Quantile autoregressive conditional heteroscedasticity

Qianqian Zhu[a], Songhua Tan[a], Yao Zheng[b] and Guodong Li[c]
[a]*Shanghai University of Finance and Economics,*
[b]*University of Connecticut and* [c]*University of Hong Kong*

Abstract

This paper proposes a novel conditional heteroscedastic time series model by applying the framework of quantile regression processes to the ARCH(∞) form of the GARCH model. This model can provide varying structures for conditional quantiles of the time series across different quantile levels, while including the commonly used GARCH model as a special case. The strict stationarity of the model is discussed. For robustness against heavy-tailed distributions, a self-weighted quantile regression (QR) estimator is proposed. While QR performs satisfactorily at intermediate quantile levels, its accuracy deteriorates at high quantile levels due to data scarcity. As a remedy, a self-weighted composite quantile regression (CQR) estimator is further introduced and, based on an approximate GARCH model with a flexible Tukey-lambda distribution for the innovations, we can extrapolate the high quantile levels by borrowing information from intermediate ones. Asymptotic properties for the proposed estimators are established. Simulation experiments are carried out to access the finite sample performance of the proposed methods, and an empirical example is presented to illustrate the usefulness of the new model.

Key words: Composite quantile regression; Conditional quantile estimation; GARCH model; Strict stationarity; Tukey-lambda distribution.

1 Introduction

Since the appearance of autoregressive conditional heteroscedastic (ARCH) (Engle, 1982) and generalized ARCH (GARCH) models (Bollerslev, 1986), GARCH-type models have become popular and powerful tools to capture the volatility of financial time series; see Francq and Zakoian (2010) for an overview. Volatility modeling plays an important role in financial risk management. In particular, it is a key ingredient for the calculation of quantile-based risk measures such as the value-at-risk (VaR) and expected shortfall. As estimating these measures is essentially a quantile estimation problem (Artzner et al., 1999; Wu and Xiao, 2002; Francq and Zakoian, 2015), considerable research has been devoted to the development of quantile regression (QR) methods for GARCH-type models, such as Taylor's (2008) linear ARCH (Koenker and Zhao, 1996) and linear GARCH models (Xiao and Koenker, 2009), Bollerslev's (1986) GARCH model (Lee and Noh, 2013; Zheng et al., 2018), and asymmetric power GARCH models (Wang et al., 2022).

A common feature of the above research is that the global structure of the volatility process is captured by a parametric GARCH-type model with distribution-free innovations. This implies that the conditional quantile process will be the product of the volatility process and the quantile of the innovation. Consider the following linear GARCH$(1,1)$ model (Taylor, 2008):

$$y_t = \varepsilon_t h_t, \quad h_t = a_0 + a_1|y_{t-1}| + b_1 h_{t-1}, \tag{1.1}$$

where $\{y_t\}$ is the observed series, and $\{\varepsilon_t\}$ are independent and identically distributed $(i.i.d.)$ innovations with mean zero. The τth conditional quantile function of y_t is

$$Q_\tau(y_t|y_{t-1}, y_{t-2}, \dots) = (a_0 + a_1|y_{t-1}| + b_1 h_{t-1})\, Q_\tau(\varepsilon_t) = \boldsymbol{\theta}_\tau' \boldsymbol{z}_t,$$

where $Q_\tau(\varepsilon_t)$ is the τth quantile of ε_t, $\boldsymbol{\theta}_\tau = (a_0, a_1, b_1)' Q_\tau(\varepsilon_t)$, and $\boldsymbol{z}_t = (1, |y_{t-1}|, h_{t-1})'$. Thus, $Q_\tau(y_t|y_{t-1}, y_{t-2}, \dots)$ can be estimated by replacing $\boldsymbol{\theta}_\tau$ and the volatility h_t with their estimates; see Xiao and Koenker (2009) and Zheng et al. (2018). Note that $Q_\tau(y_t|y_{t-1}, y_{t-2}, \dots)$ is dependent on τ only through $Q_\tau(\varepsilon_t)$, whereas the GARCH parameters remain invariant across different τ. However, in practice the GARCH parameters may vary across quantile levels. The above framework would fail to capture this phenomenon, potentially resulting in poor forecast accuracy; see Section 6 for empirical evidence. To address this limitation, a natural idea is to allow the GARCH parameters to be τ-dependent.

Recently random-coefficient time series models built upon quantile regression have attracted growing attention. By assuming that the AR coefficients are monotonic functions of a standard uniform random variable, the quantile AR model in Koenker and Xiao (2006) allows for asymmetric dynamic structures across quantile levels; see, e.g., Ferreira (2011) and Baur et al. (2012) for various empirical applications of this model. There have been many extensions of the quantile AR model, such as the quantile self-exciting threshold AR model (Cai and Stander, 2008), the threshold quantile AR model (Galvao et al., 2011), and the quantile double AR model (Zhu and Li, 2022). However, as far as we know, the approach of Koenker and Xiao (2006) has not been explored for GARCH-type models. To fill this gap, this paper proposes the quantile GARCH model, where the GARCH parameters are allowed to vary across quantile levels.

Our main contributions are threefold. First, we develop a more flexible QR framework for conditional heteroscadastic time series, namely the quantile GARCH model, and establish a sufficient condition for its strict stationarity. As the volatility process of the GARCH model is latent and defined recursively, a direct extension of Koenker and Xiao (2006) would be infeasible. Instead, by exploiting the ARCH(∞) form (Zaffaroni, 2004) of the GARCH model, we introduce a random-coefficient GARCH process, where the GARCH parameters are functions of a standard uniform random variable. It can be written as a weighted sum of past information across all lags, where the weights are exponentially decaying random-coefficient functions. The proposed model can capture asymmetric dynamic structures and varying persistence across different quantile levels, while including the linear GARCH model as a special case.

Secondly, for the proposed quantile GARCH model, we introduce the self-weighted QR estimator. The uniform convergence theory of the estimator, including uniform consistency and weak convergence, is established for the quantile process with respect to the quantile level τ. Note that the weak convergence of the unweighted QR estimator would require $E(|y_t|^3) \leq \infty$. By contrast, the self-weighted estimator only requires $E(|y_t|^s) \leq \infty$ for an arbitrarily small $s > 0$ and thus is applicable to very heavy-tailed financial data. The major theoretical difficulty comes from the non-convex and non-differentiable objective function of self-weighted QR estimator. To overcome it, we adopt the bracketing method in Pollard (1985) to derive the pointwise Bahadur representation of the self-weighted QR estimator for each fixed τ, hence the pointwise $\sqrt{n}$-consistency and asymptotic normality. Then, we strengthen the pointwise convergence to uniform

convergence for all τ, by deriving the Bahadur representation uniformly in τ and proving the asymptotic tightness of its leading term. In addition, to check whether the persistence coefficient is τ-independent, we construct a Cramér-von Misses (CvM) test. Based on the weak convergence result, we obtain the limiting null distribution of the CvM test statistic and propose a feasible subsampling method to calculate its critical values.

Finally, to remedy the possible inefficiency of the QR at high quantile levels due to data scarcity, we further introduce the self-weighted composite quantile regression (CQR) estimator. High quantile levels are of great interest in financial risk management. A common approach to extremal QR (Chernozhukov, 2005) is to estimate the quantiles at multiple intermediate levels and then extrapolate those at high levels (Wang et al., 2012; Li and Wang, 2019). We adopt such an approach for the quantile GARCH model. Since this model is similar to Taylor's (2008) GARCH model, we can conveniently make use of the latter for the extrapolation under a chosen innovation distribution such that an explicit quantile function is available. We choose the Tukey-lambda distribution (Joiner and Rosenblatt, 1971), since it not only has an explicit quantile function, but is flexible in fitting heavy tails and approximating many common distributions such as the Gaussian distribution (Gilchrist, 2000). For the proposed weighted CQR estimator, we derive asymptotic properties under possible model misspecification and provide practical suggestions for computational issues. In addition, our simulation studies and empirical analysis indicate that the CQR outperforms the QR at high quantile levels.

The rest of this paper is organized as follows. Section 2 introduces the quantile GARCH(1,1) model and studies its strict stationarity. Section 3 proposes the self-weighted QR estimator, together with the convergence theory for the corresponding quantile process and a CvM test for checking the constancy of the persistence coefficient across all quantile levels. Section 4 introduces the CQR estimator and derives its asymptotic properties. Simulation studies and an empirical example are provided in Sections 5 and 6, respectively. Conclusion and discussion are given in Section 7. A section on the generalization to the quantile GARCH(p, q) model, all technical proofs, and additional numerical results are given in the Supplementary Material. Throughout the paper, $\to_d$ denotes the convergence in distribution, $\rightsquigarrow$ denotes weak convergence, and $o_p(1)$ denotes the convergence in probability. Moreover, $\|\cdot\|$ denotes the norm of a matrix or column vector, defined as $\|A\| = \sqrt{\mathrm{tr}(AA')} = \sqrt{\sum_{i,j} a_{ij}^2}$. In addition, $\ell^\infty(\mathcal{T})$ denotes the space of all uniformly bounded functions on $\mathcal{T}$. The dataset in Section 6 and computer programs

for the analysis are available at https://github.com/Tansonghua-sufe/QGARCH.

2 Proposed quantile GARCH(1,1) model

2.1 Motivation

For succinctness, we restrict our attention to the quantile GARCH(1,1) model in the main paper, while the generalization to the quantile GARCH(p, q) model is detailed in the Supplementary Material.

To motivate the proposed model, first consider a strictly stationary GARCH$(1, 1)$ process in the form of

$$x_t = \eta_t h_t^{1/2}, \quad h_t = a_0 + a_1 x_{t-1}^2 + b_1 h_{t-1}, \tag{2.1}$$

where $a_0 > 0$, $a_1 \geq 0$, $b_1 \geq 0$, and the innovations $\{\eta_t\}$ are $i.i.d.$ random variables with mean zero and variance one. The ARCH(∞) representation (Zaffaroni, 2004) of model (2.1) can be written as

$$x_t = \eta_t \left(\frac{a_0}{1 - b_1} + a_1 \sum_{j=1}^{\infty} b_1^{j-1} x_{t-j}^2 \right)^{1/2}. \tag{2.2}$$

Then, the τth conditional quantile function of x_t in model (2.2) is given by

$$Q_\tau(x_t | x_{t-1}, x_{t-2}, \dots) = Q_\tau(\eta_t) \left(\frac{a_0}{1 - b_1} + a_1 \sum_{j=1}^{\infty} b_1^{j-1} x_{t-j}^2 \right)^{1/2}, \quad \tau \in (0, 1), \tag{2.3}$$

where $Q_\tau(\eta_t)$ denotes the τth quantile of η_t. The parameters a_0, a_1 and b_1, which are independent of the specified quantile level τ, control the scale of the conditional distribution of x_t, while the distribution of η_t determines its shape. As a result, if the GARCH coefficients are allowed to vary with τ and thus capable of altering both the scale and shape of the conditional distribution, we will have a more flexible model that can accommodate asymmetric dynamic structures across different quantile levels.

However, note that (2.3) is nonlinear in the coefficients of the x_{t-j}^2's. Consequently, a direct extension from (2.1) to a varying-coefficient model is undesirable, since it will result in a nonlinear conditional quantile function whose estimation is computationally challenging. Alternatively, we will consider the linear GARCH$(1, 1)$ model in (1.1), in which case (2.2) is revised to

$$y_t = \varepsilon_t \left(\frac{a_0}{1 - b_1} + a_1 \sum_{j=1}^{\infty} b_1^{j-1} |y_{t-j}| \right). \tag{2.4}$$

Then, its corresponding conditional quantile function has the following linear form:

$$Q_\tau(y_t|y_{t-1}, y_{t-2}, \dots) = Q_\tau(\varepsilon_t)\left(\frac{a_0}{1-b_1} + a_1\sum_{j=1}^{\infty} b_1^{j-1}|y_{t-j}|\right), \quad \tau \in (0,1). \tag{2.5}$$

We will adopt (2.5) to formulate the proposed quantile GARCH model.

Remark 2.1. As shown in Zheng et al. (2018), the traditional GARCH(1, 1) model in (2.1) has an equivalent form of the linear GARCH(1, 1) model in (1.1) up to a one-to-one transformation $T(\cdot)$. Specifically, for any x_t following model (2.2), if we take the transformation $y_t = T(x_t) = x_t^2\text{sgn}(x_t)$, then it can be shown that y_t satisfies (2.4) with $\varepsilon_t = T(\eta_t) = \eta_t^2\text{sgn}(\eta_t)$. Note that $E(\varepsilon_t)$ may not be zero although $E(\eta_t) = 0$, and this will not affect our derivation since the conditional quantile function at (2.5) depends on $Q_\tau(\varepsilon_t)$ rather than $E(\varepsilon_t)$.

2.2 The proposed model

Let $\mathcal{F}_t$ be the σ-field generated by $\{y_t, y_{t-1}, \dots\}$. To allow the GARCH parameters to vary with τ, we extend model (2.5) to the following conditional quantile model:

$$Q_\tau(y_t|\mathcal{F}_{t-1}) = \omega(\tau) + \alpha_1(\tau)\sum_{j=1}^{\infty}[\beta_1(\tau)]^{j-1}|y_{t-j}|, \quad \tau \in (0,1), \tag{2.6}$$

where $\omega : (0,1) \to \mathbb{R}$ and $\alpha_1 : (0,1) \to \mathbb{R}$ are unknown monotonic increasing functions, and $\beta_1 : (0,1) \to [0,1)$ is a non-negative real-valued function. Note that both the scale and shape of the conditional distribution of y_t can be altered by the past information $|y_{t-j}|$. Assuming that the right hand side of (2.6) is monotonic increasing in τ, then (2.6) is equivalent to the following random-coefficient process:

$$y_t = \omega(U_t) + \alpha_1(U_t)\sum_{j=1}^{\infty}[\beta_1(U_t)]^{j-1}|y_{t-j}|, \tag{2.7}$$

where $\{U_t\}$ is a sequence of *i.i.d.* standard uniform random variables; see a discussion on the monotonicity of $Q_\tau(y_t|\mathcal{F}_{t-1})$ with respect to τ in Remark 2.2. We call model (2.6) or (2.7) the quantile GARCH(1, 1) model.

Similar to the GARCH model which requires the innovations to have mean zero, the quantile GARCH model also needs a location constraint. For the conditional quantile function (2.6), we may impose that

$$Q_{0.5}(y_t|\mathcal{F}_{t-1}) = 0. \tag{2.8}$$

Since $\beta_1(\cdot)$ is non-negative, condition (2.8) holds if and only if

$$\omega(0.5) = \alpha_1(0.5) = 0. \tag{2.9}$$

For the quantile GARCH$(1,1)$ model, we impose condition (2.9) throughout this paper.

Recall that the functions $\omega(\cdot)$ and $\alpha_1(\cdot)$ are monotonic increasing and $\beta_1(\cdot)$ is non-negative. Under (2.9) the quantile GARCH$(1,1)$ model (2.7) can be rewritten into

$$\begin{aligned} y_t &= \operatorname{sgn}(U_t - 0.5)|y_t|, \\ |y_t| &= |\omega(U_t)| + \sum_{j=1}^{\infty} |\alpha_1(U_t)|[\beta_1(U_t)]^{j-1}|y_{t-j}|, \end{aligned}$$

where y_t, $U_t - 0.5$, $\omega(U_t)$ and $\alpha_1(U_t)$ have the same sign at each time t. For simplicity, denote $\phi_{0,t} = |\omega(U_t)|$ and $\phi_{j,t} = |\alpha_1(U_t)|[\beta_1(U_t)]^{j-1}$ for $j \geq 1$. Then the quantile GARCH$(1,1)$ model (2.7) is equivalent to

$$y_t = \operatorname{sgn}(U_t - 0.5)|y_t|, \quad |y_t| = \phi_{0,t} + \sum_{j=1}^{\infty} \phi_{j,t}|y_{t-j}|, \quad j \geq 1. \tag{2.10}$$

This enables us to establish a sufficient condition for the existence of a strictly stationary solution of the quantile GARCH$(1,1)$ model in the following theorem.

Theorem 2.1. *Suppose that condition (2.9) holds. If there exists $s \in (0,1]$ such that*

$$E(\phi_{0,t}^s) < \infty \quad \textit{and} \quad \sum_{j=1}^{\infty} E(\phi_{j,t}^s) < 1, \tag{2.11}$$

or $s > 1$ such that

$$E(\phi_{0,t}^s) < \infty \quad \textit{and} \quad \sum_{j=1}^{\infty} [E(\phi_{j,t}^s)]^{1/s} < 1, \tag{2.12}$$

then there exists a strictly stationary solution of the quantile GARCH$(1,1)$ equations in (2.10)*, and the process $\{y_t\}$ defined by*

$$y_t = \operatorname{sgn}(U_t - 0.5)\left(\phi_{0,t} + \sum_{\ell=1}^{\infty} \sum_{j_1,\ldots,j_\ell=1}^{\infty} \phi_{0,t-j_1-\cdots-j_\ell}\phi_{j_1,t}\phi_{j_2,t-j_1}\cdots\phi_{j_\ell,t-j_1-\cdots-j_{\ell-1}}\right) \tag{2.13}$$

is the unique strictly stationary and $\mathcal{F}_t^U$-measurable solution to (2.10) *such that $E|y_t|^s < \infty$, where $\mathcal{F}_t^U$ is the σ-field generated by $\{U_t, U_{t-1}, \ldots\}$.*

Theorem 2.1 gives a sufficient condition for the existence of a unique strictly stationary solution satisfying $E|y_t|^s < \infty$. The proof relies on a method similar to that of Theorem 1 in Douc et al. (2008); see also Giraitis et al. (2000) and Royer (2022).

Remark 2.2 (Monotonicity conditions for quantile and coefficient functions)**.** As discussed in Koenker and Xiao (2006) and Phillips (2015), it is very difficult to derive a necessary and sufficient condition on random-coefficient functions to ensure the monotonicity of $Q_\tau(y_t|\mathcal{F}_{t-1})$ in τ for the quantile GARCH$(1,1)$ model in (2.6). Given that $\omega(\cdot)$ and $\alpha_1(\cdot)$ are monotonic increasing, a sufficient condition for monotonicity of $Q_\tau(y_t|\mathcal{F}_{t-1})$ is that the non-negative function $\beta_1(\cdot)$ is monotonic decreasing on $(0, 0.5)$ and monotonic increasing on $(0.5, 1)$. However, since $Q_\tau(y_t|\mathcal{F}_{t-1})$ could be monotonic increasing even if $\beta_1(\cdot)$ does not satisfy the above constraint (e.g., if $\beta_1(\tau)$ is constant over τ), we refrain from imposing any monotonicity constraint on $\beta_1(\cdot)$ in order to avoid overly restricting the function space.

Remark 2.3 (Special cases of Theorem 2.1)**.** When $\omega(U_t) = a_0\varepsilon_t/(1-b_1), \alpha_1(U_t) = a_1\varepsilon_t$, and $\beta_1(U_t) = b_1$, the quantile GARCH$(1,1)$ model in (2.7) reduces to the linear GARCH$(1,1)$ model in (2.4). Then, (2.11) can be simply written as $a_1^s E|\varepsilon_t|^s + b_1^s < 1$ for $s \in (0,1]$, while (2.12) reduces to $a_1(E|\varepsilon_t|^s)^{1/s} + b_1 < 1$ with $E|\varepsilon_t|^s < \infty$ for $s > 1$. In particular, when $s = 1$, the stationarity condition becomes $a_1 + b_1 < 1$, which is exactly the necessary and sufficient condition for the existence of a second-order stationary solution to the GARCH$(1,1)$ model in (2.1). If $s = 2$, then the condition becomes $a_1[E(\eta_t^4)]^{1/2} + b_1 < 1$ with $E(\eta_t^4) < \infty$, which is slightly stronger than the necessary and sufficient condition for the existence of a fourth-order stationary solution to the GARCH$(1,1)$ model in (2.1); see also Bollerslev (1986) and Zaffaroni (2004).

Remark 2.4 (Extension to asymmetric quantile GARCH models)**.** There are numerous variants of the GARCH model, such as the exponential GARCH (Nelson, 1991) and threshold GARCH (Zakoian, 1994) models. The quantile GARCH model in this paper can be extended along the lines of these variants. For example, to capture leverage effects in quantile dynamics, as the quantile counterpart of the threshold GARCH model (Zakoian, 1994), the threshold quantile GARCH$(1,1)$ model can be defined as

$$Q_\tau(y_t|\mathcal{F}_{t-1}) = \omega(\tau) + \alpha_1^+(\tau)\sum_{j=1}^{\infty}[\beta_1(\tau)]^{j-1}y_{t-j}^+ - \alpha_1^-(\tau)\sum_{j=1}^{\infty}[\beta_1(\tau)]^{j-1}y_{t-j}^-,$$

where $\omega : (0,1) \to \mathbb{R}$ and $\alpha_1^+, \alpha_1^- : (0,1) \to \mathbb{R}$ are monotonic increasing, $\beta_1 : (0,1) \to [0,1)$, $y_{t-j}^- = \min\{y_{t-j}, 0\}$, and $y_{t-j}^+ = \max\{y_{t-j}, 0\}$. We leave this interesting extension for future research.

3 Quantile regression

3.1 Self-weighted estimation

Let $\boldsymbol{\theta} = (\omega, \alpha_1, \beta_1)' \in \Theta$ be the parameter vector of the quantile GARCH(1, 1) model, which belongs to the parameter space $\Theta \subset \mathbb{R}^2 \times [0, 1)$. From (2.6), we can define the conditional quantile function below,

$$q_t(\boldsymbol{\theta}) = \omega + \alpha_1 \sum_{j=1}^{\infty} \beta_1^{j-1} |y_{t-j}|.$$

Since the function $q_t(\boldsymbol{\theta})$ depends on observations in the infinite past, initial values are required in practice. In this paper, we set $y_t = 0$ for $t \leq 0$, and denote the resulting function by $\widetilde{q}_t(\boldsymbol{\theta})$, that is, $\widetilde{q}_t(\boldsymbol{\theta}) = \omega + \alpha_1 \sum_{j=1}^{t-1} \beta_1^{j-1} |y_{t-j}|$. We will prove that the effect of the initial values on the estimation and inference is asymptotically negligible.

Let $\psi_\tau(x) = \tau - I(x < 0)$, where the indicator function $I(\cdot) = 1$ if the condition is true and 0 otherwise. For any $\tau \in \mathcal{T} \subset (0, 1)$, we propose the self-weighted quantile regression (QR) estimator as follows,

$$\widetilde{\boldsymbol{\theta}}_{wn}(\tau) = (\widetilde{\omega}_{wn}(\tau), \widetilde{\alpha}_{1wn}(\tau), \widetilde{\beta}_{1wn}(\tau))' = \underset{\boldsymbol{\theta} \in \Theta}{\operatorname{argmin}} \sum_{t=1}^{n} w_t \rho_\tau \left(y_t - \widetilde{q}_t(\boldsymbol{\theta}) \right), \tag{3.1}$$

where $\{w_t\}$ are nonnegative random weights, and $\rho_\tau(x) = x\psi_\tau(x) = x[\tau - I(x < 0)]$ is the check function; see also Ling (2005), Zhu and Ling (2011), and Zhu et al. (2018).

When $w_t = 1$ for all t, (3.1) reduces to the unweighted QR estimator. In this case, the consistency and asymptotic normality of the estimator would require $E|y_t| < \infty$ and $E|y_t|^3 < \infty$, respectively. A sufficient condition for the existence of these moments is provided in Theorem 2.1. However, higher order moment conditions will make the stationarity region much narrower. Moreover, financial time series are usually heavy-tailed, so these moment conditions can be easily violated. By contrast, using the self-weighting approach (Ling, 2005), we only need a finite fractional moment of $|y_t|$.

Denote the true parameter vector by $\boldsymbol{\theta}(\tau) = (\omega(\tau), \alpha_1(\tau), \beta_1(\tau))'$. Let $F_{t-1}(\cdot)$ and $f_{t-1}(\cdot)$ be the distribution and density functions of y_t conditional on $\mathcal{F}_{t-1}$, respectively. To establish the asymptotic properties of $\widetilde{\boldsymbol{\theta}}_{wn}(\tau)$, we need the following assumptions.

Assumption 1. $\{y_t\}$ *is strictly stationary and ergodic.*

Assumption 2. *(i) The parameter space* Θ *is compact; (ii)* $\boldsymbol{\theta}(\tau)$ *is an interior point of* Θ.

Assumption 3. *With probability one, $f_{t-1}(\cdot)$ and its derivative function $\dot{f}_{t-1}(\cdot)$ are uniformly bounded, and $f_{t-1}(\cdot)$ is positive on the support $\{x : 0 < F_{t-1}(x) < 1\}$.*

Assumption 4. *$\{w_t\}$ is strictly stationary and ergodic, and w_t is nonnegative and measurable with respect to $\mathcal{F}_{t-1}$ such that $E(w_t) < \infty$ and $E(w_t|y_{t-j}|^3) < \infty$ for $j \geq 1$.*

Assumption 5. *The functions $\omega(\cdot), \alpha_1(\cdot)$ and $\beta_1(\cdot)$ are Lipschitz continuous.*

Theorem 2.1 provides a sufficient condition for Assumption 1. In Assumption 2, condition (i) is standard for the consistency of estimator, while condition (ii) is needed for the asymptotic normality; see also Francq and Zakoian (2010) and Zheng et al. (2018). Assumption 3 is commonly required for QR processes whose coefficients are functions of a uniform random variable; see Assumption A.3 in Koenker and Xiao (2006) for quantile AR models and Assumption 4 in Zhu and Li (2022) for quantile double AR models. Specifically, the positiveness and continuity of $f_{t-1}(\cdot)$ are required to show the uniform consistency of $\widetilde{\boldsymbol{\theta}}_{wn}(\tau)$ in Theorem 3.1, while the boundedness of $f_{t-1}(\cdot)$ and $\dot{f}_{t-1}(\cdot)$ is needed for the weak convergence in Theorem 3.2. In the special case where the quantile GARCH$(1,1)$ model in (2.7) reduces to model (2.4), Assumption 3 can be simplified to conditions similar to Assumption (A2) in Lee and Noh (2013) and Assumption 4 in Zhu et al. (2021). Assumption 4 on the self-weights $\{w_t\}$ is used to reduce the moment requirement on $\{y_t\}$ in establishing asymptotic properties of $\widetilde{\boldsymbol{\theta}}_{wn}(\tau)$; see more discussions on $\{w_t\}$ in Remark 3.1. Assumption 5 is required to establish the stochastic equicontinuity for weak convergence in Theorem 3.2.

Let $\boldsymbol{T}_n(\tau) = n^{-1/2}\sum_{t=1}^{n} w_t\dot{q}_t(\boldsymbol{\theta}(\tau))\psi_\tau(y_t - q_t(\boldsymbol{\theta}(\tau)))$ and $\Sigma_w(\tau_1,\tau_2) = (\min\{\tau_1,\tau_2\} - \tau_1\tau_2)\Omega_{1w}^{-1}(\tau_1)\Omega_{0w}(\tau_1,\tau_2)\Omega_{1w}^{-1}(\tau_2)$, where $\Omega_{0w}(\tau_1,\tau_2) = E\left[w_t^2\dot{q}_t(\boldsymbol{\theta}(\tau_1))\dot{q}_t'(\boldsymbol{\theta}(\tau_2))\right]$ and $\Omega_{1w}(\tau) = E\left[f_{t-1}(F_{t-1}^{-1}(\tau))w_t\dot{q}_t(\boldsymbol{\theta}(\tau))\dot{q}_t'(\boldsymbol{\theta}(\tau))\right]$. Theorems 3.1 and 3.2 below establish the uniform consistency and weak convergence for the QR process $\widetilde{\boldsymbol{\theta}}_{wn}(\cdot)$, respectively.

Theorem 3.1. *For $\{y_t\}$ generated by model (2.7) with condition (2.9), suppose $E|y_t|^s < \infty$ for some $s \in (0,1)$. If Assumptions 1, 2(i), 3 and 4 hold, then $\sup_{\tau\in\mathcal{T}}\|\widetilde{\boldsymbol{\theta}}_{wn}(\tau) - \boldsymbol{\theta}(\tau)\| \to_p 0$ as $n \to \infty$.*

Theorem 3.2. *For $\{y_t\}$ generated by model (2.7) with condition (2.9), suppose $E|y_t|^s < \infty$ for some $s \in (0,1)$ and the covariance kernel $\Sigma_w(\tau_1,\tau_2)$ is positive definite uniformly for $\tau_1 = \tau_2 = \tau \in \mathcal{T}$. If Assumptions 1–5 hold, as $n \to \infty$, then we have*

$$\sqrt{n}(\widetilde{\boldsymbol{\theta}}_{wn}(\cdot) - \boldsymbol{\theta}(\cdot)) = \Omega_{1w}^{-1}(\cdot)\boldsymbol{T}_n(\cdot) + o_p(1) \rightsquigarrow \mathbb{G}(\cdot) \text{ in } (\ell^\infty(\mathcal{T}))^3, \tag{3.2}$$

where the remainder term is uniform in $\tau \in \mathcal{T}$*, and* $\mathbb{G}(\cdot)$ *is a zero mean Gaussian process with covariance kernel* $\Sigma_w(\tau_1, \tau_2)$*.*

Owing to the self-weights, the above results hold for very heavy-tailed data with a finite fractional moment. The proof of Theorem 3.2 is nontrivial. The first challenge comes from the non-convex and non-differentiable objective function of QR. Specifically, we need to prove the finite dimensional convergence of $\widetilde{\boldsymbol{\theta}}_{wn}(\tau)$, i.e., the $\sqrt{n}$-consistency of $\widetilde{\boldsymbol{\theta}}_{wn}(\tau)$ for each τ in the form of $\sqrt{n}(\widetilde{\boldsymbol{\theta}}_{wn}(\tau) - \boldsymbol{\theta}(\tau)) = O_p(1)$. We overcome this challenge by adopting the bracketing method in Pollard (1985). The second challenge is to obtain the Bahadur representation uniformly in $\tau \in \mathcal{T}$ and prove the asymptotic tightness of the leading term $\Omega_{1w}^{-1}(\cdot)\boldsymbol{T}_n(\cdot)$ in this representation. The key to accomplishing this is to verify the stochastic equicontinuity for all remainder terms and $\boldsymbol{T}_n(\cdot)$.

In particular, when a fixed quantile level $\tau \in \mathcal{T}$ is considered, by the martingale central limit theorem (CLT), we can obtain the asymptotic normality of $\widetilde{\boldsymbol{\theta}}_{wn}(\tau)$ without the Lipschitz condition in Assumption 5 as follows.

Corollary 3.1. *For* $\{y_t\}$ *generated by model* (2.7) *with condition* (2.9)*, suppose* $E|y_t|^s < \infty$ *for some* $s \in (0, 1)$ *and* $\Sigma_w(\tau, \tau)$ *is positive definite. If Assumptions 1–4 hold, then* $\sqrt{n}(\widetilde{\boldsymbol{\theta}}_{wn}(\tau) - \boldsymbol{\theta}(\tau)) \to_d N(\mathbf{0}, \Sigma_w(\tau, \tau))$ *as* $n \to \infty$*.*

To estimate the asymptotic covariance $\Sigma_w(\tau, \tau)$ in Corollary 3.1, we first estimate $f_{t-1}(F_{t-1}^{-1}(\tau))$ in $\Omega_{1w}(\tau)$ using the difference quotient method (Koenker, 2005). Let $\widetilde{Q}_\tau(y_t|\mathcal{F}_{t-1}) = \widetilde{q}_t(\widetilde{\boldsymbol{\theta}}_{wn}(\tau))$ be the fitted τth conditional quantile. We employ the estimator $\widetilde{f}_{t-1}(F_{t-1}^{-1}(\tau)) = 2\ell[\widetilde{Q}_{\tau+\ell}(y_t|\mathcal{F}_{t-1}) - \widetilde{Q}_{\tau-\ell}(y_t|\mathcal{F}_{t-1})]^{-1}$, where ℓ is the bandwidth. As in Koenker and Xiao (2006), we consider two commonly used bandwidths for ℓ as follows:

$$\ell_B = n^{-1/5}\left\{\frac{4.5 f_N^4(F_N^{-1}(\tau))}{[2F_N^{-2}(\tau)+1]^2}\right\}^{1/5} \quad \text{and} \quad \ell_{HS} = n^{-1/3} z_\alpha^{2/3}\left\{\frac{1.5 f_N^2(F_N^{-1}(\tau))}{2F_N^{-2}(\tau)+1}\right\}^{1/3}, \tag{3.3}$$

where $f_N(\cdot)$ and $F_N(\cdot)$ are the standard normal density and distribution functions, respectively, and $z_\alpha = F_N^{-1}(1-\alpha/2)$ with $\alpha = 0.05$. Then the matrices $\Omega_{0w}(\tau, \tau)$ and $\Omega_{1w}(\tau)$ can be approximated by the sample averages:

$$\widetilde{\Omega}_{0w}(\tau, \tau) = \frac{1}{n}\sum_{t=1}^{n} w_t^2 \dot{\widetilde{q}}_t(\widetilde{\boldsymbol{\theta}}_{wn}(\tau))\dot{\widetilde{q}}_t'(\widetilde{\boldsymbol{\theta}}_{wn}(\tau)) \quad \text{and}$$

$$\widetilde{\Omega}_{1w}(\tau) = \frac{1}{n}\sum_{t=1}^{n} \widetilde{f}_{t-1}(F_{t-1}^{-1}(\tau)) w_t \dot{\widetilde{q}}_t(\widetilde{\boldsymbol{\theta}}_{wn}(\tau))\dot{\widetilde{q}}_t'(\widetilde{\boldsymbol{\theta}}_{wn}(\tau)),$$

where $\dot{\widetilde{q}}_t(\boldsymbol{\theta}) = (1, \sum_{j=1}^{t-1} \beta_1^{j-1}|y_{t-j}|, \alpha_1 \sum_{j=2}^{t-1}(j-1)\beta_1^{j-2}|y_{t-j}|)'$. Consequently, a consistent estimator of $\Sigma_w(\tau,\tau)$ can be constructed as $\widetilde{\Sigma}_w(\tau,\tau) = \tau(1-\tau)\widetilde{\Omega}_{1w}^{-1}(\tau)\widetilde{\Omega}_{0w}(\tau,\tau)\widetilde{\Omega}_{1w}^{-1}(\tau)$.

Remark 3.1 (Choices of self-weights)**.** The goal of the self-weights $\{w_t\}$ is to relax the moment condition from $E|y_t|^3 < \infty$ to $E|y_t|^s < \infty$ for $s \in (0,1)$. If there is empirical evidence that $E|y_t|^3 < \infty$ holds, then we can simply let $w_t = 1$ for all t. Otherwise, the self-weights are needed. There are many choices of random weights $\{w_t\}$ that satisfy Assumption 4. Note that the main role of $\{w_t\}$ in our technical proofs is to bound the term $w_t y_{t-j}^{\delta}$ for $\delta \geq 1$ by $O(|y_{t-j}|^s)$ for some $s \in (0,1)$. Following He et al. (2020), we may consider

$$w_t = \left(\sum_{i=0}^{\infty} e^{-\log^2(i+1)} \left\{ I\left[|y_{t-i-1}| \leq c\right] + c^{-1}|y_{t-i-1}| I\left[|y_{t-i-1}| > c\right] \right\} \right)^{-3} \tag{3.4}$$

for some given $c > 0$, where y_s is set to zero for $s \leq 0$. In our simulation and empirical studies, we take c to be the 95% sample quantile of $\{y_t\}_{t=1}^n$.

Remark 3.2 (The quantile crossing problem)**.** If we are only interested in estimating $Q_\tau(y_t|\mathcal{F}_{t-1})$ at a specific quantile level τ, the L-BFGS-B algorithm (Zhu et al., 1997) can be used to solve (3.1) with the constraint $\beta_1 \in (0,1)$. Then the estimate $\widetilde{Q}_\tau(y_t|\mathcal{F}_{t-1}) = \widetilde{q}_t(\widetilde{\boldsymbol{\theta}}_{wn}(\tau))$ can be obtained for $Q_\tau(y_t|\mathcal{F}_{t-1})$. As a more flexible approach, we may study multiple quantile levels simultaneously, say $\tau_1 < \tau_2 < \cdots < \tau_K$. However, the pointwise estimates $\{\widetilde{Q}_{\tau_k}(y_t|\mathcal{F}_{t-1})\}_{k=1}^K$ in practice may not be a monotonic increasing sequence even if $Q_\tau(y_t|\mathcal{F}_{t-1})$ is monotonic increasing in τ. To overcome the quantile crossing problem, we adopt the easy-to-implement rearrangement method (Chernozhukov et al., 2010) to enforce the monotonicity of pointwise quantile estimates $\{\widetilde{Q}_{\tau_k}(y_t|\mathcal{F}_{t-1})\}_{k=1}^K$. By Proposition 4 in Chernozhukov et al. (2010), it can be shown that the rearranged quantile curve has smaller estimation error than the original one whenever the latter is not monotone; see also the simulation experiment in Section 3.2 of the Supplementary Material.

Remark 3.3 (Rearranging coefficient functions)**.** The proposed model in (2.6) assumes that $\omega(\cdot)$ and $\alpha_1(\cdot)$ are monotonic increasing. In practice, we can apply the method in Chernozhukov et al. (2009) to rearrange the estimates $\{\widetilde{\omega}_{wn}(\tau_k)\}_{k=1}^K$ and $\{\widetilde{\alpha}_{1wn}(\tau_k)\}_{k=1}^K$ to ensure the monotonicity of the curves across τ_k's. It is shown in Chernozhukov et al. (2009) that the rearranged confidence intervals are monotonic and narrower than the original ones.

3.2 Testing for constant persistence coefficient

In this subsection, we present a test to determine if the persistence coefficient $\beta_1(\tau)$ is independent of the quantile level τ for $\tau \in \mathcal{T} \subset (0,1)$. This problem can be cast as a more general hypothesis testing problem as follows:

$$H_0: \ \forall \tau \in \mathcal{T}, \ R\boldsymbol{\theta}(\tau) = r \quad \text{versus} \quad H_1: \ \exists \tau \in \mathcal{T}, \ R\boldsymbol{\theta}(\tau) \neq r, \tag{3.5}$$

where R is a predetermined row vector, and $r \in \Gamma$ denotes a parameter whose specific value is unknown, but it is known to be independent of τ. Here the parameter space Γ contains all values $R\boldsymbol{\theta}(\tau)$ can take under the proposed model. Then, we can write the hypotheses for testing the constancy of $\beta_1(\tau)$ in the form of (3.5) by setting $R = (0,0,1)$ and $r = \beta_1 \in \Gamma = (0,1)$. In this case, the null hypothesis in (3.5) means that $\beta_1(\tau)$ does not vary cross quantiles.

For generality, we present the result for the general problem in (3.5). Under H_0, we can estimate the unknown r using $\widetilde{r} = \int_{\mathcal{T}} R\widetilde{\boldsymbol{\theta}}_{wn}(\tau)d\tau$. Define the inference process $v_n(\tau) = R\widetilde{\boldsymbol{\theta}}_{wn}(\tau) - \widetilde{r} = R[\widetilde{\boldsymbol{\theta}}_{wn}(\tau) - \int_{\mathcal{T}} \widetilde{\boldsymbol{\theta}}_{wn}(\tau)d\tau]$. To test H_0, we construct the Cramér-von Misses (CvM) test statistic as follows:

$$S_n = n \int_{\mathcal{T}} v_n^2(\tau)d\tau. \tag{3.6}$$

Let $\sigma(\tau_1,\tau_2) = R[\Sigma_w(\tau_1,\tau_2) + \int_{\mathcal{T}}\int_{\mathcal{T}} \Sigma_w(\tau,\tau')d\tau d\tau' - \int_{\mathcal{T}} \Sigma_w(\tau_1,\tau)d\tau - \int_{\mathcal{T}} \Sigma_w(\tau,\tau_2)d\tau]R'$. Denote $v_0(\tau) = R[\mathbb{G}(\tau) - \int_{\mathcal{T}} \mathbb{G}(\tau)d\tau]$ with $\mathbb{G}(\tau)$ defined in Theorem 3.2.

Corollary 3.2. *Suppose the conditions of Theorem 3.2 hold. Under H_0, then we have $S_n \to_d S \equiv \int_{\mathcal{T}} v_0^2(\tau)d\tau$ as $n \to \infty$. If the covariance function of $v_0(\cdot)$ is nondegenerate, that is, $\sigma(\tau,\tau) > 0$ uniformly in $\tau \in \mathcal{T}$, then $Pr(S_n > c_\alpha) \to Pr(S > c_\alpha) = \alpha$, where the critical value c_α is chosen such that $Pr(S > c_\alpha) = \alpha$.*

Corollary 3.2 indicates that we can reject H_0 if $S_n > c_\alpha$ at the significance level α. In practice, we can use a grid of values $\mathcal{T}_n$ in place of $\mathcal{T}$. Similar to Corollary 3 in Chernozhukov and Hansen (2006), we can verify that Corollary 3.2 still holds for the discretization if the largest cell size of $\mathcal{T}_n$, denoted as δ_n, satisfies $\delta_n \to 0$ as $n \to \infty$.

Note that the CvM test in (3.6) is not asymptotically distribution-free due to the estimation of r, which is commonly known as the Durbin problem (Durbin, 1973). This complicates the approximation of the limiting null distribution of S_n and the resulting critical value c_α. We suggest approximating the limiting null distribution by subsampling the linear approximation of the inference process $v_n(\tau)$; see also Chernozhukov

and Hansen (2006). This approach is computationally efficient as it avoids the repeated estimation over the resampling steps for many values of τ. Specifically, by Theorem 3.2, under H_0 we have

$$\sqrt{n}v_n(\tau) = \frac{1}{\sqrt{n}}\sum_{t=1}^{n} z_t(\tau) + o_p(1), \tag{3.7}$$

where $z_t(\tau) = R[m_t(\tau) - \int_{\mathcal{T}} m_t(\tau)d\tau]$, with $m_t(\tau) = w_t\Omega_{1w}^{-1}(\tau)\dot{q}_t(\boldsymbol{\theta}(\tau))\psi_\tau(y_t - q_t(\boldsymbol{\theta}(\tau)))$. By the consistency of $\widetilde{\boldsymbol{\theta}}_{wn}(\tau)$ in Theorem 3.1, we can estimate $z_t(\tau)$ using $\widetilde{z}_t(\tau) = R[\widetilde{m}_t(\tau) - \int_{\mathcal{T}} \widetilde{m}_t(\tau)d\tau]$, where $\widetilde{m}_t(\tau) = w_t\widetilde{\Omega}_{1w}^{-1}(\tau)\dot{\widetilde{q}}_t(\widetilde{\boldsymbol{\theta}}_{wn}(\tau))\psi_\tau(y_t - \widetilde{q}_t(\widetilde{\boldsymbol{\theta}}_{wn}(\tau)))$. Thus, a sample of estimated scores $\{\widetilde{z}_t(\tau), \tau \in \mathcal{T}, 1 \le t \le n\}$ is obtained, where n is the sample size. Then a subsampling procedure is conducted as follows. Given a block size b_n, we consider $L_n = n - b_n + 1$ overlapping blocks of the sample, indexed by $B_k = \{k, k+1, \ldots, k+b_n-1\}$ for $k = 1, \ldots, L_n$. For each block B_k, we compute the inference process $v_{k,b_n}(\tau) = b_n^{-1}\sum_{t\in B_k}\widetilde{z}_t(\tau)$ and define $S_{k,b_n} = b_n\int_{\mathcal{T}} v_{k,b_n}^2(\tau)d\tau$. Then the critical value c_α can be calculated as the $(1-\alpha)$th empirical quantile of $\{S_{k,b_n}\}_{k=1}^{L_n}$.

To establish the asymptotic validity of the subsampling procedure above, we can use a method similar to the proof of Theorem 5 in Chernozhukov and Hansen (2006). This is possible under the conditions of Theorem 3.2 and an α-mixing condition on y_t, provided that $L_n \to \infty$, $b_n \to \infty$, and $b_n/n \to 0$ as $n \to \infty$. However, we leave the rigorous proof for future research. Following Shao (2011), we consider $b_n = \lfloor cn^{1/2} \rfloor$ with a positive constant c, where $\lfloor x \rfloor$ stands for the integer part of x. Our simulation study shows that the CvM test has reasonable size and power when $c = 0.5, 1$ or 2.

4 Composite quantile regression

4.1 Self-weighted estimation

It is well known that the QR can be unstable when τ is very close to zero or one due to data scarcity (Li and Wang, 2019). However, estimating high conditional quantiles is of great interest in financial risk management. As a remedy, this section proposes the composite quantile regression (CQR). To estimate the conditional quantile at a target level $\tau_0 \in \mathcal{T} \subset (0, 0.01] \cup [0.99, 1)$, the main idea is to conduct extrapolation based on estimation results of intermediate quantile levels at the one-sided neighbourhood of τ_0.

Suppose that $\{y_t\}$ follows the quantile GARCH(1, 1) model in (2.7). Note that the conditional quantile function $Q_\tau(y_t|\mathcal{F}_{t-1})$ cannot be extrapolated directly due to the

unknown nonparametric coefficient functions. To develop a feasible and easy-to-use extrapolation approach, we leverage the close connection between the linear GARCH$(1,1)$ process in (2.4) and quantile GARCH$(1,1)$ process in (2.7). First, we approximate y_t in (2.7) by the linear GARCH$(1,1)$ model in (2.4). Then, the τth conditional quantile of y_t in (2.6) can be approximated by that of the linear GARCH$(1,1)$ model in (2.5):

$$Q_\tau(y_t|\mathcal{F}_{t-1}) \approx Q_\tau(\varepsilon_t)\left(\frac{a_0}{1-b_1} + a_1\sum_{j=1}^{\infty} b_1^{j-1}|y_{t-j}|\right), \tag{4.1}$$

where ε_t's are the *i.i.d.* innovations of the linear GARCH$(1,1)$ model. If the quantile function $Q_\tau(\varepsilon_t)$ has an explicit parametric form, then (4.1) will be fully parametric and hence can be easily used for extrapolation of conditional quantiles of y_t at high levels. While this parametric approximation will induce a bias, the gain is greater estimation efficiency at high quantile levels; see more discussions on the bias-variance trade-off in Section 4.3.

Next we need a suitable distribution of ε_t such that the tail behavior can be flexibly captured. There are many choices such that $Q_\tau(\varepsilon_t)$ has an explicit form, including distributions in lambda and Burr families (Gilchrist, 2000). We choose the Tukey-lambda distribution since it provides a wide range of shapes. It can not only approximate Gaussian and Logistic distributions but also fit heavy Pareto tails well. Given that ε_t follows the Tukey-lambda distribution with shape parameter $\lambda \neq 0$ (Joiner and Rosenblatt, 1971), $Q_\tau(\varepsilon_t)$ has a simple explicit form given by

$$Q_\tau(\lambda) := Q_\tau(\varepsilon_t;\lambda) = \frac{\tau^\lambda - (1-\tau)^\lambda}{\lambda}. \tag{4.2}$$

Combining (4.1) and (4.2), we can approximate the conditional quantile $Q_\tau(y_t|\mathcal{F}_{t-1})$ by

$$q_{t,\tau}(\boldsymbol{\varphi}) = Q_\tau(\lambda)\left(\frac{a_0}{1-b_1} + a_1\sum_{j=1}^{\infty} b_1^{j-1}|y_{t-j}|\right) := Q_\tau(\lambda)h_t(\boldsymbol{\phi}),$$

where $\boldsymbol{\varphi} = (\boldsymbol{\phi}', \lambda)' = (a_0, a_1, b_1, \lambda)'$ is the parameter vector of linear GARCH$(1,1)$ model with ε_t following the Tukey-lambda distribution. Note that $Q_{0.5}(\lambda) = 0$ for any λ. Thus, $q_{t,0.5}(\boldsymbol{\varphi}) = 0$ holds for any $\boldsymbol{\varphi}$, i.e., the location constraint on $Q_\tau(y_t|\mathcal{F}_{t-1})$ in (2.8) is satisfied.

Since $q_{t,\tau}(\boldsymbol{\varphi})$ depends on unobservable values of y_t in the infinite past, in practice we initialize $y_t = 0$ for $t \leq 0$ and define its feasible counterpart as

$$\widetilde{q}_{t,\tau}(\boldsymbol{\varphi}) = Q_\tau(\lambda)\left(\frac{a_0}{1-b_1} + a_1\sum_{j=1}^{t-1} b_1^{j-1}|y_{t-j}|\right) := Q_\tau(\lambda)\widetilde{h}_t(\boldsymbol{\phi}).$$

The initialization effect is asymptotically negligible, as we verify in our technical proofs. Note that $\widetilde{q}_{t,\tau}(\boldsymbol{\varphi})$ is fully parametric. Since $\boldsymbol{\varphi}$ is independent of τ, we can approximate the nonparametric function $Q_{\tau_0}(y_t|\mathcal{F}_{t-1})$ by the parametric function $\widetilde{q}_{t,\tau_0}(\boldsymbol{\varphi})$, where we replace $\boldsymbol{\varphi}$ with an estimator obtained by fitting the above Tukey-lambda linear GARCH$(1,1)$ model at lower quantile levels.

Let $\Phi \subset (0,\infty) \times [0,\infty) \times [0,1) \times \Lambda$ be the parameter space of $\boldsymbol{\varphi}$, where $\Lambda = (-\infty, 0) \cup (0,\infty)$ is the parameter space of λ. To estimate $\boldsymbol{\varphi}$ locally for the target level τ_0, we utilize the information at lower quantile levels in the one-sided neighborhood of τ_0, namely $\mathcal{T}_h = [\tau_0, \tau_0 + h] \subset (0, 0.5)$ if τ_0 is close to zero and $\mathcal{T}_h = [\tau_0 - h, \tau_0] \subset (0.5, 1)$ if τ_0 is close to one, where $h > 0$ is a fixed bandwidth; see Section 4.3 for discussions on the selection of bandwidth h. If $Q_\tau(y_t|\mathcal{F}_{t-1})$ is well approximated by $q_{t,\tau}(\boldsymbol{\varphi})$ for $\tau \in \mathcal{T}_h$, then we can estimate $\boldsymbol{\varphi}$ by the weighted CQR as follows:

$$\check{\boldsymbol{\varphi}}_{wn} = (\check{\boldsymbol{\phi}}'_{wn}, \check{\lambda}_{wn})' = \underset{\boldsymbol{\varphi}\in\Phi}{\operatorname{argmin}} \sum_{t=1}^{n} \sum_{k=1}^{K} w_t \rho_{\tau_k} \left(y_t - \widetilde{q}_{t,\tau_k}(\boldsymbol{\varphi})\right), \tag{4.3}$$

where $\{w_t\}$ are the self-weights defined as in (3.1), and $\tau_1 < \cdots < \tau_K$ are fixed quantile levels with $\tau_k \in \mathcal{T}_h$ for all $1 \le k \le K$; see also Zou and Yuan (2008). In practice, equally spaced levels are typically used. That is, $\tau_k = \tau_0 + h(k-1)/(K-1)$ if τ_0 is close to zero, whereas $\tau_k = \tau_0 - h(k-1)/(K-1)$ if τ_0 is close to one. As a result, the conditional quantile $Q_{\tau_0}(y_t|\mathcal{F}_{t-1})$ can be approximated by $\widetilde{q}_{t,\tau_0}(\check{\boldsymbol{\varphi}}_{wn})$.

4.2 Asymptotic properties

Note that the approximate conditional quantile function $q_{t,\tau}(\boldsymbol{\varphi})$ can be rewritten using the true conditional quantile function $q_t(\cdot)$ as follows:

$$q_{t,\tau}(\boldsymbol{\varphi}) = \frac{a_0 Q_\tau(\lambda)}{1 - b_1} + a_1 Q_\tau(\lambda) \sum_{j=1}^{\infty} b_1^{j-1} |y_{t-j}| := q_t(\boldsymbol{\theta}^*_\tau), \tag{4.4}$$

where $\boldsymbol{\theta}^*_\tau = g_\tau(\boldsymbol{\varphi}) = (a_0 Q_\tau(\lambda)/(1-b_1), a_1 Q_\tau(\lambda), b_1)'$, and $g_\tau : \mathbb{R}^4 \to \mathbb{R}^3$ is a measurable function such that $q_{t,\tau} = q_t \circ g_\tau$. Let $\check{\boldsymbol{\theta}}^*_{wn}(\tau) := g_\tau(\check{\boldsymbol{\varphi}}_{wn})$ be the transformed CQR estimator. In view of (4.4) and the fact that $Q_\tau(y_t|\mathcal{F}_{t-1}) = q_t(\boldsymbol{\theta}(\tau))$, $\check{\boldsymbol{\theta}}^*_{wn}(\tau)$ can be used as an estimator of $\boldsymbol{\theta}(\tau)$; see (2.6) and the definition of $q_t(\cdot)$ in Section 3.1. The pseudo-true parameter vector $\boldsymbol{\varphi}^*_0 = (\boldsymbol{\phi}'_0, \lambda_0)' = (a_{00}, a_{10}, b_{10}, \lambda_0)'$ is defined as

$$\boldsymbol{\varphi}^*_0 = \underset{\boldsymbol{\varphi}\in\Phi}{\operatorname{argmin}} \sum_{k=1}^{K} E[w_t \rho_{\tau_k} \left(y_t - q_{t,\tau_k}(\boldsymbol{\varphi})\right)], \quad \tau_k \in \mathcal{T}_h. \tag{4.5}$$

In other words, for $\tau \in \mathcal{T}_h$, the best approximation of the nonparametric function $Q_\tau(y_t|\mathcal{F}_{t-1}) = q_t(\boldsymbol{\theta}(\tau))$ via the fully parametric function $q_{t,\tau}(\cdot)$ is given by $q_{t,\tau}(\boldsymbol{\varphi}_0^*) = q_t(g_\tau(\boldsymbol{\varphi}_0^*))$.

In general, $Q_\tau(y_t|\mathcal{F}_{t-1})$ may be misspecified by $q_{t,\tau}(\boldsymbol{\varphi}_0^*)$, and $\boldsymbol{\theta}(\tau) = g_\tau(\boldsymbol{\varphi}_0^*)$ may not hold for all τ. Thus, asymptotic properties of the CQR estimator $\check{\boldsymbol{\varphi}}_{wn}$ and its transformation $\check{\boldsymbol{\theta}}_{wn}^*(\tau) = g_\tau(\check{\boldsymbol{\varphi}}_{wn})$ should be established under possible model misspecification. The following assumptions will be required.

Assumption 6. *$\{y_t\}$ is a strictly stationary and α-mixing time series with the mixing coefficient $\alpha(n)$ satisfying $\sum_{n\geq 1}[\alpha(n)]^{1-2/\delta} < \infty$ for some $\delta > 2$.*

Assumption 7. *(i) The parameter space Φ is compact and $\boldsymbol{\varphi}_0^*$ is unique; (ii) $\boldsymbol{\varphi}_0^*$ is an interior point of Φ.*

Note that Assumption 1 is insufficient for the asymptotic normality of $\check{\boldsymbol{\varphi}}_{wn}$ under model misspecification, since $E[\psi_\tau(y_t - q_{t,\tau}(\boldsymbol{\varphi}_0^*))|\mathcal{F}_{t-1}] \neq 0$ in this case, which renders the martingale CLT no longer applicable. Instead, we rely on Assumption 6 to ensure the ergodicity of $\{y_t\}$ and enable the use of the CLT for α-mixing sequences; see Fan and Yao (2003) and more discussions in Remark 4.1. Assumption 7 is analogous to Assumption 2, which is standard in the literature on GARCH models (Francq and Zakoian, 2010; Zheng et al., 2018). If there is no model misspecification, i.e. $Q_\tau(y_t|\mathcal{F}_{t-1})$ is correctly specified by $q_{t,\tau}(\boldsymbol{\varphi}_0^*)$ for all $\tau \in \mathcal{T}_h$, then the uniqueness of $\boldsymbol{\varphi}_0^*$ can be guaranteed for $K \geq 3$ and $\lambda < 1$.

Let $\dot{q}_{t,\tau}(\boldsymbol{\varphi})$ and $\ddot{q}_{t,\tau}(\boldsymbol{\varphi})$ be the first and second derivatives of $q_{t,\tau}(\boldsymbol{\varphi})$ with respect to $\boldsymbol{\varphi}$, respectively, given by

$$\dot{q}_{t,\tau}(\boldsymbol{\varphi}) = (Q_\tau(\lambda)\dot{h}_t'(\boldsymbol{\phi}), \dot{Q}_\tau(\lambda)h_t(\boldsymbol{\phi}))' \quad \text{and} \quad \ddot{q}_{t,\tau}(\boldsymbol{\varphi}) = \begin{pmatrix} Q_\tau(\lambda)\ddot{h}_t(\boldsymbol{\phi}) & \dot{Q}_\tau(\lambda)\dot{h}_t(\boldsymbol{\phi}) \\ \dot{Q}_\tau(\lambda)\dot{h}_t'(\boldsymbol{\phi}) & \ddot{Q}_\tau(\lambda)h_t(\boldsymbol{\phi}) \end{pmatrix},$$

where $\dot{Q}_\tau(\lambda)$ and $\dot{h}_t(\boldsymbol{\phi})$ (or $\ddot{Q}_\tau(\lambda)$ and $\ddot{h}_t(\boldsymbol{\phi})$) are the first (or second) derivatives of $Q_\tau(\lambda)$ and $h_t(\boldsymbol{\phi})$, respectively. Denote $\boldsymbol{X}_t = \sum_{k=1}^K w_t \dot{q}_{t,\tau_k}(\boldsymbol{\varphi}_0^*)\psi_{\tau_k}(y_t - q_{t,\tau_k}(\boldsymbol{\varphi}_0^*))$ and $\Omega_{0w}^* = E(\boldsymbol{X}_t\boldsymbol{X}_t') + n^{-1}\sum_{t\neq s}^n E(\boldsymbol{X}_t\boldsymbol{X}_s')$. Define the matrices

$$\Omega_{11}^* = \sum_{k=1}^K E\left[w_t \ddot{q}_{t,\tau_k}(\boldsymbol{\varphi}_0^*)\psi_{\tau_k}(y_t - q_{t,\tau_k}(\boldsymbol{\varphi}_0^*))\right] \text{ and } \Omega_{12}^* = \sum_{k=1}^K E\left[w_t f_{t-1}(q_{t,\tau_k}(\boldsymbol{\varphi}_0^*))\dot{q}_{t,\tau_k}(\boldsymbol{\varphi}_0^*)\dot{q}_{t,\tau_k}'(\boldsymbol{\varphi}_0^*)\right].$$

Let $\Omega_{1w}^* = \Omega_{12}^* - \Omega_{11}^*$ and $\Sigma_w^* = \Omega_{1w}^{*-1}\Omega_{0w}^*\Omega_{1w}^{*-1}$.

Theorem 4.1. *For $\{y_t\}$ generated by model (2.7) with condition (2.9), suppose $E|y_t|^s < \infty$ for some $s \in (0,1)$ and Σ_w^* is positive definite. If Assumptions 3, 4, 6, 7(i) hold, then as $n \to \infty$, we have (i) $\check{\boldsymbol{\varphi}}_{wn} \to_p \boldsymbol{\varphi}_0^*$. Moreover, if Assumption 7(ii) further holds, then (ii) $\sqrt{n}(\check{\boldsymbol{\varphi}}_{wn} - \boldsymbol{\varphi}_0^*) \to_d N(\mathbf{0}, \Sigma_w^*)$; and (iii) $\sqrt{n}(\check{\boldsymbol{\theta}}_{wn}^*(\tau) - \boldsymbol{\theta}(\tau) - B(\tau)) \to_d N(\mathbf{0}, g_\tau(\boldsymbol{\varphi}_0^*)\Sigma_w^* g_\tau'(\boldsymbol{\varphi}_0^*))$, where $B(\tau) = g_\tau(\boldsymbol{\varphi}_0^*) - \boldsymbol{\theta}(\tau)$ is a systematic bias.*

Theorem 4.1(iii) reveals that $\check{\boldsymbol{\theta}}_{wn}^*(\tau)$ is a biased estimator of $\boldsymbol{\theta}(\tau)$ if $g_\tau(\boldsymbol{\varphi}_0^*) \neq \boldsymbol{\theta}(\tau)$ i.e., when $Q_\tau(y_t|\mathcal{F}_{t-1})$ is misspecified by $q_{t,\tau}(\boldsymbol{\varphi}_0^*)$. Moreover, the systematic bias $B(\tau)$ depends on the bandwidth h, which balances the bias and variance of $\check{\boldsymbol{\theta}}_{wn}^*(\tau)$; see Section 4.3 for details. However, at the cost of introducing the systematic bias, the proposed CQR method can greatly improve the estimation efficiency at high quantile levels, as it overcomes the inefficiency due to data scarcity at tails. Similar to Theorem 3.1, we employ the bracketing method in Pollard (1985) to tackle the non-convexity and non-differentiability of the objective function. However, due to the possible model misspecification, the mixing CLT is used instead of the martingale CLT; see Assumption 6. We will discuss the estimation of the covariance matrix Σ_w^* in the Supplementary Material.

Remark 4.1 (Mixing properties)**.** The proof of the mixing property in Assumption 6 is challenging. For a stationary Markovian process, a common approach to proving that it is geometrically β-mixing and thus α-mixing is to establish its geometric ergodicity (Doukhan, 1994; Francq and Zakoian, 2006). Note that the proposed quantile GARCH process can be regarded as a random-coefficient ARCH(∞) process. However, ARCH(∞) processes are not Markovian in general (Fryzlewicz and Subba Rao, 2011). Thus, the above approach is not feasible. Fryzlewicz and Subba Rao (2011) provides an alternative method to establish mixing properties. By deriving explicit bounds for mixing coefficients using conditional densities of the process, they obtain mixing properties of stationary ARCH(∞) processes and show that the bound on the mixing rate depends on the decay rate of ARCH(∞) parameters. This method potentially can be applied to the quantile GARCH process. However, it is challenging to derive the conditional density of y_{k+s} given $\{\ldots, y_0, U_1, \ldots, U_{k-1}, y_k, \ldots, y_{k+s-1}\}$ due to the random functional coefficients driven by U_t. Thus, we leave this for future research.

4.3 Selection of the bandwidth h

As shown in Theorem 4.1(iii), the bandwidth h plays an important role in balancing the bias and efficiency of the estimator $\check{\boldsymbol{\theta}}^*_{wn}(\tau)$. In the extreme case that $h = 0$, (4.3) will become a weighted quantile regression at the fixed quantile level τ_0, and $\check{\boldsymbol{\theta}}^*_{wn}(\tau_0)$ will be equivalent to the QR estimator $\widetilde{\boldsymbol{\theta}}_{wn}(\tau_0)$. Then we have $g_{\tau_0}(\boldsymbol{\varphi}^*_0) = \boldsymbol{\theta}(\tau_0)$ and $B(\tau_0) = 0$. Although $B(\tau)$ does not have an explicit form with respect to h, our simulation studies show that a larger h usually leads to larger biases but smaller variances of $\check{\boldsymbol{\theta}}^*_{wn}(\tau)$ when the true model is misspecified; see Section 5.4 for details.

In practice, we can treat h as a hyperparameter and search for h that achieves the best forecasting performance from a grid of values via cross-validation. Specifically, we can divide the dataset into training and validation sets, and choose the value of h that minimizes the check loss in the validation set for the target quantile level τ_0:

$$h^{opt} = \underset{h\in(0,d)}{\operatorname{argmin}} \sum_{t=n_0+1}^{n_0+n_1} \rho_\tau \left(y_t - \widetilde{q}_{t,\tau_0}(\check{\boldsymbol{\varphi}}_{wn}(h))\right), \tag{4.6}$$

where n_0 and n_1 are the sample sizes of the training and validation sets, respectively, $\check{\boldsymbol{\varphi}}_{wn}(h)$ is the CQR estimator calculated by (4.3) with bandwidth h, and $d > 0$ determines the range of the grid search. Usually we take d to be a small value such as 0.1 to avoid large biases. The chosen bandwidth h^{opt} will be used to conduct CQR for rolling forecasting of the conditional quantile at time $t = n_0 + n_1 + i$ for any $i \geq 1$.

5 Simulation studies

5.1 Data generating processes

This section conducts simulation experiments to examine the finite sample performance of the proposed estimators and CvM test. The data generating process (DGP) is

$$y_t = \omega(U_t) + \alpha_1(U_t) \sum_{j=1}^{\infty} [\beta_1(U_t)]^{j-1} |y_{t-j}|, \tag{5.1}$$

where $\{U_t\}$ are *i.i.d.* standard uniform random variables. For evaluation of the QR and CQR estimators, we consider two sets of coefficient functions as follows:

$$\omega(\tau) = 0.1F^{-1}(\tau),\ \alpha_1(\tau) = 0.1F^{-1}(\tau),\ \beta_1(\tau) = 0.8, \tag{5.2}$$

and

$$\omega(\tau) = 0.1F^{-1}(\tau),\ \alpha_1(\tau) = \tau - 0.5 + 0.1F^{-1}(\tau),\ \beta_1(\tau) = 0.3 + 0.6|\tau - 0.5|, \quad (5.3)$$

where $F(\cdot)$ is the distribution function of the standard normal distribution or Tukey-lambda distribution in (4.2) with the shape parameter $\lambda = -0.2$, denoted by $F_N(\cdot)$ and $F_T(\cdot)$ respectively. Note that F_T has heavy Pareto tails and does not have the finite fifth moment (Karian et al., 1996). For coefficient functions in (5.3), the strict stationarity condition (2.11) with $s = 1$ in Theorem 2.1 can be verified for $F = F_N$ or F_T by direct calculation or simulating 10^5 random numbers for U_t, respectively. Note that the DGP with coefficient functions in (5.2) is simply the following GARCH(1, 1) process:

$$y_t = \varepsilon_t \left(0.1 + 0.1 \sum_{j=1}^{\infty} 0.8^{j-1} |y_{t-j}| \right),$$

where ε_t follows the distribution F. As a result, the model is correctly specified for the CQR under (5.2) with F being the Tukey-lambda distribution (i.e. $F = F_T$), whereas it is misspecified under all other settings. Two sample sizes, $n = 1000$ and 2000, are considered, and 1000 replications are generated for each sample size.

In addition, for the CvM test in (3.6), we consider the following coefficient functions:

$$\omega(\tau) = 0.1F^{-1}(\tau),\ \alpha_1(\tau) = 0.1F^{-1}(\tau),\ \beta_1(\tau) = 0.3 + d(\tau - 0.5)^2, \quad (5.4)$$

where $d = 0, 1$ or 1.6, and all other settings are the same as those for (5.2). We can similarly verify that the strict stationarity condition holds with $s = 1$ under this setting. Note that the case of $d = 0$ corresponds to the size of the test, whereas the case of $d = 1$ or 1.6 corresponds to the power.

The computation of QR and CQR estimators and the CvM test involves a infinite sum. For computational efficiency, we adopt an exact algorithm based on the fast Fourier transform instead of the standard linear convolution algorithm; see Nielsen and Noël (2021) for details.

5.2 Self-weighted QR estimator

The first experiment focuses on the self-weighted QR estimator $\widetilde{\boldsymbol{\theta}}_{wn}(\tau)$ in Section 3.1. For the estimation of the asymptotic standard deviation (ASD) of $\widetilde{\boldsymbol{\theta}}_{wn}(\tau)$, we employ the two bandwidths (3.3). The resulting ASDs with respect to bandwidths ℓ_B and ℓ_{HS} are denoted by ASD_1 and ASD_2, respectively. Tables 1 and 2 display the biases, empirical

standard deviations (ESDs) and ASDs of $\widetilde{\boldsymbol{\theta}}_{wn}(\tau)$ at quantile level $\tau = 0.5\%, 1\%$ or 5% for (5.2) and (5.3) with F being the standard normal distribution F_N or Tukey-lambda distribution F_T, respectively. We have the following findings. First, as the sample size increases, most of the biases, ESDs and ASDs decrease, and the ESDs get closer to the corresponding ASDs. Secondly, the ASDs calculated using ℓ_{HS} are marginally smaller than those using ℓ_B and closer to the ESDs. Thus, we use the bandwidth ℓ_{HS} in the following for stabler performance. Thirdly, when τ is closer to zero, the performance of $\widetilde{\boldsymbol{\theta}}_{wn}(\tau)$ gets worse with larger biases, ESDs and ASDs, which indicates that the self-weighted QR estimator tends to deteriorate as the target quantile becomes more extreme.

The above results are obtained based on the self-weights in (3.4) with c being the 95% sample quantile of $\{y_t\}_{t=1}^n$. We have also considered the 90% sample quantile for the value of c, and the above findings are unchanged. In addition, simulation results for the unweighted QR estimator are given in the Supplementary Material. It is shown that the unweighted estimator is less efficient than the self-weighted one when $E|y_t|^3 = \infty$.

5.3 The CvM test

The second experiment evaluates the performance of the CvM test in Section 3.2. Since we are particularly interested in the behavior of persistence coefficient function $\beta_1(\tau)$ at tails, we consider $\mathcal{T} = [0.7, 0.995]$ and $[0.8, 0.995]$. To calculate S_n in (3.6), we use a grid $\mathcal{T}_n$ with equal cell size $\delta_n = 0.005$ in place of $\mathcal{T}$. Moreover, ℓ_{HS} in (3.3) is employed to calculate $\widetilde{z}_t(\tau)$ in the subsampling procedure. The rejection rates of S_n at 5% significance level are summarized in Table 3. Firstly, observe that the size is close to the nominal rate when $b_n = \lfloor n^{1/2} \rfloor$. The case with $b_n = \lfloor 0.5n^{1/2} \rfloor$ tends to be undersized, while that with $b_n = \lfloor 2n^{1/2} \rfloor$ tends to be oversized. Secondly, the power generally increases as the sample size n or departure level d increases. Thirdly, a larger subsampling block size b_n or wider interval $\mathcal{T}$ tends to result in a greater power. Hence, we recommend using $b_n = \lfloor n^{1/2} \rfloor$ since it leads to reasonable size and power. For a fixed $\mathcal{T}$, we have also considered other settings for $\mathcal{T}_n$, and the above findings are unchanged. This indicates that the CvM test is not sensitive to the choice of the grid.

5.4 Self-weighted CQR estimator

In the third experiment, we examine the performance of the proposed CQR method in Section 4 via the transformed estimator $g_\tau(\check{\boldsymbol{\varphi}}_{wn}) = \check{\boldsymbol{\theta}}^*_{wn}(\tau) = (\check{\omega}^*_{wn}(\tau), \check{\alpha}^*_{1wn}(\tau), \check{\beta}^{*\prime}_{1wn}(\tau))'$. The DGP is preserved from the first experiment. To obtain the weighted CQR estimator $\check{\boldsymbol{\varphi}}_{wn}$ in (4.3), we let $\mathcal{T}_h = \{\tau_k : \tau_k = \tau_0 + h(k-1)/(K-1)\}_{k=1}^K$, where $K = 19$, $\tau_0 = 0.5\%, 1\%$ or 5% is the target quantile level, and $h > 0$ is the bandwidth.

To investigate the influence of bandwidth h on the CQR, we obtain the estimator $g_\tau(\check{\boldsymbol{\varphi}}_{wn})$ for each $h \in \{0.01, 0.02, \ldots, 0.10\}$ at quantile level $\tau = 0.5\%, 1\%$ or 5% for the DGP in (5.1) with (5.2) or (5.3), $F = F_T$, and sample size $n = 2000$. Figures 1 and 2 illustrate the empirical squared bias, variance and mean squared error (MSE) of $g_\tau(\check{\boldsymbol{\varphi}}_{wn})$ versus h for coefficient functions in (5.2) and (5.3), respectively. Note that the model is correctly specified under coefficient functions in (5.2) with $F = F_T$ and misspecified under (5.3) with $F = F_T$. Figure 1 shows that the squared bias is close to zero, which is because the model is correctly specified. Meanwhile, as h increases, the variance and MSE get smaller, indicating the efficiency gain from using more data for the estimation. On the other hand, Figure 2 shows that a larger h leads to larger biases but smaller variances under model misspecification. Consequently, as h increases, the MSE first decreases and then increases. Moreover, it can be observed that the CQR estimator can have much smaller MSE than the QR estimator (i.e., the case with $h = 0$) especially for the high quantiles. This corroborates the usefulness of the CQR for high quantile levels.

Next we verify the asymptotic results of the CQR estimator by focusing on a fixed bandwidth $h = 0.1$. The ASD of $g_\tau(\check{\boldsymbol{\varphi}}_{wn})$ is calculated based on $\dot{g}_\tau(\check{\boldsymbol{\varphi}}_{wn})\check{\Sigma}^*_w\dot{g}'_\tau(\check{\boldsymbol{\varphi}}_{wn})$, where $\check{\Sigma}^*_w$ is obtained as in Section 2 of the Supplementary Material. Specifically, to estimate Ω^*_{1w}, the bandwidth ℓ_k for quantile level τ_k is set to ℓ_{HS} defined in (3.3) with τ replaced by τ_k. To obtain the kernel estimator $\check{\Omega}^*_{0w}$ in (S.8), we consider the QS kernel in (S.10) with the automatic bandwidth $\widehat{B}_n = 1.3221[n\widehat{\alpha}(2)]^{1/5}$, $0.1\widehat{B}_n$ or $10\widehat{B}_n$ for B_n, where the latter two choices of B_n correspond to under- or over-smoothing in comparison to $\widehat{B}_n$, respectively. The resulting ASDs with respect to $\widehat{B}_n, 0.1\widehat{B}_n$ and $10\widehat{B}_n$ are denoted as ASD_a, ASD_b and ASD_c, respectively. Tables 4 and 5 report the biases, ESDs and ASDs of $g_\tau(\check{\boldsymbol{\varphi}}_{wn})$ for the DGP with coefficient functions in (5.2) and (5.3), respectively. The quantile levels $\tau = 0.5\%, 1\%$ and 5% and distributions $F = F_N$ and F_T are considered.

We first examine the results in Table 4, which corresponds to the DGP with (5.2) and

covers two scenarios: correctly specified (when $F = F_T$) and missspecified (when $F = F_N$) models. For both scenarios, we have three main findings as follows. Firstly, as the sample size increases, most of the biases, ESDs and ASDs become smaller, and the ESDs get closer to the corresponding ASDs. Secondly, as τ approaches zero, the biases, ESDs and ASDs of $\check{\omega}^*_{wn}(\tau)$ and $\check{\alpha}^*_{1wn}(\tau)$ get larger, while that of $\check{\beta}^*_{1wn}(\tau)$ is almost unchanged. This is expected since $\check{\omega}^*_{wn}(\tau)$ and $\check{\alpha}^*_{1wn}(\tau)$ are τ-dependent, and their true values have larger absolute values as τ goes to zero. However, $\check{\beta}^*_{1wn}(\tau) = \check{b}_{1wn}$ is independent of τ. Thirdly, the results of ASD_a, ASD_b and ASD_c are very similar, which suggests that the kernel estimator in (S.8) is insensitive to the selection of bandwidth B_n.

It is also interesting to compare the results under the two scenarios in Table 4. However, it is worth noting that the true values of $\omega(\tau)$ and $\alpha_1(\tau)$ for the correctly specified model (i.e., when $F = F_T$) are larger than those for the misspecified model (i.e., when $F = F_N$) in absolute value. As a result, the absolute biases, ESDs and ASDs of $\widetilde{\omega}_{wn}(\tau)$ and $\widetilde{\alpha}_{1wn}(\tau)$ are much smaller for F_N than that for F_T in Table 4. On the other hand, note that the true values of $\beta_1(\tau)$ are the same for F_N and F_T. Thus, the comparison of the results for $\widetilde{\beta}_{1wn}(\tau)$ under F_N and F_T can directly reveal the effect of model misspecification. Indeed, Table 4 shows that the absolute biases, ESDs and ASDs of $\widetilde{\beta}_{1wn}(\tau)$ for F_T are much smaller than those for F_N. This confirms that the CQR performs better under correct specification (i.e., $F = F_T$) than misspecification (i.e., $F = F_N$).

Note that the above misspecification is only due to the misspecified innovation distribution F, whereas the coefficient function (i.e., model structure) is correctly specified via (5.2). By contrast, the DGP with (5.3) have a misspecified model structure, which is more severe than the former. As a result, Table 5 shares the three main findings from Table 4 for the ESDs and ASDs but not for the biases. In particular, most biases do not decrease as the sample size increases. This is consistent with Theorem 4.1(iii), which shows that $g_\tau(\check{\boldsymbol{\varphi}}_{wn})$ is in general a biased estimator of $\boldsymbol{\theta}(\tau)$ under model misspecification. It also indicates that the misspecification in the model structure is systematic and has greater impact on the bias than that in the innovation distribution F.

We have also considered other choices of the number of quantile levels K and the kernel function $K(\cdot)$. The above findings are unchanged. To save space, these results are omitted.

5.5 Comparison between QR and CQR estimators

We aim to compare the in-sample and out-of-sample performance of QR and CQR in predicting conditional quantiles. The self-weights $\{w_t\}$ in (3.4) are employed for both QR and CQR, and the set $\mathcal{T}_h$ with $K = 19$ and $h = 0.1$ is used for CQR as in the third experiment.

For evaluation of the prediction performance, we use $\widetilde{q}_t(\boldsymbol{\theta}(\tau))$ as the true value of the conditional quantile $Q_\tau(y_t|\mathcal{F}_{t-1})$. Based on the QR estimator $\widetilde{\boldsymbol{\theta}}_{wn}(\tau)$ and the transformed CQR estimator $g_\tau(\check{\boldsymbol{\varphi}}_{wn})$, $Q_\tau(y_t|\mathcal{F}_{t-1})$ can be predicted by $\widetilde{q}_t(\widetilde{\boldsymbol{\theta}}_{wn}(\tau))$ and $\widetilde{q}_t(g_\tau(\check{\boldsymbol{\varphi}}_{wn}))$, respectively. Note that estimates of $Q_\tau(y_t|\mathcal{F}_{t-1})$ for $t = 1, \ldots, n$ are in-sample predictions, and that of $Q_\tau(y_{n+1}|\mathcal{F}_n)$ is the out-of-sample forecast. We measure the in-sample and out-of-sample prediction performance separately, using the biases and RMSEs of conditional quantile estimates by averaging individual values over all time points and replications as follows:

$$\begin{aligned}
\text{Bias}_{In}(\boldsymbol{\theta}_\tau) &= \frac{1}{Mn}\sum_{k=1}^{M}\sum_{t=1}^{n}[\widetilde{q}_t^{(k)}(\boldsymbol{\theta}_\tau) - \widetilde{q}_t^{(k)}(\boldsymbol{\theta}(\tau))], \\
\text{Bias}_{Out}(\boldsymbol{\theta}_\tau) &= \frac{1}{M}\sum_{k=1}^{M}[\widetilde{q}_{n+1}^{(k)}(\boldsymbol{\theta}_\tau) - \widetilde{q}_{n+1}^{(k)}(\boldsymbol{\theta}(\tau))], \\
\text{RMSE}_{In}(\boldsymbol{\theta}_\tau) &= \left\{\frac{1}{Mn}\sum_{k=1}^{M}\sum_{t=1}^{n}[\widetilde{q}_t^{(k)}(\boldsymbol{\theta}_\tau) - \widetilde{q}_t^{(k)}(\boldsymbol{\theta}(\tau))]^2\right\}^{1/2}, \\
\text{RMSE}_{Out}(\boldsymbol{\theta}_\tau) &= \left\{\frac{1}{M}\sum_{k=1}^{M}[\widetilde{q}_{n+1}^{(k)}(\boldsymbol{\theta}_\tau) - \widetilde{q}_{n+1}^{(k)}(\boldsymbol{\theta}(\tau))]^2\right\}^{1/2},
\end{aligned}$$

where $M = 1000$ is the total number of replications, $\widetilde{q}_t^{(k)}(\boldsymbol{\theta}_\tau)$ represents the conditional quantile estimate at time t in the kth replication, and $\boldsymbol{\theta}_\tau$ is the QR estimator $\widetilde{\boldsymbol{\theta}}_{wn}(\tau)$ or the transformed CQR estimator $g_\tau(\check{\boldsymbol{\varphi}}_{wn})$.

Table 6 reports the above measures for the DGP in (5.1) with coefficient functions in (5.2) and (5.3). Firstly, note that most of the biases and RMSEs decrease as the sample size increases. Secondly, the QR and CQR perform similarly for (5.2) and (5.3) with $F = F_N$ in terms of the bias and RMSE. However, when $F = F_T$, obviously the CQR outperforms the QR in biases and RMSEs especially for high quantiles. This confirms that the CQR can be more favorable than the QR at high quantile levels if the data is heavy-tailed, yet can be comparable to the latter if otherwise. This is also consistent with the findings in Figures 1 and 2. Lastly, although the CQR estimator is biased under model misspecification, the biases of its conditional quantile predictions are very

close to or even smaller than those of the QR. This suggests that the CQR can provide satisfactory approximation of conditional quantiles, possibly owing to the flexibility of the Tukey-lambda distribution.

In the Supplementary Material, we also provide a simulation experiment to investigate the effect of quantile rearrangement on the prediction performance.

6 An empirical example

This section analyzes daily log returns of the S&P500 Index based on the proposed quantile GARCH model. The daily closing prices from July 1, 2015 to December 30, 2021, denoted by p_t, are downloaded from the website of Yahoo Finance. Let $y_t = 100\,(\ln p_t - \ln p_{t-1})$ be the log return in percentage, which has $n = 1637$ observations in total. The time plot of $\{y_t\}$ suggests that the series exhibits volatility clustering, and it is very volatile at the beginning of 2020 due to COVID-19 pandemic; see Figure 3. Table 7 displays summary statistics of $\{y_t\}$, where the sample skewness with value -1.053 and kurtosis with value 23.721 indicate that the data are left-skewed and very heavy-tailed. The above findings motivate us to fit $\{y_t\}$ by our proposed quantile GARCH model to capture the conditional heteroscedasticity of the return series and possible asymmetric dynamics over its different quantiles.

We fit a quantile GARCH$(1,1)$ model to $\{y_t\}$. Since the data are very heavy-tailed, the self-weighted QR estimator in (3.1) is used to obtain estimates of $\boldsymbol{\theta}(\tau) = (\omega(\tau), \alpha_1(\tau), \beta_1(\tau))'$, where the self-weights in (3.4) are employed with c being the 95% sample quantile of $\{y_t\}$. The estimates of $\boldsymbol{\theta}(\tau)$ for $\tau \in (0.7, 1)$ together with their 95% pointwise confidence intervals are plotted against the quantile level in Figure 3. Note that $\boldsymbol{\theta}(\tau)$ of our model corresponds to $\boldsymbol{\theta}_\tau = (a_0 Q_\tau(\varepsilon_t)/(1-b_1), a_1 Q_\tau(\varepsilon_t), b_1)$ in the linear GARCH$(1,1)$ model in (2.4). To compare the fitted coefficients of our model with those of model (2.5), we also provide estimates of $\boldsymbol{\theta}_\tau$ using the filtered historical simulation (FHS) method (Kuester et al., 2006) based on the Gaussian quasi-maximum likelihood estimation (QMLE). Specifically, a_0, a_1 and b_1 are estimated by Gaussian QMLE of the linear GARCH$(1,1)$ model in (2.4), and then $Q_\tau(\varepsilon_t)$ is estimated by the empirical quantile of resulting residuals $\{\widehat{\varepsilon}_t\}$.

From Figure 4, we can see that the confidence intervals of $\omega(\tau)$, $\alpha_1(\tau)$ and $\beta_1(\tau)$ do not include the FHS estimates of $\boldsymbol{\theta}_\tau$ for $\tau \in (0.7, 0.8), (0.9, 1)$ and $(0.9, 1)$ respectively. Since

the quantile GARCH model includes the linear GARCH model as a special case, this indicates that the model with constant coefficients fails to capture the asymmetric dynamic structures across different quantiles. In addition, we apply the CvM test in Section 3.2 to check whether $\beta_1(\tau)$ is constant for $\tau \in \mathcal{T}_1 = [0.700, 0.850]$, $\tau \in \mathcal{T}_2 = [0.850, 0.950]$, $\tau \in \mathcal{T}_3 = [0.950, 0.980]$, $\tau \in \mathcal{T}_4 = [0.980, 0.995]$, and $\tau \in \mathcal{T} = [0.700, 0.995] = \cup_{i=1}^4 \mathcal{T}_i$. The CvM test statistic S_n is calculated using a grid $\mathcal{T}_n$ with equal cell size $\delta_n = 0.005$. Its critical value is approximated using the proposed subsampling procedure with $b_n = \lfloor n^{1/2} \rfloor$. The p-values of S_n for $\mathcal{T}_1, \ldots, \mathcal{T}_4$, and $\mathcal{T}$ are 0.585, 0.054, 0.555, 0.017, and 0.150, respectively. Therefore, it is likely that $\beta_1(\tau)$ is varying over $[0.850, 0.950]$ and $[0.980, 0.995]$.

Since the 5% VaR is of common interest in practice, we report the fitted quantile GARCH model at $\tau = 0.05$ as follows:

$$\widetilde{Q}_{0.05}(y_t|\mathcal{F}_{t-1}) = -0.380_{0.100} - 0.341_{0.075} \sum_{j=1}^{\infty} 0.790_{0.033}^{j-1} |y_{t-j}|, \tag{6.1}$$

where the standard errors are given in the corresponding subscripts of the estimated coefficients. We divide the dataset into a training set ($\mathcal{S}_{\text{train}}$) with size $n_0 = 1000$ and a test set ($\mathcal{S}_{\text{test}}$) with size $n - n_0 = 637$. Then we conduct a rolling forecast procedure at level $\tau = 0.05$ (i.e. negative 5% VaR) with a fixed moving window of size n_0 from the forecast origin $t_0 = n_0 + 1$ (June 24, 2019). That is, we first obtain the one-step-ahead conditional quantile forecast for t_0 (i.e., the first time point in $\mathcal{S}_{\text{test}}$) based on data from $t = 1$ to $t = n_0$, using the formula $\widetilde{Q}_{0.05}(y_{t_0}|\mathcal{F}_{n_0}) = \widetilde{\omega}_{wn_0}(0.05) + \widetilde{\alpha}_{1wn_0}(0.05) \sum_{j=1}^{n_0} [\widetilde{\beta}_{1wn_0}(0.05)]^{j-1} |y_{t_0-j}|$. Then for each $i = 1, \ldots, n - n_0 - 1$, we set the forecast origin to $t_0 + i$ and conduct the forecast based on data from $t = 1 + i$ to $t = n_0 + i$. These forecasts are displayed in the time plot in Figure 3. It is clear that the VaR forecasts keep in step with the returns closely, and the return falls below the corresponding negative 5% VaR forecasts occasionally.

We also thoroughly compare the forecasting performance of the proposed model with that of existing conditional quantile estimation methods as follows:

- FHS: The FHS method (Kuester et al., 2006) based on the linear GARCH$(1, 1)$ model in (2.4), where the coefficients are estimated by the Gaussian QMLE, and the residual empirical quantiles are used to approximate the innovation quantiles.

- XK: The two-step estimation method QGARCH2 of Xiao and Koenker (2009) based on linear GARCH(1,1) model (1.1). Specifically, the initial estimates of

$\{h_t\}$ are obtained by combining the conditional quantile estimates of sieve ARCH approximation $h_t = \gamma_0 + \sum_{j=1}^{m} \gamma_j |y_{t-j}|$ over multiple quantile levels, $\tau_k = k/20$ for $k = 1, 2, \ldots, 19$, via the minimum distance estimation. Here we set $m = 3n^{1/4}$ as in their paper.

- Hybrid: The hybrid estimation method proposed in Zheng et al. (2018) based on Bollerslev's GARCH(1, 1) model in (2.1) with $x_t = y_t$.

- CAViaR: The indirect GARCH(1, 1)-based CAViaR method in Engle and Manganelli (2004), where we use the same code and settings for the optimization as in their paper.

We consider the lower and upper $1\%, 2.5\%$ and 5% quantiles and conduct the above rolling forecast procedure for all competing methods. The forecasting performance is evaluated via the empirical coverage rate (ECR), prediction error (PE), and VaR backtests. The ECR is calculated as the percentage of observations in the test set $\mathcal{S}_{\text{test}}$ that fall below the corresponding fitted conditional quantiles. The PE is calculated as follows:

$$PE = \frac{1}{\sqrt{\tau(1-\tau)/(n-n_0)}} \left| \frac{1}{n-n_0} \sum_{t=n_0+1}^{n} I\{y_t < \widehat{Q}_\tau(y_t|\mathcal{F}_{t-1})\} - \tau \right|,$$

where $n-n_0$ is the size of $\mathcal{S}_{\text{test}}$, and $\widehat{Q}_\tau(y_t|\mathcal{F}_{t-1})$ is the one-step-ahead conditional quantile forecast based on each estimation method.

We conduct 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). The null hypothesis of the CC test is that, conditional on $\mathcal{F}_{t-1}$, $\{H_t\}$ are *i.i.d.* Bernoulli random variables with the success probability being τ, where $H_t = I(y_t < Q_\tau(y_t|\mathcal{F}_{t-1}))$ is the hit series. For the DQ test in Engle and Manganelli (2004), we consider the regression of H_t on a constant and four lagged hits $H_{t-\ell}$ with $1 \le \ell \le 4$. The null hypothesis is that the intercept equals to τ and the regression coefficients are zero. If we fail to reject the null hypotheses of the VaR backtests, then the forecasting method is satisfactory. Table 8 reports the ECRs, PEs and p-values of VaR backtests for the one-step-ahead forecasts. In terms of ECRs and backtests, all methods perform reasonably well, since the ECRs are close to the corresponding nominal levels, and at least one backtest is not rejected at the 5% significance level. However, it is clear that the proposed QR estimator has the smallest PEs.

Furthermore, we compare the performance of the proposed self-weighted QR and CQR estimators at high quantile levels, including the lower and upper $0.1\%, 0.25\%$ and 0.5% quantiles. For a more accurate evaluation, we enlarge the S&P500 dataset to cover the period from February 23, 2000 to December 30, 2021, which includes $n = 5500$ observations in total. Moreover, since the self-weighted CQR requires a predetermined bandwidth h, we divide the dataset into a training set ($\mathcal{S}_{\text{train}}$) with size $n_0 = 1000$, a validation set ($\mathcal{S}_{\text{val}}$) with size $n_1 = 500$, and a test set ($\mathcal{S}_{\text{test}}$) with size $n_2 = n - n_0 - n_1$. We choose the optimal h that minimizes the check loss in (4.6) for $\mathcal{S}_{\text{val}}$; see Section 4.3 for details. Then based on the chosen h, we conduct a moving-window rolling forecast procedure similar to the previous one. The window size is n_0, and the forecast origin is $t_0 = n_0 + n_1 + 1 = 1501$. That is, we first obtain the conditional quantile forecast for t_0 (i.e., the first time point in $\mathcal{S}_{\text{test}}$) based on data from $t = t_0 - n_0 = 501$ to $t = t_0 - 1 = 1500$ (i.e., the last 500 observations in $\mathcal{S}_{\text{train}}$ and all observations in $\mathcal{S}_{\text{val}}$). We repeat this procedure by advancing the forecast origin and moving window until the end of $\mathcal{S}_{\text{test}}$ is reached. Table 9 displays the results for the proposed QR, CQR and other competing methods. Notably, the CQR method has the smallest PE and the most accurate ECR at almost all quantile levels, while the QR method is generally competitive among the other methods. In summary, for the S&P 500 dataset, the proposed quantile GARCH model has superior forecasting performance than the original GARCH model, and the proposed CQR estimator outperforms the QR estimator at high quantile levels.

Finally, to remedy the quantile crossing problem, we have further conducted the quantile rearrangement (Chernozhukov et al., 2010) for the proposed QR method. There are only inconsequential changes to Tables 8 and 9, while all main findings summarized earlier remain the same. In addition, for Figure 4, we can also rearrange the self-weighted QR estimates $\{\widetilde{\omega}_{wn}(\tau_k)\}_{k=1}^K$ and $\{\widetilde{\alpha}_{1wn}(\tau_k)\}_{k=1}^K$ to ensure the monotonicity of the curves. After the rearrangement, the curves for $\omega(\cdot)$ and $\alpha_1(\cdot)$ become smoother than those in Figure 4. The corresponding confidence intervals are slightly narrower than the original ones; see Section 4 of the Supplementary Material for details.

7 Conclusion and discussion

This paper proposes the quantile GARCH model, a new conditional heteroskedastic model whose coefficients are functions of a standard uniform random variable. A suf-

ficient condition for the strict stationarity of this model is derived. To estimate the unknown coefficient functions without any moment restriction on the data, we develop the self-weighted QR and CQR methods. By efficiently borrowing information from intermediate quantile levels via a flexible parametric approximation, the CQR method is more favorable than the QR at high quantile levels. Our empirical analysis shows that the proposed approach can provide more accurate conditional quantile forecasts at high or even extreme quantile levels than existing ones.

The proposed approach can be improved and extended in the following directions. Firstly, the estimation of the asymptotic covariance matrices for the QR and CQR estimator are complicated due to the unknown conditional density function. As an alternative to the kernel density estimation, an easy-to-use bootstrap method such as the block bootstrap and random-weight bootstrap may be developed, and asymptotically valid bootstrap inference for the estimated coefficient functions and conditional quantiles can be further studied. Secondly, it is worth investigating whether it is possible to construct a debiased CQR estimator that is provably no less efficient than the proposed biased estimator at high quantile levels. Thirdly, the expected shortfall, defined as the expectation of the loss that exceeds the VaR, is another important risk measure. It is also of interest to forecast the ES based on the proposed quantile GARCH model. Lastly, the parametric method to model the tails based on the flexible Tukey-lambda distribution is efficient and computationally simple. It can be generalized to other high quantile estimation problems for various data settings.

8 Supplementary material

The Supplementary Material contains the generalization to the quantile GARCH(p, q) model, all technical proofs, and additional results for the numerical studies in this paper.

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. Zhu's research was supported by an NSFC grant 12001355. Li's research was partially supported by a Hong Kong RGC grant 17306121 and an NSFC

grant 72033002.

References

Artzner, P., Delbaen, F., Eber, J. M., and Heath, D. (1999). Coherent measures of risk. *Mathematical Finance*, 9:203–228.

Baur, D. G., Dimpfl, T., and Jung, R. C. (2012). Stock return autocorrelations revisited: a quantile regression approach. *Journal of Empirical Finance*, 19:254–265.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroscedasticity. *Journal of Econometrics*, 31:307–327.

Cai, Y. and Stander, J. (2008). Quantile self-exciting threshold autoregressive time series models. *Journal of Time Series Analysis*, 29:186–202.

Chernozhukov, V. (2005). Extremal quantile regression. *The Annals of Statistics*, 33:806–839.

Chernozhukov, V., Fernández-Val, I., and Galichon, A. (2009). Improving point and interval estimators of monotone functions by rearrangement. *Biometrika*, 96:559–575.

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

Chernozhukov, V. and Hansen, C. (2006). Instrumental quantile regression inference for structural and treatment effect models. *Journal of Econometrics*, 132:491–525.

Christoffersen, P. (1998). Evaluating interval forecasts. *International Economic Review*, 39:841–862.

Douc, R., Roueff, F., and Soulier, P. (2008). On the existence of some ARCH(∞) processes. *Stochastic Processes and their Applications*, 118:755–761.

Doukhan, P. (1994). *Mixing: Properties and Examples.* Lecture Notes in Statistics 85, Berlin: Springer.

Durbin, J. (1973). Weak convergence of the sample distribution function when parameters are estimated. *The Annals of Statistics*, 1:279–290.

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

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.

Fan, J. and Yao, Q. (2003). *Nonlinear Time Series: Nonparametric and Parametric Methods*. Springer, New York.

Ferreira, M. S. (2011). Capturing asymmetry in real exchange rate with quantile autoregression. *Applied Economics*, 43:327–340.

Francq, C. and Zakoian, J.-M. (2006). Mixing properties of a general class of garch(1,1) models without moment assumptions on the observed process. *Econometric Theory*, 22:815–834.

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

Francq, C. and Zakoian, J.-M. (2015). Risk-parameter estimation in volatility models. *Journal of Econometrics*, 184:158–174.

Fryzlewicz, P. and Subba Rao, S. (2011). Mixing properties of arch and time-varying arch processes. *Bernoulli*, 17:320–346.

Galvao, A. F., Montes-Rojas, G., and Olmo, J. (2011). Threshold quantile autoregressive models. *Journal of Time Series Analysis*, 32:253–267.

Gilchrist, W. G. (2000). *Statistical Modelling with Quantile Functions*. CHAPMAN & HALL/CRC.

Giraitis, L., Kokoszka, P., and Leipus, R. (2000). Stationary arch models: Dependence structure and central limit theorem. *Econometric Theory*, 16:322.

He, Y., Hou, Y., Peng, L., and Shen, H. (2020). Inference for conditional value-at-risk of a predictive regression. *The Annals of Statistics*, 48:3442–3464.

Joiner, B. L. and Rosenblatt, J. R. (1971). Some properties of the range in samples from tukey's symmetric lambda distributions. *Journal of the American Statistical Association*, 66:394–399.

Karian, Z. A., Dudewicz, E. J., and Mcdonald, P. (1996). The extended generalized lambda distribution system for fitting distributions to data: history, completion of theory, tables, applications, the "final word" on moment fits. *Communications in Statistics-Simulation and Computation*, 25:611–642.

Koenker, R. (2005). *Quantile regression*. Cambridge University Press, Cambridge.

Koenker, R. and Xiao, Z. (2006). Quantile autoregression. *Journal of the American Statistical Association*, 101:980–990.

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.

Lee, S. and Noh, J. (2013). Quantile regression estimator for GARCH models. *Scandinavian Journal of Statistics*, 40:2–20.

Li, D. and Wang, H. J. (2019). Extreme quantile estimation for autoregressive models. *Journal of Bussiness and Economic Statistics*, 37:661–670.

Ling, S. (2005). Self-weighted least absolute deviation estimation for infinite variance autoregressive models. *Journal of the Royal Statistical Society: Series B*, 67:381–393.

Nelson, D. B. (1991). Conditional heteroskedasticity in asset returns: a new approach. *Econometrica*, 59:347–370.

Nielsen, M. Ø. and Noël, A. L. (2021). To infinity and beyond: Efficient computation of ARCH(∞) models. *Journal of Time Series Analysis*, 42:338–354.

Phillips, P. C. B. (2015). Halbert White Jr. memorial JFEC lecture: Pitfalls and possibilities in predictive regression. *Journal of Financial Econometrics*, 13:521–555.

Pollard, D. (1985). New ways to prove central limit theorems. *Econometric Theory*, 1:295–314.

Royer, J. (2022). Conditional asymmetry in Power ARCH(∞) models. *Journal of Econometrics*.

Shao, X. (2011). A bootstrap-assisted spectral test of white noise under unknown dependence. *Journal of Econometrics*, 162:213–224.

Taylor, S. J. (2008). *Modelling financial time series*. World Scientific, New York.

Wang, G., Zhu, K., Li, G., and Li, W. K. (2022). Hybrid quantile estimation for asymmetric power GARCH models. *Journal of Econometrics*, 227:264–284.

Wang, H. J., Li, D., and He, X. (2012). Estimation of high conditional quantiles for heavy-tailed distributions. *Journal of the American Statistical Association*, 107:1453–1464.

Wu, G. and Xiao, Z. (2002). An analysis of risk measures. *Journal of Risk*, 4:53–75.

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

Zaffaroni, P. (2004). Stationarity and memory of ARCH(∞) models. *Econometric Theory*, 20:147–160.

Zakoian, J. M. (1994). Threshold heteroskedastic models. *Journal of Economic Dynamics and Control*, 18:931–955.

Zheng, Y., Zhu, Q., Li, G., and Xiao, Z. (2018). Hybrid quantile regression estimation for time series models with conditional heteroscedasticity. *Journal of the Royal Statistical Society: Series B*, 80:975–993.

Zhu, C., Byrd, R. H., Lu, P., and Nocedal, J. (1997). Algorithm 778: L-BFGS-B: Fortran subroutines for large-scale bound-constrained optimization. *ACM Transactions on mathematical software (TOMS)*, 23:550–560.

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

Zhu, Q. and Li, G. (2022). Quantile double autoregression. *Econometric Theory*, 38:793–839.

Zhu, Q., Li, G., and Xiao, Z. (2021). Quantile estimation of regression models with GARCH-X errors. *Statistica Sinica*, 31:1261–1284.

Zhu, Q., Zheng, Y., and Li, G. (2018). Linear double autoregression. *Journal of Econometrics*, 207:162–174.

Zou, H. and Yuan, M. (2008). Composite quantile regression and the oracle model selection theory. *The Annals of Statistics*, 36:1108–1126.

Table 1: Biases, ESDs and ASDs of the self-weighted QR estimator $\widetilde{\boldsymbol{\theta}}_{wn}(\tau)$ at quantile level $\tau = 0.5\%, 1\%$ or 5% for DGP (5.1) with Setting (5.2). ASD_1 and ASD_2 correspond to the bandwidths ℓ_B and ℓ_{HS}, respectively. F is the standard normal distribution F_N or Tukey-lambda distribution F_T.

		$F = F_N$					$F = F_T$				
	n	True	Bias	ESD	ASD_1	ASD_2	True	Bias	ESD	ASD_1	ASD_2
						$\tau = 0.5\%$					
ω	1000	-0.258	-0.002	0.073	0.099	0.066	-0.942	-0.391	1.011	1.573	1.207
	2000	-0.258	-0.002	0.058	0.066	0.049	-0.942	-0.204	0.721	1.055	0.787
α_1	1000	-0.258	-0.031	0.151	0.212	0.151	-0.942	-0.159	0.609	0.808	0.598
	2000	-0.258	-0.027	0.128	0.144	0.109	-0.942	-0.104	0.421	0.530	0.376
β_1	1000	0.800	-0.052	0.141	0.257	0.169	0.800	-0.043	0.114	0.169	0.118
	2000	0.800	-0.043	0.129	0.159	0.114	0.800	-0.027	0.082	0.103	0.071
						$\tau = 1\%$					
ω	1000	-0.233	-0.007	0.062	0.069	0.052	-0.755	-0.270	0.726	0.911	0.712
	2000	-0.233	-0.004	0.049	0.050	0.037	-0.755	-0.178	0.510	0.629	0.495
α_1	1000	-0.233	-0.023	0.135	0.160	0.122	-0.755	-0.095	0.377	0.472	0.332
	2000	-0.233	-0.016	0.098	0.111	0.084	-0.755	-0.060	0.271	0.306	0.234
β_1	1000	0.800	-0.056	0.145	0.232	0.169	0.800	-0.033	0.093	0.118	0.084
	2000	0.800	-0.042	0.127	0.140	0.101	0.800	-0.020	0.068	0.075	0.057
						$\tau = 5\%$					
ω	1000	-0.164	-0.008	0.038	0.040	0.033	-0.405	-0.130	0.315	0.305	0.257
	2000	-0.164	-0.004	0.030	0.030	0.027	-0.405	-0.063	0.202	0.218	0.185
α_1	1000	-0.164	-0.015	0.085	0.090	0.077	-0.405	-0.029	0.135	0.144	0.118
	2000	-0.164	-0.008	0.060	0.063	0.057	-0.405	-0.016	0.090	0.098	0.083
β_1	1000	0.800	-0.063	0.156	0.178	0.150	0.800	-0.022	0.065	0.069	0.055
	2000	0.800	-0.033	0.109	0.105	0.093	0.800	-0.011	0.042	0.046	0.038

Table 2: Biases, ESDs and ASDs of the self-weighted QR estimator $\widetilde{\boldsymbol{\theta}}_{wn}(\tau)$ at quantile level $\tau = 0.5\%, 1\%$ or 5% for DGP (5.1) with Setting (5.3). ASD_1 and ASD_2 correspond to the bandwidths ℓ_B and ℓ_{HS}, respectively. F is the standard normal distribution F_N or Tukey-lambda distribution F_T.

		$F = F_N$					$F = F_T$				
	n	True	Bias	ESD	ASD_1	ASD_2	True	Bias	ESD	ASD_1	ASD_2
						$\tau = 0.5\%$					
ω	1000	-0.258	-0.016	0.058	0.063	0.042	-0.942	-0.142	0.496	0.689	0.485
	2000	-0.258	-0.007	0.040	0.038	0.028	-0.942	-0.100	0.365	0.440	0.321
α_1	1000	-0.753	-0.002	0.120	0.145	0.103	-1.437	-0.053	0.476	0.621	0.477
	2000	-0.753	-0.006	0.088	0.096	0.074	-1.437	-0.033	0.346	0.431	0.306
β_1	1000	0.597	-0.024	0.079	0.086	0.061	0.597	-0.028	0.112	0.149	0.109
	2000	0.597	-0.013	0.053	0.053	0.040	0.597	-0.019	0.085	0.100	0.070
						$\tau = 1\%$					
ω	1000	-0.233	-0.012	0.046	0.043	0.033	-0.755	-0.110	0.334	0.396	0.288
	2000	-0.233	-0.006	0.032	0.031	0.024	-0.755	-0.058	0.247	0.278	0.211
α_1	1000	-0.723	-0.006	0.106	0.114	0.089	-1.245	-0.041	0.330	0.402	0.285
	2000	-0.723	-0.003	0.077	0.083	0.067	-1.245	-0.018	0.240	0.266	0.202
β_1	1000	0.594	-0.021	0.070	0.069	0.053	0.594	-0.025	0.096	0.110	0.078
	2000	0.594	-0.010	0.049	0.048	0.038	0.594	-0.011	0.068	0.071	0.053
						$\tau = 5\%$					
ω	1000	-0.164	-0.006	0.033	0.032	0.029	-0.405	-0.034	0.147	0.157	0.130
	2000	-0.164	-0.002	0.022	0.023	0.021	-0.405	-0.013	0.100	0.109	0.096
α_1	1000	-0.614	-0.000	0.093	0.096	0.087	-0.855	-0.009	0.150	0.159	0.136
	2000	-0.614	-0.003	0.066	0.068	0.063	-0.855	-0.006	0.105	0.110	0.098
β_1	1000	0.570	-0.012	0.071	0.073	0.065	0.570	-0.011	0.069	0.072	0.061
	2000	0.570	-0.006	0.049	0.051	0.046	0.570	-0.005	0.047	0.050	0.044

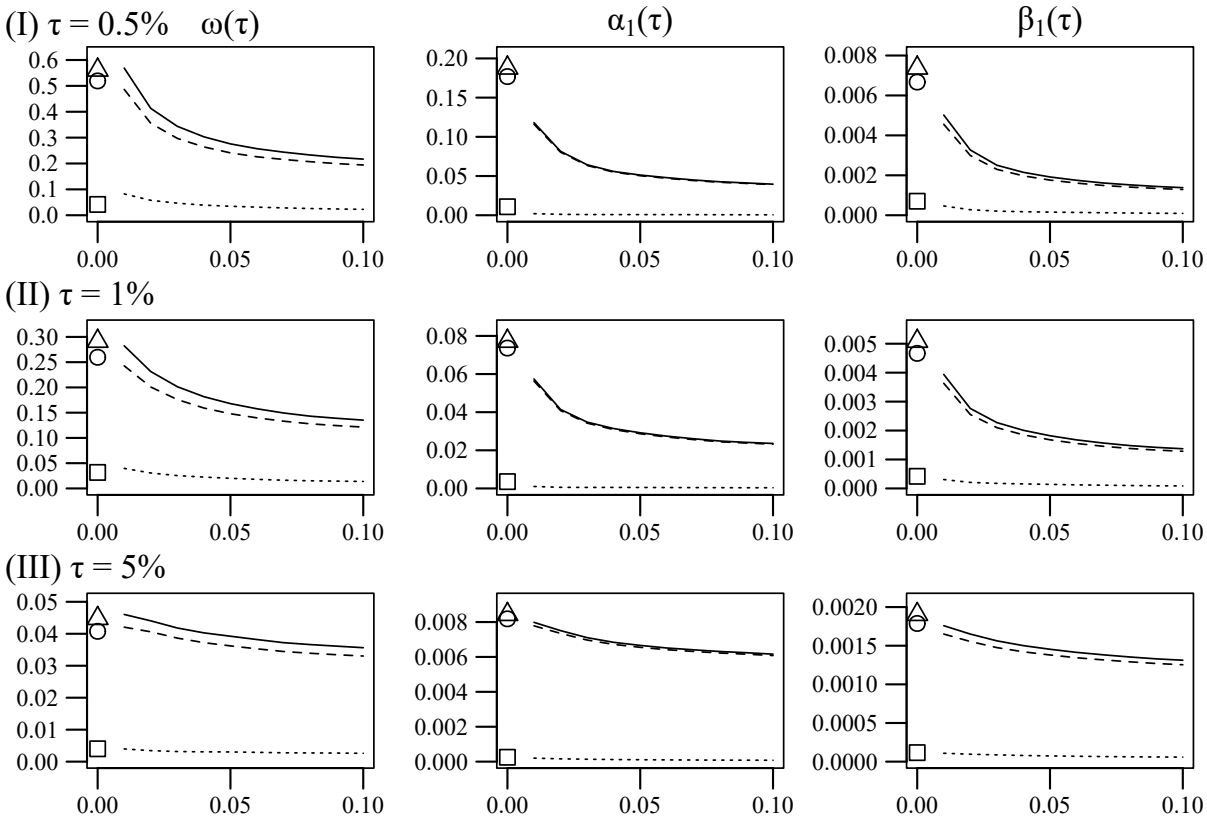

Figure 1: Empirical squared bias (dotted line), variance (dashed line) and MSE (solid line) of the transformed CQR estimator $\check{\boldsymbol{\theta}}^*_{wn}(\tau)$ versus the bandwidth h at quantile level $\tau = 0.5\%, 1\%$ or 5% for DGP (5.1) with Setting (5.2) and F being the Tukey-lambda distribution F_T. Empirical squared bias (square), variance (circle) and MSE (triangle) of the QR estimator are also labeled at $h = 0$ for comparison.

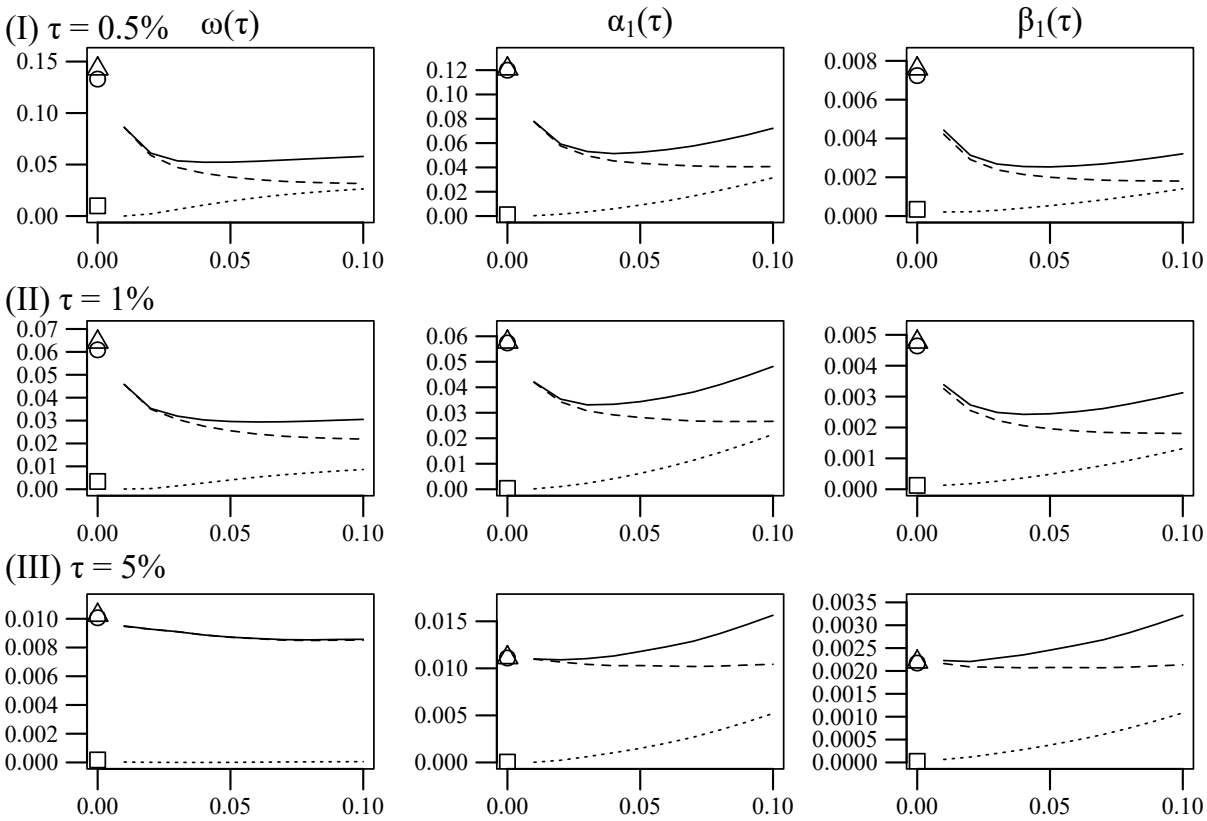

Figure 2: Empirical squared bias (dotted line), variance (dashed line) and MSE (solid line) of the transformed CQR estimator $\check{\boldsymbol{\theta}}^*_{wn}(\tau)$ versus the bandwidth h at quantile level $\tau = 0.5\%, 1\%$ or 5% for DGP (5.1) with Setting (5.3) and F being the Tukey-lambda distribution F_T. Empirical squared bias (square), variance (circle) and MSE (triangle) of the QR estimator are also labeled at $h = 0$ for comparison.

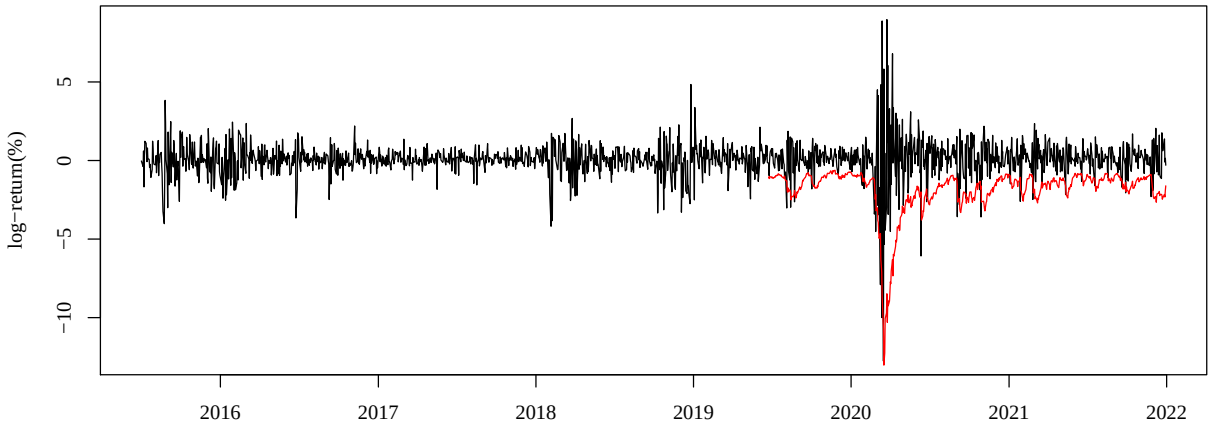

Figure 3: Time plot for daily log returns in percentage (black line) of S&P500 Index from July 2, 2015 to December 30, 2021, with negative 5% VaR forecasts (red line) from June 24, 2019 to December 30, 2021.

Table 3: Rejection rates of the CvM test at the 5% significance level for $\mathcal{T} = [0.7, 0.995]$ and $[0.8, 0.995]$, where b_1, b_2 and b_3 correspond to $\lfloor cn^{1/2} \rfloor$ with $c = 0.5, 1$ and 2, respectively. F is the standard normal distribution F_N or Tukey-lambda distribution F_T.

			$F = F_N$			$F = F_T$		
$\mathcal{T}$	n	d	b_1	b_2	b_3	b_1	b_2	b_3
		0	0.045	0.058	0.084	0.028	0.041	0.056
	1000	1	0.101	0.116	0.140	0.098	0.122	0.167
$[0.7, 0.995]$		1.6	0.236	0.278	0.332	0.259	0.325	0.403
		0	0.047	0.055	0.064	0.034	0.049	0.062
	2000	1	0.190	0.214	0.236	0.240	0.274	0.334
		1.6	0.571	0.597	0.656	0.689	0.739	0.784
		0	0.036	0.046	0.071	0.026	0.040	0.054
	1000	1	0.071	0.093	0.124	0.061	0.073	0.121
$[0.8, 0.995]$		1.6	0.169	0.223	0.284	0.147	0.191	0.274
		0	0.028	0.037	0.056	0.027	0.038	0.055
	2000	1	0.143	0.161	0.202	0.131	0.173	0.213
		1.6	0.481	0.558	0.601	0.473	0.530	0.610

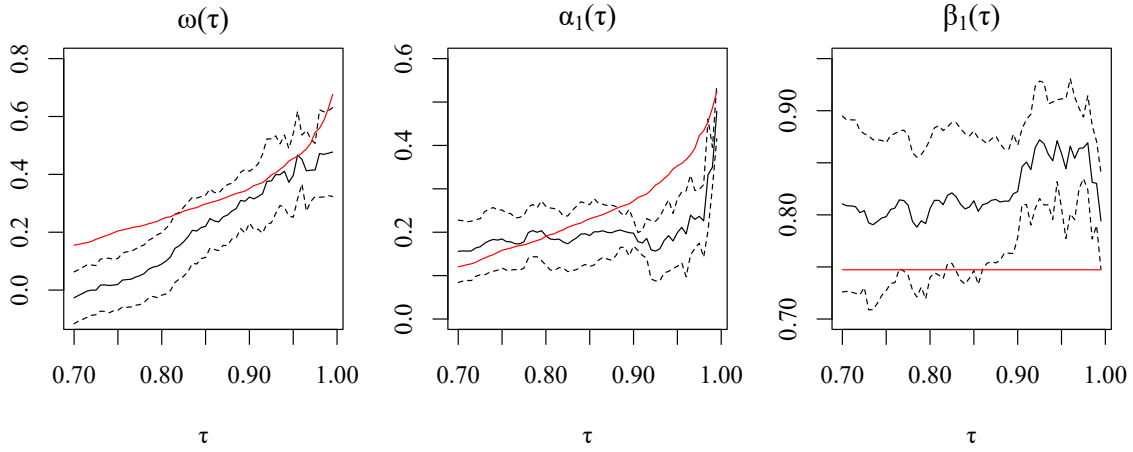

Figure 4: Self-weighted QR estimates of $\boldsymbol{\theta}(\tau) = (\omega(\tau), \alpha_1(\tau), \beta_1(\tau))'$ (black solid), together with their 95% confidence intervals (black dotted) at $\tau_k = k/200$ with $140 \leq k \leq 199$, and estimates of $\boldsymbol{\theta}_\tau = (a_0 Q_\tau(\varepsilon_t)/(1-b_1), a_1 Q_\tau(\varepsilon_t), b_1)$ (red solid) for the linear ARCH(∞) model in (2.4) using the FHS method.

Table 4: Biases, ESDs and ASDs of the transformed CQR estimator $\check{\boldsymbol{\theta}}^*_{wn}(\tau)$ with bandwidth $h = 0.1$, at quantile level $\tau = 0.5\%, 1\%$ or 5% for DGP (5.1) with Setting (5.2). ASD_a, ASD_b and ASD_c correspond to the optimal, under-smoothing and over-smoothing bandwidths $\widehat{B}_n, 0.1\widehat{B}_n$ and $10\widehat{B}_n$, respectively. F is the standard normal distribution F_N or Tukey-lambda distribution F_T.

		$F = F_N$						$F = F_T$					
	n	True	Bias	ESD	ASD_a	ASD_b	ASD_c	True	Bias	ESD	ASD_a	ASD_b	ASD_c
						$\tau = 0.5\%$							
ω	1000	-0.258	-0.009	0.054	0.061	0.061	0.060	-0.942	-0.331	0.697	0.642	0.640	0.639
	2000	-0.258	-0.004	0.041	0.045	0.045	0.045	-0.942	-0.151	0.441	0.437	0.437	0.436
α_1	1000	-0.258	-0.020	0.122	0.127	0.127	0.125	-0.942	-0.047	0.303	0.315	0.316	0.307
	2000	-0.258	-0.008	0.085	0.088	0.088	0.087	-0.942	-0.023	0.198	0.208	0.208	0.206
β_1	1000	0.800	-0.059	0.144	0.153	0.153	0.150	0.800	-0.020	0.055	0.058	0.058	0.057
	2000	0.800	-0.029	0.098	0.095	0.095	0.095	0.800	-0.010	0.036	0.038	0.038	0.037
						$\tau = 1\%$							
ω	1000	-0.233	-0.008	0.048	0.053	0.053	0.053	-0.755	-0.256	0.542	0.499	0.498	0.496
	2000	-0.233	-0.004	0.037	0.041	0.041	0.041	-0.755	-0.117	0.348	0.343	0.343	0.342
α_1	1000	-0.233	-0.016	0.110	0.112	0.112	0.111	-0.755	-0.037	0.229	0.238	0.239	0.232
	2000	-0.233	-0.008	0.078	0.080	0.080	0.079	-0.755	-0.019	0.152	0.159	0.159	0.157
β_1	1000	0.800	-0.054	0.140	0.134	0.134	0.133	0.800	-0.019	0.055	0.057	0.057	0.056
	2000	0.800	-0.029	0.100	0.093	0.093	0.092	0.800	-0.009	0.036	0.037	0.038	0.037
						$\tau = 5\%$							
ω	1000	-0.164	-0.007	0.036	0.053	0.053	0.052	-0.405	-0.114	0.281	0.249	0.249	0.246
	2000	-0.164	-0.004	0.028	0.030	0.030	0.030	-0.405	-0.051	0.182	0.175	0.175	0.174
α_1	1000	-0.164	-0.015	0.086	0.121	0.119	0.117	-0.405	-0.016	0.112	0.114	0.114	0.110
	2000	-0.164	-0.007	0.060	0.060	0.060	0.059	-0.405	-0.009	0.078	0.078	0.078	0.077
β_1	1000	0.800	-0.063	0.160	0.304	0.298	0.292	0.800	-0.016	0.055	0.054	0.054	0.052
	2000	0.800	-0.033	0.112	0.100	0.100	0.099	0.800	-0.008	0.035	0.036	0.036	0.035

Table 5: Biases, ESDs and ASDs of the transformed CQR estimator $\check{\boldsymbol{\theta}}^*_{wn}(\tau)$ with bandwidth $h = 0.1$, at quantile level $\tau = 0.5\%, 1\%$ or 5% for DGP (5.1) with Setting (5.3). ASD_a, ASD_b and ASD_c correspond to the optimal, under-smoothing, and over-smoothing bandwidths $\widehat{B}_n, 0.1\widehat{B}_n$ and $10\widehat{B}_n$, respectively. F is the standard normal distribution F_N or Tukey-lambda distribution F_T.

		$F = F_N$						$F = F_T$					
	n	True	Bias	ESD	ASD_a	ASD_b	ASD_c	True	Bias	ESD	ASD_a	ASD_b	ASD_c
						$\tau = 0.5\%$							
ω	1000	-0.258	0.017	0.043	0.040	0.040	0.040	-0.942	0.126	0.260	0.244	0.243	0.240
	2000	-0.258	0.022	0.029	0.028	0.028	0.028	-0.942	0.162	0.178	0.169	0.169	0.167
α_1	1000	-0.753	-0.115	0.122	0.115	0.115	0.114	-1.437	-0.174	0.295	0.277	0.277	0.270
	2000	-0.753	-0.119	0.085	0.081	0.081	0.081	-1.437	-0.177	0.202	0.192	0.192	0.190
β_1	1000	0.597	-0.045	0.067	0.062	0.062	0.061	0.597	-0.042	0.063	0.059	0.059	0.058
	2000	0.597	-0.038	0.045	0.043	0.043	0.042	0.597	-0.037	0.042	0.041	0.041	0.040
						$\tau = 1\%$							
ω	1000	-0.233	0.009	0.040	0.037	0.037	0.037	-0.755	0.062	0.217	0.202	0.202	0.199
	2000	-0.233	0.014	0.027	0.026	0.026	0.026	-0.755	0.093	0.148	0.140	0.140	0.139
α_1	1000	-0.723	-0.094	0.114	0.110	0.110	0.109	-1.245	-0.146	0.239	0.228	0.228	0.222
	2000	-0.723	-0.096	0.080	0.078	0.078	0.077	-1.245	-0.147	0.163	0.158	0.158	0.156
β_1	1000	0.594	-0.044	0.069	0.063	0.063	0.062	0.594	-0.041	0.063	0.059	0.059	0.058
	2000	0.594	-0.037	0.046	0.044	0.044	0.043	0.594	-0.036	0.043	0.041	0.041	0.040
						$\tau = 5\%$							
ω	1000	-0.164	-0.003	0.034	0.032	0.032	0.032	-0.405	-0.009	0.138	0.125	0.125	0.123
	2000	-0.164	0.000	0.023	0.023	0.023	0.023	-0.405	0.007	0.092	0.087	0.088	0.087
α_1	1000	-0.614	-0.047	0.104	0.105	0.105	0.103	-0.855	-0.070	0.145	0.147	0.148	0.143
	2000	-0.614	-0.049	0.073	0.074	0.074	0.074	-0.855	-0.072	0.102	0.104	0.104	0.102
β_1	1000	0.570	-0.043	0.085	0.076	0.076	0.075	0.570	-0.038	0.069	0.063	0.063	0.061
	2000	0.570	-0.034	0.055	0.053	0.053	0.052	0.570	-0.033	0.046	0.043	0.043	0.043

Table 6: Biases and RMSEs for conditional quantile estimates of the QR and CQR with bandwidth $h = 0.1$, at quantile level $\tau = 0.5\%, 1\%$ or 5% for DGP (5.1) with Settings (5.2) and (5.3). F is the standard normal distribution F_N or Tukey-lambda distribution F_T.

			DGP1				DGP2			
			Bias		RMSE		Bias		RMSE	
F	n	Method	In	Out	In	Out	In	Out	In	Out
						$\tau = 0.5\%$				
F_N	1000	QR	0.000	-0.001	0.039	0.038	0.003	0.001	0.057	0.062
	1000	CQR	0.006	0.005	0.034	0.034	-0.004	-0.005	0.060	0.062
	2000	QR	0.000	0.000	0.029	0.029	0.001	0.000	0.040	0.038
	2000	CQR	0.003	0.003	0.024	0.023	-0.007	-0.005	0.049	0.053
F_T	1000	QR	-0.423	-0.544	14.818	10.154	-0.007	0.358	24.762	13.228
	1000	CQR	-0.265	-0.101	9.607	3.810	-0.052	0.169	8.168	6.090
	2000	QR	-0.240	-0.181	8.028	6.216	0.020	0.038	16.072	1.978
	2000	CQR	-0.089	-0.082	5.500	2.782	-0.073	-0.014	4.051	2.115
						$\tau = 1\%$				
F_N	1000	QR	0.000	-0.001	0.031	0.032	0.001	0.001	0.048	0.050
	1000	CQR	0.004	0.003	0.029	0.028	-0.004	-0.004	0.053	0.053
	2000	QR	0.000	-0.000	0.023	0.022	0.000	0.001	0.034	0.034
	2000	CQR	0.002	0.001	0.020	0.019	-0.005	-0.004	0.042	0.046
F_T	1000	QR	-0.338	-0.401	6.027	7.031	-0.022	0.052	17.539	6.121
	1000	CQR	-0.235	-0.096	6.139	2.938	-0.105	0.059	7.412	3.842
	2000	QR	-0.157	-0.061	4.591	2.476	-0.013	0.048	9.131	1.705
	2000	CQR	-0.067	-0.059	4.096	2.003	-0.107	-0.064	3.838	1.690
						$\tau = 5\%$				
F_N	1000	QR	-0.001	-0.001	0.019	0.019	-0.000	0.001	0.038	0.040
	1000	CQR	0.001	0.001	0.020	0.019	-0.000	0.001	0.042	0.043
	2000	QR	-0.000	-0.001	0.014	0.014	-0.000	-0.001	0.027	0.031
	2000	CQR	0.000	-0.000	0.014	0.014	-0.002	-0.001	0.031	0.033
F_T	1000	QR	-0.151	-0.019	2.454	1.652	-0.042	0.052	4.947	1.792
	1000	CQR	-0.079	0.035	1.922	1.853	-0.052	0.007	7.302	1.268
	2000	QR	-0.057	-0.052	1.604	0.925	-0.029	-0.020	2.204	0.913
	2000	CQR	-0.038	-0.015	1.387	0.871	-0.060	-0.035	3.092	0.889

Table 7: Summary statistics for S&P500 returns.

Mean	Median	Std.Dev.	Skewness	Kurtosis	Min	Max
0.051	0.074	1.161	-1.053	23.721	-12.765	8.968

Table 8: Empirical coverage rates (ECRs) in percentage, prediction errors (PEs) and p-values for correct conditional coverage (CC) and the dynamic quantile (DQ) tests for five estimation methods at lower and upper $1\%, 2.5\%, 5\%$ quantile levels. The ECR closest to the nominal level τ and the smallest PE are marked in bold.

τ		QR	FHS	XK	Hybrid	CAViaR
1%	ECR	1.26	**1.10**	1.26	1.26	1.26
	PE	0.65	**0.25**	0.65	0.65	0.65
	CC test	0.74	0.90	0.74	0.74	0.74
	DQ test	0.96	0.99	0.05	0.96	0.96
2.5%	ECR	**2.98**	3.30	3.14	3.14	3.45
	PE	**0.78**	1.29	1.03	1.03	1.54
	CC test	0.42	0.44	0.32	0.32	0.33
	DQ test	0.75	0.12	0.26	0.72	0.72
5%	ECR	6.12	6.12	6.12	**5.65**	5.97
	PE	1.30	1.30	1.30	**0.75**	1.12
	CC test	0.42	0.27	0.27	0.32	0.30
	DQ test	0.01	0.05	0.00	0.36	0.58
95%	ECR	94.51	95.92	92.94	94.19	**95.45**
	PE	0.57	1.06	2.39	0.94	**0.52**
	CC test	0.11	0.18	0.06	0.07	0.22
	DQ test	0.62	0.70	0.08	0.44	0.67
97.5%	ECR	**97.65**	97.96	96.86	97.80	**97.65**
	PE	**0.23**	0.74	1.03	0.49	**0.23**
	CC test	0.68	0.57	0.55	0.64	0.68
	DQ test	0.68	0.72	0.64	0.79	0.90
99%	ECR	**99.06**	98.90	99.22	98.90	98.74
	PE	**0.15**	0.25	0.55	0.25	0.65
	CC test	0.93	0.90	0.82	0.90	0.74
	DQ test	1.00	0.99	0.99	0.99	0.96

Table 9: Empirical coverage rates (ECRs) in percentage and prediction errors (PEs) for six estimation methods at lower and upper $0.1\%, 0.25\%, 0.5\%$ quantile levels. CQR represents the composite quantile regression with the optimal h by minimizing the check loss in (4.6) for the validation set. The ECR closest to the nominal level τ and the smallest PE are marked in bold.

τ		CQR	QR	FHS	XK	Hybrid	CAViaR
0.1%	ECR	**0.15**	0.27	0.32	0.70	0.62	0.45
	PE	**1.00**	3.50	4.50	12.01	10.51	7.00
0.25%	ECR	**0.35**	0.55	0.52	0.90	0.80	0.62
	PE	**1.27**	3.80	3.48	8.23	6.97	4.75
0.5%	ECR	**0.75**	0.90	0.78	1.23	1.12	0.88
	PE	**2.24**	3.59	2.47	6.50	5.60	3.36
99.5%	ECR	99.48	99.42	99.40	99.33	**99.47**	99.35
	PE	**0.22**	0.67	0.90	1.57	**0.22**	1.34
99.75%	ECR	**99.70**	99.60	**99.70**	99.45	99.62	99.62
	PE	**0.63**	1.90	**0.63**	3.80	1.58	1.58
99.9%	ECR	**99.85**	99.78	99.83	99.60	99.80	99.72
	PE	**1.00**	2.50	1.50	6.00	2.00	3.50