2016/30

On Asymptotic Theory for ARCH (∞) Models

Christian M. Hafner and Arie Preminger

50 YEARS OF CORE

DISCUSSION PAPERS

CORE
Voie du Roman Pays 34, L1.03.01
Tel (32 10) 47 43 04
Fax (32 10) 47 43 01
Email: immaq-library@uclouvain.be
http://www.uclouvain.be/en-44508.html

On asymptotic theory for ARCH(∞) models

Christian M. Hafner*[1] and Arie Preminger[2]

[1] *Université catholique de Louvain*
[2] *Ben Gurion University*

August 12, 2016

Abstract

ARCH(∞) models nest a wide range of ARCH and GARCH models including models with long memory in volatility. The existing literature on such models is quite restrictive in terms of existence of moments. However, the popular FIGARCH, one version of a long memory in volatility model, does not have finite second moments and rarely satisfies the moment conditions of the existing literature. This paper considerably weakens the moment assumptions of a general ARCH(∞) class of models, and develops the theory for consistency and asymptotic normality of the quasi maximum likelihood estimator.

Key words: volatility, long memory, fractional integration, quasi maximum likelihood

JEL Classification Number: C12, C13, C14.

*Corresponding author, Institute of statistics, biostatistics and actuarial sciences, and CORE, Université catholique de Louvain, Voie du Roman Pays 20, 1348 Louvain-la-Neuve, Belgium, email: `christian.hafner@uclouvain.be`.

We would like to thank Paolo Zaffaroni for helpful comments and suggestions.

1 Introduction

ARCH(∞) type models generalize the class of ARCH(p) models, introduced by Engle (1982), and GARCH(p, q) by Bollerslev (1986), to allow for a flexible form of the decay of the coefficients in the volatility equation. In particular, it is possible within the ARCH(∞) model class to include a long memory effect in the squared series of observations, which is often observed in financial returns. Examples of parsimonious models having an ARCH(∞) representation with hyperbolic decay of the coefficients are the FIGARCH model of Baillie et al (1996) and alternative versions of fractional GARCH models introduced by Robinson (1991) and Ding and Granger (1996). Davidson (2004) discusses the non-trivial moment and memory properties of the FIGARCH model and proposes an extension, the hyperbolic GARCH (HYGARCH) model with a more natural transition from a short memory GARCH to integrated GARCH (IGARCH). For a review of long memory ARCH type models, see e.g. Tayefi and Ramanathan (2012) and Giraitis et al (2009).

Berkes et al (2003) show the existence and uniqueness of the ARCH(∞) representation of a GARCH(p, q) model, and prove consistency and asymptotic normality of the quasi maximum likelihood estimator under mild conditions. These processes have short memory with exponentially decaying coefficients. Douc et al (2008) provide sufficient conditions for the existence of a stationary causal solution of an ARCH(∞) process, which allows coefficients to decay at a power-law and hence includes the FIGARCH and fractional GARCH models. In the FIGARCH case, this solution necessarily implies an infinite variance of the process, as shown by Kazakevičius and Leipus (2002) and Kazakevičius and Leipus (2003).

Strong consistency and asymptotic normality of quasi-maximum likelihood estimators (QMLE) in a general class of models including ARCH(∞) have been derived by Bardet and Wintenberger (2009). However, they assume finite second moments which excludes e.g. the FIGARCH model. Robinson and Zaffaroni (2006) relax this restriction, albeit at the cost of having other, stronger conditions than Bardet and Wintenberger (2009). The moment condition of Robinson and Zaffaroni (2006) is $\mathbb{E}|x_0|^{2\rho} < \infty$, where x_0 is the stationary solution of the model and $\rho \in (4/(2d_0+3), 1)$, where $d_0 > 1/2$ is the fractional decay parameter. Note that for d_0 close to $1/2$, ρ will be close to one, so that essentially a finite moment close to second order is required. This appears restrictive, as not all possible FIGARCH models are included, depending on the fractional parameter d_0.

In this paper, we show consistency and asymptotic normality of QMLE in a general ARCH(∞) class of models with only a mild fractional moment condition, and otherwise similar assumptions as in Robinson and Zaffaroni (2006). This extends the applicability of the results to a much broader class of models, and in particular includes all FIGARCH-type models, as we will show explicitly in a separate section.

Throughout the paper we use the following notation. The Euclidean norm of a matrix or a vector X is denoted by $|X| = \sqrt{\mathrm{tr}(X'X)}$, and $||X||_r = (\mathbb{E}|X|^r)^{1/r}$ denotes the L^r norm of a random matrix or vector X.

The remainder of the paper is organized as follows. The following section presents the model, the assumption, and the main results. Section 3 shows that the FIGARCH model is a special case of our model framework and satisfies our assumptions. The last section concludes. Proofs are delegated to the appendix.

2 The model and main results

The ARCH(∞) process comprises a wide class of models for conditional heteroscedasticity in time series. Consider the model

$$y_t = h_{0t}^{1/2} \varepsilon_t \tag{1}$$

$$h_{0t} = \psi_{00} + \sum_{j=1}^{\infty} \psi_{0j} y_{t-j}^2 \tag{2}$$

where $\{\varepsilon_t\}$ is a sequence of independent and identically distributed (i.i.d.) random variables with $\mathbb{E}(\varepsilon_t) = 0$, $\mathbb{E}(\varepsilon_t^2) = 1$, and

$$\psi_{00} > 0, \quad \psi_{0j} \geq 0, \quad \sum_{j=1}^{\infty} \psi_{0j} < \infty.$$

We shall assume that a strictly stationary solution to the model exists almost surely (a.s.). We consider a parametric version, in which we know functions $\psi_j(\theta_0)$ of a $p \times 1$ vector θ_0 such that, for $j \geq 0$, $\psi_{0j} = \psi_j(\theta_0)$. Let $\{y_t\}_{t=1}^T$ be a realization of the stationary solution of (22). The process h_{0t} is approximated by the process

$$\tilde{h}_t(\theta_0) = \psi_0(\theta_0) + \sum_{j=1}^{t-1} \psi_j(\theta_0) y_{t-j}^2 \; .$$

The Gaussian likelihood is often used in practice for estimation. Thus, the log likelihood, multiplied by -2 and ignoring additive constants, takes the form

$$\tilde{L}_T(\theta) = \frac{1}{T}\sum_{t=1}^{T} \tilde{l}_t, \quad \tilde{l}_t = \tilde{l}_t(\theta) = \frac{y_t^2}{\tilde{h}_t(\theta)} + \log \tilde{h}_t(\theta)$$

and the quasi maximum likelihood estimator (QMLE) is defined as $\hat{\theta}_T = \arg\min_{\theta\in\Theta} \tilde{L}_T(\theta)$.

The likelihood using the true, unobserved volatility process (23) is given by

$$L_T(\theta) = \frac{1}{T}\sum_{t=1}^{T} l_t, \quad l_t = \frac{y_t^2}{h_t(\theta)} + \log h_t(\theta),$$

where

$$h_t(\theta) = \psi_0(\theta) + \sum_{j=1}^{\infty} \psi_j(\theta) y_{t-j}^2 \ .$$

We make the following assumptions.

(A1) The ε_t are i.i.d. random variables with $\mathbb{E}(\varepsilon_t) = 0$, and $\mathbb{E}(\varepsilon_t^2) = 1$, and $\lim_{\lambda\to 0} \lambda^{-\alpha} P(\varepsilon_t^2 \leq \lambda) < \infty$ for some $\alpha > 0$.

(A2) $\theta_0 \in \Theta$, where $\Theta \subset \mathbb{R}^p$ is a compact set.

(A3) For all $j \geq 0$, $\inf_{\theta\in\Theta} \psi_j(\theta) > 0$ and $\sup_{\theta\in\Theta} \psi_j(\theta) \leq K j^{-d-1}$ for some $K > 0$ and $d \in (0, 1)$.

(A4) There exists a stationary and ergodic solution to (22)-(23) such that $\mathbb{E}(|y_t^2|^\delta) < \infty$ for some $\delta > 0$.

Assumption (A1) allows an asymmetric distribution of ε_t. The second part of (A1) excludes the case where the distribution of ε_t has some mass concentrated around zero, see Shinki and Zhang (2012) and Robinson and Zaffaroni (2006) for similar assumptions. A sufficient condition is that the density function of ε_t is bounded from above. Assumption (A2) is a standard assumption to show consistency. Assumption (A3) is a memory condition implying that $\psi_{0j} = o(j^{-1})$ as $j \to \infty$. Assumption (A4) is needed to apply the ergodic theorem and the martingale difference central limit theorem for stationary processes to establish the consistency and asymptotic normality of QMLE. Sufficient conditions for (A4) have been given by Douc et al (2008).

Theorem 1 *Under Assumptions (A1)-(A4), $\hat{\theta}_T \to \theta_0$ almost surely.*

To obtain asymptotic normality, we use the following additional assumptions.

(A5) $\mathbb{E}\left[\beta + \psi_{01}\varepsilon_t^2\right]^{-1} \leq 1$, where $\beta = \inf_{j\geq 0} \beta_j$ and $\beta_j = \psi_{0,j+1}(\theta)/\psi_{0j}(\theta)$.

(A6) $\left|\frac{\partial \psi_j(\theta)}{\partial \theta_{i_1}}\right| \leq K\psi_j(\theta)^{1-\eta}$, $\left|\frac{\partial^2 \psi_j(\theta)}{\partial \theta_{i_1}\partial \theta_{i_2}}\right| \leq K\psi_j(\theta)^{1-\eta}$, $i_1, i_2 = 1, \ldots p$ for all $\eta > 0$.

(A7) There does not exist $\lambda \in \mathbb{R}^p$, $\lambda \neq 0$, such that $\lambda' \frac{\partial h_t}{\partial \theta} = 0$ a.s.

(A8) θ_0 is an interior point of Θ.

(A9) $\mathbb{E}(\varepsilon_t^4) = \kappa < \infty$.

For a given model, Assumption (A5) can easily be verified by numerical calculation, see the next section for an example. Assumption (A6) assumes continuous twice differentiability of the coefficient functions, which is weaker than what is needed e.g. in Theorem 2 of Robinson and Zaffaroni (2006). Similar to Assumption G of Robinson and Zaffaroni (2006), (A7) ensures that the asymptotic covariance matrix of the estimator is positive definite. Again, we show in the next section that this condition can be checked for a particular model. (A8) and (A9) are standard assumptions.

The next theorem establishes the asymptotic normality of the quasi-maximum likelihood estimators. Define

$$V = \mathbb{E}\left(\frac{\partial^2 l_t(\theta_0)}{\partial \theta \partial \theta'}\right) = \mathbb{E}\left(\frac{1}{h_t^2(\theta_0)} \frac{\partial h_t(\theta_0)}{\partial \theta} \frac{\partial h_t(\theta_0)}{\partial \theta'}\right)$$

Theorem 2 *Under Assumptions (A1)-(A9),*

$$\sqrt{T}\left(\hat{\theta}_T - \theta_0\right) \to_d N(0, (\kappa - 1)V^{-1}).$$

The asymptotic covariance matrix can be estimated using the following estimators,

$$\hat{V}_T = \frac{1}{T}\sum_{t=1}^{T} \frac{\partial^2 \tilde{l}_t(\hat{\theta}_T)}{\partial \theta \partial \theta'}, \quad \hat{\kappa}_T = \frac{1}{T}\sum_{t=1}^{T} \hat{\varepsilon}_t^4, \quad \hat{\varepsilon}_t = h_t^{-1/2}(\hat{\theta}_T) y_t.$$

Using standard arguments it is straightforward to show that $\hat{V}_T^{-1} \to V^{-1}$ and $\hat{\kappa}_T \to \kappa$ in probability. Note that QMLE is consistent, but not efficient in the case of non-Gaussian innovations. However, one could use it as a first step estimator, which in a second step could be improved e.g. via a nonparametric estimation of the efficient score function, see e.g. Drost and Klaassen (1997) for classical GARCH models.

3 Example: FIGARCH

The ARCH(∞) specification of Section 2 is sufficiently general to nest many types of long memory models, see e.g. Davidson (2004) for an overview. In the following we discuss a popular example. Consider the FIGARCH model of Baillie et al (1996) given by

$$\begin{aligned} y_t &= h_t^{1/2}\varepsilon_t, \quad \varepsilon_t \sim i.i.d.\ N(0,1) \\ h_t &= \omega + \psi(L)y_t^2 \end{aligned} \tag{3}$$
$$\psi(L) = 1 - \frac{1-a(L)}{b(L)}(1-L)^d$$

where $0 < d < 1$, L is the lag operator, $a(L) = \sum_{j=1}^{p} a_j L^j$, $b(L) = 1 - \sum_{j=1}^{q} b_j L^j$. The lag polynomials $1 - a(L)$ and $b(L)$ are coprime, i.e. do not have common roots. Furthermore, $a_j > 0,\ b_j > 0$ and $b(z) \neq 0$ for $z \leq 1$. For the fractional differencing term, there is the well-known expansion

$$(1-L)^d = \sum_{j=0}^{\infty} \frac{\Gamma(j-d)}{\Gamma(-d)\Gamma(j+1)} L^j$$

We have the following low level assumptions. Note first that the normality assumption of ε_t implies (A1) and (A9). There is a $d_0 > 1/2$ such that (A6) holds, which is needed for the asymptotic normality of the QMLE. It is directly satisfied for GARCH models. Furthermore, θ_0 is an interior point of the compact set Θ such that (A8) holds. Theorem 3 of Conrad and Haag (2006) gives sufficient conditions under which the first part of (A3) holds for the FIGARCH(p, d, q) model. Stirling's approximation indicates that the coefficients of the expansion of $(1-L)^d$ are bounded by Kj^{-d-1}, so that the same holds for $\psi_j(\theta)$, and the second part of (A3) follows. Assumption (A4) can be shown using the results of Douc et al (2008). To check (A5), we set for simplicity $a(L) = 0$ and $b(L) = 1$, and simulated the process 10,000 times for values $d \in (0,1)$ and Gaussian innovations ε_t. We found that (A5) holds in these cases.[1] For Assumption (A6), using similar arguments as Robinson and Zaffaroni (2006) for their fractional GARCH model, note that $\partial\psi(z)/\partial d = -(1-a(z))/b(z)(1-z)^d \ln(1-z)$. The coefficient of z^j in the expansion of $-(1-z)^d \ln(1-z)$ is bounded by $K \ln(j) j^{-d-1} \leq Kj^{-(d+1)(1-\eta)} \leq K\psi_j^{1-\eta}$ for

[1] A GAUSS program to calculate the condition in (A5) is available from the authors upon request.

any $\eta > 0$. Derivatives with respect to a_i and b_j are dominated, and the second derivatives can be treated in a similar way. This shows (A6).

We now want to check (A7) for the general FIGARCH model given in (3). First, define

$$\phi(L) = (1-L)^d b(L)^{-1},\ \gamma(L) = b(L)^{-1}(1-a(L)).$$

Let us take the derivative with respect to the model parameters,

$$\begin{aligned}
\frac{\partial h_t}{\partial \omega} &= 1 \\
\frac{\partial h_t}{\partial a_j} &= L^j \phi(L) y_t^2 \\
\frac{\partial h_t}{\partial b_j} &= L^j \phi(L)\gamma(L) y_t^2 \\
\frac{\partial h_t}{\partial d} &= -\gamma(L)(1-L)^d \log(1-L) y_t^2
\end{aligned}$$

We need to show that, for any $\lambda \in \mathbb{R}^{p+q+2}$, $\lambda' \frac{\partial h_t}{\partial \theta} = 0 \Rightarrow \lambda = 0$. Let $\lambda = (\lambda^\omega, \lambda^{a\prime}, \lambda^{b\prime}, \lambda^d)'$, where $\lambda^a = (\lambda_1^a, \ldots, \lambda_p^a)'$ corresponds to the derivative with respect to a, and the remaining elements of λ are defined accordingly. We observe that

$$\begin{aligned}
\lambda' \frac{\partial h_t}{\partial \theta} &= \lambda^\omega + \sum_{j=1}^p \lambda_j^a L^j \phi(L) y_t^2 + \sum_{j=1}^q \lambda_j^b L^j \phi(L)\gamma(L) y_t^2 - \lambda^d \gamma(L)(1-L)^d \log(1-L) y_t^2 \\
&= \lambda^\omega + \pi(L) y_t^2 = 0
\end{aligned}$$

where $\pi(L) = \sum_{i=0}^\infty \pi_i L^i$. It is clear that $\pi_0 = 0$, otherwise y_t^2 would be measurable with respect to ε_{t-i}^2 for $i \geq 1$. Repeated application of this argument will show that $\pi_i = 0$ for all i. Hence, $\pi(L) = 0$ which implies that $\lambda^\omega = 0$. It remains to show that $\pi(L) = 0$ implies that $\lambda = 0$. We have

$$\begin{aligned}
0 &= \sum_{j=1}^p \lambda_j^a L^j \phi(L) + \sum_{j=1}^q \lambda_j^b L^j \phi(L)\gamma(L) - \lambda^d \gamma(L)(1-L)^d \log(1-L) \\
&= \sum_{j=1}^p \lambda_j^a L^j \phi(L) + \sum_{j=1}^q \lambda_j^b L^j \phi(L)\gamma(L) \\
&\Rightarrow \sum_{j=1}^p \lambda_j^a L^j b(L) + \sum_{j=1}^q \lambda_j^b L^j (1-a(L)) = 0
\end{aligned}$$

The last equality follows since otherwise $\log(1-L)$ is a fraction of finite degree lag polynomials. The result follows since $1-a(L)$ and $b(L)$ are coprime. If $1-a(z)$ has

p roots $[a_1,\ldots,a_p]$ we have that $\sum_{j=1}^{p}\lambda_j^a a_i^j b(a_i)=0,\;\; i=1,\ldots,p$, where $b(a_i)\neq 0$ and

$$\begin{bmatrix} a_1^j & 0 & \cdots & 0 \\ 0 & a_2^j & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & a_p^j \end{bmatrix}\begin{bmatrix} \lambda_1^a \\ \lambda_2^a \\ \vdots \\ \lambda_p^a \end{bmatrix}=0 \Rightarrow \begin{bmatrix} \lambda_1^a \\ \lambda_2^a \\ \vdots \\ \lambda_p^a \end{bmatrix}=0 \tag{4}$$

The matrix on the left hand side of (4) is invertible since $a_i^j \neq 0$. Similarly, plugging the roots of $b(L)$ into the equation above we get that $\lambda_j^b = 0\;\; j=1,\ldots,q$.

To conclude, under suitable low-level conditions that are easy to check, (A1)-(A9) hold for the FIGARCH model, which is therefore a special case of our framework.

4 Conclusions

We have derived the asymptotic theory of the quasi maximum likelihood estimator of a general class of ARCH(∞) models under weak assumptions. All assumptions are low level in the sense that they can easily be checked for a given model. We show that our framework includes the popular FIGARCH model as a special case.

A potential route for future research could be to investigate semiparametric estimators of the model, based on first step QMLE, in the spirit of Drost and Klaassen (1997). Such estimators use nonparametric estimates of the efficient score function to construct a more efficient estimator of the model parameters in the case of non-Gaussian innovations. Extending the results of Drost and Klaassen (1997) for short memory GARCH to potentially long memory ARCH(∞) would be a theoretical and practical challenge. Our paper contributes by delivering the theory for the first step estimator under weak assumptions, which could help in developing the theory of a semiparametric estimator.

Appendix

Let K be a generic constant whose values will be modified along the proofs. To simplify notation we set $h_t = h_t(\theta)$, $h_{0t} = h_t(\theta_0)$, and

$$\dot{h}_{it} = \frac{\partial h_t}{\partial \theta_i},\ \ddot{h}_{ijt} = \frac{\partial^2 h_t}{\partial \theta_i \partial \theta_j},\ \dot{\tilde{h}}_{it} = \frac{\partial \tilde{h}_t}{\partial \theta_i},\ \ddot{\tilde{h}}_{ijt} = \frac{\partial^2 \tilde{h}_t}{\partial \theta_i \partial \theta_j}$$

Let $\nabla l_t(\theta) = \frac{\partial l_t(\theta)}{\partial \theta}$, $\nabla l_{it}(\theta) = \frac{\partial l_t(\theta)}{\partial \theta_i}$ and $\nabla^2 l_t(\theta) = \frac{\partial^2 l_t(\theta)}{\partial \theta \partial \theta'}$, $\nabla^2 l_{ijt}(\theta) = \frac{\partial^2 l_t(\theta)}{\partial \theta_i \partial \theta_j}$ denote the first and second derivatives of $\ell_t(\theta)$ (and their elements), respectively. The same notation is used for $\tilde{\ell}_t(\theta)$ and its derivatives.

Lemma 1 *Under Assumption (A1), the model (22)-(23) is uniquely identifiable.*

Proof: Let $h_t(\theta) = \psi_0(\theta) + \psi_\theta(L) y_t^2$ and suppose that $h_t(\theta) = h_t(\theta_0)$ a.s. for all t. Then $h_t(\theta) - h_t(\theta_0) = 0 \Rightarrow \psi_0(\theta_0) - \psi_0(\theta) = [\psi_\theta(L) - \psi_{\theta_0}(L)]\, y_t^2$ a.s. for all t. If $\psi_\theta(L) \neq \psi_{\theta_0}(L)$, the right hand side consists of some linear combination of ε_{t-j}^2, $j > 0$, which has a nondegenerate distribution and, hence, is not constant. Therefore, $\psi_\theta(L) = \psi_{\theta_0}(L)$ and $\psi_0(\theta) = \psi_0(\theta_0)$. $\square$

Lemma 2 *Under Assumptions (A3),(A4), (A6), for some $a \in (1-d, 1)$ and $r \in (0, 2\delta)$,*

i) $T^{-a} \sum_{t=1}^T \sup_{\theta \in \Theta} ||h_t - \tilde{h}_t||_r \ = o(1)$

ii) $T^{-a} \sum_{t=1}^T \sup_{\theta \in \Theta} ||\dot{h}_{it} - \dot{\tilde{h}}_{it}||_r \ = o(1)$ *for all i*

iii) $T^{-a} \sum_{t=1}^T \sup_{\theta \in \Theta} ||\ddot{h}_{ijt} - \ddot{\tilde{h}}_{ijt}||_r \ = o(1)$ *for all i, j*

Proof:
i) Note that

$$h_t - \tilde{h}_t = \sum_{j=0}^{\infty} \psi_{j+t} y_{-j}^2 \tag{5}$$

Hence,

$$\begin{aligned} T^{-a} \sum_{t=1}^{T} \sup_{\theta \in \Theta} ||h_t - \tilde{h}_t||_r &\leq T^{-a} \sum_{t=1}^{T} \sum_{j=t}^{\infty} \sup_{\theta \in \Theta} \psi_j(\theta) ||y_{t-j}||_r \\ &\leq KT^{-a} \sum_{t=0}^{\infty} \sum_{j=t+1}^{t+T} \sup_{\theta \in \Theta} \psi_j(\theta) \\ &\leq KT^{-a} \sum_{t=0}^{\infty} \sum_{j=t+1}^{t+T} j^{-d-1} \end{aligned}$$

$$
\begin{aligned}
&\leq \quad KT^{-a}\sum_{t=0}^{\infty}\min(t+1,T)(t+1)^{-(d+1)} \\
&\leq \quad KT^{-a}\sum_{t=0}^{T-1}(t+1)^{-d}+KT^{1-a}\sum_{t=T}^{\infty}(t+1)^{-(1+d)} \\
&\leq \quad KT^{1-(a+d)}\int_0^1 r^{-d}dr+K\sum_{t=T}^{\infty}(t+1)^{-(a+d)} \\
&\leq \quad KT^{1-(a+d)}. \qquad (6)
\end{aligned}
$$

The first two inequalities result from (5), Assumption (A4) and the Minkowski inequality, respectively. Assumption (A3) implies the third inequality. Similar to Robinson and Zaffaroni (2006, p.1070), we use the fact that $\sum_{j=t+1}^{t+T} j^{-d-1} \leq \sum_{t=0}^{\infty}\min(t+1,T)(t+1)^{-d-1}$ to obtain the next two inequalities. The last term tends to zero as $T \to \infty$.

ii-iii) Using similar arguments as above and setting $\eta < (d+a-1)/(1+d)$, the desired results follow. □

Lemma 3 *Under Assumptions (A1) and (A5),*

i) $\mathbb{E}\left(\frac{h_{0t-k}}{h_{0t}}\right)^r \leq 1$*, for all* $r \geq 1$

ii) $\mathbb{E}\left(\sum_{i=1}^{n}\varepsilon_{t+i}^2\right)^{-r} \leq K$ *for all* $\alpha n > r$*, where* α *is given in (A1).*

iii) $\mathbb{E}\left(\sup_{\theta\in\Theta}\frac{h_{0t-j}}{h_t(\theta)}\right) \leq K$ *for all* $j \geq 0$

Proof:

i) By (23), we get

$$
\begin{aligned}
\frac{h_{0t-1}}{h_{0t}} &= \frac{h_{0t-1}}{\psi_{00}+\sum_{j=1}^{\infty}\psi_{0j}y_{t-j}^2} = \frac{h_{0t-1}}{\psi_{00}+\psi_{01}h_{t-1}\varepsilon_{t-1}^2+\beta h_{t-1}+\sum_{j=2}^{\infty}\psi_{0j}y_{t-j}^2-\beta h_{t-1}} \\
&= \frac{h_{0t-1}}{\psi_{00}(1-\beta)+\psi_{01}h_{0t-1}\varepsilon_{t-1}^2+\beta h_{0t-1}+\sum_{j=1}^{\infty}\psi_{0j+1}y_{t-j-1}^2-\beta\sum_{j=1}^{\infty}\psi_{0j}y_{t-j-1}^2} \\
&= \frac{h_{0t-1}}{\psi_{00}(1-\beta)+\psi_{01}h_{t-1}\varepsilon_{t-1}^2+\beta h_{0t-1}+\sum_{j=1}^{\infty}y_{t-j-1}^2(\psi_{0j+1}-\beta\psi_{0j})} \\
&\leq \frac{h_{0t-1}}{\psi_{01}h_{0t-1}\varepsilon_{t-1}^2+\beta h_{0t-1}} = \frac{1}{\psi_{01}\varepsilon_{t-1}^2+\beta} \qquad (7)
\end{aligned}
$$

and

$$
\frac{h_{0t-k}}{h_{0t}} = \frac{h_{0t-1}}{h_{0t}}\frac{h_{0t-2}}{h_{0t-1}}\cdots\frac{h_{0t-k}}{h_{0t-k-1}} \leq \prod_{i=1}^{k}\frac{1}{\psi_{01}\varepsilon_{t-i}^2+\beta} \qquad (8)
$$

Hence, by Assumptions (A1) and (A5),

$$\mathbb{E}\left(\frac{h_{0t-k}}{h_{0t}}\right)^r \leq \prod_{i=1}^{k} \mathbb{E}(\psi_{01}\varepsilon_{t-i}^2 + \beta)^{-r} \leq 1 \tag{9}$$

ii) Let $Y_t = \varepsilon_t^2$ and $X_n = \sum_{j=1}^n Y_{t+i}$. By Assumption (A1) there exists some $K > 0$ and $\alpha > 0$, for some sufficiently small $\lambda > 0$, such that

$$\Pr(X_n \leq \lambda) \leq \Pr(\cap_{i=1}^n \{Y_{t+i} \leq \lambda\}) = \Pr(Y_{t+i} \leq \lambda)^n \leq (K\lambda^\alpha)^n \tag{10}$$

Further, by the definition of expectation and integration by parts

$$\begin{aligned}
\mathbb{E}(X_n)^{-r} &= r\int_0^\infty x^{-r-1}P(X_n < x)dx \\
&= r\int_\lambda^\infty x^{-r-1}P(X_n < x)dx + r\int_0^\lambda x^{-r-1}P(X_n < x)dx \\
&\leq K + r\int_0^\lambda x^{-r-1}(Kx^\alpha)^n dx \\
&\leq K + K\int_0^\lambda x^{-r+\alpha n-1}dx
\end{aligned} \tag{11}$$

Note that the first term after the second equality is finite for all $\lambda > 0$. The first inequality follows using (10) for sufficiently small $\lambda > 0$ and the integral is finite if $\alpha n > r$.

iii) Let $\bar{\psi} = \min_{i=1,\ldots,M}\{\inf_{\theta\in\Theta}\psi_i(\theta)\}$. By Assumption (A3), $\bar{\psi} > 0$. For $j > M$,

$$\frac{h_{0t-j}}{h_t} = \frac{h_{0t-j}}{\psi_0 + \sum_{i=1}^\infty \psi_i y_{t-i}^2} \leq \frac{h_{0t-j}}{\sum_{i=1}^M \psi_i h_{0t-i}\varepsilon_{t-i}^2} \leq \frac{h_{0t-j}}{\bar{\psi}\sum_{i=1}^M h_{0t-i}\varepsilon_{t-i}^2} \tag{12}$$

Next, define $m = \arg\min_{l=1,\ldots M}\{h_{0t-l}\}$. It is clear that this variable is a well-defined random variable. Also, let $I(\cdot)$ be the indicator function which equals one if $m = l$ and zero otherwise. Now, by (12),

$$\frac{h_{0t-j}}{h_t} \leq \sum_{l=1}^M \frac{h_{0t-j}}{\bar{\psi}\sum_{i=1}^M h_{0t-i}\varepsilon_{t-i}^2} I(m = l) \leq K\sum_{l=1}^M \frac{h_{0t-j}}{h_{0t-l}\sum_{i=1}^M \varepsilon_{t-i}^2} \tag{13}$$

From the c_r inequality for $k \geq 1$,

$$\mathbb{E}\left(\sup_{\theta\in\Theta}\frac{h_{0t-j}}{h_t}\right)^k \leq K\sum_{l=1}^M \mathbb{E}\left(\frac{1}{\sum_{i=1}^M \varepsilon_{t-i}^2}\frac{h_{0t-j}}{h_{0t-l}}\right)^k \tag{14}$$

Let $k' = k(1+\delta)$ where $\delta > 0$, applying the Hölder inequality for each of the M terms yields,

$$\mathbb{E}\left(\sup_{\theta\in\Theta}\frac{h_{0t-j}}{h_t}\right)^k \leq K\sum_{l=1}^{M}\left(\mathbb{E}\left(\sum_{i=1}^{M}\varepsilon_{t-i}^2\right)^{-k'/\delta}\right)^{1-k/k'}\left(\mathbb{E}\left(\frac{h_{0t-j}}{h_{0t-l}}\right)^{k'}\right)^{k/k'} \tag{15}$$

Since $j > l$, by (9),

$$\mathbb{E}\left(\frac{h_{0t-j}}{h_{0t-l}}\right)^{k'} \leq \prod_{l=1}^{j-l}\mathbb{E}[\psi_{01}\varepsilon_{t-l}^2+\beta]^{-k'} \leq (\mathbb{E}[\psi_{01}\varepsilon_{t-l}^2+\beta]^{-k'})^{j-l} < \infty \tag{16}$$

If $M > k'/\delta\alpha$ the first term on the RHS of (15) is bounded as well by Lemma 3(ii). For $j \leq M$, we generalize Lemma 5 of Robinson and Zaffaroni (2006). We have

$$\begin{aligned} h_{0t} &= \psi_{00}+\psi_{01}h_{0t-1}\varepsilon_{t-1}^2+\sum_{i=2}^{\infty}\psi_{0i}h_{0t-i}\varepsilon_{t-i}^2 \\ &\leq \psi_{00}+\psi_{01}h_{0t-1}\varepsilon_{t-1}^2+Kh_{0t-1} = \psi_{00}+h_{0t-1}(\psi_{01}\varepsilon_{t-1}^2+K) \end{aligned} \tag{17}$$

Thus, $h_{0t-j}/h_{0t-j-1} \leq K(\varepsilon_{t-j-1}^2+1)$, and for $i \geq 1$, $h_{0t-j}/h_{0t-j-i} \leq K\prod_{\ell=1}^{i}(\varepsilon_{t-j-\ell}^2+1)$. It follows that

$$\begin{aligned} \frac{h_{0t-j}}{h_t} &\leq \left\{\inf_{\theta\in\Theta}\left(\frac{\sum_{i=1}^{M}\psi_{i+j}(\theta)h_{0t-i-j}\varepsilon_{t-i-j}^2}{h_{0t-j}}\right)\right\}^{-1} \\ &\leq K\left\{\inf_{\theta\in\Theta}\sum_{i=1}^{M}\psi_{i+j}(\theta)\frac{\varepsilon_{t-i-j}^2}{\prod_{\ell=1}^{i}(\varepsilon_{t-j-\ell}^2+1)}\right\}^{-1} \\ &\leq K\frac{\prod_{\ell=1}^{M}(\varepsilon_{t-j-\ell}^2+1)}{\sum_{i=1}^{M}\varepsilon_{t-i-j}^2}\left\{\inf_{\theta\in\Theta,i=1,\ldots M}\psi_{i+j}(\theta)\right\}^{-1} \end{aligned} \tag{18}$$

The proof can be completed by applying the Hölder inequality and Lemma 3(ii) as above. □

Lemma 4 *Under Assumptions (A1), (A5) and (A6), for all* $r > 0$

i) $\mathbb{E}\sup_{\theta\in\Theta}\left(\dot{h}_{it}/h_t\right)^r < \infty$ *for all* i

ii) $\mathbb{E}\sup_{\theta\in\Theta}\left(\ddot{h}_{ijt}/h_t\right)^r < \infty$ *for all* i, j

Proof:
i) We have

$$\dot{h}_{it} = \tilde{\psi}_0 + \sum_{j=1}^{\infty} \tilde{\psi}_{ij}(\theta) y_{t-j}^2$$

where $\tilde{\psi}_{ij}(\theta) = \partial\psi_j(\theta)/\partial\theta_i$. Let $r' = r(1+\epsilon)$ where $\epsilon > 0$. By the Hölder inequality,

$$\begin{aligned}
&\left(h_t^{-1}\sum_{j=1}^{\infty} \tilde{\psi}_{ij}(\theta) y_{t-j}^2\right)^r \\
&\leq \left(h_t^{-1}\sum_{j=1}^{\infty}\left[|\tilde{\psi}_{ij}(\theta)|\psi_j(\theta)^{-(1-1/r')}y_{t-j}^{2/r'}\right]\cdot\left[\psi_j(\theta)^{(1-1/r')}y_{t-j}^{2(1-1/r')}\right]\right)^r \\
&\leq \left\{\left(h_t^{-1}\sum_{j=1}^{\infty}|\tilde{\psi}_{ij}(\theta)|^{r'}\psi_j(\theta)^{1-r'}y_{t-j}^2\right)^{1/r'}\left(h_t^{-1}\sum_{j=1}^{\infty}\psi_j(\theta)y_{t-j}^2\right)^{1-1/r'}\right\}^r \\
&\leq \left(h_t^{-1}\sum_{j=1}^{\infty}|\tilde{\psi}_{ij}(\theta)|^{r'}\psi_j(\theta)^{1-r'}y_{t-j}^2\right)^{r/r'} \qquad (19)
\end{aligned}$$

Since $r/r' < 1$, by the c_r inequality,

$$\left(\frac{\sum_{j=1}^{\infty}\tilde{\psi}_{ij}(\theta)y_{t-j}^2}{h_t}\right)^r \leq K\sum_{j=1}^{\infty}|\tilde{\psi}_{ij}(\theta)|^r\psi_j(\theta)^{(1-r')r/r'}\left(\frac{y_{t-j}^2}{h_t}\right)^{r/r'} \qquad (20)$$

By Assumption (A6),

$$\sup_{\theta\in\Theta}|\tilde{\psi}_{ij}(\theta)|^r\psi_j(\theta)^{(1-r')r/r'} \leq K\sup_{\theta\in\Theta}\psi_j(\theta)^{(1-r')r/r'+r(1-\eta)} \leq Kj^{-(1+d)[r/r'-r\eta]} = Kj^{-(1+a)}$$

If $\epsilon > 0$ and $\eta > 0$ are small enough, the term $r/r' - \eta r$ is arbitrarily close to 1, which implies that $a > 0$. Hence,

$$\begin{aligned}
\mathbb{E}\sup_{\theta\in\Theta}\left(\frac{\sum_{j=1}^{\infty}\tilde{\psi}_{ij}(\theta)y_{t-j}^2}{h_t}\right)^r &\leq K\sum_{j=1}^{\infty}j^{-1-a}\mathbb{E}\sup_{\theta\in\Theta}\left(\frac{y_{t-j}^2}{h_t}\right)^{r/r'} \qquad (21)\\
&\leq K\sum_{j=1}^{\infty}j^{-1-a}\mathbb{E}\sup_{\theta\in\Theta}\left(\frac{h_{0t-j}^2}{h_t}\right)^{r/r'}\mathbb{E}\left(\varepsilon_{t-j}^2\right)^{r/r'} < \infty
\end{aligned}$$

By the c_r inequality, Assumption (A1) and since $r/r' < 1$, we get $\mathbb{E}\left(\varepsilon_{t-j}^2\right)^{r/r'} \leq K\mathbb{E}\left(\varepsilon_{t-j}^2\right) = K$, and the result follows by Lemma 3(iii).

ii) The proof is similar to that of Lemma 4(i), hence it is omitted. □

Lemma 5 *Under Assumptions (A3) and (A4),*

i) $T^{-1}\sum_{t=1}^{T}\sup_{\theta\in\Theta}|l_t - \tilde{l}_t| \;=\; o_{a.s.}(1)$ *for* $d > 0$.

If additionally Assumption (A6) holds, then

ii) $T^{-1/2}\sum_{t=1}^{T}\sup_{\theta\in\Theta}|\nabla l_{it}(\theta) - \nabla\tilde{l}_{it}(\theta)| \;=\; o_P(1)$ *for* $d > \frac{1}{2}$ *and for all* i.

iii) $T^{-1}\sum_{t=1}^{T}\sup_{\theta\in\Theta}|\nabla^2 l_{ijt}(\theta) - \nabla^2\tilde{l}_{ijt}(\theta)| \;=\; o_P(1)$ *for* $d > 0$ *and for all* i, j.

Proof:

i) First note that

$$L_T - \tilde{L}_T = \frac{1}{T}\sum_{t=1}^{T}(l_t - \tilde{l}_t) = \frac{1}{T}\sum_{t=1}^{T}\ln\left(\frac{h_t}{\tilde{h}_t}\right) - \frac{1}{T}\sum_{t=1}^{T}\left(\frac{y_t^2}{h_t} - \frac{y_t^2}{\tilde{h}_t}\right) \tag{22}$$

Assumption (A2) implies that $h_t, \tilde{h}_t \geq \psi_L > 0$, and an application of the mean value theorem leads to $|h_t^{-1} - \tilde{h}_t^{-1}| \;\leq\; \psi_L^{-2}|h_t - h_t|$ and $|\log h_t - \log\tilde{h}| \;\leq\; \psi_L^{-1}|h_t - \tilde{h}_t|$. Because of

$$h_t(\theta) = \tilde{h}_t(\theta) + \sum_{j=0}^{\infty}\psi_{j+t}(\theta)y_{-j}^2,$$

it follows from (A3) and the Minkowski inequality that for some $\delta > 0$

$$\begin{aligned}
\sup_{\theta\in\Theta}|L_T - \tilde{L}_T| \;&\leq\; \frac{1}{T}\sum_{t=1}^{T}\sup_{\theta\in\Theta}|l_t - \tilde{l}_t| \\
&\leq\; \frac{K}{T}\sum_{t=1}^{T}(1 + y_t^2)\sup_{\theta\in\Theta}(h_t - \tilde{h}_t).
\end{aligned} \tag{23}$$

From Assumptions (A3), (A4) and the Minkowski inequality we have that

$$\begin{aligned}
\left\|\sup_{\theta\in\Theta}\sum_{j=1}^{\infty}\psi_j(\theta)(1 + y_{-j}^2)\right\|_\delta \;&\leq\; \sum_{j=1}^{\infty}\sup_{\theta\in\Theta}\psi_j(\theta)\left(1 + ||y_{-j}^2||_\delta\right) \\
&\leq\; K\sum_{j=1}^{\infty}j^{-1-d} < \infty
\end{aligned}$$

which, by Loeve (1977, p.121) is a.s. finite. Thus, by the Cesaro mean convergence theorem (or the Toeplitz lemma), the right hand side of the second inequality in (23) tends to zero a.s. as $T \to \infty$.

ii) By direct calculation,

$$\frac{\partial l_t}{\partial \theta} = \left\{1 - \frac{y_t^2}{h_t}\right\}\left\{\frac{1}{h_t}\frac{\partial h_t}{\partial \theta}\right\} \tag{24}$$

Hence,

$$\begin{aligned} T^{-1/2}\sum_{t=1}^{T}\left(\frac{\partial l_t}{\partial \theta_i} - \frac{\partial \tilde{l}_t}{\partial \theta_i}\right) &= T^{-1/2}\sum_{t=1}^{T}\left\{\frac{y_t^2}{\tilde{h}_t} - \frac{y_t^2}{h_t}\right\}\frac{1}{h_t\dot{h}_{ti}} \\ &+ \left\{1 - \frac{y_t^2}{\tilde{h}_t}\right\}\left\{\frac{1}{h_t - \tilde{h}_t}\right\}\dot{h}_{ti} \\ &+ \left\{1 - \frac{y_t^2}{\tilde{h}_t}\right\}\frac{1}{\tilde{h}_t}\left\{\dot{h}_{it} - \dot{\tilde{h}}_{it}\right\} \end{aligned} \tag{25}$$

By repeated applications of the Hölder inequality, the Minkowski inequality and Assumption (A4), we obtain

$$\begin{aligned} & T^{-1/2}\sum_{t=1}^{T}\sup_{\theta\in\Theta}\left\|\nabla l_{it} - \nabla \tilde{l}_{it}\right\|_{r/3} \\ \leq\; & T^{-1/2}K\sum_{t=1}^{T}\left\{\sup_{\theta\in\Theta}||h_t - \tilde{h}_t||_r\left(2\left\|\frac{1}{h_t}\frac{\partial h_t}{\partial \theta_i}\right\|_r\right) + K\left\|\frac{\partial h_t}{\partial \theta_i} - \frac{\partial \tilde{h}_t}{\partial \theta_i}\right\|_r\right\} \end{aligned} \tag{26}$$

The last equality follows from Lemma 2(ii) by setting $a = 0.5$ and $d > 0.5$. The desired result follows from the Markov inequality.

iii) The proof uses similar arguments, hence it is omitted. □

Lemma 6 *Under Assumptions (A1)-(A4),*

i) $\liminf_{T\to\infty}\inf_{\theta\in\mathcal{N}}\frac{1}{T}\sum_{t=1}^{T} l_t(\theta) \;\geq\; \mathbb{E}\inf_{\theta\in\mathcal{N}} l_t(\theta)$, *for any compact set* $\mathcal{N} \subseteq \Theta$.

ii) $\lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T} l_t(\theta_0) = \mathbb{E}l_t(\theta_0)$

iii) $\mathbb{E}(l_t(\theta_0)) \;\leq\; \mathbb{E}(l_t(\theta))$ *with equality if and only if* $\theta = \theta_0$.

Proof:

i) For any compact set $\mathcal{N} \subseteq \Theta$ we have,

$$\liminf_{T\to\infty}\inf_{\theta\in\mathcal{N}} T^{-1}\sum_{t=1}^{T} l_t(\theta) \;\geq\; \liminf_{n\to\infty} T^{-1}\sum_{t=1}^{T}\inf_{\theta\in\mathcal{N}} l_t(\theta) \tag{27}$$

Further, note that $\mathbb{E}l_t(\theta) < \infty$ is well-defined and belongs to $\mathbb{R}\cup\{+\infty\}$ because $\mathbb{E}\log^- h_t(\theta) \;\leq\; \max(0, -\ln\psi_0(\theta)) \;<\; \infty$. Hence, by Assumption (A4), we can apply the ergodic theorem (see Billingsley (1995) p.284) to the stationary and ergodic sequence $\{\inf_{\theta\in\mathcal{N}} l_t(\theta)\}_t$ to obtain

$$\liminf_{T\to\infty} \inf_{\theta\in\mathcal{N}} T^{-1}\sum_{t=1}^{T} l_t(\theta) \;\geq\; \mathbb{E}\left(\inf_{\theta\in\mathcal{N}} l_t(\theta)\right) \tag{28}$$

ii) Assumption (A4) and Jensen's inequality imply that $\mathbb{E}\log h_{0t} \leq r^{-1}\log \mathbb{E}h_{0t}^{r} < \infty$ for some $r > 0$. Hence, $\mathbb{E}(l_t(\theta_0)) < \infty$ by Assumption (A4), and the desired results follow from the ergodic theorem.

iii) Set $x = h_{0t}/h_t$ and note that

$$\mathbb{E}(l_t(\theta)) - \mathbb{E}(l_t(\theta_0)) \;=\; \mathbb{E}\left[\frac{h_{0t}}{h_t} - \log\left(\frac{h_{0t}}{h_t}\right) - 1\right] = \mathbb{E}\left[x - \log(x) - 1\right]$$

Since $\log(x) \leq x - 1$ for $x > 0$, with equality if and only if $x = 1$, the results follows directly from Lemma 1. $\square$

Lemma 7 *Under Assumptions (A1)-(A9),*

i) $T^{-1/2}\sum_{t=1}^{T} \nabla l_t(\theta_0) \to_D N(0, \kappa J)$ *where* J *is a positive definite matrix, and* $\kappa = \mathbb{E}(\varepsilon_t^4) - 1$.

ii) If $\tilde{\theta}_T \to_p \theta_0$, $T^{-1}\sum_{t=1}^{T} \nabla^2 l_t(\tilde{\theta}_n) \to_p -J$

Proof:

i) From (24),

$$\mathbb{E}\left\{\nabla l_t(\theta_0)|\mathcal{F}_{t-1}\right\} = h_{0t}^{-1}\dot{h}_{0t}\mathbb{E}\left\{1 - \varepsilon_t^2\right\} = 0$$

where $\mathcal{F}_t = \sigma(y_t, y_{t-1}, \ldots)$. We also have

$$var\left\{\nabla l_t(\theta_0)\right\} = \mathbb{E}\left\{\nabla l_t(\theta_0)\nabla l_t'(\theta_0)\right\} = \mathbb{E}\left(1 - \varepsilon_t^2\right)^2 \mathbb{E}\left\{h_{0t}^{-2}\dot{h}_{0t}\dot{h}_{0t}'\right\} = \kappa J.$$

By applying the Cauchy-Schwarz inequality and Lemma 4(i) we can show that the variance is finite. These results imply that $\{\nabla l_t(\theta_0), \mathcal{F}_t\}$ is a stationary and ergodic martingale difference sequence with finite variance which is positive definite by using similar arguments used in Lemma 9 of RZ(2006). Thus, Theorem 23.1 of Billingsley (1968) and the Cramér-Wold device imply that $T^{-1/2}\sum_{t=1}^{T} \nabla^2 l_t(\theta_0) \to_D N(0, \kappa J)$.

ii) The second derivative of the log likelihood function is given by

$$
\begin{aligned}
\nabla^2 l_{ijt} &= \left\{1-\frac{y_t^2}{h_t}\right\}\frac{\ddot{h}_{ijt}}{h_t}+\left\{2\frac{y_t^2}{h_t}-1\right\}\frac{\dot{h}_{it}\dot{h}_{jt}}{h_t^2} \\
&= \left\{1-\frac{h_{0t}\varepsilon_t^2}{h_t}\right\}\frac{\ddot{h}_{ijt}}{h_t}+\left\{2\frac{h_{0t}\varepsilon_t^2}{h_t}-1\right\}\frac{\dot{h}_{it}\dot{h}_{jt}}{h_t^2} \qquad (29)
\end{aligned}
$$

We note that $\mathbb{E}\left(\nabla^2\ell_t(\theta_0)\right)=-J$ and we can show by applying the Cauchy-Schwarz and Minkowski inequalities that

$$
\mathbb{E}\left(\sup_{\theta\in\Theta}\left|\nabla^2 l_{ijt}\right|\right)\leq\left\|1+\frac{h_{0t}}{h_t}\varepsilon_t^2\right\|_2\left\|\frac{\ddot{h}_{ijt}}{h_t}\right\|_2+\left\|2\frac{h_{0t}}{h_t}\varepsilon_t^2+1\right\|_2\left\|\frac{\dot{h}_{it}}{h_t}\right\|_4\left\|\frac{\dot{h}_{jt}}{h_t}\right\|<\infty
$$

The last inequality follows from Lemma 4(ii). From the ergodic theorem (see e.g., Billingsley (1995)),

$$
\sup_{\theta\in\Theta}\left|\frac{1}{T}\sum_{t=1}^{T}\nabla^2 l_t(\theta)-\mathbb{E}\left(\nabla^2 l_t(\theta)\right)\right| \to_{a.s.} 0
$$

Hence, given $\epsilon>0$

$$
\left|\frac{1}{T}\sum_{t=1}^{T}\nabla^2 l_t(\tilde{\theta}_T)-\mathbb{E}\left(\nabla^2 l_t(\tilde{\theta}_T)\right)\right| < \frac{1}{2}\epsilon
$$

in probability for T sufficiently large. Since $\mathbb{E}\left(\nabla^2\ell_t(\theta)\right)$ is continuous,

$$
\left|\mathbb{E}\left(\nabla^2 l_{ijt}(\tilde{\theta}_T)\right)-\mathbb{E}\left(\nabla^2 l_{ijt}(\theta_0)\right)\right| < \frac{1}{2}\epsilon
$$

since $\tilde{\theta}_T\to\theta_0$ in probability, and the desired result follows from an application of the triangle inequality since ϵ is arbitrary. $\square$

Proof of Theorem 1:

We use similar arguments as in Theorem 5.3.1 of Straumann (2005, p.101) showing strong consistency by contradiction. Suppose that almost sure convergence of $\hat{\theta}_T$ to θ_0 does not hold, so that for some arbitrary $\gamma>0$, the compact set $\mathcal{F}=\{\omega\in\Omega|\limsup_{T\to\infty}||\hat{\theta}_T-\theta_0||\ \geq\gamma,\ \hat{\theta}_T\in\Theta\}$ has a positive probability. Since the set $\mathcal{N}=\Theta\cap\{\theta:|\hat{\theta}_T-\theta_0|\ \geq\ \gamma\}$ is compact, there exists a non-null subset $\bar{\mathcal{F}}\subset\mathcal{F}$ such that for every $\omega\in\bar{\mathcal{F}}$, one can find in $\mathcal{N}$ a convergent subsequence $\hat{\theta}_{T_i}(\omega)\to\theta\in\mathcal{N}$. By definition of the QMLE,

$$
\begin{aligned}
\liminf_{i\to\infty}\frac{1}{T_i}\sum_{t=1}^{T_i}\tilde{l}_t(\theta_0) &\geq \liminf_{i\to\infty}\inf_{\theta\in\mathcal{N}}\frac{1}{T_i}\sum_{t=1}^{T_i}\tilde{l}_t(\theta) \\
&= \liminf_{i\to\infty}\frac{1}{T_i}\sum_{t=1}^{T_i}\tilde{l}_t(\hat{\theta}_{T_i}) \qquad (30)
\end{aligned}
$$

By Lemma 5(i), with probability one we have

$$\liminf_{i\to\infty} \frac{1}{T_i}\sum_{t=1}^{T_i} l_t(\theta_0) \;\geq\; \liminf_{i\to\infty} \frac{1}{T_i}\sum_{t=1}^{T_i} l_t(\hat{\theta}_{T_i}) \tag{31}$$

The inequality above and Lemmas 6(i)-(ii) imply that with positive probability $\mathbb{E}l_t(\theta_0) \;\geq\; \mathbb{E}\inf_{\theta\in\mathcal{N}} l_t(\theta)$. This result contradicts Lemma 6(iii) which states that in the limit $L_T(\theta)$ is uniquely minimized at θ_0. Since $\gamma > 0$ is arbitrary, consistency follows. □

Proof of Theorem 2: By using a mean-value expansion of $\tilde{Q}_T(\hat{\theta}_T) = \sum_{t=1}^{T} \tilde{l}_t(\hat{\theta}_T)$ around θ_0, we have

$$\begin{aligned}
0 &= T^{-1/2}\sum_{t=1}^{T} \nabla\tilde{l}_t(\hat{\theta}_T) \\
&= T^{-1/2}\sum_{t=1}^{T} \nabla\tilde{l}_t(\theta_0) + \left(\frac{1}{T}\sum_{t=1}^{T}\nabla^2\tilde{l}_t(\bar{\theta}_T)\right)\sqrt{T}(\hat{\theta}_T-\theta_0) \\
&= T^{-1/2}\sum_{t=1}^{T} \nabla\tilde{l}_t(\theta_0) + \left[\left(\frac{1}{T}\sum_{t=1}^{T}\nabla^2\tilde{l}_t(\bar{\theta}_T) - \frac{1}{T}\sum_{t=1}^{T}\nabla^2 l_t(\bar{\theta}_T)\right)\right. \\
&+ \left.\left(\frac{1}{T}\sum_{t=1}^{T}\nabla^2 l_t(\bar{\theta}_T) + J\right) - J\right]\sqrt{T}(\hat{\theta}_T-\theta_0)
\end{aligned} \tag{32}$$

where $\bar{\theta}_T$ lies on the chord between $\hat{\theta}_T$ and θ_0.

Lemmas 5(ii), 7(i), and the asymptotic equivalence lemma (see e.g. White (1994), p.172) imply that $\frac{1}{\sqrt{T}}\sum_{t=1}^{T} \partial\tilde{\ell}_t(\theta_0)\Big/\partial\theta \to_D N(0,\kappa J)$ where J is a positive definite matrix. Further, note that $\bar{\theta}_T \to \theta_0$ almost surely by Theorem 1. Hence, Lemmas 5(iii) and 7(ii) imply that the first and second terms, inside the square brackets in (32), converge in probability to zero. To complete the proof it suffices to solve (32) with respect to $\sqrt{T}(\hat{\theta}_T-\theta_0)$ and apply Slutsky's theorem. □

References

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

Bardet, J. M. and Wintenberger, O. (2009). Asymptotic normality of the quasi-maximum likelihood estimator for multidimensional causal processes, *The Annals of Statistics* 37, 2730-2759.

Berkes, I., Horvath, L. and Kokoszka, P. (2003). GARCH processes: structure and estimation, *Bernoulli*, 9, 201-227.

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

Billingsley (1995). *Probability and Measure.* John Wiley, New York.

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

Chung, K. L. (1974). A Course in Probability Theory. Orlando: Academic Press, Inc.

Conrad, C. and Haag, B. (2006). Inequality constraints in the fractionally integrated GARCH model, *Journal of Financial Econometrics* 4, 413-449.

Davidson, J. (2004). Moment and memory properties of linear conditional heteroskedasticity models, and a new model, *Journal of Business and Economic Statistics*, 22, 16-29.

Ding, Z. and Granger, C. W. J. (1996). Modeling volatility persistence of speculative returns: A new approach, *Journal of Econometrics*, 73, 185-215.

Douc, R., Roueff, F. and Soulier, P. (2008). On the existence of some ARCH(∞) processes, *Stochastic Processes and their Applications*, 118, 755-761.

Drost, F. C. and Klaassen, C. A. J. (1997). Efficient estimation in semiparametric GARCH models, *Journal of Econometrics* 81, 193-221.

Engle, R. F. (1982), Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation, *Econometrica*, 50, 987-1007.

Giraitis, L., Leipus, R. and Surgailis, D. (2009). ARCH(∞) models and long memory properties, *Handbook of Financial Time Series*, ed. by T.G. Andersen, R.A. Davis, J.-P. Kreiß, T. Mikosch, pp. 71-84, Berlin Heidelberg: Springer Verlag.

Kazakevičius, V. and Leipus, R. (2002). On stationarity in the ARCH(∞) model, *Econometric Theory*, 18, 1-16.

Kazakevičius, V. and Leipus, R. (2003). A new theorem on existence of invariant distributions with applications to ARCH processes, *Journal of Applied Probability*, 40, 147-162.

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

Linton, O. B. and Mammen, E. (2005). Estimating semiparametric ARCH(∞) models by kernel smoothing methods, *Econometrica*, 73, 771-836.

Loeve, M. (1977): Probability Theory 1, 4th ed. Springer, New York.

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

Nelson, D. B. (1990). Stationarity and Persistence in the GARCH(1,1) Model, *Econometric Theory*, 6, 318-334.

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. (2006). Pseudo-maximum likelihood estimation of ARCH(∞) models, *The Annals of Statistics*, 34, 1049–1074.

Shinki, K. and Zhang, Z. (2012). Asymptotic theory for fractionally asymmetric power ARCH models, *Journal of Statistical Planning and Inference*, 142, 2871-2890.

Straumann, D. (2005). *Estimation in Conditionally Heteroscedastic Time Series Models.* Lecture Notes in Statistics 181, Springer Verlag, Berlin.

Tayefi, M. and Ramanathan, T. V. (2012). An overview of FIGARCH and related time series models, *Austrian Journal of Statistics*, 41, 175-196.

White, H. (1994). *Asymptotic Theory for Econometricians.* Academic Press, New York.