International Journal of Sustainable Construction Engineering & Technology
Vol 2, Issue 1, June 2011

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] Faculty of Civil and Environmental Engineering, Universiti Tun Hussein Onn Malaysia, 86400 Parit Raja - Batu Pahat, Johor, Malaysia
[2] Loughborough University, Leicestershire, UK

* Corresponding author, e-mail: nryasmin@uthm.edu.my

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

Keywords: *Low-cost housing demand, ANN*

**Corresponding Author*

1.0 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 [1]. The main objective of the housing programs during the Eighth Malaysia Plan period was in line with the housing policy of providing adequate, quality and affordable houses to all Malaysians [2]. During the Ninth Plan period, the development of the housing sector continues to focus on the provision of adequate, affordable and quality houses for all Malaysians [3].

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. The housing development is played the main sectors in expanding national economy in our country. The housing category is divided into four main categories: (1) low cost, (2) low medium cost, (3) medium cost, and (4) high cost housing. By develop low cost and low medium cost housing it can reduce housing growth illegally on government’s land and also prevent the public creating other new squatters.

Public and private sector housing targets and achievements from year 2001 until 2005 show that the total number of medium and high cost houses constructed by the private sector far exceeded the Plan target. A total number of 222,023 units medium cost and 274,973 units of high cost houses were constructed. On the other hand, in the low medium cost housing category, a total of 83,910 units were completed, achieving 63.9% of the Plan target. Of this total, the private sector constructed 61,084 units or 72.8%. The overall performance in this category was better than the 20.7% of the target achieved during the previous Plan period. It shows the imbalance of housing construction in Malaysia which leads to high demand pressure for low cost houses.

During the Ninth Malaysia Plan, requirement for new houses is expected to be about 709,400 units of which 6.3% will be in Pahang. According to the housing requirements by state from 2006 to 2010 [3], of the total requirement, 92.8% will be for new houses while 7.2% for replacement. The private sector is expected to supply 72.1% of the total requirement. 38.2% will be a combination of low and low medium cost houses as well as houses for the poor while 61.8% of medium and high cost houses.

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.

2.0 OBJECTIVE

The Aim of this paper is to develop a model to forecast low-cost housing demand in Pahang, Malaysia Using Artificial Neural Networks. The actual and forecasted data will be compared and validate using Mean Absolute Percentage Error (MAPE).

3.0 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 December 2006.

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.

4.0 SIGNIFICANT INDICATORS

To perform neural network model, the important or significant of the independent indicators should be determined to avoid longer generalization [4]. 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 exist 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 1304.06 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 1304.06 > 71.64. From the scree plot (refer Figure 1), eigenvalue for the principal component (PC) two to nine are close to zero which they can be ignored. Therefore, only one PC is used as the input in the NN with total variation of 95.0%.

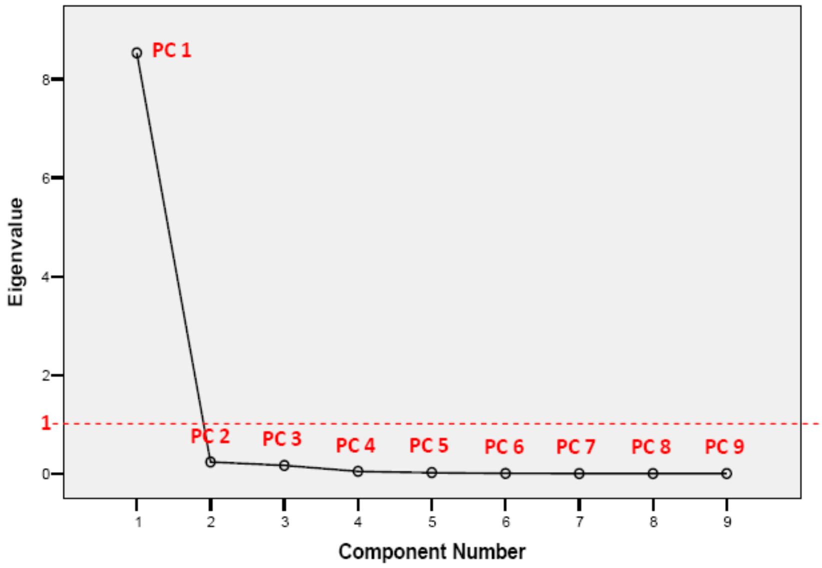

Figure 1: Scree plot for Pahang

Table 1: Component score coefficient matrix for Pahang

Indicator	PC1
Population growth	-0.995
Birth rate	-0.997
Mortality baby rate	0.968
Inflation rate	0.890
Income rate	0.995
Housing stock	0.988
GDP rate	0.954
Unemployment rate	-0.972
Poverty rate	-0.995

According to [5], the number of component is to be equal to the number of eigen value 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 1 show that the most significant indicator for PC1 is income rate.

5.0 MODEL DEVELOPMENT

In this study, a simple network that is a feed forward structure with one hidden layer is used since the structure have proved most useful in solving real problems [6]. The NN is then trained using back propagation which is the best known example of supervised learning algorithm [7]. 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 72 data, 50 data is used as the training data that is from January 2001 to February 2005. Another 20 data which is from March 2005 to October 2006 is used as the testing data and the last 2 data is used to validate. Validation is done by comparing the actual and forecasted demand for November 20006 and December 2006. 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 [8-9]. The average error used is 0.001 and 40,000 learning epochs. The number for the input node for state of Pahang is one since it has only one PC as the input. The number of the output neuron for this task is one which is the housing demand. Figure 2 shows the Neural Network topology with one inputs and one output.

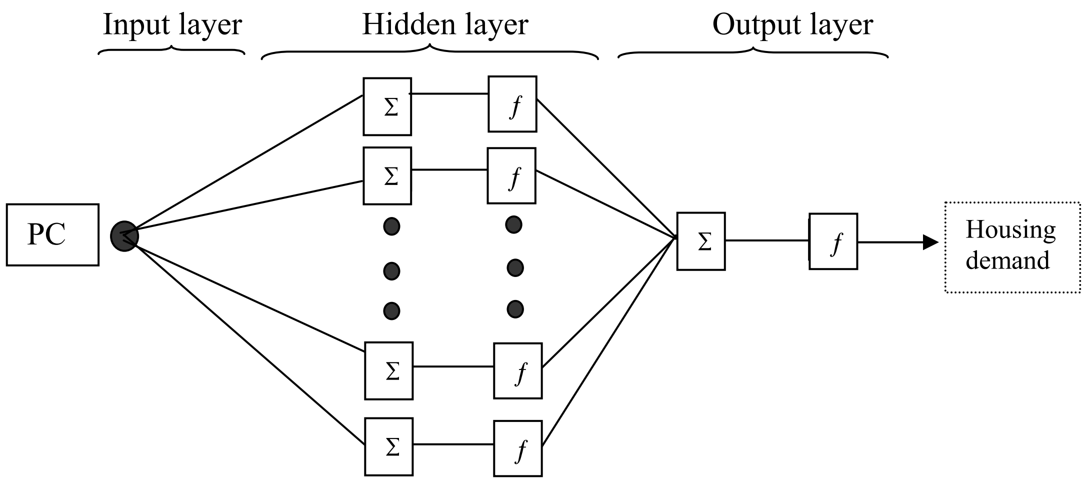

Figure 2: Neural Network topology with 1 input for state of Pahang

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 2 shows the networks performance with different number of phases and neurons in hidden layer. The highest R^2 is 0.48 at Phase 2 with 22 neurons. Thus, the best Neural Network for Pahang to forecast low cost housing demand is 1-22-1, which is 1 number of neurons in input layer, 22 numbers of neurons in hidden layer and 1 number of neuron in output layer with 0.7 learning rate and 0.4 momentum rate.

Table 2: Networks performance with different phases and neurons 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	0.46	0.48	0.46	0.46
Neurons	9	22	9	6

The ability of forecasting is very good if MAPE value is less than 10% while MAPE for less than 20% is good [10]. Therefore, evaluation using MAPE value shows that Neural Network capable to forecast low-cost housing demand in Pahang quite good for November and December 2006 (refer Table 3).

Table 3: Actual and forecasted demand on low-cost housing in Pahang for November and December 2006

Time series	**Actual data**	**Forecasted data**	**PE (%)**
November 2006	114	117	2.63
December 2006	213	140	34.27
		MAPE (%)	18.45

6.0 CONCLUSION

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. Comparison between the actual and forecasted data shows that NN can forecast low-cost housing demand in Pahang very good on November 2006 with MAPE value of 2.63%. By developing this model, it is hoped that there will be no more under or over construction of low-cost houses in Malaysia. It is also hoped that it can be 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 urban area in Malaysia towards the future as there is no model have been created yet. Furthermore, a lot of advantages if a better planning of low cost housing construction is done such as save in budget, time, manpower and also paper less.

7.0 REFERENCES

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

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

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

[4] 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.

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

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

[7] 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.

[8] 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.

[9] 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.

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