

FORECASTING HOUSE PRICE INDEX USING ARTIFICIAL NEURAL NETWORK

Mohamad Shukry b. Mohd Radzi¹, Chitrakala Muthuveerappan², Norhaya Kamarudin³, and Izran Sarrazin b. Mohammad⁴

^{1,2,3,4} Department of Property Management
Faculty of Geoinformation and Real Estate
University Technology of Malaysia
81310, Johor Bahru, Malaysia

Abstract

Property forecasting is an important component in the decision-making process for investors and governments in supporting asset allocation, formulating property funding strategies, and determining suitable policies. Multiple regression analysis has been accepted as the most commonly used technique for property forecasting. However, this technique has a limited ability to effectively deal with relations among variables, nonlinearity, and multicollinearity. A number of alternative techniques have been applied to improve the reliability of forecasting information, one of which is artificial neural network. This study applies this technique to forecast residential property prices in Malaysia. Unemployment rate, population size, mortgage rate, and household income were chosen as the independent variables while Housing Price Index (HPI) was chosen as the dependent variable. Quarterly time-series data from 2000 to 2009 were used for training and testing the ANN model. Subsequently, data for 2010 and 2011 were used to validate the model. Model validation has resulted in the mean absolute percentage error (MAPE) of 8% thereby suggesting a high level accuracy of ANN in its ability to learn, generalize, and converge time series data efficiently and to produce reliable forecasting information.

Keywords: house price index, artificial neural network (ANN), demand forecasting

1.0 INTRODUCTION

Property price volatility can exert a significant impact on the economy of a country (Ge and Runeson, 2004). For instance, an economic crisis in the United States of America occurred in 2008. Major investment and financial corporations in America were declared bankrupt due to significant amount of debt mostly triggered by the housing bubble since 2006. This created a chain of reaction causing the country's economy to collapse. Since then, countries including Malaysia are learning from the crisis by adjusting themselves from perilous events such as housing bubble.

Understanding price trends is important for investor, economists, practitioners and even private individuals. In Malaysia, Housing Price Index (HPI) is one of the common tools used to evaluate housing price trends. It also acts as an analytical tool for estimating changes in the rates of mortgage defaults, prepayments and housing affordability. Therefore, forecasting price index would enable users to make important and appropriate decision to reduce the risk of loss.

Property forecasting is an important component in property investment decision-making process for institutional investors, supporting asset allocation, property fund strategy and stock selection in a mixed-asset portfolio (Tsatsaronis and Zhu, 2004). Forecasting has some degree of uncertainty. However, a high degree of sophistication has been developed recently, with a range of advanced quantitative and qualitative procedure used by institutional investors in property forecasting, including judgemental procedures, causal or econometric procedures, and time series and trend analysis procedures.

2.0 MULTIPLE REGRESSION ANALYSIS VS ARTIFICIAL NEURAL NETWORK

Multiples regression analysis (MRA) has been the most commonly used technique in property forecasting due to its established methodology, long history of application, and wide acceptance among both practitioners and academicians (Kontrimas and Verikas, 2011).

However, the limitations of MRA model have also been identified (Quang and Gary, 1992). These limitations partly arise due to the model's inability to effectively deal with relations among variables, nonlinearity, and multicollinearity.

Several studies were carried out to discover better models for either valuing or forecasting property. Khalafallah (2008) applied artificial neural network (ANN) to evaluate the current market situation during the world economic crisis in 2008 and predicted the future performance of property in order to help investors and other market players in making important decisions. Although ANN has been limitedly used for valuation or forecasting property price, studies carried out to compare the accuracy of MRA and ANN discovered that the latter has superiority compared to the former.

Nguyen and Cripps (2001) compared MRA and ANN in predicting housing value. Ng and Skitmoreb (2008) used ANN and compared its accuracy with that of MRA in predicting of housing price. Khashei and Bijari (2011) also compared MRA and ANN in the mass appraisal context. Julio and Esperanza (2004) simulated a hypothesis in relation to valuing real estate value in Madrid. The majority of the studies computed the percentage error of both competing models using unknown output data. Generally, the results have shown that ANN has an average error of 5 to 10% whereas MRA has a higher average error of 10 to 15%.

2.1 Artificial Neural Network

In general, artificial neural network is a segmentation of human neural brain processing data paradigm in which the human biological system is based on. It is information-processing model designed to discover and track the relationship among various data sets autonomously (Yasmin, 2004). The basic element of a human brain is a specific type of cells and these cells are known as neurons. All natural neurons have four basic components, namely dendrites, soma, axon and synapses. A biological neuron receives data from other sources, combines them, performs a nonlinear operation on the result,

and generates information as a final result. Artificial neuron adopts the same concept of biological neuron, whereby it simulates the four basic functions from biological neuron. Figure 1.1 shows the basics of an artificial neuron.

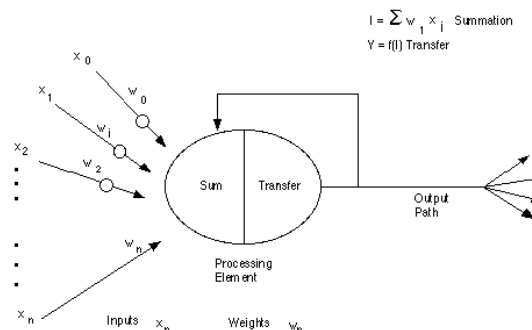


Figure 1.1: The basics of an artificial neuron

The raw data which are the input are presented as $x(n)$ and each of the data is multiplied by a connection weight which is represented by $w(n)$. The data are then summed, fed through a transfer function to generate a result, and then to produce an output. ANN normally has three layers although they can be more. The first layer is the input layer which consists of neurons received from the external environment. The output layer comprises neurons that communicate the output of the system to the external environment or user. When the input layer sends the input to the next layer, its neurons produce an output, which becomes input to other layers of the systems. The process continues until certain criteria are achieved (Yasmin, 2004). Figure 1.2 shows the structure of the system.

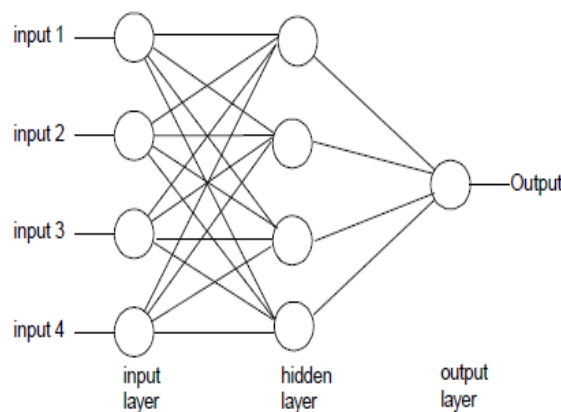


Figure 1.2: Artificial neural network simplified structure

To ensure a complete system, each neuron in the layer needs to communicate with other input in the next layer and to create a maximum relationship between the neurons before producing an output. The more is the connection between the neurons, the easier for them to be trained (Yusof, 2006). Training here means the process of presenting raw data into the system so that it can learn and memorize the knowledge among inputs through a learning role. In order to perform a network, there are several neural network model itemisations that should be known, namely learning algorithm, weights and biases, summation function, activation function, learning rate, and momentum rate (Yasmin, 2004).

2.1 Artificial Neural Network in Demand Forecasting

ANN gets so much attention among practitioners and academicians as a forecasting tool due to its ability of self-learning which allows it to analyse almost incomprehensible amount of data, test for the discovery relationship or connections among the data, and use the discovered data for predictions of future trends or event. NeuroShell 2 software is used in this study to simulate the network.

The selection of indicators influencing real estate housing price is made according to previous studies including Ng et al., (2004); Reichert (1990); Wilson (2002). Ng et al. (2003) explained that the behaviour of housing price in Hong Kong was affected by macroeconomic factors, housing-related elements, demographic attributes and supply-related variables. Eighteen variables were identified to represent all the four groups of factors, namely population, household income, land supply, number of property transactions, construction cost index, GDP, interest rate, unemployment rate, number of marriages, number of birth, rental index, GDP of construction, Hang Seng Index, number of private completion of units, private consumption expenditure, number of households, policy or political events or happenings, and population aged 20–60 years old who were the source of income groups. Reichert (1990) identified the important differences in the way new housing prices

reacted to local and national economic factors. The study disclosed that regional housing prices reacted consistently to certain national economic factors, such as mortgage rates. On the other hand, local factors such as population shifts, employment, and income trends often have a unique impact on housing prices. The study rejected the hypothesis of a single national housing market in favour of one that allows for broad national trends to be superimposed upon a unique regional market.

Wilson (2002) explained that the theoretical market models indicated that the main variables influencing house prices at both national and regional levels were income, real or nominal interest rates, the general level of prices, household wealth, demographic variables, tax structure, financial liberalisation, and housing stock. Kostas and Haibin (2004) explained that the demand and supply factors that drive real housing prices were those that have a longer-term influence and those that affect shorter-term dynamics. Factors that influence the demand for housing over longer prospect include growth in household disposable income, gradual shifts in demographics, and permanent features of the tax system and the average level of interest rates. Where else the availability and cost of land, and construction cost and investments on existing housing stocks were the longer-term determinants in housing supply.

3.0 RESEARCH AIM AND OBJECTIVES

This research aims at specifying a house price index forecasting model using ANN. To ensure the viability of the work, three objectives are set forth: a) to identify the appropriate variables that are necessary to represent the housing price; b) to collect identified variables' statistical data; and c) to develop house price index forecasting model by training the identified variables using ANN.

4.0 RESEARCH METHODOLOGY

In order to specify the ANN model, the network structure was determined first. This

included the training process, stop training criteria, testing and model validation. This research was focused on the independent and dependent variables that reflect the real property market structure. The dependent variable was house price index while the independent variables were interest rate, household income, employment rate, and population size.

Data for the dependent variable covered a period of 2000 to 2011 and were collected from various sources. The Malaysian house price index data were collected from the National Property Information Centre (NAPIC). Data on employment rate and population were collected from the Department of Statistic Malaysia based on the 2010 census. The Malaysian property interest rate data, based on the average rate taken from Malaysian main banking institutions, were collected from the Bank Negara. Household income data were collected from the Household Income Survey (HIS), conducted by the Department of Statistic Malaysia and later published by the Malaysia Economic Planning Unit under the Prime Minister Department Malaysia.

Using NeuroShell 2, the back propagation method was applied to train the model, and Test Set Extract Detail Module was used to extract the test set data from the training patterns. While developing the model, the model itself must be trained for defining the nodes and hidden layers. Following Rossini (1997) and Zhang and Patuwo (1997), the number of hidden neuron was determined by a trial-and-error process.

A very few hidden nodes for a given problem would have caused back-propagation not to converge to a solution and while many hidden nodes would have caused a much longer learning period. Too few or too many hidden neurons will not produce a satisfactory model. The study applied four (4) hidden layers while training the model.

While the ANN model did the training, there were two common criteria to stop training a network, namely the training cycles (epochs), and the desired errors. Ge and Runeson (2004) set the network training phase to 30,000 cycles with no other stopping criteria and suggested an optimum level of 20,000 to 40,000 cycles. Small and Wong (2001), Yasmin (2004) used both criteria to stop training the network. In this study, only the network training phase was selected where the training cycles were applied to 20,000 to 40,000 events.

The network weights were continually being adjusted until the output error congregated to an acceptable point. The network was then tested by applying the models to the test patterns. The errors between the calculated output and the actual output were calculated. To evaluate the accuracy of the network performance, the mean absolute error (MAE) and mean squared error (MSE) were applied as the network performance indicators together with the correlation coefficient.

5.0 EMPIRICAL RESULT AND VALIDATION

For the purpose of model validation, the input data for 2010-2011 were fed into the model to forecast the housing price index for that period.

Table 1.1: Summary of regression outputs

Training Data (%)	Testing Data (%)	R^2	r^2	Mean squared error	Mean absolute error	Correlation coefficient r
90	10	0.9667	0.9747	3.986	1.413	0.9873
80	20	0.9932	0.9932	0.819	0.718	0.9966
70	30	0.9923	0.9924	0.922	0.766	0.9962
60	40	0.9923	0.9924	0.921	0.774	0.9962

The regression outputs were then compared with the actual data to evaluate the performance of the competing models. The results show that the optimum data combination comprised 80% of training data and 20% of testing data with the highest $r = 0.9966$ and the lowest mean squared error of 0.819. The R^2 also shows that the selected training and testing data were able to reflect 99.32% of the actual value.

This validation was verified by calculating the mean absolute percentage error (MAPE) of the forecast data. Previous studies have ruled that an ANN model with a 10% MAPE is considered very good and that with a 20% MAPE is considered average.

The training model shows that all of the networks perform well with a value of correlation coefficient, r , between 0.9887 and 0.9966 for testing data and 0.9935 and 0.9975 for testing data. The highest correlation coefficient for the network was five hidden neurons which was 0.9966 with the mean square error of 0.818.

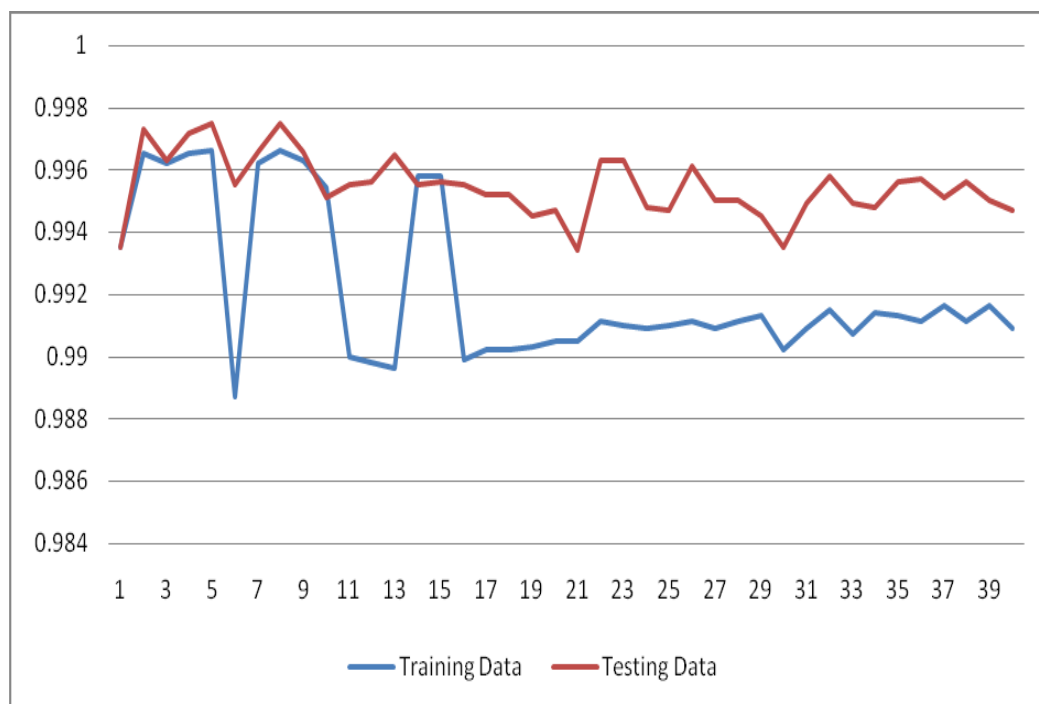
The analysis shows that the best neural network to forecast the housing price index in

Malaysia in this study was 4-5-1, i.e. 4 input layers, 5 neurons in the hidden layer and 1 neuron in the output layer.

6.0 CONCLUSION

This study employed recurrent back-propagation neural network to produce housing price prediction model in Malaysia. The analysis has shown that network was a good alternative method in lieu of the traditional multiple regression analysis (MRA) for predicting house price. It was able to map the complicated non-linear relationship between house price index that represented price movement and factors influencing house price. However, there were some limitations of ANN. One main problem faced was the limited amount of output data for verification. Since the National Property Information Centre (NAPIC) Malaysia only produces house price index on a 10-year basis, it was hard to gather index from the year before 2000. This study recommends that more dependent variables be included while training the forecasting model to produce highly accurate results.

Figure 1.3: Network performance of training and testing data with different number of neurons



REFERENCES

- Gallego, J. and Mora-Esperanza (2004). Artificial Intelligence Applied To Real Estate Valuation: An Example for the Appraisal of Madrid. *Catastro*: 255-265.
- Ge, X. J. and G. Runeson (2004). Modeling Property Prices Using Neural Network Model for Hong Kong. *International Real Estate Review*, 7(1): 121-138.
- Ian Wilson, S. Paris (2002). Residential Price Forecasting at National and Regional Levels. *FIG XXII International Congress*, Washington, D.C. USA.
- Kakar, M. K. and M. A. Kakar (2011). Factors Affecting the Real Estate Prices in Pakistan. *Asian Economic and Financial Review*, 1(2): 57-66.
- Khalafallah, A. (2008). Neural Network Based Model for Predicting Housing Market Performance. *Tsinghua Science and Technology*, 13(1): 325-328.
- Khashei, M. and M. Bijari (2010). An artificial neural network (p, d, q) model for timeseries forecasting. *Expert Systems with Applications*, 37: 479-489.
- Kontrimas, V. and A. Verikas (2011). The Mass Appraisal of the Real Estate by Computational Intelligence. *Applied Soft Computing*, 11: 443-448.
- Ng, S. T., M. Skitmoreb (2008). Using genetic algorithms and linear regression analysis for private housing demand forecast. *Building and Environment*, 43: 1171-1184.
- Nguyen, N. and A. Cripps (2001). Predicting Housing Value: A Comparison of Multiple Regression Analysis and Artificial Neural Network. Incomplete...
- Quang, D. A. and G. Gary (1992). A neural network approach to residential property appraisal. *Real Estate Appraiser*, 58(3): 38...incomplete.
- Reichert, A. K. (1990). The Impact of Interest Rates, Income, and Employment Upon Regional Housing Prices. *Journal of Real Estate Finance and Economics*, 3: 375-391.
- Rossini, P. (1997). Application of Artificial Neural Network to the Valuation of Residential Property. *Third Annual Pacific-Rim Real Estate Society Conference*. Palmerston North, New Zealand.
- Small, G. R. and R. Wong (2001). The Validity of Forecasting. *Pacific Rim Real Estate Society International Conference*. Christchurch, New Zealand.
- Tsatsaronis, K. and H. Zhu (2004). What Drives Housing Price Dynamics: Cross-country Evidence. incomplete...
- Yasmin, Z. and Muhd, Z. A.M. (2004). Computerized Forecasting Model Based on Artificial Neural Network for Low Income Group Housing in Urban Area, Universiti Teknologi Malaysia...incomplete...
- Yusof, N. (2006). *Time Series Modeling and Designing of Artificial Neural Network for Revenue Forecasting*. Faculty of Computer Science and Information System, Universiti Teknologi Malaysia. Master of Science (Computer Science).
- Zhang, G., and B. E. Patuwo. (1997). Forecasting with Artificial Neural Network. *International Journal of Forecasting*, 14: 35-62.