Brazilian Journal of Probability and Statistics (2006), 20, pp. 111–132.
©Associação Brasileira de Estatística

Non-parametric volatility estimation in continuous time

Aluísio Pinheiro, Neale A. El-Dash and **Luiz K. Hotta**
Universidade Estadual de Campinas

Abstract: The idea of volatility is fundamental to precise definition of risk and, hence, its estimation (or prediction) is a very important task in finance applications. We present some ideas on non-parametric estimation of volatility function in diffusion models. A nonlinear wavelet estimate of the volatility function is proposed and its performance is compared to three kernel estimators in both simulated and real data. Simulation is developed for eight volatility shapes and some interesting, but not unexpected, results are presented. Some issues such as online estimation and prediction, robustness to oversmoothing and performance under sudden changes in pattern of volatility are also discussed.

Key words: High frequency financial series, non-parametric curve estimation, volatility estimation, wavelets.

1 Introduction

Conditional volatility estimation is of great importance in finance applications. Parametric models such as the ARCH family are quite used as well as stochastic volatility models. References on parametric volatility modeling are Ghysels et al. (1996), Bollerslev et al. (1992) and Shephard (1996). Even though parametric procedures are still the dominating statistical tools in volatility estimation, non-parametric alternatives have increased their presence in the last decade (Bosq, 1998, Antoniadis and Oppenheim, 1994 and Van Es et al., 2005). The reason for that change in venues is the fast development of computational power and new (and faster) paradigms such as the wavelet decomposition as well as the search for more robust models.

Non-parametric applications range from historic simulations to sieves (Darolles and Gouriéroux, 2001), passing through semiparametrics (Yang, 2000), local polynomial smoothing (Härdle and Tsybakov, 1997 and Yang et al., 1999) and conditional distribution quantiles via kernel (Abberger, 1997).

A closely related paper is that of Hoffmann (1999), that uses nonlinear wavelet techniques for discrete AR(1) models and shows the adaptable nature of those estimators, on the sense of Donoho and Johnstone (1995). A recent and important work in non-parametric estimation of volatility is Mercurio and Spokoiny (2004), in which local power fitting estimators are used for changing point types of volatility models.

High frequency data bring two additional challenges to statistical methodologies. The dependence structure is amplified as well as the effects of computational costs can be dramatically increased. Computational efficiency is a big advantage for wavelets application on those cases.

This work has as initial methodological motivations Bertrand (1996), Avesani and Bertrand (1997) and specially Genon–Catalot et al. (1992), whose theoretical results either motivated or challenged wavelet application in financial data. We compared kernel and wavelet based procedures in both simulated and real data sets. Emphasis was put in the study of the presence of outlier and structural changes in the volatility function. Traditionally in these cases we could take two approaches. In the first approach we could initially use a statistical procedure in order to detect the presence of outlier and structural changes and formulate the model to take them into account. Although this is not the approach considered the estimated volatility function will give some insights about the presence of outlier and structural changes as will be seen later. In the second approach, which is adopted here, we look for estimators which can handle these facts, i.e. estimators which are robust in handling with outlier and structural changes. We illustrate the natural and automatic adaptation of wavelets to outliers (as can be seen for a simpler and theoretical set-up in Walter (1992a,b). Some discrepancies between our results and the one found in Bertrand (1996) and Avesani and Bertrand (1997) are pointed out and discussed as well as some indications simulation provide of broader use of both wavelet bases and kernel functions not considered on the literature.

The text goes as follows. In Section 2, the underlying stochastic structure is provided by the presentation of diffusion models, non-parametric estimators are revised and their adaptations to volatility diffusion models are presented. In Section 3, an automatic data-driven thresholding method for the nonlinear wavelet estimator is proposed. Illustrations by simulation and real data analysis are made in Section 4. Some final remarks, conclusions and directions for future research are presented in Section 5.

2 Nonparametric volatility estimation in diffusion models

Formal definition and mathematical treatment of diffusion models are fundamentally based on the Wiener process and Ito's integral and process.

Definition 2.1 (Diffusion Model). *For every functions* $b : [0, \infty) \times \omega \to \mathbb{R}$ *and* $\sigma : [0, \infty) \times \omega \to \mathbb{R}$ *for which Ito's integral exist, one defines the following diffusion model:*

$$X_t = X_0 + \int_0^t b(s, X_s)ds + \int_0^t \sigma(s, X_s)dW_s, \tag{2.1}$$

where W_s *is a Wiener process.*

The function $b(\cdot,\cdot)$, known as *drift*, measures the *level* of the diffusion while $\sigma(\cdot,\cdot)$, known as diffusion coefficient, relates to the *speed* and *size* of the series oscillations. The diffusion model (2.1) can be written as:

$$dX_t = b(t, X_t)dt + \sigma(t, X_t)dW_t, \tag{2.2}$$

which is the most accepted notation and the one adopted here.

One can show that the volatility in diffusion models, defined as the variance of dX_t conditioned on $\{X_s\}_{s\leq t}$ is given by

$$Var(dX_t/\{X_s\}_{s\leq t}) = \sigma^2(t, X_t).$$

For diffusion model simulations, a discretization of the continuous process is necessary. We implemented the Euler's method.

Definition 2.2 (Euler's Method). *Let $dX_t = b(t, X_t)dt + \sigma(t, X_t)dW_t$ for $t \in [T_0, T]$. Euler's method for simulation of X_t uses the discretized time $T_0 = t_0 < t_1 < ... < t_n = T$, given recursively as:*

$$X_{t_{k+1}} = X_{t_k} + b(t_k, X_{t_k})\Delta_{k+1} + \sigma(t_k, X_{t_k})Y_{t_{k+1}}$$

where $X_{t_0} = x_0$ is the initial guess, $\sigma(.,.)$ is the standard deviation level, $\Delta_k = t_k - t_{k-1}$ and $Y_{t_{k+1}} \sim \mathcal{N}(0, \Delta_{k+1})$.

Suppose (2.2) and take $\sigma^2(t, X_t) = \sigma^2(t)$, i.e.,

$$dX_t \quad = \quad b(t, X_t)dt + \sigma(t)dWt, \qquad t \in [0, T], \tag{2.3}$$

where $b(\cdot,\cdot)$ and $\sigma(\cdot)$ are unknown and $P(X_0 = x_0) = 1$, for some unknown x_0.

Several error measures for non-parametric functional estimators have been proposed in the literature. We refer the reader to Härdle (1990) and Bosq (1998) for a thorough treatment of those measures and some specific results.

2.1 Kernel volatility estimation

Suppose $\{X_t\}_{t\in T}$ is observed with $T = 1$, at times $(t_1, ..., t_n)$, $t_i = i/n$, where $\Delta = 1/n$ is the sampling interval. We define with no loss of generality (on T) the kernel estimator of $\sigma^2(\cdot)$ as follows, as originally proposed by Florens-Zmirou (1997).

Definition 2.3 (Kernel Estimator). *The volatility function $\sigma^2(t)$, as defined by (2.3), can be estimated by:*

$$S_n(t) = \Delta^{-1} \frac{\sum_{i=1}^{n-1} K_h(X_{t_i} - t)[X_{t_{i+1}} - X_{t_i}]^2}{\sum_{i=1}^{n} K_h(X_{t_i} - t)}, \tag{2.4}$$

where $K_h(.)$ is the selected Kernel function and h is the smoothing parameter.

The estimator defined by (2.4) is precisely the Nadaraya-Watson regression estimator when the response variable $Y_i = \Delta^{-1}[X_{t_{i+1}} - X_{t_i}]^2$ is the empirical quadratic variation process. The following theorem shows some of its most relevant asymptotic properties (due to Florens-Zmirou, 1993).

Theorem 2.1. *Let $\sigma^2(t)$ as defined in (2.3) and $S_n(t)$ given by (2.4). Suppose that $b(\cdot,\cdot)$ is limited, twice differentiable, with limited derivatives and that $\sigma(\cdot)$ has three limited and continuous derivatives and that there are two constants k and K such that $0 < k \leq \sigma(t) \leq K$. If, moreover, $nh^3 \to 0$, then:*

$$\sqrt{nh}\left(\frac{S_n(t)}{\sigma^2(t)} - 1\right) \xrightarrow{\mathcal{D}} L^{-1/2}(t)Z,$$

where Z has a standard normal distribution independent of $L(t)$ and

$$L(t) = \lim_{\delta\to 0}\frac{1}{2\delta}\int_0^1 1_{(|X_s - t|<\delta)}ds.$$

To our knowledge there are no general asymptotic results such as Theorem 2.1 on the Integrated Mean Squared Error (IMSE). For an estimates $\hat{m}(.)$ of $m(.)$, the IMSE is defined as

$$IMSE = E\left(\int_{\mathbb{R}}(\hat{m}(x) - m(x))^2 w(x)dx\right),$$

where $w(.)$ is a nonnegative weight function. We present below some estimators due to Bertrand (1996) which reduce the Nadaraya-Watson estimator to moving average, S_n^{MA}, and centered moving average, S_n^{CMA}, as follows.

$$\begin{aligned} S_n^{MA}(t) &= \frac{\displaystyle\sum_{j=0}^{n-1}\left(\sum_{i=1}^{A-1}[X_{t_{j-i+1}} - X_{t_{j-i}}]^2\right)\mathbf{1}_{[t_j,t_{j+1})}(t)}{A\Delta}, \\ S_n^{CMA}(t) &= \frac{\displaystyle\sum_{j=0}^{n-1}\left(\sum_{i=-\frac{A}{2}}^{\frac{A}{2}-1}[X_{t_{j-i+1}} - X_{t_{j-i}}]^2\right)\mathbf{1}_{[t_j,t_{j+1})}(t)}{A\Delta}, \end{aligned}$$

where A indicates the number of observations effectively used for estimating volatility in a single time point.

Consistency of S_n^{CMA} is clear as long as t is a continuity point and for those points for which $\sigma^2(t_-) = \lim_{x\nearrow t}\sigma^2(x)$. It is also shown in Bertrandt (1996) that, if $\sigma(\cdot)$ satisfies $\sigma(\cdot) = \sum_{i=1}^N \sigma_i 1_{[t_i,t_{i+1})}(\cdot)$, where σ_i $\forall i$ is some positive constant then, as long as $\Delta \to 0$, $A \to \infty$ and $A\Delta \to 0$,

$$IMSE_{A,\Delta} \to 0.$$

One additional practical concern is the optimal choice of the smoothing parameter h, for which we implemented crossvalidation.

3 Wavelets estimation

3.1 General ideas of wavelets functional estimation

For a thorough treatment of wavelets and their applications in statistics, see Vidakovic (1999). There are several ways of defining wavelets analysis in $L_2(\mathbb{R})$. The most constructive and easier to interpret is the so-called Multiresolution Analysis (MRA), due to Mallat (1989), as follows.

Definition 3.1. *A Multiresolution Analysis, $\{V_j\}_{j\in\mathbb{Z}}$, in $L_2(\mathbb{R})$ is an increasing sequence of closed subspaces of $L_2(\mathbb{R})$ which satisfies the following conditions:*

(i) $\cap_{j\in\mathbb{Z}} V_j = 0$ and $\cup_{j\in\mathbb{Z}} V_j$ is dense in $L_2(\mathbb{R})$.

(ii) $\forall g \in L_2(\mathbb{R})$ and $\forall j \in \mathbb{Z}$, $g(x) \in V_j \iff g(2x) \in V_{j+1}$.

(iii) $\forall g \in V_0$ and $\forall k \in \mathbb{Z}$, $g(x-k) \in V_0$.

(iv) $\exists g \in V_0$ such that $\{g(x-k)\}_{k\in\mathbb{Z}}$ is a Riesz basis of V_0.

An intuitive presentation for wavelets follows. Wavelets are a set of orthonormal basis functions in $L_2(\mathbb{R})$. The basis is generated from translation or dilation of two specific functions: a scaling function ϕ (or father wavelet) and the mother wavelet ψ. The father wavelet is computed as a solution of the so-called dilation equation and the mother wavelet is obtained from the father wavelets by filters.

Given the father and mother wavelets, one obtain functions ϕ_{jk} as $\phi_{jk}(x) = 2^{j/2}\phi(2^j x - k)$ and ψ_{jk} as $\psi_{jk}(x) = 2^{j/2}\psi(2^j x - k)$. One can derive from a MRA that for each $j \in \mathbb{Z}$, $\{\phi_{jk}, k \in \mathbb{Z}\}$ is an orthonormal basis for V_j, the *approximation* space of resolution j. A *detail* space of resolution j, call it W_j, can de defined by $V_{j+1} = V_j \oplus W_j$. A basis for W_j is easily drawn from the MRA as $\{\psi_{jk}\}, k \in \mathbb{Z}$. Moreover,

$$\cup_{j\in\mathbb{Z}} W_j = V_{j_0} \cup (\cup_{j\geq j_0} W_j) = lim_{j\to\infty} V_j,$$

where $j_0 \in \mathbb{Z}$. Therefore, $\{\psi_{jk}, j, k \in \mathbb{Z}\}$ and $\{\{\phi_{j_0 k}, k \in \mathbb{Z}\}, \{\psi_{jk}, j \geq j_0, k \in \mathbb{Z}\}\}$ are both orthonormal bases for $L_2(\mathbb{R})$. Finally, any square integrable function can be written as

$$f(t) = \sum_{j\in\mathbb{Z}} \sum_{k\in\mathbb{Z}} \beta_{j,k}\psi_{j,k}(t) = \sum_{k\in\mathbb{Z}} \alpha_{j_0,k}\phi_{j_0,k}(t) + \sum_{j\geq j_0} \sum_{k\in\mathbb{Z}} \beta_{j,k}\psi_{j,k}(t), \tag{3.1}$$

where equalities are meaningful in L_2 norm.

Desirable features of wavelets bases (from either the theoretical or applied point of view) include compact support and adaptable smoothness. A family that has both properties is Daubechies wavelets, that will be employed here. The bases are indexed by the number of null moments ($N \geq 1$) and we will call them DAUB# N. In general ($N \geq 2$) they have no closed form but fast filters are available and no practical problem rise from that fact.

Following (3.1), estimating f is equivalent to estimating its coefficients. Construction of estimation algorithms is specific to the nature of the sample $\{X_t\}$.

For instance, suppose the sample is regularly spaced (the fixed sampling interval called Δ) and that the sample size $n = 2^J$ for some $J \in \mathbb{N}$. There is a maximum number of coefficients that can be estimated and that means also that there is a maximum level of resolution that can be used. Therefore, the actual wavelet approximation for f is given by

$$f(t) \simeq \sum_{k \in K_{j_0}} \alpha_{j_0,k}\phi_{j_0,k}(t) + \sum_{j=j_0}^{j_1} \sum_{k \in K_j} \beta_{j,k}\psi_{j,k}(t), \tag{3.2}$$

where j_0 and j_1 are, respectively, the minimum and maximum empirical levels of resolution and K_j are finite sets whose union has cardinality n.

Defining the minimum level of resolution is easier and less determinant than choosing j_1. Since we use compactly supported wavelets, we can simply choose j_0 to be the largest resolution for which $\phi_{j_0,k}$ still covers the sampled interval. In our case, it means that the support of $\phi_{j_0,k}(t)$ covers $[0, T]$ and for DAUB# N, that is

$$j_0 = \lceil \log_2(2N - 1) \rceil.$$

The cascade algorithm (also known as pyramidal algorithm) can then be employed for fast computation of the coefficients $\hat{\alpha}_{j,k}$ and $\hat{\beta}_{j,k}$, whose values will provide an estimate of (3.2) based on the data. For details on the cascade algorithm and related references, see Vidakovic (1999). From now on, we will use hat to denote estimators or estimates.

Eventhough (3.2) is a nice approximation, specially for smooth f and ϕ, linear projection estimators tend to be asymptotically biased. For that reason, nonlinear procedures involving shrinkage are desirable. In practice shrinking the function in wavelets domain is equivalent to smoothing it in time domain. Shrinkage in wavelet analysis can be attained through level limitation and coefficients thresholding. Thresholding is usually divided in two major categories. Hard thresholding is given by a rule such as

$$\delta^d(\hat{\beta}_{j,k}, \lambda) = \hat{\beta}_{j,k}\mathbf{1}_{(|\hat{\beta}_{j,k}|>\lambda)}.$$

Soft thresholding has rules such as

$$\delta^s(\hat{\beta}_{j,k}, \lambda) = (\hat{\beta}_{j,k} - sign(\hat{\beta}_{j,k})\lambda)\mathbf{1}_{(|\hat{\beta}_{j,k}|>\lambda)},$$

for some $\lambda > 0$.

In either case, one will have a series of thresholded estimated coefficients such that

$$\hat{\beta}^{thr}_{j,k} = \delta^t(\hat{\beta}_{j,k}, \lambda),$$

where $t = d$ or $t = s$ if using respectively a hard or soft rule.

The art in thresholding is choosing the value (or values) of λ and proposing a rule which poses as little weight as possible on that choice. In other words, one wants an optimal λ but do not want to be strongly penalized by a suboptimal choice. For thresholding values, see Vidakovic (1999) and Donoho et al. (1996).

The so-called universal thresholding, motivated by additive normal error is given by $\sigma\sqrt{2\log n}$, where σ is the empirical coefficients standard deviation.

In Pinheiro and Vidakovic (1997), another thresholding procedure is proposed, based on energy retention via Lorentz curve. One knows that the total empirical energy is given by

$$\|f\|_{L_2}^2 = \sum_j \sum_k \beta_{j,k}^2$$

and it is easy to compute specific level by level energy.

Definition 3.2 (Lorentz Curve). *Let $\{\beta_{j,k} : j = 0, .., n \ k \in \mathbb{Z}\}$ be the coefficients of a signal f. The Lorentz curve based on their empirical level energies is:*

$$L(p) = \frac{\sum_{j=0}^{\lfloor np \rfloor} \sum_k \beta_{j,k}^2}{\sum_{j=0}^{n} \sum_k \beta_{j,k}^2} = \frac{\sum_{j=0}^{\lfloor np \rfloor} d_j^2}{\sum_{j=0}^{n} d_j^2}, \qquad p \in [0,1],$$

where $d_j^2 = \sum_k \beta_{j,k}^2$.

Theorem 3.1 (Meyer, 1992, p.51). *A function $f \in C^s(\mathbb{R})$ if and only if*

$$|\beta_{j,k}| \leq C 2^{-j/2} 2^{-j(r \wedge s)}$$

on a r-regular MRA.

Notice that Theorem 3.1 shows that wavelet analysis concentrate the energy of a signal on very few low levels. Therefore, a good decomposition should have very few *relevant* coefficients. One can then compute the empirical level energies, by $\hat{d}_j^2 = \sum_k \hat{\beta}_{j,k}^2$, and the following empirical version of Theorem 3.1 is obtained.

Theorem 3.2 (Pinheiro and Vidakovic, 1997). *Let $\hat{\beta}_{j,k}$ the estimated coefficients for a sample of size n, $B_N = \min_{x \in \mathbb{R}} \max_k \psi_{0,k}^2(x) > 0$. Then,*

$$\hat{d}_j^2 \geq \frac{B_N 2^j}{C}.$$

At first, Theorems 3.1 and 3.2 look contradictory. However, the former claims that wavelet decomposition is very parsimonious and shows that signal energies are concentrated in a very few low levels of resolution. On the other hand, Theorem 3.2 shows that empirical coefficients are actually arbitrarily large and, worst, coefficients estimates from high levels tend to be even larger. The correct interpretation of that apparent contradiction is that estimated coefficients must be conformed to Theorem 3.1. Based on this shortcoming of linear wavelet estimators, Pinheiro and Vidakovic (1997) proposed the following thresholding procedure.

Take the empirical Lorentz curve, i.e.,

$$\hat{L}(p) = \frac{\sum_{j=0}^{\lfloor np \rfloor} \sum_{k} \hat{\beta}_{j,k}^2}{\sum_{j=0}^{n} \sum_{k} \hat{\beta}_{j,k}^2} = \frac{\sum_{j=0}^{\lfloor np \rfloor} \hat{d}_j^2}{\sum_{j=0}^{n} \hat{d}_j^2}, \qquad p \in [0,1].$$

One can then adjust the maximum level of resolution for (3.2) to follow the behavior in Theorem 3.1. That means that levels are acceptable if they show a geometrical decreasing. Theorem 3.2 tell us that eventually one will have the empirical Lorentz curve showing an exponential increase. The first procedure is therefore to limit use to levels whose behavior conforms to Theorem 3.1. After choosing j_1, one applies a coefficients thresholding as follows.

Definition 3.3 (Thresholding 1 by Lorentz Curve). *Suppose a sequence of estimated coefficients $\{\beta_{j,k} : j = 0, .., n \; k \in \mathbb{Z}\}$ and a positive parameter κ. The Lorentz curve thresholded coefficients (LC1) is given by:*

$$\hat{\beta}_{j,k}^{thr} = \delta^d(\hat{\beta}_{j,k}, \sqrt{\kappa \overline{d}^2}) = \hat{\beta}_{j,k} \mathbf{1}_{(\hat{\beta}_{j,k}^2 > \kappa \overline{d}^2)},$$

where $\overline{d}^2 = n^{-1} \sum_{j=0}^{n} \sum_{k} \hat{\beta}_{j,k}^2$.

The cut-off parameter κ in Definition 3.3 has a very simple interpretation and its choice has very different conceptual interpretations. It balances two conflicting measures: parsimonious and energy retention. If $\kappa \leq 1$, one tends to preserve energy retaining more coefficients whilst, for $\kappa > 1$, preference is given to fewer coefficients over energy preservation.

This procedure does not take into account the level of resolution from which a coefficient is taken. In this work, we present a slightly modified Lorentz curve thresholding procedure as well as a hands-free maximum level choice. Coefficients thresholding is performed level by level as follows.

Definition 3.4 (Thresholding 2 by Lorentz Curve). *Let κ and κ_j be positive real numbers and $\{\beta_{j,k} : j = 0, .., n \; k \in \mathbb{Z}\}$ be the sequence of estimated coefficients. The LC2 thresholded coefficients are given by:*

$$\hat{\beta}_{j,k}^{thr} = \delta^d(\hat{\beta}_{j,k}, \sqrt{\kappa \overline{d}^2 + \kappa_j \overline{d}_j^2}) = \hat{\beta}_{j,k} \mathbf{1}_{(\hat{\beta}_{j,k}^2 > \kappa \overline{d}^2 + \kappa_j \overline{d}_j^2)},$$

where $\overline{d}_j^2 = n^{-1} \sum_{k} \hat{\beta}_{j,k}^2$.

Notice that

$$\kappa \overline{d}^2 + \kappa_j \overline{d}_{j'}^2 = \left(\frac{n_{j'}}{n} \kappa + \kappa_j \right) \overline{d}_{j'}^2 + \kappa \overline{d}_{-j'}^2, \tag{3.3}$$

where n_j is the number of coefficients in level W_j, n is the total number of coefficients, and $\overline{d}^2_{-j'} = \sum_{j=0;j\neq j'}^{n} \sum_k \hat{\beta}^2_{j,k}/(n-1)$.

From (3.3), one sees that different thresholding levels for each resolution give the procedure more leverage to weight in or out coefficients from specific levels. That is very important because procedures can be tailored depending on the regularity supposed or the *piece* of the function one intends to represent. Moreover, direct comparisons between level energy and its coefficients can be done. The wavelet LC2 thresholded is written as:

$$\hat{f}(t) = \sum_{k\in\mathbb{Z}} \hat{\alpha}_{j_0,k}\phi_{j_0,k}(t) + \sum_{j=j_0}^{j_1} \sum_{k\in\mathbb{Z}} \hat{\beta}^{thr}_{j,k}\psi_{j,k}(t).$$

3.2 Wavelet volatility estimators

Consider a sample of size $N = 2^n$, with $t_i = i2^{-n}$, where $i = 0, 1, ..., N$, i.e., one observes the diffusion in $[0, T]$ with sampling interval $\triangle_n = 2^{-n}T$. Some theoretical asymptotic results are known for a linear wavelet estimator of $\sigma(\cdot)$ under conditions (3.4), on $b(\cdot,\cdot)$ and $\sigma(\cdot)$, as seen in Genon-Catalot et al. (1992).

$$\left.\begin{array}{ll} \textbf{(i)} & b(\cdot,\cdot) \in C^1([0,+\infty)\times\mathbb{R}); \\ \textbf{(ii)} & \forall T,\ \exists K_T \text{ s.t. } \forall t\ \epsilon\ [0,T],\ \mid b(t,u) \mid \leq K_T(1+\mid u \mid); \\ \textbf{(iii)} & \sigma(\cdot) \in C^m([0,\infty)), \text{ with } m \geq 1, \text{ and } \sigma(t) > 0\ \forall t \geq 0. \end{array}\right\} \tag{3.4}$$

Under (3.4), and $m \geq 0$, solution of (2.2) is unique in $[0, \infty)$. As we use a regular MRA in $L_2(\mathbb{R})$ and $\sigma^2(\cdot)$ is not necessarily an element of that space, we can replace it with a new function $\bar{\sigma}^2(\cdot) \in L_2(\mathbb{R})$, such that:

$$\bar{\sigma}(t) = \sigma(t), \qquad \forall t\epsilon[0,T].$$

Suppose also that $\bar{\sigma}(t) \in C^m(\mathbb{R})$ and that $supp\{\bar{\sigma}\} \subset [-\epsilon, T+\epsilon]$, for some $\epsilon \geq 0$. We define a new diffusion process $\bar{X}_t$ as:

$$d\bar{X}_t = b(t,\bar{X}_t)dt + \bar{\sigma}(t)dW_t, \qquad \bar{X}_0 = x_0$$

that would for all practical purposes be identical to X_t.

Therefore, from now on, we will consider $\sigma^2(\cdot)$ to be an element of $L_2(\mathbb{R})$ and we can project it in wavelets approximation and detail spaces. An empirical estimate of approximation coefficient $\alpha_{j(n),k}$ is given by

$$\hat{\alpha}_{j(n),k} = \sum_{i=0}^{N-1} \phi_{j(n),k}(t_i)(X_{t_{i+1}} - X_{t_i})^2, \tag{3.5}$$

and the wavelet linear estimator of $\sigma^2(t)$ on level j_1 is given by

$$\hat{\sigma}^2(t) = \sum_{k\epsilon\mathbb{Z}} \hat{\alpha}_{j(n),k}\phi_{j(n),k}(t). \tag{3.6}$$

The estimator (3.6) is proposed by Genon-Catalot et al. (1992) where the following Integrated Mean Error (IME) and Integrated Mean Squared Error (IMSE) properties are presented:

$$2^{\frac{n}{2}} \int_0^T h(t) \left(\hat{\sigma}^2(t) - \sigma^2(t)\right) dt \overset{n\to\infty}{\to} \mathcal{N}\left(0, 2\int_0^T h^2(t)\sigma^4(t)dt\right) \tag{3.7}$$

$$\begin{aligned} R_n &= E\left(\int_0^T (\hat{\sigma}^2(t) - \sigma^2(t))^2 \gamma(t)dt\right) \leq C(2^{4j(n)-2n} + 2^{-2j(n)(m\wedge r)} + 2^{-n}) + \\ &\sup_{t\epsilon\mathbb{R}} \gamma(t) \left(2^{j(n)-n} 2 \int_0^T \sigma^4(t)dt + o(2^{j(n)-n})\right), \end{aligned} \tag{3.8}$$

where C is a constant depending only on ϕ, γ and σ^2 and h is any test function.

Theorems 2.1 shows how well kernel-based estimators can perform for dependent data sampled from a continuous time process. Theorem 2.1 deals specifically with the volatility function. This specification, however, comes with a cost: the volatility function must be very well-behaved. On the other hand, (3.7) and (3.8) investigate asymptotics for linear wavelet estimators. One can see, from (3.7) and Theorem 2.1 that kernel convergence is attained with a $\sqrt{nh}$, where n is the sample size and h is the smoothing parameter, whilst, for wavelets, the convergence rate is $\sqrt{n}$ for the IMSE. In theory, this would mean that kernel results would be stronger, because they can be taken pointwise while the wavelets convergence would be true only when the volatility function is smeared by a test function $h(\cdot)$. In practice, one can take $h(\cdot)$ as narrowly supported as one wishes, as long as $h(\cdot)$ is still a test function. The point to be made is that both techniques would ensure a rate of convergence which is a product of $\sqrt{n}$ and a smoothing factor ($1/\sqrt{h}$ for the kernel and $\|h(\cdot)\|_2$ for the wavelets). Therefore, the problem of comparison of these paradigms would be concerned with the constants, which are quite complex to be dealt with analytically.

As for iid observations, the linear wavelet estimator helps us as guidelines but some kind of thresholding should be performed. We propose in Definition 3.5 the following nonlinear estimator.

Definition 3.5 (Nonlinear Wavelet Estimator). *The nonlinear wavelet estimator for the volatility function σ^2 is given by*

$$\hat{\sigma}^2(t) = \sum_{k\epsilon\mathbb{Z}} \hat{\alpha}_{j_0,k}\phi_{j_0,k}(t) + \sum_{j=j_0}^{j_1} \sum_{k\in\mathbb{Z}} \hat{\beta}_{j,k}^{thr} \psi_{j,k}(t),$$

where (being N the sample size)

- *j_0 and j_1 are respectively the minimum and maximum resolutions allowed by the data, i.e., $j_0 = \lceil \log_2(2N-1) \rceil$ and j_1 is chosen accordingly to Theorems 3.1 and 3.2, i.e., the first three levels are always maintained and j_1 is the smallest $j \geq 3$ for which the ratio between the $(j+1)$-th and the j-th levels energies is larger than some pre-determined number (in our case, 1 or 3);*

- $\hat{\alpha}_{j_0,k}$ *are the empirical approximation coefficients, given by (3.5);and*
- *the estimated wavelet coefficients* $\hat{\beta}^{thr}_{j,k}$ *are the LC2 thresholded estimates for the detail coefficients, i.e., given by the rule of Definition 3.4 applied on* $\hat{\beta}_{j,k} = \sum_{i=0}^{N-1} \phi_{j,k}(t_i)(X_{t_{i+1}} - X_{t_i})^2$.

4 Applications

We study the non-parametric estimators on a real series and on simulations. In the simulations, we used the wavelet estimators and three kernel estimators: Gaussian, MA and CMA. Crossvalidations were performed only for MA and CMA, due to computational costs. Gaussian smoothing parameter was chosen in order to minimize the Integrated Squared Error (ISE), as seen in Härdle (1990). The ISE is the IMSE without taking the expectation. Another comment on the Gaussian kernel is that, although it violates some of the conditions on asymptotic results for dependent data (see Bosq, 1998, Bertrand, 1996 and Avesani and Bertrand, 1997), its known good properties motivate its use and simulation results show that its behavior is quite excellent.

Wavelets estimators used three different bases: DAUB#2, DAUB#5 and DAUB#8, and LC2 thresholding procedures. For the automatic definition of maximum levels, we used two different ratio criteria. Three levels were always maintained and, after the third, as long as the i-th level empirical energy would be at most the same (or at most three times as much) as the $(i-1)$-th level's, the i-th level would stay. For the coefficients thresholding we used four different combinations of parameters $\{\kappa_j\}_{j=1,2,3}$ and κ, described in Table 1. For levels higher than 3, $\kappa_j = \kappa_3$.

Table 1 *Thresholding parameters.*

Combinations	κ_1	κ_2	κ_3	κ
Comb 1	2	2	2	2
Comb 2	1	2	3	1
Comb 3	1	3	5	3
Comb 4	.5	0.75	1	1

4.1 Simulation study

We use in this simulation study the following model

$$dX_t = b(X_t)dt + \theta(t)dWt \qquad X_0 = 0 \qquad t\epsilon[0,T],$$

where $b(X_t) = 0.75 - X_t$ and $\theta(\cdot)$ is one of the eight different types, illustrated in Figures 1 and 2, and with no loss of generality, $T = 1$.

As nonparametric estimators tend to differ mainly on jumps, we used two almost norm-equivalent groups of volatility functions, one with smooth and the other with their respective *irregular* versions. Figures 1 and 2 show respectively

the volatility functions with jumps and the *regular* volatility functions. The eight models were generated in 4096 time points with 1000 simulated data series each.

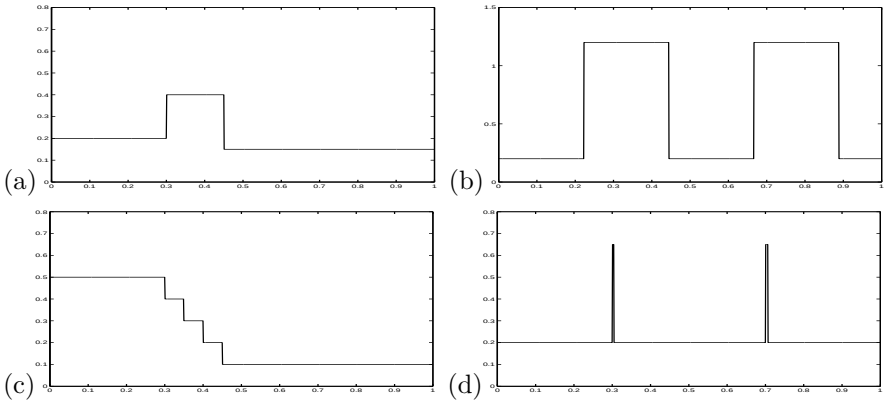

Figure 1 *$\theta(\cdot)$ Functions with jumps: (a) Vol1 (b) Vol2 (c) Vol3 (d) Vol4.*

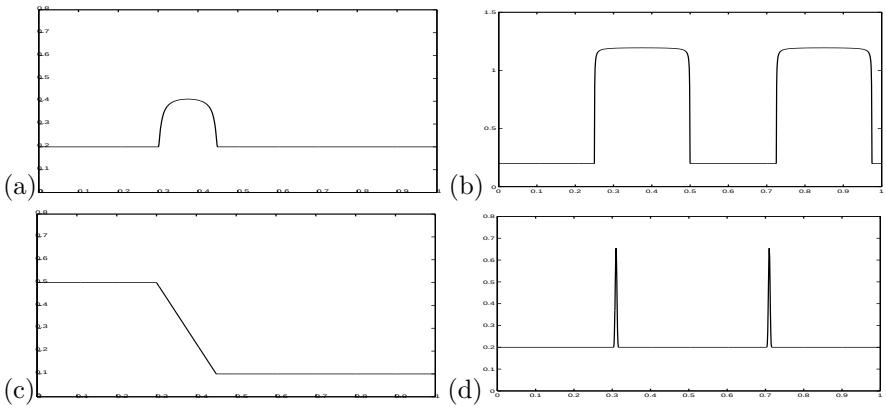

Figure 2 *Regular $\theta(\cdot)$ Functions: (a) Vol5 (b) Vol6 (c) Vol7 (d) Vol8.*

It is important to observe that Vol1-Vol4 do not satisfy some of the regularity conditions required for either the kernel or linear wavelet estimators theoretical results. All the volatility functions, except Vol4 and Vol8, are related to structural changes while Vol4 is clearly an outlier and Vol8 can be considered as two groups of outliers or as two short structural change periods. One can argue that an empirical volatility function will not behave as a pure Vol4 or Vol8 type, but our goal is to test the estimators in limit cases. Although that can be of some concern, the feature in which one will usually be interested is still the recovery (quantitatively and qualitatively) of the highest peaks of volatility. However, in every case we would like to have the estimation error as small as possible. A small error will allow the researcher to have a clear idea of the volatility function and take the

right conclusion. For error comparisons we used the discretized Integrated Squared Error and Integrated Absolute Error (ISE and IAE, respectively) measures. The ISE is defined as:

$$ISE = n^{-1} \sum_{i=1}^{n} w(i/n) \left(\hat{\theta}(i/n) - \theta(i/n) \right)^2,$$

where $w(.)$ ia a nonnegative weight function. The IAE is defined similarly taking the absolute values instead of the squared ones. Those measures are very global and hinder finer comparisons of estimators performance exactly at the points (or regions) of most interest such as jumps, cusps or high curvatures. Four weight functions were employed in order to emphasize the error in different regions:

- *Total*, that put weight equal to one everywhere.
- *Q100*, that puts uniform weight on change points and on its 200 nearest neighbors and zero on the rest
- *Q50*, that puts uniform weight on change points and on its 100 nearest neighbors and zero on the rest
- *Q0*, that puts uniform weight on change points and zero on the rest.

The *Total* statistics test the overall performance of the estimators, while the other statistics measure the performance of the estimators near the change points. For instance, Q50 measure how fast the estimators can detect the change without spreading the influence to neighbor observations. Results are summarized in tables and plots, using means and medians as descriptive tools. In Table 2, the smallest Integrated Errors (ISE's) are highlighted.

Table 2 presents the mean and median of the discrete square error for each model. Smoothed histograms for each estimator empirical distribution are presented in Figures 3-4 for Q0 and Q100 ISE's. Table 3 shows the computational time quartiles for each estimation type, based on the empirical distribution of 8000 series (8 types times 1000 replications).

The overall picture of the results presented is that Gaussian kernel shows the best performance in terms of ISE for all volatility types, but Vol4 and Vol8, the functions which present local perturbations. In these cases, wavelets produce the best performance, with ISE's almost four times smaller than the Gaussian ISE's. Wavelet estimator came a closely second to Gaussian kernel for the other six cases.

The wavelet estimator clearly outperformed the MA and CMA techniques, with larger differences in the volatility functions with jumps (Vol4) or high curvatures (Vol8). The wavelet ISE's are in general a quarter to a third of the MA and CMA ISE's. This show that this estimator is more robust to this type of jumps.

Among wavelet estimators, two points should be made. DAUB#2 shows general superior performance, which means that, since some of the functions are not smooth, an artificial basis smoothness will do more harm than good. Ratio 1 and thresholding combination 4 show slight superiority, mostly by ratio than by thresholding rule. That means that an automatic thresholding rule can be used

without performance concerns but for level adequacy, it is better to cut early (ratio 1) than to cut late (ratio 3). Notwithstanding, the wavelet procedure is quite robust to thresholding combinations, in the sense that their relative position to MA, CMA and kernel will be usually the same.

Table 2 *Integrated squared errors medians and means - Vol1-Vol8 ($x10^{-4}$).*

Vol	Error	Wavelets		Gaussian		CMA		MA	
Type	Measure	Median	Mean	Median	Mean	Median	Mean	Median	Mean
	Q0	1.23	1.68	**0.83**	**0.88**	2.30	2.41	2.38	2.49
Vol1	Q50	1.89	2.27	**1.31**	**1.37**	3.00	3.21	3.16	3.35
	Q100	2.11	2.41	**1.35**	**1.41**	3.05	3.38	3.21	3.50
	Total	2.70	2.95	**1.47**	**1.52**	3.82	4.02	3.84	4.07
	Q0	196.78	222.52	**164.10**	**166.66**	516.11	520.31	463.75	471.48
Vol2	Q50	310.84	335.47	**290.63**	**293.29**	895.24	879.54	697.04	704.16
	Q100	326.51	350.68	**321.76**	**324.72**	1045.47	1006.54	722.20	727.78
	Total	339.17	364.03	**323.01**	**326.21**	1083.82	1088.03	1015.96	1019.81
	Q0	0.74	0.87	**0.52**	**0.59**	1.00	1.03	0.98	1.06
Vol3	Q50	1.15	1.26	**0.71**	**0.76**	1.57	1.57	1.53	1.64
	Q100	1.40	1.53	**0.78**	**0.85**	1.84	1.89	1.79	1.89
	Total	5.41	7.26	**2.73**	**2.90**	3.55	3.53	3.55	3.53
	Q0	**2.97**	**3.01**	13.24	13.19	14.18	13.93	14.00	14.17
Vol4	Q50	**3.32**	**3.44**	13.29	13.27	14.49	14.31	14.28	14.59
	Q100	**3.39**	**3.50**	13.31	13.29	14.54	14.37	14.34	14.65
	Total	**4.18**	**4.38**	13.47	13.45	15.06	14.80	14.98	15.00
	Q0	0.85	1.15	**0.43**	**0.47**	1.68	1.79	1.73	1.83
Vol5	Q50	1.00	1.26	**0.49**	**0.54**	1.80	1.98	1.86	2.01
	Q100	1.06	1.32	**0.50**	**0.55**	1.82	2.04	1.89	2.07
	Total	1.43	1.68	**0.68**	**0.73**	1.89	2.19	1.91	2.19
	Q0	333.30	349.77	**169.16**	**174.51**	502.78	510.80	527.21	535.13
Vol6	Q50	378.81	398.87	**244.79**	**249.99**	585.97	593.32	629.73	636.53
	Q100	379.45	399.92	**249.63**	**254.87**	586.19	593.55	629.93	636.75
	Total	381.86	402.28	**249.79**	**255.02**	950.11	959.00	950.13	956.87
	Q0	0.29	0.51	**0.22**	**0.28**	0.93	1.07	0.94	1.05
Vol7	Q50	0.36	0.68	**0.30**	**0.37**	1.20	1.34	1.17	1.32
	Q100	0.47	0.84	**0.37**	**0.47**	1.48	1.62	1.44	1.61
	Total	4.14	5.70	**2.30**	**2.50**	3.57	3.55	3.57	3.55
	Q0	**2.48**	**2.72**	6.81	6.85	8.26	8.27	8.66	8.89
Vol8	Q50	**2.65**	**2.90**	7.18	7.23	8.80	8.87	9.30	9.68
	Q100	**2.71**	**2.95**	7.22	7.27	8.99	9.01	9.51	9.83
	Total	**3.71**	**3.95**	7.41	7.45	9.63	9.94	10.59	10.70

The fact that wavelets outperforms CMA and MA contradicts the interpretation in Bertrand (1996). There, Haar based procedures were outperformed by CMA and MA, and CMA was found to be uniformly the best procedure. We

should point out that we are using more regular wavelets and drawing conclusions from a broader set of volatility functions. Moreover, although the results from the IAE's are not presented, they are qualitatively the same as those analyzed from the ISE's.

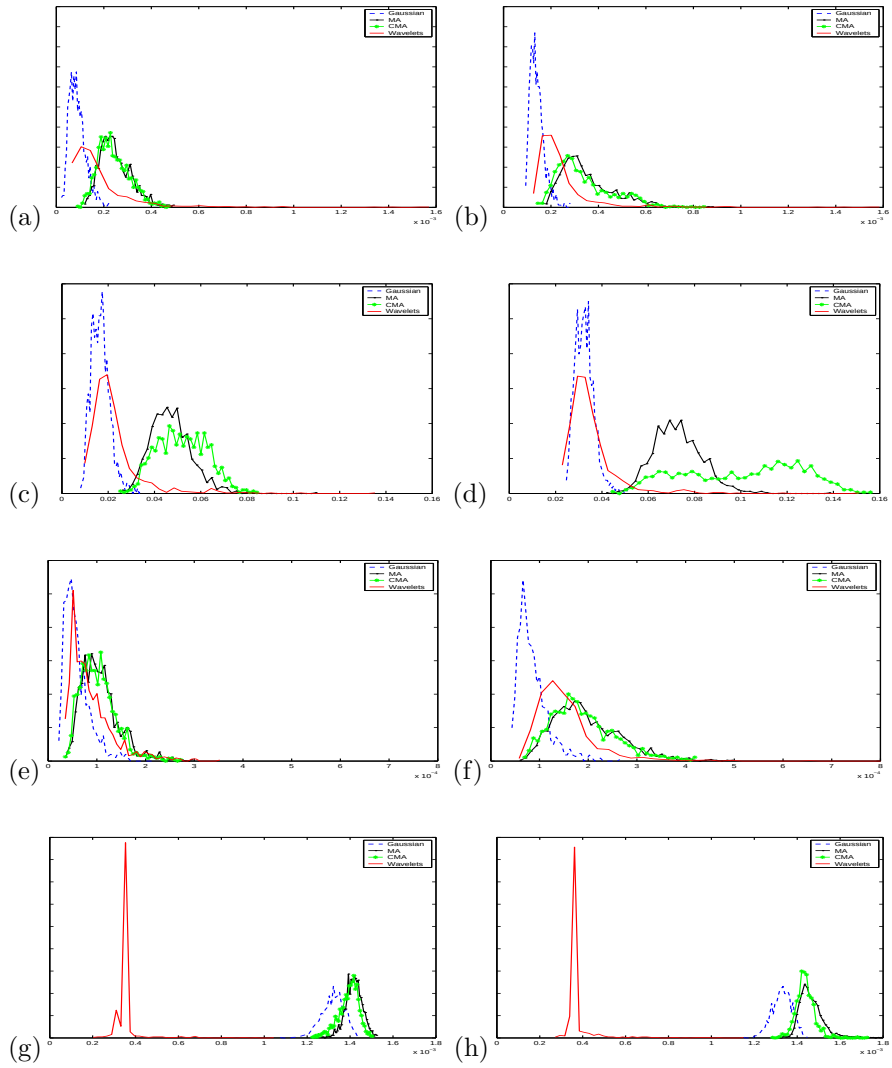

Figure 3 *Smoothed Histogram for the Weighted Integrated Squared Error (ISE) - (a) Vol1 Q0 (b) Vol1 Q100 (c) Vol2 Q0 (d) Vol2 Q100 (e) Vol3 Q0 (f) Vol3 Q100 (g) Vol4 Q0 (h) Vol4 Q100.*

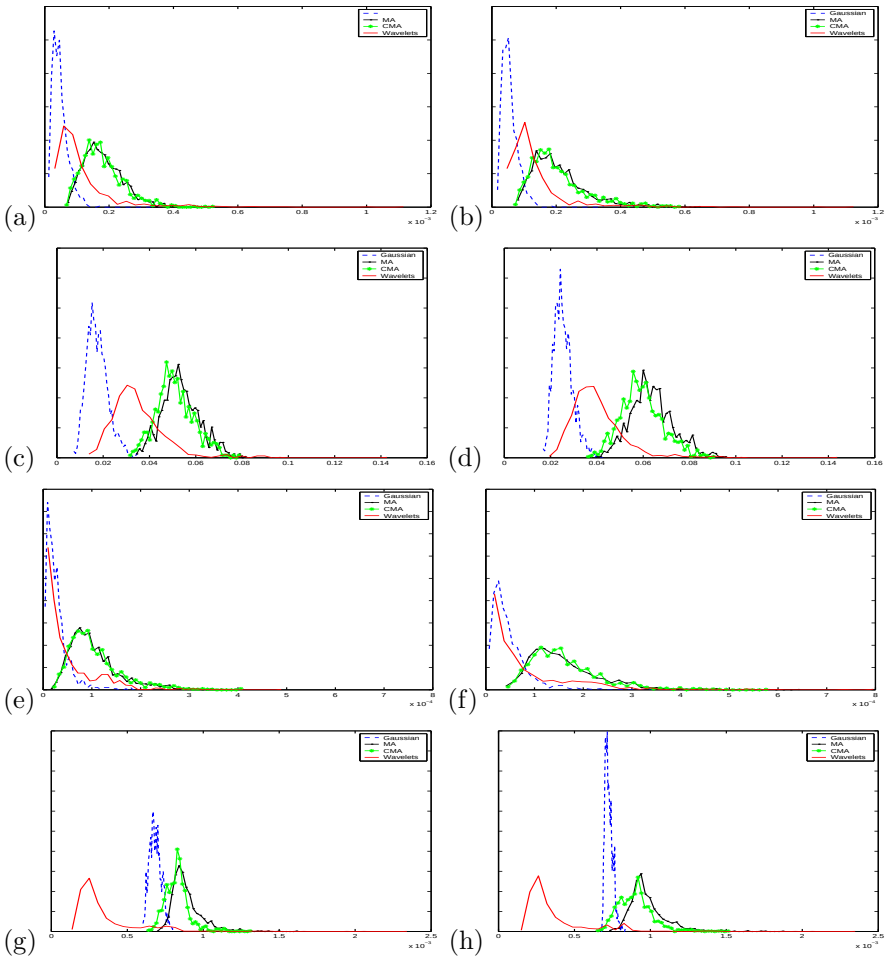

Figure 4 *Smoothed Histogram for the Integrated Squared Errors (ISE) - (a) Vol5 Q0 (b) Vol5 Q100 (c) Vol6 Q0 (d) Vol6 Q100 (e) Vol7 Q0 (f) Vol7 Q100 (g) Vol8 Q0 (h) Vol8 Q100*

We should write a few lines concerning computational performance and feasibility. By the numbers in Table 3, wavelets are clearly superior to any of the other methods. Some numbers are that wavelets procedures are usually 100 times faster than either MA or CMA and even faster when compared to Gaussian kernel. That difference can mean the feasibility of online computation at ease. Moreover we should point that faster and more powerful computers were used for Gaussian kernel estimators, because of time and memory issues. Considering that wavelet estimators came closely second to Gaussian kernel estimators for six cases, outperform it by almost four times in the two jump-type volatility functions, and its computational outstanding performance, one could argue for the nonlinear wavelet

estimator to be the most appropriate technique among the tested ones.

Table 3 *Distribution of Computer Time for the Simulated Series (in Seconds).*

Estimator	Q_1	Median	Q_3
Wavelets	0.2	0.2	0.2
MA	20.4	23.3	25.1
CMA	21.6	26.5	28.3
Gaussian	32.7	36.7	38.9

4.2 Real series

In this section we estimate the volatility of BOVESPA index (São Paulo's stock exchange market), minute by minute, from 11am on December 18 2001 to 6pm on February 18 2002[1]. One can see the series in Figure 5.

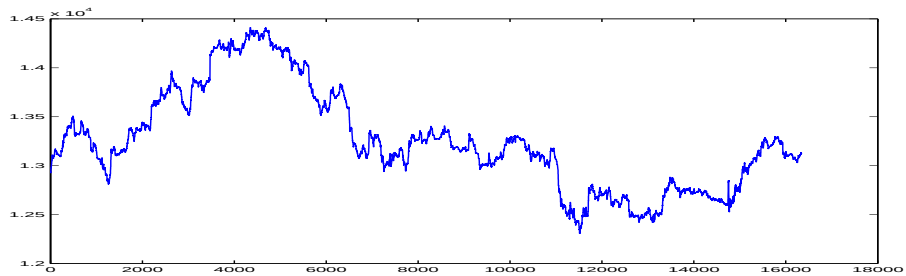

Figure 5 *BOVESPA index - 11am on December 18 2001 to 6pm on February 18 2002*

The aim is to compare nonparametric estimates based on the returns. However, we should point out that we were unable to compute normal kernel estimates due to computational memory issues. Moreover, MA estimators will not be considered because they are too slow for sudden changes in volatility, as seen in Avesani and Bertrand (1997). Therefore, CMA and three wavelet bases are considered: DAUB#1, DAUB#5 and DAUB#8.

One can notice that this series has a large isolated value around the 3200-th observation which could be classified as Vol4 type (see Figures 1 and 2). However, this classification could change to Vol8 type when the sampling frequency is increased. In either case on the simulation studies, nonlinear wavelet estimators were quite superior to the other methods and similar results for peak reconstruction should be expected. There is also a clear structural change with increase in the level around the 14600-th observation and other smaller Vol8 type outliers.

[1]This series has 16348 observations for which some minor adjustments are made in order to use cascade algorithms (since $2^{14} = 16384$)

The estimates of the volatility function is presented in Figure 6, which illustrates clearly that nonlinear wavelet estimators will recover very well the most relevant peaks while CMA will perform very poorly, for single peaks and even for wider plateaus. Recovery from wavelets will be possible for a wide variety of wavelet bases, as illustrated with DAUB#1, DAUB#5 and DAUB#8, on Figure 6 (c), (d) and (e) respectively. The quick recover of the volatility estimation emphasizing outliers and structural change is very important in volatility analysis as commented previously.

Finally, besides recovering peaks much better than the CMA estimator, wavelets can also do it with much less noise; for that, look at the baseline noise in Figure 6(b) and compare it to the baseline noise of the remaining plots in Figure 6. This is expected because of the limiting maximum level procedure combined to the thresholding of statistically meaningless coefficients.

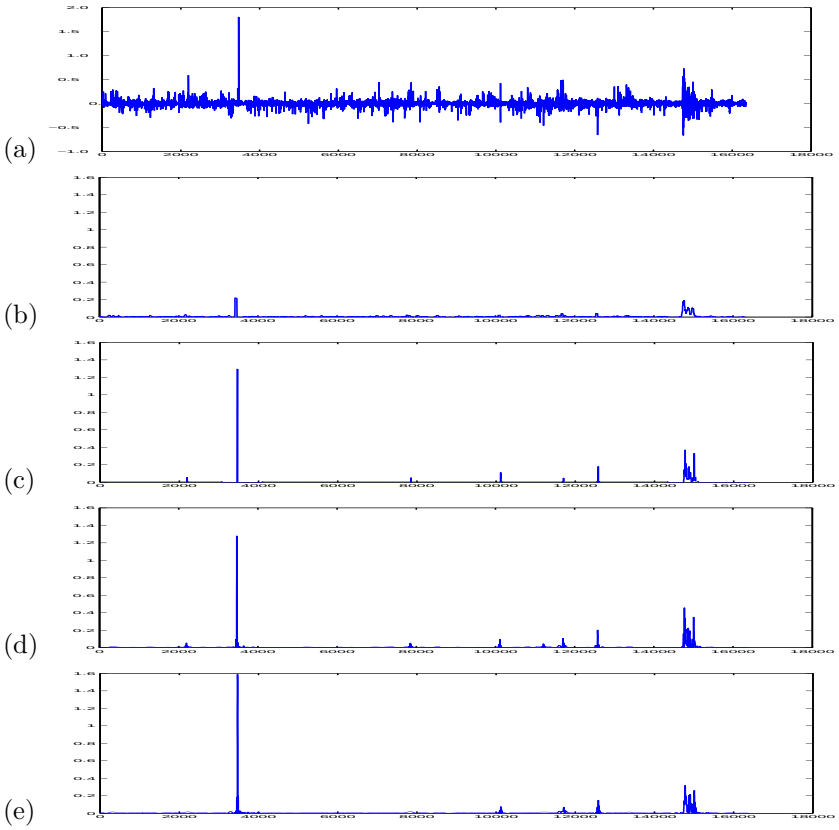

Figure 6 *(a)BOVESPA returns - 11am on December 18 2001 to 6pm on February 18 2002 ($\times 10^{-2}$) BOVESPA Volatility Estimates ($\times 10^{-2}$): (b) CMA (c)DAUB# 1 (d) DAUB# 5 (e) DAUB# 8*

5 Final remarks

From the results shown on Sections 4.1 and 4.2, one concludes that wavelet estimates are more efficient for estimating volatility functions, whenever they present peaks, jumps or cusps. This efficiency is easily seen with local error measures as well as by visual comparisons. For more *regular* volatility functions, Gaussian kernel procedures have a better performance. This Gaussian kernel performance is much better than MA and CMA but, although uniformly, only slightly better than wavelets estimators. It is important to reinforce the fact that no asymptotic results for Gaussian kernel estimators are known as well as for some of the most challenging volatilities used in the simulations. Finally, on the performance issues, although MA and CMA are theoretically consistent they do not perform up to wavelets in any situation or Gaussian kernel in most situations (all but one). Our simulation studies also show a very negative performance for the MA estimator.

We can summarize our conclusions on the following considerations

- Wavelet estimators are exceptionally faster than any other one of the three; moreover, it needs much less memory resources. Those two properties make them very suitable for online operations.

- Gaussian kernel estimators, eventhough having no theoretically proved performance, present results that shows their ability in dealing with volatility functions with some degree of regularity.

- Wavelet estimators Lorentz curve procedures can be automatized and generalized with no major loss in performance which confirm their versatility and robustness to misspecification.

- Wavelet estimators have a clear capacity of representing irregular function with parsimony as illustrated in Vol4 and Vol8 and the real data set.

- Differences in wavelet and Gaussian kernel performances are small in regular functions and favorable to kernel estimator; they are larger in irregular functions and favorable to wavelets.

Those considerations gives us the understanding that theoretical incursions should be made into the asymptotic performance of Gaussian kernel and wavelet estimators as well as most nonparametric estimators for less regular (and more complex) volatility functions and, from an applied point of view, wavelets should be seriously considered given its overall good performance (with excellency in most cases) and its computational convenience.

A recent work on non-parametric volatility estimation is given by Mercurio and Spokoiny (2004). A model is constructed in a change-point setup and some powerful results for local power fitting volatility estimator are shown. We believe that performance comparisons between these estimators and the proposed nonlinear wavelet estimator should be considered for future research.

Acknowledgements

The authors were partially supported by CAPES, CNPq and FAPESP. We thank *Finance Lab*, of *Ibmec Business School*, Brazil, for graciously providing the data to us. We would also like to thank the anonymous reviewers for their suggestions that truly helped to improve the paper.

(Received August, 2004. Accepted June, 2006.)

References

Abberger, K. (1997). Quantile smoothing in financial time series. *Statistical Papers*, **38**, 125-148.

Antoniadis, A. and Oppenheim, G., editors (1995). *Wavelets and Statistics.* Papers from the conference held in Villard de Lans, November 16–18, 1994. New York: Springer-Verlag.

Avesani, R. G. and Bertrand, P. (1997). Does volatility jump or just diffuse? A statistical approach. In *Numerical Methods in Finance*, **270–289**. Cambridge: Cambridge University Press.

Bertrand, P. (1996). Estimation of the stochastic volatility of a diffusion process i. comparison of haar basis estimator and kernel estimators. *Technical Report - Inria Sophia Antipolis.*

Bollerslev, T., Chou, R. Y. and Kroner, K. K. (1992). Arch modelling in finance: a review of the theory and empirical evidence. *Journal of Econometrics*, **52**, 5–59.

Bosq, D. (1998). *Nonparametric Statistics for Stochastic Processes: Estimation and Prediction.* Lecture Notes in Statistics. New York: Springer-Verlag.

Darolles, S. and Gouriéroux, C. (2001). Truncated dynamics and estimation of diffusion equations. *Journal of Econometrics*, **102**, 1–22.

Donoho, D. and Johnstone, I. (1995). Adapting to unknown smoothness via wavelet shrinkage. *Journal of the American Statistical Association*, **90**, 1200–1224.

Donoho, D., Johnstone, I., Kerkyacharian, G. and Picard, D. (1996). Density estimation by wavelet thresholding. *Annals of Statistics*, **24**, 508–539.

Florens-Zmirou, D. (1993). On estimating the diffusion coefficient from discrete observations. *Journal of Applied Probability*, **30**, 790–804.

Genon-Catalot, V., Laredo, C. and Picard, D. (1992). Nonparametric estimation of the diffusion coefficient by wavelet methods. *Scandinavian Journal of Statistics*, **19**, 317–335.

Ghysels, E., Harvey, C. and Renault, E. (1996). Stochastic volatility. In *Statistical Methods in Finance*, **14**, 119-191, of *Handbook of Statistics*. Amsterdam: North-Holland.

Härdle, W. (1990). *Applied Non-parametric Regression*. Cambridge: Cambridge University Press.

Härdle, W. and Tsybakov, A. (1997). Local polynomial estimators of the volatility function in nonparametric autoregression. *Journal of Econometrics*, **81**, 223–242.

Hoffmann, M. (1999). On nonparametric estimation in nonlinear ar(1)-models. *Statistics and Probability Letters*, **44**, 29–45.

Mallat, S. (1989). A theory for multi-resolution signal decomposition: the wavelet representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **11**, 674–693.

Mercurio, D. and Spokoiny, V. (2004). Statistical inference for time-inhomogeneous volatility models. *Annals of Statistics*, **32**, 577–602.

Meyer, Y. (1992). *Wavelets and Operators*. Cambridge: Cambridge University Press.

Pinheiro, A. and Vidakovic, B. (1997). Estimating the square root of a density via compactly supported wavelets. *Computational Statistics and Data Analysis*, **25**, 399–415.

Shephard, N. (1996). Statistical aspects of arch and stochastic volatility. In *Time Series Models in Econometrics, Finance and Other Fields*, 1–67. London: Chapman and Hall.

Van Es, B., Spreij, P. and van Zanten, H. (2005). Nonparametric volatility estimation for discrete time series models. *Journal of Nonparametric Statistics*, **17**, 237–251.

Vidakovic, B. (1999). *Statistical Modeling by Wavelets*. New York: John Wiley and Sons.

Walter, G. (1992a). Approximating of the delta function by wavelets. *Journal of Approximation Theory*, **71**, 329–343.

Walter, G. (1992b). Wavelets and generalized functions. In C. K. Chui, editor, *Wavelets - A Tutorial in Theory and Applications*, 51–70. London: Academic Press.

Yang, L. (2000). Finite nonparametric garch model for foreign exchange volatility. In C-P Han and W. W. S. Wei, editors, *Selected Papers from the 1999 Applied Statistical Symposium of the International Chinese Statistical Association*, 1347–1365. Monticello, New York: Marcel Dekker.

Yang, L., Härdle, W. and Nielsen, J. P. (1999). Nonparametric autoregression with multiplicative volatility and additive mean. *Journal of Time Series Analysis*, **20**, 579–604.

Aluísio Pinheiro, **Neale A. El-Dash** and **Luiz K. Hotta**
Departamento de Estatística
Universidade Estadual de Campinas
Caixa Postal 6065
13083-970 Campinas/SP, Brazil
E-mails: pinheiro@ime.unicamp.br, neale@ime.usp.br and hotta@ime.unicamp.br