ROYAL STATISTICAL SOCIETY
DATA | EVIDENCE | DECISIONS

Journal of the Royal Statistical Society
Statistical Methodology
Series B

J. R. Statist. Soc. B (2018)
80, *Part* 5, *pp.* 975–993

Hybrid quantile regression estimation for time series models with conditional heteroscedasticity

Yao Zheng,

University of Hong Kong, People's Republic of China

Qianqian Zhu,

Shanghai University of Finance and Economics, People's Republic of China

Guodong Li

University of Hong Kong, People's Republic of China

and Zhijie Xiao

Boston College, Chestnut Hill, USA

[Received December 2016. Final revision April 2018]

Summary. Estimating conditional quantiles of financial time series is essential for risk management and many other financial applications. For time series models with conditional heteroscedasticity, although it is the generalized auto-regressive conditional heteroscedastic (GARCH) model that has the greatest popularity, quantile regression for this model usually gives rise to non-smooth non-convex optimization which may hinder its practical feasibility. The paper proposes an easy-to-implement hybrid quantile regression estimation procedure for the GARCH model, where we overcome the intractability due to the square-root form of the conditional quantile function by a simple transformation. The method takes advantage of the efficiency of the GARCH model in modelling the volatility globally as well as the flexibility of quantile regression in fitting quantiles at a specific level. The asymptotic distribution of the estimator is derived and is approximated by a novel mixed bootstrapping procedure. A portmanteau test is further constructed to check the adequacy of fitted conditional quantiles. The finite sample performance of the method is examined by simulation studies, and its advantages over existing methods are illustrated by an empirical application to value-at-risk forecasting.

Keywords: Bootstrap method; Conditional quantile; Generalized auto-regressive conditional heteroscedasticity; Non-linear time series; Quantile regression

1. Introduction

Time series models with conditional heteroscedasticity have been known to be greatly successful at capturing the volatility clustering of financial data since the appearance of Engle's (1982) auto-regressive conditional heteroscedastic (ARCH) model and Bollerslev's (1986) generalized auto-regressive conditional heteroscedastic (GARCH) model; see Francq and Zakoian (2010). One of many popular applications of these models is to estimate quantile-based risk measures such as the value at risk, VaR, and the expected shortfall, and, for such problems, quantile regression

Address for correspondence: Yao Zheng, Department of Statistics and Actuarial Science, University of Hong Kong, Pokfulam Road, Hong Kong, People's Republic of China.
E-mail: yaozheng@connect.hku.hk

 1369–7412/18/80975

(Koenker and Bassett, 1978) naturally makes an appealing tool (Engle and Manganelli, 2004; Francq and Zakoïan, 2015).

In the literature, feasible quantile regression has remained challenging for the arguably most important conditional heteroscedastic time series model: Bollerslev's (1986) GARCH model

$$x_t = \eta_t \sqrt{h_t}, \qquad h_t = \alpha_0 + \sum_{i=1}^{q} \alpha_i x_{t-i}^2 + \sum_{j=1}^{p} \beta_j h_{t-j}, \tag{1.1}$$

where $\{\eta_t\}$ are independent and identically distributed (IID) innovations with mean 0 and variance 1. Denote the τth quantile of η_t by $Q_{\tau,\eta}$ and the information set that is available at time t by $\mathcal{F}_t$. In estimating the conditional quantile of x_t in model (1.1), i.e.

$$Q_\tau(x_t|\mathcal{F}_{t-1}) = Q_{\tau,\eta} \sqrt{\left(\alpha_0 + \sum_{i=1}^{q} \alpha_i x_{t-i}^2 + \sum_{j=1}^{p} \beta_j h_{t-j} \right)}, \qquad 0 < \tau < 1, \tag{1.2}$$

there are two key challenges that make quantile regression highly intractable.

(a) The *square root* in equation (1.2), along with the check function $\rho_\tau(x) = x\{\tau - I(x < 0)\}$, leads to a non-smooth objective function which is non-convex even for the ARCH case.
(b) The *recursive* form of the unobservable $\{h_t\}$ in model (1.1) adds another layer of difficulty to the already complicated theoretical derivation and numerical optimization.

Before introducing our approach to addressing these challenges, consider the following variant of model (1.1), i.e. Taylor's (1986) linear GARCH model

$$y_t = \sigma_t \varepsilon_t, \qquad \sigma_t = \alpha_0 + \sum_{i=1}^{q} \alpha_i |y_{t-i}| + \sum_{j=1}^{p} \beta_j \sigma_{t-j}, \tag{1.3}$$

where $\{\varepsilon_t\}$ are IID innovations with mean 0. Denote the τth quantile of ε_t by $Q_{\tau,\varepsilon}$. Note that challenge (a) is never an issue for model (1.3), as

$$Q_\tau(y_t|\mathcal{F}_{t-1}) = \left(\alpha_0 + \sum_{i=1}^{q} \alpha_i |y_{t-i}| + \sum_{j=1}^{p} \beta_j \sigma_{t-j} \right) Q_{\tau,\varepsilon}, \qquad 0 < \tau < 1.$$

If there were no σ_{t-j} in model (1.3), the problem would be just a linear quantile regression, which was considered in Koenker and Zhao (1996). For the general case, Xiao and Koenker (2009) proposed to replace the σ_{t-j}s with some initial estimates obtained by quantile regression for sieved ARCH models and thereby circumvented challenge (b). Unfortunately, because of challenge (a), easy-to-implement quantile regression procedures for Bollerslev's (1986) original GARCH model (1.1) have been seemingly impossible.

In this paper, we tackle this open problem by applying a simple transformation to the conditional quantile in equation (1.2). With the square root in equation (1.2) in mind, we naturally look for a transformation $T(\cdot)$ which is

(a) the inverse of the square-root function *in some sense* and
(b) a continuous and non-decreasing function from $\mathbb{R}$ to $\mathbb{R}$.

This interestingly leads to $T(x) = x^2 \operatorname{sgn}(x)$, where $\operatorname{sgn}(\cdot)$ is the sign function and, then,

$$T\{Q_\tau(x_t|\mathcal{F}_{t-1})\} = Q_\tau\{T(x_t)|\mathcal{F}_{t-1}\} = \left(\alpha_0 + \sum_{i=1}^{q} \alpha_i x_{t-i}^2 + \sum_{j=1}^{p} \beta_j h_{t-j} \right) T(Q_{\tau,\eta}). \tag{1.4}$$

The linearity of equation (1.4) enables a convenient hybrid three-step estimation procedure as follows.

Step 1: obtain initial estimates of $\{h_t\}$ by fitting the GARCH model (1.1) with the Gaussian quasi-maximum-likelihood method.
Step 2: estimate $Q_\tau\{T(x_t)|\mathcal{F}_{t-1}\}$ by a weighted linear quantile regression.
Step 3: use the relationship $Q_\tau(x_t|\mathcal{F}_{t-1}) = T^{-1}[Q_\tau\{T(x_t)|\mathcal{F}_{t-1}\}]$ to estimate $Q_\tau(x_t|\mathcal{F}_{t-1})$, where $T^{-1}(x) = \sqrt{|x|}\ \mathrm{sgn}(x)$ is the inverse function of $T(\cdot)$.

The hybrid procedure proposed contains two main estimation steps with different purposes. As a preliminary estimation of the global model structure, step 1 exploits the general suitability of the GARCH model in volatility modelling. Subsequently, the quantile regression in step 2 targets a particular quantile level of interest and allows a more flexible characterization of the conditional quantile structure while inheriting the GARCH modelling strategy. In the literature, there are conditional quantile estimation methods that essentially utilize only step 1 or step 2, and the leading examples are the filtered historical simulation (FHS) method (Kuester *et al.*, 2006) and the conditional auto-regressive VaR-method called 'CAViaR' (Engle and Manganelli, 2004). Roughly speaking, the FHS method uses the GARCH structure only for global estimation of the volatility, but not for quantile estimation. In contrast, CAViaR focuses on the local approximation at a particular quantile level, and it adopts the GARCH-type structure only for quantile estimation. The current paper tries to exploit the GARCH structure in both the global estimation of the volatility and the local estimation of quantiles, and the hybrid method proposed can have superior performance in practice, since the actual 'truth' usually lies somewhere in between the global model and the quantile model. More specifically, as the FHS method is reliant solely on GARCH modelling, it is less robust than the proposed method when the quantile structure actually varies in shape across the quantile levels, which is a feature that is frequently encountered in practice (Engle and Manganelli, 2004). Although CAViaR imposes the structure at only a particular quantile level and offers full flexibility, it can lack efficiency at commonly used quantile levels, e.g. $\tau = 0.05$ and $\tau = 0.01$, where the data are very sparse. Moreover, the computation of the CAViaR method is generally challenging. The hybrid method proposed combines the advantages of both approaches and is supposed to be more potent in practice.

As the estimation of the asymptotic covariance matrix of the estimator is complicated by the innovation density function that is involved, a bootstrapping procedure is needed. A straightforward approach is to adopt the random-weighting bootstrap method in Jin *et al.* (2001) in both steps 1 and 2, where the minimands of the corresponding objective functions are perturbed by random weights. By replacing the first perturbation with sample averaging, we alternatively propose a novel mixed method to avoid repeating the optimization in step 1 many times. As a result, the computation time is reduced significantly. Furthermore, we construct a portmanteau test to check the adequacy of fitted conditional quantiles based on the residual quantile auto-correlation function (QACF) in Li *et al.* (2015).

The rest of the paper is organized as follows. Sections 2 and 3 propose the hybrid estimation and mixed bootstrapping procedures, and Section 4 proposes the portmanteau test. Section 5 presents the simulation experiments, and Section 6 provides an empirical analysis on VaR-forecasting. Section 7 concludes with a short discussion. Appendix A presents proof sketches of the theorems and, for brevity, the detailed proofs are provided in the on-line supplementary material. Throughout the paper, '$\rightarrow_\mathrm{d}$' denotes convergence in distribution, $o_p(1)$ denotes a sequence of random variables converging to 0 in probability and $o_p^*(1)$ corresponds to the bootstrap probability space.

The data that are analysed in the paper and the programs that were used to analyse them can be obtained from

http://wileyonlinelibrary.com/journal/rss-datasets

2. Hybrid conditional quantile estimation

2.1. The proposed hybrid estimation procedure

Let $\{x_t\}$ be a strictly stationary and ergodic process generated by model (1.1) with parameter vector $\theta=(\alpha_0,\alpha_1,\ldots,\alpha_q,\beta_1,\ldots,\beta_p)'$, where $\alpha_0>0$, $\alpha_i\geqslant 0$ for $1\leqslant i\leqslant q$ and $\beta_j\geqslant 0$ for $1\leqslant j\leqslant p$; see Bollerslev (1986). The necessary and sufficient condition for the existence of a unique strictly stationary and ergodic solution to the model is given in Bougerol and Picard (1992). Let $\mathcal{F}_t$ be the σ-field that is generated by $\{x_t,x_{t-1},\ldots\}$, and let $b_\tau=T(Q_{\tau,\eta})$ and $\theta_\tau=b_\tau\theta$, where $Q_{\tau,\eta}$ is the τth quantile of η_t and $T(x)=x^2\,\mathrm{sgn}(x)$. Then, the τth quantile of the transformed variable $y_t=T(x_t)$ conditional on $\mathcal{F}_{t-1}$ is

$$Q_\tau(y_t|\mathcal{F}_{t-1})=b_\tau\left(\alpha_0+\sum_{i=1}^{q}\alpha_i x_{t-i}^2+\sum_{j=1}^{p}\beta_j h_{t-j}\right)=\theta_\tau' z_t,\qquad 0<\tau<1,\tag{2.1}$$

where $z_t=(1,x_{t-1}^2,\ldots,x_{t-q}^2,h_{t-1},\ldots,h_{t-p})'$. Note that, if $\{h_t\}$ were observable, then we would be able to estimate $Q_\tau(y_t|\mathcal{F}_{t-1})$ by linear quantile regression.

For $0<\underline{w}<\overline{w}$ and $0<\rho_0<1$ with $p\underline{w}<\rho_0$, define $\Theta=\{\theta:\beta_1+\ldots+\beta_p\leqslant\rho_0,\ \underline{w}\leqslant\min(\alpha_0,\alpha_1,\ldots,\alpha_q,\beta_1,\ldots,\beta_p)\leqslant\max(\alpha_0,\alpha_1,\ldots,\alpha_q,\beta_1,\ldots,\beta_p)\leqslant\overline{w}\}\subset\mathbb{R}_+^{p+q+1}$, where $\mathbb{R}_+=(0,\infty)$; see Berkes and Horváth (2004). Let the true value of θ be $\theta_0=(\alpha_{00},\alpha_{01},\ldots,\alpha_{0q},\beta_{01},\ldots,\beta_{0p})'$, and let $\theta_{\tau 0}=b_\tau\theta_0$. Define $h_t(\theta)$ recursively by

$$h_t(\theta)=\alpha_0+\sum_{i=1}^{q}\alpha_i x_{t-i}^2+\sum_{j=1}^{p}\beta_j h_{t-j}(\theta).\tag{2.2}$$

Then $h_t(\theta_0)=h_t$. As $h_t(\theta)$ in equation (2.2) depends on infinite past observations, initial values for $\{x_0^2,\ldots,x_{1-q}^2,h_0,\ldots,h_{1-p}\}$ are needed. We set them to $m^{-1}\Sigma_{t=1}^m x_t^2$ for a fixed number m, say $m=5$ in our numerical studies, and denote the resulting $h_t(\theta)$ by $\tilde{h}_t(\theta)$; fixing the initial values will not affect our asymptotic results.

We propose the hybrid conditional quantile estimation procedure as follows.

Step 1 (estimation of the global model structure): perform the Gaussian quasi-maximum likelihood estimation of model (1.1),

$$\tilde{\theta}_n=\arg\min_{\theta\in\Theta}\sum_{t=1}^{n}\tilde{l}_t(\theta),\tag{2.3}$$

where $\tilde{l}_t(\theta)=x_t^2/\tilde{h}_t(\theta)+\log\{\tilde{h}_t(\theta)\}$; see Francq and Zakoian (2004). Then compute the initial estimates of $\{h_t\}$ as $\tilde{h}_t=\tilde{h}_t(\tilde{\theta}_n)$.

Step 2 (quantile regression at a specific level): perform the weighted linear quantile regression of y_t on $\tilde{z}_t=(1,x_{t-1}^2,\ldots,x_{t-q}^2,\tilde{h}_{t-1},\ldots,\tilde{h}_{t-p})'$ at quantile level τ:

$$\hat{\theta}_{\tau n}=\arg\min_{\theta_\tau}\sum_{t=1}^{n}\tilde{h}_t^{-1}\rho_\tau(y_t-\theta_\tau'\tilde{z}_t).\tag{2.4}$$

Thus the τth conditional quantile of y_t can be estimated by $\hat{Q}_\tau(y_t|\mathcal{F}_{t-1})=\hat{\theta}_{\tau n}'\tilde{z}_t$.

Step 3 (transforming back to x_t): estimate the τth conditional quantile of the original observation x_t by $\hat{Q}_\tau(x_t|\mathcal{F}_{t-1})=T^{-1}(\hat{\theta}_{\tau n}'\tilde{z}_t)$, where $T^{-1}(x)=\sqrt{|x|}\,\mathrm{sgn}(x)$.

Assumption 1.

(a) θ_0 is in the interior of Θ;
(b) η_t^2 has a non-degenerate distribution with $E[\eta_t^2]=1$;
(c) the polynomials $\Sigma_{i=1}^q \alpha_i x^i$ and $1-\Sigma_{j=1}^p \beta_j x^j$ have no common root;
(d) $E[\eta_t^4]<\infty$.

Assumption 2. The density $f(\cdot)$ of $\varepsilon_t = T(\eta_t)$ is positive and differentiable almost everywhere on $\mathbb{R}$, with its derivative $\dot{f}$ satisfying that $\sup_{x\in\mathbb{R}} |\dot{f}(x)|<\infty$.

Assumption 1 was used by Francq and Zakoian (2004) to ensure the consistency and asymptotic normality of the Gaussian quasi-maximum-likelihood estimator (QMLE) $\tilde{\theta}_n$, which is known as the sharpest result. It implies only a finite fractional moment of x_t, i.e. $E|x_t|^{2\delta_0}<\infty$ for some $\delta_0>0$ (Berkes *et al.*, 2003; Francq and Zakoian, 2004). For the GARCH model, imposing a higher order moment condition on x_t would reduce the available parameter space Θ; see Francq and Zakoian (2010), chapter 2.4.1. Assumption 2 is made for brevity of the technical proofs, while it suffices to restrict the positiveness of $f(\cdot)$ and the boundedness of $|\dot{f}(\cdot)|$ in a small and fixed interval $[b_\tau - r, b_\tau + r]$ for some $r>0$.

Let $\kappa_1 = E[\eta_t^2 I(\eta_t < Q_{\tau,\eta})]-\tau$ and $\kappa_2 = E[\eta_t^4]-1$. Define the $(p+q+1)\times(p+q+1)$ matrices $J=E[h_t^{-2}\{\partial h_t(\theta_0)/\partial\theta\}\{\partial h_t(\theta_0)/\partial\theta'\}]$, $\Omega_0 = E[z_t z_t']$, $\Omega_i = E[h_t^{-i} z_t z_t']$, $H_i = E[h_t^{-i} z_t \partial h_t(\theta_0)/\partial\theta']$ and $\Gamma_i = E[h_t^{-i} z_t \Sigma_{j=1}^p \beta_{0j}\partial h_{t-j}(\theta_0)/\partial\theta']$ for $i=1$ and $i=2$,

$$\Sigma_1 = \Omega_2^{-1}\left\{\frac{\tau-\tau^2}{f^2(b_\tau)}\Omega_2 + \frac{\kappa_1 b_\tau}{f(b_\tau)}(\Gamma_2 J^{-1} H_2' + H_2 J^{-1}\Gamma_2') + \kappa_2 b_\tau^2 \Gamma_2 J^{-1}\Gamma_2'\right\}\Omega_2^{-1} \tag{2.5}$$

and

$$\Sigma_2 = \Omega_1^{-1}\left\{\frac{\tau-\tau^2}{f^2(b_\tau)}\Omega_0 + \frac{\kappa_1 b_\tau}{f(b_\tau)}(\Gamma_1 J^{-1} H_1' + H_1 J^{-1}\Gamma_1') + \kappa_2 b_\tau^2 \Gamma_1 J^{-1}\Gamma_1'\right\}\Omega_1^{-1}.$$

Theorem 1. If assumptions 1 and 2 hold, then $\sqrt{n}(\hat{\theta}_{\tau n}-\theta_{\tau 0}) \to_{\mathrm{d}} N(0,\Sigma_1)$.

The weights $\{\tilde{h}_t^{-1}\}$ in equation (2.4) are used to improve the efficiency, as $y_t - Q_\tau(y_t|\mathcal{F}_{t-1}) = h_t(\varepsilon_t - b_\tau)$. Removing the weights gives the unweighted estimator

$$\check{\theta}_{\tau n} = \arg\min_{\theta_\tau} \sum_{t=1}^n \rho_\tau(y_t - \theta_\tau' \tilde{z}_t),$$

and, as the following corollary shows, the asymptotic normality of $\check{\theta}_{\tau n}$ requires a higher order moment condition on x_t, which will entail a smaller available parameter space.

Corollary 1. If $E|x_t|^{4+\iota_0}<\infty$ for some $\iota_0>0$, and assumptions 1 and 2 hold, then $\sqrt{n}(\check{\theta}_{\tau n} - \theta_{\tau 0}) \to_{\mathrm{d}} N(0,\Sigma_2)$.

For the ARCH case, we can show that $\Sigma_2-\Sigma_1$ is always non-negative definite, i.e. $\hat{\theta}_{\tau n}$ is asymptotically more efficient than $\check{\theta}_{\tau n}$. A general comparison of Σ_1 and Σ_2 for the GARCH model is very difficult because of the complicated forms of the two matrices. However, given the true parameter vector, the innovation distribution and τ, we can obtain theoretical values of the constants b_τ, $f(b_\tau)$, κ_1 and κ_2 and estimate all matrices that are involved in Σ_1 and Σ_2 by the corresponding sample averages, based on a generated sequence with a large sample size. Then, we can obtain the asymptotic relative efficiency of $\hat{\theta}_{\tau n}$ to $\check{\theta}_{\tau n}$, defined as $\mathrm{ARE}(\hat{\theta}_{\tau n},\check{\theta}_{\tau n}) = (|\Sigma_2|/|\Sigma_1|)^{1/(p+q+1)}$, where '$|\cdot|$' is the determinant of a matrix; see Serfling (1980). As shown in the on-line supplementary material, the weighted estimator is always asymptotically more efficient than the unweighted estimator, i.e. $\text{ARE}(\hat{\theta}_{\tau n}, \check{\theta}_{\tau n}) > 1$, for GARCH(1, 1) models with different parameter values, innovation distributions and quantile levels. Therefore, we shall focus on the weighted estimator $\hat{\theta}_{\tau n}$ from now on.

Corollary 2. If the conditions in theorem 1 hold, then

$$\hat{Q}_\tau(y_{n+1}|\mathcal{F}_n) - Q_\tau(y_{n+1}|\mathcal{F}_n) = u'_{n+1}(\tilde{\theta}_n - \theta_0) + z'_{n+1}(\hat{\theta}_{\tau n} - \theta_{\tau 0}) + o_p(n^{-1/2}),$$

where $u_{n+1} = b_\tau \Sigma_{j=1}^p \beta_{0j} \partial h_{n+1-j}(\theta_0)/\partial\theta$.

When $b_\tau \neq 0$, it further holds for the τth conditional quantile estimator of x_{n+1} that

$$\hat{Q}_\tau(x_{n+1}|\mathcal{F}_n) - Q_\tau(x_{n+1}|\mathcal{F}_n) = \frac{u'_{n+1}(\tilde{\theta}_n - \theta_0) + z'_{n+1}(\hat{\theta}_{\tau n} - \theta_{\tau 0})}{2\sqrt{|b_\tau h_{n+1}|}} + o_p(n^{-1/2}). \tag{2.6}$$

In practice, multiple quantile levels are often considered simultaneously, say $\tau_1 < \tau_2 < \ldots < \tau_K$. Although $\{\hat{Q}_{\tau_k}(y_{n+1}|\mathcal{F}_n)\}_{k=1}^K$ from the proposed procedure may not be monotonically increasing in k, it is convenient to employ the rearrangement method in Chernozhukov *et al.* (2010) to fix the quantile crossing problem after the estimation.

2.2. Relationship with existing methods

In this subsection, we discuss the relationship between the hybrid method proposed and two important approaches in the literature: the FHS method (Kuester *et al.*, 2006) and CAViaR (Engle and Manganelli, 2004).

We first consider the FHS method. Note that $Q_\tau(y_t|\mathcal{F}_{t-1}) = \theta'_\tau z_t = b_\tau h_t$. If we ignore the GARCH structure and consider a simple weighted linear quantile regression only for the parameter b_τ in the second stage, we have

$$\tilde{b}_{\tau n} = \arg\min_b \sum_{t=1}^n \tilde{h}_t^{-1} \rho_\tau(y_t - b\tilde{h}_t). \tag{2.7}$$

It is not difficult to see that $\tilde{b}_{\tau n}$ is just the τth empirical quantile of $\{y_t/\tilde{h}_t\}$. Thus, the corresponding procedure, with a simplified second-stage estimation, reduces to the FHS method, with the estimates $\hat{Q}_\tau(y_t|\mathcal{F}_{t-1}) = \tilde{b}_{\tau n}\tilde{h}_t = \tilde{\theta}'_{\tau n}\tilde{z}_t$, where $\tilde{\theta}_{\tau n} = \tilde{b}_{\tau n}\tilde{\theta}_n$ is the corresponding FHS estimator of θ_τ. The FHS method relies heavily on the global GARCH structure to fit the conditional quantiles. Specifically, as it allows only b_τ to change across the quantiles, it will suffer from inflexibility in practice, since the real data rarely behave exactly like a GARCH model. The additional simulation results in the on-line supplementary material also demonstrate that the FHS method always has much larger biases than the method proposed.

In contrast, applying the CAViaR method of Engle and Manganelli (2004) to the transformed observations y_t by assuming the linear form (2.1), we have

$$\hat{\vartheta}_{\tau n} = \arg\min_\vartheta \sum_{t=1}^n \rho_\tau\{y_t - \vartheta' v_t(\vartheta)\}, \tag{2.8}$$

where $v_t(\vartheta) = (1, x_{t-1}^2, \ldots, x_{t-q}^2, q_{t-1}(\vartheta), \ldots, q_{t-p}(\vartheta))'$ with $q_s(\vartheta) = \vartheta' v_s(\vartheta)$. Unlike the proposed $\hat{\theta}_{\tau n}$ and the FHS estimator $\tilde{\theta}_{\tau n}$ which both converge to $\theta_{\tau 0} = b_\tau \theta_0$, the CAViaR estimator $\hat{\vartheta}_{\tau n}$ converges to $\vartheta_{\tau 0} := (b_\tau \alpha_{00}, b_\tau \alpha_{01}, \ldots, b_\tau \alpha_{0q}, \beta_{01}, \ldots, \beta_{0p})'$. This approach will lead to the unweighted estimator $\check{\theta}_{\tau n}$ in the Section 2.1 if we first obtain initial estimates of $\{q_t(\vartheta)\}$, and hence those of $v_t(\vartheta)$ in equation (2.8), by replacing ϑ with the more efficient Gaussian QMLE $\tilde{\theta}_n$, and then perform quantile regression in equation (2.8). As a result, CAViaR is even less efficient than the unweighted method in Section 2.1, although it enjoys greater flexibility than the FHS method since it imposes a structure at only the quantile level τ. Moreover, the computation of CAViaR is generally challenging, which actually requires grid search.

We may interpret the method proposed as a hybrid version of FHS and CAViaR. It combines the efficiency of the former and the flexibility of the latter and hence may perform better in practice. However, when the data are exactly generated by a GARCH model, the proposed estimator $\hat{\theta}_{\tau n}$ may be less efficient than the FHS estimator $\tilde{\theta}_{\tau n}$. Let

$$\Sigma_3 = \frac{\tau - \tau^2}{f^2(b_\tau)}\theta_0\theta_0' + \frac{\kappa_1 b_\tau}{f(b_\tau)}\Sigma_0 + \kappa_2 b_\tau^2(\Sigma_0 + J^{-1} - \theta_0\theta_0'),$$

where $\bar{\beta}_0 = (0, \ldots, 0, \beta_{01}, \ldots, \beta_{0p})' \in \mathbb{R}^{p+q+1}$ and $\Sigma_0 = \theta_0\bar{\beta}_0' + \bar{\beta}_0\theta_0'$. If the conditions in theorem 1 hold, we can show that $\sqrt{n}(\tilde{\theta}_{\tau n} - \theta_{\tau 0}) \to_{\text{d}} N(0, \Sigma_3)$; see also Gao and Song (2008) and Francq and Zakoïan (2015). In particular, for ARCH models, Σ_1 and Σ_3 reduce to $(\tau - \tau^2)J^{-1}/f^2(b_\tau)$ and $(\tau - \tau^2)\theta_0\theta_0'/f^2(b_\tau) + \kappa_2 b_\tau^2(J^{-1} - \theta_0\theta_0')$ respectively. Then, we can further show that $\Sigma_3 - \Sigma_1$ is non-negative definite if and only if $(\tau - \tau^2)/f^2(b_\tau) - \kappa_2 b_\tau^2 \leqslant 0$, which depends on the specific innovation distribution and quantile level τ. For the GARCH model, similarly to our discussion on the unweighted estimator in Section 2.1, we have computed the ARE of the proposed estimator $\hat{\theta}_{\tau n}$ to the FHS estimator $\tilde{\theta}_{\tau n}$ for GARCH(1, 1) models for various parameter settings, innovation distributions and quantile levels. As expected, the FHS estimator $\tilde{\theta}_{\tau n}$ is asymptotically more efficient in general, whereas the proposed estimator $\hat{\theta}_{\tau n}$ can be asymptotically more efficient when $\{\eta_t\}$ become more heavy tailed; see the on-line supplementary material for details.

3. A mixed bootstrapping procedure

To circumvent difficulties due to the density function in the asymptotic covariance matrix in theorem 1, we propose a bootstrapping procedure to approximate the asymptotic distribution of $\hat{\theta}_{\tau n}$, which benefits from both the convenience of the random-weighting bootstrap method in Jin *et al.* (2001) and the time efficiency of sample averaging.

From the proof of theorem 1, the Gaussian QMLE $\tilde{\theta}_n$ affects the asymptotic distribution of the proposed estimator $\hat{\theta}_{\tau n}$ through the relationship $\sqrt{n}(\hat{\theta}_{\tau n} - \theta_{\tau 0}) = \Omega_2^{-1}T_{1n}/f(b_\tau) - b_\tau\Omega_2^{-1}\Gamma_2\sqrt{n}(\tilde{\theta}_n - \theta_0) + o_p(1)$, where $T_{1n} = n^{-1/2}\Sigma_{t=1}^n \psi_\tau(\varepsilon_t - b_\tau)z_t/h_t$, with $\psi_\tau(x) = \tau - I(x < 0)$. Apparently, the random-weighting bootstrap should be incorporated in both steps 1 and 2, leading to the following bootstrapping procedure.

Step 1: perform the randomly weighted Gaussian QMLE

$$\tilde{\theta}_n^* = \arg\min_{\theta \in \Theta} \sum_{t=1}^n \omega_t \tilde{l}_t(\theta), \tag{3.1}$$

where $\{\omega_t\}$ are IID non-negative random weights with mean and variance both equal to 1, and then compute the initial estimates of $\{h_t\}$ as $\tilde{h}_t^* = \tilde{h}_t(\tilde{\theta}_n^*)$.
Step 2: perform the randomly weighted quantile regression

$$\hat{\theta}_{\tau n}^* = \arg\min_{\theta_\tau} \sum_{t=1}^n \omega_t \tilde{h}_t^{*-1} \rho_\tau(y_t - \theta_\tau' \tilde{z}_t^*), \tag{3.2}$$

where $\tilde{z}_t^* = (1, x_{t-1}^2, \ldots, x_{t-q}^2, \tilde{h}_{t-1}^*, \ldots, \tilde{h}_{t-p}^*)'$.
Step 3: calculate the conditional quantile estimate $\hat{Q}_\tau^*(x_t|\mathcal{F}_{t-1}) = T^{-1}(\hat{\theta}_{\tau n}^{*\prime}\tilde{z}_t^*)$.

The purpose of the bootstrapping procedure is to avoid estimating the density $f(b_\tau)$ that is involved in the asymptotic covariance matrix Σ_1. Observe that no density actually appears in the asymptotic covariance matrix of the Gaussian QMLE $\tilde{\theta}_n$. This motivates us to replace the optimization in step 1 above with a simple sample averaging. Note that

$$\begin{aligned} \sqrt{n}(\tilde{\theta}_n^* - \tilde{\theta}_n) &= \frac{1}{\sqrt{n}} \sum_{t=1}^{n} (\omega_t - 1)\xi_t + o_{p^*}(1), \\ \sqrt{n}(\tilde{\theta}_n - \theta_0) &= \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \xi_t + o_p(1), \end{aligned} \tag{3.3}$$

where $\xi_t = J^{-1}(|y_t|/h_t - 1)h_t^{-1}\{\partial h_t(\theta_0)/\partial\theta\}$ and $\tilde{\theta}_n^*$ is defined as in expression (3.1); see also Francq and Zakoian (2004). The matrix $J = E[h_t^{-2}\{\partial h_t(\theta_0)/\partial\theta\}\{\partial h_t(\theta_0)/\partial\theta'\}]$ can be estimated consistently by $\tilde{J} = n^{-1}\Sigma_{t=1}^n \tilde{h}_t^{-2}\{\partial\tilde{h}_t(\tilde{\theta}_n)/\partial\theta\}\{\partial\tilde{h}_t(\tilde{\theta}_n)/\partial\theta'\}$. Therefore, step 1 can be replaced by the following step.

Step 1′: calculate the estimator $\tilde{\theta}_n^*$ by

$$\tilde{\theta}_n^* = \tilde{\theta}_n - \frac{\tilde{J}^{-1}}{n} \sum_{t=1}^{n} (\omega_t - 1)\left(1 - \frac{|y_t|}{\tilde{h}_t}\right) \frac{1}{\tilde{h}_t} \frac{\partial\tilde{h}_t(\tilde{\theta}_n)}{\partial\theta}. \tag{3.4}$$

Combining steps 1′, 2 and 3, we have a mixed bootstrapping procedure.

Assumption 3. The random weights $\{\omega_t\}$ are IID non-negative random variables with mean and variance both equal to 1, satisfying $E|\omega_t|^{2+\kappa_0} < \infty$ for some $\kappa_0 > 0$.

Theorem 2. Suppose that $E|\eta_t|^{4+2\nu_0} < \infty$ for some $\nu_0 > 0$ and assumptions 1–3 hold. Then, conditionally on $\mathcal{F}_n$, $\sqrt{n}(\hat{\theta}_{\tau n}^* - \hat{\theta}_{\tau n}) \to_{\mathrm{d}} N(0, \Sigma_1)$ in probability as $n \to \infty$, where Σ_1 is defined as in theorem 1.

Corollary 3. Under the conditions of theorem 2, it holds that

$$\hat{Q}_\tau^*(y_{n+1}|\mathcal{F}_n) - \hat{Q}_\tau(y_{n+1}|\mathcal{F}_n) = u_{n+1}'(\tilde{\theta}_n^* - \tilde{\theta}_n) + z_{n+1}'(\hat{\theta}_{\tau n}^* - \hat{\theta}_{\tau n}) + o_p^*(n^{-1/2}),$$

where u_{n+1} is defined as in corollary 2.

By corollaries 2 and 3, along with the asymptotic results for $\tilde{\theta}_n^*$ and $\hat{\theta}_{\tau n}^*$ in the proof of theorem 2, the confidence interval for the conditional quantile $Q_\tau(x_{n+1}|\mathcal{F}_n)$ can be easily constructed on the basis of the bootstrap sample $\{\hat{Q}_\tau^*(x_{n+1}|\mathcal{F}_n)\}$, where $\hat{Q}_\tau^*(x_{n+1}|\mathcal{F}_n) = T^{-1}\{\hat{Q}_\tau^*(y_{n+1}|\mathcal{F}_n)\}$; see also Spierdijk (2016).

The first-order validity of the mixed bootstrapping procedure proposed is established by theorem 2 and corollary 3. Unfortunately, the second-order correctness (Lahiri, 2003) is almost impossible to achieve. In fact, as long as quantile regression is employed, because of the non-smoothness of the loss function $\rho_\tau(\cdot)$, it will be very difficult to attain second-order correctness for the bootstrapping procedure; see also Horowitz (1998). Note also that the $o_p(1)$ term in expression (3.3) plays a non-negligible role in the Edgeworth expansion of $\sqrt{n}(\tilde{\theta}_n - \theta_0)$ (Linton, 1997) but is ignored by $\tilde{\theta}_n^*$ in step 1′. Hence, the second-order correctness has already been lost when we use the much faster sample averaging method in step 1′ to replace the optimization in step 1. However, the sacrifice is worthwhile, as the second-order correctness is unachievable anyway because of the non-smooth objective function in step 2. Actually, in the literature, bootstrap methods with second-order correctness are still limited to the GARCH(1, 1) model and are unavailable for the general GARCH model (Corradi and Iglesias, 2008; Jeong, 2017).

4. Diagnostic checking for conditional quantiles

On the basis of the proposed procedures in Sections 2 and 3, we next construct a portmanteau test to check the adequacy of fitted conditional quantiles.

Let $\varepsilon_{t,\tau} = h_t^{-1}\{y_t - Q_\tau(y_t|\mathcal{F}_{t-1})\} = \varepsilon_t - b_\tau$. We define the QACF of $\{\varepsilon_{t,\tau}\}$ at lag k as

$$\rho_{k,\tau} = \text{qcorr}_\tau(\varepsilon_{t,\tau}, |\varepsilon_{t-k,\tau}|) = \frac{E[\psi_\tau(\varepsilon_{t,\tau})|\varepsilon_{t-k,\tau}|]}{\surd\{(\tau-\tau^2)\sigma_{a,\tau}^2\}}, \qquad k = 1, 2, \ldots,$$

where $\sigma_{a,\tau}^2 = \text{var}(|\varepsilon_{t,\tau}|) = E[|\varepsilon_{t,\tau}| - \mu_{a,\tau}]^2$, with $\mu_{a,\tau} = E|\varepsilon_{t,\tau}|$; see also the QACF in Li *et al.* (2015) and the absolute residual ACF in Li and Li (2005). If $Q_\tau(x_t|\mathcal{F}_{t-1})$ is correctly specified by model (1.2), then $E[\psi_\tau(\varepsilon_{t,\tau})|\mathcal{F}_{t-1}]$ is 0 and so is $\rho_{k,\tau}$ for any $k \geqslant 1$.

Accordingly, let $\hat{\varepsilon}_{t,\tau} = \tilde{h}_t^{-1}(y_t - \hat{\theta}_{\tau n}'\tilde{z}_t)$, and then the corresponding residual QACF at lag k can be calculated as $r_{k,\tau} = (\tau-\tau^2)^{-1/2}\hat{\sigma}_{a,\tau}^{-1} n^{-1} \Sigma_{t=k+1}^n \psi_\tau(\hat{\varepsilon}_{t,\tau})|\hat{\varepsilon}_{t-k,\tau}|$, where $\hat{\sigma}_{a,\tau}^2 = n^{-1}\Sigma_{t=1}^n (|\hat{\varepsilon}_{t,\tau}| - \hat{\mu}_{a,\tau})^2$, with $\hat{\mu}_{a,\tau} = n^{-1}\Sigma_{t=1}^n |\hat{\varepsilon}_{t,\tau}|$. For a predetermined positive integer K, we first derive the asymptotic distribution of $R = (r_{1,\tau}, \ldots, r_{K,\tau})'$.

Let $\epsilon_t = (|\varepsilon_{t,\tau}|, |\varepsilon_{t-1,\tau}|, \ldots, |\varepsilon_{t-K+1,\tau}|)'$ and $\Xi = E[\epsilon_t \epsilon_t']$, and define the $K \times (p+q+1)$ matrices $D_1 = E[h_t^{-1}\epsilon_{t-1} z_t']$, $D_2 = E[h_t^{-1}\epsilon_{t-1}\Sigma_{j=1}^p \beta_{0j}\partial h_{t-j}(\theta_0)/\partial\theta']$ and $D_3 = E[h_t^{-1}\epsilon_{t-1}\partial h_t(\theta_0)/\partial\theta']$. In addition, let $P = D_2 - D_1\Omega_2^{-1}\Gamma_2$, $Q = D_3 - D_1\Omega_2^{-1}H_2$, $\Omega_3 = D_1\Omega_2^{-1}D_1'$ and

$$\Sigma_4 = \sigma_{a,\tau}^{-2}\left\{\Xi - \Omega_3 + \frac{\kappa_1 b_\tau f(b_\tau)}{\tau-\tau^2}(QJ^{-1}P' + PJ^{-1}Q') + \frac{\kappa_2 b_\tau^2 f^2(b_\tau)}{\tau-\tau^2} PJ^{-1}P'\right\}. \tag{4.1}$$

Theorem 3. If $E|\eta_t|^{4+2\nu_0} < \infty$ for some $\nu_0 > 0$ and assumptions 1 and 2 hold, then $\surd n R \to_{\text{d}} N(0, \Sigma_4)$, where Σ_4 is a positive definite matrix.

Theorem 3 implies that the portmanteau test statistic $Q(K) = nR'\hat{\Sigma}_4^{-1}R$ converges to a χ^2-distribution with K degrees of freedom as $n \to \infty$, where $\hat{\Sigma}_4$ is a consistent estimator of Σ_4. Note that, even for the ARCH case, the asymptotic covariance matrix $\Sigma_4 = \sigma_{a,\tau}^{-2}(\Xi - D_1 J^{-1} D_1')$ still depends on the parameter vector θ_0, the density $f(\cdot)$ and the quantile level τ in a complicated way.

We next employ the bootstrap method to approximate Σ_4. Let $\hat{\varepsilon}_{t,\tau}^* = \tilde{h}_t^{-1}(y_t - \hat{\theta}_{\tau n}^{*\prime}\tilde{z}_t^*)$, $r_{k,\tau}^* = (\tau-\tau^2)^{-1/2}\hat{\sigma}_{a,\tau}^{-1} n^{-1}\Sigma_{t=k+1}^n \omega_t \psi_\tau(\hat{\varepsilon}_{t,\tau}^*)|\hat{\varepsilon}_{t-k,\tau}^*|$ and $R^* = (r_{1,\tau}^*, \ldots, r_{K,\tau}^*)'$.

Theorem 4. Suppose that the conditions in theorem 2 hold. Then, conditionally on $\mathcal{F}_n$, $\surd n(R^* - R) \to_{\text{d}} N(0, \Sigma_4)$ in probability as $n \to \infty$, where Σ_4 is defined as in theorem 3.

In step 3 in Section 3, we can calculate R^* and $T^{(1)} = \surd n(R^* - R)$. Then, repeating steps 1′ and 2 for $B-1$ times yields $\{T^{(1)}, \ldots, T^{(B)}\}$, and Σ_4 can be approximated by the sample covariance matrix Σ_4^* of $\{T^{(i)}\}_{i=1}^B$. Therefore, we reject the null hypothesis that $r_{k,\tau}$ with $1 \leqslant k \leqslant K$ are jointly insignificant if $Q(K)$ exceeds the 0.95th theoretical quantile of χ_K^2. In addition, we reject the null hypothesis that $r_{k,\tau}$ is individually insignificant if $\surd n r_{k,\tau}$ falls outside the range between the 0.025th and 0.975th empirical quantiles of $\{T_k^{(i)}\}_{i=1}^B$, where $T_k^{(i)}$ is the kth element of $T^{(i)}$.

5. Simulation studies

This section contains three simulation experiments for evaluating the finite sample performance of the estimation, bootstrapping and diagnostic checking procedures proposed.

In the first experiment, we focus on the proposed estimator $\hat{\theta}_{\tau n}$ and the bootstrapping approximation of its asymptotic distribution. The data are generated from the GARCH(1, 1) model with $(\alpha_0, \alpha_1, \beta_1) = (0.1, 0.15, 0.8)$, where the innovations $\{\eta_t\}$ are standard normal or follow the standardized Student t_5-distribution with unit variance. We consider three sample sizes, $n = 500$, 1000, 2000, with 1000 replications generated for each sample size, and two quantile levels: $\tau = 0.05$ and $\tau = 0.1$. Four distributions for the random weights $\{\omega_t\}$ in the bootstrapping

Table 1. Bias, ESD and ASD for $\hat{\theta}_{\tau n}$ at $\tau = 0.05$ or $\tau = 0.1$, for normal or Student t_5-distributed innovations, where ASD_i corresponds to random weight W_i for $i = 1, 2, 3, 4$, and α_0, α_1 and β_1 represent corresponding elements of $\hat{\theta}_{\tau n}$

n		*Results for normal distribution*						*Results for Student's t_5-distribution*					
		Bias ($\times 10$)	*ESD* ($\times 10$)	ASD_1 ($\times 10$)	ASD_2 ($\times 10$)	ASD_3 ($\times 10$)	ASD_4 ($\times 10$)	*Bias* ($\times 10$)	*ESD* ($\times 10$)	ASD_1 ($\times 10$)	ASD_2 ($\times 10$)	ASD_3 ($\times 10$)	ASD_4 ($\times 10$)
$\tau = 0.05$													
500	α_0	−0.24	10.20	11.48	11.77	11.28	11.57	−0.61	10.42	13.88	14.85	13.71	15.13
	α_1	−0.07	3.05	3.26	3.25	3.26	3.26	−0.75	3.89	4.53	4.08	4.27	4.34
	β_1	0.03	7.52	8.15	8.65	8.12	8.34	0.32	8.33	11.38	13.63	11.21	12.61
1000	α_0	0.20	6.06	7.00	7.09	7.03	7.06	−0.30	6.84	8.46	8.12	7.81	8.18
	α_1	0.08	2.24	2.31	2.29	2.30	2.30	−0.25	2.60	2.89	2.73	2.79	2.81
	β_1	−0.25	4.76	5.25	5.34	5.28	5.30	−0.04	5.81	7.06	7.30	6.72	7.18
2000	α_0	0.24	4.38	4.68	4.71	4.69	4.70	−0.05	4.72	5.18	5.11	5.00	5.20
	α_1	0.07	1.59	1.62	1.61	1.61	1.61	−0.16	1.84	1.98	1.91	1.94	1.95
	β_1	−0.24	3.48	3.60	3.63	3.61	3.61	−0.09	4.20	4.50	4.59	4.41	4.62
$\tau = 0.1$													
500	α_0	−0.09	6.47	7.22	7.28	7.12	7.31	−0.34	5.28	7.40	7.98	7.19	7.65
	α_1	0.00	1.90	2.07	2.04	2.06	2.06	−0.32	1.86	2.10	1.99	2.04	2.05
	β_1	−0.14	4.75	5.16	5.33	5.14	5.26	0.21	4.23	6.11	7.28	5.88	6.45
1000	α_0	0.00	4.11	4.39	4.43	4.41	4.42	−0.14	3.55	4.33	4.30	4.17	4.37
	α_1	0.06	1.38	1.44	1.43	1.44	1.44	−0.10	1.26	1.38	1.34	1.36	1.36
	β_1	−0.13	3.17	3.30	3.33	3.31	3.32	0.00	2.92	3.66	3.86	3.59	3.83
2000	α_0	0.07	2.74	2.98	2.99	2.98	2.98	0.08	2.54	2.75	2.71	2.67	2.79
	α_1	0.04	0.96	1.01	1.01	1.01	1.01	−0.07	0.89	0.95	0.94	0.94	0.95
	β_1	−0.14	2.14	2.29	2.30	2.29	2.29	−0.14	2.24	2.39	2.43	2.34	2.46

procedure are considered: the standard exponential distribution W_1, the Rademacher distribution W_2, which takes the value 0 or 2, each with probability 0.5 (Li *et al.*, 2014), Mammen's two-point distribution W_3, which takes the value $(-\sqrt{5}+3)/2$ with probability $(\sqrt{5}+1)/2\sqrt{5}$ or the value $(\sqrt{5}+3)/2$ with probability $1-(\sqrt{5}+1)/2\sqrt{5}$ (Mammen, 1993), and a mixture of the standard exponential distribution and the Rademacher distribution W_4, with mixing probability 0.5.

The bias, empirical standard deviation (ESD) and asymptotic standard deviation (ASD) for $\hat{\theta}_{\tau n}$ are reported in Table 1, where the ASDs are estimated by the proposed bootstrapping procedure using different distributions for the random weights. We have the following findings:

(a) the biases are all small;
(b) as n or τ increases, the bias and standard deviations decrease, and the ASDs become closer to the corresponding ESDs;
(c) the performance of the bootstrapping approximation is insensitive to the choice of random weights;
(d) the ASDs appear to be closer to the corresponding ESDs when $\{\eta_t\}$ are normal than when they follow the Student t_5-distribution;
(e) when $\tau = 0.05$, the standard deviations for the normal distribution are smaller than those for the Student t_5-distribution, whereas the opposite holds for most cases when $\tau = 0.1$.

Generally speaking, for GARCH models, heavier tails of $\{\eta_t\}$ will lead to lower efficiency of the Gaussian QMLE and higher efficiency of the quantile regression, which results in mixed perfor-

Table 2. Bias, ESD and ASD for the residual QACF $r_{k,\tau}$ at $\tau = 0.05$ or $\tau = 0.1$ and $k = 2, 4, 6$, for normal or Student t_5-distributed innovations

n	*k*	*Results for normal distribution*						*Results for Student's t_5-distribution*					
		$\tau = 0.05$			$\tau = 0.1$			$\tau = 0.05$			$\tau = 0.1$		
		Bias ($\times 100$)	*ESD* ($\times 100$)	*ASD* ($\times 100$)	*Bias* ($\times 100$)	*ESD* ($\times 100$)	*ASD* ($\times 100$)	*Bias* ($\times 100$)	*ESD* ($\times 100$)	*ASD* ($\times 100$)	*Bias* ($\times 100$)	*ESD* ($\times 100$)	*ASD* ($\times 100$)
500	2	1.27	4.88	6.72	0.67	4.35	5.34	0.78	4.36	5.91	0.69	4.32	4.82
	4	0.90	4.88	6.83	0.47	4.59	5.43	0.69	4.67	5.94	0.42	4.31	4.84
	6	1.04	4.91	6.81	0.61	4.64	5.44	0.37	4.75	6.03	0.08	4.52	4.90
1000	2	0.48	3.24	4.05	0.36	3.13	3.44	0.30	3.13	3.57	0.25	3.14	3.26
	4	0.50	3.34	4.09	0.15	3.19	3.51	0.35	3.13	3.54	0.30	3.01	3.17
	6	0.43	3.29	4.13	0.30	3.16	3.54	0.18	3.35	3.66	−0.01	3.20	3.29
2000	2	0.29	2.23	2.59	0.20	2.23	2.33	0.28	2.15	2.30	0.09	2.21	2.23
	4	0.15	2.26	2.62	0.02	2.14	2.36	0.10	2.26	2.31	0.10	2.19	2.21
	6	0.16	2.25	2.63	0.14	2.19	2.38	0.15	2.20	2.32	0.04	2.18	2.23

Table 3. Rejection rate of the test statistic $Q(K)$ for $K = 6$ at the 5% level of significance, for normal or Student t_5-distributed innovations and $d = 0$, 0.3, 0.6

n	*Results (%) for normal distribution*						*Results (%) for Student's t_5-distribution*					
	$\tau = 0.05$			$\tau = 0.1$			$\tau = 0.05$			$\tau = 0.1$		
	$d = 0.0$	$d = 0.3$	$d = 0.6$	$d = 0.0$	$d = 0.3$	$d = 0.6$	$d = 0.0$	$d = 0.3$	$d = 0.6$	$d = 0.0$	$d = 0.3$	$d = 0.6$
500	2.8	4.8	7.4	3.4	6.9	27.0	1.9	3.8	7.8	3.4	6.5	21.0
1000	3.3	7.2	21.6	4.0	15.7	60.9	3.0	10.6	29.4	4.3	16.3	46.8
2000	4.5	16.1	55.2	4.9	36.5	92.5	5.3	27.9	69.8	4.3	34.3	83.2

mance of the proposed method under different innovation distributions, and the performance is further affected by the specific parameter values and quantile level.

The second experiment considers the proposed residual QACF $r_{k,\tau}$ and the bootstrapping approximation of its asymptotic distribution. The data and all other settings are the same as in the previous experiment. For brevity, we present only results for W_1 from now on, and the results for W_2, W_3 and W_4 are provided in the on-line supplementary material, where it is found that the performance is insensitive to the choice of random weights. Table 2 provides the bias, ESD and ASD for $r_{k,\tau}$ at lags $k = 2, 4, 6$. Findings (a) and (b) in the previous experiment are also observed in Table 2. Furthermore, we have repeated the first two experiments by using $\tau = 0.01$ and have found that the sample size may have to be as large as 5000 to achieve a good approximation.

The third experiment examines the empirical size and power of the test statistic $Q(K)$. The data are generated from

$$x_t = \sqrt{h_t}\eta_t, \qquad h_t = 0.4 + 0.2x_{t-1}^2 + dx_{t-4}^2 + 0.2h_{t-1},$$

where the departure $d=0$, 0.3, 0.6. We conduct conditional quantile estimation based on the GARCH(1, 1) model assumption; thus, $d=0$ corresponds to the size of the test, and $d\neq 0$ corresponds to the power. All other settings are preserved from the previous experiment. Table 3 reports the rejection rate at the maximum lag $K=6$. It can be seen that the rejection rate increases as either n or the departure d increases. To make the size close to the nominal rate 5%, the sample size n needs to be as large as 2000 at $\tau=0.05$, whereas $n=1000$ is sufficient for $\tau=0.1$. Moreover, as τ increases from 0.05 to 0.1, the increase in the power is larger for the normal distribution than for the Student t_5-distribution. When τ grows closer to 0, the actual departure in the quantile regression, namely $|b_\tau d|$, increases, whereas the density $f(b_\tau)$ decreases as the data around b_τ become more sparse. Consequently, the overall effect of τ on the power is mixed and depends on the specific innovation distribution.

6. Empirical analysis

In this section, we analyse the daily log-returns of three stock market indices from January 2nd, 2008, to June 30th, 2016: the Standard&Poors S&P500-index, the Dow Jones 30-index and the Hang Seng index (HSI). The sample sizes are $n=2139$, 2139, 2130 respectively.

We begin by illustrating the proposed method with the S&P500 data for $\tau=0.05$, i.e. the 1-day 5% VaR. Fig. 1 gives the time plot of the log-returns $\{x_t\}$. By the estimation procedure proposed, the initial estimates of $\{h_t\}$ are calculated by $\tilde{h}_t=2.646\times 10^{-6}+0.126x_{t-1}^2+0.858\tilde{h}_{t-1}$, and the fitted conditional quantile function is

$$\hat{Q}_{0.05}(y_t|\mathcal{F}_{t-1})=-4.713\times 10^{-7}-0.124x_{t-1}^2-3.007\tilde{h}_{t-1}.$$

Fig. 1 shows that the residual QACF falls only slightly outside the corresponding 95% confidence interval at lags 3, 21 and 24, and is well within it at all the other lags. By the diagnostic checking procedure proposed, the p-values of the portmanteau test $Q(K)$ are all larger than 0.257 for $K=6$, 12, 18, 24, 30, which suggests the adequacy of the fitted conditional quantiles.

Next we examine the forecasting performance of the proposed method for all stock market indices by using the following rolling procedure: first, conduct the estimation by using the first 2 years' data and compute the conditional quantile forecast for the next trading day, i.e.

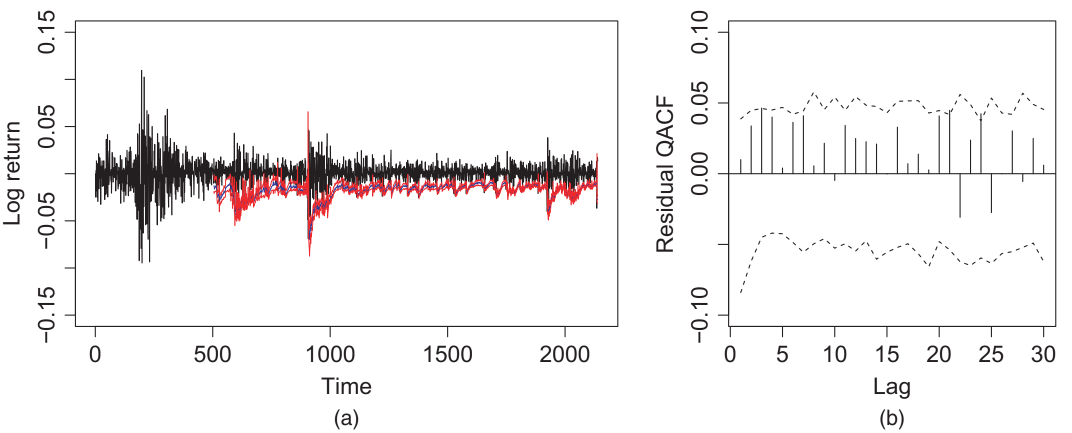

Fig. 1. (a) Time plot of daily log-returns (——) of the S&P500-index from January 2nd, 2008, to June 30th, 2016, and rolling forecasts of the conditional quantiles (——) at $\tau=0.05$ from January 4th, 2010, to June 30th, 2016, with corresponding 95% confidence bounds (——) and (b) residual QACF of the fitted GARCH model at $\tau=0.05$, with corresponding 95% confidence bounds

the forecast of $Q_\tau(x_{n+1}|\mathcal{F}_n)$; then, advance the forecasting origin by 1 to include one more observation in the estimation subsample, and repeat the foregoing procedure until the end of the sample has been reached. See Fig. 1 for an illustration of the rolling forecasts at $\tau=0.05$ for the S&P500 data, where the corresponding 95% confidence intervals are constructed by the bootstrapping procedure proposed.

To compare the forecasting performance of the proposed method with existing conditional quantile estimation methods, we also conduct rolling forecasting for the FHS method that was discussed in Section 2.2 and four other methods which we call XK_1, XK_2, CAViaR and RiskM in what follows. In particular, XK_1 and XK_2 are adapted versions of the 'QGARCH1' and 'QGARCH2' methods in Xiao and Koenker (2009) for the GARCH model, where we first apply the transformation $T(\cdot)$ to the observed sequence $\{x_t\}$ as in estimation step 1 of the proposed procedure in Section 2.1. For XK_1, the initial estimates of $\{h_t\}$ are obtained by linear quantile regression at quantile level τ by using the sieve approximation, $h_t=\gamma_0+\Sigma_{j=1}^m \gamma_j x_{t-j}^2$, where we set $m=3n^{1/4}$ as in Xiao and Koenker (2009). For XK_2, the initial estimates of $\{h_t\}$ are obtained by combining the sieve-approximation-based estimation in XK_1 over multiple quantile levels, $\tau_i=i/20$ for $i=1,2,\ldots,19$, via minimum distance estimation. CAViaR refers to the indirect GARCH(1, 1) based CAViaR method in Engle and Manganelli (2004), and we use the MATLAB code from them for the grid search optimization and the same settings of initial values for the optimization as in Engle and Manganelli (2004). Finally, RiskM refers to the conventional RiskMetrics method, which assumes that the data follow the integrated GARCH(1, 1) model $x_t=\sqrt{h_t}\eta_t$, $h_t=0.06x_{t-1}^2+0.94h_{t-1}$, where $\{\eta_t\}$ are IID standard normal; see Morgan and Reuters (1996) and Tsay (2010).

We use VaR-backtesting as the primary criterion, and the empirical coverage performance as the secondary criterion. Specifically, we adopt the following two measures:

(a) the minimum of the p-values of the two VaR-backtests, the likelihood ratio test for correct conditional coverage in Christoffersen (1998) and the dynamic quantile test in Engle and Manganelli (2004);
(b) the empirical coverage error, namely the empirical coverage rate (i.e. the proportion of observations that exceed the corresponding VaR-forecast) minus the corresponding nominal rate τ.

For the dynamic quantile test, following Kuester *et al.* (2006), the regressor matrix contains four lagged hits, $\text{Hit}_{t-1},\ldots,\text{Hit}_{t-4}$, and the contemporaneous VaR-estimate, where Hit_t is the indicator of exceedance for the observation at time t. We consider the smaller of the two p-values, because the conditional coverage and dynamic quantile tests have different null hypotheses and hence are complementary to each other.

Table 4 presents the results of the two measures for the six estimation methods at the lower L and upper U 0.01th, 0.025th and 0.05th conditional quantiles, i.e. the 1%, 2.5% and 5% VaRs for long and short positions. For the S&P500 and Dow 30-data, it can be seen that none of the methods performs satisfactorily at the lower quantiles. For the upper quantiles of these two data sets, both XK_1 and XK_2 perform poorly, whereas the other methods are generally adequate: all p-values for the hybrid method proposed and RiskM are larger than 0.2 and, despite the small p-value at U5.0 for the Dow 30-data, the FHS method performs fairly well. For the HSI-data, the FHS method is adequate at all quantiles, and the hybrid method proposed performs well except for the case of U1.0. In contrast, RiskM performs poorly at the lower quantiles, and CAViaR is unsatisfactory at U2.5 and U1.0. Therefore, it is clear that, in terms of the backtesting performance, the method proposed and the FHS method dominate the other competitors. Indeed, for the three data sets at the six quantile levels, among all methods,

Table 4. Minimum p-value of two VaR-backtests and empirical coverage error for six estimation methods for three stock market indices at the lower L and upper U 0.01th, 0.025th and 0.05th conditional quantiles

Index	τ	*Minimum p-value of VaR-backtests*						*Empirical coverage errors (%)*					
		Hybrid	*FHS*	XK_1	XK_2	*CAViaR*	*RiskM*	*Hybrid*	*FHS*	XK_1	XK_2	*CAViaR*	*RiskM*
S&P500	L1.0	0.000	0.082	0.000	0.000	0.030	0.000	−0.02	0.04	−0.57	−0.45	−0.45	1.57
	L2.5	0.001	0.005	0.000	0.000	0.005	0.000	−0.48	−0.36	−1.77	−1.52	−0.79	1.84
	L5.0	0.017	0.016	0.000	0.000	0.006	0.000	−0.90	−1.15	−2.31	−1.82	−1.39	1.12
	U5.0	0.245	0.244	0.012	0.018	0.253	0.855	0.54	0.84	1.51	1.45	0.66	0.05
	U2.5	0.356	0.222	0.010	0.220	0.502	0.557	0.30	0.42	1.03	0.48	0.18	−0.19
	U1.0	0.275	0.342	0.130	0.653	0.626	0.382	0.08	0.33	0.45	0.14	0.20	−0.04
Dow 30	L1.0	0.063	0.115	0.000	0.000	0.001	0.000	−0.14	0.16	−0.57	−0.45	−0.45	1.63
	L2.5	0.000	0.000	0.000	0.000	0.001	0.000	−0.54	−0.24	−1.40	−0.97	−0.60	1.90
	L5.0	0.000	0.027	0.000	0.000	0.002	0.006	−0.72	−0.78	−2.37	−2.06	−1.02	0.87
	U5.0	0.273	0.064	0.000	0.002	0.135	0.555	0.84	1.21	2.00	1.51	1.02	0.23
	U2.5	0.568	0.806	0.011	0.044	0.671	0.762	0.11	0.24	1.03	0.36	0.11	0.05
	U1.0	0.418	0.221	0.031	0.741	0.217	0.754	−0.28	0.39	0.57	0.14	−0.41	−0.10
HSI	L1.0	0.393	0.425	0.004	0.632	0.827	0.000	0.11	−0.02	−0.57	−0.20	−0.08	1.34
	L2.5	0.362	0.290	0.000	0.072	0.355	0.006	−0.04	0.14	−1.15	−0.65	−0.22	1.01
	L5.0	0.421	0.159	0.000	0.026	0.095	0.003	−0.69	−0.69	−1.99	−1.31	−0.94	1.46
	U5.0	0.766	0.635	0.003	0.014	0.169	0.818	0.14	−0.23	1.68	1.43	0.51	−0.29
	U2.5	0.477	0.631	0.012	0.083	0.038	0.492	0.04	0.35	1.02	0.78	−0.08	−0.08
	U1.0	0.048	0.492	0.033	0.010	0.036	0.019	−0.17	0.26	0.57	0.32	−0.23	−0.35

the method proposed has the largest number of cases where the minimum p-value exceeds 0.2, whereas the FHS method has the smallest number of cases where the minimum p-value is less than 0.05.

To determine whether the method proposed or the FHS method is superior, we next take into account the secondary criterion: the empirical coverage error. To do so, for each method we count the numbers of cases (among the total of 18 cases) where the absolute value of its corresponding empirical coverage error is the smallest and second smallest among all methods. From the right-hand panel of Table 4, the results are 9 and 6 for the method proposed, and 4 and 5 for the FHS method. For the other competitors, the numbers are all much smaller. In the on-line supplementary material, we also conduct a case-by-case comparison of these two methods based on a more comprehensive analysis of the backtesting and empirical coverage results, and it is shown that the method proposed does have a clearly better performance than the FHS method.

Moreover, we have also performed the foregoing analysis again by using the rearrangement method of Chernozhukov *et al.* (2010) to avoid any quantile crossing for the method proposed. We find that both the corresponding backtesting and the empirical coverage results are almost unchanged; see the on-line supplementary material for details.

7. Conclusion and discussion

In this paper, our idea of transforming the quantiles enables us first to turn a highly intractable quantile regression problem into a much simpler linear quantile regression, making the conditional quantile estimation for the GARCH model an easy job. The major novelty of this paper also lies in the hybrid nature of the estimation method proposed, which enables the conditional quantile estimator to provide a good balance between the efficiency of Gaussian QMLE and the flexibility of quantile regression. The hybrid method remedies the different drawbacks of two important approaches in the literature, i.e. FHS and CAViaR. Consequently, better forecasting performance can be achieved, as confirmed by our empirical evidence.

Our method can be extended in several directions. First, it is well known that financial time series can be so heavy tailed that $E[\eta_t^4]=\infty$ (Mikosch and Stărică, 2000; Mittnik and Paolella, 2003; Hall and Yao, 2003). For such cases, we may alternatively consider methods that are more robust than Gaussian QMLE for initial estimation of the conditional variances, e.g. the least absolute deviations estimator of Peng and Yao (2003). Second, our procedure can be applied to conditional quantile estimation for other conditional heteroscedastic models, including the asymmetric Glosten–Jagannathan–Runkle–GARCH model (Glosten *et al.*, 1993). Third, although the multivariate GARCH model has been widely used for volatility modelling of multiple asset returns (Engle and Kroner, 1995), conditional quantile estimation for the corresponding portfolio return is still an open problem. This paper offers some preliminary ideas on this, which we leave for future research.

Acknowledgements

We are deeply grateful to the Joint Editor, the Associate Editor and two referees for their valuable comments that led to the substantial improvement in the quality of this paper. The first version of this paper was written when Zheng and Zhu were doctoral students at the University of Hong Kong. Li's research was partially supported by Hong Kong Research Grant Council grant 17304617, and Xiao thanks Boston College and National Natural Science Foundation of China grant 71571110 for research support.

Appendix A: Proof sketches of theorems 1–4

The following lemma 1 establishes some important moment conditions which are used repeatedly in our proofs. All detailed proofs are provided in the on-line supplementary material.

Lemma 1. Under assumption 1, for any $\kappa>0$, there is a constant $c>0$ such that

(a) $E[\sup[\{h_t(\theta_2)/h_t(\theta_1)\}^\kappa : \|\theta_1-\theta_2\|\leqslant c,\ \theta_1,\theta_2\in\Theta]]<\infty$,
(b) $E[\sup\{\|h_t^{-1}(\theta_1)\partial h_t(\theta_2)/\partial\theta\|^\kappa : \|\theta_1-\theta_2\|\leqslant c,\ \theta_1,\theta_2\in\Theta\}]<\infty$,
(c) $E[\sup\{\|h_t^{-1}(\theta_1)\partial^2 h_t(\theta_2)/\partial\theta\partial\theta'\|^\kappa : \|\theta_1-\theta_2\|\leqslant c,\ \theta_1,\theta_2\in\Theta\}]<\infty$,
(d) $E[\sup\{|h_t^{-1}(\theta_1)\partial^3 h_t(\theta_2)/\partial\theta_i\partial\theta_k\partial\theta_l|^\kappa : \|\theta_1-\theta_2\|\leqslant c,\ \theta_1,\theta_2\in\Theta\}]<\infty$,

for all $1\leqslant i,k,l\leqslant p+q+1$, where '$\|\cdot\|$' is the norm of a matrix or column vector, defined as $\|A\|=\sqrt{\mathrm{tr}(AA')}=\sqrt{\Sigma_{i,j}|a_{ij}|^2}$.

A.1. Proof sketch of theorem 1

Let $z_t(\theta)=(1,x_{t-1}^2,\ldots,x_{t-q}^2,h_{t-1}(\theta),\ldots,h_{t-p}(\theta))'$, $\tilde{z}_t(\theta)=(1,x_{t-1}^2,\ldots,x_{t-q}^2,\tilde{h}_{t-1}(\theta),\ldots,\tilde{h}_{t-p}(\theta))'$. Write $z_t=z_t(\theta_0)$, $\breve{z}_t=\tilde{z}_t(\theta_0)$ and $\tilde{z}_t=\tilde{z}_t(\tilde{\theta}_n)$. Let $L_n(\theta)=\Sigma_{t=1}^n \tilde{h}_t^{-1}\rho_\tau(y_t-\theta'\tilde{z}_t)$, $\breve{L}_n(\theta)=\Sigma_{t=1}^n \tilde{h}_t^{-1}\rho_\tau(y_t-\theta'\breve{z}_t)$ and $\breve{e}_{t,\tau}=y_t-\theta_{\tau0}'\breve{z}_t$. Applying the identity (Knight, 1998)

$$\rho_\tau(x-y)-\rho_\tau(x)=-y\psi_\tau(x)+\int_0^y I(x,s)\mathrm{d}s,\qquad x\neq 0,\tag{A.1}$$

where $\psi_\tau(x)=\tau-I(x<0)$ and $I(x,s)=I(x\leqslant s)-I(x\leqslant 0)$, we have that, for any fixed $u\in\mathbb{R}^{p+q+1}$, $L_n(\theta_{\tau0}+n^{-1/2}u)-\breve{L}_n(\theta_{\tau0})=-L_{1n}(u)+L_{2n}(u)$, where

$$L_{1n}(u)=\sum_{t=1}^n \psi_\tau(\breve{e}_{t,\tau})\tilde{h}_t^{-1}\xi_{nt}(\tilde{\theta}_n)$$

and

$$L_{2n}(u)=\sum_{t=1}^n \tilde{h}_t^{-1}\int_0^{\xi_{nt}(\tilde{\theta}_n)} I(\breve{e}_{t,\tau},s)\mathrm{d}s,$$

with $\xi_{nt}(\theta)=(\theta_{\tau0}+n^{-1/2}u)'\tilde{z}_t(\theta)-\theta_{\tau0}'\breve{z}_t$. It is worth noting that we define $\breve{z}_t=\tilde{z}_t(\theta_0)$ deliberately to cancel the effect of the initial values in $\tilde{z}_t=\tilde{z}_t(\tilde{\theta}_n)$, which is a crucial step of our proof; see also Zheng *et al.* (2016). If we use $z_t=z_t(\theta_0)$ instead of $\breve{z}_t$, then the effect of the initial values, in the order of $C\rho^t\zeta$ by lemma S.1 in the on-line supplementary material, will remain inside the summations of $L_{1n}(u)$ and $L_{2n}(u)$, making the effect asymptotically non-negligible.

To handle $L_{1n}(u)$ and $L_{2n}(u)$, we consider the decomposition $\xi_{nt}(\tilde{\theta}_n)=\xi_{1nt}(\tilde{\theta}_n)+\xi_{2nt}(\tilde{\theta}_n)+\xi_{3nt}(\tilde{\theta}_n)$, with $\xi_{1nt}(\theta)=n^{-1/2}u'z_t+\Sigma_{j=1}^p\beta_{\tau0}^{(j)}(\theta-\theta_0)'\partial h_{t-j}(\theta_0)/\partial\theta$, $\xi_{2nt}(\theta)=n^{-1/2}\Sigma_{j=1}^p u^{(j)}\{h_{t-j}(\theta)-h_{t-j}\}+\Sigma_{j=1}^p\beta_{\tau0}^{(j)}\{h_{t-j}(\theta)-h_{t-j}-(\theta-\theta_0)'\partial h_{t-j}(\theta_0)/\partial\theta\}$ and $\xi_{3nt}(\theta)=n^{-1/2}\Sigma_{j=1}^p u^{(j)}\{\tilde{h}_{t-j}(\theta)-h_{t-j}(\theta)\}+\Sigma_{j=1}^p\beta_{\tau0}^{(j)}[\{\tilde{h}_{t-j}(\theta)-h_{t-j}(\theta)\}-\{\tilde{h}_{t-j}(\theta_0)-h_{t-j}\}]$, where $u^{(j)}$ is the $(j+q+1)$th element of u and $\beta_{\tau0}^{(j)}=b_\tau\beta_{0j}$, for $j=1,\ldots,p$. By carefully decomposing $L_{1n}(u)$ and $L_{2n}(u)$ and handling the remaining initial value effects in $\tilde{h}_t^{-1}$, as well as repeatedly applying lemmas 1 and S.1, we can show that

$$\begin{aligned}L_n(\theta_{\tau0}+n^{-1/2}u)-\breve{L}_n(\theta_{\tau0})=&-u'\{T_{1n}-b_\tau f(b_\tau)\Gamma_2\sqrt{n}(\tilde{\theta}_n-\theta_0)\}+\tfrac{1}{2}f(b_\tau)u'\Omega_2u\\&-T_{2n}+T_{3n}+o_p(1),\end{aligned}$$

where $T_{1n}=n^{-1/2}\Sigma_{t=1}^n\psi_\tau(\varepsilon_t-b_\tau)z_t/h_t$, $T_{2n}=(\tilde{\theta}_n-\theta_0)'\Sigma_{t=1}^n\psi_\tau(\varepsilon_t-b_\tau)\Sigma_{j=1}^p\pi_t^{(j)}$ and $T_{3n}=0.5f(b_\tau)(\tilde{\theta}_n-\theta_0)'\times\Sigma_{t=1}^n\Sigma_{j_1=1}^p\Sigma_{j_2=1}^p\pi_t^{(j_1)}\pi_t^{(j_2)\prime}(\tilde{\theta}_n-\theta_0)$, with $\pi_t^{(j)}=\beta_{\tau0}^{(j)}h_t^{-1}\partial h_{t-j}(\theta_0)/\partial\theta$.

Applying equation (3.3), the central limit theorem and corollary 2 in Knight (1998), together with the convexity of $L_n(\cdot)$, we have

$$\sqrt{n}(\hat{\theta}_{\tau n}-\theta_{\tau0})=\frac{\Omega_2^{-1}}{f(b_\tau)}T_{1n}-b_\tau\Omega_2^{-1}\Gamma_2\sqrt{n}(\tilde{\theta}_n-\theta_0)+o_p(1)\to_{\mathrm{d}}N(0,\Sigma_1),\tag{A.2}$$

and the proof is complete.

A.2. Proof sketch of theorem 2

Similarly to the proof of theorem 1, we first let $L_n^*(\theta)=\Sigma_{t=1}^n \omega_t \tilde{h}_t^{-1}\rho_\tau(y_t-\theta' \tilde{z}_t^*)$ and $\breve{L}_n^*(\theta)=\Sigma_{t=1}^n \omega_t \tilde{h}_t^{-1}\rho_\tau(y_t-\theta'\breve{z}_t)$. Applying identity (A.1), for any fixed $u\in\mathbb{R}^{p+q+1}$, we have $L_n^*(\theta_{\tau 0}+n^{-1/2}u)-\breve{L}_n^*(\theta_{\tau 0})=-L_{1n}^*(u)+L_{2n}^*(u)$, where

$$L_{1n}^*(u)=\sum_{t=1}^n \omega_t \psi_\tau(\breve{e}_{t,\tau})\tilde{h}_t^{-1}\xi_{nt}^*$$

and

$$L_{2n}^*(u)=\sum_{t=1}^n \omega_t \tilde{h}_t^{-1}\int_0^{\xi_{nt}^*} I(\breve{e}_{t,\tau},s)\mathrm{d}s,$$

with $\xi_{nt}^*=(\theta_{\tau 0}+n^{-1/2}u)'\tilde{z}_t^*-\theta_{\tau 0}'\breve{z}_t$. Then, by carefully dealing with decompositions of $L_{1n}^*(u)$ and $L_{2n}^*(u)$ in a way similar to that for the proof of theorem 1, we can show that

$$\begin{aligned}L_n^*(\theta_{\tau 0}+n^{-1/2}u)-\breve{L}_n^*(\theta_{\tau 0})=&-u'\{T_{1n}^*-b_\tau f(b_\tau)\Gamma_2\sqrt{n}(\tilde{\theta}_n^*-\theta_0)\}+\tfrac{1}{2}f(b_\tau)u'\Omega_2 u\\&-T_{2n}^*+T_{3n}^*+o_p^*(1),\end{aligned}$$

where $T_{1n}^*=n^{-1/2}\Sigma_{t=1}^n\omega_t\psi_\tau(\varepsilon_t-b_\tau)z_t/h_t$, $T_{2n}^*=(\tilde{\theta}_n^*-\theta_0)'\Sigma_{t=1}^n\omega_t\psi_\tau(\varepsilon_t-b_\tau)\Sigma_{j=1}^p\pi_t^{(j)}$ and $T_{3n}^*=0.5f(b_\tau)(\tilde{\theta}_n^*-\theta_0)'\Sigma_{t=1}^n\Sigma_{j_1=1}^p\Sigma_{j_2=1}^p\pi_t^{(j_1)}\pi_t^{(j_2)\prime}(\tilde{\theta}_n^*-\theta_0)$. Then, by verifying Liapounov's condition, we can show that, conditionally on $\mathcal{F}_n$, $T_{1n}^*-T_{1n}\rightarrow_{\mathrm{d}} N\{0,\tau(1-\tau)\Omega_2\}$ in probability as $n\rightarrow\infty$. By the convexity of $L_n^*(\cdot)$ and corollary 2 of Knight (1998),

$$\sqrt{n}(\hat{\theta}_{\tau n}^*-\theta_{\tau 0})=\frac{\Omega_2^{-1}}{f(b_\tau)}T_{1n}^*-b_\tau\Omega_2^{-1}\Gamma_2\sqrt{n}(\tilde{\theta}_n^*-\theta_0)+o_p^*(1),$$

which, in conjunction with equations (3.3) and (A.2), yields

$$\sqrt{n}(\hat{\theta}_{\tau n}^*-\hat{\theta}_{\tau n})=\frac{\Omega_2^{-1}}{f(b_\tau)}(T_{1n}^*-T_{1n})+\frac{b_\tau\Omega_2^{-1}\Gamma_2 J^{-1}}{\sqrt{n}}\sum_{t=1}^n(\omega_t-1)\frac{1-|\varepsilon_t|}{h_t}\frac{\partial h_t(\theta_0)}{\partial\theta}+o_p^*(1).$$

Applying Lindeberg's central limit theorem and the Cramér–Wold device, the proof is complete.

A.3. Proof sketch of theorems 3 and 4

Observe that

$$\frac{1}{\sqrt{n}}\sum_{t=k+1}^n\psi_\tau(\hat{\varepsilon}_{t,\tau})|\hat{\varepsilon}_{t-k,\tau}|=\frac{1}{\sqrt{n}}\sum_{t=k+1}^n\psi_\tau(\varepsilon_{t,\tau})|\varepsilon_{t-k,\tau}|+\sum_{t=k+1}^n\mathcal{E}_{1nt}+\sum_{t=k+1}^n\mathcal{E}_{2nt}+\sum_{t=k+1}^n\mathcal{E}_{3nt},$$

where $\mathcal{E}_{1nt}=n^{-1/2}\{\psi_\tau(\hat{\varepsilon}_{t,\tau})-\psi_\tau(\varepsilon_{t,\tau})\}|\varepsilon_{t-k,\tau}|$, $\mathcal{E}_{2nt}=n^{-1/2}\psi_\tau(\varepsilon_{t,\tau})(|\hat{\varepsilon}_{t-k,\tau}|-|\varepsilon_{t-k,\tau}|)$ and $\mathcal{E}_{3nt}=n^{-1/2}\times\{\psi_\tau(\hat{\varepsilon}_{t,\tau})-\psi_\tau(\varepsilon_{t,\tau})\}(|\hat{\varepsilon}_{t-k,\tau}|-|\varepsilon_{t-k,\tau}|)$. By Taylor series expansions, the fact that $\sqrt{n}(\hat{\theta}_{\tau n}-\theta_{\tau 0})=O_p(1)$ and $\sqrt{n}(\tilde{\theta}_n-\theta_0)=O_p(1)$, lemma 1 and the finite covering theorem, we can show that $\Sigma_{t=k+1}^n\mathcal{E}_{2nt}=o_p(1)$, $\Sigma_{t=k+1}^n\mathcal{E}_{3nt}=o_p(1)$ and

$$\sum_{t=k+1}^n\mathcal{E}_{1nt}=-f(b_\tau)\{d_{1k}'\sqrt{n}(\hat{\theta}_{\tau n}-\theta_{\tau 0})+b_\tau d_{2k}'\sqrt{n}(\tilde{\theta}_n-\theta_0)\}+o_p(1),$$

where $d_{1k}=E[h_t^{-1}|\varepsilon_{t-k,\tau}|z_t]$ and $d_{2k}=E[h_t^{-1}|\varepsilon_{t-k,\tau}|\Sigma_{j=1}^p\beta_{0j}\partial h_{t-j}(\theta_0)/\partial\theta]$. Then, by the law of large numbers we can verify that $\hat{\mu}_{a,\tau}=\mu_{a,\tau}+o_p(1)$ and $\hat{\sigma}_{a,\tau}^2=\sigma_{a,\tau}^2+o_p(1)$, which, together with equations (3.3) and (A.2) and the decomposition of $n^{-1/2}\Sigma_{t=k+1}^n\psi_\tau(\hat{\varepsilon}_{t,\tau})|\hat{\varepsilon}_{t-k,\tau}|$ above, yields $R=(\tau-\tau^2)^{-1/2}\sigma_{a,\tau}^{-1}n^{-1}\Sigma_{t=k+1}^n\varpi_t+o_p(n^{-1/2})$, where

$$\varpi_t=\psi_\tau(\varepsilon_{t,\tau})\left(\epsilon_{t-1}-D_1\Omega_2^{-1}\frac{z_t}{h_t}\right)+b_\tau f(b_\tau)(D_2-D_1\Omega_2^{-1}\Gamma_2)J^{-1}\frac{1-|\varepsilon_t|}{h_t}\frac{\partial h_t(\theta_0)}{\partial\theta},$$

with $D_i=(d_{i1},\ldots,d_{iK})'$ for $i=1$ and $i=2$. Applying the central limit theorem and the Cramér–Wold device, we have $\sqrt{n}R\rightarrow_{\mathrm{d}} N(0,\Sigma_4)$. Furthermore, by a method similar to that for the proof of theorem 8.2 in Francq and Zakoian (2010), we can show that Σ_4 is positive definite, and hence theorem 3 follows. Finally, by methods that are similar to those for the proofs of theorems 2 and 3, we have that R^*-

$R=(\tau-\tau^2)^{-1/2}\sigma_{a,\tau}^{-1}n^{-1}\Sigma_{t=k+1}^{n}(\omega_t-1)\varpi_t+o_p^*(n^{-1/2})$, and then the proof is complete similarly to that for theorem 2.

References

Berkes, I. and Horváth, L. (2004) The efficiency of the estimators of the parameters in GARCH processes. *Ann. Statist.*, **32**, 633–655.

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

Bollerslev, T. (1986) Generalized autoregressive conditional heteroscedasticity. *J. Econmetr.*, **31**, 307–327.

Bougerol, P. and Picard, N. (1992) Stationarity of GARCH processes and of some nonnegative time series. *J. Econmetr.*, **52**, 115–127.

Chernozhukov, V., Fernández-Val, I. and Galichon, A. (2010) Quantile and probability curves without crossing. *Econometrica*, **78**, 1093–1125.

Christoffersen, P. (1998) Evaluating interval forecasts. *Int. Econ. Rev.*, **39**, 841–862.

Corradi, V. and Iglesias, E. M. (2008) Bootstrap refinements for QML estimators of the GARCH(1,1) parameters. *J. Econmetr.*, **144**, 500–510.

Engle, R. F. (1982) Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica*, **50**, 987–1007.

Engle, R. F. and Kroner, K. F. (1995) Multivariate simultaneous generalized ARCH. *Econmetr. Theory*, **11**, 122–150.

Engle, R. F. and Manganelli, S. (2004) CAViaR: conditional autoregressive value at risk by regression quantiles. *J. Bus. Econ. Statist.*, **22**, 367–381.

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

Francq, C. and Zakoian, J.-M. (2010) *GARCH Models: Structure, Statistical Inference and Financial Applications*. Chichester: Wiley.

Francq, C. and Zakoïan, J.-M. (2015) Risk-parameter estimation in volatility models. *J. Econmetr.*, **184**, 158–173.

Gao, F. and Song, F. (2008) Estimation risk in GARCH VaR and ES estimates. *Econmetr. Theory*, **24**, 1404–1424.

Glosten, L. R., Jagannathan, R. and Runkle, D. E. (1993) On the relation between the expected value and the volatility of the nominal excess return on stocks. *J. Finan.*, **48**, 1779–1801.

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

Horowitz, J. L. (1998) Bootstrap methods for median regression models. *Econometrica*, **66**, 1327–1351.

Jeong, M. (2017) Residual-based GARCH bootstrap and second order asymptotic refinement. *Econmetr. Theory*, **33**, 779–790.

Jin, Z., Ying, Z. and Wei, L. J. (2001) A simple resampling method by perturbing the minimand. *Biometrika*, **88**, 381–390.

Knight, K. (1998) Limiting distributions for l_1 regression estimators under general conditions. *Ann. Statist.*, **26**, 755–770.

Koenker, R. and Bassett, G. (1978) Regression quantiles. *Econometrica*, **46**, 33–49.

Koenker, R. and Zhao, Q. (1996) Conditional quantile estimation and inference for ARCH models. *Econmetr. Theory*, **12**, 793–813.

Kuester, K., Mittnik, S. and Paolella, M. S. (2006) Value-at-risk prediction: a comparison of alternative strategies. *J. Finan. Econmetr.*, **4**, 53–89.

Lahiri, S. N. (2003) *Resampling Methods for Dependent Data*. New York: Springer.

Li, G., Leng, C. and Tsai, C.-L. (2014) A hybrid bootstrap approach to unit root tests. *J. Time Ser. Anal.*, **35**, 299–321.

Li, G. and Li, W. K. (2005) Diagnostic checking for time series models with conditional heteroscedasticity estimated by the least absolute deviation approach. *Biometrika*, **92**, 691–701.

Li, G., Li, Y. and Tsai, C.-L. (2015) Quantile correlations and quantile autoregressive modeling. *J. Am. Statist. Ass.*, **110**, 246–261.

Linton, O. (1997) An asymptotic expansion in the GARCH(1,1) model. *Econmetr. Theory*, **13**, 558–581.

Mammen, E. (1993) Bootstrap and wild bootstrap for high dimensional linear models. *Ann. Statist.*, **21**, 255–285.

Mikosch, T. and Stărică, C. (2000) Limit theory for the sample autocorrelations and extremes of a GARCH (1,1) process. *Ann. Statist.*, **28**, 1427–1451.

Mittnik, S. and Paolella, M. S. (2003) Prediction of financial downside-risk with heavy-tailed conditional distributions. In *Handbook of Heavy Tailed Distributions in Finance* (ed. S. T. Rachev), pp. 385–404. New York: Elsevier.

Morgan, J. and Reuters (1996) RiskMetrics: technical document. Morgan Guaranty Trust Company, New York.

Peng, L. and Yao, Q. (2003) Least absolute deviations estimation for ARCH and GARCH models. *Biometrika*, **90**, 967–975.

Serfling, R. (1980) *Approximation Theorems of Mathematical Statistics*. New York: Wiley.

Spierdijk, L. (2016) Confidence intervals for ARMA-GARCH value-at-risk: the case of heavy tails and skewness. *Computnl Statist. Data Anal.*, **100**, 545–559.

Taylor, S. (1986) *Modelling Financial Time Series*. New York: Wiley.

Tsay, R. S. (2010) *Analysis of Financial Time Series*, 3rd edn. New York: Wiley.

Xiao, Z. and Koenker, R. (2009) Conditional quantile estimation for generalized autoregressive conditional heteroscedasticity models. *J. Am. Statist. Ass.*, **104**, 1696–1712.

Zheng, Y., Li, Y. and Li, G. (2016) On Fréchet autoregressive conditional duration models. *J. Statist. Planng Inf.*, **175**, 51–66.

Supporting information

Additional ‘supporting information’ may be found in the on-line version of this article:

‘Supplementary material for “Hybrid quantile regression estimation for time series models with conditional heteroscedasticity”’.