SPATIAL HEDONIC MODELLING (SHM) FOR MASS VALUATION

[2]Suriatini Ismail, [1]Taher Buyong, [2]Ibrahim Sipan,
[2]Mohamad Ghazali Hashim and [2]Renuga Navaneethan

[1]Institute of Advanced Technology
Universiti Putra Malaysia
43400 UPM Serdang, Malaysia
Email:
rehat_55@yahoo.com

[2]Faculty of Geoinformation Science and Engineering
Universiti Teknologi Malaysia
81310 UTM Skudai, Malaysia
Email:
suriatini@fksg.utm.my, ibrahim@fksg.utm.my & renu_utm2004@yahoo.com

ABSTRACT

The literature has criticized that conventional hedonic model is not capable of taking into account of spatial effects on house prices even with the inclusion of locational variables. In response to this, several advances of multiple regression methods that incorporate spatial dependence have recently emerged. One of the methods is known as spatial hedonic modelling (SHM). There are three common spatial hedonic models namely General Spatial Model (SAC), Spatial Autoregressive Model (SAR) and Spatial Error Model (SEM). The literature has noted that these models are superior compared to the conventional model. This paper presents the preliminary results of a comparison between the performance of the conventional model and the best spatial hedonic model in the prediction of residential property values in Malaysia. The sample comprises 196 residential properties located within the jurisdiction of Majlis Perbandaran Kulai, Johor, Malaysia. Seven independent variables are used to construct both the conventional and spatial hedonic models. The outcomes of this study support the literature in the context of Malaysia.

Keywords: *Spatial dependence, Ordinary Least Square (OLS), spatial regression, mass appraisal*

1.0 INTRODUCTION

The common procedure of mass appraisal of housing values for the purpose of property taxes uses manual methods to estimate the value of a single property. Therefore, the procedure is not only inefficient and costly but most importantly, its ability to provide fair and equitable property taxes are always questioned.

Attempts to develop a better procedure for mass appraisal of property values in Malaysia are numerous and these are pioneered by Azhari Husin (1990). For example, Computer Aided Mass Appraisal (CAMA) systems for property taxes have been developed for some local authorities such as Majlis Perbandaran Kuantan (MPK) (Dzurlkanian et al., 2006). However, most of these procedures are based on conventional hedonic model which utilizes Ordinary Least Square (OLS) method.

The literature has criticized that conventional hedonic model is not capable of taking account for spatial dependence on house prices even with the inclusion of locational variables. In parallel to that, spatial hedonic modelling has been suggested as a more suitable technique for house price analysis if spatial dependence is present in the data. Hence, the primary aim of this paper is to compare the relative performance of the conventional model and spatial model in prediction of residential property values in Malaysia.

The next section of this paper discusses briefly spatial hedonic modelling. Data and analysis procedure are discussed in the third section of this paper followed by results and discussion in the fourth section. The final section concludes the paper.

2.0 SPATIAL HEDONIC MODELLING (SHM)

The hedonic price approach is a method of estimating the implicit prices of the characteristics of a differentiated product (Kim, 2004). The price of a differentiated product can be expressed as a function of the quantities of its various characteristics of physical and locational aspects. This relationship is called the hedonic price equation. These techniques rely on observable market transactions to obtain values for various characteristics of heterogeneous products (Boxall et al., 2005).

Spatial hedonic modelling is a method that incorporates spatial dependence into a regression model (Boxall et al., 2005 and Kim, 2004). Spatial autocorrelation, also expressed as spatial dependence refers to the possible occurrence of the interdependence among the observations viewed in the geographical space, which violates the assumptions made in the conventional hedonic model estimation (Hua et al., 2005). According to Suriatini (2006), there are two types of spatial autocorrelation namely spatial error dependence and spatial lag dependence.

Suriatini (2006) states that spatial error dependence refers to the correlated errors that occur among the independent variables. It can rise from omitted variables, variable measurement error or misspecification of the functional form. Meanwhile, spatial lag dependence refers to the correlated errors that occur between the dependent variables.

There are three common spatial hedonic models known as General Spatial Model (SAC), Spatial Autoregression Model (SAR) and Spatial Error Model (SEM) (De Silva et al., forthcoming and LeSage, 1999a). These are presented as follows.

2.1 GENERAL SPATIAL MODEL

General Spatial Model is used to deal with both types of spatial dependence, namely spatial lag dependence and spatial error dependence. The model takes the following form (De Silva et al., forthcoming):

$$Y = \rho W_1 Y + X\beta + \mu$$
$$\mu = \lambda W_2 \mu + \varepsilon$$

where $W_1 Y$ is the spatial lag term, $W_2\mu$ is the spatial error term, W_1 and W_2 are $n \times n$ spatial weight matrices to identify the geographical relationship among observations using geo-coordinate information. Y is an $n \times 1$ vector of dependent variables; and X is an $n \times k$ matrix of explanatory variables. Both μ and ε are vectors of error terms. μ is the vector of spatially correlated error terms; and ε is the vector of uncorrelated error terms.

2.2 SPATIAL ERROR MODEL (SEM)

Spatial Error Model is used to handle the spatial dependence due to the omitted variables or errors in measurement through the error term. SEM takes the following form (De Silva et al., forthcoming):

$$Y = X\beta + \mu$$
$$\mu = \lambda W_2 \mu + \varepsilon$$

where $W_2\mu$ is the spatial error term and W_2 is $n \times n$ spatial weight matrix to identify the geographical relationship among observations using geo-coordinate information. Y is $n \times 1$ vector of dependent variables and X is an $n \times k$ matrix of explanatory variables. Both μ and ε are $n \times 1$ vectors of error terms. μ is the vector of spatially correlated error terms and ε is the vector of uncorrelated error terms.

2.3 SPATIAL AUTOREGRESSIVE MODEL (SAR)

Spatial Autoregressive Model (SAR) is also known as Spatial Lag Model. Under this model, the housing price is partially explained by the characteristics of the "neighbours" apart from the numbers of its associated attributes (Eddie Hui et al. 2007). SAR takes the following form (De Silva et al., forthcoming):

$$Y = \rho W_1 Y + X\beta + \varepsilon$$

where $W_1 Y$ is the spatially lagged term and W_1 is $n \times n$ spatial weight matrix to identify the geographical relationship among observations using geo-coordinate information. Y is $n \times 1$ vector of dependent variables and X is an $n \times k$ matrix of explanatory variables. ε is the vector of uncorrelated error terms.

3.0 DATA AND ANALYSIS FRAMEWORK

The data comprises 196 single-storey terrace houses transacted between 2002 and 2006 within Majlis Perbandaran Kulai (MPKu) area. Table 1 summarises the variables, its description and expected signs.

Table 1: List of variables, its descriptions and expected signs

A) DEPENDENT VARIABLE			
	VARIABLE	**DESCRIPTION**	**EXPECTED SIGNS**
1.	PRICE	Price (RM/unit)	-

B) INDEPENDENT VARIABLES			
	VARIABLE	**DESCRIPTION**	**EXPECTED SIGNS**
1.	LA	Land area (sq.m)	+ve
2.	MFA	Main floor area (sq.m)	+ve
3..	AFA	Ancillary floor area (sq.m)	+ve
4..	ADD_FA	Additional floor area (sq.m)	+ve
5.	POSITION	Type of lot (1 = Intermediate, 2 = End and 3 = Corner)	+ve
6.	FLOOR	Floor finishes (1 = Cement, 2 = Ceramic tiles, 3 = combination of ceramic tiles and cement)	+ve
7.	TITLE	Types of title (1 = Others, 2 = HSM (Freehold), 3 = HSD (Freehold) and 4 = Grant)	+ve

Originally, there were about 1,300 transacted houses as obtained from the Jabatan Perkhidmatan Harta (JPPH), in Johor Bahru. After the first round of data screening, about 500 of the data were discarded primarily due to non-existence of lot numbers or grant numbers (i.e. the identifier to link the attribute components of the houses to their spatial component in order to utilize Geographical Information System (GIS) functionalities).

The second round of data screening focused on identifying outliers. This stage looked for abnormalities in the transacted prices (such as a corner lot house being transacted at a much lower price, in the same year, than its neighbouring same row intermediate houses and transacted price is too high/low in comparison to its neighbouring houses with the same size of land and floor area), land area (such as an intermediate house having land area bigger than its corner lot neighbour of the same row), and floor area (such as a house with main floor area greater than its land area). Finally, only 196 observations were available to be analysed.

One hundred and ninety-six observations were then used in the estimation of the coefficients of the independent variables using conventional OLS method and Spatial Hedonic Modelling (SHM). Firstly, the Ordinary Least Square (OLS) regression was performed to obtain the OLS residuals and coefficients for each variable. The determined coefficients were then used in the estimation of the prices of the in-sample houses. Then, 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. These tests are available in Spatial Econometrics Toolbox (SET) for Matlab (www.spatial-econometrics.com).

Next, three types of spatial hedonic regressions, namely Spatial Error Model (SEM), Spatial Autoregressive Model (SAR) and General Spatial Model (SAC) were performed to estimate the coefficients of the independent variables. Once the spatial models were estimated, the LM-Sar test was performed to detect spatial autocorrelation in the SAR model. The LM-Sar test result is one of the criteria for spatial model selection. Then, best spatial model was selected.

According to Rutt (2007), there are no decisive criteria explicitly outlining the steps to follow in selecting a spatial model but there are guidelines that can be used to justify selecting one model over another. Thus, the model selection procedures in this study have followed the general suggestions provide in LeSage (1999a) based upon the testing routines included in his manual and resulting signs and significance of the spatial coefficients.

The first stage of model selection, as outlined in LeSage (1999a), is based on the statistical significance of the spatial autocorrelation coefficients. He suggests that insignificant coefficient for ρ in the SAC model would indicate that SEM model is preferred, while the same circumstances for λ suggest that SAR model is better. The seecond stage of model selection is based on the LM-Sar test result. This test examines spatial autocorrelation in the residuals of SAR model. LeSage (1999a) suggests that the SAC model might be the most appropriate model to handle the data if the LM-Sar test result indicates spatial autocorrelation in the residuals of the SAR model. The final stage of model selection is based on the value of log-likelihood while considering its sign. The higher the value of log-likelihood, the better is the model. After selecting the best spatial hedonic model, the determined coefficients of the independent variables in the model were used in the estimation of the prices of the in-sample houses.

Finally, the performance of the OLS models and the best spatial hedonic model are compared based on the adjusted R^2, Standard Error of Estimate (SEE), and sign and magnitude of coefficients. In addition, the estimated prices of OLS and the spatial models were compared with the actual transacted prices to measure the predictive capability of both models. In this case, Coefficient of Dispersion (COD) is computed, whereby the lower the value, the better is the model.

4.0 RESULTS AND DISCUSSIONS

Descriptive statistics of the sample data are presented in Table 2. The single-storey terrace houses in the study area are priced between RM105,000 and RM193,000. The descriptive result indicates that the range of land area is between 111.48 square meters and 532.28 square meters. The mean main floor area, ancillary floor area and additional floor area are about 84 square meters, 15.44 square meters and 6.41 square meters, respectively. The majority of the properties is intermediate lot with tile floor finishing and has been issued with HSM (Kekal) title (i.e. temporary freehold ownership).

Table 2: Descriptive Statistics

	Minimum	Maximum	Mean	Std. Deviation
PRICE	105000.00	193000.00	132765.3061	15459.07817
LA	111.48	532.28	159.3797	55.39745
MFA	62.59	118.54	83.9883	8.95704
AFA	.00	69.49	15.4404	10.07598
ADD_FA	.00	54.76	6.4065	12.58834
POSITION	1	3	1.15	.503
FLOOR	1	3	2.02	.781
TITLE	1	4	2.36	.856

Table 3 summarises the test results of Moran's I, LM-Error and LM-Sar tests. Moran's I and LM-Error tests results indicate that spatial autocorrelation is present in the OLS residuals. Therefore, spatial model is necessary to capture the spatial dependence. Meanwhile, the LM-Sar test result indicates that the inclusion of spatially lagged dependent variables failed to eliminate spatial dependence in the SAR residuals.

Table 3: Spatial autocorrelation test results

	Moran's I	**LM-Error**	**LM-Sar**
Value	0.183	18.293	67.427
Statistic	5.0726	-	-
Marginal Probability	0.00000039	0.00001895	0.00000000

Table 4 summarises the coefficient estimations obtained for the OLS, SEM, SAR and SAC models. The OLS, SEM, SAR and SAC models explain about 61.1%, 62.2%, 61.1% and 66.1% variation in housing prices respectively. The OLS and SAR model explain the property prices at the same level.

Regression results presented in Table 4 show that the spatial autocorrelation parameters in all three spatial models are significant. The SAC model produced estimates of ρ and λ that were both significant at the 90% level. This means that, the SAC model might be the most appropriate model to deal with the spatial autocorrelation. Besides that, the SAC model also has greater log-likelihood values compared to the SEM and SAR models. In addition, the LM-Sar test result indicates spatial autocorrelation in the SAR residuals. In couple with the results from LM-Sar test, log-likelihood and the significance of both spatial autocorrelation parameters in the SAC model, SAC model is shown to be more appropriate for handling the data. Thus, SAC model is selected as the best spatial hedonic model and accordingly the results of SAC and OLS models discussed in details in the following paragraph.

The OLS model shows that 61.1% of variation in property prices is explained by the independent variables. All variables have expected signs except for TITLE in the OLS model. The type of title probably represents the property age which has not been included in the modelling. Properties issued with "grant" might be older than the properties issued with "HSM (Freeholdl)", "HSD (Freehold)" and "other titles". Based on literature review, property age normally has negative effect on property prices. Thus, this might cause the type of title to have a negative sign. Besides that, the result indicates that AFA is not significant in explaining the property prices in the study area. As indicated in Table 4, LA is a significant determinant of residential property prices. A unit increase in land area will cause about RM78 increase in property price. MFA, ADD_FA POSITION, FLOOR and TITLE are also found to be good explanatory variables to explain the variation in property price. The unstandardized coefficient indicates that a unit increase in main floor area and additional floor area will cause about RM168 and RM548 increase in property price, respectively.

Table 4: Comparison of the four competing models

Model	**OLS**	**SEM**	**SAR**	**SAC**
Model form	Linear	Linear	Linear	Linear
R^2	0.6252	0.6745	0.6251	0.6732
Adjusted R^2	0.6112	0.6624	0.6111	0.6610
SEE	9639.1442	8797.4916	8918.0157	8815.2252
Log-Likelihood	-	-1995.097	-1995.0375	-1881.0203
Dependent variable (PRICE)	House price (PRICE)	House price (PRICE)	House price (PRICE)	House price (PRICE)
Independent variables:	**Coefficient (t-statistic)**	**Coefficient (t-statistic)**	**Coefficient (t-statistic)**	**Coefficient (t-statistic)**
(Constant)	92157.05	82987.69	53985.89	59798.51
	(11.54)***	(8.68)***	(5.23)***	(3.86)***
Land Area (LA)	78.46	85.33	84.39	89.55
	(2.61)***	(3.13)***	(3.04)***	(3.24)***
Main Floor Area (MFA)	168.44	298.42	181.36	247.84
	(2.02)**	(3.07)***	(2.34)**	(2.76)***
Ancillary Floor Area	110.16	-44.24	-36.00	-59.48
(AFA)	(1.30)	(-0.46)	(-0.44)	(-0.64)
Additional Floor Area	548.16	496.78	484.09	495.11
(ADD_FA)	(8.39)***	(7.81)***	(7.79)***	(7.79)***
Position (POSITION)	5236.35	5492.46	5356.77	5307.09
	(1.64)*	(1.92)*	(1.81)*	(1.83)*
Floor Finishes (FLOOR)	3353.77	2583.12	1668.81	1861.69
	(3.56)***	(2.05)**	(1.80)*	(1.56)*
Title (TITLE)	-1719.21	-1090.10	-1088.03	-1038.03
	(-2.00)**	(-1.04)	(-1.35)	(-1.10)
Lambda (λ)		0.44		0.27
		(6.08)***		(1.86)*
Rho (ρ)			0.31	0.21
			(4.77)***	(1.91)*

***Significant at 1%; **Significant at 5%; *Significant at 10%

The SAC model explains about 66.1% of total variation in property prices. All variables in this model have expected signs except for AFA and TITLE. The possibility for TITLE to has unexpected sign has been explained in the previous paragraph. AFA is shown to have a negative effect on property prices. This is perhaps due to the reason that for the same size of land area, larger AFA means less MFA. As MFA becomes more significant in the SAC model, the effect of AFA might has been controlled by MFA. These variables are also found to be insignificant in explaining the housing prices. Other variables in this model significantly affect the housing prices. A unit increase in land area will cause about RM90 increase in property price. MFA and ADD_FA are also found to be good explanatory variables to explain the variation in property price. The unstandardized coefficient indicates that a unit increase in main floor area and additional floor area will cause about RM248 and RM495 increase in property prices, respectively.

A comparison between the OLS and SAC models reveals that three out of seven explanatory variables (LA, MFA and POSITION) become more significant in the spatial model while four independent variables (AFA, ADD_FA, FLOOR and TITLE) become less significant. However, the SAC model has greater adjusted R^2 and lower SEE. This suggests that, the SAC model is better compared to the OLS model.

An in-sample testing was carried out for both OLS and SAC models to measure the predictive capability of the models. The determined coefficients in the OLS and SAC models were used in the estimation of the prices of the in-sample houses. Coefficient of Dispersion (COD), is computed in order to measure the predictive capability of the models. According to McMillan and Weber (2007), The International Association of Assessing Officers (IAAO) calls the coefficient of dispersion "the most generally useful measure of variability". It measures the average percentage deviation of the assessment ratios from the median ratio. COD is defined as follows:

$$COD = \frac{100}{n} \sum_{i=1}^{n} \frac{|R_i - R_m|}{R_m}$$

where R_i is the assessment ratio for observation i, R_m is the median of ratio and n is number of observations.

Table 5 summarises the results of the prediction accuracy.

Table 5 : Predictive capability of OLS and SAC models

	OLS	**SAC**
Max deviation	50498	48020
Min deviation	143	31
COD	5.12	4.88

Based on Table 5, the SAC model slightly outperforms the OLS model in terms of predictive capability. The value of adjusted R^2 of the SAC and OLS models show that the SAC model outperforms the OLS model by about 5%[1]. The SAC model has greater adj R^2, and lesser SEE and COD. Lesser COD means the predicted property values are constant with the market value (actual price). Thus, the SAC model has better capability to predict property prices in the study area.

[1] Based on a review of 11 studies mentioned in Suriatini (2005), the past spatial hedonic models outperform the past OLS model by 2.4% to 23%.

5.0 CONCLUSION

The primary objective of this paper was to compare the performance of conventional model and spatial model in the prediction of residential property values in Malaysia. In doing so, the conventional OLS model and spatial models namely SEM, SAR and SAC model were estimated. Out of the three spatial models, the best model was selected for a comparison with the OLS model. The SAC model is found to be the most appropriate model to handle the data compared to SEM and SAR. Comparison between the OLS and SAC model revealed that some variables are more significant in the OLS model and some are more significant in the SAC model. However, the SAC model has greater adjusted R^2 and lower SEE values. This means that, the SAC is better compared to the OLS model. The in-sample testing confirmed that the SAC model outperforms the OLS model in predicting the house prices in the study area. This suggests that development of a mass appraisal model for the study area should consider spatial hedonic modelling.

Nonetheless, given that these are preliminary results of a bigger study, analysis could be repeated to include further investigation on the quality of data. Additionally, different functional forms could be tried out other than the use of different data codings. Apart from that, issues like multicollinearity, and heteroscedasticity, could also be examined.

ACKNOWLEDGEMENT

Financial support of this research under Agreement No: NAPREC (R&D 1/07) is gratefully acknowledged.

REFERENCES

Azhari Husin (1990), The construction of regression-based mass appraisal models : a methodological discussion, PhD Thesis, University of Reading. (Unpublished)

Boxall, P. C., Chan, W. H. and McMillan, M. L. (2005), The Impact of Oil and Natural Gas Facilities on Rural Residential Property Values: A Spatial Hedonic Analysis, *Resource and Energy Economics*, Vol. 27, pp. 248-269.

De Silva, D. G., Kruse, J. B. and Yongsheng, W. (Forthcoming), Spatial Dependencies in Wind Related Housing Damage, *Natural Hazards (Journal of the International Society for the Prevention and Mitigation of Natural Hazards)*

Daud, Dzurllkanian @ Zulkarnain, Eboy, O. V., Buang Alias, Mohd Safie and Mustafa Omar (2006), *The Prospect of Applying CAMA System for rating Valuation in Malaysia.* The First Symposium of the World's Best Practices in Mass Appraisal for Asia , August 25-26, Bangkok, Thailand (Asia Hotel Bangkok) . (Unpublished)

Eddie Hui, C. M., Chau, C. K., Pun, L. and Law, M. Y. (2007), Measuring the Neighboring and Environmental Effects on Residential Property Value: Using Spatial Weighting Matrix, *Building and Environment*, Vol. 42, pp. 2333-2343.

Hua, S., Yong, T. and Shi-Ming, Y. (2005), A Spatio-Temporal Autoregressive Model for Multi Unit Residential Market Analysis, *Journal of Real Estate Finance and Economics*, Vol. 31, No. 2, pp. 155-187

Kim, J. (2004), An Assessment of the Discommodity Effects of Swine Production on Rural Property Values, PhD Thesis, University of Illinois. (Unpublished)

LeSage, J.P. "*The Theory and Practice of Spatial Econometrics*." University of Toledo, February 1999a. at http://www.spatial-econometrics.com/. (Accessed on 12th February, 2008.)

McMillan, D. P. and Weber, R. N. (2007), Property Tax Assessment Inequities and Thin Markets at http://tigger.uic.edu/~mcmillen/papers/Property%20Tax%20Assessment%20Inequities%20and%20Thin%20Markets.pdf (accessed on 1st April, 2008)

Rutt, M.E. (2007), On the Moo-ve Testing for Spatial Agglomeration Economies in the U.S. Diary Industry, Master Thesis, Kansas State University. (Unpublished)

Suriatini Ismail (2006), Spatial Autocorrelation and Real Estate Studies: A Literature Review, *Malaysian Journal of Real Estate*, vol. 1, no. 1, pp. 1-13.

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