MaPhySto
The Danish National Research Foundation:
Network in Mathematical Physics and Stochastics

Research Report
no. 11 March 2004

Thomas Mikosch and Daniel Straumann:

Stable Limits of Martingale Transforms with Application to the Estimation of Garch Parameters

ISSN 1398-2699

STABLE LIMITS OF MARTINGALE TRANSFORMS WITH APPLICATION TO THE ESTIMATION OF GARCH PARAMETERS

THOMAS MIKOSCH AND DANIEL STRAUMANN

Abstract. In this paper we study the asymptotic behavior of the Gaussian quasi maximum likelihood estimator of a stationary GARCH process with heavy-tailed innovations. This means that the innovations are regularly varying with index $\alpha \in (2,4)$. Then, in particular, the marginal distribution of the GARCH process has infinite fourth moment and standard asymptotic theory with normal limits and $\sqrt{n}$-rates breaks down. This was recently observed by Hall and Yao (2003). It is the aim of this paper to indicate that the limit theory for the parameter estimators in the heavy-tailed case nevertheless very much parallels the normal asymptotic theory. In the light-tailed case, the limit theory is based on the CLT for stationary ergodic finite variance martingale difference sequences. In the heavy-tailed case such a general result does not exist, but an analogous result with infinite variance stable limits can be shown to hold under certain mixing conditions which are satisfied for GARCH processes. It is the aim of the paper to give a general structural result for infinite variance limits which can also be applied in situations more general than GARCH.

1. Introduction

The motivation for writing this paper comes from Gaussian quasi maximum likelihood estimation (QMLE) for GARCH (*generalized autoregressive conditionally heteroscedastic*) processes with regularly varying noise; we refer to Section 4 for a detailed description of the problem. Recall that the process

$$X_t = \sigma_t\, Z_t\,, \quad \text{with} \quad \sigma_t^2 = \alpha_0 + \sum_{i=1}^{p} \alpha_i X_{t-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2\,, \quad t \in \mathbb{Z}\,, \tag{1.1}$$

is said to be a GARCH(p,q) process (GARCH process of *order* (p,q)). Here (Z_t) is an iid sequence with $EZ_1^2 = 1$ and $EZ_1 = 0$, and α_i, β_j are non-negative constants. GARCH processes and their parameter estimation have been intensively investigated over the last few years; see Mikosch [19] for a general overview and Straumann and Mikosch [28] and the references therein for parameter estimation in GARCH and related models. In the context of QMLE the asymptotic behavior of the parameter estimator is essentially determined by the limiting behavior of the following quantity, see (4.21),

$$L_n'(\boldsymbol{\theta}_0) = \frac{1}{2} \sum_{t=1}^{n} \frac{h_t'(\boldsymbol{\theta}_0)}{\sigma_t^2}\,(Z_t^2 - 1)\,,$$

where L_n' is the derivative of the underlying log-likelihood, h_t' is the derivative of σ_t^2 when considered as a function of the parameter $\boldsymbol{\theta}$, and $\boldsymbol{\theta}_0$ is the true parameter (consisting of the α_i and β_j values)

Date: November 21, 2003.

Key words and phrases. GARCH process, Gaussian quasi-maximum likelihood, regular variation, infinite variance, stable distribution, stochastic recurrence equation, mixing

AMS 2000 Subject Classification: Primary: 62F12 Secondary: 62G32 60E07 60F05 60G42 60G70.

Thomas Mikosch's research was supported by MaPhySto, the Danish Research Network for Mathematical Physics and Stochastics, DYNSTOCH, a research training network under the Improving Human Potential Programme financed by the Fifth Framework Programme of the European Commission, and by the Danish Natural Science Research Council (SNF) grant no. 21-01-0546.

in a certain parameter space. In this context,

$$\mathbf{G}_t = \frac{h_t'(\boldsymbol{\theta}_0)}{\sigma_t^2}, \quad t \in \mathbb{Z},$$

is a stationary ergodic sequence of vector-valued random variables which is adapted to the filtration $\mathcal{F}_t = \sigma(Y_{t-1}, Y_{t-2}, \ldots)$, $t \in \mathbb{Z}$, where $Y_t = Z_t^2 - 1$ constitutes an iid sequence.

If $\mathbf{G}_t$ has a finite first moment the sequence $(\mathbf{G}_t Y_t)$ is a transform of the martingale difference sequence (Y_t), hence a stationary ergodic martingale difference sequence with respect to $(\mathcal{F}_t)$. If $E|\mathbf{G}_1|^2 < \infty$ and $EY_1^2 < \infty$, an application of the central limit theorem (CLT) for finite variance stationary ergodic martingale differences (see Billingsley [4], Theorem 23.1) yields

$$n^{-1/2} \sum_{t=1}^{n} \mathbf{G}_t Y_t \xrightarrow{d} N(\mathbf{0}, \boldsymbol{\Sigma}),$$

where $\boldsymbol{\Sigma}$ is the covariance matrix of $\mathbf{G}_1 Y_1$. This result does not require any additional information about the dependence structure of $(\mathbf{G}_t Y_t)$. It implies the asymptotic normality of the parameter estimator based on QMLE.

If $EY_1^2 = \infty$ a result as general as the CLT for stationary ergodic martingale differences is not known. However, some limit results for stationary sequences with marginal distribution in the domain of attraction of an infinite variance stable distribution exist. We recall two of them in Section 2. Our interest in infinite variance stable limit distributions for $\sum_{t=1}^n \mathbf{G}_t Y_t$ is again closely related to parameter estimation for GARCH processes. Recently, Hall and Yao [15] gave the asymptotic theory for QMLE in GARCH models when $EZ_1^4 = \infty$. To be more specific, they assume regular variation with index $\alpha \in (1,2)$ for the distribution of Z_1^2. It is our aim to show that their results can be obtained by a general limit result for the martingale transforms $\sum_{t=1}^n \mathbf{G}_t Y_t$ when the iid noise (Y_t) is regularly varying with index $\alpha \in (1,2)$. The key notions in this context are *regular variation* of the finite-dimensional distributions of $(\mathbf{G}_t Y_t)$ and strong mixing of this sequence, see Section 2 for these notions.

Our objective is twofold. *First,* we want to show that the theories on parameter estimation for GARCH processes with heavy- or light-tailed innovations (Z_t) parallel each other. We use the recent structural approach to GARCH estimation by Berkes et al. [3] in order to show that such a unified approach is possible. *Second,* our approach to the asymptotic theory for parameter estimators is not restricted to GARCH processes. In the light-tailed case, Straumann and Mikosch [28] extended the approach by Berkes et al. [3], including among others AGARCH and EGARCH processes. The main difficulty of our approach when infinite variance limits occur is the verification of certain mixing conditions. In contrast to the case of asymptotic normality, such conditions cannot be avoided. However, it is difficult to check for a given model that these conditions hold; see Section 4.4 in order to get a flavor of the task to be solved.

GARCH processes and their parameter estimation give the motivation for this paper. The corresponding limit theory for the QMLE with heavy-tailed innovations can be found in Section 4. Our main tool for achieving these limit results is based on asymptotic theory for martingale transforms with infinite variance stable limits. This theory is formulated and proved in Section 3. It is based on more general results for sums of stationary mixing vector sequences with regularly varying finite-dimensional distributions. This theory is outlined in Section 2.

2. Preliminaries

In this section we collect some basic tools and notions to be used throughout this paper. First we want to formulate a classical result on infinite variance stable limits for iid vector-valued summands due to Rvačeva [26]. Before we formulate this result we recall the notions of *stable random vector* and *multivariate regular variation.* The class of stable random vectors coincides with the class of possible limit distributions for sums of iid random vectors, and multivariate regular variation is the domain of attraction condition for sums of iid random vectors. Then we continue with an analog of Rvačeva's result for stationary ergodic vector sequences. In this context, we also need to recall some *mixing conditions*.

Stable random vectors. Recall that a vector $\mathbf{X}$ with values in $\mathbb{R}^d$ is said to be α-stable for some $\alpha \in (0,2)$ if its characteristic function is given by

$$Ee^{\,i(\mathbf{x},\mathbf{X})} = \begin{cases} \exp\left\{-\int_{\mathbb{S}^{d-1}} |(\mathbf{x},\mathbf{y})|^\alpha \;\left(1 - i\,\mathrm{sign}((\mathbf{x},\mathbf{y}))\;\tan(\pi\alpha/2)\right)\Gamma(d\mathbf{y}) + i\,(\mathbf{x},\boldsymbol{\mu})\right\} & \alpha \neq 1\,, \\ \exp\left\{-\int_{\mathbb{S}^{d-1}} |(\mathbf{x},\mathbf{y})|\;\left(1 + i\,\frac{2}{\pi}\mathrm{sign}((\mathbf{x},\mathbf{y}))\;\log|(\mathbf{x},\mathbf{y})|\right)\Gamma(d\mathbf{y}) + i\,(\mathbf{x},\boldsymbol{\mu})\right\} & \alpha = 1\,, \end{cases}$$

see Samorodnitsky and Taqqu [27], Theorem 2.3.1. The *index of stability* $\alpha \in (0,2)$, the *spectral measure* Γ on the unit sphere $\mathbb{S}^{d-1}$ and the location parameter $\boldsymbol{\mu}$ uniquely determine the distribution of an infinite variance α-stable random vector $\mathbf{X}$.

Multivariate regular variation. If $\mathbf{X}$ is α-stable for some $\alpha \in (0,2)$, it is regularly varying with index α. This means the following. The random vector $\mathbf{X}$ with values in $\mathbb{R}^d$ is *regularly varying with index* $\alpha \geq 0$ if there exists a random vector $\boldsymbol{\Theta}$ with values in the unit sphere $\mathbb{S}^{d-1}$ of $\mathbb{R}^d$ such that for any $t > 0$, as $x \to \infty$,

$$\frac{P\left(|\mathbf{X}| > tx\,, \widetilde{\mathbf{X}} \in \cdot\right)}{P(|\mathbf{X}| > x)} \stackrel{v}{\to} t^{-\alpha}\, P(\boldsymbol{\Theta} \in \cdot)\,, \tag{2.1}$$

where for any vector $\mathbf{x} \neq \mathbf{0}$,

$$\widetilde{\mathbf{x}} = \mathbf{x}/|\mathbf{x}|\,,$$

and $\stackrel{v}{\to}$ denotes *vague convergence* in the Borel σ-field of $\mathbb{S}^{d-1}$; see Resnick [23, 24] for its definition and details. The distribution of $\boldsymbol{\Theta}$ is called the *spectral measure* of $\mathbf{X}$. Alternatively, (2.1) is equivalent to

$$\frac{P(\mathbf{X} \in x\cdot)}{P(|\mathbf{X}| > x)} \stackrel{v}{\to} \mu\,,$$

where $\stackrel{v}{\to}$ denotes vague convergence in the Borel σ-field of $\overline{\mathbb{R}}^d \backslash \{\mathbf{0}\}$ and μ is a measure on the same σ-field satisfying the homogeneity assumption $\mu(tA) = t^{-\alpha}\mu(A)$ for $t > 0$.

Stable limits for sums of iid random vectors. Now let $(\mathbf{Y}_t)$ be an iid sequence of random vectors with values in $\mathbb{R}^d$. According to Rvačeva [26], there exist sequences of constants $a_n > 0$ and $\mathbf{b}_n \in \mathbb{R}^d$ such that

$$a_n^{-1} \sum_{t=1}^n \mathbf{Y}_t - \mathbf{b}_n \stackrel{d}{\to} \mathbf{X}_\alpha$$

for some α-stable random variable $\mathbf{X}_\alpha$ with $\alpha \in (0,2)$ if and only if $\mathbf{Y}_1$ is regularly varying with index α, and the normalizing constants a_n can be chosen as

$$P(|\mathbf{Y}_1| > a_n) \sim n^{-1}\,. \tag{2.2}$$

For a stationary sequence $(\mathbf{Y}_t)$ a similar result can be found in Davis and Mikosch [12] as a multivariate extension of one-dimensional results in Davis and Hsing [11]. For its formulation one needs regular variation of the summands and a particular mixing condition, called $\mathcal{A}(a_n)$ which was introduced in Davis and Hsing [11].

Mixing conditions. We say that the *condition* $\mathcal{A}(a_n)$ holds for the stationary sequence $(\mathbf{Y}_t)$ of random vectors with values in $\mathbb{R}^d$ if there exists a sequence of positive integers r_n such that $r_n \to \infty$, $k_n = [n/r_n] \to \infty$ as $n \to \infty$ and

$$E \exp\left\{-\sum_{t=1}^{n} f(\mathbf{Y}_t/a_n)\right\} - \left(E \exp\left\{-\sum_{t=1}^{r_n} f(\mathbf{Y}_t/a_n)\right\}\right)^{k_n} \quad \to \quad 0\,, \qquad n \to \infty\,, \quad \forall f \in \mathcal{G}_s\,, \tag{2.3}$$

where $\mathcal{G}_s$ is the collection of bounded non-negative step functions on $\overline{\mathbb{R}}^d \backslash \{\mathbf{0}\}$. The convergence in (2.3) is not required to be uniform in f. This is indeed a very weak condition and is implied by many known mixing conditions, in particular the strong mixing condition which is relevant in the context of GARCH processes; see Section 4. We refer to Davis and Mikosch [12] for a comparison of $\mathcal{A}(a_n)$ with other mixing conditions.

For later use we also recall the definition of a *strongly mixing stationary sequence* $(\mathbf{Y}_t)$ of random vectors with *rate function* (ϕ_k), see Rosenblatt [25], cf. Doukhan [13] or Ibragimov and Linnik [16]:

$$\sup_{A \in \sigma(\mathbf{Y}_s\,,\, s \le 0)\,,\, B \in \sigma(\mathbf{Y}_s\,,\, s > k)} |P(A \cap B) - P(A)\,P(B)| \;=:\; \phi_k \to 0 \quad \text{as } k \to \infty.$$

If (ϕ_k) decays to zero at an exponential rate then $(\mathbf{Y}_t)$ is said to be *strongly mixing with geometric rate.*

Recall that *absolute regularity* (or *β–mixing*) is a mixing notion which is slightly more restrictive than strong mixing:

$$E\left(\sup_{B \in \sigma(\mathbf{Y}_t\,,\, t > k)} |P(B \mid \sigma(\mathbf{Y}_s\,,\, s \le 0)) - P(B)|\right) =: \; b_k \to 0, \qquad k \to \infty. \tag{2.4}$$

Indeed, β-mixing implies strong mixing with the same rate function.

Stable limits for sums of stationary random variables. The following result is a combination of Theorem 2.8 and Proposition 3.3 in [12]. It gives conditions under which an α-stable weak limit occurs for the sum process of a stationary sequence. In what follows, we write

$$\mathbf{S}_0 = \mathbf{0} \quad \text{and} \quad \mathbf{S}_n = \mathbf{Y}_1 + \cdots + \mathbf{Y}_n\,, \quad n \ge 1\,,$$

and for any Borel set $B \subset \mathbb{R}$,

$$\mathbf{S}_n B = (S_n^{(h)}(B))_{h=1,\ldots,d}\,,$$

where

$$S_n^{(h)}(B) = \sum_{t=1}^{n} Y_t^{(h)} I_B(|Y_t^{(h)}|/a_n)\,, \quad n \ge 1\,.$$

Theorem 2.1. *Let $(\mathbf{Y}_t)$ be a strictly stationary sequence of random vectors with values in $\mathbb{R}^d$ and the real sequence (a_n) be defined by (2.2). Assume that the following conditions are satisfied:*

(a) *The finite-dimensional distributions of $(\mathbf{Y}_k)$ are regularly varying with index $\alpha > 0$. To be specific, let $\operatorname{vec}(\theta_{-k}^{(k)}, \ldots, \theta_k^{(k)})$ be the $(2k+1)d$-dimensional random row vector with values in the unit sphere $\mathbb{S}^{(2k+1)d-1}$ that appears in the definition (2.1) of regular variation of $\operatorname{vec}(\mathbf{Y}_{-k}, \ldots, \mathbf{Y}_k)$, $k \ge 0$, with respect to the max-norm $|\cdot|$ in $\mathbb{R}^{(2k+1)d}$.*

(b) *The mixing condition $\mathcal{A}(a_n)$ holds for $(\mathbf{Y}_t)$.*

(c)

$$\lim_{k\to\infty}\limsup_{n\to\infty} P\left(\bigvee_{k\le |t|\le r_n} |\mathbf{Y}_t| > a_n y \,\middle|\, |\mathbf{Y}_0| > a_n y\right) = 0\,, \quad y>0\,, \tag{2.5}$$

where (r_n) appears in the formulation of $\mathcal{A}(a_n)$.

Then the limit

$$\gamma = \lim_{k\to\infty} E\left(|\theta_0^{(k)}|^\alpha - \bigvee_{j=1}^k |\theta_j^{(k)}|^\alpha\right)_+ \Big/ \, E|\theta_0^{(k)}|^\alpha \tag{2.6}$$

exists. If $\gamma>0$, then the following results hold.

(i) *If $\alpha\in(0,1)$, then*

$$a_n^{-1}\mathbf{S}_n \xrightarrow{d} \mathbf{X}_\alpha\,,$$

for some α-stable random vector $\mathbf{X}_\alpha$.

(ii) *If $\alpha\in[1,2)$ and for all $\delta>0$,*

$$\lim_{y\to 0}\limsup_{n\to\infty} P\left(|\mathbf{S}_n(0,y] - E\mathbf{S}_n(0,y]| > \delta\, a_n\right) = 0\,, \tag{2.7}$$

then

$$a_n^{-1}\left(\mathbf{S}_n - E\mathbf{S}_n(0,1]\right) \xrightarrow{d} \mathbf{X}_\alpha\,,$$

for some α-stable random vector $\mathbf{X}_\alpha$.

The structure of the limiting vectors $\mathbf{X}_\alpha$ is given by some functional of the points of a limiting point process. The proof of this result makes heavily use of point process convergence results which are appropriate tools in the context of regularly varying distributions when extremely large values may occur in the sequence $(\mathbf{Y}_t)$; see Davis and Mikosch [12] for details.

3. Stable limits for martingale transform

In this section we want to derive infinite variance stable limits for sums of strictly stationary random vectors which have the particular form

$$\mathbf{Y}_t = \mathbf{G}_t\, Y_t\,,$$

where (Y_t) is an iid sequence and $(\mathbf{G}_t)$ is a strictly stationary sequence of random vectors with values in $\mathbb{R}^d$ such that $(\mathbf{G}_t)$ is adapted to the filtration given by the σ-fields $\mathcal{F}_t = \sigma(Y_{t-1}, Y_{t-2},\ldots)$, $t\in\mathbb{Z}$. If $EY_1 = 0$ and $E|\mathbf{G}_1|<\infty$, $E(\mathbf{G}_t Y_t|\mathcal{F}_t) = \mathbf{0}$ a.s., and therefore $(\mathbf{G}_t\, Y_t)$ is a martingale difference sequence and

$$\mathbf{S}_0 = \mathbf{0}, \quad \mathbf{S}_n = \mathbf{Y}_1 + \cdots + \mathbf{Y}_n, \qquad n\ge 1,$$

is the *martingale transform* of the martingale $(\sum_{t=1}^n Y_t)_{n\ge 0}$ by the sequence $(\mathbf{G}_t)$. We keep this name even if $E|\mathbf{Y}_1| = \infty$.

3.1. **Basic assumptions.** We impose the following assumptions on the sequences (Y_t) and $(\mathbf{G}_t)$.

A.1 Y_1 is regularly varying with index $\alpha\in(0,2)$.

A.2 $E|\mathbf{G}_1|^{\alpha+\epsilon}<\infty$ for some $\epsilon>0$.

A.3 $(\mathbf{G}_t Y_t)$ satisfies condition $\mathcal{A}(a_n)$, see (2.3), where $P(|Y_1|>a_n)\sim n^{-1}$ and (r_n), defined in (2.3), is such that

$$n\, r_n \left(\frac{a_{r_n}}{a_n}\right)^{\alpha+\epsilon} \to 0\,, \tag{3.1}$$

where ϵ is the same as in **A.2**.

Remark 3.1. Regular variation of Y_1 with index α and the iid property of (Y_t) imply that

$$P\left(a_n^{-1}\max_{1\le t\le n}|Y_t|\le x\right)\to\Phi_\alpha(x)=\mathrm{e}^{-x^{-\alpha}}\,,\quad x>0\,,$$

for the Fréchet distribution Φ_α; see Embrechts et al. [14], Chapter 3.

In this setting, the heaviness of the tails of the distribution of $\mathbf{G}_1Y_1$ is essentially determined by the distribution of Y_1; see Remark 3.3 below.

3.2. **Main result.** We are now ready to formulate our main result on the asymptotic behavior of the sum process $(\mathbf{S}_n)$.

Theorem 3.2. *Consider the martingale transform* $(\sum_{t=1}^n\mathbf{Y}_t)_{n\ge0}=(\sum_{t=1}^n\mathbf{G}_tY_t)_{n\ge0}$ *defined above. Assume that the conditions* **A.1-A.3** *are satisfied. Moreover, if* $\alpha\in(1,2)$ *assume that* $EY_1=0$ *and, if* $\alpha=1$*, that* Y_1 *is symmetric. Then the finite-dimensional distributions of* $(\mathbf{Y}_t)$ *are regularly varying with index* α *and the limit* γ *in* (2.6) *exists. If* $\gamma>0$*, then*

$$a_n^{-1}\mathbf{S}_n\stackrel{d}{\to}\mathbf{X}_\alpha\,,\tag{3.2}$$

where the sequence (a_n) *is given by*

$$P(|Y_1|>a_n)\sim n^{-1}\,.$$

and $\mathbf{X}_\alpha$ *is an* α*-stable random vector.*

Remark 3.3. It is not difficult to see that $\mathbf{Y}_t$ is regularly varying with index α. For the proof we need a result of Breiman [10]. It says that if one has two independent random variables $\xi,\eta>0$ a.s., ξ is regularly varying with index $\alpha>0$ and $E\eta^\nu<\infty$ for some $\nu>\alpha$, then

$$P(\xi\,\eta>x)\sim E\eta^\alpha\,P(\xi>x)\,,$$

i.e., $\xi\eta$ is regularly varying with the same index α. Now observe that for $t,x>0$ and a Borel set $S\subset\mathbb{S}^{d-1}$, by multiple application of Breiman's result,

$$\begin{aligned}
&\frac{P\left(|\mathbf{G}_1|\,|Y_1|>tx\,,\dfrac{\mathbf{G}_1\,Y_1}{|\mathbf{G}_1|\,|Y_1|}\in S\right)}{P(|\mathbf{G}_1|\,|Y_1|>x)}\\
&=\frac{P\left(|\mathbf{G}_1|\,|Y_1|>tx\,,\mathrm{sign}(Y_1)\,\widetilde{\mathbf{G}}_1\in S\right)}{P(|\mathbf{G}_1|\,|Y_1|>x)}\\
&=\frac{P\left(|\mathbf{G}_1|\,Y_1>tx\,,\widetilde{\mathbf{G}}_1\in S\right)}{P(|\mathbf{G}_1|\,|Y_1|>x)}+\frac{P\left(|\mathbf{G}_1|\,Y_1<-tx\,,-\widetilde{\mathbf{G}}_1\in S\right)}{P(|\mathbf{G}_1|\,|Y_1|>x)}\\
&\sim\frac{E\left(|\mathbf{G}_1|^\alpha\,I_S\left(\widetilde{\mathbf{G}}_1\right)\right)\,P(Y_1>tx)}{E|\mathbf{G}_1|^\alpha\,P(|Y_1|>x)}+\frac{E\left(|\mathbf{G}_1|^\alpha\,I_S\left(-\widetilde{\mathbf{G}}_1\right)\right)\,P(Y_1\le-tx)}{E|\mathbf{G}_1|^\alpha\,P(|Y_1|>x)}\,.
\end{aligned}$$

Writing for some $p,q\ge0$ with $p+q=1$ and a slowly varying function $L(x)$,

$$P(Y_1>x)=p\,L(x)\,x^{-\alpha}\quad\text{and}\quad P(Y_1\le-x)=q\,L(x)\,|x|^{-\alpha}\,,\quad x>0\,,$$

we can read off the spectral measure of the vector $\mathbf{Y}_1$:

$$P(\mathbf{\Theta}\in S)=p\,\frac{E\left(|\mathbf{G}_1|^\alpha\,I_S\left(\widetilde{\mathbf{G}}_1\right)\right)}{E|\mathbf{G}_1|^\alpha}+q\,\frac{E\left(|\mathbf{G}_1|^\alpha\,I_S\left(-\widetilde{\mathbf{G}}_1\right)\right)}{E|\mathbf{G}_1|^\alpha}\,.\tag{3.3}$$

By regular variation, $a_n = n^{1/\alpha}\ell(n)$ for some slowly varying function ℓ. By Breiman's result and since $E|\mathbf{G}_1|^{\alpha+\epsilon} < \infty$ for some $\epsilon > 0$, it also follows that

$$P(|\mathbf{G}_1|\,|Y_1| > x) \sim E|\mathbf{G}_1|^\alpha\, P(|Y_1| > x)\,,$$

and therefore $P(|\mathbf{Y}_1| > c\, a_n) \sim n^{-1}$ for some constant $c > 0$. Moreover, we have

$$n\, P(a_n^{-1}\mathbf{Y}_1 \in \cdot) \xrightarrow{v} \mu_1\,, \tag{3.4}$$

for some measure μ_1 on $\overline{\mathbb{R}}^d\backslash\{\mathbf{0}\}$ which is determined by α and the spectral measure.

Remark 3.4. It follows from the proof below that

$$\begin{aligned} (3.5)\qquad & n\, P(a_n^{-1}\,(\mathbf{Y}_1,\ldots,\mathbf{Y}_h) \in d(\mathbf{x}_1,\ldots,\mathbf{x}_h)) \\ &\xrightarrow{v}\ \mu_1(d\mathbf{x}_1)\,\varepsilon_{\mathbf{0}}(d(\mathbf{x}_2,\ldots,\mathbf{x}_h)) + \cdots + \mu_1(d\mathbf{x}_h)\,\varepsilon_{\mathbf{0}}(d(\mathbf{x}_1,\ldots,\mathbf{x}_{h-1})) \\ &=:\ \mu_h(d(\mathbf{x}_1,\ldots,\mathbf{x}_h))\,. \end{aligned}$$

where μ_1 is defined by (3.4), $\varepsilon_{\mathbf{0}}$ is Dirac measure at $\mathbf{0}$ and

$$(\mathbf{Y}_1,\ldots,\mathbf{Y}_h) := \mathrm{vec}(\mathbf{Y}_1,\ldots,\mathbf{Y}_h) \quad\text{and}\quad (\mathbf{x}_1,\ldots,\mathbf{x}_h) := \mathrm{vec}(\mathbf{x}_1,\ldots,\mathbf{x}_h)\,. \tag{3.6}$$

This means in particular that the limiting measure in the definition of regular variation for $(\mathbf{Y}_1,\ldots,\mathbf{Y}_h)$ is the same as in the definition of regular variation for $\mathrm{vec}(\mathbf{Y}_1',\ldots,\mathbf{Y}_h')$, where $\mathbf{Y}_i'$ are iid copies of $\mathbf{Y}_1$. This part of the theorem is valid for any $\alpha > 0$.

Proof. We verify the conditions of Theorem 2.1. Since **A.3** implies $\mathcal{A}(a_n)$ and since we require $\gamma > 0$, it remains to check (a) and (c) in Theorem 2.1.

(a) *Regular variation of the finite-dimensional distributions.* We show regular variation of the vector $(\mathbf{Y}_1,\ldots,\mathbf{Y}_h)$ defined in (3.6), i.e., we show that (3.5) holds.

We restrict ourselves to prove regular variation of the pairs $(\mathbf{Y}_1,\mathbf{Y}_2) := \mathrm{vec}(\mathbf{Y}_1,\mathbf{Y}_2)$; the case of general finite-dimensional distributions is completely analogous. The regular variation of $\mathbf{Y}_1$ was explained in Remark 3.3. Let now B_1 and B_2 be two Borel sets in $[0,\infty]^d\backslash\{\mathbf{0}\}$, bounded away from zero. In particular, there exists $M > 0$ such that $|\mathbf{x}| > M$ for all $\mathbf{x} \in B_1$ and $\mathbf{x} \in B_2$. Then for any $\epsilon > 0$,

$$\begin{aligned} & \left\{a_n^{-1}\mathbf{Y}_1 \in B_1\,,\, a_n^{-1}\mathbf{Y}_2 \in B_2\right\} \\ \subset\ & \{|\mathbf{G}_1|\ |Y_1| > M\, a_n\,, |\mathbf{G}_2|\ |Y_2| > M\, a_n\} \\ \subset\ & \{\epsilon\,|Y_1| > M\, a_n\,, \epsilon\,|Y_2| > M\, a_n\} \\ & \cup\left\{|\mathbf{G}_1|\, I_{(\epsilon,\infty)}(|\mathbf{G}_1|)\,|Y_1| > M\, a_n\,, \epsilon\,|Y_2| > M\, a_n\right\} \\ & \cup\left\{|\mathbf{G}_2|\, I_{(\epsilon,\infty)}(|\mathbf{G}_2|)\,|Y_2| > M\, a_n\,, \epsilon\,|Y_1| > M\, a_n\right\} \\ & \cup\left\{|\mathbf{G}_1|\, I_{(\epsilon,\infty)}(|\mathbf{G}_1|)\,|Y_1| > M\, a_n\,, |\mathbf{G}_2|\, I_{(\epsilon,\infty)}(|\mathbf{G}_2|)\,|Y_2| > M\, a_n\right\} =: \bigcup_{i=1}^4 D_i\,. \end{aligned}$$

By independence and an application of Breiman's result, $nP(D_1) \to 0$ and $nP(D_2) \to 0$. Similarly,

$$\begin{aligned} n\, P(D_3) &\le n\, P\left(|\mathbf{G}_2|\, I_{(\epsilon,\infty)}(|\mathbf{G}_2|)\,|Y_2| > M a_n\right) \\ &\sim n\, P(|Y_2| > M\, a_n)\, E\left(|\mathbf{G}_2|^\alpha\, I_{(\epsilon,\infty)}(|\mathbf{G}_2|)\right)\,, \end{aligned}$$

thus, by Lebesgue's dominated convergence theorem,

$$\lim_{\epsilon\uparrow\infty}\limsup_{n\to\infty} n\, P(D_3) = 0\,,$$

and $nP(D_4) \to 0$ can be proved in the same way. We conclude that

$$n\,P\left(a_n^{-1}(\mathbf{Y}_1,\mathbf{Y}_2)\in d(\mathbf{x}_1,\mathbf{x}_2)\right) \stackrel{v}{\to} \mu_1(d\mathbf{x}_1)\,\varepsilon_{\mathbf{0}}(d\mathbf{x}_1)+\mu_1(d\mathbf{x}_2)\,\varepsilon_{\mathbf{0}}(d\mathbf{x}_2)=\mu_2(d(\mathbf{x}_1,\mathbf{x}_2))\,,$$

see Resnick [24]. This proves the regular variation of the 2-dimensional finite-dimensional distributions. The higher-dimensional case is completely analogous.

(c) *The condition* (2.5). We have for any $y>0$,

$$\begin{aligned}
& P\left(\max_{k\le t\le r_n}|\mathbf{G}_t|\,|Y_t|>ya_n\,\middle|\,|\mathbf{G}_0|\,|Y_0|>y\,a_n\right)\\
\le\;& P\left(\max_{k\le t\le r_n}|\mathbf{G}_t|>y\,a_n/(s_k\,a_{r_n})\,\middle|\,|\mathbf{G}_0|\,|Y_0|>y\,a_n\right)+P\left(\max_{k\le t\le r_n}|Y_t|>s_k\,a_{r_n}\right)\\
=:\;& I_1+I_2\,,
\end{aligned}$$

where (s_k) is any sequence such that $s_k\to\infty$. In what follows, all calculations go through for any $y>0$; for ease of notation we set $y=1$. Then, by Remark 3.1,

$$\lim_{k\to\infty}\lim_{n\to\infty} I_2=\lim_{k\to\infty}(1-\Phi_\alpha(s_k))=0\,.$$

An application of Markov's inequality yields for some constant $c>0$ and $\epsilon>0$ as in **A.2** (here and in what follows, c denotes any positive constant whose value is not of interest),

$$\begin{aligned}
I_1 &\le \sum_{t=k}^{r_n} P\left(|\mathbf{G}_t|>a_n/(s_k a_{r_n})\mid |\mathbf{G}_0|\,|Y_0|>a_n\right)\\
&\le \left(\frac{s_k a_{r_n}}{a_n}\right)^{\alpha+\epsilon}\sum_{t=k}^{r_n}\frac{E[|\mathbf{G}_t|^{\alpha+\epsilon} I_{\{|\mathbf{G}_0|\,|Y_0|>a_n/(s_k a_{r_n})\}}]}{P(|\mathbf{G}_0|\,|Y_0|>a_n)}\\
&\le c\,n\,r_n\left(\frac{s_k a_{r_n}}{a_n}\right)^{\alpha+\epsilon} E\,|\mathbf{G}_0|^{\alpha+\epsilon}\\
&\to 0 \quad\text{as } n\to\infty.
\end{aligned}$$

Here we used Breiman's result [10] to show that

$$P(|\mathbf{G}_0|\,|Y_0|>a_n)\sim E\,|\mathbf{G}_0|^\alpha\;P(|Y_0|>a_n)\,,$$

condition (3.1) and the fact that $E|\mathbf{G}_1|^{\alpha+\epsilon}<\infty$; see **A.2**.

Now we turn to

$$\begin{aligned}
& P\left(\max_{-r_n\le t\le -k}|\mathbf{G}_t|\,|Y_t|>a_n\,\middle|\,|\mathbf{G}_0|\,|Y_0|>a_n\right)\\
\le\;& P\left(\max_{-r_n\le t\le -k}|\mathbf{G}_t|>a_n/(s_k a_{r_n})\,\middle|\,|\mathbf{G}_0|\,|Y_0|>a_n\right)\\
& +P\left(\max_{-r_n\le t\le -k}|Y_t|>s_k a_{r_n}\,\middle|\,|\mathbf{G}_0|\,|Y_0|>a_n\right)\\
=:\;& I_3+I_4\,.
\end{aligned}$$

The quantity I_3 can be treated in the same way as I_1 to show that $I_3 \to 0$ a.s. as $n \to \infty$. We turn to I_4. Fix $0 < M < \infty$. Then

$$\begin{aligned} I_4 &\le \frac{P\left(\max_{-r_n \le t \le -k} |Y_t| > s_k a_{r_n}\,, M\,|Y_0| > a_n\right)}{P(|\mathbf{G}_0|\,|Y_0| > a_n)} \\ &\quad + \frac{P\left(|\mathbf{G}_0|\, I_{(M,\infty)}(|\mathbf{G}_0|)\,|Y_0| > a_n\right)}{P(|\mathbf{G}_0|\,|Y_0| > a_n)} \\ &=: I_{41} + I_{42}\,. \end{aligned}$$

By independence of the Y_i's, Breiman's [10] result and since $r_n \to \infty$,

$$\begin{aligned} I_{41} &\sim \frac{P\left(\max_{-r_n \le t \le -k} |Y_t| > s_k a_{r_n}\right)\, M^\alpha\, P(|Y_0| > a_n)}{E|\mathbf{G}_0|^\alpha\, P(|Y_0| > a_n)} \\ &\sim c\,(1 - \Phi_\alpha(s_k)) \quad \text{as } n \to \infty \\ &\to 0 \quad \text{as } k \to \infty. \end{aligned}$$

By virtue of Breiman's [10] result,

$$I_{42} \sim \frac{E\left(|\mathbf{G}_0|^\alpha\, I_{(M,\infty)}(|\mathbf{G}_0|)\right)\, P(|Y_0| > a_n)}{E|\mathbf{G}_0|^\alpha\, P(|Y_0| > a_n)}\,.$$

Since $|\mathbf{G}_0|$ has finite moments of order greater than α, an application of the Lebesgue dominated convergence theorem yields

$$\lim_{M\to\infty} \lim_{n\to\infty} I_{42} = 0\,.$$

This proves (2.5). □

Thus the conditions (a)-(c) and $\gamma > 0$ of Theorem 2.1 are satisfied. In the case $\alpha < 1$, Theorem 2.1 immediately yields (3.2). In the case $\alpha \in [1,2)$ we have to check condition (2.7). It suffices to show it for the components $\mathbf{S}_n^{(i)}(0,y]$, $i = 1,\ldots,d$, of $\mathbf{S}_n(0,y]$. Since the components can be handled in the same way, we suppress the dependence on i and, for the ease of notation, write $G_t Y_t$ for the summands of the ith component.

We start with the case $\alpha \in (1,2)$. As before, write $\mathcal{F}_t = \sigma(Y_{t-1}, Y_{t-2},\ldots)$. Then, for $z > 0$, since $EY_1 = 0$,

$$\begin{aligned} E[G_t\, Y_t\, I_{(0,z]}(|G_t\, Y_t|/a_n) \mid \mathcal{F}_t] &= G_t\, E[Y_t\, I_{(0,z]}(|G_t\, Y_t|/a_n) \mid G_t] \\ &= -G_t\, E[Y_t\, I_{(z,\infty)}(|G_t\, Y_t|/a_n) \mid G_t]\,. \end{aligned}$$

Consider the decomposition

$$\begin{aligned} & a_n^{-1} \sum_{t=1}^n \left[G_t\, Y_t\, I_{(0,z]}(|G_t\, Y_t|/a_n) - E[G_1\, Y_1\, I_{(0,z]}(|G_1\, Y_1|/a_n)]\right] \\ = \;& a_n^{-1} \sum_{t=1}^n \left[G_t\, Y_t\, I_{(0,z]}(|G_t\, Y_t|/a_n) - G_t\, E[Y_t\, I_{(0,z]}(|G_t\, Y_t|/a_n) \mid G_t]\right] \\ & - a_n^{-1} \sum_{t=1}^n \left[G_t\, E[Y_t\, I_{(z,\infty)}(|G_t\, Y_t|/a_n) \mid G_t] - E[G_1\, Y_1 I_{(z,\infty)}(|G_1\, Y_1|/a_n)]\right] =: T_1 + T_2\,. \end{aligned}$$

For fixed n, T_1 is a sum of stationary mean zero martingale differences. An application of Karamata's theorem to the regularly varying random variable $G_1 Y_1$ with index α yields for some constant

$c > 0$,

$$\begin{aligned}
\operatorname{var}(T_1) &= n\,a_n^{-2}\,E\left[G_1\,Y_1\,I_{(0,z]}(|G_1\,Y_1|/a_n) - G_1 E[Y_1\,I_{(0,z]}(|G_1\,Y_1|/a_n) \mid G_1]\right]^2 \\
&\le c\,n\,a_n^{-2}\,E\left[G_1\,Y_1\,I_{(0,z]}(|G_1\,Y_1|/a_n)\right]^2 \\
&\sim c\,z^{2-\alpha} \quad \text{as } n \to \infty \\
&\to 0 \quad \text{as } z \downarrow 0. \qquad (3.7)
\end{aligned}$$

Next we treat T_2. Fix $0 < \delta < M < \infty$ to be chosen later. Notice that by Karamata's theorem and the uniform convergence theorem for regularly varying functions uniformly for $c \in [\delta, M]$,

$$\frac{E[Y_1\,I_{(cx,\infty)}(|Y_1|)]}{cx\,P(|Y_1| > cx)} \to C$$

for some constant C. Taking this into account, the strong law of large numbers yields, with probability 1,

$$\begin{aligned}
& a_n^{-1}\sum_{t=1}^{n} G_t I_{[\delta,M]}(|G_t|)\,E[Y_t\,I_{(z,\infty)}(|G_t\,Y_t|/a_n) \mid G_t] \qquad (3.8)\\
&= a_n^{-1}\sum_{t=1}^{n} G_t I_{[\delta,M]}(|G_t|)\;\left[(za_n/G_t)\,P(|Y_t| > za_n/|G_t| \mid G_t)\,(C + o(1))\right] \\
&= (C + o(1))\,z^{1-\alpha}\,n^{-1}\sum_{t=1}^{n} |G_t|^{\alpha}\,I_{[\delta,M]}(|G_t|) \\
&\to C\,z^{1-\alpha}\,E[|G_1|^{\alpha} I_{[\delta,M]}(|G_1|)]\,.
\end{aligned}$$

On the other hand, since $G_1 I_{[\delta,M]}(|G_1|)\,Y_1$ is regularly varying with index $\alpha \in (1, 2)$, by the same argument and Breiman's result,

$$\begin{aligned}
& n\,a_n^{-1}\,E[G_1 I_{[\delta,M]}(|G_1|)\,Y_1\,I_{(z,\infty)}(|G_1\,Y_1|/a_n)] \qquad (3.9)\\
&= n\,a_n^{-1}\;\left[(C + o(1))\,(z\,a_n)\,P(G_1 I_{[\delta,M]}(|G_1|)\,|Y_1| > za_n)\right] \\
&= (C + o(1))\,z^{1-\alpha}\,E[|G_1|^{\alpha} I_{[\delta,M]}(|G_1|)]\,.
\end{aligned}$$

This shows that (3.8) and (3.9) cancel asymptotically as $n \to \infty$ for every fixed z.

A similar argument shows that, with probability 1,

$$\begin{aligned}
& a_n^{-1}\left|\sum_{t=1}^{n} G_t I_{[0,\delta]}(|G_t|)\,E[Y_t\,I_{(z,\infty)}(|G_t\,Y_t|/a_n) \mid G_t]\right| \qquad (3.10)\\
&\le a_n^{-1}\sum_{t=1}^{n} |G_t| I_{[0,\delta]}(|G_t|)\,E[|Y_1| I_{(z,\infty)}(\delta\,|Y_1|/a_n)] \\
&\to c\,(z/\delta)^{1-\alpha}\,E[|G_1| I_{[0,\delta]}(|G_1|)]\,.
\end{aligned}$$

Moreover,

$$\begin{aligned}
& n\,a_n^{-1}\,|E[G_1 I_{[0,\delta]}(|G_1|)\,Y_1\,I_{(z,\infty)}(|G_1\,Y_1|/a_n)]| \qquad (3.11)\\
&\le n\,a_n^{-1} E[|G_1| I_{[0,\delta]}(|G_1|)\,|Y_1|\,I_{(z,\infty)}(\delta\,|Y_1|/a_n)] \\
&\sim c\,(z/\delta)^{1-\alpha}\,E[|G_1| I_{[0,\delta]}(|G_1|)]\,.
\end{aligned}$$

Now choose $\delta = z^2$. Then, first letting $n \to \infty$ and then $z \downarrow 0$, both (3.10) and (3.11) vanish asymptotically.

Finally, we consider

$$\begin{aligned} & a_n^{-1} E \left| \sum_{t=1}^{n} G_t I_{(M,\infty)}(|G_t|) \, E[Y_t \, I_{(z,\infty)}(|G_t \, Y_t|/a_n) \mid G_t] \right| \\ \leq \quad & a_n^{-1} \, n \, E[|G_1| I_{(M,\infty)}(|G_1|) \, |Y_1| \, I_{(z,\infty)}(|G_1 \, Y_1|/a_n)] \, . \end{aligned}$$

An application of Breiman's result to the regularly varying random variable $G_1 I_{[M,\infty)}(|G_1|) Y_1$ gives that the right-hand side is asymptotically equivalent as $n \to \infty$ to

$$c \, z^{1-\alpha} \, E[|G_1|^\alpha I_{[M,\infty)}(|G_1|)] \, .$$

Choosing M large enough, the right-hand side is smaller than z, say. The same argument can be applied to

$$n \, a_n^{-1} \, |E[G_1 I_{[M,\infty)}(|G_1|) \, Y_1 \, I_{(z,\infty)}(|G_1 \, Y_1|/a_n)]| \, .$$

Collecting the bounds above, we see that

$$\lim_{z \downarrow 0} \limsup_{n \to \infty} P(|T_2| > r) = 0 \, , \quad r > 0 \, .$$

This together with (3.7) concludes the proof of (2.7) for $\alpha \in (1,2)$.

For $\alpha = 1$ we use the additional condition of symmetry of Y_t. Then $E\mathbf{S}_n(0,y] = 0$ and the same argument as for $\mathrm{var}(T_1)$ above shows that (2.7) holds in this case as well. This concludes the proof of (2.7).

Since the conditions of Theorem 2.1 are satisfied for $\alpha \in [1,2)$ we conclude that

$$a_n^{-1}(\mathbf{S}_n - E\mathbf{S}_n(0,1]) \xrightarrow{d} \mathbf{X}_\alpha$$

for some α-stable random vector in $\mathbb{R}^d$. For $\alpha = 1$ we can drop $E\mathbf{S}_n(0,y]$ because of the symmetry of $\mathbf{G}_t Y_t$. For $\alpha \in (1,2)$, $\mathbf{G}_t Y_t$ is regularly varying with index α. Since $E(\mathbf{G}_t Y_t) = 0$, Karamata's theorem yields

$$a_n^{-1} \, E\mathbf{S}_n(0,1] \to \mathbf{b}$$

for some constant $\mathbf{b}$ which can be incorporated in the stable limit, and therefore centering in (3.2) can be avoided. This concludes the proof of Theorem 3.2. □

Remark 3.5. If the roles of $\mathbf{G}_1$ and Y_1 are interchanged in the sense that $\mathbf{G}_1$ is regularly varying with index α and $E|Y_1|^{\alpha+\epsilon} < \infty$ for some $\epsilon > 0$, $\mathbf{Y}_1$ is regularly varying with index α, as the following calculations show. Observe that the random variables $|\mathbf{G}_1| I_S(\widetilde{\mathbf{G}}_1)$ and $|\mathbf{G}_1| I_S(-\widetilde{\mathbf{G}}_1)$ are regularly varying for any Borel set $S \subset \mathbb{S}^{d-1}$ for which $S, -S$ are continuity sets with respect to the spectral measure $P(\mathbf{\Theta}_G \in \cdot)$ of $\mathbf{G}_1$ and which satisfy $P(\mathbf{\Theta}_G \in \pm S) > 0$. For such a set S, any

$t, x > 0$, by multiple application of Breiman's result,

$$\frac{P\left(|\mathbf{G}_1|\,|Y_1| > tx\,, \dfrac{\mathbf{G}_1\,Y_1}{|\mathbf{G}_1|\,|Y_1|} \in S\right)}{P(|\mathbf{G}_1|\,|Y_1| > x)}$$

$$= \frac{P\left(|\mathbf{G}_1|\,|Y_1| > tx\,, \operatorname{sign}(Y_1)\,\widetilde{\mathbf{G}}_1 \in S\right)}{P(|\mathbf{G}_1|\,|Y_1| > x)}$$

$$= \frac{P\left(|\mathbf{G}_1|\,Y_1 > tx\,, \widetilde{\mathbf{G}}_1 \in S\right)}{P(|\mathbf{G}_1|\,|Y_1| > x)} + \frac{P\left(|\mathbf{G}_1|\,Y_1 < -tx\,, -\widetilde{\mathbf{G}}_1 \in S\right)}{P(|\mathbf{G}_1|\,|Y_1| > x)}$$

$$\sim \frac{E[Y_1^\alpha I_{(0,\infty)}(Y_1)]\,P\left(|\mathbf{G}_1| > tx\,, \widetilde{\mathbf{G}}_1 \in S\right)}{E|Y_1|^\alpha\,P(|\mathbf{G}_1| > x)} + \frac{E[|Y_1|^\alpha I_{(-\infty,0)}(Y_1)]\,P(\left(|\mathbf{G}_1| > tx\,, \widetilde{\mathbf{G}}_1 \in -S\right)}{E|Y_1|^\alpha\,P(|\mathbf{G}_1| > x)}$$

$$x \to \infty\,.$$

Letting $x \to \infty$, one can read off the spectral measure of $\mathbf{Y}_1$:

$$P(\mathbf{\Theta} \in S) = \frac{E[Y_1^\alpha I_{(0,\infty)}(Y_1)]}{E|Y_1|^\alpha}\,P(\mathbf{\Theta}_G \in S) + \frac{E[|Y_1|^\alpha I_{(-\infty,0)}(Y_1)]}{E|Y_1|^\alpha}\,P(\mathbf{\Theta}_G \in -S)\,. \tag{3.12}$$

The spectral measure of $\mathbf{Y}_1$ under the A-conditions, see (3.3), is completely different from (3.12). One might hope that a result similar to Theorem 3.2 can be derived simply under the conditions that $\mathbf{G}_1$ is regularly varying with index $\alpha < 2$ and $E|Y_1|^{\alpha+\epsilon} < \infty$. However, it is not clear whether $(\mathbf{Y}_t)$ has regularly varying finite-dimensional distributions, and it is not clear how to verify condition (2.5). Therefore one cannot expect to derive a result as general as Theorem 3.2 without additonal conditions on the sequences $(\mathbf{G}_t)$ and (Y_t). Examples of limit results when Y_t is light tailed and $\mathbf{G}_t$ has regularly varying tails are given by the sample autocovariances of GARCH processes; see Basrak et al. [2].

4. Gaussian quasi maximum likelihood estimation for GARCH processes with heavy–tailed innovations

In this section we apply Theorem 3.2 to Gaussian quasi maximum likelihood estimation (QMLE) in GARCH processes. The limit properties of the QMLE were studied by Berkes et al. [3]. They proved strong consistency of the QMLE under the moment condition $E|Z_1|^{2+\delta} < \infty$ for some $\delta > 0$ and established asymptotic normality under $EZ_1^4 < \infty$. Here (Z_t) is an iid innovation sequence; see Section 4.1 below for the definition of the GARCH model and the QMLE. Hall and Yao [15] refined these results and also allowed for innovations sequences, where Z_1^2 is regularly varying with index $\alpha \in (1,2)$. Then the speed of convergence is slower than the usual $\sqrt{n}$ rate and the limiting distribution of the QMLE is (multivariate) α–stable.

It is our objective to show that the asymptotic theories for the QMLE under light- and heavy-tailed innovations parallel each other and that very similar techniques can be applied in both cases. However, in the light-tailed case (see [3]) an application of the CLT for stationary ergodic martingale differences is the basic tool which establishes the asymptotic normality of the QMLE. In the heavy-tailed situation one depends on an analog of the CLT which is provided by Theorem 3.2.

As a matter of fact, the structure of the proofs shows that the asymptotic properties of the QMLE are not dependent on the particular structure of the GARCH process if one can establish the regular variation of the finite-dimensional distributions of the underlying process (X_t) and the mixing condition $\mathcal{A}(a_n)$. Therefore the results of this section have the potential to be extended to more general models, including, for example, the AGARCH or EGARCH models whose QMLE properties in the light-tailed case were treated in Straumann and Mikosch [28]. The most intricate step in the proof is, however, the verification of this mixing condition for a given time series model. We establish this condition for a GARCH process by an adaptation of Theorem 4.3 in Mokkadem [22]; this yields strong mixing with geometric rate of the relevant sequence. We devote Section 4.4 to the solution of this problem.

Before we start, we introduce some notation. If $K \subset \mathbb{R}^d$ is a compact set, we write $\mathbb{C}(K, \mathbb{R}^{d'})$ for the space of continuous $\mathbb{R}^{d'}$–valued functions equipped with the sup–norm $\|v\|_K = \sup_{s\in K} |v(s)|$. The space $\mathbb{C}(K, \mathbb{R}^{d_1\times d_2})$ consists of the continuous $d_1 \times d_2$–matrix valued functions on K; in $\mathbb{R}^{d_1\times d_2}$ we work with the operator norm induced by the Euclidean norm $|\cdot|$, i.e.,

$$\|\mathbf{A}\| = \sup_{|x|=1} |\mathbf{A}x|, \qquad \mathbf{A} \in \mathbb{R}^{d_1\times d_2}.$$

4.1. **Definition of the QMLE.** Recall the definition of a GARCH(p, q) process (X_t) from (1.1). As before, (Z_t) is an iid innovation sequence with $EZ_1^2 = 1$ and $EZ_1 = 0$, and α_i, β_j are non-negative constants. GARCH processes have been intensively investigated over the last few years. Assumptions for strict stationarity are complicated: they are expressed in terms of Lyapunov exponents of certain random matrices; see Bougerol and Picard [5] for details. A necessary condition for stationarity is

$$\beta_1 + \cdots + \beta_q < 1\,. \tag{4.1}$$

(Corollary 2.3. in [5]). We will make use of this condition later.

In what follows, we always assume strict stationarity of the GARCH processes. As a matter of fact, the observation X_t is always a measurable function of the past and present innovations $(Z_t, Z_{t-1}, Z_{t-2}, \ldots)$; hence (X_t) is automatically ergodic.

In what follows, we review how an approximation to the conditional *Gaussian* likelihood of a stationary GARCH(p, q) process is constructed, i.e., a conditional likelihood under the *synthetic* assumption Z_t iid $\sim \mathcal{N}(0, 1)$. Given $X_0, \ldots, X_{-p+1}$ and $\sigma_0^2, \ldots, \sigma_{-q+1}^2$, the random variables $X_1, \ldots, X_n$ are conditionally Gaussian with mean zero and variances $h_t(\boldsymbol{\theta})$, $t = 1, \ldots, n$, where $\boldsymbol{\theta} = (\alpha_0, \alpha_1, \ldots, \alpha_p, \beta_1, \ldots, \beta_q)^T$ denotes the presumed parameter and

$$\check{h}_t(\boldsymbol{\theta}) = \begin{cases} \sigma_t^2 & t \le 0, \\ \alpha_0 + \alpha_1 X_{t-1}^2 + \cdots + \alpha_p X_{t-p}^2 + \beta_1 \check{h}_{t-1}(\boldsymbol{\theta}) + \cdots + \beta_q \check{h}_{t-q}(\boldsymbol{\theta}) & t > 0. \end{cases}$$

The conditional Gaussian log–likelihood has the form

$$\begin{aligned} &\log f_{\boldsymbol{\theta}}(X_1, \ldots, X_n \mid X_0, \ldots, X_{-p+1}, \sigma_0^2, \ldots, \sigma_{-q+1}^2) \\ &= -\frac{n}{2}\log(2\pi) - \frac{1}{2}\sum_{t=1}^{n}\left(\frac{X_t^2}{\check{h}_t(\boldsymbol{\theta})} + \log \check{h}_t(\boldsymbol{\theta})\right). \end{aligned} \tag{4.2}$$

Since $X_0, \ldots, X_{-p+1}$ are not available and the squared volatilities $\sigma_0^2, \ldots, \sigma_{-q+1}^2$ unobservable, the conditional Gaussian log–likelihood (4.2) cannot be numerically evaluated without a certain initialization for $\sigma_0^2, \ldots, \sigma_{-p+1}^2$ and $X_0, \ldots, X_{-q+1}$. The initial values being asymptotically irrelevant, we set the X_t's equal to zero and $\hat{h}_t(\boldsymbol{\theta}) = \alpha_0/(1 - \beta_1 - \cdots - \beta_q)$ for $t \le 0$. We arrive at

$$\hat{h}_t(\boldsymbol{\theta}) = \begin{cases} \alpha_0/(1 - \beta_1 - \cdots - \beta_q) & t \le 0, \\ \alpha_0 + \alpha_1 X_{t-1}^2 + \cdots + \alpha_{\min(p,t-1)} X_{\max(t-p,1)}^2 & \\ \quad + \beta_1 \hat{h}_{t-1}(\boldsymbol{\theta}) + \cdots + \beta_q \hat{h}_{t-q}(\boldsymbol{\theta}) & t > 0. \end{cases} \tag{4.3}$$

The function $(\hat{h}_t(\boldsymbol{\theta}))^{1/2}$ can be understood as an estimate of the volatility at time t and under parameter hypothesis $\boldsymbol{\theta}$. It can be established that $|\hat{h}_t - \check{h}_t| \overset{\text{a.s.}}{\to} 0$ with a geometric rate of convergence and uniformly on the compact set K defined in (4.4) below. This suggests that by replacing $\check{h}_t(\boldsymbol{\theta})$ by $\hat{h}_t(\boldsymbol{\theta})$ in (4.2) we obtain a good approximation to the conditional Gaussian log–likelihood. Since the constant $-n\log(2\pi)/2$ does not matter for the optimization, we define the QMLE $\hat{\boldsymbol{\theta}}_n$ as a maximizer of the function

$$\hat{L}_n(\boldsymbol{\theta}) = \sum_{t=1}^n \hat{\ell}_t(\boldsymbol{\theta}) = -\frac{1}{2}\sum_{t=1}^n \left(\frac{X_t^2}{\hat{h}_t(\boldsymbol{\theta})} + \log \hat{h}_t(\boldsymbol{\theta})\right)$$

with respect to $\boldsymbol{\theta} \in K$, and K being the compact set

$$K = \left\{ \boldsymbol{\theta} \in \mathbb{R}^{p+q+1} \,\middle|\, m \le \alpha_i, \beta_j \le M,\ \beta_1 + \cdots + \beta_q \le \bar{\beta} \right\}, \tag{4.4}$$

where $0 < m < M < \infty$ and $0 < \bar{\beta} < 1$ are such that $qm < \bar{\beta}$.

Remark 4.1. From a comparison with [3], one might think at first sight that our definition of the QMLE is different from theirs. To see that $\hat{h}_t$ coincides with $\tilde{w}_t$ in [3], introduce the polynomials

$$\boldsymbol{\alpha}(z) = \alpha_1 z + \cdots + \alpha_p z^p \quad \text{and} \quad \boldsymbol{\beta}(z) = 1 - \beta_1 z - \cdots - \beta_q z^q$$

for every $\boldsymbol{\theta} = (\alpha_0, \alpha_1, \ldots, \alpha_p, \beta_1, \ldots, \beta_q)^T \in K$. Then one can show by induction on t that

$$\hat{h}_t(\boldsymbol{\theta}) = \frac{\alpha_0}{\boldsymbol{\beta}(1)} + \sum_{j=1}^{t-1} \psi_j(\boldsymbol{\theta}) X_{t-j}^2, \tag{4.5}$$

where the coefficients $\psi_j(\boldsymbol{\theta})$ are defined through

$$\frac{\boldsymbol{\alpha}(z)}{\boldsymbol{\beta}(z)} = \sum_{j=1}^{\infty} \psi_j(\boldsymbol{\theta}) z^j, \qquad |z| \le 1. \tag{4.6}$$

Note that the latter Taylor series representation is valid because $\beta_i \ge 0$ and $\beta_1 + \cdots + \beta_q \le \bar{\beta} < 1$ imply $\boldsymbol{\beta}(z) \ne 0$ on K for $|z| \le 1 + \epsilon$ and $\epsilon > 0$ sufficiently small. We choose (4.3) rather than (4.5) as a first definition for the squared volatility estimate under parameter hypothesis $\boldsymbol{\theta}$, because the recursion (4.3) is natural and computationally attractive. In [3], starting point for the definition of the QMLE is Theorem 2.2, which says that for all $t \in \mathbb{Z}$ one has $h_t(\boldsymbol{\theta}_0) = \sigma_t^2$, where $\boldsymbol{\theta}_0$ is the true parameter and

$$h_t(\boldsymbol{\theta}) = \frac{\alpha_0}{\boldsymbol{\beta}(1)} + \sum_{j=1}^{\infty} \psi_j(\boldsymbol{\theta}) X_{t-j}^2. \tag{4.7}$$

In [3] this leads to the definition of a squared volatiliy estimate at time t under parameter $\boldsymbol{\theta}$ based on $(X_1 \ldots, X_n)$, which is given by (4.5). Note also that $(h_t(\boldsymbol{\theta}))$ obeys

$$h_{t+1}(\boldsymbol{\theta}) = \alpha_0 + \alpha_1 X_t^2 + \cdots \alpha_p X_{t+1-p}^2 + \beta_1 h_t(\boldsymbol{\theta}) + \cdots + \beta_q h_{t+1-q}(\boldsymbol{\theta}), \qquad \boldsymbol{\theta} \in K. \tag{4.8}$$

4.2. **Limit distribution in the case $EZ_1^4 < \infty$.** First we list the conditions employed by [3] for establishing consistency and asymptotic normality of $\hat{\boldsymbol{\theta}}_n$. Write $\boldsymbol{\theta}_0 = (\alpha_0^\circ, \alpha_1^\circ, \ldots, \alpha_p^\circ, \beta_1^\circ, \ldots, \beta_q^\circ)^T$ for the true parameter.

C.1 There is $\delta > 0$ such that $E|Z_1|^{2+\delta} < \infty$.

C.2 The distribution of $|Z_1|$ is not concentrated in one point.

C.3 There is $\mu > 0$ such that $P(|Z_1| \le t) = o(t^\mu)$ as $t \downarrow 0$.

C.4 The true parameter $\boldsymbol{\theta}_0$ lies in the interior of K.

C.5 The polynomials $\boldsymbol{\alpha}^\circ(z) = \alpha_1^\circ z + \cdots + \alpha_p^\circ z^p$ and $\boldsymbol{\beta}^\circ(z) = 1 - \beta_1^\circ z - \cdots - \beta_q^\circ z^q$ do not have any common roots.

Now we are ready to quote the main result of [3]. We cite it in order to be able to compare the assumptions and assertions both in the light- and heavy-tailed cases; cf. Theorem 4.4 below.

Theorem 4.2 (Theorem 4.1 of Berkes et al. [3])**.** *Let (X_t) be a stationary* GARCH(p,q) *process with true parameter vector $\boldsymbol{\theta}_0$. Suppose the conditions* **C.1 – C.5** *hold. Then the* QMLE $\hat{\boldsymbol{\theta}}_n$ *is strongly consistent, i.e.,*

$$\hat{\boldsymbol{\theta}}_n \stackrel{\text{a.s.}}{\to} \boldsymbol{\theta}_0, \qquad n \to \infty.$$

If in addition $EZ_0^4 < \infty$, then $\hat{\boldsymbol{\theta}}_n$ is also asymptotically normal, i.e.,

$$\sqrt{n}\,(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \stackrel{d}{\to} \mathcal{N}(0, \boldsymbol{B}_0^{-1}\mathbf{A}_0\boldsymbol{B}_0^{-1}),$$

where the $(p+q+1)\times(p+q+1)$–matrices $\mathbf{A}_0$ and $\boldsymbol{B}_0$ are given by

$$\begin{aligned} \mathbf{A}_0 &= \frac{E(Z_0^4-1)}{4}\,E\left(\frac{1}{\sigma_1^4}\,h_1'(\boldsymbol{\theta}_0)^T h_1'(\boldsymbol{\theta}_0)\right), \\ \boldsymbol{B}_0 &= -\frac{1}{2}E\left(\frac{1}{\sigma_1^4}\,h_1'(\boldsymbol{\theta}_0)^T h_1'(\boldsymbol{\theta}_0)\right). \end{aligned} \tag{4.9}$$

4.3. **Limit distribution in the case $EZ_1^4 = \infty$.** First we identify the limit determining term for the QMLE. To this end, we set analogously to [3],

$$L_n(\boldsymbol{\theta}) = \sum_{t=1}^n \ell_t(\boldsymbol{\theta}) = -\frac{1}{2}\sum_{t=1}^n\left(\frac{X_t^2}{h_t(\boldsymbol{\theta})} + \log h_t(\boldsymbol{\theta})\right)$$

and define $\tilde{\boldsymbol{\theta}}_n$ as a maximizer of L_n with respect to $\boldsymbol{\theta} \in K$. It is a slightly simpler problem to analyze $\tilde{\boldsymbol{\theta}}_n$ because (ℓ_t) is stationary ergodic, in contrast to $(\hat{\ell}_t)_{t\in\mathbb{N}}$. As is shown in Proposition 4.3 below, $\hat{\boldsymbol{\theta}}_n$ and $\tilde{\boldsymbol{\theta}}_n$ are asymptotically equivalent. It turns out that the asymptotic distribution of the QMLE is essentially determined by the limit behavior of $L_n'(\boldsymbol{\theta}_0)/n$, up to multiplication with the matrix $-\boldsymbol{B}_0^{-1}$. Actually, these results follow by a careful analysis of the proofs in Berkes et al. [3]. To give some guidance to the reader who wants to verify all details, we briefly repeat the necessary steps and arguments. Compare also with the similar reference Straumann and Mikosch [28], where the case of processes with a more general volatility structure than GARCH is treated.

Proposition 4.3. *Let (X_t) be a stationary* GARCH(p,q) *process with true parameter vector $\boldsymbol{\theta}_0$. Suppose the conditions* **C.1 – C.5** *apply. If there is a positive sequence $(x_n)_{n\ge1}$ with $x_n = o(n)$ as $n\to\infty$ and*

$$x_n\,\frac{L_n'(\boldsymbol{\theta}_0)}{n} \stackrel{d}{\to} \mathbf{D}, \qquad n \to \infty, \tag{4.10}$$

for an $\mathbb{R}^{p+q+1}$–valued random variable $\mathbf{D}$, then the QMLE $\hat{\theta}_n$ satisfies the limit relation

$$x_n(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \stackrel{d}{\to} -\boldsymbol{B}_0^{-1}\mathbf{D}, \tag{4.11}$$

where $\boldsymbol{B}_0$ is given by (4.9)*.*

Proof. We first demonstrate $x_n(\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \stackrel{d}{\to} -\boldsymbol{B}_0^{-1}\mathbf{D}$. In a second step we establish $x_n(\tilde{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}}_n) \stackrel{\text{a.s.}}{\to} \mathbf{0}$. Finally, the assertion (4.11) is an immediate consequence of Slutsky's lemma.

Step 1. Since for all $\boldsymbol{\theta} \in K$ the polynomials $\boldsymbol{\beta}(z)$ have no root in the disc $\{\,|z| \le 1+\epsilon\,\}$, $\epsilon > 0$ sufficiently small, the coefficients $\psi_j(\boldsymbol{\theta})$, see (4.6), and their first and second derivatives decay exponentially fast, uniformly on K (Lemmas 3.2 and 3.3 in [3]). From this together with the fact that $E\log^+|X_0| < \infty$ (Lemma 2.3 in [3]), one shows by means of Lemma 2.2 in [3] that the functions h_t are twice continuously differentiable on K a.s. and that differentiation and infinite sum in (4.6) may be interchanged. As a consequence, (h_t') and (h_t'') are measurable functions of $(X_{t-1}, X_{t-2}, \ldots)$. Since (X_t) is stationary ergodic, so are (h_t') and (h_t'') (see e.g. Proposition 2.3 in [28]). This means that L_n is twice continuously differentiable, i.e., $L_n'' = \sum_{t=1}^n \ell_t''$, where

$$\ell_t''(\boldsymbol{\theta}) = -\frac{1}{2}\frac{1}{h_t(\boldsymbol{\theta})^2}\left((h_t'(\boldsymbol{\theta}))^T h_t'(\boldsymbol{\theta})\left(2\frac{X_t^2}{h_t(\boldsymbol{\theta})} - 1\right) + h_t''(\boldsymbol{\theta})(h_t(\boldsymbol{\theta}) - X_t^2)\right), \qquad \boldsymbol{\theta} \in K. \tag{4.12}$$

An inspection of the proof of Theorem 4.1 in [3] reveals that $\tilde{\boldsymbol{\theta}}_n \stackrel{\text{a.s.}}{\to} \boldsymbol{\theta}_0$. Consequently, for large enough n the following Taylor expansion is valid:

$$L_n'(\tilde{\boldsymbol{\theta}}_n) = L_n'(\boldsymbol{\theta}_0) + L_n''(\boldsymbol{\zeta}_n)(\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0), \tag{4.13}$$

where $|\boldsymbol{\zeta}_n - \boldsymbol{\theta}_0| < |\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0|$. Since $\tilde{\boldsymbol{\theta}}_n$ is the maximizer of L_n and $\boldsymbol{\theta}_0$ lies in the interior of K, one has $L_n'(\tilde{\boldsymbol{\theta}}_n) = 0$. Therefore (4.13) is equivalent to

$$n^{-1} L_n''(\boldsymbol{\zeta}_n)\,(\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) = -n^{-1} L_n'(\boldsymbol{\theta}_0). \tag{4.14}$$

Our aim is to apply a uniform strong law of large numbers for proving the uniform convergence of $n^{-1}L_n'' = n^{-1}\sum_{t=1}^n \ell_t''$ (see e.g. Theorem 2.5 in [28]). Since ℓ_t'' is a measurable function of $(X_t, X_{t-1}, \ldots)$, it is stationary ergodic. If we can verify $E\|\ell_1''\|_K < \infty$, then in $\mathbb{C}(K, \mathbb{R}^{d\times d})$,

$$L_n''/n \stackrel{\text{a.s.}}{\to} L'', \qquad n \to \infty, \tag{4.15}$$

where $L''(\boldsymbol{\theta}) = E[\ell_1''(\boldsymbol{\theta})]$, $\boldsymbol{\theta} \in K$. For showing $E\|\ell_1''\|_K < \infty$, first note that $E\|X_1^2/h_1\|_K^{1+s} < \infty$ if $s < \delta/2$ (Lemma 5.1 of [3]) and that $\|h_1'/h_1\|_K$ and $\|h_1''/h_1\|_K$ have finite moments of any order (Lemma 5.2 of [3]). Then the desired relation is obtained from an application of the triangle inequality to the norm of (4.12), followed by the use of Hölder's inequality. Relation (4.15) together with $\boldsymbol{\zeta}_n \stackrel{\text{a.s.}}{\to} \boldsymbol{\theta}_0$ implies

$$L_n''(\boldsymbol{\zeta}_n)/n \stackrel{\text{a.s.}}{\to} E[\ell_1''(\boldsymbol{\theta}_0)], \qquad n \to \infty.$$

Take into account that h_1, h_1' and h_1'' are independent of Z_1, $h_1(\boldsymbol{\theta}_0) = \sigma_1^2$ a.s. and $X_1 = \sigma_1 Z_1$, in order to conclude

$$E[\ell_1''(\boldsymbol{\theta}_0)] = -2^{-1} E\left((h_1'(\boldsymbol{\theta}_0))^T h_1'(\boldsymbol{\theta}_0)/\sigma_1^4\right) = \boldsymbol{B}_0.$$

It is shown in Lemma 5.7 of [3] that $\boldsymbol{B}_0$ is negative definite and hence invertible. Thus the matrix $L_n''(\boldsymbol{\zeta}_n)/n$ has inverse $\boldsymbol{B}_0^{-1}(1 + o_P(1))$, $n \to \infty$. Therefore equation (4.14) is equivalent to

$$(\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) = -\boldsymbol{B}_0^{-1}(1 + o_P(1)) L_n'(\boldsymbol{\theta}_0)/n, \tag{4.16}$$

which shows that $x_n(\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \stackrel{d}{\to} -\boldsymbol{B}_0^{-1}\mathbf{D}$.

Step 2. The relation $x_n(\tilde{\boldsymbol{\theta}}_n - \hat{\boldsymbol{\theta}}_0) \stackrel{\text{a.s.}}{\to} \mathbf{0}$ follows along the lines of proof of Theorem 4.4 in [3] or Lemma 7.5 in [28]. With the help of the mean value theorem together with the facts that $\|\hat{h}_t - h_t\|_K \stackrel{\text{a.s.}}{\to} 0$ and $\|\hat{h}_t' - h_t'\|_K \stackrel{\text{a.s.}}{\to} 0$ with exponential rate (Lemmas 5.8 and 5.9 in [3]), one can bound $\|\hat{L}_n' - L_n'\|_K \le \sum_{t=1}^\infty \|\hat{\ell}_t' - \ell_t'\|_K$ to show $\sup_{n\in\mathbb{N}} \|\hat{L}_n' - L_n'\|_K < \infty$ a.s. Consequently, since $x_n = o(n)$,

$$\frac{x_n}{n}\,\|\hat{L}_n' - L_n'\|_K \stackrel{\text{a.s.}}{\to} 0, \qquad n \to \infty. \tag{4.17}$$

From Taylor's theorem,

$$L_n'(\tilde{\boldsymbol{\theta}}_n) - L_n'(\hat{\boldsymbol{\theta}}_n) = L_n''(\boldsymbol{\xi}_n)(\tilde{\boldsymbol{\theta}}_n - \hat{\boldsymbol{\theta}}_n), \tag{4.18}$$

where $\boldsymbol{\xi}_n$ lies on the line segment connecting $\hat{\boldsymbol{\theta}}_n$ and $\tilde{\boldsymbol{\theta}}_n$. This line segment is completely contained in the interior of K provided n is large enough. Since $L_n'(\tilde{\boldsymbol{\theta}}_n) = \hat{L}_n'(\hat{\boldsymbol{\theta}}_n) = 0$, equation (4.18) is equivalent to

$$\frac{x_n}{n}\,(\hat{L}_n'(\hat{\boldsymbol{\theta}}_n) - L_n'(\hat{\boldsymbol{\theta}}_n)) = \frac{L_n''(\boldsymbol{\xi}_n)}{n}\,x_n(\tilde{\boldsymbol{\theta}}_n - \hat{\boldsymbol{\theta}}_n). \tag{4.19}$$

By virtue of relation (4.17), both sides of (4.19) tend to $\mathbf{0}$ a.s. when $n \to \infty$. From (4.15) and $\boldsymbol{\xi}_n \stackrel{\text{a.s.}}{\to} \boldsymbol{\theta}_0$, we conclude $L_n''(\boldsymbol{\xi}_n)/n \stackrel{\text{a.s.}}{\to} E[\ell_1''(\boldsymbol{\theta}_0)] = \boldsymbol{B}_0$. Since $\boldsymbol{B}_0$ is invertible, we can deduce $x_n\,(\tilde{\boldsymbol{\theta}}_n - \hat{\boldsymbol{\theta}}_n) \stackrel{\text{a.s.}}{\to} \mathbf{0}$. This completes the proof of the proposition. □

Now we can state the main theorem of this section. We remind once again that Hall and Yao [15] derived the identical result by means of different techniques.

Theorem 4.4. *Let (X_t) be a stationary* GARCH(p,q) *process with true parameter vector $\boldsymbol{\theta}_0$. Suppose that Z_1^2 is regularly varying with index $\alpha \in (1,2)$ and that* **C.3** – **C.5** *hold true. Moreover, assume that Z_1 has a Lebesgue density f, where the closure of the interior of the support $\{f > 0\}$ contains the origin. Define $(x_n) = (n a_n^{-1})$, where*

$$P(Z_1^2 > a_n) \sim n^{-1}, \qquad n \to \infty.$$

Then the QMLE $\hat{\boldsymbol{\theta}}_n$ *is consistent and*

$$x_n(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \stackrel{d}{\to} \mathbf{D}_\alpha, \qquad n \to \infty, \tag{4.20}$$

for some non–degenerate α–stable vector $\mathbf{D}_\alpha$.

Before proving the theorem, we discuss its practical consequences for parameter inference:

- The rate of convergence x_n has — roughly speaking — magnitude $n^{1-1/\alpha}$, which is less than $\sqrt{n}$. The heavier the tails of the innovations, i.e., the smaller α, the slower is the convergence of $\hat{\boldsymbol{\theta}}_n$ towards the true parameter $\boldsymbol{\theta}_0$.
- The limit distribution of the standardized differences $(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)$ is α–stable and hence non–Gaussian. The exact parameters of this α–stable limit are not explicitly known.
- Confidence bands based on the normal approximation of Theorem 4.2 are false if $EZ_1^4 = \infty$.
- By the definition of a GARCH process, the distribution of the innovations Z_t is unknown. Therefore assumptions about the heaviness of the tails of its distribution are purely hypothetical. As a matter of fact, the tails of the distribution of X_t can be regularly varying even if Z_t has light tails, such as for the normal distribution; see Basrak et al. [2]. Depending on the assumptions on the distribution of Z_1, one can develop different asymptotic theories for QMLE of GARCH processes: asymptotic normality as provided by Theorem 4.2 or infinite variance stable distributions as provided by Theorem 4.4.

Proof. The proof follows by combining Theorem 3.2 and Proposition 4.3. Indeed, setting

$$\mathbf{G}_t = h_t'(\boldsymbol{\theta}_0)/\sigma_t^2\,, \quad Y_t = (Z_t^2 - 1)/2 \quad \text{and} \quad \mathbf{Y}_t = \mathbf{G}_t\, Y_t\,,$$

one recognizes that

$$L_n'(\boldsymbol{\theta}_0) = \frac{1}{2}\sum_{t=1}^{\infty} \frac{h_t'(\boldsymbol{\theta}_0)}{\sigma_t^2}\,(Z_t^2 - 1) = \sum_{t=1}^{n} \mathbf{G}_t Y_t \tag{4.21}$$

is a martingale transform. Regular variation of Z_1^2 with index $\alpha \in (1,2)$ implies **A.1**, but also **C.1** and **C.2**. Condition **A.2** is fulfilled because $\|h_1'/h_1\|_K$ has finite moments of any order (Lemma 5.2 of [3]), and so has $\|\mathbf{G}_1\|$. The condition **A.3** holds true if we can show that $(\mathbf{Y}_t)$ is strongly mixing with geometric rate, in which case we choose $r_n = n^\delta$ in $\mathcal{A}(a_n)$ for any small $\delta > 0$ so that (3.1) immediately follows. This choice of (r_n) is justified by the arguments given in Basrak et al. [2]. The strong mixing condition with geometric rate of $(\mathbf{Y}_t)$ will be verified in Section 4.4. Finally, we have to give an argument for $\gamma > 0$. The latter quantity has interpretation as the extremal index of the sequence $(|\mathbf{Y}_t|)$ (see Remark 2.3 in [12]; cf. Leadbetter et al. [18] for the definition and properties of the extremal index). According to Theorem 3.7.2 in [18], if $\gamma = 0$, then if for some sequence (u_n) the relation $\liminf_{n\to\infty} P(\tilde{M}_n \le u_n) > 0$ holds, one neccessarily has $\lim_{n\to\infty} P(M_n \le u_n) = 1$. Here $M_n = \max(|\mathbf{Y}_1|, \ldots, |\mathbf{Y}_n|)$ and $(\tilde{M}_n)$ is the corresponding sequence of partial maxima for an iid sequence (R_i) where R_1 has the same distribution as $|\mathbf{Y}_1|$. Assume $\gamma = 0$. The random variable

$|\mathbf{Y}_1|$ is regularly varying with index α since $\mathbf{Y}_1$ is regularly varying with index α. Hence $(a_n^{-1}\tilde{M}_n)$ has a Fréchet limit distribution Φ_α, but $P(M_n \le x a_n) \to 1$ does not hold for all positive x. Indeed, straightforward arguments exploiting

$$\sum_{j=1}^{\infty} \frac{\partial \psi_j(\boldsymbol{\theta})}{\partial \alpha_i}\, z^j = \frac{z^i}{\boldsymbol{\beta}(z)}, \qquad |z| \le 1,$$

for all $i = 1, \ldots, p$, show that

$$\frac{\partial h_t(\boldsymbol{\theta})}{\partial \alpha_i} \ge 0 \qquad \text{for all } i = 0, \ldots p, \tag{4.22}$$

and

$$\sum_{i=0}^{p} \alpha_i \, \frac{\partial h_t(\boldsymbol{\theta})}{\partial \alpha_i} = h_t(\boldsymbol{\theta}). \tag{4.23}$$

Since the Euclidean norm is equivalent to the 1–norm $|\boldsymbol{x}| = \sum_{i=1}^{p+q+1} |x_i|$ and $\alpha_i \le M$ on K, there is $c > 0$ such that

$$\frac{|h_t'(\boldsymbol{\theta})|}{h_t(\boldsymbol{\theta})} \ge \frac{c}{h_t(\boldsymbol{\theta})} \sum_{i=0}^{p} \alpha_i \left| \frac{\partial h_t(\boldsymbol{\theta})}{\partial \alpha_i} \right| = \frac{c}{h_t(\boldsymbol{\theta})} \sum_{i=0}^{p} \alpha_i \, \frac{\partial h_t(\boldsymbol{\theta})}{\partial \alpha_i} = c.$$

Note that the last two equalities in the latter display are a consequence of (4.22) and (4.23). Hence we have $P(M_n \le x\, a_n) \le P(\max_{t \le n} |Y_t| \le c^{-1} x a_n)$, and the right–hand side converges to a Fréchet limit and is never equal to 1 for all positive x. From this contradiction we may conclude that $\gamma > 0$. All conditions of Theorem 3.2 have been verified so that

$$2a_n^{-1} L_n'(\boldsymbol{\theta}_0) = 2x_n \, \frac{L_n'(\boldsymbol{\theta}_0)}{n} \xrightarrow{d} \tilde{\mathbf{D}}_\alpha,$$

where $\tilde{\mathbf{D}}_\alpha$ is α–stable (notice that $P((Z_0^2 - 1)/2 > a_n/2) \sim P(Z_0^2 > a_n) \sim n^{-1}$). Since $x_n/n = a_n^{-1} \to 0$, Proposition 4.3 implies

$$x_n(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \xrightarrow{d} -2^{-1} \boldsymbol{B}_0^{-1} \tilde{\mathbf{D}}_\alpha = \mathbf{D}_\alpha.$$

Recalling that a linear transform of an α–stable random vector is again α–stable concludes the proof of the theorem. $\square$

4.4. **Verification of strong mixing with geometric rate of $(\mathbf{Y}_t)$.** To begin with, we quote a powerful result due to Mokkadem [22], which allows one to establish strong mixing in stationary solutions of so–called polynomial linear stochastic recurrence equations (SRE's). A sequence $(\mathbf{Y}_t)$ of random vectors in $\mathbb{R}^d$ obeys a linear SRE if

$$\mathbf{Y}_t = \mathbf{P}_t \mathbf{Y}_{t-1} + \mathbf{Q}_t \tag{4.24}$$

where $((\mathbf{P}_t, \mathbf{Q}_t))$ constitutes an iid sequence with values in $\mathbb{R}^{d\times d} \times \mathbb{R}^d$. A linear SRE is called *polynomial* if there exists an iid sequence $(\mathbf{e}_t)$ in $\mathbb{R}^{d'}$ such that $\mathbf{P}_t = \mathbf{P}(\mathbf{e}_t)$ and $\mathbf{Q}_t = \mathbf{Q}(\mathbf{e}_t)$, where $\mathbf{P}(\mathbf{x})$ and $\mathbf{Q}(\mathbf{x})$ have entries and coordinates, respectively, which are polynomial functions of the coordinates of $\mathbf{x}$. The existence and uniqueness of a stationarity solution to (4.24) has been studied by Brandt [9], Bougerol and Picard [6] and Babillot et al. [1] and others. The following set of conditions is sufficient: $E \log^+ \|\mathbf{P}_1\| < \infty$, $E \log^+ |\mathbf{Q}_{\mathbf{1}}| < \infty$, and the top Lyapunov coefficient associated with the operator sequence $(\mathbf{P}_t)$ is strictly negative, i.e.,

$$\rho = \inf\{\, t^{-1} E \log \|\mathbf{P}_t \cdots \mathbf{P}_1\| \;\; | \; t \ge 1\} < 0. \tag{4.25}$$

Here $\|\cdot\|$ is the operator norm corresponding to an arbitrary fixed norm $|\cdot|$ in $\mathbb{R}^d$, e.g. the Euclidean norm. The following result is a slight generalization of Theorem 4.3 in [22]; see the beginning of the proof below for a comparison.

Theorem 4.5. *Let* $(\mathbf{e}_t)$ *be an iid sequence of random vectors in* $\mathbb{R}^{d'}$. *Then consider the polynomial linear* SRE

$$\mathbf{Y}_t = \mathbf{P}(\mathbf{e}_t)\mathbf{Y}_{t-1} + \mathbf{Q}(\mathbf{e}_t), \tag{4.26}$$

where $\mathbf{P}(\mathbf{e}_t)$ *is a random* $d \times d$ *matrix and* $\mathbf{Q}(\mathbf{e}_t)$ *a random* $\mathbb{R}^d$*–valued vector. Suppose:*

1. $\mathbf{P}(\mathbf{0})$ *has spectral radius strictly smaller than* 1 *and the top Lyapunov coefficient* ρ *corresponding to* $(\mathbf{P}(\mathbf{e}_t))$ *is strictly negative.*
2. *There is* $s > 0$ *such that*
$$E\|\mathbf{P}(\mathbf{e}_1)\|^s < \infty \quad \text{and} \quad E|\mathbf{Q}(\mathbf{e}_1)|^s < \infty.$$
3. *There is a smooth algebraic variety* $V \subset \mathbb{R}^{d'}$ *such that* $\mathbf{e}_1$ *has a density* f *with respect to Lebesgue measure on* V. *Assume that* $\mathbf{0}$ *is contained in the closure of the interior of the support* $\{f > 0\}$.

Then the polynomial linear SRE (4.26) *has a unique stationary ergodic solution* $(\mathbf{Y}_t)$, *which is absolutely regular with geometric rate and consequently strongly mixing with geometric rate.*

Remark 4.6. As regards the definition of a *smooth algebraic variety*, we first introduce the notion of an *algebraic subset.* An algebraic subset of $\mathbb{R}^{d'}$ is a set of form

$$V = \{\mathbf{x} \in \mathbb{R}^{d'} \mid F_1(\mathbf{x}) = \cdots = F_r(\mathbf{x}) = \mathbf{0}\},$$

where $F_1, \ldots, F_r$ are real multivariate polynomials. An *algebraic variety* is an algebraic subset which is not the union of two proper algebraic subsets. An algebraic variety is *smooth* if the Jacobian of $\mathbf{F} = (F_1, \ldots, F_r)^T$ has identical rank everywhere on V. Examples of smooth algebraic varieties in $\mathbb{R}^{d'}$ are the hyperplanes of $\mathbb{R}^{d'}$ or $V = \mathbb{R}^{d'}$. Recall also the definition of absolute regularity (2.4).

Proof. There is nothing to prove if $E\|\mathbf{P}(\mathbf{e}_1)\|^{\tilde{s}} < 1$ for some $\tilde{s} > 0$ as this special case is the content of Theorem 4.3 in [22]. For the general case it suffices to prove the absolute regularity with geometric rate for some subsequence $(\mathbf{Y}_{tm})_{t\in\mathbb{Z}}$, where $m \geq 1$ is fixed. Indeed, the mixing coefficient b_k is nonincreasing and since $(\mathbf{Y}_t)$ is a Markov process, the simpler representation

$$b_k = E\left(\sup_{B\in\sigma(\mathbf{Y}_{k+1})} |P(B \mid \sigma(\mathbf{Y}_0)) - P(B)|\right)$$

is also valid, see e.g. Bradley [8]. Since $\rho < 0$, there is $m \geq 1$ with $E\log\|\mathbf{P}(\mathbf{e}_m)\cdots\mathbf{P}(\mathbf{e}_1)\| < 0$. From the fact that the map $u \mapsto E\|\mathbf{P}(\mathbf{e}_m)\cdots\mathbf{P}(\mathbf{e}_1)\|^u$ has first derivative equal to $E\log\|\mathbf{P}(\mathbf{e}_m)\cdots\mathbf{P}(\mathbf{e}_1)\|$ at $u = 0$, we deduce that there is $0 < \tilde{s} \leq s$ with $E\|\mathbf{P}(\mathbf{e}_m)\cdots\mathbf{P}(\mathbf{e}_1)\|^{\tilde{s}} < 1$. Then note that $(\tilde{\mathbf{Y}}_t) = (\mathbf{Y}_{tm})$ obeys a linear SRE:

$$\tilde{\mathbf{Y}}_t = \tilde{\mathbf{P}}(\tilde{\mathbf{e}}_t)\tilde{\mathbf{Y}}_{t-1} + \tilde{\mathbf{Q}}(\tilde{\mathbf{e}}_t),$$

where
$$\tilde{\mathbf{e}}_t = \begin{pmatrix} \mathbf{e}_{tm} \\ \vdots \\ \mathbf{e}_{(t-1)m+1} \end{pmatrix}$$

and

$$\begin{aligned}
\tilde{\mathbf{P}}(\tilde{\mathbf{e}}_t) &= \mathbf{P}(\mathbf{e}_{tm})\cdots\mathbf{P}(\mathbf{e}_{(t-1)m+1}),\\
\tilde{\mathbf{Q}}(\tilde{\mathbf{e}}_t) &= \mathbf{Q}(\mathbf{e}_{tm}) + \sum_{j=1}^{m-1}\left(\prod_{i=1}^{j}\mathbf{P}(\mathbf{e}_{tm+1-i})\right)\mathbf{Q}(\mathbf{e}_{tm-j}).
\end{aligned}$$

Since both the matrix $\tilde{\mathbf{P}}(\tilde{\mathbf{e}}_t)$ and the vector $\tilde{\mathbf{Q}}(\tilde{\mathbf{e}}_t)$ are polynomial functions of the coordinates of $\tilde{\mathbf{e}}_t$ and the sequence $(\tilde{\mathbf{e}}_t)$ is iid, $(\tilde{\mathbf{Y}}_t)$ obeys a polynomial linear SRE. Observe that $\tilde{\mathbf{P}}(\mathbf{0}) = (\mathbf{P}(\mathbf{0}))^m$ has spectral radius strictly smaller than 1, that $E\|\tilde{\mathbf{P}}(\tilde{\mathbf{e}}_1)\|^{\tilde{s}} < 1$ and $E\|\tilde{\mathbf{Q}}(\tilde{\mathbf{e}}_1)\|^{\tilde{s}} < \infty$ and that $\tilde{\mathbf{e}}_1$

has a density with respect to Lebesgue measure on V^m, where V^m is a smooth algebraic variety (see A.14 in [22]). Thus an application of Theorem 4.3 in [22] yields that $(\tilde{\mathbf{Y}}_t)$ is absolutely regular with geometric rate. This proves the assertion. □

The following two facts will also be needed.

Lemma 4.7. *Let $(\mathbf{P}_t)$ be an iid sequence of $k \times k$–matrices with $E\|\mathbf{P}_1\|^s < \infty$ for some $s > 0$. Then the associated top Lyapunov coefficient $\rho < 0$ if and only if there exist $c > 0$, $\tilde{s} > 0$ and $\lambda < 1$ so that*

$$E\|\mathbf{P}_t \cdots \mathbf{P}_1\|^{\tilde{s}} \le c\lambda^t, \qquad t \ge 1. \tag{4.27}$$

Proof. For the proof of necessity, observe that there exists $n \ge 1$ such that $E \log \|\mathbf{P}_n \cdots \mathbf{P}_1\| < 0$. From the fact that the map $u \mapsto E\|\mathbf{P}_n \cdots \mathbf{P}_1\|^u$ has first derivative equal to $E \log \|\mathbf{P}_n \cdots \mathbf{P}_1\|$ at $u = 0$, we deduce that there is $\tilde{s} > 0$ with $E\|\mathbf{P}_n \cdots \mathbf{P}_1\|^{\tilde{s}} = \tilde{\lambda} < 1$. Since the operator norm $\|\cdot\|$ is submultiplicative and the factors in $\mathbf{P}_t \cdots \mathbf{P}_1$ are iid,

$$E\|\mathbf{P}_t \cdots \mathbf{P}_1\|^{\tilde{s}} \le \tilde{\lambda}^{t/n-1} \left(\max_{\ell=1,\dots,n-1} E\|\mathbf{P}_\ell \cdots \mathbf{P}_1\|^{\tilde{s}} \right) \le c\lambda^t, \qquad t \ge 1,$$

for $c = \tilde{\lambda}^{-1} \left(\max_{\ell=1,\dots,n-1} E\|\mathbf{P}_\ell \cdots \mathbf{P}_1\|^{\tilde{s}} \right)$ and $\lambda = \tilde{\lambda}^{1/n}$. Regarding the proof of sufficiency, use Jensen's inequality and $\lim_{t\to\infty} t^{-1} E \log \|\mathbf{P}_t \cdots \mathbf{P}_1\| = \rho$ to conclude

$$\begin{aligned} \rho = \lim_{t\to\infty} \frac{1}{t\tilde{s}} E \log \|\mathbf{P}_t \cdots \mathbf{P}_1\|^{\tilde{s}} &\le \limsup_{t\to\infty} \frac{1}{t\tilde{s}} \log E\|\mathbf{P}_t \cdots \mathbf{P}_1\|^{\tilde{s}} \\ &\le \limsup_{t\to\infty} \frac{1}{t\tilde{s}} (\log c + t \log \lambda) = \frac{\log \lambda}{\tilde{s}} < 0. \end{aligned}$$

This completes the proof of the lemma. □

Lemma 4.8. *Suppose that*

$$\mathbf{P}_t = \begin{pmatrix} \mathbf{A}_t & \mathbf{0}_{r\times(k-r)} \\ \mathbf{B}_t & \mathbf{C}_t \end{pmatrix}, \qquad t \in \mathbb{Z}, \tag{4.28}$$

forms an iid sequence of $k \times k$–matrices with $E\|\mathbf{P}_1\|^s < \infty$, $s > 0$, where $\mathbf{A}_t \in \mathbb{R}^{r\times r}$, $\mathbf{B}_t \in \mathbb{R}^{(k-r)\times r}$ and $\mathbf{C}_t \in \mathbb{R}^{(k-r)\times(k-r)}$. Then its associated top Lyapunov coefficient $\rho_{\mathbf{P}} < 0$ if and only if the sequences $(\mathbf{A}_t)$ and $(\mathbf{C}_t)$ have top Lyapunov coefficients $\rho_{\mathbf{A}} < 0$ and $\rho_{\mathbf{C}} < 0$.

Proof. For the proof of sufficiency of $\rho_{\mathbf{A}} < 0$ and $\rho_{\mathbf{C}} < 0$ for $\rho_{\mathbf{P}} < 0$, it is by Lemma 4.7 enough to derive a moment inequality of form (4.27) for $(\mathbf{P}_t)$. By induction we obtain

$$\mathbf{P}_t \cdots \mathbf{P}_1 = \begin{pmatrix} \mathbf{A}_t \cdots \mathbf{A}_1 & \mathbf{0}_{r\times(k-r)} \\ \mathbf{Q}_t & \mathbf{C}_t \cdots \mathbf{C}_1 \end{pmatrix},$$

where

$$\begin{aligned} \mathbf{Q}_t &= \mathbf{B}_t\mathbf{A}_{t-1} \cdots \mathbf{A}_1 + \mathbf{C}_t\mathbf{B}_{t-1}\mathbf{A}_{t-2} \cdots \mathbf{A}_1 + \mathbf{C}_t\mathbf{C}_{t-1}\mathbf{B}_{t-2}\mathbf{A}_{t-3} \cdots \mathbf{A}_1 \\ &\quad + \cdots + \mathbf{C}_t \cdots \mathbf{C}_3\mathbf{B}_2\mathbf{A}_1 + \mathbf{C}_t \cdots \mathbf{C}_2\mathbf{B}_1 . \end{aligned}$$

Observe that

$$\max(\|\mathbf{A}_t \cdots \mathbf{A}_1\|, \|\mathbf{C}_t \cdots \mathbf{C}_1\|) \le \|\mathbf{P}_t \cdots \mathbf{P}_1\| \le \|\mathbf{A}_t \cdots \mathbf{A}_1\| + \|\mathbf{C}_t \cdots \mathbf{C}_1\| + \|\mathbf{Q}_t\| . \tag{4.29}$$

It is sufficient to show (4.27) for each block in the matrix $\mathbf{P}_t \cdots \mathbf{P}_1$. Because of $\rho_{\mathbf{A}} < 0$, $\rho_{\mathbf{C}} < 0$ and $E\|\mathbf{A}_1\|^s, E\|\mathbf{C}_1\|^s \le E\|\mathbf{P}_1\|^s < \infty$, Lemma 4.7 already implies moment bounds of form (4.27) for $(\mathbf{A}_t)$ and $(\mathbf{C}_t)$. Thus we are left to bound $\|\mathbf{Q}_t\|$. Without loss of generality we may assume that the constants $\lambda < 1$ and $\tilde{s}, c > 0$ in the inequality (4.27) are equal for $(\mathbf{A}_t)$ and $(\mathbf{C}_t)$ and that

$\tilde{s} \le s \le 1$. From an application of the Minkowski inequality and exploiting the independence of the factors in each summand of $\mathbf{Q}_t$, we receive the desired relation

$$E\|\mathbf{Q}_t\|^{\tilde{s}} \le c^2 t\, E\|\mathbf{B}_1\|^{\tilde{s}} \lambda^{t-1} \le \tilde{c}\,\tilde{\lambda}^t\,,$$

some $\tilde{\lambda} \in (\lambda, 1)$, $\tilde{c} > 0$. For the proof of necessity, assume $\rho_{\mathbf{P}} < 0$. Then the left-hand estimates in (4.29) and Lemma 4.7 imply that $\rho_{\mathbf{A}} < 0$ and $\rho_{\mathbf{C}} < 0$. □

We now exploit Theorem 4.5 in order to establish strong mixing with geometric rate of the sequence $(\mathbf{Y}_t) = (\mathbf{G}_t Y_t)$, where $\mathbf{G}_t = h_t'(\boldsymbol{\theta}_0)/\sigma_t^2$ and $Y_t = (Z_t^2 - 1)/2$.

Proposition 4.9. *Let (X_t) be a stationary* GARCH(p,q) *process with true parameter vector $\boldsymbol{\theta}_0$. Moreover, assume that Z_1 has a Lebesgue density f, where the closure of the interior of the support $\{f > 0\}$ contains the origin. Then $(\mathbf{Y}_t)$ is absolutely regular with geometric rate.*

Proof. For the proof of this result we first embed $(\mathbf{Y}_t)$ in a polynomial linear SRE. Without loss of generality assume $p, q \ge 3$. Write

$$\begin{aligned}
\tilde{\mathbf{Y}}_t &= \Big(\sigma_{t+1}^2, \ldots, \sigma_{t-q+2}^2, X_t^2, \ldots, X_{t-p+2}^2, \\
&\qquad \frac{\partial h_{t+1}(\boldsymbol{\theta}_0)}{\partial \alpha_0}, \ldots, \frac{\partial h_{t-q+2}(\boldsymbol{\theta}_0)}{\partial \alpha_0}, \ldots, \frac{\partial h_{t+1}(\boldsymbol{\theta}_0)}{\partial \alpha_p}, \ldots, \frac{\partial h_{t-q+2}(\boldsymbol{\theta}_0)}{\partial \alpha_p}, \\
&\qquad \frac{\partial h_{t+1}(\boldsymbol{\theta}_0)}{\partial \beta_1}, \ldots, \frac{\partial h_{t-q+2}(\boldsymbol{\theta}_0)}{\partial \beta_1}, \ldots, \frac{\partial h_{t+1}(\boldsymbol{\theta}_0)}{\partial \beta_q}, \ldots, \frac{\partial h_{t-q+2}(\boldsymbol{\theta}_0)}{\partial \beta_q}\Big)^T.
\end{aligned}$$

Since $Z_t^2 = X_t^2/\sigma_t^2$, we have

$$\sigma(\mathbf{Y}_t,\, t > k) \subset \sigma(\tilde{\mathbf{Y}}_t, t > k) \quad \text{and} \quad \sigma(\mathbf{Y}_t,\, t \le 0) \subset \sigma(\tilde{\mathbf{Y}}_t,\, t \le 0).$$

Consequently, it is enough to demonstrate abolute regularity with geometric rate of the sequence $(\tilde{\mathbf{Y}}_t)$. We introduce various matrices. Write $\mathbf{0}_{d_1 \times d_2}$ for the $d_1 \times d_2$ matrix with all entries equal to zero and let $\mathbf{I}_d$ denote the identity matrix of dimension d. Then set

$$\mathbf{M}_1(Z_t) = \begin{pmatrix} \boldsymbol{\tau}_t & \beta_q^\circ & \boldsymbol{\alpha}^\circ & \alpha_p^\circ \\ \mathbf{I}_{q-1} & \mathbf{0}_{(q-1)\times 1} & \mathbf{0}_{(q-1)\times(p-2)} & \mathbf{0}_{(q-1)\times 1} \\ \boldsymbol{\xi}_t & \mathbf{0}_{1\times 1} & \mathbf{0}_{1\times(p-2)} & \mathbf{0}_{1\times 1} \\ \mathbf{0}_{(p-2)\times(q-1)} & \mathbf{0}_{(p-2)\times 1} & \mathbf{0}_{(p-2)\times(p-2)} & \mathbf{0}_{(p-2)\times 1} \end{pmatrix},$$

where

$$\begin{aligned}
\boldsymbol{\tau}_t &= (\beta_1^\circ + \alpha_1^\circ Z_t^2, \beta_2^\circ, \ldots, \beta_{q-1}^\circ) \in \mathbb{R}^{q-1}, \\
\boldsymbol{\xi}_t &= (Z_t^2, 0, \ldots, 0) \in \mathbb{R}^{q-1}, \\
\boldsymbol{\alpha}^\circ &= (\alpha_2^\circ, \ldots, \alpha_{p-1}^\circ) \in \mathbb{R}^{p-2}.
\end{aligned}$$

Moreover, define

$$\mathbf{M}_2(Z_t) = \begin{pmatrix} \mathbf{0}_{q\times(p+q-1)} \\ \mathbf{U}_1 \\ \vdots \\ \mathbf{U}_p \end{pmatrix} \qquad \text{and} \qquad \mathbf{M}_4 = \begin{pmatrix} \mathbf{V}_1 \\ \vdots \\ \mathbf{V}_q \end{pmatrix},$$

where $\mathbf{U}_i \in \mathbb{R}^{q\times(p+q-1)}$ and $\mathbf{V}_j \in \mathbb{R}^{q\times(p+q-1)}$ are given by

$$\begin{aligned}
[\mathbf{U}_1]_{k,\ell} &= \delta_{k\ell,11}\, Z_t^2, \\
[\mathbf{U}_i]_{k,\ell} &= \delta_{k\ell,1(q+i-1)}, \qquad i \ge 2, \\
[\mathbf{V}_j]_{k,\ell} &= \delta_{k\ell,1j}.
\end{aligned}$$

Here, $\delta.$ denotes the Kronecker symbol. Also introduce the $q\times q$ matrix

$$\mathbf{C} = \begin{pmatrix} \beta_1^\circ & \cdots & \beta_q^\circ \\ & & \\ \mathbf{I}_{q-1} & & \mathbf{0}_{(q-1)\times 1} \end{pmatrix},$$

and let

$$\mathbf{M}_3 = \operatorname{diag}(\mathbf{C}, p+1), \qquad \mathbf{M}_5 = \operatorname{diag}(\mathbf{C}, q)$$

be the block diagonal matrices consisting of $p+1$ (or q) copies of the block C. Finally, we define

$$\mathbf{P}(Z_t) = \begin{pmatrix} \mathbf{M}_1(Z_t) & \mathbf{0}_{(p+q-1)\times(p+1)q} & \mathbf{0}_{(p+q-1)\times q^2} \\ \mathbf{M}_2(Z_t) & \mathbf{M}_3 & \mathbf{0}_{(p+1)q\times q^2} \\ \mathbf{M}_4 & \mathbf{0}_{q^2\times(p+1)q} & \mathbf{M}_5 \end{pmatrix}$$

and $\mathbf{Q} \in \mathbb{R}^{p+q-1+q(p+q+1)}$ by $[\mathbf{Q}]_k = \alpha_0\delta_{k,1} + \delta_{k,p+q}$. Differentiating both sides of (4.8) at the true parameter $\boldsymbol{\theta} = \boldsymbol{\theta}_0$, we recognize that

$$h'_{t+1}(\boldsymbol{\theta}_0) = (1, X_t^2, \ldots, X_{t+1-p}^2, \sigma_t^2, \ldots, \sigma_{t+1-q}^2)^T + \beta_1^\circ h'_t(\boldsymbol{\theta}_0) + \cdots + \beta_q^\circ h'_{t+1-q}(\boldsymbol{\theta}_0).$$

From this recursive relationship together with $\sigma_{t+1}^2 = \alpha_0^\circ + \alpha_1^\circ X_t^2 + \cdots + \alpha_p^\circ X_{t+1-p}^2 + \beta_1^\circ\sigma_t^2 + \cdots + \beta_q^\circ\sigma_{t+1-q}^2$ we derive a polynomial linear SRE for $(\tilde{\mathbf{Y}}_t)$:

$$\tilde{\mathbf{Y}}_t = \mathbf{P}(Z_t)\tilde{\mathbf{Y}}_{t-1} + \mathbf{Q}. \tag{4.30}$$

The proof of Proposition 4.9 follows from the following lemma. □

Lemma 4.10. *Under the assumptions of Proposition 4.9, the polynomial linear* SRE (4.30) *has a strictly stationary solution* $(\tilde{\mathbf{Y}}_t)$ *which is absolutely regular with geometric rate.*

Proof. The aim is to show that (4.30) obeys the conditions of Theorem 4.5. Since $EZ_1^2 = 1$ it is immediate that $E\|\mathbf{P}(Z_1)\| < \infty$ since this statement is true for the Frobenius norm, and all matrix norms are equivalent. Treat the blocks $\mathbf{M}_1(Z_t)$, $\mathbf{M}_3$ and $\mathbf{M}_4$ separately. Observe that the matrix $\mathbf{M}_1(Z_t)$ appears in the linear SRE for the vector $\mathbf{S}_t = (\sigma_{t+1}^2, \ldots, \sigma_{t-q+2}^2, X_t^2, \ldots X_{t-p+2}^2)^T$, namely

$$\mathbf{S}_t = \mathbf{M}_1(Z_t)\mathbf{S}_{t-1} + (\alpha_0^\circ, 0, \ldots, 0)^T.$$

Theorem 1.3 of [5] says that (1.1) admits a unique stationary solution if and only if $(\mathbf{M}_1(Z_t))$ has strictly negative top Lyapunov coefficient; consequently $\rho_{\mathbf{M}_1} < 0$. Moreover, arguing by recursion on p and expanding the determinant with respect to the last column, it is easily verified that $\mathbf{M}_1(0)$ has characteristic polynomial

$$\det(\lambda\mathbf{I}_{p+q-1} - \mathbf{M}_1(0)) = \lambda^{p+q-1}\left(1 - \sum_{i=1}^q \beta_i^\circ\lambda^{-i}\right).$$

Since (4.1) holds for a stationary GARCH(p,q) process, by the triangle inequality

$$\left|1 - \sum_{i=1}^q \beta_i^\circ\lambda^{-i}\right| \ge 1 - \sum_{i=1}^q \beta_i^\circ\lambda^{-i} \ge 1 - \sum_{i=1}^q \beta_i^\circ > 0$$

if $|\lambda| \geq 1$, and hence $\mathbf{M}_1(0)$ has spectral radius < 1. Observe that the building block $\mathbf{C}$ has characteristic polynomial

$$\det(\lambda \mathbf{I}_q - \mathbf{C}) = \lambda^q \left(1 - \sum_{i=1}^{q} \beta_i^\circ \lambda^{-i}\right),$$

showing that its spectral radius is strictly smaller than 1 (use the same argument as before). Thus the *deterministic* matrices $\mathbf{M}_3$ and $\mathbf{M}_5$ have spectral radius < 1, which also implies that their associated top Lyapunov coefficients are stricly negative. Combining these results, we deduce that $\mathbf{P}(0)$ has spectral radius < 1 and conclude by twice applying Lemma 4.8 that $(\mathbf{P}(Z_t))$ has strictly negative top Lyapunov coefficient. Hence by Theorem 4.5 the stationary sequence $(\tilde{\mathbf{Y}}_t)$ is absolutely regular with geometric rate. □

Remark 4.11. Since (X_t^2, σ_t^2) is a subvector of $\tilde{\mathbf{Y}}_t$, stationary GARCH(p, q) processes are absolutely regular with geometric rate; this result has previously been established by Boussama [7].

Acknowledgement. Daniel Straumann would like to thank RiskLab Zürich for the opportunity to conduct this research. In particular, he is grateful to Paul Embrechts and Alexander McNeil for stimulating discussions.

References

[1] Babillot, M. and Bougerol, P. and Elie, L. (1997) The random difference equation $X_n = A_n X_{n-1} + B_n$ in the critical case. *Ann. Probab.* **25**, 478–493.

[2] Basrak, B., Davis, R.A. and Mikosch. T. (2002) Regular variation of GARCH processes. *Stoch. Proc. Appl.*, **99**, 95–115.

[3] Berkes, I., Horváth, L. and Kokoszka, P. (2003) GARCH processes: structure and estimation. *Bernoulli* **9**, 201–228.

[4] Billingsley, P. (1968) *Convergence of Probability Measures.* (1968) Wiley, New York.

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

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

[7] Boussama, F. (1998) *Ergodicité, mélange et estimation dans les modèles GARCH.* PhD Thesis, Université 7 Paris.

[8] Bradley, R.C. (1986) Basic properties of strong mixing conditions. In: *E. Eberlein and M. Taqqu (Eds.): Dependence in Probability and Statistics* Birkhäuser, Boston.

[9] Brandt, A. (1986) The stochastic equation $Y_{n+1} = A_n Y_n + B_n$ with stationary coefficients. *Adv. in Appl. Probab.* **18**, 211–220.

[10] Breiman, L. (1965) On some limit theorems similar to the arc-sin law. *Theory Probab. Appl.* **10**, 323–331.

[11] Davis, R.A. and Hsing, T. (1995) Point process and partial sum convergence for weakly dependent random variables with infinite variance. *Ann. Probab.* **23**, 879–917.

[12] Davis, R.A. and Mikosch, T. (1998) The sample autocorrelations of heavy-tailed processes with applications to ARCH. *Ann. Statist.* **26**, 2049–2080.

[13] Doukhan, P. (1994) *Mixing. Properties and Examples.* Lecture Notes in Statistics **85**. Springer Verlag, New York.

[14] Embrechts, P., Klüppelberg, C. and Mikosch, T. (1997) *Modelling Extremal Events for Insurance and Finance.* Springer, Berlin.

[15] Hall, P.G. and Yao, Q. (2003) Inference in ARCH and GARCH models with heavy–tailed errors. *Econometrica* **71**, 285–317.

[16] Ibragimov, I.A. and Linnik, Yu.V. (1971) *Independent and Stationary Sequences of Random Variables.* Wolters–Noordhoff, Groningen.

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

[18] Leadbetter, M.R., Lindgren, G. and Rootzén, H. (1983) *Extremes and Related Properties of Random Sequences and Processes.* Springer, Berlin.

[19] Mikosch, T. (2003) Modeling dependence and tails of financial time series. In: Finkenstädt, B. and Rootzén, H. (Eds.) *Extreme Values in Finance, Telecommunications, and the Environment,* pp.185–286. Chapman and Hall, Boca Raton.

[20] Mikosch, T. and Stărică, C. (2000) Limit theory for the sample autocorrelations and extremes of a GARCH(1,1) process. *Ann. Statist.* **28**, 1427–1451. An extended version is available under www.math.ku.dk/∼mikosch.

[21] Mikosch, T. and Straumann, D. (2001) Whittle estimation in a heavy-tailed GARCH(1,1) model. *Stoch. Proc. Appl.*, **100**, 187–222.

[22] Mokkadem, A. (1990) Propriétés de mélange des processus autorégressifs polynomiaux. *Ann. Inst. H. Poincaré Probab. Statist.*, **26**, 219–260.

[23] Resnick, S.I. (1986) Point processes, regular variation and weak convergence. *Adv. Appl. Prob.* **18**, 66–138.

[24] Resnick, S.I. (1987) *Extreme Values, Regular Variation, and Point Processes.* Springer, New York.

[25] Rosenblatt, M. (1956) A central limit theorem and a strong mixing condition. *Proc. Nat. Acad. Sci. U.S.A.* **42**, 43–47.

[26] Rvačeva, E. L. (1962) On domains of attraction of multi-dimensional distributions. *Select. Transl. Math. Statist. and Probability* American Mathematical Society, Providence, R.I. **2**, 183–205.

[27] Samorodnitsky, G. and Taqqu, M.S. (1994) *Stable Non–Gaussian Random Processes. Stochastic Models with Infinite Variance.* Chapman and Hall, London.

[28] Straumann, D. and Mikosch, T. (2003) Quasi–maximum–likelihood estimation in conditionally heteroscedastic time series: a stochastic recurrence equations approach. To appear in *Ann. Statist.*

Laboratory of Actuarial Mathematics, Department of Applied Mathematics and Statistics, University of Copenhagen, Universitetsparken 5, DK-2100 Copenhagen Ø, Denmark, and Maphysto, the Danish Research Network for Mathematical Physics and Stochastics

Risklab, Department of Mathematics, ETH Zürich, ETH Zentrum, CH-8092 Zürich, Switzerland

E-mail address: `mikosch@math.ku.dk` *and* `strauman@math.ethz.ch`