Fifth International Conference on Construction in the 21st Century (CITC-V)
"Collaboration and Integration in Engineering, Management and Technology"
May 20-22, 2009, Istanbul, Turkey

Comparison of Fuzzy Logic, Artificial Neural Network and Multiple Regression Analysis Methods in the Determination of Selling Prices of Residences

Hakan Kuşan, Osman Aytekin, İlker Özdemir
Eskişehir Osmangazi University,Eskişehir, Turkey
hkusan@ogu.edu.tr, oaytekin@ogu.edu.tr, iozdemir@ogu.edu.tr

Abstract
Housing price is a determination related to the residence properties, this determination is not just the total of the residence unit itself and the constitutional properties of the residence, and on the other hand, it is also the combination of the region's properties and location. Therefore, the determination of selling price of residence depends on several factors. These factors can be classified such as house factors, environmental factors, transportation factors and regional socio-economic factors. Additionally, they are variable and have different specific properties. At the same time, the prediction of the real selling price of residence shows variety with respect to the estate salesmen and used prediction methods. In this study, a questionnaire study has been applied including some input variables such as the properties of residence, city plans, nearness to central trade, shopping, cultural, medical, social, and training centers, public transportation systems and other factors in order to determine the selling price of residence. Data obtained from questionnaire application have been used as input variables in the methods of fuzzy logic system, artificial neural network and multiple regression analysis. The selling prices of some residences located in the different regions of Eskişehir city have been predicted by using these three methods. Prediction results have been compared to the real selling prices of these residences and the error percentages have been determined. Finally, in the end of the study, it is suggested that such methods can be capable and usable for similar applications in the determination of selling price of residence according to the obtained results from three different methods.

Keywords
House price, Fuzzy logic, Artificial neural network, Multiple regression analysis

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 and Park, 2005).

In this study, it is aimed to predict the house unit prices (UPs) by using different models. 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 models have been obtained from a prepared questionnaire application. 160 of these data have been used for training set and 40 of them have been used for testing set in the constructed models. UPs obtained from the models and the real UP values have been compared to each other to determine the precision of the models suggested in this study.

2. Related Previous Research

Many researchers 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 real estate 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).

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 and Rosen, 1982). 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 and Park 2005).

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 *et al.*, 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, 2008). 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 *et al.*, 2006).

3. Adapted Network Based Fuzzy Inference System

The concept of "fuzzy set", instead of Aristotelian logic which has only two possibilities, was preliminarily asserted by Zadeh (Zadeh, 1965). The concept of "fuzzy set" has been developed rapidly by both Zadeh himself and indefinite researchers. Real applications of this concept are actualized successfully at the same time. Fuzzy systems, generally, are the systems using fuzzy sets to provide the transformation between the input and output variables (Zadeh, 1965). The variables, which do not have numerical representatives, can be directly transferred to the model with fuzzy logic by using the opinions and experiences of specialists.

Fuzzy logic provides the ease of constructing models by using the data sets of input and output with the help of adapted techniques such as Adapted-Network-Based Fuzzy Inference System (ANFIS). The ANFIS system of Jang is one of the oldest and the most common used systems in the neural-fuzzy systems. The architecture of ANFIS demonstrates a network architecture having the capability of neural learning of the Sugeno type fuzzy systems. This network is composed of the nodes being set as layers for the purpose of performing a specific function for each layer (Jang, 1993).

4. Artificial Neural Network

Artificial Neural Network (ANN) is a type of calculation way inspired from the biological neural networks in which the solution is being learned by using a training set and the property of generalizing solution related to the learned system is present (McCulloch and Pitts, 1943). ANNs are the mathematical systems consisting of many operational members that are connected to each other by means of being weighted. An operational member receives the signals from other neurons; combines these signals, transforms them and exhibits a numerical result. Operational members in the system generally represent the real neurons and connect to each other in a net work; the obtained structure constructs the ANNs.

Typically, the architecture of ANN is constructed and the values of weights are determined in order to maximize the accuracy of produced outputs by choosing one of the different mathematical algorithms.

ANNs reveal the relationship between the input variables and the predicted variables by using previous examples to determine the weights. Since the relationships between the variables were revealed, ANN can have been run by using different data and the predictions can be produced. The performance of a network is measured by way of aimed signal and error rate. The error rate is determined by comparing the output of network with the aimed output. As a result of capability of solving complicated problems, ANN with these properties mentioned above, has recently being used in many application areas.

5. Multiple Regression Method

Multiple Regression Method (MR) is generally used for investigating the relationship between the variables and summarizing the data. Regression is basically a combination of the forward and back propagated methods and probably is the most common method (Draper and Smith, 1981). In the Multiple Regression the cause-effect relationship between multiple independent variables ($x_1, x_2, \ldots\ldots, x_n$) and one dependent variable (y) can be expressed with the Equation1 as a mathematical model.

$$y = a + b_1x_1 + b_2x_2 + \ldots\ldots + b_nx_n \quad (1)$$

The differences between the real y and *theoretical* y values are minimized by utilizing *the least squares method* for calculating the coefficients in this equation.

6. 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 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)

The questionnaire study is the main reference for obtaining the data used in the models. 200 residences from 40 different urban regions of Eskişehir city in Turkey have been chosen as samples after interviewing with real estate agencies. The evaluation of the chosen factors for sample residences by rating them between 1-10 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.

For training and testing the models, RF, BF, FF, RCF, PF, SCF, TSF, PTF, SEF have been used as input data, UP has been used as output data. 200 data have been obtained from questionnaire application, 160 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 models are listed in Table 1.

Table 1: The 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

In this study, while the prediction model for UPs is constructed using ANFIS, Sugeno type fuzzy inference system has been used. In the rule base, fuzzy variables have been connected with "prod" (fuzzy and) operators and rules have been associated using "max-min" decomposition technique. Furthermore, 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 are the triangular types and premise parameter sub-spaces are determined by using clustering of the training data set. The output membership function has been obtained with a constant expressed as f=r. In this way, 78 rules are obtained as following.

R_i: If (AG is $AGmf_i$) and (C is Cmf_i) and (W is Wmf_i) and (S is Smf_i) and (A is Amf_i) and (RA is $RAmf_i$) and (SP is $SPmf_i$) and (SF is $SFmf_i$) Then (f_c is $f_c mf_i$) i=1,2…78.

The ANN model, with 9 neurons in the input layer and 1 neuron in the output layer, has been developed and demonstrated in Figure 1. It has been attempted to determine the required number of neurons for hidden layer by using different numbers of neurons, after then, 10 neurons has been used in the hidden layer of multilayer neural network due to its minimum absolute percentage error values of training and testing sets. The values have been used in network in the range of [0.1-0.9] by normalizing them. The model has been trained with feed-forward and back-propagation training algorithm. The non-linear sigmoid function has been used in the hidden layer and the cell outputs at the output layer. Momentum rate (η) and learning rate (α) values have been taken into account as 0.9 and 0.6, respectively. After then, the model has been trained through 10000 iterations.

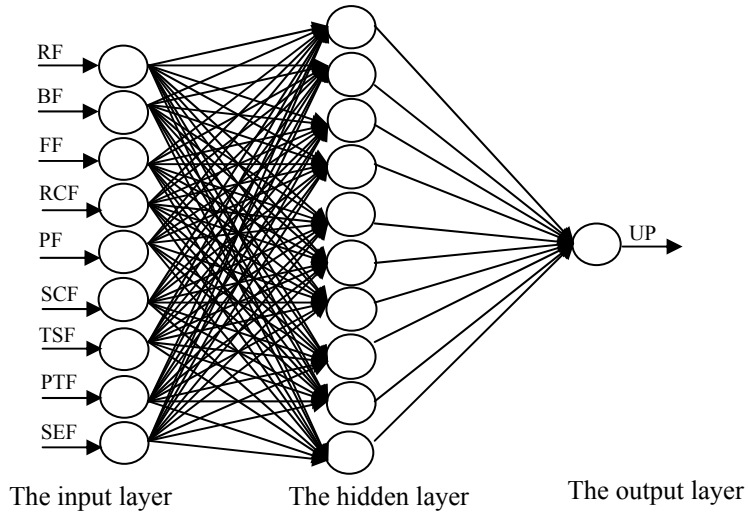

Figure 1: The Architecture of the Neural Network Model

Unit price values have been calculated by using multiple regression analysis as another approach method related to the input values defined above. The results of model analysis have been calculated by Microsoft Excel software is shown in Table 2.

Table 2: Results of Regression Analysis

	Coefficients	Std Error	t Stat
Constant	-585.389	52.210	-11.212
RF	30.869	4.304	7.1715
BF	8.964	4.564	1.963
FF	11.999	2.029	5.912
RCF	11.628	3.130	3.715
PF	21.769	4.888	4.453
SCF	20.209	5.402	3.740
TSF	-0.655	5.944	-0.110
PTF	9.078	5.850	1.551
SEF	94.294	4.310	21.873

7. Results and Conclusions

In this study, the use of regression analysis as one of the common prediction methods and fuzzy logic-artificial neural network as new flexible modeling approach methods, in predicting house selling prices, have been compared to each other by using the reference data obtained from a questionnaire application. UPs values obtained from models are given in Table 3.

Table 3: Comparison of Real Unit Price with Testing Results Obtained from UPs Models ($/m^2)

Cases	Real	ANFIS	ANN	MR	Cases	Real	ANFIS	ANN	MR
1	530.00	551.30	514.35	507.78	21	776.50	778.20	672.86	756.08
2	700.00	723.50	645.57	670.71	22	670.00	677.10	610.89	661.36
3	618.20	611.70	586.63	580.05	23	630.00	637.40	588.75	641.35
4	555.60	517.50	558.48	559.00	24	590.90	596.00	605.12	549.13
5	571.40	578.00	492.92	557.62	25	680.00	701.60	653.32	687.11
6	653.90	685.00	700.61	743.13	26	750.00	707.40	704.18	742.49
7	666.70	720.50	693.14	721.06	27	863.60	862.10	824.37	862.13
8	814.80	832.30	825.75	884.52	28	928.60	854.60	919.12	950.59
9	793.10	863.30	824.20	856.81	29	950.00	908.40	922.96	942.89
10	800.00	799.30	774.98	820.26	30	1040.00	1068.30	1081.82	1066.46
11	584.70	594.40	628.97	614.91	31	1050.00	987.30	967.18	965.96
12	689.70	701.10	658.28	664.67	32	900.00	849.60	876.88	907.61
13	720.30	680.90	826.11	769.18	33	1045.50	1041.70	1029.82	1033.97
14	730.80	695.20	750.31	778.03	34	952.40	922.40	844.30	878.74
15	650.00	644.00	652.58	697.71	35	1050.00	989.20	1019.93	998.56
16	750.00	723.80	684.12	698.20	36	1050.00	1016.20	975.08	996.81
17	875.00	871.60	832.25	863.41	37	833.30	888.60	901.68	924.96
18	785.70	757.50	733.15	758.24	38	1074.10	1081.20	1105.12	1061.83
19	840.00	786.00	760.91	784.30	39	1047.60	1115.20	1116.73	1118.14
20	843.80	780.00	775.62	762.15	40	1111.10	1059.20	1067.74	1067.32

The errors occurred during the training and testing processes in Ups models can be expressed as a root-mean-squared error (RMSE) and is calculated by using Equation 2.

$$RMSE = \sqrt{\frac{1}{n}\sum_{j=1}^{n}(P_{(ij)} - T_j)^2} \tag{2}$$

Additionally, the absolute fraction of variance (R^2) and mean absolute percentage error (MAPE) are calculated by using Equations 3 and 4 respectively.

$$R^2 = \frac{(n\sum_{i=1}^{n} T_i P_i - \sum_{i=1}^{n} T_i \sum_{i=1}^{n} P_i)^2}{(n\sum_{i=1}^{n} T_i^2 - (\sum_{i=1}^{n} T_i)^2)(n\sum_{i=1}^{n} P_i^2 - (\sum_{i=1}^{n} P_i)^2)} \tag{3}$$

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{T_i - P_i}{T_i}\right| \tag{4}$$

Where $P_{(ij)}$ is the value predicted by UPs model, i for sample case j (out of n sample cases) and T_j is the target value for sample case j.

In Table 4, RMSE, R^2 and MAPE values regarding ANFIS, ANN and MR are presented. When the results are examined, it is seen that the performance of ANFIS is better than the performances of other models, R^2 values of ANFIS are higher and RMSE values are lower compared to the results of the other models. It is also observed that the performance of ANN model leads to similar results to that one of MR model. Besides, it is noticed that ANFIS and ANN methods can be used as alternative methods instead of constructing model equations by regression analysis, more precise and fit for purpose solutions can be produced using fuzzy logic and artificial neural network modeling because of adapting capability of the parameters resiliently by users.

Table 4: Comparison of the ANFIS, ANN and MR Models' Performance

	ANFIS			ANN			MR		
	RMSE	R^2	*MAPE (%)*	*RMSE*	R^2	*MAPE (%)*	*RMSE*	R^2	*MAPE (%)*
Train	0.0048	0.9839	0.7138	0.0435	0.9654	4.4091	0.0432	0.9543	4.3435
Test	0.0375	0.9509	3.6580	0.0525	0.9158	5.6453	0.0453	0.9251	4.6761

In the scope of this study, it is required to emphasize that fuzzy logic, artificial neural network and multiple regression models are discussed, in which the determinations of UPs are 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, ANN and ANFIS can not 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. Consequently, it can be pointed out that the ANN, ANFIS and MR models can be applied to predict the house sale price for different cities in the world.

8. References

Alonso, W. (1964). *Location and Land Use,* Harvard University Press, Cambridge.

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

Brown, N.J., and Rosen, S. (1982). "On the estimation of structural hedonic price model". *Econometrica*, Vol.50, No.3, pp. 765-768.

Din, A., Hoesli, M., and Bender, A. (2001). "Environmental variables and real estate prices", *Urban Studies*, Vol.38, No.11, pp. 1989–2000.

Draper, N.R., and Smith, H. (1981). *Applied Regression Analysis*, Wiley. New York.

Jang J.S.R. (1993) "ANFIS: Adaptive-network-based fuzzy inference system". *IEEE Trans. on Syst. Man Cybern*, Vol.23, No.3, pp. 665–685.

Kim, K., and Park, J. (2005). "Segmentation of the housing market and its determinants: Seoul and its neighboring new towns in Korea", *Australian Geographer*, Vol.36, No.2, pp. 221–232.

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

Liu, J., Zhang, G.X.L., and Wu, W.P. (2006). "Application of fuzzy neural network for real estate prediction", *Third International Symposium on Neural Networks*, Editors: J. Wang, Chengdu, pp. 1187-1191.

McCulloch, W. S. and Pitts, W. (1943). "A logical calculus of the ideas immanent in nervous activity", *Bull. Math. Biophysics*, Vol. 5, pp. 115-133.

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

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

Selim, H. (2009). "Determinants of house prices in Turkey: Hedonic regression versus artificial neural network", *Expert Systems with Applications*, Vol. 36, No. 2, pp. 2843-2852.

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

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*, Vol.3, pp. 247-256.

Zadeh, L.A. (1965). "Fuzzy sets", *Information and Control*, Vol.8, pp. 338-353.