A Multiresolution Analysis of Stock Market Volatility Using Wavelet Methodology

DRAFT OF LICENTIATE THESIS

"SHORT" VERSION!*

Tommi A. Vuorenmaa†
Bank of Finland (Research Department)
Research Unit on Economic Structures and Growth
Department of Economics, University of Helsinki

Previous version: May 13, 2004
This version: May 27, 2004

Abstract

The non-stationary character of stock market returns manifests itself through the volatility clustering effect and large jumps. An efficient way of representing a time series with such complex dynamics is given by wavelet methodology. With the help of a wavelet basis, the discrete wavelet transform (DWT) is able to break a time series with respect to a time-scale while preserving the time dimension and energy. Time-scale specific information is important if one accepts the view that stock market consists of heterogenous investors operating at different time-scales. Considerable more insight into the volatility dynamics is gained by looking at the data at several different time-scales. At small time-scales, in particular, the locality of wavelet analysis allows one to fully exploit high-frequency data. In addition, the DWT is even faster than the fast Fourier transform, so it is ideally suited for analyzing large data sets. The "large-scale aim" of this paper is to first introduce wavelet methodology and then to analyze high-frequency stock market volatility with it. In more detail, the data consists of 5-minute observations of Nokia Oyj at the Helsinki Stock Exchange

*The main results of the empirical section of this particular version are to be presented at the *"Economics and Econometrics of the Market Microstructure Summer School"* (June 7–11 2004; Constance, Germany). To save space and to make the paper a little bit more focused, Sections 2, 3, and 4 have been excluded (they contain wavelet theory). The full draft is available from the author upon request. The thesis is expected to be ready in August, 2004.

†Comments are welcome! Email: tvuorenm@cc.helsinki.fi

(HEX) spanning 4 years (1999-2002). Several microstructure problems have to be dealt with, some characteristic of the HEX. Pre-filtered data is then being analyzed by the "maximal overlap" DWT to study both the global and local scaling laws in turbulent and calm periods, separately. In particular, this gives local long-memory estimates of a stochastic volatility model with time-varying parameters.

1 Introduction

Financial time series share common characteristics. For example, in stock market return data one typically observes *discontinuities*, i.e. sudden big changes, and *clusters of volatility*, i.e. alternating of highly volatile and tranquil periods (see Fig. 1). These, and many other, phenomena are so widely recognized that they are called *stylized facts* (see e.g. Cont (2001) or Dacorogna *et al.* (2001)). Clearly a good model of financial returns would at least have to capture the non-Gaussian behavior and time-varying volatility. Most notably, the elegant Gaussian random walk framework of Bachelier (1900) suffers seriously from the neglection of these empirical regularities.

A lot of effort has been put into the study of discontinuities since the pioneering work of Mandelbrot (1963) and Fama (1965). In fact, back in the 1970s the study of non-Gaussian (especially Paretian) heavy-tailed distributions dominated the empirical finance literature. Although there is nowadays a vital new line of research of jumps (see e.g. Ait-Sahalia (2003), Barndorff-Nielsen and Shephard (2003a, 2003b), Andersen *et al.* (2003), and Huang and Tauchen (2003)), for the last two decades the main emphasis has been put into the research of volatility clustering phenomena or what has become to known as the "ARCH-effect". In particular, the seminal articles of Engle (1982) and Bollerslev (1986) launched a huge interest in different kinds of *(generalized) autoregressive conditional heteroskedastic* ((G)ARCH) models.[1] (For a review of these models, see Bollerslev *et al.* (1992) or Bollerslev *et al.* (1994).) In addition, an ever more important complementary class of models is the class of *stochastic volatility* models (see e.g. Ghysels *et al.* (1995)) which are able to produce the ARCH-effect, too. The immense interest in the conditional variance stems from the fact that a correctly specified volatility model is important in valuation of stocks and stock options and in designing optimal dynamic hedging strategies for options and futures, among other reasons (see e.g. Engle and Ng (1993)).

But as important as these models and their numerous extensions have been for the newborn field of *financial econometrics* (for an essay, see Bollerslev (2001)), they have not helped much in *explaining* the stylized facts. True enough, they are only *models*, and as such perhaps only meant for succesful data fitting, but are they missing on something crucial? In a way they are, since they are modeling only one time-scale

[1] In fact, in 2003 Engle was given a (half of) Nobel Prize in Economic Sciencies "for methods of analyzing economic time series with time-varying volatility (ARCH)" (see the "Advanced Information" at Http://www.nobel.se/economics/laureates/2003).

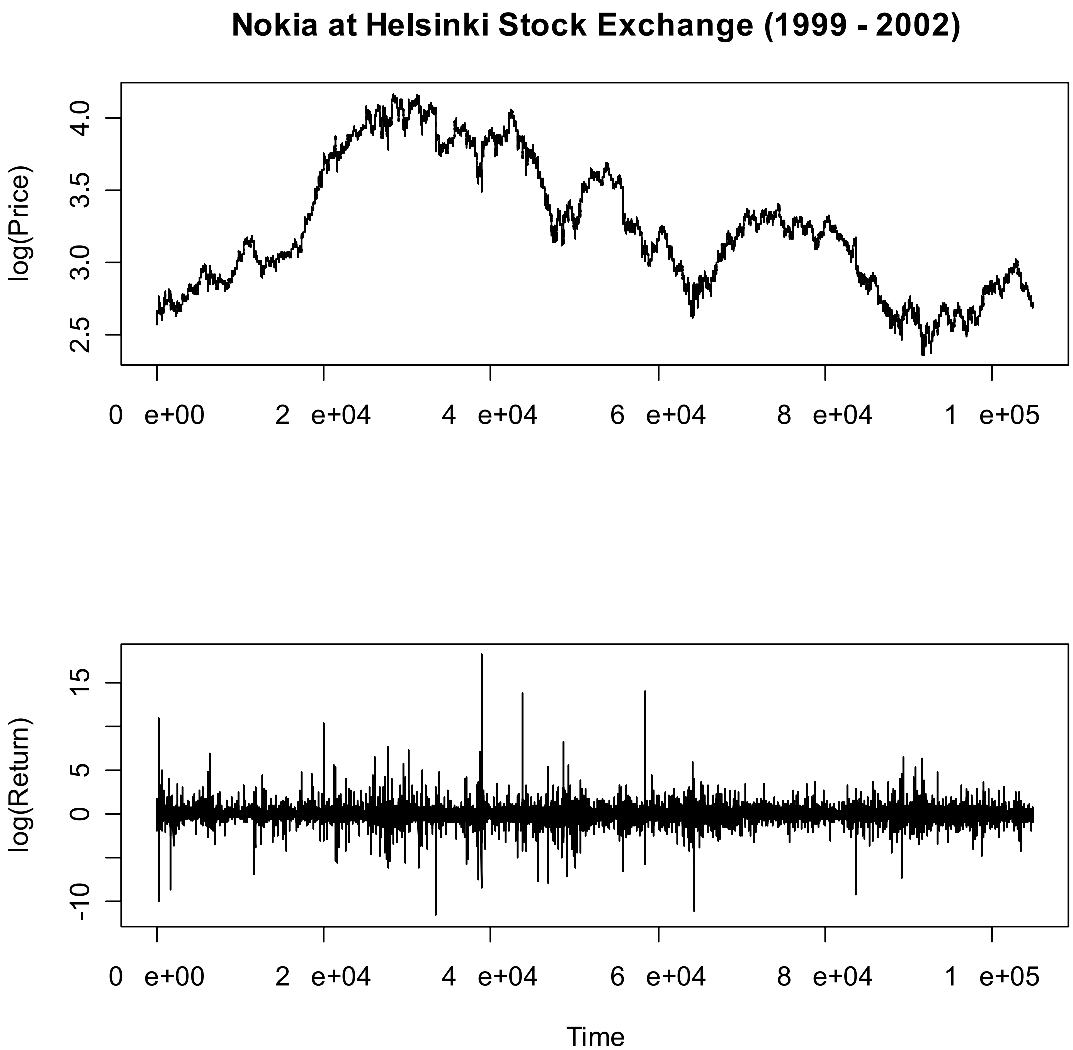


Figure 1: Logarithmic price and return history of Nokia sampled every 5-minutes from 1999 to 2002. Notice the jumps and the clusters of volatility.

(usually a day, or larger). But stock market data has no specific time-scale to analyze! A notable exception in this respect is the *heterogenous* ARCH model introduced in Müller *et al.* (1997) based on the hypothesis of a heterogenous market (see Müller *et al.* (1993), Peters (1994)). According to this hypothesis the stock market consists of *multiple layers* of investment horizons (time-scales) varying from an extremely short (minutes) to long (years). The small time-scales are commonly thought to be related to speculative activity and the bigger time-scales to investment activity. Indeed, the players in the stock market form a heterogenous group with respect to other reasons as well, such as perceptions of the market, risk profiles, institutional constraints, degree of information, prior beliefs, and other characteristics such as geographical locations. Interestingly though, Müller *et al.* (1993) argue that many of these differences translate to sensitivity to different time-scales. Indeed, Müller *et al.* (1997) show some support for the view that time-scale is *"one of the most important aspects in which trading behaviours differ"* (further support is given by Lynch and Zumbach (2003)). For example, big institutional investors have relatively long trading horizons and they trade on economic fundamentals. On the other hand, some of the market participants, the so-called day-traders, don't keep open positions over night and they simply trade on market sentiment. These small time-scales, in particular, have become increasingly important because of the recent availability of *high-frequency* data. In fact, Goodhart and O'Hara (1997) conjecture that *"the ability to analyze higher frequency data may be particularly useful in pursuing* [why volatility persistance endures]*"*. It is therefore obvious that to better capture the dynamics of a stock market one must analyze data at multiple, perhaps even a continuum of, time-scales.

Multiple time-scales are especially important from the view point of risk management. For example, in the risk management industry one often needs to *scale* a risk measure (e.g. standard deviation) of one time-scale to another. The industry standard is to scale by the square-root of time, familiar from Brownian motion (i.e. continuous-time random walk). But one is then implicitly assuming that the *data generating process* (DGP) is made of *identically and independently distributed* (IID) random variables. This assumption is *not* reasonable for financial time series; just consider volatility clustering, for example. The existance of serial correlation in the *conditional* second moments is usually obvious even by eye (as in Fig. 1). In fact, the persistance is universally found so strong and long-lasting that volatility is said to exhibit *long-memory* (or long-range dependence). Under such non-IID circumstances square root scaling may indeed lead to wrong conlusions (see Diebold *et al.* (1997)). Interestingly, scaling laws have been mainly studied with foreign exchange rate data only (e.g., Andersen *et al.* (2000) and Gençay *et al.* (2001)). This is probably because of its larger turnover, higher liquidity, and lower transaction costs compared to stock markets.

Another commonly used and overly simplifying assumption is *stationarity*, often of second-order. Most of the parametric GARCH models belong to this group, for instance. In particular, *spectral analysis* requires "covariance stationarity". Spectral analysis is a powerful non-parametric method that allows one to *represent* a stationary time series in *frequency domain (* in which the frequency aspects of data can be easily studied). This view point should be contrasted with *time domain* presentation – the customary way of presenting stock market data – which hides the frequency information. In the context of spectral analysis, in particular, the requirement of stationarity stems from the fact that a power spectrum is just a Fourier transform of the corresponding sample autocorrelation function.[2] As is well-known, autocorrelation is incapable of detecting non-stationarities. And because Fourier transform is looking only for sinusoids "globally", spectral analysis is *not* suitable for transient and evolving behavior of stock markets. Clean sinusoid components are rarely (never, dare to say!) encountered in *empirical finance*.

Clearly, therefore, one needs to use more flexible tools than the traditional ones to study stock market data. Considering the arguments made this far, the most straightforward way of increasing flexibility would be to use a non-parametric multiscale approach. This is exactly where *wavelet analysis* enters the picture. Wavelet analysis offers a non-parametric, mathematically concise way of studying the heterogeneity of stock markets. But what is wavelet analysis exactly and how is it able to provide a time-scale perspective on a complex (deterministic) function or, in the case of time series, on a non-stationary realization of a stochastic process? Loosely speaking, wavelet methodology extends Fourier methodology (i.e. spectral methods) by replacing frequency by time-scale while still preserving time dimension. As the reader might suspect at this point, time-scale and frequency have an intimate relationship. Indeed, if the data is stationary, then time-scale can be regarded as the reciprocal of frequency (see Priestley (1996)): when frequency is high (low), then time-scale is small (large). So in this way wavelet and spectral methods are complementary tools, just differing in angle. "Rivals" they become as soon as one requires *local* instead of global analysis. And in fact, local adaptiviness is the best asset of wavelet analysis. This stems from the mathematical fact that the basis functions used in wavelet analysis, called wavelets, are well-localized in *both* time and scale. This gives wavelets a distinct advantage over standard frequency domain methods when analyzing complex dynamics, such as the one found in stock markets.

There are a wide variety of wavelets available today. However, in the empirical analysis (Sec. 7) only wavelets belonging to the family of *Daubechies* wavelets are applied, mostly because of their compact support. In particular, the *least asymmetric* wavelets are used to analyze the volatility of Nokia Oyj at Helsinki Stock Exchange. The results are interesting in several respects. First, the multiresolution analysis conducted offers qualititative insight into the volatility dynamics. Next, the analysis

[2]This is the famous Wiener–Khintchine Theorem (see e.g. Priestley (1981)). In certain cases one can also analyze non-stationary series with Fourier methods. Furthermore, it has to be noted that Fourier transform *can* be time-localized to a certain degree. This is done by the so-called windowed Fourier transform (see Sec. 3.1.2).

reveals interesting changes in global and local scaling laws which gives information on the behavior of different type of investors in time. Quantitatively, the semiparametric wavelet approach allows the estimation of a stochastic volatility model with a time-varying long-memory parameter. This approach is ideally suited in the high-frequency context because of the existing complex volatility dependencies and large number of observations. And contrary to wide held beliefs, high-frequency data is very informative about longer-run volatility dependencies (see Andersen and Bollerslev (1997b)). The analysis also reveals the consequences of microstructure effects (especially the intraday periodicity in volatility) to long-memory estimation. Overall, the results clearly suggest the presence of long-memory volatility dependencies. Indeed, the present paper is about more than just *"deployment of wavelet technique to financial data"* of which Norsworthy *et al.* (2000), and perhaps justifiably so, have criticized some early authors of.

Before turning to the results of the empirical analysis, however, it is essential to go through the basics of classical Fourier theory (Sec. **??**) as well as wavelet theory (Sec. **??**). These two subsections serve as the backbone to the idea and theory of multiresolution analysis, discussed next in somewhat technical manner (Sec. 4). Only after then, I believe, one is prepared well enough to handle stochasticity (Sec. 5). There is also a quick overview of volatility measures and modeling (Sec. 6) before the actual analysis. I begin this rather long tour by taking a quick glance at the history of wavelet methodology, a form of *atomic decomposition*.

2 A historical glance at wavelet methodology

Sections 2, 3, and 4 have been excluded from this version.

3 Essentials of Fourier and wavelet theories

The purpose of this section is to give the necessary technical background for the wavelet multiresolution analysis (to be defined mathematically in Sec. 4). It is suggested that also those familiar with the basics of Fourier and wavelet theories would at least skim through this section to be acquintant with the notation used.

4 Multiresolution analysis

This section explains, mathematically, how wavelet methodology is used to decompose a *deterministic* finite-energy function with respect to resolution (time-scale). The key concept of *multiresolution analysis* (due to Mallat (1988)) is introduced. Importantly, wavelets are shown to act as linear *filters*. Moreover, the construction of *compactly supported* wavelets belonging to the family of Daubechies is discussed.

5 Decomposing time series

This section aims to show how multiresolution analysis is done in a time series context. More specifically, in this section the function-to-analyzed is not deterministic but instead *a realization of a stochastic process*. Special attention is paid to the energy preservation property that allows the decomposition of the total variance into time-scale specific *wavelet variances*. At the moment, at least, the most comprehensive coverage of wavelet analysis in statistical time series context is provided by Percival and Walden (2000). A nice complementary review article is Nason and von Sachs (1999).

5.1 Practical issues

In order to come up with a useful wavelet analysis of a time series, one must take into account several issues. The most important ones are:

1. the choice of a wavelet filter;
2. handling boundary conditions;
3. sample sizes that are not a power of two.

These three points are next being shortly discussed in turn (for details, see Percival and Walden (2000, Ch. 4.11)).

Choice of a wavelet filter. The problem with the smallest width wavelet filters is that they can sometimes introduce undesirable artifacts into the resulting analysis, such as unrealistic blocks, "sharks' fins", etc. The wider width wavelet filters indeed can better match to the characteristic features in a time series. Unfortunately though, as the width gets wider, (i) more coefficients are being unduly influenced by boundary conditions, (ii) there is some decrease in the degree of localization of the DWT coefficients, and (iii) there is an increase in computational burden. Thus one should search for the smallest L that gives reasonable results. In practice, if one also wants to have the DWT coefficients be alignable in time, the optimal choice is often LA(8) (this is the filter used in Sec. 7, in fact).

Handling boundary conditions. The DWT uses circular filtering which means that the time series is treated as a portion of a periodic sequence with period N. For financial time series this is problematic since there is rarely evidence to support this assumption. Furthermore, there may be a large discontinuity between the last and first observations. The extent that circularity influences the DWT coefficients and corresponding MRA is quantified in Percival and Walden (2000, pp. 145–9). Notice that the Haar wavelet yields coefficients that are free of the circularity assumption.

As a minimal way of dealing with the circularity assumption, Percival and Walden suggest to exactly indicate on plots the DWT coefficients and MRAs that are affected by the boundary. However, they continue to argue that the influence of circularity can be quite small, particularly when the discrepancy between the beginning and end of the series is not too large. Thus the marked regions are usually quite conservative measures of the influence of circularity. One way of reducing the impact of circularity is to *reflect* the time series about its end point. The resulting series of length $2N$ has the same mean and variance as the original series. This method eliminates the effects due a serious mismatch between the first and last values. The cost is increased, but "quite acceptable", computational burden. There are also other ways to cope with circularity like polynomial extrapolation at both ends of the time series and specially designed "boundary wavelets" that are zero outside the range of the data (for details, see e.g. Bruce and Gao (1996)).

Handling sample sizes that are not a power of two. The "full" DWT requires N to be a power of two and the "partial" DWT requires N to be an integer multiple of 2^{J_0}. In reality, however, it rarely happens that the data at hand is of dyadic length or even an integer multiple of it. There are some *ad hoc* methods for dealing with this problem. The most obvious one is to truncate the series to the closest integer multiple of 2^{J_0}. One also needs to consider what choice J_0 is reasonable. An easy alternative – familiar from Fourier analysis – is to "pad" the series with zeros or the sample mean. Note that padding with the sample mean does not change the sample mean from the original series. One could also pad by replicating a data value (typically the last one). In fact, Ogden (1996) has compared various ways of preconditioning data not meeting the criteria of power of two. He found no method to be clearly superior in every respect but dictated by the particular application of interest. As one might suspect, though, extending the data by padding is the easiest to implement and the least computationally expensive too. Ogden points out that the wavelet coefficients resulting from preconditioned data should be used cautiously, though. For example, padding by repeating the last value will introduce a flat artifact towards the end of the interval causing new problems that remain even with very large samples.

In the next subsections, I will consider only discrete wavelet transforms for two reasons: (i) financial data is inherently discrete, and (ii) discrete transforms are computationally less demanding than continuous ones. This is the standard way in time series applications.

5.2 Discrete wavelet transform

For a throughout description of the DWT, see Percival and Walden (2000, Sec. 4). For a quick and "dirty" treatment, see Gençay *et al.* (2002a, Sec. 4.4). My presentation parallels with the latter because of space limitations.

Construction. An easy way to introduce the DWT is through a matrix operation. Consider a dyadic length (i.e., $N = 2^J$) column vector of observations $\mathbf{x}$. The length N column vector of discrete wavelet coefficients $\mathbf{w}$ is obtained via

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

where $\mathcal{W} \in M(N \times N)$ is an *orthonormal* matrix defining the DWT (see App. B). The matrix $\mathcal{W}$ is composed of the wavelet and scaling filter coefficients arranged on a row-by-row basis. The structure of the resulting $\mathbf{w}$ and the matrix $\mathcal{W}$ may be seen through the subvectors $\mathbf{w}_1, \mathbf{w}_2, ..., \mathbf{w}_J, \mathbf{v}_J$ and submatrices $\mathcal{W}_1, ..., \mathcal{W}_J, \mathcal{V}_J$, respectively:

$$\mathbf{w} = \begin{pmatrix} \mathbf{w}_1 \\ \mathbf{w}_2 \\ \vdots \\ \mathbf{w}_J \\ \mathbf{v}_J \end{pmatrix} \text{ and } \mathcal{W} = \begin{bmatrix} \mathcal{W}_1 \\ \mathcal{W}_2 \\ \vdots \\ \mathcal{W}_J \\ \mathcal{V}_J \end{bmatrix},$$

where $\mathbf{w}_j$ is a length $N/2^j$ column vector of wavelet coefficients associated with changes on a scale of length $\lambda_j = 2^{j-1}$ and $\mathbf{v}_J$ is a length $N/2^J$ column vector of scaling coefficients associated with averages on a scale of length $2^J = 2\lambda_J$. Similarly, $\mathcal{W}_j \in M(N/2^j \times N)$ and $\mathcal{V}_J \in M(N/2^J \times N)$.

As an illustration of the structure of the matrix $\mathcal{W}$, consider a filter of length $L = 2$ and a signal of length $N = 8$. Then the matrix $\mathcal{W}_1 \in M(4 \times 8)$ is

$$\mathcal{W}_1 = \begin{bmatrix} h_1 & h_0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & h_1 & h_0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & h_1 & h_0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & h_1 & h_0 \end{bmatrix} = \begin{bmatrix} \mathbf{h}_1^{(2)} \\ \mathbf{h}_1^{(4)} \\ \mathbf{h}_1^{(6)} \\ \mathbf{h}_1 \end{bmatrix},$$

where $\mathbf{h}_1^{(k)}$, $k \in \{2, 4, 6\}$, is the vector of zero-padded unit scale wavelet filter coefficients in reverse order, and which is circularly shifted to the right by k.

Similarly, by letting $\mathbf{h}_2$ and $\mathbf{h}_4$ denote the vector of zero-padded scale two and four wavelet filter coefficients, respectively, one can construct the matrix $\mathcal{W}_2 \in M(2 \times 8)$ and the row vector $\mathcal{W}_3 \in M(1 \times 8)$. In this case, the circular shift is by factors of four and eight (i.e., no change), respectively:

$$\mathcal{W}_2 = \begin{bmatrix} \mathbf{h}_2^{(4)} \\ \mathbf{h}_2 \end{bmatrix} \text{ and } \mathcal{W}_3 = \mathbf{h}_3.$$

Finally, the matrix $\mathcal{V}_3 \in M(1 \times 8)$ is a row vector whose elements are all equal to $1/\sqrt{N}$ (Gençay *et al.* (2002a, p. 120)).

Of course, one must be able to explicitly compute the wavelet filter coefficients for level $j = 1, ..., J$ to complete the construction of the matrix $\mathcal{W}$. Given the FRFs of the unit scale wavelet and scaling filters, it is possible to recover the wavelet filter $h_{j,l}$ for scale $\lambda_j = 2^{j-1}$ by the inverse DFT of

$$H_{j,k} = H_{1,2^{j-1}k \bmod N} \prod_{l=0}^{j-2} G_{1,2^l k \bmod N}, \text{ for } k = 0, ..., N-1.$$

The length of the resulting wavelet filter is $L_j = (2^j - 1)(L - 1) + 1$. Similarly, one can recover the scaling filter g_J for scale λ_J by the inverse DFT of

$$G_{J,k} = \prod_{l=0}^{J-1} G_{1,2^l k \bmod N}, \text{ for } k = 0, ..., N-1.$$

(Gençay *et al.* (2002a, p. 121).)

Example 1 (Haar) *In the case of $L = 2$ and $N = 8$, the matrix $\mathcal{W}$ is*

$$\mathcal{W} = \begin{bmatrix} \mathcal{W}_1 \\ \mathcal{W}_2 \\ \mathcal{W}_3 \\ \mathcal{V}_3 \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{2} & \frac{1}{2} & -\frac{1}{2} & -\frac{1}{2} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{2} & \frac{1}{2} & -\frac{1}{2} & -\frac{1}{2} \\ \frac{1}{\sqrt{8}} & \frac{1}{\sqrt{8}} & \frac{1}{\sqrt{8}} & \frac{1}{\sqrt{8}} & -\frac{1}{\sqrt{8}} & -\frac{1}{\sqrt{8}} & -\frac{1}{\sqrt{8}} & -\frac{1}{\sqrt{8}} \\ \frac{1}{\sqrt{8}} & \frac{1}{\sqrt{8}} & \frac{1}{\sqrt{8}} & \frac{1}{\sqrt{8}} & \frac{1}{\sqrt{8}} & \frac{1}{\sqrt{8}} & \frac{1}{\sqrt{8}} & \frac{1}{\sqrt{8}} \end{bmatrix}.$$

Boundary effects. It is important to know which wavelet coefficients have been computed using observations across the boundary (cf. Sec. 5.1). Clearly the number of affected coefficients grow as the level j and length L grow. More precisely, Percival and Walden (2000, Ch. 4.11) show that the number of the affected DWT coefficients is

$$L'_j = \left\lceil (L-2)\left(1 - 2^{-j}\right) \right\rceil,$$

where $\lceil x \rceil$ is the smallest integer greater than or equal to x. Thus, for display purposes, it is possible to approximately line up the DWT coefficient vectors with the original time series by circularly shifting the level j vector of DWT coefficients appropriately. Percival and Walden (2000, Ch. 4.11) give a precise table of integer shifts for the least asymmetric wavelet filter. The following heuristic can be used, however: If the number of coefficients affected by the boundary is even, then place half of the of the boundary coefficients at each end of the series. If, on the other hand, the number of coefficients affected by the boundary is odd, then place the "extra" coefficient at the beginning of the series. Notice that shifting coefficients from the extremal phase wavelet filter is not as straigthforward given its poor phase properties (see Sec. **??**). (Gençay *et al.* (2002a, p. 145.)

Multiresolution analysis. An additive decomposition of a time series can be obtained using the DWT by first defining the jth level WAVELET DETAIL

$$\mathbf{d}_j \doteq \mathcal{W}_j^T \mathbf{w}_j, \text{ for } j = 1, ..., J,$$

which is associated with changes in $\mathbf{x}$ at scale λ_j. The wavelet coefficients $\mathbf{w}_j = \mathcal{W}_j \mathbf{x}$ represent the portion of the wavelet analysis attributable to scale λ_j. Thus $\mathcal{W}_j^T \mathbf{w}_j$

is the portion of the wavelet synthesis attributable to scale λ_j. For a dyadic length $N = 2^J$ time series, the final wavelet detail $\mathbf{d}_{J+1} = \mathcal{V}_J^T \mathbf{v}_J$ is equal to the sample mean of observations. On the other hand, define the jth level WAVELET SMOOTH as

$$\mathbf{s}_j \doteq \sum_{k=j+1}^{J+1} \mathbf{d}_k \text{ for } 0 \le j \le J,$$

where $\mathbf{s}_{J+1}$ is defined to be a vector of zeros. In contrast to the wavelet detail $\mathbf{d}_j$ which is associated with variations at a particular scale, the wavelet smooth $\mathbf{s}_j$ becomes smoother as more details are summed. Indeed, it holds that $\mathbf{x} - \mathbf{s}_j = \sum_{k=1}^{j} \mathbf{d}_k$. The jth level WAVELET ROUGH characterizes the remaining lower-scale details through

$$\mathbf{r}_j \doteq \sum_{k=1}^{j} \mathbf{d}_k, \text{ for } 1 \le j \le J+1,$$

where $\mathbf{r}_0$ is defined to be a vector of zeros. Thus a time series $\mathbf{x}$ may be decomposed as

$$\mathbf{x} = \mathbf{r}_j + \mathbf{s}_j = \sum_{k=1}^{j} \mathbf{d}_k + \sum_{k=j+1}^{J+1} \mathbf{d}_k = \sum_{k=1}^{J+1} \mathbf{d}_k,$$

where $\mathbf{d}_{J+1}$ is the sample mean of observations.

Variance decomposition. One of the most important properties of the DWT is to decompose the sample variance of a time series on a scale-by-scale basis. This is possible because the DWT is an energy (variance) preserving transform:

$$\|\mathbf{x}\|^2 = \mathbf{x}^T\mathbf{x} = (\mathcal{W}\mathbf{w})^T \mathcal{W}\mathbf{w} = \mathbf{w}^T \mathcal{W}^T \mathcal{W}\mathbf{w} = \mathbf{w}^T\mathbf{w} = \|\mathbf{w}\|^2,$$

where $\mathcal{W}$ is an orthonormal matrix defining the DWT. In another words,

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

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

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

where $\|\mathbf{w}_j\|^2$ 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. Because $\mathcal{W}$ and $\mathcal{V}$ are orthonormal matrices, we have $\mathbf{d}_j^T\mathbf{d}_j = \mathbf{w}_j^T\mathbf{w}_j$ for $1 \le j \le J$ and $\mathbf{s}_J^T\mathbf{s}_J = \mathbf{v}_J^T\mathbf{v}_J$, so alternatively,

$$\|\mathbf{x}\|^2 = \sum_{j=1}^{J} \|\mathbf{d}_j\|^2 + \|\mathbf{s}_J\|^2.$$

The theoretical counterpart of the variance decomposition in the context of stationary long-memory processes will be discussed later (Sec. 5.5).

5.3 Partial DWT

For a bit closer look at the partial DWT than presented below, see either Percival and Walden (2000, Sec. 4.7) or Gençay *et al.* (2002a, Sec. 4.4.2). I will be rather brief because the partial DWT is a straightforward generalization of the DWT.

Partial discrete wavelet transform (pDWT) offers more flexibility than the "full" DWT due the choice of a scale beyong which a wavelet analysis into individual large scales is no longer of real interest. A practical benefit of this method is that the sample size no longer needs to be a of dyadic length. Indeed, it is enough that the sample size to be a multiple of 2^{J_0}. Naturally, the choice of J_0 depends on the goals of the analysis.

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

$$\mathcal{W} = \begin{bmatrix} \mathcal{W}_1 \\ \mathcal{W}_2 \\ \vdots \\ \mathcal{W}_{J_p} \\ \mathcal{V}_{J_p} \end{bmatrix},$$

except that the matrix of scaling filter coefficients $\mathcal{V}_{J_p} \in M(N/2^{J_p} \times N)$ is a matrix of circularly shifted scaling coefficients vectors.

Multiresolution analysis. For a level $J_0 < J$ partial DWT, one has the MRA

$$\mathbf{x} = \sum_{j=1}^{J_0} \mathbf{d}_j + \mathbf{s}_{J_0},$$

where the details $\mathbf{d}_j$ are related to changes on a scale of $\lambda_j = 2^{j-1}$ and the smooth $\mathbf{s}_{J_0}$ to averages of a scale of $\lambda_{J_0} = 2^{J_0}$.

Variance decomposition. The energy decomposition is

$$\begin{aligned} \|\mathbf{x}\|^2 &= \sum_{j=1}^{J_0} \|\mathbf{w}_j\|^2 + \|\mathbf{v}_{J_0}\|^2 \\ &= \sum_{j=1}^{J_0} \|\mathbf{d}_j\|^2 + \|\mathbf{s}_{J_0}\|^2 . \end{aligned}$$

5.4 Maximal overlap DWT

For a throughout discussion of the maximal overlap DWT, see Percival and Walden (2000, Sec. 5). Once again, Gençay *et al.* (2002a, Sec. 4.5) give a compact introduction. I will mostly follow the latter because of the increased complexity of the transform. Notice that in practice a pyramid algorithm similar to that of the DWT is utilized (see Percival and Mojfeld (1997)). This algorithm requires $O(N \log_2 N)$

multiplications, so it is computationally a bit heavier to execute than the DWT, but still only as heavy as the FFT.

The increased complexity stems mostly from the fact that the MAXIMAL OVERLAP DISCRETE WAVELET TRANSFORM (MODWT) gives up orthogonality in order to gain features that the DWT (the full or partial) does not posses.[3] The following properties distinguish the MODWT from the DWT (Percival and Walden (2000, pp. 159–60)):

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 (to be defined in Sec. 5.5) estimator is asymptotically more efficient than the same estimator based on the DWT.

Construction. Let $\mathbf{x}$ be a length N column vector of observations. The length $(J+1)N$ column vector of MODWT coefficients $\widetilde{\mathbf{w}}$ is obtained via

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

where $\widetilde{\mathcal{W}} \in M((J+1)N \times N)$ is a *non-orthogonal* matrix defining the MODWT. The resulting $\widetilde{\mathbf{w}}$ and the matrix $\widetilde{\mathcal{W}}$ consist of column subvectors $\widetilde{\mathbf{w}}_1, ..., \widetilde{\mathbf{w}}_J, \widetilde{\mathbf{v}}_J$, each of length N, and submatrices $\widetilde{\mathcal{W}}_1, ..., \widetilde{\mathcal{W}}_J, \widetilde{\mathcal{V}}_J \in M(N \times N)$:

$$\widetilde{\mathbf{w}} = \begin{pmatrix} \widetilde{\mathbf{w}}_1 \\ \widetilde{\mathbf{w}}_2 \\ \vdots \\ \widetilde{\mathbf{w}}_J \\ \widetilde{\mathbf{v}}_J \end{pmatrix} \text{ and } \widetilde{\mathcal{W}} = \begin{bmatrix} \widetilde{\mathcal{W}}_1 \\ \widetilde{\mathcal{W}}_2 \\ \vdots \\ \widetilde{\mathcal{W}}_J \\ \widetilde{\mathcal{V}}_J \end{bmatrix},$$

where $\widetilde{\mathbf{w}}_j$ is associated with changes on a scale of length $\lambda_j = 2^{j-1}$ and $\widetilde{\mathbf{v}}_J$ is associated with averages on a scale of length $2^J = 2\lambda_J$. In the special case of a dyadic length time series, the MODWT may be subsampled and rescaled to obtain the DWT wavelet and scaling coefficients via

$$w_{j,t} = 2^{j/2}\widetilde{w}_{j,2^j(t+1)-1} \text{ and } v_{J,t} = 2^{J/2}\widetilde{v}_{J,2^J(t+1)-1}, \text{ for } t = 0, ..., N/2^j - 1.$$

[3] The MODWT is also known as the "stationary DWT", the "translation-invariant DWT" and the "time-invariant DWT".

In this dyadic case the DWT and MODWT filter coefficients are related in the following way: instead of using the wavelet and scaling filters from the previous section, the MODWT uses the rescaled filters

$$\widetilde{\mathbf{h}}_j = \frac{\mathbf{h}_j}{2^j} \text{ and } \widetilde{\mathbf{g}}_J = \frac{\mathbf{g}_J}{2^J}, \text{ for } j = 1, ..., J.$$

The construction of the submatrix $\widetilde{\mathcal{W}}_1$ is done by circular shifting the rescaled wavelet filter vector $\widetilde{\mathbf{h}}_1$ by integer units. The matrices $\widetilde{\mathcal{W}}_2$ and $\widetilde{\mathcal{W}}_3$ are formed similarly by replacing $\widetilde{\mathbf{h}}_1$ by $\widetilde{\mathbf{h}}_2$ and $\widetilde{\mathbf{h}}_3$. Thus revisiting the case of $L = 2$ and $N = 8$ of Example 1, one has

$$\widetilde{\mathcal{W}}_1 = \begin{bmatrix} \widetilde{h}_1 & 0 & 0 & 0 & 0 & 0 & 0 & \widetilde{h}_2 \\ \widetilde{h}_2 & \widetilde{h}_1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \widetilde{h}_2 & \widetilde{h}_1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \widetilde{h}_2 & \widetilde{h}_1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \widetilde{h}_2 & \widetilde{h}_1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \widetilde{h}_2 & \widetilde{h}_1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \widetilde{h}_2 & \widetilde{h}_1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \widetilde{h}_2 & \widetilde{h}_1 \end{bmatrix} = \begin{bmatrix} \widetilde{\mathbf{h}}_1^{(1)} \\ \widetilde{\mathbf{h}}_1^{(2)} \\ \widetilde{\mathbf{h}}_1^{(3)} \\ \widetilde{\mathbf{h}}_1^{(4)} \\ \widetilde{\mathbf{h}}_1^{(5)} \\ \widetilde{\mathbf{h}}_1^{(6)} \\ \widetilde{\mathbf{h}}_1^{(7)} \\ \widetilde{\mathbf{h}}_1 \end{bmatrix}.$$

Example 2 (Haar) *Insert* $\widetilde{h}_1 = -1/\sqrt{8}$ *and* $\widetilde{h}_2 = 1/\sqrt{8}$ *in the above matrix* $\widetilde{\mathcal{W}}_1$. *Unfortunately, the matrix* $\widetilde{\mathcal{W}} \in M(32 \times 8)$ *is too large to write down here, but it follows the same logic.*

For any positive integer J_0, the level J_0 MODWT of $\mathbf{x}$ is a transform consisting of the $J_0 + 1$ vectors $\widetilde{\mathbf{w}}_1, ..., \widetilde{\mathbf{w}}_{J_0}$ and $\widetilde{\mathbf{v}}_{J_0}$ which are all N-dimensional. The vector $\widetilde{\mathbf{w}}_j$ contains the MODWT wavelet coefficients associated with changes on scale $\lambda_j \doteq 2^{j-1}$, while $\widetilde{\mathbf{v}}_{J_0}$ contains the MODWT scaling coefficients associated with averages on scale $\lambda_{J_0} \doteq 2^{J_0}$.

Boundary effects. The MODWT uses integer translates of the wavelet and scaling filters, both of length $L_j = (2^j - 1)(L - 1) + 1$. This causes there to be a total of L_j wavelet coefficients affected by the boundary at each j. The time-alignment property of an MRA (Property 2) does not hold for the MODWT wavelet and scaling coefficients anymore without a proper adjustment (determined in McCoy *et al.* (1995)), however. This adjustment is different from the DWT (see Sec. 5.2). Precise integer shifts for the least asymmetric low-pass filter $g_{j,l}$ and high-pass filter $h_{j,l}$ are given by

$$\xi_j^g = \begin{cases} -\frac{(L_j-1)(L-2)}{2(L-1)} \text{ if } \frac{L}{2} \text{ is even} \\ -\frac{(L_j-1)L}{2(L-1)} \text{ if } L = 10 \text{ or } 18 \\ -\frac{(L_j-1)(L-4)}{2(L-1)} \text{ if } L = 14 \end{cases},$$

$$\xi_j^h = \begin{cases} -\frac{L_j}{2} \text{ if } \frac{L}{2} \text{ is even} \\ -\frac{L_j}{2} + 1 \text{ if } L = 10 \text{ or } 18 \\ -\frac{L_j}{2} - 1 \text{ if } L = 14 \end{cases},$$

respectively. (Gençay *et al.* (2002a, p. 145).)[4]

Multiresolution analysis. A MODWT MRA can be written as

$$x_t = \sum_{j=1}^{J+1} \widetilde{d}_{j,t} \text{ for } t = 0, ..., N-1,$$

where $\widetilde{d}_{j,t}$ is the tth element of the the jth level MODWT detail $\widetilde{\mathbf{d}}_j \doteq \widetilde{\mathcal{W}}_j^T \widetilde{\mathbf{w}}_j$, for $j = 1, ..., J$. The MODWT wavelet smooth and rough are, respectively,

$$\widetilde{s}_{J,t} = \sum_{k=j+1}^{J+1} \widetilde{d}_{k,t} \text{ and } \widetilde{r}_{j,t} = \sum_{k=1}^{j} \widetilde{d}_{k,t}, \text{ for } t = 0, ..., N-1.$$

Importantly, although MODWT is not an orthonormal transformation, the MRA

$$\mathbf{x} = \sum_{j=1}^{J_0} \widetilde{\mathbf{d}}_j + \widetilde{\mathbf{s}}_{J_0}, \tag{1}$$

where $\widetilde{\mathbf{s}}_{J_0} \doteq \mathcal{V}_{J_0}^T \widetilde{\mathbf{v}}_{J_0}$ is the J_0 level MODWT smooth, still holds true. This is useful in practice because as stated before, a special property of the MODWT wavelet details and smooths is that they are associated with zero-phase filters, i.e. features in the original time series are alignable with the wavelet details and smooth. Thus there is no need for an adjustment.

Variance decomposition. In order to retain the variance preserving property of the DWT, the wavelet and scaling coefficients must be rescaled properly as seen above. Unfortunately, although the MODWT is capable of producing a scale-by-scale analysis of variance upon the energy decomposition (Percival and Mojfeld (1997)),

$$\|\mathbf{x}\|^2 = \sum_{j=1}^{J_0} \|\widetilde{\mathbf{w}}_j\|^2 + \|\widetilde{\mathbf{v}}_{J_0}\|^2, \tag{2}$$

energy preservation does *not* hold for the MODWT details and smooths in general:

$$\|\mathbf{x}\|^2 \neq \sum_{j=1}^{J_0} \|\mathbf{d}_j\|^2 + \|\mathbf{s}_{J_0}\|^2. \tag{3}$$

This is because the MODWT is not an orthonormal transformation. Indeed, Percival and Walden (2000, Ch. 5.3) show for example that $\|\mathbf{d}_1\|^2 \leq \|\widetilde{\mathbf{w}}_1\|^2$. Thus when using the MODWT, one is restricted to analyzing the wavelet and scaling coefficients in order to quantitatively study the scale-dependent variance properties.

[4] An alternative definition of shifts for both extremal phase and least asymmetric wavelet filters is provided by Hess-Nielsen and Wickerhauser (1996), but Gençay *et al.* (2002a, p. 145) argue the differences be of minor importance.

5.5 Wavelet variance

The DWT and the MODWT can decompose the sample variance of a time series on a scale-by-scale basis. A wavelet based analysis of variance is sometimes called a WAVELET SPECTRUM. Such a spectrum may be of interest for several reasons (Percival and Walden (2000, p. 296)):

1. a scale-by-scale decomposition of variance is useful if the phenomena consists of variations over a range of different scales;
2. wavelet variance (to be defined below) is closely related to the concept of SPECTRAL DENSITY FUNCTION (SDF, Fourier spectrum);
3. wavelet variance is a useful substitute for the variance of a process for certain processes with infinite variance.

Consider a discrete parameter real-valued stochastic ARFIMA process $\{X_t\}$ (see App. C) whose dth order backward difference Y_t is a *stationary* process with mean μ_Y (not necessarily zero). Then a Daubechies wavelet filter $\widetilde{h}_l$ of width $L \geq d$ results in the jth wavelet coefficient process

$$\overline{w}_{j,t} \doteq \sum_{l=0}^{L_j-1} \widetilde{h}_{j,l} X_{t-l},$$

being a stationary process (for a stationary process any L would suffice). Now define the (time-independent, global) WAVELET VARIANCE for $\{X_t\}$ at scale $\lambda_j \doteq 2^{j-1}$ to be

$$\nu_X^2(\lambda_j) \doteq \mathbb{V}\left\{\overline{w}_{j,t}\right\},$$

which represents the contribution to the total variability in $\{X_t\}$ due to changes at scale λ_j. Then by summing up these time-scale specific wavelet variances, one gets the variance of $\{X_t\}$:

$$\sum_{j=1}^{\infty} \nu_X^2(\lambda_j) = \mathbb{V}\{X_t\}. \tag{4}$$

Notice that the wavelet variance is well-defined for both stationary and non-stationary processes with stationary dth order backward differences as long as the width L of the wavelet filter is large enough. In the non-stationary case, in particular, the sum of the wavelet variances diverges to infinity. An advantage of the wavelet variance is that it handles both types of processes equally well. (Percival and Walden (2000, Ch. 8.2).)

The wavelet coefficients affected by the boundary are a source of a bias and warrant special attention (see Sec. 5.1). Namely, by taking into account only coefficients that are *not* affected by the periodic boundary conditions, an *unbiased* estimator of $\nu_X^2(\lambda_j)$ is

$$\widetilde{\nu}_X^2(\lambda_j) = \frac{1}{M_j} \sum_{t=L_j-1}^{N-1} \widetilde{w}_{j,t}^2,$$

where $M_j \doteq N - L_j + 1 > 0$ and $\widetilde{w}_{j,t} \doteq \sum_{l=0}^{L_j-1} \widetilde{h}_{j,l} X_{t-l \bmod N}$ are the (periodically extended) MODWT coefficients. Furthermore, if a sufficiently long wavelet filter is used, i.e. if $L > 2d$ (or if $\mu_Y = 0$), then $\mathbb{E}\left(\overline{w}_{j,t}\right) = 0$, which in turn implies that

$$\nu_X^2(\lambda_j) = \mathbb{E}\left\{\overline{w}_{j,t}^2\right\} = \mathbb{E}\left\{\widetilde{w}_{j,t}^2\right\},$$

where the last equality follows from using coefficients only for the range not affected by the boundary. In fact, if the sample mean over *all* possible t would be calculated, then in general, one would get a *biased* estimator of $\nu_X^2(\lambda_j)$. (Percival and Walden (2000, Ch. 8).)

Now consider $\{X_t\}$ to be a stationary process with $S_X(f)$ defined over the frequency interval $[-1/2, 1/2]$. A fundamental property of the Fourier spectrum is that

$$\int_{-1/2}^{1/2} S_X(f)df = \mathbb{V}\{X_t\},$$

i.e. the SDF decomposes the variance of a series across different frequencies (Percival and Walden (2000, p. 296)). On the other hand, the wavelet spectrum decomposes the variance of a series across different scales (see Eq. (4)). By recalling the close (although fragile) relationship between frequency and time-scale (see Sec. **??**), it is then no surprise that the estimates of the wavelet variance can be turned into SDF estimates. In specific, the band-pass nature of the MODWT wavelet filter implies that

$$\nu_X^2(\lambda_j) \approx 2 \int_{1/(2^{j+1}\Delta t)}^{1/(2^j \Delta t)} S_X(f)df,$$

where S_X is the SDF estimated by the squared magnitude of the coefficients of the DFT, called *periodogram*,

$$\widehat{S}_X(f_k) = \frac{1}{N} \left| \sum_{t=0}^{N-1} X_t e^{-2\pi f_k t} \right|^2,$$

and $f_k = k/N$ denotes the kth Fourier frequency, $k = 0, ..., \lfloor N/2 \rfloor$. This approximation improves as the width L of the wavelet filter increases because then $\widetilde{h}_{j,l}$ becomes a better approximation to an ideal band-pass filter. In fact, if the filter is wide enough, one can estimate S_X using piecewise constant functions over each interval $[1/2^{j+1}\Delta t, 1/2^j \Delta t]$. Notice that in the case of *long-memory*, however, this approximation underestimates the lowest frequencies (see Percival and Walden (2000, Ch. 8.5)).

The SDF can be used to construct a confidence interval for $\nu_X^2(\lambda_j)$. Namely, assuming that $\{\overline{w}_{j,t}\}$ is a Gaussian process, then for large M_j the random variable $\widetilde{\nu}_X^2(\lambda_j)$ is approximately Gaussian distributed with mean $\widetilde{\nu}_X^2(\lambda_j)$ and variance $2A_j/M_j$, where $A_j \doteq \int_{-1/2}^{1/2} S_j^2(f)df$, provided that A_j is finite and $S_j(f) > 0$ almost everywhere. The downside of this contruction is that the confidence interval may have a *negative* lower limit which is problematic when plotting wavelet variance estimates in a double-logarithmic scale. Furthermore, Gençay *et al.* (2002a, p. 244) point out that an incorrect Gaussian assumption would produce too narrow intervals that would not reflect the true variability of the point estimate.[5] There does exist other ways of constructing confidence intervals, but these are a bit more complex. Confidence intervals could for example constructed using an "equivalent degrees of freedom" argument and chi-squared distribution (see Percival and Walden (2000, pp. 336–7)) or multitaper spectrum estimation (Serroukh *et al.* (2000)).

It is well-known that the periodogram is an *inconsistant* estimator of the Fourier spectrum (see Priestley (1992, p. 425)). Likewise, the popularly used GPH-estimator (Geweke and Porter-Hudak (1983)) based on an ordinary least squares (OLS) regression of the log-periodogram for frequencies close to zero (see App. C), is in general an inconsistant estimator of the long-memory parameter from a fractionally integrated process with $|d| < 1/2$. Other asymptotic properties of this estimator are problematic too (see Hurvich and Beltrao (1993) and Robinson (1995)).[6]

On the other hand, it has been argued above that wavelet variance is a regularization of the Fourier spectrum (see Percival (1995) or McCoy and Walden (1996)) so that those scales that contribute the most to the variance of the series are associated with those coefficients with the largest variance. Using this fact, Jensen (1999) showed that an OLS-based wavelet estimator is indeed *consistant* when the sample variance of the wavelet coefficients is used in the regression. Namely, using the wavelet variance of the DWT coefficients $w_{j,t}$,

$$\overline{\nu}_X^2(\lambda_j) = \frac{1}{2^j} \sum_{k=0}^{2^j-1} w_{j,k}^2, \tag{5}$$

one has that

$$\mathbb{V}\{w_{j,t}\} = \nu_X^2(\lambda_j) \to \sigma^2 2^{j(2d-1)},$$

as $j \to \infty$. Here σ^2 is a finite constant. If a large number of wavelet coefficients are available for scale j, then the sample wavelet variance provides a consistent estimator

[5] In the context of volatility modeling (see Sec. 7.4), the assumption of Gaussinity is admittedly not correct; absolute returns are known to be exponentially distributed (Granger and Hyung (1999)). The assumption of Gaussianity is not considered critical, though.

[6] However, in the case of $0 < d < 1/2$ and under certain regularity conditions – Gaussianity, in particular – Robinson (1995) has proven that the GPH-estimator is consistant and asymptotically Gaussian. But as argued in the previous Footnote, volatility is not distributed normally.

of the true wavelet variance (Jensen (1999, p. 22)). Thus, by taking logarithms on both sides, one obtains the log-linear relationship

$$\log \nu_X^2(\lambda_j) = \log \sigma^2 + (2d-1)\log 2^j, \tag{6}$$

from which the unknown d can be estimated consistently by OLS-regression by replacing ν_X^2 with its sample variance $\overline{\nu}_X^2$ of Eq. (5).

The asymptotic variance of the estimator of d has been derived in Jensen (1999). In particular, Jensen found that the (negative) bias found in this estimator is offset by its low variance. Indeed, in *mean square error* (MSE) sense the wavelet OLS-estimator fared significantly better than the GPH-estimator. To further reduce the MSE of $\widehat{d}$, a weighted least squares (WLS) estimator could be applied (see e.g. Abry *et al.* (1993) and Abry and Veitch (1998)). In particular, Percival and Walden (2000, Ch. 9.5) have shown that the WLS estimator reduces the MSE by a factor of two in comparison to the OLS estimator in simulation studies.

Wavelet variance can be defined also *locally.* But unlike in Jensen (1999), where all wavelet coefficients were used in calculating the wavelet variance, now only those "close" to the time point t are used. Namely, given $L > 2d(u)$, an *unbiased* estimator of LOCAL WAVELET VARIANCE for $\{X_t\}$ at scale λ_j based upon the MODWT is

$$\widetilde{\nu}_X^2(u, \lambda_j) = \frac{1}{K_j} \sum_{s=\tau_j}^{\tau_j + K_j} \widetilde{w}_{j,t+s,T}^2, \tag{7}$$

where u represents a time point in the rescaled time domain $[0,1]$ (i.e. $u = t/T$), K_j is a *"cone of influence",* and τ_j is an *"offset"* described below (Whitcher and Jensen (2000, p. 98)).

In principle, the "cone of influence" K_j includes only those wavelet coefficients where the corresponding observation made a significant contribution. Indeed, as Percival and Walden argue (2000, p. 103), the width of the filter L is *not* a very good measure of the effective width of the filter because coefficients around $l = 0$ and $l = L_j - 1$ are very close to zero and thus do not significantly contribute to the calculation of the wavelet coefficient. Whitcher and Jensen (2000) suggest K_j (the central portion of a filter) as a more reasonable choice. A slight inconvience in using K_j is that it varies across scales and different filters. The tabulated values for Daubechies family of wavelets are given in Whitcher and Jensen (2000, Table 1). Also the values of "offsets" τ_j for each wavelet filter $L > 2$ are needed to indicate where the width K_j begins (given in Whitcher and Jensen (2000, Table 2)).

Whitcher and Jensen (1999) have shown that when the MODWT is being applied to a LOCALLY STATIONARY (in the sense of Dahlhaus (1996, 1997), see App. D) long-memory process $\{X_{t,T}\}$, then the level-j MODWT wavelet coefficients $\{\widetilde{w}_{j,t,T}\}$ form a locally stationary process with mean zero and time-varying variance

$$\mathbb{V}\{\widetilde{w}_{j,t,T}\} = \nu_X^2(u, \lambda_j) \to \sigma^2(u) 2^{j[2d(u)-1]},$$

as $j \to \infty$. Thus, analogously to Eq. (6),

$$\log \nu_X^2(u, \lambda_j) = \log \sigma^2(u) + [2d(u) - 1] \log 2^j, \tag{8}$$

from which the unknown $d(u)$'s can be estimated consistently by OLS by replacing ν_X^2 with its time-varying sample variance $\widetilde{\nu}_X^2$ from Eq. (7). The exact expression for $\sigma^2(u)$ is given in Whitcher and Jensen (2000). In general, Gençay *et al.* (2002a, p. 172) argue that the parameter estimation for a non-stationary long-memory time series model through OLS (or ML) *"should benefit greatly"* from wavelet-based methods.

Through simulations, Whitcher and Jensen (2000) showed that the *median* of $\widehat{d}(u)$ accurately estimates the true value of the fractional differencing parameter (with a slight negative bias near the boundaries) in the case of globally stationary ARFIMA. Because less information is used to construct the local estimator than the global one, $\widehat{d}(u)$ also exhibited a slight increase in its MSE. Interestingly, when the ARFIMA process was disturbed by a sudden shift in the long-memory parameter to imitate local stationarity, the estimated fractional differencing parameter still performed well (on both sides of the change) although with a slight bias and increase in MSE at the boundaries.

6 Volatility modeling

Volatility, interpreted as uncertainty, is one of the key variables in most models in modern finance.[7] The explosive growth in derivative markets and the recent availability of high-frequency data have only highlighted its relevance. In option pricing, for example, volatility of the underlying asset (which may be volatility itself, by the way) must be known from now until the option expires as accurately as possible. In financial risk management, volatility forecasting has even become compulsary after the "1996 Basle Accord Amendment" which sets the minimal capital requirements in banks. In a more general level, evidence of how strongly financial market volatility can affect the economy as a whole was recently acquired when the terrorist attack in New York on September 11, 2001 took place; periods of high uncertainty have economically paralyzing consequences. In a way, then, volatility estimates can be considered *"as a barometer for the vulnerability of financial markets and the economy"* (Poon and Granger (2003, p. 479)).

In this section some of the most popular measures of volatility and types of models are discussed. There are basicly two strands of volatility models, those assuming that conditional variance depends on *past* values (i.e. observation driven models) and

[7]Volatility is *not* the same as risk, however. In particular, risk is usually associated with small or negative returns (the so-called "downside risk") whereas most measures of dispersion (e.g. standard deviation) make no such distinction. Furthermore, standard deviation is a useful risk measure only when it is attached to a distribution or a pricing dynamic. For further details on the conceptual differences between volatility, risk, and standard deviation, see Poon and Granger (2003, Sec. 2.1).

those assuming that conditional variance is *stochastic* (made precise later). Although this section starts with the former approach, most attention is paid to the more flexible case of "stochastic volatility" (flexibility is discussed in Carnero *et al.* (2001), for instance). Poon and Granger (2003) provide an extensive up-to-date review of different types of volatility models used for forecasting in financial markets.

6.1 Measures of volatility

It is an indisputable stylized fact that volatility tends to cluster so that the variance is time-varying and shows persistant behavior. A systematic search for the *causes* of serial correlation in conditional second moments is in its infancy, though. Diebold and Nerlove (1989) discuss the possibility of a serially correlated news arrival process as the generating mechanism for which Engle *et al.* (1990) find some evidence. Further support is given by Tauchen and Pitts (1983) who have argued that such a news process would probably induce a strong contemporaneous relationship between volume and volatility, confirmed by e.g. Lamoureux and Lastrapes (1990a) and surveyd in Karpoff (1987). In particular, using the mixture-of-distribution hypothesis (see Clark (1973)), Andersen and Bollerslev (1997b) show that long-memory features of volatility (the slowly decaying autocorrelation function) may indeed arise through the interaction of a large number of heterogenous information arrivals. Such a finding is important because then long-memory characteristics reflect inherent properties of the DGP, rather than structural shifts as suggested e.g. by Lamoureux and Lastrapes (1990b). Still, part of the reason for the uncertainty of the causes is the lack of unique, universally accepted definition of volatility. The ambiguity stems from its latency, i.e., volatility is not a directly observable measure in general.

In volatility estimation, the basic paradigm is that "volatility" (in this case approximated by the square of returns r_t) can be decomposed into predictable and unpredictable components,

$$r_t = \sigma_t \varepsilon_t,$$

where ε_t are IID disturbances with mean 0 and variance 1. By definition, then, the predictable component is the conditional variance σ_t^2 of a series. The determinants of the predictable part are of special interest in finance because the risk premium is a function of it.

In a seminal paper, Engle (1982) proposed that the conditional variance depends linearly on the past squared values of the process,

$$\sigma_t^2 = \sigma^2 + \sum_{k=1}^{q} \alpha_k r_{t-k}^2;$$

the well-known ARCH(q) model. Bollerslev (1986) generalized this to the parsimonious GARCH(p,q) model,

$$\sigma_t^2 = \sigma^2 + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2 + \sum_{k=1}^{q} \alpha_k r_{t-k}^2,$$

where the volatility is a linear function of both lagged squares of returns and lagged volatilities. Bollerslev showed that this equation defines a second-order stationary solution if (and only if) $\sum_{j=1}^{p} \beta_j + \sum_{k=1}^{q} \alpha_k < 1$ (and $\sigma^2 > 0$). In practice, GARCH$(1,1)$ often suffices. Moreover, it is usually found (in long time series, at least) that the estimated parameters β and α sum close to one suggesting the presence of a "unit root" in the volatility equation. This so-called "Taylor-effect" (Taylor (1986)) has commonly been interpreted as evidence of volatility persistance (see e.g. Poterba and Summers (1986)), although this interpretation has faced a considerably amount of criticism lately (see e.g. Mikosch and Stărică (2004)). In any case, Engle and Bollerslev (1986) extended the model to the case where $\sum_{k=1}^{p} \beta_j + \sum_{k=1}^{q} \alpha_k = 1$, i.e. the *integrated* GARCH (IGARCH) model.[8] Counterintuitively the autocorrelation function for IGARCH$(1,1)$ is exponentially (not hyperbolically) decreasing, which is itself indicative of "short-memory" although the effect of a shock to expectation is permanent (for an insightful discussion along these lines, see Ding and Granger (1996)).

A more succesful (in the stock markets at least, not necessarily in the FX markets) extension of GARCH is to let σ_t^2 be an *asymmetric* function of the past data. Namely, to model the stylized fact that in stock markets volatility is negatively correlated with lagged returns, the so-called *leverage-effect* (first noted by Black (1976)), Nelson (1988) proposed *exponential* GARCH (EGARCH) that model the logarithm of the variance $\log \sigma_t^2$. Added flexibility to this model is achieved by applying a "fractional differencing operator" d (a parameter that is going to play a big role in the following sections, see App. C), resulting in *fractionally integrated* EGARCH (FIEGARCH) model (Bollerslev and Mikkelsen (1996)) that nests the conventional EGARCH for $d = 0$. In general, a fractional generalization has proved to be empirically useful in modeling long-term dependence in conditional variances (see e.g. Vilasuso (2002)). (This generalization corresponds to the generalization of the standard ARIMA class of models to fractionally integrated ARMA (ARFIMA) (see App. C) models that model long-term dependence in mean.) The usefulness is mainly attributable to the fact that the weakly stationary FIEGARCH succeeds in modeling the slow hyperbolic rate of decay of a shock to the forecast of $\log \sigma_{t+T}^2$ if $0 < d < 1/2$, and thus captures the observed "long-memory" in volatility (confirmed again in Sec. 7). Of course, there exist numerous other extensions such as regime switching models, for example.

[8]Here the prefix "integrated" does *not* imply non-stationarity as in the case of random walk, however (proved in Bougerol and Picard (1992)). More precisely, *strict* stationarity still holds but because the marginal variance of r_t is infinite, *weak* stationarity does not (see e.g. Gourieroux and Jasiak (2001, Ch. 6.2.4)).

For some of the most popular ones, see Bollerslev *et al.* (1992), Hentschel (1995), and (in the multivariate case) Kroner and Ng (1998).[9]

Several other type of measures have been used to measure volatility, as well. It is for example known that normalized squared or absolute returns over an appropriate horizon provide an *unbiased* estimate for volatility. Although the majority of time series volatility models are squared returns models (as above), absolute returns based models seem to produce better volatility forecasts in practice (see e.g. Taylor (1986) and McKenzie (1999)). Indeed, for example Ding *et al.* (1993) suggest measuring volatility directly from absolute returns (this is my choice in Sec. 7, too). The use of squares is most likely a reflection of the Gaussian assumption made regarding the data (McKenzie (1999, p. 50)). But the error distribution of stock market returns is not Gaussian and therefore higher moments than the second one must be considered. In fact, Davidian and Carroll (1987) have shown that absolute returns specification is more robust against asymmetry and non-normality. Unfortunately, though, in either case the signal-to-noise ratio is diminutive when evaluations are conducted over daily time-spans (Andersen and Bollerslev (1997)); i.e. absolute (or squared) returns over longer horizons provide a very noisy estimate for volatility. Blair *et al.* (2001) have reported a significant increase in forecasting ability for 1-day ahead forecast when intraday 5-minute squared returns are used instead of daily ones. Another closely related variance estimate is to average the squared returns over a fixed horizon (see e.g. Poterba and Summers (1986)). Standard time series techniques could then be applied to assess the temporal dependence. As explained in Bollerslev *et al.* (1992, pp. 17–18), this two-stage procedure is subject to criticism, too. First, it does not make efficient use of all the data and the conventional standard errors from the second-stage estimation may not be appropriate. Second, there is the possibility that the actual parameter estimates may be inconsistant. Third, this kind of a procedure may lead to misleading conclusions about the true underlying dependence in the second-order movements of the data.

Another popular method of assessing volatility is based on the implied volatility from options prices. This particular approach is *indirect*, however. Namely, the traditional Black–Scholes formula can be inverted to give an estimate of volatility under the assumption of a constant variance (i.e. volatility). Engle and Mustafa (1992), among others, have considered the case of ARCH-volatility in this setting. Unfortunately, when using stochastic volatility (to be discussed in the next two subsections) several complications arise in option theory (see e.g. Wiggins (1987) and Melino and Turnbull (1990)). Moreover, on practical side, *not every* asset of interest have actively traded options and implied volatilities derived from frictionless market models may be affected by institutional factors distorting the time series analysis (Fung and Hsieh (1991)). In fact, option markets may not be sufficienty developed to allow for meaningful variations in intra-day implied volatility to be derived as Goodhart

[9] An estimate of volatility in these models is attained via (quasi) *maximum likelihood* (ML) or via *generalized method of moments* (GMM).

and O'Hara (1997) note. Still, volatility extracted from options can be very informative because it uses a potentially richer information set and could therefore lead to improved forecasting performance.

Yet another method is the use of historical "highs and lows" as in Parkinson (1980) who assumes that prices follow a continuous time random walk with constant variance. By developing this idea a little further, Garman and Klass (1980) derive several efficient estimator of volatility using highs, lows, opening, closing, and transactions volume. Beckers (1983) test the accuracy of these estimators and suggest an adjustment which could even be improved upon by including variances implied in option prices. However, the generalization of these ideas to other stochastic processes allowing for time-varying variances is not straightforward (Bollerslev *et al.* (1992, p. 19)).

Theoretically the most attractive way of measuring volatility is the sum of short-term intraday squared (or absolute) returns of a predetermined horizon (usually a day) resulting in *realized volatility* (explicitly considered in Fung and Hsieh (1991) and Andersen and Bollerslev (1997)). The idea that high-frequency data might be useful in estimating the variance is not that new, however. In particular, Schwert (1989) has estimated monthly volatility from daily returns and, most importantly, Merton (1980) has postulated that the variance of returns can be estimated far more accurately from the available time series of realized returns than can the expected return (under certain general assumptions). This makes sense because for a random walk a minimal exhaustive statistic for volatility is essentially given by the full set of increments (Corsi *et al.* (2001)). Indeed, under the assumption of a continuous-time diffusion, the realized volatility is a *consistant* estimator of the 1-day *integrated volatility*. In fact, it can be shown that the stochastic error of the measure can be reduced arbitrarily by increasing the sampling frequency of returns. Integrated volatility is a natural population measure of the volatility as it is fundamental in pricing of derivative securities (see e.g. Hull and White (1987)).

The use of realized volatility has some interesting consequences. First, it allows one to treat volatility as an observable. Furthermore, Andersen and Bollerslev (1997) have shown that realized variance takes the beloved ARCH models "back into business" in the sense that ARCH models again seem to serve as good forecasting devices (which has been put into serious doubt lately). However, the trouble with the realized volatility approach is that in practice one does not observe the price continuously. In addition, the small time intervals are contaminated by microstructure effects such as "bid-ask bounce" (see e.g. Roll (1984)) and volatility seasonalities (see e.g. Andersen and Bollerslev (1997)). Thus the precision to which one can measure the volatility using high-frequency data depends on the characteristics of the return series analyzed (Bai *et al.* (2001)). Finally, the realized variance approach is computationally quite expensive.

6.2 Stochastic volatility and long-memory

Most of the modern financial theory is based on continuous-time semimartingales (see e.g. Shiryaev (1999) for an excellent account). In particular, stochastic volatility (SV) models of the form

$$dX(t) = \mu(t)dt + \sigma(t)dW(t) \tag{9}$$

belong to this family. Here the drift $\mu(t)$ and the instantaneous standard deviation $\sigma(t)$ are time-varying random functions and $W(t)$ is a standard Brownian motion. Many specifications for $\sigma(t)$ are available (see Taylor (1994)). It may be, for example, that the logarithm of the volatility follows an Ornstein–Uhlenbeck process (as in Wiggins (1987)).[10] Continuous-time SV models have been proposed and applied also in Melino and Turnbull (1990) and Harvey *et al.* (1994), for example. See Ghysels *et al.* (1995) for a good review.

Although continuous-time models are more elegant to work with in theory, in practice one often settles for a discrete model. A discrete-time SV model may be written as

$$y_t = \sigma_t \varepsilon_t,$$

where y_t denotes the demeaned return process $y_t = \log(S_t/S_{t-1}) - \mu$, $\{\varepsilon_t\}$ is a series of IID random disturbances with mean 0 and variance 1, and the conditional variance $\{\sigma_t^2\}$ is modeled as a stochastic process $\{\log \sigma_t^2\} \doteq \{h_t\}$. The modeling of volatility as a stochastic variable immediately leads to heavy tailed distributions for returns (e.g. Poon and Granger (2003, p. 485)). Here the logarithm ensure that $\{\sigma_t^2\}$ is always positive, similarly to the EGARCH specification above (Sec. 6.1). The important difference is, however, that $\{\sigma_t^2\}$ is *not* directly observable. Furthermore, $\{h_t\}$ is now independent of $\{\varepsilon_t\}$. Introduction of correlation between $\{h_t\}$ and $\{\varepsilon_t\}$ would produce volatility asymmetry (Hull and White (1987)).

From the point of view of financial theory, a particularly attractive and simple model for $\{h_t\}$ is an AR(1)-process $h_t = \gamma + \phi h_{t-1} + \eta_t$, where $\eta_t \sim IID(0, \sigma_\eta^2)$, and $|\phi| < 1$ ensures that $\{h_t\}$ (and hence $\{y_t\}$) is strictly stationary.[11] Such an autoregressive term introduces persistance (i.e. volatility clustering). In general, SV models are more flexible than GARCH models because of the extra volatility noise term η_t in the volatility equation. For example, the simple ARSV(1) specification has been shown by Carnero *et al.* (2001) to be empirically more adequate than the most popularly used GARCH$(1,1)$. Interestingly, Carnero *et al.* demonstrate

[10] In specific, Wiggins (1987) supposes that $d(\log\sigma(t)) = \lambda(\xi - \log\sigma(t))\,dt + \gamma dW_2(t)$, where $dW(t)dW_2(t) = \rho dt$ and γ is called the "volatility of volatility". When $\rho = 0$, the price process and the volatility process are not correlated at all. Interestingly, γ alone is enough to produce heavy tails. When $\rho < 0$, large negative return corresponds to high volatility and stretches the left tail into the left. The opposite happens when $\rho > 0$.

[11] The attractiveness of this "ARSV(1)" specification (proposed by Taylor (1986)) stems from the fact that an AR(1)-process is the natural discrete-time approximation to a continuous-time Ornstein–Uhlenbeck process.

that ARSV(1) produces smoother volatility estimates. Not all of the results are that unanimous, though (see Poon and Granger (2003, Sec. 5.3) and the references there-in). Finally notice that the model for $\{y_t\}$ is usually represented in the form

$$y_t = \sigma \exp(h_t/2)\varepsilon_t,$$

where the scale parameter $\sigma > 0$ removes the need for the constant term γ in the first-order autoregression.[12]

Naturally there exist many specifications for the volatility scheme $\{h_t\}$, such as ARMA or random walk. In particular, a *long-memory stochastic volatility* (LMSV) model was proposed in Breidt *et al.* (1998) (and independently in Harvey (1998)). In their model $\{h_t\}$ (i.e. log-volatility) is generated by *fractionally integrated Gaussian noise*,

$$(1-B)^d h_t = \eta_t,$$

where $|d| < 1/2$ and $\eta_t \sim NID\left(0, \sigma_\eta^2\right)$. More generally, $\{h_t\}$ can be modeled as an ARFIMA(p,d,q) process,

$$\phi(B)(1-B)^d h_t = \theta(B)\eta_t, \tag{10}$$

where $\phi(z) = 1 - \phi_1 z - ... - \phi_p z^p$ for $|z| \leq 1$ is an autoregressive polynomial of order p, $\theta(z) = 1 + \phi_1 z + ... + \phi_q z^q$ is a moving average polynomial of order q, both $\phi(z)$ and $\theta(z)$ have all of their roots outside the unit circle, and $\theta(z)$ has no roots in common with $\phi(z)$. Notice that this model encompasses a "short-memory" model when $d = 0$.

Breidt *et al.* (1998) argued the LMSV model to have certain advantages over observation driven models (e.g. FIEGARCH). For example, because it is built from the widely used ARFIMA class of long-memory models, LMSV inherits most of the statistical properties of ARFIMA models and is therefore analytically tractable. Even the limiting distribution of the GPH-estimator of d has been derived (see Deo and Hurvich (1999) and Velasco (1999)). Moreover, the estimation of d is *not* crucially dependent on the choice of a unit discrete time interval, as noted in Bollerslev and Wright (2000, p. 87). Namely, although the LMSV (like ARFIMA) model is not closed under temporal aggregation, the *rate* of decay of the autocovariance function of squared (or absolute) returns is invariant to the length of the return interval (see Chambers (1998)).

The LMSV model is still a stationary model, however. Thus it ignores *"the known intraday volatility patterns and the irregular occurances of market crashes, mergers and political coups"* as Jensen and Whitcher (2000) note. In particular, the long-memory parameter d may not be constant over time. This motivated them to introduce a *non-stationary* class of long-memory stochastic volatility models with *time-varying* parameters. In their model, the logarithmic transform of the squared

[12] However, the estimation of SV models is notoriously difficult and usually done by variants of the method of moments (as in Melino and Turnbull (1990), for example). For a survey of estimation methods for stochastic volatility models, see Broto and Ruiz (2002).

returns is a locally stationary process that has a time-varying spectral representation (see App. D). This means that the level of persistance associated with a shock to conditional variance (which itself is allowed to vary in time) is dependent on when the shock takes place. The shocks themselves, of course, still produce responses that persist in the long-memory hyperbolic sense.

In specific, Jensen and Whitcher defined $y_{t,T}$ to be

$$\begin{aligned} y_{t,T} &= \exp\left(H_{t,T}/2\right)\varepsilon_t, & \text{(11a)} \\ \mathbf{\Phi}(t/T,B)(1-B)^{d(t/T)}H_{t,T} &= \mathbf{\Theta}(t/T,B)\eta_t, & \text{(11b)} \end{aligned}$$

where $|d(u)| < 1/2$, $\varepsilon_t \sim NID(0,1)$ and $\eta_t \sim NID(0,\sigma_\eta^2)$ are independent of each other. The functions $\mathbf{\Phi}(u,B)$ and $\mathbf{\Theta}(u,B)$ are, respectively, order p and q polynomials whose roots lie outside the unit circle uniformly in u and whose coefficients functions, $\phi_j(u)$, for $j = 1, ..., p$, and $\theta_k(u)$, for $k = 1, ..., q$, are continuous on $\mathbf{R}$. The coefficient functions satisfy $\phi_j(u) = \phi_j(0)$, $\theta_k(u) = \theta_k(0)$ for $u < 0$, and $\phi_j(u) = \phi_j(1)$, $\theta_k(u) = \theta_k(1)$ for $u > 1$, and are differentiable with bounded derivatives for $u \in [0,1]$. Notice by setting $\Phi(u,B) = \Phi(B)$, $\Theta(u,B) = \Theta(B)$, and $d(u) = 0$ for all $u \in [0,1]$, then one gets the SV model of Harvey *et al.* (1994). If, on the other hand, one sets $d(u) = d$ for all $u \in [0,1]$, then one gets the LMSV model (Eq. (10)) of Breidt *et al.* (1998).

7 Empirical analysis

The analysis below is *not* yet complete. More emphasis will be put on "microstructure effects" and economic interpretation. In particular, the FFF (see App. E) is currently being used to remove the strong intraday periodicity in volatility. The results of the re-runned analysis *compared* to the ones without the periodicity removed. This comparison should shed some light on whether the (i) scaling law (using wavelet variances across different levels) and consequently (ii) the OLS-estimate of the fractional differencing parameter d changes when accounting for the seasonality.

7.1 Data description

The original data set included all stock transactions done at the Helsinki Stock Exchange (HEX) between January 4 (1999) and December 30 (2002), i.e. it was so-called "tick-by-tick" data. Because of its highest liquidity, the stock of Nokia Oyj was chosen and the data was discretized (i.e. homogenized): 5-minute prices were extracted using the closest transaction price to the relevant time mark.[13] Discretizing is necessary for the wavelet decomposition to be interpretable in terms of time-scales that capture a band of frequencies (cf. spectral analysis). From a theoretical perspective

[13] The HEX is by far the most liquid market place trading Nokia: In year 2003, for example, the HEX accounted for 62,1% of the total *number* of shares traded while the percentage for New York Stock Exchange (NYSE) was only 20,3%. (Source: the HEX (May 4, 2004).)

discretizing can be justified by assuming that the DGP does not vary significantly over short time intervals.[14] It is worth acknowledging, however, that the statistical behavior of the sampled data *could* differ significantly from the behavior of the DGP from which the sample was obtained. In the current context this principle might manifest itself through the so-called "non-synchronous trading" and possibly inducing negative serial correlation (see Campbell *et al.* (1997, Ch. 3.1) and Lo and MacKinlay (1999, Ch. 4)). To better account for such microstructure effects, one could have also used the last transaction before the relevant time mark (a method originally introduced by Wasserfallen and Zimmermann (1985) and used in Hol and Koopman (2002), among others) or linearly interpolated price (introduced by Andersen and Bollerslev (1997c)). But because there was occasional liquidity problems, the closest one was considered the best compromise. This choice should not have any significant affect to the conclusions of the subsequent analysis.

The interval of 5 minutes has been used in many earlier studies (e.g. Andersen and Bollerslev (1998)). It has been found "optimal" in the sense that it is often the smallest interval that doesn't suffer too badly from microstructure effects such as "bid-ask bounce" (see Campbell *et al.* (1997, Ch. 3) or Gourieroux and Jasiak (2002, Ch. 14)). Concerning missing observations for a specific time mark (such as technical breaks and incomplete trading days), the previous price standing was always used. The 5-minute returns were then calculated as the scaled difference between successive log-prices, i.e.,

$$r_{t,d} = 100 \left(\ln P_{t,d} - \ln P_{t,d-1}\right),$$

where $r_{t,d}$ denotes the return for intraday period d on trading day t, with $d \geq 1$ and $t = 1, ..., T$. Notice that the prices $P_{t,d}$ were adjusted for splits but *not* for dividends. This is because there were only four divident paying days in the whole four year period 1999–2002 and their impact were very small. As a general rule, the empirical analysis is done *including* overnight returns if not otherwise mentioned (as in Sec. 7.6). Finally, the so-called "block trades" were not removed, possibly causing a few "artificially generated" jumps per month. These trades are currently considered to be of minor importance, however.

At the HEX, an electronic trading system called Helsinki Stock Exchange Automated Trading and Information System (HETI) has been in use since 1990. This means that there is no "floor" but brokers trade electronically, the smallest "tick-size" (i.e. price change) being 0.01.[15] As a general rule, all banking days are also market days at the HEX. From the point of view of data handling, however, one of

[14] For some other ways to deal with the "sporadic nature" of trading, see the discussion in Goodhart and O'Hara (1997) or Dacorogna *et al.* (2001, Sec. 3.2.1).

[15] At this point it is worthwhile to acknowledge that there are several different types of market places. Obviously, different systems potentially affect the dynamics of the price differently, too. So to be precise: the HEX is a continuous, order-driven limit-order-book (LOB) market place, although there is call *auction* at the market opening. (In year 2004, the HEX is going to be integrated to the SAXESS trading system, however. See Http://www.hex.fi.). For comparison, the NYSE is an order-driven, floor-based, continuous market with a *specialist* (acting as the "market maker"). In

the main problems were that the HEX did not have constant trading hours during the four years. In fact, the trading day was first extended to include evening hours like in many other countries, but then this trend was reversed. These changes were mainly caused by an international pressure towards harmonization of exchange open hours. The long-run trend of longer trading days has been suppressed by the weak market conditions during the last few years.[16] For example, at the HEX the evening trading hours currently form only around 6% of the daily trading volume. Therefore, in the following analysis *four* different time periods are analyzed. These periods are next being described in detail (see also Table ??). Notice that Periods I and III both contain approximately the same number of observations (although Period III does not last that long). In fact, these two periods are being analyzed in more detail in the next subsections.

Time period	**Trading (I)**	**AMT (I)**	**Trading (II)**	**AMT (II)**
4.1.1999 – 31.8.2000	10.30–17.30	17.30–18.00	–	9.00–9.30
1.9.2000 – 11.4.2001	10.00–18.00	18.00–18.15	–	8.30–9.00
17.4.2001 – 27.3.2002	10.00–18.00	18.03–18.30	18.03–21.00	8.30–9.00
2.4.2002 – 30.12.2002	10.00–18.00	18.03–18.30	18.03–20.00	8.30–9.00

In Period I, from January 4 (1999) to August 31 (2000), *continuous* trading took place between 10.30 a.m. and 5.30 p.m., totaling to 7 hours and 85 intraday 5-minute prices. Transactions between 8 a.m. and 10.30 a.m. were discarded, most of them belonging to the *after market trading* II (AMT (II)) taking place between 9.00 a.m. and 9.30 a.m.[17] Likewise, transactions between 5.30 p.m. and 6 p.m. were discarded because they belonged to AMT (I). Only one day, namely April 20 (2000), was an incomplete day and the missing observations were substituded by the last observed price. In total, therefore, there were 419 trading days resulting in 35,615 ($= 419 * 85$) price observations (i.e. 35,614 return observations).

In Period II, from September 1 (2000) to April 11 (2001), continuous trading was extended from both ends by half an hour.[18] Thus trading took place between 10 a.m. and 6 p.m., totaling to 8 eight hours and 91 intraday 5-minute prices. December 12 (2000) was an incomplete trading day and it was corrected by substitution. In total, therefore, there were 155 trading days resulting in 14,105 price observations.

In Period III, from April 17 (2001) to March 27 (2002), continuous trading was extended further by including evening hours from 6. p.m. to 9 p.m. A *technical break* (when no transactions took place), occured every day between 6 p.m. and

general, one particular advantage of a continuous market is that it provides good intraday market information. For more details on different systems, see Gourieroux and Jasiak (2001, Ch. 14.1).

[16]For example, Deutsche Börse cut its trading hours from 9.00 – 20.00 to 9.00 – 17.30 in the beginning of November 2003.

[17]During AMT, the trading price can fluctuate between the trading range established during continuous trading for round-lot trades (Http://www.porssisaatio.fi).

[18]Actually, this period ended the day before, but the data of day April 11 (2001) included transactions only up to 6.52 p.m. For simplicity, therefore, this day was ended at 6 p.m.

6.03 p.m. Importantly, continuous trading and AMT (I) took place *simultaneously.* This simultaneity required very careful filtering. Especially when the trading day experienced a big cumulative price change, then artificially big returns (even around 20%) could appear (see Fig. 2). An example of such a trading day was April 18 (2002) when Nokia announced its 1st quartal result that triggered a significant price drop earlier on that day. To safeguard from such artificial returns, the following filtering rule was applied: *prices that had a percentage price change of more than* 3% *relatively to the last genuine price recorded (before the technical break at 6 p.m.) were detected as artificial and replaced by the previous genuine price.* This particular rule was based on a careful inspection of the data (for some other rules often employed in high-frequency finance, see Dacorogna *et al.* (2001)).[19] And in fact, the noise reduction obtained with this filter was so considerable that the difference to the non-filtered series was evident by eye. In summary, continuous trading took place for 11 hours (including the 3-minute break) and produced 133 intraday 5-minute prices. There were no incomplete trading days. In total, therefore, there were 237 trading days resulting in 31,521 price observations.

In Period IV, from April 2 (2002) to December 30 (2002), continuous trading was cut from the end by an hour. That is, continuous trading took place between 10 a.m. and 8 p.m. – apart from the technical break and simultaneity just described – totaling to 10 hours and 121 intraday 5-minute prices. The same 3%-filter was employed. There were no incomplete trading days here either. In total, therefore, there were 188 trading days resulting in 22,748 price observations.

7.2 Preliminary data analysis

Statistical key figures of all the four periods are summarized below (Table ??). In what follows, however, only the first and the third period are being analyzed. There are at least three reasons for preferring these two periods over the other two. First, Periods I and III are of approximately equal size and also contain the greatest number of observations. Second, Periods I and III represent turbulent and calm regimes, respectively. In specific, Period I (1/99–8/00) is representative of the "IT-bubble" and Period III (4/01–3/02) of its aftermath. In fact, Polzehl *et al.* (2004) find that the "2001 recession" in the U.S. might have started as early as October 2000 and ended as late as the summer of 2003 which neatly supports these two categories. Notice that the volatilities of Periods I and III seem to differ significantly by simply eyeballing

[19]Notice that the percentage change was calculated relatively to the last genuine price because there is no guarantee that two artificial prices could not be adjacent. In fact, *if* the percentage would be have been calculated simply from adjacent prices, an artificial price would then have survived the filter. Admittedly, however, there is a small "defect" in this 3%-filter. Namely, fixing the denominator to the last genuine price is not reasonable if there is a strong price *trend* in either direction. But because the AMT (I) lasted only for 27 minutes, a trend was regarded of minor importance.

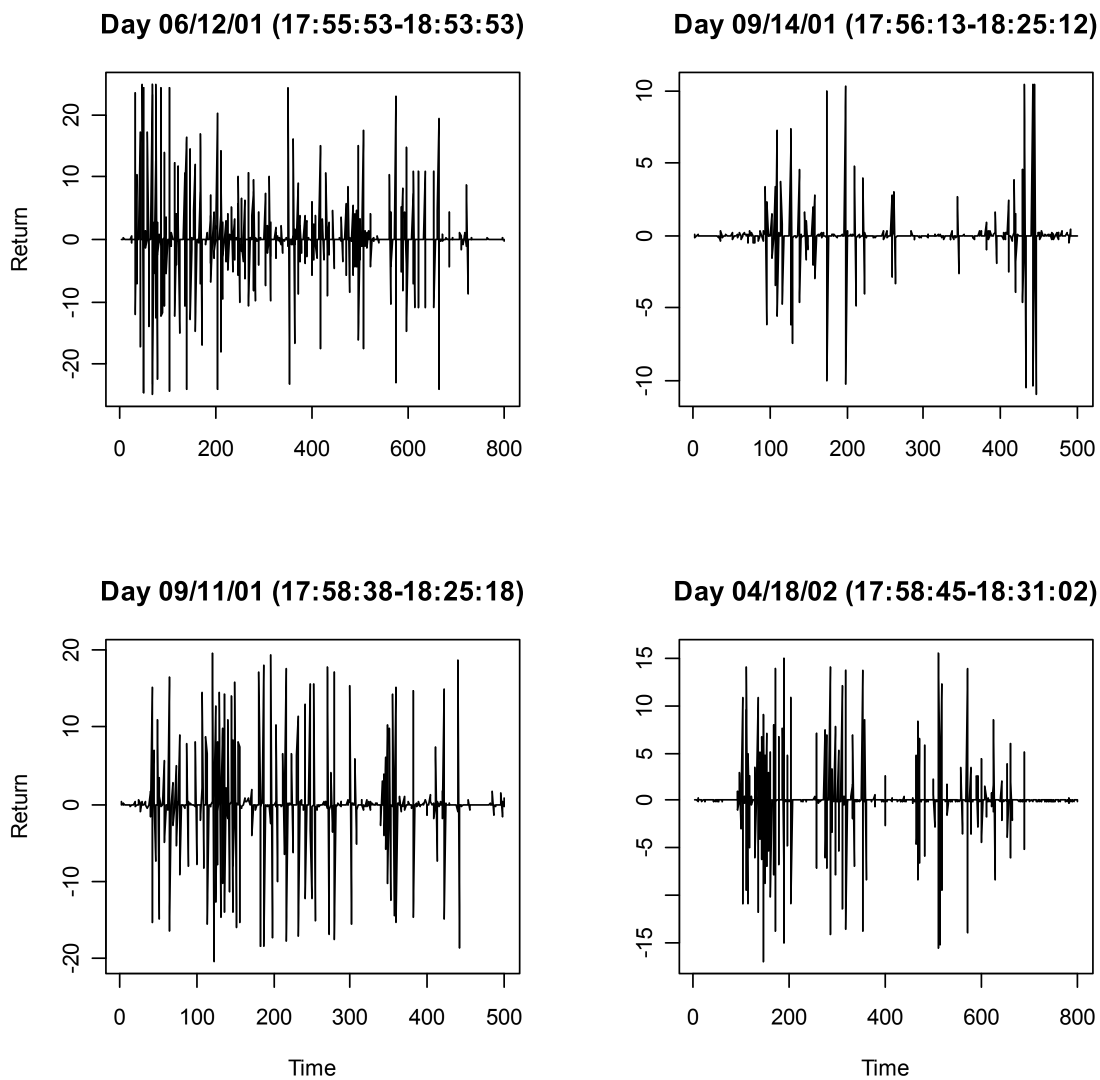


Figure 2: Four examples of trading days that experienced extremely high return variability during the AMT (I), 6.03 − 6.30 p.m (notice that *all* transactions are included here). Such volatility is not genuine, however, and must be filtered out before any reasonable analysis can be conducted.

the log-return series (see the bottom plots of Figs. 3 and 4).[20] This is important because it is known that structural breaks can generate artificial long-memory (see e.g. Diebold (1986), Lamoureux and Lastrapes (1990b), Granger and Hyung (1999), and Diebold and Inoue (2001)). In particular, Mikosch and Stărică (2004)) have argued that long-memory might be due to non-stationarity, thus lending stationary models (GARCH-type models, in particular) inappropriate over longer horizons. It is thus safer to analyze these periods separately. Third, although Periods I and III supposedly belong to different regimes, they have an interesting characteristic in common: they both contain an upward trend although the trend is much more pronounced in Period I (see the top plots of Figs. 3 and 4). This is relevant to the current study for example because a trend is known to be another possible source of spurious long-memory (see e.g. Bhattacharya *et al.* (1983)). Thus in many aspects it should be fruitful to compare the results of these two periods.

The sample autocorrelation functions (ACFs) of returns in Periods I and III differ interestingly from each other (see the top plots of Figs. 5 and 6). First, it seems that there is a statistically significant pattern in Period I: the opening of the HEX as well the the U.S. markets (New York) at 5.30 p.m. (CET+1) have caused some linear dependence. Of course, this finding does not necessarily imply any kind of arbitrage opportunities in economic sense. Indeed, when transactions costs are included a minor amount of autocorrelation is totally consistent with a *martingale* process and the notion of *efficient markets*. Considering the results of Period III, it seems that markets have become more liquid and efficient. A bit surprisingly, though, in Period I no significant negative autocorrelation of MA(1) type at lag one exists that is typically found and attributed to bid-ask bounce. In Period III, a significant negative first-lag autocorrelation -0.08 does appear (Andersen and Bollerslev (1997b) found it to be -0.04 in the FX markets with 5-minute data). It is then somewhat puzzling how increased liquidity would be related to the appearance of negative first-lag autocorrelation. One possibility is that the sizes of the trades have increased as well, thus causing the microstructure effects to last longer. In the subsequent analysis the MA(1) type of dynamics have not been considered that essential and therefore they have not been filtered out.

To proxy volatility, I decided to use absolute returns (as e.g. in Granger and Ding (1996) and Andersen and Bollerslev (1997c)) for several reasons. First, absolute returns are relatively outlier-resistant compared to squared returns (used e.g. in Lobato and Savin (1998)). In particular, *log*-squared returns would suffer from an "inlier" problem because a return very close to zero would result in a large negative number (although there are ways to deal with this kind of problem, see e.g. Fuller

[20]However, the two big "outliers" in Period III tend to balance the total difference so that the standard deviation of the returns of Period I is actually smaller than that of Period III: they are 0.3789182 and 0.386855, respectively. Similarly, the means of absolute returns (proxying volatility) are 0.1874 and 0.22870, respectively. Wavelet variances will shed more light on this apparently counterintuitive finding (see Sec. 5.5).

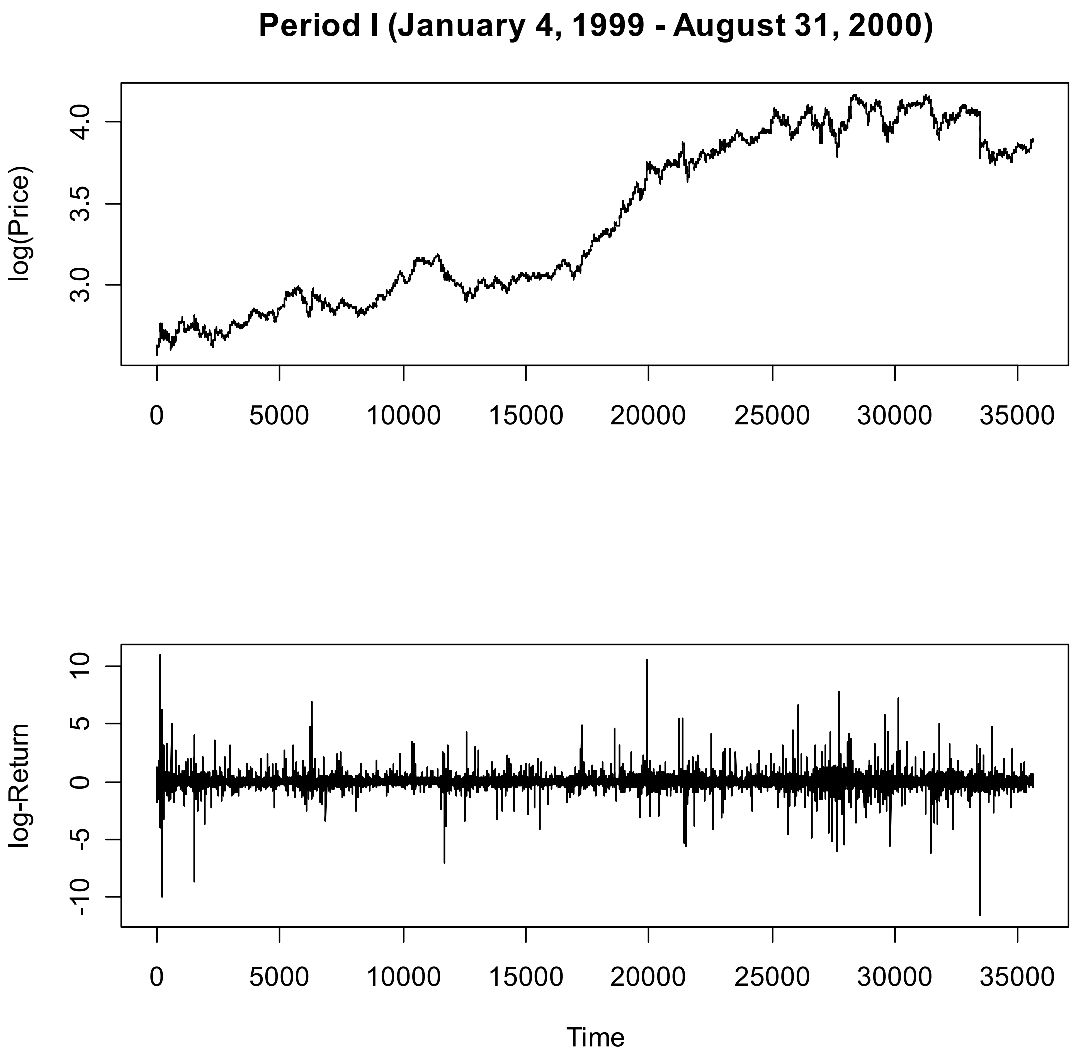


Figure 3: The price and return series of Period I. This period has a relatively restless outlook because of the large amount of intermediate sized jumps. Still, small pockets of tranquility are also visible.

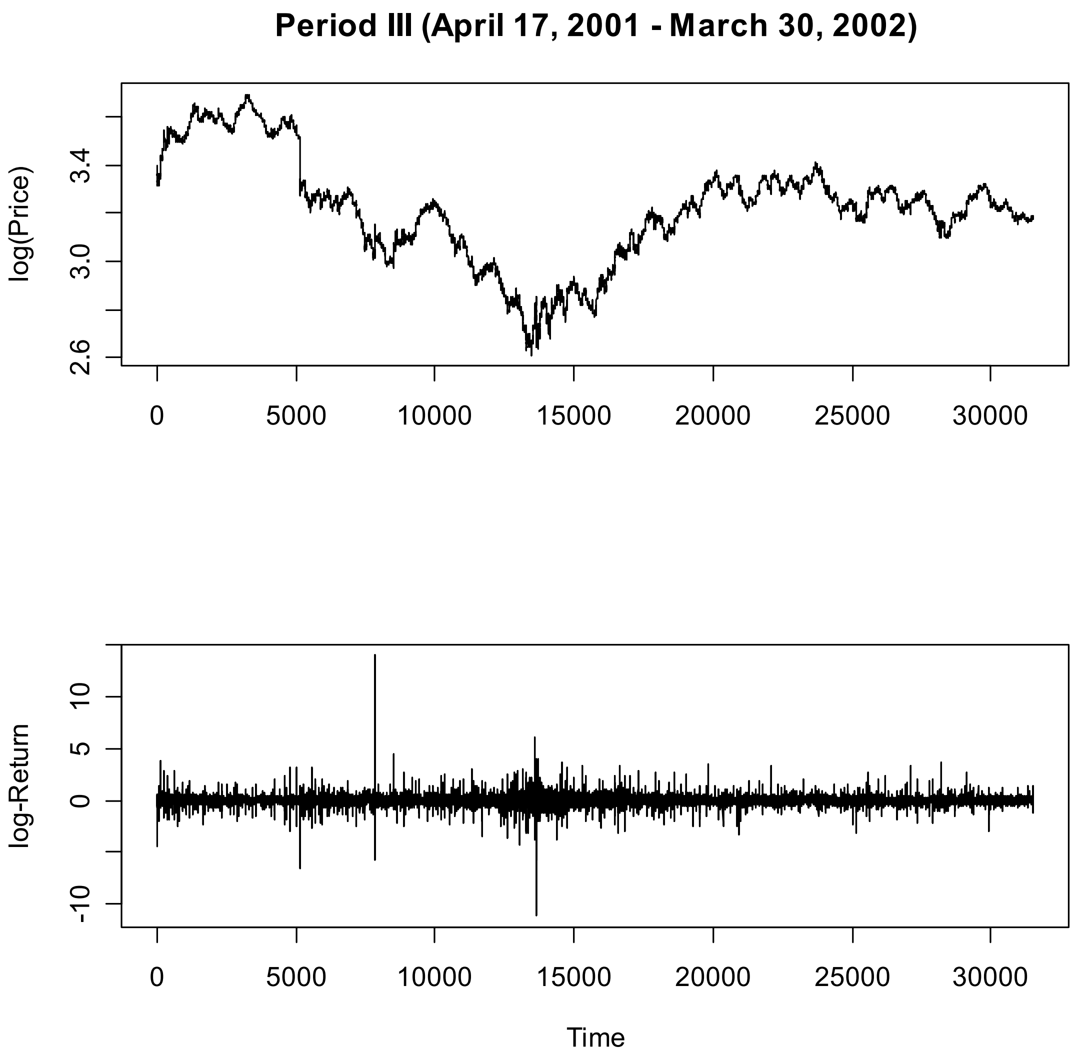


Figure 4: The price and return series of Period III. This period is relatively calm, there is only one major cluster of larger volatility between observation numbers $10,000$ and $15,000$.

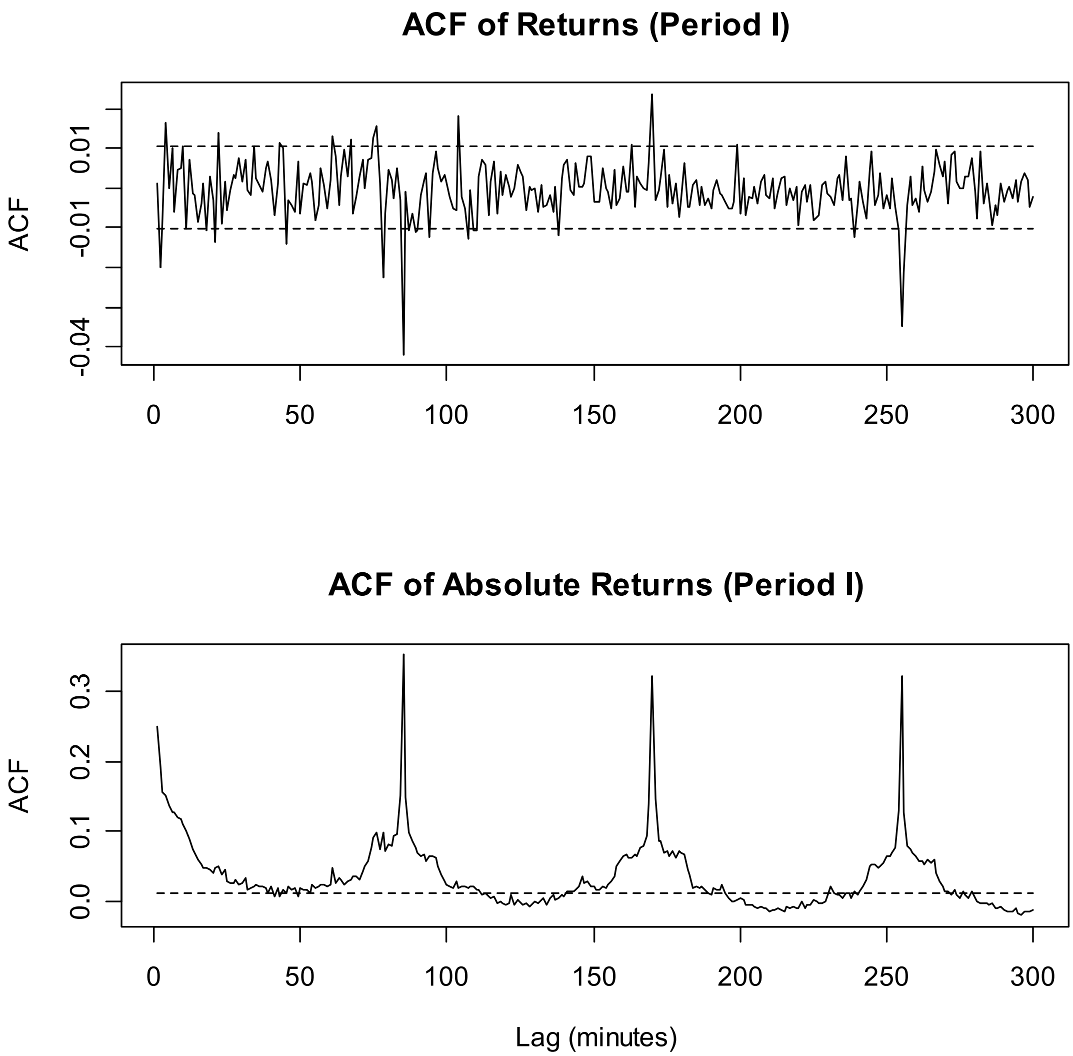


Figure 5: The sample autocorrelation function of returns and absolute returns in Period I. The 95% confidence interval (dashed line) is for Gaussian white noise: $\pm 1.96/\sqrt{N}$.

(1996)). Secondly, Wright (2000) has demonstrated that squared returns result in large downward bias when using semiparametric methods to estimate long-memory in the context of conditionally heavy-tailed data such as stock returns. He showed that this bias does not to appear with absolute returns. This is consistant with the result of Ding *et al.* (1993), Ding and Granger (1996), and Lobato and Savin (1998) that the long-memory property is strongest for absolute returns (in stock markets, at least). Finally, the well-known advantages of the more accurate realized variance could not be exploited in just 5 minutes, so absolute returns as an unbiased measure of the latent volatility remains the best alternative (see Sec. 6.1).

As expected, the results show that the sample autocorrelations of absolute returns stay significantly positive for a long time, in statistical as well as in economic sense (see the bottom plots of Figs. 5 and 6). In particular, in Period III the first-lag autocorrelation 0.32 is well above the confidence interval (Andersen and Bollerslev (1997b) found 0.309). Clearly, then, returns are *not* independent. Interestingly, the emerging intradaily "U-pattern" (sometimes referred to as the "inverse J") is quite similar in both periods. There are some important differences, however. First, the ACF peaks higher in Period I. This peak is caused by the large (in average) overnight return which is calculated over a much longer interval than just 5 minutes (this will be discussed in more detail in Sec. 7.6). The adjacent lags experience high autocorrelation in both periods. Second, the sudden drop before the peak in Period III seems to be caused by the closure of continuous trading. The reason why this slump does not exist in Period I is simply that evening trading did not take place then (see Sec. 7.1). Third, in Period I, the smaller peaks prior to the highest peak are attributable to the opening of the New York stock markets, i.e. they are the "New York effect".[21] Interestingly, in Period III no distinct New York effect exists in autocorrelations (see Sec. 7.6, however). This is probably due the weaker link between the U.S. and European markets after the burst of the IT-bubble. At this point one should be reminded that volatility spillover effects have been reported elsewhere, too. In fact, Engle *et al.* (1990) termed such contagion effects in the FX markets as *"meteor showers"* suggesting that different market places affect each other. A possible source of meteor showers are heterogoneous expectations (Hogan and Melvin (1994)). Similar U-pattern has been observed in the New York stock markets (see e.g. Wood *et al.* (1985), Harris (1986) and Lockwood and Linn (1989)) and the FX markets (see e.g. Andersen and Bollerslev (1997, 1997b), Baillie and Bollerslev (1991), Dacorogna *et al* (1993), Ito *et al.* (1996), and Bollerslev and Wright (2000)), although in the FX markets this periodicity is associated with the opening and closing of the various financial centers around the world.

The initial rapid decay of the sample autocorrelation followed by a very slow rate of dissipation (although the strong periodicity confounds this in the raw absolute re-

[21]Notice that all these effects would be blurred (not totally lost, however) if the sample autocorrelation would be calculated over the whole period of four years (1999–2002). This is the consequence of the New York effect and the changes in trading hours at the HEX.

turns, see Sec. 7.6) is characteristic of slowly mean-reverting fractionally integrated processes that exhibit hyperbolic rate of decay. As noted in several occasions concerning stock market volatility (see e.g. Bollerslev and Mikkelsen (1996), Breidt *et al.* (1998), Ding *et al.* (1993), and Granger and Ding (1996)), the common ARCH-models exhibiting exponential rate of decay (i.e. "short-memory") fail in this respect. The hyperbolic rate of decay will be quantified later with the fractional differencing parameter d (Sec. 7.4).

7.3 Multiresolution decomposition

This subsection takes a first-hand multiresolution look at Periods I and III. Such a decomposition is expected to give more insight into the evolving dynamics of stock market volatility. The analysis conducted below is also intended to highlight some of the aspects of wavelet theory described in the previous sections.

A first-hand look at the long-run dynamics of the data is achieved conveniently by an MRA. In specific, a MODWT MRA($J = 14$) of price series of Periods I and III using a LA(8) filter (with reflecting boundary) produces a set of wavelet smooths with varying amount of details included (see Figs. 7 and 8). These smooths show, for example, that Period I has a strong upward trend. Notice that all the smooths are automatically aligned in time with the original series. Furthermore, these smooths converge to the original price series as more and more details are being added.[22] Concerning volatility, however, very little can be inferred from these figures.

In order to study volatility, the MODWT($J = 12$) is performed to absolute returns using LA(8) (with reflecting boundary). Instead of focusing on the scaling coefficients or smooths as above, the wavelet coefficients are now of interest (see Fig. 9 for even levels $j = 2, 4$ and Fig. 10 for $j = 6, 8, 10$ in Period I). Now the approximate zero-phase filter property (i.e. alignment in time) becomes much more apparent: big changes in volatility stand out at the smallest scales (i.e. highest frequencies). As the scale gets bigger (i.e. frequency lower), rapid changes tend to be smoothed out because a wider filter averages more. For example, in Period III the large spike in volatility between observations $5,000$ and $10,000$ has died out already at the 6th level (see Fig. 13). On the other hand, the spike between $10,000$ and $15,000$ continues to prevail even at the 10th level. This means that the former spike was a high-frequency event only, while the latter was a more severe and long-lasting burst of volatility. Thus from a purely non-firm specific angle, short-time speculators and long-term investors would have to react differently in such an event: the former would be a concern for speculators while the latter would interest investors as well. Moreover, because the wavelet coefficients should in theory form a stationary series at each level (see Sec. 5.5), the same general characteristics should persist in the future, as well. This does not imply, however, that for example the "cycle" visible at the end of the 10th level

[22]These "moving averages" could be applied in, e.g., forecasting in the spirit of "double" and "triple crossing" methods (methods that are shortly discussed in Gençay *et al.* (2002a, pp. 48–49)).

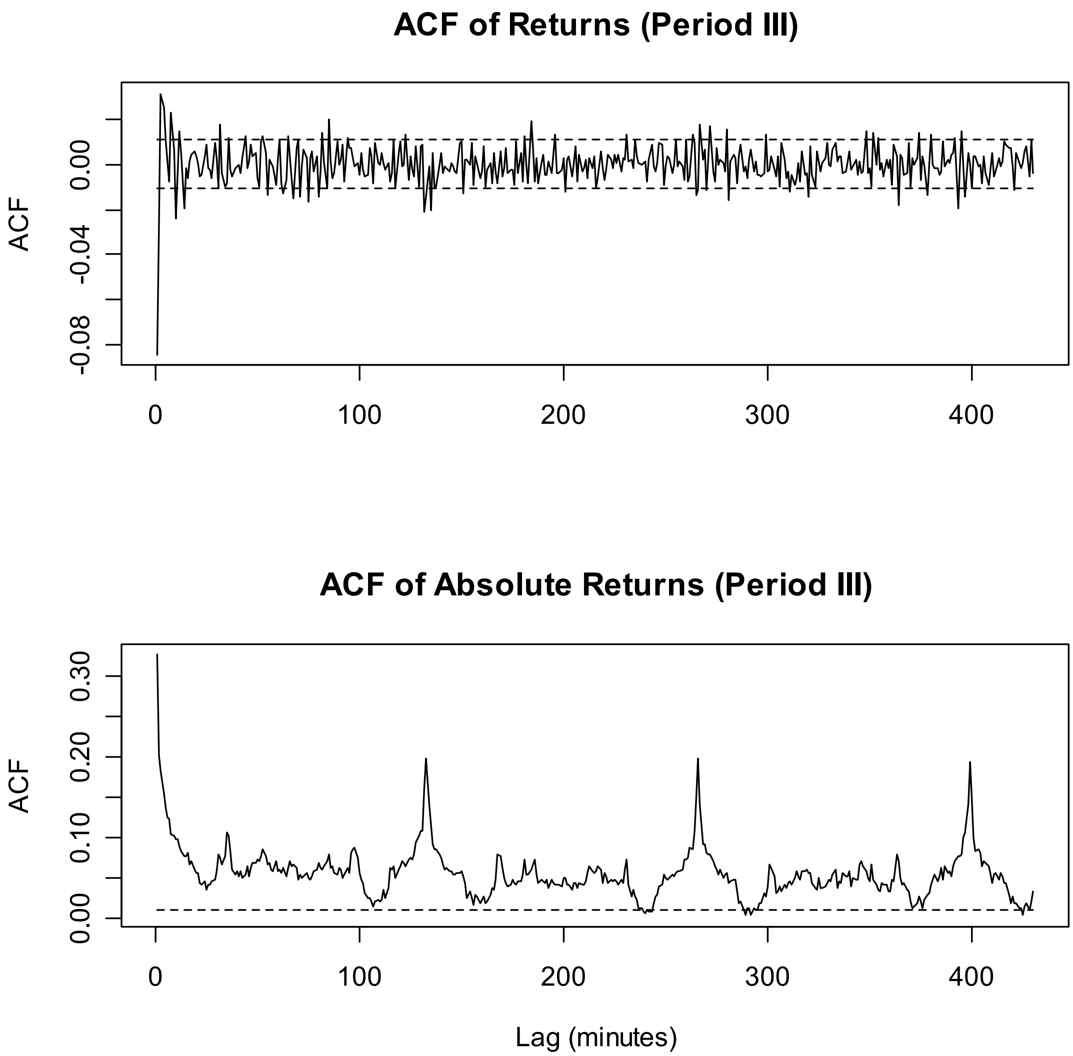


Figure 6: The sample autocorrelation function of returns and absolute returns in Period III. The 95% confidence interval (dashed line) is for Gaussian white noise: $\pm 1.96/\sqrt{N}$.

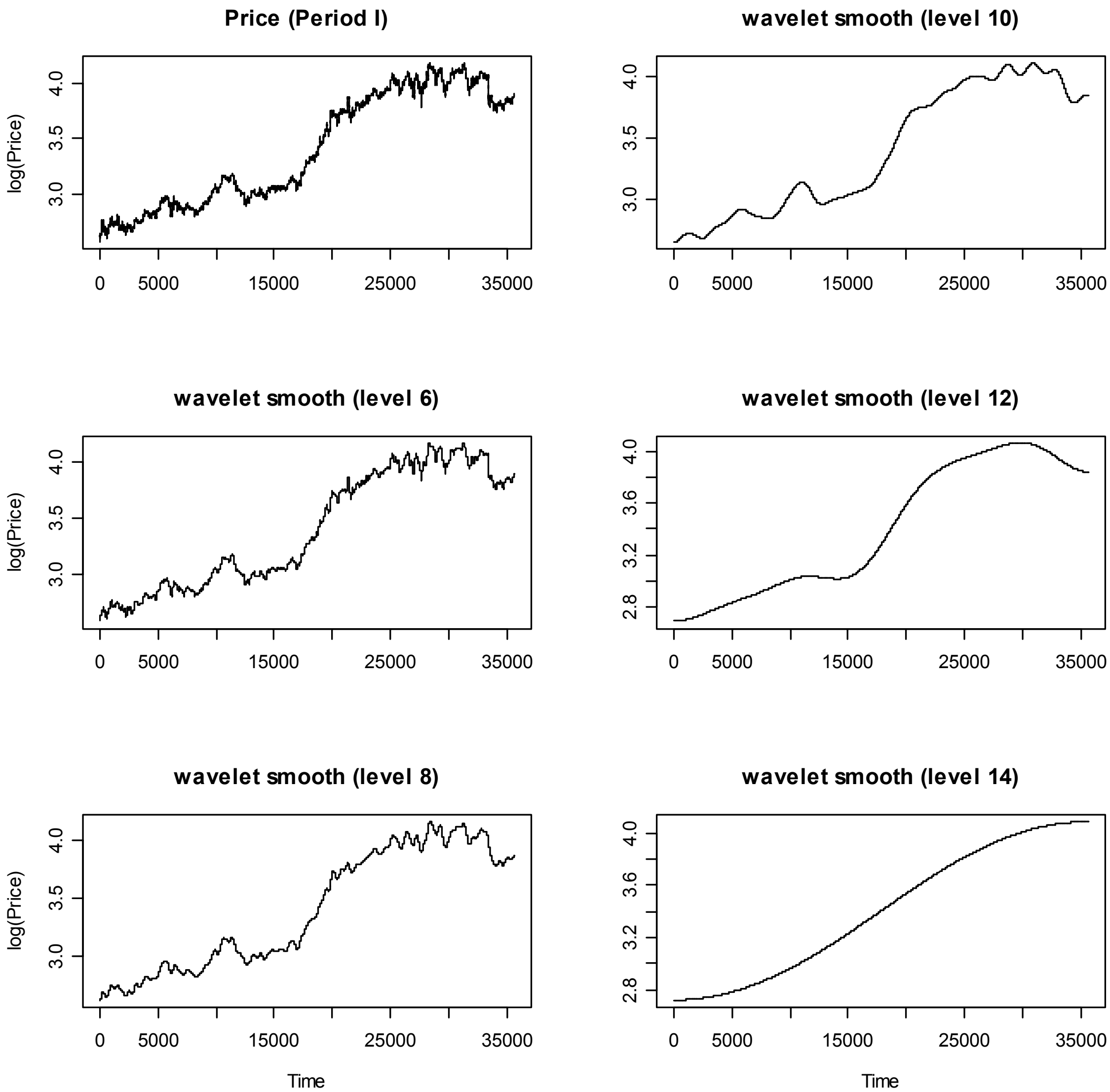


Figure 7: The original price series of Period I (left upper corner) and its wavelet smooths of varying levels. As less and less details are included, the smoother the outcome (right lower corner).

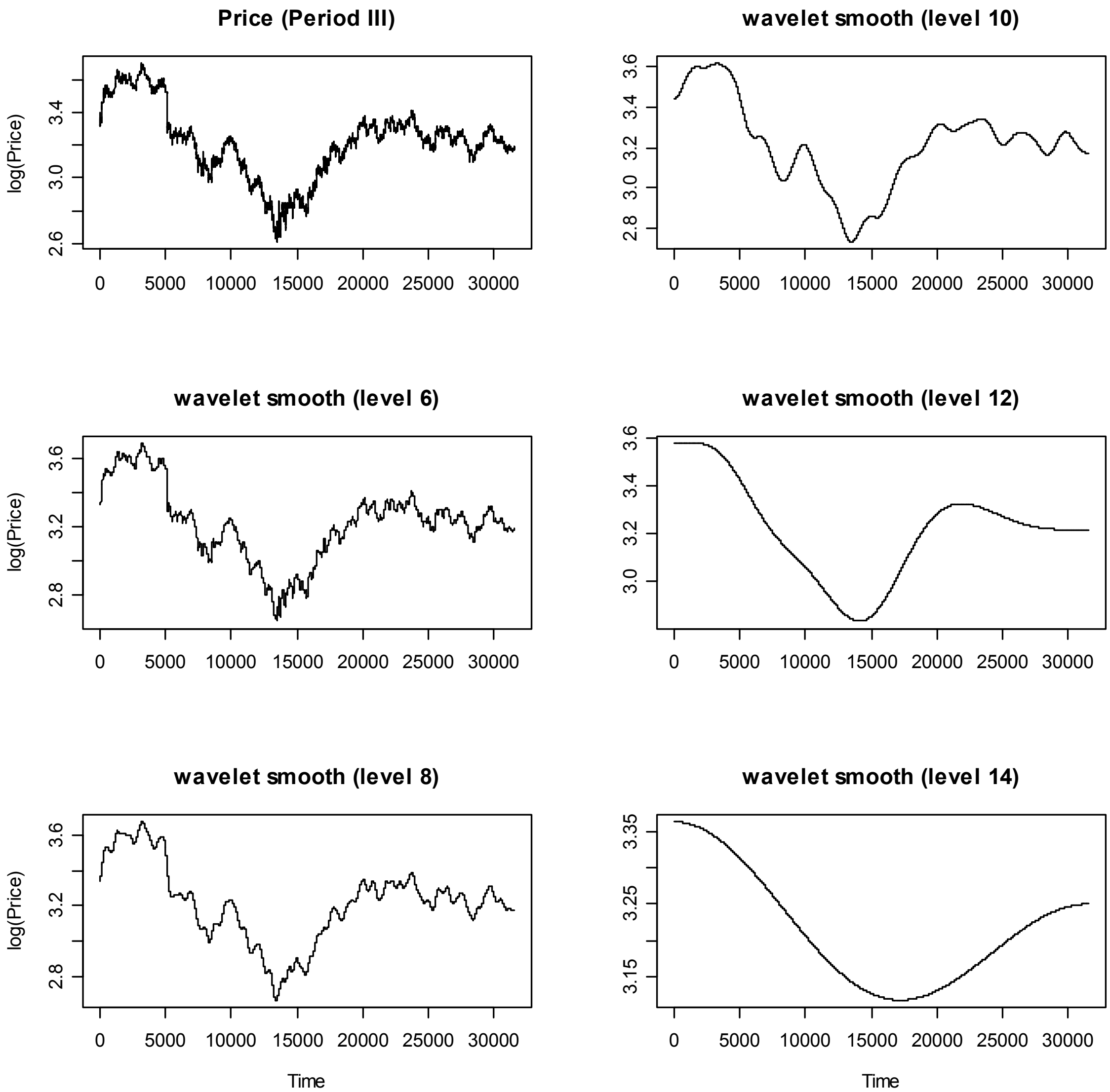


Figure 8: The original price series of Period III (left upper corner) and its wavelet smooths of varying levels. As less and less details are included, the smoother the outcome (right lower corner).

of Period I is something that could be used for forecasting. Although it is not a pure artifact of the wavelet filter, there is simply no reason why such a cycle should persist in an efficient stock market.

The first 12 wavelet levels with the corresponding scales and associated changes are listed below (see Table ??). Notice that I follow Daubechies' indexing so that a bigger level is associated with larger time-scale (in contrast to what I used in Sec. 4). An unfortunate consequence of dyadic scaling is that it becomes coarse fast, so that some interesting dynamics might be lost (alternative scaling has been proposed by Pollock (??)). In interpreting the time-scales in "market time", one should be careful with the period in question since the length of the trading day has varied considerably during the four years (see Sec. 7.1). For instance, in Period I the first 6 levels correspond to intraday dynamics capturing frequencies $1/64 \leq f \leq 1/2$, i.e. oscillations with a period of $10-320$ minutes (approx. 5 hours). In Period III, on the other hand, the first 7 levels correspond to intraday dynamics capturing frequencies $1/128 \leq f \leq 1/2$, i.e. oscillations with a period of $10-640$ minutes (approx. 11 hours). In particular, in Period I, the 6th level corresponds to approximately a half of a trading day. In Period III a half of a trading day corresponds to the 7th level. These levels are possibly a watershed between intraday and interday dynamics.

Level	**Scale**	**Associated with changes of**
1	1	5 min.
2	2	10 min.
3	4	20 min.
4	8	40 min.
5	16	80 min.
6	32	160 min. $\approx$ 3 h.
7	64	320 min. $\approx$ 5 h.
8	128	640 min. $\approx$ 11 h.
9	256	1280 min. $\approx$ 21 h.
10	512	2560 min. $\approx$ 43 h.
11	1024	5120 min. $\approx$ 85 h.
12	2048	10240 min. $\approx$ 171 h.

This far the analysis has been mainly descriptive. Although no scale-specific quantitative conclusion concerning the energy could not be made based on the MODWT MRA, the MODWT coefficients lend themselves to a quantitative study (see Eq. (2)). Few general observations can be made immediately. It is for example seen that the unconditional distributions at the 12 levels show convergence from a highly leptokurtic distribution to a Gaussian one (see Figs. 11 and 14) which is confirmed by the Jarque–Bera test statistic (see Table ??). Furthermore, statistical key figures confirm that the mean stays zero at all levels (a consequence of the zero average property of wavelets) while the changes get smaller in absolute value (see Table ??).

A closer quantitative look at the unconditional wavelet variance is the topic of the next subsection.

Period I

j	**Min.**	**1st Q.**	**Median**	**Mean**	**3rd Q.**	**Max.**
1	-4.868	-0.05475	$-4.776e-03$	$-9.334e-18$	0.04743	6.067
2	-2.931	-0.04164	$-2.733e-03$	$-9.391e-18$	0.03772	4.182
3	-1.769	-0.03490	$-1.913e-03$	$-9.389e-18$	0.03158	2.641
4	-1.007	-0.03249	$-2.538e-03$	$-9.381e-18$	0.02580	1.442
5	-0.7617	-0.03709	$-7.627e-03$	$-9.356e-18$	0.02695	1.181
6	-0.6765	-0.05384	$-2.553e-03$	$-9.555e-18$	0.05114	0.7713
7	-0.3751	-0.01752	$-5.763e-04$	$-9.356e-18$	0.01723	0.5056
8	-0.1974	-0.01794	$-7.311e-04$	$-9.336e-18$	0.01547	0.3057
9	-0.1334	-0.01359	$-1.204e-03$	$-7.994e-18$	0.01070	0.1726
10	-0.08500	-0.01775	$-1.569e-03$	$-7.069e-18$	0.01372	0.08984
11	-0.06619	-0.01798	$-3.173e-03$	$-4.563e-18$	0.01349	0.08682
12	-0.06628	-0.02302	$-1.341e-03$	$-1.228e-17$	0.02361	0.06722

Period III

j	**Min.**	**1st Q.**	**Median**	**Mean**	**3rd Q.**	**Max.**
1	-4.964	-0.07072	$-6.457e-03$	$-1.157e-17$	0.06400	6.669
2	-3.909	-0.05304	$-3.357e-03$	$-1.155e-17$	0.04888	5.670
3	-2.329	-0.04397	$-2.212e-03$	$-1.149e-17$	0.03913	3.390
4	-1.215	-0.03808	$-3.502e-03$	$-1.140e-17$	0.03178	1.768
5	-1.065	-0.03987	$-3.898e-03$	$-1.092e-17$	0.03720	1.411
6	-0.6380	-0.03441	$-1.539e-03$	$-1.179e-17$	0.03263	0.7801
7	-0.4323	-0.02711	$-9.540e-04$	$-1.190e-17$	0.02490	0.4504
8	-0.2017	-0.01848	$-1.213e-03$	$-1.170e-17$	0.01658	0.3167
9	-0.1679	-0.01843	$-1.911e-03$	$-1.278e-17$	0.01632	0.2068
10	-0.07587	-0.01253	$-5.021e-04$	$-1.240e-17$	0.01174	0.1299
11	-0.09427	-0.01174	$-8.843e-04$	$-1.798e-17$	0.01048	0.1225
12	-0.07697	-0.01674	$-1.734e-03$	$-1.062e-17$	0.01239	0.1108

7.4 Global scaling laws and long-memory

There is no privileged time horizon in financial decision making. As a consequence, financial risk is often assessed at different horizons varying from small (minutes) to long (months). In risk management industry it is typical to convert risk measures (such as standard deviation) at small time-scales to large time-scales by square-root scaling. For example, a standard deviation calculated from daily returns is converted to weekly standard deviation by multiplying the daily standard deviation by $\sqrt{5}$. However, this simple square-root scaling is valid only if the underlying data is IID

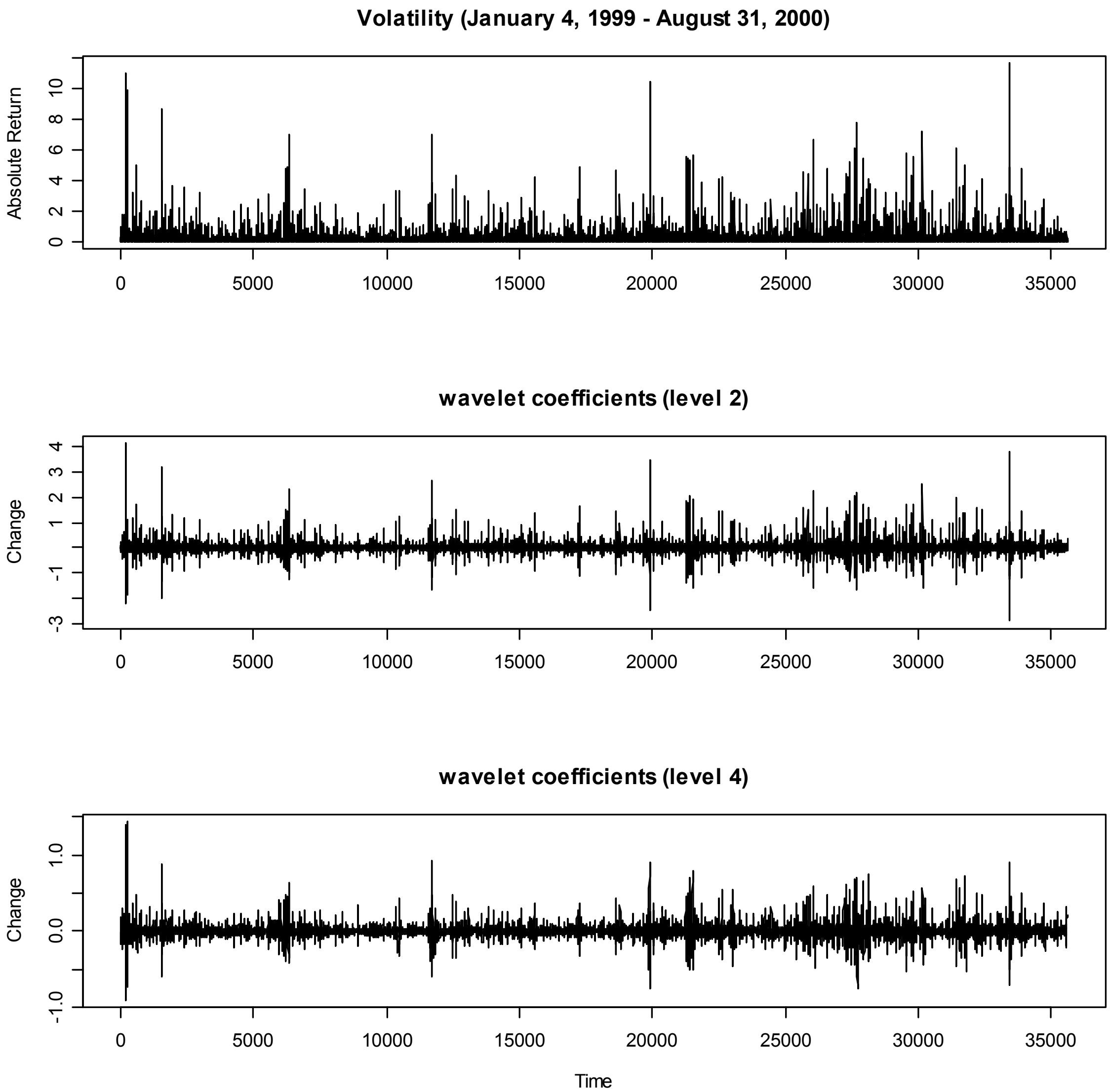


Figure 9: The original absolute return series proxying volatility in Period I (the top subplot) and the MODWT($J = 12$) wavelet coefficients of levels $j = 2$ and 4 using LA(8) and reflecting boundary.

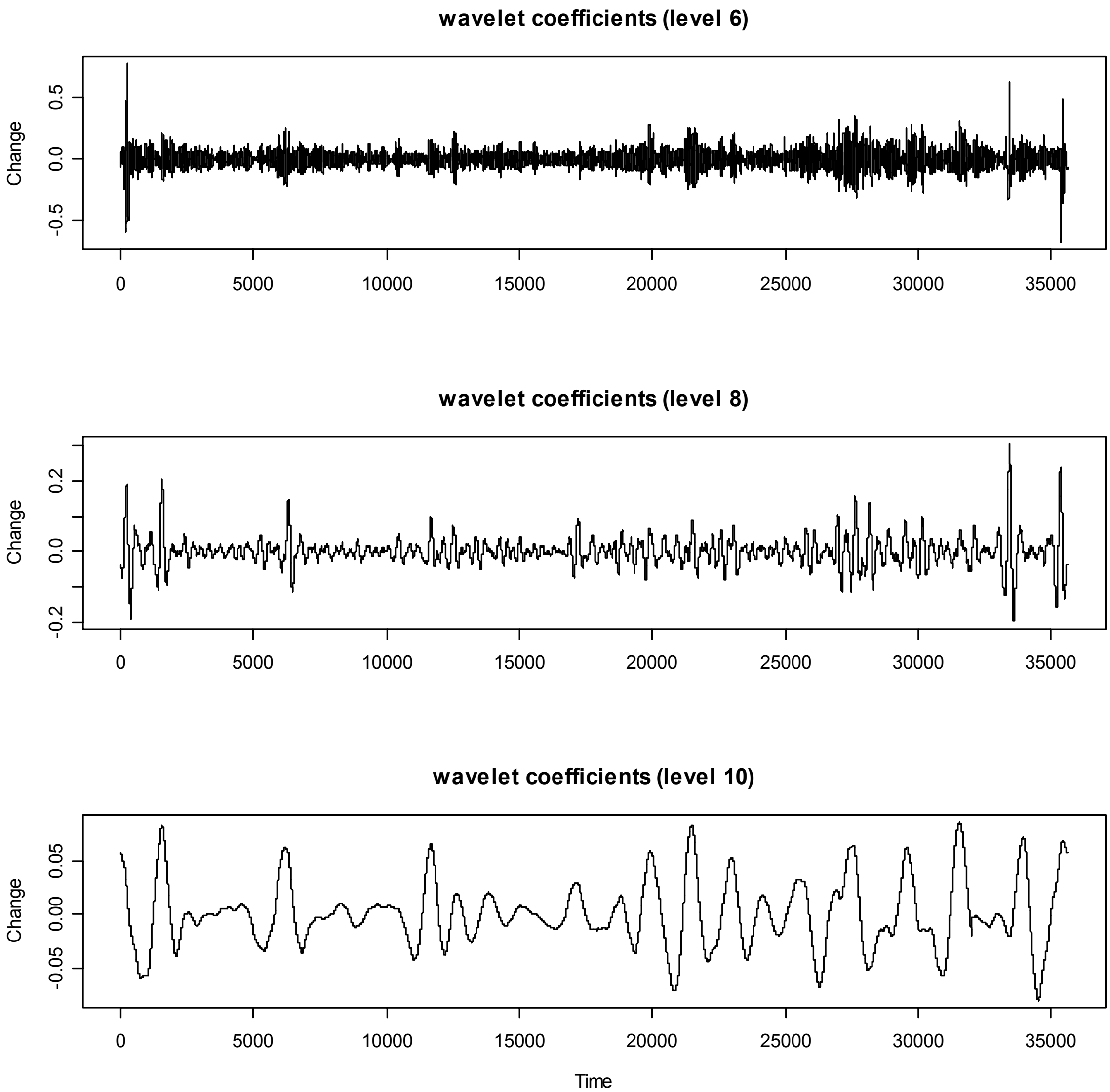


Figure 10: The Period I MODWT($J = 12$) wavelet coefficients of levels $j = 2$ and 4 using LA(8) and reflecting boundary.

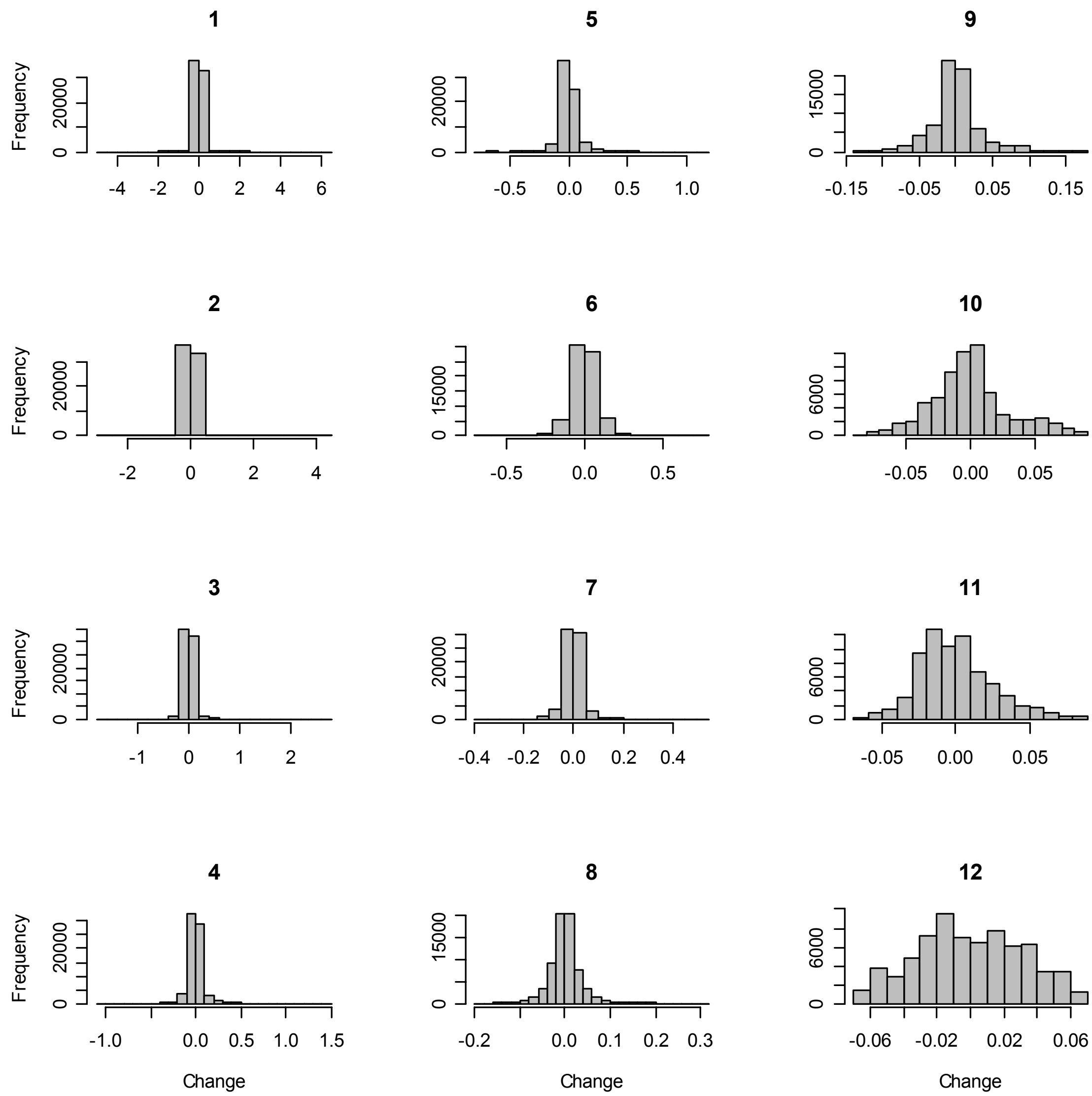


Figure 11: The unconditional distributions of the Period I MODWT wavelet coefficients across levels $j = 1, ..., 12$. Notice the change in the shape of the distribution from the small levels to large levels.

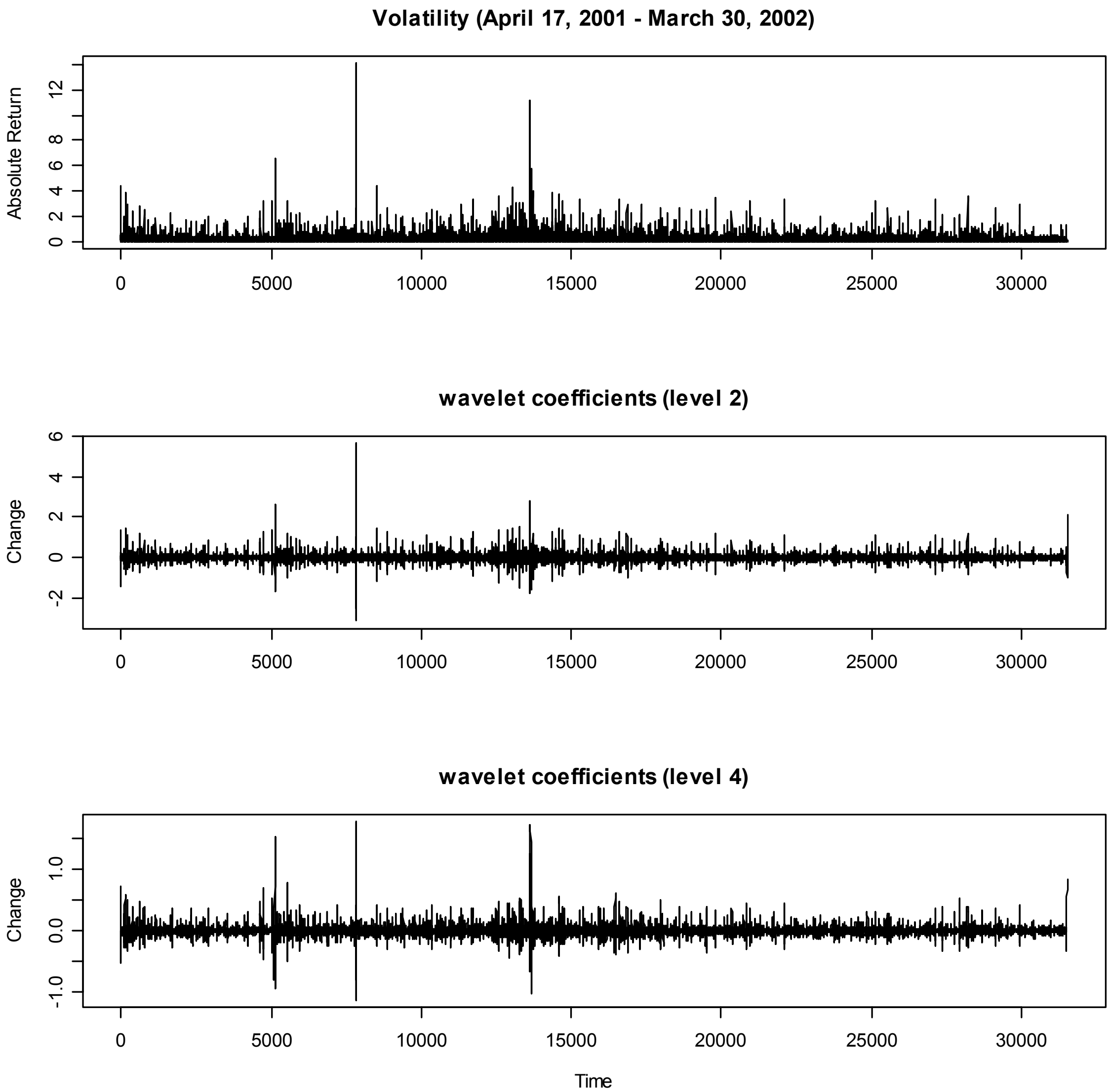


Figure 12: The original absolute return series proxying volatility in Period III (the top subplot) and the MODWT(J = 12) wavelet coefficients of levels j = 2 and 4 using LA(8) and reflecting boundary.

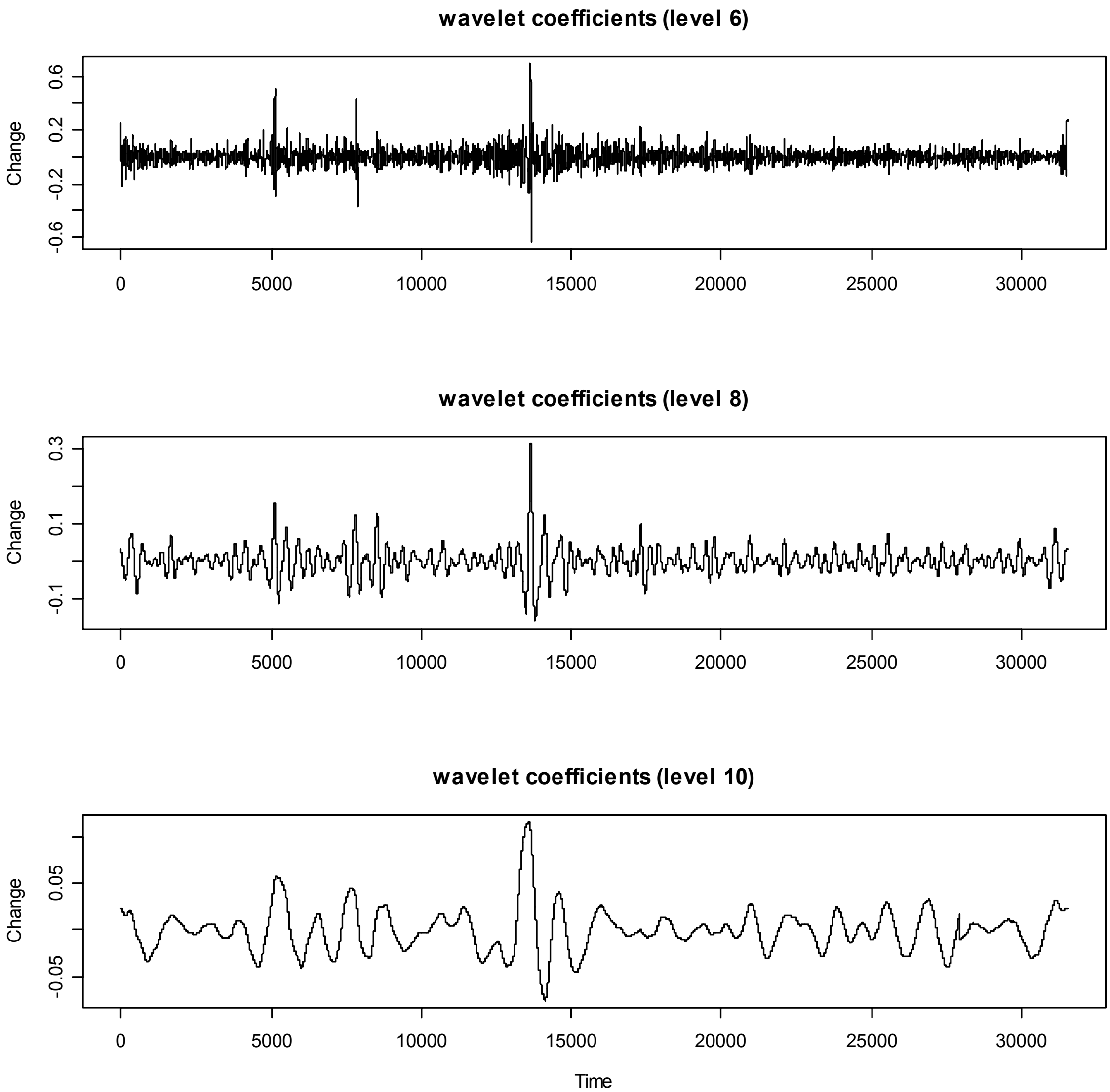


Figure 13: The Period III MODWT($J = 12$) wavelet coefficients of levels $j = 6, 8$ and 10 using LA(8) and reflecting boundary.

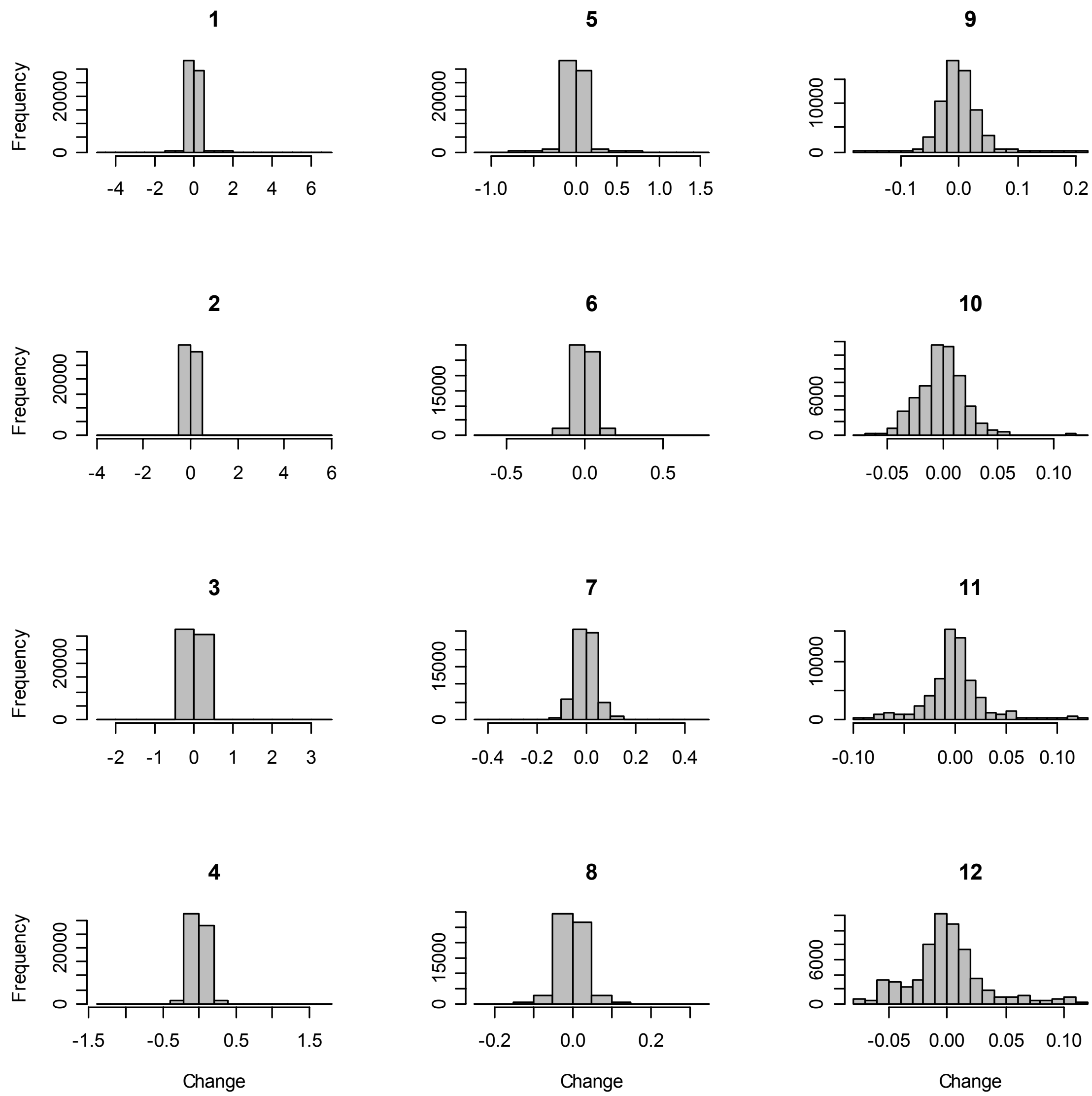


Figure 14: The unconditional distributions of the Period III MODWT wavelet coefficients across levels $j = 1, ..., 12$. Notice the change in the shape of the distribution from the small levels to large levels.

which clearly does not hold for financial time series (see Diebold *et al.* (1998)). Several authors have provided evidence of scaling laws in the FX markets (for example Müller *et al.* (1990, 1995), Guillaume *et al.* (1997), and Andersen *et al.* (2000)) but a bit less so in the stock markets.

However, a single scaling factor might be appropriate only in a subset of time-scales. Namely, it has been argued that intradaily dynamics are different from the dynamics of larger time-scales. To study this "multi-scaling" in the FX markets, Gençay *et al.* (2000) used wavelet methodology based on the argument that it is specification free and robust to a possible misspecification of the distribution of returns. Using absolute returns as volatility proxy, they confirmed that a different scaling regime exists for intradaily time-scales than does for interday and larger time-scales. Based on their findings, it is thus interesting to see if (i) the same phenomena appears in stock market data, (ii) the scaling factor is stable in time (i.e. the same in Periods I and III), and if (iii) there is any reasonable explanation for its magnitude and structure. Although the first point has already been studied in the literature, the next two have not (and the first one rarely with wavelets).

In contrast to Gençay *et al.* (2000) who used returns sampled every 20 minutes to reduce the impact of microstructure effects (originating from the buying and selling intentions of quoting institutions), I decided to use 5-minute returns. As the bid-ask bounce effect may be now more severe, special care has to be taken when interpreting the results. So although no large first-lag negative autocorrelation was observed in neither period (significant only in Period III, see Sec. 7.2), the results from the smallest time-scales (especially the 1st) should be interpreted cautiously. Fortunately, wavelet methodology is very flexible in this respect and there is no *ex ante* reason in exlucing any particular time-scale from the analysis (it may be done *ex post*). In fact, including a time-scale as small as possible (5-minutes here) is quite reasonable to see all the effects. The second important point to keep in mind is that the strong intraday periodicity (reported in Sec. 7.2) may have confounding effects. Thus the results concerning the intraday levels from 1 to 7 should be regarded as preliminary until the periodicity has been taken care of (in Sec. 7.6). Finally (as explained in Sec. 6.1), absolute returns provide a rather noisy measure of the latent volatility, so the following results might a bit different under a different volatility measure.[23]

Because the MODWT coefficients preserve the total energy of a time series, it is sensible to analyze time-scale specific energy (cf Eq. (2)). In the current analysis, the MODWT coefficients of absolute returns of Periods I and III were formed by the MODWT($J = 12$) using LA(8) filter. Interestingly, by assuming the boundary to be reflecting instead of periodic resulted in noticiable different results at the largest

[23]Realized volatility (e.g. daily) cannot be comfortably used in this context because intradaily dynamics are the subject of interest and very frequently sampled returns would suffer excessively from microstructure effects. Recently, however, Hansen and Lunde (2004) have shown that a Newey–West type of correction of realized variance yields an *unbiased* measure of volatility even if intraday returns are sampled every second!

scales (see the top plots of Fig. 15). Most likely this is because at the largest scales the filter is so wide that neither type of specification holds very well and the boundary effects prevail (seen as a sudden growth in the wavelet variance at levels 11 and 12). However, in the current context the reflecting boundary assumption was considered as more natural. The Gaussian confidence bands were calculated only in this case (see the bottom plots of Fig. 15). Notice that the non-Gaussian confidence intervals would be a bit wider.

From the results of Period I (see the left hand part of Fig. 15) it is obvious that most of the total energy of volatility is located at the smallest time-scales, i.e. the highest frequencies.[24] It is the good localization properties of wavelets that are able to reveal this. Moreover, it is seen that energy decreases as the function of frequency: the lower is the frequency, the lower its energy content. In fact, the relationship is approximately hyperbolic which is observed as an approximate linear relationship on a double-logarithmic scale. More precisely, there would seem to exist two different scaling regions in Period I, namely levels $j_1 = 1, ..., 6$ and $j_2 = 7, ..., 10$. The first six levels capture frequencies $1/64 \leq f \leq 1/2$, i.e. oscillations with a period of 10–320 minutes. In Period I these levels correspond to intraday dynamics. Interestingly, there is a visible break in the scaling law at the seventh level associated with 320-minute changes or oscillations with a period of approximately 640 minutes. The seventh and higher levels are related to one day and higher dynamics in Period I.

To study if the break in the scaling law is stable, Period III is analyzed similarly. Unfortunately, there is no similar "kink" at any level in Period III (see the right hand part of Fig. 15). One might have naively expected a break at the 8th level because of the considerable longer trading day (see Sec. 7.1). But the relationship is almost perfectly linear across all levels up to 10; there is only a very slight kink at the 6th level. The difference between the scaling laws of Periods I and III is most evident at this particular level: the wavelet variance of Period I seem to be significantly larger than in Period III (see Fig. 16). Indeed, the 6th level wavelet variance of Period I lies outside the Period III's Gaussian 95% confidence interval. The difference is visible by eye when the wavelet coefficients are plotted together (see Fig. 17). It seems that Period I experienced more turbulence in the corresponding time-scale than Period III did. However, this simple observation is not yet enough to explain the more jumpy look of Period I. Because sudden jumps are high-frequency events, they should be well captured by the 1st level. This intuition is confirmed by the 1st level wavelet variance of Period I which lies outside the confidence interval, as well (see Fig. 16). In fact, the difference is now even more pronounced (see Fig. 18). Based on these findings it can then be concluded that the "more volatile" outlook of Period I is mainly caused by the different dynamics at levels 1 and 6 corresponding to 5-minute and approximately 5-hour changes, respectively. Thus the difference in the overall level of volatility can be attributed to specific time-scales and intraday speculators.

[24] Or as D.S.G. Pollock suggested to me at the "Workshop on Computational Econometrics and Statistics" (Neuchâtel, Switzerland 2004), financial markets tend to "shriek" under stress.

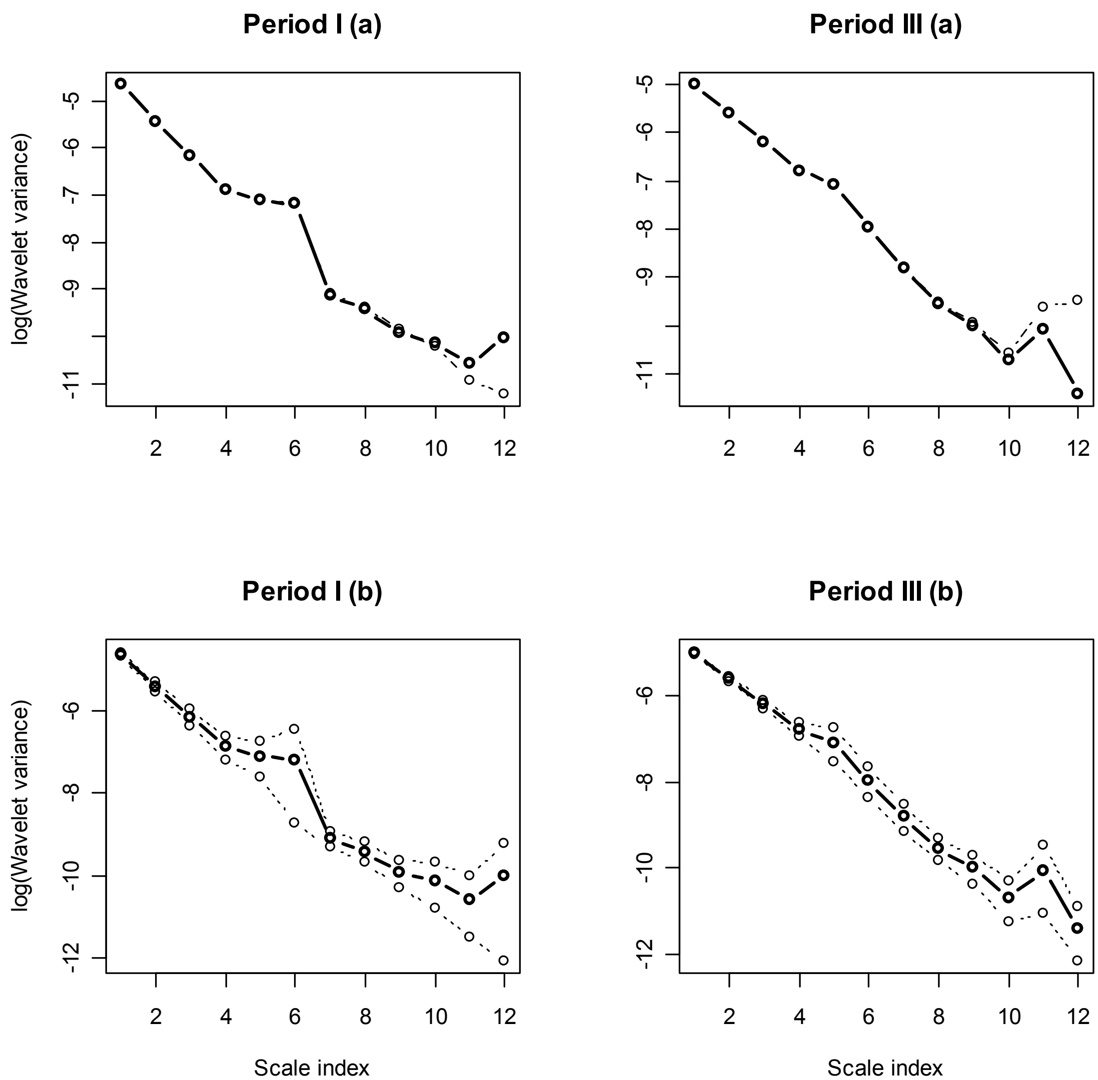


Figure 15: Wavelet variances of Periods I (on left) and III (on right) on a double-logarithmic scale. The upper plots show the result using reflecting (continuous line) and periodic boundary (dotted line). The lower plots show the Gaussian 95% confidence intervals with reflecting boundary only.

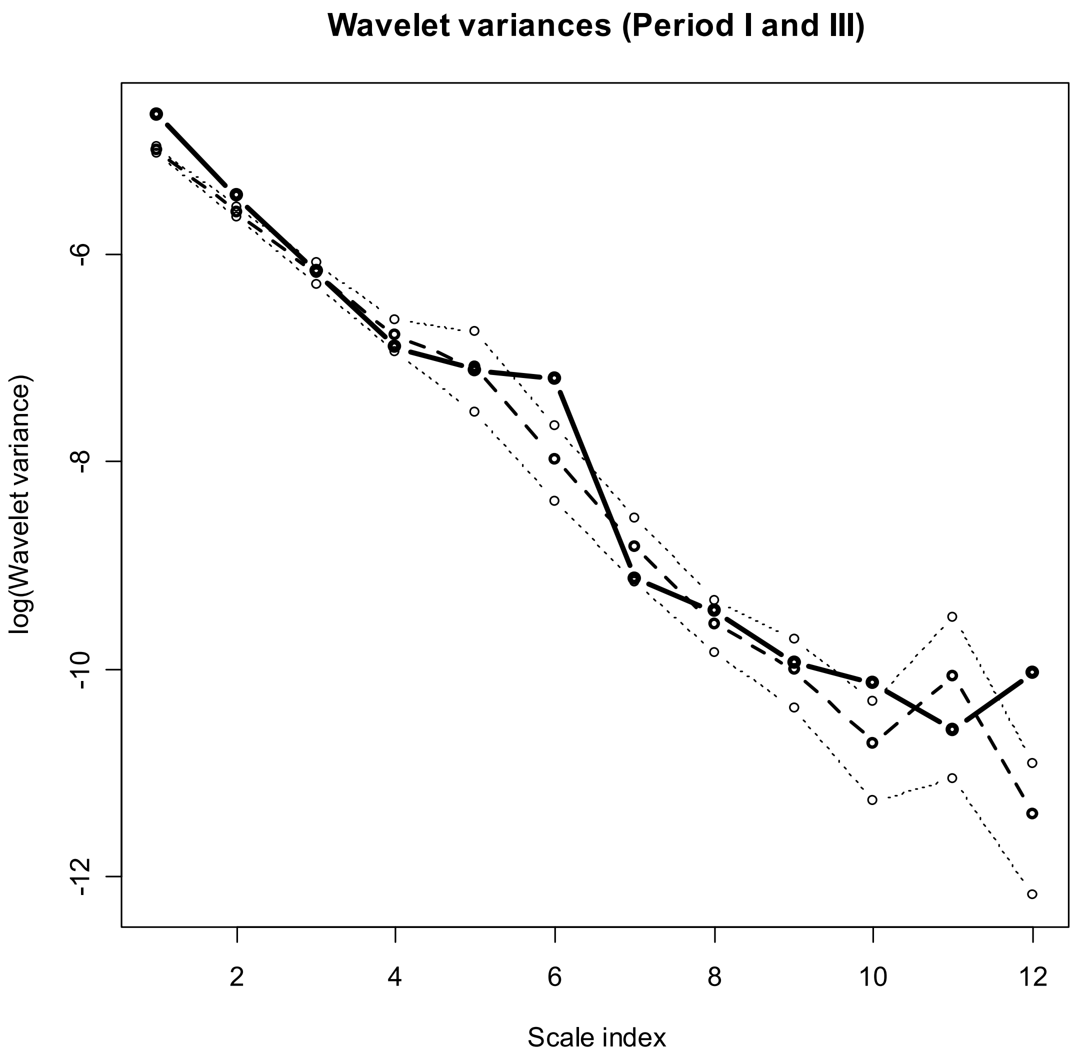


Figure 16: The wavelet variances of Periods I (continuous line) and III (dashed line) on a double-logarithmic scale using LA(8) with reflecting boundary. Gaussian 95% confidence interval (dotted line) of Period III has been drawn, too. In particular, notice that the 6th level wavelet variance of Period I is significantly different.

The jumps at the 1st level measure the flow of new information and the general level of nervousness. The wavelet variance at level 6 is not so easily interpretable, however. The difference in this particular level may be just an artifact of the strong volatility seasonality (see Sec. 7.2). The effect of this seasonality is studied later (in Sec. 7.6). Notice, however, the wavelet variances at levels 3 and 5 are almost identical.

As noted earlier (in Sec. 7.2), the slow decay of the sample autocorrelations of the absolute returns is suggestive of long-range dependence. This dependence is now being quantified using a wavelet variance approach. From a statistical point of view this quantification is important as standard statistical tools for inference are invalid in the case of long-memory. For example, standard errors for the estimates of the coefficients of ARCH or stochastic volatility models would be incorrect and hence the confidence intervals for predictions (Lobato and Savin (1998); see also Beran (1994)). Economically long-memory has consequences for option pricing. For instance, long-memory has a significant impact upon the term structure of implied volatilities (see Taylor (2000)). And of course, estimation of the fractional differencing parameter d allows the use of long-memory ARCH (e.g. FIEGARCH) and stochastic volatility (e.g. LMSV) models for simulation and forecasting.

The semiparametric wavelet domain method (see Sec. 5.5) is able to ignore the intraday periodicities and hence it is in principle ideally suited for estimating the rate of this decay. However, the relatively short time-span used (approx. 1.5 years) could well be criticized and it has been a topic of hot debate in past years. It is true that when estimating long-memory dependencies in the *mean*, the small sample bias depends crucially on the time span of the data. But the most recent evidence (see Andersen and Bollerslev (1997, 1997b) and Bollerslev and Wright (2000)) suggests that the performance of the estimates from the *volatility* series may be greatly enhanced by increasing the observation frequency instead of time-span. In particular, Bollerslev and Wright (2000) have argued that high-frequency data allows for vastly superior and nearly unbiased estimation of the fractional differencing parameter d. For example, Andersen and Bollerslev (1997) (using realized volatility as their volatility measure) found that already 3 months may be enough to obtain an accurate estimate of d when the intraday periodicity has been first properly taken care of. Thus relatively short time spans of high-frequency data should be very informative about long-run volatility dependencies, too. This argument is intimately related to the seminal observation of Merton (1980) that high-frequency data may greatly enhance estimates of volatility (cf. Sec. 6.1).

Using the linear relationship of $\log \nu_X^2(\lambda_j)$ and $\log 2^j$ (see Eq. (6)), the estimation of d is done for Periods I and III by OLS. The same approach has been used before by Jensen (2000) and Tkacz (2000), for example. Furthermore, a similar type of OLS procedure using the Fourier periodogram has been carried out by Andersen and Bollerslev (1997b, 1997c), Ray and Tsay (2000), and Wright (2002), among others. Following Ray and Tsay (2000), the standard errors obtained from regression theory are now being used to judge the significance. This is because in the context

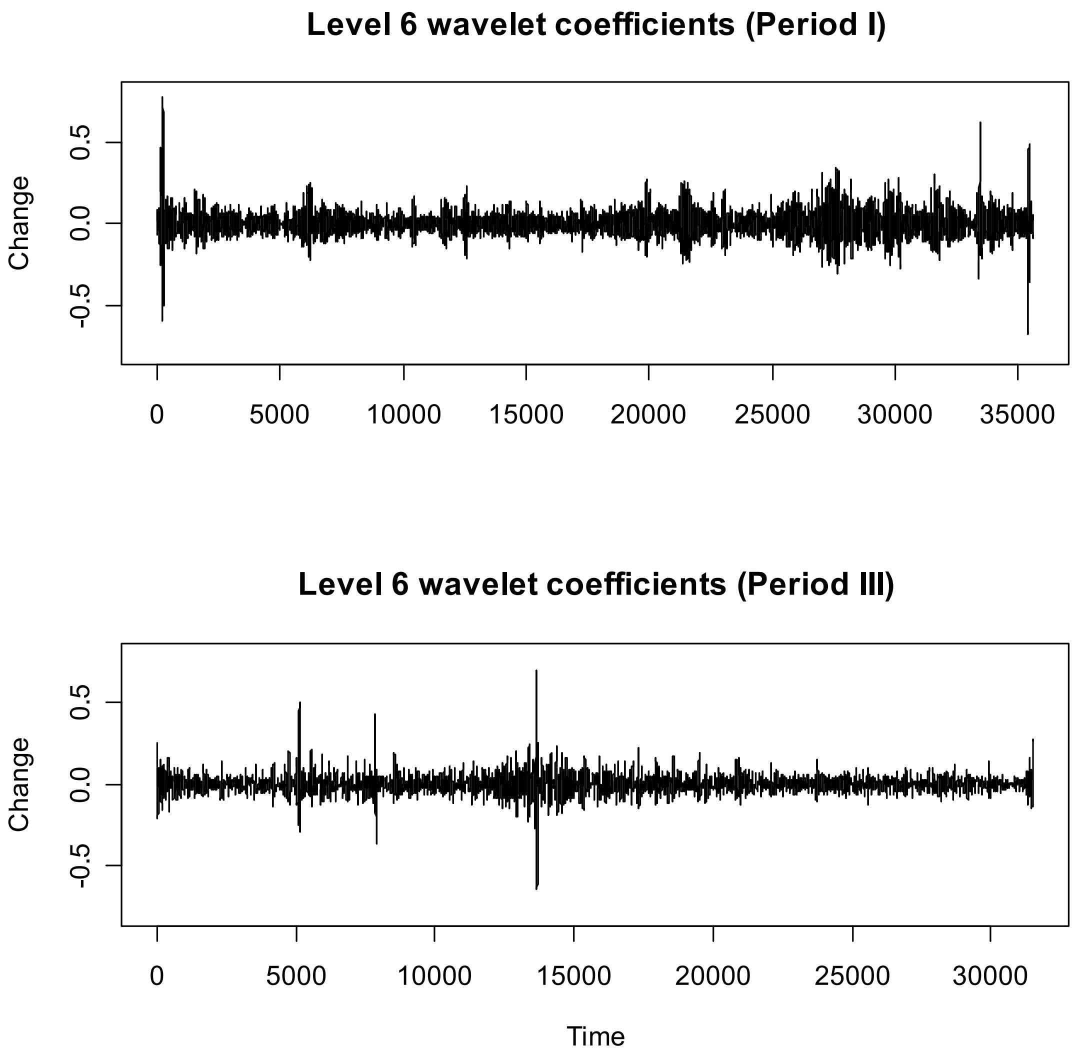


Figure 17: The 6th level MODWT wavelet coefficients of absolute returns of (a) Period I and (b) Period III. The LA(8) filter with reflecting boundary was used. Notice that the outlook of (a) is fatter. Indeed, its wavelet variance is significantly larger than in (b).

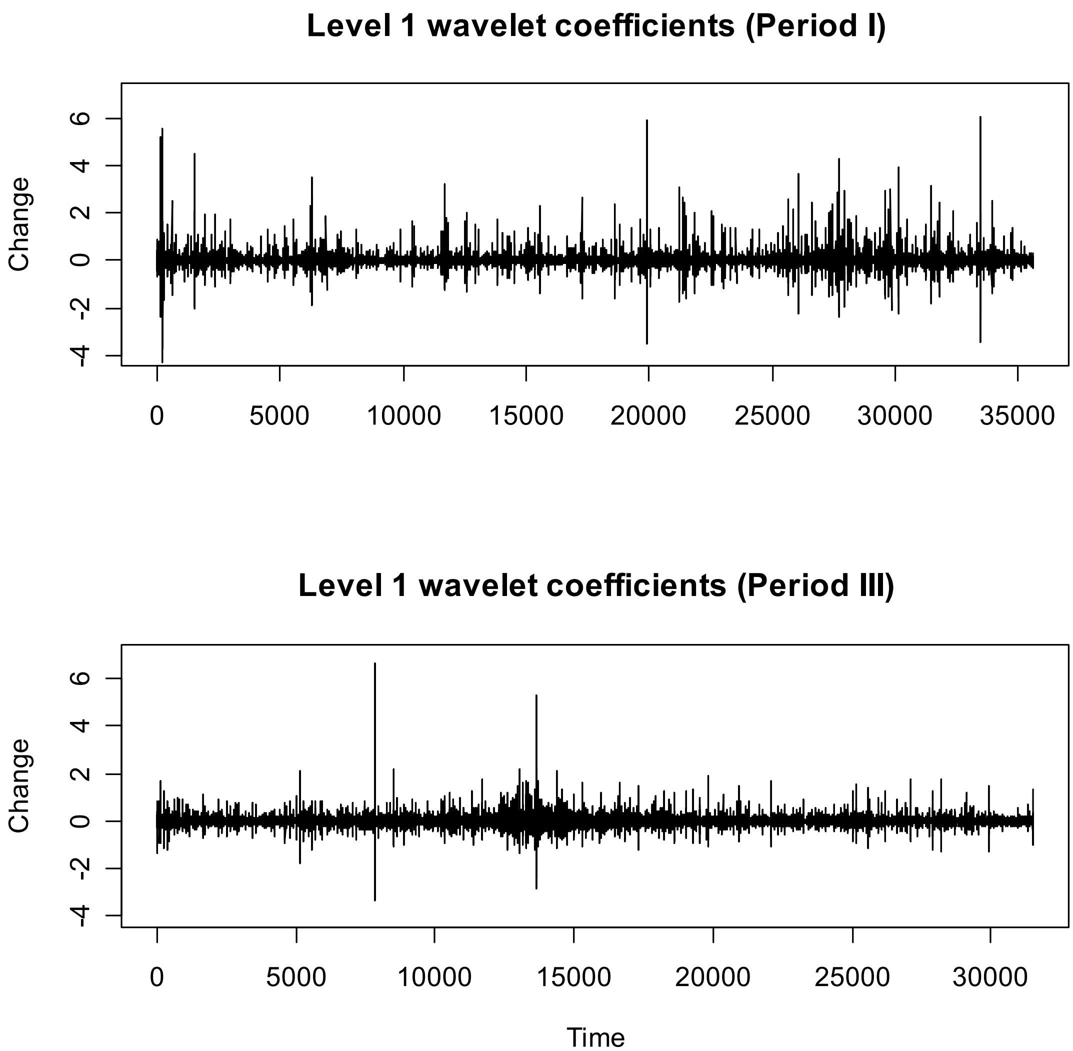

Level 1 wavelet coefficients (Period I)
Change
6
4
2
0
-2
-4
0
5000
10000
15000
20000
25000
30000
35000
Level 1 wavelet coefficients (Period III)
Change
6
4
2
0
-2
-4
0
5000
10000
15000
20000
25000
30000
Time

Figure 18:

of GPH-estimator Deo and Hurvich (1998) have shown that these standard errors are much closer to the standard errors obtained in finite-sample simulations than the asymptotical standard errors obtained from theory. The estimates of d using different levels j in the regression support the conjectured long-memory (see Table ??). Notice that coefficients using levels $j_1 = 2, ..., 6$ and $j_2 = 7, ..., 10$ in Period III do not differ statistically. This would seem to imply that the strong intraday periodicity of volatility has not influenced the estimate of d to a crucial extent. The results of Period I are not as evident, though. To be sure that the intraday periodicity has not confounded the estimate of the d, the periodicity was removed and the estimation re-run (see Sec. 7.6).[25]

Period I

Levels	**Coefficient**	**Std. error**	t**-value**	$\mathbf{P}(>\lvert\mathbf{t}\rvert)$	$\widehat{d}$
$2-10$	-0.62044	0.05016	-12.37	$5.19e-06$	0.18978
$2-6$	-0.44433	0.09108	-4.879	0.01646	0.277835
$7-10$	-0.37730	0.02267	-16.65	0.003590	0.31135

Period III

Levels	**Coefficient**	**Std. error**	t**-value**	$\mathbf{P}(>\lvert\mathbf{t}\rvert)$	$\widehat{d}$
$2-10$	-0.64082	0.02061	-31.10	$9.18e-09$	0.17959
$2-6$	-0.56431	0.04396	-12.84	0.00102	0.217845
$7-10$	-0.57769	0.03998	-14.45	0.00475	0.211155

7.5 Local scaling laws and long-memory

The assumption of a constant long-memory structure may not always be reasonable. Bayraktar *et al.* (2003) tackled the problem of time-varying long-memory by segmenting the data before the estimation of the Hurst coefficient $H(t)$ (a closely related measure of long-memory, see e.g. Beran (1994)). But this scheme might not always be sufficient as Whitcher and Jensen (2000) have pointed out. In particular, they argued that *"the ability to estimate local behavior by applying a partitioning scheme to a global estimating procedure is inadequate when compared with an estimator designed to capture time-varying features"*. In this respect, the work of Gonçalves and Abry (1997), who estimated a local scaling exponent for a continuous-time multifractal Brownian motion, seems more appropriate. Unfortunately, the approach of Gonçalves and Abry involves the contruction of non-standard wavelets which hinders practical implementation. To overcome this difficulty, Whitcher and Jensen (2000)

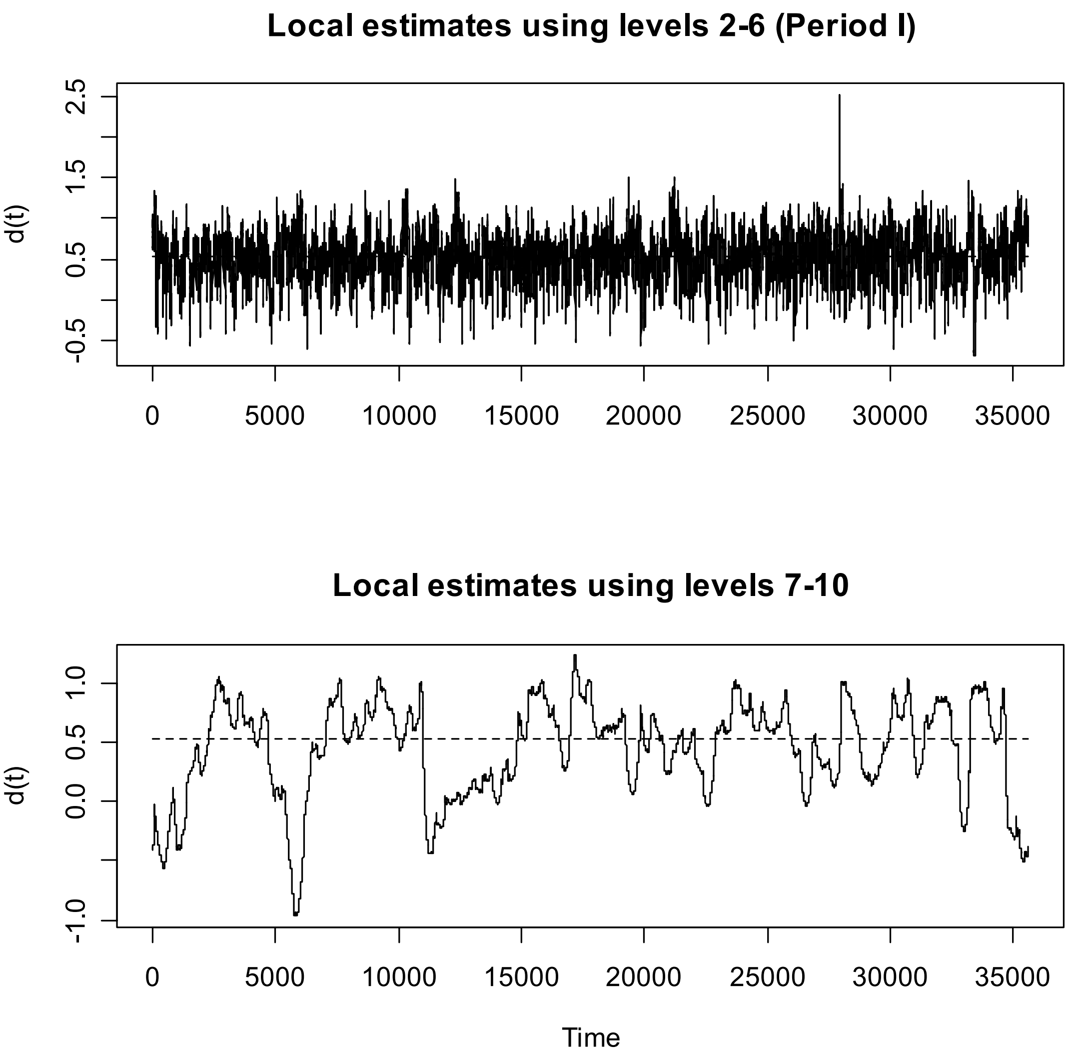
[25]In fact, the wavelet method could have been used to annihilate the intraday dependencies in a similar way that was done in Andersen and Bollerslev (1997b) who used a low-pass filtering technique based on a two-sided weighted average of both past and future absolute returns. Unfortunately however, by only considering the interdaily and longer dynamics (i.e. wavelet smooth of level J), as in Gençay *et al.* (2000), I was not able to reproduce the hyperbolic decay in the sample autocorrelation function of the filtered series. So the FFF was used instead.

introduced a local long-memory parameter estimator based on the MODWT allowing them to stay in the more traditional ARFIMA framework (Gilbert *et al.* (1998) use the DWT). I follow the methodology laid out in Whitcher and Jensen (2000) and Jensen and Whitcher (2000).

In contrast to Jensen and Whitcher (2000) who use log-squared returns to proxy volatility, I continue to use absolute returns (for the reasons stated in Sec. 7.2). To let the linear relationship of $\log \nu_X^2(u, \lambda_j)$ and $\log 2^j$ (see Eq. (8)) hold, I then implicitly assume that absolute results are generated by a locally stationary process. Considering the jumps and clustering of volatility, this assumption seems more reasonable than covariance stationarity. The estimation of time-varying long-memory parameter is therefore done for Periods I and III by OLS. The choice of levels j to be included in the regression was first based on the approximately linear portion of the global scaling law; the global analysis suggested that there exists two linear regions, $j_1 = 1, ..., 6$ and $j_2 = 7, ..., 10$, in Period I. However, using only the smallest levels resulted in estimates of $d(u)$ that varied in a white noise fashion (see the top plots of Figs. 19 and 20). This is in agreement with the argument of Jensen and Whitcher (2000) that intraday levels are irrelevant to long-memory phenomena. However, this contradicts the findings of Andersen and Bollerslev (1997b) that intraday volatility can be informative even in the long-run. Because of this contradiction, and the fact that 5-minute returns are still subject to microstructure effects (bid-ask bounce at the first lag) in Period III, I considered only the first level as uninformative. On the other hand, using only the larger levels resulted in severe unstability in the estimate (see the bottom plots of Figs. 19 and 20). This unstability can be explained by the small number of levels (only 4) used in the OLS-regression. Yet, larger than 10 levels were seriously affected by the boundary.[26]

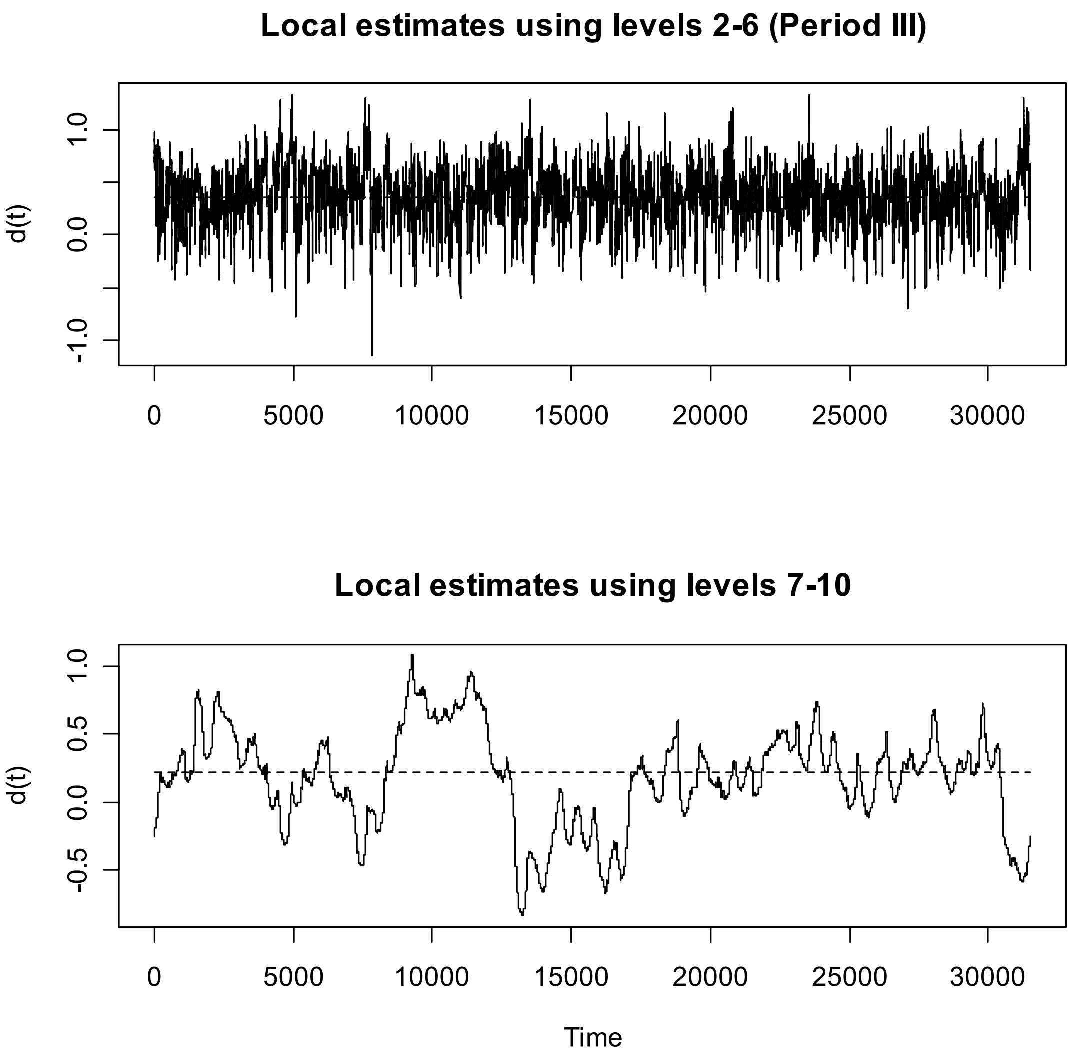
It turned out that using the small and large levels together, i.e. levels $j_3 = 2, ..., 10$,gave the most stable results. This is to expected because of the larger number of observations included. The local long-memory parameter estimates of Periods I and III show similar characteristics (see Figs. 21 and 22). Namely, most of the time the estimate of $d(t)$ stays in the interval $(0, 1/2)$ indicative of long-memory although "outliers" tend to drag the estimate downwards and "out of bounds". Now the approximate zero-phase filter property of the MODWT wavelet filters (after a proper adjustment, see Sec. 5.4) shows it usefulness: the drops in the estimate of $d(t)$ are alignable in time with the original series. For example, it can be inferred from the figure that in Period I the sudden low estimates of $d(t)$ between observations $5,000$ and $10,000$ are due to high returns at that moment in time (see Fig. 21). However, there are also instants when the estimate jumps upwards. In these peculiar cases the regression using levels j_1 and j_2 separately can be informative. For example, in Period I the high jump between observations $25,000$ and $30,000$ is the result of the

[26]The use of levels 11 and 12 is computationally costly too. (These levels can still be included if deemed useful, however! It might be that the use of more levels in the regression could further stabilize the estimate of d.)



Local estimates using levels 2-6 (Period I)
d(t)
-0.5
0.5
1.5
2.5
0
5000
10000
15000
20000
25000
30000
35000
Local estimates using levels 7-10
d(t)
-1.0
0.0
0.5
1.0
0
5000
10000
15000
20000
25000
30000
35000
Time

Figure 19:



Local estimates using levels 2-6 (Period III)
d(t)
-1.0
0.0
1.0
0
5000
10000
15000
20000
25000
30000
Local estimates using levels 7-10
d(t)
-0.5
0.0
0.5
1.0
0
5000
10000
15000
20000
25000
30000
Time

Figure 20:

wavelet variances at the smallest levels being extremely unstable (see the top plot of Fig. 19).[27] In general, wavelet variances at the smallest levels react more powerfully to jumps in volatility. This is the downside in including the smallest levels in the regression. At this point, then, some crititicm of the range of estimates can be raised. It is not at all obvious how one would use estimates that do not belong to a sensible interval of $(0, 1/2)$. This is of course crucial from the point of view modeling using the LMSV model (see Eq. (11a)–(11b)) and will be shortly discussed next.

The estimates of $d(t)$ are of little use if they behave uncontrollably. It is however reliefing to note some structure in the behavior of the long-memory estimate. Namely, the estimate seems to become more stable during less turbulent times. In particular, during the long upward trends the estimate of long-memory rises before reaching stability at about $0.3 - 0.4$. This is in line with the definition of long-memory: the long-memory parameter should be larger when stronger persistance is found. From the view point of modeling the dynamics of $d(t)$, it is interesting to note that the estimates are unconditionally Gaussian (see Fig. 23). This is confirmed by the Jarque–Bera test of normality (see Table ??). At this stage, however, it is too early to say if there is any specific structure in $d(t)$ (such as ARFIMA) that could be exploited in forecasting, for example. This problem is hampered by the possibility of long-memory being spurious; i.e. structural breaks might have affected the estimate of $d(t)$ upwards. Indeed, it has been shown by Granger and Hyung (1999) that there exists a positive relationship between the number of breaks and the value of d. More precisely, in their study of daily S&P 500 absolute returns (from January 4, 1928 to August 30, 1991), d varied in the range of 0.154 and 0.715 calculated over 10 non-overlapping subperiods. They found that d was possibly affected by the magnitude of the break, too, and concluded that the long-memory property is more likely caused by the presence of breaks. These findings would then imply that the timing and size of the breaks is important. Thus, regarding forecastability, it would then be equally important to know if the breaks are endogenous and therefore forecastable. Whether $d(u)$ carries inherent structure or the breaks are endogenous remains an interesting research problem and will be studied in the near future.

The possibility of misspecification has to be acknowledged, too. Namely, it is disturbing that the medians of $d(t)$s in both periods are different (larger) from the ds (cf. Sec. 5.5) although they are expected to match. This anomaly is currently under study.

Period I

Statistical key figures						Jarque–Bera		
Min.	**1st Q.**	**Median**	**Mean**	**3rd Q.**	**Max.**	X^2	**df**	p
−0.3426	0.2084	0.3117	0.3020	0.4056	1.0313	5.5039	2	0.0638

[27] To be precise, this sudden instability was caused by sudden drop in the wavelet variance at the very low levels. This forced me to ignore the 2nd level wavelet variances in OLS for a while. Ignoring was mandatory because taking a logarithm of 0 is not well-defined and certainly not applicable in regression.

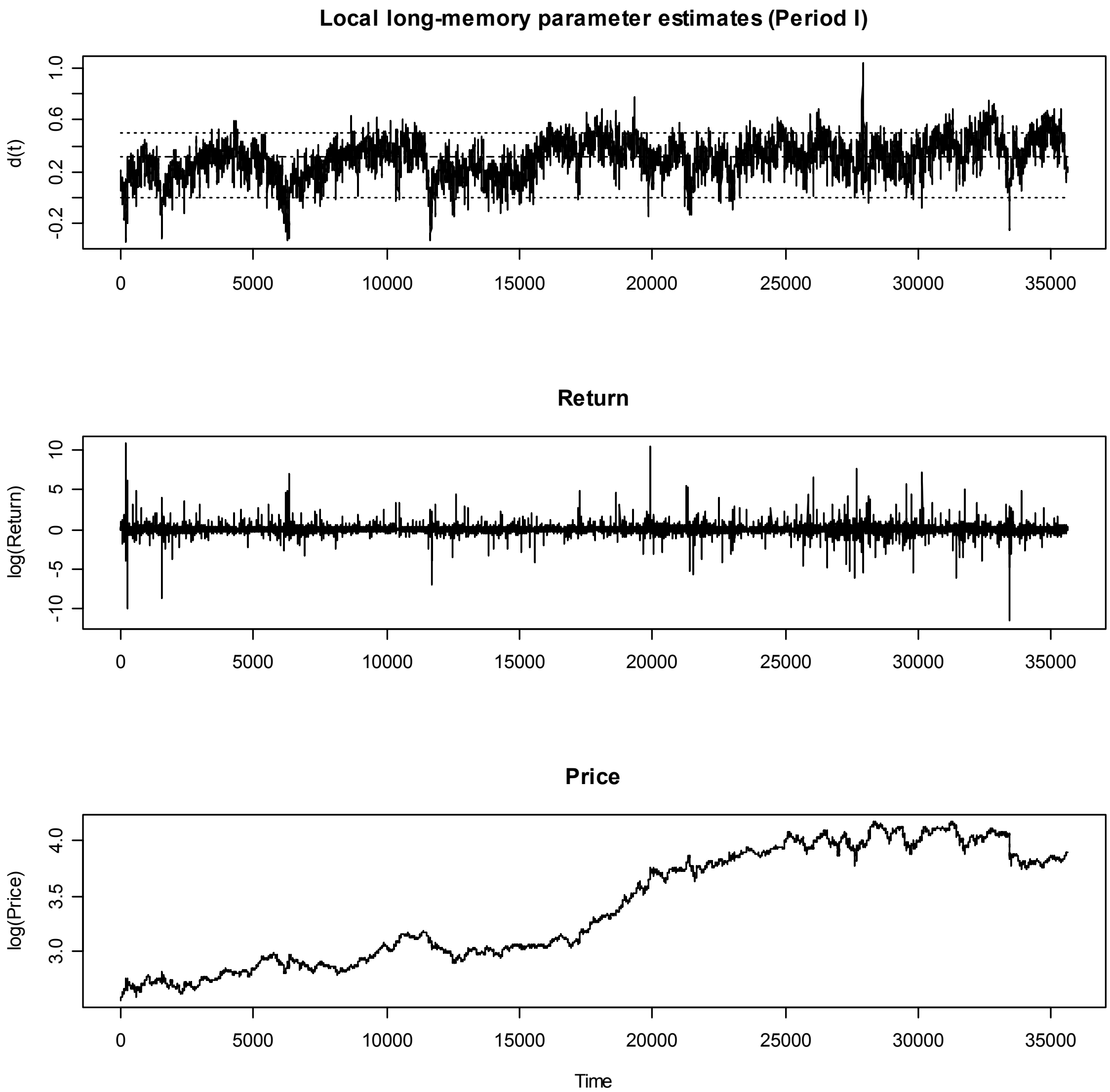


Figure 21: Local long-memory parameter estimates in Period I (the top subplot). The return and price series are plotted below to align the features in time. In particular, notice how the estimate of $d(t)$ drops whenever a big price change takes place.

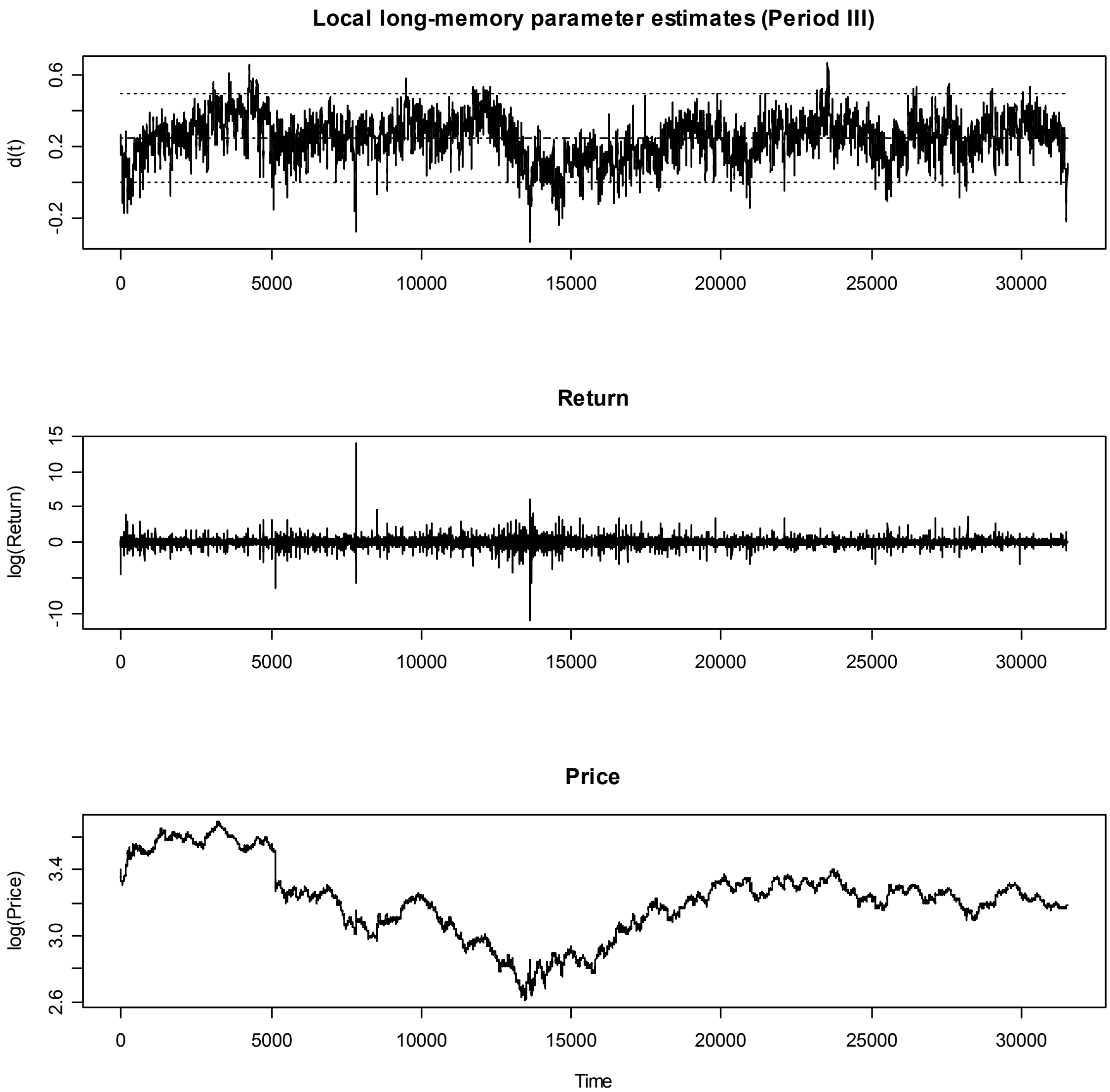


Figure 22: Local long-memory parameter estimates in Period III (the top subplot). The return and price series are plotted below to align the features in time. In particular, notice how the estimate of $d(t)$ drops whenever a big price change takes place.

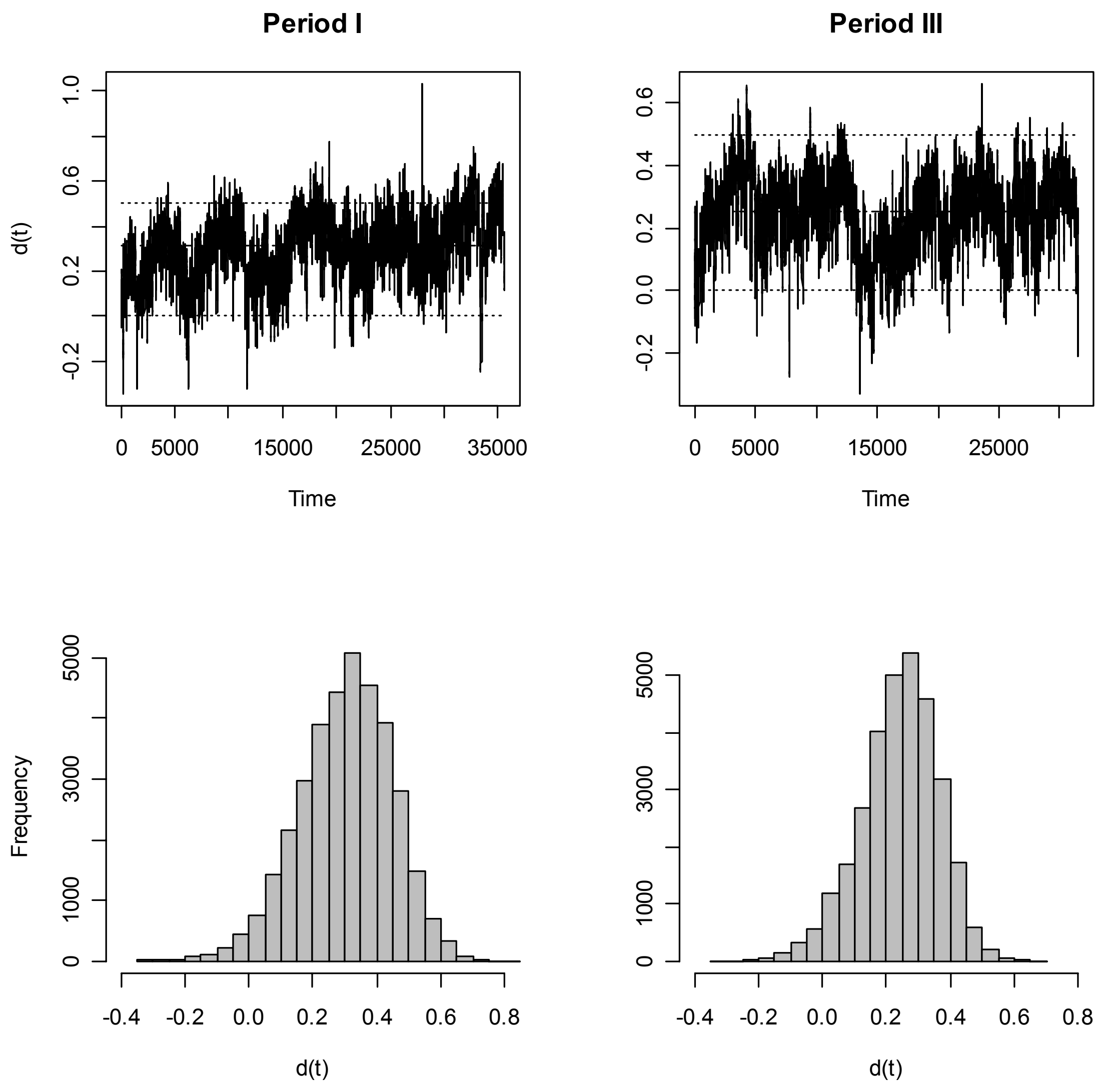


Figure 23: The unconditional distributions of the estimate of $d(t)$ is Gaussian in both periods (see Jarque–Bera test statistic).

Period III

Statistical key figures						Jarque–Bera		
Min.	**1st Q.**	**Median**	**Mean**	**3rd Q.**	**Max.**	X^2	**df**	p
−0.3287	0.1663	0.2505	0.2412	0.3262	0.6624	3.6219	2	0.1635

7.6 Effects of volatility periodicity

The effect of intraday volatility periodicity on the above results (especially the scaling laws) will now be studied. From now on, the overnight 5-minute returns will be excluded from the analysis. In average, the general shape of volatility is similar in Periods I and III (see Figs. 24 and 25). It is obvious that the exclusion of the overnight return interval has not removed all the overnight effects: the first interval exhibits considerably larger volatility than the rest of the intraday intervals. However, after the highly volatile first 5 minutes the average volatility tends to calm down smoothly (although some predetermined days have significant turbulence at midday because of the announcement of quartal reports). At afternoon hours the behavior of volatility becomes abrupt again. The first regular peak occurs at 3.35 p.m. (futures market opening?) and a larger one just an hour later at 4.35 p.m. The latter peak is the New York effect (see Sec. 7.2). There is also a small peak half an hour later at 5.05 p.m. (why?). In Period III, the biggest peak is at 6.05 p.m. (and right after it) when AMT (I) starts (at 6.03 p.m.). It is still a bit unclear if this particular peak is genuine or not; it might well just be an artifact that the 3%-filter was unable to get totally rid of.

The removal of the intraday seasonalities was done using *Fourier Flexible Form* (FFF) (see App. E). The results depend largely on how many sinusoidal terms are included in the parametric form. It seems that more terms does not necessarily give a better result in every respect. For example, by including many terms the overall fit becomes better but then the filtered returns tend to become autocorrelated. On the other hand, while a small number of terms does not touch the autocorrelation structure that much, the inferior fit obtained does not remove all the intraday patterns. Balancing between these two goals suggests to settle for a minimal amount of sinusoids. In Period III, setting $P = 4$ and $J = 0$ and 5 dummies gives a nice fit in average (see Fig. 25):

$$
\begin{aligned}
\widehat{f}(\theta; n) \;=\;& \underset{(0.87)^{*}}{1.85} - \underset{(2.54)^{***}}{11.17}\,\frac{n}{N_1} + \underset{(0.84)^{***}}{3.54}\,\frac{n^2}{N_2} + \underset{(0.27)^{***}}{1.62}\,\mathbf{1}_{n=d_1} + \underset{(0.22)^{***}}{0.74}\,\mathbf{1}_{n=d_2} \\
& + \underset{(0.23)}{0.14}\,\mathbf{1}_{n=d_3} + \underset{(0.23)^{***}}{1.32}\,\mathbf{1}_{n=d_4} + \underset{(0.23)^{***}}{0.91}\,\mathbf{1}_{n=d_5} + \Bigg(- \underset{(0.50)^{***}}{2.19}\,\cos\frac{n2\pi}{N} \\
& - \underset{(0.07)^{***}}{0.55}\,\sin\frac{n2\pi}{N} - \underset{(0.13)^{***}}{0.63}\,\cos\frac{n4\pi}{N} + \underset{(0.04)^{***}}{0.54}\,\sin\frac{n4\pi}{N} + \underset{(0.06)^{*}}{0.13}\,\cos\frac{n6\pi}{N} \\
& + \underset{(0.04)^{***}}{0.18}\,\sin\frac{n6\pi}{N} - \underset{(0.04)}{0.03}\,\cos\frac{n8\pi}{N} - \underset{(0.03)}{0.04}\,\sin\frac{n8\pi}{N} \Bigg),
\end{aligned}
$$

where the numbers in parentheses are standard errors and the asterixes are significance codes (for 0.001, 0.01, and 0.05). Using this estimate resulted in filtered returns that have the overall same autocorrelation characteristics as the original returns with the exception that the first-lag halfed in size, being still statistically significant (see Fig. 26). Although the filtered absolute returns continue to show some signs of periodicity, this is considered adequate in the present context. Interestingly, allowing daily variability in the regression (i.e. $J \geq 1$) did not make the fit signifantly better but instead wrongly emphasized few "outliers". Daily volatility was estimated by realized variance.

The analysis of this subsection is currently being done.

8 Conclusions

To be done.

A Some functional analysis

Supports sections 3 and 4 and is therefore skipped in this version.

B Orthonormal transforms

This section borrows from Percival and Walden (2000, Sec. 3.1).

Let $\mathcal{O}$ be an orthonormal real-valued $N \times N$ matrix, i.e. $\mathcal{O}^T\mathcal{O} = \mathcal{I}_N$. Let $\mathcal{O}_{j\bullet}$ and $\mathcal{O}_{\bullet k}$ refer to the jth row vector and kth column vector, respectively. Then

$$\overset{[N\times N]}{\mathcal{O}} = \begin{bmatrix} \mathcal{O}_{0\bullet}^T \\ \mathcal{O}_{1\bullet}^T \\ \vdots \\ \mathcal{O}_{N-1\bullet}^T \end{bmatrix} = [\mathcal{O}_{\bullet 0}, \mathcal{O}_{\bullet 1}, ..., \mathcal{O}_{\bullet N-1}].$$

One can use this matrix to *analyze* an arbitrary real-valued time series, given by the column vector $\mathbf{x}$, in the following way:

$$\overset{[N\times 1]}{\mathbf{O}} = \mathcal{O}\mathbf{x} = \begin{bmatrix} \mathcal{O}_{0\bullet}^T \\ \mathcal{O}_{1\bullet}^T \\ \vdots \\ \mathcal{O}_{N-1\bullet}^T \end{bmatrix} \mathbf{x} = \begin{bmatrix} \mathcal{O}_{0\bullet}^T\mathbf{x} \\ \mathcal{O}_{1\bullet}^T\mathbf{x} \\ \vdots \\ \mathcal{O}_{N-1\bullet}^T\mathbf{x} \end{bmatrix} = \begin{bmatrix} \langle \mathbf{x}, \mathcal{O}_{0\bullet} \rangle \\ \langle \mathbf{x}, \mathcal{O}_{1\bullet} \rangle \\ \vdots \\ \langle \mathbf{x}, \mathcal{O}_{N-1\bullet} \rangle \end{bmatrix},$$

since $\mathcal{O}_{j\bullet}^T\mathbf{x} = \left\langle \mathcal{O}_{j\bullet}, \mathbf{x} \right\rangle = \left\langle \mathbf{x}, \mathcal{O}_{j\bullet} \right\rangle$. The column vector $\mathbf{O}$ consists of the transform coefficients for $\mathbf{x}$ with respect to the orthonormal transform $\mathcal{O}$. Specifically, the jth transform coefficient O_j is given by the inner product $\left\langle \mathbf{x}, \mathcal{O}_{j\bullet} \right\rangle$.

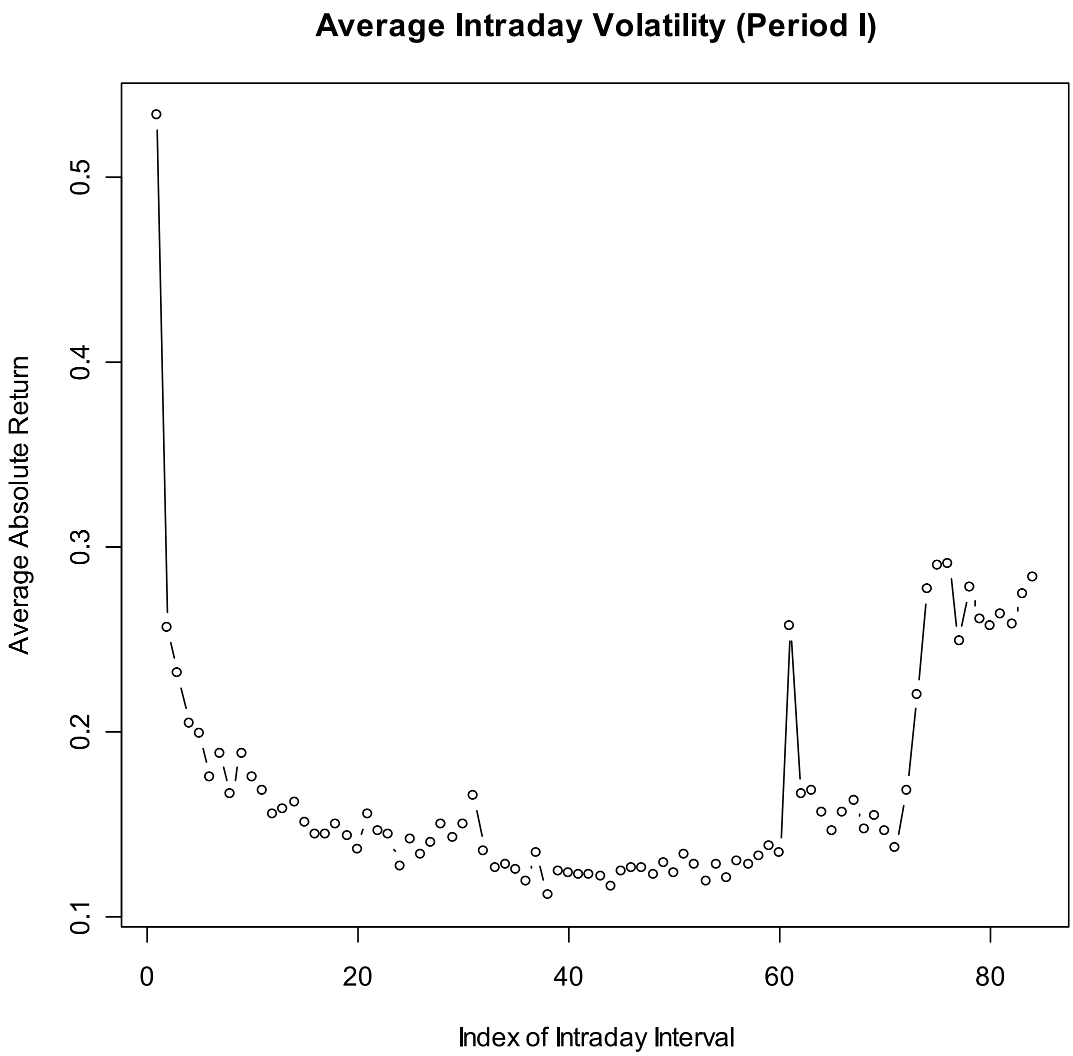

Average Intraday Volatility (Period I)
Average Absolute Return
0.1
0.2
0.3
0.4
0.5
0
20
40
60
80
Index of Intraday Interval

Figure 24:

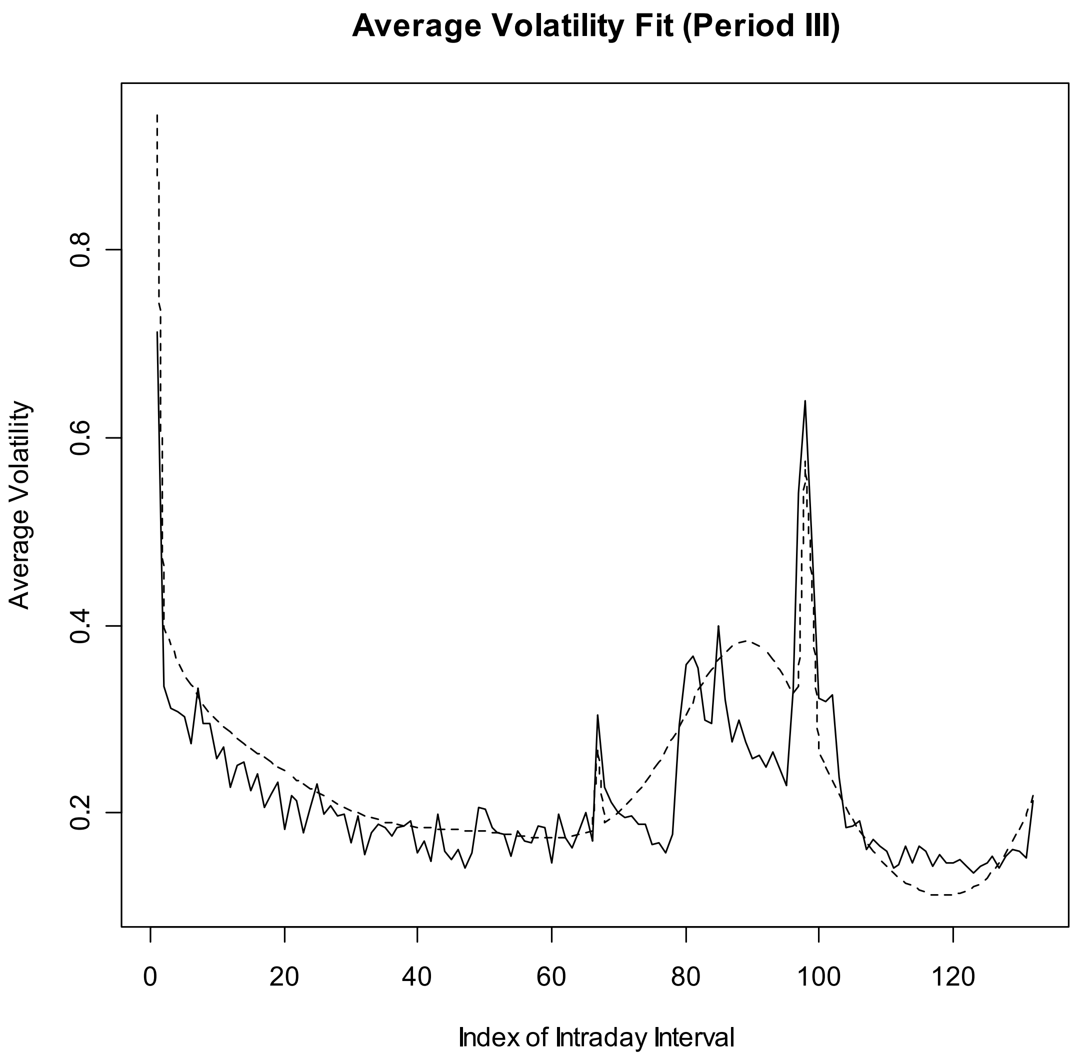

Average Volatility Fit (Period III)
Average Volatility
0.2
0.4
0.6
0.8
0
20
40
60
80
100
120
Index of Intraday Interval

Figure 25:

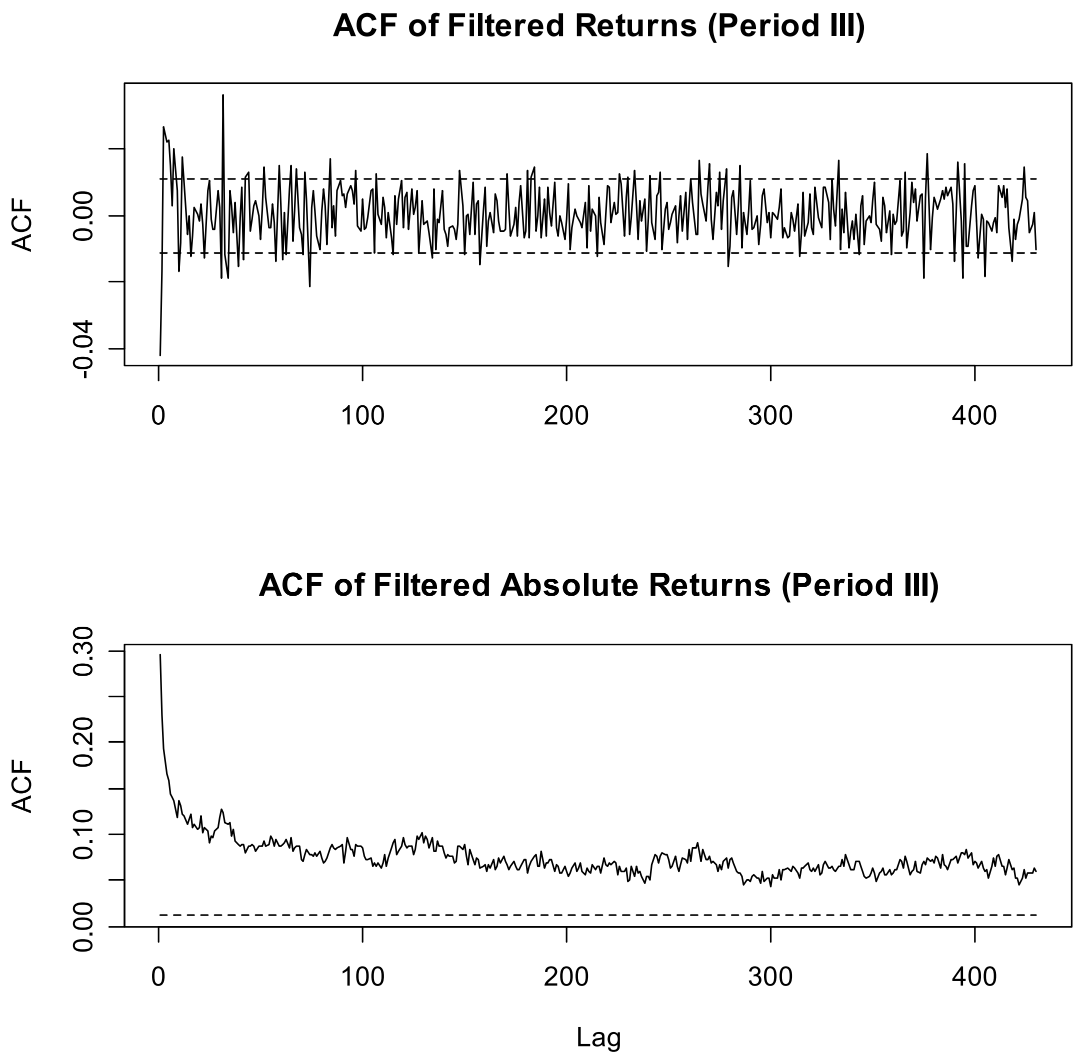

ACF of Filtered Returns (Period III)
ACF
0.00
-0.04
0
100
200
300
400
ACF of Filtered Absolute Returns (Period III)
ACF
0.00
0.10
0.20
0.30
0
100
200
300
400
Lag

Figure 26:

On the other hand, premultiplying both sides of the above equation by $\mathcal{O}^T$, and using orthonormality, one can *synthesize* the time series $\mathbf{x}$ as

$$\underset{[N\times 1]}{\mathbf{x}} = \mathcal{O}^T \mathbf{O} = [\mathcal{O}_{0\bullet}, \mathcal{O}_{1\bullet}, ..., \mathcal{O}_{N-1\bullet}] \begin{bmatrix} O_0 \\ O_1 \\ \vdots \\ O_{N-1} \end{bmatrix} = \sum_{j=0}^{N-1} O_j \mathcal{O}_{j\bullet}.$$

Furthermore, since $O_j = \langle \mathbf{x}, \mathcal{O}_{j\bullet} \rangle$, it is possible to re-express the time series $\mathbf{x}$ as a unique linear combination of $\mathcal{O}_{0\bullet}, \mathcal{O}_{1\bullet}, ..., \mathcal{O}_{N-1\bullet}$:

$$\mathbf{x} = \sum_{j=0}^{N-1} \langle \mathbf{x}, \mathcal{O}_{j\bullet} \rangle \mathcal{O}_{j\bullet}.$$

That is, one has recovered the ability to write a vector as the sum of its projections on the basis vectors.

C Fractional differencing and long-memory

The *fractional differencing operator* $(1-B)^d$ is formally defined by its infinite Maclaurin series expansion,

$$(1-B)^d \doteq \sum_{k=0}^{\infty} \frac{\Gamma(k-d)}{\Gamma(k+1)\Gamma(-d)} B^k,$$

where B and $\Gamma(\cdot)$ denote the lag-operator and the gamma function, respectively (e.g. Breidt *et al.* (1998, p. 328)).

A real-valued discrete parameter *fractional* ARIMA (ARFIMA) process $\{X_t\}$ is often defined with a binomial series expansion (Gençay *et al.* (2002a, p. 163)),

$$(1-B)^d X_t \doteq \sum_{k=0}^{\infty} \binom{d}{k} (-1)^k X_{t-k}.$$

where

$$\binom{a}{b} \doteq \frac{a!}{b!(a-b)!} = \frac{\Gamma(a+1)}{\Gamma(b+1)\Gamma(a-b+1)}.$$

These models were originally introduced by Granger and Joyeux (1980) and Hosking (1981).

In ARFIMA models, the "long-memory" dependency is characterized solely by the *fractional differencing parameter d*. A time series is said to exhibit *long-memory* when it has a covariance function $\gamma(j)$ and a spectrum $f(\lambda)$ such that they are of the same order as j^{2d-1} *and* λ^{-2d}, as $j \to \infty$ and $\lambda \to 0$, respectively.[28] For $0 < d < 1/2$, an ARFIMA model exhibits long-memory, and for $-1/2 < d < 0$ it exhibits *antipersistance*. In practice, the range $|d| < 1/2$ is of particular interest because then an ARFIMA model is stationary and invertible (Hosking (1981)).

More detailed definitions of long-memory can be found from Beran (1994), for example. Concerning fractionally integrated processes in econometrics in particular, see Baillie (1996).

D Locally stationary process

Dahlhaus (1996, 1997) defines a locally stationary process $X_{t,T}$ $(t = 0, 1, ..., T-1)$ as the triangular array, with transfer function A^0, drift μ, and spectral representation

$$X_{t,T} = \mu\left(t/T\right) + \int_{-\pi}^{\pi} e^{i\omega t} A^0_{t,T}(\omega) dZ(\omega),$$

where the components satisfy certain technical conditions (see Dahlhaus (1997, Definition 2.1) or Jensen and Whitcher (2000)). For example, autoregressive processes with time-varying coefficients are locally stationary (Dahlhaus (1996, Theorem 2.3)).

Jensen and Whitcher (2000) give another example of a locally stationary process. It is constructed by considering a stationary, invertible moving average process Y_t with spectral representation

$$Y_t = \int_{-\pi}^{\pi} e^{i\omega t} A(\omega) dZ(\omega),$$

where the transfer function is $A(\omega) = \left(1 + \theta e^{-i\omega t}\right)/2\pi$ and $|\theta| < 1$. If the process $X_{t,T}$ is now defined as

$$X_{t,T} = \mu\left(t/T\right) + \sigma\left(t/T\right) Y_t,$$

where $\mu, \sigma : [0, 1] \to \mathbf{R}$ are continuous functions, then $X_{t,T}$ is a locally stationary process with the time varying transfer function

$$A(u, \omega) = A^0_{t,T}(\omega) = \frac{\sigma(u)}{2\pi}\left(1 + \theta e^{-i\omega t}\right).$$

Thus the time-path of $X_{t,T}$ exhibits the periodic behavior of a stationary moving average process but with time-varying amplitude equal to $\sigma(u)$.

[28] The rate of decay in covariance does *not necessarily* imply the rate of decay in spectrum, as noted in Bollerslev and Wright (2000, p. 87). Formal conditions for the equivalance are discussed in Beran (1994), for example (see also Granger and Ding (1996)).

E Fourier Flexible Form

Following Andersen and Bollerslev (1997c), intraday returns can be decomposed as

$$r_{t,n} = \mathbb{E}\left(r_{t,n}\right) + \frac{\sigma_t s_{t,n} Z_{t,n}}{\sqrt{N}},$$

where in their notation N refers to the number of return intervals n per *day* (i.e. not to the total length of the series!), σ_t is the *daily* volatility factor, and $Z_{t,n}$ is an IID random variable with mean 0 and variance 1. Notice that $s_{t,n}$, the periodic component for the nth intraday interval, depends on the characteristics of trading day t. By then squaring both sides and taking logarithms, define $x_{t,n}$ to be

$$x_{t,n} \doteq 2\log\left[\left|r_{t,n} - \mathbb{E}\left(r_{t,n}\right)\right|\right] - \log\sigma_t^2 + \log N = \log s_{t,n}^2 + \log Z_{t,n}^2,$$

so that $x_{t,n}$ consists of a deterministic and a stochastic component.

The modeling of $x_{t,n}$ is done via non-linear regression in n and σ_t,

$$x_{t,n} = f\left(\theta; \sigma_t, n\right) + u_{t,n},$$

where $u_{t,n} \doteq \log Z_{t,n}^2 - \mathbb{E}\left(\log Z_{t,n}^2\right)$ is an IID random variable with mean 0. In practice, the estimation of f is implemented by the following parametric expression:

$$\begin{aligned} f\left(\theta; \sigma_t, n\right) &= \sum_{j=0}^{J} \sigma_t^j \left[\mu_{0j} + \mu_{1j}\frac{n}{N_1} + \mu_{2j}\frac{n^2}{N_2} + \sum_{i=1}^{D} \lambda_{ij} \mathbf{1}_{n=d_i} \right. \\ &\quad \left. + \sum_{p=1}^{P}\left(\gamma_{pj}\cos\frac{pn2\pi}{N} + \delta_{pj}\sin\frac{pn2\pi}{N}\right)\right], \end{aligned}$$

where $N_1 \doteq (N+1)/2$ and $N_2 \doteq (N+1)(N+2)/6$ are normalizing constants. If one sets $J = 0$ and $D = 0$, then this reduces to the standard FFF proposed by Gallant (1981, 1982). The trigonometric functions are ideally suited for smooth varying patterns. Andersen and Bollerslev (1997c) have argued that in equity markets allowing for $J \geq 1$ might be important, however. Namely, by including cross-terms in the regression allows $s_{t,n}$ to depend on the overall level of volatility on trading day t which is often the case in stock market data. The actual estimation of f is most easily accomplished using a two-step procedure described in Andersen and Bollerslev (1997c, App. B).

After a proper normalization, the estimator of the intraday periodic component for interval n on day t is found to be

$$\widehat{s}_{t,n} = \frac{T\exp\left(\widehat{f}_{t,n}/2\right)}{\sum_{t=1}^{[T/N]}\sum_{n=1}^{N}\exp\left(\widehat{f}_{t,n}/2\right)},$$

where T is the *total* length of the sample and $[T/N]$ denotes the number of trading days. The filtered returns (returns free from the volatility periodicity) are then obtained via

$$\widetilde{r}_{t,n} \doteq r_{t,n}/\widehat{s}_{t,n}.$$

References

[1] Aït-Sahalia, Yacine (2003): Disentangling Volatility from Jumps. NBER Working Paper No. w9915.

[2] Abry, P. – Gonçalves, P. – Flandrin, P. (1993): Wavelet-Based Spectral Analysis of $1/f$ processes.

[3] Abry, P. – Veitch, D. (1998): Wavelet Analysis of Long-Range-Dependent Traffic. *IEEE Transactions on Information Theory* **44**, 2–15.

[4] Andersen, Torben G. – Bollerslev, Tim – Diebold, Francis X. – Labys, Paul (2000): The Distribution of Realized Exchange Rate Volatility.

[5] Andersen, Torben G. – Bollerslev, Tim (1997): Answering the Critics: Yes, ARCH Models Do Provide Good Volatility Forecasts. NBER Working Paper 6023, Cambridge.

[6] Andersen, Torben G. – Bollerslev, Tim (1997b): Heterogenous Information Arrivals and Return Volatility Dynamics: Uncovering the Long-run in High Frequency Returns. *Journal of Finance* **52**, 975–1005.

[7] Andersen, Torben G. – Bollerslev, Tim (1997c): Intraday Periodicity and Volatility Persistance in Financial Markets. *Journal of Empirical Finance* **4**, 115–158.

[8] Andersen, Torben G. – Bollerslev, Tim – Diebold, Francis X. (2003): Some Like it Smooth, and Some Like it Rough: Untangling Continuous and Jump Components in Measuring, Modeling, and Forecasting Asset Return Volatility. Manuscript, Northwestern University, Duke University, University of Pennsylvania.

[9] Atkins, Frank J. – Sun, Zhen (2003): Using Wavelets to Uncover the Fisher Effect. Draft, Department of Economics, Discussion Paper 2003-09.

[10] Bachman, George – Narici, Lawrence – Beckenstein, Edward (2000): Fourier and Wavelet Analysis. Springer-Verlag, New York, USA.

[11] Bai, Xuezheng – Russell, Jeffrey R. – Tiao, George C. (2001): Beyond Merton's Utopia (I): Effects of Non-Normality and Dependence on the Precision of Variance Estimates Using High-Frequency Financial Data. Working Paper, University of Chicago.

[12] Baillie, Richard T. – Myers, Robert F. (1991): Bivariate Garch Estimation of the Optimal Commodity Futures Hedge. *Journal of Applied Econometrics* **6**, 109–124.

[13] Baillie, Richard T. – Bollerslev, Tim (1991): Intra-Day and Inter-Market Volatility in Foreign Exchange Rates. *Review of Economic Studies* **58**, 565–585.

[14] Barndorff-Nielsen, Ole E. – Shephard, Neil (2003a): Power and Bipower Variation with Stochastic Volatility and Jumps. Manuscript, Oxford University.

[15] Barndorff-Nielsen, Ole E. – Shephard Neil (2003b): Econometrics of Testing for Jumps in Financial Economics Using Bipower Variation. Manuscript, Oxford University.

[16] Beran, Jan (1994): Statistics for Long Memory Processes, Volume 61 of Monographs on Statistics and Applied Probability. Chapman and Hall, New York.

[17] Bhattacharya, R.N. – Gupra, V.K. – Waymire, E. (1983): The Hurst Effect under Trends. *Journal of Applied Probability* **20**, 649–662.

[18] Black, Fischer (1976): Studies of Stock Price Volatility Changes. *Proceedings of the 1976 Meetings of the American Statistical Association, Business and Economics Statistics Section*, 177–181.

[19] Bollerslev, Tim (1986): Generalized Autoregressive Conditional Heteroskedasticity. *Journal of Econometrics* **31**, 307–327.

[20] Bollerslev, Tim – Chou, Ray Y. – Kroner, Kenneth F. (1992): ARCH Modeling in Finance; A Review of the Theory and Empirical Evidence. *Journal of Econometrics* **52**, 5–59.

[21] Bollerslev, Tim – Engle, Robert F. – Nelson, Daniel B. (1994): ARCH Models. In Engle and McFadden (eds.): Handbook of Econometrics, Vol IV, 2959–3038. Elsevier Science.

[22] Bollerslev, Tim – Mikkelsen, Hans-Ole (1996): Modeling and Pricing Long-Memory in Stock Market Volatility. *Journal of Econometrics* **73**, 151–184.

[23] Bollerslev, Tim – Wright, Johathan H. (2000): Semiparametric Estimation of Long-Memory Volatility Dependencies: The Role of High-Frequency Data. *Journal of Econometrics* **98**, 81–106..

[24] Bollerslev, Tim (2001): Financial Econometrics: Past Developments and Future Challenges. *Journal of Econometrics* **100**, 41–51.

[25] Bougerol, Philippe – Picard, Nico (1992): Stationarity of GARCH Processes and of Some Nonnegative Time Series. *Journal of Econometrics* **52**, 115–127.

[26] Breidt, F. Jay – Crato, Nuno – de Lima, Pedro (1998): The Detection and Estimation of Long Memory in Stochastic Volatility. *Journal of Econometrics* **83**, 325–348.

[27] Broto, Carmen – Ruiz, Esther (2002): Estimation Methods for Stochastic Volatility Models: A Survey. Working Paper 02-54 (14), Universidad Carlos III de Madrid.

[28] Bruce, A. G. – Gao, H.-Y. (1996): Applied Wavelet Analysis with S-PLUS. Springer, New York.

[29] Burda, Michael – Wyplosz, Charles (2001): Macroeconomics - A European Text. Third Edition, Oxford University Press.

[30] Campbell, John Y. – Lo, Andrew W. – MacKinlay, A. Craig (1997): The Econometrics of Financial Markets. Second corrected printing. Princeton University Press.

[31] Carnero, M. Angeles – Peña, Daniel – Ruiz, Esther (2001): Is Stochastic Volatility More Flexible than GARCH? Working Paper 01-08, Universidad Carlos III de Madrid.

[32] Chambers, Marcus J. (1998): Long Memory and Aggregation in Macroeconomic Time Series. *International Economic Review* **39**, Symposium on Forecasting and Empical Methods in Macroeconomics and Finance, 1053–1072.

[33] Chew, Cyrene (2001): The Money and Income Relationship of European Countries by Time Scale Decomposition Using Wavelets.

[34] Cont, Rama (2001): Empirical Properties of Asset Returns: Stylized Facts and Statistical Issues. *Quantitative Finance* **1**, 223–236.

[35] Corsi, Fulvio – Zumbach, Gilles – Müller, Ulrich – Dacorogna, Michel (2001): Consistent High-Precision Volatility from High-Frequency Data. Olsen & Associates Working Paper.

[36] Craigmile, Peter F. – Percival, Donald B. (2002): Wavelet-Based Trend Detection and Estimation.

[37] Craigmile, Peter F. – Percival, Donald B. (200?): Asymptotic Decorrelation of Between-Scale Wavelet Coefficients.

[38] Craigmile, Peter F. – Percival, Donald B. – Guttorp, Peter (2000): The Impact of Wavelet Coefficient Correlations on Fractionally Differenced Process Estimation. Technical Report No. 049, NRCSE.

[39] Curci, Giuseppe – Corsi, Fulvio (2003): A Discrete Sine Transform Approach for Realized Volatility Measurement. Working Paper No. 44, National Centre of Competence in Research, Financial Valuation and Risk Management.

[40] Dacorogna, Michel M. – Müller, Ulrich A. – Nagler, Robert J. – Olsen, Richard B. – Pictet, Olivier V. (1993): A Geographical Model for the Daily and Weekly Seasonal Volatility in the Foreign Exchange Market. *Journal of International Money and Finance* **12**, 413–438.

[41] Dacorogna, Michel M. – Gençay, Ramazan – Müller, Ulrich A. – Olsen, Richard B. – Pictet, Olivier V. (2001): An Introduction to High-frequency Finance. Academic Press.

[42] Dahlhaus, R. (1996): On the Kullback–Leibler Information Divergence of Locally Stationary Processes. *Stochastic Processes and their Applications* **62**, 139–168.

[43] Dahlhaus, R. (1997): Fitting Time Series Models to Nonstationary Processes. *The Annals of Statistics* **25**, 1–37.

[44] Daubechies, I. (1988): Orthonormal Bases of Compactly Supported Wavelets. *Communications of Pure and Applied Mathematics* **41**, 909–996.

[45] Daubechies, I. (1990): The Wavelet Transform, Time-Frequency Localization and Signal Analysis. *IEEE Transactions on Information Theory* **36**, 961–1005.

[46] Davidian, Marie – Carroll, Raymond J. (1987): Variance Function Estimation. *Journal of American Statistical Association* **82**, 1079–1091.

[47] Deo, Rohit S. – Hurvich, Clifford M. (1999): On the Log Periodogram Regression Estimator of the Memory Parameter in Long Memory Stochastic Volatility Models. New York University.

[48] Diebold, Francis X. (1986): Modeling the Persistance of the Conditional Variances: A Comment. *Econometric Reviews* **5**, 51–56..

[49] Diebold, Francis X. (2004): The Nobel Memorial Prize for Robert F. Engle. Working paper, Wharton Financial Institutions Center.

[50] Diebold, Francis X. – Inoue, Atsushi (2001): Long Memory and Regime Switching. *Journal of Econometrics* **105**, 131–159.

[51] Diebold, Francis X. – Hickman, Andrew – Inoue, Atsushi – Schuermann, Til (1997): Converting 1-Day Volatility to h-Day Volatility: Scaling by $\sqrt{h}$ is Worse than You Think. Working paper, Wharton Financial Institutions Center.

[52] Ding, Zhuanxin – Granger, Clive W.J. – Engle, Robert F. (1993): A Long Memory Property of Stock Market Returns and a New Model. *Journal of Empirical Finance* **1**, 83–106.

[53] Ding, Zhuanxin – Granger, Clive W.J. (1996): Modeling Volatility Persistance of Speculative Returns: A New Approach. *Journal of Econometrics* **73**, 185–215.

[54] Donoho, David L. (1992): De-noising via Soft-thresholding. Technical Report. Department of Statistics, Stanford University, Stanford.

[55] Donoho, David L. – Johnstone, Iain M. (1992): Ideal Spatial Adaptation via Wavelet Shrinkage. Technical Report. Department of Statistics, Stanford University, Stanford.

[56] Donoho, David L. – Johnstone, Iain M. – Kerkyacharian, Gerard – Picard, Dominique (1993): Density Estimation by Wavelet Thresholding. Technical Report 426. Department of Statistics, Stanford University, Stanford.

[57] Donoho, David L. – Johnstone, Iain M. – Kerkyacharian, Gerard – Picard, Dominique (1995): Wavelet Shrinkage: Asymptopia? *Journal of the Royal Statistical Society, Series B (Methodological)* **57**, 301–369.

[58] Engle, Robert F. (2000): The Econometrics of Ultra High Frequency Data. *Econometrica* **68**, 1–22.

[59] Engle, Robert F. (1982): Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of United Kingdom Inflation. *Econometrica* **50**, 987–1008.

[60] Engle, Robert F. – Russell, J.R. (1998): Autoregressive Conditional Duration: A New Model for Irregularly Spaced Transaction Data. *Econometrica* **66**, 1127–1162.

[61] Engle, Robert F. – Mustafa, Chowdhury (1992): Implied ARCH Models from Options Prices. *Journal of Econometrics* **52**, 289–311.

[62] Engle, Robert F. – Ng, Victor K. (1993): Measuring and Testing the Impact of News on Volatility. *Journal of Finance* **48**, 1749–1778.

[63] Engle, Robert F. – Ito Takatoshi – Lin, Wen-Ling (1990): Meteor Showers or Heat Waves? Heteroskedastic Intra Daily Volatility in the Foreign Exchange Market. *Econometrica* **58**, 525–542.

[64] Fama, Eugene F. (1965): The Behavior of Stock Market Prices. *Journal of Business* **38**, 34–105.

[65] Feinstone, Lauren J. (1987): Minute by Minute, Normality, and Randomness in Intra-Daily Asset Prices.

[66] Franses, Philip Hans – van Dijk, Dick (2000): Non-Linear Time Series Models in Empirical Finance. Cambridge University Press, Cambridge.

[67] Fuller, Wayne A. (1996): Introduction to Statistical Time Series. 2nd Edition, John Wiley & Sons.

[68] Fung, William K.H. – Hsieh, David A. (1991): Empirical Analysis of Implied Volatility: Stocks, Bonds and Currencies. Working Paper, Fuqua School of Business. Internet: Http://faculty.fuqua.duke.edu/~dah7/.

[69] Gallant, A. Ronald (1981): On the Bias in Flexible Functional Forms and an Essentially Unbiased Form. *Journal of Econometrics* **15**, 211–245.

[70] Gallant, A. Ronald (1982): Unbiased Determination of Production Technologies. *Journal of Econometrics* **20**, 285–323.

[71] Gençay, Ramazan – Selçuk, Faruk – Whitcher, Brandon (2001): Scaling Properties of Foreign Exchange Volatility. *Physica A* **289**, 249–266.

[72] Gençay, Ramazan – Selçuk, Faruk – Whitcher, Brandon (2002a): An Introduction to Wavelets and Other Filtering Methods in Finance and Economics. Academic Press.

[73] Gençay, Ramazan – Selçuk, Faruk – Whitcher, Brandon (2002b): Asymmetry of Information Flow Between Volatilities Across Time Scales.

[74] Gençay, Ramazan – Selçuk, Faruk – Whitcher, Brandon (2002c): Robustness of Systematic Risk Across Time Scales.

[75] Genon-Catalot, V. – Larédo, C. – Picard, D. (1992): Non Parametric Estimation of the Diffusion Coefficient by Wavelet Methods. *Scandinavian Journal of Statistics* **19**, 319–335.

[76] Geweke, J. – Porter-Hudak, S. (1983): The Estimation and Application of Long Memory Time Series Models. *Journal of Time Series Analysis* **4**, 221–238.

[77] Ghysels, Eric – Harvey, Andrew – Renault, Eric (1995): Stochastic Volatility. CIRANO Scientific Series.

[78] Ghysels, Eric – Santa-Clara, Pedro – Valkanov, Rossen (2003): Predicting Volatility: Getting the Most out of Return Data Sampled at Different Frequencies. Presented at the Conference "New Frontiers of Financial Volatility Modeling", Florence, Italy, May 25–27.

[79] Goodhart, Charles A.E. – O'Hara, Maureen (1997): High Frequency Data in Financial Markets: Issues and Applications. *Journal of Empirical Finance* **4**, 73–114.

[80] Gourieroux, Christian – Jasiak, Joann (2001): Financial Econometrics. Princeton University Press, Princeton.

[81] Granger, Clive W.J. – Ding, Zhuanxin (1996): Varieties of Long Memory Models. *Journal of Econometrics* **73**, 61–77.

[82] Granger, Clive W.J. – Hyung, Namwon (1999): Occasional Structural Breaks and Long Memory. Discussion Paper 99-14, University of California San Diego.

[83] Guillaume, D. M. – Dacorogna, M. M. – Dave, R. D. – Müller, U. A. – Olsen, R. B. – Pictet, O. V. (1997): From the Bird's Eye to the Microscope: a Survey of New Stylized Facts of the Intra-daily Foreign Exchange Markets. *Finance and Stochastics* **1**, 95–129.

[84] Hall, P. – Kerkyacharian, G. – Picard, D. (1998): Block Threshold Rules for Curve Estimation Under Kernel and Wavelet Methods. *Annals of Statistics* **26**, 922–942.

[85] Hamilton, James D. (1989): A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle. *Econometrica* **57**, 357–384.

[86] Hamilton, James D. (1994): Time Series Analysis. Princeton University Press.

[87] Harris, Lawrence (1986): A Transaction Data Study of Weekly and Intradaily Patterns in Stock Returns. *Journal of Financial Economics* **16**, 99–117.

[88] Harvey, Andrew (1998): Long Memory in Stochastic Volatility. In Stephen Satchell and John Knight (eds.), *Forecasting Volatility in Financial Markets*, 307–320. Butterworth-Heinemann, Oxford.

[89] Hentschel, Ludger (1995): All in the Family Nesting Symmetric and Asymmetric GARCH Models. *Journal of Financial Economics* **39**, 71–104.

[90] Huang, Xin – Tauchen, George (2003): The Relative Contribution of Jumps to Total Price Variation. Working Paper, Duke University.

[91] Hogan, Kedreth C. – Melvin, Michael T. (1994): Sources of Meteor Showers and Heat Waves in the Foreign Exchange Market. *Journal of International Economics* **37**, 239–247.

[92] Hol, Eugenie – Koopman, Siem Jan (2002): Stock Index Volatility Forecasting with High Frequency Data. University of Birmingham, Free University Amsterdam.

[93] Holscheider, M. (1995): Wavelets – An Analysis Tool. Oxford University Press, USA.

[94] Hubbard, Barbara Burke (1998): The World According to Wavelets: The Story of a Mathematical Technique in the Making. Second edition. A. K. Peters, USA.

[95] Hull, John – White, Alan (1987): The Pricing of Options on Assets with Stochastic Volatilities. *Journal of Finance* **42**, 281–300.

[96] Härdle, Wolfgang – Kerkyacharian, Gerard – Picard, Dominique – Tsybakov, Alexander (1998): Wavelets, Approximation, and Statistical Applications. Springer, New York.

[97] Ito, Takatoshi – Lyons, Richard K. – Melvin, Michael T. (1996): Is There Private Information in the FX market? The Tokyo Experiment. Working Paper, University of California at Berkeley.

[98] Jensen, Mark J. (1998): An Approximate Wavelet MLE of Short and Long-Memory Parameters. *Studies in Nonlinear Dynamics & Econometrics* **3**, Article 5.

[99] Jensen, Mark J. (2000): An Alternative Maximum Likelihood Estimator of Long-Memory Processes Using Compactly Supported Wavelets. *Journal of Economic Dynamics and Control* **24**, 361–387.

[100] Jensen, Mark J. – Whitcher, Brandon (2000): Time-Varying Long-Memory in Volatility: Detection and Estimation with Wavelets. Technical Report, University of Missouri at Columbia and EURANDOM.

[101] Johnstone, Iain M. – Kerkyacharian, Gerard – Picard, Dominique (1992): Estimation d'une densité de probabilité par méthode d'ondelettes. *Comptes Rendus Acad. Sciences Paris A* **315**, 211–216.

[102] Karpoff, Jonathan M. (1987): The Relation Between Price Changes and Trading Volume: A Survey. *Journal of Financial and Quantitative Analysis* **22**, 109–126.

[103] Kroner, Kenneth F. – Ng, Victor K. (1998): Modeling Asymmetric Comovements of Asset Returns. *Review of Financial Studies* **11**, 817–844.

[104] Lamoureux, Christopher G. – Lastrapes, William D. (1990a): Heteroskedasticity in Stock Return Data: Volume versus GARCH Effects. *Journal of Finance* **45**, 221–229.

[105] Lamoureux, Christopher G. – Lastrapes, William D. (1990b): Persistance in Variance, Structural Change, and the GARCH Model. *Journal of Business and Economic Statistics* **8**, 225–234.

[106] Lo, Andrew W. – MacKinlay, A. Craig (1999): A Non-Random Walk Down Wall Street. Princeton University Press, New Jersey.

[107] Lobato, L.N. – Savin, N.E. (1998): Real and Spurious Long-Memory Properties of Stock-Market Data. *Journal of Business and Economic Statistics* **16**, 261–267.

[108] Lockwood, Larry J. – Linn, Scott C. (1990): An Examination of Stock Market Return Volatility During Overnight and Intraday Periods, 1964–1989. *Journal of Finance* **45**, 591–601.

[109] Lynch, Paul E. – Zumbach, Gilles O. (2003): Market Heterogeneities and the Causal Structure of Volatility.

[110] Malliavin, Paul – Mancino, Maria Elvira (2002): Fourier Series Method for Measurement of Multivariate Volatilities. *Finance and Stochastics* **6**, 49–61.

[111] Mandelbrot, Benoit (1963): The Variation of Certain Speculative Prices. *Journal of Business* **36**, 394–419.

[112] McKenzie, Michael D. (1999): Power Transformation and Forecasting the Magnitude of Exchange Rate Changes. *International Journal of Forecasting* **15**, 49–55.

[113] Merton, Robert C. (1980): On Estimating the Expected Return on the Market: An Exploratory Investigation. *Journal of Financial Economics* **8**, 323–361.

[114] MEYER, YVES (1994): Wavelets: Algorithms and Applications. Second printing. Society for Industrial and Applied Mathematics.

[115] MÜLLER, U. A. – DACOROGNA, M. M. – OLSEN, R. B. – PICTET, O. V. – SCHWARZ, M. – MORGENEGG, C. (1990): Statistical Study of Foreign Exchange Rates, Empirical Evidence of a Price Change Scaling Law, and Intraday Analysis. *Journal of Banking and Finance* **14**, 1189–1208.

[116] MÜLLER, U. A. – DACOROGNA, M. M. – DAVE, R. D. – PICTET, O. V. – WARD, R. B. (1995): Fractals and Intrinsic Time, a Challenge to Econometricians. Olsen & Associates Discussion Paper.

[117] MÜLLER, ULRICH A. – DACOROGNA, MICHEL M. – DAVÉ, RAKHAL D. – OLSEN, RICHARD B. – PICTET, OLIVIER V. – VON WEIZSÄCKER JACOB E. (1997): Volatilities of Different Time Resolutions – Analyzing the Dynamics of Market Components. *Journal of Empirical Finance* **4**, 213–239.

[118] NASON, GUY P. – VON SACHS, RAINER (1999): Wavelets in Time Series Analysis. *Phil. Trans. R. Soc. Lond.* A ????

[119] NORSWORTHY, JOHN R. – LI, DING – GORENER, RIFAT (2000): Wavelet-based Analysis of Time Series: An Export from Engineering to Finance. Proceedings of the 2000 IEEE International Engineering Management Society Conference, Albuquerque, New Mexico. Http://www.norsworthy.net/papers.php.

[120] OGDEN, TODD (1996): On Preconditionng the Data for the Wavelet Transform When the Sample Size is Not a Power of Two.

[121] OSLER, C. L. (1995): Exchange Rate Dynamics and Speculator Horizons.

[122] PERCIVAL, DONALD B. – WALDEN, ANDREW T. (2000): Wavelet Methods for Time Series Analysis. Cambridge University Press.

[123] POLZEHL, JÖRG – SPOKOINY, VLADIMIR – STĂRICĂ, CĂTĂLIN (2004): When did the 2001 Recession Really End?

[124] POON, SER-HUANG – GRANGER, CLIVE W.J. (2003): Forecasting Volatility in Financial Markets: A Review. *Journal of Economic Literature* **XLI**, 478–539.

[125] POTERBA, JAMES M. – SUMMERS, LAWRENCE H. (1986): The Persistance of Volatility and Stock Market Fluctuations. *The American Economic Review* **76**, 1142–1151.

[126] PRESS, WILLIAM H. – TEUKOLSKY, SAUL A. – VETTERLING, WILLIAM T. – FLANNERY, BRIAN P. (1992): Numerical Recipes in Fortran 77: The Art of Scientific Computing. Cambridge University Press. Online version: http://www.library.cornell.edu/nr/bookfpdf.html.

[127] PRIESTLEY, M. B. (1992): Spectral Analysis and Time Series. Academic Press, San Diego.

[128] PRIESTLEY, M. B. (1996): Wavelets and Time-Dependent Spectral Analysis. *Journal of Time Series Analysis* **17**, 85–103.

[129] RAMSEY, JAMES B. (1996): The Contribution of Wavelets to the Analysis of Economic and Financial Data. Unpublished paper, http://www.econ.nyu.edu/user/ramseyj/publish/publish.htm.

[130] RAMSEY, JAMES B. – LAMPART, CAMILLE (1998a): Decomposition of Economic Relationships by Time Scale Using Wavelets: Money and Income. *Macroeconomic Dynamics* **2**, 49–71.

[131] RAMSEY, JAMES B. – LAMPART, CAMILLE (1998b): The Decomposition of Economic Relationships by Time Scale Using Wavelets: Expenditure and Income. *Studies in Nonlinear Dynamics and Econometrics* **3**, 23–42.

[132] Rogers, L. C. G. – Williams, D. (19??): Diffusions, Markov Processes and Martingales Volume 2: Ito Calculus. John Wiley & Sons.

[133] ROLL, RICHARD (1984): A Simple Implicit Measure of the Effective Bid-Ask Spread in an Efficient Market. *Journal of Finance* **39**, 1127–1139.

[134] SCHLEICHER, CHRISTOPH (2002): An Introduction to Wavelets for Economists. Working Paper 2002-3, Monetary and Financial Analysis Department, Bank of Canada.

[135] SHANN, W. C. – YEN, C. C. (1999): On the Exact Values of Orthonormal Scaling Coefficients of Lengths 8 and 10. *Applied and Computational Harmonic Analysis* **6**, 109–112.

[136] SHIRYAEV, A. N. (1999): Essentials of Stochastic Finance: Facts, Models, and Theory. World Scientific, Singapore.

[137] SHLEIFER, ANDREI – VISHNY, ROBERT W. (1990): Equilibrium Short Horizons of Investors and Firms.

[138] SCHWERT, G. WILLIAM (1989): Why Does Stock Market Volatility Change Over Time? *Journal of Finance* **44**, 1115–1153.

[139] TAYLOR, STEPHEN (1986): Modeling Financial Time Series. John Wiley & Sons, New York.

[140] TAYLOR, STEPHEN (2000): Consequences for Option Pricing of a Long Memory in Volatility. Manuscript, Lancaster University.

[141] Tkacz, Greg (2000): Estimating the Fractional Order of Integration of Interest Rates Using a Wavelet OLS Estimator. Working Paper 2000-5, Bank of Canada.

[142] Tsay, W. J. – Chung, C. F. (2000): The Spurious Regression of Fractionally Integrated Processes. *Journal of Econometrics* **96**, 155–182.

[143] Vidakovic, Brani (1999): Statistical Modeling by Wavelets. John Wiley & Sons.

[144] Vilasuso, Jon (2002): Forecasting Exchange Rate Volatility. *Economics Letters* 76, 59–64.

[145] Walter, Gilbert G. (1994): Wavelets and Other Orthogonal Systems With Applications. CRC Press, USA.

[146] Wasserfallen, W. – Zimmermann, H. (1985): The Behavior of Intraday Exchange Rates. *Journal of Banking and Finance* **9**, 55–72.

[147] Whitcher, Brandon – Jensen, Mark J. (2000): Wavelet Estimation of a Local Long Memory Parameter. *Exploration Geophysics* **31**, 94–103.

[148] Wiggins, James B. (1987): Option Values Under Stochastic Volatility. *Journal of Financial Economics* **19**, 351–372.

[149] Woljtaszczyk, P. (1997): A Mathematical Introduction to Wavelets. Cambridge University Press, Cambridge, UK.

[150] Wood, Robert A. – McInish, Thomas H. – Ord, J. Keith (1985): An Investigation of Transaction Data for NYSE Stocks. *Journal of Finance* **40**, 723–739.

[151] Wright, Jonathan H. (2000): Log-Periodogram Estimation of Long Memory Volatility Dependencies with Conditionally Heavy Tailed Returns. International Finance Discussion Papers No. 685.