Proceedings of the International MultiConference of Engineers and Computer Scientists 2013 Vol II,
IMECS 2013, March 13 - 15, 2013, Hong Kong

Determination of Housing Price in Taipei City Using Fuzzy Adaptive Networks

Wei-Tes Lee, Jian-Jiun Chen and Kuentai Chen*,

Abstract—The main purpose of this research is to forecast pre-owned housing prices using Fuzzy Adaptive Network (FAN). Fuzzy adaptive network (FAN) combines fuzzy inference system with neural networks and utilizes the strengths of both methods. It has the power of fuzzy inference and the learning ability from neural networks. Fuzzy variables are first introduced to the area of predicting housing prices. The results are compared to back-propagation neural network (BPNN) and adaptive network fuzzy inference system (ANFIS). Fuzzy variables used in this study include livability, ambient conditions and expected development potential.

Conclusions of this research are: (1) the housing prices in the Taipei Administrative Region fall into three clusters; (2) the overall predictions in the Neihu District are better than the Da-an District; (3) the predictions of FAN, BPNN and ANFIS are MAPE of 4.54%, 8.27% and 15.07%, respectively. That is, FAN performs better than BPNN and ANFIS in predicting the housing prices of Taipei city.

Keywords — real estate prices, fuzzy adaptive network, back-propagation neural network and adaptive network fuzzy inference system to forecast.

I. Introduction

Taipei City, as the capital of Taiwan, is well-developed and its constructions are fairly of advanced. The citizens have privileges of very good living conditions and convenient traffic network, which in return brings up much higher housing prices. With such a high price, each purchase of a house is an important decision to make for a family. The price of a house is usually determined by both objective conditions of the house and the subjective preferences of the buyer. The objective house conditions include size of living area, parking space, distance to school or metro system or bus station, and so on. On the other hand, the subjective preferences vary by different buyers. There are many subjective factors affecting the price of real estates, such as the disgust facilities, convenience of living, and the neighborhood safety. Traditional predicting methods concentrated on the objective conditions with the attempt to provide an unbiased reference price for the buyers. However, the problem is that the price a buyer is willing to pay is in natural never unbiased because of many personal reasons.

Manuscript submitted on December 28, 2012. This work was supported by the Ming Chi University of Technology. All authors are with the Department of Industrial Engineering and Management, Ming Chi University of Technology, Taipei, Taiwan. (+886-2-29089899-3111; fax: +886-2-29085900) The email of wei-Tes Lee is fgh1110@hotmail.com, the email of Jian-Jiun Chen is sing0702@hotmail.com while the e-mail of Kuentai Chen (corresponding author) is: kuentai@mail.mcut.edu.tw.

To address this problem, we propose to build a model with both fuzzy and crisp variables. The introduction of fuzzy subjective factors is in hope of improving the accuracy of prediction to Taipei's housing prices. In this research, the prices of pre-owned houses are selected as the target, and the collected trading records are from January 2009 to June.

II. Literature

A. Variable Selection

The real estates are inherently permanent commodities regarding political, economic, supply and demand, global and local considerations, environmental conditions and many other factors. There are many direct and indirect factors that can affect their sales price, which makes it difficult to predict. Some researches are conducted to the price model building, but the uses of impact variables are never the same. This study tries to collect variables used in the past articles, and the result is shown in Table 1.

Studying the previous research related to the real estates, we categorize the independent variables as objective and subjective factors. These variables are then divided into two categories: fuzzy variables and crisp variables.

1) fuzzy variables: we propose to use three fuzzy variables, namely the Livability, Ambient conditions, and the Expected development potential.

2) crisp variables: limited to the data collection difficulty, only part of crisp variables are available to us. These variables include the floor space, number of floor, total number of floors in the building, land stakeholders and housing age.

Chang (Chang and Chang, 2005) pointed out that the back-propagation neural network utilizes the Widrow-Hoff learning rule generalized from the multi-layered with nonlinear differential transfer function of the network. Back-propagation neural network architecture can be divided into three layers, the input layer, a hidden layer and output layer. The input layer basically does sample input, hidden layer has a double curved line transfer function, and the output layer has a linear transfer function. Therefore, the output value of the network can be more than one non-consecutive points of any function, which explains problem of the non-linear input variables.

A well-known study named Adaptive network fuzzy inference system (ANFIS), introduced by Zhang, has its model architecture combined by the fuzzy inference system and fuzzy inference system extension developed by fuzzy theory. It collects the given input and output variables to generate a set of fuzzy rules and membership functions

TABLE I
Real estate-related variable selection

Author	Research themes	Select a variable
Chin-Oh Chang 、Steven C. L. Farr(1993)	A STUDY OF Real Estate Transaction Price	Period, location, use patterns, housing age, on the ground floor, where the floor area
Mok et al. (1995)	A Hedonic Price Model for Private Properties in Hong Kong	Away from the commercial center, floor, field of view, floor area, housing age, fitness activities, shopping malls
LIN-SU-CHING (2002)	The Elasticity of Consumption and Investment for Housing Demand in Taiwan	Floor number, housing age, residential structure, residential use patterns, its own kitchen, household currently living satisfaction
Chen Hung-Chou (2002)	A Research of Applying Multicriteria Evaluation with Qualitative and Quantitative Data Model in Real Estate Appraisals	Regional environmental factors:Frontage road width, traffic route distribution and the adjacent building of the degree of Transportation: Close to the center of the degree, nearly to the extent of the business district and close to the parking level Individual characteristics factors:Degrees of vision near and far, housing patterns, lighting, housing, age, maintenance facilities as well as the Public than
Chen Hui-Chieh (2007)	A Study on the Factors Attributing to the Prices of Residential Buildings	The tsubo Price, floors, the face of the width of the road, the straight-line distance of the residential building and park
Li (2007)	Using Fuzzy Neural Network in Real Estate Prices Prediction	The characteristics of the real estate, quality, location, environment, and residents pay levels

parameters. Fuzzy inference fuzzy system is the most important core mode, it can simulate human thinking approach to solve the problem and thus achieve the intended purpose. (Zhang, 1993)

III. Research methods

This study uses Matlab 2008a programming software to code FAN model. Then FAN is compared with BPNN and ANFIS prediction model. The core purpose of the study is committed to improve the quality of the forecast of real estate prices. We collected real transaction data from the real estate transaction Quotes Bulletin Taipei from January 2009 to June period.

Fuzzy adaptive network (FAN) is introduced by Cheng and Lee (1999). The proposed framework is formed by combining the fuzzy inference skills and neural network learning ability.

The Fuzzy adaptive network is formed basing on past use of fuzzy adaptive networks in different research fields. FAN uses linguist terms to represent how human spoken language to convey the message. The training ability of FAN is obtained from the error backpopagation of the network. Some research have successfully applied FAN in different fields (Jiao et al, 2006; Chen et al, 2006).

Fuzzy adaptive network is a network model of a total of five-layer network architecture (Figure 1). The structure is similar to adaptive network fuzzy inference system but different in how the minimal error is reached. The nodes in the architecture are mainly divided into two parts: circles and square nodes. Circle nodes are fixed nodes that perform certain transformation function, while the square nodes can be trained. FAN calculates premise parameters as well as the solution of fuzzy linear programming and feedback the error to push the front layers of the network through backpropagation. The correction of the premise parameters continues training until it reaches the desired minimum allowable error or maximum epochs to train. Fuzzy inference system basically consists of three concepts: (1) the rule base: a set of fuzzy If-Then rules; (2) database: using fuzzy rules to define its membership function; (3) the inference mechanism: based inference rules to perform the inference process conditions given to draw reasonable output value. Various fuzzy inference systems can build a framework based on the If-Then rules and aggregation process, while the FAN is derived via Takagi and Sugeno inference model, which retains the skills of fuzzy systems and neural network learning ability to provide an adaptive network system. The FAN has five layers:

The first layer: This layer consists of adaptive node containing the premise parameters mean and standard deviation for the premise part of the fuzzy rules.

Second layer: the nodes in the layer are fixed nodes denoting the attribution of different semantic fuzzy rules (Fuzzy Rule). Multiplying is performed in order to calculate the intensity value of each node and then output to the next network layer.

Third layer: The network layer node are fixed nodes to calculate the weights of the previous links. Normalization is performed so that the output values of weights are between0-1.

Fourth layer: The network layer nodes are adaptive nodes. Y^l is the inference part of fuzzy rules, and the nromalized weights $\overline{w}^l$ are multiplied.

Fifth Layer: Summation.

This research applies Mean Absolute Percentage Error (MAPE) as the error measure. The smaller the MAPE is, the

ISBN: 978-988-19252-6-8
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

better the model performs. To estimate the prediction error, the MAPE formula is as follows:

$$MAPE = \left[\Sigma \left| \frac{P_i - T_i}{T_i} \times 100 \right| \right] \div N$$

where P_i is predictive value, T_i is actual value, and N is total number of samples.

In this research, in addition of MAPE, we also calculated mean square error (MSE) as a basis to determine the predictive ability of the network model. The MSE mainly estimated the prediction accuracy of the network model. The smaller the MSE is, the smaller the degree of dispersion between the predicted value and the actual value is. The MSE calculation can be expressed by the following formula:

$$MSE = \frac{\sum_{i=1}^{N} (y_i - \hat{y}_i)^2}{N}$$

where y_i is the actual value, $\hat{y}_i$ is the prediction, and N is the Total number of samples.

The referenced experts suggested using road blocks in major districts in Taipei City when designing a questionnaire for collecting information of fuzzy variables. For example, the roads divide a single administrative area into several blocks for experts to grade the living function of different blocks, as well as the surrounding environment conditions and the expected development potential. In this way, scores are obtained more efficiently and it allows experts to judge the extent of the pros and cons of the different blocks of the same administrative region.

Clustering analysis is also performed in this research. As there exists twelve administrative regions in Taipei City, they are divided according to their properties. The blocks of similar nature are combined into the same cluster such that a high degree of heterogeneity exists between different clusters.

IV. RESULTS

As shown in Table 3, the correlation coefficient matrix of all crisp variables used in this research is calculated. It suggests that area of use and the trading price of the transaction has a linear relationship, while no significant linear relationship exists between other variables.

A. Residual analysis

Total number of 1255 trading recording are considered in this research. Residual analysis shows that the samples are not normally distributed and there exists some outliers. Therefore, outliers are identified and deleted form the samples for further analysis.

B. BPNN forecast results

In this study, BPNN is set to use the Bayesian function, one hidden layer, 10 neurons in hidden layer, and 3000 training epochs. The following examples selected two districts, Neihu District and Daan District, to demonstrate the price prediction. The research results show that: (1) Neihu District, using 117 samples for the training generates MAPE 16.34% as the training error, and the MSE of 51 987 (million). For 13 testing samples, the predicted results in MAPE of 15.58% and MSE of 37,920.5 (million), as shown in Figure 3. (2) Da-an District, using 171 training samples, results in a training error of 15.82% MAPE, and 76,044.14 (million) MSE. For 20 testing samples, the MAPE is 14.56% and the MSE is 63541.71, (million). The predictions are as shown in Figure 2 and Figure 3.

C. ANFIS prediction results

In this research, the ANFIS is set to use Gaussian membership functions, training 3000 iterations. The same two districts, Neihu District and Daan District, are analyzed for comparison. The results show that: (1) Neihu District, MAPE is 7.14% and the MSE is 11,187.39 (million). For the 13 testing samples, the MAPE is 6.89% and the MSE is 13,871.87 (million), as shown in Figure 3. (2) Daan District, the MAPE is 11.62% and the MSE is 39,483.44 (million). For the 20 testing samples, the MAPE is 9.65% and the MSE is 28,335.22 (million). The predictions are as shown in Figure 4 and Figure 5.

D. FAN forecast results

After screening of the best changes in parameters of the Daan District and Neihu District, we found the best combination of parameters are the same for both two districts with confidence level = 0.6, fuzzy width e_k = 0.1, and learning rate L-rate = 0.03. The minimum allowable error is set to E = 0.01. The results are as follows:

1) For the Neihu District, the training error converged with 3436.977 (10K). The 13 testing samples result in MAPE of 3.79% and MSE of 1941.222 (million), while the training error of MAPE is 3.70% and MSE is 2598.852 (million). The predictions are as shown in Figure 6.

2) For Daan District, the training error converged with 7257.618(10K). Using 20 test samples for prediction, the prediction results in MAPE of 5.28% and MSE of 7257.618 (million), while the 171 training samples with MAPE of 2.60% and the MSE of 1150.229 (million). Figure 7 shows the predictions are very close to the target values.

ISBN: 978-988-19252-6-8
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

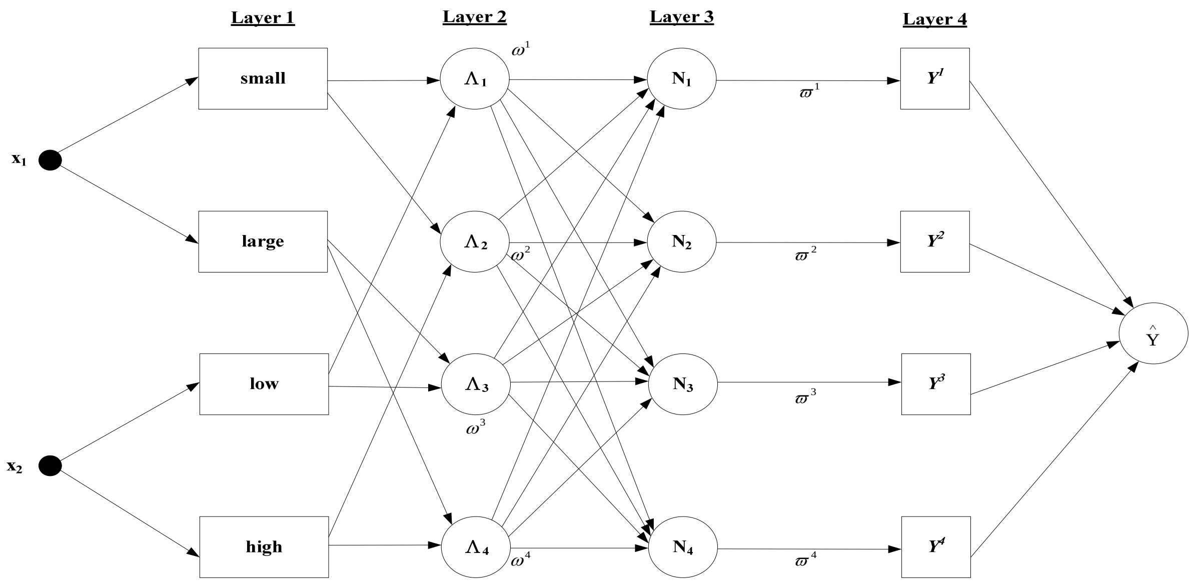
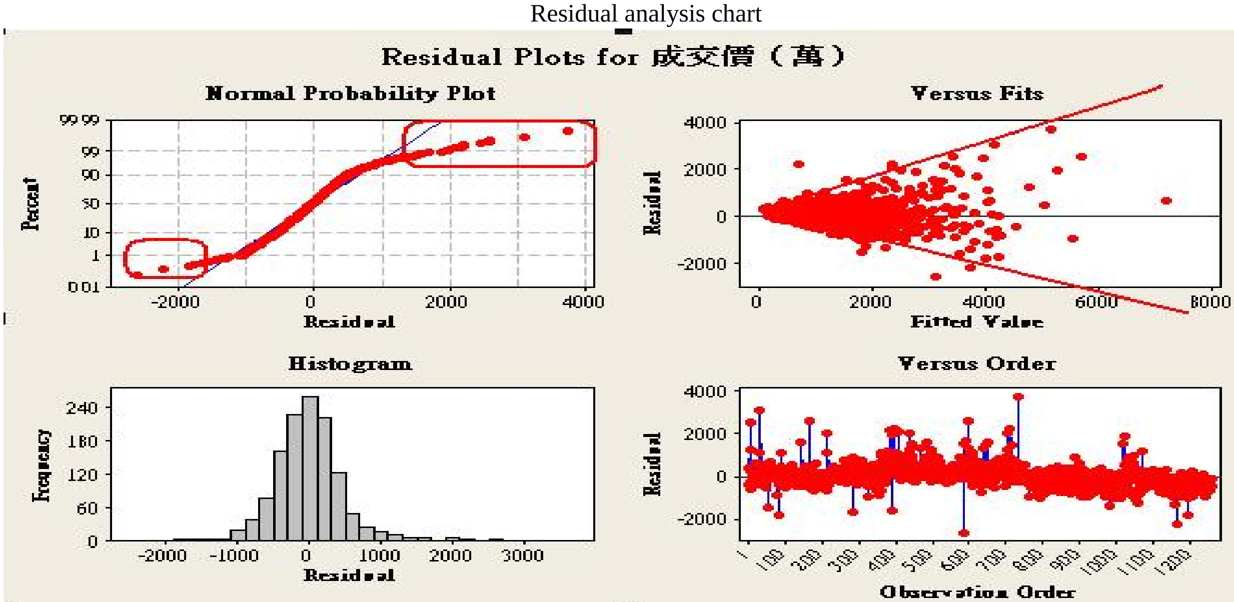


Fig. 1. Adaptive Network Fuzzy Inference System Architecture

TABLE II Correlation coefficient matrix

	Area of use	Land stakeholders	Floor	Total number of Floor	Year of construction	Turnover Total price
Land stakeholders	0.342					
Floor	0.036	-0.089				
Total number of Floor	0.083	-0.146	0.490			
Year of construction	-0.021	-0.042	-0.087	0.258		
Turnover Total price	**0.856**	0.276	0.070	0.132	-0.020	
per pyeong Price	0.024	-0.028	0.092	0.102	0.057	0.463

TABLE III
Residual analysis chart



ISBN: 978-988-19252-6-8
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

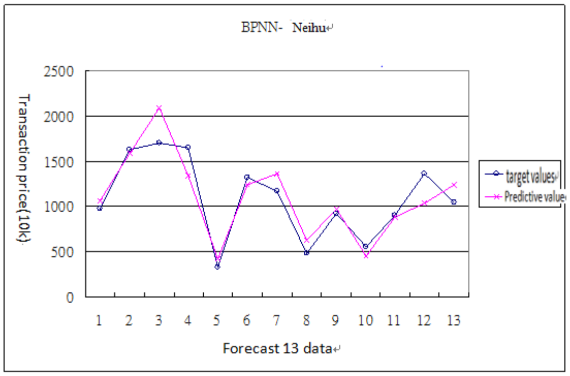


Fig. 2. BPNN forecast results in the Neihu District

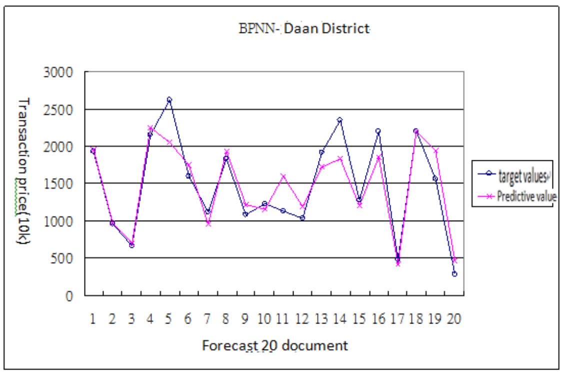


Fig. 3. BPNN forecast results in the Daan District

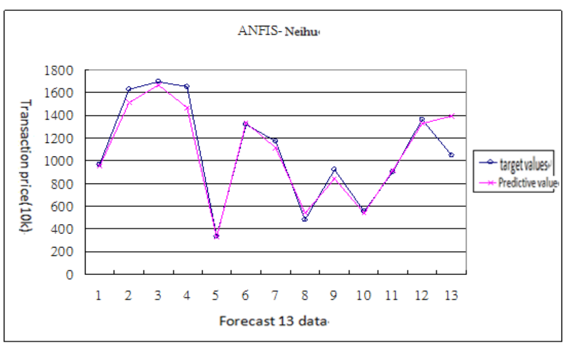


Fig. 4. ANFIS in Neihu forecast results

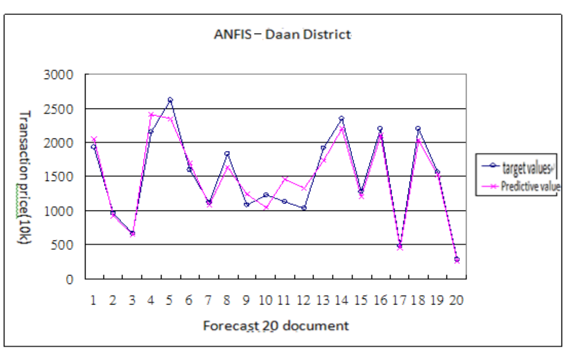


Fig. 5. ANFIS forecast results in the Daan District

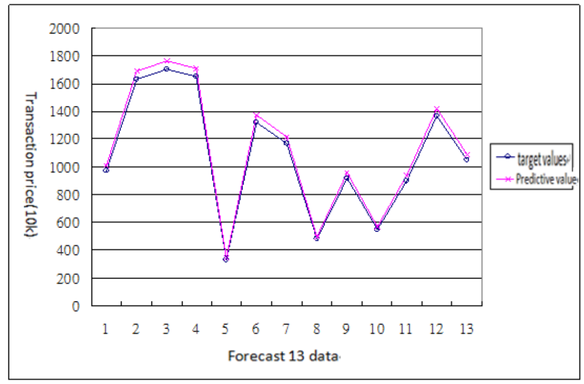


Fig. 6. The target value and the predictive value of the 13 test samples in Neihu District

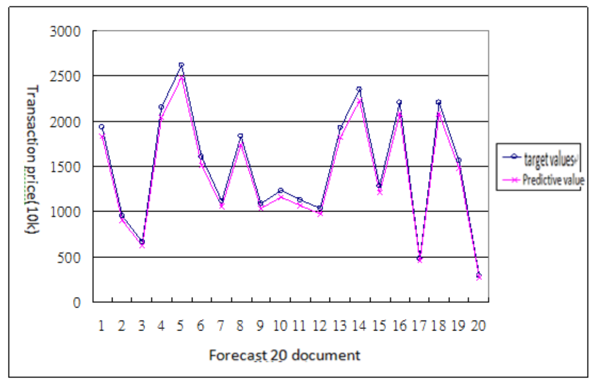


Fig. 7. The prediction of Daan District housing prices using 20 test samples

E. *Comparison*

This research uses three predictive models, namely FAN, BPNN, and ANFIS, to predict the housing prices. A total of six variables were included in these models, including three crisp variables (area to use, land stakeholders, house age) and three fuzzy variables (living functions, the surrounding environmental conditions, expected development potential). For demonstration, Neihu District and Daan District are selected as examples. In comparison, we found that:

1) FAN model has the best performance among three methods when predicting both Districts. Its MAPE decreased to 3.79% and mean square error (MSE) to 1941.22 (million). In other words, the performance is FAN>ANFIS>BPNN.

2) The overall predictions in Neihu District are better than in Da-an District for all methods. As we know that the numbers of samples used for training are 117 and 171, respectively, it suggests that the size of 117 samples is good enough for training.

TABLE Ⅳ
Three forecasting models predict the effect of finishing

District / Model	Neihu District, predicted results		Daan District forecast results	
	MAPE %	MSE(Ten thousand)	MAPE %	MSE(Ten thousand)
FAN	3.79	1941.22	**5.28**	**7257.618**
BPNN	15.58	37920.50	14.56	63541.71
ANFIS	6.89	13871.87	9.65	28335.22

ISBN: 978-988-19252-6-8
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

V. Conclusion

In recent years, real estate prices in Taipei are much higher than many other cities, a rising trend of housing prices in Taipei city can be easily observed. More and more people and government agencies become aware of this problem. With this high price, it is important to obtain more information for the buyers before they make the decisions. Thus an efficient decision support will be desired and demanded. In this study, a fuzzy adaptive network (FAN) is proposed to predict the real estate prices in Taipei City through both crisp and fuzzy variables. The results are compared with BPNN and ANFIS. The contribution of this study and the results can be summarized as the following:

1) In this study, the advantages of incorporating fuzzy variables in prediction of housing prices are demonstrated. In other words, both objective variables and subjective variables should be considered.

2) For prediction, Taipei's twelve Administrative Region should be clustered before model construction.

3) Sample size of 117 for training using FAN can provide satisfactory results.

4) In this study, FAN performed better.

5) Predictions in Neihu District are better than in Da-an District using FAN, ANFIS, or BPNN.

Acknowledgement

This research is funded by Ming Chi University of Technology, Taiwan.

References

[1]. LIN-SU-CHING (2002), "The Elasticity of Consumption and Investment for Housing Demand in Taiwan", The ROC Residential Society 11th Annual Conference Proceedings,parer268 of paper277.

[2]. Chin-Oh Chang、Steven C. L. Farr(1993),"A STUDY OF Real Estate Transaction Price", Chaoyang University of Technology Architecture and Urban Design, Vol.

[3]. Hui-Chieh Chen (2007),"A Study on the Factors Attributing to the Prices of Residential Buildings", Chaoyang University of Technology, Department of Finance Thesis.

[4]. Chin-Oh Chang、Steven C. L. Farr(1993),"A STUDY OF Real Estate Transaction Price", Residence of a 75 -97 page.

[5]. Chen, K. T., and E. S. Lee (2006). "Fuzzy Adaptive Networks in Credit Rating and Loan Approval."

[6]. Li, Z. X. (2007). "Using Fuzzy Neural Network in Real Estate Prices Prediction." Proceedings of the 26th Chinese Control Conference, Hunan, China.

[7]. Li, D. Y., W. Xu, H. Zhao, and R. Q. Chen (2009). "A SVR Based Forecasting Approach for Real Estate Price Prediction." Proceedings of the Eighth International Conference on Machine Learning and Cybernetics, Baoding.

[8]. Mok, H., P. Chan, and Y. S. Cho (1995). "A Hedonic Price Model for Private Properties in Hong

[9]. Jang, Jyh-Shing, (1993), "ANFIS : Adaptive-Network-Based Fuzzy Inference System", IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS, VOL. 23, NO. 3, MAYIJUNE 1993