Fields Institute Communications
Volume **00**, 0000

Probabilistic and statistical properties of GARCH processes

István Berkes
A. Rényi Institute of Mathematics, Hungarian Academy of Sciences
P.O. Box 127, H-1364 Budapest, Hungary
berkes@renyi.hu

Lajos Horváth
Department of Mathematics, University of Utah
155 South 1440 East, Salt Lake City, UT 84112-0090, USA
horvath@math.utah.edu

Piotr Kokoszka
Department of Mathematics and Statistics
Utah State University, 3900 Old Main Hill
Logan, UT 84322-3900, USA
piotr@math.usu.edu

Dedicated to Miklós Csörgő on the occasion of his 70th birthday

Abstract. GARCH type models are extensively used in modeling returns on speculative assets. We review recent advances in the theory of GARCH(p,q) models. First we focus on the existence and the uniqueness of a solution to GARCH(p,q) equations and on a convenient representation of the solution in terms of the past observations. We then consider the asymptotic theory for the partial sum and the empirical process. After the review of the estimation methods, with special emphasis on the pseudo maximum likelihood estimation, we describe the asymptotic theory for the sequential empirical process and the autocorrelation function of the squared residuals. We conclude with a discussion of long memory ARCH processes.

1 Probabilistic properties

Analysis of financial data has received considerable attention in the literature during the last 20 years. Several models have been suggested to capture 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 well known and most often used examples is the autoregressive

2000 *Mathematics Subject Classification.* Primary 60G10, 60F17; Secondary 62M10, 62G20.
István Berkes was supported by the Hungarian National Foundation for Scientific Research, Grants T 29621, T 37886 and by NSF grant INT-0223262.
Lajos Horváth was supported by NATO grant PST.CLG.977607 and by NSF grant INT-0223262.
Piotr Kokoszka was supported by NATO grant PST.CLG.977607, USU New Faculty grant and by NSF grant INT-0223262 .

conditionally heteroscedastic (ARCH) process introduced by Engle [27]. The ARCH model has been investigated and generalized by several authors, cf. for example Bollerslev [14] and Gouriéroux [38]. In this paper we survey some of the basic results on the probabilistic as well as statistical properties of GARCH(p,q) processes. We say that $y_k, -\infty < k < \infty$ is a GARCH(p,q) sequence if it satisfies the following equations:

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

and

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

where

$$\omega > 0, \quad \alpha_i \ge 0, \quad 1 \le i \le p, \quad \beta_j \ge 0, \quad 1 \le j \le q \tag{1.3}$$

are constants. If $\beta_1 = \ldots = \beta_q = 0$, the sequence is called ARCH(p).

The first question is the existence of a process satisfying (1.1)– (1.3).

Theorem 1.1 *If $p = q = 1$,*

$$\epsilon_i, \quad -\infty < i < \infty \quad \textit{is a stationary and ergodic sequence,} \tag{1.4}$$

$$E|\log(\beta_1 + \alpha_1 \epsilon_0^2)| < \infty \tag{1.5}$$

and (1.3) holds, then (1.1) and (1.2) have a unique solution if

$$E \log(\beta_1 + \alpha_1 \epsilon_0^2) < 0. \tag{1.6}$$

Moreover, the solution is stationary and ergodic.

Proof Using the recursions (1.1) and (1.2), Nelson [56] showed that if a solution exists it must be given by

$$\sigma_\ell^2 = \omega \left[1 + \sum_{k=1}^{\infty} \prod_{i=1}^{k} (\beta_1 + \alpha_1 \epsilon_{\ell-i}^2) \right]. \tag{1.7}$$

The existence of σ_ℓ^2 follows from the ergodic theorem (cf. Breiman [19], p. 118). The stationarity and ergodicity are immediate consequences of Theorem 3.5.8 of Stout [61]. □

Theorem 1.2 *If $p = q = 1$, (1.3) and (1.5) are satisfied and*

$$\epsilon_i, \quad -\infty < i < \infty \quad \textit{are independent and identically distributed random variables,} \tag{1.8}$$

then (1.1) and (1.2) have a unique solution if and only if (1.6) holds.

Proof The result was obtained by Nelson [56]. □

Theorem 1.3 *If $p = q = 1$, (1.1)-(1.3), (1.5), (1.6) , (1.8) hold and there is $\nu_* > 0$ such that $E(\beta_1 + \alpha_1 \epsilon_0^2)^{\nu_*} = 1$ and $E(\beta_1 + \alpha_1 \epsilon_0^2)^t < 1$ for all $0 < t < \nu_*$, then*

$$E(\sigma_0^2)^\nu \begin{cases} < \infty, & \textit{if } \nu < \nu_* \\ = \infty, & \textit{if } \nu > \nu_*. \end{cases}$$

Proof The characterization of the number of moments of σ_0^2 is also due to Nelson [56] . □

For the estimation of ν_* we refer to Berkes, Horváth and Kokoszka [10].

We have to introduce further notation to obtain similar results for the general GARCH(p,q). Let

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

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

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 that $\min(p,q) \geq 2$.) 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

$$\gamma_L = \inf_{0 \leq n < \infty} \frac{1}{n+1} E \log \| A_0 A_1 \ldots A_n \|,$$

assuming that

$$E(\log \| A_0 \|) < \infty. \tag{1.9}$$

(We note that $\| A_0 (1, 0, \ldots,)^T \| \geq 1$ and therefore $\| A_0 \| \geq 1$.)

Theorem 1.4 *If (1.3), (1.4) and (1.9) hold, then (1.1) and (1.2) have a unique solution if*

$$\gamma_L < 0. \tag{1.10}$$

Moreover, the solution is stationary and ergodic.

Proof Following Bougerol and Picard [15], [16], we introduce $\mathbf{X}_n = (\sigma_n^2, \ldots, \sigma_{n-q-1}^2, y_{n-1}^2, \ldots, y_{n-p-1}^2)^T \in \mathbb{R}^{p+q-1}$ and $\mathbf{D} = (\omega, 0, \ldots, 0)^T \in \mathbb{R}^{p+q-1}$. Equations (1.1) and (1.2) can be written equivalently as

$$\mathbf{X}_{n+1} = A_n \mathbf{X}_n + \mathbf{D}. \tag{1.11}$$

Brandt [18] and Bougerol and Picard [15], [16] showed that if (1.11) holds, then the solution must be given by

$$\mathbf{X}_n = \mathbf{D} + \sum_{0 \leq k < \infty} A_n \cdots A_{n-k} \mathbf{D}. \tag{1.12}$$

By the subadditive ergodic theorem (cf. Kingman [46]) the sum on the right hand side of (1.12) exists a.s. if (1.10) holds. The stationarity and ergodicity are immediate consequences of Theorem 3.5.8 in Stout [61]. □

Theorem 1.5 *If (1.3), (1.8) and (1.9) hold, then (1.1) and (1.2) have a unique solution if and only if (1.10) holds.*

Proof The proof can be found in Bougerol and Picard [15]. □

Remark 1.6 If the conditions of Theorem 1.5 are satisfied and $E\epsilon_0^2 = 1$, then $\alpha_1 + \ldots + \alpha_p + \beta_1 + \ldots + \beta_q < 1$ is the necessary and sufficient condition for the existence of Ey_0^2 (cf. Bollerslev [14]). Ling [51] and Ling and McAleer [53] obtained necessary and sufficient conditions for the existence of the moments of σ_0^2 in general GARCH(p,q) models.

Remark 1.7 The extension of Theorems 1.4 and 1.5 to the characterization of ARCH(∞) with slowly decaying coefficients were provided by Kazakevičius and Leipus [44], [45]. It is interesting that even in the case when the Lyapunov exponent is zero, one can have stationary solution. This is a case when the coefficients are hyperbolically decaying and satisfy the condition

$$E\epsilon_0^2 \sum_{1 \le j < \infty} < 1.$$

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

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

and

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

Here and in the sequel we do not need $\min(p,q) \ge 2$, and the possible inclusion of extra terms in (1.2); this was needed only for computing the Lyapunov exponent γ_L.

Lemma 1.8 *If (1.10) holds, then*

$$\beta_1 + \beta_2 + \ldots + \beta_q < 1. \tag{1.13}$$

Relation (1.13) is equivalent to

$$\textit{all roots of } \mathcal{B}(x) = 0 \textit{ are outside of the unit disk.} \tag{1.14}$$

Proof The first statement is due to Bougerol and Picard [16]. The second part was obtained by Berkes, Horváth and Kokoszka [12].

□

The next technical lemma is used to show that σ_k^2 is a linear function of $y_i^2, i < k$. For the proof we refer to Berkes, Horváth and Kokoszka [12]. Let $\log^+ x = \log x$, if $x > 1$ and 0 otherwise.

Lemma 1.9 *If $\{\xi_k, 0 \le k < \infty\}$ is a sequence of identically distributed random variables satisfying*

$$E \log^+ |\xi_0| < \infty, \tag{1.15}$$

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

We now establish a representation for σ_k^2 in terms of the y_{k-i}^2, $i \ge 1$. Since by Lemma 1.8 $\mathcal{B}(x)$ has all roots outside the unit disk, we have

$$\sum_{j=0}^{\infty} d_j x^j = \frac{1}{\mathcal{B}(x)}, \qquad |x| \le 1 \tag{1.16}$$

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

$$c_0 = \omega \sum_{0 \le m < \infty} d_m \tag{1.17}$$

and

$$c_j = \alpha_1 d_{j-1} + \alpha_2 d_{j-2} + \cdots + \alpha_p d_{j-p}, \;\; 1 \le j < \infty. \tag{1.18}$$

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

$$\frac{\mathcal{A}(x)}{\mathcal{B}(x)} = \sum_{1 \le i < \infty} c_i x^i, \qquad |x| \le 1 \tag{1.19}$$

and by (1.18) the coefficients $c_1, c_2, \ldots$ decay exponentially fast.

Theorem 1.10 *If (1.1)-(1.4), (1.9), (1.10) are satisfied and*

$$E \log \sigma_0^2 < \infty, \tag{1.20}$$

then

$$\sigma_k^2 = c_0 + \sum_{1 \le i < \infty} c_i y_{k-i}^2 \quad \textit{for all } k \tag{1.21}$$

with probability one.

Proof Since $||A_0|| \ge ||\,\boldsymbol{\xi}_0|| = \epsilon_0^2$, (1.9) yields that $E \log^+ \epsilon_0^2 \le E \log^+ ||A_0|| < \infty$. Since by (1.1) we have $E \log^+ y_0^2 \le E \log^+ \sigma_0^2 + E \log^+ \epsilon_0^2$, from (1.20) we get that

$$E \log^+ y_0^2 < \infty. \tag{1.22}$$

Since the sequence $c_1, c_2, \ldots$ decays exponentially fast, Lemma 1.9 yields that the series in (1.21) is absolutely convergent with probability one. Using the backward shift operator $\mathbf{S}$, we can write (1.2) succinctly as

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

By (1.14) we have that $1/\mathcal{B}(\mathbf{S})$ exists and therefore

$$\sigma_k^2 = \frac{\mathcal{A}(\mathbf{S})}{\mathcal{B}(\mathbf{S})} \left(\frac{\omega}{\mathcal{A}(1)} + y_k^2 \right). \tag{1.23}$$

Now by (1.19) we have (1.21). □

Next we consider the uniqueness of the representation in (1.21). Let $\mathcal{F}_i$ be the σ-algebra generated by $\epsilon_j, -\infty < j \le i$.

Theorem 1.11 *If (1.1)-(1.4). (1.9), (1.10), (1.20) are satisfied and*

$$\textit{there is } 0 < C \le \infty \textit{ such that } P\{E(\min(\epsilon_0^2, C)|\mathcal{F}_{-1}) \ne \min(\epsilon_0^2, C)\} > 0, \tag{1.24}$$

then the representation in (1.21) is unique.

Proof Let assume that there is an other representation for σ_0^2 :

$$\sigma_0^2 = c_0^* + \sum_{1 \le i < \infty} c_i^* y_{-i}^2.$$

If $c_i = c_i^*$ for all $1 \le i < \infty$ then $c_0 = c_0^*$. Let i^* be the smallest positive integer satisfying $c_{i^*} \ne c_{i^*}^*$. Then

$$\epsilon_{-i^*}^2 = \left\{ c_0 - c_0^* + \sum_{i^* < i < \infty} (c_i - c_i^*) y_{-i}^2 \right\} \Big/ \left\{ (c_{i^*}^* - c_{i^*}) \sigma_{-i^*}^2 \right\}. \tag{1.25}$$

By (1.2) and (1.3) we have $\sigma_{-i^*}^2 > 0$ and thus the random variable on the right-hand side of (1.25) is well-defined and therefore ϵ_{-i^*} is measurable with respect to $\mathcal{F}_{-i^*-1}$. So by the stationarity of ϵ_i we conclude that ϵ_0 is measurable with respect to $\mathcal{F}_{-1}$, conradicing (1.24). □

Next we provide two special cases when (1.24) holds.

Corollary 1.12 *We assume that (1.1)-(1.4). (1.9), (1.10), (1.20) are satisfied and*

$$\epsilon_0^2 \ \textit{is a non-degenerate random variable.} \tag{1.26}$$

(i) If (1.8) is also satisfied, then (1.24) holds.
(ii) If $E\epsilon_0^2 < \infty$ and $\epsilon_i^2 - E\epsilon_0^2$ is a martingale difference sequence with respect to $\mathcal{F}_i$, then (1.24) holds.

Proof (i) Since ϵ_0^2 and $\mathcal{F}_{-1}$ are independent, we get that

$$E(\min(\epsilon_0^2, C)|\mathcal{F}_{-1}) = E(\min(\epsilon_0^2, C)) \ \text{ for all } \ C.$$

By (1.26) we have that

$$P\{E(\min(\epsilon_0^2, C)) \neq \min(\epsilon_0^2, C)\} > 0 \ \text{ for some } \quad C,$$

establishing (1.24).
(ii) If $\epsilon_i^2 - E\epsilon_0^2$ is a martingale difference sequence, then by the stationarity $E(\epsilon_0^2|\mathcal{F}_{-1}) = E\epsilon_0^2$. Using (1.26) we get that (1.24) holds with $C = \infty$. □

Under the conditions of Theorem 1.11 the coefficients c_i in (1.21) are unique. The next result shows that under mild conditions these coefficients are uniquely determined by $\boldsymbol{\theta}$. First we need the notion of minimal GARCH. We say that the definition (1.1)-(1.2) is *minimal* if that there is no pair (p^*, q^*) such that $p^* < p$ or $q^* < q$ and

$$\sigma_k^2 = \omega^* + \sum_{1 \le i \le p^*} \alpha_i^* y_{k-i}^2 + \sum_{1 \le j \le q^*} \beta_j^* \sigma_{k-j}^2 \tag{1.27}$$

for some (not necessarily non–negative) $\omega^*, \alpha_i^*, 1 \le i \le p^*$ and $\beta_j^*, 1 \le j \le q^*$.

Theorem 1.13 *If (1.1)-(1.4), (1.9), (1.10) (1.20) and (1.24) hold, then the definition (1.1)-(1.2) is minimal if and only if*

$$\textit{the polynomials } \mathcal{A}(x) \textit{ and } \mathcal{B}(x) \textit{ are coprimes in the set of polynomials with real coefficients.} \tag{1.28}$$

Proof Let us assume that (1.28) holds and assume indirectly that there are (p^*, q^*), $p^* < p$ or $q^* < q$, and $\omega^*, \alpha_i^*, 1 \le i \le p^*, \beta_j^*, 1 \le j \le q^*$ such that (1.27) holds. Let $\mathcal{A}^*(x) = x\alpha_1^* + \cdots + x^{p^*}\alpha_{p^*}^*$ and $\mathcal{B}^*(x) = 1 - (\beta_1^* x + \cdots + \beta_{q^*}^* x^{q^*})$. By Theorem 1.11 the coefficients c_i must satisfy (1.19). Similarly to (1.19) we have

$$\sum_{1 \le i < \infty} c_i x^i = \frac{\mathcal{A}^*(x)}{\mathcal{B}^*(x)}. \tag{1.29}$$

Consequently, we have that $\mathcal{A}(x)/\mathcal{B}(x) = \mathcal{A}^*(x)/\mathcal{B}^*(x)$. By (1.28) $\mathcal{A}(x)$ 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 $p^* \ge p, q^* \ge q$, a contradiction.

Let us assume conversely that the definition (1.1)-(1.2) is minimal but (1.28) 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 $\mathcal{B}(x) = \mathcal{B}^*(x)\mathcal{P}(x)$. Now we show that with $\tilde{\omega} = \omega\mathcal{B}^*(1)/\mathcal{B}(1)$, we have

$$\sigma_k^2 = \tilde{\omega} + \sum_{1 \le i \le p^*} \alpha_i^* y_{k-i}^2 + \sum_{1 \le j \le q^*} \beta_j^* \sigma_{k-j}^2. \tag{1.30}$$

Indeed, by (1.23) and (1.29) we get 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 $\mathcal{B}$, (1.30) contradicts the minimality assumption. □

2 Approximations for the empirical processes and partial sums of y_i

The empirical process of the sequence y_i is defined by

$$R(s,t) = \sum_{1 \le k \le t} Y_k(s),$$

where

$$Y_k(s) = I\{y_k \le s\} - P\{y_k \le s\}.$$

The distribution function of ϵ_0 is denoted by H.

Theorem 2.1 *We assume that (1.1)-(1.3), (1.8), (1.10) hold,*

$$|H(t) - H(s)| \le C|t-s|^\theta \textit{ with some } 0 < C < \infty \textit{ and } 0 < \theta \le 1$$

and

$$E(\log \| A_0 \|)^\mu < \infty \textit{ with some } \mu > 2 + 16/\theta. \tag{2.1}$$

Then the series

$$\Delta(s,s') = EY_0(s)Y_0(s') + \sum_{1 \le n < \infty} EY_0(s)Y_n(s') + \sum_{1 \le n < \infty} EY_0(s')Y_n(s)$$

is absolutely convergent for any $-\infty < s, s' < \infty$ and there is a Gaussian process $K^*(s,t)$ with $EK^*(s,t) = 0$, $EK^*(s,t)K^*(s',t') = \min(t,t')\Delta(s,s')$, such that

$$\sup_{0 \le t \le T} \sup_{-\infty < s < \infty} |R(s,t) - K^*(s,t)| = o(T^{1/2}(\log T)^{-\lambda}) \qquad a.s.$$

with some $\lambda > 0$.

Proof The approximation was obtained by Berkes and Horváth [3], [6]. □

Remark 2.2 If $E(\log^+ |\epsilon_0|)^\mu < \infty$ with some $\mu > 2 + 16/\theta$, then condition (2.1) holds in any GARCH(p,q) model.

Remark 2.3 Basrak, Davis and Mikosch [2] stated that if (1.10) holds, $E|\epsilon_0|^\mu < \infty$ with some $\mu > 0$, ϵ_0 has a positive density on an interval containing 0, then $\{y_n^2, -\infty < n < \infty\}$ is β-mixing with a geometric rate. For the proof we refer to Boussama [17]. Combining the mixing property of $\{y_n^2\}$ with the main result in Philipp and Pinzur [59] and Philipp [58] one could get strong approximations for $\sum_{1 \le k \le t}(I\{y_k^2 \le s\} - P\{y_k^2 \le s\})$. However, the method used by Basrak et al. [2] does not give a similar result for $R(s,t)$. Also, to get the β-mixing of $\{y_n^2\}$, Basrak et al. [2] assume that $E|\epsilon_0|^\mu < \infty$ with some $\mu > 0$, which is stronger than (2.1). For further results on mixing properties of GARCH(p,q) and related sequences we refer to Carrasco and Chen [20] who also assumed that ϵ_0 has a smooth density. Most GARCH sequences can be viewd as generalized hidden Markov models. Hence results on mixing properties of Markov processes can be utilized (cf. Davydov [21] and Meyn and Tweedie [55]).

Remark 2.4 In the proof of Theorem 2.1, Berkes and Horváth [3] used a totally different approach and established a new structural property of GARCH sequences which is easier to verify than strong mixing and is much more convenient in applications. Indeed, they proved that if (1.1)-(1.3), (1.8), (1.10) and (2.1) hold then there is a sequence $\{y_n'\}$ which is close to $\{y_n\}$ (in the sense that $y_n - y_n' \to 0$ rapidly) and for each $n \geq 1$ the finite sequence $\{y_1', \dots, y_n'\}$ is not only mixing, but in fact is m-dependent with $m = n^\rho$ for some $0 < \rho < 1$, i.e. the terms of this sequence with indices differing at least by n^ρ are independent. This property does not imply strong mixing for $\{y_n\}$ but it is more convenient to use. In fact, it permits us to deduce asymptotic properties of $\{y_n\}$ directly from known results for independent random variables via standard blocking techniques. We note also that this approach requires weaker moment conditions on $\{\epsilon_n\}$ than the strong mixing technique. It seems likely that the same method will be applicable in many other situations.

Let $N(a,b)$ be a normal random variable with parameters a and b.

Theorem 2.5 *If (1.1)-(1.3), (1.8), (1.10) hold, $E\epsilon_0^4 < \infty$ and*

$$(E\epsilon_0^4)^{1/2} \sum_{1 \leq j < \infty} c_j < 1, \tag{2.2}$$

then

$$n^{-1/2} \sum_{1 \leq i \leq n} (y_i^2 - Ey_0^2) \ \xrightarrow{\mathcal{D}}\ N\left(0, \sum_{-\infty < i < \infty} \operatorname{cov}(y_i^2, y_0^2)\right).$$

Proof This result is a special case of Theorem 5.1 of Giratis, Kokoszka and Leipus [30]. □

Remark 2.6 The central limit theorem in Theorem 2.5 can be derived from the mixing property of GARCH(p,q) if $E|\epsilon_0|^\mu < \infty$ with some $\mu > 0$ and $Ey_0^\kappa < \infty$ with some $\kappa > 4$. However, it is difficult to connect the moment condition on y_0 with (2.2) and assumptions on $\alpha_1, \dots, \alpha_p$, $\beta_1, \dots \beta_q$.

To illustrate the martingale properties of $\sum y_i$ we provide a short proof of the weak convergence of the partial sums of y_i.

Theorem 2.7 *If (1.1)-(1.3),(1.8), (1.9) hold, $E\epsilon_0 = 0, E\epsilon_0^2 = 1$ and*

$$\alpha_1 + \dots + \alpha_p + \beta_1 + \dots + \beta_q < 1, \tag{2.3}$$

then

$$n^{-1/2} \sum_{1 \leq i \leq nt} y_i \ \xrightarrow{\mathcal{D}[0,1]}\ \left(\frac{\omega}{1 - (\alpha_1 + \dots + \alpha_p + \beta_1 + \dots + \beta_q)}\right)^{1/2} W(t),$$

where $\{W(t), 0 \leq t < \infty\}$ denotes a Wiener process.

Proof Relation 2.3 implies that (1.10) holds and $Ey_0^2 < \infty$. Using (1.21) we also conclude that $E\sigma_0^2 < \infty$. Assumptions $E\epsilon_0 = 0, E\epsilon_0^2 = 1$ and (1.1)-(1.2) yield that $Ey_0 = 0$ and

$$Ey_0^2 = \frac{\omega}{1 - (\alpha_1 + \dots + \alpha_p + \beta_1 + \dots + \beta_q)}.$$

According to Theorem 1.4, y_i is a stationary and ergodic sequence. Also, since $E(y_i|\mathcal{F}_{i-1}) = \sigma_i E\epsilon_i = 0$ we get that $\{\sum_{1 \leq j \leq i} y_i, \mathcal{F}_i\}$ is a martingale. Hence Theorem 23.1 of Billingsley [13], p. 206, implies Theorem 2.7. □

Remark 2.8 Doukhan and Louhichi [23] introduced $(\theta, \mathcal{L}, \psi)$-weak dependence. Using Dufflo [26] (cf. Proposition 3.1 in Ango Nze and Doukham [1]) we get that a GARCH$(1,1)$ sequence is $(\theta, \mathcal{L}, \psi)$-weakly dependent with $\theta_r = O((E(\beta_1 + \alpha_1\epsilon_0^2))^r)$, if (1.1)-(1.3), (1.6), (1.8) and $E\epsilon_0^2 < \infty$ hold. Note that there is no condition on the smoothness of the distribution of ϵ_0.

Theorem 2.9 *If $p = q = 1$, (1.1)-(1.3), (1.6), (1.8), $E\epsilon_0^2 < \infty$ hold and*

$$E|y_0|^\kappa < \infty \ \textit{with some} \ \kappa > 8,$$

then

$$n^{-1/2} \sum_{1 \le i \le nt} (y_i^2 - Ey_0^2) \ \stackrel{\mathcal{D}[0,1]}{\longrightarrow} \ \Big(\sum_{-\infty < j < \infty} \mathrm{cov}(y_j^2, y_0^2)\Big)^{1/2} W(t), \tag{2.4}$$

where $\{W(t), 0 \le t < \infty\}$ denotes a Wiener process.

Proof We pointed out in Remark 2.8 that y_i^2 is $(\theta, \mathcal{L}, \psi)$-weakly dependent stationary sequence, so Theorem 5.1 in Ango Nze and Doukhan [1] can be used to establish the functional limit theorem in Theorem 2.9. □

Remark 2.10 Theorem 2.1 of Giraitis, Kokoszka and Leipus [31] implies that (2.4) holds for general GARCH(p,q) processes under the assumptions of Theorem 2.5, but with (2.2) strengthened to

$$(E\epsilon_0^8)^{1/4} \sum_{1 \le j < \infty} c_j < 1 \tag{2.5}$$

(only $E\epsilon_0^8 < \infty$ is required). Their proof uses assumptions on the cumulants of y_i^2 which are shown to hold under (2.5) and are used to verify tightness.

3 Estimation of the parameters

In applications the parameter $\boldsymbol{\theta} = (\omega, \alpha_1, \ldots, \alpha_p, \ \beta_1, \ldots, \beta_q)$ is fixed but is unknown to us. 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 $c_i, 0 \le i < \infty$, will be functions of $\mathbf{u}$ and satisfy the following recursion formulas: If $q \ge p$, then

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

and if $q < p$ the equations above are replaced with

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

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

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

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

$$\tilde{L}_n(\mathbf{u}) = \sum_{1<k\le n} -\frac{1}{2}\left\{\log \hat{w}_k(\mathbf{u}) + \frac{y_k^2}{\hat{w}_k(\mathbf{u})}\right\}, \tag{3.1}$$

where

$$\hat{w}_k(\mathbf{u}) = c_0(\mathbf{u}) + \sum_{1\le i<k} c_i(\mathbf{u}) y_{k-i}^2. \tag{3.2}$$

If we were to assume that ϵ_0 is standard normal, then conditionally on $\mathcal{F}_{k-1}, y_k/\sigma_k(\boldsymbol{\theta})$ is standard normal. The approximation of the likelihood in (3.1) is derived under this assumption. However, we will show that the quasi-maximum likelihood estimator $\tilde{\boldsymbol{\theta}}_n$, defined as

$$\tilde{\boldsymbol{\theta}}_n = \arg\max_{\mathbf{u}\in U} \tilde{L}_n(\mathbf{u}) \tag{3.3}$$

will be consistent and asymptotically normal without assuming that ϵ_0 is standard normal. The set U is chosen as follows:

$$
\begin{aligned}
U = \{\mathbf{u} : t_1 + t_2 + \cdots + t_q \le \rho_0 \quad &\text{and} \quad \underline{u} \le \min(x, s_1, s_2, \ldots, s_p, t_1, t_2, \ldots, t_q) \\
&\le \max(x, s_1, s_2, \ldots, s_p, t_1, t_2, \ldots, t_q) \le \overline{u}\}
\end{aligned}
$$

with some $0 < \underline{u} < \overline{u}$, $0 < \rho_0 < 1, q\underline{u} < \rho_0$. We assume that

$$\boldsymbol{\theta} \text{ is in the interior of } U. \tag{3.4}$$

Theorem 3.1 *We assume that (1.1), (1.2), (1.8), (1.9), (1.10), (3.4) hold and*

$$E|\epsilon_0^2|^{1+\delta} < \infty \qquad \textit{with some } \delta > 0, \tag{3.5}$$

$$\lim_{t\to 0} t^{-\mu} P\{\epsilon_0^2 \le t\} = 0 \qquad \textit{with some } \mu > 0 \tag{3.6}$$

and

$$E\epsilon_0^2 = 1. \tag{3.7}$$

Then

$$\tilde{\boldsymbol{\theta}}_n \overset{a.s.}{\to} \boldsymbol{\theta}, \quad \textit{as} \quad n \to \infty. \tag{3.8}$$

Proof Lumsdaine [54] proved (3.8) under stronger conditions if $p = q = 1$. Theorem 3.1 is also a special case of Lee and Hansen [48] when $p = q = 1$. For the proof of (3.8) we refer to Berkes, Horváth and Kokoszka [12]. □

Next we discuss the asymptotic normality of $n^{1/2}(\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta})$. Let

$$\ell_k(\mathbf{u}) = -\frac{1}{2}(\log w_k(\mathbf{u}) + y_k^2/w_k(\mathbf{u})) \tag{3.9}$$

with

$$w_k(\mathbf{u}) = c_0(\mathbf{u}) + \sum_{1 \le i < \infty} c_i(\mathbf{u}) y_{k-i}^2, \tag{3.10}$$

and introduce the matrices

$$\mathbf{A}_0 = \operatorname{cov}(\ell_0'(\boldsymbol{\theta})) \text{ and } \mathbf{B}_0 = E(\ell_0''(\boldsymbol{\theta})). \tag{3.11}$$

Theorem 3.2 *We assume that (1.1), (1.2), (1.8), (1.9), (1.10), (3.4), (3.6), (3.7) hold and*

$$E\epsilon_0^4 < \infty. \tag{3.12}$$

Then

$$\mathbf{A}_0 \text{ and } \mathbf{B}_0 \text{ are non–singular,}$$

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

and

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

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

Proof The asymptotic linearity as well as the asymptotic normality were proved by Berkes et al. [12] and Berkes and Horváth [7]. Earlier results for the GARCH$(1,1)$ case were obtained by Lee and Hansen [48] and Lumsdaine [54]. □

Remark 3.3 If $p = q = 1$, Lee and Hansen [48] replaced (1.8) with the assumption that y_i^2 is a martingale difference sequence.

Remark 3.4 Berkes and Horváth [8] proved that Theorem 3.2 cannot be true without assumption (3.12).

Remark 3.5 The definition of $\tilde{L}_n(\mathbf{u})$ is based on the standard normal density. Berkes and Horváth [7] showed that other densities can also be used to define $\tilde{L}_n(\mathbf{u})$. They proved analogues of Theorems 3.1 and 3.2 for a large class of $\tilde{L}_n(\mathbf{u})$ and pointed out that the estimator is "optimal" if the density of ϵ_0 is used to define $\tilde{L}_n(\mathbf{u})$.

Remark 3.6 Drost and Klaassen [24] used a semiparametric approach to estimate $\boldsymbol{\theta}$. They found a reparametrization such that the efficient score functions in the parametric model of the autoregression parameters are orthogonal to the tangent space generated by the nuisance parameters. They constructed adaptive and hence efficient estimators for $\boldsymbol{\theta}$.

Remark 3.7 For ARCH(p) processes a family of L_r-estimators is introduced in Horváth and Liese [43]. The construction is based on the concept of weighted M-estimators. Under weak assumptions on the error distribution, consistency is established. The asymptotic normality is proved for the L_1 and L_2 estimators. To prove the asymptotic normality of the L_1 estimator, one needs the existence of the density of ϵ_0^2 at the median of ϵ_0^2, whereas for the L_2 estimator $E\epsilon_0^4 < \infty$ is needed.

4 Squared residuals

Relation $\sigma_k^2 = w_k(\boldsymbol{\theta})$ suggests that $\hat{w}_k(\hat{\boldsymbol{\theta}}_n)$ can be used as an estimator for σ_k^2, where $\hat{\boldsymbol{\theta}}_n$ is an estimator for $\boldsymbol{\theta}$. The squared residuals are

$$\hat{\epsilon}_k^2 = \frac{y_k^2}{\hat{w}_k(\hat{\boldsymbol{\theta}}_n)}, \quad 2 \le k \le n.$$

In this section we will investigate the asymptotic properties of the empirical distribution function

$$\hat{F}_k(t) = \frac{1}{k} \sum_{2 \le i \le k} I\{\hat{\epsilon}_i^2 \le t\}, \quad 0 \le t < \infty, \ \ 2 \le k \le n$$

and the corresponding (sequential) empirical process of the squared residuals

$$\hat{\alpha}_n(t,s) = n^{1/2} s(\hat{F}_{[ns]}(t) - F(t)), \quad 0 \le t < \infty, \ \ 0 \le s \le 1,$$

where

$$F(t) = P\{\epsilon_0^2 \le t\}, \quad 0 \le t < \infty.$$

The next result shows that if $\hat{\boldsymbol{\theta}}_n$ is a strongly consistent estimator for $\boldsymbol{\theta}$, i.e.

$$\hat{\boldsymbol{\theta}}_n \to \boldsymbol{\theta} \qquad \text{a.s.,} \tag{4.1}$$

then $\hat{F}_n$ satisfies the Glivenko-Cantelli theorem.

Theorem 4.1 *If (1.1)-(1.3), (1.8), (1.9), (1.10), (3.4), (3.6), (4.1) hold,*

$$E|\epsilon_0^2|^\gamma < \infty \ \ \textit{with some} \ \ \gamma > 0 \tag{4.2}$$

and F is continuous, then

$$\lim_{n\to\infty} \sup_{0 \le t < \infty} |\hat{F}_n(t) - F(t)| = 0 \qquad a.s.$$

Proof For details we refer to Berkes and Horváth [5]. Horváth, Kokoszka and Teyssiére [41] obtained the result earlier for the empirical process of ARCH(p) squared residuals. □

The Glivenko-Cantelli theorem stated in Theorem 4.1 holds under the condition that F is continuous. The weak convergence of $\hat{\alpha}_n(t,s)$ will require the existence of a smooth density, $f = F'$. Namely, we assume that

$$\lim_{s\to 0} sf(s) = 0, \tag{4.3}$$

$$\lim_{s\to\infty} sf(s) = 0 \tag{4.4}$$

and

$$f \text{ is continuous on } (0,\infty). \tag{4.5}$$

Let

$$\mathbf{d}_0 = (d_1, \ldots, d_{p+q+1}) = E\frac{w_0'(\boldsymbol{\theta})}{w_0(\boldsymbol{\theta})}. \tag{4.6}$$

We assume that the estimator $\hat{\boldsymbol{\theta}}_n$ admits the representation

$$\begin{aligned}\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta} = \frac{1}{n}\sum_{1\le i\le n} \Big(& g_1(\epsilon_i)\ell_1(\epsilon_{i-1},\epsilon_{i-2},\ldots), g_2(\epsilon_i)\ell_2(\epsilon_{i-1},\epsilon_{i-2},\ldots), \\ & \ldots, g_{p+q+1}(\epsilon_i)\ell_{p+q+1}(\epsilon_{i-1},\epsilon_{i-2},\ldots)\Big) + o_P(n^{-1/2}).\end{aligned} \tag{4.7}$$

We pointed out in Section 3 that the most often used estimators of $\boldsymbol{\theta}$ satisfy (4.7). The random variables in (4.7) are also assumed to have at least two finite moments:

$$Eg_i(\epsilon_0) = 0 \quad \text{and} \quad Eg_i^2(\epsilon_0) < \infty, \quad 1 \le i \le p+q+1 \tag{4.8}$$

and

$$E\ell_i^2(\epsilon_{-1}, \epsilon_{-2}, \ldots) < \infty, \;\; 1 \le i \le p+q+1. \tag{4.9}$$

Let

$$\bar{q}_{ij} = Eg_i(\epsilon_0)g_j(\epsilon_0), \;\; \bar{\ell}_i = E\ell_i(\epsilon_{-1}, \epsilon_{-2}, \ldots),$$

$$\bar{\ell}_{ij} = E\ell_i(\epsilon_{-1}, \epsilon_{-2}, \ldots)\ell_j(\epsilon_{-1}, \epsilon_{-2}, \ldots), \;\; 1 \le i, j \le p+q+1$$

and

$$q_i(t) = E\{I\{\epsilon_0^2 \le t\}g_i(\epsilon_0)\}, \;\; 1 \le i \le p+q+1,$$

and

$$r(t, t', s, s') = (s \wedge s')(F(t \wedge t') - F(t)F(t')) + tf(t)ss' \sum_{1 \le i \le p+q+1} q_i(t')\bar{\ell}_i d_i$$

$$+ t'f(t')ss' \sum_{1 \le i \le p+q+1} q_i(t)\bar{\ell}_i d_i + tf(t)t'f(t')ss' \sum_{1 \le i,j \le p+q+1} \bar{q}_{ij}\bar{\ell}_{ij} d_i d_j,$$

where $t \wedge t' = \min(t, t')$.

Theorem 4.2 *If (1.1)-(1.3), (1.8)–(1.10), (3.4), (3.6), (4.2)-(4.5) and (4.7)–(4.9) hold, then*

$$\hat{\alpha}_n(t, s) \to \Gamma(t, s),$$

where the convergence is in the Skorokhod space $\mathcal{D}([0, \infty] \times [0, 1])$ and Γ is a Gaussian process with $E\Gamma(t, s) = 0$ and $E\Gamma(t, s)\Gamma(t', s') = r(t, t', s, s')$.

Proof Horváth, Kokoszka and Teyssiére [41] obtained the result earlier for ARCH(p) squared residuals. The general case was proven by Berkes and Horváth [5]. □

Remark 4.3 The process $\Gamma(t, s)$ in Theorem 4.2 depends on the estimator for $\boldsymbol{\theta}$ and on f on an intricate way. However,

$$\{\Gamma(t, s) - s\Gamma(t, 1), 0 \le t < \infty, 0 \le s \le 1\} \overset{\mathcal{D}}{=} \{K(F(t), s) - sK(F(t), 1), 0 \le t < \infty, 0 \le s \le 1\},$$

where $K(x, y), 0 \le x, y \le 1$ is a Kiefer process. Hence under the conditions of Theorem 4.2

$$\sup_{0 \le t < \infty} \sup_{0 \le s \le 1} |\hat{\alpha}_n(t, s) - s\hat{\alpha}_n(t, 1)| \;\; \xrightarrow{\mathcal{D}} \;\; \sup_{0 \le x \le 1} \sup_{0 \le y \le 1} |K(x, y) - yK(x, 1)|$$

so the limit does not depend on $\boldsymbol{\theta}$ nor on f.

Bootstrap specification and change-point tests based on the squared residuals are investigated by Horváth, Kokoszka and Teyssiére [42] and Kokoszka and Teyssiére [47].

In the rest of this section we assume that $\tilde{\boldsymbol{\theta}}_n$, the quasi-maximum likelihood estimator of Section 3 is used to estimate $\boldsymbol{\theta}$. If the innovations $\epsilon_1, \ldots, \epsilon_n$ were observable, an estimator for the autocovariance function of the squared innovations would be

$$r_n(k) = \frac{1}{n} \sum_{k < i \le n} (\epsilon_i^2 - 1)(\epsilon_{i-k}^2 - 1). \tag{4.10}$$

(We assume in (4.10) that (3.7) holds, i. e. $E\epsilon_0^2 = 1$.) Since $\epsilon_1^2, \ldots, \epsilon_n^2$ are not observable, we replace them with the squared residuals

$$\tilde{\epsilon}_i^2 = y_i^2 / \hat{w}_i(\tilde{\boldsymbol{\theta}}_n), \;\; 1 \le i \le n,$$

where $\hat{w}_n(\mathbf{u})$ is defined in (3.2) and we shall use

$$\hat{r}_n(k) = \frac{1}{n}\sum_{k<i\le n}(\hat{\epsilon}_i^2 - 1)(\hat{\epsilon}_{i-k}^2 - 1).$$

Our next result compares $\hat{r}_n(k)$ to $r_n(k)$. Let

$$\mathbf{c}_k = E\left[(\epsilon_{-k}^2 - 1)\frac{w_0'(\boldsymbol{\theta})}{w_0(\boldsymbol{\theta})}\right],$$

where $w_0(\mathbf{u})$ is defined in (3.10). We use $\mathbf{u}^T$ to denote transpose of a vector or a matrix.

Theorem 4.4 *If (1.1)–(1.3), (1.8)–(1.10), (1.26), (1.28), (3.4), (3.6), (3.7), (3.12) hold, then for any* $0 \le k < \infty$

$$\hat{r}_n(k) = r_n(k) + (\boldsymbol{\theta} - \tilde{\boldsymbol{\theta}}_n)\mathbf{c}_k^T + o_P(n^{-1/2}).$$

Proof Horváth and Kokoszka [40] obtained the result for the correlation between estimated squared residuals of ARCH(p) sequences. The present version was proven by Berkes, Horváth and Kokoszka [9]. □

According to Theorem 4.4 the first order properties (i.e. the mean, law of large numbers) are the same for $\hat{r}_n(k)$ as for $r_n(k)$. However, they differ in the second order properties (i.e. variance, central limit theorem).

Theorem 4.4 implies the asymptotic normality of $\hat{r}_k(k)$ for any $1 \le k < \infty$. In order to get χ^2–statistics we shall prove the joint asymptotic normality of $n^{1/2}(\hat{r}_n(i_1), \ldots, \hat{r}_n(i_K))$ for any $1 \le i_1 < i_2 < \cdots < i_K$. We start with the description of the asymptotic covariance matrix of $n^{1/2}(\hat{r}_n(i_1), \ldots, \hat{r}_n(i_K))$. Let

$$d_0^2 = E(\epsilon_0^2 - 1)^2$$

and $\mathbf{C}$ be the $K \times (p+q+1)$ matrix defined as

$$\mathbf{C} = \begin{pmatrix} \mathbf{c}_{i_1} \\ \mathbf{c}_{i_2} \\ \vdots \\ \mathbf{c}_{i_K} \end{pmatrix}$$

and let $\mathbf{I}_K$ denote the $K \times K$ identity matrix. The asymptotic covariance matrix is

$$\mathbf{D} = d_0^4(\mathbf{I}_K + \mathbf{CSC}^T),$$

where

$$\mathbf{S} = \frac{1}{d_0^2}\mathbf{B}_0^{-1} + \frac{1}{d_0^4}\mathbf{B}_0^{-1}\mathbf{A}_0\mathbf{B}_0^{-1}.$$

Theorem 4.5 *If (1.1)–(1.3), (1.8)–(1.10), (1.26), (1.28), (3.4), (3.6), (3.7), (3.12) hold, then*

$$n^{1/2}(\hat{r}_n(i_1), \ldots, \hat{r}_n(i_K)) \xrightarrow{\mathcal{D}} N(\mathbf{0}, \mathbf{D}),$$

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

Proof Li and Mak [50] and Ling and Li [52] assumed that conditionally on the past the y's have normal distribution. In our setting, this reduces to standard normal innovations ϵ_i. Berkes, Horváth and Kokoszka [9] obtained the central limit theorem without assuming the normality of the innovations. □

Theorem 4.5 yields the convergence in distribution of quadratic forms of $\hat{r}_n(i_1), \hat{r}_n(i_2), \ldots, \hat{r}_n(i_K)$.

Theorem 4.6 *If (1.1)–(1.3), (1.8)–(1.10), (1.26), (1.28), (3.4), (3.6), (3.7), (3.12) hold and*

$$P\{\epsilon_0^2 \le t\} \text{ is continous,} \tag{4.11}$$

then

$$\mathbf{D} \text{ is a positive definite matrix} \tag{4.12}$$

and

$$n(\hat{r}_n(i_1), \ldots, \hat{r}_n(i_K))\mathbf{D}^{-1}(\hat{r}_n(i_1), \ldots, \hat{r}_n(i_K))^T \xrightarrow{\mathcal{D}} \chi^2(K), \tag{4.13}$$

where $\chi^2(K)$ denotes a chi-square random variable with K degrees of freedom.

Proof Li and Mak [50] and Ling and Li [52] claimed (4.13) when the innovations are standard normal. However, the existence of $\mathbf{D}^{-1}$ was used in Li [49], Li and Mak [50] and Ling and Li [52] without proving it. For the proof of Theorem 4.6 we refer to Berkes, Horváth and Kokoszka [9]. □

Remark 4.7 For a weighted version of (4.13) we refer to Berkes, Horváth and Kokoszka [11].

5 Further results

The GARCH$(1,1)$ sequence of Section 1 is a special case of the so-called augmented GARCH$(1,1)$ process introduced by Duan [25]:

$$y_k = \sigma_k \epsilon_k \tag{5.1}$$

and

$$\Lambda(\sigma_k^2) = g_1(\epsilon_{k-1})\Lambda(\sigma_{k-1}^2) + g_2(\epsilon_{k-1}), \tag{5.2}$$

where $\Lambda(\cdot)$ is increasing and continous on $[0, \infty)$ and $g_1(\cdot), g_2(\cdot)$ are measurable functions. By Brandt [18], if (1.4) holds, $E|\log|g_1(\epsilon_0)|| < \infty, E|\log|g_2(\epsilon_0)|| < \infty$ and $E\log|g_1(\epsilon_0)| < 0$, then (5.1) and (5.2) have a stationary and ergodic solution. The augmented GARCH$(1,1)$ is a generalization of the following models: LGARCH of Bollerslev [14], NGARCH of Engle and Ng [28], MGARCH of Geweke [29], EGARCH of Nelson [57] and VGARCH of Engle and Ng [28]. A different generalization of the GARCH$(1,1)$ sequence of Section 1 is considered by He and Teräsvirta [39]. The method in Section 3 can be adopted to estimate the parameters of these processes.

A common property of ARCH(p) and GARCH(p,q) sequences is that they are defined by finite recursions and their autocorrelations decrease very rapidly, implying short memory behavior of these sequences. Short memory behavior holds even for ARCH(∞) models defined by the infinite recursion

$$y_k = \sigma_k \epsilon_k \tag{5.3}$$

$$\sigma_k^2 = a + \sum_{i=1}^{\infty} b_i y_{k-i}^2, \qquad k \in \mathbf{Z} \tag{5.4}$$

where (b_k) is a sequence of non-negative numbers and (ϵ_k) is an independent, identically distributed sequence of random variables having suitable moments. In fact, under reasonable conditions implying the existence of a covariance stationary solution of (5.3) and (5.4), we automatically have

$$\sum_{k=1}^{\infty} |\mathrm{Cov}\,(y_0^2, y_k^2)| < +\infty$$

implying short range dependence of the sequence (cf. Giraitis, Kokoszka and Leipus [30] and Kazakevičius and Leipus [44]). On the other hand, empirical evidence suggests in many typical financial situations a much greater degree of persistence in the process, indicating a long memory behavior of y_k. A model describing such long memory behavior was suggested by Giraitis, Robinson and Surgailis [34]. They introduced a model called LARCH ("Linear ARCH") defined by

$$y_k = \sigma_k \epsilon_k \tag{5.5}$$

$$\sigma_k = a + \sum_{i=1}^{\infty} b_i y_{k-i}. \tag{5.6}$$

As they showed, if $a \neq 0$, the ϵ_i are independent, identically distributed with $E\epsilon_0 = 0, E\epsilon_0^2 = 1$ and

$$\sum_{i=1}^{\infty} b_i^2 < 1 \tag{5.7}$$

then (5.5) and (5.6) have a unique stationary solution admitting a Volterra expansion

$$\sigma_n = a + a\sum_{k=1}^{\infty} \sum_{j_1,\ldots,j_k=1}^{\infty} b_{j_1}\ldots b_{j_k}\epsilon_{n-j_1}\ldots\epsilon_{n-j_1-\ldots-j_k} \tag{5.8}$$

and if

$$b_j \sim cj^{-\beta}, \qquad 1/2 < \beta < 1 \tag{5.9}$$

then σ_n has long memory behavior. (Here $\sim$ means that the ratio of the left and right sides tends to 1.) In particular,

$$N^{-(\frac{3}{2}-\beta)} \sum_{j=1}^{[Nt]} (\sigma_j - a) \qquad 0 \le t \le 1$$

converges, as $N \to \infty$, not to the standard Brownian motion as it should be the case in a short memory situation, but to the fractional Brownian motion $W_{3/2-\beta}$. Let us recall that $\{W_\gamma(t),\, t \ge 0\}$, the fractional Brownian motion with parameter γ $(0 < \gamma < 1)$, is a Gaussian process with mean 0 and covariance

$$EW_\gamma(s)W_\gamma(t) = \frac{1}{2}(|s|^{2\gamma} + |t|^{2\gamma} - |s-t|^{2\gamma}).$$

See e.g. Samorodnitsky and Taqqu [60], Chapter 7.

Surgailis [62] (see also Giraitis and Surgailis [35] [36] [37]) extended the Dobrushin-Major-Taqqu theory (cf. Taqqu [63], [64] and Dobrushin-Major [22]) to linear (moving average) processes defined by

$$\xi_n = \sum_{-\infty<j<\infty} b_j \epsilon_{n-j},$$

where $\epsilon_i, -\infty < i < \infty$ are independent, identically distributed random variables with $E\epsilon_0 = 0, E\epsilon_0^2 = 1$ and b_j are real numbers with $\sum b_j^2 < +\infty$. Long memory behavior holds here if (5.9) is satisfied. In this case Surgailis [62] showed that the class of the limiting processes of $a_N^{-1}\sum_{k=1}^{[Nt]} h(\xi_k)$ is the same as in the Gaussian case, just the role of the Hermite polynomials is played by another polynomial sequence, the so-called Appell polynomials.

Berkes and Horváth [4] investigated the limiting behavior of sums $\sum h(\sigma_n)$, $\sum h(y_n)$, where (y_n, σ_n) is a long memory LARCH sequence defined by (5.5) and (5.6), where $a \neq 0$, ϵ_i are independent, identically distributed random variables with $E\epsilon_0 = 0$, $E\epsilon_0^2 = 1$ and b_j are positive numbers satisfying (5.9). The strong similarity between linear sequences and the basic recursion formula for LARCH sequences was exploited by Giraitis, Robinson and Surgailis [34] to determine the limit distribution of $\sum(\sigma_n - a)$. In fact, they proved that the asymptotic behavior of $\sum(\sigma_n - a)$ is the same as that of $\sum \sigma_n^*$, where σ_n^* is the linear process defined by

$$\sigma_n^* = \sum_{j=1}^{\infty} b_j \delta_{n-j}, \tag{5.10}$$

where δ_j are independent, identically distributed random variables with $\delta_0 \stackrel{\mathcal{D}}{=} \epsilon_0\sigma_0$.

Berkes and Horváth [4] showed that this phenomenon does not extend to $\sum h(\sigma_n)$ with general h: the variances of $\sum_{n=1}^N h(\sigma_n)$ and $\sum_{n=1}^N h(\sigma_n^*)$ can grow with different speed and the sums have, in general, different limit distributions. However, they showed that $\sum h(\sigma_n)$, $\sum h(y_n)$ exhibit a long memory behavior, similar to that of Gaussian and linear processes, if $E(\sigma_n h'(\sigma_n)) \neq 0$ and $E(y_n h'(y_n)) \neq 0$. In this case the variances of the sums grow as $CN^{3-2\beta}$ and the sums, properly normalized, converge to the fractional Brownian motion.

Theorem 5.1 *We assume that (5.5), (5.6) hold, $E\epsilon_0 = 0$, $E\epsilon_0^2 = 1$, $E|\epsilon_0|^\nu < +\infty$ for some $\nu > 4$ and (5.9) is satisfied with*

$$b^2 = \sum_{n=1}^{\infty} b_n^2 < \frac{\nu - 1}{3(6\nu)^3 ||\epsilon_0||_\nu^2} \tag{5.11}$$

where $||\cdot||_\nu$ denotes the L_ν norm. Let $h : \mathbb{R} \to \mathbb{R}$ be a twice continuously differentiable function with

$$|h''(x)| \le C(|x|^\alpha + 1) \qquad x \in \mathbb{R}$$

where

$$0 < \alpha < (\nu - 4)^2/(2\nu).$$

Then

$$N^{-(\frac{3}{2}-\beta)} \sum_{n=1}^{[Nt]} (h(\sigma_n) - Eh(\sigma_n)) \stackrel{\mathcal{D}[0,1]}{\longrightarrow} \gamma d W_{3/2-\beta}(t)$$

and

$$N^{-(\frac{3}{2}-\beta)} \sum_{n=1}^{[Nt]} (h(y_n) - Eh(y_n)) \stackrel{\mathcal{D}[0,1]}{\longrightarrow} \gamma_1 d W_{3/2-\beta}(t)$$

where

$$\gamma = \frac{1}{a} E(\sigma_0 h'(\sigma_0)), \quad \gamma_1 = \frac{1}{a} E(y_0 h'(y_0))$$

and

$$d = \left(\frac{B(1-\beta, 2\beta - 1)}{(1-\beta)(3-2\beta)} \right)^{1/2} \frac{ac}{(1-b^2)^{1/2}}.$$

Here a, b, c are from (5.6), (5.9), (5.11) and $B(.,.)$ is the beta function.

Proof We refer to Berkes and Horváth [4]. □

Remark 5.2 In particular, under the conditions of Theorem 5.1 we have

$$N^{-(\frac{3}{2}-\beta)} \sum_{n=1}^{N} (h(\sigma_n) - Eh(\sigma_n)) \stackrel{\mathcal{D}}{\longrightarrow} N(0, d^2\gamma^2)$$

and

$$N^{-(\frac{3}{2}-\beta)} \sum_{n=1}^{N} (h(y_n) - Eh(y_n)) \stackrel{\mathcal{D}}{\longrightarrow} N(0, d^2\gamma_1^2).$$

Giraitis et al. [33] study the cross-covariance function $\mathrm{Cov}(\sigma_k^2, y_0)$ in the LARCH model and establish conditions for the finiteness of higher order moments of y_k.

Giraitis et al. [32] consider R/S-type tests for distinguishing between LARCH processes described above and their short memory counterparts for which the b_j decay exponentially fast.

References

[1] P. Ango Nze and P. Doukhan. Weak dependence: models and applications. In: H. Dehling, T. Mikosch, and M. Sorensen, editors, *Empirical Process Techniques for Dependent data*, pages 117–136. Birkhäuser, 2002.

[2] B. Basrak, R. Davis, and T. Mikosch. Regular variation of GARCH processes. *Stochastic Processes and their Applications*, 99:95–115, 2002.

[3] I. Berkes and L. Horváth. Strong approximation of the empirical process of GARCH sequences. *The Annals of Applied Probability*, 11:789–809, 2001.

[4] I. Berkes and L. Horváth. Asymptotic results for long memory LARCH sequences. *The Annals of Applied Probability*, 2003, forthcoming.

[5] I. Berkes and L. Horváth. Limit results for the empirical process of squared residuals in GARCH models. *Stochastic Processes and Their Applications*, 2003, forthcoming.

[6] I. Berkes and L. Horváth. Correction to 'Strong approximation of the empirical process of GARCH sequences, The Annals of Applied Probability 11 (2001) 789–809'. *The Annals of Applied Probability*, 2003, forthcoming.

[7] I. Berkes and L. Horváth. The efficiency of the estimators of the parameters in GARCH processes. *Annals of Statistics*, 2003, forthcoming.

[8] I. Berkes and L. Horváth. The rate of consistency of of the quasi-maximum likelihood estimator. *Statistics and Probability Letters*, 61:133–143, 2003.

[9] I. Berkes, L. Horváth and P. S. Kokoszka. Asymptotics for GARCH squared residual correlations. *Econometric Theory*, 2003, forthcoming. Preprint available at `http://math.usu.edu/~piotr/research.html`.

[10] I. Berkes, L. Horváth and P. S. Kokoszka. Estimation of the maximal moment exponent of a GARCH(1,1) sequence. 2002. Preprint available at `http://math.usu.edu/~piotr/research.html`.

[11] I. Berkes, L. Horváth and P. S. Kokoszka. A weighted χ^2-test for GARCH(1,1) squared residual correlations,, 2002. Preprint available at `http://math.usu.edu/ piotr/research.html`.

[12] I. Berkes, L. Horváth and P. S. Kokoszka. GARCH processes: structure and estimation. *Bernoulli*, 2003, forthcoming. Preprint available at `http://math.usu.edu/~piotr/research.html`.

[13] P. Billingsley. *Convergence of Probability Measures.* Wiley, New York, 1968.

[14] T. Bollerslev. Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31:307–327, 1986.

[15] P. Bougerol and N. Picard. Strict stationarity of generalized autoregressive processes. *The Annals of Probability*, 20:1714–1730, 1992a.

[16] P. Bougerol and N. Picard. Stationarity of GARCH processes and of some nonnegative time series. *Journal of Econometrics*, 52:115–127, 1992b.

[17] F. Boussama. *Ergodicity, Mixing and Estimation in GARCH Models.* PhD thesis, University of Paris 6, 2000. In French; translated by F. Comte as "Probabilistic properties of multivariate GARCH processes".

[18] A. Brandt. The stochastic equation $y_{n+1} = a_n y_n + b_n$ with stationary coefficients. *Advances in Applied Probability*, 18:211–220, 1986.

[19] L. Breiman. *Probability.* Addison-Wesley, Redding, Mass., 1968.

[20] M. Carrasco and X. Chen. Mixing and moment properties of various GARCH and stochastic volatility models. *Econometric Theory*, 18:17–39, 2002.

[21] Yu. A. Davydov. Mixing conditions for Markov chains. *Theory of Probability and its Applications*, 18:312–328, 1973.

[22] R. L. Dobrushin and P. Major. Non-central limit theorems for non-linear functions of Gaussian fields. *Zeitshrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 50:27–52, 1979.

[23] P. Doukhan and S. Louhichi. A new weak dependence and applications to moment inequalities. *Stochastic Processes and their Applications*, 84:313–343, 1999.

[24] F. C. Drost and C. A. Klaassen. Efficient estimation in semiparametric GARCH models. *Journal of Econometrics*, 81:193–221, 1997.

[25] J. C. Duan. Augmented GARCH(p,q) process and its diffusion limit. *Journal of Econometrics*, 79:97–127, 1997.

[26] M. Dufflo. *Méthodes Récursives Aléatoires.* Masson, Paris, 1996. English Edition, Springer Verlag, 1996.

[27] R. F. Engle. Autoregressive conditional heteroskedasticity with estimates of the variance of United Kingdom inflation. *Econometrica*, 50:987–1007, 1982.
[28] R. F. Engle and V. Ng. Measuring and testing the impact of news on volatility. *Journal of Finance*, 48:1749–1778, 1993.
[29] J. Geweke. Modeling the persistence of conditional variances: a comment. *Econometric Reviews*, 5:57–61, 1986.
[30] L. Giraitis, P. S. Kokoszka and R. Leipus. Stationary ARCH models: dependence structure and Central Limit Theorem. *Econometric Theory*, 16:3–22, 2000.
[31] L. Giraitis, P. S. Kokoszka and R. Leipus. Testing for long memory in the presence of a general trend. *Journal of Applied Probability*, 38:1033–1054, 2001.
[32] L. Giraitis, P. S. Kokoszka, R. Leipus and G. Teyssiere. Rescaled variance and related tests for long memory in volatility and levels. *Journal of Econometrics*, 2003. forthcoming.
[33] L. Giraitis, R. Leipus, P. Robinson and D. Surgailis. LARCH, leverage and long memory. Technical report, London School of Economics, 2002.
[34] L. Giraitis, P. Robinson, and D. Surgailis. A model for long memory conditional heteroskedasticity. *The Annals of Applied Probability*, 10:1002–1024, 2000.
[35] L. Giraitis and D. Surgailis. Multivariate Appell polynomials and the central limit theorem. In: E. Eberlein and M. S. Taqqu, editors, *Dependence in Probability and Statistics*. Birkhäuser, New York, 1986.
[36] L. Giraitis and D. Surgailis. A limit theorem for polynomials of linear processes with long range dependence. *Lietuvos Matematikos Rinkinys*, 29:290–311, 1989.
[37] L. Giraitis and D. Surgailis. A central limit theorem for the empirical process of a long memory linear sequence. *Journal of Statistical Planning and Inference*, 80:81–93, 1999.
[38] C. Gouriéroux. *ARCH Models and Financial Applications*. Springer-Verlag, New York, 1997.
[39] C. He and T. Teräsvirta. Properties of moments of a family of GARCH processes. *Journal of Econometrics*, 92:173–192, 1999.
[40] L. Horváth and P. S. Kokoszka. Large sample distribution of ARCH(p) squared residual correlations. *Econometric Theory*, 17:283–295, 2001.
[41] L. Horváth, P. S. Kokoszka and G. Teyssiére. Empirical process of the squared residuals of an ARCH sequence. *The Annals of Statistics*, 29:445–469, 2001.
[42] L. Horváth, P. S. Kokoszka and G. Teyssiére. Bootstrap specification tests for ARCH based on the empirical process of squared residuals, 2002. Preprint available at `http://www.math.usu.edu/~piotr/research.html`.
[43] L. Horváth and F. Liese. L_p–estimators in ARCH models. *Journal of Statistical Planning and Inference*, 2003, forthcoming.
[44] V. Kazakevičius and R. Leipus. On the stationarity in the ARCH(∞) model. *Econometric Theory*, 18:1–16, 2002.
[45] V. Kazakevičius and R. Leipus. A new theorem on existence of invariant distributions with applications to ARCH processes. *Journal of Applied Probability*, 40:147–162, 2003.
[46] J. F. C. Kingman. Subaditive ergodic theory. *The Annals of Probability*, 1:883–909, 1973.
[47] P. S. Kokoszka and G. Teyssière. Change point detection in GARCH models: asymptotic and bootstrap tests. Working paper available at `http://stat.usu.edu/~piotr/research.html`.
[48] S. W. Lee and B. E. Hansen. Asymptotic theory for the GARCH(1,1) quasi-maximum likelihood estimator. *Econometric Theory*, 10:29–52, 1994.
[49] W. K. Li. On the asymptotic standard errors of residual autocorrelations in nonlinear time series modeling. *Biometrika*, 79:435–437, 1992.
[50] W. K. Li and T. K. Mak. On the squared residual autocorrelations in non–linear time series with conditional heteroskedasticity. *Journal of Time Series Analysis*, 15:627–636, 1994.
[51] S. Ling. On the probabilistic properties of a double threshhold ARMA conditional heteroskedastic model. *Journal of Applied Probability*, 36:688–705, 1999.
[52] S. Ling and W. K. Li. On fractionally integrated autoregressive moving–average time series models with conditional heteroscedasticity. *Journal of the American Statistical Association*, 92:1184–1194, 1997.
[53] S. Ling and M. McAleer. Necessary and sufficient moment conditions for GARCH(r, s) and asymmetric power GARCH(r, s) model. *Econometric Theory*, 18:722–729, 2002.
[54] R. L. Lumsdaine. Consistency and asymptotic normality of the quasi–likelihood estimator in IGARCH(1,1) and covariance stationary GARCH(1,1) models. *Econometrica*, 64:575–596, 1996.
[55] S. P. Meyn and R. L. Tweedie. Stability of Markov processes I: criteria for discrete-time chains. *Advances in Applied Probability*, 24:542–574, 1982.
[56] D. B. Nelson. Stationarity and persistence in the GARCH($1,1$) model. *Econometric Theory*, 6:318–334, 1990.

[57] D. B. Nelson. Conditional heteroskedasticity in asset returns: a new approach. *Econometrica*, 59:347–370, 1991.

[58] W. Phillipp. Invariance principles for sums of mixing random elements and the multivariate empirical process. In: P. Révész, editor, *Limit Theorems in Probability and Statistics*, pages 843–873. North Holland, 1984.

[59] W. Phillipp and L. Pinzur. Almost sure approximation theorems for multivariate empirical processes. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 54:1–13, 1980.

[60] G. Samorodnitsky and M. S. Taqqu. *Stable Non-Gaussian Random Processes: Stochastic Models with Infinite Variance*. Chapman & Hall, 1994.

[61] W. F. Stout. *Almost Sure Convergence*. Academic Press, New York, 1974.

[62] D. Surgailis. Zones of attraction of self-similar multiple integrals. *Lithuanian Mathematical Journal*, 22:327–340, 1982.

[63] M. S. Taqqu. Weak convergence to fractional Brownian motion and to the Rosenblatt process. *Zeitshrift für Wahrscheinlichkeitstheorie und verwandte Gebiet*, 31:287–302, 1975.

[64] M. S. Taqqu. Convergence of integrated processes of arbitrary Hermite rank. *Zeitshrift für Wahrscheinlichkeitstheorie und verwandte Gebiet*, 50:53–83, 1979.