PSEUDO-MAXIMUM LIKELIHOOD ESTIMATION OF ARCH(∞) MODELS[1]

Peter M. Robinson[a] **and Paolo Zaffaroni**[b]
[a]London School of Economics
[b] Banca d'Italia

December 2003

Abstract

Strong consistency and asymptotic normality of the Gaussian pseudo-maximum likelihood estimate of the parameters in a wide class of ARCH(∞) processes are established. The conditions are shown to hold in case of exponential and hyperbolic decay in the ARCH weights, though in the latter case a faster decay rate is required for the central limit theorem than for the law of large numbers. Particular parameterizations are discussed.

1 Introduction

ARCH(∞) processes comprise a wide class of models for conditional heteroscedastacity in time series. Consider an observable process x_t satisfying

$$x_t = \gamma_0 + \epsilon_t = \gamma_0 + z_t\sigma_t, \quad t \in \mathbb{Z}, \tag{1}$$

$$\sigma_t^2 = \omega_0 + \sum_{j=1}^{\infty} \psi_{0j}\epsilon_{t-j}^2, \quad \text{almost surely } (a.s.), \tag{2}$$

$$\omega_0 > 0, \psi_{0j} \geq 0\,(j \geq 1), \quad \sum_{j=1}^{\infty} \psi_{0j} < \infty, \tag{3}$$

where z_t is a sequence of independently identically distributed ($i.i.d.$) unobservable random variables and $\mathbb{Z} = \{t : t = 0, \pm 1, ...\}$. The coefficients ω_0, γ_0,

[1]The first author's research was supported by a Leverhulme Trust Personal Research Professorship and ESRC Grant R000238212 We thank Fabrizio Iacone for help with the numerical calculations referred to in Section 2.
AMS 1991 Subject Classification. Primary 62M10, Secondary 62F12.
Key words and phrases. ARCH(∞) models, pseudo-maximum likelihood estimation, asymptotic inference.

ψ_{0j} are unknown, but we know functions $\psi_j(\zeta)$ of the $r \times 1$ vector ζ, where $r < \infty$, such that for some unknown ζ_0, $\psi_j(\zeta_0) = \psi_{0j}$, $j \geq 1$. The case when γ_0 is known, e.g. $\gamma_0 = 0$, is covered by a simplified version of our treatment. If the x_t were instead unobserved regression errors we have $\gamma_0 = 0$ but would then need to replace x_t by least squares regression residuals in what follows; the details of this extension would be relatively straightforward. On the other hand another relatively straightforward extension of our theory would cover the simultaneous estimation of regression parameters, ω_0 and the ψ_{0j}, after replacing γ_0 by a more general parametric function; in general an efficiency gain is afforded by simultaneous estimation.

To discuss possible choices of the $\psi_j(\lambda)$, define

$$\psi(z;\zeta) = \sum_{j=1}^{\infty} \psi_j(\zeta) z^j. \tag{4}$$

Consider the class of functions

$$\psi(z;\zeta) = c(\zeta)\left\{1 - \frac{a(z;\zeta)}{b(z;\zeta)}(1-z)^{d(\zeta)}\right\}, \tag{5}$$

where $c(\zeta)$ and $d(\zeta)$ are known functions of ζ, and $a(z;\zeta)$ and $b(z;\zeta)$ are polynomials in z of known degrees whose coefficients are known functions of ζ, which have no zeros in common. With ζ_i denoting the ith element of ζ, suppose in particular that $b(z;\zeta)$ is given by

$$\begin{aligned} b(z;\zeta) &= 1, \text{ if } n = 0, \\ &= 1 - \sum_{j=1}^{n} \zeta_{j+m} z^j, \zeta_{j+m} > 0,\ j = 1, ..., n, \text{ if } n > 0, && (6) \\ b(z;\zeta) &\neq 0, |z| \leq 1. && (7) \end{aligned}$$

If for $m > 0$ we also take

$$\begin{aligned} d(\zeta) &\equiv 0, && (8) \\ c(\zeta) &\equiv 1, && (9) \\ a(z;\zeta) &= b(z;\zeta) - \sum_{j=1}^{m} \zeta_j z^j, \zeta_j > 0,\ j = 1, ..., m, && (10) \end{aligned}$$

we have

$$\psi(z;\zeta) = \left\{ \sum_{j=1}^{m} \zeta_j z^j \right\} / \left\{ 1 - \sum_{j=1}^{n} \zeta_{j+m} z^j \right\} \tag{11}$$

with $r = m + n$. For $m \geq 1, n \geq 1$, this is the GARCH(n, m) model of Bollerslev (1987) when

$$a(z;\zeta) \neq 0, \ \ |z| \leq 1,$$

and the IGARCH(n, m) model of Engle and Bollerslev (1986) when

$$a(1;\zeta) = 0.$$

(Often in the literature these are both referred to as GARCH.) It reduces to the ARCH(n) model of Engle (1982) when $m = 0$. In these models $\psi_j(\zeta)$ decays exponentially in j. Nelson (1990), Bougerol and Picard (1992a,b), Davis, Mikosch and Basrak (1999), Mikosch and Starica (2000) and Berkes, Horvath and Kokoszka (2001) have investigated theoretical properties of GARCH(n, m) and IGARCH(m, n) processes.

If we instead combine (6) and (7) with

$$d(\zeta) = \zeta_{m+n+1} \in (0, 1), \tag{12}$$

$$\begin{aligned} a(z;\zeta) &= 1, \text{ if } \ n = 0, \\ &= 1 - \sum_{j=1}^{m} \zeta_j z^j, \quad \zeta_j > 0, \ j = 1, ..., m, \text{ if } m > 0, \end{aligned} \tag{13}$$

so $r = m + n + 1$, we have the FIGARCH(n, ζ_{m+n+1}, m) model of Baillie, Bollerslev and Mikkelsen (1996), for $m \geq 0, n \geq 0$. When $\omega = 0$, where ω is any admissible value of ω_0, such specifications fall in a class of ARCH(∞) models earlier considered by Robinson (1991), who employed this class in testing for the null of no-ARCH against long memory alternatives; see also Ding and Granger (1996). Such models are motivated by the slow decay in autocorrelations of squared returns found in a variety of financial time series (e.g. Whistler, 1990, Ding, Granger and Engle, 1993). Giraitis, Kokoszka and Leipus (2000), Zaffaroni (2003) have investigated theoretical properties of ARCH(∞) processes. For $\omega_0 > 0$, as required in the FIGARCH specification, ϵ_t cannot have finite variance. The 'fractional' parameter $d(\zeta)$ should not be identified too closely with the fractional parameter d arising in fractional

ARIMA (FARIMA) models; in particular a FARIMA is covariance stationary for $0 < d < 1/2$, whereas x_t is not covariance stationary for all $d(\zeta) > 0$.

A modified version of FIGARCH combines (6), (7), (12) and (13) with

$$c(\zeta) = \zeta_{m+n+2} \in (0,1), \tag{14}$$

so $r = m+n+2$. We call this 'mFIGARCH(n, ζ_{m+n+1}, m)', for $m \geq 0, n \geq 0$. The only difference from the FIGARCH specification is the replacement of (9) by (14). The scale parameter ζ_{m+n+2} may afford some additional flexibility, but more importantly, under (9) we have

$$\psi(1;\zeta) = \sum_{j=1}^{\infty} \psi_j(\zeta) = 1, \text{ all } \zeta, \tag{15}$$

whereas under (14)

$$\psi(1;\zeta) = \sum_{j=1}^{\infty} \psi_j(\zeta) < 1, \text{ all } \zeta. \tag{16}$$

Notice also that (15) holds for IGARCH models, but (16) for GARCH models. Under (15), ϵ_t has infinite variance and checking one of our conditions in the following section (Assumption F) for FIGARCH and IGARCH models is problematic. Under (16), ϵ_t has finite variance and we are able to give a relatively convincing verification of our conditions in case of mFIGARCH and GARCH models. Despite the differences between moment properties across these models of interest, our theoretical treatment for the general class of models which we cover does not require ϵ_t to have finite variance, recognizing evidence of fat-tailedness found in much financial data.

As with FIGARCH, the $\psi_j(\zeta)$ for mFIGARCH decay only at hyperbolic rate, like $j^{-\zeta_{m+n+1}-1}$. This suggests a more direct modelling of this behaviour, indicated most simply by the 'power law' model

$$\psi_j(\zeta) = \zeta_2 j^{-\zeta_1 - 1}, \quad \zeta_1 > 0, \ \zeta_2 > 0, \tag{17}$$

which, unlike FIGARCH and mFIGARCH, imposes no upper bound on the rate of decay (cf (12)).

We wish to estimate the $(r+2) \times 1$ vector $\theta_0 = (\omega_0, \gamma_0, \zeta_0')'$ on the basis of observations x_t, $t = 1, ..., T$. For Gaussian z_t an approximate maximum likelihood estimate is defined as follows. Denote by $\theta = (\omega, \gamma, \zeta')'$ any admissible value of θ_0 and define, for $t \in \mathbb{Z}$,

$$\begin{aligned} \epsilon_t(\gamma) &= x_t - \gamma, \qquad \hat{\sigma}_t^2(\theta) = \omega + \sum_{i=1}^{t-1} \psi_i(\zeta)\epsilon_{t-i}^2(\gamma), \\ \hat{q}_t(\theta) &= \frac{\epsilon_t^2(\gamma)}{\hat{\sigma}_t^2(\theta)} + \ln\, \hat{\sigma}_t^2(\theta), \quad \hat{Q}_T(\theta) = T^{-1}\sum_{t=1}^{T} \hat{q}_t(\theta), \end{aligned}$$

the hats referring to the truncation due to x_t, $t \leq 0$, being unobservable. Let Θ be a prescribed compact subset of R^{r+2} and define

$$\hat{\theta}_T = \arg\min_{\theta \in \Theta} \hat{Q}_T(\theta).$$

We do not assume Gaussianity in our asymptotic theory, and thence refer to $\hat{\theta}_T$ as a pseudo-maximum likelihood estimate (PMLE).

We discuss work on the weak consistency and asymptotic normality (with convergence rate $T^{1/2}$) of $\hat{\theta}_T$ and other estimates. Such properties were first established by Weiss (1986) in the ARCH(n) model, while Lee and Hansen (1994), Lumsdaine (1996) dealt with GARCH$(1,1)$ and IGARCH$(1,1)$ models. The proofs of the latter two sets of authors make significant use of the simple structure of their processes, and do not readily extend to the GARCH(n,m) for general m,n. Giraitis and Robinson (2001) were recently able to cover general GARCH(n,m) processes, and more general members of the class (1)-(3), but not for $\hat{\theta}_T$, rather for Whittle estimates based on the squared observations x_t^2 (with γ_0 known to be zero). Such estimates are asymptotically equivalent to least squares regression of x_t^2 on past x_{t-s}^2, $s > 0$, which is computationally especially simple in case of ARCH(n) models (see Engle, 1982); they were also considered by Bollerslev (1987) in case of GARCH(n,m) models. However they have a number of disadvantages, as indeed discussed by Giraitis and Robinson (2001): they are not only asymptotically inefficient under Gaussian z_t but they are never asymptotically efficient; they require finiteness of fourth moments of x_t for consistency and of eighth moments for asymptotic normality, which are considered unacceptable by financial analysts dealing with financial data; their limit covariance matrix is relatively complicated to estimate; they are less well motivated in ARCH models than in stochastic volatility models such as those of Taylor (1986), Robinson and Zaffaroni (1997, 1998), Harvey (1998), Breidt et al

(1998), Zaffaroni (2002) where the actual likelihood is computationally less tractable, while Whittle estimation also plays a less special role in the short-memory-in-x_t^2 ARCH models of Giraitis and Robinson (2001) than in the long-memory-in-x_t^2 models of the previous four references as it entails automatic 'compensation' for possible lack of square-integrability of the spectrum of x_t^2. Mikosch and Straumann (2002) have shown that a finite fourth moment is necessary for consistency of Whittle estimates, and that even with finite fourth moment convergence rates are slowed by fat tails. Hall and Yao (2003) have established that the asymptotic distribution of the PMLE $\hat{\theta}_T$ for GARCH(n, m) models is non-normal when the distribution of the z_t^2 is in the domain of attraction of a stable law with exponent between 1 and 2. Berkes, Horvath and Kokoszka (2001) have established the strong consistency and asymptotic normality of the PMLE $\hat{\theta}_T$ for GARCH(n, m) and IGARCH(n, m) models, their treatment entailing ϵ_t having unconditional (fractional) moment of only small degree. Berkes, Horvath and Kokoszka (2001) also comment on some of the work mentioned above, as well as related work of Jeantheau (1998), Comte and Lieberman (2000). Another recent contribution, of Ling and McAleer (2001), considers the PMLE within the framework of a vector ARMA process with GARCH(n, m) errors; they establish weak consistency in case $E\epsilon_t^2 < \infty$ and asymptotic normality in case $E\epsilon_t^6 < \infty$.

The present paper establishes strong consistency and asymptotic normality for $\hat{\theta}_T$ under conditions that cover a wide variety of parametric forms of (1)-(3), comprising both exponentially and hyperbolically decaying coefficients $\psi_j(\zeta)$. The following section lists our assumptions, with discussion. Section 3 presents the main results, with partial proof details, the remainder being in the form of a series of lemmas contained in Section 4.

2 Assumptions

We list first a series of regularity conditions. With some abuse of notation we shall write u^v in place of $|u|^v$ even when u is negative and v is non-integral. Define

$$\mathcal{N}_\varepsilon(\zeta) = \left\{ \widetilde{\zeta} : \left\| \widetilde{\zeta} - \zeta \right\| \leq \varepsilon;\ \widetilde{\zeta} \in \Upsilon \right\},\ \Psi^{(s)}(\zeta) = \sum_{j=1}^{\infty} \psi_j^s(\zeta),\ \psi_j^{(1)}(\zeta) = \frac{\partial \psi_j(\zeta)}{\partial \zeta},$$

and denote by K a generic positive constant.

Assumption $A(q)$**.** *The* z_t *are symmetric i.i.d. variates with* $Ez_0 = 0$, $Ez_0^2 = 1,\ Ez_0^q < \infty$ *and probability density function* $f(z)$ *satisfying*

$$f(z) = O\left(L(z^{-1})z^b\right), \quad \textit{as } z \to 0^+,$$

for $b > -1$ *and a function* L *that is slowly varying at the origin.*

Assumption B**.** *There exist* $\omega_L, \omega_U, \gamma_L, \gamma_U$ *such that* $0 < \omega_L < \omega_U < \infty$, $-\infty < \gamma_L < \gamma_U < \infty$, *and a compact set* $\Upsilon \in R^r$,*such that* $\Theta = [\omega_L, \omega_U] \times [\gamma_L, \gamma_U] \times \Upsilon$.

Assumption C**.** θ_0 *is an interior point of* Θ.

Assumption D**.** *For some* $\underline{d} > 0$,

$$\sup_{\zeta \in \Upsilon} \psi_j(\zeta) \leq K j^{-\underline{d}-1}. \tag{18}$$

For some $\eta > 0$ and all $\zeta \in \Upsilon$, $t \geq 1$

$$\max_{j \geq t} \ \psi_j(\zeta) j^{1+\eta} \leq K \psi_t(\zeta) t^{1+\eta}. \tag{19}$$

For all $n < \infty$ *there exists* $\eta_n > 0$ *such that*

$$\inf_{1 \leq j \leq n, j \notin S} \ \inf_{\zeta \in \Upsilon} \ \psi_j(\zeta) \geq \eta_n > 0, \tag{20}$$

where $S = \left\{ j : \psi_j(\zeta) = 0, \text{ all } \zeta \in \Upsilon \right\}$.

Assumption E**.** *For all* $\eta > 0$ *there exists* $\varepsilon > 0$ such that, for all $\zeta \in \Upsilon$, $j \geq 1$

$$\sup_{\widetilde{\zeta} \in \mathcal{N}_\varepsilon(\zeta)} \ \psi_j(\widetilde{\zeta}) \leq K \psi_j(\zeta)^{1-\eta}. \tag{21}$$

Assumption F**.** *For*

$$\rho \in ((\underline{d} + 1)^{-1}, 1] \tag{22}$$

we have

$$Ez_0^{2\rho} \Psi^{(\rho)}(\zeta_0) < 1. \tag{23}$$

Assumption $G(l)$**.** *For all* $j \geq 1$, $\psi_j(\zeta)$ *has continuous* k*th derivative on* Υ such that

$$\left| \frac{\partial^k \psi_j(\zeta)}{\partial \zeta_{i_1} ... \partial \zeta_{i_k}} \right| \leq K \psi_j(\zeta)^{1-\eta} \tag{24}$$

for all $\eta > 0$ and all $i_j = 1, ..., r$, $j = 1, ..., k$, $k \leq l$.

Assumption H**.** *For each* $\zeta \in \Upsilon$ there exist integers $j_i(\zeta)$, $i = 1, ..., r$, *such that* $1 \leq j_1(\zeta) < ... < j_r(\zeta) < \infty$ *and the matrix*

$$\left\{ \psi_{j_1}^{(1)}(\zeta), ..., \psi_{j_r}^{(1)}(\zeta) \right\} \tag{25}$$

has full rank.

Assumption I**.** *There exists*

$$d_0 > \frac{1}{2} \tag{26}$$

such that

$$\psi_{0j} \leq K j^{-1-d_0}, \tag{27}$$

and (23) holds for

$$\rho \in (4/(2d_0 + 3), 1]. \tag{28}$$

It is useful to discuss these assumptions in relation to the following rate specifications of $\psi_j(\zeta)$ and its derivatives. First suppose that, for all $\zeta \in \Upsilon$, $j \geq 1$

$$\beta^{j-1}(\zeta)/K \leq \psi_j(\zeta) \leq K \, \beta^{j-1}(\zeta), \tag{29}$$

for a boundedly differentiable function $\beta(\zeta) \in (0, 1)$, and

$$\left| \frac{\partial^k \psi_j(\zeta)}{\partial \zeta_{i_1} ... \partial \zeta_{i_k}} \right| \leq K \, j^k \, \beta^j(\zeta) \, , \; k \leq l. \tag{30}$$

(29) and (30) hold for GARCH and IGARCH models. Second, suppose that, for all $\zeta \in \Upsilon$,

$$j^{-d(\zeta)-1}/K \leq \psi_j(\zeta) \leq K \, j^{-d(\zeta)-1}, \tag{31}$$

for a boundedly differentiable function $d(\zeta) > 0$, and

$$\left| \frac{\partial^k \psi_j(\zeta)}{\partial \zeta_{i_1} \ldots \partial \zeta_{i_k}} \right| \leq K \left\{ \ln(j+1) \right\}^k j^{-d(\zeta)-1}, \quad k \leq l. \tag{32}$$

(31) and (32) hold for FIGARCH, mFIGARCH and 'power law' models.
We remark on the assumptions as follows.

1. In Assumption $A(q)$ we will require $q > 2$ for strong consistency of $\widehat{\theta}_T$ and $q = 4$ for asymptotic normality. Symmetry of z_t simplifies the exposition but is not essential to our results.

2. To satisfy Assumption B, we choose Υ such that

$$0 < \beta_L \leq \beta(\zeta) \leq \beta_U < 1, \; 0 < d_L \leq d(\zeta) \leq d_U < \infty,$$

in cases (29) and (31) respectively, though for FIGARCH or mFIGARCH we take $d_U < 1$.

3. In Assumption D, (18) and (19) are satisfied in both cases (29) and (31), In (29) we can take $\underline{d}$ arbitrarily large, and in (31), $\underline{d} = d_L$. The restriction (20) implies that, if $\psi_j(\zeta_0) = 0$, we must know this and specify the model accordingly, for example we cannot over-parameterize an ARCH(n) model. In our proofs we therefore simplify matters by acting as if S is empty.

4. Assumption E can be checked as follows. Under (29),

$$\sup_{\mathcal{N}_\varepsilon(\zeta)} \psi_j(\tilde{\zeta}) \leq K \left\{ \sup_{\mathcal{N}_\varepsilon(\zeta)} \beta(\tilde{\zeta}) \right\}^j$$

and by bounded differentiability of $\beta(\zeta)$,

$$\sup_{\mathcal{N}_\varepsilon(\zeta)} \beta(\tilde{\zeta}) \leq \beta(\zeta) \left\{ 1 + \frac{K\varepsilon}{\beta_L} \right\} \leq K \beta(\zeta)^{1-\eta}$$

for $\varepsilon \leq \beta_L \left(\beta_U^{-\eta} - 1 \right) / K$. Thus

$$\sup_{\mathcal{N}_\varepsilon(\zeta)} \psi_j(\tilde{\zeta}) \leq K \left\{ \beta^j(\zeta) \right\}^{1-\eta} \leq K \psi_j(\zeta)^{1-\eta}.$$

Under (31),

$$\sup_{\mathcal{N}_\varepsilon(\zeta)} \psi_j(\tilde{\zeta}) \leq K j^{-\inf_{\mathcal{N}_\varepsilon(\zeta)} d(\tilde{\zeta})-1}$$

and by bounded differentiability of $d(\zeta)$,

$$\inf_{\mathcal{N}_\varepsilon(\zeta)} d(\tilde{\zeta}) \geq d(\zeta) - K\varepsilon \geq \{d(\zeta)+1\}(1-\eta) - 1$$

for $\varepsilon \leq \eta(d_L+1)/K$. Thus

$$\sup_{\mathcal{N}_\varepsilon(\zeta)} \psi_j(\tilde{\zeta}) \leq K\left\{j^{-d(\zeta)-1}\right\}^{1-\eta} \leq K\psi_j(\zeta)^{1-\eta}.$$

5. The bounds (18) and (27) together imply $d_0 \geq \underline{d}$, while (28) entails no additional restriction over (22) when $d_0 \geq 2\underline{d} + 1/2$, and (18) and (22) together imply

$$\sum_{j=1}^{\infty} \sup_{\zeta \in \Upsilon} \psi_j^s(\zeta) < \infty,\ s \geq \rho. \tag{33}$$

6. Assumption F is satisfied whenever $\Psi^{(1)}(\zeta_0) < 1$ (16), as in GARCH and mFIGARCH models, and here we may choose $\rho = 1$, whence from Lemma 2 (see Section 4) $E\epsilon_0^2 < \infty$ is implied. The other possibility admitted by (23) is $\Psi^{(1)}(\zeta_0) = 1$ (15), which arises in IGARCH and FIGARCH models. Here we must have $\rho < 1$, whence (again from Lemma 2 of Section 4) the milder unconditional moment condition $E\epsilon_0^{2\rho} < \infty$ is implied. Whereas $Ez_0^{2\rho} < 1$ for $\rho < 1$, and $Ez_0^{2\rho}$ decreases in ρ, $\Psi^{(\rho)}(\zeta_0)$ increases in ρ. Analytic verification of (23) for given ζ_0, ρ seems impossible, and numerical verification highly problematic due to the slow convergence of the ψ_{0j}^{ρ}. However, consider the family of densities

$$f(z) = \exp\left[-\{\alpha(\gamma)\,|z|\}^{1/\gamma}\right] / \{2\gamma\Gamma(\gamma)\alpha(\gamma)\} \tag{34}$$

for $\gamma > 0$, where $\alpha(\gamma) = \{\Gamma(\gamma)/\Gamma(3\gamma)\}^{1/2}$. (Such a family was also considered by Nelson (1991) to model the innovation of the exponential GARCH model.) We have $Ez_0 = 0$, $Ez_0^2 = 1$ as necessary, and assume $A(q)$ is satisfied for all $q > 0$ with

$$Ez_0^{2\rho} = \frac{\Gamma((2\rho+1)\gamma)}{\Gamma(\gamma)^{1-\rho}\Gamma(3\gamma)^{\rho}}.$$

In case $\gamma = 0.5$, (34) is the normal density, for which $\widehat{\theta}_T$ is asymptotically efficient. Here $E(z_0^{2\rho}) = 2^\rho\Gamma(\rho + .5)/\sqrt{\pi}$, and numerical calculations for the FIGARCH$(0, \zeta_1, 0)$ model, in which $\psi_j(\zeta) = \Gamma(j - \zeta_1)/\left\{\Gamma(\zeta_1)\Gamma(j+1)\right\}$, cast doubt on (23). In case $\gamma = 1$, (34) is the Laplace density, with $E(z_0^{2\rho}) = 2^{\rho-1}\Gamma(2\rho+1)$. As γ increases, $Ez_0^{2\rho}$ can be made small for fixed $\rho < 1$, for example with $p = 0.95$ it is 0.64 when $\gamma = 10$ and 0.42 when $\gamma = 20$.

7. Assumption $G(l)$ holds under (30) and (32) because, for all n and $\zeta \in \Upsilon$, $\beta(\zeta)^{-j\eta} \geq \beta_U^{-j\eta} \geq j^n$ and $j^{\{d(\zeta)+1\}\eta} \geq j^{(d_L+1)\eta} \geq (\log j)^\eta$ for any $\eta > 0$ and j sufficiently large.

8. Assumption H is an identifiability condition, used to prove that $\hat{Q}_T(\theta)$ has, asymptotically, a unique minimum. We check it first for mFIGARCH models. Note that the restrictions on ζ in (6), (7), (10) and (12)-(14) are assumed to hold throughout Υ; thus Υ excludes ζ such that either both of $a(z;\zeta)$ and $b(z;\zeta)$ are over-specified (which would require $\zeta_m = 0$ and/or $\zeta_{m+n} = 0$), in particular, as in ARMA models, ones such that $a(z;\zeta)$ and $b(z;\zeta)$ have one or more zeroes in common (which would require $\zeta_m = \zeta_{m+n} = 0$). We suppose that there are no restrictions linking elements of ζ, because a proof for this case implies that Assumption H will hold in the presence of restrictions. Fix ζ, and for brevity drop the ζ argument and write $d = \zeta_{m+n+1}$, $c = \zeta_{m+n+2}$. We have

$$\frac{\partial\psi(z)}{\partial\zeta_k} = -cz^k\phi(z), \qquad k = 1, ..., m, \tag{35}$$

$$\frac{\partial\psi(z)}{\partial\zeta_k} = cz^k\rho(z)\phi(z), \;\; k = m+1, ..., m+n, \tag{36}$$

$$\frac{\partial\psi(z)}{\partial d} = -ca(z)\log(1-z)\phi(z), \tag{37}$$

$$\frac{\partial\psi(z)}{\partial c} = 1 - a(z)\phi(z), \tag{38}$$

where

$$\phi(z) = b(z)^{-1}(1-z)^d, \qquad \rho(z) = b(z)^{-1}a(z).$$

Suppose that $m + n > 0$. Choose $j_i = i$ for $i = 1, ..., m+n$, leaving j_{m+n+1} and j_{m+n+2} to be determined subsequently. Denote (25) by U, and partition it in the ratio $m+n:2$, calling its (i,j)th submatrix U_{ij}. We first show that the $(m+n)\times(m+n)$ matrix U_{11} is non-singular. Write I_m for the m-rowed identity matrix, R for the $n\times(m+n)$ matrix with (i,j)th element ρ_{j-i}, and S for the $n+m\times n+m$ matrix with (i,j)th element $-c\phi_{j-i}$, where $\phi_j=\rho_j=0$ for $j<0$, $\phi_0=\rho_0=1$, and for $j>0$, ϕ_j and ρ_j are respectively given by

$$\phi(z)=1+\sum_{j=1}^{\infty}\phi_j z^j,\ \rho(z)=1+\sum_{j=1}^{\infty}\rho_j z^j,$$

these series converging absolutely in view of (7). Noting that $\psi_j^{(1)}$ is given by $(\partial/\partial\zeta)\psi(z)=\sum_{j=1}^{\infty}\psi_j^{(1)}z^j$, and applying (35) and (36), we find that when $m\geq 1$, the first m rows of U_{11} can be written $(I_m,O)S$, where O is the $m\times n$ matrix of zeroes, and when $n\geq 1$, the last n rows of U_{11} can be written RS. Now S is upper-triangular with non-zero diagonal elements. Thus for $n=0$, $U_{11}=S$ is non-singular, while for $n\geq 1$ U_{11} is non-singular if and only if the matrix, R_2, consisting of the last n column of R, is non-singular. We have

$$R_2=\begin{bmatrix}\rho_m & \rho_{m+1} & \cdots & \rho_{m+n-1}\\ \vdots & & & \vdots\\ \rho_{m-n+1} & \rho_{m-n+2} & \cdots & \rho_m\end{bmatrix}.$$

Suppose $m\geq n$. Then R_2 is singular if and only if the ρ_j, $j=m,...,m+n-1$, are generated by a homogeneous linear difference equation of degree $n-1$, that is if there exist scalars $\lambda_1,...,\lambda_{n-1}$, not all zero, such that

$$\rho_j-\sum_{i=1}^{n-1}\lambda_i\rho_{j-i}=0,\qquad j=m,...,m+n-1.$$

But it follows from (5), (6) and (13) that they are instead generated by the inhomogeneous linear difference equation

$$\rho_j-\sum_{i=1}^{n-1}\zeta_{m+i}\rho_{j-i}=\pi_j,\qquad j=m,...,m+n-1,$$

where $\pi_m=\zeta_{m+n}-\zeta_n$, $\pi_j=\zeta_{m+n}\rho_{j-m}-\zeta_{n+m-j}$ for $j=m+1,...,\min(m+n-1,2m-n)$, and $\pi_j=\zeta_{m+n}\rho_{j-m}$ for $j=2m-n+1,...,m+n-1$, the last case being relevant only when $m+2\leq 2n$. The fact that the π_j are not all zero follows from the initial conditions, also derivable from (5), (6) and (13), and the implication that $a(z)$ and $b(z)$ have no common zeroes. For example when $m = n$ we have $\pi_m = \zeta_{m+n} - \zeta_m \neq 0$. The demonstration that R_2 is non-singular for $m < n$ follows similar lines, while when $m = 0$, R_2 is upper-triangular with unit diagonal elements. Given non-singularity of U_{11}, non-singularity of U follows if $U_{22} - U_{21}U_{11}^{-1}U_{12}$ is non-singular. This can be achieved because of the form of (37) and (38) and because we are free to choose j_{m+n+1} and j_{m+n+2} (depending on ζ) from all integers that exceed $m + n$. This completes verification of Assumption H when $m + n > 0$. In case $m = n = 0$ we can choose $j_1 = 1$, $j_2 = 2$, whence U has determinant $-cd^2/2 \neq 0$. Though FIGARCH, GARCH and IGARCH are strictly not special cases of the mFIGARCH specification, in view of the inequality restrictions (12) and (14), nevertheless the above argument indicates that Assumption H is verified in these cases also. The relevant difference between FIGARCH and mFIGARCH in the present context is that the former contains one less parameter, while for GARCH and IGARCH the only difference in the argument above with respect to $U = U_{11}$ is that I_m is replaced by $-I_m$ and the (i,i)th element of R is replaced by zero, for $i = 1, ..., n$, while the remaining elements are changed in sign. We finally observe that for the 'power law' model (17) we may choose $j_1 = 2$, $j_1 = 3$, say, whence U has determinant $\zeta_2 6^{-\zeta_1 - 1}(\log 3 - \log 2) \neq 0$.

9. Assumption I is not needed for the consistency of $\hat{\theta}_T$, but seems to be needed for the central limit theorem. In (29) it entails no additional restriction because d_0 can be arbitrarily large. In (31) we choose $d_0 = d(\zeta_0)$, and here it rules out FIGARCH and mFIGARCH models with $d_0 = \zeta_{0,m+n+1} \in (0, 1/2]$, and 'power law' models with $d_0 = \zeta_{0,1} \in (0, 1/2]$. There are good reasons to suppose that when $d_0 \in (0, 1/2]$ the asymptotic bias in $\hat{\theta}_T$ is of order at least $T^{-1/2}$, and thus prevents a central limit theorem centered at θ_0. This is due to replacement of the uncomputable $\sigma_t^2(\theta)$ by $\hat{\sigma}_t^2(\theta)$ in the objective function, entailing a truncation whose error varies inversely with d_0. In particular, looking ahead to the proof of Theorem 2 in the next section, the source of bias is found in the term $H^{-1}B_{1T}$, where H is positive definite and

$$B_{1T} = -\sum_{t=1}^{T} z_t^2 \left\{ \frac{\hat{\sigma}_t^{2(1)}}{\hat{\sigma}_t^4} \right\} \left(\sigma_t^2 - \hat{\sigma}_t^2 \right), \tag{39}$$

where $\hat{\sigma}_t^2 = \widehat{\sigma}_t^2(\theta_0)$, $\hat{\sigma}_t^{2(1)} = (\partial/\partial\theta)\, \widehat{\sigma}_t^2(\theta_0)$. For B_{1T} not to affect the central limit theorem, we must have $B_{1T} = o_p(T^{1/2})$, and this is shown under Assumption I in the proof of Theorem 2. To consider, more informally, the possibility of a lower bound for B_{1T} for any $d_0 > 0$, note that

$$\sigma_t^2 - \hat{\sigma}_t^2 = \sum_{j=t}^{\infty} \psi_{0j} \epsilon_{t-j}^2. \tag{40}$$

This is nonnegative, and in (31) we have $\psi_{0j} \geq j^{-d_0-1}/K$, suggesting that (40) exceeds t^{-d_0}/K as $t \to \infty$ with probability approaching one. Also $Ez_0^2 > 0$, so the only possibility that $B_{1T} = o_p(T^{1/2})$ for $d_0 \leq 1/2$ would be due to 'cancellation' produced by the factor in braces in (39), where

$$\hat{\sigma}_t^{2(1)} = (1, -2\sum_{j=t}^{\infty} \psi_{0j} \epsilon_{t-j}, \sum_{j=t}^{\infty} \psi_{0j}^{(1)\prime} \epsilon_{t-j}^2)'. \tag{41}$$

The second element of $\hat{\sigma}_t^{2(1)}/\hat{\sigma}_t^4$ is thus an odd function, and the corresponding element of B_{1T} is thence $o_p(T^{1/2})$ for all $d_0 > 0$. However, the first element of $\hat{\sigma}_t^{2(1)}/\hat{\sigma}_t^4$ is an even function and while the $\psi_{0j}^{(1)}$ can have elements of either sign there seems no reason why this should lead to such cancellation that would sufficiently lower the order of magnitude of the last r elements of (39). Notice that corresponding truncation in Whittle estimation of fractional ARIMA models does not lead to a corresponding problem, essentially because the truncation error has zero mean and variance which decays slowly, but sufficiently fast, as $t \to \infty$. The latter observation relates to Baillie, Bollerslev and Mikkelsen's (1996) practical solution to the truncation problem, namely, replacing the $\epsilon_{t-j}^2(\gamma)$ in

$$\sigma_t^2(\theta) = \omega + \sum_{j=1}^{\infty} \psi_j(\zeta) \epsilon_{t-j}^2(\gamma), \tag{42}$$

for $j \geq t$, by the sample variance, s_T^2, of x_t. Some truncation is in general still necessary for practical implementation, but the practitioner is free to truncate as remotely as computational restrictions permit, and Baillie, Bollerslev and Mikkelsen (1996) provide numerical evidence that the resulting error can be very small. They are also careful in their Monte Carlo study to minimize the effects of truncation in data generation, and there is no appreciable difference between their biases for $d_0 = 0.75$ and $d_0 = 0.5$, the knife-edge case which Assumption I barely excludes - they did not consider

$d_0 \in (0, 0.5)$. Unfortunately, as they indicate, s_T^2, like ϵ_t^2, does not have finite expectation in FIGARCH models, and $E(\varepsilon_t^2 - s_T^2)^\rho$ is non-zero for $\rho < 1$. We have thus not been able to show that this modification solves the asymptotic bias problem in the central limit theorem for $d_0 \leq 1/2$ (or that it does not adversely affect other aspects of our proof). An alternative 'solution' to the bias problem simply replaces σ_t^2 in the model (2) by $\hat{\sigma}_t^2$, so that the process starts at $t = 0$ rather than $t = -\infty$, as in nonstationary models. Whether or not such a model has appeal, the details of the asymptotic theory would be substantially affected, because x_t would then be stationary in only an asymptotic sense.

3 Main Results

We find it convenient to present the proofs of both consistency and asymptotic normality of $\widehat{\theta}_T$ in two parts. First we establish these results (in Propositions 1 and 2) for the infeasible estimate $\widetilde{\theta}_T$ given by

$$\widetilde{\theta}_T = \arg\min_{\theta\in\Theta} Q_T(\theta),$$

where

$$Q_T(\theta) = T^{-1} \sum_{t=1}^{T} q_t(\theta),\ q_t(\theta) = \frac{\epsilon_t^2(\gamma)}{\sigma_t^2(\theta)} + \ln \sigma_t^2(\theta),$$

with $\sigma_t^2(\theta)$ given in (42). Then in Theorems 1 and 2 we show that $\sigma_t^2(\theta)$ can be replaced by $\hat{\sigma}_t^2(\theta)$.

Proposition 1 *For some $\delta > 0$, let Assumptions $A(2+\delta)$, B, C, D, E, F, $G(1)$ and H hold. Then*

$$\widetilde{\theta}_T \to \theta_0 \ \ a.s. \ \ as\ T \to \infty.$$

Proof. The proof follows from Lemmas 7 and 10 and a standard proof of strong consistency of implicitly-defined extremum estimates. ■

Theorem 1 *For some $\delta > 0$, let Assumptions $A(2+\delta)$, B, C, D, E, F, $G(1)$ and H hold. Then*

$$\hat{\theta}_T \to \theta_0 \ \ a.s. \ \ as\ T \to \infty.$$

.

Proof. The proof follows as in that of Proposition 1 and using also Lemma 8. ■

Let e_2 be the second column of the I_{r+2}-rowed identity matrix, and define

$$\begin{aligned}
\chi_t(\theta) &= \frac{\epsilon_t^2(\theta)}{\sigma_t^2(\theta)}, \quad \nu_t(\theta) = \frac{\partial \epsilon_t(\gamma)}{\partial \theta} = -2\epsilon_t(\gamma)e_2, \\
\sigma_t^{2(1)}(\theta) &= \frac{\partial \sigma_t^2(\theta)}{\partial \theta}, \quad \tau_t(\theta) = \frac{\partial \log \sigma_t^2(\theta)}{\partial \theta} = \frac{\sigma^{2(1)}(\theta)}{\sigma_t^2(\theta)}, \\
g_t(\theta) &= \tau_t(\theta)\tau_t'(\theta)\left\{\chi_t(\theta) - 1\right\}^2 + \sigma_t^{-4}(\theta)\nu_t(\theta)\nu_t'(\theta) \\
&\quad + \sigma_t^{-2}(\theta)\left\{1 - \chi_t(\theta)\right\}\left\{\tau_t(\theta)\nu_t'(\theta) + \nu_t(\theta)\tau_t'(\theta)\right\} \\
h_t(\theta) &= \sigma_t^{-2}(\theta)\frac{\partial^2 \sigma_t^2(\theta)}{\partial\theta\partial\theta'}\left\{1 - \chi_t(\theta)\right\} - \sigma_t^{-2}(\theta)\left\{\nu_t(\theta)\tau_t'(\theta) + \tau_t(\theta)\nu_t'(\theta)\right\} \\
&\quad + \sigma_t^{-2}(\theta)\frac{\partial^2\, \epsilon_t^2(\gamma)}{\partial\theta\partial\theta'} + \tau_t(\theta)\tau_t'(\theta)\left\{\chi_t(\theta) - 1\right\}, \\
G(\theta) &= Eg_0(\theta), \qquad H(\theta) = Eh_0(\theta).
\end{aligned}$$

Also define

$$\begin{aligned}
M(\theta) &= E\left\{\tau_0(\theta)\tau_0'(\theta)\right\}, \qquad P(\theta) = E\left\{\sigma_0^{-4}(\theta)\nu_0(\theta)\nu_0'(\theta)\right\} = 4E\left\{\sigma_0^{-4}(\theta)\epsilon_0^2(\gamma)\right\}e_2e_2', \\
M &= M(\theta_0), \qquad P(\theta_0) = 4\sigma_t^{-2}e_2e_2',
\end{aligned}$$

where we have

$$G = G(\theta_0) = (2+\kappa)M + P, \qquad H = M + P/2,$$

in which κ is the fourth cumulant of z_t. In case γ_0 is known (for example, to be zero), we have $G = (2+\kappa)M$, $H = M$. In case z_t is Gaussian, $\kappa = 0$. The formula for G uses the symmetry of z_t but holds if z_t only has zero third cumulant; indeed our results would allow a non-zero third cumulant with G redefined accordingly.

Proposition 2 *Let Assumptions A(4), B, C, D, E, F, G(3) and H hold. Then*

$$T^{\frac{1}{2}}\left(\tilde{\theta}_T - \theta_0\right) \to_d N(0, H^{-1}G\,H^{-1}), \quad \textit{as } T \to \infty.$$

Proof. Write

$$Q_T^{(1)}(\theta) = \frac{\partial Q_T(\theta)}{\partial \theta} = \frac{1}{T}\sum_{t=1}^{T} u_t(\theta),$$

where

$$u_t(\theta) = \tau_t(\theta)(1 - \chi_t^2(\theta)) + \sigma_t^{-2}(\theta)\nu_t(\theta). \tag{43}$$

By the mean value theorem

$$0 = Q_T^{(1)}(\tilde{\theta}_T) = Q_T^{(1)}(\theta_0) + \widetilde{H}_T(\tilde{\theta}_T - \theta_0), \tag{44}$$

where $\widetilde{H}_T$ has as its ith row the ith row of $H_T(\theta) = T^{-1}\sum_{t=1}^{T} h_t(\theta)$ evaluated at $\theta = \tilde{\theta}_T^{(i)}$ where $\left\|\tilde{\theta}_T^{(i)} - \theta_0\right\| \leq \left\|\tilde{\theta}_T^{(i)} - \tilde{\theta}_T\right\|$. Now $u_t(\theta_0) = \tau_t(\theta_0)(1 - z_t^2) - 2e_2 z_t/\sigma_t$ is, by Lemmas 2, 3 and 7 a stationary ergodic martingale difference vector with finite variance, so from Brown (1971) and the Cramer-Wold device, $T^{\frac{1}{2}}Q_T^{(1)}(\theta_0) \to_d N(0, G)$ as $T \to \infty$. Finally, by Lemma 7 and Theorem 1 $\widetilde{H}_T \to_p H$, whence the proof is completed in standard fashion. ■

Define

$$\begin{aligned}
\widehat{u}_t(\theta) &= (\partial/\partial\theta)\widehat{q}_t(\theta), \widehat{g}_t(\theta) = \widehat{u}_t(\theta)\ \widehat{u}_t'(\theta), \quad \widehat{h}_t(\theta) = \frac{\partial^2\widehat{q}_t(\theta)}{\partial\theta\partial\theta'}, \\
\widehat{G}_T(\theta) &= \frac{1}{T}\sum_{t=1}^{T}\widehat{g}_t(\theta), \quad \widehat{H}_T(\theta) = \frac{1}{T}\sum_{t=1}^{T}\widehat{h}_t(\theta).
\end{aligned}$$

Theorem 2 *Let Assumptions* $A(4)$, B, C, D, E, F, $G(3)$, H *and* I *hold. Then*

$$T^{\frac{1}{2}}\left(\hat{\theta}_T - \theta_0\right) \to_d N(0, H^{-1}G\,H^{-1}), \quad \text{as } T \to \infty. \tag{45}$$

and $H^{-1}G\,H^{-1}$ *is strongly consistently estimated by* $\widehat{H}_T^{-1}(\hat{\theta}_T)\widehat{G}_T(\hat{\theta}_T)\widehat{H}_T^{-1}(\hat{\theta}_T)$.

Proof. We have

$$0 = \widehat{Q}_T^{(1)}(\hat{\theta}_T) = \widehat{Q}_T^{(1)}(\theta_0) + \widehat{H}_T(\hat{\theta}_T - \theta_0),$$

where $\widehat{H}_T$ has as its ith row the ith row of $\widehat{H}_T(\theta)$ evaluated at $\theta = \hat{\theta}_T^{(i)}$ where $\left\|\hat{\theta}_T^{(i)} - \theta_0\right\| \leq \left\|\hat{\theta}_T - \hat{\theta}_0\right\|$. Thus from (44)

$$\hat{\theta}_T - \widetilde{\theta}_T = (\widetilde{H}_T^{-1} - \widehat{H}_T^{-1})\widehat{Q}_T^{(1)}(\theta_0) - \widetilde{H}_T^{-1}\left\{\widehat{Q}_T^{(1)}(\theta_0) - Q_T^{(1)}(\theta_0)\right\}.$$

In view of Proposition 2 and Lemma 8, (45) follows on showing that

$$\widehat{Q}_T^{(1)}(\theta_0) - Q_T^{(1)}(\theta_0) = o_p(T^{-1/2}).$$

The left hand side can be written $(B_{1T}+B_{2T}+B_{3T})/T$, where

$$B_{1T}=\sum_{t=1}^{T}z_t^2\,b_{1t},\ B_{2T}=-\sum_{t=1}^{T}(z_t^2-1)b_{2t},\ \ B_{3T}=-2e_2\sum_{t=1}^{T}z_tb_{3t},$$

with

$$b_{1t}=-\frac{\hat{\sigma}_t^{2(1)}(\sigma_t^2-\hat{\sigma}_t^2)}{\hat{\sigma}_t^4},\ b_{2t}=\frac{\sigma_t^{2(1)}}{\sigma_t^2}-\frac{\hat{\sigma}_t^{2(1)}}{\hat{\sigma}_t^2},\ b_{3t}=\frac{\sigma_t^2-\hat{\sigma}_t^2}{\hat{\sigma}_t^2\sigma_t}\ ,$$

and $\hat{\sigma}_t^{2(1)}$ given in (41). We show that $B_{iT}=o_p(T^{1/2})$, $i=1,2,3$.

Henceforth we drop the zero subscript in ψ_{0j}. Using (41) and Assumption $G(l)$

$$\frac{1}{\widehat{\sigma}_t^2}\left\|\hat{\sigma}_t^{2(1)}\right\|\leq\frac{1}{\widehat{\sigma}_t^2}\left\{2\sum_{j=1}^{t-1}\psi_j\left|\epsilon_{t-j}\right|+1+K\sum_{j=1}^{t-1}\psi_j^{1-\eta}\epsilon_{t-j}^2\right\}.$$

Now

$$\sum_{j=1}^{t-1}\psi_j\left|\epsilon_{t-j}\right|\leq\left(\sum_{j=1}^{t-1}\psi_j\epsilon_{t-j}^2\right)^{1/2}\left(\sum_{j=1}^{\infty}\psi_j\right)^{1/2}\leq K\widehat{\sigma}_t$$

so since $\widehat{\sigma}_t\geq\omega_L>0$

$$\widehat{\sigma}_t^{-2}\sum_{j=1}^{t-1}\psi_j\left|\epsilon_{t-j}\right|\leq K\widehat{\sigma}_t^{-1}<\infty.$$

From (19),

$$\psi_k\leq K\psi_j,\ \ k\geq j, \tag{46}$$

so

$$\sum_{j=1}^{t-1}\psi_j^{1-\eta}\epsilon_{t-j}^2\leq K\psi_t^{-\eta}\widehat{\sigma}_t^2.$$

It follows that

$$\left\|\hat{\sigma}_t^{2(1)}\right\|/\widehat{\sigma}_t^2\leq K\psi_t^{-\eta}. \tag{47}$$

On the other hand

$$E(\sigma_t^2-\hat{\sigma}_t^2)^{\rho}\leq K\sum_{j=t}^{\infty}\psi_j^{\rho}E\epsilon_{t-j}^{2\rho}\leq K\sum_{j=t}^{\infty}\psi_j^{\rho}. \tag{48}$$

Thus using (46) again

$$E\left\|b_{1t}\right\|^{\rho} \le K\psi_t^{-\eta}\sum_{j=t}^{\infty}\psi_j^{\rho} \le K\sum_{j=t}^{\infty}\psi_j^{\rho(1-\eta)} \le Kt^{1-\rho(d_0+1)(1-\eta)}.$$

It follows that

$$\begin{aligned} E\|B_{1T}\|^{\rho} &\le K\sum_{t=1}^{T}Ez_0^{2\rho}Eb_{1t}^{\rho} \le KT^{2-\rho(d_0+1)(1-\eta)} \\ &\le KT^{\rho/2-\{1+2(d_0+1)(1-\eta)\}[\rho/2-2/\{1+2(d_0+1)(1-\eta)\}]} = o(T^{1/2}), \end{aligned}$$

using (28) (which requires (26)) and arbitrariness of η. Next, by elementary inequalities

$$E\left\|B_{2T}\right\|^{2\rho} \le K\sum_{t=1}^{T}(Ez_0^{2\rho}+1)E\left\|b_{2t}\right\|^{2\rho} \le K\sum_{t=1}^{T}\left\{E\left\|b_{4t}\right\|^{2\rho}+E\left\|b_{5t}\right\|^{2\rho}\right\},$$

where

$$b_{4t} = \frac{\sigma_t^{2(1)}-\hat{\sigma}_t^{2(1)}}{\sigma_t^2} = \sigma_t^{-2}\left(-2\sum_{j=t}^{\infty}\psi_j\epsilon_{t-j}, 0, \sum_{j=t}^{\infty}\psi_j^{(1)\prime}\epsilon_{t-j}^2\right)^{\prime},\ b_{5t} = \frac{\hat{\sigma}_t^{2(1)}(\sigma_t^2-\hat{\sigma}_t^2)}{\hat{\sigma}_t^2\sigma_t^2}.$$

Now

$$\begin{aligned} \|b_{4t}\| &\le \left(2\sum_{j=t}^{\infty}\psi_j\left|\epsilon_{t-j}\right| + \sum_{j=t}^{\infty}\left\|\psi_j^{(1)}\right\|\epsilon_{t-j}^2\right)/\sigma_t^2. \\ &\le \sigma_t^{-2}\left[2\left\{\sum_{j=t}^{\infty}\psi_j\right\}^{1/2} + \left\{\sum_{j=t}^{\infty}\left(\left\|\psi_j^{(1)}\right\|^2/\psi_j\right)\epsilon_{t-j}^2\right\}^{1/2}\right]\left\{\sum_{j=t}^{\infty}\psi_j\epsilon_{t-j}^2\right\}^{1/2} \\ &\le K\left\{\left(\sum_{j=t}^{\infty}j^{-d_0-1}\right)^{1/2} + \left(\sum_{j=t}^{\infty}\psi_j^{1-2\eta}\epsilon_{t-j}^2\right)^{1/2}\right\} \\ &\le K\left[t^{-d_0/2} + \left\{\sum_{j=t}^{\infty}j^{-(d_0+1)(1-2\eta)}\epsilon_{t-j}^2\right\}^{1/2}\right], \end{aligned}$$

so

$$E\left\|b_{4t}\right\|^{2\rho} \leq Kt^{-\rho d_0} + K\sum_{j=t}^{\infty} j^{-(d_0+1)\rho(1-2\eta)} \leq Kt^{1-(d_0+1)\rho(1-2\eta)}.$$

Thus

$$\sum_{t=1}^{T} E\left\|b_{4t}\right\|^{2\rho} \leq KT^{2-(d_0+1)\rho(1-2\eta)} \leq KT^{\rho-(d_0+2)\{\rho-2/(d_0+2)\}+2(d_0+1)\rho\eta} = o(T^{\rho})$$

from (28) and arbitrariness of η. Also $\|b_{5t}\| \leq K\left\|\hat{\sigma}_t^{2(1)}/\hat{\sigma}_t^2\right\|(\sigma_t^2 - \hat{\sigma}_t^2)^{1/2}$ so from (47) and (48) we have $E\left\|b_{5t}\right\|^{2\rho} \leq Kt^{1-(d_0+1)\rho(1-2\eta)}$, and as before

$$\sum_{t=1}^{T} E\left\|b_{5t}\right\|^{2\rho} = o(T^{\rho})$$

and thence $B_{2T} = o_p(T^{1/2})$. Next we have

$$E\left\|B_{3T}\right\|^{2\rho} \leq KE\left|\sum_{t=1}^{T} z_t b_{3t}\right|^{2\rho} \leq K\sum_{t=1}^{T} Ez_0^{2\rho}Eb_{3t}^{2\rho}. \tag{49}$$

from Von Bahr and Esseen (1965). Now $b_{3t} \leq (\sigma_t^2 - \hat{\sigma}_t^2)^{1/2}\hat{\sigma}_t^{-2}$ so (49) is bounded by

$$K\sum_{t=1}^{T} E\left\{(\sigma_t^2 - \hat{\sigma}_t^2)^{\rho}\hat{\sigma}_t^{-4\rho}\right\} \leq K\sum_{t=1}^{T} E(\sum_{j=t}^{\infty}\psi_j\epsilon_{t-j}^2)^{\rho} \leq KT^{2-\rho(d_0+1)}$$

as before. This is $o(T^{\rho})$, and so $B_{3T} = o_p(T^{1/2})$ by Markov's inequality. It remains to consider the last statement of the theorem, which follows on standard application of Propositions 1 and 2, Theorem 1 and Lemmas 7 and 8. ■

4 Technical Lemmas

Define

$$\sigma_t^{*2}(\theta) = \omega + \sum_{j=1}^{\infty} \psi_j(\zeta)\epsilon_{t-j}^2, \quad \sigma_t^{*2} = \omega_U + \sum_{j=1}^{\infty} \sup_{\zeta \in \Upsilon} \psi_j(\zeta)\epsilon_{t-j}^2.$$

Lemma 1 *Under Assumptions B and D, for all $\theta \in \Theta$, $t \in \mathbb{Z}$*

$$K^{-1}\sigma_t^{*2}(\theta) \leq \sigma_t^2(\theta) \leq K\, \sigma_t^{*2}(\theta) \ \ a.s.$$

Proof. A simple extension of Lee and Hansen (1994, Lemma 1). ■

Lemma 2 *Under Assumptions $A(2), B$, D and F,*

$$E\epsilon_t^{2\rho} < E\sigma_t^{2\rho} \leq \sup_{\theta \in \Theta} \sigma_t^{2\rho}(\theta) \leq KE\sigma_t^{*2\rho} < \infty. \tag{50}$$

$$E \sup_{\theta \in \Theta} \left| \ln \sigma_t^2(\theta) \right| \leq K < \infty. \tag{51}$$

Proof. We first prove (50), in which the first inequality follows from Jensen's inequality, the second is obvious and the third follows from Lemma 1. Writing $\overline{\psi}_j = \sup_{\zeta \in \Upsilon} \psi_j(\zeta)$, by recursive substitution

$$\sigma_t^{*2} \leq K + K \sum_{l=1}^{\infty} \left(\sum_{j_1=1}^{\infty} ... \sum_{j_l=1}^{\infty} \overline{\psi}_{j_1} \psi_{j_2} ... \psi_{j_l} z_{t-j_1}^2 z_{t-j_1-j_2}^2 ... z_{t-j_1 ... -j_l}^2 \right)$$

and so by Hardy, Littlewood and Polya (1964, Theorem 27),

$$\sigma_t^{*2\rho} \leq K + K \sum_{l=1}^{\infty} \left(\sum_{j_1=1}^{\infty} ... \sum_{j_l=1}^{\infty} \overline{\psi}_{j_1}^{\rho} \psi_{j_2}^{\rho} ... \psi_{j_l}^{\rho} z_{t-j_1}^{2\rho} z_{t-j_1-j_2}^2 ... z_{t-j_1 ... -j_l}^{2\rho} \right).$$

Thus

$$E\sigma_t^{*2\rho} \leq K + K \left(Ez_t^{2\rho} \sum_{j=1}^{\infty} \overline{\psi}_j^{\rho} \right) \sum_{l=0}^{\infty} \left(Ez_0^{2\rho} \sum_{j=1}^{\infty} \psi_j^{\rho} \right)^l < \infty,$$

in view of (23) and (33), to complete the proof of (50). To prove (51), we have $|\ln x| \leq x + x^{-1}$ for $x > 0$ and $\sigma_t^2(\theta) \geq \omega_L > 0$, so

$$E \sup_{\theta \in \Theta} \left| \ln \sigma_t^2(\theta) \right| \leq \rho^{-1} E \sup_{\theta \in \Theta} \sigma_t^{2\rho}(\theta) + \omega_L^{-1} \leq K < \infty.$$

■

Lemma 3 *Under Assumptions* $A(2)$ *and* D, $\{\sigma_t^2\}$, $\{\epsilon_t\}$ *are strictly stationary and ergodic. Under also Assumptions* F *and* $G(l)$,

$$\inf_{\theta \in \Theta} \sigma_t^2(\theta) > 0\,, \qquad \sup_{\theta \in \Theta} \sigma_t^2(\theta) < \sigma_t^{*2} < \infty \quad a.s., \tag{52}$$

and forall $\theta \in \Theta$,$\{\sigma_t^2(\theta)\}$, $\{q_t(\theta)\}$ *and their first* l *derivatives are strictly stationary and ergodic.*

Proof. The first part of (52) follows from $\omega_L > 0$ and the second from Lemma 2 and Lòeve (1977, p. 121). The strict stationarity and ergodicity follows by adapting Nelson (2000, proof of Theorem 2). ■

Lemma 4 *Under Assumption* $A(0)$, *for positive integer* $k < (b+1)n/2$,

$$E\left(\sum_{t=1}^{n} z_t^2\right)^{-k} < \infty. \tag{53}$$

Proof. Denote by $M_X(t) = E(e^{tX})$ the moment-generating function of a random variable X. By Cressie et al (1981) the left side of (53) is proportional to

$$\begin{aligned} \int_0^\infty t^{k-1} M_{\Sigma z_t^2}(-t)dt &= \int_0^\infty t^{k-1} M_{z_0^2}^n(-t)dt \\ &\leq \int_0^1 t^{k-1}dt + \int_1^\infty t^{k-1} M_{z_0^2}^n(-t)dt. \end{aligned} \tag{54}$$

It suffices to show that the last integral is bounded. For all $\delta > 0$, there exists $\varepsilon > 0$ such that $L(z^{-1}) \leq z^{-\delta}$, $z \in (0, \varepsilon)$, so

$$M_{z_0^2}(-t) = 2\int_0^\infty e^{-tz^2} f(z)dz \leq K \int_0^\varepsilon e^{-tz^2} z^{b-\delta} dz + 2e^{-t\varepsilon^2}.$$

The last integral is bounded by

$$K\,t^{(\delta-b-1)/2}\int_0^\infty e^{-z}z^{(\delta-b-1)/2-1}dz \le K\,t^{(\delta-b-1)/2}.$$

Thus (54) is finite if $k+n(\delta-b-1)/2<0$, that is, since δ is arbitrary, if $k<(b+1)n/2$. ■

Lemma 5 *Under Assumptions $A(q)$, B, C and D, for $p<q/2$,*

$$E\sup_{\theta\in\Theta}\left(\frac{\sigma_t^2}{\sigma_t^2(\theta)}\right)^p \le K<\infty.$$

Proof. By repeated substitution, for $j>0$,

$$\begin{aligned}\sigma_t^2 &= \omega_0+\sum_{k=1}^{\infty}\psi_{j+k}\epsilon_{t-j-k}^2+\sum_{r_1=1}^{j}\psi_{r_1}z_{t-r_1}^2\left(\omega_0+\sum_{k=1}^{\infty}\psi_{j+k-r_1}\epsilon_{t-j-k}^2\right)\\ &\quad+\sum_{r_1=}^{j-1}\sum_{r_2=1}^{j-r_1}\psi_{r_1}\psi_{r_2}z_{t-r_1}^2z_{t-r_1-r_2}^2\left(\omega_0+\sum_{k=1}^{\infty}\psi_{j+k-r_1-r_2}\epsilon_{t-j-k}^2\right)\\ &\quad+\ldots+\psi_1^j\prod_{r=1}^{j}z_{t-r}^2\sigma_{t-j}^2\\ &\le h_{tj}\,\sigma_{t-j}^2\ a.s.\end{aligned}$$

where $h_{tj}=\Pi_{i=1}^{j}(1+z_{t-i}^2)$ and we apply (46) and

$$1+\sum_{r_1=1}^{j}\psi_{r_1}z_{t-r_1}^2+\sum_{r_1=1}^{j-1}\sum_{r_2=1}^{j-r_1}\psi_{r_1}\psi_{r_2}z_{t-r_1}^2z_{t-r_1-r_2}^2+...+\psi_1^j\prod_{r=1}^{j}z_{t-r}^2\le h_{tj}.$$

Thus

$$\begin{aligned}\frac{\sigma_t^2}{\sigma_t^2(\theta)} &\le \frac{K\sigma_t^2}{\sigma_t^{2*}(\theta)}\le K\left(\frac{\omega}{\sigma_t^2}+\sum_{j=1}^{\infty}\psi_j(\zeta)\,z_{t-j}^2\,\frac{\sigma_{t-j}^2}{\sigma_t^2}\right)^{-1}\\ &\le K\left(\sum_{j=1}^{\infty}\psi_j(\zeta)\frac{z_{t-j}^2}{h_{t,j}}\right)^{-1}\le K\left(\sum_{j=1}^{n}\psi_j(\zeta)\frac{z_{t-j}^2}{h_{t,j}}\right)^{-1}a.s\,,\end{aligned}$$

for all n, so

$$\sup_{\theta\in\Theta}\frac{\sigma_t^2}{\sigma_t^2(\theta)}\le\sup_{\theta\in\Theta}K\frac{h_{tn}/\psi_n(\zeta)}{\sum_{j=1}^n z_{t-j}^2}\le\frac{K}{\epsilon_n}\frac{h_{tn}}{\sum_{j=1}^n z_{t-j}^2}\quad a.s.,$$

from (20). By Hölder's inequality,

$$E\left(\frac{h_{tn}}{\sum_{j=1}^n z_{t-j}^2}\right)^p\le K\left(Eh_{tn}^{q/2}\right)^{2p/q}E\left\{\left(\sum_{j=1}^n z_{t-j}^2\right)^{-pq/(q-2p)}\right\}^{1-2p/q}. \quad (55)$$

Now

$$Eh_{tn}^{q/2}=\prod_{i=1}^n E(1+z_{t-i}^2)^{q/2}\le 2^{n(q/2-1)}(1+Ez_0^q)$$

so the first expectation on the right of (55) is finite, while by Lemma 4 the second expectation is finite on choosing $n>2pq/[(b+1)(q-2p)]$. ∎

Lemma 6 *Under Assumptions* $A(2)$, B, C, D, E, F *and* $G(l)$, *for all* $p>0$*and* $k\le l$,

$$E\sup_{\theta\in\Theta}\left|\frac{1}{\sigma_t^2(\theta)}\frac{\partial^k\sigma_t^2(\theta)}{\partial\theta_{i_1}...\partial\theta_{i_k}}\right|^p<\infty, \quad (56)$$

$$E\sup_{\theta\in\Theta}\left|\frac{1}{\widehat{\sigma}_t^2(\theta)}\frac{\partial^k\widehat{\sigma}_t^2(\theta)}{\partial\theta_{i_1}...\partial\theta_{i_k}}\right|^p<\infty. \quad (57)$$

Proof. By compactness of Θ, (56) follows if for some $\varepsilon>0$,

$$E\left\{\sup_{\widetilde{\theta}:\|\widetilde{\theta}-\theta\|<\varepsilon}\left|\frac{1}{\sigma_t^2(\widetilde{\theta})}\frac{\partial^k\sigma_t^2(\widetilde{\theta})}{\partial\theta_{i_1}...\partial\theta_{i_k}}\right|^p\right\}<\infty.$$

Take $i_1\le i_2\le...\le i_k$. First assume $i_1\ge 3$, whence, for given k and $i_1,...i_k$

$$\frac{\partial^k\sigma_t^2(\theta)}{\partial\theta_{i_1}...\partial\theta_{i_k}}=\sum_{j=1}^\infty\xi_j(\zeta)\epsilon_{t-j}^2(\gamma)$$

where $\xi_j(\zeta)=\partial^k\psi_j(\zeta)/\partial\zeta_{i_1-2}...\partial\zeta_{i_k-2}$. Now

$$\left|\sum_{j=1}^\infty\xi_j(\zeta)\epsilon_{t-j}^2(\gamma)\right|\le 2\sum_{j=1}^\infty|\xi_j(\zeta)|\left(\epsilon_{t-j}^2+\gamma^2\right),$$

so using Lemma 1

$$\left|\frac{1}{\sigma_t^2(\theta)}\frac{\partial^s \sigma_t^2(\theta)}{\partial\theta_{i_1}...\partial\theta_{i_k}}\right| \leq \frac{2\sum_{j=1}^{\infty}|\xi_j(\zeta)|\,\epsilon_{t-j}^2}{\sigma_t^{*2}(\theta)} + K\sum_{j=1}^{\infty}|\xi_j(\zeta)|\,.$$

Denote by $\widetilde{\zeta}$ the column vector consisting of the last r elements of $\tilde{\theta}$. It suffices to take $p > 1$. By Hölder's inequality

$$\sum_{j=1}^{\infty}\left|\xi_j(\widetilde{\zeta})\right|\epsilon_{t-j}^2 \leq \left\{\sum_{j=1}^{\infty}\left|\xi_j(\widetilde{\zeta})\right|^{p/\rho}\psi_j(\widetilde{\zeta})^{1-p/\rho}\epsilon_{t-j}^2\right\}^{\rho/p}\left\{\sum_{j=1}^{\infty}\psi_j(\widetilde{\zeta})\epsilon_{t-j}^2\right\}^{1-\rho/p},$$

so

$$\left\{\frac{\sum_{j=1}^{\infty}\left|\xi_j(\widetilde{\zeta})\right|\epsilon_{t-j}^2}{\sigma_t^{*2}(\overline{\theta})}\right\}^p \leq K\sum_{j=1}^{\infty}\left|\xi_j(\widetilde{\zeta})\right|^p\psi_j(\widetilde{\zeta})^{\rho-p}\epsilon_{t-j}^{2\rho}.$$

By Assumptions E and $G(l)$, for all $\eta_1 > 0$, $\eta_2 > 0$ there exists ε such that

$$\begin{aligned}\sup_{\widetilde{\zeta}\in\mathcal{N}_\varepsilon(\zeta)}\left|\xi_j(\widetilde{\zeta})\right|^p\psi_j(\widetilde{\zeta})^{\rho-p} &\leq K\sup_{\widetilde{\zeta}\in\mathcal{N}_\varepsilon(\zeta)}\psi_j(\widetilde{\zeta})^{p(1-\eta_1)+\rho-p}\\ &\leq Kj^{-(\underline{d}+1)(\rho-p\eta_1)(1-\eta_2)}.\end{aligned}$$

We may choose η_1 and η_2 such that $(\underline{d}+1)(\rho - p\eta_1)(1-\eta_2) > 1$ so that

$$E\sup_{\widetilde{\theta}:\left\|\widetilde{\theta}-\theta\right\|<\varepsilon}\left\{\frac{\sum_{j=1}^{\infty}\left|\xi_j(\widetilde{\zeta})\right|\epsilon_{t-j}^2}{\sigma_t^{*2}(\widetilde{\theta})}\right\}^p < \infty.$$

The above proof implies that also

$$\sup_{\widetilde{\zeta}\in\mathcal{N}_\varepsilon(\zeta)}\left\{\sum_{j=1}^{\infty}\left|\xi_j(\widetilde{\zeta})\right|\right\}^p < \infty$$

whence the proof of (56) with $i_1 \geq 3$ is concluded. Next take $i_1 = 2$. If $i_2 > 2$

$$\frac{\partial^k\sigma_t^2(\theta)}{\partial\theta_{i_1}...\partial\theta_{i_k}} = -2\sum_{j=1}^{\infty}\xi_j(\zeta)\epsilon_{t-j}(\gamma) \tag{58}$$

where $\xi_j(\zeta) = \partial^{k-1}\psi_j(\zeta)/\partial\zeta_{i_2-2}...\partial\zeta_{i_k-2}$, while if $i_2 = 2$, $i_3 > 2$

$$\frac{\partial^k \sigma_t^2(\theta)}{\partial\theta_{i_1}...\partial\theta_{i_k}} = -2\sum_{j=1}^{\infty} \xi_j(\zeta)$$

where $\xi_j(\zeta) = \partial^{k-2}\psi_j(\zeta)/\partial\zeta_{i_3-2}...\partial\zeta_{i_k-2}$. In the first of these cases the proof is seen to be very similar to that above after noting that by the Cauchy inequality (58) is bounded by

$$K\left\{\sum_{j=1}^{\infty} |\xi_j(\zeta)|\,\epsilon_{t-j}^2 \sum_{j=1}^{\infty} |\xi_j(\zeta)|\right\}^{1/2} + K\sum_{j=1}^{\infty} |\xi_j(\zeta)|\,,$$

while in the second it is more immediate; we thus omit the details. We are left with the cases $i_1 = i_2 = i_3 = 2$ and $i_1 = 1$, both of which are trivial. The details for (57) are very similar (because the truncations in numerator and denominator match) and are thus omitted. ■

Define

$$Q(\theta) = Eq_0(\theta),\ G_T(\theta) = T^{-1}\sum_{t=1}^{T} g_t(\theta),\ H_T(\theta) = T^{-1}\sum_{t=1}^{T} h_t(\theta).$$

Lemma 7 *For some $\delta > 0$, under Assumptions $A(2+\delta)$, B, C, D, E, F and $G(1)$,*

$$\sup_{\theta\in\Theta} |Q_T(\theta) - Q(\theta)| \to 0 \ \ a.s. \quad as\ T \to \infty, \tag{59}$$

and $Q(\theta)$is continuous in θ.
If also Assumption $G(2)$ holds,

$$\sup_{\theta\in\Theta} \|G_T(\theta) - G(\theta)\| \to 0 \ \ a.s. \quad as\, T \to \infty, \tag{60}$$

If also Assumption $G(3)$ holds,

$$\sup_{\theta\in\Theta} \|H_T(\theta) - H(\theta)\| \to 0 \ \ a.s. \quad as\, T \to \infty, \tag{61}$$

and $H(\theta)$ is continuous in θ.

Proof. To prove (59), note first that by Lemmas 1, 2, 3 and 5

$$\sup_{\Theta} E\left|q_0(\theta)\right| \leq \sup_{\Theta} E\left|\log \sigma_0^2(\theta)\right| + \sup_{\Theta} E\chi_0(\theta) < \infty.$$

Thus by ergodicity

$$Q_T(\theta) \to Q(\theta), \qquad a.s.,$$

for all $\theta \in \Theta$. Then uniform convergence follows on establishing the equicontinuity property

$$\sup_{\tilde{\theta}:\left\|\tilde{\theta}-\theta\right\|<\varepsilon} \left|Q_T(\tilde{\theta}) - Q_T(\theta)\right| \to 0, \qquad a.s.,$$

as $\varepsilon \to 0$, and continuity of $Q(\theta)$. By the mean value theorem it suffices to show that

$$\sup_{\Theta} \left\|\frac{\partial Q_T(\theta)}{\partial \theta}\right\| + \sup_{\Theta} \left\|\frac{\partial Q(\theta)}{\partial \theta}\right\| < \infty, \quad a.s.,$$

which, by Lòeve (1977, p. 121) and identity of distribution, is implied by $E\sup_{\Theta} \left\|u_0(\theta)\right\| < \infty$. Using (43) and $\epsilon_t^2(\gamma) \leq K(\epsilon_t^2 + 1)$, $\left\|\nu_t(\theta)\right\| \leq 2(|\epsilon_t| + 1)$ we have

$$\left\|u_t(\theta)\right\| \leq K\left[\left\|\tau_t(\theta)\right\|\left\{1 + z_t^2 \frac{\sigma_t^2}{\sigma_t^2(\theta)}\right\} + |z_t| \frac{\sigma_t}{\sigma_t(\theta)} + 1\right].$$

Thus $E\sup_{\Theta} \left\|u_0(\theta)\right\|$ is bounded by a constant times

$$E\sup_{\Theta} \left\|\tau_0(\theta)\right\| + \left[E\sup_{\Theta}\left\{\frac{\sigma_0^2}{\sigma_0^2(\theta)}\right\}^p\right]^{1/p} \left[E\sup_{\Theta} \left\|\tau_0(\theta)\right\|^{p/(p-1)}\right]^{1-1/p}$$
$$+ E\sup_{\Theta}\left\{\frac{\sigma_0}{\sigma_0(\theta)}\right\} + 1$$

for all $p > 1$. On choosing $p < q/2$, this is finite, by Lemmas 5 and 6. This completes the proof of (59). Then (60) and (61) follow by applying analogous arguments to those above, and so we omit the detail, indeed (60) and (61) are only used in the proof of consistency of $\widehat{G}$, $\widehat{H}$ for G, H, where convergence over only a neighbourhood of θ_0 would suffice. ■

Lemma 8 *Under Assumptions $A(2+\delta)$, B, C, D, E, F and $G(1)$,*

$$\sup_{\theta\in\Theta}\left|Q_T(\theta)-\hat{Q}_T(\theta)\right|\to 0 \ \ a.s. \quad as\ T\to\infty. \tag{62}$$

If also Assumption $G(2)$ holds,

$$\sup_{\theta\in\Theta}\left\|G_T(\theta)-\widehat{G}_T(\theta)\right\|\to 0 \ \ a.s. \quad as\ T\to\infty. \tag{63}$$

If also Assumption $G(3)$ holds,

$$\sup_{\theta\in\Theta}\left\|H_T(\theta)-\widehat{H}_T(\theta)\right\|\to 0 \ \ a.s. \quad as\ T\to\infty. \tag{64}$$

Proof. We have $\hat{Q}_T(\theta)-Q_T(\theta)=A_T(\theta)+B_T(\theta)$, where

$$A_T(\theta)=T^{-1}\sum_{t=1}^{T}\ln\left[\frac{\hat{\sigma}_t^2(\theta)}{\sigma_t^2(\theta)}\right],\qquad B_T(\theta)=T^{-1}\sum_{t=1}^{T}\epsilon_t^2(\gamma)\left\{\hat{\sigma}_t^{-2}(\theta)-\sigma_t^{-2}(\theta)\right\}.$$

Because

$$\sigma_t^2(\theta)=\hat{\sigma}_t^2(\theta)+\sum_{j=0}^{\infty}\psi_{j+t}(\zeta)\epsilon_{-j}^2(\gamma),$$

$|\ln(1+x)|\le|x|$, and $\sigma_t^2(\theta)\ge\omega_L>0$, it follows that

$$\begin{aligned}|A_T(\theta)| &\le KT^{-1}\sum_{t=1}^{T}\left\{\sigma_t^2(\theta)-\hat{\sigma}_t^2(\theta)\right\}\le KT^{-1}\sum_{t=1}^{T}\sum_{j=t}^{T}\psi_j(\zeta)\epsilon_{t-j}^2(\gamma)\\ &\le KT^{-1}\sum_{t=0}^{\infty}\left\{\sum_{j=t+1}^{t+T}\psi_j(\zeta)\right\}\epsilon_{-t}^2(\gamma).\end{aligned}\tag{65}$$

Now from (19)

$$\begin{aligned}\sum_{j=t+1}^{t+T}\psi_j(\zeta) &\le K\psi_{t+1}(\zeta)(t+1)^{\eta+1}\sum_{j=t+1}^{t+T}j^{-\eta-1}\\ &\le K\min(t+1,T)\psi_{t+1}(\zeta),\end{aligned}$$

so (65) is bounded by

$$\frac{K}{T}\sum_{t=0}^{T-1}(t+1)\psi_{t+1}(\zeta)\epsilon_{-t}^2(\gamma)+K\sum_{t=T}^{\infty}\psi_t(\zeta)\epsilon_{-t}^2(\gamma).$$

Thus

$$\sup_{\Theta} A_T(\theta)\le\frac{K}{T}\sum_{t=0}^{T}(t+1)\overline{\psi}_{t+1}(\epsilon_{-t}^2+1)+K\sum_{t=T}^{\infty}\overline{\psi}_t(\epsilon_{-t}^2+1). \tag{66}$$

From (33) and Lemma 2, $\sum_{t=1}^{\infty}\overline{\psi}_t(\epsilon_{-t}^2+1)<\infty$ *a.s.* Thus the second term of (66) tends to zero *a.s.* as $T\to\infty$ while the first does so for the same reasons combined with the Kronecker lemma. Next

$$\begin{aligned}|B_T(\theta)| &\le KT^{-1}\sum_{t=1}^{T}\chi_t(\theta)\sum_{j=t}^{\infty}\psi_j(\zeta)\epsilon_{t-j}^2(\gamma)\\ &\le KT^{-1}\sum_{t=1}^{T}\chi_t(\theta)\sum_{j=t}^{\infty}\overline{\psi}_j(\epsilon_{t-j}^2+1).\end{aligned} \tag{67}$$

From (33) and Lemma 2, $\sum_{j=t}^{\infty}\overline{\psi}_j(\epsilon_{t-j}^2+1)\to 0$, *a.s.* Also, for each θ, *a.s.*

$$T^{-1}\sum_{t=1}^{T}\chi_t(\theta)\to E\chi_0(\theta)\le K\left\{E\left(\frac{\sigma_0^2}{\sigma_0^2(\theta)}\right)+1\right\}\le K<\infty$$

by ergodicity and Lemma 5. Thus (67) $\to 0$ *a.s.* by the Toeplitz lemma. The convergence is uniform in θ because, from the proof of Lemma 7, for all $\theta\in\Theta$

$$\sup_{\tilde{\theta}:\left\|\tilde{\theta}-\theta\right\|<\varepsilon}\left\|\chi_0(\tilde{\theta})-\chi_0(\theta)\right\|\to 0,\quad \text{a.s.},$$

as $\varepsilon\to 0$. This completes the proof of (62). We omit the proofs of (63) and (64) as they involve the same arguments. ■

Lemma 9 *For some $\delta>0$, under Assumptions $A(2+\delta)$, B, C, D, E , F, $G(1)$ and H, for all $\theta\in\Theta$, $M(\theta)$ is finite and positive definite.*

Proof. Fix $\theta\in\Theta$. Finiteness of $M(\theta)$ follows from Lemma 6. Positive definiteness follows if, for all non-null $(r+2)\times 1$ vectors λ, $\lambda' M(\theta)\lambda=$

$E\{\lambda'\tau_0(\theta)\}^2 > 0$, that is, that

$$\lambda'\tau_0(\theta)\sigma_0^2(\theta) \neq 0, \quad a.s. \tag{68}$$

since $0 < \sigma_0^2(\theta) < \infty$ $a.s.$ Define

$$\begin{aligned}
\tau_{t\omega}(\theta) &= \frac{\partial}{\partial\omega}\ln\sigma_t^2(\theta) = \sigma_t^{-2}(\theta), \\
\tau_{t\gamma}(\theta) &= \frac{\partial}{\partial\gamma}\ln\sigma_t^2(\theta) = -2\sigma_t^{-2}(\theta)\sum_{j=1}^{\infty}\psi_j(\zeta)\epsilon_{t-j}(\gamma), \\
\tau_{t\zeta}(\theta) &= \frac{\partial}{\partial\zeta}\ln\sigma_t^2(\theta) = \sigma_t^{-2}(\theta)\sum_{j=1}^{\infty}\psi_j^{(1)}(\zeta)\epsilon_{t-j}^2(\gamma),
\end{aligned}$$

so that $\tau_t(\theta) = (\tau_{t\omega}(\theta),\, \tau_{t\gamma}(\theta),\, \tau'_{t\zeta}(\theta))'$. Write $\lambda = (\lambda_1, \lambda_2, \lambda_3')'$, where λ_1 and λ_2 are scalar and λ_3 is $r \times 1$. Consider first the case $\lambda_1 = \lambda_2 = 0$, $\lambda_3 \neq 0$. Suppose (68) does not hold. Then we must have

$$\sum_{j=1}^{\infty}\lambda_3'\psi_j^{(1)}(\zeta)\epsilon_{t-j}^2(\gamma) = 0, \quad a.s.$$

If $\lambda_3'\psi_1^{(1)}(\zeta) \neq 0$ it follows that

$$(\sigma_{t-1}z_{t-1} + \gamma_0 - \gamma)^2 = -\left\{\lambda_3'\psi_j^{(1)}(\zeta)\right\}^{-1}\sum_{j=2}^{\infty}\lambda_3'\psi_j^{(1)}(\zeta)\epsilon_{-j}^2(\gamma). \tag{69}$$

However since $\sigma_{t-1} > 0$ $a.s.$ the left side involves the non-degenerate random variable z_{t-1}, which is independent of the right side, so (69) cannot hold. Thus $\lambda_3'\psi_j^{(1)}(\zeta) = 0$. Repeated application of this argument indicates that, for all ζ, $\lambda_3'\psi_j^{(1)}(\zeta) = 0$, $j = 1, ..., J$. This is contradicted by Assumption H, so that (68) cannot hold. Next consider the case $\lambda_1 = 0$, $\lambda_2 \neq 0$, $\lambda_3 = 0$. If (68) does not hold we must have

$$\sum_{j=1}^{\infty}\psi_j(\zeta)\epsilon_{t-j}(\gamma) = 0, \quad a.s. \tag{70}$$

Let k be the smallest integer such that $\psi_k(\zeta) \neq 0$. Then (70) implies

$$z_{t-k} = \sigma_{t-k}^{-1}(\theta) \left\{ \gamma - \gamma_0 - \psi_k^{-1}(\zeta) \sum_{j=k+1}^{\infty} \psi_j(\theta) \epsilon_{t-j}(\gamma) \right\}.$$

But the left side is nondegenerate and independent of the right side, so (70) cannot hold. Next consider the case $\lambda_1 = 0$, $\lambda_2 \neq 0$, $\lambda_3 \neq 0$. If (68) is not true then, taking $\lambda_2 = 1$, we must have

$$\sum_{j=1}^{\infty} \left\{ \lambda_3' \psi_j^{(1)}(\zeta) \epsilon_{t-j}(\gamma) - 2\psi_j(\zeta) \right\} \epsilon_{t-j}(\zeta) = 0, \quad a.s. \tag{71}$$

Let k be the smallest integer such that either $\lambda_3' \psi_k^{(1)}(\zeta) \neq 0$ or $\psi_k(\zeta) \neq 0$; the preceding argument indicates that there exists such k. Then we have

$$\left\{ 2\psi_k(\zeta) - \lambda_3' \psi_k^{(1)}(\zeta)(\sigma_{t-k} z_{t-k} + \gamma - \gamma_0) \right\} \{ \sigma_{t-k} z_{t-k} + \gamma - \gamma_o \}$$

$$= \sum_{j=k+1}^{\infty} \left\{ \lambda_3' \psi_j^{(1)}(\zeta) \epsilon_{t-j}(\gamma) - 2\psi_j(\zeta) \right\} \epsilon_{t-j}(\zeta), \quad a.s.$$

The left side is $a.s.$ non-zero and involves the non-degenerate random variable z_{t-k} that is independent of the right side, so (71) cannot hold. We are left with the cases where $\lambda_1 = 0$, $\sigma_t^2(\theta)\tau_{t\omega}(\theta) \equiv 1$, and the preceding arguments indicate that there exist no λ_2 and λ_3 such that

$$\lambda_2 \tau_{t\gamma}(\theta) + \lambda_3' \tau_{t\zeta}(\theta) = 1, \quad a.s.$$

■

Lemma 10 *For some $\delta > 0$, under Assumptions $A(2+\delta)$, B, C, D, E, F, $G(1)$ and H,*

$$\inf_{\substack{\theta \in \Theta \\ \theta \neq \theta_0}} Q(\theta) > Q(\theta_0).$$

Proof. We have

$$Q(\theta) - Q(\theta_0) = E\left[\frac{\sigma_0^2}{\sigma^2(\theta)} - \ln\left\{ \frac{\sigma_0^2}{\sigma^2(\theta)} \right\} - 1 \right] + (\gamma - \gamma_0)^2 E\left[\frac{1}{\sigma_0^2(\theta)} \right].$$

The second term on the right hand side is zero only for $\gamma = \gamma_0$ and is positive otherwise. Because $x - \ln x - 1 \geq 0$, with equality only when $x = 1$, it remains to show that

$$\ln \sigma_0^2(\theta) = \ln \sigma_0^2, \quad a.s., \text{ some } \theta \neq \theta_0. \tag{72}$$

By the mean value theorem, (72) implies that $(\theta - \theta_0)' \tau_0(\bar{\theta}) = 0$, a.s., for $\theta \neq \theta_0$ and some $\bar{\theta}$ such that $\left\|\bar{\theta} - \theta_0\right\| \leq \|\theta - \theta_0\|$. But by Lemma 9 there is no such $\bar{\theta}$. ■

References

Baillie, R.T., Bollerslev, T. and Mikkelsen, H.O. (1996). Fractionally integrated generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* 74, 3-30.

Berkes, L., Horvath, L. and Kokoszka P. (2001). GARCH processes: structure and estimation. *Bernoulli*, forthcoming.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroscedasticity. *Journal of Econometrics* 31, 302-327.

Bougerol, P. and Picard, N. (1992a). Strict stationarity of generalized autoregressive processes. *Annals of Probability 20, 1714-30.*

Bougerol, P. and Picard, N. (1992b) Stationarity of GARCH processes and of some nonegative time series. *Journal of Econometrics* 52, 115-127.

Breidt, F.J., Crato, N. and de Lima P. (1998). The detection and estimation of long memory in stochastic volatility. *Journal of Econometrics* 73, 325-34.

Brown, B. (1971). Martingale central limit theorems. *The Annals of Mathematical Statistics* 35, 1296–1303.

Comte, F. and Lieberman, O. (2000). Asymptotic theory for multivariate GARCH processes. *Journal of Multivariate Analysis*, forthcoming.

Cressie, N., Davis, A., Folks, J. and Policello II, G. (1981). The moment-generating function and the negative integer moments. *The American Statistician* 35, 148–150.

Davis, R.A., Mikosch, T. and Basrak, B. (1999). Sample ACF of mutivariate stochastic recurrence equations with applications to GARCH. Preprint.

Ding, Z. and Granger, C.W.J. (1996). Modelling volatility persistence of speculative returns: a new approach. *Journal of Econometrics* 73, 185-215.
Ding, Z., Granger, C.W.J. and Engle, R.F. (1993). A long memory property of stock market returns and a new model. *Journal of Empirical Finance* 1, 83-106.
Engle, R.F. (1982). Autoregressive conditional heteroscedasticity with Estimates of United Kingdom Inflation. *Econometrica 50, 987-1007.*
Engle, R.F. and Bollerslev, T. (1986). Modelling the persistence of conditional variances. *Econometric Reviews* 5, 1-50.
Giraitis, L., Kokoszka, P. and Leipus, R. (2000). Stationary ARCH models:dependence structure and central limit theorem. *Econometric Theory* 16, 3-22.
Giraitis, L. and Robinson, P.M. (2001). Whittle estimation of ARCH models. *Econometric Theory* 17, 608-631.
Hall, P. and Yao, Q. (2003). Inference in ARCH and GARCH models with heavy-tailed errors. *Econometrica* 71, 285-317.
Hardy, G., Littlewood, J. and Polya, G. (1964). *Inequalities*, Cambridge University Press, Cambridge (UK).
Harvey, A.C. (1998). Long memory in stochastic volatility. In: *Forecasting volatility in financial markets* (J. Knight and S. Satchell, eds.), Butterworth-Heinemann, London.
Jeantheau, T. (1998). Strong consistency of estimators for multivariate ARCH models. *Econometric Theory* 14, 70-86.
Lee, S., and Hansen, B. (1994). Asymptotic theory for the GARCH$(1,1)$ quasi-maximum likelihood estimator. *Econometric Theory* 10, 29–52.
Ling, S. and McAleer, M. (2001). Asymptotic theory for a vector ARCH-GARCH model. *Econometric Theory* , forthcoming.
Lòeve, M. (1977). *Probability theory 1* (4th Edition), Springer-Verlag, New York.
Mikosch, T. and Starica, C. (2000). Limit theory for the sample autocorrelations and extremes of a GARCH$(1,1)$ process. *Annals of Statistics 28, 1427-1451.*
Mikosch, T. and Straumann, D. (2002). Whittle estimation in a heavy-tailed GARCH(1,1) model. *Stochastic Processes and Their Applications 100, 187-222.*
Nelson, D. (1990). Stationarity and persistence in GARCH(1,1) models. *Econometric Theory 6, 318-334.*

Nelson, D. (1991). Conditional heteroskedasticity in asset returns: a new approach. *Econometrica 59, 347-370.*

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

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

Taylor, S.J. (1986). *Modelling financial time series.* Chichester, UK.

Von Bahr, B. and Esseen, C.G. (1965). Inequalities for the rth absolute moment of asum of random variables $1 \leq r \leq 2$. *Annals of Mathematical Statistics 36, 299-303.*

Weiss, A. (1986). Asymptotic theory for ARCH models. *Econometric Theory 2, 107-131.*

Whistler, D.E.N. (1990). *Semiparametric models of daily and intradaily exchange rate volatility.* Ph.D. thesis, University of London.

Zaffaroni, P (2002). Gaussian inference on certain long-range dependent volatility models. *Journal of Econometrics*, forthcoming.

Zaffaroni, P (2003). Stationarity and memory of ARCH(∞) models. *Econometric Theory*, forthcoming.

Department of Economics
London School of Economics
Houghton Street
London WC2A 2AE
UK
E-mail: p.m.robinson@lse.ac.uk

Servizio Studi
Banca d'Italia
Via Nazionale, 91
00194 Roma
ITALY
E-mail: paolo.zaffaroni@econ.cam.ac.uk