Least absolute deviations estimation for ARCH and GARCH models

By LIANG PENG

School of Mathematics, Georgia Institute of Technology, Atlanta, GA 30332-0160, U.S.A.

peng@math.gatech.edu

AND QIWEI YAO

Department of Statistics, The London School of Economics and Political Science

Houghton Street, London, WC2A 2AE, U.K.

q.yao@lse.ac.uk

Summary

The class of ARCH/GARCH models is arguably the most frequently used family for modelling conditional second moments, and has proved particularly valuable in modelling highly volatile time series. These include financial data, which can be particularly heavy tailed. Hall & Yao (2003) showed that, for ARCH/GARCH models with heavy-tailed errors, the conventional maximum quasilikelihood estimator suffers from complex limit distributions and slow convergence rates. In this paper three types of absolute deviations estimator have been examined, and the one based on logarithmic transformation turns out to be particularly appealing. We have shown that this estimator is asymptotically normal and unbiased. Furthermore it enjoys the standard convergence rate of $n^{1/2}$ regardless of whether the errors are heavy-tailed or not. Simulation lends further support to our theoretical results.

Some key words: ARCH; Asymptotic normality; GARCH; Gaussian likelihood; Heavy tail; Least absolute deviations estimator; Maximum quasilikelihood estimator; Time series.

1. Introduction

With the motivation of explaining and forecasting risk in financial time series, ARCH and GARCH models were proposed for modelling explicitly the conditional second moments; see Engle (1982), Bollerslev (1986) and Taylor (1986). Early successful applications of ARCH/GARCH models were confined to the case of normal errors. On the other hand, empirical evidence suggests that financial data may have heavy tails (Mittnik et al. 1988; Mittnik & Rachev, 2000) and models with heavy-tailed errors have also been adopted in practice. Excellent surveys of ARCH/GARCH modelling for financial data are available in Shephard (1996) and Rydberg (2000). For their theoretical properties, we refer to §4.2 of Fan & Yao (2003).

When the errors in GARCH models are normal, an explicit conditional likelihood function is readily available to facilitate parameter estimation. In practice, the error distribution is typically unknown. Nevertheless, conditional Gaussian likelihood still motivates parameter estimators, which may be called maximum quasilikelihood estimators. The asymptotic properties of maximum quasilikelihood estimators were established for ARCH(p) models by Weiss (1986), for GARCH(1,1) models by Lee & Hansen (1994) and Lumsdaine (1996), and for general GARCH(p, q) models by Hall & Yao (2003). In fact Hall & Yao (2003) showed that when the error distribution is heavy-tailed with an infinite fourth moment, the estimators may not be asymptotically normal, the range of possible limit distributions is extraordinarily large, and the convergence rate is slower than the standard rate of $n^{1/2}$. Complex asymptotic properties were also observed from a Whittle estimator (Giraitis & Robinson, 2001) for heavy tailed GARCH(1,1) models in an unpublished University of Copenhagen report by T. Mikosch and D. Straumann.

Note that quasi-maximum likelihood estimation based on a Gaussian likelihood may be viewed as an extended version of least squares estimation, which is known to be sensitive to heavy tails. In contrast, a least absolute deviations method would be more robust; see, for example, Davis et al. (1992), Adler et al. (1997) and the references within. In this paper, we explore in §2 three types of least absolute deviations estimator for ARCH and GARCH models and advocate the one based on logarithmic transformation. Our theoretical result in §4 shows that this estimator is asymptotically normal and unbiased. Furthermore, it enjoys the $n^{1/2}$ convergence rate regardless of the tail-weight of error distributions; see Remark 3 in §4 below. This is in marked contrast to the conventional Gaussian maximum likelihood estimator. The simulation results in §3 lend further support to our theoretical results.

2. Models and estimators

A generalised autoregressive conditional heteroscedastic, GARCH, model with orders $p \geq 1$ and $q \geq 0$ is defined as

$$X_t = \sigma_t \varepsilon_t, \quad \text{and} \quad \sigma_t^2 \equiv \sigma_t(\theta)^2 = c + \sum_{i=1}^{p} b_i X_{t-i}^2 + \sum_{j=1}^{q} a_j \sigma_{t-j}^2, \tag{1}$$

where $c > 0$, $b_j \geq 0$ and $a_j \geq 0$ are unknown parameters, $\theta = (c, b_1, \cdots, b_p, a_1, \cdots, a_q)^{\mathrm{T}}$, $\{\varepsilon_t\}$ is a sequence of independent and identically distributed random variables with mean 0 and variance 1, and ε_t is independent of $\{X_{t-k}, k \geq 1\}$ for all t. When $q = 0$, (1) reduces to an autoregressive conditional heteroscedastic, ARCH, model. The necessary and sufficient condition for (1) to define a unique strictly stationary process $\{X_t,\ t = 0, \pm 1, \pm 2, \cdots\}$ with $EX_t^2 < \infty$ is that

$$\sum_{i=1}^{p} b_i + \sum_{j=1}^{q} a_j < 1. \tag{2}$$

Furthermore, for such a stationary solution, $EX_t = 0$ and $\mathrm{var}(X_t) = c/(1 - \sum_{i=1}^{p} b_i - \sum_{j=1}^{q} a_j)$; see Giraitis et al. (2000), and also Theorem 4.4 of Fan & Yao (2003).

The maximum quasilikelihood estimation method can be motivated by temporarily assuming that $\varepsilon_t \sim N(0,1)$. Given $\{(X_k, \sigma_k^2), 1 \leq k \leq \nu\}$ with $\nu \geq \max(p,q)$, the conditional density function of $X_{\nu+1}, \cdots, X_n$ is then proportional to

$$\Big(\prod_{t=\nu+1}^{n} \sigma_t^2 \Big)^{-1/2} \exp\Big(-\frac{1}{2} \sum_{t=\nu+1}^{n} \frac{X_t^2}{\sigma_t^2} \Big). \tag{3}$$

Under condition (2), $\sigma_t^2 = \sigma_t(\theta)^2$ may be expressed as

$$\sigma_t(\theta)^2 = \frac{c}{1 - \sum_{j=1}^{q} a_j} + \sum_{i=1}^{p} b_i X_{t-i}^2 + \sum_{i=1}^{p} b_i \sum_{k=1}^{\infty} \sum_{j_1=1}^{q} \cdots \sum_{j_k=1}^{q} a_{j_1} \cdots a_{j_k} X_{t-i-j_1-\cdots-j_k}^2, \tag{4}$$

where the multiple sum vanishes if $q = 0$; see Hall & Yao (2003). This leads to the following approximation for σ_t^2 based on $X_1, \cdots, X_t$:

$$\begin{aligned} \widetilde{\sigma}_t(\theta)^2 &= \frac{c}{1 - \sum_{j=1}^{q} a_j} + \sum_{i=1}^{\min(p,t-1)} b_i X_{t-i}^2 + \sum_{i=1}^{p} b_i \sum_{k=1}^{\infty} \sum_{j_1=1}^{q} \cdots \sum_{j_k=1}^{q} a_{j_1} \cdots a_{j_k} \\ &\times \quad X_{t-i-j_1-\cdots-j_k}^2 I(t - i - j_1 - \cdots - j_k \geq 1). \end{aligned} \tag{5}$$

Maximising (3) with σ_t^2 replaced by $\widetilde{\sigma}_t^2$, we obtain the quasi-maximum likelihood estimator

$$\widehat{\theta}_{\rm ml} = \arg\min_{\theta} \sum_{t=\nu+1}^{n} \Big[\frac{X_t^2}{\widetilde{\sigma}_t(\theta)^2} + \log\{\widetilde{\sigma}_t(\theta)^2\}\Big] \tag{6}$$

where the minimisation is taken over all the nonnegative values of the parameters. The asymptotic properties of the estimator $\widehat{\theta}_{\rm ml}$ were derived in Hall & Yao (2003). In particular, when the distribution of ε_t is heavy-tailed in the sense that $E(|\varepsilon_t|^d) = \infty$ for some $2 < d \leq 4$, the convergence rate of $\widehat{\theta}_{\rm ml}$ is slower than the standard rate of $n^{1/2}$.

Now we reparameterise the model (1) in such a way that the median of ε_t^2, instead of the variance of ε_t, is equal to 1 while $E\varepsilon_t = 0$ unchanged. Under this new parameterisation the parameters c and $b_1, \cdots, b_p$ differ from those in the old setting by a common positive constant factor while the parameters $a_1, \cdots, a_q$ remain unchanged. Furthermore, the form of expansion (4) is also unchanged. Write

$$X_t^2/\sigma_t(\theta)^2 = 1 + e_{t,1}, \tag{7}$$

where $e_{t,1} = (\varepsilon_t^2 - 1)$ which has median 0. This leads to an absolute deviations estimator

$$\widehat{\theta}_1 = \arg\min_{\theta} \sum_{t=\nu+1}^{n} |X_t^2/\widetilde{\sigma}_t(\theta)^2 - 1|, \tag{8}$$

which is an L_1 estimator based on regression relationship (7). Although the idea behind the above estimation is simple, the estimator $\widehat{\theta}_1$ is, unfortunately, biased; see Remark 4 in §4 below. To overcome this shortcoming, we define a modified form of least absolute deviations estimator as follows:

$$\widehat{\theta}_2 = \arg\min_{\theta} \sum_{t=\nu+1}^{n} |\log(X_t^2) - \log\{\widetilde{\sigma}_t(\theta)^2\}|, \tag{9}$$

which is motivated by the regression relationship

$$\log(X_t^2) = \log\{\sigma_t(\theta)^2\} + e_{t,2}, \tag{10}$$

where $e_{t,2} = \log(\varepsilon_t^2)$. Note that median of $e_{t,2}$ is equal to $\log\{\text{median}(\varepsilon_t^2)\}$, which is 0 under the reparameterisation. The distribution of X_t^2 is confined to the nonnegative half axis and is typically skewed. Intuitively the log-transform will make the distribution less skewed. Theorems 1 and 2 below show that the estimator $\widehat{\theta}_2$ is in fact asymptotically normal and unbiased under very mild conditions.

Our third estimator is motivated by the simple regression equation

$$X_t^2 = \sigma_t^2 + e_{t,3}, \tag{11}$$

where $e_{t,3} = \sigma_t^2(\varepsilon_t^2 - 1)$. Again under the new parameterisation, the median of $e_{t,3}$ is 0. This leads to the estimator

$$\widehat{\theta}_3 = \arg\min_{\theta} \sum_{t=\nu+1}^{n} |X_t^2 - \widetilde{\sigma}_t(\theta)^2|. \tag{12}$$

Intuitively we prefer the estimator $\widehat{\theta}_2$ to $\widehat{\theta}_3$ since the error terms $e_{t,2}$ in regression model (10) are independent and identically distributed while the errors $e_{t,3}$ in model (11) are not independent. Therefore, ideally the sum on the right-hand side of (12) should be replaced by a weighted sum with weights reflecting the dependence, which is typically intractable. In fact the asymptotic normality of $\widehat{\theta}_3$ requires more conditions; see Remark 5 in §4.

The minimisation in (8), (9) and (12) should be taken over all $c > 0$ and all nonnegative b_i's and a_j's. For a pure ARCH process, i.e. $q = 0$, it is easy to see from (5) that $\widetilde{\sigma}_t(\theta)^2 \equiv \sigma_t(\theta)^2$ for all $t > p$. Thus we may let $\nu = p$ in the definitions of the above estimators.

Remark 1. All our three least absolute deviations estimators were derived from relevant regression relationships. Like least squares estimators, they make no use of distribution information. For heavy-tailed data, a plausible pseudolikelihood approach may assume that ε_t has a Laplace distribution. The resulting estimator will be derived from minimising

$$\sum_t \log\{\sigma_t(\theta)\} + \sum_t |X_t/\sigma_t(\theta)|.$$

Unfortunately its asymptotic properties are as complex as those of $\widehat{\theta}_{\rm ml}$ defined in (6), and therefore we do not pursue this direction.

3. Numerical properties

In this section, we compare numerically the three least absolute deviations estimators with the conditional Gaussian maximum likelihood estimator for ARCH(2) and GARCH(1,1) models. In both cases we took the errors ε_t to have either a standard normal distribution or a standardised Student's t-distribution with $d = 3$ or $d = 4$ degrees of freedom. We standardised the t-distributions to ensure that their first two moments are, respectively, 0 and 1. Note that, when $\varepsilon_t \sim t(d)$, $E|\varepsilon_t|^d = \infty$. We used $c = 3$, $b_1 = 0.5$ and $b_2 = a_1 = 0.4$ in the models. Setting the sample size $n = 300$, we drew 500 samples respectively for each setting. We used $\nu = 20$ in the estimation for GARCH models. To ensure a fair comparison, we employed an exhaustive search procedure to find estimates. Since the values of parameters c and b_i estimated by the least absolute deviations methods differ from the numerical values specified above by a common factor, we define the average absolute error as $(|\widehat{b}_1/\widehat{c} - b_1/c| + |\widehat{b}_2/\widehat{c} - b_2/c|)/2$ for ARCH(2) and $(|\widehat{b}_1/\widehat{c} - b_1/c| + |\widehat{a}_1 - a_1|)/2$ for GARCH(1,1).

Figure 1 presents the boxplots for the average absolute errors. For models with heavy-tailed errors, i.e. $\varepsilon_t \sim t_d$ with $d = 3, 4$, the least absolute deviation estimator $\widehat{\theta}_2$ performed best. Furthermore, the gain from using $\widehat{\theta}_2$ was more pronounced when the tails were very heavy, i.e. $\varepsilon_t \sim t_3$. Note that, when $\varepsilon_t \sim t_4$, the Gaussian maximum likelihood estimator $\widehat{\theta}_{\rm ml}$ was almost as good as $\widehat{\theta}_2$, and was better than both $\widehat{\theta}_1$ and $\widehat{\theta}_3$. However, when $\varepsilon_t \sim t_3$, $\widehat{\theta}_{\rm ml}$ was no longer desirable. On the other hand, when the error ε_t was normal, $\widehat{\theta}_{\rm ml}$ was of course the best. In fact the average absolute error of $\widehat{\theta}_{\rm ml}$ was larger when the tail of the error distribution was heavier, which reflects the fact that, the heavier the tails are, the slower the convergence rate is; see Hall & Yao (2003). However, this is not always the case for the least absolute deviations estimators as they are more robust against heavy tails.

The above patterns were also observed in simulations with other models. In general, our numerical results suggest that we should use the least absolute deviations estimator $\widehat{\theta}_2$ when ε_t has heavy and especially very heavy tails, e.g. $E(|\varepsilon_t|^3) = \infty$, while in general the Gaussian maximum likelihood estimator $\widehat{\theta}_{\rm ml}$ is desirable as long as ε_t is not very heavy-tailed.

4. Asymptotic properties

4·1. *A central limit theorem*

In this section, we show that asymptotically also $\widehat{\theta}_2$ is a better estimator than $\widehat{\theta}_1$ and $\widehat{\theta}_3$. We establish the asymptotic normality of $\widehat{\theta}_2$. The properties of both $\widehat{\theta}_1$ and $\widehat{\theta}_3$ will be briefly stated.

Let $\theta^0 = (c^0, b_1^0, \cdots, b_p^0, a_1^0, \cdots, a_q^0)^{\rm T}$ be the true value under which the median of ε_t^2 equals 1, or equivalently the median of $\log(\varepsilon_t^2)$ equals 0. Define

$$A_{t0}(\theta) = \frac{1}{\sigma_t(\theta)^2}\frac{1}{1-\sum_{j=1}^q a_j},$$

$$A_{ti}(\theta) = \frac{1}{\sigma_t(\theta)^2}\Big(X_{t-i}^2 + \sum_{k=1}^{\infty}\sum_{j_1=1}^{q}\cdots\sum_{j_k=1}^{q} a_{j_1}\cdots a_{j_k} X_{t-i-j_1-\cdots-j_k}^2\Big), \quad i = 1, \cdots, p,$$

$$\begin{aligned} A_{t,p+j}(\theta) &= \frac{1}{\sigma_t(\theta)^2}\Big\{\frac{c}{(1-\sum_{j=1}^q a_j)^2} + \sum_{i=1}^{p} b_i X_{t-i-j}^2 \\ &+ \sum_{i=1}^{p} b_i \sum_{k=1}^{\infty}(k+1)\sum_{j_1=1}^{q}\cdots\sum_{j_k=1}^{q} a_{j_1}\cdots a_{j_k} X_{t-i-j-j_1-\cdots-j_k}^2\Big\}, \quad j = 1, \cdots, q. \end{aligned}$$

Let $U_t \equiv U_t(\theta) = \{A_{t0}(\theta), \cdots, A_{t,p+q}(\theta)\}^{\mathrm{T}}$. Put $\Sigma = E_0\{U_t(\theta^0)U_t(\theta^0)^{\mathrm{T}}\}$, a $(1+p+q)\times(1+p+q)$ matrix, where E_0 denotes expectation under $\theta = \theta^0$. Some regularity conditions are now in order.

Condition 1. There exists a unique strictly stationary solution $\{X_t\}$ of model (1) with $E_0(X_t^2) < \infty$.

Condition 2. All $b_1^0, \cdots, b_p^0$ are positive, and all $a_1^0, \cdots, a_q^0$ are positive if $q \geq 1$.

Condition 3. Σ is nonsingular.

Condition 4. $\log(\varepsilon_t^2)$ has a median zero, and its density function f is continuous at zero.

Remark 2. Condition 1 holds if and only if the true parameters (before the reparameterisation) satisfy inequality (2); see Theorem 4.4 of Fan & Yao (2003). The conditions which ensure the existence of a strictly stationary solution for model (1) have been established by, among others, Kesten (1973), Bougerol & Picard (1992), Chen & An (1998) and Giraitis et al. (2000). Note that (2) is not a necessary condition since a strictly stationary process may have an infinite the second moment. Conditions 1 – 3 were employed by Hall & Yao (2003).

For simplicity and clarity, we shall first consider the estimators defined with the complete conditional variance function; i.e. we let $\nu = p$ and employ σ_t^2, instead of $\widetilde{\sigma}_t^2$, in the definitions of the estimators $\widehat{\theta}_1, \widehat{\theta}_2$ and $\widehat{\theta}_3$, so that, insofar as calculation of σ_t^2 $(1 \leq t \leq n)$ is concerned, we may use values of X_u for $-\infty < u \leq n$. The estimators defined in terms of the truncated variance $\widetilde{\sigma}_t^2$ will be dealt with in §4.2 below. There we show that our main result does not change when the truncated approximation is employed as long as $\nu \to \infty$ at a proper rate.

THEOREM 1. *Under Conditions 1 – 4, there exists a local minimiser $\widehat{\theta}_2$ within radius η of θ^0 for which*

$$n^{1/2}(\widehat{\theta}_2 - \theta^0) \to N[0,\ \Sigma^{-1}/\{4f(0)^2\}]$$

in distribution, as $n \to \infty$, where $\eta > 0$ is a sufficiently small but fixed constant.

Remark 3. The above theorem indicates that the least absolute deviations estimator $\widehat{\theta}_2$ is asymptotically normal with convergence rate $n^{1/2}$ under very mild conditions. In particular, the tail-weight of the distribution of ε_t is irrelevant as we have imposed no condition on the moments of ε_t beyond $E(\varepsilon_t^2) < \infty$. In contrast, the asymptotic normality for the Gaussian maximum likelihood estimator $\widehat{\theta}_{\mathrm{ml}}$ is only possible if $E(|\varepsilon_t|^{4-\delta}) < \infty$ for any $\delta > 0$, and furthermore the convergence rate $n^{1/2}$ is only observable when $E(\varepsilon_t^4) < \infty$; see Hall & Yao (2003).

Remark 4. Similarly to Theorem 1, $\sqrt{n}(\widehat{\theta}_1 - \theta^0)$ is also asymptotically normal with mean

$$E\{\varepsilon_t^2 I(\varepsilon_t^2 > 1) - \varepsilon_t^2 I(\varepsilon_t^2 < 1)\}\{E|\sigma_{11}|, \cdots, E|\sigma_{(1+p+q)(1+p+q)}|\}^{\mathrm{T}},$$

which is unlikely to be 0. This shows that $\widehat{\theta}_1$ is often a biased estimator.

Remark 5. It may be shown that $\sqrt{n}(\widehat{\theta}_3 - \theta^0)$ is also asymptotically normal under the additional condition $EX_t^4 < \infty$. The latter will also ensure that the maximum quasilikelihood estimator $\widehat{\theta}_{\mathrm{ml}}$ converges to normality in distribution at the standard rate $n^{1/2}$.

4·2. *A central limit theorem with truncated conditional variances*

In practice we may only employ $\widetilde{\sigma}_t^2$ rather than σ_t^2. For small t the accuracy of this approximation is severely curtailed, suggesting that when conducting inference we should avoid early terms in the series; that is, we require that the integer $\nu = \nu(n)$ diverges with n but at a rate sufficiently slow to ensure that $\nu/n \to 0$ as $n \to \infty$. Theorem 2 below shows that, for an appropriate choice of ν, Theorem 1 continues to hold.

THEOREM 2. *Let $\nu/\log n \to \infty$ and $\nu/n \to 0$ as $n \to \infty$. Then Theorem 1 holds with $\widehat{\theta}_2$ defined in (9).*

Acknowledgement

We are grateful to the editor and a reviewer for helpful comments. Q. Yao's work was partially supported by the U.K. Engineering and Physical Sciences Research Council.

Appendix

Proofs

Proof of Theorem 1. Put $Z_t(\theta) = \log X_t^2 - \log \sigma_t(\theta)^2$. For any $v = (v_0, \cdots, v_{p+q})^{\mathrm{T}} \in R^{p+q+1}$, let

$$S_n(v) = \sum_{t=1+p}^{n} \{|Z_t(\theta^0 + n^{-1/2}v)| - |Z_t(\theta^0)|\},$$

$$S_n^*(v) = \sum_{t=1+p}^{n} \{|Z_t(\theta^0) - n^{-1/2}v^{\mathrm{T}} U_t(\theta^0)| - |Z_t(\theta^0)|\},$$

where $U_t(\theta) = \{A_{t0}(\theta), \cdots, A_{t,p+q}(\theta)\}^{\mathrm{T}}$ is defined in §4.1. It holds that, for $z \neq 0$,

$$|z-y| - |z| = -y\,\mathrm{sgn}(z) + 2(y-z)\{I(0<z<y) - I(y<z<0)\}.$$

Hence,

$$\begin{aligned} S_n^*(v) &= -n^{-1/2} \sum_{t=1+p}^{n} \Big\{ \sum_{l=0}^{p+q} A_{tl}(\theta^0) v_l \Big\} \,\mathrm{sgn}\{Z_t(\theta^0)\} \\ &+ 2 \sum_{t=1+p}^{n} \Big\{ \sum_{l=0}^{p+q} A_{tl}(\theta^0) n^{-1/2} v_l - Z_t(\theta^0) \Big\} I\Big\{ 0 < Z_t(\theta^0) < \sum_{l=0}^{p+q} A_{tl}(\theta^0) n^{-1/2} v_l \Big\} \\ &- 2 \sum_{t=p+1}^{n} \Big\{ \sum_{l=0}^{p+q} A_{tl}(\theta^0) n^{-1/2} v_l - Z_t(\theta^0) \Big\} I\Big\{ \sum_{l=0}^{p+q} A_{tl}(\theta^0) n^{-1/2} v_l < Z_t(\theta^0) < 0 \Big\}. \end{aligned}$$

Write the three terms on the right-hand side of the above expression as I_1, I_2 and I_3 respectively.

Let $\mathcal{F}_t = \sigma(\varepsilon_s, s \leq t)$. Then $[\sum_{l=0}^{p+q} A_{tl}(\theta^0) v_l \mathrm{sgn}\{Z_t(\theta^0)\},\ t \geq p+1]$ is a martingale difference sequence. It can be shown that $E_0\{A_{ti}(\theta)\}^u < \infty$ for any $u > 0$, $i = 0, \cdots, p+q$ and θ within radius η of θ^0, where $\eta > 0$ is a sufficiently small but fixed constant. Consequently, we may show that $I_1 \to N(0, v^{\mathrm{T}} \Sigma v)$ in distribution. Let

$$W_{nt} = \Big\{ \sum_{l=0}^{p+q} A_{tl}(\theta^0) n^{-1/2} v_l - Z_t(\theta^0) \Big\} I\Big\{ 0 < Z_t(\theta^0) < \sum_{l=0}^{p+q} A_{tl}(\theta^0) n^{-1/2} v_l \Big\},$$

and let F and G denote the distribution functions of $\log(\varepsilon_t^2)$ and $B_t = \sum_{l=0}^{p+q} A_{tl}(\theta^0) v_l$, respectively. Then

$$\begin{aligned} \limsup_{n\to\infty} nE\big(W_{nt}^2\big) &= \limsup_{n\to\infty} \Big\{ n \int_0^{\epsilon n^{1/2}} \int_0^{n^{-1/2}y} (n^{-1/2}y - z)^2 \, dF(z) dG(y) \\ &+ n \int_{\epsilon n^{1/2}}^{\infty} \int_0^{n^{-1/2}y} (n^{-1/2}y - z)^2 \, dF(z) dG(y) \Big\} \\ &\leq \limsup_{n\to\infty} \Big[n \int_0^{\epsilon n^{1/2}} \int_0^{n^{-1/2}y} (n^{-1/2}y - z)^2 \{f(0) + \delta\} \, dz dG(y) \\ &+ n \int_{\epsilon n^{1/2}}^{\infty} n^{-1} y^2 \, dG(y) \Big] \\ &= O\Big\{ \limsup_{n\to\infty} n \int_0^{\epsilon n^{1/2}} n^{-3/2} y^3 \, dG(y) \Big\} \\ &= O\big[\epsilon E\{B_1^2 I(B_1 > 0)\}\big], \end{aligned}$$

which converges to 0 as $\epsilon \to 0$. We may show that

$$E(W_{nt} | \mathcal{F}_{t-1}) \simeq \frac{1}{2} n^{-1} B_t^2 f(0) I(B_t > 0),$$

see Davis and Dunsmuir (1997). Hence

$$\sum_{t=p+1}^{n} E(W_{nt}|\mathcal{F}_{t-1}) \to \frac{f(0)}{2} E\{ B_1^2 I(B_1 > 0)\},$$

in probability. Since

$$\mathrm{var}\Big[\sum_{t=p+1}^{n} \{W_{nt} - E(W_{nt}|\mathcal{F}_{t-1})\}\Big] = \sum_{t=p+1}^{n} \mathrm{var}\{W_{nt} - E(W_{nt}|\mathcal{F}_{t-1})\} \leq \sum_{t=p+1}^{n} EW_{nt}^2 \to 0,$$

we have that

$$\sum_{t=p+1}^{n} W_{nt} \to \frac{f(0)}{2} E\{ B_1^2 I(B_1 > 0)\},$$

in probability. Therefore we could show that $I_2 + I_3 \to f(0)EB_1^2$, in probability. Thus

$$S_n^*(v) \to f(0) v^{\mathrm{T}} \Sigma v + v^{\mathrm{T}} \xi,$$

in distribution, uniformly on any compact set in R^{1+p+q}, where $\xi \sim N(0, \Sigma)$. Now put $D = \frac{\partial^2}{\partial\theta\partial\theta^{\mathrm{T}}} \log \sigma_t^2(\theta)$. Then it is easy to see that $D = \frac{\partial}{\partial\theta^{\mathrm{T}}} U_t(\theta)$. Note that, for $1 \leq j, l \leq q$,

$$\begin{aligned} \frac{\partial}{\partial a_l} A_{t,p+j}(\theta) &= -A_{t,p+j}(\theta) A_{t,p+l}(\theta) + \frac{1}{\sigma_t(\theta)^2} \Big\{ \frac{2c}{(1 - \sum_{j=1}^{q} a_j)^3} + 2 \sum_{i=1}^{p} b_i X_{t-i-j-l}^2 \\ &+ \sum_{i=1}^{p} b_i \sum_{k=1}^{\infty} (k+2)(k+1) \sum_{j_1=1}^{q} \cdots \sum_{j_k=1}^{q} a_{j_1} \cdots a_{j_k} X_{t-i-j-l-j_1-\cdots-j_k}^2 \Big\}. \end{aligned}$$

We may show that Conditions 1 & 2 imply that $E\{X_{t-i-j-l}^2/\sigma_t(\theta)^2\} < \infty$ and

$$E\{\frac{1}{\sigma_t(\theta)^2} \sum_{i=1}^{p} b_i \sum_{k=1}^{\infty} (k+2)(k+1) \sum_{j_1=1}^{q} \cdots \sum_{j_k=1}^{q} a_{j_1} \cdots a_{j_k} X_{t-i-j-l-j_1-\cdots-j_k}^2\} < \infty$$

for $i = 1, \cdots, p$ and $j, l = 1, \cdots, q$. We therefore have $E\{\frac{\partial}{\partial a_l} A_{t,p+j}(\theta)\} < \infty$ for any θ within radius η of θ^0. Similarly we could show that the expectation of every element in D is finite, i.e. $E(v^{\mathrm{T}} D v) < \infty$, for such a θ. As in Davis & Dunsmuir (1997), we further have that

$$S_n(v) \to f(0) v^{\mathrm{T}} \Sigma v + v^{\mathrm{T}} \xi,$$

in distribution. Hence the required central limit theorem follows from Lemma 2.2 and Remark 1 of Davis et al. (1992). This completes the proof of Theorem 1.

Proof of Theorem 2. From the proof of Theorem 1 and the fact that $\nu/n \to 0$ as $n \to \infty$, it suffices to show that

$$\sup_{\theta \in \mathcal{N}} \sum_{t=1+\nu}^{n} \Big| \log \frac{\widetilde{\sigma}_t(\theta)^2}{\sigma_t(\theta)^2} \Big| = o_p(1),$$

where $\mathcal{N}$ denotes a sufficiently small, but fixed, open neighbourhood of the true parameter value θ^0. We therefore only need to show that

$$\sup_{\theta\in\mathcal{N}} \sum_{t=1+\nu}^{n} \sum_{i=1}^{p} b_i \sum_{k=1}^{\infty} \sum_{j_1=1}^{q} \cdots \sum_{j_k=1}^{q} a_{j_1} \cdots a_{j_k} X^2_{t-i-j_1-\cdots-j_k} I(t-i-j_1-\cdots-j_k<1) = o_p(1).$$

This is true because $\nu/\log n \to \infty$, $\sum_{j=1}^q a_j < 1$, $E(X_t^2) < \infty$ and the fact that for any $\delta > 0$

$$\sup_{1\le i\le p,\, t\ge 1+\nu} \mathrm{pr}\Big\{ \sum_{k=1}^{\infty} \sum_{j_1=1}^{q} \cdots \sum_{j_k=1}^{q} a_{j_1} \cdots a_{j_k} X^2_{t-i-j_1-\cdots-j_k} I(t-i-j_1-\cdots-j_k<1) > \delta \Big\}$$
$$\le \delta^{-1} \sum_{k=[\nu/q]-p}^{\infty} \sum_{j_1=1}^{q} \cdots \sum_{j_k=1}^{q} a_{j_1} \cdots a_{j_k} E(X_1^2) = \delta^{-1} E(X_1^2) \sum_{k=[\nu/q]-p}^{\infty} \Big(\sum_{j=1}^{q} a_j \Big)^k.$$

The proof is completed.

References

Adler, R.J., Feldman, R.E. & Gallagher, C. (1997). Analysing stable time series. In *A User's Guide to Heavy Tails: Statistical Techniques For Analysing Heavy Tailed Distributions and Processes*, Ed. R.J. Adler, R.E. Feldman and M. Taqqu, pp.133-58. Boston: Birkhäuser.

Bollerslev, T. (1986). Generalised autoregressive conditional heteroscedasticity. *J. Economet.* **31**, 307-27.

Bougerol, P. & Picard, N. (1992). Strict stationarity of generalized autoregressive processes. *Ann. Prob.* **20**, 1714-30.

Chen, M. & An, H.Z. (1998). A note on the stationarity and the existence of moments of the GARCH model. *Statistica Sinica* **8**, 505-10.

Davis, R.A. & Dunsmuir, W.T.M. (1997). Least absolute deviation estimation for regression with ARMA errors. *J. Theor. Prob.* **10**, 481-97.

Davis, R.A., Knight, K. & Liu, J. (1992). M-estimation for autoregressions with infinite variances. *Stoch. Proces. Applic.* **40**, 145-80.

Engle, R.F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of U.K. inflation. *Econometrica* **50**, 987-1008.

Fan, J. & Yao, Q. (2003) *Nonlinear Time Series: Nonparametric and Parametric Methods.* New York: Springer.

Giraitis, L., Kokoszka, P. & Leipus, R. (2000). Stationary ARCH models: dependence structure and central limit theorem. *Economet. Theory* **16**, 3-22.

Giraitis, L. & Robinson, P.M. (2001). Whittle estimation of ARCH models. *Economet. Theory* **17**, 608-23.

HALL, P. & YAO, Q. (2003). Inference in ARCH and GARCH models with heavy-tailed errors. *Econometrica* **71**, 285-317.

KESTEN, H. (1973). Random difference equations and renewal theory for products of random matrices. *Acta Math.* **131**, 207-48.

LEE, A.W. & HANSEN, B.E. (1994). Asymptotic theory for GARCH(1,1) quasi-maximum likelihood estimator. *Economet. Theory* **10**, 29-52.

LUMSDAINE, R. (1996). Consistency and asymptotic normality of the quasi-maximum likelihood estimator for IGARCH(1,1) and covariance stationary GARCH(1,1) models. *Econometrica* **16**, 575-96.

MITTNIK, S. & RACHEV, S.T. (2000). *Stable Paretian Models in Finance.* New York: Wiley.

MITTNIK, S., RACHEV, S.T. & PAOLELLA, M.S. (1988). Stable Paretian modeling in finance: some empirical and theoretical aspects. In *A Practical Guide to Heavy Tails*, Ed. R.J. Adler, R.E. Feldman and M.S. Taqqu, pp.79–110. Boston: Birkhäuser.

RYDBERG, T. (2000). Realistic statistical modelling of financial data. *Int. Statist. Rev.* **68**, 233-58.

SHEPHARD, N. (1996). Statistical aspects of ARCH and stochastic volatility. In *Time Series Models in Econometrics, Finance and Other Fields*, Ed. D.R. Cox, D.V. Hinkley and O.E. Barndorff-Nielsen, pp.1-67. London: Chapman & Hall.

TAYLOR, S.J. (1986). *Modelling Financial Time Series.* New York: Wiley.

WEISS, A. (1986). Asymptotic theory for ARCH models: estimation and testing. *Economet. Theory* **2**, 107-31.

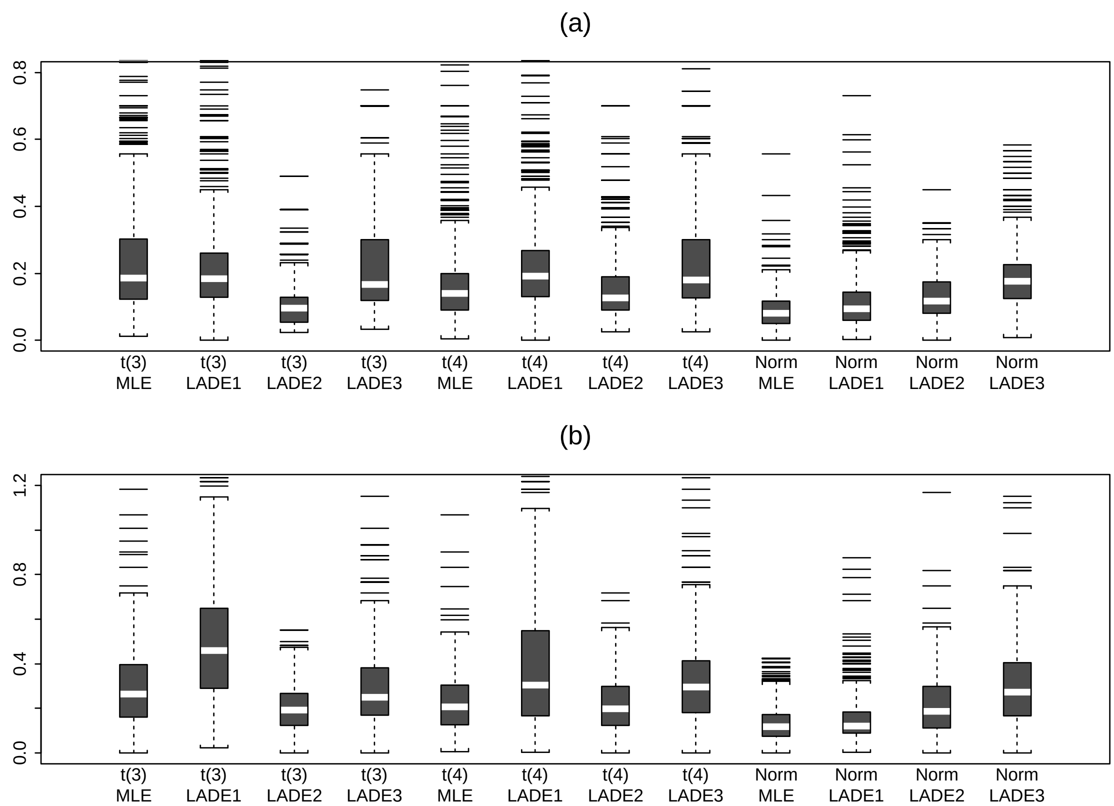

Figure 1: *Boxplots of the average absolute errors of the maximum likelihood estimates (MLE) and the three least absolute deviations estimates: LADE1 –* $\widehat{\theta}_1$, *LADE2 –* $\widehat{\theta}_2$ *and LADE3 –* $\widehat{\theta}_3$. *Labels* $t(3)$, $t(4)$ *and Norm indicate that error* ε_t *has, respectively, t-distributions with 3 and 4 degrees of freedom, and a normal distribution. (a) ARCH(2), (b) GARCH(1,1).*