A Neuro-wavelet Method for the Forecasting of Financial Time Series

Luis F. Ortega

***Abstract*—We propose a wavelet neural network model (neuro-wavelet) for the short-term forecast of stock returns from high-frequency financial data. The proposed hybrid model combines the inherent capability of wavelets and artificial neural networks to capture non-stationary and non-linear attributes embedded in financial time series. A comparison study was performed on the modeling and predictive power among two traditional econometric models and four different dynamic recurrent neural network architectures. Several statistical measures and tests were performed on the forecasting estimates and standard errors to evaluate the predictive performance of all models. A Jordan net which used as input to the neural network the coefficients resulting from a non-decimated Haar wavelet-based decomposition of the high and low stock prices showed consistently to have a superior modeling and predictive performance over the other models. Reasonable forecasting accuracy for one, three, and five step-ahead horizons was achieved by the Jordan neuro-wavelet model.**

***Keywords*—dynamic neural networks, neuro-wavelets, wavelet decomposition, time series forecasting.**

I. Introduction

Financial time series, in general, are characterized by a nonstationary and nonlinear behavior. These undesired characteristics are mainly fashioned by the participation of a wide range of agents operating simultaneously in financial markets at different time-horizons of investment or scales [5]. The aggregate result is a combination of long and short memory processes embedded in a single complex signal that is not easy to analyze [13]. Additionally, financial time series, in particular those generated by high-frequency data, rarely depict a regular probability distribution, and involve numerous, unobvious, and indiscernible variables. The challenge, under this complicated scenario, is to develop suitable methods for the process of information extraction that needs to be performed in order to build adequate models for simulation and forecasting purposes.

This article presents a methodology that is sui generis in that it makes use of the synergy happening between tools borrowed from disciplines others than finance and economics to develop a resourceful approach to the analysis of high-frequency financial data. The methodology is an amalgamation of a wavelet transformation and a dynamic neural network that leads to the proposed wavelet neural network (WNN) or neuro-wavelet net model.

Discrete wavelet transforms (DWT) offer the capability of capturing key features of a process with a limited number of coefficients by using a set of orthogonal basis for the transformation of the signal. Particularly, a DWT is able to capture features that are localized in time such as the ones found in *nonstationary* signals. Some articles describing applications of wavelets to finance are [1], [3], [12], and [17]. Neural networks (NN), on the other hand, are data-driven self-adaptive methods that have the capability to extract essential parameters from complex high-dimensional data. This capability facilitates the task of fitting arbitrarily intricate *nonlinear* statistical models to data with significant precision. Examples of financial applications of NN can be found in [2], [14], and [15]. In a WNN, a complex signal is first subjected to a wavelet-based decomposition process, so that unclear temporal structures can be exposed for further and easier evaluation. The enhanced decomposed signal is then used as an input element to a NN to discern and capture valuable information during the knowledge discovery process of the training phase. For articles covering the application of WNN, we cite [7], [9], and [16].

Section I of this article presents a brief introduction. Section II covers all aspects of the experiment including a description of the procedure for the generation of the results. Section III presents the corresponding analysis of results, conclusions and recommendations for future research.

II. EXPERIMENT

The experiment aimed to demonstrate that short-term predictions for future stock returns can be estimated with reasonable accuracy based on a dynamic recurrent neural network that uses as input current and lagged detail and smooth coefficients resulting from a non-decimated Haar wavelet-based decomposition of the high and low stock prices. The predictive power of the one, three, and five step-ahead forecasts of the proposed neuro-wavelet net was compared against the performance of five other models by using a set of predefined criteria.

A. Source Data

The raw data consisted of the trade prices for Apple ordinary stock (ticker: AAPL). The data was directly obtained from the TAQ3 database of the NYSE and included the period of September 1st through November 7th of the year 2008. Only trade transactions that occurred during normal hours of trading operation were included (i.e. between 9:30AM and 4:00PM EST). A total of 49 trading days encompassing 14.8 million records made up the original high-frequency time series.

The raw data was first filtered to verify *reliability*, *consistency*, and *liquidity*. Price levels and stamped dates

Manuscript received May 22, 2012; revised June 7, 2012.
L. F. Ortega is a Financial Engineering Ph.D. candidate with the School of Systems & Enterprises, Stevens Institute of Technology, Hoboken, NJ 7030 USA (e-mail: lortega@stevens.edu).

and times were tested to ensure that their values were within the expected logical ranges. A stock was defined as liquid if it registered an average of at least one trade transaction of 200 stock units every 5 minutes. This equated to a daily volume requirement of at least 15,600 stock units and a minimum average of 12 trade transactions per trading hour. The average daily volume and average number of transactions per minute for AAPL were 50,510,619 and 776 respectively. The filtered raw data was then tested for the existence of *outliers.* Only additive outliers[1] were evaluated [4]. Trade-price levels in the filtered raw data did not exhibit any unexplainable spikes.

Three time series were constructed from the depurated source data: The high and the low price s were computed for the 390 minutes conforming each of the 49 trading days. The resulting *high* and *low price series* were 19,110 long. The *log-returns series* was computed after converting the depurated trade-price series whose inherent nature was the irregularity of the time intervals between consecutive trade transactions to a time sequence of equally-spaced intervals of one minute. The resulting one-minute interval log-return series was also 19,110 long.

B. Pre-processing

Only the high and low price series were subjected to a scale-based decomposition process using a Haar Maximum Overlap Discrete Wavelet Transform (MODWT)[2] with a critical resolution level J = 8. The intent was to utilize the resulting detail and smooth coefficients as an exogenous signal to be input to the Jordan and Elman neuro-wavelet nets during training.

The Haar wavelet is a good differencing filter and as such is appropriate for capturing fluctuations in adjacent observations similar to the changes occurring on the stock returns [8]. The MODWT was the selected wavelet method for its ability to cope with the *circular shift effect*[3] [11]. In this experiment, we were interested on forecasts that fell under the domain of intra-day activity. A critical resolution level of J = 8 is associated with a time interval, given in minutes, that can capture intra-day price changes, [128, 256] [5]. Also note that a total of 510 boundary coefficients were removed from the high price, low price, and log-return time series to account for the *circularity effect*[4] affecting those wavelet coefficients associated with locations at the beginning and end of the high and low price series [11].

The result of the multi-resolution decomposition of the high and low price series was a collection of 18 sets of coefficients corresponding to eight levels of detail $d_j(t)$ and a single level of smooth $s_J(t)$ coefficients for each of the two series [10]. The 18 sets of coefficients were grouped as a sequence of feature vectors and at each time step t, the high and low price level were given by:

$$Price(t) = \sum_{j=1}^{J} d_j(t) + s_J(t) \tag{3}$$

The resulting time series for the high prices, low prices, and log-returns were all arrays of dimension 1x18,600. The vector sequence containing the exogenous signal conformed of the detail and smooth coefficients resulted in an array of dimension 18x18,600. All time series as well as the vector sequence were split into two parts. The first 16,000 elements of the series and vector sequence made up the in-sample data which was used for modeling while the last 2,600 elements constituted the out-of-sample data used for the evaluation of the one, three, and five step-ahead forecasting.

C. Processing

The experiment consisted of two phases: modeling and forecasting.

1. Modeling Phase

The in-sample data was used to fit two econometric models and to train four neural network topologies. The models under consideration were: an ARIMA(p,d,q), an ARIMAX(p,d,q)[5], a Jordan and Elman NN that used the high and low stock price series as an exogenous signal presented to the NN as an input source, and a Jordan and Elman WNN that used the wavelet coefficients contained in a feature vector sequence fed as an input to the WNN.

The in-sample portion of the log-returns produced an ARIMA(3,1,1) and ARIMAX(3,1,1) as the best fitted econometric models. The resulting closed formulas were, following the same order:

$$(r_t - r_{t-1}) - \phi_1(r_{t-1} - r_{t-2}) = e_t + \theta_1 e_{t-1} + \theta_2 e_{t-2} \tag{5}$$

$$(r_t - r_{t-1}) - \phi_1(r_{t-1} - r_{t-2}) = e_t + \theta_1 e_{t-1} + \theta_2 e_{t-2} + \beta_1(H_t - H_{t-1}) + \beta_2(L_t - L_{t-1}) \tag{7}$$

Where, the innovation series $\{e_t\}$ is a white noise, and $\{r_t\}$, $\{H_t\}$, and $\{L_t\}$ correspond to the one-minute time series of the log-returns, high and low prices respectively. The selection of the best ARIMA and ARIMAX models was based on the lowest AIC criteria.

[1] Additive outliers are those data points that show unusual characteristic only at one particular point in time. This means that any adverse effect caused by the outlier is not carried over through time.

[2] The MODWT is a modified version of the DWT that sacrifices orthogonality for the ability to remain unaffected by the circular shift effect. As a consequence, a time series decomposed using the MODWT is always aligned with respect to time.

[3] A DWT is orthogonal and depends critically on the starting point of the signal being analyzed. A shift in the starting point will yield to different results that are reflected on dissimilar sets of wavelet coefficient values.

[4] The MODWT is an operation that assumes circularity which could be a questionable assumption in some instances leading to unreliable values for those wavelet coefficients affected by the circularity assumption.

[5] The ARIMAX is an ARIMA model which allows for external regressors to account for the effect of outer signals on the model.

ISBN: 978-988-19251-6-9
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

Similarly, the Jordan and Elman NN as well as the Jordan and Elman WNN made use of the in-sample portion of the log-return series as the target for training the networks on the forecasting of the $t + 1$, $t + 3$, and $t + 5$ log-returns.

The architectures associated with the Jordan and Elman networks were as follows: Each of the two network topologies was conformed of one input source, one hidden layer, and one output layer. The input accommodated the corresponding exogenous signal accordingly; this is the high and low prices for the case of the Jordan and Elman NN, and the detail and smooth coefficients for the case of the Jordan and Elman WNN. The synapse between the input and the hidden layer had attached a tapped delay line to account for seven time delays. The single hidden layer was comprised of 20 neurons using a tan-sigmoid function as the activation function for each of the neurons. The tan-sigmoid function was chosen because of its recognized ability to address pattern recognition and forecasting problems. The number of neurons in the hidden layer was selected based on two considerations: the number of total inputs feeding the hidden layer and the performance error of the network. The output layer was made up of one neuron which used a linear function as the activation function. The single output layer aggregated the results of the hidden layer and generated a time series corresponding to the forecasted one, three, and five step-ahead log-return values.

In a Jordan architecture, the output of the output layer is connected back to the network as an additional input element. In an Elman architecture, the feedback is from the output of a hidden layer to the same layer (see Fig. 1).

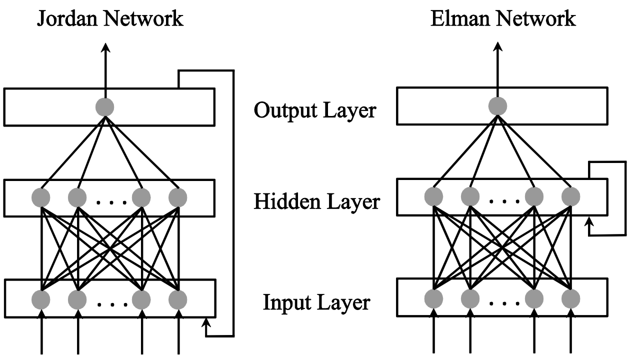

Fig 1. Dynamic Recurrent Neural Network Architectures

In both cases and for the purpose of this experiment, the feedback connections had attached a tapped delay line to account for two time delays corresponding to lags 1 and 2.

The Elman networks were trained using the Scaled Conjugate Gradient algorithm and the Jordan networks using the Levenberg-Marquardt algorithm. The first algorithm is a gradient-descent-based method while the second algorithm is a Jacobian-based method. The difference in training algorithm is justified by the additional computational demand that the topology of an Elman network imposed.

2. Forecasting Phase

All resulting models were tested for accuracy of the one, three, and five step-ahead forecasts using the corresponding out-of-sample data

The out-of-sample portion of the data was conformed of the last 2,600 elements of the one-minute interval log-returns, high and low price series, and the wavelet feature vector sequence. Log-returns estimates for the one, three, and five step-ahead forecasts were produced for the two econometric models, the Jordan and Elman neural networks, and the Jordan and Elman neuro-wavelet nets.

III. RESULTS

A comparative analysis of the predictive power of all six models was carried out based on a set of performance parameters defined as follows: The mean square error (MSE), self-explanatory, is the average of the squared errors resulting from the difference between the estimated and actual log-return values. The directional accuracy (DA), expressed as a percentile, is just the success ratio of the sign predictions. Thus, the DA is indicative of the number of times that the forecasted values have matched the direction specified by the sign followed by the actual log-returns. The directional change accuracy (DCA), expressed also as a percentile, measures the correctness of the predicted variation on log-return direction. The missed confidence interval (MCI), expressed as a percentile, is indicative of the number of times that a target log-return value failed to fall within the 95% confidence interval defined by standard errors produced by the corresponding predicted log-returns. For an expanded explanation of the performance criteria use in this experiment see [6].

A. Comparative Forecasting Analysis

The comparative forecasting analysis was performed using the results from the five, three, and one step-ahead forecasts of the ARIMA (3,1,1), ARIMAX (3,1,1), and the Elman and Jordan topologies with and without the use of wavelet coefficients as an input source to the networks. All performance measures were computed using the results generated by the corresponding out-of-sample data. Tables I, II, and III present the results for the one, three, and five step-ahead forecasts for all models under evaluation.

We can observe that the mse, DA, and DCA values for the different scenarios pointed in the direction of a consistent superior prediction performance of the Jordan network that used wavelet coefficients as an input source. Only in the scenario of the one step-ahead forecast, the Jordan neuro-wavelet was not able to achieve superiority in the DCA measure. However, the resulting 58.80713% was not extremely far from the highest DCA value of 65.01669%. All models achieved approximately the same performance for the MCI measure; this value was around 5.0%.

ISBN: 978-988-19251-6-9
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

TABLE I
1 STEP-AHEAD FORCASTING

	mse	*DA*	*DCA*	*MCI*
Arima(3,1,1)	0.03919277	69.44979 %	65.01669 %	5.00000 %
Arimax(3,1,1)	0.03933833	68.60331 %	63.7931 %	5.038462 %
Elman neural network	0.03950338	58.23029 %	41.51786 %	5.059869 %
Elman neuro-wavelet net	0.02799383	67.9289 %	40.34598 %	4.943994 %
Jordan neural network	0.03999068	53.80162 %	37.79264 %	5.285494 %
Jordan neuro-wavelet net	0.01780613	75.49209 %	58.80713 %	5.169753 %

TABLE II
3 STEP-AHEAD FORCASTING

	mse	*DA*	*DCA*	*MCI*
Arima(3,1,1)	0.0395666	50.71236 %	36.67223 %	5.04234 %
Arimax(3,1,1)	0.0397037	49.17212 %	35.72621 %	5.119323 %
Elman neural network	0.03959135	51.97216 %	19.03964 %	5.2184 %
Elman neuro-wavelet net	0.03123017	58.46868 %	22.38973 %	5.06378 %
Jordan neural network	0.04007157	50.28969 %	28.57143 %	5.135135 %
Jordan neuro-wavelet net	0.02119616	70.76091 %	48.3817 %	5.135135 %

TABLE III
5 STEP-AHEAD FORCASTING

	mse	*DA*	*DCA*	*MCI*
Arima(3,1,1)	0.03953439	50.63584 %	37.02673 %	5.046225 %
Arimax(3,1,1)	0.03966897	49.90366 %	32.85078 %	5.046225 %
Elman neural network	0.03952317	51.78019 %	21.62011 %	4.990329 %
Elman neuro-wavelet net	0.03472186	54.06347 %	21.56425 %	4.951644 %
Jordan neural network	0.04020202	50.9857 %	37.40927 %	5.448223 %
Jordan neuro-wavelet net	0.02000625	72.01392 %	50.08375 %	5.061824 %

The measures presented in tables I, II, and III also made obvious the superiority of those neural networks that used wavelet coefficients as an input source over those that did not. This superiority surfaced during the modeling phase as illustrated in Fig. 2 and Fig. 3. Fig. 2 presents the modeling results for the Elman and Jordan neural networks which utilized the high and low price series as an input source. Under this setting, better fittings of the in-sample data were achieved in all cases by the Jordan net. The actual data was represented in black color and the fittings in gray. Similarly, Fig. 3 presents the modeling results for the Elman and Jordan networks that utilized the coefficients of a wavelet-based decomposition of the high and low price series as an input source. Under this setting, again, the Jordan net achieved better fittings of the in-sample data in all cases.

A quick look at both figures also exposed a better modeling performance of the neuro-wavelet networks over the plain neural networks as their modeling behavior is depicted by larger gray areas in the corresponding plots. Larger gray areas imply better fittings of the in-sample data given that the distances between the actual and estimated log-return values decreased.

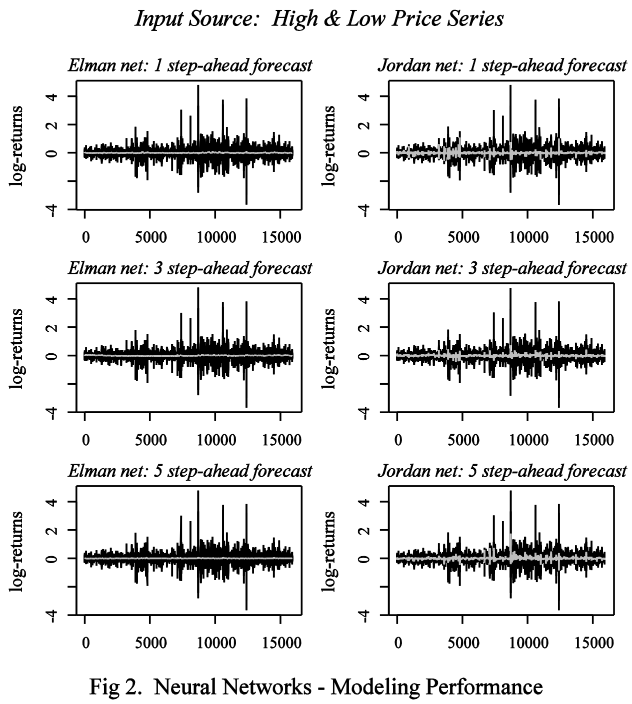

Fig 2. Neural Networks - Modeling Performance

ISBN: 978-988-19251-6-9
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

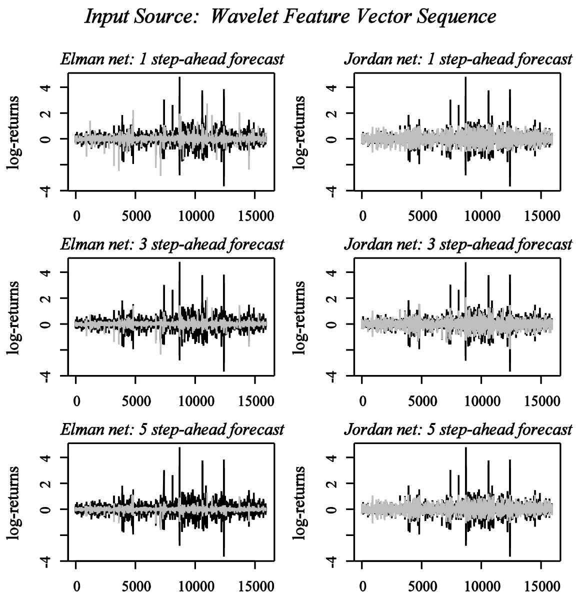

Fig 3. Neuro-Wavelet Networks - Modeling Performance

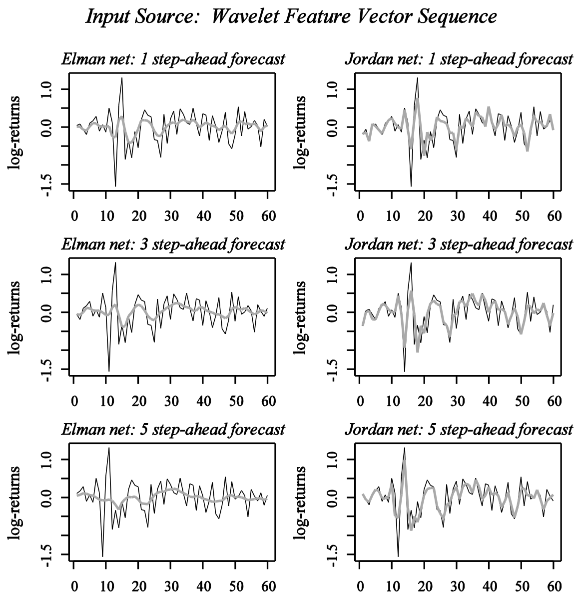

Fig 5. Neuro-Wavelet Networks - Forecasting Performance

The next four figures corresponding to the first 60 estimates of the one, three, and five step-ahead log-return forecasts are presented to assist with a comparative visualization of the predictive performance between those neural networks that used wavelet coefficients as an input source and those that did not.

Fig. 4 shows the scenario where the Elman and Jordan nets used the high and low price series as an input source to the network while Fig. 5 illustrates the scenario where the input source was a feature vector sequence containing coefficients resulting from a wavelet-based decomposition of the high and low price series. The first 60 elements of the out-of-sample data are portrayed in black color and their corresponding forecasted values in gray.

Better forecasting results were generated by the Jordan and Elman WNN. In particular, the Jordan neuro-wavelet networks achieved the best forecasting performance overall. The superior forecasting performance of the Jordan neuro-wavelet nets is better noted in Fig. 5 which illustrates how the paths depicted by the forecasts (gray lines) produced by the Jordan neuro-wavelet nets mirrored with relative accuracy the log-returns changes in direction observed in the paths portrayed by the actual values (black lines); this superiority was reflected in higher DA and DCA values.

Likewise, Fig. 6 and Fig. 7 illustrate the number of times that a target log-return value (black points) failed to fall within the 95% confidence interval defined by the corresponding forecast standard errors (gray band).

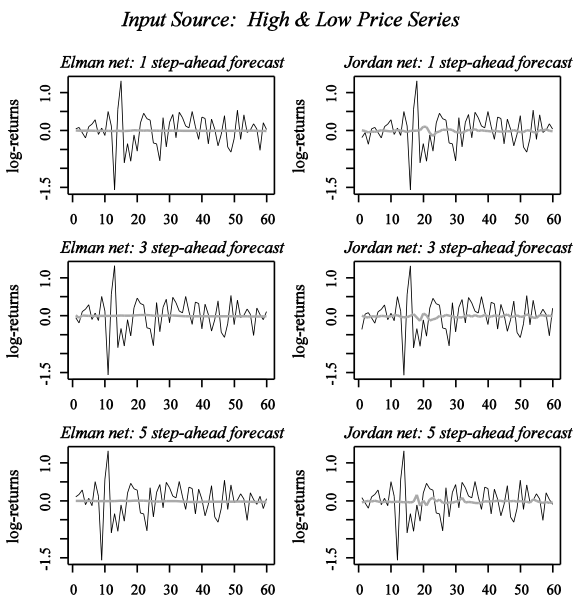

Fig 4. Neural Networks - Forecasting Performance

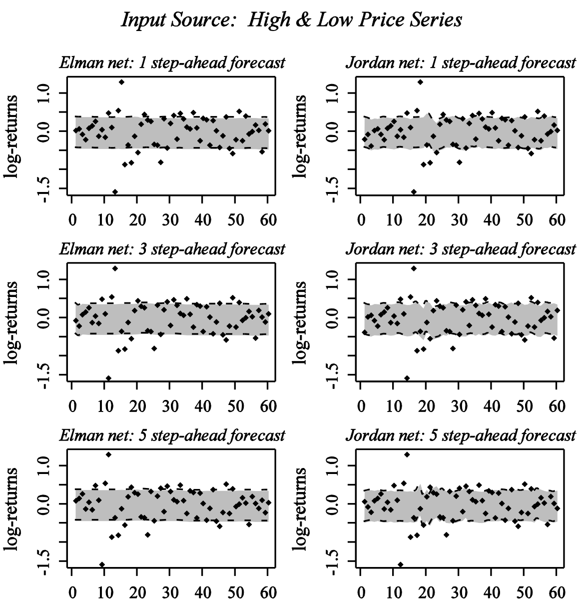

Fig 6. Neural Networks - Forecasting Performance

ISBN: 978-988-19251-6-9
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

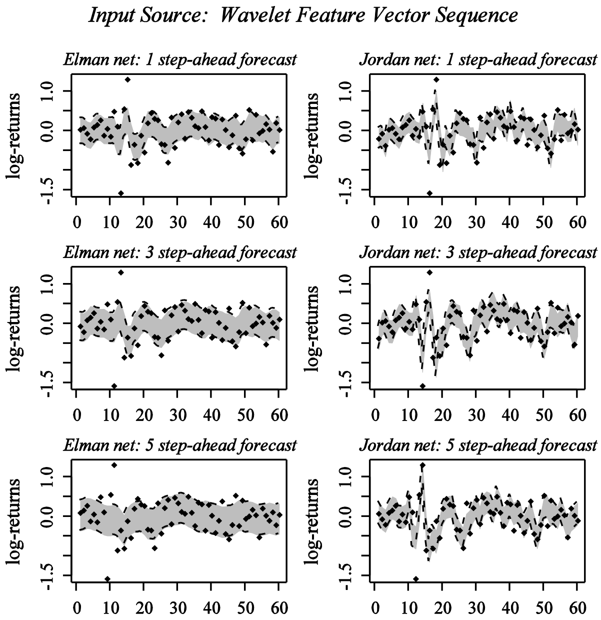

Fig 7. Neuro-Wavelet Networks - Forecasting Performance

The plots in Fig 6 and Fig. 7 corroborated the MCI values presented in tables I, II, and, III (i.e., approximately 5.0% for all cases). The errors generated by all forecasting models were regularly distributed and the distributions were stationary. These attributes facilitated the definition of the 95% confidence interval boundaries needed to generate the MCI measure.

IV. CONCLUSIONS

A hybrid model based on a Jordan network topology that used the coefficients resulting from a non-decimated Haar wavelet-based multi-resolution decomposition as an external input source to the network showed a consistent superior performance for the modeling and forecasting of high-frequency stock returns.

An external signal decomposed via a Haar MODWT as opposed to a non-decomposed version of the same signal appears to offer richer information that can be used for modeling and forecasting. The better quality of the decomposed data can be attributed to the wavelet ability to capture structural nonstationary characteristics at different time scales. The relevant contribution of the wavelet coefficients was authenticated when two neuro-wavelet network topologies were compared against equivalent neural network topologies that did not use wavelet coefficients.

The proposed Jordan neuro-wavelet net achieved reasonable accurate results, even though it did not need to have a predefined specific parametric model to initiate a simulation process. In particular, the proposed neuro-wavelet net showed superior modeling and forecasting performance when compared against two parametric methods.

The proposed wavelet-based procedure proposed in this article can be applied to any exogenous signal to evaluate its structural relation with the target signal that is used during the training phase of a neural network. Ideally, strong cross-correlation relations between the exogenous and target signals are desired at all scales.

A possibility for future research could be the study of the predictive power, short and long-term, of each level of wavelet resolution as a function of its contribution to the time series generated by the forecasts. In finance, this is equivalent to modeling a financial instrument behavior on a scaled-based basis and testing each scale independently. Another possibility for research could be the utilization of a different exogenous signal or a combination of several exogenous signals as a function of the impact exerted over the model performance results.

REFERENCES

[1] V. Bjorn, “Multiresolution Methods for Financial Time Series Prediction,” in *Proc. of the IEEE/IAFE 1995 Conf. on Computational Intelligence for Financial Engineering*.
[2] T. Chenoweth, and Z. Obradovic, “A Multicomponent Nonlinear Prediction System for the S&P 500 Index,” *Neurocomputing*, vol. 10, no. 3: pp. 275–290, 1996.
[3] I. M. Dremin, and A. V. Leonidov, “Volatility Dynamics of Wavelet-Filtered Stock Prices,” *Bulletin of the Lebedev Physics Institute*, vol. 35, no. 1: pp. 1–5, 2008.
[4] P. H. Franses, and N. Haldrup, “The Effects of Additive Outliers on Test for Unit Root and Cointegration,” *J. of Business & Economics*, vol. 12, no. 4: p. 471, 1994.
[5] R. Gencay, F. Selcuk, and B. Whitcher, *An introduction to Wavelets and Other Filtering Methods in Finance and Economics*. San Diego, CA: Academic Press, 2002.
[6] P. D. McNelis, *Neural Networks in Finance*. Burlington, MA: Elsevier Academic Press, 2005.
[7] K. K. Minu, M. C. Lineesh, and C. Jessy John, “Wavelet Neural Networks for Nonlinear Time Series Analysis,” *Applied Mathematical Sciences*, vol. 4, no. 50: pp. 2485-2595, 2010.
[8] M. Misiti, Y. Misiti, G. Oppenheim, and J. M. Poggi, *Wavelets and their Application*. London, UK: Wiley-ISTE Publishing, 2007.
[9] F. Murtagh, J. L. Starck, and O. Renaud, “On Neuro-wavelet Modeling,” *Decision Support Systems*, vol. 37: pp. 475-484, 2004.
[10] R. T. Odgen, *Essential Wavelets for Statistical Applications and Analysis*. Boston, MA: Birkhauser, 1997.
[11] D. B. Percival, and A. T. Walden, *Wavelet Methods for Time Series Analysis.* Cambridge University Press, 2006.
[12] O. Renaud, J. L. Stark, and F. Murtagh, “Prediction based on a Multiscale Decomposition,” *Int. J. of Wavelets, Multiresolution and Information Processing*, vol. 1, no. 2: pp. 217-232, 2003.
[13] S. Soltani, D. Boichu, P. Simard, and S. Canu, “The long-term memory prediction by multiscale decomposition,” *Signal Processing*, vol. 80, no. 10: pp. 2195– 2205, 2000.
[14] M. Zekic, “Neural Network Applications in Stock Market Predictions - A Methodology Analysis,” in *Proc. of the 9th Int. Conf. on Information and Intelligent Systems ‘98*, 1998, pp. 255-263.
[15] G. Zhang, B. E. Patuwo, and M. Y. Hu, “Forecasting with Artificial Neural Networks: The State of the Art,” *International Journal of Forecasting*, vol. 14: pp. 35-62, 1998.
[16] Q. Zhang, and A. Benveniste, “Wavelet Networks,” *IEEE Transactions on Neural Networks*, vol. 3, no. 6: pp. 889-898, 1992.
[17] G. Zheng, J. L. Starck, J. G. Campbell, and F. Murtagh, “Multiscale transforms for filtering financial data streams,” *Journal of Computational Intelligence in Finance*, vol. 7: pp. 18– 35, 1999.

ISBN: 978-988-19251-6-9
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)