Determining Location Influence for Shop Houses Rental Value Using Geographical Weighted Regression (GWR)

Oliver Valentine Eboy [1], **Ibrahim bin Sipan** [2], **Buang bin Alias** [2]

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

[2] *Faculty of Geoinformation Science and Engineering,*
University Technology Malaysia, 81310 Skudai,
Johor, Malaysia

Abstract

This paper examines the spatial relationship between the rental value of shop house and the influence of location using Geographically Weighted Regression (GWR). GWR attempts to capture spatial variation by calibrating a multiple regression model fitted at each site of shop house, weighting the locational factors from the subject shop house. GWR produces a set of parameter estimates and statistics for the shop houses in the study area. It is evident that the GWR model provides useful information on rental value caused by the surrounding factors. The GWR model is also compared with the traditional ordinary least squares (OLS) model to show the differences of the two models. The parameter estimates and statistics of the GWR and OLS models are then mapped using the Geographic Information system (GIS). Consequently, the influence of site location, bank facilities, shopping complexes and other factors can be evaluated, tested, modelled, and readily visualised. The results show that the location of bank gives rise to a higher significant spatial variation in the rental value of shop house than other factors. It is concluded that, GWR is a useful tool that provides much more information on spatial relationships to assist in model development and in furthering our understanding of spatial processes.

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

1. Introduction

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

This paper introduces the Geographically Weighted Regression (GWR) to determine the rental value of commercial property in Johor Bahru. We apply this local modelling technique to estimate a linear regression model of shop houses. The improvement of GWR over OLS in model fitting is tested and compared. Then, the locational influence, which gives the most significant effect on spatial variation is illustrated using ArcGIS.

2. Influence of Location Factor on the Shop House Value

Location is an important factor determining property values. The influence of location can be expressed 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 influence of locational factors may be different from that of residential. This is illustrated by Wyatt and Ralphs

(2003), who discover locational factors that influence commercial property such as accessibility to the market place, proximity to suppliers of raw materials and important nodes such as railway stations, car parks and open spaces. The study shows 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 the residential property.

3. GWR and Regression Analysis

Geographically 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 variation 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 ordinary least squares (OLS) to explore the spatial relationship between rental values and locational factors. The formula can be stated as follows:

$$\boxed{RV = B_0 + B_1X_1 + B_2X_2 \ldots\ldots B_nX_n}$$

Where, RV is the estimated rental value for each shop house which is calculated as the sum of B0 (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:

$$\boxed{RV = B_0(u_i,v_i) + B_1(u_i,v_i)\,X_1 + B_2(u_i,v_i)\,X_2 \ldots\ldots B_n(u_i,v_i)\,X_n}$$

where (u_i,v_i) denotes the coordinate of the ith point in space and $B_n(u_i,v_i)$ is a realization of the continous 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).

4. Methodology and Data

Ninety observations of ground floor shop house rental values for the year 2004 through 2005 were collected from twenty-nine localities in Johor Bahru. Meanwhile, seventeen locational 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 ninety shop houses. Correlation analysis was carried out to identify the multicollinearity between the seventeen locational factors. The result indicates that the factors can be considered independent.

The attribute data for the locational factors are coded using "dummy" variables (e.g. 1 or 0) that depict the impact of location on properties involved. For example, code "1" represents property that receives the impact of a locational factor, otherwise, code "0" will be given. So, the more the locational factors be present near the shop house, the more the locational influence the property will receive. In this context, no measurement tools such as buffer or network analysis available in GIS was used to construct the spatial variables concerned.

The F test was used to measure the overall goodness of fit of the model, the spatial autocorrelation test for spatial variation of the model and finally, the Monte Carlo test for finding out the significance of spatial variation in each local parameter estimated.

5. Results and Discussion

Table 1 compares the results from the OLS and the GWR.

Table 1: **Comparison of the statistics of OLS and GWR models**

	OLS	**GWR**
Residual sum of squares (RSS)	18740127.350	13910118.641
Effective number of parameters	18.000	27.745
Sigma	510.176	472.692
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.

Table 2: **ANOVA of the GWR and OLS model**

Source	**SS**	**DF**	**MS**	**F**
OLS Residuals	18740127.4	18		
GWR Improvement	4830008.5	9.74	495643.1392	
GWR Residuals	13910118.6	62.26	223437.5289	2.2183

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

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.

As Lee and Wong (2001) indicate, positive spatial autocorrelation occurs if residuals of the same sign cluster together while negative spatial autocorrelation occurs if residuals of different signs cluster together. Figure 1 shows how the spatial variation patterns is formed based on the Moran's I test for spatial autocorrelation.

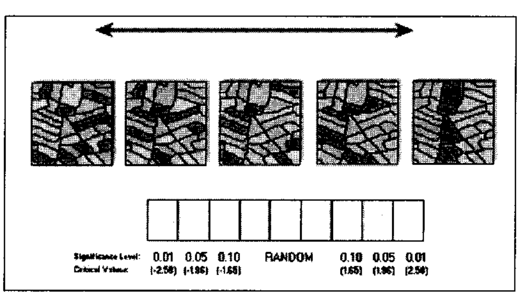

Fig. 1: Spatial variation form (dispersed or clustered) using Moran's I test for spatial autocorrelation

By using the Moran's I test in the ArcGIS 9 software, we identified the spatial autocorrelation based on the residual from both of the model. The result is shown in Table 3.

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

	OLS	GWR
Moran's Index	-0.04	-0.07
Expected Index	-0.011	-0.011
Variance	0.003	0.003
Z Score	-0.43	-1.12

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 GWR has further reduced the Moran's Index from -0.04 to -0.07. While 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, may be due to random chance.

The differences of spatial autocorrelation between GWR and OLS can be seen if we compare maps of the residuals from the two models. Figure 2 and 3 show the GIS visualisation of the residuals for OLS and GWR.

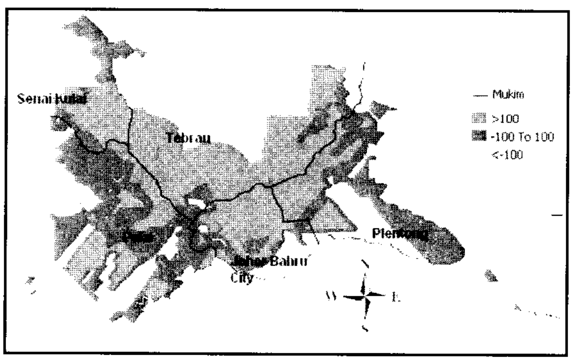

Fig. 2: Residuals visualisation from the OLS model

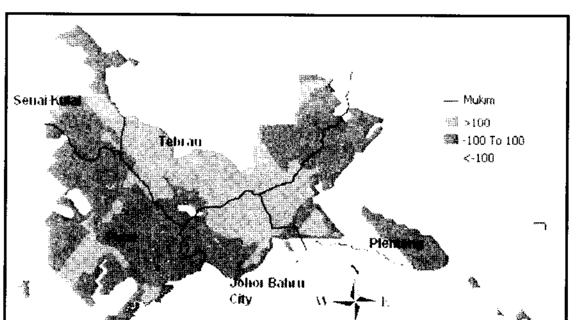

Fig. 3: Residuals visualisation from the GWR model

From the maps, the green coloured pattern shows the minimal residual between -100 to 100, which is acceptable since the differences are not too much. 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 patterns and this may be caused by the variety of locational factors.

It is important to test whether the local model (GWR) offers a statistically significant improvement over the global model (OLS). Thus, the Monte Carlo test was conducted to determine the significance of the spatial variation in each of the local estimates from the model. The result of Monte Carlo test based on the GWR software is shown in Table 4.

Table 4: **Test for spatial variability of the GWR parameters**

Parameter	P-Value
Intercept	0.33000 n/s
Site_Loc	0.05000 *
Rd_Type	0.38000 n/s
Rd_Direc	0.89000 n/s
Car_Park	0.15000 n/s
School	0.85000 n/s
Univrsty	0.95000 n/s
CBD	0.38000 n/s
Industry	0.29000 n/s
Construc	0.71000 n/s
Shp_Ctr	0.56000 n/s
Sprt_Ctr	0.52000 n/s
Rctrn_Ct	0.72000 n/s
Office	0.98000 n/s
Bank	0.00000 ***
Post_Off	0.67000 n/s
Bus_Taxi	0.27000 n/s
View	0.13000 n/s

*** = significant at .1% level
** = significant at 1% level
* = significant at 5% level
Note : Tests based on the Monte Carlo significance test procedure due to Hope (1968)

The results of Monte Carlo test on the local estimates indicate that the GWR model fits significantly ($\alpha = 0.01$) for Bank. This means that there is significant spatial variation in the local parameter estimates for the variable of Bank. The spatial variation for the other variables is either not or have low significant 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 of 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 of Bank. Figure 4 demonstrates the results of the local parameter estimates using GIS interpolation for visualisation purposes.

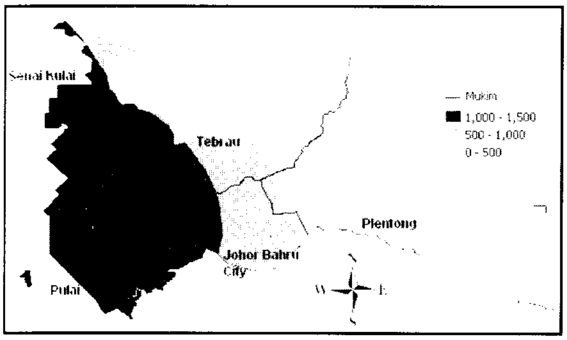

Fig. 4: Local estimates of the bank parameter

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 to the community centre as the number of banks over there is still average which is about 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. Conclusion

This study has shown that the GWR can produce a set of local estimates for the model coefficients at each point in the defined geographic area. These model coefficients can be visualised 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 modelling procedure.

This study however, has been carried out without using any measurement tools of GIS analysis such as buffer or network analysis to determine the influence of locational factors on shop houses. Furthermore, the weight of the locational factors also needs to be determined as we believe that its inclusion can improve the model.

References

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.*

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. Program Human Geography* Vol. 21.

Fotheringham, A.S., C. Brunsdon, And M. Charlton, (1998). *Geographically weighted regression: A natural evolution of the expansion method for spatial data analysis. Environ. Plan. A.*

Fotheringham, A.S., And C. Brunsdon. (1999).

Local forms of spatial analysis. Geographical Analysis Vol. 31.

Fotheringham, A.S., C. Brunsdon, And M. Charlton. (2000). *Quantitative geography: perspectives on spatial data analysis. SAGE Publications.*

Fotheringham, A.S., C. Brunsdon, And M. Charlton, (2002). *Geographically weighted regression: The analysis of spatially varying relationships. John Wiley & Sons, New York, NY.*

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.

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

Hope, A. C. A., (1968), *A simplified Monte Carlo significance test procedure, Journal Of The Royal Statistical Society Series B* Vol. 30.

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.*

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

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.

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

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.

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., and Ralphs, M., (2003). *GIS in land and property management, Spon Press.*

Wyatt, P. J. (1997). *The Development of Property Information System for Real Estate Valuation. International Journal of Geographical Information Systems.* Vol. 11 N