Modeling property rating valuation using Geographical Weighted Regression (GWR) and Spatial Regression Model (SRM): The case of Kota Kinabalu, Sabah

Oliver Valentine Eboy¹, Narimah Samat²

¹School of Social Science, Universiti Malaysia Sabah, 88999 Kota Kinabalu, Sabah, Malaysia, ²School of Humanities, Universiti Sains Malaysia, 11800 Penang, Malaysia

Correspondence: Oliver Valentine Eboy (email: oliver@ums.edu.my)

Abstract

Property revaluation or reassessment is a compulsory activity for property tax to be imposed on all properties. It was conducted manually, involving exhaustive, time consuming and costly processes. As such there is a growing need to develop alternative valuation models capable of estimating property values of large numbers in a short time with little manpower and low costs. The spatial statistics of geographical weighted regression (GWR) and spatial regression model (SRM) are two of them. This study demonstrates the development of the GWR and SRM in estimating residential property value in areas under the Kota Kinabalu City Hall (DBKK) jurisdiction. It collected and cleaned 5,524 data items. Five valid and significant variables were identified and utilized in the modeling exercise. By using GWR and SRM various tests were conducted to identify and remove modeling errors such as multicollinearity, heteroscedasticity and spatial autocorrelation. It was found that the SRM stood out as the best property rating valuation model for DBKK area compared to the GWR. The SRM analysis also identified the building quality as the main positive influence of the property rates while the location factor provides the least influence. In short, this study had proved the effectiveness of SRM in producing a property rating valuation model even with problematic dataset. It could also, in addition, easily produce property value maps to indicate variations in property rates and thus improve the management of property rating valuation in local authority areas.

Keywords: geographical weighted regression (GWR), Kota Kinabalu, model error, property rating, property valuation model, spatial regression model (SRM)

Introduction

All taxes, rates, rents, license fees, dues and other sum or charges payable to the local authority are some of the main source of revenue to the local authority. Among the sources stated above, the rates or property assessment accounted 60 to 70 percent to the total revenue of the local authorities (Ahmad Atory Hussain, 1991). However, the assessment needs to be updated from time to time to keep up with the current market value. In order to do this, revaluation of the rates need to be conducted every five years which is in accordance with the Local Government Act 1976. Unfortunately, the revaluation was normally carried out after 10 or 20 years (Dzurllkanian Daud et al., 2008). Table 1 shows the pending revaluation exercise by Local Government in Malaysia. As stated in Table 1 below, only 16 local authorities had performed revaluation within 1 – 5 years after the end of last revaluation, while 29 others conducted the revaluation after 6 years or more.
Table 1. The pending revaluation exercise by local governments in Malaysia

<table>
<thead>
<tr>
<th>Pending Revaluation (After Last Revaluation)</th>
<th>Frequency</th>
<th>Percentage (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-5 years after 5-year end of last revaluation</td>
<td>16</td>
<td>35.6</td>
</tr>
<tr>
<td>6-10 years after 5-year end of last revaluation</td>
<td>11</td>
<td>24.4</td>
</tr>
<tr>
<td>10-15 years after 5-year end of last revaluation</td>
<td>7</td>
<td>15.6</td>
</tr>
<tr>
<td>More than 15 years after 5-year end of last revaluation</td>
<td>11</td>
<td>24.4</td>
</tr>
<tr>
<td>Total</td>
<td>45</td>
<td>100</td>
</tr>
</tbody>
</table>

Source: Dzurillkanian Daud et al. (2012)

Revaluation has not been conducted regularly since it is time consuming and costly process to be undertaken manually (Tretton, 2007; Mustafa Omar, 2004). In addition to that, there are inadequacies in tax administration such as lack of assessment tools and absence of technically qualified personnel (Dzurillkanian Daud et al., 2008). Consequently, the rating values of the property were generally behind the current market value. Although, computers has been used in producing property rating maps and running daily administrative operation such as tax collection in most local governments, it is not used for the analyzing or calculating the property rating.

Property valuation model was then introduced to overcome this problem. It is capable to performed valuation for the property in large quantity for taxation purpose and in a very short time. The model would enable the authority to produce a faster and cheaper revaluation process with accurate property value predicted. This technique also provides uniformity and consistency in ad valorem valuations particularly when revaluations of large number of parcels at the same time (Deddis, 2002). Such an approach potentially could help local authority to speed up revaluation process and reduce cost.

Unfortunately, the usage of this approach have yet to be materialized in Malaysia as it is still developed and tested at the academic level even though such approach has been adopted by various countries such as United Kingdom, Australia, U.S, Africa, New Zealand and Europe (Dzurillkanian et al., 2006:2). Therefore, new method and new study is needed to be conducted in order to convince the local authority in Malaysia to adopt this approach.

This paper examined the capability of spatial statistics specifically the geographical weighted regression (GWR) and spatial regression model (SRM) in developing a property value model for tax purpose in a local authority jurisdiction area. However, the study for this paper focused on the residential properties excluding apartment and condominium as different modelling approach required for these type of properties and thus pose difficulty in model comparison. The objective of this study was, firstly, to collect and identified the property rating modelling variables. Secondly, to conduct modelling analysis using GWR and SRM and then, perform some test for model error. Thirdly, to assessed and compared the model performance as to identify which model is suitable for DBKK’s property rating model.

Property valuation model

Traditionally in property valuation, five valuation methods which was used consistently are comparable method, cost method, residual method, investment method and income method (Scarrett, 2008; Richmond, 1985; Ismail Omar, 1992; Appraisal Institute, 1992). However, there is another valuation method that gains momentum at this time which is the regression method (Brown, 1974; Gloudemans and Miller, 1978; Mark and Goldberg, 1988; Cannaday, 1989; Ismail Omar, 1992).

Recent studies showed that an advance regression method in a form of GWR developed by Brunson et al. (1996, 1998) was used for property valuation (Hernandez et al., 2003; Bitter et al., 2006; Long et al., 2007; McCluskey and Borst, 2011) including in Malaysia (Taher Buyong, 2011; Ibrahim Sipan et al., 2012). Most of the studies managed to prove that the GWR, with the capability to include the geographical coordinate in the regression equation, was a better property valuation model than the
traditional method. It managed to provide better accuracy than the traditional regression model that been used in determining the properties’ value at that time.

However, GWR is still vulnerable to produce model error especially spatial autocorrelation error (Löchl and Axhausen, 2010; McCluskey and Borst, 2011). Spatial autocorrelation is one of three main modelling errors that could occur in a property valuation model (Des Rosiers et al., 2001). The other two are multicollinearity and spatial heterogeneity. These model errors, if unchecked, would lead to bias, misleading or misspecification to the property valuation models (Rosenshein et al., 2011). In other word, the model would be inaccurate. Therefore, another modelling method based on the regression technique namely spatial regression model (SRM) specifically used to address the spatial autocorrelation error (Suriatini Ismail, 2005; Löchl and Axhausen, 2010). It has the capability to detect the spatial autocorrelation in two different forms namely, spatial error model and spatial lag model using the lagrange multiplier (LM) test (Wilhelmsson, 2002). A spatial lag model or a mixed regressive, spatial autoregressive model is appropriate when the focus of interest is the assessment of the existence and strength of spatial interaction. In this model, the property value would be estimated partially from nearby or neighboring observations of other property values. This model would assume that the property value of each property was affected by the property values in the neighborhood in a form of spatial weighted average (Suriatini Ismail, 2005). This is in addition to the other variables that provide indirect effect to the property value which represent the property and neighbourhood characteristics. The spatial error model was used for spatially autocorrelated model which occurred because of the error term in the model. Thus, the spatial error model is capable to rectify any potential bias influence of spatial autocorrelation due to the use of spatial data. It helps to find the most suitable coefficients estimation in the model and ensure that the correct inference is adopted. It is however, not appropriate for model which indicates no spatial interaction (Suriatini Ismail, 2005). Thus, SRM managed to provide good estimation in some property value model studies (Suriatini Ismail, 2005; Löchl and Axhausen, 2010) and potentially managed to eliminate the model error.

Study area

Property within Dewan Bandaraya Kota Kinabalu or Kota Kinabalu City Hall (DBKK) jurisdiction is the city council which administers the city and district of Kota Kinabalu in the state of Sabah, Malaysia was used as study area. DBKK area was chosen as recently, from year 2011 to 2012, it attained the highest figure in malaysian property index (JPPH, 2012) and produced rapid increase in property tax collection from year 1998 to 2010 (DBKK 2011). Figure 1 shows the location of Kota Kinabalu in Sabah. It covers a large area in Kota Kinabalu that consists of many zones. However, due to data constraint, only the selected zones in city and urban area were used for modelling purpose which includes Kota Kinabalu, Luyang, Luyang Timur, Teluk Likas, Sembulan, Tanjung Aru, Damai, Kolam, Ridge, Kepayan, Dah Yeh and Signal Hill.

Methodology and data

A modeling framework was outlined for this study to produce the property rating valuation model as shown in Figure 2 below. The first stage involved acquisition of property value including its contributing factors and the spatial elements in property valuation that needed for valuation. The attributes consists of physical building, geographical aspect, neighborhood, external facilities and legality represent the non-spatial data were compiled. While the spatial data consist of location factor which was derived using GIS where distance from each property location to the nearest location factor such as bank, tourism attraction, market and school was measured. The selected relevant data were then gathered and examined using various steps such as verification, cleaning and conversion to prepare database suitable for analysis. Data gathered during first stage would be brought in to the second stage, where analysis was performed using
GWR method based on the data acquired. The SRM analysis would only be conducted if the GWR indicates spatial autocorrelation error. The model developed was run through an assessment to obtain a property rating valuation model suitable for the residential properties in the study area.

Figure 1. Location of Kota Kinabalu in Sabah
Before the assessment process can begin, the spatial autocorrelation test need to be conducted in the stage of GWR and SRM analysis as shown in the diagram in Figure 3 below. The test was initiated once the GWR analysis was processed. If the GWR output indicates that spatial autocorrelation was not present in the model, the GWR then can proceed to the assessment stage. However, if the spatial autocorrelation
exists, the SRM analysis needs to be applied. The SRM is categorized into two models namely the spatial error and spatial lag model. Using the LM test, the model which attain significant or the highest value would be selected as the property rating model for this study.

Due to unavailability of recent data, the data from year 1997 was used in which it is still valid and currently applied in DBKK at the time of this study (DBKK, 2012). Originally, the study collected 14,520 observations for the whole area of DBKK through selection of residential property valuation data excluding apartments, flats and condominiums within the urban area. However, after data cleaning and removing of missing or incomplete data, only 5,524 records were retained for the analysis. This was enough to be used as data sample to develop the property rating model.

Model development

In the starting of the analysis, the first model was developed by using the GWR. This model has been rewritten in equation (1) based on the traditional regression formula as follows (Charlton and Fotheringham, 2009):

\[ y_i(u) = \beta_{0i}(u) + \beta_{1i}(u)x_{1i} + \beta_{2i}(u)x_{2i} + \ldots + \beta_{mi}(u)x_{mi} \]  

for \( i = 1..n \)  

Where;
- \( y \) is the vector of observed values
- \( \beta \) is the vector of estimated parameters,
- \( x \) is the design matrix which contains the values of the independent variables,
- \( u \) is the vector of location (coordinate)

The notation \( \beta_{0i}(u) \) indicates that the parameter describes a relationship around location \( u \) and is specific to that location. A prediction may be made for the dependent variable if measurements for the independent variables are also available at the location \( u \).
In the event of which the spatial autocorrelation exists and unable to be eliminated from the model, the SRM would then be conducted. Two types of SRM model can be produced which is the spatial lag and spatial error model. A spatial lag model can be expressed in equation (2) as follows (Anselin, 2001:316):

\[ y = \rho Wy + x\beta + \varepsilon \]  

(2)

Where;
- \( y \) = Dependent Variable
- \( \rho \) = spatial coefficient
- \( Wy \) = weight matrix for dependent variable
- \( x \) = matrix of observations on the independent variables
- \( \varepsilon \) = vector of error terms

While a spatial error model can be written in equation (3) as follows (Lehner, 2011:5):

\[ y = \beta_0 + \beta_1 x_1 + \ldots + \beta_n x_n + u \]
\[ u = \lambda W u + \varepsilon \]
\[ \varepsilon \sim N(0, \sigma^2 I_n) \]  

(3)

Where;
- \( y \) = vector of dependent variable
- \( \beta_0 \) = Constant term
- \( \beta_1, \ldots, \beta_n \) = Independent Variable Component
- \( u \) = vector of spatially correlated error
- \( \lambda \) = spatial autoregressive coefficient
- \( W \) = spatial weight matrix
- \( E \) = random error

Variables to be used in determining property rating value were identified and would be discussed in the following section.

**Model variable selection**

After undergoing data preprocessing and cleaning, five independent variables was selected to be used to estimate the dependent variable. The dependent variable is the property rating value that was imposed by the DBKK to the property owner. This variable is measured based on currency scale in ringgit Malaysia (RM). The five independent variables chosen for the model were RCA, land area, building type, building quality and location factor which also called as the property value influence factor. The reduced coverage area (RCA) represents the main floor area of the property but was recalculated to be better suited for valuation purpose. While the land area referring to the land size available in the property area. Both variables were measured using square feet unit. Next, the location factor variable was obtained based on the GIS analysis conducted and the measurement was based on meter unit from the property location to the nearest location factor consists of public institutions, tourism centers, public recreations, public facilities, commercial areas, government offices and religious centers. As for the building type it represent the type of residential property consists of semi-detached, terrace, town house and kampung house. This is followed by building quality that provides the condition level of the building. Both building type and building quality variables was converted into interval scale measurement using expert judgement feedbacks from the DBKK’s property valuers. The summary descriptions of the variables selected for the model are shown in Table 3 below.
Table 3. Summary description of the GWR and SRM model variables.

<table>
<thead>
<tr>
<th>Variable Name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>PropRate_Value</td>
<td>Current Property Rating Value</td>
</tr>
<tr>
<td>Bld_type</td>
<td>Building type</td>
</tr>
<tr>
<td>RCA</td>
<td>Reduced Covered Area Estimation</td>
</tr>
<tr>
<td>Bldq</td>
<td>Building Quality</td>
</tr>
<tr>
<td>Land_area</td>
<td>Land area estimation</td>
</tr>
<tr>
<td>Location_Factor</td>
<td>Distance from property to the nearest location factor</td>
</tr>
</tbody>
</table>

Once the variables were selected and analyzed, the output from both GWR and SRM could then be assessed and compared to determine which model that best represent the DBKK area for property rating purpose.

Results and discussion

To test the spatial autocorrelation formally, this study adopted the spatial statistics of Moran’s I to determine the existence of significant spatial autocorrelation. This test enables identification of the three forms of spatial autocorrelation, of positive, negative or random. Moran’s I value of the GWR model indicates positive spatial autocorrelation (Z score = 74.080, p-value = 0.00) meaning that similar residuals cluster together. This means that it is more likely for the spatial autocorrelation detected to occur out of missing variables for important property characteristics. Subsequently, SRM analysis needs to be conducted and the type of spatial autocorrelation of spatial error and spatial lag need to be identified.

Based on Table 4, it shows that both LM (Error) and LM (Lag) were significant (p-value of 0.000). Hence, this would require the consideration of a robust form of the statistics as decision unable to be made based on the previous result. However, both robust LM (Error) and robust LM (Lag) also produced significant result. Therefore, if both robust LM produced significant result in spatial autocorrelation, the model with the higher value prevails (Anselin, 2005). In this case, the robust LM (error) achieved higher value of 1420.9258 compare to robust LM (lag) with 31.3767. The spatial autocorrelation error detected shows that some missing variables occurred from the model that were not included in the model. The missing variables might come from the variables that had been removed from the model because of missing records or produced multicollinearity error. As a result, the SRM’s spatial error model would be used for this study as the residential property rating valuation model for the entire zone of Kota Kinabalu area.

Table 4. Output from the LM spatial autocorrelation test of the study area

<table>
<thead>
<tr>
<th></th>
<th>Lagrange Multiplier (Error)</th>
<th>Lagrange Multiplier (Lag)</th>
<th>Robust Lagrange Multiplier (Error)</th>
<th>Robust Lagrange Multiplier (Lag)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Value</td>
<td>7919.8632</td>
<td>6530.3141</td>
<td>1420.9258</td>
<td>31.3767</td>
</tr>
<tr>
<td>Probability</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

In the final step, the model performance was then examined. Hence, the measurement of $R^2$ values was referred to, in which, the higher its value, the better the accuracy of the model. High accuracy of the property value estimation would be produced if the measurement of $R^2$ was high. In this study, the $R^2$ achieved 0.78 value indicated that the SRM model explains approximately 78% of the property rating value. This figure indicates good accuracy estimation of the model. It was also higher than the $R^2$ of GWR which obtains 0.72 or 72%. This study also conducted model performance comparison using
Akaike Information Criterion (AIC). The rule of thumb is that, any of the models that produce the lowest value is the better model. The AIC value for SRM (86263.6) was smaller than the GWR (86800.46) which further strengthen the suitability of SRM compared to GWR. Therefore the SRM was selected as the property rating model for the DBKK area.

To determine the strength and type of relationship the independent variable has to the property rating value, the coefficient for each of the independent variable were measured. Table 5 shows the coefficient value of each independent variable which also called as property value influence factor. The coefficient reflects the expected change in the property rating value for every one unit change in the property value influence factor. For example a coefficient of 443.656 associated with building quality (BLDQ) representing RM currency may be interpreted as RM443.656 of property rating value. This shows that BLDQ gives a high increase to the residential property value in the study. Another independent variable that provided a high positive increase to the residential property value is the building type (BLD_TYPE) with coefficient value of 249.069. The other factors of RCA (RCA), land area (LAND AREA) and location factor (LOC_FAC) also gave positive increase albeit lower coefficient value of 0.1095 0.005 and 0.267 respectively. All the independent variables of BLDQ, BLD_TYPE, RCA, LAND AREA, LOC_FAC and including the Intercept were statistically significant at 95% confidence level based on the probability measurement which means the coefficient value for all the variables were eligible to be used to explain the model.

<table>
<thead>
<tr>
<th>Property Value Influence Factor</th>
<th>Coefficient (B)</th>
<th>Relationship with Property Rating Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>-37.3567</td>
<td>Moderate negative relationship</td>
</tr>
<tr>
<td>BLDQ</td>
<td>443.6557</td>
<td>Strong positive relationship</td>
</tr>
<tr>
<td>BLD_TYPE</td>
<td>240.0685</td>
<td>Strong positive relationship</td>
</tr>
<tr>
<td>RCA</td>
<td>0.105202</td>
<td>Weak positive relationship</td>
</tr>
<tr>
<td>LAND AREA</td>
<td>0.005199</td>
<td>Weak positive relationship</td>
</tr>
<tr>
<td>LOC_FAC</td>
<td>0.26748</td>
<td>Weak positive relationship</td>
</tr>
</tbody>
</table>

Based on the Figure 4 below, the distribution of the property rating value estimated by the SRM’s spatial error model can be clearly visualize using GIS tool. The distribution of the property rating value in the map shows that parts of Bukit Padang and Tanjung Aru zones (dark color) contributed highest property values in the area. Based on the result in Table 5, there is a high probability that the high values occurred because of the high influence from the building quality and building type in that area. Additionally, this could also attribute due to the location factor as the affected zones are situated nearby attractive places such as hillside view, recreational parks and beach. On the contrary, large parts of Ridge and Kepayan zones (light color) obtain lowest values in the area. The SRM model unable to provide the reason behind this as none of the variables included in the model provide negative effect except the Intercept. The negative value in the intercept shows that there are missing variables that contributed to the negative value influence in the area which was not included in the model.

Based on the discussion with the DBKK authority, the reason of spatial autocorrelation error occurred, in the DBKK valuation data was probably due to two factors. Firstly, the different type of building structure was not addressed properly. For example, the property type of detached house was not categorized as temporary, semi-permanent or permanent structure although this information can affect the property value. Secondly, some residential properties were used for commercial purpose has made the model confused as although the size of the area is big but low in value or vice versa. These residential properties were mainly used either as play school or showroom cum office. Most of these houses were
located along the main road or can be clearly seen from the main road. Inconsistency in the recording of this data may have contributed to the error in the model.

![Figure 4. Property rating map using SRM spatial error model for DBKK area.](image)

**Conclusion**

This paper exemplifies the property rating model for tax purpose developed using GWR and SRM. This model is capable of estimating a large-scale property value in the area. Using the samples of 5,524 data from the property valuation data from DBKK, the model successfully estimated the property values and displayed it in a value map using GIS tool. Despite the existence of spatial autocorrelation error in GWR, the SRM manage to overcome the error to produce a suitable property rating model for DBKK. Eventually, the SRM was chosen as the property rating model for DBKK. The performance of the SRM model was also good with 78% accuracy and this was valid to be used as a rough references or guideline for the authority to apply rating value in the area. This study also takes into account of spatial autocorrelation test and shows the relevance of using SRM as the property rating model. Although there is still much to be done especially to overcome the spatial autocorrelation problem in the DBKK data but this could be one of the early step in producing property valuation model for DBKK. Therefore, this study has proved that spatial statistics can be used to assists the local authority in determining the property rating value of the area. This is also a major contribution to improve revaluation exercise such that accurate property rating could be obtained and able to minimize the cost, time and manpower.
Acknowledgements

Authors would like to acknowledge Universiti Malaysia Sabah for providing financial support to undertake this research and Kota Kinabalu City Hall (DBKK) for providing data and information to be used in this study.

References


