LONG AND SHORT MEMORY CONDITIONAL HETEROSCEDASTICITY IN ESTIMATING THE MEMORY PARAMETER OF LEVELS[1]

by

P M Robinson and M Henry
London School of Economics and Political Science

Contents:

Discussion Paper
No. EM/98/357
August 1998

The Suntory Centre
Suntory and Toyota Internatianal Centres for Economics and Related Disciplines
London School of Economics and Political Science
Houghton Street
London WC2A 2AE
Tel.: 0171-405 7686

[1] Financial support from ESRC grant R000235892, from the Ecole Polytechnique, and from the Financial Markets Group is gratefully acknowledged. The paper was revised while the first author held a Leverhulme Personal Research Professorship and the second author was supported by a grant from the CREST. We are grateful for the comments of the referees.

Abstract

Semiparametric estimates of long memory seem useful in the analysis of long financial time series because they are consistent under much broader conditions than parametric estimates. However, recent large sample theory for semiparametric estimates forbids conditional heteroscedasticity. We show that a leading semiparametric estimate, the Gaussian or local Whittle one, can be consistent and have the same limiting distribution under conditional heteroscedasticity as under conditional homoscedasticity assumed by Robinson (1995a). Indeed, noting that long memory has been observed in the squares of financial time series, we allow, under regularity conditions, for conditional heteroscedasticity of the general form introduced by Robinson (1991) which may include long memory behaviour for the squares, such as the fractional noise and autoregressive fractionally integrated moving average form, as well as standard short memory ARCH and GARCH specifications.

Keywords: long memory; dynamic conditional heteroscedasticity; semiparametric estimation.

JEL Nos.: C13, C14, C15, C22, G15.

1 Introduction

In recent years, tools for investigating possible long memory in time series have been considerably developed. Early work of Mandelbrot (1969) considered the possibility of long memory modelling in economic and financial time series, his work and that of Adenstedt (1974) began parametric modelling of long memory, while Geweke and Porter-Hudak (1983) introduced semiparametric procedures, and empirical applications have become numerous. A review of the literature from an econometric standpoint is in Robinson (1994). Very long, approximately stationary series, such as series of asset returns and other financial measurements, are best analyzed, at least at an initial stage, by semiparametric estimates. They have the advantage of avoiding precise specification in that they parametrically model only the low frequency part of the spectral density (or the long-lagged autocovariances), thus avoiding inconsistency in estimation of even the low frequency structure that would be caused by misspecification (or overfitting) of the short memory dynamics. Semiparametric estimates have a slower rate of convergence than parametric ones, but with sufficient data this concern may be outweighed by their greater robustness properties.

We semiparametrically model long memory in a covariance stationary series x_t, $t = 0, \pm 1, \ldots$, by

$$f(\lambda) \sim G\lambda^{1-2H} \text{ as } \lambda \to 0^+, \tag{1.1}$$

where $\frac{1}{2} < H < 1$ and $0 < G < \infty$, $f(\lambda)$ being the spectral density of x_t satisfying

$$\gamma_j = \mathrm{cov}(x_t, x_{t+j}) = \int_{-\pi}^{\pi} f(\lambda)\cos(j\lambda)\, d\lambda, \quad j = 0, \pm 1, \ldots . \tag{1.2}$$

Under (1.1), $f(\lambda)$ has a pole at $\lambda = 0$ for $1/2 < H < 1$ (when there is long memory in x_t), $f(\lambda)$ is positive and finite for $H = 1/2$ (which we identify with short memory in x_t) and $f(0) = 0$ for $0 < H < 1/2$ (which we describe as negative dependence or antipersistence). Two leading semiparametric estimates of the memory parameter H are the log periodogram estimate of Geweke and Porter-Hudak (1983) and the Gaussian or local Whittle estimate of Künsch (1987). Only recently has asymptotic distributional theory of these estimates been laid down, by Robinson(1995a,b), though earlier attempts in case of the log periodogram estimate appear in the literature, and in fact, the version of the log periodogram estimate considered by Robinson (1995b) differs from the original, and also provides efficiency improvements. Even with such improvements, the Gaussian semiparametric estimate is the more efficient. Unlike the log periodogram estimate, it is not defined in closed form, but nonlinear optimization is only needed with respect to a single parameter, H, and can be accomplished rapidly.

The asymptotic theory of Robinson(1995a,b) rules out the possibility of conditional heteroscedasticity, and this seems a drawback in case of financial series for which semiparametric estimates otherwise seem appropriate. Indeed, Robinson (1995b) analyzed the log periodogram under the assumption that x_t is Gaussian. For the Gaussian semiparametric estimate he made the weaker assumption

$$x_t \;=\; E(x_t) \;+\; \sum_{j=0}^{\infty} \alpha_j \varepsilon_{t-j}, \quad \sum_{j=0}^{\infty} \alpha_j^2 \;< \infty, \tag{1.3}$$

where the ε_t satisfy at least

$$E(\varepsilon_t|\mathcal{F}_{t-1}) = 0 \quad \text{almost surely (a.s.)}, \tag{1.4}$$

$$\sigma_t^2 \overset{\text{def.}}{=} V(\varepsilon_t|\mathcal{F}_{t-1}) = \sigma^2 \quad \text{a.s.}, \tag{1.5}$$

for all t, where $\mathcal{F}_t$ is the σ-field of events generated by $(\varepsilon_s,\ s \le t)$ and σ^2 is a constant. We would like to relax (1.5) to allow for the possibility of autocorrelation in the ε_t^2, for example in some financial applications, the levels x_t can be approximated by a martingale sequence (so $\alpha_j = 0,\ j > 0$) but the squares $x_t^2 = \varepsilon_t^2$ cannot, so that the sequence x_t is not a sequence of independent random variables. In fact, empirical evidence (see, e.g. Ding, Granger, and Engle (1993)) has also suggested that dependence in the squares can fall off very slowly, in a way that is possibly more consistent with long memory than with standard short memory ARCH and GARCH specifications.

In fact, prior to Ding, Granger, and Engle (1993), GARCH-type models admitting the possibility of long memory had already been proposed by Robinson (1991) and applied to financial time series by Whistler (1990). Robinson (1991) considered the specifications

$$\sigma_t^2 = \sigma^2 + \sum_{j=1}^{\infty} \psi_j (\varepsilon_{t-j}^2 - \sigma^2) \tag{1.6}$$

and

$$\sigma_t^2 = \left(\sigma + \sum_{j=1}^{\infty} \psi_j \varepsilon_{t-j} \right)^2$$

We shall discuss only the "ARCH(∞)" specification (1.6). This can be reparameterized as

$$\sigma_t^2 = \beta + \sum_{j=1}^{\infty} \psi_j \varepsilon_{t-j}^2$$

and includes both standard ARCH (when $\psi_j = 0,\ j > p$, for finite p) and GARCH (when the ψ_j decay exponentially) models. More generally, if, for complex valued z,

$$\psi(z) = 1 - \sum_{j=1}^{\infty} \psi_j z^j \tag{1.7}$$

satisfies

$$|\psi(z)| \neq 0,\ |z| \leq 1, \tag{1.8}$$

define

$$\phi(z) = \sum_{j=0}^{\infty} \phi_j z^j = \psi(z)^{-1},\ \phi_0 = 1. \tag{1.9}$$

Then Robinson (1991) rewrote (1.6) as

$$\varepsilon_t^2 - \sigma^2 = \sum_{j=0}^{\infty} \phi_j \nu_{t-j}, \tag{1.10}$$

where

$$\nu_t = \varepsilon_t^2 - \sigma_t^2 \tag{1.11}$$

satisfies

$$E(\nu_t | \mathcal{F}_{t-1}) = 0 \quad \text{a.s.}, \tag{1.12}$$

by construction. The requirement

$$0 < \sum_{j=0}^{\infty} \phi_j^2 < \infty \tag{1.13}$$

includes the traditional long memory specifications of moving average coefficients, for example the autoregressive fractionally integrated moving average (ARFIMA) case

$$\phi(z) = (1-z)^{-d}\, \frac{b(z)}{a(z)}, \tag{1.14}$$

for $0 < d < \frac{1}{2}$ and finite order polynomials $a(z)$ and $b(z)$ whose zeros are outside the unit circle in the complex plane, and the fractional noise case

$$\text{corr}\left(\varepsilon_t^2, \varepsilon_{t+j}^2\right) = \frac{\sum_{i=0}^{\infty} \phi_i \phi_{i+j}}{\sum_{i=0}^{\infty} \phi_i^2} = \frac{1}{2}\left\{|j-1|^{2d+1} - |j|^{2d+1} + |j+1|^{2d+1}\right\}. \tag{1.15}$$

Robinson (1991) developed Lagrange multiplier tests for no-ARCH against alternatives consisting of general finite parameterization of (1.6), specializing to (1.14) and (1.15). In both these cases, the autoregressive weights ϕ_j satisfy

$$\sum_{j=0}^{\infty} |\psi_j| < \infty. \tag{1.16}$$

Under

$$\max_t E\left(\varepsilon_t^4\right) < \infty, \tag{1.17}$$

it follows that

$$\begin{aligned} E\left(\nu_t^2\right) &\leq 2\left[E\left(\varepsilon_t^4\right) + E\left\{E\left(\varepsilon_t^2|\mathcal{F}_{t-1}\right)\right\}^2\right] \\ &\leq 4E\left(\varepsilon_t^4\right) \leq K, \end{aligned} \tag{1.18}$$

where K is a generic finite constant, so that the innovations ν_t in (1.10) are square integrable martingale differences, ε_t^2 is well defined as a covariance stationary process and its autocorrelations can exhibit the usual long memory structure implied by (1.14) or (1.15). Even if (1.17) does not hold, the "autocorrelations" $\sum_{i=0}^{\infty} \phi_i \phi_{i+j} / \sum_{i=0}^{\infty} \phi_i^2$ are well defined under (1.13). Giraitis, Kokoszka, and Leipus (1998) have derived sufficient conditions for a stationary solution of (1.6), given that $\varepsilon_t = \eta_t \sigma_t$ for i.i.d. η_t and $\psi_j \geq 0$ for all j, which do not cover long memory in ε_t^2, so the character of solutions of (1.6) remains open to further study.

Subsequent to Robinson (1991), similar long memory versions of (1.6) have been pursued by Baillie, Bollerslev, and Mikkelsen (1996), Ding and Granger (1996) and others, for example, the model labelled (4.27) in Ding and Granger (1996) was discussed in Section 5 of Robinson (1991), being the case $a = b \equiv 1$ in (1.10), (1.14) above. Alternative models which provide long memory in squares and short memory in levels were proposed by Robinson and Zaffaroni (1997,1998).

In view of the empirical evidence of Whistler (1990) and Ding, Granger, and Engle (1993), it seems appropriate to allow for possible long memory in ε_t^2 in inference on long memory in x_t. In this paper, we consider the Gaussian semiparametric estimate of H in these circumstances, partly because it is well motivated by superior efficiency properties under the previous conditions, and because the log periodogram estimate (and some others) are technically more complex and cumberome to handle when Gaussianity is relaxed, due to their highly nonlinear structure.

The following section describes the Gaussian semiparametric estimate of H. Because the estimate is of the implicitly defined extremum type, one has to establish consistency prior to deriving limiting distributional behaviour, and these tasks are carried out in Sections 3, the proofs appearing in an appendix. Section 4 reports a small Monte Carlo study of finite sample behaviour. Section 5 contains some concluding comments.

2 Semiparametric Gaussian estimate

On the basis of observations x_t, $t = 1, \ldots, n$, define the periodogram

$$I(\lambda) = \frac{1}{2\pi n} \left| \sum_{t=1}^{n} x_t e^{it\lambda} \right|^2,$$

and consider estimating H by

$$\hat{H} = \mathrm{argmin}_{\Delta_1 \leq h \leq \Delta_2} R(h),$$

where $0 < \Delta_1 < \Delta_2 < 1$ and

$$R(h) = \log \left\{ \frac{1}{m} \sum_{j=1}^{m} \frac{I(\lambda_j)}{\lambda_j^{1-2h}} \right\} - (2h-1) \frac{1}{m} \sum_{j=1}^{m} \log \lambda_j,$$

in which $m \in (0, [n/2])$ and $\lambda_j = 2\pi j/n$.

As explained in Robinson (1995a), for $m = [n/2]$, $\hat{H}$ is a form of Gaussian or Whittle estimate under the parametric model $f(\lambda) = G|\lambda|^{1-2H}$, all $\lambda \in (-\pi, \pi]$, and its asymptotic properties would be approximately covered by Fox and Taqqu (1986), Giraitis and Surgailis (1990) and others, under Gaussianity, or more generally the assumption that x_t is linear with independent and identically distributed innovations. (These authors considered continuous, rather than discrete, averaging over frequencies.) When $m < [n/2]$ such that, as $n \to \infty$,

$$\frac{1}{m} + \frac{m}{n} \to \infty, \tag{2.1}$$

$\hat{H}$ can be viewed as a semiparametric estimate based on (1.1), and can be derived by concentrating out the scale factor from a narrow-band form of Whittle objective function. Under (1.1), (1.3)-(1.5) and (2.1), and other regularity conditions, Robinson (1995a) showed that $\hat{H}$ is consistent for H, and under further conditions that

$$m^{\frac{1}{2}}(\hat{H} - H) \to_d N(0, \frac{1}{4}) \quad \text{as} \quad n \to \infty. \tag{2.2}$$

The bandwidth parameter m is analogous to that employed in weighted periodogram estimates of the spectral density of short memory processes. Clearly (2.1) is a minimal requirement for consistency under (1.1). Henry and Robinson (1996) discussed optimal choices of m in the determination of $\hat{H}$.

The compact set $[\Delta_1, \Delta_2]$ of admissible h values in Robinson (1995a) can include ones between $\frac{1}{2}$ and 1, where there is long memory, ones between 0 and $\frac{1}{2}$, where there is negative dependence or antipersistence, and $h = \frac{1}{2}$, where there is short memory. It seems desirable to avoid assuming, say, $\frac{1}{2} < H < 1$, a priori, but rather to allow also for the possibility that $H \leq \frac{1}{2}$, especially in view of the very mixed evidence of the existence of long memory in levels of financial series (see, e.g. Lo (1991), Lee and Robinson (1996)), in view of the efficient markets hypothesis, under which $H = \frac{1}{2}$, and in view of the possibility that log price levels may be nonstationary with less than a unit root, in which case returns can exhibit negative dependence (as in Henry and Payne (1997)). By contrast, the bulk of asymptotic theory relevant to long memory assumes a priori that long memory exists.

It turns out that not only is $\hat{H}$ still consistent for H in the presence of the (possibly long memory) ARCH behaviour described in the previous section (although with stronger moment conditions), but (2.2) holds in detail with the same asymptotic variance, so that no features of the ARCH structure defined by (1.6) or (1.10) enter. This outcome is not entirely predictable, since ARCH-type behaviour can affect limiting distributional properties (see, e.g. Weiss (1986), Kuersteiner (1997)). It is especially desirable in the present case. This is in the first place due to the simplicity of the limiting variance in (2.2), which is independent of both H and G. Moreover, although maximum likelihood estimation of parametric versions of (1.10), such as (1.14) and (1.15), is implicit in the derivation of LM tests by Robinson (1991), no rigorous asymptotic theory exists for such estimates, apart from the ARCH and GARCH special cases studied by Weiss (1986), Lee and Hansen (1994) and Lumsdaine (1996). Thirdly, there is no asymptotic theory available for semiparametric estimation of the memory parameter determining the asymptotic behaviour of the ψ_j or ϕ_j in (1.6) and (1.10). We will return to this last point in section 5. Our derivation of the asymptotic properties of $\hat{H}$ follows the main steps of the proof in Robinson (1995a), and uses a number of properties established there, but it also differs significantly, posing new challenges. This appears to be the first paper which develops asymptotic theory in a long memory context that allows for ARCH structure. Long memory is not covered by the mixing conditions stressed in much econometric literature, the long memory literature featuring either Gaussian processes (e.g. Fox and Taqqu (1986), Robinson (1995b)), nonlinear functions of Gaussian processes (e.g. Taqqu (1979)), linear filters of independently and identically distributed sequences (e.g. Giraitis and Surgailis (1990)), nonlinear functions of such linear filters ('Appel polynomials', see Giraitis and Surgailis (1986)), as well as the model (1.3)-(1.5). None of these approaches represents conditional heteroscedasticity in a martingale difference sequence.

3 Consistency and asymptotic normality of the Gaussian semiparametric estimate.

We introduce the following assumptions.

Assumption A1 For $H \in [\Delta_1, \Delta_2]$, $0 < \Delta_1 < \Delta_2 < 1$, and $0 < G < \infty$, $f(\lambda)$ satisfies (1.1).

Assumption A2 In a neighbourhood $(0, \delta)$ of the origin, $f(\lambda)$ is differentiable and

$$\frac{d}{d\lambda} \log f(\lambda) = O(\lambda^{-1}) \quad \text{as} \quad \lambda \to 0^+.$$

Assumption A3 x_t satisfies (1.3), (1.4) and (1.17) with σ_t^2 given by (1.6) such that (1.16) holds and the ϕ_j defined by (1.7)-(1.9) satisfy (1.13). In addition either

$$E\left(\varepsilon_t^3 | \mathcal{F}_{t-1}\right) = E(\varepsilon_t^3) \text{ a.s.}, \quad t = 0, \pm 1 \ldots, \tag{3.1}$$

or

$$\sum_{j=0}^{\infty} |\phi_j| < \infty. \tag{3.2}$$

Assumption A4 m satisfies (2.1).

Assumptions A1, A2 and A4 are identical to the equivalently-numbered ones of Robinson (1995a). We stress that only local (to zero) assumptions are made on $f(\lambda)$, so that it need not be smooth, or even bounded (or nonzero) outside a neighbourhood of the origin. In place of the current Assumption A3, Robinson (1995a) assumed (1.3)-(1.5) with a homogeneity condition, so that we require more moments while allowing for ARCH behaviour, possibly with long memory. The requirement (3.1) that conditional third moments be nonstochastic is restrictive, but satisfied if ε_t has a conditionally symmetric density, or, more specially, if

$$\varepsilon_t | \mathcal{F}_{t-1} \sim N(0, \sigma_t^2). \tag{3.3}$$

The alternative requirement (3.2) rules out long memory in ε_t^2 but covers standard ARCH and GARCH specifications (that is (1.14) with $d = 0$), as well as many processes for which autocorrelation in squares decays more slowly than exponentially. Note that (1.17) itself entails a restriction on the magnitude of the ϕ_j; see for instance the results of Engle (1982), Bollerslev (1986) for ARCH(1) and GARCH(1,1) processes under (3.3), and of Nelson (1990) under more general distributional assumptions. However, (1.17) is not a necessary condition, and indeed, under (3.2) it can be shown to be unnecessary by means of a longer argument, involving truncations, than that in the proof of the following theorem.

Theorem 1 Under Assumptions A1-A4,

$$\hat{H} \rightarrow_p H, \quad \text{as } n \rightarrow \infty.$$

The limiting distributional properties of $\hat{H}$ rest on stronger conditions than those sufficient for consistency.

Assumption A1' For some $\beta \in (0, 2]$,

$$f(\lambda) \sim G\lambda^{1-2H}\left(1 + O(\lambda^{\beta})\right) \quad \text{as } \lambda \rightarrow 0^+,$$

where $G \in (0, \infty)$ and $H \in [\Delta_1, \Delta_2]$.

Assumption A2' In a neighbourhood $(0, \delta)$ of the origin, $\alpha(\lambda)$ is differentiable and

$$\frac{d}{d\lambda}\log|\alpha(\lambda)| = O(\frac{|\alpha(\lambda)|}{\lambda}) \quad \text{as } \lambda \rightarrow 0^+,$$

where $\alpha(\lambda) = \sum_{j=0}^{\infty}\alpha_j e^{ij\lambda}$.

Assumption A3' The first sentence of Assumption A3 holds, and

$$\max_t E\varepsilon_t^8 < \infty, \tag{3.4}$$

$$E\left(\varepsilon_t^2\varepsilon_u\varepsilon_{v-1}\right) = 0,\ E\left(\varepsilon_t^4\varepsilon_u|\mathcal{F}_{u-1}\right) = E\left(\varepsilon_t^4\varepsilon_u^2\varepsilon_v|\mathcal{F}_{v-1}\right) = 0,\ \text{a.s.},\ t \geq u \geq v, \tag{3.5}$$

$$\phi_j = O(j^{d-1}), \quad \text{as } j \rightarrow \infty, \quad d < \frac{1}{2}, \tag{3.6}$$

$$\alpha_j = O(j^{H-\frac{3}{2}}), \quad \text{as } j \rightarrow \infty, \tag{3.7}$$

and the α_j are quasi-monotonically decreasing.

Assumption A4' As $n \rightarrow \infty$

$$\frac{1}{m} + \frac{m^{1+2\beta}(\log m)^2}{n^{2\beta}} + \frac{(m\log m)^2}{n} \rightarrow 0, \tag{3.8}$$

and, if (3.2) does not hold, for the same d as in Assumption A3'

$$\frac{m(\log m)}{n^{\frac{1}{2}-d}} \rightarrow 0. \tag{3.9}$$

Compared to the corresponding assumptions in Robinson (1995a), Assumptions A1' and A2' are unchanged (still restricting $f(\lambda)$ only near the origin, such that β indicates the smoothness of $f(\lambda)/G\lambda^{1-2H}$ there), but Assumptions A3' and A4' trade off the relaxation of constant conditional innovations variances and fourth moments with some strengthening of conditions. The eighth moment condition (3.4) replaces the fourth moment condition of Robinson (1995a), while, when there is long memory in the ε_t^2, extension of (3.1) to (3.5) is again satisfied in case (3.3). The strengthening of moment conditions is a matter both of practical concern, in view of the characteristics of much financial data, and of theoretical concern in view of the results of Engle (1982), Bollerslev (1986), Nelson (1990), for example. As with Theorem 1, it is likely that Theorem 2 below can be established under a milder moment condition by a more detailed argument. Note, however, that Davis and Mikosch (1997) have shown that the sample autocorrelations of squares of ARCH(1) sequences have non-degenerate probability limits when fourth moments do not exist. Condition (3.5) is seen to hold under (3.3), on noting that then

$$E\left(\varepsilon_t^4|\mathcal{F}_{t-1}\right) = 3\sigma_t^2, \quad E\left(\varepsilon_t^6|\mathcal{F}_{t-1}\right) = 15\sigma_t^6, \quad \text{a.s.},$$

and applying these properties and (1.4), (1.6) and (1.16) recursively. Condition (3.6) strengthens (1.13) while being satisfied in the examples (1.14) and (1.15). d can be arbitrarily close to $\frac{1}{2}$, so that (3.6) is not of great concern in itself, except that (3.9) strengthens (3.8) unless $d \le (1-2\beta)/(4\beta+2)$, which is possible only when $\beta < \frac{1}{2}$ is chosen in (3.8), whereas when the levels x_t themselves have fractional noise or ARFIMA long memory (analogous to models (1.15) and (1.14) for ε_t^2), $\beta = 2$ is available in Assumption A1'. In (3.8), the requirement $(m \log m)^2/n \to 0$ was not in Robinson (1995a), but it does not bind when $\beta \le \frac{1}{2}$. Fractional noise and ARFIMA x_t satisfy (3.7), which is consistent with Assumption A1', and also satisfy the quasi-monotonicity assumption on the α_j, which entails (see Yong (1974)), for all sufficiently large j

$$|\alpha_j - \alpha_{j+1}| \le K\frac{|\alpha_j|}{j}. \tag{3.10}$$

In fact, we believe that this requirement, and (3.9), could be removed or relaxed by a more detailed proof, but the quasi-monotonicity requirement does not seem very onerous, while (3.9) is also needed when the ε_t^2 have long memory, and there always exists an m sequence satisfying both (3.8) and (3.9).

Theorem 2 Under Assumptions A1'-A4', (2.2) holds.

The most notable aspect of this Theorem 2 is that the asymptotic variance, 1/4, achieved by Robinson (1995a) is not affected by the conditional heteroscedasticity.

For readers not wishing to go through the proof of Theorem 2 in the appendix, we provide here a briefer, more intuitive explanation of this outcome, in case of the simple ARCH(1) model

$$\sigma_t^2 = \beta + \psi_1 \varepsilon_{t-1}^2. \tag{3.11}$$

The most likely way in which conditional heteroscedasticity could affect the asymptotic variance is through the variance of the normalized score $m^{\frac{1}{2}} dR(H)/dh$. It turns out (see Robinson (1995a)) that this can be approximated by a quantity proportional to

$$\sum_{t=2}^{n} z_t, \tag{3.12}$$

where

$$z_t = \varepsilon_t \xi_t, \qquad \xi_t = \sum_{s=1}^{t-1} \varepsilon_s c_{t-s} \tag{3.13}$$

$$c_s = \frac{2}{n m^{\frac{1}{2}}} \sum_{j=1}^{m} b_j \cos(s\lambda_j), \qquad b_j = \log j - \frac{1}{m} \sum_{i=1}^{m} \log i. \tag{3.14}$$

Now the asymptotic variance of (3.12) is unaffected by conditional heteroscedasticity if

$$\sum_{t=2}^{n} E\left\{\left(\sigma_t^2 - \sigma^2\right) \xi_t^2\right\} \to 0, \quad \text{as} \quad n \to \infty. \tag{3.15}$$

Under (3.11), (3.15) is proportional to

$$\sum_{t=2}^{n} E\left(\chi_{t-1}\xi_t^2\right) = \sum_{t=2}^{n} E\left(\chi_{t-1} \sum_{s=1}^{t-1} c_{t-s}^2 \chi_s\right) + \sum_{t=2}^{n} E\left(\chi_{t-1} \sum_{v \neq s}^{t-1} \varepsilon_v \varepsilon_s c_{t-v} c_{t-s}\right), \tag{3.16}$$

where $\chi_t = \varepsilon_t^2 - \sigma^2$. The second term on the right is zero on applying (1.10), nested conditional expectations, (3.1) and (1.4). The first term on the right of (3.16) is bounded in absolute value by

$$\sum_{t=2}^{n} \sum_{s=1}^{t-1} c_{t-s}^2 |\gamma_{t-s-1}|, \tag{3.17}$$

where $\gamma_j = \operatorname{cov}(\varepsilon_t^2, \varepsilon_{t+j}^2)$. (3.17) tends to zero by the Toeplitz lemma because $\sum_{t=2}^{n} \sum_{s=1}^{t-1} c_{t-s}^2 \to 1$ (see Robinson (1995a)) and $\gamma_j \to 0$ as $j \to \infty$ under (3.11); in fact arbitrarily slow decay in the autocorrelations of the squares ε_t^2 suffices.

4 Finite sample comparison

While the asymptotic properties of $\hat{H}$ which we have established are highly desirable, and reassuring in applications to long financial series, it is of interest to examine their relevance to series of more moderate length. For example, conditional heteroscedasticity might worsen the normal approximation in (2.2), and if there is considerable persistence, of the ARCH or GARCH type or especially of the long memory type which our asymptotics may also permit, the variance of $\hat{H}$ might differ considerably from $1/4m$. It is also of interest to consider robustness to departures from the moment conditions of Theorems 1 and 2. Finite sample performance of $\hat{H}$ was examined under the presumption of no conditional heteroscedasticity by Robinson (1995a), and compared with that of a version of the log-periodogram estimate, while Taqqu and Teverovsky (1995) include such estimates in a more comprehensive simulation study, but again restricted to conditionally homoscedastic environments. We report a Monte Carlo study of $\hat{H}$ applied to simulated series x_t following an ARIMA(0,$H-\frac{1}{2}$,0) parametric version of (1.3), for various H and various forms of conditional heteroscedasticity in ε_t.

We first took $\varepsilon_t = \sigma_t \eta_t$, where the η_t are NID(0,1), so that (3.3) is satisfied, and σ_t follows one of the specifications below.

(i) IID: $\sigma_t^2 = \sigma^2$. The ε_t are independent and identically distributed, so that there is no conditional heteroscedasticity. We can take $\sigma^2 = 1$ with no loss of generality.

(ii) ARCH: $\sigma_t^2 = .5 + .5\varepsilon_{t-1}^2$. The ε_t are ARCH(1) with modest autocorrelation in the ε_t^2; they satisfy (1.17), but not (3.4) (Engle (1982)).

(iii) GARCH: $\sigma_t^2 = .05 + .5\varepsilon_{t-1}^2 + .45\sigma_{t-1}^2$. The ε_t are GARCH(1,1), with strong autocorrelation in the ε_t^2 at "short" lags (nearly IGARCH); they do not satisfy (1.17) (Bollerslev (1986)).

(iv) LMARCH: $\sigma_t^2 = \left\{1-(1-L)^{.25}\right\}\varepsilon_t^2$. The ε_t have (moderate) long memory ARCH structure satisfying (1.6)-(1.9) and (1.14) with $a(z) = b(z) = 1$, so that the ε_t^2 follow the ARFIMA(0,d,0) structure discussed in Section 5 of Robinson (1991), with $d = .25$.

(v) VLMARCH: $\sigma_t^2 = \left\{1-(1-L)^{.45}\right\}\varepsilon_t^2$. The ε_t have "very long memory" ARCH structure, such that the ε_t^2 follow the same type of model as in (iv) but with $d = .45$, close to the stationarity boundary.

The model specification (1.6) adopted here for σ_t^2 does not allow for asymmetric response of conditional variances to positive and negative returns, which is reported in the empirical finance literature as the leverage effect. We have nevertheless also considered a form of Nelson's EGARCH (Nelson (1991)), which models the leverage effect.

(vi) EGARCH: $\ln \sigma_t^2 = -.5 + .9 \ln \sigma_{t-1}^2 - .5\eta_{t-1} + .5|\eta_{t-1}|$, with η_t still NID(0,1). The coefficient of η_{t-1} induces a strong leverage effect, volatility rising in response to unexpectedly low returns. In case of unexpectedly high returns, the volatility behaves as in an AR(1) stochastic volatility model, with AR coefficient calibrated on typical values in the empirical literature on financial volatilities (which are nearly always larger than .9, see e.g. Ghysels, Harvey, and Renault (1996)). The innovations ε_t have finite unconditional moments of all orders.

So far as the ARFIMA($0,H-\frac{1}{2},0$) model for x_t is concerned (so that in relation to (1.3), $\sum_{j=0}^{\infty}\alpha_j L^j = (1-L)^{\frac{1}{2}-H}$), we consider:

(a) "Antipersistence": H=.25,

(b) "Short memory": H=.5,

(c) "Moderate long memory": H=.75,

(d) "Very long memory": H=.95.

We study each of (i)-(vi) with (a)-(d), covering a range of short/long/negative memory in ε_t and a range of short/long memory in ε_t^2.

Tables 1-4 deal respectively with each of the four H values (a)-(d). In each case the results are based on n=64, 128 and 256 observations, with bandwidths m= $n/16$, $n/8$, $n/4$, and 10000 replications, as in the Monte Carlo study of Robinson (1995a) with conditionally homoscedastic ε_t. In each table we report, for the conditional variance specifications (i)-(vi), Monte Carlo bias of the Gaussian semiparametric estimate; Monte Carlo root mean squared error (MSE); 95% coverage probabilities based on the N(H, $1/4m$) approximation (2.2) for $\hat{H}$; and also the efficiency of the log-periodogram estimate relative to the Gaussian estimate, that is the ratio of the Monte Carlo mean squared errors, and we can compare this with the ratio of the asymptotic variances $\sqrt{6}/\pi \simeq .78$. We make the comparison with the log periodogram estimate (the version in Robinson (1995b), but with no trimming) because it has been popularly used, but we do not otherwise report the results for this estimate.

The innovations ε_t were generated recursively with starting values subsequently discarded. In particular, $\varepsilon_t = \sigma_t\eta_t$ with $\sigma_t^2 = 1$, $t = -1000, \ldots, 0$, and $\sigma_t^2 = \sigma^2 P(L)\varepsilon_t^2$, $t = 1, \ldots, 2n$, where $\eta_t \sim$ NID(0,1) and σ^2 and $P(L)$ are the relevant intercept and operator in cases (i) to (v), the latter being truncated to 1000 lags in the two long memory cases (iv) and (v). In case (vi), $\ln \sigma_t^2$ was generated recursively according to the formula. The Gauss random number generator RNDN was used with random seed starting at the value 12145389. A method based on the Cholevsky decomposition $(m_{i,j})_{i,j=1}^{2n}$ of the Toeplitz matrix $\left(\rho_{|i-j|}\right)_{i,j=1}^{2n}$, where ρ_j are the autocovariances of an ARFIMA($0,H-\frac{1}{2},0$), was then used to simulate x_t from the errors ε_t as $x_t = \sum_{i=1}^{t} m_{ti}\varepsilon_i$, $t = 1,\ldots,2n$, the first n values being subsequently discarded. For each series simulated, the periodogram was computed by the Gauss Fast Fourier Transform algorithm and $\hat{H}$ computed using a simple gradient algorithm. The optimization was constrained to the compact set $[.001, .999]$ (chosen values for Δ_1 and Δ_2 respectively) and for selected replications, $R(h)$ was plotted on the interval $[-1, 2]$ and was always found to be very smooth with a single relative minimum.

Perhaps the most striking feature of the results is the relatively poor performance of $\hat{H}$ and of the normal inference rule (2.2) provided by Theorem 2 in the GARCH case, compared to the other processes. Out of the 36 H, m, n combinations, the GARCH bias is largest in 18 cases, while its MSE ties largest in 3 cases and is outright largest in 28. Moreover the deviation of 95% coverage probabilities from their normal values ties largest 3 times and is outright largest 28 times, for GARCH. Relative efficiencies to the log periodogram estimate are also most out of line with their asymptotic values for the GARCH: it ties with the largest discrepancy 12 times and has the outright largest 10 times. To further investigate this relatively poor performance of $\hat{H}$ in case of GARCH errors, Monte Carlo empirical distribution functions of $2\sqrt{m}(\hat{H}-H)$ are plotted for all four values of H against the standard normal distribution function in figures 1 to 3, which correspond to three different choices of the pair (n, m), namely (64,4), (128,16) and (256,64). These empirical distributions are truncated because the estimate is restricted to the interval $[0.001, 0.999]$. In the case where $n = 64$ and $m = 4$, the empirical distributions are highly leptokurtic and a high proportion of estimates hit a boundary. When n and m increase, the tails become thinner.

Looking at the other heteroscedastic specifications, VLMARCH leads to a slightly worse performance than LMARCH, but with no reliable evidence that this is significantly worse than ARCH, or indeed IID. Failure of the moment conditions (1.17) and (3.4) has no evident effect. In our series of modest length, the relatively poor behaviour under GARCH may be better explained by the impact of a near unit root; for much larger values of n, LMARCH and VLMARCH would presumably do worse than GARCH, but in such samples this is unlikely to be a matter of great practical concern. In absolute terms, even GARCH does not perform so badly for us to question the usefulness of the asymptotic robustness results in moderate sample sizes. When $H = \frac{1}{2}$, $\hat{H}$ has almost identical root MSE and 95% coverage probabilities for EGARCH and ARCH. In the EGARCH case, Monte Carlo biases are typically larger when there is antipersistence and smaller in case of very long memory in levels. As expected, MSE decreases monotonically, as n and m increase. The decay in bias with increasing n is less noticable, while the typical decay in bias with increasing m is somewhat surprising but broadly in line with results of Robinson (1995a), in case of fractional Gaussian noise levels (which has similar spectral shape to that of the ARFIMA($0,H-\frac{1}{2},0$)). As in the no-ARCH finite sample results of Robinson (1995a), coverage probabilities are markedly sensitive to choice of m, and this problem clearly requires further study beyond that of Henry and Robinson (1996), though for larger n this is likely to be less of a problem.

Finally, the effect of heavy-tailed conditional distributions for ε_t is investigated in Tables 5 and 6 in case of short memory levels ($H = 0.5$). Monte Carlo biases, root MSEs, coverage probabilities and relative efficiencies of the log periodogram estimate are reported as before for models (i) to (v) only with $\varepsilon_t = \sigma_t \eta_t$, where the η_t are i.i.d. t_4 in Table 5 and i.i.d. t_2 in Table 6, so that η_t has respectively infinite fourth moment and infinite second moment. Relative efficiency of the log periodogram estimate seems unaffected by heavy-tailedness. However, when there is no conditional heteroscedasticity, $\hat{H}$ on the whole performs better when η_t is t_4 than when it is normal, and better still when it is t_2, in terms of Monte Carlo bias, MSE and coverage probability. Conditional heteroscedasticity produces a reverse picture. The results for t_4 η_t are better than those for normal η_t in only 7 cases in respect of bias, 4 in respect of MSE and 2 in respect of coverage probability. The results for t_2 are better than those for t_4 in only 1 case in respect of bias, 4 in respect of MSE and 4 in respect of coverage probability. Moreover, these exceptions are mostly for the $n = 64$, $m = 8$ combination, and frequently the deterioration produced by extreme heavy-tailedness is substantial. And although bias and MSE typically decrease with increasing n and m for t-distributed η_t, suggesting that consistency of $\hat{H}$ is maintained, there is some tendency for coverage probabilities to actually worsen (become smaller) especially for t_2, so that not only is the heavy-tailedness reflected in the distribution of $\hat{H}$ but there is evidence that the limit distribution of Theorem 2 may not hold under this violation of the moment conditions, in line with the evidence of Davis and Mikosch (1997) referred to earlier.

Overall the results suggest that the possibility of conditional heteroscedasticity can be a cause for concern in moderate sample sizes, especially for IGARCH-like behaviour and when the conditional distribution of ε_t has heavy tails. On the other hand, some forms of conditional heteroscedasticity appear to have little effect and in these circumstances, use of $\hat{H}$ and the associated large sample inference rules of Robinson (1995a) seems warranted at least for reasonably large samples, though as is typically the case with smoothed nonparametric estimation, reporting results for a range of bandwidths is a wise precaution.

5 Final comments.

This paper seems to be the first attempt to study the impact of conditional heteroscedasticity on the behaviour of semiparametric estimates of long memory. Moreover, we have allowed in the asymptotic theory not only for standard ARCH and GARCH specifications of conditional heteroscedasticity, but for the ARCH(∞) model for squared innovations introduced by Robinson (1991), which covers ARFIMA structure. The fact that the limiting distribution has the same simple form as under conditional homoscedasticity not only implies that existing rules of large sample statistical inference remain valid (including the test for $I(0)$ based on the objective function $R(h)$ recently developed by Lobato and Robinson (1998)), but also suggests that the formulae for asymptotic mean squared errors of $\hat{H}$ provided by Henry and Robinson (1996) will remain valid, and the consequent rules for the optimal choice of bandwidth m. So far as the technical contribution of the current paper is concerned, it seems that very similar methods can be used to investigate the large sample distribution theory of other statistics in the presence of (possibly long memory) conditional heteroscedasticity, such as nonparametric estimates of the spectral density of a process with short memory in levels, as well as more elaborate statistics.

The Gaussian semiparametric estimate can be used at an initial stage in the analysis of a series x_t, perhaps to test for a specific value of H such as $\frac{1}{2}$ (as in Lobato and Robinson (1998)), or to create a fractionally differenced series $\Delta^{\hat{H}-\frac{1}{2}}x_t$, where Δ is the differencing operator. This represents an asymptotically valid aproximation to an $I(0)$ series without any parametric assumption on the autocorrelation of the underlying $I(0)$ process $\Delta^{H-\frac{1}{2}}x_t$, so we might then proceed to identify the order of a parametric model such as an ARMA on the basis of the $\Delta^{\hat{H}-\frac{1}{2}}x_t$, possibly then carrying out estimation of the ARFIMA model for x_t by a parametric Gaussian method. A question that then arises is whether the innovations in the model (equivalent to our ε_t) have conditional heteroscedasticity, and if so, what is the nature and extent of it. This is of interest whether or not x_t has long memory, and even if x_t is a martingale difference, $x_t = \varepsilon_t$. If (1.10) is parameterized, say by (1.15) or (1.14), then we can estimate the unknown parameters by applying the conditional Gaussian loglikelihood underlying the LM tests developed by Robinson (1991), though asymptotic properties of the parameter estimates remain to be established in the long memory case, and indeed in many short memory ones. However, such a procedure carries the disadvantage that even the memory parameter d will be inconsistently estimated if the short memory dynamics of the squares is misspecified, while we may in any case prefer an exploratory approach at the initial stage.

One may thus consider applying a semiparametric procedure for estimating d to the

ε_t^2, or their proxies. For example, the Gaussian method appears to be a candidate, because, although the ε_t^2 cannot be Gaussian, Gaussianity of x_t was not assumed by Robinson (1995a), or in the current paper. However, while some of the analysis of these papers will be relevant, and (1.10) represents ε_t^2 as a linear filter of martingale differences ν_t, not only do the ν_t have conditional heteroscedasticity but their odd conditional moments are perforce stochastic, so that no conditions analogous to (3.1) or (3.5) can be imposed. The form of the limiting distribution of the Gaussian semiparametric estimate of d, as well as its derivation, are thus open questions.

Appendix

Proof of Theorem 1 The main part of the proof of the corresponding Theorem 1 of Robinson (1995a) applies except for the proof that

$$\sum_{r=1}^{m-1} \left(\frac{r}{m}\right)^{2(\Delta-H)+1} \frac{1}{r^2} \left| \sum_{j=1}^{r} \left(2\pi J(\lambda_j) - \sigma^2\right) \right| \to_p 0, \tag{A.1}$$

where

$$J(\lambda) = \frac{1}{2\pi n} \left| \sum_{t=1}^{n} \varepsilon_t e^{it\lambda} \right|^2$$

and $\Delta = \Delta_1$ when $H < \frac{1}{2} + \Delta_1$ and $\Delta \in (H - \frac{1}{2}, H]$ otherwise. (Note that unlike in Robinson (1995a), we take the unconditional variance of ε_t to be σ^2, not unity.)

The justification for the above claim rests on the fact that the remainder of the aforementioned proof depends only on unconditional second moment properties. In view of (3.18) of Robinson (1995a), (A.1) is implied if

$$\sum_{t=1}^{n} (\varepsilon_t^2 - \sigma^2) = o_p(n) \tag{A.2}$$

and

$$\sum_{\substack{s \neq t \\ 1}}^{n} \varepsilon_s \varepsilon_t A_{st}^{(r)} = o_p(r^{1-\eta} n), \text{ some } \eta > 0, \tag{A.3}$$

uniformly in $r \in [1, m-1]$, where $A_{st}^{(r)} = \sum_{j=1}^{r} \cos\left[(s-t)\lambda_j\right]$. The left side of (A.2) has mean zero and variance

$$\sum_{t,s=1}^{n} \sum_{j,k=0}^{\infty} \phi_j \phi_k E(\nu_{t-j} \nu_{s-k}) = \sum_{t,s=1}^{n} \sum_{j=0}^{\infty} \phi_j \phi_{j+s-t} E(\nu_{t-j}^2) \tag{A.4}$$

in view of (1.4), with $\phi_j = 0,\ j < 0$. In view of (1.18) and the Cauchy inequality, (A.4) is, with $\Phi_j = \left(\sum_{i=j}^{\infty} \phi_i^2\right)^{\frac{1}{2}}$,

$$O\left(n\sum_{j=0}^{\infty}\phi_j^2 + n\Phi_0 \sum_{j=1}^{n-1} \Phi_j\right) = o\left(n^2\right)$$

by the Toeplitz lemma and (1.9), thus verifying (A.2). To prove (A.3), the left hand side has variance

$$4E\left(\sum_{\substack{u<v\\1}}^{n}\sum_{\substack{s<t\\1}}^{n} \varepsilon_s\varepsilon_t\varepsilon_u\varepsilon_v A_{st}^{(r)} A_{uv}^{(r)}\right). \tag{A.5}$$

In view of (1.4) of Assumption A3, it is clear that no summands for which $t \neq v$ can contribute. Thus, (A.5) is

$$4E\left(\sum_{\substack{s<t\\1}}^{n} \varepsilon_t^2\varepsilon_s^2 A_{st}^{(r)2}\right) + 8E\left(\sum_{\substack{u<s<t\\1}}^{n} \varepsilon_t^2\varepsilon_s\varepsilon_u A_{st}^{(r)} A_{ut}^{(r)}\right). \tag{A.6}$$

The first term in (A.6) is bounded by

$$4\max_t E(\varepsilon_t^4)\sum_{\substack{s<t\\1}}^{n} A_{st}^{(r)2} = O(rn^2),$$

from (3.20) of Robinson (1995a). Substituting (1.10) in the second term of (A.6) gives

$$\begin{aligned}
8E&\left(\sum_{\substack{u<s<t\\1}}^{n}\left(\sigma^2 + \sum_{j=0}^{\infty}\phi_j\nu_{t-j}\right)\varepsilon_u\varepsilon_s A_{st}^{(r)} A_{ut}^{(r)}\right)\\
&= 8\sum_{\substack{u<s<t\\1}}^{n} \phi_{t-s} E\left(\nu_s\varepsilon_u\varepsilon_s\right) A_{st}^{(r)} A_{ut}^{(r)}\\
&= 8\sum_{\substack{u<s<t\\1}}^{n} \phi_{t-s} E\left(\varepsilon_s^3\varepsilon_u\right) A_{st}^{(r)} A_{ut}^{(r)}.
\end{aligned}$$

Under (3.1), this is identically zero. Under (3.2), it is bounded in absolute value by

$$8r\max_t E(\varepsilon_t^4)\sum_{j=0}^{\infty}|\phi_j|\sum_{\substack{s<t\\1}}^{n}\left|A_{st}^{(r)}\right| \leq Krn\left(\sum_{\substack{s<t\\1}}^{n} A_{st}^{(r)2}\right)^{\frac{1}{2}} = O\left(r^{\frac{3}{2}}n^2\right)$$

because $\left|A_{st}^{(r)}\right| \leq r$. Thus, (A.3) is verified.

As explained by Robinson (1995a), there is a lack of uniformity in the convergence of $R(h)$ around $h = H - \frac{1}{2}$ which is of concern when $H \geq \frac{1}{2} + \Delta$, and then one has to show also that

$$\frac{1}{m}\sum_{j=1}^{m}(a_j - 1)\left(2\pi J(\lambda_j) - \sigma^2\right) \rightarrow_p 0 \tag{A.7}$$

where $a_j = (\frac{j}{p})^{2(\Delta-H)}$ for $1 \leq j \leq p$, and $a_j = (\frac{j}{p})^{2(\Delta_1-H)}$ for $p < j \leq m$, where $p = \exp(\frac{1}{m}\sum_{j=1}^{m}\log j)$. However, by similar arguments to those used above we establish (A.7) under Assumption A3, in view of the proposition, established in Robinson (1995a), that $\sum_{t=1}^{n}\sum_{s\neq t}^{n}\left[\sum_{j=1}^{m}(a_j-1)\cos\{(s-t)\lambda_j\}\right]^2 = o(mn^2)$.

Proof of Theorem 2 Again, the basic structure of the proof of Robinson (1995a) is unchanged, and a number of properties established there are still of use. Again a mean value theorem argument is applied, and the scores approximated by a martingale. The approximation, and the treatment of second derivatives of $R(h)$, are affected by the changed conditions, but we postpone discussion of this until after we have established the asymptotic normality of the approximating martingale, whose proof is considerably affected.

With the definitions (3.12) and (3.14), $\sum_2^n z_t$ is a martingale and we wish to show, as in Robinson (1995a), that as $n \rightarrow \infty$

$$\sum_{t=1}^{n} E(z_t^4) \rightarrow 0, \tag{A.8}$$

$$\sum_{t=1}^{n} E\left(z_t^2|\mathcal{F}_{t-1}\right) \rightarrow_p \sigma^4. \tag{A.9}$$

By the Schwarz inequality, $E\left(z_t^4\right) \leq (E\varepsilon_t^8)^{\frac{1}{2}}(E\xi_t^8)^{\frac{1}{2}}$. Because the ε_t are martingale differences, by Burkholder's (Burkholder (1973)) and c_r-inequalities

$$E\left(\xi_t^8\right) \leq KE\Big(\sum_{s=1}^{t-1}c_{t-s}^2\varepsilon_s^2\Big)^4 \leq \left(\max_s E\varepsilon_s^8\right)r_n^4 = O\left((\log m)^8/n^4\right)$$

uniformly in t by (4.22) of Robinson (1995a), with $r_t = c_1^2 + \ldots + c_t^2$. Thus,

$$\sum_{t=1}^{n} E\left(z_t^4\right) \leq K\frac{(\log m)^4}{n} \rightarrow 0$$

to verify (A.8). To check (A.9), write

$$E\left(z_t^2|\mathcal{F}_{t-1}\right) = \sigma_t^2\xi_t^2 = \sigma^2\xi_t^2 + (\sigma_t^2 - \sigma^2)\xi_t^2.$$

From (4.14) and (4.15) of Robinson (1995a),

$$\sum_{t=1}^{n} \xi_t^2 - \sigma^2 = \sum_{t=1}^{n-1} \chi_t r_{n-t} + \sigma^2 \left\{ \sum_{t=1}^{n-1} r_{n-t} - 1 \right\} + \sum_{t=2}^{n} \sum_{r \neq s} \varepsilon_r \varepsilon_s c_{t-r} c_{t-s}, \tag{A.10}$$

with $\chi_t = \varepsilon_t^2 - \sigma^2$. The first term on the right has mean zero and variance

$$\sum_{t=1}^{n-1} \sum_{u=1}^{n-1} \gamma_{t-u} r_{n-t} r_{n-u}. \tag{A.11}$$

Now

$$|\gamma_j| = O\left(j^{2d-1}\right), \quad \text{as } j \to \infty \tag{A.12}$$

by (3.4) and (3.6), and

$$\sum_{t=1}^{n-1} r_t \to 1, \quad \text{as } n \to \infty \tag{A.13}$$

established by Robinson (1995a). It follows from the Toeplitz lemma that (A.11) tends to zero. Clearly, the second term in (A.10) thus tends to zero, whereas the last term has mean zero and variance bounded by

$$2 \left(\max_t E\varepsilon_t^4 \right) \sum_{t,u=2}^{n} \sum_{\substack{r \neq s \\ 1}}^{\min(t-1,u-1)} |c_{t-r} c_{t-s} c_{u-r} c_{u-s}|. \tag{A.14}$$

This follows from the corresponding derivation in Robinson (1995a), but upper bounding $E(\varepsilon_t^2 \varepsilon_s^2)$ by the Schwarz inequality. The absolute value did not arise in Robinson (1995a) but it is clear from his derivation that the bound established there applies to (A.14), namely $O\left((\log m)^4 (n^{-1} + m^{-1/3})\right) \to 0$. It remains to show that

$$\sum_{t=2}^{n} (\sigma_t^2 - \sigma^2) \xi_t^2 \to_p 0. \tag{A.15}$$

The left side is

$$\sigma^2 \sum_{t=2}^{n} (\sigma_t^2 - \sigma^2) r_{t-1} + \sum_{t=2}^{n} (\sigma_t^2 - \sigma^2) \sum_{s=1}^{t-1} c_{t-s}^2 \chi_s + \sum_{t=2}^{n} (\sigma_t^2 - \sigma^2) \sum_{\substack{v \neq s \\ 1}}^{t-1} \varepsilon_v \varepsilon_s c_{t-v} c_{t-s}. \tag{A.16}$$

The first term is

$$\sigma^2 \sum_{t=2}^{n} \sum_{j=1}^{\infty} \psi_j \chi_{t-j} r_{t-1} = \sigma^2 \left(S_1 + S_2 \right),$$

where

$$S_1 = \sum_{j=1-n}^{n-1} \chi_j \sum_{t=1}^{n-1} r_t \psi_{t-j+1}, \quad S_2 = \sum_{j=-\infty}^{-n} \chi_j \sum_{t=1}^{n-1} r_t \psi_{t-j+1},$$

and $\psi_j = 0,\ j \leq 0$. Now S_1 has mean zero and variance

$$\begin{aligned} \sum_{j,k=1-n}^{n-1} \gamma_{j-k} \sum_{s,t=1}^{n-1} r_s r_t \psi_{s-j+1} \psi_{t-k+1} &\leq Knr_{n-1}^2 \Big(\sum_{j=1}^{\infty} |\psi_j|\Big)^2 \sum_{j=0}^{2n-2} |\gamma_j| \\ &= O\left(\frac{(\log m)^8}{n^{1-2d}}\right) \to 0, \quad \text{as } n \to \infty, \end{aligned}$$

using (1.16), (A.12) and $r_n = O\left((\log m)^4/n\right)$, which was established by Robinson (1995a). On the other hand

$$E\,|S_2| \leq K \sum_{t=1}^{n-1} r_t \sum_{j=n}^{\infty} |\psi_j| \to 0, \quad \text{as } n \to \infty,$$

from (1.16) and (A.13), so that the first term in (A.16) is $o_p(1)$. The second term in (A.16) is

$$\sum_{t=2}^{n} \sum_{v=-\infty}^{0} \psi_{t-v} \sum_{s=1}^{t-1} c_{t-s}^2 \chi_v \chi_s \tag{A.17}$$

$$+ \sum_{t=2}^{n} \sum_{v=1}^{t-1} \psi_{t-v} \sum_{s=1}^{t-1} c_{t-s}^2 \chi_v \chi_s. \tag{A.18}$$

The expectation of the absolute value of (A.17) is bounded by

$$K\left(\max_t E\varepsilon_t^4\right) \sum_{t=2}^{n} \sum_{j=t}^{\infty} |\psi_j| r_{t-1} \to 0$$

using (1.16), (A.13) and the Toeplitz lemma. (A.18) includes the component

$$\sum_{t=2}^{n} \sum_{s=1}^{t-1} \psi_{t-s} c_{t-s}^2 \chi_s^2,$$

whose absolute value has expectation which likewise tends to zero. The remainder of (A.18) can be written

$$\sum_{t=2}^{n} \sum_{v=1}^{t-1} \psi_{t-v} \sum_{s=1}^{v-1} c_{t-s}^2 \chi_v \chi_s + \sum_{t=2}^{n} \sum_{v=1}^{t-1} \psi_{t-v} \sum_{s=v+1}^{t-1} c_{t-s}^2 \chi_v \chi_s. \tag{A.19}$$

The first term in (A.19) has mean square

$$\sum_{t,u=2}^{n} \sum_{v=1}^{t-1} \psi_{t-v} \sum_{s=1}^{v-1} c_{t-s}^2 \sum_{q=1}^{u-1} \psi_{u-q} \sum_{p=1}^{q-1} c_{u-p}^2 E\left(\chi_v \chi_s \chi_q \chi_p\right). \tag{A.20}$$

Now each (v,s,q,p) such that $s < v$, $p < q$ satisfies one of the relations $v = q$, $s \le q < v$, $q < s < v$, $p \le v < q$ or $v < p < q$. The contribution from summands in (A.20) such that $v = q$ is bounded by

$$\begin{aligned} & K\left(\max_t E\chi_t^4\right) \sum_{t,u=2}^{n} \sum_{v=1}^{\min(t-1,u-1)} |\psi_{t-v}\psi_{u-v}| \sum_{s=1}^{v-1} c_{t-s}^2 \sum_{p=1}^{v-1} c_{u-p}^2 \\ & \le K\left(\max_t E\varepsilon_t^8\right) r_{n-1}^2 n \Big(\sum_{j=1}^{\infty} |\psi_j|\Big)^2 = O\left((\log m)^4/n\right) \to 0. \end{aligned}$$

Next, for $v > q \ge s$, $p < q$,

$$E\left(\chi_v\chi_s\chi_q\chi_p\right) = E\left\{\sum_{j=-\infty}^{q} \phi_{v-j}\nu_j\chi_s\chi_q\chi_p\right\}, \tag{A.21}$$

because

$$E\left(\chi_v|\mathcal{F}_q\right) = \sum_{j=-\infty}^{q} \phi_{v-j}\nu_j, \quad \text{a.s.}, \quad v > q, \tag{A.22}$$

as follows from (1.10) and

$$E\left(\nu_j|\mathcal{F}_q\right) = E\left(\varepsilon_j^2|\mathcal{F}_q\right) - E\left(E\left(\varepsilon_j^2|\mathcal{F}_{j-1}\right)|\mathcal{F}_q\right) = 0, \quad \text{a.s.}, \quad q < j.$$

Now (A.21) is bounded in absolute value by

$$\begin{aligned} E\left|\Big(\sum_{j=-\infty}^{q} \phi_{v-j}\nu_j\Big)\chi_s\chi_q\chi_p\right| &\le \left\{E\left(\sum_{j=-\infty}^{q} \phi_{v-j}\nu_j\right)^4 \left(\max_t E\chi_t^4\right)^3\right\}^{\frac{1}{4}} \\ &\le K\left\{E\left(\sum_{j=-\infty}^{q} \phi_{v-j}^2\nu_j^2\right)^2\right\}^{\frac{1}{4}} \\ &\le K\Phi_{v-q}^{\frac{1}{2}}\left(\sum_{j=-\infty}^{q} \phi_{v-j}^2 E\left(\nu_j^4\right)\right)^{\frac{1}{4}} \\ &\le K\Phi_{v-q}, \end{aligned}$$

where the second inequality employs Burkholder's (1973) inequality and the final one $E\left(\nu_j^4\right) \le 8\left[E\left(\varepsilon_j^8\right) + E\left\{E\left(\varepsilon_j^2|\mathcal{F}_{j-1}\right)\right\}^4\right] \le K$, by (3.4). Considering similarly the three cases $\{p < q < s < v\}$, $\{p \le v < q \text{ and } s < v\}$ and $\{s < v < p < q\}$, we have

$$|E\left(\chi_v\chi_s\chi_q\chi_p\right)| \le K\left(\Phi_{v-q} + \Phi_{v-s} + \Phi_{q-v} + \Phi_{q-p}\right)$$

whenever $s < v$, $p < q$ and $v \ne q$, where $\Phi_j = 0$ for $j < 0$. Thus the contribution to (A.20) for $v \ne q$ is bounded in absolute value by

$$K\sum_{t,u=2}^{n}\sum_{v=1}^{t-1} |\psi_{t-v}| \sum_{s=1}^{v-1} c_{t-s}^2 \sum_{q=1}^{u-1} |\psi_{u-q}| \sum_{p=1}^{q-1} c_{u-p}^2 \left(\Phi_{v-q} + \Phi_{v-s} + \Phi_{q-v} + \Phi_{q-p}\right)$$

$$\leq \quad K \sum_{t,u=2}^{n} \left\{ \sum_{v=1}^{t-1} \sum_{q=1}^{u-1} |\psi_{t-v}\psi_{u-q}| \Phi_{v-q} \right\} r_{t-1} r_{u-1}$$

$$+ \quad K \sum_{j=1}^{\infty} |\psi_j| \sum_{u=2}^{n} r_{u-1} \sum_{t=2}^{n} \sum_{s=1}^{t-1} c_{t-s}^2 \left\{ \sum_{v=1}^{t-1} |\psi_{t-v}| \Phi_{v-s} \right\}. \tag{A.23}$$

The terms in braces are bounded respectively by

$$\sum_{i,j=0}^{\infty} |\psi_i \psi_{i+j+u-t}| \Phi_j, \quad \sum_{i=1}^{\infty} |\psi_i| \Phi_{t-s-i},$$

which tend to zero as $|u-t| \to \infty$ and $|t-s| \to \infty$ respectively, in view of (1.13) and (1.16) and the Toeplitz lemma. Thus, (1.16), (A.13) and the Toeplitz lemma further imply that (A.23)$\to 0$ as $n \to \infty$, completing the proof that the first term of (A.19) is $o_p(1)$. The second term of (A.19) can be treated in the same way to conclude that (A.18) is $o_p(1)$. The last term of (A.16) is

$$2 \sum_{t=2}^{n} \sum_{j=-\infty}^{t-1} \psi_{t-j} \chi_j \sum_{\substack{v<s \\ 1}}^{t-1} \varepsilon_v \varepsilon_s c_{t-v} c_{t-s}. \tag{A.24}$$

Now, note that

$$E(\chi_j \varepsilon_s \varepsilon_v \chi_k \varepsilon_r \varepsilon_u) = 0, \quad v < s, \ u < r, \ v \neq u \text{ or } s \neq r.$$

This follows by proceeding recursively using (1.6) and nested conditional expectations, and the fact that $E(\varepsilon_t | \mathcal{F}_{t-1})$, $E(\varepsilon_t^3 | \mathcal{F}_{t-1})$, $E(\varepsilon_t^4 \varepsilon_u | \mathcal{F}_{u-1})$, $t \geq u$ and $E(\varepsilon_t^4 \varepsilon_u^2 \varepsilon_v | \mathcal{F}_{v-1})$, $t \geq u \geq v$, are all a.s. zero under A3'. On the other hand, for all indices,

$$|E(\chi_j \varepsilon_s \varepsilon_v \chi_k \varepsilon_r \varepsilon_u)| \leq \max_t E\left(\varepsilon_t^8\right) < \infty$$

by Hölder's inequality. It follows that (A.24) has second moment

$$4 \sum_{t,u=2}^{n} \sum_{j=-\infty}^{t-1} \psi_{t-j} \sum_{k=-\infty}^{u-1} \psi_{u-k} \sum_{\substack{v<s \\ 1}}^{min(t,u)-1} c_{t-v} c_{t-s} c_{u-v} c_{u-s} E\left(\chi_j \chi_k \varepsilon_v^2 \varepsilon_s^2\right)$$

$$\leq K \sum_{t,u=2}^{n} \sum_{\substack{v<s \\ 1}}^{min(t,u)-1} |c_{t-v} c_{t-s} c_{u-v} c_{u-s}| = O\left(\frac{(\log m)^4}{m^{\frac{1}{3}}}\right)$$

as in (A.14), to complete the proof that (A.10) $\to_p 0$ and thus of (A.9).

Application of the remainder of the proof of Robinson (1995a) requires estimation of $U_r - r\sigma^2$ and $V_r - U_r$, where $U_r = 2\pi \sum_{j=1}^{r} J(\lambda_j)$, and $V_r = \sum_{j=1}^{r} I(\lambda_j)/G\lambda_j^{1-2H}$, for $1 \leq r \leq m$. In Robinson (1995a) it is shown that $U_r - r\sigma^2 = O_p(r^{\frac{1}{2}})$, but inspection of the only use that is made of this bound indicates that $O_p(r^{1-\eta})$ would suffice, for any $\eta > 0$. From Robinson (1995a),

$$U_r - r\sigma^2 = \frac{r}{n}\sum_{t=1}^{n}(\varepsilon_t^2 - \sigma^2) + \sum_{t=2}^{n}\varepsilon_t\sum_{s=1}^{t-1}\varepsilon_s d_{t-s}, \tag{A.25}$$

where $d_s = \frac{2}{n}\sum_{j=1}^{r}\cos s\lambda_j$. The first term of (A.25) has mean zero and variance

$$O\left(\frac{r^2}{n}\sum_{j=1}^{n}|\gamma_j|\right) = O\left(r^2 n^{2d-1}\right) = O\left(r^{2(1-\eta)}\frac{r^{2\eta}}{n^{1-2d}}\right),$$

and this is $O\left(r^{2(1-\eta)}\right)$ under (3.8) on taking $\eta \leq \frac{1}{2} - d$. The second term in (A.25) has mean zero and variance

$$E\left\{\sum_{t=2}^{n}\sigma_t^2\sum_{s=1}^{t-1}\varepsilon_s^2 d_{t-s}^2\right\} + E\left\{\sum_{t=2}^{n}\sigma_t^2\sum_{v\neq s}^{t-1}\varepsilon_s\varepsilon_v d_{t-s}d_{t-v}\right\}.$$

The first term is $O_p\left(n(\max_t E\varepsilon_t^6)\sum_{t=1}^{n}d_t^2\right) = O(r)$ from Robinson (1995a), whereas the second term is zero from (3.5). Thus, $U_r - r\sigma^2 = O_p(r^{1-\eta})$, some $\eta > 0$. The bound established for $V_r - U_r$ by Robinson (1995a) was

$$O_p\left(r^{1/3}(\log r)^{2/3} + r^{\beta+1}n^{-\beta} + r^{1/2}n^{-1/4}\right), \tag{A.26}$$

where (3.8) was assumed. Again, this bound is stronger than necessary, and it will suffice to establish the bound (A.26) $+O_p\left(rn^{d-\frac{1}{2}}\right)$. To approximate the scores by a suitable martingale it is sufficient that

$$\sum_{j=1}^{m}\nu_j\left(\frac{I(\lambda_j)}{G\lambda_j^{1-2H}} - \sigma^2 J(\lambda_j)\right) = o_p(m^{\frac{1}{2}}), \tag{A.27}$$

and the left side is, by summation by parts and $|\log r - \log(r+1)| \leq r^{-1}$, bounded by

$$\sum_{r=1}^{m-1}\frac{1}{r}|V_r - U_r| + 2\log m|V_m - U_m|.$$

We can then invoke (3.8) and (3.9) to establish (A.27), if indeed $V_r - U_r =$ (A.26) $+O_p\left(rn^{d-\frac{1}{2}}\right)$. In fact, part of the proof in Robinson (1995a) that $U_r - V_r$ has bound (A.26) continues to hold, but not that relating to the contribution to the variance of (A.25) from fourth cumulants. Under the conditions of Robinson (1995a) that second and fourth conditional moments are constant, $\text{cum}(\varepsilon_r, \varepsilon_s, \varepsilon_t, \varepsilon_u) = \text{cum}(\varepsilon_r, \varepsilon_r, \varepsilon_r, \varepsilon_r)$ if $r = s = t = u$, and zero otherwise. However, under the present asssumptions, we have

$$\begin{aligned} \text{cum}(\varepsilon_r, \varepsilon_s, \varepsilon_t, \varepsilon_u) &= \text{cum}(\varepsilon_r, \varepsilon_r, \varepsilon_r, \varepsilon_r), \quad r = s = t = u, \\ &= \gamma_{r-s}, \quad r = t \neq s = u, && \text{(A.28)} \\ &= \gamma_{r-t}, \quad r = s \neq t = u, && \text{(A.29)} \\ &= \gamma_{r-t}, \quad r = u \neq t = s, && \text{(A.30)} \end{aligned}$$

and zero otherwise. The contributions from (A.28)-(A.30) to the variance of $V_r - U_r$ will thus be studied. In view of (A.28)-(A.30) the contribution of fourth cumulants to the variance of V_r includes terms such as

$$\left(\frac{G}{r}\right)^2 \sum_{j,k}^{r} (\lambda_j \lambda_k)^{2H-1} \sum_{v \neq s} \gamma_{v-s} \alpha_v(\lambda_j) \alpha_s(-\lambda_j) \alpha_s(\lambda_k) \alpha_v(-\lambda_k), \tag{A.31}$$

where $\alpha_v(\lambda) = \sum_{t=1}^{n} \alpha_{t-v} e^{it\lambda}$ and we take $\alpha_t = 0$, $t < 0$. Now $\alpha_v(\lambda)$ is identically zero when $v > n$. On the other hand when $v < 0$ such that $(-v)^{-1} = O(|\lambda|)$ we have by summation by parts, (3.7) and (3.10), that

$$\begin{aligned} |\alpha_v(\lambda)| &\leq \sum_{t=1-v}^{n-v-1} |\alpha_t - \alpha_{t+1}| \left| \sum_{s=1-v}^{t} e^{is\lambda} \right| + |\alpha_{n-v}| \left| \sum_{s=1-v}^{n-v} e^{is\lambda} \right| \\ &\leq K \frac{(1-v)^{H-3/2}}{|\lambda|} = O\Big(|\lambda|^{\frac{1}{2}-H}\Big), \end{aligned}$$

whereas for $v < 1$ such that $-v = O(1/|\lambda|)$

$$|\alpha_v(\lambda)| \leq \sum_{t=1-v}^{1-v+s} |\alpha_v| + \left| \sum_{t=1-v+s}^{n-v} \alpha_t e^{it\lambda} \right| \tag{A.32}$$

for $1 \leq s < n$. Applying summation by parts in the same way as above to the second term of (A.32) indicates that it is $O((1-v+s)^{H-3/2}/|\lambda|)$, while the first term is $O((1-v+s)^{H-1/2})$. Choosing s such that $1 - v + s \sim 1/|\lambda|$ indicates that (A.32) is also $O(|\lambda|^{\frac{1}{2}-H})$. In the same way, it follows that for $1 \leq v \leq n$, $\alpha_v(\lambda) = O(|\lambda|^{\frac{1}{2}-H})$. It immediately follows that (A.31) is $O(r^2 n^{-1} \sum_{j=1}^{n} |\gamma_j|) = O(r^2 n^{2d-1})$ as desired. The other fourth cumulant contributions to the variance of V_r are treated in the same way, and those to the covariance between V_r and U_r and to the variance of U_r follow if anything more easily, to complete the proof that the fourth cumulant contribution to $V_r - U_r$ is $O_p(rn^{d-\frac{1}{2}})$. We have of course not assumed (3.2) in the above, but if we do then $\sum_{j=0}^{\infty} |\gamma_j| < \infty$, so it is easily seen that (A.31) is $O(r^2/n)$, whence (3.9) is not required.

References

Adenstedt, R. (1974): "On large sample estimation for the mean of a stationary random sequence," *Annals of Statistics*, 2, 1095–1107.

Baillie, R., T. Bollerslev, and H. Mikkelsen (1996): "Fractionally integrated generalized autoregressive conditional heteroscedasticity," *Journal of Econometrics*, 74, 3–30.

BOLLERSLEV, T. (1986): "Generalized autoregressive conditional heteroscedasticity," *Journal of Econometrics*, 31, 307–327.

BURKHOLDER, D. (1973): "Distribution function inequalities for martingales," *Annals of Probability*, 1, 19–42.

DAVIS, R. A., and T. MIKOSCH (1997): "The sample ACF of heavy-tailed stationary processes and applications to ARCH," Preprint, Colorado State University.

DING, Z., and C. W. J. GRANGER (1996): "Modelling volatility persistence of speculative returns: a new approach," *Journal of Econometrics*, 23, 185–215.

DING, Z., C. W. J. GRANGER, and R. ENGLE (1993): "A long memory property of stock market returns and a new model," *Journal of Empirical Finance*, 1, 83–106.

ENGLE, R. (1982): "Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation," *Econometrica*, 50, 987–1007.

FOX, R., and M. S. TAQQU (1986): "Large sample properties of parameter estimates for strongly dependent stationary Gaussian time series," *Annals of Statistics*, 14, 517–132.

GEWEKE, J., and S. PORTER-HUDAK (1983): "The estimation and application of long memory time series models," *Journal of Time Series Analysis*, 4, 221–238.

GHYSELS, E., A. HARVEY, and E. RENAULT (1996): "Stochastic Volatility," in *Handbook of Statistics,* Vol. 14. Elsevier, New York.

GIRAITIS, L., P. KOKOSZKA, and R. LEIPUS (1998): "Stationary ARCH models: dependence structure and central limit theorem," Preprint, London School of Economics.

GIRAITIS, L., and D. SURGAILIS (1986): "Multivariate Appel polynomials and the central limit theorem," in *Dependence in Probability and Statistics*, ed. by E. Eberlein, and M. S. Taqqu, pp. 21–71. Birkhäuser, Boston.

GIRAITIS, L., and D. SURGAILIS (1990): "A central limit theorem for quadratic forms in strongly dependent linear variables and its application to asymptotic normality of Whittle's estimate," *Probability Theory and Related Fields*, 86, 87–104.

HENRY, M., and R. PAYNE (1997): "An investigation of long range dependence in intra-day foreign exchange volatility," Discussion Paper 264, Financial Markets Group, London School of Economics.

HENRY, M., and P. M. ROBINSON (1996): "Bandwidth choice in Gaussian semiparametric estimation of long range dependence," in *Athens Conference on Ap-*

plied Probability and Time Series Analysis, Volume II: Time Series Analysis, In Memory of E.J. Hannan,, ed. by P. M. Robinson, and M. Rosenblatt. Springer Verlag, New York, pp. 220-232.

KUERSTEINER, G. (1997): "Efficient IV estimation for autoregressive models with conditional heterogeneity," preprint, Yale University.

KÜNSCH, H. (1987): "Statistical aspects of self-similar processes," in *Proceedings of the First World Congress of the Bernouilli Society*, pp. 67–74. VNU Science Press.

LEE, D., and P. M. ROBINSON (1996): "Semiparametric exploration of long memory in stock prices," *Journal of Statistical Planning and Inference*, 50, 155–174.

LEE, S.-W., and B. E. HANSEN (1994): "Asymptotic theory for the GARCH(1,1) quasi-maximum likelihood estimator," *Econometric Theory*, 10, 29–52.

LO, A. (1991): "Long term memory in stock market prices," *Econometrica*, 59, 1279–1313.

LOBATO, I., and P. M. ROBINSON (1998): "A nonparametric test for I(0)," *Review of Economic Studies*, forthcoming.

LUMSDAINE, R. (1996): "Consistency and asymptotic normality of the quasi-maximum likelihood estimator in IGARCH(1,1) and covariance stationary GARCH(1,1) models," *Econometrica*, 64, 575–596.

MANDELBROT, B. B. (1969): "Long-run linearity, locally Gaussian processes, H-spectra and infinite variances," *International Economic Review*, 10, 82–113.

NELSON, D. B. (1990): "Stationarity and persistence in the GARCH(1,1) model," *Econometric Theory*, 6, 318–334.

NELSON, D. B. (1991): "Conditional heteroscedasticity in asset pricing: a new approach," *Econometrica*, 59, 347–370.

ROBINSON, P. M. (1991): "Testing for strong serial correlation and dynamic conditional heteroscedasticity in multiple regression," *Journal of Econometrics*, 47, 67–84.

ROBINSON, P. M. (1994): "Time series with strong dependence," in *Advances in Econometrics*, ed. by C. Sims, vol. 1, pp. 97–107. Cambridge University Press.

ROBINSON, P. M. (1995a): "Gaussian semiparametric estimation of long range dependence," *Annals of Statistics*, 23, 1630–1661.

ROBINSON, P. M. (1995b): "Log periodogram regression of time series with long range dependence," *Annals of Statistics*, 23, 1048–1072.

Robinson, P. M., and P. Zaffaroni (1997): "Modelling nonlinearity and long memory in time series," in *Nonlinear Dynamics and Time Series*, ed. by C. D. Cutler, and D. T. Kaplan, pp. 161–170. Fields Institute Communications 11.

Robinson, P. M., and P. Zaffaroni (1998): "Nonlinear time series with long memory: a model for stochastic volatility," *Journal of Statistical Planning and Inference*, 68, 359–371.

Taqqu, M. S. (1979): "Convergence of integrated processes of arbitrary Hermite rank," *Zeitschrift für Wahrscheinlichkeits-theorie und Verwandte Gebiete*, 50, 53–83.

Taqqu, M. S., and V. Teverovsky (1995): "Estimators for long range dependence: an empirical study," *Fractals*, 3, 785–798.

Weiss, A. A. (1986): "Asymptotic theory for ARCH models: Estimation and testing," *Econometric Theory*, 2, 107–131.

Whistler, D. E. N. (1990): "Parametric models of daily and intra-daily exchange rate volatility," Ph.D. Thesis, University of London.

Yong, C. (1974): *Asymptotic Behaviour of Trigonometric Series*. Chinese University of Hong Kong.

"Antipersistence": H=.25, $\eta_t \sim$ N(0,1)

TABLE 1.1: Monte Carlo BIASES for the Gaussian semiparametric estimate of long memory applied to an ARFIMA(0, −.25, 0) series with five specified innovation structures.

	n=64			n=128			n=256		
MODEL	m=4	m=8	m=16	m=8	m=16	m=32	m=16	m=32	m=64
IID	0.060	0.014	-0.001	-0.006	-0.011	-0.004	-0.028	-0.017	-0.004
ARCH	0.062	0.010	-0.001	-0.003	-0.016	-0.007	-0.028	-0.016	-0.006
GARCH	0.065	0.020	0.005	-0.004	-0.010	-0.003	-0.026	-0.018	-0.006
LMARCH	0.064	0.012	0.002	-0.001	-0.012	-0.004	-0.022	-0.014	-0.003
VLMARCH	0.064	0.018	0.001	-0.002	-0.010	-0.004	-0.020	-0.013	-0.004
EGARCH	-0.107	-0.054	-0.039	-0.033	-0.012	-0.017	-0.002	-0.002	-0.007

TABLE 1.2: Monte Carlo ROOT MSE for the Gaussian semiparametric estimate of long memory applied to an ARFIMA(0, −.25, 0) series with five specified innovation structures.

	n=64			n=128			n=256		
MODEL	m=4	m=8	m=16	m=8	m=16	m=32	m=16	m=32	m=64
IID	0.34	0.24	0.16	0.23	0.16	0.11	0.16	0.11	0.07
ARCH	0.34	0.23	0.17	0.23	0.16	0.12	0.16	0.11	0.08
GARCH	0.34	0.25	0.19	0.24	0.19	0.14	0.18	0.14	0.11
LMARCH	0.34	0.24	0.17	0.24	0.16	0.12	0.16	0.12	0.08
VLMARCH	0.34	0.25	0.18	0.24	0.17	0.13	0.17	0.13	0.10
EGARCH	0.37	0.26	0.18	0.25	0.17	0.13	0.17	0.11	0.08

TABLE 1.3: 95% COVERAGE PROBABILITIES for the Gaussian semiparametric estimate of long memory applied to an ARFIMA(0, −.25, 0) series with five specified innovation structures.

	n=64			n=128			n=256		
MODEL	m=4	m=8	m=16	m=8	m=16	m=32	m=16	m=32	m=64
IID	0.85	0.90	0.84	0.91	0.84	0.89	0.83	0.88	0.91
ARCH	0.85	0.90	0.82	0.92	0.84	0.85	0.84	0.88	0.85
GARCH	0.84	0.88	0.75	0.90	0.76	0.76	0.77	0.77	0.74
LMARCH	0.84	0.90	0.82	0.91	0.83	0.85	0.82	0.86	0.86
VLMARCH	0.85	0.89	0.79	0.91	0.79	0.80	0.79	0.81	0.80
EGARCH	0.81	0.86	0.80	0.88	0.83	0.84	0.84	0.88	0.86

TABLE 1.4: RELATIVE EFFICIENCY of the log periodogram estimate compared to the Gaussian semiparametric estimate of long memory applied to an ARFIMA(0, −.25, 0) series with five specified innovation structures.

	n=64			n=128			n=256		
MODEL	m=4	m=8	m=16	m=8	m=16	m=32	m=16	m=32	m=64
IID	0.56	0.68	0.73	0.68	0.76	0.78	0.76	0.80	0.78
ARCH	0.57	0.67	0.74	0.67	0.74	0.79	0.75	0.79	0.81
GARCH	0.57	0.67	0.74	0.66	0.74	0.80	0.73	0.80	0.84
LMARCH	0.57	0.68	0.74	0.67	0.75	0.80	0.76	0.81	0.81
VLMARCH	0.56	0.68	0.75	0.67	0.75	0.81	0.75	0.82	0.83
EGARCH	0.56	0.67	0.73	0.67	0.74	0.80	0.75	0.80	0.81

"Short memory": $H = .5$, $\eta_t \sim$ N(0,1)

TABLE 2.1: Monte Carlo BIASES for the Gaussian semiparametric estimate of long memory applied to white noise with five specified error structures.

	n=64			n=128			n=256		
MODEL	m=4	m=8	m=16	m=8	m=16	m=32	m=16	m=32	m=64
IID	-0.035	-0.029	-0.025	-0.027	-0.026	-0.013	-0.020	-0.013	-0.008
ARCH	-0.034	-0.030	-0.021	-0.030	-0.024	-0.016	-0.021	-0.015	-0.009
GARCH	-0.033	-0.034	-0.019	-0.037	-0.022	-0.018	-0.026	-0.019	-0.012
LMARCH	-0.031	-0.034	-0.020	-0.032	-0.021	-0.013	-0.019	-0.011	-0.009
VLMARCH	-0.032	-0.032	-0.025	-0.033	-0.024	-0.016	-0.022	-0.015	-0.007
EGARCH	-0.030	-0.036	-0.031	-0.031	-0.025	-0.020	-0.018	-0.015	-0.010

TABLE 2.2: Monte Carlo ROOT MSE for the Gaussian semiparametric estimate of long memory applied to white noise with five specified error structures.

	n=64			n=128			n=256		
MODEL	m=4	m=8	m=16	m=8	m=16	m=32	m=16	m=32	m=64
IID	0.37	0.27	0.18	0.27	0.18	0.11	0.18	0.11	0.07
ARCH	0.36	0.27	0.19	0.27	0.18	0.13	0.17	0.11	0.08
GARCH	0.36	0.29	0.21	0.28	0.20	0.15	0.20	0.15	0.11
LMARCH	0.37	0.28	0.19	0.27	0.18	0.12	0.18	0.12	0.08
VLMARCH	0.37	0.28	0.20	0.28	0.19	0.13	0.19	0.13	0.10
EGARCH	0.36	0.27	0.19	0.27	0.18	0.13	0.17	0.11	0.09

TABLE 2.3: 95% COVERAGE PROBABILITIES for the Gaussian semiparametric estimate of long memory applied to white noise with five specified error structures.

	n=64			n=128			n=256		
MODEL	m=4	m=8	m=16	m=8	m=16	m=32	m=16	m=32	m=64
IID	0.63	0.76	0.84	0.77	0.84	0.89	0.83	0.88	0.92
ARCH	0.65	0.77	0.81	0.77	0.83	0.84	0.85	0.88	0.86
GARCH	0.65	0.72	0.76	0.74	0.77	0.75	0.79	0.77	0.74
LMARCH	0.64	0.75	0.81	0.76	0.82	0.85	0.82	0.86	0.87
VLMARCH	0.64	0.75	0.79	0.75	0.80	0.81	0.80	0.81	0.81
EGARCH	0.65	0.77	0.80	0.78	0.84	0.84	0.85	0.88	0.86

TABLE 2.4: RELATIVE EFFICIENCY of the log periodogram estimate compared to the Gaussian semiparametric estimate of long memory applied to white noise with five specified error structures.

	n=64			n=128			n=256		
MODEL	m=4	m=8	m=16	m=8	m=16	m=32	m=16	m=32	m=64
IID	0.60	0.78	0.82	0.78	0.84	0.80	0.84	0.82	0.77
ARCH	0.60	0.77	0.80	0.78	0.83	0.82	0.83	0.82	0.82
GARCH	0.60	0.76	0.81	0.77	0.84	0.84	0.84	0.86	0.85
LMARCH	0.60	0.78	0.82	0.78	0.84	0.82	0.84	0.83	0.81
VLMARCH	0.60	0.76	0.82	0.78	0.83	0.83	0.84	0.85	0.84
EGARCH	0.61	0.78	0.82	0.79	0.83	0.82	0.83	0.81	0.82

"Moderate long memory": $H = .75$, $\eta_t \sim$ N(0,1)

TABLE 3.1: Monte Carlo BIASES for the Gaussian semiparametric estimate of long memory applied to an ARFIMA(0, .25, 0) series with five specified innovation structures.

	n=64			n=128			n=256		
MODEL	m=4	m=8	m=16	m=8	m=16	m=32	m=16	m=32	m=64
IID	-0.108	-0.050	-0.027	-0.040	-0.012	-0.010	-0.004	0.001	-0.007
ARCH	-0.112	-0.053	-0.031	-0.035	-0.014	-0.015	-0.003	-0.004	-0.005
GARCH	-0.113	-0.057	-0.033	-0.043	-0.020	-0.020	-0.014	-0.007	-0.006
LMARCH	-0.110	-0.051	-0.026	-0.038	-0.013	-0.011	-0.005	0.001	-0.006
VLMARCH	-0.104	-0.052	-0.034	-0.044	-0.015	-0.010	-0.005	-0.004	-0.006
EGARCH	-0.107	-0.054	-0.039	-0.033	-0.012	-0.017	-0.002	-0.002	-0.007

TABLE 3.2: Monte Carlo ROOT MSE for the Gaussian semiparametric estimate of long memory applied to an ARFIMA(0, .25, 0) series with five specified innovation structures.

	n=64			n=128			n=256		
MODEL	m=4	m=8	m=16	m=8	m=16	m=32	m=16	m=32	m=64
IID	0.38	0.26	0.17	0.26	0.17	0.11	0.17	0.11	0.07
ARCH	0.37	0.26	0.18	0.25	0.17	0.12	0.16	0.11	0.08
GARCH	0.37	0.28	0.20	0.27	0.20	0.15	0.19	0.14	0.11
LMARCH	0.38	0.27	0.18	0.26	0.17	0.12	0.17	0.12	0.08
VLMARCH	0.37	0.27	0.19	0.27	0.18	0.13	0.18	0.13	0.10
EGARCH	0.37	0.26	0.18	0.25	0.17	0.12	0.17	0.11	0.08

TABLE 3.3: 95% COVERAGE PROBABILITIES for the Gaussian semiparametric estimate of long memory applied to an ARFIMA(0, .25, 0) series with five specified innovation structures.

	n=64			n=128			n=256		
MODEL	m=4	m=8	m=16	m=8	m=16	m=32	m=16	m=32	m=64
IID	0.80	0.86	0.83	0.87	0.84	0.88	0.84	0.89	0.91
ARCH	0.81	0.86	0.80	0.88	0.84	0.85	0.85	0.88	0.86
GARCH	0.80	0.84	0.75	0.86	0.76	0.76	0.79	0.77	0.75
LMARCH	0.80	0.85	0.81	0.87	0.83	0.85	0.82	0.86	0.87
VLMARCH	0.80	0.85	0.79	0.86	0.80	0.81	0.80	0.82	0.81
EGARCH	0.81	0.86	0.80	0.88	0.83	0.84	0.84	0.88	0.86

TABLE 3.4: RELATIVE EFFICIENCY of the log periodogram estimate compared to the Gaussian semiparametric estimate of long memory applied to an ARFIMA(0, .25, 0) series with five specified innovation structures.

	n=64			n=128			n=256		
MODEL	m=4	m=8	m=16	m=8	m=16	m=32	m=16	m=32	m=64
IID	0.61	0.75	0.79	0.74	0.78	0.79	0.79	0.81	0.79
ARCH	0.62	0.75	0.78	0.74	0.79	0.80	0.78	0.82	0.80
GARCH	0.60	0.74	0.79	0.74	0.79	0.81	0.80	0.82	0.83
LMARCH	0.61	0.76	0.78	0.74	0.80	0.80	0.79	0.81	0.81
VLMARCH	0.61	0.75	0.80	0.74	0.79	0.81	0.79	0.82	0.81
EGARCH	0.61	0.75	0.80	0.74	0.79	0.81	0.78	0.80	0.80

"Very long memory": $H = .95$, $\eta_t \sim$ N(0,1)

TABLE 4.1: Monte Carlo BIASES for the Gaussian semiparametric estimate of long memory applied to an ARFIMA(0, .45, 0) series with five specified innovation structures.

	n=64			n=128			n=256		
MODEL	m=4	m=8	m=16	m=8	m=16	m=32	m=16	m=32	m=64
IID	-0.201	-0.102	-0.059	-0.087	-0.044	-0.027	-0.035	-0.015	-0.013
ARCH	-0.190	-0.107	-0.070	-0.085	-0.047	-0.033	-0.034	-0.017	-0.018
GARCH	-0.210	-0.132	-0.088	-0.110	-0.073	-0.053	-0.060	-0.043	-0.037
LMARCH	-0.210	-0.117	-0.076	-0.101	-0.060	-0.043	-0.052	-0.030	-0.024
VLMARCH	-0.218	-0.121	-0.081	-0.112	-0.064	-0.047	-0.056	-0.037	-0.032
EGARCH	-0.187	-0.105	-0.070	-0.084	-0.046	-0.034	-0.034	-0.017	-0.017

TABLE 4.2: Monte Carlo ROOT MSE for the Gaussian semiparametric estimate of long memory applied to an ARFIMA(0, .45, 0) series with five specified innovation structures.

	n=64			n=128			n=256		
MODEL	m=4	m=8	m=16	m=8	m=16	m=32	m=16	m=32	m=64
IID	0.38	0.23	0.14	0.22	0.13	0.09	0.12	0.08	0.06
ARCH	0.37	0.23	0.16	0.21	0.14	0.10	0.12	0.08	0.07
GARCH	0.38	0.25	0.17	0.23	0.15	0.11	0.14	0.10	0.08
LMARCH	0.38	0.23	0.15	0.21	0.13	0.09	0.13	0.08	0.06
VLMARCH	0.38	0.24	0.16	0.22	0.14	0.10	0.13	0.09	0.07
EGARCH	0.37	0.23	0.16	0.21	0.13	0.10	0.12	0.08	0.07

TABLE 4.3: 95% COVERAGE PROBABILITIES for the Gaussian semiparametric estimate of long memory applied to an ARFIMA(0, .45, 0) series with five specified innovation structures.

	n=64			n=128			n=256		
MODEL	m=4	m=8	m=16	m=8	m=16	m=32	m=16	m=32	m=64
IID	0.80	0.86	0.89	0.88	0.91	0.93	0.93	0.94	0.95
ARCH	0.81	0.86	0.87	0.89	0.91	0.91	0.93	0.94	0.92
GARCH	0.81	0.85	0.85	0.87	0.88	0.87	0.90	0.90	0.86
LMARCH	0.81	0.86	0.88	0.89	0.91	0.91	0.92	0.94	0.93
VLMARCH	0.80	0.86	0.87	0.87	0.90	0.89	0.91	0.92	0.89
EGARCH	0.82	0.87	0.87	0.89	0.92	0.91	0.93	0.94	0.92

TABLE 4.4: RELATIVE EFFICIENCY of the log periodogram estimate compared to the Gaussian semiparametric estimate of long memory applied to an ARFIMA(0, .45, 0) series with five specified innovation structures.

	n=64			n=128			n=256		
MODEL	m=4	m=8	m=16	m=8	m=16	m=32	m=16	m=32	m=64
IID	0.61	0.65	0.64	0.62	0.61	0.63	0.57	0.59	0.65
ARCH	0.59	0.67	0.67	0.62	0.62	0.63	0.57	0.59	0.66
GARCH	0.62	0.67	0.65	0.63	0.61	0.61	0.57	0.57	0.60
LMARCH	0.61	0.65	0.64	0.61	0.59	0.61	0.56	0.54	0.59
VLMARCH	0.62	0.65	0.65	0.61	0.60	0.60	0.56	0.56	0.62
EGARCH	0.61	0.67	0.68	0.61	0.61	0.64	0.57	0.58	0.66

“Short memory”: $H = .5$, $\eta_t \sim t_4$

TABLE 5.1: Monte Carlo BIASES for the Gaussian semiparametric estimate of long memory applied to white noise with five specified error structures.

	n=64			n=128			n=256		
MODEL	m=4	m=8	m=16	m=8	m=16	m=32	m=16	m=32	m=64
IID	-0.028	-0.031	-0.020	-0.026	-0.022	-0.011	-0.021	-0.011	-0.005
ARCH	-0.033	-0.041	-0.035	-0.028	-0.030	-0.022	-0.025	-0.020	-0.019
GARCH	-0.041	-0.043	-0.027	-0.042	-0.037	-0.027	-0.043	-0.029	-0.024
LMARCH	-0.035	-0.030	-0.027	-0.031	-0.023	-0.016	-0.021	-0.022	-0.013
VLMARCH	-0.031	-0.036	-0.028	-0.029	-0.029	-0.019	-0.030	-0.024	-0.019

TABLE 5.2: Monte Carlo ROOT MSE for the Gaussian semiparametric estimate of long memory applied to white noise with five specified error structures.

	n=64			n=128			n=256		
MODEL	m=4	m=8	m=16	m=8	m=16	m=32	m=16	m=32	m=64
IID	0.37	0.27	0.17	0.28	0.17	0.11	0.17	0.11	0.07
ARCH	0.35	0.26	0.21	0.25	0.18	0.16	0.17	0.13	0.13
GARCH	0.36	0.30	0.24	0.30	0.25	0.21	0.26	0.22	0.18
LMARCH	0.36	0.28	0.20	0.28	0.20	0.15	0.22	0.16	0.11
VLMARCH	0.36	0.29	0.22	0.29	0.22	0.17	0.24	0.19	0.15

TABLE 5.3: 95% COVERAGE PROBABILITIES for the Gaussian semiparametric estimate of long memory applied to white noise with five specified error structures.

	n=64			n=128			n=256		
MODEL	m=4	m=8	m=16	m=8	m=16	m=32	m=16	m=32	m=64
IID	0.64	0.76	0.85	0.77	0.85	0.89	0.85	0.89	0.91
ARCH	0.69	0.78	0.76	0.80	0.82	0.76	0.86	0.84	0.72
GARCH	0.66	0.68	0.69	0.69	0.65	0.61	0.63	0.58	0.53
LMARCH	0.65	0.74	0.78	0.73	0.78	0.77	0.74	0.75	0.74
VLMARCH	0.65	0.72	0.74	0.72	0.72	0.72	0.69	0.67	0.64

TABLE 5.4: RELATIVE EFFICIENCY of the log periodogram estimate compared to the Gaussian semiparametric estimate of long memory applied to white noise with five specified error structures.

	n=64			n=128			n=256		
MODEL	m=4	m=8	m=16	m=8	m=16	m=32	m=16	m=32	m=64
IID	0.60	0.77	0.80	0.78	0.81	0.78	0.83	0.81	0.77
ARCH	0.61	0.77	0.80	0.78	0.83	0.82	0.82	0.83	0.83
GARCH	0.60	0.74	0.81	0.74	0.82	0.85	0.81	0.85	0.88
LMARCH	0.60	0.77	0.82	0.77	0.84	0.84	0.84	0.86	0.85
VLMARCH	0.60	0.76	0.83	0.76	0.83	0.85	0.83	0.87	0.86

"Short memory": $H = .5,\ \eta_t \sim t_2$

TABLE 6.1: Monte Carlo BIASES for the Gaussian semiparametric estimate of long memory applied to white noise with five specified error structures.

	n=64			n=128			n=256		
MODEL	m=4	m=8	m=16	m=8	m=16	m=32	m=16	m=32	m=64
IID	-0.018	-0.027	-0.019	-0.024	-0.018	-0.010	-0.017	-0.009	-0.006
ARCH	-0.043	-0.047	-0.042	-0.042	-0.039	-0.037	-0.036	-0.032	-0.034
GARCH	-0.047	-0.042	-0.035	-0.051	-0.048	-0.040	-0.055	-0.047	-0.038
LMARCH	-0.036	-0.038	-0.032	-0.040	-0.034	-0.028	-0.047	-0.038	-0.028
VLMARCH	-0.042	-0.036	-0.037	-0.052	-0.043	-0.037	-0.054	-0.048	-0.037

TABLE 6.2: Monte Carlo ROOT MSE for the Gaussian semiparametric estimate of long memory applied to white noise with five specified error structures.

	n=64			n=128			n=256		
MODEL	m=4	m=8	m=16	m=8	m=16	m=32	m=16	m=32	m=64
IID	0.35	0.25	0.16	0.26	0.16	0.10	0.16	0.10	0.07
ARCH	0.33	0.26	0.23	0.24	0.21	0.20	0.17	0.17	0.19
GARCH	0.35	0.31	0.26	0.31	0.28	0.24	0.28	0.26	0.23
LMARCH	0.36	0.29	0.23	0.30	0.25	0.21	0.28	0.24	0.20
VLMARCH	0.35	0.30	0.25	0.31	0.27	0.23	0.29	0.26	0.22

TABLE 6.3: 95% COVERAGE PROBABILITIES for the Gaussian semiparametric estimate of long memory applied to white noise with five specified error structures.

	n=64			n=128			n=256		
MODEL	m=4	m=8	m=16	m=8	m=16	m=32	m=16	m=32	m=64
IID	0.68	0.81	0.87	0.80	0.88	0.91	0.87	0.91	0.93
ARCH	0.74	0.78	0.71	0.83	0.78	0.65	0.86	0.76	0.56
GARCH	0.68	0.67	0.62	0.66	0.57	0.53	0.59	0.50	0.42
LMARCH	0.65	0.71	0.70	0.67	0.65	0.62	0.59	0.55	0.50
VLMARCH	0.68	0.68	0.65	0.66	0.61	0.56	0.58	0.50	0.45

TABLE 6.4: RELATIVE EFFICIENCY of the log periodogram estimate compared to the Gaussian semiparametric estimate of long memory applied to white noise with five specified error structures.

	n=64			n=128			n=256		
MODEL	m=4	m=8	m=16	m=8	m=16	m=32	m=16	m=32	m=64
IID	0.61	0.76	0.79	0.78	0.80	0.78	0.83	0.80	0.77
ARCH	0.64	0.75	0.79	0.77	0.81	0.81	0.81	0.83	0.83
GARCH	0.62	0.73	0.78	0.73	0.80	0.83	0.78	0.82	0.84
LMARCH	0.60	0.74	0.80	0.74	0.81	0.84	0.80	0.85	0.86
VLMARCH	0.61	0.73	0.80	0.73	0.80	0.84	0.78	0.83	0.85