한국지능정보시스템학회논문지 제 5 권 제 1 호 (1999.6)

A Hybrid System of Joint Time-Frequency Filtering Methods and Neural Network Techniques for Foreign Exchange Rate Forecasting

Taeksoo Shin*, Ingoo Han*

환율예측을 위한 신호처리분석 및 인공신경망기법의 통합시스템 구축

신택수, 한인구

Abstract

Input filtering as one of preprocessing methods is so much crucial to get good performance in time-series forecasting. There are a few preprocessing methods (i.e. ARMA outputs as time domain filters, and Fourier transform or wavelet transform as time-frequency domain filters) for handling time-series. Specially, the time-frequency domain filters describe the fractal structure of financial markets better than the time domain filters due to theoretically additional frequency information. Therefore, we, first of all, try to describe and analyze specially some issues on the effectiveness of different filtering methods from viewpoint of the performance of a neural network based forecasting. And then we discuss about neural network model architecture issues, for example, what type of neural network learning architecture is selected for our time series forecasting, and what input size should be applied to a model. In this study an input selection problem is limited to a size selection of the lagged input variables. To solve this problem, we simulate on analyzing and comparing a few neural networks having different model architecture and also use an embedding dimension measure as chaotic time-series analysis or nonlinear dynamic analysis to reduce the dimensionality (i.e. the size of time delayed input variables) of the models. Throughout our study, experiments for integration methods of joint time-frequency analysis and neural network techniques are applied to a case study of daily Korean won / U.S. dollar exchange returns and finally we suggest an integration framework for future research from our experimental results.

Key words: R/S, Embedding Dimension, Signal Decomposition, Fourier Transform, Wavelet Transform, Neural Networks

---

* Graduate School of Management, Korea Advanced Institute of Science and Technology

1. Introduction

Making predictions and building trading models are central goals for financial institutions as an investor or financial manager. However, the difficulty in forecasting chaotic time series such as economic or financial data makes them fall into a dilemma. It is sometimes attributed to the limitation of many conventional forecasting models, while it would reversely give many researchers a opportunity to develop more predictable forecasting models.

For example, the models using artificial intelligence such as neural network techniques have been recognized as more useful forecasting models than the conventional statistical forecasting models (Hill *et al*., 1994; Tang *et al*., 1991; Tang and Fishwick, 1993).

Recently, more intelligent forecasting models have been developed through integration methods between neural network techniques and other learning algorithms. This study introduces joint time-frequency analysis and focuses on the integration of signal processing algorithms (such as Fourier or wavelet analysis) and neural network techniques to gain more meaningful time series features for the efficient and effective learning. For example, a discrete wavelet transform (DWT) allows us to compute wavelet coefficients from coarse to fine scale levels efficiently and so reduces the training time of the neural network (Tsui *et al.*, 1995).

To date, economic and financial data are often analyzed in either time domain or frequency domain. If the data is stationary, then these are useful approaches. However, economic and financial data are usually non-stationary or non-homogeneous in some sense. In these cases, it is instructive to look at the data either on the time-frequency dimension or on multiple scales over time. Time- frequency or time-scale methods allow us to observe the changes in behavior over time.

As noted by Ville (1948) there are two basic approaches to time-frequency analysis. The first approach is to initially cut the signal into slices in time, and then to analyze each of these slices separately to examine their frequency content. The other approach is to first filter different frequency bands, and then cut these bands into slices in time and analyze their energy content.

The first of these approaches is used for the construction of the short time Fourier transform and the Wigner-Ville transform, while the second leads to the wavelet transform. The wavelet transform is a mechanism used to dissect or breakdown a signal into its constituent parts, thus enabling analysis of data in different frequency domains with each component resolution matched to its scale.

Alternatively this may be seen as a decom- position of the signal into its set of basis functions (wavelets), analogous to the use of sines and cosines in Fourier analysis to represent other functions. These basis functions are obtained from dilations or contractions (scaling), and translations of the mother wavelet. The important difference that distinguishes the wavelet transform from Fourier analysis is its time and frequency localization properties.

When analyzing signals of a non-stationary nature, it is often beneficial to be able to acquire a correlation between the time and frequency domains of a signal. In contrast to the Fourier transform, the wavelet transform allows exceptional localization in both the time domain via translations of the mother wavelet, and the scale (frequency) domain via dilations.

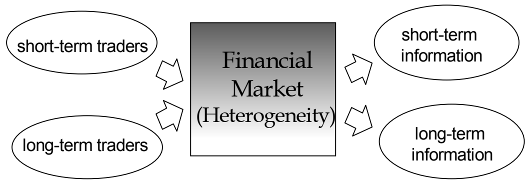
Specially, the wavelet analysis is a robust tool that may be used to obtain qualitative information for highly nonstationary time series, i.e. to detect a small-amplitude harmonic forcing term even when is chaotic and even for short total times (Permann and Hamilton, 1992).

The purpose of this study is to introduce new filtering methodologies based on Fourier and wavelet transforms that use signal decomposition algorithms and so can forecast with greater accuracy than others models. First of all, the original series of daily Korean won / U.S. dollar returns are applied to several models including random walk, mean reverting, ARMA, and artificial neural networks as a benchmark model. And then the Fourier filtered and wavelet filtered time series decomposed by time scale are applied to our neural network models.

This paper is organized as follows. The next section reviews the financial market heterogeneity. The third and the fourth section describe time series decomposition methods and filtering methods. The fifth section presents the integration methods of joint time-frequency analysis and neural networks learning and then their experimental results and the conclusion contains final comments.

2. Financial Market Heterogeneity

Market heterogeneity suggests that the different intentions among market participants result in sensitivity by the market to several different time-scales (Refer to Fig. 1). Müller *et al.* (1993, 1995) present that the heterogeneous market hypothesis has been associated with fractal phenomena in the empirical behavior of foreign exchange rate (FX) markets. Based on the hypothesis, a scaling law relating time horizon and size of price movements (volatility) has been identified in Müller *et al.* (1990).



[Figure 1] Heterogeneous Financial Market Structure

For example, short-term traders are constantly watching the market; they re-evaluate the situation and execute transactions at a high frequency. Long-term traders may look at the market less frequently. A quick price increase of 0.5% followed by a quick decrease of the same size, for example, is a major event for an FX intraday trader but a non-event for central banks and long-term investors.

Sometimes, small short-term price moves may have a certain influence on the timing of long-term traders transactions but never on their investment decisions as such. Long-term traders are interested only in large price movements and these normally happen only over long time intervals.

Therefore, long-term traders with open positions have no need to watch the market every minute. In other words, they judge the market, its prices and also its volatility with a coarse time grid. It means that a coarse time grid reflects the view of long-term trader and a fine time grid reflects that of a short-term trader.

In summary, these different types of traders create the multi-scale dynamics of the time series. Specially the multi-scale nature of wavelet analysis makes it useful for detection and characterization of selfsimilar or scaling behavior, such as is common for chaotic processes.

3. Time Series Decomposition Methods

Conventionally, time series have been thought to consist of a mixture of trend (T_t), seasonal (S_t), cycle (C_t), and irregular components (e_t). We can write the time series Z_t as

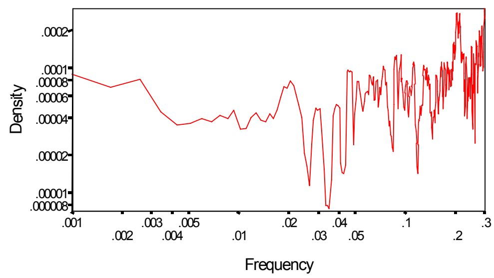
$$Z_t = f(T_t, S_t, C_t, e_t) \qquad (1)$$

But, when the coverage of forecasting is much short-term, time series mainly consist of cyclic and irregular components. A traditional way of decomposing a time series into cycles of different frequencies has been through spectral analysis. The spectrum of a series gives an alternative way of looking at the series in the frequency domain instead of the time domain. The spectrum of a series can be interpreted as the decomposition of the variance of the series. A peak in the spectrum at specific frequency ω indicates that a cycle of frequency ω is present in the series and that it gives an important contribution to the variance of the series. If at a frequency ω, the spectrum of the series is almost zero, it means the existence of no cycle of frequency ω with substantial contribution to the variance of the series. Taking in a series the cycles of high frequency and those of low frequency as detected by the spectrum it is possible to decompose the series into components of high and low frequency fluctuations.

Secondly, in time series analysis, the focus has also been on regular, periodic cycles. In Fourier analysis, we assume that irregularly shaped time series are the sum of a number of periodic sine waves, each with differing frequencies and amplitudes. Spectral analysis attempts to break an observed irregular time series, with no obvious cycle, into these sine waves (Refer to Fig. 4). Peaks in the power spectrum are considered evidence of cyclical behavior. Spectral analysis imposes an unobserved periodic structure on the observed nonperiodic time series (Peters, 1994).

There is no intuitive reason for believing that the underlying basis of market or economic cycles has everything to do with sine waves or any other periodic cycle. Spectral analysis would be an inappropriate tool for market cycle analysis.

On the contrary, in chaos theory, nonperiodic cycles exist. These cycles have an average duration, but the exact duration of a future cycle is unknown. In that situation, we need a more robust tool for cycle analysis, a tool that can detect both periodic and nonperiodic cycles. That is, rescaled range (R/S) analysis can perform that function (Refer to Section 6). R/S analysis is useful for uncovering periodic cycles, even when the cycles are superim- posed on one another. R/S analysis can discern cycles within cycles.



[Figure 2] Fourier power spectrum on a log-log scale of Korean Won/U.S. Dollar (1992. 1. - 1996. 6.)

Fig. 2 shows an example of the cyclic features from the viewpoint of R/S analysis. That is, the power spectra of Korean won / U.S. dollar series are flat and represent a broadband power spectrum. It shows that the initially straight downward trend of the power spectrum on a log-log plot is characteristic for a fractal (i.e. possibly chaotic) signal.

4. Signal Decomposition Filtering Methods

A filter in our study is defined as a means of separating the various periodic components of a time series into individual series.

The following sections describe several filtering methods.

4.1. Time Domain Filters-ARMA Filter

Time domain filters mean linear filters using statistical parameters of such statistical models as ARMA in our study.

4.1.1. Moving average (MA) models

$$Y_t = \sum_{j=0}^{q} b_j e_{t-j} = b_0 e_t + b_1 e_{t-1} + \ldots + b_q e_{t-q} \qquad (2)$$

Equation (2) describes a convolution filter: the new series y_t is generated by an qth-order filter with coefficients b_0, ..., b_q from the series e_t, ..., e_{t-q}. It is called an qth-order moving average model, MA(q).

The moving averages technique performs quite well when the market is in a state of trend but performs rather poorly around turning points and/or oscillations, because it receives delayed signals of the abrupt changes (Refenes, 1995).

4.1.2. Autoregressive (AR) models

MA filters operate in an open loop without feedback; they can only transform an input that is applied to them. If we do not want to drive the series externally, we need to provide some feedback (or memory) in order to generate internal dynamics:

$$Y_t = \sum_{i=1}^{p} a_i Y_{x_{t-i}} + e_t \qquad (3)$$

Equation (3) is called a pth-order autoregressive model, AR(p) or an infinite impulse response (IIR) filter (because the output can continue after the input ceases). Depending on the application, et can represent either a controlled input to the system or noise.

4.1.3. ARMA models (Weigend, 1994)

The next step in complexity is to allow both AR and MA parts in the model. This is called an ARMA(p,q) model.

$$Y_t = \sum_{i=1}^{p} a_i Y_{t-i} + \sum_{j=0}^{q} b_j e_{t-j} \qquad (4)$$

Equation (4) has the mixed features of AR and MA model shown from Equation (2) and (3).

4.2. Joint Time-Frequency Domain Filters

4.2.1. Fourier Transform

Knowledge of the frequency components provides a means of estimating where in each cycle the present time series reached, with important consequences for predicting future behavior. Thus a Fourier spectrum indicates the frequencies, and their strengths, inherent in a time series. However this assumes the signal, or signals, to be stationary. For a general nonstationary time series the Fourier transform provides no information on the time localization of spectral components, rather the frequency spectrum would reveal wide band features characteristic of noise. A single abrupt change in the time series would, for example, affect all the components of the frequency spectrum. A transform designed for stationary signals cannot resolve features of a nonstationary signal.

In the beginning, the communications industry has developed many different types of filters. These filters not only are used in electronic communications, but also have an application base that includes radar and sonar imaging, electronic warfare, medical technology and so on.

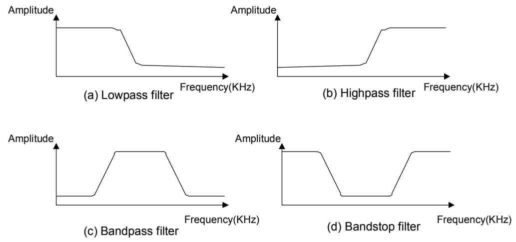
$$y_t = \sum_{k=0}^{N/2} [a_k \cos(\omega_k t) + b_k \sin(\omega_k t)] + \varepsilon_t, \quad t = 1,2,\ldots,N. \quad (5)$$

$$s.t. \begin{pmatrix} a_k = R_k \cos(\theta_k) \\ b_k = -R_k \sin(\theta_k) \end{pmatrix}$$

ω_j : sinusoidal frequency
R_j : amplitude of the variation
θ_j : phase
ε : stationary random series

Equation (5) is called the Fourier series of the sequence y_t. The a_k and b_k are called Fourier coefficients.

By the Fourier coefficients, all the application-specific filter implementations can be grouped into four general filter types: lowpass, highpass, bandpass, and bandtop filter. The characteristic frequency response of these filters is depicted in Fig. 3. Especially, the adaptive filter can reproduce the characteristics of any of the four basic filter types, alone or in combination (Freeman and Skapura, 1991).



[Figure 3] Frequency-Response Characteristics of the Four Basic Filter Types

One of popular Fourier transforms is a Fast Fourier transform (FFT). It works by recursively splitting the series in half, transforming each half separately in a quarter of the time that would be taken to transform the entire series, then quickly merging the results. FFT has three requirements about the time series data as follows (Hartle, 1994).

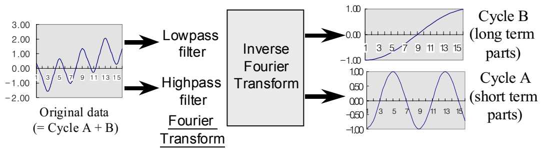
First, the data must have the major trend removed, as a cycle length longer than the data being analyzed will skew the values of the power spectrum. One method to remove the trend is to measure the trend using the linear least-squares method and subtract the trend values from the original data.

Second, the beginning and end points of the data must have approximately the same values for FFT.

An FFT assumes that the cycles continue to repeat into the future. If you apply an FFT to data with different beginning and ending points, the output of the FFT will be incorrect. To adjust the data so it has the same beginning and end points, the data must be processed with a Hanning window. This step will eliminate endpoint discontinuities.

Finally, the last requirement for FFTs: the data period must be a power of 2 - that is, 64, 128, 256 or greater periods. Thus, we must augment the data by adding zero to the data so the final array has a value equal to a power of 2.

Fig. 4 shows that the FFT decomposes summed cycles such as combined cycle A and B into separate cycles according to the power and frequency of each cycle.



[Figure 4] An Example of Using Fourier Transforms

4.2.2. Wavelet Transform

A new decomposition method, wavelet analysis is significantly different from Fourier analysis as follows. While Fourier analysis gives us only frequency information, wavelet analysis gives us both frequency information and time information and so can represent a nonstationary process better than Fourier analysis by allowing us to look at the series through wavelets of variable sizes. While the wavelet theory has brought about significant advancements in representation of functions, not much work on its applicability to forecasting has been made. Although an initial attempt to provide the statistical framework for wavelet analysis was made by Basseville *et al.* (1992a, 1992b), the applicability of wavelet analysis to forecasting needs to be further developed. Tak (1995) introduced new methodologies based on wavelet decomposition that can forecast with greater accuracy than existing models and utilized the simplest wavelets and a univariate case.

Besides, there has been a rapidly-growing literature on the use of wavelets for denoising and smoothing (Donoho, 1992; Donoho and Johnstone, 1994, 1995a, 1995b; Donoho *et al.*, 1993, 1994). Wavelets are well suited to the above problem because of time (or space) localization. When approximating a signal, wavelets can preserve local features (discontinuities, turning points, etc.) while still removing noise.

There are many different types of DWTs which have been explored since the original work in the 1980s. That is, unlike sines and cosines, which define a unique Fourier transform, there is not a unique set of wavelets; in fact, there are infinitely many possible sets (i.e. scaling functions). Roughly, the different sets of wavelets make different trade-offs between how compactly they are localized in space and how smooth they are.

The DWT shown in Fig. 5, as developed by Daubechies (1988), is similar to FFT. Both take an input vector whose length is normally a power of two and output a different vector of the same length. The entire process is also reversible, which means that the transform data can be used to reconstruct the original input at any point in the procedure. But, the wavelet transform yields a decomposition which is neither continuous nor unique. So unlike the FFT, the DWT does not have the limited time-frequency resolution of the FFT and thus provides more accurate representation of the input (Rioul and Vetterli, 1991).

The DWT consists of applying a wavelet coefficient matrix hierarchically, first to the full data vector of length N, then to the smooth vector of length N/2, then to the smooth-smooth vector of length N/4, and so on until only a trivial number of smooth-,...-smooth components (usually 2) remain. The procedure is sometimes called a pyramidal algorithm. The output of the DWT consists of these remaining components and all the detail components that were accumulated along the way.

Suppose the finest scale provides the original data, $x_N = x$, and the approximation at scale m is xm where usually $m = 2_0, 2_1, \ldots, 2_N$. The incremental detail added in going from x_m to x_{m+1}, the detail signal, is yielded by the wavelet transform. If em is this detail signal, then the following holds:

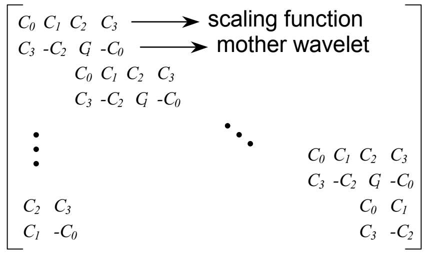
$$x_{m+1} = H^T(m)x_m + G^T(m)e_m \qquad (6)$$

where H(m) (i.e. defined as scaling function or father wavelet in Fig.5-(c)) and G(m) (i.e. defined as mother wavelet in Fig.5-(d)) are matrices (linear transformations) depending on the wavelet chosen, and T denotes transpose (adjoint). An intermediate approximation of the original signal is immediately possible by setting detail components $e_{m'}$ to zero for $m' \geq m$ (thus, for example, to obtain x_2, we use only x_0, e_0, and e_1). Alternatively we can de-noise the detail signals before reconstituting x and this has been termed wavelet regression (Bruce and Gao, 1994; Murtagh, 1996).

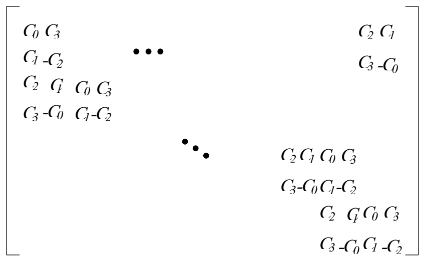
Define e as the row-wise juxtaposition of all detail components, $\{e_m\}$, and the final smoothed signal, x_0, and consider the wavelet transform W given by

$$Wx = e = \left[e_{N-1} \cdots e_0 x_0\right]^T \qquad (7)$$

Taking $W^TW = I$ (the identity matrix) is a strong condition for exact reconstruction of the input data, and is satisfied by the orthogonal wavelet transform.

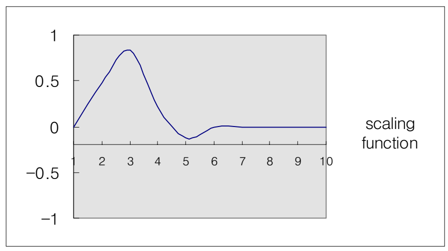


(a) Wavelet Transform Matrix

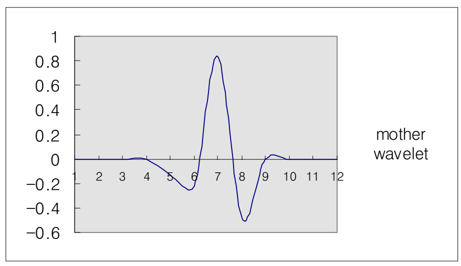


(b) Inverse Wavelet Transform Matrix

$$c_0 = \frac{1+\sqrt{3}}{4\sqrt{2}} \quad c_1 = \frac{3+\sqrt{3}}{4\sqrt{2}} \quad c_2 = \frac{3-\sqrt{3}}{4\sqrt{2}} \quad c_3 = \frac{1-\sqrt{3}}{4\sqrt{2}}$$



(c) Scaling Function (or Father Wavelet)



(d) Mother Wavelet

[Figure 5] Daubechies Wavelet with Order 4

Throughout our study, we use the Daubechies wavelet transform with order 4 (DAUB4) as examples of these orthogonal wavelets (Fig. 5).

5. Integration of Filtering Methods and Neural Networks

In this section, we present three types of neural networks for univariate time series forecasting as follows.

(1) NN technique using the real values
(2) NN technique using the time domain filters (AR, MA, and ARMA Models)
(3) NN technique using the time-frequency domain filters (Fourier transform, and wavelet transform)

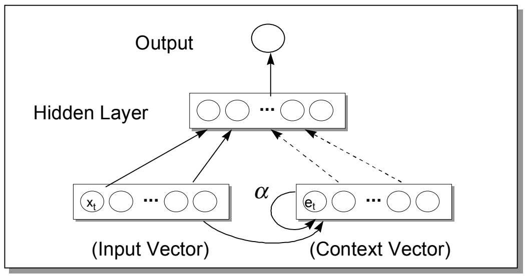
In this study, we use a fast Fourier transform (FFT) as a Fourier transform method and Daubechies wavelet as a wavelet transform method of the log price change data.

In this section, we suggest two integration methods of filtering methods and neural network techniques.

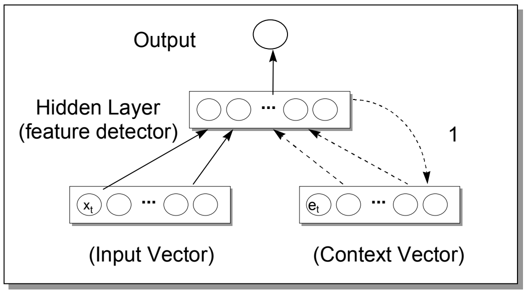
One integration method is to forecast one period-ahead return using single neural network model. The other integration method forecasts each scale (i.e. lowpass filtered data and highpass filtered data) using multiple neural network models, followed by combining of the individual forecasts.

5.1. Recurrent Neural Networks (RNN)

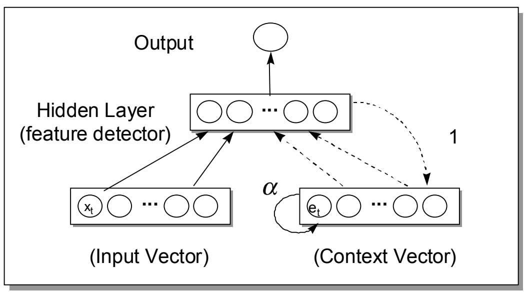
Several researchers confirm the superiority of RNNs over feedforward networks when performing nonlinear time series prediction (Connor and Atlas, 1991; Logar *et al.*, 1993). Especially, RNNs can yield good results because of the rough repetition of similar patterns present in time series. These regular but subtle sequences can provide beneficial forecast ability. However, a large network size is often needed and its training generally requires an excessively long history of input. In general, context units of RNNs accumulate a weighted moving average or trace of the past output values.



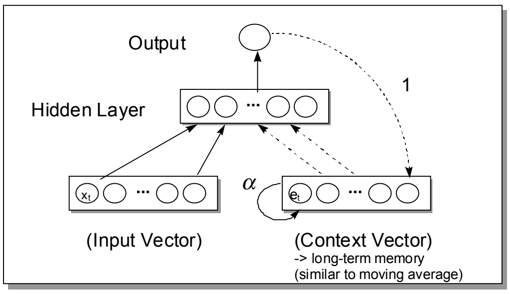
(a) RNN(1) : Stornetta *et al.* Model[1988]



(b-1) RNN(2) : Elman Model[1988,90]



(b-2) RNN(2) : Modified Elman Model



(c) RNN(3) : Jordan Model[1986,89]

[Figure 7] Recurrent Neural Network Architectures

Stornetta *et al.* Model (1988) in Fig. 7- (a) RNN(1) can perform sequence recognition tasks. The only feedback is now from the context units to themselves, give them decay properties, but their input is now the network input itself, which only reaches the rest of the network via the context units.

Elman (1990) suggested the architecture shown in Fig. 7- (b-1) RNN(2). The input layer is divided into two parts: the true input units and the context units. The context units simply hold a copy of the activations of the hidden units from the previous time step. Modified Elman model in Fig. 7- (b-2) modified RNN(2) has additional parts, i.e. the feedback from the context units to themselves.

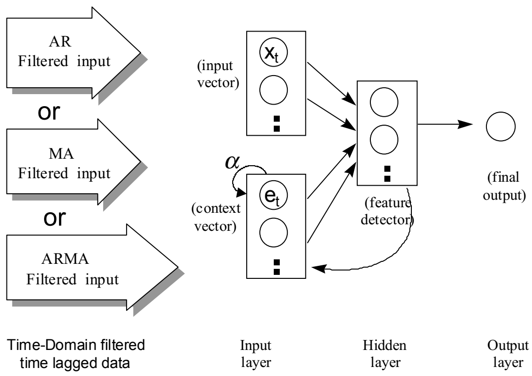
Fig.7- (c) RNN(3) shows the Jordan (1986, 1989) architecture. It has the context units fed from the output layer and also from themselves.

Generally RNN(1) is less well suited than RNN(2) and RNN(3) to generating or reproducing sequences.

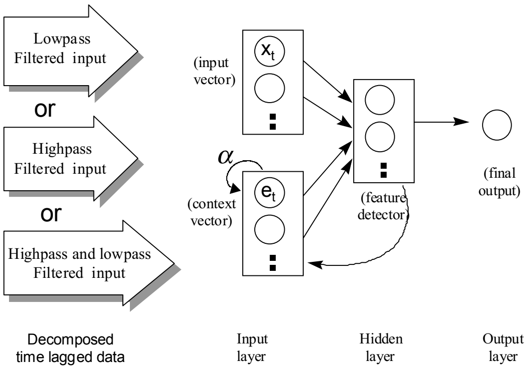
Based on the above models, we suggest the following integration models. First of all, we use a general integration model of the time domain filtering methods and neural networks as benchmark integration model (Fig. 8A).

Next, we describe a new integration model of joint time-frequency filtering methods and neural networks. It consists of two types of integration approach, i.e. (a) single recurrent neural network model (RNN) combined with the filtering methods and (b) multiple recurrent neural network model (MRNN) combined with the filtering methods (Fig. 8B).

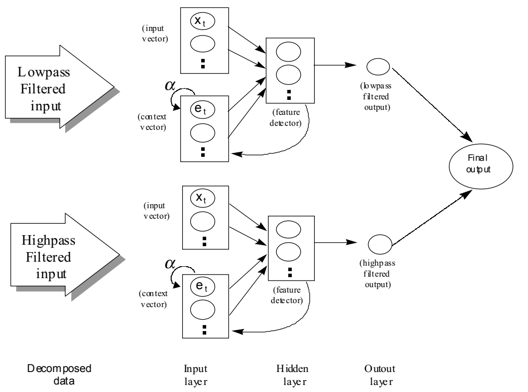
In our integration models (Fig. 8B), we use each filter type such as highpass type (short-term part) and lowpass type (long-term part) as joint time-frequency domain filters applied to single RNN models. And also we design the second model of two integration methods on a condition that time series consist of highpass filtered parts and lowpass filtred parts (i.e. time series = highpass part + lowpass part). That is, two sub-integration models separately forecast one day ahead each lowpass output and highpass output and then two forecast filtered outputs are summed up for forecasting final output.



[Figure 8A] Integration of Time Domain Filtering Methods and Neural Network Learning Algorithms



(a) Single Recurrent Neural Network (RNN) Model Combined with Frequency Domain Filtering Method



(b) Multiple Recurrent Neural Network (MRNN) Model Combined with Time-Frequency Domain Filtering Method (Output = lowpass filtered data + highpass filtered data)

[Figure 8B] Recurrent Neural Network Model Combined with Time-Frequency Filtering Analysis

6. Empirical Analysis and Results: A Case Study of Korean Won / U.S. Dollar Exchange Rate Market

The daily Korean won / U.S. dollar exchange rates are transformed to the returns and then standardized from January 10, 1992 to June 25, 1996. The returns are defined as the logarithm of today's exchange rate divided by the logarithm of yesterday's exchange rate. The learning phase involves observations from January 10, 1992 to August 4, 1995, while the testing phase runs from August 7, 1995 to June 25, 1996.

In our study, the returns are basically decomposed into an approximation part and a detail part of the daily series by the FFT and the DAUB4 filter.

Theoretically, the joint time-frequency filters decomposes the input signal into detail signals, and a residual or the time series into varying scales of temporal resolution. On the contrary, the original signal can be expressed as an additive combination of the filter coefficients, at the different resolution levels.

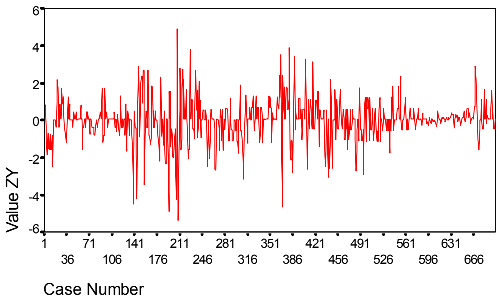
We show the effectiveness (performance) of different forecasting models by two comparative analyses of different filtering methods in the following sections.

6.1. Highpass and Lowpass Filters

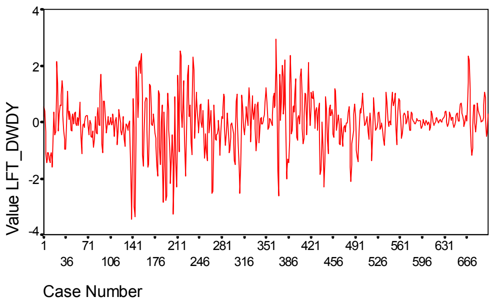
In this paper, we select the input size based on chaos analyses from the experimental results. That is, The number of input nodes used in the neural networks is chosen by the nonlinear dynamic analysis (estimating embedding dimension) and the analysis was proved useful from a set of values we tested, because it gave nearly the smallest in-sample prediction error.

Therefore, throughout our study, the neural network structure we used basically has 4 inputs (Refer to Section 6.2.2.). The inputs have several types, i.e. lowpass and highpass filtered data by FFT and DWT (DAUB4) including original time series data as shown in Fig. 6A, 6B, and 6C.

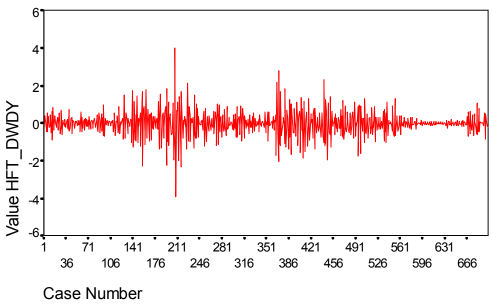
Lowpass filters pass all frequencies below the specified frequency, and they are usually employed for smoothing. Highpass filters pass all frequencies above the specified frequency. They are usually used to extract information on local variation while suppressing overall signal levels.



[Figure 6A] Original Time Series - Daily Korean Won / U.S. Dollar Returns ($\ln X_t - \ln X_{t-1}$)

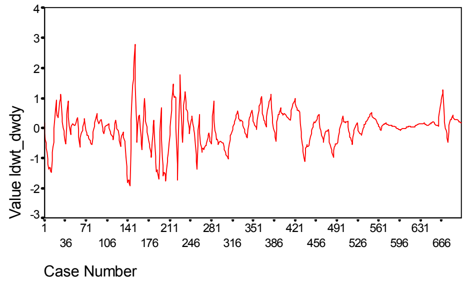


(a) Lowpass Filter

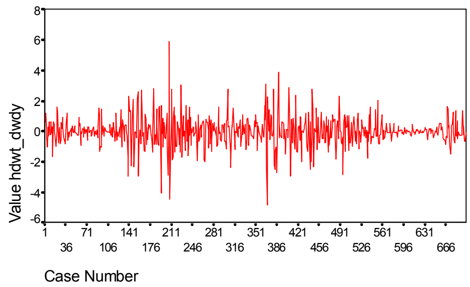


b) Highpass Filter

[Figure 6B] Fourier Transformed Time Series



(a) Lowpass Filter



(b) Highpass Filter

[Figure 6C] Wavelet Transformed Time Series

6.2. Nonlinear Dynamic Analysis

6.2.1. Rescaled Range Analysis

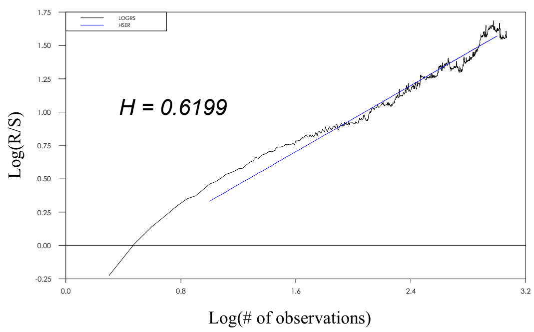
Hurst exponent (Hurst *et al.*, 1965) is a measure of predictability of time series that has interesting characteristics. The exponent is derived using so called R/S analysis. Given a time series X containing a number of points, n, and choosing an integer divisor p where for convenience: 10 <= p < n/2, the data can be divided into n/p blocks. For each block the average value is calculated, then the maximum range of each block and the standard deviation of each block. The value (range)/(standard deviation) is calculated for each block and then averaged over the blocks.

The average value R/S is related to the Hurst exponent by the following formula:

$$R / S = \left(\frac{p}{2} \right)^{H} \qquad (8)$$

where H is the Hurst exponent. In order to gain a more reliable estimate the value of R/S is calculated for all the possible values of p, and the resulting tuples are logged and a linear regression is performed on them.

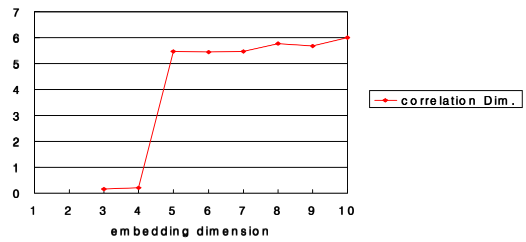
Fig. 9 shows the results of applying rescaled range analysis to Korean won/ U.S. dollar returns. The Hurst coefficient for the FX is 0.6199 greater than 0.5. It is calculated as a global least-mean-squares estimate for the slope of the curve of the logarithm of the correlation integral in the linear region as a function of n (the number of observations) during the given period (1992.1. - 1996. 6.). It means that the coefficient indicates a bias or memory effect that is biased toward reinforcement of the trend, which is called persistence. It also helps analyzing a predictability of our time series in advance.



[Figure 9] R/S (Rescaled Range) Analysis

6.2.2. Embedding dimension

Neural networks for univariate time series forecasting are a kind of nonlinear AR model and so the choice of order for the models is based on the embedding dimension of the series in our study because the chaos analysis is a good method to analyze nonlinear dynamics in the series.



[Figure 10] Correlation Dimension vs. Embedding Dimension

That is, neural networks provide a reliable basis for nonlinear and dynamic market modeling. Nonlinear dynamics and chaos theory can also provide an information about important input sizes (i.e. time lags) for the design of forecasting systems using neural networks (Embrechts *et al.,* 1994).

The results of the chaos analysis in Fig. 10 indicate a saturating tendency for the correlation dimension, leading to a fractal dimension of about 6. The embedding dimension (i.e. the dimension of the phase space for which saturation in the correlation dimension occurs) is 5.

The embedding dimension of 5 indicates that 4 time-lag data must be shown to a neural network to predict the 5th data point of the time series. But, while the short-term predictions are in principle possible for a chaotic time series, they might prove difficult in practice because of the high value of the correlation dimension.

6.3. Selecting the Order of the ARMA Model

In this section, we show the model building process of conventional ARMA models and ARMA filters. There are several heuristics to find the right order such as the Akaike Information Criterion (AIC), Schwartz's Bayesian Criterion (SBC) and so on. First, based on the standard tools, we select ARMA(0,1) or MA(1) Model as the optimal ARMA model as shown in Table 1.

<Table 1> Estimates of ARMA Parameter Identification

ARMA(p,q)	*Standard error*[a]	*Log likelihood*	*AIC*[b]	*SBC*[c]
ARMA(1,0)	1 0	10	0 0	0 0
ARMA(,0)	1 0	10 1 0	0 1	10
ARMA(0,1)	1.0767	10 1.0	**2086.07**	**20 .16**
ARMA(0,2)	1.077	10 0. 1	20 7.	2101. 7
ARMA(1,1)	1.077	10 0. 2	20 7.	2101.
ARMA(2,1)	1.07 0	10 0.	20 .7	2107. 7
ARMA(1,2)	1.077	10 0. 7	20 .1	2107.
ARMA(2,2)	1.07	10 .	20 6.7	210 .

a. *Residual Standard Deviation*,
b. *AIC = -2ln(maximum likelihood) +2k,*
c. *SBC = -2ln(maximum likelihood) + kln(n),*
(*k= the number of total parameters, n = sample size*).

Second, we also evaluate ARMA models as optimal time domain filters in terms of their neural network performance. To this purpose, we use alternative time domain filters such as AR(p) and ARMA(p,q) model ($p,q \neq 0$). Based on AIC and SBC statistics, we select AR(1), MA(1), and ARMA(1,1) as a alternative time domain filter. Table 3 shows AR(1) filter has a little better performance than any other time domain filter in test samples.

6.4. Determining Thresholding Criteria of Time-Frequency Filter

Whether Fourier transformation is lowpass, highpass, or bandpass is characterized by two parameters, i.e. frequency and width. For lowpass filters, the frequency is the cutoff below which periodic components are passed and above which periodic components are obstructed. The width is the transition range over which the response of the filter goes from one extreme (unimpeded passage) to the other extreme (total cutoff).

It is difficult to specify the width parameter. No simple calculation provides the correct width parameter. It is arbitrary choice. Unfortunately, a real tradeoff is involved.

A specific DWT, based on the Daubechies wavelet, produces a set of wavelet coefficients (or decomposed series) from coarse to fine scale levels.

The simplest wavelet shrinkage technique is so-called hard thresholding. Wavelet coefficients are replaced by 0 if they are smaller in absolute value than a fixed threshold as a filtering criteria. The threshold is a tuning parameter of wavelet shrinkage. Donoho and Johnstone (1994, 1995a) propose several thresholds (i.e. universal, SURE), as well as several thresholding policies. Nason (1994) suggests the well-known cross-validation method to find a proper threshold, which minimizes the mean integrated error for use with wavelets. A few other references in threshold selection and wavelet shrinkage applications are Gao (1993) and Vidakovic (1994, 1995).

The ability of the network to capture dynamic behavior over higher resolution levels deteriorates quite fast. The higher the order of the resolution scale, the smoother the curve, and thus the less information the network can retrieve.

These issues of the optimal filters deviate from our main topics in this study (Refer to Shin and Han (1999a, 1999b)). In this study, we select 25% and 50% as a threshold value (i.e. band levels of threshold filtered coefficients divided by total bands of threshold coefficients) of Daubechies wavelet transform and FFT based on root mean square errors (RMSE) of our models changing the threshold value from 10 % to 90%.

6.5. Comparative Analysis

We make the following comparative analyses in this section. They consist of (1) the comparative effects of non-filters and filters on the model performances, (2) the comparative effects of different filter types (ARMA filters as a time domain filter and, Fourier filters and wavelet filters as a time- frequency domain filter) on the model performances.

<Table 2> Performance Comparison of Non-Filtered Forecasting Models as Benchmark Models Using Root Mean Square Error (RMSE) Measure

Model	*(I-C-H-O)[d]*	*Train Set*	*Test Set*
BPN	(4-0-4-1)	0.835638	0.877505
RNN(1)[a]	(4-4-4-1)	0.854616	0.894718
RNN(2)[b]	(4-4-4-1)	0.835809	**0.871192**
RNN(3)[c]	(4-1-4-1)	0.846807	0.885184
ARMA(0,1)	-	0.854402	0.882932
Random Walks	-	1.070294	1.062600
Mean Reverting	-	0.946250	0.945900

a. Stornetta *et al.* Model (1988),
b. Modified Elman Model (1988, 1990),
c. Jordan Model (1986, 1989),
d. Model Structure (I: Input Units, C: Context Units, H: Hidden Units, O: Output Units).

Table 2 summarizes the RMSE (root mean square error) on the two distinct subsets of the time series in each model.

First of all, we use random walks, mean reverting, ARMA(p,q), back-propagation neural networks (BPN), and RNNs without filtering as a benchmarking model and compare each other in the predictability of the models. That result shows RNN(2), i.e. a modified Elman model had a little better performance than any other model in test samples. So we use this model as our main model to use filtered data in the following experiments.

<Table 3> The Integration Performance of ARMA Filters and Modified Elman RNN Models (Unit: RMSE)

Filter Type	*(I-C-H-O)[a]*	*Train Set*	*Test Set*
AR(1)	(4-4-4-1)	0.886288	**0.947860**
MA(1)	(4-4-4-1)	0.886420	0.948424
ARMA(1,1)	(4-4-4-1)	0.886114	0.949303

a. Model Structure (I: Input Units, C: Context Units, H: Hidden Units, O: Output Units).

In Table 3, by trial and error methods we found the optimal pth or qth order of AR(p), MA(q), or ARMA(p,q) filter combined to a modified Elman model. Based on its experimental results, we evaluated the final model performances shown in Table 1. That is, a AR(1) or ARMA(1,0) filter among ARMA(p,q) filters had the best model performance. But the result of the AR(1) didn't show a significant difference as compared with that of the other ARMA(p,q) filters.

Table 4 compares the performances of different joint time-frequency filters integrated with RNNs and shows the following results. First, as compared with Table 3, the RNNs combined with joint time-frequency filters outperformed both RNNs without filtering and RNNs with time domain filters. Namely all the integration of joint time- frequency filtering methods and neural networks were significantly superior to the other models such as the statistical models (ARMA, random walk, or mean reverting) and the integration of time domain filters and neural network models. Second, The performance of each time-frequency filter type presented different validation results in integration with the RNNs. Before explaining the results, we assumed that each filter type, i.e. highpass type (short-term part) or lowpass type (long-term part) of joint time-frequency domain filters has its own information on the RNN forecasting models.

<Table 4> The Integration Performance of Joint Time-Frequency Filters and Modified Elman Models (Unit: RMSE)

Model	*(I-C-H-O)*[i]	*Train Set*	*Test Set*
LFT_RNN(2)[a]	(4-4-4-1)	0.889336	0.939239
HFT_RNN(2)[b]	(4-4-4-1)	0.889659	0.948427
LWT_RNN(2)[c]	(4-4-4-1)	0.842508	0.939097
HWT_RNN(2)[d]	(4-4-4-1)	0.889990	0.948492
FT_RNN(2)[e]	(8-4-4-1)	0.263899	0.26952
FT_MRNN(2)[f]	(4-4-4-1)[j]	0.251940	0.253377
WT_RNN(2)[g]	(8-4-4-1)	0.735958	0.80823
WT_MRNN(2)[h]	(4-4-4-1)[j]	0.718174	0.792379

a. Lowpass Fourier Filter + Single Modified RNN(2),
b. Highpass Fourier Filter + Single Modified RNN(2),
c. Lowpass Wavelet Filter + Single Modified RNN(2),
d. Highpass Wavelet Filter + Single Modified RNN(2),
e. Both Fourier Filters + Single Modified RNN(2),
f. Both Fourier Filters + Multiple Modified RNN(2),
g. Both Wavelet Filters + Single Modified RNN(2),
h. Both Wavelet Filters + Multiple Modified RNN(2),
i. Model Structure (I: Input Units, C: Context Units, H: Hidden Units, O: Output Units),
j. Subordinate Model Structure of MRNN(2).

Our experimental results show that the RNN considering both highpass and lowpass filer types significantly outperformed the RNNs considering partially highpass or lowpass types. Finally in predictability between FT_MRNN and WT_MRNN, FT_MRNN outperformed WT_MRNN.

Table 5 and Fig. 11 show the following summarized results of Table 2, 3, and 4. They consist of five types of models, i.e. ARMA(0,1) and a modified RNN(2) without any filters, a modified RNN(2) with a time domain filter, a multiple modified RNN(2) model with FFT filters, and a multiple modified RNN(2) model with DAUB4 filters. That is, first, the modified RNN(2) models basically outperformed the statistical ARMA model and second, the modified RNN(2) with time-frequency domain filters out- performed the modified RNN(2) models without any filters. Third, the time-frequency domain filtered modified RNN(2) models outperformed the time domain filtered modified RNN(2) models. Finally, the modified RNN(2) model with the FFT filters outperformed the modified RNN(2) model with the DAUB4 filters.

Based on the above results, we propose a new integration model framework of joint time-frequency analysis and neural networks for the future study (Fig. 12). That is, our proposed model, first of all, considers chaos analysis for the dimensionality of NN models. After the chaos analysis, our frame- work shows that according to filtering criteria, two types of control parameters, i.e. 1) thresholding criteria for time-frequency filters and 2) learning parameters for the neural networks are simultaneously optimized by one evaluation function (e.g. RMSE) within our model.

7. Conclusions

Our study utilized several filtering methods in the context of NN forecasting using daily Korean won / U.S. dollar series. That is, from our experimental study, we obtained the following results. First, we found that using the approximation part (or lowpass filtered data) and the detail part (or highpass filtered data) in forecasting got better results than applying the original series directly to the models. Second, in integration with neural network models we detected that the joint time-frequency filtering (Fourier or wavelet transform) outperformed the time domain filtering. In summary, the neural network models combined with joint time-frequency analysis consistently outperformed the other benchmark models.

Third, even though wavelet transform methods theoretically have an advantage over Fourier transform methods, the performances of the Fourier transform method got much better results than that of the wavelet transform method. Some of the reasons are related to the following.

First, a wavelet transform has a feature detecting sharply defined structures (such as abrupt change, singularities, discontinuities, edges and bumps) better than a Fourier transform. Its performance is different according to unique time series distribution or its features.

Second, we used a Daubechies wavelet as a popular wavelet transform method. However, there are a lot of wavelet transform methods in wavelet analysis. That is, we didn't fully consider that each wavelet method has its own features in our study.

Third, the wavelet filter frequencies and widths in this study were chosen arbitrarily.

Therefore considering the above issues, in future research, we will analyze the differences of various wavelet transform types and then try to detect the optimal features (e.g. filtering criteria) using artificial intelligent techniques for financial time series forecasting.

<Table 5> Comparative Analysis of Different Forecasting Models

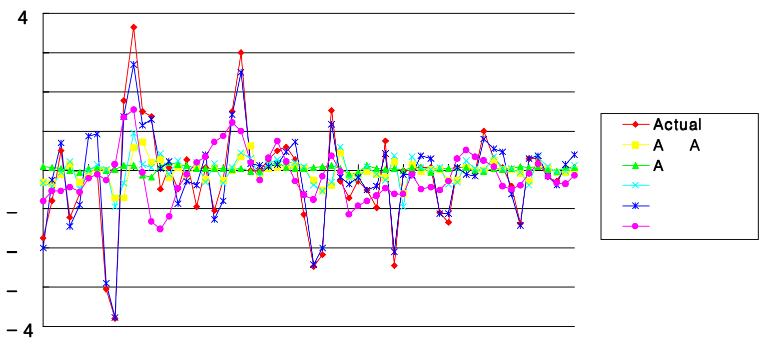
(a) The Average Performances (RMSE) of Different Forecasting Models

Model	*Filter type*	*Train Set*	*Test Set*
ARMA(0,1)	-	0.854402	0.882932
Modified RNN(2)	-	0.835809	0.871192
Modified RNN(2)	AR(1)	0.886288	0.947859
Multiple Modified RNN(2)	Fast Fourier Transform	0.251940	0.253377
Multiple Modified RNN(2)	Daubechies Wavelet Transform (DAUB4)	0.718174	0.792379

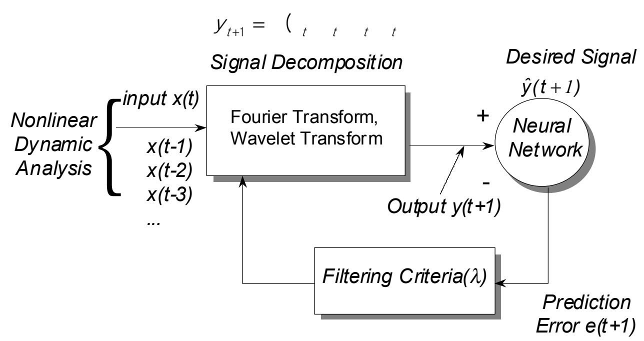
(b) Paired Samples t -Test for the Differences in RMSE Using Test Samples

Model	*RNN(2)*	*AR(1)_RNN(2)*	*FT_MRNN(2)*	*WT_MRNN(2)*
ARMA(0,1)	1.21 (.228)	-2.89 (.004)***	6.98 (.000)***	2.03 (.043)**
RNN(2)		2.29 (.004)***	7.27 (.000)***	1.83 (.069)*
AR(1)_RNN(2)[a]			6.78 (.000)***	3.61 (.000)***
FT_MRNN(2)				-5.87 (.000)***

*. significant at 10% level , **. significant at 5% level, ***. significant at 1% level,
a. AR(1) Filter + Modified RNN(2).



[Figure 11] Actual and Predicted Value Distributions of Differently Analyzed Forecasting Models According to Test Time Horizon



[Figure 12] A Proposed Integration Model Framework of Joint Time-Frequency Analysis and Neural Networks

References

[1] Alsberg, B.K., A.M. Woodward, and D.B. Kell, "An Introduction to Wavelet Transforms for Chemometricians: A Time-Frequency Approach," *Chemometrics and Intelligent Laboratory Systems* 37, (1997), 215~239.

[2] Ambardar, A., *Analog and Digital Signal Processing*, PWS Publishing Company, (1995).

[3] Azoff, E., *Neural Network Time Series Forecasting of Financial Markets*, John Wiley & Sons, (1994), 33-35.

[4] Basseville, M., A. Benveniste, and A.S. Willsky, "Multiscale Autoregressive Processes, Part I: Schur-Levinson Parameterizations," *IEEE Transactions on Signal Processing* 40, (1992a), 1915~1934.

[5] Basseville, M., A. Benveniste, and A.S. Willsky, "Multiscale Autoregressive Processes, Part II: Lattice Structures for Whitening and Modeling," *IEEE Transactions on Signal Processing* 40, (1992b), 1935~1953.

[6] Baxter, M., and R.G. King, "Measuring Business Cycles: Approximate Band-pass Filters for Economic Time Series," *Working paper*, University of Virginia, (1995).

[7] Brock, W. and W. Dechart, "Theorems on Distinguishing Deterministic from Random Systems," in *Dynamic Econometric Modeling,* Barnett, W., Berndt, R.J., and White, P.J. (eds.), Cambridge: Cambridge University Press, (1991).

[8] Bruce, A. and H.Y. Gao, *S+Wavelets User's Manual, Version 1.0*, Seattle, WA: StatiSci Division, MathSoft Inc. (1994).

[9] Daubechies, I., "Orthogonal Bases of Compactly Supported Wavelets," *Communications on Pure and Applied Mathematics*, Vol. 41, (1988), 909~996.

[10] Deboeck, G.J. (ed.), *Trading on the Edge - Neural, Genetic, and Fuzzy Systems for Chaotic Financial Markets*, John Wiley& Sons, Inc., (1994).

[11] Donoho, D.L., "De-noising by Soft-thresholding," *Technical Report 409*, Dept. of Statistics, Stanford University, CA, (1992).

[12] Donoho, D.L. and I.M. Johnstone, "Ideal Spatial Adaptation by Wavelet Shrinkage," *Biometrika*, 81, (1994), 425~455.

[13] Donoho, D.L. and I.M. Johnstone, "Adapting to Unknown Smoothing via Wavelet Shrinkage," *Journal of American Statist. Association*, 90, (1995a), 1200~1224.

[14] Donoho, D.L. and I.M. Johnstone, "Minimax Estimation via Wavelet Shrinkage," *Submitted to the Annals of Statistics,* (1995b).

[15] Donoho, D.L., I.M. Johnstone, G. Kerkyacharian, and D. Picard, "Density Estimation by Wavelet Thresholding," *Preprint*, Dept. Statistics, Stanford University, (1993).

[16] Donoho, D.L., I.M. Johnstone, G. Kerkyacharian, and D. Picard, "Wavelet Shrinkage: Asymptopia?" *Journal of Royal Statistical Society, B.*, (1994).

[17] Elman, J.L. and D. Zipser, "Learning the Hidden Structure of Speech," *Journal of the Acoustical Society of America* 83, (1988), 1615~1626.

[18] Elman, J.L., "Finding Structure in Time," *Cognitive Science* 14, (1990), 179~211.

[19] Embrechts, M., M. Cader, and G.J. Deboeck, "Nonlinear Dimensions of Foreign Exchange, Stock, and Bond Markets," in *Trading on the Edge - Neural, Genetic, and Fuzzy Systems for Chaotic Financial Markets*, Deboeck, G.J. (ed.), John Wiley& Sons, Inc., (1994), 297~313.

[20] Episcopos, A. and J. Davies, "Predicting Returns on Canadian Exchange Rates with Artificial Neural Networks and EGARCH-M Models," *Neural Computing and Application*, (1995).

[21] Freeman, J.A. and D.M. Skapura, *Neural Networks Algorithms, Applications, and Programming Techniques*, Addison-Wesley Publishing Company, (1991), 49~53.

[22] Gao, H.Y., "Choice of Thresholds for Wavelet Estimation of the Log Spectrum," *Technical Report 438*, Dept. Statistics, Stanford University, (1993).

[23] Granger, C., *Spectral Analysis of Economic Time Series*, Princeton, NJ: Princeton University Press, 1964.

[24] Grauwe, D., H. Dewachter, and M. Embrechts,

Exchange Rate Theory - Chaotic Models of Foreign Exchange Markets, Blackwell Oxford UK & Cambridge USA, (1993), 94~95.

[25] Hartle, T., "Preprocessing Data and Fast Fourier Transform," *Technical Analysis of Stock and Commodities*, (1994).

[26] Hill, T., L. Marquez, M. O'Connor, and W. Remus, "Artificial Neural Networks for Forecasting and Decision Marking," *International Journal of Forecasting*, 10, (1994), 5~15.

[27] Hurst, H.E., R.P. Black, and Y.M. Simaika, *Long-Term Storage: An Experimental Study*, Constable - London U.K. (1965).

[28] Jordan, M.I., "Attractor Dynamics and Parallelism in a Connectionist Sequential Machine," In *Proceedings of the Eighth Annual Conference of the Cognitive Science Society* , Amherst, (1986), 531~546.

[29] Jordan, M.I., "Serial Order: A Parallel, Distributed Processing Approach," *In Advances in Connectionist Theory: Speech*, J.L. Elman and D.E. Rumelhart (eds.), (1989).

[30] Kim, S.H. and H. Noh, "A Comparative Study of ARIMA and Neural Network Models: Case Study in Korean Corporate Bond Yields," *Korean Management Science and Operations Research '96 Fall Conf.* [Korean title], Seoul, Sept. (1996), 19~22.

[31] Kuan, C.M. and T. Lim, "Forecasting Exchange Rates Using Feedforward and Recurrent Neural Networks," BEBR *Working Paper*, 92-0128, University of Illinois, Urbana-Champaign, (1992).

[32] Masters, T., *Neural, Novel & Hybrid Algorithms for Time Series Prediction*. John Wiley & Sons, Inc. New York, (1995).

[33] Müller, U. A., M.M. Dacorogna, R.D. Dave, O.V. Pictet, R.B. Olsen, and J.R. Ward, "Fractals and Intrinsic Time - A Challenge to Econometricans," *XXXIXth International Conference of the Applied Econometrics Association*, Real Time Ecnometrics – Submonthly Time Series, 14-15th Oct. (1993), in Luxembourg, and at *the 4th International PASE Workshop*, in Ascona (Switzerland), 22-26th Nov. (1993).

[34] Müller, U. A, M.M. Dacorogna, R.B. Olsen, O.V. Pictet, M. Schwarz, and C. Morgenegg, "Statistical Study of Foreign Exchange Rates, Empirical Evidence of A Price Change Scaling Law, and Intraday Analysis," *Journal of Banking and Finance*, 14, 1189~1208.

[35] Müller, U. A, M.M. Dacorogna, R.B. Olsen, O.V. Pictet, and J.E. von Weizsacker, "Volatilities of Different Time Resolutions - Analyzing the Dynamics of Market Components," *The First International Conference on High Frequency Data in Finance*, Zurich, March, (1995).

[36] Murtagh, F., "Wedding the Wavelet Transform and Multivariate Data Analysis," *Journal of Classification*, Submitted, (1996).

[37] Nason, G.P., "Wavelet Regression by Cross-Validation," *Technical Report 447*, Dept. Statistics, Stanford University, (1994).

[38] Peters, E., *Fractal Market Analysis: Applying Chaos Theory to Investment and Economics*, John Wiley&Sons, Inc. (1994), 87~88.

[39] Press, W., S.A. Teukolsky, W.T. Vetterling, and B.P. Flannery, *Numerical Recipes in C: The Art of Scientific Computing*, Cambridge Univ. Press, (1992).

[40] Refenes, A. (ed.), *Neural Networks in the Capital Markets*, John Wiley & Sons, (1995).

[41] Rioul, O. and M. Vetterli, "Wavelet and Signal Processing," *IEEE Signal Processing Magazine*, October (1991), 14~35.

[42] Shin, T. and I. Han, "Wavelet Thresholding Techniques to Support Multi-Scale Decomposition for Financial Forecasting Systems," *Korea Intelligent Information System Society '99 Spring Joint Conf.* [Korean title], Seoul, June (1999a), 175~186.

[43] Shin, T. and I. Han, "A Hybrid System of Wavelet Transformations and Neural Networks Using Genetic Algorithms: Applying to Chaotic Financial Markets," *Korea Intelligent Information System Society '99 Spring Joint Conf.* [Kor. title], Seoul, June (1999b), 271~280.

[44] Stornetta, W.S., T. Hogg, and B.A. Huberman, "A Dynamical Approach to Temporal Pattern Processing," In *Neural Information Processing*

Systems (Denver 1987), ed. D.Z. Anderson, New York: American Institute of Physics., (1988), 750~759.

[45] Strang, G., “Wavelets,” *American Scientist 82*, April, (1994), 250~255.

[46] Tak, B., "A New Method for Forecasting Stock Prices Using Artificial Neural Network and Wavelet Theory," *Ph.D. Dissertation*, Depart. Economics, University of Pennsylvania, (1995).

[47] Tang, Z., C. Almeida, and P.A. Fishwick, “Time Series Forecasting Using Neural Networks vs Box-Jenkins Methodology,” *Simulation* 57 (5), (1991), 303~310.

[48] Tang, Z. and P.A. Fishwick, “Feedforward Neural Nets as Models for Time Series Forecasting,” *ORSA Journal on Computing* 5 (4), (1993), 374~385.

[49] Torrence, C. and G. Compo, “ A Practical Guide to Wavelet Analysis,” *Bull Amer. Met. Soc.,* (1997).

[50] Tsui, F., M. Sun, C. Li, and R.J. Sclabassi, “Recurrent Neural Networks and Discrete Wavelet Transform for Time Series Modeling and Prediction,” *Proceedings of ICASSP-95*, Vol.5, (1995).

[51] Vidakovic, B., "Nonlinear Wavelet Shrinkage with Bayes Rules and Bayes Factors," *Discussion Paper* 94-24, ISDS, Duke University, (1994).

[52] Vidakovic, B., "Unbalancing Data with Wavelet Transformations," *Proceedings SPIE 1995, 2569, Wavelet Applications in Signal and Image Processing III*, (1995).

[53] Ville, W., "Theorie et Applications de la Notion de Signal Analytique", *Cables et Transmission*, Vol. 2a, (1948), 61~74. (Translated into English by I. Selin, RAND Corp. Report T-92, Santa Monica, CA, Aug, 1958.).

[54] Wei, W.W.S., *Time Series Analysis - Univariate and Multivariate Methods*, Addison-Wesley Publishing Company, (1990).

[55] Weigend, A., et al. (eds.), Time Series Prediction: Forecasting the Future and Understanding the Past, Addison-Wesley Publishing Company, (1994), 14~15.