A Spatial Autoregressive Specification with a Comparable Sales Weighting Scheme

Author Claude Besner

Abstract This research incorporates a Spatial Autoregressive Variable with Similarity components (SARS) within a traditional hedonic model. The behavior of economic agents and the spatial dependence price structure are linked to the real estate appraisal paradigm. The SARS variable's similarity components generate anisotropies that deform concentric circles of spatial dependence so as to designate the influence exerted by "comparables." The incorporation of similarity components improves the predictive capacity and reduces the spatial dependence among residuals in the SAR model. The research determines for the Montreal Urban Community the underlying distance parameters of spatial dependence as well as anisotropic factors specific to price interdependence for two single-family house archetypes: the condominium and the individual house.

Introduction

Hedonic residential analysis lends support to the fact that the utility—and thus the value—of a house is a function of the utility of its various characteristics. The traditional econometric approach, which makes it possible to estimate the value of each characteristic, is subject to methodological and statistical problems resulting from the spatial dependence of errors.

Spatial autocorrelation of residential prices relates to a situation where the price of a house at one location is correlated with the price of neighboring houses. This dependence originates in part from the fact that each house shares with its neighbors influences from location factors that are nearly identical. This study assumes that residential price autocorrelation can be more generally explained by the behavior of economic agents and therefore by the real estate pricing mechanism.

When making decisions, each economic actor (seller, buyer and any representative thereof) takes the value of neighboring residences into consideration. They use information from comparable sales that took place in the neighborhood of the house to evaluate. This research assumes that this behavior structures the interdependence of residential prices (in parallel and beyond the strictly spatial interdependence resulting from the sharing of characteristics associated with a common location). The dependence phenomenon thus appears to be multidimensional and is measured by an autocorrelation function that incorporates distance and similarity components.

This study adds to a traditional hedonic model a spatial lag variable, "SARS," that takes observation similarities into account. The SAR model (Anselin, 1988) is a special case of the more general SARS model, where the influence of the degree of similarity is nil.

Literature Review

Spatial dependence among hedonic regression residuals was initially revealed by Brigham (1965), who carried out topographic error projections, and then underscored by Ball (1973) and Richardson (1974). In 1975, Sibert built a model of residential values based on spatial autocorrelation, but did not adopt the hedonic conceptual framework. Jackson (1979) seemed to be the first to suggest using a regression model with an autoregressive structure, although he did not specify it. Anas and Eum (1984) assumed the absence of spatial autocorrelation but implicitly used a spatial autoregressive term. They used the most recent nearby sale as a temporal proxy. Dubin (1988) carried out a formal verification of the existence of spatial dependence among the hedonic regression error terms.

Since the end of the 1980s, there has been a marked increase in studies highlighting concerns about spatial autocorrelation: Goodman (1989), Des Rosiers (1992), Waddell, Berry and Hoch (1993), Maylere (1995), Rodriguez, Sirmans and Marks (1995) and Wiltshaw (1996). Anselin (1984, 1988), Odland (1988) and Haining (1990), spatial modeling specialists, illustrated their views with the help of residential real estate prices.

Can (1989, 1990, 1992) appears to be the first to propose a systematic procedure within the framework of hedonic analysis for verifying and specifying the phenomenon by adding a spatial autoregressive variable. Can and Megbolugbe (1997) followed the same path while Pace and Gilley (1997) and Pace, Barry, Clapp, et al. (1998) adopted the same conceptual framework. Pace, Sirmans and Slawson (2001) impose a set of restrictions on the spatial weight matrix and find results in accordance with the present research.

Dubin, Pace and Thibodeau (1999) discuss alternative spatial autoregression model specifications. Some authors introduced spatial dependence effects in a statistical perspective (geostatistical models) in contrast to the econometric approach (lattice models) adopted by this study. Dubin (1992, 1998), Olmo (1995), as well as Basu and Thibodeau (1998) used the geostatistic technique of kringing (for more on this technique, see Ripley, 1981; Lam, 1983; and Haining, 1990). The approach in this study follows that of Anselin (1988) and Can (1989, 1990, 1992).

About the Time Dimension

The temporal vector seems to cause an asymmetrical relationship among residential prices. In other words, the influence of the selling price of a nearby house j on a buyer's bid for a house i would only be exerted if the sale of j precedes the purchase of i. Should this be the case, the absence of the temporal relationship introduces a specification bias that distorts the causal relationship. While Can (1989, 1992) does not take this asymmetry into account, Can and Megbolugbe (1997) and Pace, Barry, Clapp, et al (1998) do condition predictions strictly on past data, but doing so reduces the number of potential observations used to compute spatial dependence. Such asymmetric conditioning can aid forecasting. However, if the locational premia and discounts remain the same over the time period, the use of both past and future observations can provide a better estimate of the spatial dependence. The SARS model suggests using all available information; the matrix W thus does not formally respect the temporal "constraint." The *SARS* variable indirectly incorporates the temporal dimension by considering the local similarity of the temporal adjustment factor (*NBMONTH*). The incorporation of this temporal variable into the similarity component of W creates anisotropies based on the modification of local factors in time.

Modeling

This research uses an autoregressive simultaneous model inspired by the seminal work of Whittle (1954).

$$Y_i = \rho W Y_j + \mu, \tag{1}$$

where WY_j denotes the vector of weights based on the value of neighboring observations, ρ is the coefficient of the autocorrelation variable and μ the error term. The dependent variable Y_i depends on the neighboring values $Y_j (i \neq j)$. The addition of exogenous variable vectors to Whittle's equation produces a linear autoregressive hedonic model:

$$Y = \alpha + \rho W Y + X\beta + \mu, \tag{2}$$

where Y represents the price of a house; X a matrix of $n \times K$ observations of the structural, locational and temporal exogenous explanatory variables; β the parameters of characteristics k; a is a constant; μ an $n \times 1$ vector of normal *iid* errors; WY is a vector of $n \times 1$ weight factors constructed from price and nearby house attributes, the autocorrelation variable; and ρ represents the autocorrelation variable coefficient. The effects of spatial dependence can also be controlled by error modeling, but the results are less conclusive (Besner, 1999), a slightly better fit is produced by the mixed regressive autoregressive model of Equation (2).

Three models are considered. The research first estimate a traditional (TRAD) hedonic model (without an autoregressive variable). The model uses a limited number of variables, those most frequently used within the framework of hedonic analyses (Des Rosiers, 1988). A temporal adjustment variable is added and also one or two dichotomous location variables to indicate inclusion in a specific municipality. The TRAD model thus includes eight or nine variables, according to the sample; the fifty-three hedonic models studied by Des Rosiers (1988) include eight variables on average (the median is seven).

Second, the base model is augmented by a SAR autoregressive term (SARS variable without the similarity component). At this point, only the distance component is used to build the weight matrix *W*.

Third, the construction of a SARS model allows an evaluation of the specific contribution made by inserting a similarity component into the matrix *W*.

Functional Form and Estimator

The research chooses the semi-log functional form. From a statistical point of view, this transformation corrects the asymmetrical distribution of the price variable. The normal (or quasi-normal) distribution obtained supports the normality and homoscedasticity of the residuals. The second aspect, which justifies adopting a semi-logarithmic functional form for the dependent variable, concerns the relatively direct interpretation of regression coefficients.

Estimating a simultaneous model requires a nonlinear optimization. Anselin (1995) adapts the maximum likelihood (ML) estimator to the spatial case, by adding a term that introduces eigenvalues ω_i of the weight matrix, to obtain:

$$LIK = \Sigma_i \ln(1 - \rho\omega_i) - \left(\frac{n}{2}\right)\ln(2\pi) - \left(\frac{n}{2}\right)\ln(\sigma^2) - (Y - \rho WY - X\beta)'(Y - \rho WY - X\beta)/2\sigma^2, \quad (3)$$

where n is the number of observations in the sample and $\sigma^2 = SSE/n$, *SSE* is the Sum of Squared Errors.

Specification of the Weight Matrix

The "Distance" component of matrix *W* has a decreasing exponential effect within the limits of a given range $D_{\text{threshold}}$:

$$w_{ij} = (Distance_{ij})^{-m} \text{ with } m \geq 0 \text{ and } j \in v_i, \tag{4}$$

where w_{ij} is the weight between observation i and a neighbor j, v_i denotes all the houses $j \in$ of the vicinity formed by $D_{\text{threshold}}$. The research considers different values of m and $D_{\text{threshold}}$ to find the best weighting scheme. The weighted function of the influence of nearby houses is written:

$$WY_i = \sum_{j \in v_i} Y_j(w_{ij}/\Sigma w_{ij}), \tag{5}$$

where WY_i represents the average weighted price of houses close to house i, Y_j the transaction price of the nearby house j, v_i all houses j considered to be neighbors to house i, $1/\Sigma wij$ a normalization factor.[1]

Also, the more house j is similar to house i, the more its price influences i. The similarity can be defined as follows:

$$S_{kij} = |X_{ki} - X_{kj}|. \tag{6}$$

A similarity component is calculated for each characteristic k ($k = 1$ to K) of the exogenous explanatory vector X. The similarity components act as local anisotropy generators within the weighted matrix. The combined effects of distance and similarity give:

$$WY_i = \sum_{j \in v_i} Y_j \frac{(D_{ij}^{-m} S_{k1ij}^{-q_{k_1}} S_{k2ij}^{-q_{k_2}} \cdots S_{Kij}^{-q_K})^2}{\Sigma w_{ij}}, \tag{7}$$

where Wi represents the average price of houses neighboring i weighted by distance and similarity, q is an exponential factor which varies from zero to 1 in 0.1 increment, z is a general exponential factor for fine-tuning the multiplicative design that permits trading off the effect of distance and similarity.

The weight of a neighboring observation j is a function of the inverse deviation between subject property i and j. The research sets $S_{kij} = 0.1$ when $X_k - X_{kj} = 0$, the weight of any neighbor j, measured on the basis of a dichotomous variable, is then (1/1) or (1/0.1). For example, suppose $i \in$ municipality A then: $j_1 \in A$ has a weight ten times greater than $j_2 \notin A$. The exponential factor q provides for weight modulation, for example: if $q = 0.6$, the weight of j_1 will only be

$(1/0.1)^{0.6}$, four times greater. The same principle applies for nondichotomous variables.

The research postulates that the matrix of the model, which best fits the sample on the basis of the *LIK* criterion, best reflects the spatial dependence structure. The search process for the best matrix supposed that proximity has preponderance over similarity, *i.e.*, in the small vicinity surrounding the house, the micro-neighborhood's spatial factors measured by proximity are statistically more important than the "local similarity"[2] of the houses themselves. The similarity components are therefore conceptualized as anisotropic factors altering the main spatial process. The search protocol for the best weighting scheme consists of three principal steps. First, the best $D_{\text{threshold}}$ and distance exponential factor m is determined by trying different combination of those factors.

Second, each similarity component variable is determined individually. The distance threshold, determined beforehand, indicates the observations for which the degree of similarity is measured. In turn, each variable k is used to build a matrix W. Each matrix multiplies the similarity component S_k by the previously determined distance component; the best weighting scheme for S_k is found using the exponential factor q. The research makes eleven estimates to determine the best value of q.

Third, five different values of the exponential factor z are considered within a matrix W incorporating the predetermined distance component multiplied by each similarity component S_k. Only those S_k that improve the fit of the model are retained during this last series of estimates. If there is no similar house near the house to be appraised, the appraiser suggests a price adjustment in accordance with the conformity principle (Appraisal Institute, 1992; Boyce, 1975). The conformity principle states that a luxurious house built in a sector that is not so, undergoes a de facto depreciation. The opposite reasoning applies to a modest house located in an affluent district.[3] The equations defining the matrix W in the SARS model respect this principle.

This research uses two specialized software: Spacestat™ version 1.8© (Anselin, 1995) and Mathsoft S-Plus™ (2000). Routines programmed in the "S-Plus" language automate: (1) the construction of the matrices W; (2) the estimations of *LIK*; and (3) the search process for the matrices. For the most part, the coding developed is based on articles by Anselin and Hudak (1992), Anselin, Hudak and Dodson (1992) and the S-Plus programming handbook. The program uses standard nonlinear optimization functions (quasi-Newton method). Each estimate takes around seven minutes, including the construction of W (routines programmed in S-Plus). The research systematically re-estimates the best models using the Spacestat™ software in which a bisection search is used to identify the rho parameter. This redundancy validates the estimates achieved with the program specifically designed for this research. A similar redundancy is created for the matrix construction.

Data and Results

Data were extracted from the database of the Montreal Urban Community (MUC), which is used to establish the assessment roll for municipalities in the urban community. The data includes the x and y coordinates for the the centroid of each sale parcel. This database is of very good quality and compares favorably with other data sources most often used, such as Multiple Listing Services (MLS) (see comparative study by Hamilton and Dale-Johnson, 1991).

Two samples were tested: the one nearest to the Central Business District (CBD) is composed exclusively of condominiums (*CONDO*). The other sample (*WEST*) is made up only of individual single-family dwellings. Exhibit 1 shows characteristics of these samples and Exhibit 2 gives descriptive statistics.

The Best Weighting Scheme

The search for the best distance component requires an exploration of both the axes defining it: the distance threshold and the exponential factor m from Equation (7) (see Exhibit 3).

The results show that the distance threshold must undoubtedly be higher than the minimal distance (D_{min}), defined as the shortest radius that allows each house i to have at least one neighbor j. The counter-performance of the matrices where $D_{threshold} < D_{min}$ clearly illustrates this. In short, as long as the chosen radius allows each house i to have at least one neighbor j, the shortest distance is generally the best, while the choice of the negative exponential to apply to the distance (*e.g.*, inverse distance or inverse square distance) is apparently not critical.

The distance component parameters are fixed as follows: $D_{threshold}$ = 400 meters for both samples,[4] $m = 1$ for the *CONDO* sample and $m = 1.5$ for the *WEST* sample. The search for the best matrix is pursued by introducing the similarity

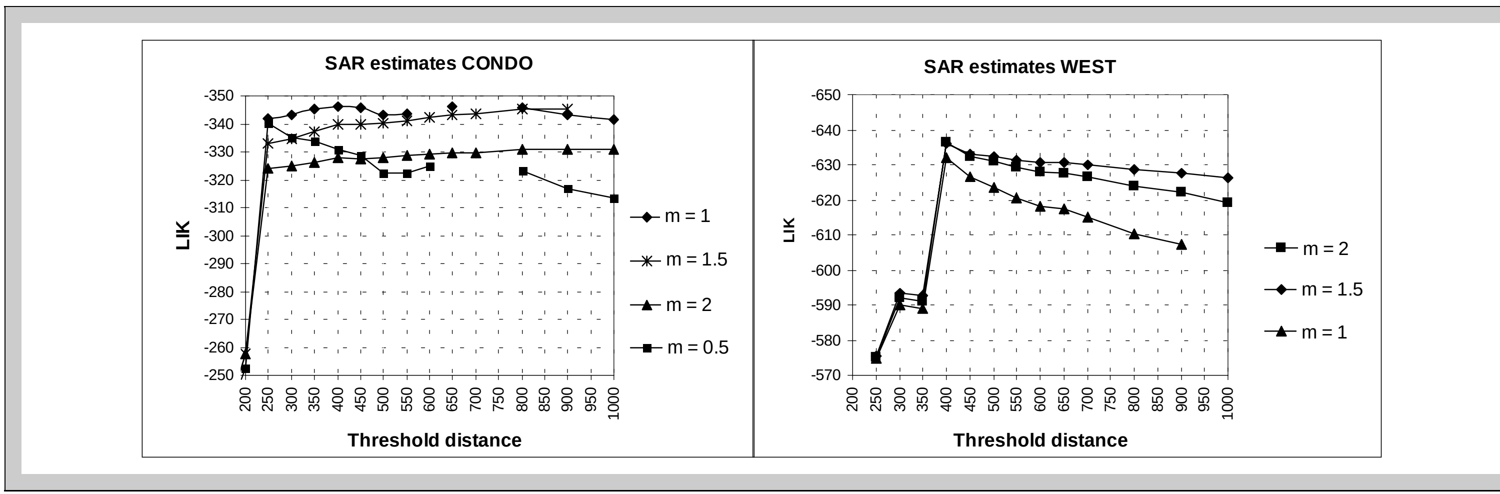
Exhibit 1 | Characteristics of the Samples

Characteristics	*CONDO*	*WEST*
Number of observations	1,008	959
Mean distance to CBD	3 km	24 km
Sample district surface	12.5 km	24 km
Approximate density of sales	80/km	40/km
Temporal boundary	93–96	95–96

Exhibit 2 | Descriptive Statistics

Variable	Description	Mean	Std. Dev.	Min.	Max.
Panel A: *CONDO*					
LIVAREA	Living Area S.F.	975.75	289.99	181.00	2,250.00
MUN17	Municipality 17	0.15	0.35	0.00	1.00
INTGARAGE	Interior Garage S.F	0.07	0.25	0.00	1.00
NBMONTH	*Months from transac. to 98*	34.13	13.57	12.00	60.00
BASEMNT	*Basement S.F.*	6.29	72.16	0.00	1,246.00
3RDSTORY	Condo on third story	0.29	0.45	0.00	1.00
4UPSTORY	*Condo 4 story up*	*0.09*	*0.29*	*0.00*	*1.00*
FIREPLACE	Fireplace	0.05	0.23	0.00	1.00
LNPRICE	Log of price	11.51	0.34	10.42	12.68
PRICE	Selling price	105,455.82	36,354.19	33,350.00	320,000.00
Panel B: *WEST*					
LOTSIZE	Lot size S.F.	7,920.81	2,754.51	2,000.00	24,890.00
NBTOILET	Number of toilets	2.31	0.78	1.00	5.00
FIREPLACE	Fireplace	0.86	0.57	0.00	3.00
INTGARAGE	Interior Garage S.F	141.28	185.59	0.00	753.00
AGE	Years from origin	32.00	13.11	0.00	73.00
MUN10	Municipality 10	0.13	0.34	0.00	1.00
MUN20	Municipality 20	0.35	0.48	0.00	1.00
LIVAREA	Living area S.F.	1,788.07	598.76	576.00	4,228.00
NBMONTH	Fireplace	22.88	6.83	12.00	36.00
LNPRICE	Log of price	11.88	0.31	10.97	13.03
PRICE	Selling price	151,602.66	51,682.17	58,000.000	456,500.00

Exhibit 3 | SAR Estimates



Each point represents a different estimate.

components. This time each estimate uses a matrix *W* with the same distance component, but with a variant of the similarity component. The latter successively incorporates each variable (one at a time).

Exhibit 4 shows that only two variables (*LIVAREA* and *NBMONTH*) contribute to improve the representation of the spatial dependence structure for the *CONDO* sample. For the *WEST* sample, the similarity component highlights the dominating effect of variable *AGE*. Most interestingly, two other samples of individual houses show the same result in Besner (1999).

The research verifies that it can, with a high degree of accuracy ($R^2 > .95$), adjust a second-degree polynomial curve to the estimation points related to each similarity component variables in Exhibit 4. For the *CONDO* and *WEST* samples, the final matrix *W* specification is defined respectively by Equation (8) and (9).

In order to verify whether or not these parameters do relate to the best weighting scheme, the distance threshold and each parameter are modified, one by one. Each of these attempts reduces the LIK criterion value. This confirms, in an imperfect manner, the equation's relevance. It should however be noted that Equations (8) and (9) undoubtedly constitute a "local minimum" on the surface of possible matrices (Besner, 1999).[5]

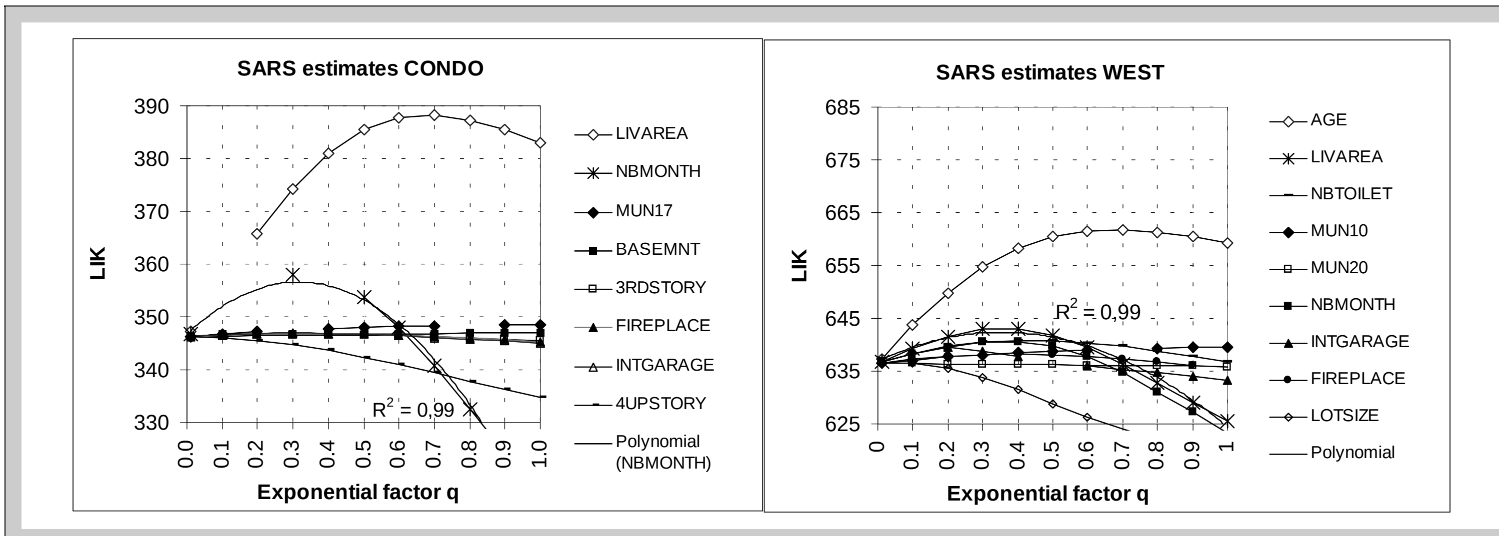
$$WY_i = \sum_{j \in v_i = 400} Y_j \frac{\begin{bmatrix} D_{ij}^{-1}(|LIVAREA_i - LIVAREA_j|)^{-0.7} \\ (|NBMONTH_i - NBMONTH_j|)^{-0.3} \end{bmatrix}^{0.9}}{\Sigma\, w_{ij}}. \quad (8)$$

$$WY_i = \sum_{j \in v_i = 400} Y_j \frac{\begin{bmatrix} D_{ij}^{-1.5}\,(|AGE_i - AGE_j|)^{-0.7} \\ (|LIVAREA_i - LIVAREA_j|)^{-0.4} \end{bmatrix}^{0.8}}{\Sigma w_{ij}} \quad (9)$$

TRAD, SAR and SARS Models

Exhibit 5 shows that the coefficients for the traditional specification all have positive signs, consistent with the theory. The positive sign of the *NBMONTH* variable corresponds well with the weak deflation of residential real estate prices in the area during the 1990s. The autoregressive term improves the prediction capacity; the LIK of the SAR and SARS models for the *CONDO* sample increased by 54% and 76% respectively, compared to the traditional model's index.

Exhibit 4 | SARS Estimates



Each point represents a different estimate. The R^2 is for a second-degree polynomial curve that fits the estimation points.

Exhibit 5 | Regression Statistics

	CONDO Models			*WEST* Models		
INDICE	TRAD	SAR	SARS	TRAD	SAR	SARS
R^2	0.683			0.810		
Adj. R^2	0.681			0.809		
LIK	225.122	346.136	397.428	570.677	636.509	667.978
AIC	−432.244	−672.273	−770.857	−1,121.350	−1,251.020	−1,309.960
Cor^2	0.683	0.757	0.793	0.810	0.838	0.851
Panel A: Parameter Estimates						
Variables						
CONSTANT	10.5730	5.2058	4.8544	11.0620	7.2566	6.7193
WPRICE	na	0.4874	0.5261	na	0.3275	0.3783
LIVAREA	0.0008	0.0006	0.0005	0.0003	0.0020	0.0002
LOTSIZE				0.0000	0.0000	0.0000
BASEMNT	0.0004	0.0030	0.0003			
3RDSTORY	0.0667	0.0703	0.0661			
4UPSTORY	0.0471	0.0628				
	(0.0302)					
FIREPLACE	0.1162	0.0752	0.0600	0.0607	0.0491	0.0508
INTGARAGE	0.0135	0.0200	0.1086	0.0001	0.0001	0.0001
NBTOILET				0.0574	0.0507	0.0430
AGE				−0.0027	−0.0020	−0.0019
NBMONTH	0.0025	0.0020	0.0016	0.0042	0.0042	0.0044
MUN17	0.1749	0.0365	0.0410			
		(0.0525)	(0.0128)			
MUN10				−0.7769	−0.0536	−0.0415
MUN20				−0.0673	−0.0046	0.0050
					(0.7720)	(0.6588)

Notes: Brackets indicate the probability associated with an asymptotic t test; absence of brackets means a near zero probability.

The total improvement corresponds to 11 percentage points on the more familiar square correlation scale.[6] For the *WEST* sample, the predictive gain is also substantial if one takes into account the already high precision of the TRAD model.

The introduction of autoregressive variables influences the coefficients of the other variables, but they all remain significant at the 1% threshold, except for the coefficient *MUN17* variable from the *CONDO* sample and MUN20 from the WEST sample. The vicinity segments (centered on each observation) defined by *W* compete with these geographical segments. The values for most of the other coefficients also decrease, but it should be noted that the coefficient of *4UPSTORY* increases slightly.

The Likelihood Ratio (LR) test (Anselin, 1995) on ρ from Exhibit 6 confirms the significant contribution made by the autoregressive variables. The Lagrange Multiplier test (Anselin, 1995) for spatial error autocorrelation shows that the SARS model clearly reduces spatial dependence. While there is still autocorrelation in the residuals of the SAR models and the SARS *CONDO* model, the value of the Lagrangian statistic is systematically smaller for the SARS models. The statistics for heteroscedasticity state however that the effects of heterogeneity remain partially uncontrolled.

Out-of-Sample Data

The predictive performance of the models is verified on out-of-sample data. The models were constructed using data (in-sample) from sales before January 1, 1997. Additional data (out-of-sample) originate from the same source, but were taken from sales that occurred between January 1 and June 30, 1997. The price prediction for out-of-sample houses is calculated by applying the linear regression equation to the vector that defines characteristics of the residence to be estimated. The price prediction using a SAR or SARS model implies that the matrix *W* is calculated beforehand. So that no out-of-sample data intervenes in the calculations, *W* is constructed using only in-sample observations that are potentially close to the out-of-sample house being estimated. The product of the lag value by the coefficient ρ is then added to the other products.

The concordance of the predictive capacity of out-of-sample data and that of in-sample data confirms the model's potential for generalization (Exhibit 7).

Interpretation of Results

The autoregressive term significantly improves the predictive capacity of the models. Moreover, results show that the SARS model systematically allows for more precise statistical fitting and for better control of the spatial dependence phenomenon within the hedonic model residuals than does the SAR model. Spatial dependence is not limited to the sharing influence of location factors from nearby houses. The structure of spatial dependence is influenced by other factors. The results thus confirm certain fundamental premises of the real estate appraisal paradigm.

Indirectly, this research raises questions regarding conclusions made by numerous research projects in which traditional hedonic models were used. It does seem more difficult to demonstrate the significant influence of any one variable, an economic externality for example, when controlling spatial dependence effects.

Exhibit 6 | Models Diagnostics

Breusch-Pagan		SAR		SARS	
		Test	Prob.	Test	Prob.
		18.740	0.016	20.172	0.010
Panel A: SAR & SARS *CONDO* Models					
L-R Ratio					
Distance threshold	m				
Dt400	1	242.029	0.000	344.613	0.000
Distance threshold	m				
Dt400	1	33.343	0.000	18.472	0.000
Dt400 sim	1	109.317	0.000	11.849	0.001
Dt500	1	33.283	0.000	20.218	0.000
Dt600	1	33.542	0.000	23.096	0.000
Dt700	1	36.820	0.000	27.912	0.000
Dt800	1	36.282	0.000	29.506	0.000
Dt900	1	35.769	0.000	29.611	0.000
Dt1000	1	36.408	0.000	30.476	0.000
		SAR		SARS	
		Test	Prob.	Test	Prob.
		27.007	0.001	23.698	0.005
Panel B: SAR & SARS *WEST* Models					
L-R ratio					
Distance threshold	m				
Dt400	1.5	131.665	0.000	194.603	0.000
Distance threshold					
Dt400	1.5	11.588	0.001	1.211	0.271
Dt400 sim	1.5	47.847	0.000	4.548	0.033
Dt500	1.5	14.549	0.000	1.693	0.193
Dt600	1.5	13.760	0.000	2.297	0.130
Dt700	1.5	15.316	0.000	3.064	0.080
Dt800	1.5	16.015	0.000	3.292	0.070
Dt900	1.5	18.058	0.000	4.262	0.039
Dt1000	1.5	19.428	0.000	4.815	0.028

Notes: The "sim" letters refer to the matrix that incorporates similarity components.

Exhibit 7 | Validation from Out-of-Sample Data

	CONDO			*WEST*		
	TRAD	SAR	SARS	TRAD	SAR	SARS
Panel A: Squared correlation index						
In sample	0.683	0.757	0.793	0.810	0.838	0.851
Out of sample	0.686	0.757	0.780	0.736	0.777	0.758
Panel B: Median absolute % error						
In sample (%)	12.29	10.44	9.54	8.71	8.54	8.10
Out of sample (%)	13.37	12.25	11.15	9.11	5.89	8.07

Notes: CONDO: n = 283; *WEST: n* = 529. In sample: data used to construct the model.

The theoretical search process used to determine the best weighting scheme supposed that proximity has preponderance over similarity when considering the spatial dependence of residential prices. The results confirm this hypothesis. It is also interesting to point out the marked difference existing between the *CONDO* and the individual dwelling segments. Research reveals the variables that characterize these groups both *globally* and *locally*: the significant regression variables and the variables incorporated within similarity components respectively.

In addition to the distance separating observations from each other, only some variables proved to be useful for representing the structure of spatial dependence. The *LIVAREA* variable clearly dominates in the case of *CONDO*s, while it plays only a secondary role in the other sample. Location and structural factors condition the local homogeneity of this variable for single houses. Local similarity of floor area in individual units limits the existence of a strong dependence link. Two nearby houses rarely have floor areas that differ from those of other neighboring homes; while this situation often exists in the case of *CONDO*s.

The *AGE* variable constitutes the most effective similarity measure for representing the spatial dependence of individual house prices. It breaks up locally the space continuity built up over the years. The results thus reminds us about the qualitative nature of this variable. Residences built during a particular period share specific technological and architectural features. Construction, at various different periods, of small local series of houses with similar physical characteristics creates a mosaic of micro-districts, each with their own specific qualities.

Limits and Extensions of Research

This research limited the search process of the best matrices to the degree of similarity of significant variables in the model. Other variables could possibly contribute to defining the structure of spatial dependence. Other regression variables might still have ''local'' impact while not being significant at the ''global'' level.

The stability, or even stagnation, of the Montreal market does not allow for a complete verification of the SARS model's full theoretical potential. It would be very interesting, for example, to test the SARS model on data originating from a sector experiencing strong local ups and downs. In theory, the model should automatically adapt to the local spatio-temporal context through incorporation of the temporal adjustment variable within the similarity component.

Conclusion

This study has suggested a new hedonic residential model with a spatial lag variable that incorporates a similarity component (SARS). This variable reflects an important facet of the mechanism of real estate pricing. The SARS model enhances the predictive power and offers better control over spatial autocorrelation in residuals compared to the basic SAR model. The SARS variable, a factor of local quality, translates both physical and locational characteristics of a neighborhood centered on each observation. Within the framework of hedonic analysis, the coefficient of the autoregressive variable estimates the implicit value of this local quality factor.

Traditional regression models do not explicitly translate the particular interest that real estate appraisers attach to comparable neighboring houses. The SARS model fills this gap and consequently validates certain aspects of the paradigm. Moreover, the analysis reconfirms some of the most important principles associated with classic location theories, especially those put forth by sociologists. The particular nature of residential real estate and the importance of historical factors are also brought into play during the analysis of the local influences on prices. Furthermore, the latter point brings out the qualitative nature of variables representing the age of a house.

The SARS variable constitutes the key element of a parsimonious specification, since by the introduction of a ''proxy'' variable (whose construction can be automated), it allows for ''measuring'' the effect of local externalities on prices. The research shows it to be especially true in the case of condominium markets for which it is not uncommon to find strong physical dissimilarities between two neighboring units.

The SARS model improves the dependence relationship definition by generating anisotropies expressed on the basis of the degree of similarity between nearby houses and the house to be estimated. The similarity component breaks the isotropy imposed by the distance component and refines the modeling of the structure of local price dependence.

Endnotes

[1] This standardization of weight allows for more direct interpretation of the W_i function, which then represents a weighted average (due to their respective distances relative to *i*) of values of neighboring houses.

[2] Global scale similarity among residences is controlled by the hedonic variables of the regression.

[3] This conformity reasoning can be found in the literature in various forms. Davis and Whinston (1961) base their reasoning on ''the prisoner dilemma'' of game theory, which leads to similar conclusions. Waddell, Berry and Hoch (1993) and Wheaton and DiPasquale (1996) also suggest this influence; they label owners who benefit from external factors caused by their neighbors as ''free-riders.''

[4] The choice of a 250-meter distance threshold in the *CONDO* sample would also be a logical choice.

[5] Besner (1999) explored another sample for which better values of the LIK criterion were found when the distance threshold was modified a posteriori. The out-of-sample performance, however, deteriorated. If one considers similarities without a priori assumptions concerning distance threshold and distance parameter ''m,'' the non linearities of space combined with the wide range of possible matrices that can be specified by Equation (8) makes it possible to find another optima and overfit sample data. The theoretical restrictions imposed in the present paper concerning the process of finding the best weighting scheme should assure generalization.

[6] Note the relative inaccuracy of this index.

References

Anas, A. and S. J. Eum, Hedonic Analysis of a Housing Market in Disequilibrium, *Journal of Urban Economics*, 1984, 15, 87–106.

Anselin, L., Specification of the Weight Structure in Spatial Autoregressive Models, *Modeling and Simulation*, 1984, 15, 99–103.

——., *Spatial Econometrics, Methods and Models,* Dordrecht: Kluwer, 1988.

——., *SpaceStat version 1.80: A program for the Statistical Analysis of Spatial Data*, Department of Geography and Department of Economics, University of California, Santa Barbara, CA, 1995.

Anselin, L. and S. Hudak, Spatial Econometrics in Practice: A Review of Software Options, *Regional Science and Urban Economics*, 1992, 22:3, 509–36.

Anselin, L., S. Hudak and R. Dodson, Spatial Data Analysis and GIS, National Center for Geographic Information and Analysis, University of California, Santa-Barbara, CA, 1992.

Appraisal Institute, *The Appraisal of Real Estate*, Illinois, 1992.

Ball, M. J., Recent Empirical Work on the Determinants of Relative House Prices, *Urban Studies*, 1973, 10, 213–33.

Basu, S. and T. Thibodeau, Analysis of Spatial Autocorrelation in House Prices, *Journal of Real Estate Finance and Economics*, 1998, 17:1, 61–85.

Besner, C., *Un Modèle Spatial Autorégressif de prix Immobiliers Résidentiels*, Thèse de Doctorat, Université du Québec à Montréal, 1999.

Boyce, B. N., *Real Estate Appraisal Terminology*, Cambridge, MA: Ballinger, 1975.

Brigham, E. F., The Determinants of Residential Land Values, *Land Economics*, 1965, 41, 325–34.

Can, A., *Modeling Spatial Variation in Housing Price Determinants*, Ph.D. dissertation, Ohio State University. Columbus, OH, 1989.

——., The Measurements of Neighborhood Dynamics in Urban House Prices, *Economic Geography*, 1990, 66, 254–72.

——., Specification and Estimation of Hedonic Housing Price Models, *Regional Science and Urban Economics*, 1992, 22, 453–74.

Can, A. and I. Megbolugbe, Spatial Dependence and House Price Index Construction, *Journal of Real Estate Finance and Economics,* 1997, 14, 203–22.

Davis, O. A. and A. B. Whinston, Economics of Urban Renewal, *Law and Contemporary Problems*, 1961, 26, 105–17.

Des Rosiers, F., *Resivalu II: Modèle Econométrique des Valeurs Résidentielles, Région de Québec 1986*, Québec: Université Laval, Faculté des Sciences de L'administration, 1988.

Des Rosiers, F. and M. Theriault, Integrating Geographic Information Systems to Hedonic Price Modeling: An Application to the Quebec Region, *Property Tax Journal*, 1992, 11:1, 29.

Dubin, R., Estimation of Regression Coefficients in the Presence of Spatially Autocorrelated Error Terms, *Review of Economics and Statistics*, 1988, 70:3, 466–74.

Dubin, R., Spatial Autocorrelation and Neighborhood Quality, *Regional Science and Urban Economics,* 1992, 22, 434–52.

Dubin, R., Predicting House Prices Using Multiple Listings Data, *Journal of Real Estate Finance and Economics*, 1998, 17, 35–59.

Dubin, R., R. K. Pace and T. G. Thibodeau, Spatial Autoregression Techniques for Real Estate Data, *Journal of Real Estate Literature,* 1999, 7:1, 79–95.

Goodman, A. C., Topics in Empirical Urban Housing Research, in *The Economics of Housing Markets*, New York, NY: Harwood Academic Publishers, 1989.

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

Hamilton, S. W. and D. Dale-Johnson, Multiple Listing Service Sales Data as an Indicator of Market Behavior: Prices, Volume, and Submarket Activity, *Property Journal*, 1991, 10: 2, 175–97.

Jackson, J. R., Intra-urban Variation in the Price of Housing, *Journal of Urban Economics*, 1979, 6, 464–79.

Lam, N. S., Spatial Interpolation Methods: A Review, *American Cartographer*, 1983, 10, 129–49.

Maleyre, I., L'analyse de la Demande Pour les Caractéristiques du Logement: Application à Abidjan, *Revue d'Economie Régionale et Urbaine*, 1995, 3, 44–80.

Mathsoft S-Plus, *User's Manual*, Version 2000 Seattle, WA: Mathsoft, 2000.

Odland, J., *Spatial Autocorrelation*, California: Sage, 1988.

Olmo, J. C., Spatial Estimation of Housing Prices and Locational Rents, *Urban Studies*, 1995, 32:8, 1331–44.

Pace, R. K., C. F. Sirmans and V. C. Slawson, Jr., Are Appraisers Statisticians?, *Real Estate Valuation Theory*, Research in Real Estate Monograph Series, 2001.

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, 333–40.

Pace, R. K., R. Barry and J.M. Clapp et al., Spatiotemporal Autoregressive Models of Neighborhood Effects, *Journal of Real Estate Finance and Economics*, 1998, 17, 15–33.

Richardson, H. W., J. Vipond and R. A. Furbey, Determinants of Urban House Prices, *Urban Studies*, 1974, 2, 189–99.

Ripley, B., *Spatial Statistics*, New York, NY: Wiley, 1981.

Rodriguez, M., C. F. Sirmans and A. R. Marks, Using Geographic Information Systems to Improve Real Estate Analysis, *Journal of Real Estate Research*, 1995, 163–73.

Sibert, J. L., *Spatial Autocorrelation and the Optimal Prediction of Assessed Values*, Ann Arbor, MI: University of Michigan, 1975.

Waddell, P., B. J. L. Berry and I. Hoch, Residential Property Values in a Multinodal Urban Area: New Evidence on the Implicit Price of Location, *Journal or Real Estate Finance and Economics*, 1993, 7, 117–41.

Wheaton, W. C. and D. DiPasquale, *Urban Economics and Real Estate Markets*, Englewood Cliffs, NJ: Prentice-Hall, 1996.

Whittle, P., On Stationary Processes in the Plane, *Biometrika*, 1954, 41, 434–49.

Wiltshaw, D. G., Spatial Autocorrelation and Valuation Accuracy: A Neglected Methodological Problem, *Journal of Property Research*, 1996, 13, 275–86.

Claude Besner, University of Quebec at Montreal, Montreal, Quebec, Canada, H3C 4R2 or besner.claude@uqam.ca.