

Stock Market Prediction System: A Wavelet based Approach

S. Saravanan and S. Mala*

Department of Management Studies, University College of Engineering, Anna University, BIT Campus, Tiruchirappalli, Tamil Nadu, India.

Received: 8 Sep. 2017, Revised: 13 Mar. 2018, Accepted: 23 Mar. 2018

Published online: 1 May 2018

Abstract: This work concentrates on computational approach for predicting the interval (number of trading days), a significant feature of stock market analysis using Haar Wavelet. A distinct model is proposed for predicting the high value of returns. The prime objective is to understand the trends using Haar wavelet and use this information to determine the interval for future direction. This model used 85 securities closing price and validated 4355 trading days. The results reported at 200 recent trading days with an average accuracy of 45.88% on 85 securities over a period of 15 years.

Keywords: Haar wavelet, National Stock Exchange (NSE), Stock Market Prediction, Wavelet Transform (WT).

1 Introduction

The objective of this work is to capture trends and similarities in order to help investors understand the current trading scenario, aiming the test of hypothesis, 'whether trading intervals have any significant impact on stock market data'. Studies suggest that, the stock analysis is still evolving, and these exist an opportunity for developing new evaluation technique and use of various data mining methods, which involves interval based analysis to make inference about the current market. Wavelet function is increasingly applied to these stock data sets because it permit experts to focus on time scale where the trading patterns would be important. Wavelets can be very much useful for distinguish trading patterns from other price movements. This work presents an interval based stock market analysis using Haar wavelet in NSE data set which can identify significant informative financial characteristics. This method can be opted as an alternative to analyze financial data which involves the effective use of stock market data and assist investors to make better decisions based on intervals.

Since lesser models are present to predict the state of stock market based on trading intervals, the efficient diagnosis of current stock market and its relationship with various financial trades will be available. These models

can become more crucial for the recent stock market changes and increases the need for financial care of stock-market. The stock market usually improve economical well-being of people. The financial market system typically generates large amount of fluctuating data, which might be analyzed for developing a better decision support system. To make a cutting edge on stock market, the hidden information in stock-market data is considered, which is largely untapped and to be identified. Financial data including numbers and charts must be analyzed for effective trading decisions, further assessing the feasibility of stock market data analysis using wavelet and specifically predicting the stock data based on trading intervals. This work introduces wavelets as a technique for stock prediction, the development and characterization of a method to assess repetitive stock patterns for detecting basic stock characteristics from NSE data set.

This paper first survey on the related work, introduce wavelets and describe the main stock market characteristics. We then detail the stock processing required to detect basic characteristics in stock market data and describe the extracted features and elaborate on the experiment. Finally, discuss the results with future work.

* Corresponding author e-mail: ursmala@gmail.com

2 Theoretical Background

Wavelets are applied in number of fields including biomedical signals, medical images, aural signal and video-signal [1,2,3]. Despite the considerable research in economics and finance, only limited researchers have identified the application of wavelets in stock markets [4, 5]. Wavelet analysis is a mathematical model very suitable for non-stationary data that transforms the time domain signal into a different domain for analysis and processing [6]. Haar wavelet are used in different applications. This work makes use of haar wavelet since most of the financial time series data are non-stationary. The economic applications of wavelets provide different insights into those application and report fruitful results. Transformations are applied to signals which obtain further information which is not visible in the raw signal. Traditional method time-series analysis focuses on time domain which requires a very strong assumption that data must be stationary and spectral analysis, which focuses on frequency-domain makes sense when the market activity is stable across the whole period [7].

A financial time series is a combination of components that operate on different frequencies. Unlike traditional methods, wavelet Analysis, provide useful information to economics and finance that help to overcome the limitations in traditional methods. U.S stock price behavior are analyzed by various waveform dictionaries to decompose the time series contained within three tick-by- tick foreign exchange rates [8,9]. The authors conclude that waveform dictionaries are most useful for analyzing data that are not stationary. Wavelets are used orthogonally to decompose the KOSPI and DJIA daily stock market indices into different respective time-scale components [10].

To quantify price spillovers among a wide range of regional stock markets on different time horizons wavelets are employed [11]. The relationship between future output and a variety of financial indicators using scale by scale based on wavelet multi resolution analysis [12]. Wavelets are used to study growth and volatility of GDP series over different time horizons, where the focus is on changes in the growth rates as well as the levels of GDP[13].

In the presence of wavelet analysis, the lead-lag relationship between financial variables and real economic activity and stock returns inflation on different time scales are studied [14,15]. Wavelet-based methods are used to forecast foreign exchange rates [16]. To detect and separate periodic components in time series the DWT and Scalo gram are used in spectral analysis [17]. The wavelets are applied to investigate the issue of market efficiency in the future markets for oil [18]. A wavelet-based prediction procedure is introduced and used to forecast market data for crude oil over different forecasting horizons. The CWT avoids one particular problem: in most of the literature from frequency domain, cut-off frequency band is arbitrary for the analysis. CWT

provides a continuous assessment of relationships or structures, as well as other observations. The wavelet analysis has the variety of algorithms. Depending on the data-analysis different wavelet algorithm are applied. In PVI non-linear system, adaptive wavelet neural network is employed to inherit time-varying system characteristics [19].

This work applies Haar-Wavelet (in close price for a stock), which is very fast and works well for the financial time series. Financial-time-series are non-stationary, cannot be described within a window by combination of sin and cos terms. Financial time series cannot be cyclical in a predictable fashion. Financial-time- series lend themselves to Haar wavelet analysis since graphs of financial time series tend to jagged, without a lot of smooth detail. The present work adopts CWT (for interval analysis) & DWT (for noise removal) using Haar wavelet as the mother-wavelet. By using CWT, from a very complicated function the user can extract components with a simpler structure. The user can study each small component instead of studying original function as the whole. Main advantage of DWT is that it inherits the same benefit of CWT but uses a limited number of translated and dilated versions of the original wavelet [20].

The stock market researchers have attempted to predict stock price using machine learning techniques and statistical models. The problem of parameter selection and over-fitting occurs in machine learning techniques [21,22]. The statistical models result in large error due to the assumption of time series data to be linear and stationary. Use of continuous time transform, every possible translation and scale-based decomposition models has been found to improve the accuracy of stock market prediction.

3 Empirical Tests

3.1 Sample Selection

The data set for this work is collected from the historical data on National Stock Exchange from the website of quandl

<https://www.quandl.com/data/XNSE-National-Stock-Exchange-of-India-Prices>.

Quandl has the richest collection of publicly available stock market data on the Internet. Quandl is a platform for economical, financial and other data, serving professional investors. More than 2000 stock data from National Stock Exchange of India is available in quandl. This work used the daily closing price of more than 80 security index. This data set involves an average of more than 3000 trading days.

3.2 Prediction Framework

Prediction analysis is needed to provide the knowledge and guidance to investors in terms of when to buy, sell or

hold the shares. This work attempts to design simple wavelet-based model for stock prediction in the future in terms of interval based on trading days. Issues in data preprocessing technique, architectural design, wavelet selection and performance measures have been considered carefully. The methods were all implemented off-line using MATLAB. The framework for stock prediction is illustrated in Fig. 1. The daily returns calculated from closing price are the significant input for this processing chain. Firstly, the noise from returns are removed by using Haar wavelet. The denoised-returns are given as input to the next stage, where intervals for positive stock market prediction is calculated using Haar wavelet. The results are further evaluated using confusion matrix.

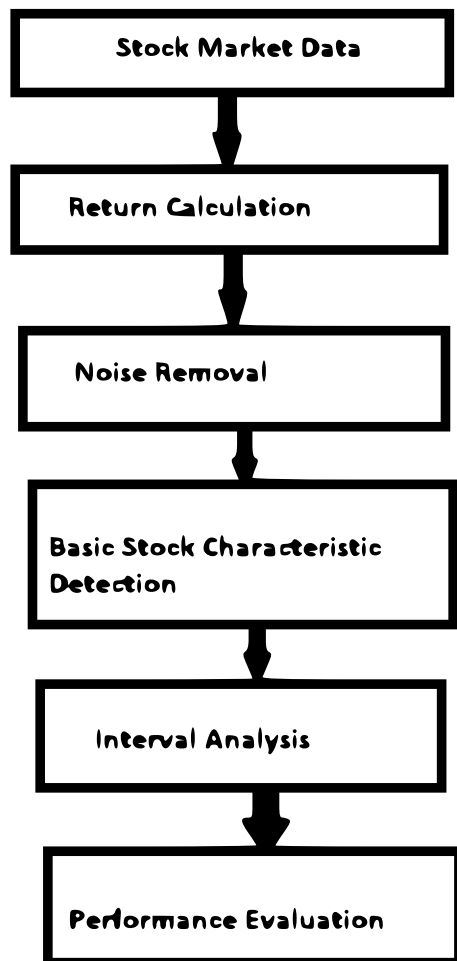


Fig. 1: Stock Market Prediction Framework

3.3 Daily Returns From Closing Price

The input to processing chain is the closing price of stock market historical data for a particular security. In the first stage, the rate of returns from closing price are calculated by the following formula depicted in (1). Returns from Closing price quotes the i th return of an asset from the period $T(i)$ to $T(i+1)$ and normalizes it with time interval, which is 1 between successive price observations.

$$Return(i) = \log \frac{times(i+1)}{(times(i+1) - times(i))} \quad (1)$$

The sample returns calculated using MATLAB for NSE_RCF security from March 1998 to June 2016 is shown in Fig. 2.

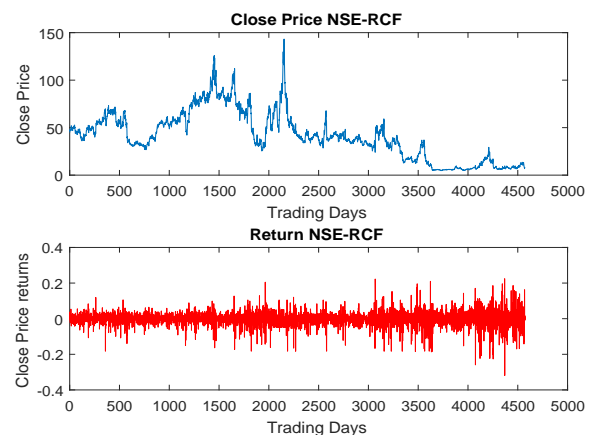


Fig. 2: NSE_RCF Closing Price and Returns

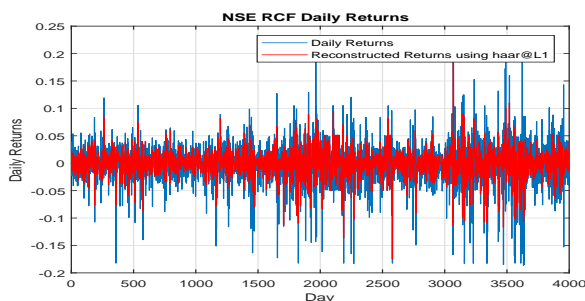
3.4 Noise Removal

Several stock market characteristics (peaks) are needed to be preserved from the noise removal methods. The steepness of stock market data edge needs to be retained in order to differentiate between profit/loss characteristics in stock. During diagnosis, return series are processed to remove noise (also known as 'market noise' caused by program trading) that tends to confuses or misrepresents genuine trends. We have used wavelet transform based noise removal and identified wavelet-based methods are suitable for visual inspection, Peak Signal to Noise Ratio (PSNR) and Mean Square Error (MSE). The algorithm used Haar Wavelet decomposition at level 1 on each component for removing noise. The DWT using Haar as mother wavelet has, Approximation Coefficients and Detail Coefficients. The choice of mother wavelet

Table 1: PSNR, MSE for sample securities using Haar Wavelet.

SECURITY	PSNR	MSE
1	81.476	0.000463
2	77.504	0.001155
3	83.937	0.000263
4	79.928	0.000661
5	81.978	0.000412
6	80.221	0.000618
7	84.013	0.000258
8	81.185	0.000495
9	80.308	0.000606

becomes important in analyzing a given time series for two reasons. Firstly, length of the discrete wavelet function determines how well it would be able to isolate features to specific frequency intervals. Secondly, it is being used to represent the hidden information contained in time series. The Detail Coefficients indicate the short bursts in the financial data which are more of impulsive reactions to news and events [22]. On the other hand, the Approximation Coefficients denote the average behavior of the indices during a long run and are considered for determining the relationship of movement between stock indices. The reconstructed signals and the daily raw returns for security NSE.RCF is shown in Fig. 3. The PSNR for NSE.RCF 79.447 and MSE for NSE.RCF 0.00073847 are calculated after reconstruction, and this shows that Haar wave function can be used for denoising daily stock market returns. Sample PSNR, MSE for some securities are listed in Table 1.

**Fig. 3:** NSE.RCF: Raw Signal and Reconstructed Signal

The PSNR and MSE for 85 different securities are analyzed before the detection of basic stock price characteristics (profit/loss) analysis. The results obtained for these securities show the current trading intervals playing the significant role in stock market prediction. 39 out of 85 securities show the impact of current 200 trading days. Fig. 4 list the highest values of PSNR(encircled) for 2 different sample securities.

PSNR	MSE	Data_From	Data_To	5 Average Peak Interval values	PSNR	MSE	Data_From	Data_To	1 Average Peak Interval values
1	82.6480	3.5333e-04	1	200	73	83.5469	2.8740e-04	1	200
2	83.9659	2.5911e-04	1	400	73	82.9521	3.2651e-04	1	400
3	84.1128	2.5223e-04	1	600	111.8000	82.2279	3.8830e-04	1	600
4	84.4863	2.3251e-04	1	800	108.7143	82.8328	3.5465e-04	1	800
5	84.9135	2.0977e-04	1	1000	108.7143	83.0256	3.2366e-04	1	1000
6	85.1776	1.9739e-04	1	1200	143.2500	83.0817	3.2204e-04	1	1200
7	85.3446	1.8594e-04	1	1400	133.7000	83.0343	3.2333e-04	1	1400
8	85.1076	2.0060e-04	1	1600	123.5000	82.8534	3.5289e-04	1	1600
9	84.9871	2.0624e-04	1	1800	117.3857	82.3244	3.8075e-04	1	1800
10	84.3532	2.3865e-04	1	2000	81.5417	81.5308	4.5944e-04	1	2000

Fig. 4: PSNR,MSE for 2 different securities with Intervals

3.5 Basic Stock Characteristic Detection

From the reconstructed returns, stock market prediction intervals are detected. The corresponding positive/negative characteristics identified by the detection algorithms are the basis for this work. Based on the proposed method results are evaluated for NSE.RCF data set. The profit/loss can be detected from the processed stock market returns. Intervals forming the basis of all stock features are used for prediction. Robustness of the algorithm for detecting is a key to achieve good performance prediction. Profit detection is particularly important because the entire stock market data are reliable. For the profit identification, we used Continuous wavelet transform Profit Detection (CWTPD). The input to CWTPD are the denoised stock market returns. CWTPD first computes the continuous 1D wavelet coefficients at scale 10 using Haar as mother wavelet.

Let R be one of the return components and ϕ the mother wavelet. The wavelet coefficient $C_p^s(R)$, at scale s and position p , is defined as in (2).

$$C_p^s(R) = \int R(t) \frac{1}{\sqrt{s}} \phi\left(\frac{t-p}{s}\right) dt. \quad (2)$$

By application of specific threshold th_{pd} on the coefficients $C_p^s = C_p^{10}(R)$, CWT-PD creates a vector T with elements T_i as in (3).

$$T_i = \begin{cases} 1, & \forall i : C_i(R) > th_{pd} \\ 0, & \text{otherwise.} \end{cases} \quad (3)$$

This step divides returns into profit and non-profit segments. Profit segments greater than the threshold are considered. CWT-PD checks the return direction of each segment. The direction is derived from the sign of corresponding element in T . The Fig. 5 illustrates raw returns (R) and Profit segments with application specific threshold 0.05.

Table 2: Various security features with average intervals.

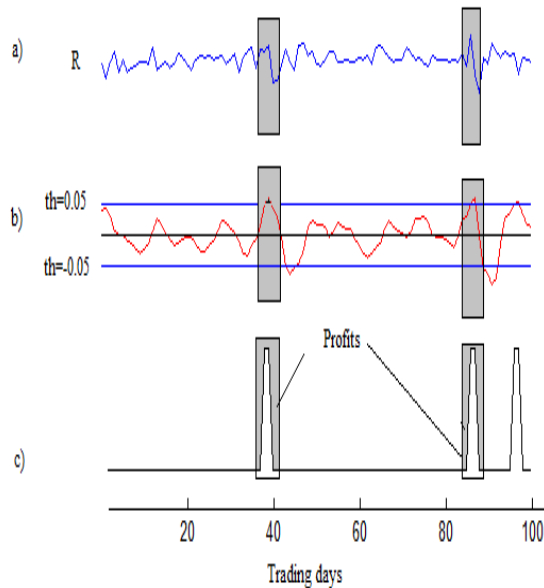
Security	PSNR	MSE	Data.Interval	AvgPeakDist
1	81.476	0.000463	1-1400	42
2	77.504	0.001155	1-400	32
3	83.937	0.000263	1-400	55
4	79.928	0.000661	1-200	23
5	81.978	0.000412	1-200	49
6	80.221	0.000618	1-1600	26
7	84.013	0.000258	1-200	183
8	81.185	0.000495	1-1200	48
9	80.308	0.000606	1-2400	31

Table 3: Confusion matrix(Actual Vs Calculated Results).

Actual Results	TRUE	FALSE	TOTAL
Positive	1066(TP)	1134 (FP)	2200
Negative	1071(TN)	1085(FN)	2156
Total (TP+FP+TN+FN)	2137	2219	4356
Correct Predictions (TP+FN)			2151
Accuracy ((TP+FN)/Total)			49.38

Table 4: Evaluation Results.

1	NumberOfObservations:	4356
2	ControlClasses:	2
3	CorrectRate: Accuracy	0.4938
4	ErrorRate:	0.5062
5	Sensitivity:	0.48455
6	Specificity:	0.50325
7	PositivePredictiveValue:	0.49883
8	NegativePredictiveValue:	0.48896

**Fig. 5:** Continuous Wavelet Transform - Profit Detection (CWT_PD) a) Raw Returns (NSE_RCF) b) Transformed wavelet with threshold.05 c) Marker vectors for profits

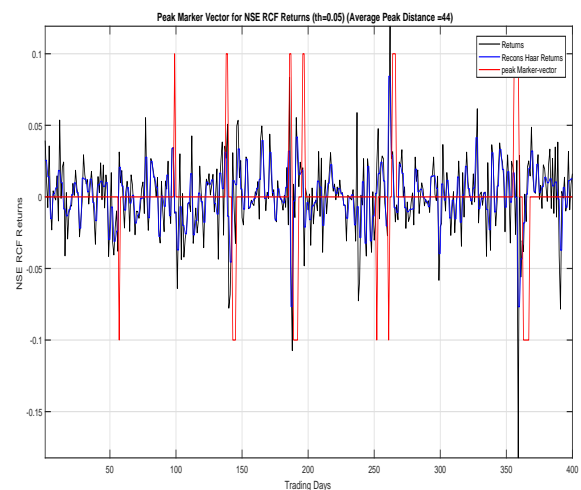
3.6 Interval Analysis

Various security features along with its data intervals and average peak distance are listed in Table 2. To evaluate the CWT-PD algorithm, we performed an experiment considering NSE_RCF stock market data with more than 4000 trading days. Using a fixed sequence to simplify labeling of individual profit, the sequence was comprised of 0 for non-profit segments and 1 for profits. The predictions are listed in Table 3. Comparing profits with the annotated ground truth, we calculated true positives (TP), true negative (TN), false positives (FP) and false negatives (FN), to calculate the accuracy.

We have taken totally 4356 trading days of NSE_RCF for interval based analysis. By using the proposed method, 2151 observations are predicted correctly.

Table 4 suggests that trading day intervals can be used to predict the future direction of stock market analysis. Fig.

1 shows average peak distance for security NSE_RCF is 44. This implies that an user may expect profit after 44 trading days for the security NSE_RCF.

**Fig. 6:** NSE_RCF Interval Analysis (AvgPeakDist 44)

3.7 Results And Discussion

This work proposed a framework for stock prediction using trading day intervals with 49.38% accuracy in the real data set. The objective of this work is to capture the trends and similarities in the movements and activities of the Indian Stock Market NSE in order to help the

investors understand, the current trading scenario. Aiming the test of hypothesis, 'whether trading intervals have any significant impact on stock market data'. Considering the hypothesis that these would be an impact of trading intervals with respect to returns for desirable investments and profitability. The developed algorithm for detecting profit using trading intervals in stock market data has been proved achieving the accuracy of 49.38% for NSE_RCF data set. The experimental evaluation shows sensitivity 48.45% and specificity 50.32% for NSE_RCF data set. The Interval analysis shows the number of trading days for various securities have significant impact on stock market data.

4 Conclusion

Application of Haar wavelet on stock market data was effectively implemented. The result indicates the important relationship between stock market prediction and trading intervals which was one of the main objectives of this work. With increased frequency of stock market data set we expect further testing on effectiveness of proposed work. We would like to extend our work further for on-line stock analysis. Trading intervals have provided new insight into stock market problem. The most important idea of this work is the feasibility of stock market prediction by trading intervals using wavelet. We have used only a selected number of stock market data for the experimental scenario which can be considered further to on-line stock.

Acknowledgement

The authors are grateful to the anonymous referee for a careful checking of the details and for helpful comments that improved this paper.

References

- [1] Seung-Min Park, Tae-Ju Lee and Kwee-Bo Sim, Heuristic Feature Extraction Method for BCI with Harmony Search and Discrete Wavelet Transform, *International Journal of Control, Automation, and Systems*, vol. 14, no. 6, pp. 1582-1587 (2016).
- [2] T. Sivalakshmi and G. Sreenivasulu, Comparative Analysis of Different Wavelets for EEG Signal Denoising, *International Journal of Innovative Research in Science, Engineering and Technology*, Vol. 6, No. 1, pp. 594-599 (2017).
- [3] S. Mala and K. Latha, Electrooculography de-noising: wavelet based approach to reduce noise, *International Journal of Advanced Engineering Technology*, Vol. 7, No. 2, pp. 482-487 (2016).
- [4] Ramsey, J. B., and Anderson, H. M., U.S. and Canadian industrial production indices as coupled oscillators, *Journal of Economic Dynamics and Control*, Vol. 26, pp. 33-67 (2002).
- [5] S. Skander, On the Use of the Wavelet Decomposition for Time Series Prediction, *Neurocomputing*, Vol. 48(1-4), pp. 267-277 (2002).
- [6] Y. Motohiro, Measure Business Cycle: A Wavelet Analysis of Economic Time Series, *Economics Letters*, vol. 100, pp. 208-212 (2008).
- [7] Tian-Xiao He, Wavelet Analysis and Applications in Economics and Finance, Research & Reviews, *Journal of Statistics and Mathematical Sciences*, Vol. 1, no. 1, pp. 22-37 (2015).
- [8] Ramsey, J. B., Usikov, D. and Zaslavsky, G. M., An Analysis of U. S. Stock Price Behavior Using Wavelets, *Fractals*, vol. 3, no. 2, pp. 377-389 (1995).
- [9] Ramsey, J. B., and Semmler, W., Interest Rate Spreads and Output: A Time Scale Decomposition Analysis Using Wavelets, *Computational Statistics and Data Analysis*, Vol. 76, no. 2, pp. 282-290 (2014).
- [10] Lee, H. S., International Transmission of Stock Market Movements: A Wavelet Analysis, *Applied Economics Letters*, Vol. 11, no. 3, pp. 197-201 (2004).
- [11] James. B. Ramsey, Wavelets in Economics and Finance: Past and Future, *Studies in Nonlinear Dynamics & Econometrics*, Vol. 6, no. 3, pp. 1-29 (2002).
- [12] Fernandez, V., The International CAPM and a Wavelet-Based Decomposition of Value at Risk, *Studies in Nonlinear Dynamics & Econometrics*, Vol. 9, no. 4, pp. 1-37 (2005).
- [13] Maslova, I., Onder, H. and Sanghi, A., Growth and Volatility Analysis Using Wavelets, *Policy Research Working Paper 6578*, The World Bank, (2014).
- [14] Kim, S. and In, F. H., The Relationship Between Stock Returns and Inflation: New Evidence from Wavelet Analysis, *Journal of Empirical Finance*, Vol. 12, no. 3, pp. 435-444 (2005).
- [15] Kim and In, F. H., The Relationship Between Financial Variables and Real Economic Activity: Evidence From Spectral and Wavelet Analyses, *Studies in Nonlinear Dynamics & Econometrics*, Vol. 7, no. 4, pp. 1-18 (2003).
- [16] Wong, H., Ip, W., Xie, Z. and Lui, X., Modelling and Forecasting by Wavelets, and the Application to Exchange Rates, *Journal of Applied Statistics*, Vol. 30, no. 5, pp. 537-553 (2003).
- [17] Arino, M. A., Morettin, P. A. and Vidakovic, B., Wavelet Scalograms and Their Applications in Economic Time Series, *Brazilian Journal of Probability and Statistics*, Vol. 18, pp. 37-51 (2004).
- [18] Yousefi, S., Weinreich, I. and Reinartz, D., Wavelet-Based Prediction of Oil Prices, *Chaos, Solitons & Fractals*, Vol. 25, no. 2, pp. 265-275 (2005).
- [19] Lai, W., Wavelet Theory and Its Applications in Economics and Finance, PhD Thesis., University of Leicester, (2015).
- [20] Hung-Yi Chen and Jin-Wei Liang, Adaptive Wavelet Neural Network Controller for Active Suppression Control of a Diaphragm-Type Pneumatic Vibration Isolator, *International Journal of Control, Automation, and Systems*, Vol. 15, no. 3, pp. 1456-1465 (2017).
- [21] Masset p., Analysis of Financial Time-Series Using Fourier and Wavelet Methods, University of Fribourg (Switzerland)- Faculty of Economics and Social Science, (2008).
- [22] Jothimani, D., Shankar, R and Yadav, S.S., Discrete Wavelet Transform-Based Prediction of Stock Index: A Study on National Stock Exchange Fifty Index, *Journal of Financial Management and Analysis*, Vol. 28, no. 2, pp. 35-49 (2015).



S. Saravanan is Head, Dept. of Management Studies, Bharathidasan Institute of Technology, Anna University, Tiruchirappalli. He received his Ph.D. in Business Administration from Anna University, India. His area of specialization is Finance and operations. He

has published research articles in reputed international journals of finance and management. He has published 3 books.



S. Mala pursuing her Ph.D in Anna University, Chennai and awarded NIXI Fellowship for Young Internet Professionals from National Internet eXchange of India (NIXI) in 2011. Her research interests include Data Mining, Feature Selection, and Stock analysis.