Least Absolute Deviations Estimation for ARCH and GARCH Models

Liang Peng
School of Mathematics
Georgia Institute of Technology
Atlanta, GA 30332-0160
Email: peng@math.gatech.edu

Qiwei Yao
Department of Statistics
London School of Economics
London, WC2A 2AE, UK
Email: q.yao@lse.ac.uk

Abstract

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 and Yao (2001) showed that for ARCH/GARCH models with heavy-tailed errors, the conventional quasi-maximum likelihood estimator suffers from complex limit distributions and slow convergence rates. In this paper three types of absolute deviations estimators have been examined, and the one based on logarithmic transform is particularly appealing. We have shown that the estimator is asymptotically normal and unbiased. Further it enjoys the standard convergence rate $n^{1/2}$ regardless whether the errors are heavy-tailed or not. Simulation lends further support to our theoretical results.

Key words and phrases: ARCH, asymptotic normality, GARCH, Gaussian likelihood, heavy tails, least absolute deviations estimator, quasi-maximum likelihood estimator, time series.

1 Introduction

Motivated by the need to explain and to forecast risk in financial time series, ARCH and GARCH models were proposed to model explicitly the conditional second moments; see Engle (1982), Bollerslev (1986) and Taylor (1986). Early successes of ARCH/GARCH application were confined

to the case of normal errors. On the other hand, empirical evidence suggests that financial data may have heavy tails (Mittnik, Rachev and Paolella, 1988; Mittnik and Rachev, 2001) and the models with heavy-tailed errors have also been adopted in practice. An excellent survey on ARCH/GARCH modelling for financial data is available in Shephard (1996) and Rydberg (2000). For their theoretical properties, we refer to section 4.2 of Fan and Yao (2002).

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 might be called quasi-maximum likelihood estimators. The asymptotic properties of quasi-maximum likelihood estimators are established for ARCH(p) models by Weiss (1986), for GARCH(1,1) models by Lee and Hansen (1994) and Lumsdaine (1996), and for general GARCH(p, q) models by Hall and Yao (2001). In fact Hall and Yao (2001) showed that when the error distribution is heavy-tailed the estimators are no longer asymptotic normal, the range of possible limit distribution is extraordinarily large, and the convergence rate is slower than the standard rate $n^{1/2}$. Complex asymptotic properties were also observed from a Whittle estimator (Giraitis and Robinson 2001) for heavy tailed GARCH(1,1) models by Mikosch and Straumann (2001).

Note that quasi-maximum likelihood estimation based on 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 most robust. See, for example, Davis, Knight and Liu (1992), Adler, Feldman and Gallagher (1997) and the references within. In this paper, we explore three types of least absolute deviations estimators for ARCH and GARCH models and advocate the one based on logarithmic transform. Our theoretical result shows that this estimator is asymptotically normal and unbiased. Further it enjoys the $n^{1/2}$ convergence rate regardless the tail-weight of error distributions (see Remark 3 in section 4 below). This is in marked contrast to the conventional Gaussian maximum likelihood estimator. Our simulation lends further support to our theoretical results.

The paper is organised as follows. The three absolute deviations estimators are defined in section 2, and their asymptotic properties are presented in section 4. We report some numerical results in section 3. Technical proofs are relegated to the Appendix.

2 Models and estimators

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

$$X_t = \sigma_t \varepsilon_t, \quad \text{and} \quad \sigma_t^2 \equiv \sigma_t(\boldsymbol{\theta})^2 = c + \sum_{i=1}^{p} b_i X_{t-i}^2 + \sum_{j=1}^{q} a_j \sigma_{t-j}^2, \tag{2.1}$$

where $c > 0$, $b_j \geq 0$ and $a_j \geq 0$ are unknown parameters, $\boldsymbol{\theta} = (c, b_1, \cdots, b_p, a_1, \cdots, a_q)^\tau$, $\{\varepsilon_t\}$ 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. The necessary and sufficient condition for (2.1) defining a unique strictly stationary process $\{X_t,\, t = 0, \pm 1, \pm 2, \cdots\}$ with $EX_t^2 < \infty$ is

$$\sum_{i=1}^{p} b_i + \sum_{j=1}^{q} a_j < 1. \tag{2.2}$$

Further for such a stationary solution, $EX_t = 0$ and $\text{Var}(X_t) = c/\{1 - \sum_{i=1}^p b_i - \sum_{j=1}^q a_j\}$. See Giraitis, Kokoszka and Leipus (2000), and also Theorem 4.4 of Fan and Yao (2002).

The most popular method for estimating GARCH models is arguably the quasi-maximum likelihood estimation based on a (conditional) Gaussian likelihood. It 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

$$\left\{ \prod_{t=\nu+1}^{n} \sigma_t(\boldsymbol{\theta})^2 \right\}^{-1/2} \exp\left\{ -\frac{1}{2} \sum_{t=\nu+1}^{n} \frac{X_t^2}{\sigma_t(\boldsymbol{\theta})^2} \right\}. \tag{2.3}$$

Under condition (2.2), σ_t^2 may be expressed as

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

where the multiple sums vanishes if $q = 0$; see Hall and Yao (2001). This leads to the following approximation for σ_t^2 based on $X_1, \cdots, X_t$

$$\begin{aligned} \widetilde{\sigma}_t(\boldsymbol{\theta})^2 &= \frac{c}{1 - \sum_{j=1}^q a_j} + \sum_{i=1}^{\min(p,t-1)} b_i X_{t-i}^2 + \sum_{i=1}^{p} b_i \sum_{k=1}^{\infty} \sum_{j_1=1}^{q} \cdots \sum_{j_k=1}^{q} a_{j_1} \cdots a_{j_k} \\ &\times \quad X_{t-i-j_1-\cdots-j_k}^2 I(t - i - j_1 - \cdots - j_k \geq 1). \end{aligned} \tag{2.4}$$

Maximising (2.3) with σ_t^2 replaced by $\widetilde{\sigma}_t^2$, we obtain the quasi-maximum likelihood estimator

$$\widehat{\boldsymbol{\theta}}_{ml} = \arg\min_{\boldsymbol{\theta}} \sum_{t=\nu+1}^{n} \left\{ \frac{X_t^2}{\widetilde{\sigma}_t(\boldsymbol{\theta})^2} + \log\{\widetilde{\sigma}_t(\boldsymbol{\theta})^2\} \right\},$$

where the minimisation is taken over all the non-negative values of parameters. The asymptotic properties of the estimator $\widehat{\boldsymbol{\theta}}_{ml}$ were derived in Hall and Yao (2001). In particular, when the distribution of ε_t is heavy tailed in the sense that $E(|\varepsilon_t|^d) = \infty$ for some $2 < d \le 4$, the convergence rate of $\widehat{\boldsymbol{\theta}}_{ml}$ is slower than the standard rate $n^{1/2}$.

If we reparametrise the model (2.1) in such a way that $E\varepsilon_t = 0$ and the median of ε_t^2 is equal to 1, the parameters c, b_i's and a_j's differ from those in the old setting by a common positive constant factor. Further, $E\left|X_t^2/\sigma_t^2 - 1\right|$ obtains the minimum value at the new true values of the parameters. This leads to an absolute deviation estimator

$$\widehat{\boldsymbol{\theta}}_1 = \arg\min_{\boldsymbol{\theta}} \sum_{t=\nu+1}^{n} |X_t^2/\widetilde{\sigma}_t(\boldsymbol{\theta})^2 - 1|. \tag{2.5}$$

Although the idea behind the above estimation is simple, the estimator $\widehat{\boldsymbol{\theta}}_1$ is, unfortunately, asymptotically biased; see Remark 4 in section 4 below.

To overcome this shortcoming, we define a modified form of least absolute deviations estimator as follows

$$\widehat{\boldsymbol{\theta}}_2 = \arg\min_{\boldsymbol{\theta}} \sum_{t=\nu+1}^{n} |\log(X_t^2) - \log\{\widetilde{\sigma}_t(\boldsymbol{\theta})^2\}|. \tag{2.6}$$

Note that the distribution of X_t^2 is confined to the non-negative half axis and is typically skewed. Intuitively the log-transform will make the distribution less skewed. Theorem 1 below shows that the estimator $\widehat{\boldsymbol{\theta}}_2$ is in fact asymptotically normal and unbiased under very mild conditions. The estimator $\widehat{\boldsymbol{\theta}}_2$ is motivated by the regression relationship

$$\log(X_t^2) = \log\{\sigma_t(\boldsymbol{\theta})\}^2 + \log(\varepsilon_t^2), \tag{2.7}$$

in which the median of $\log(\varepsilon_t^2)$ is equal to 0 under the reparametrisation.

It we write

$$X_t^2 = \sigma_t^2 + e_t, \tag{2.8}$$

where $e_t = \sigma_t^2(\varepsilon_t^2 - 1)$ (for all t) form a sequence of martingale differences; see (2.1). Again under the new parametrisation, the median of e_t is 0. This leads to our third least absolute deviations estimator

$$\widehat{\boldsymbol{\theta}}_3 = \arg\min_{\boldsymbol{\theta}} \sum_{t=\nu+1}^{n} |X_t^2 - \widetilde{\sigma}_t(\boldsymbol{\theta})^2|. \tag{2.9}$$

Intuitively we prefer the estimator $\widehat{\boldsymbol{\theta}}_2$ to $\widehat{\boldsymbol{\theta}}_3$ since the error terms $\log(\varepsilon_t^2)$ in regression model (2.7) are independent and identically distributed while the errors e_t in model (2.8) are not independent.

Therefore ideally the sum on the RHS of (2.9) should be replaced by a weighted sum with weights reflecting the dependence, which is typically intractable. In fact the asymptotic normality of $\widehat{\boldsymbol{\theta}}_3$ requires more conditions; see Remark 5 in section 4.

Remark 1. The minimisation in (2.5), (2.6) and (2.9) 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 (2.4) that $\widetilde{\sigma}_t^2 \equiv \sigma_t^2$ for all $t > p$. Thus we may let $\nu = p$ in the definitions of the above estimators.

3 Numerical properties

In this section, we compare numerically the three least absolute deviations estimators (LADE) with the conditional Gaussian maximum likelihood estimator (MLE) for an ARCH(2) and a GARCH(1,1) models. In both cases we took the errors ε_t to have either standard normal distribution or standardised Student's t distribution with d degrees of freedom with $d = 4$ or 3. Note that when $\varepsilon_t \sim t(d)$, $E|\varepsilon_t|^d = \infty$. We employed $c = 3$, $b_1 = 0.5$ 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 exhausting search to find estimates. Note that the definition of least absolute deviations estimators implies a reparametrisation. We define the average absolute error (AAE) as $(|\widehat{b}_1/\widehat{c} - b_1/c| + |\widehat{b}_2/\widehat{c} - b_2/c|)/2$ for ARCH(2), $(|\widehat{b}_1/\widehat{c} - b_1/c| + |\widehat{a}_1/\widehat{c} - a_1/c|)/2$ for GARCH(1,1).

Figure 1 presents the boxplots for the AAEs. For models with heavy-tailed errors (i.e. $\varepsilon_t \sim t_d$ with $d = 3, 4$), the least absolute deviation estimator $\widehat{\boldsymbol{\theta}}_2$ performed best. Further the gain from using $\widehat{\boldsymbol{\theta}}_2$ was more pronounced when the tails are very heavy (i.e. $\varepsilon_t \sim t_3$). Note that when $\varepsilon_t \sim t_4$, Gaussian MLE $\widehat{\boldsymbol{\theta}}_{ml}$ was almost as good as $\widehat{\boldsymbol{\theta}}_2$, and was better than both $\widehat{\boldsymbol{\theta}}_1$ and $\widehat{\boldsymbol{\theta}}_3$. However when $\varepsilon_t \sim t_3$, $\widehat{\boldsymbol{\theta}}_{ml}$ was no longer desirable. On the other hand, when the error ε_t is normal, $\widehat{\boldsymbol{\theta}}_{ml}$ was of course the best. In fact the AAE of $\widehat{\boldsymbol{\theta}}_{ml}$ is larger when the tail of the error distribution is heavier, which reflects the fact that the heavier the tails are, the slower the convergence rate is. (See Hall and Yao 2001.) However this is not always the case for the LADEs as they are more robust against heavy tails.

The above pattern were also observed in the simulation with other models. In general, our numerical results suggest that we should use the least absolute deviations estimator $\widehat{\boldsymbol{\theta}}_2$ when ε_t has heavy and especially very heavy tails (i.e. $E(|\varepsilon_t|^3) = \infty$), while in general the quasi-maximum

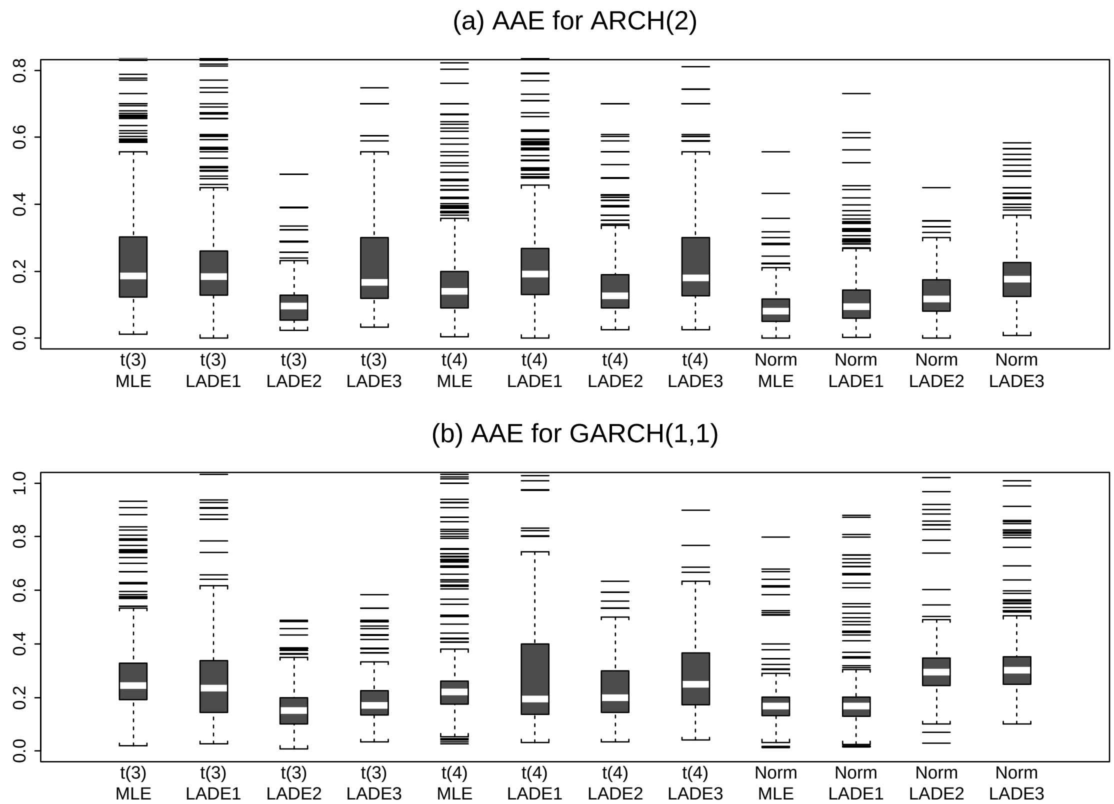

Figure 1: *Boxplots of the average absolute errors of the maximum likelihood estimates (MLE) and the three least absolute deviations estimates: LADE1 –* $\widehat{\boldsymbol{\theta}}_1$*, LADE2 –* $\widehat{\boldsymbol{\theta}}_2$ *and LADE3 –* $\widehat{\boldsymbol{\theta}}_3$.

likelihood estimator $\widehat{\boldsymbol{\theta}}_{ml}$ is desirable as long as that ε_t is not very heavy-tailed.

4 Asymptotic properties

In this section, we justify the superior performance of $\widehat{\boldsymbol{\theta}}_2$ over $\widehat{\boldsymbol{\theta}}_1$ and $\widehat{\boldsymbol{\theta}}_3$ for heavy tailed models in terms of asymptotic theory. We establish the asymptotic normality of the estimator $\widehat{\boldsymbol{\theta}}_2$. The properties of both $\widehat{\boldsymbol{\theta}}_1$ and $\widehat{\boldsymbol{\theta}}_3$ will be stated without proofs. To minimise the technical argument, we prove the result for ARCH models only although similar results hold for general GARCH models. Hence the second equation in (2.1) is of the form

$$\sigma_t^2 = c + \sum_{j=1}^{p} b_j X_{t-j}^2,$$

where $c > 0$ and $b_j \geq 0$ $(j = 1, \cdots, p)$. Let $\boldsymbol{\theta}^0 = (c^0, b_1^0, \cdots, b_p^0)^\tau$ be the true value under which the median of $\log(\varepsilon_t^2)$ equals 0.

Define $\sigma_{11}^2 = 1/(c^0 + \sum_{i=1}^p b_i^0 X_{p+1-i}^2)^2$, and for $2 \le i, j \le p+1$,

$$\sigma_{1i}^2 = \frac{X_{p+2-i}^2}{(c^0 + \sum_{l=1}^p b_l^0 X_{p+1-l}^2)^2}, \qquad \sigma_{ij}^2 = \frac{X_{p+2-i}^2 X_{p+2-j}^2}{(c^0 + \sum_{l=1}^p b_l^0 X_{p+1-l}^2)^2}.$$

Let $\mathbf{\Sigma}$ be the $(p+1) \times (p+1)$ symmetric matrix with $E(\sigma_{ij}^2)$ as its (i,j)-th element. Some regularity conditions are now in order.

A1. There exists a unique strictly stationary solution of model (2.1).

A2. There exists a positive number δ such that $E(\sigma_{ij}^2)^{1+\delta} < \infty$ for all $1 \le i, j \le p+1$.

A3. $\mathbf{\Sigma}$ is nonsingular.

A4. $\log \varepsilon_t^2$ has a unique median zero, and its density function f is continuous at zero.

Remark 2. Different conditions to ensure the existence of a strictly stationary solution of (2.1) have been established by, for example, Kesten (1973), Bougerol and Picard (1992), and Giraitis, Kokoszka and Leipus (2000). Note that (2.2) is not a necessary condition for A1. Further, condition A2 is implied by the assumption that none of $b_1, \cdots, b_p$ vanishes, which, together with condition A3, were imposed in Hall and Yao (2001).

Theorem 1. Under conditions A1 - A4, $n^{1/2}(\widehat{\boldsymbol{\theta}}_2 - \boldsymbol{\theta}^0)$ is asymptotically normal with mean 0 and variance $\mathbf{\Sigma}^{-1}/\{4f(0)^2\}$.

Remark 3. By Theorem 1, the least absolute deviations estimator $\widehat{\boldsymbol{\theta}}_2$ is asymptotically normal with the convergence rate $n^{1/2}$ under very mild conditions. For example we only require $E(\varepsilon_t^2) < \infty$. In contrast, the asymptotic normality for the Gaussian maximum likelihood estimator $\widehat{\boldsymbol{\theta}}_{ml}$ 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 and Yao (2001).

Remark 4. The asymptotic normality stated in Theorem 1 also holds for the estimator $\widehat{\boldsymbol{\theta}}_1$ except the limit distribution has the mean

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

instead of 0. Therefore $\widehat{\boldsymbol{\theta}}_1$ is asymptotically biased.

Remark 5. Theorem 1 would also hold for the estimator $\widehat{\boldsymbol{\theta}}_3$ if we replace σ_{11}^2 by 1, σ_{1i}^2 by X_{p+2-i}^2, and σ_{ij}^2 by $X_{p+2-i}^2 X_{p+2-j}^2$. But now condition A2 implies $E\varepsilon_t^4 < \infty$, under which the quasi-maximum likelihood estimator $\widehat{\boldsymbol{\theta}}_{ml}$ is also asymptotically normal with the convergence rate $n^{1/2}$.

Appendix: Proof of Theorem 1.

Put $Z_t(\boldsymbol{\theta}) = \log X_t^2 - \log(c + \sum_{i=1}^p b_i X_{t-i}^2)$, and

$$A_{t0}(\boldsymbol{\theta}) = \frac{1}{c + \sum_{l=1}^p b_l X_{t-l}^2}, \qquad A_{ti}(\boldsymbol{\theta}) = \frac{X_{t-i}^2}{c + \sum_{l=1}^p b_l X_{t-l}^2} \quad i = 1, \cdots, p,$$

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

where $\mathbf{v} = (v_0, \cdots, v_p)^\tau$. It holds that for $z \neq 0$,

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

Hence,

$$\begin{aligned} S_n(\mathbf{v}) \; &= -n^{-1/2} \textstyle\sum_{t=p+1}^n \{\sum_{l=0}^p A_{tl}(\boldsymbol{\theta}^0) v_l\} \operatorname{sgn}\{Z_t(\boldsymbol{\theta}^0)\} \\ &+2 \textstyle\sum_{t=p+1}^n \{\sum_{l=0}^p A_{tl}(\boldsymbol{\theta}^0) n^{-1/2} v_l - Z_t(\boldsymbol{\theta}^0)\} I\{0 < Z_t(\boldsymbol{\theta}^0) < \sum_{l=0}^p A_{tl}(\boldsymbol{\theta}^0) n^{-1/2} v_l\} \\ &-2 \textstyle\sum_{t=p+1}^n \{\sum_{l=0}^p A_{tl}(\boldsymbol{\theta}^0) n^{-1/2} v_l - Z_t(\boldsymbol{\theta}^0)\} I\{\sum_{l=0}^p A_{tl}(\boldsymbol{\theta}^0) n^{-1/2} v_l < Z_t(\boldsymbol{\theta}^0) < 0\}. \end{aligned}$$

Write the three terms on the RHS of the above expression as I_1, I_2 and I_3 respectively.

Let $\mathcal{F}_t = \sigma(\varepsilon_s, s \leq t)$. Note that $[\sum_{l=0}^p A_{tl}(\boldsymbol{\theta}^0) v_l \operatorname{sgn}\{Z_t(\boldsymbol{\theta}^0)\},\ t \geq p+1]$ is a martingale difference. By condition A2 we may show that $I_1 \xrightarrow{d} N(0, \mathbf{v}^\tau \boldsymbol{\Sigma} \mathbf{v})$. Let

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

and F and G denote the distribution function of $\log(\varepsilon_t^2)$ and the joint distribution function of $(X_p^2, \cdots, X_1^2)$, respectively. Put

$$B(x_p, \cdots, x_1) = \frac{v_0}{c^0 + \sum_{l=1}^p b_l^0 x_{p+1-l}^2} + \sum_{i=1}^p \frac{v_i x_{p+1-i}^2}{c^0 + \sum_{l=1}^p b_l^0 x_{p+1-l}^2}.$$

Then

$$\begin{aligned} & \limsup_{n\to\infty} n E W_{nt}^2 \\ = \; & \limsup_{n\to\infty} n \int_0^\infty \cdots \int_0^\infty \int_{-\infty}^\infty \{n^{-1/2} B(x_p, \cdots, x_1) - z\}^2 \\ & \times I\{0 < z < n^{-1/2} B(x_p, \cdots, x_1)\}\, dF(z) dG(x_p, \cdots, x_1) \\ \leq \; & \limsup_{n\to\infty} n^{-\delta} E\{B(X_p, \cdots, X_1)^{2(1+\delta)}\} = 0. \end{aligned}$$

It is easy to check that

$$E(W_{nt}|\mathcal{F}_{t-1}) \sim \frac{1}{2} n^{-1} B(X_{t-1}, \cdots, X_{t-p})^2 f(0) I\{B(X_{t-1}, \cdots, X_{t-p}) > 0\}.$$

Hence

$$\sum_{t=p+1}^{n} E(W_{nt}|\mathcal{F}_{t-1}) \xrightarrow{p} \frac{f(0)}{2} E[B(X_p, \cdots, X_1)^2 I\{B(X_p, \cdots, X_1) > 0\}].$$

Since

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

we have

$$\sum_{t=p+1}^{n} W_{nt} \xrightarrow{p} \frac{f(0)}{2} E[B(X_p, \cdots, X_1)^2 I\{B(X_p, \cdots, X_1) > 0\}].$$

Therefore we could show that $I_2 + I_3 \xrightarrow{p} \gamma^2 f(0)$. Thus

$$S_n(\mathbf{v}) \xrightarrow{d} f(0)(m^0)^{-2} \mathbf{v}^\tau \mathbf{\Sigma} \mathbf{v} + (m^0)^{-1} \mathbf{v} \tau \boldsymbol{\xi}$$

uniformly on any compact set in R^{p+1}. Hence the theorem follows from the same arguments as in Davis and Dunsmuir (1997).

References

Adler, R.J., Feldman, R.E. and 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. Birkhäuser, Boston.

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

Bougerol, P. and Picard, N. (1992). Strict stationarity of generalized autoregressive processes. *Ann. Probab.*, **20**, 1714 - 1730.

Davis, R.A. and Dunsmuir, W.T.M. (1997). Least absolute deviation estimation for regression with ARMA errors. *Journal of Theoretical Probability*, **10**, 481 - 497.

Davis, R.A., Knight, K. and Liu, J. (1992). M-estimation for autoregressions with infinite variances. *Stoch. Processes Appl.* **40**, 145-180.

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

Fan, J. and Yao, Q. (2002) *Nonlinear Time Series: Nonparametric and Parametric Methods.* Springer, New York.

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

Giraitis, L. and Robinson, P.M. (2001). Whittle estimation of ARCH models. *Econometric Theory*, **17**, 608-623.

Hall, P. and Yao, Q. (2001). Inference in ARCH and GARCH models with heavy-tailed errors. Revised for *Econometrica*.

Hannan, E. J. (1973). The asymptotic theory of linear time-series models. *J. Appl. Prob.* **10**, 130-145.

Kesten, H. (1973). Random difference equations and renewal theory for products of random matrices. *Acta Math.*, **131**, 207 - 248.

Lee, A.W. and Hansen, B.E. (1994). Asymptotic theory for GARCH(1,1) quasi-maximum likelihood estimator. *Econometric 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**, 575-596.

Mikosch, T. and Straumann, D. (2001). Whittle estimation in a heavy-tailed GARCH $(1,1)$ model. Manuscript.

Mittnik, S. and Rachev, S.T. (2000). Stable Paretian Models in Finance. New York: Wiley.

Mittnik, S., Rachev, S.T. and 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. *Inter. Statist. Review*, **68**, 233-258.

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. Chapman & Hall, London.

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

Weiss, A. (1986). Asymptotic theory for ARCH models: estimation and testing. *Econometric Theory*, **2**, 107-131.