NON PARAMETRIC MODELING OF STOCK INDEX

Sujatha K.V.[1] and Meenakshi Sundaram S.[2]

[1]Research Scholar, Sathyabama University, Chennai 600 119.
[2]Department of Mathematics, Sathyabama University, Chennai 600 119.
Email: [1]sujathacenthil@gmail.com, [2]sundarambhu@rediffmail.com

Abstract

An attempt is made to predict the daily closing prices of BSE data which is highly fluctuating. The variables considered were found to be non normal as evidenced from Multivariate Omnibus test. Hence instead of Classical Multivariate Statistical procedures, the Non parametric Neural Network model with the new set of independent variables using Principal Component Analysis was used to predict the daily prices. The predictive ability of each model is measured using standardized error measures.

Key Words: Normality, Prediction, nonparametric, PCA, Error Measure.

I. INTRODUCTION

Financial Forecasting or specifically Stock Market prediction is one of the hottest fields of research lately due to its commercial applications owing to the high stakes and the kinds of attractive benefits that it has to offer. Mining stock market tendencies is a challenging task due to its high volatility and noisy environment. Technically the stocks prices are evaluated by analyzing statistics generated by market activity, past prices, and volume. It looks for peaks, bottoms, trends, patterns, and other factors affecting a stock's price movement. Future values of stock prices often depend on their past values and the past values of other correlated variables.

There are many technical indices used in stock market prediction. Moving Average, Exponential Moving Average, Weighted Moving Average, Moving Average Difference Oscillator, Relative Strength Index, Volume, Volume Change, Moving Average Convergence-Divergence, Momentum, Rate of Return, Advance-Decline, Upside-Downside Volume Ratio, High-Low Differential Index, High-Low Ratio, Volume, and Historical Volatility are some examples. There are many variations of these and new index terms may be derived from them. Each index has its own meaning and interpretation. A comprehensive description of technical stock market indicators can be found Robert[12]. For technically analyzing the BSE Stock Index data the variables daily opening, high, low and volume of transaction were considered from 1st January 2009 till 31st January 2010(260).

Traditionally forecasting research and practice had been dominated by statistical methods Robert [11], Richard [12]. All traditional methods require the data to be multivariate normally distributed. Testing for normality is a common procedure in much applied work and many tests have been proposed Mardia [16], D'Agostino [3] and Small [13]. The need for testing normality in a multivariate setting is discussed by Ganadesikan (1977), Cox and Small [1] and Cox and Wermuth [2]. A test frequently used is the sum of squares of the standardized sample skewness and kurtosis, which is asymptotically distributed as a X^2 variate. Multivariate Omnibus test proposed by Mardia was considered to check the condition of normality of the predictors.

The variables under analysis were found to be multi collinear which is overcome by non parametric Principal Component Analysis. The new set of uncorrelated principal components was used as predictors for predicting the daily closing prices of BSE sensex data. As the dependent variables were violating the basic assumption of normality, nonparametric neural network modeling of the multivariate data was accomplished. The power of neural networks is its ability to model a nonlinear process without a priori knowledge about the nature of the process. The idea of prediction using neural networks is to find an approximation of mapping between the input and output data values through training.

The remainder of this paper is organized as follows. In Section II brief discussion of the theoretical issues of testing multivariate normality, Principal component analysis and details on research design and methodology are provided. Results are discussed in Section III. Finally, concluding remarks are offered in Section IV.

II. RESEARCH DESIGN AND METHODOLOGY

A. Multivariate Normality

It is well known that many multivariate statistical procedures, including MANOVA, discriminant analysis, and canonical correlations, call upon the assumption of multivariate normality. Although at least 50 tests of multivariate normality exist, relatively little is known about the power of these procedures. Mardia's skewness and kurtosis measures are recommended for diagnosing possible deviations from normality.

Let $X' = (X_1, \ldots, X_n)$ be a $p \times n$ matrix of n observations on a $p-$ dimensional vector with sample mean and covariance $\overline{X} = n^{-1}(X_1 + \ldots + X_n)$ and $S = n^{-1}\overline{X}'\,\overline{X}$ where $\overline{X}' = (X_1 - X, \ldots X_n - X.$

Create the $n \times n$ matrix:

$$D = (d_{ij}) = \overline{X}\, S^{-1}\, \overline{X}' \quad [1]$$

And define multivariate measures of skewness and kurtosis as:

$$b^*_{1p} = \frac{1}{n^2} \Sigma^n_{i=1} \Sigma^n_{j=1} d^3_{ij} \quad [2]$$

and

$$b^*_{2p} = \frac{1}{n} \Sigma^n_{i=1} d^4_{ii} \quad [3]$$

An omnibus test based on these measures is given by

$$M_p = \frac{nb^*_{1p}}{6} + \frac{n\,(b^*_{2p} - p\,(p+2))^2}{8p\,(p+2)} \quad [4]$$

$$- X^2 \left(\frac{p\,(p+1)\,(p+2)}{6} + 1 \right)$$

A. Principal Component Analysis

Principal component analysis (PCA) technique consists in rewriting the coordinates in a data set in other coordinates system which will be more convenient for analysis. This new coordinates are represented on orthogonal axis, being obtained in decreasing variance order. The total amount of principal components is equal to the amount of original variables and presents the same statistical information. The PCA is defined as follows:

Let $X' = (x_1, x_2, \ldots x_p)$ be a p dimensional random variable. The i^{th} principal component of X' is

$$y_i = e'_i\, x = e_{1i}\, x_1 + e_{2i}\, x_2 + \ldots + e_{\pi}\, x_p \quad [5]$$

$$(i = 1, 2, \ldots, p, e'_i\, e_i = 1)$$

and it must satisfy the following conditions:

- The variable y_1 is the one whose variance is maximum among all of the variance of $y = e'\, x$
- The variable y_k is not correlative with $y_1, y_2, \ldots y_{k-1}\, (k = 2, 3, \ldots, p)$

Therefore, p principal components of p variables are p linear combinations of the p variables, where the coefficient vectors of the linear combinations are unit vectors. The first principal component y_1 is the variable whose variance is maximal among the variances of the linear combinations. The second principal component y_2 is the variable whose variance is maximal among the variances of the linear combinations and irrelevant with y_1. The third principal component y_3 is the variable whose variance is maximal among the variances of the linear combinations and irrelevant with both y_1 and y_2, and so on.

Proportion between the variance of the k[th] principal component and the sum of all deviations:

$$\frac{\lambda_k}{\lambda_1 + \lambda_2 + \ldots + \lambda_p}\,(k = 1, 2, \ldots p) \quad [5]$$

It is called contribution rate of the principal component y_k

Correlation coefficient between y_i and x_j,

$$\rho_{y_i, x_j} = \frac{e_{ji}\sqrt{\lambda_i}}{\sqrt{\sigma_{jj}}}\,(i, j = 1, 2, \ldots, p) \quad [6]$$

It is called as the factorial loading or the principal component loading.

B. Multilayer Perceptron

Neural Networks [NN] have been used in function estimation such as stock price prediction, option price modeling, portfolio optimization and currency exchange rate estimation (Steiner and Wittkemper [14]; Yao and Tan [20]; Galindo [4]; Leigh et al. [7]; Hutchinson et al. [6]; Trafalis et al. [17]. NN is a learning machine that is designed to model the way in which the brain performs the particular tasks. The multi-layer perceptron (MLP) is the most widely used type of NN for function approximation. It is both simple and based on solid mathematical grounds. Input quantities are processed through successive layers of "neurons". There is always an input layer, with a number of neurons equal to the number of variables of the problem, and an output layer, where the perceptron response is made available, with a number of neurons equal to the desired number of quantities computed from the inputs.

The layers in between are called "hidden" layers. All problems which can be solved by a perceptron can be solved with only one hidden layer, but it is sometimes more efficient to use two hidden layers. Each neuron of a layer other than the input layer computes first a linear combination of the outputs of the neurons of the previous layer, plus a bias. The coefficients of the linear combinations plus the biases are called the weights. They are usually determined from examples to minimize, on the set of examples, the (Euclidian) norm of the desired output – net output vector. Neurons in the hidden layer then compute a non-linear function of their input. The two main activation functions used in current applications are hyperbolic tangent and sigmoid, and are described by

$$\phi(y_i) = \tan h(v_i)$$

$$\phi(y_i) = (1 + e^{-v_i})^{-1} \quad [7]$$

in which the former function is a hyperbolic tangent which ranges from – 1 to 1, and the latter is equivalent in shape but ranges from 0 to 1. Here y_i is the output of the i^{th} node (neuron) and v_i is the weighted sum of the input synapses. The MLP divides the data set in to three parts

Training - To train the Network

Testing - To prevent over training

Holdout - To access the final network.

Multilayer Layer Perceptron has rescaling option which is done to improve the network training. There are three rescaling options: standardization, normalization, and adjusted normalization. All rescaling is performed based on the training data, even if a testing or holdout sample is defined.

$$\text{Standardization} = \frac{\text{X} - \text{Mean}}{\text{S}} \quad [8]$$

$$\text{Normalization} = \frac{\text{X} - \text{Minimum}}{\text{Maximum} - \text{Minimum}} \quad [9]$$

Adjusted Normalization

$$2 \times \frac{\text{X} - \text{Minimum}}{\text{Maximum} - \text{Minimum}} - 1 \quad [10]$$

The units in the output layer can use any one of the following activation function - Identity, Sigmoid, Softmax or Hyperbolic Tangent. The activation functions are given below

$$\gamma(c) = \frac{e^c - e^{-c}}{e^c + e^{-c}} \quad [11]$$

$$\gamma(c) = \frac{1}{1 + e^{-c}} \quad [12]$$

$$\gamma(c_k) = \frac{\exp(c_k)}{\Sigma_j \exp(c_j)} \quad [13]$$

$$\gamma(c) = c \quad [14]$$

Error Functions that are used are sum of square error and relative error.

Sum of square error is defined as the sum of the squared deviation between observed and the model predicted value. Sum of Square Error,

$$E_T(c) = \sum_{m=1}^{M} E_m(c) \quad [15]$$

Where

$$E_m(c) = \frac{1}{2}\sum_{r=1}^{R}(Y_r^{(m)} - a_{1-r}^{m})^2 \quad [16]$$

$Y_r^{(m)} =$ Target vector, pattern

a_{1-r}^{m} Unit j for layer i, for $i = 0, 1, 2, \ldots 1$

Relative Error is the ratio of an absolute error to the true, specified, or theoretically correct value of the quantity that is in error,

$$\text{Relative Error} = \frac{\sum_{m=1}^{M}\left(Y_r^{(m)} - \hat{Y}_r^{(m)}\right)^2}{\sum_{m=1}^{M}\left(Y_r^{(m)} - \bar{Y}_r\right)^2} \quad [17]$$

$\bar{Y}_r =$ The mean of $Y_n^{(m)}$

Standardized Error Measures used for comparing the ability of the MLP models are

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|A_t - P_t| \quad [18]$$

$$MAPE = \frac{1}{n}\sum_{t=1}^{n}\frac{|A_t - P_t|}{A_t} \quad [19]$$

$$SMAPE = \frac{1}{n}\sum_{t=1}^{n}\frac{|A_t - P_t|}{A_t + P_t} \quad [20]$$

were A_t is the actual value and P_t is the predicted value.

III. RESULTS AND FINDINGS

The stock prices are highly volatile and have a chaotic behavior. Figure 1 shows the daily closing price curve for the period considered.

Multivariate Skewness and Kurtosis for the 260 observations,

Multivariate Skewness = 48.5176

Multivariate Kurtosis = 4.6701e + 005

Mardia's test statistics value

= 2.9531e + 011

From the test statistic value it was clear that the variables are not normally distributed.

The variables daily opening price, high price, low price and volume of transaction of BSE Sensex data are found to be multi collinear in nature. Principal component Analysis, using the covariance matrix of the data resulted into a new set of four uncorrelated independent variables.

Table 1 Optimal Weight of Principal Component Variables

Variable	PCA1	PCA2	PCA3	PCA4
Open	0.061	0.576	0.597	− 0.554
High	0.061	0.576	0.177	0.796
Low	0.063	0.570	− 0.782	− 0.243
Volume	− 0.994	0.106	− 0.002	− 0.001

Figure 1 Daily closing Prices of BSE Sensex

New set of four independent variables are taken as input variables and the daily closing prices of BSE is considered to be the target variable of MLP. Out of the 260 observations 130 are considered for training, 32 for testing and 78 as hold out.

Initially the activation function of the hidden layer is set to hyperbolic tangent, Identity is taken as the activation function of the output layer, the target variable is standardized and then the input variables are successively smoothed (standardized, normalized and adjusted normalized). With this architecture the sum of square error and relative error values are found for training, testing and hold out. These error values are also measured after fixing the activation function of the output layer as hyperbolic tangent with rescaled variables.

Finally the activation function of the output layer is set to sigmoid. The error values are also found for the different network as formulated above with the activation function of the hidden layer fixed to sigmoid. The different combinations of the activation function of the output and the hidden layer with the three rescaling options of the input and target variables resulted in 30 models. The models with minimum error value are given in Table 2. The network structure of the models with minimum error values is provided in Table 3.

Table 2 Sum of Square and Relative Error Values

Model	Training		Testing		Holdout
	Sum of Square Error	Relative Error	Sum of Square Error	Relative Error	Relative Error
Model 1	0.07371751	9.89E-04	0.01148669	0.00100463	0.00111982
Model 2	0.0071049	8.85E-04	0.00132279	0.00107367	0.00120048
Model 3	0.00725588	9.04E-04	1.01E-03	8.17E-04	1.30E-03

Table 3 Network Structure of Optimal Models

Models	Activation Function of the Hidden Layer	Activation Function of the Output Layer	Rescaling of Dependent Variable	Rescaling of Covariates
Model 3	Hyperbolic Tangent	Identity	Normalized	Adjusted Normalized
Model 6	Hyperbolic Tangent	Identity	Standardized	Adjusted Normalized
Model 21	Sigmoid	Identity	Normalized	Adjusted Normalized

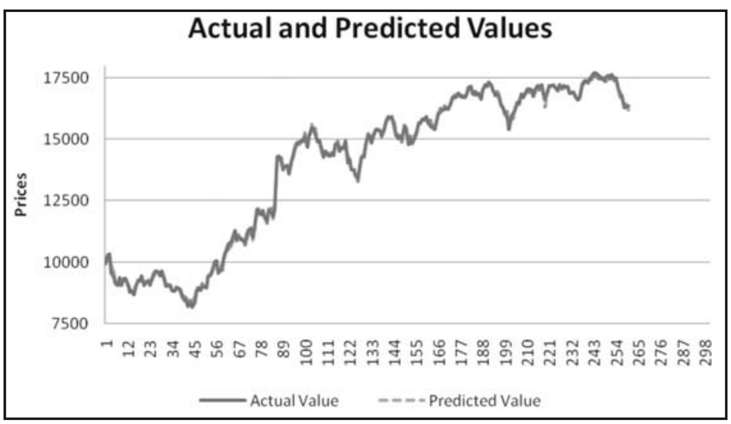

Figure 2 Actual and Predicted Values

Table 4 MAE, MAPE and SMAPE Values

Partition	Model 1			Model 2			Model 3		
	Mae	Mape	Smape	Mae	Mape	Smape	Mae	Mape	Smape
Training	78.59152	0.00610	0.00304	129.408	0.00992	0.00496	89.3765	0.00706	0.003535
Testing	73.71611	0.00578	0.00289	119.912	0.00921	0.00460	83.9202	0.00675	0.003382
Holdout	58.41691	0.00392	0.00196	107.985	0.00720	0.00359	60.0960	0.00410	0.00205

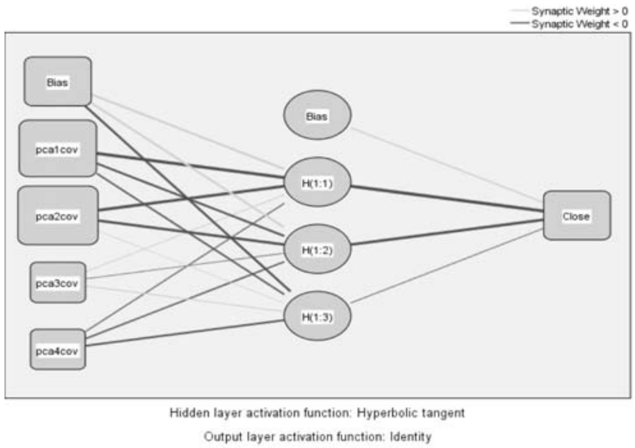

Figure 3 Network Diagram

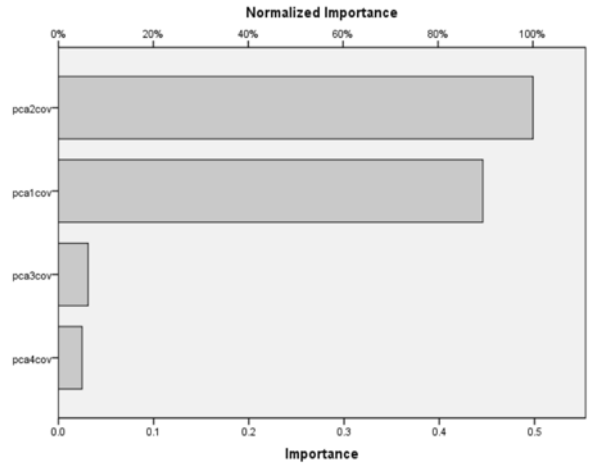

Figure 4 Independent Variable Importance Chart

The daily closing values were predicted with the MLP Networks having minimum error. With the predicted values Error Measure – Mean Average Error, Mean Absolute Percentage Error and Symmetric Mean Absolute Percentage Error were found. From the Error values of the MLP models given in the Table 4, the error values are least for the Multilayer Perceptron Model 3. The parameter estimates of the best fitted MLP model are given in Table 5.

IV. CONCLUSION

Some procedure for assessing the assumption of multivariate normality should be used, even if the subsequent multivariate analyses are robust to violations of Multivariate Normality.

The Multilayer Perceptron Model with hyperbolic tangent as activation function hidden layer, identity as activation function of the output layer with adjusted normally smoothed input variables along with target variables rescaled normally found to be more efficient in predicting daily closing price.

Stock prices are very sensitive to the information of politics, economy and society which are not considered in this paper, which may influence the prediction error if considered. Obviously the further study of this paper is to add up additional variables to improve the predictive ability of the model.

References

[1] Cox, D.R. and Small, N.J.H. 1978, Testing Multivariate normality, Biometrika, 65, pp. 263-272.

[2] Cox, D.R. and Wermuth, N 994, Tests of linearity, multivariate normality and the adequacy of linear scores, Appl. Statistic, 43, pp. 347-355.

[3] D'Agostino, R.B., 1982, Departure from normality, testing for. In Kotz S., Johnson, N.L. and Read, C.B. (eds), Encyclopedia of Statistical Sciences, vol. 2. Amsterdam: North-Holland.

[4] Galindo J, 1998, A framework for comparative analysis of statistical and machine learning methods: An application to the black scholes option pricing equations. Techincal report Banco de Mexico, Mexico, DF, 04930.

[5] Haykin S 1994, Neural networks: a comprehensive foundation, MacMillan New York.

[6] Hutchinson JM, Loaw, Poggio T, 1994, A nonparametic approach to pricing and hedging derivative securities via learning networks,Journal of Finance XLIX, pp. 851–889.

[7] Leigh W, Paz M, Purvis R, 2002, An analysis of a hybrid neural network and pattern recognition technique for predicting short- term increases in the nyse, International Journal of Management Science 30, pp. 69-76.

[8] Lean Yu, Shouyang Wang, KinKeung Lai, 2005, Mining Stock Market Tendency Using GA- Based Support Vector Machines, WINE 2005, LNCS 3828, Springer Verlag Berlin, Heidelberg pp 336- 345.

[9] Mardia, K.V. 1970, Measures of multivariate skewness and kurtosis, Biometrika 57, pp. 519- 530.

[10] Mardia, K.V., 1980, Tests of multivariate and univariate normality, In Krishniah, P. R. (edition), Handbook of Statistics, vol. 1, ch.9. Amsterdam: North Holland.

[11] Richard Johnson and Dean Wichern, 1998, Applied Multivariate Statistical Analysis, 4th Ed. Prentice Hall, New Jersey.

[12] Robert W. Colby, and Thomas A. Meyers, 1988, The Encyclopedia of technical Market Indicators, Dow Jones- Irwin, Homewood Illinoise.

[13] Small, N.J.H, 1985, Multivariate normality testing for. In kotz, S., Johnson, N.L. and Read, C.B. (edtions), Encyclopedia of Statistical Sciences, vol. 6. Amsterdam: North-Holland.

[14] Steiner M, Wittkemper H, 1997, Portfolio optimization with a neural network implementation of the coherent market hypothesis., European Journal of Operations Research 100, pp. 27–40.

[15] Sujatha K.V. and S. Meenakshi Sundaram, 2010, A MLP, RBF Neural Network Model for Prediction in BSE SENSEX Data Set, Proceedings of National Conference on Applied Mathematics.

[16] Trafalis TB, Ince H, 2002, Benders decomposition technique for support vector regression, Proceedings of the 2002 International Joint Conference, IEEE, vol. 3, pp. 2767–2772.

[17] Trafalis TB, Ince H, Mishina T, 2003, Support vector regression in option pricing, Proceedings of Conference on Computational Intelligence and Financial Engineering, Hong Kong.

[18] Xiaoping Yang, 2005, The Prediction of Stock Prices Based on PCA and BP Neural Networks Chinese Business Review, ISSN 1537-1506, USA Vol 4, No.5 (Serial No.23) pp. 64 – 68.

[20] Yao J, Tan CL, 2000, A case study on using neural networks to perform technical forecasting of forex., Neurocomputing vol 34 pp 79–98.

K.V. Sujatha is working in the department of Mathematics, Sathyabama University, Chennai. She has a teaching experience of 9 years. Her field of interest Non parametric models, Multivariate statistical methods and statistical inference.