Statistica Sinica Preprint No: SS-2019-0003	
Title	Quantile Estimation of Regression Models with GARCH-X Errors
Manuscript ID	SS-2019-0003
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202019.0003
Complete List of Authors	Qianqian Zhu Guodong Li and Zhijie Xiao
Corresponding Author	Qianqian Zhu
E-mail	zhu.qianqian@mail.shufe.edu.cn
Notice: Accepted version subject to English editing.	

Quantile Estimation of Regression Models with GARCH-X Errors

Qianqian Zhu, Guodong Li and Zhijie Xiao

Shanghai University of Finance and Economics,

University of Hong Kong and Boston College

Abstract

Conditional quantile estimation is an essential ingredient in modern risk management and many other applications, while the conditional heteroscedastic structure is usually assumed to capture the volatility in financial time series. This paper studies linear quantile regression models with GARCH-X errors, which includes the most popular generalized autoregressive conditional heteroscedasticity (GARCH) as a special case, and is also able to incorporate additional covariates into the conditional variance. Three conditional quantile estimators are proposed, and their asymptotic properties are established under mild conditions. A bootstrap procedure is further developed to approximate their asymptotic distributions. The finite-sample performance of the proposed estimators is examined via simulation experiments. An empirical application illustrates the usefulness of the proposed methodology.

Key words: Bootstrap method; GARCH-X errors; Joint estimation; Quantile regression; Two-step procedure; Value-at-Risk.

1 Introduction

The linear model provides a powerful tool in exploring the relationship between response and predictive variables (Kutner et al., 2005). For example, one may aim to predict stock returns based on related economic variables such as crude oil and gold prices; see Chernozhukov and Umantsev (2001) and Gay (2016). In economics and finance, considerable attention has been devoted to regression models with autoregressive errors for time series data; see Durbin (1960), Wang, Li, and Tsai (2007) and references therein. Stylized facts indicate that volatility clustering is a common feature for financial time series such as daily stock returns and foreign exchange rates (Ryden, Terasvirta, and Asbrink, 1998; Taylor, 2008; Tsay, 2010). As a result, it is necessary to take the conditional heteroscedasticity into account when a linear model is fitted to financial time series.

Since the appearance of autoregressive conditional heteroscedastic (ARCH) and generalized autoregressive conditional heteroscedastic (GARCH) models (Engle, 1982; Bollerslev, 1986), time series models with ARCH-type errors have become very common in empirical studies (Li, Ling, and McAleer, 2002). Motivated by the stylized facts and great success of GARCH-X models in interpreting the volatility for financial data, this paper focuses on the linear model,

$$Y_t = \boldsymbol{\phi}' \boldsymbol{X}_{t-1} + u_t,$$

where $Y_t \in \mathbb{R}$ is the response, $\boldsymbol{X}_{t-1} = (x_{1,t-1}, \ldots, x_{m,t-1})' \in \mathbb{R}^m$ consists of m covariates which can be endogenous or exogenous, and $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_m)' \in \mathbb{R}^m$ is the vector of linear coefficients. Regression errors u_t's follow a GARCH-X model (Apergis, 1998),

$$u_t = \sigma_t^* \varepsilon_t^*, \quad \sigma_t^{*2} = \omega^* + \sum_{i=1}^{q} \alpha_i^* u_{t-i}^2 + \sum_{j=1}^{p} \beta_j^* \sigma_{t-j}^{*2} + \boldsymbol{\pi}^{*\prime} \boldsymbol{V}_{t-1}, \tag{1.1}$$

where $\omega^* > 0$, $\alpha_i^* \geqslant 0$ for $i = 1, \ldots, q$, $\beta_j^* \geqslant 0$ for $j = 1, \ldots, p$, $\boldsymbol{V}_{t-1} = (v_{1,t-1}^2, \ldots, v_{d,t-1}^2)' \in \mathbb{R}^d$ includes d exogenous covariates, $\boldsymbol{\pi}^* = (\pi_1^*, \ldots, \pi_d^*)' \in \mathbb{R}^d$ is the coefficient vector with $\pi_k^* \geqslant 0$ for $1 \leqslant k \leqslant d$, and innovations $\{\varepsilon_t^*\}$ are independent and identically distributed ($i.i.d.$) random variables with mean zero and unit variance. Model (1.1) is very general, and includes ARCH and GARCH models as special cases. It reduces to the GARCH-X model studied by Han and Kristensen (2014) when $p = q = d = 1$, to Bollerslev's GARCH model when $d = 0$, and to Engle's ARCH model when $p = d = 0$. In practice, $\boldsymbol{V}_t$ may compose of realized volatility measures (Engle and Gallo, 2006; Hwang and Satchell, 2005), economic and financial indicators (Glosten, Jagannathan, and Runkle, 1993). Model (1.1) has attracted increasing attention in modeling economic and financial series; see Shephard and Sheppard (2010), Hossain and Ghahramani (2016) and Medeiros and Mendes (2016).

As a widely used measure of market risk, Value-at-Risk (VaR) plays an essential role for risk management and capital regulation in financial industry (Duffie and Pan, 1997; Taylor, 2019). Since VaR is a tail quantile of the conditional return distribution, the evaluation of VaR is explicitly a conditional quantile estimation problem; see Wu and Xiao (2002), Kuester, Mittnik, and Paolella (2006), Francq and Zakoian (2015), Wang and Zhao (2016) and Martins-Filho, Yao, and Torero (2018). Several methods have been proposed to estimate and forecast VaR, e.g., the parametric approach using a specific parametric model with known innovation distribution; the semiparametric approach using the filtered historical simulation or quantile regression; the nonparametric approach using the conditional auto-regressive VaR-method or kernel density estimation; see, e.g., Engle and Manganelli (2004), Wang and Zhao (2016) and Taylor (2019) for details. Specifically, quantile regression (Koenker and Bassett, 1978) provides a suitable tool to model the VaR based on a specific parametric model without assuming distribution form on innovations. Moreover, it possesses robustness to extreme values and facilitates distribution-free inference. This motivates us to focus on the conditional quantile estimation and VaR prediction for the linear model with GARCH-X errors. Many literatures have studied quantile regression for conditional heteroscedastic models. For examples, Koenker and Zhao (1996) and Xiao and Koenker (2009) considered quantile regression for linear (G)ARCH models proposed by Taylor (2008); Lee and Noh (2013) and Zheng et al. (2018) investigated quantile regression for Bollerslev's GARCH models; Noh and Lee (2016) studied quantile regression for ARMA models with asymmetric GARCH errors. However, little attention has been paid to the quantile estimation for linear models with GARCH-X errors. This paper aims to fill this gap, and the main contributions are summarized below. Section 2 contains the methods and theoretical results.

(a) Section 2.1 proposes three conditional quantile estimators, i.e. a jointly weighted estimator, a jointly unweighted estimator and a two-step estimator for linear models with GARCH-X errors. Specifically, the joint estimators are obtained by simultaneously estimating regression coefficients and GARCH-X parameters via quantile regression, while the two-step estimator is the hybrid of a least squares estimator for linear coefficients and a conditional quantile estimator for GARCH-X parameters. Moreover, to take into account the conditional heteroscedasticity, we introduce a set of weights into the joint estimation to improve efficiency.

(b) Section 2.2 establishes the root-n consistency and asymptotic normality for the proposed estimators. Due to the quadratic GARCH-X structure and the non-smoothness of the quantile loss function, the objective function with respect to the parameter vector is neither differentiable nor convex, which makes the theoretical derivation and numerical optimization intractable. This paper adopts the bracketing method (Pollard, 1985) to overcome this difficulty. In addition, only $E(Y_t^2) < \infty$ is required in deriving the asymptotic normality for AR-GARCH models, which makes proposed estimating methods suitable for heavy-tailed data.

(c) To circumvent difficulties in estimating the density function $f_\varepsilon(b_\tau)$ in the asymptotic covariance matrices, Section 2.3 introduces a random-weighting bootstrap method to directly approximate covariance matrices. The theoretical justification of the bootstrap method is also provided.

Section 3 conducts simulation experiments to evaluate the finite-sample performance of three proposed estimators, and a real example on VaR prediction is illustrated in Section 4. Section 5 gives a short conclusion and discussion. Technical proofs of all theorems and corollaries are relegated to a separate supplementary file. Throughout the paper, we denote by $\|\cdot\|$ the norm of a matrix or column vector, defined as $\|A\| = \sqrt{\mathrm{tr}(AA')} = \sqrt{\sum_{i,j} a_{ij}^2}$.

2 Model, methodology and asymptotic results

2.1 Quantile regression estimation

Consider a linear model with GARCH-X errors,

$$Y_t = \boldsymbol{\phi}'\boldsymbol{X}_{t-1} + u_t, \tag{2.1}$$

and

$$u_t = \sigma_t\varepsilon_t, \quad \sigma_t^2 = 1 + \sum_{i=1}^{q}\alpha_i u_{t-i}^2 + \sum_{j=1}^{p}\beta_j\sigma_{t-j}^2 + \boldsymbol{\pi}'\boldsymbol{V}_{t-1}, \tag{2.2}$$

where $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_m)'$ is an m-dimensional coefficient vector of the covariates $\boldsymbol{X}_{t-1} = (x_{1,t-1}, \ldots, x_{m,t-1})'$ in regression model, u_t's are regression errors, $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_d)'$ is a

d-dimensional coefficient vector of covariates $\boldsymbol{V}_{t-1} = (v_{1,t-1}^2, \ldots, v_{d,t-1}^2)'$ in the volatility model, $\alpha_i \geqslant 0$ for $1 \leqslant i \leqslant q$, $\beta_j \geqslant 0$ for $1 \leqslant j \leqslant p$, $\pi_k \geqslant 0$ for $1 \leqslant k \leqslant d$, and $\{\varepsilon_t\}$ are *i.i.d.* random variables with mean zero and finite variance. In practice, $\boldsymbol{X}_{t-1}$ may include lagged values of Y_t.

Let $\mathcal{F}_t$ be the σ-field generated by $\{\boldsymbol{X}_t, \boldsymbol{X}_{t-1}, \ldots; \boldsymbol{V}_t, \boldsymbol{V}_{t-1}, \ldots; \varepsilon_t, \varepsilon_{t-1}, \ldots\}$ and b_τ be the τth quantile of ε_t. Assume that ε_t is independent of $\mathcal{F}_{t-1}$, and then the τth quantile of Y_t, conditional on $\mathcal{F}_{t-1}$, has the form of

$$Q_{Y_t}(\tau|\mathcal{F}_{t-1}) = \boldsymbol{\phi}'\boldsymbol{X}_{t-1} + b_\tau \sigma_t, \tag{2.3}$$

where σ_t is defined in (2.2). Let $\omega^* = \text{var}(\varepsilon_t)$, $\varepsilon_t^* = \varepsilon_t/\sqrt{\omega^*}$, $\sigma_t^* = \sigma_t\sqrt{\omega^*}$, $\alpha_i^* = \omega^*\alpha_i$, $\beta_j^* = \beta_j$ and $\boldsymbol{\pi}^* = \omega^*\boldsymbol{\pi}$. The GARCH-X error in (2.2) then has a standard form of (1.1). Note that the GARCH-X model is an extension of Bollerslev's GARCH model by including additional predictors. Since model (1.1) will suffer from an identifiability problem in the quantile estimation (Xiao and Koenker, 2009; Lee and Noh, 2013; Noh and Lee, 2016), this paper will use the GARCH-X form at (2.2).

Denote the parameter vector of models (2.1) and (2.2) by $\boldsymbol{\lambda} = (\boldsymbol{\gamma}', \boldsymbol{\phi}')'$, where $\boldsymbol{\gamma} = (\alpha_1, \ldots, \alpha_q, \beta_1, \ldots, \beta_p, \pi_1, \ldots, \pi_d)'$. Define functions $u_t(\boldsymbol{\phi}) = Y_t - \boldsymbol{\phi}'\boldsymbol{X}_{t-1}$ and $\sigma_t^2(\boldsymbol{\lambda}) = 1 + \sum_{i=1}^q \alpha_i u_{t-i}^2(\boldsymbol{\phi}) + \sum_{j=1}^p \beta_j \sigma_{t-j}^2(\boldsymbol{\lambda}) + \boldsymbol{\pi}'\boldsymbol{V}_{t-1}$. Note that functions $\sigma_t^2(\boldsymbol{\lambda})$'s are defined recursively and thus depend on infinite past observations. Therefore, initial values are required in practice. In this paper, we set $u_t(\boldsymbol{\phi}) = 0$ and $\sigma_t^2(\boldsymbol{\lambda}) = 1$ for $t \leqslant 0$, and denote the resulting function of $\sigma_t(\boldsymbol{\lambda})$ as $\widetilde{\sigma}_t(\boldsymbol{\lambda})$; see also Lee and Noh (2013). To estimate $Q_{Y_t}(\tau|\mathcal{F}_{t-1})$ at (2.3), it is natural to simultaneously estimate the regression coefficients and the GARCH-X parameters by quantile regression, and a joint conditional quantile estimator can be defined as

$$\widetilde{\boldsymbol{\theta}}_{\tau n} = (\widetilde{b}_{\tau n}, \widetilde{\boldsymbol{\lambda}}_n')' = \underset{b,\boldsymbol{\lambda}}{\text{argmin}} \sum_{t=1}^{n} \rho_\tau\{Y_t - \boldsymbol{\phi}'\boldsymbol{X}_{t-1} - b\widetilde{\sigma}_t(\boldsymbol{\lambda})\}, \tag{2.4}$$

where $\rho_\tau(x) = x[\tau - I(x < 0)]$ is the check function. However, $\widetilde{\boldsymbol{\theta}}_{\tau n}$ may suffer from efficiency loss due to the presence of conditional heteroscedasticity in regression errors. As a result, we may consider a jointly weighted conditional quantile estimator,

$$\widehat{\boldsymbol{\theta}}_{\tau n} = (\widehat{b}_{\tau n}, \widehat{\boldsymbol{\lambda}}_n')' = \underset{b,\boldsymbol{\lambda}}{\text{argmin}} \sum_{t=1}^{n} \widehat{\sigma}_t^{-1}\rho_\tau\{Y_t - \boldsymbol{\phi}'\boldsymbol{X}_{t-1} - b\widetilde{\sigma}_t(\boldsymbol{\lambda})\}, \tag{2.5}$$

where the weight $\widehat{\sigma}_t^{-1} = \widetilde{\sigma}_t^{-1}(\widehat{\boldsymbol{\lambda}}_n^{int})$, and $\widehat{\boldsymbol{\lambda}}_n^{int}$ is an appropriate estimator of $\boldsymbol{\lambda}_0$. The objective functions in (2.4) and (2.5) are both non-convex with respect to $\boldsymbol{\theta} = (b, \boldsymbol{\lambda})'$, even if models (2.1) and (2.2) are reduced to the ARCH(1) models where $\boldsymbol{\theta} = (b, \alpha_1)'$. This make challenging the theoretical derivation and numerical optimization.

As in Koenker and Zhao (1996), a two-step procedure can also be applied to models (2.1) and (2.2). Specifically, the first step uses the least squares estimation to obtain an estimator $\check{\boldsymbol{\phi}}_n$ for model (2.1), and computes regression residuals by $\check{u}_t = u_t(\check{\boldsymbol{\phi}}_n)$. Then the second step performs the conditional quantile estimation for model (2.2) below

$$\check{\gamma}_{\tau n} = (\check{b}_{\tau n}, \check{\gamma}_n')' = \arg\min_{b,\boldsymbol{\gamma}} \sum_{t=1}^{n} \rho_\tau\left\{\check{u}_t - b\check{\sigma}_t(\boldsymbol{\gamma})\right\},$$

where $\check{\sigma}_t^2(\boldsymbol{\gamma})$'s are calculated recursively by $\check{\sigma}_t^2(\boldsymbol{\gamma}) = 1 + \sum_{i=1}^{q} \alpha_i \check{u}_{t-i}^2 + \sum_{j=1}^{p} \beta_j \check{\sigma}_{t-j}^2(\boldsymbol{\gamma}) + \boldsymbol{\pi}'\boldsymbol{V}_{t-1}$ given initial values $\check{u}_t = 0$ and $\check{\sigma}_t^2(\boldsymbol{\gamma}) = 1$ for $t \leqslant 0$. It can be shown that the preliminary estimator $\check{\boldsymbol{\phi}}_n$ is involved in the Bahadur representation of $\check{\gamma}_{\tau n}$; see Corollary 2 in Section 2.2. Denote $\check{\boldsymbol{\theta}}_{\tau n} = (\check{\gamma}_{\tau n}', \check{\boldsymbol{\phi}}_n')'$. We call $\widehat{\boldsymbol{\theta}}_{\tau n}$, $\widetilde{\boldsymbol{\theta}}_{\tau n}$ and $\check{\boldsymbol{\theta}}_{\tau n}$ the jointly weighted estimator, jointly unweighted estimator and two-step estimator, respectively. We can verify that initial values of $\breve{u}_t$, $u_t(\boldsymbol{\phi})$, $\breve{\sigma}_t^2(\boldsymbol{\gamma})$ and $\sigma_t^2(\boldsymbol{\lambda})$ have no effects on the asymptotic distributions of three proposed estimators.

For the jointly weighted estimator $\widehat{\boldsymbol{\theta}}_{\tau n}$, we next define a Bayesian information criterion (BIC) to select the orders of m, d, p and q in model (2.3),

$$\text{BIC}_\tau(m,d,p,q) = 2n\log\widehat{\sigma}_{\tau n} + (1+m+d+p+q)\log n, \tag{2.6}$$

where $\widehat{\sigma}_{\tau n} = n^{-1}\sum_{t=1}^n \widehat{\sigma}_t^{-1}\rho_\tau\{Y_t - q_t(\widehat{\boldsymbol{\theta}}_{\tau n})\}$ with $\widehat{\sigma}_t = \widetilde{\sigma}_t(\widehat{\boldsymbol{\lambda}}_n^{int})$ and $\widehat{\boldsymbol{\theta}}_{\tau n}$ being defined by (2.5); see Zhu, Zheng, and Li (2018). Let $(\widehat{m},\widehat{d},\widehat{p},\widehat{q}) = \arg\min_{m,d,p,q}\text{BIC}_\tau(m,d,p,q)$, where $1\leqslant m\leqslant m_{\max}$, $1\leqslant d\leqslant d_{\max}$, $1\leqslant p\leqslant p_{\max}$, $1\leqslant q\leqslant q_{\max}$, and $m_{\max}$, $d_{\max}$, $p_{\max}$, and $q_{\max}$ are predetermined integers. By a method similar to the proof of Theorem 5 in Zhu, Zheng, and Li (2018), we can show that the proposed BIC at (2.6) is consistent when the true orders satisfy that $m_0\leqslant m_{\max}$, $d_0\leqslant d_{\max}$, $p_0\leqslant p_{\max}$ and $q_0\leqslant q_{\max}$. Similarly we can define the BIC for the jointly unweighted estimator $\widetilde{\boldsymbol{\theta}}_{\tau n}$, and verify its consistency as well.

Note that the estimating procedure should be repeated $m_{\max}\times d_{\max}\times p_{\max}\times q_{\max}$ times to search for the orders of m, d, p and q, which is time-consuming when $m_{\max}$, $d_{\max}$, $p_{\max}$ and $q_{\max}$ are large. Alternatively, we may fix some orders in advance and search for the others by the BIC. For example, we may select orders based on the background of data and other quantitative tools such as ACF and PACF.

2.2 Asymptotic properties

Let $\boldsymbol{\theta} = (b,\boldsymbol{\lambda}')'$ be the parameter vector of model (2.3), and $\boldsymbol{\theta}_{\tau 0} = (b_\tau,\boldsymbol{\lambda}_0')' = (b_\tau,\boldsymbol{\gamma}_0',\boldsymbol{\phi}_0')'$ be its true value, where $\boldsymbol{\phi}_0 = (\phi_{10},\ldots,\phi_{m0})'$ and $\boldsymbol{\gamma}_0 = (\alpha_{10},\ldots,\alpha_{q0},\beta_{10},\ldots,\beta_{p0},\pi_{10},\ldots,\pi_{d0})'$. Denote by $q_t(\boldsymbol{\theta}) = \boldsymbol{\phi}'\boldsymbol{X}_{t-1} + b\sigma_t(\boldsymbol{\lambda})$ and $\widetilde{q}_t(\boldsymbol{\theta}) = \boldsymbol{\phi}'\boldsymbol{X}_{t-1} + b\widetilde{\sigma}_t(\boldsymbol{\lambda})$ the conditional quantile functions of Y_t without and with initial values, respectively. Suppose that the parameter space $\Theta \subset \mathbb{R} \times \mathbb{R}_{+}^{p+q+d} \times \mathbb{R}^{m}$ is a compact set satisfying

$$\underline{b} \leqslant |b| \leqslant \overline{b}, \sum_{j=1}^{p} \beta_j \leqslant \rho_0, \ \underline{w} \leqslant \min(\alpha_1, \ldots, \alpha_q, \beta_1, \ldots, \beta_p, \pi_1, \ldots, \pi_d) \leqslant \max(\alpha_1, \ldots, \alpha_q, \beta_1, \ldots, \beta_p, \pi_1, \ldots, \pi_d) \leqslant \overline{w},$$

where $\mathbb{R}_{+} = (0, +\infty)$, $0 < \underline{b} < \overline{b}$, $0 < \underline{w} < \overline{w}$, $0 < \rho_0 < 1$ and $p\underline{w} < \rho_0$. We further assume that $\boldsymbol{\theta}_{\tau 0}$ is an interior point of Θ. Moreover, denote by $F_{\varepsilon}(\cdot)$ and $f_{\varepsilon}(\cdot)$ the distribution and density functions of ε_t, respectively.

We first discuss the asymptotic properties for the jointly weighted estimator $\widehat{\boldsymbol{\theta}}_{\tau n}$. Since the objective function in (2.5) is non-convex and non-differentiable, the convexity lemma of Pollard (1991) cannot be applied directly. Instead, we derive the asymptotic properties by verifying the stochastic differentiability condition defined by Pollard (1985). As a result, we first prove the consistency of $\widehat{\boldsymbol{\theta}}_{\tau n}$ in Theorem 1 and then establish its asymptotic normality in Theorem 2.

Assumption 1. *(i) $\{\boldsymbol{X}_t\}$, $\{\boldsymbol{V}_t\}$ and $\{u_t\}$ are strictly stationary and ergodic with $E(\|\boldsymbol{X}_t\|^2) < \infty$ and $E(\|\boldsymbol{V}_t\|) < \infty$; (ii) The polynomials $\alpha(x) = \sum_{i=1}^{q} \alpha_i x^i$ and $\beta(x) = 1 - \sum_{j=1}^{p} \beta_j x^j$ have no common root.*

Assumption 2. *ε_t has a continuous density function $f_{\varepsilon}(\cdot)$ at a neighborhood of b_{τ}.*

This paper focuses on the model with stationary covariates, hence Assumption 1(i) assumes that $\{\boldsymbol{X}_t\}$ and $\{\boldsymbol{V}_t\}$ are strictly stationary. Assumption 1(ii) is the identifiability condition for GARCH-X model (2.2). Moreover, for model identification the intercept should not be included in the regression model, i.e. $\boldsymbol{X}_{t-1}$ does not incorporate one.

Theorem 1. *Under Assumptions 1 and 2, if $\widehat{\boldsymbol{\lambda}}_n^{int} - \boldsymbol{\lambda}_0 = o_p(1)$ and $E(u_t^2) < \infty$, then $\widehat{\boldsymbol{\theta}}_{\tau n} \to \boldsymbol{\theta}_{\tau 0}$ in probability as $n \to \infty$.*

Denote $\Sigma_i(\tau) = E[\sigma_t^{-i}\partial q_t(\boldsymbol{\theta}_{\tau 0})/\partial\boldsymbol{\theta}\partial q_t(\boldsymbol{\theta}_{\tau 0})/\partial\boldsymbol{\theta}']$ for $i = 0$, 1 and 2, where

$$\frac{\partial q_t(\boldsymbol{\theta}_{\tau 0})}{\partial \boldsymbol{\theta}} = \left(\sigma_t, \frac{b_\tau}{2\sigma_t}\frac{\partial \sigma_t^2(\boldsymbol{\lambda}_0)}{\partial \boldsymbol{\gamma}'}, \boldsymbol{X}_{t-1}' + \frac{b_\tau}{2\sigma_t}\frac{\partial \sigma_t^2(\boldsymbol{\lambda}_0)}{\partial \boldsymbol{\phi}'}\right)'.$$

To study the asymptotic normality of $\widehat{\boldsymbol{\theta}}_{\tau n}$, the following assumptions are required.

Assumption 3. *$E(\|\boldsymbol{X}_t\|^{4+\delta}) < \infty$ for some $\delta > 0$. The matrices $E(\boldsymbol{X}_t\boldsymbol{X}_t')$ and $\Sigma_0(\tau)$ are positive definite.*

Assumption 4. *The density function $f_\varepsilon(\cdot)$ is positive and differentiable almost everywhere on $\mathbb{R}$, with $f_\varepsilon(\cdot)$ satisfying $\sup_{x\in\mathbb{R}} f_\varepsilon(x) < \infty$ and its derivative $\dot{f}_\varepsilon(\cdot)$ satisfying $\sup_{x\in\mathbb{R}} |\dot{f}_\varepsilon(x)| < \infty$.*

Assumption 3 is required to verify the root-n consistency and asymptotic normality of $\widehat{\boldsymbol{\theta}}_{\tau n}$. Assumption 4 is made to simplify technical proofs, while it suffices to restrict the boundedness of $f_\varepsilon(\cdot)$ and $|\dot{f}_\varepsilon(\cdot)|$ in a small but fixed neighborhood of b_τ. Moreover, Assumption 4 implies Assumption 2.

Theorem 2. *Suppose that $\sqrt{n}(\widehat{\boldsymbol{\lambda}}_n^{int} - \boldsymbol{\lambda}_0) = O_p(1)$ and $E(u_t^2) < \infty$. If Assumptions 1, 3 and 4 hold, then*

(i) $\sqrt{n}(\widehat{\boldsymbol{\theta}}_{\tau n} - \boldsymbol{\theta}_{\tau 0}) = O_p(1)$; *and*

(ii) $\sqrt{n}(\widehat{\boldsymbol{\theta}}_{\tau n} - \boldsymbol{\theta}_{\tau 0}) \to N(\mathbf{0}, \Xi_1)$ *in distribution as* $n \to \infty$, *where* $\Xi_1 = \tau(1-\tau)f_\varepsilon^{-2}(b_\tau)\Sigma_2^{-1}(\tau)$.

To prove Theorem 2, we apply the bracketing method to verify the stochastic differentiability condition (Pollard, 1985). This, together with standard arguments for conditional quantile estimators, implies the root-n consistency of $\widehat{\boldsymbol{\theta}}_{\tau n}$, and then the asymptotic normality follows. Moreover, different from the condition on $\widehat{\boldsymbol{\lambda}}_n^{int}$ in Theorem 1, Theorem 2 requires that $\widehat{\boldsymbol{\lambda}}_n^{int}$ is a root-n consistent estimator of $\boldsymbol{\lambda}_0$.

Let $\boldsymbol{W}_t = (\sigma_t, 0.5\sigma_t^{-1}b_\tau\partial\sigma_t^2(\boldsymbol{\lambda}_0)/\partial\boldsymbol{\gamma}')'$, $\boldsymbol{M}_t = \boldsymbol{X}_{t-1} + 0.5\sigma_t^{-1}b_\tau\partial\sigma_t^2(\boldsymbol{\lambda}_0)/\partial\boldsymbol{\phi}'$. It is clear that $\partial q_t(\boldsymbol{\theta}_{\tau 0})/\partial\boldsymbol{\theta} = (\boldsymbol{W}_t', \boldsymbol{M}_t')'$. Define matrices $D_i = E(\sigma_t^i\boldsymbol{X}_{t-1}\boldsymbol{X}_{t-1}')$ for $i = 0$ and 2, $\Omega_i = E(\sigma_t^{-i}\boldsymbol{W}_t\boldsymbol{W}_t')$ for $i = 0$ and 1, $\Gamma_1 = E(\sigma_t^{-1}\boldsymbol{W}_t\boldsymbol{M}_t')$ and $\Gamma_2 = E(\sigma_t\boldsymbol{W}_t\boldsymbol{X}_{t-1}')$. Let $\omega^* = \text{var}(\varepsilon_t)$ and $\kappa = E[\varepsilon_t I(\varepsilon_t < b_\tau)]$. Define the matrices

$$\Xi_2 = \frac{\tau(1-\tau)}{f_\varepsilon^2(b_\tau)}\Sigma_1^{-1}(\tau)\Sigma_0(\tau)\Sigma_1^{-1}(\tau) \quad \text{and} \quad \Xi_3 = \begin{pmatrix} \Sigma_{11}(\tau) & \Sigma_{12}(\tau) \\ \Sigma_{12}'(\tau) & \Sigma_{22} \end{pmatrix},$$

where $\Sigma_{12}(\tau) = \kappa f_\varepsilon^{-1}(b_\tau)\Omega_1^{-1}\Gamma_2 D_0^{-1} - \Omega_1^{-1}\Gamma_1\Sigma_{22}$, $\Sigma_{22} = \omega^* D_0^{-1}D_2D_0^{-1}$ and

$$\Sigma_{11}(\tau) = \Omega_1^{-1}\left[\frac{\tau(1-\tau)}{f_\varepsilon^2(b_\tau)}\Omega_0 + \frac{\kappa}{f_\varepsilon(b_\tau)}(\Gamma_2 D_0^{-1}\Gamma_1' + \Gamma_1 D_0^{-1}\Gamma_2') + \Gamma_1\Sigma_{22}\Gamma_1'\right]\Omega_1^{-1}.$$

Using the same technique as for Theorem 2, we can derive the asymptotic properties for the jointly unweighted estimator $\widetilde{\boldsymbol{\theta}}_{\tau n}$ and the two-step estimator $\check{\boldsymbol{\theta}}_{\tau n}$ below.

Corollary 1. *Suppose that $E(|u_t|^{2+\delta}) < \infty$ for some $\delta > 0$. If Assumptions 1, 3 and 4 hold, then*

(i) $\sqrt{n}(\widetilde{\boldsymbol{\theta}}_{\tau n} - \boldsymbol{\theta}_{\tau 0}) = O_p(1)$; *and*

(ii) $\sqrt{n}(\widetilde{\boldsymbol{\theta}}_{\tau n} - \boldsymbol{\theta}_{\tau 0}) \to N(\mathbf{0}, \Xi_2)$ *in distribution as* $n \to \infty$.

Corollary 2. *Suppose that matrices Ω_0 and Ω_1 are positive definite, and $E(|u_t|^{2+\delta}) < \infty$ for some $\delta > 0$. If Assumptions 1, 3 and 4 hold, then*

(i) $\sqrt{n}(\check{\boldsymbol{\gamma}}_{\tau n} - \boldsymbol{\gamma}_{\tau 0}) = O_p(1)$, *where* $\boldsymbol{\gamma}_{\tau 0} = (b_\tau, \boldsymbol{\gamma}_0')'$; *and*

(ii) $\check{\boldsymbol{\gamma}}_{\tau n}$ *has the following Bahadur representation*

$$\sqrt{n}(\check{\boldsymbol{\gamma}}_{\tau n} - \boldsymbol{\gamma}_{\tau 0}) = \frac{\Omega_1^{-1}}{f_\varepsilon(b_\tau)}\frac{1}{\sqrt{n}}\sum_{t=1}^n \boldsymbol{W}_t\psi_\tau(\varepsilon_t - b_\tau) - \Omega_1^{-1}\Gamma_1\sqrt{n}(\check{\boldsymbol{\phi}}_n - \boldsymbol{\phi}_0) + o_p(1),$$

where $\psi_\tau(x) = \tau - I(x < 0)$ *and*

$$\sqrt{n}(\check{\boldsymbol{\phi}}_n - \boldsymbol{\phi}_0) = \left(\frac{1}{n}\sum_{t=1}^{n} \boldsymbol{X}_{t-1}\boldsymbol{X}'_{t-1}\right)^{-1} \frac{1}{\sqrt{n}}\sum_{t=1}^{n} \boldsymbol{X}_{t-1}\sigma_t\varepsilon_t.$$

Moreover, it holds that $\sqrt{n}(\check{\boldsymbol{\theta}}_{\tau n} - \boldsymbol{\theta}_{\tau 0}) \to N(\mathbf{0}, \Xi_3)$ *in distribution as* $n \to \infty$.

Note that Corollaries 1-2 require stronger moment condition on u_t than that in Theorem 2. Corollary 2 provides a theoretical justification that $\sqrt{n}(\check{\boldsymbol{\lambda}}_n - \boldsymbol{\lambda}_0) = O_p(1)$, where $\check{\boldsymbol{\lambda}}_n = (\check{\gamma}'_n, \check{\boldsymbol{\phi}}'_n)'$. Hence, $\check{\boldsymbol{\lambda}}_n$ can be used to construct the weights $\{\widehat{\sigma}_t^{-1}\}$ in (2.5) for the jointly weighted estimation. Specifically, we may set $\widehat{\sigma}_t = \sqrt{1 + \check{\gamma}'_n \check{\boldsymbol{z}}_t}$, where $\check{\boldsymbol{z}}_t = (\check{u}^2_{t-1}, \ldots, \check{u}^2_{t-q}, \check{\sigma}^2_{t-1}(\check{\gamma}_n), \ldots, \check{\sigma}^2_{t-p}(\check{\gamma}_n), \upsilon^2_{1,t-1}, \ldots, \upsilon^2_{d,t-1})'$.

A general theoretical comparison of the three proposed estimators is complicated due to the iterative form of σ_t. However, given τ, the true parameter vector $\boldsymbol{\theta}_{\tau 0}$ and density function $f_\varepsilon(\cdot)$, we can obtain theoretical values of b_τ, $f_\varepsilon(b_\tau)$, ω^* and κ, and estimate all matrices in Ξ_i $(i = 1, 2, 3)$ by sample averages based on a generated sequence with large size. Then we can compute the asymptotic relative efficiency of $\widehat{\boldsymbol{\theta}}_{\tau n}$ to $\widetilde{\boldsymbol{\theta}}_{\tau n}$ and $\check{\boldsymbol{\theta}}_{\tau n}$, defined as $\text{ARE}(\widehat{\boldsymbol{\theta}}_{\tau n}, \widetilde{\boldsymbol{\theta}}_{\tau n}) = (|\Xi_2|/|\Xi_1|)^{1/(p+q+d+1)}$ and $\text{ARE}(\widehat{\boldsymbol{\theta}}_{\tau n}, \check{\boldsymbol{\theta}}_{\tau n}) = (|\Xi_3|/|\Xi_1|)^{1/(p+q+d+1)}$, respectively, where $|\cdot|$ is the determinant of a matrix; see Serfling (2009). Simulation results in Section 3 indicate that the jointly weighted estimator $\widehat{\boldsymbol{\theta}}_{\tau n}$ is asymptotically more efficient than the jointly unweighted estimator $\widetilde{\boldsymbol{\theta}}_{\tau n}$, while the relative performance of $\widehat{\boldsymbol{\theta}}_{\tau n}$ and the two-step estimator $\check{\boldsymbol{\theta}}_{\tau n}$ is mixed in asymptotic efficiency; see more details in Section 3.2.

Based on $\widehat{\boldsymbol{\theta}}_{\tau n}$, $\widetilde{\boldsymbol{\theta}}_{\tau n}$ and $\check{\boldsymbol{\theta}}_{\tau n}$, the conditional quantile of Y_t given $\mathcal{F}_{t-1}$ can be estimated by $q_t(\widehat{\boldsymbol{\theta}}_{\tau n})$, $q_t(\widetilde{\boldsymbol{\theta}}_{\tau n})$ and $q_t(\check{\boldsymbol{\theta}}_{\tau n})$, respectively. Note that $\widehat{\boldsymbol{\theta}}_{\tau n} = (\widehat{\gamma}'_{\tau n}, \widehat{\boldsymbol{\phi}}'_n)'$ and $\widetilde{\boldsymbol{\theta}}_{\tau n} = (\widetilde{\gamma}'_{\tau n}, \widetilde{\boldsymbol{\phi}}'_n)'$, where $\widehat{\gamma}_{\tau n} = (\widehat{b}_{\tau n}, \widehat{\gamma}'_n)'$ and $\widetilde{\gamma}_{\tau n} = (\widetilde{b}_{\tau n}, \widetilde{\gamma}'_n)'$. The following corollary provides the theoretical results for the τth conditional quantile of Y_{n+1} based on three approaches.

Corollary 3. *If the conditions of Theorem 2 and Corollaries 1-2 hold, then, conditional on $\mathcal{F}_n$,*

$$\sqrt{n}[q_{n+1}(\widehat{\boldsymbol{\theta}}_{\tau n}) - q_{n+1}(\boldsymbol{\theta}_{\tau 0})] = \boldsymbol{W}'_{n+1}\sqrt{n}(\widehat{\boldsymbol{\gamma}}_{\tau n} - \boldsymbol{\gamma}_{\tau 0}) + \boldsymbol{M}'_{n+1}\sqrt{n}(\widehat{\boldsymbol{\phi}}_n - \boldsymbol{\phi}_0) + o_p(1),$$

$$\sqrt{n}[q_{n+1}(\widetilde{\boldsymbol{\theta}}_{\tau n}) - q_{n+1}(\boldsymbol{\theta}_{\tau 0})] = \boldsymbol{W}'_{n+1}\sqrt{n}(\widetilde{\boldsymbol{\gamma}}_{\tau n} - \boldsymbol{\gamma}_{\tau 0}) + \boldsymbol{M}'_{n+1}\sqrt{n}(\widetilde{\boldsymbol{\phi}}_n - \boldsymbol{\phi}_0) + o_p(1),$$

and

$$\sqrt{n}[q_{n+1}(\check{\boldsymbol{\theta}}_{\tau n}) - q_{n+1}(\boldsymbol{\theta}_{\tau 0})] = \boldsymbol{W}'_{n+1}\sqrt{n}(\check{\boldsymbol{\gamma}}_{\tau n} - \boldsymbol{\gamma}_{\tau 0}) + \boldsymbol{M}'_{n+1}\sqrt{n}(\check{\boldsymbol{\phi}}_n - \boldsymbol{\phi}_0) + o_p(1).$$

Theorem 2 and Corollaries 1-3 still hold when model (2.2) reduces to an ARCH model or a GARCH model. To establish Theorem 2, $E(\|\boldsymbol{X}_t\|^{4+\delta}) < \infty$ is necessary if $\{\boldsymbol{X}_{t-1}\}$ includes exogenous variables. However, when $d = 0$ and $\{\boldsymbol{X}_{t-1}\}$ only contains lagged values of Y_t, i.e. models (2.1) and (2.2) reduce to AR-GARCH models or AR(m)-ARCH(q) models with $m \leqslant q$, the moment condition on $\boldsymbol{X}_t$ can be relaxed to $E(\|\boldsymbol{X}_t\|^2) < \infty$. Moreover, for Corollaries 1-2, the moment condition on $\boldsymbol{X}_t$ can be reduced to $E(\|\boldsymbol{X}_t\|^{2+\delta}) < \infty$ for AR-GARCH models and AR(m)-ARCH(q) models with $m \leqslant q$. In addition, when the GARCH-X errors reduce to ARCH errors, we can show that $\Xi_2 - \Xi_1$ is non-negative definite, i.e. $\widehat{\boldsymbol{\theta}}_{\tau n}$ is asymptotically more efficient than $\widetilde{\boldsymbol{\theta}}_{\tau n}$.

2.3 Bootstrapping approximation

To circumvent difficulties in estimating the density function $f_\varepsilon(b_\tau)$, we propose a bootstrapping procedure to directly approximate the asymptotic distributions of $\widehat{\boldsymbol{\theta}}_{\tau n}$, $\widetilde{\boldsymbol{\theta}}_{\tau n}$ and $\check{\boldsymbol{\theta}}_{\tau n}$, respectively.

For the joint estimators $\widehat{\boldsymbol{\theta}}_{\tau n}$ and $\widetilde{\boldsymbol{\theta}}_{\tau n}$, we define the corresponding randomly weighted bootstrapping estimators below,

$$\widehat{\boldsymbol{\theta}}_{\tau n}^{*} = (\widehat{b}_{\tau n}^{*}, \widehat{\boldsymbol{\lambda}}_{n}^{*\prime})^{\prime} = \underset{b,\boldsymbol{\lambda}}{\operatorname{argmin}} \sum_{t=1}^{n} \omega_t \widehat{\sigma}_t^{-1} \rho_\tau \{Y_t - \boldsymbol{\phi}^{\prime} \boldsymbol{X}_{t-1} - b\widetilde{\sigma}_t(\boldsymbol{\lambda})\} \tag{2.7}$$

and

$$\widetilde{\boldsymbol{\theta}}_{\tau n}^{*} = (\widetilde{b}_{\tau n}^{*}, \widetilde{\boldsymbol{\lambda}}_{n}^{*\prime})^{\prime} = \underset{b,\boldsymbol{\lambda}}{\operatorname{argmin}} \sum_{t=1}^{n} \omega_t \rho_\tau \{Y_t - \boldsymbol{\phi}^{\prime} \boldsymbol{X}_{t-1} - b\widetilde{\sigma}_t(\boldsymbol{\lambda})\}, \tag{2.8}$$

where $\{\omega_t\}$ are $i.i.d.$ non-negative random weights with mean and variance both equal to one; see also Zheng et al. (2018) and Zhu, Zeng, and Li (2019).

For the two-step estimator $\check{\boldsymbol{\theta}}_{\tau n}$, the randomly weighted bootstrapping is involved in both steps. In the first step, a randomly weighted least squares estimator is obtained by $\check{\boldsymbol{\phi}}_n^{*} = \left(\sum_{t=1}^{n} \omega_t \boldsymbol{X}_{t-1} \boldsymbol{X}_{t-1}^{\prime}\right)^{-1} \sum_{t=1}^{n} \omega_t \boldsymbol{X}_{t-1} Y_t$ and the bootstrapped residuals are computed by $\check{u}_t^{*} = u_t(\check{\boldsymbol{\phi}}_n^{*})$. Then a randomly weighted quantile estimation is performed below

$$\check{\gamma}_{\tau n}^{*} = (\check{b}_{\tau n}^{*}, \check{\gamma}_{n}^{*\prime})^{\prime} = \arg\min_{b,\boldsymbol{\gamma}} \sum_{t=1}^{n} \omega_t \rho_\tau \left\{\check{u}_t^{*} - b\check{\sigma}_t^{*}(\boldsymbol{\gamma})\right\}, \tag{2.9}$$

where, given initial values $\check{u}_t^{*} = 0$ and $\check{\sigma}_t^{*2}(\boldsymbol{\gamma}) = 1$ for $t \leqslant 0$, $\check{\sigma}_t^{*2}(\boldsymbol{\gamma})$'s are calculated recursively by $\check{\sigma}_t^{*2}(\boldsymbol{\gamma}) = 1 + \sum_{i=1}^{q} \alpha_i \check{u}_{t-i}^{*2} + \sum_{j=1}^{p} \beta_j \check{\sigma}_{t-j}^{*2}(\boldsymbol{\gamma}) + \boldsymbol{\pi}^{\prime} \boldsymbol{V}_{t-1}$. As a result, the randomly weighted bootstrapping estimator for $\check{\boldsymbol{\theta}}_{\tau n}$ is defined as $\check{\boldsymbol{\theta}}_{\tau n}^{*} = (\check{\gamma}_{\tau n}^{*\prime}, \check{\boldsymbol{\phi}}_{n}^{*\prime})^{\prime}$.

Assumption 5. *The random weights $\{\omega_t\}$ are i.i.d. non-negative random variables with mean and variance both equal to one, satisfying $E|\omega_t|^{2+\delta} < \infty$ for some $\delta > 0$.*

Theorem 3. *Suppose that Assumption 5 and the conditions in Theorem 2 and Corollaries 1-2 hold. Then, conditional on $\mathcal{F}_n$:*

(i) $\sqrt{n}(\widehat{\boldsymbol{\theta}}_{\tau n}^{*} - \widehat{\boldsymbol{\theta}}_{\tau n}) \to_d N(\mathbf{0}, \Xi_1)$ *in probability as $n \to \infty$;*

(ii) $\sqrt{n}(\widetilde{\boldsymbol{\theta}}_{\tau n}^{*} - \widetilde{\boldsymbol{\theta}}_{\tau n}) \to_d N(\mathbf{0}, \Xi_2)$ *in probability as $n \to \infty$; and*

(iii) $\sqrt{n}(\check{\boldsymbol{\theta}}^*_{\tau n} - \check{\boldsymbol{\theta}}_{\tau n}) \to_d N(\mathbf{0}, \Xi_3)$ *in probability as* $n \to \infty$*;*

where Ξ_i *for* $i = 1, 2$ *and 3 are defined in Theorem 2 and Corollaries 1-2.*

From Theorem 3, we can approximate the covariance matrices of $\widehat{\boldsymbol{\theta}}_{\tau n}$, $\widetilde{\boldsymbol{\theta}}_{\tau n}$ and $\check{\boldsymbol{\theta}}_{\tau n}$ by the bootstrapped covariance matrices of $\sqrt{n}(\widehat{\boldsymbol{\theta}}^*_{\tau n} - \widehat{\boldsymbol{\theta}}_{\tau n})$, $\sqrt{n}(\widetilde{\boldsymbol{\theta}}^*_{\tau n} - \widetilde{\boldsymbol{\theta}}_{\tau n})$ and $\sqrt{n}(\check{\boldsymbol{\theta}}^*_{\tau n} - \check{\boldsymbol{\theta}}_{\tau n})$, respectively. As a result, confidence intervals of the estimators can be constructed by plugging in the approximated asymptotic standard deviations (ASDs) calculated by the bootstrap method. Moreover, hypothesis tests for detecting significance of parameters can also be conducted by replacing the covariance matrices with their bootstrapping approximations.

For the random weights, there are many distributions satisfying Assumption 5, such as the standard exponential distribution and the Rademacher distribution which takes the value 0 or 2 with probability 0.5. According to the simulation findings in Zheng et al. (2018) and Zhu, Zeng, and Li (2019), the performance of the randomly weighted bootstrapping approximation is insensitive to the choice of random weights. As a result, this paper simply uses the random weights generated from the standard exponential distribution in the following sections.

3 Simulation studies

3.1 Finite-sample performance of three proposed estimators

The first experiment aims to evaluate the finite-sample performance of the three proposed estimators $\widehat{\boldsymbol{\theta}}_{\tau n}$, $\widetilde{\boldsymbol{\theta}}_{\tau n}$ and $\check{\boldsymbol{\theta}}_{\tau n}$, and their bootstrapping approximations. The data $\{Y_t\}_{t=1}^n$ are generated from a linear model with GARCH-X errors,

$$Y_t = 0.5X_{t-1} + u_t, \quad u_t = \sigma_t \varepsilon_t, \quad \sigma_t^2 = 1 + \alpha u_{t-1}^2 + \beta \sigma_{t-1}^2 + 0.1 v_{t-1}^2, \tag{3.1}$$

where $\{X_{t-1}\}$ and $\{v_{t-1}\}$ are *i.i.d.* standard normal random variables, $(\alpha, \beta) = (0.15, 0.8)$, and $\{\varepsilon_t\}$ are *i.i.d.* standard normal or standardized Student's t_5 random variables with variance one. Three sample sizes, $n = 500$, 1000 and 2000, are considered, and 1000 replications are generated for each sample size.

We apply the three estimating methods in Section 2.1 to the data, and obtain $\widehat{\boldsymbol{\theta}}_{\tau n}$, $\widetilde{\boldsymbol{\theta}}_{\tau n}$ and $\check{\boldsymbol{\theta}}_{\tau n}$ at two quantile levels $\tau = 0.05$ and 0.10, respectively. The bootstrapping procedure in Section 2.3 is conducted to approximate the covariance matrices Ξ_i for $i = 1, 2, 3$, where the size of bootstrapped samples is $B = 500$, and the random weights $\{\omega_t\}$ are generated from the standard exponential distribution. Then the asymptotic standard deviations (ASDs) can be calculated with the bootstrapping approximation. We can also construct confidence intervals (CIs) for each parameter based on three estimators and their ASDs. Specifically, given $\widehat{\boldsymbol{\theta}}_{\tau n}$, the 95% CI of $\theta_{\tau 0,j}$ $(j = 1, \ldots, 5)$ can be constructed by $\widehat{\theta}_{\tau n,j} \pm 1.96 \times \text{ASD}(\widehat{\theta}_{\tau n,j})$, where $\theta_{\tau 0,j}$ and $\widehat{\theta}_{\tau n,j}$ are the j-th element of $\boldsymbol{\theta}_{\tau 0}$ and $\widehat{\boldsymbol{\theta}}_{\tau n}$ respectively, and $\text{ASD}(\widehat{\theta}_{\tau n,j})$ is the ASD of $\widehat{\theta}_{\tau n,j}$. Likewise, the CIs based on other two estimation methods can also be constructed.

Tables 1-3 report the biases, empirical standard deviations (ESDs), and ASDs of $\widehat{\boldsymbol{\theta}}_{\tau n}$, $\widetilde{\boldsymbol{\theta}}_{\tau n}$ and $\check{\boldsymbol{\theta}}_{\tau n}$, respectively, as well as the empirical coverage rates (ECR) of 95% CIs. To save space, simulation results for the setting with $n = 1000$ are provided in the supplement. It can be seen that, as the sample size n increases, the biases, ESDs and ASDs decrease, and ESDs and ASDs get closer to each other, and, as innovations get more heavy-tailed, the standard deviations with respect to $\boldsymbol{\lambda}$ generally become larger while those related to b_τ get smaller. It is as expected since the value of $|b_\tau|$ is smaller for t_5 distributed innovations than that for normal distributed innovations. Moreover, as the quantile level τ increases from 0.05 to 0.10, the performance of the three estimators generally gets better when $\{\varepsilon_t\}$ follow t_5 distribution. However, when $\{\varepsilon_t\}$ follow normal distribution, the performance with respect to b_τ and ϕ becomes better, while that with respect to $\boldsymbol{\gamma}$ gets a little bit worse as τ increases. It may be due to the fact that, as τ gets closer to the center, more observations are available but b_τ approaches zero. Finally, except for α, the ECRs of 95% CIs for other parameters are close to the nominal level 0.95 for all settings. This may be due to the fact that the true value of α is relative small, and the inclusion of v_{t-1} into the GARCH model may hinder the accurate estimation for α.

For the comparison among three estimators, we have findings below. First, as the sample size increases, the jointly weighted estimator has smaller standard deviations than the jointly unweighted estimator. This is as expected since the efficiency gain from the weighting procedure becomes more evident when the sample size is larger. Secondly, the two-step method outperforms the jointly weighted method in estimating b_τ and ϕ, while performs a little bit worse for the other parameters. Note that ϕ is estimated by the least squares method in the two-step estimation while by the conditional quantile method in the joint estimation, which leads to more available observations for the former. Moreover, the better performance for b_τ is probably due to the better performance for ϕ. Finally, in general, the accuracy of CIs for three estimators is comparable.

3.2 Theoretical comparison among three estimators

The second experiment compares the asymptotic efficiency of the jointly weighted estimator $\widehat{\boldsymbol{\theta}}_{\tau n}$ with that of the jointly unweighted estimator $\widetilde{\boldsymbol{\theta}}_{\tau n}$ and the two-step estimator $\check{\boldsymbol{\theta}}_{\tau n}$, respectively. We generate a sequence of sample size $n = 10,000$ from model (3.1), where $\{\varepsilon_t\}$ are *i.i.d.* standard normal or standardized Student's t_5 random variables with variance one. For covariates X_{t-1} and v_{t-1}, we consider two cases: (1) $X_{t-1} = v_{t-1}$ and they are *i.i.d.* standard normal random variables; (2) $X_{t-1} = Y_{t-1}$ and $\{v_{t-1}\}$ are *i.i.d.* standard normal random variables. We consider different values for (α, β) and conduct the estimation at two quantile levels $\tau = 0.05$ and 0.10. The calculated asymptotic relative efficiency, $\text{ARE}(\widehat{\boldsymbol{\theta}}_{\tau n}, \widetilde{\boldsymbol{\theta}}_{\tau n})$ and $\text{ARE}(\widehat{\boldsymbol{\theta}}_{\tau n}, \breve{\boldsymbol{\theta}}_{\tau n})$, are given in Table 4.

From Table 4, it can be seen that $\text{ARE}(\widehat{\boldsymbol{\theta}}_{\tau n}, \widetilde{\boldsymbol{\theta}}_{\tau n}) > 1$ for all cases; i.e. the jointly weighted estimator is asymptotically more efficient than the jointly unweighted estimator. While, for the ARE of estimator $\widehat{\boldsymbol{\theta}}_{\tau n}$ to the two-step estimator $\breve{\boldsymbol{\theta}}_{\tau n}$, the observations are mixed. First, the jointly weighted estimator becomes more efficient as the coefficient α or the quantile level τ increases. This is as expected since larger α results in larger volatility and more data are available as τ increase, which leads to better performance of the weighting procedure. Moreover, the efficiency gain from the jointly weighted estimator is more evident when X_{t-1} is endogenous, while it becomes smaller as innovations get more heavy-tailed. Based on the simulation findings, we will focus on the jointly weighted and two-step estimating methods in the next section.

4 Empirical analysis

This section analyzes the daily log return of the Occidental Petroleum security. The data of daily closing prices (NYSE:OXY), denoted as p_t, span from January 2, 2008 to December 29, 2017, with 2470 observations in total. The time plot of centered log returns in percentage, i.e. $Y_t = r_t - n^{-1}\sum_{t=1}^{n} r_t$ with $r_t = 100(\ln p_t - \ln p_{t-1})$, shows evident volatility clustering; see Figure 1. Summary statistics of $\{Y_t\}$ are provided in Table 5. The negative sample skewness, together with the kurtosis greater than three, implies that the data are skewed and heavy-tailed.

The Occidental Petroleum Corporation is an oil and gas producer, and its stock returns can probably be affected by the lagged values of Y_t and oil price (Chernozhukov and Umantsev, 2001). Moreover, literature studies indicate that gold can be a hedge against stock returns (Baur and McDermott, 2010; Iqbal, 2017). This paper focuses on the effects of $\{Y_{t-1}\}$, the lagged values of crude oil returns and gold returns on $\{Y_t\}$. We use WTI Crude Oil Price and Gold Fixing Price 10:30 A.M. in London Bullion Market as the price series, which span from January 2, 2008 to December 29, 2017, and are downloaded from website of Federal Reserve Economic Data (FRED, *https://fred.stlouisfed.org/*). Figure 2 gives the time plots of their log returns in percentage, denoted as Oil_t and Gold_t. We first regress Y_t on lagged returns Y_{t-1}, Oil_{t-1} and Gold_{t-1}. The linear model is fitted by least squares method, and the regression residuals are calculated by $\breve{u}_t = Y_t - \breve{\boldsymbol{\phi}}_n' \boldsymbol{X}_{t-1}$, where $\boldsymbol{X}_{t-1} = (Y_{t-1}, \text{Oil}_{t-1}, \text{Gold}_{t-1})'$ and $\breve{\boldsymbol{\phi}}_n$ is the least squares estimate. The ACF and PACF plots of $\{\breve{u}_t^2\}$ show strong ARCH effects, which implies that a linear model with ARCH-type errors can be applied to $\{Y_t\}$. To further capture the possible influence of market volatility on $\{Y_t\}$, this paper includes the realized kernel variance ($\times 100^2$) of S&P 500 index, denoted by $\{\upsilon_t^2\}$, as the covariate in the GARCH model, where the realized variance series are downloaded from the Oxford-Man Institutes realized library (*http://realized.oxford-man.ox.ac.uk/*); see the time plot of $\{\upsilon_t^2\}$ in Figure 2. It was shown by Hansen and Lunde (2005) that GARCH$(1,1)$ model has satisfying performance in most practical applications. Finally, we consider a linear model with GARCH(1,1)-X errors for $\{Y_t\}$, where the regressors are Y_{t-1}, Oil_{t-1} and Gold_{t-1}, and the covariate in the GARCH-X model is $\{\upsilon_{t-1}^2\}$.

We aim to estimate the VaR for $\{Y_t\}$. Since 5% VaR is of common interest for practitioners, we focus on the conditional quantile of Y_t at $\tau = 0.05$, i.e. the negative 5% VaR. We first apply the two-step method, and obtain the conditional quantile below,

$$\begin{aligned}\breve{Q}_{Y_t}(0.05|\mathcal{F}_{t-1}) =& -0.071_{0.044} Y_{t-1} - 0.029_{0.034}\text{Oil}_{t-1} + 0.074_{0.063}\text{Gold}_{t-1} - 0.766_{0.202}\breve{\sigma}_t,\\ \breve{\sigma}_t^2 =& 1 + 0.203_{0.311}\breve{u}_{t-1}^2 + 0.765_{0.100}\breve{\sigma}_{t-1}^2 + 2.989_{0.710}\upsilon_{t-1}^2, \qquad (4.1)\end{aligned}$$

where the standard errors are in subscripts and calculated by the bootstrap method in Section 2.3. Estimates of b_τ and γ are significant at the 5% significance level, while regressors in the linear model are insignificant. Based on the weights calculated from model (4.1), we employ the jointly weighted estimation, and obtain the fitted model below,

$$\begin{aligned}\widehat{Q}_{Y_t}(0.05|\mathcal{F}_{t-1}) =& -0.050_{0.072}Y_{t-1} + 0.029_{0.051}\text{Oil}_{t-1} - 0.213_{0.070}\text{Gold}_{t-1} - 0.686_{0.163}\widehat{\sigma}_t,\\ \widehat{\sigma}_t^2 =& 1 + 0.367_{0.144}\widehat{u}_{t-1}^2 + 0.766_{0.097}\widehat{\sigma}_{t-1}^2 + 3.455_{0.131}\upsilon_{t-1}^2,\end{aligned} \tag{4.2}$$

where the standard errors in subscripts are also computed by the bootstrapping procedure. Compared with model (4.1), the estimate of Gold_{t-1} in model (4.2) is significantly negative at the 5% significance level, which implies that gold can be a safe haven in bearish market (lower quantiles). It is noteworthy that regression coefficients of models (4.1) and (4.2) are quite different both in magnitude and sign. This is as expected since they are estimated by different methods, and conditional heteroscedasticity is taken into account when performing estimation in model (4.2) while ignored by model (4.1). Moreover, as the inclusion of market volatility in GARCH models, the estimates of the GARCH parameter β in models (4.1) and (4.2) are smaller than usual estimates in GARCH models without exogenous variables. Similar finding was also mentioned in Hwang and Satchell (2005), and they concluded that the long persistency frequently found in volatility process may be due to missing time-varying components. In addition, the estimate of top Lyapunov component for model (4.2) is -0.194, and hence the process defined by model (4.2) is stationary.

We next evaluate the forecasting performance of the jointly weighted (JW) and two-step (TS) estimating methods by using a rolling procedure for conditional quantile forecasts at $\tau = 0.01$ and $\tau = 0.05$, which are negative 1% VaR and 5% VaR, respectively. The fixed moving window of size 1000 is used for the rolling forecasting procedure. Specifically, we conduct the conditional quantile estimation using the linear model with GARCH$(1,1)$-X errors for each moving window, and compute the one-step-ahead conditional quantile forecast for the next trading day, i.e. the forecast of $Q_{Y_{n+1}}(\tau|\mathcal{F}_n)$. The model estimates are updated by moving forward the window until we reach the end of the data set. Finally, we obtain 1469 one-day-ahead 1% (or 5%) VaRs in total. For illustration, the rolling forecasts at $\tau = 1\%$ and 5% for $\{Y_t\}$ are displayed in Figure 1, where the rolling forecasts are obtained by the jointly weighted approach. Obviously the magnitudes of VaRs increase as the volatility of data gets larger. It can also be seen that Y_t falls below its one-day negative 5% VaR occasionally and its one-day negative 1% VaR even more rarely.

To compare the forecasting performance of the proposed methods with existing conditional quantile estimation, we also perform the rolling forecasting procedure using the fully parametric (PAR) method, the filtered historical simulation (FHS) method (Kuester, Mittnik, and Paolella, 2006) and the conditional auto-regressive VaR-method called CAViaR (Engle and Manganelli, 2004). For PAR and FHS, a linear model with GARCH(1,1)-X errors defined by (2.1) and (1.1) is fitted to the data, and the parameters are estimated by the maximum likelihood estimation (MLE) with innovations $\{\varepsilon_t^*\}$ being skewed Student's t distributed. Figure 3 gives the Q-Q plot of residuals $\{\breve{\varepsilon}_t^*\}$ against the fitted skewed Student's t distribution, and their density plots are also presented. It can be seen that they are very close to each other, and we may argue that the PAR and FHS methods reach almost their maximum power. The $100\tau\%$ negative VaR for PAR is computed by $\breve{\boldsymbol{\phi}}_n'\boldsymbol{X}_{t-1} + \breve{Q}_\tau\breve{\sigma}_t^*$, where $\breve{\boldsymbol{\phi}}_n$ and $\breve{\sigma}_t^* = \sigma_t^*(\breve{\boldsymbol{\lambda}}_n)$ are maximum likelihood estimates and $\breve{Q}_\tau$ is the τth quantile of the estimated skewed Student's t distribution. While the $100\tau\%$ negative VaR for FHS is calculated with $\breve{Q}_\tau$ replaced by the sample τth quantile of the filtered residuals. CAViaR refers to the indirect GARCH(1,1)-based CAViaR method in Engle and Manganelli (2004).

To evaluate the forecasting performance of aforementioned five VaR estimating methods, we calculate the empirical coverage rate (ECR), and perform the VaR backtests for VaR forecasts. Specifically, ECR is calculated as the proportion of observations that fall below the corresponding conditional quantile forecast for the last 1469 data points. Two VaR backtests, i.e. the likelihood ratio test for correct conditional coverage (CC) in Christoffersen (1998) and the dynamic quantile (DQ) test in Engle and Manganelli (2004), are employed. Denote the hit by $H_t = I(Y_t < Q_{Y_t}(\tau|\mathcal{F}_{t-1}))$. The null hypothesis of CC test is that, conditional on $\mathcal{F}_{t-1}$, $\{H_t\}$ are *i.i.d.* Bernoulli random variables with success probability being τ. For the DQ test, following Engle and Manganelli (2004), we regress H_t on regressors including a constant, four lagged hits H_{t-i}, $i = 1, 2, 3, 4$, and the contemporaneous VaR forecast. The null hypothesis of DQ test is that all regression coefficients are zero and the intercept equals to the quantile level τ.

Table 6 reports ECRs and p-values of two VaR backtests for five estimating methods at the upper and lower 1% and 5% conditional quantiles. It can be seen that none of the five methods perform well at the lower 5% quantile with p-values smaller than 0.05, however they are adequate at other three quantiles. In terms of backtests, JW and TS methods are comparable with other three methods. With respect to ECRs, it can be seen that the ECRs of JW or TS approach are closest to the nominal quantile level τ. We can conclude that, overall the proposed JW and TS methods outperform other three competitors in forecasting VaRs for the considered Occidental Petroleum returns. Moreover, our estimating methods, especially the JW method, only use the information at one quantile level, and the corresponding estimator of the parameter vector $\boldsymbol{\lambda}$ can be much less efficient when the quantile level is near to zero or one (Zou and Yuan, 2008). This hence further demonstrates the usefulness of two proposed methods.

5 Conclusion and discussion

This paper focuses on the conditional quantile estimation for linear models with GARCH-X errors. Three conditional quantile estimators, i.e. jointly weighted estimator, jointly unweighted estimator and two-step estimator are proposed. The root-n consistency and asymptotic normality are established for the three proposed estimators, where the bracketing method (Pollard, 1985) is adopted to overcome the theoretical difficulties due to the non-convex and non-differentiable objective functions. Simulation results indicate that the jointly weighted approach generally outperforms its unweighted counterpart when sample size is large. Compared with the two-step estimating method, the jointly weighted method is preferred when the data are much volatile, and the efficiency gain is highlighted especially when linear regressors are endogenous and the quantile level is not too far from the center. Better forecasting performance on Value-at-Risk can be achieved by the proposed methods, as confirmed by empirical evidence.

It is also of interest to consider the linear model with conditional heteroscedasticity of unknown form, $Y_t = \boldsymbol{\phi}'\boldsymbol{X}_{t-1} + \sigma(\boldsymbol{X}_{t-1})\varepsilon_t$, where $\sigma(\cdot)$ is an unknown function; see Zhao (2001). Then, conditional on $\mathcal{F}_{t-1}$, the τth quantile of Y_t has a form of $Q_{Y_t}(\tau|\mathcal{F}_{t-1}) = \boldsymbol{\phi}'\boldsymbol{X}_{t-1} + b_\tau\sigma(\boldsymbol{X}_{t-1})$. As a result, an adaptive weighted conditional quantile estimation (WCQE) can be constructed below,

$$(\widehat{\boldsymbol{\phi}}_n', \widehat{b}_{\tau n}, \widehat{\sigma}(\cdot)) = \arg \min_{\boldsymbol{\phi}, b, \sigma(\cdot)} \sum_{t=1}^{n} \widehat{w}_t \rho_\tau \{Y_t - \boldsymbol{\phi}'\boldsymbol{X}_{t-1} - b\sigma(\boldsymbol{X}_{t-1})\},$$

where the weights $\{\widehat{w}_t\}$ are the initial estimators of $\{\sigma^{-1}(\boldsymbol{X}_{t-1})\}$. We may employ some nonparametric methods, such as the k-nearest neighbors and kernel smoothing approaches, to fit the unknown function $\sigma(\cdot)$. We leave it for future research.

Supplementary Materials

Technical details for all theorems and corollaries, together with additional simulation results for Section 3.1, can be found in the supplementary materials.

Acknowledgements

We are deeply grateful to the Co-Editor, Associate Editor and two anonymous referees for their valuable comments and suggestions, leading to the substantial improvement of the article. Zhu's research was partially supported by Shanghai Pujiang Program 2019PJC051, and Key Laboratory for Applied Statistics of MOE, Northeast Normal University grant 130028906. Li's research was partially supported by Hong Kong Research Grant Council grant 17304617.

References

Apergis, N. (1998). Stock market volatility and deviations from macroeconomic fundamentals: evidence from garch and garch-x models. *Kredit und Kapital 31*, 400–412.

Baur, D. G. and T. K. McDermott (2010). Is gold a safe haven? International evidence. *Journal of Banking & Finance 34*, 1886–1898.

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

Chernozhukov, V. and L. Umantsev (2001). Conditional value-at-risk: Aspects of modeling and estimation. *Empirical Economics 26*, 271–292.

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

Duffie, D. and J. Pan (1997). An overview of value at risk. *The Journal of Derivatives 4*, 7–49.

Durbin, J. (1960). Estimation of parameters in time-series regression models. *Journal of the Royal Statistical Society. Series B*, 139–153.

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

Engle, R. F. and G. M. Gallo (2006). A multiple indicators model for volatility using intra-daily data. *Journal of Econometrics 131*, 3–27.

Engle, R. F. and S. Manganelli (2004). CAViaR: conditional autoregressive value at risk by regression quantiles. *Journal of Business & Economic Statistics 22*, 367–381.

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

Gay, R. D. (2016). Effect of macroeconomic variables on stock market returns for four emerging economies: Brazil, russia, india, and china. *International Business & Economics Research Journal 15*, 119–126.

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

Han, H. and D. Kristensen (2014). Asymptotic theory for the QMLE in GARCH-X models with stationary and nonstationary covariates. *Journal of Business & Economic Statistics 32*, 416–429.

Hansen, P. R. and A. Lunde (2005). A forecast comparison of volatility models: does anything beat a GARCH (1, 1)? *Journal of Applied Econometrics 20*, 873–889.

Hossain, S. and M. Ghahramani (2016). Shrinkage Estimation of Linear Regression Models with GARCH Errors. *Journal of Statistical Theory and Applications 15*, 405–423.

Hwang, S. and S. E. Satchell (2005). GARCH model with cross-sectional volatility: GARCHX models. *Applied Financial Economics 15*, 203–216.

Iqbal, J. (2017). Does gold hedge stock market, inflation and exchange rate risks? An econometric investigation. *International Review of Economics & Finance 48*, 1–17.

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

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

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

Kutner, M. H., C. J. Nachtsheim, J. Neter, and W. Li (2005). *Applied linear statistical models* (5th ed.). McGraw-Hill.

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

Li, W. K., S. Ling, and M. McAleer (2002). Recent theoretical results for time series models with GARCH errors. *Journal of Economic Surveys 16*, 245–269.

Martins-Filho, C., F. Yao, and M. Torero (2018). Nonparametric estimation of conditional value-at-risk and expected shortfall based on extreme value theory. *Econometric Theory 34*, 23–67.

Medeiros, M. C. and E. F. Mendes (2016). L(1)-regularization of high-dimensional time-series models with non-Gaussian and heteroskedastic errors. *Journal of Econometrics 191*, 255–271.

Noh, J. and S. Lee (2016). Quantile Regression for Location-Scale Time Series Models with Conditional Heteroscedasticity. *Scandinavian Journal of Statistics 43*, 700–720.

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

Pollard, D. (1991). Asymptotics for least absolute deviation regression estimators. *Econometric Theory 7*, 186–199.

Ryden, T., T. Terasvirta, and S. Asbrink (1998). Stylized facts of daily return series and the hidden markov model. *Journal of Applied Econometrics 13*, 217–244.

Serfling, R. J. (2009). *Approximation theorems of mathematical statistics*, Volume 162. New York: John Wiley & Sons.

Shephard, N. and K. Sheppard (2010). Realising the future: forecasting with high-frequency-based volatility (HEAVY) models. *Journal of Applied Econometrics 25*, 197–231.

Taylor, J. W. (2019). Forecasting value at risk and expected shortfall using a semiparametric approach based on the asymmetric Laplace distribution. *Journal of Business & Economic Statistics 37*, 121–133.

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

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

Wang, C.-S. and Z. Zhao (2016). Conditional Value-at-Risk: Semiparametric estimation and inference. *Journal of Econometrics 195*, 86–103.

Wang, H., G. Li, and C.-L. Tsai (2007). Regression coefficient and autoregressive order shrinkage and selection via the lasso. *Journal of the Royal Statistical Society, Series B 69*, 63–78.

Wu, G. and Z. Xiao (2002). An analysis of risk measures. *Journal of Risk 4*, 53–76.

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

Zhao, Q. (2001). Asymptotically efficient median regression in the presence of heteroskedasticity of unknown form. *Econometric Theory 17*, 765–784.

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

Zhu, Q., R. Zeng, and G. Li (2019). Bootstrap inference for GARCH models by the least absolute deviation estimation. *Journal of Time Series Analysis*, DOI: 10.1111/jtsa.12474.

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

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

Table 1: Biases, ESDs, ASDs and ECRs of 95% confidence intervals for $\widehat{\boldsymbol{\theta}}_{\tau n}$ at $\tau = 0.05$ and 0.10, for normally distributed X_{t-1} and v_{t-1}. The innovations are normally or Student's t_5 distributed.

			Normal				t_5			
τ	n		Bias	ESD	ASD	ECR	Bias	ESD	ASD	ECR
0.05	500	b_τ	-0.405	0.604	0.476	0.895	-0.282	0.523	0.420	0.911
		α	-0.023	0.068	0.067	0.890	-0.011	0.084	0.079	0.881
		β	-0.074	0.114	0.108	0.934	-0.065	0.123	0.119	0.939
		π	0.148	0.326	0.256	0.933	0.157	0.346	0.291	0.932
		ϕ	-0.007	0.375	0.334	0.931	0.005	0.400	0.384	0.946
	2000	b_τ	-0.122	0.255	0.259	0.938	-0.110	0.261	0.249	0.941
		α	-0.007	0.040	0.039	0.917	-0.004	0.049	0.044	0.905
		β	-0.023	0.044	0.051	0.956	-0.025	0.054	0.059	0.951
		π	0.066	0.171	0.161	0.922	0.070	0.205	0.181	0.924
		ϕ	-0.002	0.183	0.180	0.943	0.024	0.213	0.208	0.946
0.10	500	b_τ	-0.358	0.539	0.462	0.909	-0.237	0.427	0.377	0.933
		α	-0.025	0.070	0.069	0.883	-0.017	0.077	0.074	0.876
		β	-0.079	0.129	0.126	0.953	-0.066	0.129	0.130	0.949
		π	0.121	0.295	0.266	0.943	0.111	0.296	0.276	0.942
		ϕ	0.000	0.303	0.302	0.942	0.008	0.287	0.295	0.951
	2000	b_τ	-0.104	0.231	0.230	0.950	-0.081	0.216	0.198	0.945
		α	-0.008	0.041	0.040	0.915	-0.004	0.046	0.042	0.903
		β	-0.023	0.048	0.056	0.967	-0.023	0.056	0.059	0.961
		π	0.054	0.181	0.165	0.917	0.056	0.189	0.171	0.910
		ϕ	0.003	0.149	0.152	0.947	0.014	0.153	0.150	0.940

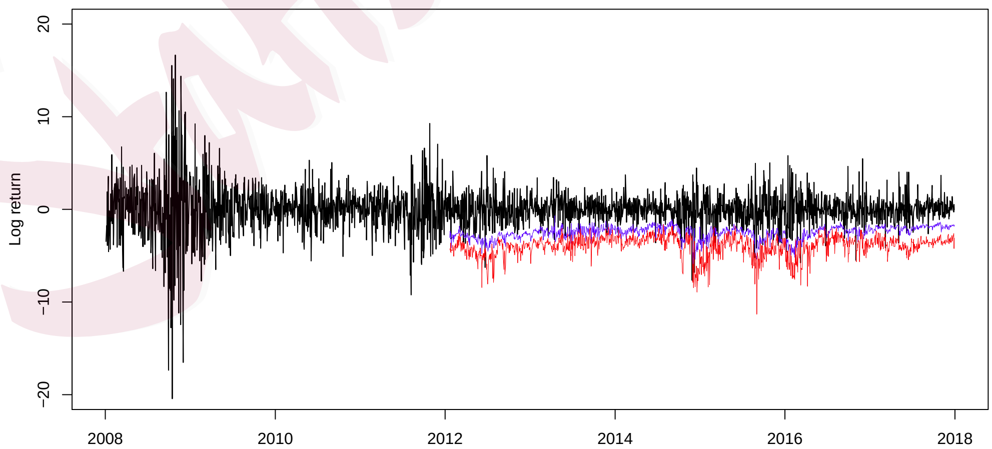

Figure 1: Time plot for centered daily log returns in percentage (black line) of NYSE:OXY stock from January 3, 2008 to December 29, 2017, with one-day negative VaR forecasts at levels of 1% (red line) and 5% (blue line) from February 3, 2012 to December 31, 2017.

Table 2: Biases, ESDs, ASDs and ECRs of 95% confidence intervals for $\widetilde{\boldsymbol{\theta}}_{\tau n}$ at $\tau = 0.05$ and 0.10, for normally distributed X_{t-1} and v_{t-1}. The innovations are normally or Student's t_5 distributed.

			Normal				t_5			
τ	n		Bias	ESD	ASD	ECR	Bias	ESD	ASD	ECR
0.05	500	b_τ	-0.256	0.503	0.390	0.938	-0.184	0.449	0.386	0.944
		α	-0.016	0.069	0.066	0.900	-0.013	0.081	0.077	0.870
		β	-0.048	0.103	0.096	0.951	-0.041	0.113	0.113	0.946
		π	0.149	0.365	0.302	0.942	0.154	0.363	0.338	0.955
		ϕ	0.001	0.366	0.349	0.941	0.015	0.408	0.408	0.949
	2000	b_τ	-0.076	0.250	0.230	0.936	-0.082	0.264	0.243	0.955
		α	-0.004	0.040	0.040	0.911	-0.004	0.049	0.047	0.910
		β	-0.015	0.045	0.047	0.950	-0.018	0.057	0.060	0.938
		π	0.064	0.187	0.187	0.943	0.071	0.215	0.203	0.941
		ϕ	-0.004	0.192	0.187	0.936	0.024	0.220	0.217	0.950
0.10	500	b_τ	-0.223	0.423	0.382	0.947	-0.153	0.362	0.339	0.959
		α	-0.018	0.072	0.069	0.894	-0.014	0.078	0.073	0.878
		β	-0.049	0.108	0.113	0.963	-0.040	0.113	0.122	0.954
		π	0.139	0.360	0.322	0.957	0.123	0.354	0.321	0.948
		ϕ	-0.001	0.307	0.309	0.938	0.009	0.290	0.305	0.961
	2000	b_τ	-0.070	0.233	0.218	0.952	-0.064	0.220	0.200	0.952
		α	-0.004	0.044	0.042	0.920	-0.003	0.048	0.045	0.912
		β	-0.015	0.050	0.054	0.966	-0.018	0.060	0.061	0.948
		π	0.063	0.217	0.194	0.932	0.053	0.204	0.191	0.922
		ϕ	0.006	0.154	0.158	0.944	0.016	0.159	0.155	0.945

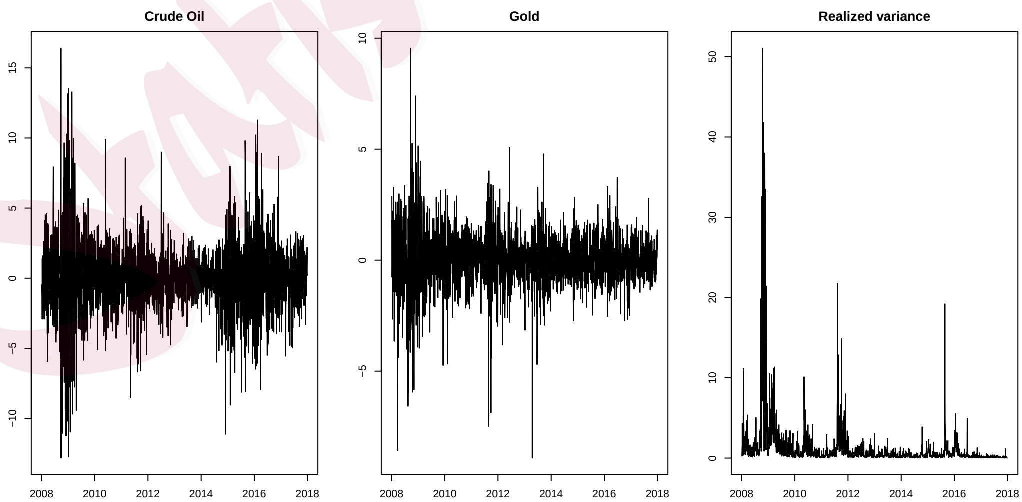

Figure 2: Time plots for daily log returns in percentage of WTI Crude Oil Prices (left) and LBMA Gold Prices (middle), and realized variance($\times 100^2$) of S&P 500 index (right) from January 3, 2008 to December 29, 2017.

Table 3: Biases, ESDs, ASDs and ECRs of 95% confidence intervals for $\check{\boldsymbol{\theta}}_{\tau n}$ at $\tau = 0.05$ and 0.10, for normally distributed X_{t-1} and v_{t-1}. The innovations are normally or Student's t_5 distributed.

			Normal				t_5			
τ	n		Bias	ESD	ASD	ECR	Bias	ESD	ASD	ECR
0.05	500	b_τ	-0.229	0.437	0.383	0.951	-0.170	0.423	0.376	0.948
		α	-0.017	0.067	0.066	0.901	-0.013	0.080	0.076	0.875
		β	-0.041	0.096	0.095	0.953	-0.037	0.111	0.111	0.947
		π	0.142	0.333	0.312	0.958	0.148	0.379	0.353	0.962
		ϕ	0.005	0.204	0.208	0.953	0.009	0.208	0.203	0.953
	2000	b_τ	-0.070	0.241	0.224	0.944	-0.080	0.257	0.237	0.952
		α	-0.004	0.040	0.040	0.918	-0.005	0.048	0.046	0.911
		β	-0.014	0.044	0.047	0.952	-0.017	0.056	0.059	0.940
		π	0.067	0.193	0.187	0.940	0.066	0.214	0.202	0.935
		ϕ	0.003	0.106	0.104	0.947	0.009	0.109	0.103	0.947
0.10	500	b_τ	-0.214	0.420	0.374	0.943	-0.146	0.346	0.331	0.960
		α	-0.018	0.071	0.068	0.889	-0.015	0.077	0.072	0.877
		β	-0.046	0.105	0.111	0.967	-0.038	0.111	0.120	0.957
		π	0.129	0.336	0.331	0.963	0.128	0.362	0.334	0.960
		ϕ	0.005	0.204	0.208	0.953	0.009	0.208	0.203	0.953
	2000	b_τ	-0.068	0.227	0.215	0.955	-0.062	0.213	0.197	0.955
		α	-0.005	0.043	0.042	0.922	-0.003	0.047	0.045	0.920
		β	-0.015	0.049	0.053	0.962	-0.018	0.058	0.060	0.942
		π	0.065	0.209	0.196	0.943	0.055	0.206	0.191	0.922
		ϕ	0.003	0.106	0.104	0.947	0.009	0.109	0.103	0.947

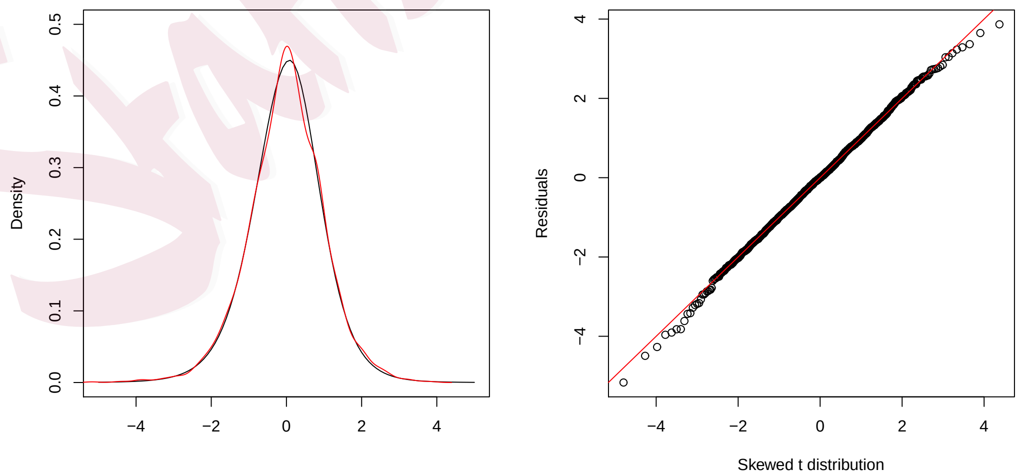

Figure 3: Density plots (left) of model residuals $\breve{\varepsilon}_t$ (red line) and the fitted skewed t distribution (black line), and Q-Q plot (right) for model residuals $\breve{\varepsilon}_t$ against the fitted skewed t distribution.

Table 4: ARE($\widehat{\boldsymbol{\theta}}_{\tau n}, \widetilde{\boldsymbol{\theta}}_{\tau n}$) and ARE($\widehat{\boldsymbol{\theta}}_{\tau n}, \check{\boldsymbol{\theta}}_{\tau n}$) for regression model with GARCH(1, 1)-X errors of different values for (α, β). The innovations $\{\varepsilon_t\}$ follow the standard normal and Student's t_5 distributions, and $\tau = 0.05$ or 0.10, based on a generated sequence of $n = 10,000$. ARE$_1$ and ARE$_2$ represent ARE($\widehat{\boldsymbol{\theta}}_{\tau n}, \widetilde{\boldsymbol{\theta}}_{\tau n}$) and ARE($\widehat{\boldsymbol{\theta}}_{\tau n}, \check{\boldsymbol{\theta}}_{\tau n}$), respectively.

		β	0.15			0.30		0.80
	τ	α	0.40	0.60	0.80	0.40	0.60	0.15
			X_{t-1} and v_{t-1} are normal distributed					
ARE$_1$	0.05	Normal	1.076	1.201	1.599	1.099	1.311	1.089
		t_5	1.087	1.171	1.376	1.111	1.281	1.115
	0.10	Normal	1.075	1.200	1.596	1.098	1.310	1.089
		t_5	1.086	1.169	1.373	1.110	1.279	1.115
ARE$_2$	0.05	Normal	0.813	1.017	1.684	0.845	1.194	0.817
		t_5	0.736	0.873	1.202	0.766	1.022	0.759
	0.10	Normal	0.895	1.107	1.810	0.930	1.298	0.905
		t_5	0.850	0.996	1.357	0.884	1.164	0.884
			$X_{t-1} = Y_{t-1}$ and v_{t-1} is normal distributed					
ARE$_1$	0.05	Normal	1.095	1.285	1.998	1.125	1.459	1.100
		t_5	1.118	1.243	1.550	1.151	1.403	1.142
	0.10	Normal	1.095	1.284	1.992	1.124	1.458	1.100
		t_5	1.117	1.242	1.553	1.150	1.402	1.141
ARE$_2$	0.05	Normal	0.930	1.638	4.172	0.992	2.242	0.856
		t_5	0.865	1.168	1.963	0.912	1.495	0.841
	0.10	Normal	1.019	1.756	4.150	1.085	2.388	0.946
		t_5	0.991	1.315	2.172	1.046	1.682	0.979

Table 5: Summary statistics for centered log returns in percentage of NYSE:OXY stock.

Min	Max	Mean	Median	Std. Dev.	Skewness	Kurtosis
-20.436	16.656	0.000	0.027	2.271	-0.259	11.025

Table 6: Empirical coverage rate (%) and p-values of two VaR backtests of five estimation methods at the 1%, 5%, 95% and 99% conditional quantiles. JW, TS, PAR, FHS and CAV represent the jointly weighted method, the two-step method, the parametric method, the filtered historical simulation method and CAViaR method, respectively.

	$\tau = 1\%$			$\tau = 5\%$			$\tau = 95\%$			$\tau = 99\%$		
	ECR	CC	DQ	ECR	CC	DQ	ECR	CC	DQ	ECR	CC	DQ
JW	0.95	0.86	0.70	4.77	0.65	0.01	94.96	0.92	0.76	98.64	0.32	0.69
TS	0.95	0.86	0.36	4.97	0.77	0.02	95.23	0.65	0.77	98.91	0.79	0.84
PAR	0.88	0.80	0.98	4.56	0.64	0.01	95.64	0.03	0.38	98.77	0.56	0.81
FHS	1.16	0.69	0.37	4.56	0.64	0.03	95.23	0.65	0.78	98.84	0.69	0.96
CAV	1.23	0.56	0.25	4.70	0.55	0.01	96.19	0.01	0.28	99.18	0.69	0.99