Whittle estimation of ARCH models

Liudas Giraitis & Peter M. Robinson

This is an electronic version of an Article published in Econometric theory, 17 (3). pp. 608-631 © 2001 Cambridge University Press.
http://uk.cambridge.org/journals/ect/

Econometric Theory, **17**, 2001, 608–631. Printed in the United States of America.

WHITTLE ESTIMATION OF ARCH MODELS

LIUDAS GIRAITIS AND PETER M. ROBINSON
London School of Economics

For a class of parametric ARCH models, Whittle estimation based on squared observations is shown to be $\sqrt{n}$-consistent and asymptotically normal. Our conditions require the squares to have short memory autocorrelation, by comparison with the work of Zaffaroni (1999, "Gaussian Inference on Certain Long-Range Dependent Volatility Models," Preprint), who established the same properties on the basis of an alternative class of models with martingale difference levels and long memory autocorrelated squares.

1. INTRODUCTION

Conditional heteroskedasticity arises in much analysis of economic and financial time series data. Even series that appear not to be autocorrelated may exhibit dependence in their squares, a notable example being daily asset returns. For a covariance stationary process, x_t, $t = 0, \pm 1, \ldots$, suppose that, almost surely,

$$E(x_t | \mathcal{F}_{t-1}) = 0, \tag{1.1}$$

$$h_t = E(y_t | \mathcal{F}_{t-1}) = \psi_0 + \sum_{j=1}^{\infty} \psi_j y_{t-j}, \tag{1.2}$$

where

$$y_t = x_t^2 \tag{1.3}$$

and $\mathcal{F}_t$ is the σ-field of events generated by x_s, $s \le t$. The requirement $\psi_0 > 0, \psi_j \ge 0, j \ge 1$, ensures positivity of the conditional variance h_t, whereas convergence conditions on the ψ_j will be imposed in the sequel. The x_t are observable in some applications, whereas in others they could be innovations in a time series model or regression errors.

In case $\psi_j \neq 0$ for some $j > 0$, we say that x_t has autoregressive conditional heteroskedasticity (ARCH). The original ARCH process is the ARCH(p) proposed by Engle (1982), wherein for known p, $\psi_j = 0$ for all $j > p$. Bollerslev (1986) proposes the more general GARCH(p,q) process in which

This research was supported by ESRC Grant R000235892. The second author's work was also supported by a Leverhulme Trust Personal Research Professorship. We thank the referee for a careful reading. Address correspondence to: Peter M. Robinson, Department of Economics, London School of Economics, London WC2A 2AE, United Kingdom.

  0266-4666/01 $9.50

$$h_t = \omega + \sum_{j=1}^{p} \alpha_j y_{t-j} + \sum_{j=1}^{q} \beta_j h_{t-j}. \tag{1.4}$$

Formally, h_t generated by (1.4) is seen to be a special case of (1.2), with $\psi_0 = \omega/(1-\beta(1))$, and, for $j > 0$, ψ_j is the coefficient of z^j in the expansion of $\alpha(z)/(1-\beta(z))$, where

$$\alpha(z) = \sum_{j=1}^{p} \alpha_j z^j, \qquad \beta(z) = \sum_{j=1}^{q} \beta_j z^j. \tag{1.5}$$

In the literature the term *ARCH* is not now restricted to h_t that are quadratic in x_t, as in (1.2) and (1.4), but applies also to the wide variety of other nonlinear forms that have been found to be of interest; further information can be found in several reviews of the subject, for example, Bollerslev, Chou, and Kroner (1992). Nevertheless, Engle's ARCH(p) and Bollerslev's GARCH(p,q) have attracted considerable theoretical attention, notably Nelson's (1990a) demonstration of convergence to diffusion process used in the option pricing literature, in addition to being featured in countless empirical studies, and the present paper focuses on the quadratic ARCH model (1.2) and its special cases.

The general "ARCH(∞)" form (1.2) is considered by Robinson (1991) in a hypothesis testing context. Following Engle's (1982) and Weiss's (1986) Lagrange multiplier (LM) tests of no-ARCH against ARCH(p) alternatives, Robinson (1991) justifies the asymptotic validity of χ^2 LM tests of no-ARCH against arbitrary finite parameterizations of the ψ_j in (1.2), where, for some explicitly or implicitly defined functions $\psi_j(\theta)$, $j \geq 1$, of a $p \times 1$ column vector θ, we have $\psi_j(\theta_0) = \psi_j$, $j \geq 1$, for some unknown $\theta_0 \in \mathbf{R}^p$. Robinson (1991) also justifies joint tests of no-autocorrelation in x_t and no-ARCH in this context, in addition to tests of no-autocorrelation in x_t (cf. (1.1)) that are robustified to allow for the presence of general conditional heteroskedasticity as represented by (1.2), without parameterizing the ψ_j. On the other hand, Robinson and Henry (1999) have found circumstances when robustification is unnecessary: when the x_t are innovations of a possibly long memory series they showed that a certain semiparametric estimate of the memory parameter of the latter can have the same limiting distribution under (1.1) and (1.2) as when x_t has constant conditional variance. Giraitis, Kokoszka, and Leipus (2000) have derived sufficient conditions for the existence of a stationary solution of (1.2) when the ψ_j are constrained to be non-negative, under which they also established a central limit theorem for partial sums of y_t. Their conditions effectively require y_t to have short memory autocorrelation.

None of these papers discusses parameter estimation in the setup described in the previous paragraph. However, the maximum likelihood estimate (MLE) based on the assumption of conditionally Gaussian x_t, which was considered by Engle (1982) and Bollerslev (1986) for the ARCH(p) and GARCH(p,q) models, extends readily to (1.2). Given observations x_t, $t = 1, \ldots, n$, the log-likelihood is, apart from an additive constant, approximately

$$\ell_n(\theta,\psi) = -\frac{1}{2}\sum_{t=1}^{n}\left\{\log h_t^*(\theta,\psi) + \frac{y_t}{h_t^*(\theta,\psi)}\right\}, \tag{1.6}$$

where

$$h_t^*(\theta,\psi) = \psi + \sum_{j=1}^{t-1}\psi_j(\theta)y_{t-j} \tag{1.7}$$

and ψ is any admissible value of ψ_0. We describe (1.6) as only approximate because $h_t^*(\theta,\psi)$ is not equivalent to $E_{\theta,\psi}(y_t|x_{t-1},\dots,x_1)$; other conventions can be used, which effectively correspond to different proxies for the unobservable x_t, $t \le 0$, and given suitably rapid decay of the ψ_j numerical differences should be slight for large n. In fact (1.6) was the basis for the ARCH(∞) LM tests of Robinson (1991).

The MLE of θ_0,ψ_0 is given by

$$\tilde{\theta},\tilde{\psi} = \arg\max \ell_n(\theta,\psi), \tag{1.8}$$

where the optimization is carried out over a suitable subset of $\mathbf{R}^{p+1}$. To conduct inference, the limiting distribution of $\tilde{\theta},\tilde{\psi}$ is of interest. Weiss (1986) shows that $\tilde{\theta},\tilde{\psi}$ is $\sqrt{n}$-consistent and asymptotically normal in the case of the ARCH(p) model for finite p, whereas Lee and Hansen (1994) and Lumsdaine (1996) establish the same properties in the case of the GARCH(1,1), where $p = q = 1$ is known a priori in (1.4). The asymptotic theory of these authors makes significantly weaker assumptions than the conditional Gaussianity motivating $\tilde{\theta},\tilde{\psi}$, so that $\ell_n(\theta,\psi)$ is viewed as a quasi-log-likelihood. Unfortunately, the analysis becomes considerably more complicated in the GARCH(p,q) model (1.4) for general p and q, and no corresponding results seem yet to be available here, let alone for other parameterizations of the ARCH(∞) (1.2). Bollerslev and Wooldridge (1993) derive the limit distribution in general models under high-level conditions but do not verify these for the GARCH(p,q). Fortunately, the GARCH(1,1) model (and the IGARCH(1,1), where $\alpha_1 + \beta_1 = 1$ in (1.4)), also covered in the asymptotics of Lee and Hansen (1994) and Lumsdaine (1996), have themselves proved useful in modeling a variety of data series. On the other hand these simple models will not always suffice, and one would like an asymptotic theory of inference that covers not only the general GARCH(p,q) (1.4) but also other parameterizations of (1.2), in particular ones that permit greater persistence than (1.4). Under (1.4), y_t has autocovariances that decay exponentially (see Bollerslev, 1986), but there is empirical evidence of sample autocovariances that decay more slowly (see, e.g., Ding, Granger, and Engle, 1993), and it is possible to choose ψ_j in (1.2) to describe only a power law decay, for example.

In such models, other methods of estimation may afford an easier asymptotic theory. In particular, because a principal stylized fact motivating models for conditional heteroskedasticity is the autocorrelation in squares y_t, a fairly natural approach matches theoretical and sample second moments of the y_t, in the same way as if one were dealing with a linear autocorrelated series. This prompts consideration of Gaussian or Whittle estimation based on the y_t, an idea that is far from new in relation to processes with conditional heteroskedasticity. Harvey (1998) and Robinson and Zaffaroni (1997, 1998) employ it for certain stochastic volatility and nonlinear moving average processes, whereas Zaffaroni (1999) has established consistency and asymptotic normality of Whittle estimates in the latter case. Indeed the idea is not new in the GARCH case, especially as Bollerslev (1986) points out that y_t generated by (1.4) have an $\mathrm{ARMA}(\max(p,q),q)$ representation, albeit with conditionally heteroskedastic innovations.

To fix ideas, rewrite (1.2) as

$$y_t = \psi_0 + \sum_{t=1}^{\infty} \psi_j y_{t-j} + \nu_t, \tag{1.9}$$

where $\nu_t = y_t - h_t$ are martingale differences. Assuming x_t is a fourth-order stationary sequence (for which conditions are given subsequently), y_t has spectral density

$$f(\lambda) = \frac{\sigma^2}{2\pi} g(\lambda), \qquad -\pi < \lambda \le \pi, \tag{1.10}$$

where

$$g(\lambda) = \left|1 - \sum_{j=1}^{\infty} \psi_j e^{ij\lambda}\right|^{-2} \tag{1.11}$$

and

$$\sigma^2 = E(\nu_t^2) = E(x_t^4) - E(h_t^2). \tag{1.12}$$

Notice that $E(\nu_t^2|\mathcal{F}_{t-1}) = E(x_t^4|\mathcal{F}_{t-1}) - h_t^2 \neq \sigma^2$, so the ν_t do not behave like an independent sequence up to second moments. Nevertheless we can consider Whittle-type procedures originally designed for processes with the latter desirable property.

Consider the objective function

$$w_n(\theta) = \sum_{j=1}^{n-1} \frac{I(\lambda_j)}{g(\lambda_j;\theta)}, \tag{1.13}$$

where $I(\lambda)$ is the periodogram of the y_t,

$$I(\lambda) = \frac{1}{2\pi n}\left|\sum_{t=1}^{n} y_t e^{it\lambda}\right|^2, \tag{1.14}$$

$\lambda_j = 2\pi j/n$, and (cf. (1.11))

$$g(\lambda;\theta) = \left|1 - \sum_{j=1}^{\infty} \psi_j(\theta) e^{ij\lambda}\right|^{-2}. \tag{1.15}$$

Then we define the estimate

$$\hat{\theta} = \arg\min_{\Theta} w_n(\theta), \tag{1.16}$$

where Θ is a compact subset of $\mathbf{R}^p$. The discrete frequency form is preferred over others such as the continuous form and the actual Gaussian likelihood, as a result of the direct use it makes of the fast Fourier transform and of $g(\lambda;\theta)$, which is usually explicitly specified, for example in the ARCH(p) and GARCH(p,q) models, where, following Bollerslev (1986), we have from (1.4) and (1.5)

$$g(\lambda) = \frac{|\alpha(e^{i\lambda}) - \beta(e^{i\lambda})|^2}{|1 - \beta(e^{i\lambda})|^2}. \tag{1.17}$$

Another feature of the discrete frequency form (1.13) is that mean-correction of y_t is taken care of by omission of summands $j = 0$ (and n).

Asymptotic theory for various Whittle forms has been given by Hannan (1973), Dzhaparidze (1974), and various subsequent authors, from the 1970's onward. Although the techniques used by these authors are relevant to our setting, the central limit theorem for quadratic forms (e.g., sums of finitely many sample autocovariances) that is involved in the proof of asymptotic normality has not previously been checked in the case of squares of ARCH sequences. Like Hannan (1973) and others, we require y_t to have short memory autocorrelation, but in our case it cannot be linear in conditionally homoskedastic martingale differences nor is it known to satisfy suitable mixing conditions, so that a direct proof of asymptotic normality of quadratic forms of ARCH squares is provided. The main results are presented, with discussion, in the following section, with the bulk of the proof left to Section 3.

It is important to point out the drawbacks of Whittle estimation in an ARCH setting. The term $\hat{\theta}$ has a different limiting variance from $\tilde{\theta}$, in view of the work of Lee and Hansen (1994) and Lumsdaine (1996), so that at least when the x_t are conditionally Gaussian it is asymptotically less efficient than $\tilde{\theta}$. Moreover, whereas in the context of Hannan (1973) the y_t can be Gaussian, so that $\hat{\theta}$ has the same limit distribution as the Gaussian MLE, it is impossible for our squares y_t to be Gaussian. Therefore the objective function $w_n(\theta)$ cannot possibly approximate the actual log-likelihood for any conceivable distribution of the x_t, and so in no circumstances can $\hat{\theta}$ be asymptotically efficient. As a related point, the limiting covariance matrix of $\hat{\theta}$ is considerably more complicated in our setting than both that of $\tilde{\theta}$, (1.8), and of $\hat{\theta}$ in the setting of Hannan (1973), essentially as a result of the conditional heteroskedasticity in the innovations ν_t. Moreover, Whittle estimation based on the squares y_t is less well motivated in our ARCH models than in the stochastic volatility and nonlinear moving average models considered by Harvey (1998) and Robinson and Zaffaroni (1997, 1998), because in their cases the actual likelihood, under any parent innovation distribution, is relatively intractable computationally, let alone theoretically, whereas the MLE $\tilde{\theta}$ for (1.4) is relatively easy to compute. Moreover, Harvey (1998) and Robinson and Zaffaroni (1997, 1998) envisage long memory in the squares, when Whittle estimation has the desirable feature of compensating for possible lack of square integrability of the spectrum, so as to produce $\sqrt{n}$-consistency and asymptotic normality. Our asymptotics only handles short memory in the y_t, and so Whittle estimation plays a less special role: a variety of estimates, including simple method of moment estimates in the GARCH(p,q) case, can be $\sqrt{n}$-consistent and asymptotically normal, and indeed over part of the parameter space they could even be more efficient than $\hat{\theta}$. As a final drawback, we require finiteness of at least eighth unconditional moments of x_t, unlike in the work by Lee and Hansen (1994) and Lumsdaine (1996) on $\tilde{\theta}$, whereas a body of opinion believes that fourth moments are infinite in much financial data. These considerations may well restrict practical interest in $\hat{\theta}$, and certainly we can identify no circumstances in which it might be preferred on theoretical grounds to $\tilde{\theta}$ in the case of ARCH(p) and GARCH(1,1) models, where rigorous asymptotic theory for $\tilde{\theta}$ is available, as indeed it is for adaptive estimates (see Linton, 1993; Drost and Klassen, 1997). However, at least until such theory can be extended to the general GARCH(p,q) and other cases of (1.2), it is to be hoped that our study of $\hat{\theta}$ will fill some gap and add to our knowledge of the performance of Whittle estimation in nonstandard situations.

2. MAIN RESULTS

We introduce first an assumption, one version of which $(J = 4)$ will be employed in our proof of consistency of $\hat{\theta}$ and another, stronger version $(J = 8)$ in our proof of asymptotic normality.

Assumption 1(J). For $t = 0, \pm 1, \dots,$

$$x_t = h_t^{1/2} \varepsilon_t, \tag{2.1}$$

where the ε_t are strictly stationary and ergodic with finite Jth moment and, almost surely,

$$E(\varepsilon_t | \mathcal{F}_{t-1}) = 0, \tag{2.2}$$

$$E(\varepsilon_t^2 | \mathcal{F}_{t-1}) = 1, \tag{2.3}$$

$$E(\varepsilon_t^{2j} | \mathcal{F}_{t-1}) = \mu_{2j}, \qquad j = 2, \dots, J/2. \tag{2.4}$$

for constants μ_{2j}, whereas h_t is given by (1.2) with

$$\psi_0 > 0, \qquad \psi_j \geq 0, \qquad j \geq 1, \tag{2.5}$$

$$|\mu_J|^{2/J} \sum_{j=1}^{\infty} \psi_j < 1. \tag{2.6}$$

Properties (2.1)–(2.3) imply the conditional moment restrictions (1.1) and (1.2). With (2.4), they indicate that ε_t behaves like an independent and identically distributed (i.i.d.) sequence up to Jth moments. Property (2.5) implies $h_t > 0$, as earlier noted, whereas, when $J = 4$, (2.5) is sufficient for (2.1) to have a unique covariance stationary solution for y_t, in terms of ε_s, $s \leq t$, by a slight extension of the argument of Giraitis et al. (2000). It also follows from Assumption 1(4), as in Giraitis et al. (2000), that, defining $\gamma(j) = \text{Cov}(y_0, y_j)$,

$$\gamma(j) \geq 0, \qquad j \geq 0, \qquad \sum_{j=0}^{\infty} \gamma(j) < \infty. \tag{2.7}$$

This in turn implies that y_t has short memory in the sense that $f(\lambda)$ is bounded. Consequently, the present paper does not cover long memory autocorrelation in y_t.

The remaining conditions for consistency are essentially taken from Hannan (1973).

Assumption 2.

(i) Θ in (1.16) is compact.
(ii) $\theta_0 \in \Theta$ and $\sigma^2 > 0$.
(iii) For all $\theta \in \Theta$

$$\int_{-\pi}^{\pi} \log g(\lambda;\theta) d\lambda = 0. \tag{2.8}$$

(iv) $g(\lambda;\theta)^{-1}$ is continuous in $(\lambda,\theta) \in [-\pi,\pi] \times \Theta$.
(v) The set $\{\lambda : g(\lambda;\theta) \neq g(\lambda;\theta_0)\}$ has positive Lebesque measure, for all $\theta \in \Theta/\{\theta_0\}$.

THEOREM 2.1. *Under Assumptions 1(4) and 2, as $n \to \infty$*

$$\hat{\theta} \to_p \theta_0. \tag{2.9}$$

Proof. Assumption 1(4) and (2.8) imply the representation $y_t - Ey_t = \sum_{j=0}^{\infty} a_j \eta_{t-j}$, where $\sum_{j=0}^{\infty} a_j^2 < \infty$ and $\{\eta_j\}$ is a sequence of uncorrelated, homoskedastic variables. On the other hand we also have $y_t = f(\varepsilon_t, \varepsilon_{t-1}, \ldots)$ for measurable f. Thus (cf. Stout, 1974, Theorem 3.5.8) ergodicity of $\{\varepsilon_j\}$ implies ergodicity of y_t. The proof now follows from that of Theorem 1 of Hannan (1973). ∎

For the central limit theorem, we introduce the following assumption.

Assumption 3.

(i) θ_0 is an interior point of Θ.
(ii) In a neighborhood of θ_0, $(\partial/\partial\theta)g^{-1}(\lambda;\theta)$ and $(\partial^2/\partial\theta\partial\theta')g^{-1}(\lambda;\theta)$ exist and are continuous in λ and θ.
(iii) $(\partial/\partial\theta)g^{-1}(\lambda;\theta_0) \in Lip(\eta)$, $\eta > \frac{1}{2}$.
(iv) The matrix

$$W = \frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{\partial \log g(\lambda;\theta_0)}{\partial\theta} \frac{\partial \log g(\lambda;\theta_0)}{\partial\theta'}\, d\lambda \tag{2.10}$$

is nonsingular.

The proof of the following theorem (see Corollary 3.1 of Section 3) implies that under our conditions y_t has fourth cumulant spectrum $f(\lambda,\omega,\nu)$, for $\lambda,\omega,\nu \in (-\pi,\pi]$, given by

$$f(\lambda,\omega,\nu) = \frac{1}{(2\pi)^3}\sum_{j,k,\ell=-\infty}^{\infty} e^{-ij\lambda - ik\omega - i\ell\nu}\,\mathrm{Cum}(y_0,y_j,y_k,y_\ell), \tag{2.11}$$

where the final factor in the summand is the cumulant of y_0,y_j,y_k,y_ℓ, and also that the matrix

$$V = \frac{2\pi}{\sigma^2}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi} \frac{\partial g(\lambda;\theta_0)^{-1}}{\partial\theta} \frac{\partial g(\omega;\theta_0)^{-1}}{\partial\theta'}\, f(\lambda,-\omega,\omega)\,d\lambda d\omega \tag{2.12}$$

is finite.

THEOREM 2.2. *Under Assumptions 1(8), 2, and 3, as* $n \to \infty$

$$n^{1/2}(\hat\theta - \theta_0) \to_d N(0, 2W^{-1} + W^{-1}VW^{-1}). \tag{2.13}$$

The proof of Theorem 2.2 is considerably longer than that of Theorem 2.1 as a result of the central limit theorem we establish for quadratic forms of y_t. Thus the proof appears in the following section. Meanwhile, we discuss implications of Theorems 2.1 and 2.2.

Remark 2.1. The form of asymptotic covariance matrix in the theorem is standard in the literature on Whittle estimation in the absence of Gaussianity or linearity assumptions (see, e.g., Robinson, 1978; Chiu, 1988; and in a more specialized setting, Cameron and Hannan, 1979). Of course in the event, impossible under the present circumstances, that y_t were Gaussian, V would vanish because $f(\lambda,\omega,\nu)$ would identically vanish. The term V would also vanish in the likewise impossible circumstances that y_t were linear in martingale difference innovations whose first four conditional moments are constant because then (from, e.g., Brillinger, 1975, p. 39) $f(\lambda,-\omega,\omega)$ is proportional to $g(\lambda)g(\omega)$ and (2.8) holds. Unfortunately we have no reason to believe that $V = 0$ under our ARCH model, an unattractive feature of Whittle estimation in this context. Presumably Assumption 1(8) imposes structure on $f(\lambda,-\omega,\omega)$ and hence on V, but we have not analyzed this.

Remark 2.2. Thus Theorem 2.2 is only useful in inference if V, and also W, can be consistently estimated. A consistent estimate of W is easily shown to be

$$\hat{W} = \frac{1}{n}\sum_{j=1}^{n-1} \frac{\partial \log g(\lambda_j;\hat{\theta})}{\partial \theta} \frac{\partial \log g(\lambda_j,\hat{\theta})}{\partial \theta'}. \tag{2.14}$$

An estimate of V was proposed by Taniguchi (1982) and one of $2W + V$ by Chiu (1988), both of which can readily be used along with (2.14) in estimating $2W^{-1} + W^{-1}VW^{-1}$. However, these authors established consistency of their estimates under Brillinger-mixing conditions, and we have no evidence that these hold under our ARCH model. A proof of consistency under our setup would likely be very lengthy; indeed the corresponding proofs of Taniguchi (1982) and Chiu (1988) were almost entirely omitted as a result of pressure of space. Mean square consistency, the property considered by these authors, would unfortunately require finiteness of sixteenth moments of x_t, a dubious proposition in the case of much financial data.

Remark 2.3. Theorem 2.2 is silent about limit distributional behavior when (2.6) holds with $J = 4$ (when $\hat{\theta}$ is consistent) but not in the more limited situation when $J = 8$. Moreover, though (2.5) and (2.6) only restrict θ_0, they should ideally be reflected in our choice of Θ. This is problematic because, despite the scale restriction (2.3), μ_4 and μ_8 are unknown because we have imposed no distributional assumption on ε_t. For Gaussian ε_t, $\mu_4^{1/2} \simeq 1.732$ and $\mu_8^{1/4} \simeq 3.2$. In this case we can compare (2.6) with the necessary and sufficient conditions for finiteness of Jth moments of GARCH(1,1) x_t due to Bollerslev (1986) (his ε_t is our x_t). In particular, for $J = 4$ (2.6) gives $3\alpha_1^2 + 2.3^{1/2}\alpha_1\beta_1 + \beta_1^2 < 1$, whereas Bollerslev's condition is $3\alpha_1^2 + 2\alpha_1\beta_1 + \beta_1^2 < 1$. For the MLE $\tilde{\theta}$, only $\alpha_1 + \beta_1 < 1$ is needed. Notice that Whittle estimation based on squares could doubtless be justified under many other assumptions besides ARCH ones, when (2.6) would not be relevant; indeed not only has this been done under an alternative stochastic volatility model by Zaffaroni (1999), but, unlike the MLE described in the previous section, the uncorrelatedness of levels property (1.1) is not essential; for example, x_t, and thus y_t, could be strongly mixing at the same rate.

Remark 2.4. Recent empirical evidence suggests that in many financial data sets sample autocorrelations decay more slowly than the exponential rate prescribed by GARCH(p,q) models. Although (2.7) rules out long memory, Giraitis et al. (2000) showed that it permits

$$\gamma(j) \sim c_1 j^{-\zeta} \quad \text{as } j \to \infty, \qquad c_1 > 0, \qquad \zeta > 1, \tag{2.15}$$

and that this occurs when

$$\psi_j \sim c_2 j^{-\zeta}, \quad \text{as } j \to \infty, \qquad c_2 > 0. \tag{2.16}$$

We could thus take $\theta = (\theta_1, \theta_2)'$ and

$$g(\lambda;\theta) = \left|1 - (\theta_1 - 1)\theta_2 \sum_{j=1}^{\infty} j^{-\theta_1} e^{ij\lambda}\right|^{-2}, \tag{2.17}$$

where the true θ_2 is upper-bounded by unity because $\sum_{j=1}^{\infty} j^{-\theta_1}$ is nearly $(\theta_1 - 1)^{-1}$, whereas μ_4 and μ_8 are at least unity. Clearly (2.17) satisfies (2.8), and we conjecture that it satisfies our other conditions for suitable θ_1, θ_2, though the lack of a closed form representation of the infinite series in (2.17) is a practical disadvantage. Automatic truncation of this series, similar to that in (1.6) and (1.7), is embodied in the alternative Whittle objective function to (1.13),

$$\sum_{t=2}^{n} \left\{ y_t - \bar{y} - \sum_{j=1}^{t-1} \psi_j(\theta)(y_{t-j} - \bar{y}) \right\}^2, \tag{2.18}$$

where $\bar{y} = n^{-1} \sum_{t=1}^{n} y_t$. Box and Jenkins (1971) considered (2.18) in the context of ARMA estimation, where the ψ_j decay exponentially, but it seems possible to show that the minimizer of (2.18) has the properties of Theorems 2.1 and 2.2 in the case of (2.17). Alternatively, (2.15) can be described by the alternative model

$$g(\lambda;\theta) = \exp\left\{\theta_2 \left(\frac{\pi^{\theta_1 - 1} - |\lambda|^{\theta_1}}{\theta_1} \right)\right\}, \qquad \theta_1 > \frac{3}{2}, \qquad \theta_2 > 0, \tag{2.19}$$

which is convenient for use in (1.13). For $\theta_1 < 3$, (2.19) has a peak at $\lambda = 0$ that is finite but not very smooth, thus approaching long memory behavior. With $\theta_1 = 2$ a priori, (2.19), or a continuous time version thereof, was considered by Lumley and Panofsky (1964) in modeling atmospheric turbulence, and in connection with Whittle estimation by Robinson (1978), and also by Chiu (1988) in connection with an alternative method of estimation. For $\theta_1 = 2$, it is readily shown that if θ_2 is the true value

$$\gamma(j) \propto (\theta_2 + j^2)^{-1}, \quad \text{all } j, \tag{2.20}$$

satisfying (2.15) for $\zeta = 2$. For $\frac{3}{2} < \theta_1 < 2$ an analytic formula is unavailable, but from Yong (1974, Theorem III-31) we deduce that

$$\gamma(j) \sim c_3 j^{-\theta_1}, \quad \text{as } j \to \infty, \qquad c_3 > 0. \tag{2.21}$$

The requirement $\theta_1 > \frac{3}{2}$ in (2.19) is to satisfy Assumption 3(iii). It is easily seen that the remaining relevant parts of Assumptions 2 and 3 are satisfied, though we are unable to check (2.5) or (2.6), the Kolmogorov–Wiener formulae admitting no closed form solution. Note that Assumptions 2 and 3 will also be satisfied if we generalize (2.19) by multiplying it by a factor corresponding to the g for an ARMA model with standard parameterization, or a Bloomfield (1973) model; though undoubtedly many members of this family will not satisfy (2.5) and (2.6), nevertheless its practical usefulness in modeling financial and other data may be worth exploring. Some spectral models do not satisfy (2.8) in that the prediction error variance σ^2 is not an explicitly known function of the parameters in f (see (1.10)), as for example when ARCH x_t are observed subject to measurement error, and it is the squares of these noise-corrupted observations that are analyzed. In such cases we could replace $w_n(\theta)$ by

$$\sum_{j=1}^{n-1} \left\{ \log f(\lambda_j;\tau) + \frac{I(\lambda_j)}{f(\lambda_j;\tau)} \right\}, \tag{2.22}$$

where τ is the full set of spectral parameters and the extension of our asymptotic properties for $w_n(\theta)$ is standard.

Remark 2.5. Undoubtedly the asymptotic properties of Theorems 2.1 and 2.2 will hold for other versions of Whittle estimation under (2.8), besides (1.13) and (2.18). We have stressed $w_n(\theta)$ because it both exploits the fact that $g(\lambda;\theta)$ is more often a known, convenient form than formulae for autocovariances or autoregressive coefficients and makes ready use of the fast Fourier transform, which can significantly aid the processing of long financial time series. We can show that alternative estimates that are not of the Whittle family but are also functions of quadratic forms of y_t are $\sqrt{n}$-consistent and asymptotically normal, for example, simple method of moments estimates such as that for (2.19) in Robinson (1978). Also, as in Robinson (1978), it is possible to show that a single Newton-type step from such an estimate, based on the objective function $w_n(\theta)$, will achieve the limiting variance of Theorem 2.2. Unfortunately, however, in the present circumstances we cannot assert that this necessarily corresponds to an efficiency improvement in view of the Whittle approach's guaranteed inefficiency under current circumstances; whether matters are made better or worse will typically depend on the actual values of the true parameters, no general efficiency ordering of these inefficient estimates being possible. We have chosen to study Whittle estimation based on squares because of the relative difficulty of a general asymptotic theory to cover the maximum likelihood approach described in Section 1 and also because of the immediate availability of $g(\lambda;\theta)$ in ARCH models of form (1.2) and the familiarity of the method and availability of software to workers in time series analysis.

3. PROOF OF THEOREM 2.2

By-now-familiar arguments from the literature on Whittle asymptotics of Hannan (1973) and subsequent authors leave us with the task of establishing that

$$n^{1/2}\nu' \int_{-\pi}^{\pi} \frac{\partial}{\partial\theta} g(\lambda,\theta_0)^{-1}(I(\lambda) - EI(\lambda))d\lambda \Rightarrow_d N(0,\sigma^2) \tag{3.1}$$

for any non-null $p \times 1$ vector ν, with $\sigma^2 = \nu' V \nu$. With

$$d(j) = \nu' \int_{-\pi}^{\pi} \frac{\partial}{\partial \theta} g(\lambda, \theta_0)^{-1} e^{ij\lambda} d\lambda$$

the left side of (3.1) is $n^{-1/2} Q_n$, where

$$Q_n = \sum_{t,s=1}^{N} d(t-s)(X_t X_s - E(X_t X_s)),$$

where

$$X_t = y_t - E y_t.$$

We shall in fact establish (3.1) under the mild requirement

$$\sum_{t=-\infty}^{\infty} d(t)^2 < \infty, \tag{3.2}$$

which is equivalent to square integrability of $(\partial/\partial\theta) g(\lambda, \theta_0)^{-1}$, whereas Assumption 3(iii) implies that (see Zygmund, 1979, p. 240)

$$\sum_{t=-\infty}^{\infty} |d(t)| < \infty.$$

From Giraitis et al. (2000) and Nelson, (1990b), y_t has the unique second-order stationary solution

$$y_t = \psi_0 \Bigg\{ \varepsilon_t^2 + \sum_{j_1=1}^{\infty} \psi_{j_1} \varepsilon_t^2 \varepsilon_{t-j_1}^2 + \psi_0 \sum_{j_1, j_2=1}^{\infty} \psi_{j_1} \psi_{j_2} \varepsilon_t^2 \varepsilon_{t-j_1}^2 \varepsilon_{t-j_1-j_2}^2 + \sum_{j_1, j_2, j_3=1}^{\infty} \psi_{j_1} \psi_{j_2} \psi_{j_3} \varepsilon_t^2 \varepsilon_{t-j_1}^2 \varepsilon_{t-j_1-j_2}^2 \varepsilon_{t-j_1-j_2-j_3}^2 + \cdots \Bigg\} \tag{3.3}$$

after relaxing the i.i.d. assumption on ε_t of these authors. Writing $\xi_t = \varepsilon_t^2$,

$$X_t = \psi_0 \sum_{l=0}^{\infty} (m_l(t) - E m_l(t)),$$

where $m_0(t) = \xi_t$ and

$$m_l(t) = \sum_{j_1, \ldots, j_l=1}^{\infty} \psi_{j_1} \ldots \psi_{j_l} \xi_t \xi_{t-j_1} \ldots \xi_{t-j_1-\cdots-j_l}$$

$$= \sum_{j_l < \cdots < j_1 < t}^{\infty} \psi_{t-j_1} \psi_{j_1-j_2} \ldots \psi_{j_{l-1}-j_l} \xi_t \xi_{j_1} \ldots \xi_{j_l}, \qquad (l \geq 1). \tag{3.4}$$

Therefore

$$Q_n = \psi_0^2 \sum_{l,k=0}^{\infty} Q_n^{(l,k)}, \tag{3.5}$$

where

$$Q_n^{(l,k)} = \sum_{t,s=1}^{n} d(t-s) : m_l(t) :: m_k(s) :$$

with the definition $:x: = x - Ex$. Hence

$$\mathrm{Var}(Q_n) = \psi_0^4 \sum_{k_1,\dots,k_4=0}^{\infty} \mathrm{Cov}(Q_n^{(k_1,k_2)}, Q_n^{(k_3,k_4)}). \tag{3.6}$$

By Lemma 3.2, (3.21), which follows,

$$n^{-1} |\mathrm{Cov}(Q_n^{(k_1,k_2)}, Q_n^{(k_3,k_4)})| \le C \left\{ \sum_t d^2(t) \right\} [(k_1+1)\dots(k_4+1)]^2 D_1^{k_1+\cdots k_4} \tag{3.7}$$

uniformly in $n, k_1, \dots, k_4$ with $D_1 = \mu_8^{1/4} \sum_{j=1}^{\infty} \psi_j$, so that from (3.6),

$$n^{-1}\,\mathrm{Var}(Q_n) \le C \left(\sum_{k=0}^{\infty} (k+1)^2 D_1^k \right)^4 < \infty, \tag{3.8}$$

and thus the series (3.5) converges in L^2.

Put

$$Q_n^{(l,k)} = V_n^{(l,k)} + R_n^{(l,k)},$$

where

$$V_n^{(l,k)} = \sum_{t,s=1}^{n} d^-(t-s) : m_l^-(t) :: m_k^-(s) :$$

and

$$m_l^-(t) = \sum_{j_l < \cdots < j_1 < t} \psi_{t-j_1}^- \psi_{j_1-j_2}^- \dots \psi_{j_{l-1}-j_l}^- \xi_t \xi_{j_1} \dots \xi_{j_l}$$

with

$$d^-(t) := d(t) 1(|t| \le L); \qquad \psi_t^- := \psi_t 1(1 \le t \le L),$$

where $L > 0$ is a fixed large number. Now write

$$Q_n^- = \psi_0^2 \sum_{l,k=0}^{L} V_n^{(l,k)}; \qquad Q_n^+ = \psi_0^2 \sum_{l,k=0}^{L} R_n^{(l,k)} + \psi_0^2 \sum_{l,k=0\,:\,\max(l,k)>L}^{\infty} Q_n^{(l,k)} \tag{3.9}$$

so that

$$Q_n =: Q_n^- + Q_n^+.$$

The proof of (3.1) now follows immediately from Proposition 3.1 and Lemma 3.1, which appear subsequently, where the former result also uses the auxiliary Lemma 3.2.

LEMMA 3.1. *Let Assumption 1(8) and (3.2) hold. Then for any fixed $L > 1$,*

$$n^{-1/2}(Q_n^- - EQ_n^-) \Rightarrow N(0, \sigma_L^2) \qquad (n \to \infty), \tag{3.10}$$

where

$$\sigma_L^2 \to \sigma^2 \qquad (L \to \infty). \tag{3.11}$$

Proof. We shall consider decompositions of the form

$$Y_n = \sum_{t=1}^{n} v_t + R_n \tag{3.12}$$

for given sequence Y_n, with v_t a sequence of martingale differences and T_n a remainder satisfying

$$ER_n^2 = O(1). \tag{3.13}$$

With some abuse of notation, but for ease of presentation, we shall employ the same notation v_t, R_n, even when the form of Y_n changes. We show later that $Y_n = Q_n^- - Q_n^-$ has a decomposition (3.12) where

$$v_t = (\xi_t - E\xi_t)f_t + (\xi_t^2 - E\xi_t^2)g_t, \qquad Ev_t^2 < \infty, \tag{3.14}$$

where $f_t = f(\xi_{t-1}, \xi_{t-2}, \ldots, \xi_{t-K})$, $g_t = g(\xi_{t-1}, \xi_{t-2}, \ldots, \xi_{t-K})$, $Ef_t^2 < \infty$, $Eg_t^2 < \infty$, and f, g are polynomials with $K \geq 1$. Clearly,

$$E(v_t | \xi_{t-1}, \xi_{t-2}, \ldots) = 0$$

almost surely, and by the same argument as in the proof of Theorem 2.1, v_t is also stationary and ergodic. It follows that by Theorem 23.1 of Billingsley (1968)

$$n^{-1/2} \sum_{t=1}^{n} v_t \Rightarrow \mathcal{N}(0, Ev_0^2) \qquad (n \to \infty),$$

so given (3.12) and (3.13),

$$n^{-1/2}(Q_n^- - EQ_n^-) \Rightarrow \mathcal{N}(0, Ev_0^2) \qquad (n \to \infty)$$

and

$$n^{-1} \operatorname{Var}(Q_n^-) \to \sigma_L^2 \equiv Ev_0^2 \qquad (n \to \infty),$$

to prove (3.10). By (3.22) of Lemma 3.2, which follows

$$
\begin{aligned}
n^{-1}\operatorname{Var}(Q_n^-) &= \sum_{k_1,\dots,k_4=0}^{L} n^{-1}\operatorname{Cov}(V_n^{(k_1,k_2)}, V_n^{(k_3,k_4)}) \\
&\to \sum_{k_1,\dots,k_4=0}^{L}\left\{\sum_{u,v,k=-\infty}^{\infty} d(u)d(v)\operatorname{Cov}(:m_{k_1}^-(0)::m_{k_2}^-(u):,:m_{k_3}^-(k)::m_{k_4}^-(k+v):)\right\} \\
&= \sigma_L^2.
\end{aligned}
$$

From relations (3.21)–(3.23) of Lemma 3.2, which appears subsequently, it follows easily that

$$
\begin{aligned}
\sigma_L^2 &\to \sum_{k_1,\dots,k_4=0}^{\infty}\left\{\sum_{u,v,k=-\infty}^{\infty} d(u)d(v)\operatorname{Cov}(:m_0^{(k_1)}::m_u^{(k_2)}:,:m_k^{(k_3)}::m_{k+v}^{(k_4)}):\right\} \\
&= \sum_{u,v,k=-\infty}^{\infty} d(u)d(v)\operatorname{Cov}(X_0X_u, X_kX_{k+v}) < \infty \qquad (L\to\infty),
\end{aligned}
$$

to prove (3.11). It remains to establish (3.12) and (3.13) for $Y_n = Q_n^- - EQ_N^-$. From (3.9) it suffices to consider $Y_n = V_n^{(l,k)} - E[V_n^{(l,k)}]$ for arbitrary l,k. Because $d^-(t) = 0$ if $|t| > L$, we have

$$
\begin{aligned}
V_n^{(l,k)} &= \sum_{t,s=1}^{n} d^-(t-s):m_k^-(t)::m_l^-(s): \\
&= \widetilde{V}_n^{(l,k)} + R_n,
\end{aligned}
$$

where

$$
\widetilde{V}_n^{(l,k)} = \sum_{t=1}^{n}\sum_{s:|s-t|\le L}[\dots], \qquad R_n = -\sum_{t=1}^{n}\sum_{s:|s-t|\le L,\, s\le 0 \text{ or } s>n}[\dots].
$$

Because

$$
\begin{aligned}
ER_n^2 &\le \max_t |d(t)|^2 E\left(\sum_{t=1}^{L}\sum_{s=-L}^{0}|m_k^-(t)m_l^-(s)| + \sum_{t=n-L}^{L}\sum_{s=n+1}^{n+L}|m_k^-(t)m_l^-(s)|\right)^2 \\
&\le \max_t |d(t)|^2 8L^2(Em_k^-(0)^2)^{1/2}(Em_l^-(0)^2)^{1/2} < \infty
\end{aligned}
$$

and $\widetilde{V}_n^{(l,k)}$ is a linear combination of finitely many sums $T_n(v) := \sum_{t=1}^{n} :m_k^-(t)::m_l^-(t-v):$ it suffices to establish (3.12) and (3.13) for $Y_n = T_n(v)$. By definition

$$
\begin{aligned}
:m_k^-(t) := \sum_{j_1,\dots,j_k=1}^{L} \psi_{j_1}\dots\psi_{j_k}(&\xi_t\xi_{t-j_1}\xi_{t-j_1-j_2}\dots\xi_{t-j_1-\dots-j_k} \\
&- E[\xi_t\xi_{t-j_1}\xi_{t-j_1-j_2}\dots\xi_{t-j_1-\dots-j_k}]).
\end{aligned}
$$

Because $E[\xi_t \xi_{t-j_1} \xi_{t-j_1-j_2} \dots \xi_{t-j_1-\dots-j_k}]$ does not depend on t, $T_n(v)$ can be written as the sum of a constant not depending on t and a linear combination of the finitely many sums

$$s_n(u_1,\dots,u_{k^*}) = \sum_{t=1}^{n} \xi_{t-u_1} \xi_{t-u_2} \dots \xi_{t-u_{k^*}}$$

with $1 \le k^* \le k + l$, where $u_1 \le u_2 \le \dots \le u_{k^*}$ and no u_i can be repeated in $(u_1,\dots,u_{k^*})$ more than twice. Therefore it suffices to show that $s_n(u_1,\dots,u_{k^*})$ admits the decomposition of type (3.12):

$$s_n(u_1,\dots,u_{k^*}) - E[s_n(u_1,\dots,u_{k^*})] \equiv \sum_{t=1}^{n} :\xi_{t-u_1} \xi_{t-u_2} \xi_{t-u_2} \dots \xi_{t-u_{k^*}} := \sum_{t=1}^{n} v_t + R_n. \quad \textbf{(3.15)}$$

We prove this by induction. Let $k^* = 1$. Then

$$:s_n(u_1) := \sum_{t=1}^{n} (\xi_{t-u_1} - E\xi_{t-u_1}) = \sum_{t=1}^{n} (\xi_t - E\xi_t) + R_n,$$

where

$$R_n = \sum_{t=1}^{n} (\xi_t - E\xi_t) - \sum_{t=1}^{n} (\xi_{t-u_1} - E\xi_{t-u_1}).$$

Clearly (3.15) and (3.13) hold.

It remains to show that (3.15) and (3.13) hold for $k^* = p \ge 2$ if they hold for $k^* = 1,\dots,p-1$. Indeed, if $u_1 < u_2$ then $E[\xi_{t-u_1} \xi_{t-u_2} \dots \xi_{t-u_{k^*}}] = E[\xi_{t-u_1}]E[\xi_{t-u_2} \dots \xi_{t-u_{k^*}}]$ and

$$\xi_{t-u_1} \xi_{t-u_2} \dots \xi_{t-u_{k^*}} - E[\xi_{t-u_1} \xi_{t-u_2} \dots \xi_{t-u_{k^*}}] = (\xi_{t-u_1} - E[\xi_{t-u_1}])\xi_{t-u_2} \dots \xi_{t-u_{k^*}} + E[\xi_{t-u_1}] :\xi_{t-u_2} \dots \xi_{t-u_{k^*}}:. \quad \textbf{(3.16)}$$

Because $u_1 < u_2 \le \dots \le u_{k^*}$, the sum over t of the first term on the right satisfies (3.15):

$$\sum_{t=1}^{n} (\xi_{t-u_1} - E[\xi_{t-u_1}])\xi_{t-u_2} \dots \xi_{t-u_{k^*}} = \sum_{t=1}^{n} (\xi_t - E[\xi_t])\xi_{t-(u_2-u_1)} \dots \xi_{t-(u_{k^*}-u_1)} + R_n,$$

where clearly $ER_n^2 = O(1)$ and $E(\xi_{t-(u_2-u_1)} \dots \xi_{t-(u_{k^*}-u_1)})^2 < \infty$. For the sums $\sum_{t=1}^{n} :\xi_{t-u_2} \dots \xi_{t-u_{k^*}}:$ from the second term of (3.16), (3.13) holds by assumption. If $u_1 = u_2$ then $u_2 < u_3 \le \dots \le u_{k^*}$, $E[\xi_{t-u_1} \xi_{t-u_2} \xi_{t-u_3} \dots \xi_{t-u_{k^*}}] = E[\xi_{t-u_1}^2]E[\xi_{t-u_3} \dots \xi_{t-u_{k^*}}]$, and

$$\xi_{t-u_1} \xi_{t-u_2} \xi_{t-u_3} \dots \xi_{t-u_{k^*}} - E[\xi_{t-u_1} \xi_{t-u_2} \xi_{t-u_3} \dots \xi_{t-u_{k^*}}] = (\xi_{t-u_1}^2 - E[\xi_{t-u_1}^2])\xi_{t-u_3} \dots \xi_{t-u_{k^*}} + E[\xi_{t-u_1}^2]\xi_{t-u_3} \dots \xi_{t-u_{k^*}},$$

which gives (3.15) by assumption. ■

PROPOSITION 3.1. *Let Assumption 1(8) and (3.2) hold. Then*

$$n^{-1}\operatorname{Var}Q_n^{+} \to 0 \qquad (L \to \infty)$$

uniformly in n.

Proof. From (3.9),

$$\operatorname{Var} Q_n^{+} \le 2\psi_0^4 \left\{ \operatorname{Var}\left(\sum_{l,k=0}^{L} R_n^{(l,k)} \right) + \operatorname{Var}\left(\sum_{l,k=0\,:\,\max(l,k)>L}^{\infty} Q_n^{(l,k)} \right) \right\} =: q_n^{(1)} + q_n^{(2)},$$

so it suffices to show that

$$n^{-1} q_n^{(j)} \to 0 \qquad (L \to \infty), \qquad (j = 1,2) \tag{3.17}$$

uniformly in $n \ge 1$. For $j = 2$, the bound (3.21) of Lemma 3.2, which follows, gives

$$\begin{aligned} n^{-1} q_n^{(2)} &= n^{-1}\psi_0^4 \sum_{k_1,k_2=0:\max(k_1,k_2)>L,\ \max(k_3,k_4)>L}^{\infty} \operatorname{Cov}(Q_n^{(k_1,k_2)}, Q_n^{(k_3,k_4)}) \\ &\le C\left\{ \sum_{k=L}^{\infty} (k+1)^2 D_1^k \right\} \left\{ \sum_{k=0}^{\infty} (k+1)^2 D_1^k \right\}^3 \to 0 \qquad (L \to \infty) \end{aligned} \tag{3.18}$$

uniformly in $n \ge 1$ because $D_1 < 1$.

We now prove (3.17) for $j = 1$. Denote $m_l^{+}(t) = m_l(t) - m_l^{-}(t), : m_l^{+}(t) := m_l^{+}(t) - E m_l^{+}(t)$, and $d^{+}(t) = d(t) - d^{-}(t)$. Write

$$\begin{aligned} R_n^{(l,k)} &\equiv \sum_{t,s=1}^{n} [d(t-s) : m_l(t) :: m_k(s) : - d^{-}(t-s) : m_l^{-}(t) :: m_k^{-}(s) :] \\ &= \sum_{t,s=1}^{n} d^{+}(t-s) : m_l(t) :: m_k(s) : + \sum_{t,s=1}^{n} d^{-}(t-s) : m_l^{+}(t) :: m_k(s) : \\ &\quad + \sum_{t,s=1}^{n} d^{-}(t-s) : m_l^{-}(t) :: m_k^{+}(s) : \\ &=: r_n^{(l,k)}(1) + r_n^{(l,k)}(2) + r_n^{(l,k)}(3). \end{aligned}$$

Then

$$q_n^{(1)} \le 3\left\{ \operatorname{Var}\left(\sum_{l,k=0}^{L} r_n^{(l,k)}(1) \right) + \operatorname{Var}\left(\sum_{l,k=0}^{L} r_n^{(l,k)}(2) \right) + \operatorname{Var}\left(\sum_{l,k=0}^{L} r_n^{(l,k)}(3) \right) \right\}.$$

It remains to show that

$$n^{-1} \operatorname{Var}\left(\sum_{l,k=0}^{L} r_n^{(l,k)}(j) \right) \to 0 \qquad (L \to 0), \qquad j = 1,2,3, \tag{3.19}$$

uniformly in n. Set

$$g^{(l)}_{j_1,\dots,j_l} = \psi_{j_1}\dots\psi_{j_l} 1(j_1 \ge 1,\dots,j_l \ge 1),$$

$$g^{-,(l)}_{j_1,\dots,j_l} = \psi^-_{j_1}\dots\psi^-_{j_l}; \quad g^{+,(l)}_{j_1,\dots,j_l} = g^{(l)}_{j_1,\dots,j_l} - g^{-,(l)}_{j_1,\dots,j_l}.$$

If $l = 0$ define $g^{(l)}_{j_1,\dots,j_l} = g^{-,(l)}_{j_1,\dots,j_l} = 1, g^{+,(l)}_{j_1,\dots,j_l} = 0$.

Then

$$r_n^{(l,k)}(1) = \sum_{t,s=1}^{n} d^+(t-s) g^{(l)}_{t-j_1,j_1-j_2,\dots,j_{l-1}-j_l} g^{(k)}_{s-s_1,s_1-s_2,\dots,s_{k-1}-s_k} \times (\xi^J - E\xi^J)(\xi^S - E\xi^S),$$

$$r_n^{(l,k)}(2) = \sum_{t,s=1}^{n} d^-(t-s) g^{+,(l)}_{t-j_1,j_1-j_2,\dots,j_{l-1}-j_l} g^{(k)}_{s-s_1,s_1-s_2,\dots,s_{k-1}-s_k} \times (\xi^J - E\xi^J)(\xi^S - E\xi^S),$$

$$r_n^{(l,k)}(3) = \sum_{t,s=1}^{n} d^-(t-s) g^{-,(l)}_{t-j_1,j_1-j_2,\dots,j_{l-1}-j_l} g^{+,(k)}_{s-s_1,s_1-s_2,\dots,s_{k-1}-s_k} \times (\xi^J - E\xi^J)(\xi^S - E\xi^S),$$

where $\xi^J = \xi_t \xi_{j_1}\dots\xi_{j_l}$, $\xi^S = \xi_s \xi_{s_1}\dots\xi_{s_k}$, $J = \{t, j_1,\dots,j_l\}$, $S = \{s, s_1,\dots,s_k\}$. We have

$$n^{-1}\operatorname{Var}\left(\sum_{l,k=0}^{L} r_n^{(l,k)}(j)\right) = n^{-1} \sum_{k_1,\dots,k_4=0}^{L} \operatorname{Cov}(r_n^{(k_1,k_2)}(j), r_n^{(k_3,k_4)}(j)), \qquad j = 1,2,3.$$

By Lemma 3.2, (3.21)

$$\operatorname{Cov}(r_n^{(k_1,k_2)}(1), r_n^{(k_3,k_4)}(1)) \le C \sum_{t\in Z} (d^+(t))^2 \prod_{l=1}^{4} \{||g^{(k_l)}||_1 (k_l+1)^2 (E\xi_0^4)^{k_l/4}\},$$

$$\operatorname{Cov}(r_n^{(k_1,k_2)}(2), r_n^{(k_3,k_4)}(2)) \le C \sum_{t\in Z} (d^-(t))^2 ||g^{+,(k_1)}||_1 ||g^{+,(k_3)}||_1 ||g^{(k_2)}||_1 ||g^{(k_4)}||_1 \times \prod_{l=1}^{4} \{(k_l+1)^2 (E\xi_0^4)^{k_l/4}\},$$

$$\operatorname{Cov}(r_n^{(k_1,k_2)}(3), r_n^{(k_3,k_4)}(3)) \le C \sum_{t\in Z} (d^-(t))^2 ||g^{-,(k_1)}||_1 ||g^{-,(k_3)}||_1 ||g^{+,(k_2)}||_1 ||g^{+,(k_4)}||_1 \times \prod_{l=1}^{4} \{(k_l+1)^2 (E\xi_0^4)^{k_l/4}\},$$

where $||\cdot||_1$ denotes the L^1 norm (see (3.20), which follows). Because

$$||g^{-,(k)}||_1 \le ||g^{(k)}||_1 = \sum_{j_1,\dots j_k=1}^{\infty} \psi_{j_1}\dots\psi_{j_k} = \left(\sum_{j=1}^{\infty}\psi_j\right)^k,$$

$$||g^{+,(k)}||_1 \le \sum_{p=1}^{k}\sum_{j_1,\dots,j_k=1}^{\infty} \psi_{j_1}\dots\psi_{j_k} 1(j_p \ge L) \le \left(\sum_{j\ge L}\psi_j\right) k \left(\sum_{j=1}^{\infty}\psi_j\right)^{k-1},$$

and $D_1 = \mu_8^{1/4}\sum_{j=1}^{\infty}\psi_j < 1$, we have

$$n^{-1}\,\mathrm{Var}\left(\sum_{l,k=0}^{L} r_n^{(l,k)}(1)\right) \le C\left(\sum_{|t|\ge L} d^2(t)\right)\left(\sum_{k=0}^{\infty}(k+1)^2 D_1^k\right)^4 \to 0 \qquad (L\to 0),$$

$$n^{-1}\,\mathrm{Var}\left(\sum_{l,k=0}^{L} r_n^{(l,k)}(j)\right) \le C\left[\sum_{t\in Z} d^2(t)\right]\left[\sum_{j\ge L}\psi_j\right]\left(\sum_{k=0}^{\infty}(k+1)^3 D_1^k\right)^4 \to 0 \qquad (L\to 0);\quad j=2,3$$

to prove (3.19). ■

We now provide the auxiliary Lemma 3.2 used in the proof of Proposition 3.1.

LEMMA 3.2. *Define the quadratic forms*

$$Z_n = \sum_{t,s=1}^{n} d(t-s) Y_t^{(k_1)} Y_s^{(k_2)}, \quad Z_n' = \sum_{t,s=1}^{n} d(t-s) Y_t^{(k_3)} Y_s^{(k_4)},$$

where $k_1,\dots k_4 \ge 0$,

$$Y_t^{(k_i)} = \sum_{j_{k_i}<\dots<j_1<t} g^{(k_i)}_{t-j_1,j_1-j_2,\dots,j_{k_i-1}-j_{k_i}}(\xi_t\xi_{j_1}\dots\xi_{j_{k_i}} - E\xi_t\xi_{j_1}\dots\xi_{j_{k_i}}), \quad i=1,\dots,4,$$

and $Y_t^{(0)} = \xi_t$.

Suppose that for $i = 1,\dots,4$, *(3.2) holds and*

$$||g^{(k_i)}||_1 = \sum_{j_1,\dots,j_{k_i}\in Z} |g^{(k_i)}_{j_1,\dots,j_{k_i}}| < \infty. \qquad \textbf{(3.20)}$$

Then

$$|n^{-1}\mathrm{Cov}(Z_n,Z_n')| \le C\left(\sum_t d^2(t)\right)\prod_{i=1}^{4}\{||g^{(k_i)}||_1(k_i+1)^2(E\xi_0^4)^{k_i/4}\}, \qquad \textbf{(3.21)}$$

where $C > 0$ *does not depend on* $n, g^{(k_i)}$, *and d.*

Moreover,

$$n^{-1}\,\mathrm{Cov}(Z_n,Z_n') \to \sum_{u,v,k=-\infty}^{\infty} d(u)d(v)\mathrm{Cov}(Y_u^{(k_1)}Y_0^{(k_2)}, Y_k^{(k_3)}Y_{k+v}^{(k_4)}) < \infty \qquad \textbf{(3.22)}$$

as $n \to \infty$, and

$$\sum_{t_3, t_4=-\infty}^{\infty} |\mathrm{Cov}(Y_{t_1}^{(k_1)} Y_{t_2}^{(k_2)}, Y_{t_3}^{(k_3)} Y_{t_4}^{(k_4)})| \le C \prod_{i=1}^{4} \{\|g^{(k_i)}\|_1 (k_i+1)^2 (E\xi_0^4)^{k_i/4}\} \tag{3.23}$$

uniformly in k_1, k_2.

Proof. Set

$$c(t_1, \dots, t_4) := \mathrm{Cov}(Y_{t_1}^{(k_1)} Y_{t_2}^{(k_2)}, Y_{t_3}^{(k_3)} Y_{t_4}^{(k_4)}).$$

Because

$$i_n := n^{-1} \mathrm{Cov}(Z_n, Z_n') = n^{-1} \sum_{t_1, \dots, t_4=1}^{n} d(t_1 - t_2) d(t_3 - t_4) c(t_1, \dots, t_4) \tag{3.24}$$

it follows that for $n \ge 2$

$$|i_n| \le \sum_{t_1, \dots, t_4=1}^{n} (d^2(t_1 - t_2) + d^2(t_3 - t_4)) |c(t_1, \dots, t_4)|. \tag{3.25}$$

Suppose that (3.23) holds. Then

$$|i_n| \le C \sum_{t=-\infty}^{\infty} d^2(t) \left(\sup_{t_1, t_2} \sum_{t_3, t_4=1}^{\infty} |c(t_1, \dots, t_4)| + \sup_{t_3, t_4} \sum_{t_1, t_2=1}^{\infty} |c(t_1, \dots, t_4)| \right)$$
$$\le C \sum_{t=-\infty}^{\infty} d^2(t) \prod_{i=1}^{4} \{\|g^{(k_i)}\|_1 (k_i+1)^2 (E\xi_0^4)^{k_i/4}\} < \infty.$$

Thus (3.21) holds. From (3.24) and (3.23) (3.22) follows easily.

It remains to show (3.23). Put $J_p = \{j_{p,0}, j_{p,1}, \dots, j_{p,k_p}\}$, $p = 1, \dots, 4$. We can write $c(t_1, \dots, t_4)$ as

$$c(t_1, \dots, t_4) = \sum_{(j)} \prod_{p=1}^{4} g_{t_p - j_{p,1}, j_{p,1} - j_{p,2}, \dots, j_{p,k_p-1} - j_{p,k_p}}^{(k_p)} \mathrm{Cov}(:\xi^{J_1}::\xi^{J_2}:, :\xi^{J_3}::\xi^{J_4}:),$$

where the sum $\sum_{(j)}$ is taken over indexes $(j) = (j_{p,0}, \dots, j_{p,k_p} : p = 1, \dots, 4)$ such that $j_{p,k_p} < \dots < j_{p,1} < j_{p,0} \equiv t_p, p = 1, \dots, 4$; $\xi^{J_p} = \xi_{j_{p,0}} \xi_{j_{p,1}} \dots \xi_{j_{p,k_p}}$ and $:\xi^{J_p} := \xi^{J_p} - E\xi^{J_p}$ for $j = 1, \dots, 4$.

Using the Cauchy inequality it is easy to verify that

$$|\mathrm{Cov}(:\xi^{J_1}::\xi^{J_2}:, :\xi^{J_3}::\xi^{J_4}:)| \le 2 \prod_{i=1}^{4} (E|:\xi^{J_i}:|^4)^{1/4} \le 32 \prod_{i=1}^{4} (E|\xi^{J_i}|^4)^{1/4}$$
$$\le 32 \lambda_4^4 \lambda_4^{k_1 + \dots + k_4}, \tag{3.26}$$

where $\lambda_4 = (E\xi_0^4)^{1/4} \equiv (E\varepsilon_0^8)^{1/4}$.

Now observe that

$$\mathrm{Cov}(:\xi^{J_1}::\xi^{J_2}:,:\xi^{J_3}::\xi^{J_4}:)$$
$$\equiv E[:\xi^{J_1}::\xi^{J_2}::\xi^{J_3}::\xi^{J_4}:] - E[:\xi^{J_1}::\xi^{J_2}:]E[:\xi^{J_3}::\xi^{J_4}:] = 0 \qquad \textbf{(3.27)}$$

in both the following cases

(a) The sets $J_1 \cup J_2$ and $J_3 \cup J_4$ do not have common elements, because then from condition (2.4) it follows that $E[:\xi^{J_1}::\xi^{J_2}::\xi^{J_3}::\xi^{J_4}:] = E[:\xi^{J_1}::\xi^{J_2}:]E[:\xi^{J_3}::\xi^{J_4}:]$.

(b) $J_i \cap (\cup_{l=1:l\neq i}^{4} J_l) = \varnothing$ for some $i = 1,\dots,4$, because then condition (2.4) implies

$$E[:\xi^{J_1}::\xi^{J_2}::\xi^{J_3}::\xi^{J_4}:] = E[:\xi^{J_i}:]E\left[\prod_{l=1:l\neq i}^{4}:\xi^{J_l}:\right] = 0,$$

$$E[:\xi^{J_i}::\xi^{J_l}:] = E[:\xi^{J_i}:]E[:\xi^{J_l}:] = 0 \qquad (i \neq l).$$

Suppose neither (a) nor (b) is satisfied. Then the index $(j) = (J_1, J_2, J_3, J_4)$ has at least one of the following properties.

(1) $J_3 \cap (J_1 \cup J_2) \neq \varnothing$ and $J_4 \cap (J_1 \cup J_2) \neq \varnothing$ (when we write $(j) \in \mathcal{M}_1$); or
(2) $J_3 \cap (J_1 \cup J_2) \neq \varnothing$ and $J_4 \cap J_3 \neq \varnothing$ (when we write $(j) \in \mathcal{M}_2$); or
(3) $J_4 \cap (J_1 \cup J_2) \neq \varnothing$ and $J_3 \cap J_4 \neq \varnothing$ (when we write $(j) \in \mathcal{M}_3$).

Using (3.26) we get

$$|c(t_1,\dots,t_4)| \le 32\lambda_4^4 \lambda_4^{k_1+\cdots+k_4} \sum_{(j)\in\mathcal{M}_1\cup\mathcal{M}_2\cup\mathcal{M}_3} \left| \prod_{p=1}^{4} g^{(k_p)}_{t_p-j_{p,1},j_{p,1}-j_{p,2},\dots,j_{p,k_p-1}-j_{p,k_p}} \right|.$$

Therefore (3.23) follows if we show that for $j = 1,2,3$

$$T^{(i)}(t_1,t_2) := \sum_{t_3,t_4=1}^{\infty} \sum_{(j)\in\mathcal{M}_i} \left| \prod_{p=1}^{4} g^{(k_p)}_{t_p-j_{p,1},j_{p,1}-j_{p,2},\dots,j_{p,k_p-1}-j_{p,k_p}} \right|$$
$$\le ((k_1+1)\dots(k_4+1))^2 \prod_{p=1}^{4} ||g^{(k_p)}||_1. \qquad \textbf{(3.28)}$$

By definition of $\mathcal{M}_1$,

$$T^{(1)}(t_1,t_2) \le \sum_{(3,u)\in I_3;(4,v)\in I_4} \sum_{(i,l),(i',l')\in I_1\cup I_2}$$
$$\times \left\{ \sum_{t_3,t_4=1}^{\infty} \sum_{j_{p,k_p}<\cdots<j_{p,1}<t_p:p=1,\dots,4} 1(j_{3,u}=j_{i,l}, j_{4,v}=j_{i',l'}) \right.$$
$$\left. \times \prod_{p=1}^{4} |g^{(k_p)}_{t_p-j_{p,1},j_{p,1}-j_{p,2},\dots,j_{p,k_p-1}-j_{p,k_p}}| \right\}.$$

Taking the sum over $t_4; j_{4,s}, 1 \le s \le k_4 (s \ne v)$ and then over $t_3, j_{3,s'}, 1 \le s' \le k_3 (s \ne u)$ we obtain

$$\begin{aligned} T^{(1)}(t_1, t_2) \le & \sum_{(3,u)\in I_3;(4,v)\in I_4} \sum_{(i,l),(i',l')\in I_1\cup I_2} \\ & \Big\{ \|g^{(k_3)}\|_1 \|g^{(k_4)}\|_1 \sum_{j_{p,k_p}<\cdots<j_{p,1}<t_p:p=1,2} \\ & \prod_{p=1}^{2} |g^{(k_p)}_{t_j - j_{p,1}, j_{p,1}-j_{p,2},\dots,j_{p,k_p-1}-j_{p,k_p}}| \Big\} \\ \le & \Big\{ \sum_{(3,u)\in I_3;(4,v)\in I_4} \sum_{(i,l),(i',l')\in I_1\cup I_2} 1 \Big\} \|g^{(k_1)}\|_1 \dots \|g^{(k_4)}\|_1 \\ \le & \{(k_1+1)\dots(k_4+1)\}^2 \|g^{(k_1)}\|_1 \dots \|g^{(k_4)}\|_1. \end{aligned}$$

Thus (3.28) holds for $i = 1$.

Similarly using the definition of $\mathcal{M}_2$, we obtain

$$\begin{aligned} T^{(2)}(t_1, t_2) \le & \sum_{(3,u)\in I_3;(4,v)\in I_4} \sum_{(i,l)\in I_1\cup I_2,(i',l')\in I_3} \\ & \Big\{ \sum_{t_3,t_4=1}^{\infty} \sum_{j_{p,k_p}<\cdots<j_{p,1}<t_p:p=1,\dots,4} 1(j_{3,u} = j_{i,l}, j_{4,v} = j_{i',l'}) \\ & \prod_{p=1}^{4} |g^{(k_p)}_{t_p - j_{p,1}, j_{p,1}-j_{p,2},\dots,j_{p,k_p-1}-j_{p,k_p}}| \Big\}. \end{aligned}$$

Taking the sum over $t_4; j_{4,s}, 1 \le s \le k_4 (s \ne v)$ and then over $t_3; j_{3,s'}, 1 \le s' \le k_3 (s \ne u)$ we obtain (3.28):

$$\begin{aligned} T^{(2)}(t_1, t_2) \le & \sum_{(3,u)\in I_3;(4,v)\in I_4} \sum_{(i,l)\in I_1\cup I_2,(i',l')\in I_3} \\ & \Big\{ \|g^{(k_3)}\|_1 \|g^{(k_4)}\|_1 \sum_{j_{p,k_p}<\cdots<j_{p,1}<t_p:p=1,2} \\ & \prod_{p=1}^{2} |g^{(k_p)}_{t_p - j_{p,1}, j_{p,1}-j_{p,2},\dots,j_{p,k_p-1}-j_{p,k_p}}| \Big\} \\ \le & \{(k_1+1)\dots(k_4+1)\}^2 \|g^{(k_1)}\|_1 \dots \|g^{(k_4)}\|_1. \end{aligned}$$

The proof of (3.28) for $T^{(3)}(t_1, t_2)$ is similar to that for $T^{(2)}(t_1, t_2)$. ■

COROLLARY 3.1. *Suppose that Assumption 1(8) holds. Then*

$$\sum_{u,v=-\infty}^{\infty} |\mathrm{Cov}(: y_t :: y_s :, : y_u :: y_v :)| < \infty \quad \textbf{(3.29)}$$

and

$$\sum_{u,v=-\infty}^{\infty} |\mathrm{Cum}(y_t, y_s, y_u, y_v)| < \infty \tag{3.30}$$

uniformly in t, s.

Proof. By Lemma 3.2, (3.23),

$$\sum_{u,v=-\infty}^{\infty} |\mathrm{Cov}(:y_t :: y_s :, :y_u :: y_v :)|$$

$$\le \psi_0^4 \sum_{u,v=-\infty}^{\infty} \sum_{k_1,\dots,k_4=0}^{\infty} |\mathrm{Cov}(:m_{k_1}(t) :: m_{k_2}(s) :, :m_{k_3}(u) :: m_{k_4}(v) :)|$$

$$\le C \sum_{k_1,\dots,k_4=0}^{\infty} \{(k_1+1)^2 \dots (k_4+1)^2 (E[\xi_4])^{(k_1+\dots+k_4)/4}\} = C \sum_{k=0}^{\infty} D_1^k < \infty.$$

Because

$$\mathrm{Cov}(:y_t :: y_s :, :y_u :: y_v :) = \mathrm{Cum}(y_t, y_s, y_u, y_v) + \gamma(t-u)\gamma(s-v) + \gamma(t-v)\gamma(s-u)$$

and $\sum_{t \in Z} |\gamma(t)| < \infty$, this and (3.29) give (3.30). ∎

REFERENCES

Billingsley, P. (1968) *Convergence of Probability Measures*. New York: Wiley.

Bloomfield, P. (1973) An exponential model for the spectrum of a scalar time series. *Biometrika* 60, 217–226.

Bollerslev, T. (1986) Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* 31, 307–327.

Bollerslev, T., R. Chou, & K. Kroner (1992) ARCH modelling in finance: A review of the theory and empirical evidence. *Journal of Econometrics* 52, 5–59.

Bollerslev, T. & J. Wooldridge (1993) Quasi-maximum likelihood estimation and inference in dynamic models with time-varying coefficients. *Econometric Reviews* 11, 143–172.

Box, G.E.P. & G.M. Jenkins (1971) *Time Series Analysis, Forecasting and Control*. San Francisco: Holden-Day.

Brillinger, D.R. (1975) *Time Series: Data Analysis and Theory*. New York: Holt, Rinehart and Winston.

Cameron, M.A. & E.J. Hannan (1979) Transient signals. *Biometrika* 66, 243–258.

Chiu, S.-T. (1988) Weighted least squares estimators on the frequency domain for the parameters of a time series. *Annals of Statistics* 16, 1315–1326.

Ding, Z., G.W.J. Granger, & R.F. Engle (1993) A long memory property of stock market returns and a new model. *Journal of Empirical Finance* 1, 83–106.

Drost, F.C. & C.A.J. Klassen (1997) Efficient estimation in semiparametric GARCH models. *Journal of Econometrics* 81, 193–221.

Dzhaparidze, K.O. (1974) A new method for estimating spectral parameters of a stationary time series. *Theory of Probability and Its Applications* 19, 122–132.

Engle, R.F. (1982) Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica* 50, 987–1008.

Giraitis, L., P. Kokoszka, & R. Leipus (2000) Stationary ARCH models: Dependence structure and central limit theorem. *Econometric Theory* 16, 3–22.

Hannan, E.J. (1973) The asymptotic theory of linear time series models. *Journal of Applied Probability* 10, 130–145. (Corrigendum: 10, 913.)

Harvey, A. (1998) Long memory in stochastic volatility. In J. Knight & S. Satchell (eds.), *Forecasting Volatility in the Financial Markets*. Oxford: Butterworth and Heinemann.

Lee, A.W. & B.E. Hansen (1994) Asymptotic theory for the GARCH(1,1) quasi-maximum likelihood estimator. *Econometric Theory* 10, 29–52.

Linton, O. (1993) Adaptive estimation in ARCH models. *Econometric Theory* 9, 539–569.

Lumley, J.L. & H.A. Panofsky (1964) *The Structure of Atmospheric Turbulence*. New York: Wiley-Interscience.

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* 16 (3), 575–596.

Nelson, D.B. (1990a) ARCH models as diffusion approximations. *Journal of Econometrics* 45, 7–38.

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

Robinson, P.M. (1978) Alternative models for stationary stochastic processes. *Stochastic Processes and Their Applications* 8, 141–152.

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

Robinson, P.M. & M. Henry (1999) Long and short memory conditional heteroscedasticity in estimating the memory parameter of levels. *Econometric Theory* 15, 299–336.

Robinson, P.M. & P. Zaffaroni (1997) Modelling nonlinearity and long memory in time series. *Fields Institute Communications* 11, 161–170.

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

Stout, W.F. (1974) *Almost Sure Convergence*. New York: Academic Press.

Taniguchi, M. (1982) On estimation of the integrals of the fourth order cumulant spectral density. *Biometrica* 69, 117–122.

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

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

Zaffaroni, P. (1999) Gaussian Inference on Certain Long-Range Dependent Volatility Models. Preprint.

Zygmund, A. (1979) *Trigonometric Series*. Cambridge: Cambridge University Press.