Soft computing algorithms in Price of Taiwan Real Estates

HONG –YU LIN[1], KUENTAI CHEN[2*]
[1]Department of Business and Management,
Mingchi University of Technology
84 Gungjuan Road, Taishan, New Taipei 243, TAIWAN
hylin@mail.mcut.edu.tw
[2*]Department of Industrial Engineering and Management,
Mingchi University of Technology
84 Gungjuan Road, Taishan, New Taipei 243, TAIWAN
kuentai@mail.mcut.edu.tw

Abstract：- The prediction of real estate price is important because it concerns both individuals and government. Other than traditional statistic methods, Neural networks and Support Vector Regression have demonstrated their advantages in previous research, and thus are applied and compared in this study. Variables are first summarized and concluded from earlier research and than selected by stepwise procedure and trial-and-error methods. It is found that SVR with trial-and-error method performed the best with MAPE=4.466% and R^2=0.8540. In addition, Rediscount rate, Money supply, and Price of last month are the three common variables for both BPNN and SVR. The economic explanation and relations to the housing price for selected variables are also provided.

Keyword：- Forecasting, Real estate, Back propagation neural network, Data mining, Support Vector Regression

1 Introduction

The real estate price concerns both individuals and government and thus usually can be treated as in important index to economic. Economic bubbles were born and burst in many countries since 90's, causing their economic growths went up and down dramatically. Taiwan in 1990s and United States in 2008 are two typical examples. One of the main factors in formation and rupture of these economic bubbles is the price volatility of real estates (Kaashoek and Dijk, 2002). One would be prepared and make correct decision when the prices vary if mastering the price trend. Therefore, to capture the changing tendency and then predict the trend of real estate price precisely has been an important research direction for Governments and researchers.

Traditional prediction in the real estate price is to build a regression model, which is usually confined by certain function and thus does not always provide a desired solution (Tay and Ho, 1992; Do and Grudnitski, 1992). Another common tool since 90s is the neural network, which is capable of dealing with nonlinear problems. It also has advantages in requiring less pre-assumptions and having learning ability, as long as enough data are supplied (Tsoukalas and Uhrig, 1997; Kuan and White, 1994; Lin and Lee, 1996). Some successful applications are reviewed below. For examples, Do and Grundnistski (1992) applied neural networks and multi-regression analysis to test the models with 105 data of house trading, resulting in absolute error of 6.9% and 11.3%, respectively. Tay and Ho (1992) adopted the same methods to predict prices of departments in Singapore using 822 training data and demonstrated that neural networks performed better than regression analysis. Similar applications can be found in McCluskey et al. (1997), McGreal et al. (1998), and Wong et al. (2001). A common character in these papers is that Back Propagation Neural Network (BPNN) is the primal prediction tool. However, some researchers argued that BPNN has some disadvantages such as requiring large amount of data, hard to find stable optimal solution, and possible over-fitting problems (Worzala et al. 1995). But it is still a popular and valid method in many applications.

As a recent popular method, Support Vector Machine (SVM) was proposed by Vapnik (1995) with principles of empirical risk minimization (ERM) and Structural Risk Minimization (SRM), which ensure the unique optimal solution without over-fitting. SVM was first applied in recognitions of handwriting, image, and voice. Vapnik et al.(1997) further proposed Support Vector Regression (SVR) for predictions and led to many successful applications in different aspects, such as prediction in stock market (Trafalis and Ince, 2000), Futures Contract (Tay and Cao, 2001), seasonal GDP of industrial machinery (Pai and Lin, 2005), and Tokyo Nikkei 225 index (Huang, Nakamori and Wang, 2005). With these successful applications, it is our attempt to apply BPNN and SVR to predict the price of real estates.

2 Data and Research methods

It is now widely accepted that the price of real estate represents the real estate climate in many aspects. In Taiwan, the price of Taipei's real estate differs significantly from the rest of cities, as Taipei city is the capital city. Therefore, data from other well-developed cities are collected to represent the real estate climate.

The data adopted in this study are mainly purchased from Taiwan's Real Estate Portal. Data are collected and averaged by month, including Taipei county, Taichung city and county, and Kaohsiung city and county. The time of collected data is from 2004/1 to 2008/12. Figure 1 shows the increasing trend of Taiwan real estate price by month, the Y-axis is the prices in 10K.

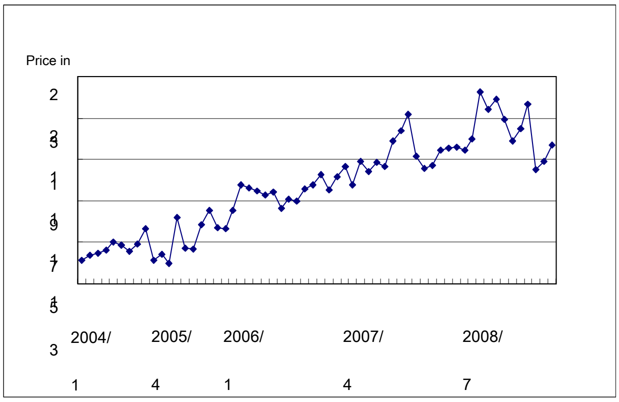

Figure 1. The trend of Taiwan real estate price

In this study, two popular prediction tools in soft computing, namely Back Propagation Neural Networks (BPN) and Support Vector Regression (SVR), are implemented to perform the training and predictions. Soft computing technologies have been widely implemented in many fields, including finance. Readers are referred to Chen and Lee (2006) for a BPN application and Holimchayachotikul et al. (2011) for a SVR application. Since these methods are now widely applied in many domains, only brief review and references are provided below.

The BPNN is a three-layer, feed-forward network with backward error propagation that learns from examples. The three layers are Input layer, Hidden layer, and Output layer as shown in Figure 2. The advantages of BPNN include the ability of training from examples, dealing with nonlinear data, error-tolerance, parallel computing, and being a universal approximator -- as long as enough data are available. Detailed description can be found in Hornik (1989) and a successful application of BPNN in finance can be found in Chen et al.(2009). SVR is originally from support vector machine (SVM), proposed by Vapnik (1995) and is modified for regression purpose in 1997 (Vapnik et al., 1997). The SVR model produced by support vector classification depends on a subset of the training data. SVR model can also cope with nonlinear data and has only a unique optimal solution with each set of kernel parameter and soft margin parameter. In our study, SVR is combined with Grid algorithm for the best combination of parameters.

Both Mean Absolute Percentage Error (MAPE) and Coefficient of Determination (R^2) are adopted as the indices of efficiency to compare the three methods. The equation of MAPE is as shown in the following

$$MAPE = \frac{\sum_{k=1}^{N} \left| \frac{d_k - y_k}{d_k} \right|}{N} \qquad \text{EQ(1)}$$

where d_k is the k^{th} target value; y_k is the k^{th} output; and N is the total number of data.

And the equation to calculate R2 is as shown in EQ(2).

$$R^2 = \frac{SSR}{SST} = \frac{\sum_{i=1}^{N} \left(t_i - \overline{y}\right)^2}{\sum_{i=1}^{N} \left(y_i - \overline{y}\right)^2} \qquad \text{EQ(2)}$$

where y_i is the i^{th} target value, t_i is the i^{th} output, and N is the total number of data.

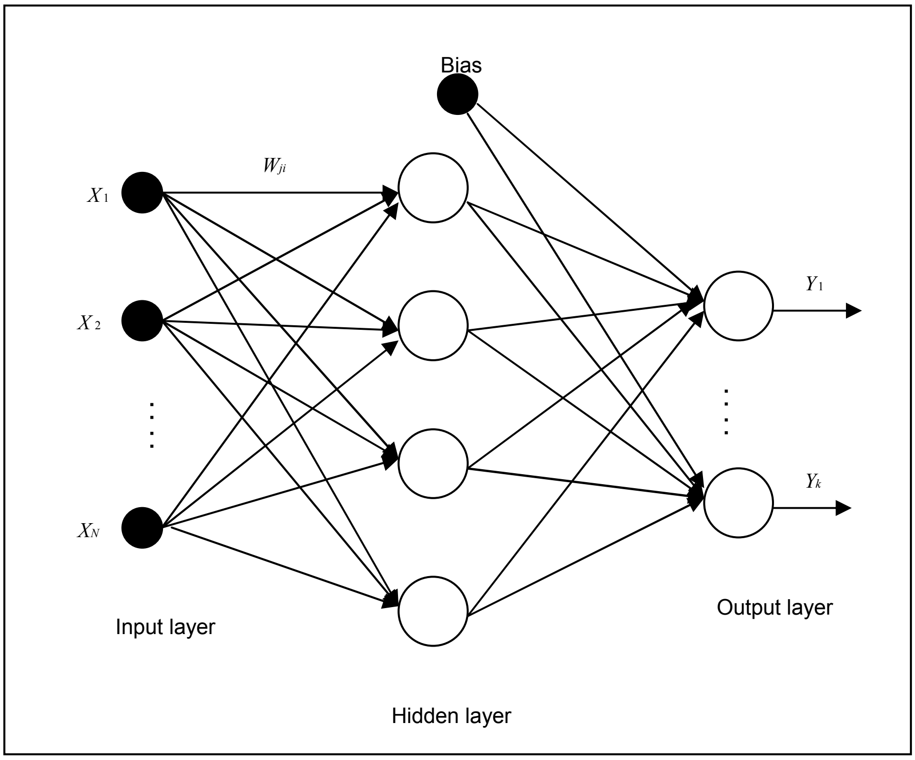

Figure 2. The structure of BPN

Original figure: Wei, H.(2004).

3 Data Analysis and Comparison

3.1 Variable Selection

According to the official publication "Real Estate Cycle Indicators of Taiwan Ministry of Interior" (in Chinese), the effective variables can be classified into three categories: leading indices, simultaneous indices, and lagging indices as shown in Table 1.

For prediction, only leading and simultaneous variables are considered in this research. Among the variables list in Table 1, due to the difficulty of obtaining data, three variables, namely *Volume index of prime land*, *Residential use rate*, and *Building Permit of Residence*, are not included. In addition, *Benchmarking lending rate* is replaced by *Rediscount rate* because the later one is an important tool when Taiwan Central Bank changes currency policy. In Taiwan, the measurement of *Money supply* includes three variables: *M1a*, *M1b*, and *M2*. The definitions are as the following:

M1a = Net currency + check deposit + current account

M1b = M1a + current savings account

M2 =M1a + current savings account + Quasi-money, or *M1b+ Quasi-money.*

Another important variable included is the so-called *transaction volume per month*. Although the *transaction volume per month* is a lagging index, it is considered because that, according to Hua and Chang (1997), "in the long run, the fluctuation of transaction volume causes the price to fluctuate in housing market".

3.2 Results

As a result, 11 variables plus "price in last month" (namely, *t-1 price*) are included in our models. These 12 variables are: *transaction volume per month, Gross Domestic Product, Rediscount rate, The Standard Unit Price of New Cases, Money supply M1a, Money supply M1b, Money supply M2, New House- purchasing Loans, Consumer Price Index, Construction stock indices, Movement of Loans for Construction, t-1 price*,

Stepwise procedure and trial-and-error methods are applied for the variable selection in BPNN. The results are as shown in Table 2. From Table 2, it can be observed that *Rediscount rate* and *t-1 price* are common variables of both selecting methods. Also, the best prediction is by trial-and-error selection with R^2 of 0.7382.

Repeat the same procedure using SVR, the result is as shown in Table 3. Again, the common variables are *Rediscount rate* and *t-1 price*. *Money supply* can be also considered as a common variable since *M2* and *M1a* are closely related：*M2 = M1a + current savings account + Quasi-money*. The best R^2 is 0.8540 using trial-and-error, which is better than using BPNN. It can be observed form Figure 3 that most predictions of SVR are better than BPN.

In fact, the most contributing variable is *t-1 price*, which is usually a very important reference before we purchase a real estate property. To verify this statement, we performed a test using it as the only variable. The result shows its R^2 is 0.6165 for BPNN (MAPE 5.867%) and 0.5545 for SVR (MAPE 6.152%). The performance of using only *t-1 price*, which is as shown in Figure 4, dose not seem to be satisfactory because it failed to predict the changing points efficiently. Therefore more variables would be desired.

3.3 Discussion

The best result was generated by SVR with trial-and-error variable selection in predicting the price of Taiwan real estate. In this well-performed model, the six contributing variables are *Gross Domestic Product, Rediscount rate, The Standard Unit Price of New Cases, Money supply M1a, New*

House-purchasing Loans, and *t-1 price*. Since these variables are selected, we try to explore the relationship between these variables and the price of Taiwan real estate in the following.

First, *Gross Domestic Product* is an important index used by many countries to evaluate economic development, and is usually treated as the national income level. That is, as the average income of people increase, the housing price for people to afford would also increase.

Second, *Rediscount rate* and *Money supply M1a* are generally the key points when Central Bank of the Republic of China is to execute its currency policy. Meanwhile, they are also critical indices of sufficient/insufficient capital in housing market. We can further examine the impact of *Rediscount rate* by observing the recent *Rediscount rate* as shown in Figure 5 and compare it with Figure 3. Figure 5 shows an obvious drop since 2008/9, and the housing price in Figure 3 also dropped drastically at the same time.

The selection of above three variables by SVR somehow substantiates that the price of Taiwan real estate is, the same with other countries, close-related to economic development, income, and money supply. However, since *Money supply M1a* is more about enterprise legal, it would be crucial to confirm whether the capital flows to enterprise legal if the Central Bank wishes to influence the price of real estate by adjusting currency policy.

In addition, *The Standard Unit Price of New Cases* consists of new units and presale units. Although they differ from the research target (exist units) in this study, they influent the price of pre-owned houses because their unit prices has function of price discovery. In other words, if the prices of new cases increase, the prices of pre-owned houses follow.

The *t-1 price*, which means the previous price of real estate, is naturally very significant information for buyers to make decisions. It plays an important role when buyers determine the price to offer, especially when Taiwan's real estate market usually lacks of accurate information or only partial information is available for buyers. It seems to be straightforward to include *t-1 price* in the prediction model.

Lastly, *New House-purchasing Loans* literally are subjected to interest rate and percentage of loans. Loans of low interest rate and high percentage would attract potential buyers to purchase; otherwise, it would lower the willing of purchasing. If we further investigate the amount of change, it can be found that in a way the amount of change represents the buyers' expectation to future housing market. In other words, a greater amount implies positive expectation in future housing market and stronger attention to enhance the price to deal; hence the housing price would attend to increase. In the other hand, a decrease of amount suggests negative opinions in the housing market; therefore the housing price might be trapped to correction or finishing.

4 Conclusions

The main purpose of this study is to predict the real estate price in Taiwan efficiently. Neural networks and Support Vector Regression are applied and compared. Variables are first selected from previous research and than chose by stepwise procedure and trial-and-error methods. It is found that SVR with trial-and-error method performed the best with MAPE = 4.466% and R^2 = 0.8540. That is, stepwise regression is efficient but not the best variable selection method with both BPNN and SVR. Possible reason is that the existence of nonlinearity in price prediction.

In addition, *Rediscount rate*, *Money supply*, and *t-1 price* are the three common variables for both BPNN and SVR. In our opinion, these variables are the most important variables in predicting real estate prices.

Observing from Table 2 and Table 3, no matter we apply stepwise or trial-and-error to select independent variables in housing price model using BPNN or SVR, *Rediscount rate* and *t-1 price* will be chosen. It means that, for Taiwan's housing price, *Rediscount rate* and *t-1 price* are the most impacting variables. The previous sold price is obviously the most straightforward reference for buyers and has drawn many attentions. Adjusting rediscount rate has been an essential policy tool for Taiwan Central Bank to claim future tendency of interest rate level. The level of interest rate reflects how much interest for the buyers to pay from house loans and how high the investors' opportunity costs are. Therefore, the rediscount rate plays an important role in impacting Taiwan housing price, it in a degree implies that Taiwan's housing market is of arbitrage or self occupied, but not a speculate market.

Furthermore, to have the best prediction using SVR, variables of *Gross Domestic Product*, *Rediscount rate*, *The Standard Unit Price of New Cases*, *Money supply M1a*, *New House-purchasing Loans*, *t-1 price* should be included in the model.

Table 1 Index variables

Real Estate aspects Climate indices	Investing	Manufacturing	Trading	Using
leading indices	*Gross Domestic Product*	*Movement of Loans for Construction*	*Consumer Price Index, CPI*	
	Money supply			
	Construction stock indices			
simultaneous indices	*Volume index of prime land*	*Building Permit of Residence*	*The Standard Unit Price of New Cases*	*Residential use rate*
	Benchmarking lending rate		*New House-purchasing Loans*	
lagging indices		*Permit for Occupancy of Residence (Floor Area,m2)*	*Registration of Translation of Building*	*House Rent Price Index*
		Construction industry average salary employees	*Land Value Increment Tax*	*Annual growth rate of households*

Table 2 BPNN result

Selected Variables		MAPE (%)	R^2
	t-1 price	5.867	0.6165
stepwise	*Rediscount rate, New House-purchasing Loans, t-1 price*	7.6836	0.5509
	Rediscount rate, Money supply M2, t-1 price	6.832	0.7222
trial-and-error	*Rediscount rate, CPI, The Standard Unit Price of New Cases, t-1 price*	**5.805***	0.7382

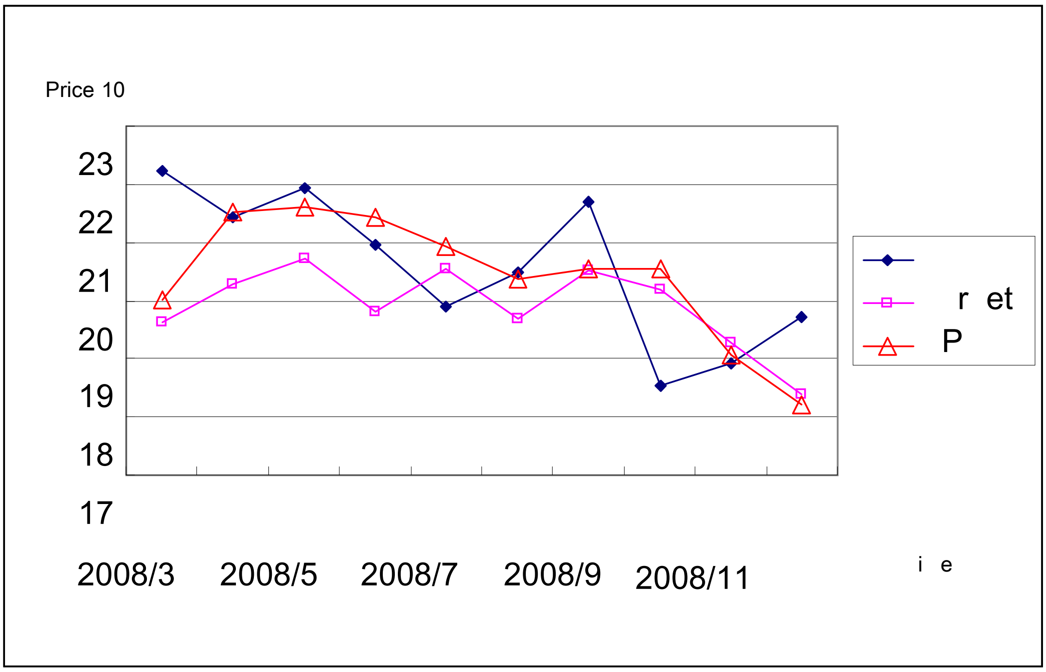

Figure 3. The predictions of BPN and SVR

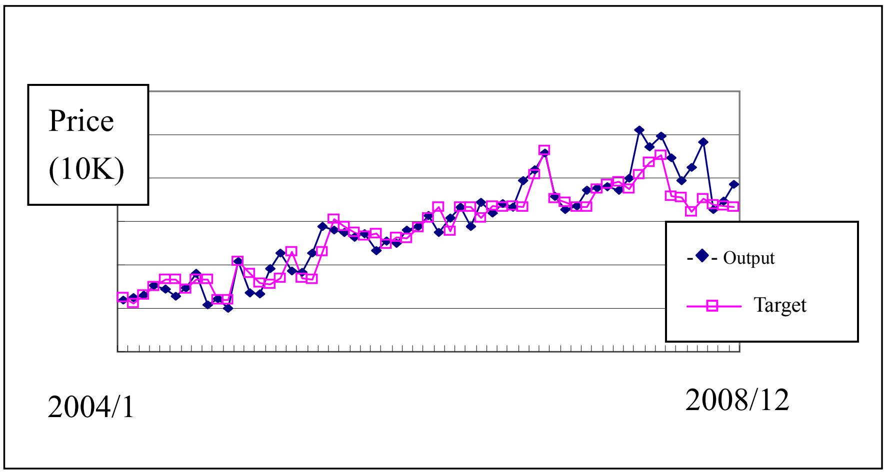

Figure 4. The prediction using only *t-1 price*

Table 3 SVR results

Variables		MAPE (%)	R^2
	t-1 price	6.152	0.5545
stepwise	*Rediscount rate, Money supply M2, t-1 price*	5.911	0.6448
Trial-and-error	*Gross Domestic Product, Rediscount rate, The Standard Unit Price of New Cases, Money supply M1a, New House-purchasing Loans, t-1 price*	**4.466***	0.8540

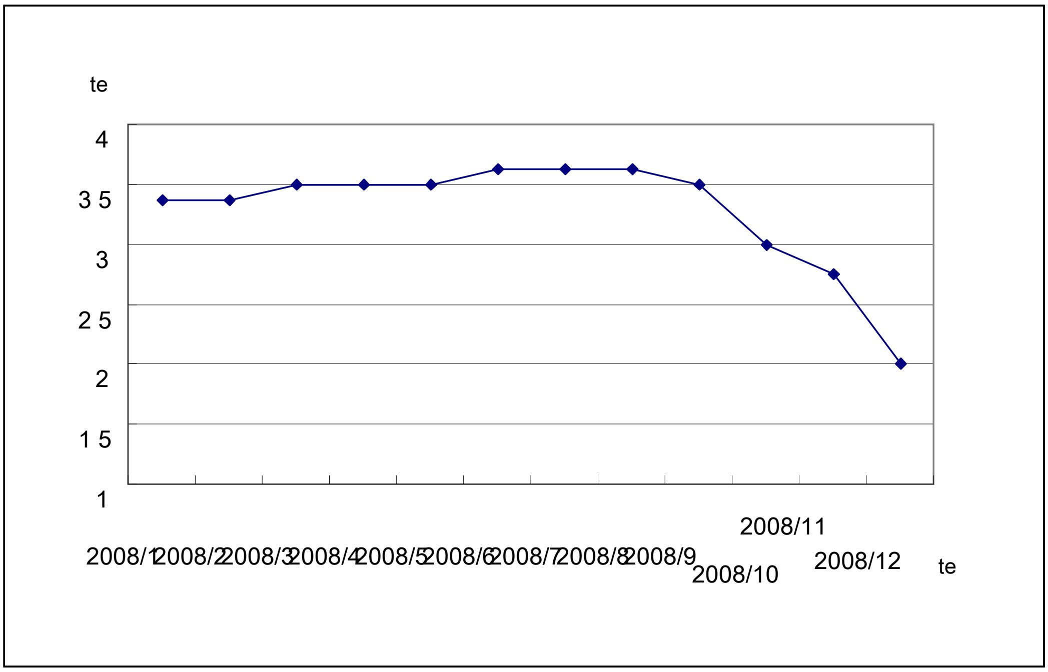

Figure 5. The trend of rediscount rate in 2008

Reference

1. Bennie, W. and Milam, A. (1999), "Pricing of Homeowner Association Dues with a Neural Network", *International Journal of Information and Management Sciences*, 10(1), pp.73-80.
2. Chen K.T. and Lee E. S. (2006), "Apply Fuzzy Adaptive Networks to credit rating", *WSEAS Transactions on Business and Economics*, Issue 8, Volume 3, August 2006, pp597-604
3. Chen K.T., Lin, H.-Y. and Huang, T.-C., "The Prediction of Taiwan 10-Year Government Bond Yield", *WSEAS Transactions on Systems*, Issue 9, Volume 8, September 2009, pp1051-1060
4. Do, Q. and Gary, G. (1992), "A Neural Network Approach to Residential Property Appraisal."

Real Estate Appraiser, 58(3), pp. 38-45.

5. Hagan, M.T. and Menhaj, M.B. (1994), "Training feedforward networks with the Marquardt algorithm." *IEEE Transactions on Neural Networks*, 5(6), pp. 989-993.
6. Holimchayachotikul, P., Leksakul, K., and Guizzi, G.(2011), "Robust Design for Etching Process Parameters of Hard Disk Drive Slider Fabrication Using Data Mining and Multi Response Optimization", *WSEAS TRANSACTIONS on SYSTEMS and CONTROL*, issue 1, volume 6, Jan. 2011, pp15-24.
7. Hopfield, J. J. (1982), "Neural networks and physical systems with emergent collective computational abilities", *Proc. Natl. Acad. Sci.* USA 79, pp2554-2558.
8. Hornik, K., Stinchcombe, M. and White, H, (1989), "Multilayer feedforward networks are universal approximators", *Neural Networks*, Volume 2, Issue 5, 1989, Pages 359-366
9. Hua, C.C. and Chang, C.O. (1997), "Housing Fluctuation Patterns Between Transaction Price & Volume", *Journal of housing studies*, 5, pp. 1-15 (In Chinese, with English Abstract)
10. Huang, W., Nakamori, Y., and Wang, S.Y. (2005), "Forecasting stock market movement direction with support vector machine." *Computers and Operations Research*, 32(10), pp. 2513-2522.
11. Kaashoek, J.F. and Dijk, V. K.H. (2002), "Neural Network Pruning Applied to Real Exchange Rate Analysis", *Journal of Forecasting,* 21(8), pp. 559-577.
12. Kuan, C.M. and White, H. (1994), "Artificial Neural Networks: An Econometric Perspective." *Econometric Reviews*, 13(1), pp. 1-91.
13. Lewis, C.D. (1982), "Industrial and Business Forecasting Methods: a practical guide to exponential smoothing and curve fitting", London: *Butterworth Scientific*.
14. Lin, C.T. and Lee, G. C.S. (1996), "Neural Fuzzy Systems: a neuro-fuzzy synergism to intelligent systems." New Jersey: *Prentice Hall*.
15. Makridakis, S. (1993), "Accuracy measures: Theoretical and practical concerns," *International Journal of Forecasting*, 9(4), pp. 527-529.
16. McCluskey, W.J., Dyson, K. and Mcfall, D. (1997), "The Mass Appraisal of Residential Property in Northern Ireland." *Computer Assisted Mass Appraisal: an international review 3*, ed. By McCluskey, W. J. and A.S. Adair, pp.59-77. Aldershot: Ashgate.
17. McGreal, S., Adair, A., McBurney, D. and Patterson, D. (1998), "Neural Networks: the Prediction of Residential Values." *Journal of Property Valuation and Investment*, 16(1), pp. 57-70.
18. Pai, P.F., and Lin, C.S. (2005), "A hybrid ARIMA and support vector machines model in stock price forecasting." *Omega*, 33(6), pp. 497-505.
19. Sfetsos, A. and Siriopoulos, C. (2004), "Combinatorial Time Series Forecasting Based on Clustering Algorithms and Neural Networks", *Neural Computing and Applications*, 13(1), pp. 56-64.
20. Tay, H. D.P. and Ho, H. D.K. (1992), "Artificial Intelligence and the Mass Appraisal of Residential Apartments." *Journal of Property Valuation and Investment*, 10(2), pp. 520-540.
21. Tay, H. F.E. and Cao, L. (2001), "Application of support vector machines in financial time series forecasting." *Omega*, 29(4), pp. 309-317.
22. Trafalis, T.B. and Ince, H. (2000), "Support Vector Machine for Regression and Applications to Financial Forecasting." *Proceedings of the IEEE-INNS-ENNS International Joint Conference on Neural*

Networks (IJCNN'00), 6, pp. 6348-6353.

23. Tsoukalas, L.H. and Uhrig, R.E. (1997), “Fuzzy and Neural Approaches in Engineering.” New York: *Wiley and Sons*.
24. Vapnik, V. N. (1995), “The nature of statistical learning theory.” New York: *Springer*.
25. Vapnik, V. N., Golowich, S., and Smola, A. (1997), “Support vector method for function approximation, regression estimation, and signal processing.” *Advances in Neural Information Processing Systems* 9, ed. by Mozer, M., Jordan, M. and Petsche, T., pp. 281–287, Cambridge, MA: *MIT Press*.
26. Wei, H.(2004). “Forecasting Foreign Exchange Rates with Artificial Neural Networks: A Review”, *International Journal of Information Technology & Decision Making*, Vol. 3, No. 1, pp. 145–165.
27. Wong, K., So, P. AT. and Hung, Y.C. (2001), “Neural Network vs. Hedonic Price Model: Appraisal of High-density Condominiums.” *Real Estate Valuation Theory*, ed. by Wang, Ko and Wolverton, M.L., pp 181-198, Boston, Mass.: *Kluwer Academic*.
28. Worzala, E., Lenk, M., and Silva, A. (1995), “An Exploration of Neural Networks and its Application to Real Estate Valuation.” *Journal of Real Estate Research*, 10(2), pp. 185-201.
29. Zhang, G. and Hu, M.Y. (1998), “Neural Network Forecasting of the British Pound /US Dollar Exchange Rate”, *Journal of Management Science*, 26(4), pp. 495–506