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, 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 just the heavy-tailed but also the skewed innovations. Under strict stationarity and some weak moment conditions, the strong consistency and asymptotic normality of the PQMLE are obtained. With no further efforts, the PQMLE can be applied to other conditionally heteroskedastic models. A simulation study is carried out to assess the performance of the PQMLE. Two applications to four major stock indexes and two exchange rates further highlight the importance of our new method. Heavy-tailed and skewed innovations are often observed together in practice, and the PQMLE now gives us a systematical way to capture these two co-existing features.

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 seminal 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 studies 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 asymptotically normal. However, the GQMLE can not capture the heavy-tailedness and skewness of ε_t, which are two well-observed features of financial data in GARCH model applications; 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 Pearson's Type IV distribution in Premaratne and Bera (2001), the Gram-Charlier distribution in Leon, Rubio, and Serna (2005) and Cheng et al. (2011), the mixture normal distribution in Bai, Russell, and Tiao (2003) and many others. However, the true distribution of ε_t is unknown a priori in practice, and as shown in White (1982) and 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 an efficiency advantage over GQMLE when ε_t is heavy-tailed. Generally, there are two ways to obtain a consistent NGQMLE.

First, one can assume a different identification condition other than $E\varepsilon_t^2 = 1$. For instance, Peng and Yao (2003) proposed the least absolute deviation estimator (LADE) under the identification condition that median$(\varepsilon_t^2) = 1$, and the consistency and asymptotic normality of the LADE was proved in Chen and Zhu (2014) 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 asymptotically normal when ε_t has a finite second moment; see also Li and Li (2008) and Zhu and Ling (2011) for more discussions in this context. Second, one can retain the identification condition $E\varepsilon_t^2 = 1$ for the NGQMLE by re-parameterizing 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, Qi, and Xiu (2014). 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 strict stationarity and a finite fractional moment of ε_t, the strong consistency and asymptotic normality of the PQMLE are obtained. Therefore, the PQMLE is applicable to all of the aforementioned non-Gaussian distributions used in the MLE method. Furthermore, we show that the PQMLE can be easily applied to other conditionally heteroskedastic models. A simulation study is carried out to assess the performance of the PQMLE, and two applications to four major stock indexes and two exchange rates further highlight the importance of our new method. Compared to the existing NGQMLEs, the PQMLE captures not only the heavy-tailed but also the skewed innovations. Heavy-tailed and skewed innovations are often observed together in practice, but none of the existing QMLE methods has focused on these co-existing features in the literature. The PQMLE method, which 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 properties. Simulation results are reported in Section 3. Applications are 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, $\|A\|$ becomes $|A|$ when A is a scalar, $O(1)$ denotes a bounded generic constant, and "$\rightarrow_d$" denotes the convergence in distribution.

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 parameter of the model given by (1)-(2) and its true value be θ_0. Denote the parameter space by Θ, where Θ is a subset of $\mathcal{R}_0^{1+p+q}$ with $\mathcal{R}_0 = [0, \infty)$. Then, we need the following assumptions:

Assumption 1. y_t is strictly stationary.

Assumption 2. (i) Θ is compact; (ii) 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 the technical conditions for proving our asymptotic theory. Note that only a finite fractional moment of ε_t is required in this case, and so our method 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 $x \in \mathcal{R}$ with $\mathcal{R} = (-\infty,\infty)$, and (λ,a,ν,m) are real parameters with $m > 1/2$ and $a > 0$. Here, K is the normalizing constant given by

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

where $i = \sqrt{-1}$ is the imaginary number and $\Gamma(\cdot)$ is the complex Gamma function. From Nagahara (1999), we know that if $\varepsilon_t \sim$ PIV(λ,a,ν,m), the mean, variance, skewness, and kurtosis of ε_t are

$$\begin{aligned}
\text{mean}(\varepsilon_t) &= \lambda - \frac{a\nu}{r} \quad \text{for } m > 1,\\
\text{var}(\varepsilon_t) &= \frac{a^2(r^2+\nu^2)}{r^2(r-1)} \quad \text{for } m > 1.5,\\
\text{skew}(\varepsilon_t) &= \frac{-4\nu}{r-2}\sqrt{\frac{r-1}{r^2+\nu^2}} \quad \text{for } m > 2,\\
\text{kurt}(\varepsilon_t) &= \frac{3(r-1)\left[(r+6)(r^2+\nu^2)-8r^2\right]}{(r-2)(r-3)(r^2+\nu^2)} \quad \text{for } m > 2.5,
\end{aligned}$$

respectively, where $r = 2(m-1)$. It is easy to see that $f(x;\lambda,a,\nu,m) = f(x-\lambda;0,a,\nu,m)$ and $f(x;0,a,\nu,m) = a^{-1}f(x/a;0,1,\nu,m)$. Thus, as in the conventional way, we treat λ and a as the location and the scale parameters, respectively. Meanwhile, it is straightforward to see that 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 $f(x;0,1,\nu,m)$ densities. From Figure 1, we can find that PIV distribution has a heavier left tail or right tail than the N$(0,1)$ distribution. This is reasonable since the j-th moment of PIV distribution exists only when $j < r+1$, while all the moments of the N$(0,1)$ distribution are finite. Moreover, when we draw the 3-dimensional plots of $\{(\nu,m,\text{skew}(\varepsilon_t))\}$ and $\{(\nu,m,\text{kurt}(\varepsilon_t))\}$ (not displayed here and available from us), we can

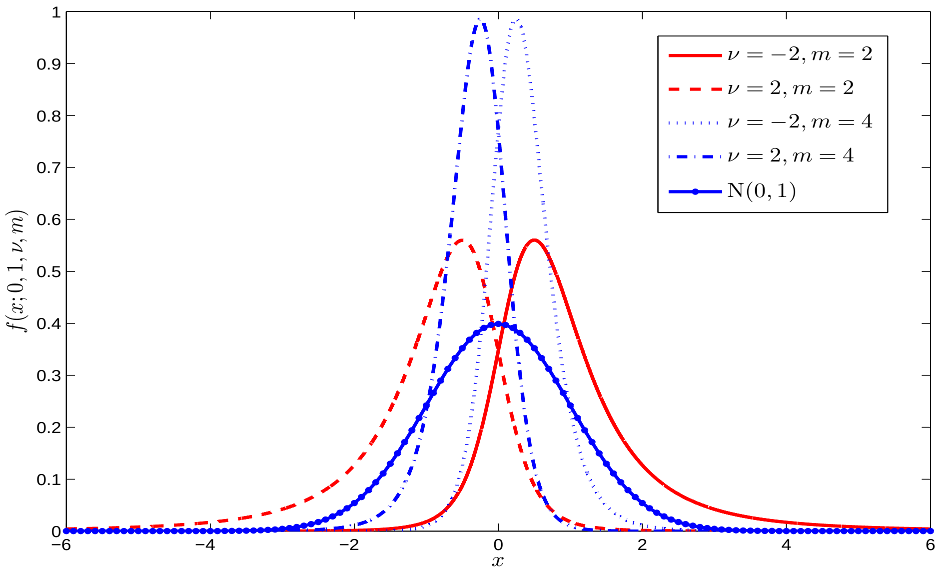

Fig. 1. The plot of four different densities $f(x; 0, 1, \nu, m)$ for the Pearson's Type IV distribution (the solid star line is the density of the N(0, 1) distribution).

see that when $|\nu|$ (or m) increases, the absolute value of skew(ε_t) increases (or decreases) for fixed m (or ν); and the same conclusion holds for kurt(ε_t). 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 the PIV distribution, we refer to Bauwens and Laurent (2005), Yan (2005), and Grigoletto and Lisi (2009).

Next, we are interested in the case when ε_t in models (1)-(2) follows the PIV distribution. Figure 2 plots one realization for each pair of (ν, m) 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 $\varepsilon_t \sim$ PIV$(0, 1, \nu, m)$ with $(\nu, m) = (\pm 2, 2)$, $(0, 2)$, $(\pm 2, 4)$, and $(0, 4)$. From Figure 2, we find that no matter how heavy-tailed ε_t is, y_t has a higher probability to be positive (or negative) 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 larger, especially for its extreme values. All of these findings indicate that the GARCH model with PIV$(0, 1, \nu, m)$ innovations can capture a very large range of asymmetry and leptokurtosis in the data set.

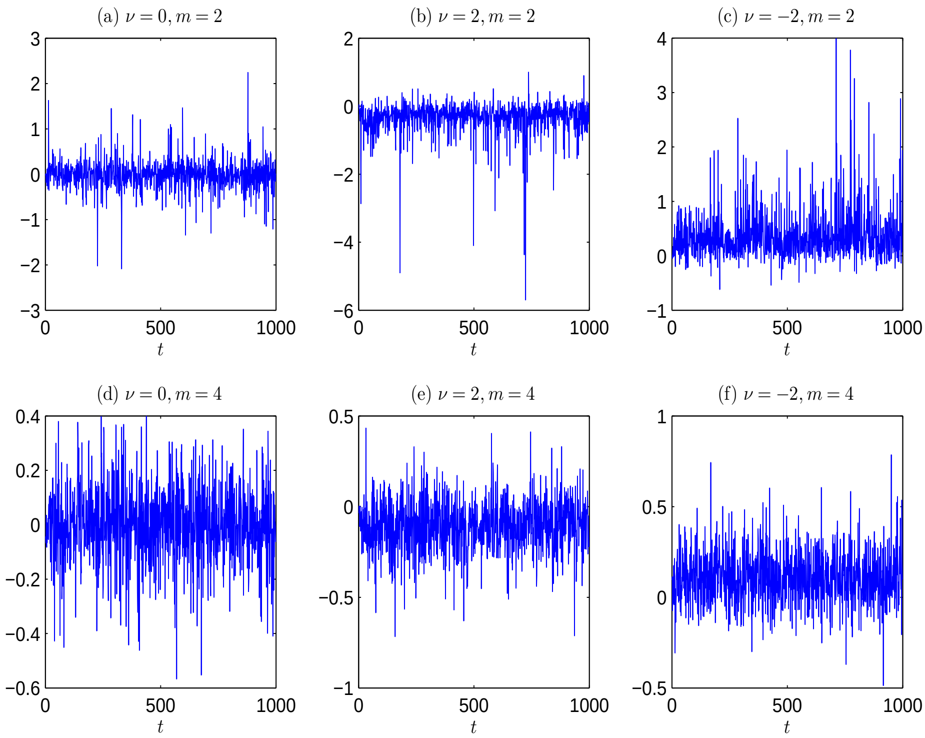

Fig. 2. One realization $\{y_t\}_{t=1}^{1000}$ from model (4), when $\varepsilon_t \sim \text{PIV}(0, 1, \nu, m)$.

2·3. *The PQMLE*

Given the observations $\{y_n, \cdots, y_1\}$ and the initial values $Y_0 := \{y_i; i \le 0\}$, we first rewrite 145
the parametric models (1)-(2) as

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

where all expressions for $c_i(\theta)$ $(i \ge 0)$ 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 150
can replace them by zeros, and then use $\tilde{h}_t(\theta)$ instead of $h_t(\theta)$ to calculate our estimator, where

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

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

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

where $\Gamma = \mathcal{R} \times (1/2, \infty)$. We look for the maximizer of $\tilde{L}_n(\theta)$ on Θ, that is,

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

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 tuning 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 properties of $\tilde{\theta}_n$, we need one more assumption below:

Assumption 4. The equation $u(c) = 1$ has a unique positive solution at $c = 1$, where

$$u(c) = E\left[\frac{2m(c\varepsilon_t)^2 + \nu c\varepsilon_t}{1 + (c\varepsilon_t)^2} \right] \quad \text{for } c > 0.$$

Assumption 4 is the identification condition for the PQMLE. From Assumption 4, we know that

$$E\left[\frac{2m\varepsilon_t^2 + \nu\varepsilon_t}{1 + \varepsilon_t^2} \right] = 1, \quad (8)$$

under which the conditional variance of y_t is $\sigma_t^2 var(\varepsilon_t)$, provided that $E\varepsilon_t^2 < \infty$. It is easy to see that the condition in (8) is the identifiability condition appeared in Berkes and Horváth (2004), and hence our identifiability condition is not exactly the same as theirs. When $\nu = 0$, we know that Assumption 4 and the condition in (8) are equivalent. However, when $\nu \neq 0$, the condition in (8) alone can not guarantee the identification of the PQMLE. To see the reason, we rewrite models (1)-(2) as

$$y_t = \sigma_t^* \varepsilon_t^* \ \text{ and } \ (\sigma_t^*)^2 = \omega^* + \sum_{i=1}^{p} \alpha_i^* y_{t-i}^2 + \sum_{j=1}^{q} \beta_j (\sigma_{t-j}^*)^2,$$

where $c > 0$, $\sigma_t^* = \sigma_t / c$, $\varepsilon_t^* = c\varepsilon_t$, $\omega^* = \omega / c^2$, $\alpha_i^* = \alpha_i / c^2$, and ε_t satisfies the condition in (8). Then, if there exists a $c \neq 1$ such that ε_t^* satisfies the condition in (8) (i.e., $u(c) = 1$ for some positive $c \neq 1$), the PQMLE can not be identified under (8). Thus, we need Assumption 4 to rule out this un-desirable situation. Note that for given $(\nu, m) \in \Gamma$, the condition in (8) does not directly hold for many often used distributions of ε_t. We now give a sufficient condition below for the standardization of ε_t:

Condition 1. For given $(\nu, m) \in \Gamma$, (i) $u(c)$ is a continuous function on $\{c : c > 0\}$; and (ii) there exists a constant $\underline{c} > 0$ such that $u(c) < 1$ on $\{c : c < \underline{c}\}$, and $u(c)$ is a strictly increasing function on $\{c : c \geq \underline{c}\}$.

Note that Condition 1(i) holds when ε_t has a continuous density function, and then Condition 1(ii) holds when $u(c)$ is a strictly increasing function on $\{c : c > 0\}$. Because $\lim_{c \to \infty} u(c) = 2m > 1$ for any given $(\nu, m) \in \Gamma$, there must exist a unique $c_0 > 0$ such that $u(c_0) = 1$ under Condition 1. This means that we can always standardize ε_t into ε_t^*, where $\varepsilon_t^* = c_0 \varepsilon_t$, and ε_t^* satisfies Assumption 4. For $(\nu, m) = (\pm 2, 4)$, Figure 3 plots the function $u(c)$ for N$(0, 1)$, Laplace$(0, 1)$, t_3, t_4, STB$(1.77, 0.2, 0.9, -0.02)$, STB$(1.82, -0.15, 1.1, -0.02)$, PIV$(0, 1, 2, 4)$, or PIV$(0.02, 0.9, -2, 4)$ distribution. Here, the STB$(\breve{\alpha}, \breve{\beta}, \breve{c}, \breve{\mu})$ distribution [see, e.g., Nolan (1997) and Andrews, Calder, and Davis (2009)] has the following characteristic function:

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

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

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

From Figure 3, we find that although $u(c)$ is not a strictly increasing function on $\{c : c > 0\}$ when $\varepsilon_t \sim$ PIV$(0, 1, 2, 4)$ for $(\nu, m) = (2, 4)$ or $\varepsilon_t \sim$ PIV$(0.02, 0.9, -2, 4)$ for $(\nu, m) = (-2, 4)$, each $u(c)$ satisfies Condition 1, and hence all of these distributions can be standardized to satisfy Assumption 4. Thus, for each $(\nu, m) \in \Gamma$, we conjecture that Assumption 4 holds for a large enough family of innovation distributions. In application (see, e.g., Section 4), we can always easily check the validation of this assumption by plotting $\tilde{u}_n(c)$ [i.e., the sample counterparts of $u(c)$], where

$$\tilde{u}_n(c) = \frac{1}{n} \sum_{t=1}^{n} \frac{2m(c\tilde{\varepsilon}_t)^2 + \nu c\tilde{\varepsilon}_t}{1 + (c\tilde{\varepsilon}_t)^2} \tag{9}$$

for $c > 0$, and $\tilde{\varepsilon}_t = y_t / \sqrt{\tilde{h}_t(\tilde{\theta}_n)}$ being the residual of models (1)-(2).

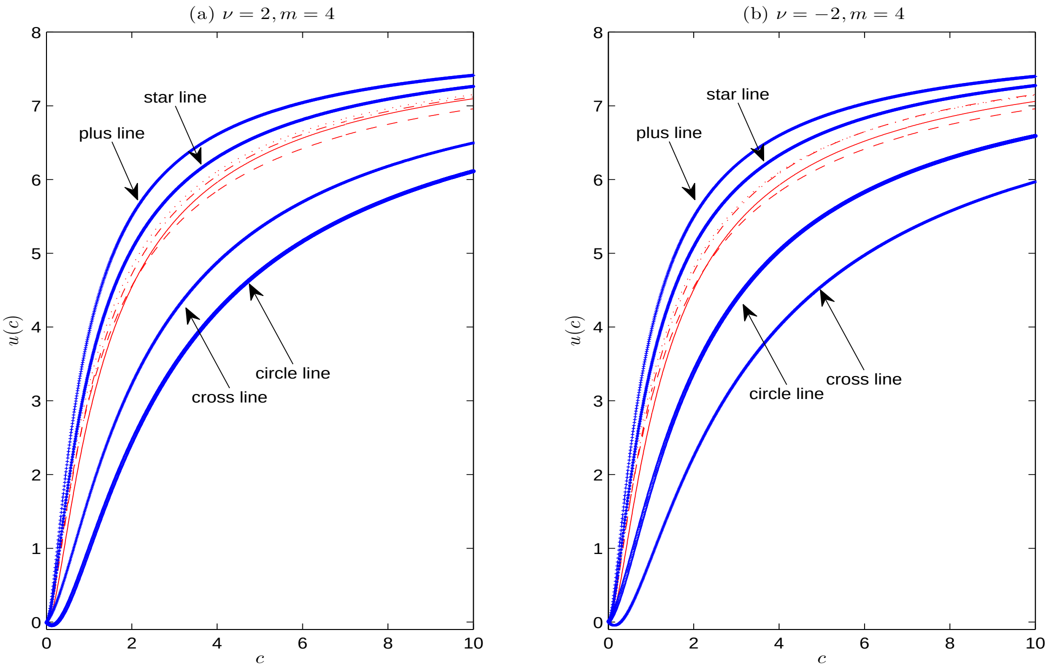

Fig. 3. The plots of $u(c)$ for N$(0, 1)$ (solid line), Laplace$(0, 1)$ (dashed line), t_3 (dotted line), t_4 (dash-dot line), STB$(1.77, 0.2, 0.9, -0.02)$ (star line), STB$(1.82, -0.15, 1.1, -0.02)$ (plus line), PIV$(0, 1, 2, 4)$ (circle line), and PIV$(0.02, 0.9, -2, 4)$ (cross line).

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 are now ready to give our main results:

THEOREM 1. *Suppose that Assumptions* 1-4 *hold. Then, as* $n \to \infty$, *(i)* $\tilde{\theta}_n \to \theta_0$ *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, Qi, and Xiu (2014). Note that as shown in Chen and Zhu (2014), 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 formally, and the simulation comparison in Section 3 is necessary.

Remark 3. To calculate the PQMLE in (7), we need to first select the tuning 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), Bhattacharyya, Chaudhary, and Yadav (2008), 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

$$\text{LLF}_P(\nu, m, \theta) = \tilde{L}_n(\theta) + n \log K. \tag{10}$$

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 will be illustrated by simulation studies in Section 3 and some empirical examples in Section 4.

Remark 4. In the construction of (6), we set the location parameter $\lambda = 0$ and scale parameter $a = 1$. The reason is that when $\lambda \neq 0$, we need to modify $g(y, s)$ as

$$g(y, s) = \log s - m \log \left[1 + \left(\frac{ys - \lambda}{a} \right)^2 \right] - \nu \tan^{-1} \left(\frac{ys - \lambda}{a} \right),$$

and then we can not verify conditions (1.14), (1.17), and (1.19) in Berkes and Horváth (2004). Thus, it is convenient to set $\lambda = 0$, since there is no theoretical result to guarantee that Theorem 1 holds when $\lambda \neq 0$. Moreover, when $\lambda = 0$ and $\varepsilon_t \sim$ PIV$(0, a, \nu, m)$, model (1) can be written as $y_t = \sigma_t^* \varepsilon_t^*$, where $\sigma_t^* = a\sigma_t$ and $\varepsilon_t^* = \varepsilon_t / a$ such that $\varepsilon_t^* \sim$ PIV$(0, 1, \nu, m)$. Thus, based on

$\varepsilon_t \sim$ PIV$(0, a, \nu, m)$ and $\varepsilon_t^* \sim$ PIV$(0, 1, \nu, m)$, the full log-likelihood functions as in (10) are the same. In view of this, we can set $a = 1$ for simplicity.

Remark 5. Note that the value of (ν, m) can be anywhere in Γ, and a different value of (ν, m) will imply a different stationarity region of y_t. To see this, Figure 4 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$(0, 1, \nu, m)$ with $(\nu, m) = (1, 2), (2, 2), (1, 4)$, and $(2, 4)$. As a comparison, the stationarity regions for the cases that $\varepsilon_t \sim N(0, 1)$ in Nelson (1990) and $\varepsilon_t \sim$ Laplace$(0, 1)$ in Zhu and Ling (2011) are also re-plotted in Figure 4. Here, we do not normalize ε_t in each case to satisfy the identifiability condition of one estimation method, since the stationarity region of y_t is invariant to the normalization of ε_t. From Figure 4, 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 for $\varepsilon_t \sim$ PIV$(0, 1, 2, 2)$, each strict stationarity region for other PIV distributions is much larger than that for $\varepsilon_t \sim N(0, 1)$ or that for $\varepsilon_t \sim$ 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 PQMLE for the following conditionally heteroskedastic models:

$$y_t = \sigma_t \varepsilon_t \;\text{ and }\; \sigma_t = \sigma(y_{t-1}, y_{t-2}, \cdots; \theta_0), \tag{11}$$

where ε_t being independent of $\{y_j; j < t\}$ is a sequence of i.i.d. random variables, the parameter space Θ is a subset of $\mathcal{R}^l$, 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 the GARCH model in (1)-(2), the asymmetric power GARCH model in Ding, Granger, and Engle (1993), and the asymmetric log-GARCH model in Geweke (1986), can be embedded into model (11); see, e.g., Bollerslev, Chou, and Kroner (1992), Tsay (2005), and Francq and Zakoïan (2010) for more discussions in this context.

As for (5), let $h_t(\theta) = [\sigma(y_{t-1}, y_{t-2}, \cdots; \theta)]^2$ and define $\tilde{h}_t(\theta)$ in the same way as $h_t(\theta)$ by replacing Y_0 by zeros. Then, based on $\{\tilde{h}_t(\theta)\}$, we can define the PQMLE for model (11) as in (7). To derive the asymptotic properties of the PQMLE, three more assumptions are needed.

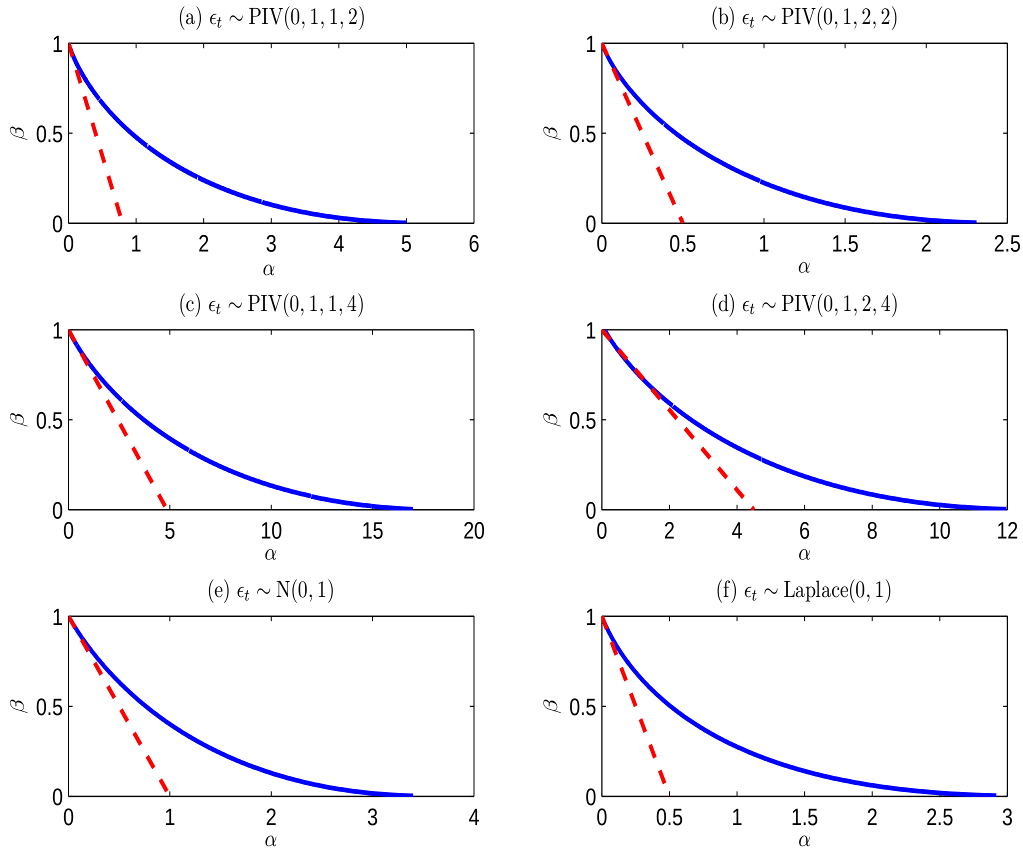

Fig. 4. The regions bounded by the solid and dashed curves are for the strict stationarity (i.e., $E[log(\alpha\varepsilon_t^2 + \beta)] < 0$) and for $Ey_t^2 < \infty$ (i.e., $E\varepsilon_t^2\alpha + \beta < 1$), respectively.

Assumption 5. (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 6. (i) $E[\sup_{\theta\in\Theta} |\log h_t(\theta)|] < \infty$;

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

Assumption 7. (i) $\tilde{h}_t(\theta) \geq \underline{z}$ (a.s.) for some $\underline{z} > 0$ and all $\theta \in \Theta$;

$$\text{(ii)} \sup_{\theta\in\Theta} |\tilde{h}_t(\theta) - h_t(\theta)| \leq O(\rho^t)R_t,$$

$$\text{(iii)} \sup_{\theta\in\Theta} \left\| \frac{1}{\tilde{h}_t(\theta)} \frac{\partial \tilde{h}_t(\theta)}{\partial \theta} - \frac{1}{h_t(\theta)} \frac{\partial h_t(\theta)}{\partial \theta} \right\| \leq O(\rho^t)R_t,$$

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

for some constant $\rho \in (0, 1)$, where R_t is a positive random variable such that $ER_t^{2\kappa_1} = O(t)$ for some $\kappa_1 \in (0, 1)$.

Assumption 5 imposes some basic requirements on the function $h_t(\theta)$, and they are satisfied by most of the conditionally heteroskedastic models; see, e.g., Francq and Zakoïan (2004, 2013). Assumption 6 gives some technical moment conditions for our proofs, and Assumption 7 makes the initial values Y_0 ignorable. Both assumptions 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). Corollary 1 below gives the strong consistency and asymptotic normality the PQMLE for model (11), 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 (*11*). If Assumptions* 1, 2*(i), and* 3-7 *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:

$$y_t = \sigma_t \varepsilon_t \text{ and } \sigma_t^2 = \omega_0 + \alpha_0 y_{t-1}^2 + \beta_0 \sigma_{t-1}^2, \tag{12}$$

where we choose $(\omega_0, \alpha_0, \beta_0) = (0.25, 0.15, 0.3)$ as in Fan, Qi, and Xiu (2014), and ε_t is chosen to be the PIV distributions, the STB distributions, and the Student's t distributions, respectively. Particularly, for the STB distribution, the parameters $(\breve{\alpha}, \breve{\beta}, \breve{\mu})$ are set as those for the British pound and the Japanese Yen in Table III of Liu and Brorsen (1995).

In order to implement the PQMLE for each replication, we choose (ν, m) as in Remark 3, and then the PQMLE $\tilde{\theta}_n^*$ is the estimator of $(\omega_0/c_0^2, \alpha_0/c_0^2, \beta_0)$, where c_0 is the unique solution of the function $u(c) = 1$ on $\{c : c > 0\}$, and $u(c)$ is defined as in Assumption 4. Consequently, we let

$$\tilde{\theta}_n = (c_0^2 \tilde{\omega}_n^*, c_0^2 \tilde{\alpha}_n^*, \tilde{\beta}_n^*)$$

be the PQMLE of $(\omega_0, \alpha_0, \beta_0)$, where c_0 is chosen to be the numerical solution of $\hat{u}_n(c) = 1$ on $\{c : c > 0\}$, and $\hat{u}_n(c)$ is defined in the same way as $\tilde{u}_n(c)$ in (9) with ε_t replacing $\tilde{\varepsilon}_t$. Here, we have assumed that ε_t can be standardized for each replication. Similarly, since the other four estimation methods require different identification conditions for model (12), the GQMLE $(\bar{\theta}_{1n}^*)$, LQMLE $(\bar{\theta}_{2n}^*)$, LADE $(\bar{\theta}_{3n}^*)$, and GNGQMLE $(\bar{\theta}_{4n}^*)$ are estimators of $(c_1\omega_0, c_1\alpha_0, \beta_0)$ with $c_1 = E\varepsilon_t^2, (E|\varepsilon_t|)^2$, median$(\varepsilon_t^2)$ and $E\varepsilon_t^2$ respectively, and then we let

$$\bar{\theta}_{1n} = \left(\frac{\bar{\omega}_{1n}^*}{E\varepsilon_t^2}, \frac{\bar{\alpha}_{1n}^*}{E\varepsilon_t^2}, \bar{\beta}_{1n}^* \right), \quad \bar{\theta}_{2n} = \left(\frac{\bar{\omega}_{2n}^*}{(E|\varepsilon_t|)^2}, \frac{\bar{\alpha}_{2n}^*}{(E|\varepsilon_t|)^2}, \bar{\beta}_{2n}^* \right),$$
$$\bar{\theta}_{3n} = \left(\frac{\bar{\omega}_{3n}^*}{\text{median}(\varepsilon_t^2)}, \frac{\bar{\alpha}_{3n}^*}{\text{median}(\varepsilon_t^2)}, \bar{\beta}_{3n}^* \right), \text{ and } \bar{\theta}_{4n} = \left(\frac{\bar{\omega}_{4n}^*}{E\varepsilon_t^2}, \frac{\bar{\alpha}_{4n}^*}{E\varepsilon_t^2}, \bar{\beta}_{4n}^* \right)$$

be the GQMLE, LQMLE, LADE, and GNGQMLE of $(\omega_0, \alpha_0, \beta_0)$, respectively. The estimated asymptotic standard deviations of all estimators are 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, Qi, and Xiu (2014). Note that the PQMLE 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 $E\varepsilon_t^2 < \infty$.

Table 1 reports the bias and root mean square error (RMSE) of all estimators for model (12). From Table 1, we find that all estimators have very small bias. Thus, we compare the performance of all estimators in terms of the minimized RMSE. When $\varepsilon_t \sim$ PIV$(0, 1, 2, 4)$, the PQMLE has the smallest RMSE. This is not surprising since it is the efficient estimator due to the fact that the selected value of (ν, m) is close to the optimal value $(2, 4)$. When $\eta_t \sim$ PIV$(0.02, 0.9, -2, 4)$, the PQMLE still has the smallest RMSE even when the location parameter λ and scale parameter a deviate from 0 and 1, respectively. For both cases of PIV distributions, the GNGQMLE has the second smallest RMSE, and the GQMLE or LADE has the largest RMSE. Next, when ε_t follows the STB distribution, only the PQMLE and LADE are applicable, and the PQMLE has smaller RMSE than the LADE in both examined cases. Thirdly, when ε_t follows the t distribution, the

Table 1. *The bias and RMSE of all estimators for model* (12)

$\varepsilon_t \sim$ PIV$(0, 1, 2, 4)$

	PQMLE			GQMLE			LQMLE			LADE			GNGQMLE		
	ω	α	β	ω	α	β	ω	α	β	ω	α	β	ω	α	β
Bias	0.0013	-0.0012	-0.0032	-0.0051	-0.0018	0.0139	0.0025	-0.0021	-0.0050	0.0066	0.0233	-0.0216	-0.0029	-0.0022	0.0076
RMSE	**0.1070**	**0.1016**	**0.2875**	0.1152	0.1204	0.3116	0.1132	0.1078	0.3040	0.1161	0.1497	0.3083	0.1089	0.1057	0.2937

$\varepsilon_t \sim$ PIV$(0.02, 0.9, -2, 4)$

	PQMLE			GQMLE			LQMLE			LADE			GNGQMLE		
	ω	α	β	ω	α	β	ω	α	β	ω	α	β	ω	α	β
Bias	-0.0018	0.0069	0.0030	-0.0018	0.0170	0.0003	0.0049	0.0146	-0.0158	0.0104	0.0414	-0.0363	-0.0050	0.0093	0.0101
RMSE	**0.1079**	**0.1177**	**0.2921**	0.1146	0.1368	0.3108	0.1127	0.1258	0.3039	0.1191	0.1713	0.3184	0.1107	0.1238	0.3013

$\varepsilon_t \sim$ STB$(1.77, 0.2, 0.9, -0.02)$

	PQMLE			GQMLE			LQMLE			LADE			GNGQMLE		
	ω	α	β	ω	α	β	ω	α	β	ω	α	β	ω	α	β
Bias	0.0069	-0.0009	-0.0085		N.A.			N.A.		0.0079	0.0001	-0.0101		N.A.	
RMSE	**0.0463**	**0.0306**	**0.0927**		N.A.			N.A.		0.0607	0.0417	0.1255		N.A.	

$\varepsilon_t \sim$ STB$(1.82, -0.15, 1.1, -0.02)$

	PQMLE			GQMLE			LQMLE			LADE			GNGQMLE		
	ω	α	β	ω	α	β	ω	α	β	ω	α	β	ω	α	β
Bias	0.0075	-0.0010	-0.0070		N.A.			N.A.		0.0075	-0.0010	-0.0058		N.A.	
RMSE	**0.0429**	**0.0236**	**0.0745**		N.A.			N.A.		0.0572	0.0334	0.1003		N.A.	

$\varepsilon_t \sim t_4$

	PQMLE			GQMLE			LQMLE			LADE			GNGQMLE		
	ω	α	β	ω	α	β	ω	α	β	ω	α	β	ω	α	β
Bias	0.0063	0.0009	-0.0123		N.A.		0.0078	0.0016	-0.0149	0.0045	0.0014	-0.0061	0.0047	0.0001	-0.0138
RMSE	0.0555	0.0377	0.1211		N.A.		0.0574	0.0393	0.1250	0.0720	0.0490	0.1567	**0.0549**	**0.0374**	**0.1210**

$\varepsilon_t \sim t_3$

	PQMLE			GQMLE			LQMLE			LADE			GNGQMLE		
	ω	α	β	ω	α	β	ω	α	β	ω	α	β	ω	α	β
Bias	0.0061	0.0007	-0.0095		N.A.		0.0074	0.0010	-0.0118	0.0074	0.0033	-0.0134	-0.0004	-0.0032	-0.0105
RMSE	**0.0495**	**0.0337**	**0.0980**		N.A.		0.0550	0.0397	0.1115	0.0625	0.0413	0.1242	0.0506	0.0350	0.1019

† The invalid estimation results are labeled as “Not Available (N.A.)”.

‡ The smallest RMSE for each case is in boldface.

performance of all estimation methods is mixed. Specifically, when $\varepsilon_t \sim t_4$, the GNGQMLE has the smallest RMSE due to its adaptive property under symmetry, but the PQMLE follows closely behind by a negligible margin; while when $\varepsilon_t \sim t_3$, the tail of ε_t becomes heavier, and the PQMLE has the smallest RMSE among all estimators. Overall, the simulation study shows that the PQMLE together with the selection procedure of (ν, m) in Remark 3 has 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 four major stock indexes in the world. The data sets we considered are the daily DJIA, FTSE, HSI, and NASDAQ indexes from January 3, 2000 to December 27, 2007. We denote the log-return ($\times 100$) of each data set by $\{y_t\}_{t=1}^n$, and the summary statistics for each y_t is given in Table 2. From this table, we find that each y_t is skewed and has a heavier tail than the $\mathrm{N}(0,1)$ distribution. Hence, we use a GARCH$(1,1)$ model with the PQMLE estimation method to fit each return series. As a comparison, we also apply the GQMLE, LQMLE, or GNGQMLE estimation method to obtain the fitted GARCH$(1,1)$ model for each return series. For the PQMLE method, ν and m are chosen as in Remark 3. For the GNGQMLE method, the auxiliary likelihood function is based on the standardized t_3, t_5, or t_7 distribution such that it has a variance equal to 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 Table 3, in which the full log-likelihood function of the PQMLE is defined as in (10), and the full log-likelihood functions of the GQMLE (LLF$_G$), LQMLE (LLF$_L$), and GNGQMLE (LLF$_{GNG}$) are defined as follows:

$$\mathrm{LLF}_G = -\sum_{t=1}^{n}\left[\log\sqrt{\tilde{h}_t(\bar{\theta}_{1n})} + \frac{y_t^2}{2\tilde{h}_t(\bar{\theta}_{1n})}\right] + n\log\left(\frac{1}{\sqrt{2\pi}}\right),$$

$$\mathrm{LLF}_L = -\sum_{t=1}^{n}\left[\log\sqrt{\tilde{h}_t(\bar{\theta}_{2n})} + \frac{|y_t|}{\sqrt{\tilde{h}_t(\bar{\theta}_{2n})}}\right] + n\log\left(\frac{1}{2}\right),$$

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

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

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

Here, $\hat{\eta}_k$ measures the discrepancy between the correct 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) tailed than the true innovation. Furthermore, Table 3 also reports the estimated values of the identification condition c_1 for each estimation method, that is, c_1 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 (i) all fitted models are adequate by looking at the ACF and PACF plots (not displayed here) of the squared and absolute residuals; (ii) the plots of $\tilde{u}_n(c)$ in Figure 5 indicate that Assumption 4 holds for the PQMLE in each return series, where $\tilde{u}_n(c)$ is defined as in (9).

Table 2. *Summary statistics of four major stock indexes*

y_t	n	mean	standard deviation	skewness	kurtosis
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

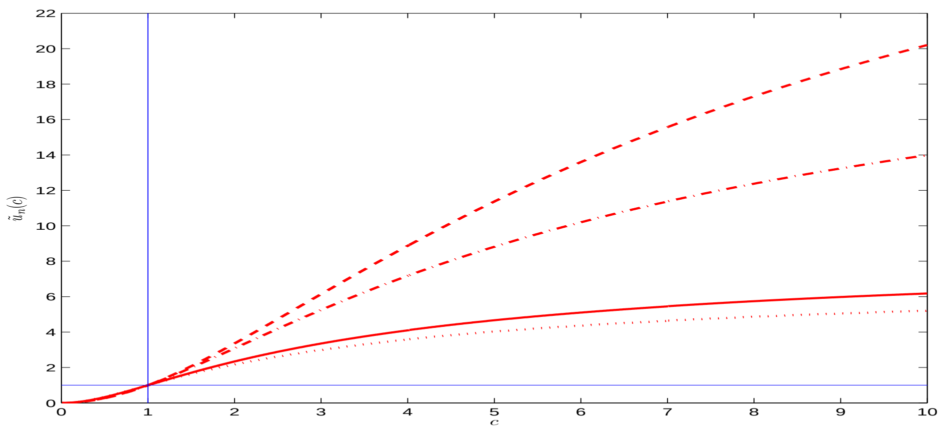

Fig. 5. The plots of $\tilde{u}_n(c)$ for DJIA (solid line), FTSE (dashed line), HSI (dotted line), and NASDAQ (dash-dot line).

From Table 3, we find that (i) all the values of c_1 are close to 1 as expected; (ii) for each return series, the PQMLE always has the best fitting in terms of the maximized LLF among all estimation methods; (iii) the GNGQMLE estimation with a t_5 or t_7 likelihood gives the second best fitted models for the DJIA and HSI return series in which the values of m are smaller, while the GQMLE estimation gives the second best fitted models for the FTSE and NASDAQ return series in which the values of m are larger; this is not surprising as a smaller m implies a heavier tail of the innovation and vice versa; (iv) the LQMLE has the worst fitting for FTSE and NASDAQ return series with larger m (thinner tail), and the GQMLE has the worst fitting for DJIA and HSI return series with smaller m (heavier tail); (v) the GNGQMLE estimation with a t_3 likelihood always has the largest value of $\hat{\eta}_k$ among all GNGQMLE estimations, and hence

Table 3. *Summary of all estimations for four major stock indexes*

y_t		PQMLE	GQMLE	LQMLE	GNGQMLE$_1$	GNGQMLE$_2$	GNGQMLE$_3$
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
	c_1	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
	c_1	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
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
	c_1	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
	c_1	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

† The standard deviations are in parentheses.

‡ The largest LLF for each case is in boldface.

it implies that the auxiliary t_3 innovation is heavier tailed than the true innovation, while the auxiliary t_5 or t_7 innovation has the similar 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 may suggest that the innovation for each return series has finite fourth moment. Overall, we know that all estimation methods are applicable, and the PQMLE estimation method taking into account both asymmetry and leptokurtosis of the innovation gives the best fitted models for all return series in terms of the maximized LLF.

Next, we use the conditional coverage test LR_{CC} in Christoffersen (1998, page 847) to examine whether each of the estimation methods can provide us with a good interval forecast for its one-step-ahead prediction. For each return series, the out-of-sample data set we used is a realization of n_0 consecutive observations starting after the last observation of the in-sample data set. Following Christoffersen (1998), the upper-tail predictive interval (UPI) and lower-tail predictive interval (LPI) for each out-of-sample data y_t at the significance level $\bar{p}$ are defined as

$$\text{UPI}_{t|t-1}(\bar{p}) = \left(F^{-1}(1-\bar{p})\bar{\sigma}_t, \infty\right) \text{ and } \text{LPI}_{t|t-1}(\bar{p}) = \left(-\infty, F^{-1}(\bar{p})\bar{\sigma}_t\right),$$

respectively, where $\bar{\sigma}_t$ is the one-step-ahead prediction of σ_t from each estimation method, and $F(\cdot)$ is the cdf of the PIV$(0,1,\nu,m)$, N$(0,1)$, Laplace$(0,1)$, or standardized t_i (for $i=3,5,7$) distribution for the PQMLE, GQMLE, LQMLE, or GNGQMLE_i estimation method, respectively. Table 4 reports all the results of LR_{CC} with $\bar{p}=0.95$, which examine whether the UPI or LPI from each estimation method gives us a good conditional coverage rate (CR). From Table 4, we find that (i) the p-value of LR_{CC} based on the LQMLE or GNGQMLE_1 method is always close to zero, and hence the CR constructed from these two methods is not satisfactory; (ii) for the DJIA or HSI return series, the CR based on the PQMLE or GNGQMLE_3 method is satisfactory in both directions, while the LPI based on the GQMLE method for the DJIA or HSI return series and the UPI based on the GNGQMLE_2 method for the DJIA return series are not satisfactory; (iii) the PQMLE and GQMLE methods indicate that only the LPI is satisfactory for the FTSE return series, and this can not be indicated by all of the GNGQMLE methods; actually, UPI of PQMLE or GQMLE is much closer to the nominal one than the others; (iv) all PQMLE, GQMLE, GNGQMLE_2, and GNGQMLE_3 methods indicate that only the LPI is satisfactory for the NASDAQ return series. Overall, the CRs of PQMLE are closest to the nominal 0.95 in 5 out

Table 4. *The results of* LR_{CC} *and out-of-sample CR with* $\bar{p} = 0.95$ *for four major stock indexes.*

y_t	n_0		PQMLE	GQMLE	LQMLE	GNGQMLE_1	GNGQMLE_2	GNGQMLE_3
DJIA	1487	UPI	5.9593	4.3344	30.276	33.411	8.5153	5.9593
			(**0.0508**)	(**0.1145**)	(0.0000)	(0.0000)	(0.0142)	(**0.0508**)
			[0.9381]	[0.9401]	[0.9771]	[0.9153]	[0.9354]	[0.9381]
		LPI	2.9401	7.2116	61.514	34.740	3.4362	2.7723
			(**0.2299**)	(0.0272)	(0.0000)	(0.0000)	(**0.1794**)	(**0.2500**)
			[0.9509]	[0.9549]	[0.9872]	[0.9159]	[0.9489]	[0.9523]
FTSE	1493	UPI	8.3814	8.3814	44.653	55.661	17.443	12.717
			(0.0151)	(0.0151)	(0.0000)	(0.0000)	(0.0002)	(0.0017)
			[0.9330]	[0.9330]	[0.9826]	[0.9029]	[0.9257]	[0.9297]
		LPI	5.1764	5.1764	73.956	38.891	13.370	8.5952
			(**0.0752**)	(**0.0752**)	(0.0000)	(0.0000)	(0.0012)	(0.0136)
			[0.9451]	[0.9451]	[0.9900]	[0.9149]	[0.9357]	[0.9404]
HSI	1490	UPI	0.1211	3.1785	38.049	20.341	1.6155	0.1405
			(**0.9412**)	(**0.2041**)	(0.0000)	(0.0000)	(**0.4459**)	(**0.9322**)
			[0.9497]	[0.9443]	[0.9805]	[0.9228]	[0.9430]	[0.9490]
		LPI	1.7182	7.6223	56.443	13.132	0.9994	1.1968
			(**0.4235**)	(0.0221)	(0.0000)	(0.0014)	(**0.6067**)	(**0.5497**)
			[0.9564]	[0.9577]	[0.9859]	[0.9302]	[0.9517]	[0.9544]
NASDAQ	1489	UPI	11.414	11.931	35.262	39.303	18.612	14.280
			(0.0033)	(0.0026)	(0.0000)	(0.0000)	(0.0001)	(0.0008)
			[0.9436]	[0.9429]	[0.9792]	[0.9174]	[0.9362]	[0.9402]
		LPI	4.3780	4.3780	84.023	36.208	3.6488	2.9362
			(**0.1120**)	(**0.1120**)	(0.0000)	(0.0000)	(**0.1613**)	(**0.2304**)
			[0.9597]	[0.9597]	[0.9919]	[0.9174]	[0.9483]	[0.9510]

† The p-values of LR_{CC} are in open brackets, and the values of CR are in square brackets.

‡ The p-values of LR_{CC} larger than $(1 - \bar{p})$ are in boldface.

of 8 cases (including ties), while GQMLE is 3 out of 8, and GNGQMLE_2 or GNGQMLE_3 is 1 out of 8. Hence, the PQMLE method is applicable in giving us a good prediction interval for each side with respect to a wide range of tail thickness of the innovation, but without relying on the selection of the auxiliary likelihood function.

4·2. *Application to exchange rates*

In this subsection, we apply the PQMLE estimation method to HKD/USD and TWD/USD exchange rates. For each exchange rate series, the period of the data we considered is listed in the second column of Table 5. 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 LADE method in Zhu and Ling (2014). Consequently, we denote the residuals from each fitted ARMA$(2, 2)$ model by y_t. Table 5 gives the summary statistics for each y_t, from which we find that each y_t is skewed and has a heavier tail than the $\mathrm{N}(0, 1)$ 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 the estimation results are summarized in Table 6, and the plots of $\tilde{u}_n(c)$ in Figure 6 imply that Assumption 4 holds for the PQMLE in each return series.

Table 5. *Summary statistics of two exchange rates*

y_t	Time Period	n	mean	standard deviation	skewness	kurtosis
HKD/USD	Jan 24, 1996–Jan 08, 2004	2000	0.0000	0.0268	-4.3767	98.122
TWD/USD	Jan 19, 1996–Jan 10, 2000	1000	0.0116	0.4504	1.4054	28.731

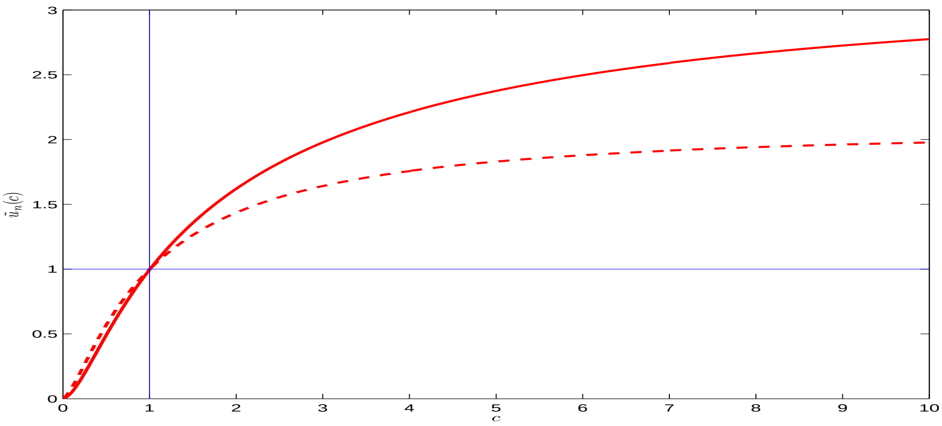

Fig. 6. The plots of $\tilde{u}_n(c)$ for HKD/USD (solid line) and TWD/USD (dashed line).

From Table 6, we find that (i) the GQMLE has the worst performance for both return series in terms of maximized LLF, and this may be because the GQMLE is not applicable when the values of m are smaller than 2.5, which implies a very heavy tail of the innovation; (ii) for HKD/USD return series, the values of c_1 deviate significantly from one in all GNGQMLE methods, meaning that all GNGQMLEs are not well identified in this case; (iii) the TWD/USD return series has a very heavy tail because the value of m is smaller than 1.5, from which we may conclude that the innovation has infinite variance, and hence only the PQMLE method is valid; (iv) for both cases, the PQMLE has the best fit in terms of the maximized LLF among all estimation methods, and this advantage of PQMLE over LQMLE or GNGQMLE may be explained by its ability to take into account both asymmetric and leptokurtic effects.

Table 6. *Summary of all estimations for two exchange rates*

y_t		PQMLE	GQMLE	LQMLE	GNGQMLE_1	GNGQMLE_2	GNGQMLE_3
HKD/USD	ω	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
		(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
	α	0.4529	0.2464	0.2718	1.2262	1.0586	0.9734
		(0.0431)	(0.2333)	(0.0544)	(0.0000)	(0.0000)	(0.0000)
	β	0.6526	0.7837	0.6599	0.6795	0.6893	0.6975
		(0.0190)	(0.1265)	(0.0398)	(0.1108)	(0.1126)	(0.1125)
	ν	-0.0248					
	m	1.6612					
	$\hat{\eta}_k$				0.6952	0.6004	0.5977
	c_1	0.9956	0.9923	0.9956	1.3948	1.3786	1.3966
	LLF	**6525.1**	5356.8	6389.7	6518.7	6469.4	6421.8
TWD/USD	ω	0.0000	0.0001	0.0017	0.0002	0.0007	0.0008
		(0.0000)	(0.0005)	(0.0001)	(0.0001)	(0.0003)	(0.0004)
	α	0.4274	1.0917	0.2539	1.1664	1.0186	0.9788
		(0.0615)	(0.6140)	(0.0636)	(0.0003)	(0.0007)	(0.0008)
	β	0.5154	0.6908	0.6843	0.6797	0.6931	0.6958
		(0.0302)	(0.1067)	(0.0482)	(0.0736)	(0.0821)	(0.0848)
	ν	-0.0223					
	m	1.4076					
	$\hat{\eta}_k$				0.6887	0.6003	0.5996
	c_1	0.9987	0.9975	0.9972	1.0011	0.9949	1.0044
	LLF	**294.9**	-219.2	247.2	265.2	225.6	193.8

† The standard deviations are in parentheses.

‡ The largest LLF for each case is in boldface.

Next, as in Subsection 4.1, we use the conditional coverage test LR_{CC} to examine whether each of the estimation methods can provide us with a good interval forecast for its one-step-ahead prediction. Table 7 reports all the results of LR_{CC} and CR with $\bar{p} = 0.95$. From this table, we find that for the HKD/USD and TWD/USD return series, only the PQMLE method gives us a satisfactory CR in both directions, although the GQMLE can also give us a satisfactory CR for UPI. Particularly, the CRs of the PQMLE are always within one percent from the 95% value for both return series, while this is not the case in other methods. Overall, compared with other methods, the performance of PI constructed from the PQMLE method is often satisfactory, and it is not affected by the selection of the auxiliary likelihood function. This advantage of PQMLE becomes more significant when the return series has a smaller value of m.

Table 7. *The results of LR_{CC} and out-of-sample CR with $\bar{p} = 0.95$ for two exchange rates series.*

y_t	n_0		PQMLE	GQMLE	LQMLE	GNGQMLE_1	GNGQMLE_2	GNGQMLE_3
HKD/USD	2000	UPI	0.9936	3.0993	20.664	33.411	34.606	34.606
			(**0.6085**)	(**0.2123**)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
			[0.9525]	[0.9575]	[0.9685]	[0.9685]	[0.9750]	[0.9750]
		LPI	3.9696	6.4482	55.627	49.185	63.514	65.891
			(**0.1374**)	(0.0398)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
			[0.9520]	[0.9585]	[0.9800]	[0.9785]	[0.9835]	[0.9840]
TWD/USD	1000	UPI	1.1373	23.754	15.057	10.989	21.803	23.754
			(**0.5663**)	(0.0000)	(0.0000)	(0.0041)	(0.0000)	(0.0000)
			[0.9550]	[0.9780]	[0.9730]	[0.9700]	[0.9770]	[0.9780]
		LPI	1.0623	21.660	13.516	14.773	19.437	23.386
			(**0.5879**)	(0.0000)	(0.0012)	(0.0006)	(0.0001)	(0.0000)
			[0.9450]	[0.9770]	[0.9720]	[0.9690]	[0.9770]	[0.9790]

† The p-values of LR_{CC} are in open brackets, and the values of CR are in square brackets.

‡ The p-values of LR_{CC} larger than $(1 - \bar{p})$ are in boldface.

5. CONCLUDING REMARKS

In this paper, we propose a PQMLE for GARCH models. Under strict stationarity and some weak moment conditions, the strong consistency and asymptotic normality of the PQMLE are obtained. Meanwhile, the PQMLE can be applied to other conditionally heteroskedastic models with no further efforts. Simulation study demonstrates that the PQMLE can achieve better efficiency than other estimators, especially when the innovation is heavy-tailed and skewed. Two applications to stock indexes and exchange rates further highlight the importance of the PQMLE method with respect to in-sample fit and out-of-sample predictions.

As is well known, the asymmetry and leptokurtosis of the innovation are two often observed co-existing features in financial and economic data sets. How to capture these two features of the innovation has attracted considerable interest in the literature. The old way to do this is using the MLE method by pre-assuming an asymmetric and leptokurtic distribution of the innovation. A plausible and well-known example is the GARCH-stable model in Liu and Brorsen (1995), who fitted the innovation of GARCH model by a stable distribution $\text{STB}(\breve{\alpha}, \breve{\beta}, \breve{c}, \breve{\mu})$. Specifically, they fitted y_t by the following model:

$$y_t = \sigma_t \varepsilon_t, \tag{13}$$

$$\sigma_t^\delta = \omega + \alpha |y_{t-1}|^\delta + \beta \sigma_{t-1}^\delta, \tag{14}$$

where $\omega > 0$, $\alpha \geq 0$, $\beta \geq 0$, $\delta > 0$, and $\varepsilon_t \sim \text{STB}(\breve{\alpha}, \breve{\beta}, 1, \breve{\mu})$ is a sequence of i.i.d. random variables. Although models (13)-(14) can well handle the heavy-tailed and skewed innovation, it has three disadvantages. First, according to Lemma 2.1 in Zhang and Ling (2014), the tail index of y_t in models (13)-(14), denoted by α^*, is the solution of the following equation:

$$E\left[\alpha|\varepsilon_t|^\delta + \beta\right]^{\alpha^*/\delta} = 1 \text{ for } \alpha^* > 0.$$

Since $E|\varepsilon_t|^{\breve{\alpha}} = \infty$, it follows that $\alpha^* < \breve{\alpha}$. Hence, models (13)-(14) only apply for the case that $E|y_t|^{\breve{\alpha}} = \infty$, while our PQMLE method is suitable as long as y_t is strictly stationary. See also Rachev and Mittnik (2000, p.284). Second, the GARCH-stable method used in (13)-(14) has a large chance in getting inconsistent estimates with incorrect standard errors for GARCH models if the true distribution of ε_t is not $\text{STB}(\breve{\alpha}, \breve{\beta}, 1, \breve{\mu})$; see, e.g., White (1982) and Newey and Steigerwald (1997) for the discussions of MLE when the true distribution of ε_t is misspecified. This is the disadvantage for all of the MLE methods, but not for our PQMLE method. Third, the GARCH-stable method is computationally expensive in getting estimates and standard deviations (e.g., the objective function and its gradients involve complex integrals when the characteristic exponent $\breve{\alpha} > 1$, and all of them even do not have a closed form when $\breve{\alpha} \leq 1$; see Liu and Brorsen (1995, pages 275-277)), but this is not the case for our PQMLE method, which can be calculated as fast as the GQMLE. Overall, although the PQMLE method is not the first try in the literature to take into account both asymmetry and leptokurtosis of the innovation, it is the first try for this problem in the context of QMLE, and compared to the existing MLE methods, it has desirable properties to be used in practice.

ACKNOWLEDGEMENT

The authors greatly appreciate the very constructive suggestions and comments of the Editor and two anonymous referees that greatly improve this article. This work is supported by Research Grants Council of the Hong Kong SAR Government, GRF grant HKU703711P, National Natural Science Foundation of China (No.11201459), Key Lab of Random Complex Structures and Data Science of Chinese Academy of Sciences, and National Center for Mathematics and Interdisciplinary Sciences of Chinese Academy of Sciences.

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 some simple algebra, we can show that

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

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

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

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

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

Fourth, it is straightforward to see that $w_1(s) := \partial w(s)/\partial s = E[g_1(\varepsilon_t, s)]$, and then by Assumption 4, we know that $w_1(s)=0$ has a unique solution at $s=1$. Meanwhile, it is not hard to see that $\lim_{s\to 0} w(s) = -\infty$ and $\lim_{s\to\infty} w(s) = -\infty$. By the continuity of $w_1(s)$, it follows that for $s>0$, $w(s)$ is a strictly concave function with a unique maximum at $s=1$.

Therefore, under Assumptions 1-4, 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

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

[2] ANDREWS, B., CALDER, M. & DAVIS, R.A. (2009). Maximum likelihood estimation for α-stable autoregressive processes. *Annals of Statistics* **37**, 1946–1982.

[3] 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.

[4] BAUWENS, L. & LAURENT, S. (2005). A new class of multivariate skew densities, with application to generalized autoregressive conditional heteroscedasticity models. *Journal of Business and Economic Statistics* **23**, 346–354.

[5] 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.

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

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

[8] BHATTACHARYYA, M., CHAUDHARY, A. & YADAV, G. (2008). Conditional VaR estimation using Pearson's type IV distribution. *European Journal of Operational Research* **191**, 386–397.

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

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

[11] 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.

[12] 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.

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

[14] CHEN, M. & ZHU, K. (2014). Sign-based portmanteau test for ARCH-type models with heavy-tailed innovations. Forthcoming in *Journal of Econometrics*.

[15] CHENG, X., LI, W.K., YU, P.L.H., ZHOU, X., WANG, C. & LO, P.H. (2011). Modeling threshold conditional heteroscedasticity with regime-dependent skewness and kurtosis. *Computational Statistics and Data Analysis* **55**, 2590-2604.

[16] CHRISTOFFERSEN, P. (1998). Evaluating interval forecasts. *International Economic Review* **39**, 841-862.

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

[18] 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.

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

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

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

[22] FAN, J., QI, L. & XIU, D. (2014). Quasi maximum likelihood estimation of GARCH models with heavy-tailed likelihoods (with discussion). *Journal of Business and Economic Statistics* **32**, 178-205.

[23] FRANCQ, C., WINTENBERGER, O. & ZAKOÏAN, J.M. (2013). Garch models without positivity constraints: exponential or log garch? *Journal of Econometrics* **177**, 34-46.

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

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

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

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

[28] GRIGOLETTO, M. & LISI, F. (2009). Looking for skewness in financial time series. *Econometrics Journal* **12**, 310-323.

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

[30] 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.

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

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

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

[34] LEON, A., RUBIO, G. & SERNA, G. (2005). Autoregressive conditional volatility, skewness and kurtosis. *The Quarterly Review of Economics and Finance* **45**, 599-618.

[35] LI, G. & LI, W.K. (2008). Least absolute deviation estimation for fractionally integrated autoregressive moving average time series models with conditional heteroscedasticity. *Biometrika* **95**, 399-414.

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

[37] LIU, S.-M. & BRORSEN, B.W. (1995). Maximum likelihood estimation of a GARCH-stable model. *Journal of Applied Econometrics* **10**, 273–285.

[38] 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.

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

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

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

[42] NOLAN, J.P. (1997). Numerical calculation of stable densities and distribution functions. *Communications in Statistics, Stochastic Models* **13**, 759–774.

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

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

[45] RACHEV, S. & MITTNIK, S. (2000). Stable Paretian models in finance. Chichester: Wiley.

[46] TSAY, R.S. (2005). Analysis of financial time series (2nd ed.). New York: John Wiley & Sons, Incorporated.

[47] VERHOEVEN, P. & MCALEER, M. (2004). Fat tails and asymmetry in financial volatility models. *Mathematics and Computers in Simulation* **64**, 351-361.

[48] WHITE, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrica* **50**, 1-25.

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

[50] ZHANG, R.M. & LING, S. (2014). Asymptotic inference for AR models with heavy-tailed G-GARCH noises. Forthcoming in *Econometric Theory*.

[51] 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.

[52] ZHU, K. & LING, S. (2014). LADE-based inference for ARMA models with unspecified and heavy-tailed heteroscedastic noises. Forthcoming in *Journal of the American Statistical Association*.