ATHENS UNIVERSITY OF ECONOMICS AND BUSINESS

DEPARTMENT OF ECONOMICS

WORKING PAPER SERIES **8-2015**

Martingale Transforms with Mixed Stable Limits and the QMLE for Conditionally Heteroskedastic Models

Stelios Arvanitis Alexandros Louka

76 Patission Str., Athens 104 34, Greece
Tel. (++30) 210-8203911 - Fax: (++30) 210-8203301
www.econ.aueb.gr

Martingale Transforms with Mixed Stable Limits and the QMLE for Conditionally Heteroskedastic Models

Stelios Arvanitis[*] Alexandros Louka[†]

October 26, 2015

Dept. of Economics, AUEB, Patision str. 80, Athens P.O. Box 10434, GREECE.

Acknowledgement. This research was funded by the Research Centre of the Athens University of Economics and Business, in the framework of "Research Funding at AUEB for Excellence and Extroversion".

Abstract

We derive a limit theorem for appropriately centered and scaled martingale transforms $\sum_{i=1}^{n} \xi_i V_i$ to mixed-stable limits when (ξ_i) is an iid sequence in the domain of attraction of an α-stable distribution where $\alpha \in (0, 2]$. Using the Principle of Conditioning we recover and extend known results in the literature while imposing weaker conditions. The results are particularly useful in determining the limit theory of the Gaussian QMLE in conditionally heteroskedastic models when the squared innovations are heavy-tailed. We provide the framework for the QMLE limit theory which in the ergodic case is based on the stochastic recurrence approach used in the relevant literature and we furthermore allow for the parameter vector to lie on the boundary of the parameter space. Then we show that the QMLE weakly converges to an α-stable distribution when $\alpha \in [1, 2]$ and is inconsistent when $\alpha < 1$. We relax the assumption on ergodicity and provide analogous results for the QMLE in the non-stationary GARCH(1,1) case. We investigate the limit theory of the usual Wald statistic and provide with the asymptotic exactness and consistency of the relevant testing procedure based on subsampling. In the context of the stationary GARCH(1,1) we construct a testing procedure for weak stationarity and derive its asymptotic properties and numerically evaluate its performance.

[*]Email address: stelios@aueb.gr
[†]Email address: loukaalex@aueb.gr

1 Introduction

It is empirically known that distributions of financial asset returns exhibit fat tail behavior. Modeling the conditional moments of such processes using GARCH-type models has only partly explained this behavior and therefore considering heavy-tailed distributions for the innovation process is of particular interest for applications in finance. The use of the Gaussian QMLE for the parameter estimation of such models is very convenient as it has been shown to be consistent and asymptotically normal under mild conditions and thus reducing the risk of model misspecification. However, asymptotic normality with the usual $\sqrt{n}$ rate breaks down when the fourth moment of the error process is infinite and diverges in a slowly varying fashion.

In the relative literature, Hall and Yao [18] obtained the asymptotic distribution of the QMLE in GARCH models by examining the asymptotic behavior of sums of the form $\sum_{i=1}^{n} \xi_i V_i$ where $(\xi_i)_{i \in \mathbb{N}}$ is an i.i.d. sequence and $(V_i)_{i \in \mathbb{N}}$ is a stationary ergodic sequence of essentially bounded random variables and the distribution of ξ_1 lies in the domain of attraction of an α-stable distribution (say S_α) with $\alpha \in [1, 2]$. Mikosch and Straumann [30] derive a limit theorem for martingale transforms when ξ_t is in S_α for $\alpha \in (1, 2)$, and $\alpha = 1$ under symmetry. They assume $\mathbb{E}|V|^{\alpha+\delta} < \infty$ and impose a mixing condition for V_i. Then they use the result to derive the limit theory of the QMLE in GARCH$(1, 1)$. Surgailis [40] derives an analogous limit theorem using the Principle of Conditioning and uses characteristic function expansions for distributions (see Ibragimov [25]) in the domain of normal attraction of α-stable distributions for $\alpha \in (1, 2)$. He assumes $\mathbb{E}|V_1|^{\alpha+\delta} < \infty$ and stationarity and ergodicity for $(V_i)_{i \in \mathbb{N}}$. Jakubowski [22] shows that Surgailis' result can be obtained by assuming $\mathbb{E}|V_1|^{\alpha} < \infty$ instead.

In this paper we extend the previous results and provide weaker conditions for a limit theorem for martingale transforms with mixed α-stable limit. We use Surgailis' approach to recover existing results with $\alpha \in (0, 2]$ while allowing for *non-normal* domains of attraction. The use of the Principle of Conditioning and the characteristic function expansions provided by Ibragimov and Linnik [25] for the cases where $\alpha \in (0, 1) \cup (1, 2]$ and Aaronson and Denker [1] for the case where $\alpha = 1$ enables us to impose relatively weak conditions on the sequence of V_i's in order to obtain as limits stable distributions. Then the rate of convergence will be $n^{1/\alpha} r_n^{1/\alpha}$ where r_n depends on the behavior of the slowly varying function that appears in the characteristic function expansion of ξ_0 and co-represents with α the tail behavior of the distribution.

Next, we provide the framework for the limit theory of the QMLE in conditionally heteroskedastic models that relies on Straumann's [38] stochastic recurrence equation (SRE) approach while allowing for distributions of the innovation process in the domain of attraction of an α-stable distribution. In doing so, we allow the true parameter vector to lie on the boundary of the parameter space motivated by the work of Andrews [4]. We derive the weak limit of the QMLE to an α-stable distribution when $\alpha \in [1, 2]$ and inconsistency when $\alpha < 1$.

Finally, we determine the limit behavior of the classical self-normalized Wald test when $\alpha \in (1, 2]$ by showing the joint convergence of the QMLE in the spirit of Hall and Yao

[18] and design a testing procedure for the existence of the unconditional variance in the GARCH(1,1) model using the method of subsampling (see Politis et al. [34]). Then we evaluate the performance of the previous testing procedure by means of Monte Carlo simulations.

The structure of the remaining paper is organized as follows. In the next section, the martingale limit theorem (MLT) is derived for martingale transforms with mixed α-stable limits. In the third section we provide the framework and the limit theory for the QMLE using the MLT. In the fourth section we investigate the limit theory of the usual Wald test and provide an example testing procedure and discuss its theoretical properties and its numerical performance.

2 A MLT with Mixed Stable Limits

Our framework is constructed upon a complete probability space $(\Omega, \mathcal{G}, \mathbb{P})$. In what follows the abbreviation $\mathbb{P}$ a.s. stands for an almost sure argument with respect to the underlying measure. We denote convergence in distribution of sequences of random elements with $\rightsquigarrow$, exponential almost sure convergence w.r.t. $\mathbb{P}$ with $\overset{\text{eas}}{\to}$, and the Painleve-Kuratowski limit of sequences of sets with $\mathrm{PK}-\lim$. All limits are considered as $n \to \infty$ unless otherwise specified. We are interested in the asymptotic behavior of the properly translated and scaled partial sums of a process of the form $(\xi_i V_i)_{i\in\mathbb{N}}$ which due to the properties of the constituent processes $(\xi_i)_{i\in\mathbb{N}}$ and $(V_i)_{i\in\mathbb{N}}$ can be abusively perceived as a multiplicative "martingale transform".[1] This transform is directly related to the form of the Quasi-Likelihood function in GARCH-type models. The following assumptions describe those properties. The first one specifies the first factor as an iid sequence with stationary distribution closely related to an α-stable law.

Assumption 1. *$(\xi_t)_{t\in\mathbb{N}}$ is an iid sequence of random variables, and the $\log$-characteristic function of the distribution of ξ_1 has the following local representation around zero:*

$$\begin{cases} \gamma it - c|t|^{\alpha} h(|t|^{-1}) \left(1 - i\beta \operatorname{sgn}(t) \tan\left(\frac{1}{2}\pi\alpha\right)\right) & , \alpha \in (0,1) \cup (1,2] \\ \left(\gamma + H(|t|^{-1})\right) it - c|t| h(|t|^{-1}) \left(1 - 2Ci\frac{\beta}{\pi} \operatorname{sgn}(t)\right) & , \alpha = 1 \end{cases} \tag{1}$$

where h is a slowly varying function at infinity associated with the tail behavior of the cdf, say F, of ξ_0 and $H(\lambda) = \int_0^{\lambda} \frac{x^2}{1+x^2} \left(1 - F(x) - F(-x)\right) dx = \int_0^{\lambda} \frac{x}{1+x^2} h(x)(2\beta c\pi^{-1} + k(x))dx$ where $k(x) \to 0$ as $x \to \infty$ (see Theorems 1 and 2 of Aaronson and Denker [1]). Also, $\beta \in [-1,1]$, $c \in \mathbb{R}_{++}$, $\gamma \in \mathbb{R}$ and $-C$ is the Euler-Mascheroni constant.

Remark 1. The representations appearing in (1) are equivalent to that the distribution of ξ_1 lies in the domain of attraction of an α-stable law, due to Theorem 2.6.5 of Ibragimov and Linnik [25] for $\alpha \neq 1$ and Theorem 2 of Aaronson and Denker [1] for $\alpha = 1$, i.e.

[1] The term is in some cases abusive due to the non-existence of appropriate moments for any or both the random variables appearing in the product. We adopt it in the spirit of Mikosch and Straumann [30].

when appropriately translated and scaled, the partial sums of $(\xi_i)_{i=1}^n$ weakly converge to α-stable random variables (see inter alia Remark 2 of the latter paper). This law has index of stability equal to α, skewness parameter equal to β and scale parameter equal to c. The parameter γ appearing in the local representations corresponds to location and it is equal to $\mathbb{E}[\xi_1]$ when $\alpha > 1$. The aforementioned Tauberian type theorems imply that α and the slowly varying function h represent the asymptotic behavior of the tails of the distribution of ξ_1. Hence they determine the form of the scaling in order to obtain the aforementioned weak limit. More precisely the scaling factor is of the form $\frac{1}{n^{1/\alpha} r^{1/\alpha}}$ where $(nr_n)^{-1/\alpha} = \inf\{x > 0 : x^\alpha h(x^{-1}) = 1/n\}$ which implies that $r_n = h^*(n)$ for all n where h^* is also slowly varying, i.e. r_n defines a slowly varying sequence (see Paragraph 2.2 of Ibragimov and Linnik [25] and Paragraph 1.9 of Bingham et al. [8]). When h converges then the distribution of ξ_0 is said to belong to the domain of *normal* attraction to the relevant α-stable law. Notice that when $\alpha < 2$ the possibility of $h(x) \to 0$ as $x \to +\infty$ is also allowed, something that permits the consideration of cases where $\mathbb{E}|\xi_1|^\alpha < +\infty$ which is precisely true if and only if $\int^{+\infty} \frac{h(x)}{x} dx$ converges, e.g. $h(x) = \log^{-2}(x)$. H is closely related to the truncated α-moment of ξ_1 (see Remark 1 of Aaronson and Denker [1]). The location parameter alone when $\alpha \neq 1$ and all the aforementioned parameters along with H and C when $\alpha = 1$ determine the form of the translating constants. Furthermore, when $\alpha = 2$ we have that h is the second truncated moment of ξ_1, i.e. $h(z) = \int_{-z}^{z} x^2 dF(x)$.

The second assumption concerns the asymptotic behavior of the partial sums of the properly transformed scaling process, as those appear inside a product of conditional expectations of the terms that appear in the local representation of the characteristic function of the partial sum of the martingale transform, when analogously scaled and translated.

Assumption 2. *$(V_t)_{t\in\mathbb{N}}$ is a non trivial $\mathbb{R}$-valued sequence of random elements. For the filtration $(F_t)_{t\in\mathbb{N}}$ with $\mathcal{F}_t \triangleq \sigma(\xi_t V_t, \xi_{t-1}V_{t-1}, \xi_{t-2}V_{t-2}, \ldots)$, ξ_t is independent of $\mathcal{F}_{t-1}$ and V_t is measurable w.r.t. $\mathcal{F}_{t-1}$. Furthermore*

$$\frac{1}{n}\sum_{i=1}^{n} |V_i|^\alpha \quad \to \quad \bar{v}_\alpha \quad \mathbb{P} \text{ a.s.} \tag{2}$$

$$\frac{1}{n}\sum_{i=1}^{n} \operatorname{sgn}(V_i)|V_i|^\alpha \quad \to \quad v_\alpha \quad \mathbb{P} \text{ a.s.} \tag{3}$$

where $v_\alpha, \bar{v}_\alpha$ are random variables assuming non-zero values $\mathbb{P}$ a.s.

Remark 2. Eq. (2) Assumption 2 along with uniform integrability, implies that $\lim_{\varepsilon\to\infty}\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} |V_i|^\alpha 1\{|V_i| < \varepsilon\} = \bar{v}_\alpha$ $\mathbb{P}$ a.s. and analogously for 3. Equivalently, $\lim_{\varepsilon\to\infty}\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} |V_i|^\alpha 1\{|V_i| \geq \varepsilon\} = 0$ $\mathbb{P}$ a.s. Those results can be the basis of useful truncation arguments for the proof of the main result of the current section.

Our results do not require the existence of higher than α moment for the scaling process in the case where this is stationary. Hence, in this respect they generalize the analogous results of Hall and Yao [18], Mikosch and Straumann [30], and Surgailis [40]. However it presently seems that we cannot easily get rid of the following assumption which posits the existence of the α moment.

Assumption 3. *Assume that $(V_t)_{t\in\mathbb{N}}$ is strictly stationary and let I denote its invariant σ-field. Furthermore*

$$\mathbb{E}|V_1|^\alpha < \infty. \tag{4}$$

Remark 3. Note that (2) and (3) can be implied in a variety of cases, notably:

- If Assumption 3 holds then (2) and (3) hold with $\bar{v}_\alpha \equiv \mathbb{E}\left[|V_1|^\alpha|I\right]$ and $v_\alpha \equiv \mathbb{E}\left[|V_1|^\alpha \operatorname{sgn}(V_1)|I\right]$ respectively, as Doob's Theorem applies (see Davidson [11] p. 196).
- If in addition to the above $(V_t)_{t\in\mathbb{N}}$ is ergodic then (2) and (3) hold with $\bar{v}_\alpha \equiv \mathbb{E}\left[|V_1|^\alpha\right]$ and $v_\alpha \equiv \mathbb{E}\left[|V_1|^\alpha \operatorname{sgn}(V_1)\right]$ respectively as Birkhoff's LLN applies.
- Suppose that $V_t \to v$, $\mathbb{P}$ a.s. as $t \to \infty$ where v is a random variable that is not zero with $\mathbb{P}$ probability 1. Then, by the Cesàro mean theorem (2) and (3) hold with $\bar{v}_\alpha \equiv |v|^\alpha$ and $v_a \equiv |v|^\alpha \operatorname{sgn}(v)$ respectively.

The next assumption essentially bounds the rate at which the running maximum of the absolute scaling process may diverge to infinity, by a rate closely related to the rate that we will acquire for the weak convergence of the partial sums of the martingale transform. This is among others useful for the local representation of the characteristic function appearing in equation 1, to be asymptotically usable for the derivation of the results, or the facilitation of several truncation arguments in the proof of the main theorem below.

Assumption 4. *Assume that for any $M > 0$*

$$\mathbb{P}\left(\max_{1\leq i\leq n}|V_i| > M r_n^{\frac{1}{\alpha}} n^{\frac{1}{\alpha}}\right) \to 0. \tag{5}$$

Remark 4. Assumption 4 implies that $\mathbb{P}\left[(\max|V_i|)^{-1} n^{1/\alpha} r_n^{1/\alpha} < M^{-1}\right] \to 0$ for any $M > 0$. Since this is true for a general M, there exists a sequence $M_n \to 0$ such that 5 still holds if we replace M by M_n (see e.g. Lemma 22 of chapter 7 in Pollard [35]). The latter implies that $(\max|V_i|)^{-1} n^{1/\alpha} r_n^{1/\alpha} \to \infty$ in $\mathbb{P}$ probability.

The following lemma provides a list of dependent, but sufficient conditions for 4. Essentially all of them work via the appropriate comparison of the tail behavior of the distribution of $|V_0|$ with the one of $|\xi_0|$. Other sufficient conditions can be established by restrictions on the dependence structure of the scaling process in conjunction with the existence of the a moment of the scaling process, in the stationary case.

Lemma 1. *Each of the following suffices for Assumption 4:*

1. *Assumptions 2 and 3 along with*

$$\mathbb{P}\left(|V_0| > x\right) = o\left(\mathbb{P}(|\xi_0| > x)\right) \text{ as } x \to \infty. \tag{6}$$

2. *Assumptions 1, 2, and 3 along with* $\mathbb{E}\left|\xi_1\right|^{\alpha} = \infty$.

3. *Assumptions 1, 2, and 3 hold and* $\mathbb{E}|V_1|^{\alpha+\delta} < \infty$ *for some* $\delta > 0$.

Proof. 1. We have that for $M > 0$, $\mathbb{P}\left(\max_{1\leq i\leq n}|V_i| > Mr_n^{\frac{1}{\alpha}}n^{\frac{1}{\alpha}}\right) = P(\cup_{i=1}^{n}\left\{|V_i| > Mr_n^{\frac{1}{\alpha}}n^{\frac{1}{\alpha}}\right\}) \leq \sum_{i=1}^{n}\mathbb{P}\left(|V_i| > Mr_n^{\frac{1}{\alpha}}n^{\frac{1}{\alpha}}\right) = n\mathbb{P}\left(|V_1| > Mr_n^{\frac{1}{\alpha}}n^{\frac{1}{\alpha}}\right)$ due to stationarity. Now, since $\mathbb{P}(|\xi_0| > x) = \frac{c_1+c_2+o(1)}{x^{\alpha}}h(x)$ as $x \to \infty$ for some non negative constants c_1, c_2 (see Theorem 2.6.1 of [25]) the latter equals $n\frac{c_1+c_2+o(1)}{M^{\alpha}r_n n}h(Mr_n^{\frac{1}{\alpha}}n^{\frac{1}{\alpha}})o(1) = o(1)$. 2. Observe that $\mathbb{E}|V_1|^{\alpha} \equiv \int_0^{\infty}\alpha t^{\alpha-1}\mathbb{P}(|V_1| > t)dt < \infty$ while $\mathbb{E}|\xi_1|^{\alpha} \equiv \int_0^{\infty}\alpha t^{\alpha-1}\mathbb{P}(|\xi_1| > t)dt = \infty$, which implies Condition 6. 3. Analogously to 2. the result follows by using the fact that $\mathbb{E}|\xi_1|^{\alpha+\delta} \equiv \int_0^{\infty}(\alpha+\delta)t^{\alpha+\delta-1}\mathbb{P}(|\xi_1| > t)dt = \infty$. □

Remark 5. Notice that if $h(x) \to c > 0$ or $\limsup h(x) = \infty$ as $x \to \infty$, then Assumption 3 implies Assumption 5 (see the following lemma). If $h(x) \to 0$ then (6) would be implied if $\mathbb{E}|V_1|^{\alpha+\delta} < \infty$ for some $\delta > 0$. In the case where the previous does not hold for any $\delta > 0$, whence $\mathbb{P}(|V_0| > x) = \frac{c_1^{\star}+c_2^{\star}+o(1)}{x^{\alpha}}h^{\star}(x)$ for an analogous pair of constants and $h^{\star}$ a slowly varying function such that $h^{\star}(x) \to 0$, (6) would be implied by $\frac{h^{\star}(x)}{h(x)} \to 0$ as $x \to \infty$. The fact that Assumption 4 can hold in cases where $\mathbb{E}|V_1|^{\alpha+\delta} = +\infty$ for any $\delta > 0$, essentially allows the extension of analogous results (see for example Mikosch and Straumann [30], or Surgailis [40], or Hall and Yao [18]) without the need to impose strict moment existence conditions.

The following assumption posits the existence of an auxiliary regularly varying function that is used for asymptotic comparison with the asymptotic behavior of h. Notice that essentially the existence of such a function, permits the non requirement of the existence for higher than the a moment for the scaling process.

Assumption 5. *For any* $\varepsilon > 0$ *and some increasing regularly varying function* $f : (0,\infty) \to (0,\infty)$,

$$\limsup_{n\to\infty}\frac{1}{h(n^{1/\alpha}r_n^{1/\alpha})}\sup_{x\in\left[\left(\max_{1\leq i\leq n}|V_i|\right)^{-1},\varepsilon^{-1}\right)}\frac{h(n^{1/\alpha}r_n^{1/\alpha}x)}{f(x^{-1})} < \infty \quad \mathbb{P}\text{ a.s.} \tag{7}$$

$$\limsup_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}|V_i|^{\alpha}f(|V_i|) < \infty \quad \mathbb{P}\text{ a.s.} \tag{8}$$

Remark 6. The fact that f is regularly varying and increasing implies that we can substitute $f(x^{-1})$ with $f(tx^{-1})$ for any $t > 0$ in (7) without affecting its applicability. Notice that when f is an appropriate power function, and Assumption 3 holds, then Assumption 5 is implied by $\mathbb{E}|V_1|^{\alpha+\delta} < \infty$ for some sufficient $\delta > 0$.

The following lemma is the basic tool we use to prove Theorem 1 below and it relies on the properties of the slowly varying function h.

Lemma 2. *For any $\varepsilon > 0$ and any slowly varying function h,*

$$\frac{1}{nr_n}\sum_{i=1}^{n}|V_i|^{\alpha}\,h\left(n^{1/\alpha}r_n^{1/\alpha}\,|tV_i|^{-1}\right)1\{|V_i|\le\varepsilon\} = \frac{1}{n}\sum_{i=1}^{n}|V_i|^{\alpha}\,1\{|V_i|\le\varepsilon\} + o(1).$$

Proof. Let $f(x) = x^{-\alpha}h(x)$. Then

$$\begin{aligned}&\frac{1}{nr_n}\sum_{i=1}^{n}|V_i|^{\alpha}\,h\left(n^{1/\alpha}r_n^{1/\alpha}\,|tV_i|^{-1}\right)1\{|V_i|\le\varepsilon\}\\ &= \frac{h\left(n^{1/\alpha}r_n^{1/\alpha}\,|t|^{-1}\right)}{r_n}\frac{1}{n}\sum_{i=1}^{n}\frac{f(n^{1/\alpha}r_n^{1/\alpha}|tV_i|^{-1})}{f(n^{1/\alpha}r_n^{1/\alpha}|t|^{-1})}1\{|V_i|\le\varepsilon\}\end{aligned}$$

where $\frac{h\left(n^{1/\alpha}r_n^{1/\alpha}|t|^{-1}\right)}{r_n} = \frac{h\left(n^{1/\alpha}r_n^{1/\alpha}|t|^{-1}\right)}{h\left(n^{1/\alpha}r_n^{1/\alpha}\right)}\frac{h\left(n^{1/\alpha}r_n^{1/\alpha}\right)}{r_n} \to 1$. Also, note that

$$\begin{aligned}&\frac{1}{n}\sum_{i=1}^{n}\left|\frac{f(n^{1/\alpha}r_n^{1/\alpha}|tV_i|^{-1})}{f(n^{1/\alpha}r_n^{1/\alpha}|t|^{-1})} - |V_i|^{\alpha}\right|1\{|V_i|\le\varepsilon\}\\ &\le \sup_{\lambda\in[\varepsilon^{-1},\infty)}\left|\frac{f(\lambda n^{1/\alpha}r_n^{1/\alpha}|t|^{-1})}{f(n^{1/\alpha}r_n^{1/\alpha}|t|^{-1})} - \lambda^{-\alpha}\right|\frac{1}{n}\sum_{i=1}^{n}1\{|V_i|\le\varepsilon\}\end{aligned}$$

which converges to 0 as $n\to\infty$ by the Uniform Convergence Theorem (UCT) for regularly varying functions (see Theorem 1.5.2 of [8]). □

Our last assumption again concerns the asymptotic behavior of partial sums of the scaling process, as those appear in the relevant terms that occur as conditional expectations based on the local representation of the characteristic function of ξ_0 when $\alpha = 1$, as established by Aaronson and Decker [1].

Assumption 6. *When $\alpha = 1$ assume that*

$$\begin{aligned}\limsup_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}|V_i|\log|V_i| &< \infty \quad \mathbb{P}\text{ a.s.},\\ \limsup_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}|V_i|\log|V_i|f(|V_i|) &< \infty \quad \mathbb{P}\text{ a.s.}.\end{aligned}$$

Remark 7. Again under the premises of Assumption 3, Assumption 6 would follow from $\mathbb{E}|V_1|^{\alpha+\delta} < \infty$ for some $\delta > 0$.

In what follows $S_\alpha(\beta, c, \gamma)$ denotes an α-stable distribution with parameters β, c, γ. Furthermore the notation $\mathbb{E}S_\alpha(\beta, c, \gamma)$ denotes the mixture of the distributions of $S_\alpha(\beta, c, \gamma)$ w.r.t. $\mathbb{P}$ given since the parameters are generally allowed to be $\mathcal{G}$-measurable functions. The main result is presented in the following theorem. It establishes the limiting behavior of the partial sums of the martingale transform, when appropriate scaled and translated.

Theorem 1. *Suppose that Assumptions 1, 2, 4, 5 and 6 hold.*

Then if $\alpha \neq 1$

$$\frac{1}{n^{1/\alpha} r_n^{1/\alpha}} \sum_{i=1}^{n} (\xi_i - \gamma) V_i \rightsquigarrow \mathbb{E} S_\alpha \left(\beta \frac{v_\alpha}{\bar{v}_\alpha}, c\bar{v}_\alpha, 0 \right) \tag{9}$$

and if $\alpha = 1$

$$\frac{1}{nr_n} \sum_{i=1}^{n} (\xi_i - \gamma - H(nr_n)) V_i - 2\beta c \pi^{-1} \left(C v_1 + \frac{1}{n} \sum_{i=1}^{n} V_i \log |V_i| \right) \rightsquigarrow \mathbb{E} S_1 \left(\beta \frac{v_1}{\bar{v}_1}, c\bar{v}_1, 0 \right). \tag{10}$$

Remark 8. ....

Proof. By the "Main Lemma for Sequences" of Jakubowski [22] the result would follow if we would prove that for all $t \in \mathbb{R}$

$$\prod_{i=1}^{n} \mathbb{E} \left(\exp \left(it \frac{1}{n^{\frac{1}{\alpha}} r_n^{\frac{1}{\alpha}}} \rho_{i,\alpha} \right) / \mathcal{F}_i \right) \tag{11}$$

converges in probability to the characteristic function of $S_\alpha \left(\beta \operatorname{sgn}(v), c\bar{v}_\alpha, 0 \right)$, where

$$\rho_{i,\alpha} = \begin{cases} (\xi_i - \gamma) V_i, & \alpha \neq 1 \\ (\xi_i - \gamma - H(nr_n)) V_i - r_n 2\beta c \pi^{-1} (C v_1 - \frac{1}{n} \sum_{i=1}^{n} V_i \log |V_i|), & \alpha = 1 \end{cases}$$

Let the representation described in Assumption 1 hold for all $t \in (-t_0, t_0)$, where $t_0 > 0$. Then notice that for any $t \neq 0$ by defining the event

$$C_{n,K} := \left\{ \omega \in \Omega : |V_i| \leq K_t \left(n r_n \right)^{\frac{1}{\alpha}} \quad , \forall i = 1, \ldots, n \right\}$$

where $K_t < \frac{t_0}{|t|}$, we have that $\mathbb{P}(C_{n,K}^c)$ which by Lemma 4 tends to 0 as $n \to \infty$. When $\alpha \neq 1$, due to Assumption 1 if $\omega \in C_{n,K}$ then

$$\textstyle\sum_{i=1}^{n} \log \mathbb{E} \left(\exp \left(it \frac{1}{n^{\frac{1}{\alpha}} r_n^{\frac{1}{\alpha}}} (\xi_i - \gamma) V_i \right) / \mathcal{G}_n \right)$$

equals

$$\begin{aligned} & -\frac{c|t|^\alpha}{nr_n} \sum_{i=1}^{n} |V_i|^\alpha h \left(n^{1/\alpha} r_n^{1/\alpha} |tV_i|^{-1} \right) \left(1 - i\beta \operatorname{sgn}(tV_i) \tan \left(\frac{1}{2} \pi \alpha \right) \right) \\ &= -\frac{c|t|^\alpha}{nr_n} \sum_{i=1}^{n} |V_i|^\alpha h \left(n^{1/\alpha} r_n^{1/\alpha} |tV_i|^{-1} \right) \\ & + \frac{|t|^\alpha}{nr_n} i\beta c \operatorname{sgn}(t) \tan \left(\frac{1}{2} \pi \alpha \right) \sum_{i=1}^{n} |V_i|^\alpha h \left(n^{1/\alpha} r_n^{1/\alpha} |tV_i|^{-1} \right) \operatorname{sgn}(V_i). \end{aligned}$$

Notice that

$$\frac{1}{nr_n}\sum_{i=1}^{n}|V_i|^{\alpha}h\left(n^{1/\alpha}r_n^{1/\alpha}|tV_i|^{-1}\right)=\frac{h(n^{1/\alpha}r_n^{1/\alpha})}{r_n}\frac{1}{n}\sum_{i=1}^{n}|V_i|^{\alpha}\frac{h\left(n^{1/\alpha}r_n^{1/\alpha}|t|^{-1}|V_i|^{-1}\right)}{h\left(n^{1/\alpha}r_n^{1/\alpha}\right)}1\{|V_i|\le\varepsilon\}$$
$$+\frac{h(n^{1/\alpha}r_n^{1/\alpha})}{r_n}\frac{1}{n}\sum_{i=1}^{n}|V_i|^{\alpha}\frac{h\left(n^{1/\alpha}r_n^{1/\alpha}|t|^{-1}|V_i|^{-1}\right)}{h\left(n^{1/\alpha}r_n^{1/\alpha}\right)}1\{|V_i|>\varepsilon\}.$$

By Lemma 2, the first term on the right hand side equals $\frac{1}{n}\sum_{i=1}^{n}|V_i|^{\alpha}1\{|V_i|\le\varepsilon\}+o(1)$, which converges to $\bar{v}_{\alpha}$ $\mathbb{P}$ a.s. if we let $\varepsilon\to\infty$. Regarding the second term we have that

$$\frac{1}{n}\sum_{i=1}^{n}|V_i|^{\alpha}\frac{h\left(n^{1/\alpha}r_n^{1/\alpha}|t|^{-1}|V_i|^{-1}\right)}{h\left(n^{1/\alpha}r_n^{1/\alpha}\right)}1\{|V_i|>\varepsilon\}$$
$$=\frac{1}{n}\sum_{i=1}^{n}|V_i|^{\alpha}f(|V_i|)1\{|V_i|>\varepsilon\}\frac{h\left(n^{1/\alpha}r_n^{1/\alpha}|t|^{-1}|V_i|^{-1}\right)}{h\left(n^{1/\alpha}r_n^{1/\alpha}\right)}\frac{1}{f(|V_i|)}$$
$$\le\sup_{x\in\left[\left(\max_{1\le i\le n}|V_i|\right)^{-1},\varepsilon^{-1}\right)}\frac{h\left(n^{1/\alpha}r_n^{1/\alpha}|t|^{-1}x\right)}{h\left(n^{1/\alpha}r_n^{1/\alpha}\right)}\frac{1}{f(x^{-1})}\frac{1}{n}\sum_{i=1}^{n}|V_i|^{\alpha}f(|V_i|)1\{|V_i|>\varepsilon\},$$

which, due to Assumption 5, converges to 0 $\mathbb{P}$ a.s. Combining the above results we have that

$$\frac{1}{nr_n}\sum_{i=1}^{n}|V_i|^{\alpha}h\left(n^{1/\alpha}r_n^{1/\alpha}|tV_i|^{-1}\right)\to\bar{v}_{\alpha}\quad\mathbb{P}\text{ a.s.}$$

When $\alpha=1$, by Assumption 1, if $\omega\in C_{n,K}$ then $\sum_{i=1}^{n}\log\mathbb{E}\left(\exp\left(it\frac{1}{nr_n}\left(\xi_i-\gamma-H(nr_n)\right)V_i\right)/\mathcal{G}_n\right)$ equals

$$-c|t|\frac{1}{nr_n}\sum|V_i|h\left(nr_n|tV_i|^{-1}\right)+i2\beta c\pi^{-1}Ct\frac{1}{nr_n}\sum V_ih\left(nr_n|tV_i|^{-1}\right)$$
$$+it\frac{1}{nr_n}\sum V_i\left[H\left(nr_n|tV_i|^{-1}\right)-H(nr_n)\right],$$

where the first two terms of the above expression can be treated analogously to obtain their $\mathbb{P}$ a.s. limit as

$$-c|t|\bar{v}_1+i2\beta c\pi^{-1}Ctv_1=-c\bar{v}_1|t|\left[1-i2\beta\pi^{-1}C\,\mathrm{sgn}(t)\frac{v_1}{\bar{v}_1}\right].$$

Regarding the third term, first notice that

$$H(k\lambda)-H(\lambda)=\int_{\lambda}^{k\lambda}\frac{x}{1+x^2}\left(c_1-c_2+\underbrace{k(x)}_{=o(1)}\right)h(x)dx$$
$$=\int_{1}^{k}\frac{\lambda^2x}{1+\lambda^2x^2}\left(c_1-c_2+k(\lambda x)\right)h(\lambda x)dx.$$

Then we have that

$$\frac{1}{nr_n}\sum V_i\left[H\left(nr_n\,|tV_i|^{-1}\right)-H(nr_n)\right]$$
$$=\frac{h(nr_n)}{r_n}\frac{1}{n}\sum V_i\int_1^{|tV_i|^{-1}}\frac{x}{\frac{1}{n^2r_n^2}+x^2}\frac{h(xnr_n)}{h(nr_n)}\left(\frac{2\beta c}{\pi}+k(nr_nx)\right)dx$$
$$=\frac{h(nr_n)}{r_n}\left(2\beta c\pi^{-1}+o(1)\right)\frac{1}{n}\sum V_i\int_1^{|tV_i|^{-1}}\frac{x}{\frac{1}{n^2r_n^2}+x^2}\frac{h(xnr_n)}{h(nr_n)}dx,$$

since for any constant A, $\sup_{x\in[(\max|V_i|)^{-1},A]}k(nr_n|t|^{-1}x)=k(nr_n|t|^{-1}x_n^*)$ for some x_n^*. Then due to Remark 4 $k(nr_n|t|^{-1}x_n^*)=o(1)$. Then, writing V_i as $V_i1\{|V_i|\le\varepsilon\}+V_i1\{|V_i|>\varepsilon\}$, first notice that

$$\frac{1}{n}\sum_{i=1}^{n}V_i1\{|V_i|\le\varepsilon\}\int_1^{|tV_i|^{-1}}\frac{x}{\frac{1}{n^2r_n^2}+x^2}\frac{h(xnr_n)}{h(nr_n)}dx$$
$$=\frac{1}{n}\sum_{i=1}^{n}V_i1\{|V_i|\le\varepsilon\}\int_1^{|tV_i|^{-1}}\frac{x}{\frac{1}{n^2r_n^2}+x^2}dx$$
$$+\frac{1}{n}\sum_{i=1}^{n}V_i1\{|V_i|\le\varepsilon\}\int_1^{|tV_i|^{-1}}\frac{x}{\frac{1}{n^2r_n^2}+x^2}\left(\frac{h(xnr_n)}{h(nr_n)}-1\right)dx.$$

Then, towards showing that the second term of the above expression is $o(1)$, note that for some $A_2=[a_1,a_2]$ with $0<a_1\le a_2$ and dependent on the choice of ε

$$\frac{1}{n}\sum_{i=1}^{n}V_i1\{|V_i|\le\varepsilon\}\int_1^{|tV_i|^{-1}}\frac{x}{\frac{1}{n^2r_n^2}+x^2}\left(\frac{h(xnr_n)}{h(nr_n)}-1\right)dx$$
$$\le\int_{x\in A_2}\frac{x}{\frac{1}{n^2r_n^2}+x^2}\left|\frac{h(xnr_n)}{h(nr_n)}-1\right|dx\frac{1}{n}\sum_{i=1}^{n}|V_i|\,1\{|V_i|\le\varepsilon\}$$
$$\to 0,$$

using the dominated convergence theorem and Assumption 2. Regarding the first term, first notice that

$$\int_1^{|tV_i|^{-1}}\frac{x}{\frac{1}{n^2r_n^2}+x^2}dx1\{|V_i|\le\varepsilon\}=\frac{1}{2}\left[\log\left(1+n^2r_n^2x^2\right)\right]_1^{|tV_i|^{-1}}1\{|V_i|\le\varepsilon\}$$
$$=\frac{1}{2}\log\left(\frac{1+n^2r_n^2|tV_i|^{-2}}{1+n^2r_n^2}\right)1\{|V_i|\le\varepsilon\}=\log|tV_i|^{-1}1\{|V_i|\le\varepsilon\}+o(1),$$

where the $o(1)$ term is also independent of V_i using the fact that

$$\sup_{x\in[(t\varepsilon)^{-1},\infty)}\left|\log\left(\frac{1+\lambda^2x}{1+\lambda^2}\right)-\log x\right|\to 0 \text{ as } \lambda\to\infty.$$

Therefore

$$\frac{1}{n}\sum_{i=1}^{n} V_i 1\{|V_i|\le\varepsilon\}\int_1^{|tV_i|^{-1}} \frac{x}{\frac{1}{n^2r_n^2}+x^2}\frac{h(xnr_n)}{h(nr_n)}dx$$
$$=\log\frac{1}{|t|}\frac{1}{n}\sum_{i=1}^{n} V_i 1\{|V_i|\le\varepsilon\}-\frac{1}{n}\sum_{i=1}^{n} V_i\log|V_i|1\{|V_i|\le\varepsilon\}+o(1).$$

Next, treating the analogous term with $V_i1\{|V_i|>\varepsilon\}$, notice that for $|tV_i|^{-1}<1$ (this can be assumed without loss of generality since ε can be chosen large enough) we have that

$$\int_1^{|tV_i|^{-1}} \frac{x}{\frac{1}{n^2r_n^2}+x^2}\frac{h(xnr_n)}{h(nr_n)}dx=\int_1^{|tV_i|^{-1}} \frac{x}{\frac{1}{n^2r_n^2}+x^2}\frac{f\left(|t|^{-1}x^{-1}\right)}{f\left(|t|^{-1}x^{-1}\right)}\frac{h(xnr_n)}{h(nr_n)}dx$$
$$\le f(|V_i|)\sup_{x\in\left[\left(\max_{1\le i\le n}|V_i|\right)^{-1},|t|\right)}\frac{1}{f(|t|^{-1}x^{-1})}\frac{h(xnr_n)}{h(nr_n)}\int_1^{|tV_i|^{-1}} \frac{x}{\frac{1}{n^2r_n^2}+x^2}dx$$
$$\le f(|V_i|)\sup_{x\in\left[\left(\max_{1\le i\le n}|V_i|\right)^{-1},|t|\right)}\frac{1}{f(|t|^{-1}x^{-1})}\frac{h(xnr_n)}{h(nr_n)}\int_1^{|tV_i|^{-1}} \frac{x}{\frac{1}{n^2r_n^2}+x^2}dx.$$

Then, since $\int_{|tV_i|^{-1}}^1 \frac{x}{\frac{1}{n^2r_n^2}+x^2}dx\le\log|tV_i|$ and due to Assumptions 5 and 6 we have that

$$\lim_{\varepsilon\to\infty}\limsup_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n} V_i 1\{|V_i|>\varepsilon\}\int_1^{|tV_i|^{-1}} \frac{x}{\frac{1}{n^2r_n^2}+x^2}\frac{h(xnr_n)}{h(nr_n)}dx=0.$$

Finally, combining the above results we obtain (10). □

Remark 9. For an easy example, consider the case where $\alpha=2$, whence $\beta=0$ and $\mathbb{E}S_2\left(0,\frac{1}{2}\bar{v}_\alpha,0\right)=\mathbb{E}\left[N\left(0,\frac{1}{2}\bar{v}_\alpha\right)\right]$. For instance, if $\xi_0\sim t_2$ then the result is specified as $\frac{1}{\sqrt{n\ln n}}\sum_{i=1}^n\xi_iV_i\rightsquigarrow\mathbb{E}\left[N\left(0,\bar{v}_\alpha\right)\right]$ by a simple calculation. Also the results can be can be easily extended when V_0 is $\mathbb{R}^d$-valued, via the use of the Cramér–Wold theorem. Then the limits are mixtures of multivariate α-stable distributions where the spectral measures are characterized by linear transformations from Theorem 2.3 of Gupta et al. [17]. Notice though that in such a case the normalizing rate must be the same across all the elements of random vector, i.e. our results do not support the case of non-trivial matrix normalization.

The following lemma describes conditions that allow for the non-consideration of the translating constants in the cases where $\alpha<1$.

Lemma 3. *When $\alpha<1$ the term $\frac{1}{n^{1/\alpha}r_n^{1/\alpha}}\gamma\sum_{i=1}^n V_i$ can be omitted from 9 when either of the following sufficient conditions hold for any $M>0$:*

$$\mathbb{P}\left(\max_{1\le i\le n}|V_i|>Mq_n^{\frac{1}{\alpha}}n^{\frac{1}{\alpha}}\right)\to 0, \tag{12}$$

where $q_n = O(r_n^{1/(1-\alpha)})$, *or*

$$\limsup_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} |V_i|^{\alpha+\delta} < \infty, \tag{13}$$

where $\delta > 0$.

Proof. Under (12), observe that

$$\frac{1}{n^{1/\alpha} r_n^{1/\alpha}} \sum |V_i| = \frac{1}{n^{1/\alpha} r_n^{1/\alpha}} \sum |V_i|^{\alpha} |V_i|^{1-\alpha} \le \frac{1}{n^{1/\alpha - 1} r_n^{1/\alpha}} \max |V_i|^{1-\alpha} \frac{1}{n} \sum |V_i|^{\alpha}$$
$$= \frac{q_n^{\frac{1-\alpha}{\alpha}}}{r_n^{1/\alpha}} \frac{1}{q_n^{\frac{1-\alpha}{\alpha}} n^{\frac{1-\alpha}{\alpha}}} \max |V_i|^{1-\alpha} \frac{1}{n} \sum |V_i|^{\alpha} \le \frac{q_n^{\frac{1-\alpha}{\alpha}}}{r_n^{1/\alpha}} M,$$

with $\mathbb{P}$ probability approaching 1 as $n \to \infty$. The result follows as we can choose M arbitrary small. Under (13), Note that for δ small enough (so that $\alpha+\delta < 1$), $\left(\frac{1}{n^{1/\alpha} r_n^{1/\alpha}} \sum_{i=1}^{n} |V_i|\right)^{\alpha+\delta} \le \frac{1}{n^{\frac{\delta}{\alpha}} r_n^{1+\frac{\delta}{\alpha}}} \frac{1}{n} \sum_{i=1}^{n} |V_i|^{\alpha+\delta}$ and the result follows since $n^k r_n \to \infty$ for any $k > 0$. □

Remark 10. Notice that we can choose $q_n = r_n$ in (12) to obtain (5) when $\liminf_{n\to\infty} r_n > 0$, e.g. when $r_n \to \infty$ or $r_n \to c > 0$.

The following corollary specializes the results of the previous theorem, in cases where stationarity for the scaling process need not hold. It shows that results such as the one in Theorem 1 of Arvanitis and Louka [6] are special cases of Theorem 1.

Corollary 1. *Suppose that Assumptions 1, 2 and 6 hold. Moreover assume that* $V_t \to v$, $\mathbb{P}$ *a.s. as* $t \to \infty$ *where* v *is a random variable. Then if* $\alpha \neq 1$

$$\frac{1}{n^{1/\alpha} r_n^{1/\alpha}} \sum_{i=1}^{n} (\xi_i - \gamma) V_i \rightsquigarrow \mathbb{E} S_{\alpha}\left(\beta \operatorname{sgn}(v), c|v|^{\alpha}, 0\right),$$

and if $\alpha = 1$

$$\frac{1}{n r_n} \sum_{i=1}^{n} \left(\xi_i - \gamma - H(n r_n)\right) V_i - 2\beta c \pi^{-1} v \left(C + \log |v|\right) \rightsquigarrow \mathbb{E} S_1\left(\beta \operatorname{sgn}(v), c|v|, 0\right).$$

Proof. By the Cesàro mean theorem we have that Conditions (2) and (3) hold with $\bar{v}_{\alpha} \equiv |v|^{\alpha}$ and $v_a \equiv |v|^{\alpha} \operatorname{sgn}(v)$ respectively. Also, Assumption 4 clearly holds in this case as $V_t \to v$ $\mathbb{P}$ a.s. Then Theorem 1 applies. □

In view of Remark 3 we also have the following result. It subsumes and extends the analogous results in Mikosch and Straumann [30], Surgailis [40], and Hall and Yao [18] since it allows for $\alpha < 1$, non-ergodicity for the scaling process and thereby mixed weak limits, and $\mathbb{E}|V_1|^{\alpha+\delta} = +\infty$ for any $\delta > 0$.

Corollary 2. *Suppose Assumptions 1-6 hold.*

Then if $\alpha \neq 1$

$$\frac{1}{n^{1/\alpha} r_n^{1/\alpha}} \sum_{i=1}^{n} (\xi_i - \gamma) V_i \rightsquigarrow \mathbb{E} S_\alpha \left(\beta \frac{\mathbb{E}\left[|V_1|^\alpha \operatorname{sgn}(V_1) | I\right]}{\mathbb{E}\left[|V_1|^\alpha | I\right]}, c\mathbb{E}\left[|V_1|^\alpha | I\right], 0 \right),$$

and if $\alpha = 1$

$$\begin{aligned} & \frac{1}{nr_n} \sum_{i=1}^{n} (\xi_i - \gamma - H(nr_n)) V_i - 2\beta c \pi^{-1} \left(C v_1 + \frac{1}{n} \sum_{i=1}^{n} V_i \log |V_i| \right) \\ \rightsquigarrow \quad & \mathbb{E} S_1 \left(\beta \frac{\mathbb{E}\left[|V_1| \operatorname{sgn}(V_1) | I\right]}{\mathbb{E}\left[|V_1| | I\right]}, c\mathbb{E}\left[|V_1| | I\right], 0 \right). \end{aligned}$$

3 Limit Theory of the QMLE

A major application of the theorem presented in the previous section concerns the characterization of the rate and the asymptotic distribution of the Gaussian QMLE in GARCH type models. In what follows we briefly describe the framework and derive the results. The derivations draw heavily on the theory developed by Straumann [38] as well as Wintenberger and Cai [44]. The differences correspond *first* to the fact that we allow for *the centralized squares* of the elements of the *structuring sequence* to lie in the domain of *non normal attraction to an α-stable distribution* and *second* to the parameter of interest to be on the boundary of the relevant parameter space.

The framework is structured as follows: first, we define the process as the unique stationary and ergodic solution of a stochastic recurrence system of equations, second we are occupied with the issue of existence, uniqueness, stationarity and ergodicity of the solution of a transformation of the aforementioned recurrence, that essentially enables the invertibility of the volatility process for any parameter value. This allows the approximation of the latter process, which is *latent*, by filters that are measurable functions of the *observed* heteroskedastic process (this is related to the notion of *observable invertibility* essentially appearing in Straumann -see definition 2 of Wintenberger and Cai). Third, we define the QMLE and given the previous, we describe sufficient conditions (e.g. existence of logarithmic moments and of universal lower bounds for the filtered processes) that establish its strong consistency. Finally, we are occupied with the issue of existence, uniqueness, stationarity and ergodicity of the solutions of recurrence equations that emerge by differentiating the previous equations, along with analogous (moment existence, linear independence etc.) conditions for those solutions that permit among others the application of the CLT of the previous section, and are in any case helpful for the establishment of the rate and the weak limit of the QMLE via the results in the last part of the Appendix.

The process Suppose that Θ is a compact subset of $\mathbb{R}$ and let θ_0 be an *arbitrary* member of Θ. Consider the conditionally heteroskedastic process (w.r.t. θ_0) defined by

$$\begin{cases} y_t = \sigma_t z_t \\ \sigma_t^2 = g_{\theta_0}\left(z_{t-1}, \dots, z_{t-p}, \sigma_{t-1}^2, \dots, \sigma_{t-l}^2\right) \end{cases}, t \in \mathbb{Z} \tag{14}$$

where the *structuring sequence* $(z_t)_{t\in\mathbb{Z}}$ is a process of iid random variables such that $\mathbb{E}z_0 = 0$ and $\mathbb{E}z_0^2 = 1$ whenever these quantities exist.

Remark 11. We use the usual convention regarding the first and second moments of z_0^2 whenever they exist, but we do not impose any assumption regarding their existence yet. As it is shown later, the introduction of such assumptions will affect the (in)consistency and asymptotic distribution of the QMLE.

Also, $g_{\cdot} \in \mathbb{C}\left(\Theta \times \mathbb{R}^p \times \mathbb{R}_+^q, \mathbb{R}_{++}\right)$ for any $\theta \in \Theta$ and $l = \max(p, q)$. Let

$$\Phi_{t,\theta_0}(x) \doteqdot \left(g_{\theta_0}\left(z_t, \dots, z_{t-p+1}, x_1, \dots, x_l\right), x_1, \dots, x_{l-1}\right)'.$$

Given the definition of $(z_t)_{t\in\mathbb{Z}}$ and the properties of g_{θ_0}, the sequence $\left(\Phi_{t,\theta_0}(x)\right)_{t\in\mathbb{Z}}$ is stationary and ergodic for any x due to Proposition 2.1.1 of Straumann [38].

Assumption 7. *Suppose that*

$$\mathbb{E}\log^+\left|g_{\theta_0}\left(z_0, \dots, z_{-p+1}, y_1, \dots, y_l\right)\right| < +\infty,$$

for some $y_1, \dots, y_l \in \mathbb{R}_{++}$, Φ_{t,θ_0} *is* $\mathbb{P}$ *a.s. Lipschitz w.r.t.* x *with coefficient* $\Lambda\left(\Phi_{t,\theta_0}\right)$ *that satisfies*

$$\mathbb{E}\log^+\Lambda\left(\Phi_{0,\theta_0}\right) < +\infty \text{ and for some } m \in \mathbb{N}^*, \mathbb{E}\log\Lambda\left(\Phi_{0,\theta_0}^{(m)}\right) < 0.$$

The previous assumption along with Theorem 2.6.1 of Straumann [38], imply that the stochastic recurrence equation (SRE) in (14) admits a unique (up to indistinguishability) stationary and ergodic solution $\left(\sigma_t^2\right)_{t\in\mathbb{Z}}$ and furthermore any other solution converges exponentially almost surely to this one as $t \to \infty$. Due to continuity those properties extend to the heteroskedastic process itself.

Continuous Invertibility and the $(h_t)_{t\in\mathbb{Z}}$ Process Given the described process, the next part of the framework concerns the issue of *continuous invertibility* (see Definition 4 of Wintenberger and Cai [44]). This is closely connected to the properties of the filtering of the latent volatility process and thereby to the optimization procedure on the relevant likelihood function. Consider g_θ from before along with the first equation of (14). Given the process $(y_t)_{t\in\mathbb{Z}}$ consider the following stochastic recursion

$$h_t(\theta) = g_\theta\left(\frac{y_{t-1}}{\sqrt{h_{t-1}(\theta)}}, \dots, \frac{y_{t-p-1}}{\sqrt{h_{t-p-1}(\theta)}}, h_{t-1}(\theta), \dots, h_{t-q-1}(\theta)\right), \tag{15}$$

where $t \in \mathbb{Z}$ and $\theta \in \Theta$. Likewise to the previous section consider

$$\Psi_{t,\theta}(x) \doteqdot \left(g_\theta \left(\frac{y_{t-1}}{\sqrt{x_1}}, \dots, \frac{y_{t-p-1}}{\sqrt{x_p}}, x_1, \dots, x_l \right), x_1, \dots, x_{l-1} \right).$$

Analogously, the sequence $\left(\Psi_{t,\theta}(x)\right)_{t\in\mathbb{Z}}$ is stationary and ergodic for any x, θ. The following assumption is essentially condition (**CI**) of Wintenberger and Cai [44].

Assumption 8. *Suppose that*

$$\mathbb{E}\log^+ \left(\sup_{\theta\in\Theta} \left| g_\theta \left(\frac{y_{t-1}}{\sqrt{x_1}}, \dots, \frac{y_{t-p-1}}{\sqrt{x_p}}, x_1, \dots, x_l \right) \right| \right) < +\infty,$$

for some $x_1, \dots, x_l \in \mathbb{R}_{++}$. $\Psi_{t,\theta}$ *is* $\mathbb{P}$ *a.s. Lipschitz w.r.t.* x *with coefficient* $\Lambda\left(\Psi_{t,\theta}\right)$ *that is* $\mathbb{P}$ *a.s. continuous w.r.t.* θ *and satisfies*

$$\mathbb{E}\log^+ \sup_{\theta\in\Theta} \Lambda\left(\Psi_{0,\theta}\right) < +\infty \text{ and for some } m \in \mathbb{N}^*, \mathbb{E}\log\Lambda\left(\Psi_{0,\theta}^{(m)}\right) < 0 \text{ for all } \theta \in \Theta.$$

The following Lemma summarizes some of the implications of the first pair of assumptions. It is essentially Theorem 3 of Wintenberger and Cai [44].

Lemma 4. *Under assumptions 7 and 8 for any* $\theta \in \Theta$ *there exists a unique stationary and ergodic solution* $\left(h_t(\theta)\right)_{t\in\mathbb{Z}}$ *to (15). Moreover* $h_t(\theta)$ *is continuous w.r.t.* θ. *Furthermore for any* $\theta \in \Theta$ *and any other solution to (15), say* $\left(\hat{h}_t(\theta)\right)_{t\in\mathbb{Z}}$, *there exists* $\varepsilon > 0$ *such that*

$$\sup_{\theta'\in B(\theta,\varepsilon)\cap\Theta} \left| h_t(\theta') - \hat{h}_t(\theta') \right| \overset{eas}{\to} 0.$$

.

This is extremely helpful since the actual evaluation at each parameter value, and thereby the computability of the optimization of the likelihood function, depends on solutions of (15) based on initial conditions. It implies that any such solution (that is in general non stationary due to its dependence on initial conditions) will converge to the stationary and ergodic solution fast enough as $t \to \infty$. The local uniformity of the approximation, the stationarity and ergodicity of the solution, along with some moment existence could imply the convergence of arithmetic means of the $\left(\hat{h}_t(\theta)\right)_{t\in\mathbb{Z}}$ process evaluated at a convergent sequence to the expectation of the ergodic solution evaluated at the limit of the aforementioned sequence. All these will be convenient for the establishment of the asymptotic properties of the estimator.

The QMLE-Definition and Existence Given a finite sample $(y_t)_{t=1,\dots,n}$ from the heteroskedastic process, the following defines the Gaussian quasi likelihood function $\hat{c}_n$. The term is used in an abusive manner since the original function would be constructed as $-\frac{1}{2} * \hat{c}_n(\theta) + \text{const}$. This form enables the characterization of the QMLE as an approximate minimizer.

Assumption 9. *Suppose that $\varsigma_{k,\theta} : \Omega \to \mathbb{R}_{++}$ is measurable for any $\theta \in \Theta$ and $\mathbb{P}$ almost surely continuous w.r.t. θ for all $k = 0, \dots, l-1$ and, $\zeta_{k,\theta} : \Omega \to \mathbb{R}$ is measurable for any $\theta \in \Theta$ and $\mathbb{P}$ a.s. continuous w.r.t. θ for all $k = 0, \dots, p-1$.*

Definition 1. Define the filter $\left(\hat{h}_t(\theta)\right)_{t=1,\dots,n}$ for $\theta \in \Theta$ by

$$\hat{h}_k(\theta) = \varsigma_{k,\theta} \text{ when } k = 0, \dots, l-1 \text{ and } y_k = \zeta_{k,\theta} \text{ when } k = 0, \dots, p-1 \text{ and}$$
$$\hat{h}_t(\theta) = g_\theta\left(\frac{y_{t-1}}{\sqrt{\hat{h}_{t-1}(\theta)}}, \dots, \frac{y_{t-p-1}}{\sqrt{\hat{h}_{t-p-1}(\theta)}}, \hat{h}_{t-1}(\theta), \dots, \hat{h}_{t-q-1}(\theta)\right).$$

We can now define the Gaussian quasi likelihood function and the subsequent estimator, as a (possibly measurable selection) of its approximate $\arg\min$.

Definition 2. The Gaussian quasi likelihood function is

$$\hat{c}_n(\theta) = \frac{1}{n}\sum_{t=1}^{n} \hat{\ell}_t(\theta)$$

where

$$\hat{\ell}_t(\theta) = \log \hat{h}_t(\theta) + \frac{y_t^2}{\hat{h}_t(\theta)}.$$

For ε_n an $\mathbb{P}$ almost surely non negative random variable the QMLE θ_n is defined by

$$\hat{c}_n(\theta_n) \le \inf_{\theta\in\Theta} \hat{c}_n(\theta) + \varepsilon_n.$$

ε_n can be perceived as an *optimization error*, and thereby the definition is wide enough to include the estimator obtained (as is usually the case) by numerical optimization of $\hat{c}_n$. The $\mathbb{P}$ almost sure continuity (w.r.t. θ) of the filter, inherited by the definition of g_θ and assumption 9 along with the compactness and the separability of Θ imply the existence of θ_n even when $\varepsilon_n = 0$ $\mathbb{P}$ a.s. This is rigorously established in the proof of the following Proposition.

Proposition 1. *Suppose that Assumption 9 holds, then the QMLE exists.*

Proof. Notice first that $\hat{c}_n$ is a Caratheodory function, i.e. continuous w.r.t. θ (due to the continuity of the filter $\hat{h}_t(\theta)$) and point-wise measurable. Then the separability of K and lemma 4.51 of Aliprantis and Border [3] imply that $\hat{c}_n$ is jointly measurable. Furthermore it is proper (i.e. it does not attain the value $-\infty$ and there exists at least one $\theta \in K$ such that $c_n(\theta) \in \mathbb{R}$) since by $\hat{c}_n$ being a Gaussian quasi likelihood function it $\mathbb{P}$ a.s. does not attain the values $\pm\infty$ $\mathbb{P}$ a.s. This implies that it is a *proper normal integrand* in the sense of definition 3.5 (Ch. 5) of Molchanov [31] due to Proposition 3.6 (Ch. 5) in the same reference. The result now follows by the Theorem of Measurable Projections in van der Vaart and Wellner [42], example 1.7.5 p. 47, Proposition 3.10.i (Ch. 5-by setting $a = \inf_K \hat{c}_n + \varepsilon_n$) and the fundamental selection theorem (Theorem 2.13-Ch. 1) of Molchanov [31] (see also the proof of Theorem 3.24.(**i**)-Ch. 5 in the same reference). $\square$

3.1 Consistency of the QMLE

We turn to the limit theory for the estimator. The aforementioned *exponentially fast* approximation of the filter by the stationary and ergodic inverted process $(h_t)_{t\in\mathbb{Z}}$ (locally uniformly) along with the consequences of Assumption 7 enable the asymptotic approximation of $\hat{c}_n$ by an average of ergodic contributions obtained as

$$c_n(\theta) = \frac{1}{n}\sum_{t=1}^{n} \ell_t(\theta),$$

with

$$\ell_t(\theta) = \log h_t(\theta) + \frac{y_t^2}{h_t(\theta)}.$$

We can address c_n as the "ergodic likelihood". Several of its properties are appropriate approximations of analogous properties of $\hat{c}_n$ and thereby they will be used for the establishment of the limit theory. In this respect, given the previous, the following Assumptions provides with sufficient conditions for strong consistency.

Assumption 10. *For the innovation process assume that* $\mathbb{E}z_0^2 = 1$.

Assumption 11. *Suppose that:*

1. $\varepsilon_n \to 0$ $\mathbb{P}$ a.s.
2. $\mathbb{E}\log^+ \sigma_0^2 < +\infty$.
3. $\inf_\Theta h_0(\theta) \geq C > 0 \quad \mathbb{P}$ a.s.
4. For any $\theta \in \Theta$:
$$h_0(\theta) = \sigma_0^2 \Leftrightarrow \theta = \theta_0.$$

Condition 11.1 implies that the optimization error vanishes asymptotically. 11.2 requires the existence of logarithmic moments for the volatility process and due to the properties of z_0, it also implies that $\mathbb{E}\log^+ y_0^2 < \infty$. By Theorem 2 of Wintenberger and Cai [44] it follows from Assumption 7 and a condition of the form

$$\mathbb{E}\left(\log^+\left|g_{\theta_0}(z_0, \dots, z_{-p+1}, y_1, \dots, y_l)\right|\right)^2 < +\infty$$

for some $y \in \mathbb{R}_{++}$. Condition 11.3 requires the existence of a universal deterministic lower bound for the volatility processes that is naturally obtained in several GARCH-type models again due to the form of the recursion, the positivity constraints and the inclusion of a strictly positive constant. In more complex cases (e.g. the EGARCH model), it could be obtained by placing further restrictions on the parameter space. 11.4 is an identification condition that can be obtained by requiring more structure on the support of the distribution of z_0 as well as on the form of the defining recursion. The result is presented in the following theorem.

Theorem 2. *Suppose that Assumptions 7, 8, 9 , 10 and 11 hold. Then the QMLE is strongly consistent.*

Notice that Assumptions 7, 8, 9 along with conditions 11.2-4 are identical to the conditions C.1-C.4 of the relevant Theorem 5.3.1 of Straumann [38] (see the proof of the second part) or Theorem 4 of Wintenberger and Cai [44]. Hence Theorem 2 is essentially an *extension* by allowing the existence of an asymptotically negligible optimization error, and thereby by providing sufficient conditions for the *consistency of approximate optimizers* of the likelihood function.

Proof. Due to Assumptions 7, 8 and 8.C.2, Lemma 4 and Proposition 5.2.12 of Straumann [38] imply that for any $\theta \in \Theta$ there exists an $\varepsilon > 0$ such that

$$\sup_{\Theta \cap B(\theta,\varepsilon)} |c_n - \hat{c}_n| \to 0 \ \mathbb{P} \text{ a.s.} \tag{16}$$

due to Part 1.(i) of the proof of Theorem 5.3.1 of Straumann [38]. This locally uniform asymptotic approximation implies the analogous asymptotic approximation w.r.t. the topology of epi-convergence by the sequential characterization of the latter (see Definitions 2.1 and 2.2 of Lachout et al. [28]). This in turns implies that if $(c_n)_{n\in\mathbb{N}}$ epi-converges to a limit function, then so does $(\hat{c}_n)_{n\in\mathbb{N}}$ to the same limit. To this end, let

$$\rho_0 \doteqdot \inf_{\theta \in K} \left(\ln h_0(\theta) + \frac{z_0^2 \sigma_0^2(\theta_0)}{h_0(\theta)} \right),$$

and notice that

$$\begin{aligned}
&\mathbb{E}\left| \inf_{\theta \in K} \left(\ln h_0(\theta) + \frac{z_0^2 \sigma_0^2(\theta_0)}{h_0(\theta)} \right) \right| \\
&= -\mathbb{E}\rho_0 1_{\rho_0 \le 0} + \mathbb{E}\rho_0 1_{\rho_0 > 0} \\
&\le -\mathbb{E} \inf_{\theta \in K} \ln h_0(\theta)\, 1_{\rho_0 \le 0} + \mathbb{E} \ln \sigma_0^2 1_{\rho_0 > 0} + \mathbb{E} \frac{\sigma_0^2(\theta_0)}{h_0(\theta_0)} 1_{\rho_0 > 0} \\
&\le C + \mathbb{E} \ln \sigma_0^2 1_{\rho_0 > 0},
\end{aligned}$$

for some $C > 0$ that exists due to Assumption 11.3-4. Similarly since σ_0^2 is bounded away from zero and due to 11.3, $\mathbb{E} \ln \sigma_0^2 1_{\rho_0 > 0} < +\infty$. Then due to Part 1.(iii) of the proof of Theorem 5.3.1 of Straumann [38] implies that $\theta_0 = \arg\min_{\Theta} \mathbb{E} \left(\ln h_0(\theta) + \frac{y_0^2}{h_0(\theta)} \right)$. Hence taking also into account 11.1 we have that Lemma 6 is applicable. $\square$

In the next section, we derive the rate of convergence and asymptotic distribution of the QMLE under the sufficient conditions we imposed earlier that ensure the consistency of the QMLE, which include the condition that $\mathbb{E}z_0^2 < \infty$. The latter together with Assumption 14.1 necessarily imposes that $\alpha \in [1,2]$[2] in which cases the moment condition can be satisfied.

3.2 Rate and Asymptotic Distribution

The remaining elements of the limit theory, i.e. the rate and the limiting distribution can be established by conditions that are local in nature. The results depend crucially on the asymptotic existence of a local to θ_0 quadratic approximation of c_n^*, as required by Theorem 5. In accordance with the differentiability properties of $\widehat{h}_t$ for a variety of heteroskedastic models, we will assume that the approximation has the form of a second order Taylor expansion. Hence due to the possibility of θ_0 being on the boundary of Θ we will need a form of differentiability for the filter (and the subsequent stationary and ergodic approximation) that is consistent with this. We will use the notion of *left/right* (*l/r*) partial derivatives as in paragraph 3.3. of Andrews [4]. This requires some further structure on the set on which θ_n at least asymptotically attains its values. The following Assumption takes care of those concepts.

Assumption 12. *Suppose that:*

1. *For some $\eta \leq \frac{\varepsilon}{m}$ for some $1 < m \in \mathbb{N}$ and the $\varepsilon > 0$ that corresponds to θ_0 in Lemma 4, $\Theta \cap \bar{B}(\theta_0, \eta)$ coincides with the closure of its interior. Furthermore, $\Theta \cap \bar{B}(\theta_0, \eta) - \theta_0$ equals the intersection of a union of orthants and an open cube.*

2. *The function*

$$(\theta^T, x_1, \ldots, x_l) \to g_\theta \left(\frac{y_1}{\sqrt{x_1}}, \ldots, \frac{y_p}{\sqrt{x_p}}, x_1, \ldots, x_l \right)$$

has continuous second order (l/r) partial derivatives differentiable on $\Theta \cap \bar{B}(\theta_0, \eta) \times \mathbb{R}^l_{++}$ for every fixed $(y_1, \ldots, y_p) \in \mathbb{R}^p$.

3. *The functions $\varsigma_{k,\theta}$ and $\zeta_{k,\theta}$ have continuous second order (l/r) partial derivatives on $\Theta \cap \bar{B}(\theta_0, \eta)$, $\mathbb{P}$ a.s., for all $k = 0, \ldots, l-1$ and $k = 0, \ldots, p-1$.*

12.1 ensures that at any point of $\Theta \cap \bar{B}(\theta_0, \eta)$, there exists enough space around each of its elements so that a left and/or right perturbation can be defined, and its second part is essentially Assumption 2^{2*}.(a) of Andrews [4]. This implies that at any such point a left and/or right partial derivative could be in principle defined. 12.2 and 12.3 ensure that both g_θ and the initial conditions have well defined and continuous left and/or right second order partial derivatives. Given those, the Taylor approximation is valid on any K that is a non

[2] Note that $\alpha > 1$ implies $\mathbb{E}z_0^2 < \infty$.

empty compact subset of $\Theta \cap \bar{B}(\theta_0, \eta)$ even if the coefficients of the relevant polynomials may depend on random elements that can take values outside K with positive $\mathbb{P}$ probability. Furthermore, since the vector $(x_1, \ldots, x_l)$ belongs to $\mathbb{R}^l_{++}$ the relevant derivatives w.r.t. to the elements of this vector are by construction left and right. Due to the chain rule (see Appendix A. of Andrews [4]), they imply that the analogous derivatives of the filter (w.r.t. θ) are also well defined. In what follows we denote the matrices of first and second order (*l/r*) partial derivatives with $\cdot'$ and $\cdot''$ respectively. Their existence along with the form of $\hat{c}_n$ and Theorem 6 of Andrews [4] imply the $\mathbb{P}$ a.s. existence of a second order Taylor expansion of the likelihood function around θ_0. This does not suffice for the second part of Assumption 17 to hold, and thereby Theorem 5 cannot be directly used. The possibility of the existence, stationarity and ergodicity of h'_t and h''_t along with the possibility that they provide geometric approximations of $\hat{h}'_t$ and $\hat{h}''_t$ respectively could enable the verification of the aforementioned conditions. The following Assumption and the subsequent Proposition takes care of this after the establishment of some notation.

Let k_i be the i-th element of the vector $(\theta^T, x_1, \ldots, x_l)$. Then for $i, j = 1, \ldots, d, \ldots, d+l$ define

$$\partial^i \psi_t\left(\theta^T, x_1, \ldots, x_l\right) = \frac{\partial}{\partial k_i} g_\theta\left(\frac{y_t}{\sqrt{x_1}}, \ldots, \frac{y_{t-p+1}}{\sqrt{x_p}}, x_1, \ldots, x_l\right)$$

and

$$\partial^{i,j} \psi_t\left(\theta^T, x_1, \ldots, x_l\right) = \frac{\partial^2}{\partial k_i \partial k_j} g_\theta\left(\frac{y_t}{\sqrt{x_1}}, \ldots, \frac{y_{t-p+1}}{\sqrt{x_p}}, x_1, \ldots, x_l\right).$$

Assumption 13. *Suppose that:*

1. *for* $i = 1, \ldots, d, \ldots, d+l$

$$\mathbb{E}\left[\ln^+\left(\sup_{\theta \in \Theta \cap \bar{B}(\theta_0, \eta)} \left|\partial^i \psi_0\left(\theta, h_0(\theta), \ldots, h_{-l}(\theta)\right)\right|\right)\right] < +\infty$$

Furthermore for every $i = 1, \ldots, d, \ldots, d+l$*, there exist a stationary sequence* $\left(\bar{C}_{i,1}(t)\right)$ *with* $\mathbb{E}\left[\ln^+ \bar{C}_{i,1}(0)\right] < \infty$ *and some function* $r_1 : \mathbb{R} \to \mathbb{R}_+$ *that is continuously differentiable in a compact neighborhood of zero and* $r_1(0) = 0$ *such that*

$$\sup_{\theta \in \Theta \cap \bar{B}(\theta_0, \eta)} \left|\partial^i \psi_t\left(\theta^T, x_1, \ldots, x_l\right) - \partial^i \psi_t\left(\theta^T, x'_1, \ldots, x'_l\right)\right| \leq \bar{C}_{i,1}(t)\, r_1\left(|x - x'|\right), \tag{17}$$

where $x = (x_1, \ldots, x_l)$ *and* $x' = (x'_1, \ldots, x'_l)$ *in* $\left(\mathbb{R}^l\right)^{++}$.

2. *For* $i, j = 1, \dots, p, \dots, p+q$

$$\mathbb{E}\left[\ln^{+}\left(\sup_{\theta\in\bar{B}(\theta_0,\eta)}\left|\partial^{i,j}\psi_0\left(\theta^T, h_0(\theta), \dots, h_{-l}(\theta)\right)\right|\right)\right] < +\infty,$$

$$\textit{and}\ \mathbb{E}\ln^{+}\left(\sup_{\theta\in\bar{B}(\theta_0,\eta)}\left|h_0'(\theta)\right|\right) < +\infty.$$

Furthermore for every $i, j = 1, \dots, d, \dots, d+l$*, there exists a stationary sequence* $\left(\bar{C}_{i,j,2}(t)\right)$ *with* $\mathbb{E}\left[\ln^{+}\bar{C}_{i,j,2}(0)\right] < \infty$ *and some function* $r_2 : \mathbb{R} \to \mathbb{R}_{+}$ *that is continuously differentiable in a compact neighborhood of zero and* $r_2(0) = 0$ *such that*

$$\sup_{\theta\in\Theta\cap\bar{B}(\theta_0,\eta)}\left|\partial^{i,j}\psi_t\left(\theta^T, x_1, \dots, x_l\right) - \partial^{i,j}\psi_t\left(\theta^T, x_1', \dots, x_l'\right)\right| \leq \bar{C}_{i,j,2}(t)\, r_2\left(|x - x'|\right). \tag{18}$$

This Assumption essentially implies the existence and uniqueness of stationary and ergodic solutions to the SRE's obtained by (*l/r*) first and second order differentiation of the second equation in (14) w.r.t. θ. Furthermore, first those solutions are identified with h_t' and h_t'' which are continuous w.r.t the parameter and $\widehat{h}_t'$, $\widehat{h}_t''$ rapidly converge to their ergodic version uniformly in a neighborhood of θ_0 which without any damage to generality and for notational simplicity we assume that it coincides with $\Theta \cap \bar{B}(\theta_0, \eta)$. The derivation of the previous along with their implications on the asymptotic relation between the Taylor expansions of $\hat{c}_n$ and c_n are obtained in the proof of the following Lemma.

Lemma 5. *Suppose that Assumptions 7, 8, 9, 12 and 13 hold. Then*

1. h_t' *and* h_t'' *are continuous w.r.t.* θ*, for all* $t \in \mathbb{Z}$*,*

$$\sup_{\Theta\cap\bar{B}(\theta_0,\eta)}\left\|h_t'(\theta) - \widehat{h}_t'(\theta)\right\| \overset{e.a.s.}{\to} 0 \ \textit{and} \sup_{\Theta\cap\bar{B}(\theta_0,\eta)}\left\|h_t''(\theta) - \widehat{h}_t''(\theta)\right\| \overset{e.a.s.}{\to} 0\textit{, and} \tag{19}$$

$$n\sup_{\Theta\cap\bar{B}(\theta_0,\eta)}\|c_n'(\theta) - \hat{c}_n'(\theta)\| \ \textit{and}\ n\sup_{\Theta\cap\bar{B}(\theta_0,\eta)}\|c_n''(\theta) - \hat{c}_n''(\theta)\| \ \textit{converge}\ \mathbb{P}\ \textit{a.s.} \tag{20}$$

2. *If for some* $(w_n)_{n\in\mathbb{N}}$ *such that* $w_n \to +\infty$*, with* $w_n = o(n)$ *and* $w_n c_n'(\theta_0) \rightsquigarrow z_{\theta_0}$ *for* z_{θ_0} *some well defined random vector, then* $w_n\hat{c}_n'(\theta_0) \rightsquigarrow z_{\theta_0}$*, and*

3. *If* $\mathbb{E}\sup_{\theta\in\bar{B}(\theta_0,\eta)}\|\ell_0''(\theta)\| < +\infty$ *then for any sequence* $\vartheta_n \to \theta_0$ $\mathbb{P}$ *a.s.,* $\hat{c}_n''(\vartheta_n) \rightsquigarrow \mathbf{J}_{\theta_0} = \mathbb{E}\ell_0''(\theta_0) = \mathbb{E}\left(\frac{h_t'(\theta_0)\left[h_t'(\theta_0)\right]^T}{\sigma_t^4}\right)$.

Proof. **1**. The implications in (19) follow in an essentially similar manner to the proofs of Propositions 5.5.1 and 5.5.2 of Straumann [38] (with the analogous use of the conventions formulated there in order to describe the SRE's that are constructed by differentiations). The differences to those proofs are the following. First Theorem 3 of Wintenberger and

Cai [44] is used in place of Proposition 5.2.12 of Straumann [38]. Second, (17), (18) are generalizations of the Hölder type continuity conditions imposed in the relevant results by Straumann. The continuous differentiability around zero also imply the implications of the conditions of Straumann by an application of the mean value theorem around zero. Third, the identification of the solutions of the SRE's obtained by differentiation with h_t' and h_t'' respectively is obtained by a lemma that prescribes that under uniform convergence and the existence of a uniform limit of the first derivatives the limit function is differentiable and the limit of the derivatives is the derivative of the limit. Via the results of Appendix A of Andrews [4], this can be also seen to hold for (*l/r*) derivatives. Given those results the first implication in (20) is obtained by an application of the mean value theorem to the function $f(a,b) = \frac{a}{b}\left(1 - \frac{y_t^2}{b}\right),\ a \in \mathbb{R},\ b > 0$, that in turn implies

$$\sup_K \left\|\widehat{\ell}_t{}'(\theta) - \ell_t'(\theta)\right\| \leq c\left(1 + y_t^2\right)\left[\sup_K \left|h_t - \widehat{h}_t\right| + \sup_K \left\|h_t' - \widehat{h}_t'\right\|\right]$$

for some $c > 0$, where K denotes a non empty compact subset of $\Theta \cap \bar{B}(\theta_0, \eta)$. The previous along with $\mathbb{E}\ln^+ y_t^2 < +\infty$ and Proposition 2.5.1 of Straumann [38] show that

$$n \sup_K \left\|\widehat{c}_n'(\theta) - c_n'(\theta)\right\| \leq \sum_{t=1}^{\infty} \sup_K \left\|\widehat{\ell}_t{}'(\theta) - \ell_t'(\theta)\right\| < +\infty$$

For the second implication we have that the triangle inequality and the mean value theorem for the functions $f(a,b) = \frac{a}{b}\left(1 - \frac{y_t^2}{b}\right)$ and $g(a,b) = \left(\frac{2y_t^2}{a} - 1\right)\frac{b}{a^2}$ imply

$$\begin{aligned}
&\sup_K \left\|\widehat{\ell}_t''(\theta) - \ell_t''(\theta)\right\| \leq c_1\left(1 + y_t^2\right)\left[\sup_K \left|h_t - \widehat{h}_t\right| + \sup_K \left\|h_t'' - \widehat{h}_t''\right\|\right] \\
&+ c_2\left(1 + y_t^2\right)\left[\sup_K \left|h_t - \widehat{h}_t\right| + \sup_K \left\|\widehat{h}_t'\left(\widehat{h}_t'\right)^T - h_t'\left(h_t'\right)^T\right\|\right].
\end{aligned}$$

for some $c_1, c_2 > 0$ which exist due to compactness of K and the uniform boundedness of the volatility filters away from zero. Analogously to the previous and due to the fact

$$n \sup_K \left\|c_n''(\theta) - \widehat{c}_n''(\theta)\right\| \leq \sum_{t=1}^{\infty} \sup_K \left\|\ell_t''(\theta) - \widehat{\ell}_t''(\theta)\right\| < \infty,$$

we obtain the needed result.

2. It is obtained by the first implication in (20), the convergence in distribution of $r_n c_n'(\theta_0)$, the assumption that $\frac{r_n}{n} \to 0$ and the triangle inequality.

3. Follows directly form the triangle inequality and the ergodic ULLN. □

In order to be able to use the results in 1, 5 and 5 for the characterization of the rate and the limit distribution we need a final Assumption that takes care of the asymptotic behavior of c_n' and c_n'' as well as of the epigraph of the local polynomial approximation of the likelihood function. In what follows K denotes a compact non empty subset of Θ of possibly small enough diameter that *contains* θ_0 and is a subset of $\Theta \cap \bar{B}(\theta_0, \eta)$, such that $\theta_n \in K$ with $\mathbb{P}$-probability that converges to one as $n \to \infty$. Given Theorem 2, K could for example be chosen as $\Theta \cap \bar{B}(\theta_0, \eta)$ itself. Furthermore, let $\mathcal{H}_n(\alpha) = \frac{n^{(\alpha-1)/\alpha}}{r_n^{1/\alpha}}(K - \theta_0)$ where r_n is as in Remark 1. The asymptotic parameter space is defined next as an appropriate limit of $\mathcal{H}_n$.

Definition 3. $\mathcal{H}(\alpha) = \text{PK} - \lim_{n\to\infty} \mathcal{H}_n(\alpha)$ i.e. it is the set containing any $x \in \mathbb{R}^d$ such that x is a cluster or a limit point of some $(x_n)_{n\in\mathbb{N}}$ with $x_n \in \mathcal{H}_n(\alpha)$.

$\mathcal{H}(\alpha)$ is essentially the limit in the Painleve-Kuratowski sense of $(\mathcal{H}_n(\alpha))_{n\in\mathbb{N}}$ (see for example Appendix B of Molchanov [31]). The definition is equivalent to that $x \in \mathcal{H}(\alpha)$ iff there exists an infinite subset of $\mathbb{N}$ (say $\mathcal{N}$) and a cofinite subset of of $\mathbb{N}$ (say $\mathcal{N}^\star$) such that for any $\varepsilon > 0$, $\mathcal{H}(\alpha) \cap B(x, \varepsilon) \neq \emptyset$ for all $n \in \mathcal{N}$ and $n \in \mathcal{N}^\star$. Notice that when $\mathcal{H}(\alpha)$ exists then it is a closed subset of $\mathbb{R}^d$ (see Proposition 4.4 of Rockafellar and Wets [36]). In our case, upon existence it always contains 0. When θ_0 is an interior point then $\mathcal{H}(\alpha) = \mathbb{R}^d$. This definition is not less general compared to Assumption 5 of Andrews [4] as Lemma 3.8 of Arvanitis and Louka [5] implies.

Assumption 14. *Suppose that:*

1. *z_0^2 lies in the domain of attraction of an (non-degenerate) α-stable distribution. Specifically, suppose that Assumption 1 holds for $\xi_0 = z_0^2 - 1$.*

2. $\mathbb{E}\left\|\frac{h_0'(\theta_0)}{\sigma_0^2}\right\|^\alpha < +\infty$.

3. *For some $\eta > 0$, $\mathbb{E}\sup_{\theta\in\bar{B}(\theta_0,\eta)} \|\ell_0''\| < \infty$.*

4. *The components of the vector* $\frac{\partial}{\partial\theta} g_\theta \left(\frac{y_{t-1}}{\sqrt{h_{t-1}(\theta)}}, \dots, \frac{y_{t-p-1}}{\sqrt{h_{t-p-1}(\theta)}}, h_{t-1}(\theta), \dots, h_{t-q-1}(\theta) \right)\bigg|_{\theta=\theta_0}$ *are linearly independent random variables.*

5. *$\mathcal{H}(\alpha)$ is convex.*

14.3 also along with Lemma 5, Theorem 2 and the ULLN for stationary and ergodic sequences imply the convergence in probability of $\widehat{c}_n''(\theta_n)$ to $\mathbb{E}\ell_0''(\theta_0)$.

Remark 12. Also, notice that (where $\|\cdot\|_B$ denotes $\sup_{\theta\in\bar{B}(\theta_0,l)}\|\cdot\|$ for some $l > 0$)

$$\begin{aligned}
\mathbb{E}\left\|\ell''(\theta)\right\|_B &= \left\|\left(\frac{2y_t^2}{h_t(\theta)} - 1\right)\frac{h_t'(\theta)[h_t'(\theta)]^T}{[h_t(\theta)]^2} + \left(1 - \frac{y_t^2}{h_t(\theta)}\right)\frac{h_t''(\theta)}{h_t(\theta)}\right\|_B \\
&\leq \left\|\frac{h_0'}{h_0}\right\|_B^2 \left(2z_0^2\left\|\frac{\sigma_0^2}{h_0}\right\|_B + 1\right) + \left\|\frac{h_0''}{h_0}\right\|_B \left(1 + z_0^2\left\|\frac{\sigma_0^2}{h_0}\right\|_B\right),
\end{aligned}$$

which implies that a sufficient condition for 14.3 is the existence of $\eta > 0$ such that

$$\mathbb{E}\left[\sup_{\theta\in\bar{B}(\theta_0,\eta)}\left\|\frac{h_0'}{h_0}\right\|^{2\lambda}+\sup_{\theta\in\bar{B}(\theta_0,\eta)}\left\|\frac{h_0''}{h_0}\right\|^{\lambda'}+\sup_{\theta\in\bar{B}(\theta_0,\eta)}\left(\frac{\sigma_0^2}{h_0}\right)^{\lambda''}\right]<+\infty,$$

with $\lambda \geq 1$, $\lambda' > 1$, $\frac{\max(\lambda,\lambda')}{\max(\lambda,\lambda')-1} \leq \lambda'' \leq 2$. 14.4 implies that $\mathbb{E}\ell_0''(\theta_0)$ is positive definite.

14.1-2 enable the use of Theorem 1. 14.5 implies the uniqueness of the limit established in the final theorem and it is analogous to Assumption 6 of Andrews [4]. The following counterexample implies that condition 14.5 is not trivial by considering a K with empty interior.

Example (K is comprised by the elements and the limit of a converging sequence.)**.** Let $(\gamma_m)_{m\in\mathbb{Z}}$ denote a real sequence that converges to zero and suppose without loss of generality that $\theta_0 = 0$. For some $x \in \mathbb{R}^d$ and $c \neq 0$ let $K = K - \theta_0 = \left\{r_m^{1/\alpha} m^{-(\alpha-1)/\alpha} cx, m \geq 1\right\} \cup \{0\}$. Then $\mathcal{H}(\alpha) = \left\{\frac{c}{\sqrt{k}}x, k = 1, 2, ...\right\} \cup \{0\}$, so obviously 14.5 fails if $x \neq 0$.

If K itself contains a set of the form $\Theta \cap \bar{B}(\theta_0, \eta^*)$ with $0 < \eta^* \leq \eta$ then condition 14.5 implies that $\mathcal{H}(\alpha)$ coincides with the closure of its interior. This is due to the fact that $K - \theta_0$ must contain a neighborhood of zero of the form $\prod_{i=1}^{k}[l_i, u_i]$ where some of the lower or upper bounds could be zero but not simultaneously for the same i. Choose an arbitrary non zero point in the previous set. It is easy to see that this belongs to $\mathcal{H}_n(\alpha)$ for all n and thereby to $\mathcal{H}(\alpha)$ which is by construction convex.

We are now ready to state the main result of this section concerning the asymptotic distribution of the QMLE. First, we treat the case where $\alpha \in (0, 2]$.

Theorem 3. *Suppose that Assumptions 7, 8, 9, 11, 12, 13 and 14.1-4, where $\alpha \in (1, 2]$.*

If $\varepsilon_n = O_p\left(\frac{r_n^{2/\alpha}}{n^{2\frac{\alpha-1}{\alpha}}}\right)$ then

$$\left(\frac{n}{r_n}\right)^{\frac{1}{\alpha}}(\theta_n - \theta_0) = O_p(1).$$

If furthermore 14.5 holds and $\varepsilon_n = o_p\left(\frac{r_n^{2/\alpha}}{n^{2\frac{\alpha-1}{\alpha}}}\right)$, then

$$\left(\frac{n}{r_n}\right)^{\frac{1}{\alpha}}(\theta_n - \theta_0) \rightsquigarrow \tilde{h}_{\theta_0}$$

where $\tilde{h}_{\theta_0}$ is uniquely defined by $q\left(\tilde{h}_{\theta_0}\right) = \inf_{h\in\mathcal{H}} q(h)$ and $q(h) := \left(h - \mathbf{J}_{\theta_0}^{-1} z_{\theta_0}\right)' \mathbf{J}_{\theta_0}\left(h - \mathbf{J}_{\theta_0}^{-1} z_{\theta_0}\right)$ for $\mathbf{J}_{\theta_0} = \mathbb{E}\ell_0''(\theta_0) = \mathbb{E}\left(\sigma_0^{-4} h_0'(\theta_0)[h_0'(\theta_0)]^T\right)$ which is positive definite and z_{θ_0} follows a multivariate α-stable distribution characterized by all its projections as: $\lambda^T z_{\theta_0} \sim S_\alpha(\beta^*(\lambda), c^*(\lambda), 0)$ where $\beta^*(\lambda) = \beta\frac{\mathbb{E}\left[\left|\lambda^T h_0'(\theta_0)\right|^{\alpha}\operatorname{sgn}(\lambda^T h_0'(\theta_0))\right]}{\mathbb{E}\left[\left|\lambda^T h_0'(\theta_0)\right|^{\alpha}\right]}$ and $c^*(\lambda) = c\mathbb{E}\left[\frac{1}{\sigma_0^{2\alpha}}\left|\lambda^T h_0'(\theta_0)\right|^{\alpha}\right]$.

Proof. Theorem 2 and Lemma 5 imply that the result would hold via Theorem 5 if the following hold. Firstly, Conditions 14.1-2 and Theorem 1 imply that

$$w_n c_n'(\theta_0) = \frac{1}{n^{1/\alpha} r_n^{1/\alpha}} \sum_{t=1}^{n} (z_t^2 - 1) \frac{h_t'(\theta_0)}{\sigma_t^2}$$

where $w_n = \frac{n}{n^{1/\alpha} r_n^{1/\alpha}}$, converges in distribution to z_{θ_0} which is characterized above. Secondly, 14.3 implies the validity of the result in the third part of Lemma 5. Finally, the last condition of the second part of Assumption 17 follows from condition 14.3 along with lemma 5.6.3 of Straumann [38] while the third part of Assumption 17 is essentially 14.4. □

Remark 13. Note that in the case where $\alpha = 2$ we have that $z_{\theta_0} \sim N\left(0, \mathbf{J}_{\theta_0}\right)$. If furthermore θ_0 is an interior point then we have that $\sqrt{\frac{n}{r_n}}(\theta_n - \theta_0) \rightsquigarrow N\left(0, \mathbf{J}_{\theta_0}^{-1}\right)$ and if furthermore $\mathbb{E}z_0^4 < \infty$ the classical result is recovered as $r_n \rightsquigarrow \mathbb{E}z_1^4 - 1$. However, we can still obtain asymptotic normality with a different rate than $\sqrt{n}$. For example consider the case where $\sqrt{2}z_1 \sim t_4$ where simple calculations show that $r_n \sim \frac{2}{3}\log n$.

Next we examine the case where $\alpha = 1$ and $\mathbb{E}z_0^2 < \infty$ assuming that θ_0 lies in the interior of the parameter space.

Theorem 4. *Suppose that Assumptions 7, 8, 9, 10, 11, 12, 13 and 14.1-4, where $\alpha = 1$. Also, suppose that θ_0 lies in the interior of Θ and $\varepsilon_n = o_p\left(r_n^2\right)$. Then, for some $\bar{\theta}_n$ between θ_n and θ_0:*

$$\frac{1}{r_n}(\theta_n - \theta_0) - \frac{\gamma + H(nr_n)}{r_n}\left[c_n''(\bar{\theta}_n)\right]^{-1} \frac{1}{n}\sum_{i=1}^{n} \frac{h_t'(\theta_0)}{\sigma_t^2} \tag{21}$$
$$\rightsquigarrow \mathbf{J}_{\theta_0}^{-1} z_{\theta_0}$$

where $\mathbf{J}_{\theta_0} = \mathbb{E}\left[\sigma_0^{-4} h_0'(\theta_0)[h_0'(\theta_0)]^T\right]$ and z_{θ_0} follows a multivariate 1-stable distribution characterized by all its projections as: $\lambda^T z_{\theta_0} \sim S_1\left(\beta^*(\lambda), c^*(\lambda), \gamma^*(\lambda)\right)$ where $\beta^*(\lambda), c^*(\lambda)$ as in 3 and $\gamma^*(\lambda) = 2\beta c\pi^{-1}\left\{C\mathbb{E}\left[\sigma_0^{-2}\lambda^T h_0'(\theta_0)\right] + \mathbb{E}\left[\sigma_0^{-2}\lambda^T h_0'(\theta_0)\log\left|\sigma_0^{-2}\lambda^T h_0'(\theta_0)\right|\right]\right\}$.

Proof. Note that the fact that θ_0 is an interior point implies that $c_n'(\theta_n) = O_p\left(\sqrt{\varepsilon_n}\right) = o_p\left(r_n\right)$. Then

$$\begin{aligned}
c_n'(\theta_n) &= c_n'(\theta_0) + c_n''(\bar{\theta}_n)(\theta_n - \theta_0) \Rightarrow \\
c_n''(\bar{\theta}_n)(\theta_n - \theta_0) &= \frac{1}{n}\sum_{t=1}^{n}(z_t^2 - 1)\frac{h_t'(\theta_0)}{\sigma_t^2} + o_p(r_n) \\
&= \frac{1}{n}\sum_{t=1}^{n}(z_t^2 - 1 - \gamma - H(nr_n))\frac{h_t'(\theta_0)}{\sigma_t^2} \\
\ldots &+ (\gamma + H(nr_n))\frac{1}{n}\sum_{t=1}^{n}\frac{h_t'(\theta_0)}{\sigma_t^2} + o_p(r_n),
\end{aligned}$$

and the result follows by an application of Theorem 1. □

Remark 14. The above results show that stable limits for the QMLE can be obtained under fairly weak assumptions for a variety of conditionally heteroskedastic models. The assumption of ergodicity is easier to verify than the mixing condition which is imposed by Mikosch and Straumann [30]. Furthermore, the parameters of the limit distributions are analytically derived as functions of the parameters of the distribution of the innovation process and functionals of the volatility process and thus the stable distribution is fully characterized. Finally, the fact that Theorem 1 allows for $\mathbb{E}\left(z_0^2\right)^{\alpha+\delta} = +\infty$ for all $\delta > 0$ allows for the extension of the set of parameter values for which the results are valid in models such as the EGARCH $(1,1)$-see below.

3.3 Inconsistency and Non-tightness of the QMLE when $\alpha < 1$

Thus far, in order to derive the asymptotic distribution of the QMLE we worked under the assumption that z_0^2 lies in the domain of attraction of an α-stable and at the same time $\mathbb{E}z_0^2 < \infty$ which implies that $\alpha \geq 1$. Below we examine the asymptotic behavior of the QMLE when $\alpha < 1$. Clearly, Theorem 2 cannot be applied in this case since $\mathbb{E}z_0^2 = +\infty$. Notice that to our knowledge, the following result on the non-tghtness of the QMLE is new in the relevant literature.

Proposition 2. *Suppose that Assumptions 7, 8, 9, 11, 14.1 with $\alpha < 1$ hold. Also let $\varepsilon_n = o_p(n^{1-1/\alpha} r_n^{-1/\alpha})$. Furthermore for any $\theta \in \Theta$, $\exists \varepsilon_\theta > 0$ such that $\forall t \in \mathbb{Z} \sup_{\Theta \cap \bar{B}(\theta,\varepsilon_\theta)} \frac{\sigma_t^2}{h_t(\theta)} < C_\theta$ $\mathbb{P}$ a.s. and $\forall \theta \in \Theta\, \exists \theta' \in \Theta$ such that $h_0(\theta') > h_0(\theta)$ $\mathbb{P}$ a.s. then the QMLE is asymptotically non-tight and thus inconsistent.*

Proof. Similarly to the first part of the proof of Theorem 2, due to Eq. 16 it suffices to examine the asymptotic behavior of c_n (instead of its non-ergodic counterpart), or equivalently of

$$\begin{aligned}
C_n(\theta) &:= \frac{n}{n^{1/\alpha} r_n^{1/\alpha}} c_n(\theta) - \frac{1}{n^{1/\alpha} r_n^{1/\alpha}} \sum \log \sigma_t^2 \\
&= \frac{1}{n^{1/\alpha} r_n^{1/\alpha}} \sum z_t^2 \frac{\sigma_t^2}{h_t(\theta)} - \frac{1}{n^{1/\alpha} r_n^{1/\alpha}} \sum \log \frac{\sigma_t^2}{h_t(\theta)}
\end{aligned}$$

as θ_n is within ε_n distance from $\inf_{\theta \in \Theta} C_n(\theta)$. But, due to the assumed bounds for $\frac{\sigma_t^2}{h_t(\theta)}$ and since $\alpha < 1$, the ergodic uniform law of large numbers gives

$$\frac{1}{n^{1/\alpha} r_n^{1/\alpha}} \sum_{t=1}^{n} \log \frac{\sigma_t^2}{h_t(\theta)} \to 0 \quad \mathbb{P} \text{ a.s.}$$

locally uniformly $\forall \theta \in \Theta$. Next notice that, by Theorem 1 $\forall \theta \in \Theta$

$$\frac{1}{n^{1/\alpha} r_n^{1/\alpha}} \sum_{t=1}^{n} z_t^2 \frac{\sigma_t^2}{h_t(\theta)} \rightsquigarrow \mathbb{E}\left|\frac{\sigma_0^2}{h_0(\theta)}\right|^{\alpha} Z$$

where $Z \sim S_\alpha(1, c, 0)$ for some $c > 0$ is a random variable with positive support. Finally, notice that

$$\mathbb{P}\left(\sup_{\theta' \in \bar{B}(\theta,\varepsilon_\theta)} \sup_{\theta'' \in \bar{B}(\theta,\varepsilon_\theta)} \frac{1}{n^{1/\alpha} r_n^{1/\alpha}} \sum z_t^2 \sigma_t^2 \left| \frac{1}{h_t(\theta')} - \frac{1}{h_t(\theta'')} \right| > \varepsilon \right)$$

$$\leq \mathbb{P}\left(\sup_{\theta' \in \bar{B}(\theta,\varepsilon_\theta)} \sup_{\theta'' \in \bar{B}(\theta,\varepsilon_\theta)} \frac{1}{n^{1/\alpha} r_n^{1/\alpha}} \sum z_t^2 \sup_{\theta \in \bar{B}(\theta,\varepsilon_\theta)} \frac{\sigma_t^2}{h_t(\theta)} \frac{||h_t'||}{h_t} ||\theta' - \theta''|| > \varepsilon \right)$$

$$\leq \mathbb{P}\left(||\theta' - \theta''|| \frac{1}{n^{1/\alpha} r_n^{1/\alpha}} \sum z_t^2 > \frac{\varepsilon}{C_\theta^2} \right), \tag{22}$$

and since by Theorem 1 $\frac{1}{n^{1/\alpha} r_n^{1/\alpha}} \sum z_t^2$ is asymptotically tight hence for any $\eta > 0\ \exists M_\eta :$ $\limsup_{n\to\infty} \mathbb{P}\left(k_n > M_\eta\right) \leq \eta$ and for a given ε and η choose $\delta : \frac{\varepsilon}{C_\theta^2 \delta} \leq M_\eta \Leftrightarrow \delta \leq \frac{\varepsilon}{C_\theta^2 M_\eta}$ which is always possible. Hence Eq. 22 and the choice of $\delta(\varepsilon, \eta)$ imply stochastic equicontinuity and therefore the weak convergence of $C_n(\theta)$ to $C(\theta) := \mathbb{E}\left|\frac{\sigma_0^2}{h_0(\theta)}\right|^\alpha Z$ in $\ell^\infty\left[\bar{B}(\theta, \varepsilon_\theta)\right]$, which has no minimizers due to the fact that $\forall \theta \in \Theta, \exists \theta' : \mathbb{E}\left|\frac{\sigma_0^2}{h_0(\theta')}\right|^\alpha Z < \mathbb{E}\left|\frac{\sigma_0^2}{h_0(\theta)}\right|^\alpha Z$ by the monotinicity of the integral. Now, suppose that θ_n is asymptotically tight. Then by Prokhorov's Theorem, there exists a random element ϑ such that $\theta_{k_n} \rightsquigarrow \vartheta$ along some subsequence. Due to separability and Skorohod representation there exists a suitable probability space (say $\mathbb{P}^*$) and random elements $C_n^* \overset{d}{=} n^{1-\frac{1}{\alpha}} r_n^{-\frac{1}{\alpha}} C_n$, $C^*(\theta) \overset{d}{=} C(\theta)$, $\varepsilon_n^* = o_p\left(n^{-\frac{1-\alpha}{\alpha}} r_n^{-\frac{1}{\alpha}}\right)$ and $C_n^* \to C^*$ $\mathbb{P}^*$ a.s.. Also, $\theta_{k_n}^* \overset{d}{=} \theta_{k_n}$ where $\theta_{k_n}^*$ satisfies $C_{kn}^*\left(\theta_{k_n}^*\right) \leq \inf_{\theta\in\Theta} C_{kn}^*(\theta) + \varepsilon_{kn}^*$ and $\theta_{k_n}^* \to \theta^*$ where $\theta^* \overset{d}{=} \vartheta$ $\mathbb{P}$ a.s. But due to the Theorem 3.4 of Molchanov [**?**], $\theta^* \in \arg\max C^*$. Since θ^* has a well defined distribution, there exists a measurable selection say $T : \theta^* = T \circ \arg\max \circ C^*$. Hence $\theta^* = T\left(\arg\max(C^*)\right) \overset{d}{=} T\left(\arg\max(C)\right)$. Thereby $\vartheta \overset{d}{=} x$ for some $x \in \arg\max C$. But $\arg\max \ell = \emptyset$ which leads to contradiction. Thus θ_n is non-tight. □

3.4 Applications

The examples below concern the verification of the assumption framework above (given the a priori validity of the Assumption 9 and of the conditions 11.1, 12.3 and 14.3) for the GQARCH(1,1) model introduced by Sentana [39].

3.4.1 GQARCH(1,1)

Let $p = q = 1$, $g_\theta(z_{t-1}, x) = \omega + \alpha\left(z_{t-1}\sqrt{x} + \frac{\gamma}{2\alpha}\right)^2 + \beta x$ and Θ is a compact subset of $\mathbb{R}^{++} \times \mathbb{R}^{++} \times \mathbb{R}^{-} \times [0, 1)$. The results referenced below are established in Arvanitis and Louka [5]. Remark R.1 implies that there exist $\theta_0 \equiv (\omega_0, \alpha_0, \gamma_0, \beta_0) \in \Theta$ such that Assumption 7 is satisfied. This is permitted even in cases where $\alpha_0 + \beta_0 > 1$ implying the existence of solutions with the required properties that are not covariance stationary. Assumption 8 is satisfied since $\beta < 1$. For Assumption 11.2 and 11.4 see lemma 2.2 and lemma 3.3. For Assumption 11.3 see lemma 2.1. The latter holds if the distribution of z_0 is not concentrated in two points. For Assumption 12.1 see lemma 3.5 and for the rest of Assumption 12 and Assumption 13 see the proofs of lemmata 4.7 and 4.8. If $\mathbb{P}\left(z_0^2 \leq t\right) = o\left(t^{\mu}\right)$ for $t \downarrow 0$, then Assumption 14.2-4 follows from lemmata 4.2, 4.3 and 4.6 due to Remark 12. Examples of Θ's that satisfy condition 14.5 of Assumption 14 can also be found in section 3.3. of Arvanitis and Louka [5].

3.4.2 AGARCH(p,q)

Several other examples can be constructed using the work of Straumann [38]. For instance the verification of Assumptions 7, 8, 11, 12, 13 and 14.2-4, for the AGARCH(p,q) model when the parameter space is appropriately restricted and under the condition $\mathbb{P}\left(z_0^2 \leq t\right) = o\left(t^{\mu}\right)$ as $t \downarrow 0$, would follow from the results appearing in examples 5.2.5 and 5.2.11 and paragraphs 5.4.2 and 5.7.1 of Straumann [38] by noticing first that for the set $[\omega_{\min}, +\infty) \times [0, +\infty)^p \times B \times [-1, 1]$ for $\omega_{\max} > 0$ and $B = \left\{x \in \mathbb{R}_+^q : \sum_{i=1}^q x_i < 1\right\}$ Assumption 12.1 follows, second that Lemma 5.7.3 of Straumann [38] can be seen to hold even when θ_0 lies on the boundary of Θ which by construction is a compact subset of the previous set with non empty interior (as long as the conditions $a_{i_0} \neq 0$ for some $i = 1, 2, \ldots, p$ and $\left(a_{p_0}, \beta_{q_0}\right) \neq (0, 0)$-simply express the linear combination w.r.t. a_{i_0} instead of a_{1_0}) and third that this is also true for lemmata 5.1 and 3.2 of Berkes et al. [7] as well as for their extensions concerning the AGARCH(p,q) case, i.e. lemmata 5.7.4 and 5.7.5 of Straumann [38]. Notice that it is easy to construct examples of Θ for which both the previous assumptions and condition 14.5 of Assumption 14 hold. Consider for example the case where $p = 2$, $q = 1$ and $\Theta = [\omega_{\min}, \omega_{\max}] \times \left[0, a_{1_{\max}}\right] \times \left[a_{2_{\min}}, a_{2_{\max}}\right] \times [\beta_{\min}, \beta_{\max}] \times [-1, 1]$ with the obvious notation, where $a_{2_0} = a_{2_{\max}}$, $\gamma_0 = -1$, and the other elements of θ_0 lie in the interior of their defining intervals, in which case $\mathcal{H} = \mathbb{R}^2 \times (-\infty, 0] \times \mathbb{R} \times [0, +\infty)$. In such a case it is easy to see that $\widetilde{h}_{\theta_0} = \left(L'\right)^{-1} \left(z_1, z_2, \min\left\{0, z_3\right\}, z_4, \max\left\{0, z_5\right\}\right)'$ where $\left(z_1, z_2, z_3, z_4, z_5\right)' \sim N\left(0, \mathbf{I}_5\right)$ and $\mathbf{J}_{\theta_0} = LL'$.

3.4.3 EGARCH(1,1)

It is also easy to extend the assumption framework so that the recursions in (14) and 1 define not the volatility processes per se but their composition with some common *bijective* transformation. If Assumptions 7 and 8 hold w.r.t the transformed processes, Assumption 9 incorporates the condition that its inverse (the *link* function as termed by Wintenberger and Cai [44]) is continuous, and Assumption 13 is augmented by the condition that the inverse has first and second derivatives that are Lipschitz continuous on the bounded away from zero-due to condition 11.3- domain of the volatilities, then Theorem 3 would also hold. Furthermore, Assumption 14.3 can be avoided as Lemma 3 in [45] shows.

3.5 Dropping Stationarity: non-stationary GARCH(1,1)

In this section we show using a motivating example that the stationarity assumption which was used above is not crucial in determining the limit theory of the QMLE. We actually extend the results of Jensen and Rahbek [24], who derive the asymptotic normality of the QMLE under non-stationarity in the GARCH(1,1) model under the assumption of finite fourth moments of the innovation process, allowing for α-stable limits. To this end, given the GARCH(1,1) model with the time indice taking values in $\mathbb{N}$ and the parameter vector denoted by $\theta = (a, \beta, \omega, \gamma)^3$ as in [24] we impose the following assumption which replaces Assumptions 7-8. To our knowledge this extension is new in the relevant literature.

Assumption 15. $\mathbb{E}\log(\alpha_0 z_t^2 + \beta_0) \geq 0.$

Then Theorem 1 in [24] can be generalized as follows regarding the asymptotic distribution of the QMLE for (α_0, β_0).

Proposition 3. *Suppose that Assumptions 15 and 14.1,5 hold with $\alpha > 1$. Also fix (ω, γ) at their true values (ω_0, γ_0). Then the results of Theorem 3 hold with the following modifications:*

1. $\mathbf{J}_{\theta_0} = \begin{pmatrix} \frac{1}{\alpha_0} & \frac{\mu_1}{\alpha_0\beta_0(1-\mu_1)} \\ \frac{\mu_1}{\alpha_0\beta_0(1-\mu_1)} & \frac{(1+\mu_1)\mu_2}{\beta_0^2(1-\mu_1)(1-\mu_2)} \end{pmatrix}$ *where* $\mu_i = \mathbb{E}(\beta_0/(\alpha_0 z_t^2 + \beta_0))^i, i = 1, 2,$ *and*

2. z_{θ_0} *follows a multivariate α-stable distribution characterized by all its projections as*

$$\lambda^T z_{\theta_0} \sim \begin{cases} S_\alpha\left(\beta^*(\lambda), c^*(\lambda), 0\right) & \textit{when} \quad \alpha > 1 \\ S_1\left(\beta^*(\lambda), c^*(\lambda), \gamma^*(\lambda)\right) & \textit{when} \quad \alpha = 1 \end{cases}$$

where $\beta^*(\lambda) = \beta\frac{\mathbb{E}\left[\left|\lambda^T U\right|^\alpha \operatorname{sgn}(\lambda^T U)\right]}{\mathbb{E}\left[\left|\lambda^T U\right|^\alpha\right]}$, $c^*(\lambda) = c\mathbb{E}\left[\left|\lambda^T U\right|^\alpha\right]$, $\gamma^*(\lambda) = 2\beta c\pi^{-1}\left\{C\mathbb{E}\left[\lambda^T U\right] + \mathbb{E}\left[\lambda^T U \log\left|\lambda^T U\right|\right]\right\}$ *and* $U := \left(\frac{1}{a_0}, \frac{1}{\beta_0}\sum_{j=1}^\infty \beta_0^j \prod_{k=1}^j \frac{1}{a_0 z_{t-k}^2 + \beta_0}\right)$.

Furthermore, if Assumption 15 holds with strict inequality, then the above holds for any (ω, γ).

Proof. As in the proof of Theorem 3 the result for the case where $(\omega, \gamma) = (\omega_0, \gamma_0)$ can be shown to hold via Theorem 5. Then, exactly as in Jensen and Rahbek [24] (see proof of Lemma 1 therein) we have that $c_n''(\bar{\theta}_n) \rightsquigarrow \mathbf{J}_{\theta_0}$. Then the result follows if we show that for any $\lambda = (\lambda_1, \lambda_2)' \in \mathbb{R}^2$

$$\frac{1}{n^{1/\alpha} r_n^{1/\alpha}} \sum_{t=1}^n (z_t^2 - 1)\frac{h_t'(\theta_0)}{\sigma_t^2} \rightsquigarrow z_{\theta_0}.$$

[3] We use a to denote the ARCH parameter here as we reserve α for the stability parameter of the stable distribution.

Analogously to Jensen and Rahbek [24], the main idea is to "asymptotically" replace the non-stationary 2×1 vector $\frac{h_t'(\theta_0)}{\sigma_t^2}$ by stationary versions. $\frac{h_t'(\theta_0)}{\sigma_t^2}$ is $(h_{1t}, h_{1t}^*)'$ in Jensen and Rahbek's notation and $(u_{1t}, u_{1t}^*)'$ are their ergodic approximations where $u_{1t} = \frac{1}{\beta_0} \sum_{j=1}^{\infty} \beta_0^j \prod_{k=1}^{j} \frac{1}{a_0 z_{t-k}^2 + \beta_0}$ and u_{1t}^* is shown to be equal to $\frac{1}{a_0}$. Then the result follows by Lemma 4 which implies that

$$\frac{1}{n^{1/\alpha} r_n^{1/\alpha}} \sum_{i=1}^{n} (z^2 - 1)(h_{1t} - u_{1t}, h_{1t}^* - u_{1t}^*)' \to 0,$$

in L^1 norm.

The fact that the same result holds for arbitrary (ω, γ) when Assumption 15 holds with strict inequality follows by using the exact same arguments as in the proof of Theorem 2 of Jensen and Rahbek [24]. □

4 Subsampling Wald Tests

4.1 Joint Convergence-Algorithm-Limit Theory

In this section we are interested with the first order limit theory of the Wald-type test that employ the QMLE examined above in the context of Theorem 3. We examine the asymptotic behavior of the usual, self normalized test statistic, commonly employed under the additional assumption that $\mathbb{E}\left(z_0^4\right) < +\infty$, when this does not hold as well as when the parameter lies on the boundary, and construct a procedure for the determination of the asymptotic rejection region based on subsampling. We derive asymptotic exactness and consistency, a result that does not hold in our general framework when the usual χ^2 critical values are used, either due to asymptotic non normality and/or due to the form of the asymptotic distribution as a non-trivial projection when the parameter lies on the boundary.

The following proposition is the basis on which the derivations that follow are founded and its proof depends on the proof of Theorem 3.1 of Hall and Yao [18]. It is worth mentioning that in the latter paper the authors use the same result to construct confidence regions for the QMLE elementwise based on the bootstrap methods instead.

Proposition 4. *Suppose that the Assumptions employed in Theorem 3 hold, $\theta_0 \in Int\Theta$, $\varepsilon_n = o_p\left(\frac{n^{2/\alpha} r_n^{2/\alpha}}{n^2}\right)$, and that $\alpha \in (1, 2]$. Then*

$$\left(\frac{n}{n^{1/\alpha} r_n^{1/\alpha}} \left(\theta_n - \theta_0\right), \frac{1}{n^{2/\alpha} r_n^{2/\alpha}} \sum_{t=1}^{n} \hat{z}_t^4 \right) \rightsquigarrow \left(\mathbf{J}_{\theta_0}^{-1} z_{\theta_0}, \zeta \right)$$

where $\zeta = 1$ if $\alpha = 2$ and ζ follows an $\frac{\alpha}{2}$-stable distribution with support $[0, +\infty)$ if $\alpha \neq 2$. The distribution of $\left(\mathbf{J}_{\theta_0}^{-1} z_{\theta_0}, \zeta\right)$ is absolutely continuous. Furthermore, if $F : \mathbb{R}^q \to \mathbb{R}^m$, continuous *and locally around θ_0 continuously differentiable with rank $[J_F(\theta_0)] = m$, then*

$$\mathcal{W}_n(F(\theta_0)) \rightsquigarrow \mathcal{W}(F(\theta_0)),$$

where

$$\mathcal{W}_n(F(\theta_0)) \doteqdot$$

$$\frac{n(F(\theta_n) - F(\theta_0))^T \left[J_F^T(\theta_n)\left(\sum_{t=1}^n \frac{\hat{h}'_t(\theta_n)\left[\hat{h}'_t(\theta_n)\right]^T}{\hat{h}_t(\theta_n)}\right)^{-1} J_F(\theta_n)\right]^{-1} (F(\theta_n) - F(\theta_0))^T}{\sum_{t=1}^n \hat{z}_t^4},$$

and

$$\mathcal{W}(F(\theta_0)) \doteqdot \frac{z_{\theta_0}^T \mathbf{J}_{\theta_0}^{-1} J_F^T(\theta_0)\left[J_F^T(\theta_0)\mathbf{J}_{\theta_0}^{-1} J_F(\theta_0)\right]^{-1} J_F(\theta_0)\mathbf{J}_{\theta_0}^{-1} z_{\theta_0}}{\zeta}$$

with absolutely continuous distribution and $\hat{z}_t \doteqdot \frac{y_t}{\sqrt{\hat{h}_t(\theta_n)}}$, $t = 1, \ldots, n$.

Proof. Firstly, the uniform law of large numbers and the fact that allows us to replace $\mathbb{E}\frac{h'_1(\theta_0)\left[h'_1(\theta_0)\right]^T}{\sigma_1^2}$ with $\frac{1}{n}\sum_{t=1}^n \frac{\hat{h}'_t(\theta_n)\left[\hat{h}'_t(\theta_n)\right]^T}{\hat{h}_t(\theta_n)}$ and $\sum_{t=1}^n z_t^4$ with $\sum_{t=1}^n \hat{z}_t^4$ (see also Remark 5.6.2 of Straumann [38] and Remark 12 of the current paper). Then given the assumption on F, the result follows exactly as in Hall and Yao [18] (see page as well as the proofs of Theorems 2.1.c,e and 3.1) along with the Continuous Mapping Theorem and the delta method. □

Given F as in the previous proposition consider for some $F_{\theta^*} \in \text{Int}F(\Theta)$ the hypothesis structure

$$\mathbf{H}_0 : F(\theta_0) = F_{\theta^*},$$
$$\mathbf{H}_{\mathbf{alt}} : F(\theta_0) \neq F_{\theta^*}.$$

Notice that in our Assumption framework the asymptotic exactness of the usual Wald test for this structure, based on the asymptotic chi-squared distribution becomes generally invalidated. Proposition 4 obviously provides with the asymptotic distribution of the self-normalized Wald test under $\mathbf{H}_0$. Notice that if $\alpha = 2$ the limit distribution is χ_q^2 even in the cases where the *second moment of* z_0 *does not exist*. Hence in this case the classical test remains asymptotically exact and consistent.

This ceases to be true when $\alpha \neq 2$. Hence under our assumption framework in order for a feasible testing procedure to be established, an approximation of the relevant quantiles of the aforementioned distribution is needed. The following algorithm provides describes the well known modification based on subsampling.

Algorithm 1. *The testing procedure consists of the following steps:*

1. *Evaluate* $\mathcal{W}_n(F_{\theta^*})$ at the original sample value.

2. *For* $0 < b_n \leq n$ *generate subsamples from the original observations* $(y_i)_{i=t,\ldots,t+b_n-1}$ *for all* $t = 1, 2, \ldots, n - b_n + 1$.

3. *Evaluate the test statistic on each subsample thereby obtaining $\mathcal{W}_{n,b_n,t}(F_{\theta^*})$ for the subsample indexed by $t = 1, 2, \dots, n - b_n + 1$.*

4. *Approximate the cdf of the asymptotic distribution under the null of $\mathcal{W}_{n,b_n,t}(F_{\theta^*})$* by $G_{n,b_n}(y) = \frac{1}{n-b_n+1} \sum_{t=1}^{n-b_n+1} 1\left(\mathcal{W}_{n,b_n,t}(F_{\theta^*}) \leq y\right)$ and for $a \in (0,1)$ calculate
$$q_{n,b_n}(1-a) = \inf_y \left\{G_{n,b_n}(y) \geq 1-a\right\}.$$

5. *Reject $\mathbf{H_0}$ at a iff $\mathcal{W}_n(F_{\theta^*}) > q_{n,b_n}(1-a)$.*

In order to derive the asymptotic properties below we finally employ the following standard assumption that restricts the asymptotic behaviour of $(b_n)_{n\in\mathbb{N}}$.

Assumption 16. $(b_n)_{n\in\mathbb{N}}$, *possibly depending on* $(y_t)_{t=1,\dots,n}$, satisfies
$$\mathbb{P}(l_n \leq b_n \leq u_n) \to 1$$
where (l_n) and (u_n) are deterministic sequences of natural numbers such that $1 \leq l_n \leq u_n$ for all n, $l_n \to \infty$ and $\frac{u_n}{n} \to 0$ as $n \to \infty$.

The main result is the following.

Proposition 5. *Suppose that the employed in Theorem 3 hold, $\varepsilon_n = o_p\left(\frac{n^{2/\alpha} r_n^{2/\alpha}}{n^2}\right)$, that $\alpha \in (1,2]$ along with Assumption 16. Furthermore suppose that*
$$a < 1 - \mathbb{P}\left(\frac{\mathbf{J}_{\theta_0}^{-1} z_{\theta_0}}{\zeta} \in \mathit{Proj}_{\mathcal{H}}^{-1}\left(\mathit{Ker}[A] \cap \mathit{Bd}[\mathcal{H}]\right)\right),$$
where $A \doteqdot \left[J_F^{Tr}(\theta_0)\, \mathbf{J}_{\theta_0}^{-1} J_F(\theta_0)\right]^{-\frac{1}{2}} J_F(\theta_0)$ and $\mathit{Proj}_{\mathcal{H}}$ is defined by
$$\mathit{Proj}_{\mathcal{H}}(x) = \inf_{y\in\mathcal{H}} (x-y)^T \mathbf{J}_{\theta_0}^{-1}(x-y).$$
Then, for the testing procedure described in Algorithm 1 we have that

1. *If $\mathbf{H_0}$ is true then*
$$\lim_{n\to\infty} \mathbb{P}\left(\mathcal{W}_n(F_{\theta^*}) > q_{n,b_n}(1-a)\right) = a.$$

2. *If $\mathbf{H_{alt}}$ is true then*
$$\lim_{n\to\infty} \mathbb{P}\left(\mathcal{W}_n(F_{\theta^*}) > q_{n,b_n}(1-a)\right) = 1.$$

Remark 15. Notice first that if Bd$[\mathcal{H}] = \emptyset$, i.e. $\theta_0 \in$ Int$[\Theta]$ then the confidence level condition becomes trivial. Second, if Ker$[A] = \{0\}$ then the condition becomes $a < 1 - \mathbb{P}\left(\frac{\mathbf{J}_{\theta_0}^{-1} z_{\theta_0}}{\zeta} \in \text{Proj}_{\mathcal{H}}^{-1}(\{0\})\right)$ and furthermore when $\mathcal{H}$ is factored as $\mathbb{R}^s \times \mathcal{K}$ for $0 < s \leq q$ and $\mathcal{K}$ suitable then it also becomes trivial due to the fact that $\frac{\mathbf{J}_{\theta_0}^{-1} z_{\theta_0}}{\zeta}$ has a density from Proposition 4.

Proof. The first result follows by a direct application of Theorem 3.5.1.i of Politis et al. [34] due to the strong mixing property of the stationary and ergodic GARCH(1,1) model with asymptotically vanishing mixing coefficients (see for example Theorem 3.4, page 71 of Francq and Zakoian [16]). The applicability of this theorem follows from Proposition 4 and the fact that the cdf of the asymptotic distribution of $\mathcal{W}_n\left(F_{\theta^*}\right)$ under the null has an atom at zero of size at most $\mathbb{P}\left(\frac{\mathbf{J}_{\theta_0}^{-1} z_{\theta_0}}{\zeta} \in \text{Proj}_{\mathcal{H}}^{-1}\left(\text{Ker}\,[A] \cap \text{Bd}\,[\mathcal{H}]\right)\right)$. This is due to the following facts. Notice first that from the previous proposition, the fact that $\text{Proj}_{\mathcal{H}}$ is continuous and Lemma 7.13.2-3 of van der Vaart [41] we have that

$$\left(\frac{n}{n^{1/\alpha} r_n^{1/\alpha}}\left(\theta_n-\theta_0\right), \frac{1}{n^{2/\alpha} r_n^{2/\alpha}} \sum_{t=1}^n \hat{z}_t^4\right) \rightsquigarrow \left(\tilde{h}_{\theta_0}, \zeta\right).$$

Second, $A\mathcal{H}$ is convex. Then, due to the fact that $\frac{\mathbf{J}_{\theta_0}^{-1} z_{\theta_0}}{\zeta}$ has a density from the previous proposition, the distribution of $\left\|A \text{Proj}_{\mathcal{H}}\left[\frac{\mathbf{J}_{\theta_0}^{-1} z_{\theta_0}}{\zeta}\right]\right\|^2$ has also a density when restricted to $(0,+\infty)$ and it has an atom at zero when $\mathbb{P}\left(\frac{\mathbf{J}_{\theta_0}^{-1} z_{\theta_0}}{\zeta} \in \text{Proj}_{\mathcal{H}}^{-1}\left(\text{Ker}\,[A] \cap \text{Bd}\,[\mathcal{H}]\right)\right)>0$. Hence if $q\,(1-a)$ denotes the relevant quantile of the distribution of $\mathcal{W}\left(F\left(\theta_0\right)\right)$, then the theorem is applicable iff $1-a>\mathbb{P}\left(\frac{\mathbf{J}_{\theta_0}^{-1} z_{\theta_0}}{\zeta} \in \text{Proj}_{\mathcal{H}}^{-1}\left(\text{Ker}\,[A] \cap \text{Bd}\,[\mathcal{H}]\right)\right)$. For the second result notice that if $\mathbf{H}_{\mathbf{alt}}$ is true then, $\mathcal{W}_n\left(\theta^*\right)=k_{1,n}+k_{2,n}^{Tr} \frac{n^{\frac{\alpha-1}{\alpha}}}{r_n^{\frac{1}{\alpha}}}\left(F\left(\theta_0\right)-F_{\theta^*}\right)+\left(\frac{n^{\frac{\alpha-1}{\alpha}}}{r_n^{\frac{1}{\alpha}}}\right)^2\left\|F\left(\theta_0\right)-F_{\theta^*}\right\|^2$ where $k_{1,n},\left\|k_{2,n}\right\|=O_p(1)$ and thereby it diverges to $+\infty$. □

4.2 A Subsampling Test For The Existence Of The Unconditional Variance In GARCH(1,1)

The previous can be accordingly modified so that a subsampling based testing procedure to be obtained for the issue of the existence of the unconditional variance in the context of the stationary and ergodic GARCH $(1,1)$ model. The test is based on the infimum of the Wald statistic presented above, where the relevant optimization is defined by the following hypotheses structure. To our knowledge no such test has been previously estabished in the relevant literature.

$$\mathbf{H}_0 : a_0+\beta_0 \geq 1,$$
$$\mathbf{H}_{\mathbf{alt}} : a_0+\beta_0<1.$$

Using the previous notation it is easy to see that for $F\left(x_1, x_2, x_3\right)=x_2+x_3$, $F_\theta=a+\beta$,

$$\mathcal{W}_n(a+\beta)=\frac{n^2\left[\left(a_n+\beta_n\right)-(a+\beta)\right]^2}{\hat{V}_n \sum_{t=1}^n \hat{z}_t^4}=\frac{1}{m_n}\left[\frac{n^{\frac{\alpha-1}{\alpha}}}{r_n^{1/\alpha}}\left(\left(a_n+\beta_n\right)-(a+\beta)\right)\right]^2,$$

where $m_n \doteqdot \frac{\hat{V}_n}{n^{2/\alpha} r_n^{2/\alpha}} \sum_{t=1}^{n} \hat{z}_t^4$ and $\hat{V}_n \doteqdot \begin{pmatrix} 0 & 1 & 1 \end{pmatrix} \left(\frac{1}{n} \sum_{t=1}^{n} \frac{\hat{h}'_t(\theta_n) \hat{h}'_t(\theta_n)}{\hat{h}_t(\theta_n)} \right)^{-1} \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}$. We readily obtain the following proposition.

Proposition 6. *Under the premises of Proposition 4 and if furthermore $\theta_0 \in$ Int $[\Theta]$ then,*

$$\mathcal{W}_n^* \doteqdot \inf_{a+\beta \geq 1} \mathcal{W}_n (a + \beta) \rightsquigarrow \inf_{a+\beta \geq 1} \mathcal{W}^\star (a + \beta) = \frac{\xi^2}{V\zeta},$$

where

$$\mathcal{W}^\star (a + \beta) = \begin{cases} +\infty, \text{ if } a + \beta \neq a_0 + \beta_0 \\ \frac{\xi^2}{V\zeta}, \text{ if } a + \beta = a_0 + \beta_0 \end{cases},$$

$\xi = \begin{pmatrix} 0 & 1 & 1 \end{pmatrix} \mathbf{J}_{\theta_0}^{-1} z_{\theta_0}$, $V = \begin{pmatrix} 0 & 1 & 1 \end{pmatrix} \mathbf{J}_{\theta_0}^{-1} \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}$ *and ζ as in Proposition 4.*

Proof. Follows directly from Proposition 4 and the CMT. □

Given this the following algorithm provides with the following testing procedure for the finiteness of the unconditional variance in the present premises.

Remark 16. Notice that

$$\mathcal{W}_n^* = \begin{cases} 0, \text{ if } a_n + \beta_n \geq 1 \\ \mathcal{W}_n (1), \text{ if } a_n + \beta_n < 1 \end{cases}.$$

Algorithm 2. *The testing procedure consists of the following steps:*

1. *Evaluate $\mathcal{W}_n^*$* at the original sample value.
2. *For $0 < b_n \leq n$ generate subsamples from the original observations $(y_i)_{i=t,\dots,t+b_n-1}$ for all $t = 1, 2, \dots, n - b_n + 1$.*
3. *Evaluate the test statistic on each subsample thereby obtaining ${\mathcal{W}^*}_{n,b_n,t}$ for the subsample indexed by $t = 1, 2, \dots, n - b_n + 1$.*
4. *Approximate the cdf of the asymptotic distribution under the null of ${\mathcal{W}^*}_{n,b_n,t}$* by $G_{n,b_n}^* (y) = \frac{1}{n-b_n+1} \sum_{t=1}^{n-b_n+1} 1 \left({\mathcal{W}^*}_{n,b_n,t} \leq y \right)$ and for $a \in (0, 1)$ calculate

$$q_{n,b_n}^* (1 - a) = \inf_y \left\{ G_{n,b_n}^* (y) \geq 1 - a \right\}.$$

5. *Reject $\mathbf{H_0}$ at a iff $\mathcal{W}_n^* > q_{n,b_n} (1 - a)$.*

The final proposition establishes the asymptotic exactness and the consistency of the previous procedure.

Proposition 7. *Under the assumptions employed in Proposition 5 and for the testing procedure described in Algorithm 2 we have that*

1. *If* $\mathbf{H_0}$ *is true then*
$$\lim_{n\to\infty} \mathbb{P}\left(\mathcal{W}_n^* > q_{n,b_n}^*(1-a)\right) = a.$$
2. *If* $\mathbf{H_{alt}}$ *is true then*
$$\lim_{n\to\infty} \mathbb{P}\left(\mathcal{W}_n^* > q_{n,b_n}^*(1-a)\right) = 1.$$

Proof. Both results follow by Proposition 6 along with Theorem 3.5.1 of Politis et al. [34]. □

Remark 17. Although loosely related, Loretan and Phillips [29] examine methods of testing the hypothesis of non constancy of the unconditional variance of a time series. They use sample split prediction tests and cusum of squares tests without explicitly modelling the volatility process and find that the latter have nonstandard limiting distributions when fourth unconditional moments are infinite. They work under the assumption that the distribution of the squared innovation process lies in the *normal* domain of attraction of an α-stable distribution for $\alpha \in [1,2]$ assuming additionally symmetry in the case where $\alpha = 1$. Then they estimate the stability parameter α (the maximal moment exponent) and compute the critical values of the limit distributions numerically. Instead our test is semi-parametric as we assume a GARCH(1,1) model and in this case observe that non-constancy of the unconditional volatility cannot happen.

Monte Carlo Results We evaluate the performance of Algorithm 2 above by performing a Monte Carlo experiment. We generate data from a GARCH(1,1) process with constant parameter $\omega_0 = 1$, GARCH parameter $\beta_0 = 0.90$ and ARCH parameter $\alpha_0 = 0.10$ ($\alpha_0 = 0.08$) under $\mathbf{H_0}$ ($\mathbf{H_{alt}}$), where the innovations are drawn from the Student's t-distribution with degrees of freedom v=3,4 and 5. The number of simulations is set to S=1000. In each simulation, we generate a sample path of size n+300 choosing the initial value for the volatility to be equal to 1 and then we drop the first 300 observations, in order to eliminate the effect of the initial value. The sample size is chosen as n=500,1000 and 5000. When n=500 we use subsample sizes n_s=200,250,300,350,400, when n=1000 we use n_s=300,400,500,600,700,800,900, when n=5000 we use n_s=1500,2500,4000. Furthermore we use all the possible subsamples that we can construct the number of which equals n-n_s+1. The experiment was implemented in MATLAB 2010a using a cluster of 10 computers each of which is equipped with an Intel Core i7-3770 processor. The parallel toolbox was used to utilize all 8 (virtual) cores. Each computer was set to perform 1/10 of the 1000 simulations. For reproducibility each simulation was assigned a specific substream (equal to the number of the simulation) of the global stream. The results are shown in Table below.

Table 1: Size and Power of the Wald Test using Subsampling

		v=3		v=4		v=5	
n	n_s	Size	Power	Size	Power	Size	Power
500	250	.226	.306	.205	.314	.201	.368
	300	.239	.304	.203	.329	.190	.356
	350	.237	.325	.183	.321	.173	.323
	400	.251	.329	.185	.297	.160	.289
1000	400	.212	.359	.185	.473	.152	.500
	500	.222	.370	.175	.455	.156	.448
	600	.226	.384	.183	.438	.150	.422
	700	.238	.380	.170	.425	.153	.389
	800	.230	.374	.163	.353	.133	.328
	900	.218	.338	.150	.277	.122	.250
5000	1500	.160	.443	.70	.626	.122	.250
	2500	.188	.440	.110	.553	.82	.651
	4000	.220	.426	.142	.437	.121	.452

5 Conclusions

In this paper we derived a limit theorem to mixed α-stable limits for "martingale transforms" for any value of the stability parameter $\alpha \in (0, 2]$ extending and improving the existing results. Then we provided a framework which relies on strict stationarity of the volatility process for the limit theory of the QMLE. We allow for the distribution of the squared innovation process to lie in the domain of attraction of stable laws and the true parameter to lie on the boundary of the parameter space. We show that when $\mathbb{E}z_0^2 < \infty$, which permits $\alpha \in [1, 2]$, then the rate of convergence of the QMLE is $\frac{n}{n^{1/\alpha} r_n^{1/\alpha}}$ where $r(n) := r_n$ is a slowly varying function, and the limit distribution is an α-stable distribution if the true parameter is an interior point; otherwise it is a projection on asymptotic parameter space. When $\alpha < 1$ and thus $\mathbb{E}z_0^2 = \infty$ we show that the QMLE is inconsistent. Furthermore we show that the stationarity assumption can be relaxed by providing an example in which we derive the limit theory of the non-stationary GARCH(1,1) model. Finally we derive the limit theory of the classical Wald test analogously to Hall and Yao [18] and construct a testing procedure for the existence of the unconditional variance in the GARCH(1,1) model. We derive its limit behavior and evaluate its finite sample performance numerically.

As a possible extension regarding the MLT theorem, we could investigate if our results are affected should we allow for $\mathbb{E}|V_1|^\alpha = \infty$ in the stationary case or the (V_t) sequence to weakly converge to a non-degenerate random variable in the non-stationary case, and thus possibly generalizing the results of Wang [43] which could be useful in other applications such as non linear cointegration. Furthermore, it could be of potential interest to examine possible ways to weaken or possibly avoid Assumption 5.

An open question with respect to limit theorem for the QMLE is whether the asymptotic distribution of the latter when $\alpha \leq 1$ can be determined even without it being consistent. A simple example where this is possible concerns the ARCH(1) model and can be found in Example 2 of Arvanitis and Louka [6]. Furthermore the determination of the limit distribution of the QMLE when the true parameter lies on the boundary and $\alpha = 1$ needs further investigation as it introduced several complexities.

Appendix

Auxiliary Results: Strong Consistency, Rate of Convergence and Asymptotic Distribution

The following are auxiliary results that concern the issues of strong consistency, rate of convergence and asymptotic distribution for approximate minimizers of appropriate criteria. To this end, suppose that Θ is a compact subset of $\mathbb{R}^d$ equipped with the relevant Euclidean topology. Let $c_n : \Omega \times \Theta \to \mathbb{R}$ be jointly measurable, θ_n be defined as a $\mathbb{P}$ a.s. approximate minimizer of c_n with optimization error ε_n a $\mathbb{P}$ a.s. non negative random variable. The following result provides with sufficient conditions that characterize the rate of convergence and the asymptotic distribution of θ_n given *consistency*. Let $\theta_0 \in \Theta$. For reasons of notational economy we suppress the dependence on ω. The following lemma provides with sufficient conditions for strong consistency when c_n has the form of an ergodic mean, allowing for cases where the analogous expectation does not exist.

Lemma 6. *Suppose that* $c_n(\theta) = \frac{1}{n}\sum_{i=1}^{n} m_i(\theta)$, $(m_i(\theta))_{i\in\mathbb{Z}}$ *is ergodic for any* θ, c_n *is jointly continuous* $\mathbb{P}$ *a.s., there exists a finite open cover of* Θ*, such that* $\mathbb{E}\left|\inf_{\theta\in A} m_0(\theta)\right| < +\infty$, *for any* A *in the cover,* $\mathbb{E}m_0(\theta)$ *assumes values in* $\overline{\mathbb{R}}$ *for any* θ *in a countable dense subset of* Θ*. Suppose furthermore that* $\theta_0 = \arg\min_{\Theta} \mathbb{E}m_0(\theta)$ *and that* $\varepsilon_n \to 0$, $\mathbb{P}$ *a.s. Then* $\theta_n \to \theta_0$ $\mathbb{P}$ *a.s.*

Proof. The first part of the Assumption framework of the Lemma implies condition $\mathcal{C}_0$ and thereby Theorem 2.3 of Choirat , Hess, and Seri, [10], which implies the joint $\mathbb{P}$ a.s. epi-convergence of c_n to $\mathbb{E}m_0$. Let $\mathbf{epi}$ denote the epigraph of a given function (see e.g. Paragraph 3.1-Ch.5 of Molchanov [31]). Then the assumed properties of c_n Proposition 3.6 and Definition 3.5 (Ch. 5) of Molchanov [31] imply that $\mathbf{epi}_n \doteqdot \mathbf{epi}(c_n)$ is a jointly measurable closed valued correspondence. Conditions 1. and 2. are essentially the sequential characterization of P a.s. *epi-convergence* of c_n to $\mathbb{E}m_0$ (see Definitions 2.1 and 2.2 of Lachout et al. [28]). It follows that $\mathbb{E}m_0$ is an lsc function (see Proposition 7.4.a of Rockafellar and Wets [36]). Hence $\mathbf{epi}(\mathbb{E}m_0)$ is a closed valued correspondence. Due to Molchanov [31], paragraph 1.1, and Klein and Thompson [26], Definition 4.5.1 this $\mathbb{P}$ a.s. epi-convergence is equivalent to the following (**i**)-(**ii**) conditions. (**i**) for large enough n, and for all ω in a measurable subset of Ω of unit $\mathbb{P}-$probability, $\mathbf{epi}_n \cap \Theta \times (\mathbb{E}m_0(\theta_0), +\infty) \neq \emptyset$ since $\Theta \times (\mathbb{E}m_0(\theta_0), +\infty)$ is open in the relevant product topology and $\mathbf{epi}(\mathbb{E}m_0) \cap \Theta \times (\mathbb{E}m_0(\theta_0), +\infty) \neq \emptyset$. Hence $\inf_\Theta c_n(\theta) \geq \mathbb{E}m_0(\theta_0)$ for all ω described previously which implies that $\liminf_n \inf_\Theta c_n \geq \mathbb{E}m_0(\theta_0)$ $\mathbb{P}$ a.s. Furthermore (**ii**) for any $\varepsilon > 0$, we have that for large n, and for all ω in a (possibly different than the previous) measurable subset of Ω of unit $\mathbb{P}-$probability, $\mathbf{epi}_n \cap \Theta \times [\mathbb{E}m_0(\theta_0) - \varepsilon, \mathbb{E}m_0(\theta_0) - 2\varepsilon] = \emptyset$ P a.s. since $\Theta \times [\mathbb{E}m_0(\theta_0) - \varepsilon, \mathbb{E}m_0(\theta_0) - 2\varepsilon]$ is compact in the relevant product topology and $\mathbf{epi}(\mathbb{E}m_0) \cap \Theta \times [\mathbb{E}m_0(\theta_0) - \varepsilon, \mathbb{E}m_0(\theta_0) - 2\varepsilon] = \emptyset$. This implies that $\limsup_n \inf_\Theta c_n(\theta) \leq \mathbb{E}m_0(\theta_0)$ $\mathbb{P}$ a.s. Now let x_n be a measurable selection from the random compact set

$$\left\{\theta \in \Theta : c_n(\theta) \leq \inf_{\Theta} c_n + \varepsilon_n\right\}$$

such that for some subsequence $\left(x_{n_k}\right)$, $x_{n_k} \rightarrow x$ $\mathbb{P}$ a.s. Its existence is guaranteed by the fundamental selection theorem (Theorem 2.13-Ch. 1 of Molchanov [31]). Then

$$\begin{aligned}
\mathbb{E}m_0(x) &\leq \liminf_{n_k} c_{n_k}\left(x_{n_k}\right) \ \mathbb{P} \text{ a.s.} \\
&\leq \limsup_{n_k} c_{n_k}\left(x_{n_k}\right) \ \mathbb{P} \text{ a.s.} \\
&= \limsup_{n_k}\left(\inf_{\Theta} c^*_{n_k} + \varepsilon_{n_k}\right) \ \mathbb{P} \text{ a.s.} \\
&\leq \mathbb{E}m_0(\theta_0) \ \mathbb{P} \text{ a.s.}
\end{aligned}$$

establishing that any $\mathbb{P}$ a.s. cluster point of such a measurable selection coincides with θ_0. The result now follows from the fact that Θ is compact. □

For $w_n \rightarrow +\infty$, we denote with $\mathcal{H}_n$ the $w_n(\Theta - \theta_0) = \{w_n(x - \theta_0), x \in \Theta\}$ and notice that $\mathcal{H}_n$ is compact and contains 0. Furthermore we denote with $\mathcal{H} = \limsup_{n\rightarrow\infty} \mathcal{H}_n$ in the sense of the obvious generalization of definition 3.

Consider the following assumption that provides more structure for the asymptotic properties of c_n.

Assumption 17. *Assume that the following hold:*

1. For any sequence (ϑ_n) with values in Θ such that $\vartheta_n \overset{p}{\rightarrow} \theta_0$, $c_n(\vartheta_n) - c_n(\theta_0) = (\vartheta_n - \theta_0)' q_n + (\vartheta_n - \theta_0)' g_n (\vartheta_n - \theta_0)$, with $\mathbb{P}$ probability that converges to 1. g_n is a random $q \times q$ matrix that can be defined in any point of the aforementioned line $\mathbb{P}$ a.s. q_n is a random $q \times 1$ matrix.

2. For some positive real sequence $w_n \rightarrow +\infty$, $w_n q_n \rightsquigarrow z_{\theta_0}$ which is a random vector whose distribution can depend on θ_0 and $g_n \overset{p}{\rightarrow} \mathbf{J}_{\theta_0}$ a non singular matrix independent of ω that may depend on θ_0.

3. $\mathcal{H}$ is convex.

The next theorem is the final result of this section.

Theorem 5. *Assume that $\theta_n \underset{p}{\rightarrow} \theta_0$. If conditions 17.1,2 hold and $\varepsilon_n = O_p\left(w_n^{-2}\right)$ then*

$$w_n(\theta_n - \theta_0) = O_p(1). \tag{23}$$

If moreover condition 17.3 holds and $\varepsilon_n = o_p\left(w_n^{-2}\right)$ then

$$w_n(\theta_n - \theta_0) \rightsquigarrow \tilde{h}_{\theta_0} \tag{24}$$

with $\tilde{h}_{\theta_0}$ defined uniquely by $q\left(\tilde{h}_{\theta_0}\right) = \inf_{h\in\mathcal{H}} q(h)$ and $q(h) := \left(h - \mathbf{J}^{-1}_{\theta_0} z_{\theta_0}\right)' \mathbf{J}_{\theta_0}\left(h - \mathbf{J}^{-1}_{\theta_0} z_{\theta_0}\right)$.

Proof. Notice that due to the definition of θ_n we have

$$c_n(\theta_n) - c_n(\theta_0) \leq O_p\left(w_n^{-2}\right).$$

From $\theta_n \underset{p}{\rightarrow} \theta_0$ and employing Assumption 17.1,2

$$\nu_n' w_n q_n + \nu_n' g_n(b_n^*) \nu_n \leq O_p(1)$$

where $\nu_n = w_n(\theta_n - \theta_0)$ and b_n^* as in 17.1. Hence due to consistency

$$\nu_n' w_n q_n + \nu_n' \left(\mathbf{J}_{\theta_0} + o_p(1)\right) \nu_n \leq O_p(1).$$

Assumption 17.2 then implies that there exists some positive $c > 0$ such that

$$\|\nu_n\| O_p(1) + c\|\nu_n\|^2 + \|\nu_n\|^2 o_p(1) \leq O_p(1)$$

which implies that

$$\|\nu_n\|^2 \left(1 + o_p(1)\right) + 2\|\nu_n\| O_p(1) \left(1 + o_p(1)\right) + O_p(1) \leq O_p(1)$$

Hence

$$\|\nu_n\| \left(1 + o_p(1)\right) \leq O_p(1)$$

establishing (23). Now given the definition of $\mathcal{H}$ consider the following. From consistency and Assumption 17 we can define $\varpi_n : \mathbb{R}^q \to \mathbb{R}$ as

$$\begin{aligned} \varpi_n(h) &\equiv w_n^2 \left(c_n\left(\theta_0 + \frac{h}{w_n}\right) - c_n(\theta_0)\right) \\ &= h' w_n q_n + h' g_n(b_n^*) h \end{aligned}$$

From the first part of the present proof we have that for U an arbitrary compact subset of $\mathbb{R}^q$

$$\varpi_n(h) \rightsquigarrow h' z_{\theta_0} + \frac{1}{2} h' \mathbf{J}_{\theta_0} h \text{ in } C(U, \mathbb{R}).$$

Hence for any A non-empty subset of $\mathbb{R}^q$,

$$\inf_{h \in A} \varpi_n(h) \rightsquigarrow \inf_{h \in A} \left(h' z_{\theta_0} + \frac{1}{2} h' \mathbf{J}_{\theta_0} h\right). \tag{25}$$

Due to (23) $h_n \doteqdot w_n(\beta_n - b(\theta_n)) \in \mathcal{H}_n \cap B(0, w_n \varepsilon) \doteqdot M_n$ with $\mathbb{P}$-probability tending to 1 for some $\varepsilon > 0$. If F is a closed non empty subset of $\mathbb{R}^q$, and $h_n \in F$, then for large enough n, either $M_n \subset F$, or $M_n \nsubseteq F$ but $M_n \cap F \neq \emptyset$. In either case due to the definitions of θ_n, β_n ϖ_n and the fact that $\varepsilon_n = o_p\left(w_n^{-2}\right)$

$$\inf_{h \in M_n \cap F} \varpi_n(h) \leq \inf_{h \in M_n} \varpi_n(h) + o_p(1)$$

and therefore due to Slutsky's lemma

$$\mathbb{P}\left(h_n \in F\right) \leq \mathbb{P}\left(\inf_{h \in M_n \cap F} \varpi_n\left(h\right) \leq \inf_{h \in M_n} \varpi_n\left(h\right) + o_p\left(1\right)\right)$$
$$\leq \mathbb{P}\left(\inf_{h \in M_n \cap F} \varpi_n\left(h\right) \leq \inf_{h \in M_n} \varpi_n\left(h\right)\right) + o\left(1\right)$$

Now notice that $M_n = M_n \cap \mathbb{R}^q$ and $\mathbb{R}^q$ is open, $\limsup_{n\to\infty} M_n = \mathcal{H}$, since PK- $\limsup \mathcal{H}_n = \mathcal{H}$ and $r_n \to \infty$. Furthermore equation (25) and the continuous mapping theorem imply that Lemma 7.13.2-3 of van der Vaart [41] is applicable, so that the last probability is less than or equal to

$$\mathbb{P}\left(\inf_{h \in \mathcal{H} \cap F} \varpi_n\left(h\right) \leq \inf_{h \in \mathcal{H}} \varpi_n\left(h\right) + o_p\left(1\right)\right) \leq \mathbb{P}\left(\inf_{h \in \mathcal{H} \cap F} \varpi_n\left(h\right) \leq \inf_{h \in \mathcal{H}} \varpi_n\left(h\right)\right) + o\left(1\right)$$

due to Slutsky's Lemma. Now from equation (25), the continuous mapping theorem and Portmanteau Lemma we have that the $\limsup$ of the probability in the right hand side of the last display is less than or equal to

$$\mathbb{P}\left(\inf_{h \in \mathcal{H} \cap F} h' z_{\theta_0} + \frac{1}{2} h' \mathbf{J}_{\theta_0} h \leq \inf_{h \in \mathcal{H}} h' z_{\theta_0} + \frac{1}{2} h' \mathbf{J}_{\theta_0} h\right)$$

which equals

$$\mathbb{P}\left(\inf_{h \in \mathcal{H} \cap F} h' z_{\theta_0} + \frac{1}{2} h' \mathbf{J}_{\theta_0} h \pm \frac{1}{2} z'_{\theta_0} \mathbf{J}^{-1}_{\theta_0} z_{\theta_0} \leq \inf_{h \in \mathcal{H}} h' z_{\theta_0} + \frac{1}{2} h' \mathbf{J}_{\theta_0} h \pm \frac{1}{2} z'_{\theta_0} \mathbf{J}^{-1}_{\theta_0} z_{\theta_0}\right)$$
$$= \mathbb{P}\left(\inf_{h \in \mathcal{H} \cap F} \left(h - \mathbf{J}^{-1}_{\theta_0} z_{\theta_0}\right)' \mathbf{J}_{\theta_0} \left(h - \mathbf{J}^{-1}_{\theta_0} z_{\theta_0}\right) \leq \inf_{h \in \mathcal{H}} \left(h - \mathbf{J}^{-1}_{\theta_0} z_{\theta_0}\right)' \mathbf{J}_{\theta_0} \left(h - \mathbf{J}^{-1}_{\theta_0} z_{\theta_0}\right)\right)$$

Since H^* is closed and convex and $\mathbf{J}_{\theta_0}$ is positive definite $\tilde{h}_{\theta_0}$ is unique, and thereby when

$$\inf_{h \in \mathcal{H} \cap F} \left(h - \mathbf{J}^{-1}_{\theta_0} z_{\theta_0}\right)' \mathbf{J}_{\theta_0} \left(h - \mathbf{J}^{-1}_{\theta_0} z_{\theta_0}\right) \leq \inf_{h \in \mathcal{H}} \left(h - \mathbf{J}^{-1}_{\theta_0} z_{\theta_0}\right)' \mathbf{J}_{\theta_0} \left(h - \mathbf{J}^{-1}_{\theta_0} z_{\theta_0}\right)$$

holds then

$$\tilde{h}_{\theta_0} \in \mathcal{H} \cap F$$

and therefore the last probability is less than or equal to

$$\mathbb{P}\left(\tilde{h}_{\theta_0} \in \mathcal{H} \cap F\right) \leq \mathbb{P}\left(\tilde{h}_{\theta_0} \in F\right)$$

hence we have proven that

$$\limsup_{n\to\infty} \mathbb{P}\left(h_n \in F\right) \leq \mathbb{P}\left(\tilde{h}_{\theta_0} \in F\right)$$

and (24) follows from the Portmanteau theorem due to the fact that F is chosen arbitrarily.

$\square$

References

[1] Aaronson, J., & Denker, M. (1998). Characteristic functions of random variables attracted to 1-stable laws. Annals of probability, 399-415.

[2] Abadir K. and J. Magnus (2004), "The Central Limit Theorem for Student's Distribution", *Econometric Theory*, Vol. 20, No. 6, pp. 1261-1263.

[3] Aliprantis C.D. and K.C. Border (1999), "*Infinite Dimensional Analysis. A Hitchhiker's Guide.*", Springer.

[4] Andrews, D. W. K., "Estimation when a parameter is on boundary", *Econometrica, 67:1341-1383, 1999.*

[5] Arvanitis, S., & Louka, A. (2015). Limit Theory for the QMLE of the GQARCH (1, 1) model. Communications in Statistics-Theory and Methods, (just-accepted)

[6] Arvanitis, S., Louka, A. (2015), A Note on the QMLE Limit Theory in the Non-stationary ARCH (1) Model. *Journal of Time Series Econometrics*.

[7] Berkes, I. , L. Horvath, and P. Kokoszka, "GARCH processes: structure and estimation", *Bernoulli,* 9:201-228, 2003.

[8] Bingham N.H., Goldie C.M. and J.L. Teugels (1992), "*Regular Variation*", Encyclopedia of Mathematics and Its Applications, Vol. 27, C.U.P.

[9] Calzolari G., Fiorentini G. and E. Sentana (2003), "Maximum Likelihood Estimation and Inference in Multivariate Conditionally Heteroscedastic Dynamic Regression Models with Student t Innovations", *Journal of Business and Economic Statistics*, 21, 4, 532-546.

[10] Choirat C., Hess C., and R. Seri (2003), "A Functional Version of the Birkhoff Ergodic Theorem for a Normal Integrand: A Variational Approach.", *The Annals of Probability*, Vol. 31, No 1, pp. 63-92.

[11] Davidson J. (1995), "*Stochastic Limit Theory.*", Advanced Texts in Econometrics, Oxford University Press.

[12] Demos A. and D. Kyriakopoulou (2009), "Asymptotic normality of the QMLE in the EGARCH (1, 1) model". working paper, 2013.

[13] Diebold, F. (1988). "*Empirical modeling of exchange rate dynamics*". Springer-Verlag.

[14] Engle, R. F. and Bollerslev, T. (1986). "Modelling the persistence of conditional variances". *Econometric Reviews*, 5 1-50.

[15] Francq, Christian, and Jean-Michel Zakoian, (2007). "Quasi-maximum likelihood estimation in GARCH processes when some coefficients are equal to zero." *Stochastic Processes and their Applications,* 117.9: 1265-1284.

[16] Francq, C., & Zakoian, J. M. (2011). GARCH models: structure, statistical inference and financial applications. John Wiley & Sons.

[17] Gupta, A. K., Nguyen, T. T., & Zeng, W. B. (1994), "Conditions for stability of laws with all projections stable." *Sankhyā: The Indian Journal of Statistics*, Series A, 438-443.

[18] Hall P. and Yao Q., "Inference in ARCH and GARCH Models with Heavy Tailed Errors", *Econometrica*, 71, 285-317, 2003.

[19] Hansen, B. E. (1994). Autoregressive conditional density estimation. International Economic Review, 705-730.

[20] Hsieh, D. A. (1989). "Modeling heteroscedasticity in daily foreign-exchange rates". *Journal of Business and Economic Statistics*, 7, 307-317.

[21] Jakubowski A. (1986), "Principle of Conditioning In Limit Theorems For Sums of Random Variables", *The Annals of Probability*, Vol. 14, No 2, pp. 902-915.

[22] Jakubowski A. (2012), "Principle of Conditioning Revisited", *Demonstratio Mathematica*, Vol. XLV, No 2.

[23] Jensen, S. T., and Rahbek, A. (2004a), "Asymptotic normality of the QMLE estimator of ARCH in the nonstationary case". *Econometrica*, 72(2), 641-646.

[24] Jensen, S. T., & Rahbek, A. (2004b). Asymptotic inference for nonstationary GARCH. Econometric Theory, 20(06), 1203-1226. ISO 690

[25] Ibragimov I.A. and Yu. V. Linnik (1971), "Independent and Stationary Sequences of Random Variables", WOLTERS-NOORDHOFF PUBLISHING GRONINGEN.

[26] Klein Er., and Thompson A.C. (1984), "*Theory of Correspondences.*", CMS Series of Monographs, Wiley Interscience.

[27] Kwapien S., W.A. Woyczynski (1992), "*Random Series and Stochastic Integrals: Single and Multiple*", Probability and Its Applications, Birkhauser.

[28] Lachout P., E. Liebsher, and S. Vogel, (2005), "Strong Convergence of Estimators as ε_n-Minimisers of Optimization Problems.", *Ann. Inst. Stat. Math.*, Vol. 57-2, pp. 291-313.

[29] Loretan, Mico, and Peter CB Phillips. "Testing the covariance stationarity of heavy-tailed time series: An overview of the theory with applications to several financial datasets." Journal of empirical finance 1.2 (1994): 211-248.

[30] Mikosch T. and D. Straumann, "Stable Limits of Martingale Transforms With Application to the Estimation of GARCH Parameters", *The Annals of Statistics*, Vol.34, No 1, 439-522, 2006.

[31] Molchanov, Ilya (2004), "*Theory of Random Sets.*", Probability and Its Applications, Springer.

[32] Nelson, D. B. (1990). "Stationarity and persistence in the GARCH(1,1) model". *Econometric theory*, 6(03), 318-334.

[33] Nelson, D. B. (1991). "Conditional heteroskedasticity in asset returns: A new approach". *Econometrica*, 59, 347-370.

[34] Politis, D. N., J. P. Romano and M. Wolf (1999). *Subsampling*. Springer New York.

[35] Pollard, D. (2002). A user's guide to measure theoretic probability (Vol. 8). Cambridge University Press.

[36] Rockafellar T.R. and Wetts J-B (1997), "*Variational Analysis.*", Springer-Verlag.

[37] Rudin, W. (2011). "*Fourier analysis on groups*." John Wiley & Sons.

[38] Straumann, D. (2004), "*Estimation in Conditionally Heteroscedastic Time Series Models.*", Springer.

[39] Sentana, E., "Quadratic ARCH Models", *The Review of Economic Studies*, 62:639-661, 1995.

[40] Surgailis, D., "A Quadratic ARCH(∞) Model with Long Memory and Levy Stable Behavior of Squares", *Adv. Appl. Prob.*, 40:1198-1222, 2008.

[41] van der Vaart, A.W., "*Asymptotic Statistics*" CUP, 1997.

[42] van der Vaart, A.W. and J.A. Wellner, "*Weak Convergence and Empirical processes*" Springer, 2000.

[43] Wang, Q. (2011). Martingale limit theorems revisited and non-linear cointegrating regression.

[44] Wintenberger, O., and S. Cai, "Parametric Inference and Forecasting for Continuously Invertible Volatility Models", http://arxiv.org/pdf/1106.4983.pdf.

[45] Wintenberger, O. (2013). Continuous invertibility and stable QML estimation of the EGARCH (1, 1) model. Scandinavian Journal of Statistics, 40(4), 846-867.