Journal of Real Estate Finance and Economics, 29:2, 149–166, 2004

The Dynamics of Location in Home Price

ALAN E. GELFAND
Institute of Statistics and Decision Sciences, Duke University, Durham, NC 27708-0251, U.S.A.
E-mail: alan@stat.duke.edu

MARK D. ECKER
Department of Mathematics, University of Northern Iowa, Cedar Falls, IA 50614-0506, U.S.A.
E-mail: ecker@math.uni.edu

JOHN R. KNIGHT
Eberhardt School of Business, University of the Pacific, Stockton, CA 95211, U.S.A.
E-mail: jknight@uop.edu

C. F. SIRMANS
Department of Finance, University of Connecticut, Storrs, CT 06269, U.S.A.
E-mail: cf.sirmans@business.uconn.edu

Abstract

It is well established that house prices are dynamic. It is also axiomatic that location influences such selling prices, motivating our objective of incorporating spatial information in explaining the evolution of house prices over time. In this paper, we propose a rich class of spatio-temporal models under which each property is point referenced and its associated selling price modeled through a collection of temporally indexed spatial processes. Such modeling includes and extends all house price index models currently in the literature, and furthermore permits distinction between the effects of time and location. We study single family residential sales in two distinct submarkets of a metropolitan area and further categorize the data into single- and multiple-transaction observations. We find the spatial component is very important in explaining house price. Moreover, the relative homogeneity of homes within the submarket and the frequency with which homes sell affects the pattern of variation across space and time. Differences between single and repeat sale data are evident. The methodology is applicable to more general capital asset pricing when location is anticipated to be influential.

Key Words: geostatistical modeling, hedonic models, index construction, spatio-temporal process

1. Introduction

The components of house prices and period-to-period changes in house prices are subjects of both academic and practical interest. The hedonic pricing model, dating to Court (1939) and having Rosen (1974) as its basis in theory, is customarily employed to measure the contribution of individual house characteristics to the overall composite value of the housing asset. Location, usually considered the most important of house characteristics, was the topic of many early applications of the hedonic technique to housing data (e.g., Ridker and Henning, 1967), and specifying a property's spatial characteristics within the error structure of pricing models is the subject of considerable recent research effort (e.g., Dubin, 1988; Basu and Thibodeau, 1998; Pace and Gilley, 1998). The hedonic model has also been used to measure the effect of time on selling prices of homes (Goodman, 1978; Thibodeau, 1989). By controlling hedonically for variations in attributes of houses that sell in different time periods, constant quality house price indices that measure the period-to-period changes in price (supply/demand equilibria) can be constructed. Alternatively, repeat sales house price indices (Bailey et al., 1963) can be developed by differencing the hedonic equation. Data for this model, however, are restricted to properties that transact more than once during the period of the index.

In contrast with the vast literature that individually treats the aspects of space and time in the housing context, research jointly considering the effects of these two important characteristics is relatively sparse. The multi-dimensional issue associated with the treatment of space and time is daunting in itself. On top of this are serious questions about the choice of price index methodology if a spatio-temporal approach is to be employed, particularly when both single and multiple sales data are available.

In this paper, we introduce a spatio-temporal model that extends currently existing house price index methods. Our specification of the spatio-temporal process permits site-specific temporal evolution of house prices, but more generally, region-wide spatial evolution of such prices. Unlike the MSA level focus of almost all existing house price index models, our approach allows indexing at arbitrarily high resolution, even down to the individual home level. We employ this model to compare data on single sales with data on repeat sales. We find that spatial variation is an important component of price index changes, even at the micro-market level, and that there are important spatio-temporal differences between markets associated with the two groups of properties.

The format of the paper is as follows. In the following section we provide a review of the spatial and spatio-temporal housing literature and justify our use of an underlying hedonic approach for modeling the process as well as for comparing single and multiple-sales data. Sections 3 and 4 present the details of our hierarchical model and the associated distributional results, respectively. In Section 5, we describe the implementation of our model for creating house price indices as well as for forecasting house prices. We apply our model to a data set whose description appears in Section 6 and the ensuing Bayesian analysis of the results appears in Section 7. Section 8 interprets these results and provides suggestions for further research. Our analysis is illustrative rather than definitive. Indeed, our methodology is applicable to more general capital asset pricing where location is anticipated to be influential.

2. Review of the spatio-temporal literature

Consistent with the axiomatic importance of location on selling price, appropriate hedonic modeling of single family homes almost always introduces characteristics of location. Such characteristics include neighborhood features and proximity or accessibility to certain externalities. Typical neighborhood features capture socio-economic, land use and quality of municipal services, while externalities include business districts, transportation networks, recreational areas and pollution. Neighborhood features as well as externalities may produce either positive or negative influences on house selling price.

Location characteristics are typically introduced into the mean structure in the hedonic model (Ridker and Henning, 1967; Li and Brown, 1980; Dubin and Sung, 1990), but sometimes important neighborhood characteristics are unavailable. Hence, due to these omitted variables, spatial association remains even after inclusion of observed location attributes in the mean structure of the model. In these cases, remaining spatial effects may be introduced into the error structure (Can, 1990; Dubin, 1988, 1992; Basu and Thibodeau, 1998; Pace and Gilley, 1998) or, more generally, as random effects (Gelfand et al., 1998). Such modeling of spatial effects proceeds from one of two possible paths. The first is the so-called geostatistical approach where the covariance between effects at pairs of spatial locations is modeled directly (Dubin, 1988, 1992; Basu and Thibodeau, 1998). The second models the inverse of the covariance matrix using simultaneous autoregressive (Pace and Gilley, 1998) or conditionally autoregressive (Gelfand et al., 1998) specifications.

Modeling the effect of the passage of time on house prices has been the focus of renewed methodological interest over the past decade. Here, time is not viewed as a cause, but rather as a univariate proxy or label for a variety of dynamic factors, economic, political, sociological, etc. Hedonic indices typically treat log selling price as the response variable in a regression model. In addition to house characteristics as explanatory variables, time dummies are utilized to measure price changes associated sale in a particular period. Coefficients of these variables are interpreted as the cumulative percentage change in constant quality house price up to the associated time period. Data for the hedonic model include all home sales in each period.

The competing basic approach to index construction is the repeat sales model, dating to Bailey et al. (1963). This model takes as the response the difference in selling price of specific homes on the log scale. A binary explanatory variable marking the time period of sale takes the value zero for the initial sale and one for each subsequent sale of a particular home. Data for this model are limited to homes that sell more than once over the time spanned by the price index. By now, the literature on both the hedonic and the repeat sales modeling approaches is considerable. Hybrid models, combining advantages of the two conventional approaches, are now appearing in the literature as well (e.g., Case and Quigley, 1991; Quigley, 1995; Hill et al., 1997).

Even though location is fixed, the contribution of location to house value changes over time in response to a variety of micro-market changes. For example, positive or negative externalities may emerge or disappear, affecting house value differentially based on proximity to the externality. Likewise, construction of roads or changes in public transit may alter the accessibility associated with a property's location. Over larger periods of time, the differential effects of dwelling age, often strongly correlated with location, may come into play. Understanding the dynamics of location is evidently important, but has received relatively little treatment to date.

Exceptions to this are Can and Megbolugbe (1997) and Pace et al. (1998) where both use the hedonic model as the basis for spatio-temporal analysis. Can and Megbolugbe (1997) identify ''recent comparable sales'' (properties within a fixed distance which sold within a fixed time period). A distance-weighted average is then entered as an explanatory variable in the mean structure. Pace et al. (1998) propose a spatio-temporal model that synthesizes models from the time series and spatial econometrics literatures. They employ a filtering process based on the spatial and temporal proximity of data, a method that greatly reduces the number of parameters to be estimated while improving estimation and prediction performance.

Archer et al. (1996) and Goetzman and Spiegel (1997) use the repeat sales methodology as the basis for spatio-temporal housing analysis. Archer et al. (1996) are concerned with extracting differential appreciation rates for various census tracts within Dade county, Florida. While tract location does provide some explanation for different house price paths over time, they find that this effect appears to be ‘‘dominated by the idiosyncratic influences of individual homes and their immediate environments.’’ (p. 334). Goetzmann and Spiegel (1997) develop a ‘‘distance-weighted repeat sales’’ procedure to examine housing returns by zip code in the San Francisco Bay area.

There are problems associated with the repeat sales model of Bailey et al. (1963). There is obvious discomfort in ignoring all single sales data since repeat sales typically constitute only 10–20 percent of all transactions (depending on the length of the index period spanned). Concern with sampling bias arises (Gatzlaff and Haurin, 1997), a bias that may be aggravated if analysis is restricted to properties whose characteristics (apart from age) do not change between sales. The differencing of the response variable with the repeat sales technique illuminates the crucial importance of the assumption of parameter stability over the indexing period.

Apart from the information loss in discarding single sale data, there is information loss for the repeat sales model in replacing two observed measurements with one difference. For instance, with usual normal distributions for the log selling price, even under all the assumptions implicit in the repeat sales model, the differences do not become sufficient statistics. The likelihood principle is violated using such models to analyze what has actually been observed. Moreover, any features which would be sought in a repeat sales model can be induced through a suitable hedonic specification.

Further problems with the repeat sales model arise when spatial effects are introduced. Since purely spatial effects in the hedonic model would cancel under differencing, effects which remain must have been spatio-temporal. Space–time interactions must be modeled with considerable care since, in general, it is not evident how to align the temporal scale with the geographic scale.

Given the concerns with the repeat sales model elaborated above, we choose the hedonic model as the basis for analyzing spatio-temporal effects in house prices at the individual property level, and for comparing single and repeat sales data. The details of our model appear in the following section.

3. Modeling details

With geocoded house locations, we have point-referenced data leading naturally to geostatistical modeling specifications (see, for example, Cressie, 1993). In this context, we formulate a rich class of spatio-temporal hedonic models for house prices. These models envision house prices as a conceptually obtainable collection of random variables at each time during the period of observation, at each location in the region of interest. However, in practice, this spatio-temporal process is only observed at a discrete set of locations and a discrete set of time points. Typically, the time scale is divided into intervals of equal length, e.g., months, quarters, years. These intervals are labeled by integer values resulting in an index set $j = 1, \ldots, T$ for time. All transactions occurring in the jth interval are assigned the value j. Since, for each transaction, the actual closing date is known, there is loss of information in such aggregation. Continuous time models for transaction times (which we mention below) might be preferable. After aggregation, the observed number of and set of spatial locations varies with j. If the total number of distinct properties which were sold over the period of observation is n, we have n conceptual time series of selling prices all spatially associated. However, we only see a ''snapshot'' of this $n \times T$ matrix—one row entry for a single sale, two if a house resold once, etc.

More precisely, consider a house, possibly conceptual, at location $\mathbf{s}$ in a region R where $\mathbf{s}$ is viewed as a coordinate pair, typically latitude and longitude, possibly rescaled using an appropriate projection. At each time t in the indexing period, denoted by $[0, T]$, this property has a conceptual log selling price which we denote as $Y(\mathbf{s}, t)$. Thus, $Y(\mathbf{s}, t)$ is a spatio-temporal stochastic process. A random realization of this process is a surface above the space $R \times [0, T]$. The entire realization is never observed. Rather, a given dataset provides a set of observed locations, $\mathbf{s}_1, \mathbf{s}_2, \ldots, \mathbf{s}_n$ in R. Associated with location $\mathbf{s}_i$ is a vector $\mathbf{t}_i$ denoting the observed times the house at $\mathbf{s}_i$ was sold during the indexing period, thus implicitly, the number of times it was sold. Hence, the spatio-temporal process is observed in snapshot form, at certain locations at certain times.

In this formulation we are modeling all transactions in R during $[0, T]$. It may be of interest to segment this population. (In particular, in Section 7, we separate the single sales $(\dim(\mathbf{t}_i) = 1)$ from the repeat sales $(\dim(\mathbf{t}_i) > 1)$.) We can then create distinct models for $Y(\mathbf{s}, t)$ for each segment with the objective of comparison.

Let $\mathbf{X}(\mathbf{s}, t)$ denote a vector consisting of the house and possibly neighborhood characteristics associated with the property at location $\mathbf{s}$ at time t which are to be used to explain $Y(\mathbf{s}, t)$.

Then, we assume

$$Y(\mathbf{s}, t) = \mathbf{X}'(\mathbf{s}, t)\boldsymbol{\beta}(t) + U(\mathbf{s}, t) \tag{1}$$

where $\boldsymbol{\beta}(t)$ is a, possibly time-varying, parameter vector and $U(\mathbf{s}, t)$ is a zero-mean spatio-temporal process. The familiar linear form in (1) is quite flexible being linear in the coefficients but not necessarily in the characteristics.

Spatio-temporal richness is captured by extending $U(\mathbf{s}, t)$ beyond a white noise process. Below, α_{s} denote temporal effects; Ws denote spatial effects. In this regard, we consider the following choices for $U(\mathbf{s}, t)$:

$$U(\mathbf{s}, t) = \alpha(t) + W(\mathbf{s}) + \varepsilon(\mathbf{s}, t), \tag{2}$$

$$U(\mathbf{s}, t) = \alpha_{\mathbf{s}}(t) + \varepsilon(\mathbf{s}, t), \tag{3}$$

$$U(\mathbf{s}, t) = W_t(\mathbf{s}) + \varepsilon(\mathbf{s}, t). \tag{4}$$

Since we are modeling on the log selling price scale, it is natural to introduce error effects in an additive fashion. The given forms avoid specification of space–time interactions. In each of (2)–(4), the $\varepsilon(\mathbf{s}, t)$ are i.i.d. $N(0, \sigma_\varepsilon^2)$ and independent of the other processes. This pure error is viewed as a residual adjustment to the spatio-temporal explanation. Expression (2) provides an additive form in temporal and spatial effects. Since we are modeling log selling prices, this implies roughly a multiplicative form on the price scale. Expression (3) provides temporal evolution at each site; temporal effects are nested within sites. Expression (4) provides spatial evolution over time; spatial effects are nested within time. Spatio-temporal modeling beyond (2)–(4) necessitates the choice of a specification to align the space and time scales. One illustrative version works with $W(\mathbf{s}, t) + \varepsilon(\mathbf{s}, t)$ where $\varepsilon(\mathbf{s}, t)$ is as above. $\mathrm{Cov}(W(\mathbf{s}, t), W(\mathbf{s}', t'))$ is modeled using a so-called separable or product form. Dependence attenuates in a multiplicative manner across space and time. The role of $\varepsilon(\mathbf{s}, t)$ in (2)–(4) is that of a measurement error process. Due to market variability, two identical houses at essentially the same location, sold at essentially the same time, need not sell for identical prices. The $\varepsilon(\mathbf{s}, t)$ process captures this variability.

Next, we consider the components in (2)–(4) in more detail. In (2), if we retain the actual sale times, hence the interval $[0, T]$, we can model $\alpha(t)$ as a one-dimensional (1D) stationary Gaussian process. In particular, for the set of actual sale times, $\{t_1, t_2, \ldots, t_m\}, \boldsymbol{\alpha} = (\alpha(t_1), \ldots, \alpha(t_m))' \sim N(\mathbf{0}, \sigma_\alpha^2 \Sigma(\phi))$ where $(\Sigma(\phi))_{rs} = \mathrm{corr}(\alpha(t_r), \alpha(t_s)) = \rho(|t_r - t_s|; \phi)$ for ρ a valid 1D correlation function. Typical choices for ρ include the exponential, $\exp(-\phi|t_r - t_s|)$ and Gaussian, $\exp(-\phi(t_r - t_s)^2)$, forms.

As noted above, in practice, t is typically confined to an indexing set, $t = 1, 2, \ldots, T$. Then we can simply view $\alpha(1), \ldots, \alpha(T)$ as the coefficients associated with a set of time dummy variables. With this assumption for the $\alpha(t)$s, if in (2), $W(\mathbf{s})$ is set to zero, $\boldsymbol{\beta}(t)$ is assumed constant over time and $\mathbf{X}(\mathbf{s}, t)$ is assumed constant over t, upon differencing, we obtain the seminal repeat sales model of Bailey et al. (1963). Also within these assumptions but restoring $\boldsymbol{\beta}$ to $\boldsymbol{\beta}(t)$ we obtain the extension of Knight et al. (1995). Alternatively, we might set $\alpha(t+1) = \rho\alpha(t) + \eta(t)$ where $\eta(t)$ are i.i.d. $N(0, \sigma_\alpha^2)$. If $\rho < 1$ we have the familiar stationary $AR(1)$ time series, a special case of the continuous time model of the previous paragraph. If $\rho = 1$ the $\alpha(t)$ follow a random walk. With a finite set of times, time dependent coefficients are handled analogously to the survival analysis setting (see, for example, Cox and Oakes, 1984, Chapter 8).

The autoregressive and random walk specifications are naturally extended to provide a model for the $\alpha_\mathbf{s}(t)$ in (3). That is, we assume $\alpha_\mathbf{s}(t+1) = \rho\alpha_\mathbf{s}(t) + \eta_\mathbf{s}(t)$ where again the $\eta_\mathbf{s}(t)$ are all i.i.d. Thus, there is no spatial modeling, only independent conceptual time series at each location. Hence, in the absence of repeat sales, there is no information in the data about ρ so the likelihood can only identify the stationary variance $\sigma_\alpha^2/(1-\rho^2)$ but not σ_α^2 or ρ. The case $\rho < 1$ with $\boldsymbol{\beta}(t)$ constant over time provides the models proposed in Hill et al. (1997, 1999). If $\rho = 1$ with $\boldsymbol{\beta}(t)$ and $\mathbf{X}(\mathbf{s}, t)$ constant over time, upon differencing we obtain the widely-used model of Case and Shiller (1989). In application, it will be difficult to learn about the $\alpha_\mathbf{s}$ processes with typically one or at most two observations for each $\mathbf{s}$.

The $W(\mathbf{s})$ are modeled as a second order stationary Gaussian process in two dimensions. In particular, for the set of locations $\{\mathbf{s}_1, \mathbf{s}_2, \ldots, \mathbf{s}_n\}, \mathbf{W} = (W(\mathbf{s}_1), W(\mathbf{s}_2), \ldots, W(\mathbf{s}_n))' \sim N(\mathbf{0}, \sigma_W^2 H(\delta))$ where $(H(\delta))_{jk} = \mathrm{corr}(W(\mathbf{s}_j), W(\mathbf{s}_k)) = \omega(\mathbf{s}_j - \mathbf{s}_k; \delta)$ for ω a valid 2D correlation function. Here, for illustration, we confine ourselves to a monotonic isotropic specification; if d_{jk} is the Euclidean distance between $\mathbf{s}_j$ and $\mathbf{s}_k$, we use $\omega(\mathbf{s}_j - \mathbf{s}_k; \delta) = \exp(-\delta d_{jk})$. Both Dubin (1992) and Basu and Thibodeau (1998) employ an isotropic stationary Gaussian model for $W(\mathbf{s})$ in (2) (but both delete $\varepsilon(\mathbf{s}, t)$ from their models). Shiller (1993) introduces terms of the form $W(\mathbf{s})$ which are not spatial in nature. Rather, they are coefficients associated with property or asset dummy variables. In the hedonic model, these variables are viewed as proxies for omitted variables.

For $W_t(\mathbf{s})$ in (4), assuming t restricted to an index set, we can, for each t, create a geostatistical model for $\mathbf{W}_t$ introducing $\sigma_W^{2(t)}$ and $\delta^{(t)}$. That is, rather than defining a dummy variable at each t, we conceptualize a separate spatial dummy process at each t. Here, the components of $\mathbf{W}_t$ correspond to the sites at which sales were observed in the time interval denoted by t. The number of transactions in interval t will typically be adequate to learn about the $W_t(\mathbf{s})$ process. Thus, we capture the dynamics of location in a very general fashion. In particular, comparison of the $\sigma_W^{2(t)}$ and $\delta^{(t)}$ reveals the nature of spatial evolution over time.

With a time dummy variable at each t, assessment of temporal effects would be provided through inference associated with these individual continuous spatial variables. For example, a plot of the point estimates against time would clarify size and trend for the effects. With distinct spatial processes, how can we see such temporal patterns? A convenient reduction of each spatial process to a univariate random variable is the block average, i.e., the average of the process over the region. We denote this average by $W_t(R) = \frac{1}{|R|}\int_R W_t(\mathbf{s})d\mathbf{s}$ where R is the region of interest and $|R|$ denotes its area. The block average is a stochastic integral and is analytically difficult to work with. However, viewed as an expectation with respect to a uniform distribution of $\mathbf{s}$ over R, a Monte Carlo integration provides an arbitrarily accurate finite sum approximation. A priori, the expected value of the process at any location, hence, any average of process, has mean 0. However, the block average has the smallest variance of any average over the region encouraging its use as a summary. A posteriori, the block average across t will move away from 0 as encouraged by the data. Hence, assessment of temporal effects using (4) would be provided through posterior inference associated with these averages.

In the above, we assume that the $\mathbf{W}_t$ are independent across t. An alternative possibility is to assume that $W_t(\mathbf{s}) = \Sigma_{j=1}^{t} V_j(\mathbf{s})$ where the $V_j(\mathbf{s})$ are i.i.d. processes, again of one of the foregoing forms. Now, for $t < t^*, \mathbf{W}_t$ and $\mathbf{W}_{t^*}$ are not independent but $\mathbf{W}_t$ and $\mathbf{W}_{t^*} - \mathbf{W}_t$ are. A version of this specification using a geostatistical form with an exponential correlation function yields, upon differencing, the model described in Goetzmann and Spiegel (1997).

4. Associated distributional results

Adopting the Bayesian approach for modeling (1) using (2), (3) or (4), one must specify both the associated likelihoods together with the prior distributions for all model parameters. Bayesian inference proceeds from the posterior distribution of all model parameters given the data. Since the posterior is a high-dimensional, analytically intractable, multi-variate distribution, such inference is implemented using simulation-based model fitting. In practice, samples from the posterior are drawn using Markov Chain Monte Carlo (MCMC) techniques (see Gelfand et al., 1998). By drawing arbitrarily many samples, we can learn arbitrarily well about any feature of the posterior distribution of any model parameter. Moreover, inference for functions of these parameters can be developed by calculating the function at each posterior sample. This gives a sample from the posterior distribution of the function.

We begin by developing the likelihood under model (1) using (2), (3) or (4). Assuming $t\varepsilon\{1, 2, \ldots, T\}$, it is convenient to first obtain the joint distribution for $\mathbf{Y}' = (\mathbf{Y}'_1, \ldots, \mathbf{Y}'_T)$ where $\mathbf{Y}'_t = (Y(\mathbf{s}_1, t), \ldots, Y(\mathbf{s}_n, t))$. That is, each $\mathbf{Y}_t$ is $n \times 1$ and $\mathbf{Y}$ is $Tn \times 1$. This joint distribution will be multi-variate normal. Thus, the joint distribution for the $Y(\mathbf{s}, t)$, which are actually observed, requires only pulling off the appropriate entries from the mean vector and appropriate rows and columns from the covariance matrix. This simplifies the computational bookkeeping, though care is still required.

In the constant $\boldsymbol{\beta}$ case, associate with $\mathbf{Y}_t$ the matrix X_t whose ith row is $\mathbf{X}(\mathbf{s}_i, t)'$. Let $\boldsymbol{\mu}_t = X_t\boldsymbol{\beta}$ and $\boldsymbol{\mu}' = (\mu'_1, \ldots, \mu'_T)$. In the time-dependent parameter case, we merely set $\boldsymbol{\mu}_t = X_t\boldsymbol{\beta}(t)$.

Under (2), let $\boldsymbol{\alpha}' = (\alpha(1), \ldots, \alpha(T)), \mathbf{W}' = (\mathbf{W}(\mathbf{s}_1), \ldots, \mathbf{W}(\mathbf{s}_n))$ and $\boldsymbol{\varepsilon}' = (\varepsilon(\mathbf{s}_1, 1), \varepsilon(\mathbf{s}_1, 2), \ldots, \varepsilon(\mathbf{s}_n, T))$. Then,

$$\mathbf{Y} = \boldsymbol{\mu} + \boldsymbol{\alpha} \otimes \mathbf{1}_{n\times 1} + \mathbf{1}_{T\times 1} \otimes \mathbf{W} + \boldsymbol{\varepsilon} \tag{5}$$

where $\otimes$ denotes the Kronecker product. Hence, given $\boldsymbol{\beta}$ along with the temporal and spatial effects,

$$\mathbf{Y}|\boldsymbol{\beta}, \boldsymbol{\alpha}, \mathbf{W}, \sigma^2_\varepsilon \sim N(\boldsymbol{\mu} + \boldsymbol{\alpha} \otimes \mathbf{1}_{n\times 1} + \mathbf{1}_{T\times 1} \otimes \mathbf{W}, \sigma^2_\varepsilon \mathbf{I}_{\mathrm{Tn}\times\mathrm{Tn}}). \tag{6}$$

Again, $\mathbf{W} \sim N(\mathbf{0}, \sigma^2_W H(\delta))$. If the $\alpha(t)$ follow an AR(1) model, $\boldsymbol{\alpha} \sim N(\mathbf{0}, \sigma^2_\alpha A(\rho))$ where $(A(\rho))_{ij} = \rho^{|i-j|}/(1-\rho^2)$. Hence, if $\boldsymbol{\alpha}$, $\mathbf{W}$ and $\boldsymbol{\varepsilon}$ are independent, marginalizing over $\boldsymbol{\alpha}$ and $\mathbf{W}$, that is, integrating (6) with regard to the prior distribution of $\boldsymbol{\alpha}$ and $\mathbf{W}$, we obtain

$$\mathbf{Y}|\boldsymbol{\beta}, \sigma^2_\varepsilon, \sigma^2_\alpha, \rho, \sigma^2_W, \delta \sim N(\boldsymbol{\mu}, \sigma^2_\alpha A(\rho) \otimes \mathbf{1}_{n\times 1}\mathbf{1}'_{n\times 1} + \sigma^2_W \mathbf{1}_{T\times 1}\mathbf{1}'_{T\times 1} \otimes H(\delta) + \sigma^2_\varepsilon I_{Tn\times Tn}). \tag{7}$$

If the $\alpha(t)$ are coefficients associated with dummy variables (now $\boldsymbol{\beta}$ does not contain an intercept) we only marginalize over $\mathbf{W}$ to obtain

$$\mathbf{Y}|\boldsymbol{\beta}, \boldsymbol{\alpha}, \sigma^2_\varepsilon, \sigma^2_W, \delta \sim N(\boldsymbol{\mu} + \boldsymbol{\alpha} \otimes \mathbf{1}_{n\times 1}, \sigma^2_W \mathbf{1}_{T\times 1}\mathbf{1}'_{T\times 1} \otimes H(\delta) + \sigma^2_\varepsilon I_{Tn\times Tn}). \tag{8}$$

The form in (6) provides conditionally independent Ys and can be viewed as the first stage of a hierarchical model as in, for example, Ecker and Gelfand (2002). At the second stage, the distributional models for $\boldsymbol{\alpha}$ and $\mathbf{W}$ are introduced.

The likelihood resulting from (6) arises as a product of independent normal densities by virtue of the conditional independence. This can facilitate model fitting but at the expense of a very high-dimensional posterior distribution. Marginalizing to (7) or (8) results in a much lower-dimensional posterior. Note, however, that while the distributions in (7) and (8) can be determined, evaluating the likelihood (joint density) requires a high-dimensional matrix inversion. In fact, evaluation of the quadratic form in the likelihood requires only a triangular decomposition of this matrix and then the solution of a linear system of equations defined with this triangular matrix.

Turning to (3), if $\boldsymbol{\alpha}'(t) = (\alpha_{s_1}(t), \ldots, \alpha_{s_n}(t))$ and now $\boldsymbol{\alpha}' = (\alpha'(1), \alpha'(2), \ldots, \alpha'(T))$ with $\boldsymbol{\varepsilon}$ as above,

$$\mathbf{Y} = \boldsymbol{\mu} + \boldsymbol{\alpha} + \boldsymbol{\varepsilon}. \tag{9}$$

Now

$$\mathbf{Y} | \boldsymbol{\beta}, \boldsymbol{\alpha}, \sigma_\varepsilon^2 \sim N(\boldsymbol{\mu} + \boldsymbol{\alpha}, \sigma_\varepsilon^2 I_{Tn \times Tn}). \tag{10}$$

If the $\alpha_{s_i}(t)$ follow an AR(1) model independently across i, then marginalizing over $\boldsymbol{\alpha}$,

$$\mathbf{Y} | \boldsymbol{\beta}, \sigma_\varepsilon^2, \sigma_\alpha^2, \rho \sim N(\boldsymbol{\mu}, A(\rho) \otimes I_{Tn \times Tn} + \sigma_\varepsilon^2 I_{Tn \times Tn}). \tag{11}$$

For (4), let $\mathbf{W}_t' = (W_t(\mathbf{s}_1), \ldots, W_t(\mathbf{s}_n))$ and $\mathbf{W}' = (\mathbf{W}_1', \mathbf{W}_2', \ldots, \mathbf{W}_T')$. Then with $\boldsymbol{\varepsilon}$ as above

$$\mathbf{Y} = \boldsymbol{\mu} + \mathbf{W} + \boldsymbol{\varepsilon}, \tag{12}$$

and

$$\mathbf{Y} | \boldsymbol{\beta}, \mathbf{W}, \sigma_\varepsilon^2 \sim N(\boldsymbol{\mu} + \mathbf{W}, \sigma_\varepsilon^2 I_{Tn \times Tn}). \tag{13}$$

If $\mathbf{W}_t \sim N(\mathbf{0}, \sigma_W^{2(t)} H(\delta^{(t)}))$ independently, $t = 1, 2, \ldots, T$ then, marginalizing over $\mathbf{W}$,

$$\mathbf{Y} | \boldsymbol{\beta}, \sigma_\varepsilon^2, \boldsymbol{\sigma}_W^2, \boldsymbol{\delta} \sim N(\boldsymbol{\mu}, D(\boldsymbol{\sigma}_W^2, \boldsymbol{\delta}) + \sigma_\varepsilon^2 I_{Tn \times Tn}) \tag{14}$$

where $\boldsymbol{\sigma}_W^{2'} = (\sigma_W^{2(1)}, \ldots, \sigma_W^{2(T)}), \boldsymbol{\delta}' = (\delta^{(1)}, \delta^{(2)}, \ldots, \delta^{(T)})$, and $D(\boldsymbol{\sigma}_W^2, \boldsymbol{\delta})$ is block diagonal with the tth block being $\sigma_W^{2(t)}(H(\delta^{(t)}))$. Because D is block diagonal, matrix inversion associated with (14) is less of an issue than for (7) and (8).

We note that with either (3) or (4), $U(\mathbf{s}, t)$ is comprised of two sources of error which the data can not directly separate. However, by incorporating a stochastic assumption on the $\alpha_{\mathbf{s}}(t)$ or on the $W_t(\mathbf{s})$, we can learn about the processes which guide the error components, as (11) and (14) reveal.

Again, for inference with regard to the foregoing models, we adopt a fully Bayesian approach, eschewing customary likelihood analysis. Usual desirable properties associated with the MLEs are predicated upon assumptions which are inappropriate. That is, typically T is not large and even if n is, the study region is usually viewed as fixed, for example, a city or a county. We have so-called infill rather than increasing-domain asymptotics (see Cressie, 1993, p. 350), so that asymptotic standard errors associated with parameter estimates are inappropriate. Instead, the Bayesian approach provides an entire posterior distribution for all model unknowns, as well as for any predictions. Using simulation-based model fitting, we can obtain these distributions to arbitrary accuracy, providing exact inference.

5. Forecasting and index construction

We now turn to forecasting under models (1) with (2), (3) or (4). Such forecasting involves prediction of log selling price at location $\mathbf{s}_0$ and time t_0, that is, of $Y(\mathbf{s}_0, t_0)$. Here $\mathbf{s}_0$ may correspond to a property already sold in $[0, T]$ or perhaps to a new location; in any event $\mathbf{s}_0 \varepsilon R$. However, typically $t_0 > T$ is of interest. Such prediction requires specification of an associated vector of characteristics $\mathbf{X}(\mathbf{s}_0, t_0)$. Also, prediction for $t_0 > T$ is possible in the fixed coefficients case but not in the time-varying coefficients case.

In general, within the Bayesian framework, prediction at $(\mathbf{s}_0, t_0)$ follows from the posterior predictive distribution of $f(Y(\mathbf{s}_0, t_0)|\mathbf{Y})$ where $\mathbf{Y}$ denotes the observed vector of log selling prices. Assuming $\mathbf{s}_0$ and t_0 are new, and for illustration, taking $U(\mathbf{s}, t)$ as in (2),

$$f(Y(\mathbf{s}_0, t_0)|\mathbf{Y}) = \int (f(Y(\mathbf{s}_0, t_0)|\boldsymbol{\beta}, \sigma_\varepsilon^2, \alpha(t_0), W(\mathbf{s}_0)) \\ \cdot f(\boldsymbol{\beta}, \boldsymbol{\alpha}, \mathbf{W}, \sigma_\varepsilon^2, \sigma_\alpha^2, \rho, \sigma_W^2, \delta, \alpha(t_0), W(\mathbf{s}_0)|\mathbf{Y}))$$

Using (15), given a random draw $(\boldsymbol{\beta}^*, \sigma_\varepsilon^{2^*}, \alpha(t_0)^*, W(\mathbf{s}_0)^*)$ from $f(\boldsymbol{\beta}, \sigma_\varepsilon^2, \alpha(t_0), W(\mathbf{s}_0)|Y)$, if we draw $Y^*(\mathbf{s}_0, t_0)$ from $N(X'(\mathbf{s}_0, t_0)\boldsymbol{\beta}^* + \alpha(t_0)^* + W(\mathbf{s}_0)^*, \sigma_e^{2^*})$, marginally, $Y^*(\mathbf{s}_0, t_0) \sim f(Y(\mathbf{s}_0, t_0)|\mathbf{Y})$.

Using sampling-based model fitting, working with (6), we obtain samples from the posterior $f(\boldsymbol{\beta}, \sigma_\varepsilon^2, \sigma_\alpha^2, \rho, \sigma_W^2, \delta, \boldsymbol{\alpha}, \mathbf{W}|\mathbf{Y})$, e.g., $(\boldsymbol{\beta}^*, \sigma_\varepsilon^{2^*}, \sigma_\alpha^{2^*}, \rho^*, \sigma_W^{2^*}, \delta^*, \boldsymbol{\alpha}^*, \mathbf{W}^*)$. But then $f(\boldsymbol{\beta}, \sigma_\varepsilon^2, \sigma_\alpha^2, \rho, \sigma_W^2, \delta, \boldsymbol{\alpha}, \mathbf{W}, \alpha(t_0), W(\mathbf{s}_0)|\mathbf{Y}) = \quad f(\alpha(t_0)|\boldsymbol{\alpha}, \sigma_\alpha^2, \rho) \cdot \quad f(W(\mathbf{s}_0)|\mathbf{W}, \sigma_W^2, \delta) \cdot f(\boldsymbol{\beta}, \sigma_\varepsilon^2, \sigma_\alpha^2, \rho, \sigma_W^2, \delta, \boldsymbol{\alpha}, \mathbf{W}|\mathbf{Y})$. If, for example, $t_0 = T + 1$, and $\alpha(t)$ is modeled as a time series, $f(\alpha(T+1)|\boldsymbol{\alpha}, \sigma_\alpha^2, \rho)$ is $N(\rho\alpha(T), \sigma_\alpha^2)$. If the $\alpha(t)$ are coefficients associated with dummy variables, setting $\alpha(T+1) = \alpha(T)$ is, arguably, the best one can do. The joint distribution of $\mathbf{W}$ and $W(\mathbf{s}_0)$ is a multi-variate normal from which $f(W(\mathbf{s}_0)|\mathbf{W}, \sigma_W^2, \delta)$ is a univariate normal. So if $\alpha(t_0)^* \sim f(\alpha(t_0)|\boldsymbol{\alpha}^*, \sigma_\alpha^{2^*}, \rho^*)$ and $W(\mathbf{s}_0)^* \sim f(W(\mathbf{s}_0)|\mathbf{W}^*, \sigma_W^{2^*}, \delta^*)$, along with $\boldsymbol{\beta}^*$ and $\sigma_\varepsilon^{2^*}$ we obtain a draw from $f(\boldsymbol{\beta}, \sigma_\varepsilon^2, \alpha(t_0), \ W(\mathbf{s}_0)|\mathbf{Y})$. (If $t_0 \varepsilon \{1, 2, \ldots, T\}, \alpha(t_0)$ is a component of $\boldsymbol{\alpha}$, then $\alpha(t_0)^*$ is a component of $\boldsymbol{\alpha}^*$. If $\mathbf{s}_0$ is one of the $\mathbf{s}_1, \mathbf{s}_2, \ldots \mathbf{s}_n, W^*(\mathbf{s}_0)$ is a component of $\mathbf{W}^*$.) Alternatively, one can work with (11). Now, having marginalized over $\boldsymbol{\alpha}$ and $\mathbf{W}$, $Y(\mathbf{s}, t)$ and $\mathbf{Y}$ are no longer independent. They have a multi-variate normal distribution from which $f(Y(\mathbf{s}, t)|\mathbf{Y}, \beta, \sigma_\varepsilon^2, \sigma_\alpha^2, \rho, \sigma_W^2, \delta)$ must be obtained. Note that for multiple predictions, $W(\mathbf{s}_0)$ is replaced by a vector, say $\mathbf{W}_0$. Now $f(\mathbf{W}_0|\mathbf{W}, \sigma_w^2, \delta)$ is a multi-variate normal distribution. No additional complications arise.

We note that prediction of $P(\mathbf{s}_0, t_0) = \exp(Y(\mathbf{s}_0, t_0))$, price rather than log price is likely of greater interest. Formally, from (15), $f(P(\mathbf{s}_0, t_0)|\mathbf{Y})$ is seen to be a mixture of lognormal distributions. But given samples $Y^*(\mathbf{s}_0, t_0)$ from$f(Y(\mathbf{s}_0, t_0)|\mathbf{Y}), P^*(\mathbf{s}_0, t_0) = \exp(Y^*(\mathbf{s}_0, t_0))$ is, immediately, a sample from $f(P(\mathbf{s}_0, t_0)|\mathbf{Y})$. These predictive posterior samples are customarily summarized with a point (mean or median) and interval estimate (lower 0.025, upper 0.025 quantiles).

Closely connected to forecasting is price index construction. For a location $\mathbf{s}$ and time t with specification of a baseline or average vector of characteristics $\mathbf{X}(\mathbf{s}, t)$ we seek the posterior distribution of $P(\mathbf{s}, t)$, the conceptual selling price of such a house. Evidently, $f(P(\mathbf{s}, t)|\mathbf{Y})$ can be obtained (sampled) as above, so we can investigate the index at any period. With a suitable stochastic model for the $\alpha(t)$ we will be able to predict the index at future time points. The choice of $\mathbf{s}$ is arbitrary, but$f(P(\mathbf{s}, t)|\mathbf{Y})$ will clearly depend upon $\mathbf{s}$ (as the foregoing discussion regarding $\mathbf{s}_0$ reveals). While we may choose to provide the price index at a specified choice of location, we have the added advantage of being able to assess the sensitivity of the index to spatial location.

6. An illustrative dataset

The data we analyze are drawn from two regions in the city of Baton Rouge, Louisiana. The two areas are known as Sherwood Forest and Highland Road. These regions are approximately the same size and have similar levels of transaction activity; they differ chiefly in the range of neighborhood characteristics and house amenities found within. Sherwood Forest is a large, fairly homogeneous neighborhood located east, southeast of downtown Baton Rouge. Highland Road, on the other hand, is a major thoroughfare connecting downtown with the residential area to the southeast. Rather than being one homogeneous neighborhood, the Highland Road area consists, instead, of heterogeneous subdivisions. Employing two regions makes a local isotropy assumption more comfortable and allows investigation of possibly differing time effects and location dynamics.

For these regions, a subsample of all single sale transactions and the sample of all repeat sale transactions during the period 1985 through 1995 are studied separately to assess whether the population of single differs from that of repeat sale houses. See Clapp et al. (1991), in this regard. The sample sizes, provided by year in Table 1, are adequate to fit the various models in Section 3 without incurring the previously mentioned matrix inversion difficulty. The location of each property is defined by its latitude and longitude coordinates, rescaled to UTM projection. In addition, a variety of house characteristics, to control for physical differences among the properties, are recorded at the time of sale. We use age, living area, other area (e.g., patios, garages and carports) and number of bathrooms as covariates in our analysis. Summary statistics for these attributes appear in Table 2. We see that the homes in the Highland Road area are somewhat newer and slightly larger than those in the Sherwood area. The greater heterogeneity of the Highland Road homes is borne out by the almost uniformly higher standard deviations for each covariate. In fact, we have more than 20 house characteristics in our dataset, but elaborating the mean with additional features provides little improvement in R^2 and introduces multicollinearity problems. So, we confine ourselves to the four explanatory variables above and turn to spatial modeling to explain a portion of the remaining variability.

Table 1. Sample size by region, type of sale and year.

	Highland		Sherwood	
Year	Repeat	Single	Repeat	Single
1985	25	40	32	29
1986	20	35	32	39
1987	27	32	27	37
1988	16	26	20	34
1989	21	25	24	35
1990	42	29	27	37
1991	29	30	25	31
1992	33	38	39	27
1993	24	40	31	40
1994	26	35	20	34
1995	26	35	21	32
Total	289	365	298	375

To support this, for each region, an ordinary least squares fit for log selling price using these explanatory variables was carried out. Then, an empirical semi-variogram (Cressie, 1993, p. 40) was obtained for each set of residuals. These are shown in Figure 1. (Distance bins are created, each labeled by its center. The average squared difference between residuals, γ, is plotted against these centers.) There is evidence of spatial association, after adjusting for house characteristics. Overall variability of the process (the sill) is much

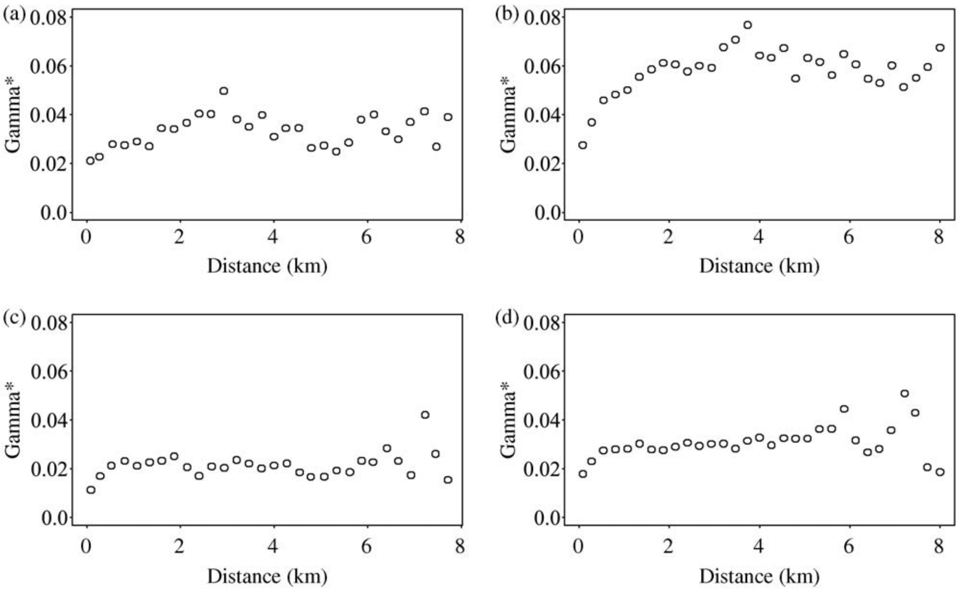

Figure 1. Empirical semivariogram based upon residuals from OLS model for Highland repeat sales in (a), Highland single sales in (b), Sherwood repeat sales in (c) and Sherwood single sales in (d).

greater in Highland than in Sherwood. The spatial variability (sill–nugget) is again larger in Highland than Sherwood. This is consistent with the greater heterogeneity in Highland compared with Sherwood. Also, the range (distance at which the sill is essentially reached) is roughly 2 km in Highland, perhaps 1 km in Sherwood. Spatial association appears to dissipate more slowly in the former.

7. Analysis of the data

We describe the results of fitting the most flexible model from Section 3, i.e., the model in (1) with fixed coefficients and the error structure in (4). This is also the preferred model using the predictive model choice approach in Gelfand and Ghosh (1998). We omit details. Fixed coefficients were justified by the shortness of the observation period. Moreover, the sample sizes precluded the use of time-dependent hedonic coefficients. Again, an exponential isotropic correlation function was adopted.

To complete the Bayesian specification, we adopt rather noninformative priors in order to resemble a likelihood/least squares analysis. In particular, we assume a flat prior on the regression parameter $\boldsymbol{\beta}$ and inverse gamma (a, b) priors for $\sigma_\varepsilon^2, \sigma_W^{2(t)}$ and $\delta^{(t)}, t = 1, \ldots, T$. The shape parameter for these inverse gamma priors was fixed at two, implying an infinite prior variance. We choose the inverse gamma scale parameter for all $\delta^{(t)}$s to be equal, that is, $b_{\delta^{(1)}} = b_{\delta^{(2)}} = \cdots = b_{\delta^{(T)}} = b_\delta$, say, and likewise for $\sigma_W^{2(t)}$. Furthermore, we set $b_{\sigma_\varepsilon} = b_{\sigma_W^2}$ reflecting uncertain prior contribution from the nugget to the sill. Finally, the exact values of $b_{\sigma_\varepsilon}, b_{\sigma_W^2}$ and b_δ vary between region and type of sale reflecting different prior beliefs about these characteristics.

Models are fitted using Gibbs sampling (Gelfand and Smith, 1990). Five chains were run to 10,000 iterations each. After assessing convergence using customary diagnostics, we burn in 5,000 observations and thin every 25th subsequent observation, resulting in a posterior sample of 1,000 observations.

Inference for the house characteristic coefficients is provided in Table 3 (point and 95 percent interval estimates). Age, living and other area are significant in all cases; number of bathrooms is significant only in Sherwood repeat sales. Significance of living area is much stronger in Highland than in Sherwood. The Highland sample is composed of homes from several heterogeneous neighborhoods. As such, living area not only measures differences in house size, but may also serve as a partial proxy for construction quality and

Table 2. Mean and standard deviation for house characteristics by region and type of sale.

	Highland		Sherwood	
Variable	Repeat	Single	Repeat	Single
Age	11.10 (8.15)	12.49 (11.37)	14.21 (8.32)	14.75 (10.16)
Bathrooms	2.18 (0.46)	2.16 (0.56)	2.05 (0.36)	2.02 (0.40)
Living area	2265.4 (642.9)	2075.8 (718.9)	1996.0 (566.8)	1941.5 (616.2)
Other area	815.1 (337.7)	706.0 (363.6)	726.0 (258.1)	670.6 (289.2)

Table 3. Parameter estimates (median and 95 percent interval estimates) for house characteristics.

Region	Variable	Parameter	Repeat	Single
Highland	Intercept	β_0	11.63 (11.59, 11.66)	11.45 (11.40, 11.50)
	Age	β_1	− 0.04 (− 0.07, − 0.02)	− 0.08 (− 0.11, − 0.06)
	Bathrooms	β_2	0.02 (− 0.01, 0.04)	0.02 (− 0.01, 0.05)
	Living area	β_3	0.28 (0.25, 0.31)	0.33 (0.29, 0.37)
	Other area	β_4	0.08 (0.06, 0.11)	0.07 (0.04, 0.09)
Sherwood	Intercept	β_0	11.33 (11.30, 11.36)	11.30 (11.27, 11.34)
	Age	β_1	− 0.06 (− 0.07, − 0.04)	− 0.05 (− 0.07, − 0.03)
	Bathrooms	β_2	0.05 (0.03, 0.07)	0.00 (− 0.02, 0.02)
	Living area	β_3	0.19 (0.17, 0.21)	0.22 (0.19, 0.24)
	Other area	β_4	0.02 (0.01, 0.04)	0.06 (0.04, 0.08)

for neighborhood location within the sample. The greater homogeneity of homes in Sherwood implies less variability in living area (as seen in Table 2) and reduces the importance of these variables in explaining house price.

Turning to the error structure, the parameters of interest for each region are the $\sigma_w^{2(t)}$, the $\delta^{(t)}$ and σ_e^2. The sill at time t is $\text{Var}(Y(s,t)) = \sigma_w^{2(t)} + \sigma_e^2$. Figure 2 plots the posterior medians of these sills. We see considerable difference in variability over the groups and time, providing support for distinct spatial models at each t. Variability is highest for Highland single sales, lowest for Sherwood repeats. These qualitative model based

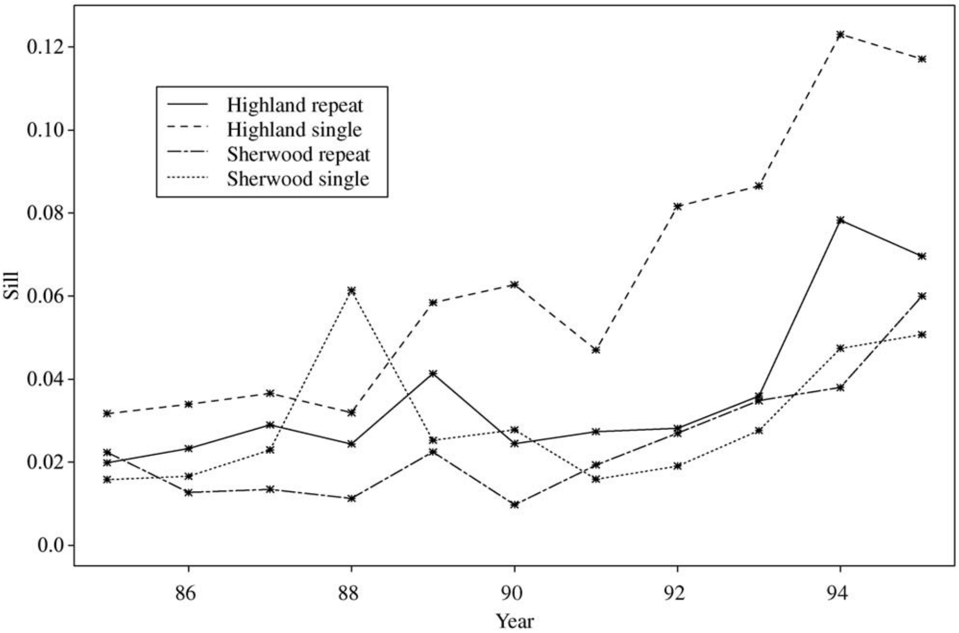

Figure 2. Posterior median sill by year.

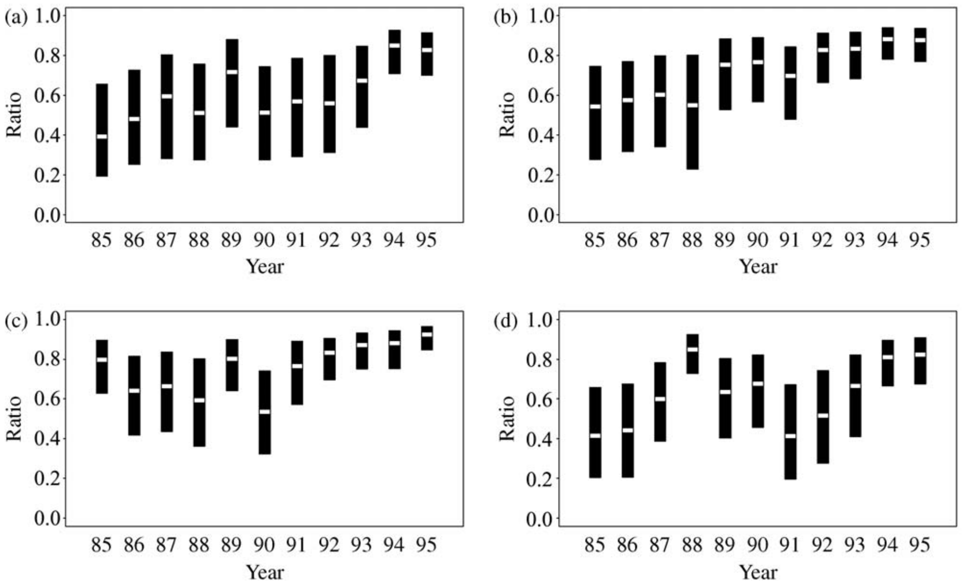

Figure 3. Posterior median and 95 percent interval estimates for the proportion of spatial variance to total variance by year for Highland repeat sales in (a), Highland single sales in (b), Sherwood repeat sales in (c) and Sherwood single sales in (d).

findings concur with the earlier exploratory conclusions from Figure 1. The additional insight is the effect of time. Variability is generally increasing over time.

In Figure 3 we obtain posterior median and interval estimates for $\sigma_w^{2(t)}/(\sigma_e^2 + \sigma_w^{2(t)})$, the proportion of spatial variance to total. Figure 3 adds insight to Figure 1 since it is done by year rather than by distance. Note that the strength of the spatial story is considerable; 40–80 percent of the variability is spatial. Notice also that the proportion of spatial variability differs between groups and over time. For all groups, the proportion is generally higher in more recent years. Spatial variability appears to be accounting for an increasing percentage of the overall variability.

In Figure 4 we provide point and interval estimates for the range. The analysis here refines our rough assessments at the end of Section 6 based on Figure 1. Formally, the range, r, for an asymptotic correlation function is the distance at which the spatial correlation between sites drops to 0.05, that is, r satisfies $e^{-\phi r} = 0.05$. Note that whether we work with single or repeat sales under model (1) using (4), there is always a spatial effect for each transaction. Also, for repeat sales, under differencing the spatial effect does not vanish (except in the unlikely situation where both the sale and the resale were in the same time period). Hence, spatial range is meaningful for all groups. The ranges for the repeat sales are quite similar for the two regions, showing some tendency to increase in the later years of observation. By contrast, the range for the Highland single sales is much different from that for Sherwood. It is typically greater and much more variable. The latter again, is a refection of the high variability in the single sale home prices in Highland. The resulting posteriors are more dispersed.

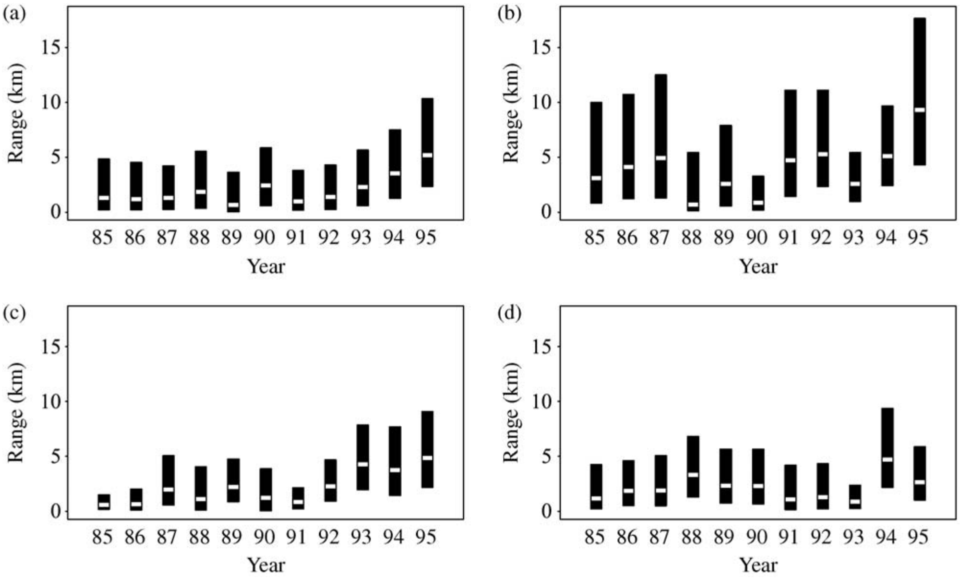

Figure 4. Posterior median and 95 percent interval estimates for the range by year for Highland repeat sales in (a), Highland single sales in (b), Sherwood repeat sales in (c) and Sherwood single sales in (d).

Finally, in Figure 5, we present the posterior distribution of the block averages (discussed at the end of Section 3) for each of the four analyses. Again, these block averages are viewed as analogues of more familiar time dummy variables. Time effects are evident. In all cases, we witness somewhat of a decline in magnitude in the 1980s and an increasing trend in the 1990s.

8. Discussion

Location has always been an important component of house prices, but only recently have the techniques of spatial statistics been brought to bear on this area of study. In this paper, we offer a flexible method for examining temporal and spatial effects on price. The range of specifications stemming from our general model of the spatio-temporal error process allows for additive effects of space and time, temporal evolution at each location and spatial evolution over time. We chose the last of these specifications to empirically study the spatio-temporal differences related to single versus multiple sales as well as differences related to the degree of homogeneity among homes within samples of data.

Under such modeling, spatial effects emerge as very important in explaining house prices. We also find that spatial effects vary substantially between single and repeat transactions. This result supports concerns about bias associated with the repeat sales models that are frequently used in house price index construction. Likewise, we see significant differences in spatial effects between housing submarkets that differ mainly in the degree of similarity among subsample observations. With the growing importance of

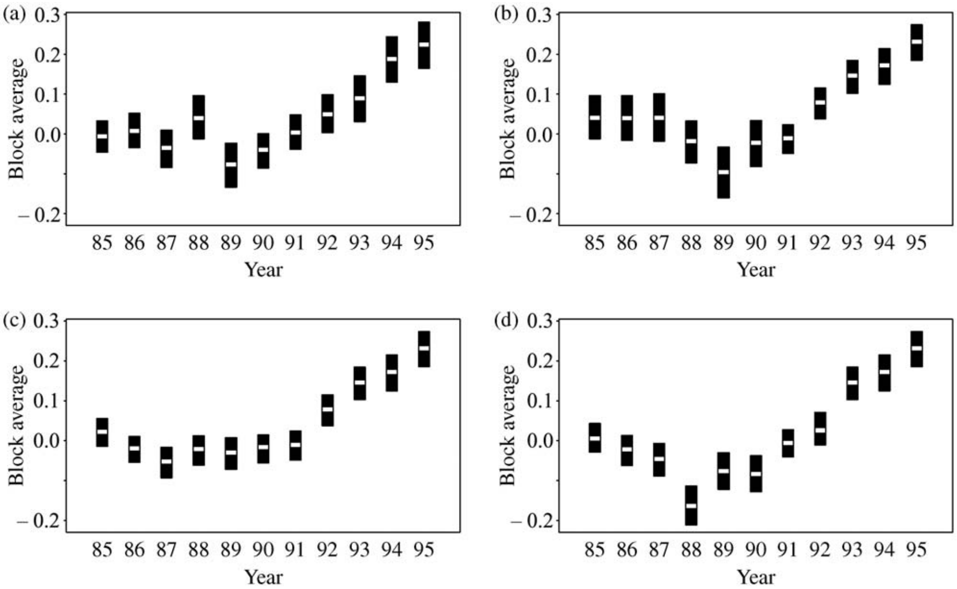

Figure 5. Posterior median and 95 percent interval estimates for the block averages by year for Highland repeat sales in (a), Highland single sales in (b), Sherwood repeat sales in (c) and Sherwood single sales in (d).

computer-generated property appraisals, this finding implies that the selection of comparable properties is location dependent. Moreover, differing spatial effects across housing submarkets holds important implications for the appropriate level of aggregation in house price index construction.

Considerable opportunity for further research remains. It is apparent that the models in (1)–(4) could be applied to pricing for any capital asset when location is expected to be influential. In our context, at the modeling level, we could extend spatial correlation beyond isotropy. We could, in fact, allow local variability in the sill, nugget and range resulting in a nonstationary error specification. We could assume a nonGaussian error process, for example, a *t*-process to capture heavier tails in the data. We could introduce time-dependent coefficients. We could also work with exact time of sale rather than with an indexing set.

References

Archer, W. R., D. H. Gatzlaff, and D. C. Ling. (1996). ''Measuring the Importance of Location in House Price Appreciation,'' *Journal of Urban Economics* 40, 334–353.

Bailey, M. J., R. F. Muth, and H. O. Nourse. (1963). ''A Regression Method for Real Estate Price Index Construction,'' *Journal of the American Statistical Association* 58, 933–942.

Basu, S., and T. G. Thibodeau. (1998). ''Analysis of Spatial Correlation in House Prices,'' *Journal of Real Estate Finance and Economics* 17, 61–85.

Can, A. (1990). ''The Measurement of Neighborhood Dynamics in Urban Housing Prices,'' *Economic Geography* 66(3), 254–272.

Can, A., and I. Megbolugbe. (1997). ''Spatial Dependence and House Price Index Construction,'' *Journal of Real Estate Finance and Economics* 14, 203–222.

Case, B., and J. M. Quigley. (1991). ''The Dynamics of Real Estate Prices,'' *The Review of Economics and Statistics* 73(3), 50–58.

Case, K. E., and R. J. Shiller. (1989). ''The Efficiency of the Market for Single Family Homes,'' *American Economic Review* 79, 125–137.

Clapp, J. M., C. Giaccotto, and D. Tirtiroglu. (1991). ''Repeat Sales Methodology for Price Trend Estimation,'' *Journal of the American Real Estate and Urban Economics Association* 19, 270–285.

Court. A. T. (1939). *The Dynamics of Automobile Demand.* New York: General Motors.

Cox, D. R., and D. Oakes. (1984). *Analysis of Survival Data*. New York: Chapman and Hall.

Cressie, N. (1993). *Statistics for Spatial Data*. New York: John Wiley and Sons.

Dubin, R. A. (1988). ''Estimation of Regression Coefficients in the Presence of Spatial Autocorrelated Error Terms,'' *Review of Economics and Statistics* 70, 466–474.

Dubin, R. A. (1992). ''Spatial Autocorrelation and Neighborhood Quality,'' *Regional Science and Urban Economics* 22, 433–452.

Dubin, R. A., and C.-H. Sung. (1990). ''Specification of hedonic Regressions: Non-nested Tests on Measures of Neighborhood Quality,'' *Journal of Urban Economics* 27, 97–110.

Ecker, M. D., and A. E. Gelfand. (2003). ''Spatial Modeling and Prediction under Stationary Non-Geometric Range Anisotropy,'' *Environmental and Ecological Statistics* 10, 165–178.

Gatzlaff, D. H., and D. R. Haurin. (1997). ''Sample Selection Bias and Repeat-Sales Index Estimates,'' *The Journal of Real Estate Finance and Economics* 14(1/2), 33–50.

Gelfand, A. E., and S. K. Ghosh. (1998). ''Model Choice: A Minimum Posterior Predictive Loss Approach,'' *Biometrika* 85, 1–11.

Gelfand, A. E., S. K. Ghosh, J. R. Knight, and C. F. Sirmans. (1998). ''Spatio-Temporal Modeling of Residential Sales Data,'' *Journal of Business and Economic Statistics* 16, 312–321.

Gelfand, A. E., and A. M. F. Smith. (1990). ''Sampling Based Approaches to Calculating Marginal Densities,'' *Journal of the American Statistical Association* 85, 398–409.

Goetzmann, W. N., and M. Spiegel. (1997). ''A Spatial Model of Housing Returns and Neighborhood Substitutability,'' *Journal of the Real Estate Finance and Economics* 14, 11–32.

Goodman, A. C. (1978). ''Hedonic Prices, Price Indices, and Housing Markets,'' *Journal of Urban Economics* 5(4), 471–484.

Hill, R. C., J. R. Knight, and C. F. Sirmans. (1997). ''Estimating Capital Asset Price Indexes,'' *The Review of Economics and Statistics* 79, 226–233.

Hill, R. C., C. F. Sirmans, and J. R. Knight. (1999). ''A Random Walk Down Main Street,'' *Regional Science and Urban Economics* 29, 89–103.

Knight, J. R., J. Dombrow, and C. F. Sirmans. (1995). ''A Varying Parameters Approach to Constructing House Price Indexes,'' *Real Estate Economics* 23(2), 87–105.

Li, M., and H. J. Brown. (1980). ''Micro-neighborhood Externalities and Hedonic Housing Prices,'' *Land Economics* 56, 125–141.

Pace, R. K., R. Barry, J. M. Clapp, and M. Rodriguez. (1998). ''Spatiotemporal Autoregressive Models of Neighborhood Effects,'' *The Journal of Real Estate Finance and Economics* 17(1), 15–33.

Pace, R. K., and O. W. Gilley. (1998). ''Generalizing OLS and the Grid Estimator,'' *Real Estate Economics* 26(2), 331–347.

Quigley, J. M. (1995). ''A Simple Hybrid Model for Estimating Real Estate Indexes,'' *Journal of Housing Economics* 4, 1–12.

Ridker, R. G., and J. A. Henning. (1967). ''The Determinants of Property Values with Special Reference to Air Pollution,'' *The Review of Economics and Statistics* 49(2), 246–257.

Rosen, S. (1974). ''Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition,'' *The Journal of Political Economy* 82(1), 34–55.

Shiller, R. J. (1993). ''Measuring Asset Values for Cash Settlement in Derivative Markets: Hedonic Repeated Measures Indices and Perpetual Futures,'' *The Journal of Finance* 98(3), 911–931.

Thibodeau, T. G. (1989). ''Housing Price Indexes from the 1974–83 SMSA Annual Housing Surveys,'' *Journal of the American Real Estate and Urban Economics Association* 17(1), 100–117.