Wavelets and estimation of long memory in nonstationary models: does anything beat the Exact Local Whittle Estimator?

H. Boubaker[a]; **M. Boutahar**[b] and **R. Khalfaoui**[c]

[a] GREQAM Aix-Marseille University

henitn2001@yahoo.fr

Centre de la Vieille Charité 2, rue de la Charité 13236 Marseille cedex 02

[b,c] Institut of Mathematics of Marseille (IMM), Aix-Marseille University

Campus de Luminy, Case 907, 13288 MARSEILLE Cedex 9

mohammed.boutahar@univ-amu.fr; r.kalfaoui@gmail.com

Keywords: Long memory; Nonstationary models; Wavelets.

Abstract

In this paper, we analyze the performance of five estimation methods for the long memory parameter d. The goal of our paper is to construct a wavelet estimate for the fractional differencing parameter in nonstationary long memory processes which dominate the well known estimate of Shimotsu and Phillips (2005). The simulation results show that the wavelet estimation method of Lee (2005) with several tapering techniques performs better under most cases in nonstationary long memory. The comparison is based on the empirical root mean squared error of each estimate.

1 Introduction

In recent years, studies about long memory have received the attention of statisticians and mathematicians. This phenomenon has grown rapidly and can be found in many fields such as hydrology, chimistry, physics, economic and finance. For instance, the studies of Tokutsu et al. (2008) for the realized volatility of stock returns in Tokyo stock exchange and Hsu (2006) for modeling the Nile River time series, Boutahar et al. (2007) for US inflation rate, Boutahar and Khalfaoui (2011) for studying the crude oil price volatility, Tsay (2009) for political time series and Guegan et al. (2012) for GDP time series. Studies based on simulation experiments in long memory domain are large, some of them we cite are studies of Tokutsu et al. (2008), Hsu (2006), Moulines et al. (2008), Boutahar and Khalfaoui (2011), Guegan et al. (2012) and McCloskey (2012).

The concept of long memory describes the property that many time series models possess, despite being stationary which is higher persistence than short memory models, such as ARMA models.

The properties of the long memory models depend on the fractional differencing parameter value, denoted by d. Several estimation techniques have been proposed in the literature for detecting the long memory phenomenon, in both time and frequency domains (see Beran (1994))). Wavelets have been used to estimate the fractional differencing parameter (see Jensen (1999) and Veitch and Abry (1999)), i.e. a log-linear relationship exists between the wavelet variance of a given long memory model and its scale equal to the long memory parameter. In this paper, we apply some estimation methods for detecting long memory in nonstationary models: Three semi-parametric wavelet-based estimates are used, that are, the wavelet ordinary least square estimate ($WOLS_1$) of Jensen (1999), the wavelet ordinary least square estimate of Veitch and Abry (1999) ($WOLS_2$) and the wavelet GPH estimate of Lee (2005) (WGPH). The Exact Local Whittle estimate (ELW) of Shimotsu and Phillips (2005) and the GPH estimate of Geweke and Porter-Hudak (1983) are also used.

Note that consistency and/or asymptotic normality of these five estimates are proved but for different intervals for d. They are proved for $WOLS_1$ of Jensen (1999), the $WOLS_2$ of Veitch and Abry (1999) and the GPH Geweke and Porter-Hudak (1983) estimates if $|d| < 1/2$. Lee (2005) proved the asymptotic normality of the WGPH estimate if $d \in (0, 3/2)$. Shimotsu and Phillips (2005) proved asymptotic normality of the ELW estimate if $d \in (\Delta_1, \Delta_2)$, with $\Delta_2 - \Delta_1 \leq 9/2$. Velasco (1999) obtained that the GPH estimate is asymptotically normal for $d \in (1/2, 3/4)$ and still consistent for $d \in (1/2, 1)$. He showed that with adequate

data tapers, the GPH estimator is consistent and asymptotically normal distributed for any d, including both nonstationary and non-invertible models, see section 3 and 4 for more detail.

There are many other interesting estimators, but we will not consider in our simulation design for simplification purpose, which handle the nonstationary long memory models. We cite, among others, the local Whittle Wavelet estimate suggested by Moulines et al. (2008). They proved that their estimator is consistent and rate optimal if the model is a linear, and is asymptotically normal if the model is Gaussian. Abadir et al. (2007) proposed the fully extended local Whittle estimator and showed its consistency and derived its asymptotic expansion, they argued that their estimator is applicable not only for the traditional cases but also for nonlinear and non-Gaussian models.

The remainder of the paper is structured as follows. The long memory definition is introduced in section 2. Wavelet analysis is presented in section 3. In section 4, we briefly describe the Geweke and Porter-Hudak (1983)'s and Shimotsu and Phillips (2005)'s long memory estimates. The tapering is briefly introduced in section 5. In section 6, we compare the performance of all proposed methods. An empirical application is proposed in section 7. Section 8 concludes the paper.

2 Long memory

Let $X(t),\ t = 1, 2, \ldots, T$ be an ARFIMA(p,d,q) process defined by

$$(1 - L)^d \Phi(L)X(t) = \Theta(L)u(t) \tag{1.1}$$

where $u(t) \sim$ i.i.d.$\mathcal{N}(0, \sigma_u^2)$, L denotes the lag operator, and $\Phi(L)$ and $\Theta(L)$ are polynomials in the lag operator L, that is

$$\Phi(L) = 1 - \phi_1 L - \phi_2 L^2 - \ldots - \phi_p L^p, \Theta(L) = 1 + \theta_1 L + \theta_2 L^2 + \ldots + \theta_q L^q. \tag{1.2}$$

$(1-L)^d$ is the fractional differencing operator for non-integer values of d. Following Granger and Joyeux (1980) and Hosking (1981), $(1-L)^d$ can be defined by

$$(1-L)^d = \sum_{i=0}^{+\infty} \begin{bmatrix} d \\ i \end{bmatrix} (-L)^i = 1 - dL - \frac{1}{2}d(1-d)L^2 - \frac{1}{6}d(1-d)(2-d)L^3 - \ldots \qquad (1.3)$$

This expression can be re-expressed as

$$(1-L)^d = \sum_{i=0}^{+\infty} \frac{\Gamma(i-d)L^i}{\Gamma(-d)\Gamma(i+1)} \qquad (1.4)$$

$X(t)$ is said to be long memory if its autocorrelation function decays to zero like a power function, that is

$$\rho(h) \sim Ch^{2d-1} \quad \text{as } h \to \infty,\ C \neq 0. \qquad (1.5)$$

For $-0.5 < d < 0.5$, $X(t)$ is stationary and invertible and for $d \geq 0.5$, $X(t)$ is nonstationary process. In this paper, we will concentrate on the fractional integrated nonstationary processes.[1]

3 Wavelets

In this section we provide a brief description of wavelets. The reader should consult Mallat (1989), Daubechies (1992) and Meyer (1993) for further details. Wavelets are *small waves* that grow and decay in a limited time period. These wavelets represent a set of functions $\{\psi_{j,k}(t)\}_{j,k\in\mathbb{Z}}$ that act as an orthonormal basis for a given time series $X(t)$ in $\mathbf{L}^2(\mathbb{R})$. The basis functions are shifted and scaled versions of the time-localized mother wavelet, $\psi(t)$, defined by

$$\psi_{j,k}(t) = 2^{j/2}\psi(2^j t - k). \qquad (1.6)$$

where k refers to translation (shifted) parameter and j refers to dilation (scaled) parameter. The wavelet series smooth to a time series $X(t)$ is defined by

$$X(t) = \sum_k s_{J,k}\phi_{J,k}(t) + \sum_k d_{J,k}\psi_{J,k}(t) + \ldots + \sum_k d_{1,k}\psi_{1,k}(t). \qquad (1.7)$$

[1]see Beran (1994) for more details on long memory processes.

where J is the number of scales and k ranges from one to the number of coefficients in the specified components. The coefficient $s_{J,k}, d_{J,k}, \ldots, d_{1,k}$ are the wavelet coefficients given by

$$s_{J,k} = \int \phi_{J,k}(t)X(t)dt. \tag{1.8}$$

$$d_{j,k} = \int \psi_{j,k}X(t)dt. \tag{1.9}$$

where $\phi(t)$ is the father wavelets defined by

$$\phi_{j,k}(t) = 2^{j/2}\phi(2^j t - k),\ j,k \in \mathbb{Z}. \tag{1.10}$$

Let a discrete fractional time series $X(t),\ t = 1, 2, \ldots, T$ with memory parameter $d \in (0, 1.5)$. Wavelet coefficients of the wavelet transform for $X(t)$ can be used for estimating d.

3.1 Jensen (1999)'s estimate

The Wavelet Ordinary Least Square (WOLS) estimate of the fractional differencing parameter was introduced by Jensen (1999). Jensen proved that the wavelet coefficients, $d_{j,k}$, associated with a mean zero ARFIMA(0,d,0) model with $|d| < 0.5$ are distributed $\mathcal{N}\left(0, \sigma^2 2^{-2jd}\right)$, where σ^2 is a finite constant.[2] The wavelet coefficient's variance at a scale j is defined by

$$\text{Var}(d_{j,k}) = R(j) = \sigma^2 2^{-2jd}. \tag{1.11}$$

By taking the algorithms on both sides of equation 1.11, we have

$$\ln\{R(j)\} = \ln\left\{\sigma^2\right\} - d\ln\left\{2^{2j}\right\}. \tag{1.12}$$

The estimate of the fractional differencing parameter d can be obtained by applying the ordinary least squares method to the equation 1.12. Following Jensen (1999), the wavelet ordinary least squares estimate of d is given by

$$\hat{d}_{WOLS_1} = \frac{\sum_{j=0}^{J-1} y_j \ln\left\{\bar{R}(j)\right\}}{\sum_{j=0}^{J-1} y_j^2}. \tag{1.13}$$

where $y_j = \ln\{2^{-2j}\} - \frac{1}{J}\sum_{j=0}^{J-1}\ln(2^{-2j})$ and $\bar{R}(j) = \frac{1}{2^j}\sum_{k=0}^{2^j-1} d_{j,k}^2$.

Jensen shows that the $\hat{d}_{WOLS_1}$ is a consistent estimate of the fractional integration parameter d.

[2]See Jensen (1999) for more details about the estimation method.

3.2 Veitch and Abry (1999)'s estimate

Based on the DWT coefficients $d_{j,k}$ defined in equation 1.9 of $X(t),\ t = 1, 2, \ldots, T$, where $X(t)$ is an ARFIMA(0,d,0). Following Veitch and Abry (1999) we have

$$\hat{\mu}_j = \frac{1}{\upsilon_j}\sum_{k=1}^{\upsilon_j} d_{j,k}^2. \tag{1.14}$$

where υ_j is the number of the wavelet coefficients at octave j available to be computed. As shown by Veitch and Abry (1999)

$$\hat{\mu}_j \sim \frac{z_j}{\upsilon_j}\chi^2_{\upsilon_j}. \tag{1.15}$$

where $z_j = c2^{2dj}$, $c > 0$, and $\chi^2_{\upsilon_j}$ is Chi-squared random variable with υ_j degrees of freedom. By taking the logarithms on both sides of equation 1.15, we have

$$\log(\hat{\mu}_j) \sim 2dj + \log_2(c) + \frac{\log(\chi^2_{\upsilon_j})}{\log 2} - \log_2(\upsilon_j). \tag{1.16}$$

The expected value and the variance of the variable $\log(\chi^2_\upsilon)$ are given by

$$\begin{aligned} \mathrm{E}\left\{\log(\chi^2_\upsilon)\right\} &= \xi(\frac{\upsilon}{2}) + \log 2, \\ \mathrm{Var}\left\{\log(\chi^2_\upsilon)\right\} &= \zeta(2, \frac{\upsilon}{2}), \end{aligned} \tag{1.17}$$

where $\xi(h) = \partial h/\partial h \log\{\Gamma(h)\}$, and $\zeta(2, \frac{\upsilon}{2})$ is the Riemann zeta function, defined by

$$\zeta(y) = \frac{1}{\Gamma(y)}\int_0^{+\infty}\frac{u^{y-1}}{\mathrm{e}^u - 1} = \frac{1}{1-2^{1-r}}\sum_{\upsilon=1}^{+\infty}\frac{(-1)^{\upsilon-1}}{\upsilon^r}. \tag{1.18}$$

The equation 1.16 can be written as

$$\vartheta_j = \alpha + \beta w_j + \varepsilon_j. \tag{1.19}$$

where $\vartheta_j = \log_2(\hat{\mu}_j) - g_j$, $\alpha = \log_2(c)$, $\beta = 2d$, $w_j = \log_2(2^j) \simeq j$ and $\varepsilon_j = \log_2\left\{\log(\chi^2_{\upsilon_j})\right\} - \log_2(\upsilon_j) - g_j$, $g_j = \xi(\upsilon_j/2) - \log(\upsilon_j/2)$. ε_j satisfies

$$\mathrm{E}(\varepsilon_j) \simeq 0, \mathrm{Var}(\varepsilon_j) = \frac{\zeta(2, \frac{\upsilon_j}{2})}{[\log 2]^2} \simeq \left\{2\upsilon_j \log^2 2\right\}^{-1}. \tag{1.20}$$

The wavelet ordinary least square estimate of Veitch and Abry (1999) is given by

$$\hat{d}_{WOLS_2} = \frac{\hat{\beta}}{2} \tag{1.21}$$

where $\hat{\beta}$ is the ordinary least square estimate obtained from equation 1.19. Veitch and Abry (1999) shows that under some regularity conditions, $\hat{d}_{WOLS_2}$ is efficient and consistent.

3.3 Wavelet GPH estimate

The wavelet GPH estimate is defined by Lee (2005). Based on the discrete wavelet transform of $X(t)$, given by

$$d_{j,k} = \sum_k X(t)\psi_{j,k}(t). \tag{1.22}$$

The spectral density of the wavelet transform at the scale j around zero frequency for $d \in (0, 1.5)$ is as follows

$$f_j(\lambda) = C_j \left|\lambda\right|^{-2d} \left|\Lambda(\lambda)\right|^2 \text{ as } \lambda \to 0 = C_j \left|\lambda\right|^{-2(d-1)} g^2(\lambda) \text{ as } \lambda \to 0. \tag{1.23}$$

where $C_j = c_j/2\pi < \infty$ is a constant term, and $|\Lambda| = \lambda^\nu g(\lambda)$ for all integer ν, with $g(t\lambda)/g(\lambda) = 1$ for all t as $\lambda \to 0$ and $0 < g(0) < \infty$.

For a fixed scale j, the periodogram of $f_j(\lambda)$ is

$$I_l(j) = \frac{1}{2\pi T} \sum_{k=0}^{2^j-1} \left|d_{j,k} \exp(i\lambda_l k)\right|^2, \; l = 1, 2, \ldots, m. \tag{1.24}$$

where $\lambda_l = 2\pi l/T$, m is the number of frequencies which restricted such that $m \to \infty$ and $m/T \to 0$ as $T \to \infty$.

The wavelet-based GPH estimate denoted as d_{WGPH}, it is obtained by a log transformation of equation 1.23, more precisely, is obtained by regressing log periodogram, $\ln\{I_l(j)\}$, on $-2\ln(\lambda_l)$ for $l = 1, 2, \ldots, m$, then by adding one to the estimate.

For $d \in (0, 1.5)$, Lee (2005) shows that $\hat{d}_{\mathrm{WGPH}}$ is consistent and asymptotically normal if $m = o(T^{4/5})$, that is

$$\sqrt{m}\left(\hat{d}_{\mathrm{WGPH}} - d\right) \to \mathcal{N}\left(0, \frac{\pi^2}{24}\right), \text{ as } T \to \infty \tag{1.25}$$

where $m = T^{4/5}$ is the optimal rate for the number of frequency in terms of the mean squared error (see Hurvich et al. (1998) and Andrews and Guggenberger (2003)).

4 Other estimates

4.1 Exact Local Whittle estimate

Let $X(t)$, $t = 1, \ldots, T$ be a time series generated by the following fractional model

$$(1-L)^d X(t) = \epsilon(t)\Upsilon\{t \geq 1\}. \tag{1.26}$$

where $\Upsilon\{.\}$ is the indicator function and $\epsilon(t)$ is a stationary process with mean zero and spectral density $f_\epsilon(\lambda) \sim G$ as $\lambda \to 0$. The discrete Fourier transform and the periodogram of $X(t)$ are given by

$$\digamma_X(\lambda_n) = \frac{1}{\sqrt{2\pi T}} \sum_{t=1}^{T} X(t)\exp(it\lambda_n),\ \lambda_n = \frac{2\pi n}{T},\ t = 0, 1, \ldots, T. \tag{1.27}$$

$$I_X(\lambda_n) = |\digamma_X(\lambda_n)|^2. \tag{1.28}$$

Shimotsu and Phillips (2005) proposed to estimate (d, G) by minimizing the objective function

$$Q_m = (d, G) = \frac{1}{m} \sum_{n=1}^{m} \left[\log(G\lambda_n^{-2d}) + \frac{1}{G} I_\epsilon(\lambda_n)\right]. \tag{1.29}$$

So that, we have

$$\left(\hat{d}, \hat{G}\right) = \underset{G\in(0,\infty),d\in[\triangle_1,\triangle_2]}{\arg\min} Q_m(d, G). \tag{1.30}$$

where $-\infty < \triangle_1 < \triangle_2 < +\infty$ are the lower and upper bounds of the admissible values of d. m is the bandwidth parameter which determines the number of periodogram ordinates used in the estimation.

Concentrating $Q_m(d, G)$ with respect to G, Shimotsu and Phillips (2005) defined the Exact Local Whittle estimate as

$$\hat{d}_{\mathrm{ELW}} = \underset{d\in[\triangle_1,\triangle_2]}{\arg\min} R_0(d). \tag{1.31}$$

where $R_0(d) = \log\left\{\hat{G}(d)\right\} - 2d\frac{1}{m}\sum_{n=1}^{m} \log\lambda_n$ and $\hat{G}(d) = \frac{1}{m}\sum_{n=1}^{m} I_\epsilon(\lambda_n)$.

Shimotsu and Phillips (2005) found that the Exact Local Whittle is consistent[3] for $d \in (\triangle_1, \triangle_2)$, and asymptotically normal

$$\sqrt{m}\left(\hat{d}_{\mathrm{ELW}} - d\right) \to \mathcal{N}\left(0, \frac{1}{4}\right), \text{ as } T \to \infty. \tag{1.32}$$

provided that $\triangle_2 - \triangle_1 \leq 9/2$.

4.2 Geweke and Porter-Hudak (1983)'s estimate

The GPH method also known as the log-periodogram method is proposed by Geweke and Porter-Hudak (1983). Following Geweke and Porter-Hudak (1983) and based on a fractional time series $X(t)$ with length T the estimated slope coefficient of

$$\ln\left\{I_X(\lambda_n)\right\} = c + d\ln\left[\left\{2\sin\left(\frac{\lambda_n}{2}\right)\right\}^{-2}\right] + \epsilon_n \tag{1.33}$$

is used as the estimate of the fractional differencing parameter d, denoted by $\hat{d}_{\mathrm{GPH}}$. Here $I_X(\lambda_n)$ is the periodogram which is normalized by 2π at the n-th Fourier frequency λ_n, where $\lambda_n = 2\pi n/T$ and $\epsilon \sim$ i.i.d.$\mathcal{N}(0, \sigma_\epsilon^2)$.[4]

5 Tapering

The idea of tapering is proposed by Cooley and Tukey (1965) in order to reduce the bias of the periodogram due to frequency domain leakage, where part of the spectrum "leaks" into adjacent frequencies. By using tapers, we reduce the leakage due to the discontinuity caused by the finiteness of the sample, therefore, tapers smooth this discontinuity. Following Cooley and Tukey (1965), it consists of multiplying the data by a sequence of non-negative weights, called "taper" or "fader" or "data window".

Definition 5.1. *A sequence of taper* $\{h(t)\}_{t=1}^{T}$ *is of order* p *if the following two conditions are satisfied*

[3] See Shimotsu and Phillips (2005) for more details about the consistency of the estimate.

[4] See Geweke and Porter-Hudak (1983) for more details.

1. $\sum_{t=1}^{T} h(t)^2 = Tb(T),\ 0 < b(T) < \infty.$

2. *For* $N = \lfloor T/p \rfloor$*, the Dirichlet Kernel* $D(\lambda)$ *satisfies*

$$D(\lambda) \equiv \sum_{t=1}^{T} h(t) exp\{i\lambda t\} = \frac{a(\lambda)}{T^{p-1}} \left(\frac{sin\lfloor T\lambda/2p \rfloor}{sin\lfloor \lambda/2 \rfloor} \right)$$

where $a(\lambda)$ *is a complex function and* $\lfloor . \rfloor$ *is the integer part.*

There exist many data tapers, in this paper we are interested in Bartlett, Hanning and Parzen windows.

Bartlett window, also known as *Triangular window.* The Bartlett window is defined as

$$h(t) = 1 - \left| \frac{2t - T}{T} \right|,\ t = 1, 2, \ldots, T \tag{1.34}$$

where T is the length of the window and $h(t)$ is the window value.

Hanning window: is also known as the *Cosine Bell* window. Usually, it is called Hanning window

$$h(t) = \frac{1}{2} \left[1 - \cos\left(\frac{2\pi T}{N} \right) \right],\ T = 1, 2, \ldots, N \tag{1.35}$$

where N is the length of the window.

Parzen window: The Parzen window is a piecewise cubic curve window obtained by the convolution of two triangles of half length. it is defined as follows:

$$h(t) = \begin{cases} 2\left\{1 - \left|\frac{2t-T}{T}\right|\right\}^3,\ 1 \leq t \leq N \ or\ 3N \leq t \leq 4N, \\ 1 - 6\left[\left\{\frac{2t-T}{T}\right\}^2 - \left|\frac{2t-T}{T}\right|^3\right],\ N < T < 3N. \end{cases} \tag{1.36}$$

with $N = \lfloor T/4 \rfloor$.

The tapered periodogram of a variable $X(t),\ t = 1, 2, \ldots, T$ for any taper sequence $h(t),\ t = 1, 2, \ldots, T$ is defined as

$$I(\lambda_n) = |F(\lambda_n)|^2. \tag{1.37}$$

where $\lambda_n = 2\pi n/T,\ \forall\ n = 1, 2, \ldots, \lfloor (T-1)/2 \rfloor,\ \lfloor . \rfloor$ is the integer part and the tapered discrete Fourier transform is given by

$$F(\lambda_n) = \left(2\pi \sum_{t=1}^{T} h(t)^2 \right)^{-1/2} \sum_{t=1}^{T} h(t)X(t)\exp(i\lambda_n t). \tag{1.38}$$

For more details about different types of tapers see Brillinger (1981), Alekseev (1996) and Velasco (1999). Figure 1 shows the Bartlett and Hanning windows.

If we denote by $\hat{d}_{\mathbf{TGPH}}$ the GPH estimator of Geweke and Porter-Hudak (1983) but obtained for tapered data, then Velasco (1999) established that

$$\sqrt{m}\left(\hat{d}_{\mathbf{TGPH}} - d\right) \to \mathcal{N}\left(0, \frac{\pi^2}{24}\right), \text{ as } T \to \infty \tag{1.25}$$

provided that the model is Gaussian and

$$[d + 1/2] \leq p \tag{1.26}$$

where p is the order of the taper and for any d, ($[x]$ is the integer part of x).

6 Simulation study

In this section, we consider sample sizes of $T = 128$, 256, 512, 1024, 2048 and a number of replications of 10000. The data generation processes are ARFIMA(0,d,0) **and ARFIMA(1,d,0)** with $d = 0.4$, 0.5, 0.6, 0.8, 0.9, 1.0, 1.1, 1.2; **the AR coefficient is equal to $\phi = 0.4$ and $\phi = 0.8$.** [5] For wavelet estimates, we consider a Least Asymmetric wavelet filter with length $L = 12$, i.e., $LA(12)$.[6] We also consider three tapers in order to show the effect of taper in the robustness of wavelet estimates, that is, Bartlett, Hanning and Parzen. **Note that the order of Bartlett, Hanning and Parzen tapers are 2, 3 and 4 respectively. Note also that all the data generating processess are such that $[d + 1/2] \leq 2$, therefore the condition (1.26) is satisfied and hence the tapers will achieve consistency. However data generating processes with $d \geq 1.2$ can also be simulated and we can define tapers of any given order p. For example the Kolmogorov taper of**

[5] In this study we are not interested in the estimation of the short-memory parameters p and q in the ARFIMA(p,d,q), for some discussing results see Boutahar and Khalfaoui (2011).

[6] We have also tried other wavelet filters, but have never any effect in the robustness of estimates (they are available upon request). For some examples see also Boubaker and Boutahar (2011).

order p is obtained from the pth convolution of the uniform density.
For the Exact Local Whittle estimate we used bandwidths $m = T^{0.7}$ and $m = T^{0.8}$. The bandwidth used for the GPH estimate is $m = T^{0.8}$.
To compare the performance of the given estimates defined in sections 3 and 4, we compute the Root Mean-Squared Error value, denoted hereafter by RMSE, i.e., RMSE $= \sqrt{T^{-1}\sum_{i=1}^{T}(\hat{d}_i - d)^2}$.

6.1 Comments on the purely long memory processes

Table 1 provides simulation results of the GPH and wavelet GPH estimates. Table 2 provides simulation results of wavelet OLS estimates. Table 3 provides simulation results of the Exact Local Whittle estimate.
From Tables 1-3 we can observe that the bias ($\bar{d} - d$) and the RMSE globally decrease as the sample size T increases, consequently all the estimators are consistent; however they have different behavior in finite sample size. Below we describe some differences.
From the simulation results shown in Tables 1 and 2, we remarked that Bartlett taper provides smaller variance values than those produced by Hanning and Parzen tapers.
From Table 3, we can observe that an estimate of the fractional differencing parameter d with a bandwidth $m = T^{0.8}$ provides smaller variance values than an estimate with bandwidth $m = T^{0.7}$.
Figures 2, 3 and 4 plot the RMSE as a function of the long memory parameter d. As it can be shown in Figure 2, wavelet estimates are superior to the Exact Local Whittle one, excepted when $d = 0.9$ and $T = 128$. We observe that the Exact Local Whittle estimate is better. We can observe also that with same values of d and T and by using the Bartlett taper wavelet estimate is superior (wavelet GPH estimate). In addition, from Figure 2 and with a sample size $T = 256$, we can observe that wavelet estimates are superior to the Exact Local Whittle estimate, excepted when $d = 0.8$. For small sample sizes: $T = 128$ and $T = 256$, we conclude that the wavelet method outperforms the Exact Local Whittle method.

From Figure 3, we also find that wavelet estimate outperforms the Exact Local Whittle estimate, excepted when $d = 0.6$, $T = 1024$ and using the Bartlett and Parzen tapers for wavelet estimates, the Exact Local Whittle estimate is superior. In order to find a better wavelet estimate, we change the taper and we can observe that by applying the Hanning taper and for the same values of d and T ($d = 0.6$ and $T = 1024$), the wavelet estimate is superior, i.e., the wavelet GPH estimate is the best. Also from Figure 3, we observe that for $T = 512$ and $T = 0.9$ the Exact Local Whittle estimate is superior to the wavelet estimates with Hanning taper, but when using a good taper, such as Bartlett and Parzen ones we find that wavelet estimate performs the Exact Local Whittle one (wavelet GPH estimate). Hence, the simulation results reveals that for sample sizes $T = 512$ and $T = 1024$ the wavelet methods outperform the Exact Local Whittle method.

As it can be shown in Figure 4, the wavelet estimates with different tapers improve with smaller RMSE and they are very competitive to the Exact Local Whittle estimate.

Figure 5 (the left panels) plots the RMSE as a function of the time T, for an ARFIMA(0,0.8,0). It seems that the $RMSE$ is not monotone but converges to zero. Note that the value of the RMSE for T=128 is always very small to the one for $T = 2048$ for all the data generating processes and for all the five estimators. For instance, in Table 1 and using the Bartlett taper we can see that the RMSE of the wavelet GPH is equal to 0.00375 for $T = 128$ and 0.00018 for $T = 2048$, hence the RMSE is reduced about 95.2% of its value. In Table 3 the RMSE of the ELW is equal to 0.00294 for $T = 128$ and 0.00039 for $T = 2048$, hence the RMSE is reduced about 86.7% of its value.

6.2 Comments on ARFIMA with short memory component

Tables 4-6 provide simulation results for the ARFIMA(1,d,0) models with $\phi = 0.4$. Tables 7-9 provide simulation results for the ARFIMA(1,d,0) models with $\phi = 0.8$. Concerning the behavior of the estimators, a similar comments as on purely long memory case can be made. The estimators remains still consistent

but they converge toward the true parameter more slowly than the purely long memory case. From Table 1,4 and 7, we can see that for $T = 2048$, the RMSE of the wavelet GPH is equal to 0.00018 for the ARFIMA(0,0.8,0), is equal to 0.00162 for the ARFIMA(1,0.8,0) with $\phi = 0.4$ and is equal to 0.00293 for the ARFIMA(1,0.8,0) with $\phi = 0.8$, therefore in presence of short memory component the RMSE converges to zero but more slowly than the purely long memory processes; the degree of correlation of the short component has also an impact on the rate of convergence of the RMSE, the rate becomes more slowly as the AR coefficient approaches to 1. This fact is also visible in figure 5 where scales of left panels are lesser that the ones of the right panels. This in fact true for all the five estimators, (compare tables 1,4,7, tables 2, 5, 8, and tables 3,6,9).
As a summary, our simulation results show that wavelet estimates dominate the well known Shimotsu's estimate for non-stationary long memory processes and the choice of the taper is crucial for providing robust estimation method. We also conclude that the wavelet GPH estimate dominates under most cases.

7 An illustrative example

In this section, the proposed methodology is applied to a real example for illustration. The data consist of weekly crude oil spot prices (in US dollars per barrel) during the period from September 10, 1993 to April 19, 2013. The data are from the *U.S. Energy Information Administration.*[7] Table 4 reports summary statistics for WTI and Brent crude oil prices. There is a total of 1024 weekly observations. As shown in Table 10, differences between standard deviations (Std.dev) of WTI and Brent is small, indicating that the contribution of WTI volatility to Brent volatility is also small. Both skewness and kurtosis statistics in the table show that the returns distribution is not distributed normally.
The estimation results of the long memory parameter by methods presented in the method-

[7] http://tonto.eia.doe.gov/dnav/pet/pet_pri_spt_s1_w.htm

ology are given in Table 11. All estimates indicate non-stationary long memory behavior of crude oil prices ($1.034 \leq \hat{d} \leq 1.216$).

8 Conclusion

We have compared five methods for the estimation of the long memory parameter d in nonstationary time series, two non-wavelet-based and three wavelet-based. These are the Geweke-Porter Hudak (GPH), Exact Local Whittle, wavelet GPH and wavelet OLS. We have introduced the tapering in the case of GPH and wavelet methods in order to show the effect of the taper for providing robust estimation method (we have employed three types of tapers: Bartlett, Hanning and Parzen ones). We have undertaken a Monte Carlo comparison. In the Monte Carlo experiments, we have focused on ARFIMA(0,d,0) and ARFIMA(1,d,0). In this study, we are not interested in the estimation of the short-memory parameters p and q in the ARFIMA(p,d,q), for some discussing results see Boutahar and Khalfaoui (2011). We conclude that wavelet GPH method of Lee (2005) is superior under most situation with respect to RMSE criterion compared to others estimates. Thus, basing in several tapers the developed wavelet estimate outperform the well known Exact Local Whittle one for non-stationary long memory time series. Moreover, according to simulation results we have observed that the tapering has an impact on the performance of the estimate. We conclude also that, to get a wavelet estimate which is superior to the Exact Local Whittle one, we must choose the optimal taper.

References

Abadir K.M., Distasoa W., Giraitis, L. (2007) Nonstationarity-extended local Whittle estimation. *Journal of Econometrics* , 141, 13531384.

Andrews, D., and Guggenberger, P. (2003). A bias-reduced log periodogram regression estimator for the long-memory parameter. *Econometrica*, **71**, 675–712.

Alekseev, V. (1996). Jackson and jackson-vallée poussin-type kernels and their probability applications. *Theory Probab Appl*, **41**, 137–143.

Beran, J. (1994). Statistics for long memory processes. *New York: (Chapman & Hall).*

Boutahar, M., and Khalfaoui, R. (2011). Estimation of the long memory parameter in non stationary models: A Simulation Study. *HAL Working Papers*, halshs-00595057.

Boutahar, M., Marimoutou, V., and Nouira, L. (2007). Estimation methods of the long memory parameter: Monte carlo analysis and application. *Journal of Applied Statistics*, **34**, 261–301.

Brillinger, D. R. (1981). Time Series: Data analysis and Theory. *Expanded Edition, 540 pp., Holden-Day. San Francisco.*

Cooley, J., and Tukey, J. (1965). An algorithm for the machine calculation of complex fourier series. *Math Comput*, **19**, 297–301.

Daubechies, I. (1992). Ten Lectures on Wavelets. *Philadelphia*, SIAM.

Geweke, J., and Porter-Hudak, S. (1983). The estimation and application of long memory time series models. *Journal of Time Series Analysis*, **4** (4), 221–238.

Granger, C. W. J., and Joyeux, R. (1980). An introduction to long-memory time series models and fractional differencing. *Journal of Time Series Analysis*, **1**, 15–29.

Guegan, D., Lu, Z., and Zhu, B. (2012). Comparaison of Several Estimation Procedures for Long Term Behavior. *Documents de travail du Centre d'Economie de la Sorbonne.*

Hosking, J. (1981). Fractional differencing. *Biometrika*, **68**, 165–176.

Hsu, N. J. (2006). Long-Memory Wavelet Models. *Statistica sinica*, **16**, 1255–1271.

Hurvich, C., Deo, R., and Brodsky, J. (1998). The mean squared error of geweke and porter-hudak's estimator of the long memory parameter of a long-memory time series. *Journal of Time Series Analysis*, **19**, 1095–1112.

Jensen, M. (1999). Using wavelets to obtain a consistent ordinary least squares estimator of the long-memory parameter. *Journal of Forecasting*, **18** (1), 17–32.

Lee, J. (2005). Estimating memory parameter in the us inflation rate. *Economics Letters*, **87**, 207–210.

McCloskey, A. (2012). Estimation of the Long-Memory Stochastic Volatility Model Parameters that is Robust to Level Shifts and Deterministic Trends. *Working Papers*, 2012–17.

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

Meyer, I. (1993). Wavelets: Algorithms and Applications. *Philadelphia: SIAM.*

Moulines, E., and Roueff, F., and Taqqu, M. S. (2008). A Wavelet Whittle Estimator of the Memory Parameter of a Nonstationary Gaussian Time Series. *The Annals of Statistics*, **4**, (36), 1925–1956.

Shimotsu, K., and Phillips, P. (2005). Exact local whittle estimation of fractional integration. *Annals of statistics*, **20**, 87–127.

Tokutsu, Y., Nagataz, S., and Maekawa, K. (2008). A Comparison of Estimators for Long-Memory Process: Simulation and Emprical Study. *http://www.hue.ac.jp/prfssr/rcfe/recent-pdf/femes2008-12.pdf.*

Tsay, W. J. (2009). Estimating long memory time-series-cross-section data. *Electoral Studies*, **28** (1), 129–140.

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.

Velasco, C. (1999). Non-stationary log-periodogram regression. *Journal of Econometrics*, **91** (2), 325–371.