Stock Market Forecasting : Comparison between Artificial Neural Networks and Arch Models

Nitin Merh*

Abstract

Data mining is the process of searching and analyzing large quantities of data for finding out meaningful patterns and rules. Artificial Neural Network (ANN) is one of the tools of data mining which is becoming very popular in forecasting the future values. Some of the areas where it is used are banking, medicine, retailing and fraud detection. In finance, artificial neural network is used in various disciplines including stock market forecasting. In the stock market time series, due to high volatility, it is very important to choose a model which reads volatility and forecasts the future values considering volatility as one of the major attributes for forecasting. In this paper, an attempt is made to develop two models - one using feed forward back propagation Artificial Neural Network and the other using Autoregressive Conditional Heteroskedasticity (ARCH) technique for forecasting stock market returns. Various parameters which are considered for the design of optimal ANN model development are input and output data normalization, transfer function and neuron/s at input, hidden and output layers, number of hidden layers, values with respect to momentum, learning rate and error tolerance. Simulations have been done using prices of daily close of Sensex. Stock market returns are chosen as input data and output is the forecasted return. Simulations of the Model have been done using MATLAB® 6.1.0.450 and EViews 4.1. Convergence and performance of models have been evaluated on the basis of the simulation results. Performance evaluation is done on the basis of the errors calculated between the actual and predicted values.

Keywords : Artificial Neural Network (ANN), Autoregressive Conditional Heteroskedasticity (ARCH), Feed Forward Back Propagation.

Received : 2011. 12. 02. Final Acceptance : 2012. 01. 27.

* Associate Professor, Computer Science and Engineering, Institute of Engineering and Technology, JK Lakshmipat University, Jaipur, e-mail : n_merh@rediffmail.com

1. Introduction

Neural network are a class of generalized non-linear non-parametric model inspired by studies of human brain. Their main advantage is that they can approximate any non-linear function to an arbitrary degree of accuracy with suitable number of hidden units [Hornik at al., 1989].

Financial forecasting is of significant practical interest. Since neural networks can mine valuable information from mass of historical information and can be efficiently used in financial areas, so the applications of neural networks to financial forecasting have been very popular over the last few years [Zhang and Li, 2004]. In finance particularly it has been applied to number of applications including development of trading model, investment selection, loan assessment, portfolio optimization, fraud detection, bankruptcy prediction, real estate assessment and so on [Zhang and Zhou, 2004]. Although many research show that neural networks is an excellent tool of data mining for forecasting the financial data but there are certain drawbacks associated with it like over fitting and poor explanation capabilities. In forecasting financial times series which is highly volatile it is quite important to find the volatility in the series under study and find techniques that can forecast such time series. ARCH and GARCH are the models which can predict changes in the volatility of the financial times series.

According to Efficient Market Hypothesis, all available market information are factored immediately into the formation of stock price and therefore, the best predictor of future stock price is the latest available price stock market [Samantha and Bordoloi, 2005; Fama, 1965, 1970]. In this paper an attempt is made to develop models for forecasting fluctuations in stock market series. This paper describes and compares Artificial Neural Network and Autoregressive Conditional Heteroskedasticity (ARCH) models.

2. Financial Data Mining

A time series is a sequence of vectors $Y_t = (y_{t-n}, \cdots\cdots, y_{t-i}, y_{t-2}, y_{t-1})$, where y_{t-i} represents past value that change with time. In stock prices the data are typical example of time series. In the current paper models will forecast the future trends on the basis of the past values. Log differenced values of Sensex is taken as the input variable used for forecasting volatility in financial markets. A model needs to be developed that will find a function F = R^i to R^o, where R^i is the input space containing i input values and R^o is the output space containing o output values.

3. Methodology Used

3.1 Data

Input data for both the models consists of data collected from August 01, 2003 to October 19, 2011 of Sensex (BSE 30). Data was downloaded from in.finance.yahoo.com on October 21, 2011. In this paper, stock market returns are defined as log returns at time t, return R_t is calculated by the following equation :

$$R_t = Ln\left(\frac{CV_t}{CV_{t-1}}\right) = LnCV_t - LnCV_{t-1}$$

Where CV_t and CV_{t-1} are the closing value of Sensex at time (day) t and t-1 respectively. Sensex return has been taken as input dataset and output being predicted volatility of Sensex. In both the models (ANN and ARCH/GARCH) available data is divided into two parts one for the training also known as in-sample data and other one is used for testing or out-of-sample/ hold-out sample.

3.2 Methods for Pre-Processing and Post-Processing for ANN Model are as Follows :

Input stock market data is collected from in.finance.yahoo.com (Date of Downloading : October 21, 2011). When data is loaded in the ANN (Artificial Neural Network), it must be preprocessed from its numeric range into the numeric range that the ANN can deal with efficiently. In this process, proper transformation of data simplifies the process of learning and may improve the generalisability of the learned results [Kim and Lee, 2004].

The preferences of range to which input and targets are normalized mainly depend on the transfer or activation function. In case of logistic functions rage is [0, 1] and for hyperbolic function limit is between [-1, 1] several researchers scale the data between [0.2, 0.8] or [0.1, 0.9] based on the fact that the non linear activation function usually have asymptomatic limits [Guoqiang Zhang at el., 1998; Tang and Fishwick, 1993; Azoff, 1994]. In the current work as at the input layer linear transfer is taken hence firstly [-1.0, +1.0] limits were taken but as the network was not trained even after 500 epochs so the data has been scaled between range [0.1, 0.9]. So before providing the data to the network, the data is to be preprocessed or scaled. Once the input scaled data is processed through the artificial neural network and the output obtained, it is de-scaled. Methods for preprocessing and post processing data are described as follows [Merh, Saxena, and Pardasani, 2008]:

(1) Procedure for Scaling of Time

$$F = \frac{y_{max} - y_{min}}{x_{max} - x_{min}}$$

Where $x_{min}\, to\, x_{max}$ are scaled from $y_{min}\, to\, x_{max}$.

$$y_{max} = 0.9 \text{ and } y_{min} = 0.1$$

Where values of $y_{max} \text{ and } y_{min}$ depends on which transfer function is used in the ANN model

$$offset = (y_{min} - (F) \times x_{min})$$

The input data is preprocessed by using the following scaling function and the output data is again processed using reverse scaling.

F × X + Offset where X is the data which is to be processed.

(2) Procedure for De-scaling of Time Series

After processing the input value through the artificial neural network model proposed in the paper data at the output layer is again scaled in the original numeric range using the following method :

$$y_{max} = 0.9 \text{ and } y_{min} = 0.1$$

$$F = \frac{y_{max} - y_{min}}{x_{max} - x_{min}}$$

Where A is the scaled value which to be converted in the original numeric value

$$val = \frac{A - y_{min}}{F + x_{min}}$$

3.3 Artificial Neural Network

The feed-forward back propagation architecture has been used as the forecasting model. The model is most widely and commonly used in various applications. Multilayer neural network is used for the prediction of the next sample point in time series databases [Takaho et al., 2002].

(1) Neural Network Architecture

In the current paper Artificial Neural Network architecture <Figure 1> consists of an input layer, one hidden layer and an output layer; index range for the above three layers is i = 1, 2, ⋯, n, h = 1, 2, ⋯, m and j = 1, 2, ⋯, p, respectively and for input data set index range is u = 1, 2, ⋯, n. In the model, selection of number of hidden layer is random. Linear transfer function is used at the input layer which receives the external real input and produces the same output. While for the output layer and the hidden layer, log sigmoid function is used as the transfer function.

More specifically, for the input layer

$$x_i^k = \sum_{u=0}^{n} w_{ui} p_i^k,$$

$$i = 1, 2, 3, \cdots \cdots n$$

$$f(x_i^k) = x_i^k, \qquad i = 1, 2, \cdots \cdots n$$

Where x_i^K is i$^{\text{th}}$ component of the input vector of X_k

For hidden layer input will be

$$z_h^k = \sum_{i=0}^{n} w_{ih}^k x_i^k, \qquad h = 1, 2, \cdots \cdots m$$

Output generated by hidden layer will be

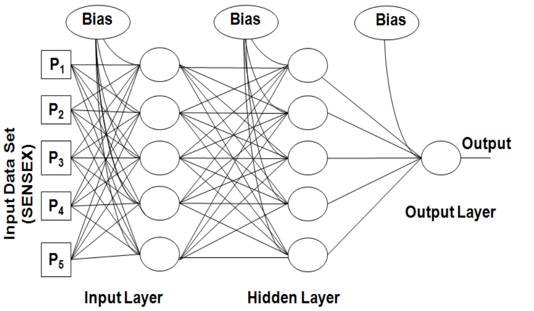
$$f(z_h^k) = \frac{1}{1 + e^{-z_h^k}}, \qquad h = 1, 2, \cdots \cdots m$$

While for the output layer input will be

$$y_j^k = \sum_{h=0}^{m} w_{hj}^k f(z_h^k), \qquad j = 1, 2, \cdots \cdots p$$

Output generated at the output layer will be as follows :

$$f(y_j^k) = \frac{1}{1 + e^{-y_k^k}}, \qquad j = 1, 2, \cdots \cdots p$$



Where P1, P2, P3, P4 and P5 are the five days close prices of SENSEX

<Figure 1> Three-Layer Feed-Forward Neural Network with One Hidden Layer

<Table 1> Data Set (Range) for In-Sample and Out-of Sample

Total Sample Size	Training Set(In-sample data)	Test Set(Out-of-sample)
2036 August 01, 2003 to October 19, 2011. Past eight years data	2000 data values August 02, 2003 to August 26, 2011.	30 data values September 07, 2011 to October 19, 2011.

In the current architecture of the Artificial Neural Network model discussed above biases are included with weights having index 0, will be treated as bias for the given layer and neuron i.e., w_{oi}^{k}, $w_{oh}^{k}$and w_{oj}^{k} are the biases for the input layer, hidden layer and output layer respectively. The stock market returns from August 01, 2003 to October 19, 2011 and the output produced will be the forecasted volatility of the stock market time series under study.

(2) Learning process

In supervised learning, information on errors is filtered back through the system and is used to adjust the connections between the layers and improves performance. Back propagation algorithm used for the training of the neural network, which uses supervised learning in which finite number of pairs of input patterns X_p and target output patterns T_p, in the form (X_1, T_1), (X_2, T_2), …… (Xp, Tp) are initialized.

While training the artificial neural network, the model adjusts the weights continuously by comparing the output generated with the target provided until the error minimizes or reaches the specified value. Here the error is the difference between the output value and the target value. The procedure of repeatedly presenting a set of input patterns and corresponding target patterns and adjusting the value of connection weights iteratively so as to minimize the average squared output derivation error function over all training patterns was used for training the neural network in back propagation algorithm [Werbos, 1974; Rumelhart et al., 1986].

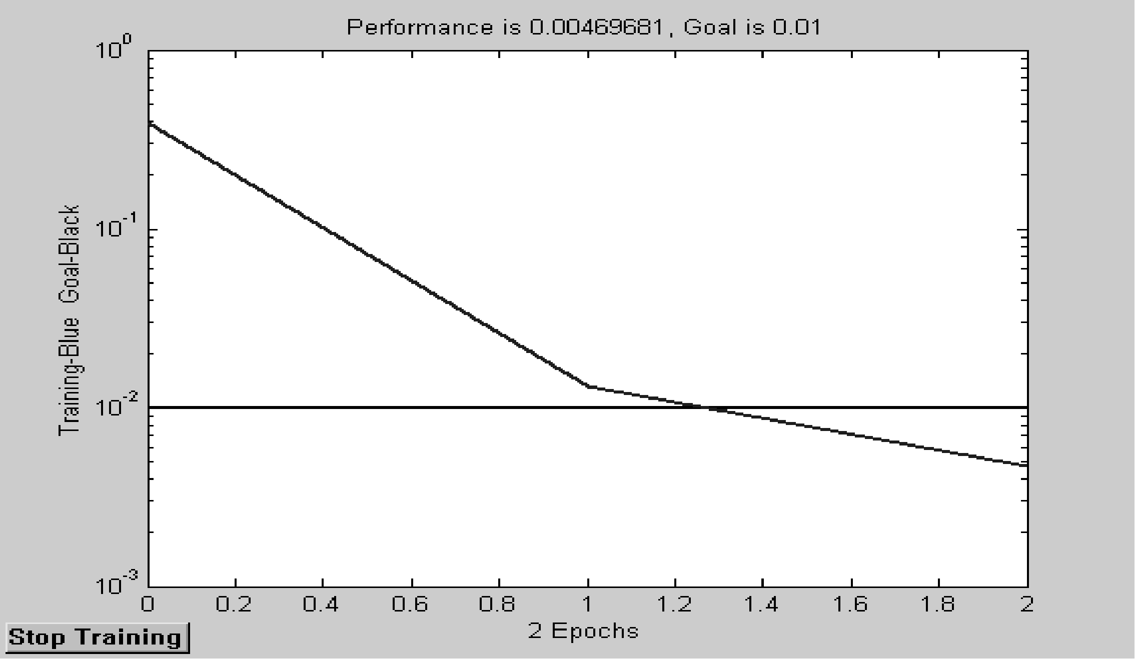
Details for training the network are as follows :

(3) Design

The next step is to evaluate the controlled parameters that are important for the development of the overall network structure. These factors include learning rate, momentum factor, initial weight, error function, bias function, target output, etc. High learning rate is good for less complex data and low learning rate with high momentum should be used for more complex data series [Tang et al., 1991; Zhang, Patuwo, and Hu, 1998]. For evaluating performance based on training, the following values were assumed : Training Learning Rate (η) = 0.01, Momentum (α) = 0.5, Epochs = 1000, and Tolerance Error = 0.01 (10^{-2}). ANN model is not able to train the network with above set of randomly chosen parameters and architecture when error tolerance is taken as 0.001 (10^{-3}) or higher.

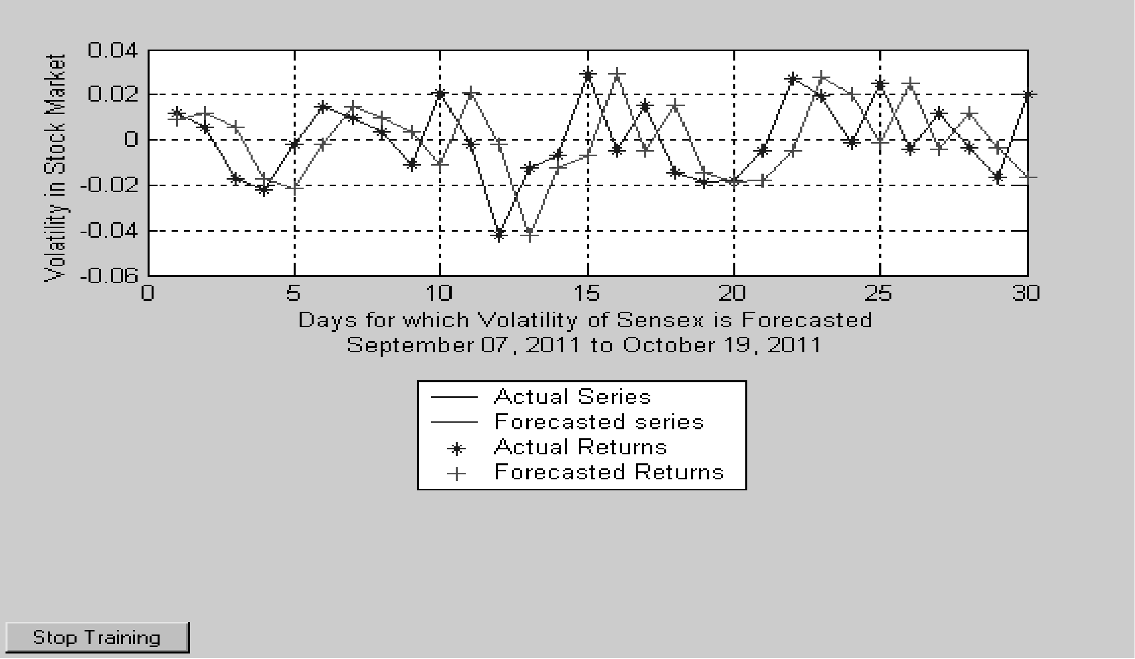
(4) Performance

The performance of the model has been evaluated by calculating errors. In the process of training, the target is the closing price of the next day. Comparison is made between the actual and predicted closing derived from the ANN model (see <Figure 2> and <Figure 3>).



Note : TRAINLM, Epoch 0/1000, MSE 0.392293/0.01, Gradient 986.04/1e-010
TRAINLM, Epoch 2/1000, MSE 0.00469681/0.01, Gradient 0.00628517/1e-010
TRAINLM, Performance goal met.

<Figure 2> Performance Evaluation Based on Training of ANN Model



<Figure 3> Comparison between Actual and Predicted Closing Prices of Sensex

4. Autoregressive Conditional Heteroskedasticity

Stock market data or financial data can be stationary or non-stationary. In such data set fluctuations are observed over time. Stock Volatility is defined as a standard deviation or conditional variance of stock returns that is not directly observable. Earlier it was assumed that volatility to be constant. Increasingly however, econometricians and forecasters are being asked to forecast and analyze the size of the errors of the model. Here we are taking the case of volatility and solution is ARCH/GARCH model. It was observed [Varma, 1999] that using daily data from 1990 to 1998 of Indian stock index (Nifty) and applying GARCH(1, 1) model with generalized error distribution performs better than the other estimators of volatility [Srinivasan, 2011].

ARCH Model is defined as :

$$R_t = c + bX_t + \varepsilon_t$$
$$\varepsilon_t = z_t \cdot \sqrt{h_t} \; where \; z_t \sim IIDN(0,1)$$
$$h_t = \omega + \alpha\varepsilon_{t-1}^2$$

GARCH Model is defined as follows :

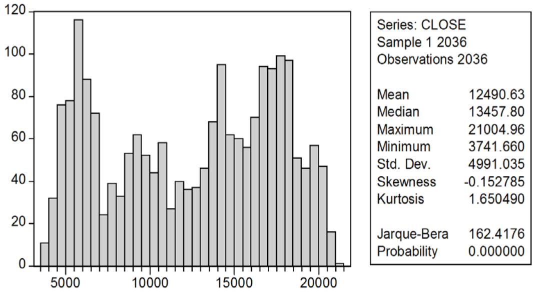
$$R_t = c + \rho R_{t-1} + \varepsilon_t$$
$$\varepsilon_t = z_t \cdot \sqrt{h_t} \; where \; z_t \sim N(0,1)$$
$$h_t = \omega + \alpha\varepsilon_{t-1}^2 + \beta h_{t-1}$$

Where c is the constant, ω is the market volatility, constants, α is news coefficient, ε is unexpected shock, z_t and h_t are parts of variance.

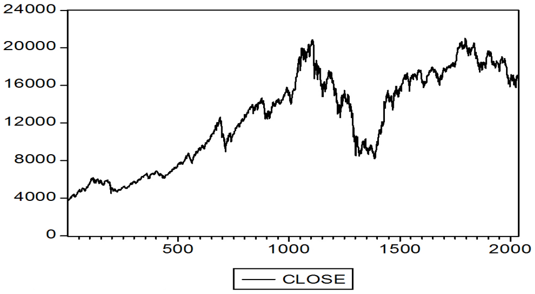
In the current paper estimation equation is as follows :

RETURN = C(1) + C(2) × RETURN(-1) where C(1) and C(2) are constants

ARCH/GARCH model has been used for forecasting volatility in the stock market data. In the development of the model first step is to verify whether time series is stationary or not, it is done with the help of Dickey- Fuller unit root test. When series statistics test and unit root test was done on the close series it was found (<Figure 4> and <Figure 5>) that neither the mean was zero nor values of t-Statistic (-1.585162) and Prob. (0.4900) indicates that times series under study is not stationary.



<Figure 4> Sequence Graph of Daily Close Price of Sensex



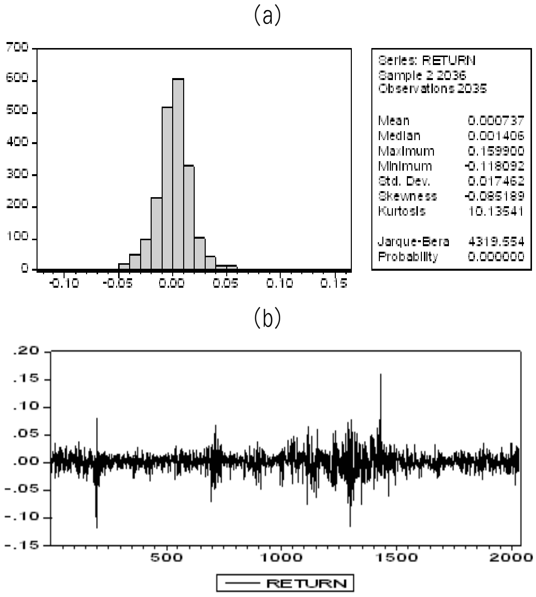
<Figure 5> Series Statistics Test on Input Times Series

After taking first difference of the time series, values of t-Statistic (-41.63303), Prob. Values (0.0000) and series statistics (Histogram and Stat) are calculated which indicates that time series under study after first difference is stationary (<Figure 6(a)>, <Figure 6(b)>). As discussed earlier stock market returns are defines as log returns which is defined as first difference of natural log of the times series under study. <Figure 6(b)> indicates high volatility in clusters which indicates heteroskedasticity.

After performing number of iterations with different orders of ARCH and GARCH it was found that the GARCH (1, 1) has best R^2 value even though R^2 is very low along with the residual test, t values and prob. values. Firstly, model selection and verification is done with the in-sample (2000 from 2036 samples) data ranging from August 01, 2003 to August 25, 2011 and out-of-sample data contains 36 values August 26, 2011 to October 19, 2011 from which only 30 days values from September 07, 2011 to October 19, 2011 are taken as the results are to be compared with the results generated from ANN (5-5-1) model. ARCH LM test indicates that there is no there is evidence of autoregressive conditional heteroskedasticity (ARCH) in the residuals. Correlogram Squared residual test and histogram Normality test indicates occurrence of shocks in some cluster of the series.

Result generated after applying GARCH (1, 1) on in-sample 2000 data return = dlog(close) is as follows :

Residual error test is performed on the residual generated so as to check whether there is presence of heteroskedasticity in the residual or not along with verifying the distribution of the residual series.


(a)

```
Series: RETURN
Sample 2 2036
Observations 2035

Mean         0.000737
Median       0.001406
Maximum      0.159900
Minimum     -0.118092
Std. Dev.    0.017462
Skewness    -0.085189
Kurtosis     10.13541

Jarque-Bera  4319.554
Probability  0.000000
```

(b)

<Figure 6> Daily Return of Sensex

<Table 2> Results of Estimated GARCH(1, 1) Model

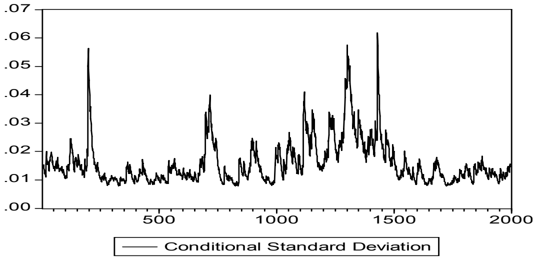
	Coefficient	Std. Error	z-Statistic	Prob.
C	0.001335	0.000278	4.802982	0.0000
RETN(-1)	0.080034	0.023915	3.346645	0.0008
Variance Equation				
C	5.78E-06	9.93E-07	5.819038	0.0000
ARCH(1)	0.135455	0.011897	11.38519	0.0000
GARCH(1)	0.849108	0.012362	68.68965	0.0000
R-squared	0.003482	Mean dependent var		0.000720
Adjusted R-squared	0.001482	S.D. dependent var		0.017474
S.E. of regression	0.017461	Akaike info criterion		-5.618686
Sum squared resid	0.607658	Schwarz criterion		-5.604672
Log likelihood	5618.067	F-statistic		1.740850
Durbin-Watson stat	2.004372	Prob(F-statistic)		0.138299

4.1 ARCH LM Test

Results of ARCH LM test indicating that there does not appear to be any ARCH up to order 7.

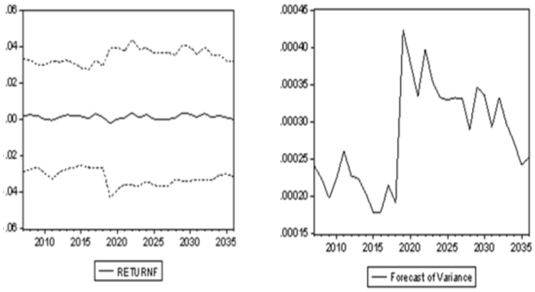
〈Table 3〉 ARCH LM Test Results on the In-Sample Data Set for Verification of Model and ARCH

ARCH Test :			
F-statistic	1.001105	Probability	0.428443
Obs*R-squared	7.011231	Probability	0.427711



〈Figure 7〉 Conditional SD graph of Residuals Generated

Forecasting of volatility for the out-of sample data for 30 samples from September 07, 2011 to October 19, 2011 has been calculated. In the current study $\alpha+\beta$ is very to 1 but is less than 1 representing that volatility shocks are persistent and volatility will die out very slowly. <Figure 7> also represents that volatility shock is persistent and shocks are dying very slowly.



〈Figure 8〉 Static Volatility Forecast and Forecast of Variance of GARCH(1, 1) Model

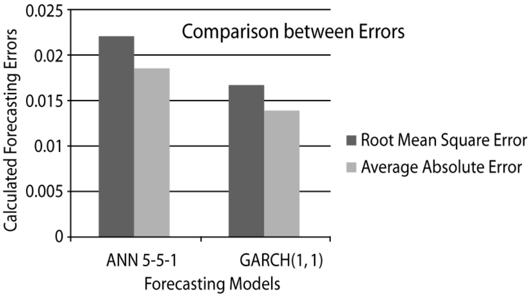
5. Comparison between ANN and GARCH Models

The results of out-of sample forecast of ANN (5-5-1) and GARCH (1, 1) models are compared. Results (<Table> 4 and <Figure 7>) show that GARCH (1, 1) model outperformed ANN (5-5-1) model in forecasting the volatility of Sensex Index. The most commonly used scale-dependent summary measures of forecast accuracy are based on the distributions of absolute errors example are Mean Square Error (MSE), Root Mean Square Error (RMSE), Mean Absolute Error and Median Absolute Error (MEDAE). Both MSE and RMSE are integral components in regression statistical models.

Various forecasting errors have been calculated out of which only RMSE and AAE are compared. Actual and forecasted values are very small (value ≈ 0) hence MAPE and MSPE are not considered. Because percentage errors are not scale-independent, they are used to compare forecast performance across different data sets. However, they have a major disadvantage in that they are infinite or undefined if value = 0 for any observation. Moreover, because the underlying error distributions of these measures have only positive values and no upper bound, percentage errors are highly prone to right-skewed asymmetry in actual practice [Smith and Sincich, 1988].

〈Table 4〉 Comparison between Errors Calculates from the Actual and Forecasted Volatility

Model	ANN 5-5-1	GARCH(1, 1)
Root Mean Square Error	0.0221	0.016780069
Average Absolute Error	0.0188	0.01390096



<Figure 9> Chart Comparing Forecasting Errors

6. Conclusion and Further work

In this paper two models one using three layered Artificial Neural Network and another using GARCH (1, 1) has been developed and compared for forecasting the return on the stock market. Returns of Sensex, which is defined as the first difference of natural log of current and previous value of the time series under study is used for forecasting volatility in the stock market. ANN model has been developed with linear function at the input layer and log sigmoid function at the hidden and output layers. Learning rules for the calculations of errors and deltas have also been derived.

Convergence and performance of models have been evaluated on the basis of the simulation results by providing values of return at the input layer and corresponding results are observed.

It was found that GARCH (1, 1) outperformed ANN (5-5-1). Volatility shocks are persistent which can be easily determined by both the models. ANN model was also able to forecast significantly.

Limitations of the models are, firstly in the case of ANN model values of learning rate, momentum and selection of layers/neurons have been chosen randomly, so a method can be developed for the calculations of learning rate, momentum factor and selection of layers/ neurons. Secondly in ARCH/GARCH modeling various other forms can be applied and their comparisons can be made.

References

[1] Azoff, E. M., “Neural Network Time Series Forecasting of Financial Markets”, John Wiley and Sons Chichester, 1994.

[2] Fama, E. F., “The Behaviour of Stock Market Prices”, *Journal of Business*, Vol. 38, 1965, pp. 34-105.

[3] Fama, E. F., “Efficient Capital Markets : A Review of Theory and Empirical Work”, *Journal of Finance*, Vol. 25, 1970, pp. 1465-1468.

[4] Hornik, K., Stinchcombe, M., and White, H., “Multilayer Feedforward Networks are Universal Approximators”, *Neural Networks*, Vol. 2, No. 5, 1989, pp, 359-366.

[5] http//in.finance.yahoo.com(Date of Downloading -October 21, 2011).

[6] Kim, K.-J. and Lee, W. B., “Stock Market Prediction Using Artificial Neural Networks with Optimal Feature Transformation”, *Neural Computing an Application*, Vol. 13, 2004, pp. 255-260.

[7] Merh, Nitin, V. P. Saxena, and Kamal Raj Pardasani, “Artificial Neural Network for Stock Market Forecasting”, *Nirma University Journal of Business and Management Studies (NUJBMS)*, Vol. 2, No. 3-4, 2008, pp. 3-19.

[8] Rumelhart, D. E., Hinton, G. E., and Williams, R. J., "Learning Internal Representations by Error Propagation, J. L. [eds.], 1986, Parallel Distributed Processing : Exploration in the Microstructure of Cognition", I, *MIT Press*, Cambridge MA, No. 1, 1986, pp. 318-362.

[9] Samantha, G. P. and Sanjib Bordoloi, "Predicting Stock Market-An Application of Artificial Neural Network Technique Through Genetic Algorithm, Finance India", Vol. XIX., No. 1, March 2005, pp. 173-188.

[10] Smith, S. and Sincich, T., "Stability over Time in the Distribution of Population Forecast Errors. Demography", Vol. 25, 1988, pp. 461-474.

[11] Srinivasan, P., "Modeling and Forecasting the Stock Market Volatility of S&P 500 Index Using GARCH Models", *The IUP Journal of Behavioral Finance*, Vol. 8, No. 1, 2011, pp. 51-69.

[12] Takaho, Hiroshi, Takayuki Arai, Tsuyoshi Otake, and MamolU Tanaka, "Prediction of the Next Stock Price Using Neural Network for Data Mining", *International Symposium on Non linear Theory and its Applications*, Xi'am, PRC, 2002.

[13] Tang, Z., Almeida, C., and Fishwick, P. A., "Time Series Forecasting Using Neural Networks vs Box-Jenkins Methodology", *Simulation*, Vol. 57, No. 5, 1991, pp. 303-310.

[14] Tang, Z. and Fishwick, P. A., "Feedforward Neural Nets as Models for Time Series Forecasting", *ORSA Journal on Computting*, Vol. 5, No. 4, 1993, pp. 374-385.

[15] Varma, J. R., "Value at Risk Models in the Indian Stock Market, Working Paper 99-07-05", *Indian Institute of Management*, Ahmedabad, 1999.

[16] Werbos, P. J., "Beyond Regression : New Tools for Prediction and Analysis in the Behaviour Sciences, Doctoral Dissertation, Applied Mathematics", Harvard University, 1974.

[17] Zhang, Dongsong and Lina Zhou, "Discovering Golden Nuggets : Data Mining in Financial Application", *IEEE Transactions on Systems*, Man, And Cybernetics, Vol. 34, No. 4, November 2004.

[18] Zhang, Guoqiang, B. Eddy Patuwao and Michael, Y. Hu, "Forecasting with Artificial Neural Networks : The state of the Art", *International Journal of Forecasting*, Vol. 14, 1998, pp. 35-62.

[19] Zhang, Qingshan Jiang Defu, and Xin li, "Application of Neural Networks in Financial Data Mining", *International Journal of Computational Intelligence*, Vol. 1, No. 2, ISSN : 2004, pp. 1304-4508.

■ Author Profile

Nitin Merh

Nitin Merh is a faculty of Computer Science and Engineering in the Institute of Engineering and Technology, JK Lakshmipat University, Jaipur, India. He has previously worked at Institute of Management, Nirma University and ICFAI Business School, India to name a few. He is recipient of Prestige Institute of Management and Research, Indore First Best Teacher Award and Best Ph. D Thesis Award in Sixth PIMR International Conference. His research interests are in Time Series Forecasting and Information Systems. He Obtained Ph. D in the area of Financial Informatics on the topic Data Mining and Artificial Neural Networks Applications in Financial Management in the Indian Context.

◈ 이 논문은 2011년 12월 02일 접수하여 2012년 01월 27일 게재확정되었습니다.