2010 International Conference on E-business, Management and Economics
IPEDR vol.3 (2011) © (2011) IACSIT Press, Hong Kong

Performance Comparison of Multiple Linear Regression and Artificial Neural Networks in Predicting Depositor Return of Islamic Bank

Saiful Anwar
Department of Management and Information Systems Science, Nagaoka University of Technology (NUT)
Niigata, Nagaoka, Kamitomioka 1603-1.
Japan. 940-2188
Email: olieanwar@gmail.com

Kenji Watanabe
Department of Disaster Management and Safety,
Nagoya Institute of Technology (NIT)
Gokiso, Showa-ku, Nagoya, Aichi, Japan. 466-8555
Email: watanabe.kenji@nitech.ac.jp

Abstract— **The utilization of artificial neural networks (ANN) in Islamic banking research is rarely reported. Therefore, this paper aims to examine the possibility of ANN utilization for predicting depositor return from time deposit product of Islamic bank. The paper compares the accuracy power of ANN and multiple linear regression (MLR) using ten years monthly data of six macroeconomic variables as independent variables, and the average rate of return from one month time deposit of all Indonesian Islamic banks (RR) as dependent variable. For this purpose, the research employs Eviews software version 5.0 to develop MLR model and Alyuda neuro intelligent software version 2.2 to develop ANN model. As a result, the accuracy power testing using in-sample data for the period of January 2000 to December 2008 indicates the ANN outperforms MLR with 94.62% accuracy rate while MLR achieves only 90.07%. Finally, using out of sample data from January 2009 until April 2010, ANN and MLR achieve averagely 78.59% and 77.44% of accuracy rate, respectively. These evidences demonstrate that ANN model provides more accurate prediction and is quite useful in providing RR prediction information to depositor. (Abstract)**

Keywords: Islamic bank, rate of return, macroeconomic variables, artificial neural networks, multiple linear regression.

I. INTRODUCTION

Mostly, the purpose of research in bank industry is to provide information and analysis for management or policy maker regarding the effect of changes of independent variables to the bank's performance. For example, [12] use regression technique to predict conventional bank failure using financial ratio variables. Additionally, [11] uses ANN to assess credit risk in Islamic bank by examining the effect of macroeconomic changes. However, this research aims to examine the superiority of ANN model for assisting depositors of Islamic bank in making decision. To do so, the paper compares the accuracy power of ANN and MLR model in making prediction.

It is expected that, the ANN can be used as adequate model due to its ability in self learning which impacts to providing a better prediction. As a result, the model will be used by depositors to predict their future return when deposited their money in particular Islamic bank. According to our knowledge, this paper is considered to be the first experiment in Islamic banking and finance subject which compares ANN and MLR model for performing depositor return prediction. For this purpose, this research employs Eviews software version 5.0 to develop MLR model and Alyuda neuro intelligent software version 2.2 to develop ANN model under Windows XP environment.

II. LITERATURE REVIEW

A. The Role of Macroeconomic variables in affecting bank's performance

This study follows the bank failure theory which issued by [8]. The theory said, "Extremely bad management may not prove fatal to a particular bank until economic condition leads to unexpected capital outflows or loan losses". Therefore, this research uses only macroeconomic variables to predict Islamic bank's performance which is represented by depositor return.

Moreover, the macroeconomic variables determine bank's profitability in following ways. Reference [7] reports that market expansion would enable banks to increase profits as represented by strong relationship between money supply and profit. Furthermore, stock indices may lead to a higher growth of the firm, industry and country level. Such condition will give more profit to the banks from financing activities. Meanwhile, [5] informs that inflation positively affects the bank's profits if the revenue accrues from business is larger than the arising of overhead cost due to inflation. Interest, on the other hand, affects majority of funding and financing activities of a bank and later on the bank's profit. Lastly, exchange rate does not affect profits of Islamic bank from foreign exchange trading, as it does to conventional bank, since it is prohibited. It benefits Islamic bank through its impact on the fluctuation of the price of goods that affect business trading and market.

B. The uniqueness of Islamic bank deposit

One of the main sources of fund in Islamic banks is called *Mudharabah* time deposit. The banks use the fund to finance the business which is not prohibited by Islamic law and lately share the profit with depositors [9]. Unlike interest that provides fix regular revenue, the return of *mudharabah* time deposit which is represented by monthly rate of return is uncertain. Additionally, the rate is varied among Islamic banks since it depends on the bank's profitability and pre-agreed profit and loss sharing (PLS) ratio. Every month, the Islamic banks publish the RR report to assist depositors comparing the current rate of return with current market interest rate.

C. Artificial Neural Networks (ANN) Model

Actually, ANN is a part of artificial intelligence that is powerful to solve problem especially with regard to pattern classification and recognition. It is a computational model which the structure and function mimic biological neuron in the human brain. Moreover, ANN consists of a group of artificial neurons, which are interconnected. Every single neuron processes information (receiving input and delivering output) using a special algorithm function.

Technically, the process of ANN is briefly explained as following. Initially, there is a neuron j which has a certain number of inputs ($x_1, x_2, x_3, \ldots x_j$) and single output ($y_j$). Each input has a weight ($w_{1j}, w_{2j}, w_{3j}, \ldots w_{ij}$) as an indicator of importance of the incoming signal into neuron (j). The net value (u_j) of the neuron is then calculated with the sum of all the input value that multiplied by their specific weight. Further, with reference to threshold value (t_j) and activation function, the neuron (j) determines output value (y_j) will be sent as output to other neuron. Each neuron has its own unique threshold value (t_j). If the net value (u_j) is greater than the threshold (t_j), the neuron (j) will send output (y_j) to other neurons. In addition, activation function is a function used to transform the activation level of a unit (neuron) to an output signal. Currently, sigmoid and logistic are the most popular activation function used. All the processes are depicted in fig.1.

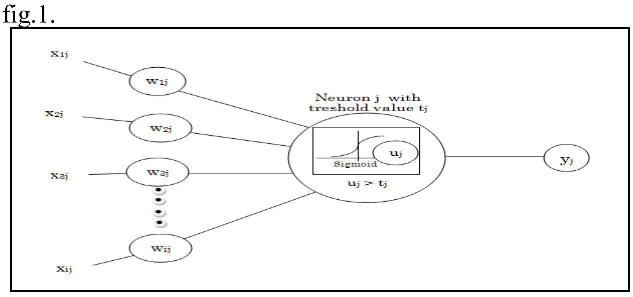

Figure 1. A model of neuron.

However, a single neuron is not useful to solve the problems. The research needs to combine some neurons into multilayer structured which so called neural networks to have the power for answering pattern classification and recognition problems. For this reason, this research employs a multi layer feed-forward networks which is the most common type of neural network currently in use. The multi layer feed-forward networks comprises of input layer, hidden layer and output layer.

Specifically, the input layer is a layer that directly connected to outside information. All data in the input layer will be feed-forward to hidden layer as the next layer. Meanwhile, the hidden layer functions as feature detectors of input signals and releases them to output layer. Finally, the output layer is considered as a collector of the features detected and producer of the response. In the networks, output from output layer is the function of the linear combination of hidden unit's activation; the hidden unit's activation function is a non-linear function of the weighted sum of inputs. Mathematically, the model can be written as in the following:

$$y = f(x,\theta) + \varepsilon \qquad (1)$$

Where x is the vector of explanatory variables, θ is weights vector (parameters) and ε is the random error component. Then, Equation (2) is the unknown function for estimation and prediction from the available data. As such, the model can be formulated as:

$$Y = f\left[v_0 + \sum_{j=1}^{m} h\left(\lambda_j + \sum_{i=1}^{n} x_i w_{ij} \right) v_j \right] \qquad (2)$$

Where:
Y = network output
f = output layer activation function
v_0= output bias
m= number of hidden units
h= hidden layer activation function
λ_j= hidden unit biases (j = 1,. . . ,m)
n= number of input units
x_i= inputs vector (i = 1,. . . ,n)
w_{ij}= weight from input unit i to hidden unit j
v_j= weights from hidden unit j to output (j = 1,. . . ,m)

D. Multiple linear regression (MLR) models

MLR is a technique used for modeling the relationship among more than two variables linearly. It is considered as the most common technique used in banking and finance area, especially for examining the financial performance in Islamic bank. For example [4] uses MLR to measure profitability of two Sudanese Islamic bank.

In building the model, certainly, the restrictive assumptions of MLR should be carefully conformed in building a valid model, such as; linearity, normality, independence of error terms, etc. The model can be simply formulated as shown in equation 3.

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \ldots + \beta_{p-1} X_{i,p-1} + \varepsilon_i \quad (3)$$

Where:
Y_i = Dependent variable
X_i= Independent variables
β = Unknown coefficient
ε_i = Error term
i = 1,2,..n

E. The superiority of ANN in doing prediction.

Accumulated studies have shown that ANN is better than MLR in doing prediction. Reference [14] reported that ANN outperforms MLR in predicting housing value. Moreover, [15] reported the same evidence that ANN is the more powerful tool in predicting bank performance by comparing the mean square error (MSE) and mean square prediction error of both model.

Moreover, [2] confirmed the superiority of ANN against multi discriminant analysis model (MDA) in predicting bankruptcy of 128 firms. In detail, ANN achieves correct classification accuracy in the range of 77.8% to 81.5%, while MDA's accuracy is in the range of 59.3% to 70.4%. Additionally, [3] review 89 studies on corporate bankruptcy prediction. They found that ANN achieves 88% accuracy rate averagely while traditional statistical models achieve 84% accuracy rate.

Although ANN model has demonstrated some successes in this area, there have been little work done on forecasting in Islamic banking and finance research using ANN model, such as; [1], and [11]

III. DATA AND METHODOLOGY

A sample data set consisting six macroeconomic variables for the period of January 2000 – April 2010 are used as independent variables. Such as; exchange rate of US dollar against Indonesian rupiah (EXCH), stock indices (STIN), money supply (represented by M1), inflation rate (INFR), market interest rate of one month time deposit (INTR), and central bank's interest rate (BIRT). Meanwhile, the data of average return of 1 month *mudharabah* time deposit of all Indonesian Islamic banks (RR) from the same period are used as dependent variable. All data are collected from Indonesian central bank (Bank of Indonesia).

Initially, the relationship between independent variables and dependent variable for the period of January 2000 – December 2008 are examined separately using MLR and ANN model. The result will indicate the significant variables among the six macroeconomic variables in determining RR fluctuation. Such information is important as preliminary understanding how ANN and MLR doing their assignment by comparing it with what theory said.

A. Working with ANN

In the beginning, all data are preprocessed to simply convert the input data into a new version for three reasons [6]. (1) To ensure the size of data reflect the importance level in determining the output. (2) To facilitate the random initialization of weights before training the networks. (3) To normalized all data to avoid different measurement due to different unit of input. Next, Alyuda provides an exhaustive search feature to design the neural networks architecture. As a result, this research use $N^{(8\text{-}6\text{-}1)}$ for learning and testing process which will be conducted later on.

The configurations used for learning process are as following: (1) The logistic function is selected for all neuron. (2) The sum-of-squared errors are selected to minimize the output errors. These are summation of squared differences between the actual values and model's output. (3) The networks outputs are set up between 0 and 1 due to logistic activation function used.

Furthermore, the ANN is trained with specific condition to avoid over fitting such as; using back propagation as learning algorithm, the learning and momentum rates are set at 0.1, and for completeness, the process should be stop when mean squared error reduces by less than 0.000001 or the model completes 20,000 iterations, whichever condition occur first. As a result, the process provides information about significant rate of each independent variable. The ANN model shows that INTR, EXCH, STIN contribute 27.12%, 26.74%, and 22.15% in explaining RR volatility, respectively. Meanwhile, other variables such as BIRT, M1 and INFR give lesser contribution for about 13.73%, 9.76% and 0.47%, respectively.

B. Working with MLR

In the beginning, all restrictive assumptions of linear regression have been trying to be fulfilled. To do so, initially, the research runs six examinations to search the best MLR model with no autocorrelation problem. However, after tested using Durbin-Watson test, all models are indicated to have autocorrelation problem. Therefore, the model number 1 is chosen based on R^2 and AIC parameter. The results of comparison can be found in table 1.

TABLE 1. PERFORMANCE OF SIX REGRESSION MODELS

No	Model	Variables	Significant Variables	R^2	AIC	Durbin-Watson
1	MLR	Intr, Infr, Birt, M1, Exch, Stin	Exch, M1, Intr	0.471056	3.375658	0.44221
2	MLR	Intr, M1, Exch	Intr, M1	0.436401	3.383562	0.390432
3	MLR	Intr	C, Intr	0.373035	3.453073	0.339401
4	ARCH (1) GARCH (1)	Intr, Infr, Birt, M1, Exch, Stin	C, Intr, Infr, Birt, M1, Exch, Stin	0.41696	2.578455	0.371109
5	ARCH (1) GARCH (1)	Intr, M1, Exch	C, Intr, M1, Exch	0.418271	2.573589	0.365073
6	ARCH (1) GARCH (1)	Intr	C, Intr	0.354912	2.596975	0.330731

Accordingly, the following MLR equation is proposed to examine the contribution of each independent variable in explaining the volatility of RR (equation 4).

$$RR_i=\beta_0+\beta_1 EXCH_{i1}+\beta_2 STIN_{i2}+\beta_3 M1_{i3}+\beta_4 INFR_{i4}+\beta_5 INTR_{i5}+\beta_6 BIRT_{i6}+\varepsilon_i \quad (4)$$

Furthermore, the 124 data are collected. The first 108 data for period of January 2000 – December 2008 are used to build the model. Meanwhile, the other 16 data for period of January 2009 – April 2010 are used to test the model as out of sample prediction. Actually, [10] recommended 30% of out of sample data to be used for testing the model. In this research, the out of sample testing uses 21% of total of initial sample size due to the limitation in collecting data.

Next, the least square method is used to do regression for the proposed MLR equation. The results of least squares regression can be seen in table 2.

Dependent Variable: RR
Method: Least Squares
Date: 08/09/10 Time: 20:20
Sample: 2000M01 2008M12
Included observations: 108
White Heteroskedasticity-Consistent Standard Errors & Covariance

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-1.832936	3.044989	-0.601951	0.5486
EXCH	0.000856	0.000365	2.345713	0.0209
STIN	0.000576	0.000386	1.492244	0.1388
M1	-6.17E-06	2.08E-06	-2.967298	0.0038
INFR	-0.089130	0.313142	-0.284630	0.7765
INTR	0.631343	0.272674	2.315377	0.0226
BIRT	-0.319181	0.235764	-1.353819	0.1788

R-squared	0.471056	Mean dependent var	7.845669
Adjusted R-squared	0.439633	S.D. dependent var	1.694119
S.E. of regression	1.268177	Akaike info criterion	3.375658
Sum squared resid	162.4357	Schwarz criterion	3.549499
Log likelihood	-175.2855	F-statistic	14.99107
Durbin-Watson stat	0.442210	Prob(F-statistic)	0.000000

TABLE 2. LEAST SQUARES REGRESSION RESULTS

Furthermore, the heteroscedasticity issue from the proposed model has been accomplished using white method which is facilitated by Eviews software. This method is so-called heteroscedasticity-corrected variances.

From the first order regression output, it was found that only three independent variables are significant in affecting RR volatility namely INTR, M1 and EXCH. This is indicated by value of t-statistic which are 2.31, -2.96 and 2.34 respectively. The value indicates should an independent variable be included in a model for 95% confident level. Or in other words, the research tolerates only a 5% chance that a particular independent variable does not belong in a model. Accordingly, a value of t-statistic which is greater than 1.98 (if the coefficient is positive) or less than -1.98 (if the coefficient is negative) will be considered to be significant statistically.

Moreover, all independent variables in the equation 5 are able to explain all variations of RR for about 47.1% as indicated by the R^2 value. Finally, the value of Akaike's information criterion (AIC) which is used to measure the goodness of fit of an estimated statistical model indicates the model is a good model (AIC value is 3.37).

$$\begin{aligned} RR &= -1.83 + 0.0008*EXCH + 0.0005*STIN - 6.17e^{-006}*M1 \\ (t) &\quad (0.60) \quad (2.34) \qquad (1.49) \qquad (-2.96) \\ &\quad -0.089*INFR + 0.631*INTR - 0.31*BIRT \\ &\quad (-0.28) \qquad (2.31) \qquad (-1.35) \end{aligned} \tag{5}$$

Additionally, all the efforts to conform to restrictive assumptions of MLR model show that developing regression model using macroeconomic variable is not an easy task since macroeconomic data are characterized as non linear time series data which violate the assumptions of linear regression [13].

C. Accuracy Performance Comparison

The research uses three methods to analyze and compare the performance of each model, such as; (1) Comparing graph of actual and predicted RR using in sample data. (2) Comparing R^2 and AIC statistical parameters. (3) Calculating the accuracy rate to measure the accurateness in predicting RR using out of sample data according to equation 6.

$$\textit{Accuracy power} = 100\% - \%\,\textit{Error of prediction} \tag{6}$$

IV. RESULTS

Fig. 2 shows the comparative performance between ANN and MLR model using in-sample data.

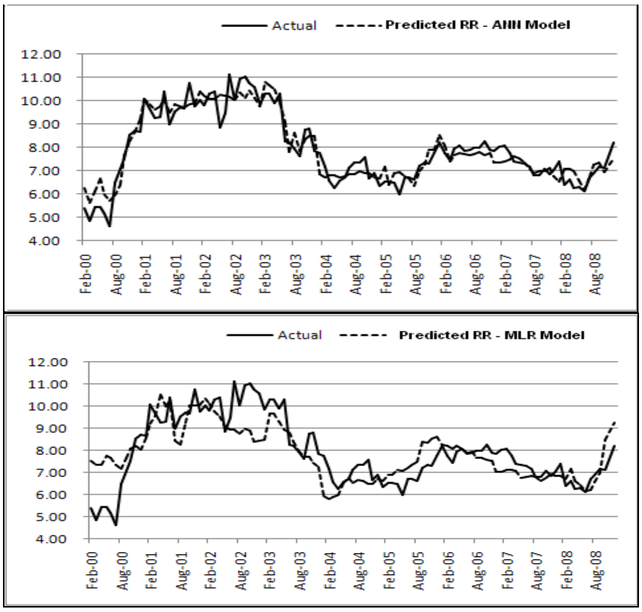

Figure 2. Graph of actual vs. predicted RR.

The figure suggests that ANN outperforms MLR model since the predicted line of ANN locates closer to actual line rather than MLR model. This is also confirmed by the accuracy rate calculation using in sample data which ANN model achieves 94.62% accuracy rate in average. It is higher than MLR model which achieves 90.07% accuracy rate.

Furthermore, using statistical criteria as shown in table 3, it supports the previous finding indicated by R^2 of ANN which is higher than MLR model. On the contrary, the AIC of ANN is much lower than MLR model. Both conditions demonstrate that ANN outperforms MLR.

TABLE 3. PERFORMANCE COMPARISON BETWEEN ANN AND MLR MODEL USING STATISTICAL CRITERION

Statistical criterion	ANN Model	MLR Model
R^2	0.878691	0.471056
AIC	-283.78632	3.375658

Finally, the out of sample data prediction is carried out using formula 6 as shown in table 4. It demonstrates that the performance of $N^{(8\text{-}6\text{-}1)}$ is better than MLR model in predicting RR for the period of January 2009 to April 2010 with 78.59% accuracy rate in average.

TABLE 4. ACCURACY RATE CALCULATION

Month	Actual RR	Predicted RR (MLR)	Predicted RR (ANN)	Accuracy rate (MLR)	Accuracy rate (ANN)
Jan-09	8.03	9.48	7.56	81.94%	94.13%
Feb-09	7.73	9.96	6.93	71.09%	89.64%
Mar-09	8.49	9.40	6.97	89.22%	82.04%
Apr-09	8.32	8.77	7.67	94.56%	92.14%
May-09	8.45	8.47	7.58	99.78%	89.74%
Jun-09	10.77	8.28	7.53	76.85%	69.87%
Jul-09	10.26	8.03	7.44	78.22%	72.56%
Aug-09	9.24	7.95	7.14	86.02%	77.30%
Sep-09	9.20	7.36	6.97	79.99%	75.74%
Oct-09	9.38	7.43	7.36	79.25%	78.47%
Nov-09	9.09	7.18	7.41	78.95%	81.55%
Dec-09	9.06	6.95	7.25	76.66%	80.04%
Jan-10	5.71	7.65	7.18	65.96%	74.21%
Feb-10	5.39	7.67	7.39	57.78%	62.88%
Mar-10	5.97	7.71	7.57	70.92%	73.21%
Apr-10	5.86	8.68	7.97	51.84%	63.99%
			Average accuracy rate	77.44%	78.59%

V. CONCLUSIONS

From the study, we may conclude that ANN model can be used for predicting RR due to its better performance compared with MLR model. Furthermore, the utilization of MLR in modeling bank's performance based on macroeconomic variables is not as easy as ANN. The research needs extra effort to conform to the underlying assumption of linear regression model. On the contrary, with Alyuda neuro intelligent software, such task is quietly simple. The ANN's ability in self learning benefits the performance measurement model which no needs such assumptions as MLR does. In fact, the ANN provides better result in making prediction.

Additionally, the ANN model provides information that market interest rate (INTR) is the leading indicator for predicting RR. Thus, the information will benefit depositor by signaling the future RR will be received. When the market interest rate tends to rise, then the RR also tends to do so or otherwise. Besides giving benefit for depositors and Islamic banks, the utilization of RR prediction model will also promote the expansion of Islamic bank industry. Because, the prediction information will provide more options for depositors to find other Islamic bank which probably provide better return. As a result, it will keep depositor funds stay longer in Islamic bank industry before shifting to conventional bank.

This research somehow has its limitation. It uses the MLR model which is still violated the autocorrelation assumption. Thus, this requires the needs to understand more sophisticated statistical technique to solve the problem which is out of the scope of this research.

ACKNOWLEDGMENT

This research is financed by Ministry of Education, Republic of Indonesia. We thank to Prof. Dr. H. Fathurrahman Djamil from STIE Ahmad Dahlan, Jakarta and Prof. Yoshiki Mikami from Nagaoka University of Technology for the very kind support.

REFERENCES

[1] Al-Osaimy, M.H. (1998), "A Neural Networks System for Predicting Islamic Banks Performance", Journal of King Abdul Azis University: Islamic Economic & Administration, vol.11, no.1, pp.33-46.

[2] Atiya, A.F. (2001), "Bankruptcy Prediction for Credit Risk Using Neural Networks: A survey and New results", IEEE Transaction on Neural Networks, vol.12, no.4, pp. 929-935.

[3] Azis, M.A. and Dar, H.A. (2004), "Predicting Corporate Bankruptcy: Whither do We Stand?", Unpublished working paper, Department of Economics, Loughborough University, UK.

[4] Bashir, A.H.M. (1999), "Risk and Profitability Measures in Islamic Banks: The Case of Two Sudanese Banks," Islamic Economic Studies, vol.6, no.2, pp. 1–24.

[5] Bashir, M. (2000), "Determinants of Profitability and Rate of Return Margins in Islamic Banks: Some Evidence from the Middle East", Economic Research Forum. October 26-29. Amman, Jordan. Available at: http://www.erf.org.eg/CMS/getFile.php?id=753. (accessed 20 December 2009).

[6] Bishop, M., C. (1995), Neural networks for pattern recognition 1st ed, Oxford University Press Inc, New York.

[7] Bourke, P. (1989), "Concentration and Other Determinants of Bank Profitability in Europa, North America and Australia", Journal of Banking and Finance, vol.13, pp. 65-67.

[8] Gambs, C.M. (1977), "Bank Failures-An Historical Perspective", Federal Reserve Bank of Kansas City Monthly Review, June, pp. 10-20.

[9] Ismal, Rifki. (2009), "Industrial Analysis of Liquidity Risk Management in Islamic Bank", Journal of Islamic and Finance, vol.26, no. 2, 2009, International Association of Islamic Bank, Karachi, Pakistan.

[10] Kutner, M., Nachtsheim C. J., & Peter, J. (2004). Applied Linear Regression Models, McGraw Hill/Irwin Series, 4th Edition.

[11] Maximilian, J.B. Hall., Muljawan D, Suprayogi. and Moorena. L. (2008), "Using the Artificial Neural Network (ANN) to Assess Bank Credit Risk: A Case Study of Indonesia", Applied Financial Economics, vol.19, pp. 1825 – 1846.

[12] Meyer, P.A. and Pifer, H.W. (1970), "Prediction of Bank Failure", Journal of Finance, vol.25, pp. 853-868.

[13] Moody, J. 2005. "Economic Forecasting: Challenges and Neural Network Solutions", International Symposium on Artificial Neural Networks. December 18-20. Hsinchu, Taiwan, available at: http://neuron-ai.tuke.sk/~daliman/Esej/2/moody95economic.pdf. (accessed 20 December 2009)

[14] Nguyen, N., & Cripps, A. (2001). "Predicting Housing Value: A Comparison of Multiple Linear Regression Analysis and Artificial Neural Networks", Journal of Real Estate Research, vol.22, no.3, pp. 313-336.

[15] Nor Mazlina Abu Bakar and Izah Mohd Tahir (2009), "Applying Multiple Linear Regression and Neural Network to Predict Bank Performance", International Business Research, vol 2, no 4, pp. 176-181.