# Non-parametric Volatility Estimation in Continuous Time*

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

Address:
Aluísio de Souza Pinheiro
Universidade Estadual de Campinas
Cidade Universitária Zeferino Vaz
Campinas SP 13083-970
Cx. Postal: 6065
BRAZIL

*2000 AMSC 62G08, 62G20, 62M10, 62P05

**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 nonparamateric estimation of volatility function in diffusion models. A nonlinear wavelet estimate of the volatiltiy 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.

Keywords: wavelets estimation, diffusion process estimation, volatiltiy estimation, non-parametric function estimation.

# 1 Introduction

Conditional volatility estimation is of great importance in finance appplications. Parametric models such as the XARCH family are quite used as well as stochastic volatility models. References on parametric volatility modelling are Ghysels et al. [1996], Bollerslev et al. [1992] and Shephard [1996]. Eventhough parametric procedures are still the dominating statistical tools in volatility estimation, non-parametric alternatives have increased their presence in the last decade (Bosq [1998] and Antoniadis and Oppenheim [1995]). 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 polinomial 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 adaptative nature of those estimators, on the sense of Donoho and Johnstone [1995].

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 dra-

matically 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 both motivated and challenged wavelet application in financial data. We compared kernel and wavelet based procedures in both simulated and real data sets. Emphasis was put in structural changes and outliers effects on the volatiltiy function. 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] e Walter [1992b]). 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 that are still not theoretically proved.

The text goes as follows. In section 2, diffusion models are presented. Non-parametric estimators are presented in section 3 and their adaptations to diffusion models are made in section 4 as well as presentation of some of their most relevant theoretical results. Illustrations by simulation and real data analysis are made in section 5. Some final remarks, conclusions and directions for future research are presented in section 6.

## 2 Diffusion Models

Formal definition and mathematical treatment of diffusion models are fundamentally based on the Wiener process and Ito's integral and process. A very important process which is the basis for the non-parametric volatility estimators we study in this paper is the so-called *Quadratic Variation*. Let $\{X_t\}_{t\in T}$ be some Ito's process. Its quadratic variation process is given by $< X, X >_t = \lim_{\Delta t_k \to 0} \sum_{t_k \leq t} |X_{t_{k+1}} - X_{t_k}|^2$, where $0 = t_1 < \cdots < t_n = t$ is some partition of $[0, t]$ and $\Delta t_k = t_{k+1} - t_k$. For every $b : [0, \infty) \times \omega \to \mathbb{R}$ and $\sigma : [0, \infty) \times \omega \to \mathbb{R}$ for which Ito's integral exist, one defines the following diffusion model:

$$dX_t = b(t, X_t)dt + \sigma(t, X_t)dW_t, \tag{1}$$

where $b(\cdot, \cdot)$, known as *drift*, measures the *level* of the process while $\sigma(\cdot, \cdot)$, known as diffusion coefficient, relates to the *speed* and *size* of the series oscillations. One can then

show that the volatility in diffusion models is given by $Var(dX_t/\{X_s\}_{s\leq t}) = \sigma^2(t, X_t)$. For diffusion model simulations, a discretisation of the continuous process is necessary. We implemented the Euler's method (see for details Oskendal [1995]).

# 3 Non-parametric Estimation

## 3.1 Kernel Estimation

Kernel function estimation origins are associated to densities as follows. Let $\{X_1, ..., X_n\}$ be a sample from $F_X(\int f_X)$ and $\widehat{f_h}(x)$ its kernel density estimator. Some usual properties of a kernel density estimator are its smoothness (of the same degree of the kernel function itself), its optimal (or near optimal) asymptotic behavior, its partial adaptation to data (via costly data-driven choices of $h$). Some of those properties will be inherited in the financial data setup. The parameter $h$ controls the estimate degree of smoothness and is therefore called the smoothing parameter.

Under a regression setup, let $\{(X_i, Y_i)\}_{i=1}^n$ be a sample from $(X, Y)$, and $m(x) = E(Y/X = x)$. Suppose $X$ and $Y$ have a joint density $f_{X,Y}$ and respective marginals $f_X$ and $f_Y$. The Nadaraya-Watson estimator for $m$ is given by

$$\widehat{m}(x) = \frac{\sum_{i=1}^{n} K\left(\frac{X_i - x}{h}\right) Y_i}{\sum_{i=1}^{n} K\left(\frac{X_i - x}{h}\right)}, \qquad x \in \mathbb{R}. \tag{2}$$

For assessing $\widehat{m}(x)$'s global performance several error measures are proposed in the literature. The integrated absolute error, IAE, the integrated squared error, ISE, the integrated mean squared error, IMSE, the discrete squared error, DSE, and the discrete mean squared error, DMSE, are usually considered to provide good performance indicators. We restrict ourselves to the ISE and IAE measures due to their feasibility in real applications. Under some regularity conditions on $f$, $m$ and $K$, bounds and convergence of IMSE, MSE and IAE results are available. Bosq [1998] and Härdle [1990] provide the reader with detailed information on properties of kernel based regression estimators for iid and dependent data.

Among the techniques for choosing the optimal $h$, crossvalidation has very nice reasoning and some robustnes against oversmoothing. The optimal $h$ by crossvalidation is given by $\widehat{h}_{otm} = arg\min_h CV(h)$, where $CV(h) = n^{-1}\sum_{j=1}^{n}[Y_j - \widehat{m}_{h,j}(X_j)]^2 w(X_j)$ and $\widehat{m}_{h,j}(\cdot)$ is the Nadaraya-Watson estimator without the $j$-th observation.

### 3.2 Wavelets Estimation

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].

**Definition 3.1.** *A Multiresolution Analysis of $L_2(\mathbb{R})$, $\{V_j,\ j \in \mathbb{Z}\}$, in is an increasing sequence of closed subespaces 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 $V_0$.*

Let $\phi_{jk}(\cdot) = \phi(2\cdot - k)$, $j,k \in \mathbb{Z}$. One can derive from a MRA that for each $j \in \mathbb{Z}$, $\{\phi_{jk}, k \in \mathbb{Z}\}$ is an ortonormal 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 fron the MRA and we will call it $\{\psi_{jk}\}, k \in \mathbb{Z}$. 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 ortonormal bases for $L_2(\mathbb{R})$. Finally, any square integrable function can be written as

$$f(t) = \sum_{j\in\mathbb{Z}}\sum_{k\in\mathbb{Z}} \beta_{j,k}\psi_{j,k}(t) = \sum_{k\in\mathbb{Z}} \alpha_{j_0,k}\phi_{j_0,k}(t) + \sum_{j\geq j_0\in\mathbb{Z}}\sum_{k\in\mathbb{Z}} \beta_{j,k}\psi_{j,k}(t), \tag{3}$$

where equalities are meaninful in $L_2$ norm.

Desirable features of wavelets bases (from either the theoretical or applied point of view) include compact support and adaptative smoothness. A family that has both properties is Daubechies wavelets, that we will be using here. The bases are indexed by the number of nule 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.

From (3) 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 $\Delta$) 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

$$f(t) \simeq \sum_{k \in K_{j_0}} \alpha_{j_0,k}\phi_{j_0,k}(t) + \sum_{j=j_0}^{j_1} \sum_{k \in K_j} \beta_{j,k}\psi_{j,k}(t). \tag{4}$$

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 $j_0 = \lceil \log_2(2N-1) \rceil$. The cascade algorithm can be then employed for fast computation of the coefficients $\alpha_{j,k}$ e $\beta_{j,k}$.

Eventhough (4) is a nice approximation, specially for smooth $f$ and $\phi$, 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. Choosing some $\lambda > 0$, Hard thresholding is given by a rule such as $\delta^d(c_{i,j}, \lambda) = c_{i,j}\mathbf{1}_{(|c_{i,j}|>\lambda)}$ and Soft thresholding has rules suchs as $\delta^s(c_{i,j}, \lambda) = (c_{i,j} - sign(c_{i,j})\lambda)\mathbf{1}_{(|c_{i,j}|>\lambda)}$.

The art in thresholding is choosing the value (or values) of $\lambda$ and proposing a rule which poses as little weight as possible on that choice. In other words, one wants an optimal $\lambda$ 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 $\sigma$ is the empirical

coefficients standard deviation.

Pinheiro and Vidakovic [1997] proposed another thresholding procedure, 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 $f$ and $d_j^2 = \sum_k \beta_{j,k}^2$. The Lorentz curve based on their empirical level energies is:*

$$L(p) = \frac{\sum_{j=0}^{\lfloor np \rfloor} \sum_k \beta_{j,k}^2}{\sum_{j=0}^{n} \sum_k \beta_{j,k}^2} = \frac{\sum_{j=0}^{\lfloor np \rfloor} d_j^2}{\sum_{j=0}^{n} d_j^2}, \qquad p \in [0,1]. \tag{5}$$

**Theorem 3.1 (Meyer).** *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. However, one has also the following empirical version of theorem 3.1.

**Theorem 3.2.** *Let $\widehat{\overline{d}}_j^2 = n^{-1} \sum_k \widehat{\beta}_{j,k}^2 n$ and $B_N = \min_{x \in \mathbb{R}} \max_k \psi_{0,k}^2(x) > 0$. Then*

$$\widehat{\overline{d}}_j^2 \geq \frac{B_N 2^j}{Cn}. \tag{6}$$

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 that apparent contradiction, Pinheiro and Vidakovic [1997] proposed the following thresholding procedure.

Take the empirical Lorentz curve, i.e., $\widehat{L}(p) = \sum_{j=0}^{\lfloor np \rfloor} \widehat{d}_j^2 / \sum_{j=0}^{n} \widehat{d}_j^2$, $p \in [0,1]$. One can then adjust the maximum level of resolution for the empirical Lorentz curve 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 those levels whose behavior conforms to theorem 3.1. After choosing $j_1$, one aplies a coefficients thresholding as follows.

**Definition 3.3 (LC1 Thresh.).** *Suppose a sequence of estimated coefficients* $\{\beta_{j,k} : j = 0,..,n \ k \in \mathbb{Z}\}$ *and a positive parameter* $\kappa$ *and let* $\overline{d}^2 = n^{-1}\sum_{j=0}^{n}\sum_k \beta_{j,k}^2$ . *The LC1 thresholded coefficients (LC1) is given by* $\beta_{j,k}^{lim} = \beta_{j,k}\mathbf{1}_{(\beta_{j,k}^2 > \kappa\overline{d}^2)}$.

The cut-off parameter $\kappa$ in Definition 3.3 has a very simple interpretation.It balances two conflicting measures: parsimonious and energy retention. If $k \leq 1$, one tends to preserve energy retaining more coefficients whilst, for $k > 1$, 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. The level choice procedure is presented in section 5.1. Coefficients thresholding is performed level by level as follows.

**Definition 3.4 (LC2 Thresh.).** *Let* $\kappa$ *and* $\kappa_j$ *be positive real numbers and* $\{\beta_{j,k} : j = 0,..,n \ k \in \mathbb{Z}\}$ *be the sequence of estimated coefficients and* $\overline{d}_j^2 = n^{-1}\sum_k \beta_{j,k}^2$. *The LC2 thresholded coefficients are given by* $\hat{\beta}_{j,k}^{lim} = \beta_{j,k}\mathbf{1}_{(\beta_{j,k}^2 > \kappa\overline{d}^2 + \kappa_j\overline{d}_j^2)}$.

Notice that

$$\kappa\overline{d}^2 + \kappa_j\overline{d}_{j'}^2 = \left(\frac{n_{j'}}{n}\kappa + \kappa_j\right)\overline{d}_{j'}^2 + \kappa\overline{d}_{-j'}^2, \tag{7}$$

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

From (7), one sees that different thresholding levels for each resolution give the procedure more leverage to weight in or out coeffcients 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:

$$\hat{f}(t) = \sum_{k\in\mathbb{Z}} \widehat{\alpha}_{j_0,k}\phi_{j_0,k}(t) + \sum_{j=j_0}^{j_1}\sum_{k\in\mathbb{Z}} \widehat{\beta}_{j,k}^{lim}\psi_{j,k}(t). \tag{8}$$

# 4 Nonparametric Volatility Estimation

Suppose (1) and take $\sigma^2(t, X_t) = \sigma^2(t)$, i.e.,

$$dX_t \quad = \quad b(t, X_t)dt + \sigma(t)dWt, \qquad t \in [0, T], \tag{9}$$

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

## 4.1 Kernel Estimator

Suppose $X_t$ is observed with $T = 1$, at times $(t_1, ..., t_n)$, $t_i = i\Delta$, where $\Delta$ 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 [1993].

**Definition 4.1 (Volatility Kernel Estimator).** *The volatility function $\sigma^2(t)$, as defined by (9), can be estimated by:*

$$S_n(t) = \frac{\sum_{i=1}^{n-1} K\left(\frac{X_{t_i} - t}{h}\right) n^{-1}[X_{t_{i+1}} - X_{t_i}]^2}{\sum_{i=1}^{n} K\left(\frac{X_{t_i} - t)}{h}\right)}. \tag{10}$$

The estimator, defined by (10) is precisely the Nadaraya-Watson regression estimator taking $Y_i = \Delta^{-1}[X_{t_{i+1}} - X_{t_i}]^2$, the empirical quadratic variation process. The following theorem shows some of its most relevant asymptotic properties (due to Florens-Zmirou [1993]).

**Theorem 4.1.** *Let $\sigma^2(t)$ be given by (9) and $S_n(t)$ by (10). Suppose that $b(\cdot)$ 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:*

$$\sqrt{nh}\left(\frac{S_n(t)}{\sigma^2(t)} - 1\right) \xrightarrow{\mathcal{D}} L^{-1/2}(t)Z, \tag{11}$$

*where* $Z \sim \mathcal{N}(0,1)$ *independent of* $L(t) = \lim_{\delta\to 0} \frac{1}{2\delta}\int_0^1 1_{(|X_s - t|<\delta)}ds$.

To our knowledge there are no general asymptotic results such as Theorem 4.1 on the IMSE. 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.

$$S_n^{MA}(t) = (A\Delta)^{-1}\sum_{j=0}^{n-1}\left(\sum_{i=1}^{A-1}[X_{t_{j-i+1}} - X_{t_{j-i}}]^2\right)\mathbf{1}_{[t_j,t_{j+1})}(t) \tag{12}$$

$$S_n^{CMA}(t) = (A\Delta)^{-1}\sum_{j=0}^{n-1}\left(\sum_{i=-\frac{A}{2}}^{\frac{A}{2}-1}[X_{t_{j-i+1}} - X_{t_{j-i}}]^2\right)\mathbf{1}_{[t_j,t_{j+1})}(t), \tag{13}$$

where $A$ is the number of observations used for estimating volatility in each point.

Consistency of $S_n^{CMA}$ is clear as long as $t$ is a continuity point and for those points for which $\sigma^2(t_-) = \lim_{x\nearrow t}\sigma^2(x)$. Bertrand [1996] also showed that, if $\sigma(\cdot) = \sum_{i=1}^N \sigma_i 1_{[t_i,t_{i+1})}(\cdot)$, for positive constants $\sigma_i\ \forall i$, and $w(\cdot)$ some weight function, an also if $\Delta \to 0$, $A \to \infty$ and $A\Delta \to 0$, then $IMSE_{A,\Delta} \to 0$.

## 4.2 Wavelet Estimators

Consider a sample of size $n$, with $t_i = i2^{-n}$, where $i = 0, 1, ..., N = [2^n T]$, i.e., one observes the diffusion in $[0,T]$ with sampling interval $\triangle_n = 2^{-n}$. Some theoretical asymptotic results are known for a linear wavelet estimator of $\sigma(\cdot)$ under conditions (14), on $b(\cdot,\cdot)$ and $\sigma(\cdot)$, as seen in Genon-Catalot et al. [1992].

$$\left.\begin{array}{ll} \textbf{(i)} & b(\cdot,\cdot) \in C^1([0,+\infty)\times\mathbb{R}); \\ \textbf{(ii)} & \forall T,\ \exists K_T \text{ s.t. } \forall t\ \epsilon\ [0,T],\ |\ b(t,u)\ | \leq K_T(1+|\ u\ |); \\ \textbf{(iii)} & \sigma(\cdot) \in C^m([0,\infty)), \text{ with } m \geq 1, \text{ and } \sigma(t) > 0\ \forall t \geq 0. \end{array}\right\} \tag{14}$$

Under (14), and $m \geq 0$, solution of (1) 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 $\bar{\sigma}(t) = \sigma(t), \quad \forall t \in [0,T]$. 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 $d\bar{X}_t = b(t,\bar{X}_t)dt + \bar{\sigma}(t)dW_t, \quad \bar{X}_0 = x$, 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

$$\hat{\alpha}_{j(n),k} = \sum_{i=0}^{N-1} \phi_{j(n),k}(t_i)(X_{t_{i+1}} - X_{t_i})^2. \tag{15}$$

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

$$\hat{\sigma}^2(t) = \sum_{k\epsilon\mathbb{Z}} \hat{\alpha}_{j(n),k}\phi_{j(n),k}(t). \tag{16}$$

The estimator (16) is proposed by Genon-Catalot et al. [1992] where the following IME and IMSE properties are presented:

$$2^{\frac{n}{2}} \int_{\mathbb{R}} h(t)\left(\hat{\sigma}^2(t) - \sigma^2(t)\right) dt \overset{n\to\infty}{\to} \mathcal{N}\left(0, 2\int_0^T h^2(t)\sigma^4(t)dt\right), \tag{17}$$

$$\begin{aligned} R_n &= E\left(\int_{\mathbb{R}} (\hat{\sigma}^2(t) - \sigma^2(t))^2\gamma(t)dt\right) \leq C(2^{4j(n)-2n} + 2^{-2j(n)(m\wedge r)} + 2^{-n}) + \\ &\quad \sup_{t\epsilon\mathbb{R}}\gamma(t)\left(2^{j(n)-n}2\int_0^T \sigma^4(t)dt + o(2^{j(n)-n})\right), \end{aligned} \tag{18}$$

where $C$ is a constant depending only on $\phi$, $\gamma$ and $\sigma^2$.

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

**Definition 4.2 (Nonlinear Wavelet Estimator).** *The nonlinear wavelet estimator for*

*the volatility function $\sigma^2$ is given by*

$$\hat{\sigma}^2(t) = \sum_{k\epsilon\mathbb{Z}} \widehat{\alpha}_{j_0,k}\phi_{j_0,k}(t) + \sum_{j=j_0}^{j_1}\sum_{k\in\mathbb{Z}} \widehat{\beta}_{j,k}^{lim}\psi_{j,k}(t), \tag{19}$$

*where $j_0$, $j_1$ are respectively the minimum and maximum resolutions given by the Lorentz curve, $\widehat{\alpha}_{j_0,k}$ are the empirical approximation coeffcients and $\widehat{\beta}_{j,k}^{lim}$ are the LC2 thresholded estimates for the detail coeffcients.*

# 5 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. In pre-simulation studies, Gaussian smoothing parameter was chosen in order to minimize ISE, as seen in Härdle [1990]. Another comment on the Gaussian kernel is that, although it violates conditions for results presented in section 3.1, 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 the 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 $\kappa$, described in table 1. For levels higher then 3, $\kappa_j = \kappa_3$.

Table 1: Thresholding Parameters

| **Combinations** | $\kappa_1$ | $\kappa_2$ | $\kappa_3$ | $\kappa$ |
|---|---|---|---|---|
| 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 |

## 5.1 Simulation Study

We use in this simulation study the following model

$$dX_t = b(X_t)dt + \theta(t)h(X_t)dWt,\ \ t\epsilon[0,T],\ \ X_0 = x_0, \tag{20}$$

where $h = 1$, $b = 0.75 - X_t$ and $\theta$ has eight different types, illustrated in figures 5.1-5.1, 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 5.1 and 5.1 shows respectively the *irregular* and *regular* volatility functions.

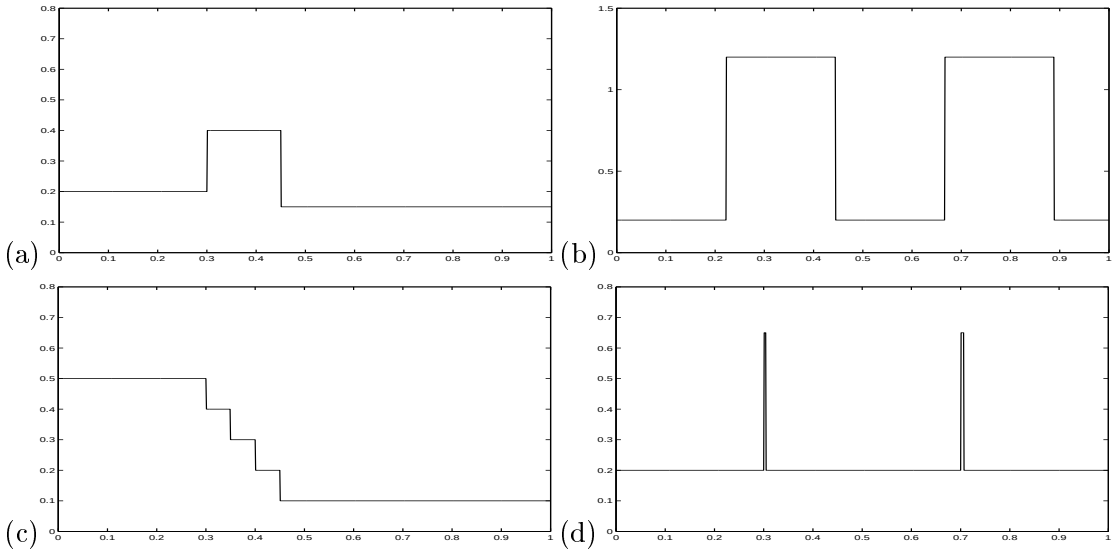


Figure 1: Irregular $\theta(\cdot)$ Functions (with jumps): (a) Vol1 (b) Vol2 (c) Vol3 (d) Vol4

The eight models were generated in 4096 time points with 1000 simulated data series each. Because of the massive output we illustrate mainly the results from volatilities Vol2 and Vol4. They represent respectively level changes and functions with local perturbations. Discussions are however carried out for all eight models.

It is important to observe that Vol2 and Vol4 do not satisfy some of the regularity conditions required for either the kernel or linear wavelet estimators theoretical results.

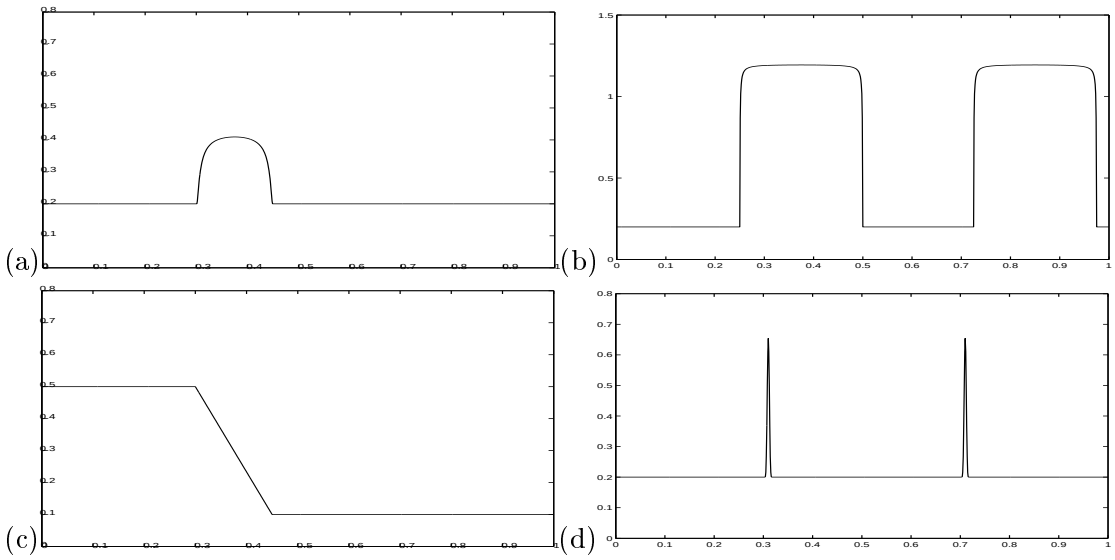


Figure 2: Regular $\theta(\cdot)$ Functions: (a) Vol5 (b) Vol6 (c) Vol7 (d) Vol8

For error comparisons we used two of the previously defined measures, ISE and IAE. 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. Therefore, two other error measures (weighed ISE and IAE) were considered. Three weight functions were used:

- *Quant100*, that puts uniform weight on change points and on its 200 nearest neighbors and zero on the rest.
- *Quant50*, that puts uniform weight on change points and on its 100 nearest neighbors and zero on the rest.
- *Quant0*, that puts uniform weight on change points and zero on the rest.

Results are summarized in tables and plots, using means and medians as descriptive tools. In all tables, the smallest IE's are highlighted. Smoothed histograms for each estimator empirical error distribution are presented as well as some computing performance considerations.

The overall picture of the tables and plots indicates that for volatilities types 1,2,3,5,6 and 7, the Gaussian kernel shows the best ISE and IAE empirical distributions. Wavelets

however have a very close performance and both techniques clearly outperform the MA and CMA techniques.

Among wavelet estimators, ratio 1 and thresholding combination 4 show slight superiority in some volatilities (1,2,5 and 6) but the procedures show some robustness to energy ratio and thresholding cutoff points specifications.

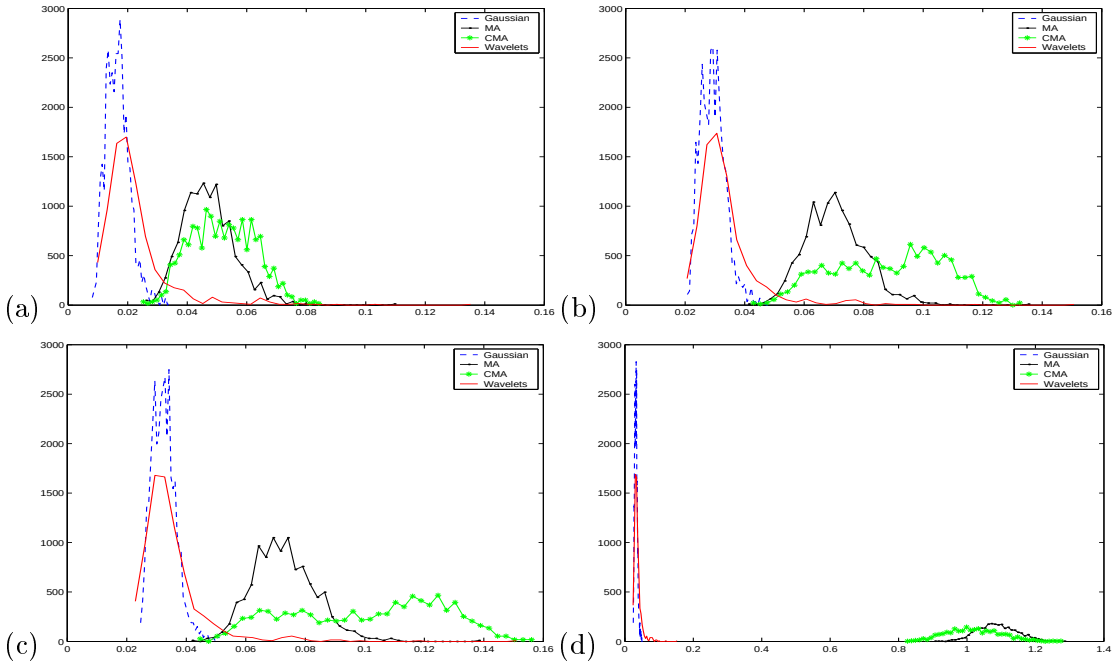


Figure 3: Vol2 ISE: (a) Quant0 (b) Quant50 (c) Quant100 (d) Total ISE

For Vol8, wavelets have a far superior performance specially when analising the weighed error measures. As expected, MA and CMA showed poor performances. For Vol4, MA and CMA showed better global performances than wavelets and much better than normal kernel estimators. That however changes (in favor of wavelets) when local performance is considered. The unexpected result is that MA outperforms CMA for Vol8, what contradicts Bertrand [1996], for jumping diffusion functions.

For Vol8, with local measures, wavelets integrated errors are less than half as big as Gaussian's whilst for Vol4 they are more than four times smaller. For the other volatilities Gaussian kernel outperforms wavelets but never close to the advantage wavelets have for

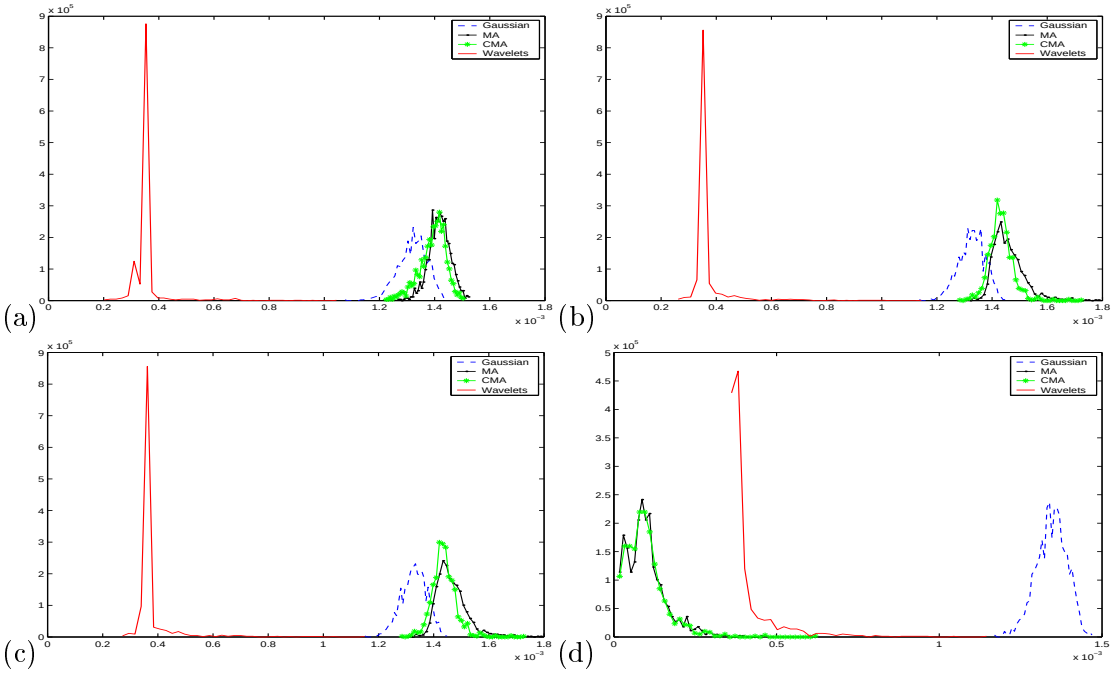


Figure 4: Vol4 ISE: (a) Quant0 (b) Quant50 (c) Quant100 (d) Total ISE

Table 2: Quant0 ISE Medians and Means - Vol1-Vol8 (x$10^{-4}$)

| Type | Wavelets | | Gaussian | | CMA | | MA | |
|---|---|---|---|---|---|---|---|---|
| | Median | Mean | Median | Mean | Median | Mean | Median | Mean |
| Vol1 | 1.23 | 1.68 | **0.83** | **0.88** | 2.38 | 2.41 | 2.30 | 2.49 |
| Vol2 | 191.50 | 204.44 | **164.10** | **166.66** | 516.11 | 5.20.31 | 463.75 | 471.48 |
| Vol3 | 0.65 | 0.76 | **0.52** | **0.59** | 0.98 | 1.03 | 1.00 | 1.06 |
| Vol4 | **2.97** | **3.01** | 13.24 | 13.19 | 14.18 | 13.93 | 14.00 | 14.17 |
| Vol5 | 0.85 | 1.15 | **0.43** | **0.47** | 1.68 | 1.79 | 1.73 | 1.83 |
| Vol6 | 333.33 | 343.97 | **169.16** | **174.51** | 502.78 | 510.80 | 527.21 | 535.13 |
| Vol7 | 0.29 | 0.51 | **0.22** | **0.28** | 0.93 | 1.07 | 0.94 | 1.05 |
| Vol8 | **2.48** | **2.72** | 6.81 | 6.85 | 8.26 | 8.27 | 8.66 | 8.89 |

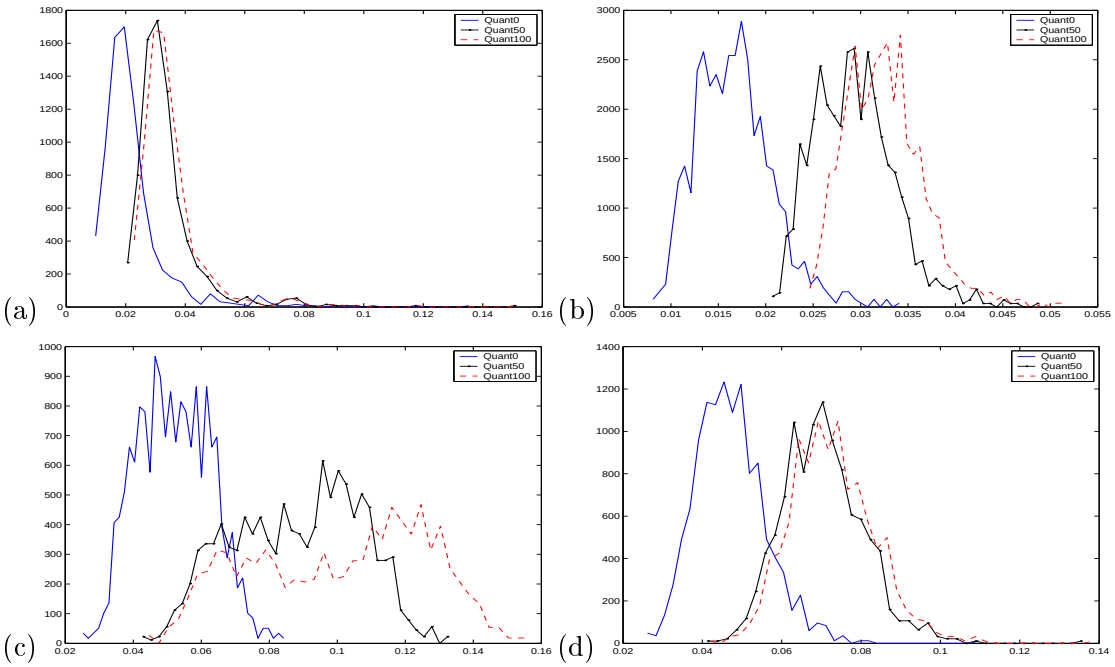


Figure 5: Vol2 ISE: (a) Wavelets (b) Gaussian (c) CMA (d) MA

those two *irregular* situations.

Another contradiction to results in Bertrand [1996] is that here wavelets outperform MA and CMA. In Bertrand [1996], 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.

One should also notice that in a sense both Vol4 and Vol8 represent local outliers but Vol8 is more regular, where Gaussian kernels can perform well while Vol4 represents a pure jump where wavelets higher performance will be enhanced.

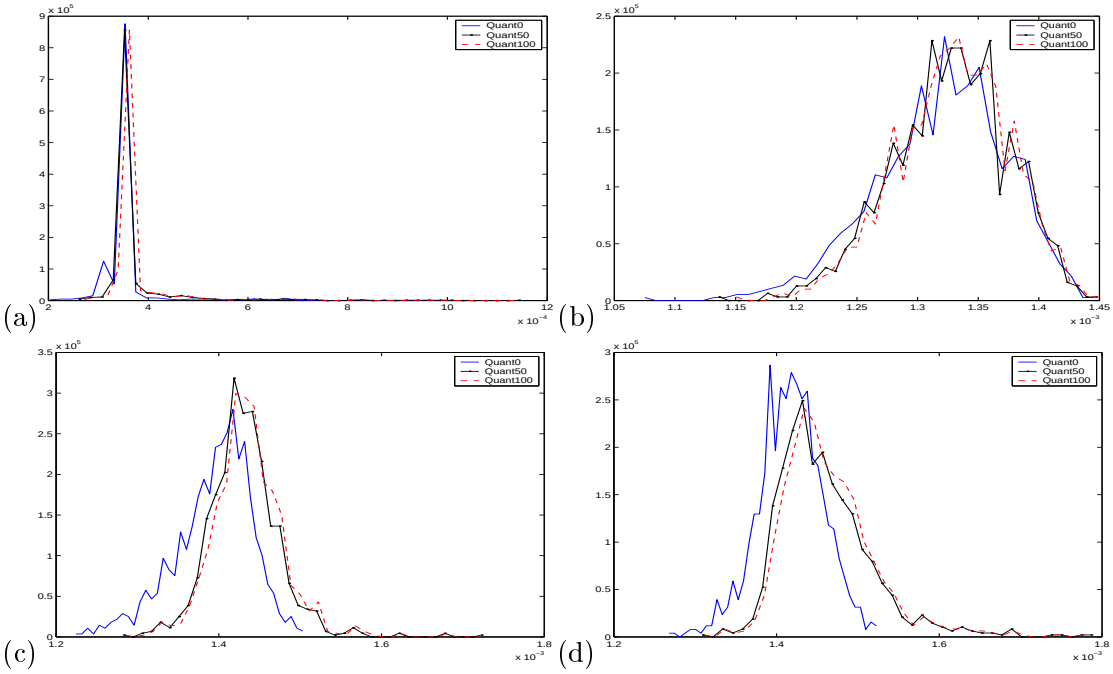


Figure 6: Vol4 ISE: (a) Wavelets (b) Gaussian (c) CMA (d) MA

Table 3: IE Mean and Median - Vol2 and Vol4 ($x10^{-2}$)

| Mean and Median | | | | Vol2 | | | | Vol4 | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | ISE | | IAE | | ISE | | IAE | |
| | | | | Mean | Med. | Mean | Med. | Mean | Med. | Mean | Med. |
| WAVELETS | DB2 | Ratio 1 | C1 | 4.69 | 4.36 | 13.8 | 13.5 | 0.039 | 0.037 | 0.524 | 0.500 |
| | | | C2 | 3.73 | 3.53 | 11.9 | 11.7 | 0.039 | 0.037 | 0.521 | 0.509 |
| | | | C3 | 3.93 | 3.78 | 12.2 | 12.2 | **0.038** | 0.037 | 0.503 | 0.497 |
| | | | C4 | 3.64 | 3.33 | 11.4 | 11.2 | 0.0407 | 0.038 | 0.598 | 0.532 |
| | | Ratio 3 | C1 | 8.89 | 10.6 | 20.7 | 23.3 | 0.0382 | 0.037 | **0.479** | 0.466 |
| | | | C2 | 8.60 | 10.6 | 20.1 | 23.3 | 0.0383 | 0.037 | 0.494 | 0.483 |
| | | | C3 | 8.68 | 10.6 | 20.2 | 23.3 | 0.0383 | 0.037 | 0.494 | 0.483 |
| | | | C4 | 8.55 | 10.6 | 19.9 | 23.3 | 0.0383 | 0.037 | 0.505 | 0.494 |
| | DB5 | Ratio 1 | C1 | 6.44 | 6.33 | 14.3 | 14.3 | 0.0445 | 0.042 | 0.988 | 0.982 |
| | | | C2 | 6.44 | 6.33 | 14.3 | 14.3 | 0.0447 | 0.042 | 0.994 | 0.983 |
| | | | C3 | 6.44 | 6.33 | 14.3 | 14.3 | 0.0446 | 0.042 | 0.991 | 0.983 |
| | | | C4 | 6.44 | 6.33 | 14.3 | 14.3 | 0.0451 | 0.042 | 1.004 | 0.983 |
| | | Ratio 3 | C1 | 6.42 | 6.33 | 14.2 | 14.3 | 0.0428 | 0.042 | 0.983 | 0.980 |
| | | | C2 | 6.42 | 6.33 | 14.2 | 14.3 | 0.0428 | 0.042 | 0.983 | 0.980 |
| | | | C3 | 6.42 | 6.33 | 14.2 | 14.3 | 0.0428 | 0.042 | 0.983 | 0.980 |
| | | | C4 | 6.42 | 6.33 | 14.2 | 14.3 | 0.0428 | 0.042 | 0.983 | 0.980 |
| | DB8 | Ratio 1 | C1 | 6.41 | 6.31 | 14.3 | 14.3 | 0.0437 | 0.042 | 0.995 | 0.987 |
| | | | C2 | 6.41 | 6.31 | 14.3 | 14.3 | 0.0440 | 0.042 | 1.002 | 0.987 |
| | | | C3 | 6.41 | 6.31 | 14.3 | 14.3 | 0.0438 | 0.042 | 0.998 | 0.987 |
| | | | C4 | 6.41 | 6.31 | 14.3 | 14.3 | 0.0445 | 0.042 | 1.014 | 0.987 |
| | | Ratio 3 | C1 | 6.41 | 6.31 | 14.3 | 14.3 | 0.0428 | 0.042 | 0.987 | 0.983 |
| | | | C2 | 6.40 | 6.31 | 14.3 | 14.3 | 0.0428 | 0.042 | 0.987 | 0.983 |
| | | | C3 | 6.40 | 6.31 | 14.3 | 14.3 | 0.0428 | 0.042 | 0.987 | 0.983 |
| | | | C4 | 6.40 | 6.31 | 14.3 | 14.3 | 0.0428 | 0.042 | 0.987 | 0.983 |
| Kernel Function | | MA | | 108.8 | 108.4 | 76.4 | 76.3 | 0.0100 | 0.0091 | 0.647 | 0.688 |
| | | CMA | | 102.0 | 101.6 | 74.6 | 74.6 | 0.0098 | **0.0089** | 0.644 | **0.686** |
| | | Gaussian | | **3.26** | **3.23** | **10.0** | **9.96** | 0.135 | 0.1345 | 0.779 | 0.774 |

Table 4: ISE Means - Vol2 and Vol4 (x$10^{-2}$)

| Mean | | | | Vol2 | | | Vol4 | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | Quant0 | Quant50 | Quant100 | Quant0 | Quant50 | Quant100 |
| WAVELETS | DB2 | Ratio 1 | C1 | 2.46 | 3.57 | 4.14 | 0.036 | 0.037 | 0.037 |
| | | | C2 | 2.13 | 3.24 | 3.57 | 0.035 | 0.037 | 0.037 |
| | | | C3 | 2.04 | 3.26 | 3.74 | 0.035 | 0.037 | 0.037 |
| | | | C4 | 2.23 | 3.36 | 3.51 | 0.036 | 0.037 | 0.038 |
| | | Ratio 3 | C1 | 4.22 | 5.98 | 7.20 | 0.035 | 0.035 | 0.036 |
| | | | C2 | 4.13 | 5.89 | 7.03 | 0.035 | 0.035 | 0.036 |
| | | | C3 | 4.12 | 5.92 | 7.10 | 0.353 | 0.035 | 0.036 |
| | | | C4 | 4.13 | 5.90 | 6.99 | 0.352 | 0.035 | 0.036 |
| | DB5 | Ratio 1 | C1 | 5.62 | 6.40 | 6.41 | 0.310 | 0.036 | 0.036 |
| | | | C2 | 5.62 | 6.40 | 6.42 | 0.310 | 0.036 | 0.036 |
| | | | C3 | 5.62 | 6.40 | 6.42 | 0.310 | 0.036 | 0.036 |
| | | | C4 | 5.62 | 6.40 | 6.42 | 0.311 | 0.036 | 0.036 |
| | | Ratio 3 | C1 | 5.61 | 6.38 | 6.40 | 0.313 | 0.035 | 0.035 |
| | | | C2 | 5.61 | 6.38 | 6.40 | 0.313 | 0.035 | 0.035 |
| | | | C3 | 5.61 | 6.38 | 6.40 | 0.313 | 0.035 | 0.035 |
| | | | C4 | 5.61 | 6.38 | 6.40 | 0.313 | 0.035 | 0.035 |
| | DB8 | Ratio 1 | C1 | 5.63 | 6.37 | 6.39 | 0.308 | 0.035 | 0.036 |
| | | | C2 | 5.63 | 6.37 | 6.39 | 0.308 | 0.035 | 0.036 |
| | | | C3 | 5.63 | 6.37 | 6.39 | 0.308 | 0.035 | 0.036 |
| | | | C4 | 5.63 | 6.37 | 6.39 | 0.309 | 0.035 | 0.036 |
| | | Ratio 3 | C1 | 5.62 | 6.37 | 6.38 | **0.301** | **0.034** | **0.035** |
| | | | C2 | 5.62 | 6.37 | 6.38 | 0.301 | 0.0344 | 0.035 |
| | | | C3 | 5.62 | 6.37 | 6.38 | 0.301 | 0.0344 | 0.035 |
| | | | C4 | 5.62 | 6.37 | 6.38 | 0.301 | 0.0344 | 0.035 |
| Kernel Function | | | MA | 4.71 | 7.042 | 7.28 | 0.142 | 0.146 | 0.146 |
| | | | CMA | 5.20 | 8.80 | 10.07 | 0. 139 | 0.143 | 0.144 |
| | | | Gaussian | **1.67** | **2.93** | **3.25** | 0.139 | 0.143 | 0.144 |

Table 5: ISE Medians - Vol2 and Vol4 (x$10^{-2}$)

| Median | | | | Vol1 | | | Vol2 | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | Quant0 | Quant50 | Quant100 | Quant0 | Quant50 | Quant100 |
| WAVELETS | DB2 | Ratio 1 | C1 | 2.15 | 3.25 | 3.84 | 0.0353 | 0.0355 | 0.0356 |
| | | | C2 | 1.94 | 3.03 | 3.37 | 0.0352 | 0.0354 | 0.0356 |
| | | | C3 | 1.91 | 3.10 | 3.58 | 0.0352 | 0.0354 | 0.0356 |
| | | | C4 | 1.97 | 3.11 | 3.27 | 0.0351 | 0.0354 | 0.0356 |
| | | Ratio 3 | C1 | 4.80 | 6.81 | 8.36 | 0.0353 | 0.0355 | 0.0356 |
| | | | C2 | 4.79 | 6.81 | 8.36 | 0.0353 | 0.0355 | 0.0356 |
| | | | C3 | 4.79 | 6.81 | 8.36 | 0.0353 | 0.0354 | 0.0356 |
| | | | C4 | 4.79 | 6.81 | 8.36 | 0.0352 | 0.0354 | 0.0356 |
| | DB5 | Ratio 1 | C1 | 5.53 | 6.29 | 6.30 | 0.0306 | 0,000334 | 0.0339 |
| | | | C2 | 5.53 | 6.29 | 6.30 | 0.0306 | 0.0334 | 0.0339 |
| | | | C3 | 5.53 | 6.29 | 6.30 | 0.0306 | 0.0334 | 0.0339 |
| | | | C4 | 5.53 | 6.29 | 6.30 | 0.0306 | 0.0334 | 0.0340 |
| | | Ratio 3 | C1 | 5.52 | 6.29 | 6.30 | 0.0307 | 0.0334 | 0.0340 |
| | | | C2 | 5.52 | 6.29 | 6.30 | 0.0307 | 0.0334 | 0.0340 |
| | | | C3 | 5.52 | 6.29 | 6.30 | 0.0307 | 0.0334 | 0.0340 |
| | | | C4 | 5.52 | 6.29 | 6.30 | 0.0307 | 0.0334 | 0.0340 |
| | DB8 | Ratio 1 | C1 | 5.54 | 6.28 | 6.29 | 0.0298 | **0.0332** | **0.0338** |
| | | | C2 | 5.54 | 6.28 | 6.29 | 0.0298 | 0.0332 | 0.0338 |
| | | | C3 | 5.54 | 6.28 | 6.29 | 0.0298 | 0.0332 | 0.0338 |
| | | | C4 | 5.54 | 6.28 | 6.29 | 0.0298 | 0.0332 | 0.0338 |
| | | Ratio 3 | C1 | 5.54 | 6.28 | 6.29 | **0.0297** | 0.0332 | 0.0339 |
| | | | C2 | 5.54 | 6.28 | 6.29 | 0.0297 | 0.0332 | 0.0339 |
| | | | C3 | 5.54 | 6.28 | 6.29 | 0.0297 | 0.0332 | 0.0339 |
| | | | C4 | 5.54 | 6.28 | 6.29 | 0.0297 | 0.0332 | 0.0339 |
| Kernel Function | | | MA | 4.64 | 6.97 | 7.22 | 0.1418 | 0.1449 | 0.1454 |
| | | | CMA | 5.16 | 8.95 | 10.45 | 0.1400 | 0.1428 | 0.1434 |
| | | | Gaussian | **1.64** | **2.91** | **3.22** | 0.1324 | 0.1329 | 0.1331 |

Table 6: IAE Means - Vol2 and Vol4 (x$10^{-2}$)

| Mean | | | | Vol2 | | | Vol4 | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | Quant0 | Quant50 | Quant100 | Quant0 | Quant50 | Quant100 |
| WAVELETS | DB2 | Ratio 1 | C1 | 7.78 | 9.76 | 10.93 | 0.194 | 0.227 | 0.248 |
| | | | C2 | 7.05 | 9.05 | 9.98 | 0.193 | 0.227 | 0.248 |
| | | | C3 | 6.93 | 9.05 | 10.26 | 0.190 | 0.223 | 0.244 |
| | | | C4 | 7.21 | 9.20 | 9.79 | 0.204 | 0.2429 | 0.269 |
| | | Ratio 3 | C1 | 10.19 | 12.94 | 15.14 | **0.181** | **0.203** | **0.223** |
| | | | C2 | 9.98 | 12.74 | 14.86 | 0.182 | 0.206 | 0.228 |
| | | | C3 | 9.97 | 12.76 | 14.96 | 0.182 | 0.206 | 0.228 |
| | | | C4 | 9.98 | 12.74 | 14.76 | 0.183 | 0.208 | 0.232 |
| | DB5 | Ratio 1 | C1 | 12.24 | 13.62 | 13.82 | 0.289 | 0.363 | 0.405 |
| | | | C2 | 12.25 | 13.62 | 13.82 | 0.290 | 0.365 | 0.407 |
| | | | C3 | 12.25 | 13.62 | 13.82 | 0.290 | 0.364 | 0.406 |
| | | | C4 | 12.25 | 13.62 | 13.82 | 0.292 | 0.367 | 0.409 |
| | | Ratio 3 | C1 | 12.23 | 13.61 | 13.81 | 0.291 | 0.363 | 0.405 |
| | | | C2 | 12.23 | 13.61 | 13.81 | 0.291 | 0.363 | 0.405 |
| | | | C3 | 12.23 | 13.61 | 13.81 | 0.291 | 0.363 | 0.405 |
| | | | C4 | 12.23 | 1.361 | 13.81 | 0.291 | 0.363 | 0.405 |
| | DB8 | Ratio 1 | C1 | 12.27 | 13.62 | 13.83 | 0.285 | 0.367 | 0.411 |
| | | | C2 | 12.27 | 13.62 | 13.83 | 0.286 | 0.368 | 0.412 |
| | | | C3 | 12.27 | 13.62 | 13.83 | 0.286 | 0.367 | 0.411 |
| | | | C4 | 12.27 | 13.62 | 13.84 | 0.288 | 0.371 | 0.415 |
| | | Ratio 3 | C1 | 12.26 | 13.62 | 13.83 | 0.283 | 0.367 | 0.410 |
| | | | C2 | 12.26 | 13.62 | 13.83 | 0.283 | 0.367 | 0.410 |
| | | | C3 | 12.26 | 13.62 | 13.83 | 0.283 | 0.367 | 0.410 |
| | | | C4 | 12.26 | 13.62 | 13.83 | 0.283 | 0.367 | 0.410 |
| Kernel Function | | MA | | 10.78 | 13.03 | 13.42 | 0.479 | 0.569 | 0.605 |
| | | CMA | | 9.90 | 13.07 | 14.47 | 0.475 | 0.559 | 0.596 |
| | | Gaussian | | **6.35** | **8.65** | **9.65** | 0.449 | 0.490 | 0.515 |

Table 7: IAE Medians - Vol2 and Vol4 (x$10^{-2}$)

| Median | | | | Vol2 | | | Vol4 | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | Quant0 | Quant50 | Quant100 | Quant0 | Quant50 | Quant100 |
| WAVELETS | DB2 | Ratio 1 | C1 | 7.47 | 9.46 | 10.66 | 0.188 | 0.214 | 0.235 |
| | | | C2 | 6.90 | 8.91 | 9.84 | 0.188 | 0.215 | 0.238 |
| | | | C3 | 6.87 | 8.95 | 10.19 | 0.186 | 0.212 | 0.234 |
| | | | C4 | 6.96 | 8.97 | 9.58 | 0.191 | 0.220 | 0.246 |
| | | Ratio 3 | C1 | 10.81 | 13.88 | 16.56 | 0.178 | 0.199 | 0.217 |
| | | | C2 | 10.80 | 13.88 | 16.56 | **0.177** | **0.201** | **0.222** |
| | | | C3 | 10.80 | 13.88 | 16.56 | 0.177 | 0.201 | 0.222 |
| | | | C4 | 10.80 | 13.88 | 16.56 | 0.178 | 0.203 | 0.226 |
| | DB5 | Ratio 1 | C1 | 12.22 | 13.64 | 13.84 | 0.287 | 0.360 | 0.399 |
| | | | C2 | 12.22 | 13.64 | 13.84 | 0.288 | 0.360 | 0.401 |
| | | | C3 | 12.22 | 13.64 | 13.84 | 0.288 | 0.360 | 0.400 |
| | | | C4 | 12.22 | 13.64 | 13.84 | 0.289 | 0.361 | 0.402 |
| | | Ratio 3 | C1 | 12.22 | 13.63 | 13.83 | 0.288 | 0.360 | 0.399 |
| | | | C2 | 12.22 | 13.63 | 13.83 | 0.288 | 0.360 | 0.399 |
| | | | C3 | 12.22 | 13.63 | 13.83 | 0.288 | 0.360 | 0.399 |
| | | | C4 | 12.22 | 13.63 | 13.83 | 0.288 | 0.360 | 0.399 |
| | DB8 | Ratio 1 | C1 | 12.31 | 13.65 | 13.87 | 0.282 | 0.362 | 0.405 |
| | | | C2 | 12.31 | 13.65 | 13.87 | 0.282 | 0.363 | 0.405 |
| | | | C3 | 12.31 | 13.65 | 13.87 | 0.282 | 0.362 | 0.405 |
| | | | C4 | 12.31 | 13.65 | 13.87 | 0.282 | 0.364 | 0.406 |
| | | Ratio 3 | C1 | 12.31 | 13.65 | 13.87 | 0.280 | 0.362 | 0.403 |
| | | | C2 | 12.31 | 13.65 | 13.87 | 0.280 | 0.362 | 0.403 |
| | | | C3 | 12.31 | 13.65 | 13.87 | 0.280 | 0.362 | 0.403 |
| | | | C4 | 12.31 | 13.65 | 13.87 | 0.280 | 0.362 | 0.403 |
| Kernel Function | | MA | | 10.80 | 13.06 | 13.44 | 0.483 | 0.573 | 0.609 |
| | | CMA | | 9.87 | 12.99 | 14.39 | 0.477 | 0.564 | 0.599 |
| | | Gaussian | | **6.31** | **8.61** | **9.59** | 0.446 | 0.487 | 0.512 |

### 5.2 Real Series

In this section we estimate the volatiltiy of BOVESPA indexes (São Paulo's exchange market), minute by minute, from 11am on 12/18/01 to 6pm on 02/18/02. This series has 16348 observations for which some minor adjustments are made in order to use casacade algorithms (since $2^{14} = 16384$). One can see the returns series in figure 5.2.

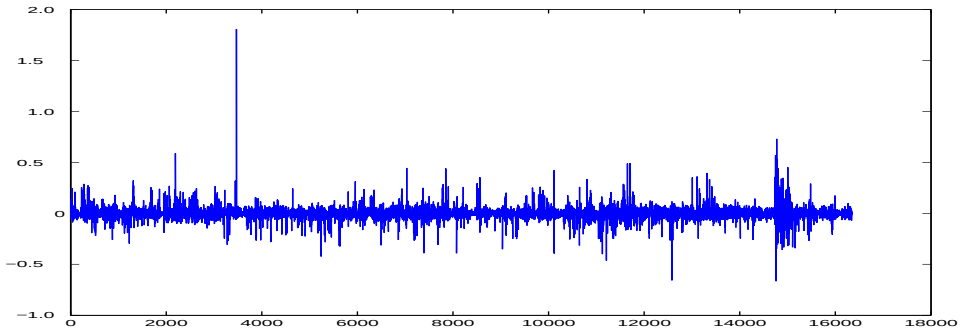


Figure 7: BOVESPA returns - 12/18/01-02/18/02 (x$10^{-2}$)

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. Therefore, we will compare MA, CMA and wavelets performances.

We observe that wavelets estimates produce, as seen in the simulation and expected from theoretical results, higher peaks (closer to the real ones) than either MA or CMA procedures. Moreover, wavelets estimated volatilty curve looks less noisy than either MA or CMA manages to produce.

## 6 Final Remarks

From section 5.1 and 5.2, wavelets 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

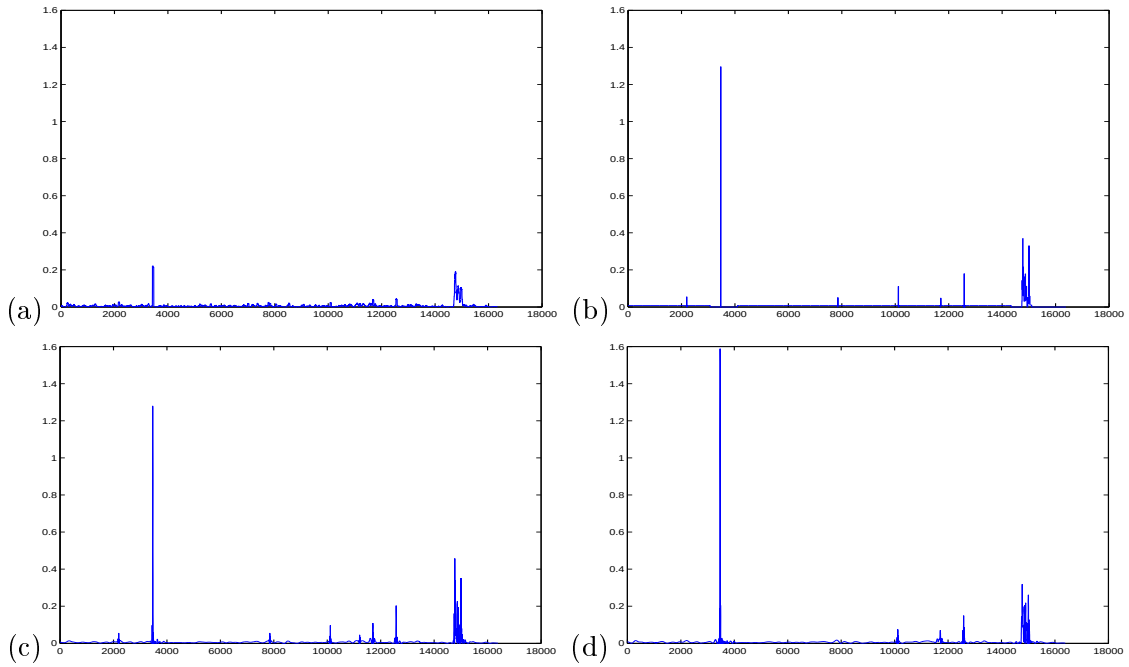


Figure 8: BOVESPA Square-root Volatility Estimates (x$10^{-2}$): (a) CMA (b) DAUB#1 (c) DAUB#5 (d) DAUB#8

results for Gaussin 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.

It is also important to write a few lines concerning computational performance and feasibility. 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.

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 confirms their versatility and robustness to misspecification.

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

- 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.

**Aknowledgements** We thank *Finance Lab*, of *Ibmec Business School*, Brasil, for graciously providing the data set to us. We also thank CAPES, CNPq ans PRONEX for partially supporting this work.

## References

ABBERGER, K. (1997). Quantile smoothing in financial time series. *Statist. Papers*, 38 (2):125–148.

ANTONIADIS, A. and OPPENHEIM, G., editors (1995). *Wavelets and statistics*. New York: Springer-Verlag.

AVESANI, G. and BERTRAND, P. (1997). Does volatility jump or just diffuse? A statistical approach. In *Numerical methods in finance*, pages 270–289. Cambridge: Cambridge Univ. 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. and KRONER, K. (1992). Arch modelling in finance: a review of the theory and empirical evidence. *J. Econometrics*, 52(1):5–59.

BOSQ, D. (1998). *Nonparametric Statistics for Stocasthic 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. *J. Econometrics*, 102(1):1–22.

DONOHO, D. and JOHNSTONE, I. (1995). Adapting to unknown smoothness via wavelet shrinkage. *J. Am. Statist. Assoc.*, 90(432):1200–1224.

DONOHO, D., JOHNSTONE, I., KERKYACHARIAN, G., and PICKARD, D. (1996). Density estimation by wavelet thresholding. *Annals of Statistics*, 24(2):508–539.

FLORENS-ZMIROU, D. (1993). On estimating the diffusion coefficient from discrete observations. *J. Appl. Probab.*, 30:790–804.

GENON-CATALOT, V., LAREDO, C., and PICARD, D. (1992). Nonparametric estimation the diffusion coefficient by wavelet methods. *Scand. J. of Statist.*, 19(4):317–335.

GHYSELS, E., HARVEY, A.C. and RENAULT, E. (1996). Stochastic volatility. In *Statistical Methods in Finance*, volume 14 of *Handbook of Statist.*, pages 119–191. Amsterdam: North-Holland.

HÄRDLE, W. (1990). *Applied Non-parametric Regression*. Cambridge: Cambridge Univ. Press.

HÄRLDE, W. and TSYBAKOV, A. (1997). Local polynomial estimators of the volatility function in nonparametric autoregression. *J. Econometrics*, 81(1):223–242.

HOFFMANN, M. (1999). On nonparametric estimation in nonlinear ar(1)-models. *Statist. Probab. Lett.*, 44(1):29–45.

MALLAT, S. (1989). A theory for multi-resolution signal decomposition: the wavelet representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11 (7):674–693.

OSKENDAL, B. (1995). *Stochastic Differential Equations: An Introduction with Applications*. New York: Springer-Verlag.

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*, pages 1–67. London: Chapman & Hall.

VIDAKOVIC, B. (1999). *Statistical Modeling by Wavelets*. New York: John Wiley & 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*, pages 51–70. New York: Academic Press.

YANG, l. (2000). Finite nonparametric garch model for foreign exchange volatility. In Chien-Pai Han and William W. S. Wei, editors, *Selected papers from the 1999 ASSIC-SAssociation*, pages 1347–1365. Monticello, NY: Marcel Dekker.

YANG, L., HÄRDLE, W., and NIELSEN, J.P. (1999). Nonparametric autoregression with multiplicative volatility and additive mean. *J. Time Ser. Anal.*, 20(5):579–604.