TECHNIQUES TO DEVELOP FORECASTING MODEL ON LOW COST HOUSING IN URBAN AREA

Noor Yasmin Zainun
Post Graduate Student, Construction Technology & Management Center,
Faculty of Civil Engineering,
Universiti Teknologi Malaysia, Skudai, Johor, Malaysia.
Email: yasminzainun@hotmail.com

Muhd. Zaimi Abd. Majid
Director, Construction Technology & Management Center,
Faculty of Civil Engineering,
Universiti Teknologi Malaysia, Skudai, Johor, Malaysia.
Email: zaimi@fka.utm.my

ABSTRACT

The number of people who will live in urban areas is expected to double to more than five billion between 1990 to 2025. Therefore, accurate predictions of the level of aggregate demand for housing are very important. Various forecasting techniques have been developed using probabilistic, statistics, simulation or artificial intelligent. Hence, there is a need to identify different techniques, in terms of accuracy, in the prediction of needs for facilities. This paper discusses the Artificial Neural Networks (ANN) technique and compaes it with other techniques in forecasting needs of housing in urban area. Investigation on previous research and literature materials will be derived and compared in terms of errors in the accuracy of the technique. The findings of this study indicates that the ANN model performs best overall.

Keywords: *urban area, accuracy, artificial neural network, forecasting.*

INTRODUCTION

Urbanization has become a global phenomenon although the degree of urbanization and the rate of urban growth vary in different parts of the world. According to Guido, in his studies on urban forestry in Asia-Pacific region, urban area is the built-up or densely populated area containing the city proper; suburbs, and continuously settled commuter areas. The definition of “urban area” varies from country to country, for example, in the USA most conservative delineation of urban land requires a population density of 620/km^2 while in Malaysia, urban areas have 10,000 inhabitants.

The first systematic major collection of statistics on housing in Peninsular Malaysia was undertaken in 1970. Since then, studies on housing have been conducted extensively in Malaysia such as socio-economic considerations of human settlements and housing (Kamal Salih, 1976), housing needs versus effective demand in Malaysia 1976-1990 (Chander, 1977), and housing needs in Peninsular Malaysia (Chander, 1974).

Malaysia has experienced spectacular urban spatial transformations from 1970 to 1997. The total Malaysian population has increased at the rate of around 2.8 per cent per year. Due to the increment of the demand for houses especially in urban area, the selection of the best method on forecasting of demand is also becoming an important factor. In view of this, there is an increasing need to objectively identify a forecasting technique which can produce an accurate demand forecast for housing.

Artificial Neural Networks have been successfully applied in numerous areas such as construction cost prediction (Li, 1995), risk analysis (Yang *et.al.* 1997) and forecasting bond ratings (Dutta and Shekhar, 1998). Neural networks also have outperformed regression in stock market returns (Kimoto *et.al.* 1990), predicting bank failures (Salchenberger *et.al.* 1992) and property values (Do and Grudnitaki, 1992). Artificial Neural Networks also have been successfully applied to time series forecasting, for example in stock prediction (Donaldson and Kamstra, 1996), currency exchange rate prediction (Refenes *et.al.* 1993) and electricity demand forecasting (Connor, 1996).

Generally, artificial intelligence is defined as the science and engineering of making intelligent machines, especially intelligent computer programs. Artificial Neural Network is a system loosely modeled on the human brain. It is known by many names, such as connectionism, parallel distributed processing, neuro-computing, natural intelligent systems, machine learning algorithms, and artificial neural networks. It is a network of many simple processors and the multiple layers of simple processing elements called neurons. Each neuron is linked to certain of its neighbours with varying coefficients of connectivity that represent the strengths or weights of these connections. These weights are obtained by a process of adaptation to a set of training patterns. Learning is accomplished by adjusting these weights to cause the overall network to output appropriate results.

The type and variety of artificial neural networks is virtually limitless. However, neural networks are classified according to two factors; (1) the topology of the network and; (2)the learning method used to train the network. The most widely used topology is the feedforward network and the most common learning method is the backpropagation of errors.

Objective

The objective of this paper is to discuss forecasting models using Artificial Neural Networks (ANN) and compare it with other forecasting techniques. The accuracy of the ANN models is discussed from the comparison test results with other models.

Methodology

The methodology of this study has included the evaluation on the various research methods conducted by previous researchers. The study involves investigation involves an extensive literature review, comparing and analysing the literature data.

Study on housing forecasting

Housing starts; a year 2000 forecast

Myers and Smith (1998) have conducted a study to forecast the nation's housing starts for the years 1999 and 2000 based on the Keynesian investment function model, assuming housing is determined by two-year-lagged long-term interest rates and real total consumption spending.

Investment in housing is a critical determinant of the direction of the economy (Myers and Smith, 1998). Whenever consumer confidence falls investment in housing and total consumption spending (TCS) would be expected to decrease, but housing investment would also be expected to fall as TCS rises whenever long-term interest rate (LTIR) rise. These events would be leading indicators of a recession. Since interest rates could rise unpredictably due to unanticipated inflation, Myers and Smith (1998) restricted the forecast horizon to two years. Therefore, the reliability of the findings increases.

In this study, regression on the forecast data were performed on two parameters. In the first regression, housing starts is compared to TCS and LTIR for the months between January 1990 and December 1998. The Keynesian investment function that has been used to determine the prediction of housing starts is :-

$$\text{Housing Starts} = f(\text{LTIR},\text{TCS}) \qquad \ldots.(1)$$

A regression was then run with the right hand side variables lagged two years comparing housing starts to TCS and LTIR between the months of January 1992 and December 1998. The modified equation is :-

$$HS_T = -828.07(-3.90) + 7.96(0.71)LTIR_{T-24} + 0.49(11.47)TCS_{T-24} \qquad \ldots..(2)$$

Forecast housing starts calculated from equation (2) are given in Table 1 below.

Forecast of Housing Starts

Date	*Housing Starts*	*Annualized Percent Change in HS*
November-98	1654.00	10%
December-98	1738.00	14%
January-99	1634.38	7%
February-99	1651.84	0%
March-99	1640.24	4%
April-99	1644.29	7%
May-99	1645.05	7%
June-99	1649.23	1%
July-99	1700.98	-1%
August-99	1697.29	5%
September-99	1698.41	8%
October-99	1690.60	0%
November-99	1717.58	4%
December-99	1711.31	-2%
January-2000	1728.50	6%
February-2000	1739.79	5%
March-2000	1740.21	6%
April-2000	1772.38	8%
May-2000	1789.14	9%
June-2000	1803.35	9%
July-2000	1789.55	5%
August-2000	1812.10	7%
September-2000	1814.82	7%
October-2000	1825.35	8%
November-2000	1823.93	6%
December-2000	1832.85	7%

Table 1: Forecasting of Housing Starts

Adapted from Myers and Smith

From Table 1 it can be seen that forecast is consistent with the past behavior of the variables. However, housing starts will continue to rise over the next two year if LTIR remains low and TCS continues to rise.

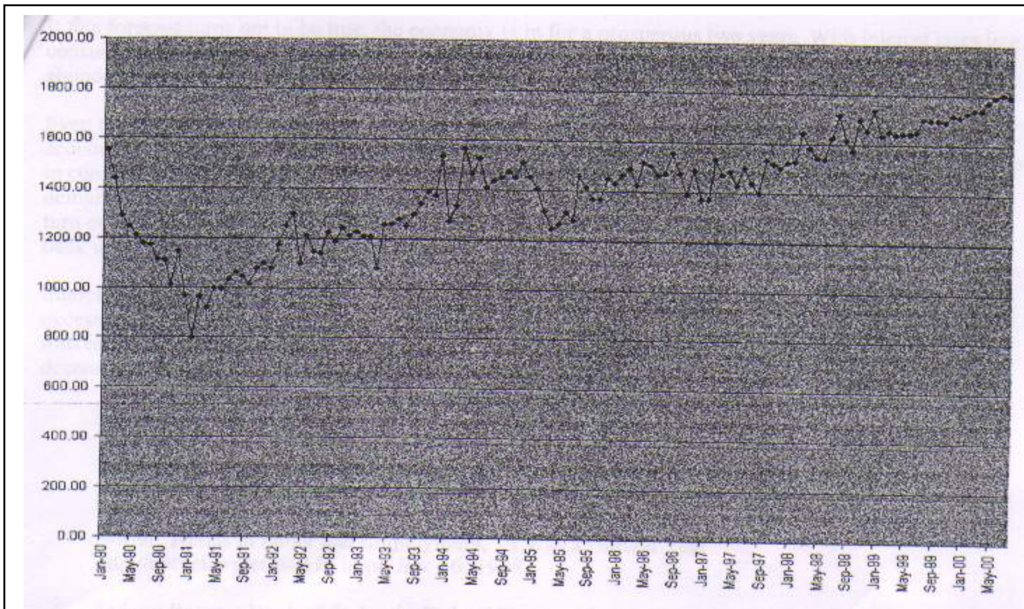

Figure 1 : Housing Starts from 1990-2000
Adapted from Myers and Smith

Figure 1 which represents the graph between the actual and forecast housing starts, shows the continual rise in housing starts. From the results it was concluded that housing starts are forecasted to rise three percent per year over the year 2001 and 2002.

2.2 Construction demand for residential properties in Thailand

A study was done by Tang *et.al.*(1990) examined the construction demand for residential properties in Thailand. The study was done using regression for the period from 1976 to 1985 using five indicators; (1) per capita income; (2) population; (3) the relative price index; (4) rate of household formation; and (5) interest rates. From the result, the study showed that population and the relative price index were significantly correlated at 5% level, and the F-test and coefficient of determination of 0.96 indicated that the combined explanatory variables had a significant impact on residential construction demand. However, no forecast of future construction was done in this study.

3.0 Application of Artificial Neural Networks

3.1 United Kingdom (UK) demand forecasting in private sector

The earliest work of modelling UK construction demand forecasting was done by Akintoye and Skitmore (1994). Ten indicators were used; (1) population; (2) interest rate; (3) shocks to economy; (4) the demand for goods; (5) surplus manufacturing capacity; (6) the ability to remodel (meeting demand through renovation); (7) government policy (monetary, fiscal, e.g. tax policies); (8) expectation of continued increased demand (demand for manufacturing goods); (9) the expectation of increased profits (on the activities of those that demand construction) and; (10) new technology; to construct the models based on multiple linear regression.

Using the same data set as Akintoye and Skitmore (1994), Yang and Parker (1997) have used two popular regression neural networks, back-propagation neural network (BPNN) and general regression neural network (GRNN) to investigate UK construction demand forecasting in the private sector.

Four parts of simulation were used, namely; one quarter ahead, two quarter ahead, three quarter ahead and four quarter (one year) ahead. Table 2 show the forecasting results of one quarter ahead and Table 3 gives the simulation results of two, three and four quarters ahead of forecasting.

From Table 2 the results in commercial sector show that ARMA-GRNN with robust estimate of the prediction errors has the best performance. In the housing sector, the results show that GRNN are better than BPNN while in the industry sector, there is no large difference between the models. All the models suffers from under estimation, where the values of MPE are negative. The minimum value of under estimation is 3.37% and the highest value of under estimation is 11.56%. Table 2 also gives a comparison between the results of neural network models and those from Akintoye and Skitmore (1994). It can be seen that except for the industry sector, new models are much better than the MR models developed by Akintoye and Skitmore (1994).

	Commercial sector		Housing sector		Industry sector	
	MPE	MAPE	MPE	MAPE	MPE	MAPE
BPNN	-11.56%	14.77%	-5.20%	16.43%	[redacted]	15.18%
AR-GRNN	-7.98%	12.68%	[redacted]	19.18%	-5.85%	14.48%
ARMA-GRNN	-8.09%	13.43%	-3.37%	19.18%	-5.83%	[redacted]
ARMA-GRNN (robust)	[redacted]		-3.92%	[redacted]	-5.81%	14.95%
Akintoye's work	50.8%	51.7%	46.4%	43.3%	-9.9%	[redacted]

Table 2: One quarter ahead forecasting
Adapted from Yang and Parker; *Applying Artificial Neural Networks to UK Construction Demand Forecasting (Private Sector)*, (1997).

Two quarters ahead	Commercial sector		Housing sector		Industry sector	
	MPE	MAPE	MPE	MAPE	MPE	MAPE
BPNN	-13.29%	17.63%	-5.19%	[redacted]		17.39%
AR-GRNN	-12.20%	15.90%	[redacted]	20.12%	-8.84%	[redacted]
ARMA-GRNN	-12.33%	16.30%	-5.43%	20.42%	-8.79%	17.23%
ARMA-GRNN (robust)	[redacted]		-5.73%	20.82%	-8.75%	17.47%
Three quarters ahead	Commercial sector		Housing sector		Industry sector	
	MPE	MAPE	MPE	MAPE	MPE	MAPE
BPNN	-14.67%	20.85%	[redacted]	21.37%	[redacted]	19.71%
AR-GRNN	-15.03%	19.73%	-6.88%	[redacted]	-10.88%	[redacted]
ARMA-GRNN	-15.44%	20.19%	-7.20%	20.18%	-10.77%	19.27%
ARMA-GRNN (robust)	[redacted]		-7.45%	20.76%	-10.69%	19.82%
Four quarters ahead	Commercial sector		Housing sector		Industry sector	
	MPE	MAPE	MPE	MAPE	MPE	MAPE
BPNN	[redacted]	23.32%	[redacted]			
AR-GRNN	-16.70%	23.34%	-7.79%	23.39%	-12.66%	20.09%
ARMA-GRNN	-16.79%	23.84%	-8.38%	23.21%	-12.67%	20.32%
ARMA-GRNN (robust)	-16.53%	[redacted]	-8.83%	23.47%	-12.68%	20.72%

Table 3: More than one quarters ahead forecasting
Adapted from Yang and Parker; *Applying Artificial Neural Networks to UK Construction Demand Forecasting (Private Sector)*, (1997).

Table 3 shows that BPNN plays a more important role in the long term forecasting than GRNN since the extrapolation ability of BPNN is better than GRNN in general.

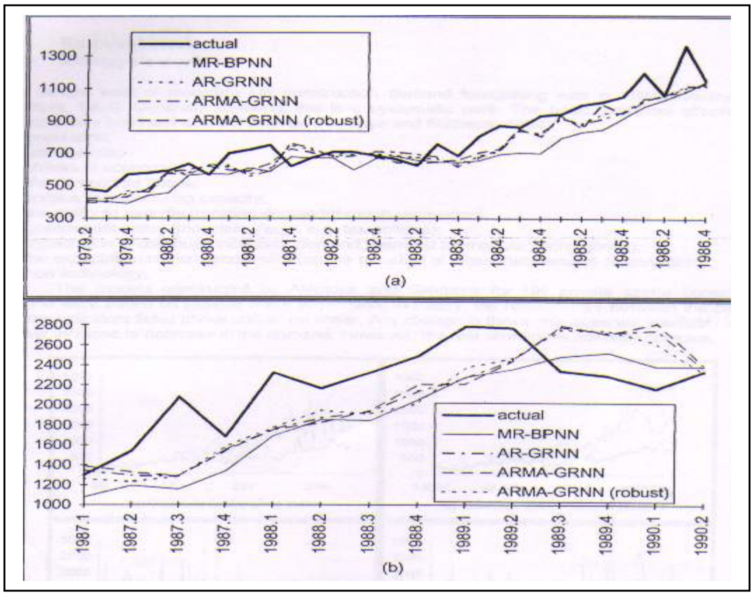

Figure 2: The prediction curves of the commercial sector

Adapted from Yang and Parker; *Applying Artificial Neural Networks to UK Construction Demand Forecasting (Private Sector)*, (1997).

Figures 2 (a) and (b) give the plots of the test results in the commercial sector that have been done by Yang and Parker (1997). It can be seen from Figure 2 (a) that before 1987, the GRNN models have a better accuracy than BPNN models, whilst after 1987, from Figure 2 (b), it can be seen that the prediction accuracy of the BPNN model is better than the GRNN model. This is because the demand before 1987 was generally increasing whilst after this date demand fluctuated. When the demand market experiences a dramatic change, neural network models are unable to catch the change immediately and has a delayed response, see Figure 2 (b). Because of the delay, the BPNN model tends to under estimate these predictions and make it less over estimate of the prediction after the second quarter in 1989 which make it becomes better than the GRNN models. Although all the models have recognized the decreasing trend and have a good prediction accuracy in the second quarter in 1990, their delayed responses have caused an over estimate when the demand decrease. Since prediction largely depends on the historical information, enough indicators should be plugged in so that neural network model could give an accurate prediction to a sudden change is a natural phenomenon.

However, from Figures 2 (a) and (b) it can be seen that all the models are able to recognize this change, in particular, GRNN has the ability to quickly turn the prediction curve to meet the change. This is further evidence that neural networks have the ability to recognize the nonlinear relationship within the data space and are able to catch the change.

3.2 Singapore construction demand forecasting

Hua (1996) applied regression neural networks, back-propagation neural networks (BPNN) to forecast Singapore's construction demand. In this case, twelve indicators were used. There are 74 quarters from the third quarter of 1975 to the fourth quarter of 1993. The BPNN is 12:5:1. 71 cases were used for constructing a model and 3 cases were used for testing. The BPNN model were trained with two randomly selected weights without validation. It is well known that BPNN tends to over-fit so validation or regularisation is an unavoidable important step for training a BPNN model. Without validation, it is hard to believe that a BPNN model has not been over trained. The other limitation of that work is that the performance of the model is measured on a very small set of data which comprise of three cases.

3.3 Singapore residential construction demand forecasting

Hua (1998) has done a comparative study of the accuracy of time series, regression and artificial neural network techniques to forecast residential construction demand in Singapore. In this study, three techniques examined. They are the univariate Box-Jenkins approach, the multiple loglinear regression and artificial neural network. Using seven indicators; (1) building tender price index; (2) bank lending; (3) population; (4) housing stock; (5) National savings; (6) gross fixed capital formation and; (7) unemployment level, Hua (1998) used the three forecasting techniques to identify which techniques can produce accurate demand forecast. To compare the performance of the models and to determine the percentage error of the forecast, percentage error (PE), mean percentage error (MPE) and absolute percentage error (MAPE) have been count. The results are shown in Table 4.

Measures of accuracy (%)	BJ	MLGR	ANN
Percentage error (PE) for 1Q93	−0.26	−3.59	−1.36
Percentage error (PE) for 2Q93	+3.52	−4.24	+0.19
Percentage error (PE) for 3Q93	+0.70	−2.60	+1.80
Percentage error (PE) for 4Q93	−0.18	+17.30	+1.23
Percentage error (PE) for 1Q94	−0.67	−3.98	−0.07
Mean percentage error (MPE) for 1Q93–1Q94	+0.62	+0.58	+0.36
Mean absolute percentage error (MAPE) for 1Q93–1Q94	1.07	6.34	0.93

Table 4: Relative measures of the accuracy of different forecasting technique
Adapted from Bee Hua, Goh; *Forecasting residential construction demand in Singapore: a comparative study of the accuracy of time serries, regression and artificial neural network technique,* (1998)

From Table 4, it can be seen that the ANN technique is the most accurate with lowest MAPE, 0.93. The Box-Jenkins was less accurate with MAPE value of 1.07 while the most inaccurate method is Multiple Loglinear Regression with the highest MAPE of 6.34.

3.4 Private residential construction forecasting in the United States (US).

Aiken *et.al.* (1998) have used artificial neural networks (ANN) to demonstrate the ability of neural networks to accurately predict private residential construction in the United States (US). In this study, they have conducted two training and testing trials. In the first trial, data from July 1949 to January 1972 were used to develop the neural network, and data from January 1972 to January 1980 were used to test the developed model. Two years of data to forecast the next semiannual value for housing starts were used. For example, data from July 1971 to January 1973 were used in the neural net training process to try to predict the housing starts value for July 1973 and the training was continued for approximately 17,000 iterations until the learning error, Mean Absolute Percent Error (MAPE) between the actual and forecasted values was reduced to 6.3%.

In the second trial, data from July 1949 to January 1980 were used to develop the neural network and data from January 1980 through July 1993 were used to test the developed model.

As a result, over the two testing period, the MAPE between the forecasted and actual values was 7.6%. The same training and testing periods also have been conducted using multi-linear regression analysis. From the results, it was found that the regression models forecasts were considerably worse than the neural networks with a MAPE of 22%.

4.0 Conclusion

Studies on previous research show that ANN is the most accurate technique. The error rate in private residential construction forecasting in the study conducted in the US was higher than those conducted in the UK and Singapore. Although the error rate was higher, the US study has utilized five input variables as compare to the study in the UK that only utilized three inputs values. The Singapore study has used the most number of input variables in the study. However, the Singapore study was an annual forecast and tested on a very small sample while the study in the US was a semi-annual forecast. In regression neural networks, it can be concluded that BPNN plays a more important role in the long term forecasting then GRNN since the extrapolation ability of BPNN is better than GRNN. The primary advantage to the GRNN is in its speed in training the network.

REFERENCES

Aiken, M. Wailer, B. and Greer, T. (2001) "Forecasting Construction Demand With A Neural Network". [online]
Available http://www.sbaer.uca.edu/docs/proceedingsII/98dsi064.htm

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

Goh, B. H. (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 pp.261-275.

Hossein Arsham (2001) "Time Series Analysis and Forecasting Techniques-Europe Mirror Site"[online]
Available http://ubmail.ubalt.edu/~harsham/stat-data/opre330Forecast.htm

Klerfors, D. (1998) "Artificial Neural Networks" [online] Available
http://hem.hj.se/~de96klda/NeuralNetworks.htm

Kamal Salih (1976) "Socio-economic Consideration in the Planning, Design and Implementation of Human Settlements and Housing For Lower Income Groups in Malaysia" The Conference of Human Settlements for the Rakyat in the Lower Income Group, Paper No. 5.

Myers, J.M. and Smith, S.C. (2001) "Housing Starts: a Year 2000 Forecast Using Long Term Interest Rates and Total Consumption Spending" [online] Available http://www3.wcu.edu/~mulligan/NCE99-13.htm

Yang and Packer (1997) "Applying Artificial Neural Networks to UK Construction Demand Forecasting (Private Sector)" COBRA '97, pp. 1-16.