

# MOVING WINDOW REGRESSION (MWR) IN MASS APPRAISAL FOR PROPERTY RATING

<sup>1</sup>Taher Buyong, <sup>2</sup>Suriatini Ismail, <sup>2</sup>Ibrahim Sipan,  
<sup>2</sup>Mohamad Ghazali Hashim and <sup>1</sup>Mohammad Firdaus Azhar

<sup>1</sup>*Institute of Advanced Technology  
Universiti Putra Malaysia  
43400 UPM Serdang, Malaysia  
Email: rehat\_55@yahoo.com*

<sup>2</sup>*Faculty of Geoinformation Science and Engineering  
Universiti Teknologi Malaysia  
81310 UTM Skudai, Malaysia  
suriatini@fksg.utm.my, ibrahim@fksg.utm.my, kp\_kuat@hotmail.com*

## ABSTRACT

Mass appraisal can be defined as an art and science of valuing properties in a defined geographic area using maps, market data, comparative information and valid statistical procedure. Mass appraisal allows valuing a large number of properties in a reasonably short period of time and for a relatively small cost per property. Several statistical methods are available for mass appraisal of properties and this paper focuses on the use of Moving Window Regression (MWR). MWR is capable of handling spatial effects peculiar to property data where this characteristic is not available in the conventional Multiple Regression Analysis (MRA). This paper presents preliminary analysis of the MWR method using 196 housing sales observation occurring between 2004 and 2006 within Majlis Perbandaran Kulai. The results are compared with the MRA method to assess the predictive performance of the method.

**Keywords:** Geographically Weighted Regression (GWR), spatial heterogeneity, spatial regression, property tax

## INTRODUCTION

Multiple Regression Analysis (MRA) is a conventional statistical technique that is commonly used in mass appraisal of properties. It is used to predict the value of a dependent variable (price) based on known values of independent variables (land area, floor area, number of bedrooms, etc.). The method starts by establishing a mathematical equation (a model) that describes how the variables are related. The specified model is calibrated against a set of houses where their prices and related independent variables are known. The outcome of the model calibration, which is a set of estimated coefficients, is used in the prediction of prices of other houses in the same area where the values of the independent variables of the houses are known.

The main problem with the MRA method is the difficulty of integrating location in the model. The importance of location in property valuation has long been established and thus cannot be ignored. It is accepted that properties are spatially unique and location is an important component that determines their values. Incorrect specifications of location effects in the model resulted in residuals that are spatially correlated and thus, violated one of the basic requirements of the MRA formulation.

Several advances of multiple regression methods that take into account of location in their models have recently emerged. The methods include spatial hedonic modeling (SHM), geographically weighted regression (GWR), moving window regression (MWR) and kriging. SHM models location explicitly through its spatial variables while GWR and MWR model location implicitly in the model formulation. The kriging method models location explicitly using local spatial autocorrelation.

A research is being conducted to assess the predictive performance of the MWR method in mass appraisal of residential properties in Malaysia. The aim of this paper is to present preliminary findings of this research. The next section presents discussions on the MWR model. This will be followed by discussions on data used in the analysis. Results and discussions of the results are presented in the subsequent section. The paper concludes by highlighting important points.

## MWR MODEL

The general linear regression model of MRA is given as

$$y(i) = b_0(i) + b_1(i)x_1 + b_2(i)x_2 + \dots + b_m(i)x_m + e(i)$$

where  $y$  is the dependent variable;  $x_1, x_2, \dots, x_m$  are the independent variables;  $b_0, b_1, b_2, \dots, b_m$  are the coefficients that will be estimated;  $e$  is the error term and  $(i)$  is the  $i^{\text{th}}$  observation.

The coefficients for the model can be estimated using ordinary least squares (OLS) and the equation, in matrix form, is

$$\hat{\beta} = (X^T X)^{-1} X^T Y$$

where  $\hat{\beta}$  is the estimated vector of coefficients,  $X$  is the matrix of independent variables with the elements of the first column set to 1, and  $Y$  is a vector of observations of the dependent variable.

The fitness of the model can be measured using residuals sum of squares, RSS, standard error,  $s_e$ , coefficient of determination,  $R^2$ , and adjusted coefficient of determination,  $\bar{R}^2$ . The equations of these measures are

Residual sum of squares (RSS)

$$RSS = \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

Standard error ( $s_e$ )

$$s_e = \left[ \frac{\sum (y_i - \hat{y}_i)^2}{(n - m - 1)} \right]$$

Coefficient of determination ( $R^2$ )

$$R^2 = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y}_i)^2}$$

Adjusted coefficient of determination ( $\bar{R}^2$ )

$$\bar{R}^2 = 1 - \frac{(n-1)}{(n-m-1)} (1 - R^2)$$

where

$y_i$  = observed value

$\hat{y}_i$  = predicted value

$\bar{y}_i$  = mean  $y_i$

$n$  = number of data points

$m$  = number of  $b$  coefficient

Simple linear regression in a moving window with equal weights assigned to all observations within the moving window (Chris Lloyd, Ian Shuttleworth 2005). The MWR method uses the same general linear regression model as the MRA. The only difference is that the regression is being applied to a smaller area (a window) instead of the whole study region. Therefore, several regressions are performed in order to cover the study region. By doing so, the MWR method is able to capture spatial variations in the study region in which the MRA method does not have the capability to do so.

For any particular window, the regression is based on the modeled relationship between the regression point and  $n$  observation points in the window. Only  $n$  neighbors of the regression point are taken into account in a regression. Observation data outside the window is not considered at all. The  $n$  observations within the window are weighted equally. It is quite common to refer the concept as “all data within the window receive a weight of one, whereas all data outside of the window have a weight of zero”.

MWR weighting scheme can be defined as follows:

$$w_{ij} = 1 \text{ if } j \text{ is one of the } n^{\text{th}} \text{ nearest neighbors of } i;$$

$$= 0 \text{ otherwise}$$

Window size might not be the same when different areas are considered because the size of the window will be dependent on the density of observations around the regression point. Window sizes have to be explored to determine the optimum size (or number of nearest neighbors) that will give the best result. In order to determine the best window size  $a$ , cross validation is performed. The equation for cross validation is given as

$$CV = \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

where  $n$  is the number of data points in a widow,  $y_i$  is the observe value for the  $i^{\text{th}}$  data point and  $\hat{y}_i$  is the estimated value of the  $i^{\text{th}}$  point obtained when  $i$  itself is omitted from the computation.

## **DATA**

For the purpose of this paper, the test dataset comprises 196 single storey terrace houses transacted between 2002 and 2006 within Majlis Perbandaran Kulai (MPKu) area in Johor. The dependent variable is the transacted price while the independent variables are:

- Land area
- Main floor area
- Ancillary floor area
- Additional floor area
- Position (intermediate, end, corner)
- Floor finishing (tiles, cement, tiles and cement)
- Title (others, HS(D)freehold, HS(M)freehold and grant)

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

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

## **RESULTS AND DISCUSSIONS**

Assessment of the prediction accuracy of the MRA and MWR models is done via in-sample data, i.e., the same dataset is used for calibration and prediction; The calibration of the models were carried out and the estimated coefficients obtained from the calibration were then used to predict the price of the houses where the observed or the actual prices are already known. Predictive accuracy of the models is determined by comparing the observed and predicted prices. Table 1 summarized the calibration output of the two models.

Table 1: Calibration Output

	MRA	MWR
Standard error	9639.1439	5979.0065
Residual Sum of Square (RSS)	17,467,662,833.31	7,189,446,651.22
Coefficient of Determination ( $R^2$ )	0.625	0.8558
Adjusted $R$ -Square ( $\bar{R}^2$ )	0.611	0.8508

In assessing the fitness of the models, the MWR model is better than the global MRA model. The adjusted  $R^2$  of the MWR and MRA models are 0.8508 and 0.611, respectively, while the  $R^2$  of the MWR and MRA models are 0.8558 and 0.625, respectively. The MWR model Residual Sum of Square (RSS) and standard error are also smaller when compared to the MRA model. These results indicate that the fitness of the MWR model is better than the MRA model.

Table 2 shows the coefficients and the values of t-statistics for the independent variables of the MRA model. The t-statistics show that all determinants for the global MRA model are statistically significant, at the least at the 90% level of confidence. *Land Area*, *MFA*, *AFA*, *Add\_FA*, *Position* and *floor* positively contribute to property price; an increase in the value of one of these variables increases the property price. Meanwhile, *Title* variable negatively contribute to property price. The analysis of t-statistics indicates that the global model specifications are appropriate.

Table 2: Coefficients and t-statistic for the independent variables of the MRA model

Independent Variables	Coefficient (t-statistics)
(Constant)	94227.2563 (11.5367)***
Land Area	75.6146 (2.6141)***
MFA	158.0275 (2.0185)**

AFA	126.9817 (1.2970)*
Add_FA	547.9209 (8.3874)***
Position	5456.5426 (1.6414)*
Floor	3458.3599 (3.5634)***
Title	-1676.8323 (-2.0023)***

Dependent Variable: Property Price

\*\*\* Significant at 99%, \*\*Significant at 95%, \*Significant at 90%

The predictive accuracy is measured by the value of the Coefficient of Dispersion (COD). COD is the average percentage of absolute dispersion around the median predicted ratio. The lower the COD, the more uniform or less disperse the prediction within the study region. The International Association of Assessing Officers (IAAO) views  $COD \leq 15.0\%$  to indicate a high quality or commendable prediction (Strauss et. al, 2003).

The equation for calculating the COD is given as

$$COD = \frac{100}{Median_{P/O}} * \left[ \frac{\sum_{i=1}^n |(P_i / O_i) - Median_{P/O}|}{n} \right]$$

where:

$P_i$  = predicted value of the  $i^{th}$  property

$O_i$  = observed price of the  $i^{th}$  property

$Median_{P/O}$  = median of the distribution of  $P_i/O_i$

Values of the COD in Table 3 indicate that the MWR model predict better than the MRA model. The MWR model has 3.3% COD while the MRA model has 5.1% COD. The MWR model prediction is less disperses than the MRA model. Maximum and minimum deviations of predicted prices relative to the observed prices are other indicators of how well a model predict. The MWR model has smaller values than the global MRA model for both measures which indicate that the predicted prices of the MWR model are closer to observed values than the predicted prices of the MRA model.

Table 3: Summary of Prediction Accuracy

	MRA	MWR
Max deviation	RM50,497	RM 26,808
Min deviation	RM142	RM0
Coefficient Of Dispersion (COD)	5.1%	3.3%

### CONCLUSIONS

The main objective of this paper is to preliminary asses the performance of the MWR model in mass appraisal of residential properties in Malaysia. The focus of the assessment is on the prediction accuracy in order for the model to be useful for mass appraisal, without compromising other aspects. This was done by comparing the performance of the MWR model and the conventional MRA global model. It was found that, despite limitations of the test dataset (small size, uses structural variables only, ignoring temporal effects, and in-sample prediction) the MWR model can predict the property price better than the global model.

This research will proceed with testing of datasets without the limitations indicated above. Larger and more representative test datasets with structural, accessibility, and neighborhood variables will be built and out-sample prediction will be carried out. The model will also incorporate temporal aspects of the data. It is hoped that the suitability of the MWR model to be used for mass appraisal of residential properties for rating in Malaysia will be ascertained.

### ACKNOWLEDGEMENT

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

## REFERENCES

- Chris Lloyd, Ian Shuttleworth (2003). "Analysing commuting using local regression techniques: scale, sensitivity, and geographical patterning" *Environment and Planning A* 2005, volume 37, pages 81 – 103
- F.Long, A.Paez and S.Faber (2007) "Spatial effect in hedonic price estimation: A case study in Toronto" Centre for Spatial Analysis, McMaster University
- Fotheringham A.S., Brunson C. and Charlton M.E. (2002) *Geographically Weighed Regression: the Analysis of Spatially Varying Relationships*. West Sussex: John Wiley & Sons Ltd
- Robert P. Strauss and David A. Strauss (2003) "Residential Real Estate Assessment Fairness in Four Urban Areas" 96<sup>TH</sup> Annual Conference on Taxation, Carnegie Mellon University
- Steven Farber (2004). "A Comparison of Localized Regression Models in a Hedonic House Price Context". Ryerson University, Toronto, Ontario, Canada