COMBINING GEOGRAPHIC INFORMATION SYSTEMS AND REGRESSION MODELS TO GENERATE LOCATIONAL VALUE RESIDUAL SURFACES IN THE ASSESSMENT OF RESIDENTIAL PROPERTY VALUES

ABDUL HAMID
Universiti Teknologi Malaysia

ABSTRACT

Locational Value Residual Surface (LVRS) has been suggested as an alternative to resolving the difficulty in the traditional modelling of locational influence on property values in a particular area. The objective of this paper was to compare the relative performance of models that apply locational value residual surface (LVRS) and the traditional multiple regression models in the prediction of residential property values. A controlled sample of 125 single- and double-storey residential properties was used to construct regression models. It was found that models applying LVRS were marginally better than the traditional models in predicting property values.

Keywords: Locational factors, locational value residual surface (LVRS), Geographical Information System (GIS), multiple regression analysis (MRA), residential property values.

INTRODUCTION

Real estate is a multi-dimensional heterogeneous commodity, characterized by durability and structural inflexibility as well as spatial immobility. It has a unique bundle of attributes such as accessibility to work, transport, amenities, physical characteristics, neighbourhood, and environmental quality (Muth, 1960; Ridker and Henning, 1967; Stegman, 1969; Kain and Quigley, 1970; Evans, 1973; So et al., 1997). Many of these attributes are spatially-related in the form of popularly known as "location, location, location" hierarchy (Pearson, 1991).

Real estate is spatially unique in which location is an intrinsic attribute that directly determines the quality and market value of the property. However, modelling the locational factors in property valuation has proved difficult because of the wide range of spatially definable attributes, which may or may not affect value at a

particular time and location. Furthermore, there is little literature consensus as to the best proxy for locational factor measurement.

Multiple regression models (MRA) are considered a classical and primary technique for explaining and predicting property values whereby locational factors can potentially be taken into account. In particular, MRA has been used to estimate residential property values in the U.S since the 1950s and in the U.K since the 1980s (Pendleton, 1965; Greaves, 1984; Adair and McGreal, 1988). It was also applied in other countries such as Australia, New Zealand, and Singapore, but has yet to be widely practiced in Malaysia.

In applying MRA, valuers identify the data to be specified and measured in quantitative form. This task becomes more complicated when the locational influence on property values need to be explicitly identified and modelled. Research that has sought to assess the determinants of property values has either ignored detailed location analysis (Wyatt, 1997; So et al., 1997) or just dealt with it only in a very general sense. Some researchers have even simply omitted the locational variables (Ferri, 1977). An interview with the local valuation-based firms and government offices such as Rahim & Co, Jurunilai Bersekutu, C.H.William & Talhar, Raja Hamzah, Ismail & Co, Zaki & Partners, T.D.Aziz, Ami and Associates, and Valuation and Property Services Offices (JPPH) disclosed that valuers infer a substantial amount of information about a property from its location, based on their local knowledge and experience.

This article discusses the use of locational value residual surface (LVRS) generated using the combination of Geographical Information System (GIS) and Multiple Regression Analysis (MRA) in a "hybrid" predictive model that utilizes LVRS to create a locational adjustment factor.

The second part of this paper briefly reviews the traditional approaches to modelling location followed by a discussion on value residual surface. A brief description of the study area is discussed in the third section. Data and analysis procedure are discussed in the fourth section. Section of results and discussion follows then. The final part of this paper concludes the study.

TRADITIONAL APPROACHES TO MODELLING LOCATIONAL FACTORS

Location is an amalgam of several factors that includes a number of spatial elements such as accessibility to shopping, employment, educational and leisure

facilities; exposure to adverse environmental effects such as traffic noise and hazard; neighbourhood amenity; perceived levels of neighbourhood security (Gallimore et al., 1996; McCluskey et al., 2000). From these, two key components of location can be isolated, i.e. neighbourhood quality and accessibility (McCluskey et al., 2000). Few, however, are capable of numeric measurement but, even that, the measures may not always be valid representation of the influence, especially because of the complex interaction of value factors. For example, the common approach to examining locational influence on property values is to include a distance variable from the central business district (CBD), assuming homocentric locations. This is based on the traditional location theory that examines the role of accessibility to central locations on property prices (Wendt, 1957). However, house prices are determined not only by accessibility but also by the environmental attributes of location (Stegman, 1969; Richardson, 1971; Pollakowski, 1982).

There are also theories of multiple-nuclei model incorporating the concentric pattern that are more appropriate for analyzing locational influence on property values. For example, pattern of property values may reflect the influence of satellite towns rather than that of regional centres (Hamid, 2001). On the basis of these theories, locational partitioning is analyzed by using locational dummy variables (Hamid, 2003). Essentially, this is to subdivide a particular geographic area into "realistic" sub-markets or neighbourhoods. However, this could pose modelling constraint in terms of data representativeness when some neighbourhoods with too few transactions give rise to "small sample" problems in the statistical estimation.

There are also researchers who employ more sophisticated location measurements such as using the type of transport, time taken per trip, and transportation cost. Table 1 shows some of the traditional locational proxy variables used in previous regression models.

Table 1: Traditional location proxy variables used in previous regression models

Authors	Location proxy variables	Location proxy variables
Jud and Winkler (1991)	Distance	The east/west and north/south distances of property from the centre of metropolitan statistical area (MSA).
Hamid (1991)	Distance	Property distance from the nearest town.
Azhari (1993)	Neighborhood quality scores	Use Thurstone paired-comparison technique base on Thurstone questionnaire.
Rodriguez (1994)	Environmental	Presence of good view.
Adair, et al. (1996)	Accessibility and environmental	Distance to work, distance to primary school, distance to secondary school, distance to corner/local shops, distance to shopping centre, distance to city centre, house on a main bus route, proximity to bus route, travel time to city centre, and travel time to work; and condition of neighbourhood, attractiveness of the area, quality of neighbouring houses, type of neighbouring houses, density of housing, wooded area/tree coverage, slope/topography of the land, attractive views, open space, non-residential uses in the area, vacant sites, traffic noise, level of owner-occupation in neighbourhood, level of education in neighbourhood, security from crime, quality of schools, and religious composition of neighbourhood.
So, et al. (1997)	Accessibility and environmental	Access to transportation, access to parks, sea view, high floor, proximity to shopping centres, presence of sports facilities, presence of swimming pools and presence of car parks; and distance to the mass transit railway (MRT), distance to bus and minibuses and evironmental quality.

Henneberry (1998)	Binary variables for neighbourhood; and accessibility	Neighbourhood characteristics base on local estate agent experience; and distance to tram stops, distance to tram lines and areas adjacent to the tram route.
Watkins (1999)	Discrete neighbourhood	neighbourhood classification base on factor analysis - household size, percentage of sole occupiers in postcode sector, percentage of households with two or more children in postcode sector, distance to city centre from centre of postcode sector in miles, presence of garden, percentage car ownership in postcode sector, no. of bedrooms in dwelling, presence of a garage, percentage of two income households in postcode sector, presence of a dining room, presence of a utility room, no. of living rooms, age band of dwelling, presence of a parking space.
Fletcher (2000)	Binary variables for neighbourhood; and discrete neighbourhood	Neighbourhood classification base on postcode location; and disaggregated models are conducted separately by postcode location, property type and age of properties.
Azhari (2001)	Accessibility; and discrete neighbourhood	Neighbourhood condition classification based on Chow test, R2 test and predictive performances of model.
Zan (2001)	Binary variables for neighbourhood; and accessibility	Neighbourhood characteristic of West, South, and North; and direct distance from CBD.
Frew and Wilson (2002)	Accessibility	Distance to city centre, distance to highway, and distance to intersection.

As the table shows, the above locational factors are represented by discrete neighbourhood variables. A problem commonly faced in the use of discrete neighbourhood variables is the requirement for subjective judgments about the boundaries of each geographic unit and the numeric indicator for geographic quality. To solve this problem, some researchers have simply asked local valuers or local experts to rank the neighbourhood quality (Hickman et al., 1984). There is little consensus, however, on which variables are the best proxy for neighbourhood quality measurement, based on actual house price or property physical characteristic or housing quality or ward boundary or should be defined in spatial

terms (Adair et al., 1996; Can, 1990). Therefore, neighbourhood quality is arguably an unobservable variable (Dubin and Sung, 1987). When an overarching model is adopted, such decisions may lead to disparities or inconsistencies where properties adjoin or close to neighbourhood boundaries. A hard edge may be implied at such boundaries, whereas in reality the varying influence of location may operate far more smoothly and the spatial trends occur as opposed to distinct areas of homogeneous property subsets (Gallimore et al., 1996; Mackmin, 1989).

LOCATIONAL VALUE RESIDUAL SURFACE (LVRS)

The limitation and highly complex process of discrete measurement of location have encouraged researchers to search for alternative approaches to derive locational compensation factors. Locational influence within an area can be established through an analysis of value residuals (errors) from a location-blind model (Anon, 2003). The residuals or the discrepancies between the actual and estimated property prices using this location-blind model can be regarded as the value of location (Jack, 1989; Gallimore et al., 1996; Figueroa, 1999; McCluskey et al. 1999). According to Ward et al. (2002), similar improved parcels will sell for different amounts if located in different parts of a town or within a same sub-division. The amount of difference between the actual sale and the estimated prices isolates the locational adjustment better than estimating the adjustment during the initial modelling phase. The coefficients of base model account for the total contribution of the independent variables to the sale price. The residual difference is made up of location as well as data inconsistencies, market variability and model inadequacies. Since location can intuitively and empirically be an influential factor on residential property values, it is possible to estimate locational influence for any point within the area by using the generated residuals (Gallimore et al., 1996).

In constructing value of location through this approach, the price prediction errors between the actual price, y_i and location-blind predicted price, $y^\wedge$ models are transformed into a 2½-dimensional locational value residual surface (LVRS). The latitude and longitude (x-y) co-ordinates indicate the geo-referenced location of each particular property, while the generated residuals ($e = y_i - y^\wedge$) are depicted on the vertical axis (z-axis). This residual surface could then be used to adjust for under- or over-prediction of any property price within the area to estimate influence of a particular location on residential property values. This variable is then included, along with other variables, in a multiple regression to capture locational influence.

The construction of the surface can be performed in a number of ways using different spatial interpolation techniques such as the Inverse Distance Weighting (IDW) and kriging techniques. A detailed description of these techniques can be found in Royle et al. (1980), Isaaks and Srivastava (1989), Oliver and Webster (1990), and Burrough and McDonnell (1998).

ANALYSIS PROCEDURE

The selected study area comprised part of three adjoining housing estates, namely Taman Pelangi, Taman Sentosa, and Taman Sri Tebrau, in the sub-urban of Johor Bahru city, Malaysia. They were selected particularly for their active sales transaction activities. The residential sales data were obtained from the Johor Bahru Valuation and Property Services Office (JPPH) covering a three-year period from January 2001 to December 2003. This was a period during which local advisers felt that the movement of property prices in the locality was quite stable. The sales records included sales price, transacted date, lot number, property characteristics and property addresses. The last element was particularly recorded to ensure that the location of each property could subsequently be digitized over the study area.

A total of one hundred and twenty-five terrace residential units in the three adjoining housing scheme were used for estimating the regression models. A holdout sample comprising twenty-one transacted units was randomly set aside for predictive purposes. The distribution of the sales data in the study area is shown in Figure 1.

Figure 1: Location map showing the distribution of residential lots
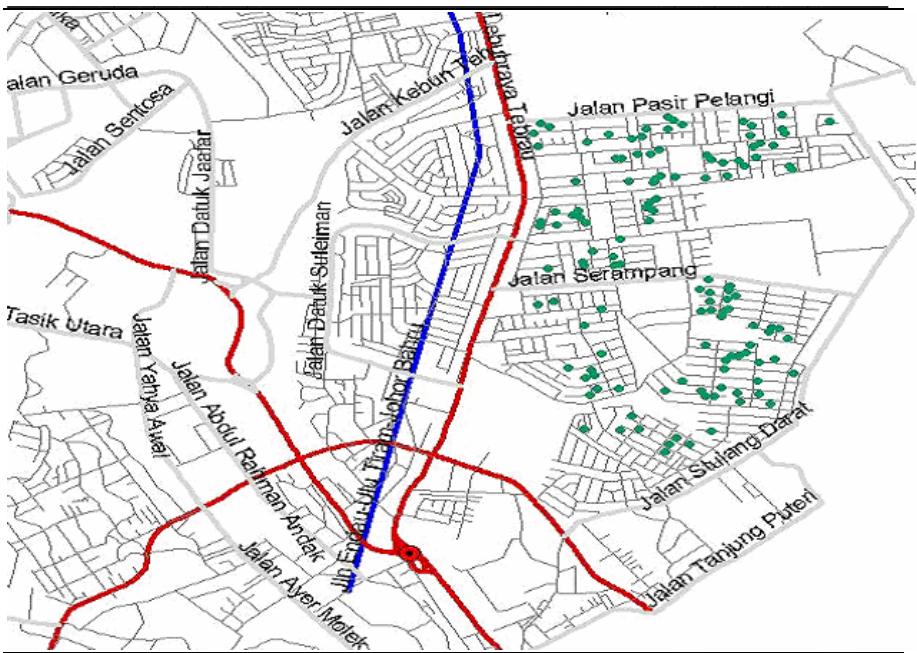
In order to calculate spatial prediction errors of residential property values, location-blind model was estimated utilizing the data describing the physical characteristics of the properties. Due to data limitation, the variables used represented only a few property characteristics and, therefore, some form of model misspecification can be expected from the modeling exercise. The *Statistical Package for the Social Science* was used to run the regression. The variables included in the location-blind model are shown in Table 2.

The Box-Cox test was applied to determine model's best functional form (Box and Cox, 1964). Besides, the diagnostic tests for multicollinearity (Farrar and Glaubar, 1967; Reichert and Moore, 1986), autocorrelation (Durbin and Watson, 1951), heteroscedasticity (Breusch and Pagan, 1979) and model misspecification (Ramsey, 1969; McGuirk, 1993) were also performed.

Table 2: List of variables used in the preliminary model

Variable	Unit of Measurement
Dependent: House price (PRIC)	Ringgit (RM)
Independent Variables:	
Land Area (LA)	Square meters (sq. m.)
Gross Floor Area (GFA)	Square meters (sq. m.)
Ancillary Area (AA)	Square meters (sq. m.)
Type of Terrace (T1)	Dummy (single-story terrace = 1)
Age (AGE)	Years
Condition of the Building (GCOND)	Dummy (good condition = 1)
Floor Finishes (FFNISH2)	Dummy (class 2 - combination of parquet, terrazzo, cement, mosaic or tiles floor finishes = 1)
No. of Bedroom (3BEDR)	Dummy (3 bedrooms = 1)

In deriving locational value adjustment factor, prediction residuals from the location-blind model were used to construct LVRS whereby the residuals were

expressed in percentage terms. The LVRS was generated by using two surface interpolation techniques – IDW and kriging. In building IDW-based LVRS, the surface was generated by calculating weights according to the reciprocal of the distance between the subject and the neighbouring residuals. Residuals within a 500-meter radius from the subject property were being identified. In the same manner, ordinary kriging was used to construct kriged-based LVRS. Both surfaces were constructed by using *ArcGis* software.

Following Ward et al. (2002) and Anon (2003), the traditional location modeling included distance from the Central Business District (CBD) and discrete boundary-based housing neighbourhood as locational variables.

The models were evaluated on the basis of comparative adjusted R^2, F-value, standard error of estimate (SEE), sum squared error (SSE), and "change of significance" of a particular variable specified, and the predictive performance of the models. The predictive performance was assessed on the basis of mean absolute percentage error (MAPE) and the proportion of accurate predication of property prices (Kennedy, 1992; Hamid, 2001).

RESULTS AND DISCUSSION

Modelling results

Table 3 shows the basic descriptive statistics of the sample. About 48% of the sample size was made up of properties in Taman Pelangi, followed by Taman Sentosa (35%), and Taman Sri Tebrau (17%). The sampled properties were located about 2.3 – 3.5 km from the Johor Bahru city centre.

The mean land area of the properties was about 164 sq. m. The range of land area was purposely controlled in the sample to avoid too much variance of this variable. On average, the ratio of gross floor area to land area was approximately 3.7:5.0. The mean ancillary area of the sampled properties was about 32 sq. m.

Table 3: Descriptive statistics of the sample variables

	Minimum	Maximum	Mean	Std. Deviation
Price (RM)	160,000.00	445,000.00	263,125.61	60,429.89
Land Area (sq. m.)	141.02	216.82	163.67	14.80
Gross Floor Area (sq. m.)	69.21	218.32	121.78	34.50
Ancillary Area (sq. m.)	7.43	91.14	31.78	12.37
Single Storey Terrace	0.00	1.00	0.46	0.50
Double Storey Terrace	0.00	1.00	0.54	0.50
Age of Building (years)	15.00	29.00	22.41	3.21
Good Building Condition	0.00	1.00	0.99	0.09
Good Finishes	0.00	1.00	0.97	0.18
Exp Finishes	0.00	1.00	0.03	0.18
3 Bedrooms	0.00	1.00	0.41	0.49
4 Bedrooms	0.00	1.00	0.59	0.49

On the basis of standard deviation, gross floor area has more variation compared to land area and ancillary area. This means people in the market could have been more concerned about differences in floor area, making this factor a vitally influential physical characteristic determining the prices of residential properties. This is evident in the different size of regression coefficients of both variables (see Table 5).

The sample was rather dominated by double-storey 4-bedroom terrace residential (i.e. 59% of the sample size). Besides, 99% of the residential properties, single- and double-storey terrace alike, were of good building condition with the mean building age of about 22 years. Since these are old housing schemes, the majority of the sampled properties have undergone some renovation. The sample shows that 97% of the properties have parquet, terrazzo, cement, mosaic or tile floor finishes.

Based on the Box-Cox transformation, the adjusted R^2 (see Table 4) shows that the log-log function was the most appropriate model to choose. For this reason, only the log-log function is reported in this paper.

Table 4: Adjusted R^2 from Box-Cox transformation on the location-blind model

λ_2 \ λ_1	-1.0	0	1.0
-1.0	0.810	0.808	0.773
0	0.806	0.816*	0.794
1.0	0.788	0.810	0.800

* The adjusted R^2 corresponding to $\lambda_1 = \lambda_2 = 0$, which refers to the log-log function.

Table 5 shows the results of the location-blind model. The model was used as a basis for generating residuals for prediction surface later. The model explained about 83 percent variation in residential property prices in the study area. Land area, gross floor area, age of the building, and the building condition were statistically significant in this model. Their coefficient signs show that all of these variables were theoretically plausible. The unstandardized regression coefficients indicate that residential property prices could have been more elastic to changes in the gross floor area compared to changes in other physical characteristics such as land area, ancillary area and age of building.

Table 5: Results for the location-blind residential property value model

R^2 / adj. R^2	0.828
Adj. R^2	0.816
F-value	63,478
SEE	0.183
SSE	0.04280

Variable	Unstandardized Coefficients	Standardized Coefficients	t-value
Dependent variable: Log of House Price (LG_PRIC)			
Independent variables:			
(Constant)	3.781		11.252**
Log of Land Area (LG_LA)	0.352	0.117	2.573*
Log of Gross Floor Area (LG_GFA)	0.520	0.074	6.691**
Log of Ancillary Area (LG_AA)	1.894E-02	0.027	0.613
Type of Terrace (Single storey) (T1)	-0.287	-0.189	-4.057**
Log of Age (LG_AGE)	0.144	-0.129	-1.446
Condition of the Building (Good) (GCOND)	-1.095E-03	0.136	3.213*
Floor Finishes (Class 2) (FFNISH2)	-0.014	0.002	0.055
No. of Bedroom (3 bedrooms) (3BEDR)	-0.118	-0.455	0.443

**Significant at .01; *Significant at .05.

The diagnostic tests for multicollinearity, autocorrelation, heteroscedasticity and model misspecification show that the model was free from the first three problems

(Tables 6). However, the RESET has indicated model misspecification. This may have been attributed to the true omission of some explanatory variables, particularly the locational variables. The exclusion of such variables can be expected to cause the coefficient estimates for the included explanatory variables in the model to be biased. Therefore, this model must be interpreted and used with caution.

Table 6: Partial correlation coefficients

	LA	GFA	AA	T1	AGE	GCOND	FFNISH2	3BEDR
LA								
GFA	0.2453							
AA	-0.2690	0.3106						
T1	-0.1654	-0.5951	-0.0345					
AGE	0.1185	-0.3680	0.2145	-0.0591				
GCOND	0.0972	-0.0775	0.1336	-0.0334	0.0559			
FFNISH2	0.0270	-0.0917	0.1790	0.0096	-0.0512	-0.0322		
3BEDR	0.3470	-0.2354	0.1763	0.4517	-0.0746	-0.0988	-0.0323	

A two-dimensional view of IDW-based and kriging-based LVRS is shown in Figure 2 and Figure 3 respectively. Similarly, Figure 4 and Figure 5 show the LVRS in a 2½-dimensional view respectively. These visual profiles provide an overall idea of at least two things. Firstly, they help valuers to find explanation for locational factors that give rise to bumps and potholes (Figure 4) or ridges and trenches (Figure 5) in the study area. This necessarily needs the knowledge and familiarity of the local surroundings and property market.

From site inspection, it was observed that the bumpy areas represent neighbourhoods that were quite strongly influenced by positive locational elements such as proximity to shopping centres, shop houses, leisure facilities and open space or park (see Figure 4 or Figure 5). The appeal of business activities, neighbourhood and environmental qualities in these areas has tended to bid up the sale prices of residential properties nearby. Examples of such areas can be found to the southeast of Taman Sri Tebrau (southwest of Jalan Serampang) and to the middle and northeast of Taman Pelangi (through which Jalan Serampang divides or area to the northwest of Jalan Stulang Darat) (see Figure 2).

On the other hand, being located to the east of Taman Sentosa (southwest of Jalan Pasir Pelangi) and to the south of Taman Pelangi (southeast of Jalan Pasir Pelangi or northeast of Jalan Serampang), the proximity of houses to negative locational elements such as oxidation ponds, industrial area and noisy main roads was seen to be associated with potholes or trenches. Besides, the structural designs of residential units in these "trenchy" neighbourhoods were not considered superior

than those in the "ridgey" neighbourhoods. Houses in these areas were also far away from any shop houses and other commercial activities.

Secondly, how locational adjustment should be made across the study area. The bumps or ridges represent under-prediction while potholes or trenches represent over-prediction of residential property prices. Consequently, the "bumpy" or "ridgey" areas need an upward while the "potholey" or "trenchy" areas need a downward locational price adjustment.

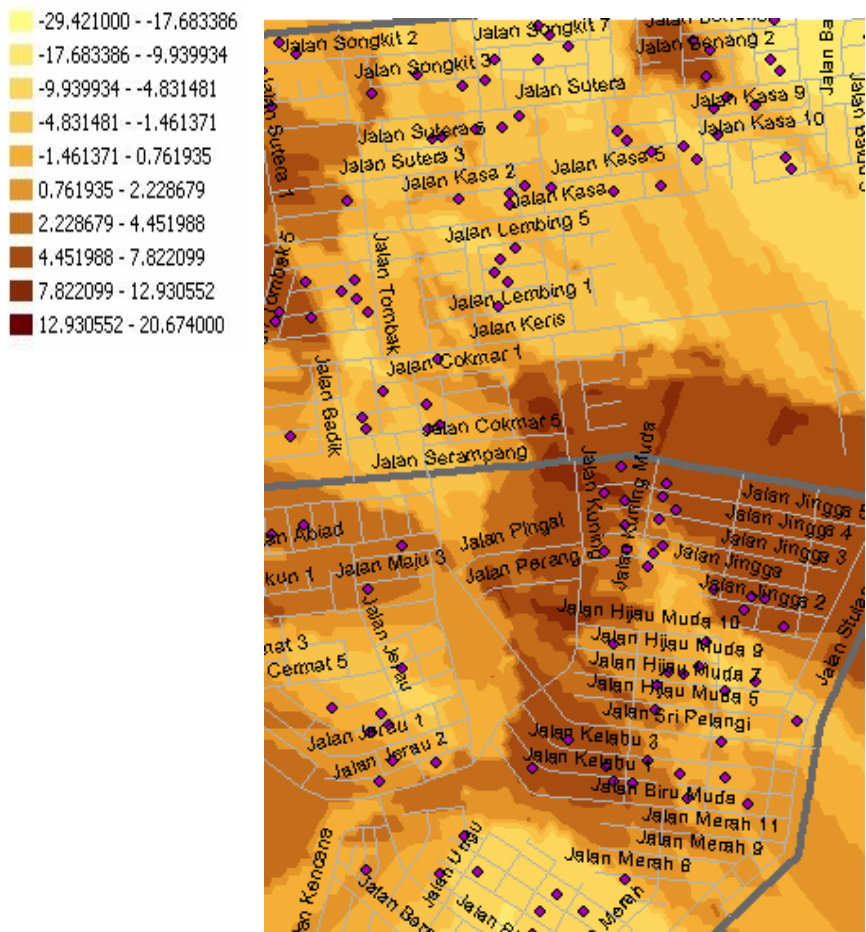**Figure 2: IDW-based 2-D view of LVRS**

Figure 3: Kriging-based 2-D view of LVRS

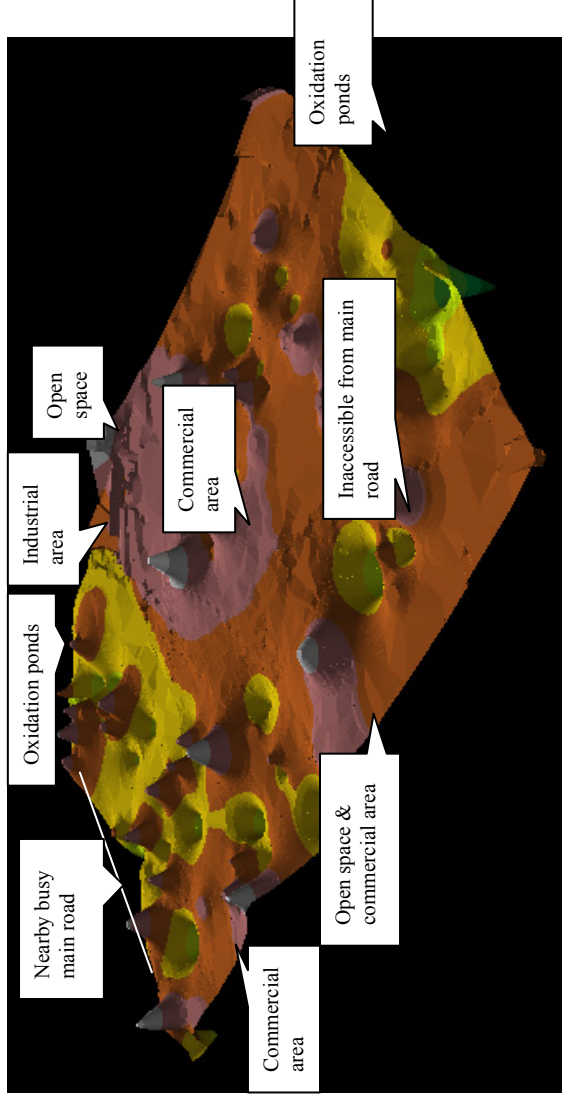**Figure 4: IDW-based 2½-D view of LVRS**

Labels in figure:
- Oxidation ponds
- Open space
- Industrial area
- Commercial area
- Inaccessible from main road
- Oxidation ponds
- Nearby busy main road
- Open space & commercial area
- Commercial area

Figure 5: Kriging-based 2½-D view of LVRS

Legend — Elevation:
- 5.849 – 7.822
- 3.875 – 5.849
- 1.901 – 3.875
- -0.072 – 1.901
- -2.046 – -0.072
- -4.019 – -2.046
- -5.993 – -4.019
- -7.966 – -5.993
- -9.940 – -7.966

Labels: Oxidation ponds; Open space; Industrial area; Commercial area; Inaccessible from main road; Nearby busy main road; Commercial area; Open space & commercial area
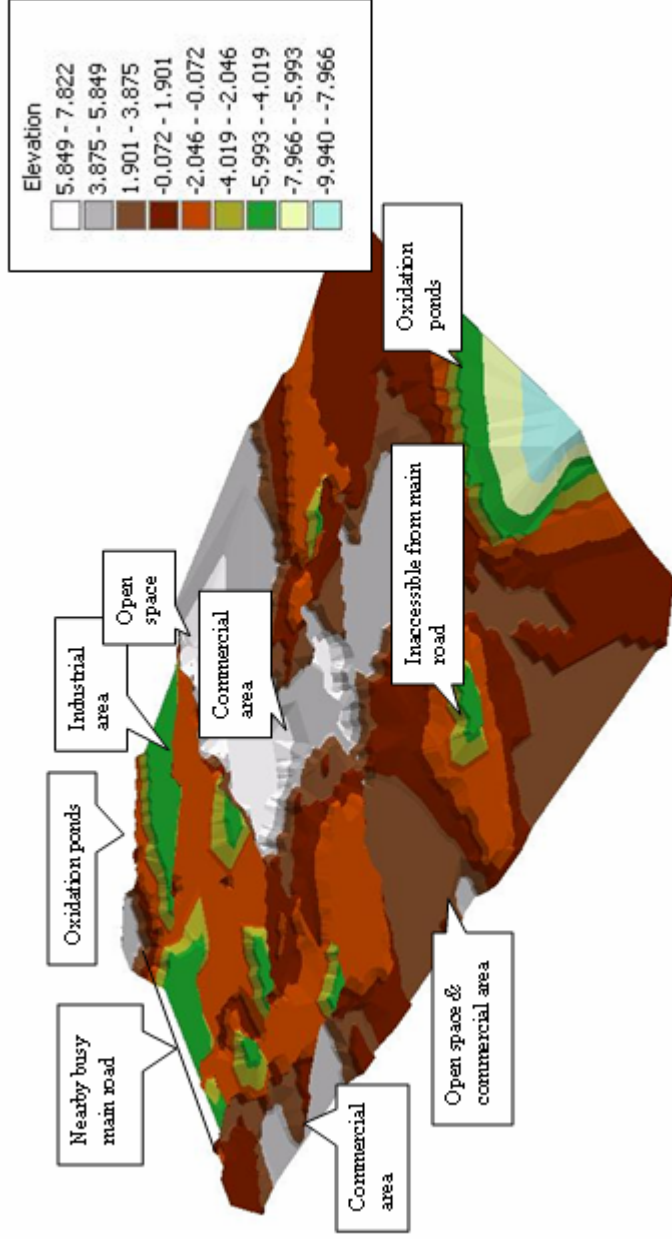
Table 7: Comparison of the competing models

Model	Model I	Model II	Model III	Model IV
Model form	Double log	Double log	Double log	Double log
R^2	0.995	0.869	0.840	0.842
Adjusted R^2	0.925	0.859	0.828	0.828
F-value	141,730.8	84.921	67.246	60.846
SEE	0.00093	3.756E-02	4.150E-02	4.143E-02
SSE	0.000	0.162	0.198	0.196
Dependent variable (LG_PRIC)	House price (LG_PRIC)	House price (LG_PRIC)	House price (LG_PRIC)	House price (LG_PRIC)
Independent variables:	**Coefficient**	**Coefficient**	**Coefficient**	**Coefficient**
(Constant)	3.776	3.726	4.220	4.238
	(165.955)***	(14.626)***	(13.120)***	(13.269)***
Log of Land Area (LG_LA)	0.356	0.367	0.221	0.159
	(37.598)***	(3.565)***	(1.829)**	(1.207)
Log of Gross Floor Area (LG_GFA)	0.520	0.494	0.465	0.449
	(91.051)***	(7.615)***	(6.244)***	(5.921)***
Log of Ancillary Area (LG_AA)	1.013E-02	0.019	0.033	0.040
	(17.199)***	(0.793)	(1.270)	(1.495)*
Type of Terrace (Single storey) (T1)	-2.607E-02	-0.035	-0.040	-0.041
	(-15.945)***	(-2.246)**	(-2.237)**	(-2.274)**
Log of Age (LG_AGE)	-0.290	-0.191	-0.248	-0.199
	(-54.666)***	(-3.195)***	(-3.830)***	(-2.586)**
Condition of the Building (Good) (GCOND)	0.144	0.117	0.124	0.120
	(42.217)***	(2.985)***	(2.840)***	(2.735)**

Floor Finishes (Class 2) (EXPFINIS)	9.248 E-04 (-9.416)***	-0.025 (-1.290)	-0.018 (-0.820)	-0.017 (-0.789)
No. of Bedroom (3 bedrooms) (BED3)	5.752E-03 (6.844)***	0.010 (0.903)	0.012 (0.904)	0.007 (0.486)
Location:				
IDW-based LVRS	5.403E-02 (137.439)***			
Kriging-based LVRS		0.005 (5.999)***		
Distance from CBD			-0.0000243 (-2.944)***	
TMN PELANGI				0.031 (1.885)**
TMN SENTOSA				-0.002 (-0.140)

***Significant at .01; **Significant at .05; *Significant at .10. Model I used a slightly different data set than other 3 models. Model II used LVRS as locational variable by applying kriging surface interpolation technique while Model III and Model IV used traditional approach with CBD and neighborhood subdivision of housing schemes as locational variables, respectively.

Table 8: Predictive capability of the models

	Model I	Model II	Model III	Model IV
Prediction errors:				
Maximum	11.89	24.55	23.67	49.36
Minimum	1.59	1.81	1.11	1.67
Mean absolute percentage error (MAPE)	7.41	11.20	12.54	22.81
$< \pm 10\%$	57	43	21	29
$> \pm 10\%$	43	57	79	71

This adjustment factor can be created and included into the original location-blind model (Table 5) utilizing information from the LVRS surface. Using the IDW-based LVRS (Figures 2 and 4) and kriging-based LVRS (Figures 3 and 5), the locational price adjustment components were created in the second-stage models. These locational components were represented by the variable Location in the hybrid models (Table 7). These models have explained about 84-99.5% variation in the prices of residential properties in the study area.

Except for the ancillary area, type of floor finishes and number of bedroom in Model II, all other variables in both models were statistically significant at $\alpha = 0.01$ and have correct signs. Quite obviously, gross floor area and land area were two most important variables positively influencing the total property prices while building age was an important detractor of the selling prices.

As in the previous models, the diagnostic tests indicated no evidence of heteroscedasticity and autocorrelation in these hybrid and traditional multiple regression models. Breusch-Pagan test showed that the four models have Q-values smaller than critical value of 2.730. The Durbin- Watson test also rejected the hypothesis that the models have negative or positive autocorrelation. However, Ramsey's test indicated model misspecification even after the inclusion of locational components. The main reason for this phenomenon could be the exclusion of some other important physical variables of the residential properties such as architectural style, car porch, number of toilets, etc in the four models. Besides, the exclusion of locational elements could be expected in the traditional regression models. Because of this, it could be expected that the residential price estimates of some residential attributes included in the models could have been biased. Again, the interpretation of the regression coefficients should be made with some caveat.

Table 7 shows the relative statistical performance of the models that use GIS-MRA-generated LVRS (Models I and II) and those that apply traditional approaches to specifying locational variables (Models III and IV). By looking at the R^2 and adjusted R^2, Model I and II have explained the variation in the

residential property prices better than Model III and IV. In terms of F-value, Models I and II also have a higher level of overall significance compared to Models III and IV. Particularly for Model I, it has shown a much higher level of F-value than other models. Further comparison shows that Models I and II have lower standard error of estimate (SEE) and sum squared errors (SSE), compared to Models III and IV. Overall, the t-values for all regressors in Models I and II were higher that those in Models III and IV. In particular, Model I has much higher t-values than other models.

The next assessment on the quality of the competing models was on their relative capability in predicting unsold residential properties in the study area. The basis of comparison was the mean absolute percentage error (MAPE) of prediction of property prices. For comparison purposes, models' predictive capability was assessed using a ±10% margin of error criterion (interpreted as "reasonably accurate" prediction) and margin of errors greater than ±10% (interpreted as "inaccurate" prediction).

By looking at the MAPE, Models I and II can be considered to be marginally better than Models III and IV. In other words, except for Model IV, both categories of GIS-MRA-based and the traditional MRA models have predicted equally well. Nevertheless, by looking at the proportion of the residential properties being predicted below ±10% margin of errors, Models I and II have a higher proportion (43-57%) of "reasonably accurate" prediction compared to Models III and IV (21-29%). In the same way, models I and II have smaller proportion (43-57%) of "less accurate" predictions compared to models III and IV (71-79%). This has given some evidence that GIS-MRA based LVRS approach to modeling locational factors has led to some improvement in the prediction of residential property values.

Model application
The GIS-MRA-generated LVRS models can be employed together with location maps such as those shown in Figures 2, 3, 4 and 5 to create a local adjustment table (LAT) that can be used to assess the market value of unsold residential properties in a particular area. LAT gives a general idea about the amount of adjustment that should be given to any particular unsold residential property located within the sample area.

Table 9: Example of application of local adjustment table for assessing property prices

Selected location	Recommended Local Value Adjustment Based on Regressed Price (%)			
	IDW technique		Kriging technique	
	Min	Max	Min	Max
Jalan Pasir Pelangi	+2.0	+4.0	+4.0	+6.0
Jalan Serampang: to the west of Jalan Kuning	-3.7	-9.0	-0.8	-2.2
: to the east of Jalan Kuning	-9.0	-14.5	-7.8	-13.0
: to the right of Jalan Serampang	-9.0	-14.5	-3.9	-5.9
: both side to the north of Jalan Sri Pelangi	+1.8	+3.7	+0.1	+2.0
: to the north of Jalan Hijau Muda 3	+7.2	+12.7	+2.0	+4.0
: to the east of Jalan Jerau	+7.2	+12.7	+2.0	+4.0
: Jalan Abiad	-9.1	-14.6	-2.0	-3.9
Jalan Stulang Darat : to the northwest area	+1.8	+3.7	+0.1	+2.0
: to the southwest area	+7.2	+12.7	+6.0	+8.0

Notes: + sign indicates that properties in a particular location need an upward price adjustment, while – sign indicates that the properties need downward price adjustment.

Table 9 shows an example of LAT for residential properties in the selected location in the study area, based on IDW-based and kriging-based LVRS. Instead of a single figure, a range of possible adjustment figures are given so that valuers can decide on, within the range, the figure that is deemed most appropriate before final predictions are made.

Table 10 shows an out-sample prediction of residential property prices in the study area. For illustrative purposes, only IDW-based LVRS technique is shown here. The predictions in the middle column used Model I in Table 7 and the IDW-based LVRS shown in Figure 2. The amount of local value adjustment that was 'deemed' most appropriate for each property was judgmentally derived by looking at the geographic position of a particular property.

On average, the prediction error using IDW-based LVRS technique in Table 10 was 7.41%, which conforms to the generally acceptable standard of ±10% of

prediction error in property valuation. Nonetheless, there was some caution in the predictions since a number of properties have indicated serious over-prediction or under-prediction such as those at No. 14, Jalan Jingga 2, Taman Pelangi (-132% prediction error); No. 194, Jalan Sutera, Taman Sentosa (-37% error); and No. 50, Jalan Songkit 8, Taman Sentosa (35% prediction error).

Table 10: Out-sample prediction of residential property prices in the study area

		Market Value		
Lot No.	Address	Actual Price (RM)	Predicted (IDW-based)	Prediction error (%)
19864	77, Jalan Sri Pelangi, Taman Pelangi	315, 000	267,964.9	14.93
19833	84, Jalan Sri Pelangi, Taman Pelangi	300, 000	224,052.4	25.326
19525	95, Jalan Kelabu, Taman Pelangi	170, 000	143,323.3	15.69
18165	16, Jalan Maju 4, Taman Pelangi	268, 000	261,256.1	2.516
19778	32, Jalan Hijau Muda 5, Taman Pelangi	230, 000	250,248.6	-8.80
17955	6, Jalan Abiad 2, Taman Pelangi	238, 000	217,994.3	8.41
18890	14, Jalan Jingga 2, Taman Pelangi	140, 000	324,323.8	-131.66
18213	16, Jalan Maju 3, Taman Pelangi	260, 000	302,880.9	-16.49
18550	18, Jalan Kuning Muda 5, Taman Pelangi	370, 000	327,902.2	11.38
20209	169, Jalan Biru Muda, Taman Pelangi	288, 000	283,192.6	1.70
11944	14, Jalan Nila 5, Taman Pelangi	275, 000	192,425.5	30.03
9788	23, Jalan Songkit 3, Taman Sentosa	270, 000	220,279.4	18.42
18563	3, Jalan Kuning Muda 5, Taman Pelangi	362, 000	331,530.8	8.42
19692	37, Jalan Hijau Muda 9, Taman Pelangi	240, 000	176,172.8	26.60
9475	27, Jalan Songkit 2, Taman Sentosa	160, 000	178,400.6	-11.5
9912	51, Jalan Songkit 8, Taman Sentosa	266, 000	237,887.0	10.60
15757	28, Jalan Jerau 4, Taman Pelangi	205, 000	249,847.7	-21.88
9515	50, Jalan Songkit 2, Taman Sentosa	250, 000	162,547.4	34.98
8117	137, Jalan Keris, Taman Sri Tebrau	168, 000	183,931.0	-9.48
11156	194, Jalan Sutera, Taman Sentosa	175, 000	239,142.6	-36.65
18490	18, Jalan Kuning Muda 1, Taman Pelangi	360, 000	267,155.0	25.79

Implications of LVRS techniques on property valuation
If valuer's main purpose of modeling is value prediction, then the LVRS serves as a source of generic variables that represent various locational factors influencing property values and it avoids the difficulty of trying to specifically identify these factors at a particular site. The LVRS, superimposed on a location map, conveniently aids valuers in pinpointing under- or over-prediction of property values on the ground and, thus, in making appropriate adjustments at known addresses or geo-referenced positions thereof. In particular, the local adjustment table (LAT) is a useful value adjustment tool that is justified based on the LVRS, whereby it can be updated from time to time according to the dynamics of local property market.

However, if specification and estimation of influence of locational factors are to be particularly envisaged, one still have to identify these factors within a particular geographic area. Further, an auxiliary model that regresses the levels of spatial residuals of property values against the influencing locational factors may need to be estimated. Certainly, this option goes back to the traditional discrete location modeling. It may have both predictive and explanatory advantages but, at the same time, it is more cumbersome while maintaining the "old problems" of discrete location modeling.

The results from this study should be able to encourage further research in various aspects of spatial modelling in property valuation. Future research may focus on the refinement of GIS-MRA LVRS techniques using a more complete model specification. The effect of sample structure on the resulting LVRS can also be examined further. Sample structure includes elements like the number, distribution, and type (single- or double-storey property value, per square foot or per unit property value) of observation points.

Further investigation on the likely spatial elements that give rise to bumps or ridges and potholes or trenches on the LVRS can be further explored in future studies. By doing so, some factors can be statistically explained with certain locational detail in the local market. As a matter of fact, the practice in the property valuation is that people naturally want to discover the specific locational factors that have actually influenced property values in the local market. This allows back-to-back approaches between GIS-MRA LVRS techniques, the traditional MRA techniques, and the traditional comparison method to be adopted in assessing property prices in a particular area.

CONCLUSION

This paper has disclosed the usefulness of LVRS in creating locational value adjustment factor in the prediction of property values by taking residential properties as a study case. The generated surface has enabled a more visualized representation of locational influence on property values and has enabled such influence to be captured at any geographic points across a particular area.

The GIS-MRA-generated LVRS has managed to improve model's overall statistical quality. On the basis of MAPE, equally good predictions were obtained from the model which used IDW-based LVRS (Model I) and that which used CBD (Model III) as a locational factor. However, there was higher proportion of "reasonably accurate" predictions from the models incorporating LVRS compared to the traditional regression models.

The local adjustment table (LAT), which is the product of LVRS was a useful tool for adjusting the market value of unsold residential properties in a particular area,

whereby it gives a general estimate of the amount of adjustment that should be given to a particular unsold property on a particular site within the sample area.

REFERENCES

Adair, A. S., Berry, J. N. and McGeal, W. S. (1996). Hedonic Modelling, Housing Submarkets and Residential Valuation. *Journal of Property Research*. 13(1): 67-83.

Adair, A. S. and McGreal, W.S. (1988). The Application of Multiple Regression Analysis in Property Valuation. *Journal of Valuation*. 6(1): 57-67.

Anon (2003). Standard on Automated Valuation Models (AVMS). *Assessment Journal*. 10(4): 109-155.

Azhari, H. (1993). The Measurement of Location: A Case Study of Eighteen Housing Schemes. *Asia Pacific Real Estate Society (APRES) Conferrence*, November 6th-7th. Langkawi Island, Malaysia.

Azhari, H. and Ghazali, M. H. (2001). The Construction of Land value Maps Using GIS and MRA: A Case Study of Residential Properties in Johor Bahru. Faculty of Surveying, Universiti Teknologi Malaysia.

Box, G.E.P. and Cox, D.R. (1964). An Analysis of Transformation. *Journal of the Royal Statistical Society*. 26(2): Series B: 211-251.

Breusch, T.S. and Pagan, A.R. (1979). A Simple Test for Heteroscedasticity and Random Coefficient Variation. *Econometrica*. 47(5): 1287-1294.

Burrough, P. A. and McDonnell R. A. (1998). *Principles of Geographical Information Systems*. New York: Oxford University Press. Chapter 6.
Can, A. (1990). The Measurement of Neighbourhood Dynamics in Urban House Price. *Economic Geography*. 66(3): 54-72.

Dubin, R. A. and Sung, C. (1987). Spatial Variation in the Price of Housing: Rent Grandients in Non-monocentric Cities. *Urban Studies*. 24: 193-204.

Durbin, J. and Watson, G.S. (1951). Testing for Serial Autocorrelation in Least Squares Regression. *Biometrika*. 38: 159-178.

Evans, A.W. (1973). *The Economics of Residential Location*. London: Macmillan.

Farrar, D.E. and Glauber, R.R. (1967). Multicollinearity in Regression Analysis. *Review of Economics and Statistics*. 49: 92-107.

Ferri, M. G. (1977). An Application of Hedonic Indexing Methods to Monthly Changes in Housing Prices: 1965-1975. *AREUEA Journal*. 5: 455-462.

Figueroa, R. A. (1999). Modelling the Value of Location in Regina Using GIS and Spatial Autocorrelation Statistics. *Assessment Journal*. 6(6): 29-37.

Fletcher, P., Gallimore, P., and Mangan, J. (2000). The Modeling of Housing Submarkets. *Journal of Property Management.* 18(5): 366-374.

Frew,J. and Wilson, B. (2002). Estimating the Connection Betweeen Location and Property Value. *Journal of Real Estate Practice and Education.* 5(1): 17.

Gallimore, P., Fletcher, M. and Carter, M. (1996). Modelling the Influence of Location on Value. *Journal of Property Valuation and Investment*. 14(1): 6-19.

Greaves, M. (1984). The Determinants of Residential Values: The Hierarchical and Statistical Approaches. *Journal of Valuation*. 3: 5-23.

Hamid, Abdul, bin Mar Iman and Mohamad Ghazali b. Hashim (1991). Application of Geographic Information System for Estimating Agricultural Land Value. *National Seminar on Research and Development in Property Management and Valuation,* Universiti Teknologi Malaysia, 8-9 July.

Hamid, Abdul, bin Mar Iman (2001). *Incorporating Geographic Information System in the Prediction of Farm Property Values*. Unpublished Ph.D thesis. Lincoln University, New Zealand.

Hamid, Abdul, bin Mar Iman (2003). Price-Contour Based Spatial Dummy Variables for Segmenting Market in the GIS-Hedonic Modelling of Residential Property Prices. *International Conference on Urban Development and Management*, Langkawi, Malaysia, 7-9 July.

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

Hickman, E. P., Gaines, J. P. and Ingram, F. J. (1984). The Influence of Neighbourhood Quality on Residential Property Values. *The Real Estate Appraiser and Analyst*. 50(2): 36-42.

Isaaks, E. H. and Srivastava, R. M. (1989). *An Introduction to Applied Geostatistics*. New York: Oxford University Press.

Jack, E. (1989). Incorporating Location into Computer-Assisted Valuation. *Property Tax Journal*. 8(2): 151-170.

Jud, G.D. and Winkler, D.T. (1991). Location and Amenities in Determining Apartment Rents: An Integer Programming Approach. *The Appraisal Journal* April: 266-275.

Kain, J. and Quigley, J. (1970). Measuring the Value of Housing Quality. *Journal of the American Statistical Association*. 45: 532-548.

Kennedy, P. (1992). *A Guide to Econometrics*, 3rd. ed. Oxford (UK) and Cambridge (USA): Basil Blackwell Ltd.

Mackmin, D. (1989). *The Valuation and Sale of Residential Property*. London: Routledge.

McCluskey, W. J., Deddis, W. G., 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*. 18(2): 162-176.

McGuirk, A.M. et al. (1993). Misspecification Testing: A Comprehensive Approach. *American Journal of Agricultural Economics*, November: 1044-1055.

Muth, R. F. (1960). *Cities and Housing*. Chicago, II: University of Chicago Press.

Oliver, M. A. and Webster, R. (1990). Kriging: A Method of Interpolation for Geographical Information Systems. *International Journal of Geographical Information Systems*. 4(3): 313-332.

Pearson, T.D. (1991). Location! Location! Location! What is Location!. *The Appraisal Journal*, LIX(1): 7-20.

Pendleton, E. C. (1965). Statistical Inference in Appraisal and Assessment Procedures. *The Appraisal Journal*. 37: 501-512.

Pollakowski, H. O. (1982). *Urban Housing Markets and Residential Location*. D.C. Lexington, MA: Health and Company.

Ramsey, J.B. (1969). Tests for Specification Errors in Classical Linear Least-Squares Regression Analysis. *Journal of the Royal Statistical Society*, Series B (Methodological). 31: 350-371.

Reichert, A.K. and Moore, J.S. (1986). Analysing the Nature of Multicollinearity in Real Estate Appraisal Models, *The Real Estate Appraiser and Analyst*. 52(3): 34-38, 55.

Richardson, H. W. (1971). *Urban Economics*. Penguin: Harmondsworth.

Ridker, R. and Henning, J. (1967). The Determinants of Residential Property Values with Special Reference to Air Pollution. *Review of Economics and Statistics*. 49: 246-257.

Rodriguez, M. and Sirmans, C.F. (1994). Quantifying the Value of View in Single-Family Housing Markets. *The Appraisal Journal*. 62(4): 600-603.

Royle, A., Clark, I., Brooker, P. I., Parker, H., Journel, A., Rendu, J. M., Sandeful, R. and Mousset-Jones, P. (1980). *Geostatistics*. New York: McGraw-Hill. Chapters 1, 2, 3 and 4.

So, H. M., Tse, R.Y. C. and Ganesan, S. (1997). Estimation the Influence of Transport on House Prices: Evidence from Hong Kong. *Journal of Property Valuation & Investment*. 15(1): 40-47.

Stegman, M. A. (1969). Accessibility Models and Residential Location. *Journal of American Institute of Planners*. 35: 22-29.

Ward, R., Guilford, J., Jones, B., et al. (2002). Piecing Together Location: Three Studies By The Lucas County Research and Development Staff. *Assessment Journal*. 9(5): 15–49.

Watkins, C. (1999). Property Valuation and Structure of Urban Housing Markets. *Journal of Property Investment and Finance*. 17: 157-175.

Wendt, P.F. (1957). Theory of Urban Land Values. *Land Economics*. 33(3): 228-240.

Wyatt, P. J. (1997). The Development of Property Information System for Real Estate Valuation. *International Journal of Geographical Information Systems*. 11(5): 435-450.

Zan, Y. (2001). An Application of the Hedonic Price Model with Uncertain Attribute – The Case of the People's Republic of China. *Property Management*. 19(1): 50-63.