Pertanika J. Sci. & Technol. **19 (S)**: 23 - 30 (2011)

ISSN: 0128-7680

A Review of Property Mass Valuation Models

Ebrahim Jahanshiri*, Taher Buyong and Abdul Rashid Mohd. Shariff
Spatial and Numerical Modeling Laboratory,
Institute of Advanced Technology, Universiti Putra Malaysia,
43400 UPM, Serdang, Selangor, Malaysia
**E-mail: e.jahanshiri@gmail.com*

ABSTRACT

Mass valuation of properties is important for purposes like property tax, price indices construction, and understanding market dynamics. There are several ways that the mass valuation can be carried out. This paper reviews the conventional MRA and several other advanced methods such as SAR, Kriging, GWR, and MWR. SAR and Kriging are good for modeling spatial dependence while GWR and MWR are good for modeling spatial heterogeneity. The difference between SAR and Kriging is the calculation of weights. Kriging weights are based on the spatial dependence or so called the semi-variogram analysis of the price data whereas the weights in SAR are based on the spatial contiguity between the sample data. MWR and GWR are special types of regression where study region is subdivided into local sections to increase the accuracy of prediction through neutralizing the heterogeneity of autocorrelations. MWR assigns equal weights for observations within a window while GWR uses distance decay functions. The merits and drawbacks of each method are discussed.

Keywords: Spatial prediction, property price indices, spatial econometrics

INTRODUCTION

For long, it has been a problem to assess property values accurately. Assessors and appraisers are known to be able to estimate values of properties through their accumulated knowledge. However, the challenges are (i) the accuracy and consistency of these valuations that refers to the weights that appraiser gives to specify the quality of the appraised value, and (ii) the speed of which the appraising process can take place. Correct and up-to-date assessment of property values is not only important to owners of the properties and real estate agencies but also to the local governments whom must define the taxes to be imposed on the properties based on their values. It is also a requirement that property values must be regularly updated in order for the taxes to be accurate and fair.

Over the past decades, property valuation has evolved from simple empirical judgments to automated valuation models and their applications have extended from single property to mass valuation (Clapp, 2003). Manual methods of expert valuation although effective, are subjective, inconsistent, and prone to errors (Adair and McGreal, 1988; Benjamin *et al*., 2004). For large jurisdictions that encompass thousands or millions of properties, manual valuation if possible, is time consuming. Therefore, automated valuation models are invented to solve for these types of problems. Automated valuation models consist of a database of property values and their characteristics and, current transactions of the properties, in a region of interest. The second major part of automated valuation models is the statistical method that is used to estimate property prices. The third part is the output and graphical user interface to do the communication and visualization of the output

Received: 1 August 2010
Accepted: 22 June 2011
*Corresponding Author

of the models. The methods of mass property valuation for so long have been confined to sales comparison method and Multiple Regression Analysis (MRA). However, apparent deficiencies of these methods have been the motivation for the usage of body of methods that are invented and borrowed from other disciplines to increase the accuracy of valuation. Improved accuracy of the new methods is possible by explaining parts of error of regression through consideration of the spatial autocorrelation and spatial heterogeneity. These effects are materialized when there is influence in terms of human communication and market demands on the property prices. Currently, these new methods are divided into two main sections that either deal with spatial autocorrelation or spatial heterogeneity. Both of these streams have their own sound theoretical basis although they may need to be merged to be more effective. A few methods like moving window kriging tries to deal with both of these effects. This paper aims to provide a review on methods of mass valuation and their improvements in the spatial domain that have been made in recent years. The composition of the paper is as follows: Section two provides the taxonomy of property mass valuation methods. Section three presents MRA, the de facto standard of mass appraisal model. Section four discusses models for spatial regression and prediction which includes spatial autoregressive models, geostatistical models, and local models. Section five provides future research directions and Section six concludes the paper by highlighting the important points.

TAXONOMY OF MASS VALUATION METHODS

MRA model is the de facto standard for mass valuation of properties. The model originated from non spatial discipline did not address peculiarities of spatial data like property data. Several other models emerge that largely aim to modify the MRA model to take care of spatial effects. The spatial econometrics research contributed the global Spatial AutoRegressive (SAR) models. These models are known as the Spatial Lag Model (SLM). Spatial Error Model (SEM), General Spatial Model, and Spatial Durbin Model (SDM). The geographic research contributed local models of Geographically Weighted Regression (GWR) and Moving Window Regression (MWR). The geostatistics research contributed the various kriging models including Regression Kriging (RK) and Moving Window Kriging (MWK). *Fig. 1* shows the taxonomy of mass valuation models.

MULTIPLE REGRESSION ANALYSIS (MRA)

MRA is a statistical methodology that utilizes the relationship between two or more independent variables (characteristics of properties like size of living area, number of bedrooms, number of bathroom, and so on) and a dependent variable (price of properties). The dependent and independent variables are regressed using properties of known prices to determine the established relationships (coefficients) between the two types of variables (Adair and McGreal, 1988). The determined coefficients are then used for the prediction of prices of unsold properties in the same stock. MRA determines the coefficients with the least possible error (Benjamin *et al*., 2004) using the Ordinary Least Squares (OLS), maximum likelihood (ML), or Weighted Least Squares (WLS) estimation techniques with OLS being the most popular (Ambrose, 1990; Beach and MacKinnon, 1978). The OLS method minimizes the sum of square of residuals or errors. The regression coefficients that are derived based on OLS shall be best linear unbiased estimator (BLUE).

However, there are some drawbacks on the use of MRA in property valuation relating to spatial autocorrelation and heteroscedasticity, the two spatial effects inherent in property data (Mark and Goldberg, 1988; Fletcher *et al*., 2000). Spatial autocorrelation means that the residuals are spatially correlated; off diagonal elements of the variance-covariance matrix of the estimated residuals deviate from zero indicating that the two observations that define the elements are spatially correlated.

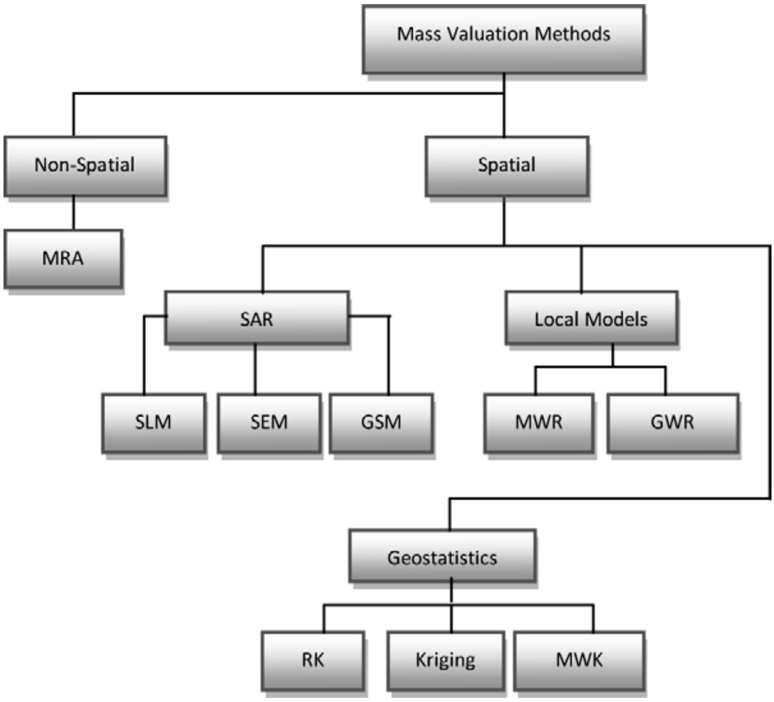

Fig. 1: Mass valuation models

Spatial autocorrelation is the result poor of specification of the regression model which may be due to incomplete or missing spatial variables that accounted spatial dependence and spatial heterogeneity in property data. The inclusion of spatial variables makes the models more complete from the point of view of regressing spatial phenomena. Unaccounted spatial dependence and spatial heterogeneity also makes the residuals deviate from normal distribution. Heteroscedasticity is partly due to spatial dependence and heterogeneity, and partly due to non-spatial reasons. It is difficult to separate the effects but accounting for spatial dependence and heterogeneity may reduce heteroscedasticity.

Varieties of ways are available for assessing the presence of spatially correlated residuals and heteroscedasticity (Belsley *et al.*, 2004). Spatially autocorrelated residuals and heteroscedasticity violate the presumption of OLS that the residuals must be uncorrelated and normally distributed with zero mean and constant variance, i.e., $e \sim N(0, \sigma^2 I)$. This makes the OLS estimated coefficients biased and unsuitable for inference. The ending effect is that the predicted property prices are unreliable.

Appraisal communities in the developed and developing countries have realized the power of computerized mass appraisal and statistical methodology. MRA technique, given its medium accuracy, flexibility, and ease of use is the preferred method that is embedded into the valuation systems especially for tax purposes (Tretton, 2007).

SPATIAL REGRESSION AND PREDICTION

The major reason for the low predictive capability of the MRA is ignoring spatial dependence and spatial heterogeneity. Spatial dependence can be seen when we consider that not only the price of a property is influenced by the prices of the surrounding properties but the characteristics of a property are also influenced by the characteristics of surrounding properties. Spatial heterogeneity can be seen when unit price of land varies from urban to suburban, to rural areas, across a region of interest. Regression methods have advance to incorporate spatial effects. It was shown that regression errors can be reduced and consequently increasing its predictive capability by adding independent variables describing the spatial characteristics of the properties like in spatial autoregressive models or devising a regression procedure to capture spatial heterogeneity like in local models (Paez *et al*., 2008; Buyong and Valivalo, 2010; Fotheringham *et al*., 2002). On the other hand, spatial dependence in property prices is exploited in the prediction of the prices in the method of geostatistical kriging (Gallimore *et al*., 1996; Bonrassa *et al*., 2003; McCluskey and Borst, 2007). We will discuss these advances in the following subsections.

Spatial Autoregressive Models

Spatial AutoRegressive (SAR) models, also called spatial models is a group of models that improves the accuracy of property price prediction of the MRA model by incorporating spatial dependence of properties in the functional model. The spatial dependence parameters are estimated along with the regression coefficients. First is the Spatial Lag Model (SLM) that models the dependency of property prices; the price of a property is dependent on the prices of its neighboring properties. Second is the Spatial Error Model (SEM) that models the spatial dependence of the error terms; an error induced by a property is dependent on the error of nearby properties. Third is the General Spatial Model (GSM) that model both the dependence of prices and errors of neighboring properties; it combines the SLM and SEM into one model. Last is the Spatial Durbin Model (SDM) that models the dependency neighboring property characteristics (Militino *et al*., 2004; Anselin, 1988; Anselin and Bera, 1998; Anselin and Lozano-Gracia, 2009). When spatial dependence are explicitly modeled, the model specifications are more complete and thus, are able to produce more accurate prediction (Ismail *et al*., 2008; Cohen and Coughlin, 2008).

SAR models incorporate spatial weight matrices that are based on the concept of spatial neighbors. Two most commonly used strategies to define spatial neighbors are Delaunay triangulation and k nearest neighbors when properties are represented by their centroids. Properties that are spatial neighbors to a subject property receive the value of one while those that are not spatial neighbors receive zero values. It is normal to try various values of k until satisfactory results are obtained when the k nearest neighbors strategy is used. When the rows and columns of the weight matrix arranged such that the subject property is at the main diagonal, the weight matrices are usually sparse and banded. Literature regarding the application of these models in the property price valuation shows improvement in the property price prediction.

Geostatistical Kriging

Geostatistical kriging is another technique to deal with the spatial autocorrelation. This technique does not fall into the category of regression models since it primarily deals with the property prices and tries to predict the price of unsold properties using the spatial relationship between the prices of sold properties. The spatial autocorrelation first needs to be rectified through a process called variography and then the information that is derived from the variography of the price data will be used to form simultaneous equations or kriging system to determine the price of unsold properties. Variography starts with calculating the differences or semi-variances between all pairs of data that are a specific distance apart. By plotting the semi-variances against different distances and modeling these relationships, we can estimate the degree of relationship (or differences to be more exact) in entire region and therefore we can use this information to predict the price of any unsold property (Chica-Olmo, 2007).

The method of geostatistical kriging has some drawbacks. First, although it gives more weight to the surrounding points, it is a global method, that is, like the MRA, it uses the entire dataset to predict the price of unsold properties while as we have mentioned in Section I, the nature and degree of spatial dependence is different in different parts of the region. To tackle this problem, the method of moving window kriging is used so that for any unsold property, we will use the spatial dependence information only in that specific window rather than the global information. Beside prices, there is other property characteristic information available. The method of co-kriging uses other important highly correlated property characteristics in the neighboring properties to predict the price of the unsold property. This method is theoretically sound since the price of a property is not only influenced by the prices of its neighboring properties but also by the characteristics of the neighboring properties. This method adds more difficulty in computational aspects because the spatial dependence information now comes from more than one variable across the region. Co-kriging, however, ignores the characteristics of the unsold property that its price is going to be predicted which can be seen as a drawback of the method. Normally, cokriging is used in the situations where the secondary variables (property characteristics) are observed less sparsely than the primary variables (property prices) which seldom happen in property data.

To deal with this problem, another method called regression kriging (RK) is used that is based on the simple MRA model but with added spatial dependence information. Regression kriging that is usually used in the literature (Dubin, 1999; 2003; Anselin *et al*., 2004) takes the residuals of simple MRA method and performs a kriging on them so that for each unsold property there will be a predicted error. This error will then be added back to the MRA analysis and then price will be calculated for that specific property using its own property characteristics (independent variables). Variations of this method could be invented using the spatial lag or spatial Durbin models.

Local Models

Spatial heterogeneity plays a major role in modeling spatial phenomena because spatial heterogeneity might be more important than the spatial dependence especially in modeling property prices. Local models have been developed to capture spatial heterogeneity; the MRA model is repeatedly regressed in several smaller areas until the region of interest is covered. If the nature of the spatial relationships is different at different places in a study region, we can estimate the coefficients and then do the prediction locally such that the determined relationships are confined in the well defined neighborhoods, called windows.

The windows of local models can be of various forms, shapes, and sizes. The most convenient for property price modeling, however, is windows of irregular-shaped boundaries with varying sizes depending on the distribution of neighboring properties to be included in the windows.

Regression windows may be centered at data points (sold houses) or non-data points (unsold houses). If data points are a lot less, as usually happen in property price modeling, it might be better to center regression windows at data points because of less total computer regression time; If regression windows are centered at non-data points, the advantage occurs during prediction; the center of regression windows, being non-data points, can be predicted directly using the determined coefficients. It is not possible to do this if regressions are centered at data points. Extra work is required to determine in which regression window a predicted point lies and use coefficients of that window in the prediction. A weighted mean is required if the predicted point falls in more than one regression window. In property price modeling, we are convinced that centering regression windows at data points is a better deal.

Window regressions necessitate the use of a subset of data points for each window where these points are the closest to the center of regressions. The issue is how many data points to be considered. If spatial heterogeneity exists in a strict sense in a region of interest, each observation should have a different value of coefficients and a global MRA model produces biased estimated coefficient. Local models reduce this biasness and require the coefficients to be the same for all observations in each window; coefficient values between local areas may vary. This can be achieved by considering observations very close to the regression points. This option, even though produces estimated coefficients with small bias, reduces the effective sample size producing coefficients with large variances and thus unreliable. Considering observations far from regression points may produce estimated coefficients with small variances and increasing reliability but with increasing bias. In line with the bias-variance trade-off is the issue of prediction accuracy. Too few observations produce prediction of lower accuracy but too many observations do not necessarily increase the accuracy of prediction significantly. The bias-variance trade-off in estimation and accuracy in prediction in local models must be effectively handled. We would like to use the optimum number of data points for each window to solve these issues. For the moment, the criterion is the accuracy of prediction and the most widely used strategy is the cross validation; it determines the optimum number of neighboring data points to be included in a regression.

Local models produced *k* sets of coefficients where *k* is the number of regression windows. As a result, local models make local statistics such as local R^2, local Moran, etc. to be available naturally. The *k* sets of coefficients also allow continuous map of coefficients to be made so that the dynamics of regression coefficients can be seen. Local models are mostly appraised in the literature for their ability to prove the non-stationarity of property prices because the different relationships in different parts of a region can be proved through mapping of regression coefficients in the region.

Geographically Weighted Regression

Geographically Weighted Regression (GWR) is the most popular local models. At each regression window, only a subset of observations nearest to the regression point enters the regression and these observations are weighted according to some distance decay functions. Observations near the regression point receive higher weight while observations further from regression point receive lower weight. Due to unequal weighting of observations, the WLS estimation is used instead of the OLS.

Moving Window Regression

Moving Window Regression (MWR) is another version of local models. The only difference between GWR and MWR is in the way weights are assigned to observations that are included in regression windows. Unlike in GWR, all observations that are included in regression windows are weighted equally in MWR. This is to say that observations will influence the subject property by the same amount no matter how far they are from the subject property. This weighting strategy makes MWR loses out to GWR because it contradicts to the theory of spatial dependence and thus make MWR less popular compared to GWR. On the other hand, MWR is simpler to implement because it uses OLS estimation due to equal weighting of observations.

FUTURE DIRECTIONS

Past research segregated spatial dependence and spatial heterogeneity in the effort to produce more accurate prediction. Since spatial effects in inherent in property data comprise both spatial dependence and heterogeneity, future research should focus on the combination of both effects on increasing the capacity of the error reduction in regression analyses.

The focus of spatial autoregressive, local and geostatistical models is primarily on the spatial domain. We know very well that property data have both the spatial dimension and time dimension. The interaction of the time and space on property data cannot be underestimated. Such effects which are called spatiotemporal effects are now being investigated and their feasibility in property price modeling are considered by the researchers. Another trend of research is on the usage of the time based geostatistics or soft geostatistics and model based geostatistics that uses the Bayesian approaches for the increase of prediction accuracy. The Bayesian approaches in the regression analysis are also being used more frequently in the literature.

Another major research area is on the software development for the ease of conduct of the appraisal using thousands of transaction data that are now being increasingly accessible to valuation professionals. Software platforms like R system and Geoda framework have readymade sections for the development of the spatial weight matrices that could used by the other proprietary and non proprietary software (Anselin *et al*., 2004).

CONCLUSIONS

Speed, consistency, and accuracy of mass valuation are now a demand that appraisal communities are challenged with. Using the traditional method of MRA will result in high margin of error and therefore for most cases is unreliable. MRA however provides a benchmark on top of which other methods are built and tested. Most common problems associated with simple MRA are ignorance of spatial effects in the model. Spatial dependence which is the influence of near properties on each other is important and should be somehow considered in MRA. The nature of these effects is not the same everywhere however and this difference will create spatial variability, spatial heterogeneity, or market segmentation. The gust of all of the spatial models is to increase the influence of nearest neighbors or prevention of farthest neighbors to influence the prediction for unsold property. Spatial models aim to improve MRA by adding spatial dependence components to the formula using the connectivity weights either in the response variable (SLM) or error terms of regression (SEM). Geostatistical kriging aims to introduce a new type of prediction using the information inherent to the geographical distribution of price or its relation to the property characteristics of nearest neighbors. Local models try to segmentize the region based on specific windows and predicting for the unsold property based on those windows. The method of MWR gives equal weight to the neighbors influencing a subject property in the windows while GWR imposes spatially varying weights that more closely resembles the data generating process. GWR is useful for ascertaining the degree of spatial heterogeneity in the area.

REFERENCES

Adair, A., & McGreal, S. (1988). The application of multiple regression analysis in property valuation. *Journal of Property Valuation and Investment*, *6*, 57-67.

Ambrose, B. (1990). An analysis of the factors affecting light industrial property valuation. *Journal of Real Estate Research*, *5*, 355-70.

Anselin, L. (1988). *Spatial econometrics: Methods and models*. Springer.

Anselin, L., & Bera, A. (1998). Spatial dependence in linear regression models with an introduction to spatial econometrics. *Handbook of Applied Economic Statistics*, *155*, 1998.

Anselin, L., Ibnu, S., & Kho, Y. (2004). GeoDa: An introduction to spatial data analysis. *Geographical Analysis*, *38*, 5-22.

Anselin, L., & Lozano-Gracia, N. (2009). Spatial Hedonic Models *K. Patterson*, 2009.

Beach, C. M., & MacKinnon, J. G. (1978). A maximum likelihood procedure for regression with autocorrelated errors. *Econometrica*, *46*(Jan. 1978), 51-58.

Belsley, D. A., Kuh, E., & Welsch, R. E. (2004). *Regression diagnostics: Identifying influential data and sources of collinearity*. Wiley-IEEE.

Benjamin, J. D., Guttery, R. S., & Sirmans, C. F. (2004). Mass appraisal: An introduction to multiple regression analysis for real estate valuation. *Journal of Real Estate Practice and Education*, *7*, 65-77.

Brasington, D. M., & Hite, D. (2005). Demand for environmental quality: A spatial hedonic analysis. *Regional Science and Urban Economics*, *35*, 57-82.

Bourassa, S .C., Hoesli, M., & Peng, V. S. (2003). Do housing submarkets really matter? *Journal of Housing Economics*, *12*, 12-28.

Buyong, T., & Valivalo, S. (2010). Modeling residential property prices using geographically weighted regression. *Pacific Rim Property Research Journal, 2010.*

Chica-Olmo, J. (2007). Prediction of housing location price by a multivariate spatial method: Cokriging. *Journal of Real Estate Research*, *29*, 92.

Clapp, J. M. (2003). A semiparametric method for valuing residential locations: Application to automated valuation. *The Journal of Real Estate Finance and Economics*, *27*(Nov. 2003), 303-320.

Cohen, J. P., & Coughlin, C. C. (2008). Spatial hedonic models of airport noise, proximity, and housing prices. *Journal of Regional Science*, *48*, 859-878.

Dubin, R. (1988). Spatial autocorrelation. *Review of Economics and Statistics*, *70*.

Dubin, R. (2003). Robustness of spatial autocorrelation specifications: Some Monte Carlo evidence. *Journal of Regional Science*, *43*, 221-248.

Dubin, R., Pace, R. K., & Thibodeau, T. G. (1999). Spatial autoregression techniques for real estate data. *Journal of Real Estate Literature*, *7*, 79-96.

Fletcher, M., Gallimore, P., & Mangan, J. (2000). Heteroscedasticity in hedonic house price models. *Journal of Property Research*, *17*, 93-108.

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

Gallimore, P., Fletcher, M., & Carter, M. (1996). Modelling the influence of location on value. *Journal of Property Valuation and Investment*, *14*, 6-19.

Ismail, S., Buyong, T., Sipan, I., Hashim, M. G., & Navaneethan, R. (2008). Spatial Hedonic Modelling (SHM) for mass valuation. *International Real Estate Research Symposium (IRERS)*. Kuala Lumpur, Malaysia.

Mark, J., & Goldberg, M. (1988). Multiple regression analysis and mass assessment: A review of the issues. *Appraisal Journal*, *56*, 89-109.

McCluskey, W. J., & Borst, R. A. (2007). Specifying the effect of location in multivariate valuation models for residential properties. *Property Management*, *25*, 312–343.

Militino, A., Ugarte, M., & García-Reinaldos, L. (2004). Alternative models for describing spatial dependence among dwelling selling prices. *Journal of Real Estate Finance and Economics*, *29*, 193-209.

Paez, A., Long, F., & Farber, S. (2008). Moving window approaches for hedonic price estimation: An empirical comparison of modelling techniques. *Urban Studies*, *45*, 1565.

Tretton, D. (2007). Where is the world of property valuation for taxation purposes going? *Journal of Property Investment and Finance*, *25*, 482-514.