Multiple Regression Analysis for house valuation using GIS: a case study

Yalpir Sukran[1] & Bunyan Unel Fatma[2]
[1]Geomatics Engineering Department, Selcuk University, Turkey
[2]Map and Cadastro Programme,
Guneysinir Vocational School of Higher Education,
Selcuk University, Turkey

Abstract

GIS is one of best systems to examine the results after a model is created for AVM. Because it is reused for the analysis of a combination of graphical and verbal data, GIS plays an important role in the creation of real-estate valuation maps.

The purpose of this study was to examine the applicability of Multiple Regression Analysis (MRA) in the determination of house (apartment/house) prices. Data was collected to form a data set from the properties of 348 houses in the Konya–Selçuklu region. The set includes 14 inputs as house attributes and the price of the house is represented as 1 output of MRA. The data is transferred to GIS so that the value map of the working area was obtained.

Keywords: real-estate valuation, Multiple Regression Analysis, Geographic Information Systems, house criteria.

1   Introduction

Real-estate valuation is a field over which not only the developing, but also the developed countries have ceaselessly been working. Real-estate valuation is one of the leading multi-disciplinary studies of today thanks to the ever-increasing demand for quality structures, and to the rearrangement of the real-estate market [1].

The strength of automated valuation methods (AVM) approach lies in large part with the ability to identify and apply the adjusted prices from comparable sales to the valuation of subject properties [2]. AVMs with a less deterministic relation between real-estate values and real-estate attributes may help. AVMs will

facilitate a quicker and optimum estimation of the real-estate values. Although the problems are highlighted, at the moment, Multiple Regression Analysis (MRA) remains the most important theoretical framework in mass appraisal [3]. MRA among AVMs, which is frequently referred to in the literature, is an approach being recognized due to the linear distribution between the values and criteria of the real-estates. For instance, while such criteria, including the surface area, number and fronts of rooms, etc. increase the value of the real-estate, age is the criterion which decreases the price of the same. These criteria have weights in the value by certain ratios. MRA method helps for the realization of a mathematical model, thanks to the estimation of these weights with the help of the data set.

As being a rapidly-developing new computer technology, geographic information system (GIS) is a system, which provides support to the decision-makers in numerous fields. Real-estate-oriented applications (i.e. commercial property investments, real-estate sector, land/plot development, etc.), as having an important role in the national economies, are effectively used in both developed, and developing countries [4]. GIS is applicable due to facilitating the generation of the maps of the real-estate values [5–8]. Value maps being drawn up via this system, which facilitates the constitution of the graphical-verbal relationship, may provide the basis for numerous applications. The determining of factors affecting real-estate value on image can be made easy formation of real-estate value maps [9, 10].

In this paper a mathematical model, which may estimate the actual values of the real-estates (houses) under the market conditions, has been generated via MRA. The results having been attained from this model were examined both statistically, and via performance analyses, and the value maps have thereby been drawn up upon being correlated in GIS environment.

2 Method

It is well known that there are many factors affecting the valuation of real-estates each of which has different effect on the value. The value estimation considering many criteria can be made with multi-regression methods. In the linear regression model, the dependent variable is assumed to be a linear function of one or more independent variables plus an error introduced to account for all other factors (Equation 1):

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} \ldots \ldots \beta_k x_{ik} + u_i \tag{1}$$

In the above regression equation,

y_i : the dependent variable, *(the value of the real-estate),*

$x_{i1}, \ldots, x_{\_iK}$: the independent *or* explanatory variables, *(number of room, age, etc.)* and

u_i : the disturbance or error term. The goal of the regression analysis is to obtain the estimations of the unknown parameters.

$\beta_1, \ldots, \beta_k$ indicates how a change in one of the independent variables affects the values taken by the dependent variable.

3   Application

The region, which is comprised of the structured blocks, is located in Bosna Hersek Quarter of Selcuklu District of the City of Konya in Turkey, 20 km away from the Centrum of Konya (Fig. 1). Students comprise most of its population density. The reason for determining this quarter as the field of study is that, it is located at a little bit of the outskirts of the city, and that it reveals integrity in itself. In other words, there is no structuring of another quarter located at the borders of this quarter.

In consideration of the 14 criteria, obtained from the houses subject to purchase and sale in market conditions, being effective on the value regarding this application, market values of 348 houses have been summed up and arranged in the format of a 14 × 348 matrix.
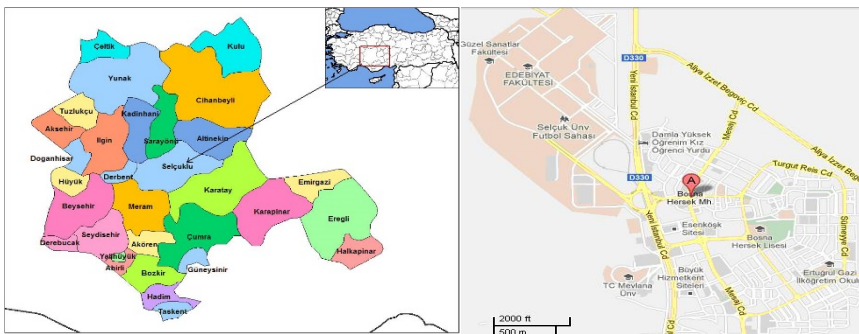

Figure 1:   Outlook of the district, where the field of study is located.

The criteria include: Number of rooms, Total and Number of floors, Age, Type of fuel and heating, Frontal status, Parking lot status, fire escape and cable TV, Distance from share taxi stop, Greenfields, shopping malls and university, Price of the house,.

In this arrangement, the criteria appertained to the real-estate have been assigned as independent variables, and the market value of the real-estate has been assigned as the dependent variable.

3.1 Regression application in residential real-estate valuation

The 348 data was used to form the mathematical model in the regression in order to determine value of the real-estate. The built model and the coefficients can be seen in the following Equation (2)

$$Y = 3.516 + 0.322*X_1 + 0.164*X_2 + 0.009*X_3 - 0.004*X_4 + 0.116*X_5 - 0.055*X_6 \\ + 0.057*X_7 - 0.056*X_8 + 0.249*X_9 + 0.025*X_{10} + 0.001*X_{11} + 0.038*X_{12} \\ + 0.006*X_{13} - 0.049*X_{14} \quad (2)$$

In the order of significance regarding the effects of the criteria on the value according to MRA model, X_1 (total number of rooms), X_9 (fire escape), X_2 (number of floors of the building), and X_5 (fuel type) are seen to be more effective in the value of the real-estate in question in comparison to the remaining criteria. X_4 (age of the structure), X_6 (method of heating), X_8 (parking lot status), X_{14} (distance to the University) are seen to be the criteria having negative effect on the value. It is logical for the variables of X_4 and X_{14} to be negative. That is so because the lower the values thereof decrease, the higher their effects on the value of the real-estate in question increase. If the age of the building (X_4) is greater, it has a lowering effect on the value of the house.

The value of R Square column in the model summary table gives 81% variance for the dependent variable "Yi" of the independent variables (Xn's) In other words, 81% of the price of housing is formed based on Xn's. ANOVA table can be utilized in the evaluation of the model. The value in the significance column of ANOVA table shows that the relationship between these variables provided as $p < 0.01$, and, therefore it was statistically significant. The relationship is formulated in as *F (14.353) = 108.607; p < 0.01* for equation (2).

3.2 Application of geographic information systems

The value having been derived via MRA method has been added to the verbal details of the system, and two separate value map have thereby been drawn up (Fig. 2).

Upon comparing the two maps drawn for the field of study, it is seen from the Fig. 3 that, there are so many similarities between the map drawn via market MRA method, and the map drawn via the subject values of purchase-sales under the market conditions.
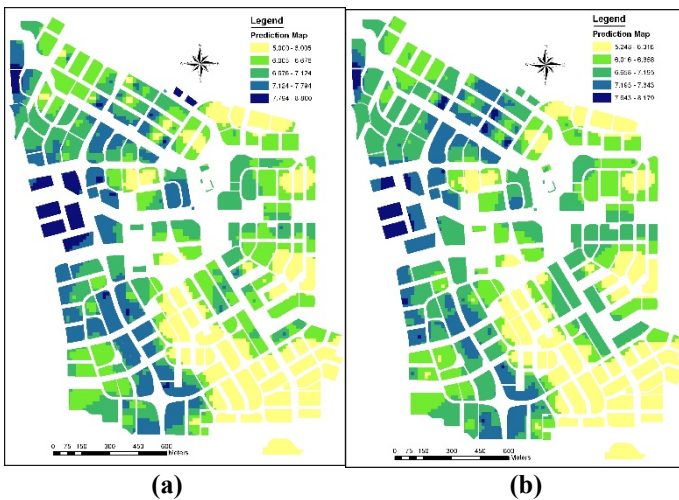

(a)   **(b)**

Figure 2:   Thematic maps having been drawn out of the Market (a) and MRA (b) values.

4   Results and discussion

As a result of the application, the test results of MRA model were primarily discussed. The estimated value derived from MRA model, and the values under the market conditions were compared. Fig. 3 shows the best curve fitting with zero intercept (y = ax line) used to define the level of accuracy between the collected market values and MRA model. According to the distribution point of estimated values and market values, MRA model in term of slope and R^2 indicates good performance and MRA model the slope of the best fits is close to 1 in data set.
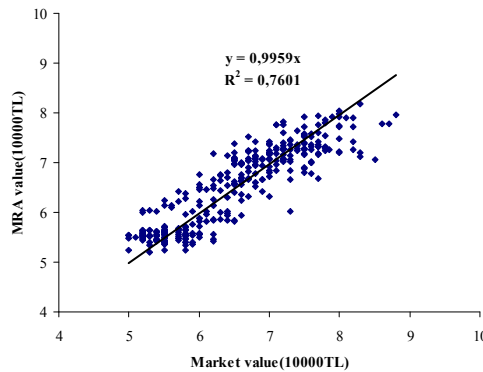

Figure 3:   The regression line between the unit market values and MRA values.

Equation 3 was applied for the absolute error calculation and Equation 4 for the Standard deviation.

$$\text{MAPE}_i = \left(\frac{\sum_{i-1}^{n} \frac{|x_p - x_i|}{x_p}}{n} \right) .* 100 \tag{3}$$

$$SD = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \overline{x}_i)^2}{(n-1)}} \tag{4}$$

x_p :Market value, x_i :Value of the model, i : {1,2,3….n}, n :Total number of the residential real-estates in the data set, SD : Standard deviation of the approximations in the data set, MAPE_i : Mean Absolute Percentage Error.

It is understood from the comparisons therein that, estimates of MRA model are successful. Ratios of approximation of MRA model to maximum, minimum, average, and SD values are above 90% in general terms. In view of these ratios in Table 1, MRA may be found as successful.

Table 1: Value ranges.

	Market value (TL)(x10⁻⁴)	MRA value (TL)(x10⁻⁴)	Percentage %
max	8.8	8.199	92.95
min	5	5.201	96.18
ave.	6.519	6.526	99.93
SD	0.892	0.807	90.52

MAPE% was found as 4.8%. In consideration with the fact that the market value of the real-estate incorporates the subjective approaches, it may well be said that, the MAPE ratio attained for MRA method is within an acceptable range.

5   Conclusion

The applicability of MRA method in real-estate value estimation was investigated in this study. The purpose of the study was to integrate GIS system of MRA method and to production house value map. It has been discussed in the study that, integration of the model generated via MRA method to GIS would allow the said model applicable not only in the drawing up of value maps, but also as a real-estate information system, on the condition to have all the required deficiencies eliminated. A database regarding the real-estates should be generated for the sound operation of the system. Generation of the database may be done more easily via computer technology. The generated database and the information system concept allow for the conduct of the analyses, and the values of the real-estates may thereby be determined approximately.

Choice of MRA method among AVM methods due to ease of understanding and operation, it may be utilized as the basis for drawing up of thematic value maps via GIS, and for all the processes (taxation, expropriation, banking transactions, etc.) which require real-estate valuation.

Acknowledgement

References

[1]   Doner, S., Alkan, R.M. CBS Destekli Taşınmaz Mal Değer Haritalarının Oluşturulması, TMMOB Harita ve Kadastro Mühendisleri Odası 13. Türkiye Harita Bilimsel ve Teknik Kurultayı, Ankara (in Turkish). 2011.
[2]   Thompson, J.F. Data Issues Involved with the Application of Automated Valuation Methods: A Case Study Mass Appraisal Methods: An International Perspective for Property Valuers, pp. 25-27. 2008.

[3]   Kauko, T., d'Amato, M. Introduction: Suitability Issues in Mass Appraisal Methodology, Mass Appraisal Methods: An International Perspective for Property Valuers, pp. 1-24. 2008.

[4]   Cete, M., Yomralioglu T. Türkiye İçin Bir Arazi İdare Sistemi Yaklaşımı, *HKMO*, 100, (1) (in Turkish). 2009.

[5]   Zeng, T.Q., Zhou, Q. Optimal Spatial Decision Making Using GIS: A Prototype of a Real Estate Geographical Information System (REGIS), *Int. J. Geographical Information Science*, 15, 4, pp. 307-321. 2001.

[6]   Tiryakioglu, I., Erdogan, S. CBS Destekli Taşınmaz Değerlemesi: Afyonkarahisar Örneği, 4. Cografi Bilgi Sistemleri Bilişim Günleri, 13-16 Eylül, İstanbul, Türkiye (in Turkish). 2006.

[7]   He, B., Tong, X., Chen, P., Qu, B. Application of Mobile GIS in Real Estate Dynamic Monitoring, International Workshop on Education Technology and Training & 2008 International Workshop on Geoscience and Remote Sensing, Shanghai, China. 2008.

[8]   Doner, S. CBS Destekli Taşınmaz Mal Değer Haritalarının Oluşturulması, Y. Lisans Tezi, İTÜ Fen Bilimleri Enstitüsü, İstanbul (in Turkish). 2010.

[9]   Pagourtzi, E., Nikolopoulos, K. Architecture for a Real Estate Analysis Information System Using GIS Techniques Integrated with Fuzzy Theory, ERES Meeting 2–5, Giugno Milan, Italy. 2004.

[10]   Bahar, M.E. Taşınmaz Değerlemesinde CBS'nin Kullanım Olanakları, *Yüksek Lisans Tezi*, İTU Fen Bilimleri Enstitusu, İstanbul (in Turkish). 2007.