J Geogr Syst (2009) 11:1–22
DOI 10.1007/s10109-008-0073-5

ORIGINAL ARTICLE

Comparing spatially varying coefficient models: a case study examining violent crime rates and their relationships to alcohol outlets and illegal drug arrests

David C. Wheeler · Lance A. Waller

Received: 15 July 2008 / Accepted: 7 November 2008 / Published online: 19 December 2008

Abstract In this paper, we compare and contrast a Bayesian spatially varying coefficient process (SVCP) model with a geographically weighted regression (GWR) model for the estimation of the potentially spatially varying regression effects of alcohol outlets and illegal drug activity on violent crime in Houston, Texas. In addition, we focus on the inherent coefficient shrinkage properties of the Bayesian SVCP model as a way to address increased coefficient variance that follows from collinearity in GWR models. We outline the advantages of the Bayesian model in terms of reducing inflated coefficient variance, enhanced model flexibility, and more formal measuring of model uncertainty for prediction. We find spatially varying effects for alcohol outlets and drug violations, but the amount of variation depends on the type of model used. For the Bayesian model, this variation is controllable through the amount of prior influence placed on the variance of the coefficients. For example, the spatial pattern of coefficients is similar for the GWR and Bayesian models when a relatively large prior variance is used in the Bayesian model.

Keywords GWR · Bayesian regression · Collinearity · Penalization methods

JEL Classification C11 · C13 · C21

D. C. Wheeler (✉)
Harvard School of Public Health, Harvard University,
677 Huntington Ave, Boston, MA 02115, USA
e-mail: dwheeler@hsph.harvard.edu

D. C. Wheeler
National Cancer Institute, 6120 Executive Blvd, Rockville, MD 20852, USA

L. A. Waller
Rollins School of Public Health, Department of Biostatistics,
Emory University, 1518 Clifton Road, Atlanta, GA 30322, USA

1 Introduction

Statistical models enable estimation of associations between an outcome of interest and a set of covariates measured on the same observational units. Statistical linear model theory provides the pervasive analytic tool of linear regression for Gaussian outcomes. When applying linear regression to spatially referenced data, one often encounters complications due to the violation of the independence of model errors assumption that underlies standard regression methodology and the assumed constant association between covariates and a response variable. The first assumption is clearly violated by spatial autocorrelation, where with positive autocorrelation observations that are spatially proximate are more likely to be similar than those farther apart. Spatial correlation may be the result of direct influences between observations, as in the case of time to first infection in neighboring areas in the study of an infectious disease outbreak due to interactions between neighboring populations, or indirect influences of unmeasured yet important environmental characteristics affecting both the response and covariate values, e.g. elevation in the case of disease vector distribution. While analytic methods for addressing spatial autocorrelation have appeared in the literature for several decades, the second issue of non-constant response–covariate associations across a study area only recently is receiving increasing attention in the statistics and geography literatures and we focus our attention here.

In this paper, we contrast two approaches for estimating spatially varying associations between alcohol sales distribution and illicit drug activity and violent crime rates in Houston, Texas. The approaches are a Bayesian spatially varying coefficient process (SVCP) model (Gelfand et al. 2003; Banerjee et al. 2004) and a geographically weighted regression (GWR) model (Fotheringham et al. 2002). In making contrasts between the two approaches, we focus on and illuminate the inherent coefficient shrinkage properties of the Bayesian SVCP model as a way to address increased coefficient variance that often follows from collinearity in the GWR model. Such collinearity often leads to correlated regression coefficients which can prove problematic for inferential purposes. Previous work (Wheeler and Tiefelsdorf 2005; Wheeler 2007; Griffith 2008) demonstrates collinearity issues with GWR, highlighting the relevance of and need for a methodological comparison. In addition, we explore flexibility in controlling the amount of coefficient spatial variation through model specification within the Bayesian SVCP model through the specification of the prior distribution for the covariance matrix of the coefficients.

One of our goals in this analysis of violent crime is to use as close a comparison as is possible between the basic GWR model and a hierarchical Bayesian random effects model in order to focus attention on the fundamental differences between the models in terms of assumptions, implementation, interpretation, and inference. Initial work (Brunsdon et al. 1999) comparing GWR and hierarchical models focused on exchangeable random effects, i.e. those with no prior spatial structure. While motivating, the conclusion drawn by the authors of this work that estimated GWR coefficients are spatially smoother than those from the non-spatial random effects model is hardly surprising. A more recent comparison of GWR and random effects

models that does consider spatial correlation is by Waller et al. (2007), who use a Bayesian spatially varying coefficient (SVC) model that has a multivariate conditionally autoregressive spatial random effects (prior) distribution in a SVC Poisson regression of crime events. Such a model induces spatial smoothing of the random effects (the local regression coefficients) by borrowing information from adjacent neighbors. Waller et al. (2007) compare results from the Bayesian SVC approach to results from a GWR model with a continuous Gaussian spatial kernel function. As a result, the spatial correlation structures are not completely comparable, with the Gaussian kernel in GWR using a larger number of observations to estimate local regression coefficients than the SVC model, and this difference in the form of spatial correlation functions could conceivably cause some differences in results.

The spatial analysis presented below expands on the analysis of violent crime in Houston by Waller et al. (2007) by using a Bayesian spatial process model which makes use of a joint specification for regression coefficients instead of the conditional specification previously considered. The joint specification allows more directly comparable exponential spatial dependence functions in the two spatially varying coefficient models, and using Gaussian linear regression to model (transformed) crime rates places the comparison on familiar ground for comparison with GWR. The spatial correlation matrix for all observations in the joint specification of the regression coefficients overcomes the use of overlapping ad hoc kernel functions in GWR. The analysis makes a contribution to the literature by focusing on and illuminating certain properties that drive behavior of the Bayesian SVCP model as well as through summarizing several advantages of the Bayesian model over GWR for constraining inflated variance of coefficients, increased model flexibility, and enhanced handling of model uncertainty for prediction. This work contributes to the simulation-study based comparison of GWR and Bayesian SVCP models in Wheeler and Calder (2007) by illuminating fundamental reasons for the observed differences in behavior of the models. Hence, in addition to analyzing crime rates themselves, we also seek to highlight and compare the underlying structures of GWR and SVCP models. In the sections below, we review the motivation for the spatial analysis of violence, illegal drug use, and alcohol; we define the inferential structure of SVC and GWR models; and we illustrate how the inferential structure impacts the conclusions drawn in our analysis.

2 Violence, drugs, and alcohol in Houston

Many theories of crime and empirical explanations of local violent crime rates focus on associations with characteristics of the people living in those neighborhoods or with characteristics of the places in which they live (Sampson and Lauritsen 1994). As a result, spatial research in the crime literature often focuses upon the integration of theories suggesting mechanisms by which characteristics of places, and aspects of local populations affect crime rates across communities. Notably, many previous analyses suggest that crime rates in one area may affect (directly or indirectly) crime rates in nearby areas, resulting in positive spatial autocorrelation among neighboring crime rates.

While incorporating spatial correlation in our models, our motivating example digs deeper and uses spatially varying coefficient models to explore whether place characteristics quantitatively impact violence differently in different geographic locations. More specifically, we compare the observed spatial distribution of alcohol outlets and illicit drug use in Houston to the general theoretical concept of crime potential, the location-specific probability that crime will be exhibited as a function of people or place characteristics in that location (Brantingham and Brantingham 1993, 1999). We focus on alcohol outlets because of previous research investigating the local rates of interpersonal violence in and around locations of alcohol outlets, particularly bars and taverns (Roncek and Maier 1991; Scribner et al. 1995; Gorman et al. 2001; Lipton and Gruenewald 2002; Gruenewald et al. 2006). Our interest here is whether the observed associations are constant across our study area or whether the associations are stronger in particular parts of the city than in others. One theoretical argument for a spatially varying coefficient model in this context is that the association between alcohol distribution and crime is likely modified by unmeasured neighborhood effects and allowing the association to vary can identify regions with stronger or weaker association in order to seek to identify the local drivers of the strength of the local association.

To investigate potentially spatially varying relationships between alcohol outlets, illegal drug activity and violent crime in Houston, we combine three datasets for 439 tracts from the 2000 US Census within the City of Houston. The first dataset used in the study comes from the Houston police department and includes monthly reports of four violent crime categories (rape, murder, robbery, and aggravated assault) for the City of Houston. Approximately 98% of the violent crime locations were geocoded and then aggregated to the census tract level. The violent crime data are based on first reports of offenses, before investigation and final classification of crimes. Such data have been used in previous studies of alcohol availability and violent crime (Nelson et al. 2001). The second dataset describes the location of alcohol outlets in Houston during the year 2000 and comes from the Texas Alcoholic Beverage Commission. It contains a total of 6,609 outlets and includes the name, geographic location, and type of permit or license of each outlet. Practically all of the outlets (99.5%) were geocoded by street address using Centrus Desktop (Group 1 Software 2003). The third dataset used is comprised of the Houston police department monthly reports of drug-law violations. Approximately 98% of the 9,985 drug crimes reported for the year 2000 were successfully geocoded and aggregated to census tracts for the analysis. We aggregate the data to census tract level to maintain confidentiality, and census tracts provide a useful and convenient spatial unit of analysis. Of course, our results are conditional on the selected scale of analysis, i.e., the familiar modifiable areal unit problem limits conclusions to the level of census tracts.

In the analysis that follows, the violent crime rate for each census tract represents the outcome variable of interest. We assume a normal approximation to the Poisson distribution for the rates. Alcohol outlets and drug-crime violations per census tract population represent the covariates. Since the distribution of violent crime rates is skewed, we take the natural logarithm of violent crime rates as the response variable. To maintain linear relationships with the transformed response variable,

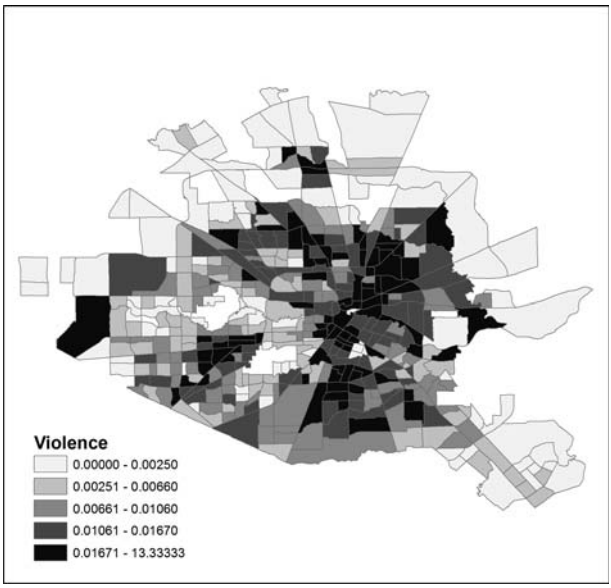

Fig. 1 Violent crime rate by census tract in the City of Houston

we also use the natural logarithm of each covariate for analysis. Violent crime rates by census tract are displayed in Fig. 1. The data reveal that some census tracts experience high rates of violence, and in some cases the number of violent crime reports exceeds the population size, for example, the high violence rate tract on the western edge of the study area contains 40 violent crime reports but only three residents. These are predominantly census tracts with very low population sizes and it is important to note that the report of a violent crime in a tract does not imply that either the victim or perpetrator of the crime were residents of the tract. While US census tracts typically contain approximately 3,000–4,000 residents, very low population count census tracts can, and do, occur in urban areas where tracts are geographically small and regions transition from residential to other uses between two census enumerations. Such tracts are often consolidated with neighboring tracts in subsequent census enumerations.

3 Overview of spatially varying coefficient models

3.1 Geographically weighted regression

The technical details underlying GWR have been described elsewhere (Fotheringham et al. 2002), but we review the basics here for completeness. In GWR, a regression model can be fitted at each observation location in the dataset. The spatial coordinates of the data points are used to calculate inter-point distances, which are input into a kernel function to calculate weights that represent spatial dependence

between observations. For each location, $s = 1, \ldots, n$, where the model is calibrated, the GWR model is

$$y(s) = \mathbf{X}(s)\boldsymbol{\beta}(s) + \varepsilon(s), \tag{1}$$

where $y(s)$ is the dependent variable at location s, $\boldsymbol{\beta}(s)$ is a column vector of regression coefficients at location s, $\mathbf{X}(s)$ is a row vector of explanatory variables at location s, and $\varepsilon(s)$ is the random error at location s. The vector of estimated regression coefficients at location s is

$$\hat{\boldsymbol{\beta}}(s) = [\mathbf{X}^T \cdot \mathbf{W}(s) \cdot \mathbf{X}]^{-1}\mathbf{X}^T \cdot \mathbf{W}(s) \cdot \mathbf{y}, \tag{2}$$

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

The weights matrix, $\mathbf{W}(s)$, is calculated from a kernel function which places more weight on observations that are closer to the model calibration location s. We use the exponential function for the kernel function in this paper to match the form of the spatial dependence function used in the Bayesian SVCP model. The weights from the exponential kernel function are calculated as

$$w_j(s) = \exp(-d_{sj}/\gamma), \tag{3}$$

where d_{sj} is the distance between the calibration location s and location j, and γ is the kernel bandwidth parameter.

To fit the GWR model to the data, we first estimate the kernel bandwidth using cross-validation, which is an iterative process that finds the kernel bandwidth that minimizes the prediction error of all the $y(s)$. Next, we calculate the weights at each calibration location using the estimated kernel function. Finally, we estimate the regression coefficients at each calibration location with Eq. 2 and then the response variable estimates by the expression $\hat{y}(s) = \mathbf{X}(s)\hat{\boldsymbol{\beta}}(s)$. The model calibration locations are census tract centroids in this study. Due to the attention in the literature (Wheeler and Tiefelsdorf 2005; Wheeler and Calder 2007; Griffith 2008) on potentially strong correlation in sets of estimated GWR coefficients, we are interested in the overall correlation of estimated coefficients for pairs of different regression terms. We refer to the covariates and the intercept as regression terms to distinguish them from the local coefficient parameters associated with them. To measure the correlation in a set of estimated regression coefficients for a pair of regression terms, we calculate the Pearson correlation coefficient and denote it r_{kl} for terms k and l over all locations in the study area.

3.2 Bayesian SVCP model

The Bayesian SVCP regression model has also been described in detail in other work (Gelfand et al. 2003; Wheeler and Calder 2007) and we provide only an

overview of the model here. The model is a hierarchical one, where the distribution of the data is specified conditional on unknown parameters, whose distribution is in turn specified conditional on other parameters. The SVCP model is specified as

$$\left[\mathbf{Y}|\boldsymbol{\beta}, \tau^2\right] = N(\mathbf{X}^*\boldsymbol{\beta}, \tau^2\mathbf{I}), \tag{4}$$

where $[\mathbf{Y}|\boldsymbol{\beta}, \tau^2]$ denotes the distribution of $\mathbf{Y}$ conditional on parameters $\boldsymbol{\beta}$ and τ^2. $\mathbf{Y}$ is a $n \times 1$ vector of responses, assumed to be Gaussian and conditional on $\boldsymbol{\beta}$ and τ^2; $\boldsymbol{\beta}$ is a $np \times 1$ vector of regression coefficient parameters with different values at each data point; and $\mathbf{X}^*$ 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 $\mathbf{X}^*$); $\mathbf{I}$ is the $n \times n$ identity matrix; and τ^2 is the error variance.

The prior distribution for the regression coefficient parameters is specified as

$$\left[\boldsymbol{\beta}|\boldsymbol{\mu}_{\boldsymbol{\beta}}, \Sigma_{\boldsymbol{\beta}}\right] = N(\mathbf{1}_{n\times 1} \otimes \boldsymbol{\mu}_{\boldsymbol{\beta}}, \Sigma_{\boldsymbol{\beta}}), \tag{5}$$

where the vector $\boldsymbol{\mu}_{\boldsymbol{\beta}} = (\mu_{\beta_0}, \ldots, \mu_{\beta_p})^T$ contains the means of the regression terms. As specified, the regression coefficients for one regression term have the same mean; the Kronecker product operator ($\otimes$) multiplies every element in $\mathbf{1}_{n \times 1}$ by $\boldsymbol{\mu}_{\boldsymbol{\beta}}$. Different specifications of the means are possible with the flexibility of the Bayesian model paradigm and we elaborate on this point later. The prior on the regression coefficients takes into account possible spatial dependence in the coefficients through the covariance, $\boldsymbol{\Sigma}_{\boldsymbol{\beta}}$, which has a separable form with two distinct components, one for the within site dependence between coefficients and one for the spatial dependence in the regression coefficients. This variance–covariance matrix captures the covariation within and between all regression coefficients simultaneously, in contrast to the repeated application of spatial kernel functions in GWR. The separable form of the covariance matrix for $\boldsymbol{\beta}$ is

$$\boldsymbol{\Sigma}_{\boldsymbol{\beta}} = \mathbf{H}(\phi) \otimes \mathbf{T}, \tag{6}$$

where $\mathbf{H}(\phi)$ is the $n \times n$ correlation matrix that captures the spatial association between the n locations, ϕ is an unknown spatial dependence parameter, and $\mathbf{T}$ is a positive-definite $p \times p$ matrix for the covariance of the regression coefficients at any spatial location. For $\boldsymbol{\beta}_k = [\beta_{1k}, \ldots, \beta_{nk}]$, we assume that each $\boldsymbol{\beta}_k$ follows a geostatistical process, specified with an exponential correlation function. In the separable form of the covariance matrix, the Kronecker product results in a $np \times np$ positive definite covariance matrix because $\mathbf{T}$ and $\mathbf{H}(\phi)$ are both positive definite. The elements of the correlation matrix $\mathbf{H}(\phi)$, $H(\phi)_{ij} = \rho(s_i - s_j; \phi)$, are calculated using the function $\rho(d; \phi) = \exp(-d/\phi)$, where d is the distance between locations s_i and s_j. In the separable covariance matrix, each of the p coefficients represented in the covariance is assumed to have the same spatial dependence structure. This aligns with the assumption of equal spatial ranges for each regression term in the GWR model, and justifies the use of the separable form.

The specification of the Bayesian SVCP model is complete with the specification of the prior distributions of the remaining parameters. The prior for the coefficient means is normal with hyperparameters $\boldsymbol{\mu}$ and σ^2, $\left[\boldsymbol{\mu}_{\boldsymbol{\beta}}\right] \sim N(\boldsymbol{\mu}, \sigma^2\mathbf{I})$, where the

hyperparameters control the mean and variance of the prior distribution. The prior for the within-site covariance matrix $\mathbf{T}$ is inverse Wishart with hyperparameters v and $\mathbf{\Omega}$, $[\mathbf{T}] \sim \text{IW}_v(\mathbf{\Omega}^{-1})$. The prior for the error variance is inverse gamma with hyperparameters a and b, $[\tau^2] \sim \text{IG}(a, b)$. 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, $[\phi] \sim \text{G}(\alpha, \lambda)$.

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

$$[\boldsymbol{\theta}|\mathbf{y}] \propto [\mathbf{y}|\boldsymbol{\theta}][\boldsymbol{\theta}]. \tag{7}$$

The posterior distribution for the parameters $\boldsymbol{\theta}$, conditional on the data, is proportional to the product of the likelihood of the data $[\mathbf{y}|\boldsymbol{\theta}]$ and the prior $[\boldsymbol{\theta}]$ for all the parameters. We use Markov chain Monte Carlo (MCMC) to sample from the posterior distribution of the parameters and base inferences on these samples. See Wheeler and Calder (2007) for implementation details of the MCMC for the SVCP model.

A feature of the Bayesian SVCP model is that the correlation between regression terms is explicitly modeled. In using the separable covariance matrix, the posterior correlation of the regression coefficients for terms k and l across all locations is $T_{kl}/\sqrt{T_{kk}T_{ll}}$. We use this expression for coefficient correlation in the next section when evaluating dependence in regression coefficients from the violent crime model.

3.3 Bayesian SVCP model and coefficient shrinkage

In comparing the GWR and SVCP models on similar footing, features in model properties become apparent. One such feature is the similarity between the Bayesian SVCP model and ridge regression, which allows us to summarize the nature of the Bayesian shrinkage in the regression coefficients induced by the SVCP model. There are different types of coefficient shrinkage, including shrinkage of a coefficient to zero, to its overall mean, or to its local mean. As a brief background, there has been discussion in the statistics literature about ridge regression solutions as Bayes estimates. Hoerl and Kennard (1970) first introduced ridge regression to overcome ill conditioned design matrices. Ridge regression coefficients minimize the residual sum of squares along with a penalty on the size of the squared coefficients as

$$\hat{\boldsymbol{\beta}}^R = \arg\min_{\boldsymbol{\beta}} \left\{ \sum_{i=1}^{n} \left(y_i - \beta_0 - \sum_{k=1}^{p} x_{ik}\beta_k \right)^2 + \lambda \sum_{k=1}^{p} \beta_k^2 \right\}, \tag{8}$$

where λ is the ridge regression parameter that controls the amount of shrinkage in the regression coefficients (Hastie et al. 2001). 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) also describe the ridge regression solutions as Bayes estimates, where

ridge regression uses independent normal distributions for the prior of each coefficient β_k. 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 (8), with $\lambda = \tau^2/\sigma^2$, where τ^2 is the error variance. Specifically, in the Bayesian regression model with $y \sim N(\mathbf{X}\boldsymbol{\beta}, \tau^2\mathbf{I})$ and an independent prior $\beta \sim N(0, \sigma^2)$ for each coefficient, the posterior for the coefficients can be expressed as

$$[\boldsymbol{\beta}|\tau^2, \sigma_\beta^2; y] \propto \exp\left(-\frac{1}{2\tau^2}\sum_{i=1}^{n}\left(y_i - \sum_{k=1}^{p} x_{ik}\beta_k\right)^2\right)\exp\left(-\frac{1}{2\sigma^2}\sum_{k=1}^{p}(\beta_k - 0)^2\right), \tag{9}$$

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

$$\sum_{i=1}^{n}\left(y_i - \sum_{k=1}^{p} x_{ik}\beta_k\right)^2 + \frac{\tau^2}{\sigma^2}\sum_{k=1}^{p}\beta_k^2, \tag{10}$$

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 when based on 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. Note that the shrinkage in ridge regression is towards 0 for each regression coefficient.

The view of ridge regression solutions as Bayes estimates suggests that we can also develop the view of the Bayesian SVCP model coefficients as spatial ridge regression estimates because of the normal distribution prior for the regression coefficients in the SVCP model. The prior in the SVCP is more complicated than the independent normal prior in the traditional Bayesian regression model, however, due to the spatial correlation component in the variance–covariance matrix. The posterior distribution for the coefficients of the SVCP model can be expressed with convenient notation as

$$\begin{aligned}[\beta|\tau^2, \boldsymbol{\mu}_{\boldsymbol{\beta}}, \Sigma_{\boldsymbol{\beta}}; y] &\propto \exp\left(-\frac{1}{2}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T(\tau^2\mathbf{I})^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})\right) \\ &\quad \times \exp\left(-\frac{1}{2}(\boldsymbol{\beta} - (1 \otimes \boldsymbol{\mu}_{\boldsymbol{\beta}}))^T\Sigma_\beta^{-1}(\boldsymbol{\beta} - (\mathbf{1} \otimes \boldsymbol{\mu}_{\boldsymbol{\beta}}))\right) \\ &\propto \exp\Big(-\frac{1}{2}[(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) + (\boldsymbol{\beta} - (\mathbf{1} \otimes \boldsymbol{\mu}_{\boldsymbol{\beta}}))^T \\ &\quad \times \tau^2(\mathbf{H}(\phi) \otimes \mathbf{T})^{-1}(\boldsymbol{\beta} - (\mathbf{1} \otimes \boldsymbol{\mu}_{\boldsymbol{\beta}}))]\Big). \end{aligned} \tag{11}$$

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

$$(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) + (\boldsymbol{\beta} - (\mathbf{1} \otimes \boldsymbol{\mu}_{\boldsymbol{\beta}}))^T\tau^2(\mathbf{H}(\phi) \otimes \mathbf{T})^{-1}(\boldsymbol{\beta} - (\mathbf{1} \otimes \boldsymbol{\mu}_{\boldsymbol{\beta}})), \tag{12}$$

where the shrinkage term is now defined by a matrix $\boldsymbol{\Lambda} = \tau^2 \cdot \mathbf{H}^{-1}(\phi) \otimes \mathbf{T}^{-1}$. Therefore, the amount of shrinkage on $\boldsymbol{\beta}$ in the SVCP model depends on τ^2, ϕ, and

T, each of which is a variance term. The shrinkage is still towards the mean vector $\boldsymbol{\mu}_{\boldsymbol{\beta}}$, but in contrast to ridge regression, could include nonzero terms, and, again in contrast to the non-spatial setting, the shrinkage now involves a spatial covariance matrix through parameter ϕ. There is considerable flexibility in the Bayesian framework to specify different values for $\boldsymbol{\mu}_{\boldsymbol{\beta}}$ through the prior. For instance, we could have as a base the ordinary least squares (OLS) regression estimates for coefficients at all locations by setting the prior mean for $\boldsymbol{\mu}_{\boldsymbol{\beta}}$ to the OLS values, in the spirit of empirical Bayesian analysis. The estimated coefficients could then be viewed as variations over space from these global values, where the variations are impacted in part by spatial correlation. This would provide an alternative to those concerned that GWR estimates are not tied in any way to the global OLS model. To be more in agreement with ridge regression, we could also shrink towards 0 by setting the prior mean for $\boldsymbol{\mu}_{\boldsymbol{\beta}}$ to zero and using a small prior variance. This corresponds to assuming no association between the covariates and outcome a priori, then allowing the data to update this prior assumption through the likelihood to arrive at posterior inference. In a similar spirit, we could use another informative prior for $\boldsymbol{\mu}_{\boldsymbol{\beta}}$ based on previous studies or beliefs. Additionally, we could use a vague prior for $\boldsymbol{\mu}_{\boldsymbol{\beta}}$ centered on a vector of zeros with large prior variances to allow the data to determine the mean values for the coefficients with little influence of prior assumptions. Finally, we could allow the $\boldsymbol{\mu}_{\boldsymbol{\beta}}$ to vary by location by changing the dimension of this mean vector to np instead of p. This option requires a change to the full conditional distributions of the parameters as outlined in Wheeler and Calder (2007) and increases the computational complexity of the associated Markov chain Monte Carlo algorithms needed to fit the models.

In keeping τ^2 and ϕ constant, the shrinkage factor is controlled by the covariance matrix **T**. In **T**, the terms that control the spatial variation in the regression coefficients are the variances, i.e. the diagonal terms in the matrix, $T_{11}, \ldots, T_{pp}$. If τ^2 and ϕ are held constant and the variances are increased in the diagonal of **T**, then the shrinkage factor will decrease and coefficients may vary more from their respective overall means. Conversely, if the variances in **T** are decreased while holding τ^2 and ϕ steady, then the shrinkage factor will increase and the regression coefficients may shrink more to their respective overall means. In the Bayesian framework, the regression coefficient variances in **T** may be effectively increased or decreased through the prior. In summary, the variance terms control the amount of shrinkage in the coefficients and the prior for $\boldsymbol{\mu}_{\boldsymbol{\beta}}$ controls toward what values the coefficients shrink.

It is worthwhile to note that there is no coefficient shrinkage in traditional GWR, and given the documented issue of inflated coefficient variances in several cases, this is a liability of the method. There is nothing in GWR to constrain the estimated regression coefficients in the situations where inflated coefficient variances artificially increase the magnitude of the coefficients. Wheeler and Tiefelsdorf (2005) show that increasing correlation in the weighted design matrix in a GWR model can lead to increased variance in estimated regression coefficients and increased correlation in estimated coefficients for pairs of regression terms which in turn results in strong co-patterning in maps of coefficients. Wheeler (2007) formalizes this inflated variance in GWR coefficients through a set of diagnostic tools that link the kernel weighted design matrix of GWR to ill-conditioning and

shared variances in the GWR coefficient covariance matrix, and we make use of these diagnostics with the Houston crime data in the next section. To control this variance in estimated regression coefficients from GWR, Wheeler (2007) developed a ridge regression version of GWR that shrinks coefficients towards zero when the weighted design matrix is ill-conditioned. In this constrained version of GWR, spatial correlation is modeled exactly as in basic GWR, through a spatial kernel function that weights observations based on geographic distance. The estimated coefficients then are a function of smoothing across observations and shrinking to zero. The shrinkage is not explicitly based on shrinking to neighboring values, but spatial correlation is nonetheless still present in the model.

The shrinkage of the Bayesian SVCP model is more theoretically appealing to some than the shrinkage in ridge regression towards zero. The technical details presented above highlight the manner in which the SVCP model can shrink local parameter estimates to compromise values defined by the prior mean and spatial covariance structure. The compromise is between unconstrained local coefficients that may have too much variation and the constant global estimates that may be unrealistic in some local areas. The model shrinks the regression coefficients to their overall means, but the variance parameter allows the posterior to move away from the overall means based on the data. Also, the shrinkage of the regression coefficients is explicitly a function of spatial dependence. It has been shown that some constraint is needed for GWR coefficients, and the Bayesian SVCP model offers constrained regression coefficients that vary locally about their global means and where spatial correlation is considered.

4 Results

4.1 Estimating Houston violent crime rates with spatially varying coefficient models

In this section, we present the results of estimating the GWR and Bayesian SVCP model parameters for the Houston violent crime data. For the violent crime data, the base model is

$$y(i) = \beta_0 + \beta_1 x_1(i) + \beta_2 x_2(i) + \varepsilon(i), \tag{13}$$

where y is the natural log of the number of violent crimes (murder, robbery, rape, and aggregated assault) per 100 persons, x_1 is the natural log of the number of drug law violations per 100 persons, x_2 is the natural log of the number of alcohol outlets per 100 persons, and i is the index for census tracts. We first fitted an OLS regression model for Eq. 13. The coefficient of determination was 0.67 and the root mean square error (RMSE) of the estimated response variable was 0.77. The estimated regression coefficients were $\hat{\beta}_0 = 0.94, \hat{\beta}_1 = 0.61, \hat{\beta}_2 = 0.21$ and all were significant at the 0.001 level. The positive relationships between violent crime and drug violations and alcohol sales are intuitive and agree with previous social science research. The OLS model explains a substantial amount of the variation in violent crime rates in Houston with two covariates.

We next fitted the GWR and Bayesian SVCP models using spatially varying coefficients in the linear expression for crime rates

$$y(i) = \beta_0(i) + \beta_1(i)x_1(i) + \beta_2(i)x_2(i) + \varepsilon(i). \quad (14)$$

Through cross-validation, the estimated GWR kernel bandwidth $\hat{\gamma}$ was 0.89. The RMSE of the response variable for the GWR model with this estimated bandwidth and regression coefficients estimated with the bandwidth is 0.34, a considerable reduction from the OLS model.

For the SVCP model, the prior specification for the model is as follows. We use a vague normal, N($\mathbf{0}$, $10^4\mathbf{I}$), for $\boldsymbol{\mu}_{\boldsymbol{\beta}}$, a vague inverse Wishart, IW($p + 1$, $0.1\mathbf{I}$), for $\mathbf{T}$, and a vague inverse gamma, IG(1, 0.01), for τ^2, where $\mathbf{I}$ is the identity matrix of dimension p. For the spatial dependence parameter ϕ, we use a gamma, Gamma(0.07, 0.01), which has a mean of 7 and variance of 700. Random starting values are used for the parameter Markov chains. Through MCMC sampling of the parameter joint posterior distribution, the spatial dependence parameter in the Bayesian SVCP model is estimated as $\hat{\phi} = 0.732$ using the posterior mean. A sample size of 4,000 was used in the MCMC, following a "burn-in" of 4,000 iterations. The posterior mean estimate of the error variance is $\hat{\tau}^2 = 0.0197$ and the posterior mean estimates for the regression coefficient means are $\hat{\mu}_\beta = (0.676, 0.572, 0.110)^T$, which have the same signs as the OLS estimates. The RMSE for the response variable is 0.04 with the SVCP model. The estimates for the response variable, $\hat{y}$, are plotted against y for the GWR and SVCP models in Fig. 2. The SVCP model fits the response variable considerably better than GWR, and has a RMSE approximately one order of magnitude below that of GWR. Both the SVCP and GWR models improve substantially on the OLS regression model.

Histograms of the estimated regression coefficients from the GWR and SVCP models are plotted in Fig. 3. The GWR estimated regression coefficients show more variation than do the SVCP model coefficients, especially for the intercept and the

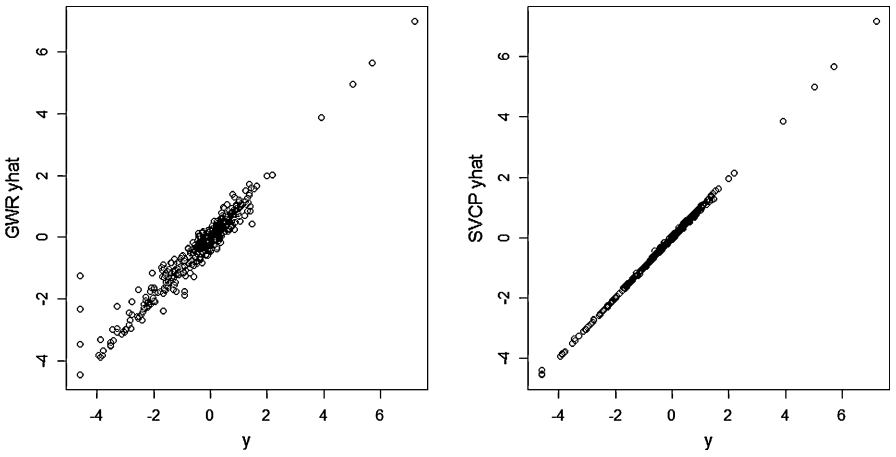

Fig. 2 The estimates of the response variable versus the response variable for the model with a geographically weighted regression (GWR) spatially varying coefficient process (SVCP) models

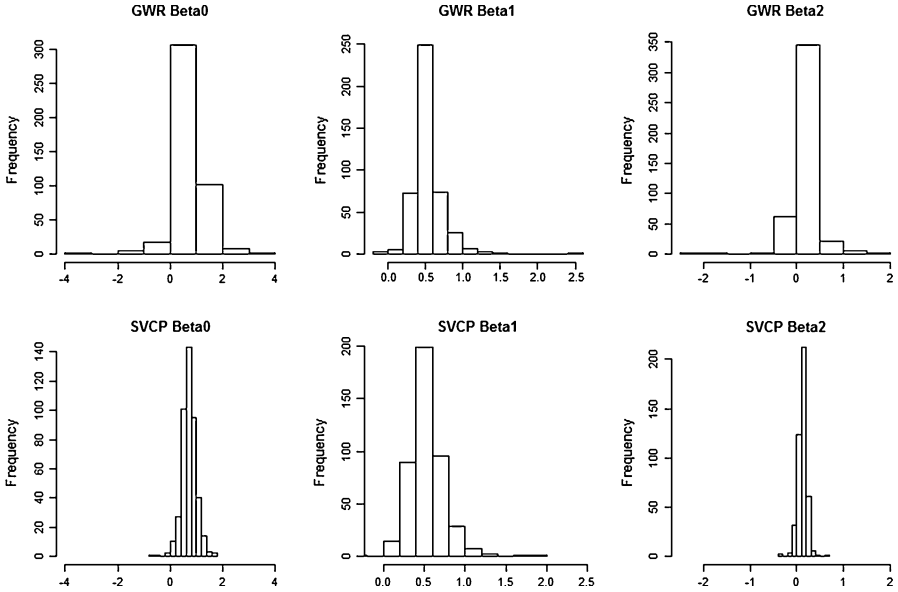

Fig. 3 Geographically weighted regression and SVCP estimated regression coefficients by census tracts for the intercept (Beta0), drug violations (Beta1), and alcohol outlets (Beta2)

alcohol sales effect, as evident from the ranges in the histograms. There are several outliers in the GWR coefficients for alcohol sales outlets, especially in the counterintuitive negative range. This suggests the SVCP model is better constraining the variation of the coefficients. It is not unexpected that the SVCP model produces a more compact set of regression coefficients than GWR, given the earlier discussion of the SVCP model as a shrinkage model in the vein of ridge regression. The extra variation in the GWR coefficients is also evident in scatter plots of $\hat{\beta}_2$ versus $\hat{\beta}_1$ (Fig. 4) for the GWR and SVCP models. Scatter plots like these have been suggested by Wheeler and Tiefelsdorf (2005) to check on the level of dependence in estimated GWR coefficients, as well as on the amount of variation in and the sign of coefficients. Indeed, it appears there is substantial dependence in the regression coefficients for these terms, particularly driven by some outliers in the GWR model. The overall correlations for the GWR coefficients are $r_{01} = 0.23$, $r_{02} = 0.59$, $r_{12} = -0.41$. The within area correlations between regression coefficients for the SVCP are $T_{01}/\sqrt{T_{00}T_{11}} = -0.35$, $T_{02}/\sqrt{T_{00}T_{22}} = 0.79$, and $T_{12}/\sqrt{T_{11}T_{22}} = -0.46$.

The estimated regression coefficients are mapped in Fig. 5 for GWR and in Fig. 6 for the SVCP model using quintiles to examine whether the quintiles of coefficients align in the two different models. Note that maps with the same breakpoints for both the GWR and SVCP coefficients would have more gray shading for the SVCP coefficients due to the larger variance of the GWR coefficients. There are clear, smooth patterns in the GWR estimated coefficients, where drug arrest effects are higher in the north and one central area in the city and alcohol outlet effects are lower in the north, overall, and higher in the central and eastern portions of the city. It is clear from the figures that the SVCP regression

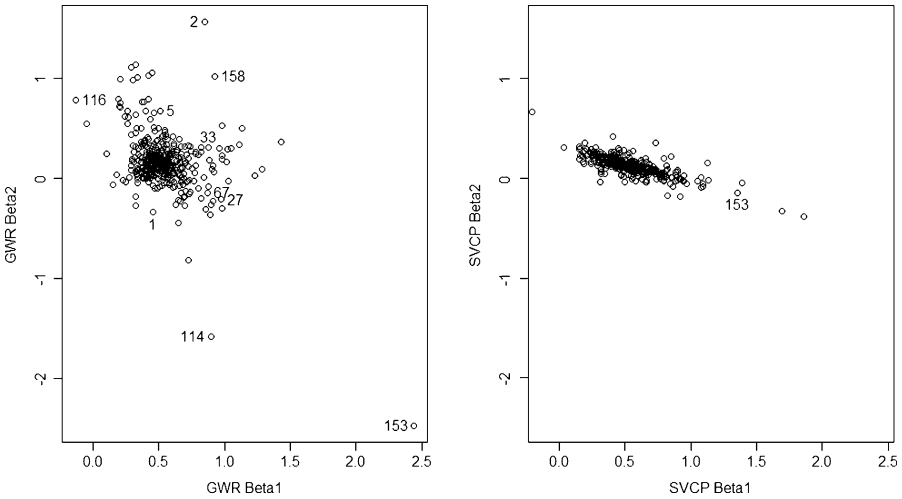

Fig. 4 Geographically weighted regression and SVCP estimated coefficients for alcohol outlets (Beta2) versus drug violations (Beta1)

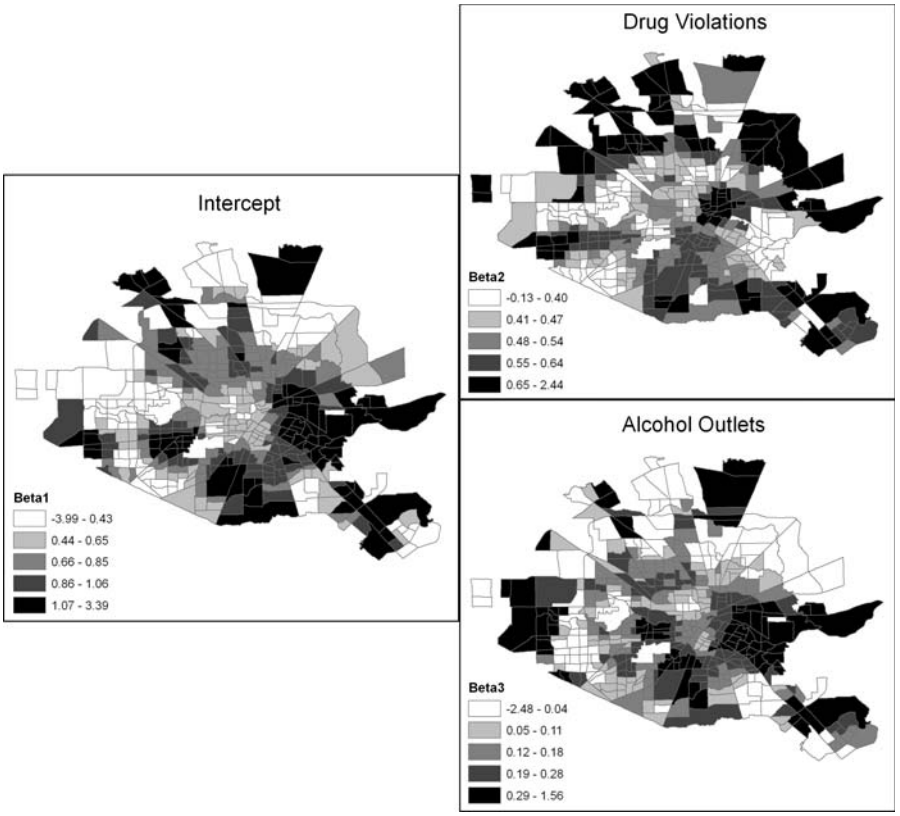

Fig. 5 Quintiles of GWR estimated coefficients for the intercept, drug violations, and alcohol outlets

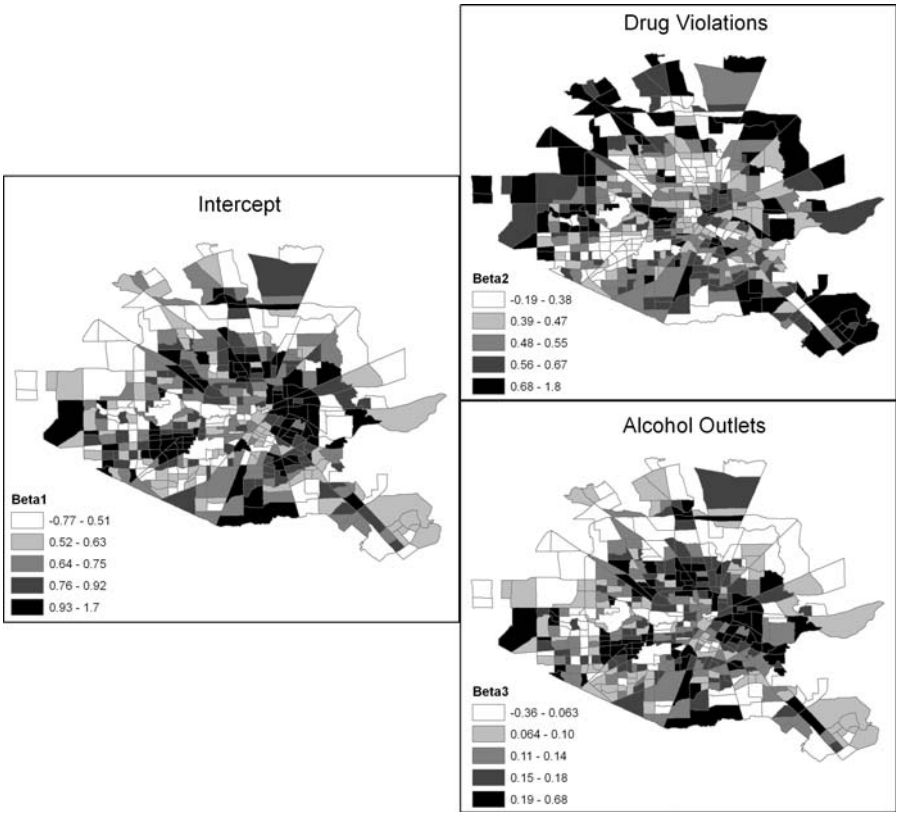

Fig. 6 Quintiles of Bayesian SVCP estimated coefficients for the intercept, drug violations, and alcohol outlets

coefficients are less smooth than the GWR coefficients. As a result, patterns in the estimated coefficients are not quite as clear. Drug arrest effects appear to be higher in the north of the city and in the southeast. Alcohol sale effects seem to be higher near the central city. The finding of GWR estimated regression coefficients that are spatially smoother than the SVCP estimated regression coefficients is in agreement with results from the analysis of the Houston crime data by Waller et al. (2007), where geographically weighted Poisson regression coefficients were smoother than those from a Bayesian spatially varying coefficient Poisson regression model that used a conditionally autoregressive spatial dependent specification for the regression coefficients. It is also clearly apparent from Fig. 6 that the SVCP model is not shrinking the coefficients to be the same for all areas, despite the shared prior means, i.e. spatial variation is maintained and the posterior estimates are indeed spatially varying. Also noticeable from the maps is some level of complementarity in the estimated coefficients for alcohol outlets and drug violations, where in some peripheral areas the alcohol coefficient is high and the drug violations coefficient is low. This pattern is stronger in the GWR coefficients than in the SVCP model

coefficients. This suggests that the complementary pattern in spatially varying coefficients may be controlled less by negative correlation between the coefficients and more by inflated variances of the coefficients. We will explore this in the next section.

To further investigate dependence in the GWR coefficients, we apply the variance–decomposition proportion and condition index diagnostic tool for collinearity introduced by Wheeler (2007). The variance–decomposition is done through singular value decomposition and it has an associated condition index, which is the ratio of the largest singular value to the smallest singular value. In diagnosing collinearity, the larger the condition index, the stronger is the collinearity among the columns of the GWR weighted design matrix. Belsley (1991) recommends a conservative value of 30 for a condition index that indicates collinearity, but suggests the threshold value could be as low as 10 when there are large variance proportions for the same component. The variance–decomposition proportion is the proportion of the variance of a regression coefficient that is affiliated with one component of its decomposition. The presence of two or more variance proportions greater than 0.5 in one component of the variance–decomposition indicates that collinearity exists between at least two regression terms, one of which may be the intercept.

The variance–decomposition proportions and condition indexes are listed in Table 1 for records with the largest condition indexes for the largest variance component. The GWR estimated bandwidth is used in the variance–decomposition of the kernel weighted design matrix to assess the collinearity in the model. These ten records are labeled in the plot of estimated GWR coefficients for the drug and alcohol covariates in Fig. 4. These labeled records comprise many of the more extreme points in the plot. Observation 153 is clearly the most extreme of the points, as it has the largest value for the drug rate effect and the smallest value for the alcohol rate effect. In Table 1, this record has large variance proportions for the same component for all three regression terms. Of the 439 records in the dataset, 5 have a condition index above 30, 10 have a condition index above 20, and 41 have a condition index above 10. There are 411 records in the data with large variance proportions (>0.5) from the same component, with the shared component being

Table 1 Record number, condition index (k), and variance–decomposition proportions (p_1 = intercept, p_2 = drug, p_3 = alcohol) for the GWR model of Houston crime

ID	k	p_1	p_2	p_3
1	27.60	0.996	0.995	0.136
2	87.66	0.992	0.992	0.001
5	21.29	0.995	0.993	0.188
27	24.25	0.997	0.690	0.947
33	35.45	0.865	0.949	0.045
67	29.49	0.994	0.982	0.371
114	40.58	0.739	0.988	0.283
116	39.45	0.579	0.996	0.922
153	38.38	0.737	0.999	0.609
158	21.94	0.955	0.942	0.006

between a covariate and the intercept for some records and between the two covariates for other records. Overall, the variance–decomposition proportions and condition index values indicate the presence of local collinearity in the GWR model. Incidentally, only one of the more extreme GWR observations in Table 1 is clearly visible in the plot for the SVCP coefficients in Fig. 4, observation 153. The other nine observations listed in the table have been shrunken towards their mean values in the SVCP model. Again, this is expected in light of the properties of the Bayesian SVCP model described previously and is clearly advantageous in constraining estimates for problematic observations.

4.2 Prior influence and SVCP model coefficient variance

In this section, we explore the sensitivity of the variation in estimated SVCP model coefficients to the specification of the prior distribution for the coefficient covariance matrix. In the Bayesian framework, it is possible to control the amount of variation and correlation in the estimated SVCP regression coefficients by selecting a prior for the within area covariance matrix that favors no correlation between the drug violations and alcohol terms. In practice, this can be implemented by specifying a relatively large scale matrix with off-diagonal elements of 0 in the inverse Wishart prior distribution for $\mathbf{T}$ and increasing the degrees of freedom to be greater than $p + 1$, where p is the number of linear regression terms. By increasing the prior variances, the posterior variance should be increased and the posterior covariance can be decreased between regression terms. For example, changing the prior for $\mathbf{T}$ from the vague IW($p + 1$, $0.1\mathbf{I}$) to the more informative IW($p + 2$, $10\mathbf{I}$) should have an increasing effect on the posterior mean variances in $\mathbf{T}$ and therefore should result in more variation in the regression coefficients $\hat{\boldsymbol{\beta}}$.

To illustrate the effect of changing the prior, we perform sampling from the joint posterior distribution using MCMC as before, but this time with a prior of IW($p + 2$, $10\mathbf{I}$) for $\mathbf{T}$. The resulting estimated regression coefficients are mapped in Fig. 7. The informative prior results in smoother regression coefficients than those in Fig. 6. The new coefficients are more similar to those from GWR in Fig. 5. Again, the Bayesian SVCP model allows for obvious spatial variation in the estimated regression coefficients. Increasing the prior variances of the regression terms allows their coefficients to vary more from their respective means. The estimated SVCP regression coefficients for alcohol outlets versus drug violations are displayed in Fig. 8 for the vague prior on the left and the informative prior on the right. The plots show that there is less overall correlation in the SVCP coefficients when the informative prior is utilized. In fact, the correlation between drug violations and alcohol outlet effects decreases in magnitude to $T_{12}/\sqrt{T_{11}T_{22}} = -0.26$. However, the RMSE of $\hat{y}$ increases to 0.061 with the informative prior. The results of changing the prior to a more informative one with more within-area variance and less covariance shows that one can control the amount of correlation in estimated regression coefficients in the SVCP model, with the potential cost of an increase in response variable estimation error. In addition, it is possible to get more smoothed coefficients, similar to those from GWR, by increasing the within area variance prior. In this case, the GWR coefficients have

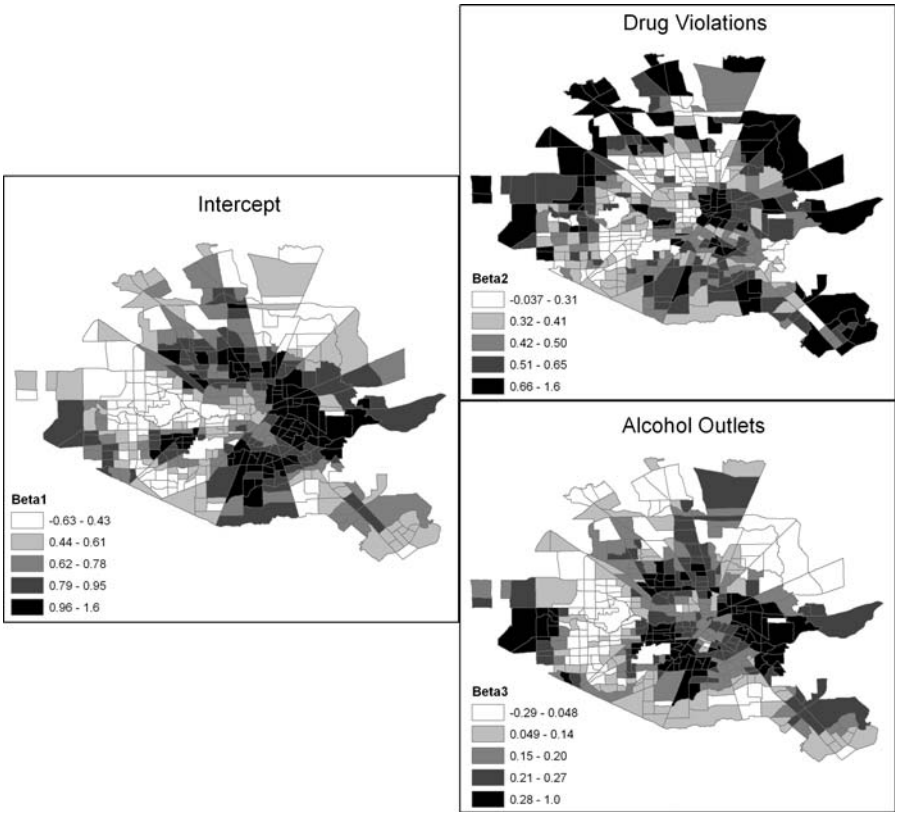

Fig. 7 Quintiles of Bayesian SVCP estimated coefficients for the intercept, drug violations, and alcohol outlets with an informative prior for **T**

some similarity to the Bayesian SCVP coefficients, but with inflated variance and less formal associated statistical inference. This view is not out of step with results of a simulation study in Wheeler and Calder (2007) that showed that estimated regression coefficients from a Bayesian SVCP model were consistently more accurate than from GWR.

4.3 Model prediction

Also of interest in spatial regression models and violent crime analysis is the prediction of the response variable for a new observation, for example the crime rate at a new census tract or for a tract for which violent crime data are missing. Both GWR and the SVCP can predict values at new observation locations in space. To compare the prediction performance of the GWR and SVCP models, we fitted the two models again using a 90% subset of the data and then predicted the response variable for the remaining hold-out random sample of approximately 10% of the original data observations using the model parameter estimates from the larger data

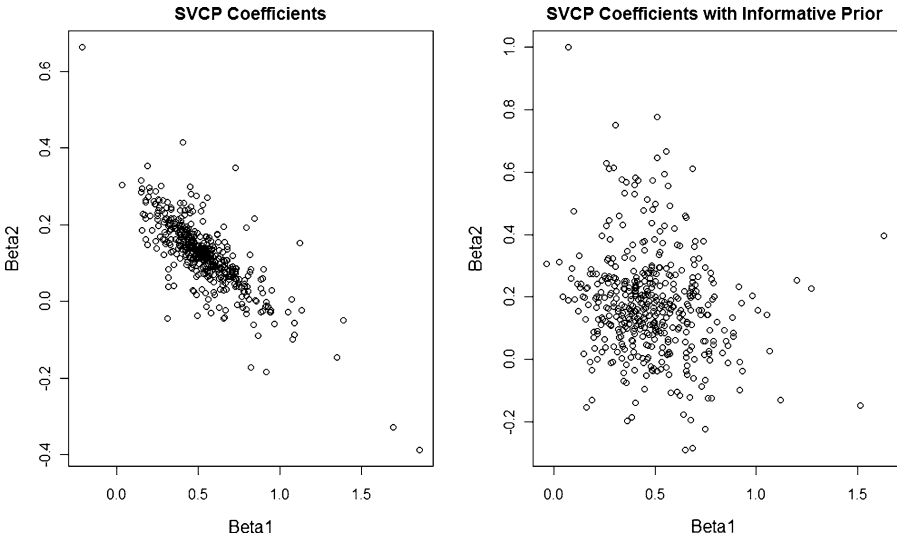

Fig. 8 Spatially varying coefficient process estimated regression coefficients for drug violations and alcohol outlets with a vague prior for **T** (*left*) and an informative prior (*right*)

sample. The sampling is done without replacement, so the two samples are mutually exclusive. There are 44 observations (census tracts) in the hold-out sample and 395 in the model calibration sample. The estimated kernel bandwidth $\hat{\gamma}$ for GWR using the calibration sample was 1.03. In the SVCP model, the regression coefficient mean estimates for the hold-out sample were drawn from the posterior prediction distribution for $\tilde{\boldsymbol{\beta}}$.

The RMSPE for $\tilde{y}$ for the GWR model is 0.659 and the RMSPE for the SVCP model is 0.655, so there is less than a 1% difference in the model prediction errors. The correlation in the predicted and actual responses is 0.85 for GWR and for the SVCP model. Clearly, the performance of the two models is similar in predicting the hold-out sample response values.

While the response variable predictions are similar in this case for the GWR and SVCP models, there is an important advantage to the SVCP model for prediction, namely the ability to produce prediction intervals. For each prediction, a prediction interval of a desired confidence level can be generated using draws from the parameter posterior distributions. See, for example, Fig. 9, which has 95% prediction intervals drawn for the SVCP model point predictions (plusses) for the response variable along with the GWR predictions (triangles) and the actual values (circles). Note that 41 of the 44 (93%) SVCP prediction intervals contain the observed violent crime rate. The GWR methodology does not directly produce prediction intervals, and it is not clear how one would generate them. The fact that one can create prediction intervals readily from the existing MCMC samples is a major advantage for the SVCP model. In fact, the quantification of such uncertainty is an advantage of the Bayesian paradigm in general, and is one of the primary objectives in the field of statistics.

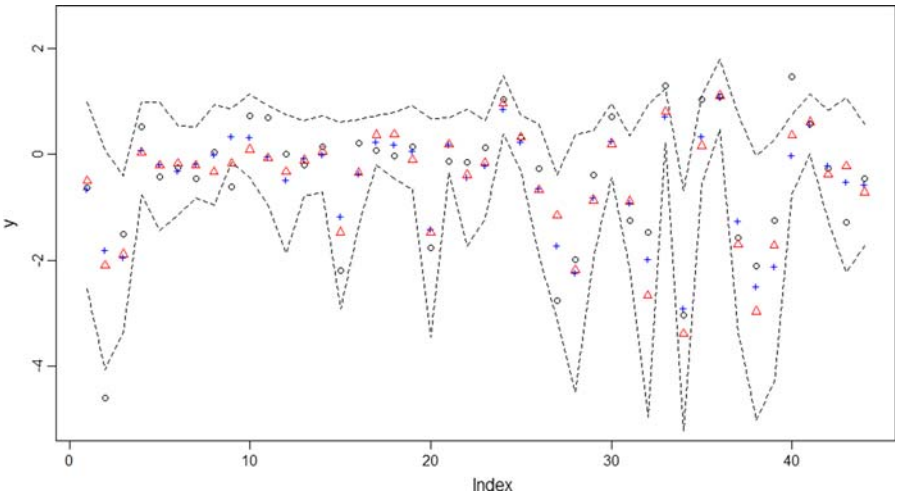

Fig. 9 Geographically weighted regression (*triangles*) and SVCP (*plusses*) predictions for the response variable (*circles*) of a hold-out sample with 95% prediction intervals from the SVCP model

It should be noted that the criterion used in model parameter estimation is not the same for the GWR and SVCP models. GWR uses leave-one-out cross-validation to minimize the prediction error of the data set where each observation is predicted by all other observations. The criterion for kernel bandwidth selection we use is explicitly a function of response variable prediction. GWR then uses the optimal bandwidth to define the kernel values for the location-specific spatial weights matrix, which in turn defines the parameter estimates. For the Bayesian SVCP model, all inference follows from the joint posterior distribution of model parameters. We sample directly from the joint posterior distribution of the parameters, where values are drawn more frequently that are more likely given the data. More likely parameter values effectively maximize the likelihood of the observed data (assuming an uninformative prior) without explicitly considering prediction of new data. Therefore, it would not be surprising if GWR with cross-validation estimation of the kernel bandwidth would perform better than the SVCP model for prediction. It would be interesting in a future study to compare the two models when using an Akaike information criterion (AIC), as discussed in Fotheringham et al. (2002), in place of cross-validation in GWR, given that the AIC is not directly based on prediction of the response variable. Also, only one random hold-out sample out of a large number of possible samples was used in this analysis. A more comprehensive comparison would include a larger number of samples.

5 Conclusions

In this paper, we have compared two different approaches, geographically weighted regression and a Bayesian SVCP model, for estimating potentially spatially varying regression coefficients for alcohol sales outlets and illegal drug violations to explain

variation in violent crime rates in Houston, Texas. GWR is a descriptive approach using spatial weights to emphasize spatially proximate observations in parameter estimation, while Bayesian SVCP models treat regression effects as spatially correlated processes within a probability model to enable statistical inference on the regression associations. In this analysis, we found spatially varying estimated regression effects for the alcohol and drug covariates with both approaches, although the GWR estimates were generally more variable. We also explored the issue of model collinearity and estimated regression coefficient correlation, as there was substantial correlation between regression coefficients in both the GWR and Bayesian SVCP models. These correlations between sets of regression coefficients are likely due to multiple causes including, but not confined to, local collinearity between the covariates of alcohols outlets and illicit drug use considered in our model for violent crime rates. As we demonstrated, however, it is possible to directly model and control the amount of variation in the estimated regression coefficients in the Bayesian SVCP model, which is not true for GWR. Case in point, when using a less informative prior distribution on the regression coefficient variance matrix, the Bayesian SVCP estimated coefficients exhibit less variation and are less spatially smooth than those from GWR. We showed that this is a result of the shrinkage properties inherent in the Bayesian SVCP model, which constrain the estimated regression coefficients and limit their associated variances. When using a more informative prior with more influence on larger variances, the SVCP estimated coefficients were more variable and exhibited a pattern more similar to those from GWR. The ability to directly control the amount of variation in regression coefficients where collinearity diagnostics indicate a problem with GWR estimates, along with readily available credible intervals for coefficients and prediction intervals for the response variable, provides evidence of several advantages of the Bayesian SVCP model over GWR. There is, however, a substantial computational cost for these modeling advantages and current methodological and computational capabilities need to be enhanced to increase general applicability of the Bayesian model.

An open research question we hope to address in the future is how much shrinkage is appropriate in spatially varying coefficient models. Certain diagnostic tools, including those for collinearity, should prove helpful to answer this question. Also, in the substantive problem studied here, it may be beneficial to include additional neighborhood contextual variables in the model to gain further insight into associations of interest.

References

Banerjee S, Carlin BP, Gelfand AE (2004) Hierarchical modeling and analysis for spatial data. Chapman & Hall/CRC, Boca Raton

Belsley DA (1991) Conditioning diagnostics: collinearity and weak data in regression. Wiley, New York

Brantingham PL, Brantingham PJ (1993) Nodes, paths and edges: considerations on the complexity of crime and the physical environment. J Environ Psychol 13:3–28. doi:10.1016/S0272-4944(05)80212-9

Brantingham PL, Brantingham PJ (1999) A theoretical model of crime hot spot generation. Stud Crime Crime Prev 8:7–26

Brunsdon C, Aitkin M, Fotheringham S, Charlton M (1999) A comparison of random coefficient modelling and geographically weighted regression for spatially non-stationary regression problems. Geogr Environ Model 3:47–62

Fotheringham AS, Brunsdon C, Charlton M (2002) Geographically weighted regression: the analysis of spatially varying relationships. Wiley, West Sussex

Gelfand AE, Kim H, Sirmans CF, Banerjee S (2003) Spatial modeling with spatially varying coefficient processes. J Am Stat Assoc 98:387–396. doi:10.1198/016214503000170

Goldstein M (1976) Bayesian analysis of regression problems. Biometrika 63(1):51–58. doi:10.1093/biomet/63.1.51

Gorman D, Speer PW, Gruenewald PJ, Labouvie EW (2001) Spatial dynamics of alcohol availability, neighborhood structure and violent crime. J Stud Alcohol 62:628–636

Griffith D (2008) Spatial filtering-based contributions to a critique of geographically weighted regression (GWR). Environ Plant A (in press)

Group 1 Software (2003) Centrus Desktop, Version 4.01.01. Boulder, CO

Gruenewald PJ, Freisthler B, Remer L, LaScala EA, Treno A (2006) Ecological models of alcohol outlets and violent assaults: crime potentials and geospatial analysis. Addiction 101:666–677. doi:10.1111/j.1360-0443.2006.01405.x

Hastie T, Tibshirani R, Friedman J (2001) The elements of statistical learning: data mining, inference and prediction. Springer, New York

Hoerl AE, Kennard RW (1970) Ridge regression: biased estimation for non-orthogonal problems. Technometrics 12:55–67. doi:10.2307/1267351

Lindley DV, Smith AFM (1972) Bayes estimates for the linear model. J Roy Statist Soc Ser B Methodol 34(1):1–41

Lipton R, Gruenewald PJ (2002) The spatial dynamics of violence and alcohol outlets. J Stud Alcohol 63:187–195

Nelson AL, Bromley RDF, Thomas CJ (2001) Identifying micro and temporal patterns of violent crime and disorder in the British city center. Appl Geogr 21:249–274. doi:10.1016/S0143-6228(01)00008-X

Roncek DW, Maier PA (1991) Bars, blocks, and crimes revisited: linking the theory of routine activities to the empiricism of "hot spots". Criminology 29:725–753. doi:10.1111/j.1745-9125.1991.tb01086.x

Sampson RJ, Lauritsen JL (1994) Violent victimization and offending: individual-, situational- and community-level risk factors. In: Reiss AJ Jr, Roth J (eds) Understanding and preventing violence: social influences, volume 3. National Academy Press, Washington, DC, pp 1–114

Scribner RA, MacKinnon DP, Dwyer JH (1995) The risk of assaultive violence and alcohol availability in Los Angeles County. Am J Public Health 85:335–340

Waller L, Zhu L, Gotway C, Gorman D, Gruenewald P (2007) Quantifying geographic variations in associations between alcohol distribution and violence: A comparison of geographically weighted regression and spatially varying coefficient models. Stochastic Environ Res Risk Assess 21(5):573–588. doi:10.1007/s00477-007-0139-9

Wheeler D (2007) Diagnostic tools and a remedial method for collinearity in geographically weighted regression. Environ Plant A 39(10):2464–2481. doi:10.1068/a38325

Wheeler D, Calder C (2007) An assessment of coefficient accuracy in linear regression models with spatially varying coefficients. J Geogr Syst 9:145–166. doi:10.1007/s10109-006-0040-y

Wheeler D, Tiefelsdorf M (2005) Multicollinearity and correlation among local regression coefficients in geographically weighted regression. J Geogr Syst 7:161–187. doi:10.1007/s10109-005-0155-6