Integration of Joint Time-Frequency Filtering Methods and Neural Network Techniques for Foreign Exchange Return Forecasting

Taeksoo Shin, Ingoo Han

Graduate School of Management, Korea Advanced Institute of Science and Technology,
207-43 Cheongryangri-Dong, Dongdaemoon-Gu, Seoul, 130-012, Korea.

Abstracts

Preprocessing is so much crucial to good performance in time-series forecasting. Preprocessing methods (i.e. time-domain filters, i.e. ARMA outputs, and frequency-domain filters, i.e., Fourier transform or wavelet transform) for handling the time-series that contain strong quasi-cyclical components are compared and contrasted within the neural network techniques. Specially, The integration approach of joint time-frequency analysis and neural network techniques is analyzed using daily Korean Won / U.S. Dollar exchange returns. And we, in addition, use the embedding dimension used in chaotic time-series analysis or nonlinear dynamic analysis for reducing the dimensionality (i.e. time lags) of the models.

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

Introduction

Making predictions and building trading models are central goals for financial institutions as a investor or financial manager. But, the difficulty in forecasting especially time series such as the economic or financial data is usually attributed to the limitation of many conventional forecasting models, but could encourage many researchers to develop the more predictable forecasting models. So the models using artificial intelligence such as neural network techniques have been recognized as more useful forecasting model than the conventional statistical forecasting models.

Recently, more intelligent forecasting models have been developed through integration method between neural network techniques and other learning algorithms. This study introduce 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, Tsui, F. *et al.*(1995) show a discrete wavelet transform (DWT) to reduce the training time of the neural network. It allows us to compute wavelet coefficients from coarse to fine scale levels efficiently.

Economic and financial data are often analyzed in either the time domain or the frequency domain. If the data is stationary, then these are useful approaches. However, it is often the case that economic and financial data are non-stationary, or non-homogenous in some sense. In these cases, it is instructive to look at the data either on the time-frequency plane or on multiple scales over time. Time-frequency or time-scale methods allow us to observe the changes in behaviour 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 decomposition 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 in the scale (frequency) domain via dilations.

The wavelet analysis is robust tool that may be used to obtain qualitative information for highly nonstationary time series. Specially it may be used 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 decomposition algorithms that can forecast with greater accuracy than existing models. First, Several models including random walk, mean reverting, ARMA, and artificial neural networks are applied to the original series of daily Korean Won / U.S. Dollar returns, and then our neural network models to the Fourier and wavelet decomposed series as a preprocessed data.

The results from the two approaches are compared to see whether the use of the decomposed series which is the special smoothed version of the original series in forecasting fields better results than applying the models directly to the original series. A neural network is incorporated with signal processing algorithms (such as Fourier and wavelet transform) to forecast the exchange rates. That is, the decomposed series are used for forecasting.

This paper is organized into four sections. The next section reviews the financial market heterogenity. The third section describes the time series decomposition method and filtering methods. The fourth section presents the integration methods of joint time-frequency analysis and neural networks training and its experimental results and the conclusion contains final comments.

Financial Market Heterogenity

The heterogeneous market hypothesis has been presented in Müller *et al.*(1993) and associated with fractal phenomena in the empirical behavior of foreign exchange rate(FX) markets. A scaling law relating time horizon and size of price movements (volatility) has been identified in Müller *et al.*(1990).

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 intra-day 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 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 reflects the view of long-term trader and a fine time grid that of a short-term trader.

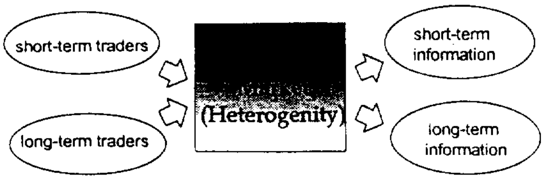


Figure 1. Heterogeneous Financial Market Structure

By the above contents, the heterogeneous markets hypothesis is characterized by the following interpretations of the empirical findings: (Müller, 1995)

The fractal structure will show up in every two-by-two currency exchange rate fluctuation, influenced by al the factors we have considered. These include the distributions of the other currencies and the constraints they experience in their fluctuation, for instance, the constraints imposed by the European Monetary System (EMS).

Since 1973, with the implementation of the floating exchange rate system by the industrialized countries, a large amount of research work has been carried out in an attempt to explain the movements of exchange rates. It is in the areas of hedging, risk management, arbitrage, mispricing, estimation of volatility, that neural network learning is likely to produce much better and quicker results.

According to the Fractal Market Hypothesis, these periods of instability occur when the market loses its fractal structure: when long-term investors are no longer participating, and risk is concentrated in one, usually short, investment horizon. In measured time, these large changes affect all investment horizons. In measured time, these large changes affect all investment horizons. Despite the fact that long-term investors are not participating during the unstable period, the turn in that horizon is still impacted, because they either have left the market or have become short-term investors. The infinite variance syndrome affects all investment horizons in measured time.

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). If these components are assumed to be independent and additive, one can write the time series Z_t as

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

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.

This method of decomposing a time series component of different frequencies has some drawbacks. One of them is that to evaluate the spectrum at a specific frequency ω_0 one need to use the value of the series along its entire range. In addition, to capture local fluctuations of a series, usually, a great amount of cycles of different frequencies may be required. In the last decade a new methodology has appeared. This methodology has been considered in many fields of pure and applied science. This is the wavelet methodology and wavelet transforms which is a more powerful alternative to the Fourier analysis and Fourier Transform.

The idea of decomposing a complicated signal into simpler forms is very attractive for both signal and system analysis. One approach to time domain analysis describes the input as a sum of weighted impulses and finds the response as a sum of weighted impulse responses. This describes the process of convolution. Since the response is, in theory, a cumulative sum of infinitely many impulse responses, the convolution operation is actually an integral. One of the most useful results is that convolution is replaced by the much simpler operation of multiplication when we move to a transformed domain.

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. Peaks in the power spectrum are considered evidence of cyclical behavior. Spectral analysis imposes an unobserved periodic structure on the observed nonperiodic time series. Instead of a circle, it is a sine or cosine wave. (Peters, 1994)

Granger (1964) was the first to suggest that spectral analysis could be applied to market time series. His results were inconclusive. Over the years, various transformations of the data were performed to find evidence of cycles that, intuitively, were felt to be there; but they could not be found. Finally, most of the field gave up and decided that the cycles were like the lucky runs of gamblers - an illusion.

Unfortunately, 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. 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, R/S analysis can perform that function.

R/S analysis has an effect in uncovering periodic cycles, even when the cycles are superimposed on one another. R/S analysis can discern cycles within cycles.

Fig. 2 shows an example of the cyclic features. 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.

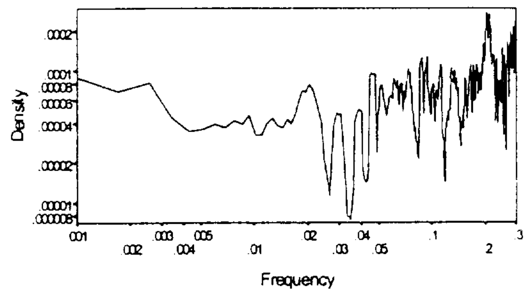


Figure 2. Fourier power spectrum on a log-log scale of Korean Won/U.S. Dollar (1992. 1.- 1996.6.)

Filtering Methods

A filter is a means of separating the various periodic components of a time series into individual series.

Many novices or technicians use moving averages because the moving average offers a smoother visual image of the market trend. In effect, the moving average removes the noise around the trend.

This concept of eliminating noise from the trend is similar to what engineers strive for in their application of digital filters. Hamming observed digital filtering includes the process of smoothing, predicting, differentiating, integrating, separation of signals, and removal of noise from a signal.

This quote applies to the vast majority of indicators in technical analysis. Moving averages, be they simple, weighted, or exponential, are low-pass filters; low-frequency components in the signal pass through with little attenuation or reduction, while high frequencies are severely reduced. Oscillator-type indicators, such as moving average convergence/divergence(MACD), momentum and relative strength index, are another type of digital filter referred to as a differentiator or high-pass filter.

The filtering of the input signal is some transformation of it, e.g. a low-pass filter, or convolution with a smoothing function. Low-pass and high-pass filters are both considered in the wavelet transform, and their complementary use provides signal analysis and synthesis.

Tsui, F. *et al.* (1995) used a discrete wavelet transform to reduce the training time of the neural network. At that time, the wavelet transforms allow us to compute wavelet coefficients from coarse to fine scale levels efficiently.

By subsampling the raw data and interpolating with the wavelet basis, we obtain the wavelet coefficients at each scale level. Therefore, fewer numbers of wavelet coefficients are obtained at coarse scale levels.

Time Domain Filters-ARMA Filter

Time domain filters mean linear Filters using statistical parameters of such statistical models as ARMA.

A. Moving average(MA) models

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

This equation describes a convolution filter: the new series x is generated by an qth-order filter with coefficients $b_0, \ldots, b_q$ from the series e. Statisticians and econometricans call this an qth-order moving average model, MA(q). Engineers call this a finite impulse response (FIR) filter, because the output is guaranteed to go to zero at q time steps after the input becomes zero.

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)

B. Autoregressive (AR) models

MA(or FIR) 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:

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

The is called an 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, e_t can represent either a controlled input to the system or noise.

C. ARMA models (Weigend, 1994)

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

ARMA models have dominated all areas of time series analysis and discrete time signal processing for more than half a century. For example, in speech recognition and synthesis, Linear Predictive Coding (Press *et al.*, 1992) compresses speech by transmitting the slowly varying coefficients for a linear model (and possibly the remaining error between the linear forecast and the desired signal) rather than the original signal. If the model is good, it transforms the signal into a small number of coefficients plus residual white noise.

Frequency Domain Filters

(1) Fourier Transform

Knowledge of the frequency components provides a means of estimating where in the cycle the present time series has 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.

Signal processing is an engineering discipline that deals primarily with the implementation of filters to remove or reduce unwanted frequency components from an information-bearing signal.

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, and medical technology. However, all the application-specific filter implementations can be grouped into four general filter types: lowpass, highpass, bandpass, bandtop. The characteristic frequency response of these filters is depicted in Figure 3. The adaptive filter has characteristics unique to the application it serves. It can reproduce the characteristics of any of the four basic filter types, alone or in combination. (Freeman and Skapura, 1991)

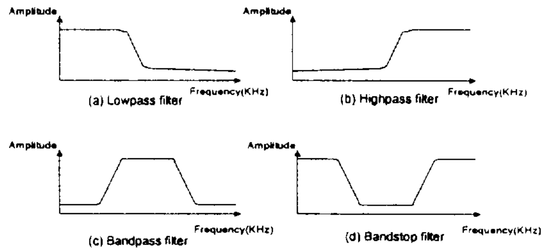


Figure 3. Frequency-response characteristics of the four basic filter types

The Fourier transform's utility lies in its ability to analyze a signal in the time domain for its frequency content. The transform works by first translating a function in the time domain into a function in the frequency domain. The signal can then be analyzed for its frequency content because the Fourier coefficients of the transformed function represent the contribution of each sine and cosine function at each frequency. An inverse Fourier transform does just what you'd expect transform data from the frequency domain into time domain.

Fast Fourier transforms (FFT) have tree requirements about the data being analyzed. (Hartle, T., 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 substract the trend values from the original data.

Second, recall the beginning and end points of the data must have approximately the same values for fast Fourier transform.

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.

The premise of data smoothing is that one is measuring a variable that is both slowly varying and also corrupted by random noise. Then it can sometimes be useful to replace each data point by some kind of local average of surrounding data points. (Press, W. *et al.*, 1992)

Since nearby points measure very nearly the same underlying value, averaging can reduce the level of noise without (much) biasing the value obtained.

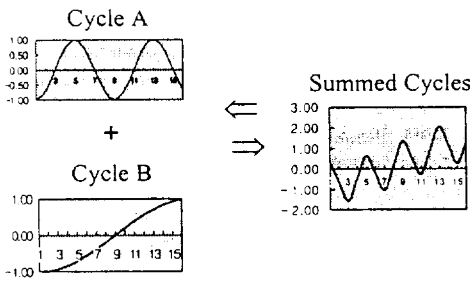


Figure 4. Fourier transform Methods

A. FIR filters (the frequency forms to MA models)

Sometimes it is more convenient to describe the filter in the frequency domain. This is useful and simple because a convolution in the time domain becomes a product in the frequency domain. If the input to a MA model is an impulse (which has a flat power spectrum), the discrete Fourier transform of the output is given by

$$Y_t = \sum_{j=0}^{q} b_j \exp(-i2\pi jf) \qquad (5)$$

The power spectrum is given by the squared magnitude of this:

$$\left|1 + b_1 e^{-i2\pi 1f} + b_2 e^{-i2\pi 2f} + \ldots + b_q e^{-i2\pi qf}\right|^2 \qquad (6)$$

B. IIR (Infinite Impulse Response or Recursive) filters

Recursive filters, whose output at a given time depends both on the current and previous inputs and on previous outputs, can generally have performance that is superior to nonrecursive filters and with the same total number of coefficients (or same number of floating operations per input point).

Probably the best known and most commonly used method for the design of IIR digital filters is the bilinear transformation of the classical analog filters (the Butterworth, Chebyshev I and II, and Elliptic filters).

The general IIR filter is given by

$$Y_t = \sum_{i=1}^{p} a_i Y_{t-i} + \sum_{j=0}^{q} b_j e_{t-j} \qquad (7)$$

If p = 0, so that there is no first sum in function ( ), then the filter is called nonrecursive or finite impulse response(FIR) filter. If $p \neq 0$, then it is called recursive or infinite impulse response (IIR) filter.

(2) Wavelet Transform

Recently, a new decomposition method known as wavelet decomposition was introduced, which is accomplished through the use of an orthogonal basis consisting of so-called 'wavelets'. Wavelet analysis is a significant advance over Fourier analysis for two reasons. First, while Fourier analysis gives us only frequency information, wavelet analysis gives us both frequency information and time information. Secondly, wavelet analysis 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, b), the applicability of wavelet analysis to forecasting needs to be further developed. Tak, B.(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, Johnsotne, Kerkyacharian, and Picard 1994a; Donoho and Johnstone, 1994; Donoho, Johnstone, Kerkyacharian, and Picard, 1994b; Donoho, Johnstone, Kerkyacharian, and Picard, 1993; Donoho and Johnstone, 1993). Wavelets are well-suited to this type problem because of their properties of time (or space) localisation. When approximating a signal, wavelets can preserve local features (discontinuities, turning points, etc) while still removing noise.

A traditional approach to spectral analysis is to assume that the time series can be composed of sinusoidal signals of varying frequency and amplitude, known as Fourier analysis. (Freeman and Skapura, 1991)

This involves representing the original time-domain series in the frequency-domain, providing a global view of the series' spectrum, or frequency components. Knowledge of the frequency components provides a means of estimating where in the cycle the present time series has 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.

There are many different types of discrete wavelet transforms which have been explored since the original work in the 1980's and each have their own advantages and disadvantages when performing data analysis.

Unlike sines and cosines, which define a unique Fourier transform, there is not one single 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.

Daubechies has discovered that the WT can be implemented with a specially designed pair of finite impulse response (FIR) filters called a quadrature mirror filter (QMF) pair. A FIR filter performs the dot product (or sum of products) between the filter coefficients and the discrete data samples. The act of passing a set of discrete samples, representing a signal, through a FIR filter is a discrete convolution of the signal with the coefficients of the filter.

The aim of a wavelet transform is to decompose any signal f into a summation of all the possible wavelet bases at the different scales.(Alsberg, *et al.*, 1997)

The discrete wavelet transforms (DWT) in Fig. 5, as developed by Daubechies, has many similarities to the fast Fourier transforms (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. The transform cosines as the transform output. The wavelet transform yields a decomposition which is neither continuous nor unique. The primary advantage of the DWT is that it 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, for obvious reasons. The output of the DWT consists of these remaining components and all the detail components that were accumulated along the way.

The continuous wavelet transform is defined as the convolution of x_n with a scaled and translated version of $\psi_0(\eta)$: (Torrence and Compo, 1997)

$$W_n(s) = \sum_{n'=0}^{N-1} x_{n'} \psi^* \left[\frac{(n'-n)\delta t}{s} \right] \quad (8)$$

where the (*) indicates the complex conjugate. By varying the wavelet scale s and translating along the localized time index n, one can construct a picture showing both the amplitude of any signals within the series and how this amplitude varies with time.

Suppose the finest scale provides the original data, $x_n = x$, and the approximation at scale m is x_m to x_{m+1}, the detail signal, is yielded by the wavelet transform. If e_m is this detail signal, then the following holds:

$$x_{m+1} = H^T(m)x_m + G^T(m)e_m \quad (9)$$

where G(m) and H(m) 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).

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

$$Wx = e = [e_{N-1} \cdots e_0 x_0]^T \quad (10)$$

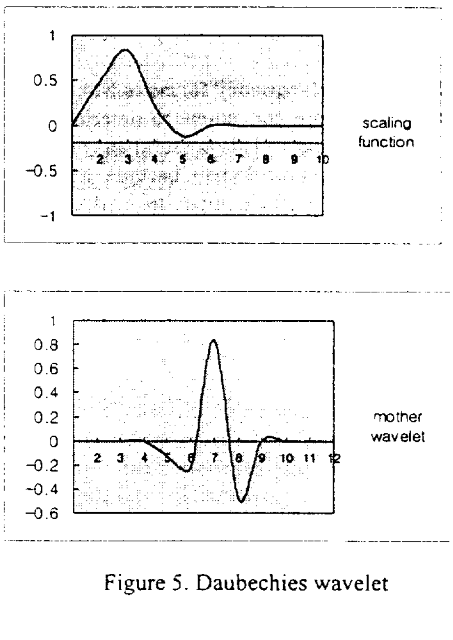


Figure 5. Daubechies wavelet

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 (ARMA)
(3) NN technique using the time-frequency domain filters (Fourier Transform, 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.
The FFT decomposes the summed cycles such as cycle A +B into separate cycles and then plots the power versus frequency of each cycle.(Fig. 4) This is called a power spectrum. Wavelets have several properties that make them useful in the sequential data analysis as follows.
First, wavelets are localized in time (in contrast to Fourier components, which extend over the entire interval being analyzed).
Second, since wavelets comprise a hierarchy of resolution scales, they are good for detecting and characterizing structure that extends over a broad range of scales (i.e. fractals, random walks or other self-similar physical processes).
Third, there is a close relationship between the magnitudes of the individual wavelet coefficients and the smoothness of the corresponding function – a relationship not always shared by Fourier analysis.
Wavelets provide a method of representing the recent past in great detail, while representing the more distant past less precisely, only keeping important low frequency information.
We suggest two integration methods of filtering methods and neural network techniques.
One of them is to forecast one period-ahead returns using single neural network model.
In the other integration method, a multiscale transform of time-varying data can allow forecasts of each scale using multiple neural network models, followed by combining of the individual forecasts.

Recurrent Neural Networks (RNN)

The RNN which has feedback connections between layers allows the dynamics of the signal to be captured.
The RNN employed in this research can be described by a nonlinear ARMA model. In this paper, we select the order based on the experimental results.
The network structure we used has 4 inputs, 4 nodes in hidden layer, and 1 output. The features of inputs are shown in Fig. 6A, 6B, and 6C.
The number of past values used in the networks, 4, was chosen from the nonlinear dynamic analysis (embedding dimension) and proved from a set of values we tested because it gave the smallest in-sample prediction error.
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. Bandpass filters pass only those periodic components in the vicinity of the specified frequency.

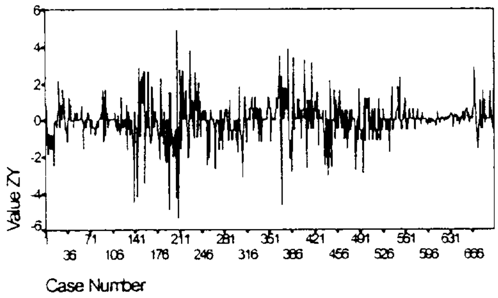
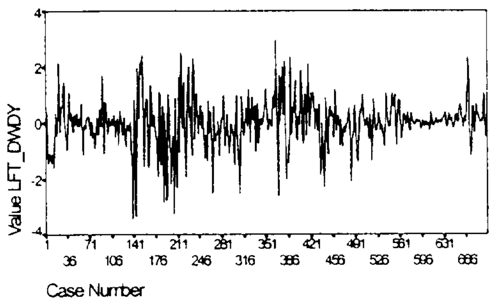
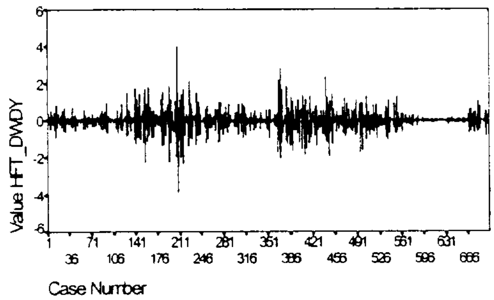


Figure 6A. Original Time Series - Daily Korean Won / U.S. Dollars return ($\ln X_t - \ln X_{t-1}$)

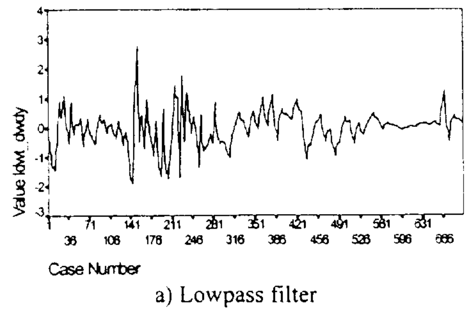


a) Lowpass filter

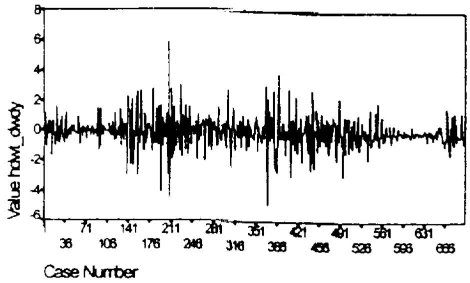


b) Highpass filter

Figure 6B. Fourier Transformed Time Series



a) Lowpass filter



b) Highpass filter

Figure 6C. Wavelet Transformed Time Series

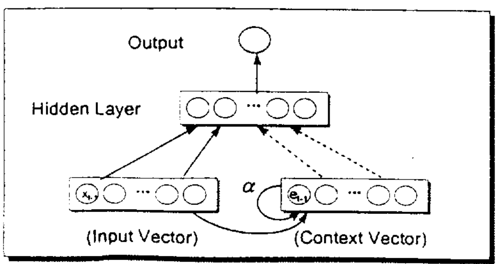
Several researchers have confirmed the superiority of RNNs over feedforward networks when performing nonlinear time series prediction. Especially, the recurrent neural network whose output is fed back to the input layer can provide longer term multi-step predictions, however, a large network size is 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. By making α in Fig. 7 closer to 1 the memory can be made to extend further back into past, at the expense of loss of sensitivity to detail. So the value of α should be chosen so that the decay rate matches the characteristic time scale of the input sequence.

Stornetta *et al.* Model (1988) in Fig. 7 – 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.

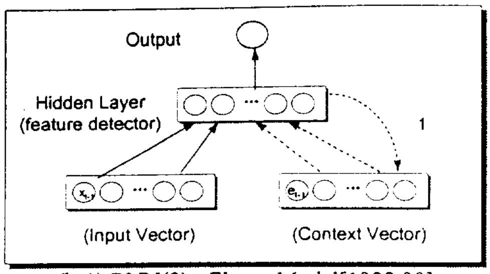
Elman (1990) suggested the architecture shown in Fig. 7 – RNN(2). The input layer is divided into two parts: the true input units and the context units. The context units simply hold a copy of the activations of the hidden units from the previous time step. Modified Elman model in Fig. 7-RNN(2) has additional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

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

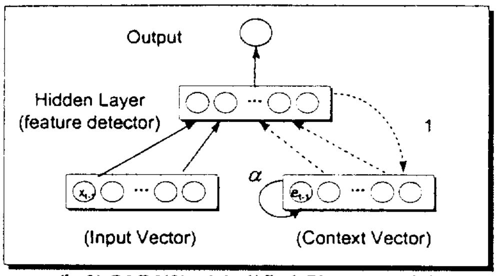
Generally RNN(1) is less well suited than RNN(2) and RNN(3) to generating or reproducing sequences.



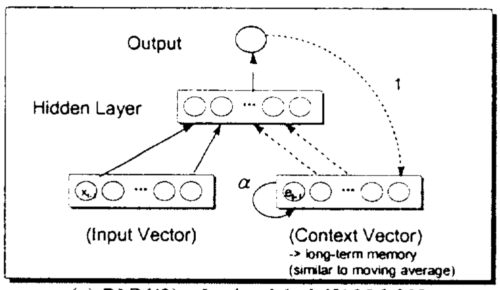
(a) RNN(1) : Stornetta *et al.* Model[1988]



(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

Fig. 8 shows two types of integration models, i.e. (1) single recurrent neural network model combined with frequency domain filtering method and (2) recurrent neural network model combined with Time-Frequency Filtering Analyses.

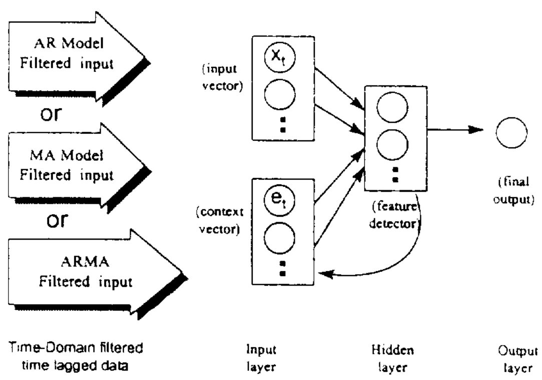
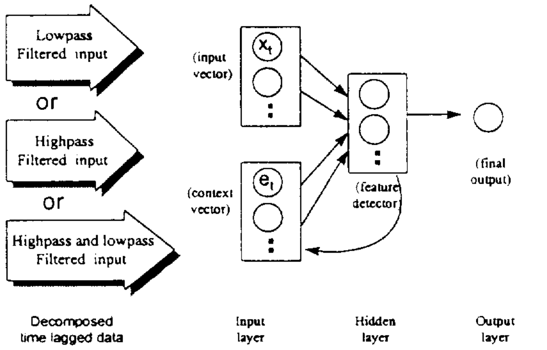
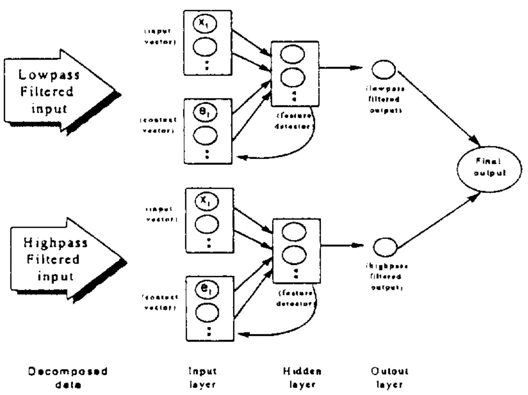


Figure 8A. Integration of Time Domain Filtering Methods and Neural Network Learning Alogrithms



(1) Single recurrent neural network model combined with frequency domain filtering method



(2) Multiple recurrent neural network models combined with frequency domain filtering method

Figure 8B. Recurrent neural network model combined with Time-Frequency Filtering Analyses

Empirical 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 using the logarithm and through standardization from January, 10 1992 to June, 25 1996. That is, the returns are defined as the logarithm of today's exchange rate divided by the logarithm of yesterday's exchange rate. The learning phase involved observations from January, 10 1992 to August, 4 1995, while the testing phase ran from August, 7 1995 to June, 25 1996.

The decomposed series, an approximation part and a detail part of the daily series

In this study, we show the effectiveness of forecasting models through two comparative analyses using different exchange return forecasting models as follows.

Nonlinearity Analysis

A. Rescaled Range Analysis

Hurst (1965) is responsible for 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 rs is related to the Hurst exponent by the following formula:

$$rs = \left(\frac{p}{2}\right)^{H} \tag{11}$$

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

The Rescaled range analysis for Korean Won/ U.S. Dollars Using log difference for the exchange rates is as follows. (Fig. 9)

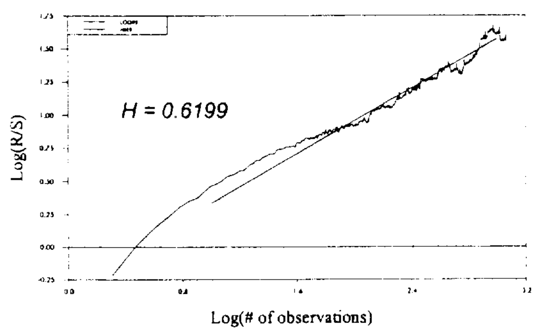


Figure 9. R/S (Rescaled range Analysis)

B. Embedding dimension (1992.1.-1996. 6.)

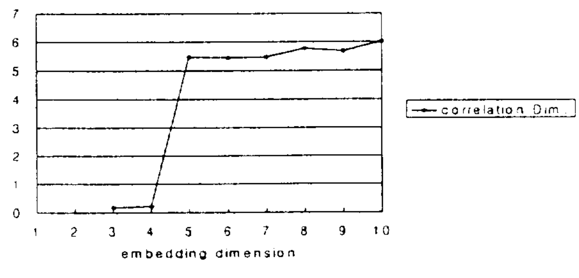


Figure 10. Correlation Dimension vs. Embedding Dimension

Fig. 10 provided the 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.

Neural Networks for univariate time series forecasting can be defined as a nonlinear AR model and the choice of order for the model is based on the embedding dimension of the series.

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 of 5 indicates that 4 data must be shown to a neural net. To predict the 5th data point of the time series. While short-term predictions are in principle possible for a chaotic time series, such predictions might prove difficult in practice because of the high value of the correlation dimension.

Neural networks provide a reliable basis for modeling nonlinear, dynamic market signals. Nonlinear dynamics and chaos theory can provide important inputs for the design of forecasting systems using neural networks.

Selecting the Order of the ARMA Model

There are several heuristics to find the right order such as the Akaike Information Criterion (AIC), Schwartz's Bayesian Criteron (SBC). But these heuristics rely heavily on the linearity of the model and on assumptions about the distribution from which the errors are drawn. When it is not clear whether these assumptions hold, a simple approach (but wasteful in terms of the data) is to hold back some of the training data and use these to evaluate the performance of competing models. According to this criteria, we selected ARMA(0,1) Model as the optimal ARMA model.

Table 1. The estimates of ARMA identification

	Standard error	*Log likelihood*	*AIC*	*SBC*
ARMA(1,0)	1.0792	-1042.65	2089.30	2098.40
ARMA(2,0)	1.0779	-1041.30	2088.61	2102.25
ARMA(0,1)	**1.0767**	**-1041.03**	**2086.07**	**2095.16**
ARMA(0,2)	1.0773	-1040.91	2087.83	2101.47
ARMA(1,1)	1.0773	-1040.92	2087.84	2101.49
ARMA(2,1)	1.0780	-1040.89	2089.78	2107.97
ARMA(1,2)	1.0775	-1040.57	2089.15	2107.34
ARMA(2,2)	1.0749	-1038.39	2086.78	2109.53

AIC = -2ln(maximum likelihood) +2k
SBC = -2ln(maximum likelihood) + kln(n)
(k= the number of total parameters, n = sample size)

Model selection is a general problem that will reappear even more forcefully in the context of nonlinear models, because they are more flexible and, hence, more capable of modeling irrelevant noise.

For nonlinear models to be useful, there must be a process that must be a process that exploits features of the data to guide the construction of the model.

Determination of the Optimal Filter

Whether Fourier transformation is lowpass, highpass, or bandpass is characterized by two parameters: 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).

The width parameter is trickier to specify. There is no simple calculation to provide the correct value. It is arbitrary choice. Unfortunately, a real tradeoff is involved.

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

The simplest wavelet shrinkage technique is so-called hard thresholding. The coordinates of d are replaced by 0 if they are smaller in absolute value than a fixed threshold λ as a filtering criteria. The threshold λ is a tuning parameter of wavelet shrinkage. Donoho and Johnstone propose several thresholds (i.e., universal, SURE), as well as several thresholding policies. Nason (1994) adjusted the well-known cross-validation method for use with wavelets. The threshold is selected by minimizing a cross-validation methods for use with wavelets. The threshold is selected by cross-validatory estimator of integrated square error (ISE). 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 dynamical 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.

This issues of the optimal filters deviate from our main topics in this study. So, in this study, we use 25% and 50% as a threshold λ value of Daubechies wavelet transform and fast Fourier transform based on root mean square errors (RMSE) of our models changing the threshold filter from 10 % to 90%.

Comparative Analysis

We have the comparative analyses as follows.

(1) The Comparative Effects of Non-Filters and Filters on the Model Performances
(2) The Comparative Effects of Filter Types on the Model Performances

- Time domain Filters and Frequency domain Filters
- Fourier Filters and Wavelet Filters
- Filtering Criteria

First of all, As shown in Table 2, we use random walks, mean reverting, ARMA(p,q), back-propagation neural networks (BPN), and recurrent neural networks without filtering method. It focuses on comparing each other in the predictability of the models. Recurrent neural networks outperformed the other models and Elman models have the best performance among them.

Table 2. Non-Filtered Forecasting Model - Benchmark Models

Model	*Train*	*Test*
BPN	0.835638	0.877505
RNN(1)	0.854616	0.894718
RNN(2)	**0.835809**	**0.871192**
RNN(3)	0.846807	0.885184
ARMA(0,1)	0.854402	0.882932
Random Walks	1.070294	1.062600
Mean reverting	0.946250	0.945900

RNN(1) : Stornetta *et al.* Model[1988]
RNN(2) : Elman Model[1988,90]
RNN(3) : Jordan Model[1986,89]

To show the results in Table 3, we have trial or error for finding the optimal pth or qth order of AR(p), MA(q), or ARMA(p,q) filter combined to the recurrent neural network models. Table 4 shows more analytical comparisons than table 3. That is, we find that Table 4 presents the results of the model validation for each experiment. In each case, the neural network models combined with joint time-frequency filtering methods outperformed both simple neural networks and the models with time-domain filtering. As expected, all the integration of joint time-frequency analysis 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 filtering and neural network models. But, an unexpected result was the difference in predictability between the FT_MRNN and WT_MRNN. That is, FT_MRNN outperformed WT_MRNN.

Generally wavelet decomposes the input signal into detail signals, and a residual or the time series into varying scales of temporal resolution. The original signal can be expressed as an additive combination of the wavelet coefficients, at the different resolution levels.

Table 3. The Integration of ARMA Filters and Elman Recurrent Neural Network Models

Filter Types	*Train*	*Test*
AR(1) filter	**0.886288**	**0.947860**
MA(1) filter	0.886420	0.948424
ARMA(1,1) filter	0.886114	0.949303

Table 4. The Integration of Time-Frequency Filtering Methods and Neural Networks

Model	*Train*	*Test*	*Fitness order*
LFT_RNN(2)	0.889336	0.939239	6
HFT_RNN(2)	0.889659	0.948427	7
LWT_RNN(2)	0.842508	0.939097	5
HWT_RNN(2)	0.889990	0.948492	8
FT_RNN(2)	0.263899	0.26952	2
FT_MRNN(2)	0.251940	0.253377	1
WT_RNN(2)	0.735958	0.80823	4
WT_MRNN(2)	0.718174	0.792379	3

LFT_RNN(2) = Lowpass Fourier Filter + Single RNN(2)
HFT_RNN(2) = Highpass Fourier Filter + Single RNN(2)
LWT_RNN(2) = Lowpass Wavelet Filter + Single RNN(2)
HWT_RNN(2) = Highpass Wavelet Filter + Single RNN(2)
FT_RNN(2) = Both Fourier Filters + Single RNN(2)
FT_MRNN(2) = Both Fourier Filters + Multiple RNN(2)
WT_RNN(2) = Both Wavelet Filters + Single RNN(2)
WT_MRNN(2) = Both Wavelet Filters + Multiple RNN(2)

Table 5. Comparative Analysis

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

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

(b) Paired samples t test for the differences in RMSE (Test Samples)

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

*: 10% level , **: 5% level, ***:1% level

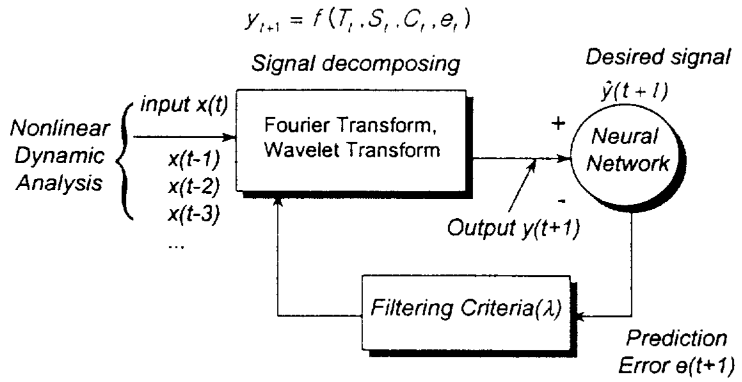


Figure 11. Integration Framework of Joint Time-Frequency Analysis and Neural Networks

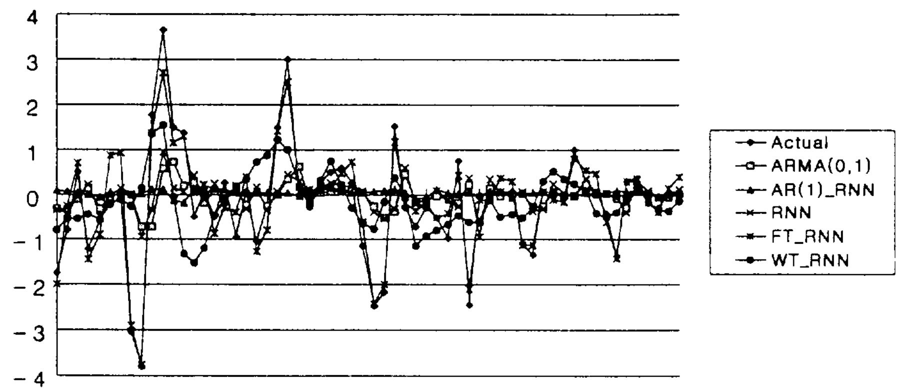


Figure 12. Comparative Analysis of Non-Filtered Forecasting Models
and Filtered Forecasting Models using the test samples

Conclusions

Signal processing is a discipline embodying a large set of methods for the extraction of information bearing signals from corrupted observations. Wavelet theory is a concept that will need to be developed further for use in economics and finance.

A number of filtering methodologies were investigated in the context of forecasting using daily Korean Won / U.S. Dollar series. First, we found that the use of the approximation part and the detail part in forecasting yields better results than applying the models directly to the original series. Second, in integration with neural network models we detected that joint time-frequency filtering (Fourier or wavelet transform) outperformed the time domain filtering. Third, even though wavelet transform methods are better methodology than Fourier transform methods, the performances of this study show the unexpected results.

As a result of our experiments, we have found that the neural network models combined with joint time-frequency analysis consistently outperform the other neural network models. But, in the comparison of Fourier transforms with wavelet transforms, the performance of wavelet transforms were not as well as Fourier transforms because this study has a few limitations as follows.

First, we used a Daubechies wavelet as wavelet transform method used popularly. But there are a lot of wavelet transform methods in wavelet analysis. Each wavelet method has its own features.

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

Therefore in the further research we will analyze the comparative analysis between different wavelet transform methods and so detect the optimal features of wavelet transform method for the financial time series forecasting.

References

Ambardar, A., *Analog and Digital Signal Processing*, PWS Publishing Company, 1995

Azoff, E., *Neural Network Time Series Forecasting of Financial Markets*, John Wiley & Sons, 1994, pp.33-35.

Brock, W. and W. Dechart, "Theorems on distinguishing deterministic from random systems," in *Dynamic Econometric Modeling*. (Editors) Barnett, W., Berndt, R.J. and White, P.J. Cambridge: Cambridge University Press, 1991.

Daubechies, I., *Communications on Pure and Applied Mathematics*, vol. 41, 1988, pp.909-996.

Daubechies, L., "Orthogonal Bases of Compactly Supported Wavelets," *Comm. Appl. Math.*, XLI, 99, 1988.

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

Elman, J.L. and D. Zipser, "Learning the Hidden Structure of Speech," *Journal of the Acoustical Society of America* 83, 1988, pp.1615-1626.

Elman, J.L., "Finding Structure in Time," *Cognitive Science* 14, 1990, pp.179-211.

Episcopos, A. and J. Davies, "Predicting returns on Canadian exchange rates with artificial neural networks and EGARCH-M models," *Neural Computing and Application*, 1995.

Freeman, J.A. and D.M. Skapura, *Neural Networks Algorithms, Applications, and Programming Techniques*, Addison-Wesley Publishing Company, 1991, pp.49-53.

Grauwe, D., H. Dewachter, and M. Embrechts, *Exchange Rate Theory - Chaotic Models of Foreign Exchange Markets*, BLACKWELL Oxford UK & Cambridge USA, 1993, pp.94-95.

Hartle, T., "Preprocessing Data and Fast Fourier Transform," *Technical Analysis of Stock and Commodities*, 1994.

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, pp.531-546.

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

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.* [Kor. title], Seoul, Sept. 1996, pp.19-22

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.

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

Müller, *et al.*, "Fractals and Intrinsic Time- A Challenge to Econometricans," 1995.

Peters, E., *Fractal Market Analysis: Applying Chaos Theory to Investment and Economics*, John Wiley&Sons, Inc. 1994, pp.87-88.

Press, W. *et al.*, *Numerical Recipes in C: The Art of Scientific Computing*, Cambridge Univ. Press, 1992

Refenes, A.(ed.), *Neural Networks in the Capital Markets*, John Wiley & Sons, 1995, p.214.

Rioul, O. and M. Vetterli, "Wavelet and Signal Processing," *IEEE Signal Processing Magazine*, October 1991, pp.14-35

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, pp.750-759.

Strang, G., "Wavelets," *American Scientist 82*, April, 1994, pp.250-255.

Torrence, C. and G. Compo, " A Practical Guide to Wavelet Analysis," *Bull Amer. Met. Soc.*, 1997

Tsui, F. *et al.*, "Recurrent Neural Networks and Discrete Wavelet Transform for Time Series Modeling and Prediction," *ICASSP-95*, 1995.

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

Wei, *Time Series Analysis - Univarate and Multivariate Methods*, Addison-Wesley Publishing Company, 1990, p.195

Weigend, A., *et al.*(eds.), *Time Series Prediction - Forecasting the Future and Understanding the Past*, Addison-Wesley Publishing Company, 1994, pp.14-15.

Baxter, M., and R.G. King, "Measuring business cycles : Approximate band-pass filters for economic time series," *Working paper*, University of Virginia,1995.