DIAGNOSTIC TOOLS AND REMEDIAL METHODS FOR COLLINEARITY IN LINEAR REGRESSION MODELS WITH SPATIALLY VARYING COEFFICIENTS

DISSERTATION

Presented in Partial Fulfillment of the Requirements for the Degree Doctor of Philosophy

in the Graduate School of The Ohio State University

By

David C. Wheeler, M.A.S., M.A.

*****

The Ohio State University
2006

Dissertation Committee:

Professor Morton O'Kelly, Advisor

Professor Catherine Calder, Advisor

Professor Prem Goel

Professor Ningchuan Xiao

Approved by

_____

Advisor
Graduate Program in Geography

ABSTRACT

The realization in the statistical and geographical sciences that a relationship between an explanatory variable and a response variable in a linear regression model is not always constant across a study area has lead to the development of regression models that allow for spatially varying coefficients. Two competing models of this type are geographically weighted regression (GWR) and Bayesian regression models with spatially varying coefficient processes (SVCP). In the application of these spatially varying coefficient models, marginal inference on the regression coefficient spatial processes is typically of primary interest. In light of this fact, there is a need to assess the validity of such marginal inferences, since these inferences may be misleading in the presence of explanatory variable collinearity. The presence of local collinearity in the absence of global collinearity necessitates the use of diagnostic tools in the local regression model building process to highlight areas in which the results are not reliable for statistical inference. The method of ridge regression and the lasso can also be integrated into the GWR framework to constrain and stabilize regression coefficients and lower prediction error. This dissertation presents numerous diagnostic tools and remedial methods for GWR and demonstrates the utility of these techniques with example datasets. In addition, I present the results of simulation studies designed to evaluate the sensitivity

of the spatially varying coefficients in the competing models to various levels of collinearity. The results of the simulation studies show that the Bayesian regression model produces more accurate inferences overall on the regression coefficients than does GWR. In addition, the Bayesian regression model is fairly robust in terms of marginal coefficient inference to moderate levels of collinearity, while GWR degrades more substantially with strong collinearity. The simulation study results also show that penalized versions of GWR models produce lower prediction and estimation error of the response variable than does GWR. In addition, penalized versions of GWR also lower the estimation error of the regression coefficients compared to GWR, particularly in the presence of collinearity.

Dedicated to my family

ACKNOWLEDGMENTS

I wish to thank my advisor in the Geography Department, Morton O'Kelly, for allowing me the academic and intellectual freedom to pursue my own interdisciplinary graduate program.

I also thank my advisor in the Statistics Department, Catherine Calder, for her technical and professional assistance, and words of encouragement.

I am grateful to Prem Goel for his many suggestions, both theoretical and practical, during discussions of this research.

I am also grateful to Noel Cressie, who allowed me access to computer resources that were invaluable for producing the results in this dissertation.

I would like to thank Larry Brown for making space for me in the department on very short notice.

I would also like to thank Electra Paskett, Bob Campbell, Jay Fisher, and the staff at the Comprehensive Cancer Center, who provided an excellent work environment during my final year of graduate studies.

I am grateful for numerous students of the geography and statistics departments who shared their knowledge with me.

Finally, I am grateful for the support and encouragement of my family.

VITA

1973 …… Born (Portsmouth, VA)

1995 …… B.A., Geography, Syracuse University
Syracuse, NY

1995 – 1997 ……. Scientific Programmer
Federal Express Corporation
Memphis, TN

1997 – 1999 ……... Graduate Research Associate
The Ohio State University
Columbus, OH

1999 ……. M.A., Geography, The Ohio State University
Columbus, OH

2000 – 2002 ……. Senior Quantitative Analyst
CVS Realty Co.
Woonsocket, RI

2002 – 2005 …….. Graduate Teaching Associate
The Ohio State University
Columbus, OH

2004 …….. M.A.S., The Ohio State University
Columbus, OH

2005 ……... Graduate Research Associate
Comprehensive Cancer Center
The Ohio State University
Columbus, OH

PUBLICATIONS

1. Wheeler D, Tiefelsdorf M (2005) “Multicollinearity and correlation among local regression coefficients in geographically weighted regression.” *Journal of Geographical Systems* 7(2): 161 - 187

2. Wheeler D, O’Kelly M (1999) “Network analysis and city accessibility of the commercial Internet.” *The Professional Geographer* 51(3): 327 - 339

3. Griffith D, Doyle P, Wheeler D, Johnson D (1998) “A tale of two swaths: Urban childhood blood-lead levels across Syracuse, NY.” *Annals of the Association of American Geographers* 88(4): 640 -655

FIELDS OF STUDY

Major Field: Geography

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Research Area

The realization in the statistical and geographical sciences that a relationship between an explanatory variable and a response variable in a regression model is not necessarily constant across a study area has lead to the development of linear regression models that allow for spatially varying coefficients. In the field of geography, the geographically weighted regression (GWR) model has become an increasingly popular tool in recent years (e.g., Fotheringham et al. 2002). GWR is similar in spirit and methodology to local linear regression models that were popularized in the statistics literature (see Loader 1999; Hastie et al. 2001), except that in GWR the weights applied to observations in a series of local weighted regression models across the study area are determined by a spatial kernel function instead of a kernel function in the variable space. GWR also differs from local regression in its focus, as it is concerned with measuring statistically significant variation in the regression coefficients and providing an interpretation of the coefficients, while local regression is concerned with fitting a curve to the data. When comparing local regression to traditional linear regression, Loader (1999, p. 19) states, “Instead of concentrating on the coefficients, we focus on the fitted

curve.” In contrast, the main motivation for GWR is to provide spatial data analysts with a method to visualize the spatial variation in relationships between a response variable and a set of explanatory variables via the estimated regression coefficients from each calibration location in a study area. Another link between local linear regression and GWR is found in the similarity of the estimation procedures for loess smoothing in Martinez and Martinez (2002, p. 292-293) and the GWR model in Fotheringham et al. (2002), which suggests viewing GWR is a local smoothing method.

In the statistics literature, Bayesian regression models with spatially varying coefficient processes have been introduced to model non-constant linear relationships between variables (Gelfand et al. 2003). For convenience, I will refer to this type of model as SVCP for spatially varying coefficient process model. As with GWR, the motivation for these models is that, in certain applications, regression coefficients may vary at the local or regional level. However, instead of fitting spatially local regression models, in the SVCP framework, the spatial varying coefficients are modeled as a multivariate spatial process. Such an approach fits naturally into the Bayesian paradigm, where parameters are treated as unknown random quantities. The SVCP and GWR models are similar in nature due to their same focus on the interpretation of the regression coefficient patterns and the overall fit to the response variable. The SVCP model differs from GWR in that it is a single statistical model specified in a hierarchical manner. In contrast, GWR is an ensemble of local spatial regression models, each fitted separately. Therefore, with the GWR model, there is an explicit disconnect between regression model coefficients locally; however, with the SVCP model, the dependence between regression coefficients is defined globally. I limit my focus in this dissertation on linear regression models with spatially varying coefficients to the models of GWR and Bayesian regression (SVCP).

Research Topic

One topic that is understudied in the literature related to both types of spatial varying coefficient regression models is the validity of marginal inference on the regression coefficients, particularly in the presence of collinearity in the explanatory variables. The inherent assumption in the published papers that apply these models to

data is that the regression coefficients are free of artifacts and strong dependence and, therefore, useful for marginal interpretation. This is, of course, an important and questionable assumption. In fact, there is no published work that assesses the validity of inferences derived from the GWR and SVCP models. However, as Gelman and Price (1999) show, statistical techniques can introduce artifacts into the estimation of parameters, and different statistical techniques introduce different artifacts. The work of Gelman and Price focuses on estimating disease rates and the spatial artifacts in these rates that result from small sample variation or correcting for this variation, but the general notion that spatial statistical artifacts distort model interpretation applies here. In addition, it is well know that in linear regression models, strong collinearity in the explanatory variables can increase the variance of the estimated regression coefficients. Typically, this increased variance leads to insignificant t-tests and counter-intuitive signs for at least one regression coefficient (Neter et al. 1996). In the local linear regression setting, this can lead to imprecise coefficient patterns with counter-intuitive signs in significant portions of the study area. For example, Wheeler (2006) shows that collinearity can degrade coefficient precision in GWR and lead to counter-intuitive signs for some regression coefficients at some locations in the study area of interest. In general, however, this is not a well understood and appreciated phenomena in the literature, as numerous applied papers present analyses using local spatial regression models without mentioning any relevant diagnostics for collinearity. Huang and Leung (2002) apply GWR to study regional industrialization in China and Nakaya (2001) uses the GWR approach for spatial interaction modeling with accessibility parameters and local distance decay. Longley and Tobón (2004) present a comparative study of several local and global

spatial estimation procedures, including GWR, to examine patterns of heterogeneity in intra-urban hardship. In all these applications, the authors interpret the local parameter patterns without reporting the level of correlation in the explanatory variables or estimated regression coefficients, even though there appears to be suspicious coefficient correlation in some map patterns. Of course, in these applied papers, the authors do not know the true regression coefficients and assume that the estimated coefficients approximate the true ones throughout the study area.

Fortunately, there are diagnostic tools one can use in the local spatial regression setting to highlight collinearity that may interfere with the interpretation of the estimated regression coefficients. These diagnostic tools are adapted from the traditional regression setting. Such methods to detect explanatory variable collinearity include variance-decomposition using singular-value decomposition (SVD), as described in Belsley (1991), and variance inflation factors (VIF), as outlined in Neter et al (1996). In addition, there are methods in the statistical literature that attempt to circumvent collinearity in traditional linear regression models with constant coefficients. These methods include ridge regression, the lasso, principal components regression, and partial least squares. Hastie et al. (2001) and Frank and Friedman (1993) independently provide performance comparisons of these methods. Ridge regression and lasso are both penalization, or regularization, methods that place a constraint on the regression coefficients, and principal components regression and partial least squares are both variable subset selection methods that use linear combinations of the explanatory variables in the regression model. Ridge regression was designed specifically to reduce collinearity effects by penalizing the size of regression coefficients and decreasing the influence in

the model of variables with relatively small variance. The lasso is a more recent development that also shrinks the regression coefficients, but shrinks the least significant variable coefficients to zero, thereby simultaneously performing model selection and coefficient penalization. The name for the lasso technique is derived from its function as a “least absolute shrinkage and selection operator” (Tibshirani 1996). Ridge regression and the lasso are deemed as better candidates than principal components regression and partial least squares to address collinearity in local spatial regression models because they more directly reduce the variance in the regression coefficients, with the cost of adding bias in the coefficients. Therefore, ridge regression and the lasso will be the focus of corrective methods in this dissertation. Interestingly, neither of these remedial methods has been discussed previously in the context of spatially varying coefficient models, and neither has been implemented in GWR. Seifert and Gasser (1999) suggest using ridging to improve the performance of local polynomial regression models by shrinking the local polynomial estimate towards the origin when the estimation location is far from the mean. The Seifert and Gasser method strikes a compromise between the variance and bias of the local linear estimator. It is worth mentioning that Seifert and Gasser were working in the setting of local regression models, so their kernels are applied in variable space, and they are not concerned with the interpretation of the parameters in the local polynomial fitting.

Research Goals

There are numerous goals for this dissertation research. As described above, collinearity in explanatory variables in linear regression models can lead to regression coefficient correlation, which if severe, can make the marginal interpretation of

coefficients dubious and can result in misleading conclusions about relationships in the phenomenon under study. Preliminary research shows that moderate collinearity in explanatory variables can lead to strongly correlated regression coefficients in GWR. The first research goal is to describe numerous diagnostic tools for the presence of collinearity in GWR models, while simultaneously demonstrating the problem of collinearity in GWR models using numerous illustrative examples. I will first review the GWR and Bayesian SVCP model methodology before introducing the diagnostic tools. The second goal of this research is to introduce several regularized, or augmented, versions of GWR to improve the GWR model results in the presence of collinearity. The final goal of this research is to evaluate the effectiveness of these regularization methods compared to the GWR and Bayesian SVCP models in terms of model performance and coefficient accuracy. The end result of this research will be the recommendation of a spatially varying coefficient regression model that has more directly interpretable coefficients than the GWR model, and that is more robust in terms of coefficient inference in the presence of collinearity.

To this end, I will evaluate coverage probabilities and accuracy of estimated regression coefficients, as well as the response error, for the GWR, augmented GWR, and SVCP models through simulation study, where the ‘true’ values of the regression coefficients are known. I will evaluate the coverage and accuracy of the regression coefficients both in the absence of collinearity and in the presence of various levels of collinearity using coverage probabilities and a measure of deviation from the true values of the coefficients used to simulate the data. I will compare the GWR, augmented GWR, and SVCP models based on these measures.

CHAPTER 2

LITERATURE REVIEW

The area of research outlined in this dissertation is inferential modeling with linear regression models that have spatially varying coefficients. I focus on two types of such regression models, geographically weighted regression (GWR) and a specific Bayesian regression model with spatially varying coefficient processes (SVCP). The topic of this dissertation is whether the local regression coefficients are valid for inference on the relationships between the explanatory variables and the response variable. The goals of this dissertation, outlined in the Introduction, include characterizing the condition of apparent increased regression coefficient covariation and its causes, including local collinearity in explanatory variables, and presenting remedial efforts to reduce any artificial variation in the coefficients in attempts to make them more interpretable and useful for inference. In this section, I present a critical review of previous efforts by researchers that either address the research area or topic of the dissertation or make preparatory steps towards the research goals explained in Chapter 1. In this review, I consider both work published in geographic books and relevant journals, such as *Geographical Analysis, Journal of Geographical Systems, Environment and Planning A,* and research published in the statistics literature. For brevity, I will not

discuss previous Bayesian regression and GWR research works that are not related to my dissertation topic and goals in some meaningful way.

Given the relatively weak penetration of Bayesian methods in general in the geography literature, it is not surprising that there have been only a few efforts by geographers in the area of Bayesian regression models with spatially varying coefficients. In general, GWR is written about in the geography literature, and the Bayesian SVCP model is written about in the statistics literature, although to a lesser extent to that of GWR in the geography literature. The specific type of Bayesian model that I focus on in this dissertation has been introduced in the statistics literature only in recent years and has not received nearly the attention in the geography literature that GWR models have received. Regardless, there are examples worth noting in this research area.

Gelfand et al. (2003) formally introduced the Bayesian SVCP model in a prominent statistics journal and applied it to an existing housing dataset. Gelfand and coauthors model the correlation between regression coefficients, but do not perform a simulation test to ensure that the model is estimating the true parameters correctly; they use only actual, not simulated, data. Therefore, the reader does not know whether the estimated coefficients are valid or confounded with artifacts. This is an important issue because two of the coefficients for the naturally negatively associated variables home living area and home other area in their model have a correlation of –0.84, and this may be an indication of enough collinearity in the model to affect inference on the parameters. It seems at least somewhat plausible that collinearity could have a systematic effect on the estimated regression coefficients in the Bayesian model.

Congdon (2003b) introduces a Bayesian regression model with spatially varying coefficients in a geography journal to model spatial heterogeneity in one integrated model, in contrast to the more typically used multi-models of GWR, and demonstrates the model with London suicide mortality data. Congdon (2004) presents various Bayesian regression models with coefficients that vary over space and time, some with and without spatially varying errors, and maps spatially varying coefficients related to suicide mortality. Congdon (2003a) also describes a Bayesian regression model with varying coefficients and presents an example using Scottish lip cancer data.

What is missing in these works, and is lacking in the Bayesian spatially varying coefficient regression model literature in general, is analysis that evaluates whether the spatially varying coefficients are accurately portraying the relationships in the data or are suffering from collinearity effects or other artifact-inducing processes. Maps of the spatially varying coefficients are frequently displayed in the literature with no mention of diagnostic checks on the validity of the model.

Due to the relative small amount of literature on Bayesian regression models with spatially varying coefficient processes, I now turn my discussion to GWR. There have been numerous papers in the geography literature regarding different versions of GWR and new statistical tests for different parameters in GWR. I will describe below ones that are relevant to the topic and goals of this dissertation.

The development of GWR started from smoothing techniques and local regression (Brunsdon et al. 1996; Fotheringham et al. 1998) and became more sophisticated by considering, for example, generalized linear model specifications (Fotheringham et al. 2002), spatial autocorrelation of the residuals (Leung et al. 2000b; Páez et al. 2002b),

maximum likelihood estimation of calibration location specific kernel bandwidths (Páez et al. 2002a), and a Bayesian approach to GWR that better accounts for the presence of outliers (LeSage 2004). There has also been the development of formal test statistics for spatial nonstationarity of the local regression coefficients (Leung et al. 2000a). Aside from coefficient maps associated with single exogenous variables and local *t*-values, however, none of the published GWR research developments at the time of this writing involve fundamental regression diagnostics, such as residual analysis or the precision of the regression coefficients.

LeSage (2004) introduces a Bayesian GWR (BGWR) model that is similar in spirit to the GWR model, as LeSage claims the BGWR model can replicate the GWR model estimates given certain conditions. LeSage's BGWR model includes variance scaling parameters to account for error variance heterogeneity. The model also has a scale parameter for smoothing the regression coefficients. When there is no spatial error heterogeneity and the smoothing parameter scale is very large, it appears the BGWR model should reproduce the GWR coefficient estimates. Based on results from presented experiments, BGWR is apparently more robust to the effects of outliers on regression coefficient estimates than is GWR, and this is due primarily to the error variance heterogeneity component. One criticism of LeSage's work is that there is no mention of collinearity effects and the interpretation of regression coefficients. In his example with the Columbus crime dataset, there are regions with counterintuitive signs for either one of the two variable coefficients in the model, and in some areas the magnitude of the counterintuitive coefficient is more with the BGWR model than the GWR model. This is especially surprising given LeSage's comments in the conclusions that the BGWR

coefficients are reliable for interpretation, and more interpretable than the GWR coefficients. In addition, the BGWR model adds numerous parameters for the variance heterogeneity to the original GWR model, which already has more parameters than observations.

Páez et al. (2002a) also are concerned with error variance heterogeneity in GWR and choose to shift the theoretical focus from parametric nonstationarity (see Fotheringham et al. 2002) to variance heterogeneity. This shift allows these authors to estimate local kernel bandwidths using maximum likelihood estimation and treating the inverse spatial weights as variance components. An interesting result of this work is the finding that the variation in the local GWR coefficients is substantially larger with one global kernel bandwidth, as is typically used with GWR, than with one local kernel bandwidth at each model calibration location. The global kernel bandwidth resulted in smaller kernels overall than with the local kernel bandwidths. This finding indicates that actual parametric nonstationarity could be artificially exaggerated when using a global kernel bandwidth. One caveat is that Páez et al. use a fixed, not adaptive, kernel function, and this could be partially responsible for the result. It is unreported and unknown, however, if the local/global kernel property generalizes beyond the example dataset. Another possible criticism is that the model that Páez et al. recommend has $(n-1)$ more kernel bandwidth parameters than the basic GWR model, which some might argue is overparameterized.

Páez et al. (2002b) report additional findings on the global versus local kernel bandwidths using the same data as in Páez et al. (2002a). Pseudo-R^2 calculations show that the GWR model with a global kernel bandwidth fits the data substantially better than

the GWR model with local kernel bandwidths (0.788 versus 0.527). They again use a fixed kernel function. The authors' conclusion is that since the global bandwidth results in smaller kernels, the individual models are fitting more locally and hence produce better fit to the data. They make the point that the smoothing methods that GWR is sometimes compared to place the focus on fitting the data and not producing an interpretable model. Páez et al. (2002b) also find that adding a spatial lag to the dependent variable reduces the variance heterogeneity and improves the GWR model fit. In summary, the Páez results that are relevant to my dissertation indicate that the basic GWR model may be overfitting the data locally at the expense of increased variation in regression coefficients, which could lead to problems with model interpretation.

There are also a few key papers that focus on statistical tests in GWR that are worth mentioning here. For example, Leung et al. (2000a) develop a distribution-based, statistical test in GWR for significant variation in a parameter from a constant level across a study area using the variance of the local coefficient variances. While this test of significant parametric stationarity is viewed as an improvement over the Monte Carlo simulation-based technique described in Fotheringham et al. (2002) because it is a formal statistical test, it does not consider or question the source of the parameter spatial variation. The spatial variation in the parameters could be real or greatly exaggerated from collinearity in the explanatory variables or from statistical artifacts added from the method itself, but this test will not distinguish between these situations. This could be an important difference because Wheeler and Tiefelsdorf (2005) show that the kernel weighting in GWR increases the effect of collinearity on the regression coefficients from what is found in traditional regression. Moreover, Wheeler (2006) shows that collinearity

can lead to increased estimated coefficient variances in the GWR models at some data locations. Leung et al. (2000a) also present a test of goodness of fit of the GWR model compared to the global ordinary least squares (OLS) model and use this test in a stepwise procedure for model building. There is an issue in the model building procedure, however, as they do not re-estimate the kernel bandwidth when variables are added and removed from the model. There is certainly a relationship between the kernel size and the nature of the relationships between the response variable and the explanatory variables. Presumably, the computational cost would be significant to re-estimate the bandwidth at each model-building step, especially with a large number of data points.

Mei et al. (2004) adapt the statistical test of Leung et al. (2000a) for determining which variables to designate global and which to designate local in a mixed GWR model and use a simulation to test the effectiveness of their technique. The term ‘mixed GWR model’ refers to the combination of explanatory variables that are constant across the study area and those that spatially vary. Hence, the test determines which variables should enter the model globally or locally; the intercept is also tested for spatially variability. One of the cases in the simulation study of Mei and coauthors has one constant coefficient and two spatially varying coefficients and the test rejects the null hypothesis of no coefficient variation for the constant variable for 15 percent of the 500 replications instead of the target of a 5 percent significance level for the largest sample size used, 256 observations. This rejection rate systematically decreases when the kernel size is decreased from the cross-validation kernel size and could suggest that GWR is adding artifacts to the estimated regression coefficients for the true constant coefficient and that the amount of artifact introduced depends on the kernel size. This result is

somewhat related to the simulation study results of Wheeler and Tiefelsdorf (2005) that show that GWR adds systematic artifacts to regression coefficients that are truly constant, and the strength of the artifacts increases as the correlation in the explanatory variables increases. The research of Mei et al. and Leung et al. is also noteworthy in relation to a goal in my dissertation, due to its emphasis on model selection in GWR, as I implement the lasso procedure in GWR to simultaneously constrain regression coefficients for collinearity effects and select significant variables in each local model.

While there has been a lack of attention to collinearity effects in local spatial regression models in geography and statistics, there have been a few contributions that motivate this topic and facilitate its analysis. A contribution in the literature that makes preparatory steps toward my goal of characterizing collinearity effects in spatial regression models is by Brunsdon et al. (2002) and Fotheringham et al. (2002). These authors provide formulas for calculating geographically weighted summary statistics, such as the mean and variance, for spatially referenced variables. I use this idea to calculate variance inflation factors (VIFs) as a diagnostic tool for collinearity effects in GWR using a weighted correlation coefficient between two model explanatory variables. Anselin (2003) presents an exploratory spatial data analysis software package called GeoDa that provides convenient tools for exploring spatial associations that include bivariate local indicators of spatial association (LISA) scatter plots and bivariate cluster maps. These tools were used early in the exploratory stage of my research to investigate spatial relationships and correlation between GWR coefficients.

One work that addresses collinearity effects in spatial models is by Tiefelsdorf (2003), who argues that correlation between origin-specific distance decay parameters

and origin population parameters in spatial interaction models interferes with the substantive interpretation of the distance decay parameters and makes the separation of spatial structure in the distance decay parameters and the regional attribute information problematic. The result of the collinearity in Tiefelsdorf's example is a distorted systematic pattern of origin-specific distance decay parameters in a misspecified model. Even though this paper is clearly not dealing with GWR-type models, it does provide an example of some awareness of and interest in the potential difficulty with model interpretation when substantial collinearity is present in the model. While Tiefelsdorf draws attention to the problem, he recommends no remedial methods to correct collinearity effects in spatial interaction models.

Interestingly, there have been no efforts by geographers to incorporate remedial methods in the vein of traditional regression for collinearity effects in local spatial regression models. There has also been no attention paid to identifiability issues in spatially varying coefficient regression models. There is, however, an established literature in statistics concerning remedial methods for collinearity in regression models using ridge regression. In the past few years, there has been a growing literature in computational statistics on the lasso as a remedial method for collinearity and a model selection tool. However, much like in the geography literature, there has been little attention to diagnostic tools and remedial efforts for collinearity in the statistics literature in the area of spatial local regression models. Therefore, the research goals outlined in Chapter 1 will make new contributes to the geography literature, and also to the statistics literature.

CHAPTER 3

LINEAR REGRESSION MODELS WITH SPATIALLY VARYING COEFFICIENTS

Geographically Weighted Regression

This section reviews the key equations used in fitting a GWR model. Portions of this chapter are taken from Wheeler and Calder (2006). The reader is referred to Fotheringham et al. (2002) for a more detailed introduction to the GWR framework. In GWR, the data are usually mean measures of aggregate data at fixed points with spatial coordinates; for example, see the Jiangsu province data in Huang and Leung (2002) and the numerous examples in Fotheringham et al. (2002). The spatial coordinates of the data are used to calculate distances that are used in a kernel function to determine weights of spatial dependence between observations. Typically, a separate regression model is fitted at each point location in the dataset, called model calibration locations. For each calibration location, $s = 1, \ldots, n$, the GWR model at location s is

$$y(\mathbf{s}) = \mathbf{X}(\mathbf{s})\boldsymbol{\beta}(\mathbf{s}) + \boldsymbol{\varepsilon}(\mathbf{s}), \tag{3.1}$$

where $y(s)$ is the dependent variable at location s, $\boldsymbol{\beta}(s)$ is the column vector of regression coefficients at location s, $\mathbf{X}(s)$ is the row vector of explanatory variables at

location s, and $\boldsymbol{\varepsilon}(s)$ is the random error at location s. The regression coefficients are estimated for each calibration location independently by weighted least squares. The vector of estimated regression coefficients at location s is calculated by

$$\hat{\boldsymbol{\beta}}(s) = [\mathbf{X}^T \cdot \mathbf{W}(s) \cdot \mathbf{X}]^{-1} \mathbf{X}^T \cdot \mathbf{W}(s) \cdot \mathbf{y}, \tag{3.2}$$

where $\mathbf{X} = [\mathbf{X}(1); \mathbf{X}(2); \ldots; \mathbf{X}(n)]^T$ is the design matrix of explanatory variables, which typically includes a column of 1's for the intercept; $\mathbf{W}(s) = diag[w_1(s), \ldots, w_n(s)]$ is the diagonal weights matrix that is calculated for each calibration location s; $\mathbf{y}$ is the $n \times 1$ vector of dependent variables; and $\hat{\boldsymbol{\beta}}(s) = \left(\hat{\beta}_{s0}, \hat{\beta}_{s1}, \ldots, \hat{\beta}_{sp}\right)^T$ is the vector of $p+1$ local regression coefficients at location s for p explanatory variables and an intercept.

The weights matrix, $\mathbf{W}(s)$, is calculated from a kernel function that places more weight on observations that are closer to the calibration location s. There are numerous choices for the kernel function, including the bi-square nearest neighbor function, the exponential function, and the Gaussian function. The bi-square nearest neighbor function is an adaptive kernel and has the form

$$w_j(s) = \begin{cases} [1 - (d_{sj}/b)^2]^2 & \text{if } j \in \{N_s\} \\ 0 & \text{if } j \notin \{N_s\} \end{cases}, \tag{3.3}$$

where d_{sj} is the distance between the calibration location s and location j, b is the distance to the N^{th} nearest neighbor, and the set $\{N_s\}$ contains the observations that are within the distance of the N^{th} nearest neighbor. The weights for observations beyond the N^{th} nearest neighbor distance are zero and the weight for observation s is 1. This kernel is adaptive because its spatial bandwidth adjusts to the density of data points across a study area. The weights from the exponential kernel function are calculated as

$$w_j(s) = \exp(-d_{sj} / \gamma), \tag{3.4}$$

where d_{sj} is the distance between the calibration location s and location j, and γ is the kernel bandwidth parameter. The exponential function is a fixed kernel function, in that the kernel does not adjust to the density of data points across the study area. I use the exponential function for the kernel later in the simulation study section of this dissertation to match the spatial dependence function used in the SVCP model, although one could also use another type in the Bayesian model.

To fit the GWR model, the kernel bandwidth is first estimated by cross-validation across all the calibration locations. Cross-validation (CV) is an iterative process that finds the kernel bandwidth with the lowest prediction error of all the $y(s)$. For each location s, it removes data for observation s in the model calibration at location s and predicts the response $y(s)$ using the other data points and the kernel weights associated with the current bandwidth. This is leave-one-out cross-validation because only one data point is

left out in each iteration of the cross-validation. For the bi-square nearest neighbor kernel, the kernel bandwidth estimate N_0 satisfies

$$N_0 = \arg\min_N \sum_{s=1}^{n} [y_s - \hat{y}_{(s)}(N)]^2 \,, \tag{3.5}$$

where $\hat{y}_{(s)}$ is the predicted value of observation s with the calibration observation s removed from the estimation, and N is the value of the kernel bandwidth that minimizes the cross-validation residual sum of squares. The summation term in the equation is the prediction squared error. After the kernel bandwidth is estimated, the kernel weights are calculated at each calibration location using the kernel function and the estimated bandwidth. Finally, the regression coefficients are estimated at each calibration location along with the response estimates by the expression $\hat{y}(s) = \mathbf{X}(s)\hat{\boldsymbol{\beta}}(s)$. These steps are similar to the steps in fitting local linear regression models (see Hastie et al, 2001).

Bayesian Regression Model with Spatially Varying Coefficient Processes

This section of the text reviews the Bayesian SVCP model and the methods used to estimate the model parameters. The Bayesian SVCP regression model is specified in a hierarchical manner. The distribution of the data is specified conditional on unknown parameters, whose distribution is in turn specified conditional on other parameters. Following Gelfand et al. (2003), the SVCP model is

$$\left[\mathbf{Y} \mid \boldsymbol{\beta},\, \tau^2\right] = N(\mathbf{X}^T\boldsymbol{\beta},\, \tau^2\mathbf{I}), \tag{3.6}$$

where the bracket notation $[A|B]$ denotes the distribution of A conditional on B. $\mathbf{Y}$ is a vector of responses assumed to be Gaussian conditional on the parameters $\boldsymbol{\beta}$ and τ^2; $\boldsymbol{\beta}$ is a $np\times 1$ vector of regression coefficient parameters; and $\mathbf{X}^T$ is the $n\times np$ block diagonal matrix of covariates where each row contains a row from the $n\times p$ design matrix $\mathbf{X}^*$, along with zeros in the appropriate places (the covariates from $\mathbf{X}^*$ are shifted p places in each subsequent row in $\mathbf{X}^T$); $\mathbf{I}$ is the $n\times n$ identity matrix; and τ^2 is the error variance.

In the second stage of the hierarchical model, the prior distribution for the regression coefficient parameters is specified as

$$\left[\boldsymbol{\beta}\,|\,\boldsymbol{\mu}_{\boldsymbol{\beta}},\,\boldsymbol{\Sigma}_{\boldsymbol{\beta}}\right] = N(\mathbf{1}_{n\times 1}\otimes\boldsymbol{\mu}_{\boldsymbol{\beta}},\boldsymbol{\Sigma}_{\boldsymbol{\beta}})\,. \tag{3.7}$$

The vector $\boldsymbol{\mu}_{\boldsymbol{\beta}} = (\mu_{\beta_0},\ldots,\mu_{\beta_p})^T$ contains the means of the regression coefficients corresponding to each of the p explanatory variables. The prior on the regression coefficients in the SVCP model takes into account the possible spatial dependence in the coefficients through the covariance, $\boldsymbol{\Sigma}_{\boldsymbol{\beta}}$. For $\boldsymbol{\beta}_p = [\beta_{p1},\ldots,\beta_{pn}]$, we can assume a priori that each $\boldsymbol{\beta}_p$ follows an areal unit model (e.g., the CAR or SAR model; see Banerjee et al. 2004) or specify the prior on $\boldsymbol{\beta}_p$ using a geostatistical approach, where a parametric distance-based covariance function is specified. I focus on a geostatistical prior specification of the regression coefficients and assume an exponential spatial dependence

function. The prior covariance matrix for the p different types of $\boldsymbol{\beta}$'s at each of n locations, $\boldsymbol{\Sigma}_{\boldsymbol{\beta}}$, can have either a separable or nonseparable form. The separable form has two distinct components, one for the spatial dependence in the regression coefficients and one for the within site dependence between coefficients of the same type. Following Gelfand et al. (2003), I assume a separable covariance matrix for $\boldsymbol{\beta}$ of the form

$$\boldsymbol{\Sigma}_{\boldsymbol{\beta}} = \mathbf{H}(\phi) \otimes \mathbf{T}, \tag{3.8}$$

where $\mathbf{T}$ is a positive-definite $p \times p$ matrix for the covariance of the regression coefficients at any spatial location, $\mathbf{H}(\phi)$ is the $n \times n$ correlation matrix that captures the spatial association between the n locations, ϕ is an unknown spatial dependence parameter, and $\otimes$ denotes the Kronecker product operator, which is the multiplication of every element in $\mathbf{H}(\phi)$ by $\mathbf{T}$. In the prior specification for $\boldsymbol{\beta}$ (equation 3.7), the Kronecker product results in a $np \times np$ positive definite covariance matrix, since $\mathbf{H}(\phi)$ and $\mathbf{T}$ are both positive definite. The elements of the correlation matrix $\mathbf{H}(\phi)$, $H(\phi)_{ij} = \rho(s_i - s_j; \phi)$, are calculated from the exponential function $\rho(h; \phi) = \exp(-h/\phi)$.

With the separable cross-covariance function, each of the p coefficients represented in the covariance is assumed to have the same spatial dependence structure. The separable cross-covariance form also has the property that the covariance between $\boldsymbol{\beta}$'s of the same type is constant across space. While a separable assumption is restrictive, one advantage to the separable covariance is that it is more convenient computationally than a nonseparable one and reduces the number of operations needed for matrix

inversion in simulating from the posterior distribution of the parameters. An example of the use of a nonseparable covariance matrix in a Bayesian regression model is Banerjee and Johnson (2005), who use a linear model of coregionalization to specify the prior on $\boldsymbol{\beta}$ (see also Banerjee et al. (2004) and Gelfand et al. (2004) for discussions of models of coregionalization).

The specification of the Bayesian SVCP model in equations (3.6) and (3.7) is complete with the specification of the prior distributions of the parameters. The prior for the error variance is inverse gamma with hyperparameters a and b, $\left[\tau^2\right] \sim IG(a,b)$. The prior for the coefficient means is normal with hyperparameters $\boldsymbol{\mu}$ and σ^2, $\left[\boldsymbol{\mu}_{\beta}\right] \sim N(\boldsymbol{\mu}, \sigma^2\mathbf{I})$. The prior for the covariance matrix $\mathbf{T}$ is inverse Wishart with hyperparameters v and $\boldsymbol{\Omega}$, $[\mathbf{T}] \sim IW_v(\boldsymbol{\Omega}^{-1})$. These priors are conjugate priors, and are used for computational convenience. The prior for the spatial dependence parameter ϕ is gamma with hyperparameters α and λ, $[\phi] \sim G(\alpha, \lambda)$. The parameterization of the gamma distribution used in this dissertation is

$$[\phi] \propto \phi^{\alpha-1} \exp(-\lambda\phi), \tag{3.9}$$

and the parameterization of the inverse Wishart distribution used in this dissertation is

$$[\mathbf{T}] \propto |\mathbf{T}|^{-(v+p+1)/2} \exp(-\frac{1}{2} trace \boldsymbol{\Omega}\mathbf{T}^{-1}). \tag{3.10}$$

Inference on the parameters in the SVCP model is based on the posterior distribution $[\boldsymbol{\theta} \mid \mathbf{y}]$ of the parameters $\boldsymbol{\theta} = (\boldsymbol{\beta}, \tau^2, \boldsymbol{\mu}_{\beta}, \phi, \mathbf{T})$, which can be obtained using Bayes Theorem:

$$[\boldsymbol{\theta} \mid \mathbf{y}] \propto [\mathbf{y} \mid \boldsymbol{\theta}] \cdot [\boldsymbol{\theta}]. \tag{3.11}$$

In other words, the posterior distribution for the parameters $\boldsymbol{\theta}$, conditional on the data, is proportional to the likelihood of the data $[\mathbf{y} \mid \boldsymbol{\theta}]$, also written as $p(\mathbf{y} \mid \boldsymbol{\theta})$, and the prior $[\boldsymbol{\theta}]$, also written as $p(\boldsymbol{\theta})$, for all the parameters. In most situations, it is usually not possible to find an analytic solution for the posterior distribution in complex Bayesian models. Instead, it is common in Bayesian statistics to use simulation-based inference tools such as Markov chain Monte Carlo (MCMC) methods to sample from the posterior distribution of the parameter and base inferences on these samples. MCMC algorithms simulate a Markov chain that has for its stationary distribution the target posterior distribution. The algorithm is run for a sufficient time so that, after a number of "burn-in" iterations, the algorithm converges and the sample path of the Markov chain can be taken to be samples from the posterior distribution of the unknown parameters. The samples from the chain after the "burn-in" are used to summarize inferences on the unknown parameters, where the sample mean or median is typically used as a point estimate of the parameter.

In order to check for convergence, it is common to run multiple MCMC algorithms with different starting values, where each is called a chain, and inspect that the

sampled posterior distributions are the same for the different chains. Another method to evaluate convergence is to use Gelman's scale reduction statistic, $\hat{R}$, which has values near 1 for each parameter if the algorithm has converged (Gelman et al. 2004).

In fitting Bayesian models, MCMC algorithms are typically based on the Gibbs sampler (e.g. Casella and George 1992), which iteratively samples from the full conditional distribution for each parameter, conditioning on the current value of the other parameters. The full conditional distribution is the distribution for a parameter given the other parameters in the model. At iteration j, the Gibbs sampler for the SVCP model would simulate successively from the following full conditional distributions:

$$
\begin{aligned}
&\phi(j) \sim [\phi \mid \boldsymbol{\beta}(j-1),\ \boldsymbol{\mu}_{\boldsymbol{\beta}}(j-1),\ \mathbf{T}(j-1),\ \tau^2(j-1),\ \mathbf{Y}] \\
&\mathbf{T}(j) \sim [\mathbf{T} \mid \boldsymbol{\beta}(j-1),\ \boldsymbol{\mu}_{\boldsymbol{\beta}}(j-1),\ \phi(j),\ \tau^2(j-1),\ \mathbf{Y}] \\
&\tau^2(j) \sim [\tau^2 \mid \boldsymbol{\beta}(j-1),\ \boldsymbol{\mu}_{\boldsymbol{\beta}}(j-1),\ \mathbf{T}(j),\ \phi(j),\ \mathbf{Y}] \\
&\boldsymbol{\mu}_{\boldsymbol{\beta}}(j) \sim [\boldsymbol{\mu}_{\boldsymbol{\beta}} \mid \boldsymbol{\beta}(j-1),\ \phi(j),\ \mathbf{T}(j),\ \tau^2(j),\ \mathbf{Y}] \\
&\boldsymbol{\beta}(j) \sim [\boldsymbol{\beta} \mid \phi(j),\ \boldsymbol{\mu}_{\boldsymbol{\beta}}(j),\ \mathbf{T}(j),\ \tau^2(j),\ \mathbf{Y}].
\end{aligned}
\tag{3.12}
$$

Clearly, to use the Gibbs sampler, one must be able to derive the full conditional distribution for each parameter. The posterior can be expressed as the full conditional up to a normalizing constant. The full conditional distribution for each parameter is derived by taking the product of the appropriate likelihood function and the priors for all the parameters and then simplifying the expression for the parameter of interest by ignoring terms that do not include the parameter of interest. If the distribution of a full conditional for a parameter is recognizable, one can sample directly from it in a Gibbs sampler. However, if the distribution of the full conditional is not recognizable, one can sample from it using a Metropolis-Hastings algorithm or slice sampling.

In order to perform inference on the model parameters, one must write the posterior distribution for each unknown parameter using the likelihood. The derivation of the full conditional distributions in this dissertation utilizes two versions of the likelihood. The likelihood for the SVCP model with $\mathbf{Y}$ as defined in equation (3.6) is

$$L(\boldsymbol{\mu}_{\boldsymbol{\beta}},\ \boldsymbol{\beta},\ \tau^2,\ \phi,\ \mathbf{T};\ \mathbf{y}) = \left|\tau^2\mathbf{I}\right|^{-1/2} \times \exp\{-\frac{1}{2}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})^T(\tau^2\mathbf{I})^{-1}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})\}\cdot \tag{3.13}$$

One can integrate this likelihood with respect to $\boldsymbol{\beta}$ to reduce the autocorrelation in the Markov chain. The likelihood with integrating over $\boldsymbol{\beta}$ is

$$L^*(\boldsymbol{\mu}_{\boldsymbol{\beta}},\ \tau^2,\ \phi,\ \mathbf{T};\ \mathbf{y}) = \left|\boldsymbol{\Psi}\right|^{-1/2} \times \exp\{-\frac{1}{2}(\mathbf{y}-\mathbf{X}^*\boldsymbol{\mu}_{\boldsymbol{\beta}})^T(\boldsymbol{\Psi})^{-1}(\mathbf{y}-\mathbf{X}^*\boldsymbol{\mu}_{\boldsymbol{\beta}})\}, \tag{3.14}$$

where $\mathbf{\Psi} = (\mathbf{X}(\mathbf{H}(\phi) \otimes \mathbf{T})\mathbf{X}^T + \tau^2\mathbf{I})$ and $\mathbf{X}^*$ is the $n \times p$ matrix of covariates. As an example of the use of both likelihoods, the full conditional distribution for $\boldsymbol{\mu}_{\boldsymbol{\beta}}$ is derived using the likelihood integrated over $\boldsymbol{\beta}$, $L^*(.;\mathbf{y})$, and the full conditional distribution for τ^2 is derived using the likelihood $L(.;\mathbf{y})$.

I first derive the full conditionals that are recognizable distributions using the likelihood and priors. The full conditional distributions using conjugate priors are next listed for $\mathbf{T}$, τ^2, $\boldsymbol{\mu}_{\boldsymbol{\beta}}$, and $\boldsymbol{\beta}$. The full conditional for the error variance is

$$[\tau^2 \mid \boldsymbol{\beta};\, \mathbf{y}] \sim L \times p(\tau^2) = IG(a + n/2,\, b + \frac{1}{2}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})). \tag{3.15}$$

The full conditional for the coefficient covariance matrix at any location is

$$[\mathbf{T} \mid \boldsymbol{\mu}_{\boldsymbol{\beta}},\, \boldsymbol{\beta},\, \phi;\, \mathbf{y}] \sim L \times p(\mathbf{T}) \times p(\boldsymbol{\beta}) = IW(v + n, \sum_i \sum_j (\mathbf{H}^{-1}(\phi))_{ij} (\boldsymbol{\beta}(s_j) - \boldsymbol{\mu}_{\boldsymbol{\beta}})(\boldsymbol{\beta}(s_i) - \boldsymbol{\mu}_{\boldsymbol{\beta}})^T + \boldsymbol{\Omega}), \tag{3.16}$$

where $\boldsymbol{\beta} = (\beta(s_1), \beta(s_2), \ldots, \beta(s_n))^T$ and $\boldsymbol{\mu}_{\boldsymbol{\beta}} = (\mu_{\beta_1}, \mu_{\beta_2}, \ldots, \mu_{\beta_p})^T$. The full conditional for the coefficient means is

$$[\boldsymbol{\mu}_{\boldsymbol{\beta}} \mid \mathbf{T},\, \tau^2,\, \phi;\, \mathbf{y}] \sim L^* \times p(\boldsymbol{\mu}_{\boldsymbol{\beta}}) = N(\mathbf{m}, \mathbf{S}), \tag{3.17}$$

where $\mathbf{S} = [(\sigma^2\mathbf{I})^{-1} + \mathbf{X}^{*T}\boldsymbol{\Psi}^{-1}\mathbf{X}^{*}]^{-1}$ and $\mathbf{m} = \mathbf{S}(\mathbf{X}^{*T}\boldsymbol{\Psi}^{-1}\mathbf{y} + (\sigma^2\mathbf{I})^{-1}\boldsymbol{\mu})$. The full conditional distribution for $\boldsymbol{\beta}$ is

$$[\boldsymbol{\beta} \mid \boldsymbol{\mu}_{\beta},\ \phi,\ \mathbf{T},\ \tau^2;\ \mathbf{y}] \sim L \times p(\boldsymbol{\beta}) = N(\mathbf{AC},\ \mathbf{A}), \tag{3.18}$$

where $\mathbf{A} = (\mathbf{X}^T\mathbf{X}/\tau^2 + \mathbf{H}^{-1}(\phi) \otimes \mathbf{T}^{-1})^{-1}$ and $\mathbf{C} = \mathbf{X}^T\mathbf{y}/\tau^2 + (\mathbf{H}^{-1}(\phi) \otimes \mathbf{T}^{-1})(\mathbf{1} \otimes \boldsymbol{\mu}_{\beta})$. Unlike the other parameters, the full conditional distribution of ϕ cannot be found in closed form.

When a conditional distribution for a parameter can only be calculated up to a normalizing constant, as is the case with ϕ, a Metropolis-Hastings (M-H) step can be used to draw a sample from the full conditional distribution of a parameter (e.g. Chib and Greenberg 1995). In the M-H step, I use a normal random walk proposal density that is centered on the current value of ϕ and has a variance s^2 that is tuned to produce an adequate acceptance rate of the proposed value of ϕ. The proposed value of ϕ is accepted if the ratio of the unnormalized full conditional distribution with the proposed value over the unnormalized full conditional distribution with the current value of ϕ is greater than 1 or greater than a randomly drawn uniform variable with a range of (0,1). The unnormalized full conditional distribution for ϕ is not a recognizable distribution and is

$$p(\phi \mid \mathbf{T},\ \boldsymbol{\beta},\ \boldsymbol{\mu}_{\beta};\ \mathbf{y}) \propto L \times p(\phi) \times p(\boldsymbol{\beta}) \sim \left|\mathbf{H}(\phi) \otimes \mathbf{T}\right|^{-1/2} \times \exp\{-\frac{1}{2}(\boldsymbol{\beta} - (\mathbf{1} \otimes \boldsymbol{\mu}_{\beta}))^T (\mathbf{H}(\phi) \otimes \mathbf{T})^{-1}(\boldsymbol{\beta} - (\mathbf{1} \otimes \boldsymbol{\mu}_{\beta}))\} \times \phi^{\alpha-1}\exp(-\lambda\phi). \tag{3.19}$$

An alternative method for simulating the spatial dependence parameter ϕ is slice sampling. Slice sampling does not require the proposal variance tuning of Metropolis-Hastings or the explicit, recognizable full conditional distributions required in Gibbs sampling. Instead, slice sampling uses a constraint that the density with the current sampled parameter must be greater than a random uniform variable that is drawn from the range of 0 to the density with the previous parameter value. If the constraint is satisfied, the new point is in the slice of the density. Two useful slice sampling references are Neal (2003) and Agarwal and Gelfand (2005). The version of slice sampling proposed by Neal samples uniformly from the unnormalized full conditional distribution of the parameter of interest and accepts the new point if the unnormalized full conditional density with the new point is greater than the previously drawn uniform random variable. To implement the slice sampling method of Neal for inference in the SVCP model, one would use the unnormalized full conditional density of ϕ in equation (3.19).

The version of slice sampling by Agarwal and Gelfand (2005) samples from the prior distribution of the parameter of interest and accepts the new point if the likelihood with the new point is greater than the previously drawn uniform random variable. It makes use of an auxiliary uniformly distributed variable U to sample from the posterior distribution of the parameters. The joint posterior distribution of the model parameters $\boldsymbol{\theta}$ and U is

$$f(\boldsymbol{\theta}, U \mid \mathbf{Y}) \propto \pi(\boldsymbol{\theta}) 1(U < L(\boldsymbol{\theta}; \mathbf{Y}), \tag{3.20}$$

where $\pi(\boldsymbol{\theta})$ is the prior for the parameters and 1(.) is an indicator function. The slice sampling algorithm of Agarwal and Gelfand is as follows:

Steps
a) Partition $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$ such that $f(\boldsymbol{\theta}_1 \mid \boldsymbol{\theta}_2)$ is easy to sample from
b) Draw $U \sim \text{Unif}(0, L(\boldsymbol{\theta}; \mathbf{Y}))$
c) Draw $\boldsymbol{\theta}_2$ from $f(\boldsymbol{\theta}_2 \mid \boldsymbol{\theta}_1, U)1(U < L(\boldsymbol{\theta}; Y))$ using shrinkage sampling
d) Draw $\boldsymbol{\theta}_1$ from $f(\boldsymbol{\theta}_1 \mid \boldsymbol{\theta}_2)$
e) Iterate steps b) through d) until appropriate number of samples drawn.

where

$$f(\boldsymbol{\theta}_2 \mid \boldsymbol{\theta}_1, U) \propto \pi(\boldsymbol{\theta}_2)1(U < L(\boldsymbol{\theta}; \mathbf{Y})) \tag{3.21}$$

and

$$f(\boldsymbol{\theta}_1 \mid \boldsymbol{\theta}_2) \propto L(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2; \mathbf{Y})\pi(\boldsymbol{\theta}_1 \mid \boldsymbol{\theta}_2) \tag{3.22}$$

is a standard distribution that is easy to sample from. In the SVCP model, $\boldsymbol{\theta}_2 = \phi$ and $\boldsymbol{\theta}_1 = (\boldsymbol{\mu}_{\boldsymbol{\beta}}, \mathbf{T}, \tau^2, \boldsymbol{\beta})$.

Shrinkage sampling is used in slice sampling to increase the efficiency of the algorithm by reducing the number of samples needed to get an acceptable $\boldsymbol{\theta}_2$. The steps for shrinkage sampling in slice sampling are as follows:

Step c)

1) Set large hyperrectangle containing current point $\boldsymbol{\theta}_{2,i}$
2) Draw $\boldsymbol{\theta}_{2,i+1}$ from $\pi(\boldsymbol{\theta}_2)$ using current hyperrectangle bounds on $\boldsymbol{\theta}_2$
3) If $\boldsymbol{\theta}_{2,i+1}$ is in slice ($1(U < L(\boldsymbol{\theta}; \mathbf{Y})$), then stop
4) Else shrink axis for each $\boldsymbol{\theta}$ in $\boldsymbol{\theta}_2$ by truncating upper or lower bound
5) Repeat steps 2) through 4) until point $\boldsymbol{\theta}_{2,i+1}$ is found in slice.

To implement the Agarwal and Gelfand slice sampling method for inference in the SVCP model, one would employ the likelihood expression integrated over beta in equation (3.14). Gelfand et al. (2003) use slice sampling to fit all the parameters associated with the SVCP model, including ϕ. I instead sample from the full conditional distributions for the non-spatial dependence parameters because it is more efficient. However, preliminary analyses show that both the Neal and Agarwal and Gelfand slice sampling methods produce acceptable results for estimating ϕ.

A feature of the SVCP model is that the correlation between explanatory variable coefficients is explicitly modeled. With the separable covariance matrix, the posterior correlation of the regression coefficients k and l across all locations is $T_{kl} \Big/ \sqrt{T_{kk} T_{ll}}$. Another expression for this correlation is

$$\frac{T_{kl}}{\sqrt{T_{kk} T_{ll}}} = \frac{\operatorname{cov}(\beta_k(s), \beta_l(s+h))}{\sqrt{\operatorname{cov}(\beta_k(s), \beta_k(s+h)) \operatorname{cov}(\beta_l(s), \beta_l(s+h))}}, \tag{3.23}$$

which does not depend on h, hence, the correlation of (β_k, β_l) does not depend on distance (h) between locations and is the same across the study area (Gelfand et al.

2004). Figure 3.1 conveys graphically the formula in equation (3.23) for two coefficients, say β_1 and β_2 at two locations, s and $s+h$. The crossing lines under the coefficient pair labels show the covariance pairings in the denominator of the equation, while the line above the coefficient pair labels shows the covariance pair in the numerator of the equation. The numerator is the covariance between two different coefficients at different locations and the denominator is the product of the standard deviation for one coefficient at different locations and the standard deviation for another coefficient at different locations. I make use of the coefficient correlation in a later section when evaluating dependence in the coefficients in the simulation studies.

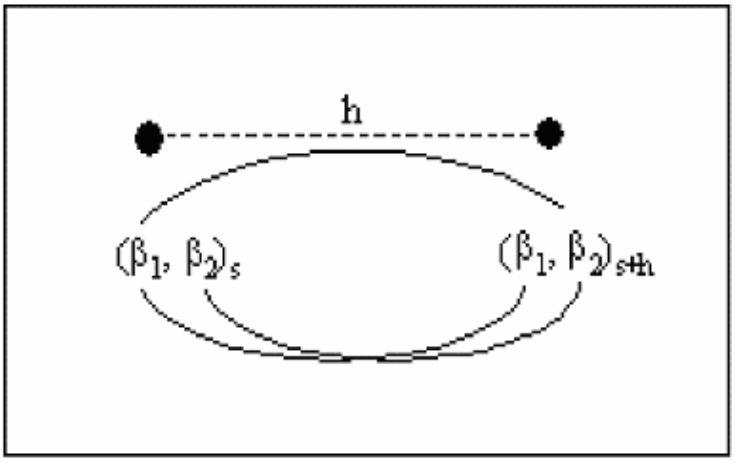

Figure 3.1. Components of the regression coefficient correlation for two types of coefficients at two hypothetical locations. The crossing lines under the coefficient pair labels show the covariance pairings in the denominator of the equation. The line above the coefficient pair labels shows the covariance pair in the numerator of the equation.

In the SVCP model, each of the p coefficients represented in the within site covariance matrix $\mathbf{T}$ has the same range in the exponential correlation function (Banerjee et al. 2004). The prior distribution specification for the regression coefficients imposes this structure, and there is naturally some question as to whether this structure will have an impact on the posterior distribution, in terms of possible influence on the correlation of the regression coefficients. Specifically, an open research question is: does the prior with the same range parameter for all variables induce dependence in the spatial process through what is in effect a constraint on the prior coefficient covariance? The extent and nature of what the impact could be is unclear at this time and is left for future research.

An extension of the Bayesian SVCP model that can address this issue uses the linear model of coregionalization (LMC), e.g. Gelfand et al. (2004). LMC allows one to use different spatial ranges in the spatial correlation function for each coefficient type, where type refers to the coefficients for one explanatory variable. Gelfand et al. (2004) use the LMC approach to have different spatial ranges for the joint response variables of selling price and income (rent) in a commercial real estate example in three cities. The LMC approach allows one to use a nonseparable covariance matrix for the regression coefficients in the SVCP model. A nonseparable covariance matrix for the regression coefficients can no longer be cleanly separated into a spatial dependence component and a within site variance component using the Kronecker product, and hence working with this type of covariance matrix usually increases computational time in finding matrix inverses. For instance, the inverse of the covariance matrix in the extended SVCP model would require $O((np)^3)$ time instead of $O(p^3)+O(n^3)$ time with the separable covariance matrix in the SVCP model (assuming the worst case scenario). Research is required to see if the additional computation time with the nonseparable covariance matrix is warranted by gains in model flexibility and improved inference.

To estimate the Bayesian SVCP model parameters with a LMC prior for the covariance of the regression coefficients, the specification of the coefficient covariance matrix must be altered to the nonseparable form

$$\boldsymbol{\Sigma}_{\boldsymbol{\beta}} = \sum_{k=1}^{p} (\mathbf{H}(\phi_k) \otimes \mathbf{T}_k), \tag{3.24}$$

where $\mathbf{H}$ is a $n \times n$ correlation matrix, $\mathbf{T}_k = a_k a_k^T$, a_k is the k^{th} column of a full rank $p \times p$ matrix $\mathbf{A}$, and ϕ_k is the spatial range parameter for the k^{th} type of regression coefficient. $\mathbf{A}$ is the lower triangular matrix of the Cholesky decomposition of $\mathbf{T}$. The index in the covariance expression for the regression coefficients starts at 1, regardless if there is an intercept or there is not. If there is an intercept, p will be increased by one and a leading column of ones added to the design matrix $\mathbf{X}^*$.

The simulation-based inference for this extended SVCP model requires relatively minor changes to the algorithm for the SVCP model. The within site covariance matrix $\mathbf{T}$ can no longer be sampled from its full conditional distribution because it is not a recognizable distribution. In addition, the one Metropolis-Hastings step for the range parameter in the SVCP model must now be replaced with p Metropolis-Hastings steps or slice sampling draws for the spatial range parameters $(\phi_1, \ldots, \phi_p)$. The expression for the coefficient covariance matrix must also be changed in the full conditional distributions for $\boldsymbol{\beta}$ and $\boldsymbol{\mu}_{\boldsymbol{\beta}}$. The subscript for the coefficient covariance matrix in this extension of the SVCP model implies that there is a different $\mathbf{T}$ for each component of the covariance matrix, where the components here are regression coefficients. However, there would be only one $\mathbf{T}$ sampled per MCMC iteration in estimating the model parameters, and then this matrix would be decomposed into p $\mathbf{T}_k$ matrices using Cholesky decomposition. It is more convenient to work with the $\mathbf{A}$ matrix when sampling $\mathbf{T}$ because $\mathbf{T} = \mathbf{A}\mathbf{A}^{\mathrm{T}}$ ensures positive definiteness (Gelfand et al. 2004). It is possible to sample $\mathbf{A}$ with either a Metropolis-Hastings step or slice sampling. To investigate the potential impacts of the separable covariance specification in the SVCP

model, I would like to in the future compare fitted results from the SVCP model and the extended SVCP model with the nonseparable covariance matrix using LMC. Coregionalization can also be used to incorporate a temporal dimension to varying coefficient models. Gelfand et al. (2005) propose an extension of the spatially varying coefficient Bayesian regression model that accommodates temporal dependence and uses coregionalization for multivariate spatial processes.

CHAPTER 4

DIAGNOSTIC TOOLS FOR COLLINEARITY

This chapter introduces numerous collinearity diagnostic tools for use with linear regression models with spatially varying coefficients, although many are specifically for GWR, and demonstrates their use with example datasets. Portions of the text in this chapter are taken from Wheeler and Tiefelsdorf (2005) and Wheeler (2006). Before describing the diagnostic tools, I will first briefly discuss the issue of collinearity in spatially varying coefficient regression models.

One potential problem with spatially varying coefficient regression models is with correlation in the estimated coefficients, at least partly due to collinearity in the explanatory variables of each local model. Wheeler and Tiefelsdorf (2005) show that while GWR coefficients can be correlated when there is no explanatory variable collinearity, the coefficient correlation increases systematically with increasing variable collinearity, and moderate collinearity of locally weighted explanatory variables can lead to potentially strong dependence in the local estimated coefficients. Wheeler and Tiefelsdorf also show that a global regression model may have acceptable coefficient correlation levels and other diagnostic levels, while its GWR counterpart may have unacceptably high levels of correlation among the local GWR coefficients. This strong

dependence in estimated coefficients can make interpretation of individual coefficients tenuous at best, and highly misleading at worst. The collinearity in variables can be exacerbated by the kernel weights applied in the GWR framework. Intuitively speaking, one is using values of a variable for each local model that are similar because they are close in space, and then applying similar weights to these nearby observations, thus intensifying the similarity in these values.

In general terms, regression model stability depends on the joint distribution of the explanatory variables, as demonstrated in the analysis by Longley (1967). In this analysis, regression coefficients changed signs depending on whether certain explanatory variables or specific observations were excluded from the model. The numerical instabilities and uncertainties in this analysis are caused by the collinearity among the explanatory variables that leads to correlation between the estimated regression coefficients. When discussing model interpretation in the face of collinearity Fox (1997, p 351) states, "collinearity deals fundamentally with the inability to separate the effects of highly correlated variables" and Greene (2000, p 256) discusses the identifiability issue of the regression coefficients by noting that "parameters are unidentified" and "different sets of parameters give the same $E(y_i)$."

In traditional regression, some commonly used exploratory tools to uncover potential collinearity among explanatory variables are bivariate scatter plots and correlation coefficients between pairs of explanatory variables, variance inflation factors (VIFs), and the correlation matrix of the estimated regression coefficients that includes the model intercept. Neter et al. (1996) provide a useful reference for these diagnostics. Belsley (1991) suggests another diagnostic tool for collinearity that uses singular value

decomposition (SVD) of the design matrix $\mathbf{X}$ to form condition indexes of this matrix and variance-decomposition proportions of the coefficient covariance matrix. These tools have not been applied to GWR models systematically by other authors. Some commonly used indicators of collinearity in a regression model are a counterintuitive sign in a regression coefficient, relatively large parameter standard errors, and large changes in magnitude or sign in one or more regression coefficients after another explanatory variable is added to the model. Just as it is important to look at diagnostic tools in a global regression analysis before interpreting the parameters, it is essential to look at these and other diagnostic tools for these effects of collinearity in local spatial regression models before interpreting the patterns of regression coefficients. The effects of collinearity will be overlooked without a proper diagnostic analysis.

Scatter Plots

To address the issue of collinearity effects in GWR, it is possible to make use of bivariate scatter plots of estimated regression coefficients, Pearson's correlation coefficient of estimated regression coefficients, and local parameter correlation maps to diagnose collinearity effects on the regression coefficients. These methods highlight both the collinearity effects on the global pattern of correlated regression coefficients across the study area and the correlated local estimated coefficients. Bivariate scatter plots show the relationship between the k^{th} and l^{th} sets of n local regression coefficients for the k^{th} and l^{th} explanatory variables and are useful for showing any strong dependencies in the coefficients.

Two Types of Coefficient Correlation

To measure the correlation in these sets of estimated regression coefficients, one can easily calculate their correlation coefficients using

$$C_{kl} \equiv Corr\left(\left\{\hat{\beta}_{1k},\ldots,\hat{\beta}_{nk}\right\},\left\{\hat{\beta}_{1l},\ldots,\hat{\beta}_{nl}\right\}\right) = \frac{\sum_{s=1}^{n}(\hat{\beta}_{sk}-\bar{\hat{\beta}}_{k})\cdot(\hat{\beta}_{sl}-\bar{\hat{\beta}}_{l})}{\sqrt{\sum_{s=1}^{n}(\hat{\beta}_{sk}-\bar{\hat{\beta}}_{k})^{2}\cdot\sum_{s=1}^{n}(\hat{\beta}_{sl}-\bar{\hat{\beta}}_{l})^{2}}}\ , \qquad (4.1)$$

where $\bar{\hat{\beta}}_{k} = \frac{1}{n}\cdot\sum_{s=1}^{n}\hat{\beta}_{sk}$. This correlation is subsequently called the *overall correlation coefficient* of two sets of local regression coefficients and is indicated C_{kl} for the correlation between variables k and l over all locations in the study area. Fotheringham et al. (2002) also present an equation for calculating what they refer to as the local regression coefficient covariance at each location. Technically, this equation is not correct because their version of GWR is not a formal statistical model with kernel weights that are part of the errors. I will, however, use this expression as an exploratory tool for correlation in the local coefficients. The equation is

$$Cov[\hat{\boldsymbol{\beta}}(s)] = \sigma^{2}\cdot[\mathbf{X}^{T}\cdot\mathbf{W}(s)\cdot\mathbf{X}]^{-1}\cdot\mathbf{X}^{T}\cdot\mathbf{W}^{2}(s)\cdot\mathbf{X}\cdot[\mathbf{X}^{T}\cdot\mathbf{W}(s)\cdot\mathbf{X}]^{-1}. \qquad (4.2)$$

A local parameter correlation matrix can be calculated from the local covariance matrix as

$$C(s) = diag^{-\frac{1}{2}}\{Cov[\hat{\boldsymbol{\beta}}(s)]\} \cdot Cov[\hat{\boldsymbol{\beta}}(s)] \cdot diag^{-\frac{1}{2}}\{Cov[\hat{\boldsymbol{\beta}}(s)]\}, \quad (4.3)$$

where $diag\{\cdot\}$ extracts the diagonal from the covariance matrix. These two equations used for the local coefficient covariance and correlation are only approximate equations because the kernel weights are calculated from the data using cross-validation before the regression coefficients are estimated from the data. The weights are inherently a function of $\mathbf{y}$, as are the regression coefficients, and the correct expression for the coefficient covariance would be non-linear. I will not attempt to derive the exact covariance formula here and instead will use the approximate formula throughout the dissertation. The correlation between coefficients for variables k and l at location s is indicated C_{kl}^{s} and comes from the (k, l) element of the $C(s)$ correlation matrix. Subsequently, I refer to these correlations as the *local coefficient correlations* at model calibration location s. The pair-wise coefficient correlations can be mapped at each calibration location for each pair of estimated coefficients.

The following matrices more clearly make the distinction between the correlations that are presented in this dissertation. The underlying $n \times (p+1)$ design matrix at the n calibration locations is

$$\mathbf{X}_{[n\times(p+1)]} = \begin{pmatrix} 1 & x_{11} & \cdots & x_{1p} \\ 1 & x_{21} & \cdots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & \cdots & x_{np} \end{pmatrix}$$

with an intercept term in the model. The resulting matrix of local GWR coefficients at calibration locations becomes

$$\mathbf{B}_{[n\times(p+1)]} = \begin{pmatrix} \beta_{10} & \beta_{11} & \cdots & \beta_{1p} \\ \beta_{20} & \beta_{21} & \cdots & \beta_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \beta_{n0} & \beta_{n1} & \cdots & \beta_{np} \end{pmatrix}.$$

The calculated local coefficient correlation matrix at one location provides the correlation among the estimated parameters in a row of the local GWR coefficient matrix

$$\mathbf{B} = \begin{pmatrix} \beta_{10} & & \beta_{11} & & \cdots & & \beta_{1p} \\ \boxed{\beta_{20}} & \leftrightarrow & \boxed{\beta_{21}} & \leftrightarrow & \cdots & \leftrightarrow & \boxed{\beta_{2p}} \\ \vdots & & \vdots & & \ddots & & \vdots \\ \beta_{n0} & & \beta_{n1} & & \cdots & & \beta_{np} \end{pmatrix}$$

and these correlations are used in local coefficient correlation maps. The overall correlation among sets of local GWR coefficients provides the correlation of pairs of parameter estimates over all locations of the local GWR coefficient matrix

$$\mathbf{B} = \begin{pmatrix} \boxed{\beta_{10}} & \leftrightarrow & \boxed{\beta_{11}} & \cdots & \beta_{1p} \\ \boxed{\beta_{20}} & \leftrightarrow & \boxed{\beta_{21}} & \cdots & \beta_{2p} \\ \vdots & & \vdots & \ddots & \vdots \\ \boxed{\beta_{n0}} & \leftrightarrow & \boxed{\beta_{n1}} & \cdots & \beta_{np} \end{pmatrix}.$$

Diagnostic Tools Example 1

I demonstrate the use of the scatter plot, overall correlation coefficient, and the local correlation coefficients in diagnosing collinearity effects in a simple model with an example dataset. The dataset comes from the Atlas of Cancer Mortality from the National Cancer Institute (Devesa et al. 1999) and contains age standardized mortality rates (per 100,000 person-years). A model was built to explain white male bladder cancer mortality rates in the 508 State Economic Areas (SEA) of the United States for the years 1970 to 1994. The model consists of the explanatory variables population density and lung cancer mortality rate, where the latter is used as a proxy for smoking, along with an intercept term. Population density is used as proxy for environmental and behavioral differences with respect to an urban/rural dichotomy. It is expected, as several studies point out, that with an increase in the population density there is an increase in the rate of bladder cancer. Lung cancer mortality rates are used as proxy for the risk factor smoking, which is a known risk factor for bladder cancer. There is epidemiological evidence that an increase in smoking elevates the risk of developing bladder cancer, thus we can expect a positive relationship between both variables. This approximation of smoking by lung cancer is reasonable, since the attributable risk of smoking for lung cancer is > 80% and the attributable risk of smoking for bladder cancer is > 55% (Mehnert et al. 1992).

A traditional regression model was first built using bladder cancer mortality as the dependent variable with population density log transformed to linearize the relationship with the dependent variable. The risk factors are significantly positively related to the rate of bladder cancer, as expected. The variance inflation factors for the two global explanatory variable parameters are less than 2 and the correlation of the global

regression parameters is moderately negative at -0.59, whereas the correlation of the two variables is 0.59. These results suggest that collinearity is not a significant problem in the global model.

The GWR model was then estimated using a bi-square nearest neighbor kernel function and I now show the previously mentioned diagnostics to check for collinearity effects in the estimated GWR coefficients. Figure 4.1 is a scatter plot of the GWR coefficient estimates for the two variables and it shows a strong negative relationship between the two sets of coefficients. The overall correlation coefficient for the sets of local coefficients is -0.85, which is much stronger than the correlation of the regression coefficients (-0.59) in the traditional regression model. The dashed reference lines are the global coefficient estimates and the plot shows that there is much variation around the global estimates. These results suggest that while collinearity is not a significant problem in the traditional regression model, it may be in the GWR model.

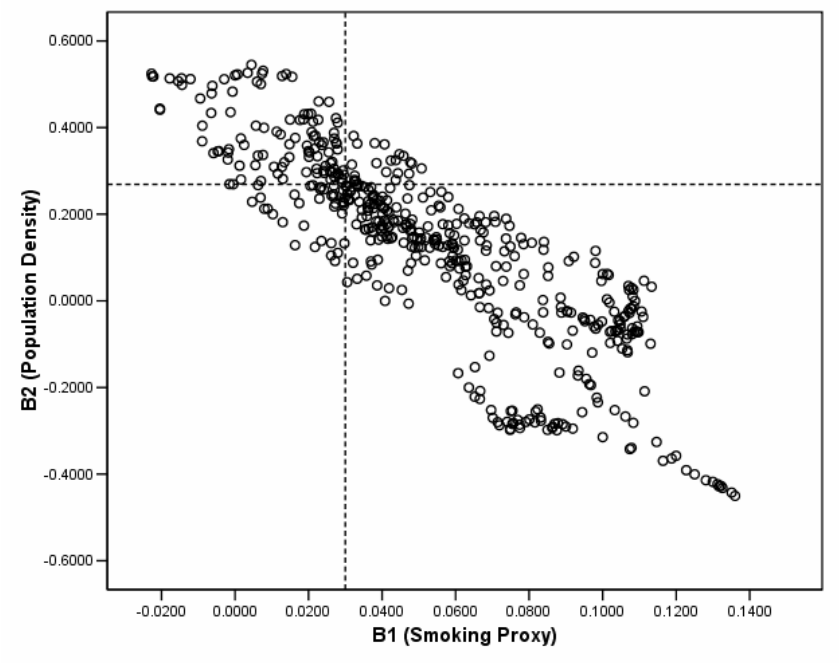

Figure 4.1. Relationship between the local GWR coefficients associated with the smoking variable and population density (C_{12} = -0.85). The dashed lines denote the levels of the related global regression coefficient estimates.

It has been argued by some that one of the primary advantages of GWR is the ability to visualize the local regression coefficient estimates in order to identify local model heterogeneities. Figure 4.2 shows the map patterns for the GWR coefficients, which are associated with different explanatory variables. The two maps show a clear inverse map pattern correlation between the sets of local regression coefficients: in general, when the local smoking proxy parameter is high, the local population density parameter is low. This pattern is most noticeable in the West, Northeast, and portions of the Midwest immediately south of Lake Michigan.

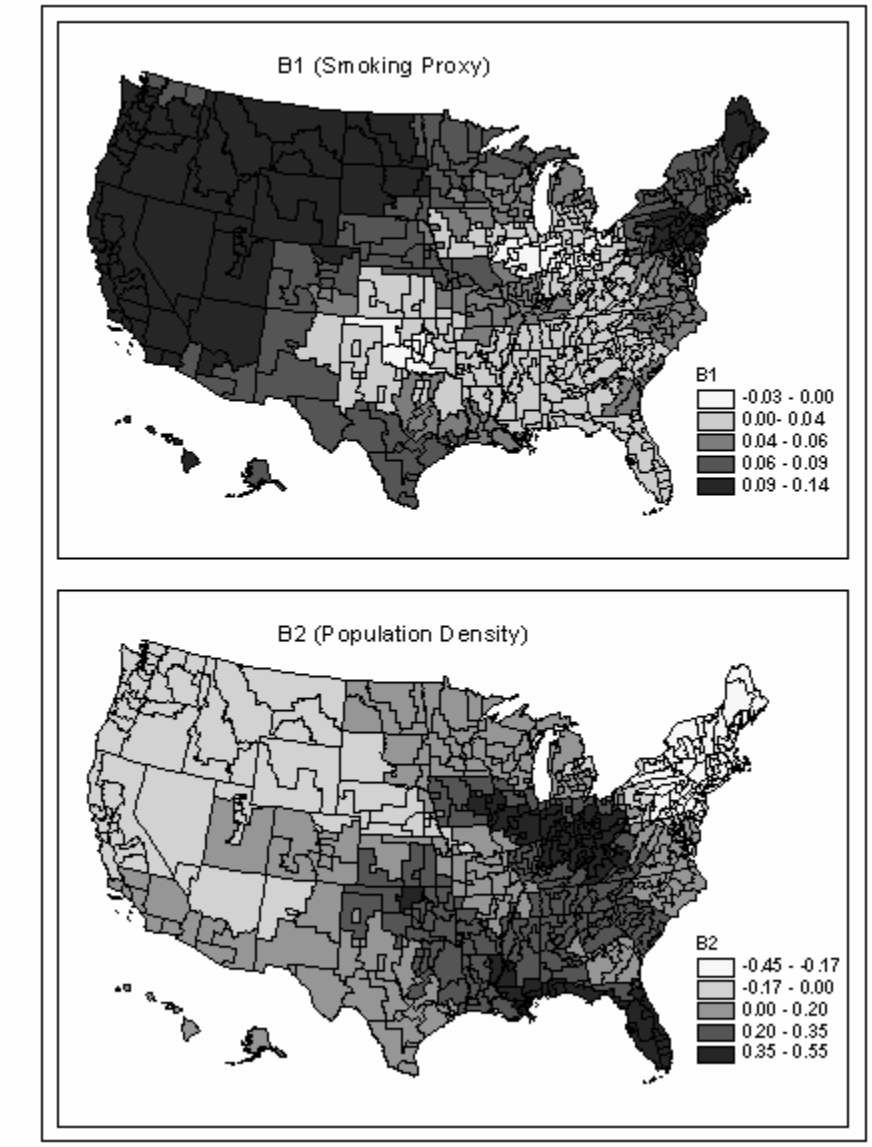

Figure 4.2. Estimated GWR coefficients for the bladder cancer mortality model. The top map displays the spatial pattern of the local regression coefficients associated with the smoking proxy variable, while the bottom map displays the spatial pattern of the local regression coefficients associated with the log population density variable.

The important question is whether this complementary relationship in the parameters is real and interpretable or a result of difficulties in the estimation of the statistical method. If the analyst does not ask this question and attempts to interpret the parameters, a likely interpretation is that in the West and Northeast smoking has a positive (statistically) relationship with bladder cancer mortality while population density has a counter-intuitive negative relationship with bladder cancer mortality. In addition, in parts of the Midwest and Oklahoma smoking has a counter-intuitive negative relationship with bladder cancer while population density has a positive relationship. If the estimated GWR coefficients are substantially affected by collinearity, this would lead to a serious false interpretation in these areas that is in gross contradiction to well-established etiological knowledge that smoking is a risk factor for bladder cancer.

Note that both choropleth map patterns of the local GWR coefficients must be cartographically symbolized by a bi-polar or diverging cartographic map theme (Brewer et al. 1997). In a bi-polar map theme a particular value denotes a common reference around which the observed values are diverging. In this case positive and negative local GWR coefficients have a substantive different interpretation. Since bi-polar map themes are difficult to display in achromatic maps, I have opted for a connotation of observations below the reference values by a light gray scale whereas observations above the reference value are encoded by a heavy gray. A noticeable gap in the middle section of the gray scale enables us to distinguish immediately between both branches of the scale.

To further explore the correlation between the sets of GWR coefficients, the local coefficient correlations for the male bladder cancer GWR model are mapped in Figure 4.3. It is clear that the strongest negative local parameter correlation is in the Midwest

and parts of the Northeast, and there are many locations in these areas with absolute magnitude correlation greater than 0.75. It is also clear from Figure 4.3 that the local coefficient correlation varies substantially over the study area and increases substantially when compared to the traditional regression coefficient correlation.

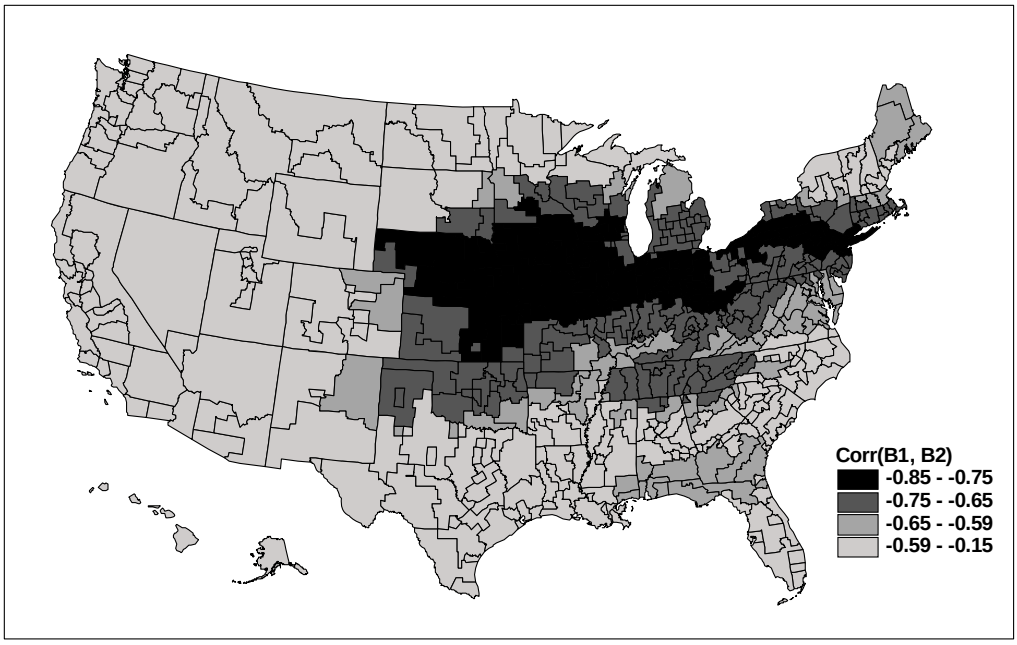

Figure 4.3. Local coefficient correlations for the GWR coefficients associated with the smoking proxy and population density variables.

I also use bivariate scatter plots to show collinearity effects in an experiment with eigenvectors of the spatial link matrix of the 159 counties from the 1990 Census layout of Georgia as explanatory variables in a GWR model with two explanatory variables and an intercept. The spatial link, or connectivity, matrix captures the mutual adjacency

relationships among the counties and the eigenvectors exhibit particular spatial patterns and the spatial autocorrelation of these patterns with respect to Moran's I is identical to the associated eigenvalue of the eigenvector. Details of the approach of generating uncorrelated spatial patterns with a given autocorrelation level can be found in Griffith (2003). The collinearity in the explanatory variables is systematically increased in the experiment, the GWR model is refitted, and then the explanatory variable coefficients are plotted. Figure 4.4 shows the estimated GWR coefficients scatter plots for four different levels of correlation in the two explanatory variables, from 0.00 to 0.72. The reference lines on the axes display the ‘true’ global parameters. The plots show the increasingly more negative relationship in the sets of coefficient estimates as the correlation in the explanatory variables increases. In addition, the variance of the local GWR coefficients also increases as the correlation in the explanatory variables increases.

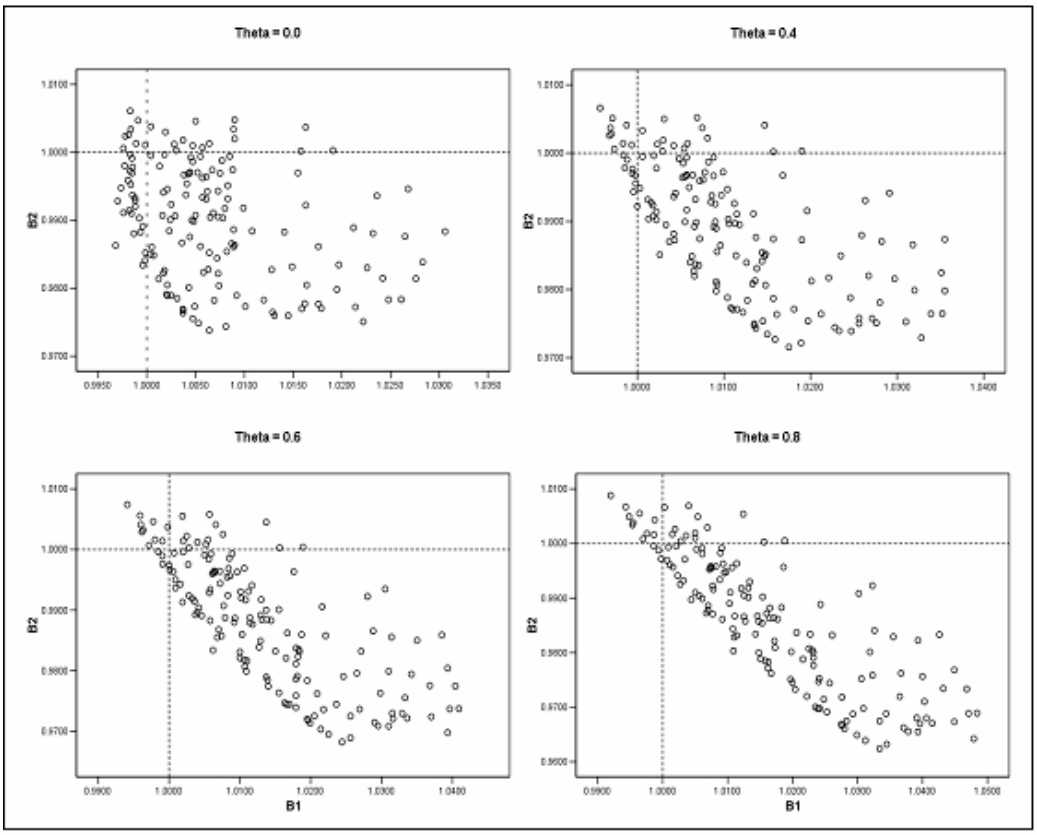

Figure 4.4. GWR coefficient estimates for two explanatory variables in a model using simulated data with correlation between the two explanatory variables at levels of 0.00 and 0.39 in the top plots (left to right) and 0.56 and 0.72 in the bottom plots. The values of theta generate the specified correlation in the explanatory variables.

I again use the overall correlation coefficient to show the relationship between estimated GWR coefficient correlation and explanatory variable collinearity through another experiment using the eigenvectors of the spatial link matrix of the counties in Georgia as explanatory variables. I systematically increase the correlation in the explanatory variables and measure the overall correlation in the estimated GWR coefficients for two different models with a different pair of explanatory variables in each

model. Figure 4.5 shows the relationship between the correlation in two pairs of explanatory variables and the estimated GWR coefficient correlation. The figure shows a clear relationship between the amount of collinearity in the explanatory variables and the overall correlation between the sets of local GWR coefficients associated with both explanatory variable pairs. The overall correlation between the coefficients becomes consistently more negative as the correlation in the exogenous variables becomes more positive. The figure also shows that there can be a fairly rapid increase in the strength of the overall correlation among the GWR coefficients as the collinearity increases.

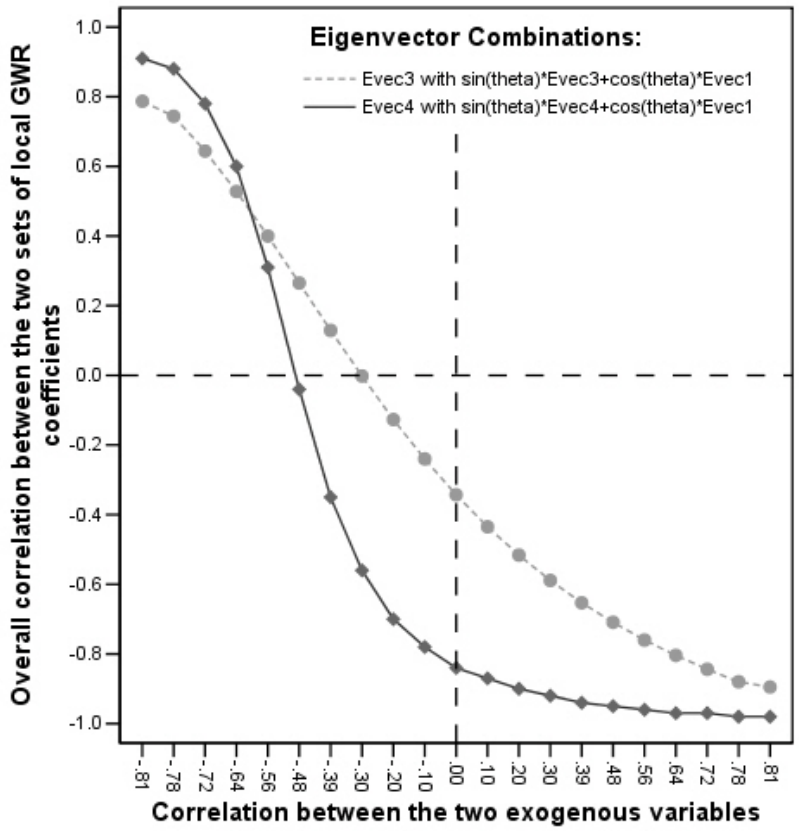

Figure 4.5. Relationship between the correlation in two pairs of explanatory variables and the overall correlation between the sets of associated GWR coefficients. There is a separate curve for each GWR model, where each model has two explanatory variables.

Variance Inflation Factor

In addition to scatter plots of GWR coefficients and maps of local coefficient correlations, it is feasible to use VIFs and variance-decomposition proportions with weighted design matrix condition numbers as diagnostic tools for collinearity in a GWR model. When using GWR models, it is possible to calculate VIF values for each explanatory variable in each local model. The VIF for a variable at location s is

$$VIF_k(s) = \frac{1}{1 - R_k^2(s)}, \tag{4.4}$$

where $R_k^2(i)$ is the coefficient of determination when x_k is regressed on the other explanatory variables at model calibration location s. The kernel size for these models is the same as in the GWR model to ensure we are diagnosing collinearity at the scale of the GWR model. For models with more than two explanatory variables, a weighted local regression of each variable on all the other variables would give the $R_k^2(i)$ needed to calculate the VIFs for the p variables. Naturally, a more efficient method to calculate the VIFs in this situation would be computationally beneficial. In a model with only two variables, the VIF is the same for both variables and is straightforward to calculate using the weighted correlation coefficient between the variables (see Fotheringham et al. 2002 for a discussion of weighted moment-based statistics). The geographically weighted correlation coefficient for two variables is

$$r_{k,l}(s) = \frac{\sum_{j=1}^{n} w_{sj}^{*}(x_{kj} - \overline{x}_{ks})(x_{lj} - \overline{x}_{ls})}{\sqrt{\sum_{j=1}^{n} w_{sj}^{*}(x_{kj} - \overline{x}_{ks})^2 \sum_{j=1}^{n} w_{sj}^{*}(x_{lj} - \overline{x}_{ls})^2}} \tag{4.5}$$

where

$$\overline{x}_{ls} = \sum_{j=1}^{n} w_{sj}^{*} x_{lj} \tag{4.6}$$

is the weighted mean for explanatory variable l at location s (similarly for variable k) and w_{sj}^{*} is the standardized kernel weight between locations s and j, where the weights are standardized to sum to one. Setting $R_k^2(s) = r_{k,l}^2(s)$ allows calculation of the VIF using equation (4.4).

Variance Decomposition Proportions and Condition Index

Two drawbacks of the VIF as a collinearity diagnostic are that it does not consider collinearity with the constant term and does not illuminate the nature of the collinearity, particularly if the collinearity is between more than two variables. Belsley (1991) suggests another diagnostic tool for collinearity that uses singular value decomposition (SVD) of the design matrix $\mathbf{X}$ to form condition indexes of this matrix and variance-decomposition proportions of the coefficient covariance matrix. It is possible to use the variance-decomposition approach in GWR by applying the GWR kernel weights to the explanatory variable design matrix. The SVD of the design matrix in the GWR framework naturally follows as

$$\mathbf{W}^{1/2}(s)\mathbf{X} = \mathbf{U}\mathbf{D}\mathbf{V}^{T}, \qquad (4.7)$$

where $\mathbf{U}$ and $\mathbf{V}$ are orthogonal $n\times(p+1)$ and $(p+1)\times(p+1)$ matrices respectively, $\mathbf{D}$ is a $(p+1)\times(p+1)$ diagonal matrix of singular values of decreasing value down the diagonal starting at position (1,1), $\mathbf{X}$ is the column scaled matrix (by its norm) of

explanatory variables including the constant, and $\mathbf{W}^{1/2}(s)$ is the square root of the diagonal weight matrix for calibration location s calculated from the kernel function.

The Páez et al. (2002a) version of GWR models variance heterogeneity over space and makes the weighted least squares (WLS) regression error assumption, which is $\boldsymbol{\varepsilon}_o \sim N(0, \sigma_o^2 \cdot \mathbf{W}^{-1})$, where for convenience $\mathbf{W}$ represents the kernel weight matrix at any location. Since the variance-decomposition diagnostic is to be applied at each model calibration location, it is natural to make this assumption on the errors here. I have followed the notation of Páez and added a subscript o to the error and error variance to indicate that these are specific to the calibration location. Using this notation, the covariance of $\hat{\boldsymbol{\beta}}$ is derived as

$$\begin{aligned} E\left[(\hat{\boldsymbol{\beta}} - E(\boldsymbol{\beta}))\cdot(\hat{\boldsymbol{\beta}} - E(\boldsymbol{\beta}))^T\right] &= E\left[(\mathbf{X}^T\mathbf{W}\cdot\mathbf{X})^{-1}\mathbf{X}^T\mathbf{W}\cdot\boldsymbol{\varepsilon}_o\cdot\boldsymbol{\varepsilon}_o{}^T\cdot\mathbf{W}\cdot\mathbf{X}\cdot(\mathbf{X}^T\mathbf{W}\cdot\mathbf{X})^{-1}\right] \\ &= (\mathbf{X}^T\mathbf{W}\cdot\mathbf{X})^{-1}\mathbf{X}^T\mathbf{W}\cdot E\left[\boldsymbol{\varepsilon}_o\cdot\boldsymbol{\varepsilon}_o{}^T\right]\cdot\mathbf{W}\cdot\mathbf{X}\cdot(\mathbf{X}^T\mathbf{W}\cdot\mathbf{X})^{-1} \\ &= \sigma_o^2\cdot(\mathbf{X}^T\mathbf{W}\cdot\mathbf{X})^{-1}\mathbf{X}^T\mathbf{W}\cdot\mathbf{W}^{-1}\cdot\mathbf{W}\cdot\mathbf{X}\cdot(\mathbf{X}^T\mathbf{W}\cdot\mathbf{X})^{-1} \\ &= \sigma_o^2\cdot(\mathbf{X}^T\mathbf{W}\cdot\mathbf{X})^{-1}, \end{aligned} \tag{4.8}$$

where $\sigma_o^2 = \frac{1}{n}(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}_o)^T\,\mathbf{W}(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}_o)$.

This is the same expression as the covariance matrix in WLS listed in Neter et al. (1996). As with equation (4.2), this equation for the local coefficient covariances is only approximate because the kernel weights are calculated from the data using cross-validation before the regression coefficients are estimated from the data. I will not

attempt to derive the exact covariance formula here and instead will use the approximate formula throughout the dissertation.

Using the SVD and matrix algebra, the variance-covariance matrix of the regression coefficients at one location can be represented as

$$Var(\hat{\boldsymbol{\beta}}) = \sigma_o^2 \mathbf{V}\mathbf{D}^{-2}\mathbf{V}^T . \tag{4.9}$$

This expression is derived by:

$$\begin{aligned} Var(\hat{\boldsymbol{\beta}}) &= \sigma_o^2 (\mathbf{X}^T\mathbf{W}\mathbf{X})^{-1} \\ \mathbf{W}^{1/2}\mathbf{X} &= \mathbf{U}\mathbf{D}\mathbf{V}^T \\ Var(\hat{\boldsymbol{\beta}}) &= \sigma_o^2 (\mathbf{X}^T\mathbf{W}^{1/2}\mathbf{W}^{1/2}\mathbf{X})^{-1} \\ &= \sigma_o^2 (\mathbf{V}\mathbf{D}\mathbf{U}^T\mathbf{U}\mathbf{D}\mathbf{V}^T)^{-1} \\ &= \sigma_o^2 (\mathbf{V}\mathbf{D}\mathbf{D}\mathbf{V}^T)^{-1} \\ &= \sigma_o^2 (\mathbf{V}\mathbf{D}^{-2}\mathbf{V}^T) \end{aligned} . \tag{4.10}$$

The variance of the k^{th} regression coefficient is

$$Var(\hat{\beta}_k) = \sigma_o^2 \sum_{j=1}^{p} \frac{v_{kj}^2}{d_j^2}, \tag{4.11}$$

where the d_j's are the singular values and the v_{kj}'s are elements of the $\mathbf{V}$ matrix. The variance-decomposition proportion is the proportion of the variance of the k^{th} regression

coefficient affiliated with the j^{th} component of its decomposition. Following from Belsley (1991), the variance-decomposition proportions are

$$\pi_{jk} = \frac{\phi_{kj}}{\phi_k}, \tag{4.12}$$

where

$$\phi_{kj} = \frac{v_{kj}^2}{d_j^2} \tag{4.13}$$

and

$$\phi_k = \sum_{j=1}^{p} \phi_{kj} . \tag{4.14}$$

The condition index for column $j = 1, ..., p+1$ of $\mathbf{W}^{1/2}(s)\mathbf{X}$ is

$$\eta_j = \frac{d_{\max}}{d_j}, \tag{4.15}$$

where d_j is the j^{th} singular value of $\mathbf{W}^{1/2}(s)\mathbf{X}$. Belsley suggests using condition indexes greater than or equal to 30, conservatively, for a column scaled $\mathbf{X}$ and variance

proportions greater than 0.50 for two or more coefficients for each variance component as an indication of collinearity in a regression model. The larger the condition number, the stronger is the collinearity among the columns of $\mathbf{X}$. The critical value of 30 is a general, and somewhat conservative, guideline from Belsley's experimentation and different values may be more appropriate for certain datasets. For example, in some of his experiments, Belsley found substantial collinearity with a condition index of 10. The presence of more than two variance proportions greater than 0.50 in one component of the variance-decomposition indicates that collinearity exists between more than two regression terms, which could include the constant. Belsley also recommends including the intercept in $\mathbf{X}$ and using uncentered explanatory variables in the SVD so as not to disguise any collinearity with the intercept.

Diagnostic Tools Example 2

I next demonstrate the utility of the VIF and variance-decomposition approach as diagnostic tools with a GWR model fitted to the Columbus crime dataset analyzed previously by Anselin (1988). This dataset contains crime rates in 49 planning neighborhoods, closely related to census tracts, in Columbus, OH. The data consist of variables for mean housing value, household income, x and y spatial coordinates of neighborhood centroids, and residential and vehicle thefts combined per thousand people for 1980. Figure 4.6 is a map of the study area with observation identifiers as labels. The dependent variable here is residential and vehicle thefts per one thousand people and is referred to as crime rate. The regression model is limited to two explanatory variables and an intercept term for ease of exposition and clarity of the methods and results. The two explanatory variables in the model are mean housing value and mean household

income, and they exhibit moderate positive correlation (r = 0.50) with one another and have clear intuitive and statistical negative relationships with crime rate (r = -0.57 and r = -0.70, respectively). One would naturally expect a negative relationship between income and crime and also housing value and crime, as more affluent neighborhoods generally have lower crime rates.

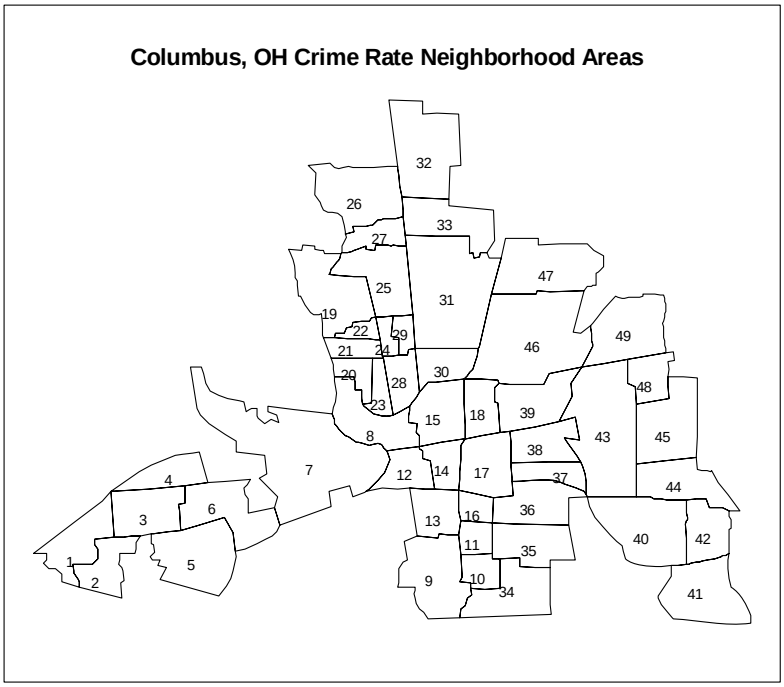

Figure 4.6. Columbus, OH 1980 crime rate neighborhood areas with identifiers.

The traditional regression model was first fitted and the results are listed in Table 4.1. The results show that collinearity is not a problem in the global model, as indicated

by the low VIF value (1.33) for the two variables, parameter correlation of –0.50 and the intuitive negative signs for the variable coefficients. The GWR model mean coefficient estimates and overall fit are listed in Table 4.2. One of the features of GWR models is typically a large increase in R^2 over the global regression model, and this is the case in Tables 4.1 and 4.2, as the R^2 increases from 0.55 to 0.92. Another noticeable difference in the two tables is the large increase in VIF values, from 1.33 to an average of 3.04 with GWR. Figure 4.7 displays the distribution of VIFs for the GWR model and shows that there are three local models with large VIFs over the conservative threshold value of 10, where VIFs greater than 10 correspond to variable correlation that considerably exceeds 0.90. These local models are at observations 37, 38, and 39, with observation 38 as the spatially connecting neighbor between 37 and 39. The local regression coefficient correlation coefficients for income and housing value at observations 37, 38, and 39 are –0.97, -0.99, and –0.98, respectively.

	Unstandardized					Standardized	
Parameter	Estimate	Standard Error	p-value	VIF	Coefficient Correlation	Estimate	Standard Error
Intercept	68.619	4.735	0.000			0.000	0.000
Inc	-1.597	0.334	0.000	1.333	-0.500	-0.544	0.114
Hoval	-0.274	0.103	0.011	1.333	-0.500	-0.302	0.114
R-square	0.55						

Table 4.1. Traditional regression model summary for unstandardized and standardized variables.

	Unstandardized				Standardized
Parameter	Mean Estimate	Mean VIF	Mean Parameter Correlation	Global Parameter Correlation	Mean Estimate
Intercept	62.670				0.000
Inc	-1.398	3.040	-0.582	-0.796	-0.477
Hoval	-0.153	3.040	-0.582	-0.796	-0.169
R-square	0.92				

Table 4.2. GWR model summary for unstandardized and standardized variables.

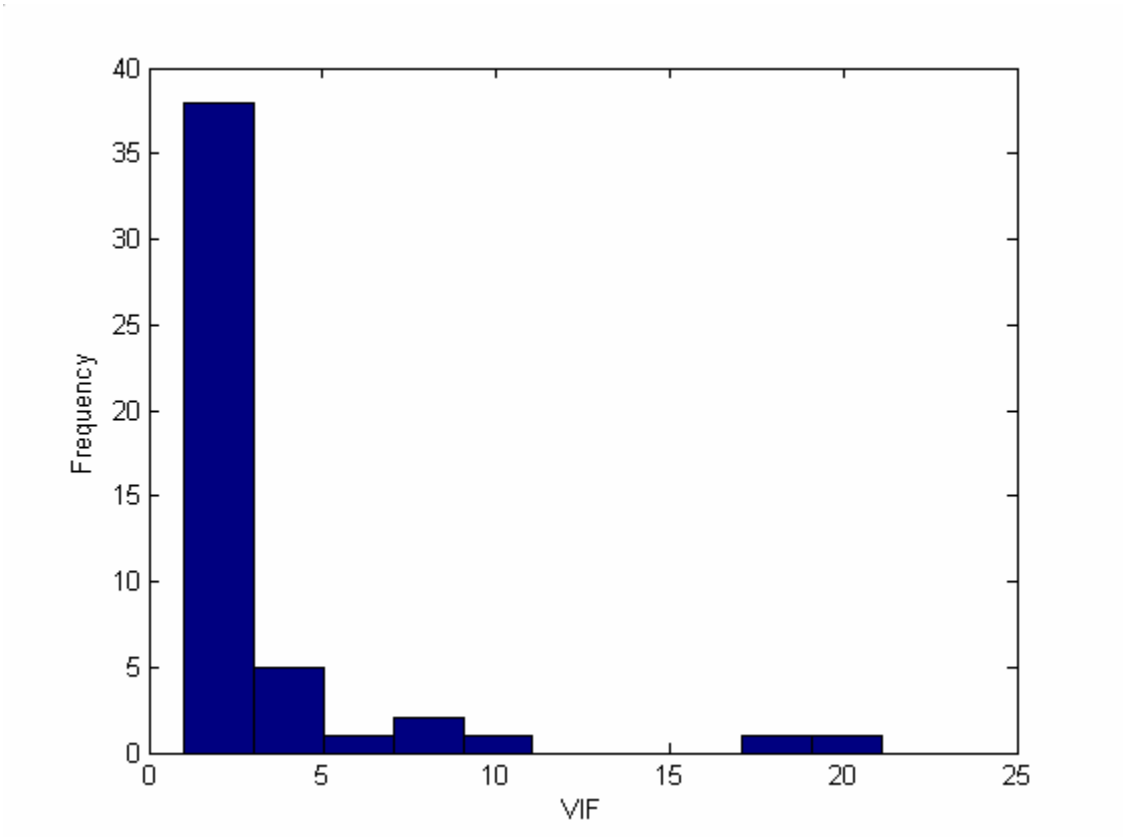

Figure 4.7. Distribution of GWR VIF values in a regression model with two explanatory variables.

The mean GWR coefficient estimates in Table 4.2 do not convey the amount of variation in the estimates and a scatter plot is beneficial to show this. Figure 4.8 shows the scatter of standardized GWR estimated coefficients for housing value versus income, along with observation identifiers, and clearly demonstrates a strong linear association between the coefficients across the study area. The correlation coefficient of the two sets of coefficients is –0.80. In Figure 4.8, it is clear that there are numerous estimated coefficients that are positive in sign, in contrast to the traditional regression model and intuition. This phenomenon is more pronounced for housing value (β_2) than for income (β_1). The three observations (37, 38, 39) with the largest VIFs are also the three observations in Figure 4.8 that have the largest housing value regression coefficients and smallest income regression coefficients. The diagnostics imply that the collinearity in the data at these model locations is increasing the variance in the estimated regression coefficients.

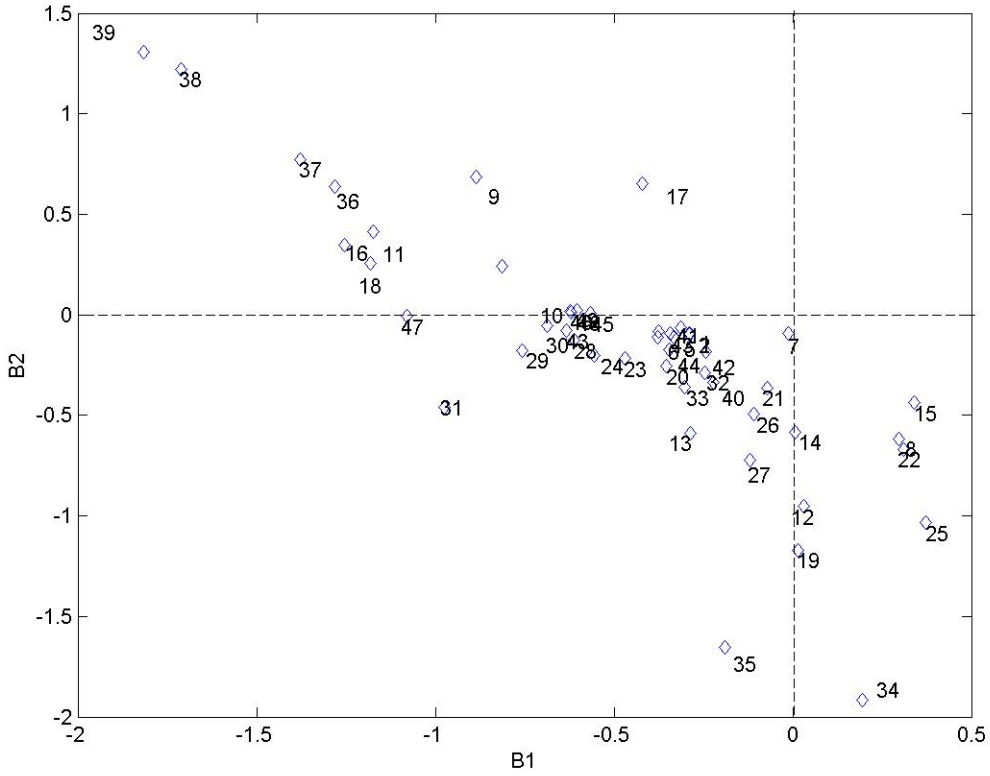

Figure 4.8. GWR estimated coefficients for housing value (β_2) versus income (β_1) with observation identifiers.

Moreover, the variance-decomposition proportions and condition indexes indicate collinearity trouble with numerous local models. Table 4.3 shows the condition index and variance-decomposition proportions for the largest variance component for all observations with a condition index greater than 25 or a VIF greater than 10. There are three observations (34, 35, 41) with a condition index over 30 and three observations (10, 39, 40) with a condition index between 25 and 30. Five of these six observations have a VIF well over 3, with observation 10 having a VIF just below 3. In all six of the observations, the variance proportions greater than 0.90 for the second and third variance

proportion columns indicate that the collinearity is between the two explanatory variables (the intercept is the first variance proportion column). Observations 34 and 35 are among the observations with the smallest estimated regression coefficients for housing value and largest estimated regression coefficients for income in Figure 4.8. Using the VIF and variance-decomposition criteria outlined above, there are indications of collinearity in at least eight of the 49 observations in the study area. These results for the GWR model suggest that collinearity is a problem with these data in the local regression models even though it is not a problem in the global regression model. Based on experience, it appears to be a general result that lack of collinearity in the global regression model will be a poor indicator for absence of collinearity in GWR models.

id	η_j	π_{j1}	π_{j2}	π_{j3}	VIF
10	28.804	0.065	0.909	0.966	2.827
34	46.472	0.388	0.948	0.992	5.602
35	41.748	0.493	0.930	0.994	4.525
37	17.106	0.021	0.975	0.964	10.881
38	21.961	0.000	0.984	0.984	21.117
39	25.999	0.053	0.988	0.981	17.815
40	25.358	0.100	0.981	0.955	8.306
41	31.024	0.038	0.975	0.956	8.382

Table 4.3. The table lists the condition indexes, variance-decomposition proportions, and VIFs for observations with either a large condition index or large VIF.

CHAPTER 5

REMEDIAL METHODS FOR COLLINEARITY

This chapter reviews the structure of the ridge regression and the lasso solutions, and then describes an implementation of each method into the GWR framework. Portions of the text in this section are taken from Wheeler (2006).

Ridge Regression

Shrinkage methods such as ridge regression place a constraint on the regression coefficients. Hoerl and Kennard (1970) first introduced ridge regression to overcome ill conditioned design matrices. The ridge regression coefficients minimize the residual sum of squares along with a penalty on the size of the squared coefficients as

$$\hat{\boldsymbol{\beta}}^{R} = \arg\min_{\boldsymbol{\beta}} \left\{ \sum_{i=1}^{n} \left(y_i - \beta_0 - \sum_{k=1}^{p} x_{ik}\beta_k \right)^2 + \lambda \sum_{k=1}^{p} \beta_k^2 \right\}, \tag{5.1}$$

where λ is the ridge regression parameter that controls the amount of shrinkage in the regression coefficients. As Hastie et al (2001) point out, an equivalent way to write the ridge regression problem that explicitly defines the constraint is

$$\hat{\boldsymbol{\beta}}^R = \arg\min_{\boldsymbol{\beta}} \sum_{i=1}^{n} \left(y_i - \beta_0 - \sum_{k=1}^{p} x_{ik}\beta_k \right)^2, \quad \text{subject to } \sum_{k=1}^{p} \beta_k^2 \le s \tag{5.2}$$

where there is a one-to-one correspondence between the parameters λ and s. The intercept is not constrained by the ridge parameter and the solutions are not invariant to scaling, so the input x variables are typically standardized to have mean 0 and equal variance and the y variable is centered before estimating λ. Hastie et al (2001) effectively remove the intercept from ridge regression by centering the x variables and estimating β_0 by the mean of y, thereby leaving only the p variable coefficients to constrain.

The ridge regression solutions are

$$\hat{\boldsymbol{\beta}}^R = \left(\mathbf{X}^T\mathbf{X} + \lambda\mathbf{I} \right)^{-1} \mathbf{X}^T\mathbf{y}, \tag{5.3}$$

where $\mathbf{I}$ is the $p \, x \, p$ identity matrix. The ridge regression parameter can be estimated before estimating the ridge regression coefficients by cross-validation or generalized cross-validation (Golub et al, 1979; Welsh 2000) by minimizing the squared prediction error.

Geographically Weighted Ridge Regression

To include ridge regression in the GWR framework, it is necessary to remove or isolate the intercept term that is customarily included in these models. There are two

approaches considered here to remove the intercept using centering of the variables, one using global centering and one using local centering. Using global centering, one first centers the x variables to remove the portion of the intercept when $\mathbf{x} = 0$, leaving the global mean of y, and then scales the x variables. Next, one centers the response variable to remove the global y mean. Then, one removes the local x and y mean deviations from the global means to get an intercept of 0 for each local model. A convenient, albeit inefficient, way to do this is to fit a GWR model to the globally centered data and then subtract the fitted intercept from the local y values. At this point, the intercept term is effectively removed from the ridge regression constraint and the penalized coefficients can be estimated. The approach allows one to compare the GWR estimates to the standardized traditional regression coefficients because the centering is the same, but the incremental estimation to remove the intercept results in additional bias in the ridge regression adjusted coefficients. It is advisable to scale the x variables by their respective standard deviations because the ridge regression solutions are scale dependent. If one does not scale the x variables, the ridge regression solution will be more influenced by variables with large variance. In other words, coefficients associated with variables of small scale should shrink more than those of larger scale.

The formula to estimate the geographically weighted ridge regression (GWRR) coefficients with global centering is

$$\hat{\boldsymbol{\beta}}(s) = \left(\mathbf{X}^{*T}\mathbf{W}(s)\,\mathbf{X}^{*} + \lambda\mathbf{I}\right)^{-1}\mathbf{X}^{*T}\mathbf{W}(s)\,\mathbf{y}^{*}, \tag{5.4}$$

where $\mathbf{X}^*$ is the $n \times p$ matrix of standardized explanatory variables, $\mathbf{y}^*$ is the standardized response variable, and other terms are as previously defined. Note that when the ridge parameter is 0, the estimated GWR and GWRR coefficients are the same. As with ridge regression, one first must estimate the ridge parameter λ before calculating the regression coefficients for each model. If one elects to use only a single λ for the entire study area, then there are now two global parameters to estimate before fitting the local models. Once the GWRR coefficients have been estimated, the response variable predictions are calculated after adjusting for the intercept term. To do so, the local mean deviation from the global $\bar{y}$ must be added back to $\hat{y}^*_{(i)} = \mathbf{X}^*_{(i)}\hat{\boldsymbol{\beta}}_{(i)}$. Bootstrapping can be used to perform inference on the regression coefficients. The bootstrap procedure to accomplish this is currently undeveloped and is left for future research.

Alternatively, one can use locally centered and globally scaled x and y values to effectively remove the local intercept. The estimation procedure is more straightforward than with global centering, but it requires centering the data for each model. The x variables are first globally scaled to have equal (unit) variance and then for each local model the x and y variables are locally centered by first calculating the weighted mean for each variable using the square root of the kernel weights $\mathbf{W}^{1/2}(s)$ and then subtracting the weighted mean from each variable. The square root of the weight is taken to correspond to the weighting of the x and y variables in equation (5.4). The weights $\mathbf{W}^{1/2}(s)$ are then applied to the centered values and this changes the coefficient estimation for the GWRR model in equation (5.4) to

$$\hat{\boldsymbol{\beta}}(s) = \left(\mathbf{X}_w^T \mathbf{X}_w + \lambda \mathbf{I}\right)^{-1} \mathbf{X}_w^T \mathbf{y}_w, \tag{5.5}$$

where $\mathbf{X}_w$ is the matrix of weighted, locally centered explanatory variables, $\mathbf{y}_w$ is the vector of weighted, locally centered responses, and other terms are as previously defined. After estimating the coefficients, the response variable predictions are calculated by adding the local mean $\bar{y}_w$ to $\hat{y}_w(s) = \mathbf{X}_w(s)\hat{\boldsymbol{\beta}}(s)$. The estimated coefficients from this approach are not as directly comparable to the standardized global regression model as the global centering results are due to the different variable centering, but the local centering should not produce coefficients that are largely dissimilar from the global centering. An advantage of this approach is that it introduces less bias in the coefficients than with the global centering approach. The local centering approach has the property that locally centered x variables will not have exactly equal scale, which means not all local models will have equal impact in the estimation of one global ridge regression parameter. In the analysis presented later in this chapter, locally centered and globally scaled variables are primarily used for the ridge parameter estimation to reduce the estimation bias in the GWRR coefficients. It is recommended that one consider the global centering approach only if a direct comparison to the global standardized regression model and traditional GWR results is of concern.

It is also possible to use local scaling of the x variables, in addition to the local centering of the variables and response. This is similar to what is done with the lasso method, which will be discussed later in this chapter, and that is why it is discussed here. With this approach, the explanatory variables are weighted, centered, and then scaled by

their norm at each model calibration location. The response variable is also weighted and centered at each model location. The ridge solutions are then calculated by equation (5.5). The solution coefficients are then scaled by the norm of the explanatory variables to rescale them to the original units so they can be used to calculate the response in the original units, using the explanatory variables in the original units. The algorithm to calculate the locally centered and scaled GWRR solutions and responses, assuming the kernel bandwidth ϕ and the ridge parameter λ have already been estimated, is:

- Calculate $\mathbf{W}$ using ϕ.
- For each location i from 1, …, n
    - Set $\mathbf{X}_\mathbf{w} = \mathbf{W}^{1/2}(i)\mathbf{X}$ and $\mathbf{y}_\mathbf{w} = \mathbf{W}^{1/2}(i)\mathbf{y}$ using the square root of the kernel weights $\mathbf{W}(i)$ at location i.
    - Calculate the mean of $\mathbf{X}_\mathbf{w}$ and set $\mathbf{X}_\mathbf{w}^c$ equal to the centered $\mathbf{X}_\mathbf{w}$. Set $\mathbf{n}_\mathbf{X}$ equal to the norm of $\mathbf{X}_\mathbf{w}^c$, and set $\mathbf{X}_\mathbf{w}^s$ equal to $\mathbf{X}_\mathbf{w}^c$ scaled by $\mathbf{n}_\mathbf{X}$. Calculate the mean of $\mathbf{y}_\mathbf{w}$ and set $\mathbf{y}_\mathbf{w}^c$ equal to the centered $\mathbf{y}_\mathbf{w}$.
    - Calculate the regression coefficients by $\boldsymbol{\beta}^s = (\mathbf{X}_\mathbf{w}^{sT}\mathbf{X}_\mathbf{w}^s + \lambda\mathbf{I})^{-1}\mathbf{X}_\mathbf{w}^{sT}\mathbf{y}_\mathbf{w}^c$.
    - Set $\boldsymbol{\beta}$ equal to $\boldsymbol{\beta}^s$ rescaled by $\mathbf{n}_\mathbf{X}$.
    - Calculate $\hat{\mathbf{y}} = \mathbf{X}\cdot\boldsymbol{\beta}$

The *a priori* thinking with the local scaling is that it may increase the model stability and, therefore, lower prediction error. The performance of local centering and scaling version of GWRR, particularly compared to the local centering and global scaling version of GWRR, will be evaluated empirically later in the simulation study chapter.

There are numerous possible schemes for estimating the kernel and ridge parameters with cross-validation: 1) estimate the kernel bandwidth first and then the ridge parameter, 2) estimate the kernel bandwidth, then estimate the ridge parameter, and then repeat using previous values until the parameters converge, 3) perform a search for the kernel bandwidth and perform a search for the ridge parameter at each evaluated value of the kernel bandwidth, and 4) estimate the kernel bandwidth and ridge parameter simultaneously with constrained optimization techniques.

Preliminary results show that there is interaction at times between the two parameters, although the kernel bandwidth dominates the squared prediction error, so scheme 1 will generally not produce optimal solutions. Scheme 2 also tends to result in a sub-optimal solution because the kernel bandwidth tends to dominate the solution and it is unlikely to move to another bandwidth in the solution by changing the ridge parameter. Scheme 3 is less efficient and scheme 4 is more complicated than the first two schemes, but these schemes will generally produce better solutions. Schemes 1 and 2 will generally produce similar solutions and schemes 3 and 4 should produce similar solutions if scheme 4 treats the kernel bandwidth as an integer variable when using the bi-square nearest neighbor kernel, as it is handled in scheme 3. While schemes 1 and 2 will not always find the minimum solution, they should yield good solutions that are almost as good as those from schemes 3 and 4 and will do so with less computational effort.

Scheme 3 is a compromise between the computational ease of schemes 1 and 2 and the complexity of scheme 4 and is therefore viewed as the best scheme for this research. Scheme 1 has some conceptual appeal in addition to its computational ease in that it takes the best kernel bandwidth found from the standard GWR estimation procedure and then effectively applies the ridge parameter to that solution. This makes the effect of the ridge parameter on GWR clearer, and for this reason the scheme 1 solution will at times be used in place of the scheme 3 solution in this dissertation. A golden section or bisection search can be used for schemes 1 and 2, while a constrained optimization algorithm is appropriate for scheme 4. A nested golden section or bisection search algorithm can be used for the two components of the estimation in scheme 3. Conveniently, Matlab software, among others, provides functions for constrained optimization and the golden section search. I also programmed the bisection search algorithm in an R implementation of GWRR.

One problem with the golden section search algorithm in Matlab is that it can terminate in a local optimum and may need to be adjusted when estimating the kernel bandwidth, as the squared prediction error function can flatten out and become stable as N increases with some datasets. This was an issue in datasets analyzed in this research and was addressed by running the golden section search routine in Matlab twice and truncating the bounds of the search space in the second run using the solution from the first run as the upper bound. A more conservative approach is to evaluate all possible values of the kernel bandwidth in the cross-validation and then select the best bandwidth by inspection. This, however, is only possible for discrete kernel bandwidths, as in the bi-square nearest neighbor kernel, and may be computationally expensive for large datasets. A faster alternative is to use a grid search, but the coarseness of the grid may miss the best bandwidth value.

There are other possible methods to estimate the kernel bandwidth. Fotheringham et al (2002) describe a generalized cross-validation (GCV) criterion for GWR that is adapted from local linear regression and also define an Akaike information criterion (AIC) for the GWR framework. Páez et al (2002a) present an alternative GWR model that can estimate local kernel bandwidths at each model calibration location by using maximum likelihood estimation and calculating the spatial weights as part of a model for variance. More attention is needed to determine the most appropriate kernel bandwidth estimation method. It is worth mentioning that the type of cross-validation used here is leave-one-out because for local regression models it is not readily justifiable to remove more than one observation for each local model. Removing anything but the i^{th} observation for the prediction at observation i seems arbitrary.

There is a modest increase in computational complexity to include the ridge regression parameter in GWR. The main computational burden in the GWR version implemented here is the CV estimation of the kernel bandwidth. The number of calculations in the CV estimation is dominated by the calculation of the kernel weights and matrix inverse for the regression coefficients at each location, not the number of iterations of the golden section search routine. An estimate of the total time required for the CV estimation of the bandwidth in GWR is $O(\log n \cdot (n^2 + np^3))$, where the number of iterations of the search routine is on the order of $\log n$ and there are n calculations of the kernel weights and matrix inverse taking $(n + p^3)$ calculations. Under estimation scheme 3, the CV estimation of λ in GWRR is nested within the CV estimation of N and this transfers the $O(n^2 + np^3)$ time from the N estimation to the λ estimation. The only additional computation is with the number of golden section search iterations needed to find λ at each value of N. GWRR model calibrations for four different sized datasets using the bi-square nearest neighbor kernel and the golden section search routine show that the number of search iterations needed for λ is approximately the same as for N. Therefore, including the ridge parameter in GWR effectively doubles the number of iterations needed in the search routine to estimate the parameters and the computational complexity of the CV estimation in GWRR is $O((\log n)^2 \cdot (n^2 + np^3))$. GWR and GWRR are both polynomial-time algorithms.

The Lasso

The lasso takes the shrinkage of ridge regression a step further by shrinking the regression coefficients of some variables to zero. The lasso specification is similar to

ridge regression, but it has a L_1 coefficient penalty in place of the ridge L_2 penalty. The lasso is defined as

$$\hat{\boldsymbol{\beta}}^R = \arg\min_{\boldsymbol{\beta}} \sum_{i=1}^{n} \left(y_i - \beta_0 - \sum_{k=1}^{p} x_{ik}\beta_k \right)^2 \quad \text{subject to} \quad \sum_{k=1}^{p} |\beta_k| \le s \tag{5.6}$$

Tibshirani (1996) notes that the lasso constraint $\sum_k |\beta_k| \le s$ is equivalent to adding the penalty term $\lambda \sum_k |\beta_k|$ to the residual sum of squares, so there is a direct correspondence between the parameters s and λ. The absolute value constraint on the regression coefficients makes the problem nonlinear and a typical way to solve this type of problem is with quadratic programming.

There are, however, ways to estimate the lasso coefficients outside of the mathematical programming framework. Tibshirani (1996) provides an algorithm that finds the lasso solutions by treating the problem as a least squares problem with 2^p inequality constraints, one for each possible sign of the β_k's, and applying the constraints sequentially. An even more attractive way to solve the lasso problem is proposed by Efron et al. (2004a), who solve the lasso problem with a small modification to the least angle regression (LARS) algorithm, which is a variation of the classic forward selection algorithm in linear regression. The modification ensures that the sign of any non-zero estimated regression coefficient is the same as the sign of the correlation coefficient

between the corresponding explanatory variable and the current residuals. Grandvalet (1998) shows that the lasso is equivalent to adaptive ridge regression and develops an EM algorithm to compute the lasso solution.

It is worthwhile to describe in more detail the LARS and lasso algorithms of Efron et al. (2004a) because these methods have not been previously introduced in the geography literature at the time of this writing. The LARS algorithm is similar in spirit to forward stepwise regression, which I now describe. The forward stepwise regression algorithm is:

(1) Start with all coefficients β_k equal to zero and set $\mathbf{r} = \mathbf{y}$.

(2) Find the predictor x_k most correlated with the residuals $\mathbf{r}$ and add it to the model.

(3) Calculate the residuals $\mathbf{r} = \mathbf{y} - \hat{\mathbf{y}}$.

(4) Continue steps 2-3 until all predictors are in the model

While the LARS algorithm is described in detail algebraically in Efron et al. (2004a), Efron et al. (2004b) restate the LARS algorithm as a purely statistical one with repeated fitting of the residuals, similar to the forward stepwise regression algorithm. The statistical statement of the LARS algorithm is:

(1) Start with all coefficients β_k equal to zero and set $\mathbf{r} = \mathbf{y}$.

(2) Find the predictor x_k most correlated with the residuals $\mathbf{r}$.

(3) Increase the coefficient β_k in the direction of the sign of its correlation with $\mathbf{r}$, calculating the residuals $\mathbf{r} = \mathbf{y} - \hat{\mathbf{y}}$ at each increase, and continue until some other predictor x_m has as much correlation with the current residual vector $\mathbf{r}$ as does predictor x_k.

(4) Update the residuals and increase (β_k, β_m) in the joint least squares direction for the regression of $\mathbf{r}$ on (x_k, x_m) until some other predictor x_j has as much correlation with the current residual $\mathbf{r}$.

(5) Continue steps 2-4 until all predictors are in the model. Stop when $\text{corr}(r, x_j) = 0 \ \forall j$, the OLS solution.

As with ridge regression, typically the response variable is centered and the explanatory variables are centered and scaled to have equal (unit) variance prior to starting the LARS algorithm. In other words, $\sum_{i=1}^{n} y_i = 0$, $\sum_{i=1}^{n} x_{ij} = 0$, and $\sum_{i=1}^{n} x_{ij}^2 = 1$ for $j = 1, \ldots, m$. Efron et al. (2004a) show that a small modification to the LARS algorithm yields the lasso solutions. In a lasso solution, the sign of any nonzero coefficient β_k must agree with the sign of the current correlation of x_k and the residual. The LARS algorithm does not enforce this, but Efron and coauthors modify the algorithm to do so by removing β_k from the lasso solution if it changes in sign from the sign of the correlation of x_k and the current residual. This modification means that in the lasso solution, the active set of variables in the solution does not necessarily monotonically increase as the routine progresses. Therefore, the LARS algorithm typically takes less iterations than does the

lasso algorithm. The modified LARS algorithm produces the entire range of possible lasso solutions, from the initial solution with all coefficients equal to zero, to the final solution, which is also the OLS solution.

In some of the lasso algorithms, such as the modified LARS algorithm and the algorithm Tibshirani describes, the shrinkage parameter s (or t) must be estimated before finding the lasso solutions. Hastie et al. (2001) estimate the parameter

$$s = \frac{\sum_{k=1}^{p} \left| \hat{\beta}_k^{ols} \right|}{t} \tag{5.7}$$

through ten-fold cross-validation, where t is some positive scalar that reduces the ordinary least squares coefficient estimates. Tibshirani (1996) uses five-fold cross-validation, generalized cross-validation, and a risk minimizer to estimate the parameter t, with the computational cost of the three methods decreasing in the same order. Efron et al. (2004a) also recommend using cross-validation to estimate the lasso parameter. If t is one or less, there is no shrinkage and the lasso solutions for the coefficients are the least squares solutions. One can also define the lasso shrinkage parameter as

$$s = \frac{\sum_{k=1}^{p} \left| \hat{\beta}_k \right|}{\sum_{k=1}^{p} \left| \hat{\beta}_k^{ols} \right|}, \tag{5.8}$$

and s ranges from 0 to 1, where 0 corresponds to the initial lasso solution with all regression coefficients shrunk to 0 and 1 corresponds to the final lasso solution, which is also the OLS solution. Then, s can be viewed as the fraction of the OLS solution that is the lasso solution. This is the definition of the lasso shrinkage parameter that I will use in the subsequent work in this dissertation.

Geographically Weighted Lasso

The lasso can be implemented in GWR relatively easily, and the result is here called the geographically weighted lasso (GWL). An efficient implementation of the GWL outlined here uses the `lars` function from the package of the same name written in the R language by Hastie and Efron (see the R Project web site: http://cran.r-project.org/). The `lars` function implements the LARS and lasso methods, where the lasso is the default method, and details are described in Efron et al. (2004a; 2004b). To make use of the `lars` function in the GWR framework, the x and y variables input to the function must be weighted by the kernel weights at each model calibration location. The `lars` function must be run at each model calibration location. This can be done in one of two ways: separate models with local scaling of the explanatory variables or one model with global scaling of the explanatory variables. The first way, local scaling, requires n calls of the `lars` function, one for each location, and the weighted x and y are centered and the x variables are scaled by the norm in the `lars` function. This effectively removes the intercept and equates the scales of the explanatory variables to avoid the problem of different scales (this problem is also avoided in GWRR). The local scaling version estimates the lasso parameter to control the amount of coefficient shrinkage at each calibration location, so there is a shrinkage parameter s_i estimated at

each location i. Since I am working here in the GWR framework, I will estimate the model shrinkage and kernel bandwidth parameters using leave-one-out cross-validation while minimizing the RMSPE. Therefore, the n s_i parameters and the kernel bandwidth ϕ must be estimated in GWL with CV before the final lasso coefficient solutions are estimated. I have chosen to estimate these parameters simultaneously, as the lasso solution will certainly depend on the kernel bandwidth. The algorithm to estimate the local scaling GWL parameters using cross-validation is:

- For each attempted bandwidth ϕ in the binary search for the lowest RMSPE
    - Calculate $\mathbf{W}$ using ϕ.
    - For each location i from 1, …, n
        - Set $\mathbf{W}^{1/2}(i)_{ii} = 0$, that is, set the (i,i) element of the diagonal weights matrix to 0 to effectively remove observation i
        - Set $\mathbf{X}_w = \mathbf{W}^{1/2}(i)\mathbf{X}$ and $\mathbf{y}_w = \mathbf{W}^{1/2}(i)\mathbf{y}$ using the square root of the kernel weights $\mathbf{W}(i)$ at location i.
        - Call `lars`($\mathbf{X}_w, \mathbf{y}_w$), save the series of lasso solutions, find the lasso solution that minimizes the error for y_i, and save this solution.
- Stop when there is only a small change in the estimated ϕ. Save the estimated ϕ.

In the previous algorithm, saving the lasso solution entails saving the estimated shrinkage fraction s_i at each location, as well as an indicator vector $\mathbf{b}$ of which variable coefficients are shrunken to zero. The algorithm uses a binary search to find the ϕ that minimizes the RMSPE. The small change in ϕ is set exogenously.

The algorithm to estimate the final local scaling GWL solutions after cross-validation estimation of the shrinkage and kernel bandwidth parameters is:

- Calculate $\mathbf{W}$ using ϕ.
- For each location i from 1, …, n
    - Set $\mathbf{X}_w = \mathbf{W}^{1/2}(i)\mathbf{X}$ and $\mathbf{y}_w = \mathbf{W}^{1/2}(i)\mathbf{y}$ using the square root of the kernel weights $\mathbf{W}(i)$ at location i.
    - Call `lars`($\mathbf{X}_w$, $\mathbf{y}_w$) and save the series of lasso solutions.
    - Select the lasso solution that matches the cross-validation solution according to the fraction s_i and the indicator vector $\mathbf{b}$.

The second GWL method, global scaling, calls the `lars` function only one time, using specially structured input data matrices. This method fits all the local models at once, using global scaling of the x variables. It also estimates only one lasso parameter to control the amount of coefficient shrinkage. The weighted design matrix for the global version is a $(n \cdot n) \times (n \cdot p)$ matrix and the weighted response vector is $(n \cdot n) \times 1$. This results in a $(n \cdot p) \times 1$ vector of estimated regression coefficients, as was the case in the Bayesian SVCP model. The weighted design matrix is such that the design matrix is

repeated n times, shifting p columns in its starting position each time it is repeated. The kernel weights for the 1st location are applied to the first n rows of the matrix, the weights for the 2nd location are applied to the next n rows of the matrix, and so forth. The weighted response vector has the response vector repeated n times, with the weights for the 1st location applied to the first n elements of the vector, and so on. The algorithm to estimate the global scaling GWL parameters using cross-validation is:

- For each attempted bandwidth ϕ in the binary search for the lowest RMSPE
    - Calculate $\mathbf{W}$ using ϕ.
    - Set $\mathbf{y}_{\mathbf{w}}^{G} = \mathbf{W}^{1/2} \times (\mathbf{1} \cdot \mathbf{y}^{T})$ using the square root of the kernel weights matrix $\mathbf{W}$ and the column unity vector $\mathbf{1}$ of length n. The operator $\times$ indicates element-by-element multiplication here. Set $k = 1$ and $m = 1$.
    - For each location i from 1, …, n
        - Set $j = k \cdot n - (n-1)$ and $l = m \cdot p - (p-1)$.
        - Set $\mathbf{X}_{\mathbf{w}} = \mathbf{W}^{1/2}(i)\mathbf{X}$ using the square root of the kernel weights $\mathbf{W}(i)$ at location i. Set $\mathbf{X}_{\mathbf{w}}^{G}(j : n \cdot k,\ l : p \cdot m) = \mathbf{X}_{\mathbf{w}}$.
        - Set $k = k + 1$ and $m = m + 1$.
    - Call `lars`$(\mathbf{X}_{\mathbf{w}}^{G}, \mathrm{vec}(\mathbf{y}_{\mathbf{w}}^{G}))$ and save the series of lasso solutions, where the vec() operator turns a matrix into a vector.

In the previous algorithm, saving the lasso solution entails saving the estimated overall shrinkage fraction s, as well as a vector $\mathbf{b}$ that indicates which of the variable coefficients are shrunken to zero. The algorithm uses a binary search to find the ϕ that minimizes the RMSPE. The small change in ϕ is again set exogenously.

The algorithm to estimate the final global scaling GWL solutions after cross-validation estimation of the shrinkage and kernel bandwidth parameters is:

- Calculate $\mathbf{W}$ using ϕ.
- Set $\mathbf{y}_{\mathrm{w}}^{G} = \mathbf{W}^{1/2} \times (\mathbf{1} \cdot \mathbf{y}^{T})$ using the square root of the kernel weights matrix $\mathbf{W}$ and the column unity vector $\mathbf{1}$ of length n. The operator $\times$ indicates element-by-element multiplication here. Set $k = 1$ and $m = 1$.
- For each location i from 1, …, n
    - Set $j = k \cdot n - (n-1)$ and $l = m \cdot p - (p-1)$.
    - Set $\mathbf{X}_{\mathrm{w}} = \mathbf{W}^{1/2}(i)\mathbf{X}$ using the square root of the kernel weights $\mathbf{W}(i)$ at location i. Set $\mathbf{X}_{\mathrm{w}}^{G}(j : n \cdot k,\, l : p \cdot m) = \mathbf{X}_{\mathrm{w}}$.
    - Set $k = k + 1$ and $m = m + 1$.
- Call `lars`$(\mathbf{X}_{\mathrm{w}}^{G}, \mathrm{vec}(\mathbf{y}_{\mathrm{w}}^{G}))$ and save the series of lasso solutions, where vec() turns the matrix into a vector.
- Select the lasso solution that matches the cross-validation solution according to the fraction s and the indicator vector $\mathbf{b}$.

In comparing the local and global scaling GWL algorithms, the global GWL algorithm requires more computational time due to the matrix inversion of a much larger matrix. The global GWR algorithm must invert a $(n \cdot p \times n \cdot p)$ matrix, while the local GWR algorithm must invert a $(p \times p)$ n times, which is clearly faster. Considering that calculating the inverse of a general $j \times j$ matrix takes between $O(j^2)$ and $O(j^3)$ time (Banerjee et al. 2004), there can be quite a difference in the computation time for the two versions of GWR. Table 5.1 shows the number of computational units required for the global and local GWL methods for various values of n and p. When n is large, global GWL can take more than two times the computation time of local GWL. In comparing the expected performance of the global and local versions of GWL, the local GWL method should produce lower prediction errors than the global GWL method, as adding more shrinkage parameters generally increases model stability and hence lowers prediction error. In summary, the local GWL should be faster than the global GWL and should have lower prediction error. The benefit of global GWL may be in lower RMSE of the regression coefficients. The local and global versions of GWL will be compared empirically in the simulation studies of the next chapter.

n	p	$(np)^2$	$n(p)^2$	$(np)^3$	$n(p)^3$
100	2	4.0E+04	4.0E+02	8.0E+06	8.0E+02
100	10	1.0E+06	1.0E+04	1.0E+09	1.0E+05
1000	2	4.0E+06	4.0E+03	8.0E+09	8.0E+03
1000	10	1.0E+08	1.0E+05	1.0E+12	1.0E+06
10000	2	4.0E+08	4.0E+04	8.0E+12	8.0E+04
10000	10	1.0E+10	1.0E+06	1.0E+15	1.0E+07

Table 5.1. Number of operations to calculate the matrix inverse in global and local GWL.

Ridge Regression and the Lasso as Bayes Estimates

Based on discussions in the literature, it is possible to view the ridge regression and lasso solutions as Bayes estimates. Works by Lindley and Smith (1972) and Goldstein (1976) show that ridge regression coefficient estimates may be viewed as Bayesian regression coefficient posterior means under specific vague priors. Hastie et al. (2001) view ridge regression and the lasso more generally as Bayes estimates with different prior distributions, where the lasso estimate uses independent double-exponential priors for each β_k and ridge regression uses independent normal distributions for each coefficient prior. Tibshirani (1996) also illustrates the differing priors for lasso and ridge regression. These authors acknowledge that the lasso solution is derived from the mode of the posterior distribution and the ridge regression solution is derived from the mean of the posterior distribution for the coefficient (it is also the mode because the posterior distribution is Gaussian).

For ridge regression, Hastie et al. (2001) point out that if the prior for each regression coefficient β_k is $N(0,\sigma^2)$, independent of the others, then the negative log posterior density of the regression coefficients $\boldsymbol{\beta}$ is equal to the expression in the braces in the ridge regression coefficient equation (5.1), with $\lambda = \tau^2/\sigma^2$, where τ^2 is the error variance. This particular Gaussian prior does not depend on direction of the regression coefficient, but instead only length, which implies that ridge regression achieves coefficient shrinkage to counter correlation present in $\mathbf{X}$, not through prior information that favors high-variance directions. In the Bayesian regression model with

$\mathbf{y} \sim N(\mathbf{X\beta}, \tau^2\mathbf{I})$ and the independent prior $\beta \sim N(0, \sigma^2)$ for each coefficient, the posterior for the coefficients can be expressed as

$$[\boldsymbol{\beta} \mid \tau^2, \sigma_\beta^2; y] \propto \exp\left(-\frac{1}{2\tau^2} \sum_{i=1}^{n} (y_i - \sum_{k=1}^{p} x_{ik}\beta_k)^2 \right) \exp\left(-\frac{1}{2\sigma^2} \sum_{k=1}^{p} (\beta_k - 0)^2 \right), \tag{5.9}$$

where for convenience of notation the variables have been centered. The negative log posterior density of $\boldsymbol{\beta}$ up to a constant is then found through algebra to be

$$\sum_{i=1}^{n} (y_i - \sum_{k=1}^{p} x_{ik}\beta_k)^2 + \frac{\tau^2}{\sigma^2} \sum_{k=1}^{p} \beta_k{}^2, \tag{5.10}$$

with the ridge shrinkage parameter $\lambda = \frac{\tau^2}{\sigma^2}$. This illustrates that the ridge regression estimate is the mean of the posterior distribution with a Gaussian prior and Gaussian data model, and that the ridge shrinkage parameter is a ratio of the error variance and common regression coefficient variance.

The view of ridge regression solutions as Bayes estimates suggests that the Bayesian SVCP model coefficients can be viewed as ridge regression estimates because of the normal distribution prior for the regression coefficients in the SVCP model. Granted, the prior in the SVCP is more complicated than the independent normal prior in the traditional Bayesian regression model due to the spatial component in the covariance matrix, but it is a normal prior nonetheless. With the Bayesian SVCP model, assuming centering to remove the intercept for convenient notation,

$$\begin{aligned} &y \sim N(\mathbf{X\beta}, \tau^2\mathbf{I}) \\ &\boldsymbol{\beta} \sim N(\mathbf{1} \otimes \boldsymbol{\mu}_{\beta}, \boldsymbol{\Sigma}_{\beta}) \\ &\boldsymbol{\Sigma}_{\beta} = \mathbf{H}(\phi) \otimes \mathbf{T}, \end{aligned} \tag{5.11}$$

the posterior distribution for the coefficients can be expressed as

$$\begin{aligned} [\beta \mid \tau^2, \boldsymbol{\Sigma}_{\beta}; y] &\propto \exp\left(-\frac{1}{2}(\mathbf{y} - \mathbf{X\beta})^T (\tau^2\mathbf{I})^{-1} (\mathbf{y} - \mathbf{X\beta}) \right) \\ &\times \exp\left(-\frac{1}{2}(\boldsymbol{\beta} - (\mathbf{1} \otimes \boldsymbol{\mu}_{\beta}))^T \boldsymbol{\Sigma}_{\beta}^{-1} (\boldsymbol{\beta} - (\mathbf{1} \otimes \boldsymbol{\mu}_{\beta})) \right) \\ &\propto \exp\left(-\frac{1}{2}[(\mathbf{y} - \mathbf{X\beta})^T (\mathbf{y} - \mathbf{X\beta}) + (\boldsymbol{\beta} - (\mathbf{1} \otimes \boldsymbol{\mu}_{\beta}))^T \tau^2 (\mathbf{H}(\phi) \otimes \mathbf{T})^{-1} (\boldsymbol{\beta} - (\mathbf{1} \otimes \boldsymbol{\mu}_{\beta}))] \right). \end{aligned} \tag{5.12}$$

The negative log posterior density of $\boldsymbol{\beta}$ up to a constant is then

$$(\mathbf{y} - \mathbf{X\beta})^T (\mathbf{y} - \mathbf{X\beta}) + (\boldsymbol{\beta} - (\mathbf{1} \otimes \boldsymbol{\mu}_{\beta}))^T \tau^2 (\mathbf{H}(\phi) \otimes \mathbf{T})^{-1} (\boldsymbol{\beta} - (\mathbf{1} \otimes \boldsymbol{\mu}_{\beta})), \tag{5.13}$$

where the shrinkage term is unconventionally a matrix $\boldsymbol{\lambda}$ that is calculated as

$$\tau^2(\mathbf{H}(\phi)\otimes\mathbf{T})^{-1} = \tau^2\cdot\mathbf{H}^{-1}(\phi)\otimes\mathbf{T}^{-1}. \quad (5.14)$$

Therefore, the amount of shrinkage on $\boldsymbol{\beta}$ towards the mean $\boldsymbol{\mu}_{\boldsymbol{\beta}}$ depends on τ^2, ϕ, and $\mathbf{T}$ in the SVCP model.

With a modification to the Bayes SVCP model regression coefficient prior to have a mean of 0 for all coefficients, $\boldsymbol{\beta} \sim N(\mathbf{0}, \boldsymbol{\Sigma}_{\boldsymbol{\beta}})$, the negative log posterior density of $\boldsymbol{\beta}$ up to a constant is

$$(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})^T(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}) + \boldsymbol{\beta}^T\tau^2(\mathbf{H}(\phi)\otimes\mathbf{T})^{-1}\boldsymbol{\beta}. \quad (5.15)$$

Therefore, the penalty on $\boldsymbol{\beta}$ with shrinkage towards zero looks like a ridge regression penalty with shrinkage to zero and depends on τ^2, ϕ, and $\mathbf{T}$.

Bayesian SVCP Model Coefficient Shrinkage Example

In Chapter 4, I fitted a GWR model for white male bladder cancer mortality in 508 State Economic Areas (SEAs) in the United States for the years 1970 to 1994 using a smoking proxy and log population density as explanatory variables. As indicated in the discussion in that chapter, collinearity appears to be a problem with these data when used in a GWR model. The GWR estimated coefficients were first mapped in Figure 4.2. The GWR estimated coefficients for these bladder cancer data are negative for each of the explanatory variables in some parts of the study area and the coefficients for the two

variables exhibit moderate to strong overall correlation. The coefficients for population density are negative for most of the Northeast. These negative coefficients are counter to previous studies, intuition, and the traditional regression estimates. As Lindley and Smith (1972) point out, when data are correlated, least-squares regression can "produce regression estimates which are too large in absolute value, of incorrect sign and unstable with respect to small changes in the data." The weighted least-squares estimation procedure of GWR likely suffers from the same condition.

To illustrate the idea of the Bayesian SVCP model coefficients as ridge regression-type shrinkage solutions with a practical example, I use the same explanatory variables as earlier in a SVCP model to explain male bladder cancer mortality rates in the SEAs of the United States. The model consists of an intercept term and the explanatory variables log population density and lung cancer mortality rate as a smoking proxy. The model is

$$y(s) = \beta_0(s) + \beta_1(s) \cdot \text{SMOKE}(s) + \beta_2 \cdot \text{LNPOP}(s) + \varepsilon(s). \tag{5.16}$$

To estimate the model parameters, I use 2000 iterations in the MCMC, with a "burn-in" of 1000 iterations. Based on trace plots and Gelman's $\hat{R}$ statistic (e.g. Gelman et al. 2003), the regression coefficients converged within 1000 iterations of the Gibbs sampler. The prior specification for this model is as follows. I use a vague normal, $N(\mathbf{0}, 10^4\mathbf{I})$, for $\boldsymbol{\mu}_\beta$, a four-dimensional inverse Wishart, $IW(4, .1\cdot\mathbf{I})$, for $\mathbf{T}$, and an inverse gamma, $IG(1, .01)$, for τ^2, where $\mathbf{I}$ is the identity matrix of dimension p. For the spatial dependence parameter ϕ, I use a gamma, $G(.103, .01)$, which has a mean of 10.37 and variance of 1037.

The SVCP model coefficients are mapped in Figure 5.1 and are all non-negative for smoking and non-negative for population density for all but two of the 508 SEAs, where the two negative population density SEAs are very close to zero. In contrast to the GWR coefficients, the SVCP model estimated coefficients do not immediately indicate a considerable problem with collinearity. While there is some overall correlation in the SVCP model coefficients for the two variables, the complimentary pattern is not as strong as with the GWR coefficients. In addition, the variance of the coefficients is not as large with the SVCP model. The SVCP model here achieves a similar penalization effect to that of ridge regression, but has the advantage that the shrinkage parameters are estimated from the data and not through a separate estimation procedure such as cross-validation.

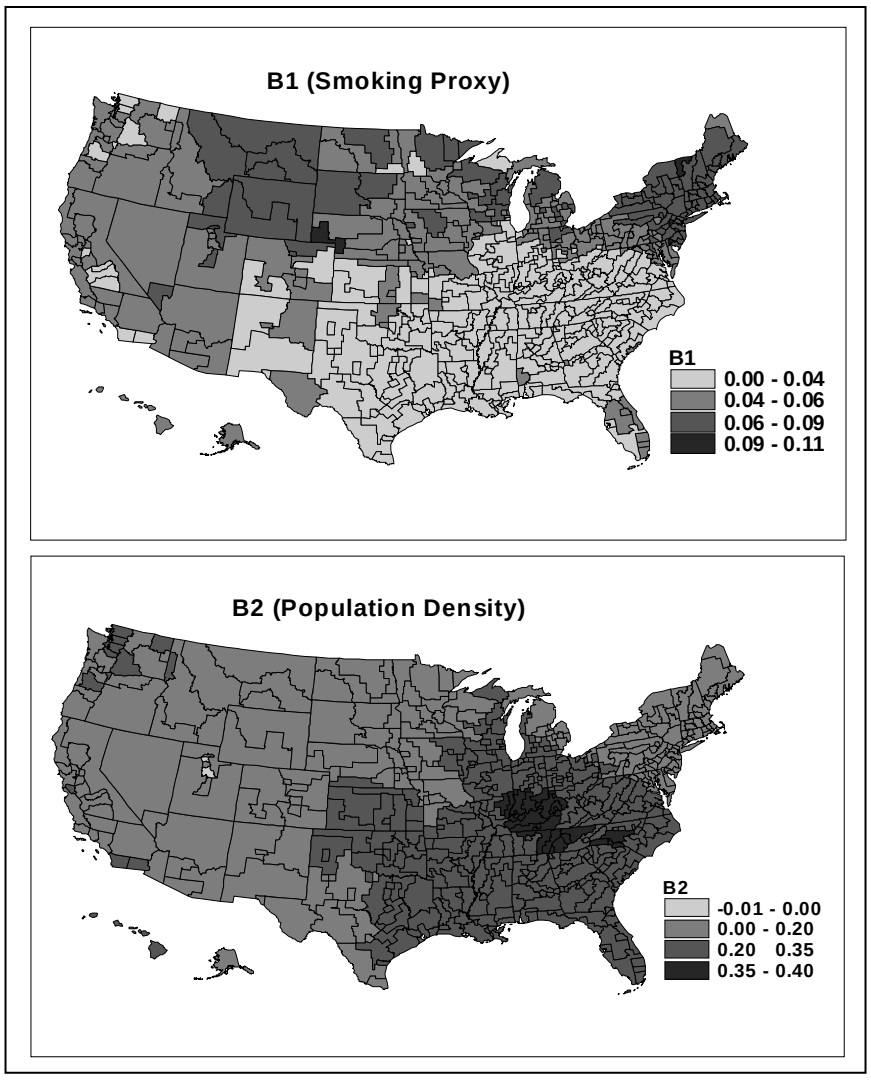

Figure 5.1. Estimated coefficients for smoking proxy (top) and population density (bottom) for the SVCP model

Geographically Weighted Ridge Regression Example

Since the VIFs and variance-decomposition proportions for the Columbus crime dataset in the previous chapter indicate problems with collinearity in parts of the study area, the dataset is a good candidate for applying the remedial method of GWRR. I now apply the GWRR model and emphasize some of its properties with the Columbus crime dataset. To fit the GWRR model using scheme 1, I first estimate the kernel bandwidth and then the ridge parameter. The estimated bi-square nearest neighbor kernel bandwidth $N = 11$ and the estimated ridge parameter $\lambda = 0.80$ with estimation scheme 1. The scheme 3 solution is the same as the scheme 1 solution in this case. The prediction error, as measured by the root mean squared prediction error (RMSPE), and the estimation error, as measured by the root mean squared error (RMSE), are plotted for the GWRR model in Figure 5.2, along with the RMSPE for the GWR model, as functions of the kernel bandwidth. The prediction error is the error using cross-validation and the estimation error uses all the observations in the model calibration. The estimation error is a global measure in that it uses the squared deviation of only the $\hat{y}(i)$ from $y(i)$ in each model at location i. The deviations from the $y(j)$ in each model at location i are not utilized.

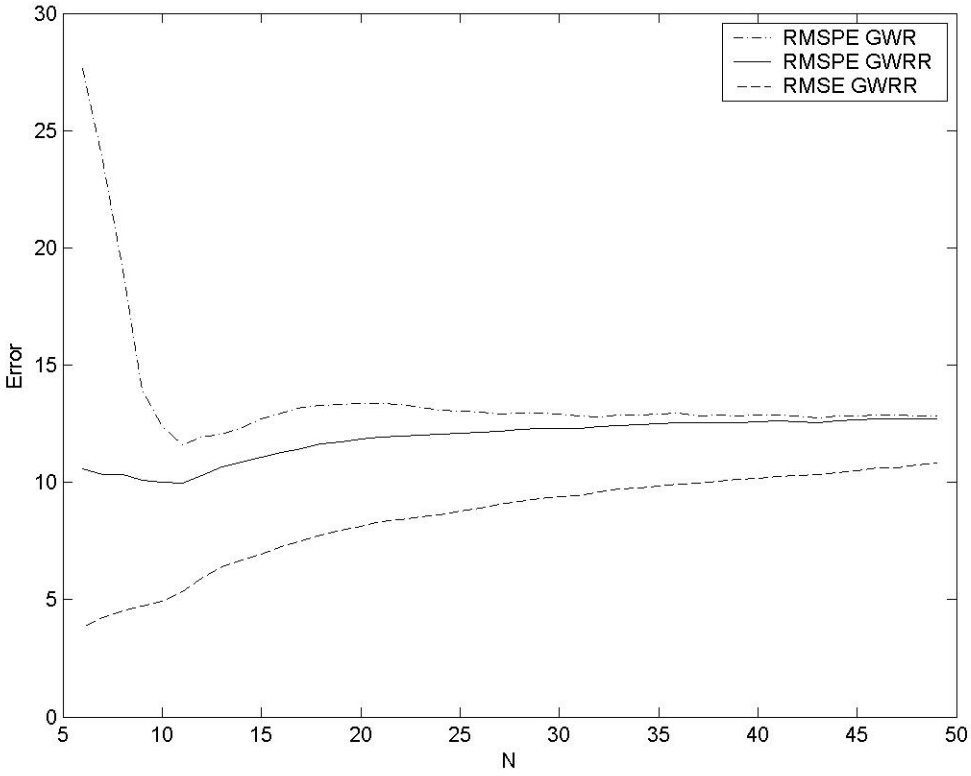

Figure 5.2. Prediction error (RMSPE) for the GWR (λ = 0) and GWRR (λ = 0.80) solutions and the estimation error (RMSE) for the GWRR solution as a function of N.

Figure 5.2 demonstrates that the estimation error is lower than the prediction error for GWRR, and this is also true for GWR. The higher prediction error is due to removing the maximally weighted observation, i, in the cross-validation for observation i, for this observation always has a weight of 1. When this observation is added back into the data used for estimating the coefficients in the model at location i, the fit naturally improves. Note that the behavior of the prediction error as N increases is stable in Figure 5.2. This behavior can be problematic for search routines and care is needed in the estimation of N to make sure a local solution with a larger N than is necessary is not selected. This is

important because initial results show that the overall GWR model fit decreases with increasing N. Note that in general there will be a perfect fit if $N \leq p$ because there are fewer observations than number of coefficients, and, for the bi-square nearest neighbor kernel function, the fit is perfect for $N \leq p+1$ because the weight of the N^{th} observation is 0. Figure 5.2 also shows that the GWRR solution has a lower prediction error than the GWR solution. This is congruous with the idea in the statistical learning literature that prediction error can usually be improved in global models by introducing some bias to reduce variability in the predictions.

Figure 5.3 shows the prediction error for a truncated range of kernel bandwidth values and four ridge parameter values. The value of λ = 0.8 corresponds to the GWRR solution. A value of 1.4 for λ serves as an upper bound reference, as no value larger than this for λ produces an optimal solution. A value of 0.0 for λ corresponds to the GWR model. The interaction of N and λ in the prediction error is apparent in the figure for λ = 0.8 and λ = 1.4. This figure shows that various values of λ improve on the GWR prediction error and that the best λ depends on N. In general, this should be the case. This evidence provides an argument for generally using scheme 3 to estimate N and λ simultaneously, however, as the scheme 1 and scheme 3 solutions are the same in this case, one might argue it is not always worth the additional computational cost.

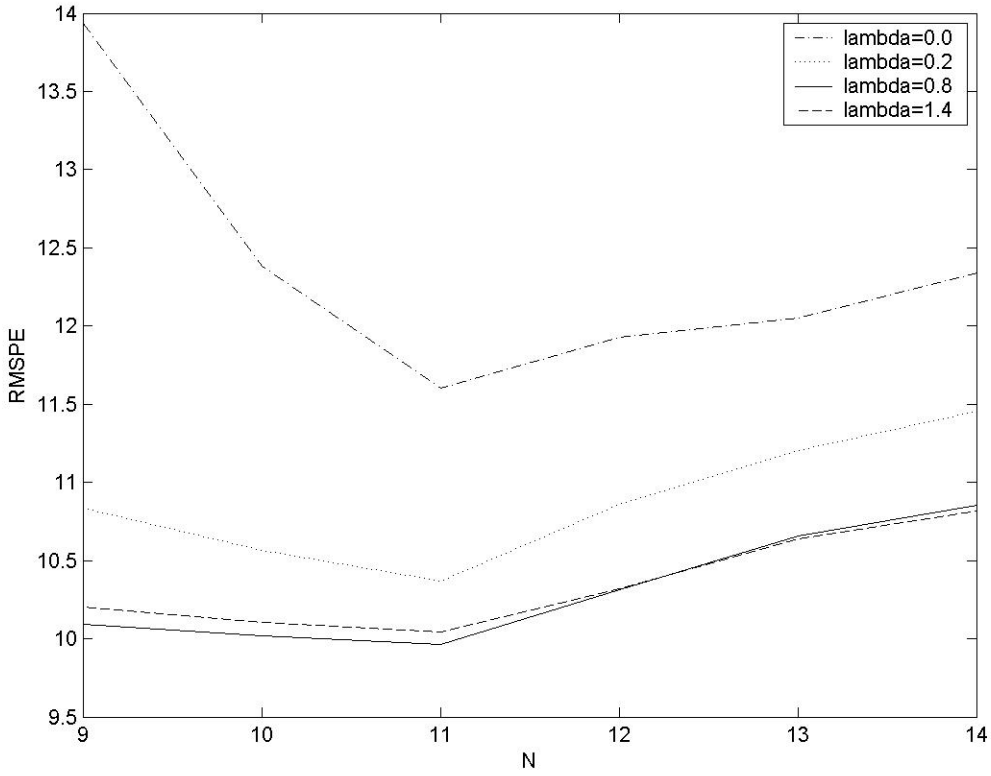

Figure 5.3. Prediction error as a function of N and λ for a truncated range of N and selected values of λ. Lambda = 0 is the GWR solution and lambda = 0.80 is the GWRR solution. Two other lambda values (0.2 and 1.4) illustrate the function behavior.

The GWRR model estimates from local centering are listed in Table 5.2. The results show a decrease in the mean local coefficient correlation and the global coefficient correlation from the GWR model. The overall correlation coefficient between the two sets of estimated regression coefficients decreases from –0.80 to -0.53 and the mean local coefficient correlation decreases from –0.58 to –0.01. The overall model fit as measured by R^2 decreases only mildly from 0.92 to 0.90. Table 5.2 also shows that for the GWRR solution, the mean coefficient estimate for housing value becomes more

negative and the mean coefficient estimate for income becomes less negative for income than with the GWR model. The Moran's I statistic is 0.054 for the GWR residuals and is 0.026 for the GWRR residuals. These statistics are not significant and indicate that there is no significant spatial autocorrelation in the GWR residuals in this case, and that the inclusion of the coefficient penalization does not significantly affect the spatial autocorrelation level in the model residuals.

	Unstandardized				Standardized
Parameter	Mean Estimate	Mean VIF	Mean Parameter Correlation	Global Parameter Correlation	Mean Estimate
Intercept	55.465				0.000
Inc	-0.745		-0.012	-0.530	-0.254
Hoval	-0.186		-0.012	-0.530	-0.205
R-square	0.90				

Table 5.2. The table lists the GWRR model summary for the Columbus crime dataset.

The regression coefficients for the GWRR models using local centering and global centering under estimation scheme 1 are plotted in Figure 5.4. The pattern of coefficients is similar for the two solutions, although there is overall more coefficient shrinkage in the local centering solution with a smaller ridge parameter value. The observations with the large VIFs and estimated GWR coefficients in the upper left corner of the corresponding figure in the previous chapter have been penalized in the GWRR solutions to now reside in the main grouping of observations that have intuitively signed coefficients. The regression coefficients for the GWR model and the GWRR local

centering model are plotted in Figure 5.5. The effect of the ridge parameter on the estimated coefficients is more clear in this figure. The coefficients have been reduced away from the positive values they had in the GWR model, especially for the housing value variable (β_2). The coefficients now have more intuitive signs considering the response variable of crime.

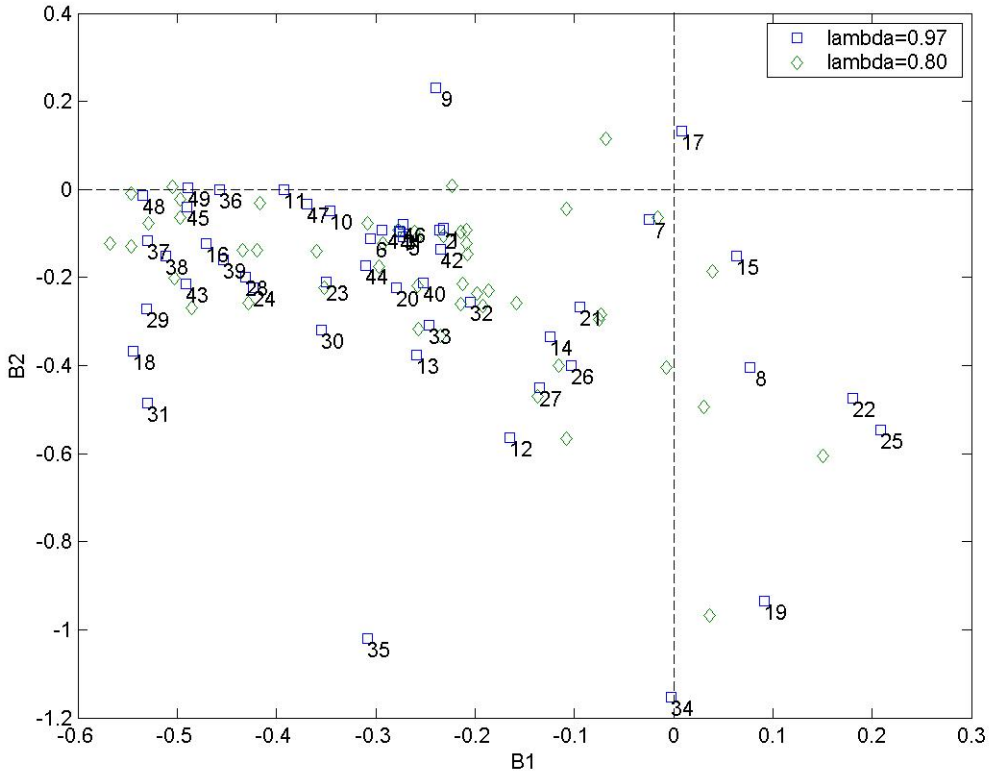

Figure 5.4. Estimated regression coefficients for the GWRR local centered (lambda = 0.80) and global centered solutions (lambda = 0.97) with observation identifiers.

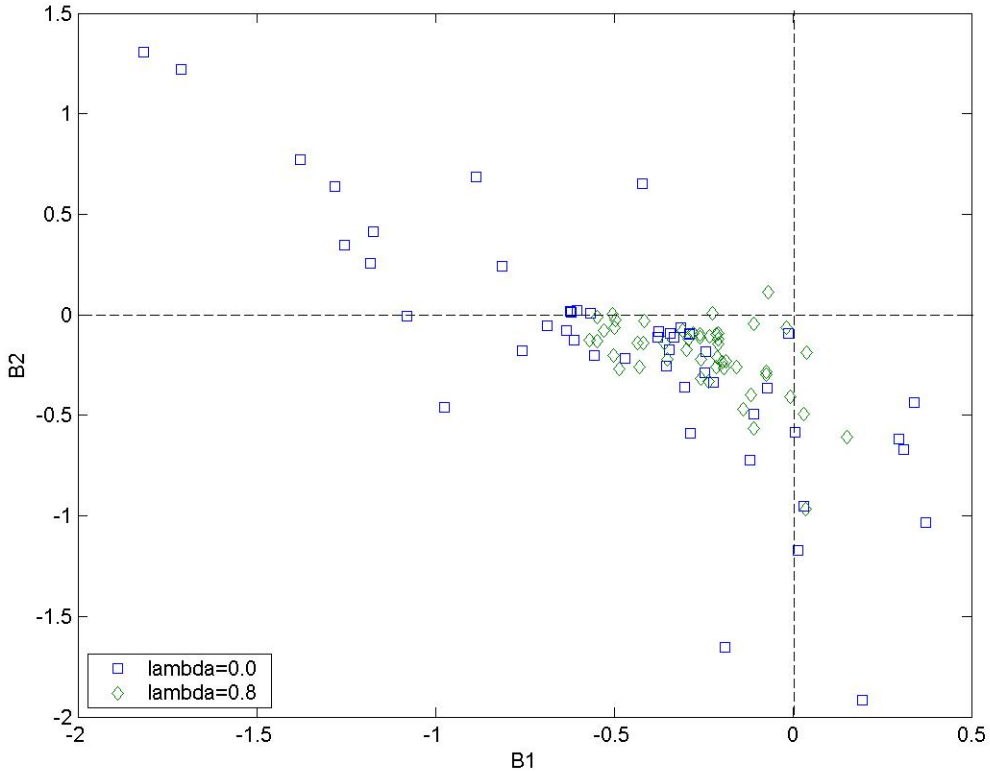

Figure 5.5. GWR (lambda = 0.0) and GWRR (lambda = 0.8) estimated regression coefficients using local centering.

The regression coefficients for the locally centered GWR model are mapped in Figure 5.6 and the corresponding GWRR coefficients are mapped in Figure 5.7. The dependence in the GWR regression coefficients in the form of negative association is clear in Figure 5.6. The areas with counter-intuitive positive regression coefficients for income are not the same areas with counter-intuitive positive regression coefficients for housing value. The two maps in Figure 5.7 show less of a complementary pattern as those in Figure 5.6, with fewer areas that have light-shaded values for the housing value parameter when the income parameter is dark-shaded (most negative), and vice-verse. The strong negative association in the GWR coefficients in the east-central portion of the study area in Figure 5.6 has been especially reduced in the GWRR coefficients in Figure 5.7.

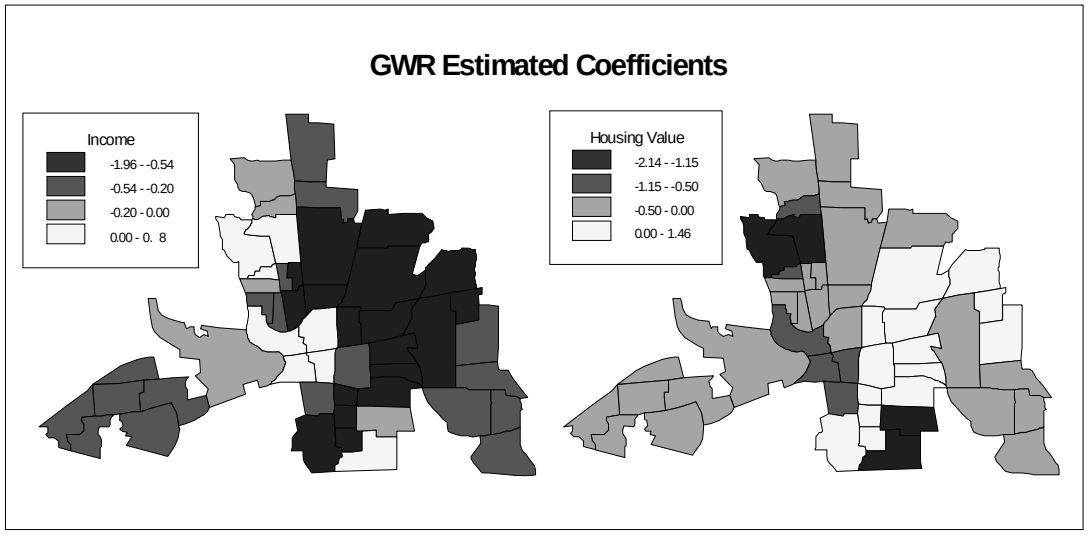

Figure 5.6. Estimated regression coefficients for the GWR model.

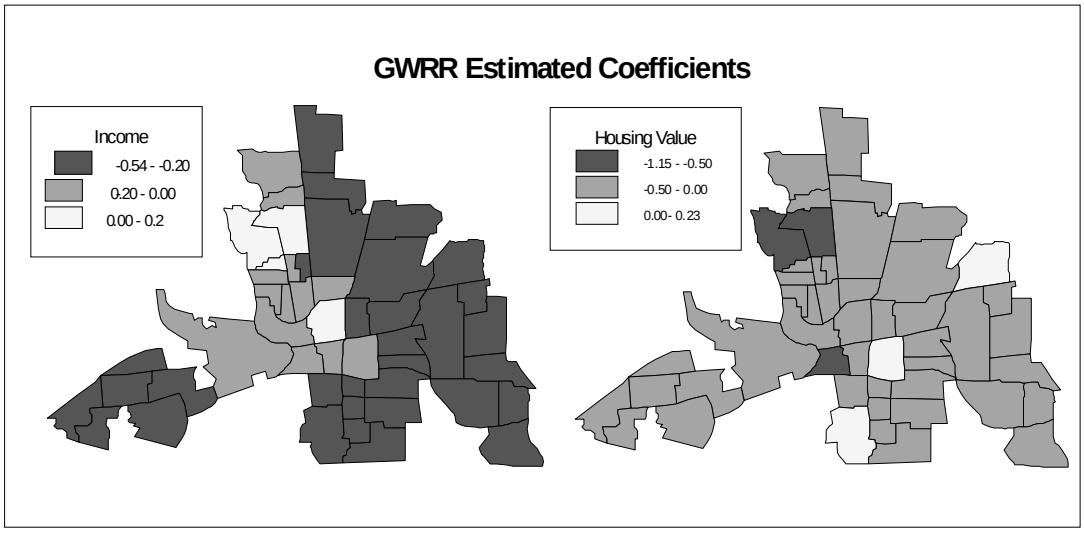

Figure 5.7. Estimated regression coefficients for the GWRR model.

A preliminary experiment to evaluate how the GWRR model responds to increasing collinearity in the explanatory variables shows the model to be quite robust to extremely collinear variables. The experiment involved increasing the level of collinearity in the Columbus crime model from the original level by replacing the standardized housing value variable in the model by a weighted combination of the standardized income and housing value variables. The new variable, $x_2^{'}$, is calculated from the standardized variables x_1^* and x_2^* as

$$x_2^{'} = ax_1^* + (1-a)x_2^*, \tag{5.17}$$

where a is a weighting scalar between 0 and 1 that controls the amount of correlation in the standardized explanatory variables. Table 5.3 contains the summary results of the experiment. The correlation in the variables ranges from 0.50 to 0.99 and the results correspond to the GWR model when $\lambda = 0$ and correspond to the GWRR model for non-zero λ. The kernel bandwidth is fixed at $N = 11$ for all values of λ to eliminate a source of variation in the experiment even though the optimal N would likely change with λ at different levels of variable correlation. The table shows that for strongly collinear variables, the GWR model behaves in such a way that it pushes the coefficients apart from one another while the GWRR model properly reflects the relationship of the variables in that the mean coefficients that are still negative and are about equal, which reflects the association structure between the explanatory variables and the response variable. Naturally, λ increases as the variable correlation increases to reduce the

coefficient variances. While the correlation levels at the bottom of the table are admittedly extreme, they are helpful in revealing the behavior of the two methods and showing the benefit of using GWRR with collinear variables, even when only two explanatory variables are included in the model.

Weight	Variable Correlation	λ	Mean Coefficient Correlation	Global Coefficient Correlation	Mean $\hat{\beta}_1$	Mean $\hat{\beta}_2$
0.00	0.50	0.00	-0.58	-0.80	-0.48	-0.17
		0.80	-0.01	-0.53	-0.25	-0.21
0.40	0.63	0.00	-0.68	-0.86	-0.44	-0.19
		0.84	-0.02	-0.55	-0.23	-0.22
0.60	0.74	0.00	-0.76	-0.91	-0.40	-0.22
		0.91	0.10	-0.57	-0.20	-0.24
0.80	0.89	0.00	-0.88	-0.97	-0.27	-0.33
		1.16	0.35	-0.48	-0.16	-0.24
0.90	0.97	0.00	-0.96	-0.99	0.01	-0.57
		1.51	0.68	0.05	-0.15	-0.21
0.95	0.99	0.00	-0.99	-1.00	0.51	-1.08
		1.77	0.89	0.72	-0.15	-0.18

Table 5.3. The table lists the mean local regression coefficient correlation and global regression coefficient correlation in the GWRR and GWR models at various levels of correlation in the explanatory variables. The weight determines the amount of variable correlation and $\lambda = 0$ corresponds to the GWR model.

CHAPTER 6

SIMULATION STUDY

In this chapter, I use simulation studies to evaluate and compare the coverage probabilities and accuracy of the regression coefficients from multiple spatially varying coefficient models. I assess inferences on the coefficients both when there is no collinearity in the explanatory variables and when there is collinearity, expressed at various levels. The motivation for doing this is to explicitly test the assumption that the inferences on the coefficients from each model are valid, in the sense that the 95% confidence interval or credible interval for each estimated coefficient contains the true value 95 percent of the time. The Bayesian 95% credible interval for a parameter is the range from the 0.025 quantile to the 0.975 quantile of the sampled posterior distribution. If the estimated coefficient intervals do not contain the true values more than 5% of the time, then there is clearly a problem with interpreting the coefficients, and the problem is more severe as the percent of intervals not containing the true values increases from 5%.

I first evaluate the assumption of acceptable coverage when there is no collinearity because this is the most favorable, although unlikely, situation. I next evaluate the assumption of acceptable coverage probabilities when there is collinearity in the model and specify systematic increases in collinearity to inspect its effect on both the

coverage probabilities and accuracy of the coefficients, as well as the strength of correlation in the estimated coefficients at each location and across the study area.

The model to generate the data for the simulation studies is

$$y(s) = \beta_1^*(s)x_1(s) + \beta_2^*(s)x_2(s) + \varepsilon(s), \qquad (6.1)$$

where the x_1 and x_2 are the first two principal components from a random sample drawn from a multivariate normal distribution of dimension ten with a mean vector of zeros and an identity covariance matrix, and the errors $\boldsymbol{\varepsilon}$ are sampled independently from a normal distribution with mean 0 and variance of τ^{2*}, which depends on the simulation study. The star notation denotes the true values of the parameters used to generate the data. Note that there is no true intercept in the model used to generate the data and I do not fit an intercept in the simulation study. The data points are equally spaced on a 10×10 grid, for a total of 100 observations. The goal of the simulation studies is to use the model in equation (6.1) to generate the data and see if the regression coefficient estimates match $\boldsymbol{\beta}^*$ for the spatially varying coefficient models.

For each simulation study, I generate a set number of realizations of the data process, where the number of realizations depends on the study. In the first simulation study, the explanatory variables and the error terms are simulated and the true regression coefficients are fixed for all realizations. The true regression coefficients used to simulate the data are based not on a model, but explicitly on the coordinates of the data points in the study area. In the rest of the simulation studies, the true regression coefficients are

simulated and thus vary from realization to realization. Another key difference between the first and other simulation studies is that there is no collinearity introduced into the explanatory variables in the first study and various levels of collinearity are used in the other simulation studies. An additional contrast is that there is only one level of spatial dependence in the regression coefficients in the first simulation study, and I consider three different levels of spatial dependence in other simulation studies. The second simulation study has a small true error variance and a small variance in the prior for the error variance. The third simulation study has a small true error variance and a larger variance in the prior for the error variance. The fourth simulation study uses a larger true error variance to generate the data. The fifth simulation study compares GWR, GWRR-global scaling, GWRR-local scaling, GWL-local, GWL-global, and the SVCP model in terms of response variable prediction and estimation error and regression coefficient estimation error. The sixth simulation study also compares these methods, but it uses a larger set of explanatory variables in the model.

Simulation Study 1

The first simulation study has one fixed pattern of true coefficients that are used to generate the data, where there is strong spatial dependence in the coefficients within each explanatory variable parameter. To achieve this, the true coefficient $\beta_1^*(s)$ is equal to the x coordinate at location s and the true coefficient $\beta_2^*(s)$ is equal to the y coordinate at location s. This results in a clear, increasing trend in each parameter across space, as shown in Figure 6.1. The regression coefficients are fixed in this way to compare the GWR and SVCP models when the true coefficients are not based on either model, thereby eliminating any predisposition towards one of the models. This gives a

baseline performance for later simulation studies. In this simulation study, the error variance τ^{2*} = .000001. This yields a very small error and effectively makes this a deterministic model. This is done to simplify the simulation study and remove any complicating effect the error term may have. In the SVCP model, simulation-based inference is carried out with 4000 iterations in the MCMC routine, discarding the first 2000 iterations as the “burn-in". Based on trace plots and Gelman’s $\hat{R}$ statistic (e.g. Gelman et al. 2003), the regression coefficients converged within 2000 iterations of the Gibbs sampler. The GWR model is fitted to the same data, using cross-validation to estimate the kernel bandwidth.

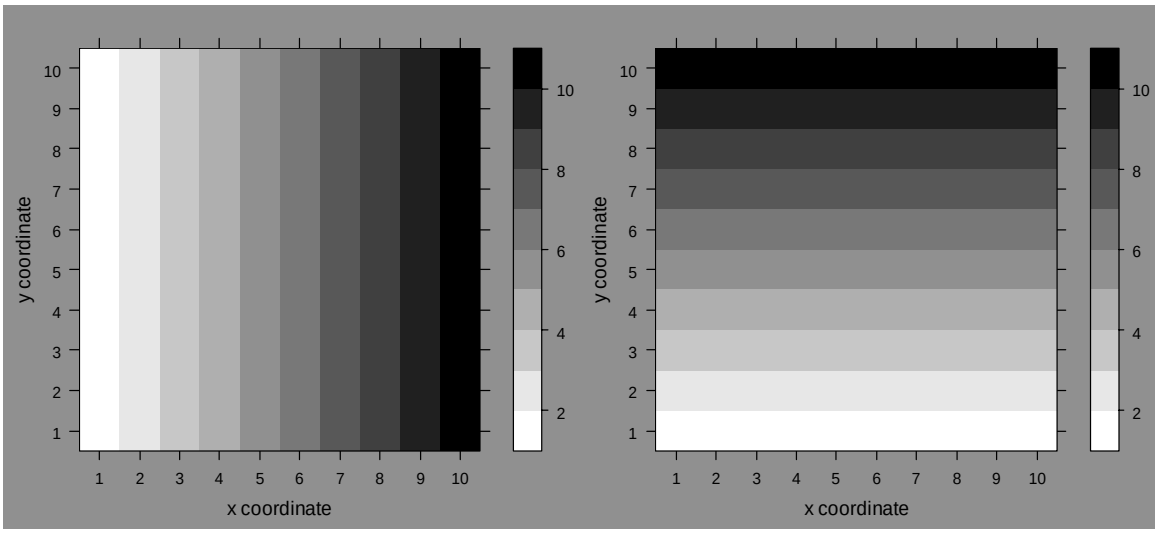

Figure 6.1. Coefficient pattern for each β^* parameter in simulation study 1. The left plot is for β_1^* and the right plot is for β_2^*. The parameter values range from 1 (lightest) to 10 (darkest).

The prior specification for this simulation study is as follows. I use a vague normal, $N(\mathbf{0}, 10^4\mathbf{I})$, for $\boldsymbol{\mu}_\beta$, a three-dimensional inverse Wishart, $IW(3, 10^{-6}\mathbf{I})$, for $\mathbf{T}$, and an inverse gamma, $IG(1, 10^{-6})$, for τ^2, where $\mathbf{I}$ is the identity matrix of dimension p. The inverse gamma prior has a small mean and variance. For the spatial dependence parameter ϕ, I use a gamma, $G(.021, .01)$, which has a mean of 2.12 and variance of 212. The hyperparameters for this gamma prior are chosen to have a large variance and a mean that solves the spatial correlation function set equal to .05, where the spatial range is set to half the maximum inter-point distance from the distance matrix $\mathbf{D}$. The spatial range is the distance beyond which the spatial association becomes negligible. Distributions that set prior mean spatial ranges to roughly half the maximum pairwise distance usually result in stable MCMC behavior (Banerjee and Johnson 2005). The calculation for the mean is found from solving $\exp(-(1/2 \cdot \max(\mathbf{D}))/\phi) = .05$ for ϕ.

In order to perform inference on the parameters in the SVCP model, starting values are required for the parameter estimates. The starting values for each SVCP model parameter estimate in each realization are assigned random values from certain distributions. The spatial dependence value is drawn from a uniform distribution over the range of pairwise distances in the study area. The coefficient means are drawn randomly from a multivariate normal distribution with zero means and the identity covariance matrix. To ensure that the starting value for the coefficient covariance matrix $\mathbf{T}$ is positive definite, a matrix is sampled from a multivariate normal distribution with a mean of the identity matrix and a diagonal matrix of 0.00001 for the covariance matrix. This gives a starting value for $\mathbf{T}$ that is close to the identity matrix.

To evaluate the coverage and accuracy of the regression coefficients, I calculate numerous summary statistics. For each realization of the process, I calculate the 95% credible intervals for each coefficient in the SVCP model and the 95% confidence intervals for each GWR coefficient. To calculate the 95% coverage probabilities, the true coefficients are compared to the 95% intervals obtained for the respective coefficient in each data realization and then the total number of realizations that contain the true values are summed and divided by the number of realizations. The means of the coverage probabilities are calculated for each explanatory variable from the corresponding coefficients to create a summary measure for all the realizations that is easy to present in a table. The accuracy of the regression coefficients is measured by calculating the root mean square error (RMSE) of all coefficients at each realization. The RMSE is the square root of the mean of the squared deviations of the estimates from the true values. The MSE is equal to the sum of the bias^2 and variance of an estimate, and it should be small for an accurate estimator. The average RMSE for all realizations is then calculated by averaging the RMSE's from all of the individual realizations.

The mean RMSE of the regression coefficients and the 95% coverage probabilities for each explanatory variable coefficient in simulation study 1 are listed in Table 6.1 for the Bayesian SVCP model and the GWR model. The results show that the Bayesian regression model has substantially lower RMSE for the coefficients than does the GWR model. In fact, the GWR RMSE is more than three times that from the SVCP model. The Bayesian regression model also has nearly 100% coverage (rounded to the nearest integer) of the true regression coefficients in the 95% confidence intervals, while the GWR model has only about 63% coverage of the true coefficients. Ideally, the

coverage probabilities would be 95% for all coefficients using the 95% credible intervals for the Bayesian model and the 95% confidence intervals for the GWR model. For the GWR model, this means that, on average, 37% of the coefficient confidence intervals do not contain the true coefficient values. This is clearly an unfavorable result for the GWR model, and if the result generalizes, casts significant doubt on inferences from this model. The results for this simulation study are from only one level of spatial dependence in the coefficients. To summarize the results from this simulation study, the Bayesian model outperforms the GWR model in terms of both regression coefficient coverage of the true values and accuracy when the true coefficients are not based on any model, there is strong spatial dependence in the regression coefficients, and there is no substantial collinearity in the explanatory variables.

Method	RMSE(β)	Mean CP (β1)	Mean CP (β2)
Bayesian	0.1097	100%	100%
GWR	0.4096	62%	63%

Table 6.1. Average root mean square error (RMSE) and average percent coverage of the 95% confidence intervals of the regression coefficients with GWR and average percent coverage of the 95% credible intervals with the SVCP model in simulation study 1.

Simulation Study 2

In the second simulation study, the true coefficients used to simulate the data are simulated from the fixed, true parameter values for the SVCP model. I fix the values for the parameters in the Bayesian model, other than the regression coefficients, and simulate

the true coefficients using the multivariate normal distribution and the fixed values for the coefficient means, covariances, and spatial dependence parameter. The coefficient covariance matrix at all locations is set to $\mathbf{T}^* = \begin{pmatrix} 0.1 & 0.0 \\ 0.0 & 0.5 \end{pmatrix}$ and the coefficient means are set to $\boldsymbol{\mu}_\beta^* = (1, 5)$. The error variance $\tau^{2*} = .000001$ again to simplify the simulation study. The spatial dependence parameter is first set to $\phi^* = 1$, then $\phi^* = 5$, and finally $\phi^* = 10$ to make three different sets of 200 realizations of the coefficient process. This is done to evaluate the performance of the GWR and SVCP model under three different levels of spatial dependence in the regression coefficients. The spatial dependence in the regression coefficients was much higher in simulation study 1 than with these three levels of ϕ^*. I use 2000 iterations in the MCMC routine for simulation-based inference of the SVCP model parameters, with a "burn-in" of 1000 iterations, and also fit the GWR model to the same data. Based on trace plots and Gelman's $\hat{R}$ statistic, the regression coefficients converged for individual realizations of the coefficient process within 1000 iterations of the Gibbs sampler.

The prior specification for this simulation study is as follows. I use a vague normal, $N(\mathbf{0}, 10^4\mathbf{I})$, for $\boldsymbol{\mu}_\beta$, a three-dimensional inverse Wishart, $IW(3, 10^{-6}\mathbf{I})$, for $\mathbf{T}$, and an inverse gamma, $IG(1, 10^{-6})$, for τ^2, where $\mathbf{I}$ is the identity matrix of dimension p. The inverse gamma prior has a small mean and variance. For the spatial dependence parameter ϕ, I use a gamma, $G(.021, .01)$, which has a mean of 2.12 and variance of 212. The same starting value distributions for the SVCP model parameter estimates are used as in simulation study 1.

Examples of the spatial pattern in individual realizations of the coefficient process are shown in Figure 6.2 for $\phi^* = 1$, Figure 6.3 for $\phi^* = 5$, and Figure 6.4 for $\phi^* = 10$. The figures show pattern in the coefficients for each variable, but the patterns are less clear and dramatic than those for the true coefficients in the first simulation study displayed in Figure 6.1, where there is more spatial dependence. In general, as ϕ^* increases there are more consistent and clear patterns in the true regression coefficients, and more global variation in the coefficients. Conversely, there is more local variation with smaller ϕ^*, resulting in less smooth patterns in the coefficients. For this simulation study, I start with no substantial collinearity in the model and systematically increase it until the explanatory variables are nearly perfectly collinear. This is done by replacing one of the original explanatory variables with one created from a weighted linear combination of the original explanatory variables, where the weight determines the amount of correlation of the variables. The formula for the new weighted variable is

$$x_2^c = c \cdot x_1 + (1-c) \cdot x_2, \tag{6.2}$$

where x_2^c replaces x_2 in the model in equation (6.1) and c is the weight between 0 and 1.

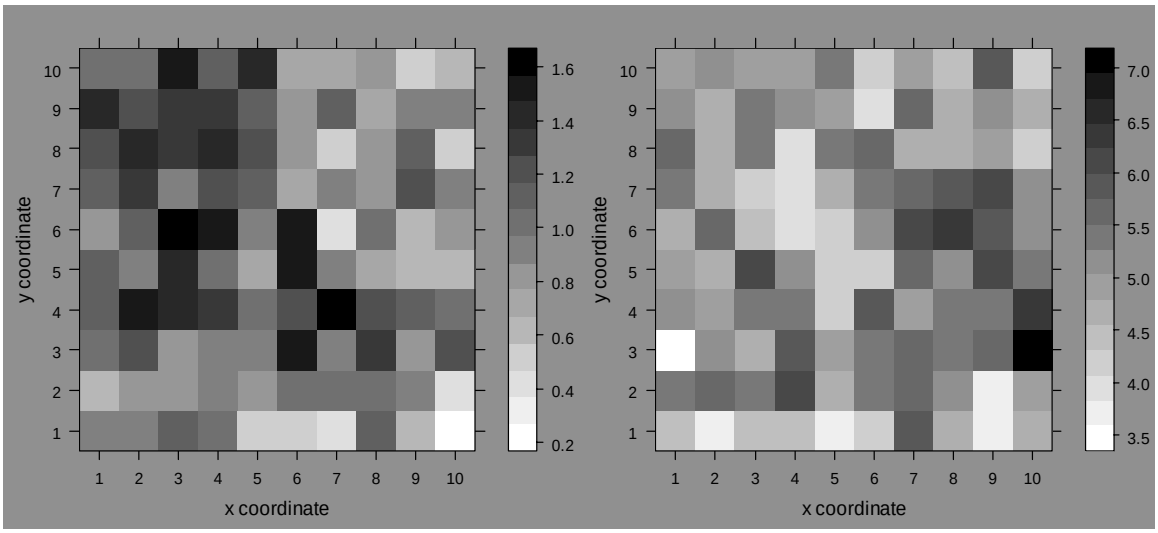

Figure 6.2. Coefficient pattern for each β^* parameter for one coefficient realization in simulation study 2 when $\phi^* = 1$. The left plot is for β_1^* and the right plot is for β_2^*.

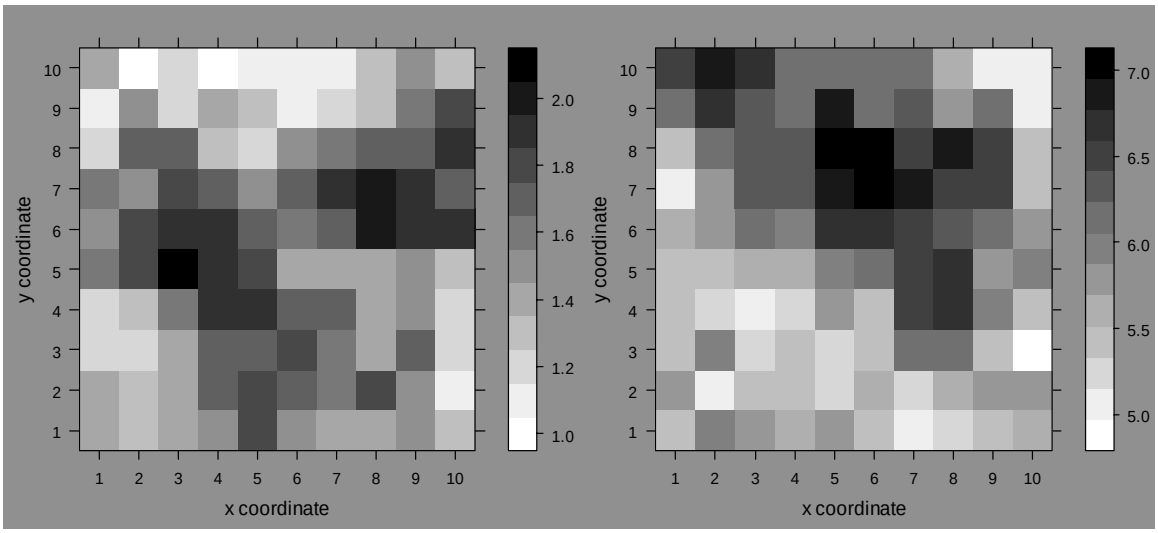

Figure 6.3. Coefficient pattern for each β^* parameter for one coefficient realization in simulation study 2 when $\phi^* = 5$. The left plot is for β_1^* and the right plot is for β_2^*.

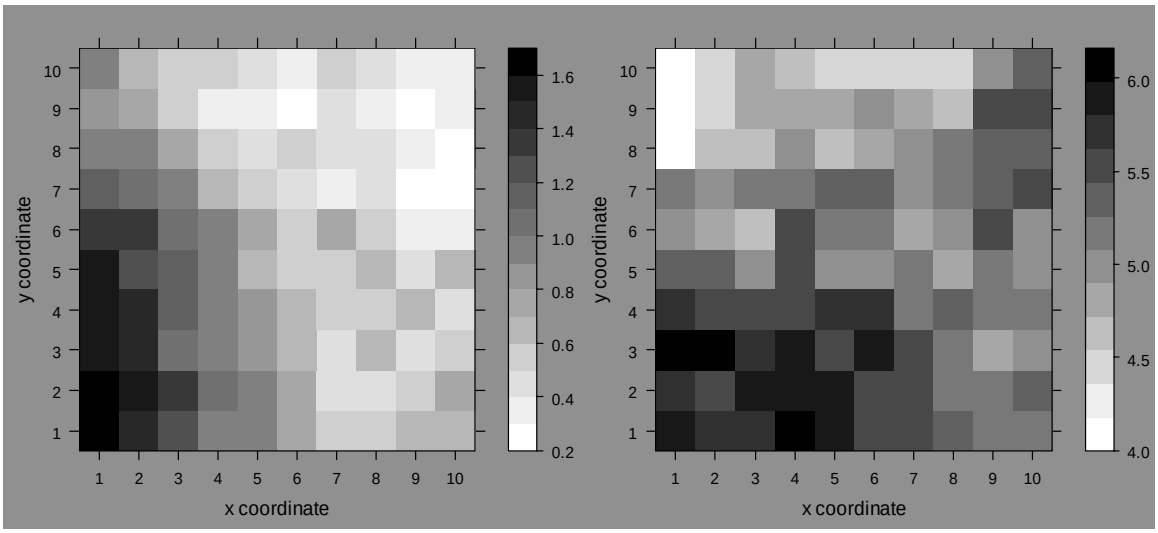

Figure 6.4. Coefficient pattern for each β^* parameter for one coefficient realization in simulation study 2 when $\phi^* = 10$. The left plot is for β_1^* and the right plot is for β_2^*.

To evaluate the coverage and accuracy of the regression coefficients from the GWR and SVCP model, I use the same coverage probability and RMSE calculations described for simulation study 1. In addition, I calculate the correlation in the estimated regression coefficients from both models. There are two types of correlation of interest in the estimate regression coefficients. One is the overall correlation coefficient (C_{12}) between the sets of estimated coefficients for two explanatory variables, and the other is the estimated local correlation coefficient (C_{12}^s) between two coefficients at any location s. The local coefficient correlation at each location is given by equation (3.23) in the SVCP model. The local coefficient correlation in the GWR model is calculated using the previously defined equation (4.3). The overall correlation between sets of coefficients is calculated for the SVCP and GWR models using previously defined equation (4.1).

The results of this simulation study are listed in Table 6.2 for the Bayesian SVCP model and Table 6.3 for the GWR model. The tables show the mean statistics calculated from the 200 realizations of the coefficient process. The coverage probabilities generally increase in both the SVCP and GWR models as ϕ^* increases. The tables also show that the SVCP model has substantially larger mean coverage probabilities than does GWR when there is no collinearity in the data ($c = 0$), regardless of the level of ϕ^*. As the collinearity in the explanatory variables increases, the coverage probabilities gradually decrease in the SVCP model and increase in the GWR model. The coverage probabilities for the SVCP model do not substantially decrease until there is moderate to strong collinearity in the variables. The coverage probabilities increase with increasing collinearity in the GWR model because the variances of the coefficients increase as well, as shown in the columns of coefficient variances in Table 6.3. It is a well-known result in statistics that regression coefficient variances increase in the presence of collinearity (Neter et al. 1996). The coverage probability increases with decreasing estimator bias and increases with increasing estimator variance. Therefore, looking only at the coverage probabilities for the GWR model is misleading because the increase in coverage probabilities is due directly to increasingly imprecise estimates.

$\phi^* = 1$

Weight	X Corr	Mean CP (β1)	Mean CP (β2)	C_{12}	C_{12}^s	RMSE(β)
0.0	0.000	0.85	0.91	0.018	-0.010	0.325
0.1	0.127	0.87	0.90	0.130	0.072	0.332
0.3	0.442	0.83	0.86	0.268	0.131	0.340
0.5	0.755	0.80	0.85	0.341	0.133	0.351
0.7	0.937	0.48	0.54	0.845	0.732	0.384
0.9	0.995	0.47	0.46	0.780	0.780	0.429

$\phi^* = 5$

Weight	X Corr	Mean CP (β1)	Mean CP (β2)	C_{12}	C_{12}^s	RMSE(β)
0.0	0.000	0.93	0.93	-0.036	-0.029	0.162
0.1	0.127	0.93	0.94	0.060	0.008	0.164
0.3	0.442	0.93	0.93	0.127	0.053	0.169
0.5	0.755	0.93	0.93	0.188	0.055	0.172
0.7	0.937	0.79	0.82	0.383	0.239	0.196
0.9	0.995	0.40	0.41	0.627	0.622	0.273

$\phi^* = 10$

Weight	X Corr	Mean CP (β1)	Mean CP (β2)	C_{12}	C_{12}^s	RMSE(β)
0.0	0.000	0.94	0.94	-0.005	-0.003	0.115
0.1	0.127	0.94	0.94	0.033	0.016	0.117
0.3	0.442	0.93	0.94	0.070	0.028	0.118
0.5	0.755	0.93	0.94	0.120	-0.018	0.123
0.7	0.937	0.81	0.83	0.308	0.195	0.143
0.9	0.995	0.42	0.44	0.549	0.534	0.206

Table 6.2. Results of simulation study 2 for the Bayesian SVCP model. The columns listed in order are the correlation weight, explanatory variable correlation, average coverage probabilities for each explanatory variable coefficient, the mean overall correlation between the two sets of variable coefficients, the mean local coefficient correlation at each location, and the average RMSE of the coefficients.

$\phi^* = 1$

Weight	X Corr	Mean CP (β1)	Mean CP (β2)	C_{12}	C_{12}^s	Var(β1)	Var(β2)	RMSE(β)
0.0	0.000	0.51	0.35	-0.075	-0.020	0.013	0.017	0.459
0.1	0.127	0.51	0.39	-0.111	-0.142	0.013	0.021	0.457
0.3	0.442	0.51	0.42	-0.242	-0.447	0.013	0.027	0.463
0.5	0.755	0.58	0.49	-0.404	-0.754	0.021	0.047	0.492
0.7	0.937	0.75	0.63	-0.747	-0.937	0.078	0.138	0.597
0.9	0.995	0.85	0.83	-0.976	-0.995	1.332	1.627	1.529

$\phi^* = 5$

Weight	X Corr	Mean CP (β1)	Mean CP (β2)	C_{12}	C_{12}^s	Var(β1)	Var(β2)	RMSE(β)
0.0	0.000	0.65	0.54	-0.046	-0.019	0.007	0.010	0.226
0.1	0.127	0.64	0.55	-0.022	-0.135	0.006	0.010	0.226
0.3	0.442	0.64	0.59	-0.088	-0.425	0.007	0.014	0.235
0.5	0.755	0.70	0.67	-0.229	-0.737	0.012	0.027	0.266
0.7	0.937	0.77	0.74	-0.549	-0.932	0.046	0.082	0.407
0.9	0.995	0.82	0.82	-0.941	-0.995	0.819	1.002	1.345

$\phi^* = 10$

Weight	X Corr	Mean CP (β1)	Mean CP (β2)	C_{12}	C_{12}^s	Var(β1)	Var(β2)	RMSE(β)
0.0	0.000	0.66	0.56	0.041	-0.019	0.004	0.005	0.164
0.1	0.127	0.64	0.57	-0.039	-0.135	0.004	0.006	0.162
0.3	0.442	0.65	0.61	-0.113	-0.422	0.004	0.008	0.165
0.5	0.755	0.69	0.65	-0.186	-0.736	0.006	0.013	0.194
0.7	0.937	0.78	0.75	-0.516	-0.932	0.024	0.043	0.289
0.9	0.995	0.81	0.81	-0.926	-0.995	0.420	0.514	0.977

Table 6.3. Results of simulation study 2 for the GWR model. The columns correspond to the correlation weight, explanatory variable correlation, average coverage probabilities for each explanatory variable coefficient, the mean overall correlation between the two sets of variable coefficients, the mean local coefficient correlation at each location, the variance of each explanatory variable coefficient, and the average RMSE of the coefficients.

The RMSE is a better measure of parameter estimate accuracy in the presence of collinearity because it increases with increasing bias and increasing estimator variance. Figure 6.5 shows the average RMSE values for the SVCP and GWR models at varying levels of collinearity for the three levels of ϕ^*. The average RMSE values at each level of collinearity show that the coefficients from the SVCP model are closer to their true values than are those from GWR. Not surprisingly, the average RMSE increases with increasing collinearity for the SVCP model, showing that the estimated coefficients move away from their true values. Intuitively, the average RMSE values for both the GWR and SVCP models decrease as the spatial dependence increases because the models are more appropriate for data with more spatial dependence. The results show that the GWR model coefficients are more biased than the SVCP model coefficients because the average RMSE for the SVCP model is consistently less than for GWR and the coverage probabilities are larger for the SVCP model than for GWR when there is no collinearity.

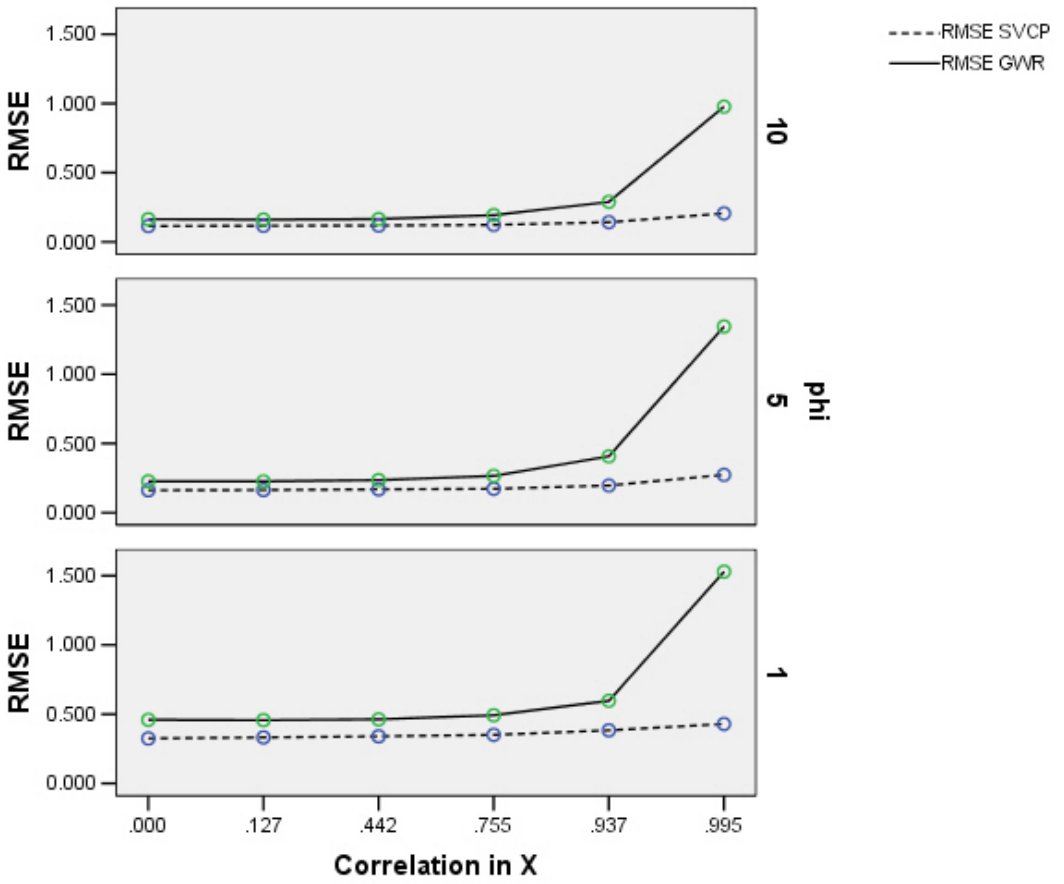

Figure 6.5. Average RMSE for regression coefficients from the SVCP (dashed) and the GWR (solid) models at specific levels of collinearity in the explanatory variables and at different levels of ϕ^* in simulation study 2.

The average overall coefficient correlation and average local coefficient correlation calculated from all realizations for both types of correlation in the SVCP and GWR models are listed in Table 6.2 and Table 6.3. The results show that the Bayesian model better controls the correlation in the regression coefficients than does the GWR model. The GWR coefficients become almost perfectly negatively correlated across space with very strong explanatory variable collinearity, while the SVCP model coefficients demonstrate only moderate levels of positive correlation. Similarly, the average local coefficient correlation is strongly negative with strong collinearity in the GWR model, while the correlation is weakly to moderately positive with the SVCP model. Furthermore, it appears the SVCP model is self-tuning with regard to the level of collinearity in the model and its subsequent effect on the estimated coefficients. The variances of the coefficients do not increase dramatically with increasing collinearity in the SVCP model as they do in the GWR model, as is evident from the decreasing coverage probabilities in the SVCP model.

Another contrast between the GWR and SVCP models became apparent from the simulation study. As ϕ^* increases in the simulations, the estimated GWR kernel bandwidth decreases. With the stronger spatial dependence, and hence, trend, in the coefficients in this simulation study, the coefficients at more distant locations tend to be more different, and GWR places less emphasis on distant data points through the spatial weights when estimating the model at one location. With smaller spatial dependence, the coefficients tend to be less dissimilar at distant locations, and GWR places more emphasis on distant locations through the local spatial weights.

Simulation Study 3

Simulation study 2 had a very small true error term in the data-generating model due to setting $\tau^{2*} = .000001$, and also had a prior for the error variance with a small mean and variance. The reason for using a small true variance was to evaluate the SVCP model performance without the complication of the error term, and the prior with small mean and variance was meant to help achieve that objective. However, given the earlier details of Chapter 5 that the SVCP model parameters may be viewed as ridge regression solutions, restricting τ^2 through its prior could affect the amount of shrinkage in the estimated regression coefficients. For that reason, a simulation study is presented here that is similar to simulation study 2, but uses a prior for τ^2 that has a larger variance and is, hence, less informative about τ^2. As in simulation study 2, the coefficient covariance matrix at all locations is set to $\mathbf{T}^* = \begin{pmatrix} 0.1 & 0.0 \\ 0.0 & 0.5 \end{pmatrix}$, the coefficient means are set to $\boldsymbol{\mu}_\beta^* = (1, 5)$, and the error variance $\tau^{2*} = .000001$. The spatial dependence parameter is set to only $\phi^* = 10$ in the interest of time. As with simulation study 2, there are 200 realizations of the coefficient process used in this simulation study.

The prior specification for this simulation study is as follows. I use a vague normal, $N(\mathbf{0}, 10^4\mathbf{I})$, for $\boldsymbol{\mu}_\beta$, a three-dimensional inverse Wishart, $IW(3, 10^{-6}\mathbf{I})$, for $\mathbf{T}$, and an inverse gamma, $IG(1, .01)$, for τ^2, where $\mathbf{I}$ is the identity matrix of dimension p. For the spatial dependence parameter ϕ, I use a gamma, $G(.021, .01)$, which has a mean of 2.12 and variance of 212. The same starting value distributions for the SVCP model parameter estimates are used as in simulation study 1.

The results of the simulation study are listed in Table 6.4 for the Bayesian SVCP model and in Table 6.5 for the GWR model. The results strongly favor the SVCP model overall, and show that the SVCP model does much better at covering the true coefficient values used to generate the data than does GWR. For example, the SVCP model average coverage probabilities are at least 0.25 higher than the ones from GWR when there is no collinearity. In addition, the average coverage probabilities are always higher for the SVCP model when there is collinearity. The coverage probabilities for the SVCP model start near the goal of .95 and decrease only slightly when there is strong collinearity. The GWR coverage probabilities increase with increasing collinearity as a result of increased estimator variance but are still much less than those from the SVCP model when there is strong collinearity. In terms of accuracy, the average RMSE is considerably lower for the SVCP model compared to the GWR model for all levels of explanatory variable correlation. The SVCP model average RMSE jumps much less compared to GWR with strong collinearity. The SVCP model also does a better job than GWR at controlling the overall and local regression coefficient correlation. The GWR estimated coefficients become strongly negatively correlated locally and overall with strong collinearity, while the SVCP model coefficients exhibit weak positive correlation locally and moderate positive correlation with strong collinearity. The dramatically better performance in estimating the coefficients when there is strong collinearity with the SVCP model over GWR helps demonstrate the Bayesian SVCP model regression coefficients as ridge regression solutions. The average RMSE's for the response variable also show that the SVCP model performs better than GWR in estimating the response.

Comparing the SVCP model results for simulation study 2 and simulation study 3 indicates that the prior for τ^2 with the small variance had an impact in the estimation of the parameters in simulation study 2. The coverage probabilities in simulation study 2 are reduced much more then in simulation study 3 when there is strong collinearity. The amount of overall correlation in the estimated coefficients is also lower in simulation study 2 compared to simulation study 3. These results suggest that a smaller posterior mean for τ^2 produces more regression coefficient shrinkage towards the coefficient means.

ϕ^* = 10

Weight	X Corr	Mean CP (β1)	Mean CP (β2)	C_{12}	C_{12}^s	RMSE(β)	RMSE(y)
0.0	0.000	0.94	0.96	0.018	-0.013	0.168	0.017
0.1	0.127	0.94	0.96	0.039	-0.007	0.170	0.017
0.3	0.442	0.94	0.95	0.098	0.015	0.176	0.017
0.5	0.755	0.93	0.94	0.206	0.069	0.188	0.018
0.7	0.937	0.90	0.92	0.399	0.165	0.211	0.018
0.9	0.995	0.92	0.91	0.744	0.271	0.282	0.016

Table 6.4. Results of simulation study 3 for the Bayesian SVCP model. The columns listed in order are the correlation weight, explanatory variable correlation, average coverage probabilities for each explanatory variable coefficient, the mean overall correlation between the variable coefficients, the mean local coefficient correlation at each location, the average RMSE of the coefficients, and RMSE of the response.

ϕ^* = 10

Weight	X Corr	Mean CP (β1)	Mean CP (β2)	C_{12}	C_{12}^s	RMSE(β)	RMSE(y)
0.0	0.000	0.67	0.54	-0.003	0.054	0.235	0.069
0.1	0.127	0.66	0.54	-0.033	-0.043	0.234	0.070
0.3	0.442	0.64	0.56	-0.116	-0.324	0.244	0.073
0.5	0.755	0.66	0.62	-0.262	-0.671	0.286	0.078
0.7	0.937	0.72	0.72	-0.546	-0.910	0.430	0.086
0.9	0.995	0.78	0.79	-0.929	-0.994	1.371	0.095

Table 6.5. Results of simulation study 3 for the GWR model. The columns correspond to the correlation weight, explanatory variable correlation, average coverage probabilities for each explanatory variable coefficient, the mean overall correlation between the variable coefficients, the mean local coefficient correlation at each location, the average RMSE of the coefficients, and the RMSE of the response.

Simulation Study 4

Simulation studies 1, 2, and 3 had a very small true error term in the data-generating model, due to setting $\tau^{2*} = .000001$. While the reason for doing so was to evaluate the SVCP model performance without the complication of the error term, it is likely that real data will have larger errors present. For that reason, a simulation study is presented here that is similar to simulation study 2, but has a $\tau^{2*} = 1$ to represent a more likely signal to noise ratio. The coefficient covariance matrix at all locations is set to $\mathbf{T}^* = \begin{pmatrix} 0.5 & 0.0 \\ 0.0 & 0.5 \end{pmatrix}$ and the coefficient means are set to $\boldsymbol{\mu}_\beta^* = (1, 5)$. In the interest of time, this simulation study was only run for $\phi^* = 10$. As in simulation studies 2 and 3, there are 200 realizations of the coefficient process used in this simulation study.

The prior specification for this simulation study is as follows. I use a vague normal, $N(\mathbf{0}, 10^4\mathbf{I})$, for $\boldsymbol{\mu}_\beta$, a three-dimensional inverse Wishart, $IW(3, .1\cdot\mathbf{I})$, for $\mathbf{T}$, and an inverse gamma, $IG(1, .01)$, for τ^2, where $\mathbf{I}$ is the identity matrix of dimension p. For the spatial dependence parameter ϕ, I use a gamma, $G(.021, .01)$, which has a mean of 2.12 and variance of 212.

The results of the simulation study are listed in Table 6.6 for the Bayesian SVCP model and in Table 6.7 for the GWR model. The results show that the SVCP model does much better at covering the true coefficient values used to generate the data than does GWR. For example, the SVCP model average coverage probabilities are about 0.20 higher than the ones from GWR when there is no collinearity. The average coverage probabilities are approximately the same for the two models only when there is very strong collinearity, and the increase in the average coverage probabilities for GWR occurs as a direct result of increased estimate variance, which results in a degradation of estimate precision. Figure 6.6 and Figure 6.7 show the coverage probabilities for the regression coefficients in the SVCP and GWR models, respectively, when there is no correlation in the explanatory variables. The plots show that there is more variation in the coverage probabilities with the GWR model than with the SVCP model. GWR does slightly better at estimating the true regression coefficients than does SVCP for low levels of collinearity, however, the SVCP is much better at estimating the coefficients when there is strong collinearity. The dramatic improvement in estimating the coefficients when there is strong collinearity with the SVCP model helps demonstrate the Bayesian SVCP model regression coefficients as ridge regression solutions. In addition, the SVCP model also does a better job than GWR at controlling the overall and local

regression coefficient correlation. One area that GWR consistently does better in than the SVCP model is with estimating the response, although the differences are not overwhelming. This is not completely surprising, given the roots of GWR in local linear regression, which is designed for response prediction and not regression coefficient inference. Wheeler and Tiefelsdorf (2005) found that GWR performed better at estimating the response when there is more spatial dependence in the errors of a simulation study. The errors used in the present simulation study are not controlled to have a specific spatial dependence, and this suggests GWR appears to perform better at estimating the response when there are non-zero errors.

ϕ^* = 10

Weight	X Corr	Mean CP (β1)	Mean CP (β2)	C_{12}	C_{12}^s	RMSE(β)	RMSE(y)
0.0	0.000	0.90	0.91	0.138	0.064	0.508	0.894
0.1	0.127	0.90	0.89	0.199	0.093	0.514	0.907
0.3	0.442	0.89	0.89	0.377	0.147	0.536	0.907
0.5	0.755	0.90	0.89	0.597	0.212	0.568	0.899
0.7	0.937	0.91	0.91	0.777	0.228	0.651	0.904
0.9	0.995	0.93	0.93	0.835	0.224	1.235	0.892

Table 6.6. Results of simulation study 4 for the Bayesian SVCP model. The columns listed in order are the correlation weight, explanatory variable correlation, average coverage probabilities for each explanatory variable coefficient, the mean overall correlation between the variable coefficients, the mean local coefficient correlation at each location, the average RMSE of the coefficients, and RMSE of the response.

$\phi^* = 10$

Weight	X Corr	Mean CP (β1)	Mean CP (β2)	C_{12}	C_{12}^s	RMSE(β)	RMSE(y)
0.0	0.000	0.70	0.70	0.058	0.019	0.487	0.884
0.1	0.127	0.69	0.72	0.011	-0.094	0.497	0.889
0.3	0.442	0.69	0.76	-0.126	-0.388	0.536	0.892
0.5	0.755	0.76	0.83	-0.365	-0.711	0.633	0.890
0.7	0.937	0.86	0.89	-0.719	-0.922	0.947	0.883
0.9	0.995	0.92	0.92	-0.973	-0.994	2.913	0.873

Table 6.7. Results of simulation study 4 for the GWR model. The columns correspond to the correlation weight, explanatory variable correlation, average coverage probabilities for each explanatory variable coefficient, the mean overall correlation between the variable coefficients, the mean local coefficient correlation at each location, the average RMSE of the coefficients, and the RMSE of the response.

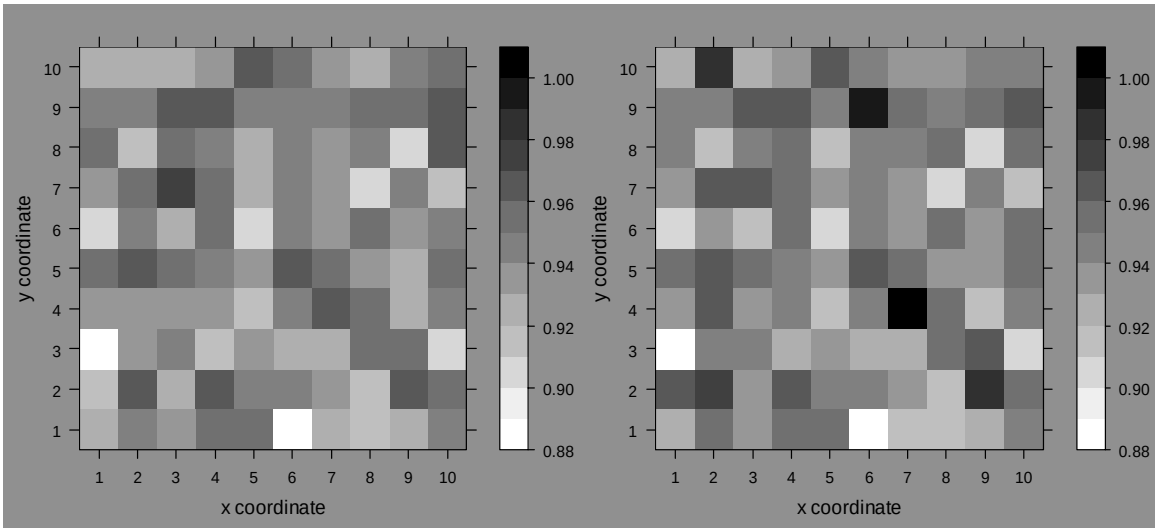

Figure 6.6. Coverage probabilities for each β^* parameter when $\phi^* = 10$ in simulation study 4 for the SVCP model. The left plot is for β_1^* and the right plot is for β_2^*.

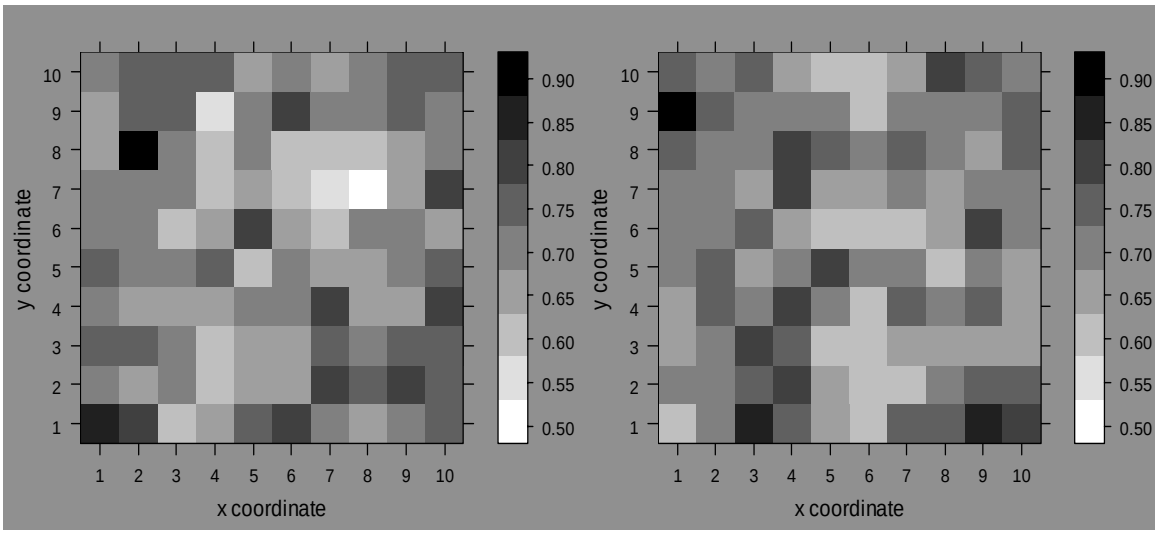

Figure 6.7. Coverage probabilities for each β^* parameter when $\phi^* = 10$ in simulation study 4 for the GWR model. The left plot is for β_1^* and the right plot is for β_2^*.

Simulation Study 5

The previous simulation studies illuminated the differences in performance between the GWR and Bayesian SVCP models, in terms of marginal regression coefficient inference. In a sense, the comparison of the two methods is not entirely an even one, as the Bayesian regression coefficients can be viewed as ridge regression solutions, as discussed earlier in the dissertation. To make the comparison more even, and to judge the benefit of adding the regularization methods of ridge regression and the lasso to the GWR framework, a simulation study is presented now that measures the prediction and estimation error of the response and the estimation error of the regression coefficients for GWR, GWRR, GWL, and the SVCP model. This simulation study uses the same model to generate the data as simulation study 4. In other words, the coefficient means are set to $\boldsymbol{\mu}_{\beta}^{*} = (1, 5)$, $\mathbf{T}^{*}$ = diag(.5, .5), and $\tau^{2*} = 1$, where diag() creates a diagonal matrix with the input numbers on the diagonal. This simulation study is run for $\phi^{*} = 10$ only in the interest of time.

The prior specification for this simulation study is as follows. I use a vague normal, $N(\mathbf{0}, 10^{4}\mathbf{I})$, for $\boldsymbol{\mu}_{\beta}$, a three-dimensional inverse Wishart, $IW(3, .1 \cdot \mathbf{I})$, for $\mathbf{T}$, and an inverse gamma, $IG(1, .01)$, for τ^{2}, where $\mathbf{I}$ is the identity matrix of dimension p. For the spatial dependence parameter ϕ, I use a gamma, $G(.021, .01)$, which has a mean of 2.12 and variance of 212.

In the interest of time, the simulation study is carried out with only four levels of explanatory variable collinearity. The weights used in equation (6.2) to create the collinearity are $c = (0.0, 0.5, 0.7, 0.9)$, which coincide with explanatory variable

correlation of $r = (0.0,\ 0.74,\ 0.93,\ 0.99)$. These levels of correlation correspond to no collinearity as a baseline, and then moderate, strong, and very strong collinearity. In this study, 100 realizations of the coefficient process are generated, and the model parameters and responses are estimated for each realization for each of the following models: GWR, GWRR–global scaling, GWRR–local scaling, GWL–global, GWL–local, and SVCP. In the interest of time, the GWRR solutions use estimation scheme 1, so the errors associated with the solutions in this study are upper bounds on the possible GWRR solution errors, where better solutions may be possible using estimation scheme 3. For each data realization, the RMSE is calculated for the responses $\mathbf{y}$ and the coefficients $\boldsymbol{\beta}$ for each model. In addition, the RMSPE is calculated for the responses $\mathbf{y}$ for GWR and the regularized GWR models. To provide summary measures for the simulation study, the RMSE's and RMSPE's are averaged over the 100 realizations of the coefficients.

The average RMSPE for $\mathbf{y}$ for each model is listed in Table 6.8. The lowest RMSPE for each level of variable correlation (column) is in bold font. The results in the table show that the GWL-local model produces the lowest prediction error of the response and is clearly the best in this category. This is an expected result, as the GWL-local model adds the most local penalization parameters to the GWR model, and a general result in penalization methods is that adding more shrinkage parameters lowers the prediction error by stabilizing the model. The next best performer in terms of RMSPE of the response is a tie between GWL-global and GWRR-local scaling, where GWL-global does better with no or moderate collinearity and GWRR-local scaling does better with strong or very strong collinearity. It is no coincidence that GWR has the highest prediction error at each level of collinearity. These results show that adding penalization terms to GWR for the regression coefficients results in lower prediction error of the response than with GWR.

	c = 0.0	c = 0.5	c = 0.7	c = 0.9
Method	RMSPE(y)	RMSPE(y)	RMSPE(y)	RMSPE(y)
GWR	1.156	1.127	1.132	1.146
GWRR - global	1.155	1.126	1.130	1.134
GWRR - local	1.155	1.125	1.127	1.130
GWL - global	1.545	1.125	1.128	1.132
GWL - local	**0.997**	**0.981**	**1.000**	**1.024**
SVCP	.	.	.	.

Table 6.8. RMSPE of the response for each model used in simulation study 5 at four levels of explanatory variable correlation.

The average RMSE for $\mathbf{y}$ for each model is listed in Table 6.9. The lowest RMSE for each level of variable correlation (column) is in bold font. The results in the table show that the GWL model produces the lowest estimation error of the response, with the GWL-local model doing better for no or moderate collinearity and the GWL-global model doing better for strong or very strong collinearity. It is noteworthy that the Bayesian SVCP model is overall the worst performer for estimating the response.

	c = 0.0	c = 0.5	c = 0.7	c = 0.9
Method	RMSE(y)	RMSE(y)	RMSE(y)	RMSE(y)
GWR	0.870	0.875	0.875	0.872
GWRR - global	0.892	0.897	0.897	0.897
GWRR - local	0.871	0.879	0.884	0.908
GWL - global	0.868	0.869	**0.863**	**0.831**
GWL - local	**0.785**	**0.813**	0.886	1.064
SVCP	0.906	0.910	0.905	0.892

Table 6.9. RMSE of the response for each model used in simulation study 5 at four levels of explanatory variable correlation.

The average RMSE for $\boldsymbol{\beta}$ for each model is listed in Table 6.10. The lowest RMSE for each level of variable correlation (column) is in bold font. The results in the table show that the Bayesian SVCP model produces the lowest estimation error of the regression coefficients overall. The GWRR models perform the next best, with the GWRR-local scaling performing better than the global scaling version when there is strong or very strong collinearity. GWRR-global scaling produces the lowest average coefficient RMSE when there is no collinearity in the model, and the SVCP model produces the lowest average RMSE at all levels of collinearity. In fact, the SVCP model

is clearly the best when there is collinearity in the model, as the RMSE's for this model are dramatically lower than with the other models when there is strong or very strong collinearity. An explanation for this is that not only can the Bayesian SVCP model coefficients be viewed as local ridge regression solutions, but the SVCP model also best captures the spatial structure of the regression coefficients through the prior covariance function. I refer to the SVCP coefficients as local ridge regression solutions because there is implicitly a matrix of shrinkage parameters in this model, in contrast to the one ridge parameter used in the implementation of GWRR introduced in this dissertation. These results suggest that one should use the Bayesian SVCP model when one is concerned with marginal inference on the regression coefficients in the presence of collinearity in the model.

	c = 0.0	c = 0.5	c = 0.7	c = 0.9
Method	RMSE(β)	RMSE(β)	RMSE(β)	RMSE(β)
GWR	0.487	0.653	0.994	3.047
GWRR - global	**0.485**	0.647	0.981	2.934
GWRR - local	0.489	0.656	0.974	2.307
GWL - global	0.490	0.656	0.989	2.817
GWL - local	0.929	1.083	1.340	3.065
SVCP	0.515	**0.582**	**0.665**	**1.325**

Table 6.10. RMSE of the regression coefficients for each model used in simulation study 5 at four levels of explanatory variable correlation.

Many times in traditional regression analyses, researchers only consider using penalization methods, such as ridge regression, when there are many more explanatory variables than two to include in the model. However, the results from this simulation

study show that one can improve on the GWR in terms of prediction and estimation of the response and estimation of the regression coefficients for even small models of two variables. The implication of this is that there appears to be no reason to continue to use GWR without penalization, as alternatives such as GWL and GWRR perform better in the same framework. I anticipate that the benefits of the penalization in GWR will only increase with an increasing number of potentially correlated explanatory variables.

Simulation Study 6

While simulation study 5 was a revealing comparison of the performance of the Bayesian SVCP model, GWR, and regularized versions of GWR, it used only two explanatory variables. Most regression problems with real data will involve more than two explanatory variables. For this reason, a simulation study similar to simulation study 5 is presented here that has four explanatory variables, with one of the true coefficients used to generate the data set equal to nearly zero. The model for the data generation is now

$$y(s) = \beta_1^*(s)x_1(s) + \beta_2^*(s)x_2(s) + \beta_3^*(s)x_3(s) + \beta_4^*(s)x_4(s) + \varepsilon(s) \qquad (6.3)$$

The true values used to generate the data are $\boldsymbol{\mu}_\beta^* = (1, 5, 5, 0)$, $\tau^{2*} = 1$, and $\mathbf{T}^* = \text{diag}(.1, .5, .5, .0000001)$, where diag() creates a diagonal matrix with the input numbers. The mean of 0 and the small variance for the fourth type of regression coefficient produce a variable effect that is effectively zero across the study area. This simulation study is run for $\phi^* = 10$ only, and uses 100 realizations of the coefficient process. In the interest of

time, the GWRR solutions use estimation scheme 1, so the errors associated with the solutions in this study are upper bounds on the possible GWRR solutions errors, where better solutions may be possible using estimation scheme 3.

The prior specification for this simulation study is as follows. I use a vague normal, $N(\mathbf{0}, 10^4\mathbf{I})$, for $\boldsymbol{\mu}_\beta$, a five-dimensional inverse Wishart, $IW(5, .1\cdot\mathbf{I})$, for $\mathbf{T}$, and an inverse gamma, $IG(1, .01)$, for τ^2, where $\mathbf{I}$ is the identity matrix of dimension p. For the spatial dependence parameter ϕ, I use a gamma, $G(.021, .01)$.

The average RMSPE for $\mathbf{y}$ for each model is listed in Table 6.11. The lowest RMSPE for each level of variable correlation (column) is in bold font. The results in the table show that the GWL-local model produces the lowest prediction error of the response and is clearly the best in this category. This finding is consistent with the simulation study 5 results. The next best performer in terms of RMSPE of the response is the GWL-global model. The better performance of the two versions of GWL relative to the other versions of GWR is not unexpected, given that the GWL methods can shrink the regression coefficients to zero to match the true values for one of the variables. As in simulation study 5, GWR has the highest prediction error at each level of collinearity. These results again show that adding penalization terms for the regression coefficients in GWR results in lower prediction error of the response than with GWR.

	C = 0.0	c = 0.5	c = 0.7	c = 0.9
Method	RMSPE(y)	RMSPE(y)	RMSPE(y)	RMSPE(y)
GWR	1.186	1.148	1.153	1.171
GWRR - global	1.185	1.147	1.151	1.160
GWRR - local	1.185	1.147	1.151	1.159
GWL - global	1.180	1.142	1.146	1.156
GWL - local	**0.935**	**0.928**	**0.938**	**0.934**
SVCP	.	.	.	.

Table 6.11. RMSPE of the response for each model used in simulation study 6 at four levels of explanatory variable correlation.

The average RMSE for $\mathbf{y}$ for each model is listed in Table 6.12. The lowest RMSE for each level of variable correlation (column) is in bold font. The results in the table show that the GWL-local model produces the lowest estimation error of the response at all levels of collinearity. This is in contrast to simulation study 5, where the GWL-global model performed better than the GWL-local model for strong collinearity. Overall, the two simulation studies show that the GWL models perform better than the other models in explaining the response variable. Taken together, the results from Table 6.11 and Table 6.12 indicate that the GWL-local model is the best for predicting and estimating the response variable in this simulation study. It is also noteworthy that the Bayesian SVCP model performs better than GWR in estimating the response at all levels of collinearity. This is a different result from simulation study 5, where GWR performed better than the Bayesian model.

	c = 0.0	c = 0.5	c = 0.7	c = 0.9
Method	RMSE(y)	RMSE(y)	RMSE(y)	RMSE(y)
GWR	0.863	0.871	0.863	0.858
GWRR - global	0.875	0.882	0.877	0.877
GWRR - local	0.865	0.873	0.867	0.875
GWL - global	0.858	0.860	0.853	0.821
GWL - local	**0.689**	**0.711**	**0.719**	**0.756**
SVCP	0.853	0.852	0.852	0.855

Table 6.12. RMSE of the response for each model used in simulation study 6 at four levels of explanatory variable correlation.

The average RMSE for $\boldsymbol{\beta}$ for each model is listed in Table 6.13. The lowest RMSE for each level of variable correlation (column) is in bold font. The results in the table show that the Bayesian SVCP model produces the lowest estimation error of the regression coefficients overall. The SVCP model has the lowest RMSE when there is collinearity in the model, and the error for this model is noticeably lower than with the other models when there is strong collinearity. This finding is consistent with the results from simulation study 5. The GWL-global model performs the next best. This finding is in contrast to the finding in simulation study 5, where the GWRR models performed better than the GWL-global model. An explanation for this difference is that the GWL model can shrink the coefficients to zero for the variable with true coefficients set to zero to effectively remove its effect from the model, while the GWRR models cannot shrink these coefficients to zero as efficiently. The results in Table 6.13 suggest that one should use the Bayesian SVCP model when one is concerned with marginal inference on the regression coefficients in the presence of collinearity in the model. If analysts want to limit their modeling to the GWR framework, and they suspect there are insignificant

explanatory variables with no value in explaining the response variable, they should use the GWL-global model based on these findings.

	c = 0.0	c = 0.5	c = 0.7	c = 0.9
Method	RMSE(β)	RMSE(β)	RMSE(β)	RMSE(β)
GWR	0.538	0.601	0.756	1.745
GWRR - global	0.540	0.602	0.755	1.718
GWRR - local	0.539	0.604	0.763	1.597
GWL - global	**0.534**	0.597	0.754	1.596
GWL - local	1.487	1.625	1.754	2.794
SVCP	0.540	**0.591**	**0.649**	**1.068**

Table 6.13. RMSE of the regression coefficients for each model used in simulation study 6 at four levels of explanatory variable correlation.

CHAPTER 7

CONCLUSIONS

There has been an increasing interest in spatially varying relationships between variables in recent years in both the statistics and geography literature. Recent attempts at modeling these relationships have resulted in geographically weighted regression (GWR) and Bayesian regression models with spatially varying coefficient processes (SVCP). While GWR models offer the potential of increased understanding of the nature of varying relationships between variables across space, collinearity in the weighted explanatory variables can produce dependence in the local regression coefficients that can distort and potentially invalidate conclusions about the relationships based on the estimated coefficients.

This dissertation makes a contribution to the literature because it is the first work to both document the issue of collinearity in geographically weighted regression models using the diagnostic tools of scatter plots, correlation coefficients, variance inflation factors, and variance-decomposition proportions and also suggest a viable alternative while retaining the GWR framework. The analysis presented here shows that it is possible to use geographically weighted regression models with a ridge regression parameter (GWRR) to reduce the effect of local variable collinearity on the model while

producing more intuitively signed coefficients and surrendering only a modest amount of overall model fit. In addition, the GWRR model has lower prediction error through the stabilized variance of the parameters. The arguments presented here also show that it is possible to use the lasso in geographically weighted regression (GWL) to perform regression coefficient shrinkage, while simultaneously performing model selection. While these methods may not address all possible statistical artifacts in GWR, they are viable tools for spatial data analysts who wish to investigate spatially varying relationships between variables in a regression setting, and consider at the same time certain model complications arising from collinearity. There is natural appeal in explaining spatial variation in relationships through estimated regression coefficients, and the work presented here should contribute to making that a more reliable exercise.

While the GWR and SVCP models have been applied to numerous real world datasets in the literature, there has been a conspicuous lack of attention to the validation of marginal inferences derived from these models. My work uses simulation study to evaluate the accuracy of the regression coefficients for both the Bayesian SVCP and GWR models, while considering the presence of collinearity. While a simulation study does not prove a general result, it does show a result for a particular situation, and that result may be generalized with additional simulation studies.

Simulation study results in this dissertation show that the Bayesian regression model provides more accurate regression coefficient estimates than does GWR in both the absence and presence of explanatory variable collinearity and produces less correlated coefficients in the presence of moderate and strong positive variable correlation. It is not completely unexpected that the Bayesian SVCP model accommodates collinearity better

than GWR does, given that ridge regression coefficient estimates may be viewed as Bayesian regression coefficient estimates under specific vague priors, and ridge regression coefficients are constrained in size to counter increased estimator variance due to the presence of collinearity in the model. Moreover, the SVCP model better captures the spatial structure in the regression coefficients through the covariance matrix in the prior distribution for all the coefficients. The Bayesian regression model handles spatial dependence in the data through one statistical model, whereas GWR is essentially an ensemble of separate models using shared data. A benefit of using the Bayesian SVCP model is that it provides the posterior distribution for each parameter through the iteratively drawn samples, while the GWR model only gives a point estimate and corresponding standard error for each parameter. In summary, the Bayesian regression model offers the spatial analyst more flexibility in modeling spatially varying relationships, and produces more interpretable and accurate inferences than does GWR. There is additional complexity in implementing the Bayesian regression model compared to the GWR model; however, the simulation study results in this paper help to justify the benefits of the additional complexity of the Bayesian model.

Other simulation study results generated here show that the penalized versions of GWR introduced in this dissertation perform better than GWR in terms of response variable prediction and estimation and regression coefficient estimation, both when there is no collinearity and where there are various levels of collinearity in the model. The GWL-local model produces the lowest prediction error of the response variable among all the methods considered. The GWL-local and GWL-global methods produce the lowest average estimation error of the response variable. These methods also perform better than

the Bayesian SVCP model in response variable prediction and estimation. GWR also performs better than the SVCP in response variable estimation when all true regression coefficients used to simulate the data are non-zero, but not when there are true regression coefficients set to zero. One can explain the case when GWR is better because GWR is based on local linear regression methods that are good for prediction, but not model interpretation, whereas the Bayesian SVCP model specifically models the regression coefficient structure and is therefore better for interpretation of the coefficients. Considering the estimation of the regression coefficients, the SVCP model produces the lowest average estimation error overall, and is dramatically better than all the other methods when there is strong collinearity. The GWRR model is a distant second in terms of regression coefficient estimation error when there are no true coefficients set to zero; the GWL-global model is second best when there are true coefficients set to zero in generating the simulation data.

The simulation study results of this research imply: 1) if researchers are interested in using a linear regression model with spatially varying coefficients for prediction of the response variable, then they should use GWL with local shrinkage parameters, 2) if researchers are interested in using a linear regression model with spatially varying coefficients for estimation of the response variable, then they should use GWL with either global or local shrinkage parameters, 3) if researchers are interested in using a linear regression model with spatially varying coefficients for marginal inference on the regression coefficients, then they should use the Bayesian SVCP model, and 4) if researchers are interested in using a linear regression model with spatially varying coefficients in the GWR framework for marginal inference on the regression coefficients,

then they should use the GWRR model with either local or global scaling of the explanatory variables when they expect all non-zero regression coefficients and use the GWL-global model when they expect some variables with no effect across the study area.

In addition to attractive performance for regression coefficient estimation, another benefit of the Bayesian SVCP model is that it is possible to extend it to consider more complicated situations. For example, one can use a different spatial dependence parameter for each variable using the LMC, and one can also add varying temporal effects to the model. However, it is an open research question as to whether adding different spatial ranges for each variable has a positive effect on reducing estimated regression coefficient correlation. Also, it is unclear whether datasets of typical size can support estimation of numerous spatial dependence parameters.

While the results presented here for the GWR regularization methods are encouraging, more experimentation is needed to verify that they generalize for larger models and larger datasets. The GWRR model as presented here with one ridge parameter was adequate to substantially correct the collinearity present in the example dataset, but it may not be as corrective for large datasets. More research is possible to study the benefit of adding multiple ridge parameters. As many as one ridge parameter for each local model could be added, or a ridge parameter could be added for a group of observations, where the observation groups could be determined endogenously. Adding a ridge parameter for each local model would make the model more similar to the GWL-local model. Clearly, formal statistical tests would be beneficial to determine the number of ridge parameters to include in the regression model. Generalized cross-validation and an Akaike information criterion could prove useful in estimating multiple ridge

parameters, as cross-validation could be computationally prohibitive for large datasets. Another area of required future research is the use of the bootstrap procedure to estimate the variances in the GWRR methods.

One argument that has been used in the past against Bayesian models has been the difficulty in implementing them in practice. However, more general-purpose software for Bayesian inference is becoming available. At the forefront of the Bayesian software movement is WinBUGS (www.mrc-bsu.cam.ac.uk/bugs/winbugs/contents.shtml), which uses Gibbs sampling and other simulation methods for inference in a wide variety of Bayesian models. GeoBUGS is available in WinBUGS and can be utilized for Bayesian spatial models commonly found in epidemiological studies and other spatial applications. There are also MCMC packages available for R. All of the packages mentioned are free and available on the Internet. Inconveniently, GeoBUGS cannot currently perform Bayesian inference on the jointly specified SVCP model used in this paper, but one can imagine that will be able to in the future, as it can currently handle the conditionally specified SVCP model. Regardless, I hope the exposition of the SVCP model provided in this dissertation makes the implementation of this type of Bayesian model more apparent.

LIST OF REFERENCES

Agarwal DK, Gelfand AE (2005) Slice sampling for simulation based fitting of spatial data models. *Statistics and Computing* 15: 61 – 69

Anselin L (1988) *Spatial Econometrics: Methods and Models.* Kluwer: Dorddrecht

Anselin L (2003) An introduction to spatial autocorrelation analysis with GeoDa. GeoDa Documentation. URL: http://www.csiss.org/clearinghouse/GeoDa/

Banerjee S, Carlin BP, Gelfand AE (2004) *Hierarchical Modeling and Analysis for Spatial Data.* Chapman & Hall: Boca Raton

Banerjee S, Johnson GA (2005) Coregionalized single- and multi-resolution spatially-varying growth curve modelling with application to weed growth. UMN Biostat Tech Report

Belsley DA (1991) *Conditioning Diagnostics: Collinearity and Weak Data in Regression*. John Wiley & Sons: New York

Brewer CA, MacEachren AM, Pickle LW, Hermann D (1997) Mapping mortality: Evaluating color schemes for choropleth maps. *Annals of the Association of American Geographers* 87: 411-438

Brunsdon C, Fotheringham AS, Charlton M (1996) Geographically weighted regression: A method for exploring spatial nonstationarity. *Geographical Analysis* 28 (4): 281-298

Brunsdon C, Fotheringham AS, Charlton M (2002) Geographically weighted summary statistics – a framework for localized exploratory data analysis. Computers, Environment and Urban Systems 26: 501–524

Casella G, George EI (1992) Explaining the Gibbs sampler. *The American Statistician* 46 (3): 167 - 174

Chib S, Greenberg E (1995) Understanding the Metropolis-Hastings algorithm. *The American Statistician* 49 (4): 327 – 335

Congdon P (2003a) *Applied Bayesian Modelling*. John Wiley & Sons: West Sussex

Congdon P (2003b) Modelling spatially varying impacts of socioeconomic predictors on mortality outcomes. *Journal of Geographical Systems* 5: 161-184

Congdon P (2004) A multivariate model for spatio-temporal health outcomes with an application to suicide mortality. *Geographical Analysis* 36 (3): 234-258

Devesa SS, Grauman DJ, Blot WJ, Pennello G, Hoover RN, Fraumeni JF Jr. (1999) *Atlas of Cancer Mortality in the United States, 1950-94.* National Cancer Institute: Bethesda. URL: http://www3.cancer.gov/atlasplus/

Efron B, Hastie T, Johnstone I, Tibshirani R (2004a) Least angle regression. *Annals of Statistics* 32 (2): 407-451.

Efron B, Hastie T, Johnstone I, Tibshirani R (2004b) Rejoinder to least angle regression. *Annals of Statistics* 32 (2): 494-499

Fotheringham AS, Brunsdon C, Charlton M (2002) *Geographically Weighted Regression: The Analysis of Spatially Varying Relationships*. John Wiley & Sons: West Sussex

Fotheringham AS, Charlton M, Brunsdon C (1998) Geographically weighted regression: A natural evolution of the expansion method for spatial data analysis. *Environment and Planning A* 30 (11): 1905-1927

Fox J (1997) *Applied Regression Analysis, Linear Models, and Related Methods*. Sage Publications: Thousand Oaks

Frank IE, Friedman JH (1993) A statistical view of some chemometrics regression tools. *Technometrics* 35 (2)

Gelfand AE, Banerjee S, Gamerman D (2005) Spatial process modeling for univariate and multivariate dynamic spatial data. *Environmetrics* 16: 465 – 479

Gelfand AE, Kim H, Sirmans CF, Banerjee S (2003) Spatial modeling with spatially varying coefficient processes. *Journal of the American Statistical Association* 98: 387 - 396

Gelfand AE, Schmit AM, Banerjee S, Sirmans CF (2004) Nonstationary multivariate process modeling through spatially varying coregionalization. *Test* 13 (2): 263-312

Gelman A, Carlin JB, Stern HS, Rubin DB (2003) *Bayesian Data Analysis* (2nd ed.). Chapman & Hall: London

Gelman A, Price PN (1999) All maps of parameter estimates are misleading. *Statistics in Medicine* 18: 3221-3234

Goldstein M (1976) Bayesian analysis of regression problems. *Biometrika* 63 (1): 51 – 58

Golub GH, Heath M, Wahba G (1979) Generalized cross-validation as a method for choosing a good ridge parameter. *Technometrics* 21 (2): 215 – 223

Grandvalet Y (1998) Least absolute shrinkage is equivalent to quadratic penalization. In Niklasson L, Boden M, Ziemske T (eds) ICANN'98, volume 1 of Perspectives in Neural Computing. Springer-Verlag: Berlin, 201-206

Greene WH (2000). *Econometric Analysis*. Prentice-Hall: Upper Saddle River

Griffith D (2003) *Spatial Autocorrelation and Spatial Filtering*. Springer-Verlag: Berlin

Hastie T, Tibshirani R, Friedman J (2001) *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer-Verlag: New York

Hoerl AE, Kennard RW (1970) Ridge regression: biased estimation for non-orthogonal problems. *Technometrics* 12: 55 – 67

Huang Y, Leung Y (2002) Analysing regional industrialisation in Jiangsu province using geographically weighted regression. *Journal of Geographical Systems* 4: 233-249

LeSage JP (2004) A family of geographically weighted regression models. In: Anselin L, Florax RJGM, Rey SJ (eds) *Advances in spatial econometrics. Methodology, tools and applications.* Springer Verlag, Berlin pp 241-264

Leung Y, Mei CL, Zhang WX (2000a) Statistical tests for spatial nonstationarity based on the geographically weighted regression model. *Environment and Planning A* 32: 9-32

Leung Y, Mei CL, Zhang WX (2000b) Testing for spatial autocorrelation among the residuals of the geographically weighted regression. *Environment and Planning A* 32: 871-890

Lindley DV, Smith AFM (1972) Bayes estimates for the linear model. Conference Proceedings, Royal Statistical Society

Loader C (1999) *Local Regression and Likelihood*. Springer: New York

Longley JW (1967) An appraisal of least squares programs from the point of the user. *Journal of the American Statistical Association* 62: 819-841

Longley PA, Tobón C (2004) Spatial dependence and heterogeneity in patterns of hardship: an intra-urban analysis. *Annals of the Association of American Geographers* 94: 503-519

Martinez WL, Martinez AR (2002) *Computational Statistics Handbook with Matlab*. Chapman & Hall: Boca Raton

Mehnert WH, Smans M, Muir CS, Möhner M, Schön D (1992) *Atlas of Cancer Incidence in the Former German Democratic Republic 1978-1982*. Oxford University Press: New York

Mei CL, He SY, Fang KT (2004) A note on the mixed geographically weighted regression model. *Journal of Regional Science* 44 (1): 143-157

Nakaya T (2001) Local spatial interaction modelling based on the geographically weighted regression approach. *GeoJournal* 53: 347-358

Neal RM (2003) Slice sampling. *The Annals of Statistics* 31 (3): 705-767

Neter J, Kutner MH, Nachtsheim CJ, Wasserman W (1996) *Applied Linear Regression Models*. Irwin: Chicago

Páez A, Uchida T, Miyamoto K (2002a) A general framework for estimation and inference of geographically weighted regression models: 1. Location-specific kernel bandwidths and a test for locational heterogeneity. *Environment and Planning A* 34: 733-754

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

Seifert B, Gasser T (2000) Data adaptive ridging in local polynomial regression. *Journal of Computational and Graphical Statistics* 9: 338 – 360

Tibshirani R (1996) Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society B* 58 (1): 267-288

Tiefelsdorf M (2003) Misspecifications in interaction model distance decay relations: A spatial structure effect. *Journal of Geographical Systems* 5: 25-50

Welsch R (2000) Is cross-validation the best approach for principal component and ridge regression? *Computing Science and Statistics* 32: 356 – 361

Wheeler D (2006) Diagnostic tools and a remedial method for collinearity in geographically weighted regression. *Environment and Planning A*, forthcoming

Wheeler D, Calder CA (2006) An assessment of coefficient accuracy in linear regression models with spatially varying coefficients. Department of Statistics Preprint No. 777, The Ohio State University

Wheeler D, Tiefelsdorf M (2005) Multicollinearity and correlation among local regression coefficients in geographically weighted regression. *Journal of Geographical Systems* 7: 161 – 187