Housing Sub-markets and Hedonic Price Analysis: A Bayesian Approach

by

David C. Wheeler1*, Antonio Páez2*, Lance A. Waller1 and Jamie Spinney3

Technical Report 07-10
November 2007

Department of Biostatistics
Rollins School of Public Health
Emory University
Atlanta, Georgia

1Department of Biostatistics
Emory University

2School of Geography and Earth Sciences
McMaster University

3Time Use Research Program
St. Mary’s University

*Correspondence Authors: Dr. David Wheeler and Dr. Antonio Páez
Telephone: (404) 727-8059     FAX: (404) 727-1370
e-mail: dcwheel@sph.emory.edu or paezha@mcmaster.ca
Housing sub-markets and hedonic price analysis: A Bayesian approach

Paper prepared for presentation at the 54th North American Meetings of the RSAI

Savannah, Georgia, November 2007

David C. Wheeler1*, Antonio Páez2*, Lance Waller1, Jamie Spinney3

1 Department of Biostatistics, Emory University
2 School of Geography and Earth Sciences, McMaster University
3 Time Use Research Program, St. Mary’s University

* Corresponding authors: dcwheel@sph.emory.edu, paezha@mcmaster.ca

Abstract. Sub-market definition has been identified as an important step in hedonic price analysis, to improve price predictions as well as estimates of marginal prices. Early market segmentation procedures were based on various definitions of housing or neighbourhood type. With the advent of new technologies (e.g. GIS), and statistical techniques (e.g. spatial models, clustering techniques), more sophisticated forms of analysis have become possible. The objective of this paper is to contribute to the literature on market segmentation by investigating the use of Bayesian models with spatially varying coefficients to delineate sub-markets based on geographical and functional considerations. The case study is the housing market in the city of Toronto in Canada. A comparison to a geographically weighted regression (GWR) model is made in terms of error in estimating and predicting hedonic prices, as well as patterns of sub-markets indicated by the spatially varying coefficient processes. Our results provide an interesting contrast to previous research that used moving window approaches to define so-called soft market segments. In addition, we describe how to introduce qualitative (i.e. expert) information obtained from the relevant property assessment corporation into hedonic price forecasting within the Bayesian framework.

Keywords. Housing prices, hedonic analysis, spatial models, Bayesian statistics, GWR
1 Introduction

Hedonic house pricing studies, a field where regression analysis has long been considered the industry standard, has experienced a remarkable evolution in the past decade thanks to technological developments (e.g. the emerging technology of geographic information systems), as well as progress in the field of spatial analysis (e.g. statistics and econometrics for spatial data). In the past few years, a number of pioneering studies have contributed to move hedonic price analysis into a promising new area, namely making more effective use of the spatial information contained in many house price datasets (e.g. Can, 1992a; Can, 1992b; Dubin, 1992; Basu and Thibodeau, 1998). Following these early studies, a specialized literature has thrived that pays special attention to two technical issues: the lack of spatial dependency often observed in spatial datasets, and patterns of heterogeneity that could be defined as geographical (e.g. Chica-Olmo, 1995; Páez et al., 2001; Tse, 2002; Farber and Yeates, 2006; Bitter et al., 2006). These issues are of interest, not only due to statistical reasons, but also because they can be associated with two substantive effects, information spillovers in housing markets (technically a form of spatial autocorrelation), and the existence of patterning of relationships that could result be identified as market segments (heterogeneity).

While autocorrelation analysis has been widely recognized as a technical issue, heterogeneity also has received considerable attention, not only within the specialized spatial hedonics literature but in the house pricing literature more in general, due to its relationships to housing submarkets definitions. Defining sub-markets has indeed been identified as an important step in hedonic price analysis, and a useful way to improve price predictions as well as estimates of marginal prices. The practice of defining sub-markets for price assessment is based on the principles of landscape compartmentalization and substitutability which combine to produce relatively homogeneous assessment units. A primary purpose of segmentation is to create sub-markets from the sample data, which are used to improve model estimates of value within each of these segments. Two broad forms of segmentation can be identified in the literature. Early market segmentation procedures were based on various definitions of housing or neighbourhood type. These procedures were used to define mutually exclusive and collectively exhaustive sub-markets, based on either structural or locational characteristics of the properties under analysis (e.g. Goodman and Thibodeau, 1998; Goodman and Thibodeau, 2003; Thibodeau, 2003). More recently, other approaches have been proposed that produce “soft” of “fuzzy” market segments. The use of moving window statistics, for example, implement “sliding neighborhoods” (see Páez et al, 2007), while fuzzy clustering techniques allow the definition of multiple membership in various markets (e.g. Hwang and Thill, 2005). An advantage of these approaches is that they avoid the necessity of defining rigid boundaries for the different sub-markets.

The objective of this paper is to contribute to the literature on market segmentation for hedonic price analysis. This research examines two approaches to geographical segmentation of the real estate market. More specifically, in this paper we investigate the use of Bayesian models with spatially varying coefficients to delineate sub-markets based on geographical and functional considerations, and the use of Geographically Weighted Regression for moving window analysis. The case study is the housing market in the city of Toronto in Canada. Comparison between the two modeling
techniques is made in terms of error in estimating and predicting hedonic prices, as well as patterns of sub-markets indicated by the spatially varying coefficient processes. Our results provide an interesting contrast to previous research that used moving window approaches to define soft market segments. In addition, we describe how to introduce qualitative (i.e., expert) information obtained from the relevant property assessment corporation into hedonic price forecasting within the Bayesian framework.

The rest of the paper is structured as follows. First we provide a brief review of the relevant literature. Section 3 describes the data and the two models used in the analysis. Section 4 presents the analysis and discusses the results, and finally section 5 provides some concluding remarks and directions for further research.

2 Housing sub-markets: Literature review

It has long been recognized that urban economists have tended to be pragmatic in their definition of functional markets (Ratcliff, 1949). Normally the market is the whole of a region in which buyers and sellers interact, which is often the same as city or regional boundaries. However, stratification of a given market is based on the idea that different goods will have different markets, whereby consumer preferences and prices are largely diversified (Bourassa et al., 1997). The idea of diversified preferences and prices led to the suggestion that the market for housing should be segmented into a series of submarkets as early as the 1950’s and 1960’s, (see Rapkin et al., 1953; Grigsby, 1963). The concept of the housing ‘submarket’ came to be seen as based on the appraisal concept of substitution, whereby the dwellings should be equally desirable substitutes, and not just those located in the same neighborhood (Grigsby, 1963; Rothenberg et al, 1991; Jones et al., 2001). However, other research suggests that while dwellings should be equally desirable substitutes, location is the main reason explaining why housing submarkets matter (Bourassa et al, 2003). Recent evidence tends to support this claim (Kestens et al., 2006). Over time, several techniques have been proposed and used to spatially delineate relatively homogeneous discrete housing submarkets.

Thibodeau (2003) suggests that a “housing market defines a geographic area where the price of housing per unit of housing service is constant”. However, the marginal hedonic prices of housing attributes tend to vary over space, which means that spatial structure is an important consideration for the estimation and specification of hedonic price models (Straszheim, 1975; Can, 1990; Can and Megbolugbe, 1997; Goodman and Thibodeau, 1998; Haurin and Brasington, 1996). In fact, location is the main reason explaining why housing submarkets matter (Bourassa et al, 2003), and may be considered at a variety of spatial scales. For example, Miron (1995) found that hedonic prices vary considerably from city to city, while Noland (1979) found that three-quarters of the variables are insignificantly different from each other but the remaining variables exhibit large differences. At the intraurban or neighbourhood scale, Straszheim (1975), Dale-Johnson (1982), Gabriel (1984), Grigsby et al. (1985), Rothenberg et al. (1991), Maclennan (1992), Feitelson (1993), Vandell (1995), Maclennan and Tu (1996), and Bourassa et al. (1999) found considerable difference in hedonic prices within a metropolitan region. However, these price variations have often been assumed away by housing economists (Jones et al., 2001).
Some commonly used procedures to determine sub-markets are briefly reviewed next.

### 2.1 Delineating Market Segmentation Boundaries

There are several techniques that have been used for the purpose of delineating market segmentation boundaries. These may be categorized into *a priori* methods, subjective/perceptual techniques, and various non-spatial statistical techniques. Each of these is briefly reviewed in the following sub-sections.

**A Priori Methods**

Residential submarkets are commonly defined by some prior view of what factors are important in terms of segmentation. Features that are typically used to define residential submarkets include socio-economic or environmental characteristics (Schnare, 1980; Harsman and Quigley, 1995; Vandell, 1995; Malpezzi, 2003), political or geographic boundaries (Schnare and Struyk, 1976; Goodman and Kawai, 1982; Adair et al., 1996), or accessibility characteristics (Vandell, 1995). In order to account for space-varying hedonic prices, the assessment jurisdiction is subdivided into discrete and mutually exhaustive relatively homogeneous geographic areas (submarkets) using one of the criteria mentioned. It is common then to estimate models separately for each submarket, to produce results that can be tested (by means of likelihood ratio or other tests) against the hypothesis of no-variation between market segments. Significant differences among segments imply the existence of heterogeneity and therefore contextual effects in the dimension of the segmentation (e.g. geographical space, income, accessibility).

While commonly used, a number of issues arise when following this approach to defining market areas. First, the most appropriate features to determine the structure of residential submarkets are not always available or sufficient to explain the demand for housing, and subjective variables such as consumers’ taste and preferences may also affect demand (Nelson and Rabianski, 1988; Bourassa et al., 1997; Thibodeau, 2003). Secondly, even when attributes are available, it is not always clear how the divide the sample, and moreover the classifications tend to be arbitrary. Thirdly, the estimation scheme of separate models for each sub-sample means that the procedure does not efficiently use the total information content, since it effectively isolates each sub-sample. And finally, the segmentation scheme implicitly assumes that the sample can be divided into discrete categories.

**Statistical Techniques**

Much of the recent research on housing market segmentation has employed statistical techniques for the purpose of defining submarkets. For example, Dale-Johnson (1982) used factor analysis, Abraham et al. (1994), Goetzmann and Wachter (1995), and Hoesli et al. (1997) used cluster analysis, Maclellan and Tu (1996) use principal component analysis (PCA) to identify variables that explain the highest proportion of the variation in the data, which are then used for cluster analysis. Bourassa et al. (Bourassa et al., 1997; Bourassa et al., 1999) also use PCA to extract the relevant factors, but then use the factor scores in cluster analyses to form submarkets. Cano-Guervós et al. (2003) integrate PCA and the theory of regionalized variables (TRV), based on housing
characteristics, in order to spatially delineate or “district” the city into a series of housing submarkets. These approaches share with the above a priori definitions the potential limitation that rigid boundaries need to be defined, however backing those boundaries with statistical evidence. Two recently proposed alternatives that do not involve such discontinuities include the use of moving windows and geographically weighted regression. Another is spatially varying coefficient process models (Gelfand et al., 2003). While moving windows offer a conceptually appealing alternative to the definition of “hard” market segments, and the approach has been shown to produce better estimates of prices, it can also be criticized for being essentially an ensemble of local models without borrowing power between subsamples. The use of fuzzy clustering techniques is also attractive conceptually (Hwang and Thill, 2005), but less straightforward to use in a regression framework for the purpose of price assessment.

Subjective/Perceptual Techniques

An alternative approach to spatially defining submarkets has been to define market areas as perceived by real estate agents (Palm, 1978; Michaels and Smith, 1990). An argument for this approach advocated by Palm (1978) is that a major weakness in attempts to define submarkets empirically is the failure to take into account the ways in which households acquire and use information about vacancies and the ways in which they match themselves with these vacancies (Bourassa et al., 1999). Nelson and Rabianski (1988, p. 138) applied this approach by using “consumer-supplied similarity measures of housing alternatives to develop a multidimensional perceptual map of the way consumers view housing choices”. Currently, it is unclear to what extent these perceptual techniques can match more objectively defined categorizations of space, and in fact research in other domains has shown that there can actually be significant variations between statistical (i.e. data-) driven approaches, and expert perceptions of space. For more information on perceptual space and measuring preferences see Menchik (1972). A potential advantage of adopting a Bayesian approach is the ability to incorporate subjective information as part of formal model development. In a subsequent section, we discuss this idea more in detail.

3 Data and Methods

3.1 Data

The analysis reported in this paper is based on data from 547 transactions of single-family detached houses sold in January 2001 in the City of Toronto. In order to reflect the true value of properties and avoid potential bias brought by distorted market forces like clearance sales, only open market sale records are included in the dataset.

The primary source of information comes from the Municipal Property Assessment Corporation (MPAC) in Ontario. Every year, MPAC prepares an assessment roll for every Ontario municipality which provides the assessed value of all the properties in a municipality or in the jurisdiction of a school board with taxing authority (www.mpac.ca). The computer file provided by MPAC contains each property’s address as well as information of each property’s structural characteristics. With ArcView, each transaction is assigned geographic coordinates and is geocoded into the study area.
We first analyze the entire dataset. Then, in order to test the performance of the models for prediction, we reserve a hold-out subsample. The full dataset for the analysis consists of 547 observations, which are divided into two groups: an estimation sample, i.e. in-sample observations; and a prediction sample, i.e. out-of-sample observations (see Case et al, 2004). The estimation sample contains 498 observations, which is 90% of the observations available, whereas the prediction sample is comprised of 49 observations, which constitutes 10% of the recorded transactions.

3.2 Geographically Weighted Regression (GWR)

In GWR models, the data are either mean measures of aggregate data at fixed points with spatial coordinates; for example, see the numerous examples in Fotheringham et al. (2002) or individual point-level data. 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

\[
y(s) = X(s)\beta(s) + \varepsilon(s), \tag{1}
\]

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

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

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

The weights matrix, \( W(s) \), 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. We use the exponential function for the kernel in this paper to match the spatial dependence function used in the SVCP model, although one could also use another type in the Bayesian 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 \( \gamma \) is the kernel bandwidth parameter.

To fit the GWR model, the kernel bandwidth is first estimated by cross-validation across all the calibration locations. Cross-validation is an iterative process that finds the
kernel bandwidth with the lowest prediction error of all the $\gamma(s)$. For each location $s$, it removes data for observation $s$ in the model calibration at location $s$ and predicts the response $\gamma(s)$ using the other data points and the kernel weights associated with the current bandwidth. Next, the weights are calculated at each calibration location using the kernel function. Finally, the regression coefficients are estimated at each calibration location along with the response estimates by the expression $\hat{\gamma}(s) = X(s)\hat{\beta}(s)$.

### 3.3 Bayesian Spatially Varying Coefficient Process (SVCP) Model

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 the specification of Gelfand et al. (2003), the SVCP model is

$$[Y | \beta, \tau^2] = N(X^T\beta, \tau^2I),$$  

where the bracket notation $[A | B]$ denotes the distribution of $A$ conditional on $B$. Note that equation (4) is identical to equation (1) under the assumption that the error terms represent an i.i.d. sample from a $N(0, \tau^2)$ distribution. $Y$ is a vector of responses assumed to be Gaussian conditional on the parameters $\beta$ and $\tau^2$; $\beta$ is a $np \times 1$ vector of regression coefficient parameters; $X^T$ is the $n \times np$ block diagonal matrix of covariates where each row contains a row from the $n \times p$ design matrix $X^*$, along with zeros in the appropriate places (the covariates from $X^*$ are shifted $p$ places in each subsequent row in $X^T$); $I$ is the $n \times n$ identity matrix; and $\tau^2$ is the error variance.

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

$$[\beta | \mu_\beta, \Sigma_\beta] = N(I_{n \times 1} \otimes \mu_\beta, \Sigma_\beta).$$  

The vector $\mu_\beta = (\mu_{\beta_0}, ..., \mu_{\beta_p})^T$ contains the means of the regression coefficients corresponding to each of the $p$ explanatory variables. The symbol $\otimes$ denotes the Kronecker product operator, which is the multiplication of every element in $I_{n \times 1}$ by $\mu_\beta$. The prior on the regression coefficients in the SVCP model takes into account the possible spatial dependence in the coefficients through the covariance, $\Sigma_\beta$. For $\beta_p = [\beta_{p_1}, ..., \beta_{p_n}]$, we can assume a priori that each $\beta_p$ follows an areal unit model (e.g., the conditional autoregressive model or simultaneous autoregressive model; see Banerjee et al, 2004) or specify the prior on $\beta_p$ using a geostatistical approach, where a parametric distance-based covariance function is specified. We 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 $\beta$’s at each of $n$ locations, $\Sigma_\beta$, 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), we assume a separable covariance matrix for $\mathbf{\beta}$ of the form:

$$
\Sigma_p = \mathbf{H}(\phi) \otimes \mathbf{T},
$$

(6)

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, and $\phi$ is an unknown spatial dependence parameter. In the prior specification for $\mathbf{\beta}$ (equation 6), 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 $\rho(d; \phi) = \exp(-d/\phi)$. The correlation between explanatory variable coefficients is explicitly modeled through the within-location covariance matrix $\mathbf{T}$, for example, the posterior correlation of the regression coefficients $k$ and $l$ across all locations is $T_{kk} / \sqrt{T_{ll} T_{ll}}$.

With the separable cross-covariance function, 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 explanatory variable in the GWR model, and justifies our use of the separable form. The separable cross-covariance form also has the property that the covariance between $\mathbf{\beta}$’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. A nonseparable form would allow for distinct spatial ranges for each coefficient process, and is therefore more flexible. 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 $\mathbf{\beta}$ (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 (4), (5), and (6) 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 $b$, $[\tau^2] \sim IG(a,b)$. Hyperparameters control the mean and variance of the prior distribution of the parameter. The prior for the coefficient means is normal with hyperparameters $\mu$ and $\sigma^2$, $[\mu_p] \sim N(\mu, \sigma^2 \mathbf{I})$. The prior for the covariance matrix $\mathbf{T}$ is inverse Wishart with hyperparameters $v$ and $\Omega$, $[\mathbf{T}] \sim IW_v(\Omega^{-1})$. These priors are conjugate priors, and are used for computational convenience. The prior for the spatial dependence parameter $\phi$ is gamma with hyperparameters $\alpha$ and $\lambda$, $[\phi] \sim G(\alpha, \lambda)$. 

8
Inference on the parameters in the SVCP model is based on the posterior distribution \( \theta \mid y \) of the parameters \( \theta = (\beta, \tau^2, \mu_p, \phi, T) \), which can be obtained using Bayes Theorem:

\[
[\theta \mid y] \propto [y \mid \theta][\theta].
\]  

In other words, the posterior distribution for the parameters \( \theta \), conditional on the data, is proportional to the likelihood of the data \( y \mid \theta \), also written as \( p(y \mid \theta) \), and the prior \( \theta \), also written as \( p(\theta) \), 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:

\[
\begin{align*}
\phi(j) & \sim [\phi \mid \beta(j-1), \mu_p(j-1), T(j-1), \tau^2(j-1), Y] \\
T(j) & \sim [T \mid \beta(j-1), \mu_p(j-1), \phi(j), \tau^2(j-1), Y] \\
\tau^2(j) & \sim [\tau^2 \mid \beta(j-1), \mu_p(j-1), T(j), \phi(j), Y] \\
\mu_p(j) & \sim [\mu_p \mid \beta(j-1), \phi(j), T(j), \tau^2(j), Y] \\
\beta(j) & \sim [\beta \mid \phi(j), \mu_p(j), T(j), \tau^2(j), Y].
\end{align*}
\]  

Clearly, to use the Gibbs sampler, one must be able to derive the full conditional distribution for each parameter. 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. Unlike the other parameters, the full conditional distribution of \( \phi \) 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 \( \phi \), 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, we use a normal random walk proposal density that is centered on the current value of \( \phi \) and has a variance \( s^2 \) that is tuned to produce an adequate acceptance rate of the proposed value of \( \phi \). The proposed value of \( \phi \) 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 \( \phi \) is greater than 1 or, if not, is greater than a randomly drawn uniform variable with a range of (0,1). See Wheeler and Calder (2007) for details related to sampling from the joint posterior for the SVCP model using MCMC.

4 Analysis and Discussion

4.1 Estimation of Hedonic Prices

The first model used for the Toronto hedonic pricing data is a simple linear model where the response variable is a function of two covariates and an intercept term. The GWR model is:

\[
y(i) = \beta_0(i) + \beta_1(i)x_1(i) + \beta_2(i)x_2(i) + \epsilon(i),
\]

where \( y \) is the natural log of the housing sale price, \( x_1 \) is the natural log of the effective site area in square feet of the property, \( x_2 \) is the Euclidean distance from the property to the nearest Light Rapid Transit (LRT), subway, or train station, and \( i \) is the index for house properties. The selection of this model was largely influenced by the analysis of this dataset due to Páez et al. (2007). The response variable and two covariates are mapped in Figure 1.

The model with non-varying regression coefficients, i.e. the ordinary least squares model, has a \( R^2 = 0.17 \) and a root mean square error (RMSE) for the estimated dependent variable of 0.455. The regression coefficient estimates are \( \hat{\beta}_0 = 4.339, \hat{\beta}_1 = 0.406, \hat{\beta}_2 = -0.086 \) and they are each significant at the 0.001 level.

Through cross-validation estimation, the estimated GWR kernel bandwidth \( \hat{\gamma} = 0.44 \). The root mean square prediction error (RMSPE) associated with this estimated bandwidth is 0.276. The RMSE of the response variable for the GWR model with this estimated bandwidth and regression coefficients estimated with the bandwidth is 0.175.

For the SVCP model, the prior specification for the model is as follows. We use a vague normal, \( N(0, 10^4 I) \), for \( \mu_p \), a somewhat vague inverse Wishart, \( IW(p+1, 0.1 I) \), for \( T \), and a vague inverse gamma, \( IG(1, 0.01) \), for \( \tau^2 \), where \( I \) is the identity matrix of dimension \( p \). For the spatial dependence parameter \( \phi \), we use a gamma, \( G(0.042, 0.01) \), which has a mean of 4.2 and variance of 420. 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 \( \hat{\phi} = 1.54 \) using the posterior mean. A sample size of 1,000 was used in the MCMC,
with a “burn-in” of 2,000 iterations. The posterior mean estimate of the error variance is \( \hat{\tau}^2 = 0.030 \) and the posterior mean estimates for the regression coefficient means are \( \hat{\mu}_\beta = (4.356, 0.318, 0.023)^T \). The posterior mean estimate of the within variance matrix for the regression coefficients is \( \hat{T} = \begin{bmatrix} 0.319 & -0.105 & 0.019 \\ 0.047 & -0.015 & 0.012 \end{bmatrix} \). The RMSE for the response variable is 0.117 with the SVCP model. The estimates for the response variable, \( \hat{y} \), are plotted against \( y \) for the GWR and SVCP models in Figure 2. The SVCP model fits the response variable considerably better than GWR, and has a RMSE that is approximately 33% lower than GWR. Both GWR and the SVCP model improve greatly on the estimation error of the dependent variable.

The distributions of each regression term in the GWR and SVCP models are illustrated in Figure 3 for the intercept, Figure 4 for area, and Figure 5 for distance to nearest public transportation. The histograms in these figures show that the GWR estimated coefficients exhibit more spread than the SVCP estimated regression coefficients. There is some skew in some of the regression coefficient distributions, most noticeably for the site area regression term for the SVCP model. The estimated regression coefficients for the GWR and SVCP models are mapped in Figure 6 for the intercept, Figure 7 for area, and Figure 8 for distance to nearest public transportation. The patterns in the estimated regression coefficients are somewhat similar for the GWR and SVCP models in portions of the study area, for example higher estimated intercepts in the northeast of the study area and higher values for the area covariate effect in the central study area, however, the GWR coefficients exhibit far more variation than the SVCP model coefficients. This is demonstrated in the lack of SVCP coefficients in the lowest and highest categories of the legend. The legends are the same for the GWR and SVCP coefficients and the map classification is based on modified natural breaks.

### 4.2 Prediction of Hedonic Prices

Also of interest in spatial regression models and hedonic price analysis is the prediction of the response variable for a new observation, for example the selling price of a new property somewhere in the study area. 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 an approximately 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 sample. The sampling is done without replacement, so the two samples are mutually exclusive. The prediction sample is the same for the GWR and SVCP models. There are 49 observations (housing sales) in the hold-out sample and 498 in the model calibration sample. The model calibration and hold-out observations are plotted in Figure 9.

The estimated kernel bandwidth for GWR using cross-validation is \( \hat{\gamma} = 0.503 \). The regression coefficient mean estimates for the hold-out sample distributions were drawn from the posterior prediction distribution for \( \hat{\beta} \) in the SVCP model. The RMSPE
for \( \tilde{y} \) for the GWR model is 0.267 and the RMSPE for the SVCP model is 0.246, with an approximately 8% difference in the model prediction errors. The model predictions are plotted in Figure 10, which conveys that the performance of the two models is similar in predicting the hold-out sample response values, but that the SVCP model performs better than GWR. The correlation in the predicted and actual responses is 0.82 for GWR and 0.85 for the SVCP model.

One key advantage in prediction of the response variable for the SVCP model is 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, Figure 11, 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 43 of the 49 (88%) prediction intervals contain the observed housing price. The GWR methodology does not directly produce prediction intervals, and it is not clear how one would generate them through resampling of the data. 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.

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 the 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 through equation (2). 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 unexpected if GWR with cross-validation estimation of the kernel bandwidth would perform better than the SVCP model for prediction, although in this case it does not.

It would be interesting 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. Therefore, we again fitted the GWR model with the calibration sample, but this time using the AIC instead of cross-validation to estimate the kernel function bandwidth and regression coefficients. Using AIC, the estimated kernel bandwidth \( \hat{\gamma} = 1.24 \), substantially larger than with cross-validation. The AIC with this kernel bandwidth is 387.43. The RMSE of the response variable with AIC is 0.284 and the RMSPE of the response variable for the hold-out sample is 0.309. The estimation and prediction errors for GWR are worse when using AIC in place of cross-validation. As suspected, the prediction error of the response variable increases when using a kernel bandwidth estimation criterion not explicitly based on prediction of the dependent variable.
One caveat of this analysis is that only one hold-out sample out of a large number of possible samples was used in this analysis. A more comprehensive comparison would use a larger number of samples.

4.3 Including Prior information in the Bayesian Model

Another advantage of the Bayesian framework is the ability to include prior information or expert views about a phenomenon under study into the model. This is typically done through an informative prior distribution for a model parameter, where the hyperparameters of the prior distribution are set to reflect characteristics of the parameter according to expert information. For example, an analyst could set the mean and variance of the prior distribution for a parameter to reflect the expected mean and range of values for the parameter, as elicited from an expert. Conceptually, such information is easier to elicit from a subject expert when the parameter is an observable quantity, such as the selling price of a house. Lele (2004) states that it is often natural for subject matter experts to provide guesses of the observable variable of interest, while considering the process that would generate the observable data. Lele (2004) also makes the point that it is easier to elicit information about observable data than the prior distribution of a statistical model parameter. To include expert opinion in a hierarchical, but not Bayesian, model, Lele (2004) combines expert guesses of the response variable with the actual observable data into one likelihood and estimates model parameters with maximum likelihood estimation.

In a Bayesian setting, Gelman et al. (2004) provide an example of constructing a prior distribution with hyperparameters for a parameter in an experiment using historical data from other experiments. These authors set the hyperparameters for the prior distribution of the parameter of interest using the mean and standard deviation of the distribution of the previously observed values. This estimation of the prior distribution using historical data effectively yields an informative prior for the parameter. This idea can be extended to the usage of expert elicited information instead of historical data, as long as the expert has not viewed the existing data prior to elicitation. There are several possible approaches to eliciting information from an expert. In a regression model setting, one could elicit beliefs about regression coefficients or about the dependent variable distribution, perhaps at numerous fixed levels of the covariates. Kadane et al. (1980) elicit expert views on quantiles of the predictive distribution of a dependent variable in a regression model conditional on fixed values of the covariates instead of eliciting beliefs about the regression coefficients directly given the perceived difficulty in thinking about what those values may be. In contrast, Chaloner et al. (1993) elicit distributional information on regression coefficients in a proportional hazards model using graphical displays of probability distributions. Garthwaite and Dickey (1988; 1991) also elicit distribution information on regression coefficients, as well as the variance, but in a linear regression model setting, using an interactive computer environment. Regardless of the approach taken to elicit some form of expert information, the incorporation of such information into the Bayesian modeling framework aligns with the philosophy of the scientific method, where knowledge that is available before collecting data (prior) is used along with the observed current data (likelihood) to inform what we know now (posterior).
The previous statistical approaches to including elicited information in regression models were for aspatial models, in that no consideration was given to local regression coefficients at different spatial locations. Different approaches to including expert opinion in a Bayesian spatial regression model are taken by Haining and Law (2007), who combine empirical violent crime event data with police perceptions of high crime areas to model the probability of an area being a high crime intensity area. Haining and Law (2007) first fit models for the police-defined (expert) high crime areas and then use the parameter posterior distributions as the prior distributions in a model for the empirically-defined high crime areas to update the expert beliefs with the observed data. Haining and Law (2007) find no differences in the posterior distributions when using the expert-informed priors or noninformative priors. This is likely due to the close correspondence between the expert-derived beliefs and the empirical data, where it is probably reasonable to assume the experts have had some level of exposure to the observed crime data. The second approach of Haining and Law (2007) is to combine the police-derived data with the observed data into one repeated measures dataset, where the somewhat questionable assumption that the police perception and observed crime are independent at each area. The first approach for including expert beliefs fits better into the Bayesian paradigm of data assimilation as described in Wikle and Berliner (2007) and that is the approach we choose to pursue in future work.

To consider what the approach for including expert opinion on housing prices would look like in the Bayesian SVCP model, we briefly outline the approach here. The approach of constructing an informative, parameterized prior distribution is outlined generally in Gelman et al. (2004, p.118) and mentioned in Haining and Law (2007). Suppose data are provided by real estate market assessors in the form of expert guesses on the housing sale price for all observations in the observed dataset, where the assessors are not shown the observed data in advance. Call the observed data vector $y_o$ and the elicited data vector $y_e$. Find the posterior distribution of the parameters $[\tau^2, T, \mu_p, \beta, \phi | y_e] \sim L(\tau^2, T, \mu_p, \beta, \phi; y_e) \times p(\tau^2) \times p(T) \times p(\mu_p) \times p(\beta) \times p(\phi)$ using MCMC with noninformative prior distributions. Then, use properties of the posterior distributions to set informative priors for the model for $y_o$. This is done by calculating the mean and variance of each parameter posterior distribution and then determining what the hyperparameters should be to have a prior distribution with the mean and variance from the MCMC for $y_e$. For example, the error variance follows an inverse gamma distribution with prior $p(\tau^2) \sim IG(a, b)$. In the SVCP model, the posterior distribution for this parameter is $[\tau^2 | \beta; y_e] \propto L \times p(\tau^2) = IG \left( a + \frac{n}{2}, b + \frac{1}{2} (y_e - X\beta)^T (y_e - X\beta) \right)$. Say the mean of $\tau^2$ from the MCMC is 2.2 and the variance is 1.1. The mean of an inverse gamma distributed parameter, $\tau^2 \sim IG(a,b)$, is $E(\tau^2) = \frac{b}{(a-1)}$ for $a > 1$ and the variance is $Var(\tau^2) = \frac{b^2}{(a-1)^2(a-2)}$ for $a > 2$ (Gelman et al. 2004). Setting $E(\tau^2) = 2.2$ and $Var(\tau^2) = 1.1$ and solving for $a$ and $b$ yields the prior $p(\tau^2)$ for the model for $y_o$. 

14
with the mean and variance of the posterior \( [\tau^2 | \beta; y_*] \). In this example, the prior is \( p(\tau^2) \sim IG(6.4, 11.88) \). Similar steps may be carried out for parameters \( (T, \mu_p, \phi) \). The prior for the regression coefficients \( \beta \) is specified entirely through other model parameters, so the other informative priors will influence the posterior for the regression coefficients in the model for \( y_* \).

As previously mentioned, it is possible in the Bayesian framework to incorporate expert opinion about model parameters directly through the parameter prior distributions by eliciting information about the parameter instead of the outcome variable. Obviously, eliciting this information will be easier when experts have a good sense for the value of a parameter though thinking about it in their everyday setting and language. We believe that is the case here when dealing with housing prices, as market assessors should have a good idea for how much the value of a home will change with an additional bathroom, bedroom, or 1,000 square feet of area. As an alternative to the approach described above, we outline here how to proceed in that direction for the SVCP model when expert opinion about the regression coefficients is available. Consider a set of expert guesses about the regression coefficients \( \beta \) at all observational units, where the expert guesses vary over observational units for each regression term (covariate or intercept). These values may be treated as prior data and incorporated into the SVCP model through the prior for the means of the regression coefficients, \( \mu_p \). This, however, requires a change in dimension of the mean vector \( \mu_p \) because in the SVCP model as previously specified, the regression terms are assumed to have the same mean at all locations. This is seen in the kronecker product of \( 1 \otimes \mu_p \) in the prior for \( \beta \). The new dimension of \( \mu_p \) with elicited information is \( np \times 1 \). The full conditional distributions for the SVCP model parameters when using expert information about the regression coefficients are listed in the Appendix. These distributions are used in sampling from the joint posterior distribution with MCMC. While including elicited prior information in the model may result in more precise regression coefficient estimates, it comes at a computational cost in the SVCP model, as the dimensionality of the coefficient mean vector increases from \( p \) to \( np \). This increase in size will slow down the MCMC sampling considerably due to slower matrix inversions.

While the approach of including prior information is outlined above for the regression coefficients, it may be applied generally to functions of parameters that are well-defined and known to a subject matter expert, for example odds-ratios in a public health analysis. Also, while the Toronto hedonic pricing data are individual-level data, the elicited prior information could be at a submarket, neighbourhood, or census geography level and could be transposed to the individual level when constructing the priors. Additionally, hierarchical models with some random effects that do not vary or are not spatially correlated could be used. A neighbourhood random effect could be used for the intercept with individual random effects that represent deviations from the neighbourhood mean. Finally, one could also investigate the use of a Bayesian cluster model (see Knorr-Held and Rasser, 2000) to divide the study area into submarkets for the response variable.
5 Conclusions

In this paper, we investigate the use of two models, a Bayesian model with spatially varying coefficients, and Geographically Weighted Regression. The objective of the paper was to compare these two models in terms of defining market segmentations for hedonic price analysis. The models operate in very different ways: whereas the Bayesian model defines clearly delineated zones for the market, GWR is based on sliding neighborhoods that produce “soft” market segments. The two models can be seen as different ways of conceptualizing and implementing sub-market definitions.

The main objective of this paper has been to compare these two approaches in terms of their ability to fit the data. Our results indicate that in terms of performance, the Bayesian approach produced better estimates and also predictions of house prices. An additional advantage of this approach is that it intervals of confidence can be easily retrieved from the models. A key issue with the use of Bayesian models is that, being based on simulation techniques, they impose a substantial computational burden. Contrast the use of a few hundred observations in the preliminary analysis reported in this paper, to the use of tens of thousands of observations in the analysis reported by Páez et al. (2007). The models used here are also relatively simple in terms of the covariates. This point suggests that an area for further research is to investigate whether gains in performance in the “base” model (basically the selection of covariates) can help to improve the two models used here, but more importantly, whether those gains are similar. Páez et al. for example report that performance gains in a base regression model are also reflected in the moving windows versions of the model. An underlying issue is the usual tension between performance and parsimony, requiring a tradeoff between approaches allowing computation using all available data, and approaches providing full specification of all unknowns, but at an appreciable cost in computing. As computing capabilities continue to advance, the methodological goal is to move toward linking the approaches by weakening assumptions made for computational convenience while maintaining analytic precision.

An important direction to extend the research in this paper is to introduce expert information as part of the estimation of Bayesian models, as outlined in section 4.3. This line of research is currently in progress, and we expect to be able to report on it in the near future.
Figure 1. Housing prices, site area, and distance to public transportation
Figure 2. GWR and SVCP model estimates for housing prices versus observed housing prices

Figure 3. GWR and SVCP estimated regression coefficients for the intercept
Figure 4. GWR and SVCP estimated regression coefficients for area

Figure 5. GWR and SVCP estimated regression coefficients for distance to public transportation
Figure 6. GWR and SVCP model estimated coefficients for the intercept
Figure 7. GWR and SVCP model estimated coefficients for area
Figure 8. GWR and SVCP model estimated coefficients for distance to nearest public transportation
Figure 9. Toronto housing sale locations for January 2001, model calibration locations (triangles) and prediction locations (plusses)
Figure 10. GWR and SVCP model predictions for housing prices versus observed values for hold-out sample
Figure 11. 95% prediction intervals for SVCP model predictions (plusses) along with actual housing prices (circles) and GWR predictions (triangles)
Appendix

In this section, we outline the full conditional distributions for the SVCP model parameters when including elicited regression coefficients estimates into the model. In order to perform inference on the model parameters, we must write the posterior distribution for each unknown parameter using the likelihood and prior distribution. The derivation of the full conditional distributions in this paper utilizes two versions of the likelihood. The likelihood for the SVCP model with $Y$ as defined in equation (4) of the paper text is:

$$L(\mu_p, \beta, \tau^2, \phi, T; y) = \left|\tau^2 I\right|^{-1/2} \times \exp\left(-\frac{1}{2}(y - X\beta)^T(\tau^2 I)^{-1}(y - X\beta)\right).$$  \hfill (1)

We can integrate this likelihood with respect to $\beta$ to reduce the autocorrelation in the Markov chain. The likelihood with integrating over $\beta$ is:

$$L'(\mu_p, \tau^2, \phi, T; y) = \left|\Psi\right|^{-1/2} \times \exp\left(-\frac{1}{2}(y - X\mu_p)^T(\Psi)^{-1}(y - X\mu_p)\right),$$  \hfill (2)

where $\Psi = (X(H(\phi) \otimes T)X^T + \tau^2 I)$ and $\mu_p$ is the $np \times 1$ vector of regression coefficient means.

The elicited guesses on the regression coefficients are incorporated through the prior of $\mu_p$, which is $[\mu_p]^\sim N(\mu, \sigma^2 I)$. The hyperparameter vector $\mu$ is set to the expert guesses prior. The prior variance $\sigma^2$ can be set reasonably low based on the confidence of the expert opinions. This specification of the prior for $\mu_p$ is in contrast to the prior when no expert information is available and noninformative values are used. The prior for the regression coefficients is now $[\beta \mid \mu_p, \Sigma_p] = N(\mu_p, \Sigma_p)$. The full conditional distribution for $\mu_p$ is derived using the likelihood integrated over $\beta$, $L'$. The full conditional for the coefficient means is

$$[\mu_p \mid T, \tau^2, \phi; y] \sim L' \times p(\mu_p) = N(\mathbf{m}, \mathbf{S}),$$  \hfill (3)

where $\mathbf{S} = [(\sigma^2 I)^{-1} + X^T \Psi^{-1} X]^{-1}$ and $\mathbf{m} = S(X^T \Psi^{-1} y + (\sigma^2 I)^{-1} \mu)$. The full conditional distributions for the other model parameters are derived using the likelihood $L$. The full conditional for the error variance is:

$$[\tau^2 \mid \beta; y] \sim L \times p(\tau^2) = IG(a + n/2, b + \frac{1}{2}(y - X\beta)^T(y - X\beta)).$$  \hfill (4)

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

$$[T \mid \mu_p, \beta, \phi; y] \sim L \times p(T) \times p(\beta) = IW(v + n, \Sigma + \sum_i (H^{-1}(\phi))_i (\beta(s_i) - \mu_p(s_i))(\beta(s_i) - \mu_p(s_i))^T + \Omega),$$  \hfill (5)

where $\beta = (\beta(s_1), \beta(s_2), \ldots, \beta(s_n))^T$ and $\mu_p = (\mu_{\beta_1}, \mu_{\beta_2}, \ldots, \mu_{\beta_n})^T$. The full conditional distribution for $\beta$ is:
\[ [\beta \mid \mu, \phi, T, \tau^2; y] \sim L \times p(\beta) = N(AC, A), \]  

(6)

where \( A = (X^T X / \tau^2 + H^{-1}(\phi) \otimes T^{-1})^{-1} \) and \( C = X^T y / \tau^2 + (H^{-1}(\phi) \otimes T^{-1}) \mu_p \). The density of the unnormalized full conditional of \( \phi \) is not a recognizable distribution and is:

\[
p(\phi ; T, \beta, \mu_p; y) \propto L \times p(\phi) \times p(\beta) - \frac{1}{2} (\beta - \mu_p)^T (H(\phi) \otimes T)^{-1} (\beta - \mu_p) \times \phi^{v-1} \exp(-\beta \mu_p).
\]

(7)

This density is used in the Metropolis-Hastings algorithm to estimate \( \phi \).

The parameterization of the gamma distribution used in this paper is:

\[
[\phi] \propto \phi^{v-1} \exp(-\lambda \phi), \quad (8)
\]

and the parameterization of the inverse Wishart distribution used in this paper is:

\[
[T] \propto |T|^{-(v+p+1)/2} \exp(-\frac{1}{2} \text{trace} \Omega T^{-1}). \quad (9)
\]
REFERENCES


Goodman AC, Thibodeau TG, 1998, "Housing market segmentation" *Journal of Housing Economics* 7 (2) 121 - 143.


Schnare AB, Struyk RJ, 1976, "Segmentation in Urban Housing Markets" *Journal of Urban Economics* 3 (2) 146 - 166.


Thibodeau TG, 2003, "Marking single-family property values to market" *Real Estate Economics* 31 (1) 1 - 22.


