AN ANALYSIS OF SPATIAL AUTOCORRELATION IN HONG KONG'S HOUSING MARKET

LI CHUN WAH, WONG SIU KEI and CHAU KWONG WING
The University of Hong Kong

ABSTRACT

Spatial autocorrelation is commonly found in the Hedonic Pricing model for real estate prices, but little is known about the time window within which nearby sales are used to form current prices. The primary objective of this study is to examine how spatial autocorrelation varies with the length of this time window. We hypothesize that more recent sales have a stronger influence than earlier sales on current prices, so a shorter time window should induce stronger spatial autocorrelation. We propose a Spatial Hedonic Pricing (SHP) model to test our hypothesis. Based on 15,500 transactions of residential units in Taikooshing, Hong Kong from 1992 to 2006, we conclude that while positive spatial autocorrelation is present in housing prices, its magnitude increases when the reference period of past sales becomes shorter. The latter is a new finding in the spatial hedonic literature that not only confirms the importance of timeliness in weighting nearby housing price information, but also calls for further research on how fast such information expires in different markets.

Keywords: Time window, hedonic pricing model, spatial autocorrelation, spatial hedonic pricing model.

INTRODUCTION

Spatial autocorrelation can be considered the co-variation of variables within a geo-space. In other words, the value of a variable observed in one location depends on the value of neighbouring variables. LeSage and Pace (2009) proposed at least three possible explanations for this. First, there is an external force causing a spatial relationship: whatever has an effect in one location also has a similar effect in nearby locations. Another one is spatial externality: something in a given location directly influences the characteristics of nearby locations. The third is spatial interaction: the movement of people, goods or information creates apparent relationships between locations. Thus, observations in close spatial proximity tend to share more similarities than those that are more spatially separated. Understanding the nature of spatial autocorrelation has important implications for how spatial information should be used to analyze location-specific variables, in particular real estate decisions that always emphasize "location, location, and location".

From the point of view of statistical modelling, spatial autocorrelation signifies a technical issue, rendering traditional econometric models like regression analysis and time series analysis insufficient. In fact, traditional econometric models cannot be applied in a straightforward way to spatially correlated data because spatial dependence violates the assumption of independence among observations. For instance, regression analysis, which usually assumes that the error term is identically and independently distributed, does not compensate for spatial dependence, and can yield unstable estimated parameters and unreliable significance tests. For time series analysis, although well-established techniques for modelling serial correlation exist, they are unidirectional and cannot be extended directly to spatial dependence, which is multidirectional in nature. Therefore, spatial dependence necessitates the use of a different methodological framework.

The study of spatial autocorrelation has gained momentum since the 1970s. Back then, geographers became concerned with the systematic spatial patterns of variables and tried to improve their geographical estimations by including a spatial process in their models. Notable discussions can be found in Silk (1979), Cliff and Ord (1981), Miron (1984), Upton and Fingleton (1985), Odland (1988), etc. A growing number of real estate researchers also noted the problem of spatial autocorrelation in property valuation and devised different spatial econometric methods to estimate property prices (e.g. Can 1990; Dubin 1992; Can and Megbolugbe 1997; Basu and Thibodeau 1998; Brasington 1999; Clapp et al. 2002; Wilhelmsson 2002, etc). Many property studies, however, assumed that spatial autocorrelation does not matter or is time-invariant. For instance, Hedonic Pricing models typically utilize location dummies to control for variations in property prices across locations without any regard for spatial autocorrelation in their regression residuals. Other pricing models took into account spatial dependence among nearby sales (e.g. Pfeifer and Deutsch 1980; Stoffer 1986; Can and Megbolugbe 1997; Pace et al. 1998; Case et al. 2004; Sun et al. 2005), but did not consider the length of the time window within which nearby sales are spatially correlated. Is it correct to assume that the current price of a house be influenced equally by nearby houses sold during a certain month and by those sold six months earlier?

The primary objective of this study is to examine the time-varying effect of spatial autocorrelation on housing prices in Hong Kong. We believe that the magnitude of spatial autocorrelation is not fixed, but varies with the time window within which buyers and sellers draw past comparables to form current prices. While more recent comparables should weigh more than older ones, the relevant length of the time window is an empirical question that cannot be pre-determined. By exploring different time windows in the spatial modelling process, we can reveal how buyers and sellers weigh old and new neighbourhood sales when they search for price information. In this paper, we constructed a Spatial Hedonic Pricing model with a spatially lagged price variable based on different time windows of previous transactions. To the best of our knowledge, no previous study has ever analysed the effect of time windows on spatial autocorrelation.

This article is divided into six sections. Section 2 is a literature review of the spatial autocorrelation and spatio-temporal models. Section 3 describes the data from Taikooshing's transactions that will be utilised for the study and provides summary statistics for housing prices and quality variables. Section 4 introduces the Spatial Hedonic Pricing model. Section 5 provides the research methodology and the empirical results of the SHP models with different time windows. Section 6 is the conclusion.

LITERATURE REVIEW

Hedonic Pricing models have been widely applied to value heterogeneous goods such as automobiles, consumer products, and real estate (e.g. Court 1939; Lancaster 1966; Griliches 1971; Rosen 1974). For real estate, the models generally assume that property prices can be decomposed into structural, neighbourhood, and locational attributes, and regression is used to estimate the implicit price of each attribute (Linneman 1980; Grether and Mieszkowski 1980; Colwell et al. 1985; Colwell 1990; Des Rosiers et al. 1996, 2001; Malpezzi 2003). However, if regression residuals are spatially correlated, Ordinary Least Squares (OLS) estimates of the implicit prices would be inefficient. Recognizing the correlation between the physical proximity of observations and omitted-variable bias, Dubin (1988) showed that a more precise estimation and prediction of housing prices can be made by allowing for spatial autocorrelation and other locational effects. Anselin (1988) provided a comprehensive review on the econometric modelling of spatial processes and illustrated how spatial effects can be viewed as special cases of a more general modelling framework.

In the spatial econometric literature, there are two main models to correct for spatial autocorrelation, namely the spatial-lag and spatial error models (Anselin 1988). Their specifications are shown below:

Spatial-lag model: $Y=\rho_1 W_1Y + X\beta+ \varepsilon$
Spatial error model: $Y=X\beta++ v$, with $v = \rho_2W_2v + \varepsilon$

In the above specifications, the dependent variable is Y (e.g. property prices). β is the vector of the parameters (e.g. implicit prices) associated with exogenous variables X (e.g. property attributes). ρ_1 is the coefficient of a spatially lagged dependent variable (i.e., W_1Y), and ρ_2 is the coefficient of the spatial autoregressive structure for the disturbance v (i.e., W_2v). W_1 and W_2 are respectively, the spatial weight matrices of the spatial autoregressive process in the dependent variable Y and in the disturbance v. The error ε is normally distributed with constant variance.

Since OLS may not be an efficient estimator for both models, Dubin (1988, 1998, 2003), Anselin (1988), Kelejian and Prucha (1998, 1999), Kelejian and Robinson (1993), and Dubin et al. (1999) suggested a maximum likelihood or generalized moments approach to estimating the parameters. Pace et al. (1998) proposed that the Generalized Least Squares (GLS) method is an alternative estimation approach for improving the loss of efficiency, predictive accuracy, and biased inference arising from an ignorance of spatial autocorrelation in OLS estimates. Further development of spatial econometric tests and estimation methods can be found in Anselin (1998, 2001a, 2001b) and Anselin and Bera (1998).

Most real estate research follows the spatial econometric approach to spatial autocorrelation. For example, Can (1990) found that the assumptions of "fixed" structural parameters in hedonic regression did not reflect the spatial dynamics of urban housing markets, and therefore proposed a spatial approach to capturing house price variabilities arising from spatial spillovers and spatial parametric drifts. Dubin (1992) suggested modelling the spatial autocorrelation of housing prices by replacing neighbourhood and accessibility variables with a spatial autoregressive error term. In analysing the effects of public school quality on housing prices, Brasington (1999) showed that the use of a spatial autoregressive model improves the overall goodness of fit and the estimated spatial parameter is highly significant. Clapp et al. (2002) proposed the Bayesian approach to modeling residuals from a Local Regression Model, which provides better predictive power than a Hedonic Pricing model. By comparing the performance of the OLS model, the spatial lag model, and the spatial error model, Wilhelmsson (2002) found that the models with a spatial structure explain real estate data better.

More sophisticated spatial econometric models, such as the Space-Time Autoregressive (STAR) model, take both space and time dimensions into account (Pfeifer and Deutsch 1980; Stoffer 1986). The STAR model was first applied to areas such as geostatistics (Kyriakidis and Journel 1999), hydrology (Deutsch and Ramos 1986), and business forecasting (Pfeifer and Bodily 1990). For real estate applications, Can and Megbolugbe (1997) incorporated both temporal and spatial dependence into their Hedonic Pricing model for housing transactions in Miami, Florida. They suggested that the transaction price of a house at any time, t, would be determined not only by its structural attributes and the desirability of the neighbourhood, but also by the price effects from prior sales within its vicinity. They incorporated these temporal and spatial dependencies into a spatial weight matrix. The extent of spatial influence was expressed as an inverse of distance between transacted properties. Can and Megbolugbe showed that their Spatial Hedonic Pricing model is useful for constructing housing price indices, especially when not all locational and neighbourhood attributes are available.

Similarly, Pace et al. (1998) proposed a spatial model that synthesized models from the time series and spatial econometrics literature and applied it to housing transactions in Fairfax County, Virginia from 1969 to 1991. They employed a filtering process[1] based on the spatial and temporal proximity of data, a method that greatly reduces the number of parameters to be estimated while improving estimation and prediction performance. Gelfand et al. (1998) adopted a similar approach by introducing the spatio-temporal component into their Hierarchical model to improve their predictions of property values. Archer et al. (1996) and Goetzmann and Spiegel (1997) used the repeat sales methodology, initiated by Bailey et al. (1963), as the basis for spatiotemporal housing analysis. They found that while tract location partly explains different housing price paths over time, this effect appears to be "dominated by the idiosyncratic influences of individual home and its immediate environment". Although several STAR models were proposed in the real estate literature, none of them considered spatial autocorrelation in terms of the length of time windows that we mentioned in the Introduction.

HONG KONG DATA

We make use of the transaction data from a private housing estate in Hong Kong for empirical analysis. We choose Hong Kong for our study because it is one of the most actively traded and efficient real estate markets in the world (Chau 1997; Hui and Lui 2002; Chau et al. 2005). Moreover, as Hong Kong is a good example of a high-rise, densely populated spatial structure, we expect that the conclusions drawn from this study should be applicable to other densely populated cities such as Shanghai, Singapore and Tokyo.

The private housing estate chosen is Taikooshing, which consists of over 14,000 apartment units in 61 buildings of similar design and quality. The units were frequently traded, providing us with 15,500 transactions from 1992 to 2006 to test our hypotheses[2]. As the units are relatively homogenous with a standard design (i.e., they were built by a single developer almost at once), it makes sense for buyers and sellers at Taikooshing to refer to the prices of other units in the same estate – a good motivation for spatial autocorrelation. As a result, we could test if the variation in prices could be explained by their spatial lags in addition to the usual hedonic variables like apartment characteristics and market conditions.

---

[1] This is a linear combination of spatial temporal process with weightings of spatial and temporal effects.

[2] While our analysis starts from 1992, transactions in 1991were used as the pre-sample data to generate the spatially lagged price variable for 1992.

The spatial dimensions, in the form of coordinates for each building were obtained from the government's Lands Department. The distance between a pair of transacted units is calculated as the Euclidean distance between their corresponding buildings. The Euclidean distance is chosen because it is the most common metric in spatial applications. After excluding 74 ineligible transactions, such as those with zero prices, missing flat sizes, or missing floor levels, we arrive at 15,426 effective transactions for our study.

The 61 buildings in Taikooshing were built on a rectangular site with the length running from the east to the west. Take the building in the centroid of the site as an example (i.e., Marigold Mansion): its average distance to the other buildings is 185m and it is 37m from the nearest building. This indicates that buildings in Taikooshing are quite close to each other, which is a rather typical arrangement for many residential developments in Hong Kong.

For the 15,426 effective transactions from 1992 to 2006, the average housing price was HK$ 4.1 million. More than 80% of the transactions ranged from HK$2 million to HK$6 million. The housing quality variables included in the study are flat size, floor level, building age, and sea view. The flat size of each transacted unit is measured in sq. ft. The floor level is the storey on which a subject unit is located. The building age represents the number of years between the transaction date of a unit and the completion date of a building. For sea view, we checked if a unit has at least one window in a certain room (i.e., living room, dining room, or bedroom) boasting a view of the sea. If such a window has an unobstructed sea view, a dummy variable known as "full sea view" would be set to 1. If the sea view is partially obstructed by another building, another dummy variable known as "partial sea view" would be set to 1. The summary statistics of the transaction prices and the above quality variables are shown in Table 1.

Table 1: Descriptive statistics of housing price, flat size, flat age, floor level, and sea view for transactions at taikooshing from 1992 to 2006

	Description	Proportion	Mean	Standard deviation
Transaction Price (HK$ Million)	Below 2	4.8%	4.1	1.8
	2 – 4	52.7%		
	4 – 6	29.6%		
	6 or above	13.0%		
Flat size (sq. ft.)	Below 650	10.9%	804.1	173.7
	650 – 800	43.0%		
	800 – 1,000	32.2%		
	1,000 or above	14.0%		
Building age (years)	Less than 10	11.3%	16.5	5.3
	10 – 15	29.4%		
	15 – 25	51.7%		
	25 or above	7.6%		
Floor level	Below 8	9.4%	15.0	7.7
	8 – 15	39.5%		
	15 – 25	36.4%		
	25 or above	14.6%		
Sea view	Full sea view	10.2%	--	--
	Partial sea view	5.0%		
	No sea view	84.9%		

Note: Figures may not add up to 100% due to rounding off.

Before turning to our empirical model, a preliminary assessment of spatial autocorrelation in our housing price data would be illustrative. Such an assessment, which is analogous to the assessment of autocorrelation in times series data, provides a preliminary check on the spatial dependence among observations before estimating our spatial econometric model. We adopt the Moran's I statistic[3], which is one of the most popular spatial autocorrelation measures, to test if spatial autocorrelation exists in our housing transaction data. The Moran's I statistic, like the correlation coefficient, ranges from -1 and 1; the higher the absolute value of Moran's I statistic, the stronger the spatial autocorrelation. It should be noted that for this preliminary assessment, the logarithm of housing prices in Hong Kong Dollars has been utilised without controlling for any housing characteristics. The spatial weight adopted is the inverse of the Euclidean distance between a pair of transactions. The Moran's I statistic is compiled by the software GeoDA developed by the Spatial Analysis Laboratory at the University of Illinois.

Table 2: Moran's I test for the spatial autocorrelation of Taikooshing's transactions from 1992 to 2006

Year	Number of transactions	Moran's I statistic	P-value
1992	1330	0.7024	0.001
1993	1518	0.7106	0.001
1994	1048	0.7563	0.001
1995	1130	0.6046	0.001
1996	2100	0.6013	0.001
1997	1800	0.6244	0.001
1998	967	0.5511	0.001
1999	730	0.5888	0.001
2000	625	0.4619	0.001
2001	600	0.4858	0.001
2002	583	0.5917	0.001
2003	625	0.5775	0.001
2004	910	0.5539	0.001
2005	831	0.5062	0.001
2006	629	0.5707	0.001

[3] Moran's I statistic is defined as I $= (N / \sum_i \sum_j w_{ij})(\sum_i \sum_j w_{ij}(p_i - \bar{p})(p_j - \bar{p}) / \sum_i (p_i - \bar{p})^2)$, in which N is the number of spatial units; p_i is a variable of interest, and w_{ij} is the spatial weight.

The Moran's I results for each year from 1992 to 2006 are shown in Table 2, which shows that the spatial autocorrelation differs significantly from zero at the 1% level of significance. Therefore, spatial autocorrelation is clearly present in Taikooshing's quality-unadjusted housing prices, and this prompted us to further test if spatial autocorrelation would still exist after adjusting for housing quality.

THE HEDONIC PRICING MODEL WITH SPATIAL AUTOCORRELATION

We construct a Spatial Hedonic Pricing (SHP) model in which a spatially lagged price variable is included as one of the independent variables to account for any spatial autocorrelation. For each transacted unit, i, at time t_i, the corresponding spatially lagged price variable is the inverse distance-weighted average price of previous transactions that took place over a time window of the past T days (i.e., transactions within $[t_i - T, t_i - 1]$) and can be expressed as $\sum_{\substack{\text{for all } j \text{ with} \\ t_j \in [t_i - T, t_i - 1]}} w_{ij} P_{jt_j}$. As mentioned in the Literature Review, although Can and Megbolugbe (1997) applied a similar model to housing transactions in Florida and found significant spatial autocorrelation, they did not explore the length of the time window. However, we consider that the magnitude of spatial autocorrelation may change when different time windows of previous transactions are used. We therefore try different values of T in various SHP models and see how spatial autocorrelation changes. If buyers and sellers put more weights on recent comparables, spatial autocorrelation should become smaller when we expand the time window. Apart from the spatially lagged price variable, we also include other independent variables, namely building age, floor level, flat size, full sea view and partial sea view. We further include a set of quarterly time dummies to account for changes in market conditions over time. We define all the variables in our SHP model below:

Dependent variable

P_{it_i} = the natural logarithm of the housing price of transacted unit i at time t_i;

Independent variables:

$\sum_{\substack{\text{for all } j \text{ with} \\ t_j \in [t_i - T, t_i - 1]}} w_{ij} P_{jt_j}$ = spatially weighted average price of units transacted within $[t_i - T, t_i - 1]$;

$bage_i$ = the building age of transacted unit i;
$level_i$ = the floor level of transacted unit i;
$size_i$ = the flat size of transacted unit i;

$$ty_{ij} = \begin{cases} 1 & if\ transacted\ unit\ i\ fell\ within\ period\ j \\ 0 & if\ transacted\ unit\ i\ did\ not\ fall\ within\ period\ j \end{cases}$$

where j = 1, 2...59 refers to the 2nd quarter of 1992, the 3rd quarter of 1992…4th quarter of 2006;

$$view_{i1} = \begin{cases} 1 & if\ transacted\ unit\ i\ has\ a\ full\ seaview \\ 0 & if\ transacted\ unit\ i\ lacks\ a\ full\ seaview \end{cases}$$

$$view_{i2} = \begin{cases} 1 & if\ transacted\ unit\ i\ has\ a\ partial\ seaview \\ 0 & if\ transacted\ unit\ i\ lacks\ a\ partial\ seaview \end{cases}$$

Equation (1) below shows the empirical form of our SHP model. We use a quadratic specification to allow for any non-linear relationship between some of the independent variables (namely $bage_i$, $level_i$, and $size_i$) and housing prices. Equation (1) will be estimated with six different time windows: T = 30, 60, 90, 120, 150, and 180 days. The parameters are assumed to be fixed for each time window. The error term is assumed to be homoskedastic $(\sigma^2(\varepsilon_{it_i}) = \sigma^2)$ with zero covariance $(\sigma(\varepsilon_{it_i}, \varepsilon_{jt_j}) = 0)$.

$$P_{it_i} = \alpha_o + \alpha_1(bage_i) + \alpha_2(level_i) + \alpha_3(size_i) + \alpha_4(bage_i^{\,2}) + \alpha_5(level_i^{\,2}) + \alpha_6(size_i^{\,2}) + \sum_{j=1}^{59}\beta_j ty_{ij} + \sum_{k=1}^{2}\gamma_k view_{ik} + \rho\sum_{j=1}^{n} w_{ij}P_{jt_j} + \varepsilon_{it_i} \qquad (1)$$

where:

n = the total number of transactions;
α_o = the intercept constant;
α_1 to α_6 = the parameter of $bage_i$, $level_i$, $size_i$, and their square terms;
β_j = the parameters of ty_{ij}, in which j=1, 2, 3…59;
γ_1 to γ_2 = the parameter of $view_{i1}$ and $view_{i2}$;
ρ = the parameter of the spatially lagged price variable (i.e., spatial autocorrelation);

$$w_{ij} = \begin{cases} 1/d_{ij} & if\ 0 < t_i - t_j \le 30 * \Gamma\ and\ coordinates\ of\ transaction\ i \ne coordinates\ of\ transaction\ j \\ 1/\min_{i \ne j}\{d_{ij}\} & if\ 0 < t_i - t_j \le 30 * \Gamma\ and\ coordinates\ of\ transaction\ i = coordinates\ of\ transaction\ j \\ 0 & if\ t_i - t_j \le 0\ or\ t_i - t_j > 30 * \Gamma \end{cases}$$

Γ = 1, 2, 3, …, 6;

d_{ij} = the horizontal distance between units *i* and *j*; and

ε_{it_i} = the error term of the model.

The spatial weight, w_{ij}, requires more explanation. It is the weight that specifies the spatial proximity (inverse distance) between transaction *i* at time t_i and transaction *j* at time t_j, in which the duration between the two transactions, $t_i - t_j$, has to be smaller than a given time window (T = 30 * Γ days, during which Γ= 1, 2, …, 6). If transaction j occurs T days before transaction i (i.e., $t_i - t_j > T$) or it occurs after or on the same date as transaction i (i.e., $t_i - t_j \leq 0$), the former is deemed irrelevant in influencing the latter and the spatial weight assigned for the pair of transactions is zero (i.e., $w_{ij} = 0$). Under this setting, the spatially lagged price variable $\sum_{\substack{\text{for all } j \text{ with} \\ t_j \in [t_i - T, t_i - 1]}} w_{ij} P_{jt_j}$, with respect to transaction i, can be expressed in terms of the *n* relevant transactions under consideration (i.e., $\sum_{j=1}^{n} w_{ij} P_{jt_j}$ in (1)), and the time window essentially becomes a classification variable Γ for specifying the spatial weight. We start with a time window of 30 days (Γ= 1) and re-estimate the model by increasing the value of Γ until the adjusted R^2 of the model starts to decline. As shown below, the adjusted R^2 starts to decline when Γ= 6 (i.e., T=180 days).

Another condition concerning w_{ij} is the coordinates of the transacted units. For units *i* and *j* with distinct coordinates, the weight is generally defined as $1/d_{ij}$. However, for two units with the same coordinates (i.e., units in the same building), the distance between them is zero by construction and their inverse distance would be undefined. In order to ensure that the weight is defined and units in close vicinity take on a large spatial weight, the spatial weight for units in the same building is assumed to be not less than the spatial weights for the nearest building ($1/\min_{i \neq j}\{d_{ij}\}$).

In Equation (1), our key interest is ρ, a measure of the overall level of spatial autocorrelation among the transactions after controlling for housing quality and time effects. Since more recent comparables should weigh more than older ones, we expect that a longer time window (T) will reduce the degree of spatial autocorrelation (ρ).

Regarding the computation of the maximum likelihood estimates (MLE) for each of the SHP models, the model specification only requires the housing price of each subject unit to depend on a fixed period of prior transactions. The transactions taking place after that for the subject unit or beyond the fixed period concerned would be weighted by zero in the spatially lagged price variable. The spatial weight matrix of the SHP model can be constructed in the form of a lower triangular matrix[4]. As a result, the MLE estimates for the SHP model are equivalent to those obtained from the Ordinary Least Squares (OLS) estimation.

EMPIRICAL RESULTS

To examine the effect of the change in the time window on the spatial autocorrelation in housing prices, we use Taikooshing's transaction data to estimate the parameters of Equation (1). We expect the coefficient of spatial autocorrelation to be significant and positive, but its magnitude would decrease with an increasing duration of the time window for previous transactions. Based on the results of previous studies, the expected impacts of the other independent variables in the SHP models are as follows:

(i) flat size has a positive effect on housing prices;
(ii) floor level has a positive effect on housing prices;
(iii) building age has a negative effect on housing prices; and
(iv) sea view has a positive effect on housing prices, and the effect of a "full sea view" is larger than that of a "partial sea view".

The estimation results of Equation (1) are summarized in Table 3. There are six sets of results, with each column representing different assumptions of time windows. Before showing the model parameters, the table first presents the adjusted R-squared (Adj. R^2) and the Root Mean Square Error (RMSE). A high Adj. R^2 and low RMSE indicate that a model provides a good goodness-of-fit. In all our cases, we obtain a high level of Adj. R^2 (from 0.81196 for a 30-day window to 0.81357 for a 150-day window) and small RMSE (from 0.18182 for a 150-day window to 0.18260 for a 30-day window). Thus, the SHP model seems to fit Taikooshing's transaction data better when a larger time window is used. This suggests that units transacted as far back as 150 days could influence current prices when the spatial dimension is taken into account. This is in stark contrast to other aggregate, notably time-series studies, which found not more than a quarter (90-day) lag in modelling the Hong Kong's efficient housing market (e.g. Newell and Chau 1996; Wong et al. 2007). Nevertheless, the improvement in Adj. R^2 and RMSE is slight, so we may not be able to draw a definitive conclusion in terms of the goodness-of-fit.

[4] A lower triangular matrix is a square matrix for which all the entries above the main diagonal are zero.

Table 3: Summary of the estimated parameters of the SHP models with different time windows

Time window	180 days		150 days		120 days		90 days		60 days		30 days	
Adj. R^2	0.81326		0.81357		0.81287		0.81221		0.81213		0.81196	
RMSE	0.18197		0.18182		0.18216		0.18248		0.18252		0.18260	
Variable	**Estimate**	**P-value**	**Estimate**	**P-value**	**Estimate**	**P-value**	**Estimate**	**P-value**	**Estimate**	**P-value**	**Estimate**	**P-value**
Spatial	3.0561E-03	<.0001	3.4603E-03	<.0001	3.7255E-03	<.0001	3.8213E-03	<.0001	4.5378E-03	<.0001	6.5653E-03	<.0001
Bage	-1.1565E-02	<.0001	-1.1841E-02	<.0001	-1.2377E-02	<.0001	-1.2950E-02	<.0001	-1.3336E-02	<.0001	-1.3775E-02	<.0001
Level	1.1222E-02	<.0001	1.1272E-02	<.0001	1.1364E-02	<.0001	1.1487E-02	<.0001	1.1507E-02	<.0001	1.1527E-02	<.0001
Size	1.8029E-03	<.0001	1.8021E-03	<.0001	1.7954E-03	<.0001	1.7906E-03	<.0001	1.7871E-03	<.0001	1.7820E-03	<.0001
$Bage^2$	-2.8723E-04	<.0001	-2.8269E-04	<.0001	-2.7616E-04	<.0001	-2.6978E-04	<.0001	-2.6545E-04	<.0001	-2.5990E-04	<.0001
$Level^2$	-2.1707E-04	<.0001	-2.1872E-04	<.0001	-2.2186E-04	<.0001	-2.2616E-04	<.0001	-2.2686E-04	<.0001	-2.2659E-04	<.0001
$Size^2$	-2.1014E-07	<.0001	-2.1038E-07	<.0001	-2.0917E-07	<.0001	-2.0864E-07	<.0001	-2.0841E-07	<.0001	-2.0777E-07	<.0001
Partial Sea view	2.5225E-02	0.0002	2.4737E-02	0.0003	2.3500E-02	0.0006	2.1639E-02	0.0017	2.0874E-02	0.0024	1.8906E-02	0.006
Full Sea view	4.9080E-02	<.0001	4.8205E-02	<.0001	4.6614E-02	<.0001	4.3604E-02	<.0001	4.2353E-02	<.0001	4.1533E-02	<.0001

The table also reports the parameters estimated from the SHP model and their significance in terms of p-value (i.e., a t-test on the null hypothesis that a parameter equals 0). We will discuss each variable in turn.

Spatial autocorrelation and the length of the time window

The first parameter (Spatial) reports the degree of spatial autocorrelation for each time window. As expected, it is positive and significantly different from zero at $p < 0.01$ across different windows. We conclude that positive spatial autocorrelation is present in the transactions in Taikooshing. That is, past transactions influence current prices, with the more recent ones having a greater influence. Also, since different time windows (ranging from 30 to 180 days) conclude the same result, the *presence* of positive spatial autocorrelation in housing prices is highly robust in our SHP model.

After establishing its presence, we want to know if the degree of spatial autocorrelation changes with the length of the time windows. Table 3 indicates that the magnitude of spatial autocorrelation decreases from the highest level of 0.00657 for a time window of 30 days to the lowest level of 0.00306 for a time window of 180 days. We conclude that spatial autocorrelation decreases as the time windows of previous transactions increases. One probable explanation for this is that the newer neighbourhood sales are weighted more than the older neighbourhood ones in the formation of current prices. In other words, buyers and sellers look back for comparable transactions when they price a subject property. Also, for two comparables located equally far from a subject property, they treat the more recently transacted one as more informative and put more weight on it. Thus, this empirical result confirms our hypothesis that the length of a time window has a bearing on spatial autocorrelation in housing prices. Specifically, a larger time window tends to reduce the magnitude of spatial autocorrelation. This implies that one must adjust for the effect of time windows when comparing the magnitudes of spatial autocorrelations across studies (e.g. in meta-analysis).

Together with the goodness-of-fit measures discussed above, our SHP model offers some insights for model specification. First, time windows should be considered when constructing a spatial econometric model. If there is no a priori knowledge of its length, one could try different lengths like what we have done here. From a goodness-of-fit perspective, the optimal length for a spatial model may not be consistent with that for time series models. Second, the goodness-of-fit measures only provide guidance on the choice of time windows that capture all relevant past information, but it does not tell which relevant information is more influential. To explicitly account for the stronger effect of more recent sales, which is what we have found, further research can be conducted to model the time decay function of past sales. For example, apart from the inverse distance weight, one could also allow for an inverse time interval weight. A larger weighting should be given to more recent comparable transactions.

The flat size, flat age, and floor level effects

The signs of the estimated parameters of flat size, building age, and floor levels in the SHP models are consistent with our hypotheses. All the estimated parameters are significantly different from zero at the $p < 0.01$ level. For flat size and floor level, the parameters are positive; while the parameter for building age is negative. Also, the estimated parameters for squared flat size, squared building age, and squared floor level are all negative and significantly different from zero. Since these parameters are smaller compared to the estimated parameters of the corresponding quality variables, we conclude that the quality variables of flat size and floor level are positively correlated with housing prices. However, because the estimated parameters of the associated squared variables are negative, if flat size is large and/or floor level is high, the effects of both on housing prices increase at a diminishing rate. As for the effect of building age on housing prices, as both the estimated parameters for building age and squared building age are negative, building age is negatively correlated with housing prices. This implies that as a building ages, its negative effect on housing prices increases at an accelerating rate.

The sea view effect

As all the estimated parameters of the SHP models for the indicator variables of "full sea view" and "partial sea view" are positive and significantly different from zero at $p < 0.01$, we conclude that when compared to the effect of "no sea view" included in the intercept, a sea view has a positive effect on housing prices and those units with the locational characteristic of "sea view" will be valued higher. Moreover, as the magnitude of the estimated parameter of a "full sea view" in each SHP model (i.e., 0.049080 to 0.041533 for models with time windows of 180 days to 30 days) is larger than that for a "partial sea view" (i.e., 0.025225 to 0.018906 for models with time windows of 180 days to 30 days), the effect of a "full sea view" on housing prices should be larger than that of a "partial sea view". The plausible reasons for these findings are attributed to the fact that units with a sea view are limited in supply, so buyers who want to purchase them have to compete with others with the same idea, which drives up housing prices. As for the findings on the effects of a "full sea view" and "partial sea view" on housing prices, they are also consistent with the general market observations because in terms of aesthetics, units with a "full sea view" are generally superior in quality to units with only a "partial sea view". Thus, if all other quality factors are held constant, buyers have to pay more for units with a "full sea view".

CONCLUSION

We hypothesize that the magnitude of spatial autocorrelation changes with the time windows of previous transactions, to which buyers and sellers refer. This nature of spatial autocorrelation has never been proposed and tested empirically. We introduce the Spatial Hedonic Pricing (SHP) model for various time windows. The empirical results show that spatial autocorrelation varies with time windows and its magnitude decreases as the time window becomes longer. Hence, newer nearby property transactions tend to have a stronger impact on the transaction price of a subject property. This suggests that newer neighbourhood sales should be weighted more than older neighbourhood sales when buyers and sellers research past price information. Moreover, as the Adj. R^2 increases with longer time windows up to a lagged period of 150 days, nearby property transactions of less than 150 days are deemed to contribute to the price searching process. This result is different from aggregate studies based on time series analyses, which often suggest a shorter lagged period (e.g. a quarter). Finally, we illustrate that time windows of previous transactions should not be treated as pre-determined. The exact duration of previous transactions included in real estate research depends on the housing market chosen and can be assessed based various goodness-of-fit statistics (e.g. Adj. R^2).

Despite our focus on the transactions in only one Hong Kong estate, the SHP model can be extended to cover a larger area of real estate transactions in Hong Kong or any other highly populated city with high-rise apartments (i.e., Tokyo or Singapore). However, as a wider catchment area of housing transactions will involve more heterogeneous properties compared to properties within a single estate, the heterogeneous nature of properties implies that there will be the more serious problem of missing variables, which would result in spatial autocorrelation. These missing variables may not decline as the time windows of previous transactions increase. One contribution of this study is that we used a sample of transactions of very homogeneous housing units, which minimize the chance of missing variables inducing spatial autocorrelation. Our study also suggests that spatial autocorrelation varies over time. More research is needed to explain this variation and the application of the SHP model to more heterogeneous housing markets.

The SHP model has many practical applications, such as property valuation and real estate price index construction. The Hong Kong housing market is active and volatile. There is a great need for quick valuations of residential properties. Thus, when compared to the subjective valuation approach of the grid adjustment method adopted by appraisers for selecting and analyzing comparable properties (Kang and Reichert 1991), the SHP model establishes an explicit relationship between housing prices, quality factors, and the housing prices of nearby transactions, and can be deployed to perform a more accurate mass valuation of residential properties so that investors, government bodies, surveyors, and bankers can make decisions based on more timely and accurate analyses.

ACKNOWLEDGEMENTS

We gratefully acknowledge the support by the Research Grants Council of the Hong Kong Special Administrative Region (RGC Reference Number: HKU 7108/08E).

REFERENCES

Anselin, L. 1988. *Spatial Econometrics: Methods and Models*. Dordrecht; Boston: Kluwer Academic Publisher.

Anselin, L. 1998. GIS Research Infrastructure for Spatial Analysis of Real Estate Markets. *Journal of Housing Research* 9(1): 113 – 133.

Anselin, L. 2001a. Rao's Score Test in Spatial Econometrics. *Journal of Statistical Planning and Inference* 97(1): 113 – 139.

Anselin, L. 2001b. *Spatial Econometrics in A Companion to Theoretical Econometrics*, ed. Baltagi, B. H., 310 – 330. Oxford: Blackwell Publisher.

Anselin, L. and Bera, A. 1998. Spatial Dependence in Linear Regression Models with an Introduction to Spatial Econometrics. *In Handbook of Applied Economic Statistics*, ed. Ullah, A. and Giles, David. E. A., 237 – 289. New York: Marcel Dekker.

Archer, W. R., Gatzlaff, D. H. and Ling, D. C. 1996. Measuring the Importance of Location in House Price Appreciation. *Journal of Urban Economics* 40(3): 334 –

353.

Bailey, M.J., Muth, R.F. and Nourse, H.O. 1963. A Regression Method for Real Estate Price Index Construction. *Journal of American Statistical Association* 58: 933 – 942.

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

Brasington, D.M. 1999. Which Measures of School Quality Does the Housing Market Value? *The Journal of Real Estate Research* 18(3): 395-414.

Can, A. 1990. The Measurement of Neighbourhood Dynamics in Urban House Prices. *Economic Geography* 66(3): 254 – 272.

Can, A. and Megbolugbe, I. 1997. Spatial Dependence and House Price Index Construction. *The Journal of Real Estate Finance and Economics* 14: 203 – 222.

Case, B., Clapp, J., Dubin, R.A. and Rodriguez, M. 2004. Modelling Spatial and Temporal House Price Patterns: A Comparison of Four Models. *The Journal of Real Estate Finance and Economics* 29(2): 167 – 91.

Chau, K. W. 1997. Political Uncertainty and the Real Estate Risk Premiums in Hong Kong. *Journal of Real Estate Research* 13(3): 297–315.

Chau, K. W., Wong, S. K., Yiu, C. Y. and Leung, H. F. 2005. Real Estate Price Indices in Hong Kong. *Journal of Real Estate Literature* 13(3): 337-356.

Clapp, John M., Kim, H., and Gelfand, Alan E. 2002. Predicting Spatial Patterns of House Prices Using LPR [Local Percentile Rank] and Bayesian Smoothing. *Real Estate Economics* 30(4): 505 – 532.

Cliff, A.D. and Ord, J.K. 1981. *Spatial Processes: Models and Applications*. London: Pion.

Colwell, P. F., Gujral, S. S., and Coley, C. 1985. The Impact of A Shopping Center on the Value of Surrounding Properties. *Real Estate Issues* 10(1): 35–39.

Colwell, P.F. 1990. Power Lines and Land Values. *Journal of Real Estate Research* 5(1): 117–127.

Court, A.T. 1939. Hedonic Price Indexes with Automotive Examples. *In The Dynamics of Automobile Demand*, ed. American Statistical Association, 99-117. New

York: General Motors Corporation.

Des Rosiers, F., Lagana, A. and Theriault, M. 2001. Size and Proximity Effects of Primary Schools on Surrounding House Values. *Journal of Property Research* 18(2): 149 – 168.

Des Rosiers, F., Lagana, A., Theriault, M. and Beaudoin, M. 1996. Shopping Centers and House Values: An Empirical Investigation. *Journal of Property Valuation & Investment* 14(4): 41–62.

Deutsch, S.J. and Ramos, J.A. 1986. Space-Time Modelling of Vector Hydrologic Sequences. *Journal of the American Water Resources Association* 22(6): 967 – 981.

Dubin, R. A. 1988. Estimation of Regression Coefficients in the Presence of Spatially Autocorrelated Error Terms. *Review of Economics and Statistics* 70(3): 466 – 474.

Dubin, R.A. 1992. Spatial Autocorrelation and Neighbourhood Quality. *Regional Science and Urban Economics* 22(3): 433 – 452.

Dubin, R.A. 1998. Predicting House Prices Using Multiple Listings Data. *The Journal of Real Estate Finance and Economics* 17(1): 35 – 59.

Dubin, R.A. 2003. Robustness of Spatial Autocorrelation Specifications: Some Monte Carlo Evidence. *Journal of Regional Science* 43(2): 221 – 248.

Dubin, R., Pace, R.K. and Thibodeau, T.G. 1999. Spatial Autoregression Techniques for Real Estate Data. *Journal of Real Estate Literature* 7(1): 79 – 96.

Gelfand, A.E., Ghosh, S.K., Knight, J.R. and Sirmans, C.F. 1998. Spatio-Temporal Modelling of Residential Sales Data. *Journal of Business and Economic Statistics* 16(3): 312 – 321.

Goetzmann, W.N. and Spiegel, M. 1997. A Spatial Model of Housing Returns and Neighbourhood Substitutability. *Journal of the Real Estate Finance and Economics* 14: 11 – 32.

Grether, D.M. and Mieszkowski, P. 1980. The Effects of Non-residential Land Uses on the Prices of Adjacent Housing: Some Estimates of Proximity Effects. *Journal of Urban Economics* 8(1): 1 – 15.

Griliches, Z. 1971. Hedonic Price Indexes for Automobiles: An Econometric Analysis of Quality Change. *In Price Indexes and Quality Change: Studies in New*

Methods of Measurement, ed. Griliches, Z., 55 – 87. Cambridge, Massachusetts: Harvard University Press.

Hui, E. and Lui, T. Y. 2002. Rational Expectations and Market Fundamentals: Evidence from Hong Kong's Boom and Bust Cycles. *Journal of Property Investment and Finance* 20(1): 9 – 22.

Kang, H. and Reichert, A.K. 1991. An Empirical Analysis of Hedonic Regression and Grid-adjustment Techniques in Real Estate Appraisal. *Journal of the American Real Estate and Urban Economics Association* Spring: 70 – 91.

Kelejian, H.H. and Prucha, I.R. 1998. A Generalised Spatial Two-Stage Least Squares Procedure for Estimating a Spatial Autoregressive Model with Autoregressive Disturbances. *The Journal of Real Estate Finance and Economics* 17(1): 99 – 121.

Kelejian, H.H. and Prucha, I.R. 1999. A Generalised Moments Estimator for the Autoregressive Parameter in a Spatial Model. *International Economic Review* 40(2): 509 – 533.

Kelejian, H.H. and Robinson, D.P. 1993. A Suggested Method of Estimation for Spatial Independent Models with Autocorrelated Errors, and an Application to a County Expenditure Model. *Papers in Regional Science* 72(3): 297 – 312.

Kyriakidis, P.C. and Journel, A.G. 1999. Geostatistical Space – Time Models: A Review. *Mathematical Geology* 31(6): 651 – 684.

Lancaster, K.J. 1966. A New Approach to Consumer Theory. *The Journal of Political Economy* 74(2): 132 – 157.

LeSage, J and Pace, R. K. 2009. *Introduction to Spatial Econometrics*. Boca Raton: Chapman & Hall/CRC.

Linneman, P. 1980. Some Empirical Results on the Nature of Hedonic Pricing Function for the Urban Housing Market. *Journal of Urban Economics* 8(1): 47 – 68.

Malpezzi, S. 2003. Hedonic Pricing Models: A Selective and Applied Review. *In Housing Economics and Public Policy: Essays in Honour of Duncan Maclennan*, ed. Gibb, K. and O'Sullivan, A. London: Blackwelling Publishing.

Miron, J. 1984. Spatial Autocorrelation in Regression Analysis: A Beginner's Guide. *In Spatial Statistics and Models*, ed. Gaile, G.L. and Willmott, C.J, 201 – 222. Boston: D. Reidel Publishing Company.

Newell, G. and Chau, K. W. 1996. Linkages between Direct and Indirect Property Performance in Hong Kong. *Journal of Property Finance* 7(4), 9 – 29.

Odland, J. 1988. *Spatial Autocorrelation*. Newbury Park, CA: Sage Publications.

Pace, R.K., Barry, R., Clapp, J.M. and Rodriguez, M. 1998. Spatiotemporal Autoregressive Models of Neighbourhood Effects. *The Journal of Real Estate Finance and Economics* 17(1): 15 – 33.

Pace, R. K. and Gilley, O. W. 1998. Generalising the OLS and Grid Estimators. *Real Estate Economics* 26(2): 331 – 347.

Pfeifer, P.E. and Bodily, S.E. 1990. A Test of Space-Time ARMA Modelling and Forecasting of Hotel Data. *Journal of Forecasting* 9(3): 255 – 272.

Pfeifer, P.E. and Deutsch, S.J. 1980. A Three-Stage Iterative Procedure for Space-Time Modelling. *Technometrics* 22(1): 35 – 47.

Rosen, S. 1974. Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition. *The Journal of Political Economy* 82(1): 34 – 55.

Silk, John. 1979. *Statistical Concepts in Geography*. London: George Allen & Unwin, Ltd.

Stoffer, D.S. 1986. Estimation and Identification of Space-Time ARMAX Models in the Presence of Missing Data. *Journal of the American Statistical Association* 81(395): 762 – 772.

Sun, H., Tu, Y. and Yu S. 2005. A Spatio-Temporal Autoregressive Model for Multi-Unit Residential Market Analysis. *The Journal of Real Estate Finance and Economics* 31(2): 155 – 187.

Upton, G. and Fingleton, B. 1985. *Spatial Data Analysis by Example*. New York: Wiley.

Wilhelmsson, M. 2002. Spatial Models in Real Estate Economics. *Housing, Theory and Society* 19(2): 92 – 101.

Wong, S. K., Chau, K. W. and Yiu, C. Y. 2007. Volatility Transmission in the Real Estate Spot and Forward Markets. *Journal of Real Estate Finance and Economics* 35: 281 – 293.

Email contact: hrrbckw@hku.hk