Journal of Real Estate Finance and Economics, Vol. 17:1, 15–33 (1998)

Spatiotemporal Autoregressive Models of Neighborhood Effects

R. KELLEY PACE
Department of Finance, E. J. Ourso College of Business Administration, Louisiana State University, Baton Rouge, LA 70803, e-mail: KelleyPace@compuserve.com kpace@unixl.sncc.lsu.edu

RONALD BARRY
Department of Mathematical Sciences, University of Alaska, Fairbanks, Alaska 99775-6660, e-mail: FFRPB@aurora.alaska.edu

JOHN M. CLAPP
University of Connecticut, 368 Fairfield Road, U-41RE, Storrs, CT 06269-2041, e-mail: john@sbaserv.sba.uconn.edu

MAURICIO RODRIQUEZ
Texas Christian University, P.O. Box 32868, Fort Worth, TX 76129, e-mail: mrodriguez@zeta.is.tcu.edu

Abstract

Using 70,822 observations on housing prices from 1969 to 1991 from Fairfax County Virginia, this article demonstrates the substantial benefits obtained by modeling the spatial as well as the temporal dependence of the data. Specifically, the spatiotemporal autoregression with twelve variables reduced median absolute error by 37.35% relative to an indicator-based model with twenty-six variables. One-step ahead forecasts also document the improved performance of the proposed spatiotemporal model. In addition, the article illustrates techniques for rapidly computing the estimates and shows how to compute indices for any location.

Key Words: STAR, STARIMA, space-time, local price indices

1. Introduction

To say that the price of a parcel of real estate depends on its location and recent market events seems all too obvious. However, the optimal way of incorporating spatial and temporal (spatiotemporal) dependencies into empirically feasible pricing models does not seem quite so obvious.

The use of indicator variables provides the easiest, although not necessarily optimal, way to model spatiotemporal dependencies. Due to the unequivocal ordering of time, one can easily determine the number of temporal indicator variables. For example, given data with a span of twenty years, one might have twenty annual indicator variables (given no intercept). Unfortunately, the lack of a unique, optimal arrangement (tessellation) of space impedes the formation of a simple set of spatial indicator variables. To capture the local structure of the market argues for employing relatively many indicator variables. To maintain parsimony argues for employing relatively few indicator variables. Modeling both space and time with interactions further raises the number of required indicators. For example, if ten indicators for time and twenty indicators for space seems adequate for modeling each separately, the joint set would require 200 spatiotemporal indicator variables (given no intercept). Hence, it appears difficult to maintain parsimony and successfully capture local spatiotemporal effects.

To better capture the effect of both spatial and temporal information on real estate prices, while overcoming some of the problems associated with indicator variable models, we introduce a spatiotemporal model that uses information from nearby, recently sold properties in predicting the value of a given property. Instead of assuming that each region has its own effect modeled by a separate parameter as with the indicator variable based models, the spatiotemporal models assume that nearby properties (''comparables'') have the same relation to the observations (''subjects'') across the entire sample. These spatiotemporal models essentially generalize conventional hedonic regressions and the adjustment grid method (Pace and Gilley, 1998). This approach leads naturally to a more parsimonious description of the data than the indicator variable approach.

This specific model falls under the very broad classification of spatiotemporal autoregressive (STAR) models. Applications to business and economic data are rare (see Pfeifer and Bodily, 1990). However, these models have been applied to other areas such as imaging or hydrology. For example, Szummer and Picard (1996) use a STAR model to examine evolving images such as moving water or fire. Deutsch and Ramos (1986) have used them to predict river heights. See Cressie (1993) for more details on STAR models.

As a specific example of the two approaches, we estimated both an indicator variable model and a spatiotemporal model using data from Fairfax County, Virginia. The selected spatiotemporal model uses twelve variables while the indicator variable-based model required twenty-six variables. Despite using fewer degrees of freedom, the spatiotemporal model showed strong improvements in goodness of fit relative to the indicator variable-based model. For example, the median absolute errors went from 0.1478 under the indicator variable model to 0.0926 under the spatiotemporal model, a reduction of 37.35%.

In addition to strong statistical performance, spatiotemporal models provide other advantages. For example, since the estimated model provides a price surface that evolves over time, one can construct indices over time for any given location or location surfaces at any given point in time.

In what follows, section 2 introduces the design of the spatiotemporal model, section 3 applies the model to housing data from Fairfax County, Virginia, and section 4 concludes with the key results.

2. A Simple Spatiotemporal Model

This section presents a simple spatiotemporal model. Section 2.1 develops the spatiotemporal model based on a generalization of the traditional autoregressive error model, section 2.2 describes the spatiotemporal and temporal weight matrices, section 2.3 discusses estimation of the model, and section 2.4 delves into computational issues.

2.1. Spatiotemporal Autoregressions

Assume the following autoregressive process

$$(I - W)Y = (I - W)X\beta + \varepsilon, \tag{1}$$

where Y denotes the n by 1 vector of observations on the dependent variable, X denotes the n by k matrix of observations on the independent variables of interest, β denotes a k by 1 vector of parameters, ε denotes an n by 1 vector of normal *iid* errors, and W denotes an n by n spatiotemporal weight matrix. The diagonal entries of W contain zeros to prevent each observation from predicting itself. Also, W contains only nonnegative elements. In addition, we assume that each row in W sums to 1. Hence, W is a linear filter (Davidson and MacKinnon, 1993, p. 691).

For further structure, we assume the observations have been ordered according to time with the first row of W corresponding to the oldest observation. As Clapp, Dolde, and Tirtiroglu (1995) documented, a given transactions price impounds information (for example, from changes in local taxes and public services) relevant to pricing of neighboring properties. Thus, it seems eminently reasonable to assume that the sales price of a neighboring property will influence the subject property only if the neighboring sale is earlier in time. The time ordering of W (and its components) greatly simplifies the matrix multiplications, equation solutions, and determinant computations necessary to estimate the model.

Essentially, multiplying the random variables X, Y by $(I - W)$ filters these variables to the resulting transformed random variables $(I - W)Y$, $(I - W)X$ do not evince autocorrelation.[1] In a temporal context, the operation $(I - W)Y$ reduces to taking the current values of Y and subtracting the previous values (or an average of the previous values) of Y scaled by a constant less than 1 (the autoregressive parameter). In a spatial context, the operation $(I - W)Y$ reduces to taking the values of Y at each location and subtracting the average of the surrounding values of Y scaled by a constant less than 1 (the spatial autoregressive parameter).

In spatiotemporal estimation, we wish to perform similar operations,—namely, to take the current value of a variable at a location and subtract an average of past, surrounding values scaled by a constant less than 1. To implement this in a flexible way, we begin by partitioning W into a matrix S that specifies spatial relations among previous observations and T that specifies temporal relations among previous observations. Hence, the matrix T is a lag operator for regularly observed data. Also, S functions in space much like a lag operator does in time. The requirement of specifying relations only among previous observations and the ordering of the observations by time means W is a lower triangular matrix (barring ties in time).

We could implement the spatiotemporal filtering in a variety of different ways. For example, we could filter for space and time additively $(I - W) = (I - \phi_S S - \phi_T T)$, where ϕ_S, ϕ_T represent autoregressive parameters related to space and time. We assume these parameters have absolute values of 1 or less. While conceptually simple, this tacitly assumes no interaction between the spatial and the temporal effects. Alternatively, we could filter for time first and then space $(I - W) = (I - \phi_S S)(I - \phi_T T)$ or space first and then time $(I - W) = (I - \phi_T T)(I - \phi_S S)$.[2] However, this presupposes we know which to filter first. On the other hand, we could generalize these by linearly combining both these forms. Equation (2) represents a further generalization of the linear combination of the two forms:

$$W = \phi_S S - \phi_T T + \phi_{ST} ST + \phi_{TS} TS. \tag{2}$$

The form in (2) subsumes the linear combination of filtering for space first and then time or time first and then space. The matrices ST and TS will not generally have the same values. Hence, it seems reasonable to allow these variables their own autoregressive parameters, ϕ_{ST}, ϕ_{TS}. The flexibility of the filtering should allow (2) to encompass the true model in many situations.

Substituting the form of the filter in (2) into (1) yields (3):

$$\begin{aligned}(I - \phi_T T - \phi_S S - \phi_{ST} ST - \phi_{TS} TS)Y \\ = (I - \phi_T T - \phi_S S - \phi_{ST} ST - \phi_{TS} TS)X\beta + \varepsilon.\end{aligned} \tag{3}$$

We could further generalize this and allow the parameters β to vary for each form of lag. Hence, we write the generalized model as

$$\begin{aligned}Y = Z\theta + X\beta_1 + TX\beta_2 + SX\beta_3 + STX\beta_4 + TSX\beta_5 + \phi_T TY + \phi_S SY \\ + \phi_{ST} STY + \phi_{TS} TSY + \varepsilon,\end{aligned} \tag{4}$$

where Z denotes the n by p_1 matrix of observations on unlagged independent variables, X denotes the n by p_2 matrix of observations on the independent variables that will also have lags, $\beta_1, \beta_2, \beta_3, \beta_4, \beta_5$ denote five p_2 by 1 vectors of parameters, and θ denotes a p_1 by 1 vector of parameters on the unlagged variables. Essentially (4) synthesizes the autoregressive distributed lag model of time series and the mixed regressive spatially autoregressive model of spatial econometrics (Ord, 1975; Anselin, 1988).

Like these antecedent models, various restrictions yield different models of interest. For example, the restrictions $\theta, \phi_T, \phi_S, \phi_{ST}, \phi_{TS}, \beta_2, \beta_3, \beta_4, \beta_5 = 0$ yield the usual regression of Y on X, while the restrictions $\beta_2 = -\phi_T\beta_1, \beta_3 = -\phi_S\beta_1, \beta_4 = -\phi_{ST}\beta_1, \beta_5 = -\phi_{TS}\beta_1$, and $\theta = 0$ yield an autoregression in errors as previously described by (1). If one believes in the general to specific model strategy, one should estimate the general model and test these restrictions before adopting a specific alternative such as the autoregression in errors specification.[3]

The above model (4) explains the dependent varliables as a function of the independent variables, the temporal effects that apply to all observations, and the effects that arise due to proximity in both time and space. Separating out the aggregate temporal and spatiotemporal effects allows for flexible specification of the global effects in time T versus the local effects in space and time in S. The spatial weight matrix S has a temporal element as well since it only has nonzero entries for previously sold observations.

Incorporating the interaction term ST and TS allows for the modeling of potentially compound spatiotemporal effects.

2.2. *Specification of the Spatial and Temporal Weight Matrices*

We restrict the rows of S and T to sum to 1

$$\left(\underset{(n \text{ by } n)}{S} \underset{(n \text{ by } 1)}{[1]} = \underset{(n \text{ by } 1)}{[1]}, \underset{(n \text{ by } n)}{T} \underset{(n \text{ by } 1)}{[1]} = \underset{(n \text{ by } 1)}{[1]} \right).$$

Hence, S and T are row-stochastic matrices. In the spatial econometrics literature such weighting matrices are said to be "standardized" (Anselin and Hudak, 1992, p. 514). One can also interpret S and T as linear filters (Davidson and MacKinnon, 1993, p. 691). The nonzero entries on the rows of S and T represent the other observations that directly interact with the observation itself. Note, we rule out nonzero entries for the diagonal of S and T to prevent observations from predicting themselves. Since the observations are temporally ordered and since we condition only on previous transactions, both S and T are strictly lower triangular ($i \leq j \leftrightarrow S_{ij} = 0$ and $T_{ij} = 0$). Hence, the first row of S and T has all zeros. Note, the first row of S, T is associated with the oldest transaction and the last row of S, T is associated with the most recent transaction. We furthermore require nonnegative values for S, T (that is, $S_{ij} \geq 0, T_{ij} \geq 0$).

We can impose some additional structure on S and T. Housing data display uneven density in both the temporal and spatial dimensions. A distance of one mile in the center of the city does not act the same as a distance of one mile in the suburbs. Similarly, transactions volume in housing markets can vary greatly over time. Real estate appraisers, who have great practical experience in predicting short-run housing prices, use a fixed number of neighbors in space ("comparables") and tend to pick a fixed number of neighbors in time as well.[4] The use of ordinal distance and time tends to reduce the problems created by uneven data densities. For example, when the density of observations over time or space is very high, T or S averages over a short time interval or small spatial area. Similarly, when observations occur infrequently, T or S may average over a long time interval or large spatial area. Nearest neighbors have often been employed in density estimation as variable bandwidth smoothers (Silverman, 1986).

A heuristic approach can sometimes clarify the logistical complexities of spatiotemporal relations. Suppose there are six observations occurring at time periods 0, 1, ... , 5 from a set of six locations $a, b, \ldots, f$ that lie on a line in that order. Further suppose $Y_a Y_b \ldots Y_f = 10\ 11 \ldots 15$. The nearest neighbor to a is b, the nearest neighbors to c are a and b, and so forth. In this example, we assume one temporal lag and two spatial neighbors ($m_s = 2, m_T = 1$). Suppose we have some temporally ordered obsrvations on Y from different locations as presented below. Since we order over time, the observations in Y will not usually have a perfect spatial ordering. Because of the zeros in the first row, the observation at time 0 does not play a direct role but does help in forming the temporal and spatial lags. Since only previously transacted observations appear in S, T, these are lower triangular:

$$S = \begin{bmatrix} 0 & \cdots & & & & \\ 1 & 0 & \cdots & & & \\ \frac{1}{2} & \frac{1}{2} & 0 & \cdots & & \\ \frac{1}{2} & \frac{1}{2} & 0 & 0 & \cdots & \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 & \cdots \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 & 0 \end{bmatrix} \quad T = \begin{bmatrix} 0 & \cdots & & & & \\ 1 & 0 & \cdots & & & \\ 0 & 1 & 0 & \cdots & & \\ 0 & 0 & 1 & 0 & \cdots & \\ 0 & 0 & 0 & 1 & 0 & \cdots \\ 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix}$$

$$Y = \begin{bmatrix} 12 \\ 15 \\ 10 \\ 13 \\ 14 \\ 11 \end{bmatrix} \begin{matrix} (0, c) \\ (1, f) \\ (2, a) \\ (3, d) \\ (4, e) \\ (5, b) \end{matrix}$$

$$TY = \begin{bmatrix} 0 \\ 12 \\ 15 \\ 10 \\ 13 \\ 14 \end{bmatrix}, \quad SY = \begin{bmatrix} 0 \\ 12 \\ 13.5 \\ 13.5 \\ 14 \\ 11 \end{bmatrix} \quad TSY = \begin{bmatrix} 0 \\ 0 \\ 12 \\ 13.5 \\ 13.5 \\ 14 \end{bmatrix}, \quad STY = \begin{bmatrix} 0 \\ 0 \\ 6 \\ 6 \\ 11 \\ 7.5 \end{bmatrix}$$

Obviously, STY and TSY do not have the same values. This means one will obtain different results by filtering for space first and time second or time first and space second. As there was a perfect linear relation between Y and location, SY for the last observations perfectly matches Y. It takes some observations before SY can perform adequately because at first only a few potential "comparables" exist. This leads to the decision of how many rows to discard.

Switching to implementation details, we used a geographic information system (GIS) to determine the decimal latitude and longitude of each house from its address, an operation known as geocoding. In the analysis, we used the unprojected decimal latitude and longitude coordinates to compute the Euclidean distance d_{ij} between every pair of observations j and i where $(i > j)$. We subsequently sorted these distances and formed the set of individual neighbor matrices $S_1, S_2, \ldots, S_{m_S}$, where S_1 represents the closest previously sold neighbor (shortest distance), S_2 represents the second previously sold neighbor (second shortest distance), and so on. The first rows of these matrices may have all zeros due to a lack of previously transacted neighbors. These very sparse matrices have a 1 in each row and contain 0s otherwise (apart from the initial rows). One can also restrict the past time interval used for neighbor selection. A priori, we decided that fifteen neighbors ($m_S = 15$) should capture the vast majority of the spatial effects. Based on both our priors and some very approximate preliminary fitting, we decided to restrict the construction of S to neighbors within five years in time. Hence, we do not use comparables that are more than five years old.

We computed the overall spatial matrix S via

$$S = \frac{\sum_{l=1}^{m_s} \lambda^l S_l}{\sum_{l=1}^{m_s} \lambda^l}, \tag{5}$$

where λ^l weighs the relative effect of the lth individual neighbor matrix. Hence, S depends on the parameters λ and m_S, as well as on the underlying metric.[5] Thus, (5) imposes an autoregressive distributed lag structure on the spatial variables. By construction, each row in S sums to 1, S is lower triangular and has zeros on the diagonal.

The use of the individual neighbor matrices, S_l, greatly speeds up investigation of the sensitivity of the results to different forms of S. The individual neighbor matrices themselves require some expense in computation. However, reweighting these as in (5) requires very little time.

Similarly, for each row of T, we give weight $1/m_T$ to the m_T immediately prior observations.[6] We enforce examination of the past by only looking at the lower triangle of T (recall the observations are sorted by time), which means observations can be nonzero only if $i > j$:

$$i > j \geq (i - m_T) \leftrightarrow T_{ij} = \frac{1}{m_T}.$$

Note, the oldest observations correspond to the initial rows of T. Also, m_T can differ from m_S.

We form all the quantities (such as TY, SY, TX, SX, STX, TSX, STY, TSY) needed in computing the estimates for the entire sample. We subsequently drop some initial number of observations from these quantities for use in the actual estimation sample. We do this for two reasons. First, this allowed us to adjust m_T without confounding its effects by changing the size of the estimation sample (we examined the interval $0 < m_T < 1600$). Second, without retaining some prior observations, the spatiotemporal estimator could perform poorly initially as it would have a very small selection of previously sold neighbors to use in the first predictions. For the data described later, we drop 1600 of the initial observations, which represent around three years of market data.

2.3. Estimation

For the model in (4), the profile log-likelihood function in the parameters ϕ becomes

$$L(\phi) - \ln|I - \phi_S S - \phi_T T - \phi_{ST} ST - \phi_{TS} TS| - \left(\frac{n}{2}\right) \ln(SSE(\phi)).$$

By the triangular nature of T and S and given these matrices have zeros on the diagonals, the matrices TS and ST are also triangular and have zeros on the diagonals. Hence, the determinant of $(I - W)$ equals 1, and the log-determinant equals 0. Thus, maximizing the likelihood equates to minimizing the SSE via OLS.[7] Lagged dependent variables can create bias problems for OLS in small samples. However, OLS consistently estimates the parameters in large samples (provided the errors are not autocorrelated).

2.4. *Computational considerations*

Pace (1997) and Pace and Barry (1997a, 1997b) have detailed some of the computational advantages of sparse data structures in spatial statistics. Sparse matrices have large proportions of zeros, and special algorithms employ this fact to greatly accelerate computations and to save memory. The use of a fixed number of neighbors in time and space ensures the sparsity of S and T. S will generally have a density of m_S/n, where m_S is the number of neighbors in space (for example, closest fifteen observations in space). Similarly, T will have average density m_T/n, where m_T is the number of neighbors in time. Hence, as n rises, S and T become progressively sparser. This greatly aids the computational feasibility of the spatiotemporal model.

The strictly lower triangular nature of T and S avoids the need to compute the determinant of a general matrix. This greatly reduces the complexity of calculating the spatiotemporal estimates.

As implemented, the spatiotemporal model possesses another major computational advantage over a straightforward execution of the stated model. For a large, active housing market T could have potentially thousands of nonzero entries in each row. This would make T rather cumbersome and potentially prohibitive to manipulate. In the spatiotemporal model, however, T does not appear by itself but only in combination with other variables (TY, TX, TSX, TSY). Thus, each column of these contains the running averages of the respective variable over time. Efficient linear filter routines exist for computing such quantities. For example, the Matlab function "filter" implements this operation. The use of running sums and averages to handle large weighting matrices is analogous to the well-known Newey-West estimator that deals with heteroskedasticity and autocorrelation (Greene 1993, pp. 375–379). The estimator uses a weighted average of products and cross-products from an OLS regression.

3. An Application to the Fairfax Housing Data

This section presents the application of the spatiotemporal modeling to housing data from Fairfax County, Virginia. The first part discusses the data, the second part details a traditional model of the data, the third part examines the general spatiotemporal model, the fourth part examines the performance of a more parsimonious model, the fifth part examines the predictive performance of the models, while the last part describes an index surface.

3.1. Fairfax Data

We began with 73,835 observations on housing transactions over 1966 through 1991 from Fairfax County Virginia. We initially discarded observations with (1) fewer than four rooms, (2) no baths, (3) more than seven baths, (4) land area over 5 acres, (5) no land area, (6) more than three half-baths, (7) having a sales price of under $10,000 or over $1,000,000, and (8) sale dates outside of 1966 to 1992. This left 72,422 observations for the overall sample.

After computing Y, X, TY, TX, SY, STX, TSY, and so forth, we dropped the initial 1600 observations from each of these quantities. This corresponds to around three years of data. We did this because of the lack of good neighbors for the initial observations. We used the remaining 70,822 observations as the actual estimation sample.[8]

3.2. A Traditional Model

For comparison, we estimated a traditional OLS hedonic pricing model using indicator variables for the years 1969 to 1991. We used a double-log specification in Age, Land Area, Other Rooms, and Bathrooms, where Other Rooms equals the total number of rooms less Bathrooms. In all, this model employed twenty-six variables as shown in table 1.

The model has an R^2 of 0.7927, all variables have the expected signs, and all seem quite significant. The time indicator variables show the expected pattern of rising values (except for the last two years). The log-likelihood was $-305{,}552.13$.

The results from the traditional model approximately match those in the literature. For example, in their study of the value of a view amenity, Rodriguez and Sirmans (1994) controlled for location by focusing their analysis to a small subarea of Fairfax County. They report an adjusted R^2 of 0.729. In addition, Case, Pollakowski, and Wachter (1991) use the same data and one area in Fairfax County (Springfield) to investigate price indices. The largest reported R^2 on their data set of 14,617 observations equals 0.828, despite using 138 variables.

Even though the traditional model seems to perform similarly to other models in the liaterature, the residuals grossly violate independence as figure 1 illustrates. The pairwise correlation in neighboring residuals varies between 0.44 and 0.25. This obvious dependence suggests the potential for using this information to improve estimation, a tact followed in the next section.

3.3. A General Spatiotemporal Model

We estimated the model in (4) where Z contains two variables, an intercept and an index going from 1 to n and where X contains the variables log(Age), log(Land Area), log(Other Rooms), and log(Bathrooms) where Other Rooms equals Total Rooms less Bathrooms. The index should capture any deterministic trends relating to the sequence of the observations (the sequence is highly correlated with time). The general spatiotemporal

Table 1. OLS Estimates.

Variables	OLS Estimates	OLS *t* Ratios
1969	9.9981	598.1301
1970	10.0432	592.5722
1971	10.1202	639.5776
1972	10.2111	659.3801
1973	10.4019	664.8944
1974	10.5260	661.0442
1975	10.5755	693.7942
1976	10.5853	737.0386
1977	10.6973	763.6197
1978	10.8159	782.0870
1979	10.9324	786.3287
1980	11.0379	778.8624
1981	11.1263	779.4617
1982	11.1310	774.2501
1983	11.1692	838.0734
1984	11.2321	859.5112
1985	11.3089	881.4452
1986	11.4195	903.4025
1987	11.5693	917.3315
1988	11.7313	937.6756
1989	11.8265	941.4974
1990	11.8198	934.3043
1991	11.7789	934.4344
Log(Age)	−0.0678	−34.4498
Log(Land Area)	0.1499	124.3250
Log(Other Rooms)	0.2999	59.7846
Log(Bathrooms)	0.5196	142.3096
R^2	0.7927	
Log likelihood	305552.13	
n	70.822	
k	27	

model has twenty-eight parameters (twenty-six in the regression equation and the two nonlinear parameters λ, m_T).[9]

Table 2 shows the estimation results. The model has an R^2 of 0.8656, the variables log(Age), log(Land Area), log(Other Rooms), and log(Bathrooms) display the expected signs, and all of these as well as their spatially lagged versions are highly significant. Note, the extremely large t ratio of 174.52 for *SY*, the spatially lagged dependent variable. Based on approximate fitting, we selected the values of 0.75 for λ and 650 for m_T (corresponding roughly to two months in time).

The log-likelihood for the spatiotemporal model was −290,199.55, vastly larger than the one for the traditional model (a difference of 15,352.58).

Finally, the higher magnitude of coefficient estimate for the variable *STY* (−0.7902) relative to the variable *TSY* (−0.0324) suggests the need to filter first for time and subsequently for space and not vice versa or these data.

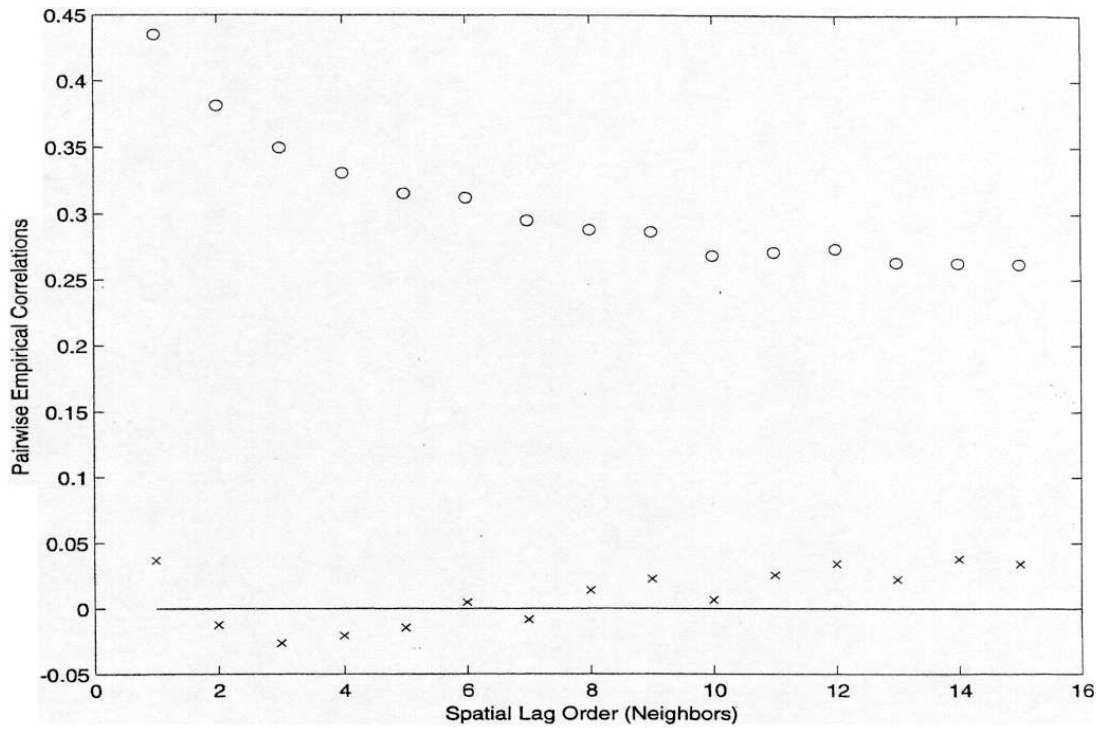

Figure 1. Pairwise correlations between sample and spatially lagged residuals (OLS = ○, spatial = ×).

While one could examine the meaning and interpretation of the estimated coefficients in more depth, the number of such terms makes this somewhat difficult. Before proceeding further in this direction, it seems worthwhile to look at a simpler model to see what performance penalty exists from reducing the dimensionality of the model.

3.4. A Parsimonious Spatiotemporal Model

In searching for a parsimonious alternative model, the relatively low significance of the intercept and the index in the general model suggests that first differences over time might perform well. Table 3 presents the estimates for a model using first differences in the variables log(Age), log(Land Area), log(Other Rooms) and log(Bathrooms) along with their spatial lags coupled with the spatially lagged, temporally differenced dependent variable $(I - T)Y$. Based on approximate fitting, we selected the values of 0.75 for λ and 650 for m_T (corresponding roughly to two months in time).

The parsimonious model uses twelve variables (ten in the regression equation and two entering into the construction of S and T) as opposed to twenty-eight for the general model.

Table 2. Spatial temporal estimates from the expanded model.

Variables	Expanded ST Estimate	Expanded ST t Ratio
Intercept	0.4764	1.3987
Index (1 : n)	0.0000	−3.6555
Log(Age)	−0.0921	−34.7458
Log(Land Area)	0.1297	84.3744
Log(Other Rooms)	0.2370	53.3393
Log(Bathrooms)	0.3342	94.3954
SLog(Age)	0.0928	26.6092
SLog(Land Area)	−0.0779	−38.1962
SLog(Other Rooms)	−0.1948	−24.5823
SLog(Bathrooms)	−0.1713	−27.3172
TLog(Age)	0.0525	1.1406
TLog(Land Area)	−0.1633	−5.0337
TLog(Other Rooms)	−0.3047	−2.9774
TLog(Bathrooms)	−0.3240	−4.3045
STLog(Age)	−0.2218	−6.9262
STLog(Land Area)	0.1921	5.7014
STLog(Other Rooms)	0.2728	1.7272
STLog(Bathrooms)	0.4158	3.5356
TSLog(Age)	0.0408	0.8056
TSLog(Land Area)	−0.1085	−3.2890
TSLog(Other Rooms)	0.2234	1.6377
TSLog(Bathrooms)	−0.4354	−3.7998
SY	0.8106	174.5224
TY	−0.0171	−0.9753
STY	−0.7902	−41.6045
TSY	−0.0324	−1.4950
λ	0.75	
m_T	650	
Log likelihood	−290,199.55	
n	70822	
k	28	

The resulting log-likelihood was −290,339.8. A statistically significant difference exists between the parsimonious and the general model based on a likelihood ratio test. However, the difference seems small given the number of observations and the number of variables.[10] Due to the relatively good performance of the parsimonious model, we adopted it for further examination.

3.5. Performance of the Models

Comparing the parsimonious spatiotemporal model to the traditional indicator-based model as in figure 1, we see the spatiotemporal model residuals show low pairwise

Table 3. Parsimonious spatiotemporal estimates.

Variables	Parsiminious ST Estimate	Parsimonious ST t Ratio
Intercept	0.0067	7.7117
$(I-T)$Log(Age)	−0.0920	−35.5535
$(I-T)$Log(Land Area)	0.1315	86.0632
$(I-T)$Log(Other Rooms)	0.2383	53.6362
$(I-T)$Log(Bathrooms)	0.3355	94.8489
$S(I-T)$Log(Age)	0.0892	25.7890
$S(I-T)$Log(Land Area)	−0.0800	−39.3564
$S(I-T)$Log(Other Rooms)	−0.1949	−24.5788
$S(I-T)$Log(Bathrooms)	−0.1732	−27.6067
$S(I-T)Y$	0.8082	174.0101
λ	0.75	
m_T	650	
Log likelihood	−290,339.8	
n	70822	
k	12	

correlations among spatial neighbors. Essentially, it has used the spatial information to improve the estimates.

Table 4 shows the empirical residual (sample) statistics for the traditional model and for the parsimonious spatiotemporal model. The spatiotemporal model shows large improvements over the traditional approach. For example, median absolute errors went from 0.1478 to 0.0926, a 37.35% decrease.

We also computed the one-step ahead forecast errors (recursive residuals) for the parsimonious spatiotemporal model.[11] As expected, the ex-sample spatiotemporal residuals are somewhat larger than the sample spatiotemporal residuals. However, the

Table 4. Residuals (e) across estimators.

	OLS	ST	Recursive ST
Min(e)	−2.7724	−2.9372	−2.4873
1st percentile	−1.0305	−0.8845	−0.7901
5th	−0.3554	−0.2607	−0.2818
10th	−0.2720	−0.1805	−0.1993
25th	−0.1476	−0.0902	−0.1023
50th	−0.0053	−0.0015	−0.0071
75th	0.1479	0.0954	0.1002
90th	0.3186	0.2127	0.2308
95th	0.4412	0.3161	0.3438
99th	0.6913	0.6121	0.6388
Max(e)	1.7365	1.6534	1.6963
Mean(e)	0.0000	0.0000	−0.0006
Median $\lvert e \rvert$	0.1478	0.0926	0.1014

ex-sample residuals from the parsimonious model have a median absolute value of only 0.1014, an increase of 9.50% over the sample residuals but still 31.39% less than the traditional model sample residuals.

Figure 2 depicts these recursive residuals ordered over Time, Age, Land Area, Other Rooms and Bathrooms. Under the null hypothesis of correct model specification, no discernible pattern should exist in any of these plots. In fact, the plots do not display any gross patterns. This adds to our confidence in the parsimonious spatiotemporal model.

3.6. *An Index Surface*

Due to the use of spatial and temporal lags, the estimated spatiotemporal models effectively create a spatial surface that evolves over time. Hence, at any point on the spatial surface we could separate out an index over time or for any given point in time we could separate out a spatial surface.

To examine the construction of a location specific index more closely, we write the equation for the estimated parsimonious spatiotemporal model at a given location:

$$\widehat{Y}_t = \text{Intercept} + (x_* - T_t X)\widehat{\beta}_1 + S_{location,t}(x_* - T_t X)\widehat{\beta}_2 + \widehat{\phi}_{ST} S_{location,t}(I - T_t)Y_t + T_t Y.$$

At a given location, the spatial weight matrix $S_{location;t}$ averages the values at nearby locations for previously sold properties for any desired point in time, t. As time progresses, the entries in the rows of $S_{location,t}$ change as nearby properties ("comparables") sell or until these comparables become too old (over sixty months in the past). The matrices, T_t contains nonzero entries for properties recently sold prior to time t. Hence, $T_t Y$ provides the average price of houses immediately prior to time t. Similarly $T_t X$ provides the typical sold house's characteristics immediately prior to time t. Finally, $x.$ represents the archetypal house associated with the index (in this case $x_* = \overline{X}$, the average of all X).

This index has several interpretable components. The term $T_t Y$ represents the simple average of housing prices over the recent past. The term $(x_* - T_t X)\widehat{\beta}_1$ represents the value of the difference between the typical house's characteristics and the recent average of house characteristics. The term $S_{location,t}(x_* - T_t X)\widehat{\beta}_2$ focuses on value of the difference between the typical house's characteristics and the recent average of house characteristics over the neighborhood. The term $\widehat{\phi}_{ST} S_{location,t}(I - T_t)Y_t$ adjusts for the changes in price of recently sold, nearby properties.

Figure 3 displays potential time indices created by taking six arbitrary locations from the Fairfax area and shows how a particular property would fair over time. In addition, the overall market index appears as well (as given by the symbol $\otimes$). As expected, some locations showed relative increases, some showed stable paths, while others varied greatly over time.

Naturally, the precision of estimation for these paths depend on the density of observations proximate in space and time. The estimation procedure optimized over the

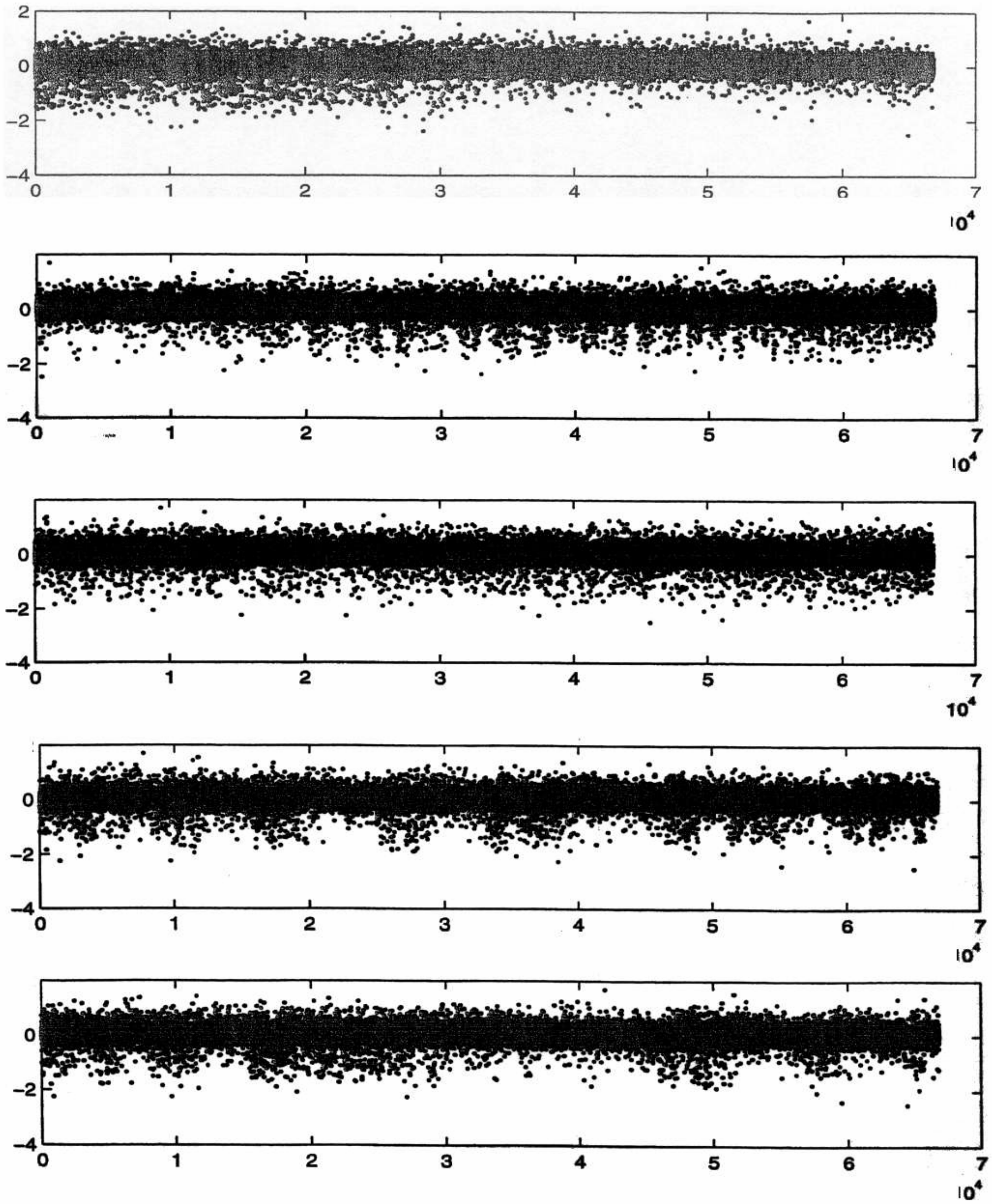

Figure 2. Recursive residual index plots for different orderings

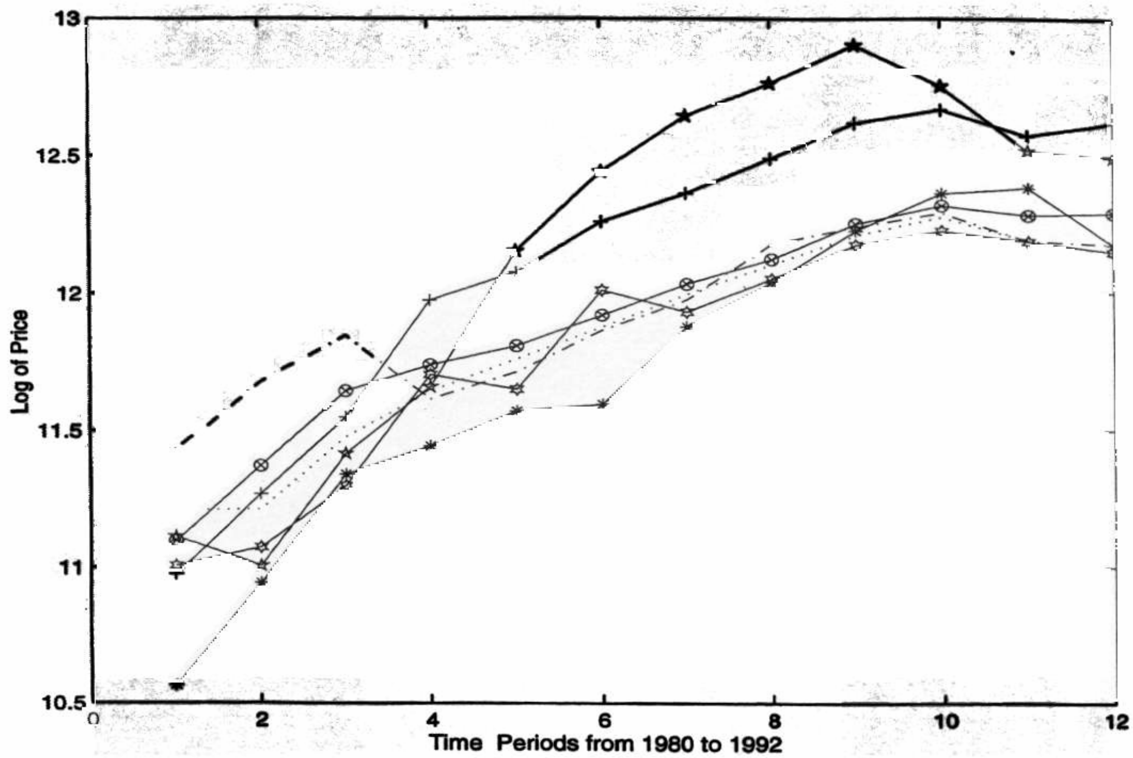

Figure 3. Individual location and overall market price indices.

number of neighbors considered. Hence, smoother indices over space (more spatial neighbors) would reduce variance at the cost of location-specific bias. The use of too many neighbors could raise mean-squared error. As a stylized fact, better specification of the regression equation often reduces the spatial correlations among the errors. Hence, better specifications would probably require fewer neighbors and result in even more location specific indices.

4. Conclusion

Real estate prices evolve over both time and space. Traditional practice involves regression of sales price on property characteristics and indicator variables from time and location (such as zip code indicators). Unfortunately, this typically requires the inclusion of more variables in the models than desired on the basis of parsimony to yield residuals without visible temporal and spatial dependencies.

This article follows a filtering or transformation approach to improve estimation. Specifically, one can filter the data for the temporal effects followed by filtering for the spatial effects, vice versa or combine the two approaches. Ideally, simple models applied to the filtered (transformed) data should not display the gross error dependencies found with the original data. The filtering approach leads to models with both temporal and spatial lags.

Applying the filtering approach to housing data from Fairfax County, Virginia, resulted in large improvements in estimation. Specifically, the R^2 went from 0.7927 using the traditional model to 0.8656 using the spatiotemporal model, the log-likelihood rose by 15,352.58, and the median absolute errors fell by 37.35%. These gains persisted even when using one-step ahead forecast errors as opposed to sample errors. Other models in the literature, such as Model L in Case, Pollakowski, and Wachter (1991, p. 301) display poorer goodness of fit ($R^2 = 0.828$) despite using more variables ($k = 138$).

The results show that the log of sales price of any property is strongly influenced by the sales prices of previously sold, neighboring properties. While this is not surprising, the extent of the neighborhood dependence is very high in Fairfax County. In fact 80.82% of the *temporal difference* in log sales prices propagates from the nearest fifteen neighbors (each weighted by a geometrically declining factor of 0.75) to the subject property. This suggests the short range spatiotemporal dependence is much greater than previously thought. Of course, the strong temporal dependence is entirely consistent with the large literature on ARIMA and related bodies.

The use of the spatial lags (based on previous sales) means that the relative premium or discount at any given location relative to the overall market changes over time depending on the sales of nearby properties. Hence, one can construct a temporal index for any given location. Also, one could construct a map of locational premia or discounts for any given time. In fact, the estimated model yields an evolving price surface.

Computationally, the spatiotemporal estimates use only the OLS estimator on suitably defined variables. Most of the effort comes from finding the spatiotemporal lags of the independent and dependent variables. However, the construction of these variables really just relies on the criteria for comparable selection employed for many years by appraisers. The application of these techniques to this very large data set shows the feasibility of such estimation: applications to smaller data sets should prove quite easy.

In conclusion, the large improvements to goodness of fit and the reduction in observed correlation among residuals of the spatiotemporal model relative to the traditional indicator-based model should earn it a place in the panoply of real estate statistical methods, especially since the computational difficulty of the spatiotemporal estimates does not appear large and the spatiotemporal model provides other benefits such as location-specific indices.

Acknowledgments

The authors gratefully acknowledge the research support they have received from their respective institutions and in particular wish to acknowledge support from the Center for Real Estate and Urban Studies at the University of Connecticut at Storrs. We also would like to especially thank Jennifer Pike for her editorial assistance.

Notes

1. This provides a sufficient condition for no autocorrelation in the errors.
2. Since we are premultiplying the random variable by $(I - W)$, the order of the filtering proceeds from right to left.
3. See Hendry, Pagan, and Sargan (1984) for details on this approach and for a variety of rich interpretations of this specification. See Anselin (1988, pp. 225–230) for a review of the spatial literature applying the general to specific approach.
4. Most appraisers select between three to ten comparable properties, depending on data availability. Ideally, these should come from the same or nearby, similar subdivisions. With respect to time, appraisers pay attention to recent trends in the overall market. The last transactions in the market provide a measure for such trends. The number considered will depend on the size of the market. For example, if a house sells each day, 100 comparables would correspond to between three to four months on average. Picking a fixed number will shorten the time interval as the market becomes more active and lengthen the time interval as the market becomes slower.
5. One can adopt different metrics such as the Euclidean, the Manhattan, or ones based on travel time as in Rodriguez, Sirmans and Marks (1995).
6. Based on some preliminary fitting, we found a constant weight for each past observation to perform acceptably and so did not examine extensively variable weights for T.
7. This ignores the effects of observations outside of the actual ones used to compute the estimates. Depending on the assumptions, proximate ex-sample observations in space and observations prior to the initial one in time could modify this determinant in which case minimizing the SSE above equates to maximizing a quasi-likelihood. Modeling jointly the problems of proximate observations in three dimensions (two for space and one for time) is quite challenging.
8. The problem of losing the first observations arises often in time series. The 1600 observations represents between two to three years of data. Naturally, we could have shortened this period. Substantially fewer observations would have resulted in more distant spatial neighbors (large numbers of prior observations leads to very close neighbors). This would have increased the error in the initial periods relative to latter periods.
9. We never really considered m_S other than fifteen and hence do not consider this a free parameter in this problem, although one could easily vary it. The parameter λ took care of the spatial weighting.
10. As a stylized fact, with enough observations almost all hypotheses become statistically significant. Various criteria such as the Akaike information criterion (AIC) or the Bayesian information criterion (BIC) penalize likelihood ratio tests to induce more parsimony. See Judge et al (1985, pp. 870–875) for more on these criteria.
11. See Judge et al (1985, pp. 173–174) for more on recursive residuals.

References

Anselin, Luc. (1988) *Spatial Econometrics: Methods and Models*. Dordrecht: Kluwer.

Anselin, Luc, and S. Hudak. (1992). "Spatial Econometrics in Practice: A Review of Software Options," *Journal of Regional Science and Urban Economics* 22, 509–536.

Case, Bradford, Henry O. Pollakowski, and Susan M. Wachter. (1991). "On Choosing Among House Price Index Methodologies," *AREUEA Journal* 19, 286–307.

Clapp, John M., Walter Dolde, and Dogan Tirtiroglu. (1995). "Imperfect Information and Investor Inferences from Housing Price Dynamics," *Real Estate Economics* 23, 239–269.

Cressie, Noel. (1993). *Spatial Statistics*. New York: Wiley.

Davidson, Russell, and James MacKinnon. (1993). *Estimation and Inference in Econometrics*. New York: Oxford University Press.

Deutsch, Stuart Jay, and Jose A. Ramos. (1986). "Space-Time Modeling of Vector Hydrologic Sequences," *Water Resources Bulletin* 22, 967–980.

Greene, William H. (1993). *Econometric Analysis* (2nd ed.). New York: Macmillan.

Hendry, David, Adrian Pagan, and Denis Sargan. (1984). "Dynamic Specification" in Z. Griliches and M. Intrilligator (eds.), *Handbook of Econometrics* (pp. 1023–1100). Amsterdam: North-Holland.

Judge, George, W. E. Griffiths, R. Carter Hill, Helmut Lütkepohl, and Tsoung-Chou Lee. (1985). *The Theory and Practice of Econometrics* (2nd ed.). New York: John Wiley.

Ord, Keith. (1975). "Estimation Methods for Models of Spatial Interaction," *Journal of the American Statistical Association* 70, 120-126.

Pace, R. Kelley. (1997). "Performing Large-Scale Spatial Autoregressions," *Economics Letters* 54, 283–291.

Pace, R. Kelley, and Ronald Barry. (1997a). "Sparse Spatial Autoregressions," *Statistics and Probability Letters* 33, 291–297.

Pace, R. Kelley, and Ronald Barry. (1997b). "Quick Computation of Regressions with a Spatially Autoregressive Dependent Variable," *Geographical Analysis* 29, 232–247.

Pace, R. Kelley, and O. W. Gilley. (1998). "Optimally Combining OLS and the Grid Estimator," *Real Estate Economics* 26.

Pfeifer, Phillip E., and Samuel E. Bodily. (1990). " A Test for Space-Time ARMA Modeling and Forecasting of Hotel Data," *Journal of Forecasting* 9, 255–272.

Rodriguez, Mauricio, and C. F. Sirmans. (1994). "Quantifying the Value of a View in Single-Family Housing Markets," *The Appraisal Journal* 62, 600-603.

Rodriguez, Mauricio, C. F. Sirmans, and Allen P. Marks. (1995). "Using Geographical Information Systems to Improve Real Estate Analysis," *Journal of the Real Estate Finance and Economics* 10, 163–173.

Silverman, B. W. (1986). *Density Estimation for Statistics and Data Analysis*. London: Chapman and Hall.

Szummer, Martin, and Rosalind W. Picard. (1996). "Temporal Texture Modeling," Manuscript, MIT Media Lab.