

---

# Wavelet Analysis of Nonlinear Long–Range Dependent Processes. Applications to Financial Time Series\*

Gilles Teyssière<sup>1</sup> and Patrice Abry<sup>2</sup>

<sup>1</sup> Laboratoire de Statistique Appliquée et MOdélisation Stochastique (SAMOS),  
Université Paris 1. [stats@gillesteysiere.net](mailto:stats@gillesteysiere.net)

<sup>2</sup> Laboratoire de Physique, École Normale Supérieure de Lyon, CNRS UMR 5672.  
[Patrice.Abry@ens-lyon.fr](mailto:Patrice.Abry@ens-lyon.fr)

**Summary.** We present and study the performance of the semiparametric wavelet estimator for the long–memory parameter devised by Veitch and Abry (1999). We compare this estimator with two semiparametric estimators in the spectral domain, the local Whittle (LW) estimator developed by Robinson (1995a) and the “log–periodogram” (LP) estimator by Geweke and Porter–Hudak (1983). The wavelet estimator performs well for a wide range of nonlinear long–memory processes in the conditional mean and the conditional variance, and is reliable for discriminating between change–points and long–range dependence in volatility. We also address the issue of selection of the range of octaves used as regressors by the weighted least squares estimator. We will see that using the feasible optimal bandwidths for either the LW and LP estimators, respectively studied by Henry and Robinson (1996) and Henry (2001), is a useful rule of thumb for selecting the lowest octave. We apply the wavelet estimator to volatility series of high frequency (intra–day) Foreign Exchange (FX) rates, and to the volatility and volume of stocks of the Dow Jones Industrial Average Index.

## 1 Introduction

The occurrence of long–range dependence, or strong dependence or long–memory, in economics is documented by numerous research works. Although it is widely accepted that squared or absolute returns on financial assets display long–range dependence, we are still unsure that what is observed is either genuine long–memory or a statistical artefact, as statistical tools used for the study of long–range dependent processes made the assumption that the process under investigation is homogeneous and stationary. Thus, it has

---

\* We thank Carlos Velasco and Olsen & Associates for providing some of the data used in this chapter, Liudas Giraitis for a very careful reading and Gabriel Lang for useful discussions. DOI:10.1007/978-3-540-34625-8\_7

been claimed, see e.g., Mikosch and Stărică (1999, 2003, 2004a, 2004b) that the observed long-range dependence is spurious and the consequence of non-stationarity in the data.

Discriminating long-range dependence from single (or multiple) change-point(s) is a very active research area in statistics, see e.g., Lavielle (1999), Berkes, Horváth, Kokoszka and Shao (2003). Change-points and long-range dependence might coexist, and finding the optimal number of change-points and their location requires a more elaborated approach than Vostrikova's (1981) binary segmentation procedure, see Lavielle and Teyssière (2005), although most econometrics research papers still resort to the binary segmentation methods, or a mix of this algorithm and the tests by Lavielle (1999) and Lavielle and Moulines (2000), a combination giving incorrect results.

We consider here a statistical method for the analysis of long-range dependence, based on wavelets, which does not require these strong stationarity assumptions and is robust to some departures from the previous hypotheses. Empirical research works by Kokoszka and Teyssière (2002) and Teyssière (2003) reached the conclusion that the intensity of strong dependence in volatility processes measured with wavelet based estimators is lower than what is usually found using estimators in either the time or the frequency domain, and then volatility processes mix a moderate level of long-range dependence and change-points.

As empirical evidence suggests that financial time series are highly nonlinear, in particular volatility processes exhibit a combination of nonlinearity, long-range dependence and change-points, we study in this chapter the estimation of the scaling coefficient for some nonlinear long-memory processes used in the statistical and econometrics literature, and compare the performance of the wavelet estimator with the one of two semiparametric estimators in the frequency domain: the local Whittle (LW) estimator by Künsch (1987) and Robinson (1995a), and the "log-periodogram" (LP) estimator by Geweke and Porter-Hudak (1983). These semiparametric estimators are of interest for researchers dealing with real data, as the Data Generating Process (DGP) of the observed data is unknown, the estimation of a misspecified parametric model might lead to serious biases in estimation. Since one has to resort to semiparametric methods, it is then useful to know the performance of these estimators for a wide range of nonlinear dependent processes.

We consider some nonlinear long-memory processes that were not studied in Abry *et al.* (2003): the linear ARCH (LARCH) process, the long-memory stochastic volatility (LMSV) process, the nonlinear moving average (NLMA) process, and some nonlinear transformations of fractionally integrated processes. We also study the performance of the wavelet estimator for the case of change-point processes, i.e., the non-stationary GARCH process, a process with non constant coefficients. The relevance of wavelet analysis for dealing with the issue of change-point and spurious strong dependence in GARCH processes has been conjectured in previous works, see e.g., Kokoszka and Teyssière (2002), Teyssière (2003), but never systematically analyzed. We

also consider the case of a volatility process mixing both strong dependence and change in regimes, the non homogeneous LMSV process, and the case of dependent processes with polynomial and broken trends.

We also address a standard issue in semiparametric estimation of strongly dependent processes, the cutoff between short-range and long-range dependence. The bandwidth selection problem has been studied for the LW and LP estimators by Henry and Robinson (1996), Hurvich, Deo and Brodsky (1998) and Henry (2001). For the wavelet estimators, we have to select the range of octaves used as regressors by the weighted least squares estimator. On the basis of some simulations, we suggest some choices for the lowest octave, and we will see that using the optimal bandwidth for the LW estimator is a useful rule of thumb for selecting this lowest octave.

We also report some simulation results for the use of the feasible optimal bandwidth for the LP estimator, and will see that this data-driven bandwidth works well for a large variety of nonlinear long-range dependent processes, and often better than the fixed bandwidths  $[T^b]$ , with  $b \in (0, 1)$ , usually considered.

We will see that although the wavelet estimator has been devised for the standard Gaussian case, it still works for a broader class of nonlinear processes, provided in the case of highly nonlinear processes, that the sample size is large enough for disentangling the long-memory component from the other nonlinear ones. Since the performance of the wavelet estimator is satisfactory enough, we consider an application of this estimator to high-frequency (intra-day) foreign exchange (FX) rates, and to trading volume.

This chapter is organized as follows: section 2 presents the long-range dependent processes that we will study with the wavelet estimator, while section 3 presents the wavelet and semiparametric spectral estimators. Section 4 compares the performance of the wavelet estimator with other standard semiparametric estimators in the spectral domain for a large variety of nonlinear strongly dependent processes. Section 6 gives several applications to high-frequency Foreign Exchange (FX) rates and to trading volume of stocks of the Dow Jones Industrial Average Index.

## 2 Long-Memory or Long-Range Dependence

### 2.1 Definition and Consequence

**Definition 1.** Let  $\{Y_t, t \in \mathbb{R}\}$  be a second-order stationary process. This process is a long-memory process if its spectrum  $f_Y(\lambda)$  is such that in a close positive neighborhood of the zero frequency,

$$f_Y(\lambda) \sim c_f \lambda^{-\alpha}, \quad \lambda \rightarrow 0_+, \quad c_f \in (0, \infty), \quad (1)$$

or equivalently, if its autocorrelation function  $\rho_Y(k)$  has the following hyperbolic rate of decay<sup>3</sup>

<sup>3</sup>  $x_k \asymp y_k$  means that  $\exists$  two constants  $C_1, C_2$  such that  $C_1 y_k \leq x_k \leq C_2 y_k, k \rightarrow \infty$ .

$$\rho_Y(k) \asymp k^{\alpha-1}, \quad (2)$$

with  $\alpha \in (0, 1)$ .

As a consequence of equation (2), the autocorrelations of a long-memory process are not summable, i.e.,

$$\sum_{k=-\infty}^{\infty} \rho_Y(k) = \infty. \quad (3)$$

Long-memory received an interesting non standard exposition based on the covariance under aggregation of the sum of covariances in Gefferth *et al.* (2003).

## 2.2 Standard Linear Long-Range Dependent Processes

### • Fractionally Integrated Process

The Fractionally Integrated process, denoted as either  $\text{FI}(d)$  or  $\text{I}(d)$ , defined as

$$(1 - L)^d(Y_t - \mu) = \varepsilon_t, \quad \varepsilon_t \sim N(0, \sigma_\varepsilon^2), \quad (4)$$

where  $\mu$  is the unknown mean of the process,  $L$  stands for the lag operator, i.e.,  $LY_t = Y_{t-1}$ , the fractional difference operator  $(1 - L)^d$  associated with the degree of fractional integration  $d \in (-0.5, 0.5)$ , is defined as

$$(1 - L)^d = \sum_{j=0}^{\infty} b_j L^j, \quad b_0 = 1, \quad b_j = \prod_{k=1}^j \left(1 - \frac{1+d}{k}\right), \quad (5)$$

$$b_j \sim -\frac{1}{\Gamma(-d)} j^{-(1+d)} \longrightarrow \infty,$$

where  $\Gamma(\cdot)$  denotes the Gamma function. When  $d \in (0, 0.5)$  (respectively,  $d \in (-0.5, 0)$ ) the process is said to be persistent (respectively, anti-persistent). For this process, one has:

$$\alpha = 2d. \quad (6)$$

### • Fractional ARIMA (FARIMA) Process

The  $\text{FI}(d)$  process is nested into the class of Fractional ARIMA processes, defined as

$$\phi(L)(1 - L)^d(Y_t - \mu) = \theta(L)\varepsilon_t, \quad \varepsilon_t \sim N(0, \sigma_\varepsilon^2), \quad (7)$$

where  $\phi(L) = 1 - \sum_{j=1}^p \phi_j L^j$  and  $\theta(L) = 1 + \sum_{j=1}^q \theta_j L^j$  are lag polynomials of respective orders  $p$  and  $q$  with root outside the unit circle. This model, which is denoted as  $\text{FARIMA}(p, d, q)$ , generalizes the class of  $\text{ARIMA}(p, d, q)$  models with integer degree of differentiation.

The spectrum of a FARIMA( $p, d, q$ ) process is equal to

$$f_Y(\lambda) = \frac{\sigma_\varepsilon^2}{2\pi} \frac{|\theta(\exp(-i\lambda))|^2}{|\phi(\exp(-i\lambda))|^2} |1 - \exp(-i\lambda)|^{-2d}. \quad (8)$$

With

$$\alpha = 2d, \quad (9)$$

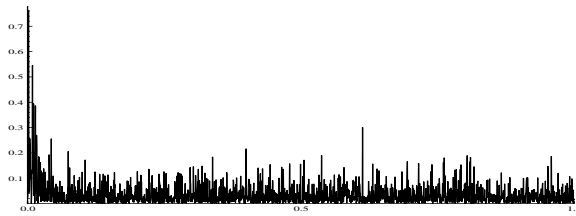
then the spectrum of a FARIMA( $p, d, q$ ) process is of the form of equation (1).

### • Estimation Issues

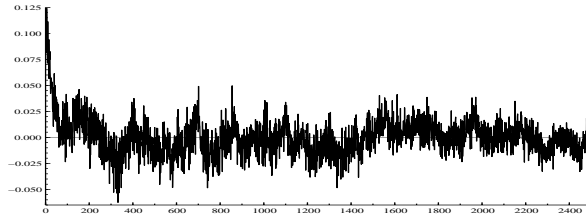
Under the hypothesis of Gaussian error terms, the parameters  $\zeta = \{\theta_i, \phi_j, \mu, d\}$  are estimated by maximizing the log-likelihood function, in either the time or the frequency domain; see Beran (1994), Robinson (1994) and Hauser (1999) for a survey. The parameters  $\hat{\zeta}$  have the rate of convergence equal to  $T^{1/2}$ , where  $T$  denotes the sample size, except the parameter  $\mu$ , the rate of convergence of which is equal to  $T^{1/2-d}$ . This illustrates the difficulty to disentangle long-range dependence from changes in the mean parameter  $\mu$  of the process.

### • Empirical Volatility: The Need for Nonlinear Long-Memory Processes

While returns on asset prices, defined as  $r_t = \log(P_t/P_{t-1})$ , where  $P_t$  denotes the asset price at time  $t$ , are uncorrelated, empirical evidence from the series of absolute returns  $|r_t|$  and squared returns  $r_t^2$  has shown that the spectrum and the autocorrelation function, henceforth ACF, for both series behave like equations (1) and (2). Both the series  $|r_t|$  or  $r_t^2$  are commonly used as empirical measures for the volatility, which hence appear to be strongly dependent. Figures 1 and 2 below display the periodogram and the ACF of absolute returns on dollar-deutschmark FX rate.



**Fig. 1.** Periodogram of absolute returns on dollar-deutschmark FX rate



**Fig. 2.** ACF of absolute returns on dollar-deutschmark FX rate

The rich dynamics of these volatility processes cannot be parsimoniously fitted by standard FARIMA processes. Moreover, as emphasized by Granger (2000, 2002) absolute and squared returns do not display trends, unlike FARIMA processes. Thus, there is a need to resort to long-range dependent nonlinear processes.

### 2.3 Nonlinear Long-Range Dependent Volatility Processes

#### • ARCH( $\infty$ ) Processes

Robinson (1991) introduced the class of ARCH( $\infty$ ) processes, further developed by Granger and Ding (1995) and other authors. The ARCH( $\infty$ ) is defined as

$$r_t = \sigma_t \varepsilon_t, \quad \varepsilon_t \sim D(0, 1), \quad \sigma_t^2 = \omega + \varphi(L)\varepsilon_t^2, \quad (10)$$

where  $D(0, 1)$  is a zero-mean unit-variance random variable, and  $\varphi(L) = \sum_{i=1}^{\infty} \varphi_i L^i$  is an infinite order lag polynomial the coefficients of which are positive and have asymptotically the following hyperbolic rate of decay

$$\varphi_j = O\left(j^{-(1+d)}\right). \quad (11)$$

However, the existence of a stationary solution to the equation (10) defining an ARCH( $\infty$ ) process imply  $\sum_{i=1}^{\infty} \varphi_i < \infty$ , then the model has moderate memory; see Giraitis, Kokoszka and Leipus (2000) for further details.

With these assumptions, the autocorrelation function (ACF) of the sequence of squared returns  $\{r_t^2\}$  satisfies

$$\text{Cov}(r_0^2, r_k^2) \asymp k^{2d-1}. \quad (12)$$

Setting  $\alpha = 2d$ , and  $Y_t = r_t^2$ , the squared returns process defined by a stationary ARCH( $\infty$ ) has moderate memory. In the applied econometrics and financial econometrics literature, the conditions for the existence of a stationary solution are disregarded, and it is wrongly claimed that the ARCH( $\infty$ ) defined by equation (10) has long-memory.

• **Linear ARCH**

Robinson (1991) introduced the linear ARCH, henceforth LARCH, further studied by Giraitis, Robinson and Surgailis (2000), and defined as

$$r_t = \sigma_t \varepsilon_t, \quad \varepsilon_t \sim D(0, 1), \quad \sigma_t = \omega + \beta(L)r_t, \quad (13)$$

where  $\beta(L) = \sum_{i=1}^{\infty} \beta_i L^i$  is an infinite order lag polynomial the coefficients of which have asymptotically the following hyperbolic rate of decay

$$\beta_j = O(j^{d-1}), \quad (14)$$

for some  $d \in (0, 1/2)$ . For instance, the lag polynomial  $\beta(L)$  can be the one of the moving average form of a FARIMA process.

Giraitis, Robinson and Surgailis (2000) proved the following theorem:

**Theorem 1.** *Suppose  $E\varepsilon_0^4 < \infty$  and  $L(E\varepsilon_0^4)^{1/2} \sum_{j=1}^{\infty} \beta_j^2 < 1$ , where  $L = 7$  if the  $\{\varepsilon_k\}$  is Gaussian and  $L = 11$  in the other cases. Then, there is a stationary solution to equations (13) and (14) given by orthogonal Volterra series*

$$r_t = \sigma_t \varepsilon_t, \quad \sigma_t = \omega \sum_{l=0}^{\infty} \sum_{j_1, \dots, j_l=1}^{\infty} \beta_{j_1} \cdots \beta_{j_l} \varepsilon_{t-j_1} \cdots \varepsilon_{t-j_1-\dots-j_l}. \quad (15)$$

The sequence  $\{r_t^2\}$  is covariance stationary and

$$\text{Cov}(r_0^2, r_k^2) \asymp k^{2d-1}. \quad (16)$$

Setting  $\alpha = 2d$ , and  $Y_t = r_t^2$ , the ACF of the squared returns process has the hyperbolic decay of a long-memory process given by equation (2). See the chapter by Giraitis, Leipus and Surgailis (2005) in this volume for further details. Giraitis, Kokoszka, Leipus and Teyssière (2000) considered “pox-plot” based estimators and the local Whittle estimator for the estimation of the scaling parameter  $\alpha$ . We consider in section 4.1 the estimation of the scaling parameter with wavelet methods.

• **The Long-Memory Stochastic Volatility Process**

The long-memory stochastic volatility process, proposed by Breidt *et al.* (1998), is defined as

$$r_t = \sigma_t \zeta_t, \quad \zeta_t \sim N(0, 1), \quad (17)$$

$$\sigma_t = \sigma \exp(X_t/2), \quad X_t \sim \text{FARIMA}(p, d, q), \quad (18)$$

where  $\sigma$  is a scale parameter, the processes  $\{X_t\}$  and  $\{\zeta_t\}$  are independent. The process  $\{r_t^2\}$  is linearized as follows

$$\log r_t^2 = \log \sigma^2 + E(\log \zeta_t^2) + X_t + (\log \zeta_t^2 - E(\log \zeta_t^2)) = \mu + X_t + \varepsilon_t, \quad (19)$$

where  $\{\varepsilon_t\}$  is i.i.d,  $E(\varepsilon_t) = 0$  and  $\text{Var}(\varepsilon_t) = \sigma_\varepsilon^2$ . Since  $\zeta \sim N(0, 1)$ , then  $\log \zeta^2$  is distributed as the logarithm of a  $\chi_1^2$  random variable, thus  $E(\log \zeta^2) = 1.27$  and  $\sigma_\varepsilon^2 = \pi^2/2$ .

The spectral density of the process  $\{\log r_t^2\}$  is given by:

$$f(\lambda) = \frac{\sigma_\varepsilon^2}{2\pi} \frac{|\theta(\exp(-i\lambda))|^2}{|\phi(\exp(-i\lambda))|^2} |1 - \exp(-i\lambda)|^{-2d} + \frac{\sigma_\varepsilon^2}{2\pi}, \quad \lambda \in (-\pi, \pi), \quad (20)$$

where  $\sigma_\varepsilon^2$  denotes the variance of the innovations of the FARIMA process  $\{X_t\}$ . Thus, the spectral density of the process  $\{\log r_t^2\}$  is the sum of the spectral density of a FARIMA( $p, d, q$ ) process and the spectral density of a white noise process. Setting  $\alpha = 2d$  and  $Y_t = \log r_t^2$ , then the spectrum  $f_Y(\lambda)$  given by equation (20) satisfies equation (1).

Deo and Hurvich (2003) studied the estimation of the long-memory parameter using the semiparametric log-periodogram (LP) estimator in the spectral domain by Geweke and Porter-Hudak (1983). In section 4.2, we compare the estimations of the long-memory parameter obtained from the wavelet and LP estimators using Henry's (2001) feasible optimal bandwidths.

#### • The Nonlinear Moving Average Process

The nonlinear moving average process, proposed by Robinson and Zaffaroni (1997), is an extension of Robinson's (1977) short range dependent nonlinear moving average process:

$$r_t = \mu + \sigma_{t-1}\varepsilon_t, \quad \sigma_{t-1} = \rho + \sum_{i=1}^{\infty} \beta_i \varepsilon_{t-i}, \quad \sum_{i=1}^{\infty} \beta_i^2 < \infty, \quad (21)$$

where the innovation process  $\{\varepsilon_t\}$  is i.i.d and satisfy the following conditions:

$$\begin{aligned} E(\varepsilon_t) &= E(\varepsilon_t^3) = 0, \\ E(\varepsilon_t^2) &= \sigma^2, \quad 0 < \sigma^2 < \infty, \\ E(\varepsilon_t^4) &= \kappa + 3\sigma^4, \end{aligned} \quad (22)$$

where  $\kappa$  is the fourth cumulant of the process  $\{\varepsilon_t\}$ . If the process  $\{\varepsilon_t\}$  is Gaussian,  $\kappa = 0$ .

The autocorrelation function of the process  $\{r_t^2\}$  is given by

$$\begin{aligned} \gamma_{r^2}(k) &= 2\sigma^8 \beta_{|k|}^2 \delta_{\beta\beta}(0) + 2\sigma^8 \delta_{\beta\beta}(k) + 4\rho^2 \sigma^6 \delta_{\beta\beta}(k) \\ &\quad + \sigma^4 \kappa \left( \beta_{|k|}^2 \delta_{\beta\beta}(0) + \delta_{\beta^2\beta^2}(k) \right) + 4\sigma^4 \mu \beta_{|k|} (\rho^2 + \sigma^2 \delta_{\beta\beta}(k)) \\ &\quad + 2\rho^2 \sigma^6 \beta_{|k|}^2 + \nu_y \delta(k, 0), \quad k = 0, \pm 1, \dots, \end{aligned} \quad (23)$$

where  $\nu_r$  is a strictly positive constant,  $\delta_{\beta\varrho} = \sum_{i=1}^{\infty} \beta_i \varrho_{i+u}$ ,  $u = 0, \pm 1, \dots$  for any square summable sequence  $\{\beta_i\}$   $\{\varrho_i\}$ ,  $\delta(\cdot, \cdot)$  is the Kronecker delta.



When the sequence  $\{\beta_i\}$  of the MA form of equation (21) verifies the convergence rate given by equation (14) and  $\rho \neq 0$ , setting  $\alpha = 2d$  and  $Y_t = r_t^2$ , then the autocorrelation function has the rate of decay given by equation (2).

**2.4 Nonlinear Transformations of Fractionally Integrated Processes**

Another approach for modeling non-Gaussian LRD processes consists in taking nonlinear transformations of Gaussian LRD processes satisfying (2); see Surgailis (1981), the third chapter by Beran (1994), Giraitis and Surgailis (1985, 2005) and references therein for further details.

We consider here the case of a Gaussian process  $\{X_t\}$  and its transformation by a function  $G(\cdot)$  such that  $EG^2(X_t) < \infty$ , then this function can be expanded in series of Hermite polynomials:

$$G(x) = \sum_{k=0}^{\infty} \frac{c_k}{k!} H_k(x), \tag{24}$$

where

$$c_k = E[G(X)H_k(X)] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} G(x)H_k(x)e^{-x^2/2}dx, \tag{25}$$

so that  $H_0(x) = 1$ ,  $H_1(x) = x$ ,  $H_2(x) = x^2 - 1$ ,  $H_3(x) = x^3 - 3x$ ,  $H_k(x) = (-1)^k e^{(x^2/2)}(e^{-x^2/2})^{(k)}$ ,  $k \geq 0$ .

If  $\{X_t\}$  is a Gaussian LRD process with scaling parameter  $\alpha$ , satisfying equation (2), then the expansion

$$Y_t = G(X_t) = \sum_{J=1}^{\infty} \frac{c_J}{J!} H_J(X_t), \tag{26}$$

has the following properties:

- The processes  $\{H_J(X_t), t \in \mathbb{Z}\}$  are orthogonal,
- For  $1 \leq J \leq 1/(1 - \alpha)$ , the process  $\{H_J(X_t)\}$  satisfies equation (2) as

$$\text{Cov}(H_J(X_0), H_J(X_k)) = J! \text{Cov}(X_0, X_k)^J \asymp J! c_\gamma^J k^{J(\alpha-1)}, \tag{27}$$

thus, the memory parameter of  $\{H_J(X_t), t \in \mathbb{Z}\}$  decreases with  $J$ .

The index of the lowest nonzero coefficient  $c_k$ ,  $k \geq 1$ , of the transformation  $G(\cdot)$  is called the Hermite rank of the transformation, denoted by  $R$ . For a transformation  $G(X_t)$  of rank 1,

$$Y_t = G(X_t) = c_1 H_1(X_t) + \sum_{J \geq 2} c_J H_J(X_t)/J! = c_1 X_t + S_t, \tag{28}$$

the linear term  $c_1 X_t$  dominates the memory of the nonlinear term  $S_t$ . As a consequence,  $Y_t$  has the same scaling parameter of  $X_t$ : the transformation  $G(X_t)$  does not increase the memory of  $X_t$ .

The intensity of long-range dependence of the nonlinear transformed process  $\{Y_t\}$  coincides with the scaling parameter of the lowest nonzero term  $k \geq 1$ , as for a transformation  $G(X_t)$  of Hermite rank  $R$ ,

$$Y_t = G(X_t) = c_R H_R(X_t) + \sum_{J \geq R+1} c_J H_J(X_t)/J!, \quad (29)$$

the memory of the leading term  $c_R H_R(X_t)$  dominates the memory of the remainder term  $\sum_{J \geq R+1} c_J H_J(X_t)/J!$ .

Dittman and Granger (2002) considered the particular case  $X_t \sim \text{FI}(d)$ , then  $\{Y_t = G(X_t), t \in \mathbb{Z}\}$  is a long-memory process with long-memory parameter  $\tilde{d} = \max\{0, (d-0.5)R+0.5\}$ . Here,  $\alpha = 2\tilde{d}$ . We consider in section 4.4 the estimation of the scaling parameter for some nonlinear transformations of  $\text{FI}(d)$  processes.

### 3 Estimation of the Long-Memory Parameter

We compare the statistical performance of the wavelet estimator with the two standard semiparametric estimators in the spectral domain used in the statistical and econometric literature for estimating the scaling parameter  $\alpha$  of the processes introduced in section 2 and of real data, namely the local Whittle estimator and the LP estimator. The remainder of this section recalls the definitions and compare the key properties of the wavelet, local Whittle and LP estimators (hereafter denoted by  $\hat{\alpha}_W$ ,  $\hat{\alpha}_{LW}$  and  $\hat{\alpha}_{LP}$ ).

*Remark 1.* All three estimators are sharing similar features:

- They are based on the same semiparametric hypothesis: in the neighborhood of the zero frequency, the spectrum of the process under investigation satisfies equation (1).
- They imply the estimation of a second-order statistical quantity from the observed times series (the periodogram  $I_Y(\lambda_j)$  or the scalogram  $S_Y(j)$ ).
- They involve the choice of a range of *frequencies* or *scales* over which the estimation is to be performed. This crucial step will be addressed with care as this is one of the main practical issue that strongly controls the actual performance of these estimators.
- For the LP and LW estimators in the spectral domain, we will select the range of frequencies using plug-in methods; see Henry (2001), Moulines and Soulier (2003) for a detailed presentation. For the wavelet estimator, we will derive the lowest octave using the optimal bandwidth provided by these plug-in methods.

### 3.1 Local Whittle Estimator

• **Definition.** The local Whittle estimator  $\hat{\alpha}_{LW}$  has been proposed by Künsch (1987) and further developed by Robinson (1995a). It is defined as

$$\hat{\alpha}_{LW} = \arg \min_{\alpha} G(\alpha, m) := \left\{ \ln \left( \frac{1}{m} \sum_{j=1}^m \frac{I_Y(\lambda_j)}{\lambda_j^{-\alpha}} \right) - \frac{\alpha}{m} \sum_{j=1}^m \ln(\lambda_j) \right\}, \quad (30)$$

where  $I_Y(\lambda_j)$  is the periodogram evaluated on a set of  $m$  Fourier frequencies  $\lambda_j = 2\pi j/T, j = 1, \dots, m \ll [T/2]$ , where  $[\cdot]$  denotes the integer part, the bandwidth parameter  $m$  tends to infinity with the sample size  $T$  but more slowly since  $1/m + m/T \rightarrow 0$  as  $T \rightarrow \infty$ . The process does not need to be Gaussian, but its spectrum is differentiable near the zero frequency, and the process has a moving average representation.

• **Performance.** The LW estimator has the following asymptotic distribution

$$\sqrt{m}(\hat{\alpha}_{LW} - \alpha) \xrightarrow{D} N(0, 1), \quad (31)$$

where  $\xrightarrow{D}$  means convergence in distribution. The normality result comes from the assumption that the spectrum

$$f_Y(\lambda) = C\lambda^{-\alpha} [1 + E_{\beta}(\alpha)\lambda^{\beta} + o(\lambda^{\beta})], \quad 0 < |E_{\beta}(\alpha)| < \infty, \quad (32)$$

as  $\lambda \rightarrow 0^+$ , with  $\beta \in (0, 2]$  controls the smoothness of the spectrum near the zero frequency,  $C$  is a strictly positive constant, and the bandwidth  $m$  satisfying

$$\frac{1}{m} + \frac{m^{2\beta+1} \log^2 m}{T^{2\beta}} \rightarrow 0, \quad T \rightarrow \infty, \quad (33)$$

see Robinson (1995a) and Henry (2001) for further details.

*Remark 2.* In this framework, Giraitis, Robinson and Samarov (1997) have demonstrated that the best attainable rate for an estimator  $\hat{\alpha}$  is  $T^{-r(\beta)}$ , with  $r(\beta) = \beta/(1 + 2\beta) < 1/2$ , thus slower than the rate  $T^{-1/2}$  of parametric estimators. For the LW estimator, the rate of convergence is  $T^{-r(\beta)} M_n$  where  $(\log m)^{-1/(1+2\beta)} M_n = o(1)$ .

• **Range of Frequencies.** In this framework, i.e., when  $f_Y(\lambda)$  satisfies (32) the optimal bandwidth in the sense of minimization of the mean square error (MSE) has been provided by Henry and Robinson (1996) as

$$m_{LW}^{opt} = \left[ \frac{(\beta + 1)^4}{2\beta^3 E_{\beta}(\alpha)^2 (2\pi)^{2\beta}} \right]^{1/(1+2\beta)} T^{2\beta/(1+2\beta)}. \quad (34)$$

With the additional assumption that  $\beta = 2$ , i.e., the smoothest case for the spectrum, we obtain the optimal bandwidth  $m_{LW}^{opt}$  with the iterative procedure:

$$\hat{\alpha}^{(k)} = \arg \min_{\alpha} G(\alpha, m^{(k)}), \quad (35)$$

$$m_{LW}^{opt(k+1)} = \left( \frac{3T}{4\pi} \right)^{4/5} \left| E_2(\alpha^{(k)}) \right|^{-2/5}, \quad (36)$$

this iterative procedure starting at  $m^{(0)} = T^{4/5}$ , with  $\alpha \in (0, 1)$ . Without additional knowledge on  $E_2(\alpha^{(k)})$ , this procedure is infeasible. If we further assume the following specification for the spectrum

$$f_Y(\lambda) = |2 \sin(\lambda/2)|^{-\alpha} f_{\star}(\lambda), \quad (37)$$

where  $f_{\star}(\cdot)$  is twice continuously differentiable and positive at  $\lambda = 0$ , then a feasible optimal bandwidth is obtained with this approximation (see Delgado and Robinson, 1996):

$$E_2(\alpha^{(k)}) = \frac{f_{\star}''(0)}{2f_{\star}(0)} + \frac{\alpha}{24} \quad (38)$$

where  $f_{\star}(0)$  and its second derivative  $f_{\star}''(0)$  are the first and last coefficients of the regression of the periodogram  $I(\lambda_j)$  against  $|1 - \exp(i\lambda)|^{-\alpha(0)}(1, \lambda_j, \lambda_j^2/2)$  for the range of Fourier frequencies  $\lambda_j$  for  $j = 1, \dots, m^{(0)}$ . This iterative procedure defined by equations (35) and (36) converges very quickly.

*Remark 3.* The choice for this optimal bandwidth is motivated by two reasons:

- This bandwidth, the theory of which has been developed for linear LRD processes, is also robust to the presence of (long-memory) conditional heteroscedasticity in the process; see Henry (2001).
- This bandwidth works well, even for nonlinear LRD processes, such as the LARCH process; see Giraitis, Kokoszka, Leipus and Teyssière (2000).

### 3.2 Log-Periodogram Regression

• **Definition.** Geweke and Porter-Hudak (1983) proposed an estimator  $\hat{\alpha}_{LP}$  for  $\alpha$  that consists in performing a least squares regression on the log-periodogram (LP) over a range of frequencies  $\lambda_j$ :

$$\log I_Y(\lambda_j) = \log c_f - \alpha \log \lambda_j + e_j, \quad j = 1, \dots, m, \quad (39)$$

where the sequence  $\{e_j\}$  is interpreted as error terms; see Henry (2005) in this volume for more details.

• **Performance.** The LP estimator  $\hat{\alpha}_{LP}$  has the following asymptotic distribution

$$\sqrt{m}(\hat{\alpha}_{LP} - \alpha) \xrightarrow{D} N\left(0, \frac{\pi^2}{6}\right). \quad (40)$$

*Remark 4.* This central limit theorem requires that the bandwidth  $m$  satisfies equation (33). As mentioned in the remark 2 above, this estimator has the slow rate of convergence  $m^{1/2}$ . The best rate of convergence  $T^{-r(\beta)}$ , with  $r(\beta) = \beta/(1 + 2\beta)$  is attained with Robinson's (1995b) version of the LP estimator.

• **Regression Range.** Usually, to select  $m$ , one chooses  $m = [T^{4/5}]$  for long-memory processes in the conditional mean, e.g., FARIMA processes or nonlinear transformations of FI( $d$ ) processes, see e.g., Dittman and Granger (2002), and  $m = [T^{1/3}]$  for the LMSV process, see e.g., Deo and Hurvich (2003). For the LMSV process, this usual choice yields estimates with a slightly lower bias than the one obtained with  $m = [T^{4/5}]$ , while the Root Mean Squared Error of the estimator is increased by a factor equal to 2. For other nonlinear processes, like the ones considered in this chapter, this choice yields strongly biased and then unreliable estimates. Under the restriction that the spectrum has the specification given by equation (37), Hurvich *et al.* (1998) proposed an optimal bandwidth  $m_{LP}^{opt}$  minimizing the MSE for the LP estimator

$$m_{LP}^{opt} = \left[ \frac{27(2f_{\star}(0))^2}{512\pi^2(f_{\star}''(0))^2} \right]^{1/5} T^{4/5}. \quad (41)$$

A feasible bandwidth is obtained by estimating  $f_{\star}(0)$  and  $f_{\star}''(0)$  as above for the case of the optimal bandwidth for the LW estimator; see the useful paper by Henry (2001) for further details.

*Remark 5.* The choice  $m = cT^{r(\beta)}$  gives the best rate, and with a suitable choice of  $c$  minimizes the MSE. However, this choice does not imply the central limit theorem of equation (40).

*Remark 6.* Unlike for the optimal bandwidth of the LW estimator, the properties of robustness of  $\hat{\alpha}_{LP}$  to heteroskedasticity have to be established, so that we tend to prefer  $\hat{\alpha}_{LW}$ .

### 3.3 Wavelet Based Estimator

#### A Short Introduction to Wavelet Analysis

For complete and thorough introductions to wavelet analysis and decompositions, the reader is referred to e.g., the books by Daubechies (1992) and Mallat (1998).

#### • Mother–Wavelet.

**Definition 2.** A wavelet is a function  $\psi(\cdot)$  defined on  $\mathbb{R}$  such that

$$\int_{\mathbb{R}} \psi(t) dt = 0, \quad (42)$$

*i.e., satisfies the admissibility condition.*

We also assume that  $\psi(t)$  satisfies some integrability condition, i.e.,  $\psi(t) \in L^1(\mathbb{R}) \cap L^2(\mathbb{R})$ . The wavelet function is then a band-pass function, i.e., a small “wave” the support of which is almost limited in both the time and frequency domains. To perform a wavelet analysis, one chooses a reference wavelet, called the *mother-wavelet*, hereafter denoted  $\psi_0$ .

• **Wavelet-Basis.** From this  $\psi_0$ , an entire family of wavelets is designed using two operators:

1. A time-shift operator:  $(\mathcal{T}_\tau \psi_0)(t) = \psi_0(t - \tau)$ ,
2. A dilation (or change of scale) operator:  $(\mathcal{D}_a \psi_0)(t) = \frac{1}{\sqrt{a}} \psi_0\left(\frac{t}{a}\right)$ .

For the particular purpose of this work, we consider here only a particular type of wavelet decomposition, the so-called *Discrete Wavelet Transform* (DWT). From the time-shift and dilation operators above, the specific collection of translated and dilated templates of  $\psi_0$ , defined as

$$\left\{ \psi_{j,k} = 2^{-j/2} \psi_0(2^{-j}t - k), \quad j \in \mathbb{Z}, k \in \mathbb{Z} \right\}, \quad (43)$$

forms (a possibly orthonormal) basis of  $L^2(\mathbb{R})$ .

• **Wavelets Coefficients of the Discrete Wavelet Transform.** The wavelet coefficients of the DWT for an analyzed process or function  $Y_t$  are labeled  $d_Y(j, k)$  and defined as

$$d_Y(j, k) = \langle Y, \psi_{j,k} \rangle = \int_{\mathbb{R}} Y(t) \psi_{j,k}(t) dt, \quad (44)$$

where  $\psi_{j,k}$  is the wavelet basis defined in (43). It is worth noting that the  $d_Y(j, k)$ s can be computed at a very low cost (of the order of that of a FFT) from a recursive pyramidal algorithm on the condition that the mother-wavelet  $\psi_0$  is chosen to belong to a *Multiresolution analysis*; see e.g., Daubechies (1992) or Mallat (1998).

• **Number of Vanishing Moments and Polynomial Trends.** The mother wavelet  $\psi_0$  is further characterized by an integer  $N$ , called the number of vanishing, or zero, moments and defined as:

$$\int_{\mathbb{R}} t^k \psi_0(t) dt \equiv 0, \quad \forall k = 0, \dots, N - 1. \quad (45)$$

Obviously, from the admissibility condition, equation (42) above, one has  $N \geq 1$ . This integer  $N$  constitutes a key degree of freedom in wavelet analysis that can be freely chosen by the user and tuned to a given purpose. We will show that the use of this degree of freedom plays a central role in the analysis of long-range dependent processes as well as in the discrimination between genuine strong dependence, and non-stationarities such as those existing in change-points processes or trended processes. Indeed, by definition of  $N$ , the

wavelets coefficients of any polynomial of order  $P < N$  will be strictly null  $d_P(j, k) \equiv 0$ . By linearity of the wavelet transform, it implies that wavelet coefficients are identical for the processes  $\{Y_t\}$  and  $\{Y_t + P_t\}$ . This implies, for instance, that when  $Y_t$  is a zero-mean long-range dependent process, the potential superposition of a non-stationary polynomial mean will not affect the measure of the long-memory parameter as long as the degree of the polynomial remain strictly lower than the number of vanishing moment, that can be varied by the user. It has been shown elsewhere, see e.g., Abry and Veitch (1999) or Veitch and Abry (2001), that increasing  $N$  reduces the impact of the superimposed trend even if it is not strictly a polynomial one. This will be further discussed in Section 5. The counterpart to the increase of  $N$  however usually lies in the corresponding increase in the time support of the wavelet, an important practical drawback as discussed below.

### Wavelet Analysis of Long-Range Dependent Processes

Let  $Y_t$  denote a second order stationary random process. It can be shown that its wavelet coefficients constitute a zero-mean random field that satisfies:

$$Ed_Y(j, k)d_Y(j', k') = \int_{\mathbb{R}} f_Y(\lambda)2^{j/2}\Psi_0(2^j\lambda)2^{j'/2}\Psi_0^*(2^{j'}\lambda)\exp(-i2\pi(2^jk - 2^{j'}k'))d\lambda, \quad (46)$$

where  $\Psi_0(\lambda)$  denotes the Fourier transform of the mother wavelet  $\psi_0$ .

From this general result, it can be shown that when  $Y_t$  is a second order stationary long-range dependent process, the coefficients  $d_Y(j, k)$  possess the two following key properties:

**P1:** The process  $\{d_Y(j, k), k \in \mathbb{Z}\}$  is stationary if  $N \geq (\alpha - 1)/2$ , and its variance reproduces in the limit of large scales the power law behavior of the spectrum of  $Y_t$ ,

$$2^j \rightarrow +\infty, \quad Ed_Y(j, \cdot)^2 \simeq 2^{j\alpha}c_f C(\alpha, \psi_0), \quad C(\alpha, \psi_0) = \int |\lambda|^{-\alpha}|\Psi_0(\lambda)|^2d\lambda, \quad (47)$$

**P2:** The process  $\{d_Y(j, k), k \in \mathbb{Z}\}$  is stationary and short-range dependent, if  $N > \alpha/2$ . The residual correlations between the elements of the sequence  $\{d_Y(j, k), k \in \mathbb{Z}\}$  is in reverse relationship with  $N$ ; see Flandrin (1989):

$$Ed_Y(j, k)d_Y(j, k') \approx C |k - k'|^{\alpha-1-2N}, \quad |k - k'| \rightarrow \infty. \quad (48)$$

Properties **P1** and **P2** above constitute the main rationale for the wavelet analysis of long-range dependent data. Indeed, **P1** yields

$$\log_2 (Ed_Y(j, \cdot)^2) = j\alpha + \log_2(c_f C(\alpha, \psi_0)). \quad (49)$$

This invites to perform a linear regression in a log-log plot. However, the expectation  $Ed_Y(j, \cdot)^2$  needs to be estimated from a (single, finite duration) observation. This is where **P2** plays a key role: the ensemble average can be efficiently replaced with the time average  $1/n_j \sum_{k=1}^{n_j} d_Y(j, k)^2$ .

### Estimation of the Long–Memory Parameter

This estimator was first proposed and studied in Abry *et al.* (1995) and Abry and Veitch (1998) and further developed and analyzed in Veitch and Abry (1999).

• **Definition.** Let  $S_Y(j) = 1/n_j \sum_{k=1}^{n_j} d_Y(j, k)^2$ , where  $n_j$  denotes the number of wavelet coefficients  $d_Y(j, k)$  available at octave  $j$ . Roughly,  $n_j$  varies as  $2^{-j}T$ , where  $T$  denotes the number of observed samples. The wavelet-based estimator consists of a (weighted) linear regression in the so-called logscale diagram:  $\log_2 S_Y(j)$  versus  $\log_2 2^j = j$ . Precisely, it is defined as:

$$\hat{\alpha}_W = \sum_{j_1}^{j_2} w_j (\log_2 S_Y(j) - (\psi(n_j/2)/\log 2 - \log_2(n_j/2))), \quad (50)$$

where  $\psi(z) := \Gamma'(z)/\Gamma(z)$  is the Psi function and the weights are chosen to satisfy the two constraints:

$$\sum_{j_1}^{j_2} j w_j \equiv 1, \quad \sum_{j_1}^{j_2} w_j \equiv 0. \quad (51)$$

A natural form for the  $w_j$ s reads:

$$w_j = \frac{1}{a_j} \frac{S_0 j - S_1}{S_0 S_2 - S_1^2}, \quad (52)$$

where  $S_p = \sum_{j_1}^{j_2} j^p / a_j$ ,  $p = 0, 1, 2$ . The ranges of octaves  $j \in [j_1, j_2]$  is assumed to be a priori chosen; this will be discussed below. The  $a_j$ s are arbitrary coefficients used to weight the estimation according to the confidence given to  $\log_2 S(j)$ . Precisely, for the estimation of the LRD parameter, we chose  $a_j \equiv \zeta(2, n_j/2)$ , with  $\zeta(2, z) := \sum_{n=0}^{\infty} 1/(z+n)^2$  a generalized Riemann Zeta function, that provides us with an approximation of  $\text{Var} \log_2 S(j)$ .  $\zeta(2, n_j/2)$  turns out to be asymptotically proportional to  $1/n_j$  (see below). It implies that the larger  $j$ , the smaller  $n_j$ , the less  $\log_2 S(j)$  contributes to the estimate of  $\alpha$ .

• **Performance.** To study approximately but analytically the performance of this estimator, the three following assumptions are assumed to hold:

- H1:** For each  $j$ , the sequences  $\{d_Y(j, \cdot)\}$  are stationary and uncorrelated,
- H2:** The processes  $\{d_Y(j, \cdot)\}$  and  $\{d_Y(j', \cdot)\}$ ,  $j \neq j'$  are uncorrelated,
- H3:** The processes  $\{d_Y(j, \cdot)\}$ ,  $j \in [j_1, j_2]$ , are Gaussian.

The two first points constitute idealizations of the decorrelation property **P2**. The third point is obviously satisfied when  $\{Y_t\}$  is a Gaussian process.

Using these assumptions, it can be shown that



$$E \log_2 S(j) = \log_2 E d_Y(j, \cdot)^2 + (\psi(n_j/2)/\log 2 - \log_2(n_j/2)), \quad (53)$$

$$\text{Var} \log_2 S(j) = \zeta(2, n_j/2)/\log^2 2. \quad (54)$$

From this, one shows that

$$E \hat{\alpha}_W = \alpha, \quad (55)$$

$$\text{Var} \hat{\alpha}_W = \frac{(1 - 2^{-J})/\log^2 2}{1 - 2^{-(J+1)}(J^2 + 4) + 2^{-2J}} \frac{1}{T}, \quad (56)$$

where  $J = j_2 - j_1 + 1$  is the width of the regression range. Veitch and Abry (1999) conjectured the asymptotic approximation of the wavelet estimator

$$\sqrt{T}(\hat{\alpha}_W - \alpha) \approx N\left(0, \frac{1}{\ln^2(2)2^{1-j_1}}\right), \quad (57)$$

where  $j_1$  is the lowest octave, the long-memory behavior being captured by the octaves larger than  $j_1$ .

This analytical yet only approximative performance turns out to be extremely close to the actual ones, even when  $Y_t$  are non Gaussian processes. In practice,  $\hat{\alpha}_W$  has a negligible bias and presents a Gaussian statistic with a variance (approximately) known without the estimation of any quantity. An effective confidence interval can hence be constructed.

Under some assumptions, i.e.,  $n_{j_1} \rightarrow \infty$ , but more slowly than the sample size  $T$ , Bardet, Lang, Moulines and Soulier (2000) proved the asymptotically Normal distribution of this estimator

$$\sqrt{n_{j_1}}(\hat{\alpha}_W - \alpha) \xrightarrow{D} N(0, K_v(\alpha)), \quad (58)$$

where the expression of  $K_v(\alpha)$  depends on the unknown scaling parameter  $\alpha$ .

• **Regression Range.** Let us turn now to the choice of the regression range  $j \in [j_1, j_2]$ . One might first use the graphical representation of the logscale diagram, i.e., the graphical plot of the  $y_j$  against the octaves  $j$  from which is estimated the scaling parameter  $\hat{\alpha}_W$ . A graphical analysis might help to select the upper and lower octaves that have a too high leverage effect on the regression line; see Belsley *et al.* (1980).

Obviously, as LRD implies a power law behavior in the limit of large scales,  $2^j \rightarrow +\infty$ , the upper bound  $j_2$  for the octaves has to be chosen as large as allowed by the observation length. We chose

$$j_2 = [\log_2 T - \log_2(2N + 1)], \quad (59)$$

where  $[\cdot]$  denotes the integer part, so that borders effects do not affect the estimation of the variance. This is equivalent to the fact that the range of frequencies for the LP and LW estimators starts at  $\lambda_j = \pi/T$ .

The choice for  $j_1$ , i.e., the cutoff between short-range dependence and long-range dependence, is similar to the bandwidth selection problem in semi-parametric/nonparametric statistics and is exactly equivalent to the choice of

the  $m$  parameter for the local Whittle and log-periodogram estimators introduced above. This choice consists in optimizing a bias-variance trade-off: a too small  $j_1$  should induce bias as the regression range would include octaves with departures from the power law due to short-range dependencies, conversely, a too large  $j_1$  decreases the bias but implies an increase of the variance, as shown in relation (57). Abry *et al.* (2003) discussed the issue of the selection for the lower octave  $j_1$  in connexion with the bandwidth selection issue for the local Whittle (LW) estimator, which has been addressed by Robinson and Henry (1996) and Henry (2001, 2005); see equation (36). The frequency cutoff, associated with the optimal bandwidth  $m_{LW}^{opt}$  for the LW estimator, is then equal to  $m_{LW}^{opt}/T$ , and corresponds to the scale  $2^{-j_1}$ . Thus, using  $m_{LW}^{opt}$  we could define the optimal lower scale as

$$j_1^{opt} = \left\lceil \frac{\log T - \log m_{LW}^{opt}}{\log 2} \right\rceil. \quad (60)$$

A similar decision rule is possible using the feasible optimal bandwidth  $m_{LP}^{opt}$  for the log-periodogram (LP) estimator; see equation (39). Indeed, estimation results given by equation (60) with  $m_{LW}^{opt}$  and  $m_{LP}^{opt}$  are very close, with a slight advantage for  $m_{LW}^{opt}$  as the RMSE is slightly lower with this bandwidth.

In Veitch, Abry and Taquq (2003), it has been proposed that the choice of the lowest octave  $j_1$  might be guided using a goodness-of-fit function, such as the generalized Pearson statistic defined as

$$Q = \sum_{j=j_1}^{j_2} \frac{(y_j - \hat{\alpha}_W j - \hat{a})^2}{\sigma_j^2}, \quad (61)$$

where  $\hat{a}$  is the unbiased estimator for  $\log_2(\widehat{c_f C})$ , and  $\sigma_j^2 = \text{Var}(y_j)$ . Under the null hypothesis of Gaussian residuals, the statistic  $Q \sim \chi^2(J-2)$ , where  $J = j_2 - j_1 + 1$ . The statistic  $Q$  can be viewed as a function of  $j_1$ , denoted  $Q(j_1)$ . Studying the evolution of  $Q$  as a function of  $j_1$ , Veitch, Abry and Taquq (2003) proposed an empirical criterion to choose automatically  $j_1$ . It turns out that this procedure amounts for most cases to select the  $j_1$  that ensures the lowest RMSE, and hence the bias-variance trade-off. Bardet *et al.* (2000) provided an analytical expression for the choice of an asymptotically optimal  $j_1$  minimizing the MSE, and conjectured that a feasible approximation of this optimal bandwidth using estimates of  $\alpha$ ,  $f_\star(0)$  and  $f_\star''(0)$  as for the LW and LP estimators might be obtained. This procedure is beyond the scope of this chapter.

• **Choosing the Number of Vanishing Moments.** On one hand, equation (48) may lead to think that the larger  $N$ , the weaker the correlation amongst wavelet coefficients, however we saw that the LRD is turn in SRD as soon as  $N > \alpha/2$ . On the other hand, as already mentioned, increasing  $N$  induces an increase of the wavelet size and hence of the border effects. In turn,

this produces a decrease of  $j_2$ , as summarized in equation (59). Therefore, an optimal practical choice for the estimation of the parameter  $\alpha$  of true LRD processes is  $N = 2$  or  $N = 3$ , as will be seen on numerical simulations conducted in next sections. However, in disentangling LRD from non-stationarities, the possibility of varying  $N$  and using larger  $N$  is fruitful.

• **Discrete Time.** By definition, the DWT applies to continuous time processes, see equation (44). However, in most practical cases, only a collection of discrete samples  $\{y_k, k = 1, \dots, n\}$  is available. This specific difficulty was carefully addressed in Veitch, Taqqu and Abry (2001). Numerical simulations such as those proposed in the next two sections taking into account this point are under investigation; see Abry and Teysnière (2005).

#### 4 Estimation of the Long-Memory Parameter for Nonlinear Long-Range Dependent Processes

• **Protocol.** In this section, we compare by means of numerical simulations, the performance of the local Whittle, log-periodogram and wavelet estimators for the long-memory parameter. Because wavelet based estimation of the long-memory parameter for linear FARIMA processes has been considered in Veitch and Abry (1999), we consider in this section the estimation of the scaling parameter of the nonlinear processes described in section 2 .

For each process of interest, 5000 replications, with sample sizes  $T$  were numerically synthetised. In general, we used  $T = 10000$ , in some cases, we went up to  $T = 20000$ , for having a better idea on the asymptotic bias for some highly nonlinear processes. For each simulation, bias and Root Mean Squared Error (RMSE) defined respectively as  $B = \langle \hat{\alpha} \rangle - \alpha$  and  $\text{RMSE} = \sqrt{B^2 + \langle \langle \hat{\alpha} \rangle \rangle}$ , where  $\langle \cdot \rangle$  and  $\langle \langle \cdot \rangle \rangle$  denote respectively the sample mean and sample variance estimators computed from the 5000 replications.

For the wavelet-based estimator, we used the Daubechies wavelets with the number of vanishing moments ranging from  $N = 2, \dots, 10$ . We considered the range of lowest octaves  $j_1 = 1, \dots, 7$  and denoted by  $\hat{\alpha}_W^{(j_1)}$  the corresponding estimates. We computed the “optimal”  $j_1$  according to equation (60),  $\hat{\alpha}_W^m$  denotes the wavelet estimator with the lowest octave  $j_1$  derived from equation (60). We also chose the  $j_1$  and  $N$  that minimize the Root Mean Squared Error (RMSE) of  $\hat{\alpha}_W$ . We denote by  $j_1^{\text{RMSE}}$  and  $\hat{\alpha}_W^{\text{RMSE}}$  the corresponding  $j_1$  and estimate. Systematically, we compare this RMSE to that obtained with the choice  $j_1 = 6$  and  $N = 2$ . As we will see below, the RMSE for this choice of  $(j_1, N)$  is not too far from the minimum one and the bias is lower. On the basis of simulation results for  $T = 20000$ , we also conjecture that for some of these nonlinear processes we could get rid off the bias with very large samples, so that we could select a higher value for  $j_1$  which minimizes the RMSE. For each simulation, Tables are reporting the results and logscale diagrams are displayed.

For the local Whittle and the log-periodogram estimators, we respectively denote by  $\hat{\alpha}_{LW}^{opt}$  and  $\hat{\alpha}_{LP}^{opt}$  those obtained with the optimal bandwidth  $m^{opt}$  as in equations (36) and (41).

#### 4.1 The Long-Memory Linear ARCH Process

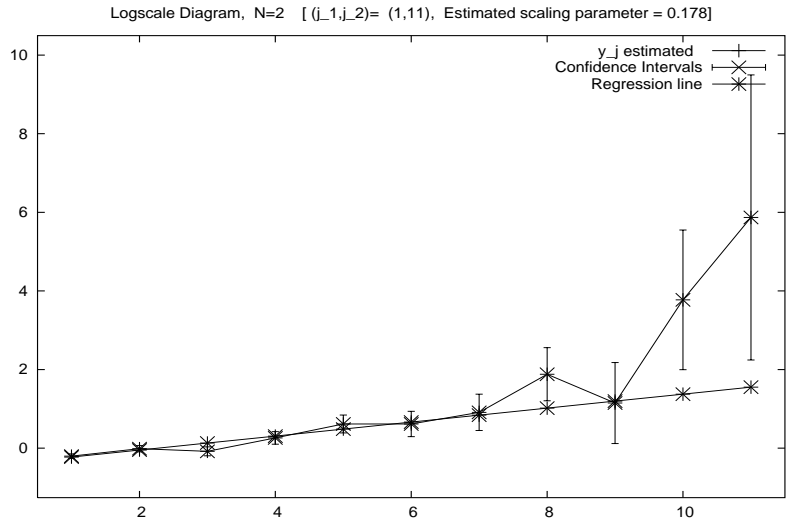
The long-memory LARCH process is defined as:

$$r_t = \sigma_t \varepsilon_t, \quad \varepsilon_t \sim D(0, 1), \quad \sigma_t = \omega + \beta(L)r_t, \quad (62)$$

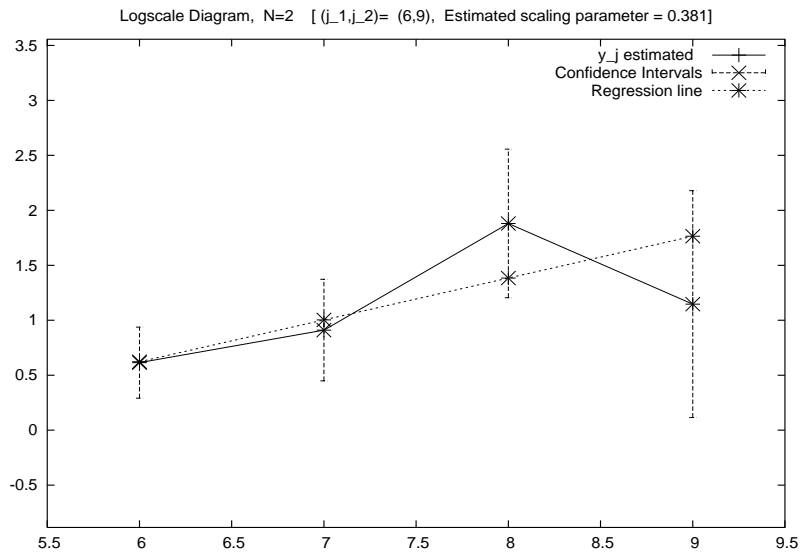
where the coefficients  $\beta_j$  of the lag polynomial  $\beta(L)$  have the rate of decay  $\beta_j = O(j^{\alpha/2-1})$ , for some  $0 < \alpha < 1$ . We choose here three parameterizations for  $\beta(L)$  from the moving average representation of a FARIMA(0,d,0), a FARIMA(1,d,0) and a FARIMA(0,d,1) processes, with  $\alpha = 2d$ , that we respectively denote as **LARCH A**, **LARCH B** and **LARCH C**. For **LARCH B** and **LARCH C**, we respectively set  $\phi = -0.20$  and  $\theta = 0.20$ .

We consider the following values for the scaling parameter  $\alpha = 0.10, 0.15, 0.20, \dots, 0.60$ . For values of  $\alpha$  over 0.60, the sequence of coefficients  $\{\beta_j\}$  must be largely rescaled for satisfying the stationarity condition of theorem 1, i.e.,  $L(E\varepsilon_0^4)^{1/2} \sum_{j=1} \beta_j^2 < 1$ . We estimate the scaling parameter using the wavelet estimator and the two spectral estimators LW and LP. We report in Table 2 the LP results only as they are very close to the LW estimator, although with a slightly higher RMSE as theoretically expected, and so far there are no results for the estimation results of the LARCH process with the LP estimator, while Giraitis, Kokoszka, Leipus and Teyssière (2000) already report estimation results for LARCH processes with the LW estimator for  $T = 3000, 6000$ . Tables 1 and 2 below report estimation results for the LARCH processes for  $T = 10,000$  while Tables 3 and 4 give the results for  $T = 20,000$ . For  $T = 20,000$ , estimation results for the LW estimator are reported on Table 4.

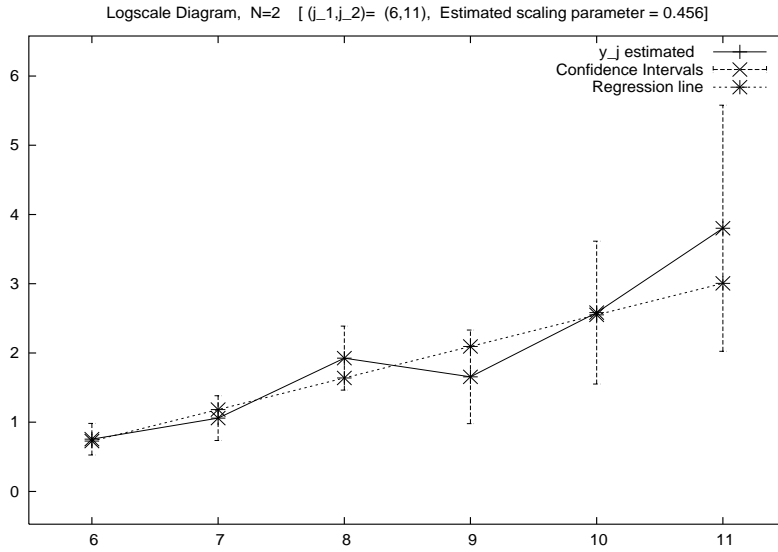
Figures 3 to 5 below display the logscale diagrams for **LARCH A** processes, with  $T = 10000$  and  $T = 20000$ . We can see that the “correct” LRD behavior is captured by the octaves  $j$  greater than 5. When  $T$  increases, the number of octaves that can be used for the wavelet regression is obviously larger, so that wavelet estimates are more reliable.



**Fig. 3.** Logscale diagram for the realization of a **LARCH A** process, with  $\alpha = 0.40$ ,  $T = 10000$ . We select here  $j_1 = 1$ ,  $j_2 = 11$ ,  $N = 2$



**Fig. 4.** Logscale diagram for the realization of a **LARCH A** process, with  $\alpha = 0.40$ ,  $T = 10000$ . We select here  $j_1 = 6$ ,  $j_2 = 9$ ,  $N = 2$



**Fig. 5.** Logscale diagram for the realization of a **LARCH A** process, with  $\alpha = 0.40$ ,  $T = 20000$ . We select here  $j_1 = 6$ ,  $j_2 = 11$ ,  $N = 2$

From the results in Tables 1 to 4, we can conclude that even for  $T = 10,000$ , both the LW and LP estimators perform slightly better than the wavelet estimator as both their bias and RMSE are slightly lower. For this sample size, choosing the wavelet estimator with  $j_1 = 6$  and  $N = 2$  reduces the bias, which becomes smaller than the one of the LW estimator, but increases the variance. However, for  $T = 20,000$ , the RMSE of the estimates for  $j_1 = 6$  is reduced and becomes closer to the one of the LW estimator. For  $T = 20,000$ ,  $j_1^{RMSE}$  is often equal to 5, the wavelet estimates are closer to the LW estimates, and we conjecture that for larger samples  $j_1 = 6$  would minimize the RMSE, so that the performance of the wavelet, LP and LW estimator would become very close.

We also note that the estimations for the wavelet estimator with  $j_1$  chosen according to equation (60) gives interesting results, not as good as those given by  $j_1^{RMSE}$ , but that can be used as a first approximation.

**Table 1.** Estimation of the scaling parameter for LARCH processes.  $T = 10000$ ,  $N = 2$ 

Model	$\alpha$	$\hat{\alpha}_W^{RMSE}$	$j_1^{RMSE}$	$E\hat{\alpha}_W^{RMSE} - \alpha$	RMSE	$\hat{\alpha}_W^{(6)}$	$E\hat{\alpha}_W^{(6)} - \alpha$	RMSE
LARCH A	0.10	0.0255	3	-0.0745	0.0824	0.0445	-0.0555	0.1350
	0.15	0.0690	4	-0.0810	0.0971	0.0966	-0.0534	0.1357
	0.20	0.1199	4	-0.0801	0.0980	0.1622	-0.0378	0.1330
	0.25	0.1785	4	-0.0715	0.0937	0.2325	-0.0175	0.1329
	0.30	0.2394	4	-0.0606	0.0891	0.3046	0.0046	0.1381
	0.35	0.3010	4	-0.0490	0.0891	0.3618	0.0118	0.1487
	0.40	0.3349	4	-0.0650	0.1028	0.3994	-0.0006	0.1549
	0.45	0.3880	5	-0.0620	0.1252	0.4228	-0.0272	0.1585
	0.50	0.4060	5	-0.0940	0.1442	0.4463	-0.0537	0.1663
	0.55	0.4234	5	-0.1266	0.1677	0.4697	-0.0803	0.1778
0.60	0.4928	6	-0.1072	0.1924	0.4928	-0.1072	0.1923	
LARCH B	0.10	0.0125	1	-0.0875	0.0898	0.0039	-0.0961	0.1554
	0.15	0.0128	2	-0.1372	0.1399	0.0314	-0.1185	0.1704
	0.20	0.0633	5	-0.1367	0.1581	0.0824	-0.1176	0.1706
	0.25	0.1222	5	-0.1278	0.1514	0.1531	-0.0969	0.1586
	0.30	0.1938	5	-0.1062	0.1354	0.2351	-0.0649	0.1441
	0.35	0.2712	5	-0.0788	0.1180	0.3190	-0.0310	0.1367
	0.40	0.3483	5	-0.0517	0.1069	0.3977	-0.0023	0.1397
	0.45	0.4202	5	-0.0298	0.1061	0.4669	0.0169	0.1504
	0.50	0.4249	4	-0.0751	0.1117	0.5216	0.0216	0.1640
	0.55	0.4901	5	-0.0599	0.1281	0.5375	-0.0125	0.1641
0.60	0.4987	5	-0.1013	0.1520	0.5530	-0.0470	0.1709	
LARCH C	0.10	0.1014	1	0.0014	0.0306	0.0725	-0.0275	0.1332
	0.15	0.1369	1	-0.0131	0.0375	0.1232	-0.0268	0.1382
	0.20	0.1477	1	-0.0523	0.0635	0.1641	-0.0359	0.1426
	0.25	0.1567	1	-0.0933	0.1004	0.2032	-0.0468	0.1479
	0.30	0.1979	4	-0.1021	0.1243	0.2406	-0.0594	0.1545
	0.35	0.2500	5	-0.1000	0.1413	0.2766	-0.0734	0.1624
	0.40	0.2803	5	-0.1197	0.1570	0.3114	-0.0886	0.1717
	0.45	0.3092	5	-0.1408	0.1745	0.3451	-0.1049	0.1824
	0.50	0.3368	5	-0.1632	0.1939	0.3777	-0.1223	0.1945
	0.55	0.4093	6	-0.1407	0.2081	0.4093	-0.1407	0.2081
0.60	0.4398	6	-0.1602	0.2230	0.4398	-0.1602	0.2230	

**Table 2.** Estimation of the scaling parameter for LARCH processes.  $T = 10000$ ,  $N = 2$ . The optimal bandwidth  $m_{LP}^{opt}$  is used for the LP estimator

Model	$\alpha$	$\hat{\alpha}_W^m$	$E\hat{\alpha}_W^m - \alpha$	RMSE	$\hat{\alpha}_{LP}^{opt}$	$E\hat{\alpha}_{LP}^{opt} - \alpha$	RMSE
LARCH A	0.10	0.0289	-0.0711	0.0847	0.0323	-0.0677	0.0821
	0.15	0.0654	-0.0845	0.0981	0.0778	-0.0722	0.0869
	0.20	0.1149	-0.0851	0.1011	0.1379	-0.0621	0.0826
	0.25	0.1756	-0.0744	0.0967	0.2085	-0.0415	0.0754
	0.30	0.2412	-0.0588	0.0934	0.2792	-0.0208	0.0749
	0.35	0.3097	-0.0403	0.0964	0.3482	-0.0018	0.0785
	0.40	0.3485	-0.0515	0.1090	0.3889	-0.0111	0.0902
	0.45	0.3672	-0.0827	0.1310	0.4138	-0.0362	0.1014
	0.50	0.4285	-0.0715	0.1317	0.4421	-0.0579	0.1089
	0.55	0.4081	-0.1419	0.1788	0.4668	-0.0831	0.1313
0.60	0.4284	-0.1715	0.2041	0.4935	-0.1064	0.1495	
LARCH B	0.10	0.0001	-0.0999	0.1102	-0.0019	-0.1019	0.1119
	0.15	0.0130	-0.1370	0.1448	0.0196	-0.1304	0.1379
	0.20	0.0415	-0.1585	0.1658	0.0605	-0.1395	0.1480
	0.25	0.0876	-0.1624	0.1711	0.1216	-0.1284	0.1408
	0.30	0.1507	-0.1493	0.1617	0.2042	-0.0958	0.1178
	0.35	0.2314	-0.1186	0.1417	0.2905	-0.0595	0.1015
	0.40	0.3228	-0.0772	0.1205	0.3803	-0.0197	0.0899
	0.45	0.4041	-0.0459	0.1127	0.4571	0.0071	0.0964
	0.50	0.4688	-0.0312	0.1153	0.5192	0.0192	0.1056
	0.55	0.4813	-0.0687	0.1318	0.5381	-0.0119	0.1075
0.60	0.4921	-0.1079	0.1565	0.5569	-0.0431	0.1178	
LARCH C	0.10	0.0640	-0.0360	0.0661	0.0631	-0.0339	0.0596
	0.15	0.1066	-0.0434	0.0751	0.1094	-0.0406	0.0643
	0.20	0.1380	-0.0620	0.0889	0.1473	-0.0527	0.0758
	0.25	0.1675	-0.0825	0.1063	0.1842	-0.0658	0.0875
	0.30	0.1971	-0.1029	0.1248	0.2206	-0.0794	0.1022
	0.35	0.2256	-0.1244	0.1459	0.2574	-0.0926	0.1160
	0.40	0.2542	-0.1458	0.1674	0.2940	-0.1060	0.1302
	0.45	0.2824	-0.1676	0.1893	0.3308	-0.1192	0.1440
	0.50	0.3113	-0.1887	0.2110	0.3662	-0.1338	0.1598
	0.55	0.3404	-0.2096	0.2327	0.4022	-0.1478	0.1736
0.60	0.3694	-0.2306	0.2541	0.4383	-0.1617	0.1874	



**Table 3.** Estimation of the scaling parameter for LARCH processes.  $T = 20000$ ,  $N = 2$ 

Model	$\alpha$	$\hat{\alpha}_W^{RMSE}$	$j_1^{RMSE}$	$E\hat{\alpha}_W^{RMSE} - \alpha$	RMSE	$\hat{\alpha}_W^{(6)}$	$E\hat{\alpha}_W^{(6)} - \alpha$	RMSE
LARCH A	0.10	0.0315	4	-0.0685	0.0767	0.0475	-0.0525	0.0925
	0.15	0.0858	5	-0.0642	0.0821	0.1022	-0.0478	0.0910
	0.20	0.1464	5	-0.0536	0.0757	0.1701	-0.0299	0.0850
	0.25	0.2132	5	-0.0368	0.0675	0.2419	-0.0081	0.0830
	0.30	0.2802	5	-0.0198	0.0642	0.3107	0.0107	0.0880
	0.35	0.3433	5	-0.0067	0.0680	0.3726	0.0226	0.0972
	0.40	0.3809	5	-0.0191	0.0746	0.4113	0.0113	0.1003
	0.45	0.4007	5	-0.0493	0.0880	0.4361	-0.0139	0.1018
	0.50	0.4202	5	-0.0798	0.1085	0.4612	-0.0388	0.1092
	0.55	0.4862	6	-0.0638	0.1213	0.4862	-0.0638	0.1213
0.60	0.5110	6	-0.0890	0.1371	0.5110	-0.0890	0.1371	
LARCH B	0.10	0.0120	1	-0.0880	0.0891	0.0065	-0.0935	0.1208
	0.15	0.0249	5	-0.1251	0.1347	0.0368	-0.1132	0.1368
	0.20	0.0913	6	-0.1087	0.1335	0.0913	-0.1086	0.1335
	0.25	0.1653	6	-0.0847	0.1157	0.1653	-0.0847	0.1157
	0.30	0.2491	6	-0.0509	0.0955	0.2491	-0.0509	0.0955
	0.35	0.3356	6	-0.0164	0.0855	0.3356	-0.0164	0.0855
	0.40	0.3612	5	-0.0388	0.0726	0.4120	0.0120	0.0895
	0.45	0.4337	5	-0.0163	0.0697	0.4808	0.0308	0.1009
	0.50	0.4943	5	-0.0057	0.0764	0.5357	0.0357	0.1122
	0.55	0.5059	5	-0.0441	0.0885	0.5533	0.0033	0.1074
0.60	0.5705	6	-0.0295	0.1121	0.5705	-0.0295	0.1121	
LARCH C	0.10	0.1019	1	0.0019	0.0218	0.0743	-0.0257	0.0851
	0.15	0.1435	1	-0.0065	0.0245	0.1270	-0.0230	0.0878
	0.20	0.1562	1	-0.0438	0.0504	0.1696	-0.0304	0.0915
	0.25	0.1910	5	-0.0590	0.0862	0.2104	-0.0396	0.0966
	0.30	0.2259	5	-0.0741	0.1243	0.2494	-0.0506	0.1033
	0.35	0.2869	6	-0.0631	0.1113	0.2869	-0.0631	0.1113
	0.40	0.3232	6	-0.0768	0.1210	0.3232	-0.0768	0.1210
	0.45	0.3585	6	-0.0915	0.1321	0.3584	-0.0915	0.1321
	0.50	0.3928	6	-0.1072	0.1446	0.3928	-0.1072	0.1446
	0.55	0.4261	6	-0.1239	0.1585	0.4261	-0.1239	0.1585
0.60	0.4583	6	-0.1717	0.1737	0.4583	-0.1417	0.1786	

**Table 4.** Estimation of the scaling parameter for LARCH processes.  $N = 2$ ,  $T = 20000$ . The optimal bandwidth  $m_{LW}^{opt}$  is used for the LW estimator

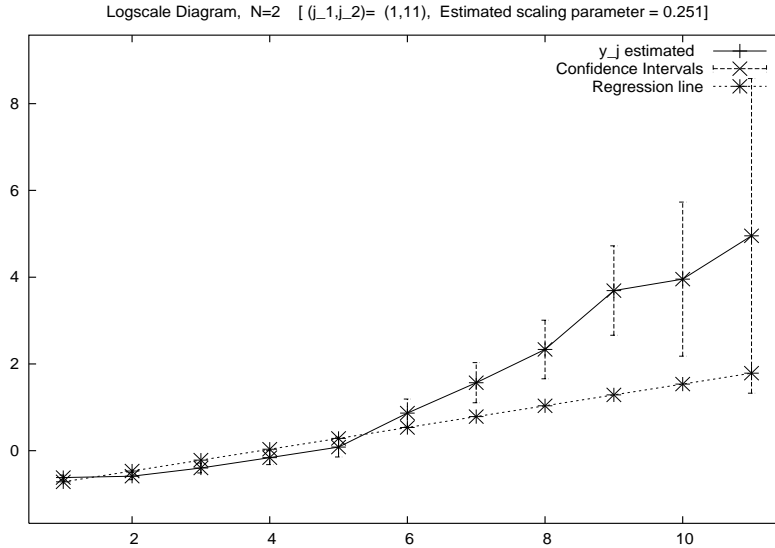
Model	$\alpha$	$\hat{\alpha}_W^m$	$E\hat{\alpha}_W^m - \alpha$	RMSE	$\hat{\alpha}_{LW}^{opt}$	$E\hat{\alpha}_{LW}^{opt} - \alpha$	RMSE
LARCH A	0.10	0.0299	-0.0701	0.0774	0.0349	-0.0651	0.0736
	0.15	0.0687	-0.0813	0.0888	0.0826	-0.0674	0.0771
	0.20	0.1239	-0.0761	0.0872	0.1470	-0.0530	0.0692
	0.25	0.1926	-0.0574	0.0780	0.2229	-0.0271	0.0584
	0.30	0.2678	-0.0322	0.0684	0.2990	-0.0010	0.0563
	0.35	0.3355	-0.0145	0.0694	0.3663	0.0163	0.0635
	0.40	0.3754	0.0246	0.0765	0.4086	0.0086	0.0659
	0.45	0.3969	-0.0531	0.0907	0.4372	-0.0128	0.0690
	0.50	0.4178	-0.0822	0.1108	0.4671	-0.0329	0.0776
	0.55	0.4381	-0.1119	0.1350	0.4975	-0.0525	0.0900
0.60	0.4573	-0.1427	0.1620	0.5285	-0.0715	0.1040	
LARCH B	0.10	0.0002	-0.0088	0.1050	-0.0003	-0.1003	0.1055
	0.15	0.0155	-0.1345	0.1385	0.0237	-0.1263	0.1308
	0.20	0.0471	-0.1529	0.1570	0.0686	-0.1314	0.1369
	0.25	0.0999	-0.1501	0.1564	0.1375	-0.1125	0.1226
	0.30	0.1800	-0.1200	0.1333	0.2305	-0.0695	0.0904
	0.35	0.2751	-0.0749	0.0972	0.3278	-0.0222	0.0647
	0.40	0.3592	-0.0408	0.0748	0.4141	0.0141	0.0652
	0.45	0.4326	-0.0174	0.0704	0.4862	0.0362	0.0782
	0.50	0.4931	-0.0069	0.0789	0.5429	0.0429	0.0789
	0.55	0.5057	-0.0433	0.0915	0.5652	0.0152	0.0790
0.60	0.5190	-0.0810	0.1160	0.5883	-0.0117	0.0806	
LARCH C	0.10	0.0650	-0.0350	0.0525	0.0650	-0.0350	0.0525
	0.15	0.1066	-0.0434	0.0751	0.1094	-0.0406	0.0640
	0.20	0.1425	-0.0575	0.0736	0.1523	-0.0477	0.0629
	0.25	0.1756	-0.0744	0.0898	0.1921	-0.0579	0.0739
	0.30	0.2094	-0.0906	0.1066	0.2321	-0.0679	0.0848
	0.35	0.2439	-0.1061	0.1229	0.2731	-0.0769	0.0952
	0.40	0.2796	-0.1204	0.1373	0.3140	-0.0860	0.1047
	0.45	0.3137	-0.1363	0.1529	0.3551	-0.0949	0.1143
	0.50	0.3460	-0.1540	0.1696	0.3955	-0.1045	0.1239
	0.55	0.3756	-0.1744	0.1888	0.4351	-0.1149	0.1343
0.60	0.4032	-0.1968	0.2098	0.4743	-0.1257	0.1451	

## 4.2 The Long–Memory Stochastic Volatility Process

We consider the long–memory stochastic volatility (LMSV) process defined by equation (17) with the scale parameter  $\sigma = 0.8$ . The process  $\{X_t\}$  is a FARIMA(0,d,0) process generated using the Durbin–Levinson algorithm. Tables 5 and 6 report the estimation results for the long–memory parameter  $\alpha = 2d$  obtained from the local Whittle, wavelet and log–periodogram estimators. The comparison with the LP estimator is of interest as the use of

this estimator for LMSV processes is theoretically justified. Deo and Hurvich (2003) report few simulation results for  $m = [T^{0.3}], [T^{0.4}], [T^{0.5}]$ : the bias increases with  $m$ , while the RMSE decreases. We use both the optimal feasible bandwidth given by (41) and  $m = [T^{0.3}]$ .

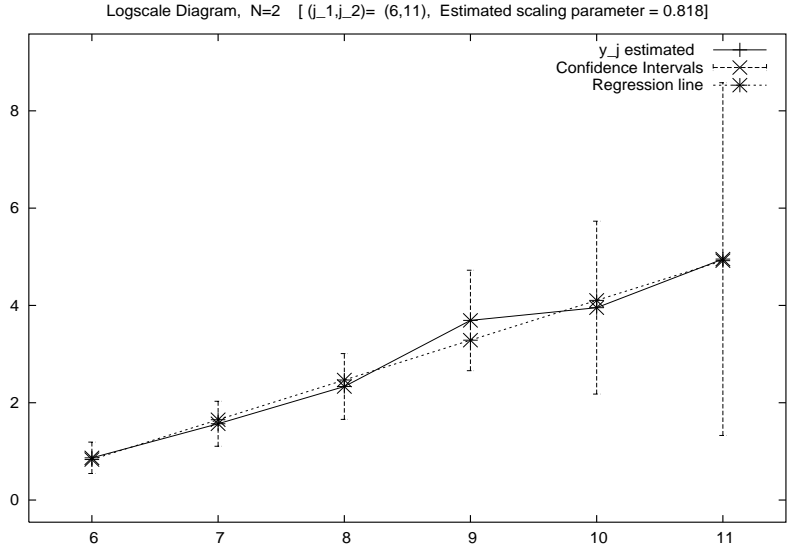
Figures 6 and 7 below display the logscale diagrams for one realization of a LMSV process, with different choices for the range of octaves  $[j_1, j_2]$ : the long-range dependent behavior of the LMSV process is captured for  $j_1 \geq 6$ .



**Fig. 6.** Logscale diagram for the realization of a LMSV process, with  $\alpha = 0.90$ . We select here  $j_1 = 1, j_2 = 11, N = 2$

These three estimators give here very similar results, when the optimal bandwidth is used for the LP estimator. This set of simulation results should stimulate further theoretical research on this wavelet estimator for volatility processes. We mention that in that case, the choice for  $j_1$  using equation (60), with either  $m_{LW}^{opt}$  or  $m_{LP}^{opt}$ , does not yield the best results.

When using the standard bandwidth  $m = [T^{0.3}]$ , the mean of the  $\hat{\alpha}_{LP}$  estimates are equal to -0.0598, -0.111, 0.0722, 0.1794, 0.2986, 0.4224, 0.5437, 0.6584, 0.7680, for scaling parameters respectively equal to 0.10, 0.20, 0.30, 0.40, 0.50, 0.60, 0.70, 0.80, 0.90, with a RMSE ranging from 0.4897 to 0.5139, i.e., the bias is slightly reduced only for  $\alpha > 0.40$ , but the RMSE is always very high so that the estimates obtained with  $m = [T^{0.3}]$  are not reliable. As we will see for other volatility processes, the choice  $m = [T^{0.3}]$  does not appear very sensible.



**Fig. 7.** Logscale diagram for the realization of a LMSV process, with  $\alpha = 0.90$ . We select here  $j_1 = 6, j_2 = 11, N = 2$

**Table 5.** Estimation of the scaling parameter for LMSV processes, wavelet estimator.  $T = 10000$

$\alpha$	$\hat{\alpha}_W^{RMSE}$	$j_1^{RMSE}$	$E\hat{\alpha}_W - \alpha$	RMSE	$\hat{\alpha}_W^{(6)}$	$E\hat{\alpha}_W^{(6)} - \alpha$	RMSE
0.10	0.0183	2	-0.0817	0.0847	0.0213	-0.0787	0.0898
0.20	0.0519	4	-0.1481	0.1559	0.0492	-0.1508	0.1569
0.30	0.1043	5	-0.1957	0.2092	0.0876	-0.2124	0.2170
0.40	0.1872	6	-0.2128	0.2441	0.1872	-0.2128	0.2441
0.50	0.2764	6	-0.2236	0.2536	0.2764	-0.2236	0.2536
0.60	0.3807	6	-0.2193	0.2499	0.3807	-0.2193	0.2499
0.70	0.4959	6	-0.2051	0.2367	0.4959	-0.2051	0.2367
0.80	0.6177	6	-0.1823	0.2184	0.6177	-0.1823	0.2184
0.90	0.7423	6	-0.1577	0.1986	0.7423	-0.1577	0.1986

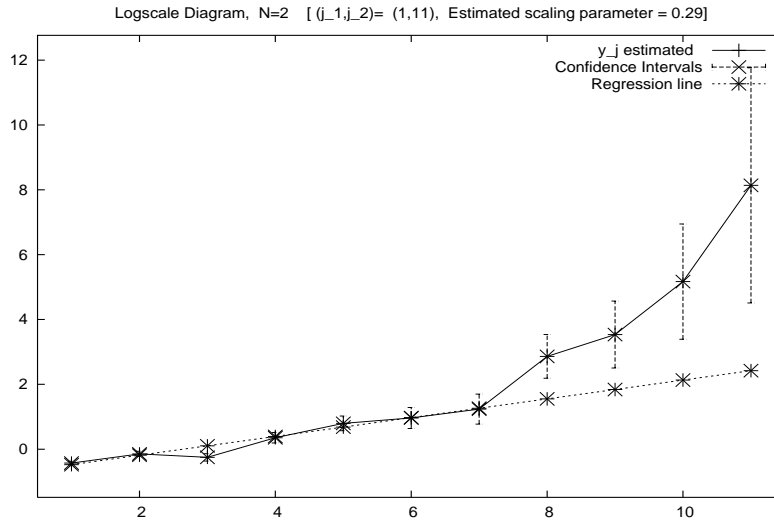
*Remark 7.* The results reported here for the wavelet estimator  $\hat{\alpha}_W^{RMSE}$  are obtained with  $N = 2$ . When  $N$  increases, both bias and RMSE (slightly) increase, but the octave which minimizes the RMSE does not depend on the choice of  $N$ .

**Table 6.** Estimation of the scaling parameter for LMSV processes, LW and LP estimator.  $T = 10000$ . The optimal bandwidths  $m_{LP}^{opt}$  and  $m_{LW}^{opt}$  are respectively used for the LP and LW estimators

$\alpha$	$\hat{\alpha}_{LP}^{opt}$	$E\hat{\alpha}_{LP}^{opt} - \alpha$	RMSE	$\hat{\alpha}_{LW}^{opt}$	$E\hat{\alpha}_{LW}^{opt} - \alpha$	RMSE
0.10	0.0153	-0.0847	0.1102	0.0186	-0.0814	0.0927
0.20	0.0516	-0.1484	0.1644	0.0508	-0.1492	0.1557
0.30	0.1049	-0.1951	0.2075	0.0971	-0.2029	0.2082
0.40	0.1764	-0.2236	0.2346	0.1622	-0.2378	0.2437
0.50	0.2656	-0.2344	0.2453	0.2525	-0.2475	0.2555
0.60	0.3715	-0.2285	0.2405	0.3679	-0.2321	0.2433
0.70	0.4914	-0.2085	0.2234	0.4978	-0.2022	0.2169
0.80	0.6198	-0.1802	0.1991	0.6313	-0.1687	0.1876
0.90	0.7511	-0.1489	0.1732	0.7634	-0.1366	0.1611

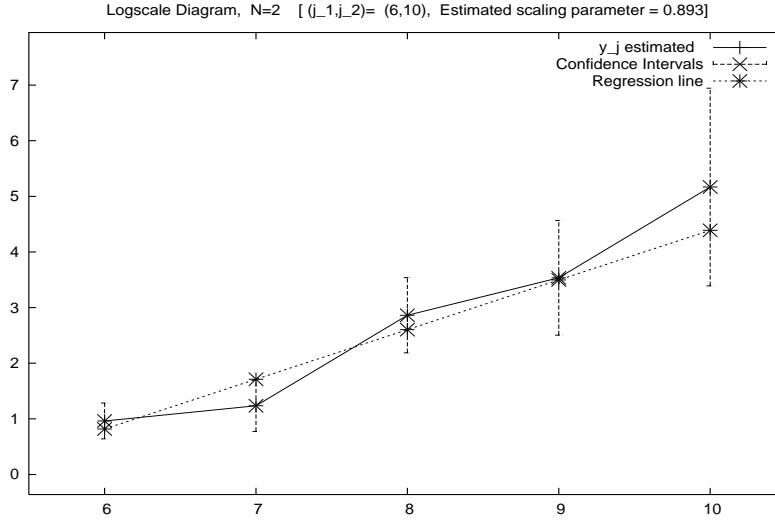
### 4.3 The Nonlinear Moving Average Process

Full Whittle estimation, i.e., with an exact specification of the spectrum, of the parameters for the Nonlinear Moving Average (NLMA) process has been considered by Robinson and Zaffaroni (1997) and studied by Zaffaroni (2003). Due to the incomplete specification used for the local Whittle estimator, see equation (1), we expect the bias for the LW estimator to be greater.



**Fig. 8.** Logscale diagram for the realization of a NLMA process, with  $\alpha = 0.90$ . We select here  $j_1 = 1, j_2 = 11, N = 2$

Figures 8 and 9 display the logscale diagrams for the realization of a NLMA process. As for the LMSV process, the long-memory behavior is captured by the octaves  $j \geq j_1 = 6$ .



**Fig. 9.** Logscale diagram for the realization of a NLMA process, with  $\alpha = 0.90$ . We select here  $j_1 = 6$ ,  $j_2 = 10$ ,  $N = 2$

**Table 7.** Estimation of the scaling parameter for NLMA processes, wavelet estimator.  $T = 10000$ ,  $N = 2$

$\alpha$	$\hat{\alpha}_W^{RMSE}$	$j_1^{RMSE}$	$E\hat{\alpha}_W^{RMSE} - \alpha$	RMSE	$\hat{\alpha}_W^{(6)}$	$E\hat{\alpha}_W^{(6)} - \alpha$	RMSE
0.10	0.2717	1	0.1717	0.1752	0.3120	0.2120	0.2529
0.20	0.3187	1	0.1187	0.1263	0.2936	0.0936	0.1807
0.30	0.3294	1	0.0294	0.0542	0.2684	0.0316	0.1692
0.40	0.3354	1	-0.0646	0.0793	0.2761	-0.1239	0.2132
0.50	0.3637	2	-0.1363	0.1487	0.3164	-0.1836	0.2565
0.60	0.4028	3	-0.1972	0.2106	0.3816	-0.2184	0.2858
0.70	0.4636	4	-0.2364	0.2554	0.4633	-0.2364	0.3033
0.80	0.5436	5	-0.2564	0.2891	0.5540	-0.2460	0.3130
0.90	0.6205	5	-0.2795	0.3094	0.6469	-0.2531	0.3197

As can be seen from Tables 7 and 8, the wavelet estimator performs slightly better than both the LW and LP estimators, as its bias and RMSE are on average smaller. The wavelet estimator with  $j_1$  selected using equation (60),

with either  $m_{LW}^{opt}$  or  $m_{LP}^{opt}$ , gives results close to the LW estimator and the wavelet estimator with  $j_1 = j_1^{RMSE}$ . Again in this case, equation (60) might serve as a rule of thumb for selecting  $j_1$ .

**Table 8.** Estimation of the scaling parameter for NLMA processes, local Whittle and log-periodogram estimators.  $T = 10000$ . The optimal bandwidths  $m_{LP}^{opt}$  and  $m_{LW}^{opt}$  are respectively used for the LP and LW estimators

$\alpha$	$\hat{\alpha}_{LP}^{opt}$	$E\hat{\alpha}_{LP}^{opt} - \alpha$	RMSE	$\hat{\alpha}_{LW}^{opt}$	$E\hat{\alpha}_{LW}^{opt} - \alpha$	RMSE
0.10	0.3028	0.2028	0.2124	0.3025	0.2025	0.2109
0.20	0.2899	0.0899	0.1157	0.2892	0.0892	0.1143
0.30	0.2689	-0.0311	0.0942	0.2679	-0.0321	0.0932
0.40	0.2732	-0.1268	0.1620	0.2731	-0.1269	0.1620
0.50	0.3092	-0.1908	0.2204	0.3091	-0.1909	0.2205
0.60	0.3744	-0.2256	0.2556	0.3714	-0.2286	0.2574
0.70	0.4557	-0.2443	0.2756	0.4537	-0.2463	0.2766
0.80	0.5470	-0.2530	0.2845	0.5469	-0.2531	0.2847
0.90	0.6458	-0.2542	0.2895	0.6456	-0.2544	0.2872

When using the standard bandwidth  $m = [T^{0.3}]$ , the mean of the  $\hat{\alpha}_{LP}$  estimates is equal to 0.3024, 0.3161, 0.3170, 0.3350, 0.3672, 0.4071, 0.4483, 0.4847, 0.5105, with a RMSE ranging from 0.0563 to 0.3927. The advantage provided by the optimal bandwidth  $m_{LP}^{opt}$  is not as obvious as for the LMSV model, but in the absence of the knowledge on the true Data Generating Process (DGP), using this optimal bandwidth does not yield any significant loss for the estimation of the parameter.

#### 4.4 Nonlinear Transformations of Fractionally Integrated Processes

We also consider some nonlinear transformations of  $FI(d)$  processes, i.e., processes  $\{Y_t\}$  defined as  $Y_t = G(X_t)$ , where  $X_t \sim FI(d)$ , and  $G(\cdot)$  can be written as a sum of Hermite polynomials. As in Dittman and Granger (2002), we choose the transformations  $G(x) = x^2$ ,  $G(x) = x^3$ ,  $G(x) = x^4$ ,  $G(x) = x^3 - 3x$ , and  $G(x) = x^4 - 6x^2$ , the Hermite rank of which are respectively equal to  $R = 2, 1, 2, 3, 4$ . We also consider some trigonometric and exponential transformations:  $G(x) = \sin(x)$ ,  $G(x) = \cos(x)$ ,  $G(x) = \exp(x)$ , and  $G(x) = (1 + \exp(-x))^{-1}$ . The Hermite rank of these transformations are all equal to 1, except for the cosine transformation, the Hermite rank of which is  $R = 2$ .

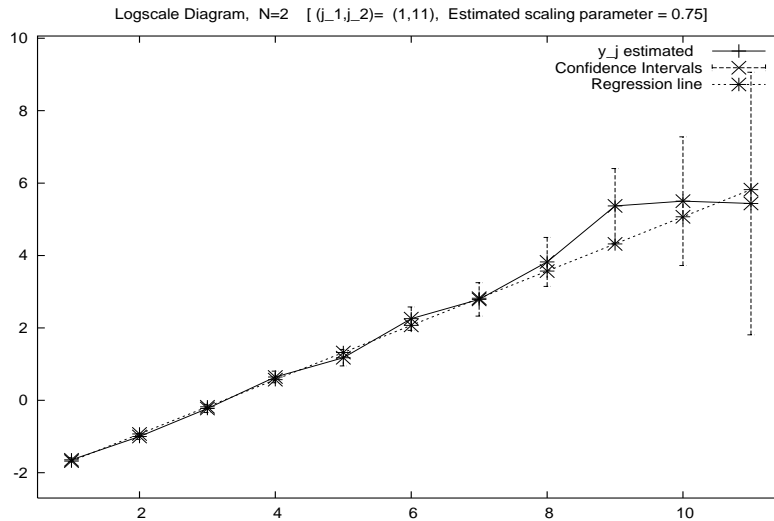
Granger and Dittman (2002) report simulation results for smaller sample sizes and the LP estimator with bandwidth  $m = [T]^{4/5}$ . The use of the optimal bandwidth  $m_{LP}^{opt}$  results for some transformations in a marginal increase of the bias and RMSE, but not very significant, so that we can reliably stick to that

optimal bandwidth. We observe that for all nonlinear transformations, the estimators  $\hat{\alpha}_{LW}$  and  $\hat{\alpha}_{LP}$  give very close results.

The  $FI(d)$  processes have been generated using the Durbin–Levinson algorithm. Consider the first series of transformations, see Tables 9 and 10: the wavelet estimator performs generally better than the LP and LW estimators, i.e., its bias and RMSE are most of the times lower than those of the LP and LW estimators, the advantage of the wavelet estimator being obvious for the largest values of  $\tilde{\alpha}$ . For the highly nonlinear transformations, e.g.,  $G(x) = x^3 - 3x$  and  $G(x) = x^4 - 6x^2$ , the bias is huge for all estimators for the lowest values of  $\tilde{\alpha}$ , and choosing a higher value for  $j_1$ , even for samples of size  $T = 20,000$ , does not yield any significant improvement. Estimation results with  $j_1$  computed from equation (60) are not as good as the ones with either  $j_1^{RMSE}$  or the LP and LW estimators, but might be considered as informative for a first analysis.

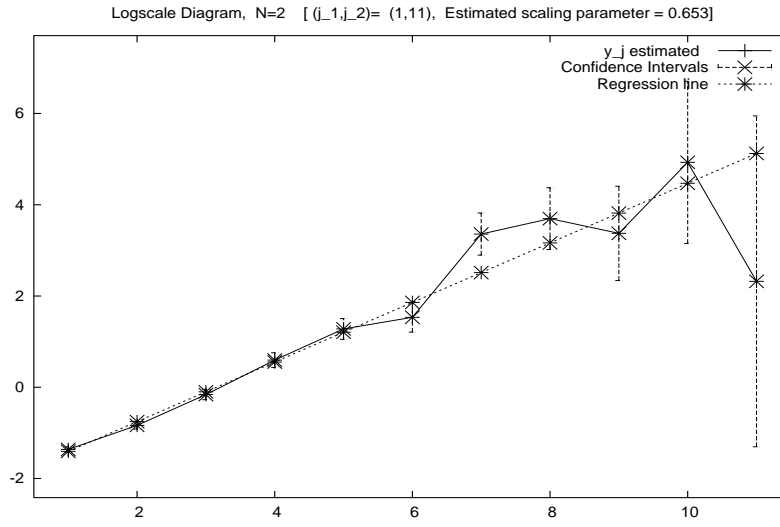
*Remark 8.* When  $N$  increases, the wavelet estimates do not differ too much in terms of bias and RMSE, except for the case  $j_1 = 7$  and  $N = 6$  for which the RMSE is very large.

Figures 10 to 18 below display the logscale diagrams for the 9 transformations  $G(x)$  considered here.

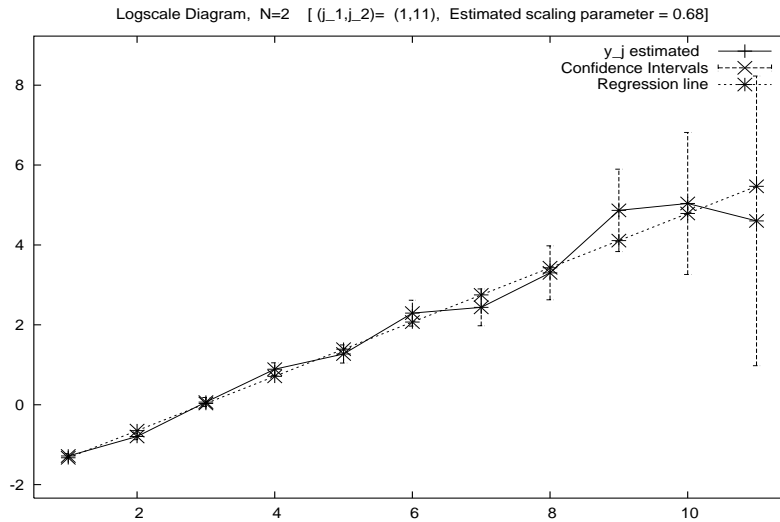


**Fig. 10.** Logscale diagram for the realization of a nonlinear transformation of a  $FI(d)$  process, with  $G(x) = x^2$ ,  $d = 0.45$  then  $\alpha = 0.80$ ,  $T = 10000$ . We select here  $j_1 = 1$ ,  $j_2 = 11$ ,  $N = 2$

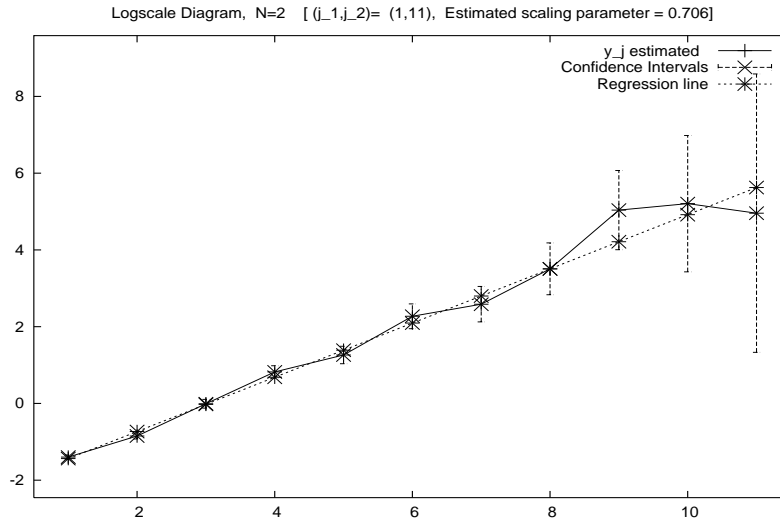




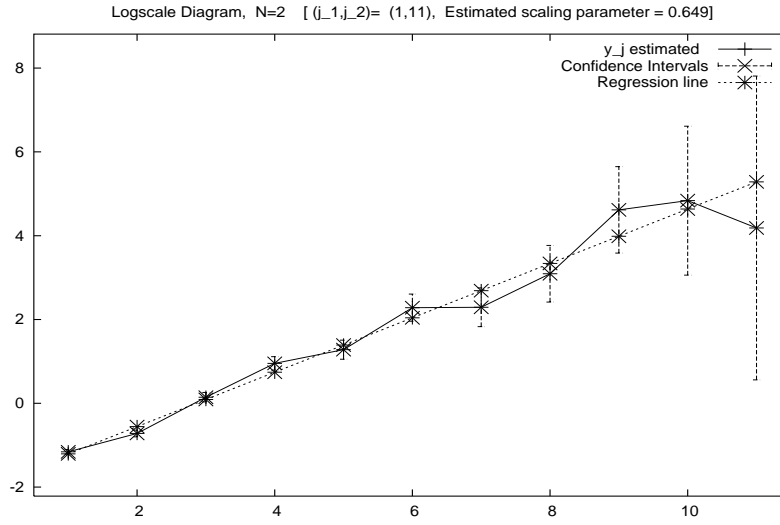
**Fig. 11.** Logscale diagram for the realization of a nonlinear transformation of a FI( $d$ ) process, with  $G(x) = x^3$ ,  $d = 0.45$  then  $\alpha = 0.90$ ,  $T = 10000$ . We select here  $j_1 = 1$ ,  $j_2 = 11$ ,  $N = 2$



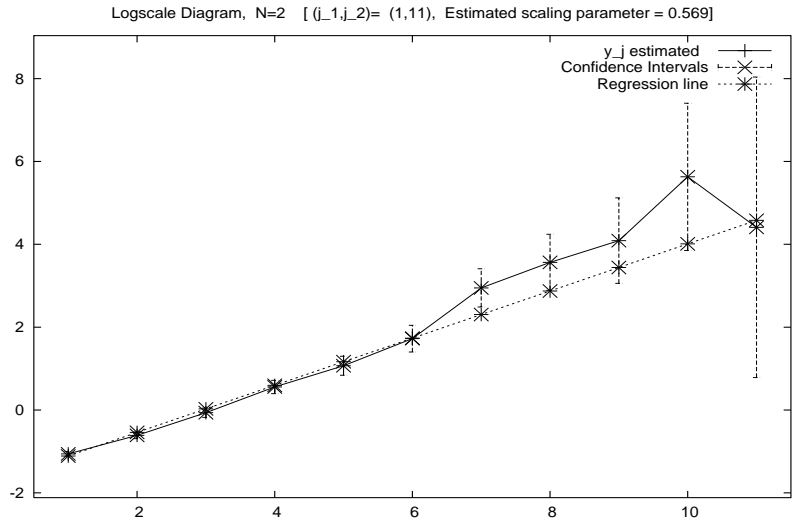
**Fig. 12.** Logscale diagram for the realization of a nonlinear transformation of a FI( $d$ ) process, with  $G(x) = x^4$ ,  $d = 0.45$  then  $\alpha = 0.80$ ,  $T = 10000$ . We select here  $j_1 = 1$ ,  $j_2 = 11$ ,  $N = 2$



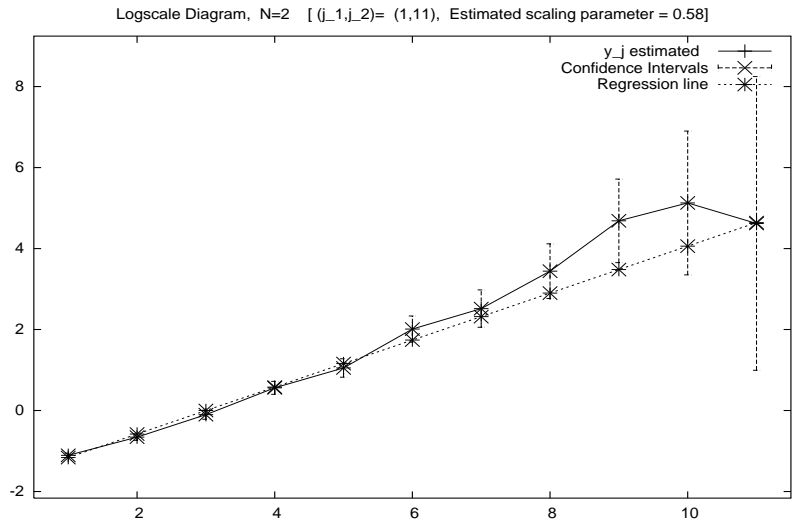
**Fig. 13.** Logscale diagram for the realization of a nonlinear transformation of a FI( $d$ ) process, with  $G(x) = x^3 - 3x$ ,  $d = 0.45$  then  $\alpha = 0.80$ ,  $T = 10000$ . We select here  $j_1 = 1$ ,  $j_2 = 11$ ,  $N = 2$



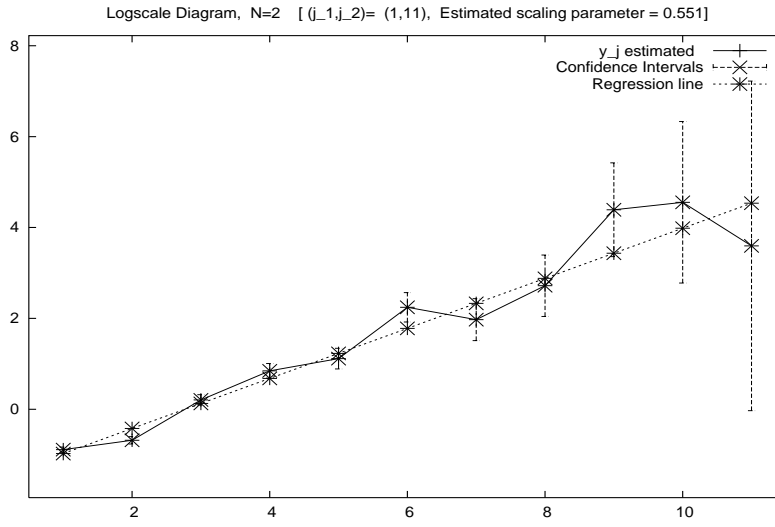
**Fig. 14.** Logscale diagram for the realization of a nonlinear transformation of a FI( $d$ ) process, with  $G(x) = x^4 - 6x$ ,  $d = 0.45$  then  $\alpha = 0.60$ ,  $T = 10000$ . We select here  $j_1 = 1$ ,  $j_2 = 11$ ,  $N = 2$



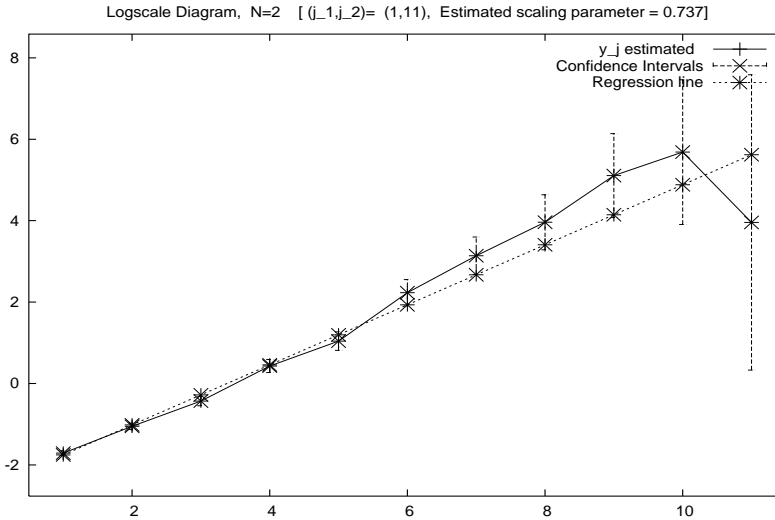
**Fig. 15.** Logscale diagram for the realization of a nonlinear transformation of a FI( $d$ ) process, with  $G(x) = \sin(x)$ ,  $d = 0.45$  then  $\alpha = 0.90$ ,  $T = 10000$ . We select here  $j_1 = 1$ ,  $j_2 = 11$ ,  $N = 2$



**Fig. 16.** Logscale diagram for the realization of a nonlinear transformation of a FI( $d$ ) process, with  $G(x) = \cos(x)$ ,  $d = 0.45$  then  $\alpha = 0.80$ ,  $T = 10000$ . We select here  $j_1 = 1$ ,  $j_2 = 11$ ,  $N = 2$



**Fig. 17.** Logscale diagram for the realization of a nonlinear transformation of a FI( $d$ ) process, with  $G(x) = \exp(x)$ ,  $d = 0.45$  then  $\alpha = 0.90$ ,  $T = 10000$ . We select here  $j_1 = 1$ ,  $j_2 = 11$ ,  $N = 2$



**Fig. 18.** Logscale diagram for the realization of a nonlinear transformation of a FI( $d$ ) process, with  $G(x) = (1 + \exp(-x))^{-1}$ ,  $d = 0.45$  then  $\alpha = 0.90$ ,  $T = 10000$ . We select here  $j_1 = 1$ ,  $j_2 = 11$ ,  $N = 2$

**Table 9.** Estimation of the scaling parameter for nonlinear transformations of FI( $d$ ) processes, with  $\alpha = 2\tilde{d}$  and  $\tilde{d} = \max\{0, (d - 0.5)R + 0.5\}$ .  $N = 2, T = 10000$

$G(x)$	$\alpha$	$\hat{\alpha}_W^{RMSE}$	$j_1^{RMSE}$	$E\hat{\alpha}_W^{RMSE} - \alpha$	RMSE	$\hat{\alpha}_W^m$	$E\hat{\alpha}_W^m - \alpha$	RMSE
$x^2$ ( $R = 2$ )	0.00	0.1176	5	0.1176	0.1488	0.1403	0.1403	0.1527
	0.20	0.2627	1	0.0627	0.0705	0.2449	0.0449	0.0919
	0.40	0.3964	1	-0.0036	0.0578	0.3786	-0.0214	0.1136
	0.60	0.5769	2	-0.0231	0.0861	0.5396	-0.0604	0.1530
	0.80	0.7447	2	-0.0553	0.0888	0.7145	-0.0855	0.1935
$x^3$ ( $R = 1$ )	0.10	0.0608	2	-0.0392	0.0464	0.0653	-0.0347	0.0599
	0.20	0.1341	3	-0.0659	0.0810	0.1385	-0.0615	0.0790
	0.30	0.2224	4	-0.0775	0.0948	0.2122	-0.0808	0.0962
	0.40	0.3087	4	-0.0913	0.1066	0.3062	-0.0938	0.1097
	0.50	0.3966	4	-0.1034	0.1178	0.3974	-0.1026	0.1213
	0.60	0.4841	4	-0.1159	0.1304	0.4882	-0.1118	0.1335
	0.70	0.5705	4	-0.1295	0.1454	0.5759	-0.1241	0.1493
	0.80	0.6596	4	-0.1404	0.1596	0.6640	-0.1360	0.1712
	0.90	0.7564	3	-0.1436	0.1587	0.7589	-0.1411	0.1971
$x^4$ ( $R = 2$ )	0.00	0.1272	1	0.1272	0.1348	0.1129	0.1129	0.1374
	0.20	0.2094	1	0.0094	0.0522	0.2047	0.0047	0.0961
	0.40	0.3392	2	-0.0608	0.0962	0.3222	-0.0778	0.1466
	0.60	0.4796	2	-0.1204	0.1469	0.4591	-0.1409	0.2143
	0.80	0.6105	2	-0.1600	0.1839	0.6097	-0.1903	0.2803
$x^3 - 3x$ ( $R = 3$ )	0.00	0.1483	1	0.1483	0.1557	0.1633	0.1633	0.1863
	0.10	0.2781	1	0.1781	0.1878	0.3188	0.2188	0.2474
	0.40	0.4368	1	0.0368	0.0782	0.4884	0.0884	0.1646
	0.70	0.6703	2	-0.0297	0.0929	0.6651	-0.0349	0.1736
$x^4 - 6x^2$ ( $R = 4$ )	0.00	0.1738	1	0.1738	0.1949	0.1676	0.1676	0.2268
	0.20	0.3333	1	0.1333	0.1664	0.3327	0.1327	0.2341
	0.60	0.5350	1	-0.0650	0.1197	0.5270	-0.0730	0.2459

**Table 10.** Estimation of the scaling parameter for nonlinear transformations of FI( $d$ ) processes, with  $\alpha = 2\tilde{d}$  and  $\tilde{d} = \max\{0, (d - 0.5)R + 0.5\}$ .  $T = 10000$ . The optimal bandwidths  $m_{LP}^{opt}$  and  $m_{LW}^{opt}$  are respectively used for the LP and LW estimators

$G(x)$	$\alpha$	$\hat{\alpha}_{LP}^{opt}$	$E\hat{\alpha}_{LP}^{opt} - \alpha$	RMSE	$\hat{\alpha}_{LW}^{opt}$	$E\hat{\alpha}_{LW}^{opt} - \alpha$	RMSE
$x^2$ ( $R = 2$ )	0.00	0.1193	0.1193	0.1371	0.1192	0.1192	0.1369
	0.20	0.2248	0.0248	0.0975	0.2225	0.0225	0.0960
	0.40	0.3644	-0.0356	0.1315	0.3599	-0.0401	0.1302
	0.60	0.5223	-0.0777	0.1735	0.5190	-0.0810	0.1736
	0.80	0.6953	-0.1047	0.2133	0.6867	-0.1132	0.2149
$x^3$ ( $R = 1$ )	0.10	0.0638	-0.0362	0.0580	0.0636	-0.0364	0.0576
	0.20	0.1410	-0.0590	0.0754	0.1409	-0.0591	0.0750
	0.30	0.2277	-0.0723	0.0882	0.2276	-0.0724	0.0881
	0.40	0.3217	-0.0783	0.0964	0.3216	-0.0784	0.0964
	0.50	0.4184	-0.0816	0.1029	0.4185	-0.0815	0.1028
	0.60	0.5134	-0.0866	0.1099	0.5134	-0.0866	0.1099
	0.70	0.6019	-0.0981	0.1261	0.6020	-0.0980	0.1245
	0.80	0.6877	-0.1123	0.1495	0.6866	-0.1134	0.1446
	0.90	0.7755	-0.1245	0.1755	0.7728	-0.1272	0.1699
$x^4$ ( $R = 2$ )	0.00	0.0981	0.0981	0.1180	0.0978	0.0978	0.1176
	0.20	0.1934	-0.0066	0.0936	0.1906	-0.0094	0.0910
	0.40	0.3158	-0.0842	0.1474	0.3157	-0.0843	0.1475
	0.60	0.4609	-0.1391	0.2043	0.4593	-0.1407	0.2048
	0.80	0.6185	-0.1815	0.2566	0.6105	-0.1895	0.2568
$x^3 - 3x$ ( $R = 3$ )	0.00	0.1856	0.1856	0.2029	0.1850	0.1850	0.2022
	0.10	0.3542	0.2542	0.2759	0.3533	0.2533	0.2743
	0.40	0.5286	0.1286	0.1789	0.5259	0.1259	0.1736
	0.70	0.6893	-0.0107	0.1423	0.6866	-0.0134	0.1377
$x^4 - 6x^2$ ( $R = 4$ )	0.00	0.1736	0.1736	0.2114	0.1732	0.1732	0.2107
	0.20	0.3506	0.1506	0.2204	0.3502	0.1502	0.2195
	0.60	0.5487	-0.0513	0.1992	0.5435	-0.0565	0.1934

For the second series of transformations, see Tables 11 and 12, the wavelet estimator performs better than the two other estimators, except for the exponential transformation.

**Table 11.** Estimation of the scaling parameter for nonlinear transformations of FI( $d$ ) processes, with  $\alpha = 2\tilde{d}$  and  $\tilde{d} = \max\{0, (d - 0.5)R + 0.5\}$ .  $N = 2$ ,  $T = 10000$

$G(x)$	$\alpha$	$\hat{\alpha}_W^{RMSE}$	$j_1^{RMSE}$	$E\hat{\alpha}_W^{RMSE} - \alpha$	RMSE	$\hat{\alpha}_W^m$	$E\hat{\alpha}_W^m - \alpha$	RMSE
$\sin(x)$ ( $R = 1$ )	0.10	0.0822	2	-0.0178	0.0352	0.0871	-0.0129	0.0450
	0.20	0.1694	2	-0.0306	0.0382	0.1777	-0.0223	0.0492
	0.30	0.2652	3	-0.0348	0.0474	0.2704	-0.0296	0.0544
	0.40	0.3556	3	-0.0444	0.0553	0.3642	-0.0358	0.0620
	0.50	0.4577	4	-0.0423	0.0639	0.4585	-0.0415	0.0711
	0.60	0.5413	4	-0.0587	0.0761	0.5496	-0.0504	0.0824
	0.70	0.6353	5	-0.0647	0.1013	0.6235	-0.0765	0.1051
	0.80	0.6648	5	-0.1352	0.1671	0.6561	-0.1439	0.1716
	0.90	0.6606	5	-0.2394	0.2606	0.6556	-0.2444	0.2654
$\cos(x)$ ( $R = 2$ )	0.00	0.1505	1	0.1505	0.1520	0.1310	0.1310	0.1421
	0.20	0.2394	1	0.0394	0.0473	0.2333	0.0333	0.0779
	0.40	0.3749	2	-0.0251	0.0519	0.3644	-0.0356	0.1015
	0.60	0.5192	3	-0.0808	0.1043	0.5177	-0.0823	0.1389
	0.80	0.6407	4	-0.1593	0.1751	0.6435	-0.1565	0.1890
$\exp(x)$ ( $R = 1$ )	0.10	0.0628	3	-0.0372	0.0598	0.0643	-0.0357	0.0648
	0.20	0.1326	3	-0.0674	0.0826	0.1361	-0.0639	0.0851
	0.30	0.2169	4	-0.0831	0.1035	0.2152	-0.0848	0.1034
	0.40	0.2986	4	-0.1014	0.1202	0.2967	-0.1033	0.1221
	0.50	0.3798	4	-0.1202	0.1389	0.3807	-0.1193	0.1414
	0.60	0.4573	4	-0.1427	0.1626	0.4599	-0.1401	0.1664
	0.70	0.5276	4	-0.1724	0.1953	0.5320	-0.1680	0.2005
	0.80	0.5868	4	-0.2132	0.2420	0.5885	-0.2114	0.2557
	0.90	0.6260	3	-0.2740	0.3022	0.6265	-0.2735	0.3384
$(1 + e^{-x})^{-1}$ ( $R = 1$ )	0.10	0.0871	1	-0.0129	0.0202	0.0980	-0.0020	0.0429
	0.20	0.1905	2	-0.0095	0.0245	0.1964	-0.0036	0.0439
	0.30	0.2879	2	-0.0121	0.0260	0.2945	-0.0055	0.0464
	0.40	0.3842	2	-0.0158	0.0277	0.3935	-0.0065	0.0531
	0.50	0.4799	2	-0.0201	0.0307	0.4946	-0.0054	0.0614
	0.60	0.5878	3	-0.0121	0.0349	0.5961	-0.0039	0.0697
	0.70	0.6848	3	-0.0152	0.0363	0.6951	-0.0049	0.0816
	0.80	0.7796	3	-0.0204	0.0395	0.7925	-0.0075	0.0958
	0.90	0.8669	3	-0.0331	0.0508	0.8796	-0.0204	0.1196

**Table 12.** Estimation of the scaling parameter for nonlinear transformations of FI( $d$ ) processes, with  $\alpha = 2\tilde{d}$  and  $\tilde{d} = \max\{0, (d - 0.5)R + 0.5\}$ .  $T = 10000$ . The optimal bandwidths  $m_{LP}^{opt}$  and  $m_{LW}^{opt}$  are respectively used for the LP and LW estimators

$G(x)$	$\alpha$	$\hat{\alpha}_{LP}^{opt}$	$E\hat{\alpha}_{LP}^{opt} - \alpha$	RMSE	$\hat{\alpha}_{LW}^{opt}$	$E\hat{\alpha}_{LW}^{opt} - \alpha$	RMSE
$\sin(x)$ ( $R = 1$ )	0.10	0.0855	-0.0145	0.0465	0.0856	-0.0144	0.0461
	0.20	0.1779	-0.0221	0.0518	0.1779	-0.0221	0.0517
	0.30	0.2736	-0.0264	0.0581	0.2736	-0.0264	0.0579
	0.40	0.3709	-0.0291	0.0649	0.3710	-0.0290	0.0647
	0.50	0.4689	-0.0311	0.0728	0.4697	-0.0303	0.0712
	0.60	0.5618	-0.0382	0.0795	0.5620	-0.0380	0.0791
	0.70	0.6417	-0.0583	0.0972	0.6425	-0.0575	0.0950
	0.80	0.6788	-0.1212	0.1582	0.6793	-0.1207	0.1569
	0.90	0.6587	-0.2413	0.2735	0.6603	-0.2397	0.2704
$\cos(x)$ ( $R = 2$ )	0.00	0.1135	0.1135	0.1303	0.1129	0.1129	0.1296
	0.20	0.2147	0.0147	0.0883	0.2141	0.0141	0.0881
	0.40	0.3512	-0.0488	0.1243	0.3510	-0.0490	0.1240
	0.60	0.5121	-0.0879	0.1633	0.5108	-0.0892	0.1622
$\exp(x)$ ( $R = 1$ )	0.10	0.0617	-0.0383	0.0592	0.0615	-0.0385	0.0588
	0.20	0.1370	-0.0630	0.0787	0.1369	-0.0031	0.1194
	0.30	0.2203	-0.0797	0.0952	0.2204	-0.0796	0.0951
	0.40	0.3085	-0.0915	0.1089	0.3086	-0.0914	0.1087
	0.50	0.3978	-0.1022	0.1226	0.3979	-0.1021	0.1223
	0.60	0.4818	-0.1182	0.1442	0.4821	-0.1179	0.1406
	0.70	0.5587	-0.1413	0.1671	0.5590	-0.1410	0.1666
	0.80	0.6206	-0.1794	0.2136	0.6201	-0.1799	0.2092
$(1 + \exp(-x))^{-1}$ ( $R = 1$ )	0.10	0.0962	-0.0038	0.0441	0.0961	-0.0039	0.0439
	0.20	0.1943	-0.0057	0.0482	0.1942	-0.0058	0.0475
	0.30	0.2922	-0.0078	0.0529	0.2923	-0.0077	0.0529
	0.40	0.3913	-0.0087	0.0636	0.3915	-0.0085	0.0601
	0.50	0.4904	-0.0096	0.0686	0.4905	-0.0095	0.0684
	0.60	0.5885	-0.0115	0.0822	0.5880	-0.0120	0.0766
	0.70	0.6834	-0.0166	0.0903	0.6833	-0.0167	0.0902
	0.80	0.7766	-0.0234	0.1053	0.7764	-0.0236	0.1041
	0.90	0.8614	-0.0386	0.1350	0.8610	-0.0390	0.1245

## 5 Robustness: Long–Memory Versus Non–Stationarity

This section aims at contributing to the issue of disentangling long–memory and non–stationarity. This can be addressed through two major questions. Can one estimate correctly the long–memory parameter of an actual long–range dependent process when non–stationarities are superimposed to it? Can



one figure out that a given time series is non-stationary and has no long-memory? The first question is illustrated in Sections 5.1, 5.2 and 5.4 where trends, change points or both are respectively superimposed to long-memory processes. The second one is considered in Section 5.3 where change points in a short-range memory process may be confused with long-memory.

### 5.1 Long-Memory Plus Trends: Polynomial Trends Superimposed to Fractionally Integrated Processes

As mentioned earlier, the wavelets coefficients  $d_Y(j, \cdot)$  are the same for the processes  $\{Y_t\}$  and  $\{Y_t + P_t\}$  where  $P_t$  is a polynomial of order at most  $N - 1$ , where  $N$  is the number of vanishing moments of the mother wavelet  $\psi_0$ . Thus, with an appropriate choice for  $N$ , we can discriminate between trended and long-range dependent processes. This separation is also possible with Fourier based methods, when replacing the periodogram by a tapered version of it, see e.g., Lobato and Velasco (2000).

We consider here the accuracy of the wavelet based estimator, as we will use this estimator in section 6.2 for estimating the dependence of financial time series with a trend, i.e., transactions volume. We consider the following process

$$Y_t = X_t + T_t, \quad X_t \sim \text{FI}(d), \quad T_t = \sum_{l=0}^q \xi_l t^l, \quad (63)$$

which additively mixes a long-range dependent process and a polynomial trend. We consider  $q = 1, 2, 3$ . Table 13 below reports simulation results for  $N = q + 1$  only.

For the relevant choices of  $N$ , i.e., those for which  $N > q$ , the wavelet estimates are the same as those for the FARIMA process without polynomial trend, since we used the same sequence of pseudo-error terms for all the  $\text{FI}(d)$  process.

*Remark 9.* In this section, we do not compare the performance of the wavelet estimator with the one of the LP and LW estimators. For the LW estimator, we would have to consider the tapered periodogram, as in Lobato and Velasco (2000), and very likely perform a similar transformation for the LP estimator.

**Table 13.** Estimation of the scaling parameter for an additive combination of a  $\text{FI}(d)$  processes and a polynomial trend of order  $q$ .  $T = 10000$ 

Order	$\alpha$	$\hat{\alpha}_W^{RMSE}$	$j_1^{RMSE}$	$E\hat{\alpha}_W^{RMSE} - \alpha$	RMSE	$\hat{\alpha}_W^{(5)}$	$E\hat{\alpha}_W^{(5)} - \alpha$	RMSE
$q = 1$	0.10	0.0860	2	-0.0140	0.0205	0.1016	0.0016	0.0744
$(N=2)$	0.20	0.1907	2	-0.0093	0.0235	0.2017	0.0017	0.0745
	0.30	0.2868	2	-0.0132	0.0254	0.3018	0.0018	0.0747
	0.40	0.3833	2	-0.0167	0.0275	0.4020	0.0020	0.0748
	0.50	0.4800	2	-0.0200	0.0297	0.5022	0.0022	0.0751
	0.60	0.5771	2	-0.0229	0.0318	0.6023	0.0023	0.0753
	0.70	0.6934	3	-0.0066	0.0338	0.7025	0.0025	0.0757
	0.80	0.7929	3	-0.0713	0.0341	0.8026	0.0026	0.0761
	0.90	0.8924	3	-0.0076	0.0345	0.9027	0.0027	0.0766
$q = 2$	0.10	0.0875	2	-0.0125	0.0196	0.1022	0.0022	0.0771
$(N=3)$	0.20	0.1930	2	-0.0070	0.0234	0.2023	0.0023	0.0772
	0.30	0.2899	2	-0.0101	0.0246	0.3023	0.0023	0.0773
	0.40	0.3870	2	-0.0130	0.0259	0.4024	0.0024	0.0776
	0.50	0.4842	2	-0.0158	0.0275	0.5024	0.0024	0.0779
	0.60	0.5817	2	-0.0183	0.0291	0.6024	0.0024	0.0783
	0.70	0.6792	2	-0.0208	0.0308	0.7042	0.0042	0.0788
	0.80	0.7769	2	-0.0231	0.0325	0.8025	0.0025	0.0794
	0.90	0.8747	2	-0.0253	0.0342	0.9026	0.0026	0.0800
$q = 3$	0.10	0.0880	1	-0.0120	0.0194	0.1030	0.0030	0.0817
$(N=4)$	0.20	0.1937	2	-0.0063	0.0237	0.2030	0.0030	0.0817
	0.30	0.2909	2	-0.0091	0.0247	0.3031	0.0031	0.0817
	0.40	0.3881	2	-0.0119	0.0259	0.4031	0.0031	0.0818
	0.50	0.4855	2	-0.0145	0.0272	0.5030	0.0030	0.0819
	0.60	0.5830	2	-0.0170	0.0287	0.6030	0.0030	0.0820
	0.70	0.6806	2	-0.0194	0.0303	0.7029	0.0029	0.0823
	0.80	0.7783	2	-0.0217	0.0319	0.8028	0.0028	0.0826
	0.90	0.8761	2	-0.0239	0.0335	0.9027	0.0027	0.0830

## 5.2 Long–Memory Plus Trends Plus Change–Points: Change–Points in Fractionally Integrated Processes

Since economic processes are subject to changes in regime, we consider the extreme case of the following change–point process

$$\begin{aligned}
 Y_t &= X_t, & X_t &\sim \text{FI}(d), & t &\leq k, \\
 Y_t &= X_t + T_t, & X_t &\sim \text{FI}(d), & T_t &= \sum_{l=0}^q \xi_l (t - k + 1)^l, & t > k, \quad (64)
 \end{aligned}$$

i.e., a process with a broken polynomial trend. In some sense, here two types of difficulties are mixed up together. The non–stationarity superimposed to long–memory results both from trends and change points. However, in that

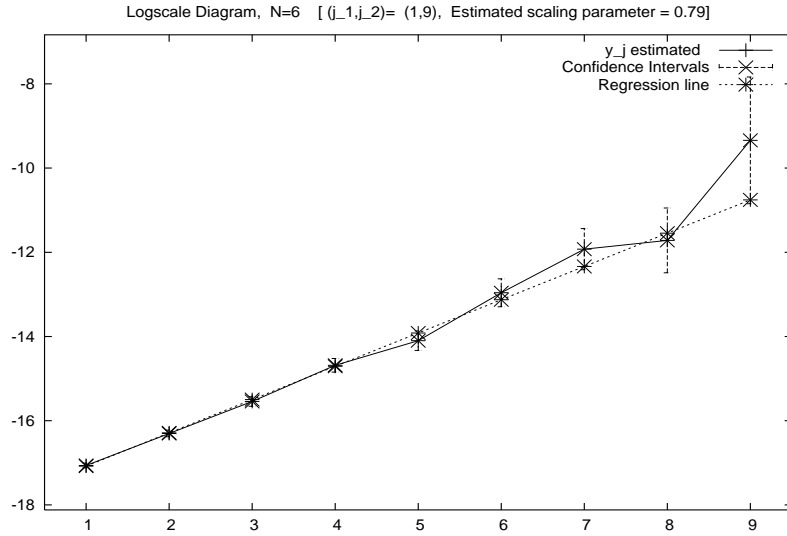
particular extreme case, a visual inspection of the series will suggest to split the series at the break points.

*Remark 10.* These results reported here obviously hold provided that the polynomial trend series are of the same order of magnitude as the series  $X_t$ .

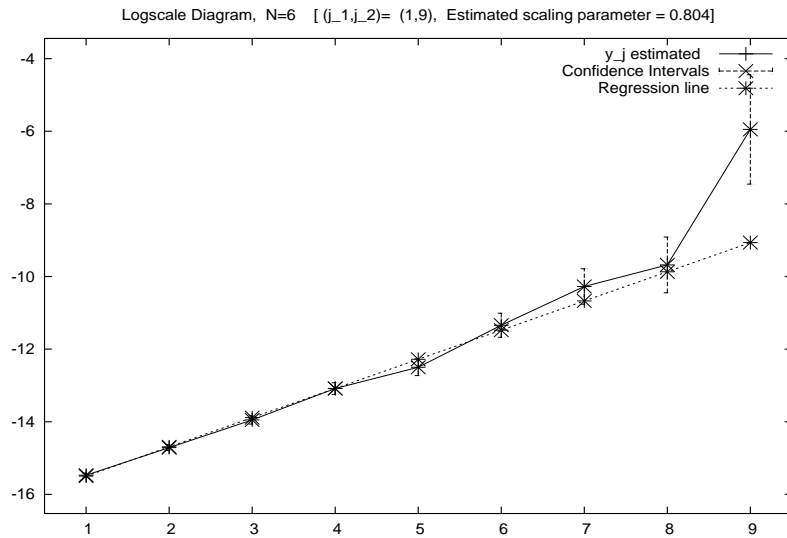
**Table 14.** Estimation of the scaling parameter for an additive combination of a  $FI(d)$  processes and with a broken polynomial trend of order  $q = 1, q = 2$  and  $q = 3$ , defined by equation (64). Columns 3 to 6 report estimation results for  $k = [T/10]$ , while columns 7 to 10 report estimation results for  $k = [T/2], T = 10000$

$N$	$\alpha$	$\hat{\alpha}_W^{RMSE}$	$j_1^{RMSE}$	$E\hat{\alpha}_W^{RMSE} - \alpha$	RMSE	$\hat{\alpha}_W^{RMSE}$	$j_1^{RMSE}$	$E\hat{\alpha}_W^{RMSE} - \alpha$	RMSE
$q = 1$	0.10	0.0970	2	-0.0030	0.0232	0.0971	2	-0.0029	0.0232
$N = 6$	0.20	0.1942	2	-0.0057	0.0238	0.1943	2	-0.0057	0.0238
	0.30	0.2915	2	-0.0085	0.0246	0.2916	2	-0.0084	0.0246
	0.40	0.3889	2	-0.0111	0.0258	0.3889	2	-0.0111	0.0258
	0.50	0.4863	2	-0.0137	0.0270	0.4863	2	-0.0137	0.0271
	0.60	0.5838	2	-0.0162	0.0285	0.5838	2	-0.0162	0.0285
	0.70	0.6814	2	-0.0186	0.0301	0.6814	2	-0.0186	0.0301
	0.80	0.7790	2	-0.0210	0.0317	0.7790	2	-0.0210	0.0317
	0.90	0.8767	2	-0.0233	0.0334	0.8767	2	-0.0233	0.0334
$q = 2$	0.10	0.0915	1	-0.0085	0.0178	0.1018	1	0.0018	0.0159
$N = 6$	0.20	0.1981	2	-0.0019	0.0232	0.2046	2	0.0046	0.0243
	0.30	0.2941	2	-0.0059	0.0240	0.2991	2	-0.0009	0.0239
	0.40	0.3906	2	-0.0094	0.0251	0.4050	2	0.0050	0.0239
	0.50	0.4874	2	-0.0126	0.0265	0.4996	2	-0.0004	0.0234
	0.60	0.5846	2	-0.0154	0.0281	0.6023	2	0.0023	0.0233
	0.70	0.6819	2	-0.0181	0.0299	0.6960	2	-0.0040	0.0237
	0.80	0.7793	2	-0.0207	0.0315	0.7903	2	-0.0097	0.0255
	0.90	0.8769	2	-0.0231	0.0332	0.8851	2	-0.0149	0.0281
$q = 3$	0.10	0.0996	2	-0.0004	0.0231	0.0981	2	-0.0019	0.0162
$N = 6$	0.20	0.1960	2	-0.0040	0.0235	0.2018	2	0.0018	0.0244
	0.30	0.2927	2	-0.0073	0.0243	0.3039	2	0.0039	0.0242
	0.40	0.3897	2	-0.0103	0.0255	0.3990	2	-0.0010	0.0238
	0.50	0.4869	2	-0.0131	0.0268	0.5041	2	0.0041	0.0237
	0.60	0.5842	2	-0.0158	0.0283	0.6037	2	0.0037	0.0235
	0.70	0.6816	2	-0.0184	0.0299	0.6972	2	-0.0028	0.0235
	0.80	0.7792	2	-0.0208	0.0315	0.7913	2	-0.0087	0.0251
	0.90	0.8768	2	-0.0232	0.0333	0.8858	2	-0.0142	0.0276

We note that for this particular process, the value of the statistic  $Q$  for  $j_1 = j_1^{RMSE}$  does not differ too much for the one when  $j_1$  is given following the arguments in Veitch, Abry, Taquq (2003) when the number of moments  $N$  is adequately chosen, i.e., large enough for fitting all possible variations of the process. Figures 19 and 20 display the logscale diagrams for trended processes with change-points.



**Fig. 19.** Logscale diagram for the realization of an additive combination of a FI( $d$ ) process ( $\alpha = 0.8$ ) and a broken polynomial trend of order  $q = 1$ , with  $k = \lceil T/10 \rceil$ . We select here  $j_1 = 1, j_2 = 9, N = 6$



**Fig. 20.** Logscale diagram for the realization of an additive combination of a FI( $d$ ) process ( $\alpha = 0.8$ ) and a broken polynomial trend of order  $q = 1$ , with  $k = \lceil T/2 \rceil$ . We select here  $j_1 = 1, j_2 = 11, N = 6$

For both cases, provided that  $N$  is large enough,  $N = 6$  there, the long-memory behavior quite always captured from the lowest octave  $j_1 = 2$ .

### 5.3 Short-Memory and Change-Points: Change-Point GARCH Process

#### Volatility, GARCH Process and Change-Points?

As mentioned in the introduction, a potentially useful application of wavelet methods is the analysis of volatility series, the volatility of asset prices  $P_t$  being commonly defined by either the absolute returns  $|r_t|$  or the squared returns  $r_t^2$ , where the returns  $r_t$  are defined as  $r_t = \ln(P_t/P_{t-1})$ .

Volatility processes are modeled using either GARCH type models, or stochastic volatility models, although GARCH type models are more often used as they are easy to estimate, parsimonious as a simple GARCH(1,1) model with three parameters can fit most of the series of asset returns, and have nice properties for asset pricing, see e.g., Duan (1995) and Kallsen and Taqqu (1998).

A standard GARCH(1,1) process is defined as

$$r_t = \sigma_t \varepsilon_t, \quad \varepsilon_t \sim N(0, 1), \quad \sigma_t^2 = \omega + \beta \sigma_{t-1}^2 + \gamma r_{t-1}^2. \quad (65)$$

The literature on the properties of GARCH-type processes is pretty impressive; interested readers are referred to the chapter by Giraitis, Leipus and Surgailis (2005) in this volume.

However, GARCH-type homogeneous processes have several drawbacks and inconsistencies:

1. They cannot fit the heavy tails of the returns  $r_t$ ,
2. When the sample size increases, the sum of estimated parameters  $\hat{\beta} + \hat{\gamma}$  tends to 1, a property called the “IGARCH effect”. As a consequence, the variance of the process  $\{r_t\}$  is infinite since  $\omega > 0$ ,
3. For large samples, the ACF of the series  $|r_t|$  and  $r_t^2$  behave like equation (2), this is the so called long-range dependence in the volatility.

Mikosch and Stărică (1999, 2003, 2004a, 2004b) emphasized in a series of research works that points 2 and 3 are inconsistent, as the ACF of the power transformation of a process with infinite variance is not properly defined. Thus, the GARCH(1,1) model might be the “right” model, provided that it is estimated on the sample for which the parameters of the process are constant, and that IGARCH, long-range dependence and tail effects might be the consequence of the non-stationarity of the GARCH process, and, for instance, of changes in the parameters so that the unconditional variance of the process is not constant. A major issue in practical time series analysis hence consists in being able to decide whether long-memory is truly present in the analyzed data or if a change-point type non-stationary property actually exists and is likely to be misinterpreted as long-memory.

### Discriminating Between True Long-Memory and Change-Points?

There is a large number of research works dealing with the issue of change-point detection, see e.g., Berkes *et al.* (2004), Berkes, Horváth and Kokoszka (2004), Kokoszka and Teyssière (2002), Mikosch and Stărică (1999, 2004b), Chu (1995). So far, none of the change-point tests proposed in research papers is based on wavelets. More precisely, the relevance of the use of the wavelet analysis in the context of non-homogeneous GARCH processes has not been studied, since as it is discussed below, it does not fit in the standard framework of change-point in the conditional mean processes. Such an analysis is the purpose of this section.

We consider a GARCH(1,1) process with change-point at time  $k$ :

$$\begin{aligned} r_t &= \sigma_t \varepsilon_t, \quad \varepsilon_t \sim N(0, 1), \\ \sigma_t^2 &= \omega + \beta \sigma_{t-1}^2 + \gamma r_{t-1}^2 \quad t \leq k, \\ \sigma_t^2 &= \omega^* + \beta^* \sigma_{t-1}^2 + \gamma^* r_{t-1}^2, \quad t > k. \end{aligned} \quad (66)$$

with  $\omega \neq \omega^*$  or  $\beta \neq \beta^*$  or  $\gamma \neq \gamma^*$ . The parameter of interest is the unconditional variance of the process  $\sigma^2 = \omega/(1 - \beta - \gamma)$ , and we consider two cases: small change and large change in the unconditional variance. We choose  $k = \lceil T/2 \rceil$  and set for the first part of the process

$$\omega = 0.1, \quad \beta = 0.3, \quad \gamma = 0.3, \quad \sigma^2 = 0.25, \quad (67)$$

while the parameters for the second part of the two processes are:

- **GARCH A** (small change):  $\omega^* = 0.125$ ,  $\beta^* = 0.6$ ,  $\gamma^* = 0.1$ ,  $\sigma^2 = 0.4667$ ,
- **GARCH B** (large change):  $\omega^* = 0.15$ ,  $\beta^* = 0.65$ ,  $\gamma^* = 0.25$ ,  $\sigma^2 = 1.5$ .

Since financial time series of size  $T = 10000$  are likely affected by several changes in regimes, we consider the following GARCH process, denoted by **GARCH C**, with two un-periodic changes in regimes at times  $k$  and  $k'$ :

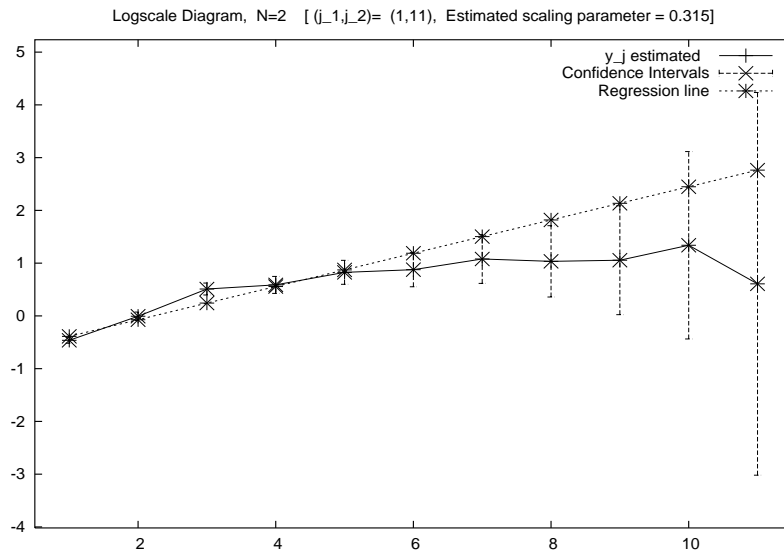
$$\begin{aligned} r_t &= \sigma_t \varepsilon_t, \quad \varepsilon_t \sim N(0, 1), \\ \sigma_t^2 &= 0.15 + 0.65\sigma_{t-1}^2 + 0.25r_{t-1}^2, \quad t \leq k \quad \sigma^2 = 1.5, \\ \sigma_t^2 &= 0.1 + 0.3\sigma_{t-1}^2 + 0.3r_{t-1}^2, \quad k < t \leq k', \quad \sigma^2 = 0.25, \\ \sigma_t^2 &= 0.25 + 0.6\sigma_{t-1}^2 + 0.2r_{t-1}^2, \quad t > k', \quad \sigma^2 = 1.25. \end{aligned} \quad (68)$$

We set  $k = \lceil T/6 \rceil$  and  $k' = \lceil 5T/6 \rceil$ , i.e., near the end-points of the sample, which is always the most difficult configuration to detect. Let us note that the absolute value and squares of change-point GARCH processes are not genuine change-point processes in the conditional mean, which makes the standard theory for change-point in the conditional mean not applicable.

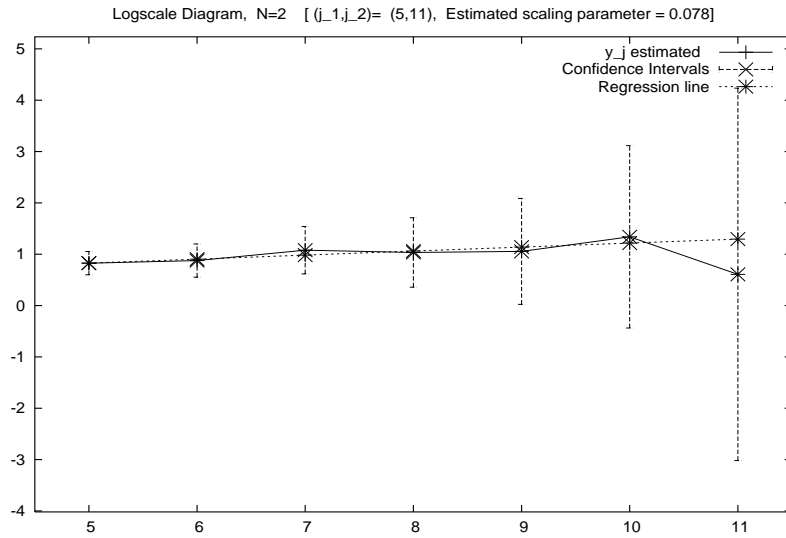
We consider also the performance of the estimators for the case without change points, i.e., the Data Generating Process (65) with parameters given by equation (67) that we denote as **DGP 0**.

Let us first analyze the logscale diagrams (presented in figures 21–26) for various realizations of the **GARCH A**, **GARCH B** and **GARCH C** processes. One can see that fitting the entire logscale diagram (i.e., starting from  $j_1 = 1$ ) would lead to the conclusion that long-memory exists in the data. However, we also see that the spurious long-memory property caused by the non-stationarity of the process is gotten rid-off by selecting the lowest octave  $j_1 \geq 5$ .

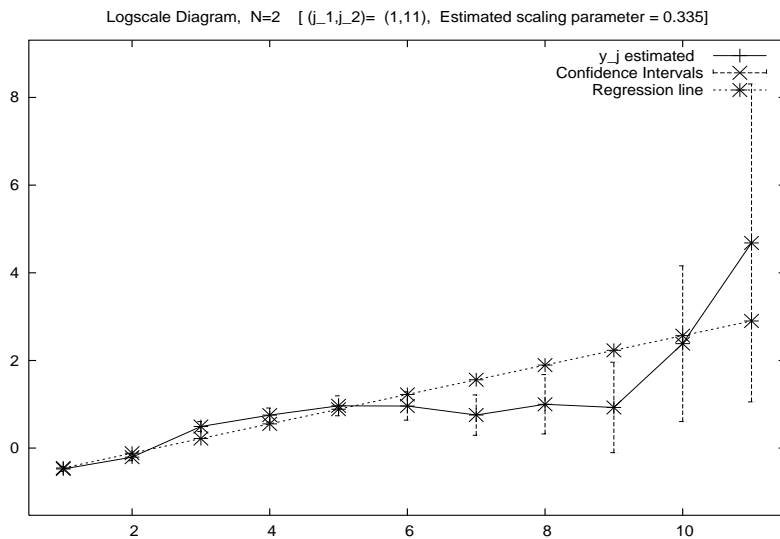
*Remark 11.* The optimal choice  $j_1^{RMSE}$  is the same for all values of the number of vanishing moments  $N$ . Thus, we cannot use lower  $j_1$  by increasing  $N$ .



**Fig. 21.** Logscale diagram for the realization of the absolute value **GARCH A** process. We select here  $j_1 = 1$ ,  $j_2 = 11$ ,  $N = 2$

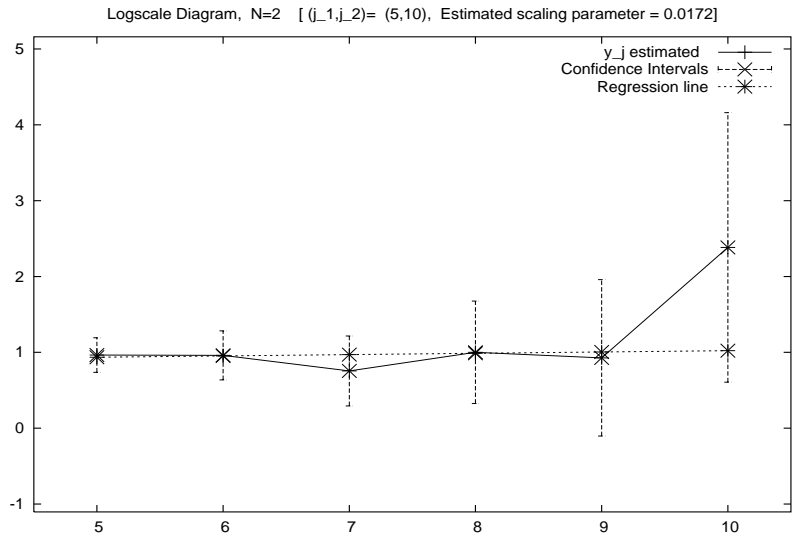


**Fig. 22.** Logscale diagram for the realization of the squares of a **GARCH A** process. We select here  $j_1 = 5$ ,  $j_2 = 11$ ,  $N = 2$

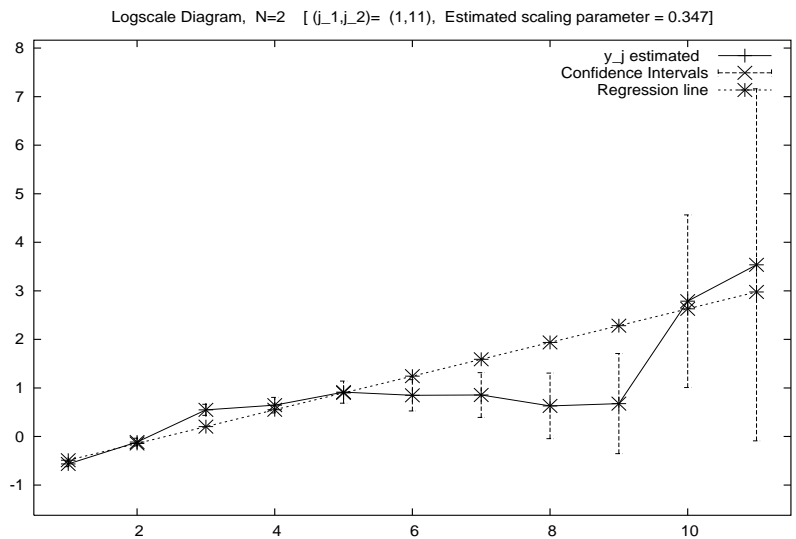


**Fig. 23.** Logscale diagram for the realization of the squares of a **GARCH B** process. We select here  $j_1 = 1$ ,  $j_2 = 11$ ,  $N = 2$

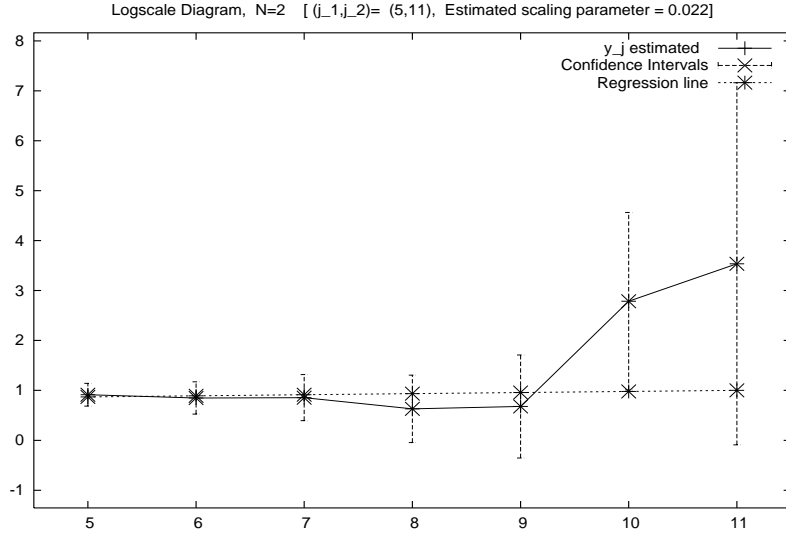




**Fig. 24.** Logscale diagram for the realization of the squares of a **GARCH B** process. We select here  $j_1 = 5$ ,  $j_2 = 10$ ,  $N = 2$



**Fig. 25.** Logscale diagram for the realization of the squares of a **GARCH C** process. We select here  $j_1 = 1$ ,  $j_2 = 11$ ,  $N = 2$



**Fig. 26.** Logscale diagram for the realization of the squares of a **GARCH C** process. We select here  $j_1 = 5$ ,  $j_2 = 11$ ,  $N = 2$

For the **GARCH A** process, i.e., with a constant unconditional variance, both the LW and the LP estimators do not detect any long-range dependence, while the wavelet estimator detects a very moderate intensity of long-memory, although the Gaussian confidence intervals contains the value zero (let us remind that  $\alpha \equiv 0$  corresponds to short memory processes). The wavelet estimate with the octave  $j_1$  selected using the optimal bandwidth, see equation (60), gives estimates from  $\hat{\alpha}_W = 0.0392$  to  $\hat{\alpha}_W = 0.0333$ , depending on the number of vanishing moments  $N$ .

**Table 15.** Estimation of the scaling parameter for change-points GARCH processes, wavelet estimators.  $T = 10000$ ,  $N = 2$

Model	$\hat{\alpha}_W^{RMSE}$	$j_1^{RMSE}$	$E\hat{\alpha}_W^{RMSE} - \alpha$	RMSE	$\hat{\alpha}_W^{(5)}$	$E\hat{\alpha}_W^{(5)} - \alpha$	RMSE
GARCH 0	0.0369	4	0.0369	0.0719	0.0151	0.0151	0.0866
GARCH A	0.0495	4	0.0495	0.0791	0.0241	0.0241	0.0900
GARCH B	0.0520	5	0.0520	0.1086	0.0865	0.0865	0.1238
GARCH C	0.0219	4	0.0219	0.0708	0.0682	0.0682	0.1126

There is a significant discrepancy between the estimators for both **GARCH B** and **GARCH C** processes. While the wavelet estimator does detect a very moderate intensity of long-range dependence, both the LW and LP estimators give a high value for  $\hat{\alpha}$ , with a RMSE far greater than the one of the

**Table 16.** Estimation of the scaling parameter for change-points GARCH processes, LP and LW estimators.  $T = 10000$ . The optimal bandwidths  $m_{LP}^{opt}$  and  $m_{LW}^{opt}$  are respectively used for the LP and LW estimators

Model	$\hat{\alpha}_{LP}^{opt}$	$E\hat{\alpha}_{LP}^{opt} - \alpha$	RMSE	$\hat{\alpha}_{LW}^{opt}$	$E\hat{\alpha}_{LW}^{opt} - \alpha$	RMSE
GARCH 0	0.0038	0.0038	0.0669	0.0015	0.0015	0.0612
GARCH A	0.0043	0.0043	0.4735	0.0065	0.0065	0.0651
GARCH B	0.1921	0.1921	1.4738	0.4015	0.4015	0.4043
GARCH C	0.2440	0.2440	1.4087	0.4188	0.4188	0.4227

wavelet estimator. For the LP, the results obtained with  $m = [T^{0.3}]$  are very high, as  $\hat{\alpha}_{LP} = 0.0595$  for **GARCH A**,  $\hat{\alpha}_{LP} = 1.4454$  for **GARCH B**, and  $\hat{\alpha}_{LP} = 1.3801$  for **GARCH C**. This advocates again the use of the optimal bandwidth (41) for volatility processes.

When selecting the lowest octave  $j_1$  that maximizes the goodness-of-fit function (61), the estimated long-memory parameter for the **GARCH B** process is  $\hat{\alpha}_W = 0.0539$  for  $N = 4$  with a maximum for  $\hat{\alpha}_W = 0.0877$  for  $N = 2$ . For the **GARCH C** process,  $\hat{\alpha}_W = 0.0069$  for  $N = 4$ , with a maximum  $\hat{\alpha}_W = 0.0723$  for  $N = 2$ . This shows that increasing  $N$  helps to get rid of spurious long-memory. The wavelet estimates with the lowest octave  $j_1$  given by equation (60) are equal to 0.0053 for the **GARCH A**, 0.0575 for the **GARCH B**, and  $-0.0281$  for the **GARCH C**, in all three cases this “optimal” octave is  $j_1 = 7$ , which yields a very large variance as well emphasizes the need for further research in this direction.

This example clearly shows that in spite of the octaves selection issue due to the nonlinearity and non Gaussianity of the sequence  $\{r_t^2\}$ , the wavelet estimator detects a very moderate level of long-range dependence, i.e.,  $\alpha < 0.0520$  in the worst case, while both the LW and the LP estimators are “fooled” by the occurrence of a large change in the unconditional variance for the short memory GARCH process. This advocates the use of the wavelet estimator, in conjunction with other estimators, for adjudicating between strong dependence and change-point for volatility processes.

#### 5.4 Long-Memory and Change-Points: The Non-Homogeneous Long-Memory Stochastic Volatility Process

We consider here a nonlinear process that displays long-range dependence in conditional variance with a change-point, defined as:

$$\begin{aligned}
 r_t &= \sigma_t \zeta_t, \quad \zeta_t \sim N(0, 1), \quad \sigma_t = \sigma \exp(X_t/2), \quad X_t \sim \text{FARIMA}(p, d, q), \\
 E X_t &= 0, \quad t \leq k, \quad E X_t = \mu, \quad t > k,
 \end{aligned} \tag{69}$$

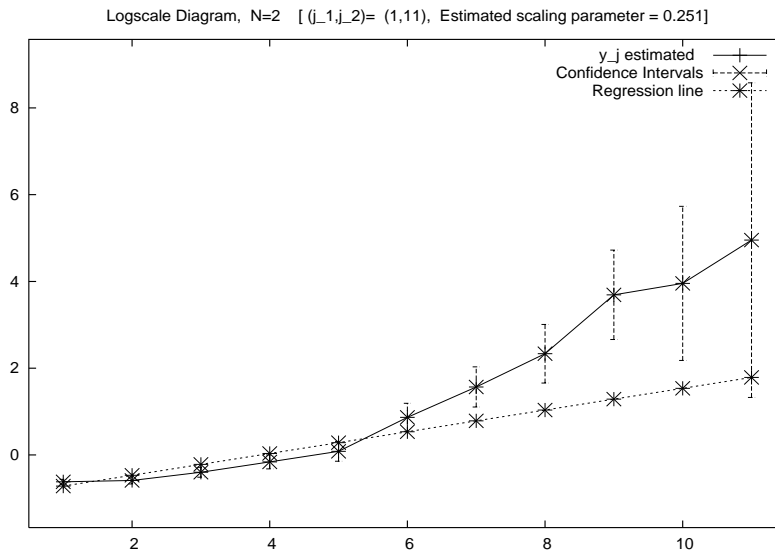
i.e., the process  $\{X_t\}$  mixes long-range dependence and a change-point at time  $k$ . We consider here samples of size  $T = 10000$ ,  $k = [T/2]$ , and  $\mu = 2.0$ .

*Remark 12.* The magnitude of the jump  $\mu = 2.0$  has to be compared with the square root of the variance of the  $X_t$  process,

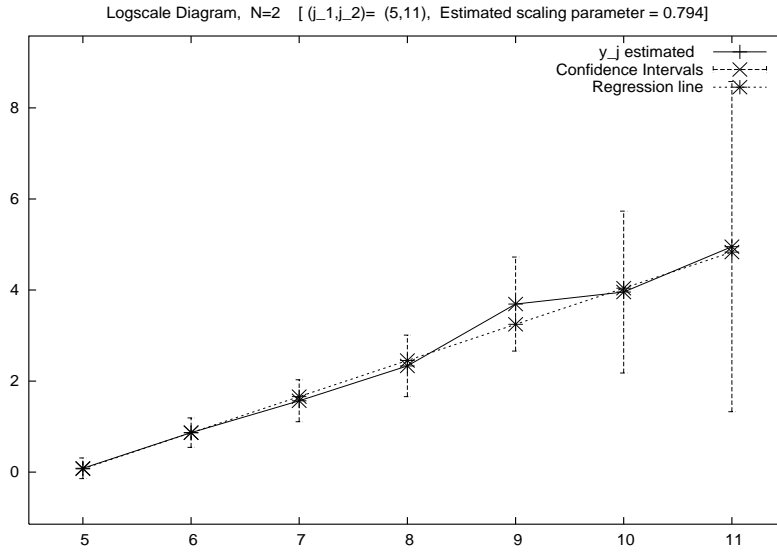
$$\text{Var}(X_t) = \sigma_\epsilon^2 \frac{\Gamma(1 - 2d)}{\Gamma^2(1 - d)}.$$

Here, we set  $\sigma_\epsilon^2 = 1$ , so that  $\sqrt{\text{Var}(X_t)}$  ranges from 1.0022 ( $d = 0.05$ ) to 1.9085 ( $d = 0.45$ ), i.e., the change is of significant magnitude.

Figures 27 and 28 display the logscale diagram for one realization of a non-homogeneous LMSV process for different regression ranges  $[j_1, j_2]$ . Provided that  $j_1 \geq 5$ , the long-range dependent behavior is satisfactorily captured.



**Fig. 27.** Logscale diagram for the realization of a change-point LMSV process, with  $\alpha = 0.90$ . We select here  $j_1 = 1, j_2 = 11, N = 2$



**Fig. 28.** Logscale diagram for the realization of a change-point LMSV process, with  $\alpha = 0.90$ . We select here  $j_1 = 5, j_2 = 11, N = 2$

**Table 17.** Estimation of the scaling parameter for change-point LMSV processes, wavelet estimator.  $T = 10000, N = 2$

$\alpha$	$\hat{\alpha}_W^{RMSE}$	$j_1^{RMSE}$	$E\hat{\alpha}_W^{RMSE} - \alpha$	RMSE	$\hat{\alpha}_W^{(6)}$	$E\hat{\alpha}_W^{(6)} - \alpha$	RMSE
0.10	0.0959	3	-0.0041	0.0317	0.4740	0.3740	0.3878
0.20	0.1777	4	-0.0223	0.0509	0.4748	0.2748	0.2939
0.30	0.3079	5	0.0079	0.0690	0.4820	0.1820	0.2109
0.40	0.3385	5	-0.0605	0.0925	0.4984	0.0984	0.1473
0.50	0.5274	6	0.0274	0.1164	0.5274	0.0274	0.1164
0.60	0.5720	6	-0.0280	0.1120	0.5720	-0.0280	0.1120
0.70	0.6337	6	-0.0663	0.1367	0.6337	-0.0663	0.1367
0.80	0.7122	6	-0.0878	0.1496	0.7122	-0.0878	0.1496
0.90	0.8047	6	-0.0953	0.1546	0.8047	-0.0953	0.1546

**Table 18.** Estimation of the scaling parameter for change–point LMSV processes, LW and LP estimators.  $T = 10000$ . The optimal bandwidths  $m_{LP}^{opt}$  and  $m_{LW}^{opt}$  are respectively used for the LP and LW estimators

$\alpha$	$\hat{\alpha}_{LP}^{opt}$	$E\hat{\alpha}_{LP}^{opt} - \alpha$	RMSE	$\hat{\alpha}_{LW}^{opt}$	$E\hat{\alpha}_{LW}^{opt} - \alpha$	RMSE
0.10	0.5308	0.4308	0.4365	0.9026	0.8026	0.8034
0.20	0.5377	0.3377	0.3451	0.8931	0.6931	0.6940
0.30	0.5514	0.2514	0.2617	0.8829	0.5829	0.5842
0.40	0.5751	0.1751	0.1899	0.8734	0.4734	0.4751
0.50	0.6103	0.1103	0.1334	0.8657	0.3657	0.3684
0.60	0.6577	0.0577	0.0968	0.8625	0.2625	0.2678
0.70	0.7173	0.0173	0.0826	0.8676	0.1676	0.1795
0.80	0.7877	-0.0123	0.0851	0.8864	0.0064	0.1160
0.90	0.8675	-0.0325	0.0947	0.9251	0.0251	0.0927

Since the sequence of pseudo–error terms is the same for LMSV processes with and without change–point, the results of Tables 17 and 18 can be directly compared to those of Tables 5 and 6. The wavelet estimator is affected by the presence of this change point, but in a far lower extent than both the LW and LP estimators. What is really informative here is the strong discrepancy between the wavelet estimator and the spectral based estimators, which suggests the presence of a break. For that case, the use of the optimal bandwidth for the LP estimator is justified as the mean estimates obtained for  $m = [T^{0.3}]$  are between 1.4114 and 1.8578, with a RMSE ranging from 0.6349 to 1.7718.

*Remark 13.* The best results, i.e., with the lowest RMSE, are obtained with  $N = 2$ .

## 6 Financial Time Series

### 6.1 Intra–day Foreign Exchange (FX) Rates

• **Financial data.** We consider four high–frequency time series on intra–day FX rates provided by Olsen and Associates, i.e., the US dollar–Swiss franc (USD–CHF), the US dollar–Japanese yen (USD–JPY), the US dollar–German deutsche mark (USD–DEM), and the British pound–US dollar (GBP–USD). The data are observed for the whole year 1996 every 30 minutes in a time scale denoted as  $\vartheta$ –time, where all intra–day periodic components have been removed: time scale with high volatility (activity) are expanded while time scale with low activity are shortened. The  $\vartheta$ –time scale can be interpreted as a business scale, and then removes the seasonality in the volatility process; see Dacorogna *et al.* (1993) for further details. Since the activity (volatility) is not the same for the series considered, they do not have the same time scale.

We consider here the logarithmic middle price  $x(t)$  defined by

$$x_t \equiv \frac{x_t^{bid} + x_t^{ask}}{2} \quad \text{with} \quad x_t^{bid} = \ln(p_t^{bid}), \quad x_t^{ask} = \ln(p_t^{ask}), \quad (70)$$

where  $p_t^{bid}$  and  $p_t^{ask}$  respectively denote the “bid” and “ask” price at time  $t$ . This variable behaves symmetrically when the price is inverted. Next, we define the returns  $r_t$  as

$$r_t = x_t - x_{t-\Delta t}, \quad (71)$$

where  $\Delta t$  represents the time interval between two consecutive observations; here,  $\Delta t = 30$  minutes. The sample sizes of the series are quite large, as  $T = 17524, 17529, 17508$  and  $17520$  for the log of returns on USD–DEM, USD–CHF, USD–JPY and GBP–USD respectively.

Figures 29 and 30 below display the ACF of the series of absolute returns  $|r_t|$  and squared returns  $r_t^2$  on US dollar–Japanese yen, up to the order 2500. The hyperbolic decay of the ACF is typical of strongly dependent processes, and is slower for the series of absolute returns than for the series of squared returns: this is the so-called “Taylor effect”, i.e., the persistence of the series  $|r_t|^\delta$  is the strongest for  $\delta = 1$ . Unlike daily returns, intra–returns are correlated, this negative correlation, which can be modeled by antipersistent processes, is the consequence of market microstructure effects.

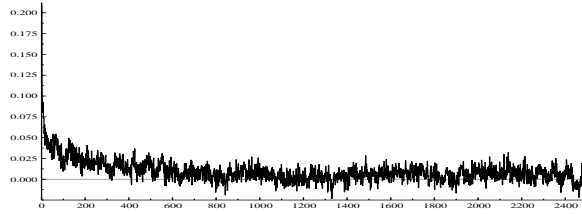


Fig. 29. ACF of absolute returns on US dollar-Japanese yen FX rate

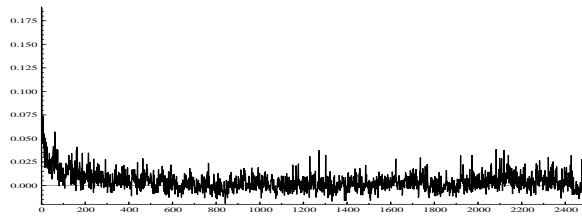


Fig. 30. ACF of squared returns on US dollar-Japanese yen FX rate

• **Long–Memory Analysis.** The estimation of the long–memory parameter for the series of high frequency financial time series has been considered

by Müller *et al.* (1990). A wavelet analysis of financial volatilities has been conducted in Kokoszka and Teyssière (2002) and Teyssière (2003). They estimated the degree of long-range dependence of absolute and squared returns on FX rates and equities using the wavelet estimator by Veitch and Abry (1999) and the local Whittle estimator. Their empirical results show that the scaling parameter estimated with the wavelet estimator is far lower than the one obtained with the local Whittle estimator.

In the present work, we estimate the degree of persistence for the series of absolute returns and squared returns for the four intra-day FX rates, denoted by  $|r_{usd-chf}|$ ,  $|r_{usd-dem}|$ ,  $|r_{usd-jpy}|$ ,  $|r_{gbp-usd}|$ ,  $r_{usd-chf}^2$ ,  $r_{usd-dem}^2$ ,  $r_{usd-jpy}^2$  and  $r_{gbp-usd}^2$ . The choices for  $j_1$  and  $N$  are guided by the inspection of the logscale diagrams.

We observe that the wavelet-based estimates of the long-memory parameter are significantly lower than the spectral-based estimates, confirming the earlier finding mentioned above. Because the wavelet estimator benefits from robustness against non-stationarities (change points and trends), such discrepancies are in favor of the conjecture that the empirical high intensity of strong dependence in asset price volatilities is a statistical artefact caused by the occurrence of change-point(s). However, the results obtained here also indicate that the estimates significantly depart from 0, i.e., the volatility series do display long-range dependence. This has implications for forecasting purposes, as mentioned by Granger and Hyung (2004) who mix both long-memory and change-point processes for forecasting absolute returns.



**Table 19.** Estimation of the long-memory parameter for volatilities of intra-day FX rates,  $N = 2$ . Standard errors are between parentheses, every other line. The optimal bandwidths  $m_{LP}^{opt}$  and  $m_{LW}^{opt}$  are respectively used for the LP and LW estimators

Series	$\hat{\alpha}_W$	$j_1$	$j_2$	$\hat{\alpha}_W^m$	$\hat{\alpha}_{LP}^{opt}$	$m_{LP}^{opt}$	$\hat{\alpha}_{LW}^{opt}$	$m_{LW}^{opt}$
$ r_{usd-chf} $	0.4309 (0.0594)	5	9	0.4837 (0.0541)	0.5956 (0.0547)	549	0.5255 (0.0443)	510
$ r_{usd-dem} $	0.4511 (0.0520)	5	11	0.4586 (0.0804)	0.5863 (0.0543)	557	0.5637 (0.0461)	470
$ r_{usd-jpy} $	0.3581 (0.0381)	4	9	0.4017 (0.0572)	0.4981 (0.0550)	543	0.5597 (0.0463)	467
$ r_{gbp-usd} $	0.2923 (0.0526)	5	11	0.3029 (0.0541)	0.4126 (0.0544)	555	0.5335 (0.0472)	447
$r_{usd-chf}^2$	0.3088 (0.0242)	3	10	0.3056 (0.0346)	0.5808 (0.0526)	594	0.5548 (0.0406)	604
$r_{usd-dem}^2$	0.2543 (0.0376)	4	9	0.2663 (0.0346)	0.2583 (0.0486)	694	0.3250 (0.0365)	749
$r_{usd-jpy}^2$	0.2566 (0.0513)	5	12	0.2794 (0.0572)	0.4163 (0.0533)	577	0.4351 (0.0436)	527
$r_{gbp-usd}^2$	0.0982 (0.0603)	5	9	0.1204 (0.0346)	0.2796 (0.0467)	753	0.2859 (0.0357)	784

**Table 20.** Wavelet estimates of the long-memory parameter for volatilities of intra-day FX, for different values of the number of moments  $N$ . Standard errors are between parentheses, every other line.

Series	$N = 3$			$N = 4$			$N = 5$			$N = 6$		
	$\hat{\alpha}_W$	$j_1$	$j_2$	$\hat{\alpha}_W$	$j_1$	$j_2$	$\hat{\alpha}_W$	$j_1$	$j_2$	$\hat{\alpha}_W$	$j_1$	$j_2$
$ r_{usd-chf} $	0.4365 (0.0380)	4	9	0.4137 (0.0385)	4	9	0.4605 (0.0621)	5	9	0.4630 (0.0631)	5	9
$ r_{usd-dem} $	0.4924 (0.0541)	5	11	0.4277 (0.0573)	5	11	0.4371 (0.0963)	6	10	0.4328 (0.1005)	6	10
$ r_{usd-jpy} $	0.3591 (0.0418)	4	8	0.3449 (0.0426)	4	8	0.3776 (0.0383)	4	10	0.3879 (0.0391)	4	10
$ r_{gbp-usd} $	0.2569 (0.0356)	4	11	0.2125 (0.0581)	5	10	0.2078 (0.0600)	5	10	0.2126 (0.0401)	4	9
$r_{usd-chf}^2$	0.3831 (0.0362)	4	10	0.3856 (0.0374)	4	10	0.4765 (0.0599)	5	10	0.4400 (0.0617)	5	10
$r_{usd-dem}^2$	0.2340 (0.0240)	3	11	0.2977 (0.0573)	5	11	0.2415 (0.0600)	5	10	0.2871 (0.0617)	5	10
$r_{usd-jpy}^2$	0.2851 (0.0380)	4	11	0.2618 (0.0390)	4	9	0.2871 (0.0383)	4	10	0.2616 (0.0391)	4	10
$r_{gbp-usd}^2$	0.1208 (0.0541)	5	11	-0.0324 (0.0573)	5	11	0.1541 (0.0383)	4	10	-0.0217 (0.0391)	4	10

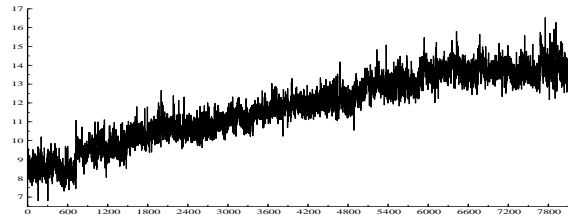
*Remark 14.* We observe that the estimated scaling parameter does not significantly vary for different values of the number of moments  $N$ , which confirms the presence of long-range dependence in the volatility series. The series of squared returns  $r_{gbp-usd}^2$  on dollar/pound FX rate, is a noticeable exception; in fact, this highly nonlinear series was always very difficult to fit, see e.g., Gallant, Hsieh and Tauchen (1991).

## 6.2 Application to Trading Volume

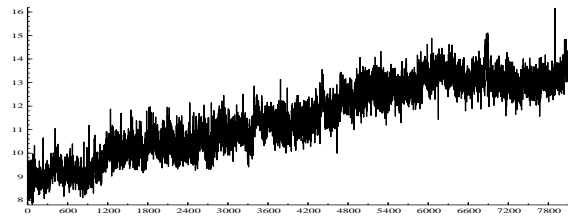
• **Bivariate Time Series.** Lobato and Velasco (2000) considered the bivariate process (volatility, log trading volume) of asset prices of stocks constituting the Dow Jones Average Industrial Index, observed from July 1962 to December 1994, i.e., 8176 daily observations, and investigated the issue of common degree of long-range dependence for these two processes.

We consider here few of the series used by Lobato and Velasco (2000), and check whether the wavelet based estimator gives the same results, i.e., the commonality of strong dependence between volatility and log of trading volume.

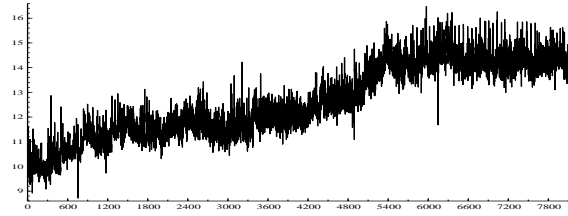
Figures 31, 32, and 33 below display the log trading volume on three stocks, Eastman Kodak, Chevron and ATT. These three series display strong and clear upward trends.



**Fig. 31.** Logarithm of trading volume on Eastman Kodak stock



**Fig. 32.** Logarithm of trading volume on Chevron stock



**Fig. 33.** Logarithm of trading volume on ATT stock

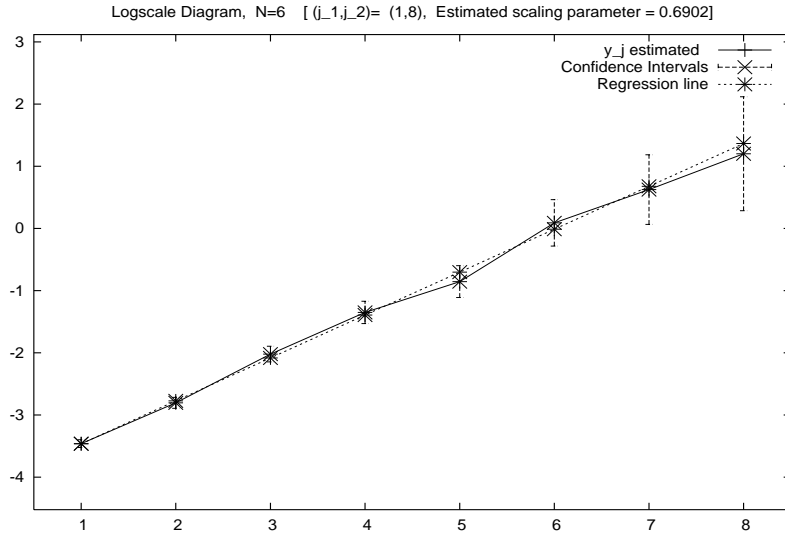
• **Long–Memory Analysis.** We estimate the long–memory parameters for the log of trading volume and the absolute value of returns of six stocks: AT & T Corp., Allied Signal Inc., Chevron Corp., Eastman Kodak Corp., Exxon Corp. and IBM.

**Table 21.** Estimation of the scaling parameter  $\alpha$  for log of trading volume, and absolute value of returns on stocks. Standard errors are between parentheses, every other line. The optimal bandwidth  $m_{LW}^{opt}$  is used for the LW estimator

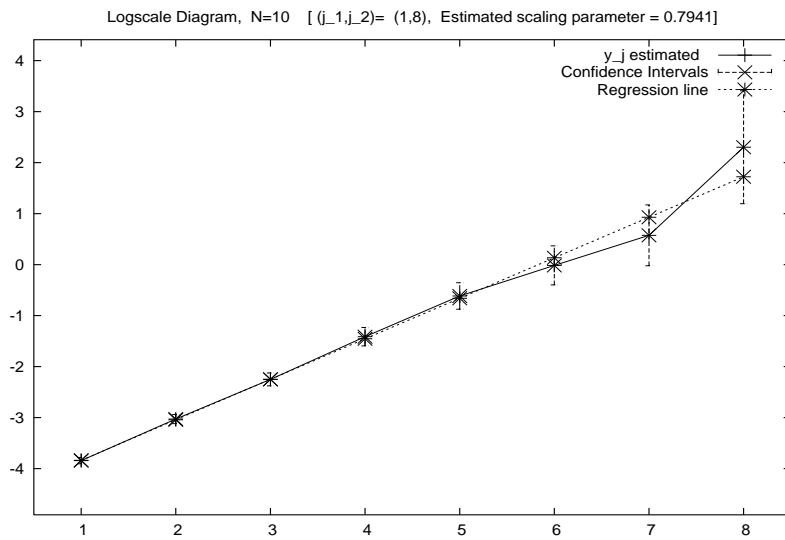
Stocks	$\hat{\alpha}_W$ (Volume)	$N$	$j_1$	$j_2$	$\hat{\alpha}_W ( r_t )$	$j_1$	$j_2$	$\hat{\alpha}_{LW} ( r_t )$
AT & T	0.7632	6	1	9	0.6422	5	10	0.3465
	(0.0179)				(0.0871)			(0.0533)
Allied Signal	0.6564	6	1	9	0.3126	5	8	0.4546
	(0.0181)				(0.1051)			(0.0546)
Chevron	0.6548	6	1	9	0.3715	7	10	0.5108
	(0.0181)				(0.2530)			(0.0572)
Eastman Kodak	0.7051	6	1	9	0.2484	5	8	0.4146
	(0.0181)				(0.1051)			(0.0516)
Exxon	0.7151	6	1	9	0.5331	6	10	0.3088
	(0.0182)				(0.1431)			(0.0515)
IBM	0.8532	6	1	9	0.3626	4	10	0.2553
	(0.0181)				(0.0556)			(0.0462)

For the series of volume, estimation results no longer differ as soon as  $N > 6$ , see also Table 22. As we can see from figures 34 and 35, the logscale diagrams consist of straight lines for all octaves, except for Chevron, see Figure 36. Together with the estimated values of the long–memory parameter, this tells that there does exist long–memory in the trading volume times series. The wavelet based estimator is not fooled by the obvious trends that also exist on top of long–memory. We also observe that the estimated long–memory parameters for the series of absolute returns are lower than the estimated scaling parameters for the series of log of volume. In fact, logscale diagrams

for the series of absolute returns are more difficult to interpret than the ones of the series of log-volume.



**Fig. 34.** Logscale diagram for the logarithm of volume on AT& T ,  $N = 6$



**Fig. 35.** Logscale diagram for the logarithm of volume on IBM ,  $N = 10$

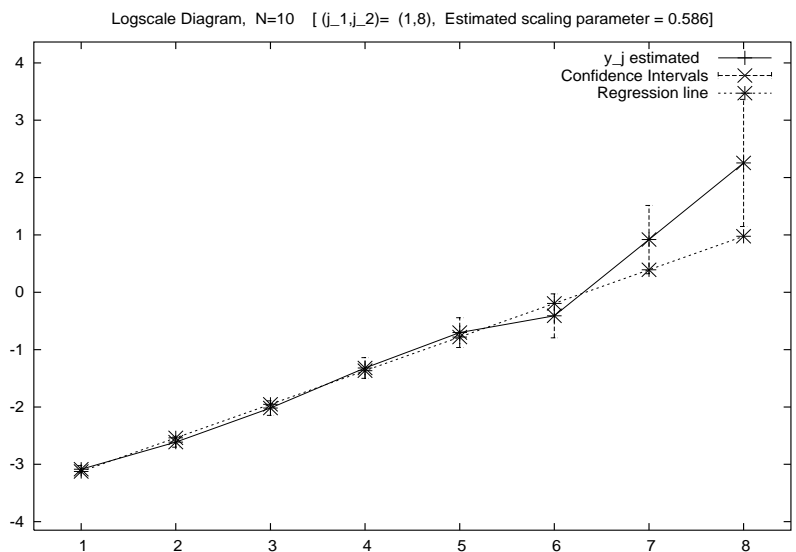


Fig. 36. Logscale diagram for the logarithm of volume on Chevron , N = 10

Table 22. Estimation of the scaling parameter  $\alpha$  for log of trading volume, and absolute value of returns on stocks, for different values of the number of moments  $N$ . Standard errors are between parentheses, every other line

Series	N = 7		N = 8		N = 9		N = 10	
	$\hat{\alpha}_W$	$j_1 j_2$	$\hat{\alpha}_W$	$j_1 j_2$	$\hat{\alpha}_W$	$j_1 j_2$	$\hat{\alpha}_W$	$j_1 j_2$
AT & T	0.7548 (0.0180)	1 9	0.7475 (0.0183)	1 8	0.7518 (0.0187)	1 8	0.7337 (0.0189)	1 8
Allied Signal	0.6405 (0.0180)	1 9	0.6385 (0.0183)	1 8	0.6367 (0.0187)	1 8	0.6316 (0.0189)	1 8
Chevron	0.6459 (0.0180)	1 9	0.6435 (0.0183)	1 8	0.6432 (0.0187)	1 8	0.6384 (0.0189)	1 8
Eastman Kodak	0.7118 (0.0180)	1 9	0.7041 (0.0183)	1 8	0.6896 (0.0187)	1 8	0.6896 (0.0189)	1 8
Exxon	0.7070 (0.0180)	1 9	0.7099 (0.0183)	1 8	0.7046 (0.0187)	1 8	0.7089 (0.0189)	1 8
IBM	0.8351 (0.0180)	1 9	0.8451 (0.0183)	1 8	0.8301 (0.0187)	1 8	0.8235 (0.0189)	1 8

## 7 Conclusion

In nominal situations, i.e., when the analyzed time series corresponds to an actual non corrupted long-range dependent process, the wavelet estimator for the long-memory parameter works satisfactorily well and this for a large variety of linear and nonlinear strongly dependent processes in conditional mean and in conditional variance. In the same situations, the local Whittle estimator slightly outperforms the log periodogram and wavelet estimators whose performance are mostly comparable. From our point of view, the key advantage of the wavelet based estimator lies in the possibility of varying  $N$  which indeed constitutes a key degree of freedom in the wavelet based analysis of long-memory: as long as the logscale diagrams and the corresponding estimates of the long-memory parameter vary with  $N$ , it shows that non-stationarities exist in the time series and may impair a correct analysis of the long-memory or confused with long-memory. When, for large enough  $N$ , no variation with  $N$  are observed, we know we have untangled long-memory from non-stationarity and can accurately estimate the corresponding parameter. Section 5 consistently showed that the wavelet estimator significantly outperformed the two other estimators in situations where non-stationarities were added to long-memory or confused with.

Also, for financial time series, the possibility of varying  $N$  enabled to show that the intensity of the long-range dependence of volatility processes is lower than the one measured with spectral based semiparametric estimators. Thus, it reveals that volatility processes mix non-stationarities with long-range dependence. It also showed however that long-memory truly exists both in the volatility and volume time series. The wavelet analysis of the series of volume and volatility finally showed that the dynamics of volatility processes is more complex than the one of volume processes.

## References

1. Abry, P., Gonçalves P. and Flandrin, P. (1995). Wavelets, spectrum analysis and  $1/f$  processes, in: Antoniadis, A. and Oppenheim, G. (Eds.), *Wavelets and Statistics*, Lecture Notes in Statistics, Vol. 103. Springer Verlag, New York, pp. 15–30.
2. Abry P. and Veitch, D. (1998). Wavelet analysis of long-range dependent traffic. *IEEE Transactions on Information Theory*, **44**, 2–15.
3. Abry, P., Flandrin, P., Taqqu, M. and Veitch, D. (2000). Wavelets for the analysis, estimation, and synthesis of scaling data, in: Park, K. and Willinger, W. (Eds.), *Self Similar Network Traffic and Performance Evaluation*. Wiley, pp. 39–88.
4. Abry, P., Flandrin, P., Taqqu, M. and Veitch, D. (2003). Self-similarity and long-range dependence through the wavelet lens, in: Doukhan, P., Oppenheim, G. and Taqqu, M.S. (Eds.), *Long-Range Dependence: Theory and Applications*. Birkhauser, Boston, pp. 527–556.

5. Abry, P. and Teyssière, G. (2005). Wavelet Estimation of the Scaling Parameter of Prefiltered Discrete Time Nonlinear LRD Processes. *Preprint*.
6. Bardet, J.M., Lang, G., Moulines, E. and Soulier, P. (2000). Wavelet estimator of long-range dependent processes. *Statistical Inference for Stochastic Processes*, **3**, 85–99
7. Belsley, D., Kuh, E. and Welsch, R.E., (1980). *Regression Diagnostics*, Wiley, New York.
8. Beran, J. (1994). *Statistics for Long-Memory Processes*. Chapman & Hall, New York.
9. Berkes, I., Horváth, L. and Kokoszka, P.S. (2004). Testing for parameter constancy in GARCH( $p, q$ ) models. *Statistics and Probability Letters*, in press.
10. Berkes, I., Gombay, E., Horváth, L. and Kokoszka, P.S. (2004). Sequential change-point detection in GARCH( $p, q$ ) models. *Econometric Theory*, **20**, 1140–1167.
11. Berkes, I., Horváth, L., Kokoszka, P.S. and Shao, Q-M. (2003). On discriminating between long-range dependence and changes in mean. *Preprint*.
12. Breidt, J., Crato, N. and de Lima, P. (1998). On the detection and estimation of long memory in stochastic volatility. *Journal of Econometrics*, **83**, 325–348.
13. Chu, C.-S.J. (1995). Detecting parameter shift in GARCH models. *Econometric Reviews*, **14**, 241–266.
14. Dacorogna, M.M., Müller, U.A., Nagler, R.J., Olsen, R.B. and Pictet, O.V. (1993). A geographical model for the daily and weekly seasonal volatility in the FX Market. *Journal of International Money and Finance*, **12**, 413–438.
15. Daubechies, I. (1992). *Ten Lectures on Wavelets*. SIAM Philadelphia, CBMS-NSF series, vol. **61**.
16. Delgado, M. and Robinson, P. (1996). Optimal spectral bandwidth for long memory. *Statistica Sinica*, **67**, 97–112.
17. Deo, R.S. and Hurvich, C.M. (2003). Estimation of long-memory in volatility, in: Doukhan, P., Oppenheim, G. and Taqqu, M.S. (Eds.), *Long-Range Dependence: Theory and Applications*. Birkhauser, Boston, pp. 313–324.
18. Dittman, I. and Granger, C.W.J. (2002). Properties of nonlinear transformations of fractionally integrated processes. *Journal of Econometrics*, **110**, 113–133.
19. Duan, J.-C. (1995). The GARCH option pricing model. *Mathematical Finance*, **5**, 13–32.
20. Flandrin, P. (1992). Wavelet analysis and synthesis of fractional Brownian motion. *IEEE Transactions on Information Theory*, **38**, 910–917.
21. Flandrin, P. (1989). On the spectrum of fractional Brownian motion. *IEEE Transactions on Information Theory*, **35**, 197–199.
22. Gallant, A.R., Hsieh, D.A. and Tauchen, G. (1991). On fitting a recalcitrant series: the pound/dollar exchange rate, 1974-1983, in: Barnett, W.A., Powell, J. and Tauchen, G. (Eds.), *Nonparametric and Semiparametric Methods in Economics and Statistics*. Cambridge University Press.
23. Gefferth, A., Veitch, D., Maricza, I., Molnár, S. and Ruzsa, I. (2003). The nature of discrete second-order self similarity. *Advances in Applied Probability*, **35**, 395–416.
24. Geweke, J. and Porter-Hudak, S. (1983). The estimation and application of long memory time series models. *Journal of Time Series Analysis*, **4**, 221–238.
25. Giraitis, L. and Surgailis, D. (2005). Lectures on long-memory processes. *In preparation*

26. Giraitis, L. and Surgailis, D. (1985). Central limit theorems and other limit theorems for functionals of Gaussian processes. *Z. Wahrscheinlichkeitstheorie verw. Gebiete* **70**, 191–212.
27. Giraitis, L., Leipus, R. and Surgailis, D. (2005). Recent advances in ARCH modelling, in: Teyssière, G. and Kirman, A. (Eds.), *Long-Memory in Economics*. Springer Verlag, Berlin. Appears in this volume.
28. Giraitis, L., Robinson, P. and Surgailis, D. (2000). A model for long memory conditional heteroskedasticity. *Annals of Applied Probability*, **10**, 1002–1024.
29. Giraitis, L., Kokoszka, P.S. and Leipus, R. (2000). Stationary ARCH Models: Dependence Structure and Central Limit Theorem. *Econometric Theory*, **16**, 3–22.
30. Giraitis, L., Kokoszka, P., Leipus, R. and Teyssière, G. (2000). Semiparametric estimation of the intensity of long-memory in conditional heteroskedasticity. *Statistical Inference for Stochastic Processes*, **3**, 113–128.
31. Giraitis, L., Robinson, P. and Samarov, A. (1997). Rate optimal semiparametric estimation of the memory parameter of the Gaussian time series with long-range dependence. *Journal of Time Series Analysis*, **18**, 49–60.
32. Granger, C.W.J. and Hyung, N. (2004). Occasional structural breaks and long memory with an application to the S&P 500 absolute stock returns. *Journal of Empirical Finance*, **11**, 399–421.
33. Granger, C.W.J. (2002). Long memory, volatility, risk and distribution. *Preprint*.
34. Granger, C.W.J. (2000). Current perspectives on long memory processes. *Academia Economic Papers*, **28**, 1–16.
35. Granger, C.W.J. and Ding, Z. (1995). Some properties of absolute returns, an alternative measure of risk. *Annales d'Économie et de Statistique*, **40**, 67–91.
36. Hauser, M.A. (1999). Maximum likelihood estimators for ARMA and ARFIMA models: a Monte Carlo study. *Journal of Statistical Planning and Inference*, **80**, 229–255.
37. Henry, M. (2005). Bandwidth choice, optimal rates and adaptivity in semiparametric estimation of long memory, in: Teyssière, G. and Kirman, A. (Eds.), *Long-Memory in Economics*. Springer Verlag, Berlin. Appears in this volume.
38. Henry, M. (2001). Robust automatic bandwidth for long-memory. *Journal of Time Series Analysis*, **22**, 293–316.
39. Henry, M. and Robinson, P.M. (1996). Bandwidth choice in Gaussian semiparametric estimation of long-range dependence, in: Robinson, P.M. and Rosenblatt, M. (Eds.), *Athens Conference on Applied Probability and Time Series Analysis, Time Series Analysis, In Memory of E.J. Hannan*. Lecture Notes in Statistics, Vol. 115. Springer Verlag, New York, pp. 220–232.
40. Hurvich, C., Deo, R. and Brodsky, J. (1998). The mean squared error of Geweke and Porter-Hudak's estimator of the memory parameter of a long memory time series. *Journal of Time Series Analysis*, **19**, 19–46.
41. Kallsen, J. and Taqqu, M.S. (1998). Option pricing in ARCH-type models. *Mathematical Finance*, **8**, 13–26.
42. Kokoszka, P.S. and Teyssière, G. (2002). Change-point detection in GARCH models: asymptotic and bootstrap tests. Under revision for the *Journal of Business and Economic Statistics*.
43. Kokoszka, P.S. and Leipus, R. (2002). Detection and estimation of changes in regime, in: Doukhan, P., Oppenheim, G. and Taqqu, M.S. (Eds.), *Long-Range Dependence: Theory and Applications*. Birkhauser, Boston, pp. 325–337.



44. Künsch, H.R. (1987). Statistical Aspects of Self-Similar Processes, in Prohorov, Yu. and Sazanov, V.V. (Eds), *Proceedings of the First World Congress of the Bernoulli Society*, **1**, 67–74. VNU Science Press, Utrecht.
45. Lang, G. and Roueff, F. (2001). Semiparametric estimation of the Hölder exponent of a stationary Gaussian process with minimax rates. *Statistical Inference for Stochastic Processes*, **4**, 283–306.
46. Lavielle, M. (1999). Detection of multiple changes in a sequence of dependent variables. *Stochastic Processes and their Applications*, **83**, 79–102.
47. Lavielle, M. and Moulines, E. (2000). Least-squares estimation of a unknown number of shifts in a time series. *Journal of Time Series Analysis*, **21**, 33–59.
48. Lavielle, M. and Teyssière, G. (2005). Adaptive detection of multiple change-points in asset price volatility, in: Teyssière, G. and Kirman, A. (Eds.), *Long-Memory in Economics*. Springer Verlag, Berlin. Appears in this volume.
49. Lobato, I.N., and Velasco, C. (2000). Long memory in stock market trading volume. *Journal of Business and Economic Statistics*, **18**, 410–427.
50. Mallat, S. (1998). *A Wavelet Tour of Signal Processing*. Academic Press, Boston.
51. Mikosch, T. and Stărică, C. (2004a). Non-stationarities in financial time series: the long-range dependence and the IGARCH effects. *Review of Economics and Statistics*, **86**, 278–290.
52. Mikosch, T. and Stărică, C. (2004b). Change of structure in financial time series and the GARCH model. *REVSTAT Statistical Journal*, **2**, 41–73.
53. Mikosch, T. and Stărică, C. (2003). Long-range dependence effects and ARCH modeling, in: Doukhan, P., Oppenheim, G. and Taqqu, M.S. (Eds.), *Long-Range Dependence: Theory and Applications*. Birkhauser, Boston, pp. 439–459.
54. Mikosch, T. and Stărică, C. (1999). Change of structure in financial time series, long range dependence and the GARCH model. *Preprint*.
55. Moulines, E. and Soulier, P. (2003). Semiparametric spectral estimation for fractional processes, in: Doukhan, P., G. and Taqqu, M.S. (Eds.), *Long-Range Dependence: Theory and Applications*. Birkhauser, Boston, pp. 251–301.
56. Müller, U.A., Dacorogna, M.M., Olsen, R.B., Pictet, O.V., Schwarz, M. and Morgeneegg, C. (1990). Statistical study of foreign exchange rates, empirical evidence of a price change scaling law, and intraday analysis. *Journal of Banking and Finance*, **14**, 1189–1208.
57. Robinson, P.M. (1995a). Gaussian semiparametric estimation of long-range dependence. *The Annals of Statistics*, **23**, 1630–1661.
58. Robinson, P.M. (1995b). Log periodogram regression of time series with long range dependence. *The Annals of Statistics*, **23**, 1048–1072.
59. Robinson, P.M. (1994). Time Series with Strong Dependence, in C.A. Sims (Ed), *Advances in Econometrics, Sixth World Congress*. Cambridge University Press, pp. 47–95.
60. Robinson, P.M. (1991). Testing for strong serial correlation and dynamic conditional heteroskedasticity in multiple regression. *Journal of Econometrics*, **47**, 67–84.
61. Robinson, P.M. (1977). The estimation of a nonlinear moving average model. *Stochastic Processes and their Applications*, **5**, 81–90.
62. Robinson, P.M. and Zaffaroni, P. (1997). Modelling nonlinearity and long-memory in time series. *Fields Institute Communications*, **11**, 161–170.
63. Surgailis, D. (1981). Convergence of sums of nonlinear functions of moving averages to self-similar processes. *Soviet Mathematics*, **23**, 247–250.

64. Teyssière, G. (2003). Interaction models for common long-range dependence in asset price volatility, in: Rangarajan, G. and Ding, M. (Eds.), *Processes with Long Range Correlations: Theory and Applications*, Lecture Notes in Physics, Vol. 621. Springer Verlag, Berlin, pp. 251–269.
65. Veitch, D. and Abry, P. (1999). A wavelet based joint estimator of the parameters of long-range dependence. *IEEE Transactions on Information Theory*, **45**, 878–897.
66. Veitch, D., Taqqu, M.S. and Abry, P. (2000). Meaningfull MRA initialisation for discrete time series. *Signal Processing*, **80**, 1971–1983. EURASIP Best Paper Award.
67. Veitch, D. and Abry, P. (2001). A statistical test for the time constancy of scaling exponents. *IEEE Transactions on Signal Processing*, **49**, 2325–2334.
68. Veitch, D., Abry, P. and Taqqu, M.S. (2003). On the automatic selection of the onset of scaling. *Fractals*, **11**, (4), 377-390.
69. Vostrikova, L.Ju. (1981). Detection of “disorder” in multidimensional random processes. *Soviet Mathematics Doklady*, **24**, 55–59.
70. Zaffaroni, P. (2003). Gaussian inference on certain long-range dependent volatility models. *Journal of Econometrics*, **115**, 199–258.