A new Pearson-type QMLE for conditionally heteroskedastic models

Zhu, Ke and Li, Wai Keung

Institute of Applied Mathematics, Chinese Academy of Sciences, Department of Statistics and Actuarial Science, University of Hong Kong

18 December 2013

A new Pearson-type QMLE for conditionally heteroskedastic models

BY KE ZHU

Institute of Applied Mathematics, Chinese Academy of Sciences, Haidian District, Zhongguancun, Beijing, China

kzhu@amss.ac.cn

AND WAI KEUNG LI

Department of Statistics and Actuarial Science, University of Hong Kong, Pokfulam Road, Kowloon, Hong Kong

hrntlwk@hku.hk

ABSTRACT

This paper proposes a novel Pearson-type quasi maximum likelihood estimator (QMLE) of GARCH(p, q) models. Unlike the existing Gaussian QMLE, Laplacian QMLE, generalized non-Gaussian QMLE, or LAD estimator, our Pearsonian QMLE (PQMLE) captures not the heavy-tailed but also the skewed innovations. Under the stationarity and weak moment conditions, the strong consistency and asymptotical normality of the PQMLE are obtained. With no further efforts, the PQMLE can apply to other conditionally heteroskedastic models. A simulation study is carried out to assess the performance of the PQMLE. Two applications to eight major stock indexes and four exchange rates further highlight the importance of our new method. To our best knowledge, the heavy-tailed and skewed innovations are observed together in practice, and the PQMLE now gives us a systematical way to capture this co-existing feature.

Some key words: Asymmetric innovation; Conditionally heteroskedastic model; Exchange rates; GARCH model; Leptokurtic innovation; Non-Gaussian QMLE; Pearson's Type IV distribution; Pearsonian QMLE; Stock indexes.

1. INTRODUCTION

After the seminar work of Engle (1982) and Bollerslev (1986), numerous volatility models have been widely used to capture the feature of conditional heteroscedasticity in economic and financial data sets; see, e.g., Bollerslev, Chou and Kroner (1992), Bera and Higgins (1993) and Francq and Zakoïan (2010). Among them, the most influential model in empirical study is the GARCH(p, q) model given by

$$y_t = \sigma_t \varepsilon_t, \tag{1}$$

$$\sigma_t^2 = \omega + \sum_{i=1}^{p} \alpha_i y_{t-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2, \tag{2}$$

where $\omega > 0$, $\alpha_i \geq 0 (i = 1, \cdots, p)$, $\beta_j \geq 0 (j = 1, \cdots, q)$, and ε_t is a sequence of i.i.d. random variables. Traditional inference for the GARCH model is based on the Gaussian quasi maximum likelihood estimator (GQMLE), which is proposed by assuming that ε_t follows a standard normal distribution. Berkes, Horváth and Kokoszka (2003) showed that when ε_t has a finite fourth moment with $E\varepsilon_t^2 = 1$ (the identification condition), the GQMLE is consistent and asymptotical normal. However, the GQMLE can not capture the heavy-tailedness and skewness of ε_t, which are two well-observed features of GARCH models in application; see, e.g., Engle and González-Rivera (1991), Christoffersen, Heston and Jacobs (2006), and Grigoletto and Lisi (2009). Motivated by this, the MLE, based on a user-chosen heavy-tailed or skewed likelihood function, so far has been largely considered. For instance, ε_t can be the Student's t distribution in Bollerslev (1987), the gamma distribution in Engle and González-Rivera (1991), the generalized error distribution in Nelson (1991), the skewed t distribution in Hansen (1994), the stable distribution in Liu and Brorsen (1995), the noncentral t distribution in Harvey and Siddique (1999), the Pearsons Type IV distribution in Premaratne and Bera (2001), the Gram-Charlier distribution in Jondeau and Rockinger (2001), the mixture normal distribution in Bai, Russell and Tiao (2003) and many others. However, the true distribution of ε_t is prior unknown in practice, and as shown in Newey and Steigerwald (1997), the MLE may lead to inconsistent estimates of models (1)-(2) if the distribution of ε_t is misspecified.

In order to obtain a consistent estimator without knowing the true distribution of ε_t, people prefer to use the non-Gaussian QMLE (NGQMLE), which has efficiency advantage over GQMLE when ε_t is heavy-tailed. Generally, there have two ways to get a consistent NGQMLE. First, one can assume a different identification condition rather than $E\varepsilon_t^2 = 1$. For instance, Peng and Yao (2003) proposed a least absolute deviation estimator (LADE) under the identification condition that median$(\varepsilon_t^2) = 1$, and the consistency and asymptotic normality of LADE was proved in Chen and Zhu (2013) under only a finite fractional moment of ε_t. By assuming that ε_t follows a standard laplace distribution, Berkes and Horváth (2004) considered the Laplacian QMLE (LQMLE) under the identification condition that $E|\varepsilon_t| = 1$, and they showed that the LQMLE is consistent and asymptotical normal when ε_t has a finite second moment. Secondly, one can remain the identification condition $E\varepsilon_t^2 = 1$ for NGQMLE but re-parameterize models (1)-(2). This method has been used for the semi-parametric estimator in Drost and Klaassen (1997), the rank-based estimator in Andrews (2012), and the generalized NGQMLE (GNGQMLE) in Fan, Li and Xiu (2013). By introducing a scale adjustment parameter, the GNGQMLE is consistent and asymptotical normal when ε_t has a finite second moment, while the semi-parametric and rank-based estimators can only estimate the heteroscedastic parameters α_i and β_j under the same re-parameterized GARCH(p, q) model. Morevoer, it is worth noting that when ε_t has an infinite fourth moment, all of LADE, LQMLE and GNGQMLE achieve root-n convergency, while the GQMLE suffers a slower convergence rate as shown in Hall and Yao (2003).

In this paper, we propose a Pearsonian QMLE (PQMLE) of models (1)-(2) by assuming that ε_t follows a Pearson's Type IV distribution. Like the LADE and LQMLE, the PQMLE requires a specified identification condition rather than $E\varepsilon_t^2 = 1$. Under the stationarity and a finite fractional moment of ε_t, the strong consistency and asymptotic normality of the PQMLE are obtained. Therefore, all of aforementioned non-Gaussian distributions used in the MLE method is applicable for the PQMLE. Furthermore, we show that the PQMLE can easily apply to other conditionally heteroskedastic models. A simulation study is carried out to assess the performance of the PQMLE, and two applications to eight major stock indexes and four exchange rates further highlight the importance of our new method. Compare to the existing NGQMLEs, the PQMLE captures not the heavy-tailed but also the skewed innovations. To our best knowledge, the heavy-tailed and skewed innovations are observed together in practice, but none of QMLE has been focus on this co-existing feature in the literature. The PQMLE method, who can capture a very large range of the asymmetry and leptokurtosis of ε_t, now gives us a systematical way to achieve this goal.

This paper is organized as follows. Section 2 proposes our PQMLE and studies its asymptotic property. The simulation results are reported in Section 3. An application is given in Section 4. Concluding remarks are offered in Section 5. The proofs are provided in the Appendix. Throughout the paper, A' is the transpose of matrix A, $|A| = (tr(A'A))^{1/2}$ is the Euclidean norm of a matrix A, $\|A\|_s = (E|A|^s)^{1/s}$ is the L^s-norm ($s \geq 1$) of a random matrix, $O(1)$ denotes a bounded generic constant, "$\to_d$" denotes convergence in distribution, and "$\to_p$" denotes convergence in probability.

2. The PQMLE and Asymptotic Theory

2·1. *Some basic assumptions*

Let $\theta = (\omega, \alpha_1, \cdots, \alpha_p, \beta_1, \cdots, \beta_q)'$ be the unknown parameters of models (1)-(2) and its true value be θ_0. Denote the parameter space by Θ, where $\Theta \in \mathcal{R}_0^{1+p+q}$ is compact and $\mathcal{R}_0 = [0, \infty)$. Then, we need the following assumptions:

Assumption 1. y_t is strictly stationary.

Assumption 2. For each $\theta \in \Theta$, $\alpha(z)$ and $\beta(z)$ have no common root, $\alpha(1) \neq 0, \alpha_p + \beta_q \neq 0$ and $\sum_{j=1}^q \beta_j < 1$, where $\alpha(z) = \sum_{i=1}^p \alpha_i z^i$ and $\beta(z) = 1 - \sum_{j=1}^q \beta_j z^j$.

Assumption 3. (i) ε_t^2 is a nondegenerate random variable; (ii) $\lim_{s\to 0} s^{-\mu} P(\varepsilon_t^2 \leq s) = 0$ for some $\mu > 0$; (iii) $E|\varepsilon_t|^{2\kappa} < \infty$ for some $\kappa > 0$.

Assumption 1 is a basic set-up for models (1)-(2), and its necessary and sufficient conditions are given in Bougerol and Picard (1992). Assumption 2 and Assumption 3(i) are the identifiability conditions for models (1)-(2) as shown in Berkes, Horváth and Kokoszka (2003). Assumptions 3(ii)-(iii) from Berkes and Horváth (2004) are technical conditions for proving our asymptotic theory. Note that only a finite fractional moment of ε_t is required in this case, and so it applies to very heavy-tailed innovations.

2·2. *The Pearson's Type IV distribution*

We briefly review the Pearson's Type IV distribution in Nagahara (1999) and Heinrich (2004). The Pearson's Type IV (PIV) distribution, as one of the asymmetric and leptokurtic distributions, has the following pdf:

$$f(x; \lambda, a, \nu, m) = K \left[1 + \left(\frac{x-\lambda}{a}\right)^2\right]^{-m} \exp\left[-\nu \tan^{-1}\left(\frac{x-\lambda}{a}\right)\right], \tag{3}$$

where $x \in \mathcal{R}$ and (λ, a, ν, m) are real parameters with $m \geq 1/2$ and $a > 0$. Here, K is the normalizing constant given by

$$K = \frac{2^{2m-2}\,|\Gamma(m + i\nu/2)|^2}{a\pi\Gamma(2m-1)},$$

where $i = \sqrt{-1}$ is the imaginary number and $\Gamma(\cdot)$ is the complex Gamma function. In (3), λ and a are the location and the scale parameters, respectively; the parameter ν is related to the asymmetry of the distribution, and a positive (or negative) ν stands for a negatively (or positively) skewed distribution; the parameter m captures the leptokurtosis of the distribution, and a smaller value of m represents a heavier tail of the distribution. To further illustrate this, Figure 1 plots four different densities $f(x; 0, 1, \nu, m)$. From Figure 1, we know that PIV(λ, a, ν, m) distribution with a small (or large) m can have a heavier (or lighter) tail than N(0,1) distribution. Also, it is

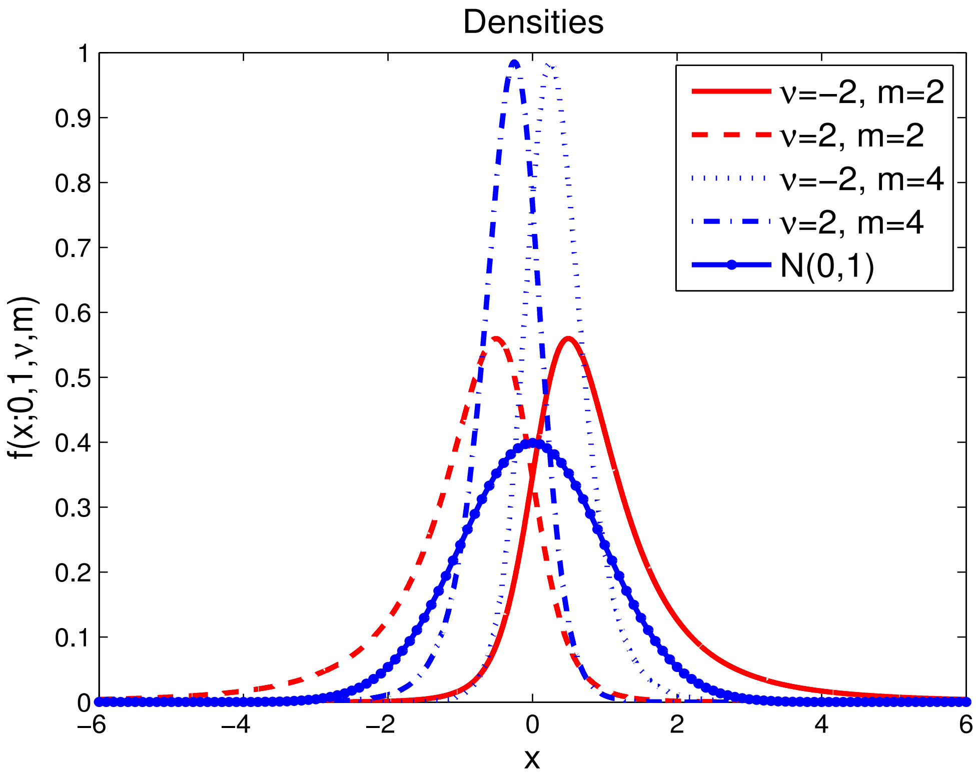

Fig. 1. The plot of four different densities $f(x;0,1,\nu,m)$ for the Pearson's Type IV distribution (the solid star line is the density of N(0,1) distribution).

worth mentioning that if $\varepsilon_t \sim \text{PIV}(\lambda, a, \nu, m)$, its j-th moment exists only when $j < r+1$ for $r = 2(m-1)$. That is, ε_t has a finite second moment when $m > 3/2$, and it has a finite fourth moment when $m > 5/2$. Particularly, the skewness and kurtosis of ε_t are as follows:

$$\text{skew}(\varepsilon_t) = \frac{-4\nu}{r-2}\sqrt{\frac{r-1}{r^2+\nu^2}} \quad \text{for } m > 2,$$
$$\text{kurt}(\varepsilon_t) = \frac{3(r-1)\left[(r+6)(r^2+\nu^2) - 8r^2\right]}{(r-2)(r-3)(r^2+\nu^2))} \quad \text{for } m > 5/2.$$

Figure 2 gives a 3-dimensional (3-D) plot of the skewness and kurtosis of ε_t. From this figure, we find that when $|\nu|$ (or m) increases, the absolute value of the skewness increase (or decrease) for fixed m (or ν); and the same conclusion holds for the kurtosis. Hence, we know that the PIV distribution can capture a very large range of the asymmetry and leptokurtosis of the innovation. For more discussions on PIV distributions, we refer to Bauwens and Laurent (2005), Yan (2005), and Grigoletto and Lisi (2009).

Next, we are interest of the case when ε_t in models (1)-(2) follows the PIV distribution. Figures 2-3 plot one realization from the following GARCH(1,1) model:

$$y_t = \varepsilon_t \sigma_t \quad \text{and} \quad \sigma_t^2 = 0.01 + 0.01 y_{t-1}^2 + 0.9\sigma_{t-1}^2, \tag{4}$$

where $\varepsilon_t \sim \text{PIV}(0, 1, \nu, m)$ with $(\nu, m) = (\pm 2, 2)$, $(0, 2)$, $(\pm 2, 4)$, and $(0, 4)$. From Figures 2-3, we find that no matter how heavy ε_t is, y_t has a higher probability to be positive (or negative)

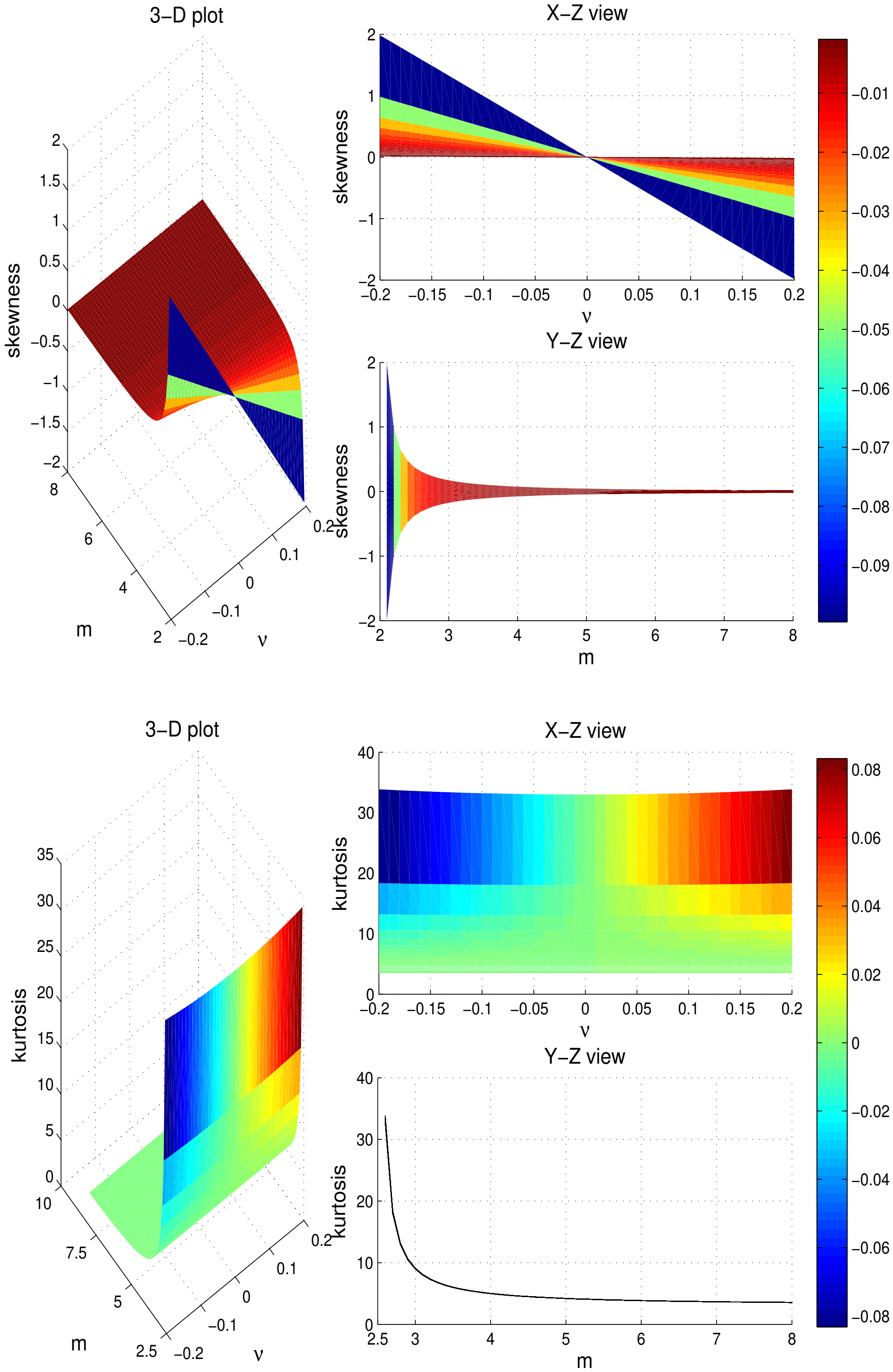

Fig. 2. (top panel) the 3-D plot of the skewness of ε_t, where $\varepsilon_t \sim \mathrm{PIV}(0,1,\nu,m)$ with $\nu \in [-0.2, 0.2]$ and $m \in (2, 8)$; (bottom panel) the 3-D plot of the kurtosis of ε_t, where $\varepsilon_t \sim \mathrm{PIV}(0,1,\nu,m)$ with $\nu \in [-0.2, 0.2]$ and $m \in (5/2, 8)$.

when $\nu < 0$(or > 0), and this asymmetric phenomena disappears when $\nu = 0$. Moreover, when m becomes smaller, the absolute value of y_t tends to be more larger, especially for its extreme values. All of these findings indicate that the GARCH model with the PIV$(0, 1, \nu, m)$ innovations can capture a very large range of the asymmetry and leptokurtosis of the data set.

2·3. *The PQMLE*

Given the observations $\{y_n, \cdots, y_1\}$ and the initial values $Y_0 =: \{y_i; i \leq 0\}$, we first rewrite the parametric models (1)-(2) as

$$\varepsilon_t(\theta) = y_t/\sqrt{h_t(\theta)} \text{ and}$$
$$h_t(\theta) = c_0(\theta) + \sum_{i=1}^{\infty} c_i(\theta) y_{t-i}^2,$$

where all expressions for $c_i(\theta)(i \geq 0)$ are given in Berkes and Horváth (2004, pages 635-636). Clearly, $\varepsilon_t(\theta_0) = \varepsilon_t$ and $h_t(\theta_0) = \sigma_t^2$. In practice, since the values of Y_0 are unobservable, we can replace them by zeros, and then use $\tilde{h}_t(\theta)$ instead of $h_t(\theta)$ to calculate our estimator, where

$$\tilde{h}_t(\theta) = c_0(\theta) + \sum_{i=1}^{t-1} c_i(\theta) y_{t-i}^2 \text{ for } t = 2, \cdots, n, \tag{5}$$

and $\tilde{h}_1(\theta) = c_0(\theta)$. For given (ν, m), when ε_t follows the PIV$(0, 1, \nu, m)$ distribution, the log-likelihood function (ignoring some constants) can be written as

$$\tilde{L}_n(\theta) = -\sum_{t=1}^{n} \left\{ \log \sqrt{\tilde{h}_t(\theta)} + m \log \left[1 + \frac{y_t^2}{\tilde{h}_t(\theta)} \right] + \nu \tan^{-1} \left(\frac{y_t}{\sqrt{\tilde{h}_t(\theta)}} \right) \right\}, \tag{6}$$

where $m \geq 1/2$. We look for the maximizer of $\tilde{L}_n(\theta)$ on Θ, that is,

$$\tilde{\theta}_n = \arg\max_{\theta \in \Theta} \tilde{L}_n(\theta). \tag{7}$$

Because we do not assume that ε_t follows the PIV$(0, 1, \nu, m)$ distribution, $\tilde{\theta}_n$ is called the Pearsonian quasi-maximum likelihood estimator (PQMLE) of θ_0. Note that equation (6) depends on the distribution parameters (ν, m), and so we should specify them before the calculation of $\tilde{L}_n(\theta)$. Particularly, when $\nu = 0$, the log-likelihood function $\tilde{L}_n(\theta)$ is symmetric. The detailed procedure to select (ν, m) is discussed in Remark 3.

Next, let $\bar{f}(x) = f(x; 0, 1, \nu, m)/K$, $g(y, s) = \log \left[s\bar{f}(ys) \right]$ and $w(s) := E\left[g(\varepsilon_t, s) \right]$, where $y \in \mathcal{R}$ and $s > 0$. Then, it is straightforward to see that

$$\tilde{L}_n(\theta) = \sum_{t=1}^{n} g\left(y_t, 1/\sqrt{\tilde{h}_t(\theta)} \right).$$

In order to derive the asymptotic property of $\tilde{\theta}_n$, we need two more assumptions below:

Assumption 4. The innovation ε_t satisfies that

$$E\left[\frac{2m\varepsilon_t^2 + \nu\varepsilon_t}{1 + \varepsilon_t^2} \right] = 1.$$

Assumption 5. $w(s)$ has a unique maximum point at $s = 1$.

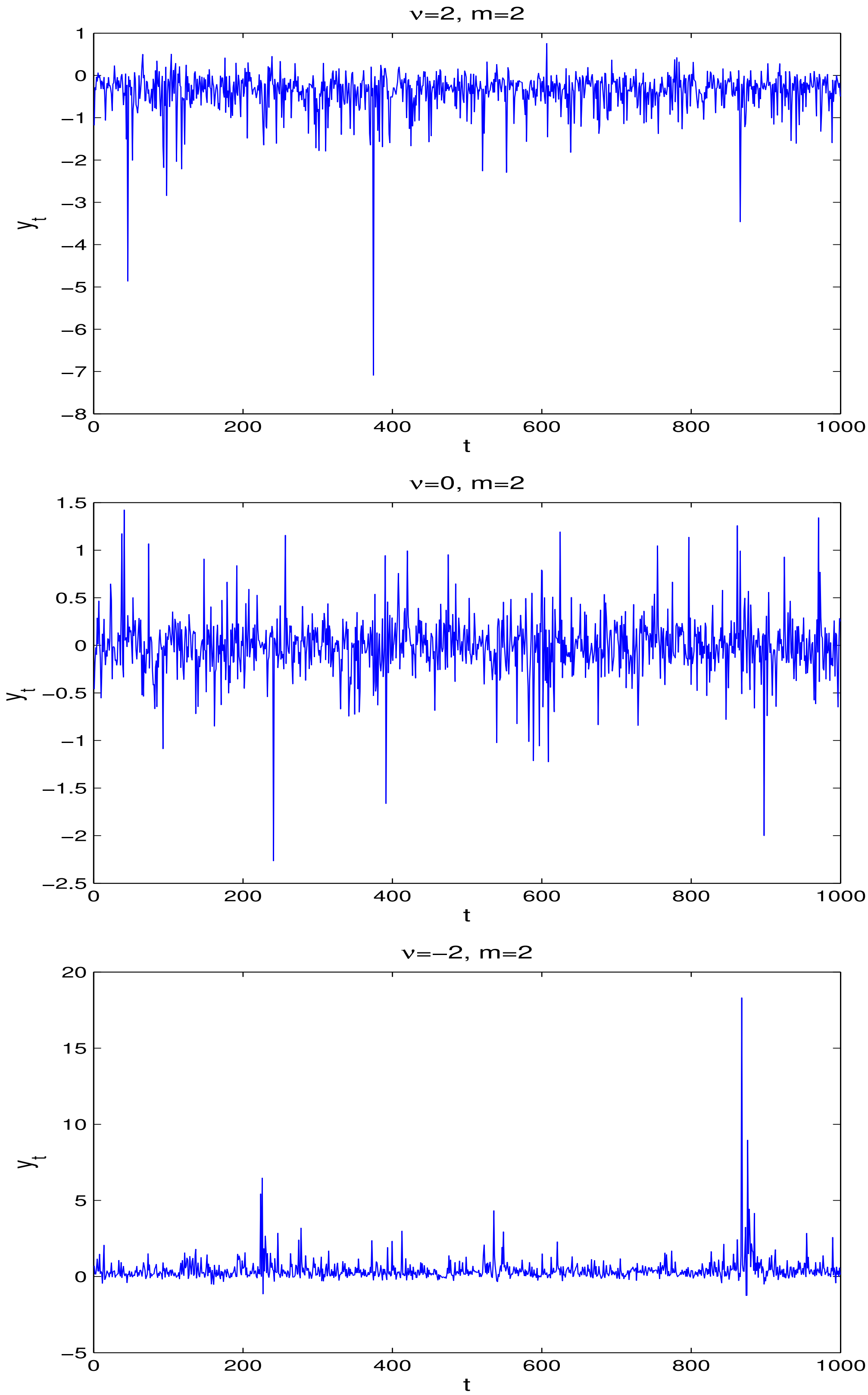

Fig. 3. One realization $\{y_t\}_{t=1}^{1000}$ from model (4), when $\varepsilon_t \sim \mathrm{PIV}(0, 1, \nu, m)$.

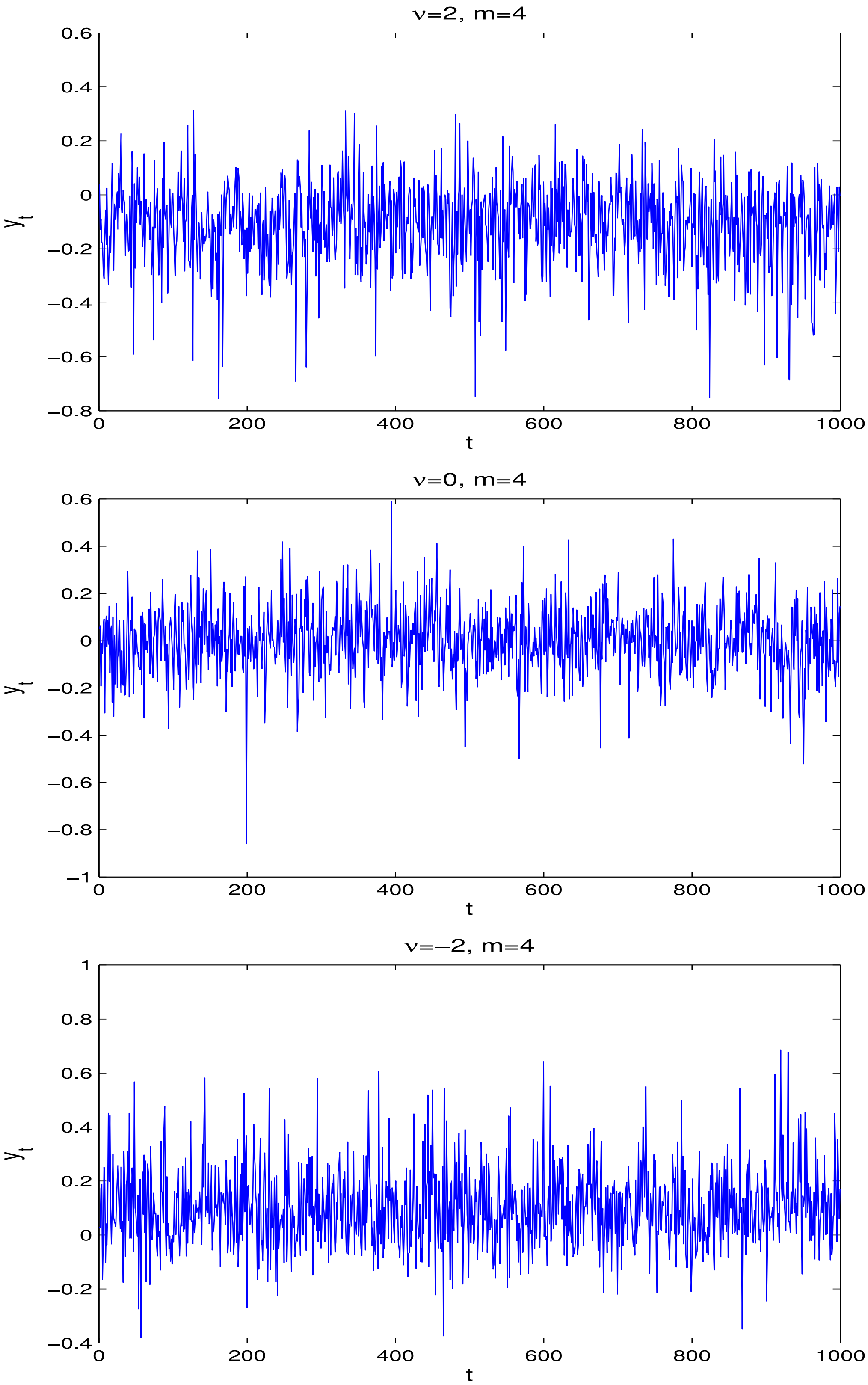

Fig. 4. One realization $\{y_t\}_{t=1}^{1000}$ from model (4), when $\varepsilon_t \sim \text{PIV}(0, 1, \nu, m)$.

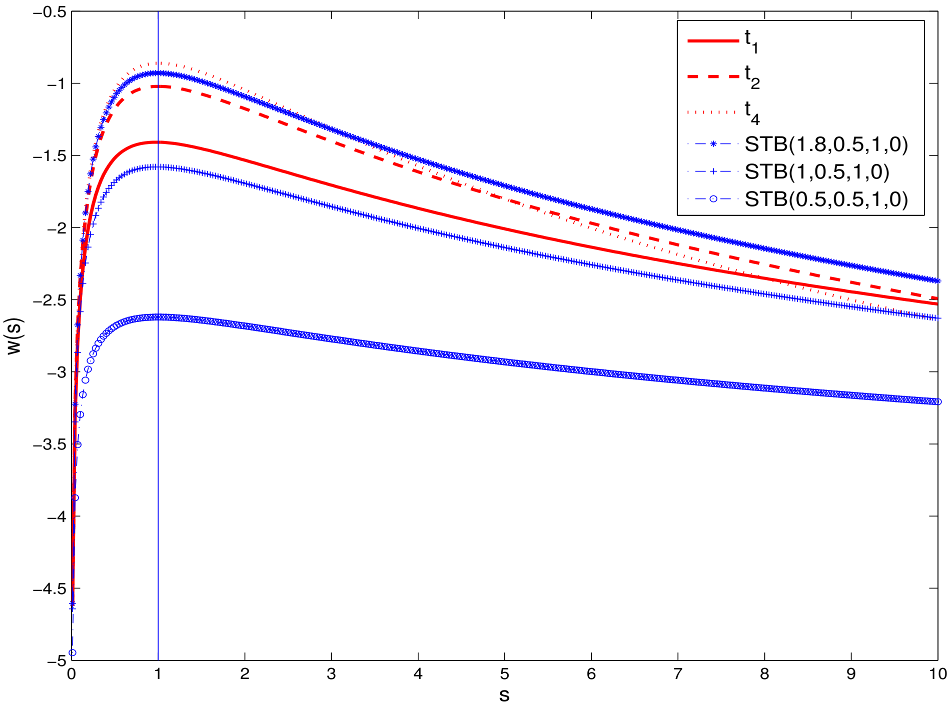

Fig. 5. The plot of $w(s)$ for Student's t and stable (STB) distributions.

Assumption 4 is the identification condition for the PQMLE. Unlike the GQMLE, the condition $E\varepsilon_t^2 = 1$ is not needed, and the conditional variance of y_t in this case is $(E\varepsilon_t^2)E\sigma_t^2$, provided that $E\varepsilon_t^2 < \infty$. Assumption 5 is a technical condition for proving the strong consistency of the PQMLE. After some simple algebra, we can show that a sufficient condition for Assumption 5 is that (i) $w(s)$ is concave on $\{s : s > 0\}$; and (ii) $E\left[\nu\varepsilon_t/(1+\varepsilon_t^2)\right] \le 0$. Figure 5 plots the function $w(s)$ for Student's t_i $(i = 1, 2, 4)$ distributions and stable (STB) distributions such that Assumption 4 holds, where (ν, m) are set to be $(-1, 1)$ for t_1, $(-1.16, 1.16)$ for t_2, $(-1.3, 1.3)$ for t_4, $(1.11, 1.11)$ for STB$(1.8, 0.5, 1, 0)$, $(0.97, 0.97)$ for STB$(1, 0.5, 1, 0)$, and $(0.76, 0.76)$ for STB$(0.5, 0.5, 1, 0)$. Clearly, $w(s)$ in Figure 5 is concave with a unique maximum point at $s = 1$ for all six distributions. Here, the STB$(\breve{\alpha}, \breve{\beta}, c, \mu)$ distribution has the following characteristic function:

$$\psi(t; \breve{\alpha}, \breve{\beta}, c, \mu) = \exp\left[it\mu - |ct|^{\breve{\alpha}}(1 - i\breve{\beta}\mathrm{sgn}(t)\Phi)\right],$$

where $\breve{\alpha} \in (0, 2]$, $\breve{\beta} \in [-1, 1]$, $c \in (0, \infty)$, $\mu \in \mathcal{R}$, and

$$\Phi = \begin{cases} \tan(\pi\breve{\alpha}/2) & \text{if } \breve{\alpha} \neq 1, \\ -(2/\pi)\log|t| & \text{if } \breve{\alpha} = 1. \end{cases}$$

Denote the first and second derivatives of $g(y, s)$ with respective to s by $g_1(y, s)$ and $g_2(y, s)$, respectively. We now are ready to give our main results:

THEOREM 1. *Suppose that Assumptions* 1-5 *hold. Then, as* $n \to \infty$, *(i)* $\tilde{\theta}_n \to \theta_0$ *almost surely (a.s.); and (ii)* $\sqrt{n}\left(\tilde{\theta}_n - \theta_0\right) \to_d N(0, 4\tau^2 A^{-1})$, *where*

$$\tau^2 = \frac{Eg_1^2(\varepsilon_t, 1)}{[Eg_2(\varepsilon_t, 1)]^2} \text{ and } A = E\left[\frac{1}{h_t^2(\theta_0)} \frac{\partial h_t(\theta_0)}{\partial \theta} \frac{\partial h_t(\theta_0)}{\partial \theta'}\right].$$

Remark 1. The PQMLE only needs a finite fractional moment of ε_t for its asymptotic normality, which is weaker than the moment condition $E\varepsilon_t^4 < \infty$ for the GQMLE in Berkes, Horváth, and Kokoszka (2003) and Francq and Zakoïan (2004), or the moment condition $E\varepsilon_t^2 < \infty$ for the LQMLE in Berkes and Horváth (2004) and the GNGQMLE in Fan, Li, and Xiu (2013). Note that as shown in Chen and Zhu (2013), the LADE in Peng and Yao (2003) also only needs a finite fractional moment of ε_t for its asymptotic normality.

Remark 2. The identification condition for the PQMLE in Assumption 4 is different from the identification condition $E\varepsilon_t^2 = 1$ for the GQMLE and the GNGQMLE, the identification condition $E|\varepsilon_t| = 1$ for the LQMLE, or the identification condition median$(\varepsilon_t^2) = 1$ for the LADE. Thus, it is not straightforward to compare the efficiency of the PQMLE with that of other estimators in formal, and the simulation comparison in Section 3 is necessary.

Remark 3. In order to calculate the PQMLE, we need to first select the parameters ν and m. This can be simply done by using the maximum likelihood estimation method; see Premaratne and Bera (2001), Verhoeven and McAleer (2004), and Bhattacharyya, Mirsa, and Kodase (2009). Assume that $\varepsilon_t \sim$ PIV$(0, 1, \nu, m)$. Then, we can estimate (ν, m, θ) jointly by maximizing the full log-likelihood function LLF$_P(\nu, m, \theta)$, where

$$\text{LLF}_P(\nu, m, \theta) = \tilde{L}_n(\theta) + n \log K. \tag{8}$$

Now, we can choose (ν, m) to be the corresponding estimators from this MLE method. Although the parameters ν and m selected by the MLE method may not be optimal, the practical usefulness of this method has been illustrated by the empirical examples in Section 4.

Remark 4. Note that the value of (ν, m) can be any one in $(-\infty, \infty) \times [1/2, \infty]$, and a different value of (ν, m) will intricate a different stationarity region of y_t. To see it, Figure 6 plots the strict stationarity region of the GARCH(1,1) model: $y_t = \varepsilon_t \sigma_t$ and $\sigma_t^2 = \omega + \alpha y_{t-1}^2 + \beta \sigma_{t-1}^2$, where $\varepsilon_t \sim$ PIV$(0, 1, \nu, m)$. As a comparison, the region for $Ey_t^2 < \infty$ is also plotted in Figure 6. From this figure, we find that the parameter region for strict stationarity is much larger than that for $Ey_t^2 < \infty$. Moreover, a smaller value of ν or a larger value of m will give a larger strict stationarity region. Particularly, except that $\varepsilon_t \sim$ PIV$(0, 1, 2, 2)$, each strict stationarity region in Figure 6 is much larger than that in Nelson (1990) when $\varepsilon_t \sim N(0, 1)$ or that in Zhu and Ling (2011) when $\varepsilon_t \sim$ Laplace$(0, 1)$. Therefore, our PQMLE can have a much larger admissible parameter region than the GQMLE, the GNGQMLE or the LQMLE.

2·4. *Extension to conditionally heteroskedastic models*

In this subsection, we study the PMLE for the following conditionally heteroskedastic models:

$$y_t = \sigma_t \varepsilon_t \text{ and } \sigma_t = \sigma(y_{t-1}, y_{t-2}, \cdots; \theta_0), \tag{9}$$

where ε_t being independent of $\{y_j; j < t\}$ is a sequence of i.i.d. random variables, the parameter space $\Theta \subset \mathcal{R}^l$ is compact, the true value θ_0 is an interior point in Θ, and $\sigma : \mathcal{R}^\infty \times \Theta \to (0, \infty)$. Many existing models, such as GARCH model in (1)-(2), asymmetric power GARCH model in

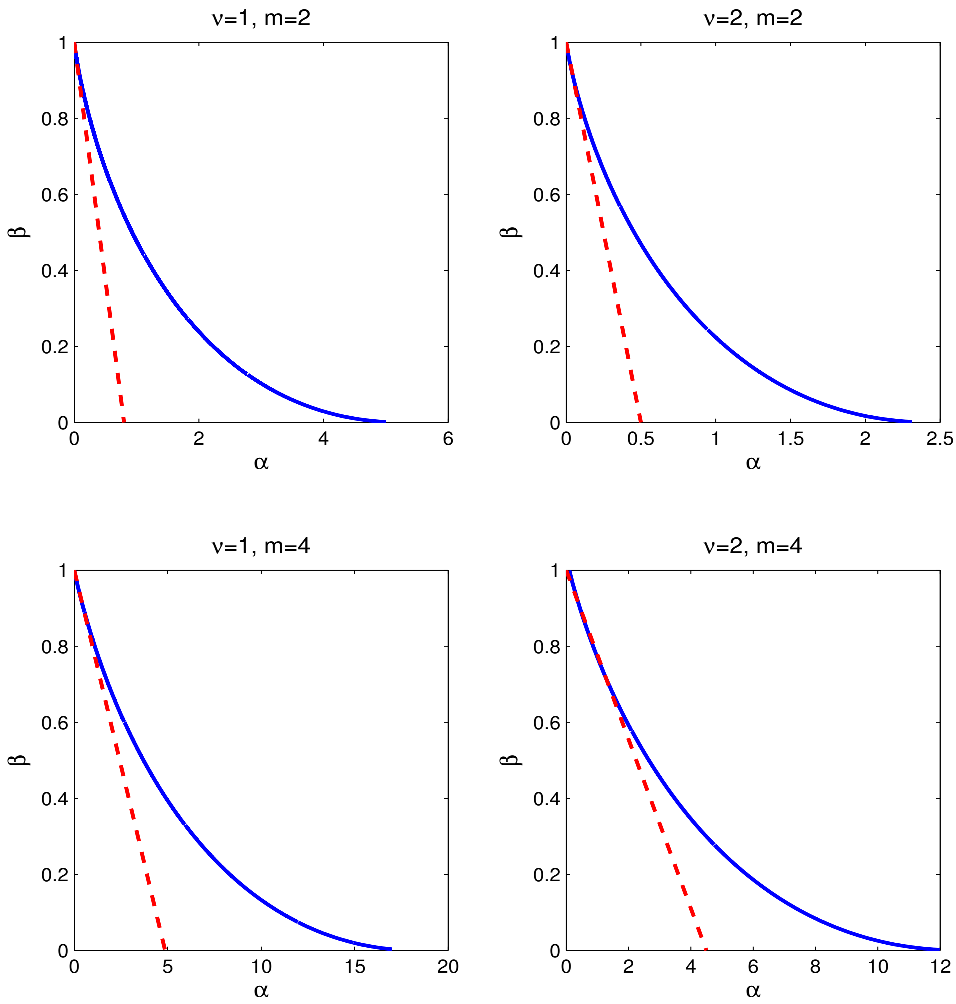

Fig. 6. The regions bounded by the solid and dashed curves are for the strict stationarity (i.e., $E[log(\alpha\varepsilon_t^2 + \beta)] < 0$) and for $Ey_t^2 < \infty$ (i.e., $E\varepsilon_t^2\alpha + \beta < 1$), respectively, where $E\varepsilon_t^2 = (r^2 + \nu^2)/(r^2(r-1))$ with $r = 2(m-1)$.

Ding, Granger, and Engle (1993) and asymmetric log-GARCH model in Geweke (1986), are embedded into model (9); see e.g., Bollerslev, Chou, and Kroner (1992) and Francq and Zakoïan (2010) for more discussions in this context.

As (5), let $h_t(\theta) = [\sigma(y_{t-1}, y_{t-2}, \cdots; \theta)]^2$ and define $\tilde{h}_t(\theta)$ in the same as $h_t(\theta)$ by replacing Y_0 by zeros. Then, based on $\{\tilde{h}_t(\theta)\}$, we can define the PMLE for model (9) as in (7). To derive the asymptotic property of the PMLE, three more technical assumptions are needed.

Assumption 6. (i) $h_t(\theta) \geq \underline{w}$ (a.s.) for some $\underline{w} > 0$ and all $\theta \in \Theta$. Moreover, $h_t(\theta) = h_t(\theta_0)$ (a.s.) if and only if $\theta = \theta_0$; (ii) if $x'(\partial h_t(\theta)/\partial\theta_i)_{i=1\cdots l} = 0$ (a.s.) for any $x \in \mathcal{R}^l$, then $x = 0$.

Assumption 7.

$$(i)\ E\left[\sup_{\theta\in\Theta}\left\|\frac{1}{h_t(\theta)}\frac{\partial h_t(\theta)}{\partial\theta}\right\|\right]^2<\infty;\ (ii)\ E\left[\sup_{\theta\in\Theta}\left\|\frac{1}{h_t(\theta)}\frac{\partial^2 h_t(\theta)}{\partial\theta\partial\theta'}\right\|\right]<\infty.$$

Assumption 8.

$$(i)\ \sup_{\theta\in\Theta}\left\|\frac{1}{\tilde{h}_t(\theta)}\frac{\partial\tilde{h}_t(\theta)}{\partial\theta}-\frac{1}{h_t(\theta)}\frac{\partial h_t(\theta)}{\partial\theta}\right\|\leq O(\rho^t)R_t,$$
$$(ii)\ \sup_{\theta\in\Theta}\left\|\frac{1}{\tilde{h}_t(\theta)}\frac{\partial^2\tilde{h}_t(\theta)}{\partial\theta\partial\theta'}-\frac{1}{h_t(\theta)}\frac{\partial^2 h_t(\theta)}{\partial\theta\partial\theta'}\right\|\leq O(\rho^t)R_t$$

for some constant $\rho\in(0,1)$ and positive random variable R_t such that $ER_t^2<\infty$.

Assumption 6 imposes some basic requirements on the function $h_t(\theta)$, and they are satisfied by most of conditionally heteroskedastic models; see, e.g., Francq and Zakoïan (2004, 2013). Assumptions 7-8 give some moment conditions, which have been verified for GARCH models in Ling (2007), asymmetric power GARCH models in Hamadeh and Zakoïan (2011) and asymmetric log-GARCH models in Francq, Wintenberger, and Zakoïan (2013). The following corollary gives the strong consistency and asymptotic normality the PQMLE for model (9), and its proof is omitted because it follows the same ones as for Theorems 1.1-1.2 in Berkes and Horváth (2004).

COROLLARY 1. *Assume that y_t follows model (9). If Assumptions* 1, 2*(iii) and* 3-8 *hold, then the conclusions in Theorem 1 hold.*

3. SIMULATION STUDY

In this section, we compare the performance of the PQMLE with those of the GQMLE, the LQMLE, the LADE and the GNGQMLE in finite samples. We generate 1000 replications of sample size n = 1000 from the following model:

$$y_t=\sigma_t\varepsilon_t \text{ and } \sigma_t^2=\omega_0+\alpha_0 y_{t-1}^2+\beta_0\sigma_{t-1}^2,\tag{10}$$

where we choose $(\omega_0,\alpha_0,\beta_0)=(0.25,0.15,0.3)$ as in Fan, Li, and Xiu (2013), and ε_t is chosen to be the PIV distributions, the STB distributions, and the Student's t distributions, respectively.

In order to implement the PQMLE, we choose $(\nu,m)=(\nu_0/\tau_0,m_0/\tau_0)$ such that Assumption 4 holds, where

$$\tau_0=E\left[\frac{2m_0\varepsilon_t^2+\nu_0\varepsilon_t}{1+\varepsilon_t^2}\right]$$

with $(\nu_0,m_0)=(2,2),(2,4),(-2,4)$ and $(0,4)$, and the corresponding PQMLEs are called the PQMLE$_1$, PQMLE$_2$, PQMLE$_3$, and PQMLE$_4$, respectively, Furthermore, since other four estimation methods require different identification conditions for model (10), the GQMLE ($\bar{\theta}_{1n}^*$), LQMLE ($\bar{\theta}_{2n}^*$), LADE ($\bar{\theta}_{3n}^*$), and GNGQMLE ($\bar{\theta}_{4n}^*$) are estimators of $(\tau_1\omega_0,\tau_1\alpha_0,\beta_0)$ with $\tau_1=E\varepsilon_t^2,(E|\varepsilon_t|)^2$, median$(\varepsilon_t^2)$ and $E\varepsilon_t^2$ respectively. In order to make our comparison feasible, we let

$$\bar{\theta}_{1n}=\left(\frac{\bar{\omega}_{1n}^*}{E\varepsilon_t^2},\frac{\bar{\alpha}_{1n}^*}{E\varepsilon_t^2},\bar{\beta}_{1n}^*\right)\quad\bar{\theta}_{2n}=\left(\frac{\bar{\omega}_{2n}^*}{(E|\varepsilon_t|)^2},\frac{\bar{\alpha}_{2n}^*}{(E|\varepsilon_t|)^2},\bar{\beta}_{2n}^*\right)$$

Table 1. *The bias and RMSE of all estimators for model (10)*

ε_t		Estimators											
PIV$(0,1,2,4)$		PQMLE$_1$			PQMLE$_2$			PQMLE$_3$			PQMLE$_4$		
		ω	α	β	ω	α	β	ω	α	β	ω	α	β
	Bias	-0.0034	0.0072	0.0071	0.0013	0.0023	-0.0050	-0.0033	0.0065	0.0049	-0.0021	0.0018	-0.0040
	RMSE	0.1110	0.1132	0.2979	0.1050	0.1010	0.2821	0.1111	0.1173	0.2983	0.1051	0.1041	0.2848
		GQMLE			LQMLE			LAD			GNGQMLE		
		ω	α	β	ω	α	β	ω	α	β	ω	α	β
	Bias	-0.0047	0.0102	0.0094	0.0069	0.0041	-0.0185	0.0075	0.0256	-0.0245	0.0016	0.0038	-0.0061
	RMSE	0.1122	0.1254	0.3029	0.1082	0.1075	0.2892	0.1156	0.1454	0.3077	0.1050	0.1049	0.2822
PIV$(0,1,2,2)$		PQMLE$_1$			PQMLE$_2$			PQMLE$_3$			PQMLE$_4$		
		ω	α	β	ω	α	β	ω	α	β	ω	α	β
	Bias	0.0059	0.0010	-0.0073	0.0047	0.0001	-0.0056	0.0037	0.0000	-0.0028	0.0041	0.0000	-0.0040
	RMSE	0.0445	0.0328	0.0881	0.0456	0.0334	0.0908	0.0547	0.0396	0.1097	0.0490	0.0358	0.0981
		GQMLE			LQMLE			LAD			GNGQMLE		
		ω	α	β	ω	α	β	ω	α	β	ω	α	β
	Bias	0.0087	0.0122	-0.0306	0.0080	0.0017	-0.0138	0.0036	0.0011	-0.0039	-0.0009	-0.0045	-0.0109
	RMSE	0.0900	0.0905	0.1728	0.0529	0.0419	0.1084	0.0554	0.0400	0.1137	0.0497	0.0354	0.1010
PIV$(0,1,2,1.6)$		PQMLE$_1$			PQMLE$_2$			PQMLE$_3$			PQMLE$_4$		
		ω	α	β	ω	α	β	ω	α	β	ω	α	β
	Bias	0.0044	0.0002	-0.0007	0.0042	0.0002	-0.0002	0.0042	0.0006	0.0005	0.0042	0.0003	0.0002
	RMSE	0.0420	0.0227	0.0477	0.0431	0.0235	0.0493	0.0498	0.0278	0.0582	0.0457	0.0252	0.0527
		GQMLE			LQMLE			LAD			GNGQMLE		
		ω	α	β	ω	α	β	ω	α	β	ω	α	β
	Bias	-0.0069	-0.0016	0.0139	0.0084	0.0024	-0.0076	0.0030	0.0011	-0.0010	-0.0189	-0.0149	-0.0045
	RMSE	0.1405	0.0829	0.2018	0.0605	0.0390	0.0767	0.0455	0.0261	0.0601	0.0523	0.0300	0.0575
PIV$(0,1,2,1.5)$		PQMLE$_1$			PQMLE$_2$			PQMLE$_3$			PQMLE$_4$		
		ω	α	β	ω	α	β	ω	α	β	ω	α	β
	Bias	0.0083	-0.0006	0.0010	0.0079	-0.0007	0.0016	0.0074	-0.0007	0.0029	0.0076	-0.0008	0.0023
	RMSE	0.0465	0.0205	0.0394	0.0477	0.0215	0.0411	0.0557	0.0256	0.0490	0.0507	0.0231	0.0442
		GQMLE			LQMLE			LAD			GNGQMLE		
		ω	α	β	ω	α	β	ω	α	β	ω	α	β
	Bias							0.0073	-0.0012	0.0020			
	RMSE							0.0515	0.0221	0.0485			
STB$(1.8,0.5,1,0)$		PQMLE$_1$			PQMLE$_2$			PQMLE$_3$			PQMLE$_4$		
		ω	α	β	ω	α	β	ω	α	β	ω	α	β
	Bias	0.0014	-0.0014	0.0025	0.0013	-0.0014	0.0018	0.0018	-0.0001	-0.0008	0.0014	-0.0010	0.0008
	RMSE	0.0565	0.0375	0.1108	0.0498	0.0335	0.0978	0.0506	0.0336	0.0981	0.0477	0.0321	0.0933
		GQMLE			LQMLE			LAD			GNGQMLE		
		ω	α	β	ω	α	β	ω	α	β	ω	α	β
	Bias							0.0016	0.0001	0.0002			
	RMSE							0.0552	0.0373	0.1070			
STB$(1.8,0.9,1,0)$		PQMLE$_1$			PQMLE$_2$			PQMLE$_3$			PQMLE$_4$		
		ω	α	β	ω	α	β	ω	α	β	ω	α	β
	Bias	0.0007	-0.0004	0.0023	0.0010	-0.0010	0.0020	0.0018	-0.0008	0.0008	0.0015	-0.0011	0.0013
	RMSE	0.0573	0.0366	0.1090	0.0518	0.0323	0.0984	0.0530	0.0328	0.1005	0.0504	0.0309	0.0953
		GQMLE			LQMLE			LAD			GNGQMLE		
		ω	α	β	ω	α	β	ω	α	β	ω	α	β
	Bias							0.0024	-0.0001	0.0004			
	RMSE							0.0595	0.0374	0.1120			

Table 2. *The bias and RMSE of all estimators for model* (10) *(con't)*

ε_t		Estimators											
STB$(1.5, 0, 1, 0)$		PQMLE$_1$			PQMLE$_2$			PQMLE$_3$			PQMLE$_4$		
		ω	α	β	ω	α	β	ω	α	β	ω	α	β
	Bias	0.0047	-0.0009	0.0007	0.0039	-0.0013	0.0015	0.0029	-0.0011	0.0023	0.0034	-0.0013	0.0019
	RMSE	0.0439	0.0305	0.0656	0.0389	0.0268	0.0580	0.0397	0.0262	0.0584	0.0376	0.0252	0.0556
		GQMLE			LQMLE			LAD			GNGQMLE		
		ω	α	β	ω	α	β	ω	α	β	ω	α	β
	Bias							0.0033	-0.0015	0.0018			
	RMSE							0.0425	0.0298	0.0644			
STB$(1.5, 0.5, 1, 0)$		PQMLE$_1$			PQMLE$_2$			PQMLE$_3$			PQMLE$_4$		
		ω	α	β	ω	α	β	ω	α	β	ω	α	β
	Bias	0.0029	-0.0001	-0.0001	0.0025	-0.0006	0.0006	0.0015	-0.0010	0.0027	0.0020	-0.0009	0.0016
	RMSE	0.0425	0.0276	0.0647	0.0391	0.0251	0.0593	0.0403	0.0259	0.0623	0.0382	0.0244	0.0583
		GQMLE			LQMLE			LAD			GNGQMLE		
		ω	α	β	ω	α	β	ω	α	β	ω	α	β
	Bias							0.0028	-0.0002	0.0009			
	RMSE							0.0430	0.0277	0.0650			
t_5		PQMLE$_1$			PQMLE$_2$			PQMLE$_3$			PQMLE$_4$		
		ω	α	β	ω	α	β	ω	α	β	ω	α	β
	Bias	0.0047	0.0073	-0.0120	0.0045	0.0041	-0.0131	0.0049	0.0025	-0.0086	0.0025	0.0016	-0.0100
	RMSE	0.0799	0.0533	0.1742	0.0716	0.0464	0.1566	0.0704	0.0453	0.1555	0.0675	0.0434	0.1496
		GQMLE			LQMLE			LAD			GNGQMLE		
		ω	α	β	ω	α	β	ω	α	β	ω	α	β
	Bias	0.0102	0.0039	-0.0237	0.0072	0.0028	-0.0154	0.0027	0.0059	-0.0078	0.0055	0.0019	-0.0151
	RMSE	0.0743	0.0519	0.1689	0.0634	0.0404	0.1440	0.0774	0.0538	0.1743	0.0622	0.0392	0.1423
t_4		PQMLE$_1$			PQMLE$_2$			PQMLE$_3$			PQMLE$_4$		
		ω	α	β	ω	α	β	ω	α	β	ω	α	β
	Bias	0.0063	0.0050	-0.0148	0.0068	0.0032	-0.0143	0.0068	0.0017	-0.0122	0.0071	0.0022	-0.0138
	RMSE	0.0700	0.0473	0.1521	0.0631	0.0415	0.1367	0.0636	0.0414	0.1363	0.0608	0.0393	0.1308
		GQMLE			LQMLE			LAD			GNGQMLE		
		ω	α	β	ω	α	β	ω	α	β	ω	α	β
	Bias				0.0081	0.0022	-0.0160	0.0052	0.0039	-0.0104	0.0060	0.0006	-0.0176
	RMSE				0.0591	0.0390	0.1289	0.0715	0.0493	0.1557	0.0564	0.0369	0.1235
t_3		PQMLE$_1$			PQMLE$_2$			PQMLE$_3$			PQMLE$_4$		
		ω	α	β	ω	α	β	ω	α	β	ω	α	β
	Bias	0.0013	-0.0018	-0.0046	0.0058	-0.0001	-0.0038	0.0014	-0.0026	0.0000	0.0073	0.0005	-0.0021
	RMSE	0.0602	0.0430	0.1181	0.0531	0.0384	0.1042	0.0507	0.0375	0.1032	0.0498	0.0366	0.0984
		GQMLE			LQMLE			LAD			GNGQMLE		
		ω	α	β	ω	α	β	ω	α	β	ω	α	β
	Bias				0.0069	0.0016	-0.0116	0.0025	-0.0013	0.0007	-0.0012	-0.0043	-0.0076
	RMSE				0.0521	0.0399	0.1080	0.0565	0.0419	0.1155	0.0472	0.0350	0.0993
t_2		PQMLE$_1$			PQMLE$_2$			PQMLE$_3$			PQMLE$_4$		
		ω	α	β	ω	α	β	ω	α	β	ω	α	β
	Bias	0.0025	0.0007	0.0005	0.0023	0.0001	0.0010	0.0024	0.0003	0.0009	0.0023	0.0000	0.0011
	RMSE	0.0476	0.0357	0.0807	0.0422	0.0319	0.0723	0.0429	0.0312	0.0740	0.0405	0.0303	0.0699
		GQMLE			LQMLE			LAD			GNGQMLE		
		ω	α	β	ω	α	β	ω	α	β	ω	α	β
	Bias							0.0012	0.0008	0.0014			
	RMSE							0.0471	0.0358	0.0829			

and

$$\bar{\theta}_{3n} = \left(\frac{\bar{\omega}^*_{3n}}{\text{median}(\varepsilon_t^2)}, \frac{\bar{\alpha}^*_{3n}}{\text{median}(\varepsilon_t^2)}, \bar{\beta}^*_{3n} \right) \quad \bar{\theta}_{4n} = \left(\frac{\bar{\omega}^*_{4n}}{E\varepsilon_t^2}, \frac{\bar{\alpha}^*_{4n}}{E\varepsilon_t^2}, \bar{\beta}^*_{4n} \right)$$

be the GQMLE, LQMLE, LADE, and GNGQMLE of $(\omega_0, \alpha_0, \beta_0)$, respectively. The estimated asymptotic standard deviations of all estimators were derived in a similar way. In all calculations, we use the true values of $E\varepsilon_t^2$, $(E|\varepsilon_t|)^2$ and median(ε_t^2), and the GNGQMLE is constructed in the same way as in Section 7.2 of Fan, Li, and Xiu (2013). Note that the PQMLEs and LADE are applicable for all innovations, but the GQMLE is only applicable when $E\varepsilon_t^4 < \infty$, and the LQMLE and GNGQMLE are only applicable when $E\varepsilon_t^2 < \infty$.

Tables 1-2 report the bias and root mean square error (RMSE) of all estimators for model (10). From them, we find that all estimators have very small bias. When $\eta_t \sim$ PIV$(0, 1, 2, 4)$, PQMLE$_2$ is the efficient estimator and so it has the smallest RMSE, while the performance of LQMLE or GNGQMLE is better than those of the remaining PQMLEs. When $\eta_t \sim$ PIV$(0, 1, 2, 2)$, PQMLE$_1$ is the efficient estimator and so it has the smallest RMSE. In this case, all PQMLEs except PQMLE$_3$ have smaller RMSEs than other estimators. This power advantage of PQMLEs becomes more significant as m becomes smaller. Note that the PQMLE$_3$ has the worse performance in all PQMLEs, and this is maybe because the sign of ν is negative for PQMLE$_3$. Next, we consider the cases that ε_t follows the STB distribution. In this case, only the PQMLEs and LADE are applicable. When $\varepsilon_t \sim$ STB$(1.8, 0.5, 1, 0)$, all PQMLEs except PQMLE$_1$ have smaller RMSEs than the LADE; when $\varepsilon_t \sim$ STB$(1.8, 0.9, 1, 0)$, the innovation becomes more skewed, and then the power advantage of all PQMLEs (including PQMLE$_1$) over LADE becomes more significant; moreover, when $\varepsilon_t \sim$ STB$(1.5, 0, 1, 0)$ or STB$(1.5, 0.5, 1, 0)$, the innovation become more heavy-tailed, and then the similar conclusions can be drawn as before. Thirdly, we consider the cases that ε_t follows the t distribution. In this case, the innovations are symmetric, and hence the QMLE$_4$ has the best performance in all PQMLEs, although its performance is worse than those of the LQMLE and GNGQMLE. Meanwhile, the GNGQMLE has the best performance in all estimators due to its adaption property, and the performance of the PQMLEs are always better than that of the LADE. Overall, the simulation study shows that all PQMLEs have a good performance in finite samples, especially for the heavy-tailed and skewed innovations.

4. APPLICATION

4·1. *Application to stock indexes*

In this subsection, we apply the PQMLE estimation method to eight major stock indexes in the world. The data sets we considered are the daily CAC40, DAX, DJTA, FTSE, HSI, NASDAQ, Nikkei225, and SP500 indexes from January 3, 2000 to December 27, 2007. As usual, we denote the log-return ($\times 100$) of each data set by $\{y_t\}_{t=1}^n$, and the summary of each y_t is given in Table 3. From this table, we find that each y_t is skewed and has a heavier tail than normal distribution. Hence, we use a GARCH(1,1) model with the PQMLE estimation method to fit each y_t. As a comparison, we also apply the GQMLE, LQMLE, or GNGQMLE estimation method to get the fitted GARCH(1,1) model for each y_t. For PQMLE method, ν and m are chosen as in Remark 3. For GNGQMLE method, the auxiliary likelihood function is written on the standardized t_3, t_5 or t_7 distribution such that it has variance one, and then the corresponding estimator is denoted by GNGQMLE$_1$, GNGQMLE$_2$ or GNGQMLE$_3$, respectively.

The detailed estimation results for each return series are given in Tables 4-5, in which the full log-likelihood function of the PQMLE is defined as in (8), and the full log-likelihood functions

Table 3. *Summary of eight major stock indexes*

y_t	n	mean	standard deviation	skewness	kurtosis
CAC40	2049	-0.0025	1.3968	-0.0930	5.9618
DAX	2031	0.0086	1.5495	-0.0455	5.7503
DJIA	2009	0.0081	1.0951	-0.0907	7.4136
FTSE	2017	-0.0012	1.1297	-0.1749	5.8796
HSI	1982	0.0238	1.3533	-0.3596	6.5512
NASDAQ	2007	-0.0216	1.8461	0.1848	7.2060
Nikkei225	1965	-0.0102	1.3796	-0.1581	4.7171
SP500	2007	0.0000	1.1155	0.0469	5.5460

of the GQMLE (LLF$_G$), LQMLE (LLF$_L$), and GNGQMLE (LLF$_{GNG}$) are defined as follows:

$$\begin{aligned}
\mathrm{LLF}_G &= -\sum_{t=1}^{n}\left[\log\sqrt{\tilde{h}_t(\bar{\theta}_{1n})}+\frac{y_t^2}{2\tilde{h}_t(\bar{\theta}_{1n})}\right]+n\log\left(\frac{1}{\sqrt{2\pi}}\right),\\
\mathrm{LLF}_L &= -\sum_{t=1}^{n}\left[\log\sqrt{\tilde{h}_t(\bar{\theta}_{2n})}+\frac{|y_t|}{\sqrt{\tilde{h}_t(\bar{\theta}_{2n})}}\right]+n\log\left(\frac{1}{2}\right),\\
\mathrm{LLF}_{GNG} &= -\sum_{t=1}^{n}\left[\log\left(\hat{\eta}_k\sqrt{\tilde{h}_t(\bar{\theta}_{4n})}\right)+\frac{k+1}{2}\left(1+\frac{y_t^2}{(k-2)\hat{\eta}_k^2\tilde{h}_t(\bar{\theta}_{4n})}\right)\right]\\
&\quad+n\log\left(\frac{\Gamma\{(k+1)/2\}}{\sqrt{(k-2)\pi}\Gamma\{k/2\}}\right)\ \text{ for } k=3 \text{ (or 5, 7)},
\end{aligned}$$

where $\bar{\theta}_{1n}$, $\bar{\theta}_{2n}$ and $\bar{\theta}_{4n}$ are the GQMLE, LQMLE and GNGQMLE, respectively, and

$$\hat{\eta}_k=\arg\max_{\eta}\sum_{t=1}^{n}\left[-\log(\eta)-\frac{k+1}{2}\left(1+\frac{y_t^2}{(k-2)\eta^2\tilde{h}_t(\bar{\theta}_{1n})}\right)\right].$$

Here, $\hat{\eta}_k$ measures the discrepancy between the true likelihood function and the given auxiliary likelihood function. Specifically, when $\hat{\eta}_k > 1$(or < 1), the given auxiliary innovation t_k is heavier (or lighter) than the true innovation. Furthermore, Tables 4-5 also report the estimated values of the identification condition τ_2 for each estimation method, that is, τ_2 is the sample mean of $(2m\varepsilon_t^2+\nu\varepsilon_t)/(1+\varepsilon_t^2)$, ε_t^2 or $|\varepsilon_t|$ for the PQMLE, GQMLE (and GNGQMLE) or LQMLE estimation method, respectively. Meanwhile, it is worth mentioning that all fitted models are adequate by looking at the the ACF and PACF plots (not depicted here) of the residuals.

From Tables 4-5, we find that (i) all the values of τ_2 are close to 1 as expected; (ii) for each return series, the PQMLE always has the best fitting in all estimation methods; (iii) the GNGQMLE estimation with a t_5 or t_7 likelihood gives the second best fitted models for the DJIA, HSI, Nikkei225 and SP500 return series in which the value of m are smaller, while the GQMLE estimation gives the second best fitted models for the CAC40, DAX, FISE and NASDAQ return series in which the value of m are larger; (iv) the LQMLE has the worse fitting in all cases except for the DJIA and HSI return series, in which the values of m are the smallest ones, and so the GQMLE has the worse fitting in these two cases; (v) the GNGQMLE estimation with a t_3 likelihood always has the largest value of $\hat{\eta}_k$ among all GNGQMLE estimations, and hence it implies the auxiliary t_3 innovation is heavier than the true innovation, while the auxiliary t_5 or t_7 innovation has the same tail as the true innovation because the values of $\hat{\eta}_k$ in these two cases are close to 1; (vi) the values of m are all larger than 2.5, and it suggests the innovation

Table 4. *Summary of all estimations for eight major stock indexes*

y_t		PQMLE	GQMLE	LQMLE	GNGQMLE_1	GNGQMLE_2	GNGQMLE_3
CAC40	ω	0.2301	0.0160	0.0071	0.0102	0.0114	0.0121
		(0.0876)†	(0.0060)	(0.0031)	(0.0019)	(0.0021)	(0.0022)
	α	1.3881	0.0851	0.0487	0.0776	0.0800	0.0812
		(0.2084)	(0.0138)	(0.0075)	(0.0003)	(0.0003)	(0.0004)
	β	0.9103	0.9068	0.9164	0.9185	0.9154	0.9137
		(0.0128)	(0.0143)	(0.0122)	(0.0100)	(0.0108)	(0.0113)
	ν	-0.0308					
	m	9.8482					
	$\hat{\eta}_k$				1.3462	1.0872	1.0372
	τ_2	0.9995	0.9995	0.9995	1.0012	1.0006	1.0003
	LLF	-3205.2	-3213.8	-3282.2	-3268.0	-3227.9	-3215.5
DAX	ω	0.3508	0.0240	0.0081	0.0095	0.0125	0.0142
		(0.1277)	(0.0081)	(0.0038)	(0.0022)	(0.0025)	(0.0027)
	α	1.8795	0.1062	0.0591	0.0925	0.0944	0.0954
		(0.2694)	(0.0161)	(0.0087)	(0.0004)	(0.0005)	(0.0005)
	β	0.8947	0.8845	0.9014	0.9074	0.9035	0.9013
		(0.0143)	(0.0167)	(0.0138)	(0.0107)	(0.0117)	(0.0123)
	ν	-0.0830					
	m	10.989					
	$\hat{\eta}_k$				1.3430	1.0880	1.0389
	τ_2	0.9995	0.9996	0.9996	1.0057	1.0036	1.0025
	LLF	-3358.9	-3366.0	-3425.4	-3420.4	-3382.2	-3370.1
DJIA	ω	0.0698	0.0261	0.0075	0.0112	0.0123	0.0128
		(0.0241)	(0.0115)	(0.0027)	(0.0031)	(0.0034)	(0.0035)
	α	0.4584	0.0847	0.0453	0.0801	0.0834	0.0845
		(0.0719)	(0.0246)	(0.0078)	(0.0005)	(0.0005)	(0.0006)
	β	0.9094	0.8934	0.9120	0.9150	0.9109	0.9094
		(0.0132)	(0.0287)	(0.0140)	(0.0186)	(0.0202)	(0.0211)
	ν	-0.0379					
	m	4.2961					
	$\hat{\eta}_k$				1.2666	1.0331	0.9909
	τ_2	0.9995	0.9995	0.9995	1.0086	1.0078	1.0077
	LLF	-2726.4	-2794.5	-2764.7	-2759.0	-2732.1	-2727.6
FTSE	ω	0.5639	0.0152	0.0091	0.0138	0.0138	0.0138
		(0.1743)	(0.0046)	(0.0029)	(0.0018)	(0.0018)	(0.0019)
	α	4.4984	0.1175	0.0699	0.1112	0.1136	0.1148
		(0.6032)	(0.0158)	(0.0099)	(0.0004)	(0.0004)	(0.0004)
	β	0.8728	0.8721	0.8774	0.8794	0.8774	0.8762
		(0.0157)	(0.0159)	(0.0161)	(0.0125)	(0.0130)	(0.0134)
	ν	-0.0028					
	m	20.676					
	$\hat{\eta}_k$				1.3533	1.0933	1.0430
	τ_2	0.9995	0.9996	0.9996	0.9996	0.9996	0.9995
	LLF	-2722.0	-2725.2	-2801.6	-2789.3	-2748.9	-2735.9

† The standard deviations are in parentheses.

Table 5. *Summary of all estimations for eight major stock indexes (con't)*

y_t		PQMLE	GQMLE	LQMLE	GNGQMLE_1	GNGQMLE_2	GNGQMLE_3
HSI	ω	0.0318	0.0414	0.0055	0.0048	0.0073	0.0087
		(0.0192)[†]	(0.0260)	(0.0036)	(0.0055)	(0.0066)	(0.0071)
	α	0.2192	0.1436	0.0378	0.0497	0.0534	0.0559
		(0.0410)	(0.0446)	(0.0079)	(0.0005)	(0.0007)	(0.0008)
	β	0.9463	0.8517	0.9319	0.9529	0.9477	0.9445
		(0.0098)	(0.0437)	(0.0138)	(0.0253)	(0.0283)	(0.0302)
	ν	-0.0741					
	m	3.5529					
	$\hat{\eta}_k$				1.2321	1.0163	0.9795
	τ_2	0.9995	1.0005	1.0002	1.1053	1.0938	1.0873
	LLF	-3174.6	-3272.3	-3191.4	-3195.3	-3177.3	-3176.7
NASDAQ	ω	0.1702	0.0104	0.0037	0.0043	0.0053	0.0059
		(0.0872)	(0.0047)	(0.0025)	(0.0014)	(0.0016)	(0.0017)
	α	1.3844	0.0650	0.0392	0.0620	0.0628	0.0634
		(0.2184)	(0.0112)	(0.0064)	(0.0002)	(0.0002)	(0.0002)
	β	0.9336	0.9319	0.9364	0.9387	0.9373	0.9363
		(0.0099)	(0.0110)	(0.0099)	(0.0080)	(0.0085)	(0.0089)
	ν	-0.0114					
	m	12.195					
	$\hat{\eta}_k$				1.3511	1.0917	1.0411
	τ_2	0.9995	0.9995	0.9995	1.0019	1.0014	1.0010
	LLF	-3576.9	-3583.7	-3652.9	-3643.0	-3602.6	-3589.5
Nikkei225	ω	0.1529	0.0292	0.0099	0.0106	0.0139	0.0161
		(0.0652)	(0.0120)	(0.0046)	(0.0026)	(0.0032)	(0.0035)
	α	0.6068	0.0940	0.0412	0.0573	0.0640	0.0687
		(0.1019)	(0.0179)	(0.0072)	(0.0003)	(0.0004)	(0.0005)
	β	0.9201	0.8960	0.9251	0.9396	0.9316	0.9261
		(0.0132)	(0.0192)	(0.0129)	(0.0093)	(0.0109)	(0.0120)
	ν	0.0013					
	m	5.6151					
	$\hat{\eta}_k$				1.3111	1.0669	1.0213
	τ_2	0.9995	0.9996	0.9996	1.0151	1.0107	1.0078
	LLF	-3289.5	-3310.9	-3333.3	-3330.3	-3299.9	-3292.5
SP500	ω	0.0579	0.0112	0.0036	0.0044	0.0057	0.0064
		(0.0257)	(0.0044)	(0.0017)	(0.0012)	(0.0014)	(0.0015)
	α	0.5753	0.0712	0.0382	0.0623	0.0664	0.0683
		(0.0914)	(0.0135)	(0.0063)	(0.0001)	(0.0002)	(0.0002)
	β	0.9265	0.9200	0.9323	0.9364	0.9311	0.9286
		(0.0111)	(0.0144)	(0.0106)	(0.0091)	(0.0102)	(0.0109)
	ν	-0.0166					
	m	5.6425					
	$\hat{\eta}_k$				1.3060	1.0637	1.0191
	τ_2	0.9995	0.9995	0.9995	1.0054	1.0031	1.0023
	LLF	-2763.5	-2786.5	-2807.6	-2804.4	-2774.1	-2766.6

[†] The standard deviations are in parentheses.

for each return series has the finite fourth moment. Overall, we know that all estimation methods are applicable, and the PQMLE estimation method taking into account both leptokurtosis and asymmetry of the innovation gives the best fitted models for all return series.

4·2. *Application to exchange rates*

In this subsection, we apply the PQMLE estimation method to four exchange rates. For each exchange rate, we use a length of 2001 daily data set up to December 13, 2013. Since the log-return ($\times 100$) of each exchange rates exhibits some correlations in its conditional mean, it is first fitted by an ARMA(2,2) model with the weighted LAD estimation method in Zhu and Ling (2013). Consequently, we denote the residual from each fitted ARMA(2,2) model by y_t. Table 6 gives the summary of each y_t, from which we find that each y_t is skewed and has a heavier tail than normal distribution. Hence, as in Subsection 4.1, we use a GARCH(1,1) model with the PQMLE, GQMLE, LQMLE and GNGQMLE estimation methods to fit each y_t. All of estimation results are summarized in Table 7, and all fitted models are adequate by looking at the the ACF and PACF plots (not depicted here) of the residuals. From Table 7, we find that the LKR/USD and TWD/USD return series have a very heavy tail because the values of m in these two cases are smaller than 2.5. This is consistent to the situation that both return series have the large values of kurtosis. Therefore, it is reasonable to see that the PQMLE method has a much better fit than other estimation methods in these two cases. Meanwhile, since the value of m for the LKR/USD return series is slightly smaller than 1.5, the innovation may have infinite variance, and hence only the PQMLE method is valid in this case. Next, for the SGD/USD and ZAR/USD return series, the values of m are larger than 2.5, and so all estimation methods are valid. In these two cases, the PQMLE method still gives the best fitted model for each return series. The advantage of PQMLE over GNGQMLE may be caused by including the asymmetry effect in the likelihood. Overall, the performance of the PQMLE is the best among all estimation methods.

Table 6. *Summary of four exchange rates*

y_t	n	mean	standard deviation	skewness	kurtosis
LKR/USD	2000	0.0047	0.2696	2.4234	52.915
SGD/USD	2000	0.0027	0.3672	0.2749	8.1305
TWD/USD	2000	0.0000	0.3140	-0.6832	18.492
ZAR/USD	2000	-0.0026	1.1102	0.2883	6.8280

5. CONCLUDING REMARKS

In this paper, we propose a PQMLE for GARCH models. Under the stationarity and weak moment conditions, the strong consistency and asymptotical normality of the PQMLE are obtained. Meanwhile, the PQMLE can apply to other conditionally heteroskedastic models with no further efforts. Unlike the existing QMLE estimators, the PQMLE is the first QMLE in the literature to take into account both leptokurtosis and asymmetry of the innovation, which are two well-known co-existing features in financial and economic data sets. Simulation study demonstrates that the PQMLE can achieve better efficiency than other estimators, especially when ε_t is heavy-tailed and skewed. Two applications to stock indexes and exchange rates further highlight the importance of the PQMLE method. All of these findings suggest that the PQMLE estimation method should have a wide application in practice.

Table 7. *Summary of all estimations for four exchange rates*

y_t		PQMLE	GQMLE	LQMLE	GNGQMLE$_1$	GNGQMLE$_2$	GNGQMLE$_3$
LKR/USD	ω	0.0011	0.0025	0.0005	0.0035	0.0027	0.0024
		(0.0002)†	(0.0018)	(0.0001)	(0.0008)	(0.0007)	(0.0007)
	α	0.2791	0.5852	0.1699	0.9243	0.8001	0.7309
		(0.0370)	(0.3491)	(0.0318)	(0.0017)	(0.0013)	(0.0011)
	β	0.4783	0.6779	0.6803	0.5545	0.6111	0.6433
		(0.0352)	(0.1140)	(0.0353)	(0.0860)	(0.0839)	(0.0823)
	ν	0.0326					
	m	1.4449					
	$\hat{\eta}_k$				0.7417	0.6423	0.6407
	τ_2	0.9995	0.9996	0.9996	1.0329	1.0118	1.0035
	LLF	1339.2	457.10	1242.9	1312.3	1239.6	1179.5
SGD/USD	ω	0.0056	0.0014	0.0006	0.0015	0.0016	0.0017
		(0.0019)	(0.0006)	(0.0003)	(0.0002)	(0.0002)	(0.0003)
	α	0.2453	0.0600	0.0339	0.0651	0.0659	0.0665
		(0.0451)	(0.0129)	(0.0062)	(0.0000)	(0.0000)	(0.0000)
	β	0.9262	0.9301	0.9330	0.9255	0.9234	0.9220
		(0.0128)	(0.0143)	(0.0117)	(0.0116)	(0.0125)	(0.0131)
	ν	0.0126					
	m	3.3869					
	$\hat{\eta}_k$				1.2530	1.0260	0.9878
	τ_2	1.0006	1.0031	1.0008	0.9897	0.9899	0.9901
	LLF	-5535.0	-6044.0	-5855.0	-5752.0	-5562.0	-5566.0
TWD/USD	ω	0.0059	0.0031	0.0017	0.0041	0.0042	0.0043
		(0.0015)	(0.0011)	(0.0005)	(0.0005)	(0.0006)	(0.0006)
	α	0.2585	0.1239	0.0718	0.1758	0.1672	0.1638
		(0.0404)	(0.0287)	(0.0121)	(0.0002)	(0.0002)	(0.0002)
	β	0.8140	0.8537	0.8314	0.8052	0.8103	0.8109
		(0.0244)	(0.0292)	(0.0241)	(0.0275)	(0.0283)	(0.0294)
	ν	-0.0047					
	m	2.2108					
	$\hat{\eta}_k$				1.1252	0.9381	0.9132
	τ_2	1.0000	1.0030	1.0007	0.9986	0.9975	0.9970
	LLF	-1159.0	-2951.0	-1233.0	-1177.0	-1254.0	-1427.0
ZAR/USD	ω	0.3595	0.0264	0.0147	0.0243	0.0257	0.0264
		(0.1000)	(0.0076)	(0.0044)	(0.0028)	(0.0031)	(0.0032)
	α	0.8461	0.0641	0.0385	0.0614	0.0625	0.0632
		(0.1680)	(0.0129)	(0.0077)	(0.0004)	(0.0004)	(0.0004)
	β	0.9111	0.9120	0.9155	0.9168	0.9142	0.9128
		(0.0160)	(0.0162)	(0.0156)	(0.0124)	(0.0133)	(0.0138)
	ν	0.0227					
	m	8.0956					
	$\hat{\eta}_k$				1.3060	1.0637	1.0191
	τ_2	0.9995	0.9998	0.9996	1.3339	1.0802	1.0322
	LLF	-2830.9	-2839.5	-2895.6	-2884.4	-2848.3	-2838.0

† The standard deviations are in parentheses.

ACKNOWLEDGEMENT

This work is supported by Research Grants Council of the Hong Kong SAR Government, GRF grant HKU703711P, and National Natural Science Foundation of China (No.11201459).

APPENDIX: PROOF OF THEOREM 1

Recall that the first, second and third derivatives of $g(y,s)$ with respective to s are $g_1(y,s)$, $g_2(y,s)$ and $g_3(y,s)$, respectively. By a simple algebra, we can show that

$$\begin{aligned} g_1(y,s) &= \frac{1}{s} - \frac{2my^2 s}{1+y^2s^2} - \frac{\nu y}{1+y^2s^2}, \\ g_2(y,s) &= -\frac{1}{s^2} - \frac{2my^2}{1+y^2s^2} + \frac{2y^2 s(2my^2 s + \nu y)}{[1+y^2s^2]^2}, \\ g_3(y,s) &= \frac{2}{s^3} + \frac{12my^4 s + 2\nu y^3}{[1+y^2s^2]^2} - \frac{16my^6s^3 + 8\nu y^5 s^2}{[1+y^2s^2]^3}, \end{aligned}$$

where $s > 0$. Next, it is straightforward to see that

$$\begin{aligned} |g_1(y,s)| &\leq \frac{1}{s} + \frac{2m}{s} + \frac{|\nu||y|}{2s|y|} = \frac{1+2m+|\nu|/2}{s}, \\ |g_2(y,s)| &\leq \frac{1}{s^2} + \frac{2m}{s^2} + \frac{4ms^2y^4}{y^4s^4} + \frac{2s|\nu||y|^3}{[1+y^2s^2]^{3/2}} \\ &\leq \frac{1+6m}{s^2} + \frac{2s|\nu||y|^3}{s^3|y|^3} = \frac{1+6m+2|\nu|}{s^2}, \\ |g_3(y,s)| &\leq \frac{2}{s^3} + \frac{12m}{s^3} + \frac{2|\nu||y|^3}{[1+y^2s^2]^{3/2}} + \frac{16m}{s^3} + \frac{8|\nu||y|^5s^2}{[1+y^2s^2]^{5/2}} \\ &\leq \frac{2+28m}{s^3} + \frac{2|\nu||y|^3}{s^3|y|^3} + \frac{8|\nu||y|^5 s^2}{s^5|y|^5} = \frac{2+28m+10|\nu|}{s^3}. \end{aligned}$$

Thirdly, for some $\kappa_0 \in (0, \kappa)$, by Assumption 3(iii) and Jansen's inequality, we have

$$\begin{aligned} E|\log \bar{f}(\varepsilon_t s)| &= E|m\log(1+\varepsilon_t^2 s^2) + \nu \tan^{-1}(\varepsilon_t s)| \\ &\leq \frac{m}{\kappa_0} E\log(1+\varepsilon_t^2 s^2)^{\kappa_0} + |\nu|\frac{\pi}{2} \\ &\leq O(1)\log[1+E|\varepsilon_t|^{2\kappa_0} s^{2\kappa_0}] + O(1) \\ &\leq O(1)(s^{2\kappa_0}+1). \end{aligned}$$

Therefore, under Assumptions 1-5, we have verified all the conditions for Theorems 1.1-1.2 in Berkes and Horváth (2004). Hence, the conclusions in Theorem 1 hold. This completes the proof.

REFERENCES

ANDREWS, B. (2012). Rank-based estimation for GARCH processes. *Econometric Theory* **28**, 1037-1064.

BAI, X., RUSSELL, J.R. & TIAO, G.C. (2003). Kurtosis of GARCH and stochastic volatility models with non-normal innovations. *Journal of Econometrics* **114**, 349–360.

BAUWENS, L. & LAURENT, S. (2005). A New Class of Multivariate Skew Densities, with Application to Generalized Autoregressive Conditional Heteroscedasticity Models. *Journal of Business & Economic Statistics* **23**, 346–354.

BERA, A.K. & HIGGINS, M.L. (1993). ARCH models: Properties, estimation and testing. *Jounal of Economic Surveys* **7**, 305-366; reprinted in Surveys in Econometrics (L. Oxley et al., eds.) 215-272. Blackwell, Oxford 1995.

BERKES, I., HORVÁTH, L. & KOKOSZKA, P. (2003) GARCH processes: Structure and estimation. *Bernoulli* **9**, 201-227.

BOUGEROL, P. & PICARD, N. (1992). Stationarity of GARCH processes and of some nonnegative time series. *Journal of Econometrics* **52**, 115-127.

BERKES, I. & HORVÁTH, L. (2004). The efficiency of the estimators of the parameters in GARCH processes. *Annals of Statistics* **32**, 633-655.

BHATTACHARYYA, M., MISRA, N. & KODASE, B. (2009). MaxVaR for non-normal and heteroskedastic returns. *Quantitative Finance* **9**, 925–935.

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

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.

BOLLERSLEV, T., CHOU, R.Y. & KRONER, K.F. (1992). ARCH modeling in finance: A review of the theory and empirical evidence. *Journal of Econometrics* **52**, 5-59

CHEN, M. & ZHU, K. (2013). Sign-based portmanteau test for ARCH-type models with heavy-tailed innovations. Working paper. Chinese Academy of Sciences.

CHRISTOFFERSEN, P., HESTON, S. & JACOBS, K. (2006). Option valuation with conditional skewness. *Journal of Econometrics* **131**, 253-284.

DING, Z., GRANGER, C.W.J. & ENGLE, R.F. (1993). A long memory property of stock market returns and a new model. *Journal of Empirical Finance* **1**, 83-106.

DROST, F.C. & KLAASSEN, C.A.J. (1997). Efficient estimation in semiparametric garch models. *Journal of Econometrics* **81**, 193-221.

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

ENGLE, R.F. & GONZÁLEZ-RIVERA, G. (1991). Semiparametric arch models. *Journal of Business and Economic Statistics* **9**, 345-359.

GRIGOLETTO, M. & LISI, F. Looking for skewness in financial time series. *Econometrics Journal* **12**, 310-323.

HANSEN, B.E. (1994). Autoregressive conditional density estimation. *International Economic Review* **35**, 705-730.

HARVEY, C.R. & SIDDIQUE, A. (1999). Autoregressive conditional skewness. *Journal of Financial and Quantitative Analysis* **34**, 465-487.

HEINRICH, J. (2004). A guide to the Pearson type IV distribution. Working paper. University of Pennsylvania.

FAN, J., QI, L. & XIU, D. (2013). Quasi maximum likelihood estimation of GARCH models with heavy-tailed likelihoods. *Journal of Business and Economic Statistics.* forthcoming.

FRANCQ, C., WINTENBERGER, O. & ZAKOÏAN, J.M. (2013). Garch models without positivity constraints: exponential or log garch? *Journal of Econometrics.* forthcoming.

FRANCQ, C. & ZAKOÏAN, J.M. (2004). Maximum likelihood estimation of pure GARCH and ARMA-GARCH processes. *Bernoulli* **10**, 605-637.

FRANCQ, C. & ZAKOÏAN, J.M. (2010). GARCH Models: Structure, Statistical Inference and Financial Applications. Wiley, Chichester, UK.

FRANCQ, C. & ZAKOÏAN, J.M. (2013). Optimal predictions of powers of conditionally heteroscedastic processes. *Journal of the Royal Statistical Society B* **75**, 345-367.

GEWEKE, J. (1986). Modeling the persistence of conditional variances: A comment. *Econometric Review* **5**, 57-61.

HAMADEH, T. & ZAKOÏAN, J.M. (2011). Asymptotic properties of LS and QML estimators for a class of nonlinear GARCH processes. *Journal of Statistical Planning and Inference* **141**, 488-507.

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

JONDEAU, E. & ROCKINGER, M. (2001). Gram-Charlier densities. *Journal of Economic Dynamics and Control* **25**, 1457-1483.

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

LIU, S.-M. & BRORSEN, B.W. (1995). Maximum likelihood estimation of a GARCH-stable model. *Journal of applied econometrics* **10**, 273–285.

NAGAHARA, Y. (1999). The PDF and CF of Pearson type IV distributions and the ML estimation of the parameters. *Statistics & Probability Letters* **43**, 251-264.

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

NELSON, D.B. (1991). Conditional heteroskedasticity in asset returns: A new approach. *Econometrica* **59**, 347-370.

NEWEY, W.K. & STEIGERWALD, D.G. (1997). Asymptotic bias for quasi-maximum likelihood estimators in conditional heteroskedasticity models. *Econometrica* **65**, 587-599.

PENG, L. & YAO, Q. (2003). Least absolute deviations estimation for ARCH and GARCH models. *Biometrika* **90**, 967-975.

PREMARATNE, G. & BERA, A.K. (2001). Modeling asymmetry and excess kurtosis in stock return data. Working paper. University of Illinois.

YAN, J. (2005). Asymmetry, fat-tail, and autoregressive conditional density in fincancial return data with systems of frequency curves. Working paper. University of Iowa.

Verhoeven, P. McAleer, M. (2004). Fat tails and asymmetry in financial volatility models. *Mathematics and Computers in Simulation* **64**, 351-361.

Zhu, K. & Ling, S. (2011). Global self-weighted and local quasi-maximum exponential likelihood estimators for ARMA-GARCH/IGARCH models. *Annals of Statistics* **39**, 2131-2163.

Zhu, K. & Ling, S. (2013). Inference for ARMA models with unknown-form and heavy-tailed G/ARCH-type noises. Working paper. Hong Kong University of Science and Technology.