Effects of Transportation Infrastructure and Location on Residential Real Estate Values

Application of Spatial Autoregressive Techniques

Murtaza Haider and Eric J. Miller

Proximity to transportation infrastructure (highways and public transit) influences residential real estate values. Housing values also are influenced by propinquity to a shopping facility or a recreational amenity. Spatial autoregressive (SAR) models were used to estimate the impact of locational elements on the price of residential properties sold during 1995 in the Greater Toronto Area. A large data set consisting of 27,400 freehold sales was used in the study. Moran's *I* was estimated to determine the effects of spatial autocorrelation that existed in housing values. SAR models, using a combination of locational influences, neighborhood characteristics, and structural attributes, explained 83 percent variance in housing values. Using the "comparable sales approach," a spatiotemporal lag variable was estimated for every property in the database. This research discovered that SAR models offered a better fit than nonspatial models. This study also discovered that in the presence of other explanatory variables, locational and transportation factors were not strong determinants of housing values. On the other hand, the number of washrooms and the average household income in a neighborhood were found to be significant determinants of housing values. Stepwise regression techniques were used to determine reduced spatial hedonic models.

This paper presents selected results from a study of a large data set of 285,000 freehold properties in the Greater Toronto Area (GTA), which were sold during 1987 and 1995. Findings from the analysis performed on the 27,400 sales recorded in 1995 are presented in this paper. Haider documents the detailed results from the spatiotemporal analyses of the entire sample of 285,000 sales (*1*). The objectives of the study were to ascertain significant determinants of housing values, including the influence of locational elements, such as proximity to subway or a mall, on housing values. The presence of spatial autocorrelation in the property values also was evaluated.

The database includes housing and census data for the GTA for the above-mentioned period. Housing sales data were obtained from the Toronto Real Estate Board (TREB). The TREB database captures almost 80 percent of all residential transactions in the study area. To determine neighborhood influences on housing values, we retrieved information on a multitude of Census Tract-level socioeconomic variables from Statistics Canada's databases. The Census Tract-based data were matched with individual properties using a geographic information system (GIS).

Department of Civil Engineering, University of Toronto, 35 St. George Street, Toronto, Ontario M5S 1A4.

The decision to use census data on the Census Tract (CT) level was not arbitrary. Census data are available for various spatial resolutions, such as Block Face and Enumeration Area (EA). Though EA-level data are more disaggregate, privacy concerns cause Statistics Canada to suppress information on certain essential variables, such as income. The benefits of disaggregation are overshadowed by the loss of intrinsic information resulting from data suppression.

In addition, CT boundaries are not entirely arbitrary, as is the case with EA boundaries. Census Tract boundaries adhere to the underlying geography of the area, and they often conform to the grid structure of major streets in the GTA. CTs are delineated in a manner to create areas, which "should be as homogenous as possible in terms of socio-economic characteristics such as similar economic status and social living conditions" (*2*). The average number of people within a CT varies between 2,500 and 8,000.

Housing data were geocoded using a modified version of geocoding algorithms available in MapInfo. We succeeded in geocoding 88 percent of all properties in our database. Properties that sold for less than $25,000 (all prices in 1995 Canadian dollars) were not used in estimating econometric models. The analysis of spatial dependency in housing values formed the basis of spatial hedonic price models developed in this study. We applied spatial autoregressive techniques after quantifying spatial autocorrelation in data.

BRIEF LITERATURE REVIEW

Locational elements, such as distance from the central business district (CBD), have long been used as an explanatory variable in hedonic models. Similarly, the impact of light-rail transit (LRT), subways, and expressways on property values also has been quantified in hedonic models. Using a sample of 235 properties, Al-Mosaind, Dueker, and Strathman observed "a positive capitalization of proximity to LRT stations for houses within 500 meters of actual walking distance" (*3*). They argued that proximity to LRT improves residents' accessibility to the CBD and other urban areas with employment opportunities. The price of properties within 500-m walking distance of LRT stations was $4,300 higher than the rest of the sample. In addition, a statistically weak negative price gradient (D_CBD) within the 500-m zone also was observed. The authors also referenced numerous studies in which proximity to an LRT line capitalized into higher property values. In addition, an earlier Toronto-based study found the price of properties located close to the Spadina LRT line to be higher than the rest of the sample (*4*).

Despite the positive correlation between higher property values and proximity to transit, structural attributes of housing units and neighborhood characteristics often turn out to be better determinants of housing values than accessibility variables. In a study of the impact of Miami metro rail on property values, the authors concluded, "Residential price effects due to rail station development are a function of neighborhood characteristics" (*5*). In a similar study, Kockelman concluded that despite the inclusion of sophisticated measures of accessibility and travel times, "the rather simplistic Distance-to-CBD measures are very strong predictors of housing price and rental rates" (*6*). Even Al-Mosaind, Dueker, and Strathman concluded, "The benefits of accessibility to a transit station may not be as great as some expect" (*3*). Our research also observed that measures of accessibility were not strong predictors of housing values. Similar to the research cited here, we also found distance from the CBD to be a strong determinant of housing values.

A similar study of 12,000 residential sales (single-family dwellings and duplexes) in Cleveland, Ohio, between 1987 and 1992 revealed that housing values increased with distance from the CBD (*7*). More often than not, housing values decline with distance from the CBD, owing to the monocentric nature of cities. The authors argued that the positive coefficient for the distance variable perhaps was due to multiple employment nodes in Cleveland.

The impact of the distance from the CBD depends upon the geography and economy of a city. In a study of house and land prices in Sydney, Australia, it was found that house and land prices fell dramatically with distance from the CBD (*8*). A negative exponential relationship between property values and distance to the CBD was discovered. The log of distance from the CBD was the most statistically significant variable in the model. Locational variables, such as accessibility to rail or to the regional center, were not significant determinants of housing values. These results are consistent with our findings.

The assumption that cities are monocentric may not hold for numerous large cities. Modern cities are fast becoming polycentric as the suburban office and retail centers experience sustained growth. In a study of travel behavior, it was discovered that suburb-to-suburb trips have increased in number, relative to suburb-to-CBD trips, due to the decentralization of employment (*9*).

In another study, the variation in housing values was explained using a distance decay function, change in population and housing stock, and change in ethnic mix (*10*). A generalized version of the repeat-sales index was used to estimate housing price appreciation. The data set consisted of 42,890 repeat sales in 79 CT groups in metropolitan Miami. The properties were geocoded to the respective CT. When the CT group identification was excluded from the model specification, the model explained 76 percent variance in house price appreciation. The addition of CT group identification explained an additional 3 percent of house price appreciation. The log-of-distance variable returned a negative coefficient, indicating that the house price appreciation rate declined with an increase in distance from the CBD.

One further study assessed the changes in housing prices in Boston, Massachusetts, against the changes in demographics, supply of new housing, distance from the CBD, employment in the manufacturing sector, and aggregate school enrollments (*11*). A weighted repeat-sales method was used to estimate models. The data set consisted of 135,000 pairs of sales between 1981 and 1994. The distance from the CBD turned out to be a significant determinant of housing values. In addition, towns located close to Boston experienced a more rapid increase in property values than the towns located at greater distances.

DERIVED LOCATION VARIABLES

A comprehensive spatial analysis is facilitated by the inclusion of implicit locational influences on housing values in econometric models. For example, it is hypothesized that housing values vary with distance from the CBD or that proximity to a large shopping facility capitalizes into higher or lower property values. To test some hypotheses, we used GIS to estimate straight-line distances between each geocoded property and downtown Toronto (the King and Bay intersection). Similarly, distances were estimated between individual properties and the 10 largest shopping centers in the GTA. In addition, binary variables, measuring a property's proximity to the transportation network, were developed. We discuss the development of other locational variables in the following paragraphs.

The proximity to Lake Ontario might influence the price of housing units. The BEACH variable assigns the value 1 for all properties located within a 2-km straight-line distance from the lakeshore. The BEACH_1 variable captured properties lying within 1 km of the lakeshore, whereas the BEACH_DO variable captured those properties located in the area between the 1-km and 2-km buffers.

To test the impact of proximity to a highway, the HWAY variable assigned a value of 1 to properties within a 2-km straight-line distance of the major expressways in the GTA. The rest of the housing stock was coded as 0. The HWAY_1 variable captured properties situated within a 1-km straight-line distance of the highway, whereas the HWAY_DO variable captured the properties situated between the 1-km and 2-km buffers.

To gauge the impact of propinquity to the transit system on housing values, three binary variables were created. The SUBWAY variable assumed the value 1 if the property is located within 1.5 km of the subway line (including the Yonge-University line, Bloor-Danforth line, and Scarborough Rapid Transit) and 0 for the rest of the stock. Two other variables, one capturing properties within a 1-km distance of the subway line (SWAY_1) and the other capturing properties located between the 1-km and 1.5-km buffers (SWAY_DO), also were created.

Three binary variables were created to gauge the impact of proximity to the shopping centers. The MALL variable captured properties within a 5-km straight-line distance from the shopping centers and was assigned a value of 1 if the property was located within the buffer and 0 otherwise. The MALL_25 variable was created for a smaller buffer of 2.5-km straight-line distance and the MALL_DO variable was created for the properties situated in the area between the two buffers.

DESCRIPTIVE ANALYSES

Figure 1 presents the spatial distribution of property values sold during 1995. The most active segment of the market could be observed from the legend, in which 9,463 properties were sold for \$170,00 to \$230,000 (all prices in 1995 Canadian dollars). More than 1,800 properties worth over \$0.4 million were sold during 1995. Each + sign presents the geocoded location of the property and its shade indicates the price range it falls under. High-value properties can be recognized from the darker shades. The spatial correlation in housing values is obvious from Figure 1 and is suggested by the fact that properties with similar prices are clustered together, instead of being randomly distributed in space.

Table 1 presents the effects of various locational variables—highway, subway, etc.—on housing values of properties sold dur-

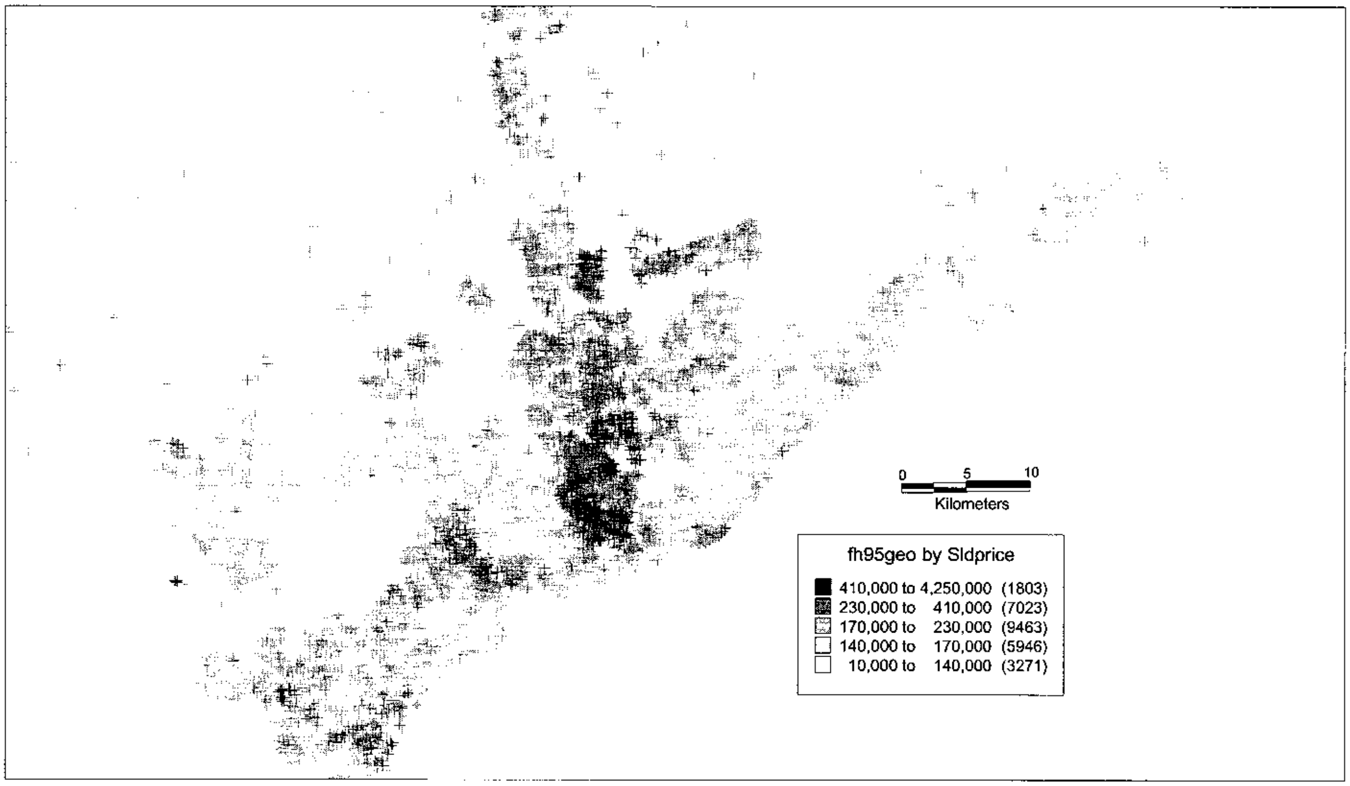

FIGURE 1 Spatial distribution of freehold property values in the GTA, 1995.

ing 1995. The average sale price of properties sold within 1.5 km of a subway line was higher than the rest of the stock. For example, properties within 1 km of a subway line averaged $267,000, generating approximately $47,000 more than the remaining sample with an average sale price of $220,000. Properties within the intersection of two buffers (SWAY_DO) averaged around $257,000. This suggests that property values decline with the distance from the subway line. These relationships were later explored in hedonic models.

We compared the mean property values for proximity variables, using the two-sample mean test (*T*-test). From the comparison in Table 2 we conclude that with the exception of HWAY_DO, we reject the null hypothesis that $\mu_1 = \mu_2$ for all variables at $\alpha = 0.05$. Based on Levene's Test for Equality of Variances (not reported), we also can assume equal variances for all variables except for HWAY_DO. However, the *T*-test is reported for assuming both equal and unequal variances.

The mean price of properties within 2 km of a highway ($223,400) was lower than the mean price of the remaining sample ($230,000). Properties that are not adjacent to the highway yet are close enough to capitalize on propinquity to a major highway reflected different trends in property values. The role of highways in influencing land use is explicitly evident from the fact that 12,154 properties out of 27,412 properties in the sample were located within 2 km of the major expressways. However, the true impact of proximity to a highway could only be determined in a multivariate analysis.

The proximity to regional shopping centers influences residential property values. There is some truth to this hypothesis. Properties within a 5-km radius of the 10 regional shopping centers were on average $4,000 more expensive than the rest of the sample (Table 1). However, the average price of properties within a 2.5-km radius of shopping centers (MALL_25) was $25,000 lower than the average price of the remaining sample at $230,000. Properties close to the shopping centers experience higher levels of noise and air pollution, more than usual traffic volume on small streets, and other discomforts that counteract the locational advantage. Properties located within the intersection of two buffers—in other words, within the doughnut (MALL_DO)—attracted $15,000 more than the rest of the sample. The impact of large retail outlets on land use can be judged from the fact that 46 percent of properties in our sample were located within a 5-km radius of the 10 regional shopping centers in the GTA.

Toronto is unique for its lakeshore real estate. Unlike that in other major metropolitan areas, Toronto's lakeshore is punctuated with industrial properties or old, small residential units. With the exception of a few neighborhoods along the lakeshore, most residential real estate near Lake Ontario is inferior in quality when compared with the rest of the stock in the GTA. This fact might explain why properties located within 2 km of the lakeshore (BEACH) were sold for a lower price ($194,000) than the rest of the sample.

SPATIAL AUTOCORRELATION

The impetus for advocating spatial autoregressive (SAR) techniques is based on the premise that spatial autocorrelation exists in housing values. Moran's *I* was calculated for housing values to quantify spatial autocorrelation. We specified a weight matrix, w_{ij}, by relying on the level of adjacency among CTs. We preferred Moran's *I* to Geary's *C*, since Moran's coefficient, in the case of a misspecified geometric weight matrix, seems to retain power better than other spatial autocorrelation test statistics (*12*).

TABLE 1 Effect of Locational Variables on Freehold Properties Sold in 1995

Variables	Description	Sales	% of Total	Mean Price
SUBWAY	Within 1.5 km of subway line	6117	22.30	263900
	Rest of the Sample	21295	77.70	216400
	Total	27412	100.00	227000
SWAY_1	Within 1 km of subway line	4204	15.30	267200
	Rest of the Sample	23208	84.70	219700
	Total	27412	100.00	227000
SWAY_DO	Within the two buffers	1913	7.00	256700
	Rest of the Sample	25499	93.00	224700
	Total	27412	100.00	227000
HWAY	Within 2 km of a Highway	12154	44.30	223400
	Rest of the Sample	15258	55.70	229800
	Total	27412	100.00	227000
HWAY_1	Within 1 km of a Highway	5241	19.10	220300
	Rest of the Sample	22171	80.90	228500
	Total	27412	100.00	227000
HWAY_DO	Within the two buffers	6913	25.20	225700
	Rest of the Sample	20499	74.80	227400
	Total	27412	100.00	227000
MALL	Within 5 km of a Mall	12574	45.90	229100
	Rest of the Sample	14838	54.10	225100
	Total	27412	100.00	227000
MALL_25	Within 2.5 km of a Mall	3086	11.30	205300
	Rest of the Sample	24326	88.70	229700
	Total	27412	100.00	227000
MALL_DO	Within the two buffers	9488	34.60	236900
	Rest of the Sample	17924	65.40	221700
	Total	27412	100.00	227000
BEACH	Within 2 km of Lake Shore	4089	14.90	194200
	Rest of the Sample	23323	85.10	232700
	Total	27412	100.00	227000
BEACH_1	Within 1 km of Lake Shore	1730	6.30	203400
	Rest of the Sample	25682	93.70	228600
	Total	27412	100.00	227000
BEACH_DO	Within the two buffers	2359	8.60	187500
	Rest of the Sample	25053	91.40	230700
	Total	27412	100.00	227000

Moran's I is defined as the following:

$$I = \frac{n\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}(y_i - \bar{y})(y_j - \bar{y})}{\left(\sum_{i=1}^{n}(y_i - \bar{y})^2\right)\left(\sum\sum_{i \neq j} w_{ij}\right)}$$

where

y_i = the variable of interest,
w_{ij} = the spatial weight matrix, and
y-bar = the mean value of y.

The calculation of Moran's I is computationally very intensive. We aggregated the property values to the CT level to reduce the sample size. This aggregation might result in a higher value for Moran's I due to the aggregation bias. However, estimating Moran's I for 27,400 observations was not possible with the available computing power. In addition, Figure 1 is a good indication of the presence of spatial autocorrelation in the disaggregated data set.

The weight matrix was based on the level of adjacency between two CTs. For example, the level of adjacency for two contiguous CTs is expressed as a function of the length of common border. Therefore, the greater the length of common border between the two CTs, the more contiguous they are. Table 3 presents results for Moran's I calculations, computed in TransCad. The Moran's I value of 0.65 suggests the presence of strong spatial autocorrelation in mean housing values.

TABLE 2 Hypothesis Test for Difference of Means

Variables	T-Statistics	
	A*	B**
HWAY	4.23	4.26
HWAY_1	4.25	4.63
HWAY_DO	1	0.98
SUBWAY	26.29	19.913
SWAY_1	22.63	16.63
SWAY_DO	10.71	7.78
MALL	2.62	2.58
MALL_2.5	10.15	12.82
MALL_DO	9.49	8.67
BEACH	18.12	22.37
BEACH_1	8.05	9.13
BEACH_DO	15.98	22.77

* Equal Variances Assumed
** Unequal Variances Assumed

METHODOLOGY

There is a consensus in the housing literature that the hedonic price method offers the best econometrics environment to estimate housing prices; the development of a hedonic model in this paper relies heavily on the model developed by Can and Megbolugbe (*13*). They have argued in the past that the existing models were insensitive to the geographic location of dwellings within the metropolitan area, thus overlooking the intermetropolitan variation in housing values. Spatial dependence varies with metropolitan area and over time. They recommend the use of spatial spillover effects, inherent in metropolitan housing markets, in the specification of housing price function.

Can and Megbolugbe adopted the "Comparable Sales Approach" in specifying the spatial lag variable (*13*). At the heart of this approach is the assumption that the price history in the immediate neighborhood of a given property will have spillover effects on its market value. Prices of the most recent sales of similar properties are considered in estimating the market value of a property, controlling for differences in their structural attributes and neighborhood characteristics. The spatial hedonic model specification is presented in the following equation:

$$P_{it} = \alpha + \rho\sum_j w_{ij}P_{j,t-m} + \sum_k \beta_k S_{ik} + \sum_l \gamma_l N_{ilt} + \xi_{it}$$

$$m = 1, \ldots, 6; j \neq i; w_{ij} = \sum_j\left[(1/d_{ij})\Big/\sum_j 1/d_{ij}\right]$$

$$j = 1, 2, \ldots, N; d_{ij} \leq 2 \text{ km}$$

The variation in house prices is explained as a function of structural characteristics (S) of housing units for $k = 1, \ldots, K$ and neighborhood characteristics (N) for $1, \ldots, L$. β and γ are the parameter vectors corresponding to S and N, while α is a constant. W_{ij} is the weight that specifies the extent of influence of price of prior sales P_j (which occurred between time $t - m$ and t) on the transaction price of the concerned property, which we would refer to as the "anchor property." Meanwhile, ρ is a measure of overall level of spatial dependence between $\{P_i, P_{j,t-m}\}$ pairs.

TABLE 3 Moran's I Calculations for Mean Property Values for CT

	Common Border Length
Observations	908
S1	22074.55
S2	213400.83
Sum of Weights	4524.29
Moran's I	**0.650**
Expected Value	-0.001
Std Error	0.033
T Statistic	19.91
95% C.I. Upper	0.714
Lower	0.586

This model incorporates both spatial and temporal functional interdependencies. The influence of prior sales is hypothesized as an inverse function of distance, d_{ij}. The lesser the distance between a prior sale and the anchor property, the more influence that prior sale will have over the transaction price of the anchor property. By introducing a spatially autoregressive term ($w_{ij} * P_{j,t-m}$) as an explanatory variable, we have explicitly controlled for the functional interdependence.

We hypothesized that the value of a property at time t is influenced by recent sales of comparable properties near the anchor property in the past 6 months. We also hypothesized that the spatial spillover effects do not extend beyond a 2-km radius of the anchor property. In other words, we assumed that housing values are not correlated if separated by a distance greater than 2 km. We also hypothesized that property values are not correlated if the sales are more than 6 months apart. The temporal cutoff point (assuming a 6-month lag) is arbitrary. However, we did employ semi-variograms to determine the extent of spatial autocorrelation. The decision to use a 2-km cutoff was based on the spatial statistical analyses documented in Haider (*1*). The large sample size affords us the opportunity to apply ordinary least squares (OLS) or weighted least squares techniques instead of maximum likelihood estimators, since OLS estimates are unbiased for large sample sizes (*14*).

INITIAL MODELING TO OBTAIN SIGNIFICANT DETERMINANTS OF HOUSING VALUES

The initial concern in model estimation was to identify those variables that explain the maximum variance in housing prices. Structural attributes of housing units, neighborhood characteristics, and derived locational variables thus were divided into numerous groups. These groups were individually regressed on the dependent variable, the natural log of housing price.

The spatial lag variable was the most significant one in explaining variance in housing values. Structural attributes of housing units, such as the number of washrooms and the parking capacity, also were significant determinants of housing prices. Variables controlling for the absence of fireplaces or the presence of multiple fireplaces also were strong predictors of housing values. Comfort variables such as centralized air-conditioning (AIR_CON) and heating arrangements were significant in explaining housing values, suggesting that air-conditioned housing units were more expensive than the ones without air-conditioning.

When housing values were regressed on trend variables, actual longitude and latitude of housing units, they explained little or no variance. Similarly, locational binary variables, controlling for accessibility premiums for individual housing, were considered weak determinants of housing values. The coefficients generally returned expected signs, a negative coefficient for the BEACH variable and a positive coefficient for the SWAY variable, yet all such variables remained weak predictors of housing values.

The distance from shopping centers explained significant variance in housing values. D_YDALE (distance from Yorkdale Mall) turned out to be a significant variable along with D_CBD (distance from the CBD). Later, during the development of spatial hedonic models, D_CBD performed better than D_YDALE and hence D_CBD was retained in the best-reduced model. The average census family income and the level of education attainment in a CT also were positively correlated with property values. Stepwise regression techniques were applied to short-list hundreds of explanatory variables to fewer than 20 variables.

DEVELOPMENT OF SPATIAL AUTOREGRESSIVE MODELS

Table 4 presents the summary statistics of explanatory variables used in reduced models. The average sales price for the sample was \$227,600. The average number of rooms in a house was 7; the average number of bedrooms was 3.3. The average number of washrooms was 2.5, with the average parking capacity at 1.2. Seventy percent of the properties in our sample were detached housing. Thirty percent of housing units reported at least one fireplace, while 10 percent of the housing units reported more than one fireplace. Almost 50 percent of the houses in the sample were centrally air-conditioned.

We used the semi-log specification to control for nonlinearity in the data set. The following equation gives the models described in this section:

$$\text{Housing values} = lag^{\rho} e^{\left[\alpha + \left(\sum_{i=1}^{n} \beta_i S_i\right) + \left(\sum_{j=1}^{n} \gamma_j N_j\right) + \epsilon\right]}$$

where

S = structural attributes (type and size of unit);
N = neighborhood characteristics;
lag = spatial lag variable; and
β, γ, and ρ = parameter vectors corresponding to S, N, and lag, respectively.

Variance inflation factors (VIFs) were estimated (results not shown) to test multicollinearity within explanatory variables. Low values for VIF suggested little or no multicollinearity within the explanatory variables. Models were weighted by the number of rooms to control for an increase in variance of residuals with the increase in the value of the dependent variable.

Table 5 shows detailed results for the best-reduced model. Also presented in the table is the nonspatial version of the same model (the lag variable is omitted from nonspatial model specification). It can be seen from this table that the spatial model explains 83 percent variance in housing values. In addition, all coefficients are significant at 95 percent confidence level and returned expected signs. The spatial lag variable is the most significant variable in the spatial hedonic price model. When the spatial lag variable is excluded from model specification, the explanatory power of the model is compromised and some variables in the nonspatial model return counterintuitive results.

TABLE 4 Summary Statistics of Explanatory Variables Used in Reduced Models

Variables	Description	Mean	Std. Dev.	Minimum	Maximum
DAYSON	No. of Days on MLS	62	57	0	973
SLDPRICE	Actual Sale Price	227600	134100	10000	4250000
ROOMS	No. of Rooms	6.93	1.95	0	90
BEDS	No. of Bedrooms	3.3	0.8	0	9
NO_WASH	No. of Wash-rooms	2.49	1.03	0	9
PARK_CAP	Parking Capacity	1.16	0.82	0	5
D_CBD	Distance from CBD	21.5	13.1	0.2	80.8
CF_AVINC	Avg. Income of Census Family in CT	68000	25000	26900	231700
LOG_PRIC	Ln of Sale Price	12.23	0.41	9.21	15.26
LOG_LAG	Ln (Lag_var)	12.27	0.32	9.8	14.5
DETACH	Binary: 1 if detached 0 otherwise	70%			
THREE_ST	Binary: 1, if three-storey , 0 otherwise	4%			
FIRE_MLT	Binary: 1, if multiple fireplace, 0 otherwise	10%			
FIRE_NO	Binary: 1, if no fireplace , 0 otherwise	30%			
AIR_CON	Bin, 1 if Cent. Air Conditioned, 0 otherwise	50%			
Percentage of Seniors	%age of seniors in the CT	9.1%	5.6%	2.0%	33.0%
Percentage of Kid_LT6	%age of families with small children in the CT	8.4%	2.6%	1.0%	17.0%
Percentage of Immigrants	%age of new immigrants in the CT	34%	12.0%	10.0%	67.0%

TABLE 5 Comparison of Spatial and Nonspatial Models for 1995 Freehold Data

	Spatial Model		Non-Spatial Model		Model with Locational Variables	
No. of Records	26922		26987		26910	
Adjusted R-Square	0.832		0.779		0.832	
Std. Error of the Estimate	0.4640		0.5343		0.4591	
Mean Predicted Value	12.2420		12.2437		12.2416	
Std. Dev. Of Pred. Value	0.3725		0.3572		0.3694	
Mean of Residuals	-4.34E-03		-5.4E-03		-4.295E-03	
Std. Dev. of Residuals	0.1723		0.1975		0.1702	
Variables	**Beta**	**T-stat**	**Beta**	**T-stat**	**Beta**	**T-stat**
(Constant)	5.11E+00	74.639	11.31	804.643	5.1214	73.55
NO_WASH	9.08E-02	67.545	0.104	67.37	0.0895	67.01
FIRE_MLT	1.26E-01	33.71	1.47E-01	34.226	0.1255	33.87
DETACH	8.79E-02	31.769	9.48E-02	29.789	0.0877	31.93
BEDS	5.50E-02	37.04	6.20E-02	36.375	0.0547	37.16
CF_AVINC	2.15E-06	31.377	6.18E-06	101.989	0.0000	30.60
D_CBD	-3.97E-03	-29.012	-7.44E-03	-49.77	-0.0037	-23.88
PARK_CAP	5.11E-02	29.111	7.72E-02	38.802	0.0507	29.12
FIRE_NO	-6.15E-02	-22.398	-9.70E-02	-31.041	-0.0629	-23.12
AIR_CON	5.15E-02	21.723	6.74E-02	24.763	0.0528	22.27
POOL_UG	5.22E-02	12.039	4.61E-02	9.267	0.0517	12.01
THREE_ST	5.16E-02	9.26	2.18E-02	3.401	0.0483	8.65
LOG_LAG	0.533	92.168			0.5326	91.60
Percentage of Seniors	0.422	14.259	0.667	19.674	0.3833	12.98
Percentage of Immigrants	-0.105	-8.383	1.23E-02	0.861	-0.0902	-6.79
Percent of KidsLT6	-0.284	-4.658	-0.452	-6.441	-0.3617	-5.91
HWAY					-0.0066	-3.03
SUBWAY					0.0165	4.77
BEACH					-0.0063	-1.88
MALL					-0.0131	-5.83

Other things being equal, the price of a unit will increase with an increase in the number of washrooms. This is also true for the number of bedrooms. An increase in parking capacity results in an increase in housing values. Binary variables representing centralized air-conditioning, detached housing, a pool, multiple fireplaces, and three-story housing showed positive influences on housing values. The percentage of older citizens in a CT is positively correlated with housing values, while the percentage of new immigrants in a CT is negatively correlated with housing values in the GTA. In addition, Figure 2 shows that the price of housing decreases with the distance from the CBD, other things being equal. [Figure 2 is based on the application of an estimated model to a sample housing unit.

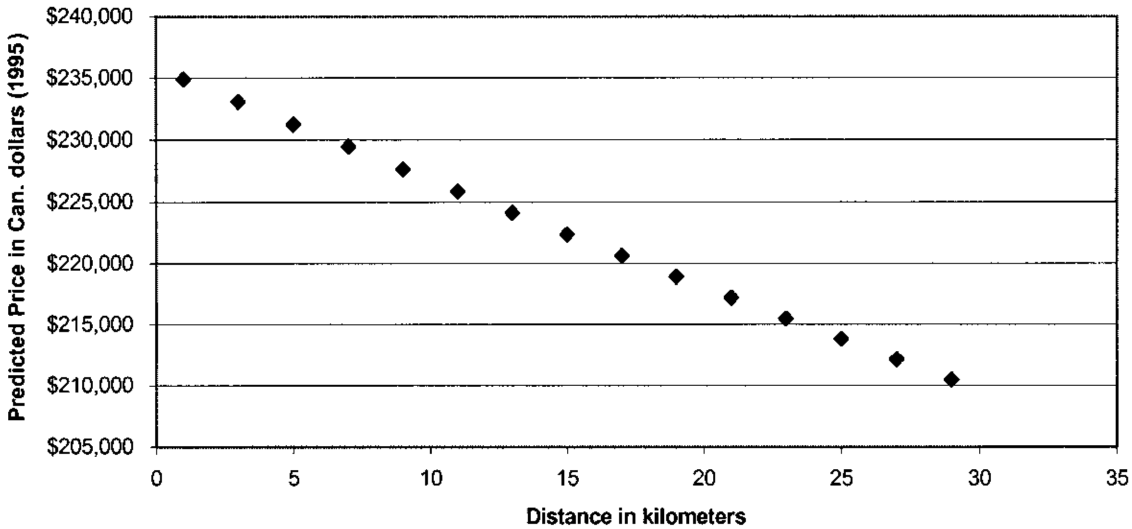

FIGURE 2 Drop in property values with an increase in distance from the CBD.

Our sample dwelling is a detached, air-conditioned, single-family dwelling with four bedrooms, two washrooms, one parking space, and a fireplace. The dwelling unit is located at a distance of 3 km from the CBD in a neighborhood with an average income of $75,000 (in 1995 Canadian dollars). Almost 15 percent of the residents in the CT are seniors, and another 14 percent are new immigrants to Canada. In addition, 35 percent of households reported at least one young child under age 6. The weighted average price of comparable sales—the lag variable—was assumed to be $225,000.]

However, a comparison with the nonspatial model reveals that omission of the lag variable reduces the adjusted R-square to 78 percent. The difference in the explanatory powers of the two models was less than our expectations. Can and Megbolugbe reported OLS estimates for both simple and spatial hedonic models (*13*). The adjusted R-square for the traditional hedonic model was 57 percent, whereas for spatial hedonic models, the adjusted R-square averaged near 75 percent (*13*). They also observed that SAR models offered better fit.

Though the difference in adjusted R-square in the two models reported in this study is not large, telltales associated with autocorrelated data are visible in the simple model. As expected, the *t*-statistics for the nonspatial model were higher than the spatial model. Spatially correlated data sets return inflated *t*-statistics if the model specification does not control for spatial autocorrelation. In addition, with the omission of the spatial lag variable, which controlled the spatial dependency in the dependent variable, D_CBD assumes greater significance in explaining the variance in housing values. Variable D_CBD almost doubles in power to -7.44×10^{-3} from -3.96×10^{-3}, thus adding more weight to the variable D_CBD. The *t*-statistic for D_CBD in the nonspatial model almost doubles in value.

The nonspatial model indicates that the percentage of immigrants in a CT is positively correlated with the housing values. This is counterintuitive since our data suggest otherwise. On the other hand, the spatial model establishes a negative correlation between the percentage of immigrants in the CT and housing values. In addition, the standard error of the estimate is smaller for the spatial model. Similarly, the standard deviation of residuals also is smaller for the spatial model.

When we brought locational variables (HWAY, SUBWAY, MALL, and BEACH) into the model, the spatial lag variable retained its explanatory power in the model. The adjusted R-square value remained unchanged at 83 percent. The positive sign for SUBWAY suggests that proximity to a subway line capitalizes into higher property values, other things being equal. Substituting the sample dwelling characteristics, mentioned earlier for Figure 2, we found that propinquity to a subway line adds approximately $4,000 to property value, other things being equal.

The negative sign for coefficients of variables HWAY and BEACH suggests that proximity to expressways and lakeshore capitalizes into lower property values, other things being equal. The negative value for the coefficient of the MALL variable was counterintuitive. In the descriptive analysis, we observed that propinquity to regional shopping centers capitalized into slightly higher property values. We calculated VIF to test for multicollinearity in the model. VIF values for the variable MALL did not suggest the presence of multicollinearity. Earlier, we observed in the explanatory analysis that an increase in property values due to proximity to regional shopping centers did not return a linear relationship with distance. Properties located very close to the shopping centers were significantly lower in value than the rest of the sample. We assume that this could be the reason for the negative sign for the variable MALL.

CONCLUSIONS

In a multivariate analysis, locational variables, with the exception of D_CBD and distance from shopping centers, explained less variance in housing values. In a properly specified model, other explanatory variables explain variance in housing values better than variables controlling for locational premiums or penalties. However, in the descriptive analysis, improved accessibility to certain desired features, such as the subway system, or proximity to a desired or a despised characteristic, such as a highway, showed an influence on property values. Though locational variables, in the presence of other explanatory variables, did not improve the adjusted R-square of the model, these variables still returned significant coefficients.

Locational variables, however, identified certain essential trends in land use. For example, 45 percent of properties in our sample were located within 2 km of major expressways in the GTA. Similarly, a significant percentage of houses were situated within a 5-km radius of the 10 largest shopping centers in the GTA.

We considered as comparable sales only those properties that were sold within a 2-km radius and within the past 6 months of the sale of the anchor property. These cutoff points are rather arbitrary. A better understanding of spatial dependency could be achieved by calculating spatial lag variables at different spatial and temporal lag intervals. For example, instead of selecting comparable properties using a 2-km buffer, one could select properties within a ½-km, 1-km, or 3-km buffer. Similarly, the temporal lag could either be expanded or reduced. Models estimated with different spatial lag variables will offer more insight into the spatial and temporal dependency in housing values.

SAR models have obvious limitations if used as a forecasting tool, given their dependence on past sales. In addition, the impact of the spatial lag variable on other spatial variables could be problematic.

ACKNOWLEDGMENTS

This study was funded by the National Sciences and Engineering Research Council Collaborative Project and Individual Operating Grants; Mr. Haider also was supported by an Ontario Graduate Scholarship. The authors would like to thank the Toronto Real Estate Board for access to its Multiple Listing Service database. Asmus Georgi, a research associate at the Department of Civil Engineering, University of Toronto, is recognized for writing the code to estimate the lag variable.

REFERENCES

1. Haider, M. *Development of Hedonic Price Indices for Freehold Properties in the Greater Toronto Area: Application of Spatial Autoregressive Techniques.* Master's thesis. Department of Civil Engineering, University of Toronto, 1999.
2. *1996 Census Dictionary*, 1st ed. Statistics Canada, Ottawa, 1997.
3. Al-Mosaind, M. A., K. J. Dueker, and J. G. Strathman. Light-Rail Transit Stations and Property Values: A Hedonic Price Approach. In *Transportation Research Record 1400*, TRB, National Research Council, Washington, D.C., 1995, pp. 90–94.
4. Bajic, V. The Effects of a New Subway Line on Housing Prices in Metropolitan Toronto. *Urban Studies*, Vol. 20, No. 2, 1983, pp. 147–158.
5. Gatzlaff, D. H., and M. T. Smith. The Impact of Metrorail on the Value of Residences Near Station Locations. *Land Economics*, Vol. 69, 1993, pp. 54–66.
6. Kockelman, K. M. Effects of Location Elements on Home Purchase Prices and Rents in San Francisco Bay Area. In *Transportation Research Record 1606,* TRB, National Research Council, Washington, D.C., 1997, pp. 40–50.
7. Simons, R. A., R. G. Quercia, and I. Maric. The Value Impact of New Residential Construction and Neighbourhood Disinvestment on Residential Sale Price. *Journal of Real Estate Research*, Vol. 15, 1998, pp. 147–161.
8. Abelson, P. House and Land Prices in Sydney from 1931 to 1989. *Urban Studies*, Vol. 34, 1997, pp. 1381–1400.
9. Levine, J. C. Decentralization of Jobs and Emerging Suburban Commute. In *Transportation Research Record 1364*, TRB, National Research Council, Washington, D.C., 1995, pp. 71–80.
10. Archer, W. R., D. H. Gatzlaff, and D. C. Ling. Measuring the Importance of Location in House Price Appreciation. *Journal of Urban Economics*, Vol. 40, 1996, pp. 334–353.
11. Case, K. E., and C. J. Mayer. Housing Price Dynamics Within a Metropolitan Area. *Regional Science and Urban Economics*, Vol. 26, 1996, pp. 387–407.
12. Florax, R., and S. Rey. The Impacts of Misspecified Spatial Interaction in Linear Regression Models. In *New Directions in Spatial Econometrics* (L. Anselin and R. Florax, eds.), 1st ed., Springer-Verlag, Berlin, 1995, pp. 111–135.
13. Can, A., and I. F. Megbolugbe. Spatial Dependence and House Price Index Construction. *Journal of Real Estate Finance and Economics*, Vol. 14, 1997, pp. 203–222.
14. Cliff, A., and J. Ord. *Spatial Processes: Models and Applications*, 1st ed. Pion Limited, London, 1981.

Publication of this paper sponsored by Committee on Transportation and Land Development.