APPLICATION OF FUZZY REGRESSION MODEL FOR REAL ESTATE PRICE PREDICTION

Abdul Ghani Sarip[1], Muhammad Burhan Hafez[2] and Md. Nasir Daud[3]

[1, 3] Faculty of Built Environment, University of Malaya, Kuala Lumpur

[2] Faculty of Computer Science & Information Technology, University of Malaya,
Kuala Lumpur

Email: garnae@um.edu.my[1], burhan.hafez@gmail.com[2], mdnasir@um.edu.my[3]

ABSTRACT

Many studies have been conducted for modeling the underlying non-linear relationship between pricing attributes and price of property to forecast the housing sales prices. In recent years, more advanced non-linear modeling techniques such as Artificial Neural Networks (ANN) and Fuzzy Inference Systems (FIS) have emerged as effective techniques to predict the house prices. In this paper, we propose a fuzzy least-squares regression-based (FLSR) model to predict the prices of real estates. A comprehensive comparison studies in terms of prediction accuracy and computational complexity of ANN, Adaptive Neuro Fuzzy Inference System (ANFIS) and FLSR has been carried out. ANN has been widely used to forecast the price of real estates for many years while ANFIS has been introduced recently. On the other hand, FLSR is comparatively new. To the best of our knowledge, no property prices prediction using FLSR was developed until recently. Besides, a detailed comparative evaluation on the performance of FLSR with other modeling approaches on property price prediction could not be found in the existing literature. Simulation results show that FLSR provides a superior prediction function as compared to ANN and FIS in capturing the functional relationship between dependent and independent real estate variables and has the lowest computational complexity.

Keywords: property price prediction, fuzzy logic, fuzzy regression, neural networks, non-linear modelling

1.0 INTRODUCTION

A real estate entity is an embodiment of the physical land and all its improvements together with all the rights, interests, benefits and liabilities arising from ownership of the entity. The valuation of real estate is thus an exercise in providing a quantitative measure of the benefits and liabilities accruing from the ownership [1]. The conduct of professional real estate valuation is the domain of the appraisers or assessors. In arriving at a value estimate, the professionals need to relate to important economic principles that underlie the operation of the real estate market. These include the principles of supply and demand, competition, substitution, anticipation and change. Common to all these principles is their direct and indirect effect on the degree of utility and productivity of a property. Consequently, it may be stated that the utility of real estate reflects the combined influences of all market forces that come to bear upon the value of a real estate parcel.

In practice, the sales comparison method or market-value approach has been the traditional and by far the most common method adopted for real estate valuation, particularly for residential real estate [2]. Using this method, the value of a subject real estate is estimated on the basis of sales of other comparable properties. This is to establish an estimate of the market value of the subject real estate, which is deemed to be the estimated amount for which a real estate should exchange on the date of valuation between a willing buyer and a willing seller in an arm's length transaction after proper marketing wherein the parties have each acted knowledgeably, prudently and without compulsion [3].

The sales comparison method rests on the notion that there exists a direct relationship between the market value of a subject real estate and the selling prices of the comparable properties whereby the latter represent competitive investment alternatives in the market. Value influences such as the physical characteristics and location qualities are considered in analysing the comparability, in a process that embeds the consideration of supply and demand in leading to the final opinion on the value estimate. In practice, valuers utilise historical

15

transaction data as benchmark. Then, they adjust the value to account for the dissimilarities that exist between subject real estate and the comparable properties.

More recently, there has been a growing interest to discover new value estimation approaches. The prime motivation has been an increased expectation from the market for speedier valuations in order to drive speedier decisions, but with no compromise on the quality. The new efficiency is expected to emanate from improving the means by which value estimates are generated. This had led to experimentation with new techniques, some of which have been found promising for real estate valuation. The new approaches include multiple regressions, intelligence-based neural networks and more complex applications such as expert systems and fuzzy logic.

2.0 THE 'NEWER' TOOLS FOR REAL ESTATE VALUATION

Multiple Regression Analysis (MRA) is a statistical method that estimates the value of a dependent variable (e.g. house price) based on given values of the independent variables (e.g., the physical attributes of the house). This approach has been widely used in prediction applications despite its limitations when it comes to dealing with the relationship between independent and the dependent variables in the presence of outliers, non-linearity and non-normality [4], [5].

Recent studies have been conducted to investigate more complex techniques for modelling the underlying non-linear relationships between the pricing factors and the price of the property, such as Artificial Neural Networks (ANN) and Fuzzy Logic (FL). For example, authors in [6] explored the correlations between pricing factors and quantified the contribution of each factor to value in real estate by using 2-hidden-layer Back-Propagation neural network. Based on this information, the impact of pricing factors and the relationship between property price and each of the factors are revealed.

A neural network approach tends to over-fit. However, the over-fitting problem can be greatly overcome through the use of monotonic neural network. This innovation is particularly useful where the relation between each of the inputs and the network output is nearly monotonic. When the number of hidden neurons increases, the monotonic neural networks are less sensitive to over-fitting than the ordinary neural networks [7]. In general, ANN outperforms MRA when the size of the training data is sufficiently large. ANN also performs better as compared to MRA when the functional form of the prediction model cannot be specified in advance [8] or when the relationships between variables are more likely to be non-linear [9].

Authors in [10] undertook a comparison of four techniques (i.e., MRA, ANN, Memory-based Reasoning and Fuzzy Logic) for estimating house value. They found that none of these methods were able to outperform the estimation ability of MRA although each method was good enough to serve as the latter's complement. Nevertheless, their study highlighted the potential strength of those other methods for improving the prediction of selling prices where quantifiable data is less.

The application of FL to predict house price has shown encouraging results in modeling the complex and non-linear behavior between input and output variables. Fuzzy Inference System (FIS) was adopted recently in several studies as a prediction model. This technique has been found useful especially when the data sample includes linguistic variables (e.g., those related to environmental attributes) or when the data is mostly drawn from non-numerical sources such as questionnaires [11]. FIS consisting of a set of IF-THEN rules that fire simultaneously when it is provided with an observation. The firing strength of each rule is relative to the degree of matching the antecedent part. There are challenges to the creation of an FIS model; these include the determination of fuzzy sets and rules. The determination requires domain knowledge from human experts. Hence, adjusting the fuzzy rules and sets can be time-consuming. Therefore, the fuzzy system is integrated with the learning capabilities of ANN, resulting in Adaptive Neuro-Fuzzy Inference System (ANFIS) to overcome the problem.

A number of studies have been carried out to investigate the application of ANFIS to predict real estate prices [12, 13]. Guan *et al.* offered an in-depth discussion on the design and implementation of ANFIS to predict the

16

Malaysian Journal of Computer Science. Vol. 29(1), 2016

sales of houses in the Midwest region of the United States. Their study proved the feasibility and effectiveness of ANFIS for that locality, arguing that more research is needed to further explore the ANFIS approach using a much larger dataset of higher dimensionality.

Meanwhile, fuzzy regression, which is a fuzzy logic-based regression model, has been used effectively to address many prediction problems such as solar networks, forensics and linguistics [14], [15], [16]. However, to the best of our knowledge, no research is known to extend the application of this technique for real estate valuation. This has thus become the underlying motivation for our undertaking of the current work.

3.0 METHODOLOGY

The aim of this work is to identify the value prediction performance of fuzzy logic approach in the context of real estate valuation. We propose two predictive fuzzy models for estimating property value, namely the Adaptive Neuro-Fuzzy Inference System (ANFIS) and Fuzzy Least-Squares Regression (FLSR).

3.1 ANFIS Model

The ANFIS model is built based on Sugeno-type fuzzy inference system [17], in which every rule adopts the following form:

$$IF\ x_1\ is\ A_{1j}\ AND\ x_2\ is\ A_{2j}\ AND\ ...\ AND\ x_n\ is\ A_{nj}$$
$$THEN\ y = c_0 + c_1 x_1 + c_2 x_2 + ... + c_n x_n \ , \tag{1}$$

where A_{ij} is the fuzzy set corresponding to the jth linguistic term of the ith variable, n is the number of input variables (i.e., eight in our study), y is the predicted value of the property and c_i denotes the consequent parameters whose values are derived from the training process.

The ANFIS model uses the training data to estimate two sets of tunable parameters, namely the antecedent parameters and consequent parameters, respectively. Antecedent parameters are the parameters determining the shape of antecedent membership function, $\mu_{A_{ij}}(x)$ and the consequent parameters are the coefficients of the linear functions of the consequent part. Fig. 1 shows the basic structure of the ANFIS model.
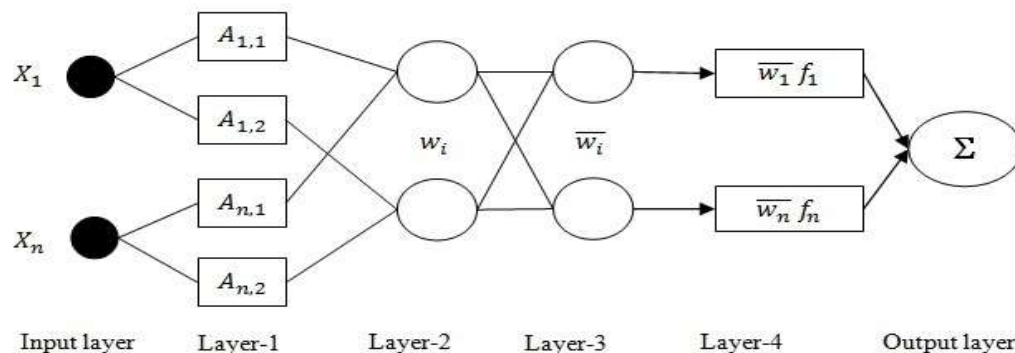

Fig. 1 Basic structure of the ANFIS model

The input layer contains nodes of the input variables representing the eight property features (see Table 1). Layer-1 contains the membership functions of the fuzzy sets associated with each input variable. Layer-2

17

includes the firing strengths (i.e., the weights) of all the rules. The firing strength of each rule is computed by the product of the membership degrees in its antecedent part available from the previous layer. Therefore, the number of nodes in this layer is identical to the number of rules. Layer-3 performs normalization over the weights resulting from the previous layer (i.e., $\overline{w_i} = \frac{w_i}{\sum w_i}$). Layer-4 consists of nodes representing linear functions, the coefficients of which are adjusted by the least-squares algorithm.

The output layer produces one value - the weighted sum of the linear functions of the previous layers, which is calculated as follows: $y = \sum_i \overline{w_i} f_i$. The training process includes two passes, namely the forward pass and the backward pass, respectively. During the forward pass, the antecedent parameters are fixed and the consequent parameters are estimated using least-squares algorithm since they are linearly related to the output. Meanwhile, during the backward pass, the consequent parameters are fixed and the antecedent parameters are optimally estimated by gradient descent algorithm.

To control the potential proliferation of rules resulting from the large number of inputs, subtractive clustering technique was employed to optimize the selection of rules and the initial parameter values based on the given training dataset.

3.2 FLSR Model

A conventional linear regression has the form shown in the following equation,

$$y = \beta_0 + \beta_1 x_1 + \ldots + \beta_n x_n + \varepsilon , \tag{2}$$

where y is the observed response variable, n is the number of the input variables (the predictors), $x_1 \ldots x_n$ are the input variables (the predictors), $\beta_0,.., \beta_n$ are the unknown model parameters and ε is the error term. Thus, we have

$$y^i = \beta_0 + \beta_1 x_1^i + \ldots + \beta_n x_n^i + \varepsilon i = 1, \ldots, m$$

for m observations.

In conventional linear regression, least squares method is used to find the β_i parameters that minimize the sum of the squared errors:

$$S(\beta_1, .., \beta_n) = \sum_{i=1}^{m} [y^i - (\beta_0 + \beta_1 x_1^1 + \ldots + \beta_n x_n^m)]^2 . \tag{3}$$

In general, the difference between the estimated and the observed values arises either from uncertainty of model structure or from imprecise observations. Embedded fuzziness (which is quite common to sales price predictions) leads to the development of Fuzzy Linear Regression (FLR) model.

Fuzzy Least-Squares Regression (FLSR) is an extension of the least-squares method to solve for fuzzy linear regression. An FLSR model for property price prediction can be formalized as:

$$\tilde{y} = \widetilde{\beta_0} + \widetilde{\beta_1} x_1 + \ldots + \widetilde{\beta_n} x_n + \tilde{\varepsilon} , \tag{3}$$

where $\tilde{y}$ is the fuzzy response, n is the number of variables, $\widetilde{\beta_0} \ldots \widetilde{\beta_n}$ are the fuzzy parameters of the model, $x_1 \ldots x_n$ are the input variables and $\tilde{\varepsilon}$ is the fuzzy error term.

We aim to find the values of the fuzzy parameters $\widetilde{\beta_i}$ that minimize the sum of the squared distances between the observed and the estimated responses $\sum_{i=1}^{m} d^2 (y_i, \tilde{y}_i)$, where m is the number of training examples and $\tilde{y}_i$ is the fuzzy response.

We use triangular fuzzy numbers (TFNs) to represent the fuzzy parameters $\widetilde{\beta_i}$ as follows

$$\widetilde{\beta_i} = (\beta_{il}, \beta_{im}, \beta_{ir}) , \tag{4}$$

18

Malaysian Journal of Computer Science. Vol. 29(1), 2016

where β_{im} is the center, β_{il} and β_{ir} are respectively the left and right points of the triangular. Fig. 2shows a representation of triangular fuzzy number.
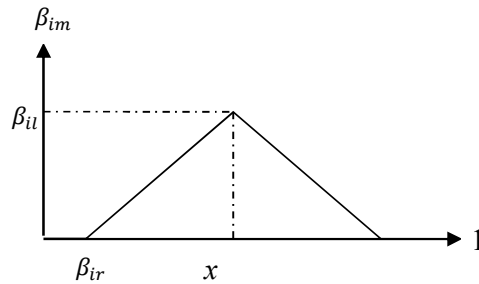

Fig. 2 A Triangular Fuzzy Number

To compute the distance between any two TFNs (i.e., $\widetilde{A_1} = (a_{1l}, a_{1m}, a_{1r})$ and $\widetilde{A_2} = (a_{2l}, a_{2m}, a_{2r})$), a distance metric known as *Diamond* distance (i.e., defined in [18]) is adopted as follow:

$$d^2(\widetilde{A_1}, \widetilde{A_2}) = (a_{1l} - a_{2l})^2 + (a_{1m} - a_{2m})^2 + (a_{1r} - a_{2r})^2 . \qquad (5)$$

Based on *Diamond* distance, we search for the parameters $\widetilde{\beta_i}$ that minimize the following sum of squared distances:

$$S(\widetilde{\beta_0}, \dots, \widetilde{\beta_n}) = \sum_{i=1}^{m} d^2(\widetilde{y_i}, \widetilde{\beta_0} + \widetilde{\beta_1} x_{i1} + \dots + \widetilde{\beta_n} x_{in})$$

$$= \sum_{i=1}^{m} [(y_{il} - \beta_{0l} - \beta_{1l} x_{i1} - \dots - \beta_{nl} x_{in})^2 + (y_{im} - \beta_{0m} - \beta_{1m} x_{i1} - \dots - \beta_{nm} x_{in})^2$$
$$+ (y_{ir} - \beta_{0r} - \beta_{1r} x_{i1} - \dots - \beta_{nr} x_{in})^2] , \qquad (6)$$

where $\widetilde{\beta_0} + \widetilde{\beta_1} x_i$ is the triangular fuzzy number $(\widetilde{\beta_{0l}} + \widetilde{\beta_{1l}} x_i, \widetilde{\beta_{0m}} + \widetilde{\beta_{1m}} x_i, \widetilde{\beta_{0r}} + \widetilde{\beta_{1r}} x_i)$.

Taking the derivatives of *S* with respect to β_{il}, β_{im}, β_{ir} (*i= 0…n*) and setting them to zero, we obtain the optimal estimates for $\widetilde{\beta_0}$ to $\widetilde{\beta_n}$ which are actually the solutions of a linear system of *3*(n+1)* equations with *3*(n+1)* unknowns.

Algorithm 1 FLSR-based real estate price prediction
1:     Input dataset D
2:     Partition D into training D_1 and testing D_2 subsets
3:     Normalize D_1 values to fall in the closed range [-1,1]
4:     Let $X \in \mathbb{R}^{m \times n}$ be the input variables matrix, $\vec{y} \in \mathbb{R}^{m \times 1}$ be the output (response) vector and B $\in$ $\mathbb{R}^{n \times 1}$ be the model parameters vector (see Equation (3))
5:     $X \leftarrow [1\ D1^{(1)} \dots D1^{(n-1)}]$: $D1^{(i)}$ is the ith column in D_1
6:     $\vec{y} \leftarrow D1^{(n)}$
7:     Define the model fuzzy parameters, $\widetilde{\beta_i}$, as triangular fuzzy numbers (see Equation (4)) // Find $B^{\wedge}$ that minimizes the sum of squared distances S (B) = $(XB - \vec{y})^T(XB - \vec{y})$, i.e., $B^{\wedge} =$ $\underset{B}{\arg\min} S(B)$
8:     Use Diamond measure to compute the distance between two TFNs (see Equation (5) and (7)

19

9:   Set the gradient of S with respect to each parameter to 0
10: Solve a linear system of equations to find the optimal parameters $B^{\wedge}$
     // Defuzzification and denormalization
11: Defuzzify each optimal triangular fuzzy number β by choosing a representative crisp value from it
     and let it be its centre β_m
12: Denormalize the values of D_1
     // Predict real estate prices using the leaned model
13: $Z \leftarrow [1\ D2^{(1)} \dots D2^{(n-1)}] : D2^{(i)}$ is the ith column in D_2
14: $\vec{p} = ZB^{\wedge}$
     // Compute Mean Absolute Error (MAE)
15: MAE = Sum $(|\vec{p}| - D2^{(n)})$

After the FLSR model is built by computing the optimal values of its parameters from the training dataset, we can apply the model to any new observation. Suppose we have the following eight-feature property which we want to predict the estimated price p:

$$\vec{x} = [x_1\, x_2\, x_3\, x_4\, x_5\, x_6\, x_7\, x_8].$$

Now we can apply our FLSR model to it as follows:

$$P = \sum_{i=0}^{8} xi\ Bi^{\wedge},$$

where $Bi^{\wedge}$ ($i=0,\dots,8$) are the optimal model parameters retrieved as explained in Equation (6) earlier, and $x_0 = 1$ (the intercept term).

4.0 DATA

The dataset of this study is extracted from a sales records of 352 properties in the District of Petaling, Kuala Lumpur. Table 1 provides description of the property in terms of eight features (or variables) and pricing in MYR. Each record in the table relates to a single property.

Table 1. Property Features

Feature/Variable	Description of the feature	Example value
LA	Land Area	170
BA	Built-up Area	87.4
BE	# Bedrooms	3
BT	# Bathrooms	3
AG	Building Age	14
RC	Repair Condition	Good
QF	Quality of Furniture	Very Good
LO	Location	Good

Data for RC, QF and LO are categorical. However, they have been converted to numerical values for the purpose of statistical computation.

A certain amount of pre-processing was necessary as part of the data preparation to detect errors in the data (e.g., missing values, duplicates and outliers). In the initial dataset, three records were identified as identical and thus two of them were removed from the dataset. Besides, two records were identified as outliers and were

20

Malaysian Journal of Computer Science. Vol. 29(1), 2016

singled out for deletion since the distance of their sale prices to the mean sale price are three times greater than the standard deviation, as shown in Table 2.

Table 2. Outlier test

	Mean sale price	Standard deviation	Outliers detected	Outlier's value	Action
1st scan for outliers	587,280	361,700	1	2,913,006	Delete
2nd scan for outliers	580,590	339,960	1	1,630,000	Delete
3rd scan for outliers	577,570	335,730	Null	Null	Null

As a result of the data screening, we obtained a dataset of 348 properties that was available for simulation models.

5.0 PERFORMANCE EVALUATION

To assess the effectiveness and performance of using fuzzy approaches to predict real estate values, we perform a series of simulation-based experiments. We also evaluate the performances of the fuzzy models, namely ANFIS and FLSR by comparing their prediction results to those of Artificial Neural Network (ANN) model. In the simulations, the pre-processed dataset used (i.e., as described in Section 4), is divided into two sets; training set and test set, with 90% of the data being assigned to the training set and 10% to the test set. The training set is used in building the models, while the test set is used to test the performance of the models.

Our evaluation criterion is based on the Mean Absolute Error (MAE) of the test set, which is defined as:

$$MAE = \frac{1}{n} \sum_{i=1}^{n} \left| y_i - \hat{y_i} \right|,$$

where n is the size of the test set, y_i is the actual price of the property and $\hat{y_i}$ is the estimated price.

We use MATLAB for data preprocessing and models implementation. Experiments are performed on a Core I3 CPU 2.13 GHz platform with 2 GB RAM. The values of the eight input variables have been normalized in [-1, 1] in order to prevent the variables from overweighting each other.

5.1 ANN Model Simulation Results

MATLAB Neural Network Toolbox is used for building the ANN model. The model parameter values are set as follows; a 2-hidden layer feed forward neural network trained using back-propagation algorithm is used to estimate the property price. The network consisting of twelve neurons in the first hidden layer and seven in the second layer. The training parameters are set up to the following values: The learning rate is 0.01, the training error goal is 0.01 (the training terminates on reaching this error), transfer functions are tangent-sigmoid for the hidden layers and Purelin for the output layer and the maximum number of epochs allowed is 1,000. These numbers are determined experimentally by performing multiple simulations to produce the best results.

After training the network, the predicted property prices are computed and plotted for the samples in the test set along with their actual values, as shown in Fig. 3.

21

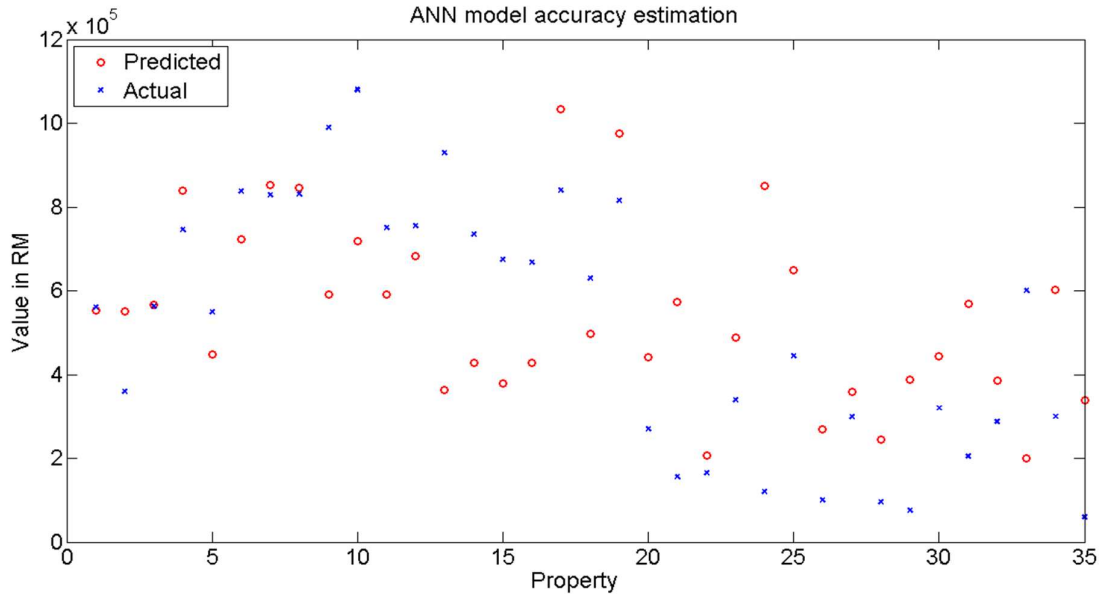Malaysian Journal of Computer Science. Vol. 29(1), 2016

Fig. 3 Result of applying ANN model to the test set

In the figure, it is observed that ANN model captures the non-linear relationship between the input variables and the output. However, it gives wrong estimated values in almost half of the test samples. The MAE for applying this ANN model on the test set is 211,580.

5.2 ANFIS Model Simulation Results

MATLAB Fuzzy Logic Toolbox is used for building the ANFIS model. Due to the relatively large number of inputs in this study, it is not feasible to generate the fuzzy rules using all possible permutations of membership functions of input variables. Owing to this reason, we use subtractive clustering, which derives the set of rules that capture the behavior of the data.

The Gaussian curve is chose as the type for all membership functions because its shape can vary easily by changing the values of mean and standard deviation values. A cluster radius of 0.88 is experimentally chosen for all the input variables. After deciding on the initial membership functions and number of the fuzzy rules, the following arrangement is made for training our Sugeno-based ANFIS. The training data is divided into 2 parts, whereby 10% is allocated for checking and 90% is allocated for training. The checking data set is used for avoiding over-fitting and ensuring generalizability. Meanwhile, the training set is used for tuning the antecedent and consequent parameters of the Sugeno model.

The maximum number of training epochs and the training error goal is set to 15 and 0.01, respectively. Fig. 4 shows that the Root Mean Square Error (RMSE) of the checking dataset hits the lowest point of 0.43 after three epochs of training. After this, the error increased steadily to 0.48 at the fifteenth epoch. Thus, the values of the model parameters after the third training epoch is kept for model testing.

22

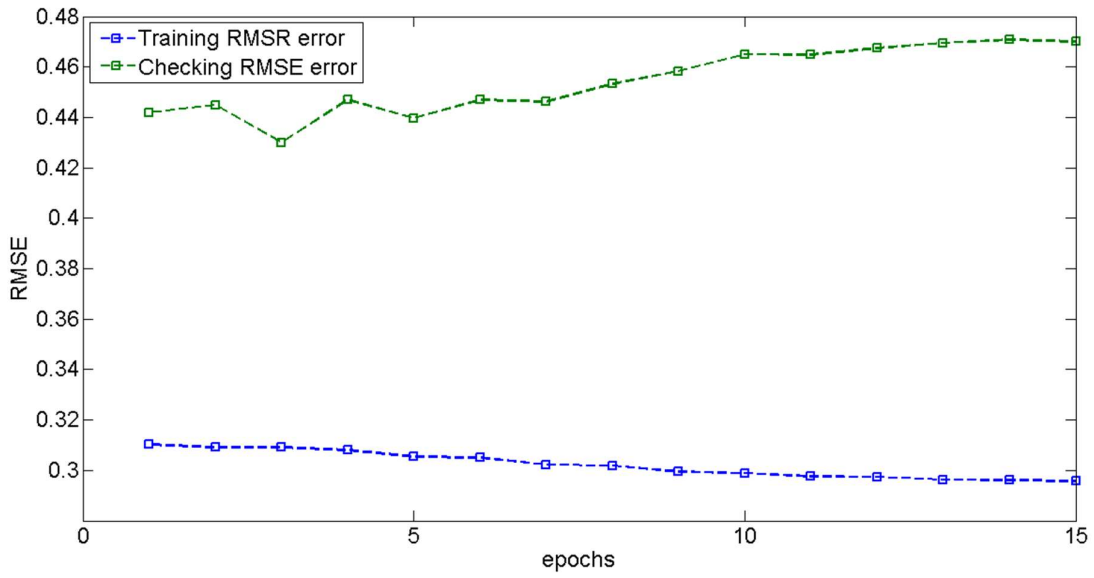Malaysian Journal of Computer Science. Vol. 29(1), 2016

Fig. 4 RMSE error

As a result of the training process, the model parameters learn the best values. We compute the predicted prices of the properties in the test set based on our trained ANFIS model and plot the results as depicted in Fig. 5.
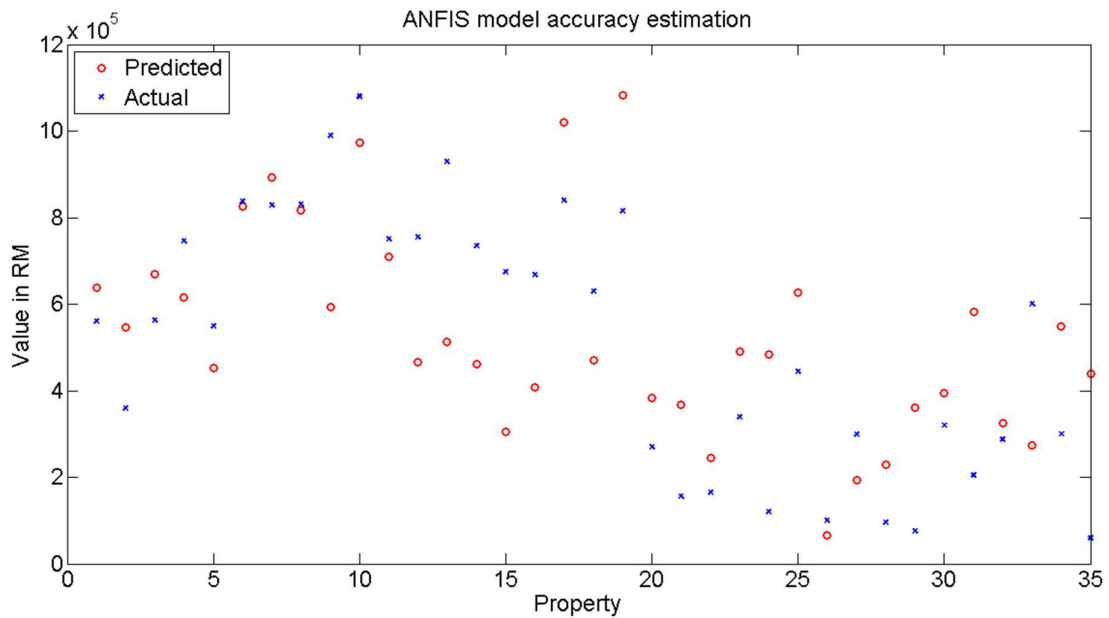

Fig. 5 Results of applying ANFIS model to the test set

The MAE for applying this ANFIS model on the test set is 187,350.

23

Malaysian Journal of Computer Science. Vol. 29(1), 2016

5.3 FLSR Simulation Results

In order to determine the optimal values of the parameters $\widetilde{\beta_0}, \widetilde{\beta_1}, .. , \widetilde{\beta_8}$, we set up a linear system of 27 equations with 27 unknowns. These unknowns are the coefficients, $\beta_l, \beta_m, \beta_r$, of each of the eight parameters $\widetilde{\beta_i}$ ($i=1,...,8$) represented as a TFN (see Fig. **2**). Since MATLAB has no built-in functions for fuzzy regression, we construct the model ourselves and solve the equations in MATLAB. The results of applying our proposed FLSR model are displayed in Fig. 6. The same test set is used, whereby the estimated prices are plotted against the actual ones.
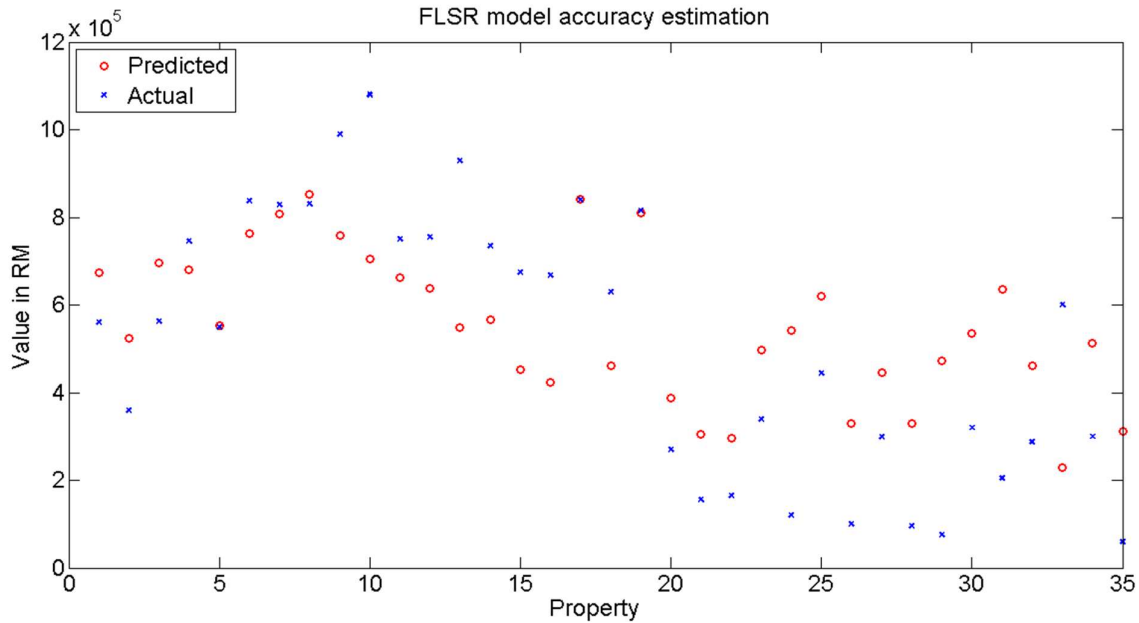

Fig. 6 Result of applying FLSR model to the test set

The figure shows how FLSR model captures the trend of the prices data in almost the whole range of the test set. The MAE for applying this FLSR model on the test set is 183,350.

We also perform another two sets of simulation experiments on two test sets of different sizes, with the three prediction models trained on 75% and 85% of the samples in the dataset. The MAE values resulted from applying the models to the test set of the remaining samples are shown together with the result of the previous simulation.

Table 3. Comparisons of MAE

	90% of data for training	85% of data for training	75% of data for training
FLSR	183,350	189,570	183,190
ANFIS	187,350	204,900	247,170
ANN	211,580	218,750	251,320

In Fig. 7, we summarize our results by comparing the three models, ANN, ANFIS and FLSR, with respect to the mean-absolute error of applying them to training sets of different sizes.
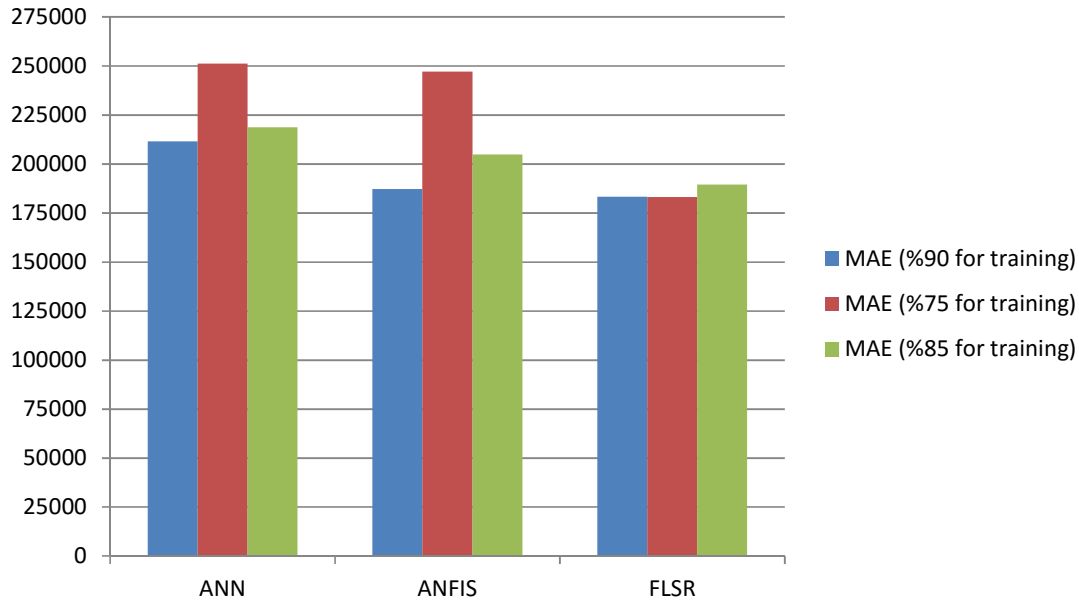
24

Fig. 7 Summary of the results

The bar chart in Fig. 7 shows that FLSR outperforms both ANFIS and ANN in all the three experiments, where the maximum reduction in MAE value is achieved in FLSR. These results support the contention that the fuzzy regression-based approach predicts the real estate property prices more accurately as compared to the ANFIS and classical ANN approach.

The only computationally expensive step in the proposed FLSR algorithm is solving the linear system of equations (LSE). The complexity of solving a LSE is O (n^2), where n is the number of input variables (i.e., real-estate features). However the computational complexity of FSLR is still lower than the complexity of the ANN model and ANFIS model. The time complexity for a standard back-propagation neural network on a single neuron is O (n^3), where n is the number of weights, which is equal to the number of inputs. Similarly, ANFIS has time complexity of O (n^3) as it uses neural networks to adjust its antecedent and consequent parameters as described in Section 3.1. This shows that the complexity of ANN and ANFIS models are at least O (n^3). The obtained results indicate that FLSR can effectively forecasting the real estate prices.

6.0 CONCLUSIONS

Real estate price prediction demand models that are capable of accurately and efficiently predicting property values. Three different modeling approaches, namely Artificial Neural Networks (ANN), Adaptive Neuro-Fuzzy Inference Systems (ANFIS) and Fuzzy Least-Squares Regression (FLSR) were applied to model the price prediction.

The superior performances of the proposed fuzzy least-squares regression-based approach hold important promise for the real estate valuation. The fact that such performance is attained on a relatively small sample size underscores the potential even more. However, it would be pre-emptive perhaps to claim conclusively on the basis of this single result. For example, could a different dataset involving a different area results in a different outcome? Further works are thus needed.

Nevertheless, we have demonstrated the utility of fuzzy logic-based approaches for real estate value prediction in this study. With this pioneering work, we hope to offer a potential new addition to the toolkits for valuation practice and more importantly, to the body of knowledge for increased incorporation of 'out-of-the-box' techniques for estimating real estate values.

25

REFERENCES

[1] Pagourtzi, E., Assimakopoulos, V., Hatzichristos, T. & French, N. "Real estate appraisal: a review of valuation methods". *Journal of Property Investment & Finance*, Vol. 21, No. 4, 2003, pp. 383-401.

[2] Ismail Omar. *Penilaian Harta Tanah.*, Kuala Lumpur, Dewan Bahasa dan Pustaka, 1997.

[3] International Valuation Standards Council. *International Valuation Standard*, London, 2013

[4] Brunson, A., Buttimer, R. J. & Rutherford, R. "Neural networks, nonlinear specifications, and industrial property values", University of Texas at Arlington, *Working Paper Series*, 1994.

[5] Do, Q. & Grudnitski, G. "A neural network approach to residential property appraisal". *The Real Estate Appraiser*, 1992, pp. 38-45.

[6] Limsombunchai, V., Gan, C. & Lee, M. "House price prediction: hedonic price model vs artificial neural network". *American Journal of Applied Sciences*, Vol. 1, No. 3, 2004, pp. 193-201.

[7] Daniels, H. & Kamp, B. "Application of MLP Networks to Bond Rating and House Pricing". *Neural Computing & Application*, Vol. 8, 1999, pp. 226–234.

[8] Nguyen, Nghiep & Cripps, Al. "Predicting housing values: A comparison of Multiple Regression Analysis and Artificial Neural Networks". *Journal of Real Estate Research*, Vol. 22, No. 3, 2001.

[9] Lai, P. & Fischer, D. "Artificial Neural networks and Computer Assisted Mass Appraisal", in *Proceedings of the 12th Annual Conference of the Pacific Rim Real Estate Society*, January 22-25, 2006.

[10] Zurada, J.M., Levitan, A.S. & Jian, G. "Non-Conventional Approaches To Property Value Assessment". *Journal of Applied Business Research*, Vol. 22, No. 3, 2006.

[11] Kusan, H., Aytekin, O. & Özdemir, I. "The use of fuzzy logic in predicting house selling price". *Expert Systems with Applications*, Vol. 37, No. 3, 2010, pp. 1808-1813.

[12] Liu J.M., R. Chen, L.M. Liu & J.L. Harris. "A semi-parametric time series approach in modeling hourly electricity loads*". Journal of Forecasting*, Vol. 25, 2006, pp. 537−559.

[13] Guan, R., Tian, D. & Wang, X. "Design and implementation of LED daylight lamp lighting system", in *Proceedings of the IEEE international conference on electronic packaging technology & high density packaging*, 2008, pp. 1–3.

[14] Qazi, A., Fayaz, H., Wadi, A., Raj, R.G., Rahim, N.A., Khan, W.A., "The artificial neural network for solar radiation prediction and designing solar systems: a systematic literature review", Journal of Cleaner Production, Volume 104, 1 October 2015, Pages 1-12, ISSN 0959-6526, http://dx.doi.org/10.1016/j.jclepro.2015.04.041. (http://www.sciencedirect.com/science/article/pii/S0959652615004096).

[15] Yeow, W.L., Mahmud, R., Raj, R.G., "An application of case-based reasoning with machine learning for forensic autopsy", Expert Systems with Applications, Volume 41, Issue 7, 1 June 2014, Pages 3497-3505, ISSN 0957-4174, http://dx.doi.org/10.1016/j.eswa.2013.10.054. (http://www.sciencedirect.com/science/article/pii/S0957417413008713).

26

[16] Moohebat, M., Raj, R.G. , Kareem, S.B.A., Thorleuchter, D., "Identifying ISI-indexed articles by their lexical usage: A text analysis approach", Journal of the Association for Information Science and Technology, Vol. 66, No. 3, pp. 501–511. doi: 10.1002/asi.23194.

[17] Takagi, T. 7 Sugeno, M. "Fuzzy identification of systems and its applications to modeling and control". *IEEE Transactions on Systems, Man and Cybernetics*, Vol. 15, No. 1 (January/February), 1985, pp. 116-132.

[18] Diamond, P. "Fuzzy least squares". *Information Sciences*, Vol. 46, No. 33, 1988, pp. 141–157.

27

Malaysian Journal of Computer Science. Vol. 29(1), 2016