DECISION SCIENCES INSTITUTE

Stock Trading Decisions Using Ensemble-based Forecasting Models

Dhanya Jothimani
Indian Institute of Technology Delhi, India
Email: dhanyajothimani@gmail.com

Ravi Shankar
Indian Institute of Technology Delhi, India
Email: ravi1@dms.iitd.ac.in

Surendra S. Yadav
Indian Institute of Technology Delhi, India
Email: ssyadav@dms.iitd.ac.in

ABSTRACT

In this paper, we propose an ensemble-based forecasting framework for making stock trading decisions. Here, two ensemble models are proposed for predicting the stock index. In first stage, Empirical Mode Decomposition (EMD) and Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (CEEMDAN) are used for decomposing the financial series into various sub-series. In second stage, each series is predicted using a machine learning algorithm, namely, Artificial Neural Network (ANN). Trading rules are defined. The models were tested on Nifty data. Return on investment based on the trading rules of proposed ensemble models was compared with Buy-and-Hold strategy.

KEYWORDS: Financial time series prediction, EMD, CEEMDAN, Trading Rules, Ensemble Models, ANN, Nifty

INTRODUCTION

Stock market attracts investors (naive, well-informed) and spectators alike. Every investor wants to gain profit at minimal risk. However, the complexity of stock market can be attributed to several factors including socio-political events and investors' behavior. This complexity, often, does not enable one to predict the direction of movement of stock prices and determine the timing of trading decision accurately. Hence, prediction of stock prices has gained attention of analysts and academicians from several domains including Finance, Mathematics and Computer Science.

Finance literature deals with two major ways to predict stock price: (i) Fundamental analysis, and (ii) Technical analysis. Fundamental analysis deals with factors like macroeconomic variables and fundamental information such as Price to Earnings (PE) ratio, Earnings Per Share (EPS) and Return on Equity (ROE) (Fischer and Jordan, 1987) to predict the movement of stock price.

Technical analysis analyzes the behavior of past stock price and volume data to predict the future price values. Technical analysis focuses on developing a model that can predict the stock price dynamics accurately. Various statistical and computationally intelligent models have been proposed for forecasting stock prices yet none seems to have captured the complex dynamics of stock market accurately (Atsalakis and Valavanis, 2009; 2013). This led to the concept of ensemble models.

Ensemble models, popular in Machine Learning and Statistics, work on the idea that prediction accuracy can be improved by using multiple predictors than using any of base predictors (Opitz and Maclin, 1999) alone. There are two types of ensemble models: pre-processing and post-processing. Pre-processing deconstructs the dataset into various components and selects a predictor for analyzing each component. Decomposition of time-series is one such preprocessing technique.

In post-processing, the predictors are selected based on data characteristics. For instance, AutoRegressive Integrated Moving Average (ARIMA) are used for modelling stationary and linear data, Generalised AutoRegressive Conditional Heteroskedasticity (GARCH) models for non-linear data and similarly, Artificial Neural Network (ANN) and Support Vector Regression (SVR) for non-stationary and non-linear data.

There are two types of decomposition models: (a) classical model, and (b) non-classical models. Classical model yields three components, namely, trend, seasonal and random. Though widely used, it suffers from few limitations: (i) it works best only with linear data but not with non-linear data, and (b) while aggregating the forecasts of components, it ignores random component leading to loss of information (Theodosiou, 2011). Few signal processing techniques like Discrete Wavelet Transform (DWT) and Empirical Mode Decomposition (EMD) fall under the category of non-classical models and are being used for decomposing time series in time-frequency domain and time domain, respectively (Liu et al, 2012; Lahmiri, 2014; Jothimani et al, 2015; 2016).

EMD was proposed by Huang et al. (1998). It uses Huang-Hilbert Transform (HHT) to decompose non-stationary and non-linear time series into a set of adaptive basis function called Intrinsic Mode Function (IMF). Since it is adaptive, it does not require any *a priori* information. DWT requires prior information regarding scale of decomposition. Despite its advantages of handling non-linear and non-stationary data, DWT suffers from leakage between scales (Crowley, 2010). EMD suffers from limitation of mode-mixing problem (Wu and Huang, 2009; Torres et al, 2011). Mode mixing is defined as either presence of signals of widely different scales in a single IMF or presence of signal of a similar scale in different IMF components. To overcome this limitation, a variation of EMD called Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (CEEMDAN) was introduced (Torres et al, 2011). Similar to EMD, CEEMDAN can be used for data preprocessing.

The paper presents two ensemble forecasting models, namely, ensemble EMD-ANN and ensemble CEEMDAN-ANN models to predict 1-step ahead forecasts for weekly Nifty price index. In first stage, EMD (for EMD-ANN model) and CEEMDAN (for CEEMDAN-ANN) was used to decompose time-series into various components, i.e., Intrinsic Mode Function (IMF) and residue. In second stage, ANN is used to predict each component. The forecasted components are aggregated to obtain the final forecast.

To help the investors to make trading decisions, trading rules are defined and illustrated using forecasts obtained from EMD-ANN and CEEMDAN-ANN models. A comparative analysis was carried out with traditional Buy-and-Hold strategy.

The contributions of the paper are three-fold. Firstly, it demonstrates the use of ensemble model of non-classical decomposition and machine learning algorithm for stock price prediction in Indian stock market. Secondly, it highlights the application of CEEMDAN as a data preprocessing technique. Thirdly, it illustrates the application of ensemble models for making trading decisions.

The paper is organized as follows: Next section discusses the ensemble framework followed by Methodology. Results are discussed followed by presenting few concluding remarks.

ENSEMBLE FRAMEWORK FOR INVESTMENT DECISIONS

The flowchart of the ensemble framework for trading decisions is shown in Figure 1. The detailed steps are as follows:

1. Decompose the original series *F(t)* into various sub-series using EMD and CEEMDAN in case of ensemble EMD-ANN model and CEEMDAN-ANN model, respectively.
2. Forecast each sub-series obtained using Artificial Neural Network (ANN).
3. Aggregate the forecasted sub-series to obtain forecasted series *F'(t)*.
4. Compare the forecasted series *F'(t)* with the original series *F(t)* to calculate error measures, i.e., Root Mean Square Error (RMSE) and Directional Accuracy (DA).
5. Test the forecasting ability of various models, namely, ANN, EMD-ANN and CEEMDAN-ANN using a statistical procedure called Wilcoxon Signed Rank Test (WSRT).
6. Define trading rules for each model and calculate return on investment (ROI).

Empirical Mode Decomposition (EMD)

The steps for decomposing the original time series *F(t)* are detailed below (Huang et al, 1998; 2003):

1. Identify all local minima and local maxima; interpolate them using cubic spline interpolation method to obtain a lower envelope *L(t)* and an upper envelope *U(t)*, respectively.
2. Calculate the mean of lower and upper envelopes using the following formula:

$$M(t) = (L(t) + U(t)) / 2 \tag{1}$$

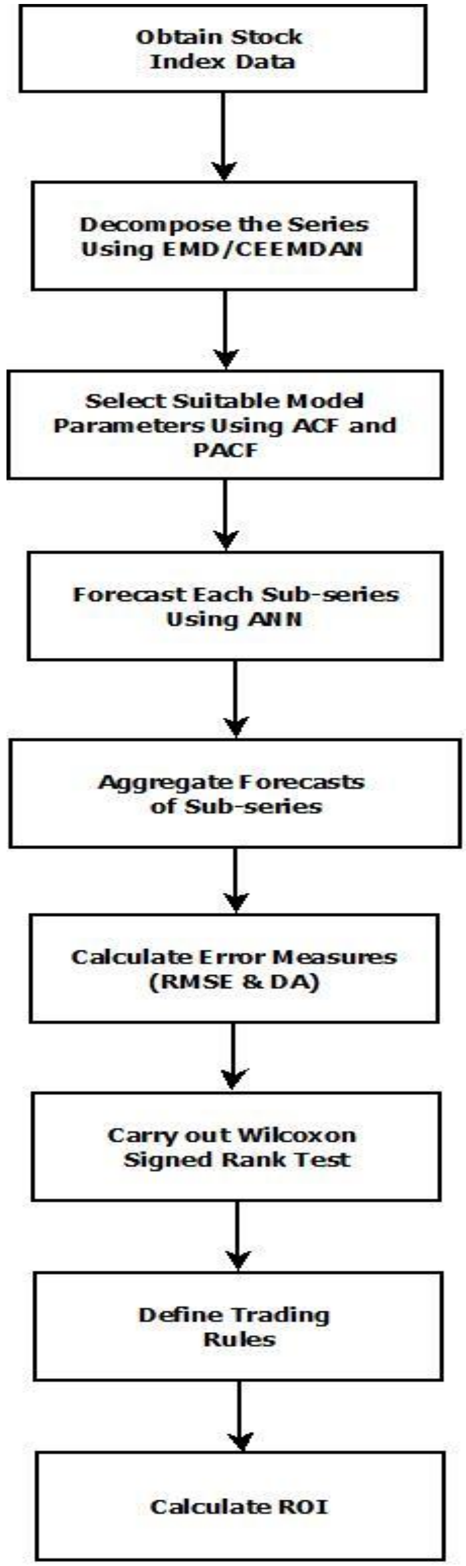
   Subtract *M(t)* from the original series to obtain a local detail *Z(t)*.

$$Z(t) = F(t) - M(t) \tag{2}$$

3. Repeat Steps 1 and 2 on *Z(t)* till: (a) the value of *M(t)* approaches zero, and (b) the difference between the number of local extrema and zero crossings is at most 1. This process is known as sifting.
   The first IMF obtained component is first intrinsic mode function $IMF_1(t)$ which is nothing but *Z(t)*.

$$IMF_1(t) = Z(t) \tag{3}$$

Figure 1: Flow Chart of Ensemble Framework for Investment Decisions



Subtract the details *Z(t)* from *F(t)* to obtain residue $R_1(t)$.

$$R_1(t) = F(t) - Z(t) \quad (4)$$

4. Repeat steps 1-3 on $R_1(t)$ to obtain the second IMF $IMF_2(t)$ and the second residue $R_2(t)$. The process is repeated on $R_i(t)$ to obtain $IMF_{i+1}(t)$ and $R_{i+1}(t)$ until $R_{i+1}(t)$ does not have more than two local extrema, where *i=1, 2, …, N-1*.

The original series *F(t)* is expressed as:

$$X(t) = \sum_{i=1}^{N} IMF_i(t) + R_N(t) \quad (5)$$

Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (CEEMDAN)

Despite the advantages of EMD as an adaptive method to decompose non-linear and non-stationary series, it suffers from a major limitation of mode mixing problem (Wu and Huang, 2009; Torres et al, 2011). Mode mixing is defined as either presence of signals of widely different scales in a single IMF or presence of signal of a similar scale in different IMF components. This affects aliasing of signals in the time-frequency distribution. To overcome this limitation, a variation of EMD known as Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (CEEMDAN) was proposed (Torres et al, 2011).

The detailed procedure of CEEMDAN is as follows:

1. To the original series *F(t)*, add white noise $w^i[t](\sim N(0,1))$ and decompose the combination by *I* realizations to obtain the first IMF of CEEMDAN using the following formula:

$$\overline{IMF_1}[t] = \frac{1}{I}\sum_{i=1}^{I} IMF_1^i[t] \quad (6)$$

2. Calculate first residue by subtracting first IMF of CEEMDAN from the original series *F(t)*

$$r_1[t] = F(t) - \overline{IMF}_1[t] \quad (7)$$

3. Obtain second mode by decomposing realizations $r_1[t] + \varepsilon_1 E_1(w^i[t])$, *i = 1, 2,…, I*.

$$\overline{IMF_2}[k] = \frac{1}{I}\sum_{i=1}^{I} E_1(r_1[t] + \varepsilon_1 E_1(w^i[t])) \quad (8)$$

   Ej(.) is an operator which produces the j-th mode of the signal obtained by EMD.

4. Calculate the k^{th} residue.

$$r_k[t] = r_{k-1}[t] - \overline{IMF}_k[t] \forall \ k=2,\dots,K \quad (9)$$

5. Calculate I realizations of $r_k[t] + \varepsilon_k E_k(w^i[t])$, *i=1, 2,…, I* to obtain $(k+1)^{th}$ mode.

$$\overline{IMF}_{(k+1)}[t] = \frac{1}{I}\sum_{i=1}^{I} E_1(r_k[t] + \varepsilon_k E_k(w^i[t])) \quad (10)$$

6. Go to Step 4 for next *k*.

7. Repeat steps 4 to 6 till it is no longer feasible to decompose the obtained residue. The final residue is defined as:

$$R[t] = F(t) - \sum_{k=1}^{K} \overline{IMF_k} \tag{11}$$

where, K is the total number of modes.

The original series can be expressed as:

$$F(t) = \sum_{k=1}^{K} \overline{IMF}_k + R[t] \tag{12}$$

CEEMDAN provides complete decomposition and exact reconstruction of original data series. It overcomes the mode mixing problem. Similar to EMD, CEEMDAN is adaptive in nature.

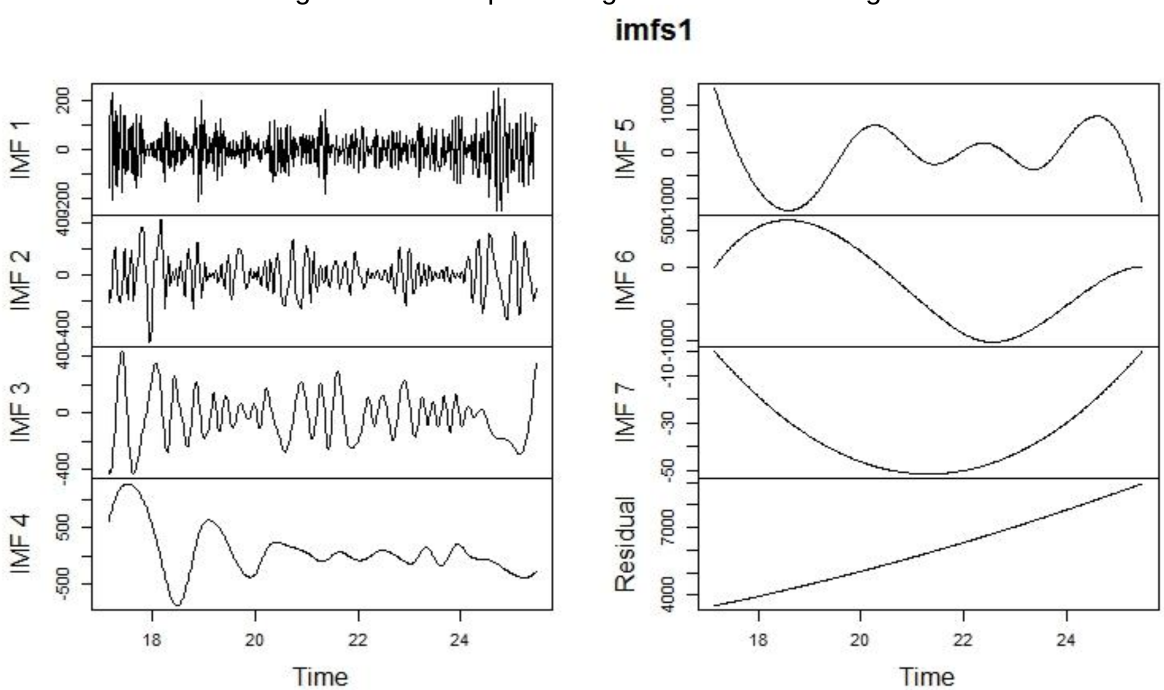
METHODOLOGY

National Stock Exchange Fifty Index, commonly known as Nifty, is the benchmark index of Indian stock market. Nifty consists of 50 companies and covers 22 sectors. The original raw data consists of weekly closing prices of Nifty. The dataset covered a period of 8 years ranging from September 2007 to December 2015.

Decomposition of Time Series

The financial time series, i.e., weekly Nifty index, is decomposed using EMD and CEEMDAN for ensemble EMD-ANN and CEEMDAN-ANN models, respectively. Seven relatively stationary IMFs along with residual component were produced in case of both EMD and CEEMDAN, which are shown in Figure 2 and Figure 3, respectively. Now, each sub-series is forecasted using Artificial Neural Network.

Figure 2: Decomposed Signals Obtained Using EMD



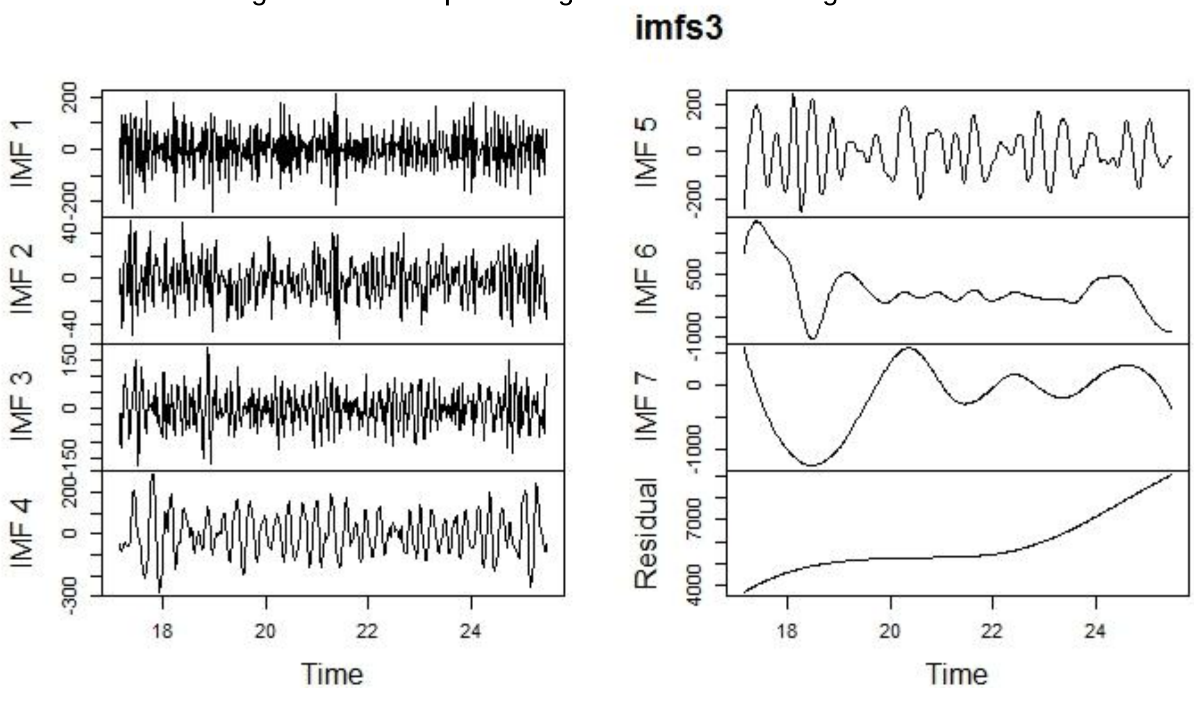
Artificial Neural Network (ANN)

ANN is used for forecasting each sub-series independently, then aggregate the forecasts of sub-series to obtain the final forecasts *F′(t)* of both EMD-ANN and CEEMDAN-ANN models.

Since ANN is a supervised learning algorithm, first 70% of dataset is used for training the network and rest 30% is used to test the validity of the model. A resilient three-layered feed forward network consisting of input layer, hidden layer and output layer is used. Each layer consists of certain number of neurons. Number of neurons in the input layer is determined using the relationship between data and its past values. This relationship is known as lag parameter and is determined using Auto Correlation Function (ACF) and Partial Auto Correlation Function (PACF) plots. The ACF and PACF plots of IMF3 obtained using CEEMDAN are shown in Figure 4. IMF3 at time *t* depends on its past six values since it cuts off at lag 6. Hence, the number of neurons in input layer is six and is expressed as:

$$X(t) = f[X(t-1), X(t-2), X(t-3), X(t-4), X(t-5), X(t-6)] \tag{13}$$

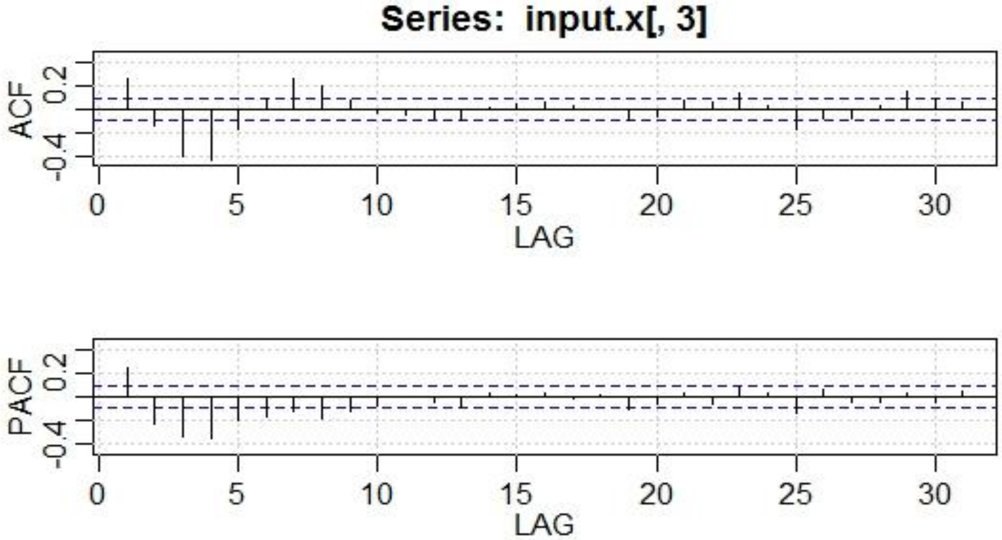
Figure 3: Decomposed Signals Obtained Using CEEMDAN



Since it is a regression problem and forecasted value is obtained, hence, the number of neurons in the output layer is one. The number of neurons in the hidden layer is determined iteratively based on the performance of the neural network. Number of neurons in each layer of neural network is shown in Table 1.

Resilient Backpropagation (RBP) algorithm is used for training the model. RBP is preferred over the most commonly preferred back propagation algorithm as it quickens the training process (Liu et al, 2012).

Figure 4: ACF and PACF Plot for IMF_3 Obtained Using CEEMDAN



The data is normalized using z scores as a preprocessing step for ANN. It helps to reduce the chances of neural network getting stuck in local optima and fastens the training process (Wu and Lo, 2010).

Similarly, the forecasts are obtained using ANN model without decomposing the data.

Table 1: Number of Neurons for Various IMFs and Residue Obtained Using EMD and CEEMDAN

IMFs	EMD			CEEMDAN		
	Input	Hidden	Output	Input	Hidden	Output
IMF_1	3	8	1	4	9	1
IMF_2	3	9	1	6	10	1
IMF_3	4	10	1	6	8	1
IMF_4	5	8	1	3	10	1
IMF_5	5	10	1	4	9	1
IMF_6	4	11	1	6	10	1
IMF_7	6	10	1	3	8	1
Residue	7	7	1	6	10	1

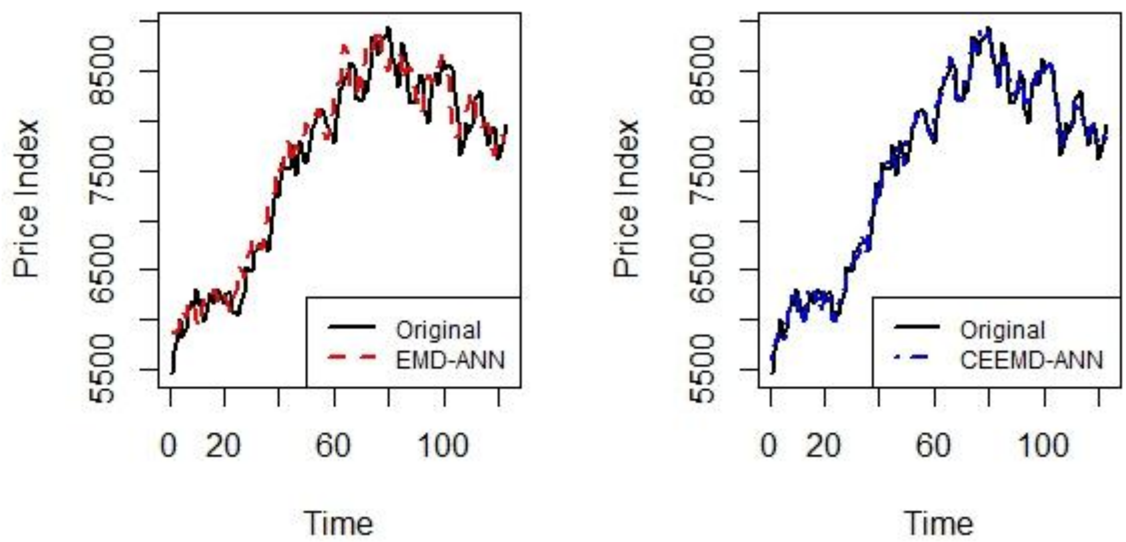
RESULTS AND DISCUSSION

Performance Measures

The performances of ANN model and both ensemble models are compared using two performance measures, namely, Root Mean Square Error (RMSE) and Directional Accuracy (DA). Error is defined as the difference between the original series and the forecasted series. Square root of mean of errors is known as Root Mean Square Error. Smaller the value of RMSE better are the forecasts. Determining whether the direction of forecasted series is same as the original series is important and is identified using directional accuracy. Higher the DA better are the forecasts. With respect to both performance measures, ensemble CEEMDAN-ANN model is better than remaining two models (Refer Table 2). Performance of EMD-ANN is slightly better than that of ANN model. We validate the results statistically using Wilcoxon Signed Rank Test (WSRT). Figure 5 represents the results of the 1-step ahead forecasts obtained using both EMD-ANN and CEEMDAN-ANN models.

Table 2: Performance Measures		
	RMSE	DA (%)
EMD-ANN	103.37	52.89
CEEMDAN-ANN	50.35	89.00
ANN	165.38	40.00

Figure 5: Forecasts Obtained Using EMD-ANN and CEEMDAN-ANN Models



Statistical Validity of Proposed Approach

The predictive capability of the proposed approach should be analyzed and validated statistically. This can be carried out using Wilcoxon Signed Rank Test (WSRT). WSRT is a distribution-free and non-parametric test that analyzes the predictive capability of two models based on their ranks and signs (Diebold and Mariano, 1995; Kao et al, 2013).

Two-tailed WSRT was carried out on RMSE values. Based on the results obtained (Refer Table 3), the null hypothesis of CEEMDAN-ANN and ANN models being same is not accepted since the z statistics value is beyond (-1.96, 1.96) at 99% confidence interval (alpha = 0.01). Similar results were observed in case of comparison between EMD-ANN and ANN models. Hence, it can be concluded that proposed ensemble models outperformed the ANN model.

Table 3: Wilcoxon Signed Rank Test between Hybrid EMD-ANN, Hybrid CEEMDAN-ANN, and ANN models				
	Ensemble EMD-ANN		Ensemble CEEMDAN-ANN	
	Z	Sign	Z	Sign
ANN	-4.188	+	-5.780	+
+: EMD-ANN > ANN, CEEMDAN-ANN > ANN -: EMD-ANN < ANN, CEEMDAN-ANN < ANN =: EMD-ANN = ANN, CEEMDAN-ANN = ANN				

Trading Decision

We can forecast the closing price of Nifty on following week using the proposed model with a reasonable accuracy. Now, using these forecasted values, investors can make their trading decisions. Few trading rules are defined. Here, the trading rules have been defined and illustrated for stock index; the same can be used for making buy/sell decisions for stocks as well.

Let F_k and F'_k be actual and forecasted close price on first trading day in the k^{th} trading week, respectively. Consider a scenario where the close price is expected to rise in the trading week *k* but it either falls or remains constant. We can use an error index E_k to represent such scenarios and is defined as follows:

$$E_K = \begin{cases} 1 & if \quad F'_k > F_{k-1} \quad and \quad F_k \le F_{k-1} \quad \forall \quad k = 2,3,...,n \\ 0 & otherwise \end{cases} \tag{14}$$

where *n* is the total number of weeks.

Based on the error index (Hsu, 2014), following three rules are used in this study:

Rule 1: IF (if an investor is not holding any stock on first trading day in the k^{th} week) AND ($F'_{k+1} > F_k$) THEN (s/he is advised to buy the stock on the next trading day in the k^{th} week)

Rule 2: IF (if an investor is holding any stock on first trading day in the k^{th} week) AND ($F'_{k+1} < F_k$) THEN (s/he is advised to sell the stock on the next trading day in the k^{th} week)

Rule 3: IF (the investor is holding any stock on first trading day in the k^{th} trading week) AND ($\sum_{j=0}^{2} E_{k-j} = 3$) THEN (s/he is advised to sell the stock on second trading day in k^{th} trading week)

According to first rule, since the price in the following week is expected to rise, the investor is advised to buy stocks if he does not hold any stock on first day of the trading week. If an investor is holding a stock on the first day of trading week and if the prices are forecasted to fall next week, then according to second rule, the investor is advised to sell his stock on next trading day. According to third rule, investor is advised to sell his held stocks if the forecasts are found to be wrong for three consecutive weeks.

The return on investment (ROI) in each trading strategy is calculated using the following equation:

$$ROI = \prod_{i=1}^{k} \frac{P_{S_i} - P_{B_i}}{P_{B_i}} \tag{15}$$

where, P_{Si} and P_{Bi} represent the selling price and buying price of the i^{th} transaction, respectively.

Here, we assume that the security (stocks or index) can be traded (sold and bought) at the opening price on the next trading day of the week as soon as buying/selling decisions are taken. The transactions based on the trading rules using the results of EMD-ANN model are illustrated in Table 4.

The closing price on August 31, 2015 is 7655.05, which is less than forecasted price of the following week (i.e., 7840.19), hence according to first rule, the investor is advised to buy stocks. On October 12, 2015, it is observed that forecast for next week is 8188.09 which is lower than the current close price (8238.15), hence applying second rule, the investor is advised to sell his stock next day (October 13, 2015) at its opening price (8121.95).

Table 4: Illustration of Trading Rules

Date	Close Price	Forecasted Close Price	Transaction	Transaction Date	E_k	Rule
31-08-2015	7655.05	7840.19	Buy at 7907.95	01-09-2015	1	1
07-09-2015	7789.30	8025.28	-	-	-	0
14-09-2015	7981.90	8036.78	-	-	-	0
21-09-2015	7868.50	8087.50	-	-	-	1
28-09-2015	7950.90	8165.05	-	-	-	0
05-10-2015	8189.70	8295.16	-	-	-	0
12-10-2015	8238.15	8188.09	Sell at 8121.95	13-10-2015	2	0
19-10-2015	8295.45	8036.79	-	-	-	0
26-10-2015	8065.80	7920.82	-	-	-	0
02-11-2015	7954.3	7928.30	-	-	-	0
09-11-2015	7762.25	7977.71	Buy at 7877.60	10-11-2015	1	1
16-11-2015	7856.55	7958.01	-	-	-	0
23-11-2015	7942.70	7809.99	Sell at 7837.00	24-11-2015	2	0
30-11-2015	7781.09	7659.66	-	-	-	0
07-12-2015	7610.45	7773.07	Buy at 7738.50	08-12-2015	1	1
14-12-2015	7761.95	7866.05	-	-	-	0
21-12-2015	7861.05	7847.61	Sell at 7829.40	22-12-2015	2	0

Similarly, since on November 9, 2015, the forecasted close price (7977.71) for next week is higher than current close price (7762.25), the investor is advised to buy stocks according to Rule 1. The process is carried out in the same way for rest of test dataset. Similarly, trading rules are applied for forecasts obtained using ensemble CEEMDAN-ANN model. ROI was calculated for EMD-ANN, CEEMDAN-ANN model and Buy-and-Hold strategy. ROI of CEEMDAN-ANN was 2.8% and 1.5% higher than that of Buy-and-Hold strategy. Trading decisions based on CEEMDAN-ANN model yielded higher ROI than EMD-ANN model. Buy-and-Hold strategy yielded the least returns among all three models.

CONCLUSION

Investment decision making is one of the difficult tasks. The paper presented two ensemble frameworks, namely, EMD-ANN and CEEMDAN-ANN model to predict close price of Nifty index for the following week. Further, trading rules were defined and illustrated to guide the investors with the trading decisions.

Ensemble models used EMD and CEEMDAN to decompose the original time series into various sub-series. Later, each decomposed series was forecasted using ANN for the respective ensemble models. The forecasted sub-series are aggregated to obtain final forecasts.

Performance measures, namely, directional accuracy and RMSE suggest that the performance of ensemble models is better in comparison to traditional ANN model. Further, the results of

WSRT confirm the superiority of predictive capability of ensemble models over traditional ANN model. Trading rules based on forecasts of ensemble models were defined and illustrated so that investors could make trading decisions. Return on investment of ensemble CEEMDAN-ANN model was found to be higher than that of EMD-ANN model and buy-and-hold strategy.

The ensemble forecasting model can be tested for high frequency data and new trading rules can be formulated for the same. A comparative study of model with data from developed and emerging economies would provide useful insights.

APPENDIX

Nomenclature

ACF	Auto Correlation Function
ANN	Artificial Neural Network
ARIMA	AutoRegressive Integrated Moving Average
CEEMDAN	Complete Ensemble Empirical Mode Decomposition with Adaptive Noise
DA	Directional Accuracy
DWT	Discrete Wavelet Transform
EMD	Empirical Mode Decomposition
GARCH	Generalised AutoRegressive Conditional Heteroskedasticity
IMF	Intrinsic Mode Function
PACF	Partial AutoCorrelation Function
RBP	Resilient Back Propagation
RMSE	Root Mean Square Error
ROE	Return On Equity
ROI	Return On Investment
SVR	Support Vector Regression
WSRT	Wilcoxon Signed Rank Test

REFERENCES

Atsalakis G, Valavanis K (2009) Surveying stock market forecasting techniques- Part II: Soft computing methods. *Expert Systems with Applications,* 36(3, Part 2):5932-5941.

Atsalakis G, Valavanis K (2013) Surveying stock market forecasting techniques- Part I: Conventional methods. Zopounidis C, ed., *Computation Optimization in Economics and Finance Research Compendium*, 49 – 104, New York: Nova Science Publishers, Inc.

Crowley P (2010) Long cycles in growth: Explorations using new frequency domain techniques with US data. Bank of Finland Research Discussion Paper No. 6/2010, Retrieved from http://dx.doi.org/http://dx.doi.org/10.2139/ssrn.1573641, February 16.

Diebold FX, Mariano RS (1995) Comparing predictive accuracy. *Journal of Business and Economic Statistics*, 13:253-265.

Fischer D, Jordan R (1987) *Security Analysis and Portfolio Management*, US: Prentice-Hall International.

Hsu CM (2014) An integrated portfolio optimisation procedure based on data envelopment analysis, artificial bee colony algorithm and genetic programming. *International Journal of Systems Science*, 45(12):2645-2664.

Huang N, Shen Z, Long S, Wu M, Shih H, Zheng Q, Yen N, Tung C, Liu H (1998) The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences,* 454(1971):903 - 995.

Huang NE, Wu ML, Qu W, Long SR, Shen SSP (2003) Applications of Hilbert Huang Transform to non-stationary financial time series analysis. *Applied Stochastic Models in Business and Industry*, 19(3):245-268.

Jothimani D, Shankar R, Yadav SS (2015) Discrete Wavelet Transform-Based Prediction of Stock Index: A Study on National Stock Exchange Fifty Index, *Journal of Financial Management and Analysis*, 28(2), 35-49.

Jothimani D, Shankar R, Yadav SS (2016) A comparative study of ensemble-based forecasting models for stock index prediction. *MWAIS 2016 Proceedings*, paper 5, Retrieved from http://aisel.aisnet.org/mwais2016/5, May 16.

Kao LJ, Chiu CC, Lu CJ, Chang CH (2013) A hybrid approach by integrating wavelet-based feature extraction with MARS and SVR for stock index forecasting. *Decision Support Systems*, 54(3):1228-1244.

Lahmiri S (2014) Wavelet low- and high-frequency components as features for predicting stock prices with backpropagation neural networks. *Journal of King Saud University - Computer and Information Sciences,* 26(2):218- 227.

Liu H, Chen C, Tian H, Li Y (2012) A hybrid model for wind speed prediction using empirical mode decomposition and artificial neural networks. *Renewable Energy,* 48:545-556.

Opitz D, Maclin R (1999) Popular ensemble methods: An empirical study. *Journal of Artificial Intelligence Research,* 11:169-198.

Theodosiou M (2011) Forecasting monthly and quarterly time series using STL decomposition. *International Journal of Forecasting,* 27(4):1178-1195.

Torres ME, Colominas MA, Schlotthauer G, Flandrin P (2011) A complete ensemble empirical mode decomposition with adaptive noise. *2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 4144-4147.

Wu G, Lo S (2010) Effects of data normalization and inherent-factor on decision of optimal coagulant dosage in water treatment by artificial neural network. *Expert Systems with Applications*, 37(7):4974-4983.

Wu Z, Huang NE (2009) Ensemble empirical mode decomposition: A noise-assisted data analysis method. *Advances in Adaptive Data Analysis*, 1(1):1-41.