Asymptotics for Quasi-Maximum Likelihood Estimators of GARCH(1,2) Model Under Dependent Innovations

Yingfu Xie and Jun Yu

Research Report
Centre of Biostochastics

Swedish University of Agricultural Sciences

Report 2003:5
ISSN 1651-8543

Asymptotics for quasi-maximum likelihood estimators of GARCH(1,2) model under dependent innovations

YINGFU XIE[1] and JUN YU

Centre of Biostochastics
Swedish University of Agricultural Sciences
SE-901 83 Umeå, Sweden

Abstract

In this paper, we investigate the asymptotic properties of the quasi-maximum likelihood estimator (quasi-MLE) for GARCH(1,2) model under stationary innovations. Consistency of the global quasi-MLE and asymptotic normality of the local quasi-MLE are obtained, which extend the previous results for GARCH(1,1) under weaker conditions.

Keywords: GARCH model, consistency, asymptotic normality, dependent error

[1]E-mail address to the correspondence author: yingfu.xie@sekon.slu.se

1 Introduction

The autoregressive conditional heteroscedastic (ARCH) model introduced by Engel (1982) has drawn much attention for its success in describing the volatility clustering, a phenomenon often happened to financial and other economic data, especially high-frequency time series, by allowing the conditional variance of current observation to depend on past innovations, leaving the unconditional variance constant. Bollerslev (1986) extended it to GARCH (Generalized ARCH), in which the conditional variance is also influenced by previous conditional variances and more flexible lag structure than ARCH is available. GARCH model has been the common tool for econometricians during the last two decades. For discussion of applications of ARCH methodology and other ARCH-type models, we refer to the paper of Bollerslev et al. (1992) and vast references therein.

On the study of the asymptotic theory of ARCH or GARCH model, Weiss (1986) firstly showed that the quasi-maximum likelihood estimator (quasi-MLE) of the ARCH model is consistent and asymptotically normal. Lumsdaine (1991) studied the asymptotic theory of GARCH(1,1) and IGARCH(1,1) (integrated GARCH) models. She initially imposed assumptions on the rescaled variable (innovation scaled by its conditional variance). But she needed a unimodal distribution and finite 32nd moment for independent and identically distributed (i.i.d.) rescaled variables. Lee and Hansen (1994), instead, assumed that the rescaled variable is dependent, namely, strictly stationary and ergodic, and provided the first consistency proof of the quasi-MLE of GARCH (1,1) model. Concerning the local consistency, they also allowed possibly integrated process or even mildly explosive GARCH processes, due to the technical assumption based on the results of Nelson (1990). For more recent theoretical results, readers are referred to the review by Li et al. (2002).

The objective of this paper is to extend the results of Lee and Hansen (1994) and investigate the asymptotic properties of quasi-MLE for GARCH(1,2) model, although they stated that their methods were valid only for GARCH(1,1) model and might be difficult to generalize. We also assume the stationarity and ergodicity of the rescaled variable z_t. However, instead of requiring uniform finiteness of the $(2+\delta)$th moment ($\delta > 0$) as Lee and Hansen (1994) did, we only need second moment condition of z_t, as well as the standard condition for second-order stationarity of GARCH model (Bollerslev, 1986), to get the consistency property of the global quasi-MLE. Asymptotic normality of the local quasi-MLE is also obtained under analogous conditions as Lee and Hansen's, mainly the fourth moment finiteness of z_t.

Section 2 will give our main results for the GARCH(1,2) model. Proofs and lemmas are given in Appendix. The motivation to this study and some discussions will be presented in Section 3.

2 Consistency and asymptotic normality

Our focus on the GARCH model is the variance structure of some stochastic process, and the mean structure is left as simple as possible. Now, suppose that the observed sequence $\{y_t\}$ is

$$y_t = \gamma_0 + \epsilon_t, \quad t = 1, 2, \ldots, n$$

with constant mean γ_0, where $E(\epsilon_t|\mathcal{F}_{t-1}) = 0$ a.s. and $\mathcal{F}_t = \sigma(\epsilon_t, \epsilon_{t-1}, \ldots)$ is the increasing σ-field containing past information up to t. Define the conditional variance $h_{0t} \triangleq E(\epsilon_t^2|\mathcal{F}_{t-1})$, and assume that ϵ_t follows a GARCH(1,2) process

$$h_{0t} = \omega_0(1 - \beta_0) + \alpha_{10}\epsilon_{t-1}^2 + \alpha_{20}\epsilon_{t-2}^2 + \beta_0 h_{0t-1} \quad a.s., \tag{1}$$

which differs from the GARCH(1,1) model in the dependence of the conditional variance h_{0t} on the square innovation ϵ_{t-2}. Further extension to a GARCH(1,q) model entails much more complex notations without essential gains. We will focus on GARCH(1,2) here. The constant term of h_{0t} process is taken as $\omega_0(1 - \beta_0)$ for convenience.

Iterate h_{0t-1} repeatedly, yielding

$$h_{0t} = \omega_0 + \alpha_{10}\epsilon_{t-1}^2 + \sum_{k=0}^{\infty}(\alpha_{20} + \alpha_{10}\beta_0)\beta_0^k \epsilon_{t-2-k}^2 \quad a.s. \tag{2}$$

The true parameter vector is $\boldsymbol{\theta}_0 = [\gamma_0, \omega_0, \alpha_{10}, \alpha_{20}, \beta_0]'$.

The model for the unknown parameters $\boldsymbol{\theta} = [\gamma, \omega, \alpha_1, \alpha_2, \beta]'$ is

$$y_t = \gamma + e_t, \quad t = 1, 2, \cdots, n.$$

The variance process is

$$h_t^*(\boldsymbol{\theta}) = \omega(1 - \beta) + \alpha_1 e_{t-1}^2 + \alpha_2 e_{t-2}^2 + \beta h_{t-1}^*(\boldsymbol{\theta}), \quad t = 3, \cdots, n$$

with the start-up conditions

$$h_1^*(\boldsymbol{\theta}) = \omega, \quad \text{and} \quad h_2^*(\boldsymbol{\theta}) = \omega + \alpha_1 e_1^2 \ ,$$

assuming ω and α_1 positive. Hence, the variance process can be written as

$$h_t^*(\boldsymbol{\theta}) = \omega + \alpha_1 e_{t-1}^2 + \sum_{k=0}^{t-3}(\alpha_2 + \alpha_1\beta)\beta^k e_{t-2-k}^2. \tag{3}$$

Define the compact parameter space

$$\begin{aligned}\Theta = \ & \{\boldsymbol{\theta} : \gamma_l \le \gamma \le \gamma_u, 0 < \omega_l \le \omega \le \omega_u, 0 < \alpha_{1l} \le \alpha_1 \le \alpha_{1u}, \\ & 0 < \alpha_{2l} \le \alpha_2 \le \alpha_{2u}, 0 < \beta_l \le \beta \le \beta_u < 1\},\end{aligned}$$

where γ_l, γ_u, ω_l, ω_u, α_{1l}, α_{1u}, α_{2l}, α_{2u}, β_l and β_u are constants; also assume $\boldsymbol{\theta}_0 \in \Theta$, which ensures nonnegative conditional variance $h_t^*(\boldsymbol{\theta})$ and a strict GARCH(1,2) model.

Define the rescaled variable $z_t = \epsilon_t / h_{0t}^{\frac{1}{2}}$, then $E(z_t|\mathcal{F}_{t-1}) = 0$ a.s. and $E(z_t^2|\mathcal{F}_{t-1}) = 1$ a.s.. Following the quasi-likelihood method, we assume that z_t is i.i.d. and standard Gaussian here to get the log quasi-likelihood function (ignoring constants)

$$L_n^*(\boldsymbol{\theta}) = \frac{1}{2n}\sum_{t=1}^{n} l_t^*(\boldsymbol{\theta}), \quad \text{where} \quad l_t^*(\boldsymbol{\theta}) = -\left(\ln h_t^*(\boldsymbol{\theta}) + \frac{e_t^2}{h_t^*(\boldsymbol{\theta})}\right).$$

Analogous to Lee and Hansen (1994), we can extend the unobserved variance into an infinite past

$$h_t(\boldsymbol{\theta}) = \omega + \alpha_1 e_{t-1}^2 + \sum_{k=0}^{\infty}(\alpha_2 + \alpha_1\beta)\beta^k e_{t-2-k}^2$$

$$h_t^{\epsilon}(\boldsymbol{\theta}) = \omega + \alpha_1 \epsilon_{t-1}^2 + \sum_{k=0}^{\infty}(\alpha_2 + \alpha_1\beta)\beta^k \epsilon_{t-2-k}^2$$

and the unobserved log-likelihood is

$$L_n(\boldsymbol{\theta}) = \frac{1}{2n}\sum_{t=1}^{n} l_t(\boldsymbol{\theta}), \quad \text{where} \quad l_t(\boldsymbol{\theta}) \triangleq -\left(\ln h_t(\boldsymbol{\theta}) + \frac{e_t^2}{h_t(\boldsymbol{\theta})}\right).$$

Because of the start-up conditions, the variance process $h_t^*(\boldsymbol{\theta})$ is not stationary. However, it is easy to show that $\sup_{\theta\in\Theta} |L_n(\boldsymbol{\theta}) - L_n^*(\boldsymbol{\theta})| \longrightarrow_p 0$ (Lemma 3). Hence its likelihood is very close to that of stationary process $h_t(\boldsymbol{\theta})$ and the non-stationarity causes no trouble. $h_t^{\epsilon}(\boldsymbol{\theta})$, which differs from $h_t(\boldsymbol{\theta})$ only in a constant scale (Lemma 1), will be instrumental in our proof.

Now, it is ready to state our assumptions and results. All limits in our paper are taken as the sample size n tends to positive infinity.

Assumption A1

(i) z_t *is strictly stationary, ergodic and* z_t^2 *is nondegenerate;*

(ii) $\alpha_{10} + \alpha_{20} + \beta_0 < 1$.

Define the global quasi-MLE

$$\hat{\boldsymbol{\theta}}_n = argmax_{\boldsymbol{\theta} \in \Theta} L_n^*(\boldsymbol{\theta}).$$

Its existence is guaranteed by compactness of the parameter space. Now the consistency property is stated as follows.

Theorem 1 *Under A1,* $\hat{\boldsymbol{\theta}}_n \longrightarrow_p \boldsymbol{\theta}_0$.

Remark 1. Assumption A1(i) is same as Lee and Hansen (1994). We don't need their assumption of uniform finiteness of the $(2+\delta)$th moment $(\delta > 0)$ of z_t for our consistency result; and A1(ii) is just the necessary and sufficient condition for second-order stationarity of GARCH(1,2) model (Bollerslev, 1986), which rules out the IGARCH process. We believe that the consistency of the global quasi-MLE of IGARCH(1,2) model can also be obtained along the line of Lumsdaine (1996), but at the possible cost of stronger condition such like the finite 32nd moment of z_t used by Lumsdaine. Lee and Hansen (1994) also used the assumption $\alpha_0 + \beta_0 < 1$ to ensure global consistency in their case. In this sense, our assumption for consistency is weaker than theirs.

Remark 2. Lee and Hansen (1994,Theorem 1) also created the consistency result for local quasi-MLE. However, it is trivial in our case in that we can get the consistency for any local quasi-MLE along the same line of our Theorem 1 due to the second order stationarity assumption on our GARCH model.

In order to obtain asymptotic normality of the local quasi-MLE, we need additional assumptions and further restrict to some subsets of the parameter space.

Assumption A2

(i) $E(z_t^4 | \mathcal{F}_{t-1}) \leq \kappa < \infty \quad a.s.$

(ii) $\boldsymbol{\theta}_0$ *is in the interior of* Θ.

Take $R_l = R(K_l^{-1}\alpha_{1l})$, where $R(\psi) = (2+\psi p)/(2+\psi) < 1$ (see Lemma 7) for $\psi > 0$, $p = 1 - \frac{1}{4\kappa} \in [0,1)$ ($\kappa \geq \frac{1}{4}$ by Cauchy-inequality) and $K_l = \dfrac{\omega_u}{\omega_0} + \dfrac{\alpha_{1u}}{\alpha_{10}} + \dfrac{\alpha_{2u}+\alpha_{1u}\beta_u}{\alpha_{2l}+\alpha_{1l}\beta_l} < \infty$.

Pick positive constants η_l and η_u, which satisfy

$$\eta_l < \beta_0(1 - R_l^{\frac{1}{6}}) \quad \text{and} \quad \eta_u < \beta_0(1 - R_0^{\frac{1}{6}}),$$

where $R_0 = R(\alpha_{10}) < 1$. Define for $1 \leq r \leq 6$ the constants

$$\beta_{rl} = \beta_0 R_l^{\frac{1}{r}} + \eta_l < \beta_0, \quad \beta_{ru} = \frac{\beta_0 - \eta_u}{R_0^{1/r}} > \beta_0,$$

and the subspaces $\Theta_l^r = \{\boldsymbol{\theta} \in \Theta : \beta_{rl} \leq \beta \leq \beta_0\}$, $\Theta_u^r = \{\boldsymbol{\theta} \in \Theta : \beta_0 \leq \beta \leq \beta_{ru}\}$ and $\Theta_r = \Theta_l^r \cup \Theta_u^r$. Note that $\Theta_r \subset \Theta_{r'}$ for $r \geq r'$ and $\boldsymbol{\theta}_0 \in \Theta_r$.

Now the asymptotic normality for the local quasi-MLE $\boldsymbol{\theta}_n^*$, defined by $argmax_{\boldsymbol{\theta}\in\Theta_2} L_n^*(\boldsymbol{\theta})$, follows:

Theorem 2 *Under A1 and A2,*

$$\sqrt{n}(\boldsymbol{\theta}_n^* - \boldsymbol{\theta}_0) \longrightarrow_D N(0, \boldsymbol{V}_0),$$

where $\boldsymbol{V}_0 = \boldsymbol{B}_0^{-1}\boldsymbol{A}_0\boldsymbol{B}_0^{-1}$, $\boldsymbol{B}_0 = \boldsymbol{B}(\boldsymbol{\theta}_0) = -E \bigtriangledown^2 l_t(\boldsymbol{\theta}_0)$, $\boldsymbol{A}_0 = E(\bigtriangledown l_t(\boldsymbol{\theta}_0) \bigtriangledown l_t(\boldsymbol{\theta}_0)')$, *and* $\bigtriangledown l_t(\boldsymbol{\theta})$ *is the vector of derivatives of* l_t *with respect to* $\boldsymbol{\theta}$.

REMARK 3. The conditions we used for obtaining the asymptotic normality of the local quasi-MLE are similar to those in Lee and Hansen (1994, Theorem 3) for GARCH(1,1).

3 Discussions

Our motivation for this study, in one aspect, came from the analysis that we did on stocks from the Center for Research in Securities Prices (CRSP) database. we selected all 367 stocks that are currently in the S&P 500 index and that have been trading on the NYSE or NASDAQ for the entire period between 1990-01-02 and 2000-12-29, and investigated their GARCH characters of monthly, weekly and daily return processes, mainly by using Akaike information criterion (AIC). We found that there are some stocks whose return processes are more complicated so that we need a GARCH(1,2) model instead of the better known GARCH(1,1). For more aggregated data such as

S&P500 Index or weekly returns, GARCH(1,1) is still a better choice. It is documented by Drost and Nijman (1993) that aggregation will simplify the GARCH character of the original process. But also note that even for aggregated data, GARCH(1,2) model may be necessary. For example, French et al. (1987) used this model for monthly stock returns in the period 1928-1984, and Pagan and Schwert (1990) for period 1835-1925.

Although Lee and Hansen (1994) stated that their methods are valid only for the simple GARCH(1,1) model and may not easily be generalized to more complicated cases, we put the results forward into GARCH(1,2) model successfully. Our main assumptions, i.e., the second moment condition for consistency and the fourth moment condition for asymptotic normality, are comparable to that of i.i.d. case (Li et al., 2002). For instance, in i.i.d. case of z_t, for the quasi-MLE of general GARCH model, consistency is obtained also under (*inter alia*) second moment condition (Ling and McAleer, 2003), or $(2+\delta)$th moment ($\delta > 0$) (Berkes et al., 2003); and asymptotic normality under (*inter alia*) the 6th moment condition (Ling and McAleer, 2003), or $(4+\delta)$th moment ($\delta > 0$) (Berkes et al., 2003).

Most theoretical studies of GARCH model so far focus on i.i.d z_t, and to the best of authors' knowledge, there is no new result published in the stationary case since Lee and Hansen (1994). The potential generalizations of our results include the extension to GARCH(1, p) model. We don't see essential difficulty for this extension, but with more complicated denotations and we think that it is not worth trying for the minor gain. Inclusion of IGARCH model is possible if similar condition of assumption A1(iv) of Lee and Hansen (1994) is used. For the general GARCH(p,q) model, the conditional variance cannot be written as a analytic function of innovations like (2) or (3). Apparently, the problem is the uniqueness of a presentation like (3) for the GARCH(p,q) model. Berkes et al. (2003) obtained the uniqueness under i.i.d z_t. If a similar result is obtained under stationary z_t, it is possible to prove the asymptotic properties of GARCH(p,q) model analogous to Berkes et al. (2003). In addition, in econometric literature and practice, GARCH models are usually assumed to describe the error process, together with an AR or ARMA process as the conditional mean. As Ling and Li (1998) and Li et al. (2002) noted, if the density function (we need a probability density) of z_t is symmetric, the MLE for parameters in the mean and GARCH structures can be obtained through a separate iteration procedure without loss of asymptotic efficiency for i.i.d. z_t. It is conjectured that this result still holds for stationary z_t. We leave it for the future.

Appendix

We list all lemmas here and proofs of the theorems and some important lemmas. All other proofs are available from the authors upon request. All inequalities and equalities hold almost surely if applicable. In addition, $|A| = (\text{tr}(A'A))^{1/2}$ denotes the Euclidean norm of a matrix or vector and $\|A\| = (E|A|^r)^{1/r}$ the L^r-norm of a random matrix or vector here.

Lemma 1

$$B^{-1}h_t^{\epsilon}(\boldsymbol{\theta}) \le h_t(\boldsymbol{\theta}) \le Bh_t^{\epsilon}(\boldsymbol{\theta}),$$

where

$$\begin{aligned} B &= 1 + 2(\gamma_u - \gamma_l)\max\{\frac{\alpha_{1u}}{\omega_l}, 1\} + \frac{(\alpha_{2u} + \alpha_{1u}\beta_u)(\gamma_u - \gamma_l)^2}{\omega_l(1-\beta_u)} + \frac{\alpha_{1u}(\gamma_u - \gamma_l)^2}{\omega_l} \\ &\quad + \frac{2(\gamma_u - \gamma_l)}{(1-\beta_u)^{1/2}}\max\{1, \frac{\alpha_{2u} + \alpha_{1u}\beta_u}{\omega_l}\}. \end{aligned}$$

Proof of Lemma 1. See Lee and Hansen (1994, Lemma 1).

Lemma 2 *Under A1, for all* $\boldsymbol{\theta} \in \Theta$

(a) $h_t(\boldsymbol{\theta})$, $l_t(\boldsymbol{\theta})$ *and its first and second derivatives are strictly stationary and ergodic;*

(b) $Eh_{0t} = \dfrac{\omega_0}{1 - \alpha_{10} - \alpha_{20} - \beta_0} < \infty$ *and* $Eh_t(\boldsymbol{\theta}) \le \bar{h} < \infty$ *for some positive constant* $\bar{h}$*.*

Proof of Lemma 2. (a) Since they are measurable functions of ϵ_t from Billingsley (1968), stationarity and ergodicity follow from Stout (1974, Theorem 3.5.8).

(b) See Lee and Hansen (1994, Theorem 2).

Lemma 3 *Under A1,*

$$\sup_{\theta \in \Theta} |L_n(\boldsymbol{\theta}) - L_n^*(\boldsymbol{\theta})| \longrightarrow_p 0.$$

Proof of Lemma 3. First, we have

$$L_n^*(\boldsymbol{\theta}) - L_n(\boldsymbol{\theta}) = \frac{1}{2n}\sum_{t=1}^{n} \ln \frac{h_t(\boldsymbol{\theta})}{h_t^*(\boldsymbol{\theta})} - \frac{1}{2n}\sum_{t=1}^{n}(\frac{e_t^2}{h_t^*(\boldsymbol{\theta})} - \frac{e_t^2}{h_t(\boldsymbol{\theta})}). \qquad (4)$$

For $h_t(\boldsymbol{\theta})$ we have the following inequality,

$$\begin{aligned} h_t(\boldsymbol{\theta}) &= \omega + \alpha_1 e_{t-1}^2 + (\alpha_2 + \alpha_1\beta)\sum_{k=0}^{\infty}\beta^k e_{t-2-k}^2 \\ &= h_t^*(\boldsymbol{\theta}) + (\alpha_2 + \alpha_1\beta)\sum_{k=t-2}^{\infty}\beta^k e_{t-2-k}^2 \\ &= h_t^*(\boldsymbol{\theta}) + (\alpha_2 + \alpha_1\beta)\beta^{t-2}\sum_{k=0}^{\infty}\beta^k e_{-k}^2 \\ &\leq h_t^*(\boldsymbol{\theta}) + \beta^{t-2}h_2(\boldsymbol{\theta}). \end{aligned}$$

The first part on the right hand side of (1) can be bounded as follows

$$\begin{aligned} 0 &\leq \frac{1}{2n}\sum_{t=1}^{n}\ln[\frac{h_t(\boldsymbol{\theta})}{h_t^*(\boldsymbol{\theta})}] \leq \frac{1}{2n}\sum_{t=1}^{n}\ln[1+\frac{\beta^{t-2}h_2(\boldsymbol{\theta})}{h_t^*(\boldsymbol{\theta})}] \leq \frac{1}{2n}\sum_{t=1}^{n}\frac{\beta^{t-2}h_2(\boldsymbol{\theta})}{h_t^*(\boldsymbol{\theta})} \\ &\leq \frac{1}{2n}\sum_{t=1}^{n}\frac{\beta^{t-2}h_2(\boldsymbol{\theta})}{\omega_l} \leq \frac{h_2(\boldsymbol{\theta})}{2n\omega_l\beta_l(1-\beta_u)} \leq \frac{h_2(\boldsymbol{\theta}^u)}{2n\omega_l\beta_l(1-\beta_u)}, \end{aligned}$$

where the third inequality follows by the fact that $\ln(1+x) \leq x$ for $x > -1$, and $\boldsymbol{\theta}^u \triangleq [\gamma_u, \omega_u, \alpha_{1u}, \alpha_{2u}, \beta_u]'$. Thus by Lemma 2(b) and Markov inequality it follows that

$$\sup_{\theta\in\Theta}\frac{1}{2n}\sum_{t=1}^{n}\ln\frac{h_t(\boldsymbol{\theta})}{h_t^*(\boldsymbol{\theta})} \longrightarrow_p 0. \tag{5}$$

Further, for the second part of the right hand side of (1) we have

$$\begin{aligned} 0 &\leq \frac{1}{2n}\sum_{t=1}^{n}(\frac{e_t^2}{h_t^*(\boldsymbol{\theta})} - \frac{e_t^2}{h_t(\boldsymbol{\theta})}) = \frac{1}{2n}\sum_{t=1}^{n}\frac{h_t(\boldsymbol{\theta}) - h_t^*(\boldsymbol{\theta})}{h_t^*(\boldsymbol{\theta})}\cdot\frac{e_t^2}{h_t(\boldsymbol{\theta})} \\ &\leq \frac{1}{2n}\sum_{t=1}^{n}\frac{\beta^{t-2}h_2(\boldsymbol{\theta})}{h_t^*(\boldsymbol{\theta})}\cdot\frac{e_t^2}{h_t(\boldsymbol{\theta})} \leq \frac{1}{2n}\frac{h_2(\boldsymbol{\theta})}{\beta\omega_l^2}\sum_{t=1}^{n}\beta^{t-1}e_t^2 \\ &\leq \frac{h_2(\boldsymbol{\theta})\sum_{t=1}^{\infty}\beta^{t-1}h_{t+1}(\boldsymbol{\theta})}{2n\beta_l\omega_l^2\alpha_{1l}} \leq \frac{h_2(\boldsymbol{\theta}^u)\sum_{t=1}^{\infty}\beta^{t-1}h_{t+1}(\boldsymbol{\theta}^u)}{2n\beta_l\omega_l^2\alpha_{1l}} \triangleq \frac{X_2(\boldsymbol{\theta}^u)}{2n\beta_l\omega_l^2\alpha_{1l}}. \end{aligned}$$

By Cauchy-Schwarz and Minkowski's inequalities and Lemma 2(b) it follows that

$$E \mid X_2(\boldsymbol{\theta}^u) \mid^{\frac{1}{2}} \leq \frac{\bar{h}}{(1-\beta_u)^{1/2}} < \infty,$$

which, together with Markov inequality and (1)-(2), implies the lemma.
REMARK. Note that in the proof of Lee and Hansen (1994, Lemma 3) the bounds were depending on $\boldsymbol{\theta}$, which was not enough for uniform convergence.

Lemma 4 (a) *For* $\beta \le \beta_0$, $\dfrac{h_t^\epsilon(\boldsymbol{\theta})}{h_{0t}} \le K_l \triangleq \dfrac{\omega_u}{\omega_0} + \dfrac{\alpha_{1u}}{\alpha_{10}} + \dfrac{\alpha_{2u} + \alpha_{1u}\beta_u}{\alpha_{2l} + \alpha_{1l}\beta_l} < \infty$;

(b) *For* $\beta > \beta_0$, $\dfrac{h_{0t}}{h_t^\epsilon(\boldsymbol{\theta})} \le H_u \triangleq \dfrac{\omega_0}{\omega_l} + \dfrac{\alpha_{10}}{\alpha_{1l}} + \dfrac{\alpha_{20} + \alpha_{10}\beta_0}{\alpha_{2l} + \alpha_{1l}\beta_l} < \infty$.

Proof of Lemma 4. They are trivial and omitted.

Lemma 5 *Under A1, for* $\boldsymbol{\theta} \in \Theta$

(a) $E\dfrac{e_t^2}{h_t(\boldsymbol{\theta})} \le H_1 \triangleq \dfrac{\bar{h}}{\omega_l} + \dfrac{(\gamma_u - \gamma_l)^2}{\omega_l} < \infty$;

(b) $L_n(\boldsymbol{\theta}) \longrightarrow_p L(\boldsymbol{\theta}) = E\dfrac{l_t(\boldsymbol{\theta})}{2}$.

Proof of Lemma 5. (a). From Lemma 2(b), we have

$$E\frac{e_t^2}{h_t(\boldsymbol{\theta})} \le \frac{E\epsilon_t^2}{\omega} + \frac{g^2}{\omega_l} = \frac{1}{\omega}Eh_t(\boldsymbol{\theta}) + \frac{g^2}{\omega_l} \le \frac{\bar{h}}{\omega_l} + \frac{(\gamma_u - \gamma_l)^2}{\omega_l} < \infty.$$

(b). By Jensen's inequality, part (a) of this Lemma, and Lemma 2, we obtain

$$E|l_t(\boldsymbol{\theta})| \le E|\ln h_t(\boldsymbol{\theta})| + E(\frac{e_t^2}{h_t(\boldsymbol{\theta})}) \le \ln Eh_t(\boldsymbol{\theta}) + H_1 < \infty.$$

This allows the application of the SLLN for stationary and ergodic sequences (see, e.g, Stout (1974, Theorem 3.5.7), which yields the desired result.

Lemma 6 *Under A1,* $\sup_{\theta\in\Theta} E|\bigtriangledown l_t(\boldsymbol{\theta})| < \infty$.

Proof of Lemma 6. We will only show $\sup_{\theta\in\Theta} E|\partial l_t(\boldsymbol{\theta})/\partial\beta| < \infty$, which is the most difficult part. For the rest, see, for example, Lumsdaine (1996) and Lee and Hansen (1994,Lemma 8). First, define $h_{\beta t}(\boldsymbol{\theta}) = \frac{\partial h_t(\boldsymbol{\theta})}{\partial\beta} \cdot \frac{1}{h_t(\boldsymbol{\theta})}$. Differentiating with respect to β both sides of

$$h_t(\boldsymbol{\theta}) = \omega(1-\beta) + \beta h_{t-1}(\boldsymbol{\theta}) + \alpha_1 e_{t-1}^2 + \alpha_2 e_{t-2}^2$$

yields

$$\frac{\partial h_t(\boldsymbol{\theta})}{\partial\beta} = -\omega + h_{t-1}(\boldsymbol{\theta}) + \beta\frac{\partial h_{t-1}(\boldsymbol{\theta})}{\partial\beta}.$$

Hence, after iterations,

$$\begin{aligned}
E|h_{\beta t}(\boldsymbol{\theta})| &= E \mid \sum_{k=0}^{\infty} \beta^k \frac{h_{t-k-1}(\boldsymbol{\theta}) - \omega}{h_t(\boldsymbol{\theta})} \mid \\
&\leq \sum_{k=0}^{\infty} \beta^k E \frac{\omega}{h_t(\boldsymbol{\theta})} + \sum_{k=0}^{\infty} \beta^k \frac{E|h_{t-k-1}(\boldsymbol{\theta})|}{\omega} \\
&\leq \frac{\bar{h} + \omega_l}{\omega_l(1-\beta_u)} < \infty.
\end{aligned}$$

Now, we can more explicitly write

$$\begin{aligned}
E|\partial l_t(\boldsymbol{\theta})/\partial \beta| &= E|\frac{e_t^2}{h_t} h_{\beta t}(\boldsymbol{\theta}) - h_{\beta t}(\boldsymbol{\theta})| \\
&\leq E[E((\epsilon_t + g)^2 | \mathcal{F}_{t-1}) \frac{|h_{\beta t}(\boldsymbol{\theta})|}{h_t(\boldsymbol{\theta})}] + E|h_{\beta t}(\boldsymbol{\theta})| \\
&\leq E|\frac{h_{0t}}{h_t(\boldsymbol{\theta})} h_{\beta t}(\boldsymbol{\theta})| + E|h_{\beta t}(\boldsymbol{\theta})|(1 + \frac{g^2}{\omega_l}).
\end{aligned}$$

For $\beta > \beta_0$, from Lemma 1 and Lemma 4(b),

$$E|\frac{h_{0t}}{h_t(\boldsymbol{\theta})} h_{\beta t}(\boldsymbol{\theta})| \leq B \cdot H_u \cdot E|h_{\beta t}(\boldsymbol{\theta})| < \infty.$$

For $\beta \leq \beta_0$,

$$\begin{aligned}
\frac{h_{0t}}{h_t^{\epsilon}(\boldsymbol{\theta})} &\leq \frac{\omega_0 + \alpha_{10}\epsilon_{t-1}^2 + (\alpha_{20} + \alpha_{10}\beta_0)\sum_{i=0}^{k-1} \beta_0^i \epsilon_{t-2-i}^2 + \beta_0^k h_{0t-k-1}}{h_t^{\epsilon}(\boldsymbol{\theta})} \\
&\leq \frac{\omega_0}{\omega} + \frac{\alpha_{10}}{\alpha_1} + \frac{(\alpha_{20} + \alpha_{10}\beta_0)}{\alpha_2 + \alpha_1\beta} k(\frac{\beta_0}{\beta})^k + \beta_0^k \frac{h_{0t-k-1}}{\omega,}
\end{aligned}$$

which follows from $(\alpha_2 + \alpha_1\beta)\beta^i \epsilon_{t-2-i}^2 \leq h_t^{\epsilon}(\boldsymbol{\theta})$ and $\sum_{i=0}^{k-1}(\beta_0/\beta)^i \leq k(\beta_0/\beta)^k$ as $(\beta_0/\beta) \geq 1$. Thus,

$$\begin{aligned}
& B^{-1} E \frac{h_{0t}}{h_t(\boldsymbol{\theta})} h_{\beta t}(\boldsymbol{\theta}) \\
&\leq E \frac{h_{0t}}{h_t^{\epsilon}(\boldsymbol{\theta})} h_{\beta t}(\boldsymbol{\theta}) \\
&\leq \sum_{k=0}^{\infty} \beta^k E \frac{h_{t-k-1}(\boldsymbol{\theta})}{h_t(\boldsymbol{\theta})} \left[\frac{\omega_0}{\omega} + \frac{\alpha_{10}}{\alpha_1} + \frac{\alpha_{20} + \alpha_{10}\beta_0}{\alpha_2 + \alpha_1\beta} k(\frac{\beta_0}{\beta})^k + \beta_0^k \frac{h_{0t-k-1}}{\omega} \right]
\end{aligned}$$

$$\leq \quad (\frac{\omega_0}{\omega} + \frac{\alpha_{10}}{\alpha_1})\frac{\bar{h}}{\omega(1-\beta_u)} + \frac{\alpha_{20}+\alpha_{10}\beta_0}{(\alpha_2+\alpha_1\beta)\omega}\sum_{k=0}^{\infty} k\beta_0^k Eh_{t-k-1}(\boldsymbol{\theta})$$
$$+\frac{1}{\omega}\sum_{k=0}^{\infty}\beta_0^k Eh_{0t-k-1}$$
$$\leq \quad (\frac{\omega_0}{\omega_l} + \frac{\alpha_{10}}{\alpha_{1l}})\frac{\bar{h}}{\omega_l(1-\beta_u)} + \frac{\alpha_{20}+\alpha_{10}\beta_0}{(\alpha_{2l}+\alpha_{1l}\beta_l)\omega_l}\cdot\frac{\beta_0\bar{h}}{(1-\beta_0)^2} + \frac{Eh_{0t}}{(1-\beta_0)\omega_l},$$

since $\beta^k h_{t-k-1}(\boldsymbol{\theta})/h_t(\boldsymbol{\theta}) \leq 1$, which completes the proof.

Proof of Theorem 1. First, note that Θ is compact by assumption. Second, in Lemma 5(b) we have shown that $L_n(\boldsymbol{\theta})$ converges to $L(\boldsymbol{\theta})$ in probability. Further, Lemma 6 implies that $L_n(\boldsymbol{\theta})$ satisfies the weak Lipschitz condition and the condition here is stronger than that in the Theorem 3 of Andrews (1992). Thus $L_n(\boldsymbol{\theta})$ tends to $L(\boldsymbol{\theta})$ in probability uniformly in Θ and $L(\boldsymbol{\theta})$ is continuous in Θ from there. Combined with Lemma 3, it yields that

$$\sup_{\theta\in\Theta}|L_n^*(\boldsymbol{\theta}) - L(\boldsymbol{\theta})| \leq \sup_{\theta\in\Theta}|L_n^*(\boldsymbol{\theta}) - L_n(\boldsymbol{\theta})| + \sup_{\theta\in\Theta}|L_n(\boldsymbol{\theta}) - L(\boldsymbol{\theta})| \longrightarrow_p 0$$

Third, the proofs in Lumsdaine (1996, Lemma 5 and Theorem 1) can also cover our case to show that the limiting likelihood $L(\boldsymbol{\theta})$ is uniquely maximized at $\boldsymbol{\theta}_0$.

Now, we have the standard conditions for consistency in nonlinear estimation. That $\hat{\boldsymbol{\theta}}_n \longrightarrow_p \boldsymbol{\theta}_0$ follows from, for example, Amemiya (1985, Theorem 4.1.1).

Lemma 7 *Under A1 and A2,*

(a) *For all $\psi > 0$ and all $r \geq 1$*

$$\beta^r E((\beta+\psi z_t^2)^{-r}|\mathcal{F}_{t-1}) < E((1+\psi z_t^2)^{-1}|\mathcal{F}_{t-1}) \leq R(\psi),$$

where $R(\psi) = (2+\psi p)/(2+\psi) < 1$, and $p = 1 - \frac{1}{4\kappa} \in (0,1)$;

(b) *For all finite r,*

$$E((\frac{h_{0t-k}}{h_{0t}})^r|\mathcal{F}_{t-k-1}) \leq (\frac{R_0}{\beta_0^r})^k,$$

where $R_0 = R(\alpha_{10}) < 1$;

Sketch of Proof of Lemma 7. Lee and Hansen (1994) used the condition

$$sup_t E(z_t^{2+\delta}|\mathcal{F}_{t-1}) \leq S_\delta < \infty$$

to get a quantity less than unity to uniformly bound the conditional probability $P(z_t^2 \le \frac{1}{2}|\mathcal{F}_{t-1})$, then, to bound the expectation $E(\frac{1}{1+\psi z_t^2}|\mathcal{F}_{t-1})$, which was said to be 'of major importance'. However, since we have now the uniform finiteness of fourth moment of z_t, simply replacing their S_δ by our $\kappa(=S_2)$, our results follows analogous to Lee and Hansen (1994, Lemma 4(2) and 4(3)).

Lemma 8 *Under A1 and A2, for* $1 \le r \le 6$

(a) $\beta^k \parallel \dfrac{h^{\epsilon}_{t-k}(\boldsymbol{\theta})}{h^{\epsilon}_t(\boldsymbol{\theta})} \parallel_r \le (R_l^{1/r})^k$ *uniformly in* $\boldsymbol{\theta} \in \Theta_l^r$;

(b) $\parallel \dfrac{h_{0t}}{h^{\epsilon}_t(\boldsymbol{\theta})} \parallel_r \le H_c \triangleq \dfrac{\omega_0}{\omega_l} + \dfrac{\alpha_{10}}{\alpha_{1l}} + \dfrac{\alpha_{20}+\alpha_{10}\beta}{\alpha_{2l}+\alpha_{1l}\beta_l} \cdot \dfrac{1}{\eta_l}$ *uniformly in* $\boldsymbol{\theta} \in \Theta_r$;

(c) $\parallel \dfrac{h^{\epsilon}_t(\boldsymbol{\theta})}{h_{0t}} \parallel_r \le K_u \triangleq \dfrac{\omega_u}{\omega_0} + \dfrac{\alpha_{1u}}{\alpha_{10}} + \dfrac{\alpha_{2u}+\alpha_{1u}\beta_u}{(\alpha_{20}+\alpha_{10}\beta_0)\eta_u}$ *uniformly in* $\boldsymbol{\theta} \in \Theta_u^r$;

(d) $\parallel \dfrac{h^{\epsilon}_{t-k}(\boldsymbol{\theta})}{h^{\epsilon}_t(\boldsymbol{\theta})} \parallel_r \le K_u H_u (\dfrac{R_0^{1/r}}{\beta_0})^k$ *uniformly in* $\boldsymbol{\theta} \in \Theta_u^r$;

(e) $\sup\limits_{\theta\in\Theta_r} \parallel h_{\beta t}(\boldsymbol{\theta}) \parallel_r \le \dfrac{1}{1-\beta_u} + B^2 max(R_l^{\frac{1}{r}}\beta_u^{-1}(1-R_l^{\frac{1}{r}})^{-1}, R_0^{\frac{1}{r}}\beta_0^{-1}H_u K_u \eta_u^{-1})$
$\triangleq H_\beta < \infty$.

Lemma 9 *Under A1 and A2,*

(a) *For all* $\boldsymbol{\theta} \in \Theta_4$, $E| \bigtriangledown l_t(\boldsymbol{\theta}) \bigtriangledown l_t(\boldsymbol{\theta})'| < \infty$;

(b) $\dfrac{1}{\sqrt{n}} \boldsymbol{A}_0^{-\frac{1}{2}} \sum\limits_{t=1}^{[nr]} \bigtriangledown l_t^*(\boldsymbol{\theta}_0) \Longrightarrow W(r)$, *where* $\boldsymbol{A}_0 = E(\bigtriangledown l_t(\boldsymbol{\theta}_0) \bigtriangledown l_t(\boldsymbol{\theta}_0)')$, $W(r)$ *denotes a Brownian motion with covariance matrix* I_4, I_4 *being the* 4×4 *identity matrix, and [.] is the integer part.*

Lemma 10 *Under A1 and A2, for* $1 \le r \le 6$

(a) $\sup_{\theta\in\Theta_{2r}} \parallel h_{\beta\beta t}(\boldsymbol{\theta}) \parallel_r \le 2H_\beta^2 < \infty$, *where* $h_{\beta\beta t} = h_t(\boldsymbol{\theta})^{-1}\partial^2 h_t(\boldsymbol{\theta})/\partial\beta^2$;

(b) $\sup_{\theta\in\Theta_{3r}} \parallel h_{\beta\beta\beta t}(\boldsymbol{\theta}) \parallel_r \le 6H_\beta^3 < \infty$, *where* $h_{\beta\beta\beta t} = h_t(\boldsymbol{\theta})^{-1}\partial^3 h_t(\boldsymbol{\theta})/\partial\beta^3$.

Lemma 11 *Under A1 and A2,*

(a) *For all* $\boldsymbol{\theta} \in \Theta_4$, $E| \bigtriangledown^2 l_t(\boldsymbol{\theta})| < \infty$;

(b) *For all* $\boldsymbol{\theta} \in \Theta_4$ *and* $1 \le i \le 5$, $E|\frac{\partial}{\partial \boldsymbol{\theta}_i} \bigtriangledown^2 l_t(\boldsymbol{\theta})| < \infty$, *where* $\boldsymbol{\theta}_i$ *is the i-th element of* $\boldsymbol{\theta}$*;*

(c) $\sup_{\theta \in \Theta_4} |\hat{\boldsymbol{B}}_n(\boldsymbol{\theta}) - \boldsymbol{B}(\boldsymbol{\theta})| \longrightarrow_p 0$ *and* $\boldsymbol{B}(\boldsymbol{\theta})$ *is continuous in* Θ_4*, where* $\hat{\boldsymbol{B}}_n(\boldsymbol{\theta}) = -\frac{1}{n}\sum_{t=1}^n \bigtriangledown^2 l_t^*(\boldsymbol{\theta})$ *and* $\boldsymbol{B}(\boldsymbol{\theta}) = -E \bigtriangledown^2 l_t(\boldsymbol{\theta})$.

After close examinations, we find that Lemma 8-11 can be prove along the same line as Lee and Hansen (1994, Lemma 5-6, Lemma 8-11) with some modifications induced by the change in the conditional variance equation (1), as shown in our lemmas. So proofs of Lemma 8-11 are omitted for sake of space.

Proof of Theorem 2. First, as indicated in remark 3 after our Theorem 2, we can show that $\boldsymbol{\theta}_n^*$ is consistent; Next, from Lemma 9(b), we have

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n \bigtriangledown l_t^*(\boldsymbol{\theta}_0) \rightarrow_D N(\boldsymbol{0}, \boldsymbol{A}_0).$$

By Lemma 11(c), we get that

$$-\frac{1}{n} \sum_{t=1}^n \bigtriangledown^2 l_t^*(\boldsymbol{\theta}) \rightarrow_p \boldsymbol{B}(\boldsymbol{\theta})$$

and $\boldsymbol{B}(\boldsymbol{\theta})$ is continuous in Θ_4. Finally, by the proof of identifiability in Theorem 1, $\boldsymbol{B}_0 > 0$. Hence, from Amemiya (1985, Theorem 4.1.3), the theorem follows.

References

Amemiya, T. (1985). *Advanced Econometrics.* Cambridge: Harvard University Press.

Andrews, D.W.K. (1992). Generic uniform convergence. *Econometric Theory* **8**, 241-257.

Berkes, I., Horvath, L. and Kokoszka, P. (2003). GARCH processes: structure and estimation. *Bernoulli* **9**, 201-227.

Billingsley, P. (1968). *Convergence of Probability Measure.* New York: Wiley.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* **31**, 307-27.

Bollerslev, T., Chou, R.Y. and Kroner, K.F. (1992). ARCH modelling in finance: A review of the theory and empirical evidence. *Journal of Econometrics* **52**, 5-59.

Drost, F.C. and Nijman, T.E. (1993). Temporal aggregation of GARCH processes. *Econometrica* **61**, 909-927.

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

French, K.R., Schwert, G.W. and Stambaugh, R.F. (1987). Expected stock returns and volatility. *Journal of Financial Economics* **19**, 3-29.

Lee, S.W. and Hansen, B.E. (1994). Asymptotic theory for the GARCH(1,1) quasi-maximum likelihood estimator. *Econometric Theory* **10**, 29-52.

Li, W.K., Ling, S. and Mcaleer, M. (2002). Recent theoretical results for time series models with GARCH errors. *Journal of Economic Surveys* **16**, 245-269.

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

Ling, S. and McAleer, M. (2003). Asymptotic theory for a vector ARMA-GARCH model. *Econometric Theory* **19**, 280-310.

Lumsdaine, R. L. (1991). *Asymptotic Properties of the Quasi-Maximum Likelihood Estimator in GARCH(1,1) and IGARCH(1,1) Models.* Dissertation, Princeton University.

Lumsdaine, R.L. (1996). Consistency and asymptotic normality of the quasi-maximum likelihood estimator in IGARCH(1,1) and covariance stationary GARCH(1,1) models. *Econometrica* **64**, 575-596.

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

Pagan, A.R. and Schwert, G.W. (1990). Alternative models for conditional stock volatility. *Journal of Econometrics* **45**(1-2), 267-290.

Stout, W.R. (1974). *Almost Sure Convergence.* New York: Academic Press.

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