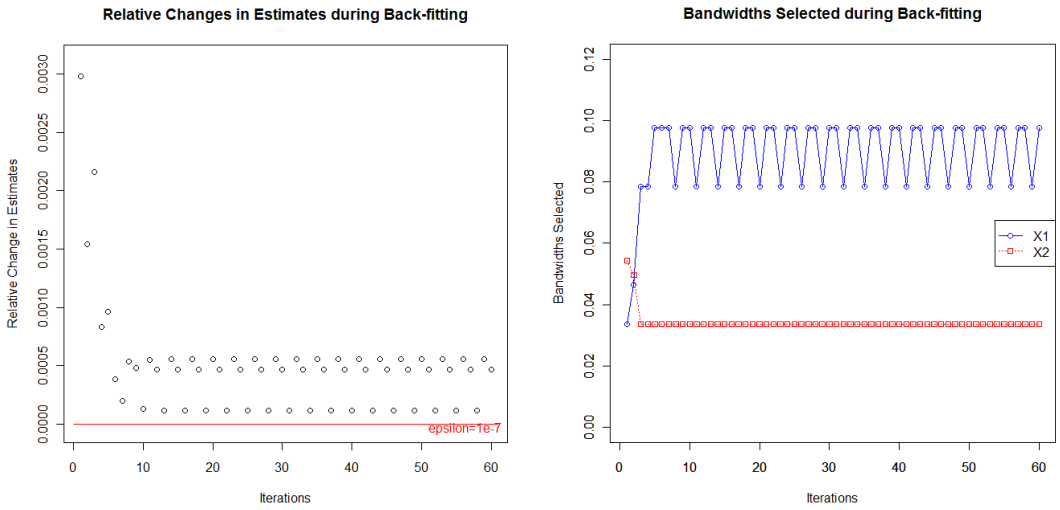
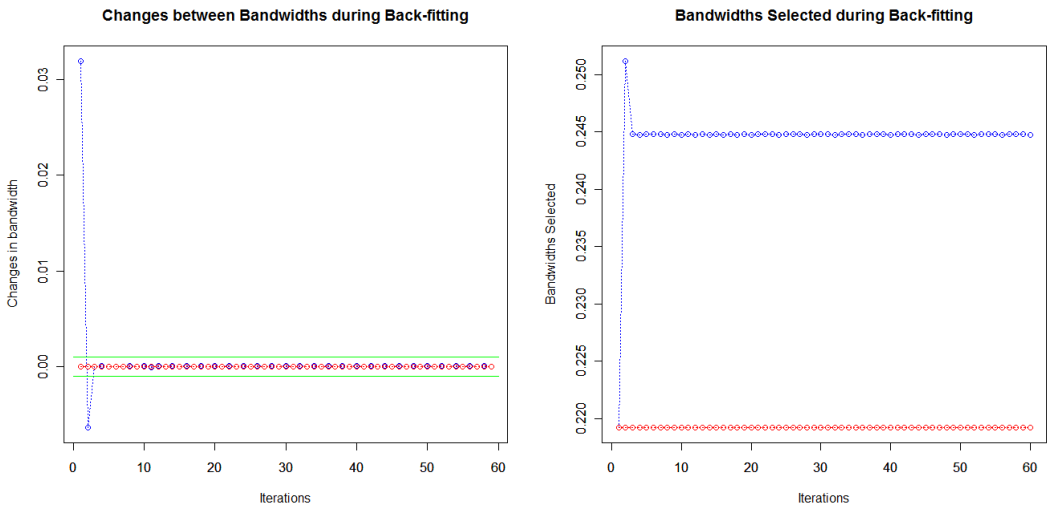


Figure 5-15 Examples of nonconvergent back-fitting

The bandwidths used in the final correction model need to be determined manually through examination of the updating tracks, as demonstrated in the above example. The basic guideline is that the bandwidths should either reach relatively steady values or exhibit regular oscillations. A method to find the steady bandwidths is to check the change between two successive bandwidth updates. If the change is relatively small, for example, below 1e-3, the bandwidth can be taken as steady. Figure 5-16 shows several cases where the steady bandwidths were found through this automatic method. The figures on the left plot the change between two successive bandwidth updates during back-fitting and the two green horizontal lines mark the thresholds of 1e-3 and -1e-3. The figures on the right plot the optimal bandwidths selected on each iterative step during back-fitting.



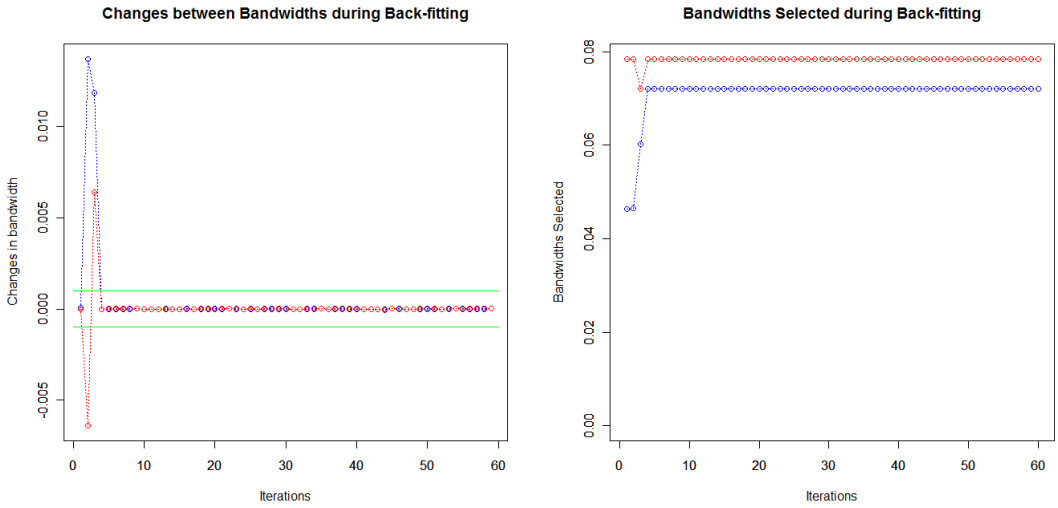


Figure 5-16 Examples of selecting steady bandwidths automatically

However, the automatic method is not always effective, as in some cases when the change between bandwidth updates is rather small but the bandwidth does not show regular oscillations. The figures below display an example of this situation. After 8 iterations, the changes between the bandwidth updates fall below 1e-3 for both bandwidths as shown in the plot on the left, while the plot on the right shows that the bandwidth for x_1 is still increasing. To correct this, the bandwidths selected after 16 iterations were used in the final model instead.

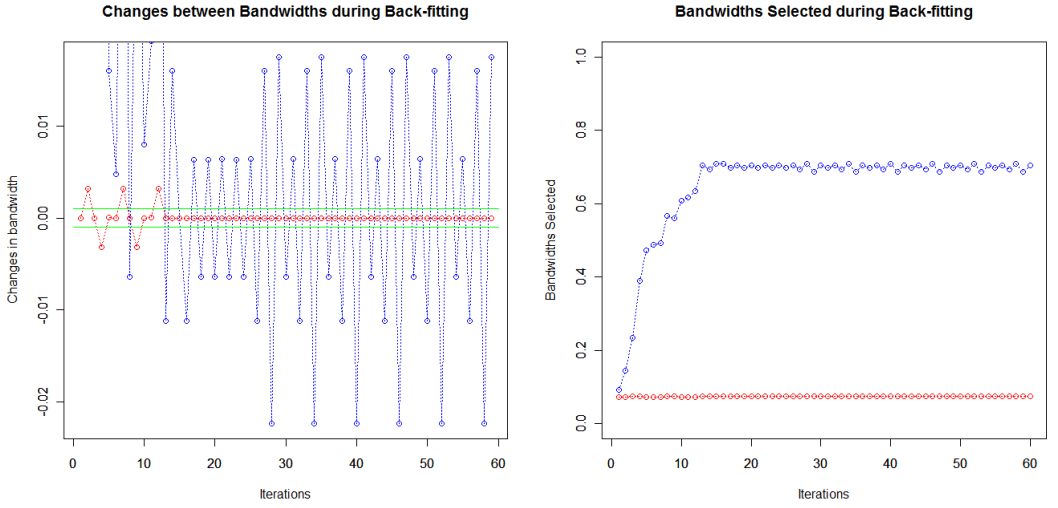


Figure 5-17 An example of selecting steady bandwidths manually (the plot on the left has been zoomed in and trimmed to show more details)

In other cases, although the bandwidth oscillates, the amplitude is so wide that the change between updates will never fall below the threshold. A typical example is shown in Figure 5-18. The plots show that optimal bandwidth for x_1 does not change

after 3 steps while the bandwidth for x_2 keeps oscillating with a large amplitude. It is useful to stop the oscillation and fix the bandwidths at reasonable values such as the ones selected after 5 steps to allow the model to converge.

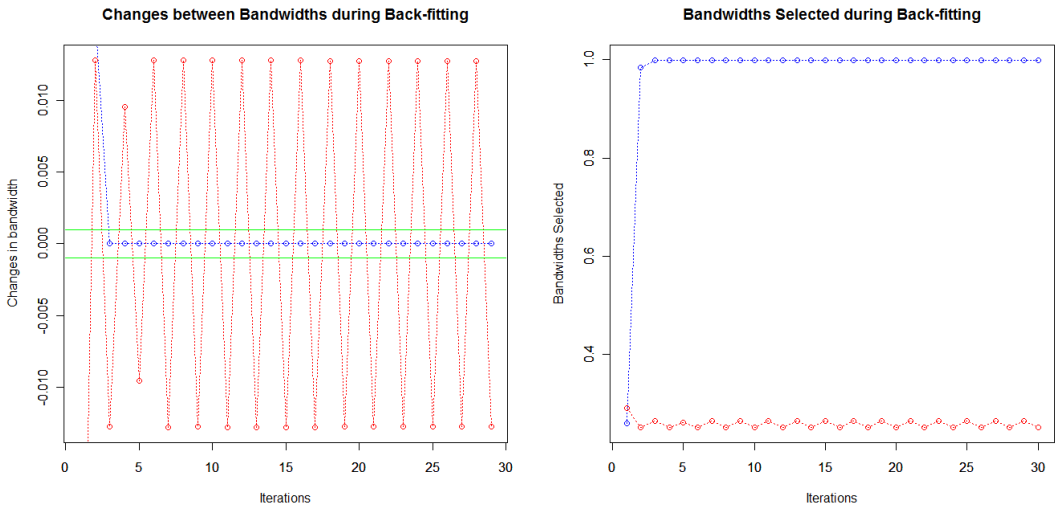


Figure 5-18 A second example of selecting steady bandwidths manually (the plot on the left has been zoomed in and trimmed to show more details)

It is difficult to determine a universal threshold for all the simulations and an adjustment has to be made manually to select the proper bandwidth. The procedure can be simpler in an empirical study, as there might be *a priori* knowledge to assist the bandwidth selection.

Although effective, this remedy is inconvenient as it is an *ad hoc* procedure. As explained before, the non-convergence has little effect on the reliability of the FBGWR model and therefore the remedy procedure was skipped in the following experiments in this chapter. Furthermore, the maximum number of iterations was raised to 60 to assure the accuracy of the model.

In conclusion, due to its efficiency and reliability, the synchronous bandwidth optimization method is adopted in the experiments in the rest of this chapter if not specified otherwise.

5.3 Choice of initial guesses

To initiate a back-fitting calibration for FBGWR, various initial guesses can be set as the value for each term to be estimated. For example, in a model:

$$\mathrm{Y} = \beta_{bw1}\mathrm{X}_1 + \beta_{bw2}\mathrm{X}_2 + \boldsymbol{\varepsilon}, \tag{5-10}$$

estimates of $\beta_{bw1}\mathrm{X}_1$ and $\beta_{bw2}\mathrm{X}_2$ can both be assumed as 0 at the beginning. In the first loop of the iteration, Y is regressed against X_1 to gain an updated estimate for $\beta_{bw1}\mathrm{X}_1$, then the residual is regressed against X_2 with the estimate of $\beta_{bw2}\mathrm{X}_2$ being updated. The regression then returns to X_1 to start a second loop. As the iteration continues, estimates of $\beta_{bw1}\mathrm{X}_1$ and $\beta_{bw2}\mathrm{X}_2$ are updated in turn until a pre-defined termination criterion is reached. This is the approach employed by Hastie and Tibshirani (1986) in their Generalized Additive Models.

In FBGWR, an alternative choice of initial guess can be used. A basic GWR with uniform bandwidth in the following form can be calibrated at the beginning:

$$\mathrm{Y} = \beta_1\mathrm{X}_1 + \beta_2\mathrm{X}_2 + \boldsymbol{\varepsilon}. \tag{5-11}$$

The resulting estimates of $\beta_1\mathrm{X}_1$ and $\beta_2\mathrm{X}_2$, denoted by $\widehat{\beta_1}\mathrm{X}_1$ and $\widehat{\beta_2}\mathrm{X}_2$ can be assumed as the initial estimates for $\beta_{bw1}\mathrm{X}_1$ and $\beta_{bw2}\mathrm{X}_2$, and the back-fitting process starts as usual with estimates updated until the process terminates. The two approaches were used on the same data and their results compared.

The datasets for the experiment were generated from a model with an intercept and one independent variable:

$$y = \beta_0 + \beta_1 x_1 + \varepsilon \tag{5-12}$$

The x_1s were randomly drawn from a uniform distribution with minimum 0 and maximum 1. εs were randomly drawn from a normal distribution with zero mean and a variance equal to 1/3 of the variance of the mean process.

The two coefficients β_0 and β_1 were defined the same as in Section 5.2:

$$\beta_0 = 1 + \left(\frac{1}{324}\right)[36 - (6-u)^2][36 - (6-v)^2], \tag{5-13}$$

$$\beta_1 = 3, \tag{5-14}$$

with the surfaces of β_0 and β_1 visualized in Figure 5-19:

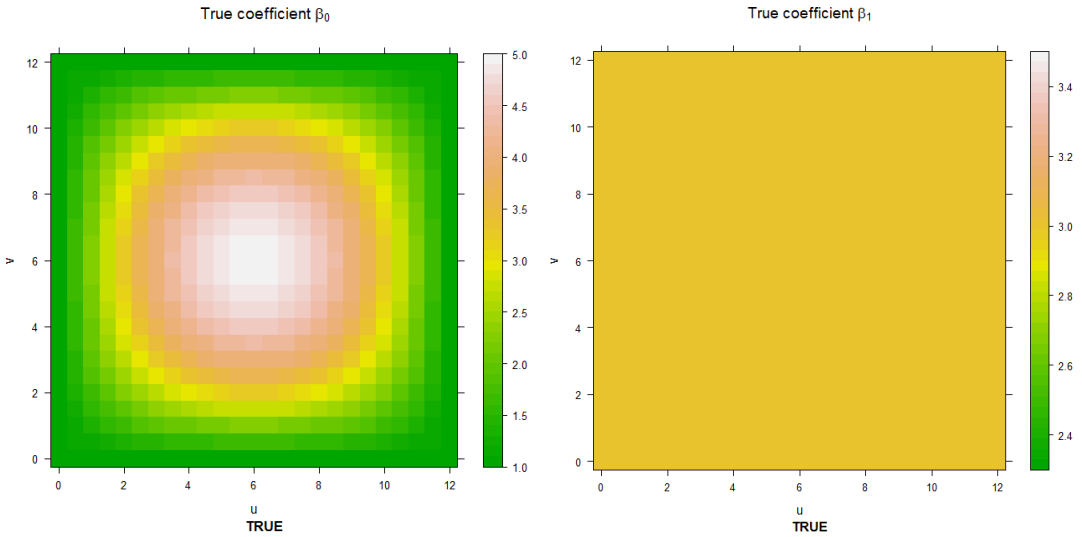


Figure 5-19 Surfaces of the true coefficients

To eliminate the effects of random factors, the experiments were replicated 10 times: in each replication, the x_1s and εs were independently generated from the same distributions but with different random numbers while β_1 and β_2 remained the same in each replication.

5.3.1 Bandwidths selected

Bandwidths were selected automatically according to the AIC criterion through the optimization method described in Section 5.2.3. In the approach where results from a basic GWR were used as initial guesses, the basic GWR was also calibrated with a bandwidth automatically selected based on AIC.

The bandwidths selected for β_0 and β_1 from both approaches and the bandwidth automatically selected for GWR are shown in Figure 5-20. FBGWR-1 is the approach where 0s were used as initial guesses, while FBGWR-2 is the one where estimates from a basic GWR were used as the initial guesses. The same notations are used in the following figures in this section.

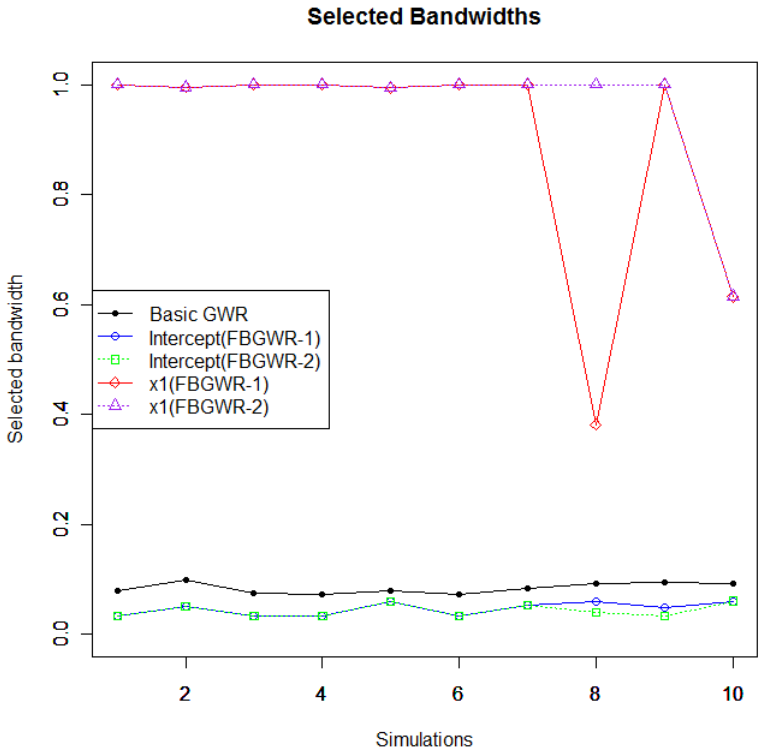
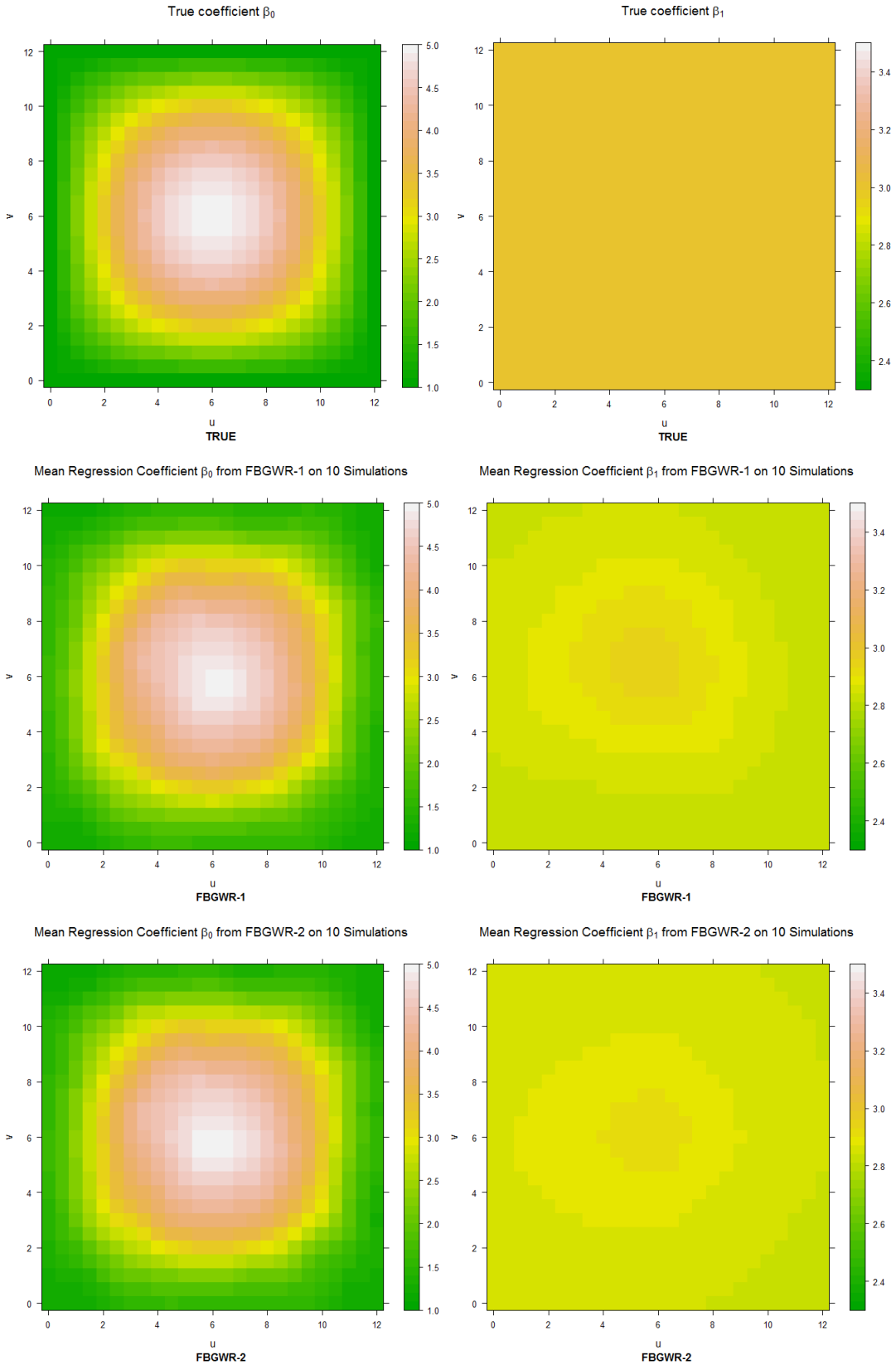


Figure 5-20 Bandwidths selected by GWR and FBGWR with different initial guesses

In most cases, both FBGWR algorithms find the proper bandwidths: a rather low bandwidth for β_0 (the intercept) showing local variations in β_0, and a bandwidth close to 1 for β_1 (the coefficient for x_1) indicating the global nature of β_1. The bandwidth from basic GWR is between the two sets of bandwidths because basic GWR reflects the average variation levels in β_0 and β_1 through a uniform bandwidth. The intercept clearly has a greater influence in the resulting GWR bandwidth than does the slope parameter. Out of the 10 simulations, differences in the two FBGWR approaches can only be seen in two simulations for β_0, and one simulation for β_1.

5.3.2 Estimated coefficients

To give an intuitive idea of how the models perform in estimating the regression coefficients, the true coefficient surfaces of β_0 and β_1 and the mean of regression coefficients estimated from the two FBGWR models as well as from the basic GWR model based on the 10 simulations are shown in Figure 5-21.


True coefficient β_0
True coefficient β_1
u
v
TRUE
Mean Regression Coefficient β_0 from FBGWR-1 on 10 Simulations
Mean Regression Coefficient β_1 from FBGWR-1 on 10 Simulations
FBGWR-1
Mean Regression Coefficient β_0 from FBGWR-2 on 10 Simulations
Mean Regression Coefficient β_1 from FBGWR-2 on 10 Simulations
FBGWR-2

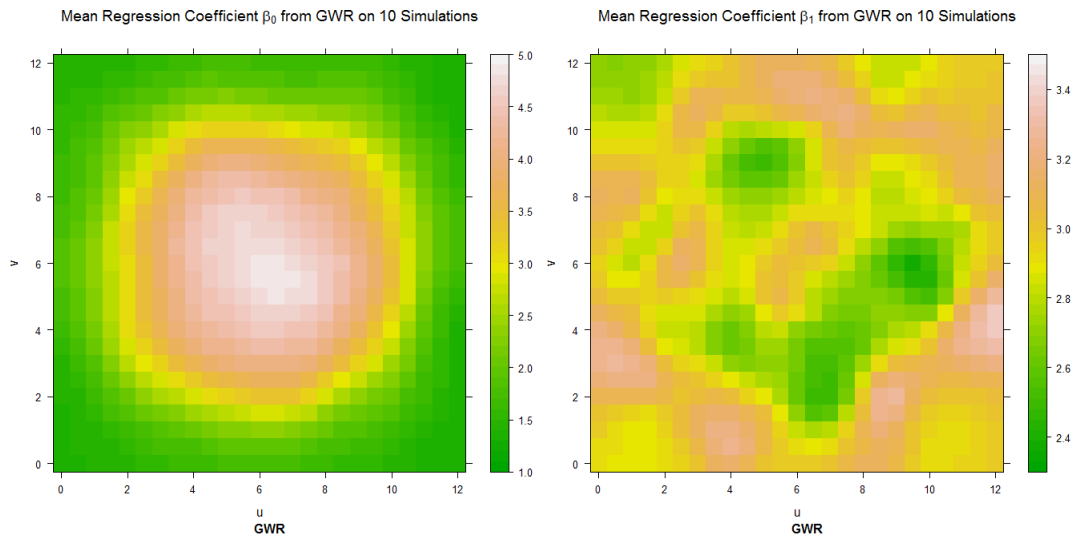


Figure 5-21 Estimated coefficient surfaces from GWR and FBGWR with different initial guesses

All three methods reproduce the surfaces for β_0 quite well. However, for β_1, the two FBGWR approaches perform better than basic GWR. The pattern in the β_1 surface estimated from basic GWR indicates that the model has introduced some artificial variations into the coefficient estimates; this can be misleading in the interpretation of the estimates. The two FBGWR models have improved the estimation by producing flatter coefficient surfaces that are closer to the true coefficient surface.

5.3.3 Algorithm efficiency

While the two FBGWR approaches were completed under the same termination setting, which requires the back-fitting process to iterate until the the relative change in estimates between two successive iterations is below a threshold of 1e-7, the actual iterations and the total time cost for each simulation are plotted in Figure 5-22.

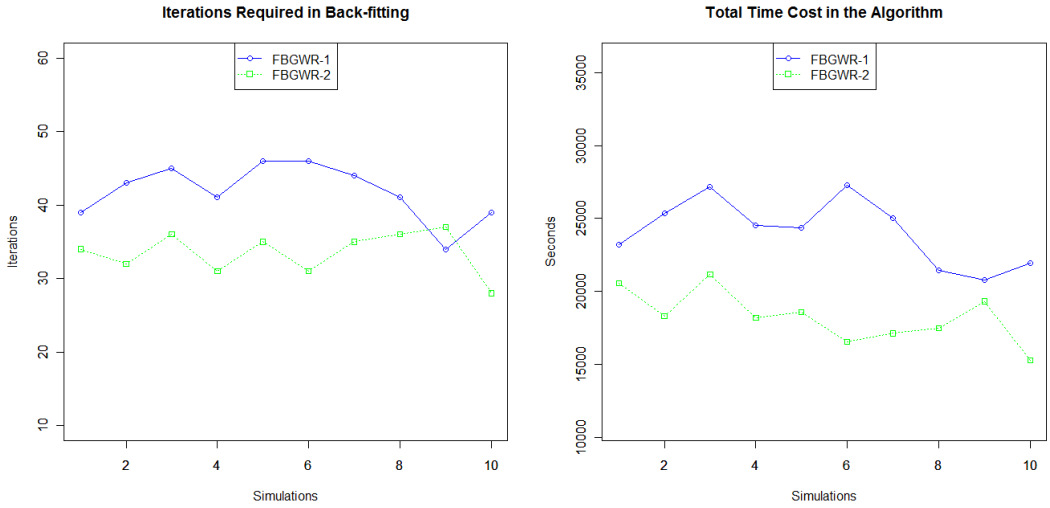


Figure 5-22 Iterations and time cost in FBGWR with different initial guesses

The average iterations required for FBGWR-1 was 42, while the FBGWR-2 procedure reduces the average to 34. Accordingly, the average time cost for FBGWR-1 was 5.7 hours (20,623 seconds) whilst for FBGWR-2 the average was 4.5 hours (16,368 seconds). This indicates that although the two approaches reached almost the same results, FBGWR-2 was more efficient than FBGWR-1. For this reason, FBGWR-2 was employed in the remaining experiments in this chapter if not specified otherwise. The same approach was used in the experiments reported in Section 5.2.

It is also reasonable to assume that the choice of initial estimates only affects the efficiency of the algorithm and not the results of the regression, given sufficient iterations are allowed until the convergence of the back-fitting algorithm.

5.4 Choice of termination threshold

As mentioned earlier in this chapter as well as in Chapter 4 (see Section 4.5.2), the termination criterion employed in the iterative process of back-fitting to calibrate FBGWR is the relative change in estimates between two successive loops. The score of change is defined as in (4-11).

The SOC is compared with a pre-defined threshold at each iterative step to detect convergence. Once SOC is below the threshold or the maximum number of iterations is reached, the process terminates.

In this section, three different values: 1e-7, 1e-5, 1e-3 were tested as termination thresholds. The datasets used here were generated from a model with an intercept and one independent variable:

$$y = \beta_0 + \beta_1 x_1 + \varepsilon \tag{5-15}$$

with the x_1s and εs were randomly drawn in the same way as before. The two coefficients β_0 and β_1 were defined as:

$$\beta_0 = 3, \tag{5-16}$$

$$\beta_1 = 1 + \left(\frac{1}{324}\right)[36 - (6-u)^2][36 - (6-v)^2], \tag{5-17}$$

The simulation was replicated 25 times; in each replication the x_1s and εs were re-drawn randomly, giving 25 independent datasets.

Figure 5-23 below shows the iterations needed for convergence. The red solid triangle pointing down marks the cases where the threshold was not reached prior to the maximum number of iterations. When 1e-3 was used as a convergence criterion, the

convergence was ensured in all cases and the iterations required were reduced considerably with the average number of iterations being only 2.16.

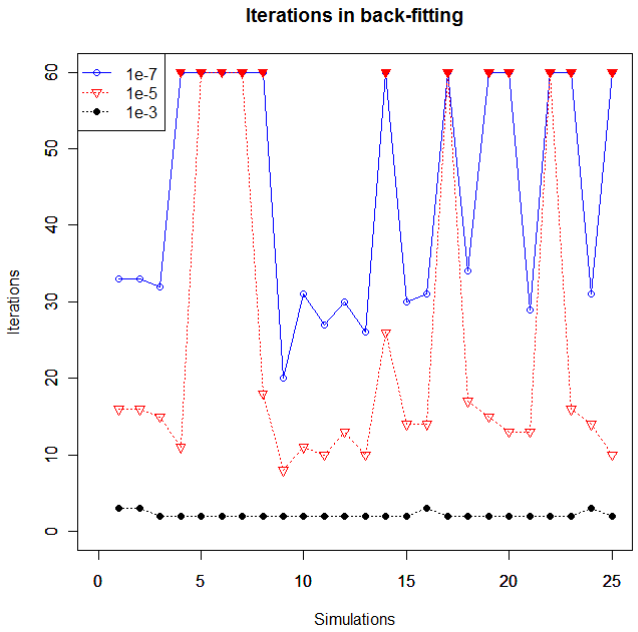
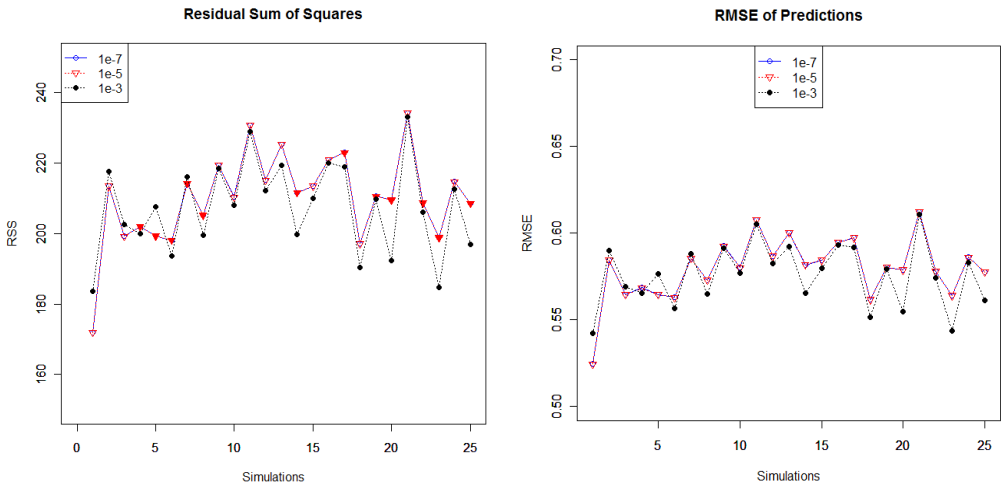


Figure 5-23 Iterations required in FBGWR with different termination thresholds

The accuracy of the algorithms in terms of RSS, RMSE of predictions, RMSE of β_0 and of β_1 are plotted and compared in Figure 5-24.



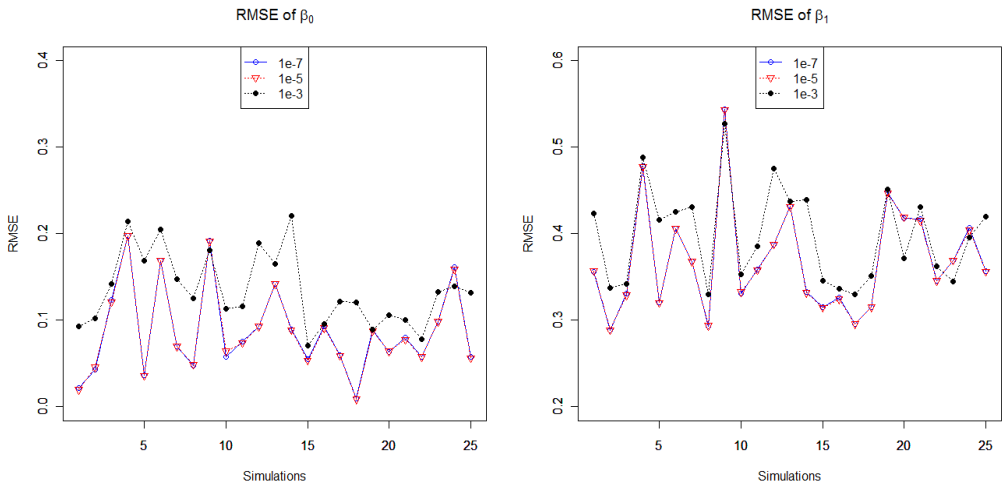


Figure 5-24 RSS and RMSEs from FBGWR with different termination thresholds

The figures of RSS and RMSE of predictions show that the threshold change does not change the overall accuracy very much. While in terms of RMSE of the coefficients β_0 and β_1, the high threshold of 1e-3 seems to have increased the RMSE considerably, and thus reduces the accuracy, while thresholds of 1e-5 and 1e-7 made little difference. Therefore, both 1e-5 and 1e-7 can be adopted in practice. When time allows, the threshold can be specified as 1e-7 first for better accuracy. If the process does not converge, then it can be switched to 1e-5, or even 1e-3, if estimation is not the essential purpose of the analysis. The strictest threshold of 1e-7 was employed in the remaining experiments in this chapter.

5.5 Flexible bandwidths in scenarios

In this section, coefficients with various heterogeneities were combined in the data generating model to simulate a series of datasets. FBGWR was calibrated on each dataset to examine whether accurate bandwidths for various coefficients could be found and whether the coefficients could be estimated accurately.

The data generating model employed two independent variables and no intercept:

$$y = \beta_1 x_1 + \beta_2 x_2 + \varepsilon \tag{5-18}$$

where x_1 and x_2 were randomly and independently drawn from a uniform distribution with minimum 0 and maximum 1 and ε was randomly drawn from a normal distribution with zero mean and a variance equal to 1/3 of the variance of the mean process. The two coefficients β_1 and β_2 were specified as one of the following functions of coordinates.

$$\beta_{zero} = 3 \tag{5-19}$$

$$\beta_{low} = 1 + (\frac{1}{6})(u + v) \tag{5-20}$$

$$\beta_{high} = 1 + \left(\frac{1}{324}\right)[36 - (6 - u)^2][36 - (6 - v)^2] \tag{5-21}$$

The three functions are visualized in Figure 5-25 below respectively as a 2-D surface, representing various levels of heterogeneity across space, with β_{zero} having zero heterogeneity, β_{low} low heterogeneity, and β_{high} high heterogeneity.

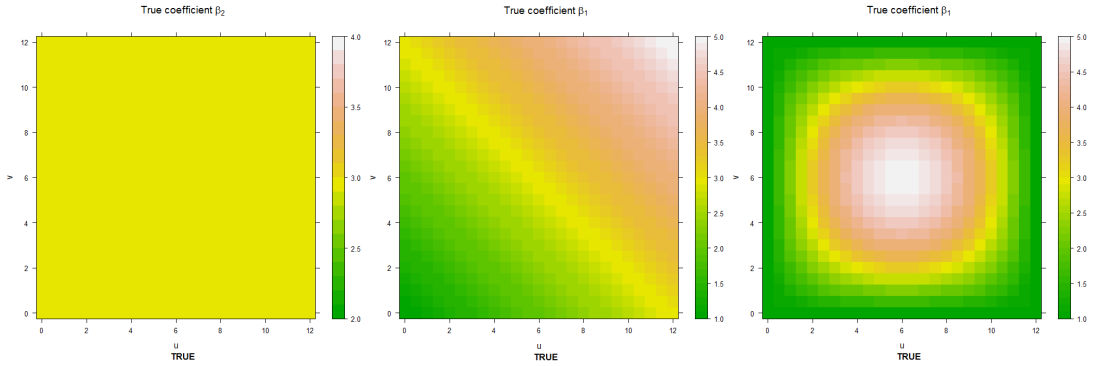


Figure 5-25 Surfaces of the simulation coefficients

Five scenarios were considered, each with a different combination of heterogeneities.

Scenario 1: high heterogeneity β_1 + high heterogeneity β_2

Scenario 2: low heterogeneity β_1 + low heterogeneity β_2

Scenario 3: zero heterogeneity β_1 + low heterogeneity β_2

Scenario 4: high heterogeneity β_1 + zero heterogeneity β_2

Scenario 5: high heterogeneity β_1 + low heterogeneity β_2

In each scenario, the data-generating process was repeated 10 times, simulating 10 datasets each with the same β_1s and β_2s but with different x_1s and εs independently and randomly drawn from the same distributions.

In the calibrated FBGWR model it is expected that the bandwidth associated with the coefficient of β_{zero} should have a large bandwidth with the ideal value of 1 if an adaptive bandwidth is employed, representing a global constant relationship. The coefficient of β_{high} should have a small bandwidth indicating a locally varying relationship; while β_{low} should have a medium bandwidth between the two.

Scenario 1 and Scenario 2 were designed to examine whether FBGWR could estimate roughly equal bandwidths for the two coefficients, while Scenarios 3-5 were employed to see whether FBGWR could distinguish the different levels of bandwidths.

Each scenario was examined and evaluated in terms of bandwidth accuracy, estimation accuracy and properties of the estimated coefficients.

5.5.1 Scenario 1: high heterogeneity β_1 + high heterogeneity β_2

In this scenario, the two coefficients are defined as:

$$\beta_1 = \beta_2 = \beta_{high} \tag{5-22}$$

5.5.1.1 Bandwidth selection

The bandwidths automatically selected for each coefficient associated with variables x_1 and x_2 from FBGWR, as well as the uniform bandwidth selected from basic GWR, are listed in Table 5-3 and plotted in Figure 5-26.

Table 5-3 Bandwidth selected in 10 simulations

	Sim1	Sim2	Sim3	Sim4	Sim5	Sim6	Sim7	Sim8	Sim9	Sim10
β_1 (FBGWR)	0.072	0.078	0.078	0.082	0.072	0.082	0.072	0.093	0.078	0.072
β_2 (FBGWR)	0.078	0.075	0.102	0.116	0.123	0.146	0.078	0.102	0.099	0.118
Basic GWR	0.072	0.059	0.094	0.102	0.093	0.094	0.078	0.093	0.072	0.093

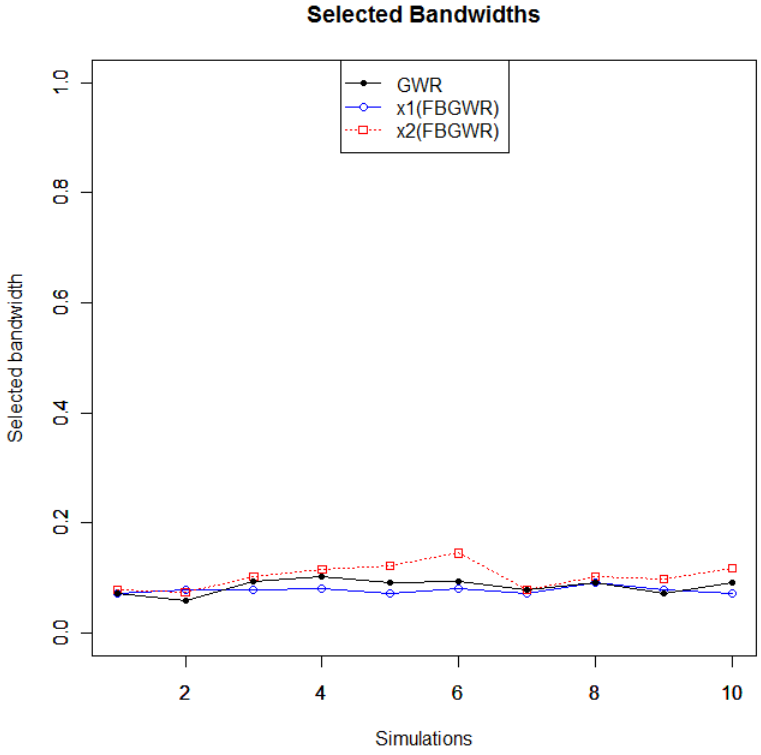


Figure 5-26 Bandwidths selected in 10 simulations

Similar bandwidths are found for the two coefficients in FBGWR, all at low values, mostly below 0.1. This successfully reflects the nature of high heterogeneity in the

two coefficients. Although in some simulations, such as simulation 5 and 6, the two bandwidths differed more than others, this can be accepted given the small scale in the difference. Also the bandwidths from FBGWR are similar to those from basic GWR. This is reasonable because in this scenario FBGWR should function just the same as basic GWR.

5.5.1.2 Algorithm accuracy

The RMSE of the two coefficients β_1 and β_2, as well as of the dependent variable y, were calculated for each simulation and are shown in Figure 5-27. The same statistics from the basic GWR model calibrated on each simulated dataset are also plotted for comparison. The scales of the three plots have been adjusted to the same levels to make them more comparable.

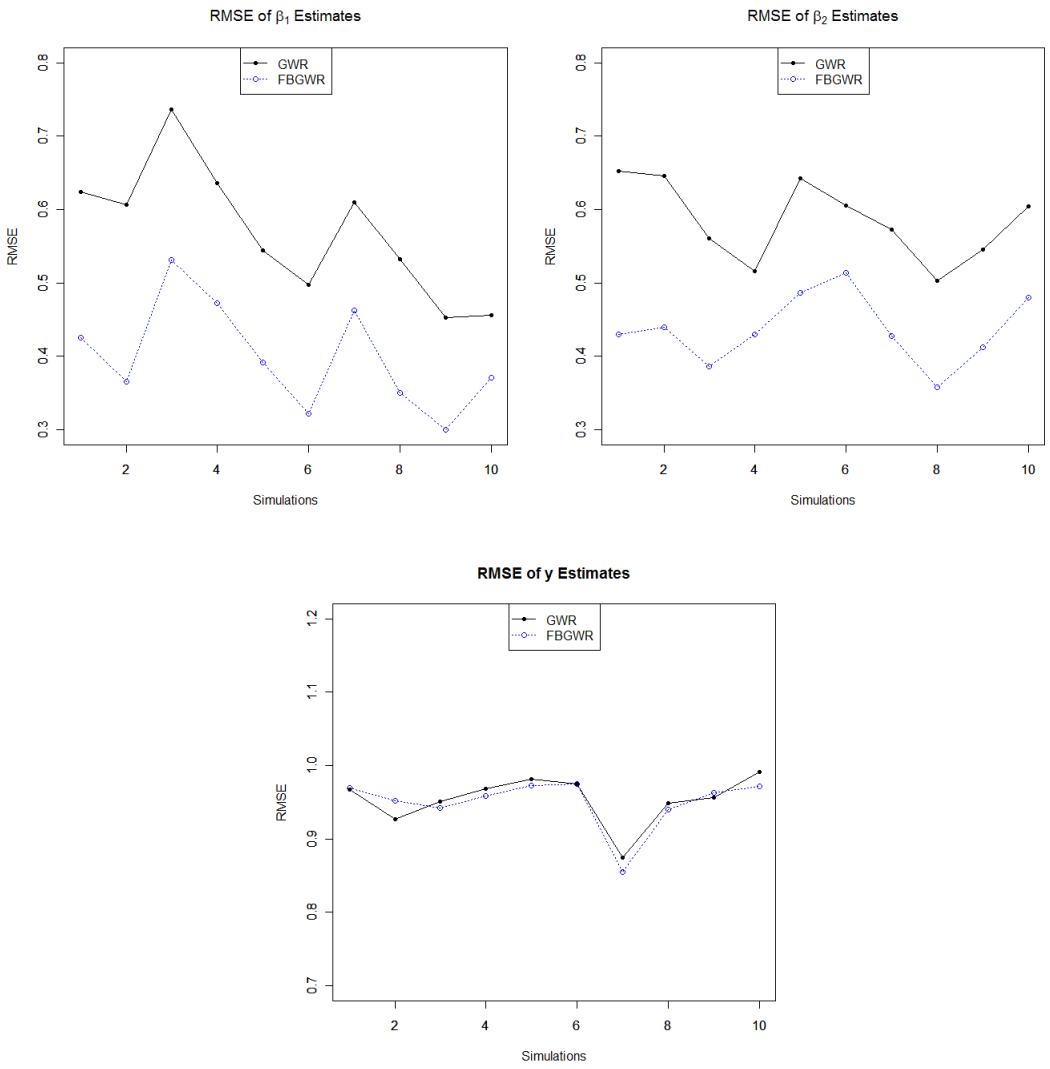


Figure 5-27 RMSEs of estimates in 10 simulations

The RMSEs of both estimated coefficients from FBGWR are much lower than those from basic GWR in all the 10 simulations while the RMSEs of the dependent variable are similar from the two models. This might suggest that although basic GWR can predict the dependent variable as well as FBGWR, it is not as accurate as FBGWR in estimating individual coefficients.

5.5.1.3 Estimated coefficients

Figure 5-28 depicts the true coefficient surfaces for β_1 and β_2, and the mean of coefficients from the 10 simulations estimated by basic GWR and FBGWR respectively.

For both coefficients, FBGWR reproduces the surfaces slightly better than GWR. This can be confirmed by the RMSEs. To make it clearer, the differences between the mean regression coefficients and the true coefficients calculated as (Mean regression coefficients – True coefficients) are also plotted in Figure 5-29.

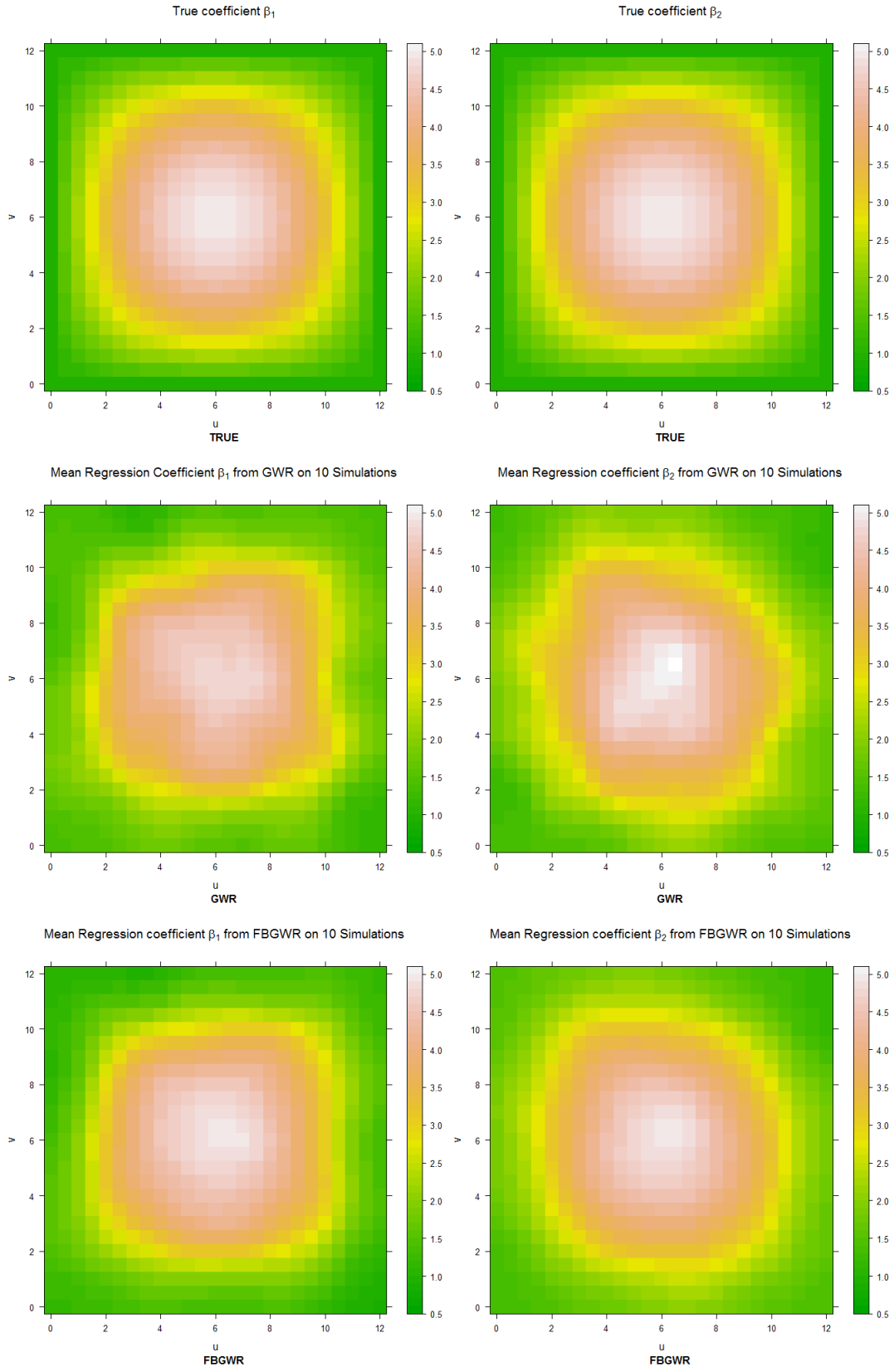


Figure 5-28 Estimated coefficient surfaces from basic GWR and FBGWR

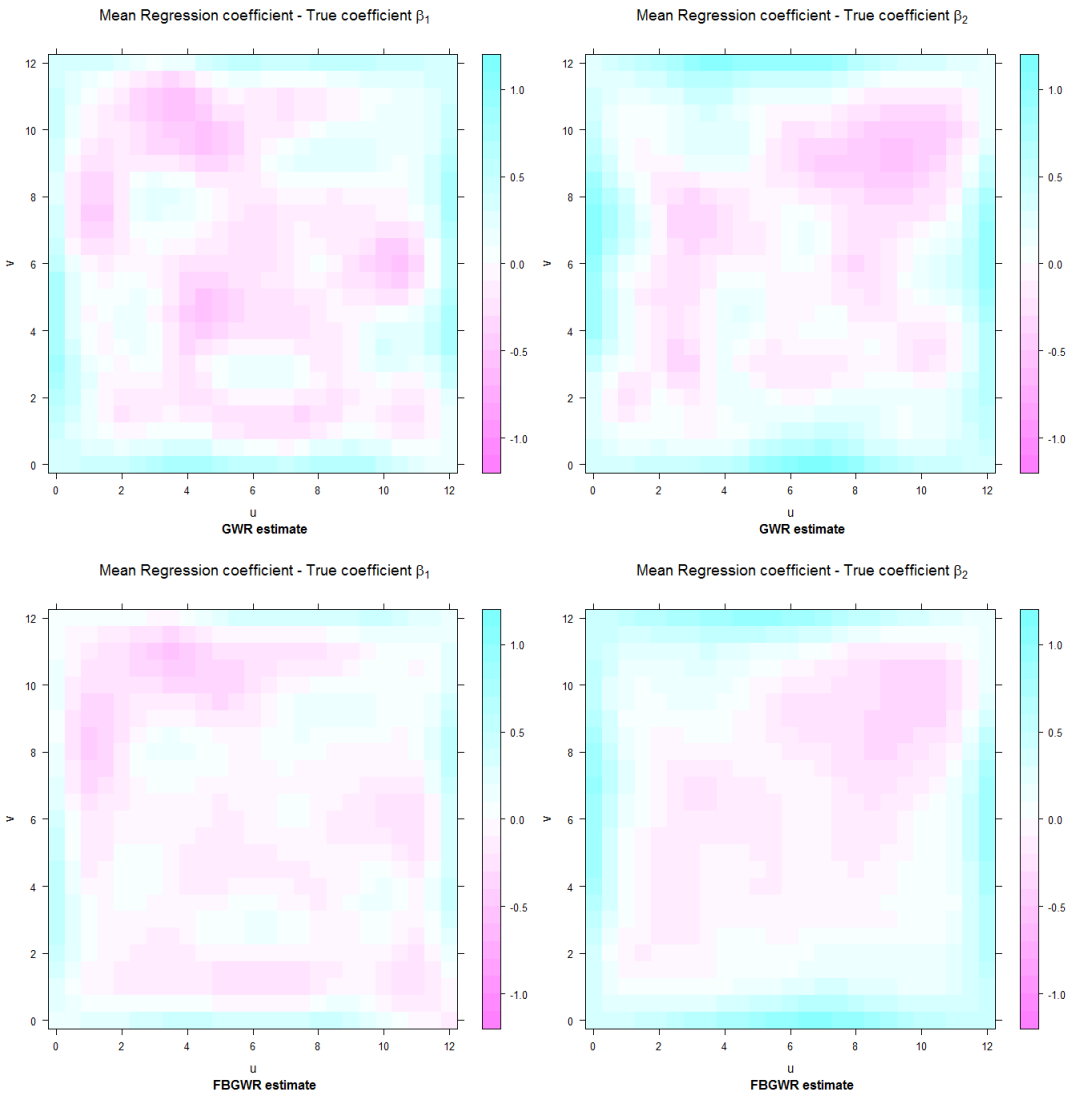


Figure 5-29 Differences between estimates and true coefficients from basic GWR and FBGWR

5.5.1.4 Changing the scale of the error term

The simulation experiment is affected by the magnitude of the random error term. If different scales of error are introduced into the data generating process, different datasets will be produced and different model calibration results will be observed.

In the data generating model, as expressed in equation (5-1), the variance of the random error term was set at 1/3 of the variance of the mean process by default. In this sub-section, difference proportions of 1/2 and 1/4 were used instead and the simulations were repeated 10 times as before. Table 5-4 summarizes the mean of the bandwidths selected and the mean of RMSEs from the 10 simulations under each configuration. The results from basic GWR models are also listed for comparison.

Table 5-4 Mean bandwidth and RMSE from datasets with various scales of error

level of variance in error	1/2 (high)	1/3 (moderate)	1/4 (low)
bandwidth for β_1	0.088	0.078	0.071
bandwidth for β_2	0.127	0.104	0.099
bandwidth (basic GWR)	0.096	0.085	0.083
RMSE of *y*	1.141	0.950	0.801
RMSE of *y* (basic GWR)	1.147	0.954	0.820
RMSE of β_1	0.436	0.399	0.377
RMSE of β_1 (basic GWR)	0.597	0.570	0.493
RMSE of β_2	0.497	0.437	0.411
RMSE of β_2 (basic GWR)	0.645	0.585	0.535

As the scale of variance in error term decreases, the optimal bandwidths as well as the RMSEs decrease in both FBGWR and basic GWR. This is because error terms with higher variability obscure the true data generating process more, making the local patterns less recognizable and therefore the regression models tend to employ larger bandwidths to reflect the patterns at a broader scale. On the contrary, error terms with lower variability make the local patterns in datasets easier to be revealed, with smaller bandwidths reported to reflect the true nature of the spatial processes. The mean uniform bandwidth from basic GWR is always between the two means of bandwidths from FBGWR.

The RMSEs of *y* are similar in FBGWR and basic GWR, while the RMSEs of β_1 and β_2 are smaller in FBGWR than in basic GWR under all configurations. This confirms the previous conclusion that FBGWR is more reliable than basic GWR in estimating individual coefficients. More details of RMSE from each of the 10 simulations under the three configurations are displayed in Figure 5-30.

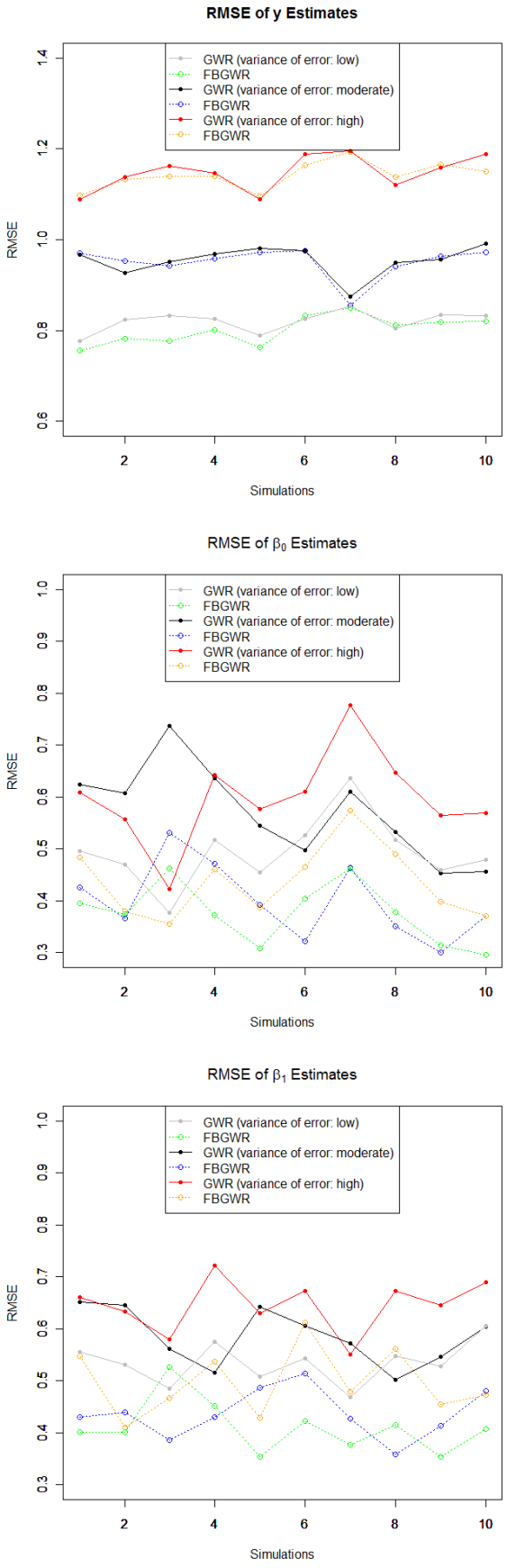


Figure 5-30 RMSEs from groups with various scales of error term

5.5.2 Scenario 2: low heterogeneity β_1 + low heterogeneity β_2

In this scenario, the two coefficients were defined as:

$$\beta_1 = \beta_2 = \beta_{low} \tag{5-23}$$

5.5.2.1 Bandwidth selection

Table 5-5 lists the bandwidths automatically selected for each coefficient associated with variables x_1 and x_2 from FBGWR as well as the uniform bandwidth selected from basic GWR. Figure 5-31 visualizes the same bandwidths.

Table 5-5 Bandwidth selected in 10 simulations

	Sim1	Sim2	Sim3	Sim4	Sim5	Sim6	Sim7	Sim8	Sim9	Sim10
β_1 (FBGWR)	0.285	0.245	0.160	0.258	0.118	0.314	0.232	0.296	0.126	0.283
β_2 (FBGWR)	0.338	0.219	0.243	0.298	0.298	0.237	0.314	0.232	0.315	0.245
Basic GWR	0.254	0.232	0.160	0.232	0.232	0.237	0.243	0.194	0.194	0.232

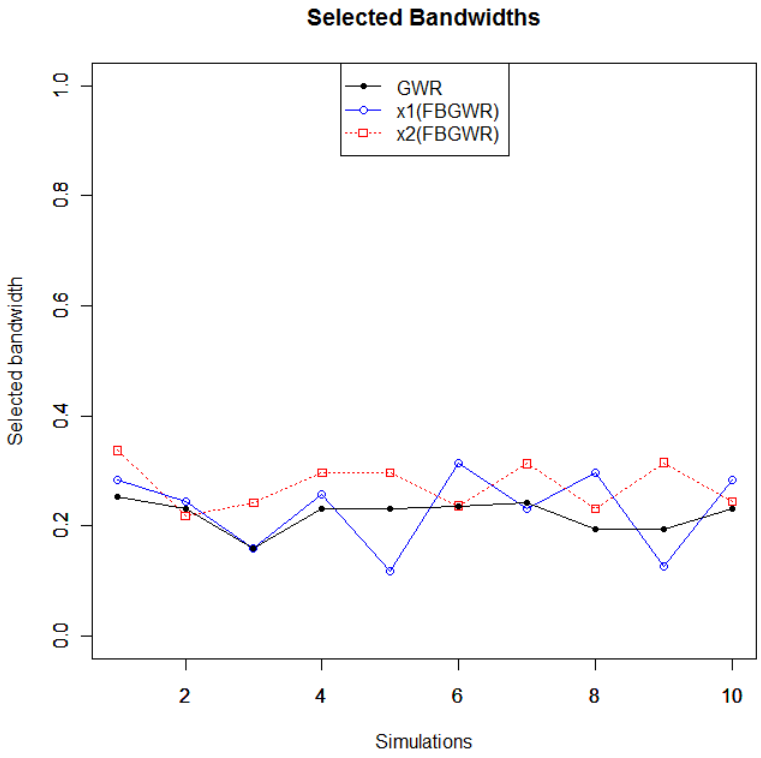


Figure 5-31 Bandwidths selected in 10 simulations

Similar bandwidths were selected for the two coefficients β_1 and β_2 in FBGWR although in some replications the two bandwidths were less similar than in others. They were all at a medium level, a level that is higher than that in scenario 1. The lowest bandwidth equals to 0.12 and the highest equals 0.34. The bandwidths from

basic GWR were similar to those from FBGWR; this is in accord with the fact that the two coefficients are actually varying at the same scale.

5.5.2.2 Algorithm accuracy

Figure 5-32 displays the RMSEs of the two coefficients and the dependent variable from both FBGWR and basic GWR for each simulation.

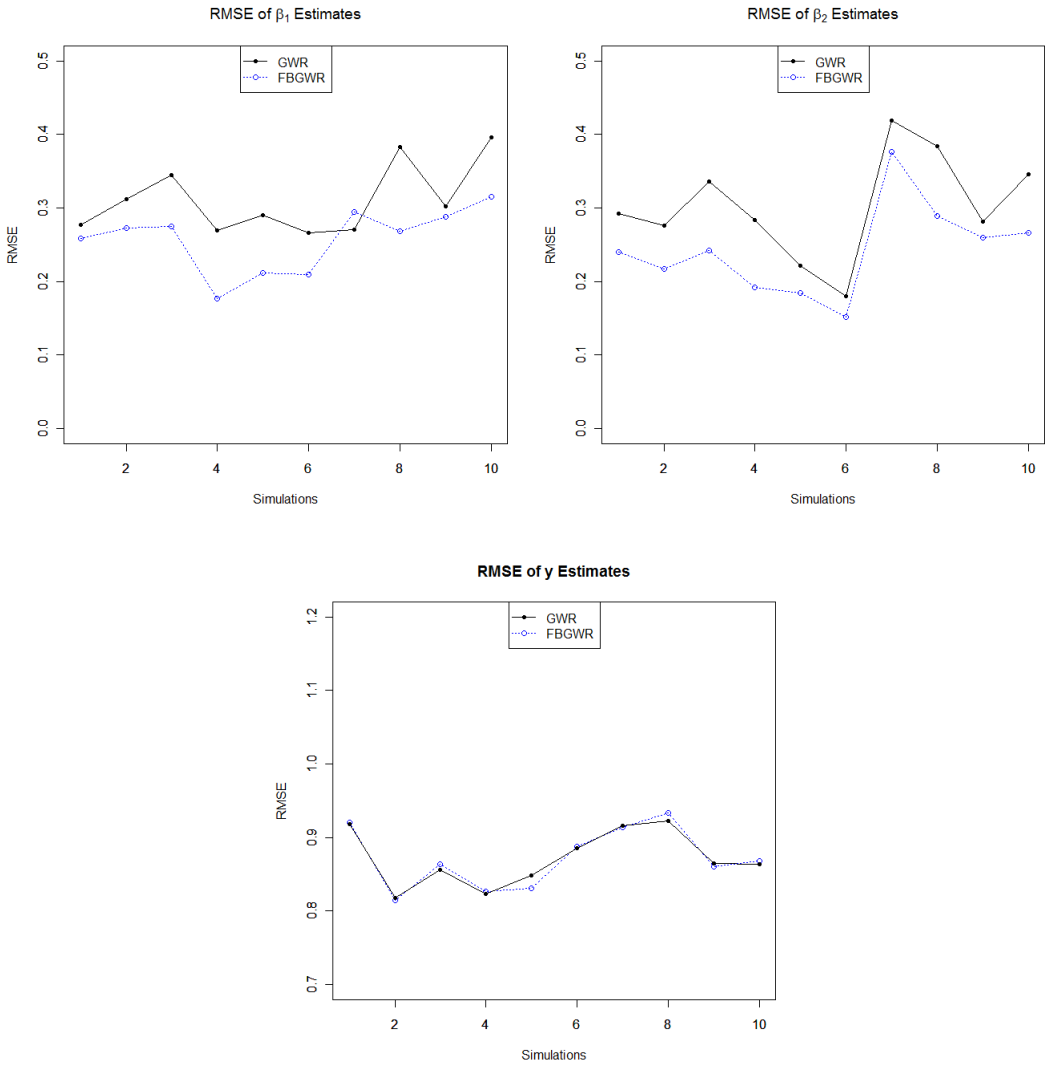


Figure 5-32 RMSEs of estimates in 10 simulations

Similar to the results in Scenario 1, the RMSEs of both estimated coefficients from FBGWR are lower than those from basic GWR, while the RMSEs of the dependent variable are similar from the two models. Again, this suggests that FBGWR is more reliable than basic GWR in estimating coefficients and explaining the underlying spatial process.

5.5.2.3 Estimated coefficients

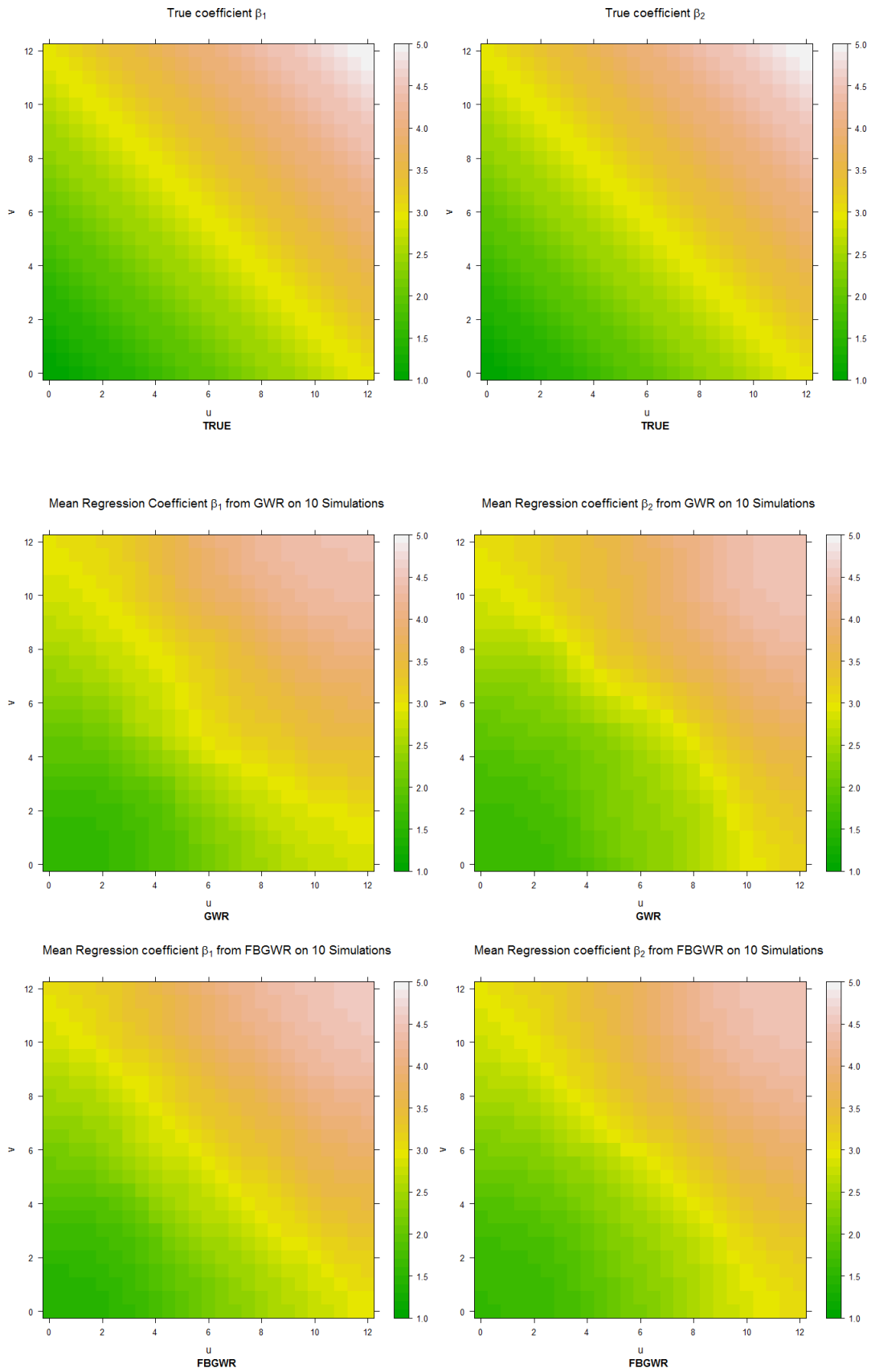


Figure 5-33 Estimated coefficient surfaces from basic GWR and FBGWR

The true surfaces of the two coefficients β_1 and β_2, and the mean of coefficients from the 10 simulations estimated by basic GWR and FBGWR respectively are displayed in Figure 5-33.

For both coefficients, FBGWR performs slightly better than basic GWR as can be seen by the differences between the mean estimated coefficients and the true coefficients (Mean estimated coefficients – True coefficients) shown in Figure 5-34.

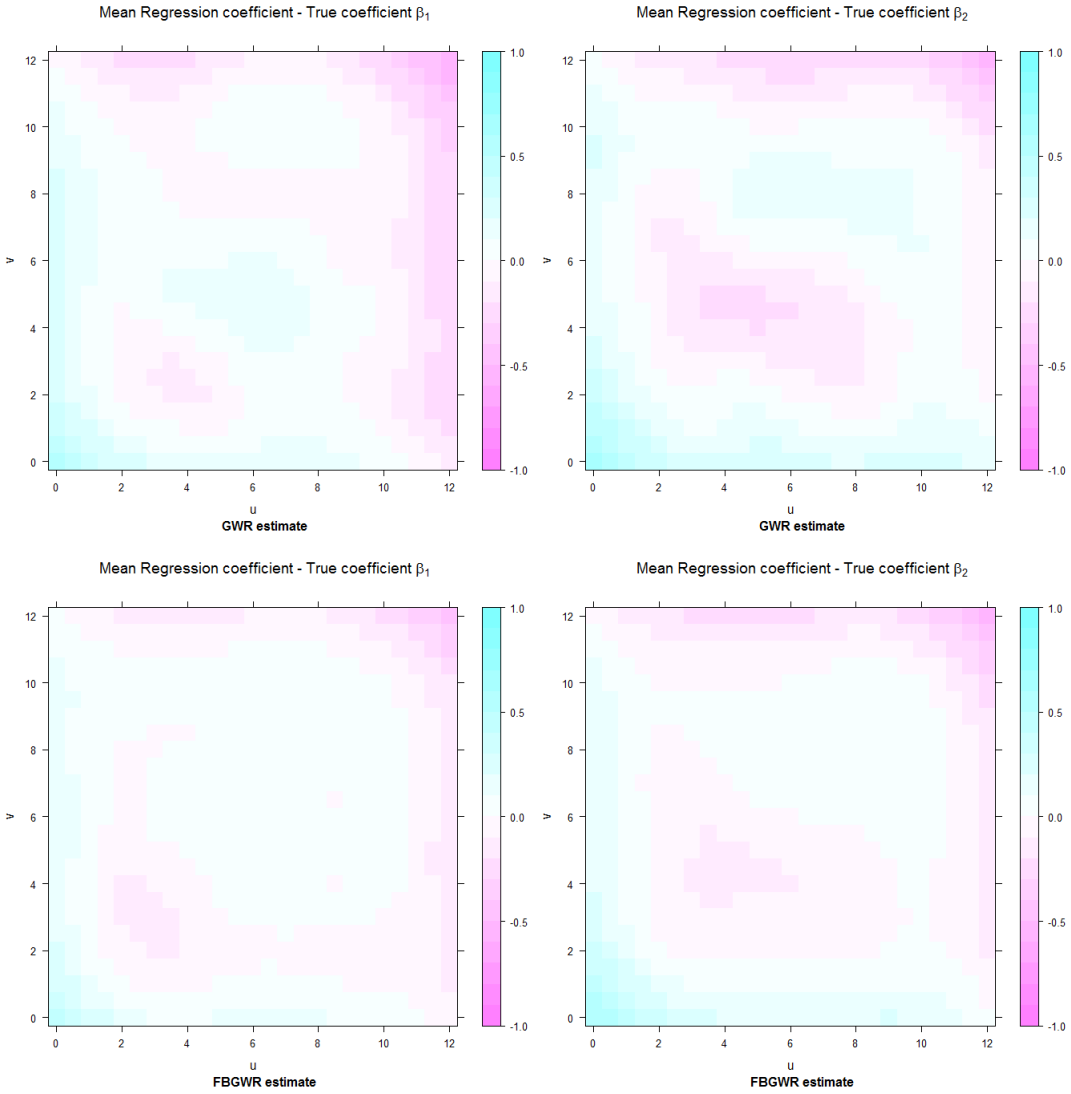


Figure 5-34 Differences between estimates and true coefficients from basic GWR and FBGWR

5.5.2.4 Changing the scale of the error term

Again, three groups of datasets were simulated with the variance of random error terms set at different scales. The mean of the bandwidths and the mean of RMSEs from each group of 10 simulations are summarized in Table 5-6.

Table 5-6 Mean bandwidth and RMSE from datasets with various scales of error

level of variance in error	1/2 (high)	1/3 (moderate)	1/4 (low)
bandwidth for β_1	0.275	0.232	0.196
bandwidth for β_2	0.309	0.274	0.242
bandwidth (basic GWR)	0.241	0.221	0.184
RMSE of y	1.055	0.872	0.742
RMSE of y (basic GWR)	1.055	0.872	0.742
RMSE of β_1	0.276	0.257	0.215
RMSE of β_1 (basic GWR)	0.325	0.311	0.270
RMSE of β_2	0.282	0.242	0.229
RMSE of β_2 (basic GWR)	0.346	0.302	0.288

The same effects as those in Scenario 1 are observed here - that the optimal bandwidths and the RMSEs decrease as the scale of variance in error term decreases. Again, this can be explained as that error terms with lower variance make the true data generating processes easier to be revealed, with smaller bandwidths found to reflect the local variations in the datasets.

The RMSEs from each of the 10 simulations in the three groups are also plotted in Figure 5-35. As in Scenario 1, the RMSEs of y are similar in the two models and the RMSEs of β_1 and β_2 are smaller in FBGWR than in basic GWR. The RMSEs of y in both models are largely affected by the scale of variance in error term, while the RMSEs of β_1 and β_2 are not affected much.

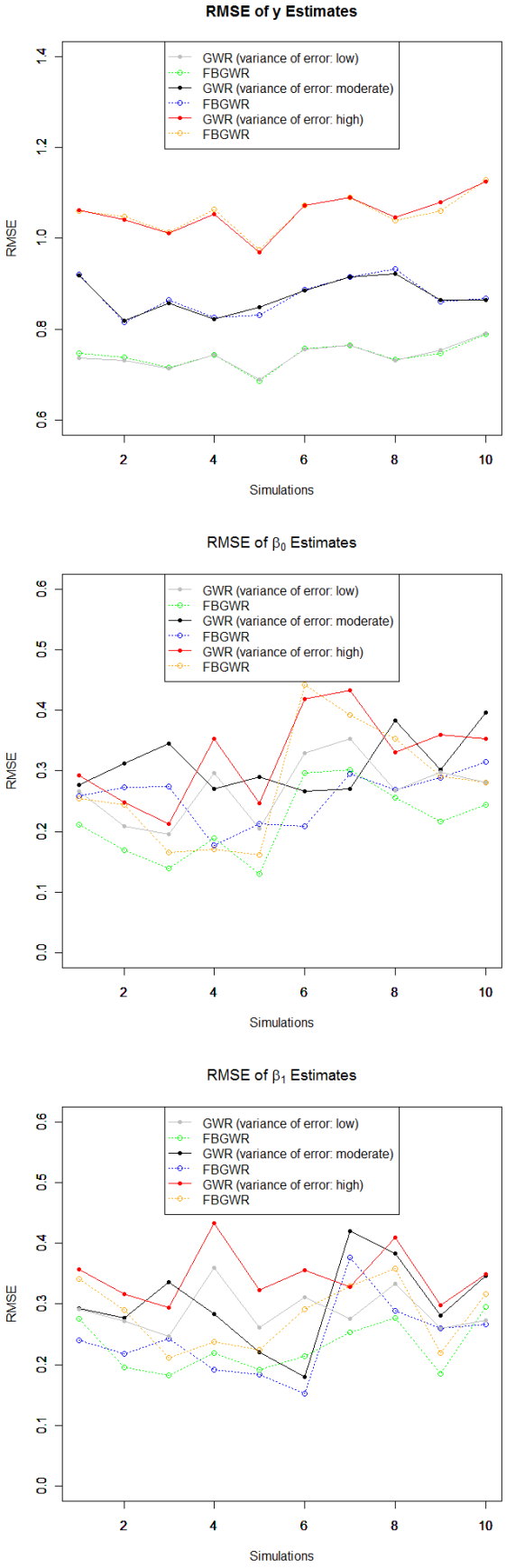


Figure 5-35 RMSEs from groups with various scales of error term

5.5.3 Scenario 3: zero heterogeneity β_1 + low heterogeneity β_2

In this scenario and the following ones, the two coefficients β_1 and β_2 were specified differently. Here, coefficient β_1 was fixed as a constant 3 having no heterogeneity, while β_2 was set with low heterogeneity.

$$\beta_1 = \beta_{zero} \quad (5\text{-}24)$$

$$\beta_2 = \beta_{low} \quad (5\text{-}25)$$

5.5.3.1 Bandwidth selection

The bandwidths automatically specified for each coefficient associated with variables x_1 and x_2 from FBGWR and the uniform bandwidth selected from basic GWR are compared in Table 5-7 and Figure 5-36.

Table 5-7 Bandwidth selected in 10 simulations

	Sim1	Sim2	Sim3	Sim4	Sim5	Sim6	Sim7	Sim8	Sim9	Sim10
β_1 (FBGWR)	1.000	0.997	0.733	1.000	1.000	1.000	0.912	1.000	0.715	0.856
β_2 (FBGWR)	0.113	0.176	0.301	0.232	0.198	0.237	0.275	0.251	0.285	0.286
Basic GWR	0.146	0.238	0.314	0.309	0.259	0.309	0.298	0.296	0.341	0.360

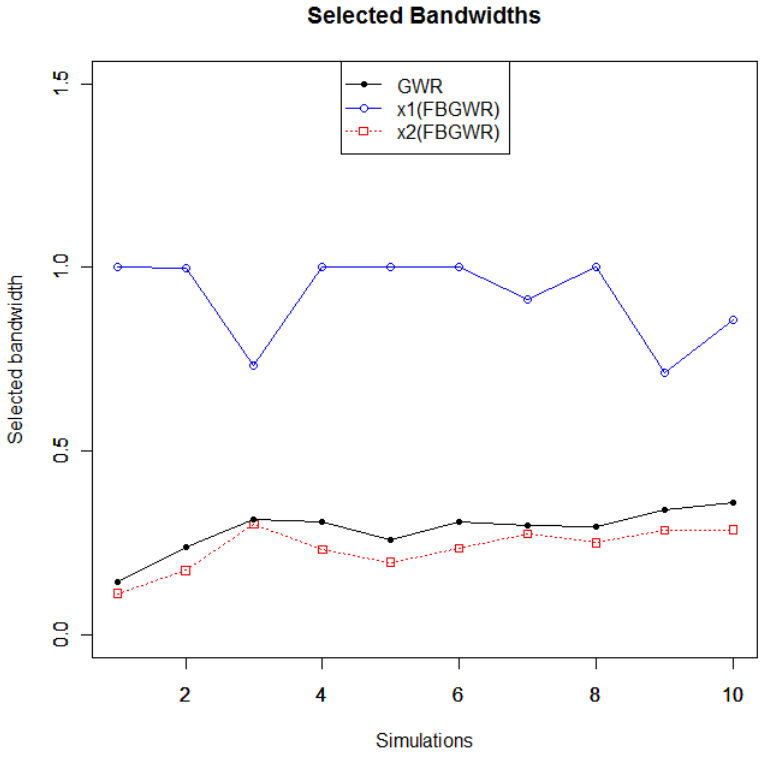


Figure 5-36 Bandwidths selected in 10 simulations

Quite distinguishable bandwidths are found for x_1 and x_2 in FBGWR with the bandwidths for x_2 being much smaller than those for x_1; most of the latter are close to 1. This is in accord with the features of the two coefficients: β_1 is a constant whilst β_2 exhibits some heterogeneity. In all 10 simulations, the bandwidth from basic GWR is between the two bandwidths from FBGWR and is closer to the smaller one, suggesting that the relationship between y and x_2 is stronger than that between y and x_1.

5.5.3.2 Algorithm accuracy

Figure 5-37 compares the RMSEs from the two models for each of the 10 simulations.

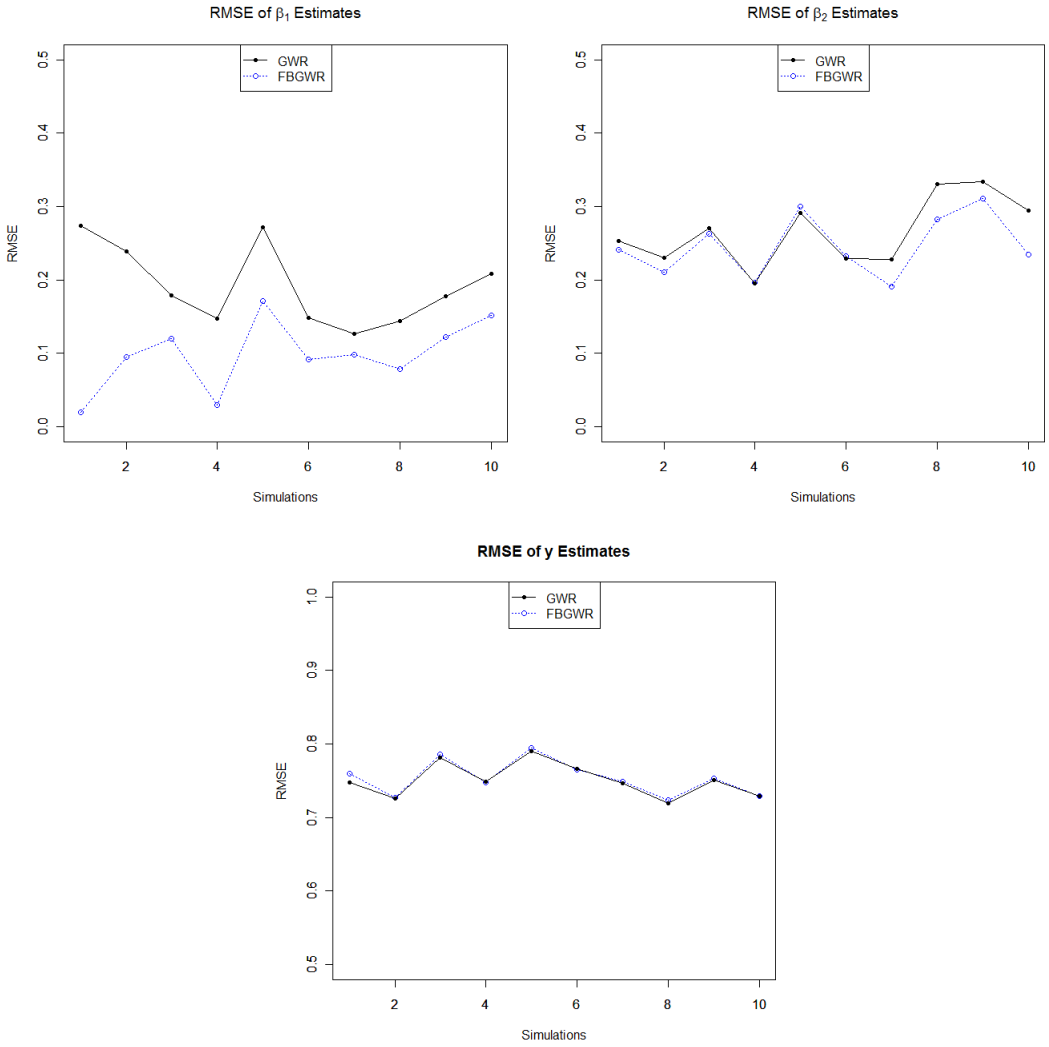


Figure 5-37 RMSEs of estimates in 10 simulations

The RMSEs of β_1 from FBGWR are clearly lower than those from basic GWR, while in the cases of β_2 and y, the two models have produced similar RMSEs. This is

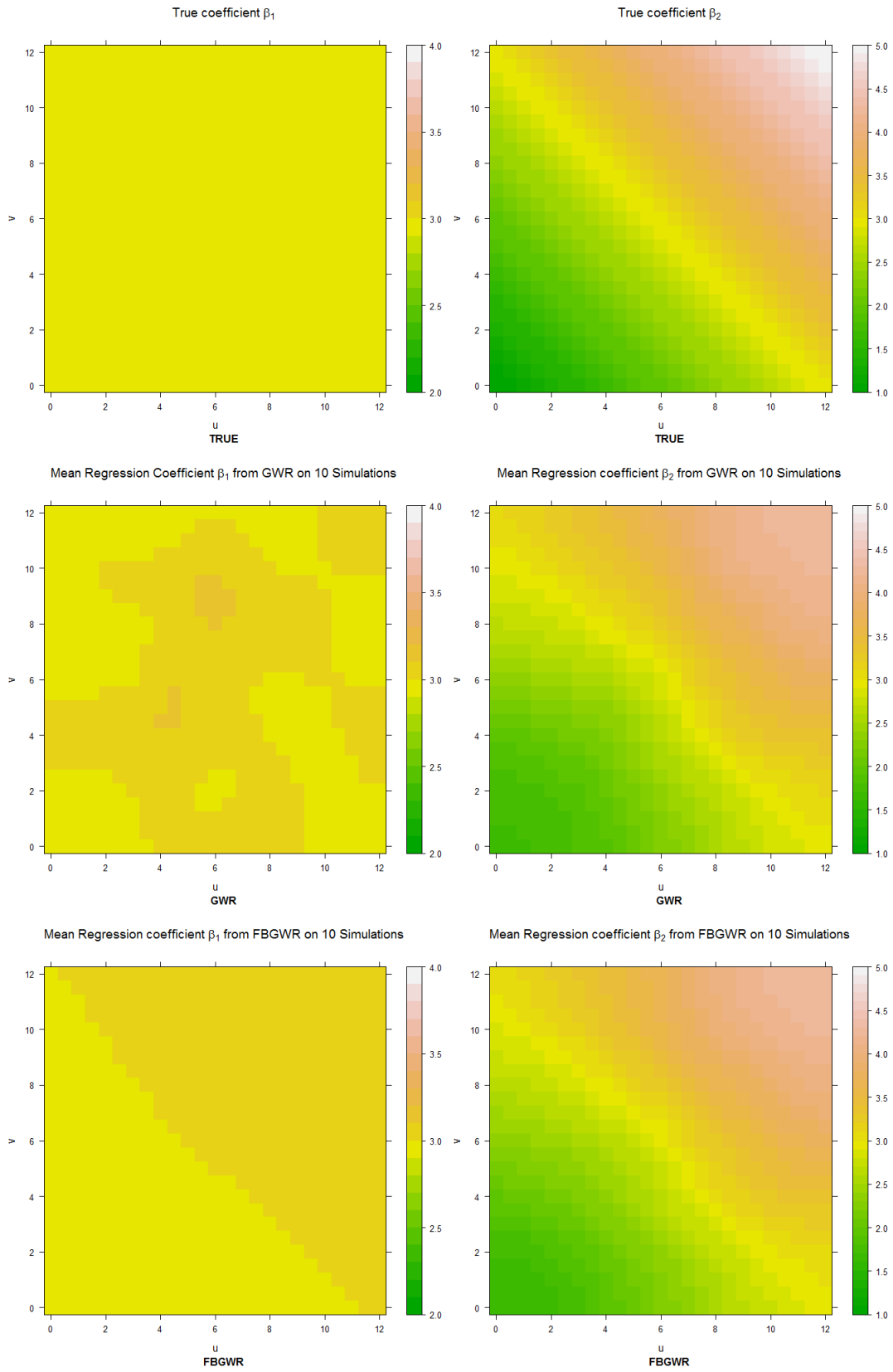


Figure 5-38 Estimated coefficient surfaces from basic GWR and FBGWR

because the large bandwidths for β_1 in FBGWR helped to estimate the coefficients more precisely than basic GWR while for β_2 both models performed well in the estimation with similar small bandwidths.

5.5.3.3 Estimated coefficients

Figure 5-38 depicts the true coefficient surfaces for β_1 and β_2 and the mean of estimated coefficients from the 10 simulations.

It is clear that FBGWR reproduces the parameter surfaces more accurately than basic GWR.

5.5.4 Scenario 4: high heterogeneity β_1 + zero heterogeneity β_2

In this scenario, coefficients β_1 were set with high heterogeneity and β_2 was fixed as a constant 3.

$$\beta_1 = \beta_{high} \quad (5\text{-}26)$$

$$\beta_2 = \beta_{zero} \quad (5\text{-}27)$$

5.5.4.1 Bandwidth selection

Table 5-8 and Figure 5-39 compare the bandwidths automatically selected for each variable x_1 and x_2 from FBGWR and the uniform bandwidth selected from basic GWR.

Table 5-8 Bandwidth selected in 10 simulations

	Sim1	Sim2	Sim3	Sim4	Sim5	Sim6	Sim7	Sim8	Sim9	Sim10
β_1 (FBGWR)	0.111	0.111	0.072	0.083	0.072	0.099	0.072	0.082	0.078	0.098
β_2 (FBGWR)	1.000	0.472	0.994	0.995	0.987	0.301	1.000	1.000	0.614	0.606
Basic GWR	0.146	0.142	0.110	0.117	0.072	0.117	0.102	0.112	0.118	0.118

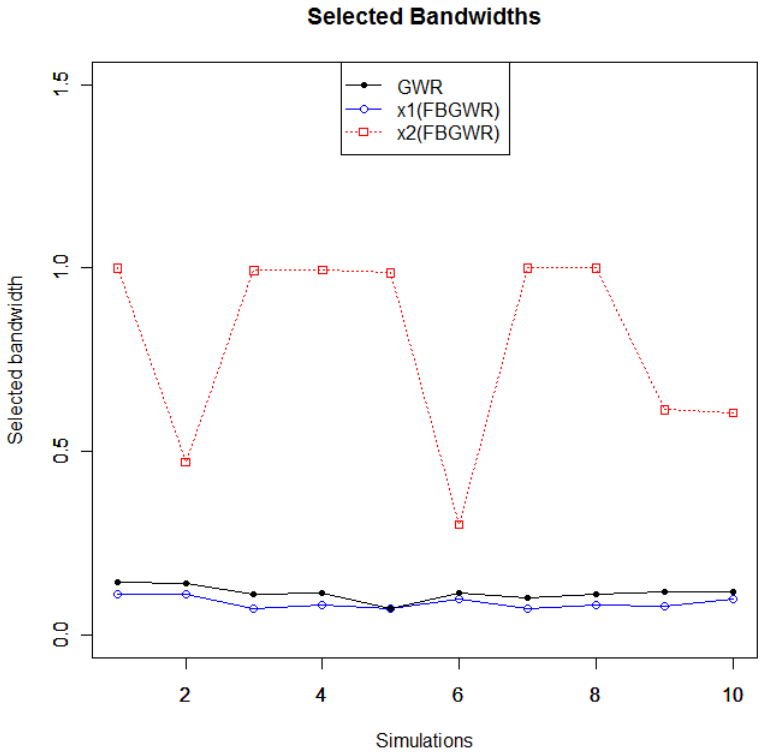
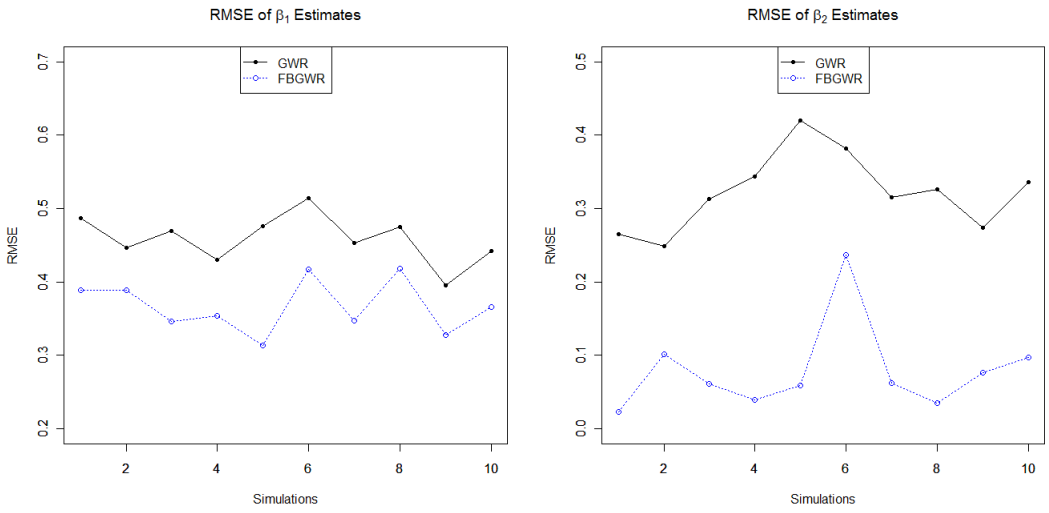


Figure 5-39 Bandwidths selected in 10 simulations

In FBGWR the bandwidths for β_1 are low in all 10 simulations while the bandwidths for β_2 are much larger, close to 1 in 6 out of 10 simulations. This reflects the nature of the two coefficients- β_1 varies at a very local spatial scale and β_2 is constant over space. The uniform bandwidths selected from GWR are between the two sets of bandwidths from FBGWR, as expected.

5.5.4.2 Algorithm accuracy

The RMSEs from FBGWR and basic GWR are compared in Figure 5-40.



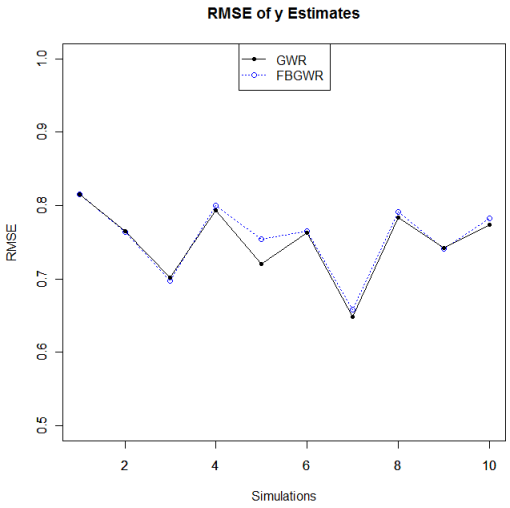


Figure 5-40 RMSEs of estimates in 10 simulations

The RMSEs of both β_1 and β_2 estimated from FBGWR are much lower than those from basic GWR. There is little difference in the RMSEs of y. It can be seen that although the two models have shown similar abilities in predicting the dependent variable, the uniform bandwidth from basic GWR is not as helpful as the flexible bandwidths in FBGWR in capturing either the local variability in β_1 or the global nature of β_2.

5.5.4.3 Estimated coefficients

Figure 5-41 visualizes the true coefficients surfaces for β_1 and β_2 and the mean of estimated coefficients from the 10 simulations.

For both β_1 and β_2, FBGWR reproduces the coefficient surfaces slightly better than basic GWR. It is interesting to notice that the β_2 surface estimated from FBGWR exhibits a weak pattern that is similar to the pattern in the β_1 surface, this is probably due to the interaction between the two regression terms during the model calibration.

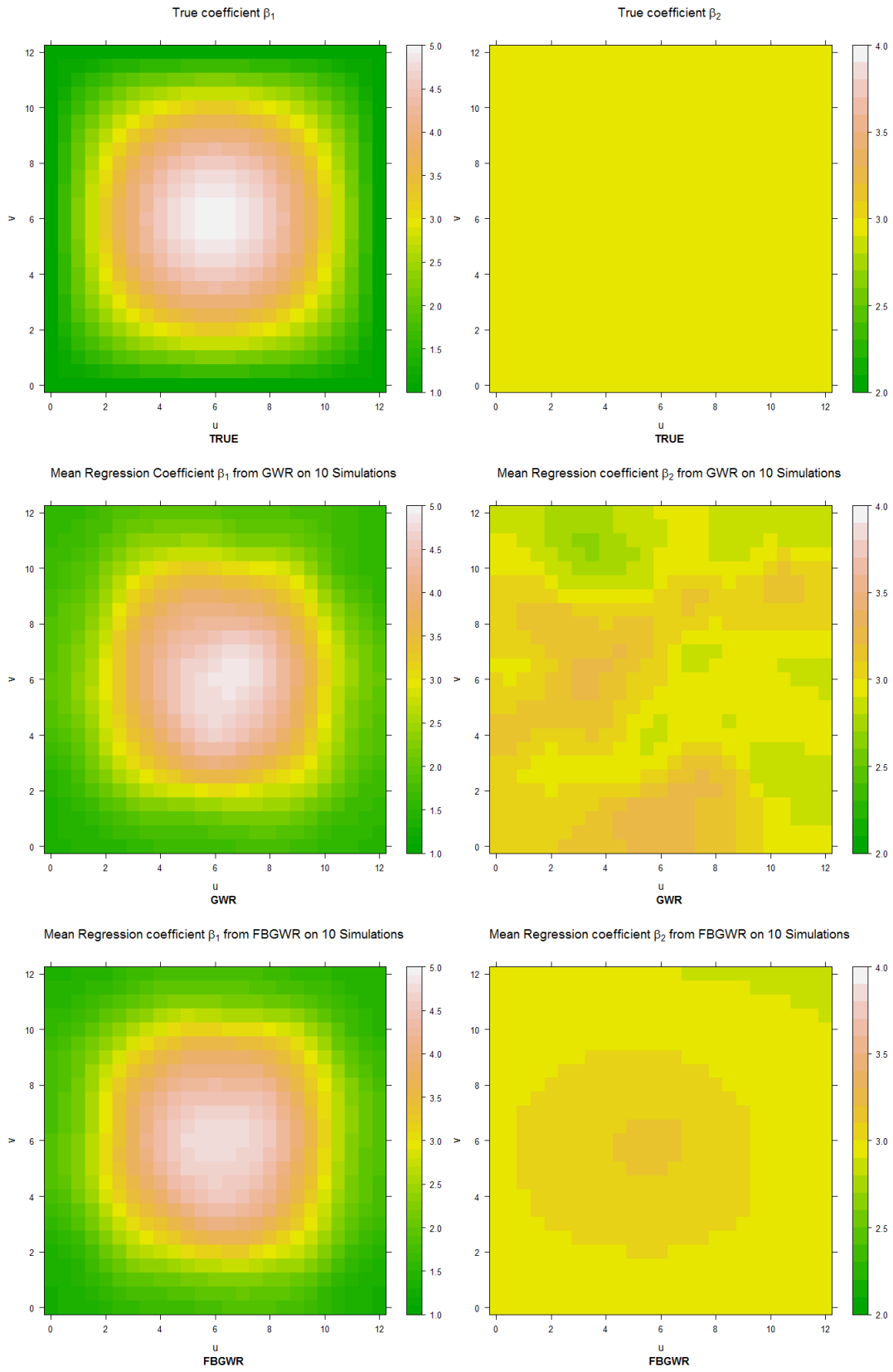


Figure 5-41 Estimated coefficient surfaces from basic GWR and FBGWR

5.5.5 Scenario 5: high heterogeneity β_1 + low heterogeneity β_2

In this scenario, the two coefficients β_1 and β_2 both vary locally but β_1 is set with higher heterogeneity than β_2.

$$\beta_1 = \beta_{high} \tag{5-28}$$

$$\beta_2 = \beta_{low} \tag{5-29}$$

5.5.5.1 Bandwidth selection

The bandwidths automatically selected for the coefficient associated with each variable x_1 and x_2 from FBGWR, as well as from the uniform bandwidth selected from basic GWR, are listed in Table 5-9 and visualized in Figure 5-42.

Table 5-9 Bandwidth selected in 10 simulations

	Sim1	Sim2	Sim3	Sim4	Sim5	Sim6	Sim7	Sim8	Sim9	Sim10
β_1 (FBGWR)	0.098	0.048	0.123	0.078	0.078	0.086	0.078	0.072	0.093	0.072
β_2 (FBGWR)	0.325	0.288	0.264	0.288	0.243	0.296	0.360	0.382	0.462	0.362
Basic GWR	0.112	0.112	0.112	0.085	0.110	0.102	0.123	0.117	0.142	0.117

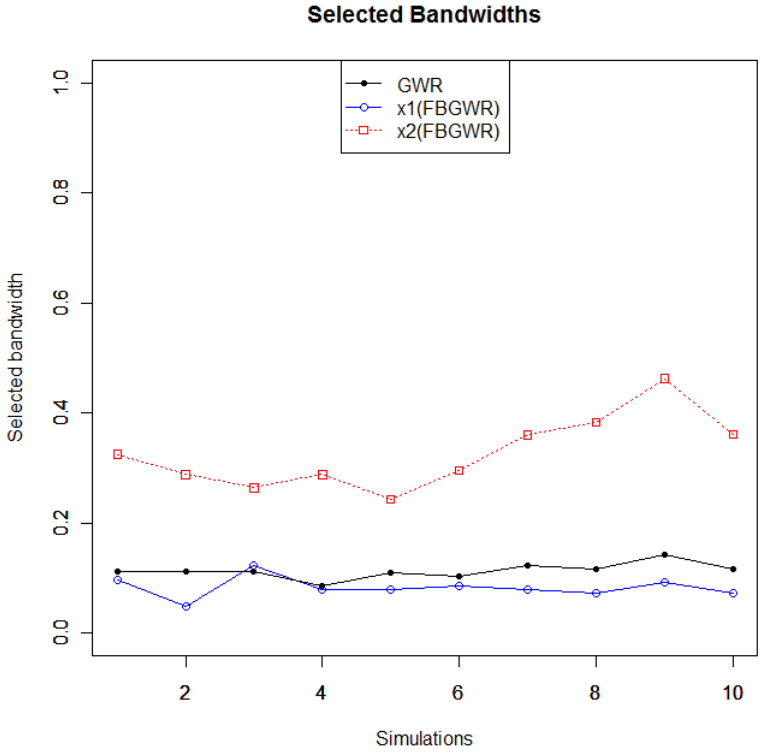


Figure 5-42 Bandwidths selected in 10 simulations

For FBGWR, the bandwidth selected for β_1 is smaller than that selected for β_2 in each simulation, reflecting the nature of the two coefficients- β_1 has a higher heterogeneity

than β_2. FBGWR therefore works successfully in distinguishing between the two levels of spatial variability in coefficients within the one model.

The uniform bandwidths selected from basic GWR are between the two sets of bandwidths from FBGWR, again demonstrating the limitation of basic GWR that it can only represent an average scale of spatial variability for all coefficients in the model.

5.5.5.2 Algorithm accuracy

The RMSEs of the two coefficients and of the dependent variable from basic GWR and FBGWR on each simulation are compared in Figure 5-43.

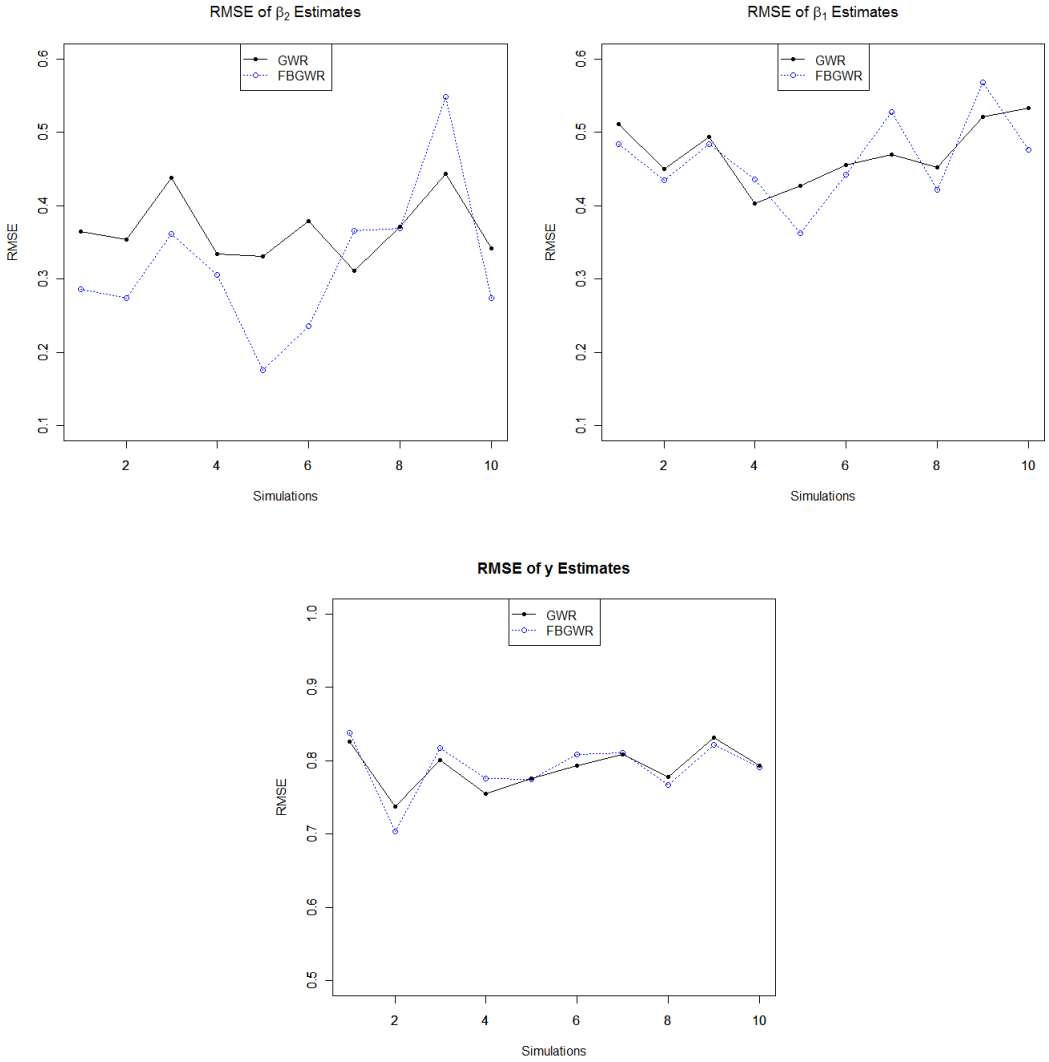


Figure 5-43 RMSEs of estimates in 10 simulations

As can be seen, the RMSEs of β_2 estimated from FBGWR are much lower than those from basic GWR in most simulations while in the case of β_1 there is not much

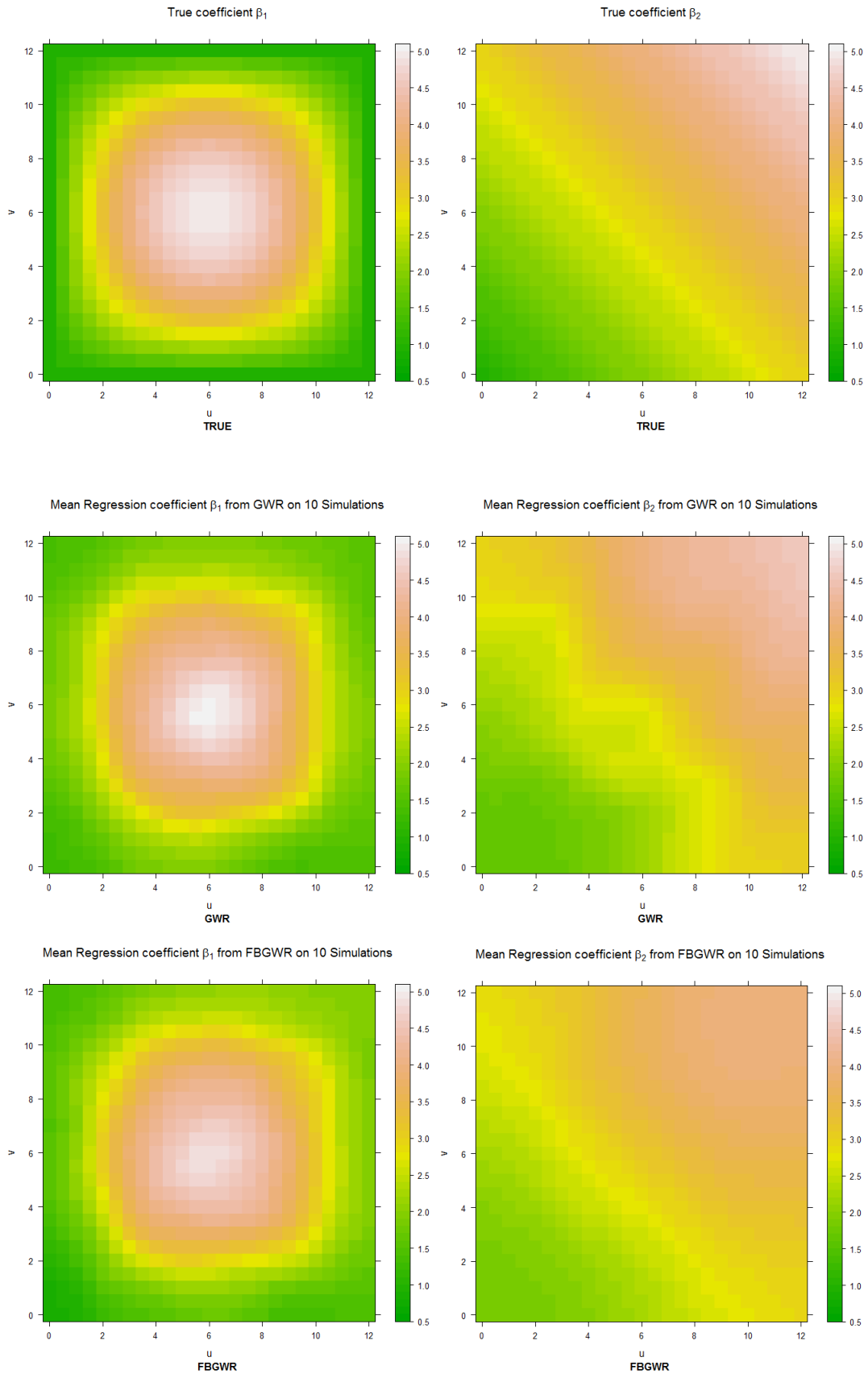


Figure 5-44 Estimated coefficient surfaces from basic GWR and FBGWR

difference. This is probably because that FBGWR specified larger bandwidths for β_2 and captured the true spatial pattern in β_2 better than basic GWR. While for β_1 there was not much difference in the bandwidths employed in the two models. Also for the dependent variable *y*, the RMSEs are similar for the two models.

5.5.5.3 Estimated coefficients

Figure 5-44 displays the true surfaces for coefficients β_1 and β_2 and the mean of the estimated coefficients from the 10 simulations by basic GWR and FBGWR respectively.

For β_1 both models reproduce the surfaces quite well, while for β_2 FBGWR performs better than GWR. This can be explained by the uniform bandwidths in basic GWR being too small to explain the true variability in β_2 which distorts the spatial pattern by introducing some anomalies. These findings are consistent with the results of the RMSEs.

5.6 Summary

Only selected experiments are presented in this chapter which is intended to demonstrate a practical algorithm for the calibration of FBGWR and to describe the performance of FBGWR in revealing the scale of spatial heterogeneity in coefficients under various situations. In most cases FBGWR performs well in finding the "correct" bandwidth for each coefficient: that is, an adaptive bandwidth of approximately 1 for a spatially stationary coefficient; a bandwidth less than 1 for a local varying coefficient; and smaller bandwidths for more locally varying processes. The ability to produce similar bandwidths for coefficients varying at the same scales demonstrates that FBGWR can serve as a basic GWR, while the ability to distinguish spatial stationary coefficients from non-stationary ones shows that FBGWR can also be a substitute for mixed GWR. Of course, FBGWR can do more beyond basic GWR and mixed GWR.

However, it is hard to define the "correct" bandwidth for a spatial varying coefficient even though the pattern of the coefficients is known. The ideal approach would be to start from determined bandwidths to generate coefficients and a dataset and then let the regression model reveal the bandwidths. This is unachievable based on current knowledge and methodologies unfortunately.

The FBGWR model has its weaknesses: the most notable one being the computational cost which will increase significantly with additional variables. Whether the model can accurately reflect the "true" coefficients also depends on the configuration of data including the scale of each variable and the error, and the comparison between them.

In real world applications, FBGWR is helpful in investigating the scale of the spatial non-stationarity in relationships between variables. On the other hand, experience and knowledge about the phenomena under study could possibly assist the determination of bandwidths and help to build a better performed model. A case study on a real dataset is described in the next chapter to demonstrate the utility of FBGWR.

Chapter 6 Case study on Irish Famine

6.1 Irish Famine data

6.1.1 Irish Famine

The Great Famine in Ireland in the late 1840s is the critical event in Irish population history: not only did it cause a severe population decline in the island of Ireland, but it also had a major effect on the populations in those countries which received Irish migrants, such as the UK, the US and Australia (Kenny, 2003). While the population decline was dramatic in the whole island, the effect was not distributed evenly (O'Grada and Eiríksson, 2006). Some areas suffered a more serious population decline than others, while a few areas, usually large towns, experienced population growth as migrants searched for relief and employment.

Although the failure of the potato crop due to potato blight is believed to be the main cause of the population decline (Bourke, 1993), this impact operated in a complex way, mixed with many other factors functioning as multiple processes. Previous research has revealed several demographic and land use factors that affected population change such as the overpopulation in Ireland before the famine (Connell, 1975), the landownership system that had forced the majority of the population to live on marginal lands (Kinealy, 1997), and locational issues such as topography and proximity to workhouses (Guinnane and O'Grada, 2000), urban areas and the coast (O'Grada and Eiríksson, 2006).

Localised research on the famine suggests that the processes by which various contributing factors affected population change might vary locally but it is impossible to quantify these processes by analysis at the national scale (Gregory and Ell, 2005, Fotheringham et al., 2013). Therefore, local rather than global models are required to help us understand the spatial variations in population decline across Ireland as a result of the famine.

6.1.2 Data source

The dataset examined here is derived from the work of Fotheringham et al., (2013) and Kelly and Fotheringham (2011) who for the first time mapped population change as well as potential determinants of population change at the spatial scale of Electoral

Divisions (EDs), the administrative units through which census data are collected in Ireland. This is a finer spatial scale than that has ever been achieved in previous studies on Irish famine. The dataset consists of 3,436 EDs, for each of which 14 potential explanatory variables as listed in Table 6-1 were collected. The variables were derived from the 1841 and 1851 Population Census and the 1851 Agricultural Census. More details on the data can be found in Fotheringham et al.(2013).

Table 6-1 List of potential explanatory variables

Variable Name	Variable Explanation
DEMOGRAPHIC	
PopCrop_41	population in 1841 per acre of cropped land in 1851 (%)
Ratio1841	male/female population ratio in 1841 (%)
Perc_Towns	percentage of population living in towns in 1841 (%)
PPB	persons per building in 1841 (%)
UNIINHABPCT	percentage of uninhabited dwellings in 1841 (%)
LAND USE	
Crops_Hold	average holding size
VALUATION	land value per hectare in 1841
Potat_Cult	percentage of crop land under potatoes in 1851 (%)
PCorn_Cult	percentage of crop land under grain in 1851 (%)
Perc_Agri	percentage of ED under agriculture 1851 (%)
LOCATIONAL	
MEAN_ELEV	mean elevation
ACC41_20	accessibility to urban areas
Coast_Dist	distance to coast
WHOUSE	proximity to workhouses

The response variable is “Perc_Chang”, percentage of population change from 1841 to 1851. Figure 6-1 maps this variable across the island of Ireland. The EDs are classified into nine classes based on the quantile classification, with each class having the same number of EDs and are plotted with gradient colors. The EDs with population growth, all falling into the last class, are marked as green to be distinguished. In general, population decline was more severe in the west and the south of the island than in the east and the north. The greatest decline in population was over 60 per cent. The majority of areas with population decline is interspersed with a few places that experienced population growth, mainly the larger settlements.

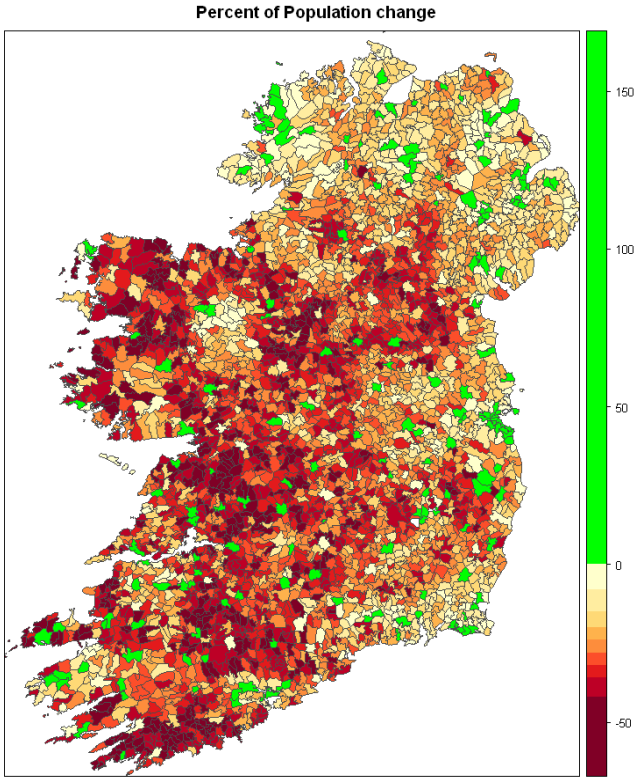


Figure 6-1 Percentage of population change from 1841 to 1851 in Ireland

The potential explanatory variables can be divided into three categories: demographic, land use and locational. The demographic data mainly comes from 1851 population census which records data for both 1841 and 1851. "PopCrop_41" measures population density on cropped land, suspected to be one of the main factors contributing to population decline; "Ratio1841" represents male/female population ratio and indicates peripherality and marginality; "Perc_Towns" stands for the clustering of population in urban settlements; "PPB" represents persons per building, an indicator of over-crowding; and "UNIINHABPCT" represents the population of uninhabited houses in 1841, indicating areas of pre-famine out-migration.

Agricultural censuses in Ireland began in 1847 but the data in early years was not complete and compatible with the population census until 1851. Because of this, the land use factors were mainly approximated from the 1851 census, which might introduce some uncertainties to the analysis. According to the analysis of Fotheringham et al. (2013), and supported by historical record (Bourke, 1965), the total acreage of crop land did not change much from 1841 to 1851, neither did the relative spatial distributions of various crops, although the absolute amounts of individual crops did. Therefore, the 1851 data can be used as a suitable source of

agriculture data, with the caveat that these variables might be "the outcome of processes that occurred before, during or after the famine" (Fotheringham et al., 2013, P. 225). “Crops_Hold” represents the average holding size with large holding size indicating places with more land-intensive and less labour-intensive farming; "VALUATION" stands for land value per hectare in 1841; "Potat_Cult" and "PCorn_Cult" represents the percentage of crop land under potatoes and grain respectively; "Perc_Agri" is the percentage of an ED under agriculture in 1851.

Locational variables are: “MEAN_ELEV”- the mean elevation, an indicator of marginality; “ACC41_20” measures the accessibility to urban areas, indicating the possibility of rural-to-urban migration. “Coast_Dist” represents distance to coast, suggesting the possibility of obtaining alternative sources of food which could help to release people from the suffering of famine; and “WHOUSE” stands for the proximity to workhouses, another source of relief.

6.1.3 Data preparation

Before regression models were built from the dataset, a sequence of preprocessing steps was undertaken on the dataset to improve model calibration.

Step 1: Excluding certain EDs

As the interest of this study lies in the processes of population loss caused by the famine in rural areas, two types of EDs were excluded before analysis to reduce bias, as guided by the previous analysis by Fotheringham et al. (2013):

- EDs with population growth during the famine years. These tend to be urban EDs or EDs containing institutions, such as workhouses, hospitals or prisons.
- EDs with population densities greater than 500 people per square kilometres in 1841. These are urban areas without an agricultural base where different population changing processes were taking place.

After excluding these two types of EDs, 3250 EDs remained in the dataset.

Step 2: Sampling

To reduce computational complexity with the FBGWR models, whilst at the same time preserving the spatial pattern of relationships as much as possible, a spatially representative sample containing 446 EDs was drawn through a stratified sampling

method. This was accomplished with the R package sp (Pebesma and Bivand, 2005, Bivand et al., 2013). Figure 6-2 describes the percent of population change in the whole dataset and the spatial sample. The blank EDs in the left map are those that have been excluded in step 1.

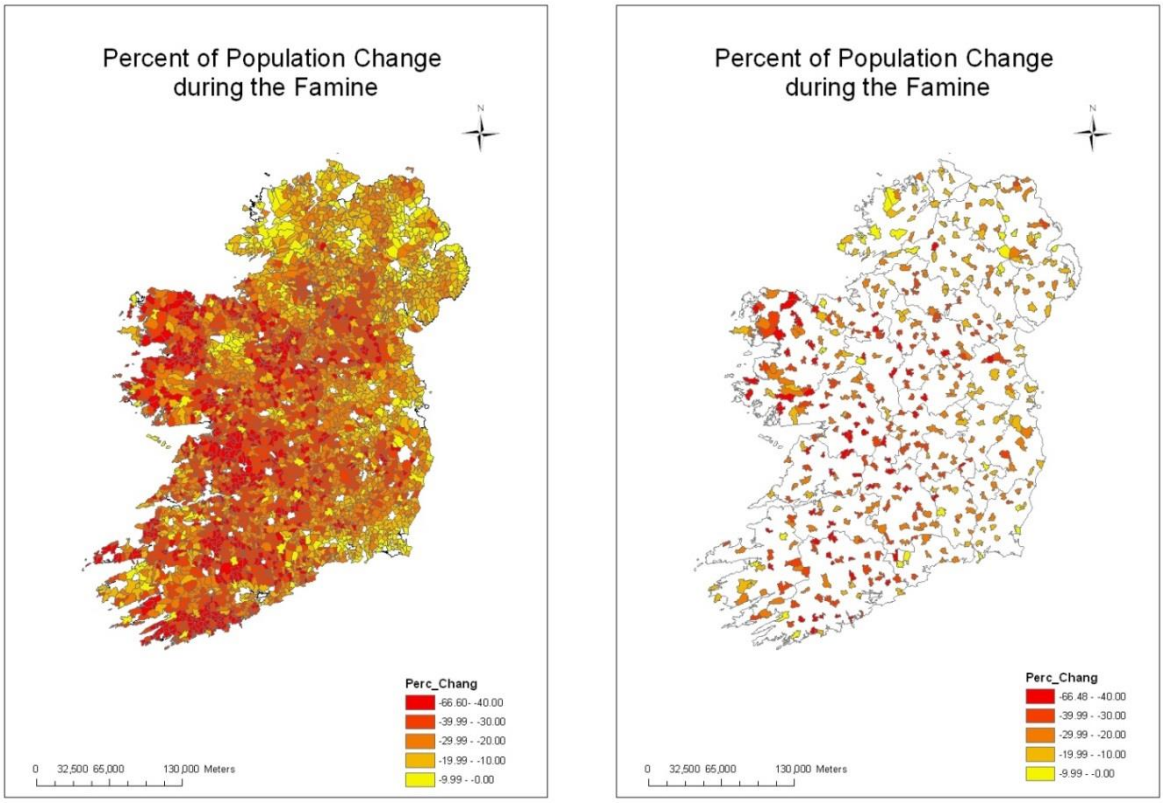


Figure 6-2 Spatial sample from the Irish Famine dataset

Left: the whole dataset (3250EDs), the sample dataset (446EDs)

Step 3: Variable Transformation

The skewness of each explanatory variable as well as the dependent variable was checked in the sample dataset, with results listed in the table below. Here the skewness is defined as the adjusted Fisher-Pearson standardized moment coefficient:

$$\frac{n}{(n-1)(n-2)}\sum_{i=1}^{n}\left(\frac{x_i-\bar{x}}{s}\right)^3 \qquad (6\text{-}1)$$

where n is the sample size and s is the standard deviation of the sample.

Two variables, “PPB” and “WHOUSE” were highly skewed, with skewness above 3. These two variables were transformed into their natural logarithms, the skewnesses after transformation being 1.13 and 1.88 respectively. The two new variables after transformation are named as “lnPPB” and “lnWHOUSE”.

Table 6-2 Skewness checking of the variables

Variable Name	Skewness	Skewness after transformation
Crops_Hold	1.07	
PopCrop_41	2.90	
Ratio1841	0.30	
Perc_Towns	2.50	
PPB	3.11	1.13
UNIINHABPCT	2.66	
VALUATION	2.06	
Potat_Cult	0.98	
PCorn_Cult	-0.41	
MEAN_ELEV	0.65	
Perc_Agri	0.34	
ACC41_20	0.68	
Coast_Dist	0.82	
WHOUSE	6.56	1.88
Perc_Chang	0.15	

Step 4: Redundancy Elimination

To eliminate the possible redundancy among the 14 variables, a multicollinearity check was carried out through VIFs (Variance Inflation Factor). Firstly, regression was performed with all the 14 variables, and VIFs were checked as below:

Table 6-3 VIF check on 14 variables

Variable	VIF
VALUATION	2.047007
UNINHABPCT	1.256381
lnPPB	1.501138
Coast_Dist	1.458678
Perc_Towns	1.498848
ACC41_20	2.210964
lnWHOUSE	1.416531
PopCrop_41	2.327902
Perc_Agri	4.081212
Potat_Cult	1.748064
PCorn_Cult	1.719914
Ratio1841	1.438118
Crops_Hold	2.423254
MEAN_ELEV	1.456236

The variable Perc_Agri had a high VIF over 3. This variable was removed and regression was repeated on the remaining 13 variables, with VIFs listed as below:

Table 6-4 VIF check on 13 variables

Variable	VIF
VALUATION	1.82413
UNINHABPCT	1.255379
lnPPB	1.493249
Coast_Dist	1.458401
Perc_Towns	1.472557
ACC41_20	1.659255
lnWHOUSE	1.397937
PopCrop_41	2.228315
Potat_Cult	1.747981
PCorn_Cult	1.65112
Ratio1841	1.385629
Crops_Hold	2.226439
MEAN_ELEV	1.320485

No VIF was greater than 3, the 13 variables were therefore retained for model building. It is worthy to notice that although the global collinearity between variables has been effectively removed, collinearity may still exist locally when geographical weighting scheme is employed in GWR. This will be examined in Section 6.4.

In Section 6.2, various forms of GWR and FBGWR models as well as a global model are built on the prepared dataset and the results are compared in Section 6.3.

6.2 Model building

6.2.1 Global model

The regression analysis starts with a global multivariate linear regression model. A stepwise model selection procedure by AIC was performed with the 13 independent variables, using function stepAIC() in R package MASS (Venables and Ripley, 2002). A subset of 9 independent variables was selected to constitute an optimal regression model. These 9 variables are: VALUATION, UNINHABPCT, Coast_Dist, ACC41_20, PopCrop_41, Potat_Cult, PCorn_Cult, Ratio1841, Crops_Hold. The parameter estimates together with other diagnostics of this model are listed in Table 6-5:

Table 6-5 Parameter estimates from the 9-variable global regression model

	Estimate	Standard Error	t value	Pr(>|t|)	
(Intercept)	19.7136	10.3739	1.900	0.058053	.
VALUATION	0.0015	0.0003	5.520	5.83E-08	***
UNINHABPCT	0.8600	0.2093	4.109	4.75E-05	***
Coast_Dist	-0.0394	0.0276	-1.426	1.54E-01	
ACC41_20	-0.0031	0.0009	-3.547	4.32E-04	***
PopCrop_41	-3.8868	0.5996	-6.482	2.45E-10	***
Potat_Cult	0.1523	0.0790	1.929	0.0544	.
PCorn_Cult	-0.1473	0.0500	-2.943	0.003429	**
Ratio1841	-0.2914	0.0927	-3.144	0.001781	**
Crops_Hold	-0.4244	0.1151	-3.687	2.55E-04	***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 10.16 on 436 degrees of freedom

Multiple R-squared: 0.2546, Adjusted R-squared: 0.2392

F-statistic: 16.54 on 9 and 436 DF, p-value: < 2.2e-16

Classic AIC: 3345.815

Five factors show the most significant impacts on population decline: VALUATION, UNINHABPCT, ACC41_20, PopCrop_41, Crops_Hold. Among these five factors, VALUATION and UNINHABPCT have positive relationships with population change, which means population decline during the famine decade was more severe in the places where the land value per hectare in 1841 and the percentage of uninhabited dwellings in 1841 was lower, in other words, where the land was poorer and where the population pressure was greater. The other three factors have negative relationships with population change suggesting that greater population decline was observed in areas which were nearer to urban areas, where population (in 1841) per acre of cropped land (in 1851) was higher, and where the average holding size was larger. Greater accessibility to urban areas facilitated migration of people from rural areas, while the latter two factors represent population pressure on the land and farming practice. The R-squared values are however low, indicating that the model does not explain the spatial variations in population change very well.

6.2.2 Basic GWR

To investigate how the spatial variation in population change might have resulted from various local processes, GWR was used to calibrate the regression model. To construct the appropriate GWR model, an approximate stepwise AIC procedure

(Fotheringham et al., 2010) adapted from the standard forward stepwise AIC procedure for multivariate linear regression was carried out.

The procedure can be described as:

Step 1: Place each of the 13 variables in turn in a basic GWR model with the dependent variable Perc_Chang regressed on single independent variable. Calculate AICc, a small sample bias corrected AIC (Fotheringham et al., 2002) in each case. Select the variable yielding the lowest AIC. Set N = 13.

Step 2: Place each of the remaining N-1 explanatory variables in the GWR model regressed on the variable(s) selected in the former step(s) and the new variable. Select the variable yielding the lowest AIC. Add this variable to the model. Set N = N - 1.

Step 3: Repeat Step2 until there is no AICc reduction.

These steps were implemented with the GWR software GWR3.0 (Charlton et al., 2007). The change in AICc during the procedure is plotted in Figure 6-3:

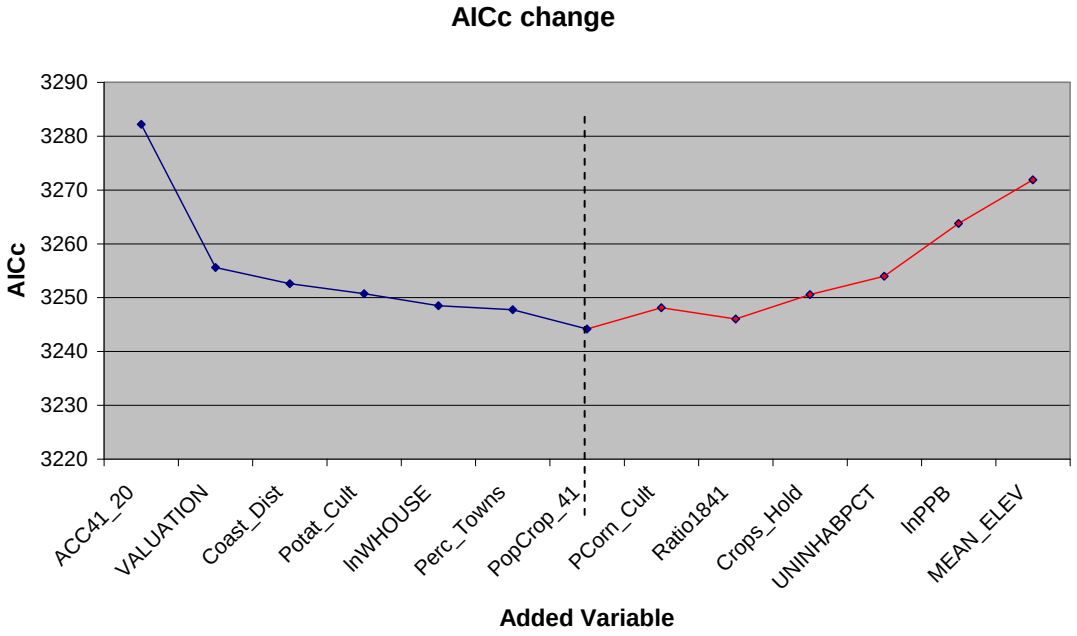


Figure 6-3 An approximate stepwise AIC procedure for GWR

After 7 variables had entered the model, the AICc started to increase when any new variable was added to the model and therefore the model selection procedure stopped and those 7 variables were retained in the model. A basic GWR model as well as a new global multivariable linear regression model was calibrated with the 7 independent variables.

The results of the global model calibration are presented in Table 6-6:

Table 6-6 Parameter estimates from the 7-variable global regression model

	Estimate	Standard Error	t value	Pr(>|t|)	
(Intercept)	-28.8165	4.0583	-7.102	5.05E-12	***
VALUATION	0.0016	0.0003	5.126	4.45E-07	***
Coast_Dist	-0.0308	0.0278	-1.109	0.2682	
Perc_Towns	0.1412	0.0462	3.058	0.0024	**
ACC41_20	-0.0031	0.0009	-3.334	0.0009	***
lnWHOUSE	-1.6090	0.7807	-2.061	0.0399	*
PopCrop_41	-2.5880	0.5134	-5.041	6.79E-07	***
Potat_Cult	0.2399	0.0749	3.201	0.0015	**

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 10.62 on 438 degrees of freedom

Multiple R-squared: 0.1824, Adjusted R-squared: 0.1693

F-statistic: 13.96 on 7 and 438 DF, p-value: 2.283e-16

Three of the five independent variables that had significant effects on population change in the previous 9-variable global model remained significant with very similar estimates and t values, while the other two independent variables did not enter the 7-variable model.

In the GWR calibration, a bi-square kernel function was employed and an adaptive bandwidth of 87 out of the 446 observations was automatically selected according to the AICc criterion. The following table gives summary statistics for the local parameter estimates from the basic GWR model.

Table 6-7 Summary statistics of parameter estimates from basic GWR model

Variable	Min	Max	Median	Mean	STD
Intercept	-97.740	20.600	-19.710	-21.830	24.868
VALUATION	-0.0030	0.0066	0.0004	0.0007	0.0015
Coast_Dist	-0.3679	0.4927	-0.0506	-0.0147	0.1762
Perc_Towns	-0.1196	0.7288	0.1635	0.1908	0.1676
ACC41_20	-0.0268	0.0059	-0.0050	-0.0055	0.0060
lnWHOUSE	-9.2040	4.6620	0.5348	-0.5137	3.3416
PopCrop_41	-9.1060	5.5220	-2.0770	-2.1260	2.7353
Potat_Cult	-0.2625	1.0670	0.2378	0.2856	0.3230

All local coefficients changed their signs across the dataset, revealing positive relationships in some local areas and negative relationships in others, while the mean of each set of local coefficients is in accordance with the global coefficient in terms of sign and magnitude. This suggests that although each factor had a general impact on population change across the country, indicated by the global analysis, there is some

local variation in this impact which needs to be investigated. Boxplots of the coefficients are displayed in Figure 6-4 to visualize the distribution of each coefficient, with the global estimates marked below each coefficient for comparison.

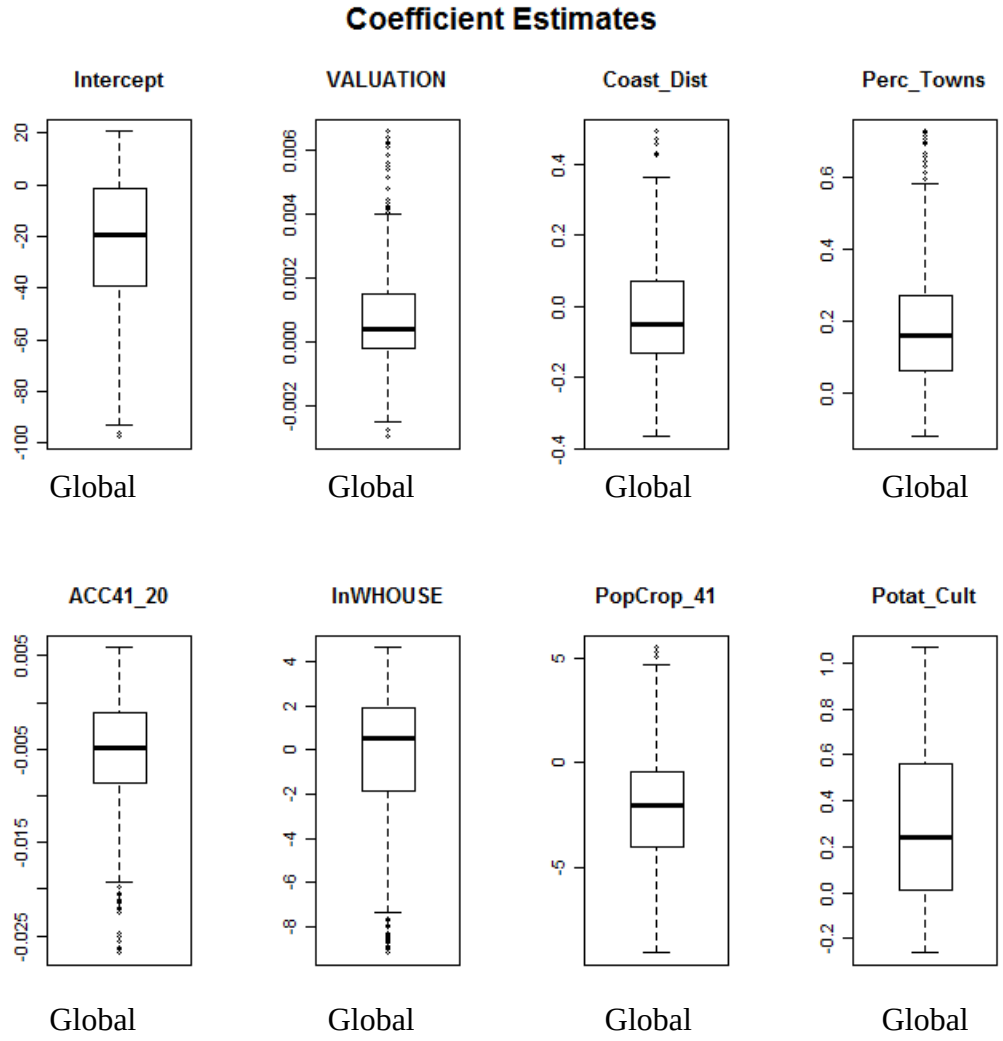


Figure 6-4 Boxplots of parameter estimates from basic GWR model

The R squared value of the basic GWR model is 0.6139, and the adjusted R squared is 0.4908, indicating an improved explanation of variability in population change by GWR over the global model.

6.2.3 FBGWR with auto-selected bandwidths

To examine whether the impact of each factor on population change might vary at different spatial scales, a FBGWR calibration was undertaken. Firstly, the bandwidth for each coefficient was automatically selected through a synchronous bandwidth optimization method as described in Section 5.2.3 in Chapter 5.

The kernel function employed was a bi-square and an adaptive bandwidth was used. Figure 6-5 illustrates how the optimal bandwidth for each coefficient was found during the algorithm. Only three parameters, the intercept and those associated with the independent variables VALUATION and Perc-Towns had "local" bandwidths; that is, the coefficients that vary at local scale, while the remaining coefficients tend to be global. It is worthwhile to note that, a distance-decay kernel function of bi-square is employed here, even if an adaptive bandwidth of 100% is used, it is still not equal to a global regression. The resulting optimal bandwidths for all coefficients expressed both as a proportion of all observations involved in the weighting scheme and as the number of nearest neighbours are listed in Table 6-8. Compared to the bandwidth of 0.1973053 (about 87 nearest neighbors out of 446 observations) in the basic GWR model as described in Section 6.2.3, the bandwidths here tend to two extremes, quite small for the intercept and rather large for the other coefficients. The very small bandwidth for the intercept, however, may suggest that the model tends to be wrapping itself around the intercept, that is, the variations in the dependent variable are mainly reflected in the local values of the intercept rather than being explained by the rest of the model.

Table 6-8 Bandwidths automatically selected from FBGWR model

Variable	Optimal Adaptive Bandwidth (proportion)	Optimal Adaptive Bandwidth (nearest neighbors)
Intercept	0.056035	25
VALUATION	0.672622	300
Coast_Dist	0.999941	446
Perc_Towns	0.914804	408
ACC41_20	0.999955	446
lnWHOUSE	0.999925	446
PopCrop_41	0.999919	446
Potat_Cult	0.999926	446

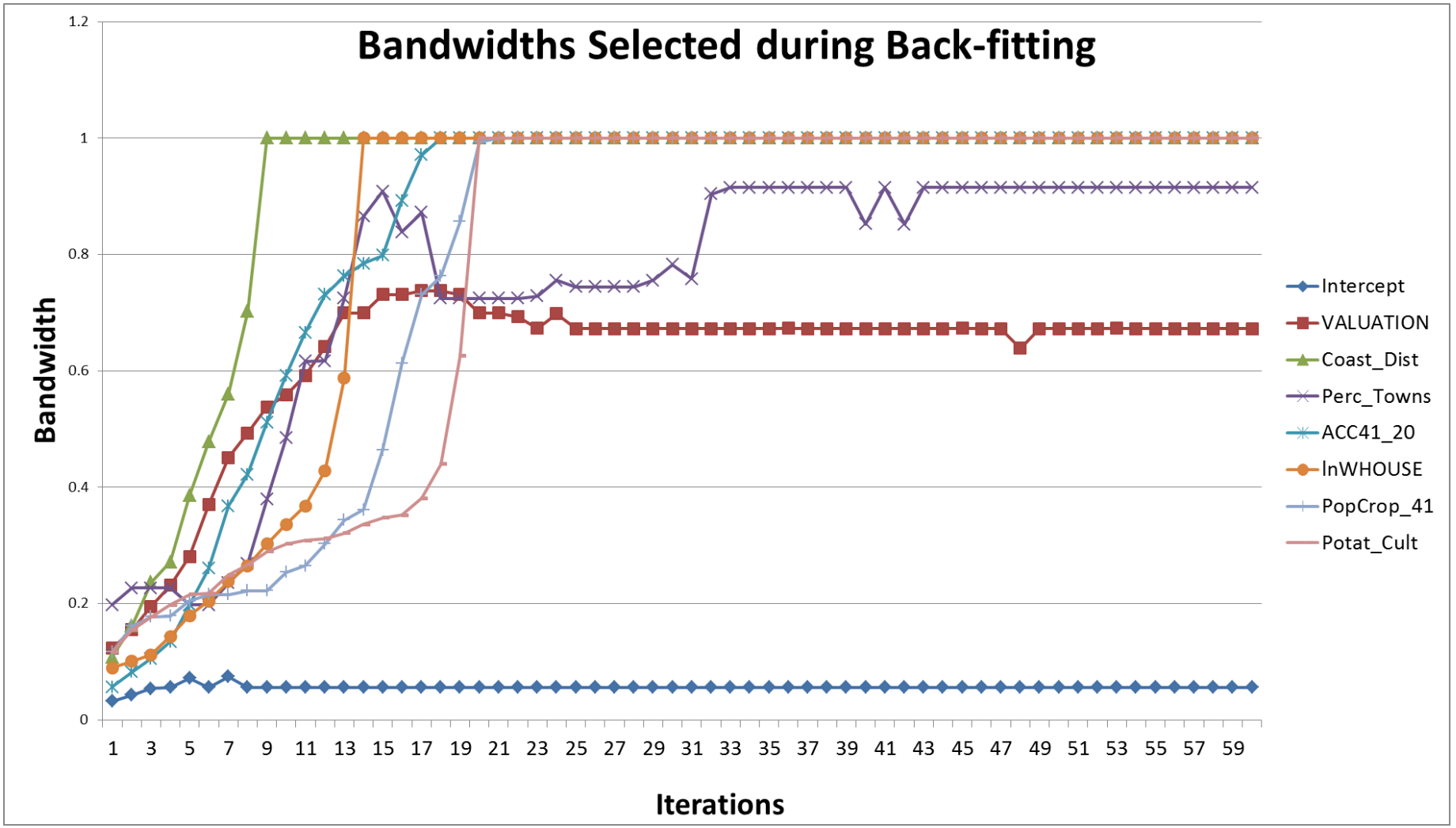


Figure 6-5 Bandwidths optimization procedure of FBGWR model

A FBGWR model with these bandwidths was calibrated and the parameter estimates and diagnostics of the model are discussed in Section 6.3.

6.2.4 FBGWR with a priori bandwidths

A different approach to specify flexible bandwidths for each coefficient in a FBGWR model is to resort to *a priori* knowledge as a guide. Here a simple method employing the results from basic GWR is demonstrated. A series of basic GWR models with a single independent variable was calibrated, each including one of the 7 independent variables and the bandwidth automatically selected from each model was assigned to that variable to be used in the FBGWR model. For the intercept, the bandwidth was taken from the 7-variable basic GWR model as described in Section 6.2.2. Table 6-9 contains the bandwidth for each coefficient.

Table 6-9 Bandwidths automatically selected from Basic GWR models

Variable	Optimal Adaptive Bandwidth (proportion)	Optimal Adaptive Bandwidth (nearest neighbors)
Intercept	0.197305	87
VALUATION	0.089627	39
Coast_Dist	0.109849	48
Perc_Towns	0.116559	51
ACC41_20	0.121061	53
lnWHOUSE	0.098661	44
PopCrop_41	0.115606	51
Potat_Cult	0.114349	50

Again, a bi-square kernel function with adaptive bandwidth was employed and the bandwidths are expressed in two forms, one a proportion and the other the number of nearest neighbours. These bandwidths are much smaller than those automatically found in the previous sub-section except the one for the intercept. This means that the coefficients will be allowed to vary at a finer spatial scale and therefore more local details will be able to be explored. A FBGWR model with these new bandwidths was calibrated and the results compared.

6.2.5 Mixed GWR

MGWR can be seen as a hybrid of global linear regression and GWR where some coefficients are allowed to vary locally over space while others are fixed globally. A

key decision in building an MGWR model is to decide whether a coefficient should be varying or fixed. The FBGWR technique could be a practical approach to assist this decision. The coefficients for which rather global bandwidths have been assigned automatically in FBGWR could be taken as fixed ones, while others could be designated as spatially varying.

Based on the bandwidth optimization results in the FBGWR model in Section 6.2.3, an MGWR model was calibrated where the intercept and coefficients of the two independent variables VALUATION and Perc-Towns were designated as varying terms, while the others were designated as fixed. The bandwidths for the varying terms were automatically selected through the AICc criterion, resulting in 0.23 in a proportion form, which is equivalent to 102 out of 446 observations. The results from all the models described in this section are compared and interpreted in the next section.

6.3 Model comparison

Five forms of regression models have been calibrated on the Irish famine dataset with 7 independent variables: a global regression model, a basic GWR model, two FBGWR models and a mixed GWR model. The bandwidths employed in each model and their overall performance in terms of residual sum of squares (RSS) and AICc are listed in Table 6-10.

Table 6-10 Model comparison on bandwidths and overall model performance

		Global regression	Basic GWR	Mixed GWR	FBGWR-1	FBGWR-2
Bandwidth	**Intercept**	global	0.20	0.23	0.06	0.20
	VALUATION	global		0.23	0.67	0.09
	Coast_Dist	global		global	1.00	0.11
	Perc_Towns	global		0.23	0.91	0.12
	ACC41_20	global		global	1.00	0.12
	lnWHOUSE	global		global	1.00	0.10
	PopCrop_41	global		global	1.00	0.12
	Potat_Cult	global		global	1.00	0.11
AICc		3383.44	3240.38	3297.34	*3132.18*	*3104.15*
RSS		49381.86	23319.25	35483.03	28026.13	24685.58

"FBGWR-1" stands for the FBGWR model with auto-selected bandwidths, while "FBGWR-2" is the model with *a priori* bandwidths based on basic GWR. The 9-

variable global regression model is not included in the table because the difference in the numbers of variables makes the comparison of little worth. When a variable has a globally constant coefficient in the model, the bandwidth is simply marked as "global". It is worth pointing out that for the FBGWR models, a normal AIC or AICc value cannot be obtained at present due to the difficulty in defining the number of parameters for the whole model and the AICc value presented in the table for FBGWR model is actually the mean of the AICc values from each fitted term in the FBGWR model. These mean AICc values are not directly comparable with those of other models.

The global regression model has the highest RSS and the highest AICc value, suggesting that the global model performs the worst in fitting the data and results in more information loss in representing the true process than other models. The mixed GWR model yields the highest RSS among the GWR models and also has a higher AICc value than the basic GWR model suggesting that the mixed GWR model fails to fit the data very well either. This indicates that it is not appropriate to assume globally constant relationships between the independent variables and population change across the whole country and that local regression is necessary to reveal local effects and to guide further detailed studies. Both the basic GWR model and the mixed GWR model results suggest rather small bandwidths for the locally varying coefficients; this also indicates that the determining factors are acting at a rather local scale. It is interesting that the basic GWR model yields the smallest RSS amongst all the models, this is possibly due to the small bandwidths used in the basic GWR which have helped to improve the model fit. Although FBGWR-2 also employs rather small bandwidths, the calibration procedure based on back-fitting has caused loss of precision, resulting in increased residuals. Due to the lack of comparable AICc values, it is hard to conclude the relative quality between the basic GWR model and the FBGWR models although the mean AICc values from both FBGWR models are smaller. As to the two FBGWR models, both result in fairly small RSS values with FBGWR-2 having a smaller RSS and slightly lower mean AICc value than FBGWR-1 Again, this is probably due to the fact that much smaller bandwidths were employed in FBGWR-2.

6.4 Collinearity diagnostic

As collinearity between variables may degrade coefficient estimate precision in GWR and potentially invalidate interpretation about the coefficients, locally weighted VIFs are checked here as a diagnostic of collinearity. At the location of each ED, a local VIF is calculated for each independent variable with respect to the corresponding geographical weighting scheme. While VIFs that exceed 10 are usually regarded as problematic (Brunsdon et al., 2012), local VIFs greater than a more strict value of 3 are counted and reported here to be more cautious and to be consistent with the global multicollinearity check in Section 6.1.3. Table 6-11 lists the count of locations that have local VIFs greater than 3 for each variable in the context of various GWR models.

Table 6-11 Count of locations with local VIFs greater than 3 in each GWR model

Variable	Basic GWR	Mixed GWR	FBGWR-1	FBGWR-2
VALUATION	19	11	0	120
Coast_Dist	15	0	0	34
Perc_Towns	16	2	0	45
ACC41_20	81	0	0	103
lnWHOUSE	2	0	0	93
PopCrop_41	0	0	0	24
Potat_Cult	0	0	0	15

The potentially problematic count increases as the bandwidth for a variable decreases. While the FBGWR model with *a priori* bandwidths raises more risky locations with high VIFs than other models, the FBGWR model with auto-selected bandwidths has reduced the number of risky locations to zero by employing relatively large bandwidths.

Moreover, there are only 3 locations with VIFs greater than 10, all for the variable "Perc_Towns" in the FBGWR model with *a priori* bandwidths. Collinearity is not too much of a problem here, although cautions should be taken when interpreting the coefficient estimates from the basic GWR and the FBGWR model with *a priori* bandwidths, especially for the variables "Perc_Towns", "ACC41_20", "VALUATION" and "lnWHOUSE".

6.5 Interpretation

Coefficients relating to various bandwidths have been estimated from five models. Figures below display the boxplots of the estimated coefficients as well as the corresponding T values for the intercept and each variable. The global parameters and T values are also included for comparison. According to the Fotheringham adjustment for multiple hypotheses testing for GWR model coefficients (Byrne et al., 2009, Fotheringham et al., 2013), T values in excess of 3.28 suggest that the corresponding parameter estimates are significantly different from zero

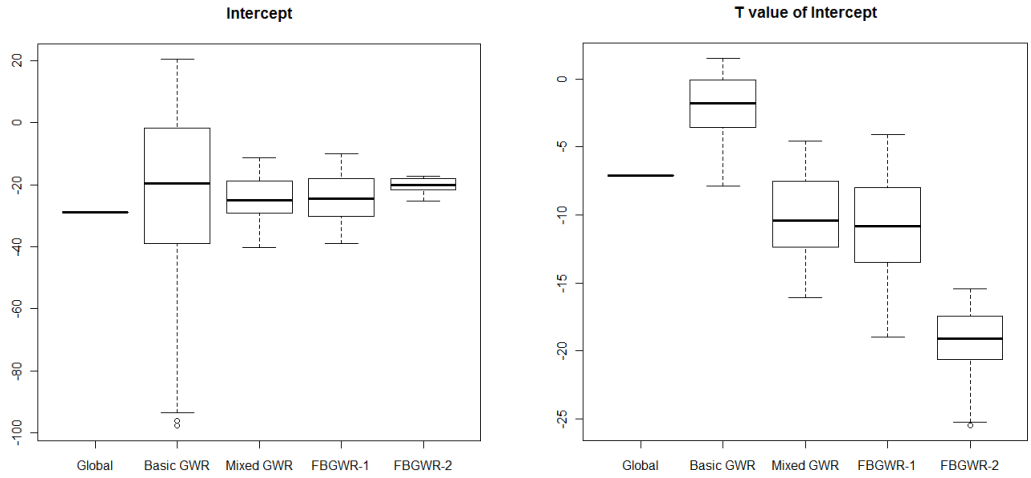


Figure 6-6 Boxplots for coefficient estimates and T values of the intercept from various models

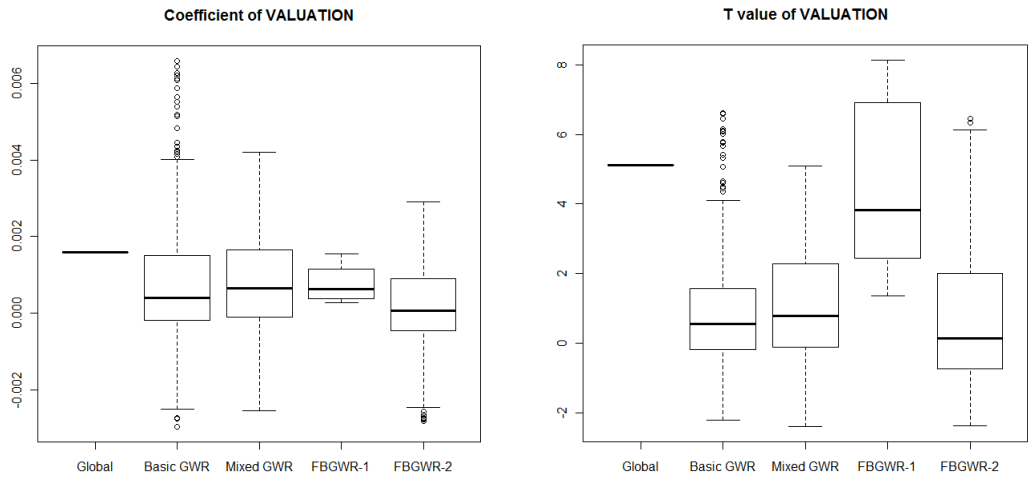


Figure 6-7 Boxplots for coefficient estimates and T values of land value per hectare in 1841 from various models

(i) land value per hectare in 1841

For the factor "VALUATION" (land value per hectare in 1841), a coefficient of 0.0016 was estimated from the global model, suggesting a positive relationship between land value in 1841 and population decline during the famine decade. That is, areas with lower land values in 1841 suffered more from population loss during the famine. The coefficients estimated from the FBGWR model with an auto-selected bandwidth vary from 0.00028 to 0.00156, also indicating positive relationships. While the other three GWR models found some negative relationships in a few locations, these relationships were not significant. The maps in Figures 6-8, 6-9, 6-10 and 6-11 display the spatial distribution of the varying coefficients together with their associated T values. For the T values, only areas with T values in excess of 3.28 are mapped. Although the four GWR models show slightly different patterns in the coefficients, all four maps suggest significant positive relationships only in the south of Ireland, where the poverty was the most severe before the famine (Mokyr, 1985). The results suggest that land value in these areas largely affected the depopulation during the famine with poorer land value caused more victims. The FBGWR with auto-selected bandwidths revealed more areas with this significant effect.

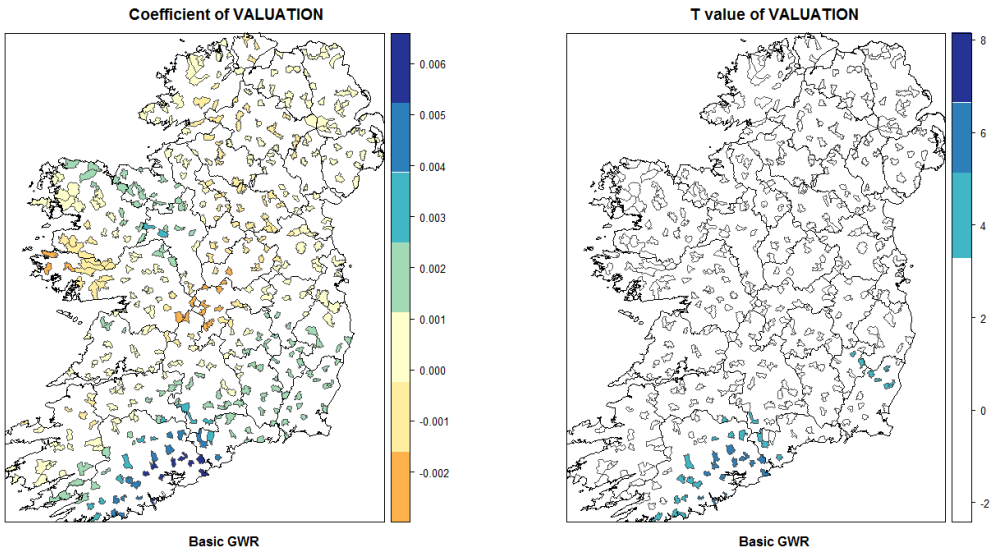


Figure 6-8 Maps of coefficient estimates and T values of land value per hectare in 1841 from basic GWR model

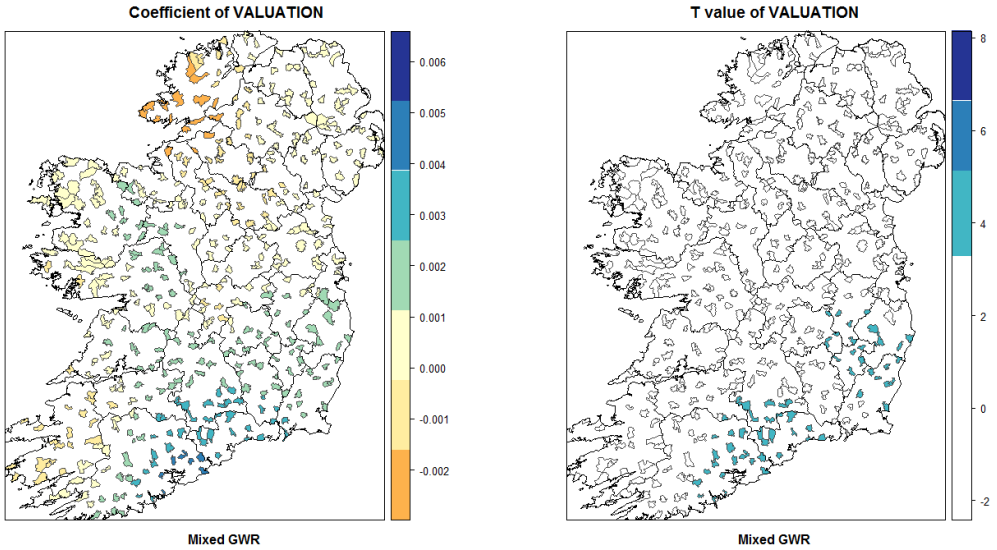


Figure 6-9 Maps of coefficient estimates and T values of land value per hectare in 1841 from mixed GWR model

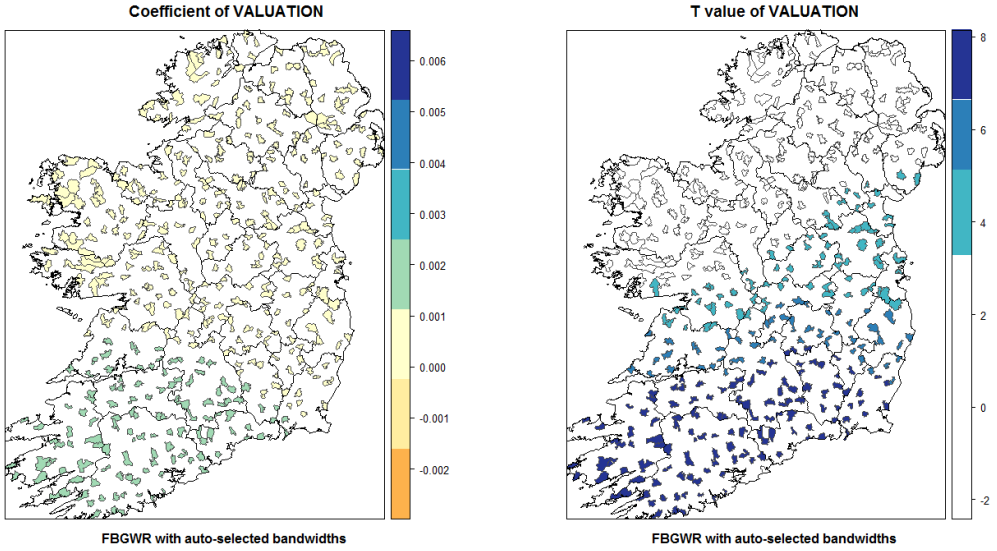


Figure 6-10 Maps of coefficient estimates and T values of land value per hectare in 1841 from FBGWR-1

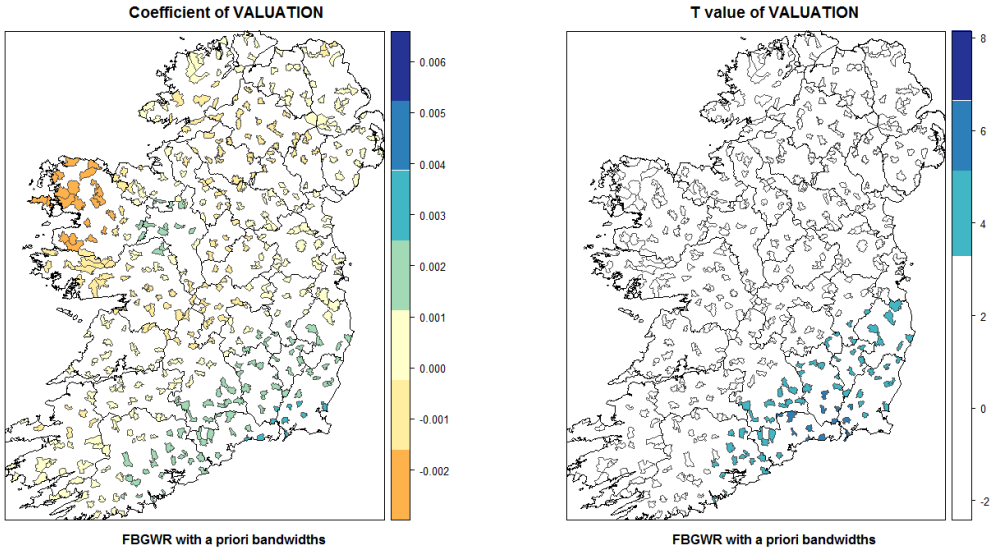


Figure 6-11 Maps of coefficient estimates and T values of land value per hectare in 1841 from FBGWR-2

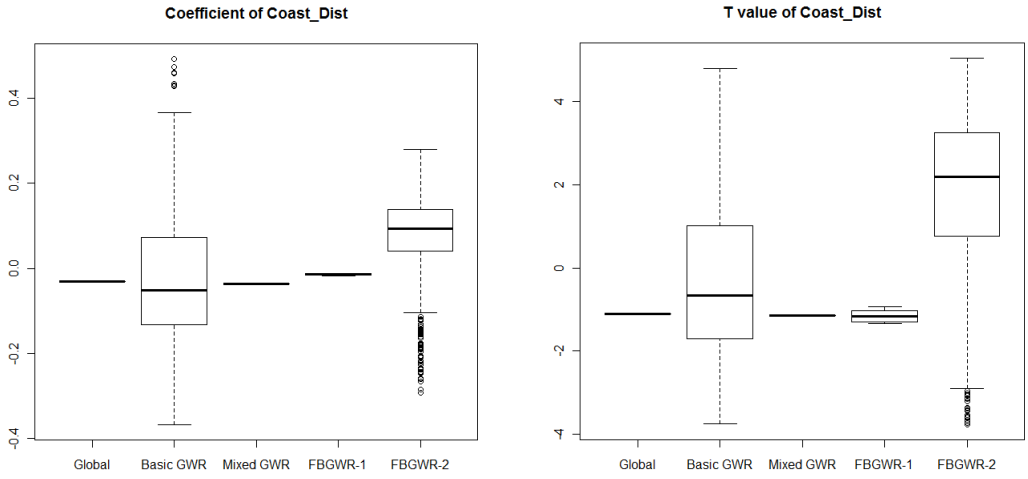


Figure 6-12 Boxplots for coefficient estimates and T values of distance to coast from various models

(ii) distance to coast

As can be seen from Figure 6-12, the effects of "Coast_Dist" (distance to coast) are largely insignificant across all five models with the exception of a few outlying values from the FBGWR-2 model. This suggests that whether an area is close to the coast or not did not affect the depopulation of this area during the famine much. As very few local parameters are significant, there is little point in mapping their spatial distributions.

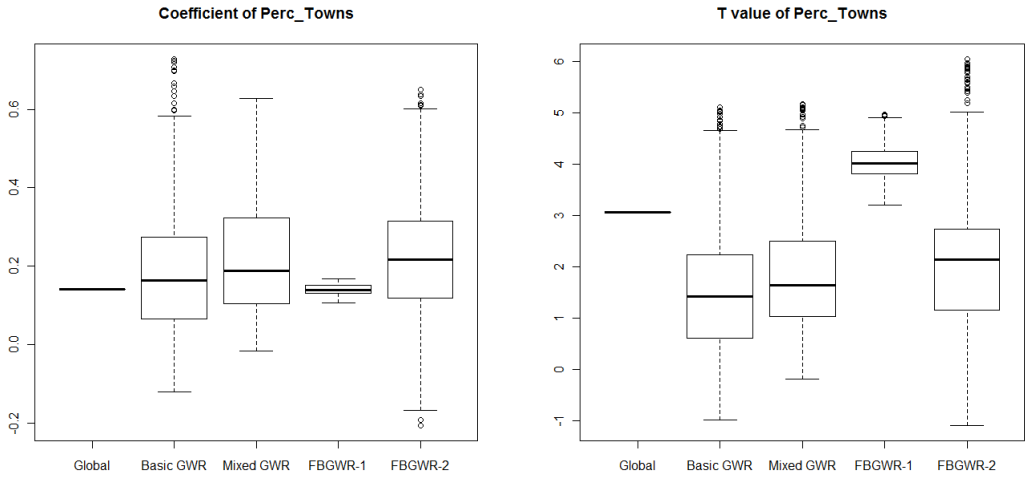


Figure 6-13 Boxplots for coefficient estimates and T values of percentage of population living in towns in 1841 from various models

(iii) percentage of population living in towns in 1841

Figure 6-13 indicates that for the variable "Perc_Town" (percentage of population living in towns in 1841) local parameter estimates from the basic GWR model, the mixed GWR model and FBGWR-2 are largely not significant, while the global model and FBGWR-1 show significant positive relationships between this variable and population decline, suggesting that areas with lower percentages of population living in towns in 1841 were more subject to population decline. This result confirms that the depopulation during the famine was mainly a rural phenomenon.

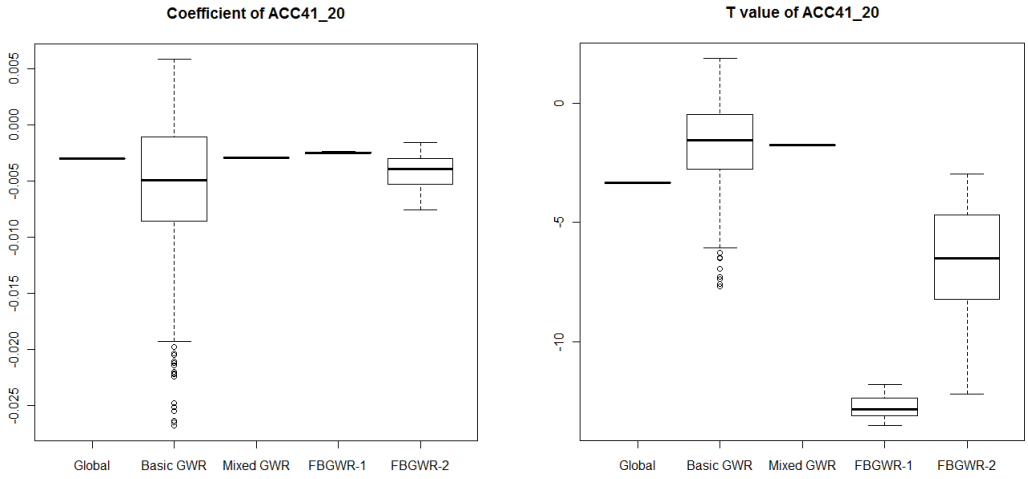


Figure 6-14 Boxplots for coefficient estimates and T values of accessibility to urban areas from various models

(iv) accessibility to urban areas

The results in Figure 6-14 suggest a negative relationship between "ACC41_20" (accessibility to urban areas) and population decline: that is, greater decline took place in areas closer to urban areas. This can be explained by the process of people's migrating from rural to urban areas seeking for relief. The maps of estimated coefficients as well as the associated T values from the basic GWR model and the two FBGWR models are shown in Figure 6-15, 6-16 and 6-17. While the basic GWR model only locates the significant effects in the south-west and north-west of Ireland, the two FBGWR models find this effect significant all over the island, suggesting that rural-to-urban migration was a common phenomenon during the famine decade. This also explains the population growth in those large towns. With a large bandwidth, FBGWR-1 masks the spatial variation in this effect, while FBGWR-2 reveals that this phenomenon was more significant in the north than in the south.

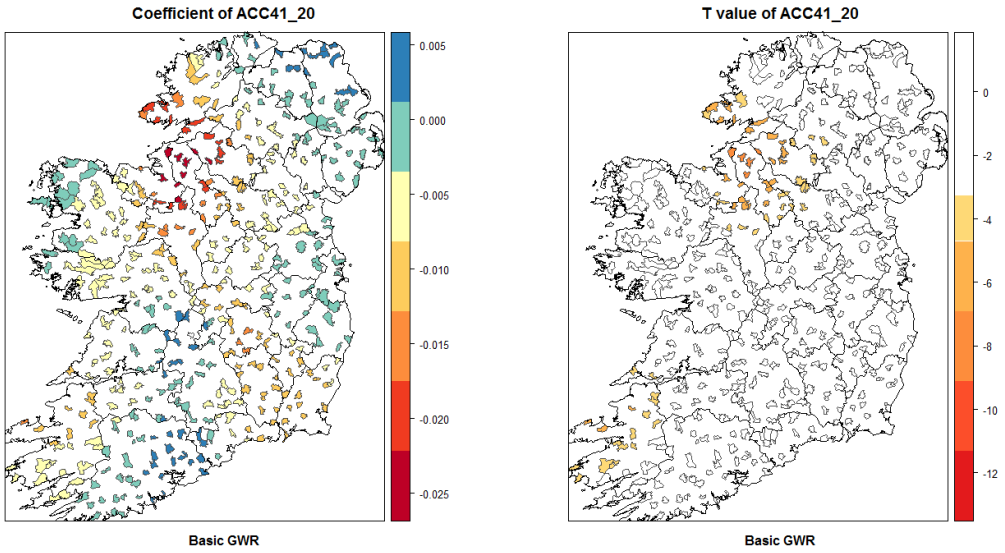


Figure 6-15 Maps of coefficient estimates and T values of accessibility to urban areas from basic GWR model

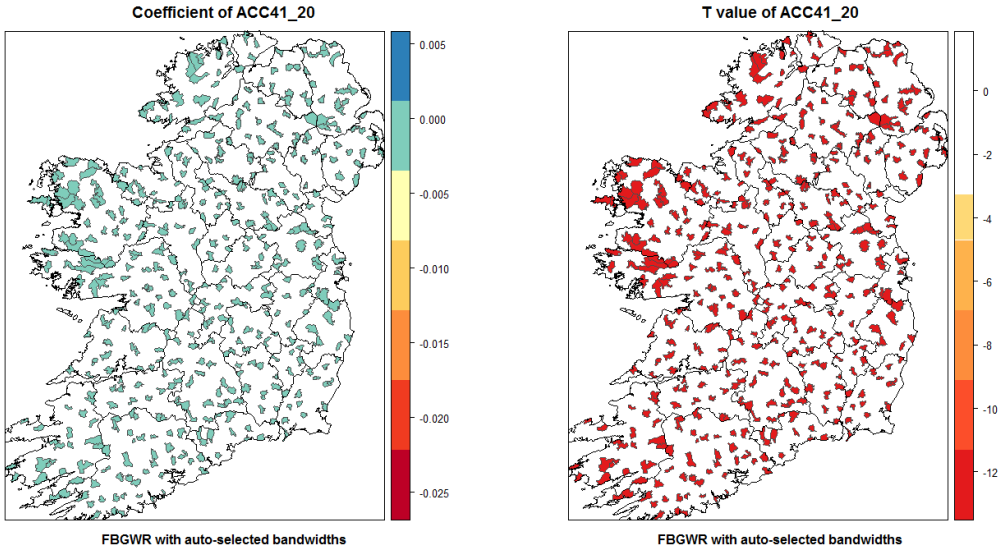


Figure 6-16 Maps of coefficient estimates and T values of accessibility to urban areas from FBGWR-1

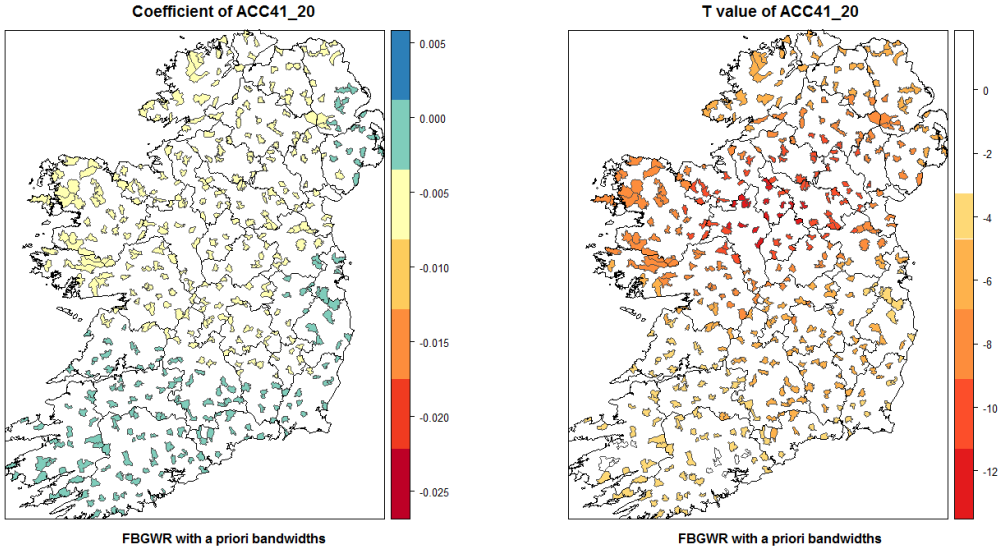


Figure 6-17 Maps of coefficient estimates and T values of accessibility to urban areas from FBGWR-2

(v) proximity to workhouses

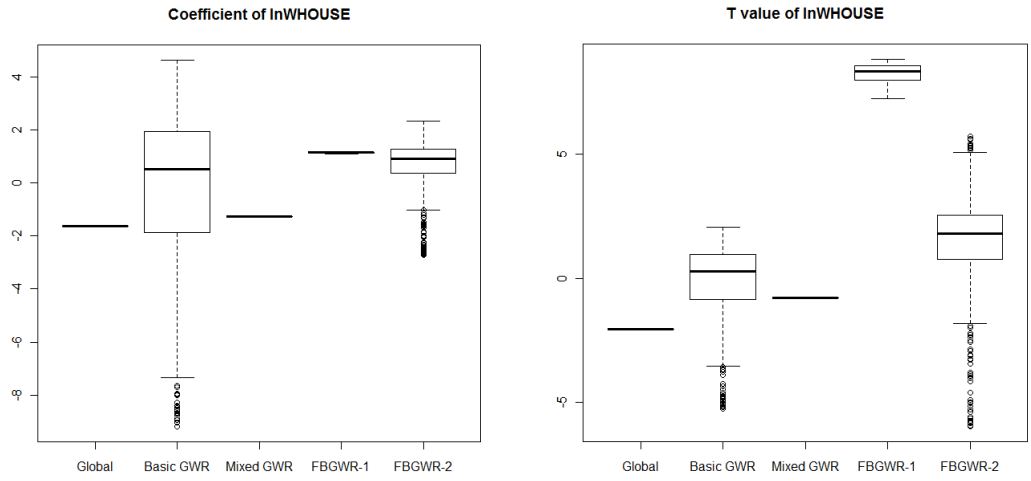


Figure 6-18 Boxplots for coefficient estimates and T values of proximity to workhouses from various models

The effects of "lnWHOUSE" (proximity to workhouses) on population decline show considerable variability across the five models as shown in Figure 6-18. Results from the global model and the mixed GWR model suggest an insignificant negative relationship and most of the local parameter estimates from basic GWR as well as from FBGWR-2 are also insignificant although some are significantly negative. On the contrary, the FBGWR model with auto-selected bandwidths shows significant positive relationships across all locations.

To examine this further, the coefficients estimated from the three models and the associated T values are mapped in Figures 6-19, 6-20 and 6-21. Again, only T values suggesting significant parameter estimates are displayed. Both basic GWR and the FBGWR with *a priori* bandwidths based on basic GWR indicate significant negative relationships in the south west coast of Ireland which means population decline was more severe in an area closer to workhouses, the latter in addition suggests positive relationships in the west of Ireland, suggesting that population decline was more severe in the areas with fewer workhouses around. For the FBGWR model with auto-selected bandwidths, the whole map exhibits a significant positive relationship. Given the fact that relief including that through workhouses was not evenly developed (O'Grada, 1992), the contrary results from different models will need to be identified and validated through more detailed local studies. It is also possible that the counterintuitive result from the FBGWR model with large bandwidth for this variable

might be due to a misspecified bandwidth although more attention needs to be paid to this result.

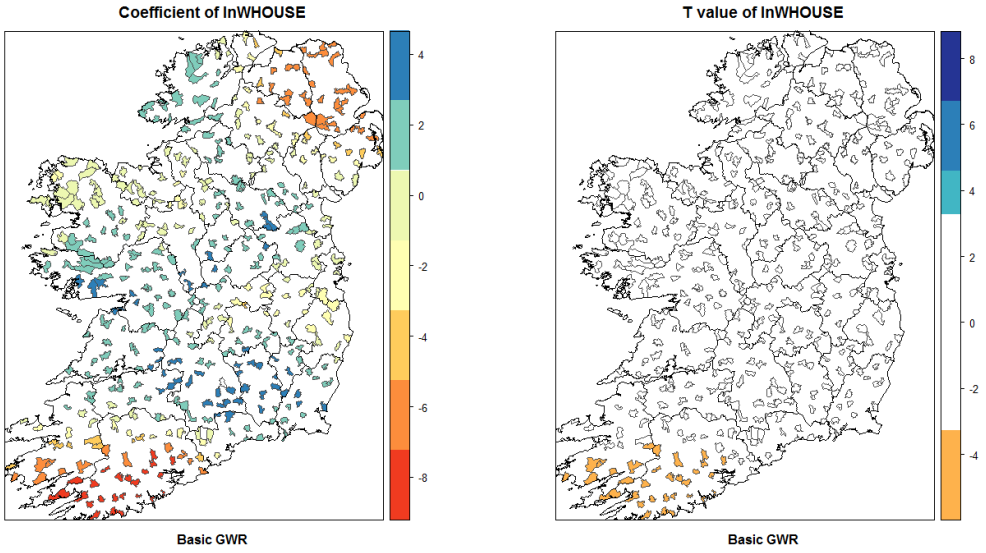


Figure 6-19 Maps of coefficient estimates and T values of proximity to workhouses from basic GWR

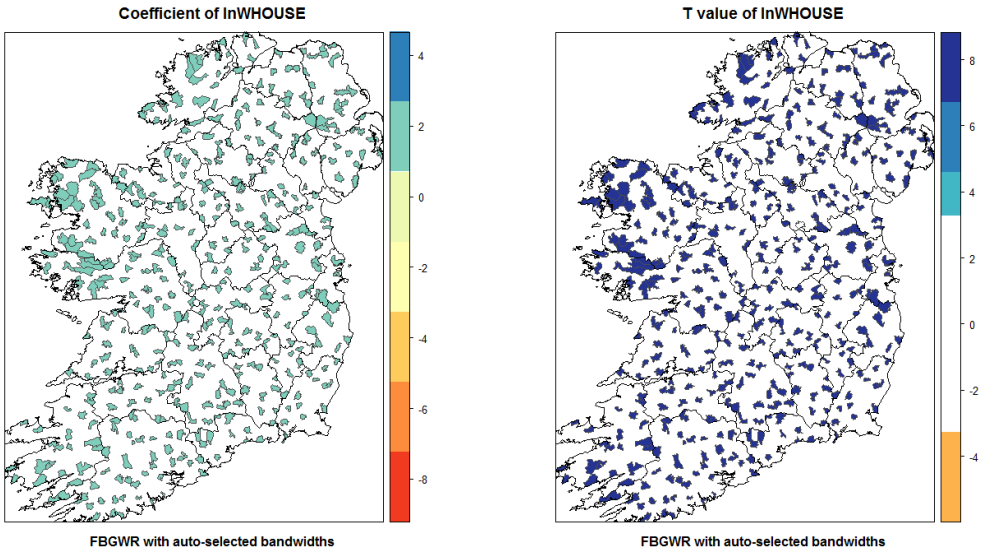


Figure 6-20 Maps of coefficient estimates and T values of proximity to workhouses from FBGWR-1

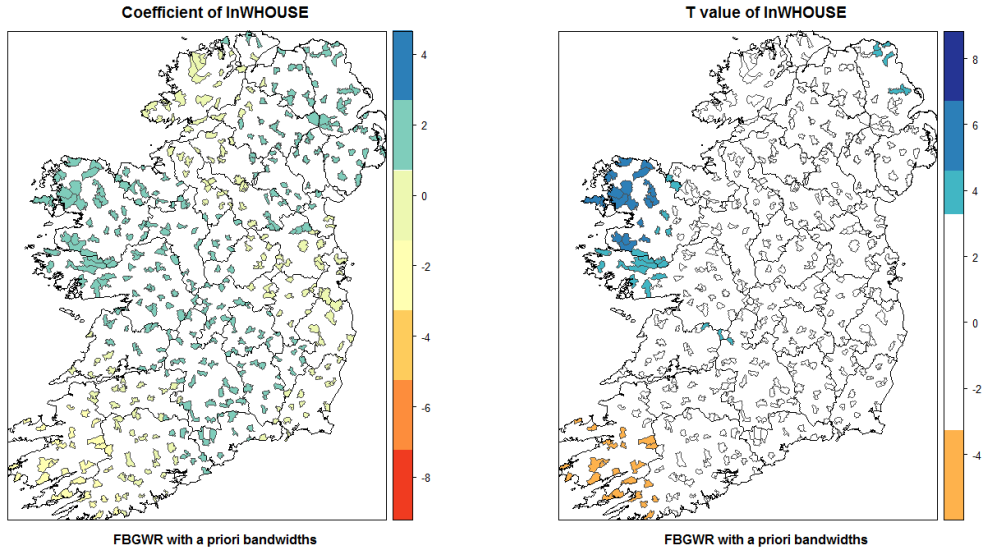


Figure 6-21 Maps of coefficient estimates and T values of proximity to workhouses from FBGWR-2

(vi) population in 1841 per acre of cropped land

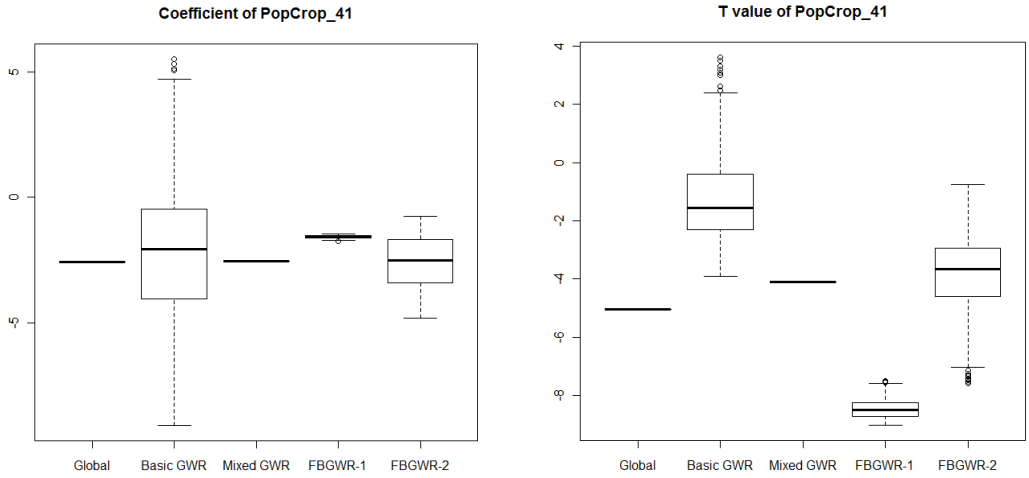


Figure 6-22 Boxplots for coefficient estimates and T values of population in 1841 per acre of cropped land from various models

The results in Figure 6-22 suggest that the factor "PopCrop_41" (population in 1841 per acre of cropped land) has a negative relationship with population change, suggesting that areas with more population per acre of cropped land suffered greater population loss during the famine. These relationships are all significant except those from the basic GWR model. This is coherent with the argument that overpopulation on cropped land in Ireland before the famine was one of the main contributing causes of the disaster (Connell, 1975). While the basic GWR model seems to have obscured

the significant relationship, mixed GWR and FBGWR models help to reveal the relationship by allowing various bandwidths in the models.

(vii) percentage of crop land under potatoes in 1851

From Figure 6-23, the parameter estimates for the variable "Potat_Cult" (percentage of crop land under potatoes in 1851) are generally positive in all models: that is, the more cropped land under potatoes an area had in 1851, the less population loss it had experienced during the famine. Suspiciously extreme significance of this relationship is evident from the FBGWR model with auto-selected bandwidths. This result seems to violate the general consensus that the disaster was caused by potato blight. However, remember that this variable represents the situation after the famine and it might be the case that the strong potato cultivation in 1851 was the result of less population loss during the famine. Also the reduction in agriculture labour force in those areas with severe population loss has led to a shift to pasture after the famine.

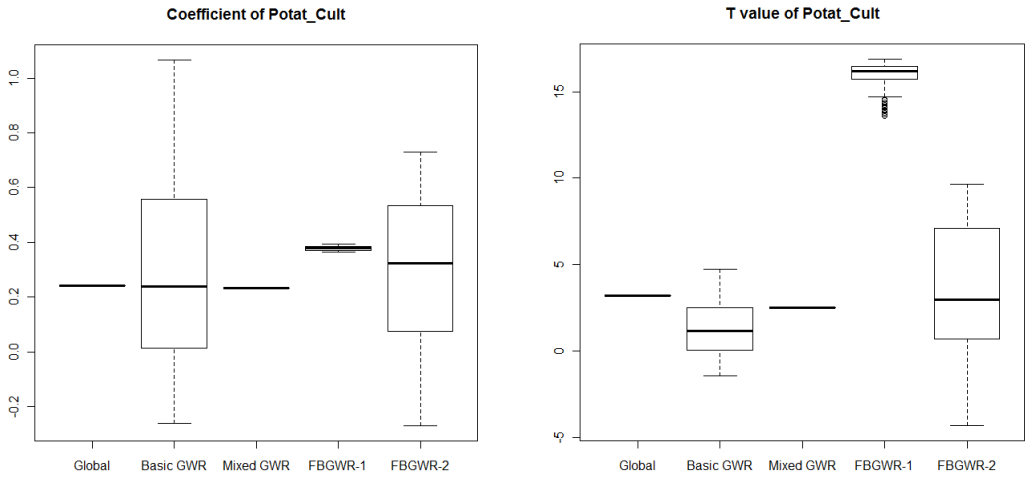


Figure 6-23 Boxplots for coefficient estimates and T values of percentage of crop land under potatoes in 1851 from various models

6.6 Summary

This chapter demonstrates an application of the FBGWR methodology on a subset of the Irish Famine data. Global linear regression was first constructed on the dataset to present a general feeling for the relationships between the selected 9 independent variables and the dependent variable of percentage population decline. An approximate stepwise AIC procedure was then executed to further select a subset of 7 independent variables for GWR model building. Using these 7 variables, basic GWR,

mixed GWR and FBGWR models as well as a global model were calibrated. Two bandwidth selecting strategies were demonstrated to construct the FBGWR model; one employed the automatic synchronous bandwidth optimization method as demonstrated in Chapter 5, the other pre-specified flexible bandwidths to the model based on the bandwidth optimization results from basic GWR. Both methods yielded quite distinguishable bandwidths for each independent variable. However, while the first method suggested approximate global bandwidths for two of the independent variables, the second method results in local bandwidths for all seven variables.

The global model and the mixed GWR model showed the poorest overall performance in fitting the data. The basic GWR model and the two FBGWR models revealed more non-stationarity in local parameter estimates, therefore serving as a guide to further local studies. The two FBGWR models, employing various combinations of bandwidths, resulted in different sets of parameter estimates. Most parameter estimates show consistency with the global results, indicating some confidence in the technique is merited. Basic GWR, employing an unchangeable local bandwidth to all processes, tends to introduce more unreliability in the estimates. For example, see the estimates for variable accessibility to urban areas. FBGWR offers the possibility of specifying bandwidths of various scales according to the various nature of each relationship and therefore discloses more fresh findings that could not be accomplished with the global model and other GWR models. For the effect of land value on population decline, the medium local bandwidth in FBGWR-1 allows more areas with significant effects to be revealed. For the variable accessibility to urban areas, both FBGWR models find more areas with significant effects, while the FBGWR with rather local bandwidth further discovers the spatial variation in these effects. For the variable percentage of population living in towns, the FBGWR with a relatively global bandwidth reveals its effect which is consistent with the global regression but has been obscured by the other GWR models. For the population density variable, FBGWR models also find the significant effect which is not found by the basic GWR model. These findings are important in that they reveal the regional dimension of the effects of the famine, some can be explained by existing literature or support the hypotheses in them while others remain to be explained through further local investigations, for example the contrary results from various models regarding the effect of proximity to workhouses.

However, as different bandwidth can lead to different model results, the estimates for the variable of proximity to workhouses as an example, the accurate specification of proper bandwidths for FBGWR remains a challenge; a misspecified bandwidth can result in misleading estimates. Although it is always helpful to reveal interesting abnormities, caution is required in their interpretation. The local collinearity diagnostic also requires more consideration in the interpretation; further actions to deal with collinearity are beyond the scope of this thesis.

Chapter 7 Conclusions

7.1 Research achievements

The main achievement of this thesis is the construction of an extended GWR model, named Flexible Bandwidth GWR (FBGWR), in which bandwidths can be specified independently for different independent variables and for the intercept in a GWR model. While basic GWR addresses spatial non-stationarity through spatially varying coefficients, FBGWR further investigates the inequality in the spatial scale of variability in various coefficients, incorporating multiple scales of analysis into one model. With FBGWR, not only can the variations in coefficients over space be captured and model fit improved, but also it yields important information on the geographical scale at which various processes operate.

There are two main challenges in this work:

1. Model Calibration
2. Bandwidth selection

The model calibration is achieved through a back-fitting-style procedure where each term in a FBGWR model is calibrated separately in turn as a basic univariate GWR model and the calibration procedure iterates and updates the fitted value for each term until the updates are small enough that the procedure is deemed to have converged. The algorithm was developed in R and tested on both simulated datasets and an empirical dataset.

Bandwidth specification is the core of FBGWR. While *a priori* knowledge and understanding of spatial processes should always be the guide in model building, an automatic bandwidth optimization strategy is developed to assist the bandwidth specification procedure. The strategy is based on AICc model selection criterion and embedded in the back-fitting model calibration procedure. Experiments on simulated datasets with two independent variables whose associated coefficients vary at different combinations of spatial scales indicate that this strategy is effective and competent in finding the appropriate bandwidths. This automatic bandwidth optimization strategy is helpful when there is not enough *a priori* knowledge about the spatial processes being studied. It helps to reveal the nature of spatial non-stationarity in relationships between the dependent variable and the independent

variables which, if followed up with empirical study, may lead to new research findings. The case study of the Irish Famine demonstrates preliminary findings about the various scales at which multiple processes affected population decline during the Irish Famine and highlights new research questions. While some of these questions can be answered by existing literatures on the Irish Famine, others may encourage further local investigations.

7.2 Applicability of FBGWR

FBGWR is useful in several ways. Firstly, it allows a better model fit by incorporating multiple bandwidths that reflect various scales of spatial non-stationarity in relationships into one model. Secondly, it enables a full exploratory investigation of processes that may vary at different spatial scales, providing information about spatial scales at which different processes operate. FBGWR with auto-selected bandwidths can be specifically valuable when there is little knowledge about the dataset to be studied, it helps to understand the underlying spatial processes and guide further investigation. Thirdly, it can assist in model selection for more parsimonious models including the global multivariate linear regression model, basic GWR and mixed GWR models. In theory, FBGWR generalizes these simpler GWR models as well as the global linear regression model. While FBGWR is more complex than these simpler models, cautions should be taken in application to avoid adding too much complexity. Coefficients that have extremely large bandwidths automatically selected from FBGWR would suggest these coefficients to be globally constant, while coefficients that have small bandwidths should be treated locally. If all coefficients in a FBGWR model are suggested to be global, then a global regression model may be preferred. Likewise, FBGWR can be a diagnostic tool for the simpler GWR models, helping to substantiate the validity of the model fit by examining the optimal bandwidths for each coefficient. Moreover, FBGWR can adjust bandwidth for individual variables in a model to circumvent adverse conditions, for example, if problematic local collinearity is found for a variable when a certain bandwidth is employed, FBGWR can be applied to raise the bandwidth for this particular variable to a safe level where local collinearity is no longer a problem.

Potential users of FBGWR include researchers who are analysing spatial data and are interested in the multiple scales of spatial processes. Applications can be made in

human geography and other social science disciplines where geo-referenced data are involved.

7.3 Future directions

7.3.1 Limitations

Being an initial attempt in the development of FBGWR, the work presented in this thesis has several limitations. The first is the lack of model diagnostics and significance testing methods. In the current FBGWR implementation, it is difficult to define the equivalent number of independent parameters or the degree of freedom for the whole model. Therefore, model comparison criteria such as AIC, AICc or adjusted R squared cannot be obtained. The sum of squared residual (RSS) is mainly used to evaluate the performance of FBGWR and to compare it with other models. While the added flexibility of regression coefficients due to flexible bandwidths brings reduced RSS, further tests are required to see whether the reduction is statistically significant.

The second limitation is the computational complexity of the current model calibration method. As an iterative approach, the back-fitting algorithm adopted in this thesis is computationally intensive and time-consuming, especially when the convergence of the algorithm is not easily reached. As shown in Chapter 5, a raised termination threshold can cut down the time cost, but will also reduce the model fit. With more independent variables and/or more regression points involved in the model, this issue will be more severe. Advanced computational techniques such as parallel computing may help to alleviate the problem. Alternatively, other calibration techniques for FBGWR can be developed.

The third limitation is also associated to the back-fitting calibration method. The mechanism of the algorithm requires the observed values of both independent variables and dependent variable to be known. This makes the algorithm capable of model fitting but incapable of prediction. This issue limits the application of the current FBGWR implementation.

7.3.2 Future work

Future work should be committed to overcoming the limitations described above. The most important task is to devise better solutions for model diagnostics and significance tests as well as model inference for FBGWR. The second target is to

work out alternative model calibration methodology which can improve efficiency and enable prediction. To speed up the current back-fitting algorithm and to investigate the conditions for convergence would also be desirable.

There are several additional topics that can be carried out under the current FBGWR framework. More simulation experiments can be designed to investigate further aspects of FBGWR. In terms of spatial systems, the current regular grid data can be extended to true geographical referenced data for a better simulation of real world situations. As to the data generation process, correlation can be introduced between independent variables to check the performance of FBGWR under various collinearity scenarios.

Another possibility is to extend the idea of flexible bandwidths further to that of flexible kernels. That is, allow various types of kernel functions to be specified for each coefficient in a GWR model according to the nature of the spatial process being examined. This is achievable under the current modelling framework which treats each regression term separately.

Further, models beyond the current Gaussian type such as Poisson and logistic models can be tried and a generalized FBGWR may be developed.

7.4 Final remarks

This thesis makes a contribution to the literature in that it is the first work to extend the widely applied local spatial regression technique of GWR by defining flexible predictor-specific bandwidths in a GWR model and developing a practical methodology to optimize the bandwidths. This new extension is important not only in the development of GWR, but also in providing a new exploratory spatial data analysis tool for geo-referenced data in human geography and other social sciences. As theories in social sciences, especially human geography, remain incomplete and open (Poon, 2004), it is difficult to fully explain the relationship between one variable and another. The ability of FBGWR to identify various scales of spatial heterogeneities in regression coefficients helps to reveal locally varying data generating mechanisms and to add more knowledge to locally varying theories.

Bibliography

AITKIN, M. & LONGFORD, N. 1986. Statistical modelling issues in school effectiveness studies. *Journal of the Royal Statistical Society, Series A,* 149**,** 1-43.

AKAIKE, H. Information theory and an extension of the maximum likelihood principle. Second international symposium on information theory, 1973. Akademinai Kiado, 267-281.

ALDSTADT, J. & GETIS, A. 2006. Using AMOEBA to create a spatial weights matrix and identify spatial clusters. *Geographical Analysis,* 38**,** 327-343.

ANSELIN, L. 1988. Spatial Econometrics: Methods and Models, Springer.

ANSELIN, L. 1995. Local Indicators of Spatial Association - LISA. *Geographical Analysis,* 27**,** 93-115.

ANSELIN, L. 2002. Under the hood: issues in the specification and interpretation of spatial regression models. *Agricultural Economics,* 27**,** 247-267.

ANSELIN, L. 2003. Spatial externalities, spatial multipliers, and spatial econometrics. *International Regional Science Review,* 26**,** 153-166.

ANSELIN, L. 2010. Thirty years of spatial econometrics. *Papers in Regional Science,* 89**,** 3-25.

BAILEY, T. C. & GATRELL, A. C. 1995. *Interactive spatial data analysis,* Longman Scientific & Technical Essex.

BANERJEE, S., GELFAND, A. E. & CARLIN, B. P. 2004. *Hierarchical Modeling and Analysis for Spatial Data,* Boca Raton, FL, Chapman and Hall/CRC.

BAVAUD, F. 1998. Models for spatial weights: A systematic look. *Geographical Analysis,* 30**,** 153-171.

BERRY, B. J. L., GRIFFITH, D. A. & TIEFELSDORF, M. R. 2008. From Spatial analysis to geospatial science. *Geographical Analysis,* 40**,** 229-238.

BERRY, B. J. L. & MARBLE, D. F. 1968. *Spatial analysis: a reader in statistical geography,* Prentice-Hall.

BITTER, C., MULLIGAN, G. F. & DALL'ERBA, S. 2007. Incorporating spatial variation in housing attribute prices: a comparison of geographically weighted regression and the spatial expansion method. *Journal of Geographical Systems,* 9**,** 7-27.

BIVAND, R. & YU, D. 2010. spgwr: Geographically weighted regression. R package version 0.6-9 ed.

BIVAND, R. S., PEBESMA, E. J. & GOMEZ-RUBIO, V. 2013. *Applied Spatial Data Analysis with R,* New York, Springer.

BLANCO‐MORENO, J., CHAMORRO, L., IZQUIERDO, J., MASALLES, R. & SANS, F. 2008. Modelling within‐field spatial variability of crop biomass–weed density relationships using geographically weighted regression. *Weed Research,* 48**,** 512-522.

BORCARD, D. & LEGENDRE, P. 2002. All-scale spatial analysis of ecological data by means of principal coordinates of neighbour matrices. *Ecological Modelling,* 153**,** 51-68.

BOURKE, A. 1993. The Visitation of God?: The Potato and the Great Irish Famine, Lilliput Press, Limited.

BOURKE, P. M. A. 1965. The Agricultural Statistics of the 1841 Census of Ireland: a Critical Review. *Economic History Review,* 18**,** 376-391.

BOWMAN, A. W. 1984. An Alternative Method of Cross-Validation for the Smoothing of Density Estimates. *Biometrika,* 71**,** 353-360.

BRANDSMA, A. S. & KETELLAPPER, R. H. 1979. Biparametric Approach to Spatial Autocorrelation. *Environment and Planning A,* 11**,** 51-58.

BREIMAN, L. & FRIEDMAN, J. H. 1985. Estimating Optimal Transformations for Multiple-Regression and Correlation. *Journal of the American Statistical Association,* 80**,** 580-598.

BROWNE, W. J., GOLDSTEIN, H. & RASBASH, J. 2001. Multiple membership multiple classification (MMMC) models. *Statistical Modelling,* 1**,** 103-124.

BRUNSDON, C., AITKIN, M., FOTHERINGHAM, S. & CHARLTON, M. 1999a. A Comparsion of Random Coefficient Modelling and Geographically Weighted Regression for Spatially Non-stationary Regression Problems. *Geographical and Environmental Modelling,* 3**,** 47-62.

BRUNSDON, C., CHARLTON, M. & HARRIS, P. 2012. Living with collinearity in Local Regression Models. *Spatial Accuracy.* Brazil.

BRUNSDON, C., FOTHERINGHAM, A. S. & CHARLTON, M. 1998a. Spatial nonstationarity and autoregressive models. *Environment and Planning A,* 30**,** 957-973.

BRUNSDON, C., FOTHERINGHAM, A. S. & CHARLTON, M. 1999b. Some notes on parametric significance tests for geographically weighted regression. *Journal of Regional Science,* 39**,** 497-524.

BRUNSDON, C., FOTHERINGHAM, A. S. & CHARLTON, M. E. 1996. Geographically weighted regression: A method for exploring spatial nonstationarity. *Geographical Analysis,* 28**,** 281-298.

BRUNSDON, C., FOTHERINGHAM, S. & CHARLTON, M. 1998b. Geographically weighted regression - modelling spatial non-stationarity. *Journal of the Royal Statistical Society Series D-the Statistician,* 47**,** 431-443.

BRUNSDON, C., MCCLATCHEY, J. & UNWIN, D. 2001. Spatial variations in the average rainfall–altitude relationship in Great Britain: an approach using geographically weighted regression. *International Journal of Climatology,* 21**,** 455-466.

BUJA, A., HASTIE, T. & TIBSHIRANI, R. 1989. Linear Smoothers and Additive-Models. *Annals of Statistics,* 17**,** 453-510.

BURRIDGE, P. 1981. Testing for a Common Factor in a Spatial Autoregression Model. *Environment and Planning A,* 13**,** 795-800.

BYRNE, G., CHARLTON, M. & FOTHERINGHAM, A. S. 2009. Multiple Dependent Hypothesis Tests in Geographically Weighted Regression. *In:* LEES, B. G. & LAFFAN, S. W. (eds.) *10th International conference on geocomputation.* Sydney

CASETTI, E. 1972. Generating Models by Expansion Method: Applications to Geographical Research. *Geographical Analysis,* 4**,** 81-91.

CASETTI, E. 1997. The expansion method, mathematical modeling, and spatial econometrics. *International Regional Science Review,* 20**,** 9-33.

CHARLTON, M., FOTHERINGHAM, A. S. & BRUNSDON, C. 2007. Geographically Weighted Regression Software: GWR3.0. 3.0 ed.

CHILÈS, J. & DELFINER, P. 1999. Geostatistics: modeling spatial uncertainty. New York, USA: John Wiley & Sons.

CHO, W. K. T. & GIMPEL, J. G. 2010. Rough terrain: spatial variation in campaign contributing and volunteerism. *American Journal of Political Science,* 54**,** 74-89.

CLEVELAND, W. S. 1979. Robust Locally Weighted Regression and Smoothing Scatterplots. *Journal of the American Statistical Association,* 74**,** 829-836.

CLEVELAND, W. S. & DEVLIN, S. J. 1988. Locally weighted regression: An approach to regression analysis by local fitting. *Journal of the American Statistical Association,* 83**,** 596-610.

CLIFF, A. D. & ORD, J. K. 1969. The Problem of Spatial Autocorrelation. *London Papers in Regional Science.* London: Pion.

CLIFF, A. D. & ORD, J. K. 2009. What Were We Thinking? *Geographical Analysis,* 41**,** 351-363.

CONNELL, K. 1975. *The population of Ireland 1700-1845,* Oxford, Oxford University Press.

CRAVEN, P. & WAHBA, G. 1979. Smoothing Noisy Data with Spline Functions - Estimating the Correct Degree of Smoothing by the Method of Generalized Cross-Validation. *Numerische Mathematik,* 31**,** 377-403.

DARMOFAL, D. 2008. The political geography of the new deal realignment. *American Politics Research,* 36**,** 934-961.

DEATON, A. & MUELLBAUER, J. 1980. *Economics and Consumer Behavior*, Cambridge University Press.

DEMŠAR, U., FOTHERINGHAM, A. S. & CHARLTON, M. 2008. Combining Geovisual Analytics with Spatial Statistics: the Example of Geographically Weighted Regression. *The Cartographic journal,* 45**,** 182-192.

FAN, J. Q. & ZHANG, W. Y. 2008. Statistical methods with varying coefficient models. *Statistics and Its Interface,* 1**,** 179-195.

FARBER, S. & PAEZ, A. 2007. A systematic investigation of cross-validation in GWR model estimation: empirical analysis and Monte Carlo simulations. *Journal of Geographical Systems,* 9**,** 371-396.

FARBER, S. & PÁEZ, A. 2007. A systematic investigation of cross-validation in GWR model estimation: empirical analysis and Monte Carlo simulations. *Journal of Geographical Systems,* 9**,** 371-396.

FARBER, S. & YEATES, M. 2006. A comparison of localized regression models in a hedonic house price context. *Canadian Journal of Regional Science,* 29**,** 405-419.

FINGLETON, B. 2003. Externalities, economic geography, and spatial econometrics: Conceptual and modeling developments. *International Regional Science Review,* 26**,** 197-207.

FINLEY, A. O. 2011. Comparing spatially-varying coefficients models for analysis of ecological data with non-stationary and anisotropic residual dependence. *Methods in Ecology and Evolution,* 2**,** 143-154.

FISCHER, M. M. & GETIS, A. 1997. Advances in Spatial Analysis. *In:* FISCHER, M. M. & GETIS, A. (eds.) *Recent Developments in Spatial Analysis.* Springer Berlin Heidelberg.

FOLMER, H. & OUD, J. 2008. How to get rid of W: a latent variables approach to modelling spatially lagged variables. *Environment and Planning A,* 40**,** 2526-2538.

FOODY, G. M. 2003. Geographical weighting as a further refinement to regression modelling: An example focused on the NDVI-rainfall relationship. *Remote Sensing of Environment,* 88**,** 283-293.

FOTHERINGHAM, A. S. 1993. On the Future of Spatial-Analysis - the Role of GIS. *Environment and Planning A***,** 30-34.

FOTHERINGHAM, A. S. 1997. Trends in quantitative methods .1. Stressing the local. *Progress in Human Geography,* 21**,** 88-96.

FOTHERINGHAM, A. S. 2009. "The Problem of Spatial Autocorrelation" and Local Spatial Statistics. *Geographical Analysis,* 41**,** 398-403.

FOTHERINGHAM, A. S. & BRUNSDON, C. 1999. Local forms of spatial analysis. *Geographical Analysis,* 31**,** 340-358.

FOTHERINGHAM, A. S., BRUNSDON, C. & CHARLTON, M. 2002. Geographically Weighted Regression: The Analysis of Spatially Varying Relationships, John Wiley & Sons.

FOTHERINGHAM, A. S., CHARLTON, M. & BRUNSDON, C. 1996. The geography of parameter space: an investigation of spatial non-stationarity. *International Journal of Geographical Information Systems,* 10**,** 605-627.

FOTHERINGHAM, A. S., CHARLTON, M. E. & BRUNSDON, C. 1998. Geographically weighted regression: a natural evolution of the expansion method for spatial data analysis. *Environment and Planning A,* 30**,** 1905-1927.

FOTHERINGHAM, A. S., KELLY, M. & CHARLTON, M. 2010. Model Selection in Geographically Weighted Regression *Joint International Conference on Theory, Data Handling and Modelling in GeoSpatial Information Science.* Hong Kong.

FOTHERINGHAM, A. S., KELLY, M. H. & CHARLTON, M. 2013. The demographic impacts of the Irish famine: towards a greater geographical understanding. *Transactions of the Institute of British Geographers,* 38**,** 221-237.

FOTHERINGHAM, A. S. & WONG, D. W. S. 1991. The Modifiable Areal Unit Problem in Multivariate Statistical Analysis. *Environment and Planning A,* 23**,** 1025-1044.

FRIEDMAN, J. H. & STUETZLE, W. 1981. Projection Pursuit Regression. *Journal of the American Statistical Association,* 76**,** 817-823.

GAO, X. L., ASAMI, Y. & CHUNG, C. J. F. 2006. An empirical evaluation of spatial regression models. *Computers & Geosciences,* 32**,** 1040-1051.

GELFAND, A. E., KIM, H. J., SIRMANS, C. F. & BANERJEE, S. 2003. Spatial modeling with spatially varying coefficient processes. *Journal of the American Statistical Association,* 98**,** 387-396.

GELMAN, A., CARLIN, J. B., STERN, H. S. & RUBIN, D. B. 2004. *Bayesian data analysis,* Chapman & Hall/CRC.

GETIS, A. 1990. Screening for Spatial Dependence in Regression Analysis. *Papers of the Regional Science Association,* 69**,** 69-81.

GETIS, A. 1995. Spatial Filtering in a Regression Framework: Examples Using Data on Urban Crime, Regional Inequality, and Government Expenditures. *In:* ANSELIN, L. & FLORAX, R. J. G. M. (eds.) *New Directions in Spatial Econometrics.* Springer Berlin Heidelberg.

GETIS, A. & ALDSTADT, J. 2004. Constructing the spatial weights matrix using a local statistic. *Geographical Analysis,* 36**,** 90-104.

GETIS, A. & GRIFFITH, D. A. 2002. Comparative spatial filtering in regression analysis. *Geographical Analysis,* 34**,** 130-140.

GETIS, A. & ORD, J. K. 1992. The Analysis of Spatial Association by Use of Distance Statistics. *Geographical Analysis,* 24**,** 189-206.

GOLDSTEIN, H. 1986. Multilevel Mixed Linear Model Analysis Using Iterative Generalized Least Squares. *Biometrika,* 73**,** 43-56.

GOLDSTEIN, H. 1987. Multilevel Models in Educational and Social Research, Charles Griffin & Company.

GOLDSTEIN, H. 1991. Multilevel Modeling of Survey Data. *Statistician,* 40**,** 235-244.

GOLDSTEIN, H. 2010. Multilevel Statistical Models, Wiley.

GOODCHILD, M. F., ANSELIN, L., APPELBAUM, R. P. & HARTHORN, B. H. 2000. Toward spatially integrated social science. *International Regional Science Review,* 23**,** 139-159.

GOODCHILD, M. F. & LONGLEY, P. A. 1999. The future of GIS and spatial analysis. *Geographical information systems: principles, techniques, management and applications***,** 567-580.

GRANGER, C. W. 1969. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica: Journal of the Econometric Society***,** 424-438.

GREGORY, I. N. & ELL, P. S. 2005. Analyzing spatiotemporal change by use of national historical geographical information systems: population change during and after the great Irish famine. *Historical Methods: A Journal of Quantitative and Interdisciplinary History,* 38**,** 149-167.

GREIG, D. M. 1980. *Optimisation,* London, Longman.

GRIFFITH, D. A. 1979. Urban Dominance, Spatial Structure, and Spatial Dynamics - Some Theoretical Conjectures and Empirical Implications. *Economic Geography,* 55**,** 95-113.

GRIFFITH, D. A. 1996. Spatial autocorrelation and eigenfunctions of the geographic weights matrix accompanying geo-referenced data. *Canadian Geographer,* 40**,** 351-367.

GRIFFITH, D. A. 2009. Celebrating 40 Years of Scientific Impacts by Cliff and Ord. *Geographical Analysis,* 41**,** 343-345.

GRIFFITH, D. A. 2010. Spatial Filtering. *In:* FISCHER, M. M. & GETIS, A. (eds.) *Handbook of Applied Spatial Analysis.* Springer Berlin Heidelberg.

GUINNANE, T. W. & O'GRADA, C. 2000. *The Workhouses and Irish Famine Mortality*, Department of Economics, University College Dublin.

GUISAN, A., EDWARDS, T. C. & HASTIE, T. 2002. Generalized linear and generalized additive models in studies of species distributions: setting the scene. *Ecological Modelling,* 157**,** 89-100.

GUO, L., MA, Z. & ZHANG, L. 2008. Comparison of bandwidth selection in application of geographically weighted regression: a case study. *Canadian Journal of Forest Research,* 38**,** 2526-2534.

HAINING, R. 1991. Bivariate Correlation with Spatial Data. *Geographical Analysis,* 23**,** 210-227.

HAINING, R. P. 2003. *Spatial Data Analysis: Theory and Practice,* Cambridge, UK, Cambridge University Press.

HAINING, R. P. 2009. Spatial Autocorrelation and the Quantitative Revolution. *Geographical Analysis,* 41**,** 364-374.

HARRIS, P., BRUNSDON, C. & FOTHERINGHAM, A. S. 2011. Links, comparisons and extensions of the geographically weighted regression model when used as a spatial predictor. *Stochastic Environmental Research and Risk Assessment,* 25**,** 123-138.

HARRIS, P., FOTHERINGHAM, A. S., CRESPO, R. & CHARLTON, M. 2010a. The Use of Geographically Weighted Regression for Spatial Prediction: An Evaluation of Models Using Simulated Data Sets. *Mathematical Geosciences,* 42**,** 657-680.

HARRIS, P., FOTHERINGHAM, A. S., CRESPO, R. & CHARLTON, M. 2010b. The Use of Geographically Weighted Regression for Spatial Prediction: An Evaluation of Models Using Simulated Data Sets (vol 42, pg 657, 2010). *Mathematical Geosciences,* 43**,** 399-399.

HARRIS, P., FOTHERINGHAM, A. S. & JUGGINS, S. 2010c. Robust geographically weighted regression: a technique for quantifying spatial relationships between freshwater acidification critical loads and catchment attributes. *Annals of the Association of American Geographers,* 100**,** 286-306.

HART, J. F. 1954. Central tendency in areal distributions. *Economic Geography,* 30**,** 48-59.

HASTIE, T. 2011. gam: Generalized Additive Models. R package version 1.06.2 ed.

HASTIE, T. & TIBSHIRANI, R. 1986. Generalized Additive Models. *Statistical Science,* 1**,** 297-310.

HASTIE, T. & TIBSHIRANI, R. 1987. Generalized Additive-Models - Some Applications. *Journal of the American Statistical Association,* 82**,** 371-386.

HASTIE, T. & TIBSHIRANI, R. 1990. *Generalized Additive Models,* London, Chapman and Hall.

HASTIE, T. & TIBSHIRANI, R. 1993. Varying-Coefficient Models. *Journal of the Royal Statistical Society Series B-Methodological,* 55**,** 757-796.

HENDERSON, R. 1924. A New Method of Graduation. *Transactions of the Actuarial Society of America,* 25**,** 29-40.

HEPPLE, L. W. 1995. Bayesian Techniques in Spatial and Network Econometrics .1. Model Comparison and Posterior Odds. *Environment and Planning A,* 27**,** 447-469.

HUANG, Y. & LEUNG, Y. 2002. Analysing regional industrialisation in Jiangsu province using geographically weighted regression. *Journal of Geographical Systems,* 4**,** 233-249.

HUANG, Y. Q. & CLARK, W. A. V. 2002. Housing tenure choice in transitional urban China: A multilevel analysis. *Urban Studies,* 39**,** 7-32.

HURVICH, C. M., SIMONOFF, J. S. & TSAI, C. L. 1998. Smoothing parameter selection in nonparametric regression using an improved Akaike information criterion. *Journal of the Royal Statistical Society Series B-Statistical Methodology,* 60**,** 271-293.

JONES, K. 1991. Specifying and Estimating Multi-Level Models for Geographical Research. *Transactions of the Institute of British Geographers,* 16**,** 148-159.

KELEJIAN, H. H. & PRUCHA, I. R. 1998. Generalized spatial two-stage least squares procedure for estimating a spatial autoregressive model with autoregressive disturbances. *Journal of Real Estate Finance and Economics,* 17**,** 99-121.

KELLY, M. & FOTHERINGHAM, A. S. 2011. The online atlas of Irish population change 1841-2002: A new resource for analysing national trends and local variations in Irish population dynamics. *Irish Geography,* 44**,** 215-244.

KENNY, K. 2003. Diaspora and comparison: The global Irish as a case study. *Journal of American History,* 90**,** 134-162.

KINEALY, C. 1997. A Death-Dealing Famine: The Great Hunger in Ireland, Pluto Press.

KRIGE, D. 1951. A Statistical Approach to Some Mine Valuation and Allied Problems on the Witwatersrand: By DG Krige. University of the Witwatersrand.

KRUGMAN, P. R. 1991. *Geography & Trade,* Mit Press.

KUPFER, J. A. & FARRIS, C. A. 2007. Incorporating spatial non-stationarity of regression coefficients into predictive vegetation models. *Landscape Ecology,* 22**,** 837-852.

LAM, N. S. N. & QUATTROCHI, D. A. 1992. On the Issues of Scale, Resolution, and Fractal Analysis in the Mapping Sciences*. *The Professional Geographer,* 44**,** 88-98.

LLOYD, C. & SHUTTLEWORTH, I. 2005. Analysing commuting using local regression techniques: scale, sensitivity, and geographical patterning. *Environment and Planning A,* 37**,** 81-103.

LLOYD, C. D. 2010a. Exploring population spatial concentrations in Northern Ireland by community background and other characteristics: an application of geographically weighted spatial statistics. *International Journal of Geographical Information Science,* 24**,** 1193-1221.

LLOYD, C. D. 2010b. Nonstationary models for exploring and mapping monthly precipitation in the United Kingdom. *International Journal of Climatology,* 30**,** 390-405.

LLOYD, C. D. 2012. Analysing the spatial scale of population concentrations by religion in Northern Ireland using global and local variograms. *International Journal of Geographical Information Science,* 26**,** 57-73.

LOADER, C. 1999. *Local Regression and Likelihood,* New York, Springer.

LOADER, C. 2012. Smoothing: Local Regression Techniques *In:* GENTLE, J. E., HÄRDLE, W. & MORI, Y. (eds.) *Handbook of Computational Statistics: Concepts and Methods.* Springer.

LONGLEY, P. A. & TOBÓN, C. 2004. Spatial dependence and heterogeneity in patterns of hardship: an intra-urban analysis. *Annals of the Association of American Geographers,* 94**,** 503-519.

LOVETT, A. & FLOWERDEW, R. 1989. Analysis of Count Data Using Poisson Regression*. *The Professional Geographer,* 41**,** 190-198.

LU, B. 2012. Geographically Weighted Regression with Non-Euclidean Distance Metrics. Ph.D. degree, National University of Ireland Maynooth.

LU, B., CHARLTON, M. & FOTHERINGHAMA, A. S. 2011. Geographically Weighted Regression Using a Non-Euclidean Distance Metric with a Study on London House Price Data. *Procedia Environmental Sciences,* 7**,** 92-97.

LU, B., HARRIS, P., CHARLTON, M., BRUNSDON, C., NAKAYA, T. & GOLLINI, I. 2013. GWmodel: Geographically weighted models. R package Version 1.2-2 ed.

MACAULAY, F. R. 1931. *The Smoothing of Time Series,* New York, National Bureau of Economic Research, incorporated.

MARTIN, R. 1999. The new 'geographical turn' in economics: some critical reflections. *Cambridge Journal of Economics,* 23**,** 65-91.

MATHERON, G. 1963. Principles of geostatistics. *Economic geology,* 58**,** 1246-1266.

MCMILLEN, D. & MCDONALD, J. 2004. Locally Weighted Maximum Likelihood Estimation: Monte Carlo Evidence and an Application. *In:* ANSELIN, L., FLORAX, R. G. M. & REY, S. (eds.) *Advances in Spatial Econometrics.* Springer Berlin Heidelberg.

MEI, C. L., HE, S. Y. & FANG, K. T. 2004. A note on the mixed geographically weighted regression model. *Journal of Regional Science,* 44**,** 143-157.

MENNIS, J. 2006. Mapping the results of geographically weighted regression. *Cartographic Journal, The,* 43**,** 171-179.

MENNIS, J. L. & JORDAN, L. 2005. The distribution of environmental equity: Exploring spatial nonstationarity in multivariate models of air toxic releases. *Annals of the Association of American Geographers,* 95**,** 249-268.

MOKYR, J. 1985. Why Ireland starved : a quantitative and analytical history of the Irish economy, 1800-1850, London ; Boston, Allen & Unwin.

MOSTELLER, F. & TUKEY, J. W. 1977. *Data analysis and regression : a second course in statistics,* Reading, Mass., Addison-Wesley Pub. Co.

NAKAYA, T. 2001. Local spatial interaction modelling based on the geographically weighted regression approach. *GeoJournal,* 53**,** 347-358.

NAKAYA, T., CHARLTON, M., FOTHERINGHAM, A. S. & BRUNSDON, C. 2009. Geographically Weighted Regression Software: GWR4.0. 4.0 ed.

NAKAYA, T., FOTHERINGHAM, A., BRUNSDON, C. & CHARLTON, M. 2005a. Geographically weighted Poisson regression for disease association mapping. *Statistics in Medicine,* 24**,** 2695-2717.

NAKAYA, T., FOTHERINGHAM, A. S., BRUNSDON, C. & CHARLTON, M. 2005b. Geographically weighted Poisson regression for disease association mapping. *Stat Med,* 24**,** 2695-2717.

NELDER, J. A. & WEDDERBURN, R. W. M. 1972. Generalized Linear Models. *Journal of the Royal Statistical Society. Series A (General),* 135**,** 370-384.

O'GRADA, C. 1992. Ireland's great famine. ReFRESH (Recent Findings of Research in Economic & Social History), 15**,** 5-8.

O'GRADA, C. & EIRÍKSSON, A. S. 2006. *Ireland's great famine : interdisciplinary perspectives,* Dublin, Ireland, University College Dublin Press.

O'LOUGHLIN, J., WARD, M. D., LOFDAHL, C. L., COHEN, J. S., BROWN, D. S., REILLY, D., GLEDITSCH, K. S. & SHIN, M. 1998. The diffusion of democracy, 1946-1994. *Annals of the Association of American Geographers,* 88**,** 545-574.

ORD, K. 1975. Estimation Methods for Models of Spatial Interaction. *Journal of the American Statistical Association,* 70**,** 120-126.

ORFORD, S. 2000. Modelling spatial structures in local housing market dynamics: A multilevel perspective. *Urban Studies,* 37**,** 1643-1671.

PAELINCK, J. H. P. & KLAASSEN, L. H. 1979. *Spatial econometrics,* Farnborough, Saxon House.

PÁEZ, A. 2004. Anisotropic variance functions in geographically weighted regression models. *Geographical Analysis,* 36**,** 299-314.

PÁEZ, A., FARBER, S. & WHEELER, D. 2011. A simulation-based study of geographically weighted regression as a method for investigating spatially varying relationships. *Environment and Planning A,* 43**,** 2992-3010.

PÁEZ, A., LONG, F. & FARBER, S. 2008. Moving window approaches for hedonic price estimation: An empirical comparison of modelling techniques. *Urban Studies,* 45**,** 1565-1581.

PÁEZ, A., UCHIDA, T. & MIYAMOTO, K. 2002. A general framework for estimation and inference of geographically weighted regression models: 1. Location-specific kernel bandwidths and a test for locational heterogeneity. *Environment and Planning A,* 34**,** 733-754.

PAVEL, P., MARTIN, K. & STEFAN, E. 2008. Application of Geographically Weighted Regression to Investigate the Impact of Scale on Prediction Uncertainty by Modelling Relationship between Vegetation and Climate. European Commission Joint Research Centre.

PEARCE, M. S. 1998. Geographically Weighted Regression: A Method for Exploring Spatial Nonstationarity. *Stata Technical Bulletin,* 46**,** 20-24.

PEBESMA, E. J. & BIVAND, R. S. 2005. Classes and methods for spatial data in R. *R News* 5.

POON, J. P. H. 2004. Quantitative methods: past and present. *Progress in Human Geography,* 28**,** 807-814.

PROPASTIN, P., KAPPAS, M. & ERASMI, S. 2008. Application of geographically weighted regression to investigate the impact of scale on prediction uncertainty by modelling relationship between vegetation and climate. *International Journal of Spatial Data Infrastructures Research,* 3**,** 73-94.

R DEVELOPMENT CORE TEAM 2011. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing.

RANGEL, T. F., DINIZ‐FILHO, J. A. F. & BINI, L. M. 2010. SAM: a comprehensive application for spatial analysis in macroecology. *Ecography,* 33**,** 46-50.

REINSCH, C. H. 1967. Smoothing by Spline Functions. *Numerische Mathematik,* 10**,** 177-183.

RIPLEY, B. D. 1981. *Spatial statistics,* New York, Wiley.

SCHWARZ, G. 1978. Estimating the dimension of a model. *The annals of statistics,* 6**,** 461-464.

SEBER, G. A. F. & WILD, C. J. 2003. *Nonlinear Regression,* Wiley.

SILVERMAN, B. W. 1984. Spline Smoothing - the Equivalent Variable Kernel-Method. *Annals of Statistics,* 12**,** 898-916.

STONE, C. J. 1977. Consistent Nonparametric Regression. *The Annals of Statistics,* 5**,** 595-620.

TAKATSUKA, M. & GAHEGAN, M. 2002. GeoVISTA Studio: a codeless visual programming environment for geoscientific data analysis and visualization. *Computers & Geosciences,* 28**,** 1131-1144.

TIEFELSDORF, M. & GRIFFITH, D. A. 2007. Semiparametric filtering of spatial autocorrelation: the eigenvector approach. *Environment and Planning A,* 39**,** 1193-1221.

TOBLER, W. R. 1970. A Computer Movie Simulating Urban Growth in the Detroit Region. *Economic Geography,* 46**,** 234-240.

VENABLES, W. N. & RIPLEY, B. D. 2002. *Modern Applied Statistics with S,* New York, Springer.

WAHBA, G. 1978. Improper Priors, Spline Smoothing and the Problem of Guarding against Model Errors in Regression. *Journal of the Royal Statistical Society Series B-Methodological,* 40**,** 364-372.

WANG, N., MEI, C. L. & YAN, X. D. 2008. Local linear estimation of spatially varying coefficient models: an improvement on the geographically weighted regression technique. *Environment and Planning A,* 40**,** 986-1005.

WATSON, G. S. 1964. Smooth Regression Analysis. Sankhyā: The Indian Journal of Statistics, Series A (1961-2002), 26**,** 359-372.

WEGMAN, E. J. & WRIGHT, I. W. 1983. Splines in Statistics. *Journal of the American Statistical Association,* 78**,** 351-365.

WHEELER, D. 2006. Diagnostic tools and remedial methods for collinearity in linear regression models with spatially varying coefficients. Doctor of Philosophy, The Ohio State University.

WHEELER, D. 2013. gwrr: Fits geographically weighted regression models with diagnostic tools. R package version 0.2-1 ed.

WHEELER, D. & TIEFELSDORF, M. 2005. Multicollinearity and correlation among local regression coefficients in geographically weighted regression. *Journal of Geographical Systems,* 7**,** 161-187.

WHEELER, D. C. 2007. Diagnostic tools and a remedial method for collinearity in geographically weighted regression. *Environment and Planning A,* 39**,** 2464-2481.

WHEELER, D. C. 2009. Simultaneous coefficient penalization and model selection in geographically weighted regression: the geographically weighted lasso. *Environment and Planning A,* 41**,** 722-742.

WHEELER, D. C. & CALDER, C. A. 2007. An assessment of coefficient accuracy in linear regression models with spatially varying coefficients. *Journal of Geographical Systems,* 9**,** 145-166.

WHITTLE, P. 1954. On Stationary Processes in the Plane. *Biometrika,* 41**,** 434-449.

WOOD, S. N. & AUGUSTIN, N. H. 2002. GAMs with integrated model selection using penalized regression splines and applications to environmental modelling. *Ecological modelling,* 157**,** 157-177.

YEE, T. W. & MITCHELL, N. D. 1991. Generalized Additive Models in Plant Ecology. *Journal of Vegetation Science,* 2**,** 587-602.