Apartment Rent Prediction Using Spatial Modeling

Authors

James Valente, ShanShan Wu, Alan Gelfand and C. F. Sirmans

Abstract

This paper provides a new model to explain local variation in apartment rents by introducing the notion of a spatial process. This model differs from those in the literature by explicitly specifying spatial association between pairs of locations as a function of distance between them. Data on apartment rents for the eight markets are used to illustrate the spatial model. Results indicate significant prediction improvement over traditional hedonic rent models that only include indicator variables to capture spatial effects.

Introduction

Real estate practitioners and academics have long sought to explain the variation in apartment rents through the use of regression models. Most of the research has focused on using a global mean function in an attempt to explain large scale or global variation, which can be attributed to the differences in physical attributes and amenities that are unique to each property, and general location variables that use proximity or zonal association to measure the effect on rent of various external amenities. Regardless, there will always be omitted unobservable variables that may carry spatial information. As a result, residual dependence structure may remain in the observed rents. Widely available GIS software and formal spatial modeling facilitate better understanding of this local variation.

This paper introduces a spatial model structure employing the highest resolution possible (*i.e.*, there is no aggregation of rental properties), at the scale of the data. Expressed in different terms, we are modeling a ''conceptual rent'' at every location in the market and the data consists of observations of this rent surface at a finite set of geo-coded locations. The model explicitly specifies spatial association between pairs of locations as a function of distance between them.[1] This enables the notion of a spatial range (*i.e.*, the distance beyond which spatial association becomes negligible). Two error terms are incorporated in the residual: one to reflect heterogeneity or non-spatial error and one to capture a purely spatial contribution. In fact, variance components for each of these terms are provided. Dummy indicator variables are introduced for neighborhoods within a market to help capture spatial structure but, again, we are still modeling at the point level. Finally, separate models for each market in the study are estimated.

Such specification enables inference about individual markets as well as comparison across markets. In particular, an inference can be drawn about the association between rents for any pair of locations in a given market. Knowledge about the decay in spatial association as a function of distance as well as the spatial range can be gained. Rents can be predicted at arbitrary locations and hence a "picture" of rents can be obtained through an estimated rent surface. Thus, a pattern can be derived, across a market, where rents tend to be elevated and where they tend to be depressed. The error terms in the residuals and their associated variance components can be examined to assess the strength of the spatial association and the usefulness of the implementation of spatial modeling. Finally, the neighborhood determinants can be analyzed to see how effectively they explain spatial association (*i.e.*, for some markets the resulting residuals reveal little spatial pattern while for others considerable pattern remains).

Background

Location variables can help explain a portion of large scale variation, and therefore should be modeled in the mean structure, because they focus on providing insight into how consumers of space, in general, value accessibility to or association with certain external amenities (submarket, employment center, highway ramp, shopping center, etc.). However, even if all available location variables are used, the rents of individual properties can still be expected to vary with each other based on their relative proximity to each other. For example, to estimate what the appropriate asking rent might be for an apartment unit in a new complex, it is correct to assume that it is more like a similar unit in a complex across the street, than it would be to a similar unit in a complex a mile away, or ten miles away in another town. At the same time, the asking rent for a unit in the property ten miles away would be more like that of a similar unit in a complex across the street from it, than it would to either of the similar units in the original town. The use of location variables cannot directly help us assess the strength of this similarity.

The fact that incorporating general location variables into a traditional Ordinary Least Squares (OLS) model does not address spatial association is evidenced by the example shown in Exhibit 1, which presents the results from a simple OLS model using the dataset for the Atlanta market (described in detail in the next section). This model contains dummy variables to account for the sub-markets, and a proximity variable measured as the distance of each property to the nearest employment center.[2]

A spatial presentation of the residuals is provided in Exhibit 2. The high degree of spatial association among properties that remains is clear. In particular, the black circles indicate positive residuals, and the hollow circles represent negative

Exhibit 1 | Simple OLS Results for Multi-Family Rents in Atlanta

	Value	Std. Error	*t*-value	Pr (>|t|)
Intercept	6.405	0.030	210.549	0.0000
Charleston	−0.214	0.022	−9.698	0.0000
Clayton	−0.274	0.020	−13.868	0.0000
Decatur	−0.140	0.022	−6.495	0.0000
120 E.	−0.180	0.035	−5.196	0.0000
120 W.	−0.257	0.028	−9.140	0.0000
Marietta	−0.164	0.019	−8.470	0.0000
N. DeKalb	−0.067	0.019	−3.495	0.0005
N. Gwinnett	−0.124	0.023	−5.384	0.0000
Roswell	−0.097	0.027	−3.661	0.0003
S. DeKalb	−0.266	0.036	−7.461	0.0000
S. Fulton	−0.296	0.019	−15.887	0.0000
S. Gwinnett	−0.152	0.055	−6.927	0.0000
Sandy Springs	−0.059	0.022	−2.753	0.0060
Smyma	−0.108	0.022	−5.037	0.0000
Size Units	−0.000	0.000	4.118	0.0000
Size Avg. Unit	0.001	0.000	20.967	0.0000
Stories	0.013	0.002	6.292	0.0000
Age	−0.007	0.000	−16.455	0.0000
Rehab	−0.007	0.009	−0.798	0.4253
Dist. Emp. Ctr.	−0.463	0.107	−4.340	0.0000

Notes: Residual standard error: 0.1288 on 931 degrees of freedom; multiple R^2 is .6715; F-Statistic is 95.13 on 20 and 931 degrees of freedom; the p-value is 0.

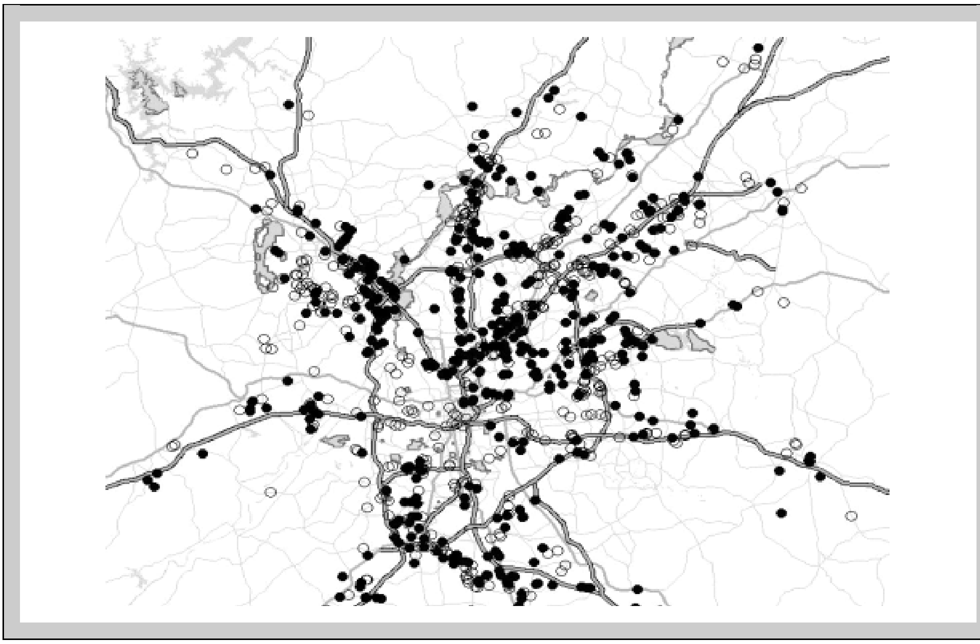
residuals. The result is that the OLS model underestimates rents just north of the center city, and overestimates rents to the southwest of the center city.

This study focuses on the question of out-of-sample rent prediction in the multi-family market. The spatial model provides an approach to explain both global trend and local variation, by explicitly addressing spatial dependency as a continuous function, which enables an analysis of how relationships between properties vary depending on their distance from each other.

Literature Review

There is a substantial body of literature focusing on which physical characteristic variables of apartments are significant in hedonic rent models. While this study

Exhibit 2 | Simple OLS Predicted Rent Residual Map of Atlanta



remains consistent with the findings of these studies in modeling the physical characteristics of the properties in the model, this is not the focus of the study, and the reader is referred to Sirmans and Benjamin (1991) for a review of this literature.

As mentioned previously, general location variables represent global trend, and therefore, their inclusion does not fully constitute a spatial model. This notion is supported by Fotheringham and Rogerson (1993) who found that many models that have been put forth as spatial are really *aspatial* models with some sort of distance variable included. Because general location variables are important in modeling the global variation, it is theoretically possible to include enough location variables in a traditional OLS model to remove all of the local variation, and thus spatial autocorrelation among the residuals. Pace, Barry and Sirmans (1998) point out, in a simple example, that the number of variables needed to remove all local variation can quickly grow out of control. This does not mean that all variables can be observed that are potential explainers of spatial association. Rather, it implies that more and more complicated functions of the location variables that have been observed can be introduced. Beyond the implications of fitting, loss of degrees of freedom, and thus the strength of the modeling, the application of this approach is market (or more accurately location) specific, and therefore provides little insight for advancing general theory.

While proximity variables have traditionally garnered significant interest from researchers, their inconsistency across markets makes them less desirable when modeling global variation than indicator variables for zonal association. In an early attempt at including both proximity and zonal location variables in a hedonic rent model, Guntermann and Norrbin (1987) included proximity proxies for access to a major highway and distance to the CBD, and indicator variables for university and non-university zones. The authors found both of the proximity variables significant, but the positive sign for the distance to CBD proxy was counter intuitive. This type of result is endemic when trying to draw broad conclusions from proximity variables.[3] The authors also attempted to subdivide the market into zones and test for the impact on individual variables. Their results indicate that a number of variables, including age, condition and additional bedrooms, were valued differently in different zones.

With similar results, Sirmans, Sirmans and Benjamin (1990) included proximity proxies for traffic congestion, access to public transportation and access to employment; they also included seven distinct indicator variables in an attempt to account for sub-market association. While they found that traffic congestion and access to public transportation were significant, they did not find that proximity to employment was significant. At the same time, they found that two of the zones exhibited a significant effect. In the present study, the findings show that even after using sub-market indicators in the model to account for global variation, substantial spatial dependence still remains between observations.

Additional studies by Asabere and Huffman (1996) for the Philadelphia market, and Frew and Wilson (2002) for the Portland, Oregon market test access to employment centers, using proximity to transportation nodes as a proxy, and have found access to employment centers to be significant. Alternatively, studies by Smith and Kroll (1989) and Des Rosiers and Thériault (1996) apply market segmentation techniques, with much success, to identify homogenous clusters of renters and zones.

While this body of research has provided significant insight into the workings of apartment markets, it fails to address the issue of local variation. In an overview of spatial statistics and its applications to real estate, Pace, Barry and Sirmans (1998) provide a good introduction to the primary spatial statistical models: conditional spatial autoregressive (CAR), simultaneous spatial autoregressive (SAR) and kriging. For additional background on these techniques, see Dubin and Sung (1987), Anselin (1990a, 1990b), Haining (1990), Cressie (1993), Bailey and Gatrell (1995) and Fotheringham, Brunsdon and Charlton (2000).

Briefly, spatial models are built in some cases for areal units and in some cases for data at point level. The former are used, for example, when measurements are available only in an aggregated fashion. For instance, responses might be observed for units such as census tracts or zip codes. They might be counts or proportions associated with the areal unit. With apartment rents, it is more attractive to work at the highest resolution of the data (*i.e.*, at the point level using geo-coded property locations as coordinates).

For point referenced data, there are two prevailing modeling approaches. For apartment rents, one approach models the expected rent for a particular building as an inverse-distance weighted average of the rents of the neighboring buildings or complexes, adjusted for characteristics of the building. This approach includes both CAR and SAR models.

Advantages of this class of models are a nearest neighbor-based smoothing of the means and convenient computation. Moreover, they certainly improve over OLS models with regard to explaining error. For instance, Pace and Gilley (1997) apply a SAR model to the Harrison and Rubinfeld (1978) hedonic pricing data with great success; they were able to reduce the SSE by 44% compared to a traditional OLS model. Additionally, for at least a couple of variables, their spatial model provided signs consistent with their a priori assumptions, where the traditional OLS model did not.

Potential limitations of CAR and SAR modeling include: (1) association between locations is not modeled directly. In particular, the customary spatial autocorrelation coefficient introduced in these models does not align with strength of direct association between locations. The former needs to be quite high in order to induce consequential direct association. (2) Spatial prediction is not built in. Only the joint distribution for rents at observed locations is modeled. Prediction for arbitrary new locations must be done in an ad hoc (sometimes deterministic) fashion.

This study describes the responses at the given locations (*e.g.*, the rents) using a spatial process (sometimes referred to as a spatial random field) model. This approach has several potential advantages. It conceptualizes a rent at every location in the region (*i.e.*, a rent surface over the region). It directly models the degree of spatial association between responses at arbitrary pairs of locations. It immediately provides the joint distribution of any number of and choice of locations with trivial marginal distributions. Hence, spatial prediction becomes straightforward. The major drawback is that it is more computationally demanding to fit to data than the CAR and SAR approaches. However, the inferential gain is worth the extra computational effort. The details of this approach and its implementation are discussed in the analysis of the apartment rental market below.

Data

The data represent a cross section of market rate apartment rental complexes that consist of forty or more units (except for California metropolitan areas, where complexes of fifteen or more units are included). The data have been provided by

REIS, Inc.,[4] a provider of real estate market information to the institutional real estate investment community. The observations represent a subset—approximately 50%—of the properties that are tracked in each market. All data are for the fourth quarter of 2002.

To test the overall robustness of the modeling technique, data from eight different metropolitan areas are used. From an analysis standpoint, each market provides potentially unique and different spatial patterns. The markets are: Atlanta, Boston, Chicago, Houston, Los Angeles, Jacksonville, San Diego and San Francisco.

Exhibit 3 presents variable definitions. Exhibit 4 contains summary statistics for the variables used in the rent estimates for each of the eight markets. The data conform to expectations with rent highest in Boston and San Francisco and lowest in Atlanta and Jacksonville. Note that the sample sizes range from 98 in Jacksonville to 1,036 in San Francisco.

Spatial Process Model

First, begin by denoting the set of all possible locations in a particular market by D. Then, denote the rent surface by $\{Y(s) : s \in D\}$, where each $Y(s)$ is viewed as a random variable. The set is an uncountable collection of random variables, indexed by location, which is referred to as a spatial (stochastic) process. The process is only observed at a finite set of locations $s_1, s_2, \ldots, s_n$. Use $Y(s_1), Y(s_2), \ldots, Y(s_n)$ to infer the unknowns in the process model (*e.g.*, the mean structure, the variability and the association between locations), and through interpolation, the entire rent surface.

Exhibit 3 | Definition of Variables

Variable	Definition
Size Units	Total number of units in the complex
Size Avg Units	Size of an average unit in the complex measured in square feet
Stories	Total number of above ground stories in the complex
Rent	Monthly asking rent per unit per month
lnRent	Natural log of the average asking rent per unit per month in the complex
Age	Age of the complex, not adjusted for renovations
Rehab	Indicator variable to identify if a project has gone through a major renovation

Exhibit 4 | Summary Statistics

	Size Units	*Size_AvgUn*	*Stories*	*Rent*	*lnRent*	*Age*	*Rehab*
Panel A: Atlanta (*n* = 952)							
Average	221	1007	3	764	6.613	22.88	0.31
Median	200	1012	2	741	6.608	25	0
Min.	52	450	1	368	5.908	1	0
Max.	1180	1803	40	1804	7.498	87	1
Range	1128	1353	39	1436	1.590	86	1
Std. Dev.	134	195	2	178	0.222	12.349	0.464
Variance	17849	37930	4	31731	0.049	152.491	0.216
Skewness	2	0	12	1	0.226	0.643	0.802
Kurtosis	5	1	178	4	0.705	1.938	−1.359
Panel B: Boston (*n* = 160)							
Average	235	862	6	1483	7.250	27.11	0.21
Median	175	851	4	1377	7.227	27	0
Min.	51	449	1	743	6.611	1	0
Max.	1283	1445	38	3633	8.198	96	1
Range	1232	996	37	2890	1.587	95	1
Std. Dev.	194	160	5	515	0.313	16.418	0.406
Variance	37597	25612	30	264785	0.098	269.534	0.165
Skewness	2	0	3	1	0.559	1.767	1.466
Kurtosis	8	2	9	2	0.084	5.347	0.150

Exhibit 4 | (continued)

Summary Statistics

	Size Units	*Size_AvgUn*	*Stories*	*Rent*	*lnRent*	*Age*	*Rehab*
Panel C: Chicago (n = 589)							
Average	235	789	7	902	6.760	33.54	0.32
Median	176	799	3	842	6.736	29	0
Min.	52	300	1	328	5.793	1	0
Max.	1869	1772	56	2275	7.730	114	1
Range	1817	1472	55	1947	1.937	113	1
Std. Dev.	206	187	9	290	0.292	21.825	0.467
Variance	42274	34876	83	83954	0.085	476330	0.218
Skewness	3	0	9	2	0.451	1.217	0.778
Kurtosis	14	2	9	4	0.620	0.804	−1.400
Panel D: Houston (n = 942)							
Average	236	848	2	626	6.406	22.89	0.27
Median	212	827	2	579	6.361	23	0
Min.	52	447	1	339	5.826	1	0
Max.	1800	1934	30	2743	7.917	81	1
Range	1748	1487	29	2404	2.091	80	1
Std. Dev.	153	148	1	186	0.246	9.423	0.445
Variance	23418	21820	2	34422	0.060	88.797	0.020
Skewness	3	2	14	3	1.135	0.002	1.034
Kurtosis	20	8	255	21	2.353	1.566	−0.933

Exhibit 4 | (continued)

Summary Statistics

	Size Units	*Size_AvgUn*	*Stories*	*Rent*	*lnRent*	*Age*	*Rehab*
Panel E: Jacksonville (n = 98)							
Average	214	913	2	616	6.403	23.18	0.35
Median	170	903	2	608	6.409	27	0
Min.	52	525	1	342	5.835	3	0
Max.	1103	1328	4	1013	6.921	54	1
Range	1051	803	3	671	1.086	51	1
Std. Dev.	142	179	1	129	0.203	10.476	0.478
Variance	20138	31903	0	16683	0.041	109.739	0.229
Skewness	3	0	0	1	0.183	0.102	0.653
Kurtosis	15	0	1	1	0.447	−0.200	−1.607
Panel F: Los Angeles (n = 1036)							
Average	134	805	3	1015	6.875	27.50	0.34
Median	93	784	3	943	6.849	28	0
Min.	50	270	1	388	5.961	1	0
Max.	4200	1870	27	3629	8.197	109	1
Range	4150	1600	26	3241	2.236	108	1
Std. Dev.	173	187	2	342	0.299	13.786	0.473
Variance	30096	34972	5	116671	0.090	190.047	0.224
Skewness	14	1	6	2	0.494	1.388	0.687
Kurtosis	296	2	41	8	0.780	3.915	−1.531

Exhibit 4 | (continued)

Summary Statistics

	Size Units	*Size_AvgUn*	*Stories*	*Rent*	*lnRent*	*Age*	*Rehab*
Panel G: San Diego (n = 593)							
Average	144	824	2	948	6.832	22.36	0.35
Median	108	826	2	912	6.816	23	0
Min.	51	332	1	519	6.252	1	0
Max.	752	1503	16	1934	7.567	84	1
Range	701	1171	15	1415	1.315	83	1
Std. Dev.	104	153	1	213	0.212	9.325	0.478
Variance	10780	23265	1	45479	0.045	86.954	0.228
Skewness	2	0	9	1	0.423	1.176	0.627
Kurtosis	6	1	102	2	0.405	5.122	−1.612
Panel H: San Francisco (n = 200)							
Average	180	761	4	1519	7.289	38.51	0.34
Median	96	760	3	1457	7.284	33	0
Min.	51	378	1	788	6.669	2	0
Max.	3483	1433	22	3441	8.144	107	1
Range	3432	1055	21	2653	1.474	105	1
Std. Dev.	339	187	4	433	0.270	22.761	0.475
Variance	115259	35148	15	187474	0.073	518.070	0.226
Skewness	8	0	2	1	0.220	1.295	0.681
Kurtosis	68	0	5	3	0.303	1.018	−1.552

Formally, the model takes the form:

$$Y(s) = \mathbf{X}^T(s)\beta + \omega(s) + \varepsilon(s), \qquad (1)$$

where $Y(s)$ is log asking rent and $\mathbf{X}(s)$ is a vector of property characteristics. In the present case, $\mathbf{X}(s)$ is a vector consisting of the number of units in the apartment complex, the average size (square feet) of each unit, the number of stories in each complex, the actual age of the complex and an indicator variable to identify if the project has gone through a major rehabilitation.[5] Also, within each market, a complex dummy variable is introduced for each of the submarkets with an associated coefficient (random effect). These dummy variables play the role of Y *location proxies*. Hence, a fairly rich mean explains the global variation. As will be seen below, in some markets the resulting residuals reveal little association, while for others there is still considerable spatial structure left.[6]

As a result, at location s in Equation (1), $\omega(s) + \varepsilon(s)$ are interpreted as the total error. It is partitioned into two independent components: $\omega(s)$ is the spatial error contribution and $\varepsilon(s)$ is the pure non-spatial error contribution. The $\varepsilon(s)$ are modeled as independent and identically distributed $N(0, \tau^2)$ as in a standard OLS regression model. Here the $\omega(s)$ are modeled as a mean 0 Gaussian spatial process with variance σ^2 and correlation function $\exp(-\phi d)$, where ϕ is an unknown positive constant and $d \geq 0$. More precisely this means that for any two locations s and s', the correlation between $\omega(s)$ and $\omega(s')$, hence between $Y(s)$ and $Y(s')$, is $\rho(s_i - s_j; \phi) = \exp-(\phi\|s_i - s_j\|)$, where $\|s_i - s_j\|$ denotes the Euclidean distance between locations s and s'. Therefore, from a particular location, correlation is constant on circles and decays in distance with ϕ the decay parameter, which is clearly endogenous to the model, determining the rate. Within a market, this seems to be plausible local behavior. Such an assumption is referred to as isotropy. It implies that association depends only on distance, but not on direction. It is customarily adopted for simplicity though it is unlikely to hold in practice. The justification is usually that it may be reasonable that it will hold locally. But then, if prediction is a primary inference objective, it will not be very sensitive to choice of correlation function.

If instead the covariance depends on the separation vector $s - s'$, then the association depends on the orientation of the vector $s - s'$ and thus can change with direction. This case is referred to as stationarity. Finally, in the most general case, calculation of association could require both s and s', not just the distance. This is referred to as nonstationarity. The foregoing Gaussian process specification uniquely determines the joint distribution of the $w(s)$'s for any number of and any set of locations. Specifically it implies that the joint distribution of $\omega(s_1)$, $\omega(s_2)$, . . . , $\omega(s_n)$ is multivariate normal with mean vector 0 and covariance matrix having $(i, j)^{th}$ entry $\sigma^2 \exp(-\phi\|s_i - s_j\|)$. As a result, the $Y(s_1)$, $Y(s)$, . . . , $Y(s_n)$, from Equation (1), have the same off diagonal covariances with diagonal entries $\sigma^2 + \tau^2$.

The spatial range is defined here as the distance beyond which spatial association becomes essentially negligible. That is, with the correlation function proposed above, interlocation distance must go to infinity before correlation reaches 0. To enable a finite choice of d, the correlation is defined to be negligible if it is ≤ 0.05. As a result, $0.05 = \exp(-\phi range)$ can be solved to obtain that the range is essentially $3/\phi$. For a given city, it will be of interest to compare this range to the largest interlocation distance in the city. Again, under Equation (1), $\text{var}(Y(s)) = \sigma^2 + \tau^2$ (*i.e.*, the sum of the spatial error variance and the pure error variance). In the geostatistical literature, $\sigma^2 + \tau^2$ is referred to as the sill while τ^2 is referred to as the nugget and σ^2 the partial sill. Under isotropy, a plot of $1/2\ \text{var}\{(Y(s) - \mathbf{X}^T(s)\beta) - (Y(s') - \mathbf{X}^T(s')\beta)\}$ vs $\|s - s'\|$ is referred to as the semi variogram. For the choice of correlation function here, this plot passes through τ^2 at $\|s - s'\| = 0$ and approaches $\sigma^2 + \tau^2$ as $\|s - s'\| \rightarrow \infty$. In this plot, the range is roughly the distance by which 95% of the sill is reached.

Again, the error under the model in Equation (1) is not expected to have an entirely spatial explanation. There will be additional heterogeneity in a market (*e.g.*, the financing and operating costs of the property), due possibly to the characteristics of the owner and/or the tenant, etc.

The model in Equation (1) can be viewed as a random effects model with $\omega(s)$ playing the role of spatial random effects (with variance component σ^2). Hence, given $\{\omega(s_i)\}$, the $Y(s_1)$ are conditionally independent and the conditional likelihood takes the product normal form:

$$\frac{1}{(\tau^2)^{n/2}} \exp\left\{-\frac{1}{2\tau^2}\sum_{i=1}^{n}(Y(s_i) - \mathbf{X}^T(s_i)\beta - \omega(s_i))^2\right\}. \qquad (2)$$

As in usual random effects models, the normality of the likelihood along with the joint normal distribution of the spatial random effects allows marginalization (*i.e.* integrate out the effects). Now the $Y(s_i)$s are dependent. Letting $\mathbf{Y}^T = (Y(s_1), Y(s_n), \ldots, Y(s_2))$ and $\mathbf{X}$ be the matrix with row $\mathbf{X}^T(s_i)$, the marginal likelihood becomes:

$$|\sigma^2 R(\phi) + \tau^2 I|^{-1/2} \exp\left\{-\frac{1}{2}(\mathbf{Y} - \mathbf{X}\beta)^T(\sigma^2 R(\phi) + \tau^2 I)^{-1}(\mathbf{Y} - \mathbf{X}\beta)\right\}. \qquad (3)$$

In Equation (3), $R(\phi)$ is an $n \times n$ matrix with (i, j) entry $\rho(s_i - s_j;\phi)$. The covariance matrix for Y, which is $\sigma^2 R(\phi) + \tau^2 I$, reveals that the spatial effects in Equation (1) can be separated from the pure error effects; σ^2, ϕ and τ^2 are all identified. Inference regarding the spatial process model for $Y(s)$ involves β [perhaps the mean $\mathbf{X}^T(s)\beta$ for a given $\mathbf{X}(s)$], σ^2, τ^2 and ϕ (or the range $3/\phi$).

Likelihood inference is possible in the form of point estimation but to accurately assess the uncertainty in these estimates, a Bayesian approach is preferable. Thus, a Bayesian model is fit to Equation (2) [or equivalently Equation (3)] using rather non-informative prior information except for the decay parameter. Experience with the markets studied suggests that the spatial range should be of the order of the distance between the centers of neighboring submarkets. So, this prior knowledge is then incorporated into the prior specification for ϕ.

Inference summaries for these parameters take the form of point 95% (credible) interval estimates. Comparison between σ^2 and τ^2 informs about how much of the variance in $Y(s)$ (which is $\sigma^2 + \tau^2$) is attributable to spatial error and how much is pure error. In other words, the relative importance of the spatial story compared to the pure error story. Arguably, primary inferential interest resides in spatial prediction (*i.e.*, the prediction of rents for unobserved locations). Under Equation (3), prediction at a new location, say S_o, arises through the predictive distribution of $Y(\mathbf{s}_0)$, given $\{Y(\mathbf{s}_i)\}$. That is, under the Bayesian framework, an entire predictive distribution can be obtained for any location in the market. Such prediction can be summarized by using the mean and standard deviation of the associated predictive distribution, or perhaps a 95% equal-tailed prediction interval.

Finally, there may be an interest in seeing the entire spatial surface. In fact, there are several possibilities here. Under Equation (1), it could be useful to see the mean surface associated with $\omega(\mathbf{s})$. This would reveal the nature of the spatial pattern (where in the market the spatial effects tend to push $Y(\mathbf{s})$ up, and where it tends to pull $Y(s)$ down), as well as the extent of spatial smoothing (whether the surface appears somewhat "spiky" or shows a lot of homogeneity). It would also be useful to create a rent surface. The idea in this case would be to identify a "typical" property (*i.e.*, a typical $\mathbf{X}(s)$ vector) to enable an expected mean $\mathbf{X}^T(s)\boldsymbol{\beta}$ combined with an expected spatial adjustment $\omega(s)$ to give an expected $Y(s)$. Exponentiating would give roughly an expected rent at location $\mathbf{s}$. This is similar to the way price indices are constructed except that this is at the point level rather than an aggregation over the whole market.

Regardless of what surface is explored, the strategy is to overlay a fine grid of points on the market D. The foregoing spatial prediction is implemented at each grid point to obtain values at these points. Then, a standard interpolation is implemented, contouring or cloropeth mapping to show the surface.

Results

Both the OLS and spatial models were tested, each with the same mean specification. In the development of the models, a regime was used that employed 80% of the observations in each market, holding back 20% for prediction. For each market, summary statistics are provided to measure how good both models were at predicting the holdouts. The summary statistics are as follows (where *m* denotes the number held out in each market):

$$\text{Predictive SSE} = \sum_{j=1}^{m} (E(y(s_j)|\underset{\sim}{y}) - y(s_{j,obs}))^2$$

$$\text{Predictive VAR} = \sum_{j=1}^{m} Var(y(s_j)|\underset{\sim}{y})$$

P(0.10) = Proportion of predictions within 10% of true value.

P(0.20) = Proportion of predictions within 20% of true value.

Predictive SSE measures how well centered the predictions, $E(y(s_j)|\underset{\sim}{y})$, are, compared with the $y(s_{j,obs})$. Predictive VAR cumulates the predictive variances for $y(\tilde{s}_j)$ in the hold out sample. Smaller predictive variances are expected from a spatial model than an OLS model because the hold out ***Ys*** are dependent on the ***Ys*** used for fitting the model in the spatial case, while in the OLS model they are not. P(0.10) and P(0.20) are familiar relative error measures.

Recall from the definition of the model above that the true error consists of two components: the spatial component $\omega(s)$ and the true non spatial error component $\varepsilon(s)$, and that the variance of the error term consists of the two components σ^2 and τ^2. The OLS model has a single error term and a single error variance. It would be incorrect to think that the fitted OLS residuals will equal the corresponding sum of error component in the fitted spatial model. Similarly, the estimates of $var\ Y(\mathbf{s})$ under the OLS model (*i.e.*, of τ^2) will not agree with the estimates of $var\ Y(\mathbf{s})$ under the spatial model (*i.e.*, of $\sigma^2 + \tau^2$). Comparison between the models is made in predictive space using the above summary measures. In addition, the benefit of the spatial model can be measured by how much of the variance in the error is captured by σ^2, the spatial component, versus how much remains in the true non-spatial error term τ^2.

The following provides a summary of the results for each of the markets in our study. Exhibit 5 compares the results from OLS and spatial models. To begin, the results are separated into three groups, which are defined by the relative success of the spatial model over the traditional OLS model. The first group contains those markets that received significant benefit from the spatial model: Atlanta, Chicago, Houston and San Diego. The second group contains those markets that received only some benefit (*i.e.*, a reduction in predictive SSE and VAR), but no significant improvements in P(0.10) or P(0.20) from the spatial model, which includes Boston and Los Angeles. Finally, the third group includes those markets that appear to have not received any benefit from the spatial model, namely Jacksonville and San Francisco.

Group 1: Significant Improvement

The Atlanta market exhibited significant spatial structure, with a spatial range of about sixteen miles. The spatial model produced a relatively strong measure for σ^2, capturing approximately 64% of the total error variance. Additionally,

Exhibit 5 | Model Information

	OLS					Spatial Model				
	Median	2.5%	97.5%	Mean	Std. Dev.	Median	2.5%	97.5%	Mean	Std. Dev.
Panel A: Atlanta Market										
Size-Units	0.0284	0.0148	0.0424	0.0283	0.0071	0.0202	0.0077	0.0320	0.0201	0.0062
Size-Avg Unit	0.0905	0.0754	0.1058	0.0905	0.0074	0.0855	0.0745	0.0964	0.0855	0.0058
Stories	0.0294	0.0150	0.0433	0.0295	0.0072	0.0139	0.0027	0.0246	0.0138	0.0057
Age	−0.0803	−0.0952	−0.0653	−0.0806	0.0079	−0.0769	−0.0893	−0.0639	−0.0768	0.0067
Rehab	−0.0072	−0.0372	0.0214	−0.0080	0.0150	−0.0078	−0.0312	0.0145	−0.0081	0.0116
σ^2	na	na	na	na	na	0.0155	0.0098	0.0269	0.0163	0.0041
τ^2	0.0186	0.0163	0.0213	0.0186	0.0013	0.0086	0.0069	0.0104	0.0086	0.0009
range	na	na	na	na	na	0.2376	0.1450	0.4756	0.2587	0.0825
μ random	6.6082	6.5405	6.6612	6.6083	0.0291	6.5849	6.4960	6.6462	6.5824	0.0360
υ random	0.0107	0.0051	0.0269	0.0121	0.0062	0.0011	0.0003	0.0045	0.0015	0.0013
Predictive SSE	1.6098					1.0732				
Predictive VAR	1.9386					0.3951				
P(0.10)	0.68					0.73				
P(0.20)	0.90					0.93				

Exhibit 5 | (continued)

Model Information

	OLS					Spatial Model				
	Median	2.5%	97.5%	Mean	Std. Dev.	Median	2.5%	97.5%	Mean	Std. Dev.
Panel B: Boston Market										
Size-Units	0.0074	−0.0218	0.0335	0.0076	0.0150	0.0128	−0.0220	0.0437	0.0125	0.0155
Size-Avg Unit	0.1117	0.0827	0.1379	0.1115	0.0147	0.1158	0.0844	0.1479	0.1160	0.0157
Stories	0.0506	0.0167	0.0832	0.0504	0.0172	0.0459	0.0135	0.0790	0.0459	0.0174
Age	−0.0588	−0.0879	−0.0237	−0.0577	0.0163	−0.0567	−0.0846	−0.0258	−0.0567	0.0153
Rehab	0.0089	−0.0702	0.0892	0.0096	0.0393	0.0123	−0.0611	0.0841	0.0128	0.0372
σ^2	na	na	na	na	na	0.0047	0.0015	0.0161	0.0058	0.0038
τ^2	0.0282	0.0222	0.0373	0.0285	0.0038	0.0237	0.0157	0.0316	0.2362	0.0041
range	na	na	na	na	na	0.0706	0.0235	0.3699	0.1091	0.2052
μ random	7.2528	7.0821	7.4175	7.2459	0.0842	7.2261	7.0609	7.3728	7.2227	0.0795
υ random	0.0563	0.0881	0.1888	0.0691	0.0621	0.0429	0.0147	0.1406	0.0517	0.0350
Predictive SSE	0.8807					0.7500				
Predictive VAR	0.6522					0.1361				
P(0.10)	0.30					0.25				
P(0.20)	0.65					0.65				

Exhibit 5 | (continued)

Model Information

	OLS					Spatial Model				
	Median	2.5%	97.5%	Mean	Std. Dev.	Median	2.5%	97.5%	Mean	Std. Dev.
Panel C: Chicago Market										
Size-Units	0.0290	0.0137	0.0439	0.0890	0.0078	0.0298	0.0125	0.0455	0.0298	0.0083
Size-Avg Unit	0.1549	0.1371	0.1707	0.1546	0.0086	0.1525	0.1369	0.1683	0.1529	0.0081
Stories	0.0681	0.0479	0.0887	0.0679	0.0105	0.0652	0.0440	0.0851	0.0650	0.0108
Age	−0.0210	−0.0397	−0.0025	−0.0210	0.0092	−0.0239	−0.0411	−0.0053	−0.0237	0.0089
Rehab	−0.0002	−0.0311	0.0334	−0.0004	0.0159	0.0042	−0.0249	0.0314	0.0045	0.0145
σ^2	na	na	na	na	na	0.0123	0.0071	0.0278	0.0140	0.0053
τ^2	0.0245	0.0216	0.0275	0.0245	0.0015	0.0181	0.0153	0.0214	0.0181	0.0016
range	na	na	na	na	na	0.1876	0.0786	0.5949	0.2196	0.1343
μ random	6.7370	6.6738	6.7963	6.7371	0.0313	6.6833	6.5819	6.7479	6.6784	0.0424
υ random	0.0195	0.0111	0.0402	0.0212	0.0074	0.0073	0.0013	0.0191	0.0081	0.0046
Predictive SSE	1.1584					0.9766				
Predictive VAR	1.1297					0.2789				
P(0.10)	0.44					0.52				
P(0.20)	0.82					0.86				

Exhibit 5 | (continued)

Model Information

	OLS					Spatial Model				
	Median	2.5%	97.5%	Mean	Std. Dev.	Median	2.5%	97.5%	Mean	Std. Dev.
Panel D: Houston Market										
Size-Units	0.0040	−0.0090	0.0167	0.0036	0.0068	−0.0017	−0.0128	0.0109	−0.0014	0.0063
Size-Avg Unit	0.1212	0.1082	0.1341	0.1214	0.0067	0.1132	0.1011	0.1247	0.1130	0.0060
Stories	0.0308	0.0174	0.0435	0.0307	0.0066	0.0273	0.0138	0.0408	0.0274	0.0068
Age	−0.0784	−0.0922	0.0646	−0.0783	0.0071	−0.8570	−0.0989	−0.7243	−0.0858	0.0068
Rehab	−0.0054	−0.0385	0.0023	−0.0063	0.0145	0.0050	−0.0224	0.0286	0.0044	0.0130
σ^2	na	na	na	na	na	0.0148	0.0099	0.0227	0.0152	0.0032
τ^2	0.0166	0.0143	0.0190	0.0166	0.0012	0.0066	0.0043	0.0094	0.0066	0.0013
range	na	na	na	na	na	0.0774	0.0382	0.1455	0.0830	0.0303
μ random	6.4041	6.3479	6.4584	6.4055	0.0275	6.3999	6.3473	6.4513	6.3997	0.0265
υ random	0.0147	0.0080	0.0307	0.0160	0.0059	0.0073	0.0023	0.0184	0.0081	0.0043
Predictive SSE	1.7771					1.2736				
Predictive VAR	1.7719					0.7097				
P(0.10)	0.52					0.59				
P(0.20)	0.89					0.93				

Exhibit 5 | (continued)

Model Information

	OLS					Spatial Model				
	Median	2.5%	97.5%	Mean	Std. Dev.	Median	2.5%	97.5%	Mean	Std. Dev.
Panel E: Jacksonville Market										
Size-Units	0.0321	0.0066	0.0588	0.0321	0.0136	0.0292	0.0061	0.0547	0.0298	0.0126
Size-Avg Unit	0.1238	0.0978	0.1483	0.1235	0.0133	0.1172	0.0953	0.1390	0.1173	0.0111
Stories	0.0199	−0.0093	0.0460	0.0193	0.0141	0.0277	−0.0018	0.0564	0.0276	0.0142
Age	−0.0753	−0.1073	−0.0463	−0.0758	0.0152	−0.0748	−0.1032	−0.0458	−0.0748	0.0146
Rehab	−0.0052	−0.0584	0.0429	−0.0071	0.0269	−0.0111	−0.0643	0.0382	−0.1292	0.0244
σ^2	na	na	na	na	na	0.0075	0.0029	0.0173	0.0081	0.0036
τ^2	0.0097	0.0068	0.0145	0.0099	0.0019	0.0043	0.0015	0.0083	0.0044	0.0017
range	na	na	na	na	na	0.0616	0.0162	0.3254	0.0897	0.0869
μ random	6.4047	6.3270	6.4679	6.4024	0.0352	6.4091	6.3317	6.4958	6.4106	0.0403
υ random	0.0101	0.0031	0.0337	0.0122	0.0081	0.0079	0.0009	0.0320	0.0100	0.0082
Predictive SSE	0.2882					0.3004				
Predictive VAR	0.2516					0.1246				
P(0.10)	0.55					0.55				
P(0.20)	0.90					0.90				

Exhibit 5 | (continued)

Model Information

	OLS					Spatial Model				
	Median	2.5%	97.5%	Mean	Std. Dev.	Median	2.5%	97.5%	Mean	Std. Dev.
Panel F: Los Angeles Market										
Size-Units	0.0558	0.0396	0.0725	0.0559	0.0085	0.5090	0.0352	0.0670	0.0508	0.0084
Size-Avg Unit	0.1416	0.1254	0.1577	0.1417	0.0083	0.1435	0.1254	0.1586	0.1428	0.0086
Stories	0.0217	0.0034	0.0394	0.0216	0.0094	0.0289	0.0116	0.0450	0.0288	0.0091
Age	−0.0572	−0.0734	−0.0394	−0.0572	0.0087	−0.0586	−0.0741	−0.0406	−0.0581	0.0086
Rehab	0.0307	−0.0022	−0.0636	0.0301	0.0170	0.0262	−0.0017	0.0567	0.0266	0.0152
σ^2	na	na	na	na	na	0.0314	0.0168	0.0583	0.0329	0.0106
τ^2	0.0280	0.0243	0.0322	0.0281	0.0020	0.0185	0.0150	0.0225	0.0186	0.0020
range	na	na	na	na	na	0.1897	0.0943	0.4530	0.2105	0.0908
μ random	6.8670	6.8040	6.9299	6.8675	0.0331	6.8534	6.7498	6.9443	6.8534	0.0473
υ random	0.0325	0.0201	0.0565	0.0340	0.0094	0.0038	0.0040	0.0208	0.0055	0.0057
Predictive SSE	2.4327					2.2571				
Predictive VAR	3.0510					1.0036				
P(0.10)	0.56					0.51				
P(0.20)	0.87					0.87				

Exhibit 5 | (continued)

Model Information

	OLS					Spatial Model				
	Median	2.5%	97.5%	Mean	Std. Dev.	Median	2.5%	97.5%	Mean	Std. Dev.
Panel G: San Diego Market										
Size-Units	0.3690	0.0264	0.0481	0.0370	0.0056	0.0319	0.0220	0.0424	0.0320	0.0055
Size-Avg Unit	0.1023	0.0924	0.0037	0.1028	0.0054	0.0983	0.0885	0.1079	0.0980	0.0051
Stories	0.0162	0.0034	0.0280	0.0158	0.0063	0.0143	0.0025	0.0256	0.0139	0.0062
Age	−0.0457	−0.0559	−0.0353	−0.0467	0.0054	−0.0475	−0.0591	−0.0640	−0.0478	0.0057
Rehab	0.0220	0.0011	0.0430	0.0218	0.0188	0.0132	−0.0054	0.0332	0.0134	0.0099
σ^2	na	na	na	na	na	0.0169	0.0082	0.0381	0.0182	0.0076
τ^2	0.1388	0.0125	0.1580	0.0139	0.0009	0.0084	0.0068	0.0102	0.0084	0.0008
range	na	na	na	na	na	0.2157	0.0860	0.5923	0.2433	0.1264
μ random	6.8820	6.8058	6.9664	6.8814	0.0387	6.8894	6.7689	6.9993	6.8899	0.0534
v random	0.0177	0.0093	0.0468	0.0200	0.0097	0.0039	0.0006	0.0153	0.0049	0.0039
Predictive SSE	0.8529					0.6647				
Predictive VAR	0.7247					0.0143				
P(0.10)	0.62					0.66				
P(0.20)	0.86					0.90				

Exhibit 5 | (continued)

Model Information

	OLS					Spatial Model				
	Median	2.5%	97.5%	Mean	Std. Dev.	Median	2.5%	97.5%	Mean	Std. Dev.
Panel H: San Francisco Market										
Size-Units	0.0182	−0.0064	0.0459	0.0186	0.0141	0.0213	−0.0055	0.0504	0.0213	0.0143
Size-Avg Unit	0.1250	0.0923	0.1571	0.1255	0.0161	0.1244	0.0892	0.1549	0.1238	0.0160
Stories	0.0837	0.0513	0.1145	0.0838	0.0163	0.0722	0.0398	0.1024	0.0715	0.0159
Age	−0.0846	−0.0059	−0.0518	−0.0847	00169	−0.0682	−0.1006	−0.0375	−0.0683	0.0166
Rehab	0.0134	−0.0383	0.0639	0.0130	0.0260	0.0047	−0.0450	0.0545	0.0060	0.0255
σ^2	na	na	na	na	na	0.0160	0.0079	0.0272	0.0164	0.0050
τ^2	0.0290	0.0231	0.0358	0.0291	0.0034	0.0152	0.0086	0.0235	0.0155	0.0038
range	na	na	na	na	na	0.0255	0.0010	0.0891	0.0319	0.0229
μ random	7.3081	7.2313	7.3936	7.3090		7.2916	7.2178	7.3640	7.2910	0.0386
υ random	0.0097	0.0022	0.0416	0.0125		0.0054	0.0010	0.0291	0.0079	0.0082
Predictive SSE	1.0044					0.8818				
Predictive VAR	0.6343					0.2399				
P(0.10)	0.55					0.45				
P(0.20)	0.80					0.80				

significant reductions were obtained in both predictive SSE and predictive VAR, indicating that not only is the predicted value closer to the observed value, but confidence about the prediction is much higher. The increases in the two proportions P(0.10) and P(0.20) also support this conclusion. Finally, the fact that the predictive interval for the REHAB value is centered at zero supports the notion that there is not a measurable rent premium associated with rehabilitating a project in this market.

Chicago is a slow-growing, spatially complex market. And, the very uniquely different spatial patterns of properties in the downtown area versus those in the suburbs—the downtown pattern is very linear, in a north/south direction along Lake Michigan, while the suburban pattern is more dispersed. Despite this, the model generated a spatial range of thirteen miles, and a relatively strong measure for σ^2, capturing approximately 42% of the total error variance. Additionally, significant reductions were obtained in both predictive SSE and predictive VAR, indicating that not only is the predicted value closer to the observed value, but confidence about the prediction is much higher. The increases in the two proportions P(0.10) and P(0.20) also support this conclusion.

Houston is a fast growing market that is political[7] in that there are minimal zoning regulations that would provide for a defined land use plan. This may have had an impact in the relatively short spatial range of about five miles. The spatial model produced a relatively strong measure for σ^2, capturing nearly 70% of the total error variance. Additionally, significant reductions were obtained in both predictive SSE and predictive VAR, indicating that not only is the predicted value closer to the observed value, but confidence about the prediction is much higher. The increases in the two proportions P(0.10) and P(0.20) also support this conclusion.

San Diego is a new market with strong growth. The model produced a spatial range of fifteen miles, and a relatively strong measure for σ^2, which captured approximately 67% of the total error variance. Additionally, significant reductions were obtained in both predictive SSE and predictive VAR, indicating that not only is the predicted value closer to the observed value, but confidence about the prediction is much higher. The increases in the two proportions P(0.10) and P(0.20) also support this conclusion.

Group 2: Marginal Improvement

Boston is a spatially complex, slow growing market, which also suffers from a political component in that it has a history of rent control, which may have resulted in an unnatural spatial arrangement. The sample size is relatively small ($n = 120$). While the spatial model produced a range of nearly five miles, σ^2 did not capture much of the variation in residuals. Additionally, while predictive SSE and predictive VAR experienced improvements, these did not translate into improvements in the proportions P(0.10) and P(0.20), indicating that the spatial model did a marginally better job in prediction. While the lack of spatial structure in this market may be the result of over specification of the mean through indicator variables for each of the submarkets, other possible explanations also include the lingering results of rent control.

Los Angeles is a multi-centered market, with considerable development constraints. The model produced a spatial range of thirteen miles. Additionally, σ^2 captured approximately 60% of the total error variance. The summary statistics provide mixed results. While noticeable improvements were observed in both the predictive SSE and predictive VAR, indicating increased confidence in the predictions, the fact that there were no improvements in the proportions P(0.10) and P(0.20) suggests that the spatial model was not able to predict any values closer to their actual observed, but that it did provide additional confidence in those that were conducted.

Group 3: No Measurable Improvement

Jacksonville is a new, fast-growing market. The spatial model produced a range of about four miles, and σ^2 captured approximately 64% of the total error variance. The submarket dummy variables specified in the mean model seem to have captured much of the spatial structure in the market. As a result, while the predictive VAR did decrease, this was offset by an increase in the predictive SSE. The fact that the two proportions, P(0.10) and P(0.20), remained the same also supports the notion that this market did not benefit much from the spatial model. As with the Boston market, the sample size was relatively small.

San Francisco is spatially complex, which also presents political considerations in both a strong antidevelopment sentiment, and lingering effects of a long period of rent control. While, the unique geographical features of the San Francisco area most likely contributed to the small spatial range of about two miles, the political influences in this market most assuredly also influenced the spatial patterns. The spatial model produced a relatively strong measure for σ^2, capturing nearly 50% of the total error variance. As with Los Angeles, the summary statistics for this market provide mixed results. While noticeable improvements were observed in both predictive SSE and predictive VAR, indicating increased confidence in the predictions, there were no improvements in the proportions P(0.10) and P(0.20).

Residual and Predicted Rent Surfaces

For illustrative purposes, Exhibits 6 and 7 contain residual and predicted rent surfaces for the Boston and Atlanta markets, respectively. Consistent with the results, the rent surface for the Boston markets is markedly different from the residual surface supporting the notion that the submarket random effects in the OLS model captured most of the spatial process in this market. At the same time, the residual and rent surface maps for the Atlanta market provide a clearer spatial process, as there is little difference between the two maps, indicating that even

Exhibit 6 | Boston

Panel A: Residual Surface of Apartment Market

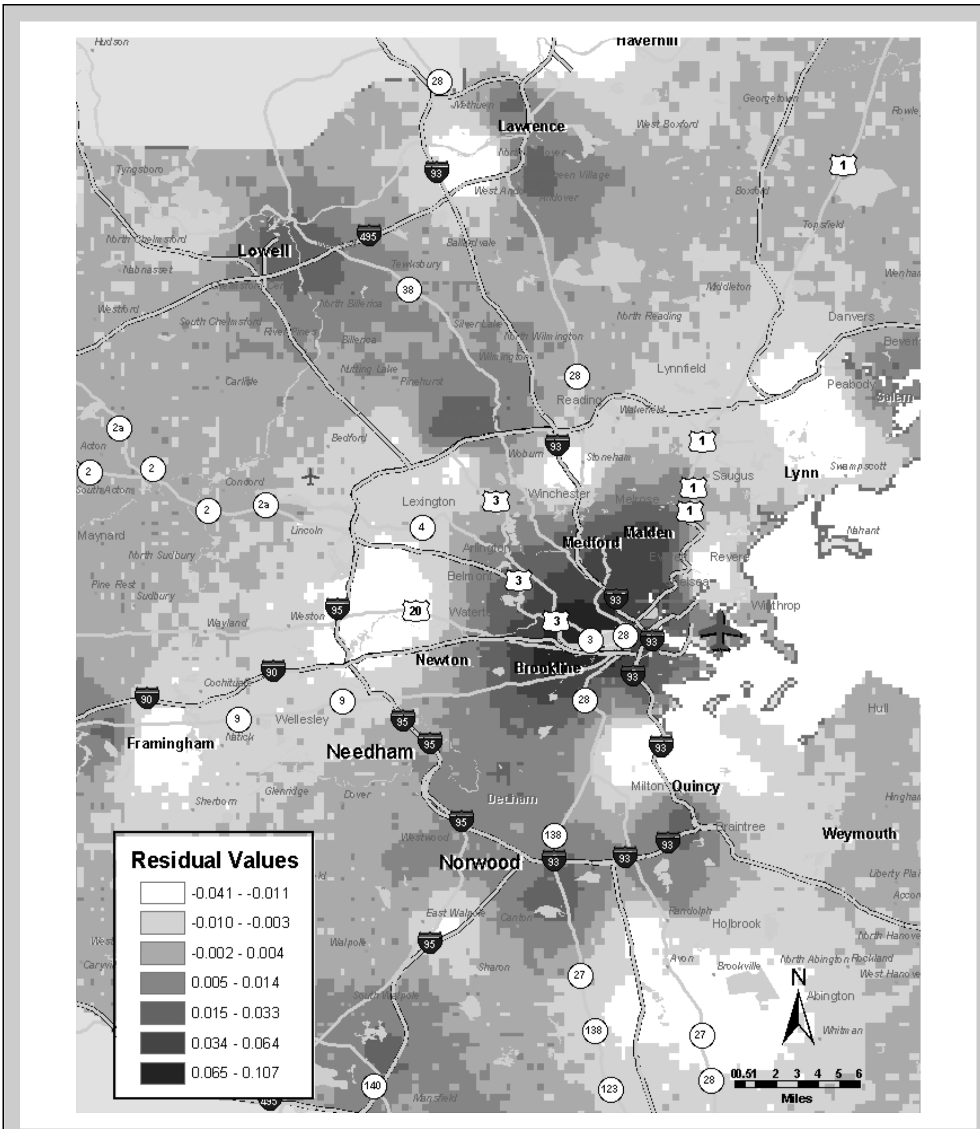


Exhibit 6 | (continued)

Boston

Panel B: Rent Surface of Apartment Market

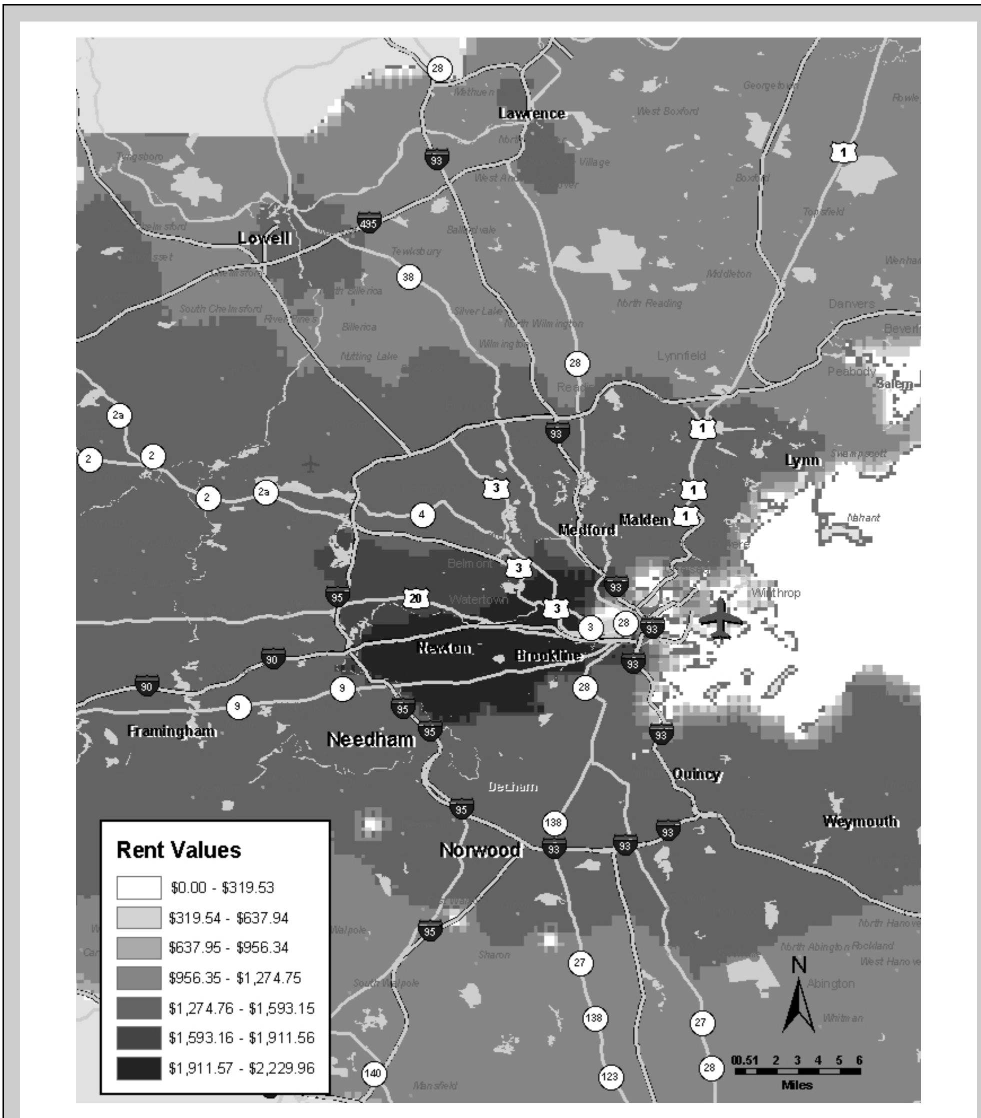


Exhibit 7 | Atlanta

Panel A: Residual Surface of Apartment Market

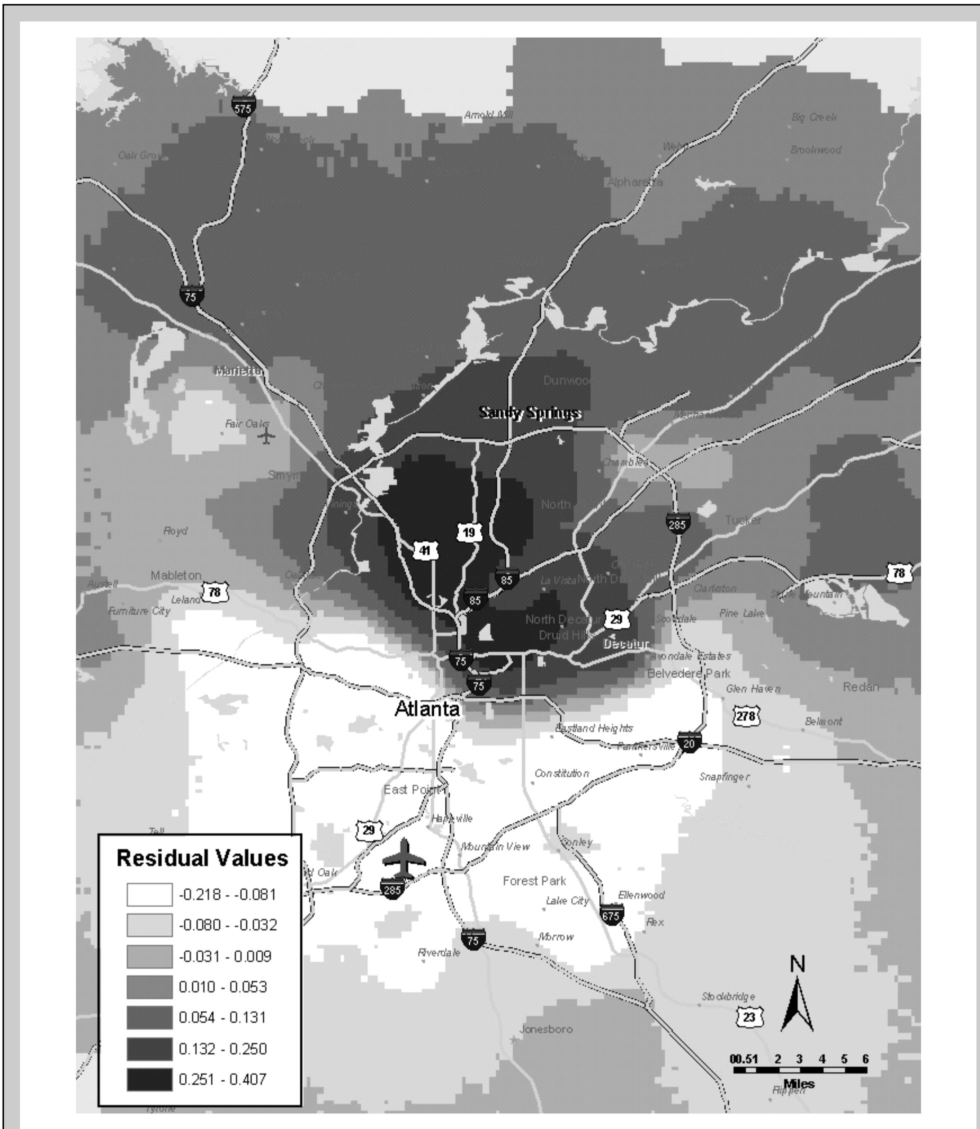
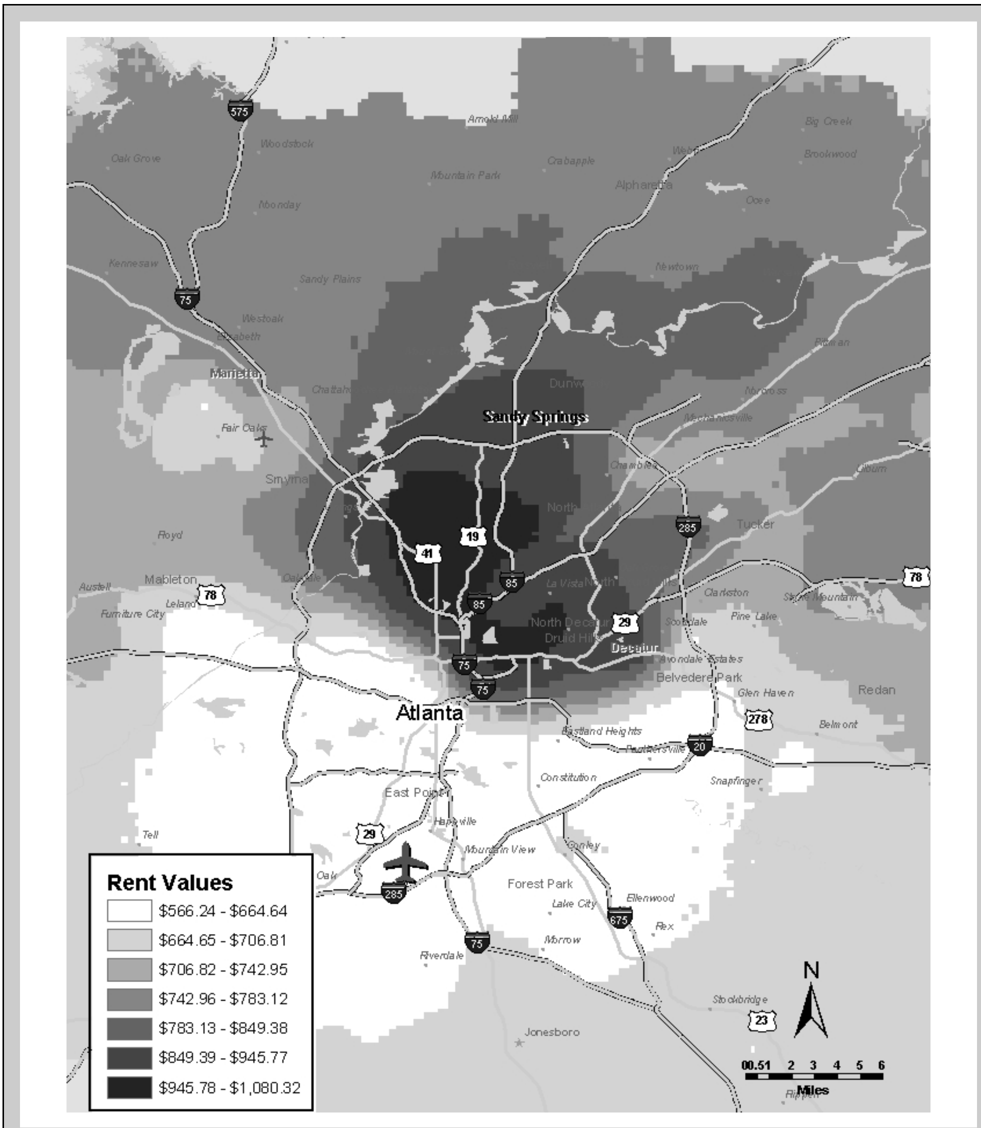


Exhibit 7 | (continued)

Atlanta

Panel B: Rent Surface of Apartment Market



though we included location variables in the mean structure, there was significant local variation left.

Conclusion

The model structure developed in this paper produces significant benefits for out-of-sample rent prediction in the apartment sector, enabling an inference of conceptual rent at any location in a market. This results in a continuous rent surface, allowing for the analysis of areas where rents appear to be elevated versus those were they seem to be lower. Additionally, a framework is provided to help understand the spatial relationships between properties based on the distance from each other, and how this relationship varies across markets. There is still much research to be done, especially in better understanding why the spatial range varies significantly between markets.

Future applications of the model structure should provide significant benefits to the valuation process, and better understanding the potential impact on rents, or value, of proposed new construction. As a result, it should also provide benefits to the property management process. We also encourage the adaptation of the model structure to time series analysis, in that the same spatial associations identified here should also benefit localized rent forecast.

Endnotes

[1] We acknowledge that such an assumption is an oversimplification (*e.g.*, that decay in dependence might change with direction). A sufficiently rich mean structure might mitigate concerns with this simplifying assumption.

[2] Nearest employment center is determined by using the square foot weighted average latitude and longitude coordinates of the office buildings in each of REIS's submarkets.

[3] While the notion of proximity to the CBD is deeply routed in the study of urban economics and economic geography, going back to the initial publication of Von Thunen's Isolated State in 1826 (translated in 1966), and put into the context of residential location patterns relative to the CBD employment center by Alonso (1964), it does not necessarily reflect current multi-nodal employment patterns in markets across the United States.

[4] Data are collected by REIS using primary surveys and secondary sources. REIS's survey department is responsible for the collection of building performance data and for verifying the accuracy of individual property information. Surveyors are responsible for contacting owners, managers and leasing agents to obtain information on property availabilities, rents and lease terms for individual apartment complexes. Necessary to the success of REIS's data collection program are: a system for training surveyors; scripts and surveying techniques that have been refined over a period of years; and the firm's database management personnel, including experts in Oracle® relational database software.

REIS subjects all survey responses to a set of quality assurance and validation processes. For example, to ensure data integrity, data are checked and validated at both the individual building level as well as the aggregate market (*i.e.*, peer group, submarket and metropolitan area) level. At the ''front-end'' of the process, surveyors compare in real time the data reported by building contacts with the previous record for the property and ask follow-up questions to verify any unusual changes in rents or vacancies. On the back-end, automated exception reports are generated to identify properties that deviate materially from peer group and or submarket averages. Follow-up telephone calls are then made for verification of clarification.

5 For a project to be identified as going through a rehab, it must have received substantial improvements. The data used included the year of the rehab, but not any other indicator that would suggest the extent.

6 This should not be taken as a criticism of employing spatial dependence structure. Without introducing spatial dependence, we could not assess whether there is benefit to such modeling.

7 We consider markets such as the Boston and San Francisco political markets because of the presence of long entitlement process, zoning, rent control, etc., which were established by governmental institutions to control the rental market. At the same time, Houston is also a political market in the sense that the local government has made a conscious decision not to enact what would be considered normal controls in most other markets.

References

Alonso, W., *Location and Land Use*, Cambridge, MA: Harvard University Press, 1964.

Anselin, L., Some Robust Approaches to Testing and Estimation in Spatial Econometrics, *Regional Science and Urban Economics*, 1990a, 20, 141–63.

——, Spatial Dependence and Spatial Structural Instability in Applied Regression Analysis, *Journal of Regional Science*, 1990b, 30:2, 185–207.

Asabere, P. K. and F. E. Huffman, Thoroughfares and Apartment Values, *Journal of Real Estate Research*, 1996, 12:1, 9–16.

Bailey, T. C. and A. C. Gatrell, *Interactive Spatial Data Analysis*, Essex, England: Pearson Education Limited, 1995.

Cressie N. A. C., *Statistics for Spatial Data*, New York, NY: John Wiley & Sons, Inc., 1993.

Des Rosiers, F. and M. Thériault, Rental Amenities and the Stability of Hedonic Prices: A Comparative Analysis of Five Market Segments, *Journal of Real Estate Research*, 1996, 12:1, 17–35.

Dubin, R. A. and C. H. Sung, Spatial Variation in the Price of Housing: Rent Gradients in Nonmonocentric Cities, *Urban Studies*, 1987, 24, 193–204.

Fotheringham A. S. and P. A. Rogerson, GIS and Spatial Analytical Problems, *International Journal of Geographic Information Systems*, 1993, 7, 3–19.

Fotheringham, A. S., C. Brunsdon and M. Charlton, *Quantitative Geography*, London, England: SAGE Publications, 2000.

Frew, J. and B. Wilson, Estimating the Connection between Location and Property Value, *Journal of Real Estate Practice and Education*, 2002, 5:1, 17–25.

Guntermann, K. L. and S. Norrbin, Explaining the Variability of Apartment Rents, *Journal of the American Real Estate and Urban Economics Association*, 1997, 15:4, 321–40.

Haining, R., *Spatial Data Analysis in the Social and Environmental Sciences*, Cambridge, England: Cambridge University Press, 1990.

Harrison, D. and D. L. Rubinfeld, Hedonic Housing Prices and the Demand for Clean Air, *Journal of Environmental Economics and Management*, 1978, 5:1, 81–102.

Pace, R. K. and O. W. Gilley, Using the Spatial Configuration of the Data to Improve Estimation, *Journal of Real Estate Finance and Economics*, 1997, 14:3, 333–40.

Pace, R. K., R. Barry and C. F. Sirmans, Spatial Statistics and Real Estate, *Journal of Real Estate Finance and Economics*, 1998, 17:1, 5–13.

Sirmans, G. S., C. F. Sirmans and J. Benjamin, Examining the Variability of Apartment Rent, *Real Estate Appraiser & Analyst*, 1990, summer, 43–8.

Sirmans, G. S. and J. Benjamin, Determinants of Market Rent, *Journal of Real Estate Research*, 1991, 6:3, 91–8.

Smith, C. A. and M. J. Kroll, Utility Theory and Rent Optimization: Utilizing Cluster Analysis to Segment Rental Markets, *Journal of Real Estate Research*, 1989, 4:1, 61–71.

Von Thunen's Isolated State, P. Hall (Ed.,), New York, NY: Pergamon Press, 1966.

James Valente, ING Clarion, New York, NY 10169 or james.valente@ingclarion.com

ShanShan Wu, ING Clarion, New York, NY 10169 or shanshan.wu@ingclarion.com.

Alan Gelfand, Duke University, Durham, NC 27708 or alan@stat.duke.edu.

C. F. Sirmans, University of Connecticut, Storrs, CT 06269-1041 or cf@sba. uconn.edu.