

Security Index Forecasting Using Multi Step Ahead Wavelet Neural Network and Wavelet Garch : A Comparative Study



Swarnava Mitra*, Atanu Das**

© Institute of Management Studies, Noida
Online access at www.publishingindia.com

Abstract

Inclusion of market noise is a common assumption in financial time series analysis. Removal of such noise significantly improves the forecasting performance of a model. This paper has used wavelet transform technique to denoise a financial time series. After denoising a multi step ahead forecasting technique based on a feed forward neural network is used for forecasting on 5 days, 10 days and 15 days horizon. The performance of the wavelet neural network model is then compared with wavelet GARCH(1,1) model empirically with data from Indian security market. The RMSE results show that former outperforms the later.

keywords — wavelet transform, feed forward neural network, early stopping, GARCH.

1. Introduction

Forecasting performance of conventional time series models has limitations because of time varying nature of included volatility. Researcher in computational finance are engaged in finding better models for forecasting problems over the years. Their accuracy can be further increased by denoising the time series in such a way that the essential information is retained. Hu Tao[1] has used a hybrid neural network model for forecasting floating foreign exchange rate. The wavelet coefficients are used as features for training the neural network. A genetic algorithm is used for selecting the best parameters for the neural network and finding the best length of the wavelet analysis coefficients. Kodogyannis et al [2] has compared the performance of fuzzy neural network and dynamic neural networks for one step and multi step ahead predictions of foreign exchange rates of US Dollar vs. Great British Pound. He et al [3] has used wavelet denoising method to separate the noise from the data and has used autoregressive moving average (ARMA) Box and Jenkins's model to find the conditional mean and generalized autoregressive conditional heteroskedastics to find the conditional variance to build a parametric Value at Risk (VaR) model, for a portfolio of asset backed securities from Chinese real estate market. But the comparative study of the forecasting power of neural network and a GARCH model fed with wavelet denoised time series has

*School Of Education Technology, Jadavpur University, Kolkata, India

**Dept. of CSE & IT, Netaji Subhash Engineering College, Kolkata, India

not been addressed to best of our knowledge.

In this work, at first daily closing values of S&P CNX Nifty index is denoised using global threshold algorithm, and the denoised index data is fed into a multi step ahead [3] time delayed [4] feed forward neural network with early stopping [5]. The time horizons considered for multi step ahead prediction are 5, 10 and 15. Then a conventional generalized autoregressive conditional heteroskedastics (GARCH)[12] model fed with same wavelet denoised time series with three different mother wavelets Daubechies, Coiflet and Symlet. A comparative study of their root mean square error (RMSE) is carried out to understand their forecasting efficiency.

Rest of the paper has been arranged in the following order, at first wavelet transform and signal denoising using wavelet transform is explained. After that feed forward neural network, concept of time delay used in feed forward neural network in time series forecasting and early stopping strategy are illustrated. Then a comparative study is done with GARCH(1,1) model. The paper ends by drawing conclusion and future scope for improvising this model.

2. Wavelet Transform

In wavelet transform the time domain signals is mapped into frequency-time domain and thus preserving both frequency and time domain information. The main idea of wavelet transform is to measure the degree of similarity between the original signal and the basic function of wavelet transform, called the mother wavelet, through wavelet coefficients computation. The calculation process is performed on shifted version of the mother wavelet. Continuous wavelet transform (CWT) is defined as the convolution between the original signal $s(t)$ and a wavelet.

$$W_{\psi}(a, b) = \int s(t) \psi_{ab}(t) dt = \frac{1}{\sqrt{a}} \int s(t) \psi\left(t - \frac{b}{a}\right) dt \quad (1)$$

where a is the scaling factor, b is the translation parameter and The mother wavelet is given by:

$$\psi_{a,b} = \frac{1}{\sqrt{a}} \psi\left(t - \frac{b}{a}\right) \quad (2)$$

The discrete wavelet transform (DWT) coefficients are usually sampled from the CWT on a dyadic grid, choosing parameters of translation $b=n \times 2^m$ and scale $a=2^m$ The mother wavelet in DWT is defined as:

$$\psi_{mn}(t) = \frac{1}{\sqrt{2^m}} \left(\frac{t - n \cdot 2^m}{2^m} \right) \quad (3)$$

DWT analyses the signal by decomposing it into its coarse and detailed information, which is accomplished by using successive high-pass and low-pass filtering operations, on the basis of the following equations:

$$Y_{high}(k) = \sum x[(n)] h(2k - n) \quad (4)$$

$$Y_{low}(k) = \sum x(n) g(2k - n) \quad (5)$$

where Y_{high} and Y_{low} are high-pass and low-pass filters with impulse response h and g , respectively. There are many wavelet families like Daubechies, Coiflet, Symlet and their properties are characterized in [7].

2.1 Denoising with Wavelets

The underlying model for the noisy signal is basically of the following form

$$y(t_i) = g(t_i) + \sigma \xi_i \quad (6)$$

where $y(t_i)$ is an empirically recorded signal with an $g(t_i)$ and ξ_i as deterministic and stochastic components. Also ξ_i s are independent Gaussian with zero mean and unit. σ represents the intensity of the noise with the signal. The objective of denoising is to suppress the noise part of the signal y and to recover g . The denoising procedure is done in the following way. Three mother wavelets are chosen Daubechies 5 or db5, Coiflet 4 or coif4 and Symlet 7 or Sym 7. Now the maximum decomposition level of the time series is obtained (in our implementation we used MATLAB 7 for this purpose), which is 6.

Signal denoising using wavelet transform is represented using periodised wavelets bases on $[0, 1]$. Also, the basis functions are generated by dilation and translation of a compactly supported scaling function Φ , also called father wavelet and a familiar mother wavelet function. The decomposition using father and mother wavelet is given in following equation (11)

$$g(t) = \sum \alpha_{j_0 k} k(t) + \sum \sum \beta_{jk} j_0 \geq 0 \quad t \in [0, 1] \quad (7)$$

where j_0 is a primary resolution level, and j_{0k} and j_k are calculated as the inner products shown in following equations (8) and (9)

$$\alpha_{j_0 k} = \langle g, \phi_{j_0 k} \rangle = \int g(t) \phi_{j_0 k}(t) dt \quad j_0 \geq 0 \quad k = 0, 1, 2, \dots, 2^{j_0-1} \quad (8)$$

$$\beta_{jk} = \langle g, \psi_{jk} \rangle = \int g(t) \psi_{jk}(t) dt \quad j \geq j_0 \geq 0 \quad k = 0, 1, 2, \dots, 2^{j-1} \quad (9)$$

When the discrete wavelet transform is used, the coefficients $C_{j_0 k}$ (discrete scaling coefficients) and d_{jk} (discrete wavelet coefficients) are used instead of the continuous parameters j_{0k} and j_k . Finally, when DWT is applied to equation (6), following expressions (10) and (11) are obtained

$$C_{j_0 k} = C_{j_0 k} + \sigma \xi_{j_0 k} \quad k = 0, 1, 2, \dots, 2^{j_0-1} \quad (10)$$

$$d_{jk} = d_{jk} + \sigma \xi_{jk} \quad k = 0, 1, 2, \dots, 2^{j-1} \quad (11)$$

Now from level 1 to 6, a threshold value is calculated. (from MATLAB 7) and Global Threshold algorithm is applied [8]. In the reconstruction phase the signal is reconstructed based on original approximation coefficients of level 6 and modified detail coefficients from level 1 to 6. Figure 1. below shows the

raw S&P CNX Nifty and the denoised version using db5 and global threshold algorithm

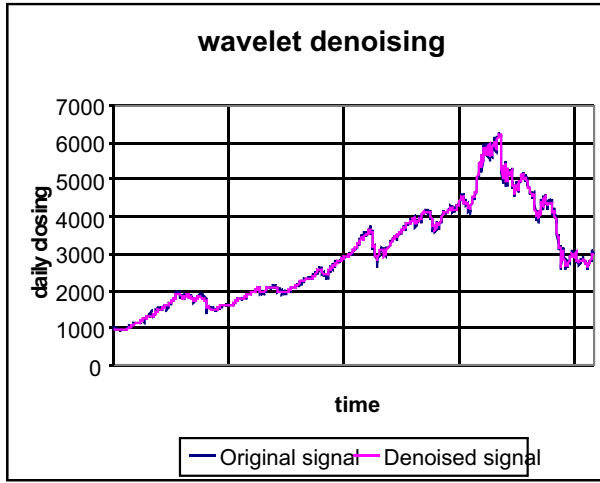


Figure 1: Denoised signal and original signal

3. Feed Forward Neural Network

Feed forward neural network is a type of neural network architecture, where input layer of neurons is connected to the output layer through one or more layers of intermediate neurons. The training process of neural networks involves adjusting the weights till a desired input/output relationship is obtained. The majority of learning algorithms used are based on the Widrow-Hoffback-algorithm. Each of these multi layered feed forward network is applied through activation functions. The overall mapping is thus characterized by a composite function relating to feed forward network's inputs to output [12] given

$$O = f_{composite} \quad (12)$$

Using p-mapping layers in a p+1 layer feed forward net yields

$$O = f^{LP} \left(f^{L_p-1} \dots (f^1(i)) \right) \quad (13)$$

Thus the interconnection weights from unit k in L_1 to unit i in L_2 are $W^{L_1 \rightarrow L_2}$. If hidden units have a sigmoid activation function, denoted as f^{sig} given by

$$O^{L_2} = \sum W^{L_1 \rightarrow L_2} \left(f^{sig} \left(\sum W^{L_0 \rightarrow L_1} \right) \right) \quad (14)$$

Above equation illustrates neural network with composition of non-linear function. The neural network forecasting can be described as follows

$$Z_{K+1} = NN(Z_k, Z_{k-1}, \dots, Z_{K-d}, e_k, e_{k-1}, \dots, e_{k-d}) \quad (15)$$

where Z_i is either original observations, and i, e are residuals. The processing of input data and the number of e 's and Z 's are

needed to determine the performance of the forecast. Standard back propagation algorithm [9] (which is a gradient descent algorithm) is used here for training the neural network, in which the network weights are moved along the negative of the gradient of the performance function. Thus for each movement we have,

$$X_{K+1} = X_k - a.k.g_i \quad (16)$$

There are a number of variations on the basic algorithm that are based on other standard optimization techniques such as conjugate gradient. But here Liebenberg-Marquardt optimization algorithm [10] has been used. Like Quasi-Newton methods, the Liebenberg-Marquardt algorithm is designed to approach second order training speed without the computation of Hessian matrix where the performance function has the form of a sum of squares that is

$$F(w) = \frac{1}{2} e^T . e = \frac{1}{2} e^T(w) e(w) = \frac{1}{2} \sum \sum (O_{ij} - R_{ij})^2 \quad (17)$$

where $W=[W_1, W_2, \dots, W_n]$ consists of all weights of the network.

3.1 Multi step ahead prediction and early stopping

The recursive relation between inputs and outputs in multi step (MS) prediction is defined as

$$\gamma_{t+p} = F(\gamma_{t+p-1} + \gamma_{t+p-2} + \dots + \gamma_t + \dots + \gamma_{t+p-s}) \quad (18)$$

where p is the MS prediction horizon, s is the input dimension, and γ_{t+p} is an estimate of the output at time-step $t+p$. The prediction accuracy deteriorates very quickly with increased p . because of which the method of early stopping is used. An approach to improve the prediction accuracy is to enhance the number observations. This approach requires the data set to be divided into three subsets: training, test, and verification sets. The training and the verification sets are the norm in all model training processes. The test set is used to verify the trend of the prediction accuracy of the model trained at some stages of the training process. At much later stages of training process, the prediction accuracy of the model may start worsening for the test set. This is the stage when the model should cease to be trained to overcome the over-fitting problem.

4. Empirical Data

The dataset considered in this investigation is S&P CNX Nifty index of National Stock Exchange of India for the period 1st Jan 2003 to 31st March 2009. This recent period is chosen because during this period Indian market has undergone a steep rise followed by recession in line with world market. This is given in following Figure 2

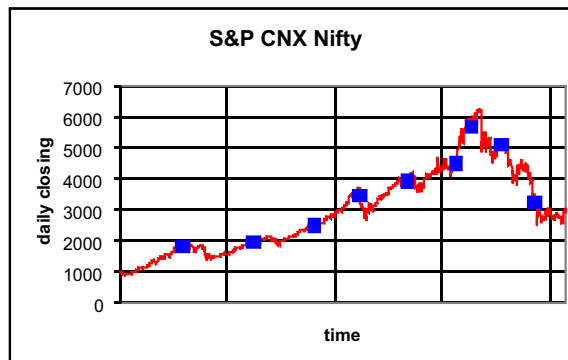


Figure 2. S&P CNX Nifty daily closing for the period 2003 to 2009

5. Outline of The Proposed Model

Based on the above theory the working flow is

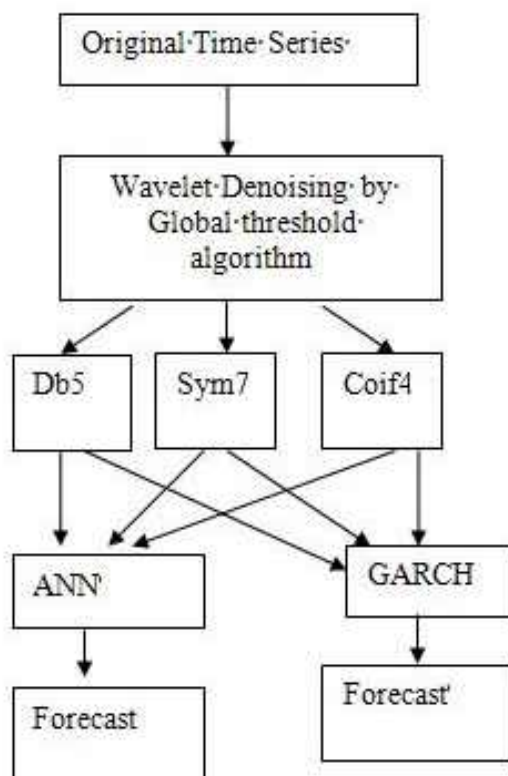


Figure 3. Work flow of Wavelet-ANN and Wavelet-GARCH methods.

6. Proposed Wavelet-ANN Method

- Load Time series data (or signal)
- Find the threshold value, threshold rule and the maximum level to which the signal can be decomposed.
- Denoise the signal using global threshold algorithm for 3 different mother wavelets: Coiflet, Symlet and Daubechies.

- Feed the denoised signal into the neural network.
- In building our neural network model the following parameter is initialized: Number of Hidden layers: 5, Maximum lag for time delay: 5, Threshold function: 'purelin', 'tansig', Learning incremental rate: 1.05, Learning decremented rate: 0.7, Number of epochs: 100, Goal: 1×10^{-5} .
- The number of time delays is then defined.
- The network is then trained using back propagation algorithm using Levenberg-Marquardt optimization.
- Use early stopping strategy as with the increase of multi step prediction horizon accuracy deteriorates.
- Calculate RMSE.

7. Results

The following figure shows the forecasting error plot of the test set and the residuals generated during forecasting.

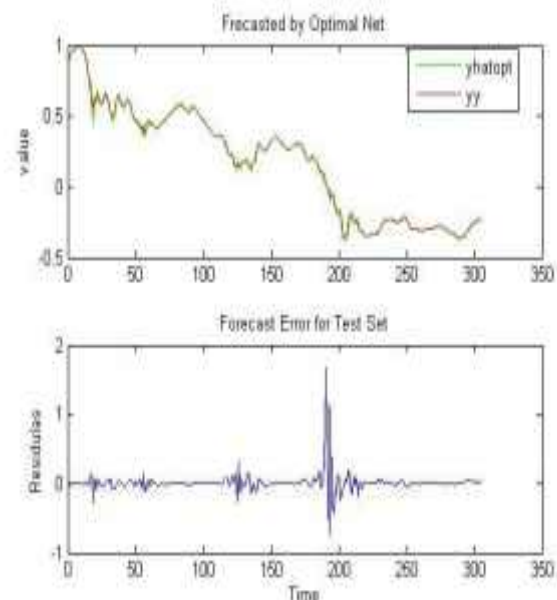


Figure 4: Forecasting error for test set and residuals generated

The performance of each model is evaluated using root mean square error for three different forecast horizons 5 day, 10 day and 15 day.

	Db5	Coif4	Sym7
ANN	0.009596	0.008390	0.008735
GARCH	0.022724	0.023123	0.023461

Table 1: For 5 days time Horizon

	Db5	Coif4	Sym7
ANN	0.009435	0.009261	0.009234
GARCH	0.022724	0.025612	0.023123

Table 2: For 10 days time Horizon

	Db5	Coif4	Sym7
ANN	0.009562	0.008418	0.009226
GARCH	0.022735	0.023567	0.023056

Table 3: For 15 days time Horizon

7. Conclusion

This work compared the forecasting efficiency of two techniques both combined with wavelet denoising. It is seen that the feed forward neural network based technique outperforms the conventionally used econometric model (like GARCH) based technique. This is particularly shown empirically with data combining a period of boom as well as recession together with high volatility. This work is based on a data of a general index of the Indian's largest stock exchange namely the NSE which incorporates almost all the features of emerging markets [11]. However other sector specific indices of NSE and BSE (Bombay Stock Exchange) can also be tested in a similar fashion and their comparison remains left for future work. Testing the efficiency of a forecasting technique (especially artificial intelligence based techniques) specific to a country or market (i.e. country specificity or market specificity) always remains an interesting research topic by which some specific behavioral issues are resolved. This work can also be taken as an effort in that direction.

8. References

1. Hu Tao "A Wavelet Neural Network Model for Forecasting Exchange Rate Integrated with Genetic Algorithm", *IJCSNS International Journal of Computer Science and Network Security*, VOL.6 No.8A, August 2006.
2. Kodogiannis V, Lolis A., "Forecasting financial time series using neural networks and fuzzy system base techniques" *Neural Computing and application* ISSN 0940-0643, vol 11, 2002.
3. Kaijian He, Chi Xie, Shou Chen, Kin Keung La "Estimating VaR in crude oil market: A novel multi-scale non-linear ensemble approach incorporating wavelet analysis and neural network" *Neurocomputing* 72, 3428–3438, 2009.
4. Aaron Mccoy, Tomas Ward, Seamus Mcloone, Declan Delaney, "Multistep-ahead neural-network predictors for network traffic reduction in distributed interactive applications" (TOMACS), Volume 17, Issue 4 Article No. 16, September 2007.
5. Renate Sitte, Joaquin Sitte, "Neural Networks Approach to the Random Walk Dilemma of Financial Time Series", *Journal of Applied Intelligence*, ISSN 0924-669X (Print) 1573-7497 (Online), Volume 16, Number 3, May, 2002.
6. Lutz Prechelt "Automatic early stopping using cross validation: quantifying the criteria", *Neural Networks* Volume 11, Issue 4, Pages 761-767, June 1998.
7. Ingrid Daubechies: *Ten Lectures on Wavelets*, SIAM 1992.
8. Donoho, D.L., "De-Noising by soft-thresholding," *IEEE Trans. on Inf. Theory*, vol. 41, 3, pp. 613-627, 1995.
9. Takeshi Kamio, Shinichiro Tanaka, Mititada Morisue, "Back propagation Algorithm for Logic Oriented Neural Networks" *IJCCN*, 2000.
10. Jinu Tian, Yan Lin, "Short term electricity price forecasting base on rough sets and improved SVM" *Proceedings of 2009 International workshop on knowledge discovery and data mining*, 2009.
11. Atanu Das, Pramatha Nath Basu, Tapan Kumar Ghoshal, "Stochastic Volatility Model for Indian Security Indices: VaR estimation and Backtesting", *The Indian Journal of Finance*, Vol III, No 9, September, 2009.
12. Atanu Das, Pramatha Nath Basu, Sudev C. Das, "A re-look at the VaR computation method recommended by National Stock Exchange of India", *The ICAI Journal of Applied Finance*, Vol. 14, No. 9, pp. 45-53, September 2008.