Bernoulli **9**(2), 2003, 201–227

GARCH processes: structure and estimation

ISTVÁN BERKES[1], LAJOS HORVÁTH[2] and PIOTR KOKOSZKA[3]

[1]*A. Rényi Institute of Mathematics, Hungarian Academy of Sciences, P.O. Box 127, H-1364 Budapest, Hungary. E-mail: berkes@renyi.hu*
[2]*Department of Mathematics, University of Utah, 155 South 1440 East, Salt Lake City UT 84112-0090, USA. E-mail: horvath@math.utah.edu*
[3]*Department of Mathematics and Statistics, Utah State University, 3900 Old Main Hill, Logan UT 84322-3900, USA. E-mail: piotr@math.usu.edu*

We study the structure of a GARCH(p, q) sequence. We show that the conditional variance can be written as an infinite sum of the squares of the previous observations and that the representation is unique. We prove the consistency and asymptotic normality of the quasi-maximum likelihood estimator of the parameters of the GARCH(p, q) sequence under mild conditions.

Keywords: asymptotic normality; consistency; GARCH(p, q) sequence; martingales; quasi-maximum likelihood

1. Introduction

The analysis of financial data has received considerable attention in the literature over the last 20 years. Several models have been suggested for capturing special features of financial data, and most of these models have the property that the conditional variance (or the conditional scaling) depends on the past. One of the best-known and most often used examples is the autoregressive conditionally heteroscedastic (ARCH) process introduced by Engle (1982). The ARCH model has been investigated and generalized by several authors, including Bollerslev (1986) and Gouriéroux (1997). The theoretical results on ARCH and related properties have played a special role in empirical work in the analysis of data on exchange rates, stock prices and so on. In this paper we study the asymptotic properties of the generalized autoregressive conditionally heteroscedastic (GARCH) process introduced by Bollerslev (1986). A GARCH(p, q) process is defined by the equations

$$y_k = \sigma_k \varepsilon_k \tag{1.1}$$

and

$$\sigma_k^2 = \omega + \sum_{1 \leqslant i \leqslant p} \alpha_i y_{k-i}^2 + \sum_{1 \leqslant j \leqslant q} \beta_j \sigma_{k-j}^2, \tag{1.2}$$

where

$$\omega > 0, \qquad \alpha_i \geqslant 0, \qquad 1 \leqslant i \leqslant p, \qquad \beta_j \geqslant 0, \qquad 1 \leqslant j \leqslant q \tag{1.3}$$

are constants. We also assume that

$$\{\varepsilon_i, -\infty < i < \infty\} \text{ are independent, identically distributed random variables.} \tag{1.4}$$

Nelson (1990) showed that in case of a GARCH(1, 1) sequence, (1.1) and (1.2) have a unique stationary solution if and only if $\mathrm{E}\log(\beta_1 + \alpha_1\varepsilon_0^2) < 0$. Bougerol and Picard (1992a; 1992b) found necessary and sufficient conditions for the existence of a unique stationary solution of (1.1) and (1.2) in the case of a general GARCH(p, q) model. To state their condition we must introduce further notation. Let

$$\boldsymbol{\tau}_n = (\beta_1 + \alpha_1\varepsilon_n^2, \beta_2, \ldots, \beta_{q-1}) \in \mathbb{R}^{q-1},$$

$$\boldsymbol{\xi}_n = (\varepsilon_n^2, 0, \ldots, 0) \in \mathbb{R}^{q-1}$$

and

$$\boldsymbol{\alpha} = (\alpha_2, \ldots, \alpha_{p-1}) \in \mathbb{R}^{p-2}.$$

(Clearly, by including extra terms with zero coefficients in (1.2) we can achieve $\min(p, q) \geqslant 2$.) Define the $(p+q-1)\times(p+q-1)$ matrix A_n, written in block form, by

$$A_n = \begin{bmatrix} \boldsymbol{\tau}_n & \beta_q & \boldsymbol{\alpha} & \alpha_p \\ I_{q-1} & 0 & 0 & 0 \\ \boldsymbol{\xi}_n & 0 & 0 & 0 \\ 0 & 0 & I_{p-2} & 0 \end{bmatrix},$$

where I_{q-1} and I_{p-2} are the identity matrices of size $q-1$ and $p-2$, respectively. The norm of any $d \times d$ matrix M is defined by

$$\|M\| = \sup\{\|M\mathbf{x}\|_d/\|\mathbf{x}\|_d : \mathbf{x} \in \mathbb{R}^d, \mathbf{x} \neq \mathbf{0}\},$$

where $\|\cdot\|_d$ is the usual (Euclidean) norm in $\mathbb{R}^d$. The top Lyapunov exponent γ_L associated with the sequence $\{A_n, -\infty < n < \infty\}$ is

$$\gamma_L = \inf_{0 \leqslant n < \infty} \frac{1}{n+1}\mathrm{E}\log\|A_0A_1 \ldots A_n\|,$$

assuming that

$$\mathrm{E}(\log\|A_0\|) < \infty. \tag{1.5}$$

(We note that $\|A_0(1, 0, \ldots,)^{\mathrm{T}}\| \geqslant 1$ and therefore $\|A_0\| \geqslant 1$.) Condition (1.5) and the subadditive ergodic theorem (cf. Kingman 1973) imply

$$\lim_{n\to\infty} \frac{1}{n+1}\log\|A_0A_1 \ldots A_n\| = \gamma_L \text{ almost surely.}$$

Bougerol and Picard (1992a; 1992b) showed that if (1.5) holds, then (1.1) and (1.2) have a unique strictly stationary solution if and only if

$$\gamma_L < 0. \tag{1.6}$$

Let $\mathbf{X}_n = (\sigma_n^2, \ldots, \sigma_{n-q-1}^2, y_{n-1}^2, \ldots, y_{n-p-1}^2)^{\mathrm{T}} \in \mathbb{R}^{p+q-1}$ and $\mathbf{D} = (\omega, 0, \ldots, 0)^{\mathrm{T}} \in \mathbb{R}^{p+q-1}$. Equations (1.1) and (1.2) can be written equivalently as

$$\mathbf{X}_{n+1} = A_n\mathbf{X}_n + \mathbf{D}.$$

Bougerol and Picard (1992a; 1992b) showed that if (1.6) holds, then

$$\mathbf{X}_n = \mathbf{D} + \sum_{0 \leq k < \infty} A_n \cdots A_{n-k}\mathbf{D}. \tag{1.7}$$

Throughout this paper we will assume that conditions (1.1)–(1.6) hold. Clearly, they are a minimal set of conditions for the existence and stationarity of the GARCH(p, q) sequence.

Let $\boldsymbol{\theta} = (\omega, \alpha_1, \alpha_2, \ldots, \alpha_p, \beta_1, \beta_2, \ldots, \beta_q)$. Assuming that $y_1, y_2, \ldots, y_n$ have been observed, Lumsdaine (1996) studied the estimation of the unknown parameter $\boldsymbol{\theta}$ in the case of $p = q = 1$. Lumsdaine (1996) advocated the quasi-maximum likelihood estimator and proved its consistency and asymptotic normality for the GARCH(1, 1) model. However, some conditions in Lumsdaine (1996) seem to be unnecessarily restrictive and should be relaxed. Lee and Hansen (1994) impose weaker conditions on the error sequence.

The main goal of our paper is to provide rigorous proofs of the consistency and asymptotic normality of the quasi-maximum likelihood estimator in GARCH(p, q) models under weaker conditions. These results will be given in Section 4, where we also compare our theorems with some earlier results. The estimation in Lumsdaine (1996) is based on a representation of σ_k^2 in terms of past observations $\{y_i, -\infty < i < k\}$. In Section 2 we obtain a similar result for GARCH(p, q) sequences and prove that this representation is unique. The representation for GARCH(p, q) is an infinite sum and the coefficients satisfy a recursion. In Section 3 we establish some basic properties of the solution of the recursions; these properties will be used in Section 4, where the asymptotic properties of the quasi-maximum likelihood estimator will be discussed.

2. Representations for GARCH(p, q)

For a strictly stationary GARCH(p, q) process with coefficients $(\omega, \alpha_1, \ldots, \alpha_p, \beta_1, \ldots, \beta_q)$, let

$$\mathcal{A}(x) = \alpha_1 x + \alpha_2 x^2 + \ldots + \alpha_p x^p$$

and

$$\mathcal{B}(x) = 1 - \beta_1 x - \beta_2 x^2 - \ldots - \beta_q x^q.$$

Here and in the following we do not need $\min(p, q) \geq 2$, and the possible inclusion of extra terms in (1.2); this was needed in Section 1 only for computing the top Lyapunov exponent γ_L. We assume that the order of $\mathcal{B}(x)$ is exactly q, i.e.

$$\beta_q \neq 0. \tag{2.1}$$

Bougerol and Picard (1992b) showed that $\gamma_L < 0$ implies

$$\beta_1 + \beta_2 + \ldots + \beta_q < 1. \tag{2.2}$$

Relation (2.2) entails that all roots of $\mathcal{B}(x) = 0$ lie outside of the unit circle. More precisely, we have:

Lemma 2.1. *Relation* (2.2) *is equivalent to*

$$|\gamma_i| > 1 \text{ for all } 1 \leqslant i \leqslant \ell, \text{ where } \gamma_1, \gamma_2, \ldots, \gamma_\ell \text{ stand for the solutions of } \mathcal{B}(x) = 0 \text{ with multiplicities } \nu_1, \ldots, \nu_\ell. \tag{2.3}$$

Proof. Let us assume first that $\beta_1 + \beta_2 + \ldots + \beta_q \geqslant 1$. Since $\mathcal{B}(0) = 1$ and $\mathcal{B}(1) = 1 - (\beta_1 + \beta_2 + \ldots + \beta_q) \leqslant 0$, we have at least one solution of $\mathcal{B}(x) = 0$ in the interval (0, 1], contradicting (2.3).

Let us assume now that (2.2) holds. Then, for any $|z| \leqslant 1$, we have $|\mathcal{B}(z)| \geqslant 1 - (\beta_1|z| + \beta_2|z|^2 + \ldots + \beta_q|z|^q) \geqslant 1 - (\beta_1 + \beta_2 + \ldots + \beta_q) > 0$, and therefore (2.3) must be true. □

We will also need the following simple lemma. Let $\log^+ x = \log x$ if $x > 1$, and 0 otherwise.

Lemma 2.2. *If* $\{\xi_k, 0 \leqslant k < \infty\}$ *is a sequence of identically distributed random variables satisfying*

$$\mathrm{E} \log^+ |\xi_0| < \infty, \tag{2.4}$$

then $\sum_{0 \leqslant k < \infty} \xi_k z^k$ *converges with probability one for any* $|z| < 1$.

Proof. By the Borel–Cantelli lemma it is enough to prove that, for any $\zeta > 1$,

$$\sum_{1 \leqslant k < \infty} P\{|\xi_k| > \zeta^k\} < \infty. \tag{2.5}$$

The distribution of ξ_k does not depend on k, so

$$\begin{aligned} \sum_{1 \leqslant k < \infty} P\{|\xi_k| > \zeta^k\} &= \sum_{1 \leqslant k < \infty} P\{\log^+ |\xi_k| > k \log \zeta\} \\ &= \sum_{1 \leqslant k < \infty} P\{\log^+ |\xi_0| > k \log \zeta\} \\ &\leqslant \mathrm{E} \log^+ |\xi_0| / \log \zeta, \end{aligned}$$

and thus (2.4) implies (2.5). □

We now establish a representation for σ_k^2 in terms of the y_{k-i}^2, $i \geqslant 1$. Since, by Lemma 2.1, $\mathcal{B}(x)$ has all roots outside the unit disc, we have

$$\sum_{j=0}^{\infty} d_j x^j = \frac{1}{\mathcal{B}(x)}, \qquad |x| \leqslant 1, \tag{2.6}$$

and the coefficients $d_0, d_1, d_2, \ldots$ decay exponentially fast. Let

$$c_0 = \omega \sum_{0 \leqslant m < \infty} d_m \tag{2.7}$$

and

$$c_j = \alpha_1 d_{j-1} + \alpha_2 d_{j-2} + \ldots + \alpha_p d_{j-p}, \qquad 1 \leqslant j < \infty. \tag{2.8}$$

We note that $c_0 = \omega/\mathcal{B}(1)$ and

$$\frac{\mathcal{A}(x)}{\mathcal{B}(x)} = \sum_{1 \leqslant i < \infty} c_i x^i, \qquad |x| \leqslant 1, \tag{2.9}$$

and by (2.8) the coefficients $c_1, c_2, \ldots$ decay exponentially fast.

Theorem 2.1. *If*

$$\mathrm{E} \log \sigma_0^2 < \infty, \tag{2.10}$$

then

$$\sigma_k^2 = c_0 + \sum_{1 \leqslant i < \infty} c_i y_{k-i}^2, \qquad \textit{for all } k, \tag{2.11}$$

with probability one.

Proof. Since $\|A_0\| \geqslant \|\boldsymbol{\xi}_0\| = \varepsilon_0^2$, (1.5) yields that $\mathrm{E} \log^+ \varepsilon_0^2 \leqslant \mathrm{E} \log^+ \|A_0\| < \infty$. Since by (1.1) we have $\mathrm{E} \log^+ y_0^2 \leqslant \mathrm{E} \log^+ \sigma_0^2 + \mathrm{E} \log^+ \varepsilon_0^2$, from (2.10) we obtain that

$$\mathrm{E} \log^+ y_0^2 < \infty. \tag{2.12}$$

Since the sequence $c_1, c_2, \ldots$ decays exponentially fast, Lemma 2.2 yields that the series in (2.11) is absolutely convergent with probability one. It is clear that the stationary sequence

$$\xi_k = \omega + \sum_{1 \leqslant i \leqslant p} \alpha_i y_{k-i}^2 \tag{2.13}$$

satisfies (2.4).

We must show that

$$\sigma_k^2 = \sum_{0 \leqslant m < \infty} d_m \xi_{k-m}. \tag{2.14}$$

The series in (2.14) is absolutely convergent with probability one on account of Lemma 2.2 and the exponential decay of d_i. It follows from (2.6) that

$$
\begin{aligned}
d_0 &= 1,\\
d_1 &= \beta_1,\\
d_2 &= d_1\beta_1 + \beta_2,\\
&\vdots\\
d_q &= d_{q-1}\beta_1 + \ldots + d_1\beta_{q-1} + \beta_q
\end{aligned}
\tag{2.15}
$$

and

$$d_i = d_{i-1}\beta_1 + \ldots + d_{i-q}\beta_q, \qquad \text{for } i > q. \tag{2.16}$$

Using (2.15), (2.16) and induction, it is not difficult to verify that, for $j \geqslant q$,

$$\xi_k + d_1\xi_{k-1} + d_2\xi_{k-2} + \ldots + d_j\xi_{k-j} = \sigma_k^2 - \sum_{1\leqslant i\leqslant q}(d_{i+j-q}\beta_q + \ldots + d_j\beta_i)\sigma^2_{k-i-j}. \tag{2.17}$$

By Lemma 2.2, as $j \to \infty$, the left-hand side of (2.17) tends a.s. to the right-hand side of (2.14). Using the exponential decay of d_j and (2.10), we obtain that

$$\sum_{1\leqslant j<\infty} P\left\{\left|\sum_{1\leqslant i\leqslant q}(d_{i+j-q}\beta_q + \ldots + d_j\beta_i)\sigma^2_{k-i-j}\right| > \delta\right\} < \infty, \qquad \text{for any } \delta > 0,$$

and therefore, applying the Borel–Cantelli lemma, we conclude (2.11). □

Using the backward shift operator $\mathbf{S}$, we can write (2.11) succinctly as

$$\sigma_k^2 = \frac{\mathcal{A}(\mathbf{S})}{\mathcal{B}(\mathbf{S})}\left(\frac{\omega}{\mathcal{A}(1)} + y_k^2\right). \tag{2.18}$$

Using partial fractions it is possible to find an explicit recursive formula for the c_i in (2.11) in terms of the coefficients $\omega, \alpha_1, \ldots, \alpha_p, \beta_1, \ldots, \beta_q$.

It is sometimes useful to express σ_k^2 in terms of finitely many values of y_{k-i}^2 and a small remainder term, as in the following theorem whose proof is omitted.

Theorem 2.2. *Let* $R = \max(p, q)$. *If* (2.10) *holds, then for any* $k \geqslant R$

$$\sigma_k^2 = \sum_{0\leqslant m\leqslant k-R}\left(\omega + \sum_{1\leqslant i\leqslant p}\alpha_i y^2_{k-i-m}\right)d_m + Q_k, \tag{2.19}$$

where

$$
\begin{aligned}
Q_k = &\left(\beta_1\sigma^2_{R-1} + \cdots + \beta_q\sigma^2_{R-q}\right)d_{k-R} + \left(\beta_2\sigma^2_{R-1} + \cdots + \beta_q\sigma^2_{R-q+1}\right)d_{k-R-1}\\
&+ \cdots + \beta_q\sigma^2_{R-1}d_{k-R-q+1}.
\end{aligned}
$$

The next result shows that if there is a representation like (2.11), it must be unique.

Theorem 2.3. *We assume that*

$$\varepsilon_0^2 \text{ is a non-degenerate random variable.} \tag{2.20}$$

If, for some k,

$$\sigma_k^2 = c_0 + \sum_{1 \leqslant i < \infty} c_i y_{k-i}^2 \text{ a.s.} \quad \text{and} \quad \sigma_k^2 = c_0^* + \sum_{1 \leqslant i < \infty} c_i^* y_{k-i}^2 \text{ a.s.},$$

then $c_i = c_i^*$ *for all* $1 \leqslant i < \infty$.

Proof. We prove the result by contradiction. Let $m > 0$ be the smallest integer satisfying $c_m \neq c_m^*$. (If $c_j = c_j^*$ for all $j > 0$, then $c_0 = c_0^*$ must also hold.) By the definition of m, we have

$$(c_m^* - c_m) y_{k-m}^2 = c_0 - c_0^* + \sum_{m < j < \infty} (c_j - c_j^*) y_{k-j}^2,$$

and (1.1) yields that

$$\varepsilon_{k-m}^2 = \frac{1}{(c_m^* - c_m)\sigma_{k-m}^2} \left\{ c_0 - c_0^* + \sum_{m < j < \infty} (c_j - c_j^*) y_{k-j}^2 \right\}. \tag{2.21}$$

Since $\sigma_{k-m}^2 \geqslant \omega > 0$, ε_{k-m}^2 is well defined. Let $\mathcal{F}_j$ be the σ-algebra generated by $\{\varepsilon_i, -\infty < i \leqslant j\}$. Relation (1.7) shows that y_j is $\mathcal{F}_j$-measurable and thus the right-hand side of (2.21) (and consequently also ε_{k-m}^2) is a real-valued random variable, measurable with respect to $\mathcal{F}_{k-m-1}$. Since the ε_j are independent, this implies that ε_{k-m}^2 is a.s. constant, contradicting (2.20). □

Putting together Theorems 2.1 and 2.3, we have the following result:

Theorem 2.4. *If* (2.10) *is satisfied, then* (2.11) *holds with exponentially decaying weights* c_i. *If* (2.20) *is also satisfied, then the representation* (2.11) *is unique.*

Lemma 2.3. *If*

$$\mathrm{E}|\varepsilon_0^2|^\delta < \infty \qquad \text{for some } \delta > 0, \tag{2.22}$$

then there is $\delta^* > 0$ *such that*

$$\mathrm{E}|y_0^2|^{\delta^*} < \infty \quad \text{and} \quad \mathrm{E}|\sigma_0^2|^{\delta^*} < \infty.$$

Proof. By (1.6) and the definition of γ_L, there is an integer $m \geqslant 1$ such that

$$\mathrm{E} \log \|A_0 A_1 \cdots A_{m-1}\| < 0. \tag{2.23}$$

Also, $\|A_0\| \leqslant C(1 + \varepsilon_0^2)$, and thus by (2.22) we have that

$$\mathrm{E}\|A_0 A_1 \cdots A_{m-1}\|^\delta \leqslant (\mathrm{E}\|A_0\|^\delta)^m < \infty. \tag{2.24}$$

Following Remark 2.6 in Basrak *et al*. (2001), we introduce the function $s(t) = \mathrm{E}\|A_0A_1 \cdots A_{m-1}\|^t$. Since $s'(0) = \mathrm{E}\log\|A_0A_1 \cdots A_{m-1}\| < 0$, $s(t)$ decreases in a neighbourhood of 0, and since $s(0) = 1$, there exists $0 < \delta^* < 1$ such that

$$\mathrm{E}\|A_0A_1 \cdots A_{m-1}\|^{\delta^*} < 1. \tag{2.25}$$

Using (1.7), we conclude that

$$\|\mathbf{X}_0\| \leqslant \|\mathbf{D}\| + \sum_{0\leqslant k<\infty} \|A_0 \cdots A_{-k}\| \, \|\mathbf{D}\|,$$

and since $0 < \delta^* < 1$, we obtain that

$$\|\mathbf{X}_0\|^{\delta^*} \leqslant \|\mathbf{D}\|^{\delta^*} + \sum_{0\leqslant k<\infty} \|A_0 \cdots A_{-k}\|^{\delta^*} \|\mathbf{D}\|^{\delta^*}.$$

Using (2.25), it follows easily that there exist $0 < \tilde{c} < \infty$ and $0 < \tilde{\rho} < 1$ such that

$$\mathrm{E}\|A_0A_1 \cdots A_k\|^{\delta^*} \leqslant \tilde{c}\tilde{\rho}^k,$$

proving that $\mathrm{E}\|X_0\|^{\delta^*} < \infty$, which implies the conclusion of Lemma 2.3. □

Theorem 2.1 and 2.3 show that there is a one-to-one correspondence between the sequence $\{\sigma_k^2, -\infty < k < \infty\}$ and the coefficients c_i, $0 \leqslant i < \infty$. However, this fact is not enough to estimate the parameters of a GARCH(p, q) sequence. For this purpose we also need the definition (1.2) to be *minimal* in the sense that there is no pair (p^*, q^*) such that $p^* < p$ or $q^* < q$ and

$$\sigma_k^2 = \omega^* + \sum_{1\leqslant i\leqslant p^*} \alpha_i^* y_{k-i}^2 + \sum_{1\leqslant j\leqslant q^*} \beta_j^* \sigma_{k-j}^2 \tag{2.26}$$

for some (not necessarily non-negative) ω^*, α_i^* $(1 \leqslant i \leqslant p^*)$ and β_j^* $(1 \leqslant j \leqslant q^*)$.

Theorem 2.5. *We assume that* (2.10) *and* (2.20) *are satisfied. Then the definition* (1.1)–(1.2) *is minimal if and only if*

$$\begin{aligned}&\textit{the polynomials } \mathcal{A}(x) \textit{ and } \mathcal{B}(x) \textit{ are coprimes}\\ &\textit{in the set of polynomials with real coefficients}.\end{aligned} \tag{2.27}$$

Proof. Let us assume that (2.27) holds and assume indirectly that there exist (p^*, q^*), $p^* < p$ or $q^* < q$, and ω^*, α_i^*, $1 \leqslant i \leqslant p^*$, β_j^*, $1 \leqslant j \leqslant q^*$ such that (2.26) holds. Let $\mathcal{A}^*(x) = x\alpha_1^* + \ldots + x^{p^*}\alpha_{p^*}^*$ and $\mathcal{B}^*(x) = 1 - (\beta_1^* x + \ldots + \beta_{q^*}^* x^{q^*})$. Similarly to (2.9), we have

$$\sum_{1 \leqslant i < \infty} c_i x^i = \frac{\mathcal{A}^*(x)}{\mathcal{B}^*(x)}, \tag{2.28}$$

and consequently $\mathcal{A}(x)/\mathcal{B}(x) = \mathcal{A}^*(x)/\mathcal{B}^*(x)$. By (2.27), $\mathcal{A}(x)$ and $\mathcal{B}(x)$ are coprimes, and thus we conclude that there is a polynomial $\mathcal{P}(x)$ such that $\mathcal{A}^*(x) = \mathcal{A}(x)\mathcal{P}(x)$ and $\mathcal{B}^*(x) = \mathcal{B}(x)\mathcal{P}(x)$, and consequently $p^* \geqslant p$, $q^* \geqslant q$, a contradiction.

Let us assume conversely that the definition (1.1)–(1.2) is minimal but (2.27) fails, i.e. there exist polynomials $\mathcal{A}^*$, $\mathcal{B}^*$ and $\mathcal{P}$ such that $\mathcal{P}$ is non-constant and $\mathcal{A}(x) = \mathcal{A}^*(x)\mathcal{P}(x)$ and $\mathcal{B}(x) = \mathcal{B}^*(x)\mathcal{P}(x)$. We now show that with $\tilde{\omega} = \omega\mathcal{B}^*(1)/\mathcal{B}(1)$, we have

$$\sigma_k^2 = \tilde{\omega} + \sum_{1 \leqslant i \leqslant p^*} \alpha_i^* y_{k-i}^2 + \sum_{1 \leqslant j \leqslant q^*} \beta_j^* \sigma_{k-j}^2. \tag{2.29}$$

Indeed, by (2.18) and (2.28) we obtain that

$$\sigma_k^2 = \frac{\omega}{\mathcal{B}(1)} + \frac{\mathcal{A}^*(\mathbf{S})}{\mathcal{B}^*(\mathbf{S})} y_k^2 = \frac{\tilde{\omega}}{\mathcal{B}^*(1)} + \frac{\mathcal{A}^*(\mathbf{S})}{\mathcal{B}^*(\mathbf{S})} y_k^2.$$

Hence

$$\mathcal{B}^*(\mathbf{S})\sigma_k^2 = \mathcal{A}^*(\mathbf{S})\left(\frac{\tilde{\omega}}{\mathcal{A}^*(1)} + y_k^2\right).$$

Observing that the degrees of $\mathcal{A}^*$ and $\mathcal{B}^*$ are less than the degrees of $\mathcal{A}$ and $\mathcal{B}$, (2.29) contradicts the minimality assumption. □

Corollary 2.1. *We assume that* (2.10), (2.20) *and* (2.27) *are satisfied. Then there is no* $(\omega^*, \alpha_1^*, \ldots, \alpha_p^*, \beta_1^*, \ldots, \beta_q^*) \neq (\omega, \alpha_1, \ldots, \alpha_p, \beta_1, \ldots, \beta_q)$ *such that*

$$\sigma_k^2 = \omega^* + \sum_{1 \leqslant i \leqslant p} \alpha_i^* y_{k-i}^2 + \sum_{1 \leqslant j \leqslant q} \beta_j^* y_{k-j}^2. \tag{2.30}$$

Proof. Let $(\omega^*, \alpha_1^*, \ldots, \alpha_p^*, \beta_1^*, \ldots, \beta_q^*)$ be a sequence satisfying (2.30) and let $\mathcal{A}^*$, $\mathcal{B}^*$ be the analogues of $\mathcal{A}$ and $\mathcal{B}$ for the sequence $(\omega^*, \alpha_1^*, \ldots, \alpha_p^*, \beta_1^*, \ldots, \beta_q^*)$. Following the first part of the proof of Theorem 2.5, we obtain that there exists a polynomial $\mathcal{P}$ such that $\mathcal{A}^*(x) = \mathcal{A}(x)\mathcal{P}(x)$ and $\mathcal{B}^*(x) = \mathcal{B}(x)\mathcal{P}(x)$. Since $\mathcal{B}$ and $\mathcal{B}^*$ have the same degree q and the same constant term 1, we obtain $\mathcal{P}(x) = 1$ and Corollary 2.1 is proven. □

The next section contains some preliminary results used in the construction of the quasi-maximum likelihood estimator.

3. Recursions related to the infinite representation of GARCH(p, q)

In Section 2 we gave explicit formulae for the coefficients c_i in (2.11) in terms of the roots $\gamma_1, \ldots, \gamma_\ell$ of the polynomial $\mathcal{B}$. However, the computation of the solutions of $\mathcal{B}(x) = 0$ may not be simple, especially if q is large, and thus our formulae are numerically impractical. By (2.9), we have

$$c_n = \frac{d^n}{dx^n}\left(\frac{\mathcal{A}(x)}{\mathcal{B}(x)}\right)_{x=0} \qquad 1 \leqslant n < \infty,$$

and computing the derivatives on the right-hand side and using $\mathcal{B}(0) = 1$, we see that c_n, $1 \leqslant n < \infty$, are actually polynomials of $\alpha_1, \ldots, \alpha_p, \beta_1, \ldots, \beta_q$ with integer coefficients. Letting $C(x) = \sum_{1 \leqslant n < \infty} c_n x^n$, (2.9) shows that $\mathcal{A}(x) = \mathcal{B}(x)C(x)$, and performing the multiplication on the right-hand side and comparing the coefficients, we obtain

$$\alpha_1 = c_1, \qquad \alpha_2 = c_2 - \beta_1 c_1, \qquad \alpha_3 = c_3 - \beta_1 c_2 - \beta_2 c_1, \ldots$$

from which $c_1, c_2, \ldots$ can be computed recursively.

In the estimation problem studied in Sections 4 and 5, the parameter $\boldsymbol{\theta} = (\omega, \alpha_1, \ldots, \alpha_p, \beta_1, \ldots, \beta_q)$ is fixed but unknown. Thus we are dealing with a class of processes whose parameter vector will be denoted by $\mathbf{u} = (x, s_1, \ldots, s_p, t_1, \ldots, t_q)$. Hence the coefficients c_i, $0 \leqslant i < \infty$, will be functions of $\mathbf{u}$ and the recursion formulae obtained above take on the following form. If $q \geqslant p$, then

$$\begin{aligned}
c_0(\mathbf{u}) &= \frac{x}{1 - (t_1 + \ldots + t_q)},\\
c_1(\mathbf{u}) &= s_1,\\
c_2(\mathbf{u}) &= s_2 + t_1 c_1(\mathbf{u}),\\
&\vdots\\
c_p(\mathbf{u}) &= s_p + t_1 c_{p-1}(\mathbf{u}) + \ldots + t_{p-1} c_1(\mathbf{u}),\\
c_{p+1}(\mathbf{u}) &= t_1 c_p(\mathbf{u}) + \ldots + t_p c_1(\mathbf{u}),\\
&\vdots\\
c_q(\mathbf{u}) &= t_1 c_{q-1}(\mathbf{u}) + \ldots + t_{q-1} c_1(\mathbf{u}).
\end{aligned}$$

If $q < p$ the equations above are replaced by

$$c_0(\mathbf{u}) = \frac{x}{1 - (t_1 + \ldots + t_q)},$$

$$c_1(\mathbf{u}) = s_1,$$

$$c_2(\mathbf{u}) = s_2 + t_1 c_1(\mathbf{u}),$$

$$\vdots$$

$$c_{q+1}(\mathbf{u}) = s_{q+1} + t_1 c_q(\mathbf{u}) + \ldots + t_q c_1(\mathbf{u}),$$

$$\vdots$$

$$c_p(\mathbf{u}) = s_p + t_1 c_{p-1}(\mathbf{u}) + \ldots + t_q c_{p-q}(\mathbf{u}).$$

If $i > R = \max(p, q)$, then

$$c_i(\mathbf{u}) = t_1 c_{i-1}(\mathbf{u}) + t_2 c_{i-2}(\mathbf{u}) + \ldots + t_q c_{i-q}(\mathbf{u}). \tag{3.1}$$

Let $0 < \underline{u} < \bar{u}$, $0 < \rho_0 < 1$, $q\underline{u} < \rho_0$ and define

$$U = \{\mathbf{u} : t_1 + t_2 + \ldots + t_q \leqslant \rho_0 \text{ and } \underline{u} \leqslant \min(x, s_1, s_2, \ldots, s_p, t_1, t_2, \ldots, t_q) \leqslant \max(x, s_1, s_2, \ldots, s_p, t_1, t_2, \ldots, t_q) \leqslant \bar{u}\}.$$

We assume that

$$\boldsymbol{\theta} \text{ is in the interior of } U. \tag{3.2}$$

(Condition (3.2) requires that p and q are known. It also rules out zero coefficients in $\boldsymbol{\theta}$.) We use $|\cdot|$ to denote the maximum norm of vectors and matrices. Let $x \vee y = \max(x, y)$.

Lemma 3.1. *For any* $\mathbf{u} = (x, s_1, \ldots, s_p, t_1, \ldots, t_q) \in U$ *and* $\mathbf{u}^* = (x^*, s_1^*, \ldots, s_p^*, t_1^*, \ldots, t_q^*) \in U$, *we have*

$$C_1 \underline{u}^i \leqslant c_i(\mathbf{u}), \qquad 0 \leqslant i < \infty, \tag{3.3}$$

$$c_i(\mathbf{u}) \leqslant C_2 \rho_0^{i/q}, \qquad 0 \leqslant i < \infty, \tag{3.4}$$

and

$$\frac{c_i(\mathbf{u}^*)}{c_i(\mathbf{u})} \leqslant C_3 \left(\max_{1 \leqslant j \leqslant q} \frac{t_j^*}{t_j} \vee 1 \right)^i, \qquad 0 \leqslant i < \infty, \tag{3.5}$$

for constants $0 < C_1, C_2, C_3 < \infty$.

Proof. The results are trivial for $i = 0$. We use induction for $i \geqslant 1$. It is clear that $c_i(\mathbf{u})$, $i \geqslant 1$, are polynomials of the coordinates of $\mathbf{u}$ with positive coefficients. Hence (3.3)–(3.5) hold for all $1 \leqslant i \leqslant R = \max(p, q)$ if C_1 is chosen small and C_2, C_3 are chosen large enough.

Let $j > R$ and assume that (3.3)–(3.5) are valid for all $i < j$. Since $\underline{u} < 1$, by (3.1) we have

$$c_j(\mathbf{u}) \geqslant \underline{u} \min_{1 \leqslant k \leqslant q} c_{j-k}(\mathbf{u}) \geqslant C_1 \underline{u}^j.$$

Also, by (3.1) we have

$$c_j(\mathbf{u}) \leqslant (t_1 + t_2 + \ldots + t_q) \max_{1 \leqslant k \leqslant q} c_{j-k}(\mathbf{u}) \leqslant \rho_0 C_2 \rho_0^{(j-q)/q} \leqslant C_2 \rho_0^{j/q}.$$

Finally, by (3.1) and the induction hypothesis we have, for $j > R$,

$$0 < \frac{c_j(\mathbf{u}^*)}{c_j(\mathbf{u})} = t_1^* \frac{c_{j-1}(\mathbf{u}^*)}{c_j(\mathbf{u})} + t_2^* \frac{c_{j-2}(\mathbf{u}^*)}{c_j(\mathbf{u})} + \ldots + t_q^* \frac{c_{j-q}(\mathbf{u}^*)}{c_j(\mathbf{u})}$$

$$= \frac{t_1^*}{t_1} \frac{c_{j-1}(\mathbf{u}^*)}{c_{j-1}(\mathbf{u})} t_1 \frac{c_{j-1}(\mathbf{u})}{c_j(\mathbf{u})} + \frac{t_2^*}{t_2} \frac{c_{j-2}(\mathbf{u}^*)}{c_{j-2}(\mathbf{u})} t_2 \frac{c_{j-2}(\mathbf{u})}{c_j(\mathbf{u})} + \ldots + \frac{t_q^*}{t_q} \frac{c_{j-q}(\mathbf{u}^*)}{c_{j-q}(\mathbf{u})} t_q \frac{c_{j-q}(\mathbf{u})}{c_j(\mathbf{u})}$$

$$\leqslant C_3 \left(\max_{1 \leqslant i \leqslant q} \frac{t_i^*}{t_i} \vee 1 \right)^{j-1} \frac{1}{c_j(\mathbf{u})} \{ t_1 c_{j-1}(\mathbf{u}) + t_2 c_{j-2}(\mathbf{u}) + \ldots + t_q c_{j-q}(\mathbf{u}) \} \left(\max_{1 \leqslant i \leqslant q} \frac{t_i^*}{t_i} \right)$$

$$\leqslant C_3 \left(\max_{1 \leqslant i \leqslant q} \frac{t_i^*}{t_i} \vee 1 \right)^{j}.$$

Thus (3.3)–(3.5) also hold for j and the proof of Lemma 3.1 is complete. □

Next we prove similar results for the vector

$$c_i'(\mathbf{u}) = \left(\frac{\partial c_i(\mathbf{u})}{\partial x}, \frac{\partial c_i(\mathbf{u})}{\partial s_1}, \ldots, \frac{\partial c_i(\mathbf{u})}{\partial s_p}, \frac{\partial c_i(\mathbf{u})}{\partial t_1}, \ldots, \frac{\partial c_i(\mathbf{u})}{\partial t_q} \right).$$

Lemma 3.2. *We assume that* $\mathbf{u} \in U$. *Then*

$$\left| \frac{\partial c_0(\mathbf{u})}{\partial x} \right| \leqslant \frac{1}{1 - \rho_0}, \tag{3.6}$$

$$\frac{\partial c_0(\mathbf{u})}{\partial s_j} = 0, \qquad 1 \leqslant j \leqslant p, \tag{3.7}$$

$$\left| \frac{\partial c_0(\mathbf{u})}{\partial t_j} \right| \leqslant \frac{\bar{u}}{(1 - \rho_0)^2}, \qquad 1 \leqslant j \leqslant q, \tag{3.8}$$

and

$$|c_i'(\mathbf{u})| / c_i(\mathbf{u}) \leqslant C_4 i, \qquad 1 \leqslant i < \infty, \tag{3.9}$$

for some constant C_4.

Proof. Since $c_0(\mathbf{u}) = x/(1-(t_1 + \ldots + t_q))$, (3.6)–(3.8) are obvious. To prove (3.9), we begin by observing that

$$\frac{\partial c_i(\mathbf{u})}{\partial x} = 0, \qquad 1 \leqslant i < \infty. \tag{3.10}$$

Next, we show that

$$\left| \frac{\partial c_i(\mathbf{u})}{\partial s_j} / c_i(\mathbf{u}) \right| \leqslant C_5, \qquad 1 \leqslant i < \infty, 1 \leqslant j \leqslant p, \tag{3.11}$$

for some constant C_5. We again use induction. Clearly, (3.11) holds for $i \leqslant R = \max(p, q)$, provided C_5 is large enough. By (3.1) we have, for $i > R$,

$$\frac{\partial c_i(\mathbf{u})}{\partial s_j} = t_1 \frac{\partial c_{i-1}(\mathbf{u})}{\partial s_j} + t_2 \frac{\partial c_{i-2}(\mathbf{u})}{\partial s_j} + \ldots + t_q \frac{\partial c_{i-q}(\mathbf{u})}{\partial s_j},$$

and since $c_\ell > 0$, $\partial c_\ell / \partial s_j \geqslant 0$ (recall that c_ℓ is a polynomial of $s_1, \ldots, s_p, t_1, \ldots, t_q$ with positive coefficients), we conclude that

$$\begin{aligned} 0 < \frac{1}{c_i(\mathbf{u})} \frac{\partial c_i(\mathbf{u})}{\partial s_j} &\leqslant \max_{1 \leqslant m \leqslant q} \frac{1}{c_{i-m}(\mathbf{u})} \frac{\partial c_{i-m}(\mathbf{u})}{\partial s_j} \{ t_1 c_{i-1}(\mathbf{u}) + \ldots + t_q c_{i-q}(\mathbf{u}) \} \frac{1}{c_i(\mathbf{u})} \\ &= \max_{1 \leqslant m \leqslant q} \frac{1}{c_{i-m}(\mathbf{u})} \frac{\partial c_{i-m}(\mathbf{u})}{\partial s_j}. \end{aligned}$$

Hence if $i > R$ and (3.11) holds for all indices less than i, then it also holds for i, completing the induction step.

We must now show that

$$\left| \frac{\partial c_i(\mathbf{u})}{\partial t_j} / c_i(\mathbf{u}) \right| \leqslant C_6 i, \qquad 1 \leqslant i < \infty, 1 \leqslant j \leqslant q, \tag{3.12}$$

for some C_6. By (3.1) we have, for $i > R$,

$$\frac{\partial c_i(\mathbf{u})}{\partial t_j} = c_{i-j}(\mathbf{u}) + t_1 \frac{\partial c_{i-1}(\mathbf{u})}{\partial t_j} + \ldots + t_q \frac{c_{i-q}(\mathbf{u})}{\partial t_j}$$

and therefore

$$\frac{1}{c_i(\mathbf{u})} \frac{\partial c_i(\mathbf{u})}{\partial t_j} = \frac{c_{i-j}(\mathbf{u})}{c_i(\mathbf{u})} + \frac{1}{c_{i-1}(\mathbf{u})} \frac{\partial c_{i-1}(\mathbf{u})}{\partial t_j} t_1 \frac{c_{i-1}(\mathbf{u})}{c_i(\mathbf{u})} + \ldots + \frac{1}{c_{i-q}(\mathbf{u})} \frac{\partial c_{i-q}(\mathbf{u})}{\partial t_j} t_q \frac{c_{i-q}(\mathbf{u})}{c_i(\mathbf{u})}.$$

Since $c_i(\mathbf{u}) \geqslant t_j c_{i-j}(\mathbf{u})$, we obtain that

$$\begin{aligned} \left| \frac{1}{c_i(\mathbf{u})} \frac{\partial c_i(\mathbf{u})}{\partial t_j} \right| &\leqslant \frac{1}{t_j} + \max_{1 \leqslant m \leqslant q} \frac{1}{c_{i-m}(\mathbf{u})} \frac{\partial c_{i-m}}{\partial t_j} \frac{1}{c_i(\mathbf{u})} \{ t_1 c_{i-1}(\mathbf{u}) + \ldots + t_q c_{i-q}(\mathbf{u}) \} \\ &\leqslant \frac{1}{\underline{u}} + \max_{1 \leqslant m \leqslant q} \frac{1}{c_{i-m}(\mathbf{u})} \frac{\partial c_{i-m}(\mathbf{u})}{\partial t_j}, \end{aligned}$$

and therefore (3.12) follows again by induction, assuming that C_6 is chosen larger than $1/\underline{u}$.

Now (3.9) follows from (3.10)–(3.12). □

Our final lemma in this section shows that an estimate similar to (3.9) holds for the matrix $c_i''(\mathbf{u})$. We recall that $|c_i''(\mathbf{u})|$ denotes the maximum norm of the matrix $c_i''(\mathbf{u})$. Similarly, $|c_i^{(3)}(\mathbf{u})|$ denotes the largest of the absolute values of the of elements of the (hyper)matrix of the third partial derivatives.

Lemma 3.3. *For all* $\mathbf{u} \in U$, *we have that*

$$|c_0''(\mathbf{u})| \leqslant C_7,$$

$$|c_i''(\mathbf{u})| \leqslant C_8 i^2 c_i(\mathbf{u}), \qquad 1 \leqslant i < \infty,$$

$$|c_0^{(3)}(\mathbf{u})| \leqslant C_9$$

and

$$|c_i^{(3)}(\mathbf{u})| \leqslant C_{10} i^3 c_i(\mathbf{u}), \qquad 1 \leqslant i < \infty,$$

for some constants C_7, C_8, C_9 *and* C_{10}.

Proof. Using induction, Lemma 3.3 can be derived along the lines of the proof of Lemma 3.2 and therefore the details are omitted.

4. The quasi-maximum likelihood estimators

The logarithm of the quasi-maximum likelihood function in GARCH(p, q) is defined as

$$L_n(\mathbf{u}) = \sum_{1 \leqslant k \leqslant n} -\frac{1}{2}\left\{\log w_k(\mathbf{u}) + \frac{y_k^2}{w_k(\mathbf{u})}\right\}, \tag{4.1}$$

where

$$w_k(\mathbf{u}) = c_0(\mathbf{u}) + \sum_{1 \leqslant i < \infty} c_i(\mathbf{u}) y_{k-i}^2. \tag{4.2}$$

The functions $c_i(\mathbf{u})$, $0 \leqslant i < \infty$, are defined in Section 3. Putting together Lemma 2.2, (2.12) and (3.4), we obtain that $w_k(\mathbf{u})$ exists with probability one. Clearly, $w_k(\boldsymbol{\theta}) = \sigma_k^2$. If we were to assume that ε_0 is standard normal, then, conditionally on $\mathcal{F}_{k-1} = \sigma\{\varepsilon_i, -\infty < i \leqslant k-1\}$, $y_k/\sigma_k(\boldsymbol{\theta})$ is also standard normal. The likelihood in (4.1) is derived under this assumption. However, we will show that the quasi-maximum likelihood estimator $\hat{\boldsymbol{\theta}}_n$, defined as

$$\hat{\boldsymbol{\theta}}_n = \arg\max_{\mathbf{u} \in U} L_n(\mathbf{u}), \tag{4.3}$$

will be consistent and asymptotically normal without assuming that ε_0 is standard normal.

Theorem 4.1. *We assume that* (2.20)*,* (2.27)*,* (3.2) *hold and*

$$\mathrm{E}|\varepsilon_0^2|^{1+\delta} < \infty, \qquad \textit{for some } \delta > 0, \tag{4.4}$$

$$\lim_{t\to 0} t^{-\mu} P\{\varepsilon_0^2 \leqslant t\} = 0, \qquad \textit{for some } \mu > 0, \tag{4.5}$$

and

$$\mathrm{E}\varepsilon_0^2 = 1. \tag{4.6}$$

Then

$$\hat{\boldsymbol{\theta}}_n \xrightarrow{\text{a.s.}} \boldsymbol{\theta}, \qquad \textit{as } n \to \infty. \tag{4.7}$$

Remark 4.1. For GARCH(1, 1), (4.7) was established by Lumsdaine (1996) under much stronger conditions. For example, Lumsdaine assumed that $\mathrm{E}\varepsilon_0^{32} < \infty$ and ε_0 has a symmetric unimodal density, bounded in a neighborhood of 0. It is clear that in this case (4.5) holds for any $\mu < 1/2$.

Remark 4.2. Jeantheau (1998) announced the weak consistency of $\hat{\boldsymbol{\theta}}_n$ as a consequence of a more general theorem. However, his conditions are stronger than ours. For example, one of his conditions implies that $\mathrm{E}y_0^4 < \infty$ (cf. Jeantheau (1998, p. 76)). This condition is not required in Theorem 4.1. We wish to point out that the conditions of Theorem 4.1 also imply (cf. Lemma 2.3) that $\mathrm{E}(y_0^2)^{\delta^*} < \infty$ for some $\delta^* > 0$, but δ^* can be very small.

We prove Theorem 4.1, as well as Theorems 4.2–4.4 below, in the next section.

We now discuss the asymptotic normality of $n^{1/2}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta})$. Let

$$\ell_k(\mathbf{u}) = -\frac{1}{2}(\log w_k(\mathbf{u}) + y_k^2/w_k(\mathbf{u})) \tag{4.8}$$

and introduce the matrices $\mathbf{A}_0 = \mathrm{cov}(\ell_0'(\boldsymbol{\theta}))$ and $\mathbf{B}_0 = \mathrm{E}(\ell_0''(\boldsymbol{\theta}))$.

Theorem 4.2. *We assume that* (2.20)*,* (2.27)*,* (3.2)*,* (4.5)*,* (4.6) *hold and*

$$\mathrm{E}|\varepsilon_0^2|^{2+\delta} < \infty, \qquad \textit{for some } \delta > 0. \tag{4.9}$$

Then

$$A_0 \textit{ and } B_0 \textit{ are non-singular}, \tag{4.10}$$

$$\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta} = \frac{1}{n}\sum_{1\leqslant k\leqslant n}\frac{1}{2}(1-\varepsilon_k^2)\frac{w_k'(\boldsymbol{\theta})}{w_k(\boldsymbol{\theta})}\mathbf{B}_0^{-1} + o_P(n^{-1/2}), \qquad \textit{as } n \to \infty, \tag{4.11}$$

and

$$n^{1/2}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}) \xrightarrow{\mathcal{D}} N(\mathbf{0},\ \mathbf{B}_0^{-1}\mathbf{A}_0\mathbf{B}_0^{-1}), \qquad \textit{as } n \to \infty, \tag{4.12}$$

where $N(\mathbf{0}, \mathbf{C})$ *stands for a multivariate normal random variable with mean* $\mathbf{0}$ *and covariance matrix* $\mathbf{C}$.

Remark 4.3. Theorem 4.2 extends the central limit theorem in Lumsdaine (1996) by removing some unnecessary conditions and allowing arbitrary p and q.

Remark 4.4. Comte and Lieberman (2000) announced the asymptotic normality of the quasi-maximum likelihood estimator in GARCH(p, q) models. However, they assume that $\alpha_1 + \ldots + \alpha_p + \beta_1 + \ldots + \beta_q < 1$, which is a much stronger condition than ours. Also they assume that ε_0 has an absolutely continuous density, positive in a neighbourhood of 0. We only need (4.5).

In practice, we observe only $y_1, \ldots, y_n$ and the logarithm of the quasi-maximum function in (4.1) cannot be computed from our data. Hence we replace $L_n(\mathbf{u})$ with

$$\tilde{L}_n(\mathbf{u}) = \sum_{1<k\leqslant n} -\frac{1}{2}\left\{\log \tilde{w}_k(\mathbf{u}) + \frac{y_k^2}{\tilde{w}_k(\mathbf{u})}\right\},$$

where

$$\tilde{w}_k(\mathbf{u}) = c_0(\mathbf{u}) + \sum_{1\leqslant i\leqslant k-1} c_i(\mathbf{u}) y_{k-i}^2.$$

Similarly to (4.3), we define

$$\tilde{\boldsymbol{\theta}}_n = \arg\max_{\mathbf{u}\in U} \tilde{L}_n(\mathbf{u}).$$

The next two theorems show that our limit theorems will remain true for $\tilde{\boldsymbol{\theta}}_n$.

Theorem 4.3. *Under the conditions of Theorem 4.1 we have*

$$\tilde{\boldsymbol{\theta}}_n \xrightarrow{\text{a.s.}} \boldsymbol{\theta}, \qquad \text{as } n \to \infty.$$

Theorem 4.4. *Under the conditions of Theorem 4.2 we have*

$$\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta} = \frac{1}{n}\sum_{1\leqslant k\leqslant n} \frac{1}{2}(1-\varepsilon_k^2)\frac{w_k'(\boldsymbol{\theta})}{w_k(\boldsymbol{\theta})}\mathbf{B}_0^{-1} + o_P(n^{1/2}), \qquad \text{as } n \to \infty,$$

and

$$n^{1/2}(\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}) \xrightarrow{\mathcal{D}} \mathrm{N}(\mathbf{0},\ \mathbf{B}_0^{-1}\mathbf{A}_0\mathbf{B}_0^{-1}), \qquad \text{as } n \to \infty.$$

Remark 4.5. We note that $\mathbf{A}_0 = -\frac{1}{2}(\mathrm{E}\varepsilon_0^4 - 1)\mathbf{B}_0$. So if $\mathrm{E}\varepsilon_0^4 = 3$, as in the case of standard normal random innovations, then $\mathbf{A}_0 = -\mathbf{B}_0$.

Remark 4.6. We note that Theorems 4.3 and 4.4 remain true if $\tilde{w}_k(\mathbf{u})$ is replaced by $\tilde{w}_k(\mathbf{u}) + v_k(\mathbf{u})$, if $\sup_{\mathbf{u}\in U}|v_k^{(3)}(\mathbf{u})| = O(\tilde{\rho}^k)$ a.s. for some $0 < \tilde{\rho} < 1$. This claim will be immediate from the proofs of Theorems 4.3 and 4.4.

5. Proof of Theorems 4.1–4.4

We start with some technical results.

Lemma 5.1. *We assume that* (3.2), (4.5) *hold and*

$$\mathrm{E}|\varepsilon_0^2|^\gamma < \infty, \qquad \textit{for some } \gamma > 0. \tag{5.1}$$

Then, for any $0 < \nu < \gamma$, *we have*

$$\mathrm{E}\left\{\sup_{\mathbf{u}\in U} \frac{\sigma_k^2}{w_k(\mathbf{u})}\right\}^\nu < \infty. \tag{5.2}$$

Proof. For any $M \geqslant 1$, we have

$$\frac{\sigma_k^2}{w_k(\mathbf{u})} \leqslant \frac{\sigma_k^2}{\sum_{1\leqslant i\leqslant M} c_i(\mathbf{u}) y_{k-i}^2} = \frac{\sigma_k^2}{\sum_{1\leqslant i\leqslant M} c_i(\mathbf{u}) \varepsilon_{k-i}^2 \sigma_{k-i}^2}. \tag{5.3}$$

Since $\sigma_{k-1}^2 > \beta_i \sigma_{k-i-1}^2$, $1 \leqslant i \leqslant q$, and $\sigma_{k-1}^2 > \alpha_i y_{k-i-1}^2$, $1 \leqslant i \leqslant p$, and since $\sigma_{k-1}^2 \geqslant \omega$, we obtain

$$\begin{aligned}\frac{\sigma_k^2}{\sigma_{k-1}^2} &= \frac{\omega + \alpha_1 y_{k-1}^2 + \alpha_2 y_{k-2}^2 + \ldots + \alpha_p y_{k-p}^2 + \beta_1 \sigma_{k-1}^2 + \ldots + \beta_q \sigma_{k-q}^2}{\sigma_{k-1}^2} \\ &\leqslant 1 + \frac{\alpha_1 \varepsilon_{k-1}^2 \sigma_{k-1}^2}{\sigma_{k-1}^2} + \frac{\alpha_2}{\alpha_1} + \frac{\alpha_3}{\alpha_2} + \ldots + \frac{\alpha_p}{\alpha_{p-1}} \\ &\quad + \beta_1 + \frac{\beta_2}{\beta_1} + \frac{\beta_3}{\beta_2} + \ldots + \frac{\beta_q}{\beta_{q-1}} \\ &\leqslant K_1(1 + \varepsilon_{k-1}^2),\end{aligned} \tag{5.4}$$

for some constant $K_1 > 1$. It follows from (5.4) that

$$\frac{\sigma_k^2}{\sigma_{k-i}^2} \leqslant K_1^M \prod_{1\leqslant j\leqslant M} (1 + \varepsilon_{k-j}^2), \qquad \text{for all } 1 \leqslant i \leqslant M,$$

and thus (5.3) yields, in view of (3.3),

$$\begin{aligned}\frac{\sigma_k^2}{w_k(\mathbf{u})} &\leqslant K_1^M \prod_{1\leqslant j\leqslant M} \frac{1 + \varepsilon_{k-j}^2}{\sum_{1\leqslant i\leqslant M} c_i(\mathbf{u}) \varepsilon_{k-i}^2} \\ &\leqslant (K_2)^M \prod_{1\leqslant j\leqslant M} \frac{1 + \varepsilon_{k-j}^2}{\sum_{1\leqslant i\leqslant M} \varepsilon_{k-i}^2},\end{aligned}$$

for some K_2, since $\mathbf{u} \in U$. The Hölder inequality yields, for any $0 < \nu < \gamma$,

$$\mathrm{E}\left(\sup_{\mathbf{u}\in U}\frac{\sigma_k^2}{w_k(\mathbf{u})}\right)^{\nu} \leqslant K_2^{M\nu}\left(\mathrm{E}\left(\prod_{1\leqslant j\leqslant M}(1+\varepsilon_{k-j}^2)\right)^{\gamma}\right)^{\nu/\gamma}\left(\mathrm{E}\left(\sum_{1\leqslant i\leqslant M}\varepsilon_{k-i}^2\right)^{-\nu\gamma/(\gamma-\nu)}\right)^{(\gamma-\nu)/\gamma}.$$

By (1.4) and condition (5.1) we have

$$\mathrm{E}\left(\prod_{1\leqslant j\leqslant M}(1+\varepsilon_{k-j}^2)^{\gamma}\right) = \left(\mathrm{E}\left(1+\varepsilon_0^2\right)^{\gamma}\right)^M < \infty,$$

and therefore it is enough to prove that

$$\mathrm{E}\left(\sum_{1\leqslant i\leqslant M}\varepsilon_i^2\right)^{-\nu\gamma/(\gamma-\nu)} < \infty. \tag{5.5}$$

Condition (4.5) implies

$$\begin{aligned} P\left\{\left(\sum_{1\leqslant i\leqslant M}\varepsilon_i^2\right)^{-\nu\gamma/(\gamma-\nu)} > t\right\} &\leqslant P\left\{\sum_{1\leqslant i\leqslant M}\varepsilon_i^2 \leqslant t^{-(\gamma-\nu)/(\nu\gamma)}\right\} \\ &\leqslant \left(P\{\varepsilon_0^2 \leqslant t^{-(\gamma-\nu)/(\nu\gamma)}\}\right)^M \\ &\leqslant K_3 t^{-M\mu(\gamma-\nu)/(\nu\gamma)} \leqslant K_3 t^{-2} \end{aligned}$$

for all $t \geqslant 1$, for some constant K_3, if $M \geqslant 2\nu\gamma/(\mu(\gamma-\nu))$. Hence (5.5) is proved, completing the proof of Lemma 5.1. □

Lemma 5.2. *We assume that* (3.2) *holds and*

$$\mathrm{E}|y_0^2|^{\delta} < \infty, \qquad \textit{for some } \delta > 0. \tag{5.6}$$

Then for any $\nu > 0$,

$$\mathrm{E}\left(\sup_{\mathbf{u}\in U}\frac{\sum_{1\leqslant i<\infty} i^3 c_i(\mathbf{u}) y_{k-i}^2}{1+\sum_{1\leqslant i<\infty} c_i(\mathbf{u}) y_{k-i}^2}\right)^{\nu} < \infty. \tag{5.7}$$

Proof. For any $M \geqslant 1$, we have

$$\begin{aligned} \frac{\sum_{1\leqslant i<\infty} i^3 c_i(\mathbf{u}) y_{k-i}^2}{1+\sum_{1\leqslant i<\infty} c_i(\mathbf{u}) y_{k-i}^2} &\leqslant \frac{\sum_{1\leqslant i\leqslant M} i^3 c_i(\mathbf{u}) y_{k-i}^2}{\sum_{1\leqslant i\leqslant M} c_i(\mathbf{u}) y_{k-i}^2} + \sum_{M<i<\infty} i^3 c_i(\mathbf{u}) y_{k-i}^2 \\ &\leqslant M^3 + \sum_{M<i<\infty} i^3 c_i(\mathbf{u}) y_{k-i}^2. \end{aligned} \tag{5.8}$$

We pick constants $\rho_0^{1/q} < \rho_* < 1$, $\rho_{**} > 1$, such that $\rho_*\rho_{**} < 1$ and take $M \geqslant M_0(C_2, \rho_*, \rho_{**})$ (C_2 is from (3.4)) large enough. Then using (3.4), (5.6) and the Markov inequality, we obtain for all $t > 0$ that

$$
\begin{aligned}
P\Bigg\{\sup_{\mathbf{u}\in U}\sum_{M<i<\infty} i^3 c_i(\mathbf{u})y_{k-i}^2 > t\Bigg\} &\leqslant P\Bigg\{C_2\sum_{M<i<\infty} i^3\rho_0^{i/q}y_{k-i}^2 > t\Bigg\} \\
&\leqslant P\Bigg\{\sum_{M<i<\infty}\rho_*^i y_{k-i}^2 > t\Bigg\} \\
&\leqslant \sum_{M<i<\infty} P\Bigg\{y_{k-i}^2 > t\rho_*^{-i}\left(\frac{\rho_{**}}{\rho_{**}-1}\right)^{-1}\rho_{**}^{-i}\Bigg\} \\
&\leqslant \sum_{M<i<\infty} P\Bigg\{(y_0^2)^\delta \geqslant t^\delta\left(\frac{\rho_{**}}{\rho_{**}-1}\right)^{-\delta}(\rho_*\rho_{**})^{-i\delta}\Bigg\} \qquad (5.9)\\
&\leqslant t^{-\delta}\mathrm{E}|y_0^2|^\delta\left(\frac{\rho_{**}}{\rho_{**}-1}\right)^{\delta}\sum_{M<i<\infty}(\rho_*\rho_{**})^{i\delta} \\
&\leqslant \frac{\mathrm{E}|y_0^2|^\delta}{1-(\rho_*\rho_{**})^\delta}\left(\frac{\rho_{**}}{\rho_{**}-1}\right)^{\delta} t^{-\delta}(\rho_*\rho_{**})^{M\delta}.
\end{aligned}
$$

Let $t > 2\max(M_0^3, 1)$ and $M = (t/2)^{1/3}$. Putting together (5.8) and (5.9), we conclude that

$$
P\Bigg\{\sup_{\mathbf{u}\in U}\frac{\sum_{1\leqslant i<\infty} i^3 c_i(\mathbf{u})y_{k-i}^2}{1+\sum_{1\leqslant i<\infty} c_i(\mathbf{u})y_{k-i}^2} > t\Bigg\} \leqslant K_4 e^{-K_5 t^{1/3}}
$$

for constants K_4 and K_5, completing the proof of Lemma 5.2. □

Lemma 5.3. *We assume that* (3.2)*,* (4.4)*,* (4.5) *and* (5.6) *hold. Then*

$$
\mathrm{E}\sup_{\mathbf{u},\mathbf{v}\in U}\frac{1}{|\mathbf{u}-\mathbf{v}|}\left|\frac{y_k^2}{w_k(\mathbf{u})}-\frac{y_k^2}{w_k(\mathbf{v})}\right| < \infty \qquad (5.10)
$$

and

$$
\mathrm{E}\sup_{\mathbf{u},\mathbf{v}\in U}\frac{1}{|\mathbf{u}-\mathbf{v}|}\,|\log w_k(\mathbf{u})-\log w_k(\mathbf{v})| < \infty. \qquad (5.11)
$$

Proof. According to the mean value theorem there exists $\boldsymbol{\xi}\in U$ satisfying $|\boldsymbol{\xi}-\mathbf{u}|\leqslant|\mathbf{u}-\mathbf{v}|$, $|\boldsymbol{\xi}-\mathbf{v}|\leqslant|\mathbf{u}-\mathbf{v}|$ such that

$$
\left|\frac{y_k^2}{w_k(\mathbf{u})}-\frac{y_k^2}{w_k(\mathbf{v})}\right| = |\mathbf{u}-\mathbf{v}|\left|\frac{y_k^2}{w_k(\boldsymbol{\xi})}\right|\left|\frac{w_k'(\boldsymbol{\xi})}{w_k(\boldsymbol{\xi})}\right|. \qquad (5.12)
$$

By Lemma 5.1, (4.4) and the independence of ε_k and σ_k, we have that

$$\mathrm{E}\left(\sup_{\mathbf{u}\in U}\frac{y_k^2}{w_k(\mathbf{u})}\right)^{1+\delta/2}=\mathrm{E}(\varepsilon_k^2)^{1+\delta/2}\mathrm{E}\left(\sup_{\mathbf{u}\in U}\frac{\sigma_k^2}{w_k(\mathbf{u})}\right)^{1+\delta/2}<\infty. \tag{5.13}$$

Using Lemma 3.2, we conclude that

$$\sup_{\mathbf{u}\in U}\left|\frac{w_k'(\mathbf{u})}{w_k(\mathbf{u})}\right|\leqslant K_6\sup_{\mathbf{u}\in U}\frac{c_0(\mathbf{u})+\sum_{1\leqslant i<\infty}ic_i(\mathbf{u})y_{k-i}^2}{1+\sum_{1\leqslant i<\infty}c_i(\mathbf{u})y_{k-i}^2}, \tag{5.14}$$

and therefore Lemma 5.2 yields

$$\mathrm{E}\left(\sup_{\mathbf{u}\in U}\left|\frac{w_k'(\mathbf{u})}{w_k(\mathbf{u})}\right|\right)^{(2+\delta)/\delta}<\infty. \tag{5.15}$$

The Hölder inequality, (5.12), (5.13) and (5.15) imply (5.10).

Using the mean value theorem again, we obtain that

$$|\log w_k(\mathbf{u})-\log w_k(\mathbf{v})|\leqslant|\mathbf{u}-\mathbf{v}|\sup_{\boldsymbol{\xi}\in U}\left|\frac{w_k'(\boldsymbol{\xi})}{w_k(\boldsymbol{\xi})}\right|,\qquad \mathbf{u},\mathbf{v}\in U,$$

and therefore (5.15) yields (5.11). □

Lemma 5.4. *We assume that* (3.2)*,* (4.4) *and* (4.5) *hold. Then*

$$\sup_{\mathbf{u}\in U}\left|\frac{1}{n}L_n(\mathbf{u})-L(\mathbf{u})\right|\overset{\text{a.s.}}{\to}0,\qquad \text{as } n\to\infty,$$

where

$$L(\mathbf{u})=-\frac{1}{2}\mathrm{E}\left\{\log w_0(\mathbf{u})+\frac{y_0^2}{w_0(\mathbf{u})}\right\}.$$

Proof. From Lemma 2.3 we have that (5.6) is satisfied. Using Lemma 3.1, we obtain that

$$0<C_1\leqslant w_k(\mathbf{u})\leqslant C_2\left(1+\sum_{1\leqslant i<\infty}\rho_0^{i/q}y_{k-i}^2\right).$$

By Lemma 2.3, $\mathrm{E}(1+\sum_{1\leqslant i<\infty}\rho_0^{i/q}y_{k-i}^2)^{\delta^*}<\infty$, with the assumption that $\delta^*<1$. Hence

$$\mathrm{E}|\log w_0(\mathbf{u})|<\infty \tag{5.16}$$

and, by (4.4) and Lemma 5.1,

$$\mathrm{E}\frac{y_0^2}{w_0(\mathbf{u})}=\mathrm{E}\varepsilon_0^2\mathrm{E}\frac{\sigma_0^2}{w_0(\mathbf{u})}<\infty. \tag{5.17}$$

By (1.7) there is a function g such that $y_k=g(\varepsilon_k,\varepsilon_{k-1},\ldots)$ and therefore, by Theorem 3.5.8 in Stout (1974), y_k is stationary and ergodic. By (5.16) and (5.17) we can use the ergodic theorem and obtain that, for any $\mathbf{u}\in U$,

$$\frac{1}{n}L_n(\mathbf{u}) \to L(\mathbf{u}) \text{ a.s.}$$

Next we write

$$\sup_{\mathbf{u},\mathbf{v}\in U} |L_n(\mathbf{u}) - L_n(\mathbf{v})| \frac{1}{|\mathbf{u}-\mathbf{v}|} \leqslant \frac{1}{2}\sum_{1\leqslant k\leqslant n} \eta_k,$$

where

$$\eta_k = \sup_{\mathbf{u},\mathbf{v}\in U} \frac{1}{|\mathbf{u}-\mathbf{v}|}\left\{|\log w_k(\mathbf{u}) - \log w_k(\mathbf{v})| + \left|\frac{y_k^2}{w_k(\mathbf{u})} - \frac{y_k^2}{w_k(\mathbf{v})}\right|\right\}.$$

Again using Theorem 3.5.8 in Stout (1974), we conclude that η_k is a stationary and ergodic sequence, and by Lemma 5.3 we have that $\mathrm{E}\eta_0 < \infty$. So, using the ergodic theorem again, we obtain that

$$\frac{1}{n}\sum_{1\leqslant k\leqslant n} \eta_k \overset{\text{a.s.}}{=} O(1),$$

showing that

$$\sup_{\mathbf{u},\mathbf{v}\in U} \left|\frac{1}{n}L_n(\mathbf{u}) - \frac{1}{n}L_n(\mathbf{v})\right| \frac{1}{|\mathbf{u}-\mathbf{v}|} = O(1) \text{ a.s.} \qquad (n\to\infty).$$

Thus the sequence $L_n(\mathbf{u})/n$ is equicontinuous with probability one, it converges a.s. to $L(\mathbf{u})$ and U is compact; hence the uniform convergence of $L_n(\mathbf{u})/n$ follows. □

Lemma 5.5. *If the conditions of Theorem* 4.1 *are satisfied, then* $L(\mathbf{u})$, $\mathbf{u}\in U$ *has a unique maximum at* $\boldsymbol{\theta}$.

Proof. Since $w_0(\boldsymbol{\theta}) = \sigma_0^2$ and $\mathrm{E}(y_0^2/w_0(\mathbf{u})) = \mathrm{E}(\sigma_0^2/w_0(\mathbf{u}))$ by (5.17) and (4.6), we have

$$L(\boldsymbol{\theta}) - L(\mathbf{u}) = -\frac{1}{2} + \frac{1}{2}\mathrm{E}\left(\frac{\sigma_0^2}{w_0(\mathbf{u})} - \log\frac{\sigma_0^2}{w_0(\mathbf{u})}\right). \tag{5.18}$$

The function $x - \log x$ is positive for all $x > 0$ and reaches its smallest value at $x = 1$. Thus, $L(\mathbf{u})$, $\mathbf{u}\in U$, has an absolute maximum at $\boldsymbol{\theta}$.

Assume that $L(\mathbf{u}^*) = L(\boldsymbol{\theta})$ for some $\mathbf{u}^*\in U$. By (5.18) and the remark following it, we have $\sigma_0^2 = w_0(\mathbf{u}^*)$ a.s. Using Theorem 2.3, we obtain that $c_i(\boldsymbol{\theta}) = c_i(\mathbf{u}^*)$ for all $0\leqslant i<\infty$ and therefore, in view of (4.2), $\sigma_k^2 = w_k(\mathbf{u}^*)$. Thus, letting $\mathbf{u}^* = (x^*, s_1^*, \ldots, s_p^*, t_1^*, \ldots, t_q^*)$, we have

$$\sigma_k^2 = w_k(\boldsymbol{\theta}) = \omega + \alpha_1 y_{k-1}^2 + \ldots + \alpha_p y_{k-p}^2 + \beta_1\sigma_{k-1}^2 + \ldots + \beta_q\sigma_{k-q}^2 \tag{5.19}$$

and

$$\sigma_k^2 = w_k(\mathbf{u}^*) = x^* + s_1^* y_{k-1}^2 + \ldots + s_p^* y_{k-p}^2 + t_1^*\sigma_{k-1}^2 + \ldots + t_q^*\sigma_{k-q}^2 \tag{5.20}$$

for all $-\infty < k < \infty$. Using Corollary 2.1, we conclude that $\mathbf{u}^* = \boldsymbol{\theta}$.

Proof of Theorem 4.1. Clearly, U is a compact set. We have also shown that $L_n(\mathbf{u})/n$ converges uniformly to $L(\mathbf{u})$ on U with probability one (Lemma 5.4) and proved that $L(\mathbf{u})$ has a unique maximum at $\mathbf{u} = \boldsymbol{\theta}$ (Lemma 5.5). Therefore standard arguments show that the locations of the maxima of $L_n(\mathbf{u})/n$ converge a.s. to that of $L(\mathbf{u})$. □

Lemma 5.6. *If the conditions of Theorem* 4.2 *are satisfied, then*

$$\sup_{\mathbf{u}\in U}\left|\frac{1}{n}L_n'(\mathbf{u}) - L'(\mathbf{u})\right| \stackrel{\text{a.s.}}{\rightarrow} 0 \tag{5.21}$$

and

$$\sup_{\mathbf{u}\in U}\left|\frac{1}{n}L_n''(\mathbf{u}) - L''(\mathbf{u})\right| \stackrel{\text{a.s.}}{\rightarrow} 0, \tag{5.22}$$

as $n \to \infty$.

Proof. From (4.8) we obtain

$$\ell_k'(\mathbf{u}) = -\frac{1}{2}\left\{\frac{w_k'(\mathbf{u})}{w_k(\mathbf{u})} - \varepsilon_k^2\frac{\sigma_k^2}{w_k(\mathbf{u})}\frac{w_k'(\mathbf{u})}{w_k(\mathbf{u})}\right\} \tag{5.23}$$

and

$$\ell_k''(\mathbf{u}) = -\frac{1}{2}\left\{-\left(\frac{w_k'(\mathbf{u})}{w_k(\mathbf{u})}\right)^{\mathrm{T}}\frac{w_k'(\mathbf{u})}{w_k(\mathbf{u})} + \frac{w_k''(\mathbf{u})}{w_k(\mathbf{u})}\right.$$

$$\left. + 2\varepsilon_k^2\frac{\sigma_k^2}{w_k(\mathbf{u})}\left(\frac{w_k'(\mathbf{u})}{w_k(\mathbf{u})}\right)^{\mathrm{T}}\frac{w_k'(\mathbf{u})}{w_k(\mathbf{u})} - \varepsilon_k^2\frac{\sigma_k^2}{w_k(\mathbf{u})}\frac{w_k''(\mathbf{u})}{w_k(\mathbf{u})}\right\}. \tag{5.24}$$

Note that $\ell_k'(\mathbf{u})$ and $\ell_k''(\mathbf{u})$ are stationary and ergodic sequences. Using Lemmas 3.2, 3.3, 5.2 and relation (5.14), we obtain that

$$\mathrm{E}\sup_{\mathbf{u}\in U}\left|\frac{w_0'(\mathbf{u})}{w_0(\mathbf{u})}\right|^{\nu} < \infty$$

and

$$\mathrm{E}\sup_{\mathbf{u}\in U}\left|\frac{w_0''(\mathbf{u})}{w_0(\mathbf{u})}\right|^{\nu} < \infty,$$

for any $\nu > 0$. Therefore by Lemma 5.1 and equations (5.23) and (5.24) we obtain that $\mathrm{E}\sup_{\mathbf{u}\in U}|\ell_0'(\mathbf{u})| < \infty$ and $\mathrm{E}\sup_{\mathbf{u}\in U}|\ell_0''(\mathbf{u})| < \infty$. Thus an application of the ergodic theorem yields $L_n'(\mathbf{u})/n \to L'(\mathbf{u})$ a.s. and $L_n''(\mathbf{u})/n \to L''(\mathbf{u})$ a.s. for any $\mathbf{u} \in U$. The proof of the uniformity follows similar lines to the proof of Lemma 5.4. □

Lemma 5.7. *If the conditions of Theorem* 4.2 *are satisfied, then* $\mathbf{A}_0$ *is non-singular.*

Proof. The existence of $\mathrm{cov}(\ell_0'(\boldsymbol{\theta}))$ is an immediate consequence of (5.15) and condition (4.9). Let $\boldsymbol{\Psi}_k = w_k'(\boldsymbol{\theta})$. From the recursion formulae for $c_i(\mathbf{u})$ at the beginning of Section 3, we easily obtain

$$w_k(\mathbf{u}) = x + \sum_{1\leqslant i\leqslant p} s_i y_{k-i}^2 + \sum_{1\leqslant j\leqslant q} t_j w_{k-j}(\mathbf{u}). \tag{5.25}$$

Differentiating (5.25) with respect to its variables $\mathbf{u} = (x, s_1, \ldots, s_p, t_1, \ldots, s_q)$ we obtain that

$$\boldsymbol{\Psi}_k = \beta_1 \boldsymbol{\Psi}_{k-1} + \ldots + \beta_q \boldsymbol{\Psi}_{k-q} + \boldsymbol{\varphi}_k, \tag{5.26}$$

where $\boldsymbol{\varphi}_k = (\varphi_{k,0}, \ldots, \varphi_{k,p+q})$ is given by

$$\varphi_{k,0} = 1,$$

$$\varphi_{k,1} = y_{k-1}^2, \ldots, \varphi_{k,p} = y_{k-p}^2,$$

$$\varphi_{k,p+1} = \sigma_{k-1}^2, \ldots, \varphi_{k,p+q} = \sigma_{k-q}^2.$$

Next we show that there exists no real vector $\boldsymbol{\lambda} \in \mathbb{R}^{p+q+1}$, $\boldsymbol{\lambda} \neq \mathbf{0}$, and integer k such that

$$\boldsymbol{\lambda}\boldsymbol{\Psi}_k^{\mathrm{T}} = 0 \text{ with probability one.} \tag{5.27}$$

Let us assume that (5.27) holds for some $\boldsymbol{\lambda} \neq \mathbf{0}$ and for some integer k. Then by stationarity, (5.27) holds for all integers k. Hence (5.26) and (5.27) imply $\boldsymbol{\lambda}\boldsymbol{\varphi}_k^{\mathrm{T}} = 0$ for all k. However, this contradicts the minimality of (1.2), which follows from the assumptions of Theorem 4.2.

We have thus proved that (5.27) and consequently $\boldsymbol{\lambda}(w_0'(\boldsymbol{\theta})/w_0(\boldsymbol{\theta}))^{\mathrm{T}} = 0$ holds with probability one only if $\boldsymbol{\lambda} = \mathbf{0}$. Hence $\mathrm{cov}(w_0'(\boldsymbol{\theta})/w_0(\boldsymbol{\theta}))$ is non-singular. Observing that

$$\ell_0'(\boldsymbol{\theta}) = \frac{1}{2}(\varepsilon_0^2 - 1)\frac{w_0'(\boldsymbol{\theta})}{w_0(\boldsymbol{\theta})} \tag{5.28}$$

and since ε_0^2 and $w_0'(\boldsymbol{\theta})/w_0(\boldsymbol{\theta})$ are independent and $\mathrm{E}(\varepsilon_0^2 - 1)^2 \neq 0$ by (2.20), we obtain that $\mathrm{cov}(\ell_0'(\boldsymbol{\theta}))$ is also non-singular.

Proof of Theorem 4.2. By (4.1) and (4.2), $L_n(\mathbf{u})$ is twice continuously differentiable on U and reaches its maximum at $\hat{\boldsymbol{\theta}}_n$, which, for sufficiently large n, is an inner point of U by (3.2) and (4.7). Thus there is a random index n_0 such that

$$L_n'(\hat{\boldsymbol{\theta}}_n) = \mathbf{0}, \qquad \text{if } n \geqslant n_0. \tag{5.29}$$

Therefore for $n \geqslant n_0$ we have $L_n'(\hat{\boldsymbol{\theta}}_n) - L_n'(\boldsymbol{\theta}) = -L_n'(\boldsymbol{\theta})$, so by the mean value theorem for the coordinates of $L_n' = (L_{n,0}', \ldots, L_{n,p+q}')$ we have

$$(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta})L_{n,i}''(\boldsymbol{\xi}_{n,i}) = -L_{n,i}'(\boldsymbol{\theta}), \qquad 0 \leqslant i \leqslant p+q,$$

where $\boldsymbol{\xi}_{n,i}$ is between $\hat{\boldsymbol{\theta}}_n$ and $\boldsymbol{\theta}$. So, using Lemma 5.6 and the continuity of $L''(\mathbf{u})$, we obtain

$$(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta})(\mathbf{B}_0 + o(1)) = -\frac{1}{n}L_n'(\boldsymbol{\theta}) \text{ a.s.} \tag{5.30}$$

By (5.23) we have

$$\ell_k'(\boldsymbol{\theta}) = \frac{1}{2}(\varepsilon_k^2 - 1)\frac{w_k'(\boldsymbol{\theta})}{w_k(\boldsymbol{\theta})}. \tag{5.31}$$

In view of (4.6), $\mathrm{E}((\ell_k'(\boldsymbol{\theta}))^{\mathrm{T}}\ell_j'(\boldsymbol{\theta}))$, $j \neq k$, is the zero matrix, and thus we conclude that

$$\operatorname{cov}\left(n^{-1/2}L_n'(\boldsymbol{\theta})\right) = \mathbf{A}_0. \tag{5.32}$$

By Lemma 5.7, $\mathbf{A}_0$ is non-singular and therefore (5.30) implies the non-singularity of B_0.

If n is large, then $(\mathbf{B}_0 + o(1))^{-1}$ exists and equals $\mathbf{B}_0^{-1} + o(1)$. By (5.32) and the Markov inequality we have $L_n'(\boldsymbol{\theta})/n^{1/2} = O_P(1)$, and thus (4.11) follows from (5.30) and (5.31).

Since $\ell_k'(\boldsymbol{\theta}) = \frac{1}{2}(\varepsilon_k^2 - 1)f(\varepsilon_{k-1}^2, \varepsilon_{k-2}^2, \ldots)$ for some measurable function f, $\ell_k'(\boldsymbol{\theta})$ is a stationary martingale difference sequence. By Theorem 3.5.8 in Stout (1974) it is also ergodic. The covariance matrix of $\frac{1}{2}(\varepsilon_k^2 - 1)(w_k'(\boldsymbol{\theta})/w_k(\boldsymbol{\theta}))B_0^{-1}$ is $\mathbf{B}_0^{-1}\mathbf{A}_0\mathbf{B}_0^{-1}$, so the Cramér–Wold device (cf. Billingsley 1968, pp. 48–49) and Theorem 23.1 of Billingsley (1968, p. 206) shows that (4.11) implies (4.12). □

Lemma 5.8. *If the conditions of Theorem* 4.1 *are satisfied, then*

$$\sup_{\mathbf{u}\in U}\left|\sum_{1\leqslant k\leqslant n}(\log w_k(\mathbf{u}) - \log \tilde{w}_k(\mathbf{u}))\right| \stackrel{\text{a.s.}}{=} O(1) \tag{5.33}$$

and

$$\sup_{\mathbf{u}\in U}\left|\sum_{1\leqslant k\leqslant n}\left(\frac{w_k'(\mathbf{u})}{w_k(\mathbf{u})} - \frac{\tilde{w}_k'(\mathbf{u})}{\tilde{w}_k(\mathbf{u})}\right)\right| \stackrel{\text{a.s.}}{=} O(1), \tag{5.34}$$

as $n \to \infty$.

Proof. Since $w_k(\mathbf{u}) \geqslant C_1$ and $\tilde{w}_k(\mathbf{u}) \geqslant C_1$ for all $\mathbf{u} \in U$ and some positive constant C_1 (cf. Lemma 3.1), by the mean value theorem and (3.4) we have

$$\begin{aligned}|\log w_k(\mathbf{u}) - \log \tilde{w}_k(\mathbf{u})| &\leqslant \frac{1}{C_1}|w_k(\mathbf{u}) - \tilde{w}_k(\mathbf{u})| \\ &\leqslant \frac{C_2}{C_1}\sum_{k\leqslant i<\infty}\rho_0^{i/q}y_{k-i}^2 \\ &= \frac{C_2}{C_1}\rho_0^{k/q}\sum_{0\leqslant j<\infty}\rho_0^{j/q}y_{-j}^2.\end{aligned} \tag{5.35}$$

By Lemmas 2.2 and 2.3 the sum $\sum_{0\leqslant j<\infty}\rho_0^{j/q}y_{-j}^2$ is convergent with probability one. Hence

$$\begin{aligned}\sup_{\mathbf{u}\in U}\left|\sum_{1\leqslant k\leqslant n}(\log w_k(\mathbf{u}) - \log \tilde{w}_k(\mathbf{u}))\right| &\leqslant K_7\sum_{0\leqslant j<\infty}\rho_0^{j/q}y_{-j}^2\sum_{1\leqslant k\leqslant n}\rho_0^{k/q} \\ &\stackrel{\text{a.s.}}{=} O(1),\end{aligned}$$

proving (5.33). It is easy to see that

$$\left|\frac{w_k'(\mathbf{u})}{w_k(\mathbf{u})} - \frac{\tilde{w}_k'(\mathbf{u})}{\tilde{w}_k(\mathbf{u})}\right| \leqslant \frac{1}{C_1} \sup_{\mathbf{u}\in U} |w_k'(\mathbf{u}) - \tilde{w}_k'(\mathbf{u})| + \frac{1}{C_1} \sup_{\mathbf{u}\in U} \left|\frac{w_k'(\mathbf{u})}{w_k(\mathbf{u})}\right| \sup_{\mathbf{u}\in U} |w_k(\mathbf{u}) - \tilde{w}_k(\mathbf{u})|.$$

By (5.35) we obtain

$$\sum_{1\leqslant k\leqslant n} \sup_{\mathbf{u}\in U} \left|\frac{w_k'(\mathbf{u})}{w_k(\mathbf{u})}\right| \sup_{\mathbf{u}\in U} |w_k(\mathbf{u}) - \tilde{w}_k(\mathbf{u})| \leqslant K_8 \sum_{0\leqslant j<\infty} \rho_0^{j/q} y_{-j}^2 \sum_{1\leqslant k<\infty} \sup_{\mathbf{u}\in U} \left|\frac{w_k'(\mathbf{u})}{w_k(\mathbf{u})}\right| \rho_0^{k/q}$$
$$< \infty \text{ a.s.}$$

for some constant K_8; in the last step we used Lemma 2.2 and (5.15). By Lemmas 3.1 and 3.2 there exist K_9 and $0 < \tilde{\rho} < 1$ such that $|c_i'(\mathbf{u})| \leqslant K_9 \tilde{\rho}^i$, $\mathbf{u} \in U$, for all $0 \leqslant i < \infty$. Hence

$$\sum_{1\leqslant k\leqslant n} \sup_{\mathbf{u}\in U} |w_k'(\mathbf{u}) - \tilde{w}_k'(\mathbf{u})| \leqslant K_9 \sum_{1\leqslant k\leqslant n} \sum_{k\leqslant i<\infty} \tilde{\rho}^i y_{k-i}^2$$
$$\leqslant K_9 \sum_{0\leqslant j<\infty} \rho^j y_{-j}^2 \sum_{1\leqslant k<\infty} \tilde{\rho}^k,$$

completing the proof of (5.34). □

Lemma 5.9. *If the conditions of Theorem* 4.1 *are satisfied, then*

$$\sup_{\mathbf{u}\in U} \left| \sum_{1\leqslant k\leqslant n} \left(\frac{y_k^2}{w_k(\mathbf{u})} - \frac{y_k^2}{\tilde{w}_k(\mathbf{u})} \right) \right| \overset{\text{a.s.}}{=} O(1) \tag{5.36}$$

and

$$\sup_{\mathbf{u}\in U} \left| \sum_{1\leqslant k\leqslant n} y_k^2 \left(\frac{w_k'(\mathbf{u})}{w_k^2(\mathbf{u})} - \frac{\tilde{w}_k'(\mathbf{u})}{\tilde{w}_k^2(\mathbf{u})} \right) \right| \overset{\text{a.s.}}{=} O(1), \tag{5.37}$$

as $n \to \infty$.

Proof. By (5.35) we have

$$\sup_{\mathbf{u}\in U} \sum_{1\leqslant k\leqslant n} \frac{y_k^2}{w_k(\mathbf{u})} \left| \frac{w_k(\mathbf{u}) - \tilde{w}_k(\mathbf{u})}{\tilde{w}_k(\mathbf{u})} \right| \leqslant \frac{C_2}{C_1} \sum_{0\leqslant j<\infty} \rho_0^{j/q} y_{-j}^2 \sum_{1\leqslant k<\infty} \sup_{\mathbf{u}\in U} \frac{y_k^2}{w_k(\mathbf{u})} \rho_0^{k/q}.$$

We note that $\sup\{y_k^2/w_k(\mathbf{u}), \mathbf{u} \in U\}$ is a stationary sequence and, by (4.4) and Lemma 5.1,

$$\mathrm{E}\left(\sup_{\mathbf{u}\in U} \frac{y_0^2}{w_0(\mathbf{u})} \right)^{\nu} = \mathrm{E}(\varepsilon_0^2)^{\nu} \mathrm{E}\left(\sup_{\mathbf{u}\in U} \frac{\sigma_0^2}{w_0(\mathbf{u})} \right)^{\nu} < \infty \tag{5.38}$$

for all $\nu < 1 + \delta$. Thus an application of Lemma 2.2 yields that

$$\sum_{1\leqslant k<\infty} \sup_{\mathbf{u}\in U} \frac{y_k^2}{w_k(\mathbf{u})}\rho_0^{k/q} < \infty \text{ a.s.,}$$

completing the proof of (5.36). The proof of (5.37) is similar to that of (5.34) and (5.36), and therefore the details are omitted. □

Proof of Theorem 4.3. Lemmas 5.8 and 5.9 imply that

$$\sup_{\mathbf{u}\in U}\left|\frac{1}{n}L_n(\mathbf{u}) - \frac{1}{n}\tilde{L}_n(\mathbf{u})\right| \overset{\text{a.s.}}{=} o(1),$$

and therefore, by Lemma 5.4,

$$\sup_{\mathbf{u}\in U}\left|\frac{1}{n}\tilde{L}_n(\mathbf{u}) - L(\mathbf{u})\right| \overset{\text{a.s.}}{=} o(1).$$

Thus $\tilde{\boldsymbol{\theta}}_n \to \boldsymbol{\theta}$ a.s. follows as in the proof of Theorem 4.1. □

Proof of Theorem 4.4. Observing that $\sup_{\mathbf{u}\in U}|L_n'(\mathbf{u}) - \tilde{L}_n'(\mathbf{u})|$ is bounded by the sum of the left-hand sides of (5.34) and (5.37), and using Lemmas 5.8 and 5.9, we obtain under the conditions of Theorem 4.4 that

$$\sup_{\mathbf{u}\in U}\left|\frac{1}{n}L_n'(\mathbf{u}) - \frac{1}{n}\tilde{L}_n'(\mathbf{u})\right| \overset{\text{a.s.}}{=} O\left(\frac{1}{n}\right). \tag{5.39}$$

Similarly to the proof of Theorem 4.2, there is a random variable n_0 such that

$$\tilde{L}_n'(\tilde{\boldsymbol{\theta}}_n) = \mathbf{0}, \qquad \text{if } n \geqslant n_0. \tag{5.40}$$

Using (5.29), (5.39) and (5.40), we obtain that

$$\frac{1}{n}L_n'(\hat{\boldsymbol{\theta}}_n) - \frac{1}{n}L_n'(\tilde{\boldsymbol{\theta}}_n) = \frac{1}{n}L_n'(\hat{\boldsymbol{\theta}}_n) - \frac{1}{n}\tilde{L}_n'(\tilde{\boldsymbol{\theta}}_n) + O\left(\frac{1}{n}\right) = O\left(\frac{1}{n}\right) \text{ a.s.} \tag{5.41}$$

By (5.22) and a coordinatewise application of the mean value theorem, we obtain that

$$\frac{1}{n}L_n'(\hat{\boldsymbol{\theta}}_n) - \frac{1}{n}L_n'(\tilde{\boldsymbol{\theta}}_n) = (\hat{\boldsymbol{\theta}}_n - \tilde{\boldsymbol{\theta}}_n)L''(\boldsymbol{\theta})(1 + o(1)) \text{ a.s.,}$$

and therefore (5.41) implies

$$|\hat{\boldsymbol{\theta}}_n - \tilde{\boldsymbol{\theta}}_n| = O\left(\frac{1}{n}\right) \text{ a.s.}$$

Hence Theorem 4.4 is an immediate consequence of Theorem 4.2. □

Proof of Remark 4.5. By (5.23) we have $\ell_0'(\boldsymbol{\theta}) = -\frac{1}{2}(1 - \varepsilon_0^2)w_0'(\boldsymbol{\theta})/\sigma_0^2$ and therefore

$$\mathrm{E}(\ell_0'(\boldsymbol{\theta})^{\mathrm{T}}\ell_0'(\boldsymbol{\theta})) = \frac{1}{4}\mathrm{E}(1 - \varepsilon_0^2)^2\mathrm{E}\left(\frac{w_0'(\boldsymbol{\theta})}{\sigma_0^2}\right)^{\mathrm{T}}\left(\frac{w_0'(\boldsymbol{\theta})}{\sigma_0^2}\right). \tag{5.42}$$

Similarly, on account of (5.24) and $\mathrm{E}\varepsilon_0^2 = 1$, we have

$$\mathrm{E}\ell_0''(\boldsymbol{\theta}) = -\frac{1}{2}\left\{-\mathrm{E}\left(\frac{w_0'(\boldsymbol{\theta})}{\sigma_0^2}\right)^{\mathrm{T}}\left(\frac{w_0'(\boldsymbol{\theta})}{\sigma_0^2}\right) + \mathrm{E}\left(\frac{w_0''(\boldsymbol{\theta})}{\sigma_0^2}\right) + 2\mathrm{E}\left(\frac{w_0'(\boldsymbol{\theta})}{\sigma_0^2}\right)^{\mathrm{T}}\left(\frac{w_0'(\boldsymbol{\theta})}{\sigma_0^2}\right) - \mathrm{E}\left(\frac{w_0''(\boldsymbol{\theta})}{\sigma_0^2}\right)\right\}$$

$$= -\frac{1}{2}\mathrm{E}\left(\frac{w_0'(\boldsymbol{\theta})}{\sigma_0^2}\right)^{\mathrm{T}}\left(\frac{w_0'(\boldsymbol{\theta})}{\sigma_0^2}\right). \tag{5.43}$$

Comparing (5.42) and (5.43), we obtain Remark 4.5. □

Acknowledgement

István Berkes was supported by grant T 29621 from the Hungarian National Foundation for Scientific Research. Lajos Horváth and Piotr Kokoszka were supported by NATO grant PST.CLG.977607. LH received additional support from DOMUS Hungarica Scientiarum et Artium. We would like to thank two anonymous referees for studying the first version of this paper in great detail. Their extensive in-depth comments led to substantial improvements both in the mathematical theory and in the presentation of results.

References

Basrak, B., Davis, R.A. and Mikosch, T. (2001) Sample ACF of multivariate stochastic recurrence equations with applications to GARCH. Preprint.

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

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

Bougerol, P. and Picard, N. (1992a) Strict stationarity of generalized autoregressive processes. *Ann. Probab.*, **20**, 1714–1730.

Bougerol, P. and Picard, N. (1992b) Stationarity of GARCH processes and of some nonnegative time series. *J. Econometrics*, **52**, 115–127.

Comte, F. and Lieberman, O. (2000) Asymptotic theory for multivariate GARCH processes. Preprint.

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

Gouriéroux, C. (1997) *ARCH Models and Financial Applications.* Berlin: Springer-Verlag.

Jeantheau, T. (1998) Strong consistency of estimators for multivariate ARCH models. *Econometric Theory*, **14**, 70–86.

Kingman, J.F.C. (1973) Subadditive ergodic theory. *Ann. Probab.*, **1**, 883–909.

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

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 GARCH(1, 1) model. *Econometric Theory*, **6**, 318–334.

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

Received January 2001 and revised May 2002