1

Variance Bounds in Financial Inverse Problems: Simulation Evidence from Time-Scale Estimators

Enrico Capobianco

Abstract

An important source of research work in applied financial mathematics comes from the idea of exploiting high frequency market returns for estimating the volatility process underlying their observed dynamics. The main result which is obtained is that arbitrarily accurate volatility estimates are found by just using more efficiently the available information set, or filtration, and while adopting increasingly finer sampling rates. The stochastic characterization of the problem relies on a semimartingale approach for representing asset prices, and on the analysis of realised and integrated volatility measures. The former volatility is an empirical approximation of the other one, and since it represents an empirical counterpart of an integral form, is given by an average of high frequency returns observed within a certain time frame, for instance intra-daily, so to obtain an estimate of the daily volatility. Semimartingales are useful because allow for the quadratic variation principle to hold, and this means the convergence of the realised to the integrated volatility can be verified. Conversely, from both theoretical and experimental standpoints, the cumulative squared high frequency returns can be seen as consistent estimators of the integrated volatility. The goal of this work is to show with simulations that the convergence properties hold also when time-scale coordinates are considered and wavelet-based estimators for the volatility process are adopted.

1 Introduction

Financial mathematics studies of volatility have inspired a vast literature from both theoretical and applied aspects, and gradually provided with a diversity of methods and instruments for modeling market dynamics and discovering volatility features.

*Research supported by the *Mathematical Sciences Research Institute*, Berkeley, and in part conducted while the author was at the *Centrum voor Wiskunde en Informatica*, Amsterdam.
MSC2000: 60G48, 62M15, 65T60, 93E03

As a consequence of the very different approaches and models which have been proposed, inspired by econometrics and financial economics, statistics and econophysics, these instruments can be more justified for certain scopes than others, depending on the application field.

In this paper, for instance, the idea is to provide the reader with simple methodological tools employed in experiments typically found in applied probability and computational harmonic analysis work.

We present simulations and numerical experiments involving the volatility process characterizing the dynamics of financial assets, stocks (indeces), portfolios, and work through modeling techniques which are not relying on particular constraints but lead to computationally efficient procedures for dealing with datasets of high-frequency information.

Coping with the informativeness of the data possibly sampled at different frequencies and keeping the problem setting general enough to embed several volatility models, is the chosen objective which stands behind our analysis.

The emphasis thus goes toward those modern statistical instruments able to explore returns and volatility in time, frequency and scale coordinates, as for instance is allowed by the multi-resolution tools for the analysis of asset prices stochastic dynamics.

2 Volatility

Given a realization sequence $\{r_t, t \in (0, T)\}$ (r_t from now) of a return generating stochastic process R_t, the underlying volatility processes describe second order dynamics of realizations of martingale differences:

$$\eta_t = r_t - E[r_t \mid \mathcal{F}_{t-1}] \tag{1}$$

where the increasing filtration of σ-fields $\mathcal{F}_{t-1}$ stands for the memory of the process, from the past up to time $t-1$. In other words, let $\mathcal{F}_{t-1}$ be the information available through the observed data r_t, usually financial returns.

Given the usual assumption $\eta_t / \mathcal{F}_{t-1} \sim N(0, \sigma_t)$, the conditional variance process σ_t is called volatility and it could be either non stochastic but simply time-varying according to the past information set $\mathcal{F}_{t-1}$ and to a certain dependence structure, or stochastic because independently driven by a disturbance process.

Linear and non-linear model formulations have been proposed, and parametric, semi-parametric and non-parametric statistical inference have been pursued in volatility studies [16].

Nowadays, one popular idea in financial mathematics is using the so-called realised volatilities as estimators of the integrated volatilites [1, 2, 6]; this aims to obtain an accurate measure based on high frequency observations. At a very general level, one finds the following relationship between the integrated σ_t^I and the realised volatility

σ_t^R, which is here expressed as a simple *signal + noise* model, being σ_t^R a noisy measure of σ_t^I:

$$\sigma_t^R = \sigma_t^I + \epsilon_t \tag{2}$$

In practical terms, given the returns $r_i(t)$ observed within a finite time interval, i.e., one trading day, week or so, these very high frequency information collected, say, every minute or five minutes, give rise to an empirical measure such as:

$$\hat{\sigma}_t^R = \sum_{i=1}^{T} r_{it}^2 \tag{3}$$

and an approximation to the integral of the unobservable variable by averaging a certain number of intraday values r_{it}^2 over a finite time interval, say one day or so. This leads to:

$$\hat{\sigma}_t^R \rightarrow \sigma_t^I = \int_0^t \sigma^I(s)ds \tag{4}$$

These two volatility measures underlie the observed return process and are linked to each other through the quadratic variation principle. Given a process R_t and the partition $T = \{t_0, t_1, \ldots, t_n\}$ of $[0, t]$, its quadratic variation over T is:

$$V_{t,T}^{(2)} = \sum_{k=1}^{n} [R_{t_k} - R_{t_{k-1}}]^2 \tag{5}$$

Then, if $\mid\mid T \mid\mid = max_{1 \leq k \leq n} \mid t_k - t_{k-1} \mid \rightarrow 0$, for the volatility case we have:

$$lim_{\mid\mid T \mid\mid \rightarrow 0} V_{t,T}^{(2)} = < R >_t \tag{6}$$

and the limit is the quadratic variation of X_t.

Furthermore, the convergence in probability to the daily volatility holds for the cumulative squared intraday returns:

$$\sum_{j=1}^{n} r_{n,j,t}^2 \rightarrow_{n \rightarrow \infty} \int_0^1 \sigma^I(t+\tau)d\tau \tag{7}$$

A very general class of stochastic processes suitable for asset price models is offered by semimartingales [24], according to which a general stochastic process X_t retaines the following classical continuous-time Doob-Meyer decomposition:

$$X_t = X_0 + A_t + M_t \tag{8}$$

with X_0 the baseline process value, A_t a predictable and bounded variation component, i.e., such that $\int_0^t \mid dA_s(w) \mid < \infty$ for $t > 0$ and $\omega \in \Omega$, and with M_t as a local martingale.

Some standard required conditions are:

- The probabilistic framework underlying this decomposition consists of a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, P)$;
- The process X_t is adapted to $(\mathcal{F}_t)_{t \geq 0}$ and *cadlag*[1];
- The σ-algebra $\mathcal{F}$ is P-complete and the right continuous flow of $\sigma-$algebras (or information) $(\mathcal{F}_t)_{t \geq 0}$ contains all the sets in $\mathcal{F}$ of P-probability zero.

The semimartingale class contains discrete-time sequences x_n of the given process, but also diffusions and other processes. It excludes instead the fractional Brownian motion B_t^H, except when the Hurst parameter $H \in (0, 1)$ is $\frac{1}{2}$, e.g., for which it corresponds to the Brownian motion.

We thus recall that by definition, a standard Brownian motion is a stochastic process B_t such that:

- B_0 and B_t is a continuous function of time;
- for each $t, s > 0$ then $B_{t+s} - B_t \sim N(0, s)$;
- $B_{t+s} - B_t$ is independent of B_t.

3 Inverse Problem Setup

One limitation of the realized volatility estimator comes from the fact that it is inherently noisy; when employing the quadratic variation idea the impact that inliers and outliers present in the data have on the statistical inference accuracy may result particularly emphasized.

Thus, robustness and de-noising issues consequently arise, likewise the role and performance of time and frequency domain estimators in handling them. It becomes key establishing statistical consistency, under these noisy conditions, and results are already available for both these classes of estimators (for general frequency-based estimators in semimartingale models see, in particular, [15], while for specific volatility applications see [19]).

In order to support the proposed plan of verifying the behavior of volatility estimators and their convergence properties with experimental simulation work, the quality of the convergence to the quadratic variation limit of volatility estimators as from Eq. 7 is now studied.

As far as concerns modeling volatility under some assumptions about the distribution of financial returns, there are several possible approaches based on optimization schemes, and likelihood or pseudo-likelihood functionals, or even other criteria.

[1]Right-continuous and with limits from the left.

However, also simulation-based methodologies have been shown to be feasible for approximating the volatility function, particularly in the presence of a stochastic characterization for the volatility process itself.

Statistical inference may thus rely on exact results under parametric assumptions or *ad hoc* simulation, or instead achieve approximate results under iterative or greedy procedures, which are typically proposed in the non-parametric statistics and machine learning approaches to the problem.

A more general strategy goes through an inverse problem setting, where one seeks to recover the latent volatility function from the observed realizations of the return process, or in other terms from the corresponding noisy realised volatilities.

This in turn brings to the attention a volatility deconvolution (or recovery) problem. A continuous-time model representation for the volatility recovery problem is:

$$dr_t = (\Theta\sigma)_t dt + \epsilon dW_t \tag{9}$$

while again r_t is a return realization and σ_t is the underlying volatility function (i.e., the unknown object we want to recover).

The operator Θ indicates that some convolution dynamics apply to the system, together with the influence of a W_t Wiener process.

A discrete-time approximated model is described:

$$r_n = (\Theta\sigma)_n + \xi_n \tag{10}$$

with r_n and σ_n acting respectively as the return and the volatility discrete sequences, and with $\xi_n \dim N(0, var_\xi)$.

Due to the relationship between integrated σ_t^I and realised volatilities σ_t^R already introduced in Eq. 2, and the fact that a return process described by a semimartingale process has an associated quadratic variation process measuring its sample-path variation, the setup corresponds to the approximation and estimation discrepancy between the realised volatility estimator and the integrated volatility, which is summarized by the error term of Eq. 2.

This error is more visible by looking at Eq. 7, where it can be evaluated locally, i.e., over certain time horizons and for measures obtained at very small $\tau = \Delta t$ discretized time intervals.

Furthermore, this general inverse problem and the last indicated equation suggest that pointwise estimates of the volatility function can be obtained from indirect measurements, which are here represented by the noisy observations of returns.

Together with the important role played by observational noise [8], another source of uncertainty comes from not knowing, in general, the form of the operator Θ, which represents the convolution law influencing the volatility dynamics.

While one may decide to turn to classical non-parametric inference techniques such as kernel or spline methods, also modern signal deconvolution techniques based on orthonormal wavelet-based approaches and other bio-othogonal variants can be adopted. These issues and the related methodological aspects are not discussed in this paper, but they were matter of investigation in [9].

The focus is instead on computing volatility estimators from the basic model described in Eq. 2, which calls for the observables quantities given in Eq. 10, and that in turn require the operational steps of Eq. 3 and Eq. 7, where the relations between squared intra-daily returns and, respectively, realized and integrated volatility are established, thus specializing the influence of the operator Θ appearing in Eq. 9-10.

The simplest way to see the above scheme applied is to simulate standard Brownian motion (BM) paths at different sampling rates, or time grids, within the 0-1 time interval, and then verifying with the quadratic variation property the associated volatility convergence to the variance bound, in this case the value 1.

Figures 1 and 2 report experiments about these simulations and describe the behavior of the time variance estimators with the pointwise (C-D) and cumulative (E-F) estimates computed at finer and finer time intervals.

One may observe that the achievement of the variance bound, meaning that a full convergence of the realized to the integrated volatility has taken place, is only almost obtained for 1000 sample points of the Brownian motion path, but achieved only for more finely sampled values, i.e., with 10^4 sample points (Figure 2). For a lower sampling frequency the bound is thus not completely achievable and the volatility estimator is not sufficiently fast.

4 Time-Scale Analysis

4.1 Spectral Background

The most important experimental tool for the anaysis of time series in the frequency domain is the periodogram $\hat{I}_T(w)$, for T data and w frequency points. The periodogram is an asymptotically unbiased estimator of the spectral density function (SDF) of the process, $f(w)$. In order to achieve consistency, a smoothed version of the estimator is usually needed [7, 22].

The periodogram of a general stochastic process is computed from one of its finite-time and discretely computed T realization, and is reported as:

$$\hat{I}_T(w) = \frac{1}{2\pi T} \mid \int_0^T x(t)e^{-iwt}dt \mid^2 \tag{11}$$

It holds that:

$$E[\hat{I}_T(w)] \rightarrow f(w), \;\; T \rightarrow \infty, \;\; w \neq 0 \tag{12}$$

The function $f(w)$ is related to the spectral decomposition of the process X into a superposition of harmonics, but such decomposition is not subject to a unique

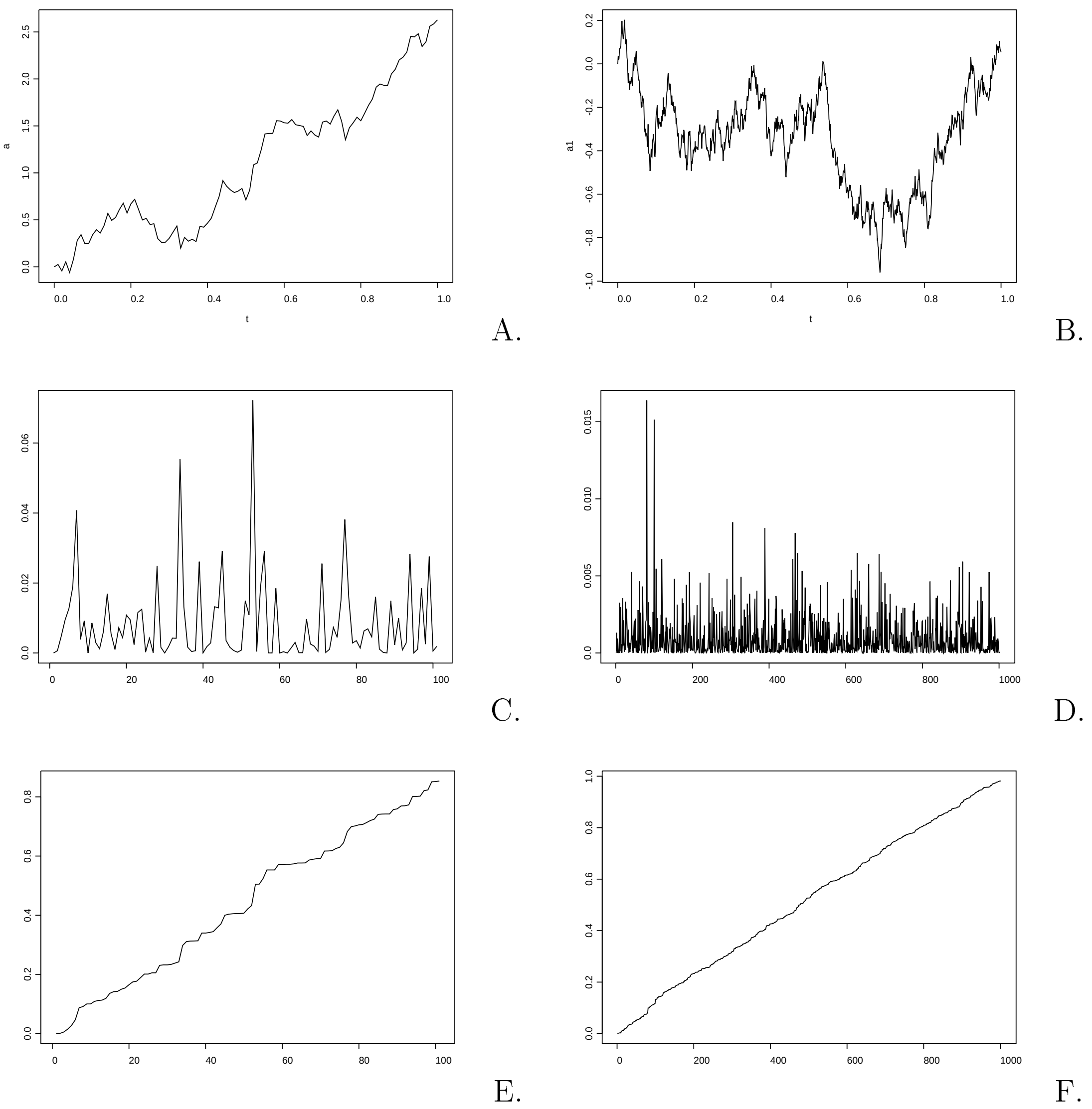
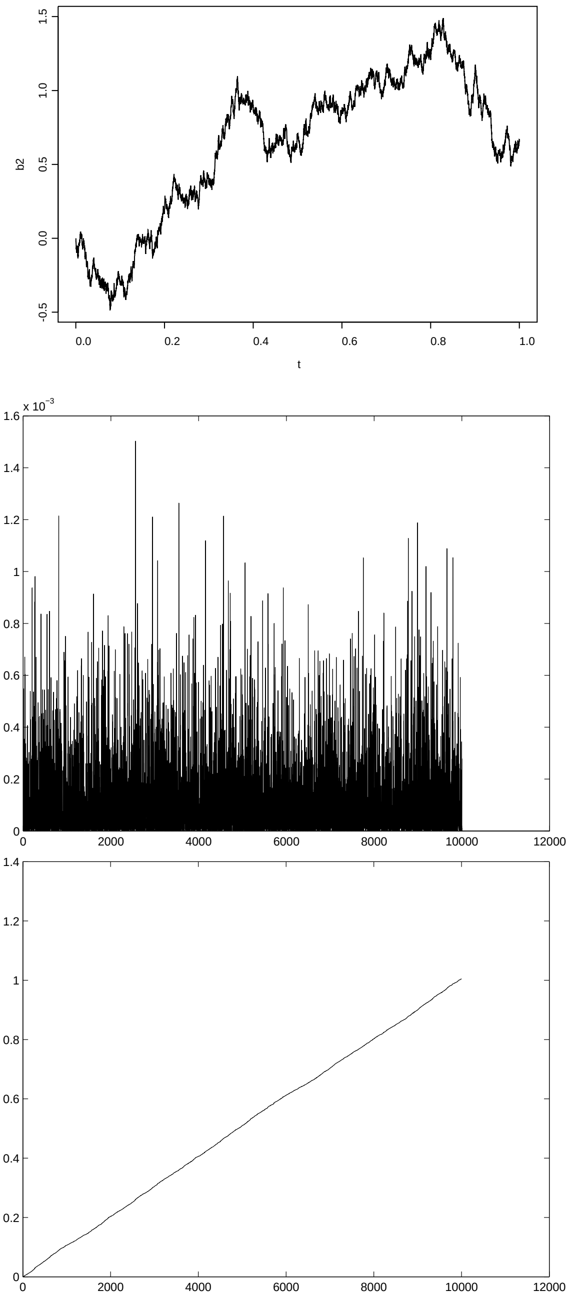


Figure 1: Brownian motion paths: (A) 100 and (B) 1000 samples. Corresponding pointwise variances in (C) and (D). Cumulative variances respectively at (E) and (F).



A.

B.

C.

Figure 2: Brownian motion path with 10000 sample points (top). Corresponding pointwise variance and cumulative variance (bottom).

restoration from the discrete observations of a continuous process, i.e., only a sampled process Δx.

The spectrum of a sequence of observations concentrates on the finite frequency band $-\frac{\pi}{\Delta} \leq w \leq \frac{\pi}{\Delta}$ that represents the main band, large enough for the power distribution belonging to frequencies higher the the Nyquist limits to offer more than just a negligible contribution to the total power of the process.

Consider the spectrum S_{X_t} of the process, which has variance function var_{X_t}. The spectrum decomposes the variance w.r.t. the continuous independent variable w (the frequency), such that:

$$\int_{-\frac{1}{2}}^{\frac{1}{2}} S_{X_t}(w)dw = var_{X_t} \tag{13}$$

The spacing Δ between observations is destined to go to 0 for consistency reasons and, accordingly, the number of discrete observations grows, so that the amount of information in the data increases too.

Another fact we know, besides the asymptotic unbiasedness of the periodogram, is that since its variance $f^2(w)$ is not depending on the sample size, in order to achieve consistency we need to smooth the periodogram across frequencies.

4.2 Wavelets

Although stationary phenomena benefit from the direct and efficient representation obtained in the frequency domain through the Fourier transform, the same instrument is not much effective for non-stationary processes.

Therefore, more accurate decompositions have been suggested and implemented with wavelets and their related multi-resolution analysis (MRA) [11, 18, 20]. The increased localization power that is achieved with wavelets, compared to the traditional Fourier transform, yields more spatial adaptivity.

This in turn might be very useful for handling all those phenomena subject to intermittent, burst-like, cluster-wise and regime switching behavior. When dealing, in particular, with non-homogeneous function classes characterization aspects, both the covariance diagonalization and the sparsity representation problems are key issues.

Before illustrating some simulations with these instruments, we offer a brief review of already well-known wavelet basics and their use for volatility representation.

A Discrete Wavelet Transform (DWT) is a time-scale sampling of a corresponding continuous transform and it suggests the following expansion for the function $f(x)$:

$$f(x) = \sum_k c_{j0,k}\phi_{j0,k}(x) + \sum_{j>j0}\sum_k d_{j,k}\psi_{j,k}(x), \tag{14}$$

where $\phi_{j0,k}$ is associated with the coarse resolution coefficients $c_{j0,k} = \int f(x)\phi_{jk}(x)dx$ and $\psi_{j,k}(x)$ refer to the detail coefficients $d_{j,k} = \int f(x)\psi_{jk}(x)dx$, and thus to the finer and finer information captured by the transform.

When approximating a realised volatility function, by relying for instance on the quadratic information as in Eq. 2, the above decomposition leads to the $L^2(R)$ volatility inner product representation obtained through wavelets as:

$$\sigma_t^{WR} = \sum_k < \sigma_t^R, \phi_{j_0,k} > \phi_{j_0,k}(t) + \sum_{j>j_0} \sum_k < \sigma_t^R, \psi_{j,k} > \psi_{j,k}(t) \tag{15}$$

where a smooth part is combined with a cumulated sequence of details obtained at different scales.

A perturbed version of the previous approximation is given, in line with Eq. 2, by the signal plus noise model:

$$\sigma_t^{WR} = \sigma_t^{WI} + \xi_t^W \tag{16}$$

where $\xi_t^W = W\epsilon_t$ is the transformed (through the wavelet matrix W) disturbance which preserves the Gaussian characteristics of the noise process.

In order to sparsify the estimator, we can apply noise shrinkage techniques [12, 13, 14], use function expansions in different bases or in overcomplete dictionaries [10].

Thus, instead of considering the L^2 estimators, we may consider other function classes suitable for handling inhomogeneous behavior, and then build non-linear estimators for the volatility function.

A key fact occurs in the wavelet coefficients domain with a Gaussian stationary and dependent (or weakly autocorrelated) process is the fact that it is decomposed in a series of detail signals which are independent across scales and temporally uncorrelated along each resolution level, but the same property of decorrelation power can only be partially preserved under non-Gaussianity, self-similarity and long range dependence features of the processes.

4.3 Multi-resolution

In order to build such a wavelet system one needs (**A**) a scaling function ϕ whose dilates and translates constitute orthonormal bases for all those V_j subspaces that are obtained as scaled versions of the subspace V_0 to which ϕ belongs, and (**B**) a mother wavelet ψ together with the ψ_{jk} generated by j-dilations and k-translations, such that $\psi_{jk}(x) = 2^{\frac{j}{2}}\psi(2^j x - k)$.

From the information obtained with the approximations computed at successively coarser resolution levels, an MRA is obtained, or in other words a sequence of closed subspaces expressed in nesting order as in a ladder of Sobolev spaces. The more negative the index, the larger the space satisfying $\ldots, V_2 \subset V_1 \subset V_0 \subset V_{-1} \subset V_{-2} \subset \ldots$, with $\bar{\cup}_{j\in Z} V_j = L_2(R)$, $\cap_{j\in Z} V_j = \{0\}$ and the additional condition $f \in V_j \iff f(2^j.) \in V_0$.

The last condition is a necessary requirement for identifying the MRA, meaning that all the spaces are scaled versions of a central space, V_0. Thus, an MRA

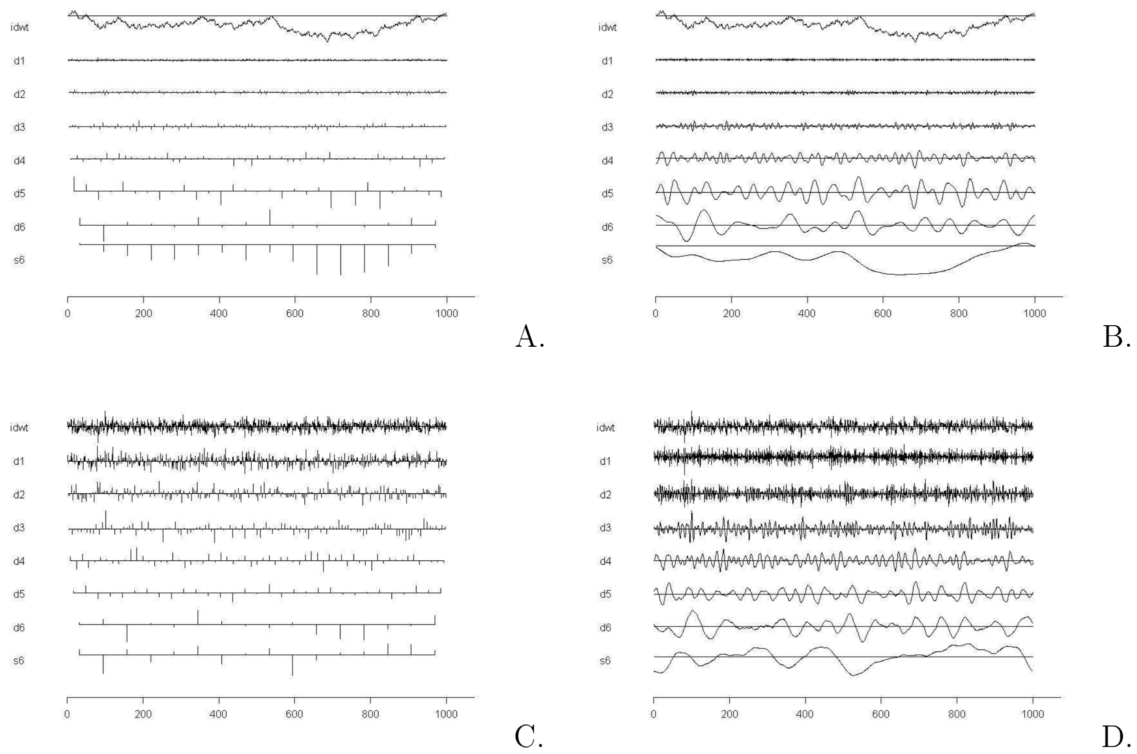


Figure 3: DWT decomposition (decimated in A-C, non-decimated in B-D) for sampled 1000 points of BM path (A-B) and corresponding increment process (C-D).

approximates for instance an $L_2[0,1]$ space through the V_j sub-spaces generated by orthonormal scaling functions ϕ_{jk}, where $k = 0, \dots, 2^j - 1$, and the sequence of 2^j wavelets ψ_{jk}, $k = 0 \dots, 2^j - 1$, represent an orthonormal basis of $L_2[0,1]$.

Figure 3 illustrates the multi-resolution decomposition obtained with decimated and non-decimated wavelets for the standard BM path sampled at 1000 points, and for the corresponding increment process. In the latter case the number of detail coefficients remains the same when changing resolution level, thus avoiding the decimation that occurs in the former case by means of the filters.

5 Simulation-based Inference

5.1 Time-scale Variance Measures

Wavelet analysis may reveal when transient phenomena occur at just some scales, thus giving rise to different regimes of dynamics.

The wavelet analysis of power-law (or $\frac{1}{f}$) processes which are well-known in several application fields, such as networks and teletraffic, geophysics, bioinformatics, economics etc., has been proposed in particular by [3, 4, 5, 23] and will not be studied further in this paper.

Here instead the attention is more on the wavelet variance and spectral counterparts, following the work, among others, of [21], and applying it to our specific volatility context, which in some ways extend previous work by [8, 9].

The first aspect to note is that the wavelet coefficients vector obtained from the application of the the wavelet transform W to the process return realization r_n delivers a smooth s and a detail d sub-vector component. Thus, we expect to be able to pick up the low frequency content of the volatility dynamics in the first component, and the other higher frequency features in the details.

The system of filters which works in the wavelet decomposition is pretty much reported in any specialized book, and here is enough to remind that the filters are distinguished in low-pass (scaling) and band-pass (wavelet), because they are characterized by different lengths and transfer functions.

The scaling filter has a passband of $-\frac{1}{4} \leq \mid w \mid \leq \frac{1}{4}$. Conversely, in terms of the λ scale band, we have that the wavelet filter has passband $\frac{1}{4\lambda} \leq \mid w \mid \leq \frac{1}{2\lambda}$ and the scaling filter has passband $-\frac{1}{4\lambda} \leq \mid w \mid \leq \frac{1}{4\lambda}$.

When considering the output d of the wavelet filter of scale λ, it results a stationary process, with zero mean and spectrum:

$$S_{W_\lambda}(w) = H_\lambda(w) S_X(w) \tag{17}$$

for H_λ as the wavelet filter transfer function.

The wavelet variance at scale λ of the new process is:

$$\sigma^2_W(\lambda) = \frac{E(d^2_{t,\lambda})}{2} = \frac{1}{2\lambda} \int_{-\frac{1}{2}}^{\frac{1}{2}} H_\lambda(w) S_X(w) dw \tag{18}$$

Being frequency and scale inversely linked, one can find an approximate expression by changing the range of the integral:

$$\sigma^2_W(\lambda) \approx 2 \int_{\frac{1}{4\lambda}}^{\frac{1}{2\lambda}} S_X(w) dw \tag{19}$$

Some form of regularization occurs when computing an average of the spectrum over a frequency interval. Conversely, one may also think of the spectrum estimation problem as an ill-posed problem, where given T we cannot uniquely determine its estimate over the w range.

The wavelet coefficients at scale $\lambda_j = 2^{j-1}$ are associated to frequencies in $(\frac{1}{2^{j+1}}, \frac{1}{2^j})$. This is particularly useful when the spectrum shows at first no particular features within the octave bands; estimating the wavelet variance can be more indicated than adopting classical non-parametric spectral estimates, due to a better localization power. We have:

$$\sigma^2_W(\lambda_j) \approx 2 \int_{\frac{1}{2^{j+1}}}^{\frac{1}{2^j}} S_X(w)dw \tag{20}$$

If $S_X(w) \propto \mid w \mid^\alpha$, over $w \in (-L, L)$, then the spectrum follows a power law. In wavelet terms one equivalently finds that $\sigma^2_W(\lambda) \propto \lambda^{-\alpha-1}$ in a certain range of scales.

One usually checks the presence of a linear variation through the plot of a log wavelet variance vs (log) scales, and the slope of the line indicates the value of the α exponent (see [5] among others).

The discrete Fourier transform (DFT) estimates the spectrum via the periodogram, as said before, according to:

$$\hat{S}_R(w_l) = \frac{1}{T} \mid \sum_{t=1}^{T} (R_t - \bar{R}) e^{-i2\pi w_l t} \mid^2 \tag{21}$$

for $w_l = \frac{l}{T}$. Then, a sampling version of the periodogram is:

$$\frac{1}{T} \sum_{l=-\frac{T}{2}+1}^{\frac{T}{2}} \hat{S}_r(w_l) = \frac{1}{T} \sum_{t=1}^{T} (r_t - \bar{r})^2 \tag{22}$$

with a computational load of the algorithm equal to $O(T log_2(T))$.

Since the DWT produces from an original realization J sequences at finer and finer scales, say $d_{t,\lambda}$, $t = 1, \ldots, \frac{T}{2\lambda}$ with $\lambda = 1, 2, 4, \ldots, 2^{J-1}$, the wavelet variance estimate is:

$$\tilde{\sigma}^2_W(\lambda) = \frac{1}{T} \sum_{t=1}^{\frac{T}{2\lambda}} d^2_{t,\lambda} \tag{23}$$

and it holds that a scale-wise decomposition is produced for the variance of the process according to:

$$\sum_{j=0}^{J-1} \tilde{\sigma}^2_W(2^j) = \frac{1}{T} \sum_{t=1}^{T} (R_t - \bar{R})^2 \tag{24}$$

It also holds that the wavelet variance of a process X_t is a convolution of the squared modulus of the Fourier transform of the wavelets:

$$\sigma^2_X(\lambda_j) = E\hat{\sigma}^2_X(\lambda_j) = \frac{2}{T} \sum_t E\hat{d}^2_{tj} = \int_{-\pi}^{\pi} S_X(w) \mid \hat{\psi}_{tj}(w) \mid^2 dw \tag{25}$$

which gives the measure of the the spectrum at frequencies $w_j = 2^{-j} w_0$, where w_0 depends on the wavelet ψ_0, i.e., the constant bandwidth wavelet filter.

With wavelets, the transform is based on constant Q band-pass filters derived from a main filter of central frequency w_ψ and correspondent bandwidth around it, i.e., Δw_ψ. Changing the scale λ we have a change in frequency like $w(\lambda) = \frac{w_\psi}{\lambda}$.

Moreover, a time-scale spectral estimator is equivalently defined as the squared modulus of the DWT:

$$\tilde{S}_X(w_j) = \frac{1}{N}\sum_{t=1}^{N_j} d_{tj}^2 = \frac{1}{N}\sum_{n=1}^{N} |< X, \psi_{nj} >|^2 \tag{26}$$

5.2 Variance Decomposition and Bounds

The DWT is faster than the DFT, requiring just $O(T)$ operations. The unbiasedness of the estimator of σ_W^2 can be built from the $d_{t,\lambda}$ sequences when no influence from boundary conditions occurs, and the consistency can be shown by adopting some non-linear thresholding in the wavelet periodogram.

Restating what said before, the following wavelet decompositon holds:

$$\sum_{j=0}^{\infty} \sigma_W^2(2^j) = \sigma_{X_t}^2 \tag{27}$$

where $\sigma_W^2(2^j)$ is the wavelet variance associated to a discrete independent variable $\lambda = 2^j$, i.e., the scale, with j octaves.

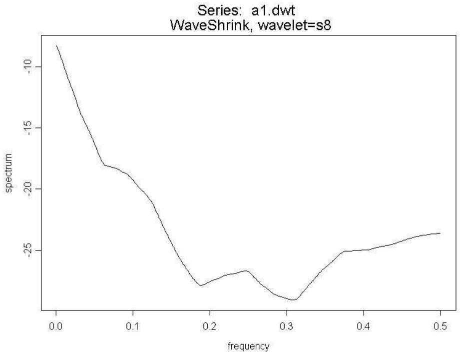
The wavelet variance replaces the idea of estimating the sample variance via the variability across observations with the variability across scales. It also regularizes the spectrum because while the latter decomposes the variance of the process across frequencies, the wavelet variance does the same with respect to the scales, measuring how much a weighted average of the process changes along time periods, given a certain bandwidth λ.

As for the periodogram estimates of the spectrum, measures are in decibels; one applies the transform $10 * log_{10}(x)$ to x, i.e., the estimated power spectrum. The plots in Figure 4 have the frequency range $0 - .5$ and compare the estimated spectra of the BM paths and the corresponding increment process with 1000 sample points. The measures are in decibels (top and middle location plots), but they are also back-transformed (E-F plots) from the logarithms.

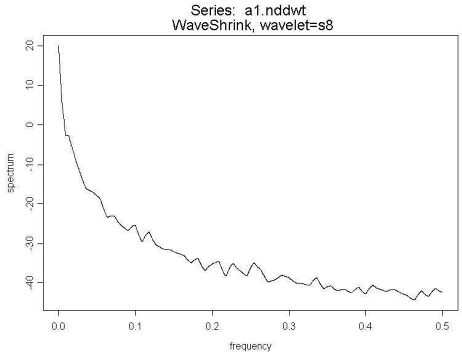
Note that one can use the stationary or non-decimated wavelet transform to get more efficient estimates; it requires $O(Tlog_2(T))$ operations and preserves the same number of wavelet coefficients at each scale.

However, the stationary wavelet decomposition is an overcomplete representation of the original process, and this redundancy shows up in the power decomposition, where substantial aliasing appears across scales, thus suggesting that it should be handled appropriately.

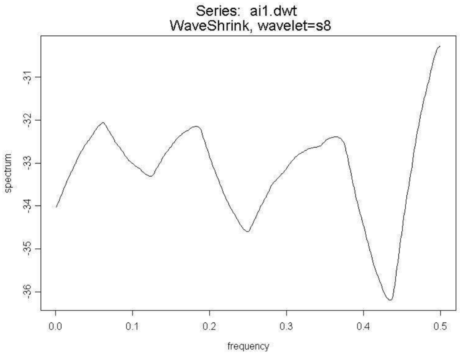
By looking at the cumulative periodograms from both the decimated and the non-decimated wavelet decompositions, one can observe that, in order to reduce the impact of the expected redundancy of the latter, the behavior in a restricted range of values has been considered together with the contribution from the finest scales to the power decomposition, and the convergence to the variance bound (see Figure 5).



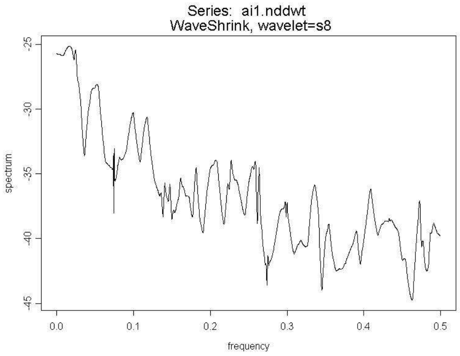
A.



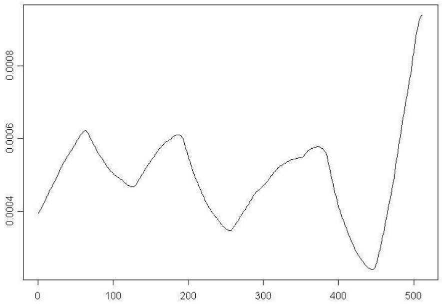
B.



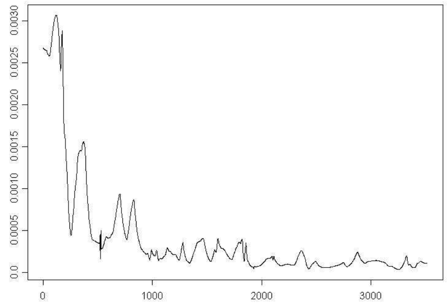
C.



D.



E.



F.

Figure 4: Spectrum estimates of decimated (left-hand) and non-decimated (right-hand) DWT of 1000 sample values of BM path (A-B), and corresponding increment process (C-D). Transformed spectrum for BM path increments with decimated DWT (E) and non-decimated ones (F).

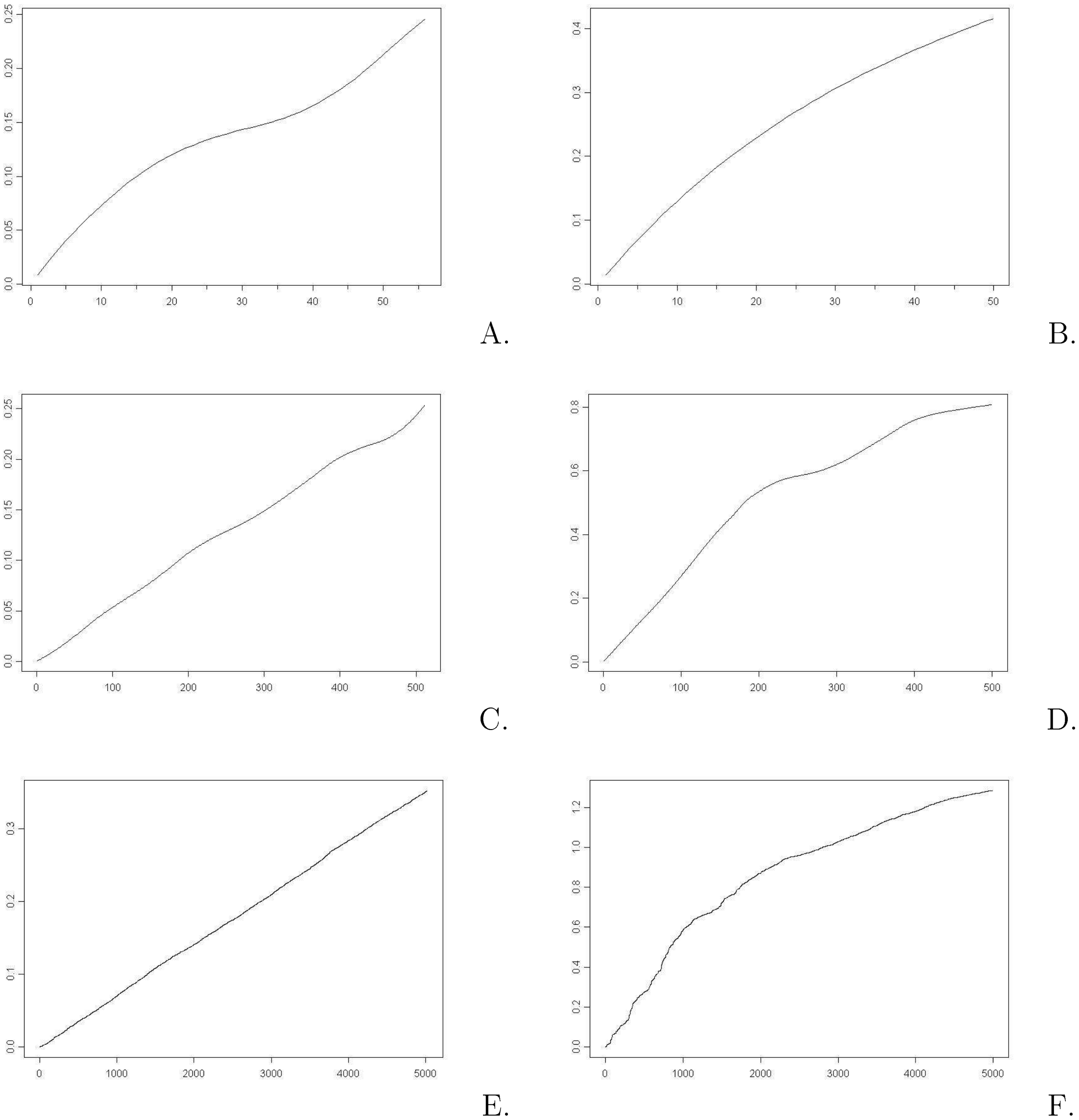


Figure 5: Cumulative periodograms from decimated and non-decimated DWT (from left to right) for increments of BM paths with 100 (top), 1000 (mid) and 10000 (bottom) samples.

The choice of the finest scales performance is clearly suffering again from a redundancy factor, as it can be seen from the overshooting behavior in Figure 5, plots (D,F). This could be due to the presence of noise mixed to more informative signal structure at very fine scales, typical of phases of the market characterized by intense and speculative trading and volatility bursts.

However, it cannot be ignored the overall performance in terms of convergence speed; by looking at the same plots above indicated, the 0.5 bound is reached in less than 1000 sample points (Fig. 5-F), which compared to the decimated wavelet (Fig. 5-E) cumulative periodogram estimator is very fast.

When compared to the previously adopted time estimator at the finest sampling time grid, i.e., 10000 sample points as in Fig. 2-C, the behavior of the wavelet estimator in terms of convergence speed is similar, i.e., of the same order of magnitude.

On the other hand, when looking at coarser scales, the same wavelet estimator seems to offer a much more limited contribution with this regard (Figure 6).

Thus, the advantage of using wavelet periodogram estimators of the cumulative volatility is the possibility of explaining the role played by the different scales into which the return process can be decomposed, or in other words, their contribution to the available filtration, in our case almost completely explained by the finest scales, apart from the noise impact which has not been considered yet.

As final remarks, what we can state is that for the new coefficient sequence d obtained by filtering the observed realization with the WT, the mean value is null; thus, an unbiased estimate of the wavelet variance is found.

Given some standard conditions on a process X, the estimator of the wavelet variance via the stationary wavelets has been shown to be asymptotically normally distributed, unbiased and with scale-wise variance equal to the spectrum of the mean corrected squared wavelet coefficients process at each scale [23]. This topic is currently under investigation for general classes of processes which include volatility dynamics.

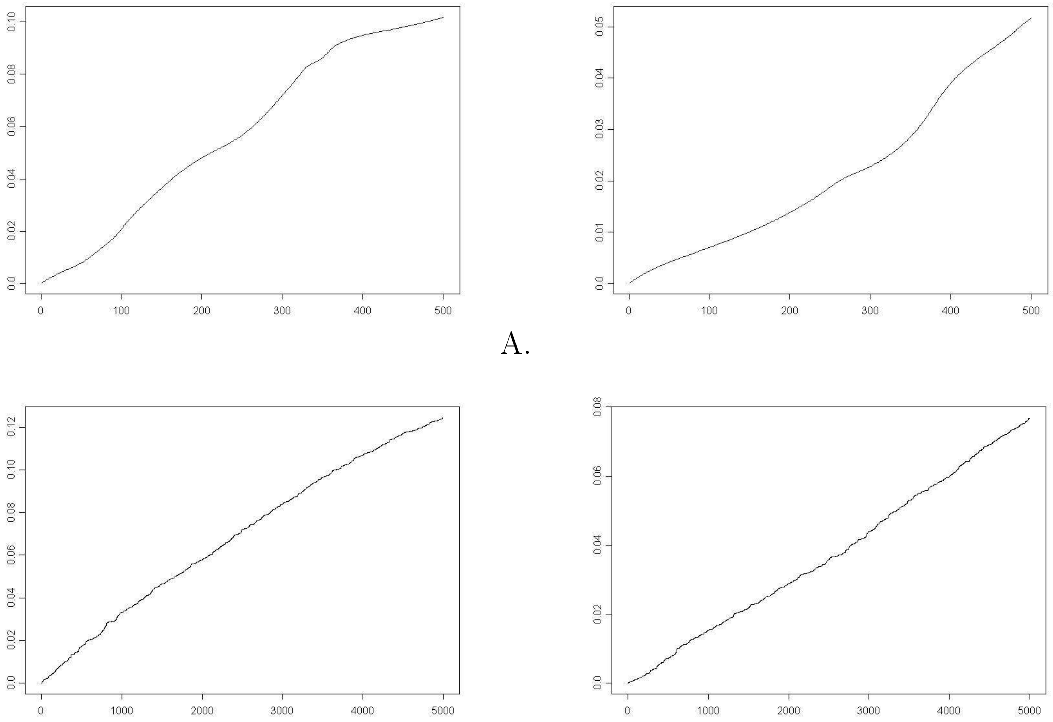
6 Conclusive Remarks

The main conclusions that we have reached from this study based on simulation experiments can be summarized as follows:

- Financial time series embedded in semimartingales can exploit the use of the quadratic variation principle for verifying the consistency of volatility estimators in both time and frequency, or conversely scale, coordinates;

- Simulating standard Brownian motion paths at different grids allows to measure the speed of convergence to the variance bounds of the volatility estimators, thus examining the discrepancy between realized and integrated volatility;

- Decimation involves losing convergence speed at different sample grids compared with time estimators, and this is evident from the plots of the cumulative volatilities and how they approach the variance bounds;



A. B. C. D.

Figure 6: Cumulative periodograms from non-decimated DWT for increments of BM paths with 1000 (A,B) and 10000 (C,D) samples. Fine (A-C) and coarse scales (B-D) are considered.

- Stationary wavelets establish good performances by achieving the variance bounds relatively fast, particularly when comparing finely sampled time estimators to the finest scales of the decomposed volatility process; however, they give rise to redundancy, and thus the power decomposition may be no longer accurate without taking care of aliasing.

In conclusion, due to the de-noising properties of wavelet-based spectral estimates, which can be investigated in an adaptive way with shrinkage-like techniques, our preliminary results indicate that one may be able to achieve a good compromise between fast convergence of the realised to the integrated volatility, while limiting at the same time the detrimental impact of the noise in the return process, together with its amplified effects when using the quadratic variation as a benchmark of volatility measures.

References

[1] T. Andersen, T. Bollerslev, F.X. Diebold and H. Ebens, The distribution of stock return volatility, *J. Finan. Econ.*, 61 (2001) pp. 43-76.

[2] T. Andersen, T. Bollerslev, F.X. Diebold and P. Labys, Modeling and forecasting Realized Volatility, *Econometrica*, 71(2) (2003) 579-625.

[3] P. Abry, P. Gonvalves and P. Flandrin, Wavelets, spectrum analysis and $\frac{1}{f}$ processes, in *Wavelets and Statistics*, A. Antoniadis and G. Oppenheim, eds., Springer-Verlag, (1995) pp. 15-30.

[4] P. Abry, D. Veitch and P. Flandrin, Long range dependence: revisiting aggregation with wavelets, *J. Time Ser. Anal.*, 19(3) (1998) pp. 253-266.

[5] P. Abry, P. Flandrin, M.S. Taqqu and D. Veitch, Wavelets for the analysis, estimation and synthesis of scaling data, in *Self-similar network traffic and performance evaluation*, C. Park and W. Willinger, eds., Wiley, New York, (2000) pp 39-88.

[6] O.E. Barndorff-Nielsen and N. Shephard, Non-Gaussian Ornstein-Uhlenbeck based models and some of their uses in financial economics (with discussion), *J. Roy. Statist. Soc. B*, 63, (2001) pp. 167-241.

[7] P.J. Brockwell and R.A. Davis, *Time Series: Theory and Methods*, Springer-Verlag, New York, 1991.

[8] E. Capobianco, Statistical Analysis of Financial Volatility by Wavelet Shrinkage, *Method. Comput. Appl. Probab.*, I(4), (1999) pp. 423-443.

[9] E. Capobianco, Multi-resolution Approximation for Volatility Processes, *Quantit. Fin.*, 2 (2002) pp. 91-110.

[10] S. Chen, D. Donoho and M.A. Saunders, Atomic Decomposition by Basis Pursuit, *SIAM Rev.*, 43(1) (2001) pp. 129-159.

[11] I. Daubechies, *Ten Lectures on wavelets*, SIAM, Philadelphia, 1992.

[12] D. Donoho and I.M. Johnstone, Ideal Spatial Adaptation via Wavelet Shrinkage, *Biometrika*, 81 (1994) pp. 425-455.

[13] D. Donoho and I.M. Johnstone, Adapting to unknown smoothness via wavelet shrinkage, *J. Amer. Statist. Assoc.*, 90 (1995) pp. 1200-1224.

[14] D. Donoho and I.M. Johnstone, Minimax Estimation via Wavelet Shrinkage, *The Ann. Statist.*, 26 (1998) pp. 879-921.

[15] K. Dzhaparidze and P. Spreij, Spectral characterization of the optional quadratic variation process, *Stoch. Proces. Appl.*, 54 (1994) pp. 165-174.

[16] E. Ghysels, A.C. Harvey and E. Renault, Stochastic Volatility, in *Handbook of Statistics* 14, G.S. Maddala and C.R. Rao, eds., Elsevier Science, Amsterdam, (1996) pp. 118-191.

[17] I. Karatzas and E. Shreve, *Brownian Motion and Stochastic Calculus*, Springer-Verlag, New York, 1998.

[18] S. Mallat, Multiresolution approximations and wavelet orthonormal bases of $L_2(R)$, *Trans. Amer. Mathem. Soc.*, 315 (1989) pp. 69-87.

[19] P. Malliavin and M.E. Mancino (2002), Fourier series method for measurement of multivariate volatilities, *Fin. Stochast.*, 6 (2002) pp. 49-61.

[20] I. Meyer, *Wavelets: algorithms and applications*, SIAM, Philadelphia, 1993.

[21] D.B. Percival, On estimation of the wavelet variance, *Biometrika*, 82(3) (1995) 619-631.

[22] M.B. Priestley, *Spectral Analysis and Time Series*, Academic Press Inc., London, 1981.

[23] A. Serroukh, A.T. Walden and D.B. Percival, Statistical properties and uses of the wavelet variance estimator for the scale analysis of time series, *J. Amer. Statist. Assoc.*, 95(449) (2000) pp. 184-196.

[24] A.N. Shiryaev, *Essentials of Stochastic Finance: Facts, Models and Theory*, World Scientific, Singapore, 1999.

Enrico Capobianco

Visiting under the Complementary Program the Mathematical Science Research Institute, 1000 Centennial Drive, Berkeley, California 94720

enrico.capobianco@libero.it