Forecasting Of Indian Stock Market Index Using Artificial Neural Network
Proposal
ABSTRACT

The objective of the study is to present the use of artificial neural network as a forecasting tool for predicting the index value of the stock market. The neural network is employed to use the homogeneous input data set which in this case is the daily returns of S&P CNX Nifty 50 Index. The data set encompassed the trading days from 1st April, 1996 to 31st March, 2008. The Daily return of the index is calculated from the daily closing prices of Nifty 50 Index. The data is collected from the historical data available on the website of National Stock Market. Accuracy of the performance of the neural network is compared using various out of sample performance measures. Modeling techniques and the architecture of the ANN will also be reported in the paper.
**Introduction**

Recently forecasting stock market return is gaining more attention, maybe because of the fact that if the direction of the market is successfully predicted the investors may be better guided and also monetary rewards will be substantial. If any system which can consistently predict the trends of the dynamic stock market be developed, would make the owner of the system wealthy. Another motivation for research in this field is that it possesses many theoretical and experimental challenges. The most important of these is the efficient market hypothesis which proposes that profit from price movement is very difficult and unlikely. In an efficient market, stock market prices fully reflect available information about the market and its constituents and thus any opportunity of earning excess profit ceases to exist any longer. So it is ascertain that no system is expected to outperform the market predictably and consistently. There has been a lot of debate about the validity of the EMH and many researchers have attempted to use neural networks to give a contradictory view to the Efficient Market Hypothesis.

Moreover, many researchers claim that the stock market is a chaos system. Chaos is a non linear deterministic process which only appears random because it is not easily expressed. These systems are dynamic, aperiodic, complicated and difficult to deal with normal analytical methods. There is not much evidence that the stock market returns are perfectly linear for the very reason that the residual variance between the predicted return and the actual is quite high. The existence of the nonlinearity of the financial market is propounded by many researchers and financial analyst. (Abhyankar et al., 1997).

During last few years there has been much advancement in the application of neural network in stock market indices forecasting with a hope that market patterns can be extracted.
A neural network is a massively parallel distributed processor made up of simple processing unit which has a natural propensity for storing experiential knowledge and making it available for use. (Simon Haykin, (1999)). Neural networks has remarkable ability to derive meaning from complicated or imprecise data, can be used to extract patterns and detect trends that are too complex to be noticed by either humans or other computer techniques. There are
several distinguished features that propound the use of neural network as a preferred tool over other traditional models of forecasting.

Neural networks are nonlinear in nature and where most of the natural real world systems are non linear in nature, neural networks are preferred over the traditional linear models. This is because the linear models generally fail to understand the data pattern and analyze when the underlying system is a non linear one. However, some parametric nonlinear model such as Autoregressive Conditional Heteroskedasticity (Engle, 1982) and General Autoregressive Conditional Heteroskedasticity have been in use for financial forecasting. But most of the non linear statistical techniques require that the non linear model must be specified before the estimation of the parameters is done and generally it happens that pre-specified nonlinear models may fail to observe the critical features of the complex system under study.

Neural networks are data driven models. Unlike, traditional time series models The novelty of the neural network lies in their ability to discover nonlinear relationship in the input data set without a priori assumption of the knowledge of relation between the input and the output (Hagen et al., 1996), rather the input variables are mapped to the output set by squashing or transforming by a special function known as activation function. They independently learn the relationship inherent in the variables which involves modification of the network parameters from a set of labeled training example. From statistical inference neural networks are analogous to nonparametric, nonlinear, regression model. However, unlike the neural network models, the traditional statistical models have limitations in understanding the relationship between the input and the output of the system because of the complex and chaos nature of the system.

Neural Networks have a built in capability to adapt the network parameters to the changes in the studied system. A neural network trained to a particular input data set corresponding to a particular environment; can be easily retrained to a new environment to predict at the same level of environment. Moreover, when the system under study is non stationary and dynamic in nature, the neural network can change its network parameters (synaptic weights) in real time.
So, neural network suits better than other models in predicting the stock market returns. This paper presents the use of artificial neural network as a forecasting tool for predicting the stock market index. The feedforward neural network with back propagation algorithm is used to forecast the daily returns of S&P CNX Nifty 50 Index. Accuracy of the performance of the neural network is compared using various out of sample performance measures.

**Literature Review**

In the last two decades forecasting of stock returns has become an important field of research. In most of the cases the researchers had attempted to establish a linear relationship between the input macroeconomic variables and the stock returns. But with the discovery of nonlinearity in the stock market index returns (A. Abhyankar et al. 1997), there has been a great shift in the focus of the researchers towards the nonlinear prediction of the stock returns. Although, there after many literatures have come up in nonlinear statistical modeling of the stock returns, most of them required that the nonlinear model be specified before the estimation is done. But for the reason that the stock market return being noisy, uncertain, chaotic and nonlinear in nature, ANN has evolved out to be better technique in capturing the structural relationship between a stock’s performance and its determinant factors more accurately than many other statistical techniques (Refenes et al., S.I. Wu et al., Schoeneburg, E.,)

Many literatures are available on application of ANN in modeling the stock market returns. Researchers have tested the accuracy of ANN in predicting the stock market index return of most developed economies across the globe. Literatures are available for forecasting index returns of U.S markets like NYSE [5], FTSE [6], DJIA [7], S&P500 [8, 9]. Studies in European context are available for markets like Euronext Paris Stock Exchange [2], German Stock Exchange [11], and Madrid Stock Exchange (Spain) [12]. Few papers are also available in context to Asian stock markets like Hang Seng Stock Exchange, Korea Stock Exchange Tokyo Stock Exchange and Taiwan Stock Exchange.

*Kim and Han (2000)* used neural network modified by Genetic Algorithm. The genetic algorithm was used to reduce the complexity of the feature space. *Kim
and Chun (1998) used refined probabilistic NN (PNN) to predict a stock market index. Pantazopoulos _et al._ (1998) presented a neurofuzzy approach for predicting the prices of IBM stock. Siekmann _et al._ (2001) implemented a network structure that contains the adaptable fuzzy parameters in the weights of the connections between the first and second hidden layers. In another paper by Rong-Jun Li; Zhi-Bin Xiong developed a fuzzy neural network that is a class of adaptive networks and functionally equivalent to a fuzzy inference system. The experiment results based on the comprehensive index of Shanghai stock market indicate that the suggested fuzzy neural network could be an efficient system to forecast financial time series.

**Data and Methodology**

The data employed in the study consists of daily closing prices of S&P CNX Nifty 50 Index. The data set encompassed the trading days from April 1st, 1996 to December 31st, 2007. The data is collected from the historical data available on the website of National Stock Market.

The study makes an attempt to design a simple neural network model where in most of the critical issues pertaining to performance of the neural network will be addressed.

The performance of the neural network largely depends on the model of the Neural Network. Issues critical to the neural network modeling like selection of input variables, data preprocessing technique, network architecture design and performance measuring statistics, should be considered carefully.

**Selection of input variables:**

Selection of input variable for the neural network model is a critical factor for the performance of the neural network because it contains important information about the complex non-linear structures of the data. It also facilitates the neural network to understand the movements in the time series. The input variables selected for this model are the lagged observation of the time series being forecasted, which in this case is the closing prices of S&P CNX Nifty 50 Index. The criticality in selecting the input variables lies in selecting the number of lagged variables of the input variables. Previous researches have indicated that with less lagged variables the information about the characteristics of the data available with the neural network increases. However, with less lagged input
variables the correlation between the lagged variable increases which may result in an over-fitting phenomenon. On the other hand, with increase in the lag between each input variable the neural network may loose out essential information of input variables, resulting in under-learning. To handle with this dilemma of over-fitting or under-learning and select an optimal structure, we will consider various lagged structure (with multiple lagged variables and varying lagged frequency) and test the performance of the neural network on a trial and error basis. A maximum of 20 lagged variables will be considered for the model.

In order to select the neural network model and for validating the performance of the network model, the whole historical data set is divided into in-sample and out of sample data set. The in-sample data set is used to construct the neural network model and for training the same. The out-of-sample data set is used to evaluate how well the network performs in forecasting with the new data set, which was not used to train and estimate the coefficients of the network. To study the performance of the network model across the whole historical range of data, the neural network performance is tested at different historic periods. Each experiment at different historic periods will be referred as different schemes with different in sample and out of sample data set. Here in every scheme, a specific ratio will be maintained between the sample size of the in sample and out of sample size of the data. In order to arrive at what will be an optimal in-sample period of study, the model will also consider various ranges (sample size) of in-sample data set for each of the scheme. To give an example, probably in each scheme the in sample period is of 2 years trading days data prior to the out of sample period.

Data Preprocessing

The performance and the reliability of a neural network model also to a large extent depend on the quality of the data used. The data preprocessing of the input variable of the neural network model facilitates de-trending of the data and highlight essential relationship, so as to facilitate proper network learning process. As neural networks are pattern recognisers, the data presented to it largely influences the accuracy of the result. Various preprocessing methods will be considered and tested for an optimal result. As again, there is no way to decide which preprocessing technique will work the best on a priori; various algorithms will be tested to achieve a stationary data series. Augmented Dickey...
fuller test will be performed to test the stationarity of the series. The data series may be scaled down at an appropriate rate in order to reduce the search space for finding the optimal coefficient estimates. However, scaling down may only be considered if the results shows a significant improvement.

**Neural Network Design and Architecture:**

The neural network architecture comprises of input layer with input neurons, hidden layer with hidden neurons in each of the hidden layers and output layer. The hidden layer of the neural network captures the data patterns and characteristics, and establishes a complex dynamic nonlinear relationship between the input and the output variable. Selection of the hidden layers in the network and the no of the neurons in each of the layers are fundamental to the structure of the neural network. The input and the output neurons can easily be determined from the no of the input and the output variables used in the model as they are equal to the input and the output variables.

While there is no specific rule that suggest the no of hidden layers required to estimate a complex non linear function, the past studies are evident that neural network with a single hidden layer is sufficient for neural network models to report successful results in the field of financial forecasting. Although, several empirical rules are also present for selection of no of neurons in the hidden layer of the neural network, but we would go by trial and error method to choose the hidden neurons.

The relationship between the input and the output of a neuron is established by the transfer function of the layer. For hidden layer the transfer function will be either tan-sigmoid function or log sigmoid function. The transfer function for the outer layer will be either linear transfer function or tan-sigmoid transfer function. Based on the performance of the network model transfer function will be finalized for the network model.

After the neural network model is constructed, training of the neural network is the next essential step of the forecasting model. Training of neural network is an iterative process of non linear optimization of the parameters like weights and bias of the network, in order to optimize the error between the input and the desired output. The result of the training process of the network depends on the algorithm used for the purpose. Two training algorithms, gradient descent adaptive backpropagation and Levenberg-Marquardt backpropagation will be considered for optimal results in our model.
The performance measure of the neural network model lies in the prediction accuracy of the model with the out of sample data. Various out of sample performance measures like NMSE and RMSE will be considered for measuring the prediction accuracy of the model. In order to test the prediction accuracy of the neural network across various time periods in history, the model will tested at various time horizon and the results of them will be compared. An Analysis of variance test (AV Test) can be carried out to test the stability of the model over a period of time.

The uniqueness of the research comes from the fact that the research employs a neural network based forecasting approach on National Stock Exchange index (CNX S&P Nifty 50) Furthermore, as not much work has been done on the forecasting of Indian stock market indices using neural network, this paper will actually help to develop neural network as another forecasting tool for highly volatile Indian market. Most importantly, this paper considers all the issues and critical factors for designing the neural network model and in order to select the appropriate network with best combination of the input nodes, lag frequency, no of input variables, data preprocessing algorithm, in-sample and out-of-sample data splitting, hidden neurons, transfer functions and learning algorithms a rigorous trial and error method is employed.

<table>
<thead>
<tr>
<th>Sl. No</th>
<th>Critical Modeling Factors</th>
<th>Various options considered for Network Design</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Input variable selection</td>
<td>• Various lagged structures with different lag (probably a lag of 1, 2 or 3) between each input variable will be considered in order to address the dilemma of over fitting and under learning of the network.</td>
</tr>
<tr>
<td>2</td>
<td>In sample and out of Sample data size.</td>
<td>• Testing with various sample size (e.g. 2, 3, 5 years or even more, of trading day’s data) of the in sample data set maintaining a fixed ratio between the sample size of the in-sample and out of sample data set.</td>
</tr>
<tr>
<td>3</td>
<td>Data Pre Processing</td>
<td>• Testing with various data preprocessing techniques (e.g. logarithmic first difference,</td>
</tr>
</tbody>
</table>
Table 1: Critical factors for network design and various options for each factor considered for design.

By trial-and-error procedure of selecting each critical factor of the network design from all plausible combinations of network structures, a final structure of neural network will be suggested. The suggested network design will be tested for prediction accuracy.

References
