Stat Infer Stoch Process (2009) 12:221–250
DOI 10.1007/s11203-008-9030-7

Estimators for the long-memory parameter in LARCH models, and fractional Brownian motion

Michael Levine · Soledad Torres · Frederi Viens

Received: 19 December 2006 / Accepted: 23 October 2008 / Published online: 9 December 2008

Abstract This paper investigates several strategies for consistently estimating the so-called Hurst parameter H responsible for the long-memory correlations in a linear class of ARCH time series, known as LARCH(∞) models, as well as in the continuous-time Gaussian stochastic process known as fractional Brownian motion (fBm). A LARCH model's parameter is estimated using a conditional maximum likelihood method, which is proved to have good stability properties. A local Whittle estimator is also discussed. The article further proposes a specially designed conditional maximum likelihood method for estimating the H which is closer in spirit to one based on discrete observations of fBm. In keeping with the popular financial interpretation of ARCH models, all estimators are based only on observation of the "returns" of the model, not on their "volatilities".

Keywords ARCH · Times series · Fractional Brownian motion · Maximum likelihood estimator · Long memory · Whittle estimator · Moving average

Mathematics Subject Classification (2000) Primary 62M09 · Secondary 60G18, 62M10, 91B84

M. Levine · F. Viens (✉)
Department of Statistics, Purdue University, 150 N. University Street,
West Lafayette, IN 47907-2067, USA
e-mail: viens@purdue.edu

M. Levine
e-mail: mlevins@stat.purdue.edu

S. Torres
Depto de Estadística – CIMFAV, Universidad de Valparaíso,
1091 Av. Gran Bretaña, Playa Ancha, Valparaiso, Chile
e-mail: soledad.torres@uv.cl

1 Introduction

Long-memory behavior is one of the most important empirical properties exhibited by financial time series, such as asset returns and exchange rates. It is well known that, for the most part, the values of such a time series r_t, $t \in \mathbf{N}$ are uncorrelated but not independent, with most of dependency "hidden" within some nonlinear functions of r_t, such as r_t^2 or $|r_t|$. Historically, this has been modeled by conditional variance (volatility) models, such as the models traditionally included in the so-called (G)ARCH framework (see Gouriéroux 1997 and also Ghysels et al. 1996). However, typically, these models possess the so-called short memory property, and more specifically, exponential decay in autocorrelations of the respective nonlinear function of r_t, such as r_t^2. A symptomatic situation is found in Dan Nelson's well-known convergence results of ARCH/GARCH models to stochastic volatility models (see Nelson 1990). The linear autoregressive conditional heteroscedasticity model (LARCH), first introduced in Robinson (1991), has long been considered a very convenient vehicle for long-memory modeling. Its name is probably due to Giraitis et al. (2000). This model can be described as

$$r_t = \sigma_t \varepsilon_t; \quad \sigma_t^2 = \left(a + \sum_{j=1}^{\infty} b_j r_{t-j} \right)^2, \quad t \in \mathbf{Z} \tag{1}$$

where $\{\varepsilon_t : t \in \mathbf{Z}\}$ are iid random variables with zero mean and unit variance. We also assume that $a \neq 0$ to avoid special cases where the solution σ_t is a sequence of uncorrelated random variables. In order to ensure weak stationarity of the LARCH process, one must require that $||b|| = \left[\sum_{j=1}^{\infty} b_j^2\right]^{1/2} < 1$. It is also easy to observe that, under the same conditions, the LARCH process r_t, as well as σ_t, is also strongly stationary, meaning that the law of r_t for fixed t does not depend on t, and that the same holds for σ_t. This model lacks the interpretation usually accorded to the volatility models, since σ_t is not necessarily positive; this is arguably irrelevant when ε_t is symmetric, a case to which we will largely restrict ourselves here. Another advantage of the LARCH model lies in the simple conditions under which the process r_t itself and its powers r_t^j, $j \geq 2$, can be understood using combinatorial diagrams; for more details, see Giraitis et al. (2000). This, and the lack of a complete understanding of the long-memory modeling potential of the standard (nonlinear) ARCH framework, has lead a number of authors into adopting LARCH as their primary long-memory modeling vehicle.

Giraitis et al. (2000) also prove that, with proper normalization, the LARCH model converges in law to the fractional Brownian motion process, that is, a zero-mean Gaussian process $B^H(t)$ with the covariance function

$$\mathbf{E} B^H(s) B^H(t) = (1/2) \left(|s|^{2H} - |t|^{2H} - |t-s|^{2H} \right)$$

where the Hurst parameter H describes the strength of dependence between the increments of the process. Their proof is very sophisticated and involves some advanced combinatorial techniques. In contrast to this, we begin our article by showing that this convergence in law can be obtained using a much simpler technique that involves the so-called moving-average representation of the fractional Brownian motion (fBm).

It is legitimate to ask whether estimating the memory parameter of a nonlinear time series process that approximates the fBm process may give some information about the Hurst parameter of the fBm process itself. Recent results in the literature, esp. Wang (2002), suggest that there is no asymptotic statistical equivalence in the sense of Le Cam between the long-memory LARCH(∞) process and any natural discretization of the limiting fBm process.

Wang (2002) proved this non-equivalence for finite order processes of GARCH type. We will attempt to give arguments showing that the same should hold for the infinite-range LARCH model we use here.

On the other hand, it is easy to construct a coupled model where the noise terms defining the LARCH process are also used to define an approximation of fBm on the same probability space, with the same long memory parameter H, in which case one may simply choose a method based on the LARCH observations to estimate this common parameter H. In addition, the convergence in distribution to fBm is a good indication of the robustness of long memory in the LARCH(∞) model. Thus we will concentrate on constructing and analyzing some estimators for its long-memory and scale parameters. We will use methods commonly employed in theory of nonlinear time series. The first method we will propose is a simple way to estimate H dynamically, via conditional MLE; we will also present a different, more involved conditional MLE, which is better adapted to the case where the LARCH(∞) and approximate fBm processes are coupled, with a common H. We will also investigate the possibility of a local Whittle estimator for H. A more detailed summary of our work is given further below in this introduction.

A strong motivation for our work herein lies in our hope that the continuous-time quantitative finance community may appreciate the use of LARCH models because it combines tractable estimation with models for stock returns r_t that are uncorrelated, but whose volatilities σ_t are random and exhibit long-memory explicitly. This is in contrast to the often criticized so-called geometric fBm (fractional Black-Scholes) model, where the log stock returns are correlated directly according to an fBm, and the volatility parameter is constant, in a naive generalization of the Black-Scholes model. Arguably, the advantage of such a model resides only in its mathematical convenience in terms of its ease of manipulation in continuous time, but it cannot be used for modeling option pricing, because of the possibilities of arbitrage which exist in continuous time. It is known from Cheridito (2003) that these arbitrage possibilities vanish when trading is forced to be done discretely in time, but then the interest of using a continuous-time model also becomes less obvious.

Let us review various candidates for time series long-memory parameter estimation, from a historical perspective. A first set of possibilities lies in the conditional maximum likelihood methods. For linear processes, Cheung (1993) showed that, under correct model specification, the various MLE methods perform better than semiparametric estimators; the picture seems to be reversed when the model is misspecified. For more details, see Boes et al. (1989). The exact MLE method was proposed in Sowell (1992) and the approximate one in Fox and Taqqu (1986) (using the frequency domain approach). Wavelet-based MLE methods for the long-memory parameter estimation were proposed in Jensen (1999) for a narrow class of fractional white noise processes and in Jensen (2000) for ARFIMA (p, d, q) processes.

The other large group of methods utilizes the frequency domain ideas; in the parametric case, such is, for example, the classical Whittle estimator. Its properties for Gaussian and linear processes were investigated by Fox and Taqqu (1986), and also by Dahlhaus (1989) and Giraitis and Surgailis (1990). The next logical step would be to relax parametric assumptions on the behavior of the spectral density estimation and only assume that

$$f(\lambda) = |\lambda|^{-\alpha_0} g(\lambda),\ |\lambda| \geq \pi$$

around the point $\lambda = 0$ where $g(\lambda) \to c$ and c and α_0 are positive constants. Semiparametric estimation methods constitute another group. These methods require little a priori information about the spectral density of the time series, except its behavior around the point $\lambda = 0$. Among those methods, the log-periodogram method of Geweke and Porter-Hudak (1983) and the local Whittle estimator of Künsch (1987) should be mentioned. They were explored in great detail by Robinson (1995a,b). Closely related are the broad-band estimators of Moulines and Soulier (1999) and of Hurvich and Brodsky (2001), as well as the exact local Whittle estimation method of Shimotsu and Phillips (2005).

There has been relatively little work done on the semiparametric estimation of the long-memory parameter for nonlinear time series. Some results for the local Whittle estimator were obtained in Hurvich et al. (2005) and in Arteche (2004). General conditions under which the local Whittle estimator of the memory parameter of a stationary (not necessarily linear) process is consistent are given in Dalla et al. (2004). They also show that these conditions are satisfied for a fairly wide class of nonlinear models that includes signal plus noise processes, nonlinear transforms of a Gaussian process and EGARCH (exponential GARCH) models. Abadir et al. (2006) obtain asymptotic confidence intervals for the trend and memory parameters in the *case of long-memory processes with trends* that are possibly nonstationary, nonlinear and non-Gaussian. They call the estimator they use the Fully-Extended Local Whittle Estimator (FELW) which is a modified, for the presence of a trend, version of the estimator these authors developed in Abadir et al. (2005).

In this article, we discuss two possible methods for estimation of the long-memory parameter of the LARCH model. One of them is based on the conditional maximum likelihood approach and it has an additional benefit of robustness to violations of distributional assumptions. As pointed out earlier in the literature review, most of the work up until now was on the MLE-based long-memory parameter estimation for *linear* processes, such as ARFIMA; we contribute an estimation method that seems to be quite robust to the observation errors based on empirical evidence and, under additional constraints on the structure of the random errors, we give certain theoretical properties that also support our claim of robustness. In addition, we show how our conditional MLE may be modified so that it may be considered as an estimator of the Hurst parameter H of observations coming approximately from a fBm, by relating such observations with the appropriately scaled accumulation of the centered squared LARCH observations $r_t^2 - \mathbf{E}r_t^2$. In addition we explain precisely how to implement this more involved conditional MLE in practice.

We also attempt to use the local Whittle method to estimate the H of the LARCH process. Similarly to the MLE case, since σ_t^2 is unobservable, we apply the method to squared asset returns process r_t^2. However, consistency of this method is not entirely clear. As is usual for local Whittle method (see, for example, Dalla et al. 2004), it is necessary for the renormalized periodograms η_j of the process to satisfy the weak law of large numbers (WLLN); a possible set of sufficient conditions is mentioned in the same paper. An alternative set of sufficient conditions can be obtained from Lahiri (2003). We show that the latter is not satisfied in the case of our LARCH model, and the former can only be satisfied if certain conjectures on the asymptotic behavior of the covariances of products of r_t^2 are satisfied, which is an open question at the moment, and non-trivially so, since the behavior of such mixed moments for the LARCH process would involve calculations that are higher in complexity than those already very delicate combinatorial arguments in Giraitis et al. (2000). Therefore, while we do provide the details of the local Whittle method in our context, we cannot guarantee that it provides a consistent estimator for H based on time series observations, and *a fortiori* based on fBm observations.

We now present the structure of our paper, along with a detailed summary of our results. In Sect. 2, we present the LARCH(∞) model, and show in Proposition 1 that $n^{-H}\sum_{i=0}^{nt}(\sigma_i - a)$ converges in distribution to a constant multiple of fBm $B^H(t)$.

In Sect. 3, we introduce a possible conditional maximum likelihood estimator $\left(\hat{a}, \hat{H}\right)_i$ for the pair of parameters (a, H) in the LARCH(∞) model with i observations that can be given by the solution of the system of two equations $\frac{\partial \log L}{\partial a} = 0$ and $\frac{\partial \log L}{\partial H} = 0$; we discuss problems that arise concerning its consistency and show that its practical implementation is still possible because all quantities in these equations are explicit, which allows us to implement the resolution of this system, yielding a practical method for estimating a and H. The numerical results based on simulated data show that the method performs very well in practice.

In Sect. 4, we investigate the robustness of our conditional MLE. We calculate the total error made in the calculation of the conditional MLE $\hat{H}$ if exogenous errors enter the observation (Proposition 2). This formula may be calculated explicitly in parallel to the calculation of $\hat{a}$ and $\hat{H}$, which is useful if some assumptions on the observation errors can be made and used in a simulation. We also provide an upper bound for the total error (Proposition 3), which is the basis for theoretical evidence that when the errors are IID centered and square integrable, the total error converges to 0 faster than any power $n^{-\alpha}$ with $\alpha < 1 - H$ (Remark 1).

In Sect. 5, we draw the connection between our conditional MLE and the estimation of H from observations of an approximate fBm. In Subsect. 5.1, we prove that the following two naive ways of proceeding do not work: working with fBm increment observations that are analogous to the r_i's themselves, and working with the fBm increment observations related directly to the volatilities as in Proposition 1. In Subsect. 5.2, after showing convergence to fBm of the partial sums of the centered squared observations $r_i^2 - \mathbf{E}r_i^2$ (Proposition 4), we calculate the system of equations needed to implement the conditional MLE based on these observations (formulas (32) and (33)), and we discuss the practical implementation of this estimator.

In Sect. 6, we first present a local Whittle estimator $\hat{\theta}$ for $\theta = 2 - 2H$, based on volatility observations, using the periodograms (36) for the discrete Fourier frequencies, defining $\hat{\theta}$ as the minimizer of the quasi-likelihood-type objective function given in (37). Then, admitting that volatilities are not directly observed, we explain what modifications need to be performed in order to base the local Whittle estimator on squared returns instead; here we run into difficulties in justifying that sufficient conditions for consistency of $\hat{\theta}$ are satisfied, and show that this issue can only be resolved by establishing long-range dependence estimates on the mixed moments of the returns.

The last section is an Appendix where a crucial technical estimate is proved.

2 The LARCH(∞) model

As in Giraitis et al. (2000), we consider the linear ARCH(∞) model (LARCH) given by

$$r_t = \sigma_t \varepsilon_t \tag{2}$$

and

$$\sigma_t = a + \sum_{j=1}^{\infty} b_j r_{t-j} \tag{3}$$

$$= a + \sum_{j=1}^{\infty} b_j \sigma_{t-j} \varepsilon_{t-j}. \tag{4}$$

In a typical financial data interpretation, the process σ_t can be understood as volatility process over an elementary time interval, while the process r_t can then represent log returns of a stock price over the same interval. In what follows, we will deviate from the standard time series notation of using $t \in \mathbf{Z}$ for our model's time parameter, using instead this letter

t, and s as well, for continuous time, while the letters k and i represent discrete time. The relation between i and k, as seen below, will typically be of the form $k = tn$ or $k = [tn]$, where n^{-1} is thus our time step.

2.1 First convergence to fBm

In order to obtain a long-memory model, we can inspire ourselves from the so-called moving-average representation of fBm: if B^H is an fBm with Hurst parameter H, there exists a standard Brownian motion W defined on all of $\mathbf{R}$ such that

$$\begin{aligned} B^H(t) &= \int_0^t (t-r)^{H-1/2} \mathrm{d}W(r) + \int_{-\infty}^0 \left((t+|r|)^{H-1/2} - |r|^{H-1/2} \right) \mathrm{d}W(r) \\ &= \int_{-\infty}^{\infty} \left((t-r)_+^{H-1/2} - (-r)_+^{H-1/2} \right) \mathrm{d}W(r). \end{aligned} \tag{5}$$

If one sums the increments $\sigma_i - a$ to obtain the mean-zero process defined by $v_0 = 0$ and

$$v_k = v_{k-1} + \sigma_k - a = \sum_{i=1}^{k} (\sigma_i - a), \tag{6}$$

one will be approximating a process whose integration over time must yield the kernel in (5), suggesting that one should take

$$b_j = c j^{H-3/2},$$

where c is a fixed constant. It is well known that if

$$\|b\|^2 := \sum_{j=1}^{\infty} |b_j|^2 < 1, \tag{7}$$

then σ and r are weakly stationary processes, meaning they have constant means, here a and 0 respectively, and constant second moment, here the common value $a^2/\left(1 - \|b\|^2\right)$. One may recognize that the long memory parameter used for instance in Giraitis et al. (2000) is denoted by $\theta = 2 - 2H$.

With this choice of b_j we proceed to giving a simple proof of convergence of (the properly normalized) v_k to fBm. In what follows and the majority of the remainder of the paper, we assume that the independent noise terms ε_j are standard normal. This allows us to present simpler proofs; we believe most results would hold for more general noise terms, but the proofs would be more involved. We mentioned in the introduction that it is easy to construct both the LARCH(∞) process and an approximate fBm on the same probability space via a coupling. We now describe this coupling, and then state and prove the convergence result.

The easiest way to couple the ε_j's and an approximate B^H is to define the Brownian motion W underlying B^H in terms of the ε_j's as the development above should suggest, as a linear interpolation of the partial sums of the ε_j's: with $k = k_t = [nt]$, the largest integer smaller than nt, we let

$$W^{(n)}(t) := \sum_{j=0}^{k-1} \frac{\varepsilon_j}{\sqrt{n}} + \frac{nt-k}{\sqrt{n}} \varepsilon_k. \tag{8}$$

Donsker's invariance principle (see Karatzas and Shreve 1991, Theorem 2.4.20) proves that, as a random element in the space of continuous functions, $W^{(n)}$ converges in distribution to a Brownian motion W, which is a key to the proof of the following.

Proposition 1 *Let n be an integer. Assume that $k = k_t = [tn]$ and that W is the Brownian motion given as the limit of the Gaussian stochastic process in (8). Define the process V on* $[0, 1]$ *that is continuous and piecewise linear, with values at multiples of* $1/n$ *equal to*

$$V\left(\frac{k}{n}\right) = n^{-H} v_k,$$

where v_k is the centered partial sum of the volatilities, as defined in (6). Then for every $t \geq 0$, $V(k/n)$ converges in distribution to the fractional Brownian motion $aB^H(t)$ at time t, as n tends to ∞, where B^H is given in (5). Moreover, as a process, $\lim_{n\to\infty} V(\cdot)$ has a continuous modification which coincides with aB^H.

Proof We have

$$\begin{aligned} V\left(\frac{k}{n}\right) &= n^{-H} \sum_{i=1}^{k} (\sigma_i - a) \\ &= n^{-H} \sum_{i=1}^{k} \sum_{j=1}^{\infty} b_j \sigma_{i-j} \varepsilon_{t-j} \\ &= n^{-H} \sum_{i=1}^{k} \sum_{j=-\infty}^{i-1} (i-j)^{H-3/2} \sigma_j \varepsilon_j. \end{aligned} \tag{9}$$

Our goal is to obtain the moving average representation of fBm. In the last expression above, we will use the sum over k to approximate a Riemann integral with respect to Lebesgue measure ds, for which we need the factor $1/n$ to represent ds, and a factor $n^{-(H-3/2)}$ to account for $(i/n - j/n)^{H-3/2}$. The sum over j, on the other hand, will approximate a Wiener-Itô integral with respect to standard Brownian motion, for which $\varepsilon_j n^{-1/2}$ will represent the Brownian increment.

With $\mathcal{F}_i^+$ defined as the sigma-field generated by $\{\varepsilon_i, \varepsilon_{i+1}, \varepsilon_{i+2}, \ldots\}$, we transform $V(k/n)$ by adding and subtracting the term $\tau_j = \mathbf{E}\left[\sigma_j | \mathcal{F}_{j-J}^+\right]$ where J is fixed:

$$\begin{aligned} V\left(\frac{k}{n}\right) &= \sum_{i=1}^{k} \frac{1}{n} \sum_{j=-\infty}^{i-1} \left(\frac{i-j}{n}\right)^{H-3/2} \sigma_j \varepsilon_j n^{-1/2} \\ &= \sum_{i=1}^{k} \frac{1}{n} \sum_{j=-\infty}^{i-1} \left(\frac{i-j}{n}\right)^{H-3/2} \left(\sigma_j - \tau_j\right) \varepsilon_j n^{-1/2} \\ &\quad + \sum_{i=1}^{k} \frac{1}{n} \sum_{j=-\infty}^{i-1} \left(\frac{i-j}{n}\right)^{H-3/2} \tau_j \varepsilon_j n^{-1/2} \\ &= V_1(k/n) + V_2(k/n). \end{aligned}$$

Since $k = [tn]$, we get that k/n converges to the fixed value $t \in [0, 1]$. The process $V_2\,(k/n)$ defined by

$$V_2\,(k/n) = \sum_{i=1}^{k} \frac{1}{n} \sum_{j=-\infty}^{i-1} \left(\frac{i-j}{n}\right)^{H-3/2} \tau_j \varepsilon_j n^{-1/2} \tag{10}$$

is an approximation of an iterated Riemann and Itô integral. Specifically, using the convergence of the process $W^{(n)}$ defined in (8) in distribution to a Wiener process W, we have the following lemma, proved in the appendix.

Lemma 1 *$V_2\,(k/n)$ converges in distribution to a* $\int\limits_{s=0}^{t} ds \int\limits_{-\infty}^{s} (s-r)^{H-3/2}\, dW\,(r)$.

This lemma allows us to say that $V_2\,(k/n)$ converges in distribution to an fBm because the process v defined by the limit in this lemma, i.e.

$$v\,(t) := a \int\limits_{s=0}^{t} \mathrm{d}s \int\limits_{-\infty}^{s} (s-r)^{H-3/2}\, \mathrm{d}W\,(s)\,. \tag{11}$$

is an fBm. Indeed, using the stochastic Fubini theorem, we can rewrite

$$\begin{aligned} v\,(t) &:= a \int\limits_{r=-\infty}^{0} \mathrm{d}W\,(r) \int\limits_{s=0}^{t} (s-r)^{H-3/2}\, \mathrm{d}s \\ &\quad + a \int\limits_{r=0}^{t} \mathrm{d}W\,(r) \int\limits_{s=r}^{t} (s-r)^{H-3/2}\, \mathrm{d}s \\ &= \frac{a}{H-1/2} \int\limits_{r=-\infty}^{0} \left((t-r)^{H-1/2} - (-r)^{H-1/2}\right) \mathrm{d}W\,(r) \\ &\quad + \frac{a}{H-1/2} \int\limits_{r=0}^{t} (t-r)^{H-1/2}\, \mathrm{d}W\,(r)\,, \end{aligned}$$

which, up to a factor, is the moving average representation (5) of fBm.

It remains to show that

$$V_1\,(k/n) = \sum_{i=1}^{k} \frac{1}{n} \sum_{j=-\infty}^{i-1} \left(\frac{i-j}{n}\right)^{H-3/2} \left(\sigma_j - \tau_j\right) \varepsilon_j n^{-1/2}$$

can be made arbitrarily small. According to the ARCH(∞) specification in (4), the random variable σ_j (and therefore τ_j) is independent of ε_j. Therefore if $j \neq j'$, $\mathbf{E}\left[\left(\sigma_j - \tau_j\right) \varepsilon_j \left(\sigma_{j'} - \tau_{j'}\right) \varepsilon_{j'}\right] = 0$, and

$$\begin{aligned} \mathbf{E}\left[V_1\,(k/n)^2\right] = \sum_{i=1}^{k} \sum_{i'=1}^{k} \frac{1}{n^2} \sum_{j=-\infty}^{i-1} \sum_{j'=-\infty}^{i'-1} \left(\frac{i-j}{n}\right)^{H-3/2} \left(\frac{i'-j'}{n}\right)^{H-3/2} \\ \times \mathbf{E}\left[\left(\sigma_j - \tau_j\right) \varepsilon_j \left(\sigma_{j'} - \tau_{j'}\right) \varepsilon_{j'}\right] \frac{1}{n} \end{aligned}$$

$$= \sum_{i=1}^{k} \sum_{i'=1}^{k} \frac{1}{n^2} \sum_{j=-\infty}^{\min(i,i')-1} \left(\frac{i-j}{n}\right)^{H-3/2} \left(\frac{i'-j}{n}\right)^{H-3/2} \mathbf{E}\left[\left(\sigma_j - \tau_j\right)^2\right] \frac{1}{n}.$$

One of the underlying assumptions is that the solution to the ARCH(∞) specifications is a stationary process σ, which implies that $\mathbf{E}\left[\left(\sigma_j - \tau_j\right)^2\right]$ does not depend on j. Thus

$$\mathbf{E}\left[V_1\left(k/n\right)^2\right] = \mathbf{E}\left[(\sigma_0 - \tau_0)^2\right] \sum_{i=1}^{k} \sum_{i'=1}^{k} \frac{1}{n^2} \sum_{j=-\infty}^{\min(i,i')-1} \left(\frac{i-j}{n}\right)^{H-3/2} \left(\frac{i'-j}{n}\right)^{H-3/2} \frac{1}{n}.$$

The limit of the above triple sum is the Riemann integral $\int_0^t \int_0^t \left(\int_{-\infty}^{\min(s,s')} (s-r)^{H-3/2} \left(s'-r\right)^{H-3/2} \mathrm{d}r\right) \mathrm{d}s\mathrm{d}s'$, which is equal to $c_H t^{2H}$ for some constant c_H depending only on H. Now letting J be arbitrarily large, we have by dominated convergence that τ_0 can be made arbitrarily close to $\mathbf{E}\left[\sigma_0|\mathcal{F}_{\{-\infty,-1\}}\right] = \sigma_0$ and therefore $\mathbf{E}\left[(\sigma_0 - \tau_0)^2\right]$ can be made arbitrarily small, so that $V\left(\frac{k}{n}\right)$ converges in distribution to the fractional Brownian motion $v(t)$ given in (11).

We have finished proving that with $k = [tn]$, $V\left(k/n\right)$ converges in distribution to $aB^H\left(t\right)$ where this fBm is defined in (5), while the Brownian motion W therein is given as the limit of the Gaussian process $W^{(n)}$ defined by (8), as stated in the proposition. To prove the last statement of proposition, one may use computations similar to the ones above, but for two-dimensional distributions, showing that such distributions for $V(k/n)$ converge to those of fBm, and then invoke Kolmogorov's continuity criterion to conclude that the continuous limit coincides with fBm at the process level; details are omitted. □

2.2 On non-equivalence of experiments

In the remainder of the article, we consider the issue of finding a strongly consistent estimator for the parameters of the discrete- and continuous-time models. Because our data typically does represent time series, it is legitimate and necessary to assume that at time j, the only available observations are those given up to that time. Section 3 shows the simplest way to do this, based on dynamic observation of the process σ. Section 5.2 draws a connection between the discrete time series and fBm by constructing a conditional maximum likelihood estimator of the Hurst parameter H based on observations which can be considered as approximate observations of fBm. Therein we also explains why the results of Sect. 5.2 are not contained in Sect. 3: it is not possible to base a conditional MLE for H on approximate observations of fBm by solely considering linear transformations of the process (σ, r).

A more difficult question is to assert whether discrete observations of a bonafide continuous time fBm can be brutally substituted for LARCH(∞) observations in a LARCH conditional MLE scheme to determine H, when there is no way to observe a LARCH(∞) process that is coupled to the fBm. This question is essentially that of equivalence of statistical experiments in the sense of LeCam for the LARCH(∞) process and its fBm limit in distribution. We finish this section by explaining why it is improbable (and difficult to check) that this equivalence holds.

Consider a regular GARCH (1,1) process

$$\begin{aligned} x_k &= \mu_k + \sigma_k \epsilon_k \\ \sigma_k^2 &= \alpha_0 + \alpha_1 y_{k-1}^2 + \beta_1 \sigma_{k-1}^2. \end{aligned} \tag{12}$$

where $k = 1, \ldots, n$ and ϵ_k is a sequence of iid standard normal random variables. The drift term μ_k is commonly parameterized as $\mu_k = c_0 + c_1\sigma_k^2$ in empirical finance applications. Nelson (1990) showed that, asymptotically, this process weakly converges to the bivariate diffusion process

$$\mathrm{d}X_t \;= (\gamma_0 + \gamma_1\sigma_t^2)\mathrm{d}t + \sigma_t \mathrm{d}W_{1,t} \tag{13}$$

$$\mathrm{d}\sigma_t^2 = (\beta_0 + \beta_1\sigma_t^2) + \beta_2\sigma_t^2 \mathrm{d}W_{2,t} \tag{14}$$

where $W_{i,t}$, $i = 1, 2$ are standard Wiener processes, $W_{1,t}$ is independent of $W_{2,t}$ and the coefficients $\gamma_0, \gamma_1, \beta_0, \beta_1$ and β_2 are the rescaled versions of $c_0, c_1, \alpha_0, \alpha_1$ and α_2, respectively. It is important to realize, however, that the weak convergence does not translate into asymptotic statistical equivalence of (12) and its diffusion limit (13). Indeed, Wang (2002) showed that, for a GARCH (1,1) model (12) and its diffusion limit the asymptotic equivalence in the sense of Le Cam does not hold unless the volatility process σ_k^2 is non-stochastic which means that $\alpha_1 \equiv 0$. The non-stochastic case is of little practical relevance since it means that the GARCH(1,1) model becomes, effectively, a Gaussian linear model. Wang (2002) gives a nice heuristic explanation of this phenomenon by noticing the different noise propagation mechanisms that the GARCH model and its diffusion limit follow. Remember that the LARCH(∞) process is conceptually similar to ARCH processes: the only difference lies in the definition of the conditional variance σ_t^2 while the main process is still defined as $x_t = \sigma_t\varepsilon_t$. Because of that, it can be expected that the noise propagation systems of the LARCH process and its fBm limit are going to be different and the asymptotic equivalence does not hold as well except, possibly, some trivial special cases.

However, similar investigation in our case is intrinsically much more difficult. First, the LARCH(∞) time series process we consider is of infinite order; in practice, whenever the maximum likelihood approach is used to compute its parameters, the truncated version of the full likelihood has to be used. In order to do this, the truncated version of the volatility function $\tilde{\sigma}_t$ has to be considered instead of σ_t as defined in (3)-(4). This truncation only assumes that the observation up to and including the moment $t-1$ can be used to compute $\tilde{\sigma}_t$; thus,

$$\tilde{\sigma}_t = \alpha_0 + \sum_{j=1}^{\infty} \alpha_j y_{t-j} I(t-j \geq 1) = \alpha_0 + \sum_{j=1}^{t-1} \alpha_j y_{t-j}.$$

This is similar to the truncation done in order to compute maximum likelihood of GARCH(p,q) model which is a special case of the general ARCH(∞) model; for details, see, for example, Fan and Yao (2003). The truncated version of the loglikelihood function of LARCH(∞) is then proportional to

$$\sum_{j=\nu(J)}^{J} \left(\tilde{\sigma}_t^2 + \frac{y_t^2}{\tilde{\sigma}_t^2}\right) \tag{15}$$

where $\nu(J) \to \infty$ as $J \to \infty$ at a rate slower than J; for example, one can suggest $\nu(J) = o(J)$. This truncated likelihood is the one that is used in practice to obtain estimates of the LARCH(∞) coefficients. Therefore, it is the possible equivalence of this likelihood (or lack thereof) with the conditional likelihood of the limiting fBm process that is of interest and not of the full LARCH(∞) likelihood that cannot be expressed in a closed form.

Second, establishing asymptotic (non)equivalence of the LARCH(∞) model and its limiting fBm process requires investigating the asymptotic behavior of the likelihood (15) and the conditional likelihood of the respective scaled fBm $aB^H(t)$. Unfortunately, an fBm is not a martingale unlike a Brownian motion-driven process (13) and, therefore, it is not easy to write down its conditional likelihood in an explicit form.

One of the possible ways to construct an approximate conditional likelihood of the fBm process $B^H(t)$ is to consider its discretization based on the moving average representation of fBm; such representation has been described in detail in Szabados (2001). That discretization allows the fBm process to be represented as a linear combination of the random walks constructed on the same probability space as the fBm process $B^H(t)$ with an infinite number of terms. Based on the above representation, it is possible to construct a truncated version of the conditional likelihood of the fBm process; however, this likelihood is very different from the conditional likelihood of the discrete version of (13) that is used in the asymptotic analysis of Wang (2002). In particular, its dependence on the Hurst parameter H is nonlinear which is quite different from the case of the regular GARCH(p,q) model considered in Wang (2002). Thus, the problem seems to be very complicated and is outside the scope of this article. It will be investigated further within the framework of our continuing research. Note that Sect. 5.1 also contains evidence pointing towards the lack of asymptotic equivalence between the LARCH(∞) model and its fBm limiting process.

3 Conditional MLE in the ARCH(∞) model

In the discrete-time model, our observations are the log returns r_j. It is easy to define the conditional QMLE estimator of the parameter $\theta = (a, H)$ formally. The formal definition is based on the assumption of the normal error distribution which is not always the case in practice; therefore, the resulting estimator is presumed to be the "quasi" MLE. At time i, that is, given past the observations $r_j : j = 1, 2, \ldots, i$, this conditional QMLE is defined as the value of the pair $\left(\hat{a}, \hat{H}\right)$ which maximizes the conditional *quasi* log-likelihood function $\log L(a, H)$ defined via

$$L(a, H) = \prod_{j=1}^{i} f\left(r_j | r_{j-k} : k = 1, 2, \ldots\right)$$

where $f\left(r | r_{j-k} : k = 1, 2, \ldots\right)$ is the conditional density at point r of the random variable r_j given the prior random variables $r_{j-1}, \ldots, r_1, r_0, r_{-1}, \ldots r_{-k}, \ldots$. By the specifications (2) and (3), it is clear that r_j is conditionally normal $N\left(0, \sigma_j^2\right)$ given $r_{j-k} : k = 1, 2, \ldots$, since σ_j is explicitly given as a function of these past observations. Hence

$$\log L\,(a, H) = -\frac{1}{2}\left(i \log\,(2\pi) + \sum_{j=1}^{i} \log \sigma_j^2 + \sum_{j=1}^{i} \frac{r_j^2}{\sigma_j^2}\right) \tag{16}$$

where

$$\sigma_i = a + \sum_{j=1}^{\infty} j^{H-3/2} r_{i-j}. \tag{17}$$

We easily calculate

$$\frac{\partial \sigma_i}{\partial a} = 1; \quad \frac{\partial \sigma_i}{\partial H} = \sum_{j=1}^{\infty} j^{H-3/2} r_{i-j} \log j, \tag{18}$$

so that

$$\frac{\partial \log L}{\partial a} = \sum_{j=1}^{i} \left(-\frac{1}{\sigma_j} + \frac{r_j^2}{\sigma_j^3} \right) \tag{19}$$

$$\frac{\partial \log L}{\partial H} = \sum_{j=1}^{i} \left(-\frac{1}{\sigma_j} + \frac{r_j^2}{\sigma_j^3} \right) \sum_{k=1}^{\infty} k^{H-3/2} r_{j-k} \log k \tag{20}$$

Therefore $\left(\hat{a}, \hat{H}\right) = \left(\hat{a}, \hat{H}\right)_i$ is defined as the solution of the two equation system $\frac{\partial \log L}{\partial a} = 0$ and $\frac{\partial \log L}{\partial H} = 0$ for fixed number of observations i, i.e. with the understanding that σ_j is given via formula (17) and each $r_j : j \leq i$ is known.

The question of whether the LARCH(∞) QMLE estimator is consistent is very interesting. While the simulations do not seem to be encountering any serious difficulties most of the time, from the theoretical viewpoint the situation is much less clear. A common set of regularity conditions used to verify consistency of QMLE estimators for sequences of dependent variables is provided in Basawa et al. (1976). It is a time honored result, used in the past, for example, to verify consistency of the ARCH (p) QMLE estimator in Weiss (1986). Recall that the LARCH model as used here is parameterized using two parameters (a, H); for convenience, let us use the notation $\theta = (a, H)$; then, $\theta_0 = (a_0, H_0)$ is the pair of true parameters. The likelihood that uses all observations up until the moment t is denoted L_t while l_t is the respective log-likelihood; the total number of observations is T. In order for a consistent root of the score equation $\frac{\partial L_T(\theta)}{\partial \theta}$ to exist, we need

1. $$\frac{1}{T} \sum_{t=1}^{T} \nabla l_t(\theta_0) \xrightarrow{p} 0. \tag{21}$$

   This is an ergodic theorem-like statement with regard to the score function $\nabla l_t(\theta_0)$ at the true parameter value θ_0.
2. There exists a nonrandom matrix $M(\theta_0) > 0$such that for all $\epsilon > 0$

   $$P\left(-\frac{1}{T} \sum_{t=1}^{T} \nabla^2 l_t(\theta_0) \geq M(\theta_0) \right) > 1 - \epsilon \tag{22}$$

   for any $T > T_1(\epsilon)$. In other words, the Hessian matrix of the log-likelihood (conditional Fisher information matrix) of the process considered needs to be bounded away from zero at least in probability.
3. There exists $\eta > 0$ such that for any $\epsilon > 0$

   $$P\left(\frac{1}{T} \left| \sum_{t=1}^{T} (\nabla^2 l_t(\theta_T^*) - \nabla^2 l_t(\theta_0)) \right| > (1+\eta)^{-1} M(\theta_0) \right) < \epsilon \tag{23}$$

   which is true for any $T > T_2(\epsilon)$ and any $\theta_T^* : \left|\theta_T^* - \theta_0\right| < \delta_1(\epsilon)$ with $\delta_1(\epsilon)$ being independent of T.

Verification of these conditions represents a serious problem in the case of LARCH(∞) process. As an example, consider condition (22). Recall that the conditional standard deviation of the LARCH(∞) process is defined as $\sigma_t = a + \sum_{j=1}^{\infty} b_j r_{t-j}$ where $a \neq 0$.

Giraitis et al. (2000) established the existence of the weakly stationary LARCH(∞) process under the condition $\sum_{j=1}^{\infty} b_j^2 \leq 1$. Note, in particular, that such a solution exists even if the coefficients b_j are not all nonnegative; consequently, conditional standard deviation is, effectively, a linear combination of zero-mean martingales with coefficients of arbitrary sign. In general, such a combination σ_t need not be bounded away from zero. This means that the $E|\nabla^2 l_t(\theta_0)|$ need not be finite and, therefore, the sum $-\frac{1}{T}\sum_{t=1}^{T} \nabla^2 l_t(\theta_0)$ from the condition (22) does not satisfy conditions of the ergodic theorem. Because of this, it is hard to see how the condition (22) can be enforced in the LARCH (∞) case. This opinion had also been conveyed to us by Prof. L. Giraitis in personal communication; he also suggested that there may exist a modification of the LARCH(∞) process for which these conditions are true but that they are almost certainly cannot be validated in the "classical" version of the LARCH (∞) process considered here.

This does not necessarily mean that the application of QMLE to the data generated by the LARCH process is always bound to fail; indeed, we have implemented this estimator on a standard personal computing platform (PC), and have observed that it performs very well using simulated data, even though the LARCH model is capable of producing significant "outliers", as can be seen from the simulated data in the Figs. 1 and 2 at the end of this article. Despite the apparent algebraic complexity of the Eqs. 19 and 20 one needs to solve to obtain $(\hat{a}, \hat{H})$, the problem poses no difficulty for standard symbolic algebra packages. Using MATLAB's simulations and algebra capabilities (Version 7.0 running on the University of Valparaíso CIMFAV cluster) yielded the best computing times. However, the consistency of the QMLE estimators does seem to be problematic in case of LARCH(∞). The same problem also exists when only a short memory LARCH process considered; such a process has finite number of terms p in the definition of the conditional standard deviation σ_t. More formally, its conditional standard deviation is defined as $\sigma_t = a + \sum_{j=1}^{p} b_j r_{t-j}$ for some integer $p > 0$. For such a process, Truquet (2008) attempts to bypass this difficulty by maximizing a version of the smoothed quasimaximum likelihood function rather than a regular quasimaximum likelihood function. We are not aware of any research in that direction for LARCH(∞) type processes.

In our implementation, which performs an iteration of the algorithm from $i = 0$ to $i = n$, we had to arbitrarily truncate the memory length so as to have a finite series in the model, replacing such summation symbols as $\sum_{j=1}^{\infty}$ by $\sum_{j=1}^{P}$ where P is thus the finite memory length. Implementation with this P also implies that the only values of observations "in the past" that are needed in first P iterations of the algorithm are $i = -P, = P+1, \ldots, -2, -1$. In addition to this new parameter P, in the table below, one recognizes the sample size n, the true values for a and H, and the conditional MLEs $\hat{a}$ and $\hat{H}$. Values are given with 4 significant digits.

n	P	a	H	$\hat{a}$	$\hat{H}$
1,000	50	0.1	0.5	0.09999	0.4960
1,000	100	1	0.7	0.9996	0.6998
1,000	500	0.8	0.8	0.7998	0.7999
1,000	500	0.9	0.6	0.8956	0.5966
1,000	500	0.2	0.9	0.1977	0.9005
1,000	1,000	0.8	0.7	0.8000	0.7000

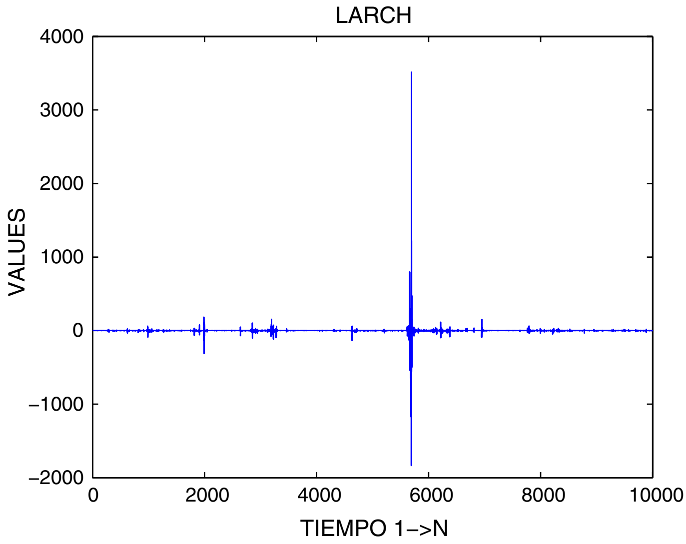

Fig. 1 Observation values for a typical LARCH time series

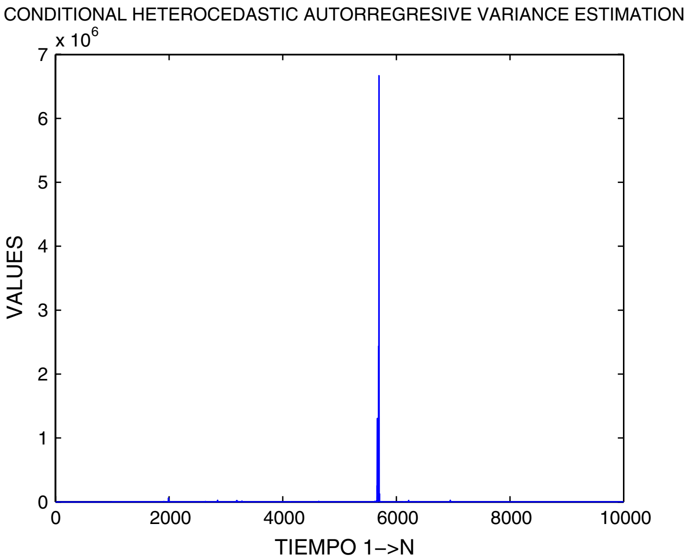

Fig. 2 Variance estimation for the above time series

Heavy-handed truncation (P small) does not seem to effect the estimator at a very significant level, although the second-to-last line shows that using a past memory $P = n$ as long as the data set (or equivalently considering half of the data set as past memory) achieves the very highest precision. Convergence as n increases seems quite rapid: $n = 1,000$ is a reasonable number of data points for the precision attained above.

4 Conditional MLE robustness

In this section we investigate what happens when there is additional exogenous uncertainty on the observations r_j. While a full stochastic-filtering-based treatment of how to extract information dynamically about the true process $(r_j)_j$ given only a noisy observation sequence is beyond the scope of this article, we may still assume that a small amount of error is present in the reported values of r_j, i.e. that we observe instead quantities $q_j = r_j + h_j$, and ask ourselves by how much our estimators $\hat{a}$ and $\tilde{H}$ will be effected by the errors h_j. We will see that this question is difficult to tract analytically, but that nevertheless there is strong mathematical and empirical evidence supporting the claim that our conditional MLE estimators are robust with respect to observation errors.

We simply propose to estimate the magnitude of the error committed on $\hat{H}$ when replacing all the r_j's by all the q_j's. It is thus best to consider that $\hat{a}$ and $\hat{H}$ are functions of the $k = nt$ variables $\bar{r}_k := (r_1, \ldots, r_k)$.Because there is no analytical way of solving the system of two equations yielding $(\hat{a}, \hat{H})$, we must invoke the mean-value theorem assisted by the implicit function theorem in order to evaluate the error

$$e_k = \hat{H}(\bar{r}_k) - \hat{H}(\bar{q}_k).$$

The implicit function theorem tells us that when a system of equations $F(X, Y, \bar{r}_k) = 0$, $G(X, Y, \bar{r}_k) = 0$ has a unique solution (X, Y), the latter can be considered as a function of the equation's parameters (here the r_j's), whose derivatives with respect to these parameters can be calculated as

$$\begin{aligned}\frac{\partial X}{\partial r_j} &= \frac{\partial F}{\partial r_j}\Big/\frac{\partial F}{\partial X} + \frac{\partial G}{\partial r_j}\Big/\frac{\partial G}{\partial X};\\ \frac{\partial Y}{\partial r_j} &= \frac{\partial F}{\partial r_j}\Big/\frac{\partial F}{\partial Y} + \frac{\partial G}{\partial r_j}\Big/\frac{\partial G}{\partial Y}.\end{aligned}$$

Here we will use $X = \hat{a}$, $Y = \hat{H}$, typically omitting the hats as is the practice in implicit function notation, and therefore the functions F and G are the expressions $\partial \log L/\partial a$ and $\partial \log L/\partial H$ given in (19) and (20), that is:

$$\begin{aligned}F(a, H, \bar{r}_k) &= \sum_{i=1}^{k}\left(\frac{1}{\sigma_i} + \frac{r_i^2}{\sigma_i^3}\right)\\ G(a, H, \bar{r}_k) &= \sum_{i=1}^{k}\left(\frac{1}{\sigma_i} + \frac{r_i^2}{\sigma_i^3}\right)\sum_{j=1}^{\infty} r_{i-j}\, j^{H-3/2}\log j\end{aligned}$$

with the understanding that each σ_j is a function of $r_{j-1}, r_{j-2}, \ldots, r_1$ given explicitly in formula (3). Note here that all further "past" observations $r_j : j \le 0$ are assumed to be known, and are not considered as variables in the calculation.

Thus we can calculate

$$\begin{aligned}\frac{\partial F}{\partial a} &= \sum_{i=1}^{k}\frac{\partial F}{\partial \sigma_i}\frac{\partial \sigma_i}{\partial a} = \sum_{i=1}^{k}\frac{\partial F}{\partial \sigma_i}\\ &= -\sum_{i=1}^{k}\left(\frac{1}{\sigma_i^2} + \frac{3r_i^2}{\sigma_i^4}\right)\end{aligned}$$

with

$$\sigma_i = a + \sum_{j=-\infty}^{0} r_j \,(i-j)^{H-3/2} + \sum_{j=1}^{i-1} r_j \,(i-j)^{H-3/2}\,;$$

also since for all $i' > i$, we have

$$\frac{\partial \sigma_{i'}}{\partial r_i} = \left(i'-i\right)^{H-3/2} \log\left(i'-i\right) =: \ell_{i'-i},$$

we get

$$\frac{\partial F}{\partial r_i} = \frac{2r_i}{\sigma_i^3} - \sum_{i'=i+1}^{k} \left(\frac{3r_{i'}^2}{\sigma_{i'}^4} + \frac{1}{\sigma_{i'}^2}\right) \ell_{i'-i}.$$

Similarly, since

$$\frac{\partial \sigma_{i'}}{\partial H} = \sum_{j=1}^{\infty} r_{i-j}\, j^{H-3/2} \log j = \sum_{j=1}^{\infty} r_{i-j} \ell_j,$$

we have

$$\frac{\partial F}{\partial H} = -\sum_{i=1}^{k} \left(\frac{1}{\sigma_i^2} + \frac{3r_i^2}{\sigma_i^4}\right) \sum_{j=1}^{\infty} r_{i-j} \ell_j.$$

For the function G we get immediately

$$\frac{\partial G}{\partial a} = -\sum_{i=1}^{k} \left(\frac{1}{\sigma_i^2} + \frac{3r_i^2}{\sigma_i^4}\right) \sum_{j=1}^{\infty} r_{i-j} \ell_j.$$

A product rule yields

$$\begin{aligned}\frac{\partial G}{\partial r_i} = \frac{2r_i}{\sigma_i^3} \sum_{j=1}^{\infty} r_{i-j} \ell_j - \sum_{i'=i+1}^{k} \left(\frac{3r_{i'}^2}{\sigma_{i'}^4} + \frac{1}{\sigma_{i'}^2}\right) \ell_{i'-i} \sum_{j=1}^{\infty} r_{i'-j} \ell_j \\ + \sum_{i'=i+1}^{k} \left(\frac{1}{\sigma_{i'}} + \frac{r_{i'}^2}{\sigma_{i'}^3}\right) \ell_j,\end{aligned}$$

and

$$\begin{aligned}\frac{\partial G}{\partial H} = -\sum_{i=1}^{k} \left(\frac{1}{\sigma_i^2} + \frac{3r_i^2}{\sigma_i^4}\right) \sum_{j=1}^{\infty} r_{i-j} \ell_j \\ + \sum_{i=1}^{k} \left(\frac{1}{\sigma_i} + \frac{r_i^2}{\sigma_i^3}\right) \sum_{j=1}^{\infty} r_{i-j}\, j^{H-3/2} \log^2 j.\end{aligned}$$

With these formulas we can now express the "error" e_k in our calculation of $\hat{H}$ based on q_i's rather than r_i's, using the Mean Value Theorem:

$$e_k = H\left(\bar{r}_k\right) - H\left(\bar{q}_k\right) = \sum_{i=1}^{k} (r_i - q_i) \left(\frac{\partial F}{\partial r_i} \Big/ \frac{\partial F}{\partial H} + \frac{\partial G}{\partial r_i} \Big/ \frac{\partial G}{\partial H}\right) (\tilde{r}_i) \tag{24}$$

where for each i, the value $\tilde{r}_i$ is in the intervals (q_i, r_i). In the above expression, the quantities a and H also appear, as is logical to expect in a formula issued from the implicit function theorem; these are to be replaced by the functions $\hat{a}$ and $\hat{H}$ evaluated at the common values $\tilde{r}_i$. Thus our calculations can be summarized in the following basic, and naive, statement.

Proposition 2 *The error committed by using an erroneous observation $q_j = r_j + h_j$ instead of r_j in the estimation $\hat{H}$ is equal to the quantity in (24) above, where the notations used therein are introduced in the previous paragraphs.*

Nevertheless, it is perhaps more intelligent to investigate in what way the quantity in (24) is related to basic convergence results such as Proposition 1. Because of the complexity of evaluating the error e_k, we have not been able to find a rigorous stochastic analytic argument to provide a simple criterion for its "smallness". Nevertheless we now present compelling theoretical calculations showing under what circumstances a convergence of e_k to 0 can be expected.

For illustrative purposes, we begin with the slightly simpler question of sensitivity of $\hat{a}$, that is, using $\partial F/\partial a$ instead of $\partial F/\partial H$, omitting tildes and hats for simplicity of notation, we can express

$$\begin{aligned}\sum_{i=1}^{k}(r_i - q_i)\frac{\partial F/\partial r_i}{\partial F/\partial a} &= \sum_{i=1}^{k}(r_i - q_i)\frac{\sum_{i'=i+1}^{k}\left(\frac{3r_{i'}^2}{\sigma_{i'}^4} + \frac{1}{\sigma_{i'}^2}\right)(i'-i)^{H-3/2}\log(i'-i)}{\sum_{i=1}^{k}\left(\frac{3r_i^2}{\sigma_i^4} + \frac{1}{\sigma_i^2}\right)} \\ &\quad - \sum_{i=1}^{k}(r_i - q_i)\frac{2r_i/\sigma_i^3}{\sum_{i=1}^{k}\left(\frac{3r_i^2}{\sigma_i^4} + \frac{1}{\sigma_i^2}\right)} \\ &:= f_1 - f_2. \end{aligned} \tag{25}$$

We estimate the coefficient of $(r_i - q_i)$ in f_1:

$$\begin{aligned} 0 &\le \frac{\sum_{i'=i+1}^{k}\left(\frac{3r_{i'}^2}{\sigma_{i'}^4} + \frac{1}{\sigma_{i'}^2}\right)(i'-i)^{H-3/2}\log(i'-i)}{\sum_{i=1}^{k}\left(\frac{3r_i^2}{\sigma_i^4} + \frac{1}{\sigma_i^2}\right)} \\ &\le \log k \sum_{i'=i+1}^{k}(i'-i)^{H-3/2}\frac{1}{1 + \sum_{i=1,\ldots,k;i\neq i'}\left(\frac{3r_i^2}{\sigma_i^4} + \frac{1}{\sigma_i^2}\right)\Big/\left(\frac{3r_{i'}^2}{\sigma_{i'}^4} + \frac{1}{\sigma_{i'}^2}\right)}. \end{aligned}$$

The random variable $\sum_{i=1,\ldots,k;i\neq i'}\left(\frac{3r_i^2}{\sigma_i^4} + \frac{1}{\sigma_i^2}\right)\Big/\left(\frac{3r_{i'}^2}{\sigma_{i'}^4} + \frac{1}{\sigma_{i'}^2}\right)$ is the sum of $k-1$ positive r.v.'s which are formed with mean- and variance-stationary r.v.'s; thus the sum can be shown to be of order k. In order to attain uniformity in i' (for large i) in this statement, one can again invoke stationarity, plus our hypothesis that all noise terms ε_k are Gaussian, to conclude, after some effort, that the same statement holds almost surely if one is willing to multiply by a power of $\sqrt{\log k}$: this comes from taking a supremum in i' of a sequence of r.v.'s which are bounded by a power of sub-Gaussian r.v.'s. Hence the coefficient of $(r_i - q_i)$ in f_1, which is positive, is bounded by a term of order

$$k^{-1}\log^{1+p/2}k\sum_{i'=i+1}^{k}(i'-i)^{H-3/2},$$

and with the notation $d_k := \sum_{i=1}^{k} i^{H-3/2}$, which is of order $k^{H-1/2}$, we get for some constant c, p and for large n,

$$\begin{aligned}|f_1| &\leq k^{-1}\log^{1+p/2}k\sum_{i=1}^{k}|r_i-q_i|\sum_{i'=i+1}^{k}(i'-i)^{H-3/2}\\ &= ct^{-1}\frac{\log^{1+p/2}n}{n}\sum_{i=1}^{k}|r_i-q_i|\,(k-i)^{H-1/2}.\end{aligned} \tag{26}$$

The term f_2 is much smaller than f_1, as the inequality $2r_i/\sigma_i^3 \leq r_i^2/\sigma_i^4 + 1/\sigma_i^2$ clearly shows. Dealing with $\partial F/\partial H$ instead of $\partial F/\partial a$ is more problematic yet, because of the presence of the mean-zero factor $\sum_{j=1}^{\infty} r_{i-j}\, j^{H-3/2}\log_j$ in the denominators of the terms f_1 and f_2. Nevertheless, since this term coincides with the expression for $\sigma_i - a$ except for the additional $\log j$, its larger magnitude than the stationary $\sigma_i - a$ helps us. Repeating the above considerations for the expressions involving G instead of F involve similar expressions as for F, with combinations of additional factors of the form ℓ_j and $\sum_{j=1}^{\infty} r_{i-j}\ell_j$, and similar conclusions hold, at further calculatory costs. These considerations yield the following more explicit statement than Proposition 2.

Proposition 3 *The error committed by using an erroneous observation $q_j = r_j + h_j$ instead of r_j in the estimation $\hat{H}$ is bounded by the quantity in (26) above.*

The formula in (26) is problematic in the sense that for uniform observation errors, it seems to diverge. Still, it stands to reason to abusively ignore the absolute values in the expression (26), and take advantage of some possible structure for the observation errors. Thus assume that these errors h_i are centered IID with unit variance, and are independent of the observations r_j. We can hence write that the global error e_k should be of the order

$$e_k \asymp \frac{\log^{1+p/2}n}{n^{1-H}}\sum_{i=1}^{k}\frac{h_i}{\sqrt{n}}\left(\frac{k-i}{n}\right)^{H-1/2}$$

Standard approximation results in stochastic calculus show that for some Brownian motion B, the series $\sum_{i=1}^{k}\frac{h_i}{\sqrt{n}}\left(\frac{k-i}{n}\right)^{H-1/2}$ converges to $\int_0^t (t-s)^{H-1/2}\mathrm{d}B(s)$, implying that e_k converges rapidly to 0. More specifically we claim the following strong robustness.

Remark 1 One can expect that, with IID centered observation errors h_i, the resulting error in the estimator $\hat{H}$ converges to 0 faster than any power $n^{-\alpha}$ with $\alpha < 1 - H$.

This remark is also supported by numerical evidence, since our explicit formula (24) for e_k allows us to compute the estimation error empirically.

5 Connection with Hurst parameter for fBm

The connection between LARCH models and fBm is known to be the convergence in distribution of normalized partial sums, for which we have given a simple proof in Sect. 2. We based this proof on a common standard Brownian motion W used to define both an approximating sequence for the fBm B^H—via Donsker's invariance principle and fBm's moving average representation (5)—, and the time series model (σ_j, r_j) from the specification (2, 3), where W and the ε_j's are related via the fact that W is the limit of the process $W^{(n)}$ defined in (8).

With this kind of coupling, where (r, σ) and B^H share the same long memory parameter, we propose in this section a variant on the conditional MLE of Sect. 3, based on observations of the LARCH process r which are close to the increments of fBm in discrete time. The motivation for this variant also comes from avoiding using the LARCH process σ, since the latter can be interpreted as the volatility of a financial time series, which is typically not observed, while the former is interpreted as the sequence of log returns, which *are* observed. This distinction is not as trivial as one may think, and indeed, the next subsection shows that a naive use of the observations r in a linear way to imitate increments of fBm cannot provide a conditional MLE for H, and a use of full information (r, σ) cannot be used for that purpose either. Our conditional MLE with approximate fBm observations based on r must use a non-linear transformation of r, in order to escape from the fact that the r_i are uncorrelated. This entire section does not infirm the conditional MLE of Sect. 3, which is also based on the r_i's, but it gives a conditional MLE with a more natural connection to fBm.

5.1 Some negative results

5.1.1 Direct use of observations r_j

To make our point that a simple-minded use of r_j as representative of fBm observations is bound to fail, consider the following decomposition of fBm, derived from the alternate form (11) of the moving average representation: for $k = k_t = tn$,

$$
\begin{aligned}
aB^H(t) &= a \int_{s=0}^{t} \mathrm{d}s \int_{-\infty}^{s} (s-r)^{H-3/2} \mathrm{d}W(s) \\
&= \sum_{i=1}^{k} \sum_{j'=-\infty}^{i-1} \int_{s=(i-1)/n}^{i/n} \left(\int_{r=(j'-1)/n}^{j'/n} (s-r)^{H-3/2} \mathrm{d}W(r) \right) \mathrm{d}s,
\end{aligned}
$$

which is asymptotically equal to the same quantity with $(s-r)$ replaced by $\left(i-j'\right)$, i.e.

$$
\begin{aligned}
aB^H(t) &= \approx a \sum_{i=1}^{k} \sum_{j'=-\infty}^{i-1} \int_{r=(j-1)/n}^{j/n} \left(\int_{s=(i-1)/n}^{i/n} (i/n - j/n)^{H-3/2} \mathrm{d}s \right) \mathrm{d}W\,(r) \\
&= n^{-H} \sum_{i=1}^{k} \sum_{j'=-\infty}^{i-1} a\sqrt{n} \left\{ W(j'/n) - W\left(\left(\left(j'-1 \right) /n \right) \right) \right\} (i-j')^{H-3/2}
\end{aligned}
$$

This is to be compared with the decomposition

$$
V\left(\frac{k}{n}\right) = n^{-H} v_k = n^{-H} \sum_{i=1}^{k} \sum_{j'=-\infty}^{i-1} r_{j'} \left(i-j'\right)^{H-3/2},
$$

because $V(k/n)$ converges to $B^H(t)$, as we saw in line (9) of the proof of Proposition 1.

Therefore, it is apparent that the analogue, in the continuous-time fBm model, of the observations r_j, are the IID terms $a\sqrt{n}\{W(j/n) - W(((j-1)/n))\} = a\varepsilon_{j-1}$. But there can be no hope, of course, of deriving any estimate of H from these IID noise terms. This negative result is symptomatic of the fact that the observations r_j are uncorrelated, and is also a point supporting the conjecture that, just as in Wang (2002)'s GARCH process study Wang (2002), the experiments of the LARCH(∞) and discretized fBm processes are not statistically equivalent.

The returns r_j are not, however, independent; this is the physical property which we exploit in Sect. 5.2 below.

5.1.2 Volatility observation

To avoid the situation of the previous paragraph, one may naively be tempted to devise a Hurst parameter estimation method based on Proposition 1, i.e. using the volatilities σ_j as observations in addition to the observation of the returns r_j, since $n^{-H}\sum_{j=1}^{k}(\sigma_j - a)$ converges to $B^H(t)$, and thus $n^{-H}(\sigma_j - a)$ can be considered as approximate increments of $B^H(t)$. Econometricians will not consider such modeling as viable, since volatilities are never directly observed. But there is a more fundamental objection to this angle: the reader will easily check that the equations yielding the conditional MLE for (a, H) at time i based on the full past observations $(r_j, \sigma_j)_{j\le i-1}$ are

$$0 = \frac{-1}{\sigma_i} + \frac{r_i^2}{\sigma_i^3},$$
$$0 = \left(\frac{-1}{\sigma_i} + \frac{r_i^2}{\sigma_i^3}\right)\sum_{j=1}^{\infty} j^{H-3/2} r_{i-k}\log k,$$

which is obviously degenerate, yielding infinitely many solutions $\hat{a} = \pm r_i - \sum_{j=1}^{\infty} j^{\hat{H}-3/2}$ r_{i-k}. We believe the phenomenon responsible for this degeneracy is the same issue at work in the previous paragraph.

5.2 Hurst parameter estimation for fBm based on LARCH observations

5.2.1 Squared observations

The following proposition provides the simplest transformation of the r_j's which yields a non-degenerate connection to fBm. It has been established previously in Giraitis et al. (2000). We have summarized and simplified the proof hereafter, because it contains a key calculation which allows us to motivate our conditional MLE.

Proposition 4 *Let n be a fixed integer, with $k = k_t = [tn]$ and define the process V_2 on* $[0, 1]$ *that is continuous and piecewise linear, with values at multiples of* $1/n$ *defined by*

$$V_2\left(\frac{k}{n}\right) = n^{-H}\sum_{i=1}^{k}\left(|r_i|^2 - \mathbf{E}\,|r_i|^2\right) = n^{-H}\sum_{i=1}^{k}\left(|r_i|^2 - \frac{a^2}{1-\|b\|^2}\right).$$

Then V_2 converges in distribution to the fractional Brownian motion $2a^2\left(1-\|b\|^2\right)^{-1}B^H$ as n tends to ∞ where B^H is defined in (5).

Proof We can write

$$\begin{aligned}|r_i|^2 &= \left(|\varepsilon_i|^2 - 1\right)|\sigma_i|^2 + |\sigma_i|^2\\ &= \nu_i + |\sigma_i|^2\end{aligned}$$

where we thus have defined a sequence ν_i of uncorrelated identically distributed random variables. The quantity which we want to show converges to fBm in distribution is

$$n^{-H}\sum_{i=1}^{k}\left(|r_i|^2 - \mathbf{E}|r_i|^2\right) = n^{-H}\sum_{i=1}^{k}(\nu_i - \mathbf{E}\nu_i) + n^{-H}\sum_{i=1}^{k}\left(|\sigma_i|^2 - \mathbf{E}|\sigma_i|^2\right).$$

The second term in the last expression above actually converges to 0 in $L^2(\Omega)$; indeed, because of the uncorrelation of the terms ν_i, and their stationarity, we obtain immediately

$$\begin{aligned}\mathbf{E}\left[\left(n^{-H}\sum_{i=1}^{k}(\nu_i - \mathbf{E}\nu_i)\right)^2\right] &= n^{-2H}\sum_{i=1}^{k}\mathbf{E}\left[(\nu_i - \mathbf{E}\nu_i)^2\right]\\ &= n^{-2H+1}\mathbf{E}\left[|\nu_0 - \mathbf{E}\nu_0|^2\right].\end{aligned}$$

It is thus sufficient to prove the convergence of $n^{-H}\sum_{i=1}^{k}\left(|\sigma_i|^2 - \mathbf{E}|\sigma_i|^2\right)$ to $\frac{2a^2}{1-\|b\|^2}B^H(t)$ in distribution.

It was established in Giraitis et al. (2000, Corollary 5.3) that for any integer ℓ, we can decompose $(\sigma_i)^\ell$ into

$$(\sigma_i)^\ell = \ell a^{-1}\mathbf{E}\left[(\sigma_0)^\ell\right]\sigma_i + y_{i\ell}$$

where $\lim_{n\to\infty}\mathbf{E}\left[\left|n^{-H}\sum_{i=1}^{k}(y_{i\ell} - \mathbf{E}[y_{i\ell}])\right|^2\right] = 0$ and $\mathbf{E}[y_{i\ell}] = -(\ell-1)\mathbf{E}\left[(\sigma_0)^\ell\right]$. Therefore with $\ell = 2$, since $\mathbf{E}|\sigma_i|^2 = a^2/\left(1 - \|b\|^2\right)$, we have

$$n^{-H}\sum_{i=1}^{k}\left(|\sigma_i|^2 - \mathbf{E}|\sigma_i|^2\right) = n^{-H}\sum_{i=1}^{k}(y_{i\ell} - \mathbf{E}[y_{i\ell}]) + \frac{2a}{1-\|b\|^2}n^{-H}\sum_{i=1}^{k}(\sigma_i - a).$$

Proposition 1 guarantees that the last term above converges in distribution to $\frac{2a^2}{1-\|b\|^2}$ times a fractional Brownian motion, while the first term converges to 0 in $L^2(\Omega)$. □

The above proof shows that each term $n^{-H}\left(|r_i|^2 - \frac{a^2}{1-\|b\|^2}\right)$ in the above proposition is asymptotically close to a fractional Brownian increment $2\frac{a^2}{1-\|b\|^2}\left(B^H((i+1)/n) - B^H(i/n)\right)$. Because of this, it is natural to propose a conditional MLE for estimating a and H based on the observations

$$x_i := n^{-H}\left(|r_i|^2 - \mathbf{E}|r_i|^2\right).$$

We will not prove consistency, since it would be mathematically significantly more involved than the proof of the robustness Proposition 3.

We now present the equations for this new conditional MLE. It presents an added difficulty that the observations $x_i = n^{-H}(|r_i|^2 - a^2/\left(1 - \|b\|^2\right))$ depend on a and H, and that the sign of r_i remains undetermined, so that there is uncertainty in the expression of σ_i using these observations x_i. More specifically, we will be obliged to write

$$\sigma_i = a + \sum_{j=1}^{\infty} Y_{i-j} b_j \left|r_{i-j}\right| = a + \sum_{j=1}^{\infty} Y_{i-j} b_j \sqrt{n^H x_i + a^2 / \left(1 - \|b\|^2\right)} \tag{27}$$

where Y_i is an IID sequence of random variables (independent of the observations) which equal $+1$ or -1 with equal probabilities, under their probability measure $\mathbf{P}_Y$. The likelihood function for r_i given $x_0, x_1, \ldots, x_{i-1}$ can thus be represented as

$$L = \mathbf{E}_Y[L_Y] := \mathbf{E}_Y\left[\exp -\frac{1}{2}\left(\log 2\pi + 2\sum_{j=1}^{i} \log\left|\sigma_j\right| + \sum_{j=1}^{i} \frac{r_j^2}{\sigma_j^2}\right)\right]$$

where σ_j is to be replaced by (27) and

$$r_j^2 = n^H x_j + \frac{a^2}{1 - \|b\|^2}. \tag{28}$$

Note that $\|b\|^2$ depends on H, and that we have

$$\frac{\mathrm{d}\,\|b\|^2}{\mathrm{d}H} = c^2 \sum_{j=1}^{\infty} 2 j^{2H-3} \log j.$$

We have the following partial derivatives:

$$\frac{\partial r_j^2}{\partial a} = \frac{2a}{1 - \|b\|^2}; \quad \frac{\partial r_j^2}{\partial H} = \frac{H}{n^{1-H}} x_j + \frac{a^2}{\left(1 - \|b\|^2\right)^2} \frac{\mathrm{d}\,\|b\|^2}{\mathrm{d}H} \tag{29}$$

and

$$\frac{\partial \sigma_j}{\partial a} = 1 + \frac{1}{1 - \|b\|^2} \sum_{j=1}^{\infty} Y_{i-j} \frac{b_j}{2\left|r_{i-j}\right|} \tag{30}$$

$$\frac{\partial \sigma_j}{\partial H} = \frac{a}{\left(1 - \|b\|^2\right)^2} \frac{\mathrm{d}\,\|b\|^2}{\mathrm{d}H} \sum_{j=1}^{\infty} Y_{i-j} \frac{b_j}{2\left|r_{i-j}\right|} + \sum_{j=1}^{\infty} Y_{i-j} \left|r_{i-j}\right| b_j \log j. \tag{31}$$

Therefore, the maximum likelihood estimator $\left(\hat{a}, \hat{H}\right)$ is obtained as the solution of the following integro-differential system

$$0 = \mathbf{E}_Y\left[L_Y \sum_{j=1}^{i}\left(\left(\frac{r_j^2}{\sigma_j^3} - \frac{1}{\sigma_j}\right)\frac{\partial \sigma_j}{\partial a} - \frac{1}{2}\frac{1}{\sigma_j^2}\frac{\partial r_j^2}{\partial a}\right)\right], \tag{32}$$

$$0 = \mathbf{E}_Y\left[L_Y \sum_{j=1}^{i}\left(\left(\frac{r_j^2}{\sigma_j^3} - \frac{1}{\sigma_j}\right)\frac{\partial \sigma_j}{\partial H} - \frac{1}{2}\frac{1}{\sigma_j^2}\frac{\partial r_j^2}{\partial H}\right)\right] \tag{33}$$

given the above formulas for the various partial derivatives.

5.2.2 Practical implementation

In practice, we use only a finite memory horizon P instead of ∞, as we did in the conditional MLE of Sect. 3 (see the description of the table of results). Thence the formulas (29)–(31) for the partial derivatives above have sums $\sum_{j=1}^{P}$ instead of $\sum_{j=1}^{\infty}$, the expectation symbols $\mathbf{E}_Y$ in (32) and (33) can be replaced by the summation symbols $\sum_{m=0}^{2^P-1} 2^{-P}$, with the notation L_m instead of L_Y, and the understanding that Y_j must be replaced by m_j where m_j is the jth term in the binary expansion of m. In order to evaluate the expressions in (32) and (33), it is useful to divide L_Y by $\mathbf{E}L_Y$ (or divide L_m by $\sum_{m=0}^{2^P-1} 2^{-P} L_m$), in order to deal with convex, rather than possibly very large, coefficients.

Additional simplification can be obtained by noting that in practice, the first summand in the expression for $\partial r_j^2/\partial H$ in (29) is of the order n^{-H}. Since the mean value theorem implies that its effect can be considered as replacing r_j by $r_j + h_j$ where the error term h_j is bounded above by $n^{-1}\left(|r_j| + \frac{a^2}{(1-\|b\|^2)|r_j|}\right)$, our robustness results in Sect. 4 show that neglecting this term should not change the estimator's consistency.

A further simplification is to replace the averaging over the Bernoulli random variables Y_i by a Monte-Carlo implementation of this average, using far fewer terms than a sum $\sum_{m=0}^{2^P-1} 2^{-P}(\cdots)$. However, one can show that the distribution of $\hat{a}$ and $\hat{H}$ is invariant with respect to the actual signs of the increments ε_j. Therefore the above implementation can be performed with a single random sequence Y_j, i.e. without any averaging. This amounts to choosing the signs of the r_j's arbitrarily, according to a distribution consistent with the model. Using $Y_j = m_j$, the jth term in the binary expansion of a pseudo-random number $m \in (0, 1)$ is of course an appropriate choice. The resulting scheme is then no more complex than the original Conditional MLE of Sect. 3.

One may also consider schemes based on moments of order $2p$ for p any integer, not just $p = 1$. Although we leave the derivation of the analogues of formulas (32) and (33) to the reader in this case, such analogues are obtained in exactly the same way, and the same simplifications apply. Indeed, a proof nearly identical to that of Proposition 4 shows that, with g_{2p} the $2p$th moment of the noise terms ε_i, the sum of the observations

$$\begin{aligned} x_i^{(p)} &:= n^{-H}\left(|r_i|^{2p} - \mathbf{E}\,|r_i|^{p}\right) \\ &= n^{-H}\,|r_i|^{2p} - n^{-H}a^{-1}g_{2p}\mathbf{E}\,|\sigma_0|^{2p} \end{aligned}$$

converge to fractional Brownian motion multiplied by the scaling factor $2pg_{2p}a^{-1}\mathbf{E}\left[|\sigma_0|^{2p}\right]$. Such higher-order moments may result in faster convergence of the conditional MLE. It is very important to realize that this method works only for even moments, under our assumption of Gaussian noises ε_j. Indeed, while the partial sum of the $x_i^{(p)}$ converges in distribution for all integer orders q, it yields convergence to 0 when $g_q = 0$.

6 A local Whittle-type estimator of the Hurst parameter

6.1 A local Whittle estimator

In this section, we revert to denoting discrete time by t instead of i or k, because the latter two letters are used in standard roles for Whittle estimators. We also use another standard notation $\theta := 2 - 2H$. Recall that the LARCH model

$$\sigma_t = a + \sum_{j=1}^{\infty} b_j r_{t-j};\ \ r_t = \sigma_t \varepsilon_t$$

with $a \neq 0$ is weakly stationary iff (7) holds. Recall that we are interested in the long-memory case, that is where the Hurst parameter $0.5 < H < 1$. Defining $b_j = O(j^{H-3/2})$, with small enough a, we ensure that (7) is true. It is also true (e.g. Corollary 2.1 in Giraitis et al. 2000) that

$$\gamma(h) = \mathrm{Cov}(\sigma_0, \sigma_h) \sim h^{2(H-1)} = h^{-\theta} \tag{34}$$

which means that the covariance and, by extension, a correlation function decreases very slowly as $h \to \infty$ since $-1 < -\theta = 2(H-1) < 0$. Suppose we want to have a consistent estimator of the Hurst parameter H. It seems that a possible candidate that converges to the true value H in probability is the localized version of the Whittle estimator described as Theorem 2.1 in Dalla et al. (2004). First, imagine that the volatility process σ_t can be observed directly. Using such an estimator means using the periodogram of the process σ_t defined as

$$I_n(\lambda) = n^{-1} \left| \sum_{t=1}^{n} \sigma_t \mathrm{e}^{-it\lambda} \right|^2.$$

The periodogram $I_n(\lambda)$ measures the contribution of the frequency λ to the overall "energy" of the process σ_t. By definition, $I_n(\lambda) = \sum_{h=-\infty}^{\infty} \gamma(h) \exp(-ih\lambda)$. On the other hand, we know that, for any frequency $\lambda_j = \frac{2\pi j}{n}$ with j integer,

$$I_n(\lambda_j) = \sum_{|h|<n} \hat{\gamma}(h) \mathrm{e}^{-ih\lambda_j}.$$

Therefore, it is tempting to say that $I_n(\lambda)$ can be used as an estimator of $2\pi f(\lambda)$. It is a well known fact, however (see any time series textbook, e.g. Brockwell and Davis 2002) that the periodogram per se is not a consistent estimator of the spectral density. If σ_t were a sequence of iid Gaussian variables, we would have the joint distribution of $\{I_n(\lambda_1), \ldots, I_n(\lambda_m)\}$ as

$$F(x_1, \ldots, x_m) = \prod_{i=1}^{m} \left(1 - \exp\left\{ \frac{-x_i}{2\pi f(\lambda_i)} \right\} \right),$$

for any integer m. Consequently, periodograms would converge to a set of independent exponential random variables with means $2\pi f(\lambda_i)$, $i = 1, \ldots, m$. In order to obtain a consistent estimator of the spectral density $f(\lambda)$, averaging over Fourier frequencies λ_j would be done, resulting in an estimator belonging to the class of *discrete spectral density estimators*. They are defined as

$$\hat{f}(\lambda) = \frac{1}{2\pi} \sum_{|j| \le m_n} W_n(j) I_n \left(g(n, \lambda) + \frac{2\pi j}{n} \right) \tag{35}$$

where $m_n \to \infty$, $m_n/n \to 0$ as $n \to \infty$ and $g(n, \lambda)$ is the multiple of $2\pi/n$ closest to λ. The weights $W_n(j)$ have to be even, non-negative, add up to 1 and be such that $\sum_{|j| \ge m_n} W_n^2(j) \to 0$ as $n \to \infty$. Again, for details see Brockwell and Davis (2002).

Thus, the periodogram has to be smoothed in order for it to be a consistent estimator of the spectral density, and, by extensions, to provide a consistent estimator of the long-memory parameter. Note that the "window" m_n used in (35) provides for a *local* as opposed to the generic Whittle method.

Now define

$$\lambda_j = \frac{2\pi j}{n}, \quad j = 1, \ldots, m,$$

i.e. the local Fourier frequencies, and

$$I_n(\lambda_j) = n^{-1} \left| \sum_{t=1}^{n} \sigma_t e^{it\lambda_j} \right|^2 \tag{36}$$

as the periodogram of the sequence σ_t, $t = 1, \ldots, n$ and $m = m_n$ is an integer bandwidth parameter such that $m \to \infty$ and $m = o(n)$ as $n \to \infty$. The local Whittle estimator of the parameter

$$\theta := 2 - 2H$$

can be defined as the minimizer

$$\alpha \equiv \hat{\theta}_n = \mathrm{argmin}_{[-1,1]} U_n(\alpha)$$

of the quasi-likelihood-type objective function

$$U_n(\alpha) = \log \left(\frac{1}{m} \sum_{j=1}^{m} \lambda_j^{\alpha} I_n(\lambda_j) \right) - \frac{\alpha}{m} \sum_{j=1}^{m} \log \lambda_j. \tag{37}$$

6.2 Local Whittle estimator based on squared returns

Unfortunately, σ_t is the "volatility" process and, as such, should not be presumed observable. Thus, the problem is to find a suitable substitute process that still allows us to extract enough information to estimate the Hurst parameter H. The simplest choice appears to be the squared returns r_t^2.

To show that the Whittle local-likelihood based estimator of $\theta = 2 - 2H$ using squared returns is consistent, we may verify assumptions A and B in the main result of Dalla et al. (2004).

- Assumption A requires the process r_t^2 to be weakly stationary and to have the spectral density of the form

$$f(\lambda) = |\lambda|^{-\alpha_0} L(\lambda) \tag{38}$$

  where $L(\lambda) \to b_0$ as $|\lambda| \to 0$, $0 < b_0 < \infty$ and $|\alpha_0| < 1$. That r_t^2 is a weakly stationary process is clear. Moreover, according to Theorem 2.2 in Giraitis et al. (2000), we have

$$\mathrm{Cov}(r_0^2, r_t^2) \sim c_2^2 t^{-\theta}$$

  when $t \to \infty$ where the constant c_2^2 does not depend on t and $0 < \theta < 1$. This implies that the spectral density function of the process r_t^2 is $f(\lambda) \sim \lambda^{-\theta}$ as $\lambda \to 0$. So we do indeed have condition (38) with $\alpha_0 = \theta = 2 - 2H$. This means that the Assumption A of Dalla et al. (2004) is fulfilled.
- Instead of the periodogram for σ in (36), we now have

$$I_n(\lambda_j) = \frac{1}{n} \left| \sum_{t=1}^{n} r_t^2 e^{-i\lambda_j t} \right|^2.$$

where $\lambda_j = \frac{2\pi j}{n}$, $j = 1, \ldots, m$. Then, Assumption B requires that renormalized periodograms of the process r_t^2, i.e. $\eta_j^* = \frac{I_n(\lambda_j)}{b_0\lambda_j^{-\theta}}$, satisfy the weak law of large number (WLLN). We now discuss the issue of verifying this condition.

Dalla et al. (2004) suggest a simple sufficient condition that enables us to claim that the Assumption B is true. Let us denote

$$\Delta_m = \max_{1\le k\le m} \mathbf{E}\left|\sum_{j=1}^{k}(\eta_j^* - \mathbf{E}\,\eta_j^*)\right|$$

Then, $\Delta_m = o(m)$ implies Assumption B; for details, see Proposition 2.2 in Dalla et al. (2004).

By definition,

$$\begin{aligned}
I_n(\lambda_j) &= \frac{1}{n}\left|\sum_{t=1}^{n} r_t^2 \mathrm{e}^{-it\lambda_j}\right|^2 \\
&= \frac{1}{n}\left[\sum_{t=1}^{n} r_t^4 + \sum_{t\neq s=1}^{n} r_t^2 r_s^2 \cos(\lambda_j|t-s|)\right] \\
&= \frac{1}{n}\left[\sum_{t=1}^{n} r_t^4 + 2\sum_{t=1}^{n}\sum_{h=1}^{n-t} r_t^2 r_{t+h}^2 \cos(\lambda_j h)\right] \qquad (39)
\end{aligned}$$

Therefore,

$$\eta_j^* - \mathbf{E}\eta_j^* = \frac{1}{nb_0\lambda_j^{-\alpha_0}}\left[\sum_{t=1}^{n}(r_t^4 - \mathbf{E}\,r_t^4) + 2\sum_{t=1}^{n}\sum_{h=1}^{n-t}(r_t^2 r_{t+h}^2 - \mathbf{E}\,r_t^2 r_{t+h}^2)\cos(\lambda_j h)\right]$$

Note also that r_t^2 and r_t^4 are strictly stationary.

Let us first handle the first term in (39). It is easy to find out that

$$\begin{aligned}
\mathbf{E}\left[\frac{1}{n}\sum_{t=1}^{n}(r_t^4 - \mathbf{E}\,r_t^4)\right]^2 &= \frac{1}{n^2}\mathbf{E}\sum_{i=1}^{n}(r_t^4 - \mathbf{E}\,r_t^4)^2 \\
&\quad + \frac{2}{n^2}\sum_{t_1,t_2=1;t_1<t_2}^{n} \mathbf{E}\,(r_{t_1}^4 - \mathbf{E}\,r_{t_1}^4)(r_{t_2}^4 - \mathbf{E}\,r_{t_2}^4)
\end{aligned}$$

Giraitis et al. (2004) can be consulted for the fact that if $\mu_{2k} := \mathbf{E}\varepsilon_t^{2k} < \infty$ and $\sum_{p=2}^{2k} ||b||_p^p|\mu_p| < 1$, then $\mathbf{E}\,r_t^{2k} < \infty$. With $k = 4$, we guarantee the existence of the 8th moment, and, therefore, $\frac{1}{n^2}\mathbf{E}\sum_{t=1}^{n}(r_t^4 - \mathbf{E}\,r_t^4)^2 = O(n^{-1})$. The remaining portion of the first term in (39) is the expression

$$\frac{2}{n^2}\sum_{t_1,t_2=1;t_1<t_2}^{n} \mathbf{E}\,(r_{t_1}^4 - \mathbf{E}\,r_{t_1}^4)(r_{t_2}^4 - \mathbf{E}\,r_{t_2}^4) = \frac{2}{n^2}\sum_{t=1}^{n}\sum_{h=1}^{n-t}\mathrm{Cov}(r_t^4, r_{t+h}^4)$$

which can be handled using the fact that for any positive integer $j > 2$, we have $Cov\,(r_0^j, r_h^j) \sim c_j^2 h^{-\theta}$ (see Giraitis et al. 2000). Here $c_4 = \frac{4c_1}{a}\mathbf{E}\,(r_0^4)$ is the constant that does not depend on the difference $h = t_1 - t_2$; c_1 depends on $a, b_j, j = 1, 2, \ldots$ and θ only. Thus, we have

$$\frac{2}{n^2}\sum_{t_1,t_2=1;t_1<t_2}^{n}\mathbf{E}\,(r_{t_1}^4-\mathbf{E}\,r_{t_1}^4)(r_{t_2}^4-\mathbf{E}\,r_{t_2}^4)=\frac{2}{n^2}\sum_{t=1}^{n}\sum_{h=1}^{n-t}\mathrm{Cov}\,(r_t^4,r_{t+h}^4)$$
$$\sim n^{-2}\sum_{t=1}^{n}\sum_{h=1}^{n-t}h^{-\theta}f=\frac{2}{n^2}\sum_{t=1}^{n}(n-t)^{1-\theta}\sim n^{-\theta}$$

Consequently, $\mathbf{E}\left[\frac{1}{n}\sum_{t=1}^{n}(r_t^4-\mathbf{E}\,r_t^4)\right]^2=o(1)$ and, by Jensen's inequality,

$$\frac{1}{n}\mathbf{E}\sum_{t=1}^{n}|r_t^4-\mathbf{E}\,r_t^4|\le\sqrt{\frac{1}{n^2}\mathbf{E}\sum_{t=1}^{n}(r_t^4-\mathbf{E}\,r_t^4)^2}=o(1)$$

as $n\to\infty$.

The second term

$$\frac{2}{nb_0\lambda_j^{-\alpha_0}}\left[\sum_{t=1}^{n}\sum_{h=1}^{n-t}(r_t^2r_{t+h}^2-\mathbf{E}\,r_t^2r_{t+h}^2)\cos(\lambda_j h)\right] \tag{40}$$

is much harder to handle. It should involve the study of more complicated moments of r_t^2 which seems to be undesirable. In particular, it should be necessary to establish a property of mixed moments analogous to Cov $(r_0^j,r_h^j)\sim c_j^2h^{-\theta}$; in other words, to investigate asymptotic behavior (as $t\to\infty$) of "mixed moments" of the form Cov $(r_0^jr_t^j,r_h^jr_{t+l}^j)$ for positive integer $j>2$ and integer $h,l>0$. Indeed, one can prove by elementary calculations that any failure to distinguish between the various covariances for different t,h,l, i.e. any attempt to use only covariances of the form Cov (r_0^j,r_h^j), yields a second term whose variance diverges. No elementary ways to solve this problem are clearly visible. The estimation of the "mixed moments" is a worthy open problem in its own right.

Appendix: Proof of Lemma 1

Let $\tau(s)$ be the limit of τ_j in $L^2(\Omega)$, where $j=[sn]$. We show first that $\tau(s)$ exists and equals the constant a. Indeed, as σ_j is independent of all noise terms $\varepsilon_{j'}$ for $j'\ge j$, we actually have

$$\tau_j=\mathbf{E}\left[\sigma_j|\mathcal{F}_{\{j-J,j-1\}}\right]$$

where $\mathcal{F}_{\{j-J,j-1\}}$ is the sigma-field generated by $\varepsilon_{j-J},\varepsilon_{j-J+1},\ldots,\varepsilon_{j-1}$. Since J is fixed, as n tends to infinity, this sigma field, which is a sub-sigma-field of $\mathcal{F}^W_{[j/n-J/n;j/n]}$, converges to the trivial sigma-field by continuity of W. Since σ_j is square-integrable, τ_j is a non-random function of the finite number of random variables generating $\mathcal{F}_{\{j-J,j-1\}}$:

$$\tau_j=g^*\left(\varepsilon_{j-J},\varepsilon_{j-J+1},\ldots,\varepsilon_{j-1}\right),$$

g^* is the function that minimizes $\mathbf{E}\left[\left(\sigma_j-g\left(\varepsilon_{j-J},\varepsilon_{j-J+1},\ldots,\varepsilon_{j-1}\right)\right)^2\right]$, and τ_j converges almost surely to a constant c. Since c thus minimizes $\mathbf{E}\left[\left(\sigma_j-c\right)^2\right]$, $c=\mathbf{E}\left(\sigma_j\right)=a$, and therefore $\lim_{n\to\infty}\tau_j=a$ almost surely and in $L^2\left(\Omega\right)$. By stationarity of σ, the convergence of τ_j to a in $L^2\left(\Omega\right)$ is uniform in j. We denote the common value of $\mathbf{E}\left[\left(\tau_j-a\right)^2\right]$ by $h\left(n\right)$, and thus we have just proved that $\lim_{n\to\infty}h\left(n\right)=0$.

Next we study $V_2(k/n)$: because ε_j is independent of τ_j and of the previous $\varepsilon_{j'}$'s for $j' < j$, we get

$$\mathbf{E}\left[\left(V_2\left(k/n\right) - a\sum_{i=1}^{k}\frac{1}{n}\sum_{j=-\infty}^{i-1}\left(\frac{i-j}{n}\right)^{H-3/2}\varepsilon_j n^{-1/2}\right)^2\right]$$
$$= \mathbf{E}\left[\left(\frac{1}{n}\sum_{j=-\infty}^{k-1}\varepsilon_j n^{-1/2}\left(\tau_j - a\right)\sum_{i=j+1}^{k}\left(\frac{i-j}{n}\right)^{H-3/2}\right)^2\right]$$
$$= \frac{h\left(n\right)}{n^3}\sum_{j=-\infty}^{k-1}\left(\sum_{i=j+1}^{k}\left(\frac{i-j}{n}\right)^{H-3/2}\right)^2. \tag{41}$$

The last sum above, together with the factor n^{-3}, is the Riemann sum approximation of the integral

$$\int_0^t\left(\int_s^t (r-s)^{H-3/2}\mathrm{d}r\right)^2 \mathrm{d}s = (H-1/2)^{-1}(H+1/2)^{-1}t^{H+1/2}.$$

Therefore, since we already proved that $\lim_{n\to\infty} h(n) = 0$, we have proved that the quantity in (41) converges to 0 as n tends to $+\infty$, which means that in the definition (10) of $V_2(k/n)$, we may replace τ_j by a as far as $L^2(\Omega)$-convergence goes.

Consequently, the lemma will be proved as soon as we can establish the following convergence in distribution:

$$\lim_{n\to\infty}\sum_{i=1}^{k}\frac{1}{n}\sum_{j=-\infty}^{i-1}\left(\frac{i-j}{n}\right)^{H-3/2}\varepsilon_j n^{-1/2} = \int_{s=0}^{t}\mathrm{d}s\int_{-\infty}^{s}(s-r)^{H-3/2}\mathrm{d}W(r) \tag{42}$$

This follows easily by first noting that the Wiener stochastic integral above can be approximated in $L^2(\Omega)$ by its Riemann sums over the partition $\{j/n : j = -\infty, \ldots, i-1\}$ of $[-\infty, i = [ns]/n]$, in which the only relevant values of increments of W are for these partition points; then one replaces W by its approximation $W^{(n)}$, which is straightforward because of the evaluation at partition points only, convergence in distribution being guaranteed by the convergence of $W^{(n)}$ to W from Donsker's invariance principle. The resulting Riemann sums coincide exactly with the discrete term in (42). The only remaining discrepancy comes from using $i = [ns]/n$ instead of $i = s$ above; this is resolved by using the Riemann-sum approximation of the Riemann integral in (42), which is easily done at the start of the evaluation described here, before discretizing the Wiener-Itô integral, by first performing a Fubini on the integrals in (42), and then replacing the Riemann integral by its Riemann sum, which causes an error in $L^2(\Omega)$ proportional to the square of the Riemann sum error, and thus also converges to 0. We omit all these cumbersome and elementary calculations.

Acknowledgments This work was supported partially by the research project Fundación Andes under the project C-13955/10, and Fondecyt Reg. N. 1050843, Chile. The second author was partially supported by the research project Nucleus Millennium P04-069-F "Information and Randomness: Fundamentals and Applications; Laboratories in Genome Mathematics and Stochastic Simulation." The third author was partially supported by NSF grants DMS-0204999 and DMS-0606615. The third author wishes to thank the second author and the Department of Statistics at the University of Valparaíso, Chile, where a portion of the research was conducted. All authors thank an anonymous referee for pointed comments and suggestions which resulted in several improvements, including the fixing of a fundamental error.

References

Abadir KM, Distaso W, Giraitis L (2005) Semiparametric estimation and inference for trending I(d) and related processes. Ann Stat 2006 (to appear)

Abadir KM, Distaso W, Giraitis L (2006) Local Whittle estimation, fully extended for nonstationarity. Preprint

Arteche J (2004) Gaussian semiparametric estimation in long memory in stochastic volatility and signal plus noise models. J Econom 119:131–154

Basawa IV, Feygin PD, Heyde CC (1976) Asymptotic properties of maximum likelihood estimators for stochastic processes. Sankhya: Indian J Stat Ser A 38(Pt.3):259–270

Boes DC, Davis RA, Gupta SN (1989) Parameter estimation in low order fractionally differenced ARMA processes. Stoch Hydrol Hydraul 3:97–110

Brockwell P, Davis R (2002) Introduction to time series and forecasting, 2nd ed. Springer, New York

Cheridito P (2003) Arbitrage in fractional Brownian motion models. Finance Stoch 7:533–553

Cheung YW (1993) Test for fractional integration: a Monte Carlo investigation. J Time Ser Anal 14:331–345

Dahlhaus R (1989) Efficient parameter estimation for self-similar processes. Ann Stat 17(4):1749–1766

Dalla V, Giraitis L, Hidalgo J (2004) Consistent estimation of the memory parameter for nonlinear time series. J Time Ser Anal 27:211–251

Fan J, Yao Q (2003) Nonlinear time series. Springer Verlag, New York

Fox R, Taqqu MS (1986) Large sample properties of parameter estimates for strongly dependent stationary Gaussian time series. Ann Stat 14:517–532

Geweke J, Porter-Hudak S (1983) The estimation and application of long-memory time series models. J Time Ser Anal 4:221–238

Ghysels E, Harvey AC, Renault E (1996) Stochastic volatility. In: Maddala G, Rao G (eds) Statistical methods in finance, vol 14 of Handbook of statistics. North-Holland, Amsterdam, pp 119–191

Giraitis L, Surgailis D (1990) A central limit theorem for quadratic forms in strongly dependent linear variables and application to asymptotical normality of Whittle's estimate. Probab Theory Relat Fields 86:87–104

Giraitis L, Robinson PM, Surgailis D (2000) A model for long memory conditional heteroskedasticity. Ann Appl Probab 10(3):1001–1024

Giraitis L, Leipus R, Robinson PM, Surgailis D (2004) LARCH, leverage, and long memory. J Financ Econom 2(2):177–210

Gouriéroux C (1997) ARCH models and financial applications, Springer Series in Statistics. Springer Verlag, New York

Hurvich CM, Brodsky J (2001) Broadband semiparametric estimation of the memory parameter of a long-memory time series using fractional exponential model. J Time Ser Anal 22:221–249

Hurvich CM, Moulines E, Soulier P (2005) Estimation of long memory in volatility. Econometrica 73: 1283–1328

Jensen MJ (1999) Using wavelets to obtain a consistent ordinary least squares estimator of the long-memory parameter. J Forecast 18:17–32

Jensen MJ (2000) An altenative maximum likelihood estimator of long-memory process using compactly supported wavelets. J Econ Dyn Control 24:361–387

Karatzas I, Shreve SE (1991) Brownian motion and stochastic calculus, 2nd edn. Springer, New York

Künsch H (1987) Statistical aspects of self-similar processes. In: Prohorov YA, Sazonov VV (eds) Proceedings of the 1st world congress of the Bernoulli society, vol 1. Science Press, Utrecht, pp 67–74

Lahiri SN (2003) A necessary and sufficient condition for asymptotic independence of discrete Fourier Transforms under short- and long- range dependence. Ann Stat 31(2):613–641

Moulines E, Soulier P (1999) Broad band log-periodogram regression of time series with long range dependence. Ann Stat 27:1415–1439

Nelson DB (1990) ARCH models as diffusion approximations. J Econom 45(1–2):7–38

Robinson PM (1991) Testing for strong serial correlation and dynamic conditional heteroskedasticity in multiple regression. J Econom 47:67–84

Robinson PM (1995a) Log-periodogram regression of time series with long range dependence. Ann Stat 23:1048–1072

Robinson PM (1995b) Gaussian semiparametric estimation of long range dependence. Ann Stat 23:1630–1661

Shimotsu K, Phillips PCB (2005) Exact local Whittle estimation of fractional integration. Ann Stat 32:656–692

Sowell F (1992) Maximum likelihood estimation of stationary univariate fractionally integrated time series models. J Econom 53:165–188
Szabados T (2001) Strong approximation of fractional Brownian motion by moving averages of simple random walks. Stoch Process Appl 92:31–60
Truquet L (2008) A new smoothed quasi maximum likelihood estimator for autoregressive process with LARCH errors. Preprint
Wang Y-Z (2002) Asymptotic nonequivalence of GARCH models and diffusions. Ann Stat 30(3):754–783
Weiss AA (1986) Asymptotic theory for ARCH models: estimation and testing. Econom Theory 2:107–131