International Journal of Forecasting 16 (2000) 229–246

international journal of forecasting

www.elsevier.com/locate/ijforecast

A method for spatial–temporal forecasting with an application to real estate prices

R. Kelley Pace[a,*], Ronald Barry[b], Otis W. Gilley[c], C.F. Sirmans[d]

[a]E.J. Ourso College of Business Administration, Louisiana State University, Baton Rouge, LA 70803, USA
[b]Department of Mathematical Sciences, University of Alaska, Fairbanks, AK 99775-6660, USA
[c]Department of Economics and Finance, College of Administration and Business, Louisiana Tech University, Ruston, LA 71272, USA
[d]Center for Real Estate and Urban Studies, 368 Fairfield Rd., Rm 426, U-41RE, Storrs, CT 06269-2041, USA

Abstract

Using 5243 housing price observations during 1984–92 from Baton Rouge, this manuscript demonstrates the substantial benefits obtained by modeling the spatial as well as the temporal dependence of the errors. Specifically, the spatial–temporal autoregression with 14 variables produced 46.9% less SSE than a 12-variable regression using simple indicator variables for time. More impressively, the spatial–temporal regression with 14 variables displayed 8% lower SSE than a regression using 211 variables attempting to control for the housing characteristics, time, and space via continuous and indicator variables. One-step ahead forecasts document the utility of the proposed spatial–temporal model. In addition, the manuscript illustrates techniques for rapidly computing the estimates based upon an interesting decomposition for modeling spatial and temporal effects. The decomposition maximizes the use of sparsity in some of the matrices and consequently accelerates computations. In fact, the model uses the frequent transactions in the housing market to help simplify computations. The techniques employed also have applications to other dimensions and metrics.

Keywords: Real estate; Spatial–temporal forecasting; Modeling

1. Introduction

Data sets have often been organized by units of time such as quarters or years as well as by geographical constructs such as regions, states, or counties. In reality, the data often represent an aggregation of individual observations which have more precise temporal and spatial characteristics. Much of the governmentally collected economic, medical, and social data falls into this category. However, the increasing capabilities of information systems and especially geographic information systems (GIS) have greatly aided work with disaggregated data having precise spatial and temporal references. For example, automated point-of-sale data from individual stores have a precise identification in

*Corresponding author. Tel.: +1-225-388-6238; fax: +1-225-334-1227.

E-mail addresses: kpace@lsu.edu (R.K. Pace), ffrpb@aurora.alaska.edu (R. Barry), gilley@cab.latech.edu (O.W. Gilley), fnceadm5@uconnvm.uconn.edu (C.F. Sirmans)

0169-2070/00/$ – see front matter
PII: S0169-2070(99)00047-3

time and space. Even governmentally collected data, where privacy concerns dictate minimum levels of aggregation, have become more precise in their identification of space or time. For example, the Home Mortgage Disclosure Act (HMDA) data for 1993 contained approximately 15 million individual transactions identified by over 60 000 locations (census tracts).

The trend towards large data sets with substantial spatial and temporal detail raises the issue of how to forecast such data. Moreover, such data raise computational issues as well as conceptual issues of how to plausibly model the spatial and temporal dependence.

Housing prices provide another example of this type of data. First, over 4.5 million houses sold during 1997 alone. Most data sources (multiple listing services or assessor databases) record the day, month, and year of the transaction. Given an address, geographic information systems can provide the corresponding latitude and longitude (or other locational coordinates) for 80% or more of the records (Johnson, 1998).

At least five common applications employ housing transaction data. First, most houses in the US (and many in other countries) have their assessed value for tax purposes determined by the predictions from statistical models calibrated using individual housing transactions (Eckert, 1990, p. 27). Second, the movement of many primary and secondary lenders to some form of automated appraisal places an added premium on prediction accuracy (Gelfand, Ghosh, Knight & Sirmans, 1998). Third, the spatial and product differentiated nature of housing makes it difficult to compare prices across time and space. The desire to make such comparisons has spurred substantial activity in creating constant quality price indices by location (Hill, Knight & Sirmans, 1997). Fourth, hedonic pricing models use housing data to estimate the costs and benefits associated with such items as pollution, growth controls, and tax policies (Case, Rosen & Hines, 1993; Brueckner, 1997). Fifth, as a house comprises a large fraction of an individual's wealth, a number of parties follow local price forecasts. Collectively, these applications involve the goals of accurate prediction, efficient coefficient estimation, valid inference, and the desire to understand both the temporal and spatial dependencies in prices.

For housing data, the benefits from modeling the error dependence over time are well-known and the benefits from modeling the error dependence over space are becoming better known.[1] Such benefits include more efficient asymptotic parameter estimation, less biased inference (positive correlations among errors artificially inflate t-statistics), and more precise predictions.

Joint modeling of errors in both time and space offers the potential for further gains.[2] However, these techniques have not been applied extensively to economic data and the best ways of specifying spatial, temporal, and spatial–temporal interactions do not appear obvi-

[1] Hill, Knight and Sirmans (1997) found annual AR(1) estimates on annual data for the autoregressive parameter of 0.54. Pace and Barry (1997a) found estimates of the spatial autoregressive parameter of 0.8536 using row-standardized weight matrices while Dubin (1988) found houses 0.5 miles apart exhibited a correlation among errors of 0.58.

[2] We have not found many applications of joint modeling of the spatial and temporal errors to housing data, although various approaches to this have been taken in different areas. For example, Pfeifer and Bodily (1990) jointly model errors to aid in the prediction of revenue for a small sample of hotels. In fact, Pfeifer and Bodily claim theirs was the only application of space–time autoregressive moving average techniques to any kind of business data (we cannot find other studies involving actual data either). Although, as they point out, these have been applied often in the physical sciences. For example, Szummer and Picard (1996) use a space–time autoregressive (STAR) model to aid in the synthesis of images of phenomenon such as moving water, fire, or other evolving textures. Deutsch and Ramos (1986) use these techniques to examine river flows. Pace, Barry, Clapp, and Rodriguez (1998) did apply a simplified version of the model developed here to a dataset from Fairfax, Virginia with over 70 000 observations. They also found great improvements from using spatial–temporal information.

ous. As Gelfand et al. (1998) state, 'Spatial–temporal interaction presents a difficult modeling challenge since it is unclear how to reconcile the two scales.' Consequently, Gelfand et al. used time indicators which made one-step ahead forecasting difficult.[3]

Ideally, the joint specification of the errors in both space and time would perform well and provide an interpretable framework for viewing the error dependence. We approach this specification question from a compound filtering perspective. Variables filtered first for time and subsequently for space could display less error dependence in either time or space. The same, however, could apply to variables filtered first for space and subsequently for time. We allow the model to nest both of these alternatives. We find the optimum is a convex combination of the two compound filters. This provides an interpretable error dependency structure which also allows for possible spatial–temporal interactions.

To demonstrate the potential of this means of joint modeling the spatial–temporal errors, we pit such a model against the opposite extreme of a model attempting to handle the effects of space and time through an extensive set of indicator variables. Both models share a common set of housing variables and data of 5243 observations on housing prices during 1985–92 from Baton Rouge, Louisiana. These data include each house's location in latitude and longitude as well as the day, month, and year the transaction closed.

In this contest the spatial–temporal model with 14 variables outperforms by 8% in terms of SSE the traditional indicator based model with 211 variables. Moreover, it maintains its accuracy in one period ex-sample forecasts.

The spatial–temporal model relies upon a rich set of spatially and temporally lagged variables for its power. It represents a hybrid between the autoregressive distributed lag model common in time series and the mixed regressive spatially autoregressive model in spatial econometrics (Ord, 1975; Anselin, 1988, p. 227). The techniques tend to support some of the procedures employed by appraisers who examine a limited set of 'comparables' in estimating the value of a house.

In addition, we introduce techniques for rapidly computing the estimates based upon a novel decomposition of the spatial and temporal effects. The decomposition maximizes the use of sparsity (proportion of zeros) in some of the matrices and consequently greatly accelerates computations. In addition, we structure the data in a way which makes it feasible to use optimized linear filter routines to simplify some of the problems of dealing with irregularly spaced data over time. Finally, the model takes advantage of the frequent transactions in the housing market to eliminate the awkward normalizing constant in the likelihood which has traditionally impeded spatial error modeling. The culmination of these improvements allows us to compute individual spatial–temporal regressions in a few seconds, thus making these practical for applied work.

Section 2 discusses the spatial–temporal model employed and provides details on an improved algorithm for computing the spatial–temporal estimates; Section 3 applies the spatial–temporal model to the Baton Rouge data, while Section 4 concludes with the key results.

2. A simple spatial–temporal model

This section describes a simple spatial–temporal model. The Section 2.1. begins with the traditional autoregressive error model, Section

[3]For example, Gelfand et al. (1998) state, 'It is not apparent how to model temporal association through the covariance matrix in such a case. Instead, we introduce usual year dummy variables to reveal time trends.' In addition they state, 'Since our cross-sectional data lacks customary time series structure we can not perform usual one-step ahead forecasting.'

2.2. describes the innovations which make this practical, Section 2.3. describes the spatial–temporal and temporal weight matrices, Section 2.4. discusses a smoothing interpretation of the model.

2.1. Spatial–temporal autoregression in errors

We wish to use a model which can, at a minimum, subsume an autoregressive error process:

$$\begin{aligned} Y &= X\beta + u \\ u &= Wu + \epsilon \end{aligned} \tag{1}$$

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 the k by 1 vectors of parameters, ϵ denotes an n by 1 vector of normal *iid* errors, u denotes the autocorrelated errors, and W denotes an n by n spatial–temporal weight matrix. We assume the observations have been ordered according to time, with the oldest observation corresponding to the first row of X and the most recent observation corresponding to the nth row of X. We can rewrite as:

$$(I - W)Y = (I - W)X\beta + \epsilon \tag{2}$$

For (2) generalized least squares and maximum likelihood provide the canonical estimators:

$$\beta_{\text{gls}} = (X'\psi^{-1}X)^{-1}X'\psi^{-1}Y \tag{3}$$

where $\psi = ((I - W)'(I - W))^{-1}$. In addition to computing the GLS estimates (conditional upon specification of W), maximum likelihood also involves computing the determinants of the n by n matrices ψ, ψ^{-1}, or $\psi^{-1/2}$. Assuming a real, finite $\ln|(I - W)|$, the concentrated log-likelihood function becomes:

$$L(W) = \ln|(I - W)| - \left(\frac{n}{2}\right)\ln(SSE(W)) \tag{4}$$

where $SSE(W) = (Y - X\beta)'\psi^{-1}(Y - X\beta) = ((I - W)(Y - X\beta))'((I - W)(Y - X\beta))$.[4]

Both statistical and computational problems arise in the specification of ψ, ψ^{-1}, or $\psi^{-1/2}$. As n becomes large, it becomes very difficult to use these matrices if all the elements are non-zero or dense. For the data studied later, a dense matrix formulation would take over 250 megabytes of storage space. This suggests searching for ways of specifying the problem to maximize the sparseness (proportion of zeros) of the matrices ψ, ψ^{-1}, or $\psi^{-1/2}$. If one of these matrices is sparse, the others typically are more dense.

For pure spatial problems, one can model W as a general sparse matrix. As Pace and Barry (1997a, b) have shown, this greatly facilitates computing the maximum likelihood estimates and conducting inference.

For pure temporal problems such as an AR(p) process with periodic observations, one can model W as a band matrix with p diagonals. Equivalently, one can difference each observation with a weighted sum of the p prior observations (subject to some separate treatments of the initial p observations). A great advantage of the strict temporal ordering is that the temporal part of ψ^{-1} is lower triangular and hence $\ln|\psi^{-1}| = \Sigma \ln(\psi_{ii}^{-1})$ which requires little computation.

Unfortunately, housing transactions occur at irregular intervals. One can follow two routes to address this facet of the data. First, one could model the temporal part of W as a band matrix. As a large market may have thousands of sales per year, modeling the temporal dependence over even one quarter could also set up the equivalent of an very high order autoregressive process if one applied the usual methods. This could lead to computational complications as

[4]Since $(\frac{1}{2})\ln|\psi| = -(\frac{1}{2})\ln|\psi^{-1}| = -\ln|\psi^{-1/2}|$, several variants of the log-likelihood function appear in the spatial literature. See Anselin (1988, p. 110) for a development of this log-likelihood.

the temporal part of W would have potentially hundreds or thousands of non-zero entries for each of the n rows.

Second, one could aggregate the data over some fixed period such as a quarter or a year and apply standard time series analysis techniques. At this point two additional choices emerge. If one uses spatial information strictly from the previous periods, this keeps the simple structure of ψ^{-1}but ignores the contemporary spatial correlation of errors associated with the houses selling in that period. See Cressie (1993, p. 450) for a more detailed criticism of ignoring the contemporaneous spatial correlation. Alternatively, one can take this into account. However, ψ^{-1} no longer would have the lower triangular form but would contain blocks along the diagonals. Hence, this requires more advanced techniques for efficient computation as laid out in Pace and Barry (1997a, b). Moreover, aggregation carries its own costs. Also, several of the idiosyncratic features of spatial statistics creep in the analysis. For example, simultaneous and conditional autoregressions have different forms and interpretations.

To acquire the advantage of strict lower triangular specification of W along with efficient computation, the next section introduces a specification for W which can achieve both ends. In addition, we introduce a more general specification than the autoregressive errors specification described above.

2.2. The spatial–temporal linear model

We begin by partitioning W into a matrix S which specifies spatial relations among previous observations and T which specifies temporal relations among previous observations. Hence, the matrix T fulfills a similar role in time with irregular periodicity as a lag operator does for time with constant periodicity. Also, S functions in space much like a lag operator does in time. The requirement of specifying relations among previous observations only and the ordering of the observations by time means W is a lower triangular matrix (barring ties). We weight the matrices in time and space by the autoregressive parameters $\phi = [\phi_S, \phi_T, \phi_{ST}, \phi_{TS}]$.

$$W = \phi_S S + \phi_T T + \phi_{ST} ST + \phi_{TS} TS \tag{5}$$

Note, this specification for W could subsume, among others, the following forms (where ω represents an arbitrary constant).

$$\begin{array}{llll} 1. & (I-W) & = & (I-\phi_S S-\phi_T T) \\ 2. & (I-W) & = & (I-\phi_S S)(I-\phi_T T) \\ 3. & (I-W) & = & (I-\phi_T T)(I-\phi_S S) \\ 4. & (I-W) & = & \omega(I-\phi_T T)(I-\phi_S S)+(1-\omega)(I-\phi_S S)(I-\phi_T T) \end{array} \tag{6}$$

The first form adjusts additively for the effects of time and space. The second form filters a variable first for temporal effects and subsequently for spatial effects.[5] The third form reverses this and the fourth form looks at a linear combination of the different types of compound filtering.

To completely generalize (1) and (5), we write the spatial–temporal linear model (STLM) as:

$$\begin{aligned} Y = Z\theta &+ X\beta_1 + TX\beta_2 + SX\beta_3 + STX\beta_4 \\ &+ TSX\beta_5 + \phi_T TY + \phi_S SY + \phi_{ST} STY \\ &+ \phi_{TS} TSY + \epsilon \end{aligned} \tag{7}$$

where Z denotes the n by p_1 matrix of observations on independent variables that do not have associated spatial, temporal or spatial–temporal lags, θ denotes the associated p_1 by 1 vector of parameters, X denotes the n by p_2 matrix of observations on the independent variables with spatial, temporal or spatial–temporal lags, and $\beta_1, \beta_2, \beta_3, \beta_4, \beta_5$ denote the p_2 by 1 vectors of parameters associated with the spatial, temporal, and spatial–temporal lagged variables. Effectively (7) synthesizes the autoregressive distributed lag model of time series and the 'mixed

[5]Since W premultiplies the variable of interest (e.g. WY, WX), the order of filtering proceeds from right to left.

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 $\phi, \theta, \beta_2, \beta_3, \beta_4, \beta_5 = 0$ 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). In addition, the restrictions $\phi = 0$ produce a model where only the various lags of the independent variables matter while the restrictions $\beta_2, \beta_3, \beta_4, \beta_5 = 0$ produce a model where only the various lags of the dependent variable matter.

If one believes in the general to specific model strategy, one should estimate the general model and test these restrictions before adopting a more specific alternative such as the autoregression in errors specification. 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.

The above model explains the dependent variables as a function of the independent variables, the temporal effects which apply to all observations, and the effects which arise due to proximity in both time and space. Separating out the aggregate temporal and spatial–temporal effects allows for flexible specification of the global effects in time via T versus the local effects in space and time via S. Incorporating the interaction terms ST and TS allows for the modeling of potential compound spatial–temporal effects.

We restrict the rows of S and T to sum to 1 (row stochastic). In the spatial econometrics literature such weighting matrices are said to be 'standardized' (Anselin & Hudak, 1992, p. 514). One can also interpret S and T as linear filters (Davidson & MacKinnon, 1993, p. 691). The non-zero entries on the rows of S and T represent the observations which directly interact with the observation subject to ruling out non-zero entries for the diagonal of S and T. Since the observations are temporally ordered and since we condition only upon previous transactions, both S and T are lower triangular ($j \geq i \leftrightarrow S_{ij} = 0$ and $T_{ij} = 0$).

After allowing for the lagged spatial effects in the dependent and independent variables, the use of maximum likelihood on the overall model assumes the true errors, ϵ, are independently and normally distributed. We summarize all of these assumptions in (8):

$$
\begin{aligned}
&\text{(a)}\ \underset{(n\text{ by }n)}{S}\ \underset{(n\text{ by }1)}{[1]} = \underset{(n\text{ by }1)}{[1]}\,,\ \underset{(n\text{ by }n)}{T}\ \underset{(n\text{ by }1)}{[1]} = \underset{(n\text{ by }1)}{[1]}\\
&\text{(b)}\ j \geq i \leftrightarrow S_{ij} = 0 \text{ and } T_{ij} = 0\\
&\text{(c)}\ -1 < \phi < 1\\
&\text{(d)}\ S_{ij} \geq 0, T_{ij} \geq 0\\
&\text{(e)}\ \epsilon \sim N(0, \sigma^2 I) \qquad (8)
\end{aligned}
$$

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.[6]

[6]This ignores the effects of observations outside of the actual ones used to compute the estimates. Depending upon the assumptions, ex-sample observations immediately prior to the initial one in time could modify this determinant in which case minimizing the SSE above equates to maximizing a quasi-likelihood.

Avoiding the necessity of computing the determinant of a general matrix greatly reduces the complexity of calculating the estimates.[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).

We impose some additional structure on S and T. Housing data displays uneven density in both the temporal and spatial dimensions. A distance of a 100 meters in the center of the city does not act the same as a distance of 100 meters in the suburbs. Similarly, housing markets go though active and thin periods. 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.[8] The use of ordinal distance and time acts to reduce the problems posed by uneven data densities.[9] For example, when the density of observations over time or space is very high, T or S averages over a short interval or area. Similarly, when observations occur infrequently, T or S may average over a long interval or large area. Nearest neighbors have often been employed in density estimation as variable bandwidth smoothers (Silverman, 1986).

The use of a fixed number of neighbors in time and space insures the sparsity of S and T. The matrix S will generally have a density of m_S/n, where m_S is the number of neighbors in space (e.g. closest 15 observations in space). However, T will have average density m_T/n, where m_T is the number of neighbors in time (e.g. previous 190 observations). Hence, as n rises, S and T become progressively sparser. This greatly aids the computational feasibility of the STLM.[10]

The STLM possesses a final major computational advantage. As mentioned earlier, for a large, active housing market T could have potentially thousands of non-zero entries in each row. This would make T rather cumbersome and potentially prohibitive to manipulate. In the

[7] In the pure spatial arena, the problems posed by the Jacobian have led to several ways of attacking this problem. First, Pace and Barry (1997a, b) use the potential sparsity of the matrix S to reduce the cost of computing the Jacobian. Kelejian and Prucha (1997) demonstrate how to estimate this by GMM, an approach which shows substantial promise when dealing with additional problems. Gelfand et al. (1998) used the Gibbs sampler to address this issue.

[8] Appraisers use the 'grid adjustment' method for estimating house prices. Pace and Gilley (1997, 1998) have shown this method has a restricted spatial autoregression representation. Appraisers use 'comparable' (i.e. neighbors) properties in their estimation of house prices and have extensively discussed the optimal number needed. Ratcliffe (1972, p. 154), for example, claims that '. . . less than 5 comparables is unsafe and more than 10 comparables is probably not necessary for purposes of home appraisal.' Furthermore, appraisers generally do not pick a fixed time span for examining past transactions. For example, they prefer to use a shorter period when conditions are more volatile (Ratcliffe, 1972, p. 153).

[9] The use of nearest neighbors is well-established in spatial statistics and econometrics. Using cardinal distances sometimes leads to pathological results. In some prior work involving another data set, using cardinal distances resulted in some observations being differenced while others stayed in levels. This created such leverage that the regression line was forced between the centroids of the data in levels and the centroid of the data in differences. Similarly, housing markets occasionally go for some time between transactions (depending upon the size of the market). This can also result in the problem of mixing differences and levels for short lag periods.

[10] A number of other minor computational advantages exist as well. For example, a priori knowledge of the exact number of non-zero elements allows pre-allocation of storage, a real advantage with interpreted languages like Matlab and Gauss.

STLM, 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.[11]

2.3. Further specification of the spatial and temporal weight matrices

For S, we initially compute the Euclidean distance d_{ij} between every pair of observations j and i for every prior observation ($j<i$). We subsequently sorted these distances and formed the set of individual neighbor matrices $S_1, S_2, \ldots, S_{15}$, 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. These very sparse matrices have a 1 in each row and contain 0s otherwise. One can also restrict the degree to which one goes into the past to find neighbors. Based on preliminary fitting, we decided to restrict the construction of S to 15 neighbors ($m_S=15$).

We computed the overall spatial matrix S via:

$$S = \frac{\sum_{l=1}^{m_S} \lambda^l S_l}{\sum_{l=1}^{m_S} \lambda^l} \tag{9}$$

where λ^l weights the relative effect of the lth individual neighbor matrix. Hence, S depends upon the parameters λ, m_S in its construction (as well as the metric). Thus, (9) 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 (the set of 15 here took 41.1 min in an unoptimized routine). However, reweighting these as in (9) requires very little time.

For example, suppose 5243 observations exist, $m_S=4$, and $\lambda=1$. For the 11th observation, S might appear as:

$$S_{11,j} = [0, 0.25, 0, 0, 0.25, 0, 0, 0, \\ 0.25, 0.25, 0, 0, 0, \ldots, 0]$$

Note, the 11th and higher entries of $S_{11,j}$ equal 0 while the row sums to 1.

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

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

Naturally, m_T can differ from m_S. Based upon some preliminary fitting, we found a constant weight for each past observation to perform acceptably and so did not examine extensively weighting elements of T in a more complicated fashion.[12]

As mentioned previously, we do not actually compute T. Rather we use linear filter routines to compute the product of T and the other variables. This avoids any need to store T,

[11] Appropriately, the Matlab command for this is named 'filter.' Such general purpose commands usually exist for fixed length averages, an additional reason to treat irregularly separated observations over time as neighbors.

[12] For a Markov process, the immediately previous sale contains the history of the process in the absence of noise. With noise, we trade noise reduction or variance (larger m_T) against bias (smaller m_T).

which could be huge for large, active markets (e.g. a market with hundreds of sales per month could have hundreds of non-zero entries in each row). Such linear filter routines have been extensively optimized and find application in many other areas.

We retained 300 observations prior to the sample to use in *TY*, *SY*, and so on. This allowed us to adjust m_T without confounding its effects by changing the sample. In addition, it allowed S to use obtain sufficient nearby observations for neighbors prior to its use in estimation. Without retaining some prior observations, the spatial–temporal estimator could perform poorly initially as it would have a very small selection of previously sold neighbors to use in the first predictions. For these data, 300 observations represents around six months in the market.

2.4. A smoothing interpretation

Traditional hedonic pricing models rely only upon the independent variables' parameter estimates for prediction. The STLM employs several other means. First, the variable *SY* provides the spatial nearest neighbor non-parametric estimate of the value of Y for the observation.[13] Second, the variable *TY* provides the temporal nearest neighbor non-parametric estimate of the value of Y for the observation. By extension, *SX* and *TX* provide spatial and temporal nearest neighbor averages for the independent variables. Combining these together allows the model to compute the spatial and temporal nearest neighbor estimates of the deviations $(Y - X\beta_1)$ from the simple parametric model $(X\beta_1)$ as well as other model variants.

Intuitively, the STLM parametrically adapts among (1) the parametric estimates of the nuisance parameters θ; (2) the parametric estimates of the independent variables; (3) a spatial nearest neighbor average of the dependent variable (with time constraints); (4) a temporal nearest neighbor average of the dependent variable; (5) a spatial nearest neighbor average of $Y - X\beta_1$ (with time constraints); and (6) a temporal nearest neighbor average of $Y - X\beta_1$. The mixture of non-parametric and parametric estimates gives the STLM a semiparametric flavor.

[13]Strictly speaking, it provides a spatial–temporal estimate as it uses prior, proximate observations in space to estimate Y. However, as its role is more spatial than temporal, we emphasize the spatial aspect.

3. An application to housing in Baton Rouge

This section illustrates the spatial–temporal model from Section 2 using the Baton Rouge data. Section 3.1. discusses the data, Section 3.2. presents the model, Section 3.3. provides the conventional trend surface model results, Section 3.4. presents the spatial–temporal linear model estimation results, Section 3.5. compares the performance of the non-spatial, trend surface, and spatial–temporal models, while Section 3.6. examines the sensitivity of the regression to different variants of the spatial and temporal weighting matrices.

3.1. Data

We selected observations from the Baton Rouge Multiple Listing Service which (1) could be geocoded; (2) had complete information on age, living area, other area, and number of baths (we also discarded negative entries); (3) had central climate control; (4) were of the most common construction type; (5) were of the most common occupancy type; (6) were of the most common zoning types; (7) had between 850 and 6000 total square feet in area; (8) had more than 500 square feet of living area; and (9) had a list price of \$20 000 or greater. Sales price was the dependent variable and ranged from $13 500 to $421 875. Note, the last filter did not reject many properties. It did, however, drop an observation for a $4000 three bedroom house. Such extremely low reported transaction prices could be the result of non-arm's length dealings or from simple clerical errors. As the intent for this exercise was to select a data set which would not cause a domination of the results by data quality issues, we used all of these filters. In total, 5543 observations satisfied these joint criteria. We retained 300 observations for prior calibration of S and T. We based all estimates on 5243 observations.

In the geocoding, we projected the decimal latitude and longitude data on each house using the 1702 Louisiana South state plane projection based upon the Clarke 1866 spheroid, and the NAD 27 datum. The end result was variables measuring the N–S and E–W location of each house in feet from a reference point.

3.2. Model

We begin by defining the following matrices:

$$X = [\ln(\text{Age}+1)\ \ln(\text{Living Area})\ \ln(\text{Other Area})\ \text{Baths}]$$
$$Z = [\text{index N–S E–W}]$$
$$Y = \ln(\text{Sales Price})$$
$$I_T = [I_{1986}\, I_{1987} \cdots I_{1992}]$$
$$I_S = [I_1\, I_2 \cdots I_{199}]$$

Where I denotes the relevant indicator variables. Given these definitions, the simple 12 variable non-spatial model 1, the more complex 211 variable trend surface model 2, and the spatial–temporal models 3–5 appear below:

$$Y = \text{intercept} + I_T\hat{\phi}_{1986-1992} + X\hat{\beta} + \epsilon \qquad \text{Model 1}$$

$$Y = \text{intercept} + I_S\hat{\phi}_{1-199} + I_T\hat{\phi}_{1986-1992} + X\hat{\beta} + \epsilon \qquad \text{Model 2}$$

$$\begin{aligned} Y - TY = {} & \text{intercept} + Z\tilde{\theta} + (X - TX)\tilde{\beta}_1 \\ & + S(X - TX)\tilde{\beta}_2 + TX\tilde{\beta}_3 \\ & + STX\tilde{\beta}_4 + TSX\tilde{\beta}_5 + \tilde{\phi}_S SY \\ & + \tilde{\phi}_T TY + \tilde{\phi}_{ST} STY + \tilde{\phi}_{TS} TSY \\ & + \epsilon \end{aligned} \qquad \text{Model 3}$$

$$\begin{aligned} Y - TY = {} & \text{intercept} + Z_2\tilde{\theta}_2 + Z_3\tilde{\theta}_3 \\ & + (X - TX)\tilde{\beta}_1 + S(X - TX)\tilde{\beta}_2 \\ & + \tilde{\phi}_S SY + \tilde{\phi}_{ST} STY \\ & + \tilde{\phi}_{TS} TSY + \epsilon \end{aligned} \qquad \text{Model 4}$$

$$\begin{aligned} Y - TY = {} & \text{intercept} + Z_2\tilde{\theta}_2 + Z_3\tilde{\theta}_3 \\ & + (X - TX)\tilde{\beta}_1 + S(X - TX)\tilde{\beta}_2 \\ & + \tilde{\phi}_S SY + \tilde{\phi}_{ST} STY + \tilde{\phi}_{TS} TSY \\ & + \epsilon \ \text{st.}\ \tilde{\phi}_S \\ = {} & -(\tilde{\phi}_{ST} + \tilde{\phi}_{TS}) \end{aligned} \qquad \text{Model 5}$$

3.3. Conventional estimation results

Table 1 contains the OLS estimates for model 1, the non-spatial model of Baton Rouge housing prices, and for model 2, an elaborate trend surface model. Model 1 provides a very simple model of housing prices over time in Baton Rouge. By conventional standards, it performs reasonably well. It shows the pattern of estimates for the annual indicator variables which accords well with other examinations of these data (Knight, Sirmans & Turnbull, 1994). The signs on the housing characteristics match prior expectations and these all seem quite significant.[14]

However, as Fig. 1 illustrates, the errors show very high levels of correlations with their individual neighbors (the correlation with a linear combination of the neighbors could naturally go higher). This high level of positive

[14] See Gilley and Pace (1995) for a further discussion of such prior expectations.

Table 1
Non-spatial and spatial trend surface estimates

Variables	Model 1 12 variable non-spatial		Model 2 211 variable trend surface	
	Estimates	t	Estimates 199 Spatial Indicators (not reported)	t
Intercept	3.0686	33.3729	4.8103	41.7600
1986	−0.0251	−2.1483	−0.0258	−2.7888
1987	−0.0731	−6.5607	−0.0751	−8.5222
1988	−0.1470	−12.1105	−0.1469	−15.2763
1989	−0.1200	−10.2652	−0.1216	−13.1839
1990	−0.0778	−6.7656	−0.0921	−10.0814
1991	−0.0602	−5.2255	−0.0667	−7.2947
1992	0.0183	1.6097	0.0021	0.2282
ln(Age+1)	−0.0683	−23.8419	−0.0514	−20.2074
ln(Living Area)	1.0179	70.6680	0.8150	62.5753
ln(Other Area)	0.0952	17.4947	0.0801	17.9154
Baths	0.0605	9.2541	0.0481	8.9927
R^2	0.7840		0.8754	
SSE	193.2509		111.5261	
Log-likelihood	−13 799.55		−12 358.43	
Schwartz IC	−3.2811		−3.5057	
n	5243		5243	
k	12		211	

spatial correlation invalidates the inferences from model 1 as the standard errors have a downward bias.

Model 2 effectively augments model 1 with a large set of spatial indicator variables. We divided the N–S and E–W locational dimensions into equal numbers of quantiles (equally spaced) and formed all the sub-regions defined by the intersection of these quantiles in the fashion of a checkerboard. As the number of quantiles grew, the number of possible sub-regions grew with the square of the number of quantiles. We gave every sub-region an indicator variable and associated parameter. Hence, ten quantiles would create 100 possible spatial sub-region indicator variables. Twice the log-likelihood increased significantly until 211 variables entered the model (199 spatial indicator variables and 12 non-spatial variables). Note, some of the possible sub-region spatial indicators had no observations and did not enter the model. Naturally, an informed practitioner could tessellate Baton Rouge more intelligently than this mechanical procedure. Nevertheless, the importance of spatial considerations leads to a large number of potential variables.

Again, the estimates appear reasonable and similar to those in model 1. Note, the magnitude of the estimates of the value of housing characteristics decreased in value as did the t statistics (except for the coefficient on ln(Other Area)). Fig. 1 documents the spatial indicator variables dramatically lowered the spatial correlations among neighbors, especially at the longer ranges. For the nearest neighbors, however, model 2 still shows positive correlations.

3.4. Spatial–temporal estimation results

In the analysis below, we selected $m_S = 15$ spatial neighbors, $m_T = 180$ temporal neighbors (about 3–4 months in time), and set λ equal to

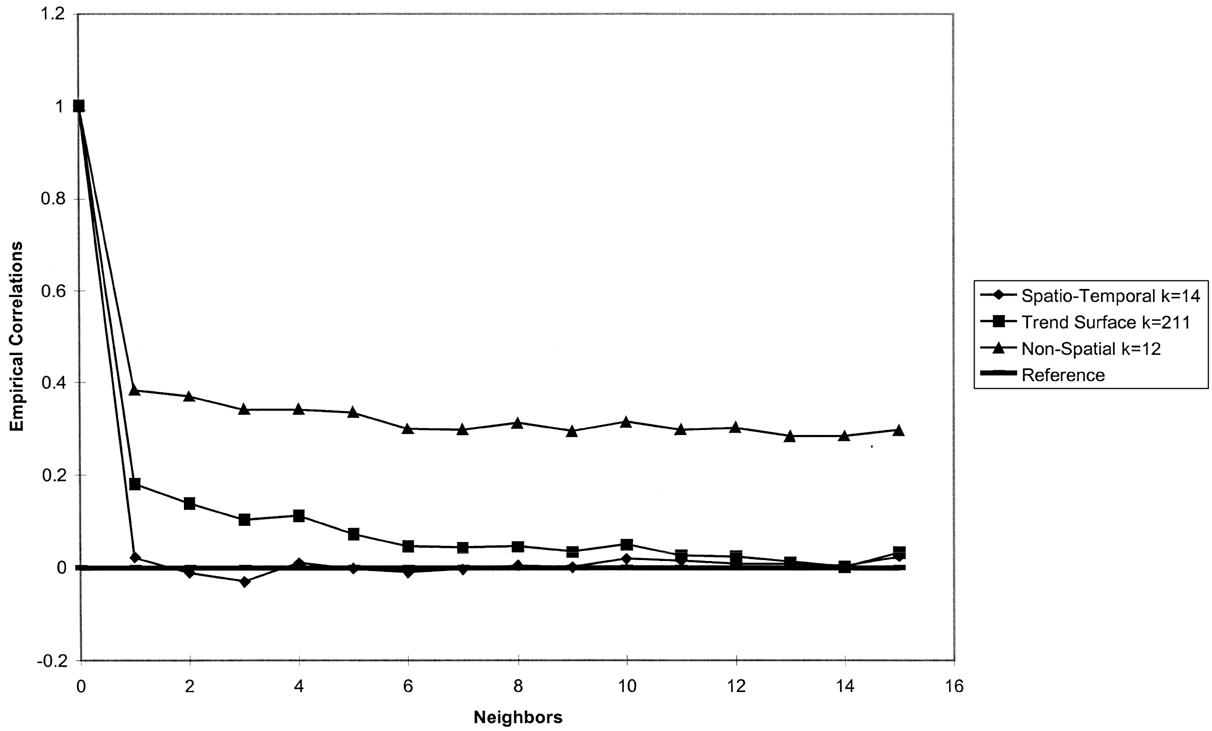

Fig. 1. Empirical error correlations versus neighbors.

0.75. Section 3.6. below discusses the sensitivity of the results to different numbers of temporal neighbors and different values of λ.

We began with an equivalent version of the full expansion of the general model as expressed by (7). Insofar as differencing over time can yield more parsimonious models, we rewrite (7) equivalently in terms of differences over time plus lags. This yielded model 3. Hence, the dependent variable is ln(Sales Price) $-$ T ln(Sales Price). Looking at the estimates in Table 2, the t values of the estimates associated do not appear extremely large for any of the terms associated with TX, STX, and TSX, T ln(Price), and the linear index (Index$(1-n)/n$). Hence, the differencing has the potential for increasing parsimony without greatly reducing the fit.

We subsequently deleted the time related independent variables, T ln(Price), and the independent trend variable (Index$(1-n)/n$). This yielded model 4. Comparing model 4 directly to model 3 does not show any statistical difference at even the 5% level. Model 4 has 14 independent variables, half the size of model 3. For model 4, the pure spatial autoregressive parameters $\tilde{\phi}_S$ has a high but plausible coefficient of 0.8833 with a very large t value of 41.28. All the variables, except the one measuring North–South effects have t values over 2 in absolute value.

The significant t statistics associated with both the coefficients on the variables ST ln(Price) and TS ln(Price) suggested the previously discussed optimal dependent variable transformation of $(I-W)=\omega(I-T)(I-\phi_S S)+(1-\omega)(I-\phi_S S)(I-T)$. Rewriting this yields $(I-W)=I-T-\phi_S S+\phi_S(\omega ST+(1-\omega)TS)$. This suggests the estimated coefficients associated with ST and TS should sum to $-\tilde{\phi}_S$.

Table 2
Spatial–temporal estimates

Variables	Model 3		Model 4		Model 5	
	Estimates	t	Estimates	t	Estimates	t
Intercept	2.1827	0.7290	1.7621	2.2449	0.2570	0.7122
Index $(1-n)/n$	−0.0267	−0.9070				
N–S	0.0000	−0.0489	0.0000	0.0151	0.0001	0.0881
E–W	−0.0024	−4.0595	−0.0024	−4.0245	−0.0023	−3.9386
$(X-TX)$:						
ln(Age+1)	−0.0351	−14.3588	−0.0354	−14.5278	−0.0352	−14.4796
ln(Living Area)	0.7315	56.9468	0.7303	57.1663	0.7306	57.1616
ln(Other Area)	0.0684	16.5354	0.0679	16.4847	0.0680	16.4844
Baths	0.0355	7.1571	0.0362	7.3365	0.0363	7.3564
$S(X-TX)$:						
ln(Age+1)	0.0118	2.7890	0.0113	2.6911	0.0111	2.6449
ln(Living Area)	−0.5065	−16.5386	−0.5032	−16.5428	−0.5058	−16.6342
ln(Other Area)	−0.0363	−3.7260	−0.0376	−3.8813	−0.0382	−3.9371
Baths	−0.0533	−5.0075	−0.0546	−5.1271	−0.0542	−5.0900
TX:						
ln(Age+1)	0.0118	0.6668				
ln(Living Area)	0.7120	2.6305				
ln(Other Area)	0.1753	2.6022				
Baths	−0.1333	−1.4481				
STX:						
ln(Age+1)	0.0055	0.1980				
ln(Living Area)	0.3151	0.9392				
ln(Other Area)	0.0736	0.4906				
Baths	0.0320	0.2067				
TSX:						
ln(Age+1)	0.1577	2.9366				
ln(Living Area)	−1.6257	−3.0615				
ln(Other Area)	0.3224	2.5639				
Baths	0.3234	1.7073				
S ln(Price)	0.8829	40.9035	0.8833	41.2803	0.8857	41.4230
T ln(Price)	−0.2691	−2.6625				
ST ln(Price)	−0.7645	−6.0032	−0.5958	−10.0057	−0.5560	−9.8151
TS ln(Price)	−0.0203	−0.0753	−0.4174	−6.1507	−0.3297	−6.0627
m_T	180		180		180	
λ	0.75		0.75		0.75	
R^2	0.8858		0.8853		0.8852	
SSE	102.1943		102.6532		102.7447	
Log-likelihood	−12 129.35		−12 141.10		−12 143.44	
Schwartz IC	−3.8920		−3.9104		−3.9112	
n	5243		5243		5243	
k	28		14		13 (14+1 restriction)	

Note, the sign switches because we are imposing the restriction on the RHS although discussing a transformation on the LHS. We imposed the restriction $\tilde{\phi}_S = -(\tilde{\phi}_{ST} + \tilde{\phi}_{TS})$ and reestimated the model. This regression appears as model 5. The restricted log-likelihood for model 5 was -12143.44 which did not differ significantly from the unrestricted log-likelihood of -12141.10 for model 4 at the 1% level. In the restricted regression, $\tilde{\phi}_{ST} = -0.5560$, $\tilde{\phi}_{TS} = -0.3297$, and naturally $\tilde{\phi}_S = 0.8857$.

Any a priori restriction on whether to filter the dependent variable first for time and then for space (e.g. $(I - W)\ln(\text{Price}) = (I - \phi_S S)(I - \phi_T T)\ln(\text{Price})$) or to filter first for space and then for time (e.g. $(I - W)\ln(\text{Price}) = (I - \phi_T T)(I - \phi_S S)\ln(\text{Price})$) is rather troubling. We avoid this dilemma by fitting the general form in (5) via models 3 and 4. Moreover, the general form in this case can be restricted to a linear combination of the above special cases for these particular data (e.g. $(I - W)\ln(\text{Price}) = (\omega(I - T)(I - \phi_S S) + (1 - \omega)(I - \phi_S S)(I - T))\ln(\text{Price})$) via model 5. The role of the restriction in model 5 is to make the general form in (5) more interpretable.

In conclusion, we accept model 5 with the restriction $\tilde{\phi}_S = -(\tilde{\phi}_{ST} + \tilde{\phi}_{TS})$. It has the second lowest amount of error coupled with an interpretable dependent variable transformation comprised of a weighted average of compound filters. Specifically, our estimated optimal transformation is $(I - W) = 0.3722 \cdot (I - T)(I - 0.8857 \cdot S) + 0.6278 \cdot (I - 0.8857 \cdot S)(I - T)$. Hence, we obtain the best results by roughly weighting the spatially filtered time differenced dependent variable almost twice as heavily as weighting the time differenced spatially filtered dependent variable. This transformation provides an intuitive yet fruitful way of dealing with the potentially daunting prospect of jointly modeling spatial and temporal effects.

In examining the estimated coefficients for model 5, the signs on the housing characteristics match prior expectations and these all seem quite significant. The E–W variable (which actually increases from west to east) has a negative sign. Hence, it measures a decline in value when going away from the Mississippi River. As the central business district (CBD) abuts the river, this most likely measures the more or less deterministic trend of a decline in housing prices as distance from the CBD increases.

3.5. Non-spatial, trend surface, and spatial–temporal comparative results

The coefficient estimates and t statistics on the housing variables show an interesting pattern across the models. Essentially, the non-spatial model estimates deviate the most from zero, followed by the trend surface model, and the spatial–temporal models. The t statistics usually show the same pattern as well. In every case the t statistics for the non-spatial model exceed the ones for the spatial–temporal model 5.

Fig. 1 documents the great reduction of the problem of error correlations with the nearest neighbors obtained by the spatial–temporal model 5 relative to those produced by the non-spatial and trend surface models. The non-spatial model displays both serious short and long range correlations, the trend surface model shows short range (but not long range) correlations while the spatial–temporal model displays low correlations with both short and long range neighbors.

To test ex-sample forecasting ability of the spatial–temporal models, we computed the one-step forecasts for model 4. We used model 4 instead of model 5 because the updating formulae for unrestricted OLS make this more computationally palatable. Table 3 shows the quantiles of the one-step or ex-sample errors as well as the sample errors for model 4. For

Table 3
Error statistics for the trend surface and spatio–temporal models

Error statistics	Trend surface (Sample errors)	Model 4 (Sample errors)	Model 4 (Ex-sample errors)
Number of variables	211	14	14
Minimum	−1.2099	−1.2531	−1.2710
1%	−0.4137	−0.4061	−0.4259
5%	−0.2388	−0.2238	−0.2275
10%	−0.1664	−0.1584	−0.1596
25%	−0.0752	−0.0712	−0.0706
50%	0.0046	0.0076	0.0085
75%	0.0824	0.0762	0.0765
90%	0.1605	0.1456	0.1471
95%	0.2095	0.2004	0.2013
99%	0.3664	0.3466	0.3343
Maximum	0.9676	1.0556	0.8938
Median $\|e\|$	0.0794	0.0735	0.0740
n	5243	5243	4000
Time neighbors	0	180	180
Space neighbors	0	15	15

model 4 the one-step forecast error quantiles do not differ greatly from their sample error counterparts.

Fitting the spatial–temporal model (model 4) with only 14 parameters led to a fit with 8% less SSE than produced by the trend surface model (model 2) with 211 parameters. Moreover, the spatial–temporal model outperformed the trend surface model over the 5th to 95th percentiles. Note, the spatial–temporal model 5 *ex-sample* prediction error beats the 211 variable trend surface model *sample* error.

Finally, all the sample error quantiles indicate a substantial amount of kurtosis in the error density. The use of robust estimators might well improve on the results further.

3.6. Sensitivity of the spatial and temporal model to weighting matrices

Table 4 contains the SSE from the spatial–temporal model fit as a function of the spatial matrix weighting λ and the number of temporal neighbors m_T. Based on preliminary fitting, we looked at the range for m_T of between 160 and 200 neighbors, and of λ between 0.6 and 0.8. This produced 25 cases.

The use of 180 temporal neighbors and a weighting of the individual spatial neighbors by λ of 0.75 produced the lowest SSE. Table 4 illustrates the relatively low sensitivity of the goodness-of-fit to small changes in the parameters λ and m_T.

The advantages from the computational innovations of employing (1) a lower triangular W; (2) sparsity in S; (3) individual neighbor matrices; and (4) linear filter routines becomes apparent through this experiment. We examined 25 separate regressions of 5243 observations where we needed to recompute SY, SX, TY, and TX repeatedly. Conditional upon already having the individual neighbor matrices, this exercise took only 58.71 s in Matlab on a 200-MHz Pentium Pro with 64 megabytes of memory.

Table 4
Spatio–temporal SSE versus number of temporal neighbors and spatial neighbor weighting

m_T	λ	SSE
160	0.65	103.5900
160	0.70	103.0298
160	0.75	102.7686
160	0.80	102.9285
160	0.85	103.6561
170	0.65	103.5743
170	0.70	103.0161
170	0.75	102.7569
170	0.80	102.9188
170	0.85	103.6485
180	0.65	103.4630
180	0.70	102.9080
180	0.75	102.6532
180	0.80	102.8207
180	0.85	103.5569
190	0.65	103.4833
190	0.70	102.9279
190	0.75	102.6722
190	0.80	102.8382
190	0.85	103.5726
200	0.65	103.4745
200	0.70	102.9257
200	0.75	102.6783
200	0.80	102.8546
200	0.85	103.6012
Execution time (s)	58.71	

4. Conclusion

This paper demonstrated some simple, easily computed techniques for taking into account both temporal and spatial information. Relative to using an extensive set of indicator variables, modeling correlations among the spatial and temporal errors produced a better goodness-of-fit coupled with much higher levels of parsimony in the regression equation. Specifically, the spatial–temporal regression with 14 variables exhibited 8% less sum-of-squared errors than the trend surface model regression with 211 variables. In this example, the extensive set of indicators over time and space did result in low long-range correlations among errors but still exhibited substantial short-range correlations among errors. In contrast, the spatial–temporal regression displayed lower levels of correlation among errors at all ranges.

The spatial–temporal regression also displayed good ex-sample forecasting, an important desideratum in many contexts. For example, most local tax assessments in the US rely upon statistical predictions of housing values. In addition, primary and secondary mortgage lenders have begun exploring the use of statistical forecasting of housing prices as opposed to employing appraisers. Indicator set approaches do not lead as naturally to such predictions.

Other applications such as creating constant quality indices could also benefit from jointly modeling the errors jointly over time and space. The continuous nature of the spatial variables in the spatial–temporal models allows the creation of a constant quality index surface as opposed to a set of separate indices. Naturally, specification of the correlations among errors across time and space could improve the quality of estimates in hedonic pricing applications.

Structuring the problem to employ lower triangular weighting matrices comprised of individual sparse neighbor matrices and linear filter routines resolves the seemingly difficult problem of differencing each error from numerous other errors and avoids lengthy determinant computations.

As many data in economics have specific spatial and temporal characteristics, these techniques could prove useful in other contexts as well. The becomes especially true when considering alternative spatial metrics which involve variables other than physical location. For example, Case, Rosen and Hines (1993) used separate metrics involving physical location, race, and income to improve on a standard regression's ability to explain public expenditures. They found significant correlations among errors based upon both ordering by physical location and by race. Note, their data spanned the period 1970–85 and hence had a temporal dimension as well. As a potential extension to our work, decomposition of the weight matrix W into more than the physical location and temporal components examined herein could potentially produce additional econometric benefits without incurring significantly greater computational costs.

Acknowledgements

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 would like to thank John Knight and John Clapp for their insightful remarks. Finally, we would like to greatly thank an anonymous Associate Editor for some insightful remarks.

References

Anselin, L. (1988). Spatial econometrics: methods and models, Kluwer, Dordrecht.

Anselin, L., & Hudak, S. (1992). Spatial econometrics in practice: a review of software options. *Journal of Regional Science and Urban Economics 22*, 509–536.

Brueckner, J. K. (1997). Testing for strategic interaction among local governments: the case of growth controls, Department of Economics, University of Illinois, Working paper.

Case, A. C., Rosen, H. S., & Hines, J. R. (1993). Budget spillovers and fiscal policy interdependence. *Journal of Public Economics 52*, 285–307.

Cressie, N. (1993). Spatial statistics, Wiley, New York.

Davidson, R., & MacKinnon, J. (1993). Estimation and inference in econometrics, Oxford University Press, New York.

Deutsch, S. J., & Ramos, J. A. (1986). Space-time modeling of vector hydrologic sequences. *Water Resources Bulletin 22*, 967–980.

Dubin, R. (1988). Estimation of regression coefficients in the presence of spatially autocorrelated error terms. *Review of Economics and Statistics 70*, 466–474.

Eckert, J. (1990). Property appraisals and assessment administration, International Association of Assessing Officers, Chicago.

Gelfand, A. E., Ghosh, S. K., Knight, J. R., & Sirmans, C. F. (1998). Spatial–temporal modeling of residential sales data. *Journal of Business and Economic Statistics 16*, 312–321.

Gilley, O. W., & Pace, R. K. (1995). Improving hedonic estimation with an inequality restricted estimator. *Review of Economics and Statistics 77*, 609–621.

Hendry, D., Pagan, A., & Sargan, D. (1984). Dynamic specification. In: Griliches, Z., & Intrilligator, M. (Eds.), Handbook of econometrics, North-Holland, Amsterdam, pp. 1023–1100.

Hill, R. C., Knight, J. R., & Sirmans, C. F. (1997). Estimating capital asset price indices. *Review of Economics and Statistics 78*, 226–233.

Johnson, S. (1998). Address matching with stand-alone geocoding engines. *Business Geographics 6*, 30–36.

Kelejian, H. H., & Prucha, I. R. (1997). A generalized spatial two-stage least squares procedure for estimating a spatial autoregressive model with autoregressive disturbances. *Journal of Real Estate Finance and Economics 17*, 99–121.

Knight, J. R., Sirmans, C. F., & Turnbull, G. (1994). List price signaling and buyer behavior in the housing market. *Journal of Real Estate Finance and Economics 9*, 177–192.

Ord, K. (1975). Estimation methods for models of spatial interaction. *Journal of the American Statistical Association 70*, 120–126.

Pace, R. K. (1997). Performing large-scale spatial autoregressions. *Economics Letters 54*, 283–291.

Pace, R. K., & Barry, R. (1997a). Sparse spatial autoregressions. *Statistics and Probability Letters 33*, 291–297.

Pace, R. K., & Barry, R. (1997b). Quick computation of regressions with a spatially autoregressive dependent variable. *Geographical Analysis 27*, 232–247.

Pace, R. K., & Gilley, O. W. (1998). Generalizing the grid and OLS estimators. *Real Estate Economics 26*, 331–347.

Pace, R. K., Barry, R., Clapp, J., & Rodriguez, M. (1998). Spatial–temporal estimation of neighborhood effects. *Journal of Real Estate Finance and Economics 17*, 15–33.

Pfeifer, P. E., & Bodily, S. E. (1990). A test of space–time ARMA modelling and forecasting of hotel data. *Journal of Forecasting 9*, 255–272.

Ratcliffe, R. (1972). Valuation for real estate decisions, Democrat Press, Washington.

Silverman, B. W. (1986). Density estimation for statistics and data analysis, Chapman and Hall, London.

Szummer, M., & Picard, R. W. (1996). Temporal texture modeling, MIT Media Lab, Cambridge, MA, Manuscript.