Consistent Estimators for the Long-Memory Parameter in LARCH And Fractional Brownian Models

by

Michael Levine
Purdue University

Soledad Torres
Universidad de Valparaiso

Frederi Viens
Purdue University

Technical Report #07-02

Department of Statistics
Purdue University
West Lafayette, IN USA

February 2007

Consistent estimators for the long-memory parameter in LARCH and fractional Brownian models*

Michael Levine [†] Soledad Torres [‡] Frederi Viens [†]

December 19, 2006

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 named fractional Brownian motion (fBm). A LARCH model's parameter is estimated using a conditional maximum likelihood method, which is proved to be consistent, to allow model validation via a portmanteau theorem, and to have good stability properties. A local Whittle estimator is also discussed. By constructing the LARCH and fBm processes on a common probability space, and showing the convergence of various partial sums of the former to the latter in L^2, the article is able to propose a specially designed conditional maximum likelihood method for estimating the fBm's Hurst parameter. 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".

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 modelled by conditional variance

***Key words and phrases:** ARCH, time series, fractional Brownian motion, maximum likelihood estimator, long memory Whittle estimator, moving average

MSC 2000 subject classifications: primary 62M09; secondary 60G18, 62M10, 91B84.

[†]Dept. Statistics, Purdue University 150 N. University St., West Lafayette, IN 47907-2067, USA
`mlevins@stat.purdue.edu` `viens@purdue.edu`

[‡]Depto de Estadística – CIMFAV, Universidad de Valparaíso, 1091 Av. Gran Bretaña, Playa Ancha, Valparaíso, Chile
`soledad.torres@uv.cl`

(volatility) models, such as the models traditionally included in the so-called (G)ARCH framework (see Gourieroux (1997) [17] and also Ghysels, Harvey, Renault (1996) [18]). However, for the most part, these models possess the so-called short memory, which means the 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) [31]). The linear autoregressive conditional heteroscedasticity model (LARCH), first introduced in Robinson (1991) [32], has long been considered a very convenient vehicle for long-term memory modeling. Its name is probably due to Giraitis, Robinson and Surgailis (2000) [14]. This model can be described as

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

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 models 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, Robinson and Surgailis (2000) [14]. This, and the lack of a complete understanding of the long-memory modeling potential of the ARCH framework, leads many authors into adopting LARCH as their primary long-memory modeling vehicle.

Giraitis, Robinson and Surgailis (2000) [14] also prove that, with proper normalization, 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

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

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 this article, we show in contrast 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). Moreover, we take this one step further and show, besides just convergence in law, that, with the right normalization, the LARCH model converges to $B^H(t)$ in $L^2(\Omega)$ where Ω is a single probability space supporting both processes. That is, we claim that there really is an underlying fBm process on the same probability space as the LARCH model that converges to it.

Due to the last result, it is legitimate to suggest estimating the memory parameter of a nonlinear time series process that approximates the fBm process instead of the Hurst parameter of the fBm process itself. In other words, we propose using methods commonly employed in theory of nonlinear time series to provide an estimate of the Hurst parameter of the original fBm process. We view this article as a first exploration of such proposals, opening new lines of parameter estimation research for long-memory continuous-time stochastic processes. In particular, we feel that the quantitative financial community should appreciate the use of approximations of such processes by 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 explicitly exhibit long-memory. This is in contrast to the often criticized so-called geometric fBm model, where the stock returns are correlated directly according to an fBm, and the volatility parameter is constant, in a naive generalization of the Black-Scholes model.

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) [9] 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) [4]. The exact MLE method was proposed in Sowell (1992)[36] and the approximate one in Fox and Taqqu (1986) [12] (using the frequency domain approach). Wavelet-based MLE methods for the long-memory parameter estimation were proposed in Jensen (1999) [21] for a narrow class of fractional white noise processes and in Jensen (2000) [22] 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 in the abovementioned article of Fox and Taqqu (1986) [12], also in Dahlhaus (1989) [10] and in Giraitis and Surgailis (1990) [16]. The next logical step would be to relax parametric assumptions on the behavior of the spectral density estimation and only assume that it has a pole at $\lambda = 0$:

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

where $g(\lambda) \to c$ for some positive finite constants c and α_0. 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) [13] and the local Whittle estimator of Künsch (1987) [23] should be mentioned. They were explored in great detail by Robinson (1995a) [33], and Robinson (1995b) [34]. Closely related are the broad-band estimators of Moulines and Soulier (1999) [30] and of Hurvich and Brodsky (2001) [19], as well as the exact local Whittle estimation method of Shimotsu and Phillips (2005) [35].

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, Moulines and Soulier (2001) [20] and in Arteche (2004) [3]. 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, Giraitis and Hidalgo (2004) [11]. 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, Distaso and Giraitis (2006) [2] obtain asymptotic confidence intervals for the trend and memory parameters in the case of trending long-memory processes 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, Distaso and Giraitis (2005) [1].

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 the underlying fBm based solely on discrete observations of this fBm's increments, and we explain precisely how to implement this conditional MLE in practice. While the proof that the resulting estimator for H is consistent will be tackled in a future publication, herein we prove that the appropriately scaled accumulation of the centered squared LARCH observations $r_t^2 - \mathbf{E}r_t^2$ converge in $L^2(\Omega)$ to the fBm.

We also attempt to use the local Whittle method to estimate the H of the LARCH process, with the same long-term objective of using it with fBm observations instead. Since σ_t^2 is unobservable, we apply the method to squared assets process r_t^2. However, consistency of this method is not entirely clear. As is usual for local Whittle method (see, for example, Dalla, Giraitis and Hidalgo (2004) [11]), 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) [25]. 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 is 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, Robinson and Surgailis (2000) [14]. 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 Section 2, we present the LARCH(∞) model, and show in Proposition 1 that $n^{-H}\sum_{i=0}^{nt}(\sigma_i - a)$ converges in $L^2(\Omega)$ to a constant multiple of an underlying fBm $B^H(t)$.

In Section 3, we prove that the 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 is given by the solution of the system of two equations $\frac{\partial \log L}{\partial a} = 0$ and $\frac{\partial \log L}{\partial H} = 0$, and that this estimator is consistent. 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; we present numerical results based on simulated data which show that the method preforms very well.

In Section 4, we justify a *portmanteau* test for model validation by proving (Theorem 1) that the vector of standardized residual autocorrelations $\{\hat{\gamma}_l : l = 1, \cdots, n\}$ defined in (19), based on the residuals $\hat{\varepsilon}_i = \varepsilon_i\left(\hat{a}_i, \hat{H}_i\right)$ where $\varepsilon_i(a, H)$ is obtained from H, a, and the observations via the model (1), converges in distribution to a centered normal with covariance matrix I^-, where this I^- is the generalized inverse of the matrix I defined in (16) and calculated in (17); the corresponding asymptotic Chi-squared distribution is also given explicitly in Theorem 1.

In Section 5, 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 4).

In Section 6, we draw the connection between our condition MLE and the estimation of H from discrete observations of the underlying fBm. In Subsection 6.1, 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 Subsection 6.2, after showing convergence to fBm in $L^2(\Omega)$ 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 Section 7, 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 8 is an Appendix where a crucial technical estimate is proved.

2 The LARCH Model; first convergence to fBm.

As in Giraitis, Robinson, and Surgailis (2000) [14], 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.

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

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

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, Robinson, and Surgailis (2000) [14] is denoted by $\theta = 2-2H$.

With this choice of b_j we proceed to proving a strong L^2 convergence result to fBm by assuming that the ε_j's are constructed on the same space as W. 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 much more involved, and beyond the scope of this article. To be more clear, the convergence in the next proposition can be established in the sense of distributions if one only assumes that the independent noise terms ε_j are identically distributed with mean 0 and variance 1, by a straightforward application of Donsker's invariance principle for Brownian motion (see [24, Theorem 2.4.20]). But this does not help us because we need our convergences to occur within the same probability space in the $L^2(\Omega)$ sense, in order to justify our ultimate goal of lifting parameter estimation techniques for discrete time series to continuous-time processes.

The easiest way to couple the ε_j's and 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,

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

Donsker's invariance principle proves that, as a random element in the space of continuous functions, $W^{(n)}$ converges in distribution to a Brownian motion W. But since our noise terms are Gaussian, more is true. The stochastic process $W^{(n)}$ is Gaussian. By the convergence in distribution, we trivially get that for each fixed t, $W^{(n)}(t)$ is a Cauchy sequence in the underlying probability space $L^2(\Omega)$ which is a Banach space (it is of course a Hilbert space). Therefore, $W^{(n)}(t)$ converges in $L^2(\Omega)$ to a random variable $W(t)$. It is then immediate that W as a stochastic process is a Brownian motion, and we have thus constructed the limit of the renormalized (and interpolated) partial sums $W^{(n)}(t)$ on the same space $L^2(\Omega)$ where the ε_j's live.

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 given 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 $L^2(\Omega)$ 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 in $L^2(\Omega)$, $\lim_{n\to\infty} V(\cdot)$ has a continuous modification which coincides with aB^H.

Proof. We have

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

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 $\{\varepsilon_i, \varepsilon_{i+1,}\varepsilon_{i+2}, \cdots\}$, we add and subtract the term $\tau_j = \mathbf{E}\left[\sigma_j|\mathcal{F}_{j-J}^+\right]$ where J is fixed. Hence

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

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

$$
V_2(k/n) = \sum_{i=1}^{k}\frac{1}{n}\sum_{j=-\infty}^{i-1}\left(\frac{i-j}{n}\right)^{H-3/2}\tau_j\varepsilon_j n^{-1/2} \tag{9}
$$

is an approximation of an iterated Riemann and Itô integral. Specifically, let W be the Brownian motion that interpolates the independent noise terms ε_j as the limit of the process $W^{(n)}$ defined in (8). We have the following lemma, proved in the appendix.

Lemma 1 $\lim_{n\to\infty}\mathbf{E}\left[\left(a\int_{s=0}^{t}ds\int_{-\infty}^{s}(s-r)^{H-3/2}dW(r) - V_2(k/n)\right)\right] = 0.$

This lemma allows us to say that $V_2(k/n)$ converges in $L^2(\Omega)$ to an fBm because the process v defined by the limit in this lemma, i.e.

$$
v(t) := a\int_{s=0}^{t}ds\int_{-\infty}^{s}(s-r)^{H-3/2}dW(s). \tag{10}
$$

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

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

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

It remains to show that

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

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 $j\neq j'$, $\mathbf{E}\left[(\sigma_j-\tau_j)\,\varepsilon_j\,(\sigma_{j'}-\tau_{j'})\,\varepsilon_{j'}\right]=0$, and

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

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

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

The limit of the above triple sum is the Riemann integral $\int_0^t\int_0^t\left(\int_{-\infty}^{\min(s,s')}(s-r)^{H-3/2}(s'-r)^{H-3/2}\,dr\right)dsds'$, 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 $V\left(\frac{k}{n}\right)$ convergence in $L(\Omega)$ to the fractional Brownian motion $v(t)$ given in (10).

We have finished proving that with $k=[tn]$, $V(k/n)$ converges in $L^2(\Omega)$ to $aB^H(t)$ 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. The proof of the last statement of proposition now follows from Remark 3 below; details are omitted. ■

Remark 1 *It is possible to go one step further in showing that V converges to aB^H as a process in $L^2(\Omega\times[0,T])$ and even in $L^2(\Omega;C([0,T]))$. To do this, one may prove the convergence to zero of the quantity $\mathbf{E}\left[\max_{t\in[0,T]}\left|V([t])-aB^H(t)\right|^2\right]$. This requires more involved arguments than the ones given above,* *including use of the regularity theory of stochastic processes. It is beyond the scope of this article. Note that such a result is much stronger than simply saying that V converges to fBm in distribution.*

Remark 2 *It is crucial to note that our $L^2(\Omega)$-convergence result of Proposition 1 implies the convergence in $L^2(\Omega)$ of any finite-dimensional vectors of values of V to the corresponding vector of values of aB^H. To then prove convergence at the process level, it is sufficient to show that the limit of V is almost surely continuous. Such a task is as technical as the one alluded to in the previous remark, and is also omitted.*

Remark 3 *However, using two-dimensional distributions, it is not difficult to invoke the Kolmogorov continuity criterion (see [24, Theorem 2.2.8]) to prove that the limit of V is an almost-surely continuous process. Indeed, the proof of convergence of $V(t)$ to $aB^H(t)$ in $L^4(\Omega)$ can be obtained using similar tools as the one we give above in $L^2(\Omega)$; then the estimates required of the limit of V to use the Kolmogorov criterion follow from the same estimates for B^H, namely $\mathbf{E}\left[\left(aB^H(t) - aB^H(s)\right)^4\right] \leq 3a^4 |t-s|^{4H}$ (in fact, equality holds). Having a process whose limit has a continuous modification equal to B^H is satisfactory for our statistical purposes.*

In this following sections, 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 6.2 draws a connection between the discrete time series and fBm by showing that the LARCH observations r_t are themselves asymptotically a set of discrete observations for a true fBm, and that a conditional maximum likelihood estimator of the Hurst parameter H for these observations is a good candidate for a consistent estimator of H for the fBm itself. Section 6.2 also explains what not to do when trying to lift estimation of H from the LARCH observations to the fBm observations.

3 Conditional MLE in the ARCH(∞) model

In the discrete-time model, our observations are the log returns r_j. Therefore we seek the conditional maximum likelihood estimator of the two parameters a and H for the ARCH(∞) model. At time i, that is, given past the observations $r_j : j = 1, 2, \cdots, i$, this conditional MLE is defined as the value of the pair $\left(\hat{a}, \hat{H}\right)$ which maximizes the conditional log-likelihood function $\log L(a, H)$ defined via

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

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

$$\log L(a,H) = -\frac{1}{2}\left(i\log(2\pi) + \sum_{j=1}^{i}\log\sigma_j^2 + \sum_{j=1}^{i}\frac{r_j^2}{\sigma_j^2}\right) \tag{11}$$

where

$$\sigma_i = a + \sum_{j=1}^{\infty} j^{H-3/2} r_{i-j}. \tag{12}$$

We easily calculate

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

so that

$$\frac{\partial\log L}{\partial a} = \sum_{j=1}^{i}\left(-\frac{1}{\sigma_j} + \frac{r_j^2}{\sigma_j^3}\right) \tag{13}$$

$$\frac{\partial\log L}{\partial H} = \sum_{j=1}^{i}\left(-\frac{1}{\sigma_j} + \frac{r_j^2}{\sigma_j^3}\right)\sum_{k=1}^{\infty} k^{H-3/2} r_{j-k}\log k \tag{14}$$

Therefore $\left(\hat{a},\hat{H}\right) = \left(\hat{a},\hat{H}\right)_i$ is the solution to the system of two equations $\frac{\partial\log L}{\partial a} = 0$ 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 (12) and each $r_j : j \leq i$ is known.

Under standard regularity conditions, this estimator is consistent, as i tends to infinity, even if the underlying distribution is not conditionally normal. In our conditionally lognormal case, the conditions are satisfied. For instance, Christian Gourieroux's textbook Gourieroux (1997) [17] can be consulted for this fact.

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 figures 1 and 2 at the end of this article. Despite the apparent algebraic complexity of the equations (13) and (13) one needs to solve to obtain $\left(\hat{a},\hat{H}\right)$, 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.

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 $i = -P, = P+1, \cdots, -2, -1$. 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	$\hat{a}$	$\hat{H}$
1000	50	0.1	0.5	0.09999	0.4960
1000	100	1	0.7	0.9996	0.6998
1000	500	0.8	0.8	0.7998	0.7999
1000	500	0.9	0.6	0.8956	0.5966
1000	500	0.2	0.9	0.1977	0.9005
1000	1000	0.8	0.7	0.8000	0.7000

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: $n = 1000$ is a reasonable number of data points for the precision attained above.

4 Model validation and hypothesis testing

In addition to having the convergence in law of $\left(\hat{a}, \hat{H}\right)_n$ to the true parameters (a, H) as the number of observations n tends to infinity, one is interested in the asymptotic distribution of $\left(\hat{a}, \hat{H}\right)$ so that one may draw some conclusions regarding hypothesis testing, such as testing whether or not $H = 1/2$, which can then serve as a way to validate or invalidate the long-memory time series model for the given time series data.

Such an asymptotic theory was proposed for parametric ARCH models by Weiss (1986) [38], Lee and Hansen (1994) [26], Ling and Li (1997) [27], Ling and McAleer (1996) [28], Lumsdaine (1996) [29], and Tjostheim (1986) [37], among others. In Weiss (1986) [38], the author studies the properties of the quasi-maximum likelihood estimator (QMLE) and related test statistics in dynamic models that jointly parametrize conditional means and conditional covariances when a normal log-likelihood is maximized but the assumption of normality is violated. It is shown that QMLE is still consistent. For consistency and asymptotic normality of the QMLE estimator in the GARCH$(1,1)$ models, see Lumsdaine (1996) [29].

The general theory of conditional MLE also implies, under appropriate regularity conditions which are satisfied in our conditionally normal setting that our estimator $\hat{\theta}_n := \left(\hat{a}, \hat{H}\right)_n$ is asymptotically normal, i.e. that the following convergence to a normal distribution holds

$$\sqrt{T}(\hat{\theta}_n - (a, H)) \to N(0, J^{-1} I J^{-1}) \tag{15}$$

as n tends to infinity, where the 2×2 matrices I and J are defined by

$$\begin{aligned} I &= \mathbf{E}_0 \left[-\frac{\partial^2 \log l_j(r;\theta)}{\partial\theta\partial\theta'} \right], \\ J &= \mathbf{E}_0 \left[\frac{\partial \log l_j(r;\theta)}{\partial\theta} \frac{\partial \log l_j(r;\theta)}{\partial\theta'} \right] \end{aligned} \tag{16}$$

and we have used the standard likelihood notation

$$\log l_j(r;\theta) := \log f\left(r_j | r_{j-k} : k = 1, 2, \cdots\right)$$

for the conditional log density of the model r at time j given past observations, and the column vector $\partial/\partial\theta$ denotes the gradient operator w.r.t the components of θ (in our case a two-dimensional column vector), while $\partial/\partial\theta'$ is the corresponding row vector, so that I and J are (in our case 2×2) matrices.

We propose a *portmanteau* statistic for testing the adequacy of fitted time series models. The study of the distribution of residual autocorrelations in linear times series started with the work [6] of Box and Pierce (1970). They showed that a portmanteau test, called Box-Pierce statistics, which is essentially a sum of squared residual autocorrelation coefficients, follows a chi-square distribution.

The required assumption for this result will be satisfied in our case of IID standard normal ε's:

Assumption (A1) ε_j is a sequence of independent and identically distributed (i.i.d.) random variables with mean zero and variance 1; and $\mathbf{E}(\varepsilon_j^4) < \infty$.

Assumption (A2) J and I are positive definite.

Given the observations $r_1, \ldots, r_n$ and the initial value $r_0 = (r_0, r_{-1}, \ldots)$, using our new likelihood notation, recall that the conditional log-likelihood function in (11) can be written as:

$$\mathcal{L}_n(\theta) := \log L\left(a, H\right) = \sum_{i=1}^{n} l_i(\theta)$$

where

$$l_i(\theta) = -\frac{1}{2}\log(2\pi) - \frac{1}{2}\log(\sigma_i^2) - \frac{r_i^2}{2\sigma_i^2}$$

Again, from Section 3, the maximizer of $\mathcal{L}_n(\theta)$ on the parameter space Θ, denoted by $\hat{\theta}_n = (\hat{a}_n, \hat{H}_n)$ is the conditional- maximum likelihood estimator of θ. The algebraic derivation of the formulas immediately following is elementary, and is omitted. We have included, among the following quantities, some values (H (not to be confused with the Hurst parameter), R, P) which are not strictly necessary to define in order to understand the asymptotic distribution of the residual autocorrelation function (Theorem 1 below), but they do enter in the calculations which lead to the formulas for the matrices I and J, in ways that should be familiar. For instance, R below turns out to be 0, and P, the term involving second derivatives w.r.t. θ, is therefore useful in the expression of the asymptotic normality. All further details are omitted.

Denote $\varepsilon_i(\theta) = \frac{r_i}{\sigma_i(\theta)}$, which, as such, depends only on θ and the observations. Denote $D_i(\theta) = \frac{\partial l_i(\theta)}{\partial\theta} =$

$-U_i(\theta)\xi_i(\theta)$, and $P_i(\theta) = \frac{\partial^2 l_i(\theta)}{\partial\theta\partial\theta'} = -U_i(\theta)U_i'(\theta) - R_i(\theta)$, where

$$\xi_i(\theta) = \left(\varepsilon_i(\theta), \frac{1}{\sqrt{2}}\left[1 - \varepsilon_i^2(\theta)\right]\right)'$$
$$U_i(\theta) = \left[\frac{1}{\sigma_i(\theta)}\frac{\partial\varepsilon_i(\theta)}{\partial\theta}, \frac{1}{\sqrt{2}\sigma_i^2(\theta)}\frac{\partial\sigma_i^2(\theta)}{\partial\theta}\right]$$
$$R_i(\theta) = \frac{1}{\sigma_i^2(\theta)}\left[\frac{\partial^2\sigma_i^2(\theta)}{\partial\theta\partial\theta'} - \frac{1}{\sigma_i^2(\theta)}\frac{\partial\sigma_i^2(\theta)}{\partial\theta}\frac{\partial\sigma_i^2(\theta)}{\partial\theta'}\right]\left[1 - \frac{r_i^2(\theta)}{\sigma_i^2(\theta)}\right]$$

We have

$$\begin{aligned} H &= \mathbf{E}\left[\xi_i(\theta_0)\xi_i'(\theta_0)\right], \\ J &= \mathbf{E}\left[U_i(\theta_0)HU_i'(\theta_0)\right], \\ I &= \mathbf{E}\left[U_i(\theta_0)U_i'(\theta_0)\right], \end{aligned} \tag{17}$$

where I and J are defined in (16). The next fact is sometimes taken as an assumption for general studies; in our case, it does hold, with the matrix I and the vectors D_i defined above.

Fact (A3) The conditional maximum likelihood estimator $\hat{\theta}_n$, which is consistent, has the expansion

$$\sqrt{n}(\hat{\theta}_n - \theta) = \frac{I^{-1}}{\sqrt{n}}\sum_{i=1}^{n} D_i(\theta) + o_p(1) \tag{18}$$

where we use the standard econometrics time series notation $o_p(1)$ for a random variable that tends to 0 in probability as n tends to ∞. Indeed, a Taylor expansion and the fact that $D_t(\hat{\theta}_n) = 0$, show that $-D_t(\theta_0) \approx (\hat{\theta_n} - \theta_0)P_t(\theta_0)$. Taking expectations yields (18).

Finally, using the next assumption, easily satisfied in our conditionally normal case, we can consistently estimate H, J, I using $\hat{\theta}_n$.

Assumption (A4) $\mathbf{E}\left[\max_{\theta\in\Theta}\varepsilon_t^4(\theta)\right] < \infty$, and $\mathbf{E}\left[\max_{\theta\in\Theta}\|U_t(\theta)\|^2\right] < \infty$.

Using the *residual*

$$\hat{\varepsilon}_i := \varepsilon_i(\hat{\theta}_n)$$

the log $-l$ (standardized) residual autocorrelation function can be defined as:

$$\hat{\gamma}_l = \frac{\sum_{i=l+1}^{n}\hat{\varepsilon}_i\hat{\varepsilon}_{i+l}}{\sqrt{\sum_{i=l+1}^{n}\hat{\varepsilon}_{i-l}^2\sum_{i=1}^{n}\hat{\varepsilon}_i^2}}. \tag{19}$$

By Theorem (3.1) in [28] we have that:

$$\frac{1}{n}\sum_{i=1}^{n}\hat{\varepsilon}_i^2 = 1 + o_p(1).$$

From the facts A1 through A4, we have that $o_p(1)$ in the above expression does indeed converge to zero in probability. Therefore we only need to consider the asymptotic distribution of

$$\hat{\rho}_l = \frac{1}{n} \sum_{i=l+1}^{n} \hat{\varepsilon}_i \hat{\varepsilon}_{i+l}. \tag{20}$$

In [27] the authors proved that:

$$\sqrt{n}\,(\hat{\rho_1}, \dots, \hat{\rho_M})' = \sqrt{n}\,(\rho_1, \dots, \rho_M)' - D\sqrt{n}(\hat{\theta}_n - \theta_0) + o_p(1)$$

where,

$$\rho_l = \frac{1}{n} \sum_{i=l+1}^{n} \varepsilon_i \varepsilon_{i+l}, \tag{21}$$

so that

$$(\rho_1, \dots, \rho_M) = \frac{1}{n} \sum_{i=1}^{n} (\varepsilon_i \varepsilon_{i-1}, \dots, \varepsilon_i \varepsilon_{i-M}) =: \frac{1}{n} \sum_{i=1}^{n} \tau_i, \tag{22}$$

and where

$$D = (D_{p,q})_{p,q} = \left(\frac{\partial \rho_p}{\partial \theta_q} \right)_{p,q}, \tag{23}$$

where now $\partial/\partial\theta_q$ is a standard partial derivative w.r.t. the qth component of $\theta = (a, H)$, $q = 1, 2$, so that the matrix D is an $M \times 2$ matrix in our case.

Lemma 2 *Let $(\rho_1, \dots, \rho_M)$ be the random vector where ρ_l is defined as in (21), then $\sqrt{n}(\rho_1, \dots, \rho_M)$ is asymptotically normally distributed with zero mean and covariance matrix $\Sigma = \Sigma_\rho = I\ \mathbf{E}\tau_t\tau_t'$,i.e. defined by*

$$\Sigma_{p,q} = I\ \mathbf{E}\left(\tau_i\left(p\right)\tau_i\left(q\right)\right) = I\ \mathbf{E}\left[\varepsilon_{i-p}\varepsilon_{i-q}\right]$$

where $\tau_t = (\varepsilon_i\varepsilon_{i-1}, \dots, \varepsilon_i\varepsilon_{i-M})$ was introduced in (22).

Proof Let $N_i = (\varepsilon_i\varepsilon_{i-1}, \dots, \varepsilon_i\varepsilon_{i-M})$. Then the sequence $(N_i)_{i\geq 0}$ is a martingale-difference array. Furthermore, if we define $\nu_i = I\ \mathbf{E}\left(N_i^2/\mathcal{F}_{i-1}\right)$ then ν_i does not depend on i. Finally by using the Central limit theorem for martingale differences the results follows. ■

Lemma 3 *The asymptotic variance of $\sqrt{n}\,(\hat{\rho}_1, \dots, \hat{\rho}_M)$ is given by*

$$\Sigma_\rho + DJ^{-1}IJ^{-1^t} + 2I^{-1}Cov\left[(\rho_1, \dots, \rho_M), \sum_{t=1}^{T} D_t(\theta_0)\right] D^t.$$

Proof: By using **A3** and equation (15) we obtain

$$\begin{aligned}
Var\left(\sqrt{n}\left(\hat{\rho}_1,\ldots,\hat{\rho}_M\right)\right) &= Var\left(\sqrt{n}\left(\rho_1,\ldots,\rho_M\right)\right) + D_{M\times 2}Var(\sqrt{n}(\hat{\theta}-\theta))D^t \\
&\quad - 2Cov\left(\sqrt{n}\left(\rho_1,\ldots,\rho_M\right),\sqrt{n}(\hat{\theta}-\theta)\right)\right) D^t \\
&= \Sigma_\rho + DJ^{-1}IJ^{-1^t}D^t \\
&\quad - 2I^{-1}Cov\left(\left(\rho_1,\ldots,\rho_M\right),\sum_{i=1}^{n}U_i\xi_i\right)D^t \\
&= \Sigma_\rho + DJ^{-1}IJ^{-1^t}D^t \\
&\quad - 2I^{-1}Cov\left(\left(\rho_1,\ldots,\rho_M\right),\sum_{i=1}^{n}D_i(\theta)\right)D^t.
\end{aligned}$$

Then the result follows. ■

Theorem 1 *As n goes to infinity, $\sqrt{n}\left(\hat{\rho}_1,\ldots,\hat{\rho}_M\right)$ converges to the normal $N(0,\Sigma_{\hat{\rho}})$ in distribution. Furthermore, $n\hat{\rho}^t\Sigma_{\hat{\rho}}^{-}\hat{\rho}$ converges to χ_r^2, the Chi-squared distribution with r degrees of freedom, where r is the rank of $\Sigma_{\hat{\rho}}$ and $\Sigma_{\hat{\rho}}^{-}$ is the generalized inverse of $\Sigma_{\hat{\rho}}$, the latter being defined in Lemma 2.*

In our preferred setting where the ε_j's are independent with mean 0 and variance 1, in the notation of Lemma 2, $\mathbf{E}\left[\varepsilon_{i-p}\varepsilon_{i-q}\right]=\delta_{p,q}$, and therefore $\Sigma_{\hat{\rho}}=I$.

Proof: It follows from Lemmas (2) and (3). ■

5 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 $\bar{r}_k := (r_1,\cdots,r_k)$.Because there is no analytical way of solving the system of two equations yielding $\left(\hat{a},\hat{H}\right)$, we must invoke the mean-value theorem assisted by the implicit function theorem in order to evaluate the error

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

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

$$\frac{\partial X}{\partial r_j}=\frac{\partial F}{\partial r_j}/\frac{\partial F}{\partial X}+\frac{\partial G}{\partial r_j}/\frac{\partial G}{\partial X};$$
$$\frac{\partial Y}{\partial r_j}=\frac{\partial F}{\partial r_j}/\frac{\partial F}{\partial Y}+\frac{\partial G}{\partial r_j}/\frac{\partial G}{\partial Y}.$$

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 (13) and (14), that is:

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

with the understanding that each σ_j is a function of $r_{j-1},r_{j-2},\cdots,r_1$ 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

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

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

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

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

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

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:

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

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 tildas and hats for simplicity of notation, we can express

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

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

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

The random variable $\sum_{i=1,\cdots,k;i\neq i'}\left(\frac{3r_i^2}{\sigma_i^4}+\frac{1}{\sigma_i^2}\right)/\left(\frac{3r_{i'}^2}{\sigma_{i'}^4}+\frac{1}{\sigma_{i'}^2}\right)$ 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,

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

The term f_2 is much smaller than f_1, as the inequality $2r_i/\sigma_i^3\le r_i^2/\sigma_i^4+1/\sigma_i^2$ 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 σ_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 $\int_0^t (t-s)^{H-1/2} dB(s)$, implying that e_k converges rapidly to 0. More specifically we claim the following strong robustness.

Remark 4 *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.

6 Connection with Hurst parameter estimation for fBm

The connection between ARCH models and fBm has in the past been phrased as a convergence in distribution of normalized partial sums. However, our calculations in Section 2 show that there really is an underlying fBm on the same probability space as the LARCH(∞) model (σ_j, r_j). Indeed, a common standard Brownian motion W was used to define both the fBm B^H, via its 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). When n is fixed, it is possible to argue that the ε_j's can be calculated from W itself (instead of the other way around as given in (8)) via the formula $\varepsilon_j = \sqrt{n}\left[W\left((j+1)/n\right) - W\left(j/n\right)\right]$. We will not make use of this representation explicitly, however; moreover, it may not be used to prove convergence theorems as n tends to infinity.

Given these pathwise connections, one may wonder whether we can use the estimators defined in Section 3 to claim that we are in fact estimating the Hurst parameter of a fBm. If the underlying fBm is not observed, but the r_j's of the ARCH(∞) model are, then answer is of course affirmative, since B^H shares the same H as $(r_j)_j$. But this is of little comfort to a practitioner oberving the fBm itself. The correct question is thus the following: if the trajectory of B^H is observed – possibly at discrete times – may one devise a conditional MLE for H similar to the one in Section 3 based on the ARCH(∞) observations r_j, but using instead true observations from B^H, for instance $B^H(j/n) : j = 1, 2, \cdots$?

This question is similar to that of robustness, since the information contained in $B^H(j/n) : j = 1, 2, \cdots$ will approximate that contained in $(r_j)_j$, but the question is not as simple, since there is no simple formula relating $B^H(j/n)$ and r_j via a "small" error term h_j. Our first task is thus to find a transformation $f(r_j)$ of the observations r_j for which the true fBm increments $B_j := B^H((j+1)/n) - B^H(j/n)$, are an approximation to the $f(r_j)$'s.

6.1 Some negative results

6.1.1 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 (10) of the moving average representation: for $k = k_t = tn$,

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

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

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

This is to be compared with the decomposition

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

because $V(k/n)$ converges to $B^H(t)$ in $L^2(\Omega)$.

Therefore, it is apparent that the analogue, in the continuous-time fBm model, of the observations r_j, are the IID terms $a\sqrt{n}\left\{W(j/n) - W(((j-1)/n))\right\} = a\varepsilon_{j-1}$. 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. They are not, however, independent, as can be seen from the behavior of their second moments.

6.1.2 Volatility observation

One may then be tempted to devise a Hurst parameter estimation method for fBm based on Proposition 1, i.e. using the "volatilities" σ_j as observations, 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

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

which is obviously degenerate, yielding infinitely many solutions $\hat{a} = \pm r_i - \sum_{j=1}^{\infty} j^{\hat{H}-3/2} r_{i-k}$.

6.2 Hurst parameter estimation for fBm based on LARCH observations

6.2.1 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, Robinson and Surgailis (2000) [14], but only as a convergence in law, which is why it could not be considered as the basis for an estimation of H for fBm.

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

Proof. Given our L^2 convergence established in Proposition 1, this proposition is largely a consequence of calculations performed in Giraitis, Robinson and Surgailis (2000) [14]. Indeed, 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, which implies

$$\begin{aligned}
&\mathbf{E}\left[\left(n^{-H}\sum_{i=1}^{k}\left(|r_i|^2-\mathbf{E}\,|r_i|^2\right)-\frac{2a^2}{1-\|b\|^2}B^H(t)\right)^2\right]\\
&=\mathbf{E}\left[\left(n^{-H}\sum_{i=1}^{k}(\nu_i-\mathbf{E}\nu_i)-+n^{-H}\sum_{i=1}^{k}\left(|\sigma_i|^2-\mathbf{E}\,|\sigma_i|^2\right)-\frac{2a^2}{1-\|b\|^2}B^H(t)\right)^2\right]\\
&\le 2n^{-2H}\sum_{i=1}^{k}\mathbf{E}\left[|\nu_i|^4\right]+2\mathbf{E}\left[\left(\frac{2a^2}{1-\|b\|^2}B^H(t)-n^{-H}\sum_{i=1}^{k}\left(|\sigma_i|^2-\mathbf{E}\,|\sigma_i|^2\right)\right)^2\right]\\
&=2n^{-(2H-1)}\mathbf{E}\left[|\nu_0-\mathbf{E}\nu_0|^4\right]+2\mathbf{E}\left[\left(\frac{2a^2}{1-\|b\|^2}B^H(t)-n^{-H}\sum_{i=1}^{k}\left(|\sigma_i|^2-\mathbf{E}\,|\sigma_i|^2\right)\right)^2\right].
\end{aligned}$$

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 0 in $L^2(\Omega)$.

It was established in Giraitis, Robinson and Surgailis (2000) [14, 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

$$\begin{aligned}
&\mathbf{E}\left[\left(\frac{2a^2}{1-\|b\|^2}B^H(t)-n^{-H}\sum_{i=1}^{k}\left(|r_i|^2-\mathbf{E}\,|r_i|^2\right)\right)^2\right]\\
&=o(1)+2\mathbf{E}\left[\left|n^{-H}\sum_{i=1}^{k}(y_{i\ell}-\mathbf{E}[y_{i\ell}])\right|^2\right]\\
&\quad+4\mathbf{E}\left[\left(\frac{2a^2}{1-\|b\|^2}B^H(t)-n^{-H}\sum_{i=1}^{k}\left(\frac{2a}{1-\|b\|^2}\sigma_i-\frac{2a^2}{1-\|b\|^2}-\right)\right)^2\right]\\
&=o(1)+\frac{8a}{1-\|b\|^2}\mathbf{E}\left[\left(aB^H(t)-n^{-H}\sum_{i=1}^{k}(\sigma_i-a)\right)^2\right].
\end{aligned}$$

The conclusion of the proposition now follows immediately from our Proposition 1. ■

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 the fractional Brownian increment $2\frac{a^2}{1-\|b\|^2}\left(B^H((i+1)/n)-B^H(i/n)\right)=2\frac{a^2}{1-\|b\|^2}B_i$. Because of this, it is natural to propose a conditional MLE for estimating a and H based on the observations

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

and use it with the observations $2a^2\left(1-\|b\|^2\right)^{-1}B_i$ instead, claiming that the resulting scheme is a consistent estimator for the parameters (a,H) of the stochastic process $2a^2\left(1-\|b\|^2\right)^{-1}B^H$ observed at discrete times. We will not prove this consistency, since it would be mathematically significantly more involved than the proof of the robustness proposition 3. Note also that we consider here the observation of a scaled fBm, with scaling parameter $2a^2\left(1-\|b\|^2\right)^{-1}$ which will be determined once a and H are estimated.

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

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

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 $x_0, x_1, \cdots, x_{i-1}$ can thus be represented as

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

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

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

We have the following partial derivatives:

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

and

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

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

Therefore, the maximum likelihood estimator $\left(\hat{a}, \hat{H}\right)$ is obtained as the solution of the 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.

6.2.2 Practical implementation

In practice, we use only a finite memory horizon P instead of ∞, as we did in the conditional MLE of Section 3 (see the description of the table of results on page 12). Thence the formulas (29), (30), and (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 Section 5 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 Section 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 the same fractional Brownian motion B^H as for $k = 1$, up to a scaling factor:

$$\lim_{n\to\infty} \mathbf{E}\left[\left(\sum_{i=1}^{k} x_i^{(p)} - 2pg_{2p}a^{-1}\mathbf{E}\left[|\sigma_0|^{2p}\right]B_H(t)\right)^2\right] = 0.$$

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 above convergence holds for moments of all integer orders q, it yields convergence to 0 when $g_q = 0$.

7 A local Whittle-type estimator of the Hurst parameter

7.1 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, Robinson and Surgailis (2000) [14]) that

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

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, Giraitis and Hidalgo (2004) [11]. 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

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

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,

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

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) [7]) 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

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

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 $\sum_{|j| \geq m_n} W_n^2(j) \to 0$ as $n \to \infty$. Again, for details see Brockwell and Davis (2002) [7]).

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

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

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

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

of the quasi-likelihood-type objective function

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

7.2 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, Giraitis and Hidalgo (2004) [11].

- 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, Robinson, and Surgailis (2000) [14], we have

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

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, Giraitis and Hidalgo (2004) [11] 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, Giraitis and Hidalgo (2004) [11] 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, Giraitis and Hidalgo (2004) [11].

By definition,

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

Therefore,

$$\eta_j^* - \mathbf{E}\eta_j^* = \frac{1}{n b_0 \lambda_j^{-\alpha_0}}\left[\sum_{t=1}^{n}(r_t^4 - \mathbf{E}\,r_t^4) + 2\sum_{t=1}^{n}\sum_{h=1}^{n-t}(r_t^2 r_{t+h}^2 - \mathbf{E}\,r_t^2 r_{t+h}^2)\cos(\lambda_j h)\right]$$

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

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

Giraitis, Leipus, Robinson, and Surgailis (2004) [15] 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

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

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, Robinson and Surgailis (2000) [14]). 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

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

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

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

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 $Cov\,(r_0^j, r_h^j) \sim c_j^2 h^{-\theta}$; in other words, to investigate asymptotic behavior (as $t \to \infty$) of "mixed moments" of the form $Cov\,(r_0^j r_t^j, r_h^j r_{t+l}^j)$ for positive integer $j > 2$ and integer $h, l > 0$. Indeed, one can prove by elementary calculations that any failiure 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.

8 Appendix: Proof of Lemma 1.

Let $\tau(r)$ be the limit of τ_j in $L^2(\Omega)$, where $j = [sn]$. We show first that $\tau(r)$ exists and equals the constant a. Indeed, as σ_j is independent of all noise terms $\varepsilon_{j'}$ for $j' \geq j$, 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 $\varepsilon_{j-J}, \varepsilon_{j-J+1}, \cdots, \varepsilon_{j-1}$. 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 $\mathcal{F}_{\{j-J,j-1\}}$:

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

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

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

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

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

$$= \frac{h(n)}{n^3}\sum_{j=-\infty}^{k-1}\left(\sum_{i=j+1}^{k}\left(\frac{i-j}{n}\right)^{H-3/2}\right)^2.$$

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

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

Therefore we have proved that the quantity in (41) converges to 0 as n tends to $+\infty$, which means that in the definition (9) 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:

$$\lim_{n\to\infty}\mathbf{E}\left[\left(\sum_{i=1}^{k}\frac{1}{n}\sum_{j=-\infty}^{i-1}\left(\frac{i-j}{n}\right)^{H-3/2}\varepsilon_j n^{-1/2} - \int_{s=0}^{t} ds\int_{-\infty}^{s}(s-r)^{H-3/2}\,dW(r)\right)\right] = 0 \tag{42}$$

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 $\{j/n : j = -\infty, \cdots, i-1\}$ of $[-\infty, i = [ns]/n]$, in which the only relevant values of increments of W are for these partition points; then one replaced W by its approximation

$W^{(n)}$, which is straightforward because of the evaluation at partition points only, further $L^2(\Omega)$-convergence being easily guaranteed by the convergence of $W^{(n)}$ to W as a Gaussian process. 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 Millenium 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.

References

[1] Abadir, K.M., Distaso, W., and Giraitis, L. (2005). Semiparametric Estimation and Inference for Trending I(d) and Related Processes. *The Annals of Statistics*, 2006 (to appear).

[2] Abadir, K.M., Distaso, W., and Giraitis, L. (2006). *Local Whittle estimation, fully extended for nonstationarity.* Preprint.

[3] Arteche, J. (2004). Gaussian semiparametric estimation in long memory in stochastic volatility and signal plus noise models. *Journal of Econometrics*, 119, 131-154.

[4] Boes, D.C., Davis, R.A., Gupta, S.N. (1989). Parameter estimation in low order fractionally differenced ARMA processes. *Stochastic Hydrology and Hydraulics*, 3, 97-110

[5] Bollerslev, Tim; Wooldridge, Jeffrey M. (1992). Quasi-maximum likelihood estimation and inference in dynamic models with time-varying covariances. *Econometric Rev.* **11**, no. 2, 143–172.

[6] Box, G.; Pierce, D.A. (1970). Distribution of Residual Autocorrelations in autoregressive-integrated moving-average Time Series Model, *J. Amer. Stat. Assoc.*, **64**, 1509–1526.

[7] Brockwell, P. and Davis, R. (2002) *Introduction to Time Series and Forecasting, 2nd Ed.* Springer.

[8] Bustos, Oscar H.; Yohai, Víctor J. (1986). *Robust estimates for ARMA models.* J. Amer. Statist. Assoc. **81**, no. 393, 155–168.

[9] Cheung, Y.W. (1993). Test for fractional integration: A Monte Carlo investigation. *Journal of Time Series Analysis*, **14**, 331-345.

[10] Dahlhaus, R. (1989). Efficient parameter estimation for self-similar processes. *Ann. Stat.*, **17** (4), 1749-1766.

[11] Dalla, V., Giraitis, L. and Hidalgo, J.(2004). Consistent estimation of the memory parameter for non-linear time series. *J. Time Series Analysis*, **27**, 211-251.

[12] Fox, R., Taqqu, M.S. (1986). Large sample properties of parameter estimates for strongly dependent stationary Gaussian time series. *The Annals of Statistics*, **14**, 517-532.

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

[14] Giraitis, L., Robinson, P.M., and Surgailis, D. (2000). A model for long memory conditional heteroskedasticity. *Ann. Appl. Prob.,* **10** (3), 1001-1024.

[15] Giraitis, L., Leipus, R., Robinson, P.M., and Surgailis, D. (2004). LARCH, leverage, and long memory. *Journal of Financial Econometrics*, **2** (2), 177-210.

[16] Giraitis, L. and Surgailis, D. (1990). A central limit theorem for quadratic forms in strongly dependent linear variables and application to asymptotical normality of Whittle's estimate. *Probability Theory and Related Fields*, **86**, 87-104.

[17] Gouriéroux, C. (1997). *ARCH Models and Financial Applications,* Springer Series in Statistics, Springer Verlag New York.

[18] Ghysels, E.; Harvey, A.C.; Renault, E. (1996). Stochastic volatility. *Statistical methods in finance*, 119–191, Handbook of Statist. **14**, North-Holland, 1996.

[19] Hurvich, C.M. and 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.

[20] Hurvich, C.M., Moulines, E. and Soulier, P. (2005) *Estimation of long memory in volatility, Econometrica*, **73**, 1283-1328.

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

[22] Jensen, M.J. (2000). An altenative maximum likelihood estimator of long-memory process using compactly supported wavelets. *Journal of Economic Dynamics and Control*, **24**, 361-387.

[23] Künsch, H. (1987). Statistical aspects of self-similar processes. In Yu.A. Prohorov and V.V. Sazonov (eds.), *Proceedings of the 1st World Congress of the Bernoulli Society*, Vol.**1**, Science Press, Utrecht, 67-74.

[24] Karatzas, I.; Shreve, S.E. *Brownian Motion and Stochastic Calculus, 2nd ed.* Springer, 1991.

[25] Lahiri, S.N. (2003). A necessary and sufficient condition for asymptotic independence of discrete Fourier Transforms under short- and long- range dependence. *The Annals of Statistics*, **31** (2), 613-641.

[26] Lee, Sang-Won; Hansen, Bruce E. (1994). Asymptotic theory for the $GARCH(1,1)$ quasi-maximum likelihood estimator. *Econometric Theory* **10**, no. 1, 29–52.

[27] Ling, S; Li, W.K. (1997). Diagnostic checking of nonlinear multivariate time series with multivariate ARCH errors. *Journal of Time Series Analysis,* **18** (5), 447–464.

[28] Ling, S; McAleer, M. (2003). Asymptotic theory for a new vector ARMA-GARCH model. *Econometric Theory* **19**, 280–310.

[29] Lumsdaine, R. L. (1996). Consistency and Asymptotic Normality of the Quasimaximum Likelihood Estimator in IGARCH(1,1) and Covariance Stationary GARCH(1,1) Models. *Econometrica*, **64**, 575-596.

[30] Moulines, E. and Soulier, P. (1999). Broad band log-periodogram regression of time series with long range dependence. *The Annals of Statistics*, **27**, 1415-1439.

[31] Nelson, D.B. (1990) ARCH models as diffusion approximations. *J. Econometrics* **45**, no. 1-2, 7–38.

[32] Robinson, P.M. (1991). Testing for strong serial correlation and dynamic conditional heteroskedasticity in multiple regression. *J. Econometrics*, **47**, 67-84

[33] Robinson, P.M. (1995a). Log-periodogram regression of toime series with long range dependence. *The Annals of Statistics*, **23**, 1048-1072.

[34] Robinson, P.M. (1995b). Gaussian semiparametric estimation of long range dependence. *The Annals of Statistics*, **23**, 1630-1661.

[35] Shimotsu, K. and Phillips, P.C.B. (2005). Exact local Whittle estimation of fractional integration. *Ann. Stat.*, **32**, 656-692.

[36] Sowell, F. (1992). Maximum likelihood estimation of stationary univariate fractionally integrated time series models. *Journal of Econometrics*, **53**, 165-188.

[37] Tjostheim, Dag. (1986). Estimation in nonlinear time series models. *Stochastic Process. Appl.* **21** (2), 251–273.

[38] Weiss, A.A. Asymptotic Theory for ARCH Models: Estimation and Testing. *Econometric Theory*, **2** (1986), pp. 107-131.

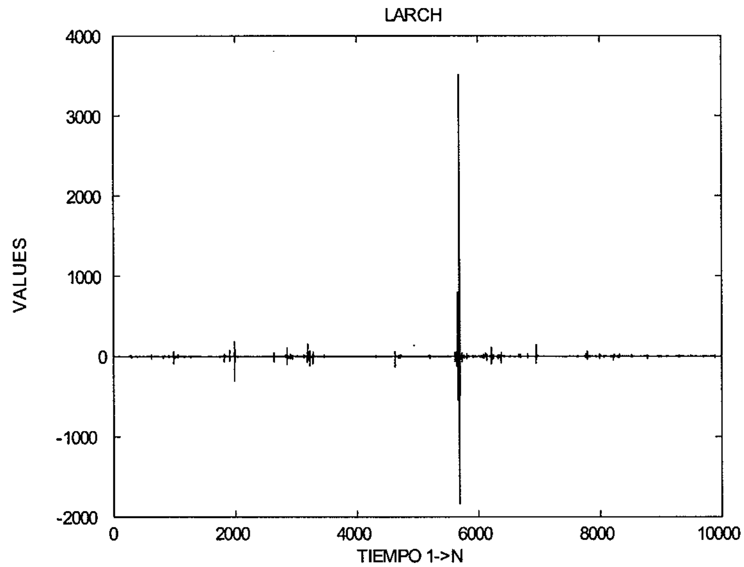

Fig. 1. Observation values for a typical LARCH time series

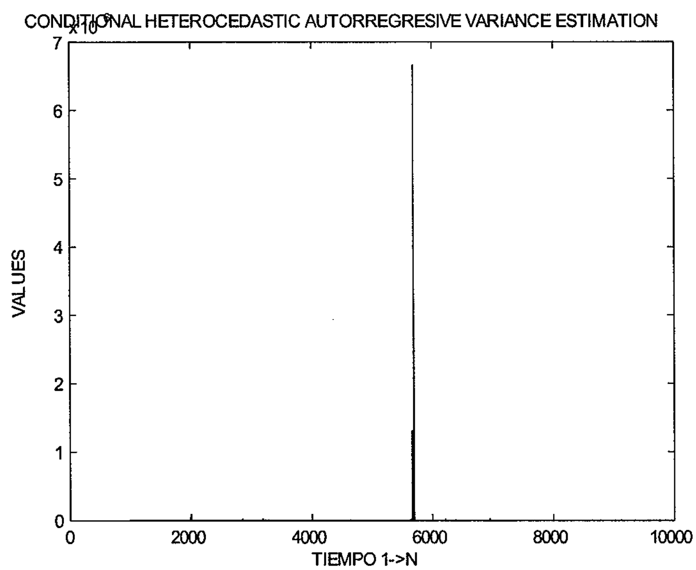

Fig. 2. Variance estimation for the above time series