MODELLING INTERACTION OF LOCATION INFLUENCE WITH RENTAL VALUE ON COMMERCIAL PROPERTIES USING SPATIAL STATISTIC TECHNIQUES

Oliver Valentine Eboy[1]
Ibrahim bin Sipan[2]

[1]School of Social Sciences,
Universiti Malaysia Sabah, Locked Bag 2073, 88999 Kota Kinabalu
Sabah, Malaysia
oliver@ums.edu.my

[2]Faculty of Geoinformation Science and Engineering,
Universiti Teknologi Malaysia, 81310 Skudai,
Johor, Malaysia

Abstract

The Multiple Regression Analysis (MRA) approach has been widely used to identify the location factor and perform value prediction of the property. However, this approach is subjective, choosing which measures to include in the model or in defining the measures itself. A possible alternative to this approach is to integrate the MRA with spatial statistic techniques to produce a better result. In this paper, the spatial relationship between the rental value of the commercial properties which is the shop house and location influence is explored using geographically weighted regression (GWR). GWR attempts to capture spatial variation by calibrating a multiple regression model fitted at each shop house in localities, weighting the location factors from the subject shop house which needs to be determined. GWR produces a set of parameter estimates and model statistics for the shop houses in the study area of Johor Bahru, Malaysia. It is evident that the GWR model provides useful information on rental value caused by surrounding factors. The GWR model was also compared with the traditional model, which is the ordinary least squares (OLS) model, to show the differences of the two models. The parameter estimates and model statistics of the GWR and OLS model were then mapped using visualisation tools, such as Geographic Information system (GIS). Consequently, the influence of site location, bank facilities, shopping complexes and others can be evaluated, tested, modelled, and readily visualised. In this study, the result shows that the bank provides a higher significant spatial variation towards the rental value of the shop house than the other influence factors. GWR is a useful tool that provides much more information on spatial relationships to assist in model development and further our understanding of spatial processes.

Keywords : Geographical Weighted Regression (GWR), Ordinary Least Squares (OLS), Geographic Information System (GIS), location, rental value, shop house.

1.0 Introduction

Mostly, ordinary least squares regression (OLS) has been used to estimate the value of the property based on the location influence factor (Scott, 1988; Wyatt, 1997; So et al., 1997). Although some researchers have improved the regression models, these still lack explanatory power particularly on why property value in certain location is over-valued or under-valued based on neighbourhood factor (Gallimore et al.,1996; Theriault et al., 2003).

In this paper, we introduce the GWR to determine the rental value of the commercial property within Johor Bahru. We applied this local modelling technique to a relatively simple linear regression model of shop houses. The improvement of GWR over OLS in model fitting is tested and compared. Then, the location influence, which gives the most significant effect on spatial variation, is illustrated using ArcGIS and WinSurfers software.

2.0 Influence of Location Factor on the Shop House Value

Location is an important factor in determining the values of properties. The influence of location may be in terms of accessibility to shopping complex, parks, petrol station, public facilities and work place; road traffic, noise and business; neighbourhood amenities; safety issues such as level of crime and security; to mention a few (Kahn, 1963; Gallimore et al., 1996; Rozana, 2004).

However, for commercial property, the location factors may be different to that of residential. This is illustrated by Wyatt et al. (2003), location factor that influence commercial property are accessibility to the market place, proximity to suppliers of raw materials and important nodes such as railway stations, car parks and open spaces. These results show that easy access to parks significantly influences residential property but not commercial property. Meanwhile, parking spaces greatly influence the value of commercial property but rarely for residential property.

Providing variety types of location factors will improve greatly the model to estimate property value (be it price or rental), rather than just putting it as only one factor in the property valuation model. This however, is not easy, as modelling locational factors has proved difficult because of the wide range of spatially defined attributes, which may affect value and maybe at only a particular time and location. This has been discussed by many researchers to determine the best locational factors, including their measurements, and how they influence property values.

Some argued that house prices are determined not only by accessibility but also by the neighbourhood quality or the environmental attributes of the location (Stegman, 1969; Pollakowski, 1982; and McCluskey et al., 2000). There are also researchers who employ measurements in measuring location such as using the type of transport, time taken per trip, and transportation cost such as Theriault et al. (2003). Moreover, Gallimore et al. (1996) had stated out that locational influences on the property value may arise from any number of sources and with this, the authors believe that different types of relevant location factors is necessary to be examined if an accurate property valuation model need to be produced.

3.0 GWR and Regression Analysis

Geographical Weighted Regression (GWR) is a modelling technique for local spatial analysis. This technique was originally proposed by Brunsdon et al., (1996; 1998).

This technique allows local, as opposed to global spatial models to be calibrated and interesting variations in relationships to be measured and mapped. Stewart Fotheringham, Martin Charlton and Chris Brunsdon of the Spatial Analysis Research Group and Department of Geography at the University of Newcastle, UK are the pioneers in this field.

In this study, we chose the normal linear regression model (i.e. ordinary least squares

(OLS)) to explore spatial relationship between rental values with location factor. The formula can be stated as follows:

$$RV = B_0 + B_1X_1 + B_2X_2.....B_nX_n$$

Where, RV is the estimated rental value for each shop house which is calculated as the sum of B_0 (constant) and location influence variables (B_1X_1 ..... B_nX_n).

However, in GWR, the inclusion of the data coordinate, which is the longitude and latitude, has rewritten the original model as follows:

$$RV = B_0(u_i,v_i) + B_1(u_i,v_i)\,X_1 + B_2(u_i,v_i)\,X_2.....\,B_n(u_i,v_i)\,X_n$$

where (u_i,v_i) denotes the coordinate of the *i*th point in space and $B_n(u_i,v_i)$ is a realization of the continuous function $B_n(u,v)$ at point *i*.

The (u,v)s are typically the locations at which data are collected. This allows a separate estimate of the parameters to be made at each data point. The resulting parameter estimates can then be mapped. Various diagnostic measures are also available such as the local standard errors, local measures of influence, and a local goodness of fit. If the (u,v)s are at the mesh points of a regular grid, then the spatial variation in the parameter estimates can be examined as a pseudo-surface. The parameters may be tested for 'significant' spatial variation. The outputs from the software provide a convenient linkage to mapping software which is ArcGIS (National Centre for Geocomputation (NCG), 2006).

Based on Fotheringham et. al., (2002), the GWR weighting is actually based on geographically weighted mean which is the starting point for thinking about geographically weighted statistics. The formula for arithmetic mean is:

$$\overline{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

This is simply the sum of the values making up a batch of numbers divided by the size of the batch. More generally, we can consider a weighted mean:

$$\overline{x} = \frac{\sum_{i=1}^{n} w_i x_i}{\sum_{i=1}^{n} w_i}$$

where the w_is are the weights. Here we multiply each value by its weight, and divide by the sum of the weights. In the case that each observation has a weight of unity, then this formula and the one above are equivalent.

In many cases the weights are integers, but they may also be non-integer numbers. In this case, we can use weights generated from the same geographical weighting scheme that we have used for geographically weighted regression. Rather than being a whole-map statistic, a geographically weighted mean is available at a particular location, say, u. Thus the formula for a geographically weighted mean at location u is:

$$\overline{x}(u) = \frac{\sum_{i=1}^{n} w(u)_i x_i}{\sum_{i=1}^{n} w(u)_i}$$

$W(u)_i$ is the geographical weight of the ith observation relative to the location u. The weights may be generated using a fixed radius or an adaptive kernel.

By analogy the local geographically weighted variance is

$$^2(u) = \frac{\sum_{i=1}^{n} w(u)_i (x_i - \overline{x}(u))^2}{\sum_{i=1}^{n} w(u)_i}$$

and the locally weighted standard deviation is the square root of this. Notice that the mean here is the geographical mean around point u and NOT the global mean of the data.

The GWR software currently supplied (GWR3.0) allows the user to compute geographically weighted means, variances, and standard deviations for a set if input data, and for either a fixed or an adaptive kernel.

4.0 Methodology and Data

Ninety observations of ground floor shop house rental values for the year 2004 through 2005 were collected from 29 localities within the Johor Bahru area as shown in Figure 1. Meanwhile, seventeen location factors such as site location, road type, road direction, car park, school, university, central business district, industrial area, construction site, shopping centre, sport centre, recreation centre, office area, bank, bus or taxi station and view of surrounding were collected for each of the 90 observations. Correlation analysis was carried out to identify the multicollinearity between the 17 independent variables of location factors. The result indicates that the factors can be considered independent.

The attribute data for the location factors is coded using "Dummy" or variable's indicator (e.g. 1 or 0) that depicts the impact of location on the properties. For example, the code "1" represents the property that receives the impact of the location factor, otherwise, code "0" will be given. So, the more the locational factors near the shop house, the more the influence it will receive. This means that no measurement tools such as buffer or network analysis available in GIS was used to construct the spatial variables involved in this study.

The tests used in this study are such as F test to measure the overall goodness of fit of the model, the spatial autocorrelation test to determine the spatial variation of the model and finally, Monte Carlo test was conducted to find out the significance of the spatial variation in each local parameter estimate from the model.

5.0 Results and Discussion

Table 1 compares the results from the OLS and the GWR. The estimation is been done using a computer software program which is SPSS 14.0 for OLS analysis and GWR 3.0 for GWR analysis. Detail information on both of this software is available at the website www.spss.com and www.ncl.ac.uk/geography/GWR.

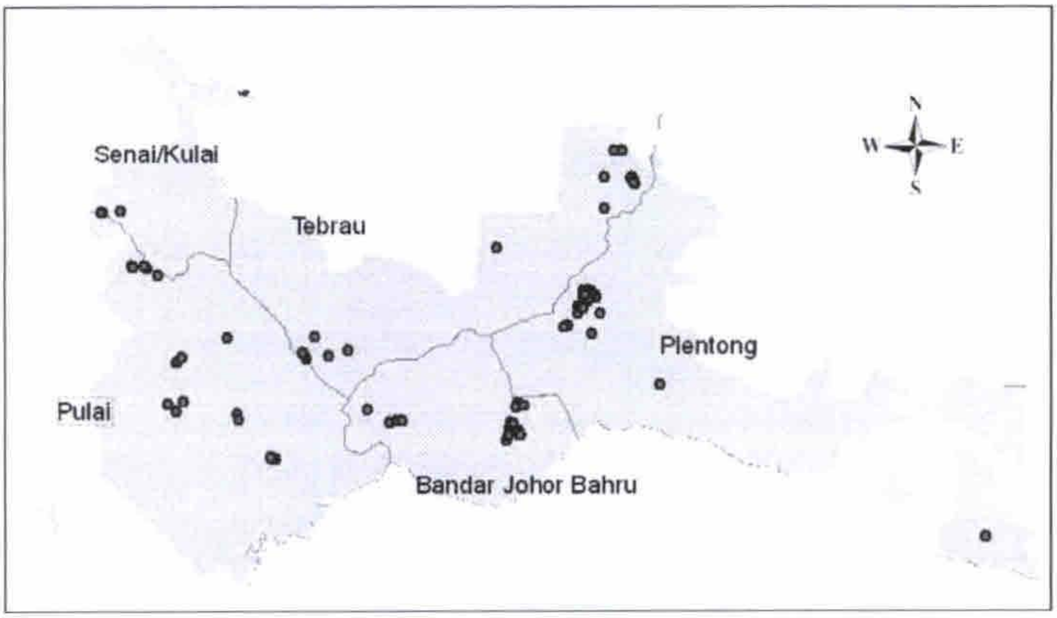

Figure 1. Area of Johor Bahru with the location of the 90 observations data

Table 1. Comparison of the statistics of OLS and GWR models

	OLS	GWR
Residual sum of squares (RSS)	18740127.350	13910118.641
Coefficient of Determination (R^2)	0.596	0.700
Adjusted r-square	0.494	0.564

Based on the above results, it can be said that the GWR is better than the OLS model, in which the coefficient (R^2) has increased from 0.596 to 0.700 and the RSS has decreased from 18740127.350 to 13910118.641. The R^2 value can be considered high and acceptable based on Tang (1997) and Carl, D. et. al. (1994) who suggest R^2 of 0.600 and above is acceptable.

Table 2 shows the results of an ANOVA in which the OLS model is compared with the GWR model. The ANOVA tests the null hypothesis that the GWR model represents no improvement over a global model.

The F test was conducted to measure the contribution of each independent variable, measuring the overall goodness of fit or correctness of the model when all the variables are considered simultaneously. The rule is that the higher the value of computed F value, the better the model will be (Gujarati, 1995). The test reveals that the value of computed F (2.22) is higher than the value of theoretical or critical F (1.75) based on the F statistical table. The F test suggests that the GWR model is a significant improvement on the global model for the Johor Bahru data.

Moran's I test indicates that positive spatial autocorrelation occurs if residuals of the same sign cluster together while negative spatial autocorrelation occurs if residuals of different signs cluster together (Lee and Wong, 2001). Figure 2 shows how the spatial variation patterns are formed.

Table 2. ANOVA of the GWR and OLS model

Source	SS	DF	F
OLS Residuals	18740127.4	18	
GWR Improvement	4830008.5	9.74	
GWR Residuals	13910118.6	62.26	2.2183

*SS = Residual Sum of Squares (RSS)
DF = Degree of Freedom

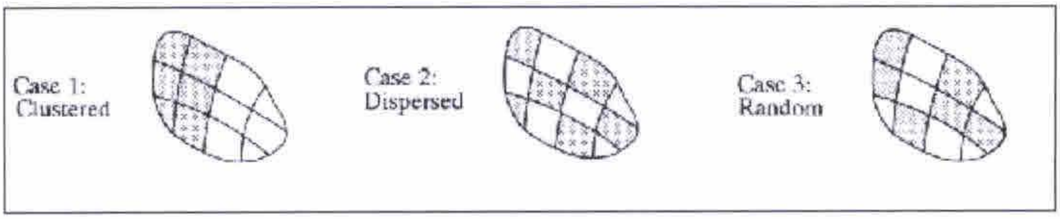

Figure 2. Spatial variation form (clustered, random or dispersed)

By using the Moran's I test in ArcGIS 9.2 software, we identified the spatial autocorrelation based on the residual from both of the model. The result is shows in Table 3.

Table 3. Moran's I spatial autocorrelation for OLS and GWR

	OLS	GWR
Moran's Index	-0.04	-0.07

Note: The Moran's I spatial autocorrelation is based on the Inverse Distance spatial relationship with Euclidean Distance method and spatial weights are standardized by row.

Based on the above results, the OLS model shows that the spatial distribution is not evident as there are no obvious patterns, clustered or dispersed, to the residuals which appear random across the study area. This means, there is no significant spatial autocorrelation from the OLS model. The GWR model however, shows that the pattern is somewhat dispersed, which could be due to random chance. This finding is actually taken from the analysis output's script in ArcGIS 9.2.

The differences of spatial variation between GWR and OLS can be seen if we compare maps of the residuals from the two models. Figure 3 and 4 show the GIS visualisation of the residuals for OLS and GWR in 3D presentation using WinSurfers software.

From the maps, the green coloured pattern shows the minimal residual between -100 to 100, which can be acceptable as the differences are not high. The high negative (yellow) and high positive (orange) residuals however, show that the rental value is under predicted and over predicted respectively. The spatial variation from GWR illustrates a much more dispersed pattern and this may be caused by the variety of location factor.

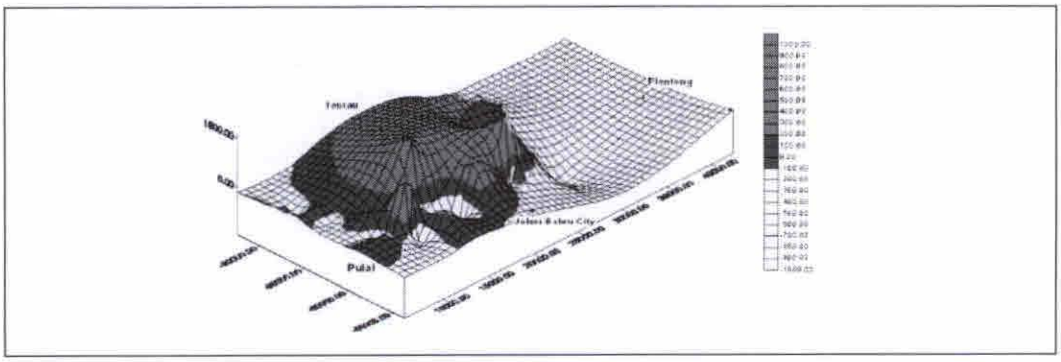

Figure 3. 3D residuals visualisation from the OLS model

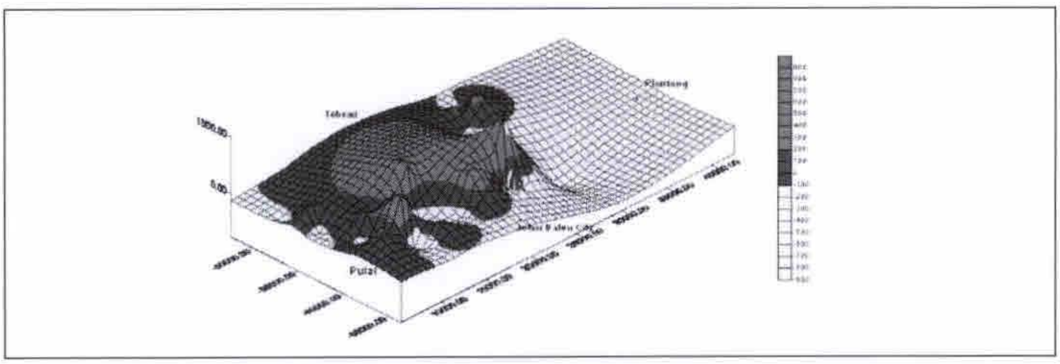

Figure 4. 3D residuals visualisation from the GWR model

Finally, a Monte Carlo test is conducted to determine the significance of the spatial variation in each of the local estimates from the model. The result of Monte Carlo test obtained by using the GWR software is as in Table 4.

Table 4. Test for spatial variability of the GWR parameters

Parameter	P-Value	Significant
Bank	0.00000	Significant at .1% level
Site Location	0.05000	Significant at 5% level
Intercept	0.33000	Not Significant
Road Type	0.38000	
Road Direction	0.89000	
Car Park	0.15000	
School	0.85000	
University	0.95000	
Centre Business District	0.38000	
Industry	0.29000	
Construction	0.71000	
Shopping Centre	0.56000	
Sport Centre	0.52000	
Recreation Centre	0.72000	
Office	0.98000	
Post Office	0.67000	
Bus/Taxi Station	0.27000	
Surrounding View	0.13000	

Note: Test based on the Monte Carlo significance test procedure due to hope (1968)

The results of a Monte Carlo test on the local estimates indicate that the GWR model fits significantly (= 0.01) for Bank. This means that there is significant spatial variation in the local parameter estimates for the variable Bank. The spatial variation for the other variables is either of no or low significance and in each case there is a reasonably high probability that the variation has occurred by chance. Based on this test, we can concentrate on the variable Bank for which the local estimates exhibit significant spatial non-stationarity. It is interesting to note that these results reinforce the conclusions reached above with the informal examination of local parameter variation for the variable Bank. Figure 5 demonstrates the results of the local parameter estimates using GIS interpolation for 3D visualisation purposes.

The GIS visualisation above shows that the value is high in the west side (Pulai) of the study area in which the influence of the bank is high. This is probably due to the bank's high influence on the community centre as the number of banks over there is still average which is 22 banks, while, in the CBD (Johor Bahru City and Tebrau) and industrial (Plentong) area, there is less influence of the bank as the number of banks in that area is already high which is about 32 and 31 banks respectively.

6.0 Conclusion

In summary, this study has shown that GWR can produce a set of local estimates for the model coefficients at each point in the defined

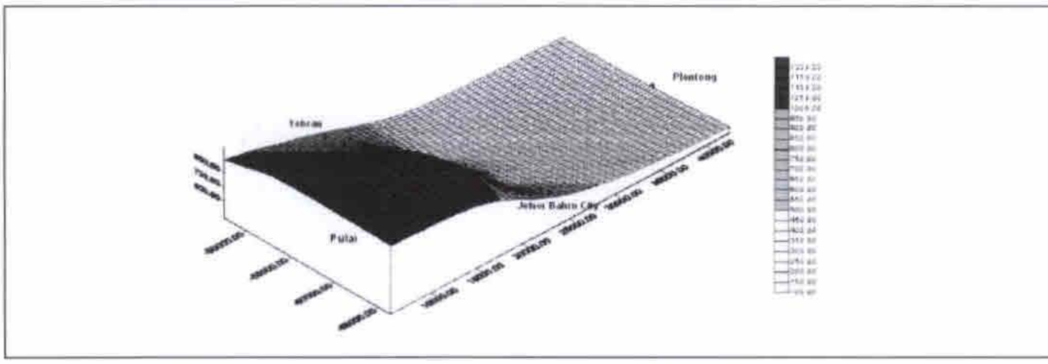

Figure 5. 3D visualisation of local estimates of the bank parameter

geographic area. These model coefficients can be visualized using tools such as GIS, which can highlight the sub-areas within the localities where rental value is higher or lower than in other sub-areas. This indicates that GIS enables researchers to use GWR to explore the spatial variation of the relationships between variables under investigation. Research can subsequently incorporate the identified geographic patterns into a formal modeling procedure.

This study however, was applied without using any measurement tools of GIS analysis such as buffer or network analysis to determine the location influence of the shop houses. Other than that, the weight of the location factor also needs to be determined as we believe that its incorporation can improve the model. Furthermore, the ninety observations data may seem too little and may affect the result especially for the non-significant parameters estimated in Monte Carlo test. Hence, the observations data should be increased in future study to produce a much more accurate result.

REFERENCES

Journal

Anselin, L., And D.A. Griffith. (1988). *Do Spatial Effects Really Matter in Regression Analysis.* Paper Regression Science Assoc. Vol 65.

Brunsdon C, Fotheringham A S, Charlton M. (1996). *"Geographically Weighted Regression: A Method for Exploring Spatial Nonstationarity"* Geographical Analysis Vol. 28 pp281-298.

Brunsdon C, Fotheringham A S, Charlton M. (1998). *Geographically Weighted Regression: Modelling Spatial Nonstationarity.* The Statistician vol. 47 pp431-443.

Carl, D.; Katz, L F; Kruegar, A.B; Newmark, D.; Wascher, W. (1994). Comment on David Newmark and William Wascher: *Employment Effects of Minimum and Subminimum Wages: Panel Data on State Minimum Wage Laws. Ithaca: Industrial and Labour Relation* Review. Vol. 47, No. 3.

Fotheringham, A.S. (1997). *Trends in Quantitative methods I: Stressing the Local.* Programme Human Geography Vol. 21.

Fotheringham, A.S., And C. Brunsdon. (1999). *Local Forms of Spatial Analysis.* Geographical Analysis Vol. 31.

Gallimore, P., Fletcher, M. and Carter, M. (1996). "Modelling The Influence of Location on Value. Journal of Property Valuation and Investment". Vol. 14 No.1.

Hope, A. C. A. (1968), "A Simplified Monte Carlo Significance Test Procedure", Journal Of The Royal Statistical Society Series B Vol. 30.

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.

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. Vol. 15 No.1.

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

Tang, T. and Li, Ping (1997). *Teaching Evaluation at a Public Institution of Higher Education: Factors Related to the Overall Teaching Effectiveness.* Washington: Public Personnel Management, Vol.2. No.3.

Theriault, Marius; Rosiers, FrancËois Des; Villeneuve, Paul; Kestens, Yan (2003). *Modelling Interactions of Location with Specific Value of Housing Attributes.* Property Management, Vol. 21 No. 1.

Wyatt, P. J. (1997). "The Development of Property Information System for Real Estate Valuation". International Journal of Geographical Information Systems. Vol. 11 No.5.

Book

Anselin, L. (1988). *Spatial Econometrics: Methods and Models.* Kluwer Academic Publishers. Dordrecht, The Netherlands.

Anselin, L. (1990). *What is Special About Spatial Data? Alternative Perspectives on Spatial Data Analysis.* P. 63-77 in Spatial statistics: Past, present, and future, Griffith, D. (ed.). Inst, of Math. Geog., Ann Arbor, MI.

Fotheringham, A.S., C. Brunsdon, And M. Charlton. (2002). *Geographically Weighted Regression: The Analysis of Spatially Varying Relationships.* John Wiley & Sons, New York, NY.

Fotheringham, A.S., C. Brunsdon, And M. Charlton. (2000). *Quantitative Geography: Perspectives on Spatial Data Analysis.* SAGE Publications.

Gujarati, D.N. (1995). *Basic Econometrics.* Third Edition. Singapore: McGraw-Hill International Edition.

Kahn, Sanders A. (1963). *Real Estate Appraisal and Investment.* New York, Ronald Press Co.

Lee, J. And Wong, D.W.S. (2001). *Statistical Analysis with ArcView GIS.* New York: John Wiley & Sons Inc.

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

Scott, I., (1988). *A Knowledge Based Approach to Computer-Assisted Mortgage Valuation of Residential Property.* Pontypridd: University of Glamorgan.

Wyatt, P., and Ralphs, M. (2003). *GIS in Land and Property Management,* Spon Press.

Thesis

Rozana Zakaria (2004). *Kesan Letakan dan Ciri-Ciri Pemandangan Terhadap Nilai Sewa Harta Tanah Komersial. (Kajian kes: Rumah Kedai Dalam Kawasan Terpilih di Johor Bahru).* Universiti Teknologi Malaysia. Bachelor of science thesis.

Website

National Centre for Geocomputation (2006). *What is GWR?.* National University of Ireland Maynooth. (http://ncg.nuim.ie/ncg/GWR/whatis.htm).