Wavelet Thresholding Techniques to Support Multi-Scale Decomposition for Financial Forecasting Systems

*Taeksoo Shin**, *Ingoo Han**

Abstract

Detecting the features of significant patterns from their own historical data is so much crucial to good performance specially in time-series forecasting. Recently, a new data filtering method (or multi-scale decomposition) such as wavelet analysis is considered more useful for handling the time-series that contain strong quasi-cyclical components than other methods. The reason is that wavelet analysis theoretically makes much better local information according to different time intervals from the filtered data. Wavelets can process information effectively at different scales. This implies inherent support for multiresolution analysis, which correlates with time series that exhibit self-similar behavior across different time scales. The specific local properties of wavelets can for example be particularly useful to describe signals with sharp spiky, discontinuous or fractal structure in financial markets based on chaos theory and also allows the removal of noise-dependent high frequencies, while conserving the signal bearing high frequency terms of the signal.

To date, the existing studies related to wavelet analysis are increasingly being applied to many different fields. In this study, we focus on several wavelet thresholding criteria or techniques to support multi-signal decomposition methods for financial time series forecasting and apply to forecast Korean Won / U.S. Dollar currency market as a case study.

One of the most important problems that has to be solved with the application of the filtering is the correct choice of the filter types and the filter parameters. If the threshold is too small or too large then the wavelet shrinkage estimator will tend to overfit or underfit the data. It is often selected arbitrarily or by adopting a certain theoretical or statistical criteria. Recently, new and versatile techniques have been introduced related to that problem.

Our study is to analyze thresholding or filtering methods based on wavelet analysis that use multi-signal decomposition algorithms within the neural network architectures specially in complex financial markets.

Secondly, through the comparison with different filtering techniques' results we introduce the present different filtering criteria of wavelet analysis to support the neural network learning optimization and analyze the critical issues related to the optimal filter design problems in wavelet analysis. That is, those issues include finding the optimal filter parameter to extract significant input features for the forecasting model.

Finally, from existing theory or experimental viewpoint concerning the criteria of wavelets thresholding parameters we propose the design of the optimal wavelet for representing a given signal useful in forecasting models, specially a well known neural network models.

Key words: Discrete Wavelet Transform, Wavelet Packet Transform, Wavelet Thresholding Techniques, Neural Networks

1. Introduction

Traditionally, the fluctuation in financial market is treated as white noise. However, it is not true when trend is properly removed and we can clearly observe some business cycles, though they evolve with time. The goal of forecasting is to identify the pattern in the time series and use the pattern to predict its future path.

The issue of generalization in this interpretation becomes one of how to extract useful information from the noise-contaminated data, and to rebuild the pattern as closely as possible, while ignoring the useless noises.

There is a wide and strong interest in feature discovery and transformation among practitioners from statistics, pattern recognition, data mining, and knowledge discovery since data preprocessing is an essential step in the knowledge discovery process for real-world applications.

Linear models can give good prediction results for simple time series, but can fail to predict time series with a wide band spectrum (i.e. called as Granger shape), a stochastic or chaotic time series, in which the power spectrum is not a useful characterization.

A number of new nonlinear techniques, such as neural networks, wavelet, and chaos analysis, promise insight that traditional linear approaches cannot provide (Geva, 1998).

Filtering techniques also have been shown to be useful in estimating coefficients of forecasting models. The principle advantage of applying filtering methods is that the techniques make it possible to isolate relevant frequencies.

While wavelets have recently been used by econometricians, they have played an important role in other fields since the early 1990s (Jensen, 1998; Ramsey and Zhang, 1998; Ramsey and Lampart, 1998). Scientists in diverse fields have observed time series where observations that are far apart (in time or space) were

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

correlated too strongly to be modeled as independent data or classical autoregressive, moving average models (ARMA). This concept of long-memory has grown rapidly and can be found in a broad scattering of fields such as astronomy, chemistry, engineering, environmental sciences, and mathematics, etc. For literature about the purpose and techniques of wavelet analysis in financial markets, refer to Table 4.

Wavelets are mathematical constructs with great potential in statistical methodology. They have been applied extensively in diverse applications, including data compression, signal processing, image analysis, object detection, turbulence, numerical analysis, neural networks, economics, astronomy, and statistics. Specially, statisticians are interested in wavelets as a modeling tool in the general nonlinear regression scheme. Some particular problems of interest are de-noising, density and function estimation, long range dependence, and change point detection. Though the wavelet regression is very attractive, it has limitations stemming from the intrinsic properties of wavelets.

Denoising by thresholding in the wavelet domain has been developed principally by Donoho *et al.* (1995). Donoho and Johnstone (1994) introduced RiskShrink with the minimax threshold, VisuShrink with the universal threshold, and discussed both hard and soft thresholds in a general context that included ideal denoising in both the wavelet and Fourier domains. Donoho and Johnstone (1995) introduced SureShrink with the SURE threshold, WaveJS with the James-Stein threshold, and LPJS also with the James-Stein threshold but in the Fourier domain instead of the wavelet domain.

The differences of All these techniques discussed above are distinguished each other as follows. That is, they can be classified by transform domain, Fourier versus wavelet, as well as by intent of use, ideal versus practical. An ideal procedure requires a prior knowledge of the noise, whereas a practical procedure does not, so that ideal procedures are only used for purposes of comparison in simulation experiments. Moreover, the procedures can be classified according to whether they use a single threshold globally for all relevant parts of the transform, or multiple thresholds locally for different parts of the transform (Fourier frequency bands or wavelet multiresolution levels).

Restricting attention to the practical procedures, SUR, WJS, and FJS appear to perform well, but it is not possible to declare any of the procedures as the best under all test cases and sample sizes.

The extensive literature on wavelet thresholding mainly focuses on two questions. The first is which threshold rule to use. Most popular are hard and soft thresholding. The second, and perhaps most important

Determining the most appropriate procedure necessarily involves experiments to compare the performance of a wavelet shrinkage denoising method (comprised of the most effective combination of wavelet transform parameters and denoising rules and thresholds for the range of sample sizes and noise levels expected) with any other methods under consideration.

Data analysis, for exploratory purposes, or prediction, is usually preceded by various data transformation and recoding. In fact, we would hazard a guess that 90% of the work involved in analyzing data lies in this initial stage of data preprocessing (Murtagh and Aussem, 1996).

This study is intended to explore the wavelet universal thresholding algorithm for denoising data and to compare its performance with that of other commonly used smoothing filters in financial forecasting.

All reconstruction filters are evaluated on the basis of the root mean square error (RMSE). That is, we evaluate the effectiveness of both these transform such as discrete wavelet transform and wavelet packet transform on a series of Korean Won/US Dollar exchange rate market.

The estimation techniques studied in this paper do not assume a priori knowledge about the underlying spectrum, besides the presumption that the signal contains significant coarse-scale coefficients. When a priori information is available, special techniques may be developed to improve performance.

The remainder of this study is organized as follows. The next section reviews time-frequency decomposition, and then discrete wavelet transform (DWT) and wavelet packet transform (WPT). Section 5 introduces thresholding techniques for financial forecasting. Section 6, 7, and 8 describes best-basis selection criteria techniques (Tree Pruning Algorithm). The fifth section describes experimental results and the conclusion contains final comments.

2. Time-Frequency Deocompositions

The Fourier transform is not to be used in case of non-stationary signals.

In the multiscale fast wavelet transform of Mallat and Zhong (1992), the time series is decomposed into different scales of the wavelets, in order to extract its internal representation. Each scale of wavelet coefficients provide a different dimension of the time series in the both time and frequency domains. Recently, due to the similarity between wavelet decomposition and one-hidden-layer NN, the idea of combining both wavelet and NN has been proposed in various works (Bakshi and Stephanopoulos, 1993; 1995; Zhang and Benveniste, 1992; 1997; Geva, 1998).

Our task is to approximate a time series at different levels of resolution using multiresolution decomposition. The individual time series resulting from the decomposition, taken together, can provide a detailed picture of the underlying processes. Nonetheless, knowing the current state of these processes may not be sufficient; that is, in order to make valuable statements about the future, additional, historical information may be required.

A naïve approach would be to apply a bank of filters, with varying frequencies and widths, to the data. Unfortunately, choosing the proper number and type of filters for this is a difficult task. Wavelet transforms provide sound mathematical principles for designing and spacing filters, and for making trade-offs between these objectives, while retaining the original relationships in the time series. These principles define a set of filters obtained by rescaling a given function several times, using what is often called a mother wavelet, which is compressed or expanding in the time domain to produce the wavelets.

3. Discrete Wavelet Transform

Recently, the wavelet transform was introduced as an alternatively technique for time-frequency decomposition.

Wavelets are building block functions and localized in time or space. They are obtained from a single function ψ(t), called the mother wavelet, by translation and dilations. Projection of the signal onto wavelet basis functions is called wavelet transform (WT).

Wavelet transform is a powerful method for multiresolution representation of signal data (Szu *et al.*, 1992).

As any transform, the WT aims to transform the signal from the original to another domain in which some operations on the signal (i.e. denoising, compression) can be carried out in an easier way. The inverse transform allows to go back to the original domain.

Wavelets are any of a set of special functions satisfying certain regularity conditions (Daubechies, 1988; 1992; 1993). Their support is finite; they are non-zero on a finite interval, and they are defined within finite frequency bands. There are two types of wavelet functions: mother wavelet, ψ(t), which describe high-frequency and detail components of a signal, and father wavelets, ϕ(t), which describe low-frequency and smooth components. The major properties of wavelets are smoothness, compactness in both time and frequency, the number of oscillations, and orthogonality, the preference for which is specified by the application. Orthogonality is a property which allows for reconstruction of the original data from the wavelet transform.

The discrete wavelet transform expresses a time series as a linear combination of scaled and translated wavelets. Knowing which wavelets appear in a transform can provide information about the frequency content of the signal for a short time period. Short-time Fourier transforms have been used in the past for this type of information, but how the wavelet transform differs is while the short-time Fourier transform must use a constant time interval for the whole analysis, the wavelet transform uses varying intervals.

The time-scale nature of the WT results in several useful properties.

We assume that N is a multiple of 2^J and consider an orthonormal, discrete wavelet transform (DWT) for siganls on the interval [0, N-1]. The transform is implemented using an iterated bank of subsampled lowpass and highpass quadrature-mirror, finite-impulse-response filters $g(l)$ and $h(l)$, respectively.

We denote by ψ(t) the (continuous) wavelet associated with the filter bank (Daubechies, 1992). In order to keep the exposition of basic concepts clear, the theoretical results which motivate our approach are presented for the simple case of periodic wavelets. This is equivalent to assuming that the signals are periodic beyond the boundaries at l=0 and l=N-1.

The two-scale relation for the Daubechies wavelets is in the form:

Given a mother wavelet ψ(t), for all real a, b (a≠0), we construct a sequence of wavelets by translations and dilations of ψ(t),

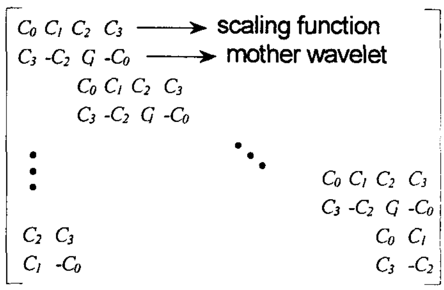
$$\psi_{a,b}(t) = |a|^{-1/2}\psi\left(\frac{t-b}{a}\right) \qquad (1)$$

where a presents the dilation parameter and b the translation parameter.

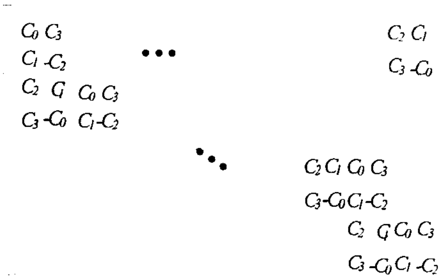
The special feature of the Daubechies wavelet family is the presence of orthogonality in combination with compact support. To achieve this, we must put up with asymmetry and low regularity.

The low-pass filter with n coefficients can be considered as a smoothing filter, which resembles a moving average of n points. In the field of signal processing the pair of filters is known as the quadrature mirror filters (QMF). The low-pass and the high-pass filters are also called the scaling and wavelet filters, respectively. These filters are used to construct the filter matrices, denoted as G and H. For the signal containing eight data points and filter number 2 (characterized by four coefficients) the G and H matrices have the following structure:

(a) Wavelet transform matrix



(b) Inverse wavelet transform matrix



(c) The filter coefficients of Daubechies 4 Wavelet

C#	*Value*
0	$(1+\sqrt{3})/4\sqrt{2} \fallingdotseq$ 0.482962913
1	$(3+\sqrt{3})/4\sqrt{2} \fallingdotseq$ 0.836516303
2	$(3-\sqrt{3})/4\sqrt{2} \fallingdotseq$ 0.224143868
3	$(1-\sqrt{3})/4\sqrt{2} \fallingdotseq$ - 0.129409522

Figure 1. Daubechies 4 wavelet

Once the filters have been defined, one applies the recursive decomposition algorithm introduced by Mallat (1989) and known as the pyramid algorithm or tree algorithm, which offers the hierarchical, multiresolution representation of function (signal).

When we think about performing a wavelet transform on a data sets, we can look at it as running the data through a smoothing (losspass) filter. This is, in effect, computing a moving average of the data. The difference between this computation and that of any other moving average is that the weights are chosen in a very particular manner. This lowpass filter is known as a scaling function ($\phi(x)$). Convolving the data with $\phi(x)$ gives us an approximation of the original series, except with some (high frequency) detail filtered out.

If instead we wish to obtain the detailed information, then we must pass the data through a differencing (highpass) filter. This highpass filter is known as the wavelet ($\psi(x)$).

Scaling functions and the corresponding wavelets are defined by the following dilation equations;

$$\phi(x) = 2\sum_{k=0}^{N} c_k \phi(2x-k). \qquad (2)$$

$$\psi(x) = 2\sum_{k=N}^{0} (-1)^{k+1} c_k \phi(2x-k). \qquad (3)$$

where N+1= p is the order of regularity of the wavelet. This higher the order of regularity of the wavelet, the smoother the wavelet is.

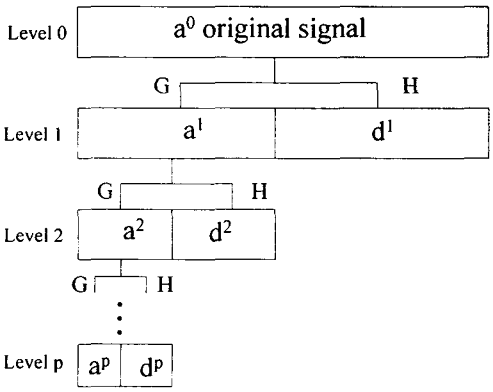


Figure 2. The tree or pyramid algorithm (Mallat, 1989) (G: The lowpass (or scaling) filter; H: The highpass (or wavelet) filter; The highpass filtered data: the wavelet transform detail coefficients ($d^1 = \{d_0^1, d_1^1, \ldots, d_{N/2-1}^1\}$) at the first level of resolution; The lowpass filtered data : the approximation coeffcients ($a^1 = \{a_0^1, a_1^1, \ldots, a_{N/2-1}^1\}$) at the first level of resolution.)

As shown in Figure 2., in the tree algorithm, the set of N input data is passed through the scaling and the wavelet filters.

4. Wavelet Packet Transform (WPT)

In the pyramid algorithm the detail branches are not used for further calculations, i.e. only the approximations at each level of resolution are treated to yield approximation and detail obtained at level m+1. Application of the transform to both the detail and the approximation coefficients results in an expansion of the structure of the wavelet transform tree algorithm to the full binary tree (Coifman and Wickerhauer, 1993; Coifman *et al.*, 1993).

Coifman and Wickerhauser (1993) developed a wavelet packet transform; this is a more general transform than the discrete wavelet transform. The main difference is that while in the discrete wavelet transform the detail coefficients are kept, and the approximation coefficients are further analyzed at each step, in the wavelet packet transform both the approximation signal and the detail signal are analyzed at each step. This results in redundant information, as each level of the transform retains n samples. The process is illustrated in Figure 3.

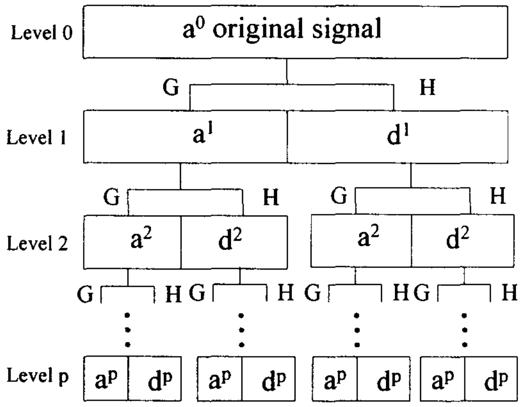


Figure 3. The wavelet packet transform as a complete binary tree (Coifman et al., 1993; Cody, 1994)

The main characteristic of the wavelet packet transform is that it produces an arbitrary frequency split, which can be adapted to the signal. While wavelet packet create arbitrary binary slicing of frequencies (with associated time resolution), they do not change over time. Often a signal is first arbitrarily segmented, and then, the wavelet packet decomposition is performed on each segment in an independent manner.

5. Wavelet Thresholding Techniques for Financial Forecasting

Over the past few years, there has grown up a tremendous literature (Greenblatt, 1995; Donoho *et al.*, 1995) on the use of the wavelet transform for a wide array of applications. It has been used extensively for approximation due to its time-frequency localization properties and its multiresolution analysis approach.

Often it is useful to use different bases at different levels of resolution. Although the wavelet decomposition can approximate at a number of resolutions at once, it isn't adaptive.

Thresholding is a rule in which the coefficients whose absolute values (energies) are smaller than a fixed threshold are replaced by zeroes. In this study, we define wavelet thresholding as denoising and smoothing techniques including best basis algorithm to extract significant multi-scale information from the original time series.

The rationale for wavelet thresholding is that usually the signal will be compressed into a few large coefficients whereas the noise component will give rise to small coefficients everywhere.

Wavelet thresholding methods in general give better results than other nonparametric smoothing methods such as spline smoothing or Fourier analysis and uses a computationally fast algorithms (Downie and Silverman, 1996). The idea is to express a function in terms of an orthonormal wavelet basis, where the basis functions are translations and dilations of a mother wavelet function. The wavelet transformation is often good at compressing a signal into a few large coefficients. If the original signal contains noise then a thresholding method will throw away the small noise coefficients and keep the large signal coefficients.

Wavelet smoothing techniques capitalize on the different properties of signal and noise wavelet components. Techniques such as thresholding or shrinkage of the noisy wavelet coefficients, and reconstruction from the local minima of the wavelet transform, have been successfully used in a variety of denoising problems (Mallat and Hwang, 1992; Lu *et al.*, 1992; Donoho and Johnstone, 1992a, 1992b; Donoho *et al.*, 1992; Donoho, 1992, Johnstone *et al.*, 1992; Moulin, 1993, 1994).

There are many ways to threshold. Basically, the process of thresholding wavelet coefficients can be divided into two steps: the policy choice and the choice of a threshold parameter.

Donoho and Johnstone (1994, 1995) and Donoho *et al.* (1995) showed that such wavelet estimators with a properly chosen threshold rule have various important optimality properties. The choice of thresholding rule, therefore, becomes a crucial step in the estimation procedure. Several approaches to thresholding have been introduced in the literature: a minimax approach (Donoho and Johnstone, 1994; 1995), multiple hypothesis testing (Abramovich and Benjamini, 1995; 1996; Ogden and Parzen, 1996a; 1996b), cross-validation (Nason, 1995; 1996; Weyrich and Warhola, 1995). The idea of thresholding has also been studied in the context of correlated errors (Wang, 1996; Johnstone and Silverman, 1997).

One of the most important problems that has to be solved with the application of digital filters is the correct choice of the filter type and the filter parameters. The most difficult choice is that of the cut-off frequency of the filter which has to be specified either explicitly or implicitly (Mittermayr *et al.*, 1996). It is often selected arbitrarily or by adopting a certain theoretical model. Thus this study focuses on a comparative study to motivate the choice of the filter parameters that are most appropriate for a specific problem.

During the last decade a new and very versatile technique, the wavelet transform (WT), has been developed as a unifying framework of a number of independently developed methods (Mallat, 1989; Meyer, 1989; Daubechies, 1992).

By setting all coefficients above a certain frequency f_c (the cutoff frequency) to zero one tries to remove only noise, while preserving the information on the signal. Gaussian white noise by definition has a constant contribution to all frequencies and thus one has to compromise when setting the cutoff frequency (Larivee and Brown, 1992).

The parsimony of wavelet transformations ensures that the high frequency features of the series can be described by a relative small number of wavelet coefficients. We can separate the high frequency component and the low frequency component of the time series by wavelet thresholding.

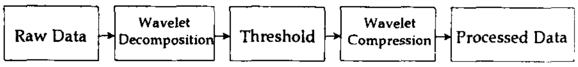


Figure 4. Inference on Wavelet Coefficients (Moulin, 1994)

Prior studies (Abramovich and Benjamini, 1994; Donoho and Johnstone, 1994; Donoho, 1995; Donoho *et al.*, 1995; Nason, 1996) discuss several different thresholding techniques for single wavelet decompositions. These aim to reduce the noise in an observed signal.

Vidakovic (1995) proposed a thresholding method based on the Lorentz curve for the energy in the wavelet decomposition. Replace the $p_0\times$ 100% of the coefficients with the smallest energy with zero.

$$p_0 = \frac{1}{n}\sum_i 1(d_i^2 \le \bar{d}^2) \quad (4)$$

where $\bar{d}^2$ is the mean of the energies $(d_1^2, d_2^2, \cdots, d_n^2)$. The value p_0 represents the proposition at which the gain by thresholding an additional element will be smaller than the loss in the energy. Besides, a few researchers studied these methods differently (Table 1).

In summary, we show a list about wavelet thresholding techniques in Table 1.

Table 1. Wavelet Thresholding Techniques

Authors	*Thresholding Methods*	*Tresholding Rules*
Donoho and Johnstone (1994)	Universal(VisuShrink) - Minimax approach	$\lambda = \sqrt{2\log(n)}\tilde{\sigma}$, $\delta(d,\lambda) = d\,1(d > \lambda$ for all the wavelet coefficients d
Donoho and Johnstone (1995)	Adaptive (SureShrink) - Minimax approach	Based on Estimator of Risk
Nason(1994,1995 ,1996), Jensen and Bultheel (1997)	Cross-Validation	$CV = \frac{1}{n}\sum_i^n (y_i - \tilde{y}_i)^2$
Abramovich and Benjamini (1995, 1996), Ogden and Parzen (1996)	Multiple hypothesis tests	Test if each wavelet coefficient is zero or not.
Vidaković (1994), Clyde *et al.* (1995), Chipman *et al.* (1997)	Bayes Rule	
Goel and Vidaković (1995)	Lorentz curve	$p_0 = \frac{1}{n}\sum 1(d_i^2 \le \bar{d}^2)$

Abramovich and Benjamini (1995)	The False Discovery Rate (FDR) approach to multiple hypo. testing	
Johnstone and Silverman (1997)	Level-dependent Threshold	

6. Best Basis Selection Criteria Techniques (Tree Pruning Algorithm)

The flexibility of the WPT is in choosing the appropriate basis for the particular problem.

Definition of the best-basis for a set of signals is problem dependent. The best basis for spectra that will be used for classification purposes can differ from the best basis for calibration purpose. If we are interested mainly in data compression, the best basis preserving best the data variance can be considered as the optimal one (Walczak and Massart, 1997).

Elimination of the small wavelet coefficients related with the noise of variance spectrum allows significant compression of the best basis.

The definition of the best-basis can also explicitly contain the criterion of the coefficient selection. For stance, the best basis can be defined as the basis with the minimal number of coefficients, whose absolute value is higher than the predefined threshold.

One way of selecting an efficient basis from all possible orthonormal bases (i.e. form a library of wavelet packet bases) is to apply the entropy or information criterion (Coifman and Wickerhauser, 1992; Coifman *et al.*, 1994), since the amount of information is a measure of inequality of distribution. A basis with coefficients all giving more or less the same values would yield a low information or high entropy value. The best basis can be defined as the basis giving the minimum entropy or maximum information for its distribution of coefficients. This definition of the best basis requires the criterion of the coefficients selection.

The chosen wavelet basis should correspond to the expected smoothness of the function or signal. The types of wavelet basis are various (i.e. Haar, Daubechies, Coiflet, Morlet, etc.).

6.1. Best Orthogonal Basis

Coifman and Wickerhauser (1992) have suggested a method for adaptively selecting the best basis. If we were to check every combination of components that make up orthogonal bases, it would be incredibly computationally intensive. Fortunately, the proposed approach does not require us to examine every possible basis. Instead, it places the decomposition into a tree structure and provides efficient algorithms for searching that tree.

They defined the best basis to be that which minimized an information cost function M and chose the Shannon entropy as their archetype for M.

The Best-Ortho-Basis methodology of adaptive time-frequency analysis (Coifman *et al.*, 1994) has, more recently, caught the interest of a wide community of applied mathematicians and signal processing engineers.

Based on ideas of recursive partitioning of the time-frequency plane, it develops, from an analysis of a given signal, a segmented basis, where the segments are terminal nodes in a data-driven recursive segmentation of the time axis.

Table 2. Best Basis Wavelet Packet: The Single-Tree Algorithm

Authors	*Best basis algorithms*	*Contents*
Daubechies (1988)	Method of Frames (MOF)	Synthesis direction approach A straight-forward linear algebra
Coifman and Wickerhauser (1992)	Best Orthogonal Basis	- Shannon entropy
Mallat and Zhang (1993)	Matching Pursuit	-Synthesis direction approach
Chen and Donoho (1995a,b), Chen (1995)	Basis Pursuit	- Similar to MOF -A large-scale constrained opt.
Donoho (1995)	CART	- Shannon entropy

6.2. Matching Pursuit

Mallat and Zhang (1992) and Davis *et al.* (1994) developed a matching pursuit algorithm to represent a function in terms of wavelet packets using the wavelet packet functions as a dictionary of atomic waveforms. The algorithm searches the data for matches to the dictionary of waveforms up to a specified number of waveforms. The gauge for matching is the highest correlation between the signal and the waveform.

The matching pursuit algorithm (Mallat and Zhang, 1992; Davis *et al.*, 1994) matches a signal to a dictionary of wavelet packets. The signal can then be broken into multi-resolution components as in the discrete wavelet transform. The method for transient analysis begins with transforming the signal into the wavelet domain, using either the discrete wavelet transform or the matching pursuit algorithm, and thresholding by retaining only those components with the largest wavelet coefficient magnitudes.

Mallat and Zhang (1993) presented a greedy algorithm for the selection for the selection of the best matching pursuit decomposition of a signal into time-frequency packets from a large dictionary of such packet waveforms.

Matching pursuit (Mallat and Zhang, 1993) is an intuitively appealing approach to decomposing a data series/function. The basic approach is simple.

When the Fast Wavelet Transform extracts an approximation from the original signal, Matching Pursuit starts with a zero vector and builds up an approximation to the signal. This approach encourages a sparser representation of the signal.

Where the Fast Wavelet Transform (FWT) starts with the original and extracts an approximation from it, Matching Pursuit starts with a zero vector and builds up an approximation to the signal. This approach encourages a sparser representation of the signal.

We determine which atom to select by ranking them according to the magnitude of each atom's inner product with the current residual. This method is a sensible approach if we look at it intuitively because the atom is always large where the magnitude of the signal is large and so the magnitude of the inner product will be large.

6.3. Method of Frames

The Method of Frames (MOF) allows us to approach the approximation problem from the direction of synthesis rather than analysis (Daubechies, 1988). In order to find a representation of a signal as synthesized from a dictionary of atoms, we must find a vector of coefficients, α that satisfies

$$\phi\alpha = s \qquad (5)$$

Computational experience shows that the coefficient vector obtained from the MOF solution provides acceptable starting values for our BP decomposition.

6.4. Basis Pursuit

Basis Pursuit (BP) finds signal representations in overcomplete dictionaries by convex optimization (Chen, 1995; Chen and David, 1995a; 1995b). With Basis Pursuit (BP), we are attempting to find the solution of the synthesis problem with minimal l^1 norm.

We avoid the myopia problem of Matching Pursuit (MP), because we optimize over all coefficients at once with BP rather than one at a time with MP. Since we are looking at all coefficients at once in a sparse framework, we are able to superresolve with BP. This accounts for another of our desirable properties.

But, there is one important disadvantage to BP as compared to MOF. BP is much more computationally intensive. MOF involves solving a straight-forward linear algebra calculation, but BP requires the solution of a large-scale constrained optimization problem.

6.5. Near-Best Basis

Search algorithms for finding signal decomposition called near-best bases using decision criteria called non-additive information costs have recently been proposed by Taswell (1994) for selecting bases in wavelet packet transforms represented as binary trees.

Once the best basis with the minimum entropy has been selected, one needs to formulate the criterion for the wavelet transform coefficients selection.

7. Data Compression

The compression features of a given wavelet basis are primarily linked to the relative scareness of the wavelet domain representation for the signal. The notion behind compression is based on the concept that the regular signal component can be accurately approximated using the following elements: a small number of approximation coefficients (at a suitably chosen level) and some of the detail coefficients.

8. Research Model Architecture

Our study is to analyze wavelet thresholding or filtering methods for extracting optimal multi-signal decomposed series (i.e. highpass and lowpass filters) as a key input variable fitting a neural network based forecasting model specially under chaotic financial markets (Figure 5).

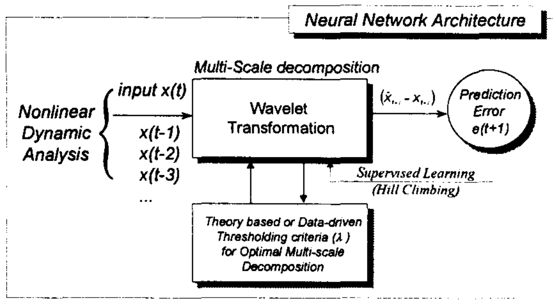


Figure 5. Integration Framework of Wavelet Transformation and Neural Networks

9. Neural Networks

For time series predictions, the most popularly used neural networks are clearly time delay neural networks (TDNN; Weigend, Huberman, and Rumelhart, 1990) and recurrent neural networks (RNN; Elman, 1990). The time delay neural networks can be analyzed by using standard methods and more the results of such analysis can be applied for time series predictions directly, but they may not be sufficient to characterize the patterns of highly dynamic time series. On the other hand, the recurrent neural networks are suited for applications that refer to the patterns of genuinely time dependent inputs such as time series predictions due to their dynamic feature.

While in the dynamic context the recurrent neural networks can outperform the time delay neural networks, they occasionally are difficult to be trained optimally by a standard backpropagation algorithm due in part to the dependence of their network parameters (Kuan and Hornik, 1991).

In this study, The basic model we experiment with is Backpropagation neural network (BPN) models which have a parsimonious 4 input nodes, 4 hidden-nodes and 1 output node with single wavelet filter, i.e. highpass, lowpass, or bandpass filter within the network structure. The other model we experiment with is BPN models which have 8 input nodes, 8 hidden-nodes and 1 output node with multiple filters.

10. Experimental Results

In this section, we evaluate prior methodology about wavelet thresholding using a case of the daily Korean Won / U.S. Dollar exchange rates are transformed to the returns using the logarithm and through standardization from January 10, 1990 to June 25, 1997. 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, 1990 to August 4, 1995, while the testing phase ran from August 7, 1995 to June 25, 1997. Further more, using a scaling function equal to 2, the 2048-point wavelet transform automatically produces ($\log_2 2048$) -1 or 10 separate filters or band 10 data because the data was symmetrically extended from points 1025-2048.

We transform the daily returns into the decomposed series such as an approximation part and a detail part by Daubechies wavelet transform with order 4 for neural network forecasting models in our study.

By the transformation, we approximate a time series at different levels of resolution using multiresolution decomposition. The individual time series resulting from the decomposition, taken together, can provide a detailed picture of the underlying processes.

The threshold λ determines the number of non-zero robust residuals. Setting λ too big will result in leakage of outliers into the signal and setting λ too small will cause distortion of the signal. We set λ so that an average of 100*p% non-zero robust residuals remain after thresholding in the Gaussian case. The tuning parameter p is set to some small value (e.g., .01). A table for λ is obtained by simulation based on the Gaussian model. This value of λ is quite insensitive to the stochastic characteristics of the underlying signal.

In summary, we use thresholding strategies shown in Table 3 and then compare each other in forecasting performance using test samples. The results are shown in Table 5-7.

Table 3. Thresholding (Data Compression or Denoising) Methods by Entropy Types

Type	*Description*
Shannon	Non-normalized entropy involving the logarithm of the squared value of each signal sample
Threshold	The number of samples for which the absolute value of the signal exceeds a threshold λ.
Norm	The concentration in l^p norm with $1 \leq p < 2$
Log Energy	The logarithm of energy, defined as the sum over all samples
SURE(Stein's Unbiased Risk Estimate)	A threshold-based method in which the threshold equals: $\sqrt{2\log_e(n\log_2(n))}$ Where n is the number of samples in the signal.

In our experiments, low-pass and high-pass filters are both considered in the wavelet transform, and their complementary use provides signal analysis and synthesis. The finest scale provides the original data, $x_N=x$, and the approximation at scale m is x_m where $m = 2^0, 2^1, \ldots, 2^N$. The incremental detail added in going from x_m to x_{m+1}, the detail signal, is yielded by the wavelet transform.

Therefore, the original signal can be expressed as an additive combination of the wavelet coefficients, at the different resolution levels.

For additivity of the wavelet transform decomposition, we must of course consider the wavelet coefficients given by the convolution of data and low-pass filter. We can think of the successive convolution as something like a moving average of N increasingly distant points.

We first try to select the most efficient basis out of the given set of bases to represent a given signal. In other words, we intend to find a basis, in which some of the coefficients attain high values (i.e. the respective basis vectors represent relevant information), while the remaining ones show low values. In this way we wish to obtain the greatest possible differentiation within the set of coefficients (Refer to Figure 6).

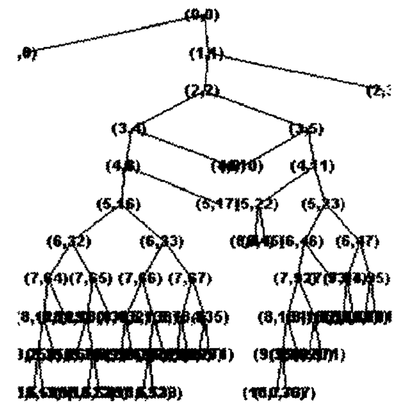


Figure 6. Best Orthogonal Basis of daily Korean Won / US Dollar returns data

Table 5, 6, and 7 compare thresholding performances from different preprocessing methods in forecasting models.

Firstly our experimental results (Table 5-7) show that wavelet transforms have proved to be very good methods for noise filtering and compressing data. This is doubtlessly due to the fact that varying resolution scales are treated, thus taking into account a range of superimposed phenomena.

Table 5 and 6 contain the comparison between hard and soft thresholding. Soft thresholding has a difference from hard thrsholding in the experimental results.

Table5-7 also show the results about the different performances among compression, denoising, best basis method, best level method, and cross-validation, etc.

But, except cross-validation method, any other method didn't outperform the others. That is, only cross-validation method significantly has the best performance among their techniques and the other methods have almost the same results.

Prior studies show that wavelet shrinkage denoising has been theoretically proven to be nearly optimal from the following perspective: spatial adaptation, estimation when local smoothness is unknown, and estimation when global smoothness is unknown. In effect, no alternative procedure can perform better without knowing a priori the smoothness class of the signal. But it is not appropriate to use a procedure that is theoretically optimal and general under most measures of local and global error for data applied to the domain specific problems.

Through our experimental results, we conclude that choosing the most appropriate shrinkage or thresholding procedure necessarily involves experiments to compare the performance of a wavelet thresholding method with any other methods under consideration. Specially, in our case study a cross-validatuion method is the best wavelet thresholding technique among all of them.

11. Conclusions and Future Research

Our research was motivated by a problem central in time series analysis: how to extract non-stationary signals which may have abrupt changes, such as level shifts, in the presence of impulsive outlier noise under short-term financial time series. A variety of techniques have been employed to deal with the problem. Our research indicates that a wavelet approach is basically an attractive alternative, offering a very fast algorithm with good theoretical properties and predictability in forecasting model design.

That reason is that the multiresolution property of the discrete wavelet transform enables the separation of transient, seasonal, and diurnal components of financial time series and also the dual time-frequency localization property of wavelets allows identification of features of transient events. This property has allowed for characterization of transients by location, duration and magnitude.

From our experiemntal results, wavelet shrinkage denoising has also been theoretically proven to be nearly optimal from the following perspective: spatial adaptation, estimation when local smoothness is unknown, and estimation when global smoothness is unknown (Taswell, 1998). In the future, the availability of these techniques will be promising more and more according to the domain features.

In summary, our experimental results show that cross-validation threshoding gives the best result in viewpoint of neural network based forecasting performance. That means that in general, the root mean square error measure of estimates as a data driven thresholding method is better than the other methods.

But, the data driven approach has some limitation as follows. That is, in fact, varying results can be obtained with different experimental conditions (signal classes, noise levels, sample sizes, wavelet transform parameters) and error measures, i.e. a cost function for global model optimization.

Ideally, the interplay between theory-based and experimental or data driven approach to implement an optimal wavelet thresholding should provide the best performance of a model according to the above experimental conditions.

Therefore, in the future we will suggest a new hybrid system of wavelet thresholding methodology and neural networks using genetic algorithms to overcome the limitation.

References

[1] Abramovich, F. and Y. Benjamini, "Adaptive Methods of Wavelet Coefficients as Multiple Hypothesis Testing Proceedure," 1994.

[2] 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, pp.215-239.

[3] Ariño, M. A. and B. Vidakovic, "On Wavelet Scalograms and Their Applications in Economic Time Series," *Discussion paper 95-21*, ISDS, Duke University, 1995.

[4] Aussem, A., J. Compbell and F. Murtagh, "Wavelet-Based Feature Extraction and Decomposition Strategies for Financial Forecasting," *Journal of Computational Intelligence in Finance*, March/April, 1998, pp.5-12.

[5] Breiman, L., J. Friedman, R. Olshen, and C.J. Stone, *Classification and Regression Trees*. Belmont, CA: Wadsworth, 1983.

[6] Chen, S.S., "Basis Pursuit," *Ph.D. Thesis*, Stanford University, Department of Statistics, Nov. 1995.

[7] Chen, S.S. and D.L. Donoho, "Atomic Decomposition by Basis Pursuit," *Tech. Report*, Stanford University, Statistics Department, 1995.

[8] Chen, S.S. and D.L. Donoho, "Basis Pursuit," *Tech. Report*, Stanford University, Statistics Department, 1995.

[9] Cody, M.A., "The Wavelet Packet Transform," *Dr. Dobb's Journal*, April, 1994, pp.44-54.

[10] Coifman, R.R., Y. Meyer, S. Quake, and M.V. Wickerhauser, "Wavelet Analysis and Signal Processing," in *Wavelets and Their Applications*, J.S. Byrnes, J.L. Byrnes, K.A. Hargreaves, and K. Berry, eds., Kluwer Academic: Boston, 1994, pp.363-380.

[11] Coifman, R.R., Y. Meyer, and M.V. Wickerhauser, *Progress in Wavelet Analysis and Applications*, in Y. Meyer and S. Roques (eds.), Editions Frontieres, France, 1993, pp.77-93.

[12] Coifman, R.R. and M.V. Wickerhauser, "Entropy Based Methods for Best Basis Selection," *IEEE Trans. on Inf. Theory* 38, No. 2, 1992, pp.719-746.

[13] Coifman, R.R. and M.V. Wickerhauser, "Wavelets and Adapted Waveform Analysis. A Toolkit for Signal Processing and Numerical Analysis," *In Processing of Symposia in Applied Mathematics* 47, 1993, pp.119-145.

[14] Davis, G., S. Mallat, and Z. Zhang, "Adaptive Time-Frequency Decompositions with Matching Pursuit," *Wavelet Applications, Proc. SPIE*, 1994, pp.402-409.

[15] Daubechies, I., "Time-Frequency Localization Operators: A Geometric Phase Space Approach," *IEEETIT* 34, 1988, pp.605-612.

[16] Daubechies, I., Ten Lectures on Wavelets, SIAM, Philadelphia, PA, 1992.

[17] Dohan, K. and P.H. Whifield, "Identification and Chararterization of Water Quality Transients Using Wavelet Analysis. I. Wavelet Analysis Methodology," *Wat. Sci. Tech.* Vol. 36, No. 5, 1997, pp.325-335.

[18] Donoho, D.L., "Denoising by Soft-Thresholding," *IEEE Transaction on Information Theory*, vol.41, 1995, pp.6130-6270.

[19] Donoho, D.L. and I.M. Johnstone, "Ideal Spatial Adaptation via Wavelet Shrinkage," *Bimetrika*, vol. 81, 1994, pp.425-455.

[20] Donoho, D.L., I.M. Johnstone, G. Kerkyacharian, and D. Picard, "Wavelet Shrinkage: Asymptopia?," *Journal of the Royal Statistical Society*, Series B, vol. 57,

1995, pp.301-369.
[21] Elman, J.L., "Finding Structure in Time," *Cognitive Science* 14, 1990, pp.179-211.
[22] Geva, A.B., "ScaleNet-Muliscale Neural-Network Architecture for Time Series Prediction," *IEEE Transaction on Neural Networks*, Vol. 9, No.5, Sep. 1998.
[23] Greenblatt, S.A., "Wavelets in Econometrics: An Application to Outlier Testing," *Computaional Economic Systems: Models, Methods, and Econometrics* (Dordrecht), Manfred Gilli (ed.), *Advances in Computaional Eonomics*, vol. 5, Kluwer Academic Publishers, 1995, pp.139-160.
[24] Høg, E., "Analyzing Continuous-time Long-memory Models with Wavelets," *International Workshop on Wavelets and Statistics*, Duke University, October 12-13, 1997.
[25] Katul, G. and B. Vidakovic, "The Partitioning of Attached and Detached Eddy Motion in the Atmospheric Surface Layer Using Lorentz Wavelet Filtering," *The Wavelet Digest* - Volume 4, Issue 2 (February 18, 1995) 1996.
[26] Kuan, C.M. and K. Hornik, "Convergence of Learning Algorithms with Constant Learning Rates," *IEEE Transactions on Neural Networks*, vol. 2, No.5, 1991, pp. 484-489.
[27] Lu, J. et al., "Noise Reduction with a multiscale edge representation and perceptual criteria," *in Proc. IEEE Int. Symp. Time-Frequency, Time-Scale Anal.* (Victoria, B.C., Canada), 1992, pp.555-558.
[28] Mallat, A., "Theory for Multiresolution Signal Decomposition: The Wavelet Representation," *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 1989, pp.674-693.
[29] Mallat, S.G. and Z. Zhang, Matching Pursuit with Time-Frequency Dictionaries, *Technical Report* 619, Courant Institute of Mathematical Sciences, 1993.
[30] Mallat, S.G. and S. Zhong, "Characterization of signal from multiscale edges," *IEEE Trans. Pattern. Anal. Machine Intell.*, vol. 10, 1992, pp.710-732.
[31] Mallat, S.G. and W.L. Hwang, "Singularity detection and processing with wavelets," *IEEE Trans. Inform. Theory, Special Issue on Wavelet Transforms and Multiresolution Analysis*, vol. 38, 1992, pp.617-643.
[32] Mittermayr, C.R., S.G. Nikolov, H. Hutter, and M. Grasserbauer, "Wavelet denoising of Gaussian peaks: a comparative study," *Chemometrics and Intelligent Laboratory Systems* 34, 1996, pp. 187-202.
[33] Morettin, P.A., "Wavelets in Statistics," *Third International Conference on Statistical Data Analysis Based on The L_1-Norm and Related Methods*, Neuchâtel, Switzerland, Aug. 11-15, 1997.
[34] Moulin, P., "Wavelet Thresholding Techniques for Power Spectrum Estimation," *IEEE Transaction on Signal Processing*, Vol. 42, No.11, November 1994, pp.3126-3136.
[35] Murtagh, F. and A. Aussem, "Using the Wavelet Transform for Multivariate Data Analysis and Time Series Forecasting," *Proc. OFCS'96*, Kobe, Springer-Verlag, 1996.
[36] Nason, G.P., "Choice of the threshold parameter in wavelet function estimation," In Wavelets and Statistics, Let
[37] Nason, G.P., "Wavelet Shrinkage using cross-validation," *Journal of the Royal Statistical Society*, Series B, vol. 58, 1996, pp.463-479.
[38] Pancham, S., "Evidence of the Multifractal Market Hypothesis Using Wavelet Transforms," *PhD Thesis*, Florida State Univ., 1994.
[39] Shin, T. and I. Han, "Integration of Joint Time-Frequency Filtering Methods and Neural Network Techniques for Foreign Exchange Return Forecasting," 한국전문가시스템학회 추계학술대회, 1998.
[40] Tak, B., "A New Method for Forecasting Stock Prices Using Artificial Neural Network and Wavelet Theory," *PhD Thesis*, Univ. of Pennsylvania, 1995.
[41] Taswell, C., "Near-Best Basis Selection Algorithms with Non-additive Information Cost Functions, " In Moeness G. Amin, (ed.) *Proceedings of the IEEE-SP International Symposium on Time-Frequency and Time-Scale Analysis*, Philadelphia, PA, Oct. 1994, pp.13-16.
[42] Taswell, C., "Near-best Basis Selection Algorithms with Non-additive Information Cost Functions," In *Proceedings of the IEEE-SP International Symposium on Time-Frequency and Time-Scale Analysis*, Philadelphia, PA, Oct. 1994, pp. 13-16.
[43] Taswell, C., The What, How, and Why of Wavelet Shrinkage Denoising, *Technical Report* CT-1998-09, Computational Toolsmiths, Stanford, 1998.
[44] Vidakovic, B., "Nonlinear Wavelet Shrinkage With Bayes Rules and Bayes Factors," *Journal of the American Statistical Association*, Volume 93, Number 441, March 1998, pp. 173-179.
[45] Walczak, B. and D.L. Massart, "Noise suppression and signal compression using the wavelet packet transform," *Chemometrics and Intelligent Laboratory Systems* 36, 1997, pp.81-94.
[46] Weigend, A.S., B.A. Huberman, and D.E. Rumelhart, "Predicting the Future: A Connectionist Approach," *Intl. J. Neur. Sys.* 1, 1990, pp.193-209.

Table 4. Prior Case Studies Using Wavelet Transform Techniques Applied to Financial Markets

Author(Year)	*Purpose*	*Data*	*Basis function*	*Methodology*	*Results*
Pancham (1994)	Test the multi-fractal market hypothesis	Monthly, weekly, daily Index	-	-	Accepted the multi-fractal market hypo.
Cody, M.(1994)	Present the concept of wavelets and the WT methods	General financial market data	DWT, WPT	Multi-scale linear prediction system	Suggested possible applications of the DWT to financial market analysis
Tak (1995)	Forecasting univariate time series	Standard & Poor's 500 index	Mexican-hat wavelet	ARIMA, detrending and AR, random walk, ANN	Outperformed than original data
Greenblatt (1996)	Analysis for structure in financial data	Foreign exchange rates	Coif-1,Coif-5	Best orthogonal basis, Matching pursuit, Method of frames, Basis pursuit	Found structure in financial data
McCabe and Weigend (1996)	Determine at which time-scale the series is most predictable	DM/US Dollar	Haar wavelet	Predictive linear models for multiresolution analysis	Rarely better than predicting the mean of the process
Høg (1996)	Estimate the fractional differencing parameter in Fractional Brownian Motion models for interest rate having the term structure	Monthly US 5-year yields on pure discount bonds (1965.11-1987.02)	Haar wavelet	ARFIMA(0,d+1,0) where H =d+1/2	$\tilde{d}$ = 0.900 95% confidence interval for d = [0.8711, 0.9289]
Høg (1997)	Analyze non-stationary but possibly mean-reverting processes	US interest rate	Haar wavelet	ARFIMA	Showed mean reversion of US interest rate
Aussem *et al.* (1998)	Predict the trend-up or down - 5 days ahead	S&P 500 closing prices	Ã trous wavelet	Dynamic recurrent NN & 1 nearest neighbors	86% correct prediction of the trend

Table 5. A Discrete Wavelet Transform Thresholding Performance Using Test Samples

Threshold Strategy&Techniques	*Entropy Type*	*Global Threshold*	*Network Structure*	*RMSE*
-	-	-	Random Walks	2.939007
-	-	-	BPN(4-4-1)[c]	1.754525
Cross-validation (HP&LP)[a]			BPN(8-8-1)	**1.676247**
Hard-Compression	-	0.5905	BPN(5-5-1)	1.768216
Soft-Denoising (LP)[b]	Fixed Form	-	BPN(4-4-1)	1.767864
	Rigorous SURE			
	Heuristic SURE			
	Minimax			
Soft-Denoising (HP&LP)	Fixed Form	-	BPN(8-8-1)	**1.751537**
	Rigorous SURE			
	Heuristic SURE			
	Minimax			
Hard-Denoising (LP)	Fixed Form	-	BPN(4-4-1)	1.766579
	Rigorous SURE			
	Heuristic SURE			
	Minimax			
Hard-Denoising (HP&LP)	Fixed Form	-	BPN(8-8-1)	1.754131
	Rigorous SURE			
	Heuristic SURE			
	Minimax			

a: Highpass + Lowpass filters, b: Lowpass filter,
c: BPN(I-H-O) = Backpropagation NN(I: # of Input nodes; H: # of Hidden nodes; O: # of output nodes).

Table 6. Wavelet Packet Transform Thresholding Performance Using Test Samples

Methods	*Entropy Type*	*Global Threshold*	*Network Structure*	*RMSE*
-	-	-	BPN(4-4-1)	1.754525
Hard-Compression (LP)	Shannon	5.717	BPN(4-4-1)	1.774456
	Threshold			
	Norm			
	Log Energy			
	SURE(Stein's Unbiased Risk Estimate)			
Hard-Compression (LP&HP)	Shannon	5.717	BPN(8-8-1)	1.759434
	Threshold			
	Norm			
	Log Energy			
	SURE(Stein's Unbiased Risk Estimate)			
Soft-Denoise (LP)	Shannon	4.336	BPN(4-4-1)	1.774456
	Threshold			
	Norm			
	Log Energy			
	SURE(Stein's Unbiased Risk Estimate)			
Soft-Denoise (LP&HP)	Shannon	4.336	BPN(8-8-1)	**1.759434**
	Threshold			
	Norm			
	Log Energy			
	SURE(Stein's Unbiased Risk Estimate)			

Table 7. The Model Performance Comparison Between Best Basis Selection and Best Level Techniques Using Test samples

Criteria	*Contents*	*Filter Types*	*BPN(#-#-1)*	*RMSE*
Best Basis	Coifman-Wickerhauser Best-Basis algorithm	LP	(4-4-1)	1.764243
		LP&HP	(8-8-1)	1.74329
Best Level	-	LP	(4-4-1)	1.767424