

Asymmetry of Information Flow Between Volatilities Across Time Scales

Ramazan Gençay* Faruk Selçuk† Brandon Whitcher‡

November 2002

Abstract

Conventional time series analysis, focusing exclusively on a time series at a given scale, lacks the ability to explain the nature of the data generating process. A process equation that successfully explains daily price changes, for example, is unable to characterize the nature of hourly price changes. On the other hand, statistical properties of monthly price changes are often not fully covered by a model based on daily price changes. In this paper, we simultaneously model regimes of volatilities at multiple time scales through wavelet-domain hidden Markov models.

We establish an important stylized property of volatility across different time scales. We call this property *asymmetric vertical dependence*. It is asymmetric in the sense that a low volatility state (regime) at a long time horizon is most likely followed by low volatility states at shorter time horizons. On the other hand, a high volatility state at long time horizons does not necessarily imply a high volatility state at shorter time horizons.

Our analysis provides evidence that volatility is a mixture of high and low volatility regimes, resulting in a distribution that is non-Gaussian. This result has important implications regarding the scaling behavior of volatility, and consequently, the calculation of risk at different time scales.

Keywords: Discrete wavelet transform, wavelet-domain hidden Markov trees, foreign exchange markets, stock markets, multiresolution analysis, scaling.

JEL No: G0, G1, C1

*Department of Economics, University of Windsor, 401 Sunset Avenue, Windsor, Ontario, N9B 3P4, Canada & HEC Genève, Uni Mail, CH-1211, Genève 4, Switzerland. Ramo Gençay gratefully acknowledges financial support from the Natural Sciences and Engineering Research Council of Canada, the Social Sciences and Humanities Research Council of Canada and the Swiss National Science Foundation under NCCR-FINRISK. We are grateful to Olsen Group (www.olsen.ch), Switzerland for providing the data, and Dr. Richard Baraniuk for providing some of the wavelet software used in this paper. E-mail: gencay@uwindsor.ca

†Bilkent University, Department of Economics, Bilkent, Ankara, 06533 Turkey and Colorado College, Colorado Springs, U.S.A. Faruk Selçuk gratefully acknowledges financial support from the Colorado College Social Science Executive Division Research and Development Fund.

‡National Center for Atmospheric Research, P.O. Box 3000, Boulder, CO 80307-3000, United States. Brandon Whitcher gratefully acknowledges the support provided by the National Center for Atmospheric Research (NCAR) Geophysical Statistics Project, sponsored by the National Science Foundation (NSF) under Grants DMS98-15344 and DMS93-12686.

1 Introduction

The fundamental properties of volatility dynamics are volatility clustering (conditional heteroscedasticity) and long memory (slowly decaying autocorrelation). Both properties might be labeled as *horizontal dependency* when viewing volatility in the time domain.¹ In this paper, we establish a third important stylized property of volatility from a time-frequency point of view – the asymmetric dependence of volatility across different time horizons. Specifically, low volatility at a long time horizon is most likely followed by low volatility at shorter time horizons. On the other hand, high volatility at long time horizons does not necessarily imply a high volatility at shorter time horizons. We call this property *asymmetric vertical dependence*.

The motivation behind the vertical dependence in volatility is the existence of traders with different time horizons. At the outer layer of the trading mechanism are the fundamentalist traders who trade on longer time horizons. At lower layers, there are short-term traders with a time horizon of a few days and day traders who may carry positions only overnight. At the next level down are the intraday traders who carry out trades only during the day but do not carry overnight positions. At the heart of trading mechanisms are the market makers operating at the shortest time horizon (highest frequency). Each of these types of traders may have their own trading tools consistent with their trading horizon and may possess a homogeneous appearance within their own class. Overall, it is the combination of these activities for all time scales that generates market prices. Therefore, market activity would not exhibit homogeneous behavior, but the underlying dynamics would be heterogeneous with each trading class at each time scale dynamically interacting across all trading classes at different time scales.² In such a heterogeneous market, a low-frequency shock to the system penetrates through all layers of the entire market reaching the market makers. The high-frequency shocks, however, would be short lived and may have no impact outside their boundaries.

Short-term traders constantly watch the market to re-evaluate their current positions and execute transactions at a high frequency. Long-term traders may look at the market only once a day or less frequently. A quick price increase followed by a quick decrease of the same size, for example, is a major event for an intraday trader but a nonevent for central banks and long-term investors.³ Long-term traders are interested

¹Clustering and long memory properties were first noted in Mandelbrot (1963) and Mandelbrot (1971). These findings remained dormant until Engle (1982) and Bollerslev (1986) proposed the ARCH and GARCH processes for volatility clustering. In the early 1990s, a comprehensive study of the long-memory properties of financial time series had started.

²The term “time scale” may be viewed as a “resolution”. At high time scales (low frequencies, long term) there is a coarse resolution of a time series while at low time scales (high frequency, short term), there exists a high resolution. Moving from low time scales to high time scales (from short term to long term) leads to a more coarse characterization of the time series due to averaging.

³Small, short-term price moves may sometimes have a certain influence on the *timing* of long-term traders’ transactions but not on their investment decisions.

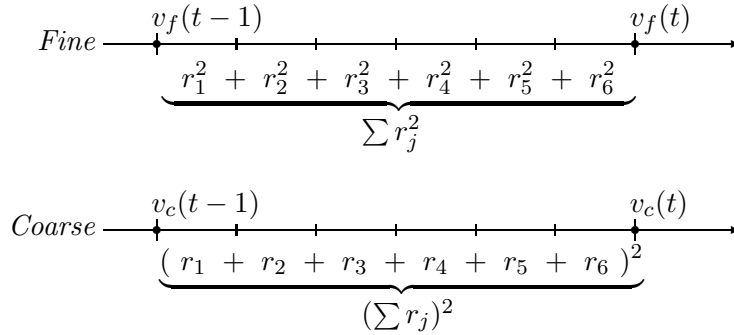


Figure 1: The coarse volatility, $v_c(t)$, captures the view and actions of long-term traders while the fine volatility, $v_f(t)$, captures the view and actions of short-term traders. The two volatilities are calculated at the same time points where returns (r_j) are measured and are synchronized.

only in large price movements and these normally happen only over long time intervals. Therefore, long-term traders with open positions have no need to watch the market at every instance.⁴ In other words, they judge the market, its prices, and also its volatility with a coarse time grid. A coarse time grid reflects the view of a long-term trader and a fine time grid that of a short-term trader.

For exploring the behavior of volatilities of different time resolution, Dacorogna *et al.* (2001) defined two types of volatility, the “coarse” volatility, v_c , and the “fine” volatility, v_f , as illustrated in Figure 1. The coarse volatility, $v_c(t)$, captures the view and actions of long-term traders while the fine volatility, $v_f(t)$, captures the view and actions of short-term traders.⁵ It has been shown in Müller *et al.* (1997) and Dacorogna *et al.* (2001) that there is an asymmetry where the coarse volatility predicts fine volatility better than the other way around.⁶

One of the goals of this paper is to investigate the propagation properties of this heterogeneity-driven asymmetry by studying the statistical properties of the flow of information from low-to-high frequency scales.⁷ Low frequency scales (fundamental-

⁴They have other means to limit the risk of rare large price movements by stop-loss limits or options.

⁵The two volatilities are calculated at the same time points where returns (r_j) are measured and are synchronized.

⁶The HARCH model of Müller *et al.* (1997) belongs to the ARCH family but differs from ARCH-type processes in a unique way of considering the volatilities of returns measured over different interval sizes. The HARCH model has the ability to capture the asymmetry in the interaction between volatilities measured at different frequencies such that a coarsely defined volatility predicts a fine volatility better than the way around.

⁷The flow of dependence, from lower resolution (low frequency content) to higher resolution (higher frequency content) can be relaxed such that an analysis from higher frequency to low frequency content can be allowed. Durand and Gonçalves (2001) comment that the directions of the directed acyclic graph (DAG) for the model of Crouse *et al.* (1998) is not necessary according to a paper by Smyth *et al.* (1996). In this paper, one can drop all directions in the graph and the conditional independence statements would still be valid.

ist type traders) are associated with traders who trade infrequently. Therefore, the framework that we study focuses on the impact of the actions of the long-term traders on short-term traders who trade more frequently. Once the regime structure (state) is identified from low-to-high frequency, this has implications for the flow of information across time scales. In particular, a high volatility regime persists longer at longer (lower frequency) trading horizons relative to short (high frequency) horizons. Alternatively, the duration of regimes tends to be longer for low-frequency trading horizons whereas high-frequency horizons have short-lived regime durations with frequent regime switching. This is not surprising since the impact of a change in long term dynamics would be short lived at higher frequencies.

Indeed, our findings indicate that a low volatility at a low frequency implies a low volatility at higher frequencies. For example, if a low volatility is observed at a weekly scale, it is more likely that there is also a low volatility at a one day scale. However, a high volatility at a low frequency does not necessarily imply a high volatility at higher frequencies. This is because the market “calms down” at higher frequencies much earlier than it does at lower frequencies.

Our modeling framework is based on wavelet-domain hidden Markov models (HMMs). The wavelet HMMs are distinct from traditional HMMs already used in time series analysis.⁸ Traditional HMMs capture the temporal dependence within a given time scale whereas wavelet HMMs capture dependencies in the two-dimensional time-frequency plane. In our analysis, we classify high-frequency data into time horizons (scales) that are consistent with the time scales in which traders operate. Each time scale is characterized with a two-state regime of high and low volatility. By connecting the state variables vertically across scales, we obtain a graph with tree-structured dependencies between state variables across different time scales. An implication of our findings is that composition of states across adjacent scales varies in time. Hence, simple aggregation of a daily volatility to obtain a monthly volatility, for instance, may not necessarily follow linear aggregation but may involve nonlinearities through state switching.

A simple way to think about wavelet multiscale analysis is the following example: The day ends and one makes the analysis of the day at various time intervals. For instance, a trader may argue that the day overall was quiet with minimal volatility except that there was high volatility within a 10-min window in the morning trading around 10:00 a.m. Such a statement requires the trader to observe the entire day and make references to specific time intervals in that trading day. Figure 2 illustrates this point with an example from the New York Stock Exchange. On January 3, 2001, the Dow Jones Industrial Average (DJIA) increased from 10,646 (previous day closing

⁸In the economics and finance literature, the persistence of mean and volatility dynamics and nonlinearities through regime switching at a given data frequency (horizontal persistence) have been examined extensively within the context of Markov switching models. The introduction of Markov switching models to the economics and finance literature is due to Hamilton (1989). Maheu and McCurdy (2002) is a recent study of high-frequency volatility using a Markov switching model.

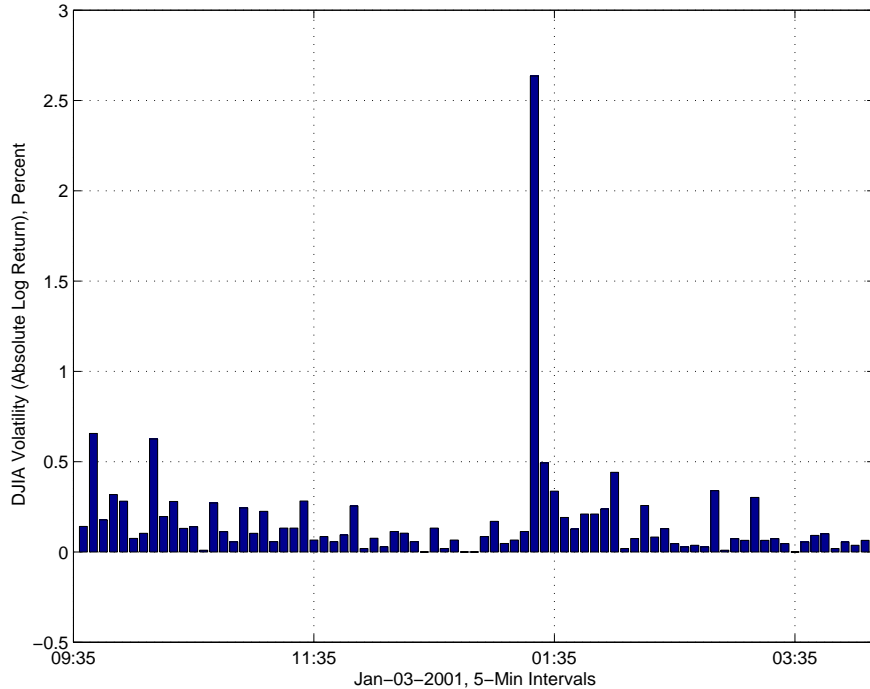


Figure 2: The Dow Jones Industrial Average (DJIA) volatility during January 3, 2001. In this day, the DJIA increased from 10,646 (previous day closing value) to 10,946 (closing value that day). The 2.82% daily increase was relatively large and the market on this day can be classified as “volatile”. However, a close inspection of 5-min DJIA values shows that the market was not volatile during the entire trading session.

value) to 10,946 (closing value that day). The 2.82% daily increase was relatively large and the market in this day can be classified as “volatile”. However, a closer inspection of 5-min DJIA values shows that the market was not volatile during the entire trading session in that day. If we look at the data from an hourly scale, the high volatility took place at the beginning of the trading session (first hour) and between 1:00-2:00 p.m. Other than these two hours, the market was not volatile on an hourly scale. Similarly, high volatility was only present for certain intervals of the intraday scales this particular day. A successful method to describe the market dynamics at different scales (monthly, daily, hourly, etc.) must be able to separate each time-scale component from the observed data.⁹ Though it is not common in economics and finance, wavelet methodology has been proved to be an excellent tool to reach this goal in several scientific areas.

Wavelet coefficients decompose the information from the original time series into

⁹One apparent question regarding the wavelet methodology is whether comparisons of volatilities are fair, and whether there is an issue of the use of future information, as low-frequency volatility uses more information on the time-domain relative to a high-frequency volatility. The usage of future information is not a concern, purely because of the fact that this study is an historical analysis describing market dynamics at different scales.

pieces associated with both time and scales. Since the wavelet coefficients capture the variation of volatility at a given scale and interval of time, we model the wavelet coefficients directly. The wavelet coefficients can be viewed as differences between weighted averages where the weights are determined by a given wavelet filter. If the concern is the total variation of the data at various time scales, it is essential to work with wavelet coefficients. For the current analysis, the important issue in a given scale is “how large a wavelet coefficient is”. If it is relatively large (relative to the average in this time scale), then it implies there was a sudden change in average volatility at that scale, meaning the system had switched to a “high volatility state”. On the other hand, if the wavelet coefficient is small, it implies no large change in volatility (relative to the average in that time scale) and that a “low volatility state” prevails.¹⁰

The outline of this paper is as follows. Section 2 introduces the discrete wavelet transform in terms of orthonormal matrices and digital filters. Multiresolution analysis, the additive decomposition of a time series based on the discrete wavelet transformation, is also introduced. Section 3 explores how the additive decomposition of high-frequency foreign exchange (FX) rates, through multiresolution analysis, accurately and efficiently isolates features in high-frequency U.S. Dollar-Deutsche Mark (USD-DEM) series. The primary model in this paper, the wavelet based hidden Markov model, is explained in Section 4 with emphasis on the hidden Markov tree formulation that allows for dependencies between wavelet coefficients across scales. Section 5 examines the wavelet hidden Markov tree modeling of high-frequency USD-DEM FX rates. In the study of the stock markets, we use a unique high-frequency stock market data set, namely the Dow Jones Industrial Average (DJIA) Index which includes the September 11, 2001 crisis. We discuss the methodology presented here along with future directions in Section 6.

2 Wavelet Methodology

The discrete wavelet transform (DWT) is a mathematical tool that projects a time series onto a collection of orthonormal basis functions (wavelets) to produce a set of wavelet coefficients. These coefficients capture information from the time series at different frequencies at distinct times. The DWT has the advantage of time resolution by using basis functions that are local in time, unlike the discrete Fourier transform whose sinusoids are infinite and hence cannot produce coefficients that vary over time. The DWT achieves this through a sequence of filtering and downsampling steps applied to a dyadic length vector of observations ($N = 2^J$ for some positive integer J) that yields N wavelet coefficients. The wavelet coefficients decompose the information from the original time series into pieces associated with both time and frequency. The DWT has proven to be useful in capturing dynamics of financial and economic time series;

¹⁰The wavelet coefficients are normalized to a unit time scale for all time scales so that comparisons are carried out in the same unit time scale.

see, for example, an excellent survey by Ramsey (2002) on wavelets in economics and finance. An in-depth introduction to the DWT with applications may be found in Gençay *et al.* (2001b). Here we provide only the essential information in order to establish notation and interpret results from models based on the DWT.

2.1 Wavelet Filters

Unlike the Fourier transform, which uses sine and cosine functions to project the data on, the wavelet transform utilizes a wavelet function that oscillates on a short interval of time. The Haar wavelet is a simple example of a wavelet function that may be used to obtain a multiscale decomposition of a time series. The Haar wavelet filter coefficient vector, of length $L = 2$, is given by $h = (h_0, h_1) = (1/\sqrt{2}, -1/\sqrt{2})$. Three basic properties characterize a wavelet filter:

$$\sum_l h_l = 0, \quad \sum_l h_l^2 = 1, \quad \text{and} \quad \sum_l h_l h_{l+2n} = 0 \quad \text{for all integers } n \neq 0. \quad (1)$$

That is, the wavelet filter sums to zero, has unit energy,¹¹ and is orthogonal to its even shifts. These properties are easily verified for the Haar wavelet filter. The first property guarantees that h is associated with a difference operator and thus identifies changes in the data. The second ensures that the coefficients from the wavelet transform preserves energy. In other words, the coefficients from the wavelet transform are properly normalized and, therefore, will have the same overall variance as the data. This would ensure that no extra information has been added through the wavelet transform nor has any information been excluded. The third property guarantees that the set of functions derived from h will form an orthonormal basis for the detail space and allows us to perform a multiresolution analysis on a finite energy signal. The complementary filter to h is the Haar scaling filter $g = (g_0, g_1) = (1/\sqrt{2}, 1/\sqrt{2})$, which possesses the following attributes:

$$\sum_l g_l = \sqrt{2}, \quad \sum_l g_l^2 = 1, \quad \text{and} \quad \sum_l g_l g_{l+2n} = 0 \quad \text{for all integers } n \neq 0,$$

and satisfies the quadrature mirror relationship $g_l = (-1)^{l+1} h_{L-1-l}$ for $l = 0, \dots, L-1$. The scaling filter follows the same orthonormality properties of the wavelet filter, unit energy and orthogonality to even shifts, but instead of differencing consecutive blocks of observations the scaling filter averages them. Thus, g may be viewed as a local averaging operator.

The transfer or gain function of the length L wavelet and scaling filter coefficients is given by their discrete Fourier transforms (DFTs), respectively,

$$H(f) = \sum_{l=0}^{L-1} h_l e^{-i2\pi fl} \quad \text{and} \quad G(f) = \sum_{l=0}^{L-1} g_l e^{-i2\pi fl} \quad (2)$$

¹¹Energy is defined to be the sum of squares.

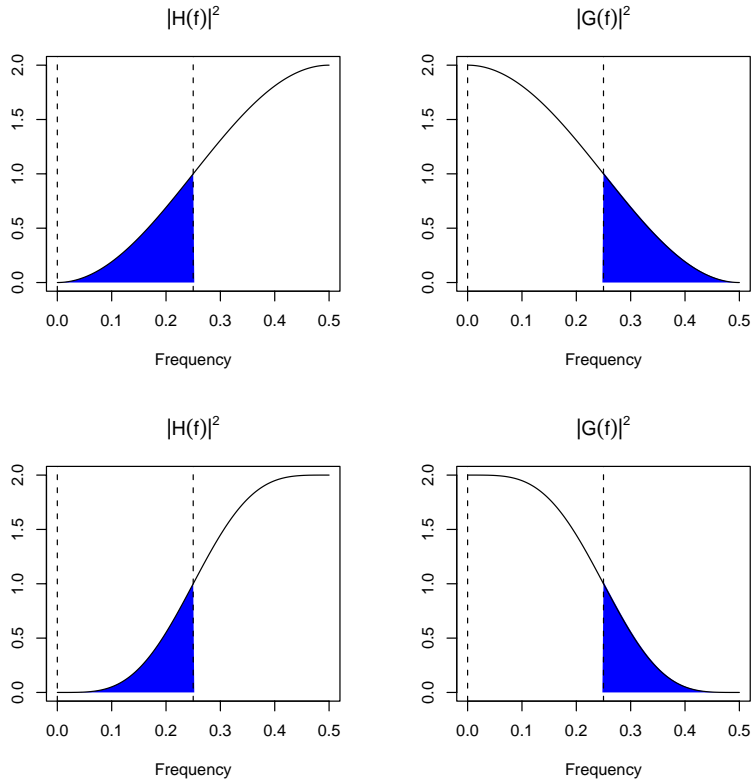


Figure 3: Squared gain functions for the wavelet and scaling filters. Each function is plotted against frequency and indicates which frequencies in the original time series are allowed in the corresponding wavelet coefficients, for $|H(f)|^2$, and scaling coefficients, for $|G(f)|^2$. The first row contains the squared gain functions calculated from the Haar wavelet, while the second row contains the squared gain functions from the D(4) wavelet. The shaded regions correspond to the leakage from an ideal filter.

where $i = \sqrt{-1}$ and f is the frequency. The fact that filter coefficients are related to their transfer function via the Fourier transform is denoted by $\{h\} \leftrightarrow H(f)$ and $\{g\} \leftrightarrow G(f)$ for the wavelet and scaling filters, respectively. To illustrate this relationship, insert the Haar wavelet and scaling coefficients into Equation 2 yielding $H(f) = (1 - e^{-i2\pi f})/\sqrt{2}$ and $G(f) = (1 + e^{-i2\pi f})/\sqrt{2}$, respectively. The transfer functions are complex valued, so for convenience we plot the squared gain functions $\mathcal{H}(f) = |H(f)H^*(f)| = 2\sin^2(\pi f)$ and¹² $\mathcal{G}(f) = |G(f)G^*(f)| = 2\cos^2(\pi f)$ in the first row of Figure 3. The squared gain function associated with the wavelet filter favors

¹² $H^*(f)$ is the complex conjugate of $H(f)$. The operator $|\cdot|$ is the modulus operator for a complex variable, i.e., $|z| = \sqrt{a^2 + b^2}$ where $z = a + ib$. In discrete time setting, the frequency $f = 1/2$, or the angular frequency $\omega = \pi$, is known as a Nyquist frequency, which is the highest possible frequency since the shortest length of a cycle would be two time periods. See Gençay *et al.* (2001b) for more information.

high frequencies and suppresses low frequencies, thus the Haar wavelet is an example of a high-pass filter. The squared gain function derived from the scaling filter does the exact opposite, favoring low frequencies and suppressing high frequencies. It is an example of a low-pass filter. Together, the Haar wavelet and scaling filters capture all the content of a signal and split it into coefficients associated with high and low frequencies. Longer wavelet filters are better approximations to ideal high-pass and low-pass filters, where the frequency axis is split into two disjoint intervals at $f = 1/4$. For example, the Daubechies extremal phase wavelet filter of length four is defined to be

$$h = \frac{1}{4\sqrt{2}} (1 - \sqrt{3}, -3 + \sqrt{3}, 3 + \sqrt{3}, -1 - \sqrt{3})^T.$$

We denote this filter as the D(4) wavelet filter. Its squared gain function is given by $\mathcal{H}(f) = 2 \sin^4(\pi f)[1 + 2 \cos(\pi f)]$ and the squared gain function of the D(4) scaling filter is $\mathcal{G}(f) = 2 \cos^4(\pi f)[1 + 2 \sin(\pi f)]$. These two squared gain functions are plotted in the second row of Figure 3 and illustrate the advantage of longer wavelets. Again, the wavelet filter is an example of a high-pass filter and the scaling filter is an example of a low-pass filter but the differentiation between frequencies above and below $f = 1/4$ is much improved over the Haar wavelet and scaling filters. This is seen by the steeper ascent (descent) of the squared gain functions for the wavelet (scaling) filters and the longer plateaus at each end of the frequency interval. Additional information regarding wavelet filters, including the Haar and longer compactly supported orthogonal wavelets, and their properties may be found in, for example, Mallat (1998) and Gençay *et al.* (2001b).

2.2 The Discrete Wavelet Transform

In this section we introduce notation and concepts in order to compute the discrete wavelet transform (DWT) of a finite-length vector of observations. There are a variety of ways to express the basic DWT. We proceed by introducing the DWT as a simple matrix operation. Let \mathbf{X} be a dyadic length vector ($N = 2^J$) 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 DWT. The vector of wavelet coefficients may be organized into $J + 1$ vectors

$$\mathbf{W} = (\mathbf{W}_1, \mathbf{W}_2, \dots, \mathbf{W}_J, \mathbf{V}_J)^T, \quad (3)$$

where \mathbf{W}_j is a length $N/2^j$ vector of wavelet coefficients associated with changes on a scale of length $\lambda_j = 2^{j-1}$, \mathbf{V}_J is a length $N/2^J$ vector of scaling coefficients associated with averages on a scale of length $2^J = 2\lambda_J$, and \mathbf{W}^T is the matrix transpose of the vector \mathbf{W} . Wavelet coefficients are obtained by projecting the wavelet filter onto the vector of observations. Since Daubechies wavelets may be considered as generalized differences (Gençay *et al.* 2001b, Sec. 4.3), we prefer to characterize the wavelet coefficients this way. For example, a unit scale Daubechies wavelet filter is a generalized difference of

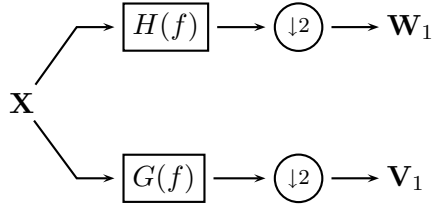


Figure 4: Flow diagram illustrating the decomposition of \mathbf{X} into the unit scale wavelet coefficients \mathbf{W}_1 and the unit scale scaling coefficients \mathbf{V}_1 using the pyramid algorithm. The time series \mathbf{X} is filtered using the wavelet filter $\{h\} \leftrightarrow H(f)$ and every other value removed (downsampled by 2) to produce the length $N/2$ wavelet coefficient vector \mathbf{W}_1 . Similarly, \mathbf{X} is filtered using the scaling filter $\{g\} \leftrightarrow G(f)$ and downsampled to produce the length $N/2$ vector of scaling coefficients \mathbf{V}_1 .

length one; that is, the wavelet filter is essentially taking the difference between two consecutive observations. We call this the wavelet scale of length $\lambda_1 = 2^0 = 1$. A scale two Daubechies wavelet filter is a generalized difference of length two; that is, the wavelet filter first averages consecutive pairs of observations and then takes the difference of these averages. We call this the wavelet scale of length $\lambda_2 = 2^1 = 2$. The scale length increases by powers of two as a function of scale.

The matrix form of the DWT is not computationally efficient and in practice the DWT is implemented via a pyramid algorithm (Mallat 1998). Starting with the data vector \mathbf{X} , filter it using h and g , subsample both filtered outputs to half their original length, and keep the subsampled output from the wavelet filter as wavelet coefficients \mathbf{W}_1 . Repeat the above filtering and downsampling operations on the subsampled output from the scaling filter \mathbf{V}_1 . The complexity of the pyramid algorithm is linear in the number of observations N , faster than the discrete Fourier transform.

Let us go into the pyramid algorithm in more detail so that the calculations are clear. For each iteration of the pyramid algorithm, we require three objects: the data vector \mathbf{X} , the wavelet filter h , and the scaling filter g . The first iteration of the pyramid algorithm begins by filtering (convolving) the data with each filter to obtain the following wavelet and scaling coefficients:

$$W_{1,t} = \sum_{l=0}^{L-1} h_l X_{2t+1-l \bmod N} \quad \text{and} \quad V_{1,t} = \sum_{l=0}^{L-1} g_l X_{2t+1-l \bmod N},$$

where $t = 0, 1, \dots, N/2 - 1$. Note that the downsampling operation has been included in the filtering step through the subscript of X_t . The N length vector of observations have been high- and low-pass filtered to obtain $N/2$ coefficients associated with this information (\mathbf{W}_1 and \mathbf{V}_1 , respectively). The second step of the pyramid algorithm starts by defining the “data” to be the scaling coefficients \mathbf{V}_1 from the first iteration and applying the filtering operations as above to obtain the second level of wavelet and

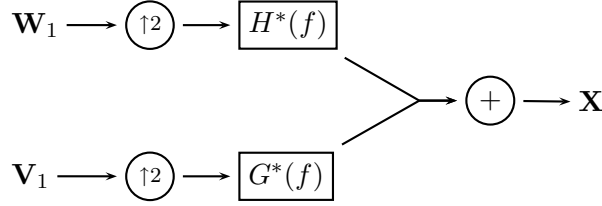


Figure 5: Flow diagram illustrating the reconstruction of \mathbf{X} from the unit scale wavelet coefficients \mathbf{W}_1 and the unit scale scaling coefficients \mathbf{V}_1 . Both \mathbf{W}_1 and \mathbf{V}_1 have zeros inserted in front of every observation (upsampling by 2). The upsampled wavelet coefficients are then filtered using the filter $H^*(f)$ and added to the upsampled scaling coefficients filtered by $G^*(f)$ to form \mathbf{X} .

scaling coefficients:

$$W_{2,t} = \sum_{l=0}^{L-1} h_l V_{1,2t+1-l \bmod N} \quad \text{and} \quad V_{2,t} = \sum_{l=0}^{L-1} g_l V_{1,2t+1-l \bmod N},$$

$t = 0, 1, \dots, N/4 - 1$. Keeping all vectors of wavelet coefficients, and the final level of scaling coefficients, we have the following length N decomposition $\mathbf{W} = (\mathbf{W}_1, \mathbf{W}_2, \mathbf{V}_2)^T$. After the third iteration of the pyramid algorithm, where we apply filtering operations to \mathbf{V}_2 , the decomposition now looks like $\mathbf{W} = (\mathbf{W}_1, \mathbf{W}_2, \mathbf{W}_3, \mathbf{V}_3)^T$. This procedure may be repeated up to J times where $J = \log_2 N$ and gives the vector of wavelet coefficients in Equation 3.

Consider the four-dimensional vector $\mathbf{X} = (2, 3, -2, -1)^T$ and apply the pyramid algorithm to produce the Haar DWT coefficient vector $\mathbf{W} = (\mathbf{W}_1, \mathbf{W}_2, \mathbf{V}_2)^T$. The first application of the pyramid algorithm yields the vectors $\mathbf{W}_1 = (1/\sqrt{2}, 1/\sqrt{2})^T$ and $\mathbf{V}_1 = (5/\sqrt{2}, -3/\sqrt{2})^T$. Notice that the wavelet coefficients are identical, since the local change was $+1$ between the adjacent observations X_0, X_1 and X_2, X_3 . The scaling coefficients may be thought of as local averages, one positive and one negative. The second application of the pyramid algorithm to \mathbf{V}_1 yields the vectors $\mathbf{W}_2 = -4$ and $\mathbf{V}_2 = 1$, where the wavelet coefficient captures the change in local averages and the scaling coefficient is proportional to the sample mean. Thus, the vector of Haar DWT coefficients for \mathbf{X} is

$$\mathbf{W} = \left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}, -4, 1 \right)^T. \quad (4)$$

Inverting the DWT is achieved through upsampling the final level of wavelet and scaling coefficients, convolving them with their respective filters (wavelet for wavelet and scaling for scaling) and adding up the two filtered vectors. Figure 5 gives a flow diagram for the reconstruction of \mathbf{X} from the first levels wavelet and scaling coefficient vectors. The symbol $\uparrow 2$ means that a zero is inserted before each observation in \mathbf{W}_1 and \mathbf{V}_1 (upsampling by 2). Starting with the final level of the DWT, upsampling the vectors

\mathbf{W}_J and \mathbf{V}_J will result in two new vectors $\mathbf{W}_J^0 = (0, W_{J,0})^T$ and $\mathbf{V}_J^0 = (0, V_{J,0})^T$. The level $J - 1$ vector of scaling coefficients \mathbf{V}_{J-1} is given by

$$V_{J-1,t} = \sum_{l=0}^{L-1} h_l W_{J,t+l \bmod 2}^0 + \sum_{l=0}^{L-1} g_l V_{J,t+l \bmod 2}^0,$$

$t = 0, 1$. Notice that the length of \mathbf{V}_{J-1} is twice that of \mathbf{V}_J , as to be expected. The next step of reconstruction involves upsampling to produce $\mathbf{W}_{J-1}^0 = (0, W_{J-1,0}, 0, W_{J-1,1})^T$ and $\mathbf{V}_{J-1}^0 = (0, V_{J-1,0}, 0, V_{J-1,1})^T$, and the level $J - 2$ vector of scaling coefficients \mathbf{V}_{J-2} is given by

$$V_{J-2,t} = \sum_{l=0}^{L-1} h_l W_{J-1,t+l \bmod 4}^0 + \sum_{l=0}^{L-1} g_l V_{J-1,t+l \bmod 4}^0,$$

$t = 0, 1, 2, 3$. This procedure may be repeated until the first level of wavelet and scaling coefficients have been upsampled and combined to produce the original vector of observations; that is,

$$X_t = \sum_{l=0}^{L-1} h_l W_{1,t+l \bmod N}^0 + \sum_{l=0}^{L-1} g_l V_{1,t+l \bmod N}^0,$$

$t = 0, 1, \dots, N - 1$. This is exactly what is displayed in Figure 5.

We now illustrate wavelet reconstruction using the four-dimensional signal \mathbf{X} and its Haar decomposition (Equation 4). First, the upsampled vectors $\mathbf{W}_2^0 = (0, -4)^T$ and $\mathbf{V}_2^0 = (0, -1)^T$ are combined to produce $\mathbf{V}_1 = (5/\sqrt{2}, -3/\sqrt{2})^T$. The second set of upsampled vectors $\mathbf{W}_1^0 = (0, 1/\sqrt{2}, 0, 1/\sqrt{2})^T$ and $\mathbf{V}_1^0 = (0, 5/\sqrt{2}, 0, -3/\sqrt{2})^T$ are then combined to produce the original vector of observations $\mathbf{X} = (2, 3, -2, -1)^T$.

To contrast different wavelet filters, Figure 6 contains a sample of $2^5 = 32$ observations with a level shift at $t = 16$ and nonstationary variance.¹³ Three wavelet filters, namely, Haar, D(4) and LA(8) wavelet filters, were used with varying lengths $L = 2, 4, 8$. All wavelet filters capture the nonstationary variance in the first scale of wavelet coefficients \mathbf{W}_1 , with the second half of the coefficients being more variable than the first half. In addition, all three capture the obvious level shift at the midpoint of the observations with a large positive wavelet coefficient in \mathbf{W}_5 . Because the signal is piecewise in nature, the Haar wavelet filter is most suitable for the analysis of this signal. It is very important to match the wavelet filter to the underlying features of the observed series.

2.3 Choice of Wavelet Filters

The selection of a particular wavelet filter is not trivial in practice and should carefully weigh several aspects of the data: length of the data, complexity of the spectral density

¹³The formula to reproduce the true vector of observations is $\mathbf{X} = 15 \cdot \mathbf{1}_{32} \cdot (-1_{[t \leq 15]} + 1_{[t > 15]}) + \mathbf{Z}_{32} \cdot (1_{[t \leq 15]} + 6 \cdot 1_{[t > 15]})$, where Z is a Gaussian random variable with mean zero and standard deviation one.

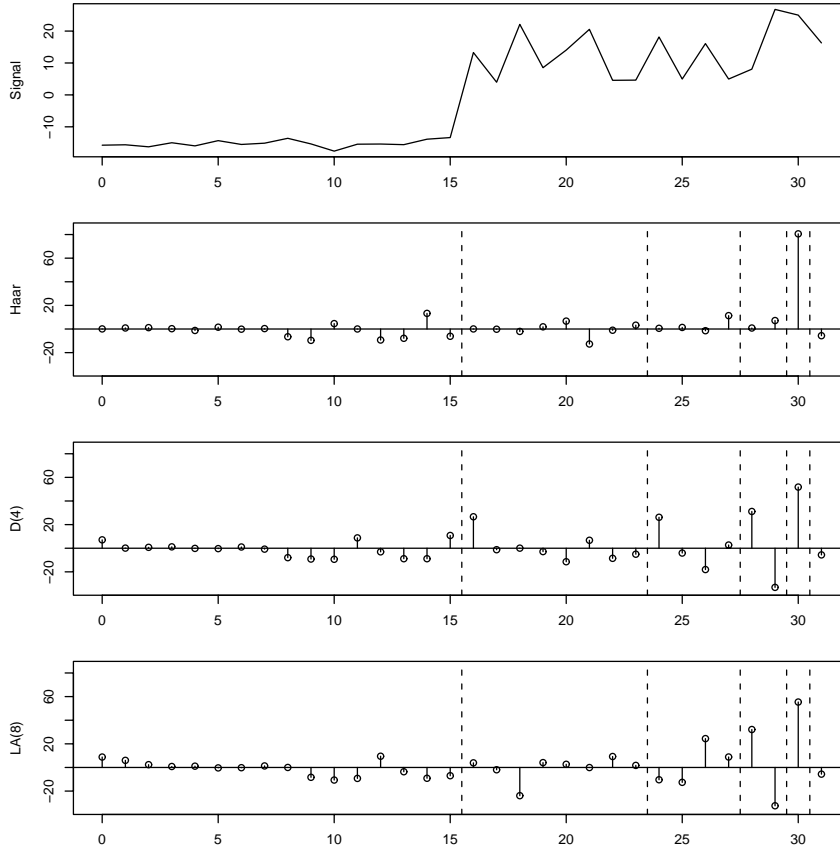


Figure 6: Example vector of $2^5 = 32$ observations. The original signal is plotted in the top row with the corresponding vectors of DWT coefficients below for the Haar, D(4) and LA(8) wavelet filters, respectively. The vertical lines delineate the scales – from left to right – \mathbf{W}_1 , \mathbf{W}_2 , \mathbf{W}_3 , \mathbf{W}_4 , \mathbf{W}_5 and \mathbf{V}_5 .

function, and the underlying shape of features in the data. The length of the original data is an issue because the distribution of wavelet coefficients computed using the boundary will be drastically different from wavelet coefficients computed from complete sets of observations. The shorter the wavelet filter, the fewer so-called “boundary” wavelet coefficients will be produced (and potentially discarded). With the luxury of high-frequency data, the effects of boundary wavelet coefficients are minimized and we are allowed to select from longer filters if necessary.

The complexity of the spectral density function is important because wavelet filters are finite in the time domain and thus infinite, although well localized, in the frequency domain. If the spectral density function is quite dynamic, then shorter wavelet filters may not be able to separate the activity between scales. Longer wavelet filters would thus need to be employed. Clearly a balance between frequency localization and time localization is needed. In most data sets of reasonable length, this balance is not

difficult. From previous studies on high-frequency FX rates (Gençay *et al.* 2001a; Gençay *et al.* 2001c), a moderate length wavelet filter, for example length eight, is adequate to deal with the stylized features in the data.

Finally, and most importantly, there is the issue of what the underlying features of the data look like. This is very important since wavelets are the basis functions, or building blocks, of the data. If one chooses a wavelet filter that looks nothing like the underlying features, then the decomposition will be quite inefficient. So care should be taken when selecting the wavelet filter and what its corresponding basis function looks like. Issues of smoothness and (a)symmetry are the most common desirable characteristics for wavelet basis functions. For this study, we chose to balance smoothness, length and symmetry by selecting the Daubechies least asymmetric wavelet filter of length eight, LA(8). This is a widely used wavelet and is applicable in a wide variety of data types.

2.4 Multiresolution Analysis

The concept of a multiresolution analysis (MRA) is that a given time series, with finite variance, may be decomposed into different approximations associated with unique resolutions (or time horizons). The difference between consecutive approximations, say at levels $J - 1$ and J in the decomposition, is the information contained in the wavelet coefficients at scale J . We have already seen that the series X_t may be decomposed and then reconstructed using straightforward matrix operations in the previous section. We now proceed to show how an MRA produces an additive decomposition of the same series.

Let us assume the level J DWT has been applied to the dyadic length vector \mathbf{X} to obtain the wavelet coefficient vector $\mathbf{W} = (\mathbf{W}_1, \dots, \mathbf{W}_J, \mathbf{V}_J)^T$. We may now formulate an additive decomposition of \mathbf{X} by reconstructing the wavelet coefficients at each scale independently. Let $\mathcal{D}_j = \mathcal{W}_j^T \mathbf{W}_j$ define the j th level *wavelet detail* associated with changes in \mathbf{X} at the scale λ_j (for $j = 1, \dots, J$). The wavelet coefficients $\mathbf{W}_j = \mathcal{W}_j \mathbf{X}$ represent the portion of the wavelet analysis (decomposition) attributable to scale λ_j , while $\mathcal{W}_j^T \mathbf{W}_j$ is the portion of the wavelet synthesis (reconstruction) attributable to scale λ_j . For a length $N = 2^J$ vector of observations, the vector $\mathcal{S}_J = \mathcal{V}_J^T \mathbf{V}_J$ is equal to the sample mean of the observations.

A multiresolution analysis (MRA) may now be defined via

$$X_t = \sum_{j=1}^J \mathcal{D}_{j,t} + \mathcal{S}_J \quad t = 0, \dots, N - 1. \quad (5)$$

That is, each observation X_t is a linear combination of wavelet detail coefficients at time t . Let $\mathcal{S}_j = \sum_{k=j}^{J+1} \mathcal{D}_k$ define the j th level *wavelet smooth* (for $1 \leq j \leq J$). Whereas the wavelet detail \mathcal{D}_j is associated with variations at a particular scale, \mathcal{S}_j is a cumulative sum of these variations and will be smoother and smoother as j increases.

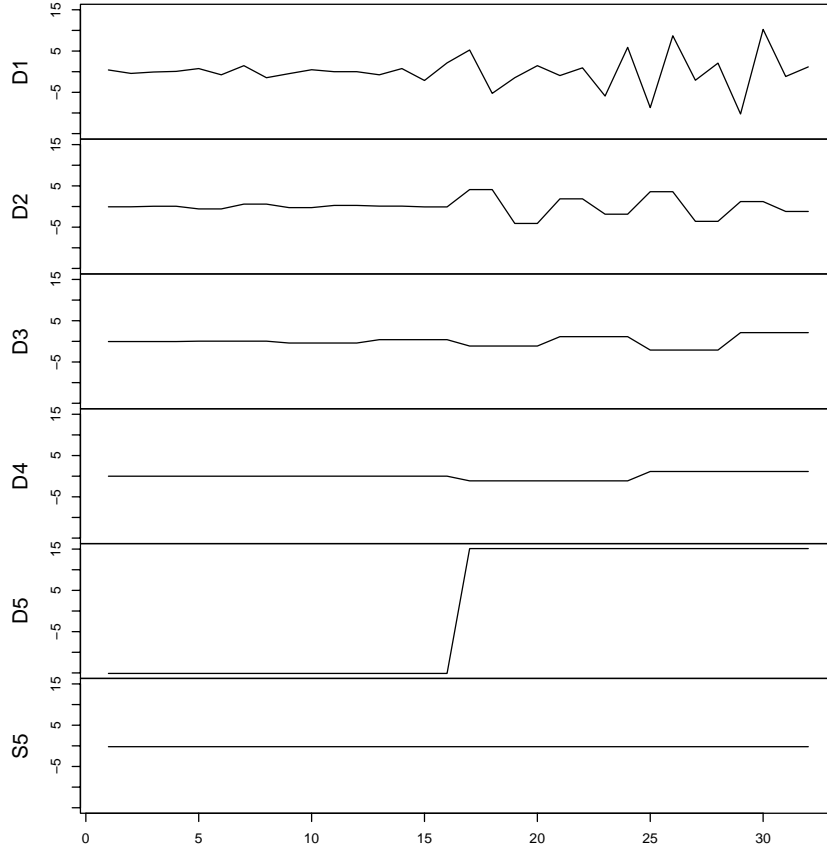


Figure 7: Multiresolution analysis of example vector of $2^5 = 32$ observations. The wavelet details and smooth form an additive decomposition of the original series, so that adding across scales produces the example vector exactly. The vertical axis is the same for each scale.

In fact, $\mathbf{X} - \mathcal{S}_j = \sum_{k=1}^j \mathcal{D}_k$ so that only lower-scale details (high-frequency features) from the original series remain. The j th level *wavelet rough* characterizes the remaining lower-scale details through

$$\mathcal{R}_j = \sum_{k=1}^j \mathcal{D}_k, \quad 1 \leq j \leq J.$$

The wavelet rough \mathcal{R}_j is what remains after removing the wavelet smooth from the vector of observations. For smaller j the wavelet rough is less smooth. A vector of observations may thus be decomposed through a wavelet smooth and rough via $X_t = \mathcal{S}_{j,t} + \mathcal{R}_{j,t}$, for all j, t , which is equivalent to Equation 5. The wavelet details are the differences between either adjacent wavelet smooths or adjacent wavelet roughs.

Given that the Haar DWT was preferred when analyzing the length 32 series \mathbf{X} in the previous section, we perform a Haar MRA with the results in Figure 7. The nonstationary variance is succinctly captured in the first two wavelet details \mathcal{D}_1 and

\mathcal{D}_2 , where the first half of coefficients are near zero and the second half of coefficients are distinctly nonzero. Wavelet details \mathcal{D}_3 and \mathcal{D}_4 indicate there is no activity in those scales associated with 4-8 and 8-16 unit changes. The level shift is captured in \mathcal{D}_5 while the overall mean is essentially zero, as seen in \mathcal{S}_5 .

3 Differentiating Time Horizons in High-Frequency Data

As previously explored in Gençay *et al.* (2001a), the DWT is an effective tool for removing seasonalities in high-frequency data series. In this section we explore the ability of a multi-scale decomposition (specifically the MRA described in Section 2.4) to reproduce the correlation structure of realized volatility found at different sampling rates of high-frequency data. The multi-scale decomposition of realized volatility is a demonstration that modeling high-frequency realized volatility in the wavelet domain captures the features at a variety of sampling rates simultaneously, whereas current methodology only models one fixed sampling frequency.

3.1 Actual Realized Volatility

We follow Andersen *et al.* (2001) and Dacorogna *et al.* (2001) to define realized volatility (or actual realized volatility) via

$$\sigma_{t,\delta}^2 = \sum_{k=0}^{\delta-1} r_{t+k/\delta}^2, \quad (6)$$

where $r_{t+k/\delta} = p_{t+k/\delta} - p_{t+(k-1)/\delta}$ are continuously compounded returns sampled δ times per day. The raw 5-min return series was obtained from Olsen & Associates spanning January 1, 1987, through December 31, 1998, and excludes weekends (defined to be Friday 21:05 GMT until Sunday 21:00 GMT). This results in a series of 901,152 high-frequency return observations, or 3,129 days of data. Hence, the sampling rate δ will vary depending on the level of aggregation needed to calculate realized volatility at coarser levels of time. For example, $\delta = 288$ for daily realized volatility but we will also consider 20-min and hourly realized volatility with $\delta = 4$ and $\delta = 12$, respectively.

Looking at the definition of realized volatility more closely, there are two operations in Equation 6. First, a filter of length δ is applied to the squared returns with each value of the filter being one. For dyadic lengths, this filter mimics the Haar scaling filter. Then a downsampling operation is performed that picks every δ th value from the smoothed return series. Hence, aggregation is related to the Haar DWT through its use of filtering and downsampling but it utilizes a non-orthogonal filter.

3.2 Wavelet Realized Volatility

In contrast to simple aggregation to produce realized volatility at different time horizons, we propose wavelet multi-scaling instead. Although the filtering of the DWT does

not allow for the filtering of arbitrary time spans (see Table 1), the fixed partitioning of 5-min volatility via multiresolution analysis (MRA) is feasible. We first perform an MRA down to level $J = \log_2(N)$ on the squared 5-min return series, where N is the longest dyadic piece of the series. For a given sampling rate δ there is a wavelet smooth that captures features that are longer than or equal to δ . For example, an hourly sampling rate of $\delta = 12$ would involve the wavelet smooth at wavelet scale λ_4 ; it is associated with 80-min averages, not 60, since the 3rd scale is associated with averages shorter than 60 minutes. Once the level, in this case $j = 4$, has been determined the wavelet smooth corresponding to that level is constructed from the wavelet details from the MRA using the formula in Section 2.4. For the example of an hourly sampling rate the wavelet smooth is given by $\mathcal{S}_{4,t} = \sum_{k=5}^{J+1} \mathcal{D}_{k,t}$.

Once the appropriate wavelet smooth has been obtained, the wavelet-based realized volatility (or wavelet realized volatility) is given by

$$\nu_{t,\delta}^2 = \sum_{k=0}^{\delta-1} \mathcal{S}_{j,t+k/\delta},$$

where $\mathcal{S}_{j,t}$ is the wavelet smooth associated with scale λ_j . That is, we aggregate the wavelet smooth in order to compare wavelet realized volatility to actual realized volatility for the sampling rate δ . From one application of an MRA, up to J distinct wavelet realized volatility series are produced. Starting from 5-min squared returns, 20-min, hourly and daily wavelet realized volatility will use scales λ_2 , λ_4 , and λ_9 . With all the information from the original 5-min time scale preserved in the wavelet details and smooth, one can look for features at multiple time horizons simultaneously.

3.3 Autocorrelation Functions for Realized Volatility

To evaluate how well the discrete wavelet transform (DWT) captures dynamics in a volatility series at different time horizons, we look at the sample autocorrelation functions (ACFs) for both actual realized volatility and wavelet realized volatility. The sample ACF for actual realized volatility is defined to be

$$\hat{\rho}_\sigma(\tau, \delta) = \frac{\sum_{t=0}^{N-\tau-1} (\sigma_{t,\delta}^2 - \bar{\mu}_\sigma)(\sigma_{t+\tau,\delta}^2 - \bar{\mu}_\sigma)}{\text{var}\{\sigma_{t,\delta}^2\}}, \quad \tau = 0, 1, \dots, N-1, \quad (7)$$

where $\text{var}\{\sigma_{t,\delta}^2\} = \sum_{t=0}^{N-1} (\sigma_{t,\delta}^2 - \bar{\mu}_\sigma)^2$ and $\bar{\mu}_\sigma$ is the sample mean of the actual realized volatility series. This is the usual definition of the covariance for the actual realized volatility series at lag τ divided by the variance (covariance at lag 0) of the actual realized volatility series. The sample ACF $\hat{\rho}_\sigma(\tau, \delta)$ estimates the true ACF of actual realized volatility at a given sampling rate.

The sample ACF for wavelet realized volatility is defined similarly via

$$\hat{\rho}_\nu(\tau, \delta) = \frac{\sum_{t=0}^{N-\tau-1} (\nu_{t,\delta}^2 - \bar{\mu}_\nu)(\nu_{t+\tau,\delta}^2 - \bar{\mu}_\nu)}{\text{var}\{\nu_{t,\delta}^2\} + \text{var}\{\nu_{t,\delta}^2(\mathcal{R}_j)\}}, \quad \tau = 0, 1, \dots, N-1, \quad (8)$$

where $\text{var}\{\nu_{t,\delta}^2\} = \sum_{t=0}^{N-1} (\nu_{t,\delta}^2 - \bar{\mu}_\nu)^2$ and $\bar{\mu}_\nu$ is the sample mean of the wavelet realized volatility series. The second term in the denominator of Equation 8 is the remainder of the variance in the MRA coefficients not accounted for by the wavelet smooth \mathcal{S}_j . Recall, the wavelet rough is computed via $\mathcal{R}_j = \sum_{k=1}^j \mathcal{D}_k$ and is then aggregated to form $\nu_{t,\delta}^2(\mathcal{R}_j) = \sum_{k=0}^{\delta-1} \mathcal{R}_{j,t+k/\delta}$. Finally, the variance of the wavelet realized volatility based on the wavelet rough is given by $\text{var}\{\nu_{t,\delta}^2(\mathcal{R}_j)\} = \sum_{t=0}^{N-1} [\nu_{t,\delta}^2(\mathcal{R}_j) - \bar{\mu}_\nu(\mathcal{R}_j)]^2$, where $\bar{\mu}_\nu(\mathcal{R}_j)$ is the sample mean of $\nu_{t,\delta}^2(\mathcal{R}_j)$.

Figure 8 shows the sample ACFs for the wavelet realized volatility $\hat{\rho}_\nu(\tau, \delta)$ and actual realized volatility $\hat{\rho}_\sigma(\tau, \delta)$ of USD-DEM returns at 20-minute, hourly and daily sampling rates. Lags up to 30 days are displayed for all sampling rates, although the number of lags in each plot differs. In both cases the wavelet realized volatility is virtually indistinguishable from the actual realized volatility at all lags except the first. For daily realized volatility, the wavelet-based version exhibits less variation from lag to lag due to the fact that the high-frequency content was removed via the MRA.

4 Wavelet-based Hidden Markov Trees

4.1 Introduction

Modeling in the wavelet domain usually ignores the correlation between wavelet coefficients, falling back on the assumption that the DWT is a whitening transform. Indeed, for a wide range of naturally occurring time series the wavelet coefficients may be treated as uncorrelated and Gaussian. These assumptions are not valid in the context of analyzing high-frequency FX rates. First, since the quantity of interest is volatility there is no opportunity to assume a Gaussian distribution for the observed series. Second, the unknown and potentially complex correlation structure in the series most likely does not produce approximately uncorrelated wavelet coefficients. We propose to borrow a probabilistic model from signal processing and apply it to the wavelet decomposition of a high-frequency volatility series. By taking advantage of the tree-based structure of the DWT, we provide an efficient representation and estimation technique of the underlying joint distribution of the wavelet coefficients. Through this representation the multiscale decomposition of the volatility series is classified into a state of high or low volatility.

In the context of signal processing applications, Crouse, Nowak, and Baraniuk (1998) proposed a variety of hidden Markov models for wavelet decompositions of one- and two-dimensional data sets (time series and images). The assumption of uncorrelated wavelet coefficients was replaced by the possibility of allowing correlation between scales of the DWT or within scales of the DWT. The assumption of Gaussianity was also replaced by specifying a small number of unobserved (hidden) states, and representing the distribution of wavelet coefficients as a mixture of Gaussian distributions conditional on the hidden state variable. Figure 9 illustrates two possible models for de-

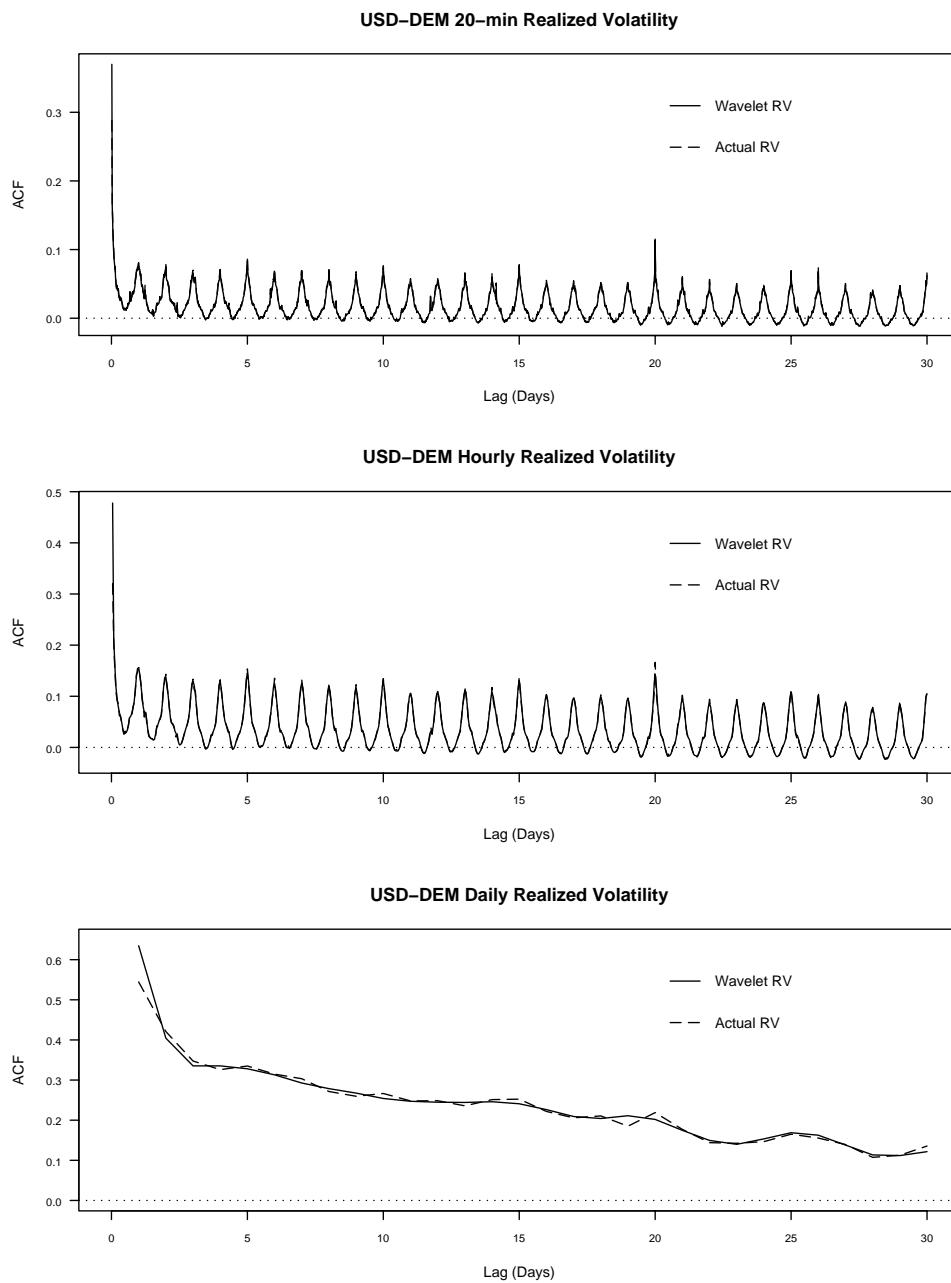


Figure 8: Sample autocorrelation functions (ACF) of the realized volatility (RV) for the U.S. Dollar - Deutsche Mark (USD-DEM) foreign exchange rate at 20-minute, hourly and daily sampling rates. The solid line denotes the ACF based on wavelet realized volatility and the dashed line denotes the actual realized volatility. The lag zero coefficient was omitted from all plots since it is identically one.

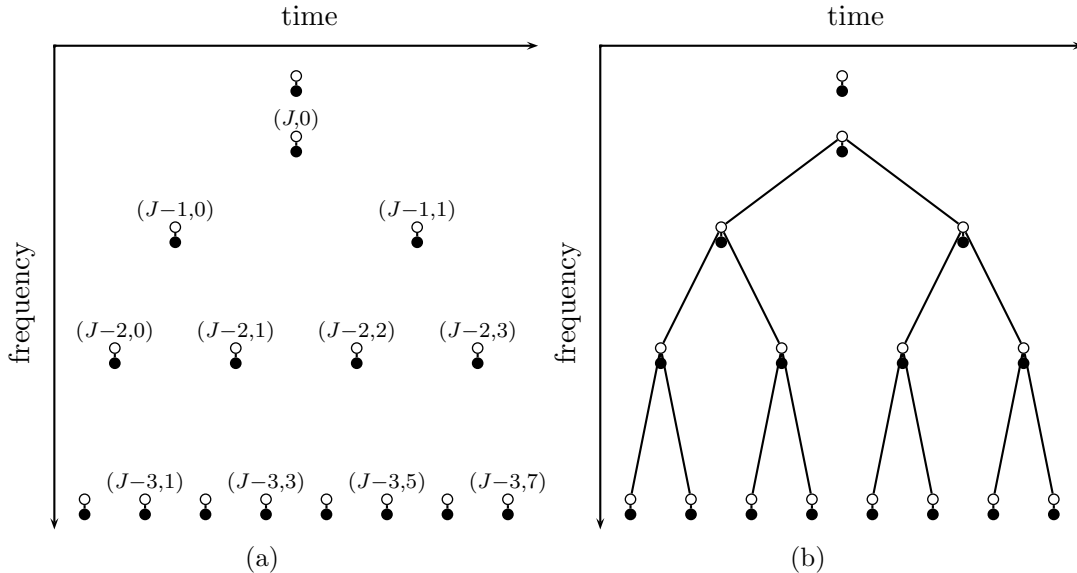


Figure 9: Models for dependence between wavelet coefficients. Each wavelet coefficient (black dot) is modeled as a mixture of two Gaussian probability density functions with a hidden state (white dot). The uppermost node is the scaling coefficient and the one below it is the coarsest wavelet coefficient (which we call the root node). From the root node downward, each subsequent level produces two children for each state and wavelet coefficient pair from the previous level. This pattern continues until level $j = 1$. For the independent mixture model (a) all state and wavelet coefficient pairs are assumed independent, while for the hidden Markov tree model (b) hidden state variables are linked in order to model the dependence between levels of the wavelet transform.

pendence between wavelet coefficients in the DWT. The first model (Figure 9a) assumes unconditional independence between all wavelet coefficients, each with an unobserved state variable. This independence is a common assumption when formulating wavelet-based models, but ignores a wealth of information contained in the local structure of the time series that is extracted through the DWT.

One possible model of dependence between wavelet coefficients is to allow association between scales but not within scales (Figure 9b). This so-called wavelet hidden Markov tree (HMT) model takes advantage of the persistence of large or small wavelet coefficients across scales with the state variables are connected vertically between scales. Let \mathbf{W} be a vector of wavelet coefficients from a dyadic length vector of observations \mathbf{X} ; see Section 2.2. The first point is that the DWT coefficients may be organized into a binary tree, denoted by $\mathcal{T} = \{(j, n) : j = 0, \dots, J; n = 0, \dots, 2^j - 1\}$. The wavelet coefficient $W_{J,0}$ is the root of the tree with children $W_{J-1,0}$ and $W_{J-1,1}$ (the only two wavelet coefficients at scale λ_{J-1}), $W_{J-1,0}$ has children $W_{J-2,0}$ and $W_{J-2,1}$, and so on. The wavelet HMT model is directional in that information from longer time horizons directly influences shorter time horizons, but not vice-versa.

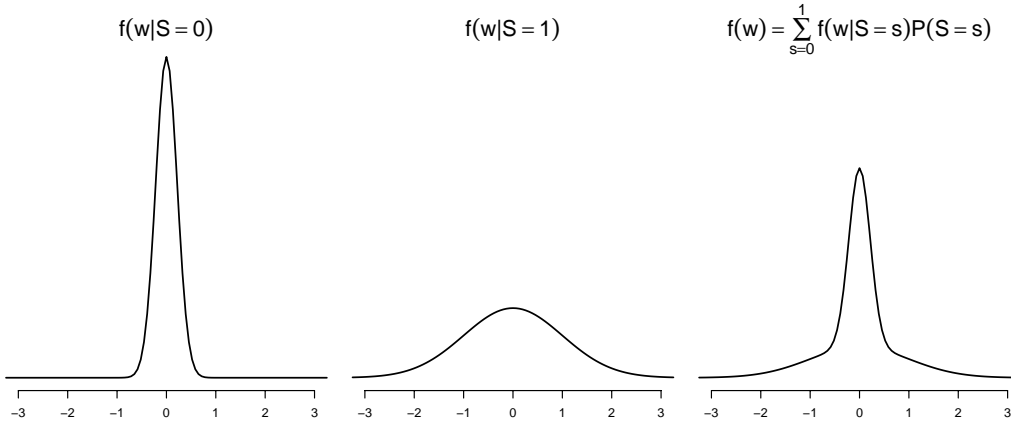


Figure 10: Gaussian mixture model ($M = 2$) for wavelet coefficient W . The Gaussian conditional probability density function (PDF) for $W|S$ are shown in the first two panels as well as the non-Gaussian mixture PDF for W . The first state ($S = 0$) corresponds to a low-variance Gaussian PDF, while the second state ($S = 1$) corresponds to a high-variance Gaussian PDF.

We impose the following five properties on the structure of our wavelet HMT model (Durand and Gonçalves 2001):

1. The wavelet coefficient W is modeled by a mixture distribution with probability density function

$$f_W(w) = \sum_{s=0}^{M-1} f_{W|S}(w | S = s)P(S = s), \quad (9)$$

where S is a discrete random variable (the hidden state) with M possible values.

2. Let $\mathbf{S} = (\mathbf{S}_1, \dots, \mathbf{S}_J, \mathbf{S}_{V_J})^T$ define the state vector associated with the vector of wavelet coefficients \mathbf{W} , indexed in the same way. Thus, the state vector may be organized as a binary tree rooted at $S_{J,0}$ and read from right to left. Since we are using two indices, one for the scale and one for the location within scale, the parent of $S_{j,n}$ is given explicitly by $S_{j+1, \lfloor n/2 \rfloor}$ for $j = 2, \dots, J$.¹⁴ The children of $S_{j,n}$ are given explicitly by $S_{j-1, 2n}$ and $S_{j-1, 2n+1}$. The notation for state variables in the binary tree is illustrated in Figure 9a using only the subscripts. The root is $S_{J,0}$, the next level down from left to right is $S_{J-1,0}$ and $S_{J-1,1}$; the next level down is $S_{J-2,0}$, $S_{J-2,1}$, $S_{J-2,2}$ and $S_{J-2,3}$; and so on. Dependence between scales in the state variables is illustrated in Figure 9b and uses an identical labeling scheme.

¹⁴ $\lfloor x \rfloor$ refers to the floor function of x which is the greatest integer in x , i.e., the largest integer less than or equal to x .

3. The state variable $S_{j,n}$ is independent of all other states given its parent and children; i.e.,

$$P(S_{j,n} | \{S_{a,b}\}_{a \neq j, b \neq n}) = P(S_{j,n} | S_{j+1, \lfloor n/2 \rfloor}, S_{j-1, 2n}, S_{j-1, 2n+1}).$$

4. The joint probability distribution of the wavelet coefficient vector \mathbf{W} is independent given the state vector \mathbf{S} ; i.e.,

$$f_{\mathbf{W}|\mathbf{S}}(\mathbf{W}) = \prod_{(j,n) \in \mathcal{T}} f_{W_{j,n}|\mathbf{S}}(w_{j,n}).$$

5. The wavelet coefficient $W_{j,n}$ is independent of all other states given its own state; i.e.,

$$f_{W_{j,n}|\mathbf{S}}(w_{j,n}) = f_{W_{j,n}|S_{j,n}}(w_{j,n}) \quad \text{for all } (j,n) \in \mathcal{T}.$$

The last two properties are known as conditional independence properties. We assume the mixture distribution for W is based on Gaussian probability density functions (PDFs) with mean μ_s and variance σ_s^2 , $s \in \{0, 1, \dots, M-1\}$. Figure 10 shows the conditional Gaussian PDFs for a two-state mixture distribution, one conditional Gaussian PDF with a lower variance and another conditional Gaussian PDF with a high variance. They may be combined with an appropriate probability mass function (Equation 9) to produce a non-Gaussian mixture distribution for a given wavelet coefficient W . The resulting distribution for W is distinctly non-Gaussian and gracefully incorporates the features from both the low-variance and high-variance conditional Gaussian PDFs.

Given the properties of the wavelet HMT, the full likelihood is given by

$$f_{\mathbf{W}}(\mathbf{W}) = \sum_{\mathbf{S}} \left\{ P(S_{V_J} = s_{V_J}) f_{V_J|S_{V_J}}(v_J) P(S_{J,0} = s_{J,0}) f_{W_{J,0}|S_{J,0}}(w_{J,0}) \right. \\ \left. \times \prod_{j=1}^{J-1} \prod_{n=0}^{N/2^j-1} f_{W_{j,n}|S_{j,n}}(w_{j,n}) P(S_{j,n} = s_{j,n} | S_{j+1, \lfloor n/2 \rfloor}) \right\}. \quad (10)$$

4.2 Implementation

The parameter vector corresponding to the wavelet HMT consists of the distribution of the root state $S_{J,0}$, the transition probabilities that $S_{j,n}$ is in state s given $S_{j-1, \lfloor n/2 \rfloor}$ is in state r , and the parameters of the Gaussian mixtures (μ, σ^2) . For the applications considered here, we assume the transition matrix is scale-dependent and model a given wavelet coefficient via a two-state Gaussian mixture distribution so that the transition matrix has the form:

$$A_j = \begin{bmatrix} p_{0,j} & 1 - p_{0,j} \\ 1 - p_{1,j} & p_{1,j} \end{bmatrix}, \quad \text{for } j = 1, \dots, J-1.$$

The conditional probabilities

$$p_{0,j} = P(\text{low volatility at scale } j \mid \text{low volatility at scale } j + 1)$$

and

$$p_{1,j} = P(\text{high volatility at scale } j \mid \text{high volatility at scale } j + 1)$$

reflect the persistence of large and small wavelet coefficients from long time horizons to shorter time horizons, respectively. We expect the transition probability from low to high volatility ($1 - p_{0,j}$) to be quite small, and therefore $p_{0,j} \approx 1$ for most scales. The main parameter of interest is $p_{i,j}$. Given a high volatility state at scale $j + 1$, how likely is a high volatility state to persist to scale j ? The complete parameter vector for the wavelet HMT model is explicitly given by

$$\theta = (p_{0,1}, p_{1,1}, \dots, p_{0,J}, p_{1,J}, \mu_{0,1}, \sigma_{0,1}^2, \dots, \mu_{0,J}, \sigma_{0,J}^2, \mu_{1,1}, \sigma_{1,1}^2, \dots, \mu_{1,J}, \sigma_{1,J}^2),$$

and includes all transition probabilities and parameters associated with the Gaussian PDFs. The DWT ensures that the expected value of all wavelet coefficients will be zero when using a wavelet filter of sufficient length. We thus make the assumption that $\mu_{s,j} = 0$ for all s and j .

Maximum likelihood estimation of the parameter vector θ cannot be performed directly on Equation 10. An adaptation of the Expected Maximization (EM) algorithm is applied to this problem where the model parameters θ and distribution of the hidden states \mathbf{S} are estimated jointly, given the observed wavelet coefficients \mathbf{W} . For the estimation step, an upward-downward algorithm for calculating the log-likelihood of the wavelet HMT was developed in Crouse *et al.* (1998). The upward-downward algorithm is similar to the well-known forward-backward algorithm for hidden Markov chains (Baum 1972).¹⁵

The wavelet HMT model is such that dependence between wavelet coefficients is allowed only between scales. That is, if one pictures a binary tree associating wavelet coefficients (Figure 9b), there are no links between adjacent wavelet coefficients within scales – only between and then only from coarse to fine resolution in time. The intuition behind this dependence structure is that if there exists a large wavelet coefficient at a given time horizon (implying a local oscillation with a large amplitude), then at least one of the wavelet coefficients computed using the same data at a shorter time horizon will also be large. That being said, the transition probabilities and parameters of the corresponding mixture distribution are estimated using all wavelet coefficients across time. In this respect, the model uses all available information for parameter estimation, since all wavelet coefficients are used.

However, one should keep in mind that each wavelet coefficient carries only local information. Once the parameter estimates $\hat{\theta}$ have been obtained for the wavelet HMT

¹⁵An alternative implementation of the upward-downward algorithm from Crouse *et al.* (1998) may be found in Durand and Gonçalves (2001).

model, the wavelet coefficients may be classified into one of the M states in the mixture distribution using the Vitterbi algorithm; see, for example, Rabiner (1989) and the references therein. Let $\delta_{j,n}$ be the sequence of states and wavelet coefficients for $W_{j,n}$ on the hidden Markov tree. The Vitterbi algorithm recursively calculates the sequence of states with highest probability from top to bottom using the transition probabilities from the wavelet HMT model. Because of the conditional dependence structure of the model, the Vitterbi algorithm only operates on the states from a small set of wavelet coefficients – the parent and children relative to $W_{j,n}$. Note, wavelet coefficients are associated with either high or low volatility since the mean is identically zero for both Gaussian PDFs.

The exact location of volatility bursts in the wavelet domain relies on how the phase information was treated in the original decomposition. Wavelet coefficients obtained from most implementations of the DWT will exhibit some translation in time. This is accounted for before classification.¹⁶

5 Wavelet HMT Models for High-Frequency Data

5.1 USD-DEM Volatility

Our variable of interest is the realized volatility at the high-frequency 5-min sampling rate. The 5-min foreign exchange (FX) return is defined as

$$r_{t,5} = \log P_t - \log P_{t-1}$$

where P_t is the FX price at time t . The FX volatility is defined by squaring 5-min return $r_{t,5}^2$. We estimated a two-state wavelet HMT model on a span of $N = 2^{19} = 524,288$ observations, from January 4, 1987 to December 27, 1993. This is the largest dyadic-length vector that is less than or equal to the available sample size of just under 1 million FX rates. For reference, Table 1 translates wavelet scales ($j = 1, \dots, 12$) into time horizons that span anywhere from several minutes to a month. We performed a second analysis on the last $N = 524,288$ (from January 9, 1992 to December 31, 1998) observations to check if the wavelet HMT model produces stable estimates for different, but not disjoint, time spans.

Table 2 provides the conditional probabilities for the scale-dependent transition matrix $A_j, j = 1, \dots, 12$. The first thing we notice is the strong vertical dependency in low volatility states. The probability of observing a low volatility state at time scale j , given that there is a low volatility state at time scale $j + 1$ is almost one at all time scales, $p_{0,j} \approx 1$ for $j = 1, \dots, 12$. For example, given that a low volatility state is observed at a 4-7 day time scale (wavelet scale 10), the probability of observing a low

¹⁶There are approximate phase shifts available for all Daubechies wavelet families in Percival and Walden (2000). Once the DWT has been applied, one must circularly shift each vector of wavelet coefficients by an integer amount.

Scale	Time Horizons		
	Minutes	Hours	Days
1	10–20		
2	20–40		
3	40–80	0.7–1.3	
4		1.3–2.7	
5		2.7–5.3	
6		5.3–10.7	
7		10.7–21.3	
8		21.3–42.7	0.9–1.8
9			1.8–3.6
10			3.6–7.1
11			7.1–14.2
12			14.2–28.4

Table 1: Translation of wavelet scales into appropriate time horizons for the USD-DEM high-frequency FX rates ($\Delta t = 5$ minutes). Each scale of the DWT corresponds to a frequency interval, or conversely an interval of periods, and thus each scale is associated with a range of time horizons.

Scales	Low-to-Low	Low-to-High	High-to-High	High-to-Low
11	0.995	0.005	0.981	0.019
10	0.865	0.135	0.996	0.034
9	0.972	0.028	0.703	0.297
8	0.960	0.040	0.782	0.218
7	0.950	0.050	0.736	0.264
6	0.986	0.014	0.478	0.522
5	0.977	0.023	0.554	0.446
4	0.985	0.015	0.452	0.548
3	0.990	0.010	0.547	0.453
2	0.995	0.005	0.505	0.495
1	0.988	0.012	0.490	0.510

Table 2: Conditional probabilities $p_{0,j}$ and $p_{1,j}$ from the scale-dependent transition matrices A_j in the wavelet hidden Markov tree (HMT) model for USD-DEM volatility. The quantity under the heading Low-to-Low, for example, is the probability of low volatility at scale j given there was low volatility at scale $j + 1$. Similarly, the quantity under the heading Low-to-High is the probability of high volatility at scale j given there was low volatility at scale $j + 1$.

volatility state at 2-4 days (wavelet scale 9) is 0.96. Similarly, if a low volatility state is experienced at a 3-5 hour time scale (wavelet scale 5), the probability of observing a low volatility state at 1-3 hours (wavelet scale 4) is 0.99.¹⁷ Technically, this means that

¹⁷Translations of wavelet scales into appropriate time scales are approximate here. For an exact

given a wavelet coefficient associated with low volatility, there is very little chance that it will produce a wavelet coefficient associated with high volatility at the lower scale. Hence, the conditional probabilities for low volatility state-to-low volatility state are almost one. Naturally, the probability of changing states (from low volatility at time scale j to high volatility at time scale $j - 1$) is almost zero as presented in Table 2. We conclude that vertical dependency in low volatility states is an extremely strong one.

The vertical dependency among high volatility states is not as strong as it is in low volatility states, especially at lower time scales. This means that a high volatility state (regime) at a given time scale will not guarantee that wavelet coefficients at the lower time scale will be associated with high volatility. For example, given a high volatility regime at a 4-7 day time scale (wavelet scale 10), the probability of being a high volatility regime at 2-4 days (wavelet scale 9) is 0.78. Similarly, if a high volatility regime prevails at a 3-5 hour time scale (wavelet scale 5), the probability of being in a high volatility regime at 1-3 hours (wavelet scale 4) is 0.55. The reason behind this property is that markets calm down at low time scales (higher frequencies) much before they do at high time scales (lower frequencies). The vertical dependency amongst the high volatility states implies that the probability of changing states from high volatility at time scale j to low volatility at time scale $j - 1$ is relatively high. In particular, the probability of changing from a high volatility regime to a low volatility regime is approximately 0.50 for time scales of approximately 12 hours or less.

These findings establish an important new stylized property of foreign exchange volatility. In addition to the well-known horizontal dependence (conditional heteroscedasticity or volatility clustering and long memory), foreign exchange volatility exhibits strong vertical dependence (persistence across different time scales). Furthermore, the vertical dependence of foreign exchange volatility is asymmetric in the sense that low-to-low volatility states exhibit a strong dependence while high-to-high volatility states possess much less dependence.

Figure 11 displays the estimates of a sequence of states, with low volatility (light rectangles) or high volatility (dark rectangles), from the wavelet HMT model estimated using the first $N = 2^{19} = 524,288$ 5-min volatilities.¹⁸ A visual inspection of the figure shows the vertical dependency in the foreign exchange volatility. Starting from wavelet scale 12 (14-28 day time scale), if the current state is one of low volatility (light rectangle) then we again observe a low volatility state at shorter time scales. However, given a high volatility state (dark rectangle), we observe high volatility states less frequently at shorter time scales. Notice that dark rectangles get narrower as one moves from a long time scale (low frequency) to shorter time scales (high frequencies) for a given period of time. The following section gives an intuitive interpretation of different

translation, see Table 1.

¹⁸A second analysis was performed on the last $N = 2^{19} = 524,288$ (from January 9, 1992 to December 31, 1998) observations from the 5-min volatility series of USD-DEM FX rates. The main findings are similar to the first data set.

time scale volatility states by zooming in on the second half of 1992 in Figure 11.

5.2 Currency Turmoil in 1992 at Different Time Scales

International foreign exchange markets experienced one of their largest turmoils in 1992. The events leading to currency turmoil in European economies and international foreign exchange markets can be tracked back to the German unification. German authorities decided to finance the cost of unification through borrowing, causing an increase in German interest rates. Other currencies of the European economies were forced to increase their interest rates to protect the value of their currencies against the Deutsche Mark (DM). Speculators were already attacking the weaker currencies by betting that they could not sustain the parity with the DM. The increase in interest rates caused further speculation. The Portuguese Escudo (PE), Spanish Peseta (SP), Italian Lira (IL), and British Pound (BP) all fell in value against the DM. On September 15, 1992, the BP and IL left the European exchange rate mechanism (ERM) and the PE and the SP were forced to devalue but stayed in the ERM.

George Soros, manager of the Quantum Fund, was reported to have held a \$10 billion USD short position on the BP and to have made \$1 billion for his fund as a result of the BP's September devaluation. Some other hedge funds were also speculating against the BP. Overall, hedge funds are estimated to have held short on the BP positions, totaling \$11.7 billion in excess of 25% of the government's official reserves in 1992 (\$40 billion). Considering the fact that central bank interventions in the ERM by September 1992 totaled roughly \$100 billion, it is clear that the total speculative position in the market against the weaker currencies was much larger than the total short positions of the hedge funds in these currencies, see Fung and Hsieh (2000).

The estimated sequence of states from the fitted wavelet HMT model clearly indicates that the currency market entered into a high volatility state towards the end of July 1992. Closer inspection of Figure 12 reveals that wavelet coefficients at scales 11 and 12 (approximately a 7-30 day time scale), between July and November 1992, were mostly from a high volatility regime. However, the high volatility state was not uniform across the scales. As we look at the lower scales, the time span of a high volatility state becomes narrower. For example, at scale 8 (approximately 1-2 days) the estimated wavelet coefficients are mostly from a high volatility regime during the month of September 1992. At scale 1 (10-20 min), the high volatility state is observed only for a few days in September. In other words, for a trader who was only interested in intraday movements (scales 1-8), the currency turmoil in 1992 lasted only a couple of days. On the other hand, for a trader (or investor) with a time horizon of 10-15 days, the turmoil started in July and ended in November.

5.3 DJIA Volatility

In our second application, we use unique high-frequency stock market data, namely the Dow Jones Industrial Average (DJIA) Index which includes the September 11, 2001 crisis. The data is in 5 minute intervals and the sample period is from January 4, 1999 to October 16, 2002.¹⁹ For each trading day, the New York Stock Exchange (NYSE) opens at 9:30 a.m. (Eastern Standard Time) and closes at 4:00 p.m. (Eastern Standard Time). There are 79 5-min values per trading day.²⁰ In addition to eliminating official holidays and weekends, we also eliminated data for days when the market was officially closed for more than 1 hour on any given business day. Overall, there are 251 trading days in 1999, 255 trading days in 2000, 249 trading days in 2001 and 200 days in 2002; a total of 955 business days with 75,446 5-min data points. Let us define the 5-min stock return via $r_{t,5} = \log P_t - \log P_{t-1}$, where P_t is the DJIA at time t . Stock market volatility is defined as the squared 5-min return $r_{t,5}^2$. In order to eliminate any price bias at smaller frequencies and to reduce the computational burden, we decided to work with 1-hour fine volatility,²¹ defined as the aggregated 5-min volatilities every hour

$$\sigma_{t,h}^2 = \sum_{i=1}^{12} r_{t,5,i}^2,$$

where the aggregation resulted in a sample size 6287 hourly fine volatilities. Since the wavelet analysis requires a dyadic sample, the largest sample size available was $2^{12} = 4096$. We therefore study the last 4096 hours in the sample from May 1, 2000 to October 16, 2002. For reference, Table 3 translates wavelet scales ($j = 1, \dots, 6$) into time horizons that span anywhere from two hours to four weeks.

Table 4 provides the conditional probabilities for the scale-dependent transition matrix $A_j, j = 1, \dots, 6$. Similar to USD-DEM volatility, there is a strong vertical dependency in low volatility states. The probability of observing a low volatility state at time scale j , given that there is a low volatility state at time scale $j+1$ is no less than 0.9 at all time scales. For example, given that a low volatility state is observed at a 8-16 hour time scale (wavelet scale 3), the probability of observing a low volatility state at 4-8 hours (wavelet scale 2) is 0.97. Similarly, if a low volatility state is experienced at a 2-4 week time scale (wavelet scale 6), the probability of observing a low volatility state at 1-2 weeks (wavelet scale 5) is 0.90.²² Technically, this means that given a wavelet coefficient associated with low volatility, there is very little chance that it will produce a wavelet coefficient associated with high volatility at the lower scale. Hence, the

¹⁹We thank Olsen & Associates, Switzerland, for providing these data.

²⁰The first new index value each day after the trading starts is registered at 9:35AM. The last index value of the day is registered at 4:05PM.

²¹A coarse volatility is based on $(\sum r_j)^2$ whereas a fine volatility is based on $\sum r_j^2$ calculated at the same data points and synchronized. More detailed analysis of the comparisons between fine and coarse volatilities are provided in Dacorogna *et al.* (2001).

²²Translations of wavelet scales into appropriate time scales are approximate here. For an exact translation, see Table 3.

Scale	Time Horizons		
	Hours	Days	Weeks
1	2–4		
2	4–8	0.6–1.2	
3	8–16	1.2–2.5	
4	16–32	2.5–4.9	
5	32–64	4.9–9.8	1–2
6	64–128	9.8–19.6	2–4

Table 3: Translation of wavelet scales into appropriate time horizons for the DJIA volatilities ($\Delta t = 1$ hour). Each scale of the DWT corresponds to a frequency interval, or conversely an interval of periods, and thus each scale is associated with a range of time horizons. A business day is based on 6.5 trading hours.

Scales	Low-to-Low	Low-to-High	High-to-High	High-to-Low
6	0.9058	0.0942	0.3414	0.6586
5	0.9012	0.0988	0.1750	0.8250
4	0.9748	0.0252	0.0691	0.9309
3	0.9469	0.0531	0.1574	0.8426
2	0.9743	0.0257	0.2485	0.7515
1	0.9487	0.0513	0.4571	0.5429

Table 4: Conditional probabilities $p_{0,j}$ and $p_{1,j}$ from the scale-dependent transition matrices A_j in the wavelet hidden Markov tree (HMT) model for DJIA volatility. The quantity under the heading Low-to-Low, for example, is the probability of low volatility at scale j given there was low volatility at scale $j + 1$. Similarly, the quantity under the heading Low-to-High is the probability of high volatility at scale j given there was low volatility at scale $j + 1$.

conditional probabilities for low-to-low volatility states are 0.90 and larger. Naturally, the probability of changing states (from low volatility at time scale j to high volatility at time scale $j - 1$) is not greater than 0.10 as presented in Table 4. We conclude that vertical dependency in low volatility states is quite strong. These findings are in line with those previously discussed for FX volatility.

The vertical dependency among high volatility states is not as strong as it is in low volatility states, especially at lower time scales. This means that a high volatility state (regime) at a given time scale will not guarantee that wavelet coefficients at the lower time scale will be associated with high volatility. For example, given a high volatility regime at a 8-16 hour time scale (wavelet scale 3), the probability of being a high volatility regime at 4-8 hours (wavelet scale 2) is 0.25. Similarly, if a high volatility regime prevails at a 2-4 week time scale (wavelet scale 6), the probability of being in a high volatility regime at 1-2 weeks (wavelet scale 5) is just 0.17. The conditional probabilities for the scale dependent transition matrix of the stock market volatility

indicates that low volatility stability is strong in the stock market. Contrary to the FX market, the probability of changing states from a high to a low state is much higher in the stock market, especially at higher scales. For example, given a high volatility state at a 2-4 week time scale (wavelet scale 6), the probability of observing a low volatility state at 1-2 weeks (wavelet scale 5) is 0.82. This indicates that switching to a low volatility state is more likely across scales in the stock market as compared to FX market(s).

Figure 13 exhibits the estimates of the sequence of states in the stock market, with low volatility (light rectangles) or high volatility (dark rectangles), from the wavelet HMT model using the sample period from May 4, 2000 to October 16, 2002. Compared to the FX market volatility states across scales, dark rectangles (high volatility states) are much narrower, indicating that the low volatility states are more likely in the stock market. Also, the relative discontinuity of the dark rectangles towards lower time scales implies that the probability switching from high to low volatility states is relatively higher in the stock market as compared to the FX market.

5.4 DJIA Volatility on and around September 11, 2001

Following a terrorist attack at the World Trade Center, trading was suspended at the NYSE from September 11, 2001 to September 14, 2001. Stock prices went down sharply during the first trading hour and afterwards when the market was re-opened. A close inspection of the stock market volatility and the most likely sequence of volatility states reveals that the effect of September 11 on the stock market was a relatively short one. In fact, Figure 13 shows that there were several volatility bursts, similar to the one experienced after September 11, during the sample period. Notice that a high volatility state across almost all scales has been in effect from mid-2002 to the end of the sampling period (October 16).

6 Conclusions

This paper establishes an important stylized property about volatility – asymmetric volatility dependence across different time horizons. We call this property *asymmetric vertical dependence*. It is asymmetric in the sense that low volatility at a long time horizon is most likely followed by low volatility at shorter time horizons. On the other hand, high volatility at long time horizons does not necessarily imply high volatility at shorter time horizons.

Our analysis provides evidence that volatility may be modeled as a mixture of high and low volatility regimes, resulting in a mixture distribution that is non-Gaussian. This result has important implications regarding the scaling behavior of volatility, and also on the calculation of risk at different time horizons. Our approach, using wavelet-based hidden Markov models for time series analysis, shows great promise in succinctly

capturing the non-Gaussian features found in high-frequency FX rates by simultaneously modeling multiple time horizons. Although we focused our attention on the wavelet hidden Markov tree (HMT) model, another possible model is to allow dependence within *and* between scales through hidden Markov chains and hidden Markov trees. In addition to taking advantage of the persistence across scales, this would allow us to explicitly model volatility clustering in time.

References

- Andersen, T. G., T. Bollerslev, F. X. Diebold, and P. Labys (2001). The distribution of realized exchange rate volatility. *Journal of the American Statistical Association*, 96, 42–55.
- Baum, L. (1972). An inequality and associated maximization technique in statistical estimation of probabilistic functions of Markov processes. *Inequalities*, 3, 1–8.
- Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31, 307–327.
- Crouse, M. S., R. D. Nowak, and R. G. Baraniuk (1998). Wavelet-based statistical signal processing using hidden Markov models. *IEEE Transactions on Signal Processing*, 46, 886–902.
- Dacorogna, M., R. Gençay, U. Müller, O. Pictet, and R. Olsen (2001). *An Introduction to High-Frequency Finance*. San Diego: Academic Press.
- Durand, J.-B. and P. Gonçalves (2001). Statistical inference for hidden Markov tree models and application to wavelet trees. Technical Report 4248, Statistical Inference for Industry and Health, INRIA.
- Engle, R. F. (1982). Autoregressive conditional heteroskedasticity with estimates of the variance of U. K. inflation. *Econometrica*, 50, 987–1008.
- Fung, W. and D. A. Hsieh (2000). Measuring the market impact of hedge funds. *Journal of Empirical Finance*, 7, 1–36.
- Gençay, R., F. Selçuk, and B. Whitcher (2001a). Differentiating intraday seasonalities through wavelet multi-scaling. *Physica A*, 289, 543–556.
- Gençay, R., F. Selçuk, and B. Whitcher (2001b). *An Introduction to Wavelets and Other Filtering Methods in Finance and Economics*. San Diego: Academic Press.
- Gençay, R., F. Selçuk, and B. Whitcher (2001c). Scaling properties of foreign exchange volatility. *Physica A*, 289, 89–106.
- Hamilton, J. D. (1989). A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica* 57, 357–384.
- Maheu, J. M. and T. H. McCurdy (2002). Nonlinear features of realized FX volatility. *Review of Economics and Statistics* 84, 1–24.
- Mallat, S. (1998). *A Wavelet Tour of Signal Processing*. San Diego: Academic Press.
- Mandelbrot, B. B. (1963). The variation of certain speculative prices. *Journal of Business*, 36, 394–419.
- Mandelbrot, B. B. (1971). When can price be arbitrated efficiently? A limit to the validity of the random walk and martingale model. *Review of Economics and Statistics*, 53, 225–236.

- Müller, U. A., Dacorogna, M. M., Davé, R. D., Olsen, R. B., Pictet, O. V., and von Weizsäcker, J. E. (1997). Volatilities of different time resolutions – analyzing the dynamics of market components. *Journal of Empirical Finance*, 4, 213–239.
- Percival, D. B. and A. T. Walden (2000). *Wavelet Methods for Time Series Analysis*. Cambridge, England: Cambridge University Press.
- Rabiner, L. R. (1989). A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77, 257–286.
- Ramsey, J. B. (2002). Wavelets in economics and finance: Past and future. *Studies in Nonlinear Dynamics and Economics*, 6, Article 1.
- Smyth, P., D. Heckerman, and M. I. Jordan (1996). Probabilistic independence networks for hidden Markov probability models. *Neural Computation* 9, 227–269.

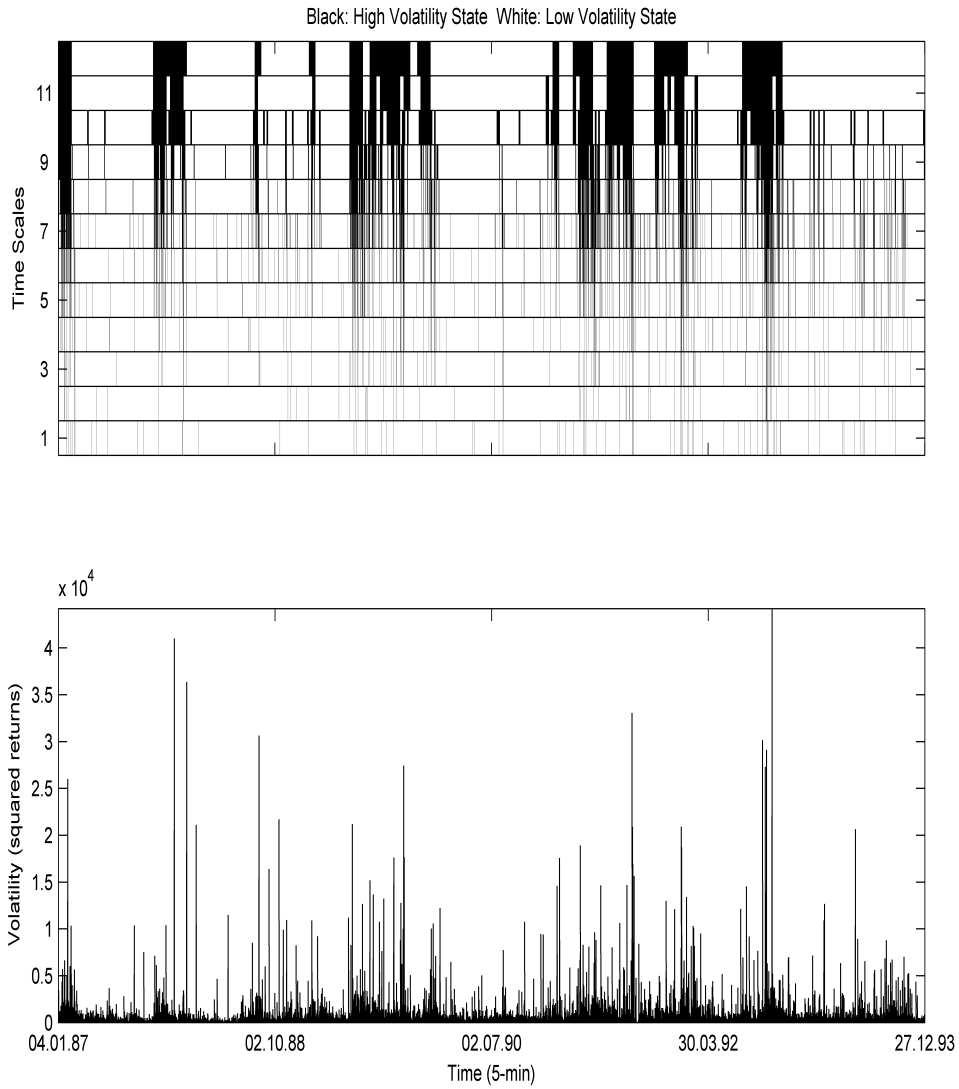


Figure 11: Most likely sequence of states from the fitted wavelet hidden Markov tree (HMT) model for the USD-DEM volatility. The first state $S = 0$ (light rectangles) indicates a low volatility (state) regime while the second state $S = 1$ (dark rectangles) indicates a high volatility regime.

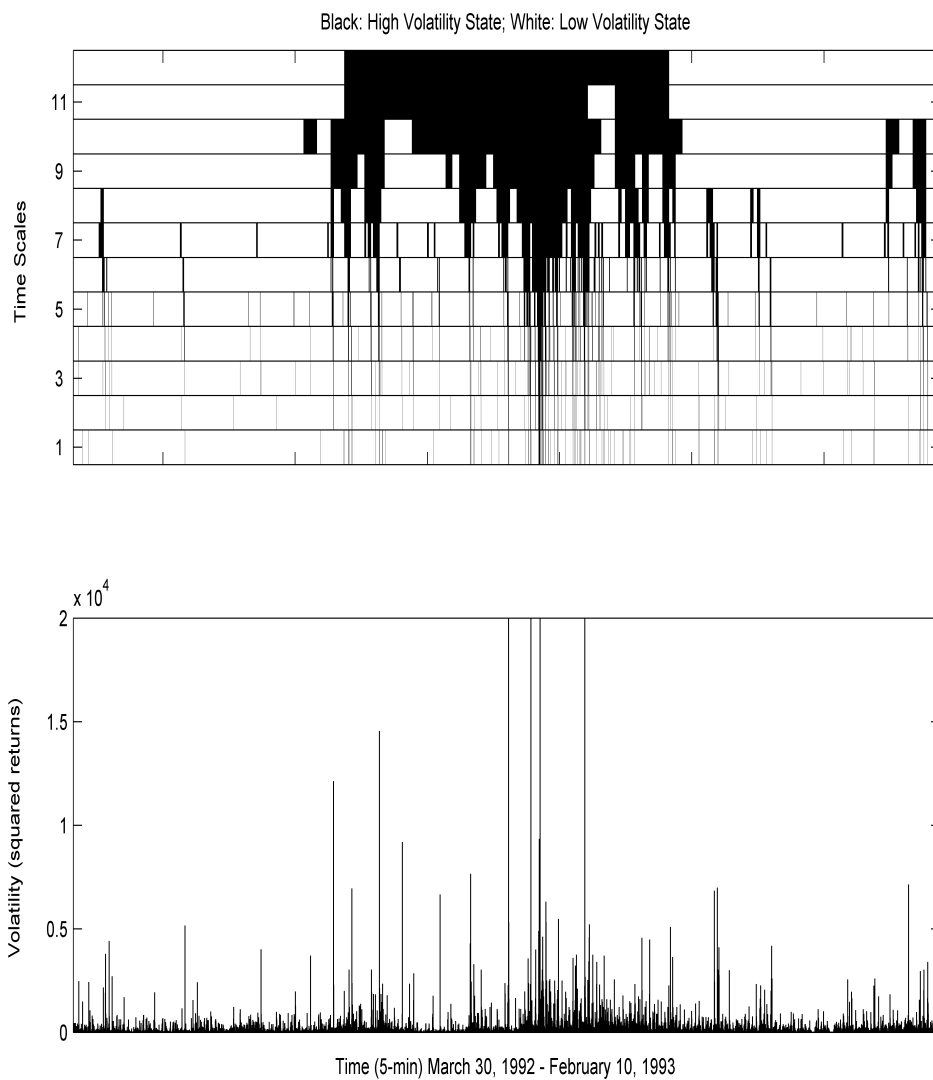


Figure 12: Most likely sequence of states from the fitted wavelet hidden Markov tree (HMT) model from March 1992 to February 1993 for the USD-DEM volatility. The first state $S = 0$ (light rectangles) indicates a low volatility (state) regime while the second state $S = 1$ (dark rectangles) indicates a high volatility regime.

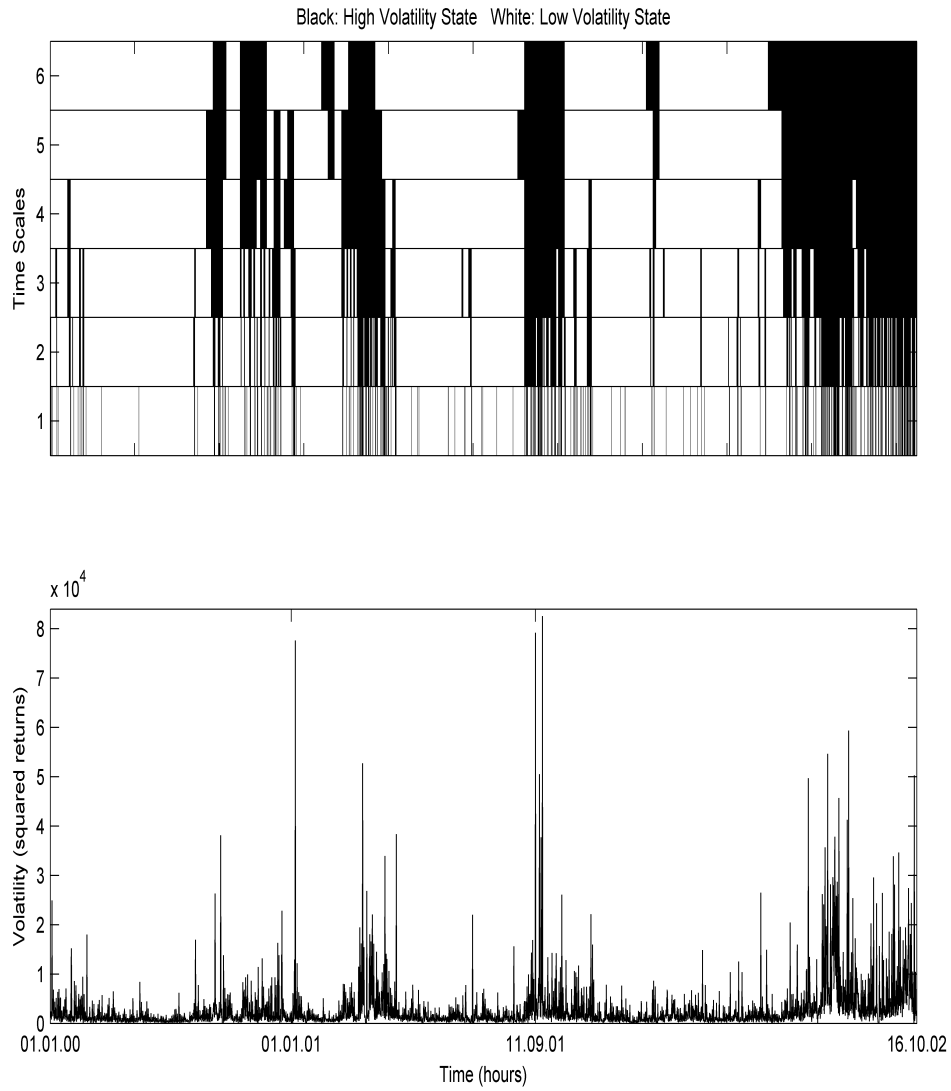


Figure 13: Most likely sequence of states from the fitted wavelet hidden Markov tree (HMT) model for the DJIA volatility. The first state $S = 0$ (light rectangles) indicates a low volatility (state) regime while the second state $S = 1$ (dark rectangles) indicates a high volatility regime.