Hybrid quantile regression estimation for time series models with conditional heteroscedasticity

Yao Zheng, Qianqian Zhu, Guodong Li and Zhijie Xiao
University of Hong Kong, Shanghai University of Finance and Economics and Boston College

Abstract

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 autoregressive conditional heteroscedastic (GARCH) model that has the greatest popularity, the quantile regression for this model usually gives rise to non-smooth non-convex optimization which may hinder its practical feasibility. This 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 proposed method takes advantage of the efficiency of the GARCH model in modeling the volatility globally as well as the flexibility of the quantile regression in fitting quantiles at a specific level. The asymptotic distribution of the proposed 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 proposed method is examined by simulation studies, and its advantages over existing methods are illustrated by an empirical application to Value-at-Risk forecasting.

Keywords and phrases: Bootstrap method; Conditional quantile; GARCH; Nonlinear 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) autoregressive conditional heteroscedastic (ARCH) model and Bollerslev's (1986) generalized autoregressive 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 (ES), and for such problems, quantile regression (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:

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

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

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

there are two key challenges that make the quantile regression highly intractable:

(i) The *square root* in (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.

(ii) The *recursive* form of the unobservable $\{h_t\}$ in (1.1) adds another layer of difficulty to the already complicated theoretical derivation and numerical optimization.

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

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

where $\{\varepsilon_t\}$ are *i.i.d.* innovations with mean zero. Denote the τth quantile of ε_t by $Q_{\tau,\varepsilon}$. Notice that Challenge (i) is never an issue for model (1.3), as

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

If there were no σ_{t-j} in (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 the quantile regression for sieved ARCH models and thereby circumvented Challenge (ii). Unfortunately, due to Challenge (i), easy-to-implement quantile regression procedures for Bollerslev's (1986) original GARCH model in (1.1) have been seemingly impossible.

In this paper, we tackle this open problem by applying a simple transformation to the conditional quantile in (1.2). With the square root in (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 nondecreasing 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,

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

The linearity of (1.4) enables a convenient hybrid three-step estimation procedure as follows: (1) obtain initial estimates of $\{h_t\}$ by fitting the GARCH model in (1.1) with the Gaussian quasi-maximum likelihood method; (2) estimate $Q_\tau[T(x_t)|\mathcal{F}_{t-1}]$ by a weighted linear quantile regression; and (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 $T^{-1}(x) = \sqrt{|x|} \operatorname{sgn}(x)$ is the inverse function of $T(\cdot)$.

The proposed hybrid procedure 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 modeling. 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 exist 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 CAViaR method (Engle and Manganelli, 2004). Roughly speaking, the FHS method uses the GARCH structure only for the global estimation of the volatility, but not for the quantile estimation. On the contrary, the CAViaR method focuses on the local approximation at a particular quantile level, and it adopts the GARCH-type structure only for the 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 proposed hybrid method 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 the 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 frequently encountered in practice (Engle and Manganelli, 2004). Although the CAViaR method 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 0.01, where the data are very sparse. Moreover, the computation of the CAViaR method is generally challenging. The proposed hybrid method 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 proposed estimator is complicated by the innovation density function 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 autocorrelation 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. The appendix gives proof sketches of the theorems, and due to space limitations, the detailed proofs are given in a separate supplementary file. Throughout the paper, $\rightarrow_d$ denotes the convergence in distribution, $o_p(1)$ denotes a sequence of random variables converging to zero in probability, and $o_p^*(1)$ corresponds to the bootstrap probability space.

2 The 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 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 $T(x) = x^2 \operatorname{sgn}(x)$. Then, the τth quantile of the transformed variable $y_t = T(x_t)$ conditional on $\mathcal{F}_{t-1}$ is

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

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

For $0 < \underline{w} < \overline{w}$ and $0 < \rho_0 < 1$ with $p\underline{w} < \rho_0$, define $\Theta = \{\theta : \beta_1 + \cdots + \beta_p \leqslant \rho_0,\ \underline{w} \leqslant \min(\alpha_0, \alpha_1, \ldots, \alpha_q, \beta_1, \ldots, \beta_p) \leqslant \max(\alpha_0, \alpha_1, \ldots, \alpha_q, \beta_1, \ldots, \beta_p) \leqslant \overline{w}\} \subset \mathbb{R}_+^{p+q+1}$, 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 (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 $m^{-1} \sum_{t=1}^{m} x_t^2$ for a fixed number m, say $m = 5$ in our numerical studies, and denote the resulting $h_t(\theta)$ by $\widetilde{h}_t(\theta)$; fixing the initial values will not affect our asymptotic results.

We propose the hybrid conditional quantile estimation procedure as follows.

- *Step E1 (Estimation of the global model structure).* Perform the Gaussian quasi-maximum likelihood estimation (QMLE) of model (1.1),

$$\widetilde{\theta}_n = \underset{\theta \in \Theta}{\operatorname{argmin}} \sum_{t=1}^{n} \widetilde{\ell}_t(\theta), \tag{2.3}$$

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

- *Step E2 (Quantile regression at a specific level).* Perform the weighted linear quantile regression of y_t on $\widetilde{z}_t = (1, x_{t-1}^2, \ldots, x_{t-q}^2, \widetilde{h}_{t-1}, \ldots, \widetilde{h}_{t-p})'$ at quantile level τ,

$$\widehat{\theta}_{\tau n} = \underset{\theta_\tau}{\text{argmin}} \sum_{t=1}^{n} \widetilde{h}_t^{-1} \rho_\tau(y_t - \theta_\tau' \widetilde{z}_t). \tag{2.4}$$

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

- *Step E3 (Transforming back to x_t).* Estimate the τth conditional quantile of the original observation x_t by $\widehat{Q}_\tau(x_t|\mathcal{F}_{t-1}) = T^{-1}(\widehat{\theta}_{\tau n}' \widetilde{z}_t)$, where $T^{-1}(x) = \sqrt{|x|}\,\text{sgn}(x)$.

Assumption 1. *(i) θ_0 is in the interior of Θ; (ii) η_t^2 has a non-degenerate distribution with $E\eta_t^2 = 1$; (iii) The polynomials $\sum_{i=1}^q \alpha_i x^i$ and $1 - \sum_{j=1}^p \beta_j x^j$ have no common root; (iv) $E\eta_t^4 < \infty$.*

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 is used by Francq and Zakoian (2004) to ensure the consistency and asymptotic normality of the Gaussian QMLE $\widetilde{\theta}_n$, 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 $\Gamma_i = E[h_t^{-i} z_t \sum_{j=1}^p \beta_{0j} \partial h_{t-j}(\theta_0)/\partial\theta']$ for $i = 1$ and 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 $\sqrt{n}(\widehat{\theta}_{\tau n} - \theta_{\tau 0}) \to_d N(0, \Sigma_1)$.*

The weights $\{\widetilde{h}_t^{-1}\}$ in (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

$$\breve{\theta}_{\tau n} = \underset{\theta_\tau}{\text{argmin}} \sum_{t=1}^{n} \rho_\tau(y_t - \theta_\tau' \widetilde{z}_t),$$

and as the following corollary shows, the asymptotic normality of $\breve{\theta}_{\tau n}$ 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 $\sqrt{n}(\breve{\theta}_{\tau n} - \theta_{\tau 0}) \rightarrow_d N(0, \Sigma_2)$.*

For the ARCH case, we can show that $\Sigma_2 - \Sigma_1$ is always nonnegative definite; i.e., $\widehat{\theta}_{\tau n}$ is asymptotically more efficient than $\breve{\theta}_{\tau n}$. A general comparison of Σ_1 and Σ_2 for the GARCH model is very difficult due to 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 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 (ARE) of $\widehat{\theta}_{\tau n}$ to $\breve{\theta}_{\tau n}$, defined as $\text{ARE}(\widehat{\theta}_{\tau n}, \breve{\theta}_{\tau n}) = (|\Sigma_2|/|\Sigma_1|)^{1/(p+q+1)}$, where $|\cdot|$ is the determinant of a matrix; see Serfling (1980). As shown in the supplementary material, the weighted estimator is always asymptotically more efficient than the unweighted estimator, i.e., $\text{ARE}(\widehat{\theta}_{\tau n}, \breve{\theta}_{\tau n}) > 1$, for GARCH$(1,1)$ models with different parameter values, innovation distributions and quantile levels. Therefore, we will focus on the weighted estimator $\widehat{\theta}_{\tau n}$ from now on.

Corollary 2. *If the conditions in Theorem 1 hold, then*

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

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

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

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

In practice, multiple quantile levels are often considered simultaneously, say $\tau_1 < \tau_2 < \cdots < \tau_K$. Although $\{\widehat{Q}_{\tau_k}(y_{n+1}|\mathcal{F}_n)\}_{k=1}^K$ 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 proposed hybrid method and two important approaches in the literature: the filtered historical simulation (FHS) method (Kuester et al., 2006) and the CAViaR method (Engle and Manganelli, 2004).

We first consider the FHS method. Notice 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

$$\widetilde{b}_{\tau n} = \underset{b}{\operatorname{argmin}} \sum_{t=1}^{n} \widetilde{h}_t^{-1} \rho_\tau(y_t - b\widetilde{h}_t). \tag{2.7}$$

It is not difficult to see that $\widetilde{b}_{\tau n}$ is just the τth empirical quantile of $\{y_t/\widetilde{h}_t\}$. Thus, the corresponding procedure, with a simplified second stage estimation, reduces to the FHS method, with the estimates $\widehat{Q}_\tau(y_t|\mathcal{F}_{t-1}) = \widetilde{b}_{\tau n}\widetilde{h}_t = \widetilde{\theta}_{\tau n}'\widetilde{z}_t$, where $\widetilde{\theta}_{\tau n} = \widetilde{b}_{\tau n}\widetilde{\theta}_n$ 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 supplementary material also demonstrate that the FHS method always has much larger biases than the proposed method.

On the other hand, applying the CAViaR method of Engle and Manganelli (2004) to the transformed observations y_t by assuming the linear form in (2.1), we have

$$\widehat{\vartheta}_{\tau n} = \underset{\vartheta}{\operatorname{argmin}} \sum_{t=1}^{n} \rho_\tau[y_t - \vartheta' v_t(\vartheta)], \tag{2.8}$$

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 $\widehat{\theta}_{\tau n}$ and the FHS estimator $\widetilde{\theta}_{\tau n}$ which both converge to $\theta_{\tau 0} = b_\tau\theta_0$, the CAViaR estimator $\widehat{\vartheta}_{\tau n}$ converges to $\vartheta_{\tau 0} := (b_\tau\alpha_{00}, b_\tau\alpha_{01}, \ldots, b_\tau\alpha_{0q}, \beta_{01}, \ldots, \beta_{0p})'$. Notice that this approach will lead to the unweighted estimator $\breve{\theta}_{\tau n}$ in the previous subsection if we first obtain initial estimates of $\{q_t(\vartheta)\}$, and hence those of $v_t(\vartheta)$ in (2.8), by replacing ϑ with the more efficient Gaussian QMLE $\widetilde{\theta}_n$, and then perform the quantile regression in (2.8). As a result, the CAViaR method is even less efficient than the unweighted method in the previous subsection, although it enjoys greater flexibility than the FHS method since it imposes a structure at only the quantile level τ. Moreover, the computation of the CAViaR method is generally challenging, which actually requires grid search.

We may interpret the proposed method as a hybrid version of the FHS and CAViaR methods. 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 $\widehat{\theta}_{\tau n}$ may be less efficient than the FHS estimator $\widetilde{\theta}_{\tau n}$. 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 $\sqrt{n}(\widetilde{\theta}_{\tau n} - \theta_{\tau 0}) \to_d N(0, \Sigma_3)$; 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 nonnegative 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, similar to our discussion on the unweighted estimator in the previous subsection, we have computed the ARE of the proposed estimator $\widehat{\theta}_{\tau n}$ to the FHS estimator $\widetilde{\theta}_{\tau n}$ for GARCH(1, 1) models for different parameter settings, innovation distributions and quantile levels. As expected, the FHS estimator $\widetilde{\theta}_{\tau n}$ is asymptotically more efficient in general, while the proposed estimator $\widehat{\theta}_{\tau n}$ can be asymptotically more efficient when $\{\eta_t\}$ become more heavy-tailed; see the 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 $\widehat{\theta}_{\tau n}$, 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 $\widetilde{\theta}_n$ affects the asymptotic distribution of the proposed estimator $\widehat{\theta}_{\tau n}$ through the relationship $\sqrt{n}(\widehat{\theta}_{\tau n} - \theta_{\tau 0}) = \Omega_2^{-1}T_{1n}/f(b_\tau) - b_\tau\Omega_2^{-1}\Gamma_2\sqrt{n}(\widetilde{\theta}_n - \theta_0) + o_p(1)$, where $T_{1n} = n^{-1/2}\sum_{t=1}^n \psi_\tau(\varepsilon_t - b_\tau)z_t/h_t$, with $\psi_\tau(x) = \tau - I(x < 0)$. Apparently, the random-weighting bootstrap should be incorporated in both Steps E1 and E2, leading to the following bootstrapping procedure:

- *Step B1.* Perform the randomly weighted Gaussian QMLE,

$$\widetilde{\theta}_n^* = \underset{\theta \in \Theta}{\operatorname{argmin}} \sum_{t=1}^n \omega_t \widetilde{\ell}_t(\theta), \tag{3.1}$$

where $\{\omega_t\}$ are *i.i.d.* non-negative random weights with mean and variance both equal to one, and then compute the initial estimates of $\{h_t\}$ as $\widetilde{h}_t^* = \widetilde{h}_t(\widetilde{\theta}_n^*)$.

- *Step B2.* Perform the randomly weighted quantile regression,

$$\widehat{\theta}_{\tau n}^{*} = \underset{\theta_\tau}{\text{argmin}} \sum_{t=1}^{n} \omega_t \widetilde{h}_t^{-1} \rho_\tau (y_t - \theta_\tau' \widetilde{z}_t^{*}), \tag{3.2}$$

where $\widetilde{z}_t^{*} = (1, x_{t-1}^2, \ldots, x_{t-q}^2, \widetilde{h}_{t-1}^{*}, \ldots, \widetilde{h}_{t-p}^{*})'$.

- *Step B3.* Calculate the conditional quantile estimate, $\widehat{Q}_\tau^{*}(x_t|\mathcal{F}_{t-1}) = T^{-1}(\widehat{\theta}_{\tau n}^{*\prime} \widetilde{z}_t^{*})$.

The purpose of the bootstrapping procedure is to avoid estimating the density $f(b_\tau)$ involved in the asymptotic covariance matirx Σ_1. Observe that no density actually appears in the asymptotic covariance matrix of the Gaussian QMLE $\widetilde{\theta}_n$. This motivates us to replace the optimization in Step B1 with a simple sample averaging. Notice that

$$\sqrt{n}(\widetilde{\theta}_n^{*} - \widetilde{\theta}_n) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} (\omega_t - 1)\xi_t + o_{p^*}(1) \quad \text{and} \quad \sqrt{n}(\widetilde{\theta}_n - \theta_0) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \xi_t + o_p(1), \tag{3.3}$$

where $\xi_t = J^{-1}(|y_t|/h_t - 1)h_t^{-1}[\partial h_t(\theta_0)/\partial \theta]$ and $\widetilde{\theta}_n^{*}$ is defined as in (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 $\widetilde{J} = n^{-1}\sum_{t=1}^{n} \widetilde{h}_t^{-2}[\partial \widetilde{h}_t(\widetilde{\theta}_n)/\partial \theta][\partial \widetilde{h}_t(\widetilde{\theta}_n)/\partial \theta']$. Therefore, Step B1 can be replaced by the following:

- *Step B1′.* Calculate the estimator $\widetilde{\theta}_n^{*}$ by

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

Combining Steps B1′, B2 and B3, we have a mixed bootstrapping procedure.

Assumption 3. *The random weights $\{\omega_t\}$ are i.i.d. non-negative random variables with mean and variance both equal to one, 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, conditional on $\mathcal{F}_n$, $\sqrt{n}(\widehat{\theta}_{\tau n}^{*} - \widehat{\theta}_{\tau n}) \to_d N(0, \Sigma_1)$ 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*

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

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

By Corollaries 2 and 3, along with the asymptotic results for $\widetilde{\theta}_n^{*}$ and $\widehat{\theta}_{\tau n}^{*}$ 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 based on the bootstrap sample $\{\widehat{Q}^*_\tau(x_{n+1}|\mathcal{F}_n)\}$, where $\widehat{Q}^*_\tau(x_{n+1}|\mathcal{F}_n) = T^{-1}[\widehat{Q}^*_\tau(y_{n+1}|\mathcal{F}_n)]$; see also Spierdijk (2016).

The first-order validity of the proposed mixed bootstrapping procedure 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 the quantile regression is employed, due to the non-smoothness of the loss function $\rho_\tau(\cdot)$, it will be very difficult to attain the second-order correctness for the bootstrapping procedure; see also Horowitz (1998). Notice also that the $o_p(1)$ term in (3.3) plays a non-negligible role in the Edgeworth expansion of $\sqrt{n}(\widetilde{\theta}_n - \theta_0)$ (Linton, 1997), but is ignored by $\widetilde{\theta}^*_n$ in Step B1′. Hence, the second-order correctness has already been lost when we use the much faster sample-averaging method in Step B1′ to replace the optimization in Step B1. However, the sacrifice is worthwhile, as the second-order correctness is anyway unachievable due to the non-smooth objective function in Step B2. Actually, in the literature, bootstrap methods with the second-order correctness are still limited to the GARCH$(1,1)$ model and unavailable for the general GARCH model (Corradi and Iglesias, 2008; Jeong, 2017).

4 Diagnostic checking for conditional quantiles

Based on 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 quantile autocorrelation function (QACF) of $\{\varepsilon_{t,\tau}\}$ at lag k as

$$\rho_{k,\tau} = \text{qcor}_\tau\{\varepsilon_{t,\tau}, |\varepsilon_{t-k,\tau}|\} = \frac{E\{\psi_\tau(\varepsilon_{t,\tau})|\varepsilon_{t-k,\tau}|\}}{\sqrt{(\tau-\tau^2)\sigma^2_{a,\tau}}}, \quad k = 1, 2, \ldots,$$

where $\sigma^2_{a,\tau} = \text{var}(|\varepsilon_{t,\tau}|) = E(|\varepsilon_{t,\tau}| - \mu_{a,\tau})^2$, 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 (1.2), then $\mathrm{E}[\psi_\tau(\varepsilon_{t,\tau})|\mathcal{F}_{t-1}]$ is zero, and so is $\rho_{k,\tau}$ for any $k \geqslant 1$.

Accordingly, let $\widehat{\varepsilon}_{t,\tau} = \widetilde{h}_t^{-1}(y_t - \widehat{\theta}'_{\tau n}\widetilde{z}_t)$, and then the corresponding residual QACF at lag k can be calculated as $r_{k,\tau} = (\tau-\tau^2)^{-1/2}\widehat{\sigma}^{-1}_{a,\tau} n^{-1}\sum_{t=k+1}^n \psi_\tau(\widehat{\varepsilon}_{t,\tau})|\widehat{\varepsilon}_{t-k,\tau}|$, where $\widehat{\sigma}^2_{a,\tau} = n^{-1}\sum_{t=1}^n(|\widehat{\varepsilon}_{t,\tau}| - \widehat{\mu}_{a,\tau})^2$, with $\widehat{\mu}_{a,\tau} = n^{-1}\sum_{t=1}^n|\widehat{\varepsilon}_{t,\tau}|$. 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 $D_1 = E(h_t^{-1}\epsilon_{t-1}z_t')$, $D_2 = E[h_t^{-1}\epsilon_{t-1}\sum_{j=1}^p \beta_{0j}\partial h_{t-j}(\theta_0)/\partial\theta']$, 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

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

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

Theorem 3 implies that the portmanteau test statistic $Q(K) = nR'\widehat{\Sigma}_4^{-1}R$ converges to the χ^2 distribution with K degrees of freedom as $n \to \infty$, where $\widehat{\Sigma}_4$ is a consistent estimator of Σ_4. Notice that, even for the ARCH case, the asymptotic covariance matrix $\Sigma_4 = \sigma_{a,\tau}^{-2}(\Xi - D_1J^{-1}D_1')$ 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 $\widehat{\varepsilon}_{t,\tau}^* = \widetilde{h}_t^{-1}(y_t - \widehat{\theta}_{\tau n}^{*\prime}\widetilde{z}_t^*)$, $r_{k,\tau}^* = (\tau-\tau^2)^{-1/2}\widehat{\sigma}_{a,\tau}^{-1}n^{-1}\sum_{t=k+1}^n \omega_t\psi_\tau(\widehat{\varepsilon}_{t,\tau}^*)|\widehat{\varepsilon}_{t-k,\tau}^*|$, and $R^* = (r_{1,\tau}^*, \ldots, r_{K,\tau}^*)'$.

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

In Step B3 in the previous section, we can calculate R^* and $T^{(1)} = \sqrt{n}(R^* - R)$. Then, repeating Steps B1$'$ and B2 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 $\sqrt{n}r_{k,\tau}$ 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 proposed estimation, bootstrapping and diagnostic checking procedures.

In the first experiment, we focus on the proposed estimator $\widehat{\theta}_{\tau n}$ 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's t_5 distribution with unit variance. We consider three sample sizes, $n = 500$, 1000 and 2000, with 1000 replications generated for each sample size, and two quantile levels, $\tau = 0.05$ and 0.1. Four distributions for the random weights $\{\omega_t\}$ in the bootstrapping 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 $\widehat{\theta}_{\tau n}$ 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: (1) the biases are all small; (2) as n or τ increases, the bias and standard deviations decrease, and the ASDs become closer to the corresponding ESDs; (3) the performance of the bootstrapping approximation is insensitive to the choice of random weights; (4) the ASDs appear to be closer to the corresponding ESDs when $\{\eta_t\}$ are normal than when they follow the Student's t_5 distribution; and (5) when $\tau = 0.05$, the standard deviations for the normal distribution are smaller than those for the Student's t_5 distribution, while 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 performance 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. Due to space limitations, we only present results for W_1 from now on, and the results for W_2, W_3 and W_4 are provided in the 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 $k = 2, 4$ and 6. Findings (1) and (2) in the previous experiment are also observed in this table. Furthermore, we have repeated the first two experiments 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

Table 1: Bias (×10), ESD (×10) and ASD (×10) for $\widehat{\theta}_{\tau n}$ at $\tau = 0.05$ or 0.1, for normal or Student's t_5-distributed innovations, where ASD_i corresponds to random weight W_i for $i = 1, 2, 3$ and 4, and α_0, α_1 and β_1 represent corresponding elements of $\widehat{\theta}_{\tau n}$.

		Normal distribution						Student's t_5 distribution					
n		Bias	ESD	ASD_1	ASD_2	ASD_3	ASD_4	Bias	ESD	ASD_1	ASD_2	ASD_3	ASD_4
							$\tau = 0.05$						
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

Table 2: Bias (×100), ESD (×100) and ASD (×100) for the residual QACF $r_{k,\tau}$ at $\tau = 0.05$ or 0.1 and $k = 2, 4$ or 6, for normal or Student's t_5-distributed innovations.

		Normal distribution						Student's t_5 distribution					
		$\tau = 0.05$			$\tau = 0.1$			$\tau = 0.05$			$\tau = 0.1$		
n	k	Bias	ESD	ASD	Bias	ESD	ASD	Bias	ESD	ASD	Bias	ESD	ASD
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% significance level, for normal or Student's t_5-distributed innovations and $d = 0$, 0.3 or 0.6.

	Normal distribution						Student's t_5 distribution					
	$\tau = 0.05$			$\tau = 0.1$			$\tau = 0.05$			$\tau = 0.1$		
n	0.0	0.3	0.6	0.0	0.3	0.6	0.0	0.3	0.6	0.0	0.3	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

$Q(K)$. 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 or 0.6. We conduct the 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's t_5 distribution. Note that when τ gets closer to zero, 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 analyze the daily log returns of three stock market indexes from January 2, 2008 to June 30, 2016: the S&P 500 index, the Dow 30 index, and the Hang Seng Index (HSI). The sample sizes are $n = 2139$, 2139 and 2130, respectively.

We begin by illustrating the proposed method with the S&P 500 data for $\tau = 0.05$, i.e., the one-day 5% VaR; see Figure 1 for the time plot of the log returns $\{x_t\}$. By the proposed estimation procedure, the initial estimates of $\{h_t\}$ are calculated by $\widetilde{h}_t = 2.646 \times 10^{-6} + 0.126x_{t-1}^2 + 0.858\widetilde{h}_{t-1}$, and the fitted conditional quantile function is

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

Figure 1 shows that the residual QACF only falls 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 proposed diagnostic checking procedure, the p-values of the portmanteau test $Q(K)$

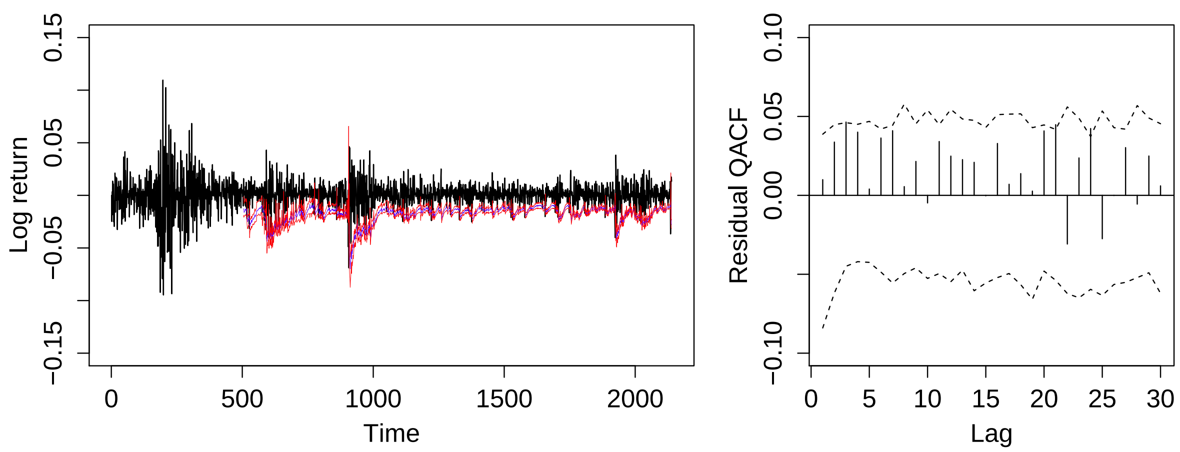

Figure 1: Left: Time plot of daily log returns (black line) of S&P 500 from January 2, 2008 to June 30, 2016 and rolling forecasts of the conditional quantiles (blue line) at $\tau = 0.05$ from January 4, 2010 to June 30, 2016 with corresponding 95% confidence bounds (red lines). Right: Residual QACF of the fitted GARCH model at $\tau = 0.05$, with corresponding 95% confidence bounds.

are all larger than 0.257 for $K = 6$, 12, 18, 24 and 30, which suggests the adequacy of the fitted conditional quantiles.

Next we examine the forecasting performance of the proposed method for all stock market indexes using the following rolling procedure: first, conduct the estimation using the first two years' data and compute the conditional quantile forecast for the next trading day, i.e., the forecast of $Q_\tau(x_{n+1}|\mathcal{F}_n)$; then, advance the forecasting origin by one to include one more observation in the estimation subsample, and repeat the foregoing procedure until the end of the sample is reached. See Figure 1 for an illustration of the rolling forecasts at $\tau = 0.05$ for the S&P 500 data, where the corresponding 95% confidence intervals are constructed by the proposed bootstrapping procedure.

To compare the forecasting performance of the proposed method with existing conditional quantile estimation methods, we also conduct the rolling forecasting for the FHS method discussed in Section 2.2 and four other methods which we call XK_1, XK_2, CAViaR and RiskM, respectively, in what follows. In particular, XK_1 and XK_2 are adapted versions of "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 Step E1 of the proposed procedure. For XK_1, the initial estimates of $\{h_t\}$ are obtained by a linear quantile regression at the quantile level τ using the sieve approximation, $h_t = \gamma_0 + \sum_{j=1}^m \gamma_j x_{t-j}^2$, 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 the 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 these authors for the grid-search optimization and the same settings of initial values for the optimization as in their paper. 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 *i.i.d.* standard normal; see Morgan and Reuters (1996) and Tsay (2010).

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

(i) the minimum of the p-values of the two VaR backtests, the likelihood ratio test for correct conditional coverage (CC) in Christoffersen (1998) and the dynamic quantile (DQ) test in Engle and Manganelli (2004);

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

		Minimum p-value of VaR backtests						Empirical coverage error					
	τ	Hybrid	FHS	XK_1	XK_2	CAViaR	RiskM	Hybrid	FHS	XK_1	XK_2	CAViaR	RiskM
S&P 500	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

(ii) 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 DQ 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 CC and DQ 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&P 500 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 proposed hybrid method 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 proposed hybrid method performs well except 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 proposed method and the FHS method dominate the other competitors. Indeed, for the three data sets at the six quantile levels, among all methods, the proposed method has the largest number of cases where the minimum p-value exceeds 0.2, while the FHS method has smallest number of cases where the minimum p-value is less than 0.05.

To determine whether the proposed method 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 totally 18 cases) where the absolute value of its corresponding empirical coverage error is the smallest and second smallest among all methods. From the right panel of Table 4, the results are 9 and 6 for the proposed method, and 4 and 5 for the FHS method, respectively. For the other competitors, the numbers are all much smaller. In the 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 proposed method does have clearly better performance than the FHS method.

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

7 Conclusion and discussion

In this paper, our idea of transforming the quantiles allows us to first 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 proposed estimation method, which enables the conditional quantile estimator to provide a good balance between the efficiency of the Gaussian QMLE and the flexibility of the quantile regression. The proposed hybrid method remedies the different drawbacks of two important approaches in the literature, i.e., the FHS and CAViaR methods. 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 more robust than the 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 the conditional quantile estimation for other conditional heteroscedastic models, including the asymmetric GJR-GARCH model (Glosten et al., 1993). Third, although the multivariate GARCH model has been widely used for volatility modeling of multiple asset returns (Engle and Kroner, 1995), the 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 anonymous 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 PhD students at the University of Hong Kong. Li's research was partially supported by the Hong Kong Research Grant Council grant 17304617, and Xiao thanks Boston College and NSFC grant 71571110 for research support.

Appendix: Proof sketches of Theorems 1–4

The following Lemma A.1 establishes some important moment conditions which are used repeatedly in our proofs. All detailed proofs are provided in the supplementary material.

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

(i) $E\sup\{[h_t(\theta_2)/h_t(\theta_1)]^\kappa : \|\theta_1-\theta_2\| \leqslant c,\ \theta_1,\theta_2 \in \Theta\} < \infty$,

(ii) $E\sup\left\{\left\|h_t^{-1}(\theta_1)\partial h_t(\theta_2)/\partial\theta\right\|^\kappa : \|\theta_1-\theta_2\| \leqslant c,\ \theta_1,\theta_2 \in \Theta\right\} < \infty$,

(iii) $E\sup\left\{\left\|h_t^{-1}(\theta_1)\partial^2 h_t(\theta_2)/\partial\theta\partial\theta'\right\|^\kappa : \|\theta_1-\theta_2\| \leqslant c,\ \theta_1,\theta_2 \in \Theta\right\} < \infty$,

(iv) $E\sup\left\{\left|h_t^{-1}(\theta_1)\partial^3 h_t(\theta_2)/\partial\theta_i\partial\theta_k\partial\theta_\ell\right|^\kappa : \|\theta_1-\theta_2\| \leqslant c,\ \theta_1,\theta_2 \in \Theta\right\} < \infty$,

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

Proof sketch of Theorem 1. Let $z_t(\theta) = (1, x_{t-1}^2, \ldots, x_{t-q}^2, h_{t-1}(\theta), \ldots, h_{t-p}(\theta))'$, $\widetilde{z}_t(\theta) = (1, x_{t-1}^2, \ldots, x_{t-q}^2, \widetilde{h}_{t-1}(\theta), \ldots, \widetilde{h}_{t-p}(\theta))'$. Write $z_t = z_t(\theta_0)$, $\breve{z}_t = \widetilde{z}_t(\theta_0)$, and $\widetilde{z}_t = \widetilde{z}_t(\widetilde{\theta}_n)$. Let $L_n(\theta) = \sum_{t=1}^n \widetilde{h}_t^{-1}\rho_\tau(y_t - \theta'\widetilde{z}_t)$, $\breve{L}_n(\theta) = \sum_{t=1}^n \widetilde{h}_t^{-1}\rho_\tau(y_t - \theta'\breve{z}_t)$, and $\breve{e}_{t,\tau} = y_t - \theta_{\tau 0}'\breve{z}_t$. Applying the identity (Knight, 1998),

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

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 $u \in \mathbb{R}^{p+q+1}$, $L_n(\theta_{\tau 0} + n^{-1/2}u) - \breve{L}_n(\theta_{\tau 0}) = -L_{1n}(u) + L_{2n}(u)$, where

$$L_{1n}(u) = \sum_{t=1}^n \psi_\tau(\breve{e}_{t,\tau})\widetilde{h}_t^{-1}\xi_{nt}(\widetilde{\theta}_n) \quad \text{and} \quad L_{2n}(u) = \sum_{t=1}^n \widetilde{h}_t^{-1}\int_0^{\xi_{nt}(\widetilde{\theta}_n)} I(\breve{e}_{t,\tau}, s)ds,$$

with $\xi_{nt}(\theta) = (\theta_{\tau 0} + n^{-1/2}u)'\widetilde{z}_t(\theta) - \theta_{\tau 0}'\breve{z}_t$. It is worth noting that we define $\breve{z}_t = \widetilde{z}_t(\theta_0)$ deliberately to cancel the effect of the initial values in $\widetilde{z}_t = \widetilde{z}_t(\widetilde{\theta}_n)$, 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 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, $\xi_{nt}(\widetilde{\theta}_n) = \xi_{1nt}(\widetilde{\theta}_n) + \xi_{2nt}(\widetilde{\theta}_n) + \xi_{3nt}(\widetilde{\theta}_n)$, with $\xi_{1nt}(\theta) = n^{-1/2}u'z_t + \sum_{j=1}^{p} \beta_{\tau 0}^{(j)}(\theta - \theta_0)'\partial h_{t-j}(\theta_0)/\partial\theta$, $\xi_{2nt}(\theta) = n^{-1/2}\sum_{j=1}^{p} u^{(j)}[h_{t-j}(\theta) - h_{t-j}] + \sum_{j=1}^{p} \beta_{\tau 0}^{(j)}[h_{t-j}(\theta) - h_{t-j} - (\theta - \theta_0)'\partial h_{t-j}(\theta_0)/\partial\theta]$, and $\xi_{3nt}(\theta) = n^{-1/2}\sum_{j=1}^{p} u^{(j)}[\widetilde{h}_{t-j}(\theta) - h_{t-j}(\theta)] + \sum_{j=1}^{p} \beta_{\tau 0}^{(j)}\{[\widetilde{h}_{t-j}(\theta) - h_{t-j}(\theta)] - [\widetilde{h}_{t-j}(\theta_0) - h_{t-j}]\}$, where $u^{(j)}$ is the $(j + q + 1)$-th element of u and $\beta_{\tau 0}^{(j)} = b_\tau \beta_{0j}$, for $j = 1, \ldots, p$. By carefully decomposing $L_{1n}(u)$ and $L_{2n}(u)$ and handling the remaining initial value effects in $\widetilde{h}_t^{-1}$, as well as repeatedly applying Lemmas A.1 and S.1, we can show that

$$L_n(\theta_{\tau 0} + n^{-1/2}u) - \breve{L}_n(\theta_{\tau 0}) = -u'\left[T_{1n} - b_\tau f(b_\tau)\Gamma_2\sqrt{n}(\widetilde{\theta}_n - \theta_0)\right] + \frac{1}{2}f(b_\tau)u'\Omega_2 u$$
$$- T_{2n} + T_{3n} + o_p(1),$$

where $T_{1n} = n^{-1/2}\sum_{t=1}^{n} \psi_\tau(\varepsilon_t - b_\tau)z_t/h_t$, $T_{2n} = (\widetilde{\theta}_n - \theta_0)'\sum_{t=1}^{n} \psi_\tau(\varepsilon_t - b_\tau)\sum_{j=1}^{p} \pi_t^{(j)}$, $T_{3n} = 0.5f(b_\tau)(\widetilde{\theta}_n - \theta_0)'\sum_{t=1}^{n}\sum_{j_1=1}^{p}\sum_{j_2=1}^{p} \pi_t^{(j_1)}\pi_t^{(j_2)\prime}(\widetilde{\theta}_n - \theta_0)$, with $\pi_t^{(j)} = \beta_{\tau 0}^{(j)} h_t^{-1}\partial h_{t-j}(\theta_0)/\partial\theta$.

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

$$\sqrt{n}(\widehat{\theta}_{\tau n} - \theta_{\tau 0}) = \frac{\Omega_2^{-1}}{f(b_\tau)}T_{1n} - b_\tau\Omega_2^{-1}\Gamma_2\sqrt{n}(\widetilde{\theta}_n - \theta_0) + o_p(1) \to_d N(0, \Sigma_1), \qquad \text{(A.2)}$$

and the proof is complete. □

Proof sketch of Theorem 2. Similarly to the proof of Theorem 1, we first let $L_n^*(\theta) = \sum_{t=1}^{n} \omega_t \widetilde{h}_t^{-1}\rho_\tau(y_t - \theta'\widetilde{z}_t^*)$ and $\breve{L}_n^*(\theta) = \sum_{t=1}^{n} \omega_t \widetilde{h}_t^{-1}\rho_\tau(y_t - \theta'\breve{z}_t)$. 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

$$L_{1n}^*(u) = \sum_{t=1}^{n} \omega_t\psi_\tau(\breve{e}_{t,\tau})\widetilde{h}_t^{-1}\xi_{nt}^* \quad \text{and} \quad L_{2n}^*(u) = \sum_{t=1}^{n} \omega_t\widetilde{h}_t^{-1}\int_0^{\xi_{nt}^*} I(\breve{e}_{t,\tau}, s)ds,$$

with $\xi_{nt}^* = (\theta_{\tau 0} + n^{-1/2}u)'\widetilde{z}_t^* - \theta_{\tau 0}'\breve{z}_t$. 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

$$L_n^*(\theta_{\tau 0} + n^{-1/2}u) - \breve{L}_n^*(\theta_{\tau 0}) = -u'\left[T_{1n}^* - b_\tau f(b_\tau)\Gamma_2\sqrt{n}(\widetilde{\theta}_n^* - \theta_0)\right] + \frac{1}{2}f(b_\tau)u'\Omega_2 u$$
$$- T_{2n}^* + T_{3n}^* + o_p^*(1),$$

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

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

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

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

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

Proof sketch of Theorems 3 and 4. Observe that

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

where $\mathcal{E}_{1nt} = n^{-1/2}[\psi_\tau(\widehat{\varepsilon}_{t,\tau}) - \psi_\tau(\varepsilon_{t,\tau})]|\varepsilon_{t-k,\tau}|$, $\mathcal{E}_{2nt} = n^{-1/2}\psi_\tau(\varepsilon_{t,\tau})(|\widehat{\varepsilon}_{t-k,\tau}| - |\varepsilon_{t-k,\tau}|)$, and $\mathcal{E}_{3nt} = n^{-1/2}[\psi_\tau(\widehat{\varepsilon}_{t,\tau}) - \psi_\tau(\varepsilon_{t,\tau})](|\widehat{\varepsilon}_{t-k,\tau}| - |\varepsilon_{t-k,\tau}|)$. By Taylor expansions, the fact that $\sqrt{n}(\widehat{\theta}_{\tau n} - \theta_{\tau 0}) = O_p(1)$ and $\sqrt{n}(\widetilde{\theta}_n - \theta_0) = O_p(1)$, Lemma A.1, and the finite covering theorem, we can show that $\sum_{t=k+1}^{n}\mathcal{E}_{2nt} = o_p(1)$, $\sum_{t=k+1}^{n}\mathcal{E}_{3nt} = o_p(1)$, and

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

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

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

with $D_i = (d_{i1}, \ldots, d_{iK})'$ for $i = 1$ and 2. Applying the central limit theorem and the Cramér-Wold device, we have $\sqrt{n}R \to_d N(0, \Sigma_4)$. 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 similar to those for the proofs of Theorems 2 and 3, we have $R^* - R = (\tau - \tau^2)^{-1/2}\sigma_{a,\tau}^{-1}n^{-1}\sum_{t=k+1}^{n}(\omega_t - 1)\varpi_t + o_p^*(n^{-1/2})$, and then the proof is complete similarly to Theorem 2. □

References

Berkes, I. and Horváth, L. (2004). The efficiency of the estimators of the parameters in GARCH processes. *The Annals of Statistics*, 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. *Journal of Econometrics*, 31:307–327.

Bougerol, P. and Picard, N. (1992). Stationarity of GARCH processes and of some nonnegative time series. *Journal of Econometrics*, 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. *International Economic Review*, 39:841–862.

Corradi, V. and Iglesias, E. M. (2008). Bootstrap refinements for QML estimators of the GARCH(1,1) parameters. *Journal of Econometrics*, 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. *Econometric Theory*, 11:122–150.

Engle, R. F. and Manganelli, S. (2004). CAViaR: conditional autoregressive value at risk by regression quantiles. *Journal of Business and Economic Statistics*, 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*. John Wiley & Sons, Chichester, UK.

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

Gao, F. and Song, F. (2008). Estimation risk in GARCH VaR and ES estimates. *Econometric 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. *The Journal of Finance*, 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. *Econometric 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. *The Annals of Statistics*, 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. *Econometric Theory*, 12:793–813.

Kuester, K., Mittnik, S., and Paolella, M. S. (2006). Value-at-risk prediction: A comparison of alternative strategies. *Journal of Financial Econometrics*, 4:53–89.

Lahiri, S. N. (2003). *Resampling methods for dependent data*. Springer, New York.

Li, G., Leng, C., and Tsai, C.-L. (2014). A hybrid bootstrap approach to unit root tests. *Journal of Time Series Analysis*, 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. *Journal of the American Statistical Association*, 110:246–261.

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

Mammen, E. (1993). Bootstrap and wild bootstrap for high dimensional linear models. *The Annals of Statistics*, 21:255–285.

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

Mittnik, S. and Paolella, M. S. (2003). Prediction of financial downside-risk with heavy-tailed conditional distributions. In Rachev, S. T., editor, *Handbook of Heavy Tailed Distributions in Finance*, pages 385–404. 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.* John Wiley & Sons, New York.

Spierdijk, L. (2016). Confidence intervals for ARMA-GARCH value-at-risk: The case of heavy tails and skewness. *Computational Statistics and Data Analysis*, 100:545–559.

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

Tsay, R. S. (2010). *Analysis of Financial Time Series.* John Wiley & Sons, 3rd edition.

Xiao, Z. and Koenker, R. (2009). Conditional quantile estimation for generalized autoregressive conditional heteroscedasticity models. *Journal of the American Statistical Association*, 104:1696–1712.

Zheng, Y., Li, Y., and Li, G. (2016). On Fréchet autoregressive conditional duration models. *Journal of Statistical Planning and Inference*, 175:51–66.