First International Conference on Construction in the 21st Century (CITC2002)
"Challenges and Opportunities in Management and Technology"
25-26 April, 2002, Miami, Florida, USA

COMPARATIVE STUDY ON FORECASTING DEMAND OF LOW COST HOUSE IN URBAN AREAS USING ARTIFICIAL NEURAL NETWORKS AND ARIMA MODEL

Muhd Zaimi Abd Majid
Director, Construction Technology and Management Centre (CTMC)
Universiti Teknologi Malaysia, Johor, Malaysia

Khairulzan Yahya
Faculty of Civil Engineering
Universiti Teknologi Malaysia, Skudai, Johor, Malaysia

ABSTRACT
As a developing nation, Malaysia is experiencing a rapid growth of urbanisation due to the expansion of urban economic activities, namely industry, commerce, construction and services. Concomitant with this, the housing demand in urban areas of Malaysia has been increased year by year especially for the low-income group. An accurate prediction of the level of demand for the low cost housing is vital to forecast the future demand. This study covers districts of Gombak in the state of Selangor, Malaysia that are among the areas that experienced a high level of urbanisation. Monthly data of housing demand was collected from the office State of Selangor, Malaysia. Trend of housing demand for five years period from February, 1996 until November, 2000 has been identified and used to forecast the housing demand. Two forecasting techniques were applied to compare the accuracy of short term forecasting, that are, Artificial Neural Network (ANN) and Autoregressive Integrated Moving Average (ARIMA). The accuracy of the models is measured by using the Mean Absolute Percentage Error (MAPE). The minimal MAPE values for both models are identified to be less than 10%.

KEYWORDS
Forecasting, ARIMA, Artificial Neural Networks

1. INTRODUCTION

Gombak, located in Selangor, which is known as the heartland of the nation, consists of 650 square km and has a population of around 553,410. The rate of population growth had decreased to 5.01% in 2000 from 6.85% in 1991, however the proportion of urban population had increased to 89.1% with only 5% of it area used for housing scheme, making it as among the most urbanised district in Malaysia. Strategically located in nearby Kuala Lumpur, the Capital city of Malaysia, it is was observed among the area which experiencing the robust economic growth through it continuous development of industrial sector as well as the establishment of government and private institutions which covered 38% of the land area.

With respect to this urbanisation, the demand on affordable houses has increased each year. Thereby, the government policy in providing shelters for every family in the country especially from the low-income group has been implemented to alleviate the demand. In this respect, a reliable demand forecast model on low cost housing (LCH) is required. However, attempt to estimate these demand are fraught difficulties, owing to the characteristic of construction and the pronounced fluctuations often caused by the changes in economy (Goh B.H, 1998). Moreover, the good forecasting is influenced by the techniques employed (Goh, B.H, 1997& 1998 and Zheng, R.Y and Packer, A, 1997).

Goh, B.H, 1997 and 1998 cited the performances of Artificial Neural Network (ANN) which using the economic data, gave the best model with outstanding accuracy for prediction of residential construction demand in Singapore, followed by the univariate ARIMA. In the United Kingdom construction demand, Zheng, R.Y and Packer, A, 1997, reported, ANN model for housing and commercial are much better than the regression model previously developed by Akitonye and Skitmore, 1994 with the exception for industry sector. Moreover, ANN has been found more robust and better in the case of long-term prediction (Zaiyong,T et. al, 1991, Roslina Salleh, 1999). In this study, ANN and univariate ARIMA approach were utilised to develop the LCH demand for Gombak district.

2. OBJECTIVE

The objective of this study is to derive the forecast demand model on LCH in the most urbanised district of Malaysia using the two different techniques; Univariate and multivariate ARIMA and ANN. The forecasting results from these two techniques are compared and the best technique is revealed in term of predictive accuracy.

3. METHODOLOGY

A list of indicators which theoretically influenced the demand of LCH in Malaysia and beyond the boundary were taken into consideration to ensure that all possible factors can be used as significant modelling variables (R. Chander, 1976, Akitonye and Skitmore, 1994, Zheng, R.Y and Packer, A, 1997, Goh, B.H, 1997 and 1998, Ahmad Zaki Yahya and R. Ramachandran 1998, Morshidi Sirat, 1999). The LCH demand and it indicators (referred as independent variable) were collected and entered to a computer for pre-processing and statistical analysis. The backward-elimination process variable selection, available on SPSS for Release 7.5 statistical package was used to select statically significant indicators. These were used as independent variables in ANN technique and multivariate ARIMA. For ARIMA modelling, only the data on the demand of low cost house was required to develop Univariate ARIMA. To compare the ability in forecasting, ANN (Neural Connection 1.0) and univariate and multivariate ARIMA model were developed using the same set of data, which consist of 49 data for training and validation while 9 remaining data were used for prediction. The forecasting accuracy was objectively measured by comparing the relative percentage errors, called MAPE, between forecast demand and real demand data for all models.

4. DATA COLLECTION AND PRE-PROCESSING

The LCH demand data were comprehensively collected from the office of Selangor Secretary of State, which consisted of monthly data from February 1996 to November 2000. The data published before February 1996 was incomplete while the data after November 2000 were not released at the point of data collection. The housing demand were represented by the number of applications whom were approved to apply for a LCH through the special open registration associated with government policy. Fig. 1 shows the pre-processed demand data that shows fairly deterministic seasonal pattern.

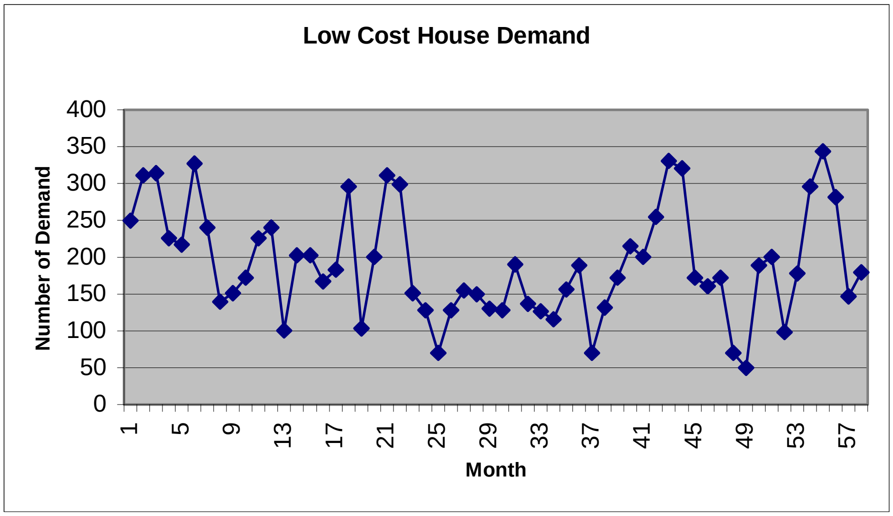

Figure 1: Low Cost House Demand (LCH) Time Series

The realistic theoretical economic data were abstracted from the data bank at district and state level, established by the Department of Statistic, Malaysia to develop an independent variable time series data set. These independent variables selected must be within the demand period considered. In summary, five independent variables associated with the LCH demand at the district and state level were found statically significant to be included in the demand model. The five independent variables were: population; infant mortality rate; GDP/Capita and LCH stock. They were found to contribute only 38% in demand variation. The statistical summary for the significant independent variable is shown in Table 1. Data were unavailable to quantify for instance the government policy and national campaign for housing, so these variables were not considered. The government policy includes an open registration system, only eligible for those with an income of about USD500 per month or below, those have to make way for development or whose houses are destroyed in disasters.

Table 1: Backward Elimination Process Parameter Estimated

Independent variable	Coefficient of regression B	Standard error	Beta	t	Sig.	R	R^2	$\check{R}^2$
Constant	-50.305	13.305		-3.781	.000	.615	.379	.319
Population	5.832	1.579	.956	3.694	.001			
Mortality rate	8.662	3.143	.318	2.756	.008			
Inflation rate	.565	.144	1.629	3.937	.000			
GDP/capita	3.104E-02	.007	.827	4.229	.000			
Stock	.226	.048	1.399	4.735	.000			

5. APPLICATION OF SEASONAL ARIMA

ARIMA modelling, which Box and Jenkins suggested was used. Both, univariate and multivariate using the same approach, i.e. statistical and it consist of three stages; identification, estimation and diagnostic checking, and application. Fig 2 shows the flow chart of ARIMA procedure (Makridakis, S, 1976, Goh, B.H, 1998). The difference between these two, univariate and multivariate is that the former using only the historical data whereas the last using the five selected significant independent variable as an input data. Once a tentative model has been identified, its parameters estimated, and its adequacy is tested against the real time series through the diagnostic checking.

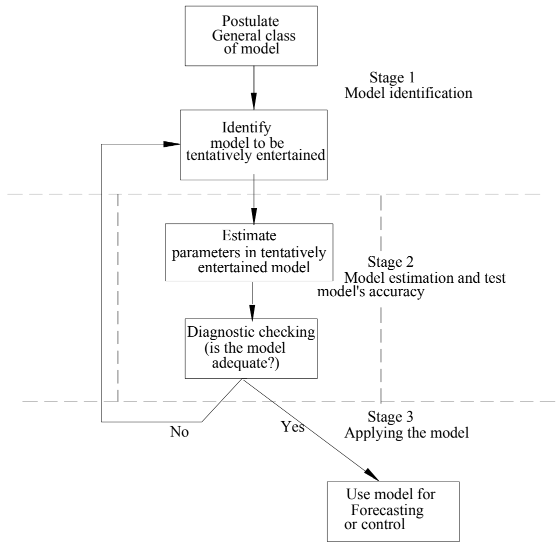

Figure 2: ARIMA Procedure (Makridakis,S, 1976, Goh,B.h, 1998)

An ARIMA model is designed using only stationary time series, thereafter; this time series was transformed (Bowermann and O'Connell, 1992), through the method of differencing. First seasonal differencing was found to give stationary time series as it's gave the constant mean and variance. The model development has been described in detail by Makridakis, S, 1976, Bowerman & O'Connell, 1976 and 1992. Using the model development guideline by Bowermann and O'Connell, 1992, the general Box-Jenkins model of order (p,d,q)(P,D,Q) has been identified. p is a non-seasonal autoregressive coefficient, P the order of seasonal autoregressive component, q the non-seasonal moving average component, Q the order of seasonal moving average component, d the degree of differencing and D the of seasonal differencing.

In this study, univariate ARIMA model of ARIMA (0,0,1)(1,1,0) was determined suitable with the parameter estimated using the SPSS Release 7.5 as shown in Table 2. Model without constant were selected based on the t-statistic of the parameter. This model was checked against the goodness of fit through the diagnostic checking. Checked on function of auto-correlation (ACF) and function of partial auto-correlation (PACF) showed that all autocorrelations were random. This confirmed the fitted model was adequate, hence, the best fit ARIMA model for LCH demand are ARIMA (0,0,1)(1,1,0) without constant.

Table 2: Parameter Estimated for Model Univariate ARIMA (0,0,1)(0,1,1)

	B	SEB	T-RATIO	APPROX. PROB.
θ_1	-.531	.154	-3.434	.001
$\phi_{1,12}$	-.521	.151	-3.433	.001

```
Number of residuals  37
Standard error       .41347356
Log likelihood       -20.899803
AIC                  45.799606
SBC                  49.021442
Residuals :   DF: 35 ;Adj. Sum of Squares: 6.693; Residual Variance :.170
```

And so for multivariate ARIMA, the same model of ARIMA (0,0,1)(1,1,0) was found capable of predicting the demand. Again, model without constant were selected, and the computed parameter of ARIMA (0,0,1)(1,1,0) is shown in Table 3. This model was checked, have been confirmed the fitted model for LCH demand.

Table 3: Parameter Estimated for Model Multivariate ARIMA (0,0,1)(0,1,1)

	B	SEB	T-RATIO	APPROX. PROB.
θ_1	-.369	.169	-2.178	.037
$\phi_{1,12}$	-.548	.144	-3.800	.0006
Population	3.872	6.250	.619	.540
Infant mortality rate	5.374	4.921	1.092	.283
inflation	.687	.392	1.749	.090
GDP/capita	.460	.179	2.560	.0157
Stock	.168	.1159	1.450	.157

```
Number of residuals  37
Standard error       .35756801
Log likelihood       -12.864156
AIC                  39.728311
SBC                  51.004737

Residuals :    DF :30 ; Adj. Sum of Squares: 4.325;Residual Variance: .1278
```

6. APPLICATION OF ARTIFICIAL NEURAL NETWORK (ANN)

Unlike ARIMA, ANN modelling uses a system, which replicates the human brain's learning process. It is a system that learns by experience and generalisation from previous example to new problem. In the present study, back propagation ANN, a gradient–decent learning algorithm and a popular supervised learning was used. Like human brain, it's made up of artificial neuron, called non-linear computational Processing Element (PE), arrange in several layers; an input layer, an output layer and at least one layer of processing element or hidden layers in between.

Based on the concept of back propagation, input and output pattern were presented to the network. In this work, two significant indicators were used as input data and so the size of the input layer is 5 nodes as shown Fig 3. The output layer consists of one node that corresponds with the output variable, demand for LCH. The number of hidden layer through which the information is processed are usually determined by trial and error according to complexity of the problem (Elhag,T.M.S and Bousabaine,A.H, 1999, Goh, B.H, 1996, Roslina Salleh, 1999). In addition, Master's technique (Master, T. 1994 and Roselina Salleh, 1999) have also been used and experimented with, and as a result, a hidden layer of two nodes was found capable to give the minimum root means square error (RMSE) and the best MAPE. This study showed a hidden layer of two nodes was associated with the Master's technique. The gradient–decent learning algorithm, sigmoid transfer function and all software default setting parameters were used for training. Training was continued until the point where the validation error gradient increases again indicating the system has experienced best network. At this point the network is trained and it should be able to provide the correct output when presented with the associated inputs. The basic propagation training algorithm has been covered widely by Rumelhart et al., 1986 and Harun. S, 1999.

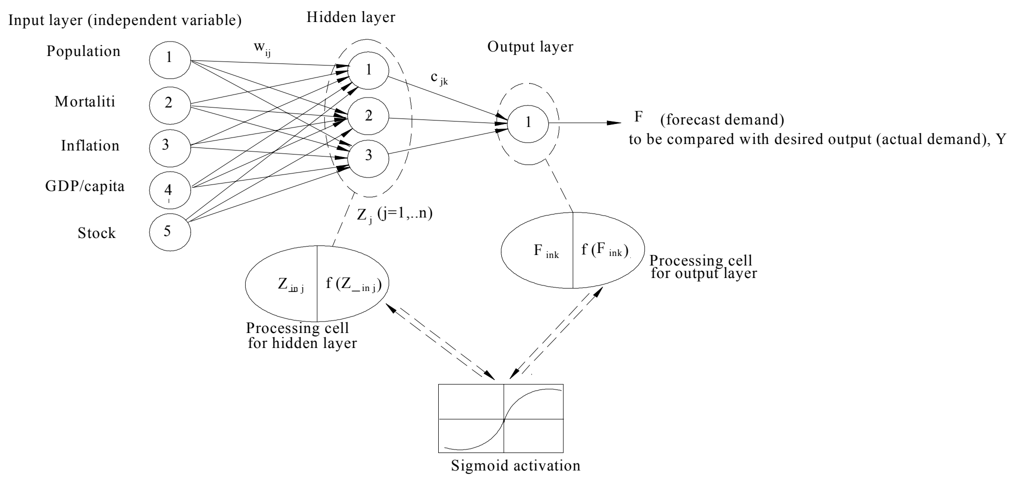

Figure 3: Artificial Neural Network Architecture for Gombak's LCH demand

7. COMPARISON OF RESULT

Table 3 shows the forecasting ability of each model, and it's found that univariate ARIMA model appears to correlate stronger than multivariate ARIMA and ANN. It was showed that multivariate ARIMA and ANN models produced under forecast. In term of accuracy, it's was measured by means absolute percentage error, MAPE. Univariate ARIMA (0,0,1)(0,1,1) models performance produced a small MAPE result of 5.61 followed by ANN model of 6.20 and Multivariate ARIMA over 10% for the forecasting 9 months data ahead. The result comparatively showed that the univariate ARIMA model marginally had the lowest MAPE.

Table 4: Forecasting ability of ARIMA and ANN

Month	Actual data	Forecast ARIMA Univariate	Forecast ARIMA Multivariate	Forecast ANN
Mar. '00	189	138	71	114
Apr. '00	200	163	86	123
May '00	98	178	100	126
Jun '00	178	160	94	126
Jul '00	296	178	109	130
Aug. '00	344	248	163	131
Sept.'00	281	206	145	133
Oct. '00	147	147	109	134
Nov. '00	179	135	102	135
MAPE	-	5.61	11.83	6.20

8. DISCUSSION

In term of accuracy, univariate ARIMA and ANN model produced comparable MAPE of less than 10%, however the univariate ARIMA did outstandingly better. The deterministic pattern of demand time series has given advantage to univariate ARIMA models to describe and carry out the linear input and out put mapping accurately.

Mathematically, the forecast for the period t is the sum of the value of the weighted term to be forecast and past values of forecast errors that in turn increases model accuracy. The lowest MAPE produced by univariate model support previous findings on the study on airline passenger data by Zaiyong,T et al. (1991).

The ANN out performed multivariate ARIMA model, showed that the ability of non-linear mapping automatically between input and out put data in ANN, although the forecast demand did not visually correlate with actual demand as strong as univariate or even with multivariate ARIMA's model. In contrast with study conducted by Goh B.H, 1998, ANN model outperformed univariate ARIMA and one of the reasons was that the independent variables used have a coefficient of determination of relatively very high. In this study, the coefficient of determination of the significant independent variable was low (0.38); therefore, this might be one reason ANN unable to perform remarkably. In this respect, even though from previous literature research denoted that ANN can works with any data, however it has been lack of information regarding on the effect of the output pattern. In addition, many researchers only dealt with the independent variable with high coefficient of determination (Goh.B.H, 1998, Aiken,W at al, 2000).

With respect to the low coefficient determination, multivariate ARIMA model unable to perform linear model as good as univariate model. In fact, the independent variable significance which previously determined by linear regression (through backward elimination process), only 2 significant independent variables have t-statistic greater than 2, i.e. inflation rate and GDP/capita. Therefore, in order to explore the superiority of the ANN, further research is required to identify the unidentified independent variables, which contributed another 62% variation in demand as well as the application of univariate ANN. In addition, the collection of more sample data may give rise to an improvement to the process of learning and generalization that result in a more accurate forecasting.

9. CONCLUSION

Therefore, the general conclusion from the comparative study is univariate ARIMA analysis was found to be the most accurate. It could be implied at this stage that the demand on LCH in urban areas in Malaysia can be forecast without its indicators and the future value of demand are assumed to be influenced only by the current demand values of the modelling indicators.

10. REFERENCES

Ahmad Zaki Yahya and R. Ramachandran (1998), "Implication of Population Profile on Housing Needs and Demand in Vision 2020", *Jurnal Perumahan*, National Housing Department. Article No.1.

Aiken et. al, (2001) "Forecasting Construction Demand With A Neural Network". [online] Available http://www.sbaer.uca.edu/docs/proceedingsII/98dsi064.htm

Akitonye and Skitmore (1994), "Model of UK Private Sector Quarterly Construction Demand", *Construction Management and Economic*, 12, 3-13.

Bowermann, B.L, dan O'connell, R.T., (1992), "Forecasting And Time Series: An Applied Approach, 3rd Ed"., Duxubury Press, Belmont, CA.

Department of Statistic, Malaysia (Mac 1995), "Laporan Am Banci Perumahan, Banci Penduduk dan Perumahan Malaysia 1991", Department of Statistic, Malaysia.

Department of Statistic, Malaysia (1999), "Bank Data Negeri/Daerah 1999 and 2000", Department of Statistic, Malaysia.

Elhag T and Boussabaine A H (1999) "Tendering price estimation: Artificial Neural Networks Vs Regression Analysis", COBRA conference, Salford,U.K, pp 114-123

Goh, B.H. (1996), "Residential Construction Demand Forecasting Using Economic Indicators: A Comparative Study of Artificial Artificial Neural Networks and Multiple Regression", *Construction Management and Economic*, 14, 25-34.

Goh, B.H. (1998), "Forecasting Residential Construction Demand in Singapore: A Comparative Study of the Accuracy of Time Series, Regression and Artificial Artificial Neural Network Technique", *Engineering Construction and Architectural Management,* 5, no.3 pp.261-275.

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

Isaac F. Megbolugbe, Allen P.Marks, Mary B. Schwartz. (1991), "The Economic Theory of Housing Demand: Critical Review", *The Journal of Real Estate Research*, Vol 6 No.3 381-393

Makridakis, S (1976), "A Survey of Time Series", *International Statistical Review*. 44 (1) : 29-70

Masters, T. (1994), " Practical Neural Network Recipies In C++.", New York : Academic Press.

Morshidi Sirat, Abdul Fatah Che Mamat, Abdul Rashid Abd Aziz, Alip Rahim, Halim Salleh, Usman Hj. Yaakob, (1999), "Low-cost Housing in Urban-Industrial Centres of Malaysia: Issue and Challenges", Penerbit Universiti Sains Malaysia.

SPSS Inc. (19). "Neural Connection 1.0 User's Guide", North Michigan Avenue, Chicago.

R. Chander (1976), "Housing Needs versus Effective Demand in Malaysia", Department of Statistic of Malaysia.

Roselina Salleh (1999), "Penggunaan Model Rangkaian Neural Dalam Peramalan Siri Masa", Msc Thesis, Universiti Teknologi Malaysia.

Rumelhart D. E., Hinton G.E., and Williams R.J. (1986). "Learning internal representations by error propagation". *Parallel distributed processing: explorations in the microstructure of cognition*; Vol. 1, D.E. Rumelhart and J.L. McClelland, eds., MIT Press, Cambridge, Mass. 318-362

SPSS Inc. (1995). "SPSS Software User's Guide R.6.0", North Michigan Avenue, Chicago.

Zaiyong,T ,Almeida. C de and Fishwick, P.A, 1991, "Time Series Forecasting Using Artificial Neural Networks vs. Box-Jenkins Methodology", *Simulation*, 57 (5) : 303-310

Zheng, R.Y and Packer, A, 1997, "Applying Artificial Artificial Neural Networks to UK Construction Demand Forecasting (Private Sector)",*RICS Foundation*,http://www.rics-foundation.org/publish/documentdetails.jsp?ID=241