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

Statistica Sinica: Preprint
doi:10.5705/ss.202019.0003

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 estimations are an essential ingredient in modern risk management, and many other applications, where the conditional heteroscedastic structure is usually assumed to capture the volatility in financial time series. This study examines linear quantile regression models with GARCH-X errors. These models include the most popular generalized autoregressive conditional heteroscedasticity (GARCH) as a special case, and 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 developed to approximate their asymptotic distributions. The finite-sample performance of the proposed estimators is examined using 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.

Statistica Sinica: Preprint
doi:10.5705/ss.202019.0003

1 Introduction

Linear models are powerful tools used to explore 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 the 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 consider conditional heteroscedasticity when a linear model is fitted to financial time series data.

Since the appearance of the autoregressive conditional heteroscedastic (ARCH) and generalized autoregressive conditional heteroscedastic (GARCH) models (Engle, 1982; Bollerslev, 1986), time series models with ARCH-type errors have become common in empirical studies (Li, Ling, and McAleer, 2002). Motivated by these stylized facts and the success of GARCH-X models in interpreting the volatility for financial data, this study focuses on the following 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 that can be endogenous or exogenous, and $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_m)' \in \mathbb{R}^m$ is a vector of linear coefficients. The regression error u_t follows the GARCH-X model (Apergis, 1998),

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

Statistica Sinica: Preprint
doi:10.5705/ss.202019.0003

where $\omega^* > 0$; $\alpha_i^* \geqslant 0$, for $i = 1, \ldots, q$; $\beta_j^* \geqslant 0$, for $j = 1, \ldots, p$; $\boldsymbol{V}_{t-1} = (\upsilon_{1,t-1}^2, \ldots, \upsilon_{d,t-1}^2)' \in \mathbb{R}^d$ 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 the 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 the 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 comprise realized volatility measures (Engle and Gallo, 2006; Hwang and Satchell, 2005), or economic and financial indicators (Glosten, Jagannathan, and Runkle, 1993). Model (1.1) has become increasingly popular for 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 in risk management and capital regulation in the financial industry (Duffie and Pan, 1997; Taylor, 2019). Because VaR is a tail quantile of the conditional return distribution, its evaluation 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: the parametric approach, using a specific parametric model with a known innovation distribution; the semiparametric approach, using a filtered historical simulation or quantile regression; and the nonparametric approach, using the conditional autoregressive VaR-method or a kernel density estimation; see, for example, Engle and Manganelli (2004), Wang and Zhao (2016), and Taylor (2019) for further detail. Specifically, a quantile regression (Koenker and Bassett, 1978) is suitable when modeling the VaR based on a specific parametric model, without assuming a distribution form on the innovations. Moreover, it is robust to extreme values and facilitates a distribution-free inference.

Statistica Sinica: Preprint
doi:10.5705/ss.202019.0003

This motivates us to focus on the conditional quantile estimation and VaR prediction for the linear model with GARCH-X errors. Many studies have examined quantile regressions for conditional heteroscedastic models. For example, Koenker and Zhao (1996) and Xiao and Koenker (2009) considered a quantile regression for the linear (G)ARCH models proposed by Taylor (2008); Lee and Noh (2013) and Zheng et al. (2018) investigated a quantile regression for Bollerslev's GARCH models; and Noh and Lee (2016) studied a quantile regression for ARMA models with asymmetric GARCH errors. However, few works have examined quantile estimations for linear models with GARCH-X errors. This study aims to fill this gap; the main contributions are summarized below. Section 2 contains the methods and theoretical results.

(a) Section 2.1 proposes three conditional quantile estimators: 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 the regression coefficients and the GARCH-X parameters using a quantile regression. The two-step estimator is a hybrid of a least squares estimator for linear coefficients and a conditional quantile estimator for GARCH-X parameters. Moreover, to take into account 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 of the proposed estimators. Owning 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 study adopts the bracketing method (Pollard, 1985) to overcome this difficulty. In addition, only $E(Y_t^2) < \infty$

Statistica Sinica: Preprint
doi:10.5705/ss.202019.0003

is required in order to derive the asymptotic normality for an AR-GARCH model; thus, the proposed estimating methods are 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 that approximates the covariance matrices directly. A theoretical justification of the bootstrap method is also provided.

Section 3 conducts simulation experiments to evaluate the finite-sample performance of the three proposed estimators. Section 4 provides a real example on VaR prediction, and Section 5 concludes the paper. Technical proofs of all theorems and corollaries are relegated to the online Supplementary Material. Throughout this 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 the regression model, u_t is a regression error, $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_d)'$ is a

Statistica Sinica: Preprint
doi:10.5705/ss.202019.0003

d-dimensional coefficient vector of covariates $\boldsymbol{V}_{t-1} = (\upsilon^2_{1,t-1}, \ldots, \upsilon^2_{d,t-1})'$ 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 ε_t is an *i.i.d.* random variable 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 let b_τ be the τth quantile of ε_t. Assume that ε_t is independent of $\mathcal{F}_{t-1}$; 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 $\omega^* = \mathrm{var}(\varepsilon_t)$, $\varepsilon^*_t = \varepsilon_t/\sqrt{\omega^*}$, $\sigma^*_t = \sigma_t\sqrt{\omega^*}$, $\alpha^*_i = \omega^*\alpha_i$, $\beta^*_j = \beta_j$, and $\boldsymbol{\pi}^* = \omega^*\boldsymbol{\pi}$. The GARCH-X error in (2.2) then has the standard form of (1.1). Note that the GARCH-X model extends Bollerslev's GARCH model by including additional predictors. Because model (1.1) suffers from an identifiability problem in the quantile estimation (Xiao and Koenker, 2009; Lee and Noh, 2013; Noh and Lee, 2016), we use the GARCH-X form given in (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 $\sigma^2_t(\boldsymbol{\lambda}) = 1 + \sum_{i=1}^q \alpha_i u^2_{t-i}(\boldsymbol{\phi}) + \sum_{j=1}^p \beta_j \sigma^2_{t-j}(\boldsymbol{\lambda}) + \boldsymbol{\pi}'\boldsymbol{V}_{t-1}$. Note that function $\sigma^2_t(\boldsymbol{\lambda})$ is defined recursively and, thus, depends on infinite past observations. Therefore, initial values are required, in practice. Here, we set $u_t(\boldsymbol{\phi}) = 0$ and $\sigma^2_t(\boldsymbol{\lambda}) = 1$, 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 using a quantile regression. Then, a joint conditional quantile estimator can

Statistica Sinica: Preprint
doi:10.5705/ss.202019.0003

be defined as

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

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

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

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 makes the theoretical derivation and numerical optimization challenging.

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

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

where $\check{\sigma}_t^2(\boldsymbol{\gamma})$ is calculated recursively using $\check{\sigma}_t^2(\boldsymbol{\gamma}) = 1 + \sum_{i=1}^{q}\alpha_i\check{u}_{t-i}^2 + \sum_{j=1}^{p}\beta_j\check{\sigma}_{t-j}^2(\boldsymbol{\gamma}) + \boldsymbol{\pi}'\boldsymbol{V}_{t-1}$, given the 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 $\check{\boldsymbol{\gamma}}_{\tau n}$; see Corollary 2 in Section 2.2. Denote $\check{\boldsymbol{\theta}}_{\tau n} = (\check{\boldsymbol{\gamma}}_{\tau n}', \check{\boldsymbol{\phi}}_n')'$. 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.

Statistica Sinica: Preprint
doi:10.5705/ss.202019.0003

We can verify that the initial values of $\breve{u}_t$, $u_t(\boldsymbol{\phi})$, $\breve{\sigma}_t^2(\boldsymbol{\gamma})$, and $\sigma_t^2(\boldsymbol{\lambda})$ have no effect on the asymptotic distributions of the 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):

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

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 $\widehat{\boldsymbol{\theta}}_{\tau n}$ defined by (2.5); see Zhu, Zheng, and Li (2018). Let $(\widehat{m},\widehat{d},\widehat{p},\widehat{q}) = \arg\min_{m,d,p,q} \mathrm{BIC}_\tau(m,d,p,q)$, 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. Using a method similar to that in the proof of Theorem 5 in Zhu, Zheng, and Li (2018), we can show that the proposed BIC in (2.6) is consistent when the true orders satisfy $m_0 \leqslant m_{\max}$, $d_0 \leqslant d_{\max}$, $p_0 \leqslant p_{\max}$, and $q_0 \leqslant q_{\max}$. We can define the BIC for the jointly unweighted estimator $\widetilde{\boldsymbol{\theta}}_{\tau n}$ and verify its consistency in a similar manner.

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 using the BIC. For example, we may select orders based on the background of the data and using other quantitative tools, such as the autocorrelation function (ACF) and the partial autocorrelation function (PACF).

2.2 Asymptotic properties

Let $\boldsymbol{\theta} = (b,\boldsymbol{\lambda}')'$ be the parameter vector of model (2.3), and let $\boldsymbol{\theta}_{\tau 0} = (b_\tau,\boldsymbol{\lambda}_0')' = (b_\tau,\boldsymbol{\gamma}_0',\boldsymbol{\phi}_0')'$ 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 condi-

Statistica Sinica: Preprint
doi:10.5705/ss.202019.0003

tional 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

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

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 $\widehat{\boldsymbol{\theta}}_{\tau n}$. Because 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.*

We focus 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 the GARCH-X model (2.2). Moreover, for the model identification, the intercept should not be included in the regression model; that is, $\boldsymbol{X}_{t-1}$ does not incorporate an intercept.

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 $n \to \infty$.*

Statistica Sinica: Preprint
doi:10.5705/ss.202019.0003

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 the technical proofs, although 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) $\sqrt{n}(\widehat{\boldsymbol{\theta}}_{\tau n} - \boldsymbol{\theta}_{\tau 0}) \to N(\boldsymbol{0}, \Xi_1)$ *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 the standard arguments for conditional quantile estimators, implies the root-n consistency of $\widehat{\boldsymbol{\theta}}_{\tau n}$. Then, the asymptotic normality follows. Moreover, in contrast to 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 $\boldsymbol{\lambda}_0$.

Statistica Sinica: Preprint
doi:10.5705/ss.202019.0003

Let $\boldsymbol{W}_t = (\sigma_t, 0.5\sigma_t^{-1} b_\tau \partial\sigma_t^2(\boldsymbol{\lambda}_0)/\partial\boldsymbol{\gamma}')'$ and $\boldsymbol{M}_t = \boldsymbol{X}_{t-1} + 0.5\sigma_t^{-1} b_\tau \partial\sigma_t^2(\boldsymbol{\lambda}_0)/\partial\boldsymbol{\phi}'$. 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 $\omega^* = \operatorname{var}(\varepsilon_t)$ 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 $\Sigma_{12}(\tau) = \kappa f_\varepsilon^{-1}(b_\tau)\Omega_1^{-1}\Gamma_2 D_0^{-1} - \Omega_1^{-1}\Gamma_1\Sigma_{22}$, $\Sigma_{22} = \omega^* D_0^{-1} D_2 D_0^{-1}$, 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 that in Theorem 2, we 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:*

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

Statistica Sinica: Preprint
doi:10.5705/ss.202019.0003

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 and 2 require a 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 set $\widehat{\sigma}_t = \sqrt{1 + \check{\gamma}_n'\check{\boldsymbol{z}}_t}$, where $\check{\boldsymbol{z}}_t = (\check{u}_{t-1}^2, \ldots, \check{u}_{t-q}^2, \check{\sigma}_{t-1}^2(\check{\gamma}_n), \ldots, \check{\sigma}_{t-p}^2(\check{\gamma}_n), v_{1,t-1}^2, \ldots, v_{d,t-1}^2)'$.

A general theoretical comparison of the three proposed estimators is complicated, owning to the iterative form of σ_t. However, given τ, the true parameter vector $\boldsymbol{\theta}_{\tau 0}$, and the density function $f_\varepsilon(\cdot)$, we can obtain theoretical values for b_τ, $f_\varepsilon(b_\tau)$, ω^*, and κ, and estimate all matrices in Ξ_i $(i = 1, 2, 3)$ using sample averages based on a large generated sequence. Then, we can compute the asymptotic relative efficiency (ARE) of $\widehat{\boldsymbol{\theta}}_{\tau n}$ to $\widetilde{\boldsymbol{\theta}}_{\tau n}$ and $\check{\boldsymbol{\theta}}_{\tau n}$, defined as $\mathrm{ARE}(\widehat{\boldsymbol{\theta}}_{\tau n}, \widetilde{\boldsymbol{\theta}}_{\tau n}) = (|\Xi_2|/|\Xi_1|)^{1/(p+q+d+1)}$ and $\mathrm{ARE}(\widehat{\boldsymbol{\theta}}_{\tau n}, \check{\boldsymbol{\theta}}_{\tau n}) = (|\Xi_3|/|\Xi_1|)^{1/(p+q+d+1)}$, respectively, where $|\cdot|$ is the determinant of a matrix; see Serfling (2009). The 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 $\widetilde{\boldsymbol{\theta}}_{\tau n}$. In contrast, the relative performance of $\widehat{\boldsymbol{\theta}}_{\tau n}$ versus that of the two-step estimator $\check{\boldsymbol{\theta}}_{\tau n}$ is mixed in terms of asymptotic efficiency; see 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 using $q_t(\widehat{\boldsymbol{\theta}}_{\tau n})$, $q_t(\widetilde{\boldsymbol{\theta}}_{\tau n})$, 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 the three

Statistica Sinica: Preprint
doi:10.5705/ss.202019.0003

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 $\{\boldsymbol{X}_{t-1}\}$ contains only lagged values of Y_t, that is, 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 the 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 nonnegative definite; that is, $\widehat{\boldsymbol{\theta}}_{\tau n}$ 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 using a bootstrapping procedure to directly approximate the asymptotic distributions of $\widehat{\boldsymbol{\theta}}_{\tau n}$, $\widetilde{\boldsymbol{\theta}}_{\tau n}$, and $\check{\boldsymbol{\theta}}_{\tau n}$.

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

Statistica Sinica: Preprint
doi:10.5705/ss.202019.0003

bootstrapping estimators, as follows:

$$\widehat{\boldsymbol{\theta}}_{\tau n}^{*} = (\widehat{b}_{\tau n}^{*}, \widehat{\boldsymbol{\lambda}}_{n}^{*\prime})' = \underset{b,\boldsymbol{\lambda}}{\operatorname{argmin}} \sum_{t=1}^{n} \omega_t \widehat{\sigma}_t^{-1} \rho_\tau \{Y_t - \boldsymbol{\phi}' \boldsymbol{X}_{t-1} - b\widetilde{\sigma}_t(\boldsymbol{\lambda})\} \tag{2.7}$$

and

$$\widetilde{\boldsymbol{\theta}}_{\tau n}^{*} = (\widetilde{b}_{\tau n}^{*}, \widetilde{\boldsymbol{\lambda}}_{n}^{*\prime})' = \underset{b,\boldsymbol{\lambda}}{\operatorname{argmin}} \sum_{t=1}^{n} \omega_t \rho_\tau \{Y_t - \boldsymbol{\phi}' \boldsymbol{X}_{t-1} - b\widetilde{\sigma}_t(\boldsymbol{\lambda})\}, \tag{2.8}$$

where $\{\omega_t\}$ are *i.i.d.* nonnegative random weights, with mean and variance both equal to one; see also Zheng et al. (2018) and **?**.

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 using $\check{\boldsymbol{\phi}}_n^{*} = \left(\sum_{t=1}^{n} \omega_t \boldsymbol{X}_{t-1} \boldsymbol{X}_{t-1}'\right)^{-1} \sum_{t=1}^{n} \omega_t \boldsymbol{X}_{t-1} Y_t$, and the bootstrapped residuals are computed using $\check{u}_t^{*} = u_t(\check{\boldsymbol{\phi}}_n^{*})$. Then, a randomly weighted quantile estimation is performed:

$$\check{\gamma}_{\tau n}^{*} = (\check{b}_{\tau n}^{*}, \check{\gamma}_{n}^{*\prime})' = \arg\min_{b,\boldsymbol{\gamma}} \sum_{t=1}^{n} \omega_t \rho_\tau \left\{\check{u}_t^{*} - b\check{\sigma}_t^{*}(\boldsymbol{\gamma})\right\}, \tag{2.9}$$

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

Assumption 5. *The random weights $\{\omega_t\}$ are i.i.d. nonnegative 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*

Statistica Sinica: Preprint
doi:10.5705/ss.202019.0003

(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, is 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 $\check{\boldsymbol{\theta}}_{\tau n}$ using 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, we can construct confidence intervals (CIs) for the estimators by substituting in the approximated asymptotic standard deviations (ASDs) calculated using the bootstrap method. Moreover, we can conduct hypothesis tests to detect the significance of the parameters by replacing the covariance matrices with their bootstrap approximations.

For the random weights, many distributions satisfy Assumption 5, including the standard exponential distribution and the Rademacher distribution, which takes the values zero or two with probability 0.5. According to the simulation findings in Zheng et al. (2018) and Zhu, Zeng, and Li (2020), the performance of the randomly weighted bootstrapping approximation is not sensitive to the choice of random weights. As a result, we simply use the random weights generated from the standard exponential distribution in the following sections.

3 Simulation studies

3.1 Finite-sample performance of the three proposed estimators

The first experiment evaluates 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 $\{\varepsilon_t\}$ is an *i.i.d.* standard normal or standardized Student's t_5 random variable 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. 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 each bootstrapped sample is $B = 500$, and the random weights $\{\omega_t\}$ are generated from the standard exponential distribution. Then, the ASDs can be calculated using the bootstrapping approximation. We can also construct CIs for each parameter based on the 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 using $\widehat{\theta}_{\tau n,j} \pm 1.96 \times \text{ASD}(\widehat{\theta}_{\tau n,j})$, where $\theta_{\tau 0,j}$ and $\widehat{\theta}_{\tau n,j}$ are the jth elements of $\boldsymbol{\theta}_{\tau 0}$ and $\widehat{\boldsymbol{\theta}}_{\tau n}$, respectively, and $\text{ASD}(\widehat{\theta}_{\tau n,j})$ is the ASD of $\widehat{\theta}_{\tau n,j}$. The CIs based on the other two estimation methods can be constructed in a similar way.

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 the 95% CIs. To save space, the simulation results for the setting with $n = 1000$ are provided in the Supplementary Material. It can be seen that, as the sample size n increases, the biases, ESDs, and ASDs decrease, and the ESDs and ASDs move closer to each other. In addition, as the innovations become more heavy-tailed, the standard deviations with respect to $\boldsymbol{\lambda}$ become larger, in general, whereas those related to b_τ become smaller. This is expected because the value of $|b_\tau|$ is smaller for t_5 distributed innovations than it is for normally distributed innovations. Moreover, as the quantile level τ increases from 0.05 to 0.10, the performance of the three estimators improves, in general, when $\{\varepsilon_t\}$ follows a t_5 distribution. However, when $\{\varepsilon_t\}$ follows a normal distribution, the performance with

Statistica Sinica: Preprint
doi:10.5705/ss.202019.0003

respect to b_τ and ϕ improves, but that with respect to $\boldsymbol{\gamma}$ worsens slightly as τ increases. This may be because, as τ gets closer to the center, more observations are available, but b_τ approaches zero. Finally, except for α, the ECRs of the 95% CIs for the other parameters are close to the nominal level of 0.95 in all settings. This may be because the true value of α is relative small, and including υ_{t-1} in the GARCH model may hinder an accurate estimation for α.

For the comparison between the three estimators, we have the following findings. First, as the sample size increases, the standard deviations of the jointly weighted estimator are smaller than those of the jointly unweighted estimator. This is expected because the efficiency gain from the weighting procedure becomes more evident when the sample size is larger. Second, the two-step method outperforms the jointly weighted method when estimating b_τ and ϕ, but performs a bit worse for the other parameters. Note that ϕ is estimated using the least squares method in the two-step estimation, but is estimated using 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 the result of the better performance for ϕ. Finally, in general, the accuracy of the CIs for the three estimators is comparable.

3.2 Theoretical comparison between the 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 $\check{\boldsymbol{\theta}}_{\tau n}$. 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 υ_{t-1}, we consider two cases: (1) $X_{t-1} = \upsilon_{t-1}$, where both are *i.i.d.* standard normal random variables; and (2) $X_{t-1} = Y_{t-1}$ and $\{\upsilon_{t-1}\}$ are *i.i.d.* stan-

Statistica Sinica: Preprint
doi:10.5705/ss.202019.0003

dard normal random variables. We consider different values for (α, β) and conduct the estimation at two quantile levels, $\tau = 0.05$ and 0.10. Table 4 shows the calculated AREs, $\text{ARE}(\widehat{\boldsymbol{\theta}}_{\tau n}, \widetilde{\boldsymbol{\theta}}_{\tau n})$ and $\text{ARE}(\widehat{\boldsymbol{\theta}}_{\tau n}, \breve{\boldsymbol{\theta}}_{\tau n})$.

The Table shows that $\text{ARE}(\widehat{\boldsymbol{\theta}}_{\tau n}, \widetilde{\boldsymbol{\theta}}_{\tau n}) > 1$ for all cases; that is, the jointly weighted estimator is asymptotically more efficient than the jointly unweighted estimator. However, the observations are mixed for the AREs of the estimator $\widehat{\boldsymbol{\theta}}_{\tau n}$ and the two-step estimator $\breve{\boldsymbol{\theta}}_{\tau n}$. First, the jointly weighted estimator becomes more efficient as the coefficient α or the quantile level τ increases. This is expected because larger α results in greater volatility, and more data become available as τ increases, leading to better performance by the weighting procedure. Moreover, the efficiency gain from the jointly weighted estimator is more evident when X_{t-1} is endogenous, although it becomes smaller as the innovations become more heavy-tailed. Based on these simulation findings, we focus on the jointly weighted and two-step estimating methods in the next section.

4 Empirical analysis

This section analyzes the daily log returns of the Occidental Petroleum security (NYSE:OXY). The data on daily closing prices, denoted as p_t, cover the period January 2, 2008, to December 29, 2017, with 2470 observations in total. A time plot of centered log returns (as percentages), that is, $Y_t = r_t - n^{-1}\sum_{t=1}^{n} r_t$ with $r_t = 100(\ln p_t - \ln p_{t-1})$, provides clear evidence of volatility clustering; see Figure 1. The summary statistics for $\{Y_t\}$ are provided in Table 5. The negative sample skewness and the kurtosis greater than three imply that the data are skewed and heavy-tailed.

The Occidental Petroleum Corporation is an oil and gas producer; therefore, its stock returns are likely to be affected by lagged values of Y_t and the oil price (Chernozhukov

Statistica Sinica: Preprint
doi:10.5705/ss.202019.0003

and Umantsev, 2001). Moreover, studies have indicated that gold can be a hedge against stock returns (Baur and McDermott, 2010; Iqbal, 2017). This study focuses on the effects of $\{Y_{t-1}\}$, the lagged crude oil returns and gold returns on $\{Y_t\}$. We use the WTI Crude Oil price and the Gold Fixing Price at 10:30 A.M. in the London Bullion Market as the price series. The data cover the period January 2, 2008, to December 29, 2017, and can be downloaded from the website of the Federal Reserve Economic Data (FRED, *https://fred.stlouisfed.org/*). Figure 2 shows time plots of their log returns as percentages, denoted as Oil_t and Gold_t, respectively. We first regress Y_t on the lagged returns Y_{t-1}, Oil_{t-1}, and Gold_{t-1}. The linear model is fitted using the least squares method, and the regression residuals are calculated using $\check{u}_t = Y_t - \check{\boldsymbol{\phi}}_n' \boldsymbol{X}_{t-1}$, where $\boldsymbol{X}_{t-1} = (Y_{t-1}, \mathrm{Oil}_{t-1}, \mathrm{Gold}_{t-1})'$ and $\check{\boldsymbol{\phi}}_n$ is the least squares estimate. The ACF and PACF plots of $\{\check{u}_t^2\}$ show strong ARCH effects, implying that a linear model with ARCH-type errors can be applied to $\{Y_t\}$. To further capture the possible influence of market volatility on $\{Y_t\}$, we include the realized kernel variance ($\times 100^2$) of the S&P 500 Index, denoted by $\{\upsilon_t^2\}$, as the covariate in the GARCH model. The realized variance series can be downloaded from the Oxford-Man Institute's realized library (*http://realized.oxford-man.ox.ac.uk/*); see the time plot of $\{\upsilon_t^2\}$ in Figure 2. Hansen and Lunde (2005) show that the GARCH$(1,1)$ model performs satisfactorily 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 $\{Y_t\}$. Because a 5% VaR is often of interest to practitioners, we focus on the conditional quantile of Y_t at $\tau = 0.05$, that is, the negative 5%

Statistica Sinica: Preprint
doi:10.5705/ss.202019.0003

VaR. We first apply the two-step method, and obtain the following conditional quantile:

$$\check{Q}_{Y_t}(0.05|\mathcal{F}_{t-1}) = -\,0.071_{0.044}Y_{t-1} - 0.029_{0.034}\text{Oil}_{t-1} + 0.074_{0.063}\text{Gold}_{t-1} - 0.766_{0.202}\check{\sigma}_t,$$

$$\check{\sigma}_t^2 = 1 + 0.203_{0.311}\check{u}_{t-1}^2 + 0.765_{0.100}\check{\sigma}_{t-1}^2 + 2.989_{0.710}\upsilon_{t-1}^2, \tag{4.1}$$

where the standard errors are shown as subscripts, and are calculated using the bootstrap method in Section 2.3. The estimates of b_τ and $\boldsymbol{\gamma}$ are significant at the 5% level, but the regressors in the linear model are nonsignificant. Based on the weights calculated from model (4.1), we employ the jointly weighted estimation, and obtain the following fitted model:

$$\widehat{Q}_{Y_t}(0.05|\mathcal{F}_{t-1}) = -\,0.050_{0.072}Y_{t-1} + 0.029_{0.051}\text{Oil}_{t-1} - 0.213_{0.070}\text{Gold}_{t-1} - 0.686_{0.163}\widehat{\sigma}_t,$$

$$\widehat{\sigma}_t^2 = 1 + 0.367_{0.144}\widehat{u}_{t-1}^2 + 0.766_{0.097}\widehat{\sigma}_{t-1}^2 + 3.455_{0.131}\upsilon_{t-1}^2, \tag{4.2}$$

where the standard errors in subscripts are computed using the bootstrapping procedure. Compared with model (4.1), the estimate of Gold_{t-1} in model (4.2) is significantly negative at the 5% level, implying that gold can be a safe haven in a bearish market (lower quantiles). Note that the regression coefficients of models (4.1) and (4.2) are quite different, in both magnitude and sign. This is expected because they are estimated using different methods. Furthermore, conditional heteroscedasticity is taken into account when performing the estimation in model (4.2), but is ignored in model (4.1). Moreover, because we include market volatility in the GARCH models, the estimates of the GARCH parameter β in models (4.1) and (4.2) are smaller than the usual estimates provided by GARCH models without exogenous variables. A similar finding was documented in Hwang and Satchell (2005), who concluded that the long persistency frequently found in volatility processes

Statistica Sinica: Preprint
doi:10.5705/ss.202019.0003

may be due to missing time-varying components. In addition, the estimate of the top Lyapunov component for model (4.2) is -0.194; 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 using a rolling procedure for the conditional quantile forecasts at $\tau = 0.01$ and $\tau = 0.05$, which are the negative 1% VaR and 5% VaR, respectively. A 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, that is, the forecast of $Q_{Y_{n+1}}(\tau|\mathcal{F}_n)$. The model estimates are updated by moving the window forward until we reach the end of the data set. Finally, we obtain 1469 one-day-ahead 1% (or 5%) VaRs. For illustration, the rolling forecasts at $\tau = 1\%$ and 5% for $\{Y_t\}$ are displayed in Figure 1; these forecasts are obtained using the jointly weighted approach. The magnitudes of the VaRs clearly increase as the volatility of the data increases. In addition, Y_t occasionally falls below its one-day negative 5% VaR, and, even more rarely, falls below its one-day negative 1% VaR.

To compare the forecasting performance of the proposed methods with that of the existing conditional quantile estimation, we perform the rolling forecasting procedure using the fully parametric (PAR) method, filtered historical simulation (FHS) method (Kuester, Mittnik, and Paolella, 2006), and conditional autoregressive VaR-method, called CAViaR (Engle and Manganelli, 2004). For the 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 using the maximum likelihood estimation, with the innovations $\{\varepsilon_t^*\}$ following a skewed Student's t distribution. Figure 3 gives the QQ plot of the residuals $\{\breve{\varepsilon}_t^*\}$ against the fitted skewed Student's t distribution, as well as their density plots. The figure shows that they

Statistica Sinica: Preprint
doi:10.5705/ss.202019.0003

are very close to each other; thus, we may argue that the PAR and FHS methods reach almost their maximum power. The $100\tau\%$ negative VaR for the PAR is computed using $\breve{\boldsymbol{\phi}}_n'\boldsymbol{X}_{t-1}+\breve{Q}_\tau\breve{\sigma}_t^*$, where $\breve{\boldsymbol{\phi}}_n$ and $\breve{\sigma}_t^* = \sigma_t^*(\breve{\boldsymbol{\lambda}}_n)$ are the maximum likelihood estimates, and $\breve{Q}_\tau$ is the τth quantile of the estimated skewed Student's t distribution. The $100\tau\%$ negative VaR for the FHS is calculated by replacing $\breve{Q}_\tau$ with 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 ECR, and perform VaR backtests for the VaR forecasts. Specifically, the ECR is calculated as the proportion of observations that fall below the corresponding conditional quantile forecast for the last 1469 data points. We use 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). Denote a hit by $H_t = I(Y_t < Q_{Y_t}(\tau|\mathcal{F}_{t-1}))$. 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 success probability τ. For the DQ test, following Engle and Manganelli (2004), we regress H_t on regressors that include a constant, four lagged hits H_{t-i}, for $i = 1, 2, 3, 4$, and the contemporaneous VaR forecast. The null hypothesis of the DQ test is that all regression coefficients are zero and the intercept is equal to the quantile level τ.

Table 6 reports the ECRs and p-values of two VaR backtests for the five estimating methods at the upper and lower 1% and 5% conditional quantiles. None of the five methods perform well at the lower 5% quantile, with p-values smaller than 0.05; however, they are adequate at the other three quantiles. In terms of the backtests, the JW and TS methods are comparable with the other three methods. With respect to the ECRs, we find that those of the JW and TS methods are closest to the nominal quantile level τ. We conclude that, overall, the proposed JW and TS methods outperform the three competitors in terms of forecasting VaRs for Occidental Petroleum returns. Moreover, our estimating methods, especially the JW method, use information at one quantile level only. Given that 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 further demonstrates the usefulness of two proposed methods.

5 Conclusion

This study examines a conditional quantile estimation for linear models with GARCH-X errors. As such, we propose three conditional quantile estimators, a jointly weighted estimator, a jointly unweighted estimator, and a two-step estimator. The root-n consistency and asymptotic normality are established for the three proposed estimators. Here, we use the bracketing method (Pollard, 1985) to overcome the theoretical difficulties due to the non-convex and non-differentiable objective functions. Simulation results indicate that, in general, the jointly weighted approach outperforms its unweighted counterpart when the sample size is large. Compared with the two-step estimating method, the jointly weighted method is preferred when the data exhibit greater volatility. This efficiency gain is especially evident when the linear regressors are endogenous and the quantile level is not too far from the center. Better VaR forecasting performance can be achieved using the proposed methods, as confirmed by our 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 is given by $Q_{Y_t}(\tau|\mathcal{F}_{t-1}) = \boldsymbol{\phi}'\boldsymbol{X}_{t-1} + b_\tau\sigma(\boldsymbol{X}_{t-1})$. As a result, an adaptive weighted conditional quantile estimation

(WCQE) can be constructed, as follows:

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

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

Supplementary Material

The online Supplementary Material provides proofs for all theorems and corollaries, together with additional simulation results for Section 3.1.

Acknowledgments

We are deeply grateful to the co-editor, associate editor and two anonymous referees for their valuable comments and suggestions. Zhu's research was supported by Shanghai Pujiang Program 2019PJC051, and the 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.

Statistica Sinica: Preprint
doi:10.5705/ss.202019.0003

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-parameter 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.

Statistica Sinica: Preprint
doi:10.5705/ss.202019.0003

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-

Statistica Sinica: Preprint
doi:10.5705/ss.202019.0003

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.

Statistica Sinica: Preprint
doi:10.5705/ss.202019.0003

Zhu, Q., R. Zeng, and G. Li (2020). Bootstrap inference for GARCH models by the least absolute deviation estimation. *Journal of Time Series Analysis 41*, 21–40.

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.

Statistica Sinica: Preprint
doi:10.5705/ss.202019.0003

Table 1: Biases, ESDs, ASDs, and ECRs of the 95% CIs for $\widehat{\boldsymbol{\theta}}_{\tau n}$ at $\tau = 0.05$ and 0.10, for normally distributed X_{t-1} and υ_{t-1}. The innovations follow a normal or a Student's t_5 distribution.

			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

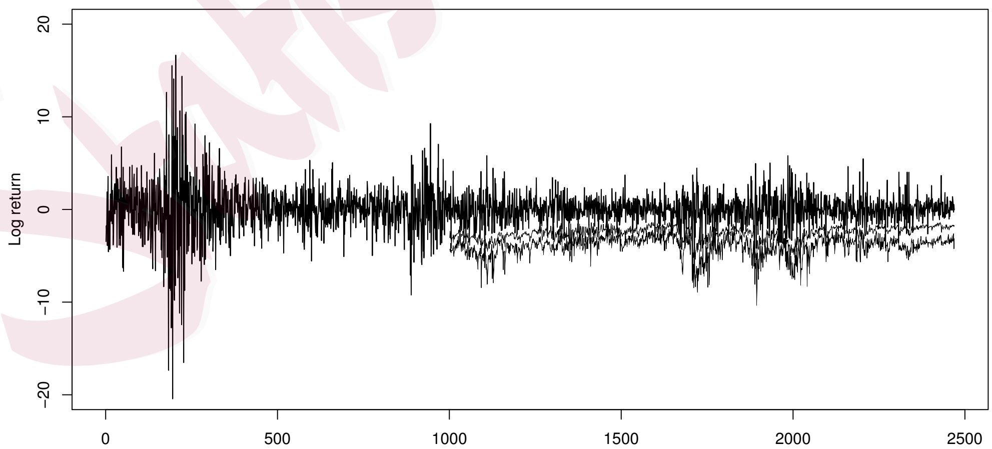

Figure 1: Time plot for centered daily log returns in percentage (solid line) of NYSE:OXY stock from January 3, 2008, to December 29, 2017, with one-day negative VaR forecasts at levels of 1% (dotted line) and 5% (dashed line) from February 3, 2012, to December 31, 2017.

Statistica Sinica: Preprint
doi:10.5705/ss.202019.0003

Table 2: Biases, ESDs, ASDs, and ECRs of the 95% CIs for $\widetilde{\boldsymbol{\theta}}_{\tau n}$ at $\tau = 0.05$ and 0.10, for normally distributed X_{t-1} and υ_{t-1}. The innovations follow a normal or a Student's t_5 distribution.

			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

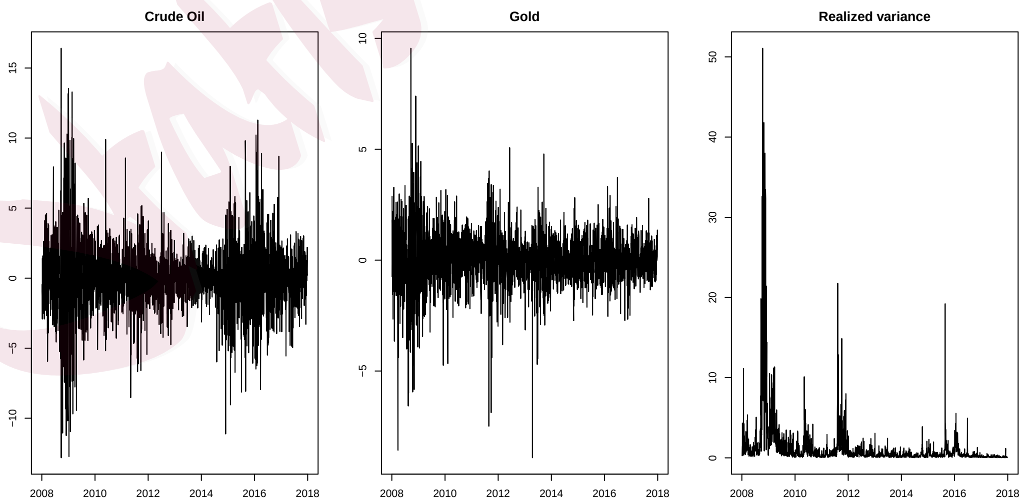

Figure 2: Time plots for daily log returns (percentage) of WTI Crude Oil prices (left) and LBMA Gold prices (middle), and the realized variance ($\times 100^2$) of the S&P 500 Index (right) from January 3, 2008, to December 29, 2017.

Statistica Sinica: Preprint
doi:10.5705/ss.202019.0003

Table 3: Biases, ESDs, ASDs, and ECRs of the 95% CIs for $\check{\boldsymbol{\theta}}_{\tau n}$ at $\tau = 0.05$ and 0.10, for normally distributed X_{t-1} and υ_{t-1}. The innovations follow a normal or a Student's t_5 distribution.

			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

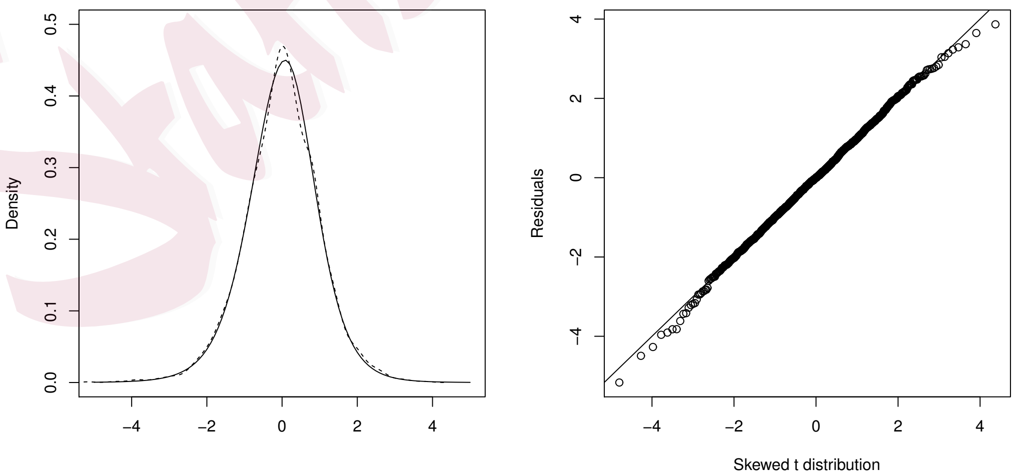

Figure 3: Density plots (left) of the model residuals $\breve{\varepsilon}_t$ (dashed line) and the fitted skewed t distribution (solid line), and a QQ plot (right) for the model residuals $\breve{\varepsilon}_t$ against the fitted skewed t distribution.

Statistica Sinica: Preprint
doi:10.5705/ss.202019.0003

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 the 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 υ_{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 υ_{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 (ECR) (%) 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, two-step method, parametric method, 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