Expert Systems with Applications 37 (2010) 1808–1813

Contents lists available at ScienceDirect

Expert Systems with Applications

journal homepage: www.elsevier.com/locate/eswa

The use of fuzzy logic in predicting house selling price

Hakan Kuşan *, Osman Aytekin, İlker Özdemir

Eskişehir Osmangazi University, Department of Civil Engineering, 26480 Eskişehir, Turkey

A R T I C L E I N F O

Keywords:
Fuzzy logic
House price
Local aspects
Prediction model
Real estate market

A B S T R A C T

In this paper, a new grading model has been developed for prediction of the selling price of house-building. Fuzzy logic systems, considering the city plans, the nearness to cultural, medical, training and educational buildings, the public transportations systems, the other environmental factors and the increased technological upgrading deals with information about construction, have been employed in order to construct the model and achieve the aim. Such factors are used as the inputs. Besides, a questionnaire application including these factors has been applied to determine the values of fuzzy training and testing sets. In this way, the constructed model has been applied to the prediction of selling prices of houses located in different regions of Eskişehir city in Turkey. The predicted values and real selling prices determined by selling market have been compared with each other. Consequently, real selling price of house-building has shown variety with respect to the regional aspects and salesmen. The suggested fuzzy logic model can be capable and usable for similar applications.

1. Introduction

Residence is kind of a wide phenomenon process, beginning with supply of refuge for people and being extended to be one of the most important financial tools, and presents a very large perspective in order to be investigated. In the housing market in which house services are taken place with the mechanism of supply and demand, changing the housing prices have been the interests of the governments, the managers and the individuals due to their influences on the socio-economic conditions and they have also other important effects on the national economic conditions. The housing price and the factors causing the irregular change of this price have been investigated for a long time by many experts from different disciplines such as economists, real estate investors, geographers, however, urban planners and politicians have begun to study related to this subject after association of the urban area usage with the housing market dynamics.

Housing price is a determination related to the residence properties, and this determination is not just the total of the residence unit itself and the constitutional properties of the residence, on the other hand, it is also the combination of the region's properties and location. In the determination of housing price, the consumers' appraisal of the residence unit's constitutional, physical and environmental qualities, and their same appraisal of neighboring properties also play effective parts. Besides, expectations of high capital income from housing investments can increase the demand for house sales which cause the possibility of the high volatility in housing prices. Since the supply of housing cannot adjust this demand in the short run, housing prices will strongly increase. Furthermore, the housing market can be influenced by other variables as macro-economic variables, spatial differences, characteristics of community structure, and environmental amenities (Kim & Park, 2005).

The valuation of real estate is required to provide a quantitative measure of the benefit and liabilities accruing from the ownership of the real estate. Valuations are required, and often carried out, by a number of different players in the marketplace such as real estate agents, appraisers, assessors, mortgage lenders, brokers, property developers, investors and fund managers, lenders, market researchers and analysts and other specialists and consultants. Market value is estimated through the application of valuation methods and procedures that reflect the nature of property and the circumstances under which the given property would most likely trade in the open market (Pagourtzi, Assimakopoulos, Hatzichristos, & French, 2003).

In literature, different methods are suggested to estimate the housing market value. Pagourtzi et al. have defined these methods in two different groups as traditional and advanced (Pagourtzi et al., 2003). In this study, the methods depending on several comparison methods such as a direct capital comparison in its simplest form in order to determine market value or the observation sets providing the determination of a regression model have been called as "Traditional". Comparable method, investment/income method, profit method, development/residual method, contractor method, multiple regression method and stepwise regression

* Corresponding author. Tel.: +90 222 239 3750; fax: +90 222 239 3613.
E-mail address: hkusan@ogu.edu.tr (H. Kuşan).

0957-4174/$ - see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.eswa.2009.07.031

method can be given as examples for traditional methods. Additionally, the other methods simulating the thought processes of the market players in which the estimation of the exchange point is being attempted have been called as "Advanced" according to this study. Artificial neural networks (ANN), hedonic pricing method, spatial analysis method, fuzzy logic (FL), and Autoregressive Moving Average (ARMA) method can be said as the examples of advanced methods.

Most of the price studies have been performed by using the other methods based on hedonic method and multiple regression analysis. Basically, these methods are generally used and suitable for the certain estimation of the relation between prices and several variables. Nevertheless, these techniques can lead problems if the process of pricing is extended to contain the points of view as outliers, non-linearity, spatial and other kinds of relations between observations, discontinuity and fuzziness. Therefore, new lookups are being continued determining housing price by profiting from recent computer technologies. In this way, some approaches, presenting the influences on the real estate by analyzing many criteria for the determination of its price and revealing the best combinations according to these influences, have been developed. Dilmore may be the first expert who accepts the potential practice of FL for the properties of real estates. For instance, one of the studies performed by Dilmore has indicated many important points about FL (Dilmore, 1993).

In this study, it is aimed to predict the house unit prices (UPs) by using FL. The factors affecting the house prices have been determined and chosen as house factors, environmental factors, transportation factors and regional socio-economic factors. After then, these factors have been rearranged into related different groups to explain and model their effects on the house prices in detail. The data about these factors to be used in FL model have been obtained from a prepared questionnaire application. One hundred and sixty of these data have been used for training set and 40 of them have been used for testing set in the constructed FL model. UPs obtained from the model and the real UP values have been compared to each other to determine the precision of the FL model suggested in this study.

2. Related previous research

Many researches have been conducted on the investigation of the factors affecting the housing prices and the relations between them. The first study about the effective factors can be accepted as Alonso's hypothesis explaining the relations between in situ values and the usage According to Alonso, if the spatial balance is existing, housing prices are decreasing since the location is getting far from the trade centers (Alonso, 1964). Muth, in his study about the housing prices at Chicago, has pointed out that the trade centers possibly far from the town center, the distance to social and cultural centers, transportation system and social structure of neighborhood have also some effects on the housing prices together with the properties of residence and the distance to town center (Muth, 1969). Wabe has studied the effects of the properties residence and urban on pricing by using regression equations. The objective of the study is to determine average residence qualities for each town observed (Wabe, 1971). Kain and Quigley have investigated dependent evaluation of quality variable of housing pricings and the properties of residences at United States of America in their study (Kain & Quigley, 1970).

The regression of housing prices evaluating the marginal contribution of the properties of a residence and the neighborhood relations is defined as the implicit and hedonic prices. Hedonic pricing method model depends on the Lancaster (1966). Since this theory has been adapted to the house market by Rosen, residential hedonic method commonly to be used as an assessment tool for the market and urban analysis. The Rosen's theoretical study about hedonic house prices is an exactly detailed investigation related to the housing market, house prices and residential properties showing the way to experimental studies. In the mentioned study, a model has been explained the housing prices based on a great many of variables by hedonic price theory. Both the choices of salesmen and producers have also included besides the nature and the mean of market equilibrium (Rosen, 1974). Brown and Rosen, has conducted a study developing the Rosen's theoretical model and reflecting the supply–demand functions and market equilibrium to the hedonic price analysis (Brown & Rosen, 1982). Rabiega et al. have researched the effect of public housing project constructed nearby the present houses in Portland on the house prices of them by using hedonic price method (Rabiega, Lin, & Robinson, 1984). Stevenson has reexamined heteroscedasticity in hedonic house price model in his study by using the average ages of houses in Boston as data. The obtained results have supported the evidence of heteroscedasticity regarding the house age in the previous findings (Stevenson, 2004). Bin has estimated a hedonic price function using semi-parametric regression. The performance of price estimation has been compared to the conventional parametric models. In order to take the location attributes of houses into account, the data from geographic information system (GIS) has been used. The results revealed that the semi-parametric regression shows better performance in both in-sample and out-of-sample price predictions and it can be used for measurement and prediction of house prices (Bin, 2004). Kim and Park have made a study defining the spatial pattern of housing price changes and their determinants in Seoul and its neighbor towns. The results of the cluster analysis have noted that the spatial pattern of housing price change rates is not correlated with house prices (Kim & Park, 2005). Fan et al. have used the tree decision approach for evaluating the relationship between the house prices and the house characteristics (Fan, Ong, & Koh, 2006).

In recent years, ANN and FL approaches have been used as alternative tools to model systems of conventional property value. Din et al., in their study, they have discussed the standard linear regression model including ordinal variables for measuring environmental quality as the reference model and they have found that price indices of ANN models exhibit a similar behavior. However, it has also concluded that the detailed price behaviors of different models show significant differences depending on the input choices of environmental variables (Din, Hoesli, & Bender, 2001). Selim has compared the hedonic regression and ANN model to each other for determining the house prices. 2004 Household Budget Survey Data for Turkey has been used as the document for the data set. At the end of the study, by reason of hedonic regression's non-linearity, it has been explained that ANN can be a better alternative modeling technique in the determination of house prices in Turkey (Selim, 2009). In another study, the fuzzy neural network prediction model based on the hedonic price theory includes a database storing hedonic characteristics and coefficients affecting the real estate price level from the recently sold typical projects which are reflecting the local environment. The experimental results of the study have shown that fuzzy neural network prediction model has a great ability for the function approximation and available for real estate price prediction with respect to the quality of attainable data (Liu, Zhang, & Wu, 2006). Lokshina has compared multi regression, ANN, FL with each other. In the evaluation of real estate price, the applicability of ANN and FL has been proved and determined that appropriate results obtained by using artificial intelligence methods. Moreover, it is concluded that the performance of the multi regression application for house prices is quite well (Lokshina, Hammerslag, & Insinga, 2003).

3. Fuzzy logic

Knowledge sources such as complexity and uncertainty generally appearing in different forms are defined as Fuzzy. The concept of fuzzy set, replacing Aristotelian logic which has only two possibilities, was preliminarily asserted by Zadeh (1965). The concept of fuzzy set has developed rapidly by both Zadeh himself and indefinite researchers. Real applications of this concept have been successfully actualized at the same time.

People agree each other by talking with linguistic data. It is possible to take linguistic uncertainties in daily spoken language into account by using FL during the modeling and calculating processes. The main area which fuzzy systems are utilized is how it would be thought in order to reach solutions in the case of such information existences. It is attempted to model any problems approximately and control the problem with mathematical in complex solutions by FL (Şen, 2004).

FL concept provides a natural way of dealing with problems in which the source of imprecision is the absence of sharply defined criteria rather than the presence of random variables (Şen, 1998). Fuzzy set theory provides a systematic calculus related to such linguistic information and can perform numerical computation by using linguistic labels stimulated by membership functions.

3.1. Fuzzy logic inference system

Fuzzy inference is the real process of mapping from a given set of input variables to an output relied upon a set of fuzzy rules. The main process of a general fuzzy inference system (FIS) includes four activities called as fuzzification, fuzzy rule base, fuzzy inference engine and defuzzification (Huang & Chiu, 2009). Fig. 1 shows a framework of FIS. Fuzzification transforms the feature value of the effort drivers into proper linguistic fuzzy information. Fuzzy rule base stores the knowledge and rules for deriving the outputs. These rules are expressed in the If–Then format. Fuzzy inference engine takes the human feelings, thoughts and logical inference into account in order to obtain a reasonable result from an already-known fact and relevant fuzzy rules. There are basically two kinds of inference operators: minimization (min) and product (prod). Defuzzification is responsible for transforming the fuzzy results from the fuzzy system into crisp values. There are many defuzzification methods such as weighted average (Wtaver) or weighted sum (Wtsum).

FIS can be said as strong tools to simulate nonlinear behaviors by employing the FL and linguistic fuzzy rules. FIS employing fuzzy "If–Then rules" can easily model the qualitative aspects of linguistic human knowledge and reasoning processes without precise quantitative analyses (Ho, Zhang, & Xu, 2001). Takagi, Sugeno and Kang have firstly investigated the fuzzy modeling or fuzzy identification, and found numerous practical applications in control, prediction and FIS (Sugeno & Kang, 1993; Takagi & Sugeno, 1985). In the Sugeno fuzzy inference systems, outcomes of fuzzy rules are characterized by function crisp outputs. The most common and well known FISs are Takagi–Sugeno (Takagi & Sugeno, 1985) and Mamdani Methods (Mamdani & Assilian, 1975).

4. Research method

It is seen that of the unit price has a very wide range distribution, while considering the qualitative and statistical properties of the unit price (UP, \$/m^2), which is the sale price per unit area of the residence as the dependent variable attempted to be

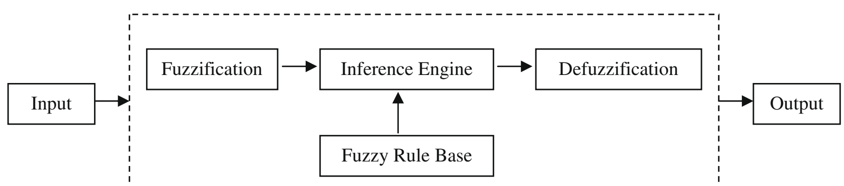

Fig. 1. A framework of FIS.

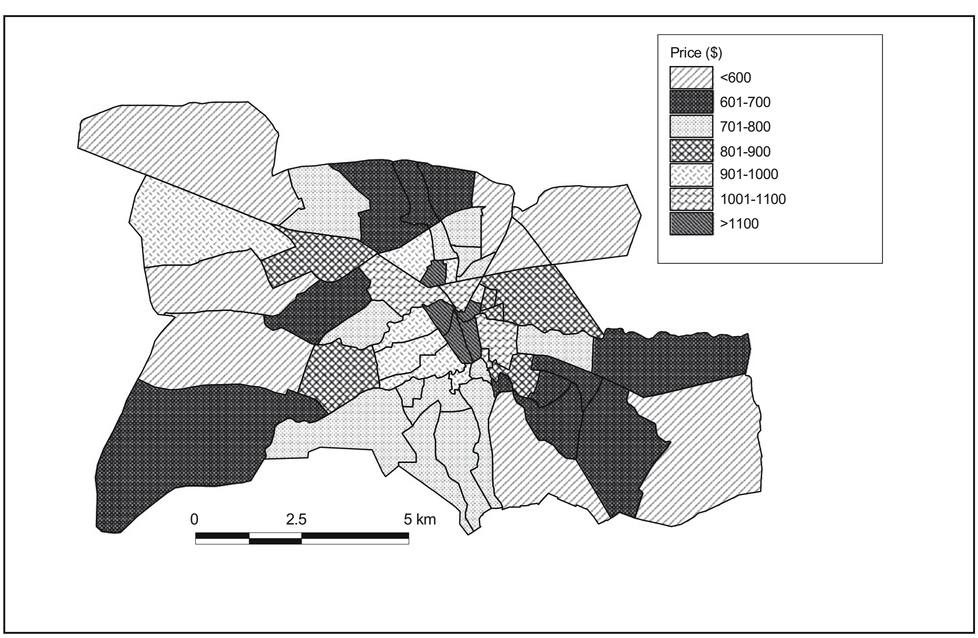

Fig. 2. Regional house UPs in Eskişehir city.

explained. At the beginning of the study, the list of factors affecting house prices and representing the inputs of the model has been formed and the factors have been grouped under main captions:

1. House factors
   1.1. Residential factors (RF): Type of residence, usability, number of rooms, number of balconies, number of open sides, view of residence.
   1.2. Factors related to building properties (BF): Elevator, hydrophore, generator, parking area, swimming pool, landscaping.
   1.3. Floor factor (FF): It presents the floor of the residence in the building.
2. Environmental factors
   2.1. Factor of regional conformity to city planning criteria (RCF).
   2.2. Factor of noise and air pollutions (PF).
3. Transportation factors
   3.1. Factor of nearness to socio-cultural centers (SCF): Library, educational, training and medical centers, etc.
   3.2. Factor of nearness to trade and shopping centers (TSF).
   3.3. Factor of public transportation systems (PTF).
4. Factors of the regional socio-economic (SEF)

Table 1
Input and output quantities used in models.

	Data used in training and testing the models	
	Minimum	Maximum
Input variables		
RF	5	10
BF	5	9
FF	3	8
RCF	4	10
PF	4	9
SCF	3	8
TSF	3	9
PTF	4	9
SEF	4	10
Output variable		
UP	520.00	1250.00

The questionnaire study is the main reference for obtaining the data used in the model application. Two hundred residences from 40 different urban regions of Eskişehir city in Turkey are chosen as samples after interviewing with real estate agencies. The evaluation of the chosen factors for sample residences by rating them between 1 and 10 points have been asked to the estate agencies for the questionnaire application. The residences mentioned in this study are apartments located in multistory buildings. Self-contained houses have been not considered in the questionnaire application. Furthermore, real prices of the houses have been collected from the house market. The unit of the price as New Turkish Lira has been converted to US Dollar in order to provide international comparison between these house prices. The distribution of house UPs in Eskişehir city with respect to regions is shown in Fig. 2.

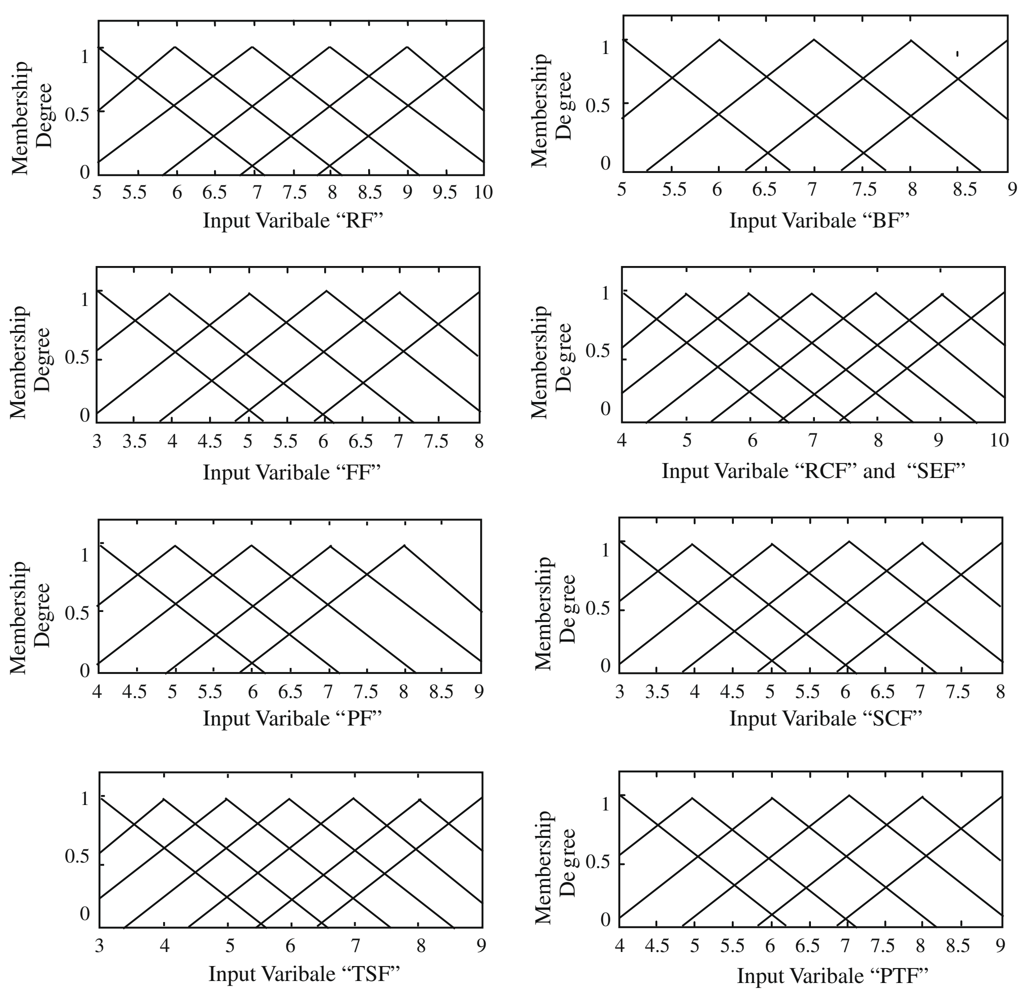

Fig. 3. Membership functions of input variables.

Fuzzy modeling is composed of two stages as structure identification and parameter prediction. Structure identification includes the processes such as selecting relevant input variables, choosing a specific type of fuzzy inference systems, determining the number of antecedent and consequent fuzzy rules, determining the type and number of membership functions. Parameter prediction is the determination of aimed values that responses to certain input values of constituted model.

While the prediction model for UPs has been constructed in this study, Sugeno type fuzzy inference system has been used. For training and testing the model, RF, BF, FF, RCF, PF, SCF, TSF, PTF, SEF have been used as input data, UP has been used as output data. Two hundred data have been obtained from questionnaire application. One hundred and sixty of them have been used for training and 40 of them have been used for testing. The limit values of the input and output variables used in Sugeno type fuzzy inference system are listed in Table 1. In the rule base, fuzzy variables were connected with "prod" (fuzzy and) operators and rules have been associated using "max-min" decomposition technique. Besides, 200 epochs have been continued and process has been terminated by the observation of the stability in error reduction. The membership functions of the training data set for the input variables of the model have been chosen the triangular types and premise parameter sub-spaces has been determined by using clustering of the training data set. The membership functions of the input variables used in the model are seen in Fig. 3, respectively. Thus, 78 rules have been obtained as following:

- R_i: If (RF is $RFmf_i$) and (BF is $BFmf_i$) and (FF is $FFmf_i$) and (RCF is $RCFmf_i$) and (PF is $PFmf_i$) and (SCF is $SCFmf_i$) and (TSF is $TSFmf_i$) and (PTF is $PTFmf_i$) and (SEF is $SEFmf_i$) Then (UP is $UPmf_i$) $i = 1, 2, \ldots, 78$.

5. Results and conclusions

In this section, results obtained from FL model have been presented in Table 2 and illustrated in Fig. 4. The errors occurred during the training and testing processes in FL model can be expressed as a root mean squared error (RMSE) and has been calculated by using Eq. (1):

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(T_i - P_i)^2} \tag{1}$$

Moreover, the absolute fraction of variance (R^2) and mean absolute percentage error (MAPE) have been calculated by using Eqs. (2) and (3), respectively,

Table 2
Comparison of real unit price with testing results obtained from UPs model ($\$/m^2$).

Cases	Real price	FL model price	Absolute error	Cases	Real price	FL model price	Absolute error
1	530.00	551.30	4.02	21	776.50	778.20	0.22
2	700.00	723.50	3.36	22	670.00	677.10	1.06
3	618.20	611.70	1.05	23	630.00	637.40	1.17
4	555.60	517.50	6.86	24	590.90	596.00	0.86
5	571.40	578.00	1.16	25	680.00	701.60	3.18
6	653.90	685.00	4.76	26	750.00	707.40	5.68
7	666.70	720.50	8.07	27	863.60	862.10	0.17
8	814.80	832.30	2.15	28	928.60	854.60	7.97
9	793.10	863.30	8.85	29	950.00	908.40	4.38
10	800.00	799.30	0.09	30	1040.00	1068.30	2.72
11	584.70	594.40	1.66	31	1050.00	987.30	5.97
12	689.70	701.10	1.65	32	900.00	849.60	5.60
13	720.30	680.90	5.47	33	1045.50	1041.70	0.36
14	730.80	695.20	4.87	34	952.40	922.40	3.15
15	650.00	644.00	0.92	35	1050.00	989.20	5.79
16	750.00	723.80	3.49	36	1050.00	1016.20	3.22
17	875.00	871.60	0.39	37	833.30	888.60	6.64
18	785.70	757.50	3.59	38	1074.10	1081.20	0.66
19	840.00	786.00	6.43	39	1047.60	1115.20	6.45
20	843.80	780.00	7.56	40	1111.10	1059.20	4.67

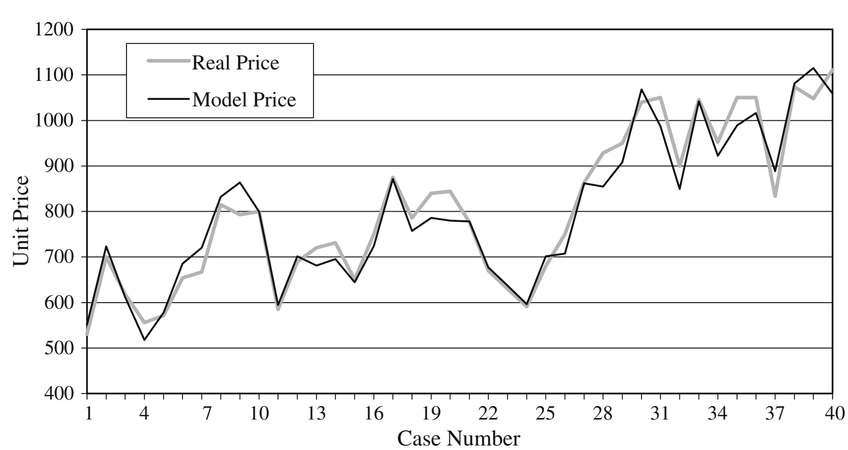

Fig. 4. Comparison of real unit price with testing results of UPs model.

$$R^2 = \frac{\left(n\sum_{i=1}^{n} T_i P_i - \sum_{i=1}^{n} T_i \sum_{i=1}^{n} P_i\right)^2}{\left(n\sum_{i=1}^{n} T_i^2 - \left(\sum_{i=1}^{n} T_i\right)^2\right)\left(n\sum_{i=1}^{n} P_i^2 - \left(\sum_{i=1}^{n} P_i\right)^2\right)} \quad (2)$$

$$\text{MAPE} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{T_i - P_i}{T_i}\right| \quad (3)$$

where P_i is the value predicted by UPs model (out of n sample cases) and T_i is the target value for sample case i. Statistical RMSE, R^2 and MAPE values have been obtained as 4.86×10^{-4}, 0.9999 and 0.007%, respectively, from training set in UPs FL model and RMSE, R^2 and MAPE values have also been found as 0.0375, 0.9305 and 3.65%, respectively, for testing set in model.

If suggested statistical values obtained from UPs model are discussed, it has been observed that the estimated UPs values can provide a close prediction to the real UPs values and the model can be available to use for UPs prediction of houses located in different regions of Eskişehir city in Turkey. In other words, the testing results in the model have also been shown that the predictions of UPs are very close to the unit real price values.

Furthermore, in the scope of this study, it is required to emphasize that suggested FL model, in which the determinations of UPs that have the main objectives and the values of criteria affecting the objectives are limited to certain range, by considering the properties of houses in different regions of Eskişehir city in Turkey. Because of these limited number of data and factors in certain narrow range, the model cannot be extended for general applications. Therefore, similar models can be produced with the new studies to be continued in this way by extending the criteria and the limit values about the house sale prices. As a final result, the FL can be applied to predict the house sale price for different cities in the world.

References

Alonso, W. (1964). *Location and land use*. Cambridge: Harvard University Press.

Bin, O. (2004). A prediction comparison of housing sales prices by parametric versus semi-parametric regressions. *Journal of Housing Economics, 13*, 68–84.

Brown, N. J., & Rosen, S. (1982). On the estimation of structural hedonic price model. *Econometrica, 50*(3), 765–768.

Dilmore, G. (1993). Fuzzy set theory: An introduction to its application for real estate analysts. In *Conference of the American real estate society in key west, Florida*.

Din, A., Hoesli, M., & Bender, A. (2001). Environmental variables and real estate prices. *Urban Studies, 38*(11), 1989–2000.

Fan, G., Ong, Z. S. E., & Koh, H. C. (2006). Determinants of house price. A decision tree approach. *Urban Studies, 43*(12), 2301–2315.

Ho, D. W. C., Zhang, P. A., & Xu, J. (2001). Fuzzy wavelet networks for function learning. *IEEE Transactions on Fuzzy Systems, 9*, 200–211.

Huang, S. J., & Chiu, N. H. (2009). Applying fuzzy neural network to estimate software development effort. *Applied Intelligence, 30*(2), 73–83.

Kain, J. F., & Quigley, J. M. (1970). Measuring and value of housing quality. *Journal of the American Statistical Association, 65*, 532–548.

Kim, K., & Park, J. (2005). Segmentation of the housing market and its determinants: Seoul and its neighboring new towns in Korea. *Australian Geographer, 36*(2), 221–232.

Lancaster, K. (1966). A new approach to consumer theory. *Journal of Political Economy, 74*, 132–157.

Liu, J., Zhang, G. X. L., & Wu, W. P. (2006). *Application of fuzzy neural network for real estate prediction. LNCS* (Vol. 3973, pp. 1187–1191). Berlin: Springer.

Lokshina, I. V., Hammerslag, M. D., & Insinga, R. C. (2003). Applications of artificial intelligence methods for real estate valuation and decision support. In *Hawaii international conference on business, Honolulu*.

Mamdani, E. H., & Assilian, S. (1975). An experiment in linguistic synthesis with a fuzzy logic controller. *International Journal of Man–Machine Studies, 7*, 1–13.

Muth, R. (1969). *Cities and housing*. Chicago: University of Chicago Press.

Pagourtzi, E., Assimakopoulos, V., Hatzichristos, T., & French, N. (2003). Real estate appraisal: A review of valuation methods. *Journal of Property Investment and Finance, 21*(4), 383–401.

Rabiega, W. A., Lin, T. W., & Robinson, L. M. (1984). The property value impacts of public housing projects in low and moderate density residential neighborhoods. *Land Economics, 60*(2), 174–179.

Rosen, S. (1974). Hedonic prices and implicit markets: Product differentiation in pure competition. *Journal of Political Economy, 82*, 34–55.

Selim, H. (2009). Determinants of house prices in Turkey: Hedonic regression. *Expert Systems with Applications, 36*(2), 2843–2845.

Şen, Z. (1998). Fuzzy algorithm for estimation of solar irradiation from sunshine duration. *Solar Energy, 63*(1), 39–49.

Şen, Z. (2004). *Fuzzy logic modeling principles in engineering*. Istanbul, Turkey: Water Foundation Press.

Stevenson, S. (2004). New empirical evidence on heteroscedasticity in hedonic housing models. *Journal of Housing Economics, 13*, 136–153.

Sugeno, M., & Kang, G. T. (1993). Structure identification of fuzzy model. *Fuzzy Sets and Systems Man and Cybernetics, 23*(3), 665–685.

Takagi, T., & Sugeno, M. (1985). Fuzzy identification of systems and its applications to modeling and control. *IEEE Transactions on Systems Man and Cybernetics, 15*, 116–132.

Wabe, J. S. (1971). A study of house prices as a means of establishing the value of journey time, the rate of time preference and the valuation of same aspects of environment in the London Metropolitan Region. *Applied Economics, 3*, 247–256.

Zadeh, L. A. (1965). Fuzzy sets. *Information and Control, 8*, 338–353.