Estimating the residence price by linear regression model and Geographical Information Systems (GIS)

Tarik Turk *, Murat Fatih Tuna b, Olgun Kitapci c

Abstract

Owning a residence is also considered as an investment tool, determining the price of a residence with the desired properties has become one of the most important questions to be answered in social life. In this study, price estimations of residences located in nine of the central districts of Ankara city (Turkey) were carried out via multi linear regression model and geographical distributions of these residences were revealed on GIS environment to perform various query, spatial analysis and documentation operations. In addition, thematic maps regarding residence prices in the study region were produced.

Keywords: estimating the residence price; Geographical Information Systems; online marketing; regression analysis

* ADDRESS FOR CORRESPONDENCE: Tarik Turk, Department of Geomatics Engineering, Faculty of Engineering, Cumhuriyet University, 58140, Sivas, TURKEY
    E-mail address: tarikturk@gmail.com / Tel.: +90-346-219-10-10 Ext: 2438
1. Introduction

Accommodation which is one of the most basic needs of people increases the tendency of residency ownership in urban life. On the other hand, buying a residence is considered to be a long term investment tool by many people. Therefore, price differences are bound to arise due to the facts that residencies are a physical need; they are used as an investment tool and have become more and more authentic due to aesthetic differences. Hence, there are various studies in literature that give answers to questions such as which residence is more logical to buy at what price and which factors affect the price of a residence to what extent.

Technological developments are important elements that are effective on the synthesis and transfer of information. Information can be considered to be an element that shapes our lives and can only be transferred to all kinds of users via a well-functioning information infrastructure. GIS provides a framework for the acquisition and organization of locational data and can show this and any other related information after an analysis carried out on a single plane (Miller, 2007).

Ozturk and Fitöz (2009) have put forth the determining parameters regarding the supply and demand of the residence market in Turkey via regression analysis and have determined that the increase in the residence prices has been in parallel with the increase in the residence demand. Holly, Pesaran and Yamagata (2011) have examined the distribution of the residence prices in the dynamic residence system based on location and time in the United Kingdom. Kiel and Zabel (2008) have put forth the hedonic residence price equation by using the 3L (Location, Location, Location) approach in the determination of residence price. Roche (2001) has mentioned the existence of basic and non-basic elements affecting residence prices in Dublin after carrying out various measurements. In conclusion, Roche has stated that there might be a speculative effect on the increase of the residence prices in Dublin which has been frequently discussed in media. Kusan, Aytekin and Ozdemir (2010) have used fuzzy logic analyses to estimate the residence prices at different regions of Eskisehir and have determined that the values they estimated were in accordance with the real values. Bin (2004) has tried to estimate the residence prices by comparing the parametric and semi-parametric regression analyses. In conclusion, it has been put forth by Bin that semi-parametric regression analysis will be beneficial in estimating the residence prices. Dumm, Sirmans and Smersh (2011) have examined the effect of zoning laws on residence prices. It has been determined as a result of this study which included the effect of the rough period during 2004-2005 that stricter zoning laws increase the sales price. Abdulai and Owusu-Ansah (2011) have carried out a study on the parameters that affect residence prices in the city of Liverpool, England. As a result of this study during which residence data between January 1990 and December 2008 was subject to cross section and time series analyses, it was determined that the parameters preferred in residences built prior to and after 2000 were different. As a result of the study during which non-residential consumption and residence expense data were used, Piazzesi, Schneider and Tuzel (2007) created an equilibrium model to estimate the residence prices. Ozsoy and Sahin (2009) have used the classification and regression tree (CART) approach in order to examine the essential parameters that affect the residence prices in Istanbul. As a result of this study, the area, elevator, security, central heating system and facade parameters have been determined to be the most important criteria for the residence prices in Istanbul.

Selim (2008) has explained the parameters determining the residence prices in Turkey via the hedonic regression model. As a result of this study, it has been emphasized that water system, pool, residence type, number of rooms, size of the residence, location characteristics and building type variables are the most significant variables. Dekkers and Koomen (2008) have carried out the hedonic residence price analysis to estimate the residence prices in the Ranstad region of Netherlands and have integrated their studies with GIS. In conclusion, they have determined that unrestricted view in a residence increases their average prices by 4 % to 8 %. In addition, they have also determined that open areas that are within 25 meters of the residence make a strongly positive effect on the residence prices.
Lake et al. (1998; 2000) have used GIS for the first time in real estate evaluation and residence pricing. This study has put forth that highways, rail road’s and industrial buildings have a negative impact on the view of the residence thereby decreasing its price Schernthanner and Asche (2010) have carried out the 2009 residence market analysis for the city of Potsdam in Germany by using GIS. Yu, Han and Chai (2007) have used GIS and three dimensional modeling and regression analysis together to examine the effect of sea view on the residence prices for the settlement areas near the eastern coast of Singapore. As a result of this study, they have determined that the prices of residences in this region that have a view of the sea are 15 % more expensive than those that do not have such a view.

Whereas in this study, multi linear regression analysis has been used to estimate the residence prices in the city of Ankara (Turkey) and the geographical distributions of the residences have been integrated on the GIS environment to carry out various query, analysis and documentation operations. In addition, user interface program has been developed via the GIS software used thereby allowing users to carry out all queries, analysis and documentation functions easily and automatically. Thematic maps regarding the residence prices in the study region have also been generated and presented to the users. Thus, a GIS based system has been put forth which can be used by those who will purchase or sell residences.

2. A Case Study in the City of Ankara (Turkey)

2.1. The Study Area

In this study, nine central districts (Akyurt, Altindag, Cankaya, Etimesgut, Golbasi, Kecioren, Mamak, Sincan, Yenimahalle) within the city boundaries of Ankara have been selected as the pilot study area. In addition to being the capital city, Ankara is also important since these nine districts have a population of about 4.250.188 based on 2011 data. As a matter of fact, the residence prices in these nine districts differ. In addition, these districts that have been selected as the pilot study area continue to develop due to constant immigration. On the other hand, the report entitled “Real Estate Sector in the Vision of 2023” published by The Association of Real Estate Investment Companies in 2012 foresees that residence demand will increase in these nine districts of Ankara along with the increase in population. Another reason for selecting these districts as the pilot study area in this study is that there is not sufficient data for residence sales in other districts in the virtual marketplace www.sahibinden.com. All of the aforementioned reasons indicate why these nine districts of Ankara were selected as the study area.

2.2. Data Used

In this study, the map that shows the borders of these nine districts located within the city limits of Ankara has been used as geographical data. Whereas the non-graphic data have been obtained from the data of the online marketplace “www.sahibinden.com” regarding the residences being sold in the pilot study area for the period between December 11, 2011 and April 23, 2012. This online market platform (www.sahibinden.com) founded in 1999 in Turkey is one of the first websites that provide ad search opportunities via the internet. Since its foundation this website has been the most visited ad and shopping platform of Turkey. On the other hand, five year neighborhood and district level population data obtained from the Turkish Statistical Institute for the period between 2007-2011 has been used due to observe the population development in the study area.

3. Methodology

This study consists of two stages which are the statistical evaluation of the obtained data and the carrying out of the required query, analysis and documentation operations following data integration on GIS environment.
3.1. Statistical Evaluation of Data

At this stage, first the properties of the houses have been examined in order to determine the factors that affect the prices of these houses. The database parameters have been considered while carrying out correlation, regression and difference analyses (Kolmogorov-Smirnov, Mann-Whitney U and Wilcoxon Rank Test). Some of the data used in the preparation of the database contain character data while some contain numerical data. Hence, ordinal scale transformation has been carried out via grouping method in order to carry out statistical analysis on the character content parameters.

Although data derived from www.sahibinden.com online market platform are sufficient for test of normality, statistics had to be conducted with non-parametric tests because of parameters’ properties. Moreover, entity of extreme values and nonnumeric expressions in some variables necessitated grouped analysis of these variables. As a general consumer perception, facades of residence as well as its properties and properties of apartment where is the residence involved in were examined with dual and multiple comparison tests whether they make difference on the residence price or not. Comparisons calculated according to the confidence interval of %95. As a result of dual comparisons conducted in 26 districts (for facades 78x4 comparisons), only 15 comparisons make significant differences on the price. The same situation was seen for parameters of heating types, security, fire escape, play field, parents’ bathroom, laundry room and vinyl floor covering. In spite of the fact that residence properties that is put forefront by seller exist in the data set, the fact that every variable has an effect on the price is unthinkable. In order to investigate this situation, feature analysis was conducted to reveal which parameters have more effect upon prices. It was hereby aimed to minimize waste of time in the course of conducting analysis with multi-dimensional data. Important parameters obtained by feature analysis and their feature values are 0.95 and higher.

Firstly, the priority levels of the parameters were tried to be calculated using data mining package software. It was determined that the model became meaningless and that the parameters were inconsistent when all the parameters were included in the model. In addition, various classification, grouping and multi regression analyses used in data mining have been used to estimate the residence prices in the pilot study area.

The numbers of samples for the districts in the study region have been determined in proportion with their population. Thereby, the proportional value of the population data for each of the districts in the study region have been calculated and data regarding the sales price and relevant parameters of the same ratio of residences have been obtained from the online marketplace platform “www.sahibinden.com”. As a result, a dataset for 1000 residences on sale has been created.

The significance levels of the parameters in the dataset have been determined and an optimum model has been tried to be obtained via multi-linear regression analysis in the SPSS statistical data evaluation software in order to estimate the price. At this point, an optimum regression model could not been obtained due to largeness of the price and area interval. Hence, the number of data has been constrained by considering the area criterion along with the price in order to reach the most suitable regression model and a more correct and consistent result. When the data intensity graph regarding the residence price and area for the thousand residences is examined, it is observed that residences with area values ranging between 75 and 200 meter square and prices ranging between 75.000 TL and 250.000 TL are more common. In this regard, the number of data has been decreased to 722. Linear regression analyses have been carried out using this generated dataset and price estimation criteria have been put forth.

Correlation table has also been generated for each of the parameters used in the dataset. Afterwards statistically significant parameters were included in the model and regression analyses were carried out. The regression analyses have been examined and statistically insignificant variables have been taken out of the model. In entered method, SPSS package program includes variables to models hardly when regression models are created. This situation does not serve decision makers’
perception of significance alterations occur in case of discharge of variables. For that reason, regression models were constituted by means of forward selection. Lastly, the model which is given in detail in Table 1 has been obtained.

As is seen in Table 1, parameters take place in the model and shown by $\beta_i$ (i=1-9) are such as to show internal and external features of residences in data set. However, parameters of districts were derived from initial data set and they are dummy variables which show effect of each district on the price.

Thus, the following equation has been obtained for the estimation of the price of the residences in the pilot study region considering the parameters listed in the above table.

Table 1. Summary of Obtained Regression Model.
Residence Price = 667,329xβ₁ + 710,792xβ₂ + 13358,141xβ₃ + 2119,075xβ₄ + 7760,370xβ₅ + 
7042,328xβ₆ + 10036,866xβ₇ + 9309,542xβ₈ + 21067,837xβ₉ + β (town coefficient)i
(town name)i

3.2. Integration of Data on the GIS Environment

In the previous section, the current and generated data along with the geographical data displaying
the boundaries of the nine districts in the pilot study region and the periodical population data of
these districts have been integrated on the GIS environment using ESRI ArcGIS 9.3 software.
Afterwards, user interface software has been developed in order to carry out various query and
analysis operations (Figure 1). Thus, it has been ensured that the system can be used easily and
efficiently by everyone.

Figure 1. The user interface program developed and their functions

When the “Calculate the Price of Residence” button from among the developed user interface
software is pressed, the user sees the screen on which significant parameters that statistically affect
the residence price as a result of the regression analysis are displayed. Here, the price of the residence
selected by entering the relevant parameters is calculated and displayed automatically after pressing the “Calculate The Residence Price” button (Figure 1). When information regarding the queried residence is desired to be known, “Show the Residences Queried in the Map” button is pressed and the district is selected on the map after which the relevant information is displayed when the “Show the Result of the Town Query” button is pressed on the “Query Town” form (Figure 1). Whereas in the other developed user interface software, “Query Town” button is pressed to carry out query according to the district name after which information regarding the queried districts is obtained and the queried districts are automatically selected on the map.

4. Results and Discussion

The $R^2$ and standard deviation values for the model obtained as a result of regression analysis have been determined as 0.96 and 28,534 TL respectively (Table 1). It is thought that there are two reasons for a standard deviation value that is obtained above the expected value. One of the reasons is the subjective effects on real residence prices. Residence owners who wish to maximize their profits from the residence sales can sometimes set their residence prices above the expected values. In this regard, it can be thought that users of the online marketplace platform www.sahibinden.com may be demanding more expensive residence prices than the prices that should be asked. Secondly, the geographical, social and cultural dimensions of the districts in the study region can have significant effects on the residence prices. For instance, when a residence with the same properties is very cheap in the Akyurt district, it can be quite expensive in the Cankaya district.

In this study, it has been determined that parameters such as the area of the residence, central system, the number of floors in the building, the elevator, parent’s bathroom, built-in closet, built-in kitchen and the district of the residence are important criterion. Only the “Building Age Group” parameter has not been excluded from the model despite the fact that it was insignificant. The reason for this is that it does not reflect the perception for residences on sale in the central districts of the city that “older residences are cheaper”, because Cankaya and Yenimahalle districts are among the oldest and most central districts of Ankara. On the other hand, Cankaya district is one of the most luxurious districts of Ankara. In addition, Cankaya has the highest number of samples in the database since it has the second largest population in the study region. In this case, the price of a new residence built with the same features as a residence in Cankaya may sometimes be cheaper than an old residence in Cankaya. The reason for this is that old residences in the central districts such as Cankaya and Yenimahalle have been built close to the center of the district. These areas that have shopping malls in addition to being very close the market with increasing population, social and cultural life have over time become centers of attention for residence selection. Hence, it should be considered that old residences located at the center of such districts may be more expensive than new residences built in districts that are away from the center. Coefficients regarding the regression model have been given in Table 2.
The prices of the residencies in the study vary due to the differing properties such as the central location of the district, its history and structural properties. As was expected, residence prices at some districts can be above or below the average residence price in the city of Ankara. According to the model obtained as a result of regression analysis, when residences in the Cankaya district are compared with residences in other districts it is observed that the Cankaya district has the most expensive residence prices.

In this model obtained as a result of regression analysis, the built in kitchen parameter has the greatest effect on residence price. When all other properties are constant, this property alone increases the residence price by about 21.067 TL. This parameter is followed by the central heating system which increases the residence price by about 13.358 TL. Parent’s bathroom which increases the residence price by 10.037 TL is in the third place.

The unit meter square price of the residences in the city of Ankara has been calculated as 667 TL. According to the Estimated Building Unit Costs declaration issued on 28.04.2012 by the Ministry of Environment and Urban Planning in the Official Gazette, the unit meter square price for the 4th Class A Group buildings encompassing residences has been determined as 615 TL. The fact that the meter square price obtained as a result of the regression analysis carried out within the scope of this study is close to the meter square price printed in this official document indicates that the study is quite realistic.

The average price of the residences in the Cankaya district are the highest despite the fact that they are the oldest (Average: 16 years old) among those at other districts. The reason for this is that this district has a central location in Ankara and that many public offices and private business centers are located here. At this point; it can be thought that it is easy to reach businesses, schools, hospitals, shopping malls etc. in comparison with other districts.

The fact that the age average of the houses in Cankaya is greater shows that the apartments that are built have more personal taste and historical features. Among these districts, it is Cankaya that has

---

**Table 2. Table of Coefficients of the Regression Model obtained**

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Unstandardized Coefficients</th>
<th>Standardized Coefficients</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>B</td>
<td>Standard Beta</td>
</tr>
<tr>
<td>Square meter</td>
<td>667,329</td>
<td>50,275</td>
</tr>
<tr>
<td>Building age group</td>
<td>710,792</td>
<td>890,258</td>
</tr>
<tr>
<td>Central heating</td>
<td>13358,141</td>
<td>3955,596</td>
</tr>
<tr>
<td>Floor number</td>
<td>2119,075</td>
<td>550,288</td>
</tr>
<tr>
<td>Elevator</td>
<td>7760,370</td>
<td>2909,911</td>
</tr>
<tr>
<td>Car park</td>
<td>7042,328</td>
<td>2411,724</td>
</tr>
<tr>
<td>Parents’ bathroom</td>
<td>10036,866</td>
<td>3280,391</td>
</tr>
<tr>
<td>Built-in wardrobe</td>
<td>9309,542</td>
<td>2625,244</td>
</tr>
<tr>
<td>Embedded kitchen</td>
<td>21067,837</td>
<td>3118,639</td>
</tr>
<tr>
<td>Akyurt</td>
<td>-6315,033</td>
<td>21286,951</td>
</tr>
<tr>
<td>Altındag</td>
<td>18006,971</td>
<td>7167,353</td>
</tr>
<tr>
<td>Cankaya</td>
<td>56356,439</td>
<td>7615,143</td>
</tr>
<tr>
<td>Etimesgut</td>
<td>5448,636</td>
<td>7483,670</td>
</tr>
<tr>
<td>Golbasi</td>
<td>20602,474</td>
<td>9675,327</td>
</tr>
<tr>
<td>Kecioren</td>
<td>22660,106</td>
<td>7086,530</td>
</tr>
<tr>
<td>Mamak</td>
<td>7638,655</td>
<td>6964,318</td>
</tr>
<tr>
<td>Sincan</td>
<td>1206,577</td>
<td>7617,922</td>
</tr>
<tr>
<td>Yenimahalle</td>
<td>24341,083</td>
<td>7872,355</td>
</tr>
</tbody>
</table>

The highest probability to have a built-in closet (~%36). Whereas more modern residences with parent’s bathroom are seen at Yenimahalle (~%37).

The prices of all residences have been calculated using the residence price equation put forth as a result of the regression analysis. These calculated values have been compared with the real residence prices. As a result of this comparison, it has been determined that there is a difference of 13 TL between the average real prices of the residences and the prices calculated using the model. In addition, 25 sample residence price data have been randomly selected at the 2011 population ratio for the districts in the study region. The residence sales prices of these selected samples have been compared with the residence sales prices calculated using the linear regression model. The prices differences obtained vary between 84 TL and 14.159 TL.

On the other hand, a GIS based system has been created as part of the study scope and all data has been integrated. Various user interface software have been developed. Parameters that affect the residence price have been selected using this software to calculate the residence price automatically, queries about the district of the residence are made and general information regarding the district is presented to the user. In addition, the distribution of the average prices for the residences in the study region with an area of 120 meter-square have been presented to the users via thematic maps that are a form of visual communication tool (Figure 2).

![Figure 2. Geographical distribution map of the residence (120 square meters) price in the city of Ankara.](image)

5. Conclusions

Accommodation, which is one of the most basic needs of humans take on the form of getting a residence in the modern urban life. On the other hand, purchasing a residence is seen as a long term investment by many people. This increases residence demand but also makes this a problem that
should be solved. Today, questions such as the geographical location of the purchased residence, its relationship with social areas and the sales price of the residence can be answered via GIS.

In this study, residence price estimations in the city of Ankara (Turkey) have been carried out using multiple linear regression analysis. It has been determined that the area of the residence, central system, the number of floors in the building, the elevator, parent’s bathroom, built-in closet, built-in kitchen and the district of the residence are the most important criterion that affect the residence price. In this process, information regarding residence prices has been presented to the user via thematic maps by making use of GIS. In addition, user interface software has been developed via GIS software in order to make it easier for the users to carry out queries, analyses and documentation operations. In addition, there is no other residence price estimation study using both statistical and geographical analyses carried out for the city of Ankara which is the capital city of Turkey and has a quite high population density. In this regard, it is thought that this system will guide both the residence buyers and the residence sellers as a decision support system. All these factors put forth that this is an original, helpful and useful study.

The results obtained in this study can contribute to the design of new residences in Ankara. Because this study also presents information as to which parameters are considered during residence selection and the ratio with which these parameters fulfill the residence demand of people.

References


