A NOTE ON THE QMLE LIMIT THEORY IN THE NON-STATIONARY ARCH(1) MODEL

STELIOS ARVANITIS[abd] AND ALEXANDROS LOUKA[acd]

Abstract: In this note we extend the standard results for the limit theory of the popular QMLE in the context of the non-stationary ARCH(1) model by allowing the innovation process not to possess fourth moments. Depending on the value of the index of stability, we either derive α-stable weak limits with non-standard rates or inconsistency and non-tightness. We obtain the limit theory by the derivation of a Limit Theorem for multiplicative "martingale" transforms with limits mixtures of α-stable distributions for any $\alpha \in (0, 2]$.

KEYWORDS: α-stable distribution, slow variation, domain of attraction, MLT with mixed limit, non stationary ARCH(1), QMLE, inconsistency, non-tightness.

1. INTRODUCTION

In this note we study the limit theory of the QMLE in the non stationary ARCH(1) model. We *extend* the results of Jensen and Rahbek [13] by allowing the innovation process not to possess fourth moments. Since this process is comprised by iid random variables, we do so by allowing the stationary distribution of their squares to belong to the domain of attraction of an α-stable law. When $\alpha > 1$, or $\alpha = 1$ and the second moment of the innovation process exists, we obtain as limiting distribution an α-stable law with non-standard rates. When $\alpha = 2$ and the fourth moment of the innovation process exists, we recover the result of Jensen and Rahbek [13], i.e. asymptotic normality with the usual $\sqrt{n}$ rate, albeit via the use of a different methodology. When $\alpha = 2$ but the fourth moment does not exist, we obtain again asymptotic normality with non-standard and slower rate. Finally, when $\alpha < 1$ we obtain inconsistency and furthermore asymptotic non-tightness for the estimator, a result that is completely novel in the relevant literature.

We show that we can obtain the limit theory by deriving an auxiliary Limit Theorem for multiplicative "martingale" transforms (denoted as MLT for simplicity) with limits mixtures of α-stable distributions for any $\alpha \in (0, 2]$. Similar MLT's, yet with non-mixed limits, have been, directly or indirectly, derived by Hall and Yao [9] and Mikosch and Straumann [17] for the cases where $\alpha \geq 1$. They were concerning stationary and ergodic transforms, and applied to the asymptotic behavior of the quasi score in the context of the ergodic GARCH model. Given the application that we have in mind, we do not require stationarity for the transform, but we require the existence of an almost sure limit for the conditionally "scaling sequence". When $\alpha < 1$ we apply the MLT to the likelihood function itself in order to obtain the result. Furthermore, we derive the MLT using the so-called "Principle of Conditioning" of Jakubowski [12] and avoid many of the complexities appearing in the relevant proofs of the aforementioned paper.

---

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

[b]Email address: stelios@aueb.gr.

[c]Email address: loukaalex@aueb.gr.

[d]Both authors acknowledge financial support from the Greek General Secretariat for Research & Technology under the grand ARISTEIA II-08-5413.

The structure of the note is as follows. In the second section we present the probabilistic framework and derive the MLT. In the third we apply it to obtain the asymptotic behavior of the QMLE. We finally discuss some parallel and future similar research and possible extensions. We gather all proofs in the Appendix.

2. A MLT WITH MIXED STABLE LIMITS

Our framework is built around 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 with $\rightsquigarrow$. 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 characteristics 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, and the log-characteristic function of the distribution of ξ_0 has the following local representation around zero:*

$$(2.1)\qquad \begin{cases} \gamma it - c|t|^{\alpha} h(1/|t|)\left(1 - i\beta \operatorname{sgn}(t)\tan\left(\frac{1}{2}\pi\alpha\right)\right) & ,\alpha \in (0,1)\cup(1,2] \\ (\gamma + H(1/|t|))\, it - c|t| h(1/|t|)\left(1 - 2Ci\frac{\beta}{\pi}\operatorname{sgn}(t)\right) & ,\alpha = 1 \end{cases}$$

where h is slowly varying function at infinity, $H(\lambda) = \int_0^{\lambda} \frac{x^2 h(x)}{1+x^2}dx$, $\beta \in [-1,1]$, $c \in \mathbb{R}_{++}$, $\gamma \in \mathbb{R}$ and $-C$ is the Euler-Mascheroni constant.

REMARK 1 The representations appearing in 2.1 are equivalent to that the distribution of ξ_0 lies in the domain of attraction of an α -stable law, due to Theorem 2.6.5 of Ibragimov and Linnik [10] for $\alpha \neq 1$ and Theorem 2 of Aaronson and Denker [1] for $\alpha = 1$, i.e. when appropriately translated and then scaled, the partial sums of $(\xi_i)_{i=0}^{n}$ 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 $\mathbb{E}\xi_0$ when $\alpha > 1$. The aforementioned theorems (that are of Tauberian type) imply that α and the slowly varying function h represent the asymptotic behavior of the tails of the distribution of ξ_0. 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 [10] and Paragraph 1.9 of Bingham et al. [5]). When h converges then the distribution of ξ_0 is said to belong to the domain of *normal* attraction to the relevant α-stable

[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 [17].

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 $\mathbb{E}\left|\xi_0\right|^{\alpha} < +\infty$ 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_0 (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.

NOTATION 1 In what follows $S_\alpha(\beta, c, \gamma)$ denotes an α-stable distribution with parameters β, c, γ.[2] 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 that we allow (for some of) those parameters to be $\mathcal{G}$-measurable non-constant functions defined on Ω.

The second assumption describes part of the dependence structure between the two constituent processes. It enables the characterization of the martingale transform as a process of random variables that are conditionally independent, yet conditionally inhomogeneous, that belong to the domain of attraction of α-stable laws (with α and h being constant with respect to the conditioning σ-algebras), and where the $(V_i)_{i\in\mathbb{N}}$ process determines the location, scaling and skewness parameters of those conditional distributions. The assumption finally specifies the asymptotic behavior of the aforementioned process.

ASSUMPTION 2 *There exists a filtration $(\mathcal{F}_t)_{t\in\mathbb{N}}$ such that, for any t, $(V_t)_{t\in\mathbb{N}}$ is measurable* w.r.t., and ξ_t *is independent of,* $\mathcal{F}_t$. *Furthermore*

(2.2) $$V_t \to v, \quad \mathbb{P} \text{ a.s. as } t \to +\infty$$

where v is a random variable assuming non-zero values $\mathbb{P}$ a.s.

REMARK 2 In cases such as the one considered in the following section $\mathcal{F}_t = \sigma\{\xi_{t-k}, V_{t-k}, k > 0\}$. More particularly for the quasi likelihood theory of the non-stationary ARCH(1) process V_t is a ratio of volatilities or of derivatives of volatilities with volatilities, etc. and satisfies 2.2 due to the asymptotic behavior of the conditional variance process. Besides this, Assumption 2 is trivially satisfied when V_t is independent of t, or more generally when there exists a filtration $(\mathcal{G}_t)_{t\in\mathbb{N}}$ w.r.t. which the process is a (super/sub-) martingale and such that $\sup_{t\in\mathbb{N}} \mathbb{E}\left|V_t\right|^p < +\infty$ for $p \geq 1$ whereas 2.2 follows from results such as Doob's convergence theorem for discrete time martingales. Notice that the previous enable the possibility that the $(V_i)_{i\in\mathbb{N}}$ process is non-stationary.

We are now ready to investigate the issue of the weak limiting behavior of $\sum_{i=1}^{n} \xi_i V_i$. The following is the main result of the section. The essential notion for the derivation of the result is the so called "Principle of Conditioning" of Jakubowski [11].

[2] Notice that when $\alpha = 2$, necessarily $\beta = 0$ and the resulting distribution is the $N(\gamma, 2c)$.

THEOREM 1 *Under Assumptions 1, 2 if $\alpha \neq 1$ then*

$$\frac{1}{n^{1/\alpha} r_n^{1/\alpha}} \sum_{i=1}^{n} (\xi_i - \gamma) V_i \rightsquigarrow \mathbb{E} S_\alpha (\beta \operatorname{sgn}(v), c|v|^\alpha, 0) \tag{2.3}$$

where $\frac{1}{n^{1/\alpha} r_n^{1/\alpha}} \sum_{i=1}^{n} \gamma V_i$ can be omitted if $\alpha < 1$, and when $\alpha = 1$ then

$$\frac{1}{n r_n} \sum_{i=1}^{n} [(\xi_i - \gamma - H(n r_n)) V_i] - 2\beta c v \pi^{-1} (C - \log |v|) \rightsquigarrow \mathbb{E} S_1 (\beta \operatorname{sgn}(v), c|v|, 0) \tag{2.4}$$

where r_n is as specified in Remark 1.[3]

REMARK 3 Notice the following:

1. In the case of $\alpha < 1$, the term $\frac{1}{n^{1/\alpha} r_n^{1/\alpha}} \gamma \sum_{i=1}^{n} V_i$ can be omitted from 2.3 since it converges to zero $\mathbb{P}$ a.s. This is due to Assumption 2, the Cesàro Theorem and the fact that r_n is slowly varying. This cannot hold in the other cases.
2. When $\alpha = 1$ and $\beta = 0$ (symmetry) the terms $H(nr_n)$ and $2\beta c v \pi^{-1} (C - \log |v|)$ vanish from 2.4.
3. The result encompasses the classical theory (see inter alia see Remark 2 of Aaronson and Denker [1]) obtained when $V_t = 1$ $\mathbb{P}$ a.s.
4. To our knowledge, the limiting mixtures are novel results attributed to the "stochasticity" of u. When the latter is constant the mixtures are obviously trivial. In this case the weak limits are α-stable distributions. Notice that as expected (see the previous remark) the properties of the $(V_i)_{i \in \mathbb{N}}$ process do not affect either the scaling sequences, which depend only on the tail behavior of the distribution of ξ_0, nor the index of stability of the limit. Those properties affect the translating sequence when $\alpha = 1$ as well as the remaining parameters of the limit in any case.
5. The result can be easily extended when V_t is $\mathbb{R}^d$-valued for $d > 1$ using the Cramér-Wold device. Then the limits would be mixtures of multivariate α-stable distributions where the (random) spectral measures (for their definition see Paragraph 2 of Mikosch and Straumann [17]) are characterized via linear transformations due to Theorem 2.3 of Gupta et. al. [8].
6. When $\alpha = 2$ and h converges then we obtain asymptotic mixed-normality with rate $\sqrt{n}$ and stochastic variance $2cv^2\rho^2$ where $\rho = \lim h$. On the other hand when h diverges to $+\infty$ then we obtain again mixed-normality with stochastic variance cv^2 albeit at the slower rate $\sqrt{\frac{n}{r_n}}$. For example when $\xi_0 \sim t_2$ then a simple calculation of the truncated second moment implies that $r_n = \log n$. This is obviously a major generalization of the results in Abadir and Magnus [2].
7. When $\alpha < 1$, the support of ξ_0 is bounded from above (below) and $V_0 > 0$ $\mathbb{P}$ a.s. then the support of the limiting distribution is $(-\infty, 0]$ ($[0, +\infty)$).

[3]The form of the result would remain essentially the same had we allowed v to assume the value zero with positive probability. The proof of such an extension would make use of the concept of the Potter bounds (see inter alia Theorem 1.5.6 in Bingham et al. [5]) and the Cesàro Theorem. Since this is not relevant to the application that we have in mind, we do not present this extension for economy of space.

Before establishing our main use of the limit theorem in the following section, we conclude the current one with a simple example in the context of a linear model.

EXAMPLE 1 Consider the simple linear regression model $y_t = \beta x_t + \varepsilon_t$ where $\varepsilon_t = \xi_t u_t$, where (ξ_t) satisfy Assumption 1 with zero mean when $\mathbb{E}|\xi_0| < \infty$. Let $\mathcal{F}_t$ be a filtration of the underlying probability space $(\Omega, \mathcal{G}, \mathbb{P})$ such that ξ_t is independent of $\mathcal{F}_t$, u_t and x_t are adapted to $\mathcal{F}_t$. Furthermore suppose that u_t, x_t are martingales w.r.t. some $(\mathcal{G}_t)_{t\in\mathbb{N}}$ and such that $\sup_{t\in\mathbb{N}} \mathbb{E}|u_t| + \sup_{t\in\mathbb{N}} \mathbb{E}|x_t| < \infty$. Then by Doob's theorem for discrete time martingales there exist random variables u, x such that $u_n \to u$ and $x_n \to x$ as $n \to \infty$ $\mathbb{P}$ a.s. where we further assume that $u, x \neq 0$ $\mathbb{P}$ a.s. Then we may obtain the asymptotic distribution of the OLS estimator β_n since $\frac{1}{n}\sum_{t=1}^{n} x_t^2 \to x^2$ $\mathbb{P}$ a.s. as $n \to \infty$ by the Cesàro mean theorem and $x_t u_t$ satisfies Assumption 2. A simple extension of Theorem 1 that handles joint convergence along with the Continuous Mapping Theorem implies that

$$\frac{n^{\frac{\alpha-1}{\alpha}}}{r_n^{\frac{1}{\alpha}}}(\beta_n - \beta_0) = \left(\frac{1}{n}\sum_{t=1}^{n} x_t^2\right)^{-1} \frac{1}{n^{\frac{1}{\alpha}} r_n^{\frac{1}{\alpha}}} \sum_{t=1}^{n} \xi_t x_t u_t \rightsquigarrow \frac{y}{x}$$

where $y \sim \mathbb{E}S_\alpha(\beta\,\mathrm{sgn}(u), c|u|^\alpha, 0)$ when $\alpha \neq 1$ and

$$\frac{1}{r_n}(\beta_n - \beta_0) - (\gamma + H(nr_n))\left(\frac{1}{n}\sum_{t=1}^{n} x_t^2\right)^{-1} \frac{1}{nr_n}\sum_{t=1}^{n} x_t u_t \rightsquigarrow \frac{y}{x}$$

where $y \sim \mathbb{E}S_1(\beta\,\mathrm{sgn}(u), c|u|, 2\beta c u \pi^{-1}(C - \log|u|))$ when $\alpha = 1$. Notice that the first result implies the inconsistency and the asymptotic non-tightness of the OLSE when $\alpha < 1$.

3. LIMIT THEORY FOR THE QMLE OF THE NON-STATIONARY ARCH(1)

We are now employing the previous result to the QMLE for the non-stationary ARCH(1) model. Define the ARCH(1) process by

$$\begin{aligned} y_t &= \sigma_t z_t, \quad t \geq 0 \\ \sigma_t^2 &= \omega_0 + a_0 y_{t-1}^2, \quad t > 0 \end{aligned}$$

and some initial value σ_0^2, where $(z_t)_{t\in\mathbb{Z}}$ is iid and such that z_0^2 lies in the domain of attraction of an (non-degenerate) α-stable distribution[4]. Thereby we henceforth assume the validity of Assumption 1 for $\xi_t = z_t^2 - 1$. We furthermore suppose that when $\alpha > 1$ then $\mathbb{E}z_0^2 = 1$. We assume that the constant $\omega_0 \geq 0$ is known while the true ARCH parameter $a_0 \geq \exp(-\mathbb{E}\ln z_0^2)$ is unknown.[5] This implies the non-stationarity (either strict or second order) of the process by (inter alia) Theorem 2.1 of Francq and Zakoian [7]. The statistical model is defined as the collection of ARCH(1) processes with ARCH parameter belonging to the parameter space $\Theta = [a_\star, a^\star]$ where $0 < a_\star < a^\star$ or $\Theta = [a_\star, +\infty)$ and in any case such that $a_0 \in \Theta$[6]. We study the asymptotic properties for the QMLE for the unknown parameter

[4]We may use the usual convention that $\mathbb{E}z_0 = 0$ whenever $\alpha > \frac{1}{2}$, or $\alpha = \frac{1}{2}$ and $\mathbb{E}|z_0| < \infty$.

[5]It is easy to see from the results of Jensen and Rahbek [13] that had ω_0 be unknown, it would also be non identifiable.

[6]Regarding the choice of the former parameter space, the choice of an $\alpha^\star > 0$ may be either the result of encompassing further information on the true parameter value, or a requirement for optimization algorithms to function in which case it could be chosen as a sufficiently large number.

via the use of the limit theorem of the previous section. To this purpose we assume the availability of the random variables $y_0, y_1, \ldots, y_n$ from the process whereas the minus quasi likelihood function is defined by

$$\ell_n(a) = \frac{1}{n}\sum_{t=1}^{n}\left[\log h_t(a) + \frac{y_t^2}{h_t(a)}\right],$$

with $h_t(a) = \omega_0 + a y_{t-1}^2$ for $t = 1, \ldots, n$ and the QMLE (say a_n) for the unknown parameter a_0 satisfies

$$\ell_n(a_n) \leq \inf_a \ell_n(a) + \varepsilon_n$$

where $\varepsilon_n = o_p(1)$.[7] The existence of the QMLE is straightforwardly verified by standard arguments of continuity and compactness for the first case of parameter space. For the second case, existence follows due to the fact that Θ is closed and bounded from below, ℓ_n is $\mathbb{P}$ a.s. continuous, and it $\mathbb{P}$ a.s. diverges to $+\infty$ as $a \to +\infty$. The following theorem is the main result of the present section. It among other things makes use of the limit theorem developed in the previous section in order to derive the asymptotic properties of the QMLE in several cases.

THEOREM 2 *For the ARCH(1) model described above suppose that* $\varepsilon_n = o_p\left(n^{(1-\alpha)/\alpha} r_n^{-1/\alpha}\right)$.

1. *Let* $\alpha \in [1, 2]$ *and* $a_0 \in \operatorname{Int}\Theta$.
   (a) *If* $\alpha > 1$ *then*

$$\frac{n^{(\alpha-1)/\alpha}}{r_n^{1/\alpha}}(a_n - a_0) \rightsquigarrow S_\alpha\left(\beta, c a_0^\alpha, 0\right).$$

   (b) *If* $\alpha = 1$ *and* $\mathbb{E}|\xi_1| < \infty$ *then*

$$\frac{1}{r_n}(a_n - a_0) - \frac{1}{r_n}\left[\gamma + H(n r_n)\right]\left[\frac{\partial^2 \ell_n}{\partial a^2}(\bar{\theta}_n)\right]^{-1} \frac{1}{n}\sum_{i=1}^{n}\frac{y_{t-1}^2}{\omega_0 + a_0 y_{t-1}^2} \rightsquigarrow S_1\left(\beta, c a_0, 2\beta c a_0 \pi^{-1}\left(C + \log a_0 - 2a_0 \log a_0\right)\right)$$

2. *Let* $\alpha \in (0, 1)$.
   (a) *If* $\Theta = [a_\star, a^\star]$ *then* $a_n \rightsquigarrow a^\star$ *and thereby the QMLE is inconsistent unless* $a_0 = a^\star$.
   (b) *If* $\Theta = [a_\star, +\infty)$ *then the QMLE is asymptotically non-tight.*

REMARK 4 Notice the following:

[7] It is easy to see that the results that follow would also hold if y_0 is chosen as an arbitrary constant.

1. The results above are obvious generalizations of those of Jensen and Rahbek [13] since they form an almost exhaustive consideration of cases where for the behavior of the fourth moment of z_0 in the context of the non-stationary ARCH(1). In some sense they generalize the results of Mikosch and Straumann [17] for the stationary and ergodic GARCH(1,1) and the results of Hall and Yao [9] for the covariance stationary GARCH(p,q) since they obtain asymptotic results for the QMLE allowing for $\alpha < 1$. Furthermore, the results of Hall and Yao for the case where $\alpha = 1$ where obtained via the imposition of a further restriction on the tail behavior of the distribution of z_0^2 which is not needed here. When we impose the same restriction, we obtain a similar to the aforementioned paper, form for the translating sequence (see Proposition 1 that follows). The results of Mikosh and Straumann for the case where $\alpha = 1$ are restricted to $\beta = 0$ something that is not useful for the consideration of the asymptotic behavior of the QMLE in GARCH-type models.
2. When $\alpha = 2$ and h converges we recover the results of Jensen and Rahbek [13], i.e. the rate is $\sqrt{n}$ and the limit is $N\left(0, \mathbb{E}\left(z_0^4 - 1\right) a_0^2\right)$. When $\alpha = 2$ but h diverges we obtain asymptotic normality but with slower rate. For example if $\sqrt{2}z_0 \sim t_4$ then $\frac{\sqrt{n}}{\sqrt{\log n}}\left(a_n - a_0\right) \rightsquigarrow N\left(0, \frac{3}{2}a_0^2\right)$, which as implied above, is novel in the context of the non-stationary ARCH(1) model.
3. A partial extension of the results in Andrews [3] via the use of Lemma 7.13.2-3 of van der Vaart [21],[8] enables the generalization of the results of 1.(a) in cases where a_0 is a boundary point. For example it is easy to see that if $a_0 = a_\star$ then we would obtain the limit distribution as an appropriate projection of the current α-stable distribution to $[0, +\infty)$, i.e. it would be supported on the latter interval, with an atom at zero of probability equal to the one attributed by the current distribution on $(-\infty, 0]$. The dual case would be analogous. When $\alpha = 1$ the fact that $\beta \neq 0$ implies the presence of the translating sequence, renders the aforementioned result inapplicable. Hence the boundary cases when $\alpha = 1$ cannot be handled via the present methodology and thereby constitute an open question.
4. Open is also the question about the relevant limit theory when $\alpha = 1$ and $\lim h \neq 0$. This cannot be handled by the present methodology and it was analogously not considered by the relevant results of Hall and Yao [9]. We suspect that in this case the estimator is generally inconsistent (see also Example 2 that follows).
5. The results for the case where $\alpha < 1$ are also novel in all the aforementioned relevant literature concerning the limit theory of the QMLE. They are obtained by the use of the concept of epi-convergence (see inter alia Knight [15]) via reductio ad absurdum. They are consistent with a "heuristic" argument that says that multiplication with the rate $\frac{n^{(\alpha-1)/\alpha}}{r_n^{1/\alpha}}$ implies asymptotic tightness in all $\alpha \neq 1$ cases. This argument essentially works both in Example 1 in the previous section as well as in a special case of the considered QMLE presented in Example 2 that follows. The question of whether we obtain generally asymptotic tightness by multiplication with the previous rate as well as a limit distribution such as the ones appearing in the examples remains also open.

[8] See Theorem 5.5 of Arvanitis and Louka [4].

As noted above, the translating sequence appearing in the case of $a = 1$ can obtain a less complex form if a Hall and Yao [9] type of condition (see Theorem 2.1) for the complement of the truncated second moment of z_0 is enforced. This is established in the following Proposition. In such a case the "centering constants" assume a form that is similar to the classical theory for the iid case.

PROPOSITION 1 *Let the conditions of Theorem 2 hold for the case where $\alpha = 1$ and additionally assume that $r_n^{-1}\left[H(nr_n) + \gamma\right]^2 \to 0$. Then*

$$\frac{1}{r_n}(a_n - a_0) - \frac{1}{r_n}\left[\gamma + H(nr_n)\right] a_0 \rightsquigarrow S_1\left(\beta, ca_0, 2\beta ca_0 \pi^{-1}\left(C + \log a_0 - 2a_0 \log a_0\right)\right)$$

We conclude this section with the case of $\omega_0 = 0$. Then the estimator assumes a known functional form from which we can derive its limit theory even when $\alpha = 1$ and the analogous first moment does not exist, or find an appropriate rate and obtain a limit distribution when $\alpha < 1$ even in the presence of non-tightness.

EXAMPLE 2 Let $\alpha \leq 1$ and $\omega_0 = 0$ and suppose that $\Theta = (0, +\infty)$. It is easy to see that $-\infty < \inf_{a\in(0,\infty)} \ell_n(a) < \infty$ $\mathbb{P}$ a.s., so that $\arg\text{zero}_{a\in(0,\infty)}\ell_n'(a) = \frac{1}{n}\sum_{t=1}^n \frac{y_t^2}{y_{t-1}^2} = a_0\frac{1}{n}\sum_{t=1}^n z_t^2$ is the QMLE $\mathbb{P}$ a.s. which is clearly non-tight when $\alpha < 1$ or when $\alpha = 1$ and $\int_{\cdot}^{+\infty} \frac{h(x)}{x} dx$ diverges. Using 1 we have that when $\alpha \neq 1$

$$\frac{n^{(\alpha-1)/\alpha}}{r_n^{1/\alpha}}(a_n - a_0) \rightsquigarrow S_\alpha\left(\beta, ca_0^\alpha, 0\right)$$

and when $\alpha = 1$

$$\frac{1}{r_n}(a_n - a_0) - \frac{a_0}{r_n}\left(\gamma + H(nr_n)\right) - 2\beta ca_0\pi^{-1}\left(C - \log a_0\right) \rightsquigarrow S_1\left(\beta, ca_0, 0\right).$$

Hence when $\alpha < 1$ or when $\alpha = 1$ and the second moment of z_0 does not exist the estimator is asymptotically non tight.

4. PARALLEL AND FURTHER RESEARCH

In this note we have demonstrated how the derivation and use of a MLT with mixed stable limits can widely extend standard results for the limit theory of the popular QMLE in the context of a simple example, i.e. the non-stationary ARCH(1) model. Given the expository character and the limited applicability of our derivation, we are currently working on a similar MLT involving stationarity, that avoids many of the strict conditions and limitations appearing in the analogous results of Hall and Yao [9] and Mikosh and Straumann [17], and is applicable to more classes of conditionally heteroskedastic models beyond the standard stationary GARCH, along with potentially interesting statistical applications such as the robustification of Wald tests. Furthermore, this in conjunction with the current MLT would possibly allow the extension of our results to the non-stationary GARCH(p,q) model in the spirit of Jensen and Rahbek [14] and of Chan and Ng [6] at least for the case where $\alpha > 1$, as otherwise, the issues of consistency and tightness would possibly require different arguments. We are also working towards coping with the open questions raised above, such as the relevant limit theory when $\alpha = 1$, yet $\mathbb{E}z_0^2 = +\infty$ and/or the parameter is on the boundary, etc. We are also interested in the obvious extension of the relevant results of Wang [22], something that would drastically widen the scope of potential applications including non-linear cointegration, and possibly allowing for the consideration of the limit theory of the QMLE in non-stationary versions of more complex models, e.g. the EGARCH. This however would require a non-trivial extension of our methodology and we leave it for future research.

REFERENCES

[1] Aaronson, J., & Denker, M. (1998). Characteristic functions of random variables attracted to 1-stable laws. Annals of probability, 399-415. 2, 3, 4, 11

[2] Abadir, K., & Magnus, J. (2004). 03.6.1. The Central Limit Theorem for Student's Distribution—Solution. Econometric theory, 20(6), 1261-1262. 4

[3] Andrews, D. W. K., "Estimation when a parameter is on boundary", Econometrica, 67:1341-1383, 1999. 7

[4] Arvanitis, S., & Louka, A. A CLT For Martingale Transforms With Slowly Varying Second Moments and the Limit Theory of the QMLE for Conditionally Heteroskedastic Models. Working paper 17/2013, Dept. of Economics, AUEB. url: http://www.econ.aueb.gr/uploadfiles/AIIdp172013.pdf. 7

[5] Bingham, N. H., Goldie, C. M., & Teugels, J. L. (1989). Regular variation (Vol. 27). Cambridge university press. 2, 4

[6] Chan, N. H., & Ng, C. T. (2009). Statistical inference for non-stationary GARCH (p, q) models. Electronic Journal of Statistics, 3, 956-992. 9

[7] Francq, C., & Zakoian, J. M. (2011). GARCH models: structure, statistical inference and financial applications. John Wiley & Sons. 5

[8] 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. 4

[9] Hall, P., & Yao, Q. (2003). Inference in ARCH and GARCH models with heavy–tailed errors. Econometrica, 71(1), 285-317. 1, 7, 8

[10] Ibragimov, I. A. (1971). Independent and stationary sequences of random variables. 2

[11] Jakubowski, A. (1986). Principle of conditioning in limit theorems for sums of random variables. The Annals of Probability, 902-915. 3

[12] Jakubowski, A. (2012). Principle of Conditioning revisited. Demonstratio Mathematica, 45(2). 1, 10

[13] Jensen, S. T., & Rahbek, A. (2004). Asymptotic normality of the QMLE estimator of ARCH in the nonstationary case. Econometrica, 72(2), 641-646. 1, 5, 7

[14] Jensen, S. T., and Rahbek, A. (2004). Asymptotic inference for Non Stationary GARCH. Econometric Theory, 20, 1203–1226. 9

[15] Knight, K. (1999). Epi-convergence in distribution and stochastic equi-semicontinuity. Working Paper, Department of Statistics, University of Toronto. 7, 15
[16] Kwapien, S., & Woyczynski, W. A. (1992). Random series and stochastic integrals. Birkhäuser.
[17] Mikosch, T., & Straumann, D. (2006). Stable limits of martingale transforms with application to the estimation of GARCH parameters. The Annals of Statistics, 493-522. 1, 2, 4, 7, 8
[18] Molchanov, Ilya (2006). Theory of random sets. Springer. 14, 15
[19] Nelson, D. B. (1990). Stationarity and persistence in the GARCH (1, 1) model. Econometric theory, 6(03), 318-334. 15
[20] Straumann, D. (2005). Estimation in conditionally heteroscedastic time series models (Vol. 181). New York: Springer. 15
[21] van der Vaart, A.W., "Asymptotic Statistics" CUP, 1997. 7
[22] Wang, Q. (2014). Martingale Limit Theory Revisited and Non Linear Cointegrating Relations. Econometric Theory, 30, 509–535. 9

APPENDIX

This first part of the appendix contains the proofs of the main results. The second part contain auxiliary results used in the proof of Theorem 1.

Main Proofs

PROOF OF THEOREM 1: By the "Main Lemma for Sequences" of Jakubowski [12] the result would follow if we proved that for all $t \in \mathbb{R}$

$$\prod_{i=1}^{n} \mathbb{E}\left(\exp\left(it\frac{1}{n^{\frac{1}{\alpha}}r_n^{\frac{1}{\alpha}}}\rho_{i,\alpha}\right)/\mathcal{F}_i\right) \tag{4.1}$$

converges in probability to the characteristic function of $S_\alpha\left(\beta\,\mathrm{sgn}(v), c|v|^\alpha, 0\right)$, with $\rho_{t,\alpha} = \begin{cases}(\xi_t-\gamma)V_t, \alpha\neq 1\\ [(\xi_t-\gamma-H(nr_n))V_t]-r_n 2\beta c v\pi^{-1}(C-\log|v|), \alpha=1\end{cases}$. Assume that the representation described in Assumption 1 holds for all $t \in (-t_0, t_0)$, and some $t_0 > 0$. Then notice that for any $t \neq 0$ by defining the event

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

where $K_t < \frac{t_0}{|t|}$ and using the same argument as in the proof of Lemma 1, we have that $\mathbb{P}(C^c_{n,K}) = \mathbb{P}\left(\max_{1\leq i\leq n}|V_i| > K_t r_n^{\frac{1}{\alpha}} n^{\frac{1}{\alpha}}\right) \to 0$. For the case of $\alpha \neq 1$ due to Assumptions

1-2, if $\omega \in C_{n,K}$ then $\sum_{i=1}^{n} \log \mathbb{E}\left(\exp\left(it\frac{1}{n^{\frac{1}{\alpha}} r_n^{\frac{1}{\alpha}}}\left(\xi_k - \gamma\right) V_k\right) / \mathcal{G}_n\right)$ equals

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

which by Lemma 1 converges $\mathbb{P}$ a.s. to $-c|t|^{\alpha}|v|^{\alpha}\left(1 - i\beta\,\mathrm{sgn}(tv)\tan\left(\frac{1}{2}\pi\alpha\right)\right)$.

For the case where $\alpha = 1$, again due to the relevant parts of Assumptions 1-2, if $\omega \in C_{n,K}$ then $\sum_{i=1}^{n} \log \mathbb{E}\left(\exp\left(it\frac{1}{nr_n}\left(\xi_i - \gamma - H(nr_n)\right) V_i\right) / \mathcal{G}_n\right)$ analogously has a local to zero representation due to Theorem 2 of Aaronson and Denker [1] as

$$
\begin{aligned}
-c|t|\frac{1}{nr_n}\sum |V_i| h\left(nr_n|tV_i|^{-1}\right) + i2\beta c\pi^{-1} Ct\frac{1}{nr_n}\sum V_i h\left(nr_n|tV_i|^{-1}\right)\\
+ it\frac{1}{nr_n}\sum V_i\left[H\left(nr_n|tV_i|^{-1}\right) - H(nr_n)\right]
\end{aligned} \tag{4.2}
$$

where, by Lemma 1, $\frac{1}{nr_n}\sum |V_i| h\left(nr_n|tV_i|^{-1}\right) \to |v|$ and $\frac{1}{nr_n}\sum V_i h\left(nr_n|tV_i|^{-1}\right) \to v$ $\mathbb{P}$ a.s. Furthermore, using Lemma 2 we have that

$$
\begin{aligned}
&\frac{1}{nr_n}\sum V_i\left[H\left(nr_n|tV_i|^{-1}\right) - H(nr_n)\right]\\
=&\frac{h(nr_n)}{r_n} 2\beta c\pi^{-1}\log\frac{1}{|t|}\frac{1}{n}\sum V_i - \frac{h(nr_n)}{r_n}\frac{1}{n}\sum V_i\log|V_i| + o(1)\\
\to& 2\beta c\pi^{-1} v\left(\log\frac{1}{|t|} - \log|v|\right)\ \mathbb{P}\ a.s.
\end{aligned}
$$

Therefore 4.2 becomes $-c|v||t|\left[1 - i2\beta c\pi^{-1}\,\mathrm{sgn}(tv)\log\frac{1}{|t|}\right] + it2\beta cv\pi^{-1}\left(C - v\log|v|\right)$ and the result follows. *Q.E.D.*

Proof of Theorem 2: For 1 we first have to establish consistency. First notice that θ_n will also satisfy $\ell_n^*(\theta_n) \le \inf_\theta \ell_n^*(\theta) + \varepsilon_n$ where $\ell_n^*(\theta) = \ell_n(\theta) - \frac{1}{n}\sum \log\sigma_t^2 = \frac{1}{n}\sum_{t=1}^{n}\left[\frac{y_t^2}{h_t(\theta)} - \log\frac{\sigma_t^2}{h_t(\theta)}\right]$. Then observe that $\forall\theta \in \Theta$ $\ell_n^*(\theta) = \frac{1}{n}\sum_{t=1}^{n}\left[\frac{y_t^2}{h_t(\theta)} - \log\frac{\sigma_t^2}{h_t(\theta)}\right] = \frac{1}{n}\sum_{t=1}^{n}\frac{\omega_0 + a_0 y_{t-1}^2}{\omega_0 + a y_{t-1}^2} z_t^2 - \frac{1}{n}\sum_{t=1}^{n}\log\frac{\omega_0 + a_0 y_{t-1}^2}{\omega_0 + a y_{t-1}^2} \rightsquigarrow \frac{a_0}{a} - \log\frac{a_0}{a}$ by 3 and the Cesàro mean theorem, which is uniquely minimized at a_0. In order to show that ℓ_n^* converges locally uniformly to the above limit, it suffices to show that ℓ_n^* is stochastically equicontinuous. We have that

$$
|\ell_n^*(a') - \ell_n^*(a)| \le \frac{1}{n}\sum_{i=1}^{n}\left|\frac{1}{h_t(a')} - \frac{1}{h_t(a)}\right| y_t^2 + \frac{1}{n}\sum_{i=1}^{n}\left|\log h_t(a) - \log h_t(a')\right|.
$$

But the first term of the right hand side of the above display equals $|a-a'|\frac{1}{n}\sum_{i=1}^{n}\frac{y_{t-1}^2 y_t^2}{(\omega_0+a'y_{t-1}^2)(\omega_0+ay_{t-1}^2)} \leq \frac{|a-a'|}{a'}\frac{1}{n}\sum_{i=1}^{n}\frac{y_t^2}{\omega_0+ay_{t-1}^2} \rightsquigarrow \frac{|a-a'|}{aa'}$. Also the second term by the mean value theorem is lower or equal to $|a-a'|\frac{1}{n}\sum_{i=1}^{n}\frac{y_{t-1}^2}{\omega_0+\min\{a,a'\}y_{t-1}^2} \xrightarrow{a.s.} \frac{|a-a'|}{\min\{a,a'\}}$. The result follows easily.

1.(a) First we show that for any $a_n^* \rightsquigarrow a_0$

$$\frac{\partial^2 \ell_n}{\partial a^2}(\theta_n^*) \rightsquigarrow \frac{1}{a_0^2}.$$

Trivial calculations show that

$$\frac{\partial^2 \ell_n}{\partial a^2}(a_n^*) = \frac{2}{n}\sum_{t=1}^{n}\frac{\left(\omega_0+a_0 y_{t-1}^2\right)y_{t-1}^4}{\left(\omega_0+a_n^* y_{t-1}^2\right)^3}z_t^2 - \frac{1}{n}\sum\frac{y_{t-1}^4}{\left(\omega_0+a_n^* y_{t-1}^2\right)^2}. \tag{4.3}$$

The second term of the latter converges $\mathbb{P}$ a.s. to $\frac{1}{a_0^2}$, since by an application of the mean value theorem we have that

$$\frac{1}{n}\sum y_{t-1}^4\left[\left(\omega_0+a_n^* y_{t-1}^2\right)^{-2}-\left(\omega_0+a_0 y_{t-1}^2\right)^{-2}\right] \leq 2\left|a_n^*-a_0\right|\sum\frac{y_{t-1}^6}{\left(\omega_0+a_\star y_{t-1}^2\right)^3} \rightsquigarrow 0$$

by the Cesàro mean theorem since $\frac{y_{t-1}^6}{\left(\omega_0+a_\star y_{t-1}^2\right)^3} \xrightarrow{\mathbb{P}a.s.} \frac{1}{a_\star}$ and the fact that $a_n^* \rightsquigarrow a_0$. Furthermore, regarding the first term of 4.3 we have that

$$\begin{aligned}\frac{2}{n}\sum\frac{\left(\omega_0+a_0 y_{t-1}^2\right)y_{t-1}^4}{\left(\omega_0+a_n^* y_{t-1}^2\right)^3}z_t^2 &= \frac{2}{n^{\frac{\alpha-1}{\alpha}}r_n^{-\frac{1}{\alpha}}}\frac{1}{n^{\frac{1}{a}}r_n^{\frac{1}{\alpha}}}\sum\frac{\left(\omega_0+a_0 y_{t-1}^2\right)y_{t-1}^4}{\left(\omega_0+a_n^* y_{t-1}^2\right)^3}\left(z_t^2-1\right)\\ &+ \frac{2}{n}\sum\frac{\left(\omega_0+a_0 y_{t-1}^2\right)y_{t-1}^4}{\left(\omega_0+a_n^* y_{t-1}^2\right)^3}.\end{aligned}$$

Now by employing Theorem 1, we can show that the first term of the right hand side is $O_p\left(n^{\frac{1-\alpha}{\alpha}}r_n^{\frac{1}{\alpha}}\right)=o_p(1)$. The second term converges to $\frac{2}{a_0^2}$ $\mathbb{P}$ a.s. since

$$\begin{aligned}&\frac{2}{n}\sum y_{t-1}^4\left(\omega_0+a_0 y_{t-1}^2\right)\left|\frac{1}{\left(\omega_0+a_n^* y_{t-1}^2\right)^3}-\frac{1}{\left(\omega_0+a_0 y_{t-1}^2\right)^3}\right|\\ &\leq \left|a_n^*-a_0\right|\frac{6}{n}\sum\frac{y_{t-1}^6}{\left(\omega_0+a_\star y_{t-1}^2\right)^3} \xrightarrow{a.s.} 0\end{aligned}$$

by the Cesàro mean theorem since $\frac{y_{t-1}^4}{\left(\omega_0+a_0 y_{t-1}^2\right)^2} \xrightarrow{a.s.} \frac{1}{a_0}$ and the fact that $a_n^* \rightsquigarrow a_0$. Furthermore by Theorem 1,

$$-n^{\frac{\alpha-1}{\alpha}}r_n^{-\frac{1}{\alpha}}\frac{\partial \ell_n(\theta_0)}{\partial a} = \frac{1}{n^{\frac{1}{\alpha}}r_n^{\frac{1}{\alpha}}}\sum_{i=1}^{n}(z_t^2-1)\frac{y_{t-1}^2}{\omega_0+a_0 y_{t-1}^2} \rightsquigarrow S_\alpha\left(\beta,\frac{c}{a_0^\alpha}\right).$$

Hence if $\varepsilon_n = o_p\left(r_n^{\frac{1}{\alpha}} n^{-\frac{2(\alpha-1)}{\alpha}}\right)$ then

$$\frac{n^{\frac{\alpha-1}{\alpha}}}{r_n^{\frac{1}{\alpha}}}(\theta_n - \theta_0) \rightsquigarrow a_0^2 S_\alpha\left(\beta, \frac{c}{a_0^\alpha}\right) = S_\alpha\left(\beta, c a_0^\alpha\right).$$

1.(b) Note that the previous result concerning the asymptotic behavior of the Hessian still holds as

$$\begin{aligned}\frac{2}{n}\sum \frac{\left(\omega_0 + a_0 y_{t-1}^2\right) y_{t-1}^4}{\left(\omega_0 + a_n^* y_{t-1}^2\right)^3} z_t^2 &= \frac{r_n}{n r_n}\sum \frac{\left(\omega_0 + a_0 y_{t-1}^2\right) y_{t-1}^4}{\left(\omega_0 + a_n^* y_{t-1}^2\right)^3}\left(z_t^2 - 1 - \gamma - H(n r_n)\right) \\ &+ \left(1 + \gamma + H(n r_n)\right)\frac{2}{n}\sum \frac{\left(\omega_0 + a_0 y_{t-1}^2\right) y_{t-1}^4}{\left(\omega_0 + a_n^* y_{t-1}^2\right)^3}\end{aligned}$$

where the first term on the right hand side is $O_p(r_n) = o_p(1)$ by Theorem 1 and the second term converges almost surely to $\left(1 + \gamma + \lim_{n\to\infty} H(n r_n)\right)\frac{2}{a_0^2} = \frac{2}{a_0^2}$ as previously and the fact that $\gamma + \lim_{n\to\infty} H(n r_n) = \mathbb{E} z_0^2 - 1 = 0$. Therefore

$$\begin{aligned}\frac{\partial^2 \ell_n}{\partial a^2}(a_n^*) &= \frac{1}{n}\sum \frac{\left(\omega_0 + a_0 y_{t-1}^2\right) y_{t-1}^4}{\left(\omega_0 + a_n^* y_{t-1}^2\right)^3}\left(z_t^2 - 1 - \gamma - H(n r_n)\right) \\ &+ \left(1 + \gamma + H(n r_n)\right)\frac{2}{n}\sum \frac{\left(\omega_0 + a_0 y_{t-1}^2\right) y_{t-1}^4}{\left(\omega_0 + a_n^* y_{t-1}^2\right)^3} \\ &- \frac{1}{n}\sum \frac{y_{t-1}^4}{\left(\omega_0 + a_n^* y_{t-1}^2\right)^2} \rightsquigarrow \frac{1}{a_0^2}.\end{aligned}$$

By Theorem 1 we obtain that

$$\frac{1}{n r_n}\sum_{i=1}^n (z_t^2 - 1 - \gamma - H(n r_n))\frac{y_{t-1}^2}{\omega_0 + a_0 y_{t-1}^2} \rightsquigarrow S_1\left(\beta, c a_0^{-1}, 2\beta c a_0^{-1}\pi^{-1}\left(C + \log a_0\right)\right).$$

But

$$-\frac{1}{r_n}\frac{\partial \ell_n(a_0)}{\partial a} = \frac{1}{n r_n}\sum_{i=1}^n (z_t^2 - 1 - \gamma - H(n r_n))\frac{y_{t-1}^2}{\omega_0 + a_0 y_{t-1}^2} + \frac{\gamma + H(n r_n)}{n r_n}\sum_{i=1}^n \frac{y_{t-1}^2}{\omega_0 + a_0 y_{t-1}^2}.$$

Furthermore, note that since $\varepsilon_n = o_p(r_n)$ we have that $\frac{\partial \ell_n(a_n)}{\partial a} = o_p(r_n)$. Thus

$$\begin{aligned}&\left[\frac{\partial^2 \ell_n}{\partial a^2}(\bar{a}_n)\right]\frac{1}{r_n}(a_n - a_0) = -\frac{1}{r_n}\frac{\partial \ell_n(a_0)}{\partial a} \\ &= \frac{1}{n r_n}\sum_{i=1}^n (z_t^2 - 1 - \gamma - H(n r_n))\frac{y_{t-1}^2}{\omega_0 + a_0 y_{t-1}^2} + \frac{\gamma + H(n r_n)}{r_n}\frac{1}{n}\sum_{i=1}^n \frac{y_{t-1}^2}{\omega_0 + a_0 y_{t-1}^2}.\end{aligned}$$

Therefore

$$\frac{1}{r_n}(a_n - a_0) - \frac{1}{r_n}[\gamma + H(nr_n)]\left[\frac{\partial^2 \ell_n}{\partial a^2}(\bar{a}_n)\right]^{-1}\frac{1}{n}\sum_{i=1}^{n}\frac{y_{t-1}^2}{\omega_0 + a_0 y_{t-1}^2}$$
$$= \left[\frac{\partial^2 \ell_n}{\partial a^2}(\bar{a}_n)\right]^{-1}\frac{1}{nr_n}\sum_{i=1}^{n}(z_t^2 - 1 - \gamma - H(nr_n))\frac{y_{t-1}^2}{\omega_0 + a_0 y_{t-1}^2}$$
$$\rightsquigarrow S_1\left(\beta, ca_0, 2\beta ca_0\pi^{-1}\left(C + \log a_0 - 2a_0\log a_0\right)\right)$$

for some $\bar{a}_n$ “between” a_n and a_0.

2.(a) Let

$$\ell_n^* = \frac{1}{n^{\frac{1}{\alpha}} r_n^{\frac{1}{\alpha}}}\sum_{i=1}^{n}\log\sigma_t^2 - \frac{n}{n^{\frac{1}{\alpha}} r_n^{\frac{1}{\alpha}}}\ell_n(a)$$
$$= \frac{1}{n^{\frac{1}{\alpha}} r_n^{\frac{1}{\alpha}}}\sum_{i=1}^{n}\log\frac{\sigma_t^2}{h_t(a)} - \frac{1}{n^{\frac{1}{\alpha}} r_n^{\frac{1}{\alpha}}}\sum_{i=1}^{n}\frac{\sigma_t^2}{h_t(a)}z_t^2$$

so that a_n is an approximate maximizer of the latter where now the approximation error equals $-\varepsilon_n$. Now notice that, due to the Cesàro mean theorem

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

since

$$\sup_{\alpha}\frac{1}{n}\left|\sum_{i=1}^{n}\log\frac{\sigma_t^2}{h_t} - \log\frac{\alpha_0}{\alpha}\right| \leq C\frac{1}{n}\sup_{\alpha}\sum\left|\frac{\sigma_t^2}{h_t} - \frac{a_0}{a}\right|$$
$$\leq C\sup_{\alpha}\frac{1}{n}\sum\left|\frac{a\omega_0 + aa_0 y_{t-1}^2 - a_0\omega_0 - aa_0 y_{t-1}^2}{\alpha h_t}\right|$$
$$= C|a^\star - a_0|\omega_0\frac{1}{n}\sum\frac{1}{\omega_0 + a_\star y_{t-1}^2}$$
$$\to 0 \quad \mathbb{P}\text{ a.s.}$$

Theorem 1 applies for $\frac{1}{n^{\frac{1}{\alpha}} r_n^{\frac{1}{\alpha}}}\sum_{i=1}^{n}\frac{\sigma_t^2}{h_t(a)}z_t^2$ since Assumptions 1 and 2 hold as $\frac{\sigma_n^2}{h_n(a)} \to \frac{a_0}{a}$. Thereby

$$\ell_n^\star(a) \rightsquigarrow -S_\alpha\left(1, c\left(\frac{a_0}{a}\right)^\alpha\right)$$

locally uniformly. Notice that by construction the support of S_α must be $[0,\infty)$ hence $\beta = 1$. Now $S_\alpha\left(1, C\left(\frac{a_0}{a}\right)^\alpha\right) = -\frac{a_0}{a}S_\alpha(1, C)$ and $S_\alpha(1, C)$ cannot assume negative values. Hence due to Theorem 3.4 of Molchanov [18] and the compactness of the parameter space, $\arg\max a_n \rightsquigarrow \arg\max\left(-\frac{a_0}{a}S_\alpha(1,C)\right) = \arg\min\left(\frac{a_0}{a}S_\alpha(1,C)\right) = a^*$ $\mathbb{P}$ a.s. Hence a_n is inconsistent unless $a_0 = a^\star$.

2.(b) Locally uniform (on compacta) weak convergence of $n^{1-\frac{1}{\alpha}}r_n^{-\frac{1}{\alpha}}\ell_n(a)$ to the previously established limit also holds. This implies that the former weakly epiconverges to the latter (see Knight [15]). Both are lower semi-continuous (lsc) $\mathbb{P}$ a.s. and the space of lower semi-continuous functions with the topology of epiconvergence can be metrized as complete and separable (see again Knight [15]). Suppose now that (a_n) is asymptotically tight. Then by Prokhorov's Theorem, there exists a random element $a : a_{k_n} \rightsquigarrow a$ along some subsequence. Due to separability and Skorohod representation there exists a suitable probability space and random elements $\ell_n^* \overset{d}{=} n^{1-\frac{1}{\alpha}}r_n^{-\frac{1}{\alpha}}\ell_n$, $\ell^* \overset{d}{=} \frac{a_0}{a}S_\alpha(\beta, ca_0^\alpha)$, $\varepsilon_n^* = o_p\left(n^{-\frac{1-\alpha}{\alpha}}r_n^{-\frac{1}{\alpha}}\right)$ and $\ell_n^* \overset{p}{\longrightarrow} \ell^*$ $\mathbb{P}^*$ a.s., and $\theta_{k_n}^* := \varepsilon_{k_n}^* - \arg\max \ell_{k_n}^* \overset{d}{=} a_{k_n}$, $a^* \overset{d}{=}$ and $a_{k_n}^* \to a^*$ $\mathbb{P}$ a.s. But due to the Theorem 3.4 of Molchanov [18], $a^* \in \arg\max \ell^*$. Since a^* has a well defined distribution, there exists a measurable selection say $T : a^* = T \circ \arg\max \circ \ell^*$. Hence $a \overset{d}{=} a^* = T(\arg\max(\ell^*)) \overset{d}{=} T(\arg\max(\ell))$. Thereby $a \overset{d}{=} x$ for some $x \in \arg\max \ell$. But $\arg\max \ell = \emptyset$. *Q.E.D.*

Proof of Proposition 1: Theorem 2 implies that $\frac{\gamma + H(nr_n)}{r_n}(a_n - a_0) \rightsquigarrow 0$. Using the mean value theorem (MVT) and the Cesàro mean theorem we can show that $\frac{\partial^2 \ell_n}{\partial a^2}(\bar{\theta}_n) = \frac{1}{n}\sum_{t=1}^n \frac{y_{t-1}^4}{(\omega_0 + a_0 y_{t-1})^2} + O(|a_n - a_0|)$, $\mathbb{P}$ a.s. Then by the MVT this implies that $\left[\frac{\partial^2 \ell_n}{\partial a^2}(\bar{\theta}_n)\right]^{-1} = \left[\frac{1}{n}\sum_{t=1}^n \frac{y_{t-1}^4}{(\omega_0 + a_0 y_{t-1})^2}\right]^{-1} + O(|a_n - a_0|)$ $\mathbb{P}$ a.s. Then notice that $\left|\frac{1}{n}\sum_{i=1}^n \frac{y_{t-1}^2}{\omega_0 + a_0 y_{t-1}^2} - \frac{1}{a_0}\right| \leq \frac{\omega_0}{a_0}\frac{1}{n}\sum_{t=1}^n \frac{1}{\omega_0 + a_0 y_{t-1}^2}$. Furthermore, a few calculations show that the distance of $\frac{1}{n}\sum_{t=1}^n \frac{y_{t-1}^4}{(\omega_0 + a_0 y_{t-1})^2}$ from its limit is of the same order of magnitude. Then if we showed that $\sum_{t=1}^n \frac{1}{y_{t-1}^2}$ converges $\mathbb{P}$ a.s. together with the MVT the result would follow. But, when $\mathbb{E}\log a_0 z_0^2 > 0$ we can use analogous arguments as in the proof of Theorem 2 of Nelson [19] to show that $\sigma_t^2 \to \infty$ exponentially fast $\mathbb{P}$ a.s. in the sense that there exists $0 < \gamma < 1$ such that $\gamma^t \sigma_t^2 \to \infty$ $\mathbb{P}$ a.s. Furthermore we have that $\mathbb{E}\log^+ z_0^{-2} = -\mathbb{E}\log z_0^2 1\{z_0^2 \leq 1\} < \infty$ since$\mathbb{E}\log z_0^2 = \mathbb{E}\log z_0^2 1\{z_0^2 \leq 1\} + \mathbb{E}\log z_0^2 1\{z_0^2 > 1\} < \infty$. Then Proposition 2.5.1 of Straumann [20] applies to show that $\sum_{t=1}^n \frac{1}{y_{t-1}^2}$ converges $\mathbb{P}$ a.s. In the case where $\mathbb{E}\log a_0 z_0^2 = 0$ note that due to the law of iterated logarithm we have that $\mathbb{P}$ a.s. $\limsup_{t\to\infty}\exp(-\sqrt{t})\sigma_t^2 \geq \omega_0\left[\limsup_{t\to\infty}\sqrt{t}\left(\log\log t \frac{1}{\sqrt{t\log\log t}}\sum \log a_0 z_0^2 - 1\right)\right] = \infty$. Also $\mathbb{E}\left(\log^+ z_0^{-2}\right)^2 = -\mathbb{E}\left(\log z_0^2 1\{z_0^2 \leq 1\}\right)^2 < \infty$ since $\int_0^1 (\log x)^2 dx < \infty$. Then, by using a modified version of Lemma 2.5.2 (replacing ρ^t by $\rho^{\sqrt{t}}$) of Straumann [20] Proposition 2.5.1 therein can be applied to show that $\sum_{t=1}^n \frac{1}{y_{t-1}^2}$ converges $\mathbb{P}$ a.s. as well. *Q.E.D.*

Auxiliary Results

Lemma 1 $\frac{1}{nr_n}\sum_{i=1}^n \operatorname{sgn}(V_i)|V_i|^\alpha h\left(n^{\frac{1}{\alpha}}r_n^{\frac{1}{\alpha}}|V_i|^{-1}\right) \to \operatorname{sgn}(v)|v|^\alpha$ $\mathbb{P}$ *a.s.*

Proof: First notice that due to Assumption 2 and with no loss of generality in assuming that $V_i \neq 0 \,\forall i \in \mathbb{N}$, for any ω contained in a subset of Ω of $\mathbb{P}$ probability 1 we have $0 < \inf_{i\in\mathbb{N}} V_i(\omega) \leq \sup_{i\in\mathbb{N}} V_i(\omega) < \infty$ so that $V_i(\omega)$ is contained in a compact set dependent on the choice of ω $\forall i \in \mathbb{N}$. Furthermore, using the fact that $r_n^{-1} h\left(n^{\frac{1}{\alpha}} r_n^{\frac{1}{\alpha}}\right) \to 1$, we have that

$$\frac{1}{n}\sum_{i=1}^{n}\left[\frac{h\left(n^{\frac{1}{\alpha}} r_n^{\frac{1}{\alpha}}\left|V_i\right|^{-1}\right)}{h\left(n^{\frac{1}{\alpha}} r_n^{\frac{1}{\alpha}}\right)}-1\right]\operatorname{sgn}(V_i)|V_i|^{\alpha}$$
$$\leq \sup_{1\leq i\leq n}\left|\frac{h\left(n^{\frac{1}{\alpha}} r_n^{\frac{1}{\alpha}}\left|V_i\right|^{-1}\right)}{h\left(n^{\frac{1}{\alpha}} r_n^{\frac{1}{\alpha}}\right)}-1\right|\frac{1}{n}\sum_{i=1}^{n}|V_i|^{\alpha} \to 0 \quad \mathbb{P} \text{ a.s.}$$

by the Uniform Convergence Theorem for slowly varying functions and the Cesàro mean theorem. The result follows as $\frac{1}{n}\sum_{i=1}^{n}\operatorname{sgn}(V_i)|V_i|^{\alpha} \to \operatorname{sgn}(v)|v|^{\alpha}$ $\mathbb{P}$ a.s. by the Cesàro mean theorem. *Q.E.D.*

LEMMA 2 *For any compact subset K of $\mathbb{R}_{++}$, we have that*

$$\sup_{k\in K}\left|H\left(k\lambda\right)-H(\lambda)-h(\lambda)\log k\right| = o(h(\lambda))$$

as $\lambda \to \infty$.

PROOF: We have that $H(k\lambda)-H(\lambda) = \int_{\lambda}^{k\lambda}\frac{xh(x)}{1+x^2}dx$. But $\int_{\lambda}^{k\lambda}\frac{xh(x)}{1+x^2}dx - \int_{\lambda}^{k\lambda}\frac{h(x)}{x}dx = -\int_{\lambda}^{k\lambda}\frac{h(x)}{x(1+x^2)}dx = -\int_{1}^{k}\frac{h(\lambda x)}{x(1+\lambda^2 x^2)}dx = -h(\lambda)\int_{1}^{k}\frac{1}{x(1+\lambda^2x^2)}\frac{h(\lambda x)}{h(\lambda)}dx$. Then it is easy to show that the supremum of the latter over $k \in K$ is $o(h(\lambda))$ by applying the bounded convergence theorem. Next notice that $\int_{1}^{k}\frac{h(\lambda x)}{x}dx - h(\lambda)\log k = h(\lambda)\int_{1}^{k}\frac{1}{x}\left[\frac{h(\lambda x)}{h(\lambda)}-1\right]dx$, so its supremum over $k \in K$ will also be $o(h(\lambda))$. *Q.E.D.*

LEMMA 3 *Suppose that $\alpha > 1$ or $\alpha = 1$ and $\mathbb{E}|\xi_1| < \infty$ together with assumptions 1 and 2. Then $\frac{1}{n}\sum_{i=1}^{n}\xi_i V_i \xrightarrow{p} v\mathbb{E}\xi_1$.*

PROOF: For the case where $\alpha > 1$ we have that $\frac{1}{n}\sum_{i=1}^{n}\xi_i V_i = \frac{1}{n}\sum_{i=1}^{n}\left(\xi_i - \mathbb{E}\xi_1\right)V_i + \mathbb{E}\xi_1\frac{1}{n}\sum_{i=1}^{n}V_i$. But by Theorem 1, $\sum_{i=1}^{n}\left(\xi_i-\mathbb{E}\xi_1\right)V_i = O_p\left(n^{\frac{1}{\alpha}} r_n^{\frac{1}{\alpha}}\right)$, thus $\frac{1}{n}\sum_{i=1}^{n}\left(\xi_i-\mathbb{E}\xi_1\right)V_i = o_p(1)$. Using the Cesàro mean theorem the result follows. For the case where $\alpha = 1$ we have that $\frac{1}{n}\sum_{i=1}^{n}\xi_i V_i = \frac{1}{n}\sum_{i=1}^{n}\left(\xi_i-\gamma-H(nr_n)\right)V_i + \left(\gamma+H(nr_n)\right)\frac{1}{n}\sum_{i=1}^{n}V_i$. Again, by Theorem 1, $\frac{1}{n}\sum_{i=1}^{n}\left(\xi_i-\gamma-H(nr_n)\right)V_i = O_p(r_n)$. Thus, $\frac{1}{n}\sum_{i=1}^{n}\left(\xi_i-\gamma-H(nr_n)\right)V_i = o_p(1)$ since $r_n \to 0$. Furthermore, by the Cesàro mean theorem together with the fact that $\gamma + \lim_{n\to\infty}H(nr_n) = \mathbb{E}\xi_1$ the result follows. *Q.E.D.*