Strathprints Institutional Repository

Wang, Hui and Pan, Jiazhu (2014) Normal mixture quasi maximum likelihood estimation for non-stationary TGARCH(1,1) models. Statistics and Probability Letters, 91. 117–123. ISSN 0167-7152 , http://dx.doi.org/10.1016/j.spl.2014.03.027

Normal Mixture Quasi Maximum Likelihood Estimation for Non-Stationary TGARCH(1, 1) Models

HUI WANG[1] and JIAZHU PAN [2*]

[1]*School of Finance, Central University of Finance and Economics, China.*

[2]*Department of Mathematics and Statistics, University of Strathclyde, UK .*

Abstract: Although quasi maximum likelihood estimator based on Gaussian density (G-QMLE) is widely used to estimate GARCH-type models, it does not perform successfully when error distribution is either skewed or leptokurtic. This paper proposes normal mixture quasi-maximum likelihood estimator (NM-QMLE) for non-stationary TGARCH(1, 1) models. We show that, under mild regular conditions, there is no consistent estimator for the intercept, and the proposed estimator for any other parameter is consistent.

AMS 2000 subject classifications. 62P05, 62M10

Keywords: Non-stationary TGARCH models; NM-QMLE; consistency

1 Introduction

Since the seminal papers by Engle (1982) and Bollerslev (1986), generalized autoregressive conditional heteroscedastic (GARCH) models have been proved particularly valuable in modelling time varying volatility. Earlier literature on inference of GARCH models is based on least-squares estimation (LSE) and maximum likelihood estimation (MLE) under the assumption that the distribution of innovations is standard Gaussian distribution, see Engle (1982) and Bollerslev (1986). Then the Gaussian quasi-maximum likelihood estimation (G-QMLE) became popular due to its simplicity. Regarding to the asymptotic inference of G-QMLE for stationary GARCH models, the consistency and asymptotic normality have been established under different conditions, see Lee and Hansen (1994), Lumsdaine (1996), Berkes et al. (2003), Hall and Yao (2003), and Francq and Zakoïan (2004).

Although G-QMLE behaves appropriately in financial applications, empirical studies have shown that when using normal innovations, the tails of the fitted GARCH(1, 1) models seem

*Address correspondence to Jiazhu Pan, e-mail: jiazhu.pan@strath.ac.uk

to be much thinner than the tails apparent in the data, see for instance Mikosch and Stărică (2000). In fact, G-QMLE does not perform successfully in cases where the error distribution of GARCH-type models is either skewed or leptokutic. In the literature, there are two approaches correcting the defect of G-QMLE: the use of mixture type models and estimating the models based on more general quasi likelihood distributions which can capture the skewness or heavy-tail of the errors. The first approach was employed in Vlaar and Palm (1993), Haas et al. (2004), Zhang et al. (2006) etc.. The second approach was used for instance in Bollerslev (1987), who considered Student's t-GARCH models; Berks and Horváth (2004), who proposed a class of QMLE for stationary GARCH models; Lee and Lee (2009), who proposed the normal mixture QMLE (NM-QMLE) which is obtained from the normal mixture quasi-likelihood (see Ha and Lee (2011) for the ARMA-GARCH model). Since the mixture type models are not GARCH models by definition, we focus on the second approch.

Nonstationarity in the volatility process has been well documented for macroeconomic and financial time series data, see Loretan and Phillips (1994) and Hwang et al. (2010). When it occurs, the prevalent stationary and conditional approaches such as GARCH-type or stochastic volatility (SV) models are inadequate and may lead to model mis-specification or poor volatility forecasts, see Stărică et al. (2005). Jensen and Rahbek (2004 a, 2004 b) are the first to consider the asymptotic theory of G-QMLE for non-stationary ARCH/GARCH(1,1) models. For further studies of estimation for non-stationary GARCH type models, see Linton et al. (2010), Francq and Zakoïan (2012) and Francq and Zakoïan (2013). On the other hand, standard GARCH models assume that positive and negative error terms have a symmetric effect on the volatility. In practice this assumption is frequently violated, in particular by stock returns, in that the volatility increases more after bad news than after good news. Among the asymmetric GARCH models, threshold GARCH (TGARCH) model is one of the most popular models in the literature, see Glosten et. al (1993) and Li and Li (1996) among others. Note that nonstationary TGARCH models capture the non-stationarity and asymmetry of the volatility of time series data simultaneously. This motivates us to study the estimation problem of nonstationary TGARCH models when the errors are skewed or leptokutic.

In this paper, adopting the ideas of Lee and Lee (2009) we propose NM-QMLE for non-stationry TGARCH(1, 1) model and demonstrate the validity of NM-QMLE by verifying its consistency. The rest of this paper is organized as follows. Section 2 presents the estimation methodology and main results. In section 3, we provide the proof of the theorems presented in section 2.

Throughout this paper, $\|x\|$ denotes the norm $\sqrt{x_1^2+\cdots+x_m^2}$ for a m-dimensional vector $x=(x_1,\cdots,x_m)'$, and $\|X\|_p=[E(|X|^p)]^{1/p}$ is the L_p norm for a random variable X.

2 Methodology and main results

Let us consider the TGARCH(1, 1) model defined by

$$X_t=\sigma_t\varepsilon_t \quad \text{and} \quad \sigma_t^2=\omega+\alpha_+(X_{t-1}^+)^2+\alpha_-(X_{t-1}^-)^2+\beta\sigma_{t-1}^2, \tag{2.1}$$

with intial values X_0 and $\sigma_0\geq 0$, where $\omega>0$, $\alpha_+\geq 0$, $\alpha_-\geq 0$, $\beta\geq 0$ and using the notation $x^+=\max(x,0)$, $x^-=\max(-x,0)$. In this model, $\{\varepsilon_t\}$ is a sequence of independent and identically distributed (iid) random variables with $E\varepsilon_t=0$ and $E\varepsilon_t^2=1$, such that ε_t is independent of $\{X_{t-k},k\geq 1\}$ for all t. According to Pan et al. (2008), there exists a unique strictly stationary and ergodic solution to model (2.1) if and only if

$$E\log\left[\alpha_+(\varepsilon_{t-1}^+)^2+\alpha_-(\varepsilon_{t-1}^-)^2+\beta\right]<0.$$

The parameter of model (2.1) is then $\phi=(\alpha_+,\alpha_-,\beta,\omega)'$ with true value $\phi_0=(\alpha_{0+},\alpha_{0-},\beta_0,\omega_0)'$. Let $\varphi=(\alpha_+,\alpha_-,\beta)'$ with the true value $\varphi_0=(\alpha_{0+},\alpha_{0-},\beta_0)'$. We wish to estimate φ_0 from observations $\{X_t,t=1,\cdots,n\}$ in the non-stationary case.

In order to obtain the NM-QMLE for model (2.1), a family of normal mixture densities is introduced first. The normal mixture (NM) density with s components is of the form

$$g_\vartheta(y)=\sum_{k=1}^{s}p_kf(y;\mu_k,\varrho_k), \tag{2.2}$$

where $\vartheta=(p_1,\cdots,p_{s-1},\mu_1,\cdots,\mu_{s-1},\varrho_1,\cdots,\varrho_{s-1})'$ and

$$f(y;\mu_k,\varrho_k)=\frac{1}{\sqrt{2\pi}\varrho_k}\exp\Big\{-\frac{(y-\mu_k)^2}{2\varrho_k^2}\Big\}$$

satisfying

$$\sum_{k=1}^{s}p_k=1,\quad \sum_{k=1}^{s}p_k\mu_k=0 \quad \textit{and} \quad \sum_{k=1}^{s}p_k(\mu_k^2+\varrho_k^2)=1. \tag{2.3}$$

In general, the s component normal mixture distribution is not identifiable, so we need the following identification condition as in Lee and Lee (2009). Denote $\tilde{\Theta}$ the set of all ϑ satisfying

(2.3) and $\mathcal{G} = \{g_\vartheta : \vartheta \in \tilde{\Theta}\}$. We assume $\mathcal{G}$ is identifiable, that is, for any $g_{\vartheta_1} \in \mathcal{G}$ and $g_{\vartheta_2} \in \mathcal{G}$ with $\vartheta_i = (p_1^{(i)}, \cdots, p_{s-1}^{(i)}, \mu_1^{(i)}, \cdots, \mu_{s-1}^{(i)}, \varrho_1^{(i)}, \cdots, \varrho_{s-1}^{(i)})'$,

$$g_{\vartheta_1} \equiv g_{\vartheta_2} \quad a.e. \Longleftrightarrow \sum_{k=1}^{s} p_k^{(1)} \delta_{(\mu_k^{(1)}, \varrho_k^{(1)})} = \sum_{k=1}^{s} p_k^{(2)} \delta_{(\mu_k^{(2)}, \varrho_k^{(2)})}, \tag{2.4}$$

where $\delta_{(\mu_k, \varrho_k)}(\cdot)$ is an indicator function with $\delta_{(\mu_k^{(i)}, \varrho_k^{(i)})}(\mu_k^{(i)}, \varrho_k^{(i)}) = 1$ and $\delta_{(\mu_k^{(i)}, \varrho_k^{(i)})}(x, y) = 0$ for all $(x, y) \neq (\mu_k^{(i)}, \varrho_k^{(i)})$, $i = 1, 2$. Furthermore, we assume $\mathcal{G}$ is nondegenerate, that is, any s-component normal mixture density in $\mathcal{G}$ can not be represented as a mixture with the number of components less than s.

The idea behind the NM-QMLE is that the estimator is constructed as if the innovations ε_t are normal-mixture random variables. Conditionally on initial values X_0, σ_0, the normal mixture quasi-likelihood is given by

$$L_n(\phi, \vartheta) = \prod_{t=1}^{n} \Big\{ \sum_{k=1}^{s} p_k \frac{1}{\sqrt{2\pi \varrho_k^2 \sigma_t^2(\phi)}} \exp\Big\{ - \frac{(X_t - \mu_k \sigma_t(\phi))^2}{2\varrho_k^2 \sigma_t^2(\phi)} \Big\} \Big\}, \tag{2.5}$$

where

$$\sigma_t^2(\phi) = \omega + \alpha_+ (X_{t-1}^+)^2 + \alpha_- (X_{t-1}^-)^2 + \beta \sigma_{t-1}^2(\phi). \tag{2.6}$$

A natural idea is to obtain an estimator of ϕ by maximizing $L_n^{NM}(\vartheta, \phi)$, and nuisance parameters ϑ are also estimated at the same time. Since the density function g of ε_t may be not in $\mathcal{G}$, then what does the true value ϑ_0 of ϑ mean? One may hope ϑ_0 can minimize the discrepancy between the true innovation density g and the quasi likelihood normal mixture density in the sense of Kullback-Leibler Information Distance (KLID), see White (1982). Thus, we define the true value $\vartheta_0 = (p_{10}, \cdots, p_{(s-1)0}, \mu_{10}, \cdots, \mu_{(s-1)0}, \varrho_{10}, \cdots, \varrho_{(s-1)0})'$ as follows,

$$\vartheta_0 = \{\vartheta \in \tilde{\Theta} : d(g, g_{\vartheta_0}) = \min_{\vartheta \in \tilde{\Theta}} d(g, g_\vartheta)\}, \tag{2.7}$$

where $d(g, g_\vartheta) = \int g(x)\big(\log g(x) - \log g_\vartheta(x)\big)dx$ is the KLID between g and g_ϑ. Note that ϑ_0 here only depends on the KLID of the two densities under consideration. Let Θ be a compact subset of $(0, \infty)^4 \times \tilde{\Theta}$. We also set $\theta = (\phi', \vartheta')'$ which belongs to the parameter space Θ. Now we define the NM-QMLE of the parameter $\theta_0 = (\phi_0', \vartheta_0')'$ by

$$\hat{\theta}_n = (\hat{\omega}_n, \hat{\varphi}_n, \hat{\vartheta}_n)' = \arg\max_{\theta \in \Theta} L_n(\phi, \vartheta) = \arg\min_{\theta \in \Theta} l_n(\theta) \tag{2.8}$$

where

$$l_n(\theta) = -n^{-1} \log L_n(\phi, \vartheta) = n^{-1} \sum_{t=1}^{n} W_t(\theta) \quad \text{and} \quad W_t(\theta) = -\log \Big\{ \frac{1}{\sigma_t(\phi)} g_\vartheta\Big(\frac{X_t}{\sigma_t(\phi)}\Big) \Big\} \tag{2.9}$$

and $L_n(\phi, \vartheta)$ is defined in (2.5). Note that normal mixture structure includes an additional parameter ϑ, and this adds complexity to $l_n(\theta)$, thereby requiring EM algorithm. For the computation of NM-QMLE for non-stationary TGARCH(1, 1) models, one may adopt the algorithm in Hwang et al.(2010).

Write $\psi = (\varphi', \vartheta')'$ and let $\hat{\psi}_n = (\hat{\varphi}_n', \hat{\vartheta}_n')'$. In order to obtain asymptotic properties of $\hat{\psi}_n$, we need the following regularity conditions:

A1. $\gamma_0 = E \log \left[\alpha_{0+}(\varepsilon_{t-1}^+)^2 + \alpha_{0-}(\varepsilon_{t-1}^-)^2 + \beta_0\right] \geq 0$;

A2. $P(\varepsilon_t = 0) = 0$. Furthermore, the support of ε_t contains at least 3 points and is not concentrated on the positive or the negative line.

A3. The functional $\vartheta_0 \in \tilde{\Theta}$ is essentially unique.

Given below are the asymptotic properties for $\hat{\psi}_n$ and $\hat{\omega}_n$. The proofs are given in Section 3.

Theorem 1. *Suppose $\mathcal{G}$ is identifiable and nondegenerate. Let assumptions A1 - A3 hold. Then the NM-QMLE defined in (2.8) satisfies the following properties*

(i) When $\gamma_0 > 0$, we have $\hat{\psi}_n \to \psi_0$, a.s. as $n \to \infty$.

(ii) When $\gamma_0 = 0$, if $\forall \theta \in \Theta$, $\beta < \left\| 1/\left(\alpha_{0+}(\varepsilon_t^+)^2 + \alpha_{0-}(\varepsilon_t^-)^2 + \beta_0\right) \right\|_p^{-1}$ for some $p > 1$, we have $\hat{\psi}_n \to \psi_0$, in probability as $n \to \infty$.

Theorem 2. *Suppose $\mathcal{G}$ is identifiable and nondegenerate, and assumptions A1 - A3 hold. Assume ε_t has continuous distribution, and Θ contains two arbitraily close points $\theta_1 = (\omega_1, \psi_1)'$ and $\theta_2 = (\omega_2, \psi_1)'$ such that $E \log \left[\alpha_{1+}(\varepsilon_{t-1}^+)^2 + \alpha_{1-}(\varepsilon_{t-1}^-)^2 + \beta_1\right] > 0$. There exists no consitent estimator of $\theta_0 \in \Theta$.*

Remark 1. One may suspect the existence of such a ϑ_0. In fact, since the parameter space for the family of univariate normal mixtures can be embedded in a compact set by using the transformation method given by Beran (1977, pp. 447-448), the existence of ϑ_0 is guaranteed if $d(g, g_\vartheta)$ is a continuous function with respect to ϑ, which is easily verified. The uniqueness of ϑ_0 is intestable and difficult to check in general KLIC, see also Lee and Lee (2009).

Remark 2. The asymptotic normality of $\hat{\theta}_n$ is difficult to obtain since the first derivative of the log normal mixture likelihood function can not be approximated by a martingale difference under the constraint (2.3) to $\tilde{\Theta}$. We leave this for further investigation.

Remark 3. Theorem 1 and Theorem 2 can be extended to the more general non-stationary models of Box-Cox transformed threshold GARCH(1, 1) models, which were proposed by Hwang and Kim (2004) and Hwang and Basawa (2004), see also Pan et al.(2008). To fix ideas, we only consider non-stationary TGARCH(1, 1) models in this paper.

3 Proofs

In this section, we provide the proofs for the Theorems prensented in section 2. To apply the standard proof for consistency of $\hat{\psi}_n$, we use a strictly stationary series ν_t to approximate the non-stationary $\sigma_t^2(\phi)/\sigma_t^2$, where and in the following $\sigma_t^2 \equiv \sigma_t^2(\phi_0)$. Define the following process

$$\nu_t(\varphi) \quad = \sum_{j=1}^{\infty} \frac{\eta_{t-j}(\varphi)}{e_{0,t-j}} \prod_{k=1}^{j-1} \frac{\beta}{e_{0,t-k}} \tag{3.1}$$

with the convention $\prod_{k=1}^{j-1} = 1$ when $j \leq 1$ and

$$\eta_t(\varphi) = \alpha_+(\varepsilon_t^+)^2 + \alpha_-(\varepsilon_t^-)^2, \quad e_{0,t} = \alpha_{0+}(\varepsilon_t^+)^2 + \alpha_{0-}(\varepsilon_t^-)^2 + \beta_0.$$

Let $\Theta_0 = \{\theta \in \Theta : \beta < e^{\gamma_0}\}$ and $\Phi_p = \{\theta \in \Theta : \beta < \|1/e_{0,1}\|_p^{-1}\}$.

Lemma 1. *Suppose assumptions A1 and A2 hold. Denote $Q_n(\theta) = l_n(\theta) - l_n(\theta_0)$. We have*

(i)For any compact subset Θ_0^* of Θ_0,

$$\lim_{n\to\infty} \sup_{\theta\in\Theta_0^*} \Big| Q_n(\theta) - E\log\Big[g_{\vartheta_0}(\varepsilon_t)/\big(\nu_t^{-1/2}(\varphi)g_\vartheta(\nu_t^{-1/2}(\varphi)\varepsilon_t)\big)\Big]\Big| = 0 \ \ a.s.;$$

(ii) For any compact subset Φ_p^* of Φ_p,

$$\lim_{n\to\infty} \sup_{\theta\in\Phi_p^*} \Big| Q_n(\theta) - E\log\Big[g_{\vartheta_0}(\varepsilon_t)/\big(\nu_t^{-1/2}(\varphi)g_\vartheta(\nu_t^{-1/2}(\varphi)\varepsilon_t)\big)\Big]\Big| = 0 \ \ in \ L^p.$$

Proof. (i) It is easy to verify that

$$Q_n(\theta) = \frac{1}{n}\sum_{t=1}^{n}\Big[\log\big(\frac{1}{\sigma_t}g_{\vartheta_0}(\varepsilon_t)\big) - \log\Big(\frac{1}{\sigma_t(\phi)}g_\vartheta\big(\frac{X_t}{\sigma_t(\phi)}\big)\Big)\Big] = Q_{1n}(\psi) + Q_{2n}(\theta),$$

where

$$Q_{1n}(\psi) = \frac{1}{n}\sum_{t=1}^{n}\log\frac{g_{\vartheta_0}(\varepsilon_t)\nu_t^{1/2}(\varphi)}{g_\vartheta\big(\nu_t^{-1/2}(\varphi)\varepsilon_t\big)} \quad \text{and} \quad Q_{2n}(\theta) = \frac{1}{n}\sum_{t=1}^{n}\log\frac{\sigma_t(\phi)g_\vartheta\big(\nu_t^{-1/2}(\varphi)\varepsilon_t\big)}{\sigma_t\nu_t^{1/2}(\varphi)g_\vartheta\big(X_t/\sigma_t(\varphi)\big)}.$$

Thus, to prove (i), we only need to establish that

$$\lim_{n\to\infty} \sup_{\theta\in\Theta_0^*} \Big| Q_{1n}(\psi) - E\log\Big[g_{\vartheta_0}(\varepsilon_t)/\big(\nu_t^{-1/2}(\varphi)g_\vartheta(\nu_t^{-1/2}(\varphi)\varepsilon_t)\big)\Big]\Big| = 0 \text{ a.s.} \tag{3.2}$$

and

$$\lim_{n\to\infty}\sup_{\theta\in\Theta_0^*}\big|Q_{2n}(\theta)\big| = 0 \text{ a.s..} \tag{3.3}$$

By Lemma 7.1 of Francq and Zakoïan (2013), $\nu_t(\varphi)$ is stationary and ergodic. Therefore, the ergodic theorem implies that

$$Q_{1n}(\psi) \to E\log\Big[g_{\vartheta_0}(\varepsilon_t)/\big(\nu_t^{-1/2}(\varphi)g_\vartheta(\nu_t^{-1/2}(\varphi)\varepsilon_t)\big)\Big] \text{ a.s..} \tag{3.4}$$

Noting that Θ_0^* is compact and $\beta < e^{\gamma_0}$ for any $\theta\in\Theta_0^*$, we have for any $k>0$

$$E\sup_{\theta\in\Theta_0^*}\left|\frac{1}{\nu_t(\varphi)}\frac{\partial\nu_t(\varphi)}{\partial\varphi_i}\right|^k \le C, \quad i=1,2,3. \tag{3.5}$$

By (3.5), Lemma 7.3 of Francq and Zakoïan (2013) and some straight calculations, we have

$$\begin{aligned} & E\sup_{\theta\in\Theta_0^*}\left|\frac{\partial Q_{1n}(\psi)}{\partial\varphi_i}\right| \\ = \ & E\sup_{\theta\in\Theta_0^*}\left|\frac{1}{2n}\sum_{t=1}^n\frac{1}{\nu_t(\varphi)}\frac{\partial\nu_t(\varphi)}{\partial\varphi_i}\Big[1+\frac{\varepsilon_t\nu_t^{-1/2}(\varphi)}{g_\vartheta(\nu_t^{-1/2}(\varphi)\varepsilon_t)}\frac{\partial g_\vartheta(x)}{\partial x}\Big|_{x=\nu_t^{-1/2}(\varphi)\varepsilon_t}\Big]\right| < \infty \end{aligned} \tag{3.6}$$

for $i=1,2,3$. Similarly, we can obtain

$$E\sup_{\theta\in\Theta_0^*}\left|\frac{\partial Q_{1n}(\psi)}{\partial\vartheta_i}\right| = E\sup_{\theta\in\Theta_0^*}\left|-\frac{1}{n}\sum_{t=1}^n\frac{\varepsilon_t\nu_t^{-1/2}(\varphi)}{g_\vartheta(\nu_t^{-1/2}(\varphi)\varepsilon_t)}\frac{\partial g_\vartheta(\nu_t^{-1/2}(\varphi)\varepsilon_t)}{\partial\vartheta_i}\right| < \infty \tag{3.7}$$

for $i=1,\cdots,3(s-1)$. For any $\theta_1=(\omega_1,\psi_1')'$, $\theta_2=(\omega_2,\psi_2')'\in\Theta_0^*$, by the mean value theorem, (3.6) and (3.7) imply that

$$\sup_{\theta_1,\theta_2\in\Theta_0^*}\big|Q_{1n}(\psi_1)-Q_{1n}(\psi_2)\big| \le \sup_{\theta_1,\theta_2\in\Theta_0^*}\left\|\frac{\partial Q_{1n}(\psi_*)}{\partial\psi}\right\|\big\|\psi_1-\psi_2\big\| = O(1)\sup_{\theta_1,\theta_2\in\Theta_0^*}\big\|\psi_1-\psi_2\big\|$$

with ψ_* between ψ_1 and ψ_2, which shows that $Q_{1n}(\psi)$ is equicontinuous. Combining this fact, (3.4) and the compact of Θ_0^*, (3.2) hold.

Next, we deal with $Q_{2n}(\theta)$. By Lemma 7.1 and 7.3 of Francq and Zakoïan (2013) and Lemma A.3 of Francq and Zakoïan (2012) , we have

$$\sup_{\theta\in\Theta_0^*}\left|\frac{\sigma_t^2(\phi)}{\sigma_t^2}-\nu_t(\varphi)\right| \to 0 \quad a.s., \ \sup_{\theta\in\Theta_0^*}\frac{\sigma_t^2}{\sigma_t^2(\phi)} \le V_t \text{ and } E\sup_{\theta\in\Theta_0^*}\Big[V_t^k+\nu_t^{-k}(\varphi)\Big] < \infty$$

for any $k>0$ and V_t is a stationary and ergodic process. Hence

$$\sup_{\theta\in\Theta_0^*}\left|\frac{1}{n}\sum_{t=1}^n\log\frac{g_\vartheta\big(\nu_t^{-1/2}(\varphi)\varepsilon_t\big)}{g_\vartheta\big(X_t/\sigma_t(\varphi)\big)}\right|$$

$$
\begin{aligned}
&= \sup_{\theta\in\Theta_0^*}\left|\frac{1}{n}\sum_{t=1}^{n}\Big(\frac{\sigma_t}{\sigma_t(\phi)}-\nu_t^{-1/2}(\varphi)\Big)\varepsilon_t\frac{1}{g_\vartheta(x_*)}\frac{\partial g_\vartheta(x_*)}{\partial x}\right|_{x_*\in\Big(\min\{\frac{\sigma_t}{\sigma_t(\phi)},\nu_t^{-1/2}(\varphi)\},\max\{\frac{\sigma_t}{\sigma_t(\phi)},\nu_t^{-1/2}(\varphi)\}\Big)}\\
&\le \frac{C}{n}\sum_{t=1}^{n}\left[\varepsilon_t^2\sup_{\theta\in\Theta_0^*}\left|\frac{\sigma_t}{\sigma_t(\phi)}+\nu_t^{-1/2}(\varphi)\right|+|\varepsilon_t|\right]\sup_{\theta\in\Theta_0^*}\left|\frac{\sigma_t}{\sigma_t(\phi)}-\nu_t^{-1/2}(\varphi)\right|\\
&\le \frac{C}{n}\sum_{t=1}^{n}\left[\big(V_t^{1/2}+\nu_t^{-1/2}(\underline{\varphi})\big)\varepsilon_t^2+|\varepsilon_t|\right]\sup_{\theta\in\Theta_0^*}\left|\frac{\sigma_t}{\sigma_t(\phi)}-\nu_t^{-1/2}(\varphi)\right|\\
&\le \frac{C\varepsilon}{n}\sum_{t=1}^{n}\left[\big(V_t^{1/2}+\nu_t^{-1/2}(\underline{\varphi})\big)\varepsilon_t^2+|\varepsilon_t|\right]
\end{aligned}
$$

for any $\varepsilon>0$, when n is large enough, where $\underline{\varphi}=\big(\inf_{\theta\in\Theta_0^*}(\alpha_+),\inf_{\theta\in\Theta_0^*}(\alpha_-),\inf_{\theta\in\Theta_0^*}(\beta)\big)'$ and C is a constant. This implies

$$
\lim_{n\to\infty}\sup_{\theta\in\Theta_0^*}\left|\frac{1}{n}\sum_{t=1}^{n}\log\frac{g_\vartheta\big(\nu_t^{-1/2}(\varphi)\varepsilon_t\big)}{g_\vartheta\big(X_t/\sigma_t(\varphi)\big)}\right|=0,\quad a.s. \tag{3.8}
$$

Similarly, we can show that

$$
\lim_{n\to\infty}\sup_{\theta\in\Theta_0^*}\frac{1}{n}\sum_{t=1}^{n}\log\frac{\sigma_t(\phi)}{\sigma_t\nu_t^{1/2}(\varphi)}=0,\quad a.s. \tag{3.9}
$$

We thus obtain (3.3) due to (3.8) and (3.9). Now the proof of (i) is completed.

(ii) The proof of (ii) is identical except that the *a.s.* convergence in (3.2) and (3.3) are replaced by L^p convergence. □

Lemma 2. *Under assumptions A1 - A3,* $E\log\big[g_{\vartheta_0}(\varepsilon_t)/(\nu_t^{-1/2}(\varphi)g_\vartheta(\nu_t^{-1/2}(\varphi)\varepsilon_t))\big]$ *attains a unique minimizer at* ψ_0*.*

Proof. By the definition of the Kullback-Leibler divergence, we have

$$
E\log\big\{ug_\vartheta(u\varepsilon_t)\big\}<E\log\big\{g_{\vartheta_0}(\varepsilon_t)\big\}
$$

for any $\vartheta(\vartheta\neq\vartheta_0)\in\tilde{\Theta}$ and $u(\neq 1)>0$. Thus,

$$
\begin{aligned}
&E\log\big[g_{\vartheta_0}(\varepsilon_t)/(\nu_t^{-1/2}(\varphi)g_\vartheta(\nu_t^{-1/2}(\varphi)\varepsilon_t))\big]\\
=\;&E\Big\{E\Big\{\log\big[g_{\vartheta_0}(\varepsilon_t)/(\nu_t^{-1/2}(\varphi)g_\vartheta(\nu_t^{-1/2}(\varphi)\varepsilon_t))\big]\big|\mathcal{F}_{t-1}\Big\}\Big\}\\
\le\;&0
\end{aligned}
$$

with the equality only if $\nu_t(\varphi)=1$ and $\vartheta=\vartheta_0$. By Lemma 7.2 of Francq and Zakoïan (2013), $\nu_t(\varphi)=1$ a.s. iff $\varphi=\varphi_0$ under assumptions A1 and A2. Combined with assumption A3, the conclusion follows. □

Lemma 3. *Suppose the conditions of Theorem 2 hold. Let* $\theta_1 = (\omega_1, \psi_1')'$ *and* $\theta_2 = (\omega_2, \psi_1')'$ *be two different points of* Θ *such that* $Elog[\alpha_{1+}(\varepsilon_t^+)^2 + \alpha_{1-}(\varepsilon_t^-)^2 + \beta_1] > 0$*. When* ω_1 *and* ω_2 *are sufficiently close, there exists no consistent test for* $H_0 : \theta_0 = \theta_1$ *against* $H_1 : \theta_0 = \theta_2$ *at the asymptotic level* $\alpha \in (0, 0.5)$*.*

Proof. We follow the lines of Francq and Zakoïan (2012). From Theorem 3.2.1 of Lehmann and Romano (2005), the most powerful test is Newman-Pearson test of rejection region $\mathcal{C} = \{\mathcal{S}_n > c_n\}$, where

$$\mathcal{S}_n = \sum_{t=1}^{n} \left[\log\left\{ \frac{1}{\sigma_t(\phi_2)} g_{\vartheta_1}\Big(\frac{X_t}{\sigma_t(\phi_2)}\Big) \right\} - \log\left\{ \frac{1}{\sigma_t(\phi_1)} g_{\vartheta_1}\Big(\frac{X_t}{\sigma_t(\phi_1)}\Big) \right\} \right]$$

and c_n is a positive constant corresponding to the α-quantile of the distribution of $\mathcal{S}_n$ under H_0. By recursion, we can obtain that $\sigma_t^2(\phi_1) \geq \omega_1 \prod_{i=1}^{t-1} \big[\alpha_{1+}(\varepsilon_i^+)^2 + \alpha_{1-}(\varepsilon_i^-)^2 + \beta_1\big]$. Noting that $\sigma_t^2(\phi_1) - \sigma_t^2(\phi_2) = \sum_{j=1}^{t} \beta_1^{j-1}(\omega_1 - \omega_2)$, under H_0, we have for any $k > 0$

$$E\Big|\frac{\sigma_t^2(\phi_1) - \sigma_t^2(\phi_2)}{\sigma_t^2(\phi_1)}\Big|^k \leq C\Big\{E\Big[\frac{\beta_1}{\alpha_{1+}(\varepsilon_i^+)^2 + \alpha_{1-}(\varepsilon_i^-)^2 + \beta_1}\Big]^k\Big\}^t = C\rho^t, \tag{3.10}$$

where $\rho = E\big\{\beta_1/\big[\alpha_{1+}(\varepsilon_i^+)^2 + \alpha_{1-}(\varepsilon_i^-)^2 + \beta_1\big]\big\}^k < 1$ due to assumption A2. By the proof of Lemma 7.3 of Francq and Zakoïan (2013), under H_0 there exists a sequence of stationary and ergodic process V_t such that

$$\frac{\sigma_t^2(\phi_1)}{\sigma_t^2(\phi_2)} \leq V_t \quad \text{and} \quad EV_t^k < +\infty \tag{3.11}$$

for any $k > 0$. Denote

$$\bar{\mathcal{S}}_n = \sum_{t=1}^{n} \left| \log\left\{ \frac{1}{\sigma_t(\phi_2)} g_{\vartheta_1}\Big(\frac{X_t}{\sigma_t(\phi_2)}\Big) \right\} - \log\left\{ \frac{1}{\sigma_t(\phi_1)} g_{\vartheta_1}\Big(\frac{X_t}{\sigma_t(\phi_1)}\Big) \right\} \right|.$$

By (3.10) and (3.11), under H_0, it follows that

$$\begin{aligned}
E\bar{\mathcal{S}}_n &= \sum_{t=1}^{n} E\Big|\log g_{\vartheta_1}\Big(\frac{X_t}{\sigma_t(\phi_2)}\Big) - \log g_{\vartheta_1}\Big(\frac{X_t}{\sigma_t(\phi_1)}\Big)\Big| + \frac{1}{2}\sum_{t=1}^{n} E\Big|\log \sigma_t^2(\phi_2) - \log \sigma_t^2(\phi_1)\Big| \\
&\leq C\sum_{t=1}^{n} E\Big|\Big(\frac{X_t}{\sigma_t(\phi_2)} + \frac{X_t}{\sigma_t(\phi_1)} + 1\Big)\varepsilon_t\Big|\Big|\frac{\sigma_t(\phi_1) - \sigma_t(\phi_2)}{\sigma_t(\phi_2)}\Big| \\
&\quad + C\sum_{t=1}^{n} E\Big[\frac{1}{\sigma_t^2(\phi_2)} + \frac{1}{\sigma_t^2(\phi_1)}\Big]\Big|\sigma_t^2(\phi_1) - \sigma_t^2(\phi_2)\Big| \\
&\leq C\sum_{t=1}^{n} E\Big[|\varepsilon_t| + \varepsilon_t^2(1 + V_t^{1/2}) + 1 + V_t\Big]\Big|\frac{\sigma_t^2(\phi_1) - \sigma_t^2(\phi_2)}{\sigma_t^2(\phi_1)}\Big| \\
&\leq C,
\end{aligned}$$

where C is independent of n, which implies

$$\mathcal{S}_n \to \mathcal{S}_0 = \sum_{t=1}^{\infty} \log \left\{ \sigma_t g_{\vartheta_1}\Big(\frac{X_t}{\sigma_t(\phi_2)}\Big) / \Big(\sigma_t(\phi_2) g_{\vartheta_1}(\varepsilon_t)\Big) \right\} \quad a.s. \text{ under } H_0.$$

Similarly, under H_1 we can show that

$$\mathcal{S}_n \to \mathcal{S}_1 = \sum_{t=1}^{\infty} \log \left\{ \sigma_t(\phi_1) g_{\vartheta_1}(\varepsilon_t) / \Big(\sigma_t g_{\vartheta_1}\Big(\frac{X_t}{\sigma_t(\phi_1)}\Big)\Big) \right\} \quad a.s..$$

Using (3.10), (3.11) and similar method in the proof of $E\bar{\mathcal{S}}_n \le C$ above, we can conclude that

$$|\mathcal{S}_0 - \mathcal{S}_1| \le |\omega_1 - \omega_2| \sum_{t=1}^{\infty} \rho^t H_t \tag{3.12}$$

where H_t is a stationary and ergodic process with finite expectation. Noting that the laws of $\mathcal{S}_0$ and $\mathcal{S}_1$ are continuous when $\omega_1 \neq \omega_2$, the power of the Neyman -Pearson test tends to

$$\lim_{n\to\infty} P_{H_1}(\mathcal{S}_n > c_n) = P(\mathcal{S}_1 > c)$$

where c is a constant such that $P(\mathcal{S}_0 > c) = \alpha$. For any $\varepsilon > 0$, we get

$$P(\mathcal{S}_1 > c) \le P(\mathcal{S}_0 + |\mathcal{S}_1 - \mathcal{S}_0| > c) \le P(\mathcal{S}_0 > c - \varepsilon) + P(|\mathcal{S}_1 - \mathcal{S}_0| > \varepsilon).$$

By continuity, $P(\mathcal{S}_0 > c - \varepsilon)$ is close to α when ε is close to zero. Furthermore, due to (3.12), $P(|\mathcal{S}_1 - \mathcal{S}_0| > \varepsilon)$ is close to zero provided $|\omega_1 - \omega_2|$ is small. Thus, $P(\mathcal{S}_1 > c) < 1$ when $|\omega_1 - \omega_2|$ is small. Thus the inconsistency of the Neyman-Pearson test and any other test is proved. □

Proof of Theorem 1.(i) By Lemma 7.1 of Francq and Zakoïan (2013), we have for any $\theta \notin \Theta_0$, $\sigma_t^2(\phi)/\sigma_t^2 \to \infty$ a.s., which implies that $Q_n(\theta)$ defined in Lemma 1 converges to ∞ a.s.. Thus,

$$\hat{\theta}_n = \arg\min_{\theta\in\Theta} l_n(\theta) = \arg\min_{\theta\in\Theta} Q_n(\theta) = \arg\min_{\theta\in\Theta_0} Q_n(\theta).$$

Combing the results of Lemma 1 and Lemma2 and the compactness of Θ, we complete the proof by standard arguments, see for example Theorem 4.1.1 of Amemiya (1985).

(ii) The proof is identical to that of (i), except that the a.s. convergence is replaced by the L^p convergence.

Proof of Theorem 2. If there exists a consistent estimator $\hat{\theta}_n$, then the test of critical region $\mathcal{C} = \{\|\hat{\theta}_n - \theta_1\| > \|\hat{\theta}_n - \theta_2\|\}$ would have null asymptotic errors of the first and second kind, which contradicts Lemma 3.

Acknowledgments

Wang's research was supported by National Natural Science Foundation of China (No. 11101448), the Program for New Century Excellent Talents in University, the program for Young Talents of Beijing, the program for national statistics science research plan (No. 2013LY015), the project 211 of the Central University of Finance and Economics and the CUFE Young Scholar Innovation Fund.

References

[1] Amemiya, T.(1985). *Advanced Econometrics*. Harvard University Press.

[2] Beran, R. (1977). Minimum Hellinger distance estimates for parametric models. *Annals of Statistics* 5, 445-463.

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

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

[5] Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* 31, 307-327.

[6] Bollerslev, T. (1987). A conditionally heteroskedastic time series model for speculative prices and rates of return. *The Review of Economics and Statistics* 69, 542-547.

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

[8] Francq, C. and Zakoïan, J.M. (2004). Maximum likelihood estimation of pure GARCH and ARMA-GARCH processes. *Bernoulli* 10, 605C637.

[9] Francq, C. and Zakoïan, J.M. (2012). Strict stationarity testing and estimation of explosive and stationary generalized autoregressive conditional heteroscedasticity models. *Econometrica* 80, 821-861.

[10] Francq, C. and Zakoïan, J.M. (2013). Inference in non stationary asymmetric GARCH models. *Annals of Statistics*, to appear.

[11] Glosten, L.R., Jaganathan, R. and D. Runkle (1993). On the relation between the expected values and the volatility of the nominal excess return on stocks. *Journal of Finance* 48, 1779C1801.

[12] Ha, J. and Lee, T.-W. (2011). NM-QMLE for ARMA-GARCH models with non-Gaussian innovations. *Statistics and Probability Letters* 81, 694-703.

[13] Haas, M., Mittnik, S. and Paolella, M. S. (2004). Mixed normal conditional heteroskedasticity. *The Journal of Financial Economics* 2, 211-250.

[14] Hall, P. and Yao, Q. (2003). Inference in ARCH and GARCH models with heavy-tailed errors. *Econometrica* 71, 285-317.

[15] Hwang, S. Y., Baek, J.S., Park, J.A. and Choi, M.S. (2010). Explosive volatilities for threshold-GARCH processes generated by asymmetric innovations. *Statistics and probability Letters* 80, 26-33.

[16] Hwang, S. Y. and Basawa, I. V. (2004). Stationarity and moment structure for Box-Cox transformed threshold GARCH(1,1) processes. *Statistics and Probability Letters* 68, 209-220.

[17] Hwang, S. Y. and Kim, T. Y. (2004). Power transformation and threshold modeling for ARCH innovations with applications to tests for ARCH structure. *Stochastic Processes and Their Applications* 110, 295-314.

[18] Jensen, S. T. and Rahbek, A. (2004 a). Asymptotic normality of the QMLE estimator of ARCH in the nonstatinary case. *Econometrica* 72, 641-646.

[19] Jensen, S. T. and Rahbek, A. (2004 b). Asymptotic inference for nonstationary GARCH. *Econometric Theory* 20, 1203-1226.

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

[21] Lee, T.-W. and Lee, S. (2009). Normal mixture quasi-maximum likelihood estimator for GARCH models. *Scandinavian Journal of Statistics* 36,157-170.

[22] Li, C.W. and Li, W.K. (1996). On a double-threshold autoregressive heteroscedastic time series model. *Journal of Applied Econometrics* 11, 253-274.

[23] Linton, O., Pan, J.Z. and Wang, H. (2010). Estimation for a non-stationary semi-strong GARCH(1,1) model with heavy-tailed errors. *Econometric Theory* 26, 1-28.

[24] Loretan, M. and Phillips, P. C. B. (1994). Testing the covariance stationarity of heavy-tailed time series. *Journal of Empirical Finance* 1, 211-248.

[25] 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.

[26] Mikosch, T. and Stărică, C. (2000). Limit theory for the sample autocorrelations and extremes of a GARCH(1, 1) process. *The Annals of Statistics* 28, 1427-1451.

[27] Pan, J. Z., Wang, H. and Tong, H. (2008). Estimation and tests for power-transformed and threshold GARCH models. *Journal of Econometrics* 142, 352-378.

[28] Stărică, C., Herzel, S. and Nord, T. (2005). Why does the GARCH(1,1) model fail to provide sensible longer-horizon volatility forecasts? mimeo, Chalmers University of Technology.

[29] Vlaar, P. J. G. and Palm, F. C. (1993). The message in weekly exchange rates in the European monetary system: mean reversion, conditional heteroscedasaticity, and jumps. *Journal of Business & Economic Statistics* 11, 351-360.

[30] White, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrica* 20, 1-25.

[31] Zhang, Z., Li, W. K. and Yuen, K. C. (2006). On a mixture GARCH time-series model. *Journal of Time Series Analysis* 27, 577-597.