



Forecasting Low-Cost Housing Demand in Johor Bahru, Malaysia Using Artificial Neural Networks (ANN)

Noor Yasmin Bt Zainun

Faculty of Civil and Environmental Engineering

Universiti Tun Hussein Onn Malaysia

86400 Parit Raja, Batu Pahat, Johor, Malaysia

Tel: 60-1-2779-1789 E-mail: nryasmin@uthm.edu.my

Ismail Abdul Rahman

Faculty of Civil and Environmental Engineering

Universiti Tun Hussein Onn Malaysia

86400 Parit Raja, Batu Pahat, Johor, Malaysia

Tel: 60-1-2787-6360 E-mail: ismailar@uthm.edu.my

Mahroo Eftekhari

Faculty of Civil and Building Engineering

Loughborough University, Leicestershire LE11 3TU, UK

Tel: 440-751-055-5056 E-mail: M.M.Eftekhari@lboro.ac.uk

Abstract

There is a need to fully appreciate the legacy of Malaysia urbanization on affordable housing since the proportions of urban population to total population in Malaysia are expected to increase up to 70% in year 2020. This study focused in Johor Bahru, Malaysia one of the highest urbanized state in the country. Monthly time-series data have been used to forecast the demand on low-cost housing using Artificial Neural Networks approach. The dependent indicator is the low-cost housing demand and nine independents indicators including; population growth; birth rate; mortality baby rate; inflation rate; income rate; housing stock; GDP rate; unemployment rate and poverty rate. Principal Component Analysis has been adopted to analyze the data using SPSS package. The results show that the best Neural Network is 2-22-1 with 0.5 learning rate and momentum rate respectively. Validation between actual and forecasted data show only 16.44% of MAPE value. Therefore Neural Network is capable to forecast low-cost housing demand in Johor Bahru, Malaysia.

Keywords: Low-cost housing, Artificial neural networks, Principal component analysis

1. Introduction

Accurate predictions of the level of aggregate demand for construction are of vital importance to all sectors of this industry such as developers, builders and consultants. Empirical studies have shown that accuracy performance varies according to the types of forecasting technique and the variables to be forecast. Hence, there is a need to identify different techniques, in terms of accuracy, in the prediction of needs for facilities (Goh B. H., 1998).

Under the Seventh Malaysia Plan (1996-2000) and Eight Malaysia Plan (2001-2005), Malaysian government is committed to provide adequate, affordable and quality housing for all Malaysia, particularly the low income group. This in line with Istanbul Declaration on Human Settlement and Habit Agenda (1996) to ensure adequate shelter for all (Syafiee Shuid, 2004). The total number of housing units targeted was 800,000 units under Seventh Malaysia Plan and 782,300 units of housing is targeted to be construct under Eighth Malaysia Plan (Chapter 18, Eight Malaysia Plan, 2006). During the Ninth Malaysia Plan, requirement for new houses is expected to be about 709,400 units of which 19.2% will be in Selangor followed by Johor at 12.9%, Sarawak 9.4% and Perak 8.2%(Ninth Malaysia Plan, 2006).

Unfortunately, in 2004 there are 100,000 of low-cost houses in Selangor, Malaysia overhang (The Sun, 2004). The over construction of the low-cost at Selangor had cause million of lost while at the same time the money can be use to provide low-cost houses in other states in Malaysia. Based on Draft Kuala Lumpur Structure Plan 2020, Kuala Lumpur still lacks of 20,595 units of houses. In spite of the pricing of low-cost houses may be too high, one of the other reasons why the houses remain unsold is because they were built in undesirable locations (Salleh Buang, News Straits Time, 2004).

Therefore, there is a vital need to have a model to forecast low-cost housing demand in Malaysia so that there will be no more under or over construction of low-cost houses. At the same time, budget, time and manpower can be saved.

2. Independent and dependent indicators

The methodologies of this study are including finding out the significant indicators using Principal Component Analysis (PCA) adapted from SPSS 10.0, series of trial and error process to find out the suitable number of hidden neurons, learning rate, and momentum rate for the network and screening the result using the best Neural Network (NN) model.

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 2000 to December 2003.

3. Significant indicators

The determinant of the correlation matrix, R is 2.84×10^{-14} 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. For this case, there are nine indicators within 36 data therefore, $p=9$ and $N=36$.

$$\begin{aligned} -a \cdot \ln(v) &= -(N - 1 - (2p + 5)/6) \ln(R) \\ &= -(36 - 1 - (2 \times 9 + 5)/6) \ln(2.84 \times 10^{-14}) \\ &= 972.16 \end{aligned}$$

Therefore, the value for the test statistic for these data is 972.16 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 $972.16 > 71.64$. From the scree plot (refer Figure 1), eigenvalue for the principal component (PC) three to nine are close to zero which they can be ignored. Since the eigenvalue for PC one to two are greater than one, total variation for the two PCs is 98.0%. Therefore, two PCs are used for the analysis. 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 1 show that the most significant indicators for PC1 are income rate and PC2 is population growth.

4. Model development

According to Cattani (1994), a network is required to perform two tasks; (1) reproduce the patterns it was trained on and (2) predict the output given patterns it has not seen before, which involves interpolation and extrapolation. In order to perform these tasks, a backpropagation network with one hidden layer is used. To find out the best number of hidden neurons for the network, the default setting of backpropagation algorithm in Neuroshell2 is applied. In this study, the learning rate and momentum rate is determined by means of trial-and-error, following four phases as shown in Table 2. These rates have been stated by SPSS Inc (1995) according to experiences in various fields using neural network. This method also has been used by Sobri Harun (1999) and Khairulzan (2002). The learning process is divided into four phases and in each phase, the learning and momentum rate will be change. The average error used is 0.001 and 40,000 learning epochs. The number for the input node for Johor Bahru district is two since it have two PCs 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 2 inputs and one output. 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.

Figure 3 shows the performance of testing when different number of neurons is applied in hidden layer with different phases. From the figure, value of r shows that the performance of training and testing network almost similar with each other using different learning and momentum rate. It also shows that the network performance are good where all the values of r are uniform between 0.49 to 0.55 except using 3, 5, 22, 30, 33 and 39 numbers of neurons in all phases and 22

neurons in phase 2. Phase 3 and 4 show that the lowest network performance is when using 3 numbers of hidden neuron while the highest network performance is when using 22 numbers of neurons. Thus, the best Neural Network for Johor Bahru district to forecast low cost housing demand is 2-22-1, which is 2 numbers of neurons in input layer, 22 numbers of neurons in hidden layer and 1 number of neuron in output layer.

Evaluation using Mean Absolute Percentage Error (MAPE) shows that MAPE value using 0.5 learning rate and 0.5 momentum rate (Phase 3) has the best performance with 13.71% rather than using 0.4 learning rate and 0.6 momentum rate with 16.44% (refer Table 3 and Figure 4). The ability of forecasting is very good if MAPE value is less than 10% while MAPE for less than 20% is good (Sobri Harun, 1999). Therefore, the best Neural Network for Johor Bahru district to forecast demand on low cost housing is 2-22-1 with using 0.5 learning rate and 0.5 momentum rate.

5. Discussion

Out of nine indicators, PCA has derived two PCs with significant indicator for PC1 is income rate and PC2 is population growth. The best NN model to forecast low cost housing demand in Johor Bahru is 2-22-1 using 0.5 learning and momentum rate respectively. Comparison between the actual and forecasted data shows that NN capable to forecast low cost housing demand in Johor Bahru with the best value of MAPE is 13.71%.

6. Conclusions

In conclusion, NN is capable to forecast low cost housing demand in Johor Bahru, Malaysia. Currently, low cost housing which offered is not enough and cannot afford the increasing demand. Therefore, by developing this model, it is hoped 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.

References

- Abdul Ghani Salleh. (2002). *Provision of Affordable Housing in Malaysia: The Role Public and Private Sector*. Universiti Sains Malaysia.
- Callan R. (1999). *The Essence of Neural Networks*. Maylands Avenue: Prentice Hall Eroupe.
- Fausett L. (1994). *Fundamentals of Neural Networks: Architecture, Algorithm and Applications*. Englewood Cliffs, N. J.: Prentice-Hall.
- 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. (1996). *The Seventh Malaysia Plan, 1996-2000*. Kuala Lumpur: Percetakan Nasional Malaysia Berhad.
- 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.
- Graupe D. (1997). *Principle of Artificial Neural Networks*. Singapore: World Scientific.
- Jackson J. E. (2003). *A User's Guide to Principle Components*. New Jersey: John Wiley & Sons, Inc, Publication.
- Jose C. Principe, Neil R. Euliano, W. Curt Lefebvre. (2000). *Neural and Adaptive Systems: Fundamental Through Simulations*. New York: John Wiley & Sons, Inc.
- Julaihi Wahid. (2003). *Effective Policy on Affordable Housing - The Malaysian Experience*. Universiti Sains Malaysia.
- Lim, Choon Kooi, Sen, R. N. and Ahmad Rafi Mohamed. (2001). *Ergonomics and Architetural Designs: Develop to Improve Residential Low-Cost Housing in Malaysia*. Multimedia University.
- Morshidi Sirat, Abdul Fatah Che Mamat, Abdul Rashid Abd Aziz, Alip Rahim, Halim Saleh, Usman Hj. Yaakub. (1999). *Low-cost housing in Urban Industrial Centres of Malaysia: Issue and Challengers*. Penerbit Universiti Sains Malaysia.
- Negnevitsky, M. (2005). *Artificial Intelligence: A guide to intelligent systems*. 2nd ed. London: Pearson Education Limited.
- Noor Aini Yusof. (2007). *Pemaju Swasta dan Perumahan Kos Rendah*. Universiti Sains Malaysia.
- Noor Yasmin Zainun and Muhd Zaimi Abd. Majid. (2002). *Techniques To Developed Needs Model On Housing For Low Income Group: A Literature And Malaysian Experience*. 2nd International conference on Systems Thinking in

Management, UK, p. 28-37.

Noor Yasmin Zainun and Muhd Zaimi Abd. Majid. (2002). Techniques To Develop Needs Model On Housing In Urban Area: A Literature And Malaysian Experience. *First International Conference on Construction in the 21st Century (CITC2002)*, USA, p. 935-942.

Noor Yasmin Zainun and Muhd Zaimi Abd. Majid. (2003). Evaluation on Various Forecasting Models Using Artificial Neural Networks (ANN). *2nd International Conference on Innovation in Architecture, Engineering and Construction (AEC)*, UK.

Noor Yasmin Zainun and Muhd Zaimi Abd. Majid. (2003). Forecasting Demand on Low Cost Housing in Urban Area in Malaysia: Artificial Neural Network (ANN) Approach. *The Ninth East Asia-Pacific Conference on Structural Engineering and Construction*, Bali, Indonesia.

Noor Yasmin Zainun and Muhd Zaimi Abd. Majid. (2004). Low Cost Housing Demand Predictor (LOCHDEP). *International Exhibiton-Ideas-Invention-New Product (IENA)*, Nuremberg, Germany.

Ong, Han Ching and Prof. Lenard, D. (2002). Partnership between stakeholders in the Provision of an Access to Affordable Housing in Malaysia. *FIG XXII International Congress*. USA: Washington D.C.

Rao M. A. and Srinivas J. (2003). *Neural Networks: Algorithms and Application*. Pangbourne, UK: Alpha Science International Ltd.

Rencher A. C. (2002). *Methods of Multivariate Analysis*. New York: John Wiley & Sons, Inc., Publication.

Salleh Buang. (2004), Solving the low-coast housing woe. *News Straits Time*, Kuala Lumpur.

Schalkoff, R. J. (1997). *Artificial Neural Networks*. Singapore: McGraw-Hill.

Syafiee Shuid. (2004). Low Medium Cost Housing in Malaysia: *Issues and Challenges*. APNHR Conference, The University of Hong Kong.

Sundarajah N.and Saratchandran P. (1998). *Parallel Architecture for Artificial Neural Networks: Paradigms and Implementations*. Piscataway, N.J.: IEEE Computer Society.

The Sun. (2004). Government will study Selangor housing plan, says Ong. KualaLumpur.

Table 1. Component Score Coefficient matrix for Johor Bahru

	Component	
	1	2
Population growth	-.006	.634
Birth rate	-.134	-.151
Mortality baby rate	.109	.390
Inflation rate	.136	-.012
Income rate	.137	.038
Housing stock	.129	-.090
GDP rate	-.136	.096
Unemployment rate	.133	-.164
Poverty rate	-.135	.115

Table 2. Determination of learning and momentum rate.

	Phase 1	Phase 2	Phase 3	Phase 4
Learning rate	0.9	0.7	0.5	0.4
Momentum rate	0.1	0.4	0.5	0.6

Table 3. Actual and forecasted demand on low cost housing for October, November and December 2003 in Johor Bahru district

Time series	Actual data	Forecasted data	MAPE (%)	
			Phase 3	Phase 4
October 2003	203	165	18.47	20.42
November 2003	131	113	13.71	16.44
December 2003	99	81	18.22	21.29

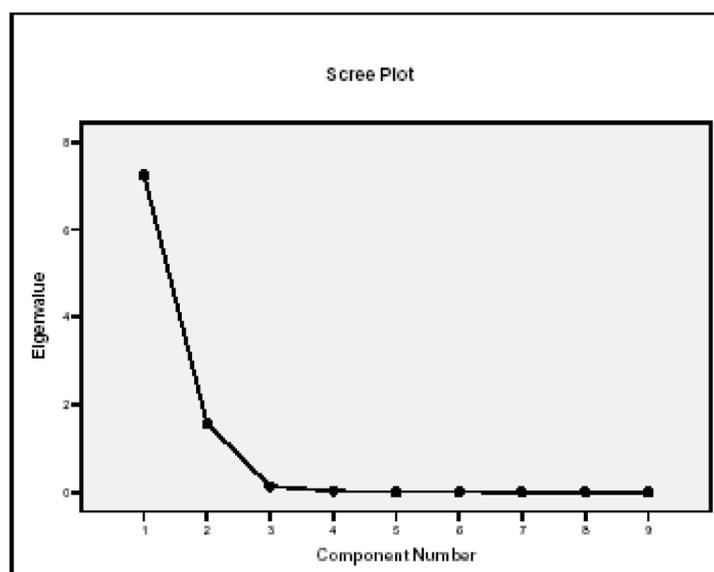


Figure 1. Scree plot for Johor Bahru

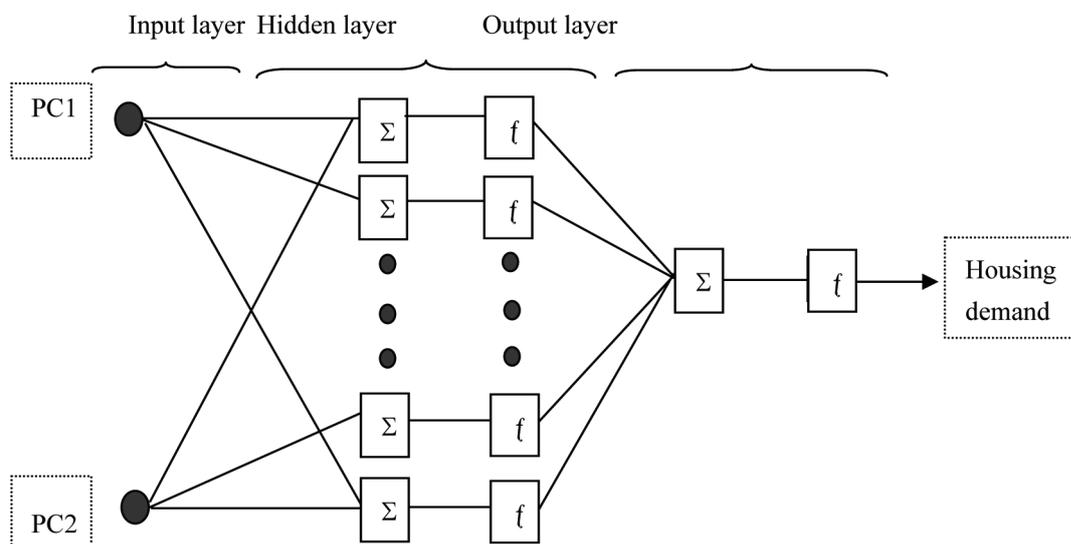


Figure 2. Neural Network topology with 2 inputs for Johor Bahru

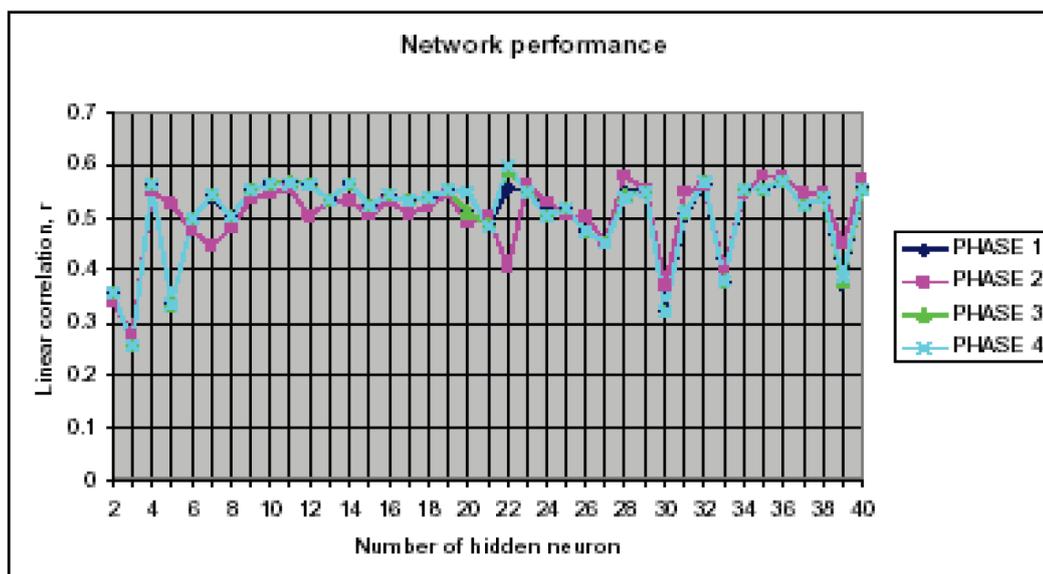


Figure 3. Network performance of testing with different number of neurons and phases

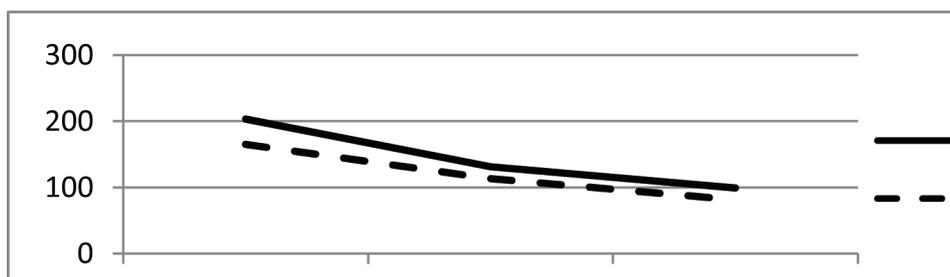


Figure 4. Graph of actual and forecasted demand data