The Discrete Wavelet Transform with Financial Time Series Applications*

Tommi Vuorenmaa†

09.04.2003

*This work was presented at the Seminar on Learning Systems (at the Rolf Nevanlinna Institute).

†RUESG / University of Helsinki. For corrections and suggestions, please contact me via tvuorenm@cc.helsinki.fi. http://www.valt.helsinki.fi/staff/tvuorenm/

Part I

Theory

1 Daubechies wavelets

This section is mainly based on [3].

Scaling functions $\varphi(t)$ and the corresponding *wavelets* $\psi(t)$ are defined by the following *dilation equations:*

$$\begin{aligned} \varphi(t) &= 2^{1/2} \sum_k g_k \varphi(2t-k), \\ \psi(t) &= 2^{1/2} \sum_k h_k \varphi(2t-k), \end{aligned}$$

where $h_k = (-1)^{k+1} \overline{g}_{1-k}$.

The scaling function acts as a low-pass filter: convolving it with data gives us an approximation (the *scaling coefficients*) of the original series, except with high-frequency detail filtered out.

On the other hand, the corresponding wavelet acts as a high-pass filter: convolving gives us only the detailed information (the *differencing* or *wavelet coefficients*).

Fourier transform of the first dilation equation gives the *filter equality*

$$\widehat{\varphi}(\omega) = m_0(\omega/2)\widehat{\varphi}(\omega/2),$$

where $m_0(\omega)$ is a trigonometric polynomial

$$m_0(\omega) = 2^{-1/2} \sum_k g_k e^{-ik\omega}$$

In general, there are no explicit time-domain formulae for the Daubechies class of wavelet filters.

Daubechies (1988) showed that [3, p. 61]

$$|m_0(\omega)|^2 = c_N \int_\omega^\pi \sin^{2N-1} x dx, \tag{1}$$

where the constant c_N is chosen so that $m_0(0) = 1$. For such functions $m_0(\omega)$ one can tabulate the scaling filter coefficients g_l.

Definition 1.1 *Wavelets constructed with the use of functions $m_0(\omega)$ satisfying (1) are called Daubechies wavelets.*

Table 4.1 (taken from [1]) provides selected Daubechies wavelet filters.

Remark 1.1 *Daubechies first considered the so-called extremal phase wavelets, usually denoted $D(L)$, where L is the length of the filter.*

Remark 1.2 *Haar wavelet can be seen as a length* $L = 2$ *Daubechies* $D(2)$ *wavelet. Then the scaling filter coefficients are* $g_0 = g_1 = 2^{-1/2}$*, or equivalently, the wavelet filter coefficients are* $h_0 = 2^{-1/2}$ *and* $h_1 = -2^{-1/2}$.

Remark 1.3 *Haar wavelet is the only symmetric compactly supported orthonormal wavelet.*

A closely related class of wavelets are called *least asymmetric* wavelets, denoted by $LA(L)$.

Table 4.2 (taken from [1]) provides selected $LA(L)$ wavelet filters.

Figure 4.10 (taken from [1]) shows wavelet filter coefficients for lengths $L = 2, 4, 8$.

Recall that a function $\psi(t)$ with P vanishing moments satisfies $\int t^p \psi(t) dt = 0$, where $p = 0, 1, ..., P - 1$.

The number of vanishing moments for Daubechies wavelets is half the filter length. This implies that longer wavelet filters may produce stationary wavelet coefficient vectors from "higher degree" nonstationary stochastic processes.

Remark 1.4 $LA(8)$ *is a much better approximation to an ideal band-pass filter than the Haar (less leakage).*

2 Discrete wavelet transform

This section borrows heavily from [1].

Let $\mathbf{x}$ be a dyadic length ($N = 2^J$) vector of observations. The length N vector of discrete wavelet coefficients $\mathbf{w}$ is obtained via

$$\mathbf{w} = \mathcal{W}\mathbf{x},$$

where $\mathcal{W}$ is an $N \times N$ orthonormal matrix defining the *discrete wavelet transform* (DWT).

The matrix $\mathcal{W}$ is composed of the wavelet and scaling filter coefficients arranged on a row-by-row basis. The structure may be seen through the submatrices $\mathcal{W}_1, ..., \mathcal{W}_J$ and $\mathcal{V}_J$ in the following way:

$$\mathcal{W} = \left[\begin{array}{cccccc} \mathcal{W}_1 & \mathcal{W}_2 & ... & \mathcal{W}_J & \mathcal{V}_J \end{array} \right]^T,$$

where $\mathcal{W}_j$ is a $N/2^j \times N$ matrix and $\mathcal{V}_J$ is a $N/2^J \times N$ matrix.*

The vector of wavelet coefficients may be organized into $J+1$ vectors:

$$\mathbf{w} = \left[\begin{array}{ccccc} \mathbf{w}_1 & \mathbf{w}_2 & ... & \mathbf{w}_J & \mathbf{v}_J \end{array} \right]^T,$$

where $\mathbf{w}_j$ is a length $N/2^j$ vector of wavelet coefficients associated with changes on a scale of length

*In the next subsection, however, the order of the matrices is reversed.

$\lambda_j = 2^{j-1}$, and $\mathbf{v}_J$ is a length $N/2^J$ vector of scaling coefficients associated with averages on a scale of length $2^J = 2\lambda_J$.

DWT is an energy (variance) preserving transform:

$$\|\mathbf{w}\|^2 = \sum_{j=1}^{J} \sum_{t=0}^{N/2^j-1} w_{j,t}^2 + v_{J,0}^2 = \sum_{t=0}^{N-1} x_t^2 = \|\mathbf{x}\|^2 .$$

This so because

$$\begin{aligned} \|\mathbf{x}\|^2 &= \mathbf{x}^T\mathbf{x} = (\mathcal{W}\mathbf{w})^T\mathcal{W}\mathbf{w} \\ &= \mathbf{w}^T\mathcal{W}^T\mathcal{W}\mathbf{w} = \mathbf{w}^T\mathbf{w} = \|\mathbf{w}\|^2 . \end{aligned}$$

Given the structure of the wavelet coefficients, $\|\mathbf{x}\|^2$ is decomposed on a scale-by-scale basis via

$$\|\mathbf{x}\|^2 = \sum_{j=1}^{J} \left\|\mathbf{w}_j\right\|^2 + \|\mathbf{v}_J\|^2 ,$$

where $\left\|\mathbf{w}_j\right\|^2$ is the energy (proportional to variance) of $\mathbf{x}$ due to changes at scale λ_j and $\|\mathbf{v}_J\|^2$ is the information due to changes at scales λ_J and higher.

2.1 DWT via a "lifting" technique

This subsection is based on [4]. The procedure known as lifting was introduced by Sweldens (1996, 1997).

Consider a signal $\mathbf{x}$ of length $N = 2^3 = 8$ (i.e. $J = 3$). This signal may be arbitrary, but for now let it be of special kind:

$$\mathbf{x} = (1, 0, 0, ..., 0).$$

To begin, we take the entries in pairs and compute the mean and the difference between the first member of the pair and the computed mean. We then save the results and repeat the procedure two more times.

We end up with the following table(s):

1	0	0	0	0	0	0	0
$\frac{1}{2}$	0	0	0	$\frac{1}{2}$	$\mathbf{0}$	$\mathbf{0}$	$\mathbf{0}$
$\frac{1}{4}$	0	$\frac{1}{4}$	$\mathbf{0}$	$\frac{1}{2}$	$\mathbf{0}$	$\mathbf{0}$	$\mathbf{0}$
$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{4}$	$\mathbf{0}$	$\frac{1}{2}$	$\mathbf{0}$	$\mathbf{0}$	$\mathbf{0}$

.

Similarly

0	1	0	0	0	0	0	0
$\frac{1}{2}$	0	0	0	$-\frac{1}{2}$	$\mathbf{0}$	$\mathbf{0}$	$\mathbf{0}$
$\frac{1}{4}$	0	$\frac{1}{4}$	$\mathbf{0}$	$-\frac{1}{2}$	$\mathbf{0}$	$\mathbf{0}$	$\mathbf{0}$
$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{4}$	$\mathbf{0}$	$-\frac{1}{2}$	$\mathbf{0}$	$\mathbf{0}$	$\mathbf{0}$

;

and so forth.

This is an example of a simple lifting scheme. The end-result (the last line) is the DWT of the signal (the first line).

From the lifting scheme we infer that the high-pass filter differences and the low-pass filter averages.

The last four ($4 = 8/2^1$) entries are the wavelet coefficients belonging to the first scale.

The next two ($2 = 8/2^2$) belong to the second scale.

Only one ($1 = 8/2^3$) wavelet coefficient belongs to the third scale.

The first entry on the last line is the scaling coefficient.

Arranging the last lines of all the eight tables into a matrix gives $\mathcal{W}_a^{(3)}$:

$$\mathcal{W}_a^{(3)} = \begin{bmatrix} \frac{1}{8} & \frac{1}{8} & \frac{1}{8} & \frac{1}{8} & \frac{1}{8} & \frac{1}{8} & \frac{1}{8} & \frac{1}{8} \\ \frac{1}{8} & \frac{1}{8} & \frac{1}{8} & \frac{1}{8} & -\frac{1}{8} & -\frac{1}{8} & -\frac{1}{8} & -\frac{1}{8} \\ \frac{1}{4} & \frac{1}{4} & -\frac{1}{4} & -\frac{1}{4} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{4} & \frac{1}{4} & -\frac{1}{4} & -\frac{1}{4} \\ \frac{1}{2} & -\frac{1}{2} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{2} & -\frac{1}{2} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{2} & -\frac{1}{2} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{2} & -\frac{1}{2} \end{bmatrix}.$$

We have essentially constructed a matrix that does the so-called *direct* Haar transform for any length 8 signal.

Remark 2.1 *The subscript "a" in the matrix* $\mathcal{W}_a^{(3)}$ *refers to "analysis", while the power* (3) *refers to the level of scale* $J = 3$.

Example 2.1 *Calculate the direct DWT such that* $\mathbf{w} = \mathcal{W}_a^{(3)}\mathbf{x}$*, where* $\mathbf{x} = (1, 0, 0, ..., 0)$*, and you will (clearly!) end up with* $\mathbf{w} = (\frac{1}{8}, \frac{1}{8}, \frac{1}{4}, 0, \frac{1}{2}, 0, 0, 0)$.

Remark 2.2 *Normalization factor is omitted here for purposes of clarity.*

Obtaining the first line of the table is done simply by reversing the steps done above.

Starting with $\mathbf{w} = (1, 0, 0, ..., 0)$, we therefore reconstruct the following table(s):

1	1	1	1	1	1	1	1
1	1	1	1	0	0	0	0
1	1	0	0	0	0	0	0
1	0	0	0	0	0	0	0

.

Similarly

1	1	1	1	−1	−1	−1	−1
1	1	−1	−1	0	0	0	0
1	−1	0	0	0	0	0	0
0	1	0	0	0	0	0	0

;

and so forth.

Arranging the first lines of the eight tables into a matrix we end up with $\mathcal{W}_s^{(3)}$:

$$\mathcal{W}_s^{(3)} = \begin{bmatrix} 1 & 1 & 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & -1 & 0 & 0 & 0 \\ 1 & 1 & -1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 1 & -1 & 0 & 0 & -1 & 0 & 0 \\ 1 & -1 & 0 & 1 & 0 & 0 & 1 & 0 \\ 1 & -1 & 0 & 1 & 0 & 0 & -1 & 0 \\ 1 & -1 & 0 & -1 & 0 & 0 & 0 & 1 \\ 1 & -1 & 0 & -1 & 0 & 0 & 0 & -1 \end{bmatrix}$$

It should be no big surprise this time that applying the above matrix to any length 8 signal performs the *inverse* Haar transform.

Remark 2.3 *The subscript "s" refers to "synthesis".*

Example 2.2 *Calculate the inverse DWT such that* $\mathbf{x} = \mathcal{W}_s^{(3)}\mathbf{w}$, *where* $\mathbf{w} = (1, 0, 0, ..., 0)$, *and you will end up with* $\mathbf{x} = (1, 1, ..., 1)$.

Remark 2.4 *The linear algebra formulation of perfect recontruction is* $\mathcal{W}_s^{(3)} \cdot \mathcal{W}_a^{(3)} = \mathcal{I}$, *and* $\mathcal{W}_a^{(3)} \cdot \mathcal{W}_s^{(3)} = \mathcal{I}$, *where* $\mathcal{I}$ *denotes the* 8×8 *identity matrix.*

Remark 2.5 *In practice, the rows of the matrix* $\mathcal{W}$ *are not explicitly constructed. Instead the DWT is implemented via a pyramid algorithm, see Mallat (1989).*

2.2 Partial DWT

This subsection is based [1].

When the length of the time series is dyadic, we do not have to implement the DWT down to level $J =$

$\log_2(N)$. A partial DWT may be performed instead that terminates at a level $J_p < J$.

The structure of the orthonormal matrix $\mathcal{W}$ is similar to above:

$$\mathcal{W} = \left[\begin{array}{ccccc} \mathcal{W}_1 & \mathcal{W}_2 & \dots & \mathcal{W}_{J_p} & \mathcal{V}_{J_p} \end{array} \right]^T,$$

except that the matrix of scaling filter coefficients will be a $N/2^{J_p} \times N$ matrix of circularly shifted scaling coefficients vectors.

3 Maximum overlap DWT

This section is based on [1]. A throughout discussion can be found in Percival and Walden (2000).

Like orthogonal wavelets, also non-orthogonal wavelets provide a basis for the space into which the function is to be projected. Biorthogonal wavelets are examples of such non-orthogonal wavelets.

They are not the only ones, however. In fact, *maximum overlap discrete wavelet transforms* (MODWT) are more commonly encountered in empirical work.†

Unlike DWT, MODWT is *not* computed by subsampling the filtered output.

The MODWT gives up orthogonality in order to gain features that the DWT does not posses.

A consequency of this is that the wavelet and scaling coefficients must be rescaled in order to retain the variance preserving property of the DWT.

The following properties are important in distinguishing the MODWT from the DWT:

†MODWT is also called *"stationary DWT"*, *"translation-invariant DWT"* and *"time-invariant DWT"*.

1. The MODWT can handle any sample size N, while the J_pth order partial DWT restricts the sample size to a multiple of 2^{J_p}.

2. The detail and smooth coefficients of a MODWT multiresolution analysis are associated with zero phase filters.

3. The MODWT is invariant to circularly shifting the original time series. This does not hold for the DWT.

4. The MODWT wavelet variance estimator is asymptotically more efficient than the same estimator based on the DWT.

To be a bit more precise:

Let $\mathbf{x}$ be a length N (not necessarily dyadic) vector of observations. The length $(J+1)N$ vector of MODWT coefficients $\widetilde{\mathbf{w}}$ is obtained via

$$\widetilde{\mathbf{w}} = \widetilde{\mathcal{W}}\mathbf{x},$$

where $\widetilde{\mathcal{W}}$ is a $(J+1)N \times N$ (non-orthogonal) matrix defining the MODWT.

Similarly to the matrix $\mathcal{W}$ defining the DWT, the matrix $\widetilde{\mathcal{W}}$ is made of $J+1$ submatrices, this time each of them $N \times N$:

$$\widetilde{\mathcal{W}} = \left[\begin{array}{ccccc} \widetilde{\mathcal{W}}_1 & \widetilde{\mathcal{W}}_2 & ... & \widetilde{\mathcal{W}}_J & \widetilde{\mathcal{V}}_J \end{array} \right]^T.$$

Instead of using the wavelet and scaling filters, the MODWT utilizes the rescaled filters ($j = 1, ..., J$)

$$\widetilde{\mathbf{h}}_j = \mathbf{h}_j/2^j \text{ and } \widetilde{\mathbf{g}}_J = \mathbf{g}_J/2^J.$$

To construct the $N \times N$ dimensional submatrix $\widetilde{\mathcal{W}}_1$, we circularly shift the rescaled wavelet filter vector $\widetilde{\mathbf{h}}_1$ by integer units to the right so that

$$\widetilde{\mathcal{W}}_1 = \left[\begin{array}{ccccc} \widetilde{\mathbf{h}}_1^{(1)} & \widetilde{\mathbf{h}}_1^{(2)} & ... & \widetilde{\mathbf{h}}_1^{(N-1)} & \widetilde{\mathbf{h}}_1 \end{array} \right]^T.$$

Here for example $\widetilde{\mathbf{h}}_1^{(1)}$ is

$$\widetilde{\mathbf{h}}_1^{(1)} = \left[\begin{array}{ccccc} h_{1,0} & h_{1,N-1} & \ldots & h_{1,2} & h_{1,1} \end{array} \right]^T.$$

Submatrices $\widetilde{\mathcal{W}}_2, \ldots, \widetilde{\mathcal{W}}_J$ are formed similarly.

The vector of MODWT coefficients may be organized into $J+1$ vectors:

$$\widetilde{\mathbf{w}} = \left[\begin{array}{ccccc} \widetilde{\mathbf{w}}_1 & \widetilde{\mathbf{w}}_2 & \ldots & \widetilde{\mathbf{w}}_J & \widetilde{\mathbf{v}}_J \end{array} \right]^T,$$

where $\widetilde{\mathbf{w}}_j$ is a length $N/2^j$ vector of wavelet coefficients associated with changes on a scale of length $\lambda_j = 2^{j-1}$, and $\widetilde{\mathbf{v}}_J$ is a length $N/2^J$ vector of scaling coefficients associated with averages on a scale of length $2^J = 2\lambda_J$ (just as with the DWT).‡

For time series of dyadic length, the MODWT may be subsampled and rescaled to obtain DWT wavelet coefficients via

$$w_{j,t} = 2^{j/2}\widetilde{w}_{j,2^j(t+1)-1} \; (t = 0, \ldots, N/2^j - 1),$$

‡This is a quote from [1, p. 135], but in my opinion there must be an error: the outcome should be of length $(J+1)N$.

and DWT scaling coefficients via

$$v_{J,t} = 2^{J/2}\widetilde{v}_{J,2^J(t+1)-1} \; (t = 0, ..., N/2^J - 1).$$

Like DWT, also MODWT is an energy (variance) preserving transform. The total variance of a time series can be partitioned as follows:

$$\|\mathbf{x}\|^2 = \sum_{j=1}^{J} \left\|\widetilde{\mathbf{w}}_j\right\|^2 + \|\widetilde{\mathbf{v}}_J\|^2 .$$

Remark 3.1 *In practice, a pyramid algorithm (similar to that of the DWT) is utilized to compute the MODWT, see Percival and Mojfeld (1997).*

Remark 3.2 *The so-called discrete wavelet packet transform (DWPT) is a further generalization of the DWT.*

Remark 3.3 *Similarly MODWPT is a generalization of the MODWT.*

Part II

Applications

4 The nature of financial time series

Financial time series are of extremely complex nature. However, there exist some universal phenomena that are called the "stylized facts". Most importantly:

Stock price evolution is of discontinuous character (see Figure 1 and 2).

Jump sizes are larger and more frequent than a Gaussian hypothesis assumes.

Volatility tends to cluster; i.e. highly volatile and tranquil times alternate. Stationarity does not easily hold.

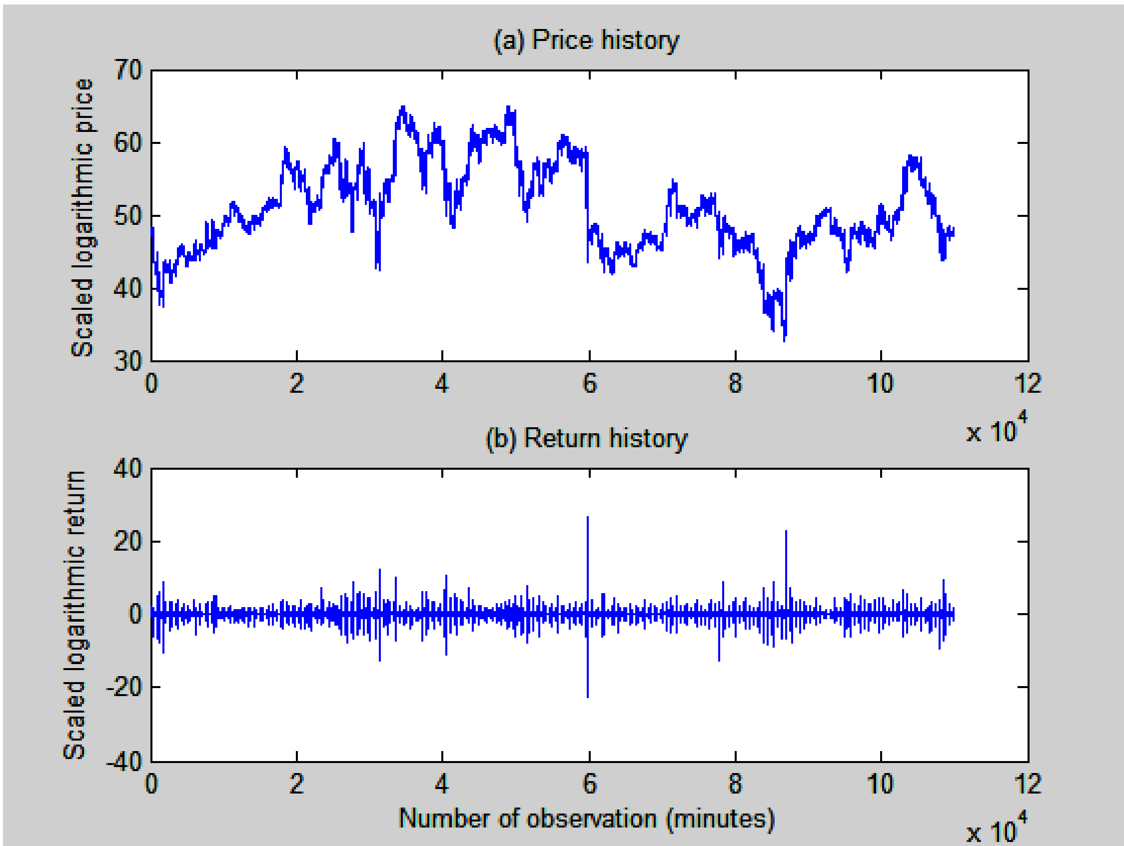

Figure 1: The (a) price and (b) return evolution of Nokia in year 2000, sampled every minute. Both variables are scaled logarithmic. Pay attention to the non-Gaussian behaviour and clustering of volatility.

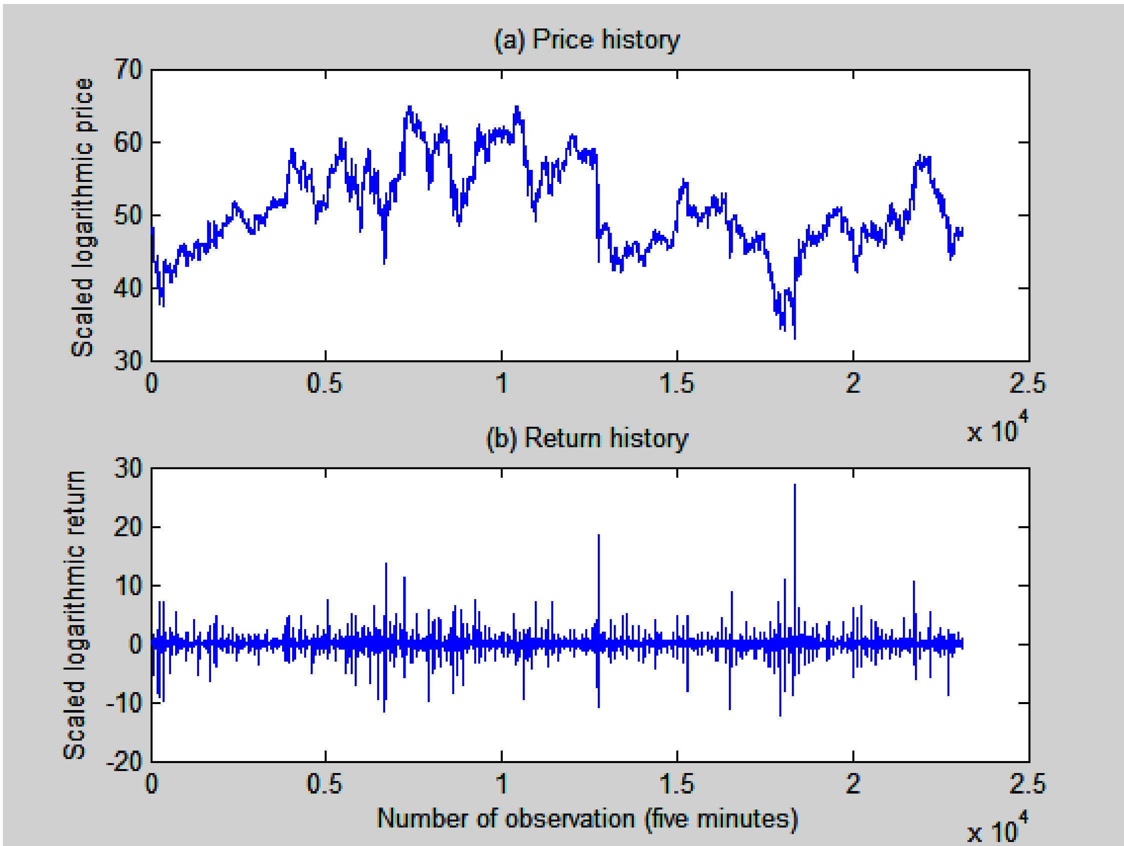

Figure 2: The (a) price and (b) return evolution of Nokia in year 2000, sampled every five minutes. Both variables are scaled logarithmic. Pay attention to the non-Gaussian behaviour and clustering of volatility.

5 Outlier testing

This section is based on [2].

The traditional view to outliers is to determine which observations come from the data generating distribution and which ones come from some contaminating distribution.

We define an outlier as an observation which appears to be "out of place". The following approach to test for an outlier is by Greenblatt (1994). His approach is based on order statistics.

While unordered observations can be assumed independent (atleast in an interval larger than a day)[§], observations ordered to ascending order of magnitude

[§]This was Greenblatt´s opinion, not mine; in my opinion daily returns can be assumed *uncorrelated*, but this does *not* necessarily imply independence.

can not. This causes the distribution theory become nontrivial. Greenblatt used Monte Carlo analysis to estimate the finite sample properties.

Greenblatt assumed that most, if not all, of the highest frequency component of the data is either noise or outliers. Thus outliers should be well captured by the smallest scale wavelet coefficients.

Let $\mathbf{w}_1 = (w_{1,0}, ..., w_{1,N/2-1})$ be the unit scale coefficient vector obtained by executing the DWT to a length N time series $\mathbf{x}$. The unit scale coefficients are associated with frequencies $1/4 \leq f \leq 1/2$ (i.e. any oscillation with a period length of 2 to 4 time units).

Haar wavelet is used because of its good local properties and because smoothess is now not needed.

We wish to find the coefficient that is the furthest away from the average. The test-statistic for testing for a single outlier therefore is

$$\mathcal{D}_2 = \frac{\max_t \left| w_{1,t} - \overline{w}_1 \right|}{\widehat{\sigma}_X(\lambda_1)},$$

where $\overline{w}_1$ is the sample mean of the $N/2$ wavelet coefficients and $\widehat{\sigma}_X(\lambda_1)$ is the unit scale sample wavelet standard deviation.

This statistic differs from the one used by David (1981) only slightly: instead of the original data we use a subset of wavelet coefficients.

Remark 5.1 *Using* $\overline{w}_1$ *is not necessary (that is, it can be set to zero) since the wavelet coefficients have mean zero as long as the spectrum of the underlying process does not vary in the interval* $1/4 \leq f \leq 1/2$.

Greenblatt first determined the 95% and 99% critical values by Monte Carlo simulation of 1,000,000 replications. Then an outlier was introduced per replication.

In Figure 3 (taken from [2]) one easily observes such an (negative) outlier at observation 347.

Figure 4 (taken from [2]) shows the same information but via highest-level wavelet coefficients.

He was able to show that the test proposed above was more powerful on *less obvious* outliers than the test performed on the data; see Tables 1 and 2 (taken from [2]).

In Table 1 and 2, the variable λ determines the magnitude of the outlier.

He also found that there was only a minor difference when an obvious outlier was present

6 Analysis of return series

This section is based on [1].

6.1 Using DWT with Haar and $LA(8)$

Consider the IBM daily (May 17, 1961 – November 2, 1962) return series $\mathbf{r}$ of length $N = 368$. This is the classic data set studied in Box and Jenkins (1976).

The length $N = 368$ is divisible by $2^4 = 16$ and therefore it is possible to perform a $J_p = 4$ *partial* DWT on it.

The first line of Figure 4.17 (taken from [1]) represent the IBM (logarithmic) returns $r_t = \log(p_{t+1}/p_t)$.

Notice the obvious increase in variance toward the latter third of the series.

Recall that the first scale of wavelet coefficients $\mathbf{w}_1$ are essentially looking at adjacent differences in the data.

Using the Haar wavelet filter shows that there is a group of rapidly fluctuating returns around observations 250 – 300.

The second scale coeffiecients do not show such a clear increase in magnitude.

From the third and fourth scale coefficients which are near zero, one infers that the return series do not exhibit low-frequency oscillations.

Using the $LA(8)$ wavelet filter gives results that have the same interpretation.

However, since the $LA(8)$ is a better approximation to an ideal band-pass filter than the Haar, the coefficients should now isolate features in a given frequency interval better than the Haar.

Remark 6.1 *The $LA(8)$ wavelet coefficient vectors are been circularly shifted in order to better align features in the wavelet coefficients with the original time series.*

6.1.1 Using MODWT with Haar and $LA(8)$

Consider the same IBM return series as above. Recall that $N = 368$, but that with MODWT we are not limited to decomposing a sample size of dyadic length.

The Haar and $LA(8)$ wavelet filters are used to perform a level $J = 4$ MODWT on the series.

Again the first line of Figure 4.23 (taken from [1]) shows the IBM return series.

The interpretation stays the same as above.

Remark 6.2 *There are N wavelet coefficients at each scale because the MODWT does not subsample after filtering.*

The first scale of wavelet coefficients $\widetilde{\mathbf{w}}_1$ contain the DWT coefficients (cf. Figure 4.17) scaled by $2^{-1/2}$, and also the DWT coefficients applied to $\mathbf{x}$ circularly shifted by one.

The MODWT coefficients are *correlated* and will appear *smoother* than the DWT coefficients.

The increased smoothness is most visible in lower frequencies.

Smoothness is furthermore increased when applying $LA(8)$. This is because the longer wavelet filter has induced significant amounts of correlation between adjacent coefficients.

Notice again (see Remark 6.1) that $LA(8)$ wavelet coefficient vectors are been circularly shifted.

7 Practical issues to consider

This section is based on [1].

7.1 Wavelet basis selection

Selecting a wavelet basis is important for two reasons:

1. The length of a discrete wavelet function determines how well it approximates an ideal band-pass filter. This in turn dictates how well the filter is able to isolate features to specific frequency intervals.

2. The wavelet basis function should be able to mimic data´s underlying features to better represent the information. For example, if the data is fairly smooth, then a longer filter such as $LA(8)$ should be used rather than the Haar.

When using the MODWT, the choice of wavelet function is not as vital as otherwise. This is because the MODWT adds correlation between adjacent wavelet coefficients.

7.2 Nondyadic length

One remedy is to "pad" the time series with zeros, sample mean, or the last value of the series.

Ogden (1997) has compared padding with zeros, repeating the last observation, interpolation, and numerical integration. No method proved to be clearly superior.

7.3 Boundary conditions

When applying DWT or MODWT to finite-length time series, one is faced by the crucial issue of wavelets affected by the boundary.

Because wavelet transform is based on filtering a time series, one has to use some kind of method for computing the remaining wavelet coefficients when one end of the vector is encountered.

One remedy is to assume the length N time series to be periodic and use the other end´s observations to finish the computations. This is applicable only for strongly seasonal time series.

A common technique is to "reflect" the time series about its last observation. Reflecting the series does not alter the sample mean nor the sample variance.

References

[1] Gençay *et al.* (2002)

[2] Greenblatt (1994)

[3] Härdle *et al.* (1998)

[4] Jensen and la Cour–Harbo (2001)