AN APPLICATION OF GEOGRAPHICALLY WEIGHTED POISSON REGRESSION

SEAN M. COLLINS

An Application of Geographically Weighted Poisson Regression

by

©Sean M. Collins

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

Department of Mathematics and Statistics
Memorial University of Newfoundland

April 2010

St. John's Newfoundland Canada

Abstract

In fitting regression models with spatial data, it is often assumed that the relationships between the response variable and explanatory variables are the same throughout the study area (i.e., the processes being modelled are stationary over space). This may be a reasonable assumption, but should not be accepted without further analysis. Geographically weighted regression (GWR) is a technique for investigating the validity of this assumption and is used to examine the presence of spatial non-stationarity. It allows relationships between a response variable and the explanatory variables to vary over space. Most studies in GWR to date have focussed on the case where the response variable is continuous and is assumed to follow a normal distribution. However, in many regression models, this is not the case. Here, the concept of geographical weighting is applied to Poisson regression, where the response variable represents a count and takes the form of any non-negative integer.

Acknowledgements

I would like to express my sincere gratitude to my supervisor Dr. Gary Sneddon for his support, guidance, encouragement and patience in the completion of this practicum, as well as for help that he provided me in completing my courses in statistics at both the undergraduate and graduate level.

I sincerely acknowledge the financial support provided by the School of Graduate Studies and Department of Mathematics and Statistics during my time as a full-time graduate student at Memorial. I also would like to thank the Economics and Statistics Branch of the Government of Newfoundland and Labrador for their support and allowing me to complete the requirements for my degree during my time there as a full-time employee.

I would like to thank Amanda Hannaford for providing support with my practicum and for being a great friend.

I would also like to thank Dr. Zhaozhi Fan and Dr. Alwell Oyet for reviewing my practicum and providing helpful comments and criticisms.

Finally, I would like to express my sincere appreciation and gratefulness to my family for their unconditional love, support and encouragement. Without them, I would not have been able to complete this endeavor.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Ordinary Least Squares (OLS) Regression

In an ordinary least squares (OLS) regression model, the dependent variable (or response variable) y is expressed as a linear function of a set of independent (or predictor) variables $x_1, \ldots, x_p$, where y is continuous and the x_k's ($k = 1, \ldots, p$) are qualitative, quantitative, or a combination of both. Based on a sample of n observations, the model can be expressed as follows:

$$y_i = \beta_0 + \sum_{k=1}^{p} \beta_k x_{ik} + \epsilon_i, \tag{1.1}$$

where $i = 1, \ldots, n$; $\beta_0, \beta_1, \ldots, \beta_p$ are regression parameters and the ϵ_i's are assumed to be independent normal random variables with zero mean and constant variance σ^2.

The OLS estimate of the coefficient parameter vector is given by:

$$\widehat{\boldsymbol{\beta}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}, \tag{1.2}$$

where $\mathbf{y} = (y_1, \ldots, y_n)^T$, $\widehat{\boldsymbol{\beta}} = (\widehat{\beta}_0, \widehat{\beta}_1, \ldots, \widehat{\beta}_p)^T$ and

$$\mathbf{X} = \begin{pmatrix} 1 & x_{11} & \cdots & x_{1p} \\ 1 & x_{21} & \cdots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & \cdots & x_{np} \end{pmatrix}$$

$\mathbf{X}$ is the $n \times (p+1)$ design matrix (with a vector of 1's in the first column for the intercept term) and $\mathbf{X}^T$ denotes the transpose of $\mathbf{X}$.

1.2 Geographically Weighted Regression (GWR)

In spatial analysis, the model represented in equation (1.1) is often referred to as a global (or spatially stationary) model since the relationship between the response variable $\mathbf{y}$ and the predictor variables $\mathbf{x} = (x_1, \ldots, x_p)$ is assumed to be constant (or stationary) across the study region; that is, the parameters do not change with geographical location (Atkinson et. al, 2003).

However, a global regression model is only appropriate when it is reasonable to assume that the relationship between $\mathbf{y}$ and $\mathbf{x}$ does not change across the study region. In some cases, there may be evidence to suggest that the relationship between the response and predictor variables depends on the geographical location and that the parameters are not constant over the study region - that is, the relationships being examined may exhibit significant spatial variation that is not accounted for in the estimation of the global parameter estimates. This variation is referred to as spatial non-stationarity (Brunsdon et. al, 1996; Fotheringham et. al, 1996, 1998). Instead of a global model, a model of the following form should be used in determining the nature of the relationship between $\mathbf{y}$ and $\mathbf{x}$:

$$y_i = \beta_{i0} + \sum_{k=1}^{p} \beta_{ik} x_{ik} + \epsilon_i, \tag{1.3}$$

where β_{ik} is the value of the k^{th} parameter at the i^{th} data point. Although this model allows the parameters to vary over space, there are problems in calibrating this type of model since there are more unknowns than observed variables - as the number of observations increases, the number of parameters increases as well (Leung et. al, 2000). To overcome this problem, Fotheringham, Charlton and Brunsdon (1998) developed a technique known as geographically weighted regression (GWR). This technique allows the regression model to be expressed in the form of equation (1.3); however, the regression coefficients are assumed to be deterministic functions of some other variables (in the case of GWR, the geographical location in space) rather than random variables.

GWR is a technique used to account for (and to examine the presence of) spatial non-stationarity by calibrating a regression model which allows different relationships to exist at different locations in space. In the process of fitting a GWR model to spatial data, the regression parameters in equation (1.3) are estimated by a weighted least squares (WLS) procedure such that the weighting system is dependent on the location in geographical space and allows local rather than global parameters to be estimated (Leung et. al, 2000). The weighting scheme used in calibrating a GWR model at the i^{th} data point is such that observed data that are near point i have more of an influence in the estimation of the local regression coefficient estimates than data located farther from i - that is, data from observations close to i have higher weights and data from observations farther away have lower weights (Fotheringham, Charlton and Brunsdon, 1998).

For the Gaussian GWR model (shown in equation (1.3)), local estimates of the parameters may be produced at locations other than those at which data are observed or sampled - that is, estimates of $\boldsymbol{\beta}$ can be computed for any point in space, regardless of whether or not that point is an actual data point (Fotheringham, Brunsdon and Charlton, 2002). Let $j = 1, \ldots, n$ represent the locations at which data have been observed (i.e., data points, or sample points) and let $i = 1, \ldots, m$ represent the locations at which the local coefficient estimates are produced (i.e., regression points). In most cases, the regression points and the data points will be the same. However, for geographically weighted Poisson or logistic regression models, the regression points must be the same as the data points - that is, there is no option of producing local parameter estimates at points other than the data points (Charlton, Fotheringham and Brunsdon, 2003). This will be discussed further in Chapter 2.

The GWR estimate of $\boldsymbol{\beta}$ at the i^{th} regression point is given by the following:

$$\widehat{\boldsymbol{\beta}}(i) = (\mathbf{X}^T\mathbf{W}(i)\mathbf{X})^{-1}\mathbf{X}^T\mathbf{W}(i)\mathbf{y}, \qquad (1.4)$$

where $\mathbf{W}(i)$ is an $n \times n$ spatial weighting matrix whose off-diagonal elements are zero and whose diagonal elements represent the weighting of each of the n observed data points for regression point i (such that each of the n data points are weighted in accordance with its proximity to point i). That is,

$$\mathbf{W}(i) = \begin{pmatrix} w_{i1} & 0 & \cdots & 0 \\ 0 & w_{i2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & w_{in} \end{pmatrix} \tag{1.5}$$

where w_{ij} is the weight given to the j^{th} data point in the calibration of the GWR model for the i^{th} regression point.

Since the weights in GWR vary according to the geographical location of the i^{th} regression point, the weighting matrix $\mathbf{W}(i)$ has to be computed for all regression points. By computing $\mathbf{W}(i)$, and evaluating (1.4) for all values of i, sets of local regression parameter estimates can be obtained. Assuming that there are m regression points and p explanatory variables, the $m \times (p+1)$ matrix $\widehat{\boldsymbol{\beta}}$ which contains the complete set of GWR parameter estimates can be expressed by the following:

$$\widehat{\boldsymbol{\beta}} = \begin{pmatrix} \widehat{\beta}_{01} & \widehat{\beta}_{11} & \cdots & \widehat{\beta}_{p1} \\ \vdots & \vdots & \vdots & \vdots \\ \widehat{\beta}_{0i} & \widehat{\beta}_{1i} & \cdots & \widehat{\beta}_{pi} \\ \vdots & \vdots & \vdots & \vdots \\ \widehat{\beta}_{0m} & \widehat{\beta}_{1m} & \cdots & \widehat{\beta}_{pm} \end{pmatrix}$$

$$= \begin{pmatrix} \widehat{\boldsymbol{\beta}}(1)^T \\ \vdots \\ \widehat{\boldsymbol{\beta}}(i)^T \\ \vdots \\ \widehat{\boldsymbol{\beta}}(m)^T \end{pmatrix}$$

where $\widehat{\boldsymbol{\beta}}(i)^T = (\widehat{\beta}_{0i}, \widehat{\beta}_{1i}, \ldots, \widehat{\beta}_{pi})$ is computed by (1.4).

1.3 Geographically Weighted Poisson Regression (GWPR)

To date, most studies in GWR have focused on the case where the dependent (or response) variable is continuous and follows (approximately) a normal, or Gaussian distribution. That is, most applications of GWR have involved calibrating a model with a Gaussian error term (which is the geographically weighted equivalent of an

OLS regression model). However, in many models, the dependent variable is actually discrete. It may be defined for integers only (e.g., the number of traffic accidents that occur at a busy intersection over a month), or it can be a binary or categorical variable (e.g., yes/no, disagree/no opinion/agree, etc.). In these cases, applying a Gaussian model is inappropriate. However, the concept of GWR can be applied to generalized linear models (GLMs), including those based on the binomial or Poisson distributions.

Poisson regression is used when the dependent variable refers to counts of the occurrences of some event over time or space and takes the form of any non-negative integer (i.e., 0, 1, 2, 3, ...). The independent variable(s) can be either qualitative, quantitative or a combination of both. In order to fit a geographically weighted Poisson regression (GWPR) model, each observation must also have information that describes its location. This information can be in the form of (x,y) coordinates based on a two-dimensional Cartesian grid, or latitude-longitude coordinates which describe a point on the Earth's surface.

The focus of this practicum is to study the technique of GWPR, its underlying theory and practical applications. Chapter 2 discusses the theoretical aspects of GWPR including how a GWPR model is calibrated, spatial weighting functions, bandwidth (or distance) selection, calculation of the GWPR model deviance and AIC (Akaike Information Criterion) as well as hypothesis testing procedures. The results of simulation studies are described in Chapter 3, where the focus is on computing the estimates of the coefficient parameters from fitting a variety of GWPR models to simulated data from a Poisson distribution. In Chapter 4, a dataset is selected for the application of GWPR methods. Conclusions and areas of further work and future research are discussed in Chapter 5.

Chapter 2

Theory of GWPR

2.1 Introduction

As in the case of a normal or Gaussian GWR model, the resulting regression coefficient estimates from a GWPR model are specific to each individual location. However, unlike Gaussian GWR, a GWPR model is calibrated using a procedure known as iteratively reweighted least squares (IRLS). As a result, the amount of time required to calibrate a GWPR model will be much longer, compared to an equivalent Gaussian GWR model. Also, the observed values of the dependent variable are required in order to compute the standard errors of the regression coefficients, so parameter estimates can only be obtained at the same location as the data points; that is, the regression points must be the same as the data points (Charlton, Fotheringham and Brunsdon, 2003).

2.2 Calibration of a GWPR Model

Using results from Nakaya et al. (2005), the regression coefficients of a GWPR model can be estimated using a modified local Fisher scoring procedure, which is a form of iteratively reweighted least squares (IRLS).

Let $\widehat{\boldsymbol{\beta}}^{(l)}(i)$ be the vector of local parameter estimates for the i^{th} location, with l representing the number of iterations, where

$$\widehat{\boldsymbol{\beta}}^{(l)}(i) = (\widehat{\beta_0}^{(l)}(i), \widehat{\beta_1}^{(l)}(i), \ldots, \widehat{\beta_p}^{(l)}(i))^T \tag{2.1}$$

The local parameter estimates for the i^{th} location and $(l+1)^{th}$ iteration is defined as follows:

$$\widehat{\boldsymbol{\beta}}^{(l+1)}(i) = (\mathbf{X}^T\mathbf{W}(i)\mathbf{V}(i)^{(l)}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{W}(i)\mathbf{V}(i)^{(l)}\mathbf{z}(i)^{(l)} \tag{2.2}$$

$\mathbf{X}$ is the $n \times (p+1)$ design matrix (with a vector of 1's in the first column for the intercept term) and $\mathbf{X}^T$ denotes the transpose of $\mathbf{X}$. $\mathbf{W}(i)$ represents the diagonal spatial weights matrix for the i^{th} location, $\mathbf{V}(i)$ denotes the diagonal variance weights matrix for the i^{th} location and $\mathbf{z}(i)$ is the vector of adjusted response variables at the i^{th} location.

The above matrix computation should be repeated to update the local parameter estimates until convergence is reached.

The diagonal variance weights matrix for the i^{th} location at the l^{th} iteration is as follows:

$$\mathbf{V}(i)^{(l)} = \begin{pmatrix} \hat{\mu}_1(\beta^{(l)}(i)) & \cdots & \cdots & 0 \\ \vdots & \hat{\mu}_2(\beta^{(l)}(i)) & \cdots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & \cdots & \hat{\mu}_n(\beta^{(l)}(i)) \end{pmatrix} \tag{2.3}$$

where

$$\hat{\mu}_j(\beta^{(l)}(i)) = \exp\left(\sum_{k=0}^{p} \hat{\beta_k}^{(l)}(i)x_{jk}\right) \tag{2.4}$$

The vector of adjusted response variables for the i^{th} location at the l^{th} iteration is as follows:

$$\mathbf{z}^{(l)}(i) = (z_1^{(l)}(i), z_2^{(l)}(i), \ \ldots, z_n^{(l)}(i))^T \tag{2.5}$$

where

$$z_j^{(l)}(i) = \sum_{k=0}^{p} \hat{\beta_k}^{(l)}(i)x_{jk} + \frac{y_j - \hat{\mu}_j(\beta^{(l)}(i))}{\hat{\mu}_j(\beta^{(l)}(i))} \tag{2.6}$$

Note that $x_{j1} = 1$ for $j = 1, 2, \cdots, n$.

At convergence, the local parameter estimates for the i^{th} location can be written as:

$$\widehat{\beta}(i) = (\mathbf{X}^T\mathbf{W}(i)\mathbf{V}(i)\mathbf{X})^{-1}\mathbf{X}^T\mathbf{W}(i)\mathbf{V}(i)\mathbf{z}(i) \tag{2.7}$$

By repeating this procedure for each regression point, sets of local parameter estimates can be obtained.

2.3 Log Likelihood Function, Deviance and AIC

A generalized linear model (GLM) has three components: a random component, a systematic component and a link function (Agresti, 2002). The random component identifies the response variable **Y** with independent observations $(y_1, \ldots, y_n)$ from a distribution in the natural exponential family. This family has probability density function or mass function of the following form:

$$f(y_i; \theta_i) = \exp\{y_i\theta_i - b(\theta_i) + c(y_i)\} \tag{2.8}$$

Several important distributions are special cases of the natural exponential family, including the Poisson distribution. When Y_i follows a Poisson distribution with positive mean μ_i, it can be shown that (Agresti, 2002):

$$f(y_i; \mu_i) = \exp\{y_i \log(\mu_i) - \mu_i - \log(y_i!)\} \tag{2.9}$$

Let $L(\boldsymbol{\mu}; \boldsymbol{y})$ represent the log-likelihood function and let $L_i = \log f(y_i; \mu_i)$ denote the contribution of y_i to the log-likelihood. Then, it can be shown that:

$$L(\boldsymbol{\mu}; \boldsymbol{y}) = \sum_i L_i = \sum_i [y_i \log(\mu_i) - \mu_i - \log(y_i!)] \tag{2.10}$$

Let $\boldsymbol{y} = (y_1, \ldots, y_n)$ represent n independent observations from a Poisson distribution with mean μ_i. The estimated log-likelihood function for the model is:

$$L(\hat{\boldsymbol{\mu}}; \boldsymbol{y}) = \sum_{i=1}^{n} \widehat{L}_i = \sum_{i=1}^{n} [y_i \log(\hat{\mu}_i) - \hat{\mu}_i - \log(y_i!)] \tag{2.11}$$

where the maximum likelihood estimate of $\boldsymbol{\mu}$ is:

$$\hat{\mu}_i = \exp\left\{\sum_{k=0}^{p} \hat{\beta}_k x_{ik}\right\} \tag{2.12}$$

For a GWPR model, the estimated log-likelihood function for the i^{th} location $(i = 1, \ldots, n)$ can be written as

$$\widehat{L}_i(\beta(i)) = y_i \log(\hat{\mu}_i(\beta(i)) - \hat{\mu}_i(\beta(i)) - \log(y_i!) \tag{2.13}$$

where

$$\hat{\mu}_i(\beta(i)) = \exp\left\{\sum_{k=0}^{p} \hat{\beta}_k(i) x_{ik}\right\} \tag{2.14}$$

The estimated log-likelihood function for a GWPR model is:

$$L(\hat{\boldsymbol{\mu}}(\boldsymbol{\beta}(\boldsymbol{i})); \boldsymbol{y}) = \sum_{i=1}^{n} \widehat{L}_i(\beta(i)) \tag{2.15}$$

where $\widehat{L}_i(\beta(i))$ reflects the dependence of $\hat{\mu}_i$ on the model parameters $\beta(i)$ estimated for the i^{th} regression point.

Let $L(\boldsymbol{y}; \boldsymbol{y})$ represent the maximum achievable log-likelihood over all possible models. This occurs for the most general model (known as the saturated model), which has a separate parameter for each observation and provides a perfect fit to the data (i.e., $\hat{\mu}_i = y_i$, for $i = 1, \ldots, n$). The saturated model can be used as a baseline for comparison with other model fits, and is given by:

$$L(\boldsymbol{y};\boldsymbol{y}) = \sum_{i=1}^{n}[y_i \log(y_i) - y_i - \log(y_i!)] \tag{2.16}$$

The deviance for a GWPR model can be expressed by the following:

$$D(\boldsymbol{y};\hat{\boldsymbol{\mu}}(\boldsymbol{\beta}(i))) = -2\left[L(\hat{\boldsymbol{\mu}}(\boldsymbol{\beta}(i);\boldsymbol{y}) - L(\boldsymbol{y};\boldsymbol{y})\right] \tag{2.17}$$

Using equations (2.13), (2.15) and (2.16) results in the following formula for the deviance:

$$D(\boldsymbol{y};\hat{\boldsymbol{\mu}}(\boldsymbol{\beta}(i))) = 2\sum_{i=1}^{n}\left[y_i \log\left(\frac{y_i}{\hat{\mu}_i(\beta(i))}\right) - y_i + \hat{\mu}_i(\beta(i))\right] \tag{2.18}$$

The *AIC* for a GWPR model is defined as (Nakaya et. al, 2005):

$$AIC = D(\boldsymbol{y};\hat{\boldsymbol{\mu}}(\boldsymbol{\beta}(i))) + 2K \tag{2.19}$$

where K is the effective number of parameters for the model. Since the trace of the hat matrix is equivalent to the number of regression parameters in a GLM (including the Poisson model), then the effective number of parameters for a GWPR model can be defined as:

$$K = trace(\mathbf{S}) \tag{2.20}$$

where $\mathbf{S}$ represents the hat matrix for the model and maps $\mathbf{y}$ to the fitted values $\hat{\mathbf{y}}$ in the following manner:

$$\hat{\mathbf{y}} = \mathbf{S}\mathbf{y} \tag{2.21}$$

where the i^{th} row of $\mathbf{S}$ is given by:

$$s_i = \mathbf{x}_i(\mathbf{X}^T\mathbf{W}(i)\mathbf{V}(i)\mathbf{X})^{-1}\mathbf{X}^T\mathbf{W}(i)\mathbf{V}(i) \tag{2.22}$$

The corrected AIC (or AIC_c) for a GWPR model is defined as (Nakaya et. al, 2005):

$$AIC_c = AIC + 2(\frac{K(K+1)}{n-K-1}) \tag{2.23}$$

There is little difference between AIC_c and AIC if the effective number of parameters K is small relative to the number of observations n. However, AIC_c is often preferred over AIC since the degrees of freedom are likely to be small in the case of local regression (Nakaya et. al, 2005).

To choose between a number of competing models (e.g., a GWPR model vs. a global Poisson regression model; GWPR models with different explanatory variables; GWPR models with different bandwidths), Nakaya et. al recommend computing the AIC_c for each model and the model with the smallest AIC_c should be selected as the best model. A common rule-of-thumb in the use of AIC_c is that there is a significant difference in the performance of two competing models if the difference in the AIC_c values between the models is at least 3 (Nakaya et. al, 2005).

2.4 Choice of Spatial Weighting Function

Let w_{ij} represent the geographical weight of the j^{th} data point at the i^{th} regression point (i.e., the weight given to the j^{th} observation in the calibration of the GWR model for the i^{th} location) and let d_{ij} represent the distance between the j^{th} observation and the i^{th} regression point. Under a global regression model with no geographical weighting, each observation has a weight of unity:

$$w_{ij} = 1,\ i, j = 1, \ldots, n$$

where j represents a point in space at which data are observed and i represents a point in space for which parameters are estimated (i.e., a regression point). In GWR models, the weights vary according to the location of the i^{th} regression point, where $0 \leq w_{ij} \leq 1$ and where w_{ij} decreases as d_{ij} increases. That is, observations that are recorded at locations close to the i^{th} regression point are given higher weighting than observations recorded at locations further away. Also, as the focal point of the regression changes, so do the weights - hence, the weighting matrix $\mathbf{W}$ has to be computed for each regression point. Some of the weighting functions encountered in the literature (Fotheringham et. al, 2002) are listed in Table 2.1 and are discussed below.

For the binary weighting function, only a subset of the data points is used to calibrate the model at each regression point. Data points that lie within a certain distance d of the regression point are given a weight of one, while all other data points are given a weight of zero. However, while many spatial processes are continuous, this weighting function is discrete. The estimates for the model coefficients could change

Table 2.1: Spatial Weighting Functions Commonly Used in Geographically Weighted Regression

Weighting Function	**Formula**	**Notes**
Binary	$w_{ij} = \begin{cases} 1, & \text{if } d_{ij} < d \\ 0, & \text{otherwise} \end{cases}$	d = the distance at which w_{ij} is set to zero
Gaussian	$w_{ij} = \exp[-1/2(\frac{d_{ij}}{b})^2]$	b = bandwidth
Fixed Bi-Square	$w_{ij} = \begin{cases} [1 - (\frac{d_{ij}}{d})^2]^2, & \text{if } d_{ij} < d \\ 0, & \text{otherwise} \end{cases}$	d = the distance at which w_{ij} is set to zero
Adaptive Bi-Square	$w_{ij} = \begin{cases} [1 - (\frac{d_{ij}}{d})^2]^2, & \text{if } d_{ij} < d \\ 0, & \text{otherwise} \end{cases}$	d = the distance to the M^{th} nearest neighbour (and j is one of the M^{th} nearest neighbours of the i^{th} regression point)

drastically as the regression point changes. The other weighting functions (Gaussian, fixed bi-square and adaptive bi-square) try to solve the problem of discontinuity by specifying w_{ij} as a continuous function of d_{ij}, the distance between regression point i and data point j.

For the Gaussian weighting function, the weights gradually decrease according to a Gaussian curve as the value of d_{ij} increases. As d_{ij} becomes larger, the degree of influence of surrounding points will decay exponentially and the value of w_{ij} will fall towards zero (i.e., observations that are far away from the i^{th} regression point will have little or no influence on the parameter estimates for location i.) The parameter b (which is known as the bandwidth or scale parameter) controls the rate at which the weight of a data point decreases as the distance between the location at which the data point is recorded and the regression point increases. If the bandwidth is large, the weights decrease slowly; if the bandwidth is small, the weights decrease rapidly.

For the fixed bi-square weighting function, the weights will also decrease according to a continuous curve as the value of d_{ij} increases. However, the weights for data points whose distance from the i^{th} regression point is greater than or equal to d will be set to zero (i.e., points that lie beyond a certain distance will be excluded from the estimation of the regression parameters for location i).

The above types of spatial weighting functions are fixed in terms of their shape and magnitude over space. However, a fixed weighting function may not be appropriate when the data points are not evenly distributed over space - in some regions, the density of data points may be high, while in other regions, the data points may be sparse. In regions where data are dense, the estimates obtained from fixed weighting functions are more likely to be biased since the number of data points used in the calibration of the local model will be large. In regions where data are sparse, the standard errors of the coefficient estimates will be high since the number of data points used in the calibration of the local model will be small; thus, the estimates obtained from fixed weighting functions will be unreliable.

An adaptive weighting function is a spatially varying weighting method which reduces the problems that can occur with a fixed weighting function when the density of the data points varies over space. The adaptive bi-square weighting function mentioned in Table 2.1 involves a function that is related to the M^{th} nearest data points (or nearest neighbours) of the i^{th} regression point. The calibration of a local model using this weighing function involves the estimation of M, which is the number of data points (besides the regression point) to be included in the model calibration.

This weighting function ensures that the number of data points to be included in the local model is the same for each regression point. The weights of each data point (up to the M^{th} point) will decrease according to a continuous curve as the distance between the i^{th} regression point and the j^{th} data point increases; weights for all data points beyond the i^{th} point are set to zero.

2.5 Choice of Neighbourhood Size

In addition to selecting a spatial weighting function, one must also determine an optimal window (or neighbourhood) size which will be used to subset the data locally for model estimation (Farber and Paez, 2007). This involves choosing a suitable bandwidth b (if the Gaussian weighting function is used), distance d (if the binary or fixed bi-square weighting function is used) or the number of nearest neighbours M (if the adaptive bi-square weighting function is used). Although the estimated regression coefficients depend, in part, on the weighting function chosen, studies have indicated that the choice of an appropriate value of b, d or M has more of an influence on the estimated parameters than the choice of the weighting function (Simonoff, 1996).

There are a number of methods for selecting an appropriate value of b, d or M that are mentioned in the literature, including cross-validation, generalized cross-validation, the Akaike Information Criterion, the Bayesian Information Criterion and the Schwartz Information Criterion. However, the methods that are most often used in practice are the cross-validation criterion (CV) and the Akaike Information Criterion (AIC).

The cross-validation (CV) approach suggested by Cleveland (1979) and Bowman (1984) involves choosing a value of b, d or M such that the following quantity is minimized:

$$CV \text{ Score} = \sum_{i=1}^{n} [y_i - \hat{y}_{\neq i}(b)]^2 \tag{2.24}$$

where $\hat{y}_{\neq i}(b)$ is the fitted value of y_i using a bandwidth of b (or distance d or number of nearest neighbours M) such that the i^{th} observation is omitted from the estimation of the GWR model. Omitting this data point is necessary since minimizing the above equation using all observations results in the fitted values tending toward their corresponding actual values (i.e., the CV score tends toward zero). This occurs when the weights for all other data points (except for the i^{th} data point) tend to zero, which occurs when the value of b, d or M tends to zero (a meaningless result since the parameter estimates are undefined in this case). The CV or 'leave-one-out' score can be regarded as the sum of the squared errors associated with estimating $\hat{y}_{\neq i}(b)$ at each data point, where each point contributes toward the total CV score (Farber and Paez, 2007).

The AIC approach involves choosing a value of b, d or M which minimizes the AIC. In addition to GWPR, the AIC can also be used for choosing a suitable value of b, d or M in logistic GWR and other geographically weighted generalized linear models. It can also be used to assess whether a particular GWR model provides a better fit than another model as mentioned previously.

Nakaya et. al (2005) mention that either the cross-validation (CV) score or AIC (conventional and corrected) can be used to select an appropriate value of b, d or

M for a GWPR model. However, they recommend using the corrected AIC (or AIC_c) given in equation (2.23) since it allows for a more suitable penalty for model complexity (i.e., models with a large number of predictors will tend to have a higher AIC_c) and is less likely to result in an over-parameterized model than using the CV score and the conventional AIC. In addition, empirical and simulation studies by Farber and Paez (2007) have shown that it is possible for the CV score to be heavily influenced by a small number of highly influential observations in the dataset (i.e., points that have relatively larger differences between y_i and $\hat{y}_{\neq i}(b)$ and thus, impact the CV score disproportionately); when this occurs, the chosen value of b, d or M may not be optimal for model calibration.

One problem that may arise in calibrating a GWPR model is when the chosen value of b, d or M is too low and the dependent variable consists of a large number of zeroes (Charlton, Fotheringham and Brunsdon, 2003). When this occurs, it is possible that all of the values for the dependent variable are zero for an individual regression point; hence, the estimated regression coefficients cannot be computed for that data point. As a result, one must use caution in selecting an optimal value of b, d or M for a GWPR model when there are a large number of zeroes in the data.

2.6 Covariance Matrix and Hypothesis Testing

Using results from Nakaya et. al (2005), the asymptotic variance-covariance matrix of the estimated regression coefficients $\widehat{\boldsymbol{\beta}}(i)$ at the i^{th} data point is given by:

$$\mathbf{Cov}(\widehat{\boldsymbol{\beta}}(i)) = \mathbf{C}(i)\mathbf{A}(i)^{-1}\mathbf{C}(i)^T \tag{2.25}$$

where

$$\mathbf{C}(i) = (\mathbf{X}^T\mathbf{W}(i)\mathbf{V}(i)\mathbf{X})^{-1}\mathbf{X}^T\mathbf{W}(i)\mathbf{V}(i) \tag{2.26}$$

The standard error of $\widehat{\beta}_k(i)$, the estimate of the k^{th} regression coefficient at the i^{th} data point, is given by:

$$\mathrm{SE}(\widehat{\beta}_k(i)) = \sqrt{\mathbf{Cov}(\widehat{\boldsymbol{\beta}}(i))_k} \tag{2.27}$$

where $\mathbf{Cov}(\widehat{\boldsymbol{\beta}}(i))_k$ is the k^{th} diagonal element of the variance-covariance matrix for the i^{th} data point.

The local t-statistic for $\widehat{\beta}_k(i)$ is then computed using the following formula:

$$t_k(i) = \frac{\widehat{\beta}_k(i)}{\mathrm{SE}(\widehat{\beta}_k(i))} \tag{2.28}$$

This quantity approximately follows the t-distribution with $n - K$ degrees of freedom, where K is the effective number of parameters for the GWPR model, and can be computed for each regression coefficient estimate at each data point. The resulting local t-statistics can then be used for conducting local tests of hypotheses for parameter significance.

In testing to see if there is evidence of spatial non-stationarity with respect to a particular predictor variable, Nakaya et, al (2005) recommend the following approach: calibrate a full GWPR model (i.e., a model where all coefficient parameters are allowed to vary spatially) and a mixed GWPR model (i.e., such that the coefficient parameter associated with the variable of interest is assumed to be constant across the study region, while all other coefficients are allowed to vary spatially). If the AIC_c from the mixed GWPR model is lower than that of the full GWPR model, then there is evidence to suggest that the relationship between the response variable and the predictor variable of interest varies spatially; otherwise, there is little evidence to suggest that the relationship varies over space.

Chapter 3

Simulation Studies

3.1 Introduction

This chapter presents results from fitting two different geographically weighted Poisson regression (GWPR) models using simulated data. The first model is one with a single predictor variable and the second is one with three predictor variables. Results from both models were obtained using the spatial weighting functions discussed in the previous chapter (i.e., binary, Gaussian, fixed bi-square and adaptive bi-square), and a variety of bandwidth/distance/nearest neighbour values were used with each of the weighting functions. Estimates of the local regression coefficients, the log-likelihood, effective number of parameters, deviance and Akaike Information Criterion (AIC) were computed from each of the fitted GWPR models. Hypothesis tests for determining the statistical significance of the regression coefficient parameters were also carried out for both the single and multiple predictor models.

3.2 Single Predictor GWPR Model

The general form of a GWPR model for a response variable y and a single predictor variable x is given by

$$y_i = Poisson(\mu_i) \tag{3.1}$$

where

$$\mu_i = \exp(\beta_{0i} + \beta_{1i}x_i) \tag{3.2}$$

and where $i = 1, \cdots, n$. The spatial region of interest consists of the coordinates (u_i, v_i) on a two-dimensional Cartesian grid. The simulation is conducted such that the grid consists of $m \times m$ lattice points with unit distance between any two adjacent points along the horizontal or vertical axes. Throughout this chapter, $i = 1$ refers to the location in the lower-left hand corner of the grid, $i = 2$ is the point above it, and so on, such that the n^{th} point is located in the upper-right hand corner.

For the single predictor GWPR model, we chose $m = 4$, so there are $n = m^2 = 16$ observations in the study region. Figure 3.1 illustrates the geographical location of the 16 points in the study region.

The values for the regression coefficients (β_{0i} and β_{1i}) and the predictor variable (x_i) for the 16 data points were chosen using the following step-changing approach:

$$\beta_{0i} = \begin{cases} 0.1, & \text{for } i = 1, \ldots, 4 \\ 0.3, & \text{for } i = 5, \ldots, 8 \\ 0.6, & \text{for } i = 9, \ldots, 12 \\ 0.9, & \text{for } i = 13, \ldots, 16 \end{cases}$$

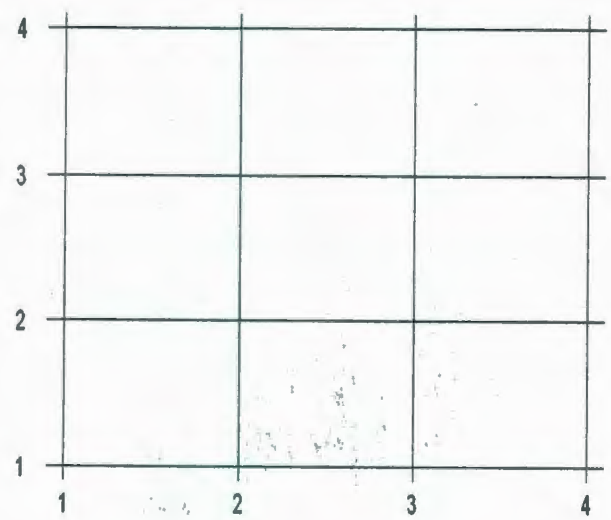

Figure 3.1: A grid with 4 × 4 lattice points

$$\beta_{1i} = \begin{cases} 1, & \text{for } i = 1, 5, 9, 13 \\ 2, & \text{for } i = 2, 6, 10, 14 \\ -1, & \text{for } i = 3, 7, 11, 15 \\ -2, & \text{for } i = 4, 8, 12, 16 \end{cases}$$

$$x_i = \begin{cases} 1.0, & \text{for } i = 1, 5, 9, 13 \\ 0.7, & \text{for } i = 2, 6, 10, 14 \\ -1.3, & \text{for } i = 3, 7, 11, 15 \\ -0.4, & \text{for } i = 4, 8, 12, 16 \end{cases}$$

The values for μ_i, $i = 1, \cdots, 16$, were calculated using equation (3.2) and the chosen values of β_{0i}, β_{1i} and x_i. The values of the response variable y_i, $i = 1, \cdots, 16$, were then randomly generated from a Poisson distribution using equation (3.1) - this process was carried out 1000 times in order to obtain the simulated datasets that were used in the analysis for the single predictor case.

3.2.1 Estimates of Regression Coefficients

The estimates of β_{0i} and β_{1i} for the 16 data points were computed from each of the 1000 simulated datasets using the iteratively reweighted least squares method described in Chapter 2 (the initial estimated values of β_{0i} and β_{1i} were obtained from fitting a global Poisson regression model to the data and were updated using the IRLS method until convergence was reached). The Euclidean distance between points on the 4×4 grid were computed and were used in the calculation of the weights w_{ij} for the binary, Gaussian, fixed bi-square and adaptive bi-square spatial weighting functions. A number of GWPR models were calibrated using a variety of bandwidths (Gaussian), distances (binary and fixed bi-square) and values of M (adaptive bi-square). A global Poisson regression model (with no geographical weighting) was also calibrated for each dataset to serve as a comparison to the results obtained from the GWPR models.

The average AIC_c from the simulated datasets was computed for each calibrated model and the values of b, d and M chosen for the four spatial weighting functions were those which produced the minimum average AIC_c. Average CV scores were also computed for each model and the values of b, d and M which gave minimum average

AIC_c values were reasonably close to those which gave minimum average CV scores. Table 3.1 shows the average AIC_c and average CV scores for a number of GWPR models that were calibrated for the single-predictor case.

The minimum average AIC_c values occur at $d = 2.0$ for the binary weighting function, $b = 1.5$ for the Gaussian weighting function, $d = 3.5$ for the fixed bi-square weighting function and $M = 15$ for the adaptive bi-square weighting function. Also, the average AIC_c values for all GWPR models shown in Table 3.1 are lower than the average AIC_c for the global Poisson regression model (31.99558). The mean and standard deviation of the regression coefficient estimates for the binary (distance $d = 2.0$), Gaussian (bandwidth $b = 1.5$), fixed bi-square (distance $d = 3.5$) and adaptive bi-square ($M = 15$) weighting functions are shown in Tables 3.2 to 3.5 respectively.

Table 3.1: Single Predictor GWPR Models: Average AIC_c Values and Average CV Scores

Weighting Function		Average AIC_c	Average CV Score
Binary	$d = 2.0$	27.31918	38.46512
Binary	$d = 2.5$	29.45442	42.52043
Binary	$d = 3.0$	30.45489	43.93526
Binary	$d = 3.5$	31.69904	46.97942
Binary	$d = 4.0$	31.89712	47.19748
Gaussian	$b = 1.0$	31.16112	373.58236
Gaussian	$b = 1.5$	28.39134	38.13858
Gaussian	$b = 2.0$	29.07540	40.67315
Gaussian	$b = 3.0$	30.35731	43.82787
Gaussian	$b = 5.0$	31.33713	46.01594
Fixed	$d = 2.5$	30.30836	64.42804
Fixed	$d = 3.0$	28.34982	37.23726
Fixed	$d = 3.5$	28.19576	38.02199
Fixed	$d = 4.0$	28.54374	39.28102
Fixed	$d = 5.0$	29.47119	41.71317
Adaptive	$M = 15$	28.44211	38.14971
Adaptive	$M = 14$	28.72668	37.33535
Adaptive	$M = 12$	28.92167	37.32353
Adaptive	$M = 11$	28.99388	37.34254
Adaptive	$M = 9$	31.58296	42.50882
Global		31.99558	47.43878

Table 3.2: Summary statistics of GWPR parameter estimates: Binary weighting function with $d = 2.0$

Points	Intercept ($\widehat{\beta_0}$) Mean	Intercept ($\widehat{\beta_0}$) St. Dev.	Slope ($\widehat{\beta_1}$) Mean	Slope ($\widehat{\beta_1}$) St. Dev.
1	1.380424	0.3186759	0.05679597	0.3390346
2	1.448374	0.1810716	0.05309431	0.1956084
3	1.379398	0.1957863	-0.10843129	0.2123751
4	1.250920	0.3022829	-0.06997141	0.3696410
5	1.577091	0.2211894	0.02592263	0.2335517
6	1.629591	0.1422096	0.02875213	0.1468863
7	1.564026	0.1504539	-0.08867626	0.1660162
8	1.424075	0.2658942	-0.08131402	0.3050974
9	1.815996	0.1933083	-0.09426530	0.1923349
10	1.799082	0.1270725	-0.05191960	0.1312799
11	1.772903	0.1390153	0.05725511	0.1495637
12	1.565832	0.2558508	-0.13390825	0.2938054
13	1.985336	0.1779333	-0.17089310	0.1698431
14	1.922429	0.1433866	-0.05631208	0.1475726
15	1.906150	0.1476894	0.07223157	0.1600729
16	1.836293	0.2329335	0.04367913	0.2900557
Global statistic	1.657375	0.1129078	-0.01274097	0.1260806

As shown in Tables 3.2 to 3.5, the means of the GWPR estimates of β_0 are much larger than their corresponding true values at each of the 16 locations for all 4 models (i.e., for all spatial weighting functions), so there is a large positive bias associated with these estimates. For most of the first 8 locations, the GWPR estimates of β_0 using the binary weighting function have the lowest bias among the 4 weighting functions while the Gaussian and adaptive bi-square weighting functions have the highest bias. However, for most of the last 8 locations, the GWPR estimates using the Gaussian weighting function have the lowest bias while the binary and adaptive bi-square weighting functions have the highest bias. The standard deviations of the

Table 3.3: Summary statistics of GWPR parameter estimates: Gaussian weighting function with $b = 1.5$

Points	Intercept ($\widehat{\beta}_0$)		Slope ($\widehat{\beta}_1$)	
	Mean	St. Dev.	Mean	St. Dev.
1	1.529052	0.1415824	-0.0520005732	0.1434242
2	1.519977	0.1380288	-0.0164573930	0.1492346
3	1.490459	0.1371564	0.0028531741	0.1635775
4	1.419947	0.1511544	-0.0150061307	0.1898297
5	1.647574	0.1242936	-0.0534318110	0.1241223
6	1.638418	0.1216429	-0.0182903097	0.1289533
7	1.609061	0.1214874	0.0008668628	0.1411059
8	1.539256	0.1342213	-0.0166939654	0.1633812
9	1.771096	0.1227802	-0.0550500454	0.1199244
10	1.762058	0.1204212	-0.0201188965	0.1244368
11	1.732917	0.1206723	-0.0010982362	0.1359728
12	1.663495	0.1337516	-0.0186496370	0.1572763
13	1.875452	0.1320706	-0.0563863611	0.1278008
14	1.866644	0.1294330	-0.0214343088	0.1325985
15	1.837687	0.1297853	-0.0024750800	0.1449892
16	1.768288	0.1444003	-0.0202116066	0.1679986
Global statistic	1.657375	0.1129078	-0.01274097	0.1260806

β_0 estimates are lowest for the Gaussian weighting function at 11 of the 16 locations and are highest for the binary weighting function at 15 of the 16 locations. The range of the standard deviations of $\widehat{\beta}_0$ among the 16 locations is small for the Gaussian and adaptive bi-square weighting functions (between 0.12 and 0.15). For the fixed bi-square weighting function, the standard deviations range between 0.12 and 0.19, while for the binary weighting function, the standard deviations among the 16 locations range between 0.12 and 0.32.

In contrast to the results obtained for $\widehat{\beta}_0$, the means of the GWPR estimates of β_1 (for all 4 models) are negatively biased for locations 1, 2, 5, 6, 9, 10, 13 and 14

Table 3.4: Summary statistics of GWPR parameter estimates: Fixed bi-square weighting function with $d = 3.5$

Points	Intercept ($\widehat{\beta_0}$) Mean	St. Dev.	Slope ($\widehat{\beta_1}$) Mean	St. Dev.
1	1.500245	0.1624557	-0.037916343	0.1605247
2	1.489722	0.1476723	-0.010770129	0.1596105
3	1.455533	0.1479899	-0.008787483	0.1749742
4	1.361697	0.1873747	-0.043775056	0.2347639
5	1.658434	0.1307112	-0.051191609	0.1262712
6	1.637787	0.1217927	-0.017829236	0.1290507
7	1.606772	0.1223666	-0.004011229	0.1415448
8	1.527697	0.1513848	-0.032877975	0.1842759
9	1.800755	0.1272578	-0.077552896	0.1192082
10	1.763157	0.1199444	-0.024454625	0.1237750
11	1.735582	0.1207566	0.006792435	0.1361976
12	1.678522	0.1475618	-0.004204512	0.1761175
13	1.925289	0.1408901	-0.089790020	0.1308413
14	1.883737	0.1318211	-0.028209851	0.1343259
15	1.856799	0.1330906	0.011770547	0.1482595
16	1.801080	0.1644313	0.001946380	0.1960232
Global statistic	1.657375	0.1129078	-0.01274097	0.1260806

(i.e., they underestimate their corresponding true values) while they are positively biased for the other 8 locations. As with the case for the estimates of β_0, the bias in the estimates of β_1 is quite large. In comparing the results between the 4 models, the GWPR estimates of β_1 using the Gaussian weighting function have the highest bias (in absolute value) for most of the first 8 locations and the lowest bias for most of the last 8 locations, while the opposite result occurs for the binary weighting function. The standard deviations of the β_1 estimates are lowest for the Gaussian weighting function at 10 of the 16 locations and are highest for the binary weighting function at 15 of the 16 locations, which is similar to the results obtained for the β_0 estimates.

Table 3.5: Summary statistics of GWPR parameter estimates: Adaptive bi-square weighting function with $M = 15$

	Intercept ($\widehat{\beta}_0$)		Slope ($\widehat{\beta}_1$)	
Points	Mean	St. Dev.	Mean	St. Dev.
1	1.553689	0.1359770	-0.021886853	0.1409639
2	1.498833	0.1440187	-0.010717806	0.1562663
3	1.466105	0.1443894	-0.010173779	0.1702049
4	1.469081	0.1422187	-0.026885850	0.1733224
5	1.659976	0.1287166	-0.048569923	0.1252412
6	1.624235	0.1352787	-0.019565822	0.1372849
7	1.577505	0.1384238	0.001434572	0.1633140
8	1.540599	0.1454307	-0.026647651	0.1768868
9	1.793438	0.1258073	-0.072648944	0.1188958
10	1.818389	0.1287063	-0.031286212	0.1259716
11	1.778821	0.1322400	0.018978195	0.1501063
12	1.682787	0.1421727	0.001316782	0.1697950
13	1.848668	0.1277117	-0.062682696	0.1251655
14	1.872723	0.1302487	-0.027578990	0.1332878
15	1.847126	0.1313918	0.011007915	0.1462333
16	1.777592	0.1341994	0.019048616	0.1571813
Global statistic	1.657375	0.1129078	-0.01274097	0.1260806

Also, the range of the standard deviations of $\widehat{\beta}_1$ among the 16 locations is small for the Gaussian and adaptive bi-square weighting functions (between 0.12 and 0.19). For the fixed bi-square weighting function, the standard deviations range between 0.12 and 0.24, while for the binary weighting function, the standard deviations among the 16 locations range between 0.13 and 0.37. In comparing the estimates of the two regression coefficients, the average $\widehat{\beta}_1$ values have a greater departure from their true values at 10 of the 16 locations (for all 4 models), while the average $\widehat{\beta}_0$ values display a greater departure from their true values at locations 1, 3, 5, 7, 9 and 11.

The last row of Tables 3.2 to 3.5 show the means and standard deviations of $\widehat{\beta}_0$

(mean = 1.657, standard deviation = 0.113) and $\widehat{\beta}_1$ (mean = -0.0127, standard deviation = 0.126) obtained from the Poisson regression model assuming the regression coefficients are stationary over the study region. Since no variation in the parameter values are assumed in the calibration process, this model is referred to as the global regression model and the statistics computed from this model are referred to as global statistics.

Figures 3.2 to 3.5 display the distribution of the $\widehat{\beta}_0$ values at locations 1, 6, 11 and 16 for the binary, Gaussian, fixed bi-square and adaptive bi-square weighting functions respectively. Figure 3.6 to 3.9 show the distribution of the $\widehat{\beta}_1$ values at the same locations for the 4 spatial weighting functions. For the binary weighting function, the shape of the distribution of $\widehat{\beta}_0$ is slightly skewed to the left at all 4 locations while the distribution of $\widehat{\beta}_1$ is slightly skewed to the right at location 1 and somewhat close to normal at locations 6 and 11. For the Gaussian, fixed bi-square and adaptive bi-square weighting functions, the distributions of both $\widehat{\beta}_0$ and $\widehat{\beta}_1$ are fairly symmetric and close to normal at all 4 locations. These plots also show the bias (which is quite large) in the GWPR estimates of $\widehat{\beta}_0$ and $\widehat{\beta}_1$ in all cases.

Figures 3.10 and 3.11 show line diagrams for the average estimates of β_0 and β_1 obtained using the 4 spatial weighting functions and their corresponding true values at the 16 points on the grid.

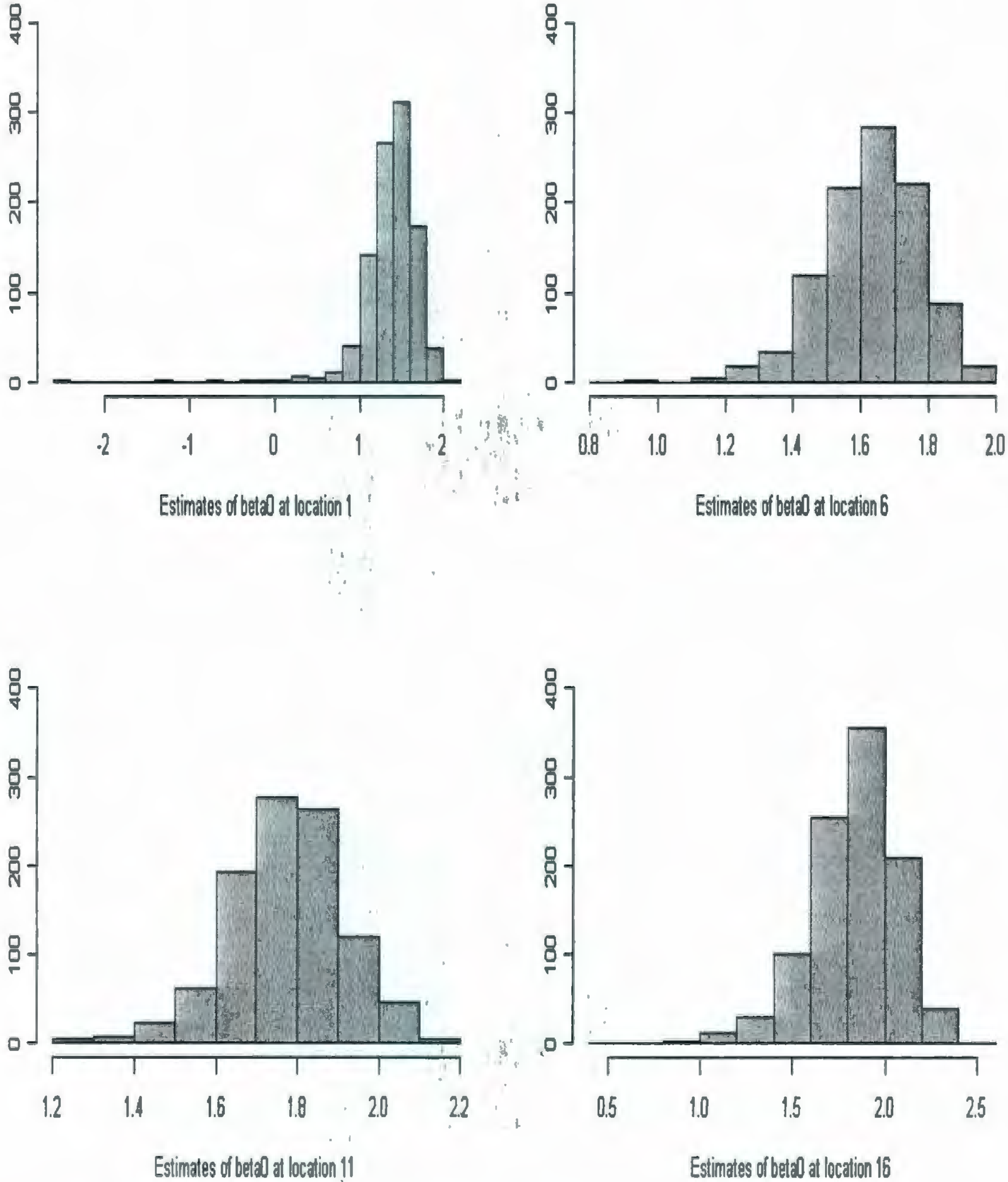

Figure 3.2: Single Predictor GWPR Models: Distribution of $\widehat{\beta}_0$ at various points on the grid - Binary weighting function

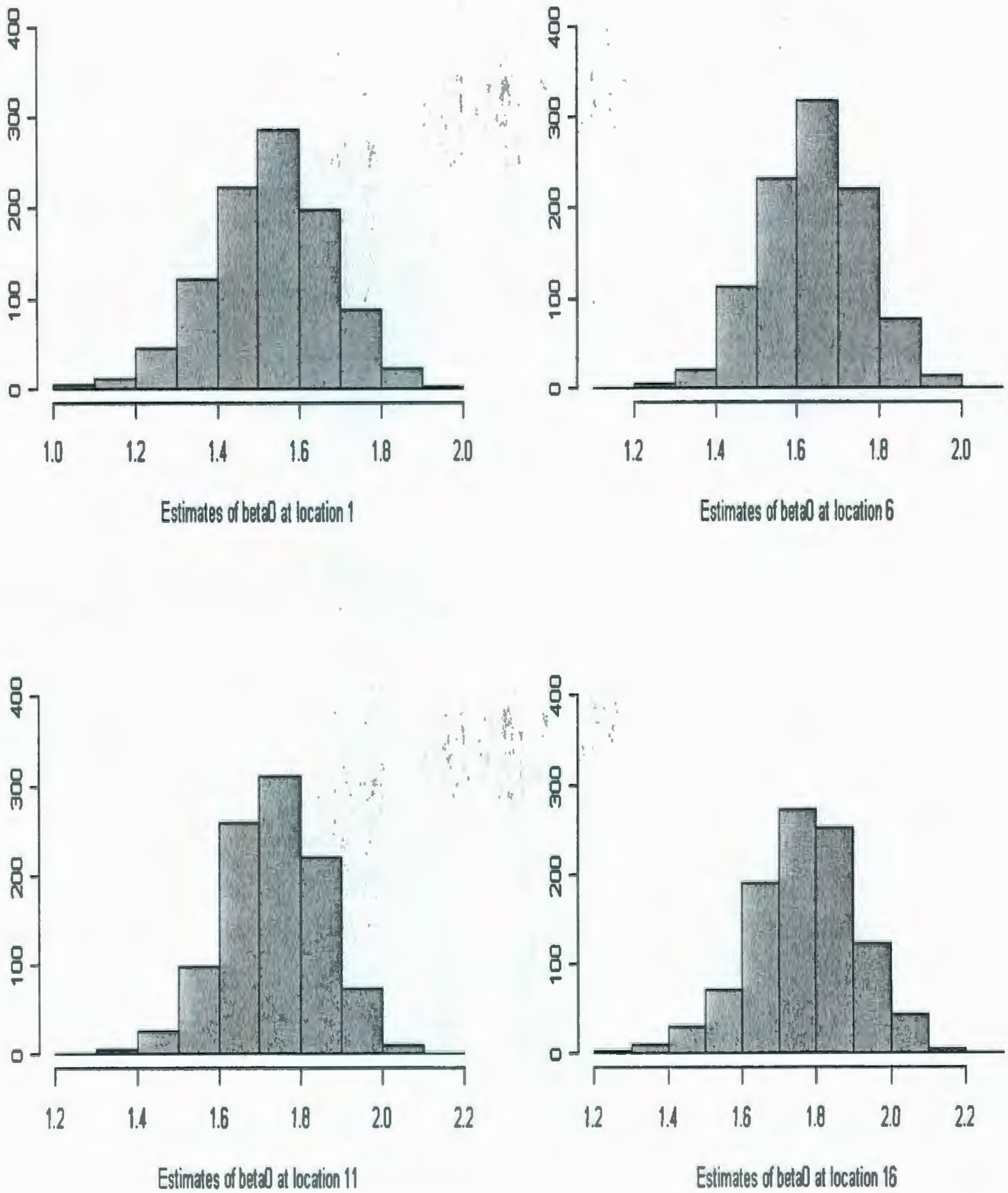

Figure 3.3: Single Predictor GWPR Models: Distribution of $\widehat{\beta}_0$ at various points on the grid - Gaussian weighting function

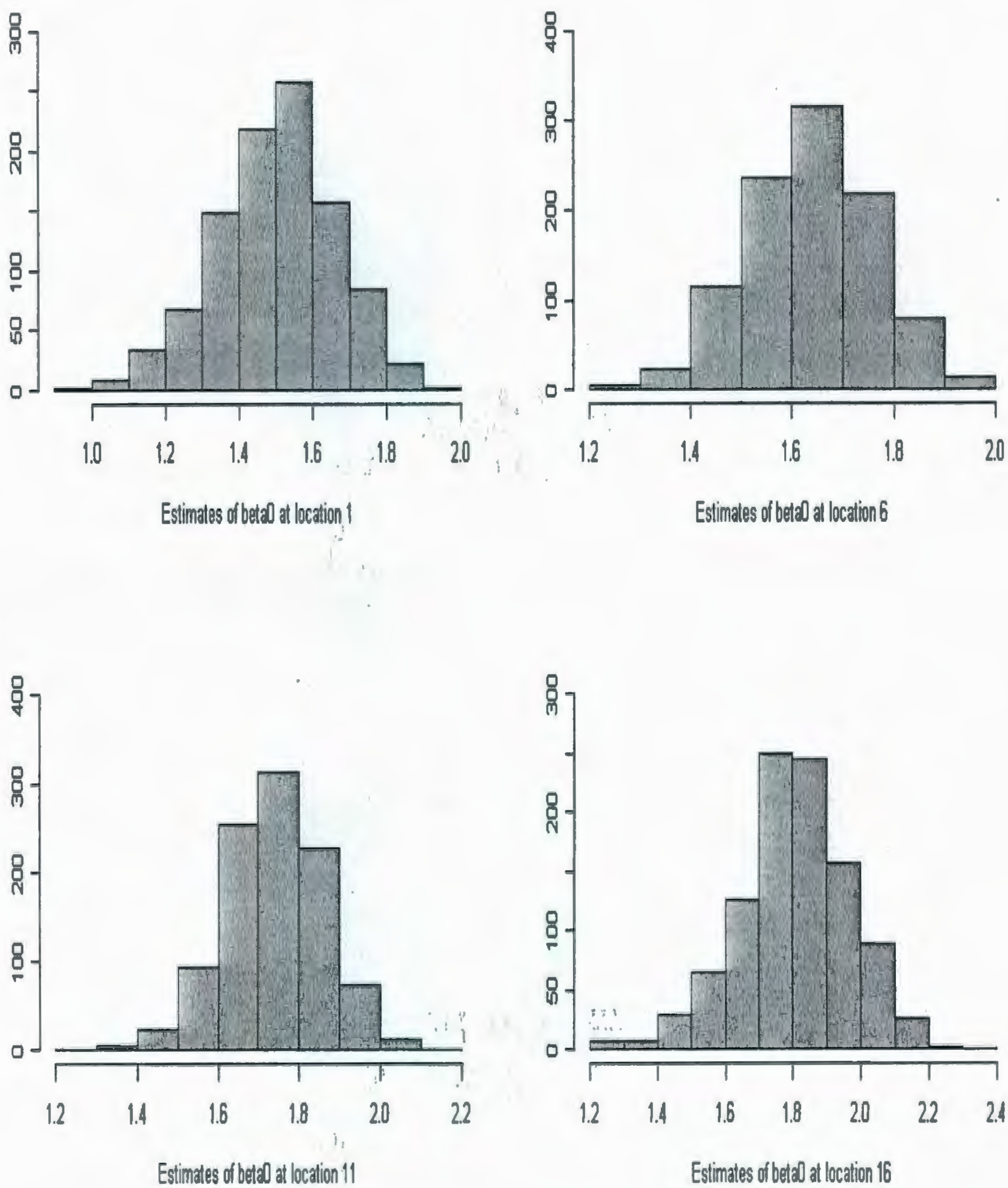

Figure 3.4: Single Predictor GWPR Models: Distribution of $\widehat{\beta_0}$ at various points on the grid - Fixed bi-square weighting function

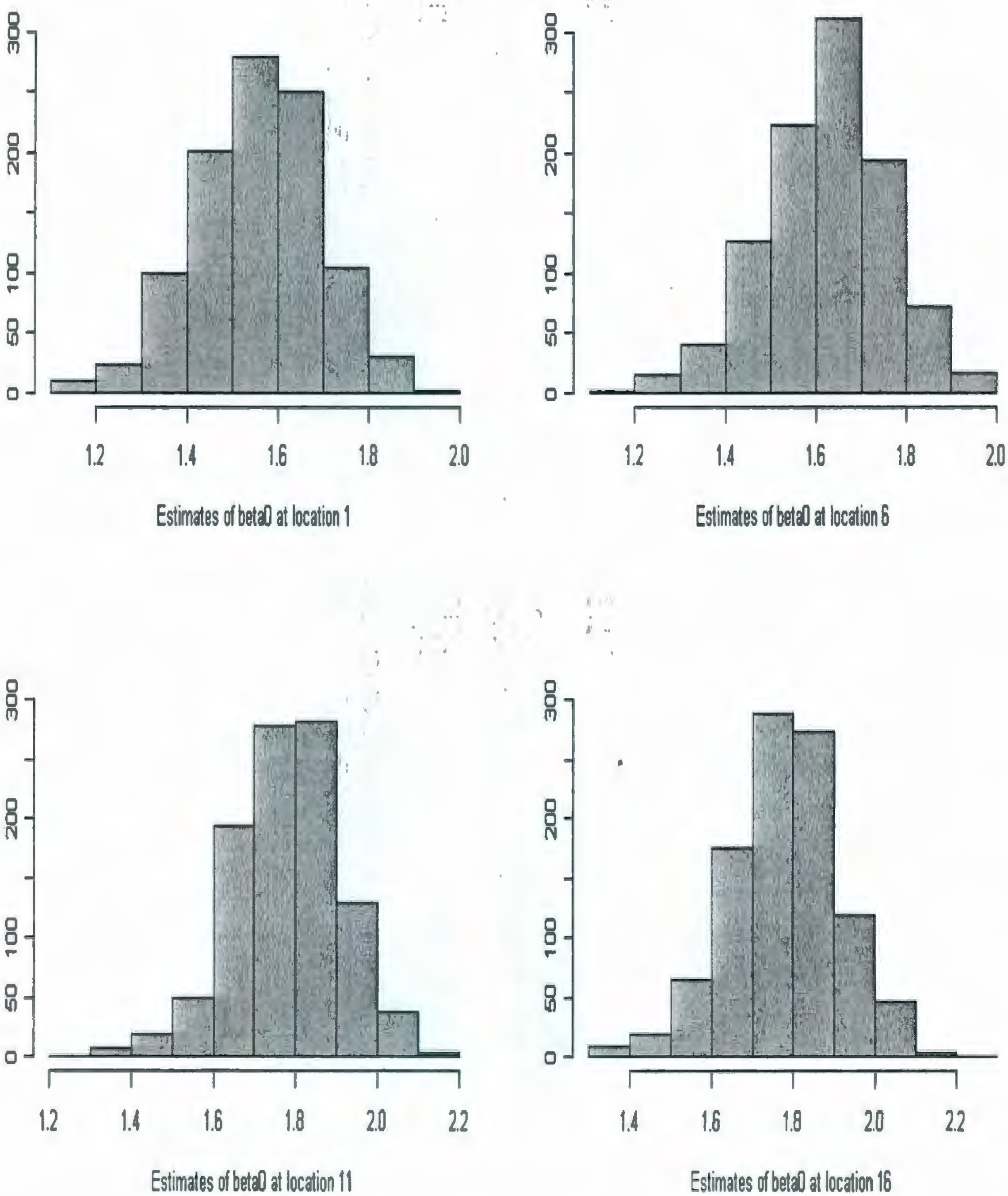

Figure 3.5: Single Predictor GWPR Models: Distribution of $\widehat{\beta}_0$ at various points on the grid - Adaptive bi-square weighting function

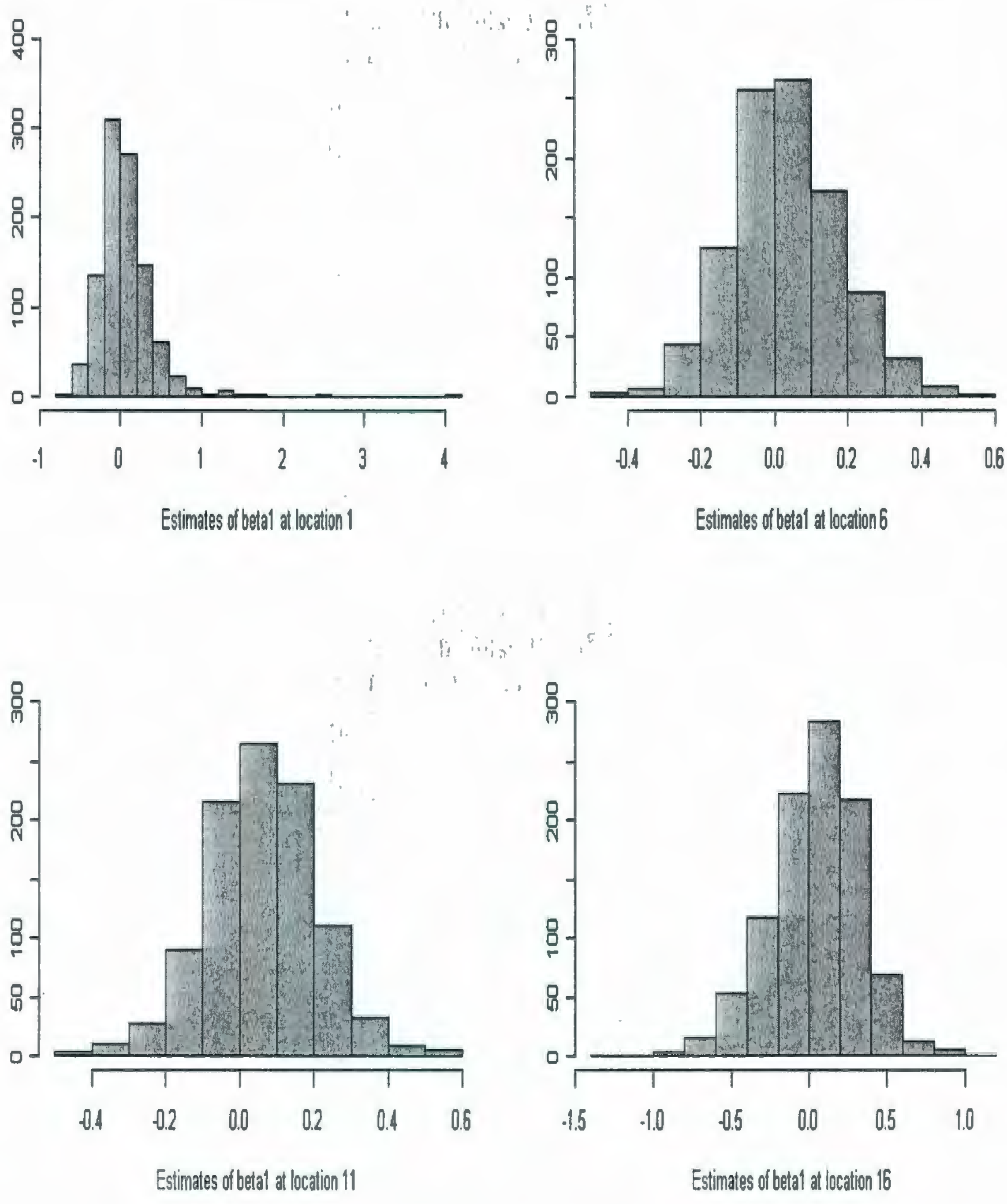

Figure 3.6: Single Predictor GWPR Models: Distribution of $\widehat{\beta}_1$ at various points on the grid - Binary weighting function

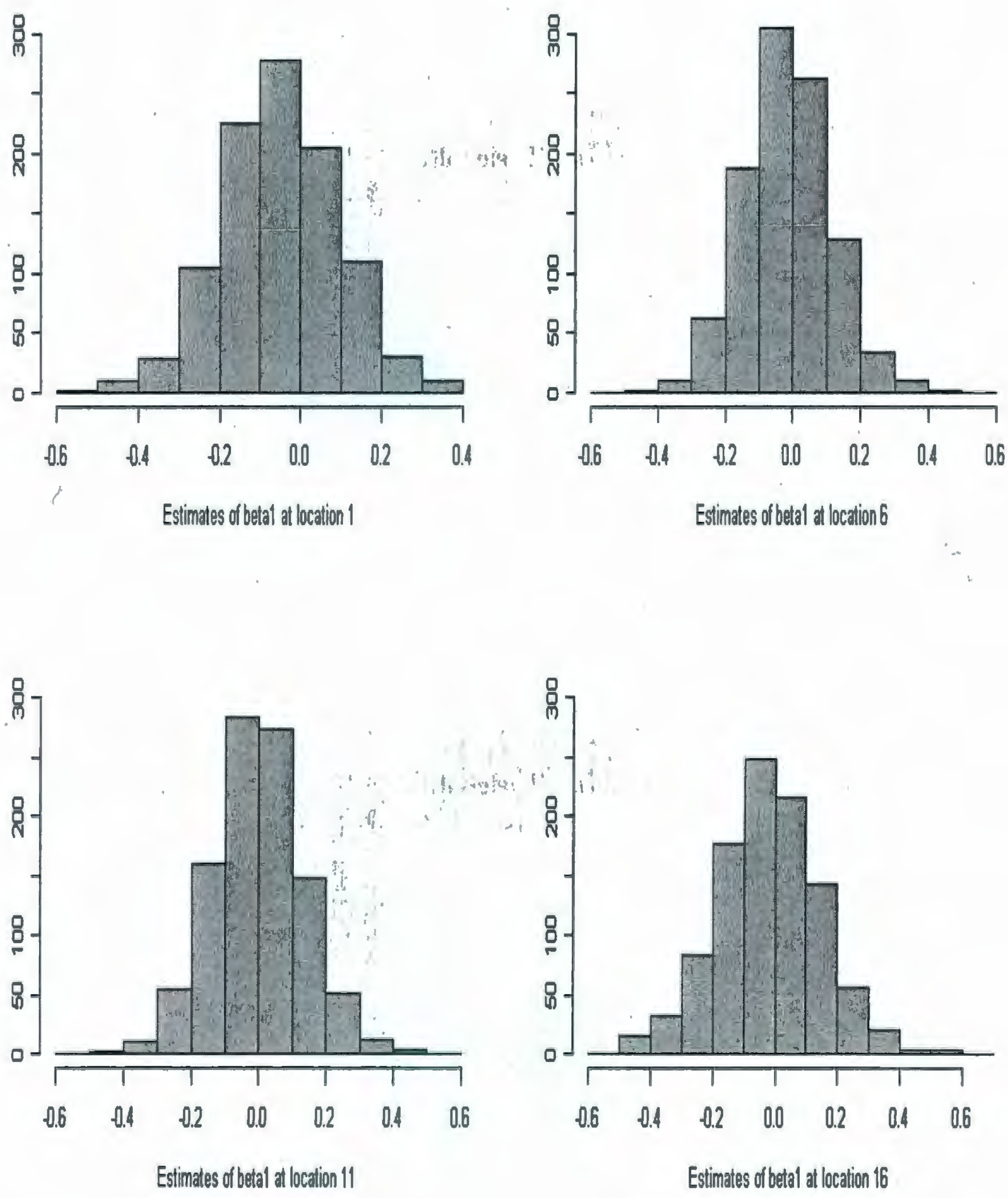

Figure 3.7: Single Predictor GWPR Models: Distribution of $\widehat{\beta}_1$ at various points on the grid - Gaussian weighting function

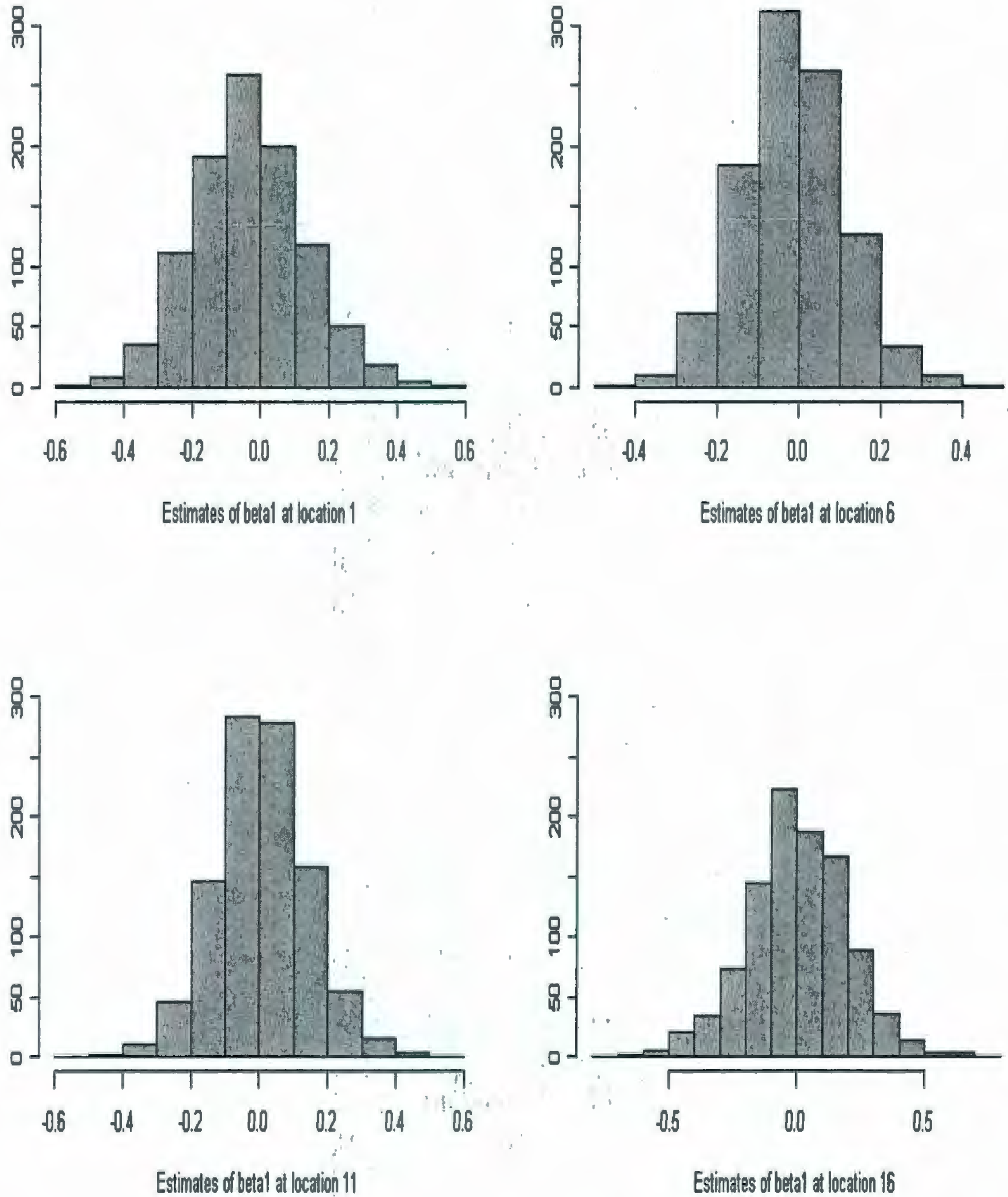

Figure 3.8: Single Predictor GWPR Models: Distribution of $\widehat{\beta}_1$ at various points on the grid - Fixed bi-square weighting function

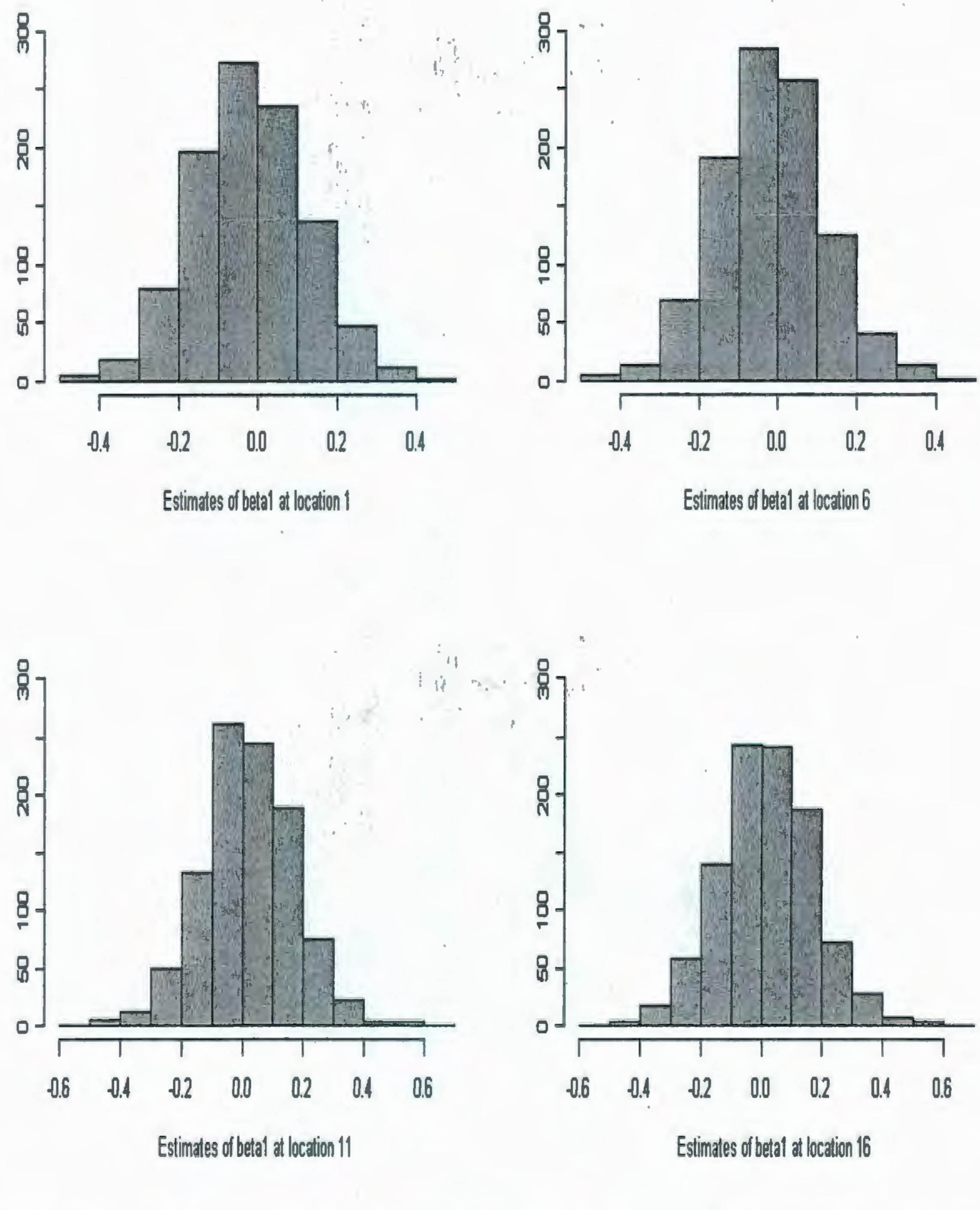

Figure 3.9: Single Predictor GWPR Models: Distribution of $\widehat{\beta}_1$ at various points on the grid - Adaptive bi-square weighting function

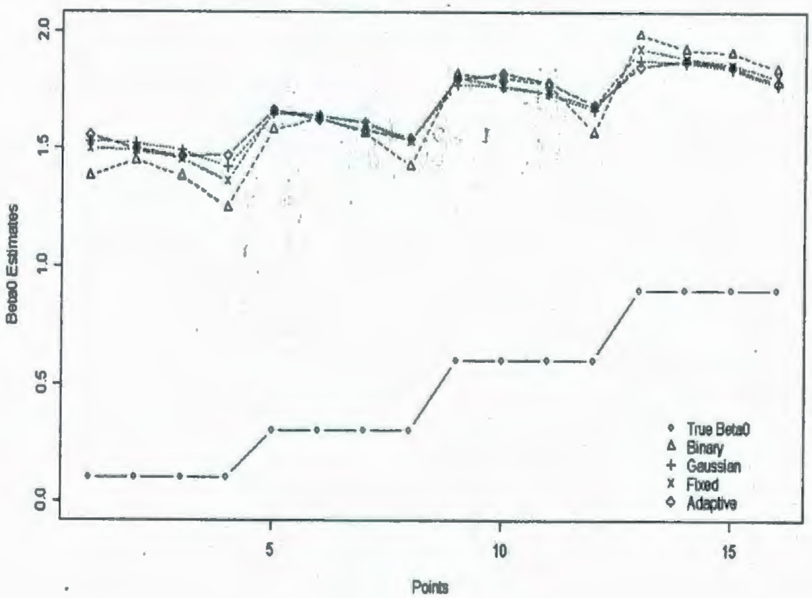

Figure 3.10: Single Predictor GWPR Models: Pattern of β_0 estimates - All spatial weighting functions

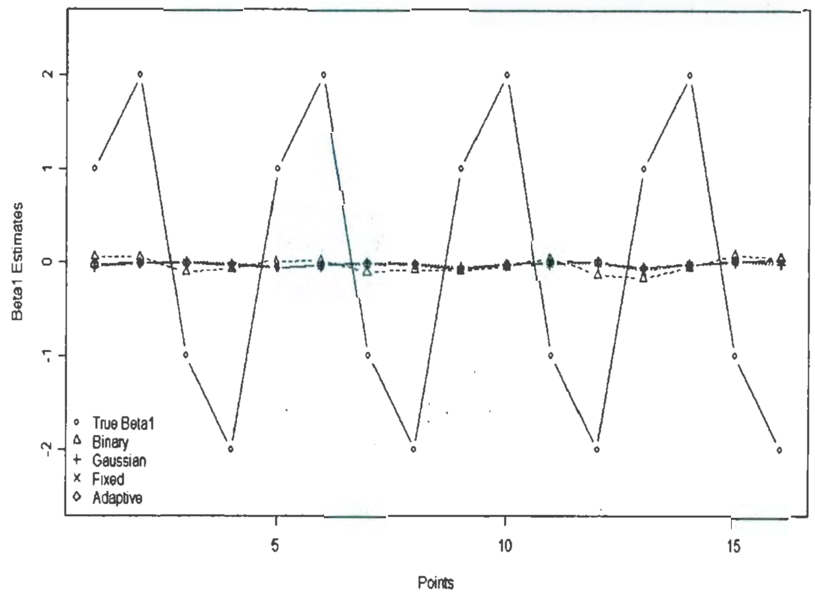

Figure 3.11: Single Predictor GWPR Models: Pattern of β_1 estimates - All spatial weighting functions

3.2.2 Hypothesis Testing

As mentioned in Section 2.6, testing for significance in the regression coefficients of a GWPR model can be carried out using local t-statistics, given by equation (2.28). To assess the performance of this testing procedure for the single predictor GWPR model, local t-tests were applied to the same simulated data generated using equations (3.1) and (3.2). The hypotheses to be tested at the i^{th} data point ($i = 1, \cdots, 16$) are

$$H_0 : \beta_{1i} = 0$$
$$H_1 : \beta_{1i} \neq 0$$

Since for the simulated data the true regression parameter values are known, the power of the local t-tests can be determined. The power of a statistical test is the probability of rejecting the null hypothesis H_0 when it is actually false and should be rejected.

For the single predictor case, the local t-tests had to be carried out 16,000 times for each GWPR model since the estimates of β_{1i} for the 16 data points were computed from each of the 1000 simulated datasets. Since the null distribution of the local t-statistic is approximated by a t-distribution with $n - K$ degrees of freedom (K is the effective number of parameters), the corresponding p-values can be obtained. All 4 spatial weighting functions described earlier were used in the analysis, where the values of b, d and M used were the same as those listed in Table 3.1. Table 3.6 provides a summary of the results from the local t-tests at the 5% level of significance ($\alpha = 0.05$).

Table 3.6: Single Predictor GWPR Models: Summary of Results from Local t-Tests for Testing $H_0 : \beta_{1i} = 0$

Weighting Function		# of Times H_0 Rejected	Power
Binary	$d = 2.0$	943	0.06
Binary	$d = 2.5$	893	0.06
Binary	$d = 3.0$	805	0.05
Binary	$d = 3.5$	832	0.05
Binary	$d = 4.0$	654	0.04
Gaussian	$b = 1.0$	703	0.04
Gaussian	$b = 1.5$	657	0.04
Gaussian	$b = 2.0$	622	0.04
Gaussian	$b = 3.0$	634	0.04
Gaussian	$b = 5.0$	632	0.04
Fixed	$d = 2.5$	895	0.06
Fixed	$d = 3.0$	774	0.05
Fixed	$d = 3.5$	723	0.05
Fixed	$d = 4.0$	690	0.04
Fixed	$d = 5.0$	638	0.04
Adaptive	$M = 15$	695	0.04
Adaptive	$M = 14$	723	0.05
Adaptive	$M = 12$	746	0.05
Adaptive	$M = 11$	746	0.05
Adaptive	$M = 9$	1161	0.07

As shown in Table 3.6, the local t-test does not appear to perform well for the single predictor case. The number of times that the null hypothesis is rejected (i.e., the correct decision) at the 5% level of significance is very small for each of the GWPR models, ranging between 600 and 1,200 (out of 16,000). As a result, the power of the local t-test is very low (ranging between 0.04 and 0.07). However, there appears to be some sort of relationship between the power of the test and the value of b, d and M; for the binary and fixed bi-square weighting functions, as the value of the distance d decreases, the power of the local t-test increases; for the Gaussian weighting function, the power of the test increases as the bandwidth b decreases; and for the adaptive bi-square weighting function, the power of the test increases as the value of M decreases. To summarize, for GWPR models with a single predictor variable, the power of the local t-test is very low regardless of the spatial weighting function used. However, the power improves slightly if the bandwidth/distance/value of M is chosen such that, in the calibration of a GWPR model around the i^{th} regression point, other surrounding points have a smaller amount of influence on the regression parameters for that regression point.

3.3 Multiple Predictor GWPR Model

The general form of a GWPR model for a response variable y and k predictor variables $x_1, x_2, \ldots, x_k$ is given by

$$y_i = Poisson(\mu_i) \tag{3.3}$$

where

$$\mu_i = \exp(\beta_{0i} + \beta_{1i}x_{1i} + \cdots + \beta_{ki}x_{ki}) \tag{3.4}$$

and where $i = 1, \cdots, n$.

For the multiple predictor case, we chose $k = 3$ and set $m = 6$, so there are $n = m^2 = 36$ observations in the study region. The values for the regression coefficients (β_{0i}, β_{1i}, β_{2i} and β_{3i}) and two of the predictor variables (x_{1i} and x_{2i}) were chosen using the following step-changing approach:

$$\beta_{0i} = \begin{cases} 0.1, & \text{for } i = 1, \ldots, 6 \\ 0.2, & \text{for } i = 7, \ldots, 12 \\ 0.3, & \text{for } i = 13, \ldots, 18 \\ 0.7, & \text{for } i = 19, \ldots, 24 \\ 0.8, & \text{for } i = 25, \ldots, 30 \\ 0.9, & \text{for } i = 31, \ldots, 36 \end{cases}$$

$$\beta_{1i} = \begin{cases} 1, & \text{for } i = 1, \ldots, 18 \\ -1, & \text{for } i = 19, \ldots, 36 \end{cases}$$

$$\beta_{2i} = \begin{cases} 3, & \text{for } i = 1, \ldots, 18 \\ -3, & \text{for } i = 19, \ldots, 36 \end{cases}$$

$$\beta_{3i} = \begin{cases} 2, & \text{for } i = 1, \ldots, 18 \\ -2, & \text{for } i = 19, \ldots, 36 \end{cases}$$

$$x_{1i} = \begin{cases} 0.7, & \text{for } i = 1, \ldots, 6 \\ 0.6, & \text{for } i = 7, \ldots, 12 \\ 0.5, & \text{for } i = 13, \ldots, 18 \\ -1.25, & \text{for } i = 19, \ldots, 24 \\ -0.8, & \text{for } i = 25, \ldots, 30 \\ -1, & \text{for } i = 31, \ldots, 36 \end{cases}$$

$$x_{2i} = \begin{cases} -0.05, & \text{for } i = 1, \ldots, 9 \\ 0.17, & \text{for } i = 10, \ldots, 18 \\ 0.1, & \text{for } i = 19, \ldots, 27 \\ -0.08, & \text{for } i = 28, \ldots, 36 \end{cases}$$

The values for the predictor variable x_{3i} were randomly generated from a uniform distribution on the interval (0, 0.5) for points 1 to 18 and from a uniform distribution on the interval (-0.2, 0.2) for points 19 to 36. Similar to the single predictor case, the values for μ_i, $i = 1, \cdots, 36$, were calculated using equation (3.4) and the chosen values for the regression coefficients and predictor variables. The values of the response variable y_i, $i = 1, \cdots, 36$, were randomly generated from a Poisson distribution using equation (3.3) - this process was also carried out 1000 times in order to obtain the simulated datasets that were used in the analysis for the multiple predictor case.

3.3.1 Estimates of Regression Coefficients

The estimates of β_{0i}, β_{1i}, β_{2i} and β_{3i} for the 36 data points were computed from each of 1000 simulated datasets using the IRLS method described in Chapter 2 (the initial estimated values of the regression coefficients were obtained from fitting a global Poisson regression model to the data and were updated using the IRLS method until convergence was reached). As with the single predictor case, a number of GWPR models were calibrated using a variety of bandwidth values (Gaussian), distances (fixed bi-square) and M values (adaptive bi-square). Table 3.7 shows the average AIC_c and average CV scores for a number of GWPR models that were calibrated for the three-predictor case.

The minimum average AIC_c values occur at $b = 2.0$ for the Gaussian weighting function, $d = 5.0$ for the fixed bi-square weighting function and $M = 31$ for the adaptive bi-square weighting function. With the exception of the model calibrated using the Gaussian weighting function with $b = 1.0$, the average AIC_c values for the

Table 3.7: Three-Predictor GWPR Models: Average AIC_c Values and Average CV Scores

Weighting Function		Average AIC_c	Average CV Score
Gaussian	$b = 1.0$	90.63627	4369800
Gaussian	$b = 1.5$	62.27233	145.7924
Gaussian	$b = 2.0$	58.75471	95.35532
Gaussian	$b = 2.5$	59.55068	96.98658
Gaussian	$b = 3.0$	60.95578	98.46995
Gaussian	$b = 5.0$	64.59212	101.8378
Fixed	$d = 3.5$	63.6522	30141770
Fixed	$d = 4.0$	59.63074	60275.03
Fixed	$d = 4.5$	58.28422	114.1461
Fixed	$d = 5.0$	58.22338	93.29386
Fixed	$d = 5.5$	58.6963	94.8348
Fixed	$d = 6.0$	59.3653	96.15068
Adaptive	$M = 35$	58.40722	93.5373
Adaptive	$M = 34$	57.40191	91.2117
Adaptive	$M = 33$	57.37232	91.36778
Adaptive	$M = 32$	57.02166	91.01119
Adaptive	$M = 31$	56.91001	104.9112
Adaptive	$M = 30$	57.26042	104.4442
Adaptive	$M = 25$	59.35501	1539372
Adaptive	$M = 21$	63.06902	1536233
Global		67.67622	104.80036

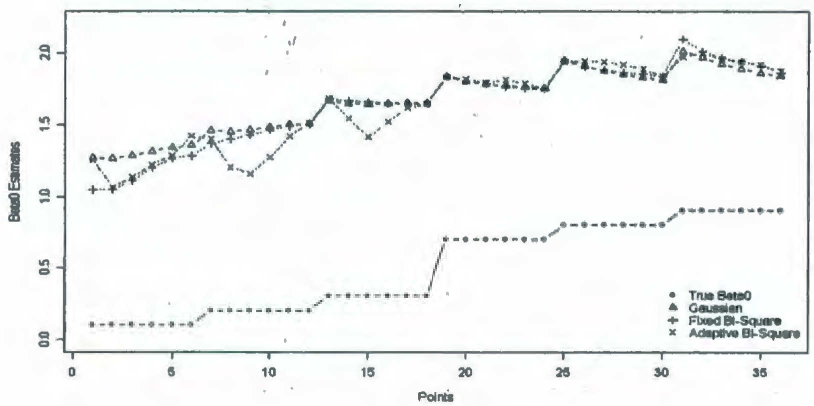

Figure 3.12: Three-Predictor GWPR Models: Pattern of β_0 estimates - Gaussian and bi-square weighting functions

GWPR models shown in Table 3.7 are lower than the average AIC_c for the global Poisson regression model. The minimum average CV scores occur at the same value of b and d for the Gaussian and fixed bi-square weighting functions, while the minimum average CV score for the adaptive bi-square weighting function occurs at $M = 32$.

Figures 3.12 through 3.15 show line diagrams for the estimates of β_0, β_1, β_2 and β_3 obtained using the Gaussian and bi-square (fixed and adaptive) weighting functions and their corresponding true values at the 36 points on the grid.

The means of the GWPR estimates of β_0 overestimate their corresponding true values by quite a large margin at all 36 locations for the Gaussian and both bi-square weighting functions. While the estimates appear to differ between the 3 models for the 1st 18 points, they appear to be reasonably close to each other for the last 18 points. For the Gaussian and fixed bi-square functions, the estimates follow an increasing trend for points 1 to 12, and appear to level off between points 13 and 18. For the adaptive bi-square weighting function, the estimates appear to follow a wave-like

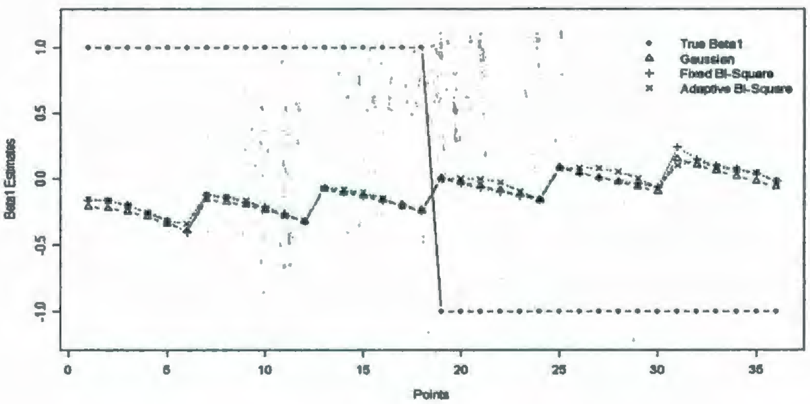

Figure 3.13: Three-Predictor GWPR Models: Pattern of β_1 estimates - Gaussian and bi-square weighting functions

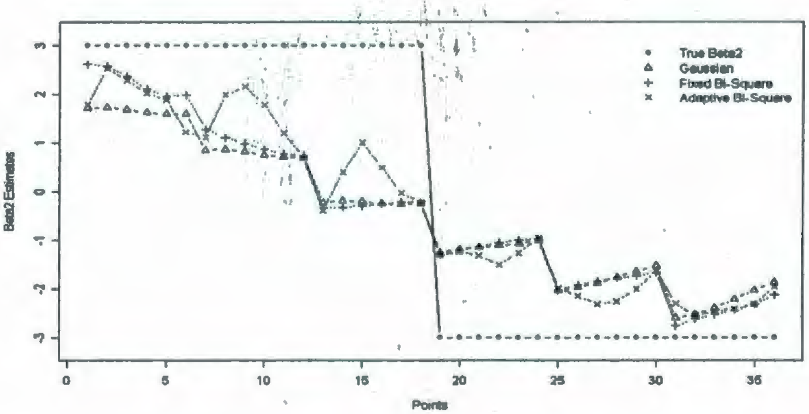

Figure 3.14: Three-Predictor GWPR Models: Pattern of β_2 estimates - Gaussian and bi-square weighting functions

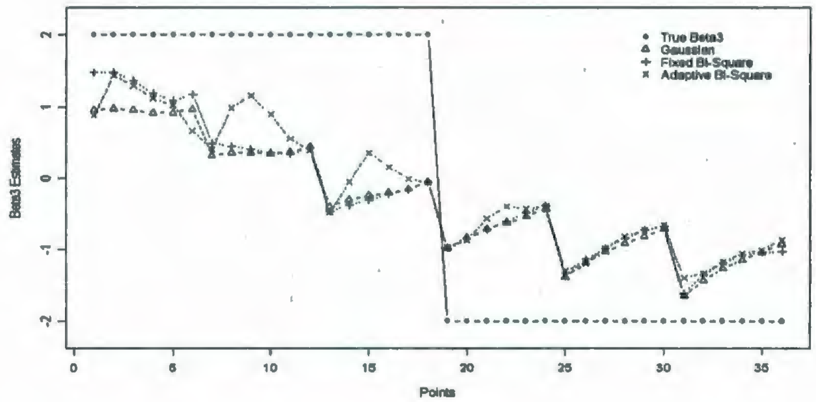

Figure 3.15: Three-Predictor GWPR Models: Pattern of β_3 estimates - Gaussian and bi-square weighting functions

pattern between points 1 and 18. For points 19 to 36, the estimates appear to follow a step-changing pattern for all 3 models, with increases occurring between points 18 and 19, points 24 and 25 and points 29 and 30. Between points 19 and 24, points 25 and 30 and points 31 and 36, the estimates follow a decreasing trend.

The means of the GWPR estimates of β_1 for the 3 spatial weighting functions underestimate their corresponding true values for points 1 to 18 (where $\beta_1 = 1$) and overestimate their corresponding true values for points 19 to 36 (where $\beta_1 = -1$). In comparing the results among the 3 weighting functions at each of the 36 locations, the estimates appear to be very close to each other and follow a similar step-changing pattern that was seen in the plot of the average estimates of β_0. Increases occur between the step-changing points (i.e., between points 6 and 7; points 12 and 13; points 18 and 19; points 24 and 25; and between points 30 and 31). The estimates follow a decreasing trend between the points where the step-changes occur (i.e., from points 1 to 6; points 7 to 12; points 13 to 18; points 19 to 24; points 25 to 30; and from points 31 to 36).

Similar to the results for β_1, the average GWPR estimates of β_2 for the 3 models underestimate their corresponding true values for points 1 to 18 (where $\beta_2 = 3$) and overestimate their corresponding true values for points 19 to 36 (where $\beta_2 = -3$). The estimates also follow a step-changing pattern, but with decreases occuring between the step-changing points. The estimates obtained from the models calibrated with the Gaussian and fixed bi-square weighting functions follow the same pattern and the average estimates of β_2 for these models appear to be very close to each other at most of the 36 points. The estimates from the 2 models follow a decreasing trend between the step-changing points for points 1 to 12, appear to level off between points 13 and 18, and then follow an increasing trend between the step-changing points for points 19 to 36. The estimates from the model calibrated with the adaptive bi-square weighting function appear to follow a wave-like pattern (similar to what was shown for this model for β_0). It should also be noted that, for points 1 to 18, the average estimates for β_2 are closer to their corresponding true values for the first 6 points and are furthest from their true values for the last 6 points. However, for points 19 to 36, the estimates are furthest away from their true values for the first 6 points and are closer to their true values for the last 6 points.

Similar to what was shown for β_1 and β_2, the average GWPR estimates of β_3 for the 3 models also underestimate their true values for points 1 to 18 (where $\beta_3 = 2$) and overestimate their true values for points 19 to 36 (where $\beta_3 = -2$). Similar to the results shown for β_2, the estimates also follow a step-changing pattern, with decreases occuring between the step-changing points. With the exception of the first 6 points, the estimates obtained from the models calibrated with the Gaussian and fixed bi-square weighting functions follow the same pattern and are close to each other at each point. The average estimates of β_3 appear to be constant between points 7 and 12 for the 2 models and follow an increasing trend between the step-changing points for points 13 to 36. The estimates from the model calibrated with the adatptive bi-square weighting function appear to follow a wave-like pattern for the first 18 points and then follow a pattern similar to the estimates obtained from the Gaussian and fixed bi-square weighting functions for the last 18 points.

The mean and standard deviation of the regression coefficient estimates for the Gaussian (bandwidth $b = 2.0$), fixed bi-square ($d = 5.0$) and adaptive bi-square ($M = 31$) weighting functions are shown in the Appendix.

3.3.2 Hypothesis Testing

To assess the performance of this testing procedure for the multiple predictor GWPR model, local t-tests were applied to the simulated data generated using equations (3.3) and (3.4). The hypotheses to be tested at the i^{th} data point ($i = 1, \cdots, 36$) are

$$H_0 : \beta_{1i} = 0$$
$$H_1 : \beta_{1i} \neq 0$$

$$H_0 : \beta_{2i} = 0$$
$$H_1 : \beta_{2i} \neq 0$$

$$H_0 : \beta_{3i} = 0$$
$$H_1 : \beta_{3i} \neq 0$$

For the multiple predictor case, the local t-tests had to be carried out 36,000 times for each GWPR model since the estimates of β_{1i}, β_{2i} and β_{3i} for the 36 data points were computed from each of 1000 simulated datasets. The Gaussian and both bi-square weighting functions (fixed and adaptive) were used in the analysis, where the values of b, d and M used were the same as those listed in Table 3.7. Tables 3.8, 3.9 and 3.10 provide a summary of the results from the local t-tests for testing the null hypothesis that β_{1i}, β_{2i} and β_{3i} are equal to 0 respectively. All tests were carried out at the 5% level of significance ($\alpha = 0.05$).

As shown in Tables 3.8, 3.9 and 3.10, the local t-test appears to perform better for the three-predictor case, compared to the single predictor case. The number of times that the null hypothesis is rejected for β_1 and β_3 (i.e., the correct decision) at the 5% level of significance is small for each of the GWPR models, ranging between 4000 and 7000 (out of 36,000). As a result, the power of the local t-test is low (ranging between 0.12 and 0.20). However, there appears to be an improvement in the results from the local t-test for β_2. The number of times that the null hypothesis is rejected at $\alpha = 0.05$ is higher, ranging between 8000 and 14000 for each of the GWPR models. As a result, the power of the test is higher, ranging between 0.20 and 0.40. It should be noted that the true values of β_2 are further away from zero, compared to the true values for the other regression coefficients.

For the tests involving β_1, the power of the local t-test increases as the values of b, d and M increase (i.e., the power increases when the values of the bandwidth/distance/number of nearest neighbours are chosen such that the calibrated GWPR models become closer to the global model). However, the opposite results occur for the tests involving β_2 and β_3, where the power increases when the values of

Table 3.8: Three-Predictor GWPR Models: Summary of Results from Local t-Tests for Testing $H_0 : \beta_{1i} = 0$

Weighting Function		# of Times H_0 Rejected	Power
Gaussian	$b = 1.0$	4202	0.12
Gaussian	$b = 1.5$	6001	0.17
Gaussian	$b = 2.0$	6413	0.18
Gaussian	$b = 2.5$	6522	0.18
Gaussian	$b = 3.0$	6629	0.18
Gaussian	$b = 5.0$	6838	0.19
Fixed	$d = 3.5$	4844	0.13
Fixed	$d = 4.0$	5591	0.16
Fixed	$d = 4.5$	5813	0.16
Fixed	$d = 5.0$	5808	0.16
Fixed	$d = 5.5$	5836	0.16
Fixed	$d = 6.0$	6045	0.17
Adaptive	$M = 35$	5735	0.16
Adaptive	$M = 34$	5421	0.15
Adaptive	$M = 33$	5405	0.15
Adaptive	$M = 32$	5330	0.15
Adaptive	$M = 31$	5337	0.15
Adaptive	$M = 30$	5319	0.15
Adaptive	$M = 25$	5283	0.15
Adaptive	$M = 21$	5420	0.15

Table 3.9: Three-Predictor GWPR Models: Summary of Results from Local t-Tests for Testing $H_0 : \beta_{2i} = 0$

Weighting Function		# of Times H_0 Rejected	Power
Gaussian	$b = 1.0$	10367	0.29
Gaussian	$b = 1.5$	12132	0.34
Gaussian	$b = 2.0$	10746	0.30
Gaussian	$b = 2.5$	9279	0.26
Gaussian	$b = 3.0$	8449	0.23
Gaussian	$b = 5.0$	8135	0.23
Fixed	$d = 3.5$	13177	0.37
Fixed	$d = 4.0$	13089	0.36
Fixed	$d = 4.5$	12644	0.35
Fixed	$d = 5.0$	11768	0.33
Fixed	$d = 5.5$	10776	0.30
Fixed	$d = 6.0$	9793	0.27
Adaptive	$M = 35$	9648	0.27
Adaptive	$M = 34$	11105	0.31
Adaptive	$M = 33$	11233	0.31
Adaptive	$M = 32$	12122	0.34
Adaptive	$M = 31$	12377	0.34
Adaptive	$M = 30$	12750	0.35
Adaptive	$M = 25$	13778	0.38
Adaptive	$M = 21$	13785	0.38

Table 3.10: Three-Predictor GWPR Models: Summary of Results from Local t-Tests for Testing $H_0 : \beta_{3i} = 0$

Weighting Function		# of Times H_0 Rejected	Power
Gaussian	$b = 1.0$	5335	0.15
Gaussian	$b = 1.5$	6393	0.18
Gaussian	$b = 2.0$	6046	0.17
Gaussian	$b = 2.5$	5680	0.16
Gaussian	$b = 3.0$	5599	0.16
Gaussian	$b = 5.0$	5630	0.16
Fixed	$d = 3.5$	6720	0.19
Fixed	$d = 4.0$	6695	0.19
Fixed	$d = 4.5$	6542	0.18
Fixed	$d = 5.0$	6195	0.17
Fixed	$d = 5.5$	5856	0.16
Fixed	$d = 6.0$	5666	0.16
Adaptive	$M = 35$	5250	0.15
Adaptive	$M = 34$	5451	0.15
Adaptive	$M = 33$	5481	0.15
Adaptive	$M = 32$	5868	0.16
Adaptive	$M = 31$	5990	0.17
Adaptive	$M = 30$	6102	0.17
Adaptive	$M = 25$	6910	0.19
Adaptive	$M = 21$	7001	0.19

b, d and M are chosen such that the calibrated GWPR models have a higher degree of geographical weighting (i.e., data points that are close to a regression point have a strong influence in the estimation of the local regression coefficients for that point, while data points that are further away from the regression point have very little or no influence).

One may feel that the power the local t-test should be higher given the biased estimates that were observed for both the single and multiple predictor cases. However, the variability of the estimates is quite high, which may suggest why the power is low.

Chapter 4

Analysis of a Socio-Economic Dataset Using GWPR

4.1 Introduction

One of the primary objectives of spatial analysis is to identify the nature of relationships that exist among variables (Brunsdon et. al, 1996). The most common type of analysis used to undertake this is regression, where relationships between a response variable and a number of predictor variables are estimated (Fotheringham et. al, 1998). These relationships are often assumed to be constant (or stationary) across space; however, this assumption is usually not verified. In this section, GWPR methods are applied to a well-known socio-economic dataset. After performing an exploratory analysis, a global Poisson regression model is then fit to the data and a suitable subset of predictor variables is chosen using a backwards-elimination stepwise procedure and a drop-in-deviance test. A number of GWPR models are then fit to the data using this subset of predictor variables in order to assess the presence of spatial non-stationarity.

4.2 Data Description

A dataset on housing prices from the Boston Standard Metropolitan Statistical Area (SMSA) is used for illustrating GWPR calibration and inference methods discussed in Chapter 2. Harrison and Rubenfeld (1978) used this dataset to analyze various methodological issues related to hedonic housing prices to estimate the demand for clean air. The hedonic price index is based on the fitted values of a regression of price on various predictor variables and is used to represent its qualitative determinants. The 1970 Census Bureau Publication is the source of the majority of the data. The dataset consists of 506 observations (1 observation per census tract) and a number of predictor variables. These variables include crime rate (CRIM), proportion of area zoned with large lots (ZN), proportion of non-retail business areas (INDUS), location contiguous to the Charles River (CHAS), levels of nitrogen oxides (NOX), average number of rooms (RM), proportion of structures built before 1940 (AGE), weighted distances to the employment centres (DIS), an index of accessibility (RAD), property tax rate (TAX), pupil-teacher ratio (PTRATIO), black population proportion (B) and lower status population proportion (LSTAT). The response variable used in their study was the median value of owner occupied homes in the census tract (MEDV).

The variables RM and AGE can be considered to represent the structural aspect of houses, where RM represents spaciousness and quantity of housing, while AGE is related to housing quality. The variables B, LSTAT, CRIM, ZN, INDUS, TAX, PTRATIO and CHAS are considered to be relating to the neighborhood amenities. In their study, Harrison and Rubenfeld (1978) observed that when the proportion of blacks (B) in the Boston area was low or moderate, an increase in the black population had a negative effect on housing prices, whie the reverse trend occurred when the proportion of back people was very high. Based on this parabolic trend, the variable B is created by shifting the proportion 0.63 towards the origin. The variable LSTAT represents the proportion of lower status people, which is obtained by averaging the proportion of adults without some high school education and the proportion of male workers classified as labourers. The variable INDUS is considered to be a proxy measure of externalities associated with industry, such as noise and heavy traffic. The variable TAX represents full value property taxes and measures the cost of public services in the communities. The local assessment ratio is used as a correction factor with the nominal tax rate and yields the full value tax rate which varies from town to town. The variable PTRATIO is treated as a measure of public sector benefits in the town, where there are better opportunities for a child's education if the value of PTRATIO is low. The variable CHAS is an indicator variable which represents the amenities of a riverside location.

Harrison and Rubenfeld (1978) used this data to investigate the willingness to pay for air quality improvements. Belsley et. al (1980), Krasker et. al (1983), Subramanian and Carson (1988), Brieman and Friedman (1985), Lange and Ryan (1989), Brieman et. al (1993) and Pace (1993) have also used this data for examining robust estimation, normality of residuals, non-parametric and semi-parametric estimation.

Table 4.1: Description of Variables in the Boston Housing Price Dataset

Variable	Description
CMEDV	Corrected median value of owner-occupied homes (in $1,000s)
CRIM	Per capita crime rate by town
ZN	Proportion of residential land zoned for lots over 25000 square feet
INDUS	Proportion of non-retail business acres per town
CHAS	Charles River dummy variable (1 if tract borders the river; 0 otherwise
NOX	Nitric oxides concentration (parts per 10 million)
RM	Average number of rooms per dwelling
AGE	Proportion of owner-occupied units built prior to 1940
DIS	Weighted distances to five Boston employment centres
RAD	Index of accessibility to radial highway
TAX	Full-value property tax rate per $10,000
PTRATIO	Pupil-teacher ratio by town
B	$1000(Bk - 0.63)^2$ where Bk is the proportion of blacks by town
LSTAT	Proportion of lower status of the population
LAT	Standardized Latitude coordinates
LONG	Standardized Longitude coordinates

In addition to the variables mentioned above, Gilley and Pace (1996) collected the location of each census tract in latitude (LAT) and longitude (LON), based on results obtained from the 1970 census. However, in the process of conducting their own study with the data, they discovered some incorrectly coded observations in the MEDV variable (8 in total) as well as a censoring problem with the same variable: they discovered that all observations (16 in total) with a value equal to or greater than $50,000 appeared as $50,000. The variable CMEDV represents the median value of housing prices with these errors corrected. A brief description of the variables is presented in Table 4.1.

As mentioned earlier, the median value of owner occupied homes (i.e., the variables

MEDV or CMEDV) was used as the response variable in most studies which used this dataset for regression analysis. However, in order to fit a Poisson regression model (global or geographically weighted) to this data, another variable would have to be used as the response since MEDV or CMEDV are continuous variables. The variable representing the average number of rooms per dwelling (RM) was chosen as a candidate for the response variable - however, a modification had to be made to this variable since the values were given to 3 decimal places. As a result, the values for RM were rounded to the nearest whole number and stored in a new variable ROOMS. This variable (ROOMS) was used as the response variable in fitting global and geographically weighted Poisson regression models.

4.3 Exploratory Analysis

A key feature of the Poisson distribution is that its variance is equal to its mean. In practice, count observations often exhibit variability exceeding that predicted by the Poisson distribution - this is referred to as overdispersion, which can result from heterogeneity among observations. In such cases, assuming a Poisson distribution for a count variable may be incorrect. When the model for the mean is correct but the true distribution is not Poisson, the maximum likelihood estimates of the regression coefficients are still consistent but their standard errors are incorrect. If there is strong evidence of overdispersion, fitting a negative binomial regression model to the data may be more appropriate. The negative binomial distribution is a related distribution for count data that has an extra parameter to better account for overdispersion by allowing the variance to exceed the mean.

Although overdispersion is common in the modelling of counts, in some cases, the variance in the count data may be smaller than what the Poisson distribution predicts, reflecting underdispersion. For the Boston housing price data, the mean and variance of the ROOMS variable are 6.2668 and 0.5326 respectively, with a variance-to-mean ratio of 0.0850. Also, as shown in Figure 4.1, although the values of ROOMS range from 4 to 9, most observations are clustered near the mean (437 of the 506 observations, or 86%, have a value of either 6 or 7). This indicates that there may be underdispersion which should be taken into account before fitting a Poisson regression model to the data with ROOMS as the response variable.

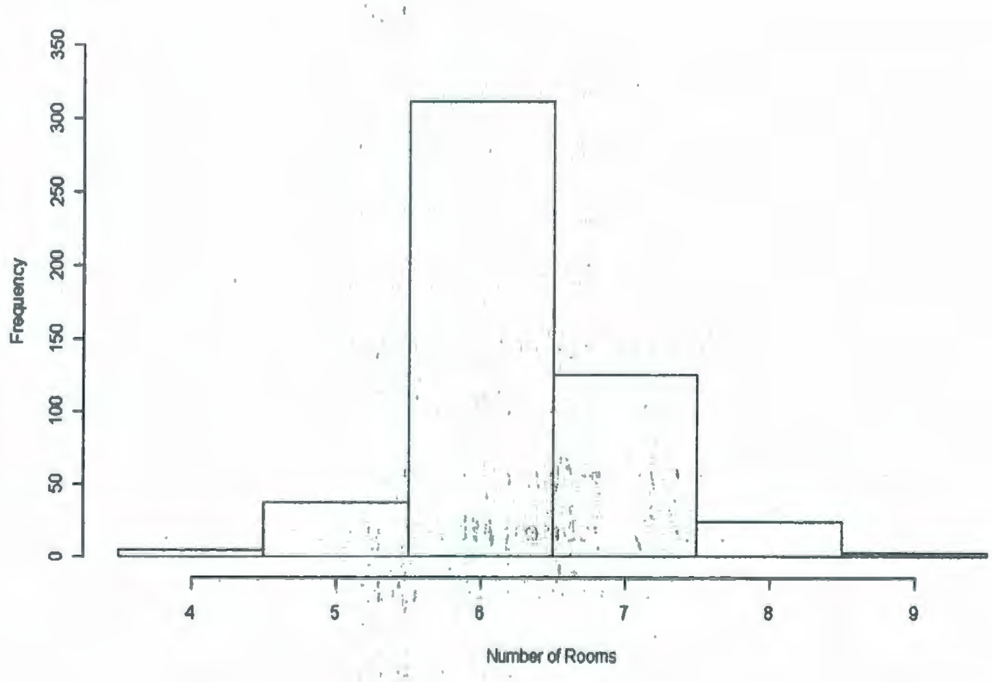

Figure 4.1: Boston Housing Price Data - Histogram of the Variable ROOMS

4.4 Global Poisson Regression Model

In this section, Poisson regression analysis was used in choosing a suitable model for the number of rooms in a dwelling. The initial step in the analysis was to fit a Poisson log-linear model using all possible predictor variables (with the exception of LAT and LON). The backwards-elimination stepwise procedure was then used to remove variables from the model that were not significant (i.e., those variables whose corresponding regression coefficients had the highest p-values). The drop-in-deviance test was then used to determine if the reduced model (i.e., the model using the subset of predictors obtained from the stepwise procedure) provided a better fit than the full model (i.e., the model using all predictors). The subset of predictors in the reduced model was then used for applying GWPR methods and techniques, which will be discussed further in the next section.

Using ROOMS as the response variable, the following global Poisson regression model was fit to the data:

$$\begin{aligned}\log(\mu) &= \beta_0 + \beta_1\text{CMEDV} + \beta_2\text{CRIM} + \beta_3\text{ZN} + \beta_4\text{INDUS} + \beta_5\text{CHAS} + \\ &\quad \beta_6\text{NOX} + \beta_7\text{AGE} + \beta_8\text{DIS} + \beta_9\text{RAD} + \beta_{10}\text{TAX} + \\ &\quad \beta_{11}\text{PTRATIO} + \beta_{12}\text{B} + \beta_{13}\text{LSTAT} \qquad (4.1)\end{aligned}$$

where μ represents the mean number of rooms per dwelling. To account for the possibility of dispersion, a quasipoisson regression model was fit to the data, using the 13 predictors indicated in model (4.1). The estimates and standard errors of the regression coefficients from this model, including their corresponding t-statistics

Table 4.2: Results from the Global Poisson Regression Model (allowing for dispersion) with 13 Predictors

Variable	$\widehat{\beta}$	SE($\widehat{\beta}$)	t-stat	p-value
Intercept	1.8410	0.0780	23.6090	0.0000
CMEDV	0.0055	0.0007	8.1220	0.0000
CRIM	0.0005	0.0006	0.9470	0.3441
ZN	0.0005	0.0002	2.1200	0.0345
INDUS	-0.0025	0.0010	-2.4210	0.0158
CHAS	-0.0027	0.0144	-0.1850	0.8535
NOX	-0.0223	0.0665	-0.3350	0.7379
AGE	0.0008	0.0002	3.8860	0.0001
DIS	-0.0008	0.0035	-0.2330	0.8158
RAD	0.0019	0.0012	1.6020	0.1099
TAX	-0.0001	0.0001	-0.8700	0.3846
PTRATIO	-0.0006	0.0023	-0.2660	0.7906
B	-0.0002	0.0000	-3.9360	0.0001
LSTAT	-0.0057	0.0009	-6.0900	0.0000

and p-values are shown in Table 4.2. As shown in Table 4.4, the estimate of the dispersion parameter and the ratio of the deviance to the degrees of freedom are both approximately 0.04. These results indicate that there is strong evidence of underdispersion in this model.

As shown in Table 4.2, there are a number of variables in model (4.1) which are not significant. There are 6 variables which have a p-value greater than 0.2: CRIM, CHAS, NOX, DIS, TAX and PTRATIO. These variables were then dropped from model (4.1), resulting in the following model with 7 predictors:

$$\begin{aligned}\log(\mu) &= \beta_0 + \beta_1\text{CMEDV} + \beta_3\text{ZN} + \beta_4\text{INDUS} + \beta_7\text{AGE} + \beta_9\text{RAD} + \\ &\quad \beta_{12}\text{B} + \beta_{13}\text{LSTAT} \qquad (4.2)\end{aligned}$$

Table 4.3: Results from the Global Poisson Regression Model (allowing for dispersion) with 7 Predictors

Variable	$\widehat{\beta}$	SE($\widehat{\beta}$)	t-stat	p-value
Intercept	1.7990	0.0283	63.6420	0.0000
CMEDV	0.0057	0.0006	9.9850	0.0000
ZN	0.0004	0.0002	2.3610	0.0186
INDUS	-0.0030	0.0008	-3.7280	0.0002
AGE	0.0008	0.0002	4.4230	0.0000
RAD	0.0012	0.0006	2.2160	0.0271
B	-0.0002	0.0000	-4.1570	0.0000
LSTAT	-0.0055	0.0009	-6.0250	0.0000

Table 4.4: Indicators for the 13-Predictor and 7-Predictor Global Poisson Regression Models

Indicator	13-Predictor Model	7-Predictor Model
Degrees of Freedom	492	498
Dispersion Estimate	0.0399	0.0396
Deviance	20.0872	20.1683
AIC_C	48.9426	36.4580

A quasipoisson regression model was then fit to the data, using the 7 predictors indicated in model (4.2). The estimates and standard errors of the regression coefficients from this model, including their corresponding t-statistics and p-values are shown in Table 4.3. In comparison to the model with 13 predictors, the estimate of the dispersion parameter and the ratio of the deviance to the degrees of freedom are both approximately 0.04. These results indicate that there is strong evidence of underdispersion in the 7-predictor global Poisson regression model.

To determine whether or not the variables CRIM, CHAS, NOX, DIS, TAX and PTRATIO can be dropped from the full global Poisson regression model (i.e., with all 13 predictors), an adjusted drop-in-deviance test (which accounts for dispersion) was used to test the following hypothesis:

$$H_0 : \beta_2 = \beta_5 = \beta_6 = \beta_8 = \beta_{10} = \beta_{11} = 0$$
$$H_1 : \text{at least one } \beta_i \neq 0 \ (i = 2, 5, 6, 8, 10, 11)$$

The test statistic for the adjusted drop-in-deviance test is given in equation (4.3).

$$F_{obs} = \frac{(\text{Deviance}_R - \text{Deviance}_F)/(df_R - df_F)}{\text{Dispersion Parameter Estimate}} \tag{4.3}$$

where Deviance_R and Deviance_F represent the deviance under the reduced and full models respectively; df_R and df_F represent the degrees of freedom under the reduced and full models respectively. The dispersion parameter is estimated by the ratio of the deviance and degrees of freedom under the full model. Under H_0, F_{obs} follows an F-distribution with numerator degrees of freedom (df_1) equal to $df_R - df_F$ and denominator degrees of freedom (df_2) equal to df_F.

From Table 4.4, with the reduced model representing the model with 7 predictors and the full model representing the model with 13 predictors, $F_{obs} = 0.338$, with p-value $= 0.9167$, using $df_1 = 6$ and $df_2 = 492$. Thus, there is no evidence against H_0 and the variables CRIM, CHAS, NOX, DIS, TAX and PTRATIO can be dropped from the global Poisson regression model.

In interpreting the regression coefficients shown in Table 4.3, a 1-unit increase in the value of a predictor variable x (with all other variables being held constant) has a multiplicative impact of $e^{\hat{\beta}}$ on $\hat{\mu}$. For example, the estimate of β_1 is 0.0057, so a

1-unit increase in the value of CMEDV (i.e., an increase of $1000 in the corrected median house price) is estimated to be associated with a 1.0057-fold change in the mean number of rooms per dwelling (where $e^{0.0057} = 1.0057$). In other words, a $1000 increase in the median house price yields an increase of 0.57% in the estimated mean. Similarly, an increase in the proportion of residential land zoned for lots over 25000 square feet (ZN) by 1% results in a 0.04% increase in the estimated mean while an increase in the proportion of non-retail business acres per town (INDUS) by 1% results in a decrease of 0.3% in the estimated mean.

4.5 GWPR Models

After presenting the underlying theory (Chapter 2) and examining results from simulation studies (Chapter 3), GWPR techniques (i.e., model calibration and hypothesis testing) are applied to the Boston housing price dataset. As mentioned earlier, there are two variables in the dataset which measure the geographical location of the houses in the sample: latitude (LAT) and longitude (LONG). These two variables are used as inputs for the spatial weighting function, where the distance between each of the datapoints is computed and the weighting matrix $\mathbf{W}(i)$ is obtained for each datapoint.

Various GWPR models were fit to the data using the seven predictor variables in model (4.2). The Gaussian spatial weighting function was used in the calibration of the models, with bandwidths ranging from 2.5 to 20. Table 4.5 shows some summary indicators (effective number of parameters, deviance and corrected AIC) from the calibrated GWPR models.

Table 4.5: Indicators for the 7-Predictor GWPR Models (Gaussian weighting function)

Bandwidth	Effective No. of Parameters	Deviance	AIC_C
2.5	99.8560	14.5819	264.0100
3.0	78.6227	15.5427	202.1526
3.5	62.9617	16.3312	160.4753
4.0	51.1924	16.9642	131.1242
5.0	35.5346	17.8431	94.4430
7.5	19.5285	18.8661	59.5747
10.0	14.4141	19.3675	49.1016
12.5	12.1081	19.6574	44.5176
15.0	10.8289	19.8295	42.0058
20.0	9.5482	19.9960	39.4989
Global	8	20.1683	36.4580

Local t-tests were also carried out for each of the GWPR models shown in Table 4.5, where the following hypotheses were tested for the 7 regression coefficients at each of the data points:

$$H_0 : \beta_{ki} = 0$$
$$H_1 : \beta_{ki} \neq 0$$

where k = 1, 3, 4, 7, 9, 12 and 13 (representing the coefficients for CMEDV, ZN, INDUS, AGE, RAD, B and LSTAT respectively) and $i = 1, \cdots, 506$. Table 4.6 provides a summary of the results from the local t-tests for each of the calibrated GWPR models. All tests were carried out at the 5% level of significance ($\alpha = 0.05$).

As shown in Table 4.5, the AIC_c values for all GWPR models are higher than the AIC_c for the global Poisson regression model. Also, there is an inverse relationship between the bandwidth and AIC_c - as the bandwidth increases, the AIC_c decreases

Table 4.6: Hypothesis Testing Results for the 7-Predictor GWPR Models (Gaussian weighting function)

	Number of times $H_0 : \beta_{ki} = 0$ rejected						
Bandwidth	β_{1i}	β_{3i}	β_{4i}	β_{7i}	β_{9i}	β_{12i}	β_{13i}
2.5	340	129	152	118	52	191	324
3.0	392	172	223	174	58	246	348
3.5	432	212	283	274	71	286	381
4.0	470	243	330	415	86	324	409
5.0	492	291	389	441	121	378	444
7.5	506	348	482	482	257	493	497
10.0	506	399	506	503	356	506	506
12.5	506	437	506	506	405	506	506
15.0	506	463	506	506	440	506	506
20.0	506	500	506	506	486	506	506
Global	506	506	506	506	506	506	506

and the model tends towards the global Poisson regression model in terms of model fit and number of parameters. This indicates that for this dataset, where ROOMS is the dependent variable and CMEDV, ZN, INDUS, AGE, RAD, B and LSTAT are the predictor variables, it appears that a GWPR model does not perform better than the global model. However, it should be noted that the GWPR models calibrated here allow all of the regression coefficients to vary over space (i.e., they are full-GWPR models). It may be the case that at least one of the coefficients does not depend on geographical location and that a mixed-GWPR model may be more appropriate for this data (i.e., result in a lower AIC_c value) compared to a full-GWPR model or the global model.

As shown in Table 4.6, it appears that when the value of the bandwidth is low, for some variables, the number of times H_0 is rejected is small (e.g., for the RAD variable, less than 20% of the coeffcients are determined to be significant when the bandwidth is less than 4). However, as the bandwidth increases, the number of significant coeffcients increases for all variables - when the bandwidth is 10, 5 of the 7 variables have significant coefficients at almost all of the 506 data points; when the bandwidth is 20 (i.e., the GWPR model becoming more similar to a global or non-geographically weighted model), practically all coefficients are significant at each of the data points for all 7 variables.

Chapter 5

Conclusions

In this practicum, the theory of GWPR has been discussed, including model calibration, calculation of various indicators (log-likelihood, effective number of parameters, deviance, AIC_C), choice of spatial weighting function and its corresponding parameter (i.e., bandwidth, distance or number of nearest neighbours) and hypothesis testing for the regression coefficients using local t-tests. The performance of GWPR methods was then assessed through its use on simulated data as well as a socio-economic dataset. The key findings are summarized below.

For the 1000 simulated datasets using one predictor variable, the calibrated GWPR models appeared to provide a better fit (i.e., lower average AIC_C values) than the global Poisson regression model. However, there was a large bias in the average values of the regression coefficients at each of the 16 data points, regardless of the spatial weighting function used. For the most part, the variability in the coefficient estimates appeared to be smallest for the model which used the Gaussian weighting function.

The local t-tests did not perform well for the single predictor case, with the power of the tests being very low (less than 10%) for all calibrated GWPR models. However, the power improved slightly when the value of the weighting function parameter decreased (i.e., such that, in the calibration of the GWPR model around the i^{th} data point, the influence of surrounding points on the regression parameters for that point became smaller and smaller, with larger weights for points close to the i^{th} data point and smaller weights for points far away from the i^{th} data point).

For the 1000 simulated datasets using three predictor variables, the calibrated GWPR models also provided a better fit than the global model. Even though the minimum average AIC_C values and cross-validation (CV) scores occurred at around the same value of the weighting function parameter (for both the single and multi-predictor case), there appeared to be more variability in the CV-scores in the multi-predictor case. As with the single predictor case, there was a large bias in the average values of all regression coefficients at each of the 36 data points. The local t-tests also did not perform well in the multi-predictor case; however, the power of the tests were higher compared to the single predictor case (between 12% and 20% for both β_1 and β_3; and between 20% and 40% for β_2, where the true values were farther away from zero, compared to the other regression coefficients). For β_1, the power of the local t-test increased as the value of the weighting function parameter increased (i.e., as the calibrated GWPR model became closer to the global model) while for β_2 and β_3, the power increases as the value of the weighting function parameter decreased. One must be cautious, however, in the use of GWPR models because of the large degree of bias in the estimators.

In the study by Nakaya et. al (2005) which examined the geographical patterns in the relationship between working-age mortality rates and a number of socio-economic covariates, it was found that the optimal model was a mixed (or semi-parametric) GWPR model (i.e., it produced the lowest AIC_C value, compared to all other fitted models). In this model, the coefficients associated with at least one of the covariates are held constant while the coefficients for all other covariates are allowed to vary over space. As discussed in Chapter 4, the calibrated GWPR models were full-GWPR models (i.e., where the coefficients for all variables are allowed to vary over space) and they did not provide a better fit to the data compared to the global model (where there was no geographical weighting); however, it is possible that a mixed GWPR model may yield a lower AIC_C value, in comparison to the full-GWPR model or global Poisson regression model. As an area of future research, more simulation studies may be needed to examine the performance of a mixed GWPR model, with respect to the degree of bias in the regression coefficients, selection of the weighting function parameter (i.e., bandwidth, distance or number of nearest neighbours) which yields the lowest AIC_C value and CV score, as well as hypothesis testing for the significance of the regression coefficients and the power of these tests.

As mentioned in Chapter 1, most applications of GWR have been used in the case where the response variable is continuous and the error terms are assumed to be normally distributed. Nakaya et. al (2005) discuss in detail the framework of GWPR while Atkinson et. al (2003) focus on geographically weighted logistic regression (GWLR), where the response variable is binary (i.e., the presence or absence of riverbank erosion). However, the concept of GWR can be extended to other GLMs and other statistical techniques - Paez (2006) examines the use of a geographically weighted probit model in determining the presence of spatially varying relationships in land use and transportation modeling while Brunsdon, Fotheringham and Charlton (2007) discuss the concept of geographically weighted discriminant analysis (GWDA), an adaptation of the GWR method which allows the modeling and prediction of categorical response variables.

Table 1: Summary statistics of GWPR parameter estimates - Three Predictors: Gaussian weighting function with $b = 2.0$

	$\widehat{\beta}_0$		$\widehat{\beta}_1$		$\widehat{\beta}_2$		$\widehat{\beta}_3$	
Points	Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.
1	1.2717	0.2089	-0.2139	0.1693	1.7140	1.2033	0.9540	0.9004
2	1.2694	0.1880	-0.2255	0.1513	1.7255	1.1038	0.9738	0.8150
3	1.2878	0.1707	-0.2498	0.1402	1.6810	1.0522	0.9518	0.7619
4	1.3169	0.1633	-0.2881	0.1377	1.6252	1.0557	0.9183	0.7761
5	1.3428	0.1681	-0.3381	0.1472	1.5948	1.1057	0.9161	0.8524
6	1.3586	0.1812	-0.3948	0.1680	1.5958	1.1926	0.9643	0.9545
7	1.4621	0.1696	-0.1570	0.1553	0.8530	1.1249	0.3129	0.8018
8	1.4534	0.1508	-0.1747	0.1398	0.8737	1.0063	0.3572	0.7289
9	1.4647	0.1366	-0.1992	0.1304	0.8229	0.9309	0.3550	0.6819
10	1.4858	0.1299	-0.2324	0.1271	0.7541	0.9064	0.3410	0.6873
11	1.5035	0.1317	-0.2749	0.1327	0.7091	0.9244	0.3618	0.7511
12	1.5114	0.1408	-0.3250	0.1488	0.6973	0.9780	0.4391	0.8531
13	1.6707	0.1415	-0.0780	0.1489	-0.2353	1.0099	-0.3957	0.7413
14	1.6492	0.1245	-0.1037	0.1343	-0.1856	0.8855	-0.2995	0.6764
15	1.6444	0.1150	-0.1301	0.1262	-0.2013	0.8074	-0.2461	0.6368
16	1.6495	0.1120	-0.1601	0.1234	-0.2362	0.7848	-0.2102	0.6394
17	1.6543	0.1149	-0.1964	0.1275	-0.2558	0.8084	-0.1528	0.6934
18	1.6530	0.1232	-0.2398	0.1402	-0.2506	0.8680	-0.0507	0.7900
19	1.8400	0.1355	0.0090	0.1509	-1.2553	0.9335	-0.9817	0.7268
20	1.8062	0.1183	-0.0254	0.1361	-1.1824	0.8113	-0.8269	0.6659
21	1.7847	0.1103	-0.0560	0.1286	-1.1418	0.7444	-0.7108	0.6353
22	1.7731	0.1095	-0.0860	0.1270	-1.1067	0.7444	-0.6246	0.6424
23	1.7648	0.1143	-0.1195	0.1315	-1.0621	0.7993	-0.5396	0.6917
24	1.7547	0.1241	-0.1586	0.1431	-1.0082	0.8886	-0.4298	0.7780
25	1.9509	0.1430	0.0891	0.1601	-2.0342	0.9515	-1.3838	0.7636
26	1.9096	0.1248	0.0471	0.1445	-1.9605	0.8357	-1.1866	0.7085
27	1.8765	0.1158	0.0111	0.1373	-1.8703	0.7787	-1.0303	0.6944
28	1.8522	0.1148	-0.0223	0.1368	-1.7586	0.7914	-0.9155	0.7150
29	1.8328	0.1208	-0.0575	0.1431	-1.6371	0.8613	-0.8184	0.7626
30	1.8150	0.1326	-0.0963	0.1562	-1.5235	0.9653	-0.7102	0.8342
31	2.0130	0.1558	0.1517	0.1755	-2.5938	1.0511	-1.6444	0.8611
32	1.9676	0.1365	0.1036	0.1587	-2.5258	0.9453	-1.4324	0.8129
33	1.9268	0.1260	0.0612	0.1509	-2.3952	0.8892	-1.2655	0.8160
34	1.8926	0.1243	0.0212	0.1513	-2.2175	0.8884	-1.1412	0.8463
35	1.8644	0.1314	-0.0194	0.1601	-2.0286	0.9408	-1.0363	0.8855
36	1.8404	0.1450	-0.0613	0.1762	-1.8615	1.0351	-0.9245	0.9354
Global	1.7475	0.1019	-0.1573	0.1140	-0.9260	0.7135	-0.6119	0.5780

Table 2: Summary statistics of GWPR parameter estimates - Three Predictors: Fixed bi-square weighting function with $d = 5.0$

	$\widehat{\beta}_0$		$\widehat{\beta}_1$		$\widehat{\beta}_2$		$\widehat{\beta}_3$	
Points	Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.
1	1.0513	0.2788	-0.1649	0.2003	2.6119	1.3370	1.4752	1.0227
2	1.0509	0.2400	-0.1655	0.1725	2.5712	1.2074	1.4832	0.9042
3	1.1148	0.2076	-0.2016	0.1533	2.3628	1.1398	1.3613	0.8185
4	1.2015	0.1905	-0.2560	0.1478	2.1106	1.1231	1.1870	0.8028
5	1.2659	0.1888	-0.3220	0.1592	1.9578	1.1740	1.0811	0.8898
6	1.2838	0.2079	-0.4084	0.1961	1.9868	1.3222	1.1737	1.0991
7	1.3719	0.1992	-0.1238	0.1666	1.2731	1.2123	0.4864	0.8596
8	1.3995	0.1639	-0.1449	0.1451	1.1144	1.0522	0.4489	0.7511
9	1.4305	0.1435	-0.1774	0.1324	0.9876	0.9622	0.4069	0.6902
10	1.4646	0.1354	-0.2167	0.1289	0.8651	0.9259	0.3613	0.6873
11	1.4925	0.1352	-0.2634	0.1348	0.7769	0.9427	0.3426	0.7570
12	1.5052	0.1433	-0.3233	0.1560	0.7436	1.0151	0.4209	0.9149
13	1.6784	0.1467	-0.0677	0.1530	-0.3356	1.0408	-0.4640	0.7526
14	1.6671	0.1252	-0.0955	0.1341	-0.3183	0.8840	-0.3792	0.6662
15	1.6590	0.1131	-0.1286	0.1238	-0.2920	0.7974	-0.3046	0.6231
16	1.6513	0.1115	-0.1633	0.1213	-0.2492	0.7804	-0.2290	0.6330
17	1.6472	0.1158	-0.2010	0.1255	-0.2121	0.8177	-0.1532	0.7013
18	1.6477	0.1241	-0.2444	0.1390	-0.2151	0.8899	-0.0496	0.8165
19	1.8347	0.1396	-0.0005	0.1529	-1.3026	0.966	-0.9798	0.7338
20	1.8078	0.1200	-0.0330	0.1352	-1.2146	0.8153	-0.8380	0.6538
21	1.7868	0.1090	-0.0658	0.1258	-1.1332	0.7364	-0.7166	0.6195
22	1.7685	0.1074	-0.0979	0.1235	-1.0523	0.7352	-0.6050	0.6394
23	1.7552	0.1129	-0.1297	0.1279	-0.9934	0.7969	-0.4879	0.7118
24	1.7514	0.1237	-0.1612	0.1413	-0.9777	0.8940	-0.3882	0.8144
25	1.9496	0.1503	0.0867	0.1655	-2.0231	0.9885	-1.3229	0.7869
26	1.9058	0.1278	0.0405	0.1458	-1.9132	0.8440	-1.1358	0.7221
27	1.8828	0.1159	0.0109	0.1359	-1.8393	0.7738	-0.9821	0.6961
28	1.8660	0.1145	-0.0145	0.1345	-1.7825	0.7954	-0.8441	0.7262
29	1.8528	0.1232	-0.0386	0.1431	-1.7335	0.8874	-0.7257	0.8021
30	1.8421	0.1412	-0.0640	0.1649	-1.6682	1.0175	-0.6751	0.9179
31	2.0974	0.1855	0.2403	0.2052	-2.7564	1.1535	-1.6123	0.9320
32	2.0108	0.1544	0.1495	0.1743	-2.6182	1.0177	-1.3522	0.8887
33	1.9653	0.1403	0.1008	0.1610	-2.5169	0.9515	-1.1892	0.8789
34	1.9412	0.1438	0.0736	0.1653	-2.4422	0.9815	-1.0772	0.9064
35	1.9132	0.1626	0.0424	0.1883	-2.3279	1.0707	-1.0394	0.9699
36	1.8651	0.1880	-0.0166	0.2198	-2.1234	1.1659	-1.0246	1.0659
Global	1.7475	0.1019	-0.1573	0.1140	-0.9260	0.7135	-0.6119	0.5780

Table 3: Summary statistics of GWPR parameter estimates - Three Predictors: Adaptive bi-square weighting function with $M = 31$

	$\widehat{\beta_0}$		$\widehat{\beta_1}$		$\widehat{\beta_2}$		$\widehat{\beta_3}$	
Points	Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.
1	1.2555	0.2060	-0.1574	0.1628	1.7774	1.1742	0.8794	0.8549
2	1.0664	0.2341	-0.1647	0.1694	2.5175	1.1978	1.4358	0.8893
3	1.1384	0.2018	-0.2021	0.1511	2.2816	1.1295	1.2905	0.8053
4	1.2239	0.1847	-0.2544	0.1459	2.0229	1.1072	1.1175	0.7884
5	1.2848	0.1829	-0.3161	0.1561	1.8694	1.1497	1.0166	0.8666
6	1.4237	0.1552	-0.3422	0.1536	1.2241	1.0595	0.6651	0.8607
7	1.4030	0.1902	-0.1235	0.1634	1.1218	1.1844	0.3992	0.8395
8	1.2053	0.2144	-0.1372	0.1643	1.9884	1.1958	0.9794	0.8579
9	1.1614	0.2056	-0.1673	0.1559	2.1591	1.1470	1.1577	0.8167
10	1.2757	0.1775	-0.2249	0.1461	1.7833	1.0970	0.8957	0.8039
11	1.4200	0.1527	-0.2806	0.1481	1.2000	1.0501	0.5508	0.8588
12	1.5138	0.1406	-0.3176	0.1524	0.6899	0.9980	0.3858	0.8941
13	1.6856	0.1441	-0.0691	0.1509	-0.3846	1.0225	-0.4826	0.7423
14	1.5445	0.1546	-0.0808	0.1527	0.3966	1.0510	-0.0547	0.7706
15	1.4159	0.1702	-0.1043	0.1520	1.0056	1.0614	0.3531	0.7969
16	1.5215	0.1413	-0.1502	0.1413	0.4940	0.9671	0.1525	0.7857
17	1.6175	0.1244	-0.2074	0.1379	-0.0225	0.8943	-0.0124	0.8047
18	1.6494	0.1231	-0.2418	0.1371	-0.2254	0.8824	-0.0675	0.8036
19	1.8327	0.1380	-0.0043	0.1510	-1.2986	0.9525	-0.9710	0.7247
20	1.8166	0.1319	0.0033	0.1496	-1.2393	0.9092	-0.8647	0.7484
21	1.7956	0.1283	-0.0031	0.1462	-1.3161	0.8583	-0.5678	0.8035
22	1.8144	0.1277	-0.0265	0.1445	-1.5038	0.8872	-0.3962	0.7908
23	1.7895	0.1244	-0.0962	0.1427	-1.2538	0.8956	-0.4307	0.8002
24	1.7487	0.1224	-0.1630	0.1393	-0.9584	0.8834	-0.3910	0.8022
25	1.9401	0.1476	0.0767	0.1626	-1.9886	0.9731	-1.3014	0.7751
26	1.9440	0.1400	0.0839	0.1577	-2.1477	0.9357	-1.1924	0.8098
27	1.9388	0.1357	0.0801	0.1548	-2.3097	0.9265	-0.9786	0.8580
28	1.9206	0.1400	0.0551	0.1598	-2.2600	0.9742	-0.8187	0.8848
29	1.8888	0.1428	0.0043	0.1650	-1.9992	1.0181	-0.7289	0.8984
30	1.8370	0.1376	-0.0698	0.1605	-1.6319	0.9956	-0.6645	0.8998
31	1.9741	0.1474	0.1048	0.1645	-2.2988	0.9552	-1.3955	0.7946
32	1.9985	0.1502	0.1354	0.1699	-2.5571	0.9933	-1.3376	0.8726
33	1.9590	0.1366	0.0923	0.1573	-2.4648	0.9311	-1.1747	0.8610
34	1.9398	0.1396	0.0696	0.1610	-2.4037	0.9652	-1.0563	0.8913
35	1.9178	0.1576	0.0444	0.1827	-2.3125	1.0575	-1.0053	0.9524
36	1.8754	0.1421	-0.0197	0.1669	-1.9553	1.0106	-0.8648	0.9143
Global	1.7475	0.1019	-0.1573	0.1140	-0.9260	0.7135	-0.6119	0.5780

Bibliography

[1] Agresti, A. (2002). *Categorical Data Analysis.* Wiley: New Jersey.

[2] Atkinson, P.M., German, S.E., Sear, D.A. & Clark, M.J. (2003). Exploring the Relations Between Riverbank Erosion and Geomorphological Controls Using Geographically Weighted Logistic Regression. *Geographical Analysis*, **35**, 58-82.

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

[4] Brunsdon, C., Fotheringham, A.S., & Charlton, M. (1996). Geographically Weighted Regression: A Method for Exploring Spatial Non-stationarity. *Geographical Analysis*, **28**, 281-298.

[5] Brunsdon, C., Fotheringham, A.S., & Charlton, M. (2007). Geographically Weighted Discriminant Analysis. *Geographical Analysis*, **39**, 376-396.

[6] Charlton, M., Fotheringham, A.S. & Brunsdon, C. (2003). *GWR 3: Software for Geographically Weighted Regression.* Department of Geography, University of Newcastle upon Tyne.

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

[8] 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.

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

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

[11] Fotheringham, A.S., Brunsdon, C. & Charlton, M. (2002). *Geographically Weighted Regression: The Analysis of Spatially Varying Relationships.* Wiley: Chichester.

[12] Gilley, O.W., & Pace, R.K. (1996). On the Harrison and Rubenfeld Data. *Journal of Environmental Economics and Management*, **31**, 403-405.

[13] Harrison, D., & Rubinfeld, D.L. (1978). Hedonic Housing Prices and the Demand for Clean Air. *Journal of Environmental Economics and Management*, **5**, 81-102.

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

[15] Nakaya, T., Fotheringham, A.S., Brunsdon, C. & Charlton, M. (2005). Geographically weighted Poisson regression for disease association mapping. *Statistics in Medicine*, **24**, 2695-2717.

[16] Pace, R.K., & Gilley, O.W. (1997). Using the Spatial Configuration of the Data to Improve Estimation. *Journal of Real Estate Finance and Economics*, **14**, 333-340.

[17] Paez, A. (2006). Exploring Contextual Variations in Land Use and Transport Analysis Using a Probit Model with Geographical Weights. *Journal of Transport Geography*, **14**, 167-176.

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