SPATIOTEMPORAL AUTOREGRESSIVE MODELLING OF RESIDENTIAL PROPERTY PRICES

Nurul Hana Adi Maimun, Suriatini Ismail, Abdul Hamid Mar Iman

Centre of Real Estate, Department of Property Management, Faculty of Geoinformation & Real Estate, Universiti Teknologi Malaysia, 81310 Skudai, Johor Bahru, Johor, Malaysia.

Abstract

The aim of this paper was to evaluate the capability of the spatiotemporal autoregressive (STAR) model in constructing house price index using Malaysian data. An improved method for constructing the Malaysian House Price Index (MHPI) was proposed by incorporating spatial and temporal information into model specification. STAR was specified based on the theoretical framework and compared with the OLS based on logical, statistical, diagnostic, and predictive performance. Comparisons were also made between the OLS and STAR indices at the aggregate and disaggregate levels. The findings suggest that the STAR model performed better than the OLS model.

Keywords: *Spatiotemporal autoregressive model; ordinary least squares; hedonic; spatial statistics; house price index; Malaysian housing market*

Jel Classification: *C13, C21, C23, C43, C51, R31*

1. INTRODUCTION

The construction of real estate price indices has always been hedonic model based (Atkinson and Crocker, 1992; Gatzlaff and Haurin, 1997; Gatzlaff and Ling, 1994; Goodman, 1978; Meese and Wallace, 1997). These indices monitor price changes in a reliable manner (Hoesli and MacGregor, 2000), improve the accuracy (Flaherty, 2004), and are stable and less prone to substantial revision in light of new information (Clapham *et al*., 2004). Nevertheless, most real estate price indices are derived from models that disregard spatial or spatiotemporal elements (Clapp and Giaccotto, 1998; Colwell *et al*., 1998; Domingo and Fulleros, 2005; Gatzlaff and Ling, 1994; Nappi-Choulet and Maury, 2009). Although property transaction data used in constructing real estate indices normally have spatial and temporal characteristics, they have not been effectively specified in the traditional hedonic models. This is especially true in the Malaysian context (see National Institute of Valuation, 1996a; 1996b; National Property Information Centre, 2009). Although distance from the nearest town centre and time dummy variables is included in the Malaysian House Price Index (MHPI) model, it is still considered inadequate. In general, distance variables included in a model do not fully constitute a spatial model, but rather an aspatial model (Fotheringham and Rogerson, 1993; Valente *et al.*, 2005). This is because these variables only represent a general trend. Further, the space and time dummies incorporated in a model are ineffective in capturing the local variations (Pace *et al*., 1998b) and generate spatial autocorrelation and biased estimates (Nappi-Choulet *et al.*, 2007) which eventually affect model's degree of freedom and strength (Valente *et al.*, 2005). Among the critical statistical implications of employing indicator variables are specification bias (De Silva *et al.*, 2008; Dubin, 2003; Herath and Maier, 2010; Pace *et al.* 1998a), parameter inefficiency (Chica-Olmo, 1995; De Silva *et al.*, 2008; Dubin, 2003; Herath and Maier, 2010; Pace *et al.* 1998a), predictive inaccuracy (Dubin, 2003; Pace *et al.* 1998a), and model misspecification (Cliff and Ord, 1969; 1973; 1981).

The drawbacks of the current modelling technique have motivated the application of the spatiotemporal autoregressive (STAR) model in the real estate field. STAR model which considers simultaneous spatiotemporal effects has been introduced in real estate price analysis by Pace *et al.* (1998b; 2000) and has shown a good potential in modelling house prices. Among the key strengths of this model are increased goodness of fit and lower median absolute predition errors (Pace *et al*., 1998b; 2000; Sun *et al.*, 2005). Besides showing a strong statistical performance, it avoids the inclusion of a huge number of indicators (Liu, 2012; Pace *et al*., 1998b, 2000; Sun *et al.*, 2005), avoids data pooling (Basu and Thibodeau, 1998; Daria, 2007; Daria *et al.*, 2008), and employs a flexible weight matrix concept (Clapp, 2004).

A preliminary analysis of the Malaysian residential property prices are conducted by plotting the average prices over a 10-year period for some selected neighbourhoods in Johor Bahru, Malaysia. Figure 1 shows that Johor Bahru's property prices vary greatly among neighbourhoods. This indicates the need to consider simultaneous spatial and temporal elements in estimating property prices.

This paper aims to evaluate the capability of the STAR model for constructing house price index using Malaysian data. Since it is able to incorporate simultaneous spatial and temporal effects into model specification, this paper hypothesizes that it will outperform the traditional hedonic model in predicting property prices.

This paper is structured as follows. Section 2 reviews the theoretical framework of the STAR model introduced by Pace *et al.* (1998b; 2000). Section 3 describes the data and process involved in specifying the model and house price indices. Section 4 presents the results and discusses the outcome of the modelling. Section 5 compares the house price indices constructed using the OLS and STAR models. The final section closes this paper with a summary on the findings and suggestions for future study.

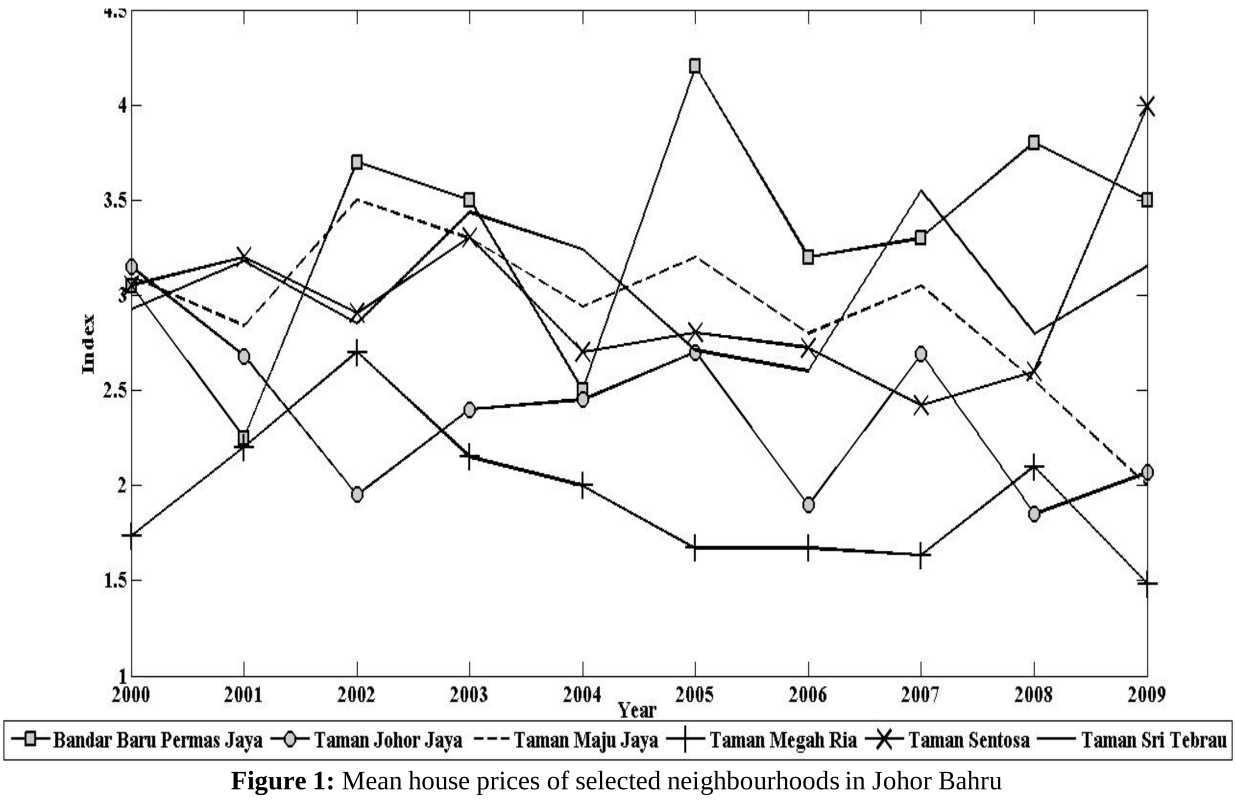

Figure 1: Mean house prices of selected neighbourhoods in Johor Bahru

2. THE THEORETICAL FRAMEWORK

Following Pace *et al.* (1998b; 2000), the general structure of a STAR model is expressed as follows:

(1)

$$Y = Z\emptyset + X\beta_1 + TX\beta_2 + SX\beta_3 + STX\beta_4 + TSX\beta_5 + \emptyset_S SY + \emptyset_T TY + \emptyset_{STY} STY + \emptyset_{TSY} TSY + \mu$$

where:

Y=*n* x *1* vector of house prices;
Z=*n* x *k* matrix of observations associated with spatial, temporal or spatiotemporal lags
$\emptyset$=Autoregressive parameters;
X=*n x k* matrix of house attributes and *n* by *p2* matrix of house attributes associated with spatial, temporal, or spatiotemporal lags;
β =Regression parameters in which $\beta 1, \beta 2, \beta 3, \beta 4, \beta 5$ are *k* by *1* vectors of parameters associated with spatial, temporal, or spatiotemporal lagged variables;
μ=Vector of white noise error term;
T=Temporal weight matrix;
S=Spatial weight matrix;
$\emptyset_S$=Spatial dependence parameter;
$\emptyset_T$=Temporal dependence parameter;
$\emptyset_{ST}$ and $\emptyset_{TS}$ =Spatiotemporal compound dependence parameters.

Equation (1) shows that the STAR model includes lagged dependent and lagged independent variables as the explanatory variables in the model specification. The spatial, *S*, and temporal, *T*, weight matrices represent the spatial and temporal relationship between the previous observations, respectively. These weight matrices are weighted by autoregressive parameters and act as space and time lag operators. In particular, the *ST* and *TS* matrices allow for modelling spatiotemporal effects.

The spatial matrix is calculated as follows:

(2)

$$S_{ij} = \frac{\sum_{i=1}^{ms} \lambda^i S^i}{\sum_{i=1}^{ms} \lambda^i} \qquad if\ j < i$$

$$S_{ij} = 0 \qquad Otherwise$$

where S_{ij} is a weighted average of the closest spatial neighbours for the respective variable, m_S is the optimal number of spatial neighbours while λ^i is the weight. Distances for each observation are sorted, starting with the closest neighbour up to the furthest. S_i represents the i^th closest neighbour sold with *i=1*... m_S. This process creates very sparse individual matrices - $S_1, S_2, S_3, \ldots, S_{ms}$.

Further, the temporal matrix is specified as follows:

(3)

$$T_{ij} = 0 \qquad Otherwise$$

$$T_{ij} = \frac{1}{m_T} \qquad if\ (i - m_T) \leq j < i$$

where T_{ij} indicates a weighted average of the closest neighbours for the respective variable for each time period. Each observation is sorted by time starting with the oldest in the first rows to the most recent in the last row. The m_T is the optimal time lag used to capture the time effect.

Both *S* and *T* matrices are standardized to sum to one. The matrices are specified as a lower triangular and contain zeros on the first rows. This is because the spatial and temporal autocorrelation effects depend on the previous observations. After including the spatial, temporal, and spatiotemporal lag effects in the dependent and independent variables, the errors are assumed to be independently and normally distributed.

3. DATA DESCRIPTION, SOURCES, AND SAMPLE

The property transaction data together with property attributes were obtained from the Valuation and Property Services Department, Johor Bahru, Malaysia. The dataset comprised property and sales information of double-storey linked houses across various housing estates in the Johor Bahru district. The Johor Bahru City Council provided the spatial reference of each property. Although the dataset contained mass information on property characteristics, only those which were theorized to have affected property values were selected. The selected properties have the following criteria: (1) sales price from RM50,000 to RM410,000, (2) transacted within 2000-2009 time period, (3) land area of 70 to 490m^2, (4) main floor area from 50 to 230m^2, (5) ancillary floor area from 0 m^2 to 131m^2, (6) have two to six bedrooms, (7) have a complete information on age of building, type of ownership, type of title, housing scheme, and geo-referencing. In total, 602 observations satisfied these criteria and, therefore, were selected to form a sample for the study. Table 1 shows a summary of the final dataset and the units of measurement.

Transaction price was used as the dependent variable to reflect the real property market. Latitude and longitude coordinates of individual housing lots were assigned to reflect spatial variation in location for each housing unit. Dummy variables were used to indicate the period of transaction. Variables used to capture structural attributes were: square metres of land area, main floor area, ancillary floor area, number of bedrooms, age of building, type of ownership, type of title, and housing scheme. The size and number of bedrooms were essential in determining the value of a property as it could have indicated space requirements for each household.

Table 1: Descriptive statistics of house transaction data

Label	Variable (unit of measurement)	Min	Max	Mean	Std. Dev.	Expected Signs
Price	House price (RM/unit)	50000	410000	226892	53924	Uncertain
YR00	Year 2000 (Control Dummy)	0	1	0.08	0.27	
YR01	Year 2001 (Dummy: 1=yes)	0	1	0.08	0.27	Negative
YR02	Year 2002 (Dummy: 1=yes)	0	1	0.18	0.39	Positive
YR03	Year 2003 (Dummy: 1=yes)	0	1	0.10	0.30	Positive
YR04	Year 2004 (Dummy: 1=yes)	0	1	0.07	0.26	Positive
YR05	Year 2005 (Dummy: 1=yes)	0	1	0.04	0.20	Positive
YR06	Year 2006 (Dummy: 1=yes)	0	1	0.03	0.17	Negative
YR07	Year 2007 (Dummy: 1=yes)	0	1	0.07	0.25	Negative
YR08	Year 2008 (Dummy: 1=yes)	0	1	0.14	0.34	Negative
YR09	Year 2009 (Dummy: 1=yes)	0	1	0.21	0.41	Negative
LA	Land area Square meters:	72.00	487.54	168.33	61.70	Positive
MFA	Main floor area Square meters:	52.10	224.20	132.96	20.20	Positive
AFA	Ancillary floor area Square meters:	0.00	130.43	24.23	11.22	Positive
Bed	Number of bedrooms (no.)	2	6	3	0.52	Positive
Age	Age of building (years)	1	42	12	10.02	Negative
Indi	Indigenous ownership (Dummy: 1=yes)	0	1	0.91	0.29	Negative
Inter	International ownership(Control Dummy)	0	1	0.09	0.29	
Free	Freehold (Control Dummy)	0	1	0.98	0.15	
Lease	Leasehold Dummy: 1=yes)	0	1	0.02	0.15	Negative
Scheme	House scheme (Dummy)	1	23	9.62	7.04	Uncertain
Long	Longitude (Decimal degree)	103.69	103.88	103.77	0.03	Uncertain
Lat	Latitude (Decimal degree)	1.47	1.57	1.53	0.03	Uncertain

The age of building could have indicated the physical conditions and vintage effects of the property during transaction. Therefore, age was expected to affect house prices negatively. Since the study was conducted in Malaysia, additional attributes affecting house prices in Malaysia were included. These were type of ownership and type of title. Both attributes exist because of the government's policies and may affect house prices negatively or positively, depending on its status.

Using this final dataset, regressions were run in the linear, semi-log, and double-log functional forms. Based on the $\bar{R}^2$, the double-log functional form was best suited to the data. Next, the STAR model was estimated via Matlab [1] guided by the equation and assumptions provided in Section 2. The OLS model was estimated with a view to comparing it with the STAR model.

The next step involved the determination of nearest neighbour matrix that, according to Dubin *et al*. (1999), was suitable for real estate transactions. Following Sun *et al.* (2005), the optimal numbers of m_S and m_T were calculated in order to derive the best spatial and temporal neighbours. Simultaneous regressions were performed using different numbers of spatial and temporal neighbours, with different numbers of λ. The lowest residual value of each regression run indicated the optimal number of spatial or temporal neighbours. Table 2 tabulates the results.

Table 2: Optimal number of spatial and temporal neighbours derivation

Number of neighbours	λ	**SSE**	
		m_S	m_T
3	0.70	8.5825	8.2908
3	0.75	8.5534	8.2261
3	0.80	8.5297	8.1740
3	0.85	8.5109	8.1378
3	**0.90**	8.4965	**8.1199**
3	0.95	8.4860	8.1216
4	0.70	8.4855	8.3315
4	0.75	8.4526	8.2796
4	0.80	8.4268	8.2374
4	0.85	8.4075	8.2072
4	0.90	8.3940	8.1913
4	0.95	8.3857	8.1912
5	0.70	8.3798	8.4533
5	0.75	8.3290	8.4068
5	0.80	8.2864	8.3701
5	0.85	8.2527	8.3456
5	0.90	8.2284	8.3361
5	0.95	8.2140	8.3433
6	0.70	8.2908	8.5529
6	0.75	8.2261	8.5212
6	0.80	8.1740	8.4989
6	0.85	8.1378	8.4871
6	**0.90**	**8.1199**	8.4872
6	0.95	8.1216	8.5001

Note: Regressions performed using double-log specification. N=587. Bolded numbers indicate optimal spatial/ temporal numbers

Based on the regression runs, the optimal number of *ms* and *mt* to be used in developing the STAR was set to 6 and 3 respectively, with λ = 0.90.

In evaluating model's quality, the R^2, $\bar{R}^2$, Log-likelihood (LL), Sum of Squared Errors (SSE), Akaike's Information Criterion (AIC), and Schwarz's Information Criterion (SIC) of the competing models were examined. Moran's I was used to identify spatial autocorrelation residuals while Breusch-Godfrey was used to identify temporal autocorrelation residuals. Since spatial autocorrelation could have caused heteroskedasticity, and not the other way round (Theriault *et al.*, 2003 and Suriatini, 2005), this study attempted to deal with the heteroskedasticity problem by addressing the spatial autocorrelation effect only. This two-in-one method followed the study by Fletcher *et al.* (2000) and Suriatini (2005). In order to analyze model's predictive performance, the mean absolute percentage error (MAPE) was adopted. The dataset containing 602 observations was divided into 542 in-sample observations and 60 out-of-sample observations, respectively. This study followed Case *et al.* (2004), Fletcher *et al.* (2004), and Goodman and Thibodeau (2007) who set aside 10% of the total observations as an ex-sample while the rest as an in-sample. The prediction residuals for both samples for each model were calculated and analyzed.

[1] Matlab was used to run the Spatial Statistics Toolbox and Spatial Econometrics Toolbox routines, downloaded from www.spatial-statistics.com and www.spatial-econometrics.com. These routines were used to estimate the OLS and STAR models.

The coefficients resulted from the OLS and STAR models were used to construct house price indices. A standard house unit with an average size of land area of 168.33 m^2, main floor area of 132.96 m^2, ancillary floor area of 24.23 m^2, three bedrooms, and of 12 years of age was chosen. The house, transacted between 2000 and 2009 was predicted by the OLS and STAR models. The value of appreciation of property between the base period (year 2000) and the subject period was calculated and converted into indices.

First, the OLS and STAR indices generated using the hedonic setup were constructed and compared with MHPI. In addition, the mean price index (unadjusted for characteristics) for a particular year was also derived. This provided a useful reference line which illustrated the real market trend. The OLS and STAR based indices were calculated by averaging the prices every year in each neighbourhood. Second, four neighbourhoods having property transactions recorded from 2001 to 2009 were selected and examined to illustrate price variations between neighbourhoods. Again, the results generated from OLS and STAR models were used to construct the neighbourhood indices.

The summary of steps involved in constructing the house price models and price movements is shown in Figure 2.

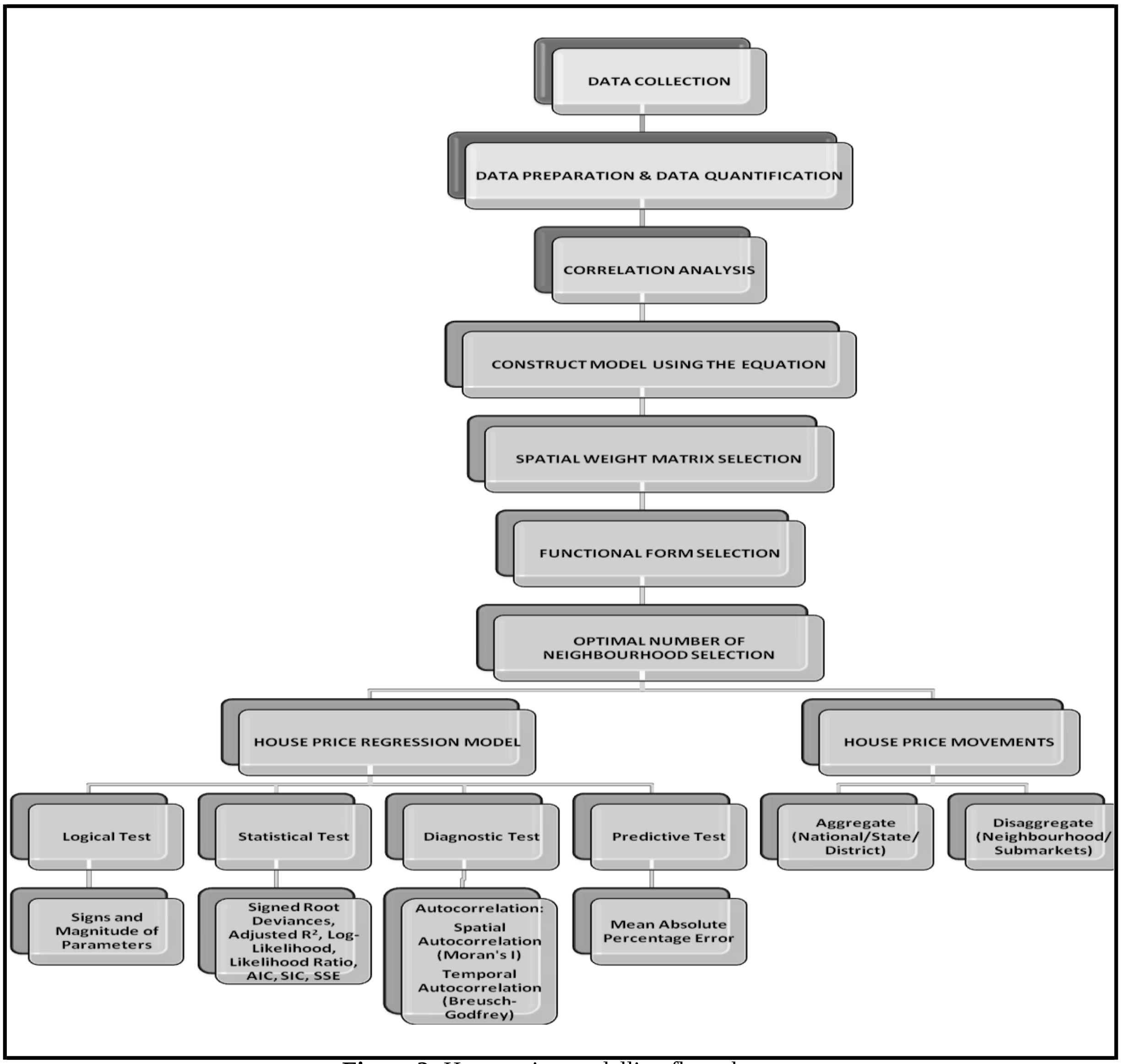

Figure 2: House price modelling flow chart

4. ANALYSIS AND DISCUSSIONS

4.1. The sign and magnitude of parameters

Table 3 tabulates the coefficients and performance of the OLS and STAR models using the double-log functional form.

Table 3: OLS and STAR regression results

Variables	OLS		STAR	
	β	**SRDS**	β	**SRDS**
Intercept	7.73	**26.93**	-5.16	-1.85
Index (1-n)/n			0.00	-0.77
Lat			-1.08	**-3.71**
Long			-0.26	**-3.85**
YR01	-0.04	-1.09		
YR02	0.01	0.48		
YR03	-0.01	-0.25		
YR04	-0.02	-0.53		
YR05	0.02	0.48		
YR06	-0.06	-1.32		
YR07	-0.10	**-2.52**		
YR08	-0.12	**-3.19**		
YR09	-0.18	**-4.93**		
LA	0.40	**16.16**		
MFA	0.47	**10.01**		
AFA	0.08	**6.19**		
Bed	-0.03	-0.60		
Age	0.03	**4.01**		
Indi	-0.14	**-6.08**		
Lease	-0.19	**-4.12**		
Scheme	0.00	**4.77**		
(X-TX) LA			0.36	**16.66**
(X-TX) MFA			0.51	**10.65**
(X-TX) AFA			0.05	**3.70**
(X-TX) Bed			0.01	0.31
(X-TX) Age			0.03	1.77
(X-TX) Indi			-0.16	**-8.20**
(X-TX) Lease			0.03	0.32
(X-TX) Scheme			0.01	**5.86**
S(X-TX) LA			-0.23	**-4.57**
S(X-TX) MFA			-0.43	**-4.40**
S(X-TX) AFA			0.06	**2.07**
S(X-TX) Bed			-0.10	-0.87
S(X-TX) Age			-0.05	-1.62
S(X-TX) Indi			0.10	1.66
S(X-TX) Lease			-0.06	-0.51
S(X-TX) Scheme			-0.02	**-5.48**
TX LA			0.39	**8.00**
TX MFA			0.26	**2.89**
TX AFA			0.02	0.70
TX Bed			0.10	1.18
TX Age			-0.01	-0.35
TX Indi			-0.18	**-4.97**
TX Lease			-0.22	-1.51
TX Scheme			0.01	**2.99**
STX LA			-0.02	-0.20
STX MFA			-0.42	**-2.97**
STX AFA			0.06	1.60
STX Bed			-0.07	-0.52
STX Age			-0.05	**-2.17**
STX Indi			0.08	1.06
STX Lease			-0.69	**-2.92**
STX Scheme			-0.01	**-4.35**
TSX LA			-0.09	-1.48
TSX MFA			0.00	0.00
TSX AFA			-0.02	-0.54
TSX Bed			0.25	**2.05**
TSX Age			0.02	0.80
TSX Indi			-0.10	-1.14
TSX Lease			0.19	1.00
TSX Scheme			0.00	1.31
S (Price)			0.68	**8.54**
T (Price)			-1.01	**-13.72**
ST (Price)			-0.22	-1.66
TS (Price)			0.13	1.32
n	587		587	
k	17		47	
ms			6	
mt			3	
λ			0.90	

Note: Cells shaded in grey indicate unexpected coefficient signs. Bolded numbers shown by SRDS statistics indicate significant variables affecting house prices. Blank cells show no information provided by the analysis.

In terms of geographical variables, the STAR model shows that house prices decreased as property location moved away from the city centre. This was in line with Alonso's (1964) classical theory of land rent. In terms of year of transaction, most year indicator results came out as expected except for 2003 and 2004 for the OLS model. In terms of other housing attributes, the OLS results show consistency with the expected results, except for the number of bedrooms and age of building. The OLS results show that additional numbers of bedrooms decreased house prices by about 3%. This contradicted the theory and previous works by Fletcher *et al.* (2000) and Li and Brown (1980), who discovered that each additional bedroom marked up the total selling price. The OLS result on building age indicates that older buildings have higher prices. This also contradicted the theory whereby older buildings have higher depreciation rates and, thus, lower prices. As for the STAR model, the result for **(X-TX)** indicates that most variables have shown the expected signs, except for age of building and leasehold status. Since the variables were integrated with time, there could be a good reason as to why the outcome did not match the prior expectations. A reduction of RM 36,571.561 recorded in house prices for bumiputra lots was logical as bumiputra lot prices in Johor Bahru were 15% lower than the international lot prices (around RM 230,800) for new developments. Almost all variables shown by **S(X-TX), STX** and **TSX** did not match prior expectations. This was not surprising because the magnitudes for **S**(Price), **ST**(Price), and **TS**(Price) were also not high. This indicates that the subject property's price did not depend on the spatial lag, spatial lag of the temporal lag, and/or temporal lag of the spatial lag of housing market prices. Instead, the STAR model shows that the Johor Bahru house prices were strongly influenced by time. Having a magnitude of -1.01 indicated by **T**(Price), house prices in Johor Bahru could have depended on three most recent neighbouring properties in time represented by **TX**. This was in line with the theory of property valuation, whereby the subject property's value depends on the most recent transacted property in time (Isakson, 1986). The spatial dependence of house prices, indicated by **S**(Price), shows that 68% of temporal difference in house prices was affected by six nearest neighbouring prices (each weighted by a geometrically declining factor of 0.9 to the subject property). Although the magnitude for **S**(Price) was lower than **T**(Price), this was still considered a fair evidence of the spatial dependency of house prices among neighbours. The higher magnitude of the coefficient estimate for **ST**(Price) compared with **TS**(Price), -0.22 and 0.13 respectively, implies the need to filter in time first, before space, and not vice versa, for this set of data. This result was in line with Pace *et al.* (1998b; 2000). On the other hand, most variables for the OLS and STAR models were significant in having two or more values in SRDS. The land area, main floor area, ancillary floor area, indigenous lot and scheme variables have significant and consistent SRDS signs in the OLS and STAR models. This shows that these variables were very important as they could have influenced property value.

Table 4: OLS and STAR statistical performance

Statistical Tests	OLS	STAR	Magnitude of Increase/ Decrease from OLS to STAR (%)
R^2	0.66	**0.79**	19.70
$\bar{R}^2$	0.65	**0.77**	18.46
SSE	12.51	**8.12**	-35.09
LL	-741.65	**-614.68**	-17.12
AIC	-3.79	**-4.12**	8.71
SIC	-3.65	**-3.76**	3.01

Note: Bolded numbers shown for different statistical tests indicate the best model performance for each statistical test. N=587.

4.2. Statistical performance

Table 4 shows that the highest increase/decrease in magnitude for all statistical tests was from the OLS to the STAR model with 20%, 18%, -35%, -17%, 9%, and 3% for R^2, $\overline{R}^2$, SSE, LL, AIC, and SIC respectively

The $\overline{R}^2$ shows that 77% variation in house prices was explained by the STAR model. In contrast, the OLS model only explained 65% of variation in house prices. The substantial percentage change of $\overline{R}^2$ from the OLS to the STAR model suggests that STAR has performed better, and was almost as good as the goodness of fit reported in Liu (2012), Pace *et al.* (1998b; 2000), and Sun *et al.* (2005). The SSE for the STAR model was much below that of the OLS model by as much as 35%. A reduction of 17% in LL value from the OLS to the STAR model also indicates an improvement in the model's goodness of fit. A value of more than three indicated by AIC and SIC for the STAR model signifies an improvement to that of the OLS.

4.3. Diagnostic performance

The above results reveal that STAR was better than OLS in terms of statistical performance. The next analysis examines whether STAR was able to capture spatial and temporal effects effectively. The Moran's I was used to identify spatial autocorrelation residuals while Breusch Godfrey was used to identify temporal autocorrelation residuals.

Table 5: OLS and STAR diagnostic performance

Diagnostic tests	OLS	STAR	Magnitude of Increase/ Decrease from OLS to STAR (%)
Moran's I	13.04	**2.12**	-84.00
Breusch-Godfrey	32.84	**0.28**	-99.00

Note: Bolded numbers shown for different statistical tests indicate the best model performance for each statistical test. Critical value for Moran I statistic is 2.33 at 1% level of significance. Breusch-Godfrey is distributed as Chi-Square χ^2 at 6.63 with 1 degree of freedom at 1% level of significance. All statistics were highly significant, at least at the 0.000000001 level. N=587.

Based on Table 5, the Moran's I value for the STAR model was 2.12, below the critical χ^2 value of 2.33 at 1% level of significance. This shows that no spatial residuals were left in the STAR regression and it managed to capture spatial autocorrelation effects effectively. In addition, the STAR model managed to reduce about 84% and 99% of spatial and temporal autocorrelation respectively compared to the OLS model. The STAR model also shows no evidence of temporal autocorrelation since the calculated value for Breusch-Godfrey, 0.28, was lower than the 1% critical χ^2 value of 6.63. In contrast, the OLS model has shown temporal autocorrelation residuals, although time indicators were included in the model. Overall, these results confirm that the STAR model has effectively captured spatial and temporal autocorrelation effects compared to the OLS model. However, these results were based on the in-sample data. To double-check, the predictive performance of the competing models were evaluated again using the out-sample data.

4.4. Predictive performance

Table 6 shows the comparison of predictive capability between the OLS and the STAR models. The STAR model has predicted the property prices with a much lower percentage of error compared to the OLS model for both in-sample and out-of-sample observations

Table 6: OLS and STAR predictive performance

Prediction Errors	OLS		STAR	
	In-Sample	Out-of-Sample	In-Sample	Out-of-Sample
MAPE	9.77	19.71	**8.21**	**5.22**
$\leq \pm$ 10%	350	22	**460**	**52**
$\geq \pm$ 10%	192	38	**82**	**8**

Note: Bolded numbers indicate the best model performance for each analysis.

About 85% of the total number of predictions of both in-sample and out-of-sample data using the STAR model fell within 10% of the original house prices. The OLS model has performed rather poorly with the majority of house prices being over-predicted with more than 10% of error from the average value. Based on the evidence provided in Table 6, it can be concluded that the STAR model was a better predictive model compared to the OLS model.

5. COMPARISON OF HOUSE PRICE INDICES

Figure 3 shows the comparison of the predicted property prices unadjusted for characteristics, based on the two competing models, their mean, and the Malaysian House Price Index (MHPI).

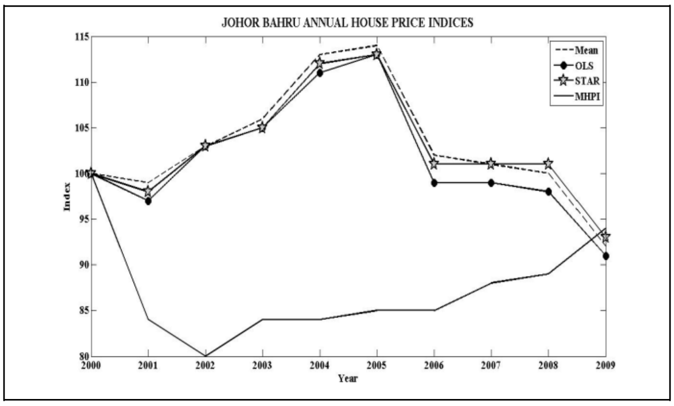

Figure 3: Johor Bahru annual house price indices

Figure 3 shows that the predicted property price indices using both competing models and their mean price indices, unadjusted characteristics, portray almost similar price index movements. This means, the two models' estimates were close with respect to the Johor Bahru housing market. The trend exhibited in the OLS and STAR indices demonstrate a more accurate house price movement over time. The comparison between the OLS and the STAR indices with MHPI indices shows an average disparity of 15% between those price indices, which was attributable to the neighbourhood effect.

Figure 4 shows that the price prediction using STAR model has tended to produced a closer trend line against the mean price index throughout the years as compared to that of the OLS modelling. This means, the incorporation of explicit spatial and temporal effects into the STAR indices has produced a more accurate prediction of neighbourhood-level house price index movements compared to its rival model

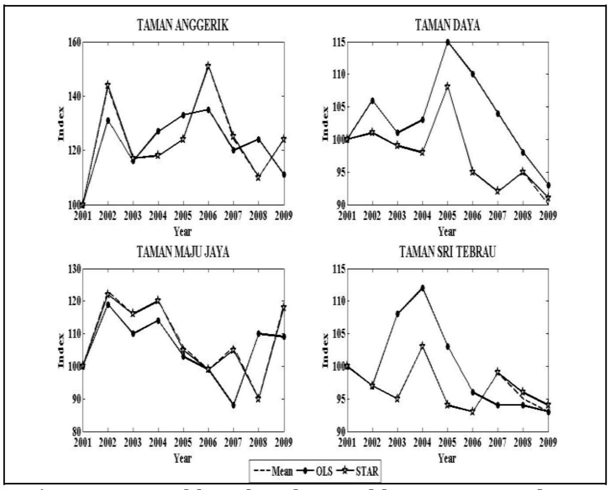

Figure 4: Neighbourhood annual house price indices

6. SUMMARY AND CONCLUSIONS

This paper has managed to establish an improved model for constructing neighbourhood-level house price index by simultaneously incorporating the spatial and temporal elements in the model specification. The $\bar{R}^2$ and SSE reported in this study show that there was a significant statistical improvement from using the OLS model to the STAR model. The value generated from Moran's I and Breusch-Godfrey tests confirmed that there were no spatial and temporal residuals left using the STAR model compared to using the OLS model. In addition, the STAR model has also performed very well in predictions, producing a relatively much lower MAPE. A 15% gap in house price movements between the disaggregated indices (mean price unadjusted for characteristics, OLS, and STAR indices) with the aggregated indices (MHPI) shows that there was a need for disaggregation of house price index at the neighbourhood level to avoid loss of information. Based on the empirical findings, it can be concluded that the STAR model could have become a very good estimation and prediction model for property valuation and property index construction purposes.

Simultaneous regressions performed based on STAR, in order to obtain the optimal number of spatial and temporal neighbours, have managed to address one of the STAR's weaknesses, namely a fixed number of hyper-parameters. Explicit tests against spatial and temporal autocorrelation using Moran's I and Breusch-Godfrey add to the current body of literature, since there was no evidence of formal diagnostic tests done on STAR as yet. Using MAPE to evaluate the predictive performance also has provided additional information for the STAR model, as previous works only reported error statistics for out-of-sample data. Despite using a small set of data for analysis, the data employed were pure neighbourhood-level transaction prices, which have provided a meaningful and more accurate information on house price movements rather than using the national-level data. This criterion has also qualified the STAR index to be regarded as an acceptable index as listed by Hwa (unpublished).

Future research may entail refinements in modelling the spatial and temporal data to study the degree of improvements in the model. This can be addressed by including varying parameters, random coefficients, and structural change. It would also be interesting to investigate various weight matrix specifications and their effects on the STAR outcomes. Future studies may also consider using geostatistical models because these models do not require weight matrix specification. This research may also be extended to include other types of properties such as commercial properties.

ACKNOWLEDGEMENTS

The authors would like to thank the Valuation and Property Services Department of Johor Bahru and the Johor Bahru City Council for providing the data and the Ministry of Higher Education (MOHE) for funding the research (Vote Number: 4F063). Our gratitude also goes to the anonymous reviewers for their constructive comments and suggestions on the early draft of this paper.

REFERENCES

Alonso, W. (1964). *Location and land use: Toward a general theory of land rent*. Cambridge, M. A.: Harvard University Press.

Atkinson, S. E. and Crocker T. D. (1992). The exchangeability of hedonic property price studies, *Journal of Regional Science*, 32(2): 169-183.

Basu, S. and Thibodeau, T. G. (1998). Analysis of spatial autocorrelation in house prices, *Journal of Real Estate Finance and Economics*, 17(1): 61-85.

Case, B., Clapp, J. M., Dubin, R., and Rodriguez, M. (2004). Modelling spatial and temporal house price patterns: A comparison of four models, *Journal of Real Estate Finance and Economics*, 29(2): 167-192.

Chica-Olmo, J. (1995). Spatial estimation of housing prices and locational rents, *Urban Studies*, 32(8): 1331-1344.

Clapham, E., Englund, P., Quigley, J. M., and Redfearn, C. L. (2004). Revisiting the past: Revision in Repeat Sales and Hedonic indexes of house prices. *May Meetings of the Homer Hoyt Institute*. May 2003. West Palm Beach, Florida: HHI.

Clapp, J. M. (2004). A semiparametric method for estimating local house price indices, *Real Estate Economics*, 32(1): 127-169.

Clapp, J. M. and Giacotto, C. (1998). Price indices based on the hedonic repeat sales method: application to the housing market, *Journal of Real Estate Finance and Economics*, 16(1): 5-26.

Cliff, A. D. and Ord, J. K. (1969). The problem of spatial autocorrelation. In Scott, A. J. (ed.), *London Papers in Regional Science, Vol. 1*, pp. 25-55, London: Pion.

Cliff, A. D. and Ord, J. K. (1973). *Spatial Autocorrelation*, London: Pion.

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

Colwell, P. F., Munneke, H. J. and Trefzger, J. W. (1998). Chicago's office market: price indices, location and time, *Real Estate Economics*, 26(1): 83-106.

Daria, S. (2007). *Spatial-temporal modelling in real estate market analysis.* Unpublished Masters Thesis. The University of Queensland, Australia
.
Daria, S., Alicia, N. R. and Rodney, S. (2008). A comparison of methods for spatial-temporal forecasting with an application to real estate prices, Paper presented at the *Australasian Meeting of the Econometric Society*, July 2008.

De Silva, D. G., Kruse, J. B. and Wang, Y. (2008). Spatial dependencies in wind-related housing damage, *Natural Hazards*, 47(3): 317-330.

Domingo, E. V. and Fulleros, R. F. (2005). Real estate price index: a model for the Philippines. In: Bank for international settlements (ed.), *Real estate indicators and financial stability*, Vol. 21, pp. 139-148. Basel, Switzerland: Bank for international settlements.

Dubin, R. A. (2003). Robustness of spatial autocorrelation specifications: some Monte Carlo evidence, *Journal of Regional Science*, 43(2): 221-248.

Dubin, R. A., Pace, R. K. and Thibodeau, T. G. (1999). Spatial autoregression techniques for real estate data, *Journal of Real Estate Literature*, 17(1): 79-95.

Flaherty, J. (2004). Measuring and evaluating changes in returns for residential property. *10th Pacific RIM Real Estate Society Annual Conference*. 25-28 January. Bangkok: PRRES.

Fletcher, P., Gallimore, P. and Mangan, J. (2000). The modelling of housing submarkets, *Journal of Property Management*, 18(5): 366-374.

Fletcher, P., Mangan, J., and Raeburn, E. (2004). Comparing hedonic models for estimating and forecasting house prices, *Property Management*, 22(3): 189-200.

Fotheringham, A. S. and Rogerson, P. A. (1993). GIS and spatial analytical problems, *International Journal of Geographic Information Systems*, 7: 3-19.

Gatzlaff, D. H. and Ling, D. C. (1994). Measuring changes in local house prices: an empirical investigation of alternative methodologies, *Journal of Urban Economics*, 35(2): 221-244.
Goodman, A. C. (1978). Hedonic prices, price indices, and housing markets, *Journal of Urban Economics*, 5(4): 471-484.

Goodman, A. C. and Thibodeau, T. G. (2007). The spatial proximity of metropolitan area housing submarkets, *Real Estate Economics*, 35(2): 209-232.

Herath, S. and Maier, G. (2010) A framework for analysing house prices using time, space, and quality criteria, Paper presented at the

European Real Estate Society Conference, June 2010.

Hoesli, M. and MacGregor, B. (2000). *Property Investment*. England: Pearson Education Limited.

Hwa, T. K. (Unpublished). *The challenges in developing a property index*.

Isakson, H. R. (1986). The nearest neighbours appraisal technique: an alternative to the adjustment grid methods, *Real Estate Economics*, 14(2): 274-286.

Li, M. M. and Brown, H. J. (1980). Micro-neighbourhood externalities and hedonic housing prices, *Land Economics*, 56(2): 125-141.

Liu, X. (2012). Spatial and temporal dependence in house price prediction, *Journal of Real Estate Finance and Economics*, In press.

Meese, R. A. and Wallace, N. E. (1997). The construction of residential housing price indices: a comparison of repeat sales, hedonic regression and hybrid approaches, *Journal of Real Estate Finance and Economics*, 14(2): 51-73.

Nappi-Choulet, I., Maleyre, I. and Maury, T-P. (2007). A hedonic model of office prices in Paris and its immediate suburbs, *Journal of Property Research*, 24(3): 241-263.

Nappi-Choulet, I. and Maury, T-P. (2009). A spatiotemporal autoregressive price index for the Paris office property market, *Real Estate Economics*, 37(2): 305–340.
National Institue of Valuation (1996a). *Data collection guidelines for house price index*, Kajang: INSPEN.

National Institute of Valuation (1996b). *Malaysian house price index: a technical summary*, Kajang: INSPEN.

National Property Information Centre (2009). *The Malaysian house price index q3-q4 2009*, Kuala Lumpur: Valuation and Property Services Department.
Pace, R. K., Barry, R., Clapp, J. M. and Rodriguez, M. (1998a). Spatial statistics and real estate, *Journal of Real Estate Finance and Economics*, 17(1): 5-13.

Pace, R. K., Barry, R., Clapp, J. M. and Rodriguez, M. (1998b). Spatiotemporal autoregressive models of neighbourhood effects, *Journal of Real Estate Finance and Economics*, 17(1): 15-33.

Pace, R. K., Barry, R., Gilley, O. W. and Sirmans, C. F. (2000). A method for spatial-temporal forecasting with an application to real estate prices, *International Journal of Forecasting*, 16(2): 229-246.

Sun, H., Tu, Y. and Yu, S. M. (2005). A spatio-temporal autoregressive model for multi-unit residential market analysis, *Journal of Real Estate Finance and Economics*, 31(2): 155-187.

Suriatini Ismail (2005). *Hedonic modelling of housing markets using Geographical Information System (GIS) and spatial statistics: a case study of Glasgow, Scotland*, Unpublished PhD Thesis, University of Aberdeen, United Kingdom.

Theriault, M., Des Rosiers, F., Villineuve, P., and Kenstens, Y. (2003). Landscaping attributes and house prices: Looking at interactions between property features and homeowners' profiles, *Property Management*, 21(1): 25-62.

Valente, J., Wu, S., Gelfand, A. and Sirmans, C. F. (2005). Apartment rent prediction using spatial modelling, *Journal of Real Estate Research*, 27(1): 105-136.