Statistica Sinica **21** (2011), 1191-1200

TAIL BEHAVIOR AND OLS ESTIMATION IN AR-GARCH MODELS

Theis Lange

University of Copenhagen

Abstract: The scope of this paper is twofold. We first describe the tail behavior for general AR-GARCH processes and hence extend the results of Basrak, Davis, and Mikosch (2002b) to another empirical relevant model class. Second, and primarily, we study properties for the OLS estimator in general AR-GARCH model. Specifically it is shown that the OLS estimator of the autoregressive parameter in the AR-GARCH model has a non-standard limiting distribution with a non-standard rate of convergence when the innovations have non-finite fourth order moment.

Key words and phrases: ARMA-GARCH, heavy tails, tail behavior.

1. Introduction

The most important univariate model in econometrics is arguably the ARMA model. Countless studies have exploited the flexibility and simplicity of this model in many areas of economics. In addition during the 1980's the presence of non-constant volatility in macroeconomic and especially financial time series was recognized. The seminal papers of Engle (1982) and Bollerslev (1986) introduced the linear autoregressive conditional heteroscedastic (ARCH) and the generalized autoregressive conditional heteroscedastic (GARCH) models. The latter is by now so widely used that it is referred to as the "workhorse of the industry", Lee and Hansen (1994).

After the introduction of the GARCH model a number of papers studied its theoretical properties. It was established that, depending on the parameters of the model, processes generated by the GARCH model could exhibit vastly different behavior ranging from degenerating to zero, to having non-finite unconditional variance, or indeed be explosive, e.g., Nelson (1990). However, it was not until recently, Basrak, Davis, and Mikosch (2002b), the literature moved from "simply" stating whether a given moments was finite or not to instead providing a precise mathematical description of the tail behavior for processes generated by the GARCH model.

Combining the ARMA model with the GARCH model for the innovations, yielding the so-called ARMA-GARCH model, provides the econometrician with a flexible yet tractable model that allows one to model both mean and variance. However, estimating the full ARMA-GARCH model by quasi maximum likelihood is still somewhat more tedious than doing simple ordinary least squares (OLS). Hence if one is interested in estimating the AR parameters of an AR-GARCH model it can be attractive to use OLS estimation, which provides closed form expressions for the estimators. Generally the presence of GARCH-type innovations does not compromise the consistency of the OLS estimators (the precise requirements for this to hold are discussed later), but it does make standard inference invalid. The severity of the departure from standard inference depends critically on the tail behavior of the underlying GARCH process. When fitting GARCH models to typical log-return data, it is often found that the estimated parameters imply a finite second order moment, but a non-finite fourth order moment, see e.g., the discussion on integrated GARCH in Engle and Bollerslev (1986), Andersen and Bollerslev (1997), and Engle and Patton (2001).

In this paper we initially establish that ARMA processes based on GARCH innovations have the same tail behavior as the GARCH innovations themselves. Hence we extend the results of Basrak, Davis, and Mikosch (2002b) to another empirically important class of models. Second, we establish that the OLS estimator of the AR parameters in an AR-GARCH model remains consistent as long as the innovations have finite second order moment, but with a non-standard rate of convergence and a stable limiting distribution if the innovations have non-finite fourth order moment. The paper proceeds as follows. Section 2 contains the results. All proofs are deferred to the Appendix.

2. The Results

The AR(s)-GARCH$(1,1)$ model can be stated as

$$y_t = \rho_1 y_{t-1} + \cdots + \rho_s y_{t-s} + \varepsilon_t(\theta), \tag{2.1}$$

$$\varepsilon_t(\theta) = \sqrt{h_t(\theta)}z_t = \sqrt{\omega + \alpha\varepsilon_{t-1}^2(\theta) + \beta h_{t-1}(\theta)}z_t, \tag{2.2}$$

with $t = 1, \ldots, T$ and z_t an i.i.d.(0,1) sequence of random variables. The parameter vector is denoted by $\theta = (\rho_1, \ldots, \rho_s, \alpha, \beta, \omega)'$ and the true parameter by θ_0. Define in addition $\bar{y}_t = (y_t, \ldots, y_{t-s+1})'$ and $\rho = (\rho_1, \ldots, \rho_s)'$. In order to ease notation we adopt the convention $\varepsilon_t = \varepsilon_t(\theta_0)$, etc., for expressions evaluated at the true parameter values.

An important aspect of tail heaviness is summarized by the so-called tail index, denoted λ, for a further discussion see Resnick (1987). Under very general conditions the tail index of a GARCH(1,1) process (e.g., given by (2.2)) can be found as the unique strictly positive solution to the equation $E[(\beta+\alpha z_t^2)^{\lambda/2}] = 1$, as shown in Basrak, Davis, and Mikosch (2002b). A tail index of λ has the interpretation that the GARCH process has finite moments of all orders below λ, but $E[|\varepsilon_t|^\lambda] = \infty$. Techniques to estimate the tail index directly from a realization of a GARCH(1,1) process are discussed in Berkes, Horvath, and Kokoszka (2003).

Assumption 1.

(i) *z_t has a density with respect to the Lebesgue measure on $\mathbb{R}$ that is bounded away from zero and infinity on compact sets,*

(ii) $E[\log(\beta + \alpha z_t^2)] < 0$, *and*

(iii) *the maximal eigenvalue of the companion form matrix corresponding to the AR part of the model is smaller than one.*

Remark 1. By Theorem 1 in Meitz and Saikkonen (2008) and Meitz and Saikkonen (2006), Assumption 1 is sufficient for the process $x_t = (\bar{y}_{t-1}, \varepsilon_t, h_t)'$ generated by the AR-GARCH model to be geometrically ergodic. Hence Assumption 1 is also sufficient for the model to have a stationary distribution.

Theorem 1. *In addition to Assumption* 1 *let*

(i) *the initial values be distributed according to the stationary distribution,*

(ii) *the GARCH parameters be such that the GARCH process has finite second order moment, but non-finite fourth order moment,*

(iii) *and z_t's distribution be symmetric.*

Then all finite-dimensional vectors $(y_t, \dots, y_{t+k})$ have regularly varying tails as defined in Resnick (1987) *with the same tail index λ as the GARCH process.*

Remark 2. The theorem can easily be extended to ARMA(s, r)-GARCH(p, q) models at the price of a somewhat more cumbersome notation and less explicit conditions for stationarity, see e.g., Ling and Li (1998) and Basrak, Davis, and Mikosch (2002b).

Theorem 2. *Under the assumptions of Theorem* 1, $T^{1-2/\lambda}(\hat{\rho}_{OLS} - \rho_0) \xrightarrow{D} \Sigma^{-1} S_1$, *where S_1 is a $\lambda/2$ stable random vector on $\mathbb{R}^s$ and $\Sigma = E[\bar{y}_t \bar{y}_t'] > 0$.*

Remark 3. By symmetry of the distributions of both ε_t and $\bar{y}_{t-1}$ it can be concluded that the location and skewness parameters of the elements of S_1 are zero. However, at present we do not have an expression for the dispersion parameter or for the dependence structure within S_1.

Acknowledgement

The author gratefully acknowledges financial support from Danish Social Sciences Research Council, Project Number 2114-04-001.

Appendix

Lemma 1. *Assume $(\xi_t)_{t\in\mathbb{Z}}$ is a stationary sequence of regular varying symmetric random variables with tail index $\gamma \in]0,1[$. Then for any $K>0$ and $\phi\in]0,1[$, it holds that* $\lim_{m\to\infty}\limsup_{x\to\infty} P\big(K\sum_{i=m}^{\infty}\phi^i|\xi_{-i}|>x\big)P(|\xi_0|>x)^{-1}=0.$

Proof. In the following, assume without lose of generality that m is so large that $K\phi^m<1$. Rewrite the numerator of the fraction as

$$\begin{aligned}
&P\big(K\sum_{i=m}^{\infty}\phi^i|\xi_{-i}|>x\big)\\
&\le P\big(\bigcup_{i=m}^{\infty}\big(K\phi^i|\xi_{-i}|>x\big)\big)+P\big(K\sum_{i=m}^{\infty}\phi^i|\xi_{-i}|1_{\{K\phi^i|\xi_{-i}|\le x\}}>x, K\bigvee_{i=m}^{\infty}\phi^i|\xi_{-i}|\le x\big)\\
&\le \sum_{i=m}^{\infty}P(|\xi_{-i}|>x\phi^{-i}K^{-1})+P\big(K\sum_{i=m}^{\infty}\phi^i|\xi_{-i}|1_{\{K\phi^i|\xi_{-i}|\le x\}}>x\big).
\end{aligned}$$

Hence by Markov's inequality,

$$\begin{aligned}
&\frac{P\big(K\sum_{i=m}^{\infty}\phi^i|\xi_{-i}|>x\big)}{P(|\xi_0|>x)}\\
&\le\sum_{i=m}^{\infty}\frac{P(|\xi_0|>x\phi^{-i}K^{-1})}{P(|\xi_0|>x)}+x^{-1}K\sum_{i=m}^{\infty}\frac{\phi^iE[|\xi_0|1_{\{|\xi_0|\le x\phi^{-i}K^{-1}\}}]}{P(|\xi_0|>x)}=I_{m,x}+II_{m,x}.
\end{aligned}$$

As ξ_0 has regular varying tails, by Proposition 0.8(ii) of Resnick (1987) it holds for all $\tau>0$ that there exists a x_0 such that for all $x>x_0$, $P(|\xi_0|>x\phi^{-i}K^{-1})/P(|\xi_0|>x)\le(1+\tau)\phi^{i(\gamma-\tau)}K^{\gamma-\tau}$. For τ adequately small this bound is summable and hence by dominated convergence, one has

$$\lim_{m\to\infty}\limsup_{x\to\infty}I_{m,x}\le\lim_{m\to\infty}(1+\tau)\sum_{i=m}^{\infty}\phi^{i(\gamma-\tau)}K^{\gamma-\tau}=0.$$

Before considering $II_{m,x}$, note that from an integration by parts that

$$\frac{E[|\xi_0|1_{\{|\xi_0|\le x\}}]}{xP(|\xi_0|>x)}\le\frac{\int_0^x P(|\xi_0|>u)du}{xP(|\xi_0|>x)}$$

and, applying Karamata's Theorem (from e.g., Resnick (1987)), this converges to $(1-\gamma)^{-1}$ as x tends to infinity. Thus the function $x\mapsto E[|\xi_0|1_{\{|\xi_0|\le x\phi^{-i}K^{-1}\}}]$ is regular varying with tail index $1-\gamma$. Applying Proposition 0.8(ii), we have that for any $\tau>0$, some constant C_1, and x sufficiently large,

$$\begin{aligned}
K\frac{\phi^iE[|\xi_0|1_{\{|\xi_0|\le x\phi^{-i}K^{-1}\}}]}{xP(|\xi_0|>x)}&=K\phi^i\big(\frac{E[|\xi_0|1_{\{|\xi_0|\le x\phi^{-i}K^{-1}\}}]}{E[|\xi_0|1_{\{|\xi_0|\le x\}}]}\big)\frac{E[|\xi_0|1_{\{|\xi_0|\le x\}}]}{xP(|\xi_0|>x)}\\
&\le C_1K\phi^i(\phi^{-i}K^{-1})^{1-\gamma+\tau}=C_1K^{\gamma-\tau}\phi^{i(\gamma-\tau)},
\end{aligned}$$

which is summable for τ adequately small. Hence we conclude

$$\lim_{m\to\infty}\limsup_{x\to\infty} II_{m,x} \le \lim_{m\to\infty}\sum_{i=m}^{\infty} C_1 K^{\gamma-\tau}\phi^{i(\gamma-\tau)} = 0,$$

which establishes the lemma.

Proof of Theorem 1. We begin by showing a tamer result, namely that y_t is regularly varying with tail index λ. Since regular variation is a property of the marginal distribution, the subscript t on y_t is omitted. In addition due to symmetry of the distribution of y_t and ε_t, all arguments are given using the absolute values only.

Since the GARCH process has finite second order moment, y has the representation $y = \sum_{i=0}^{\infty}\tilde{\phi}_i\varepsilon_{-i}$ for a deterministic sequence $(\tilde{\phi})_{i=0}^{\infty}$. In addition there exists constants $C \le 0$ and $\phi \in]0,1[$ such that $|\tilde{\phi}_i| \le C\phi^i$ for all i. Define $A_m = \sum_{i=0}^{m-1}\tilde{\phi}_i\varepsilon_{-i}$ and $B_m = \sum_{i=m}^{\infty}\tilde{\phi}_i\varepsilon_{-i}$ for any $m \ge 1$, and note that $y = A_m + B_m$. Direct considerations (as in e.g., Lemma 3.6 of Jessen and Mikosch (2006)) establish that y is regularly varying with tail index λ if the following two conditions are met.

(C.1) $\lim_{m\to\infty}\limsup_{x\to\infty} P(|B_m| > x)/P(|\varepsilon_0| > x) = 0$.

(C.2) There exists a sequence of constants $(c_m)_{m\in\mathbb{N}}$ and $c_0 > 0$ such that $P(A_m > x) \sim c_m P(|\varepsilon_0| > x)$ as $x \to \infty$, and $c_m \to c_0$ as $m \to \infty$.

To establish $(C.1)$ note that for any $x > 0$ it holds that $P(|B_m| > x) \le P(\sum_{i=m}^{\infty} C\phi^i|\varepsilon_{-i}| > x)$. In the following we therefore show

$$\lim_{m\to\infty}\limsup_{x\to\infty}\frac{P\left(\sum_{i=m}^{\infty}\phi^i|\varepsilon_{-i}| > x\right)}{P(|\varepsilon_0| > x)} = 0. \tag{A.1}$$

By Basrak, Davis, and Mikosch (2002b) the random variable ε_0 has regular varying tails and we therefore wish to use Lemma 1 to establish (A.1). This lemma requires that the tail index of the sequence belongs to the unit interval, which is not the case for ε_0. Choose therefore $\eta > \lambda$ and set $K_m = \sum_{i=m}^{\infty}\phi^i$ and $p_i = \phi^i/K_m$; then by Jensen's inequality we get

$$\big(\sum_{i=m}^{\infty}\phi^i|\varepsilon_{-i}|\big)^{\eta} = K_m^{\eta}\big(\sum_{i=m}^{\infty}p_i|\varepsilon_{-i}|\big)^{\eta} \le K_m^{\eta}\sum_{i=m}^{\infty}p_i|\varepsilon_{-i}|^{\eta} \le C_2\sum_{i=m}^{\infty}\phi^i|\varepsilon_{-i}|^{\eta}$$

for some constant $C_2 > 0$. Thus

$$\frac{P\left(\sum_{i=m}^{\infty}\phi^i|\varepsilon_{-i}| > x\right)}{P(|\varepsilon_0| > x)} \le \frac{P\left(C_2\sum_{i=m}^{\infty}\phi^i|\varepsilon_{-i}|^{\eta} > x^{\eta}\right)}{P(|\varepsilon_0|^{\eta} > x^{\eta})},$$

and by Bingham, Goldie, and Teugels (1987), Proposition 1.5.7(i), the random variable $|\varepsilon_0|^\eta$ is regularly varying with tail index $\eta^{-1}\lambda \in]0,1[$. Hence by Lemma 1

$$\lim_{m\to\infty}\limsup_{x\to\infty}\frac{P\left(\sum_{i=m}^{\infty}\phi^i|\varepsilon_{-i}|>x\right)}{P(|\varepsilon_0|>x)} \le \lim_{m\to\infty}\limsup_{x\to\infty}\frac{P\left(C_2\sum_{i=m}^{\infty}\phi^i|\varepsilon_{-i}|^\eta>x^\eta\right)}{P(|\varepsilon_0|^\eta>x^\eta)} = 0,$$

which proves (A.1), and hence $(C.1)$.

For any fixed integer m, the process ε_t can be thought of as a GARCH(m,m) process with $\alpha_i=\beta_i=0$ for $i=2,\ldots,m$. Hence by Basrak, Davis, and Mikosch (2002b), Theorem 3.1 there only requires that not all the GARCH parameters are zero, the vector $(\varepsilon_0,\ldots,\varepsilon_m)'$ is regular varying with tail index λ. Since a linear combination of a regular varying vector is itself regular varying with the same tail index, it can be concluded that A_m is regular varying with tail index λ. Thus there exists constants c_m such that $\lim_{x\to\infty}P(A_m>x)/P(|\varepsilon_0|>x)=c_m$.

Next we prove that $(c_m)_{m=1}^{\infty}$ is a Cauchy sequence with a strictly positive limit c_0. Note first that there exists a constant $\bar{c}$ such that $c_m\le\bar{c}$ for all $m\ge 1$, as can be seen by replacing m with 1 in the arguments leading to (A.1). Assume without loss of generality that $c_n>c_m$. For any $\epsilon>0$, $\delta\in]0,1[$, and $n>m$, it holds that.

$$\begin{aligned} c_n-c_m &= \lim_{x\to\infty}\frac{P(A_m>(1-\delta)x)+P(|\sum_{i=m}^{n-1}\tilde{\phi}_i\varepsilon_{-i}|>\delta x)-P(A_m>x)}{P(|\varepsilon_0|>x)} \\ &\le c_m((1-\delta)^{-\lambda}-1)+\limsup_{x\to\infty}\frac{P(C\sum_{i=m}^{\infty}\phi^i|\varepsilon_{-i}|>\delta x)}{P(|\varepsilon_0|>x)} \\ &\le \bar{c}((1-\delta)^{-\lambda}-1)+f(m). \end{aligned}$$

Now by choosing δ so small that $\bar{c}((1-\delta)^{-\lambda}-1)<\epsilon/2$, and m so large that $f(m)<\epsilon/2$, it has been established that $(c_m)_{m=1}^{\infty}$ is a Cauchy sequence. By simple symmetry arguments, c_m is bounded away from zero and hence c_m must converge to a strictly positive limit c_0, which establishes $(C.2)$.

Finally we wish to extent the result to all vectors of the form $(y_1,\ldots,y_k)'$. By Basrak, Davis, and Mikosch (2002a), Theorem 1.1(ii), it suffices to show that all linear combinations are regular varying. However, for all $v\in\mathbb{R}^k\setminus\{0\}$, $v'(y_1,\ldots,y_k)'=\sum_{i=0}^{\infty}a_i\varepsilon_{-i}$, where the coefficients tend to zero at a geometric rate. Regular variation of the vector is therefore guaranteed by the same arguments as for y_t.

Proof of Theorem 2. Note initially that

$$T^{1-2/\lambda}(\hat{\rho}_{OLS}-\rho_0)=\big(T^{-1}\sum_{t=1}^{T}\bar{y}_{t-1}\bar{y}'_{t-1}\big)^{-1}\big(T^{-2/\lambda}\sum_{t=1}^{T}\bar{y}_{t-1}\varepsilon_t\big).$$

By Remark 1 it holds that $T^{-1}\sum_{t=1}^{T}\bar{y}_{t-1}\bar{y}_{t-1}' \stackrel{a.s.}{\to} \Sigma$, which establishes the first part of the claimed convergence. In order to establish the second part of the theorem take $\mathbf{y}_t(k) = (\varepsilon_t, y_{t-1}, \ldots, y_{t-k})'$ and let a_T be a sequence such that $TP(|y_t| > a_T) \to 1$ (due to the regular variation of y_t, one can choose a_T to be proportional to the $1 - 1/T$ quantile of the distribution function of $|y_t|$, e.g., by setting $a_T = T^{1/\lambda}$). The convergence of $T^{-2/\lambda}\sum_{t=1}^{T}\bar{y}_{t-1}\varepsilon_t$ follows from Davis and Mikosch (1998), Proposition 3.3, if we can establish the following.

(A.1) $\mathbf{y}_t(k)$ is regularly varying for all $k \geq 1$.

(A.2) The mild mixing condition $\mathcal{A}(a_T)$ from Davis and Mikosch (1998), p. 2052.

(A.3) Condition (2.10) of Davis and Mikosch (1998).

(A.4) Condition (3.3) of Davis and Mikosch (1998).

$(A.1)$ follows by a trivial extension of Theorem 1. Furthermore, Remark 1 establishes that the Markov chain $(\bar{y}_{t-1}, \varepsilon_t, h_t)'$ is geometrically ergodic, which implies that the stationary version is strongly mixing with geometrically decreasing rate function. Since the condition $\mathcal{A}(a_T)$ is implied by strong mixing, $(A.2)$ is satisfied.

The two remaining conditions require a bit more work. Recall that the y_t has the representation $y_t = \sum_{i=0}^{\infty}\tilde{\phi}_i\varepsilon_{t-i}$ and that there exists constants $C > 0$ and $\phi \in]\alpha+\beta, 1[$ such that $|\tilde{\phi}_i| \leq C\phi^i$ for all $i \in \mathbb{N}_0$. For later use define the auxiliary process $\breve{y}_t = C\sum_{i=0}^{\infty}\phi^i|\varepsilon_{t-i}|$, which is clearly positive. Inspecting the proof of Theorem 1 reveals that $\breve{y}_t$ is regularly varying with tail index λ. In addition one has the relations $\breve{y}_t \geq |y_t|$ and $(\breve{y}_t^2 + 1)C_0 \geq h_t$ for some constant C_0 and for all t.

With $|\cdot|$ denoting the max norm, Davis and Mikosch (1998), condition (2.10), has

$$\lim_{m\to\infty}\limsup_{T\to\infty} P\Big(\bigvee_{m\leq|t|\leq r_T}|\mathbf{y}_t(k)| > a_T x \,\big|\, |\mathbf{y}_0(k)| > a_T x\Big) = 0, \;\; x > 0, \tag{A.2}$$

where r_T is an integer sequence such that $r_T \to \infty$ and $r_T/T \to 0$ as $T \to \infty$. By the definition of conditional probabilities, Markov's inequality, and the symmetry of the distributions, it holds for $t > 0$ that

$$\begin{aligned} P(|y_t| > a_T x \mid |y_0| > a_T x) &\leq \frac{E[1_{\{|y_0|^2 > a_T^2 x^2\}} y_t^2]}{a_T^2 x^2 P(|y_0|^2 > a_T^2 x^2)} \\ &= \frac{E[1_{\{|y_0|^2 > a_T^2 x^2\}}(\sum_{i=0}^{t-1}\tilde{\phi}_i^2\varepsilon_{t-i}^2 + (\sum_{i=t}^{\infty}\tilde{\phi}_i\varepsilon_{t-i})^2)]}{a_T^2 x^2 P(|y_0|^2 > a_T^2 x^2)} \\ &\leq C_0^2\frac{E[1_{\{|y_0|^2 > a_T^2 x^2\}}(\sum_{i=0}^{t-1}\phi^{2i}\varepsilon_{t-i}^2 + \phi^{2t}y_0^2)]}{a_T^2 x^2 P(|y_0|^2 > a_T^2 x^2)} = I_{t,T}. \end{aligned}$$

In order to bound $I_{t,T}$, note that for $t > 0$ the recursion of Nelson (1990) gives

$$E[\varepsilon_t^2 \mid h_0] = \sum_{i=0}^{t-1} \omega(\alpha+\beta)^i + (\alpha+\beta)^t h_0 \le C_1 + \phi^t h_0$$

for some positive constant C_1 independent of t. Direct calculations provide the relation $\sum_{i=0}^{t-1} \phi^{2i}\phi^{t-i} \le \phi^t/(1-\phi)$, which converges to zero as t tends to infinity. In addition it holds that

$$\frac{E[1_{\{|y_0|>a_T x\}}\breve{y}_0^2]}{a_T^2 x^2 P(|y_0| > a_T x)} \le \frac{E[1_{\{|\breve{y}_0|^2>a_T^2 x^2\}}\breve{y}_0^2]}{a_T^2 x^2 P(|\breve{y}_0^2| > a_T^2 x^2)} \, \frac{P(|\breve{y}_0| > a_T x)}{P(|y_0| > a_T x)},$$

and hence by Karamata's Theorem,

$$\limsup_{T\to\infty} \frac{E[1_{\{|y_0|>a_T x\}}\breve{y}_0^2]}{a_T^2 x^2 P(|y_0| > a_T x)} \le \frac{C_2}{\lambda - 2}$$

for some constant C_2. Applying Karamata's Theorem again it can be concluded that there exists T_0 such that, for all $T > T_0$,

$$\begin{aligned} I_{t,T} &\le C_0 \frac{E\big[1_{\{y_0^2 > a_T^2 x^2\}}\big(\phi^{2t} y_0^2 + \breve{y}_0^2 \phi^t/(1-\phi) + C_1\big)\big]}{a_T^2 x^2 P(y_0^2 > a_T^2 x^2)} \\ &\le C_3\phi^{2t} + C_3\phi^t + \frac{C_3}{a_T^2 x^2} \le 2C_3\phi^t + \frac{C_3}{a_T^2 x^2} \end{aligned}$$

for some positive constant C_3 independent of t. We are now ready to verify (A.2).

$$\begin{aligned} &\lim_{m\to\infty} \limsup_{T\to\infty} P\Big(\bigvee_{m\le|t|\le r_T} |\mathbf{y}_t(k)| > a_T x \Big| |\mathbf{y}_0(0)| > a_T x\Big) \qquad \text{(A.3)} \\ &\quad\le \lim_{m\to\infty} \limsup_{T\to\infty} 2(k+1) \sum_{t=m}^{r_T+k} P(|y_t| > a_T x \mid |y_0| > a_T x) \frac{P(|y_0| > a_T x)}{P(|\mathbf{y}_0(k)| > a_T x)} \\ &\quad\quad + \lim_{m\to\infty} \limsup_{T\to\infty} 2(k+1) \sum_{t=m}^{r_T+k} P(|y_t| > a_T x \mid |\varepsilon_0| > a_T x) \frac{P(|\varepsilon_0| > a_T x)}{P(|\mathbf{y}_0(k)| > a_T x)} \\ &\quad\quad + \lim_{m\to\infty} \limsup_{T\to\infty} 2(k+1) \sum_{t=m}^{r_T+k} P(|\varepsilon_t| > a_T x \mid |y_0| > a_T x) \frac{P(|y_0| > a_T x)}{P(|\mathbf{y}_0(k)| > a_T x)} \\ &\quad\quad + \lim_{m\to\infty} \limsup_{T\to\infty} 2(k+1) \sum_{t=m}^{r_T+k} P(|\varepsilon_t| > a_T x \mid |\varepsilon_0| > a_T x) \frac{P(|\varepsilon_0| > a_T x)}{P(|\mathbf{y}_0(k)| > a_T x)}. \end{aligned}$$

For the first term on the right hand side, the preceding arguments show that

$$\lim_{m\to\infty}\limsup_{T\to\infty} 2(k+1)\sum_{t=m}^{r_T+k} P(|y_t|>a_Tx \mid |y_0|>a_Tx)\underbrace{\frac{P(|y_0|>a_Tx)}{P(|\mathbf{y}_0(k)|>a_Tx)}}_{\le 1}$$

$$\le \lim_{m\to\infty} 4(k+1)\sum_{t=m}^{\infty} C_3\phi^t + \lim_{m\to\infty}\limsup_{T\to\infty} 2(k+1)(r_T+k)/(a_T^2x^2) = 0,$$

by choosing r_T such that $r_T/a_T^2 \to 0$. Note that negative values of t are dealt with since, by stationarity, for $t>0$,

$$\begin{aligned} P(|y_{-t}|>a_Tx \mid |y_0|>a_Tx) &= \frac{P(|y_{-t}|>a_Tx, |y_0|>a_Tx)}{P(|y_{-t}|>a_Tx)} \\ &= P(|y_t|>a_Tx \mid |y_0|>a_Tx). \end{aligned}$$

The other terms of (A.3) are zero by identical arguments. Using the geometric ergodicity of $(\varepsilon_t, \bar{y}_{t-1})$, it is easily shown that the extremal index γ, which appears in (2.10) of Davis and Mikosch (1998) and in Mikosch and Straumann (2006), is strictly positive as required. This completes the verification of $(A.3)$.

Finally $(A.4)$ is considered. In the setup of the AR-GARCH model condition (3.3) of Davis and Mikosch (1998) is

$$\lim_{x\to 0}\limsup_{T\to\infty} P\Big(\Big|\sum_{t=1}^{T}\frac{\varepsilon_t y_{t-h}}{a_T^2}1_{\{|\varepsilon_t y_{t-h}|\le a_T^2 x\}} - \underbrace{E\Big[\sum_{t=1}^{T}\frac{\varepsilon_t y_{t-h}}{a_T^2}1_{\{|\varepsilon_t y_{t-h}|\le a_T^2 x\}}\Big]}_{=0}\Big| > \delta\Big) = 0,$$

for all $\delta>0$ and $h=1,\ldots,s$. Markov's inequality and Kamarata's Theorem (the required regular variation of $\varepsilon_t y_{t-h}$ can be verified by the same arguments as for y_t) now give, for any $h=1,\ldots,s$, that

$$\begin{aligned} P(|a_T^{-2}\sum_{t=1}^{T}\varepsilon_t y_{t-h}1_{\{|\varepsilon_t y_{t-h}|\le a_T^2x\}}|>\delta) &\le \frac{1}{\delta^2}a_T^{-4}\sum_{t=1}^{T}E[\varepsilon_t^2 y_{t-h}^2 1_{\{|\varepsilon_t y_{t-h}|^2\le a_T^4x^2\}}] \\ &= \frac{1}{\delta^2}a_T^{-4}TE[\varepsilon_h^2y_0^2 1_{\{|\varepsilon_h y_0|^2\le a_T^4x^2\}}] \\ &\sim C_4x^2TP(|\varepsilon_h y_0|^2 > a_T^4x^2) \ \text{ for large } T \\ &\overset{T\to\infty}{\to} C_5x^2 \overset{x\to 0}{\to} 0. \end{aligned}$$

Hence $(A.4)$ holds. Due to $(A.1)$ - $(A.4)$, Proposition 3.3 of Davis and Mikosch (1998) is applicable and $a_T^{-2}\sum_{t=1}^{T}\bar{y}_{t-1}\varepsilon_t \overset{D}{\to} S_1$, where S_1 is a $\lambda/2$-stable random vector in $\mathbb{R}^s$.

References

Andersen, T. G. and Bollerslev, T. (1997). Intraday periodicity and volatility persistence in financial markets. *J. Empirical Finance* **4**, 115-158.

Basrak, B., Davis, R. and Mikosch, T. (2002a). A characterization of multivariate variation. *Ann. Appl. Probab.* **12**, 908-920.

Basrak, B., Davis, R. and Mikosch, T. (2002b). Regular variation of GARCH processes. *Stochastic Process. Appl.* **99**, 95-115.

Berkes, I., Horvath, L. and Kokoszka, P. (2003). Estimation of the maximal moment exponent of a GARCH(1,1) sequence. *Econometric Theory* **19**, 565-586.

Bingham, N., Goldie, C. and Teugels, J. (1987). *Regular Variation*, Vol. 27 of *Encyclopedia of Mathematics and its Applications*, Cambridge University Press, Cambridge, UK.

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

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

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

Engle, R. and Bollerslev, T. (1986). Modelling the persistence of conditional variance. *Econometric Rev.* **5**, 1-50.

Engle, R. and Patton, A. (2001). What good is a volatility model?. *Quantitative Finance* **1**, 237-245.

Jessen, A. and Mikosch, T. (2006). Regularly varying functions. Technical report, Laboratory of Actuarial Mathematics, University of Copenhagen.

Lee, S. and Hansen, B. (1994). Asymptotic theory for the GARCH(1,1) quasi-maximum likelihood estimator. *Econometric Theory* **10**, 29-53.

Ling, S. and Li, W. K. (1998). Limiting distributions of maximum likelihood estimators for unstable ARMA models with GARCH errors. *Ann. Statist.* **26**, 84-125.

Meitz, M. and Saikkonen, P. (2006). Stability of nonlinear AR-GARCH models. Technical report, SSE/EFI Working Paper Series in Economics and Finance No. 632.

Meitz, M. and Saikkonen, P. (2008). Stability of nonlinear AR-GARCH models. *J. Time Ser. Anal.* **29**, 453-475.

Mikosch, T. and Straumann, D. (2006). Stable limits of martingale transforms with application to the estimation of GARCH parameters. *Ann. Statist.* **34**, 493-522.

Nelson, D. B. (1990). Stationarity and persistence in the GARCH(1,1) model. *Econometric Theory* **6**, 318-334.

Resnick, S. (1987). *Extreme Values, Regular Variation, and Point Processes.* Applied Probability, Springer.

Department of Biostatistics, University of Copenhagen, Oester Farimagsgade 5, DK-1014 Copenhagen K, Denmark.

E-mail: t.lange@biostat.ku.dk

(Received March 2009; accepted February 2010)