e-ISSN:1985-7527

FORECASTING LOW-COST HOUSING DEMAND IN PAHANG, MALAYSIA USING ARTIFICIAL NEURAL NETWORKS

Noor Yasmin Zainun[1], Ismail Abdul Rahman[1] and Mahroo Eftekhari[2]
[1]Universiti Tun Hussein Onn Malaysia, Johor, Malaysia
[2]Loughborough University, Leicestershire, UK

Abstract

Low cost housing is one of the government main agenda in fulfilling nation's housing need. Thus, it is very crucial to forecast the housing demand because of economic implication to national interest. Neural Networks (ANN) is one of the tools that can predict the demand. This paper presents a work on developing a model to forecast low-cost housing demand in Pahang, Malaysia using Artificial Neural Networks approach. The actual and forecasted data are compared and validate using Mean Absolute Percentage Error (MAPE). It was found that the best NN model to forecast low-cost housing in state of Pahang is 1-22-1 with 0.7 learning rate and 0.4 momentum rate. The MAPE value for the comparison between the actual and forecasted data is 2.63%. This model is helpful to the related agencies such as developer or any other relevant government agencies in making their development planning for low cost housing demand in Pahang.

Keyword: *Low-cost housing demand, AN,*

1. Introduction

Low-cost housing is a popular political agenda worldwide. Housing provision in Malaysia is a planning with cooperation between government and private sector. The main objective of the housing policy is to ensure that all Malaysian particularly the low income groups have access to adequate and affordable shelter and related facilities (Ministry of Housing and Local Government Malaysia, 1999). Tenth Malaysia Plan mentions that one of the Malaysia's longstanding development objectives is the provision of affordable housing for Malaysians in both rural and urban areas, with a focus on lower income groups (The Economic Planning Unit, 2010).

Each five years National Plan, government has focused on various housing programs in both rural and urban areas with the aim of providing affordable housing. Housing development is one of the important sectors that contribute in expanding national economy in our country. The housing categories are divided into four main categories: (1) low cost, (2) low medium cost, (3) medium cost, and (4) high cost housing.

Today, the housing issue for Malaysia is no longer about insufficient housing stock, but rather about ensuring that there are enough houses for various segments of society. The challenge is to match demand for housing with supply, based on location and affordability (The Economic Planning Unit, 2010). In 2009, a total of 13,529 individuals and families applied for public housing, while a survey of states showed that there were 97,260 squatter families who were yet to be relocated to permanent housing. Currently there are 11,800 available units, with an estimated 161,000 units more to be built during the Plan period (10^{th} Malaysia Plan, 2011-2015).

Due to the increment of the demand for low cost houses is very significant and vital; the selection of the best method on forecasting of demand is also becoming an important factor. All this while, the number of unit of low cost houses have been built by practice the requirement imposed by the government which is 30% of the total development. Obviously, by following this requirement, the numbers of low cost houses to be built do not reflect the actual demand of low cost housing. Henceforth, developing a model as an alternative way to forecast the number of units of low cost houses is therefore timely and imperative for a developing nation.

2. Artificial Neural Networks

The choice of the suitable forecasting techniques is vital to generate good forecasts. Empirical studies have shown that accuracy performance varies with different forecasting techniques, and therefore, accuracy plays an important part in selecting and testing a given forecasting technique.

The use of artificial neural networks to forecast and classify has increased significantly over the last several years (Khairulzan Yahya, 2002). Neural network models also have been used successfully in numerous areas such as forecasting bond ratings (Dutta and Shekhar, 1998); classifying solvent versus non-solvent life insurance firms (Lin *et. al*., 1994); modeling of property prices (T. Kauko,2002); house price prediction (V. Limsombuncai, C. Gan and M. Lee, 2004); forecasting demand in consumer durables (P. D. Mc Nelis and J. Nickelsburg, 200); and modeling end-use energy consumption in the residential sector (M. A. Koksal, V. Ismet Ugursal, 2006).

According to Ian Flood (2008), within the civil engineering discipline, artificial neural networks appear from publications statistics to be one of the great successes of computing. In the ASCE Journal of Computing, over 12% papers published from 1995 to 2005 which is 54 out of 445 papers have used the term “neural” as part of their title. Furthermore, the distribution of these publications by year indicates that there has been no decline in interest over this period. The citations indicates similarly confirm the popularity of artificial neural networks. According to ISI Web of knowledge, three of the top five most frequently cited articles from all issues of the ASCE Journal of Computing are on artificial neural networks, including the first and second placed articles in this ranking. Table 3.1 shows the five most frequently cited articles in the Journal of Computing in Civil Engineering.

***Table 1*:** *The five most frequently cited articles in the Journal of Computing in Civil Engineering. Source: ISI Web of Knowledge, Web of science, 2006*

Article title	Number of citations
Neural networks in civil engineering	131
Neural networks for river flow prediction	97
Genetic algorithms in pipeline optimization	74
Genetic algorithms in discrete optimization of steel truss roofs	37
Damage detection in structures based on feature-sensitive neural networks	35

Artificial neural networks also have been successfully applied to forecast in construction industry such as; forecasting demand in private sector construction in United Kingdom with the minimum value of under estimate is only

3.37% and the highest value of under estimate is 11.56% compare to the actual value (Yang and Packer, 1997), Singapore resident construction demand forecasting with the value of mean absolute percentage error (MAPE) only 0.93 compare to the actual value (Goh B. H., 1998), and private residential construction forecasting in United States where the value of MAPE is 7.6 compare to the actual value (Aiken *et. al*,1998).

Several studies have demonstrated how neural networks may be far more and accurate than competing techniques including multi-linear regression and discriminant analysis. Neural networks have outperformed regression in predicting bank failures (Salchenberger *et. al*, 1992), stock market returns (Kimoto *et. al*, 1990), property values (Do and Grudnitaki, 1992; Tay and Ho, 1992), United States Treasury Bill rates (Aiken, 1995), and other problems (Aiken *et. al*, 1995). “Research is, of course, a very long way from being able to replicate human cognitive skills using artificial neural networks, but the decision to take the next tentative step towards this goal is long overdue.” (Ian Flood, 2008).

3. ANN Strengths and Weaknesses

ANNs are particularly good at filtering out noise or unnecessary information, isolating it and using only relevant information. A well-trained ANN can work with noisy or incomplete inputs and still produce correct output by making use of context and generalizing or filling gaps, which is why it has been called artificial intelligence.

In addition, ANNs are very good at solving vector mapping problems that are nonlinear in form and comprise a fixed set of independent variables. Therefore, they frequently provide more accurate solutions than the alternative modeling techniques, and do not require the user to have a good understanding of the basic shape of the function being modeled (Ian Flood, 2008).

A common criticism of ANNs is that they function like a ‘black box’ that works in rather mysterious ways. The main computations are performed in the hidden layers, and the user has no way of tracing how the ANN processed the information or reached its conclusion. Thus, there can be a certain amount of discomfort in using and relying on ANNs.

A second problem is that the training process does require substantial time and effort. The better and more thorough the training, the better the resulting ANN should perform (Review of Business, 1997). The additional complexity in these devices is there simply to provide greater precision in results, not greater functionality (Ian Flood, 2008).

Obviously, using inadequate data and training prior to actual use will not permit the ANN to fully explore and correlate the relationships among the data, resulting in at least initial loss of accuracy. However, feeding the network with too much data can actually cause the network to learn too much of the details of the data, as opposed to learning the general pattern of how the data interrelates. This is in addition to the wasted time and expense of over training. Networks designed to deal with more complex data and predictions than others require more training (Christopher, 1997).

4. Methodology

The methodologies of this study are including finding out the significant indicator using Principal Component Analysis (PCA) adapted from SPSS and a Neural Network (NN) model development adopted from NeuroShell2. PCA is used to derive new indicators; that is the significant indicators from the nine selected indicators. The indicators are:

(1) population growth; (2) birth rate; (3) mortality baby rate; (4) inflation rate; (5) income rate; (6) housing stock; (7) GDP rate; (8) unemployment rate; and (9) poverty rate. The dependent indicator is the monthly time series data on low cost housing demand starting from January 2001 to January 2007.

In NN model development, a series of trial and error process are done to find the suitable number of neurons in hidden layer, learning rate, momentum rate and screening the result using the best NN model.

5. Significant indicators

To perform neural network model, the important or significant of the independent indicators should be determined to avoid longer generalization. In this study, Principal Components Analysis (PCA) is used to decrease the 'curse in dimensionality' in NN. The determinant of the correlation matrix, R is 3.58x 10^{-15} that is very close to zero. It shows that linear dependencies are existed among the response indicators. Therefore, the PCA method can be performed. By testing from the hypothesis, populations of the correlation matrix are equal to identity matrix, which considered all the data are multivariate normal while the indicators are uncorrelated.

The value for the test statistic for these data is 2267.46 and the critical point of the chi-square distribution with $p(p-1) / 2 = 36$ for the degree of freedom, $\alpha = 0.001$, the critical point is 71.64. Clearly it shows that the hypothesis is rejected at the 0.001 significant levels because of 2267.46 > 71.64. From the scree plot (refer **Figure 1**), eigenvalues for principal component (PC) three to nine are close to zero. Since eigenvalue for PC three is less than 1, total variation for two PCs are 88.633%, and others eigenvalues are close enough to zero that they can be ignored; therefore two PCs are used for the analysis. Therefore, Pahang has two principal components, which are two significant indicators that will be used as the input to develop the neural network model.

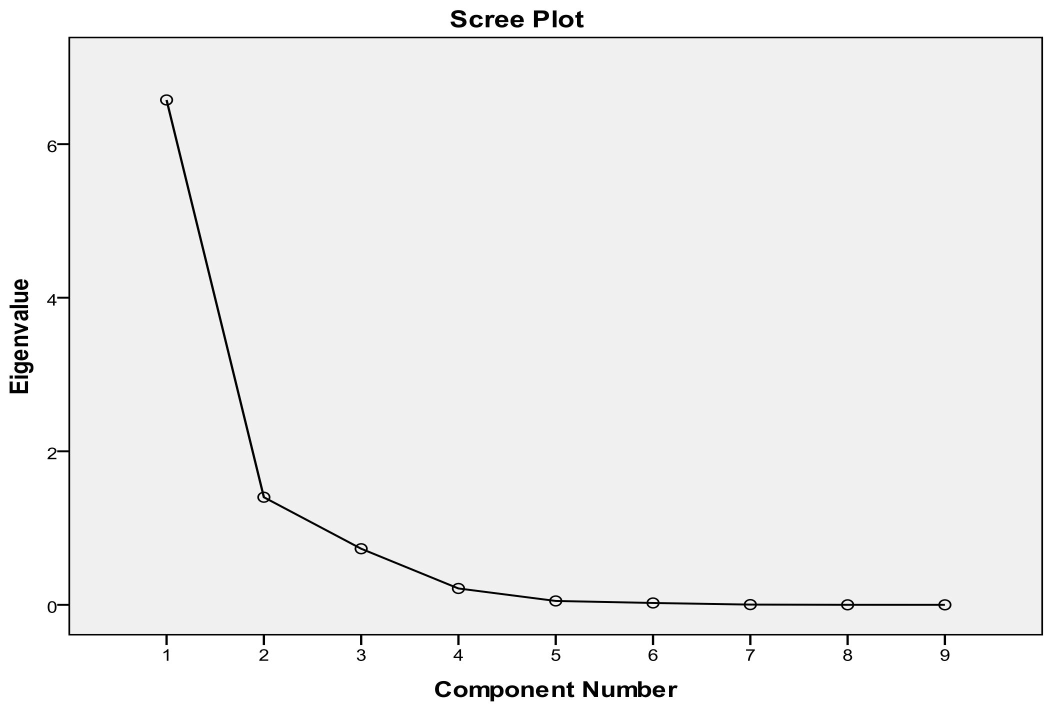


***Figure 1**: Scree plot for Pahang*

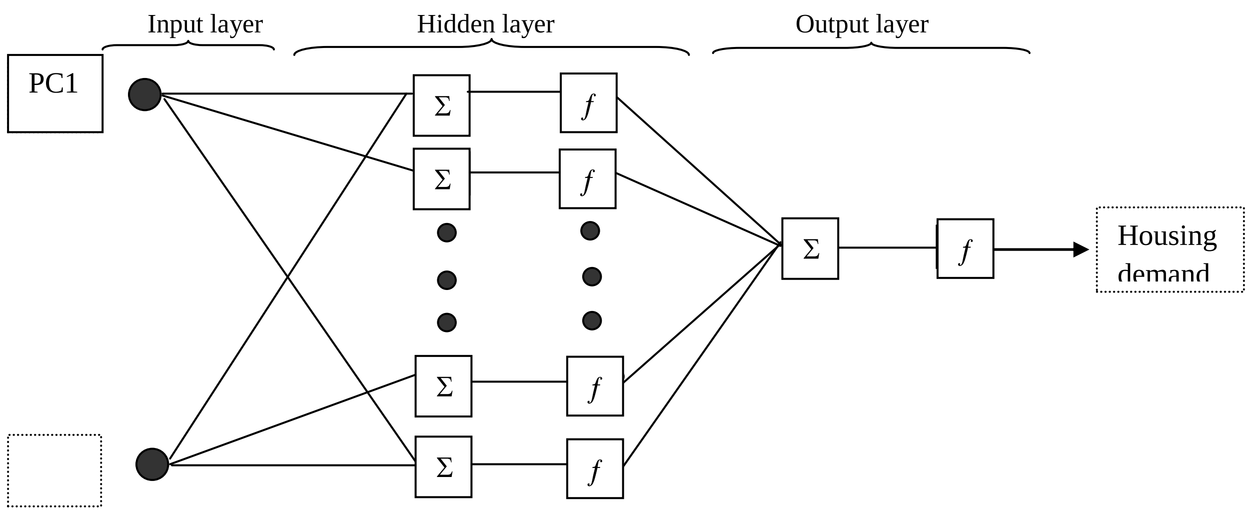
***Table 2**: Component score coefficient matrix for Pahang*

Indicator	PC1	PC2
Inflation rate	0.114	0.037
GDP	0.167	0.026
Poverty rate	0.013	-0.230
Income rate	0.193	-0.007
Unemployment rate	0.264	-0.493
Housing stock	0.121	0.085
Birth rate	-0.187	0.003
Population growth	0.400	-0.357
Mortality baby rate	-0.043	0.272

According to Johnson (1998), the number of component is to be equal to the number of eigenvalue of R, which is 1. Therefore, the significant indicators for each component are with the value of component score coefficient matrix nearest to 1. The other indicators are still considered but they give less effect compared to the significant indicators. **Table 2** shows that mortality baby rate and population growth gives the highest impact to the low cost housing demand in Pahang. Finally, PC1 and PC2 will be used as the input in NN model development.

6. Model development

In this study, a simple network that is a feed forward structure with one hidden layer is used since the structure has proved most useful in solving real problems. The NN is then trained using back propagation which is the best known example of supervised learning algorithm. This algorithm will uses the data to adjust the network's weight and thresholds so as to minimize the error in the predictions on the training set. The data are divided into three sets, which are training, testing and validation set. From 73 data, 60 data is used as the training data, 20 data is used as the testing data and the last 3 data is used to validate. Validation is done by comparing the actual and forecasted demand in November 20006, December 2006 and January 2007. Learning rate and momentum rate is determined by means of trial-and-error following four phases as shown in **Table 2**. This method also has been used by Sobri Harun (2001) and Khairulzan (2002). The average error used is 0.001 and 40,000 learning epochs. The number for the input node for state of Pahang is two since it has two PC as the input. The number of the output neuron for this task is one which is the housing demand. **Figure 2** below shows the Neural Network topology with one inputs and one output.



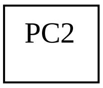


***Figure 2**: Neural Network topology with 2 input for state of Pahang*

Using the training and testing data, a series of trial and error process is conducted by varying the number of hidden neurons in order to find the suitable number of hidden neurons. The process started by applying the smallest number of hidden neurons. In this study, the hidden neuron varies from 1 to 40. Training and testing are conducted by increasing hidden neurons after each training and testing process. The network will minimize the difference between the given output and the prediction output monitored by the minimum average error while the training process is conducted. When the value is reducing, the error also will be minimizing. This process continues until 40,000 cycles of test sets were presented after the minimum average error or the minimum average reaches the convergence rate, which comes first.

***Table 3**: Networks performance with different phases in hidden layer*

Phase	**1**	**2**	**3**	**4**
Learning rate	0.9	0.7	0.5	0.4
Momentum rate	0.1	0.4	0.5	0.6
The best R^2	0.71	0.77	0.70	0.72

Table 3 shows the networks performance with different number of phases and neurons in hidden layer. The highest R^2 is 0.72 at 15 neurons in hidden layer. Therefore, the best Neural Network model to forecast low cost housing demand in Pahang is 2-15-1, which is 2 numbers of neurons in input layer, 15 numbers of neurons in hidden layer and 1 number of neurons in output layer with 0.7 learning rate and 0.4 momentum rate.

***Table 4**: Actual and forecasted demand on low-cost housing in Pahang*

Time series	**Actual data**	**Forecasted data**	**PE (%)**
November 2006	76	99	30.84
December 2006	114	114	0.25
January 2007	160	144	9.71
		MAPE	6.96

The ability of forecasting is very good if MAPE value is less than 10% while MAPE for less than 20% is good. Therefore, evaluation using MAPE value shows that Neural Network capable to forecast low-cost housing demand in Pahang very good (refer **Table 4**).

7. Conclusion

The best NN model to forecast low-cost housing in Pahang is 2-15-1 with 0.7 learning rate and 0.4 momentum rate. Comparison between the actual and forecasted data shows that Artificial Neural Network can forecast low-cost housing demand in Pahang very good with MAPE value of 6.96%. This model has high potential to produce a very good time series forecasting. By developing this model, it is hoped that this model will be improved and be used to forecast low cost housing demand in the future.

8. Acknowledgement

The authors would like to thank the Ministry of Higher Education, Malaysia for supporting this research under the Fundamental Research Grant Scheme (FRGS).

9. References

D. E. Rumelhart, G. E. Hinton, and R. J. Williams (1986). "Learning internal representations by error propagation". In D. E. Rumelhart and J. L. McClelland, editors, Parallel Distributed Processing, volume 1. MIT Press, Cambridge, MA.

Dutta, S. and Shekhar, S. (1998). "Bond Rating: A Non-Conservative Application of Neural Networks." *Proceedings of the IEEE International Conference on Neural Networks*, p. II:443-450.

Flood, I. and Kartam, N. (2008). "Artificial Neural Networks for Civil Engineers." *International Journal of Systems Science*, v.39 n 10, p. 925-946.

Goh, Bee Hua (1998). "Forecasting residential construction demand in Singapore: a comparative study of the accuracy of time series, regression and artificial neural network technique", *Engineering Construction and Architectural Management,* 5, no.3 p.261-275.

Government of Malaysia (2001). The Eight Malaysia Plan, 2001-2005. Kuala Lumpur: Percetakan Nasional Malaysia Berhad.

Government of Malaysia (2006). The Ninth Malaysia Plan, 2006-2010. Kuala Lumpur: Percetakan Nasional Malaysia Berhad.

Government of Malaysia (2010). The Tenth Malaysia Plan, 2011-2015. Kuala Lumpur: Percetakan Nasional Malaysia Berhad.

Haykin, S. (1994). "Neural Networks: A Comprehensive Foundation". New York: Macmillan Publishing.

Johnson, D.E (1998). "Applied Multivariate Methods for Data Analyst". United States of America: Brooks/Cole Publishing Company.

Khairulzan Yahya and Muhd Zaimi Abd. Majid (2002). "Forecasting of Low Cost Housing Demand In Urban Areas- Artificial Neural Network And ARIMA Model Approach." 2nd International Conference on Systems Thinking in Management, UK, p. 14-19.

Khairulzan Yahya and Muhd Zaimi Abd. Majid (2002). "Comparative Study on Forecasting Demand on Low-cost House in Urban Areas Using Artificial Neural Networks and ARIMA Model." First International Conference on Construction in the 21st Century (CITC2002), USA, p. 687-694.

Ministry of Housing and Local Government Malaysia (1999). "Housing in the New Millennium – Malaysian Perspective." http:www.kpkt.gov.my/jpn/artikel3.htm.

Sobri Harun, Nor Irwan Ahmat Nor, and Amir Hashim Mohd. Kassim (2001). "Rainfall-Runoff Modeling Using Artificial Neural Network". *Malaysian Journal of Civil Engineering*, 13 (1). pp. 37-50.

Sobri Harun (1999) "Forecasting and Simulation of Net Inflows for Reservoir Operation and Management", PhD thesis., Universiti Teknologi Malaysia, Skudai, Malaysia.