Forecasting low-cost housing demand in Johor Bahru, Malaysia using artificial neural networks (ANN)

This item was submitted to Loughborough University's Institutional Repository by the/an author.


**Additional Information:**

- This article was published in the Journal of Mathematics Research [© Canadian Center of Science and Education (CCSE)] and licensed under a Creative Commons Attribution 3.0 License. The definitive version is available at: http://www.ccsenet.org/journal/index.php/jmr/article/view/1059

**Metadata Record:** [https://dspace.lboro.ac.uk/2134/11440](https://dspace.lboro.ac.uk/2134/11440)

**Version:** Published

**Publisher:** © Canadian Center of Science and Education (CCSE)

Please cite the published version.
This item was submitted to Loughborough’s Institutional Repository (https://dspace.lboro.ac.uk/) by the author and is made available under the following Creative Commons Licence conditions.

For the full text of this licence, please go to: http://creativecommons.org/licenses/by-nc-nd/2.5/
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, $R = 2.84\times10^{-14}$ 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 $N=36$.

\[
-a.\ln(v) = -(N - 1 - (2p + 5)/6)\ln(R) = -(36 - 1 - (2\times9 + 5)/6)\ln(2.84\times10^{-14}) = 972.16
\]

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, $= 0.001$, 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 Cattan (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 22 numbers of hidden neurons 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


Table 1. Component Score Coefficient matrix for Johor Bahru

<table>
<thead>
<tr>
<th>Component</th>
<th>1</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Population growth</td>
<td>-0.06</td>
<td>0.634</td>
</tr>
<tr>
<td>Birth rate</td>
<td>-0.134</td>
<td>-0.151</td>
</tr>
<tr>
<td>Mortality baby rate</td>
<td>0.109</td>
<td>0.390</td>
</tr>
<tr>
<td>Inflation rate</td>
<td>0.136</td>
<td>-0.012</td>
</tr>
<tr>
<td>Income rate</td>
<td>0.137</td>
<td>0.038</td>
</tr>
<tr>
<td>Housing stock</td>
<td>0.129</td>
<td>-0.090</td>
</tr>
<tr>
<td>GDP rate</td>
<td>-0.136</td>
<td>0.096</td>
</tr>
<tr>
<td>Unemployment rate</td>
<td>0.133</td>
<td>-0.164</td>
</tr>
<tr>
<td>Poverty rate</td>
<td>-0.135</td>
<td>0.115</td>
</tr>
</tbody>
</table>

Table 2. Determination of learning and momentum rate.

<table>
<thead>
<tr>
<th></th>
<th>Phase 1</th>
<th>Phase 2</th>
<th>Phase 3</th>
<th>Phase 4</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Learning rate</strong></td>
<td>0.9</td>
<td>0.7</td>
<td>0.5</td>
<td>0.4</td>
</tr>
<tr>
<td><strong>Momentum rate</strong></td>
<td>0.1</td>
<td>0.4</td>
<td>0.5</td>
<td>0.6</td>
</tr>
</tbody>
</table>
Table 3. Actual and forecasted demand on low cost housing for October, November and December 2003 in Johor Bahru district

<table>
<thead>
<tr>
<th>Time series</th>
<th>Actual data</th>
<th>Forecasted data</th>
<th>MAPE (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>October 2003</td>
<td>203</td>
<td>165</td>
<td>18.47</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>20.42</td>
</tr>
<tr>
<td>November 2003</td>
<td>131</td>
<td>113</td>
<td>13.71</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>16.44</td>
</tr>
<tr>
<td>December 2003</td>
<td>99</td>
<td>81</td>
<td>18.22</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>21.29</td>
</tr>
</tbody>
</table>

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