SPATIAL AUTOCORRELATION IN HEDONIC MODEL: EMPIRICAL EVIDENCE FROM MALAYSIA[1]

Suriatini Ismail, Abdul Hamid Mar Iman, Norhaya Kamaruddin, Hishamuddin Mohd Ali, Ibrahim@Atan Sipan, Renuga Navaneethan
Department of Property Management
Faculty of Geoinformation Science & Engineering
Universiti Teknologi Malaysia
81310 Skudai, Johor, Malaysia
suriatini@fksg.utm.my, hamid@fksg.utm.my, norhaya@fksg.utm.my, hisham@fksg.utm.my, Ibrahim@fksg.utm.my, renu_utm2004@yahoo.com

Abstract

The need to consider spatial autocorrelation in hedonic price modelling is paramount since reliability and accuracy are important in any housing market analysis. Although hedonic modelling is widely accepted, the issue of spatial autocorrelation has received little attention in house price studies in Malaysia. This paper examines the presence of spatial autocorrelation in the hedonic price model using data from Malaysia. A pilot sample of eighty three-year sale observations (2001-2003) and nine property attributes were used to construct four types of double-log competing models namely ordinary least square model (OLS), spatial autoregressive model (SAR), spatial error model (SEM) and general spatial model (SAC). The results have indicated the presence of spatial lag and spatial error types of autocorrelation based on the SAC model. However, based on the out-of-sample test, the models were found to be equivalent in terms of statistical quality. Despite the findings, it is important to stress that taking into account the possible phenomenon of spatial autocorrelation is critical in using hedonic housing price model to ascertain that the model chosen is free from spatial autocorrelation problem. If the problem does exist, we can adopt the appropriate steps to improve the statistical quality of price-estimating model being chosen.

Keywords: hedonic model, house price, spatial autocorrelation

[1] This paper is based on the research carried out under the Fundamental Research Grant Scheme (FRGS) Vote 78065, received from the Ministry of Higher Education Malaysia (MOHE). We would like to thank MOHE whose funding has made the research possible.

1.0 INTRODUCTION

Past real estate studies have shown that house price hedonic modelling is popular for housing market analysis. It involves regressing three main categories of property characteristics namely structural, accessibility and neighbourhood against property prices. The inclusion of many variables representing such characteristics can give rise to multicollinearity, spatial autocorrelation and heteroscedasticity. Of these, spatial autocorrelation has received the least attention in the real estate literature (Suriatini, 2005). According to Des Rosiers *et al.* (2000) spatial autocorrelation is a problem that can cause imperfections in the application of hedonic models in housing market analysis. Therefore, it is important to consider this phenomenon in hedonic price modelling (Malpezzi, 2003).

The fact that spatial autocorrelation is less commonly given attention in real estate research is not surprising because it is more complex to deal with. Nevertheless, Suriatini (2006) presents a literature review on spatial autocorrelation in real estate studies highlighting a number of studies that have considered spatial autocorrelation in house price models in many countries except Malaysia. Unlike serial correlation which is one-dimension time-back-tracking, spatial correlation is two-dimensional and no particular direction is ruled out *a priori* (Bailey and Gatrell, 1995, 4). In this context, this paper aims at bridging the gap by examining spatial autocorrelation in hedonic model from Malaysia.

This paper is organised as follows. Section 2 reviews the literature focusing on the importance of considering spatial autocorrelation in hedonic price studies. It highlights the motivation of the paper and focuses on house price studies in Malaysia. Section 3 describes the data and analyses. Section 4 presents the results and discussions. Section 5 concludes this paper.

2.0 LITERATURE REVIEW

Hedonic modelling is a regression-based technique founded by researchers like Haas (1920), Court (1939), Griliches (1961; 1971) and Lancaster (1966), as cited in Hamid (2001). Rosen (1974) formalises the concept (Hamid, 2001; Gelfand *et al.,* 2004). Originally developed within an economic framework to study essentially composite commodities such as cars, refrigerators, washing machines and personal computers (Orford, 1999, 23), it was not until the development of Rosen's (1974) theory of implicit prices of differentiated products that the hedonic approach has been widely used as a vehicle in econometric modelling for analysing prices of heterogeneous economic goods.

2.1 Basic aspects of hedonic price modelling

Rosen (1974) set the founding theory that property is bought and sold as a bundle of attributes, each of which contributes to the price of the property. Since individual attributes do not have an explicit price, it follows that the price paid for a particular property is the sum of the implicit prices that the market ascribes to the various attributes contained in the bundle. Thus, the theory of hedonic price functions provides a framework for the analysis of differentiated goods, like housing units, whose individual features do not have observable market prices (Pagourtzi *et al.*, 2003). Hedonic price can be represented as follows (Hamid, 2001):

$$P = p(X_1, X_2, \ldots, X_k)$$

where P represents the explicit price of property, *p* represents the implicit price of a vector of k property characteristics, X, comprising structural characteristics, accessibility and neighbourhood quality.

Hedonic modelling is a well-established technique that has been applied in house pricing for many years. As the above general equation shows, it is used for identifying the influence of many factors on house prices with a view to making inferences about non-observable values of different attributes like air quality, noise, commuter access (railway, subway or highway) and neighbourhood amenities. In other words, house price hedonic modelling is concerned with market implicit pricing of each house characteristic which can be done by regressing or correlating a measure of house prices against the corresponding property characteristics (Cass, 1980; Fletcher *et al.*, 2000). Besides, the knowledge about the implicit prices of property attributes can be used to test for the existence of housing sub-markets.

Maclennan (1977, 68) outlined discernible purposes of hedonic modelling of house prices to include statistical explanation of house price determinants, relative importance of these elements through house price regression equation, deriving housing demand function, and testing alternative theories of residential location.

Hedonic modelling is commonly adopted in the valuation of residential properties in the U.S.A., particularly through mass appraisal, and has been used in the U.K. to a limited extent (Dunse and Jones, 1998). Construction of a property price index also makes extensive use of the hedonic technique.

There are two categories of variables involved in house price hedonic modelling namely dependent and independent variables. The dependent variable is a measure of house value, which can be represented by rents, sale price, asking price, or even owner's estimated price. In previous studies, the use of different measures of house value is evident. Rents are normally used in studies that involve rented properties while the rest are normally used in studies that involve owner occupied properties.

The use of sale prices is evident in Bowen *et al.* (2001), Watkins (2001), Whilhelmsson (2000), McCluskey *et al.* (2000), Fletcher *et al.* (2000), Des Rosiers *et al.* (2000), Adair *et al.* (1996) and Chen (1994). Although sale prices have shown to be the most preferable and widely used, asking prices seem to be acceptable when sales data availability is a problem (Cheshire and Sheppard, 1998; Henneberry, 1998). Owner's estimations are probably the least preferable in studies that attempt to examine market interactions of demand and supply, though, they are also pragmatic to use, provided that their tendency to be overestimated is acknowledged (Bourassa *et al.*, 1999).

The second category of variable in regression analysis consists of the independent variables that include all significant factors affecting prices. They are locational (or spatial) and structural factors.

While structural characteristics are conceptually straightforward, locational factors can be somewhat confusing. This is because different elements and different terms can represent locational factors. For example, Wilhelmsson (2000) has listed three main factors that households normally consider when buying a house namely its structural attributes, its location in relation to urban services, and its environmental attributes. The first refers to property's structural characteristics. The second refers to accessibility to urban facilities. The third can be classified as neighbourhood quality.

This study replicates that of Abdul Hamid's (2007) in terms of study area and choice of variables. This is to enable a comparison between his model, (which is free from multicollinearity, heterocedasticity, and serial autocorrelation) that did not consider spatial autocorrelation and the models in this study that do otherwise.

This study evaluates the competing models based on their comparative statistical quality (adjusted R^2, F-value, t-values, and coefficient signs) besides their predictive capability based on out-sample properties. Out-sample predictions are performed as follows (Pace, 2008: email communication):

Out-sample Prediction for OLS model: $yhat = X_{out} * \beta$

Out-sample Prediction for SEM model: $yhat = X_{out} * \beta + \rho W(y_{in} - X_{in}\beta)$

Out-sample Prediction for SAR model: $yhat = Z_{out} + \rho W(y_{in} - Z_{in})$

$Z = (I_{all} - \rho W)^{-1} X_{all} * \beta$ and use Z_{out} as the predictor

$(I - \rho W)^{-1} = \dfrac{1}{1-\rho}$ for row-standardized weight matrix

where, *yhat* represents predicted price, X_{out} represents independent variables of out-sample properties, β represents coefficients of independent variables, ρ represents spatial autocorrelation parameter, W represents spatial weight matrix, X_{in} represents dependent variables of in-sample properties and y_{in} represents actual property prices.

2.2 Main sources of hedonic problems

The purpose of regression analysis is not only to estimate the coefficient for the variables but also to draw inferences about the true coefficient. This means that the reliability of the inferences depends on the stability, and, hence, reliability of the estimated coefficient.

Statistics attempts to find structure in chaos (Cressie, 1989) and simplify complex phenomena. In implementing the Ordinary Least Squares techniques (OLS), the statistical assumptions made about the independent variables and the error term are critical to the valid interpretation of the regression estimates (Gujarati, 1995, 59). The simplest structure imposed assumes that observations of a phenomenon are taken under identical conditions, and independently from one observation to another; that is the errors are normal, independent and identically distributed (Cressie, 1989). Gujarati (2003, 66-76) lists ten assumptions of Markov Theorem in applying the Classical Linear Regression Model (CLRM). Adding another assumption of normality of errors to the list creates a Classical Normal Linear Regression Model (CNLRM) (Gujarati, 2003, 108).

In the presence of violations of the above assumptions, hypothesis tests about implicit marginal prices of individual attributes become potentially misleading (Bowen *et al*., 2001). Three main sources of hedonic problems that are related to violations of assumptions are multicollinearity, heteroscedasticity and spatial autocorrelation (Des Rosiers *et al*., 2001). When these violations occur, a substantial portion of price variability remains unexplained (Theriault *et al.,* 2003). These problems may occur due to the integration of large sets of variables into a single regression model (Des Rosiers *et al*., 2000; Theriault *et al.,* 2003). Neglecting or accounting only partially for the spatial elements of the property (Orford, 1999, 23) also can cause them. This paper focuses on the issue of spatial autocorrelation (or spatial dependence).

2.3 The issue of spatial autocorrelation in hedonic price modelling

According to Miron (1984), the effect of spatial dependence among the unobserved error term is twofold. Firstly, it makes OLS estimates of the t-test values unreliable. In other words, the t-values no longer tell accurately whether the included explanatory variables have a significant effect on the average house price. Secondly, it is no longer true that OLS estimates are relatively efficient, that is having small sampling variability associated with them (Miron, 1984, 205). The importance of these statistical elements in a hedonic price modelling substantiates the need to detect the existence of spatial autocorrelation in OLS residuals. This is due to the fact that spatial autocorrelation in sample data can alter the conclusions of statistical analyses performed without due allowance for the former. The existence of spatial autocorrelation does not provide minimum-variance unbiased linear estimators and produces a bias in the estimation of correlation coefficients and variances (Dutilleul *et al.,* 1993).

In practice, it is important to detect the existence of spatial autocorrelation in the OLS residuals. This is to help us judging the reliability of the hypothesis testing based on the model. In addition, Wiltshaw (1996) stresses that consideration of spatial autocorrelation applies to every cross-sectional empirical study of OLS. According to him, no particular case study whether imagined or real, can confirm the presence or absence of spatial autocorrelation in other market analyses. Each case must be analysed separately, just as done when testing for temporal autocorrelation, heteroscedasticity and other hedonic problems.

According to Overmars *et al*. (2003), if autocorrelation is detected in the regression residuals, this can imply that the regression model should have an autoregressive structure, or that non-linear relationships between the dependent and the independent variable are present, or that some important regressors are missing. There are two types of spatial autocorrelation namely spatial error dependence and spatial lag dependence (Suriatini, 2006). The first refers to the correlated errors that occur among the independent variables. It is also called spatial heteroscedasticity. It can rise from omitted variables, variable measurement error or misspecification of the functional form. The second refers to the correlated errors that occur between the dependent variables. It can be said to be true spatial autocorrelation.

In the study by Wilhelmsson (2002), the Moran's I test showed that real estate data are highly spatially dependent. Thus, he believed that even if one tries to account for spatial effects with the inclusion of distance and sub-markets dummies, one cannot reject the hypothesis of no spatial autocorrelation. The existence of spatial autocorrelation despite detailed hedonic specifications is evident in Harrison and Rubinfeld (1978) as well as Des Rosiers *et al.* (2001).

Ignoring spatial autococorrelation leads to serious violations of the underlying independence assumption of OLS regression (Paez *et al.*, 2001), which can result in erroneous statistical inference due to loss of much predictive power (Pace and Barry, 1997a, 1997b). The importance of detecting its existence as indicated by Bowen *et al.* (2001) is appealing. They stated that "…*without explicit investigation, the analyst has no way of knowing if there is a violation, and if so, where the violation implications lie along the 'subtle-to-severe' continuum" (Bowen et al., 2001, 472-473).*

The above statement supports Wiltshaw (1996) in that spatial autocorrelation is more an empirical issue than a theoretical one and, hence, should be tested in every case of cross sectional analysis.

According to Anselin (1998), the presence of spatial dependence in cross sectional geo-referenced data can be considered a nuisance or a substantive. He elaborates that it is a nuisance if the focus of analysis is on obtaining proper statistical inference (such as of estimation, hypotheses testing and prediction). Meanwhile, it is a substantive if the focus of analysis is on discovering the form of spatial interaction such as the precise nature of spatial spill-over, and the economic and social processes that lie behind it.

When spatial dependence is considered a nuisance, the main objective is to correct statistical procedures for the effect of the spatial autocorrelation, for example by increasing the sample size or by using robust methods or adjustments that incorporates the spatial autocorrelation in a regression error term. When it is a substantive, the main objective is to incorporate the structure of spatial dependence into a statistical model and how to estimate and interpret it. Anselin (1998) believes that the prevalence of spatial dependence in the cross-sectional data used in real estate analysis requires the application of appropriate techniques of spatial statistics and spatial econometrics for efficient estimation, valid inference and optimal prediction (Pace and LeSage, 2003).

A literature review on house price hedonic model discovers eight previous studies from Malaysia as listed in Table 1. The table shows that all of the studies involved cross sectional model. Most of them employed data of different years and areas in Johor Bahru. The fact that none of the studies has considered spatial autocorrelation serves as a motivation for this paper.

Table 1: Hedonic price studies in Malaysia

No.	Study	Study area	Date of Data	N	TS	CS	Type of Property	Notes/Availability of data for the current study
1	Hamid (2007)	Johor Bahru	2001-2003	125	√	√	Single & double-storey terraced houses	80 price observations 8 independent variables With x, y co-ordinates
2	Hamid (2006)	Johor Bahru	1997-2000	800	√	√	Single-storey terraced houses	Mean price value from property market report 18 independent variables 20 out-sample observations
3	Chau and Chin (2002)	Penang	1998-1999	120	√	√	Condominiums	No x,y co-ordinates
4	Chin et. al.(2004)	Penang	1996 &1998	422	√	√	Condominiums	No x,y co-ordinates
5	Azhari Husin (1990)	Johor Bahru	1984-1987	1522	√	√	Landed residential properties	No x,y co-ordinates
6	Azhari Husin and Mohd Ghazali Hashim (1994)	Johor Bahru	1984-1987	Not mentioned		√	Landed residential properties	Data set is missing from author's keeping though x, y co-ordinates were involved
7	Dzurllkarnain Daud *et al.* (1996)	Johor Bahru	1994-1995	125	√	√	Single & double-storey terraced houses	No x,y co-ordinates
8	Aminah Yusof (2006)	Johor Bahru	1990 1995 2000 2002	Not mentioned		√ √ √ √	Double storey terrace houses	No x,y co-ordinates

Note: TS=Time series, CS=Cross sectional, N= Number of observations

3.0 DATA AND ANALYSIS

This paper used the secondary data obtained from the study by Abdul Hamid (2007). As many as eighty three-year sale observations (2001-2003) and nine property attributes of double and single storey-terraced houses have been involved (Table 2). The data set covered three adjoining housing schemes of Taman Pelangi (48% of observations), Taman Sentosa (35% of observations), and Taman Sri Tebrau (17% of observations) located in the sub-urban of Johor Bahru City.

Table 2: Variables involved

Variable	Unit of Measurement
Dependent Variable	
House price (PRICE)	Malaysia Ringgit (RM)
Independent Variables	
Land Area (LA)	Square meters (sq. m.)
Gross Floor Area (GFA)	Square meters (sq. m.)
Ancillary Area (AFA)	Square meters (sq. m.)
Type of Terrace (TERRACE_1)	Dummy (single story terrace = 1)
Age (AGE)	Years
Condition of the Building (GCOND)	Dummy (good condition = 1)
Floor Finishes (FFINISH_2)	Dummy (class 2 - combination of parquet, terrazzo, semen, mosaic or tiles floor finishes = 1)
No. of Bedroom (BEDR_3)	Dummy (3 bedrooms = 1)
Distance from CBD (CBD)	Meters (m

Des Rosiers and Theriault (1992) believe that the identification and interpretation of complex phenomena such as spatial autocorrelation of hedonic models residuals would simply not be feasible without the help of a GIS. Nevertheless, Hamid (2001) has suggested two options for detecting spatial autocorrelation. Firstly, to estimate regression prediction errors and input these estimates to GIS data file for spatial display. Secondly, to use spatial correlation statistics such as Moran's I. This paper applies spatial statistics[2] in examining spatial autocorrelation in real estate data from Malaysia.

Test for the presence of spatial autocorrelation based on the OLS residuals was carried out using Moran's I test and Lagrange Multiplier Error (LM-Error) test. The procedure for computing statistics for these tests is available in Spatial Econometrics Toolbox (SET) for Matlab (www.spatial-econometrics.com).

[2] Spatial statistics consider spatial dependencies to provide inference that is more realistic, better prediction and more efficient parameter estimation (Pace *et al*, 1998). Spatial statistics is concerned with the methods of analysis that explicitly consider the spatial arrangement of the data (Martinez and Martinez, 2002, 465) and are appropriate tools for analysing spatial autocorrelation. They are the most useful tools for describing and analysing how various geographic objects (or events) occur or change across the study area (Lee and Wong, 2001, 132).

4.0 RESULTS AND DISCUSSIONS

Based on Table 3, the sampled properties were located about 1.8 – 3.4 km from the Johor Bahru city centre. The mean land area of the properties was about 162.2 sq. m. The range of land area was purposely controlled in the sample to avoid too much variance of this variable. On average, the gross floor area was about 72.5% the size of land area. The mean ancillary area of the properties was about 32 sq. m.

On the basis of standard deviation, gross floor area has more variation compared to land area and ancillary area. This means, the people in the market could have been more concerned about differences in floor area, making this factor a possibly more influential physical characteristic of residential properties.

The sample was rather dominated by double-storey 4-bedroom terraced residential. Besides, the majority of the properties, single- and double-storey terraced alike, were of good building condition and have either parquet, terrazzo, cement, mosaic or tile floor finishes. The mean age of the residential properties was about 22 years old. Since these are old housing schemes, the majority of the sampled residential properties have also undergone some renovation.

Table 3: Descriptive statistics

	N	Minimum	Maximum	Mean	Std. Deviation
PRICE	80	160000	433000	257111.10	64566.682
LA	80	143.06	216.82	162.1623	14.25105
GFA	80	69.21	218.32	117.6133	37.20585
AA	80	7.43	70.86	31.6084	11.07191
TERRACE_1	80	0	1	.56	.499
TERRACE_2	80	0	1	.44	.499
AGE	80	15	28	22.49	3.230
GCOND	80	0	1	.99	.112
BCOND	80	0	1	.01	.112
FFINISH_1	80	0	1	.98	.157
FFINISH_2	80	0	1	.03	.157
BED_3	80	0	1	.53	.503
BED_4	80	0	1	.48	.503
CBD	80	1842.11	3421.05	2580.2645	734.14389
Valid N (listwise)	80				

The results in Table 4 show a modest improvement in the adjusted R^2 for both SAC and SAR models compared to the OLS. The OLS and SEM seemed to be statistically equivalent in terms of R^2 and adjusted R^2. The statistical significance of regression hypersurface did not show much difference among the three models. In particular, applying SAC, SAR, and SEM models did not improve any variables that were originally statistically insignificant in the OLS model.

The results also show that spatial autocorrelation (indicated by the coefficients of rho and lambda) was statistically significant in determining property prices only in the SAC model. This was not the case for the SAR and SEM models.

Table 4: Comparison of Regression Results Based on Three Competing Models

	OLS	SAC	SAR	SEM
Dependent variable	**Natural Log price (LG_PRICE)**			
R^2	0.8713	0.8870	0.8760	0.8713
Adj, R^2	0.8547	0.8724	0.8601	0.8548
σ^2	0.0090	0.0069	0.0077	0.0079
Durbin-Watson	1.8834	-	-	
Log-likelihood	-	156.102	108.816	107.809
N	80	80	80	80
Independent Variable	**Coefficient (t-value)**	**Coefficient (t-value)**	**Coefficient (t-value)**	**Coefficient (t-value)**
constant	7.2363 (6.89)**	4.5890 (4.02)**	6.0575 (4.57)**	7.2394 (7.34)**
Log land area (LG_LA)	0.5856 (3.70)**	0.4445 (3.25)**	0.5240 (3.44)**	0.5854 (3.95)**
Log gross floor area (LG_GFA)	0.4079 (4.48)**	0.3805 (4.59)**	0.3891 (4.60)**	0.4069 (4.78)**
Log ancillary area (LG_AA)	0.0251 (0.66)	0.0211 (0.66)	0.0182 (0.52)	0.0248 (0.70)
Single-storey terraced (TERRACE_1)	-0.1334 (-2.65)**	-0.1179 (-2.85)**	-0.1341 (-2.89)**	-0.1340 (-2.84)**
Log age of house (LG_AGE)	-0.2377 (-2.88)**	-0.1477 (-2.18)**	-0.1948 (-2.38)**	-0.2373 (-3.07)**
Unit with good conditions (GCOND)	0.3136 (3.05)**	0.2915 (3.16)**	0.3083 (3.25)**	0.3133 (3.26)**
Ordinary type of floor finishes (FFINISH_2)	-0.0523 (-0.72)	-0.0492 (-0.81)	-0.0622 (-0.92)	-0.0530 (-0.78)
Three-bedroom unit (BED_3)	-0.0114 (-0.33)	-0.0095 (-0.35)	-0.0122 (-0.39)	-0.0114 (-0.36)
Log distance from central business district (LG_CBD)	0.0911 (2.08)**	0.0542 (1.66)*	0.0799 (1.92)**	0.0915 (2.22)**
ρ	-	0.2840 (3.03)**	0.1260 (1.33)	-
λ	-	-0.5530 (-2.42)**	-	0.0120 (0.06)

** Significant at 1% level; * significant at 5% level.

To partially gauge their predictive power, the models were compared based on in-sample's mean absolute percentage error (MAPE) calculations. Table 6 shows that, with the MAPE of around 7%, all models were equivalently accurate in terms of in-sample predictions with the SAC model perhaps slightly outperformed other models. Table 7 indicates that all the competing models have shown equivalent predictive capability based on the out-sample properties.

Table 5: Comparison of Basic Diagnostic Statistics

	Wald (OLS)	Lratio (OLS)	LM-Error (OLS)	LM-Sar (SAR)	Moran's I (OLS)
Computed Value	0.0018	0.0028	0.0022	0.2012	0.0032
Statistic	-	-	-	-	0.7296
Marginal Probability	0.9666	0.9582	0.9622	0.6537	0.4656
Mean	-	-	-	-	-0.0412
Standard Deviation	-	-	-	-	0.0609
Chi (critical) 0.01 value	6.6350	6.6350	17.6110	6.6350	-

Table 6: Comparison of In- and Out-Sample MAPE among the Competing Models

	MAPE	
	In-Sample	Out-Sample
Ordinary Least Squares Model (double-log) (OLS)	7.21	9.71
General Spatial Autocorrelation Model (SAC)	6.74	n.a.
Spatial Autoregressive Model (SAR)	7.04	9.73
Spatial Error Model (SEM)	7.21	9.71

Note: n.a. = algorithm for out-sample's prediction of the dependent variable is not yet available for this paper.

A question arises at this stage as to the advantage of using models that address spatial autocorrelation in analysing house prices when they do not really out-perform the traditional OLS in terms of statistical quality. As shown in Table 5, spatial autocorrelation tests for the OLS model did not indicate any spatial error or spatial lag in the data. By contrast, the SAC model has indicated otherwise. By say giving more faith to the latter and, thus, believing in the existence of spatial error and/or spatial lag, we ask a question: so what?

The answer is that, if spatial error and/or spatial error could have really occurred, the OLS model could have been more biased and inefficient than SAC, SAR, and SEM models. In this case, it is useful to test the level of biasedness and efficiency in using models that address spatial autocorrelation against the traditional OLS model. Unfortunately, we do not have any *a priori* information on the true population mean of each regression coefficient, making tests for biasedness and efficiency not possible. Nevertheless, an indirect test can be performed. Based on this indirect test, if regression coefficients of the OLS model are not statistically different from those of the corresponding spatial-autocorrelation-based models, then we can make a general conclusion that the OLS model may not likely to be more biased or more inefficient than its rival models.

The t-test for difference in regression coefficients between the OLS model and other models in Table 4 could be performed, as an indirect test, to ascertain whether SAC, SAR, and SEM models are more unbiased and more efficient than the OLS model. A simple test is given as:

$$t_c = \frac{\beta_{OLS} - \beta_{AUTO}}{0.5(\sigma_{OLS} + \sigma_{AUTO})}$$

where t_c represents calculated t-value for differential value of regression coefficients, β_{OLS} represents coefficient of independent variables in the OLS model, β_{AUTO} represents coefficient of independent variable in the spatial hedonic model, σ_{OLS} represents standard error of independent variable in the OLS model and σ_{AUTO} represents the standard error of independent variable in the spatial hedonic model.

Results of the test (Table 7) indicate that t-values for all the differential regression coefficients were insignificant and, thus, indicating that the OLS model was not likely to be more biased or inefficient than its rival models.

Table 7: Results of the t test

Variables	t_c
Constant	11.6047*
LG_LA	0.0833
LG_GFA	0.0095
LG_AA	0.0006
TERRACE_1	-0.0028
LG_AGE	-0.0271
GCOND	0.0086
FFINISH_2	-0.0008
BED_3	-0.0002
LG_CBD	0.0056

Note: * significant at 1% level

5.0 CONCLUSION

This paper has examined the existence of spatial autocorrelation in the house price data from Malaysia. Four types of test for spatial autocorrelation have been used namely the Moran's I, The Lagrange Multiplier, L-ratio, and Wald. All tests have indicated insignificant spatial autocorrelation based on the OLS model as well as other models. By contrast, the lambda dan rho in the SAC model has indicated otherwise. By giving more faith to the latter, it was assumed that both types of spatial autocorrelation namely spatial error and spatial lag exist in the data.

In spite of addressing spatial autocorrelation, the models concerned have not shown superior statistical quality over the traditional OLS model. A further indirect test for model's biasedness and inefficiency found that there was no evidence that the OLS model was more biased or more inefficient compared to its rival spatial-autocorrelation-based models.

Despite the findings, it is important to stress that taking into account the possible phenomenon of spatial autocorrelation is critical in using hedonic housing price model to: (1) ascertain that the model chosen is free from spatial autocorrelation problems; (2) if the problems do exist, we can adopt the appropriate steps to improve the statistical quality of price-estimating model being chosen.

REFERENCES

HAMID, ABDUL, BIN HJ. MAR IMAN, 2006. Price-Contour-Based Spatial Dummy Variables for Segmenting Market in the Geographic Information System Assisted Hedonic Modelling of Residential Property Price, *Malaysia Journal of Real Estate*, **1**(1), pp. 39-57

HAMID, ABDUL, BIN HJ. MAR IMAN, 2007. Combining Geographic Information System and Regression Model in Generating Locational Value Residual Surfaces in the Assessment of Residential Property Values, *Pacific Rim Property Research Journal*, **13**(1), pp. 35-62

HAMID, ABDUL, BIN HJ. MAR IMAN, 2001. *Incorporating A Geographic Information System in Hedonic Modelling of Farm Property Values*. PhD Thesis. Lincoln University. (unpublished)

ADAIR, A.S., BERRY, J.N. and MCGREAL, W.S., 1996. Hedonic Modelling, Housing Submarkets and Residential Valuation. *Journal of Property research,* **13**, pp. 67-83.

ANSELIN, L., 1998. GIS Research Infrastructure for Spatial analysis of Real Estate Markets. *Journal of Housing Research,* **9**(1), pp. 113-133.

AZHARI, HUSIN and MOHAMAD GHAZALI, HASHIM, 1994. *The Construction of Land Value Maps Using GIS & MRA: A Case Study of Residential Properties in Johor Bahru, Malaysia*, Research Paper Universiti Teknologi Malaysia..

AZHARI, HUSIN, 1990. *The Construction Of Regression-Based Mass Appraisal Models: A Methodological Discussion And An Application To Housing Submarkets In Malaysia*. PhD Thesis. University of Reading. (unpublished)

BAILEY, T.C. and GATRELL, A.C., 1995. *Interactive Spatial Data Analysis*. England: Longman Group Limited.

BOURASSA, S.C., HAMELINK, F., HOESLI, M. and MACGREGOR, B.D., 1999. Defining Housing Submarkets. *Journal of Housing Economics,* **8**, pp. 160-183.

BOWEN, W.M., MIKELBANK, A. and PRESTEGAARD, D.M., 2001. Theoretical and Empirical Considerations Regarding Space in Hedonic Housing Price Model Applications. *Growth and Change,* **32**(4), pp. 466-490.

CASS, T., 1980. *Statistical Methods in Management 2*. London: Cassell Ltd.

CHAU K. W AND CHIN T. L., 2002. A Critical Reveiw of the Literature on the Hedonic Pricing Model and Its aplication to the Housing Market in Penang, *The Seventh Asian Real Estate Society Conference. Seoul, Korea*.

CHEN, L., 1994. *Spatial Analysis of Housing Market Using Hedonic Based on GIS*. PhD Thesis. Massachusets Institute of Technology.

CHESHIRE, P. and SHEPPARD, S., 1998. Estimating the Demand for Housing, Land, and Neighbourhood Characteristics. *Oxford Bulletin of Economics and Statistics,* **60**(3), pp. 357-382.

CHIN , T. L., CHAU, K. W. and NG, F. F., 2004. The Impact of the Asian Financial Crisis on the Pricing of Condominiums in Malaysia, *Journal of Real Estate Literature*, **12**(1), pp 33-49.

COURT, A.T., 1939. Hedonic Price Indexes with Automobile Examples. *The Dynamics of Automobile Demand,* pp. 97-117.

CRESSIE, N., 1989. Geostatistics. *The American Statistician,* **43**(4), pp. 197-202.

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), pp. 149-168.

DES ROSIERS, F., THERIAULTS, M. and VILLINEUVE, P., 2000. Sorting out Access and Neighbourhood Factors in Hedonic Price Modelling. *Journal of Property Investment & Finance,* **18**(3), pp. 291-315.

DUNSE, N. and JONES, C., 1998. A Hedonic Price Models of Office Rents. *Journal of Property Valuation & Investment,* **16**(3), pp. 297-312.

DUTILLEUL, P., CLIFFORD, P., RICHARDSON, S. and HEMON, D., 1993. Modifying the t Test for Assessing the Correlation Between Two Spatial Processes. *Biometrics,* **49**(1), pp. 305-314.

DZURLLKARNAIN D., TIRUSELVAM ARUMUGAM, IBRAHIM @ ATAN SIPAN, 1996, *Pembangunan Model Penilaian Kaedah Mass-Appraisal Menggunakan Teknik Multiple Regression Analysis Bagi Maksud Kadaran : Kes Kajian : Majlis Bandaraya Johor Bahru*, Research Paper UTM

FLETCHER, M., GALLIMORE, P. and MANGAN, J., 2000. The Modelling of Housing Submarkets. *Journal of Property Investment & Finance,* **18**(4), pp. 473-487.

GELFAND, A.E., ECKER, M.D., KNIGHT, J.R. and SIRMANS, C.F., 2004. The Dynamics of Location in Home Price. *Journal of Real Estate Finance and Economics,* **29**(2), pp. 149-166.

GRILICHES, Z., 1961. Hedonic Price Indexes for Automobiles: An Econometric Analysis of Quality Change *in The Price Statistics of The Federal Government.* General Series No.73. New York: National Bureau of Economic Research.

GRILICHES, Z., ed, 1971. *Price Indexes and Quality Change.* Cambridge: Harvard University.

GUJARATI, D., 1995. *Basic Econometrics*. 3rd edn. New York: McGraw-Hill.

GUJARATI, D., 2003. *Basic Econometrics*. 4th Edition edn. Boston: McGraw-Hill Higher Education.

HAAS, G.C., 1922. *Sale Prices as a Basis for Farm Land Appraisal.* Technical Buletin 9 (November). The University of Minnesota Agricultural Experiment Station: Division of Agricultural Economics.

HARRISON, D.J. and RUBINFELD, D.L., 1978. Hedonic Housing Prices and the Demand for Clean Air. *Journal of Environmental Economics and Management,* **5**(1), pp. 81-102.

HENNEBERRY, J., 1998. Transport Investment and House Prices. *Journal of Property Valuation & Investment,* **16**(2), pp. 144-158.

LANCASTER, K.J., 1966. A New Approach to Consumer Theory. *Journal of Political Economy,* (April), pp. 132-153.

MACLENNAN, D., 1977. Some Thoughts on the Nature and Purpose of House Price Studies. *Urban studies,* **14**, pp. 59-71.

MALPEZZI, S., 2003. Hedonic Pricing Models: a Selective and Applied Review. In: T. O'SULLIVAN and K. GIBB, eds, *Housing Economics & Public Policy.* Oxford: Blackwell Publishing, .

MARTINEZ, W.L. and MARTINEZ, A.R., 2002. Computational Statistics Handbook with MATLAB. Boca Raton, Florida: Chapman & Hall/CRC.

MCCLUSKEY, W.J., DEDDIS, W.G., MANNIS, A., LAMONT, I.G. and BORST, R.A., 2000. The Application of Surface Generated Interpolation Models for the Prediction of Residential Property Values. *Journal of Property Investment and Finance,* **2**, pp. 162-176.

MIRON, J., 1984. Spatial Autocorrelation in Regression Analysis: A beginner's Guide. In: G.L. GAILE and C.J. WILLMOTT, eds, *Spatial Statistics and Models.* Dordrecht: D. Reidel Publishing Company, pp. 201-222.

ORFORD, S., 1999. Valuing The Built Environment: GIS and House Price Analysis. Aldershot: Ashgate Publishing Ltd.

OVERMARS, K.P., DE KONING, G.H.J. and VELDKAMP, A., 2003. Spatial Autocorrelation in Multi Scale Land Use Models. *Ecological Modelling,* **164**, pp. 257-270.

PACE, R.K, (2008). Email communication on 3rd April 2008

PACE, R.K. and BARRY, R., 1997a. Quick Computation of Spatial Autoregressive Estimators. *Geographical Analysis,* **29**, pp. 232-246.

PACE, R.K. and BARRY, R., 1997b. Sparse Spatial Autoregressions. *Statistics & Probability Letters,* **33**, pp. 291-297.

PACE, R.K. and LESAGE, J.P., 2003. Likelihood Dominance Spatial Inference. *Geographical Analysis,* **35**(2), pp. 133-147.

PACE, R.K., BARRY, R. and SIRMANS, C.F., 1998. Spatial Statistics and Real Estate. *Journal of Real Estate Finance and Economics,* **17**(1), pp. 5-13.

PAEZ, A., UCHIDA, T. and MIYAMOTO, K., 2001. Spatial Association and Heterogeneity Issues in Land Price Model. *Urban Studies,* **38**(9), pp. 1493-1508.

PAGOURTZI, E., ASSIMAKOPOULOS, V., HATZICHRISTOS, T. and FRENCH, N., 2003. Real Estate Appraisal: A Review of Valuation Methods. *Journal of Property Investment and Finance,* **21**(4), pp. 383-401.

SURIATINI, ISMAIL, 2005. *Hedonic Modelling of Housing Markets Using Geographical Information System (GIS) and Spatial Statistics: A Case Study of Glasgow, Scotland*, PhD Thesis, University of Aberdeen. (Unpublished)

SURIATINI, ISMAIL, 2006. Spatial Autocorrelation and Real Estate Studies: A Literature Review, *Malaysian Journal of Real Estate*, Fakulti Kejuruteraan & Sains Geoinformasi. **1**(1), pp.1-13

THERIAULT, M., DES ROSIERS, F., VILLINEUVE, P. and KESTENS, Y., 2003. Landscaping Attributes and House Prices: Looking at Interactions between Property Features and Homeowners' Profiles. *Property Management,* **21**(1), pp. 25-62.

WATKINS, C., 2001. The Definition and Identification of Housing Submarkets. *Environment and Planning A,* **33**, pp. 2235-2253.

WILHELMSSON, M., 2000. The Impact of Traffic Noise on the Values of Single-family Houses. *Journal of Environment Planning and Management,* **43**(6), pp. 799-815.

WILHELMSSON, M., 2002. Spatial Model in Real Estate Economics. *Housing Theory and Society,* **19**, pp. 92-101.

WILTSHAW, D.G., 1996. Spatial Autocorrelation and Valuation Accuracy: a Neglected Methodological Problem. *Journal of Property Research,* **13**, pp. 275-286.

www.spatial-econometrics.com (March 2008)

AMINAH, YUSOF, 2006. Modeling House Price in Malaysia: An Analysis of Hedonic Function of House-Specific Characteristics, *Proceedings of the International Real Estate Research Symposium (IRERS) 2008*, INSPEN, Malaysia