Limiting behavior of the heteroskedastic robust Wald test when the innovations have heavy tails

By Theis Lange

Department of Economics, University of Copenhagen

E-mail: *theis.lange@econ.ku.dk*

Abstract: This paper initially establishes that the usual OLS estimator of the autoregressive parameter in the first order stable autoregressive model with autoregressive conditional heteroskedastic errors, the AR-ARCH model, has a non-standard limiting distribution with a non-standard rate of convergence when the innovations have non-finite fourth order moment. Furthermore, it is shown that the robust t- and Wald test statistics of White (1980) are still consistent and have the usual rate of convergence, but a non-standard limiting distribution when the innovations have non-finite fourth order moment. The critical values for the non-standard limiting distribution are higher than the usual N(0,1) and χ_1^2 critical values, respectively, which implies that an acceptance of a hypothesis using the standard robust t- or Wald tests remains valid even if the fourth order moment condition is not met. However, the size of the test might be higher than the nominal size. Hence the analysis presented in this paper extends the usability of the robust t- and Wald tests of White (1980). Finally, a small empirical study illustrates the results.

Keywords: ARCH; Robust t- and Wald tests; Heavy tails.

1 Introduction

Given a process $(y_t)_{t=1}^T$ this paper studies the OLS estimator from the regression of y_t on y_{t-1} when the process is assumed to be generated by a stable autoregressive model with autoregressive conditional heteroskedastic errors, the AR-ARCH model. By now the presence of ARCH type effects in financial and macro economic time series is a well established fact. The seminal paper by Engle (1982) in which the linear ARCH model model was originally introduced has been followed by countless papers studying various aspects ARCH type models.

Recognizing that unmodeled heteroscedasticity in the innovations might seriously compromise the validity of traditional t- and Wald tests of the significance of parameters estimated by OLS White (1980) introduced the heteroscedastic robust t- and Wald tests. These tests are now so widely applied that they are routinely reported by many statistical software packages. However, Whites results depend on the innovations to have finite fourth order moment, which is often not met in empirical studies.

In this paper we show that the robust t- and Wald test statistics have the correct normalization, but a non-standard limiting distribution when the innovations have non-finite fourth order moment. The critical values for the non-standard limiting distributions are higher than the usual N(0,1) and χ_1^2 critical values, respectively, which implies that an acceptance of a hypothesis using the standard robust t- or Wald test procedures remains valid even if the fourth order moment condition is not met. However, the size of the test might be higher than the nominal size. Hence the analysis presented in this paper extends the usability of the robust t- and Wald tests of White (1980), which, to our knowledge, has not previously been done in the literature. In addition the paper establishes that the OLS estimator of the autoregressive parameter will have a stable limit with a non-standard rate of convergence. As the tools for handling stable distributions are less evolved than similar tools for normal distributions we are forced to restrict attention to a fairly simple first order model as the true data generating mechanism. Finally, a small empirical study shows how the evidence of no correlation between consecutive movements of interest rates critically depends on our extension of the usability of the robust Wald test.

The paper proceeds as follows. In Section 2 the model and some important properties including geometric ergodicity and tail heaviness are discussed. Section 3 presents the limiting distributions for the OLS estimator and Section 4 states the limiting distributions for the robust t- and Wald test statistics and discusses implications of the results on the standard testing procedures. Finally, Section 5 contains a small empirical study and Section 6 concludes. All proofs are contained in the Appendix.

2 The AR-ARCH Model

The model can be stated as

$$
\begin{aligned}
y_t &= \rho y_{t-1} + \varepsilon_t(\theta), && (1)\\
\varepsilon_t(\theta) &= \sqrt{h_t(\theta)} z_t && (2)\\
h_t(\theta) &= \omega + \alpha \varepsilon_{t-1}^2(\theta) && (3)
\end{aligned}
$$

with $t = 1, ..., T$ and z_t an i.i.d.(0,1) sequence of random variables. The parameter vector is denoted $\theta = (\rho, \alpha, \omega)'$ and the true parameter θ_0. In order to ease notation we adopt the convention $\varepsilon_t = \varepsilon_t(\theta_0)$ etc. for expressions evaluated at the true parameter values. The analysis is conditional on the initial values y_0 and ε_{-1}.

In the context of the AR-ARCH model heavy tails can be introduced either by choosing the value of the ARCH parameter α sufficiently large while keeping the underlying error process z_t light tailed or through the tails of the underlying error process z_t. In this paper the first approach will be explored. The second approach has been investigated in e.g. Davis & Mikosch (1998).

For a fixed value of the ARCH parameter α the tail index, denoted λ, can be found as the unique strictly positive solution to the equation $E[(\alpha z_t^2)^{\lambda/2}] = 1$ as shown in Davis & Mikosch (1998) p. 2062. Note that a tail index of λ has the implication that the ARCH process has finite moments of all orders below λ, but $E[|\varepsilon_t|^\lambda] = \infty$. Figure 1 depicts the correspondence between α and λ when z_t is assumed Gaussian. Using the moment interpretation of the tail index and Figure 1 it is evident that the ARCH process has finite fourth order moment, but non-finite second order moment if the ARCH parameter belongs to the interval]0.57, 1]. This part of the parameter space will be the focus for much of the rest of the paper.

The following lemma, which has been proved in Lange, Rahbek & Jensen (2007), establishes minimal conditions under which processes generated by the AR-ARCH model are geometrically ergodic. Geometric ergodicity, and the laws of large numbers implied by this concept, constitutes an important tool when establish-

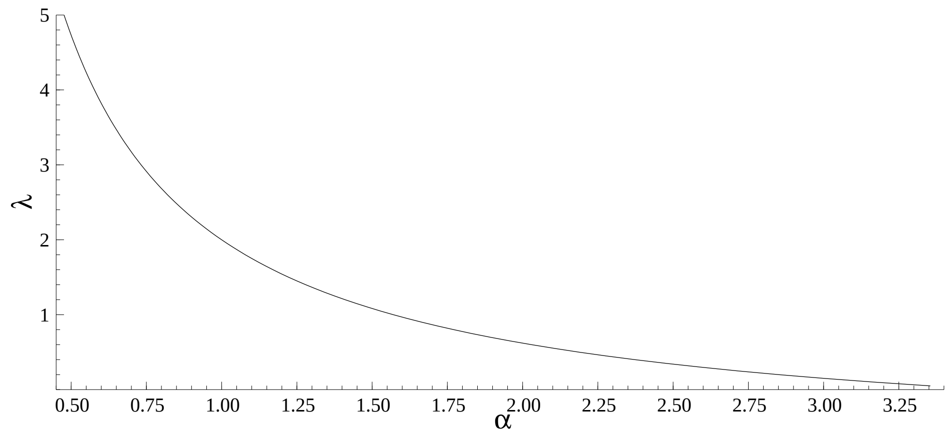

Figure 1: Correspondence between the ARCH parameter α and the tail index λ when z_t is assumed Gaussian.

ing asymptotic theory, but for our intended applications it is of equal importance that the lemma establishes minimal conditions under which there exists an initial distribution such that the process is stationary.

Lemma 1. *Assume that z_t has a density f with respect to the Lebesgue measure on $\mathbb{R}$, which is bounded away from zero on compact sets and furthermore that*

$$E[\log(\alpha_0 z_t^2)] < 0 \text{ and } |\rho_0| < 1$$

then the process $x_t = (y_{t-1}, \varepsilon_t)'$ generated by the AR-ARCH model, is geometrically ergodic. In particular there exists a stationary version and moreover if $E|g(x_t, ..., x_{t+k})| < \infty$, where expectation is taken with respect to the invariant distribution, the Law of Large Numbers given by

$$\lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} g(x_t, ..., x_{t+k}) \overset{a.s.}{=} E[g(x_t, ..., x_{t+k})], \tag{4}$$

holds irrespectively of the choice of initial distribution.

3 Limiting behavior of the OLS estimator

In this section the limiting behavior of the OLS estimator of the autoregressive parameter ρ is derived under two sets of moment assumptions on the innovation sequence. Initially the section reviews the well known result that the OLS estimator is asymptotically normal when the innovations have finite fourth order moment and states precise expressions for the parameters of the limiting distribution. In the second part of this section the limiting behavior of the OLS estimator when the innovations only have finite second order moment, is derived. To our knowledge this result has not previously been established in the literature. The final part of this section discusses some implications of the limiting results.

3.1 The normal case: Finite fourth order moment

Standard techniques combined with Lemma 1 give the following result regarding the OLS estimator of ρ when the innovations have finite fourth order moment.

Theorem 1. *In addition to the assumptions in Lemma 1 assume that*

(i) $E[\alpha_0^2 z_t^4] = \alpha_0^2 \kappa < 1$

(ii) *and z_t's distribution is symmetric.*

Then the OLS estimator of the autoregressive parameter ρ in the AR-ARCH model given by (1) - (3) is consistent with the following limiting distribution

$$\sqrt{T}(\hat{\rho}_{OLS} - \rho_0) = \frac{T^{-1/2}\sum_{t=1}^{T} y_{t-1}\varepsilon_t}{T^{-1}\sum_{t=1}^{T} y_{t-1}^2} \xrightarrow{D} N(0, \Sigma), \tag{5}$$

where

$$\Sigma = (1-\rho_0^2) + \frac{(\kappa-1)(1-\rho_0^2)^2\alpha_0}{(1-\kappa\alpha_0^2)(1-\alpha_0\rho_0^2)}.$$

Remark 1. Condition (i) implies that ε_t has finite fourth order moment under the invariant measure and that z_t has finite fourth order moment.

The proof can be found in the appendix. As the parameter values approaches the value for which the ARCH process no longer have finite fourth order moment

($\alpha_0^2\kappa = 1$) the asymptotic variance Σ converges toward infinity. One could therefore conjecture that in this case the limiting distribution would be a stable law with a slower than square root T rate of convergence. In the following section we will prove this conjecture.

3.2 The stable case: Non-finite fourth order moment

Next, we will analyze the effect of relaxing the fourth order moment condition of Theorem 1 to a second order condition. As conjectured this leads to both a non-standard limiting distribution as well as non-standard rate of convergence. Since the tools for manipulating stable laws are somewhat less evolved than similar tools for normal distributions it is necessary to assume stationarity of the process as geometric ergodicity does not suffice in the present version of the proof, which can be found in the appendix.

Theorem 2. *In addition to the assumptions in Lemma 1 assume that*

(i) *the initial values are distributed according to the stationary distribution,*
(ii) *the ARCH parameter α_0 is such that the ARCH process has finite second order moment, but non-finite fourth order moment; that is the tail index λ belongs to the interval* $]2, 4]$,
(iii) *and z_t's distribution is symmetric*

then it holds that

$$T^{1-2/\lambda}(\hat{\rho}_{OLS} - \rho_0) \xrightarrow{D} S_0, \tag{6}$$

where S_0 is a $\lambda/2$ stable random variable, with the remaining parameters unknown.

Remark 2. The existence of a stationary distribution is guaranteed by Lemma 1.

3.3 Implications

As Theorem 1 is a standard result this section will only address the implications of Theorem 2, which has the most direct implications for the construction of confidence bands. When constructing confidence bands one needs to know both the asymptotical distribution (usual the normal distribution) including parameter estimates as well as the rate of convergence (usual square root T). However, if the true α_0 is such that the innovation sequence ε_t does not have finite fourth order moment the result implies that the rate of convergence will be non-standard and unknown and the parameters of the limiting distribution unknown as well. Since the rate of convergence in Theorem 2 is slower than the usual square root T the usually constructed confidence bands, based on normality and standard rate of convergence, might be way to narrow leading to erroneous conclusions. Unfortunately, since λ in unknown in practice and Theorem 2 does not include a precise specification of the remaining parameters of the limiting stable distribution, one cannot easily derive a corrected confidence band based on the theorem. In the next section we will address this problem by considering the heteroskedastic robust$t$- and Wald tests of White (1980).

4 Limiting behavior of the robust Wald test

In this section we will examine the behavior of the heteroskedastic robust t- and Wald tests of White (1980). The robust t-test statistic for the hypothesis $H_0 : \rho_0 = 0$ is given by

$$V_T = \sqrt{T}(\hat{\rho}_{OLS} - \rho_0)\big(\frac{1}{T}\sum_{t=1}^{T}\varepsilon_t^2 y_{t-1}^2\big)^{-1/2}\big(\frac{1}{T}\sum_{t=1}^{T}y_{t-1}^2\big),$$

and the robust Wald test statistic is V_T^2. Under the hypothesis V_T can be rewritten as

$$V_T = \big(\sum_{t=1}^{T}\varepsilon_t\varepsilon_{t-1}\big)\big(\sum_{t=1}^{T}\varepsilon_t^2\varepsilon_{t-1}^2\big)^{-1/2},$$

which is sometimes refereed to as a *self normalizing sum.* If the innovations ε_t have finite fourth order moment it is well known, see e.g. White (1980), that V_T converges to standard normal distribution under H_0 as T tends to infinity irrespectively of possible heteroscedasticity of the innovations. This result forms the basis for the usual robust t- and Wald test. However, if the the innovations do not have finite fourth order moment the limiting behavior of V_T has not been examined in the litterateur. From the theory for self normalizing sums of i.i.d. random variables, see Giné, Götze & Mason (1997), it is known that a self normalized sum of i.i.d. random variables converges to a normal distribution if and only if the numerator belongs to the domain of attraction for a normal distribution (loosely speaking this corresponds to requiring that the Lindeberg condition holds for the numerator), however, since the sequence $(\varepsilon_t)_{t=1}^T$ is not i.i.d. the results of Giné et al. (1997) are not applicable in our setup. Based on this result one would still conjecture that V_T does not have a Gaussian limiting distribution since Theorem 2 establishes that the numerator belongs to the domain of attraction of a stable law. Theorem 3 below formalizes this conjecture and the proof can be found in the appendix.

Theorem 3. *In addition to the assumptions in Lemma 1 assume that*

- *(i)* *the initial values are distributed according to the stationary distribution,*
- *(ii)* *the ARCH parameter α_0 is such that the ARCH process has finite second order moment, but non-finite fourth order moment; that is the tail index λ belongs to the interval $]2,4]$,*
- *(iii)* *and the distribution of z_t is symmetric*

then under H_0 it holds that

$$V_T \xrightarrow{D} \frac{S_1}{S_2^{1/2}} \tag{7}$$

where the vector $(S_1, S_2^{1/2})$ is jointly a $\lambda/2$ stable random variable, with the remaining parameters and dependence structure unknown.

Remark 3. A natural question is what happens when $\lambda \in]0,2]$, corresponding to $E[\varepsilon_t^2] = \infty$, but in this case Theorem 2 indicates that $\hat{\rho}_{OLS}$ is not even consistent rendering a test for a particular value fruitless.

Corollary 1. *Under the conditions of Theorem 3 the heteroskedastic robust Wald test given by V_T^2 converges in distribution to S_1^2/S_2 where the vector (S_1^2, S_2) is jointly a $\lambda/4$ stable random variable, with the remaining parameters and dependence structure unknown.*

In the following section the implications of Theorem 3 on the usual testing procedure will be discussed.

4.1 Implications for the standard testing procedure

Arguably the most important implication of Theorem 3 and Corollary 1 is that the normalization required to ensure a non-degenerate limiting distribution for both the robust t- and Wald tests does not depend on the fourth order moment being finite. However, since the limiting distribution is no longer Gaussian when the innovations have non-finite fourth order moment the critical values are not the usual ones.

In the usual Gaussian case, obtained when the innovations have finite fourth order moment, it can be directly established by utilizing the properties of the normal distribution that the limiting distribution for the robust t- and Wald test statistics are nuisance parameter free (indeed even in higher order models than the one considered in this paper). In contrast to this, it is not possible to verify that the stable limits in Theorem 3 and Corollary 1 do not depend on additional parameters besides the tail index, since a precise mathematical expression for parameter values and dependence structure is not available. However, in the first order AR-ARCH model considered in this paper the only remaining unknown parameter is the scale parameter ω_0 and the test statistic V_T is clearly invariant to the scale of the innovations.

Since the scale parameter ω_0 does not affect the limiting distributions in Theorem 3 and Corollary 1 the critical values for the hypothesis H_0 will only depend on the ARCH parameter α_0 and perhaps the exact distribution of the innovations z_t. Based on simulations Figure 2 Panel A-C illustrate how the critical values change as the tail index λ is decreased. Panel D shows the correspondence between the tail parameter and the ARCH parameter for the different choices of the distribution for z_t.

From Figure 2 it is evident that the critical values for the non-standard limiting distribution are higher than usual χ_1^2 critical values and increase as the tail index decreases (corresponding to increasing the ARCH parameter). In addition it is seen that the distribution of the underlying innovations z_t only affects the limiting distribution through the tail index. This implies that an acceptance of a hypothesis using the standard robust Wald test procedure remains valid even if the fourth order moment condition is not met. However, the size of the test

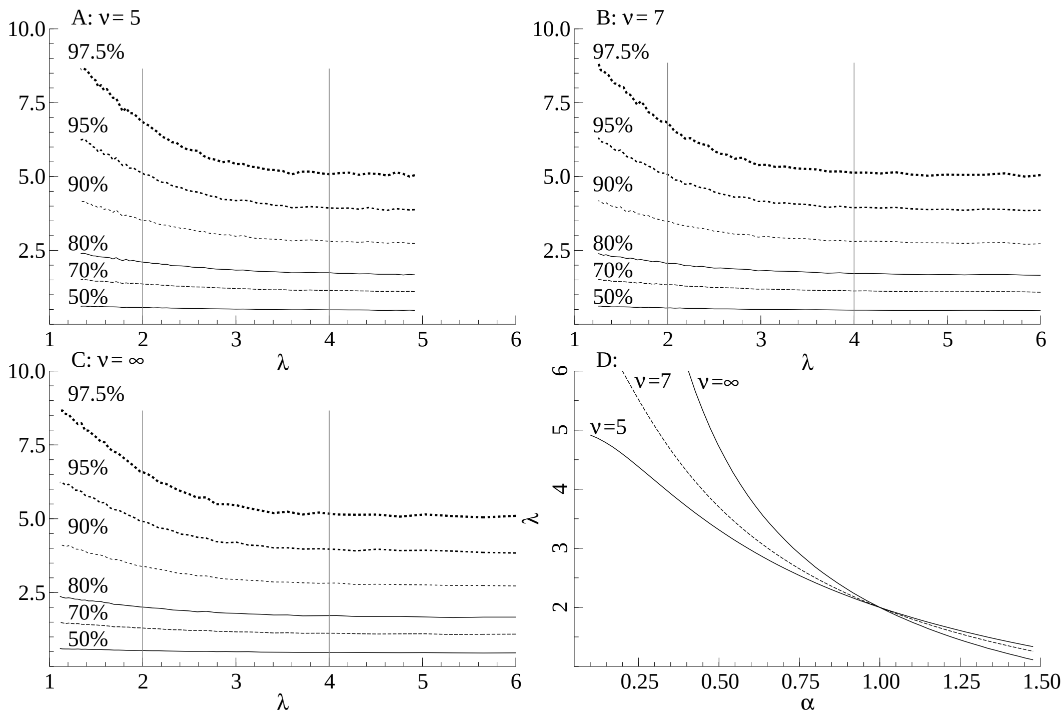

Figure 2: Quantiles for the limiting distribution for the robust Wald test statistic (S_1^2/S_2 from Corollary 1) computed by simulating from the AR-ARCH model with $\rho_0 = 0$, $\omega_0 = 1$ for a range of values for α_0 (to ease comparison the critical values are reported as functions of the tail index). In Panel A and B the innovations z_t follow a standardized student-t distribution with 5 and 7 degrees of freedom, respectively, while $z_t \sim N(0,1)$ in Panel C. Each simulated path was 5,000 data points long and 100,000 Monte Carlo replications were conducted for each value of α_0. The two vertical lines corresponds to the values of the tail index where the process does no longer have finite fourth order and second order moment, respectively. Finally, Panel D shows the correspondence between the ARCH parameter and the tail index for the different distributions.

might be higher than the nominal size. Furthermore, Theorem 2 implies that the robust $t-$ and Wald tests are still consistent as long as the innovations have finite second order moment. Hence the analysis of this paper extends the usability of the robust $t-$ and Wald tests of White (1980).

5 Empirical illustration

In this section we will reexamine the evidence of linear predictability in the daily movements of interest rates. The data set consists of daily recordings of the 3-months US t-bill rate (r_t) covering the period from the 2^{nd} of January 1990 to the 29^{th} of February 2008, yielding a total of 4,544 observations, see Figure 3.

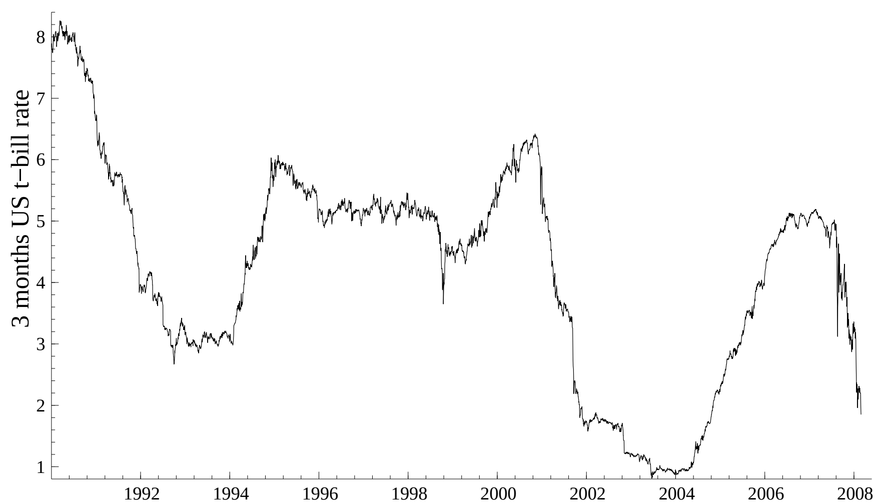

Figure 3: Daily recordings of the 3-months US t-bill rate.

To examine whether past interest rate movements are correlated with future interest rate movements we will test if the coefficient in the regression of $x_t = r_t - r_{t-1}$ on x_{t-1} is statistically significantly different from zero. As it is well documented that daily interest rates exhibit heteroscedasticity we will conduct both the usual Wald test as well as the robust Wald test of White (1980). Finally we will employ the full AR-ARCH model to estimate the magnitude of the ARCH effect by quasi maximum likelihood and thereby assess the potential size distortion of the robust Wald test caused by heavy tails.

	Wald test	Robust Wald test
Test statistic	30.27	3.44
p-value	0.000	0.065
Corrected p-value[†]	-	0.126

Table 1: Summary of test results for the hypothesis of a zero coefficient in the regression of x_t on x_{t-1}. The unrestricted OLS estimator is 0.081. [†]Computed using the estimated ARCH coefficient of 0.91 and Student-t degree of freedom of 2.45 from the full AR-ARCH model and the non-standard distribution from Corollary 1.

Based on the the test statistics and p-values presented in Table 1 it is evident that if the heteroscedasticity of the errors is ignored one would reject the hypothesis that the coefficient in the regression is zero. If the test is instead based on the robust Wald test statistic compared to the χ_1^2 distribution the hypothesis is accepted with a p-value of 0.065. Estimating the full AR-ARCH model with standardized Student-t innovations with ν degrees of freedom by quasi maximum likelihood provides the estimates $\hat{\alpha} = 0.91$ and $\hat{\nu} = 2.45$, corresponding to a tail index of 2.2. Hence the conditions for employing the robust Wald test of White (1980) are not met. However, by employing the non-standard limiting distribution from Corollary 1 the p-value increases to 0.126. Thus taking the magnitude of the ARCH effect into account reveals that the robust Wald test is somewhat size distorted, but as previously discussed this only strengthens the conclusion of no linear predictability.

It should be stressed that we do not suggest that practitioners do full quasi maximum likelihood estimation just to correct their robust Wald test inference. The purpose of this section is merely to illustrate the necessity of the extended usability of the robust tests and quantify the potential size distortion.

6 Conclusion

In this paper we have established that the usual OLS estimator of the autoregressive parameter in the AR-ARCH model has a non-standard limiting distribution with a non-standard rate of convergence if the innovation process is a realization of an ARCH(1) process with non-finite fourth order moment. Furthermore, we have established that the robust t- and Wald test statistics of White (1980) for the hypothesis $\rho_0 = 0$ have the correct normalization, but a non-standard limiting distribution when the innovations have non-finite fourth order moment. The critical values for the non-standard limiting distribution are higher than the usual N(0,1) and χ_1^2 critical values, respectively, which implies that an acceptance of the hypothesis using the standard robust t- or Wald tests remains valid even in the fourth order moment condition is not met. However, the size of the tests might be higher than the nominal size. Hence the analysis presented in this paper extends the usability of the robust t- and Wald tests of White (1980). In Figure 2 the critical values are summarized. Finally, a small empirical study shows how the extended usability of the robust test is required to establish that consecutive movements of interest rates are not correlated. In addition the empirical study quantifies the potential size distortion caused by the heavy tails of the innovations.

Appendix

Proof of Theorem 1. Note initially that since we have assumed finite fourth order moment and hence also finite second order moment, y_t has the stationary representation $y_t^* = \sum_{t=0}^{\infty} \rho^i \varepsilon_{t-i}$, which will be used when calculating expected values under the stationary distribution. The second order moments of the ARCH process and the volatility process are given by $E[\varepsilon_t^2] = E[h_t] = \frac{\omega_0}{1-\alpha_0}$. Next the fourth order moment can be derived from

$$E[\varepsilon_t^4] = E[z_t^4 h_t^2] = \kappa(\omega_0^2 + \alpha_0^2 E[\varepsilon_{t-1}^4] + 2\omega_0\alpha_0 E[\varepsilon_{t-1}^2]).$$

Since the expectation is taken with respect to the stationary distribution it holds that

$$E[\varepsilon_t^4] = \kappa \frac{\omega_0^2 + 2\omega_0^2 \frac{\alpha_0}{1-\alpha_0}}{1 - \kappa\alpha_0^2} = \kappa \frac{\omega_0^2(1+\alpha_0)}{(1-\kappa\alpha_0^2)(1-\alpha_0)},$$

and $E[h_t^2] = \omega_0^2(1+\alpha_0)(1-\kappa\alpha_0^2)^{-1}(1-\alpha_0)^{-1}$. Utilizing the representation for h_t as a function of $z_t, ..., z_{t-k}$ and h_{t-k} from Nelson (1990) it holds that for some

$k \in \mathbb{N}_0$

$$
\begin{aligned}
E[\varepsilon_{t-k}^2 \varepsilon_t^2] &= E\left[\varepsilon_{t-k}^2 z_t^2 \left(h_{t-k} \prod_{i=1}^{k} \alpha_0 z_{t-i}^2 + \omega_0 \left(1 + \sum_{k=1}^{k-1} \prod_{i=1}^{k} \alpha_0 z_{t-i}^2 \right)\right)\right] \\
&= \kappa E[\alpha_0^k h_{t-k}^2] + \omega_0 E\left[\varepsilon_{t-k}^2 \frac{1-\alpha_0^k}{1-\alpha_0}\right] \\
&= \kappa \alpha_0^k \omega_0^2 \frac{1+\alpha_0}{(1-\kappa\alpha_0^2)(1-\alpha_0)} + \omega_0^2 \frac{1-\alpha_0^k}{(1-\alpha_0)^2} \\
&= \frac{\omega_0^2}{(1-\alpha_0)^2} + \frac{\omega_0^2 \alpha_0^k (\kappa - 1)}{(1-\kappa\alpha_0^2)(1-\alpha_0)^2}.
\end{aligned}
$$

Using the symmetry of z_t's distribution and the infinite representation of y_t yields

$$
\begin{aligned}
E[y_{t-1}^2 h_t] &= E\left[\left(\sum_{i=0}^{\infty} \rho_0^i \varepsilon_{t-1-i}\right)^2 (\omega_0 + \alpha_0 \varepsilon_{t-1}^2)\right] \\
&= \omega_0 \sum_{i=0}^{\infty} \rho_0^{2i} E[\varepsilon_{t-i-1}^2] + \alpha_0 \sum_{i=0}^{\infty} \rho_0^{2i} E[\varepsilon_{t-h}^2 \varepsilon_t^2] \\
&= \frac{\omega_0^2}{(1-\alpha_0)(1-\rho_0^2)} + \frac{\alpha_0 \omega_0^2}{(1-\alpha_0)^2(1-\rho_0^2)} + \frac{\omega_0^2 \alpha_0^2 (\kappa-1)}{(1-\kappa\alpha_0^2)(1-\alpha_0)^2} \sum_{i=0}^{\infty} \alpha_0^i \rho_0^{2i} \\
&= \frac{\omega_0^2}{(1-\alpha_0^2)(1-\rho_0^2)} + \frac{\omega_0(\kappa-1)}{(1-\kappa\alpha_0^2)(1-\alpha_0)^2(1-\alpha_0\rho_0^2)},
\end{aligned}
$$

and

$$
E[y_{t-1}^2] = E\left[\left(\sum_{i=0}^{\infty} \rho_0^i \varepsilon_{t-1-i}\right)^2\right] = \frac{\omega_0}{(1-\alpha_0)(1-\rho_0^2)}.
$$

Next, define the filtration $\mathbb{F}_t = \sigma(\varepsilon_t, y_t, ...)$. In order to apply a standard CLT for martingale difference sequences (e.g. Brown (1971)) we first verify the Lindeberg condition

$$
\frac{1}{T} \sum_{t=1}^{T} E[y_{t-1}^2 \varepsilon_t^2 1_{\{|y_{t-1}\varepsilon_t| > \delta\sqrt{T}\}} \mid \mathbb{F}_t] \leq \frac{k}{\delta^\xi T^{1+\xi/2}} \sum_{t=1}^{T} (y_{t-1}\sqrt{h_t})^{2+\xi} \to 0,
$$

where k is a positive constant and $\xi > 0$ is chosen such that $E[(y_{t-1}\sqrt{h_t})^{2+\xi}]$ is finite. The constant ξ exists because the inequality which ensures finite fourth order moment is a sharp inequality, see Lange et al. (2007) for details. Furthermore

$$\frac{1}{T}\sum_{t=1}^{T} E[y_{t-1}^2 \varepsilon_t^2 \mid \mathbb{F}_{t-1}] = \frac{1}{T}\sum_{t=1}^{T} y_{t-1}^2 h_t \xrightarrow{P} E[y_{t-1}^2 h_t].$$

Hence

$$\sqrt{T}(\hat{\rho}_{OLS} - \rho_0) \xrightarrow{D} N(0, \Sigma) \quad \text{as} \quad T \to \infty,$$

where

$$\Sigma = \frac{(1-\alpha_0)^2(1-\rho_0^2)^2}{\omega_0^2} E[y_{t-1}^2 h_t] = (1-\rho_0^2) + \frac{(\kappa-1)(1-\rho_0^2)^2\alpha_0}{(1-\kappa\alpha_0^2)(1-\alpha_0\rho_0^2)}.$$

This completes the proof. □

The proof of the Theorem 2 rests to a large extent on the following lemma.

Lemma 2. *Under the assumptions of Theorem 2 all finite dimensional vectors* $\mathbf{y}_t(k) = (y_t, ..., y_{t+k})$ *have regularly varying tails as defined in Resnick (1987) with the same tail index* λ *as the ARCH process.*

The proof is inspired by the proofs of Lemma A.3.26 in Embrechts, Klüppelberg & Mikosch (1997) and Lemma 4.24 in Resnick (1987). However none of these results are directly applicable since the innovations are not independent.

Proof of Lemma 2. 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 will be omitted. In addition due to symmetry of the distribution of y_t and ε_t all arguments will be given using the absolute value of both only.

Since the ARCH process has finite second order moment y has the representation $y = \sum_{i=0}^{\infty} \rho^i \varepsilon_{-i}$. Define $y^{(m)} = \sum_{i=0}^{m-1} \rho^i \varepsilon_{-i}$ for any $m \geq 1$. We will now show that the remainder $y - y^{(m)}$ has negligible influence on the tails of y for m sufficiently large. Observe that for any $\delta \in]0,1[$ and $x > 0$ it holds that,

$$P(|y| > x) \leq P(|y^{(m)}| > (1-\delta)x) + P(\sum_{i=m}^{\infty} |\rho|^i |\varepsilon_{-i}| > \delta x) \tag{8}$$

and

$$P(|y| > x) \geq P(|y^{(m)}| > (1+\delta)x) - P(\sum_{i=m}^{\infty} |\rho|^i |\varepsilon_{-i}| > \delta x) \tag{9}$$

In the following we show

$$\lim_{m\to\infty} \limsup_{x\to\infty} \frac{P\left(\sum_{i=m}^{\infty} |\rho|^i |\varepsilon_{-i}| > x\right)}{P(|\varepsilon_0| > x)} = 0. \tag{10}$$

Rewrite the numerator as

$$\begin{aligned}
&P\left(\sum_{i=m}^{\infty} |\rho|^i |\varepsilon_{-i}| > x\right) \\
&\quad = P\left(\sum_{i=m}^{\infty} |\rho|^i |\varepsilon_{-i}| > x, \bigvee_{i=m}^{\infty} |\rho|^i |\varepsilon_{-i}| > x\right) \\
&\qquad + P\left(\sum_{i=m}^{\infty} |\rho|^i |\varepsilon_{-i}| > x, \bigvee_{i=m}^{\infty} |\rho|^i |\varepsilon_{-i}| \leq x\right) \\
&\quad \leq P\left(\bigcup_{i=m}^{\infty} \left(|\rho|^i |\varepsilon_{-i}| > x\right)\right) \\
&\qquad + P\left(\sum_{i=m}^{\infty} |\rho|^i |\varepsilon_{-i}| 1_{\{|\rho|^i |\varepsilon_{-i}| \leq x\}} > x, \bigvee_{i=m}^{\infty} |\rho|^i |\varepsilon_{-i}| \leq x\right) \\
&\quad \leq \sum_{i=m}^{\infty} P(|\varepsilon_{-i}| > x|\rho|^{-i}) + P\left(\sum_{i=m}^{\infty} |\rho|^i |\varepsilon_{-i}| 1_{\{|\rho|^i |\varepsilon_{-i}| \leq x\}} > x\right).
\end{aligned}$$

Hence by Markov's inequality it holds that

$$\begin{aligned}
&\frac{P\left(\sum_{i=m}^{\infty} |\rho|^i |\varepsilon_{-i}| > x\right)}{P(|\varepsilon_0| > x)} \\
&\quad \leq \sum_{i=m}^{\infty} \frac{P(|\varepsilon_0| > x|\rho|^{-i})}{P(|\varepsilon_0| > x)} + x^{-1} \sum_{i=m}^{\infty} \frac{|\rho|^i E[|\varepsilon_0| 1_{\{|\varepsilon_0| \leq x|\rho|^{-i}\}}]}{P(|\varepsilon_0| > x)} \\
&\quad = I + II.
\end{aligned} \tag{11}$$

By Basrak, Davis & Mikosch (2002*b*) the random variable ε_0 has regular varying tails and by Proposition 0.8(ii) of Resnick (1987) it holds that for all $\tau > 0$ there exists a x_0 such that for all $x > x_0$

$$P(|\varepsilon_0| > x|\rho|^{-i})/P(|\varepsilon_0| > x) \leq (1+\tau)|\rho|^{i(\lambda-\tau)}.$$

For τ adequately small this bound is summable and hence by dominated convergence and the regular variation of ε_0 it holds that

$$\limsup_{x\to\infty} I \leq (1+\tau)\sum_{i=m}^{\infty}|\rho|^{i(\lambda-\tau)}.$$

In considering II, suppose temporarily that $0 < \lambda < 1$ (this will never be the case when $E[\varepsilon_0^2] < \infty$, but it is a necessary steep towards proving the full result). From an integration by parts it holds that

$$\frac{E[|\varepsilon_0| 1_{\{|\varepsilon_0|\leq x\}}]}{xP(|\varepsilon_0| > x)} \leq \frac{\int_0^x P(|\varepsilon_0| > u)du}{xP(|\varepsilon_0| > x)}$$

and applying Karamata's Theorem (from e.g. Resnick (1987)) this converges to $(1-\lambda)^{-1}$ as x tends to infinity. Thus the function $x \mapsto E[|\varepsilon_0| 1_{\{|\varepsilon_0|\leq x|\rho|^{-i}\}}]$ is regular varying with tail index $1-\lambda$ and applying again Proposition 0.8(ii) we have that for any $\tau > 0$, some constant k, and x sufficiently large it holds that

$$\begin{aligned}\frac{|\rho|^i E[|\varepsilon_0| 1_{\{|\varepsilon_0|\leq x|\rho|^{-i}\}}]}{xP(|\varepsilon_0| > x)} &= |\rho|^i \left(\frac{E[|\varepsilon_0| 1_{\{|\varepsilon_0|\leq x|\rho|^{-i}\}}]}{E[|\varepsilon_0| 1_{\{|\varepsilon_0|\leq x\}}]}\right)\frac{E[|\varepsilon_0| 1_{\{|\varepsilon_0|\leq x\}}]}{xP(|\varepsilon_0| > x)} \\ &\leq |\rho|^i(|\rho|^{-i})^{-(1-\lambda+\tau)}k = k|\rho|^{i(\lambda-\tau)},\end{aligned}$$

which is summable for τ adequately small. So we conclude

$$\limsup_{x\to\infty} II \leq k\sum_{i=m}^{\infty}|\rho|^{i(\lambda-\tau)}$$

and hence when $0 < \lambda < 1$ there exists constants $\tilde{\tau} \in]0,1[$ and $\tilde{k} > 0$ such that

$$\limsup_{x\to\infty}\frac{P\left(\sum_{i=m}^{\infty}|\rho|^i|\varepsilon_{-i}| > x\right)}{P(|\varepsilon_0| > x)} \leq \tilde{k}\sum_{i=m}^{\infty}|\rho|^{i\tilde{\tau}} < \infty. \tag{12}$$

If $\lambda \geq 1$ we get a similar inequality by reducing to the case $0 < \lambda < 1$ as follows. Pick $\eta \in]\lambda, \lambda\tilde{\tau}^{-1}[$ and set $c = \sum_{i=m}^{\infty} |\rho|^i$ and $p_i = |\rho|^i/c$ then by Jensen's inequality (e.g. Feller (1971) p. 153) we get

$$\begin{aligned}\left(\sum_{i=m}^{\infty} |\rho|^i |\varepsilon_{-i}|\right)^{\eta} &= c^{\eta}\left(\sum_{i=m}^{\infty} p_i |\varepsilon_{-i}|\right)^{\eta} \\ &\leq c^{\eta} \sum_{i=m}^{\infty} p_i |\varepsilon_{-i}|^{\eta} = c^{\eta-1} \sum_{i=m}^{\infty} |\rho|^i |\varepsilon_{-i}|^{\eta}.\end{aligned}$$

Thus

$$\frac{P\left(\sum_{i=m}^{\infty} |\rho|^i |\varepsilon_{-i}| > x\right)}{P(|\varepsilon_0| > x)} \leq \frac{P\left(\sum_{i=m}^{\infty} |\rho|^i |\varepsilon_{-i}|^{\eta} > c^{1-\eta} x^{\eta}\right)}{P(|\varepsilon_0|^{\eta} > x^{\eta})}.$$

By Bingham, Goldie & Teugels (1987) Proposition 1.5.7(i) the function $P(|\varepsilon_0|^{\eta} > x^{\eta})$ is regularly varying with tail index $\eta^{-1}\lambda \in]0, 1[$. Hence (12) gives

$$\limsup_{x\to\infty} \frac{P\left(\sum_{i=m}^{\infty} |\rho|^i |\varepsilon_{-i}| > x\right)}{P(|\varepsilon_0| > x)} \leq \tilde{k} \sum_{i=m}^{\infty} |\rho|^{i\lambda\eta^{-1}} c^{\lambda(1-\eta^{-1})} < \infty. \tag{13}$$

This proves (10). Combine (8) and (9) with the above to obtain the relations

$$\begin{aligned}&\liminf_{x\to\infty} \frac{P(|y| > x)}{P(|\varepsilon_0| > x)} \\ \geq\ & \liminf_{x\to\infty} \frac{P(|y^{(m)}| > (1+\delta)x)}{P(|\varepsilon_0| > x)} - \limsup_{x\to\infty} \frac{P(\sum_{i=m}^{\infty} |\rho|^i |\varepsilon_0| > \delta x)}{P(|\varepsilon_0| > x)} \\ \to\ & \liminf_{x\to\infty} \frac{P(|y^{(m)}| > (1+\delta)x)}{P(|\varepsilon_0| > x)} \quad \text{as } m \to \infty\end{aligned}$$

and

$$\begin{aligned}&\limsup_{x\to\infty} \frac{P(|y| > x)}{P(|\varepsilon_0| > x)} \\ \leq\ & \limsup_{x\to\infty} \frac{P(|y^{(m)}| > (1-\delta)x)}{P(|\varepsilon_0| > x)} + \limsup_{x\to\infty} \frac{P(\sum_{i=m}^{\infty} |\rho|^i |\varepsilon_0| > \delta x)}{P(|\varepsilon_0| > x)} \\ \to\ & \limsup_{x\to\infty} \frac{P(|y^{(m)}| > (1+\delta)x)}{P(|\varepsilon_0| > x)} \quad \text{as } m \to \infty\end{aligned}$$

By Basrak et al. (2002b) equation (2.6) $y^{(m)}$ is regular varying with tail index λ and hence will

$$\lim_{x\to\infty} \frac{P(|y^{(m)}| > x)}{P(|\varepsilon_0| > x)} = c_m$$

for a sequence of constants c_m. Using the same type of arguments as for $y - y^{(m)}$ one can conclude that c_m tends to a finite limit c as m tends to infinity. Hence it can be concluded that

$$\begin{aligned}
(c-\delta)(1+\delta)^{-\lambda} &\leq \liminf_{x\to\infty} \frac{P(|y| > x)}{P(|\varepsilon_0| > x)} \\
&\leq \limsup_{x\to\infty} \frac{P(|y| > x)}{P(|\varepsilon_0| > x)} \\
&\leq (c+\delta)(1-\delta)^{-\lambda}.
\end{aligned}$$

Now by letting δ go towards zero it can be concluded that y is regular varying with index λ.

Finally we wish to extent this result to all vectors of the form $\mathbf{y}(k) = (y_0, ..., y_k)$. By Basrak, Davis & Mikosch (2002a) Theorem 1.1(ii) it suffices to show that all linear combinations $v \in \mathbb{R}^k \setminus \{0\}$ are regular varying. However, for all v

$$v'\mathbf{y}(k) = \sum_{i=0}^{\infty} c_i \varepsilon_{-i},$$

where the coefficients are absolutely summable and smaller than one in absolute value for i sufficiently large. Hence can regular variation of $\mathbf{y}(k)$ be verified by the same arguments as above. This completes the proof of Lemma 2 □

Proof of Theorem 2. Define the empirical autocovariance and the empirical autocorrelation as

$$\begin{aligned}
\gamma_T(r) &= \frac{1}{T}\sum_{t=1}^{T} y_t y_{t+r}, \quad r = 0, 1 \\
\rho_T(1) &= \gamma_T(1)/\gamma_T(0).
\end{aligned}$$

These are clearly closely related to the OLS estimator of the autoregressive parameter. We will therefore in the following prove that $\gamma_T(r) - E[\gamma_T(r)]$ and $\rho_T(1) - E[\gamma_T(1)]/E[\gamma_T(0)]$ are both asymptotically stable with index $\lambda/2$.

Define $\mathbf{y}_t(k) = (y_t, ..., y_{t+k})'$ and let a_T be a sequence such that $TP(|y_t| > a_T) \to 1$ (one can choose a_T to be the $1-1/T$ quantile of the distribution function for $|y_t|$). The proof is structured as the proof of Theorem 2.10 in Basrak et al. (2002*b*) and we must therefore verify that

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

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

$(A.3)$ condition (2.10) of Davis & Mikosch (1998), and

$(A.4)$ condition (3.3) of Davis & Mikosch (1998).

$(A.1)$ follows straight from Lemma 2. Furthermore Lemma 1 establishes that the Markov chain $(y_{t-1}, \varepsilon_t)'$ is geometrically ergodic, this implies in particular that the stationary version is strongly mixing (actually even β-mixing) with geometrically decreasing rate function. And since the condition $\mathcal{A}(a_T)$ is implied by strong mixing the verification of $(A.2)$ is complete.

The two remaining conditions require a bit more work. With $|\cdot|$ denoting the max norm condition (2.10) of Davis & Mikosch (1998) can be stated as

$$\lim_{m\to\infty} \limsup_{T\to\infty} P\Big(\bigvee_{m \leq |t| \leq r_T} |\mathbf{y}_t(k)| > a_T x \,\Big|\, |\mathbf{y}_0(k)| > a_T x \Big) = 0, \ \ x > 0, \qquad (14)$$

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

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

The recursion of Nelson (1990) gives

$$E_0[\varepsilon_t^2] = \varepsilon_0 \alpha^t + \omega \frac{1-\alpha^t}{1-\alpha} \leq \varepsilon_0 \alpha^t + C_0,$$

for some positive constant C_0 independent of t. Direct calculations provide the relation

$$\sum_{i=0}^{t-1} \rho^{2i} \alpha^{t-i} = \begin{cases} \frac{\alpha(\rho^{2t}-\alpha^t)}{\rho^2-\alpha} & \text{if } \alpha \neq \rho^2 \\ t\alpha^t & \text{if } \alpha = \rho^2 \end{cases} .$$

Note that the sum converges to zero as t tends to infinity for all ρ, α smaller than one in absolute value. Introduce the auxiliary process $\tilde{y}_t = \sum_{i=0}^{\infty} |\rho|^i |\varepsilon_{t-i}|$, which is clearly positive. Inspecting the proof of Lemma 2 reveals that $\tilde{y}_t$ is regularly varying with tail index λ. In addition one has the relation $\tilde{y}_t \geq |y_t|$ for all t and $\tilde{y}_0 \geq |\varepsilon_0|$. Hence it holds that

$$\frac{E[1_{\{|y_0|>a_T x\}} \varepsilon_0^2]}{a_T^2 x^2 P(|y_0| > a_T x)} \leq \frac{E[1_{\{|\tilde{y}_0|^2 > a_T^2 x^2\}} \tilde{y}_0^2]}{a_T^2 x^2 P(|\tilde{y}_0^2| > a_T^2 x^2)} \frac{P(|\tilde{y}_0| > a_T x)}{P(|y_0| > a_T x)},$$

and by Karamata's Theorem (e.g. Resnick (1987) Proposition 0.6)

$$\limsup_{T\to\infty} \frac{E[1_{\{|y_0|>a_T x\}} \varepsilon_0^2]}{a_T^2 x^2 P(|y_0| > a_T x)} \leq \frac{C_1}{\lambda - 2},$$

for some constant C_1. Applying Karamata's Theorem again it can concluded that there exists T_0 such that for all $T > T_0$ it holds

$$\begin{aligned} I_{t,T} &\leq \frac{E\left[1_{\{|y_0|>a_T x\}} \left(\rho^{2t}|y_0|^2 + \varepsilon_0^2 \frac{\alpha(\rho^{2t}-\alpha^t)}{\rho^2-\alpha} + C_2\right)\right]}{a_T^2 x^2 P(|y_0| > a_T x)} \\ &\leq C_3 \rho^{2t} + C_4 \frac{\alpha(\rho^{2t}-\alpha^t)}{\rho^2-\alpha} + \frac{C_2}{a_T^2 x^2} \\ &\leq C_5 a^t + \frac{C_2}{a_T^2 x^2}, \end{aligned}$$

for some positive constants $C_2, ..., C_5$ and $a \in]0,1[$ all independent of t, since by assumption both ρ and α are smaller that one in absolute value. Note that the special case $\alpha = \rho^2$ can be treated using the same arguments. We are now ready to verify (14).

$$
\begin{aligned}
&\lim_{m\to\infty}\limsup_{T\to\infty} P\big(\bigvee_{m\le|t|\le r_T} |\mathbf{y}_t(k)| > a_T x\big||\mathbf{y}_0(0)| > a_T x\big)\\
\le\;& \lim_{m\to\infty}\limsup_{T\to\infty} 2(k+1)\sum_{t=m}^{r_T+k} P\left(|y_t| > a_T x \mid |y_0| > a_T x\right) \underbrace{\frac{P(|y_0| > a_T x)}{P(|\mathbf{y}_0(k)| > a_T x)}}_{\le 1}\\
\le\;& \lim_{m\to\infty} 2(k+1)\sum_{t=m}^{\infty} C_5 a^t + \lim_{m\to\infty}\limsup_{T\to\infty} 2(k+1)(r_T+k)/(a_T^2 x^2)\\
=\;& 0,
\end{aligned}
$$

by choosing r_T such that $r_T/a_T^2 \to 0$. Note that negative values of t are dealt with by noting that due to stationarity the following relation holds for $t > 0$

$$
\begin{aligned}
P(|y_{-t}| > a_T x \mid |y_0| > a_T x) &= P(|y_{-t}| > a_T x, |y_0| > a_T x)/P(|y_0| > a_T x)\\
&= P(|y_{-t}| > a_T x, |y_0| > a_T x)/P(|y_{-t}| > a_T x)\\
&= P(|y_t| > a_T x \mid |y_0| > a_T x)
\end{aligned}
$$

This completes the verification of $(A.3)$. Finally $(A.4)$ is considered. In the setup of the AR-ARCH model condition (3.3) of Davis & Mikosch (1998) reads

$$
\begin{aligned}
&\lim_{x\to 0}\limsup_{T\to\infty} P\big(\big|a_T^{-2}\sum_{t=1}^{T} y_t y_{t+1} 1_{\{|y_t y_{t+1}|\le a_T^2 x\}} - E\big[a_T^{-2}\sum_{t=1}^{T} y_t y_{t+1} 1_{\{|y_t y_{t+1}|\le a_T^2 x\}}\big]\big| > \delta\big)\\
&\quad = 0,
\end{aligned}
$$

for all $\delta > 0$, which can also be found in Davis & Hsing (1995) p. 895. Markov's inequality and Kamarata's Theorem (the required regular variation of $y_t y_{t+1}$ can be verified by the same arguments as for y_t) now give

$$
\begin{aligned}
P(|a_T^{-2}\sum_{t=1}^{T} y_t y_{t+1} 1_{\{|y_t y_{t+1}| \leq a_T^2 x\}} - E[a_T^{-2}\sum_{t=1}^{T} y_t y_{t+1} 1_{\{|y_t y_{t+1}| \leq a_T^2 x\}}]| > \delta) \\
\leq \quad & \frac{1}{\delta^2} a_T^{-4} \sum_{t=1}^{T} E[(y_t y_{t+1} 1_{\{|y_t y_{t+1}| \leq a_T^2 x\}} - E[y_t y_{t+1} 1_{\{|y_t y_{t+1}| \leq a_T^2 x\}}])^2] \\
\leq \quad & \frac{4}{\delta^2} a_T^{-4} \sum_{t=1}^{T} E[y_t^2 y_{t+1}^2 1_{\{|y_t y_{t+1}|^2 \leq a_T^4 x^2\}}] \\
= \quad & \frac{4}{\delta^2} a_T^{-4} T E[y_0^2 y_1^2 1_{\{|y_0 y_1|^2 \leq a_T^4 x^2\}}] \\
\sim \quad & C_6 x^2 T P(|y_0 y_1|^2 > a_T^4 x^2) \ \text{ for large } T \\
\rightarrow \quad & C_7 x^2 \ \text{ as } T \rightarrow \infty \\
\rightarrow \quad & \ \text{ as } x \rightarrow 0.
\end{aligned}
$$

Using the same arguments, it can be shown that $(A.4)$ also holds for the sequence y_t^2. This completes the verification of $(A.4)$. Due to $(A.1)$ - $(A.4)$ one can apply Theorem 3.5 of Davis & Mikosch (1998). Note that their condition (3.4) is not meet, but by inspecting the proof it becomes clear that $(A.4)$ suffices. Hence it holds that

$$
\begin{aligned}
T a_T^{-2}(\gamma_T(r) - E[\gamma_T(r)]) & \xrightarrow{D} W_r \quad \text{as } T \rightarrow \infty, \quad r = 0, 1 \\
T a_T^{-2}(\rho_T(1) - E[\gamma_T(1)]/E[\gamma_T(0)]) & \xrightarrow{D} S_0 \quad \text{as } T \rightarrow \infty,
\end{aligned}
$$

where W_0, W_1, and S_0 are $\lambda/2$-stable random variables. As the convergence result in Theorem 3.5 of Davis & Mikosch (1998) is based on an application of the continuous mapping theorem the stated convergence results hold jointly. Since a_T can be chosen to be the $1 - 1/T$ quantile of the distribution function of $|y_t|$ one gets that a_T can be chosen as $a_T = T^{1/\lambda}$. This implies that the normalizing sequence $T a_T^{-2}$ can be chosen as $T^{1-2/\lambda}$. Hence it holds that

$$
T^{1-2/\lambda}(\hat{\rho}_{OLS} - \rho_0) = T^{1-2/\lambda}(\rho_T(1) - E[\gamma_T(1)]/E[\gamma_T(0)]) \xrightarrow{D} S_0.
$$

This completes the proof. □

Proof of Theorem 3. Since the process $(\varepsilon_t)_{t=1}^T$ is an ARCH(1) process it follows directly from Davis & Mikosch (1998) pp. 2069 - 2070 that

$$T^{-2/\lambda}\sum_{t=1}^{T}\varepsilon_t\varepsilon_{t-1}\xrightarrow{D}S_1 \quad \text{and} \quad T^{-4/\lambda}\sum_{t=1}^{T}\varepsilon_t^2\varepsilon_{t-1}^2\xrightarrow{D}S_2,$$

where S_1 is a $\lambda/2$ stable random variable and S_2 is a $\lambda/4$ stable random variable. Again the convergence hold jointly, but the remaining parameters and dependence structure are unknown. Under H_0 we can rewrite V_T as

$$V_T=\big(T^{-2/\lambda}\sum_{t=1}^{T}\varepsilon_t\varepsilon_{t-1}\big)\big(T^{-4/\lambda}\sum_{t=1}^{T}\varepsilon_t^2\varepsilon_{t-1}^2\big)^{-1/2},$$

and the continuous mapping theorem completes the proof. □

References

Basrak, B., R.A. Davis & T. Mikosch (2002*a*), 'A characterization of multivariate variation', *The Annals of Applied Probability* **12**(3), 908 – 920.

Basrak, B., R.A. Davis & T. Mikosch (2002*b*), 'Regular variation of GARCH processes', *Stochastic Processes and their Applications* **99**, 95–115.

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

Brown, B. M. (1971), 'Martingale central limit theorems', *Annals of Mathematical Statistics* **42**, 59–66.

Davis, R.A. & T. Hsing (1995), 'Point process and partial sum convergence for weakly dependent random variables with infinite variance', *The Annals of Probability* **23**(2), 879–917.

Davis, R.A. & T. Mikosch (1998), 'The sample autocorrelations of heavy-tailed processes with applications to ARCH', *The Annals of Statistics* **26**, 2049–2080.

Embrechts, P., C. Klüppelberg & T. Mikosch (1997), *Modelling Extremal Events*, Springer, Berlin.

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

Feller, W. (1971), *An Introduction to Probability Theory and Its Applications*, Vol. II, 2nd ed., Wiley, New York.

Giné, E., F. Götze & D.M. Mason (1997), 'When is the Student t-statistic asymptotically standard normal?', *The Annals of Probability* **25**, 1514–1531.

Lange, T., A. Rahbek & S.T. Jensen (2007), Estimation and asymptotic inference in AR-ARCH model. Working paper Department of Mathematical Sciences, University of Copenhagen.

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

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

White, H. (1980), 'A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity', *Econometrica* **48**, 817–838.