*Econometric Theory*, **25**, 2009, 561–570. Printed in the United States of America.
doi:10.1017/S0266466609090689

# NOTES AND PROBLEMS

# QUASI-MAXIMUM LIKELIHOOD ESTIMATION OF SEMI-STRONG GARCH MODELS

JUAN CARLOS ESCANCIANO
*Indiana University*

This note proves the consistency and asymptotic normality of the quasi–maximum likelihood estimator (QMLE) of the parameters of a generalized autoregressive conditional heteroskedastic (GARCH) model with martingale difference centered squared innovations. The results are obtained under mild conditions and generalize and improve those in Lee and Hansen (1994, *Econometric Theory* 10, 29–52) for the local QMLE in semistrong GARCH(1,1) models. In particular, no restrictions on the conditional mean are imposed. Our proofs closely follow those in Francq and Zakoïan (2004, *Bernoulli* 10, 605–637) for independent and identically distributed innovations.

## 1. INTRODUCTION

Since the seminal contributions by Engle (1982) and Bollerslev (1986), generalized autoregressive conditional heteroskedastic (GARCH) models have received considerable attention in the theoretical and empirical literature devoted to financial and economic time series. Although a vast quantity of alternative volatility models have been proposed to capture several "stylized facts" of financial series, the GARCH model is certainly still the most popular and successful model. See Straumann (2005) for a recent reference on GARCH and other volatility models.

The asymptotic properties of the quasi–maximum likelihood estimator (QMLE) were first established by Weiss (1986) for autoregressive conditional heteroskedastic (ARCH) models. Lumsdaine (1996) proved the consistency and asymptotic normality (CAN) of the *local* QMLE for the GARCH(1,1) model. These results

The note has benefited from the comments of three referees, the co-editor, Paolo Paruolo, and Emma Iglesias. This research was funded by the Spanish Plan Nacional de I+D+I, reference number SEJ2007-62908. Address correspondence to Juan Carlos Escanciano, Indiana University, Department of Economics, 100 S. Woodlawn, Wylie Hall, Bloomington, IN 47405-7104, U.S.A.; e-mail: jescanci@indiana.edu.

© 2009 Cambridge University Press 0266-4666/09 $15.00 

https://doi.org/10.1017/S0266466609090689 Published online by Cambridge University Press

for the local QMLE have been improved in, e.g., Ling (2007). Berkes, Horváth, and Kokoszka (2003) gave rigorous proofs of CAN for the *global* QMLE for GARCH$(p,q)$ models. See also Berkes and Horváth (2003, 2004), Hall and Yao (2003), and Robinson and Zaffaroni (2006). Some of the conditions in Berkes et al. (2003) were improved by Francq and Zakoïan (2004), who provided the weakest known conditions under finite fourth moments of innovations. For a recent review on the QMLE in GARCH models, in addition to other aspects of these models, see the survey by Li, Ling, and McAleer (2002).

However, all the aforementioned references assumed independent and identically distributed (i.i.d.) standardized innovations. The i.i.d. assumption on the standardized innovations restricts substantially the flexibility of GARCH models in modeling higher order conditional moments, such as conditional skewness and kurtosis. This limitation is important from an empirical point of view because there is now a growing econometrics and financial literature documenting time-varying conditional skewness and kurtosis in economic and financial time series; see, e.g., Gallant, Hsieh, and Tauchen (1991), Hansen (1994), Harvey and Siddique (1999, 2000), Premaratne and Bera (2001), Jondeau and Rockinger (2003), Leon, Rubio, and Serna (2005), and Brooks, Burke, Heravi, and Persand (2005), among many others. In this paper we are consistent with this existing literature, and hence, we do not restrict the functional form of these higher order conditional moments. As we shall show, neglected conditional kurtosis of innovations may cause misleading conclusions in statistical inferences using the QMLE; see the discussion after our Theorem 2.

To the best of our knowledge, only Lee and Hansen (1994) have studied the CAN of the local QMLE in GARCH(1,1) models with martingale difference innovations. As these authors explicitly stated (see Lee and Hansen, 1994, p. 39), their methods may not easily generalize to more complicated models, e.g., GARCH$(p,q)$. It remains then an open problem to investigate the CAN of the *global* QMLE for general GARCH$(p,q)$ models under martingale assumptions. The purpose of the present note is to carry out such an analysis under mild conditions on the data generating process.

The rest of the note is organized as follows. Section 2 studies the CAN of the global QMLE for the GARCH$(p,q)$ model. Mathematical proofs are relegated to Section 3. Henceforth, the norm of a matrix $A=(a_{ij})$ is defined by $\|A\|=\sum|a_{ij}|$. The symbols $\longrightarrow_d$ and $\longrightarrow_{a.s.}$ denote convergence in distribution and almost surely (a.s.), respectively. All limits are taken as $n\to\infty$.

## 2. CAN OF THE QMLE IN GARCH ($p,q$) MODELS

The GARCH$(p,q)$ model is defined as

$$Y_t=\sqrt{h_t}\varepsilon_t,\qquad h_t=w_0+\sum_{i=1}^{q}\alpha_{0i}Y_{t-i}^2+\sum_{j=1}^{p}\beta_{0j}h_{t-j},\qquad t\in\mathbb{Z},\qquad p,q\in\mathbb{N},$$

https://doi.org/10.1017/S0266466609090689 Published online by Cambridge University Press

where $\varepsilon_t$ is a sequence of strictly stationary and ergodic random variables satisfying

$$\mathrm{E}[\varepsilon_t^2 \mid \mathcal{F}_{t-1}] = 1 \quad \text{a.s.} \tag{1}$$

and $w_0 > 0$, $\alpha_{0i} \geq 0$, $i = 1, \ldots, q$, $\beta_{0j} \geq 0$, $j = 1, \ldots, p$. Here $\mathcal{F}_{t-1}$ is the $\sigma$-field generated by $(Y_{t-1}, Y_{t-2}, \ldots)$. Define the vector of parameters $\theta = (w, \alpha_1, \ldots, \alpha_q, \beta_1, \ldots, \beta_p)'$ and the parameter space $\Theta \subset (0, \infty) \times [0, \infty)^{p+q}$. The true parameter value is unknown, and it is denoted by $\theta_0 = (w_0, \alpha_{01}, \ldots, \alpha_{0q}, \beta_{01}, \ldots, \beta_{0p})'$. The prime denotes transposition.

The problem we tackle is, given a sample of size $n$ of $Y_t$, $\{Y_t\}_{t=1}^n$, say, to estimate the parameter $\theta$. A QMLE estimator is defined as any measurable solution $\widehat{\theta}_n$ of

$$\widehat{\theta}_n = \arg\min_{\theta \in \Theta} \widetilde{L}_n(\theta) \equiv n^{-1} \sum_{t=1}^{n} \widetilde{\ell}_t(\theta), \qquad \widetilde{\ell}_t(\theta) = \frac{Y_t^2}{\widetilde{\sigma}_t^2(\theta)} + \log \widetilde{\sigma}_t^2(\theta), \tag{2}$$

where the $\widetilde{\sigma}_t^2$ are defined recursively, for $t \geq 1$, by

$$\widetilde{\sigma}_t^2 \equiv \widetilde{\sigma}_t^2(\theta) = w + \sum_{i=1}^{q} \alpha_i Y_{t-i}^2 + \sum_{j=1}^{p} \beta_j \widetilde{\sigma}_{t-j}^2(\theta),$$

starting with initial values chosen, e.g., as

$$Y_0^2 = \cdots = Y_{1-q}^2 = \widetilde{\sigma}_0^2 = \cdots = \widetilde{\sigma}_{1-p}^2 = w \tag{3}$$

or

$$Y_0^2 = \cdots = Y_{1-q}^2 = \widetilde{\sigma}_0^2 = \cdots = \widetilde{\sigma}_{1-p}^2 = Y_1^2. \tag{4}$$

It can be shown that the initial values do not matter for the asymptotic properties of the QMLE; see Section 3. However, they may have importance from a practical point of view.

For strong consistency of $\widehat{\theta}_n$ we need the following regularity conditions. First, define the polynomials $\mathcal{A}_\theta(z) = \sum_{i=1}^{q} \alpha_i z^i$ and $\mathcal{B}_\theta(z) = 1 - \sum_{j=1}^{p} \beta_j z^i$, $z \in \mathbb{C}$. By convention, $\mathcal{A}_\theta(z) \equiv 0$ if $q = 0$ and $\mathcal{B}_\theta(z) \equiv 1$ if $p = 0$.

**Assumptions A1.**

(a) $\{\varepsilon_t\}_{t \in \mathbb{Z}}$ is a strictly stationary and ergodic process satisfying (1), and $\varepsilon_t^2$ has a nondegenerate distribution;

(b) $\{Y_t\}_{t \in \mathbb{Z}}$ is a strictly stationary and ergodic process with $\mathrm{E}[|Y_t|^{2s}] < \infty$ for some $s > 0$.

**Assumptions A2.** $\theta_0 \in \Theta$ and $\Theta$ is compact.

https://doi.org/10.1017/S0266466609090689 Published online by Cambridge University Press

**Assumptions A3.** If $p > 0$, $\mathcal{A}_{\theta_0}(z)$ and $\mathcal{B}_{\theta_0}(z)$ have no common root. For all $\theta \in \Theta$, $\mathcal{B}_\theta(z)$ has its roots outside the unit disk. Moreover, $\mathcal{A}_{\theta_0}(1) \neq 0$ and $\alpha_{0q} + \beta_{0p} \neq 0$.

We are now in a position to establish the first result of the note.

THEOREM 1. *Let $(\widehat{\theta}_n)$ be a sequence of QMLE satisfying (2), with initial conditions (3) or (4). Then, under Assumptions A1–A3, $\widehat{\theta}_n \longrightarrow_{a.s.} \theta_0$.*

Our assumptions are standard in the literature. Recently, Linton, Pan, and Wang (2009) (see also Dahl and Iglesias, 2007) found necessary and sufficient conditions for the strict stationarity in Assumption A1(b) for the semistrong GARCH(1,1) model. The sufficient conditions are as follows: $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ is a strictly stationary and ergodic process, $\varepsilon_t^2$ has a nondegenerate distribution, $\mathrm{E}[|\varepsilon_t|^s] < \infty$ for some $s > 0$, and $\gamma = \mathrm{E}[\log(\alpha_0\varepsilon_t^2 + \beta_0)] < 0$. These sufficient conditions can be easily extended to general GARCH$(p,q)$ models and, in particular, imply that the roots of $\mathcal{B}_{\theta_0}(z)$ are outside the unit disk. In view of Jensen and Rahbek (2004a, 2004b), we conjecture that our results are also valid for some nonstationary time series. Francq and Zakoïan (2004) discussed assumptions similar to A2 and A3. Unlike these authors, we do not assume that the standardized innovations are i.i.d., but rather we assume that they are strictly stationary and ergodic innovations satisfying the martingale condition (1). Lee and Hansen (1994) also required the latter assumption, but results were given for the local QMLE, i.e., solving (2) in a neighborhood of $\theta_0$. Moreover, Lee and Hansen (1994) assumed that $\mathrm{E}[Y_t \mid \mathcal{F}_{t-1}] = \gamma$ a.s. for a constant $\gamma$. We do not impose restrictions on the conditional mean. Furthermore, unlike in Lee and Hansen (1994), there is no need for $\theta_0$ to belong to the interior of $\Theta$, allowing situations of overidentification such as ARCH$(q)$ models. See Francq and Zakoïan (2004, p. 609) for remarks about Assumption A3 and its implications on the parameter space.

Let $\sigma_t^2(\theta)$ be the strictly stationary and ergodic solution (cf. Assumption A3) of

$$\sigma_t^2 \equiv \sigma_t^2(\theta) = w + \sum_{i=1}^{q} \alpha_i Y_{t-i}^2 + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2(\theta). \tag{5}$$

For asymptotic normality, we need the following additional assumptions.

**Assumptions A4.** $\mathrm{E}[|\varepsilon_t|^{4(1+\delta)}] < \infty$ for some $\delta > 0$.

**Assumptions A5.** $\theta_0 \in int(\Theta)$, where $int(\Theta)$ denotes the interior of $\Theta$.

THEOREM 2. *Under Assumptions A1–A5,*

$$\sqrt{n}(\widehat{\theta}_n - \theta_0) \longrightarrow_d N(0, J^{-1} I J^{-1}),$$

*where* $J = \mathrm{E}_{\theta_0}[(1/\sigma_t^4(\theta_0))(\partial\sigma_t^2(\theta_0)/\partial\theta)(\partial\sigma_t^2(\theta_0)/\partial\theta')]$,
$I = \mathrm{E}_{\theta_0}[(\tau_t - 1)(1/\sigma_t^4(\theta_0))\ (\partial\sigma_t^2(\theta_0)/\partial\theta)(\partial\sigma_t^2(\theta_0)/\partial\theta')]$, *and* $\tau_t = \mathrm{E}[\varepsilon_t^4 \mid \mathcal{F}_{t-1}]$.

https://doi.org/10.1017/S0266466609090689 Published online by Cambridge University Press

In Berkes et al. (2003) Assumption A4 is assumed in addition to $t^{-\mu}P(\varepsilon_t^2 \le t) \to 0$ when $t \to 0$, for some $\mu > 0$. We do not require the latter assumption. Our moment assumption is just slightly stronger than that of Francq and Zakoïan (2004), who required fourth moments in the innovations. Hall and Yao (2003) and Mikosch and Straumann (2002) discussed the case where $\mathrm{E}[\varepsilon_t^4] = \infty$; see also Linton et al. (2007). Lee and Hansen (1994) assumed that $\tau_t = \mathrm{E}[\varepsilon_t^4 \mid \mathcal{F}_{t-1}] \le C < \infty$ a.s. Under constant conditional kurtosis, i.e., $\tau_t = \mathrm{E}[\varepsilon_t^4 \mid \mathcal{F}_{t-1}] \equiv \tau$, $I = (\tau - 1)J$, and the asymptotic variance matrix becomes $(\tau - 1)J^{-1}$, as in the i.i.d. noise case. In the more general case in which the conditional kurtosis is time-varying, standard errors based on the classical QMLE (i.e., with i.i.d. noise) are invalid, as shown in our Theorem 2. The empirical evidence on conditional kurtosis supports the use of robust standard errors based on our Theorem 2. Estimation of robust standard errors can be obtained straightforwardly from our Theorem 2 in a routine fashion; see Lee and Hansen (1994) for further details.

## 3. PROOFS

For the proofs we closely follow Francq and Zakoïan (2004). We indicate the corresponding changes in their arguments resulting from the use of our conditions instead of their conditions using the i.i.d. assumption. In what follows $C$ is a generic constant that may change from one expression to another.

Define

$$L_n(\theta) = n^{-1}\sum_{t=1}^{n}\ell_t(\theta), \qquad \ell_t(\theta) = \frac{Y_t^2}{\sigma_t^2(\theta)} + \log\sigma_t^2(\theta), \tag{6}$$

where $\sigma_t^2(\theta)$ is defined in (5).

**Proof of Theorem 1.** Rewrite (5) in vector notation

$$\underline{\sigma}_t^2 = \underline{c}_t + B\underline{\sigma}_{t-1}^2, \tag{7}$$

where

$$\underline{\sigma}_t^2 = \begin{pmatrix} \sigma_t^2 \\ \sigma_{t-1}^2 \\ \vdots \\ \sigma_{t-p+1}^2 \end{pmatrix}, \qquad \underline{c}_t = \begin{pmatrix} w + \sum_{i=1}^{q}\alpha_i Y_{t-i}^2 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \qquad B = \begin{pmatrix} \beta_1 & \beta_2 & \cdots & \beta_p \\ 1 & 0 & & 0 \\ \vdots & & & \\ 0 & \cdots & 1 & 0 \end{pmatrix}. \tag{8}$$

The following intermediate results are obtained in Theorem 2.1 in Francq and Zakoïan (2004):

https://doi.org/10.1017/S0266466609090689 Published online by Cambridge University Press

(i) $\sup_{\theta\in\Theta}\left|\widetilde{L}_n(\theta)-L_n(\theta)\right|\longrightarrow_{a.s}0.$

(ii) $(\exists t\in\mathbb{Z}$ such that $\sigma_t^2(\theta)=\sigma_t^2(\theta_0)\ P_{\theta_0}-$a.s.$)\Longrightarrow\theta=\theta_0.$

(iii) $\mathrm{E}_{\theta_0}[|\ell_t(\theta_0)|]<\infty$, and if $\theta\neq\theta_0$, $\mathrm{E}_{\theta_0}[\ell_t(\theta)]>\mathrm{E}_{\theta_0}[\ell_t(\theta_0)]$.

(iv) Any $\theta\neq\theta_0$ has a neighborhood $V(\theta)$ such that $\liminf_{n\to\infty}\inf_{\theta^*\in V(\theta)}\widetilde{L}_n(\theta^*)>\mathrm{E}_{\theta_0}[\ell_1(\theta_0)]$ a.s.

An analysis of the proof of (i)–(iv) in Francq and Zakoïan (2004) reveals that the authors only use the i.i.d. assumption in the innovations for proving that $\mathrm{E}[|Y_t|^{2s}]<\infty$ for some $s>0$ (see Berkes et al., 2003, Lem. 2.3) and for establishing sufficient conditions for strict stationarity and ergodicity (i.e., the assumption that a certain matrix has strictly negative Lyapunov exponent). We have assumed these conditions directly in Assumption A1. After (i)–(iv), the proof follows standard compactness arguments. ■

**Proof of Theorem 2.** From the classical Taylor-series expansion we have

$$\begin{aligned}0&=n^{-1/2}\sum_{t=1}^{n}\widetilde{\ell}_t(\widehat{\theta}_n)\\&=n^{-1/2}\sum_{t=1}^{n}\frac{\partial}{\partial\theta}\widetilde{\ell}_t(\theta_0)+\left(n^{-1}\sum_{t=1}^{n}\frac{\partial}{\partial\theta_i\partial\theta_j'}\widetilde{\ell}_t(\theta_{ij}^*)\right)\sqrt{n}(\widehat{\theta}_n-\theta_0),\end{aligned}$$

where the $\theta_{ij}^*$ are between $\widehat{\theta}_n$ and $\theta_0$. We will show that

$$n^{-1/2}\sum_{t=1}^{n}\frac{\partial}{\partial\theta}\widetilde{\ell}_t(\theta_0)\longrightarrow_d N(0,I)$$

and

$$n^{-1}\sum_{t=1}^{n}\frac{\partial}{\partial\theta_i\partial\theta_j'}\widetilde{\ell}_t(\theta_{ij}^*)\longrightarrow_d J(i,j).$$

The theorem will straightforwardly follow. To that end, we will prove the following intermediate results:

(a) $\mathrm{E}_{\theta_0}\left[\left\|\frac{\partial\ell_t(\theta_0)}{\partial\theta}\frac{\partial\ell_t(\theta_0)}{\partial\theta'}\right\|\right]<\infty$, and $\mathrm{E}_{\theta_0}\left[\left\|\frac{\partial^2\ell_t(\theta_0)}{\partial\theta\partial\theta'}\right\|\right]<\infty$.

(b) $J$ is nonsingular, and $\mathrm{E}_{\theta_0}\left[\frac{\partial\ell_t(\theta_0)}{\partial\theta}\frac{\partial\ell_t(\theta_0)}{\partial\theta'}\right]=I$.

(c) There exists a neighborhood $V(\theta_0)$ of $\theta_0$ such that, for all $i,j,k\in\{1,\ldots,p+q+1\}$,

$$\mathrm{E}_{\theta_0}\left[\sup_{\theta\in V(\theta_0)}\left|\frac{\partial^3\ell_t(\theta)}{\partial\theta_i\partial\theta_j\partial\theta_k}\right|\right]<\infty.$$

https://doi.org/10.1017/S0266466609090689 Published online by Cambridge University Press

(d) $\|n^{-1/2}\sum_{t=1}^{n}\{(\partial\widetilde{\ell}_t(\theta_0)/\partial\theta)-(\partial\ell_t(\theta_0)/\partial\theta)\}\|\longrightarrow_d 0$ and $\sup_{\theta\in V(\theta_0)}\|n^{-1}\sum_{t=1}^{n}\{(\partial^2\widetilde{\ell}_t(\theta)/\partial\theta\partial\theta')-(\partial^2\ell_t(\theta)/\partial\theta\partial\theta')\}\|\longrightarrow_d 0$.

(e) $n^{-1/2}\sum_{t=1}^{n}(\partial/\partial\theta)\ell_t(\theta_0)\longrightarrow_d N(0,I)$.

(f) $n^{-1}\sum_{t=1}^{n}(\partial/\partial\theta_i\partial\theta_j')\ell_t(\theta_{ij}^*)\longrightarrow_{a.s.} J(i,j)$.

We start by proving (a). The derivatives of $\ell_t(\theta_0)$ are

$$\frac{\partial\ell_t(\theta_0)}{\partial\theta}=\left(1-\frac{Y_t^2}{\sigma_t^2(\theta_0)}\right)\frac{\partial\sigma_t^2(\theta_0)}{\partial\theta}\frac{1}{\sigma_t^2(\theta_0)}\quad\text{and}$$

$$\frac{\partial^2\ell_t(\theta_0)}{\partial\theta\partial\theta'}=\left(1-\frac{Y_t^2}{\sigma_t^2(\theta_0)}\right)\frac{\partial^2\sigma_t^2(\theta_0)}{\partial\theta\partial\theta'}\frac{1}{\sigma_t^2(\theta_0)}+\left(2\frac{Y_t^2}{\sigma_t^2(\theta_0)}-1\right)\frac{\partial\sigma_t^2(\theta_0)}{\partial\theta}\frac{\partial\sigma_t^2(\theta_0)}{\partial\theta'}\frac{1}{\sigma_t^4(\theta_0)}. \tag{9}$$

By Hölder's inequality, for any $\gamma>1$,

$$\begin{aligned}&\mathrm{E}_{\theta_0}\left[\left\|\frac{\partial\ell_t(\theta_0)}{\partial\theta}\frac{\partial\ell_t(\theta_0)}{\partial\theta'}\right\|\right]\\&\quad\le\left(\mathrm{E}_{\theta_0}\left[\left|\left(1-\frac{Y_t^2}{\sigma_t^2(\theta_0)}\right)\right|^{2(1+\gamma)}\right]\right)^{1/(1+\gamma)}\\&\quad\times\left(\mathrm{E}_{\theta_0}\left[\left\|\frac{1}{\sigma_t^4(\theta_0)}\frac{\partial\sigma_t^2(\theta_0)}{\partial\theta}\frac{\partial\sigma_t^2(\theta_0)}{\partial\theta'}\right\|^{(1+\gamma)/\gamma}\right]\right)^{\gamma/(1+\gamma)}.\end{aligned} \tag{10}$$

Now, it is straightforward to show that (see Francq and Zakoïan, 2004)

$$\frac{1}{\sigma_t^2(\theta_0)}\frac{\partial\sigma_t^2(\theta_0)}{\partial w}\le C,\qquad\frac{1}{\sigma_t^2(\theta_0)}\frac{\partial\sigma_t^2(\theta_0)}{\partial\alpha_i}\le\frac{1}{\alpha_i},\qquad i=1,\dots,q. \tag{11}$$

Whereas from (7),

$$\frac{\partial\underline{\sigma}_t^2(\theta_0)}{\partial\beta_j}=\sum_{k=1}^{\infty}\left\{\sum_{i=1}^{k}B^{i-1}B^{(j)}B^{k-i}\right\}\underline{c}_{t-k},$$

where $B^{(j)}$ is a $p\times p$ matrix with $(1,j)$th element 1 and all other elements equal to zero. Hence from the positivity of the parameters, $\beta_j B^{(j)}\le B$, and hence

$$\beta_j\frac{\partial\underline{\sigma}_t^2(\theta_0)}{\partial\beta_j}\le\sum_{k=1}^{\infty}\left\{\sum_{i=1}^{k}B^{i-1}BB^{k-i}\right\}\underline{c}_{t-k}=\sum_{k=1}^{\infty}kB^k\underline{c}_{t-k}.$$

https://doi.org/10.1017/S0266466609090689 Published online by Cambridge University Press

Notice that Assumption A3 implies that $\|B^k\| \leq C\rho^k$, for some $\rho \in (0,1)$ and all $k$. Using the simple inequality $x/(1+x) \leq x^s$, for $x \geq 0$ and all $s \in (0,1]$, we obtain for any $r > 1$ and $0 < \delta \leq r$,

$$\left|\frac{1}{\sigma_t^2(\theta_0)}\frac{\partial\sigma_t^2(\theta_0)}{\partial\beta_j}\right|^r \leq \frac{1}{\beta_j^r}\left(\sum_{k=1}^{\infty}\frac{kB^k(1,1)\underline{c}_{t-k}(1)}{w+B^k(1,1)\underline{c}_{t-k}(1)}\right)^r$$

$$\leq \frac{1}{\beta_j^r}\left(\sum_{k=1}^{\infty}k\left(\frac{B^k(1,1)\underline{c}_{t-k}(1)}{w}\right)^{\frac{\delta}{r}}\right)^r$$

$$\leq \frac{C}{w^\delta\beta_j^r}\left(\sum_{k=1}^{\infty}k\rho^{\frac{\delta k}{r}}\underline{c}_{t-k}^{\frac{\delta}{r}}(1)\right)^r. \tag{12}$$

Now, using Hölder's inequality, one can bound (12) by

$$\frac{C}{w^\delta\beta_j^r}\left(\sum_{k=1}^{\infty}\rho^{\frac{\delta k}{2}}\underline{c}_{t-k}^{\delta}(1)\right)\left(\sum_{k=1}^{\infty}k^{r/(r-1)}\rho^{\delta k/(2(r-1))}\right)^{(r-1)}.$$

By Assumption A1, $\mathrm{E}\left[\underline{c}_{t-k}^s(1)\right] < \infty$, for some $s \in (0,1]$, and hence choosing $\delta = s$, we conclude that for all $r > 1$

$$\mathrm{E}\left[\left|\frac{1}{\sigma_t^2(\theta_0)}\frac{\partial\sigma_t^2(\theta_0)}{\partial\beta_j}\right|^r\right] < \infty.$$

The last statement, (11), and Assumption A4 imply that the moment in (10) is bounded. By similar arguments it holds that $\mathrm{E}_{\theta_0}\left[\left\|\frac{\partial^2\ell_t(\theta_0)}{\partial\theta\partial\theta'}\right\|\right] < \infty$.

The nonsingularity of $J$ in (b) was proved in Francq and Zakoïan (2004). To prove (c), differentiating (9) and applying the Cauchy–Schwartz and Hölder inequalities (see Francq and Zakoïan, 2004), it is sufficient to bound for $\delta > 1$

$$\mathrm{E}\left[\left|\sup_{\theta\in V(\theta_0)}\frac{Y_t^4}{\sigma_t^4(\theta)}\right|\right] \leq \left(\mathrm{E}_{\theta_0}\left[|\varepsilon_t|^{4(1+\delta)}\right]\right)^{1/(1+\delta)}$$

$$\times\left(\mathrm{E}_{\theta_0}\left[\sup_{\theta\in V(\theta_0)}\left|\frac{\sigma_t^4(\theta_0)}{\sigma_t^4(\theta)}\right|^{(1+\delta)/\delta}\right]\right)^{\delta/(1+\delta)}.$$

Let $\Theta^*$ be a compact set containing $\theta_0$ and included in the $int(\Theta)$. Denote by $B_0$ the matrix defined in (8) evaluated at $\theta_0$. For all $\eta > 0$, there exists a neighborhood $V(\theta_0)$ of $\theta_0$, with $V(\theta_0) \subset \Theta^*$, such that $B_0 \leq (1+\eta)B$ for all $\theta \in V(\theta_0)$. From

https://doi.org/10.1017/S0266466609090689 Published online by Cambridge University Press

(7) we obtain

$$\sigma_t^2(\theta) = w \sum_{k=0}^{\infty} B^k(1,1) + \sum_{i=1}^{q} \alpha_i \left\{ \sum_{k=1}^{\infty} B^k(1,1) Y_{t-k-i}^2 \right\}.$$

Hence, using the same arguments as in (12), we conclude that for any $r > 1$,

$$\mathrm{E}_{\theta_0} \left[ \sup_{\theta \in V(\theta_0)} \left| \frac{\sigma_t^2(\theta_0)}{\sigma_t^2(\theta)} \right|^r \right] < \infty.$$

The rest of the proof of (c)–(f) follows the same steps as in Francq and Zakoïan (2004) in a routine fashion, and hence it is omitted. After proving (a)–(f) it suffices to apply the Slutsky lemma. ■

## REFERENCES


Berkes, I. & L. Horváth (2003) The rate of consistency of the quasi-maximum likelihood estimator. *Statistics and Probability Letters* 61, 133–143.

Berkes, I. & L. Horváth (2004) The efficiency of the estimators of the parameters in GARCH processes. *Annals of Statistics* 32, 633–655.

Berkes, I., L. Horváth & P. Kokoszka (2003) GARCH processes: Structure and estimation. *Bernoulli* 9, 201–227.

Bollerslev, T. (1986) Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* 31, 307–327.

Brooks, C., S.P. Burke, S. Heravi & G. Persand (2005) Autoregressive conditional kurtosis. *Journal of Financial Econometrics* 3, 399–421.

Dahl, C.M. & E.M. Iglesias (2007) Asymptotic Normality of the QMLE for Nonstationary GARCH with Serially Dependent Innovations. Manuscript, Michigan State University.

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

Francq, C. & J.M. Zakoïan (2004) Maximum likelihood estimation of pure GARCH and ARMA-GARCH. *Bernoulli* 10, 605–637.

Gallant, A.R., D.A. Hsieh & G. Tauchen (1991) On fitting a recalcitrant series: The pound/dollar exchange rate, 1974-1983. In W.A. Barnett, J. Powell & G. Tauchen (eds.). *Nonparametric and Semiparametric Methods in Econometrics and Statistics*, pp. 199–240. Cambridge University Press.

Hall, P. & Q.W. Yao (2003) Inference in ARCH and GARCH models. *Econometrica* 71, 285–317.

Hansen, B. (1994) Autoregressive conditional density estimation. *International Economic Review* 35, 705–730.

Harvey, C. & A. Siddique (1999) Autoregressive conditional skewness. *Journal of Financial and Quantitative Analysis* 34, 465–487.

Harvey, C. & A. Siddique (2000) Conditional skewness in asset pricing tests. *Journal of Finance* 55, 1263–1296.

Jensen, S.T. & A. Rahbek (2004a) Asymptotic normality of the QML estimator of ARCH in the nonstationary case. *Econometrica* 72, 641–646.

Jensen, S.T. & A. Rahbek (2004b) Asymptotic inference for nonstationary GARCH. *Econometric Theory* 20, 1203–1226.

Jondeau, E. & M. Rockinger (2003) Conditional volatility, skewness, and kurtosis: Existence, persistence, and comovements. *Journal of Economic Dynamics and Control* 27, 1699–1737.

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


https://doi.org/10.1017/S0266466609090689 Published online by Cambridge University Press

Leon, A., G. Rubio & G. Serna (2005) Autoregressive conditional volatility, skewness and kurtosis: *Quarterly Review of Economics and Finance* 45, 599–618.

Li, W.K., S. Ling & M. McAleer (2002) A survey of recent theoretical results for time series models with GARCH errors. *Journal of Economic Survey* 16, 245–269.

Ling, S. (2007) Self-weighted and local quasi-maximum likelihood estimators for ARMA-GARCH/IGARCH models. *Journal of Econometrics* 140, 849–873.

Linton, O., J. Pan & H. Wang (2009) Estimation for a non-stationary semi-strong GARCH(1,1) model with heavy-tailed errors. *Econometric Theory* 25, forthcoming.

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* 6, 575–596.

Mikosch, T. & D. Straumann (2002) Whittle estimation in a heavy-tailed GARCH(1,1) model. *Stochastic Processes and Their Applications* 100, 187–222.

Premaratne, G. & A.K. Bera (2001) Modeling Asymmetry and Excess Kurtosis in Stock Return Data. Working paper 00-123, Department of Economics, University of Illinois, Champaign.

Robinson, P.M. & P. Zaffaroni (2006) Pseudo-maximum likelihood estimation of ARCH($\infty$) models. *Annals of Statistics* 34, 1049–1074.

Straumann, D. (2005) *Estimation in Conditionally Heteroscedastic Time Series Models*. Lecture Notes in Statistics 181. Springer-Verlag.

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

https://doi.org/10.1017/S0266466609090689 Published online by Cambridge University Press