Statistica Sinica Preprint No: SS-2024-0167	
Title	On Efficient Estimation for Value-at-Risk via Location-Scale Time Series Models
Manuscript ID	SS-2024-0167
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202024.0167
Complete List of Authors	Chaoxu Lei and Qianqian Zhu
Corresponding Authors	Qianqian Zhu
E-mails	zhu.qianqian@mail.shufe.edu.cn
Notice: Accepted author version.	

On efficient estimation for Value-at-Risk via location-scale time series models

Chaoxu Lei and Qianqian Zhu

School of Statistics and Management

Shanghai University of Finance and Economics

Abstract: This paper proposes efficient estimation methods for Value-at-Risk (VaR) in the framework of location-scale time series models, including the semi-parametric and parametric composite quantile regression (CQR). The semi-parametric CQR does not impose any distribution assumptions on the innovations, while the parametric CQR assumes that the innovations follow some distributions with explicit and parametric quantile functions. Compared with the quantile regression, the semi-parametric CQR method improves estimation efficiency by combining data information at multiple quantile levels. The parametric CQR takes advantage of model flexibility, and can further enhance efficiency in face of data scarcity when estimating high conditional quantiles. We establish the asymptotic properties of both CQR methods for location-scale time series models, and particularly for the ARMA-GARCH, double autoregressive and NAR-GARCH type models. We also compare both CQR estimators in estimation efficiency, and compare them with the Gaussian and exponential quasi-maximum likelihood estimators. Finally, we examine the finite-sample performance of the proposed methods via simulation studies, and analyze an empirical dataset to illustrate their usefulness in modeling and forecasting VaR for financial assets.

Key words and phrases: ARMA-GARCH models; Composite quantile regression; Double autoregressive models; Location-scale; Value-at-Risk.

1. Introduction

In risk management, value-at-risk (VaR) is a widely used indicator to measure market risk. It quantifies the maximum potential loss of an asset or portfolio over a specified period at a given confidence level. Various models have been proposed for VaR and they can be classified into three groups: non-parametric, parametric and semi-parametric methods. Non-parametric methods, such as historical simulation (HS) in Barone-Adessi et al. (1999) and the Boudoukh, Richardson, and Whitelaw (BRW) method in Boudoukh et al. (1998), neither require the assumption of return distribution nor parameterize conditional volatility. As a result, they are easy to implement but may yield less accurate results when faced with highly fluctuating data. On the other hand, parametric methods, including Risk Metrics in Morgan and Reuters (1996), not only require the modeling and parameterization of conditional volatility but also make distribution assumptions on the data or innovations. While these methods have good interpretability, their predictive ability heavily relies on the accuracy of model assumptions. On the spectrum of statistical approaches, semi-parametric methods fall in between non-parametric and parametric methods, by assuming a specific parametric model but imposing no distribution constraints on the data. Examples include filtered historical simulation (FHS) in Kuester et al. (2006) and conditional auto-regressive VaR-method (CAViaR) in Engle and Manganelli (2004). In a general sense, the FHS method models the overall data, while CAViaR focuses directly on modeling the VaR at a specific quantile level.

Another notable semi-parametric method is quantile regression (QR) in Koenker (2005), which fits the return series by parametric models and utilizes QR to estimate parameters at a target quantile level. Thus QR not only captures global characteristics but also provides local quantile estimations. In the literature, QR has been extensively studied in a range of volatility models, such as ARCH (Koenker and Zhao, 1996), GARCH (Xiao and Koenker, 2009; Lee and Noh, 2013; Zheng et al., 2018), asymmetric power GARCH (Wang et al., 2022), and quantile GARCH (Zhu et al., 2023) models. Meanwhile, numerous studies have investigated QR for location-scale time series models, including ARMA-asymmetric GARCH models (Noh and Lee, 2016), linear models with GARCH-X errors (Zhu et al., 2021), linear and quantile double autoregressive models (Zhu et al., 2018; Zhu and Li, 2022). Although QR combines the advantages of the FHS and CAViaR methods, it may lack efficiency at a specific quantile level. To address this issue, we consider the composite quantile regression (CQR) (Zou and Yuan, 2008) in the framework of location-scale time series models. CQR combines conditional quantile information at multiple quantile levels, consequently it can improve estimation efficiency and maintain the advantages of QR. While CQR has been explored for volatility models, such as GARCH (Wang et al., 2018), quantile GARCH (Zhu et al., 2023) and double-threshold ARCH (Jiang et al., 2014) models, there is very limited research on CQR for general location-scale time series models. To fill this research gap, this paper investigates CQR for general location-scale time series models, and establishes the corresponding global estimator with theoretical guarantee.

Location-scale time series models can capture the dependence between the current and past observations using conditional mean and variance structures. The well-known location-scale time series models include the classes of ARMA-GARCH, double autoregressive (DAR) models and nonlinear AR models with ARCH or GARCH errors (NAR-GARCH). These models flexibly describe the autocorrelation and volatility clustering simultaneously for time series. ARMA-GARCH type models are more parsimonious than DAR type models when fitting time series, while DAR type models usually possess larger parameter spaces and more relaxed moment conditions when establishing asymptotic properties than the former. NAR-GARCH models include the threshold AR-GARCH (Liu et al., 1997; Ling, 1999) and smooth transition AR-GARCH (STAR-GARCH, Chan and McAleer (2002); Chan et al. (2015)) models as special cases. In comparison with ARMA-GARCH and DAR type models, the NAR-GARCH model is able to capture nonlinear conditional mean structures. While there are a wide range of location-scale time series models, we focus specifically on ARMA-GARCH, DAR and NAR-GARCH type models owing to their popularity. The literature have studied many ARMA-GARCH type models, including ARMA-GARCH (Francq and Zakoïan, 2004), ARMA-IGARCH (Zhu and Ling, 2011), ARMA-asymmetric GARCH (Noh and Lee, 2016) models and so on. Recently, various DAR type models have been proposed and paid growing attention, such as the DAR (Ling, 2007), threshold DAR (Li et al., 2015, 2016), mixture DAR (Li et al., 2017), linear DAR (Zhu et al., 2018), augmented DAR (Jiang et al., 2020) and asymmetric linear DAR (Tan and Zhu, 2022) models. For NAR-GARCH models, most research focuses on the probability properties (Cline and Pu, 2004; Cline, 2007) and empirical applications, with little study related to the asymptotic theory for estimation. The aforementioned literature have investigated the parameter estimation for these models, mainly focusing on quasi-maximum likelihood estimation (QMLE) (Francq and Zakoïan, 2004; Ling, 2007; Zhu and Ling, 2011; Chan et al., 2015; Tan and Zhu, 2022) and QR method (Wang and Zhao, 2016; Noh and Lee, 2016; Zhu et al., 2018; Zhu and Li, 2022). However, as far as we know, few studies have considered the CQR estimation for location-scale time series models, even less for ARMA-GARCH, DAR and NAR-GARCH type models. This motivates us to study the CQR for these models as special cases of location-scale models.

In this paper, we investigate two CQR methods for location-scale time series models. We refer to the first approach as semi-parametric CQR, which considers a location-scale model with no distribution assumptions imposed on the innovation for time series. Compared with the QR using information at one quantile level, this method can improve estimation efficiency by combining information from multiple quantile levels. It is well-known that conditional quantiles at high levels are of particular interest in financial risk management. When estimating and forecasting these high conditional quantiles, the semi-parametric CQR lacks flexibility and can be less efficient in face of data scarcity. To address this drawback, we propose the parametric CQR as the second method. In contrast with the first method, the second approach fits time series using a location-scale model but assumes a parametric distribution with an explicit quantile function for the innovation. The fully parametric setting makes this method flexible to do extrapolation for conditional quantiles at high levels. As a result, the parametric CQR can further improve estimation efficiency in estimating high conditional quantiles. There are several distributions in the literature that have explicit quantile functions, such as the Lambda, generalized extreme value, generalized Pareto, and Burr XII distributions (Gilchrist, 2000). For illustration, we choose the Tukey-lambda distribution for the innovation as it can approximate many distributions, including Gaussian, Logistic, Pareto-type, and extreme value distributions. Other distributions mentioned above can also be used in conjunction with CQR in a similar manner.

Our paper has three main contributions. Firstly, we introduce semi-parametric and parametric CQRs for location-scale time series models, which enrich the estimation methods for these models. Compared with the QR method, both CQRs can preserve its advantages in capturing both global and local characteristics of time series, whereas improve its estimation efficiency. Secondly, we establish the consistency and asymptotic normality for the aforementioned CQR estimators. Note that the objective functions for both CQR estimators are nondifferentiable and nonconvex. In order to prove the $\sqrt{n}$-convergence rate and asymptotic normality of both estimators, we employ the bracketing method (Pollard, 1985) to address the resulting challenges; see also Zhu and Ling (2011) and Zhu et al. (2023). We also compare both CQR estimators in estimation efficiency, and compare them with the Gaussian and exponential QMLEs. Owing to the popularity of ARMA-GARCH, DAR and NAR-GARCH type models, we further investigate both CQRs for these models and derive the conditions to ensure their asymptotic properties. Simulation studies indicate that the proposed methods perform well especially for heavy-tailed data, not only for parameter estimation but also for conditional quantile prediction at moderate and high quantile levels. Finally, the proposed CQRs provide useful tools for VaR prediction of financial time series. Compared with the FHS method (i.e. a semi-parametric method), the semi-parametric CQR can directly obtain local quantile estimations of innovations. Meanwhile, the parametric CQR is more flexible in comparison with the Risk Metrics method (i.e. a parametric method) which assumes the IGARCH model with standard normal innovations for time series. The real application in Section 4 demonstrates that the proposed CQRs outperform competitive methods in VaR prediction at most of the moderate and high quantile levels. Therefore, the proposed CQRs can provide more accurate VaR prediction in practice.

The remainder of this paper is organized as follows. Section 2 introduces the location-scale time series models including some popular examples, and presents the semi-parametric and parametric CQRs for these models with asymptotic properties established. Moreover, the proposed CQR estimators are compared with the Gaussian and exponential QMLEs via the asymptotic relative efficiency. In Section 3, we apply both CQRs to ARMA-GARCH, DAR and NAR-GARCH type models, and verify the conditions for their asymptotic properties. Simulation experiments are conducted in Section 4 to assess the finite-sample performance of the proposed CQRs, and compare them with the Gaussian and exponential QMLEs in estimating and forecasting conditional quantiles. Section 5 illustrates the proposed methods by analyzing the daily return of Microsoft Corp's stock. Conclusion and discussion are given in Section 6. All technical details are relegated to the Supplementary Material. Throughout the paper, $\to_p$ and $\to_{\mathcal{L}}$ denote the convergences in probability and distribution, respectively, and $o_p(1)$ denotes a sequence of random variables converging to zero in probability. We denote by $\|\cdot\|$ the norm of a matrix or column vector, defined as $\|A\| = \sqrt{tr(AA')} = \sqrt{\sum_{i,j} a_{ij}^2}$.

2. The methodology

In this section, we propose the semi-parametric and parametric CQR estimation methods with theoretical guarantee for location-scale time series models.

Consider the location-scale time series model as follows

$$y_t = \mu_t(\boldsymbol{\vartheta}) + \eta_t h_t(\boldsymbol{\vartheta}), \tag{2.1}$$

where $t \in \mathbb{Z}$, $\boldsymbol{\vartheta}$ is the parameter vector which belongs to the parameter space Θ, $\mu_t(\boldsymbol{\vartheta}) := \mu_t(y_{t-1}, y_{t-2}, \ldots; \boldsymbol{\vartheta})$ and $h_t(\boldsymbol{\vartheta}) := h_t(y_{t-1}, y_{t-2}, \ldots; \boldsymbol{\vartheta}) > 0$ are measurable functions for the conditional location and scale, respectively, and $\{\eta_t\}$ are the $i.i.d$ innovations with zero mean. Here $\mu_t(\boldsymbol{\vartheta})$ and $h_t(\boldsymbol{\vartheta})$ can be linear or nonlinear functions with respect to $\boldsymbol{\vartheta}$, thus model (2.1) includes many time series models as special cases; see Examples 1-3 for details.

Example 1 (ARMA-GARCH type models)**.** *ARMA-GARCH type models include ARMA-GARCH (Francq and Zakoïan, 2004), ARMA-IGARCH (Zhu and Ling, 2011) and ARMA-asymmetric GARCH (Noh and Lee, 2016) models as special cases. For illustration, consider* *the ARMA(p,q)-GARCH(P,Q) model defined as follows*

$$y_t = \sum_{i=1}^{p} \alpha_i y_{t-i} + \sum_{j=1}^{q} \beta_j \epsilon_{t-j} + \epsilon_t,\ \epsilon_t = \eta_t h_t,\ h_t^2 = \omega + \sum_{i=1}^{Q} \gamma_i \epsilon_{t-i}^2 + \sum_{j=1}^{P} \nu_j h_{t-j}^2, \tag{2.2}$$

where $\alpha_i, \beta_j \in \mathbb{R}$ for $i = 1, \ldots, p$ and $j = 1, \ldots, q$, $\omega > 0$, $\gamma_i \geq 0$ for $i = 1, \ldots, Q$, $\nu_j \geq 0$ for $j = 1, \ldots, P$, and $\{\eta_t\}$ are i.i.d innovations with zero means. Model (2.2) can be rewritten into model (2.1) with parameter $\boldsymbol{\vartheta}^I = (\alpha_1, \ldots, \alpha_p, \beta_1, \ldots, \beta_q, \omega, \gamma_1, \ldots, \gamma_Q, \nu_1, \ldots, \nu_P)'$, where

$$\mu_t(\boldsymbol{\vartheta}^I) = \sum_{i=1}^{p} \alpha_i y_{t-i} + \sum_{j=1}^{q} \beta_j \epsilon_{t-j} \ \ \textit{and}\ \ h_t(\boldsymbol{\vartheta}^I) = \left(\omega + \sum_{i=1}^{Q} \gamma_i \epsilon_{t-i}^2 + \sum_{j=1}^{P} \nu_j h_{t-j}^2 \right)^{\frac{1}{2}}.$$

Example 2 (DAR type models)**.** *The DAR model (Ling, 2007) has many extensions, such as the linear DAR (Zhu et al., 2018), augmented DAR (Jiang et al., 2020), asymmetric linear DAR (Tan and Zhu, 2022) and dual-asymmetric linear DAR (Tan and Zhu, 2023) models. For illustration, consider the asymmetric linear DAR (ALDAR) model of order (p,q):*

$$y_t = \sum_{i=1}^{p} \varphi_i y_{t-i} + \eta_t \left[\omega + \sum_{j=1}^{q} (\alpha_j^+ y_{t-j}^+ - \alpha_j^- y_{t-j}^-) \right], \tag{2.3}$$

where $\varphi_i \in \mathbb{R}$ for $1 \leq i \leq p$, $\omega > 0$, $\alpha_j^+, \alpha_j^- \geq 0$ for $1 \leq j \leq q$, $y_t^+ = \max\{0, y_t\}$ and $y_t^- = \min\{0, y_t\}$ are positive and negative parts of y_t, respectively, and $\{\eta_t\}$ are i.i.d innovations with zero means. Corresponding to the location-scale model (2.1), denote $\mu_t(\boldsymbol{\vartheta}^{II}) = \sum_{i=1}^{p} \varphi_i y_{t-i}$ and $h_t(\boldsymbol{\vartheta}^{II}) = \omega + \sum_{j=1}^{q} (\alpha_j^+ y_{t-j}^+ - \alpha_j^- y_{t-j}^-)$, where $\boldsymbol{\vartheta}^{II} = (\varphi_1, \varphi_2, \ldots, \varphi_p, \omega, \alpha_1^+, \alpha_2^+, \ldots, \alpha_q^+, \alpha_1^-, \alpha_2^-, \ldots, \alpha_q^-)'$ is the parameter vector of model (2.3).

Example 3 (NAR-GARCH type models)**.** *The general NAR-GARCH models include many smooth transition AR processes with GARCH errors (STAR-GARCH) as special cases, such as the multiple regime logistic or exponential STAR-GARCH model; see Section 2 in Chan et al. (2015) for more details. For illustration, consider the 2-regime exponential STAR(p)-*

GARCH$(1,1)$ *(ESTAR-GARCH) model as follows:*

$$y_t = \alpha_{00} + \alpha_{10} G(y_{t-d}; \gamma, c) + \sum_{i=1}^{p} \left[\alpha_{0i} + \alpha_{1i} G(y_{t-d}; \gamma, c)\right] y_{t-i} + \epsilon_t,$$

$$\epsilon_t = \eta_t h_t, \; h_t^2 = \omega + a\epsilon_{t-1}^2 + b h_{t-1}^2, \tag{2.4}$$

where y_{t-d} is the transition variable, $d \in \mathcal{N}$ is a delay parameter which usually takes $d = 1$ in practice, $G(y_{t-d}; \gamma, c) = 1 - \exp(-(y_{t-d} - c)^2/\gamma)$ is the smooth transition function with c and $\gamma > 0$ being the location and scale parameters of model transition, $\omega > 0, a \geq 0, b \geq 0$, and $\{\eta_t\}$ are i.i.d innovations with zero means; see also Tsay (2010). Denote by $\boldsymbol{\vartheta}^{III} = (\alpha_{00}, \alpha_{01}, \ldots, \alpha_{0p}, \alpha_{10}, \alpha_{11}, \ldots, \alpha_{1p}, \gamma, c, \omega, a, b)'$ the parameter vector, then model (2.4) *can be rewritten into model* (2.1) *with $\mu_t(\boldsymbol{\vartheta}^{III}) = \alpha_{00} + \alpha_{10} G(y_{t-d}; \gamma, c) + \sum_{i=1}^{p} \left[\alpha_{0i} + \alpha_{1i} G(y_{t-d}; \gamma, c)\right] y_{t-i}$ and $h_t(\boldsymbol{\vartheta}^{III}) = (\omega + a\epsilon_{t-1}^2 + b h_{t-1}^2)^{1/2}$.*

Let $\mathcal{F}_t$ be the σ-field generated by $\{y_s, s \leq t\}$. Note that $\mu_t(\boldsymbol{\vartheta})$ and $h_t(\boldsymbol{\vartheta})$ are $\mathcal{F}_{t-1}$-measurable. Then the τth conditional quantile of y_t in (2.1) can be written as

$$Q_\tau(y_t|\mathcal{F}_{t-1}) = \mu_t(\boldsymbol{\vartheta}) + b_\tau h_t(\boldsymbol{\vartheta}), \tag{2.5}$$

where b_τ is the τth quantile of η_t for any $\tau \in (0,1)$. To estimate the VaR at level τ and time t, i.e. the negative $Q_\tau(y_t|\mathcal{F}_{t-1})$, below we introduce the semi-parametric and parametric CQR estimation methods for model (2.1), respectively.

2.1 The semi-parametric CQR

We first consider the semi-parametric CQR for model (2.1), where no distribution assumption is imposed on η_t such that b_τ in (2.5) is treated as an unknown parameter.

To improve the estimation efficiency, we consider the CQR method to combine the information at fixed multiple quantile levels $0 < \tau_1 < \tau_2 < \ldots < \tau_K < 1$ with K being a predetermined integer. Denote the conditional quantile function at τ_k as $q_{t,\tau_k}(\boldsymbol{\phi}) = \mu_t(\boldsymbol{\vartheta}) + b_k h_t(\boldsymbol{\vartheta})$, where $\boldsymbol{\phi} = (\boldsymbol{\vartheta}', b_1, b_2, \ldots, b_K)'$. Note that $\mu_t(\boldsymbol{\vartheta})$ and $h_t(\boldsymbol{\vartheta})$ probably depend on the information in the infinite past, and then initial values of $\mathcal{F}_s$ for $s \leq 0$ are needed to obtain a feasible conditional quantile function in practice. For example, $\mu_t(\boldsymbol{\vartheta})$ and $h_t(\boldsymbol{\vartheta})$ in the ARMA-GARCH model (2.2) are defined recursively on y_{t-i}'s, ϵ_{t-j}'s and h_{t-i}'s such that information in the infinite past is involved, and thus initial values of y_s, ϵ_s and h_s for $s \leq 0$, are required; see Section 3.1 for details. Replacing $\mathcal{F}_s$ for $s \leq 0$ in $\mu_t(\boldsymbol{\vartheta})$ and $h_t(\boldsymbol{\vartheta})$ with the given initial values, we can obtain the feasible counterpart of $q_{t,\tau_k}(\boldsymbol{\phi})$, denoted as $\widetilde{q}_{t,\tau_k}(\boldsymbol{\phi})$. As a result, the semi-parametric CQR estimator is defined as follows:

$$\widehat{\boldsymbol{\phi}}_n = (\widehat{\boldsymbol{\vartheta}}_n', \widehat{b}_1, \ldots, \widehat{b}_K)' = \arg\min_{\boldsymbol{\phi}\in\Phi} \sum_{k=1}^{K} \sum_{t=1}^{n} \rho_{\tau_k}\left(y_t - \widetilde{q}_{t,\tau_k}(\boldsymbol{\phi})\right), \tag{2.6}$$

where $\rho_\tau(y) = y\{\tau - I\,(y < 0)\}$ is the check function, $\Phi = \Theta \times \mathbb{R}^K$ represents the parameter space of $\boldsymbol{\phi}$, and the selection of set $\{\tau_k\}_{k=1}^K$ is discussed in Section 3.2. We will prove that the effect of initial values on the estimation is asymptotically negligible under some regular conditions; see Assumption 6 for details. Based on the semi-parametric CQR estimator $\widehat{\boldsymbol{\phi}}_n$, then the τth conditional quantile of y_t in (2.5) can be estimated by $\widehat{Q}_\tau(y_t|\mathcal{F}_{t-1}) = \widetilde{q}_{t,\tau}(\widehat{\boldsymbol{\phi}}_n)$.

For the semi-parametric CQR, denote the true parameter vector by $\boldsymbol{\phi}_0 = (\boldsymbol{\vartheta}_0', b_{10}, \ldots, b_{K0})'$, where b_{k0} is the τ_kth quantile of η_t. Let $\dot{q}_{t,\tau}(\boldsymbol{\phi})$ and $\ddot{q}_{t,\tau}(\boldsymbol{\phi})$ (or $\dot{\widetilde{q}}_{t,\tau}(\boldsymbol{\phi})$ and $\ddot{\widetilde{q}}_{t,\tau}(\boldsymbol{\phi})$) be the first and second derivatives of $q_{t,\tau}(\boldsymbol{\phi})$ (or $\widetilde{q}_{t,\tau}(\boldsymbol{\phi})$) with respect to $\boldsymbol{\phi}$, respectively. Define $f(\cdot)$ and $F(\cdot)$ as the density and distribution functions of η_t, respectively. To establish the asymptotic properties of the semi-parametric CQR estimator $\widehat{\boldsymbol{\phi}}_n$, we need the following assumptions.

Assumption 1 (Process)**.** *$\{y_t : t = 1, 2, \ldots\}$ is strictly stationary and ergodic.*

Assumption 2 (Parameter space)**.** *(i) The parameter space Φ is compact; (ii) the true parameter $\boldsymbol{\phi}_0$ is an interior point in Φ.*

Assumption 3 (Innovation). *With probability one, $f(\cdot)$ and its derivative function $\dot{f}(\cdot)$ are uniformly bounded, and $f(\cdot) > 0$ holds on the support $\{x : 0 \leq F(x) \leq 1\}$.*

Assumption 4 (Identification). *(i) $q_{t,\tau}(\boldsymbol{\phi})$ is continuous in $\boldsymbol{\phi} \in \Phi$; (ii) if $q_{t,\tau}(\boldsymbol{\phi}) = q_{t,\tau}(\boldsymbol{\phi}_0)$, then $\boldsymbol{\phi} = \boldsymbol{\phi}_0$.*

Assumption 5 (Moments). *(i) $E\left[\sup_{\boldsymbol{\phi}\in\Phi} |q_{t,\tau}(\boldsymbol{\phi})|\right] < \infty$; (ii) $E[h_t^{-1}(\boldsymbol{\vartheta}_0)\sup_{\boldsymbol{\phi}\in\Phi} \|\dot{q}_{t,\tau}(\boldsymbol{\phi})\|^3] < \infty$; (iii) $E\left[h_t^{-1}(\boldsymbol{\vartheta}_0)\sup_{\boldsymbol{\phi}\in\Phi} \|\ddot{q}_{t,\tau}(\boldsymbol{\phi})\|^2\right] < \infty$.*

Assumption 6 (Initial values). *(i) $\sum_{t=1}^{\infty} h_t^{-1}(\boldsymbol{\vartheta}_0)\sup_{\boldsymbol{\phi}\in\Phi} |q_{t,\tau}(\boldsymbol{\phi}) - \widetilde{q}_{t,\tau}(\boldsymbol{\phi})|^2 < \infty$; (ii) $\sum_{t=1}^{\infty} h_t^{-1}(\boldsymbol{\vartheta}_0)\sup_{\boldsymbol{\phi}\in\Phi} \|\dot{q}_{t,\tau}(\boldsymbol{\phi}) - \dot{\widetilde{q}}_{t,\tau}(\boldsymbol{\phi})\|^2 < \infty$.*

Assumption 1 is general and can be verified by checking the stationary region of parameters for time series models. Assumption 2 imposes conditions on the parameter space, where (i) is used for the consistency of the CQR estimator while (ii) is required for the asymptotic normality; see also Francq and Zakoïan (2004), Noh and Lee (2016) and Zhu et al. (2018). Assumption 3 provides conditions on the innovation, which is commonly used to prove the asymptotic normality of quantile regression estimators; see Assumption (A2) in Lee and Noh (2013). Assumption 4 guarantees the uniqueness of the true parameter and is necessary for model identification. Assumption 5 gives moment conditions to derive the asymptotic properties. Specifically, Assumption 5 (i) is required for the consistency of the CQR estimator whereas (ii) and (iii) are for the asymptotic normality. These moment conditions on $\dot{q}_{t,\tau}(\boldsymbol{\phi})$ and $\ddot{q}_{t,\tau}(\boldsymbol{\phi})$ exclude the threshold models with unknown discontinuous thresholds, and thus our theory is inapplicable to these threshold models. Assumption 6 ensures that the initial values have no influence on the asymptotic properties of the CQR estimator.

Denote $\Gamma_{kk'} = \min(\tau_k, \tau_{k'})[1 - \max(\tau_k, \tau_{k'})]$, and define the following matrices for the semi-parametric CQR:

$$\Omega = \sum_{k=1}^{K}\sum_{k'=1}^{K}\Gamma_{kk'}E\left[\dot{q}_{t,\tau_k}(\boldsymbol{\phi}_0)\dot{q}'_{t,\tau_{k'}}(\boldsymbol{\phi}_0)\right] \text{ and } \Sigma = \sum_{k=1}^{K} f(b_{k0})E\left[\frac{1}{h_t(\boldsymbol{\vartheta}_0)}\dot{q}_{t,\tau_k}(\boldsymbol{\phi}_0)\dot{q}'_{t,\tau_k}(\boldsymbol{\phi}_0)\right].$$

Theorem 1. *For $\{y_t\}$ generated by model (2.1), if Assumptions 1-6 hold, then we have $\widehat{\boldsymbol{\phi}}_n \to_p \boldsymbol{\phi}_0$ as $n \to \infty$.*

Theorem 2. *Suppose $\Xi = \Sigma^{-1}\Omega\Sigma^{-1}$ is positive definite. For $\{y_t\}$ generated by model (2.1), if Assumptions 1-6 hold, then we have $\sqrt{n}(\widehat{\boldsymbol{\phi}}_n - \boldsymbol{\phi}_0) \to_{\mathcal{L}} N(0, \Xi)$ as $n \to \infty$.*

Theorems 1 and 2 establish the consistency and asymptotic normality of the semi-parametric CQR estimator $\widehat{\boldsymbol{\phi}}_n$. Note that the objective function in (2.6) is nondifferentiable and nonconvex. To prove the $\sqrt{n}$-convergence rate and asymptotic normality of $\widehat{\boldsymbol{\phi}}_n$, we employ the bracketing method (Pollard, 1985) to tackle the resulting challenges.

To estimate the asymptotic covariance of $\widehat{\boldsymbol{\phi}}_n$, we first estimate $f(b_{k0})$ in Σ using the difference quotient method in Koenker (2005). Particularly, we employ the estimator $\widehat{f}(b_{k0}) = 2h/[Q_{\tau_k+h}(\widehat{\eta}_t) - Q_{\tau_k-h}(\widehat{\eta}_t)]$, where $\widehat{\eta}_t = [y_t - \mu_t(\widehat{\boldsymbol{\vartheta}}_n)]/h_t(\widehat{\boldsymbol{\vartheta}}_n)$ is the residual, $Q_\tau(\widehat{\eta}_t)$ is the τth sample quantile of $\{\widehat{\eta}_t\}$, and h is the bandwidth. As in Koenker and Xiao (2006), we consider two commonly used bandwidths for h as follows:

$$h_B = n^{-1/5}\left\{\frac{4.5\phi^4(Q_\Phi(\tau_k))}{[2Q_\Phi^2(\tau_k)+1]^2}\right\}^{1/5} \text{ and } h_{HS} = n^{-1/3}z_\alpha^{2/3}\left\{\frac{1.5\phi^2(Q_\Phi(\tau_k))}{2Q_\Phi^2(\tau_k)+1}\right\}^{1/3}, \tag{2.7}$$

where $\phi(\cdot)$ and $Q_\Phi(\cdot)$ are the standard normal density and quantile functions, respectively, and $z_\alpha = Q_\Phi(1-\alpha/2)$ with $\alpha = 0.05$. Then we can approximate the matrices Σ and Ω using sample averages with $\boldsymbol{\phi}_0$ replaced by $\widehat{\boldsymbol{\phi}}_n$ and $\dot{q}_{t,\tau_k}(\cdot)$ replaced by $\dot{\widetilde{q}}_{t,\tau_k}(\cdot)$. Finally, we can estimate Ξ by $\widehat{\Xi} = \widehat{\Sigma}^{-1}\widehat{\Omega}\widehat{\Sigma}^{-1}$ with

$$\widehat{\Omega} = \frac{1}{n}\sum_{t=1}^{n}\sum_{k=1}^{K}\sum_{k'=1}^{K}\Gamma_{kk'}\dot{\widetilde{q}}_{t,\tau_k}(\widehat{\boldsymbol{\phi}}_n)\dot{\widetilde{q}}'_{t,\tau_{k'}}(\widehat{\boldsymbol{\phi}}_n) \text{ and } \widehat{\Sigma} = \frac{1}{n}\sum_{t=1}^{n}\sum_{k=1}^{K}\frac{\widehat{f}(b_{k0})}{\widetilde{h}_t(\widehat{\boldsymbol{\vartheta}}_n)}\dot{\widetilde{q}}_{t,\tau_k}(\widehat{\boldsymbol{\phi}}_n)\dot{\widetilde{q}}'_{t,\tau_k}(\widehat{\boldsymbol{\phi}}_n),$$

where $\widetilde{h}_t(\widehat{\boldsymbol{\vartheta}}_n)$ is the feasible version of $h_t(\widehat{\boldsymbol{\vartheta}}_n)$ given initial values.

2.2 The parametric CQR

In risk management, the focus is primarily on the tail risk of assets. Thus it is crucial to efficiently estimate the tail quantiles of the time series. This subsection proposes a parametric CQR which can take advantage of model flexibility, and further enhance efficiency in face of data scarcity when estimating high conditional quantiles. This method assumes an appropriate parametric distribution for η_t with an explicit quantile function, such that b_τ in (2.5) depends on τ and some unknown parameters explicitly. This makes it convenient to estimate high conditional quantiles by extrapolating the quantile structure to high levels.

There are various parameter distributions having quantile functions with explicit forms, such as the lambda, generalized extreme value, generalized Pareto, and Burr XII distributions (Gilchrist, 2000). Let $\boldsymbol{\lambda}$ be the parameter vector of quantile function for the innovation η_t, and then the τth quantile of η_t can be denoted by $Q_\tau(\boldsymbol{\lambda})$. Therefore, the τth conditional quantile function for $\{y_t\}$ generated by model (2.1) has the form of

$$g_{t,\tau}(\boldsymbol{\psi}) = \mu_t(\boldsymbol{\vartheta}) + Q_\tau(\boldsymbol{\lambda})h_t(\boldsymbol{\vartheta}), \tag{2.8}$$

where $\boldsymbol{\psi} = (\boldsymbol{\vartheta}', \boldsymbol{\lambda}')'$. As for $q_{t,\tau}(\boldsymbol{\phi})$, $g_{t,\tau}(\boldsymbol{\psi})$ may depend on unobservable information in the infinite past, thus we use the feasible conditional quantile function $\widetilde{g}_{t,\tau}(\boldsymbol{\psi})$ given initial values of $\mathcal{F}_s$ for $s \leq 0$ to approximate $g_{t,\tau}(\boldsymbol{\psi})$. Then the parametric CQR estimator is defined as

$$\widehat{\boldsymbol{\psi}}_n = (\widetilde{\boldsymbol{\vartheta}}_n', \widehat{\boldsymbol{\lambda}}_n')' = \arg\min_{\boldsymbol{\psi}\in\Psi} \sum_{k=1}^{K}\sum_{t=1}^{n} \rho_{\tau_k}\left(y_t - \widetilde{g}_{t,\tau_k}(\boldsymbol{\psi})\right), \tag{2.9}$$

where $\Psi = \Theta \times \Lambda$ is the parameter space of $\boldsymbol{\psi}$ with Λ being the parameter space of $\boldsymbol{\lambda}$. Based on the parametric CQR estimator $\widehat{\boldsymbol{\psi}}_n$, the τth conditional quantile of y_t in (2.5) can be estimated by $\widehat{Q}_\tau(y_t|\mathcal{F}_{t-1}) = \widetilde{g}_{t,\tau}(\widehat{\boldsymbol{\psi}}_n)$. Note that the target level τ is not necessarily in the set of $\{\tau_k\}_{k=1}^K$, because the explicit function $Q_\tau(\boldsymbol{\lambda})$ facilitates us to do extrapolation for any high level τ.

For the parametric CQR, denote its true parameter vector by $\boldsymbol{\psi}_0 = (\boldsymbol{\varphi}_0', \boldsymbol{\lambda}_0')'$. Let $\dot{g}_{t,\tau}(\boldsymbol{\psi})$ and $\ddot{g}_{t,\tau}(\boldsymbol{\psi})$ (or $\dot{\widetilde{g}}_{t,\tau}(\boldsymbol{\psi})$ and $\ddot{\widetilde{g}}_{t,\tau}(\boldsymbol{\psi})$) be the first and second derivatives of $g_{t,\tau}(\boldsymbol{\psi})$ (or $\widetilde{g}_{t,\tau}(\boldsymbol{\psi})$) with respect to $\boldsymbol{\psi}$, respectively. Define the matrices for the parametric CQR as follows:

$$M = \sum_{k=1}^{K}\sum_{k'=1}^{K} \Gamma_{kk'} E\left[\dot{g}_{t,\tau_k}(\boldsymbol{\psi}_0)\dot{g}_{t,\tau_{k'}}'(\boldsymbol{\psi}_0)\right] \text{ and } N = \sum_{k=1}^{K} f(Q_{\tau_k}(\boldsymbol{\lambda}_0)) E\left[\frac{1}{h_t(\boldsymbol{\vartheta}_0)}\dot{g}_{t,\tau_k}(\boldsymbol{\psi}_0)\dot{g}_{t,\tau_k}'(\boldsymbol{\psi}_0)\right].$$

To establish the asymptotic properties of the parametric CQR estimator $\widehat{\boldsymbol{\psi}}_n$, we still need Assumptions 1-6, but with the notations $\boldsymbol{\phi}, \boldsymbol{\phi}_0, \Phi, q_{t,\tau}(\boldsymbol{\phi})$ and $\widetilde{q}_{t,\tau}(\boldsymbol{\phi})$ in Assumptions 2 and 4-6 replaced by $\boldsymbol{\psi}, \boldsymbol{\psi}_0, \Psi, g_{t,\tau}(\boldsymbol{\psi})$ and $\widetilde{g}_{t,\tau}(\boldsymbol{\psi})$, respectively. In the following of this subsection, Assumptions 2 and 4-6 represent the conditions with the above notations replaced.

Theorem 3. *For $\{y_t\}$ generated by model (2.1), if Assumptions 1-6 hold, then we have $\widehat{\boldsymbol{\psi}}_n \to_p \boldsymbol{\psi}_0$ as $n \to \infty$.*

Theorem 4. *Suppose $\Pi = N^{-1}MN^{-1}$ is positive definite. For $\{y_t\}$ generated by model (2.1), if Assumptions 1-6 hold, then we have $\sqrt{n}(\widehat{\boldsymbol{\psi}}_n - \boldsymbol{\psi}_0) \to_{\mathcal{L}} N(0, \Pi)$ as $n \to \infty$.*

Theorems 3 and 4 establish the consistency and asymptotic normality of the parametric CQR estimator $\widehat{\boldsymbol{\psi}}_n$. Since the objective function in (2.9) is nondifferentiable and nonconvex, we also employ the bracketing method (Pollard, 1985) to prove the $\sqrt{n}$-convergence rate and asymptotic normality. To estimate the asymptotic covariance of $\widehat{\boldsymbol{\psi}}_n$, $f(Q_{\tau_k}(\boldsymbol{\lambda}_0))$ in the matrix N is similarly estimated using the difference quotient method (Koenker, 2005), i.e. $\widehat{f}(Q_{\tau_k}(\boldsymbol{\lambda}_0)) = 2h/[Q_{\tau_k+h}(\widehat{\boldsymbol{\lambda}}_n) - Q_{\tau_k-h}(\widehat{\boldsymbol{\lambda}}_n)]$ with two bandwidths in (2.7). Then matrices M and N can be approximated using sample averages with $\boldsymbol{\psi}_0$ replaced by $\widehat{\boldsymbol{\psi}}_n$ and $\dot{g}_{t,\tau_k}(\cdot)$ by $\dot{\widetilde{g}}_{t,\tau_k}(\cdot)$. As a result, the asymptotic covariance Π can be estimated by $\widehat{\Pi} = \widehat{N}^{-1}\widehat{M}\widehat{N}^{-1}$.

Note that Theorems 3-4 are established under the situation that the quantile function

$Q_\tau(\boldsymbol{\lambda})$ is correctly specified for the innovation η_t; see Assumption 4(ii). If $Q_\tau(\boldsymbol{\lambda})$ is mis-specified, then the conditional quantile function $g_{t,\tau}(\boldsymbol{\psi})$ in (2.8) is a working model, and the resulting parametric CQR estimator $\widehat{\boldsymbol{\psi}}_n$ in (2.9) will be asymptotically biased. We also establish the asymptotic properties for $\widehat{\boldsymbol{\psi}}_n$ under mis-specification in Section S1 of the Supplementary Material. Simulation results in Section 4 indicate that the parametric CQR estimator $\widetilde{\boldsymbol{\vartheta}}_n$ is insensitive to the mis-specification due to $Q_\tau(\boldsymbol{\lambda})$, while the conditional quantile estimation and forecasting can be slightly affected. As a result, in practice we can choose a distribution such as the Tukey-lambda distribution for η_t, which not only has explicit quantile function but also can approximate various distributions; see Section 3 for details.

Remark 1 (Asymptotic efficiency comparison)**.** *For general specifications of model* (2.1)*, it is difficult to determine which CQR estimator is asymptotically more efficient than the other one; see Section S2 of the Supplementary Material for details. Alternatively, we study the asymptotic relative efficiency (ARE) of $\widehat{\boldsymbol{\vartheta}}_n$ to $\widetilde{\boldsymbol{\vartheta}}_n$ via simulation studies. Simulation results in Section 4 indicate that the parametric CQR estimator $\widetilde{\boldsymbol{\vartheta}}_n$ is asymptotically more efficient than the semi-parametric CQR estimator $\widehat{\boldsymbol{\vartheta}}_n$ when the data is more heavy-tailed.*

We also compare the semi-parametric CQR with Gaussian QMLE (GQMLE) and exponential QMLE (EQMLE) in Section S2. Their AREs also depend on the distribution of η_t. When η_t follows the standard normal (or Laplace) distribution, the GQMLE (or EQMLE) is the most efficient. If η_t follows heavy-tailed distributions, the semi-parametric CQR tends to be more efficient than GQMLE and EQMLE; see Section 4 for details.

Given the CQR estimators $\widehat{\boldsymbol{\phi}}_n$ and $\widehat{\boldsymbol{\psi}}_n$, the τth conditional quantile $Q_\tau(y_{n+1}|\mathcal{F}_n)$ can be predicted by $\widetilde{q}_{n+1,\tau}(\widehat{\boldsymbol{\phi}}_n)$ and $\widetilde{g}_{n+1,\tau}(\widehat{\boldsymbol{\psi}}_n)$, respectively. The following corollary provides theoretical justification for one-step-ahead predictions of $Q_\tau(y_{n+1}|\mathcal{F}_n)$ using the CQR estimators.

Corollary 1 (VaR predictions)**.** *Under the conditions of Theorems 2 and 4, it holds that*

$$\widetilde{q}_{n+1,\tau}(\widehat{\boldsymbol{\phi}}_n) - Q_\tau(y_{n+1}|\mathcal{F}_n) = \dot{q}'_{n+1,\tau}(\boldsymbol{\phi}_0)(\widehat{\boldsymbol{\phi}}_n - \boldsymbol{\phi}_0) + o_p(n^{-1/2}) \quad \textit{and}$$

$$\widetilde{g}_{n+1,\tau}(\widehat{\boldsymbol{\psi}}_n) - Q_\tau(y_{n+1}|\mathcal{F}_n) = \dot{g}'_{n+1,\tau}(\boldsymbol{\psi}_0)(\widehat{\boldsymbol{\psi}}_n - \boldsymbol{\psi}_0) + o_p(n^{-1/2}).$$

Corollary 1 is a direct result of the Taylor expansion and Theorems 2 and 4. It indicates that the VaR prediction is determined by the representation of estimators, and thus efficiency gain in estimation can lead to improvement of VaR predictions. Simulation studies in Section 4 provide evidence that the proposed CQRs outperform the Gaussian and exponential QMLEs in one-step-ahead VaR predictions as the data gets more heavy-tailed and the target quantile level τ becomes more extreme. This is expected since the proposed CQR estimators are more efficient than the QMLEs in these situations. Moreover, simulation results also suggest the advantage of parametric CQR for predicting high conditional quantiles for heavy-tailed data, especially when the quantile function $Q_\tau(\boldsymbol{\lambda})$ is correctly specified for the innovation η_t.

3. Illustrations for CQRs

This section illustrates the application of the proposed CQRs for model (2.1) in practice. Since the ARMA-GARCH, DAR and NAR-GARCH type models in Examples 1-3 are very popular and widely used location-scale time series models, we choose them for illustration.

For the parametric CQR in Section 2.2, we need to select a parametric distribution for η_t. For illustration, we assume that η_t follows the Tukey-lambda distribution with the shape parameter λ, and then the τth quantile of η_t can be expressed as follows:

$$Q_\tau(\lambda) = \frac{\tau^\lambda - (1-\tau)^\lambda}{\lambda}. \tag{3.1}$$

The Tukey-lambda distribution not only has an explicit quantile function, but also can approximate various distributions such as the Gaussian, Logistic and heavy-tailed distributions.

3.1 CQRs in ARMA-GARCH models

This subsection examines the proposed CQRs in the framework of ARMA-GARCH type models. Here we focus on the ARMA-GARCH model in (2.2) for illustration, and the proposed CQRs can be similarly applied to other ARMA-GARCH type models, and DAR and NAR-GARCH type models in Examples 2-3; see Section S4 in the Supplementary Material.

For the ARMA(p,q)-GARCH(P,Q) model in (2.2), the location and scale functions $\mu_t(\boldsymbol{\vartheta}^{\mathrm{I}})$ and $h_t(\boldsymbol{\vartheta}^{\mathrm{I}})$ are $\mathcal{F}_{t-1}$-measurable, and they depend on observations in the infinite past. In this paper we set $y_s = \epsilon_s = 0$ and $h_s = 1$ for $s \leq 0$ as initial values, and denote the resulting conditional quantile functions as $\widetilde{q}^{\mathrm{I}}_{t,\tau_k}(\boldsymbol{\phi}^{\mathrm{I}})$ and $\widetilde{g}^{\mathrm{I}}_{t,\tau_k}(\boldsymbol{\psi}^{\mathrm{I}})$ for the semi-parametric and parametric CQRs, where $\boldsymbol{\phi}^{\mathrm{I}} = (\boldsymbol{\vartheta}^{\mathrm{I}'}, b_1, \ldots, b_K)'$ and $\boldsymbol{\psi}^{\mathrm{I}} = (\boldsymbol{\vartheta}^{\mathrm{I}'}, \lambda)'$. We will prove that the effect of initial values on the estimation is asymptotically negligible.

For model (2.2), the semi-parametric and parametric CQR estimators $\widehat{\boldsymbol{\phi}}^{\mathrm{I}}_n$ and $\widehat{\boldsymbol{\psi}}^{\mathrm{I}}_n$ are defined by (2.6) and (2.9), with $\widetilde{q}_{t,\tau_k}(\boldsymbol{\phi})$ and $\widetilde{g}_{t,\tau_k}(\boldsymbol{\psi})$ replaced by $\widetilde{q}^{\mathrm{I}}_{t,\tau_k}(\boldsymbol{\phi}^{\mathrm{I}})$ and $\widetilde{g}^{\mathrm{I}}_{t,\tau_k}(\boldsymbol{\psi}^{\mathrm{I}})$, respectively. Denote their true parameter vectors by $\boldsymbol{\phi}^{\mathrm{I}}_0 = (\boldsymbol{\vartheta}^{\mathrm{I}'}_0, b_{10}, \ldots, b_{K0})'$ and $\boldsymbol{\psi}^{\mathrm{I}}_0 = (\boldsymbol{\vartheta}^{\mathrm{I}'}_0, \lambda_0)'$, where $\boldsymbol{\vartheta}^{\mathrm{I}}_0 = (\alpha_{10}, \ldots, \alpha_{p0}, \beta_{10}, \ldots, \beta_{q0}, \gamma_{10}, \ldots, \gamma_{Q0}, \nu_{10}, \ldots, \nu_{P0})'$. We can prove that the asymptotic properties of $\widehat{\boldsymbol{\phi}}^{\mathrm{I}}_n$ and $\widehat{\boldsymbol{\psi}}^{\mathrm{I}}_n$ in Theorems 1-4 still hold for ARMA-GARCH models, with some assumptions having explicit forms. Specifically, a sufficient condition for Assumption 1 is provided in Assumptions (A2) and (A8) of Francq and Zakoïan (2004) for model (2.2). Moreover, the identification and moment conditions in Assumptions 4-5 hold if Assumptions 4′-5′ below hold; see Section S6.1 of the Supplementary Material for detailed proofs. In addition, if $\tau_k \in (0, 0.5)$ or $(0.5, 1)$ for $k = 1, 2, \ldots, K$ are chosen for the parametric CQR, then $K \geq 4$ in Assumption 4(iv) can be replaced by $K \geq 3$ for the identification. This conclusion also holds for the DAR and NAR-GARCH type models.

Assumption 4′ (Identification)**.** *(i) $\alpha(z)$ and $\beta(z)$ have no common root and $(\alpha_{p0}, \beta_{q0}) \neq (0,0)$; (ii) $\gamma(z)$ and $\nu(z)$ have no common root and $(\gamma_{Q0}, \nu_{P0}) \neq (0,0)$; (iii) $\omega = 1$ for semi-parametric CQR; (iv) $K \geq 4$ and $\lambda < 1$ for parametric CQR.*

Assumption 5′ (Moments)**.** *(i) $\underline{\alpha} \leq \alpha_i \leq \overline{\alpha}$ for $1 \leq i \leq p$, $\underline{\beta} \leq \beta_i \leq \overline{\beta}$ for $1 \leq i \leq q$, $0 < \underline{\omega} \leq \omega \leq \overline{\omega}$, $0 < \underline{\gamma} \leq \gamma_i \leq \overline{\gamma}$ for $1 \leq i \leq Q$, $0 < \underline{\nu} \leq \nu_i \leq \overline{\nu}$ for $1 \leq i \leq P$, $\underline{b} \leq b_k \leq \bar{b}$ for $1 \leq k \leq K$ and $\underline{\lambda} \leq \lambda \leq \overline{\lambda}$, where $\underline{\alpha}$, $\overline{\alpha}$, $\underline{\beta}$, $\overline{\beta}$, $\underline{\omega}$, $\overline{\omega}$, $\underline{\gamma}$, $\overline{\gamma}$, $\underline{\nu}$, $\overline{\nu}$, $\underline{b}$, $\bar{b}$, $\underline{\lambda}$ and $\overline{\lambda}$ are some constants; (ii) $E(y_t^2) < \infty$.*

3.2 Implementation issues

This subsection discusses on the computational issues for the semi-parametric and parametric CQR estimation methods for model (2.1) in practice.

First, we consider the selection of quantile levels $\{\tau_k\}_{k=1}^K$ in CQR estimations. In practice, we typically use the equally spaced quantiles, i.e. $\tau_k = k/(K+1)$ for $k = 1, 2, \ldots, K$. Simulation study indicates that both CQR methods can perform well with a reasonably large value of $K \geq 9$; see also Zou and Yuan (2008). Following the suggestion from Zou and Yuan (2008), in the simulation and empirical analysis we choose $K = 19$, i.e. $5\%, 10\%, \ldots, 95\%$ quantile levels for τ_k's. Note that the target level τ is not necessarily in the set of $\{\tau_k\}_{k=1}^K$, especially when estimating conditional quantiles at high levels. This has no effect on the parametric CQR because the explicit quantile function $Q_\tau(\boldsymbol{\lambda})$ facilitates us to do extrapolation for any high level τ. However, the τth conditional quantile $Q_\tau(y_t|\mathcal{F}_{t-1})$ in (2.5) cannot be estimated using the semi-parametric CQR since the quantile b_τ is unavailable in this situation. Instead of increasing K to cover the target level τ for the semi-parametric CQR, we include τ into the set of quantile levels to avoid high computational cost due to a very large value of K.

Second, we discuss on the optimization issues. Note that the optimization in (2.6) and (2.9) is nonlinear, and nonnegative constraints are required for the parameters in the conditional scale function $h_t(\cdot)$. To search for both CQR estimators in (2.6) and (2.9) within their parameter spaces, we use the "nlopt" function in R to handle the nonlinear optimization and impose the nonnegative constraints on the scale parameters. Moreover, the objective functions in (2.6) and (2.9) are nondifferentiable and nonconvex, thus reasonable initial values should be provided for the optimization. The initial values are chosen within the parameter space and stationary region such that Assumption 1 holds. Simulation studies indicate that both CQRs are insensitive to the selection of initial values. In simulation and empirical analysis, the Gaussian QMLEs of model (2.1) are chosen as the initial values for $\boldsymbol{\vartheta}$, monotone increasing and equally spaced values within $(-1, 1)$ for b_k with $k = 1, \ldots, K$, and 0.1 for λ.

Finally, we provide some practical suggestions for order selection of model (2.1) in the CQR estimations. Consider the order selection of ARMA-GARCH, ALDAR and ESTAR-GARCH models for illustration. For ARMA-GARCH models, the orders of the GARCH part in model (2.2) can be set as $(P, Q) = (1, 1)$, since empirical study indicates that the GARCH$(1, 1)$ model has satisfactory performance in modeling volatility (Hansen and Lunde, 2005). For the ARMA part in model (2.2), we can use the Bayesian information criterion (BIC) to select the orders (p, q). Similarly, we can use the BIC to select the orders (p, q) for the ALDAR model (2.3) and the order p for the ESTAR-GARCH model (2.4). Based on the proposed CQRs, the BIC for order selection can be defined as $\text{BIC} = 2(n - m_{\max}) \log L_n(\widehat{\boldsymbol{\upsilon}}_n) + d \log(n - m_{\max})$, where $L_n(\widehat{\boldsymbol{\upsilon}}_n) = (n - m_{\max})^{-1} \sum_{t=m_{\max}+1}^{n} \sum_{k=1}^{K} \rho_{\tau_k}(y_t - \widetilde{\zeta}_{t,\tau_k}(\widehat{\boldsymbol{\upsilon}}_n))$, $\widetilde{\zeta}_{t,\tau_k}(\widehat{\boldsymbol{\upsilon}}_n) = \widetilde{q}_{t,\tau_k}(\widehat{\boldsymbol{\phi}}_n)$ (or $\widetilde{g}_{t,\tau_k}(\widehat{\boldsymbol{\psi}}_n)$) with $\widehat{\boldsymbol{\upsilon}}_n = \widehat{\boldsymbol{\phi}}_n$ (or $\widehat{\boldsymbol{\psi}}_n$) being the semi-parametric (or parametric) CQR estimator, d is the dimension of $\widehat{\boldsymbol{\upsilon}}_n$, and $m_{\max}$ is a predetermined positive integer.

Then $(\widehat{p}_n, \widehat{q}_n) = \arg\min_{1\le p,q\le m_{\max}} \text{BIC}(p,q)$ are the selected orders for the ARMA-GARCH or ALDAR model, and $\widehat{p}_n = \arg\min_{1\le p\le m_{\max}} \text{BIC}(p)$ is the selected order for the ESTAR-GARCH model. In Section S7 of the Supplementary Material, we have established the selection consistency of the above BIC under some regular conditions.

4. Simulation

This section conducts simulation experiments to examine the finite-sample performance of the proposed CQR estimators for the ARMA-GARCH, ALDAR and ESTAR-GARCH models, and compare their performance in estimating and forecasting conditional quantiles.

Consider the following data generating processes (DGPs):

$$\text{DGP1 (ARMA(1,1)-GARCH(1,1))}: y_t = 0.2y_{t-1} + 0.1\epsilon_{t-1} + \eta_t\sqrt{1 + 0.1\epsilon_{t-1}^2 + 0.8h_{t-1}^2},$$

$$\text{DGP2 (ALDAR(1,1))}: y_t = 0.2y_{t-1} + \eta_t\left(1 + 0.2y_{t-1}^+ - 0.3y_{t-1}^-\right),$$

$$\text{DGP3 (ALDAR(1,2))}: y_t = 0.2y_{t-1} + \eta_t\left(1 + 0.3y_{t-1}^+ - 0.4y_{t-1}^- + 0.3y_{t-2}^+ - 0.4y_{t-2}^-\right),$$

$$\begin{aligned}\text{DGP4 (ESTAR(1)-GARCH(1,1))}: y_t = & -0.3 + 0.4G(y_{t-1};1,0) + \left[-0.5 + G(y_{t-1};1,0)\right]y_{t-1} \\ & + \eta_t\sqrt{0.01 + 0.2\epsilon_{t-1}^2 + 0.75h_{t-1}^2},\end{aligned}$$

where $\{\eta_t\}$ are $i.i.d$ random variables following the standard normal, Student's t_5 or Tukey-lambda distribution with the shape parameter $\lambda = 0.1$ in (3.1), denoted by F_N, F_{t_5} and F_λ, respectively. The sample size is set to $n = 500$ or 1000, with 1000 replications generated for each sample setting.

The first experiment aims to examine the finite-sample performance of the semi-parametric and parametric CQR estimators using DGP1 – DGP4. For both CQRs, the quantile levels are set to $\tau_k = k/(1+K)$ for $k = 1,\ldots,K$ with $K = 19$. Tables 1-2 list the biases, empirical standard deviations (ESDs), and asymptotic standard deviations (ASDs) of both CQRs for

DGP1, and the results for DGP2 − DGP4 are listed in Section S8 of the Supplementary Material to save space. Here, the ASD for the ith element of the semi-parametric (or parametric) CQR estimator is calculated as the average of the estimated standard deviations $\widehat{\Xi}_{ii}/\sqrt{n}$ (or $\widehat{\Pi}_{ii}/\sqrt{n}$) among 1000 replications, where A_{ii} is the ith diagonal element of a square matrix A. Our findings can be summarized as follows: (i) as the sample size increases, biases, ESDs, and ASDs generally decrease, and ESDs approach ASDs; (ii) most of the ASDs and ESDs increase as the distribution of η_t gets more heavy-tailed, indicating a lower estimation efficiency for more heavy-tailed data; (iii) the ASDs of the semi-parametric CQR using h_{HS} in (2.7) are slightly smaller than those using h_B, and closer to the corresponding ESDs. Thus we utilize h_{HS} in the empirical analysis; (iv) for the mis-specified situation of $Q_\tau(\boldsymbol{\lambda})$ that η_t follows F_N or F_{t_5} but the Tukey-lambda distribution is employed for $Q_\tau(\boldsymbol{\lambda})$, the biases of the parametric CQR estimator are still small, and the ESDs/ASDs are close to those of the semi-parametric CQR estimator. Therefore, the parametric CQR is insensitive to the mis-specification due to $Q_\tau(\boldsymbol{\lambda})$.

In the second experiment, we compare the performance of the semi-parametric and parametric CQRs, GQMLE and EQMLE in estimating and predicting conditional quantiles for ARMA-GARCH, ALDAR and ESTAR-GARCH models using DGP1, DGP2 and DGP4. To evaluate the performance of in-sample estimation and out-of-sample prediction of the above estimation methods, we use the mean biases and root mean square errors (RMSEs): $\text{Bias}_{in} = (mn)^{-1}\sum_{i=1}^{m}\sum_{t=1}^{n}[\widehat{Q}_\tau^{(i)}(y_t|\mathcal{F}_{t-1}) - Q_\tau^{(i)}(y_t|\mathcal{F}_{t-1})]$, $\text{Bias}_{out} = m^{-1}\sum_{i=1}^{m}[\widehat{Q}_\tau^{(i)}(y_{n+1}|\mathcal{F}_n) - Q_\tau^{(i)}(y_{n+1}|\mathcal{F}_n)]$, $\text{RMSE}_{in} = \{(mn)^{-1}\sum_{i=1}^{m}\sum_{t=1}^{n}[\widehat{Q}_\tau^{(i)}(y_t|\mathcal{F}_{t-1}) - Q_\tau^{(i)}(y_t|\mathcal{F}_{t-1})]^2\}^{1/2}$ and $\text{RMSE}_{out} = \{m^{-1}\sum_{i=1}^{m}[\widehat{Q}_\tau^{(i)}(y_{n+1}|\mathcal{F}_n) - Q_\tau^{(i)}(y_{n+1}|\mathcal{F}_n)]^2\}^{1/2}$, where $m = 1000$ is the number of replications, $n = 1000$ is the sample size, and $Q_\tau^{(i)}(y_t|\mathcal{F}_{t-1})$ and $\widehat{Q}_\tau^{(i)}(y_t|\mathcal{F}_{t-1})$ represent the τth conditional quantile and its estimate at time t in the ith replication, respectively. In order to compare the performance of the four estimation methods at moderate and high quantile levels, we consider two sets $\mathcal{T}_1 = \{0.05, 0.1, 0.9, 0.95\}$ and $\mathcal{T}_2 = \{0.001, 0.005, 0.995, 0.999\}$ for the target quantile level τ, respectively. For both CQRs, we choose $\{\tau_k\}_{k=1}^K$ with $\tau_k = k/(1+K)$ and $K = 19$. When estimating and predicting high conditional quantiles at $\tau \in \mathcal{T}_2$, note that the target level τ is not in the set of $\{\tau_k\}_{k=1}^K$. We additionally include τ into this set for the semi-parametric CQR.

For illustration, Table 3 and Table S.1 in the Supplementary Material list the biases and RMSEs of the in-sample estimation and out-of-sample prediction using the semi-parametric and parametric CQRs, GQMLE and EQMLE for DGP1 at $\tau \in \mathcal{T}_1$ and $\tau \in \mathcal{T}_2$, respectively. The results for DGP2 and DGP4 are listed in Section S8 of the Supplementary Material to save space. We have the following findings for the ARMA-GARCH model in DGP1: (i) as the distribution of η_t becomes more heavy-tailed or the target quantile level τ gets more extreme, the biases and RMSEs of all the estimation methods generally increase, indicating that the accuracy of estimation and prediction decreases; (ii) the RMSEs of both CQRs are smaller than those of GQMLE and EQMLE for heavy-tailed η_t or extreme target quantile level τ. This implies that the proposed CQRs are more efficient than the QMLEs in estimating and predicting high conditional quantiles of heavy-tailed data; (iii) in terms of RMSEs, the parametric CQR outperforms the semi-parametric CQR for most situations, especially when η_t is heavy-tailed or the target quantile level τ is extreme. This suggests the advantage of parametric CQR in estimating and predicting high conditional quantiles for heavy-tailed data; (iv) for the mis-specified situations that η_t follows F_N or F_{t_5} but Tukey-lambda distribution is employed for $Q_\tau(\boldsymbol{\lambda})$, the parametric CQR is slightly worse than those of the semi-parametric

CQR in biases and RMSEs. This implies that the conditional quantile estimation and prediction using parametric CQR is not sensitive to this mis-specification. For the ALDAR and ESTAR-GARCH models, it can be found that the semi-parametric CQR and EQMLE perform similarly and they have better performance than the GQMLE, while the parametric CQR outperforms the semi-parametric CQR when the quantile function $Q_\tau(\boldsymbol{\lambda})$ is correctly specified for the innovation η_t.

In the third experiment, we aim to compare the semi-parametric CQR with the parametric CQR, GQMLE and EQMLE using ARE in Section S2 of the Supplementary Material. To calculate the ARE, we generate a sequence of sample size $n = 10000$ from the model in DGP1, DGP2 or DGP4, and two scenarios are considered for η_t: (a) η_t follows the Tukey-lambda distribution with the shape parameter $\lambda = 0.1 + 0.02k$ for $k = 0, 1, \dots, 20$; (b) η_t follows the mixture distribution with the probability density function (pdf) $f(x) = (1-\delta)\phi(x) + \delta m(x)$, where $\delta = k/20$ with $k = 0, 1, \dots, 20$, $\phi(x)$ is the pdf of $N(0, 1)$, and $m(x)$ is the pdf of $N(0, 6)$, standard Laplace or t_5 distribution. Note that the Tukey-lambda distribution in scenario (a) covers the distributions with light and heavy tails, and the smaller value of λ implies more heavy-tailed distribution. To study the ARE of two CQR estimators, i.e. $\text{ARE}(\widehat{\boldsymbol{\vartheta}}_n, \widetilde{\boldsymbol{\vartheta}}_n)$, we focus on the case that the quantile function $Q_\tau(\boldsymbol{\lambda})$ is correctly specified by parametric CQR, and thus we consider scenario (a). Meanwhile, for AREs of the semi-parametric CQR to GQMLE and EQMLE, i.e. $\text{ARE}(\widehat{\boldsymbol{\vartheta}}_n, \check{\boldsymbol{\vartheta}}_n)$ and $\text{ARE}(\widehat{\boldsymbol{\vartheta}}_n, \breve{\boldsymbol{\vartheta}}_n)$, we consider scenario (b).

Figure 1 plots $\text{ARE}(\widehat{\boldsymbol{\vartheta}}_n, \widetilde{\boldsymbol{\vartheta}}_n)$, $\text{ARE}(\widehat{\boldsymbol{\vartheta}}_n, \check{\boldsymbol{\vartheta}}_n)$ and $\text{ARE}(\widehat{\boldsymbol{\vartheta}}_n, \breve{\boldsymbol{\vartheta}}_n)$ for the ARMA-GARCH model; see Section S3 in the Supplementary Material for calculation details. It can be seen that the ARE can be either larger or smaller than one, which indicates that no estimator can dominate the others for general situations. Specifically, we have the following findings: (i) as η_t becomes more heavy-tailed, ARE$(\widehat{\boldsymbol{\vartheta}}_n, \widetilde{\boldsymbol{\vartheta}}_n)$ gets smaller than one, and thus the parametric CQR tends to be more efficient than the semi-parametric CQR; (ii) the semi-parametric CQR is less efficient than GQMLE (or EQMLE) when η_t approximately follows the normal (or Laplace) distribution, but it tends to be more efficient than GQMLE and EQMLE when η_t becomes more heavy-tailed; (iii) when $\delta = 0$ such that $\eta_t \sim N(0,1)$, then ARE$(\widehat{\boldsymbol{\vartheta}}_n, \check{\boldsymbol{\vartheta}}_n) < 1$ and the GQMLE is the most efficient. This is because the GQMLE reduces to the MLE when $\eta_t \sim N(0,1)$ and its asymptotic covariance attains the Cramér-Rao lower bound; (iv) when $\delta = 1$ and $m(x)$ is the pdf of a standard Laplace distribution such that η_t follows a standard Laplace distribution, then ARE$(\widehat{\boldsymbol{\vartheta}}_n, \check{\boldsymbol{\vartheta}}_n) < 1$ and the EQMLE is the most efficient. In this situation the EQMLE reduces to the MLE with its asymptotic covariance attains the Cramér-Rao lower bound. In addtion, the results for the ALDAR and ESTAR-GARCH models are relegated to Section S8 of the Supplementary Material to save space, and the simulation findings are unchanged.

5. Empirical analysis

This section analyzes the daily closing prices of Microsoft Corp's stock (MSFT), denoted as p_t, from January 2, 2018 to December 29, 2023. The dataset is downloaded from the website of CRSP via WRDS (https://wrds-www.wharton.upenn.edu/pages/get-data/center-research-security-prices-crsp/), with a total of 1509 observations. Let $r_t = 100(\ln p_t - \ln p_{t-1})$ be the log return in percentage, and denote the centered series by $\{y_t\}$. Figure 2 illustrates volatility clustering of $\{y_t\}$, and the sample skewness of -0.23 and kurtosis of 9.82 indicate that $\{y_t\}$ is skewed and heavy-tailed; see Table 4 for summary statistics. The ACF and

PACF plots of $\{y_t\}$ in Figure S.3 of the Supplementary Material imply that $\{y_t\}$ is serial correlated. Therefore, location-scale time series models are suitable for $\{y_t\}$ to capture the serial correlation and volatility clustering.

We choose the ARMA-GARCH, ALDAR and ESTAR-GARCH models in Examples 1-3 to fit $\{y_t\}$, and estimate these models by the proposed semi-parametric and parametric CQRs. To obtain the CQR estimators in (2.6) and (2.9), we set $\tau_k = k/(K+1)$ for $k = 1, \ldots, K$ with $K = 19$. We use the BIC in Section 3.2 to select the orders of ARMA-GARCH, ALDAR and ESTAR-GARCH models. Given $m_{\max} = 5$, we select $(p, q) = (1, 1)$ for the ARMA-GARCH model while $(p, q) = (1, 2)$ for the ALDAR model and $p = 1$ for the ESTAR-GARCH model using both CQRs. The estimation results of these models are summarized in Table 5.

Value-at-Risk (VaR) is a widely used risk measure for financial assets, and it is equivalent to the negative conditional quantile of the return series $\{y_t\}$. In order to assess the forecasting performance of the fitted models, one-step-ahead predictions of VaR are conducted using a rolling forecasting procedure with a fixed moving window. Specifically, we employ both CQRs in Section 2 to fit the ARMA$(1, 1)$-GARCH$(1, 1)$, ALDAR$(1, 2)$ and ESTAR(1)-GARCH$(1, 1)$ models for each moving window of size 1000. Subsequently, we calculate the one-step-ahead forecast for the τth conditional quantile of $\{y_t\}$. To assess the forecasting performance of both CQRs, we consider two VaR backtests and the empirical coverage rate (ECR). The ECR represents the proportion of observations that fall below the corresponding conditional quantile forecasts. The VaR backtests are the likelihood ratio test for correct conditional coverage (CC) in Christoffersen (1998) and the dynamic quantile test (DQ) in Engle and Manganelli (2004). Denote the hit by $H_t = I(y_t < Q_\tau(y_t|\mathcal{F}_{t-1}))$. The null hypothesis of CC test assumes that $\{H_t\}$ given $\mathcal{F}_{t-1}$ are $i.i.d.$ Bernoulli random variables with a success probability of τ. The DQ test conducts the regression of H_t on a constant, four lagged hits $H_{t-i}(i = 1, 2, 3, 4)$ and the contemporaneous VaR forecast. Its null hypothesis is that the intercept is equal to τ and other regression coefficients are zero. If both backtests cannot reject the null hypotheses and the ECR is close to τ, then the forecasting method is satisfactory. Additionally, we employ the model confidence set (MCS) procedure in Hansen et al. (2011) to compare the fitted models. If the MCS p-value of the fitted model is larger than a given significance level, then this model can be added into the set of superior models.

We first calculate the one-step-ahead VaR forecasts at the moderate quantile levels, i.e. $\tau = 5\%, 10\%, 90\%$ and 95%. Given the semi-parametric CQR estimators $\widehat{\boldsymbol{\vartheta}}_n$ and $\widehat{b}_\tau$, the τth conditional quantile forecast of $\{y_t\}$ is calculated as $\widehat{Q}_\tau(y_{n+1}|\ \mathcal{F}_n) = \mu_{n+1}(\widehat{\boldsymbol{\vartheta}}_n) + \widehat{b}_\tau h_{n+1}(\widehat{\boldsymbol{\vartheta}}_n)$. Given the parametric CQR estimators $\widetilde{\boldsymbol{\vartheta}}_n$ and $\widehat{\lambda}_n$, the τth conditional quantile forecast of $\{y_t\}$ is calculated by $\widetilde{Q}_\tau(y_{n+1}|\mathcal{F}_n) = \mu_{n+1}(\widetilde{\boldsymbol{\vartheta}}_n) + Q_\tau(\widehat{\lambda}_n)h_{n+1}(\widetilde{\boldsymbol{\vartheta}}_n)$. Using the above procedure, we obtain 508 one-step-ahead VaR forecasts fitted by the ARMA-GARCH, ALDAR and ESTAR-GARCH models using both CQRs for each τ. We compare the forecasting performance of both CQRs with other VaR estimation methods including FHS, CAViaR, and Risk Metrics. Specifically, the FHS is based on the ARMA-GARCH model fitted by GQMLE (Francq and Zakoïan, 2004) and EQMLE (Zhu and Ling, 2011), the ALDAR model fitted by GQMLE (Tan and Zhu, 2022) and EQMLE (Tan and Zhu, 2023), and the ESTAR-GARCH model fitted by GQMLE (Chan and McAleer, 2002) and EQMLE, respectively. Moreover, the CAViaR employs an indirect GARCH$(1, 1)$ specification (Engle and Manganelli, 2004). Table 6 summarizes all the methods for comparison.

Table 7 reports the ECRs and p-values of VaR backtests and MCS for one-step-ahead forecasts. When the series $\{y_t\}$ is fitted by the ARMA-GARCH model, the semi-parametric

CQR method outperforms the other four estimation methods at $\tau = 90\%$ and 95%. This is evident since its ECRs are close to the nominal quantile levels, and the p-values of the two backtests are greater than or equal to 0.15. While the FHS methods excels solely at $\tau = 10\%$. When fitting $\{y_t\}$ by ALDAR model, both CQRs demonstrate good performance at $\tau = 90\%$ and 95% in terms of ECRs and backtests. However, the FHS method fitted by GQMLE performs acceptably only at $\tau = 5\%$, while the FHS method fitted by EQMLE fails in the backtests at all levels. When fitting $\{y_t\}$ by ESTAR-GARCH model, the semi-parametric CQR method outperforms the other estimation methods at $\tau = 5\%$, 90% and 95%, while the FHS methods perform the best only at $\tau = 10\%$. Meanwhile, Risk Metrics and CAViaR methods perform better than CQRs at $\tau = 10\%$. Overall, the ARMA-GARCH and ESTAR-GARCH models fitted by semi-parametric CQR outperform the others in terms of backtests at three quantile levels, and the ECRs of the former are close to nominal quantile levels at all quantile levels. And among all the fitted models, only the ALDAR model fitted by GQMLE does not belong to the set of superior models, as its MCS p-values are smaller than 0.1 at all quantile levels. Consequently, the ARMA-GARCH model fitted by semi-parametric CQR dominates the others in predicting VaRs of $\{y_t\}$ at moderate quantile levels.

Furthermore, to demonstrate the forecasting performance of both CQRs at high quantile levels, we calculate the one-step-ahead VaR forecasts at $\tau = 0.1\%, 0.5\%, 99.5\%$ and 99.9%. To ensure accurate evaluation, we expand the dataset and use the daily closing prices from January 2, 2002 to December 29, 2023, including 5537 observations. The data is transformed into centered series of log returns in percentage denoted by $\{y_t\}$. A moving window with size 1000 is utilized to generate a total of 4536 one-step-ahead predictions through a rolling forecasting procedure. The forecasting procedure for the parametric CQR at high quantile levels remains consistent with that at moderate levels. While we additionally include τ into the set $\{\tau_k\}_{k=1}^{K}$ with $K = 19$ for the semi-parametric CQR at high quantile levels. As for moderate quantile levels, we compare the performance of both CQRs with FHS, CAViaR, and Risk Metrics methods at high quantile levels.

Table 8 shows the ECRs and p-values of VaR backtests and MCS for one-step-ahead forecasts at high levels. When fitting $\{y_t\}$ by the ARMA-GARCH model, the proposed CQRs have superior performance than FHS at $\tau = 0.1\%, 0.5\%$ and 99.5% in terms of ECRs and backtests. When fitting $\{y_t\}$ by the ALDAR model, the proposed CQRs outperform the other methods at all quantile levels, because its ECRs are closest to the corresponding quantile levels and the p-values of two backtests are no less than 0.15. When fitting $\{y_t\}$ by the ESTAR-GARCH model, the proposed CQRs have superior performance at $\tau = 0.5\%$ and 99.5%, while the FHS performs better at $\tau = 0.1\%$ and 99.9%. Notably, both Risk Metrics and CAViaR exhibit poor performance at high quantile levels. Among all the considered methods, the ALDAR model estimated by parametric CQR exhibits the closest ECRs to nominal quantile levels for $\tau = 0.1\%$, 99.5% and 99.9%, with the p-values of two backtests no less than 0.15. Moreover, only the ALDAR model fitted by GQMLE, Risk Metrics and CAViaR do not belong to the set of superior models, since their MCS p-values are less than 0.1 at all quantile levels. In conclusion, for predicting VaRs at high quantile levels, the ALDAR model estimated using parametric CQR is more suitable than the other methods.

6. Conclusion and discussion

This paper explores efficient estimation of VaR in the framework of location-scale time series models using CQR. We propose the semi-parametric and parametric CQRs for estimation and establish their asymptotic properties. We also compare their estimation efficiency with that of the Gaussian and exponential QMLEs. Moreover, we apply both CQRs to ARMA-GARCH, DAR and NAR-GARCH type models, and establish their asymptotic properties under regular conditions. By simulation studies and empirical analysis on Microsoft Corp's stock return, we demonstrate that both CQRs can provide robust model estimation and accurate VaR forecasts at moderate and high quantile levels especially for heavy-tailed data.

Our research can be extended in several directions. Firstly, self-weights can be introduced into both CQRs to reduce the moment requirement in our framework. In this study, $E(|y_t|^s) < \infty$ for $s \geq 3$ is required to establish asymptotic normality of CQR estimators for ARMA-GARCH, ALDAR and ESTAR-GARCH models. This moment requirement is stringent for financial time series with heavy tails, whereas our methods combining self-weights can handle more heavy-tailed time series. Secondly, statistical tools such as the goodness-of-fit test can be proposed with theoretical guarantee for model diagnosis. Thirdly, the proposed framework can be extended to estimate the risk measures VaR and expected shortfall jointly, using the fact that expected shortfall is jointly elicitable with VaR (Fissler and Ziegel, 2016). Lastly, our methods can also be utilized to construct prediction intervals. It would be intriguing to compare the validity and efficiency of our approaches with other methods, such as the conformal prediction for time series (Stankeviciute et al., 2021; Barber et al., 2023).

Supplementary Materials

The online Supplementary Material includes all technical details for Sections 2–3, together with additional results for simulation studies and empirical analysis in Sections 4–5.

Acknowledgements

We are deeply grateful to the Co-Editor, the Associate Editor and two anonymous referees for their valuable comments that led to the substantial improvement in the quality of this paper. Zhu's research was supported by NSFC grants 12001355 and 72373087.

References

Barber, R. F., E. J. Candes, A. Ramdas, and R. J. Tibshirani (2023). Conformal prediction beyond exchangeability. *The Annals of Statistics 51*, 816–845.

Barone-Adessi, G., K. Giannopoulos, and L. Vosper (1999). VaR without correlations for nonlinear portfolios. *Journal of Future Markets 19*, 583–602.

Boudoukh, J., M. Richardson, and R. Whitelaw (1998). The best of both worlds. *Risk 11*, 64–67.

Chan, F. and M. McAleer (2002). Maximum likelihood estimation of STAR and STAR-GARCH models: theory and Monte Carlo evidence. *Journal of Applied Econometrics 17*, 509–534.

Chan, F., M. McAleer, and M. C. Medeiros (2015). Structure and asymptotic theory for nonlinear models with GARCH errors. *EconomiA 16*, 1–21.

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

Cline, D. B. (2007). Stability of nonlinear stochastic recursions with application to nonlinear AR-GARCH models. *Advances in Applied Probability 14*, 1920–1949.

Cline, D. B. and H.-m. H. Pu (2004). Stability and the Lyapounov exponent of threshold AR-ARCH models. *The Annals of Applied Probability 39*, 841–862.

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

Fissler, T. and J. F. Ziegel (2016). Higher order elicitability and Osband's principle. *The Annals of Statistics 44*, 1680–1707.

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

Gilchrist, W. (2000). *Statistical modelling with quantile functions.* Chapman and Hall/CRC.

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.

Hansen, P. R., A. Lunde, and J. M. Nason (2011). The model confidence set. *Econometrica 79*, 453–497.

Jiang, F., D. Li, and K. Zhu (2020). Non-standard inference for augmented double autoregressive models with null volatility coefficients. *Journal of Econometrics 215*, 165–183.

Jiang, J., X. Jiang, and X. Song (2014). Weighted composite quantile regression estimation of DTARCH models. *The Econometrics Journal 17*, 1–23.

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

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

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.

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

Li, D., S. Ling, and J.-M. Zakoïan (2015). Asymptotic inference in multiple-threshold double autoregressive models. *Journal of Econometrics 189*, 415–427.

Li, D., S. Ling, and R. Zhang (2016). On a threshold double autoregressive model. *Journal of Business & Economic*

Statistics 34, 68–80.

Li, G., Q. Zhu, Z. Liu, and W. K. Li (2017). On mixture double autoregressive time series models. *Journal of Business & Economic Statistics 35*, 306–317.

Ling, S. (1999). On the probabilistic properties of a double threshold arma conditional heteroskedastic model. *Journal of Applied probability 36*, 688–705.

Ling, S. (2007). A double AR(p) model: structure and estimation. *Statistica Sinica 17*, 161–175.

Liu, J., W. K. Li, and C. Li (1997). On a threshold autoregression with conditional heteroscedastic variances. *Journal of Statistical Planning and Inference 62*, 279–300.

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

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*, 295–313.

Stankeviciute, K., A. M Alaa, and M. van der Schaar (2021). Conformal time-series forecasting. *Advances in Neural Information Processing Systems 34*, 6216–6228.

Tan, S. and Q. Zhu (2022). Asymmetric linear double autoregression. *Journal of Time Series Analysis 43*, 371–388.

Tan, S. and Q. Zhu (2023). On dual-asymmetry linear double AR models. *Statistics and Its Interface 16*, 3–16.

Tsay, R. S. (2010). *Analysis of financial time series.* 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, G., K. Zhu, G. Li, and W. K. Li (2022). Hybrid quantile estimation for asymmetric power GARCH models. *Journal of Econometrics 227*, 264–284.

Wang, M., Z. Chen, and C. D. Wang (2018). Composite quantile regression for GARCH models using high-frequency data. *Econometrics and Statistics 7*, 115–133.

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

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: Statistical Methodology 80*, 975–993.

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

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

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

Zhu, Q., S. Tan, Y. Zheng, and G. Li (2023). Quantile autoregressive conditional heteroscedasticity. *Journal of the Royal Statistical Society Series B: Statistical Methodology 85*, 1099–1127.

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.

Chaoxu Lei, Shanghai University of Finance and Economics, School of Statistics and Data Science

E-mail: lchx@stu.sufe.edu.cn

Qianqian Zhu, Shanghai University of Finance and Economics, School of Statistics and Data Science

E-mail: zhu.qianqian@mail.shufe.edu.cn

Table 1: Biases, ASDs, and ESDs of the semi-parametric CQR estimator for DGP1, where the innovations follow the standard normal, Student's t_5 or Tukey-lambda distribution with the shape parameter $\lambda = 0.1$, denoted by F_N, F_{t_5} or F_λ, respectively. ASD_1 and ASD_2 correspond to the bandwidths h_B and h_{HS}, respectively.

	n	Bias	ASD_1	ASD_2	ESD	Bias	ASD_1	ASD_2	ESD	Bias	ASD_1	ASD_2	ESD
		$F = F_N$				$F = F_{t_5}$				$F = F_\lambda$			
α_1	500	-0.027	0.172	0.167	0.171	-0.015	0.187	0.179	0.178	-0.010	0.212	0.205	0.207
	1000	-0.010	0.125	0.124	0.116	-0.015	0.132	0.127	0.128	-0.008	0.132	0.128	0.128
β_1	500	0.025	0.176	0.171	0.172	0.013	0.191	0.182	0.182	0.000	0.220	0.213	0.213
	1000	0.008	0.122	0.120	0.117	0.010	0.133	0.128	0.131	0.005	0.134	0.129	0.130
γ_1	500	-0.001	0.063	0.061	0.061	-0.004	0.061	0.057	0.048	0.026	0.123	0.128	0.137
	1000	-0.002	0.041	0.040	0.042	-0.011	0.040	0.039	0.037	-0.025	0.075	0.072	0.072
ν_1	500	-0.021	0.122	0.118	0.132	-0.005	0.058	0.055	0.066	0.017	0.053	0.052	0.052
	1000	-0.017	0.089	0.087	0.095	-0.004	0.064	0.062	0.058	0.014	0.040	0.038	0.035

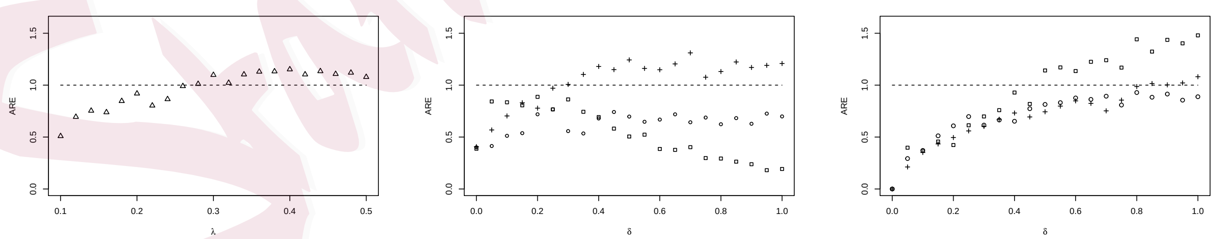

Figure 1: The $\text{ARE}(\widehat{\boldsymbol{\vartheta}}_n, \widetilde{\boldsymbol{\vartheta}}_n)$ (left), $\text{ARE}(\widehat{\boldsymbol{\vartheta}}_n, \check{\boldsymbol{\vartheta}}_n)$ (middle) and $\text{ARE}(\widehat{\boldsymbol{\vartheta}}_n, \breve{\boldsymbol{\vartheta}}_n)$ (right) for ARMA-GARCH models, where $\lambda = 0.1 + 0.02k$ and $\delta = k/20$ with $k = 0, 1, \ldots, 20$, for Tukey-lambda ($\triangle$), $N(0, 6)$ ($\square$), t_5 ($+$) or standard Laplace ($\circ$) distribution.

Table 2: Biases, ASDs, and ESDs of the parametric CQR estimator for DGP1, where the innovations follow the standard normal, Student's t_5 or Tukey-lambda distribution with the shape parameter $\lambda = 0.1$, denoted by F_N, F_{t_5} or F_λ, respectively.

	n	Bias	ASD	ESD	Bias	ASD	ESD	Bias	ASD	ESD
		$F = F_N$			$F = F_{t_5}$			$F = F_\lambda$		
α_1	500	-0.018	0.166	0.170	-0.005	0.179	0.180	0.000	0.202	0.193
	1000	-0.006	0.122	0.117	-0.011	0.131	0.123	-0.002	0.123	0.129
β_1	500	0.018	0.175	0.171	0.005	0.180	0.184	-0.006	0.196	0.197
	1000	0.005	0.119	0.118	0.008	0.134	0.123	0.000	0.125	0.131
γ_1	500	-0.049	0.030	0.028	-0.060	0.021	0.016	-0.023	0.042	0.029
	1000	-0.051	0.017	0.019	-0.061	0.012	0.012	-0.024	0.020	0.020
ν_1	500	-0.049	0.135	0.162	-0.007	0.065	0.070	0.009	0.066	0.056
	1000	-0.020	0.071	0.099	-0.003	0.046	0.049	0.010	0.037	0.033

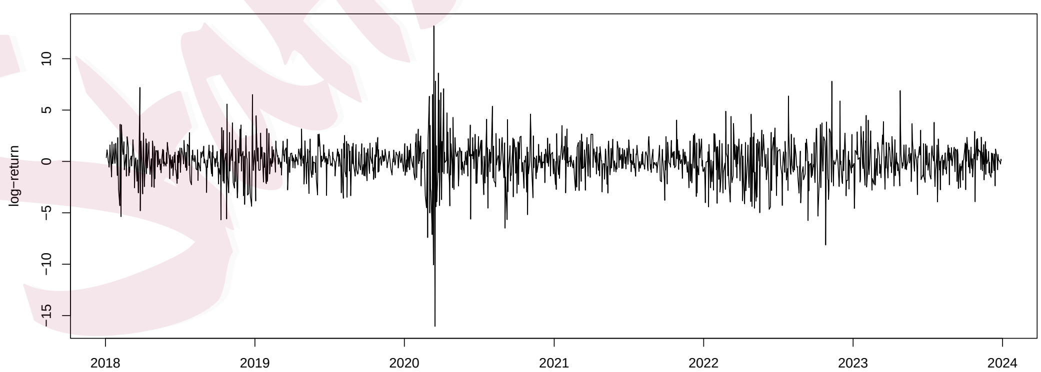

Figure 2: Time plot for daily log returns in percentage of Microsoft Corp's stock (MSFT) from January 2, 2018 to December 29, 2023.

Table 3: Biases and RMSEs for estimating and predicting conditional quantiles at $\tau = 5\%, 10\%, 90\%$ and 95% for DGP1, where the innovations follow the standard normal, Student's t_5 or Tukey-lambda distribution with the shape parameter $\lambda = 0.1$, denoted by F_N, F_{t_5} or F_λ, respectively. M1, M2, M3 and M4 represent the semi-parametric CQR, parametric CQR, GQMLE and EQMLE, respectively.

		F_N				F_{t_5}				F_λ			
		Bias		RMSE		Bias		RMSE		Bias		RMSE	
τ	F	in	out	in	out	in	out	in	out	in	out	in	out
5%	M1	0.011	0.036	0.342	0.341	0.064	0.057	0.830	0.897	1.586	2.215	2.574	3.254
	M2	0.019	0.043	0.303	0.302	0.111	0.105	0.710	0.760	1.803	2.550	2.715	3.549
	M3	0.015	0.016	0.156	0.154	-0.128	0.023	3.735	4.035	2.231	4.694	10.250	13.859
	M4	0.057	0.008	0.490	0.176	0.495	0.098	4.419	5.011	2.691	4.981	11.251	15.656
10%	M1	0.005	0.026	0.292	0.295	0.026	0.031	0.631	0.661	1.186	1.679	2.014	2.544
	M2	0.006	0.026	0.250	0.246	0.037	0.038	0.538	0.572	1.334	1.904	2.068	2.718
	M3	0.014	0.014	0.126	0.126	0.248	0.382	3.240	3.574	2.544	4.608	8.805	11.975
	M4	0.049	0.009	0.382	0.144	0.734	0.519	3.845	4.271	3.026	5.069	9.697	13.723
90%	M1	-0.012	-0.019	0.284	0.291	-0.071	-0.034	0.616	0.642	-1.509	-2.024	2.280	2.858
	M2	-0.005	-0.010	0.251	0.258	-0.037	-0.003	0.539	0.583	-1.332	-1.873	2.067	2.674
	M3	-0.009	-0.013	0.126	0.122	-0.283	-0.477	3.306	3.760	-2.649	-4.695	8.946	12.343
	M4	-0.044	-0.012	0.382	0.141	-0.757	-0.578	3.892	4.364	-3.144	-5.157	9.823	13.485
95%	M1	-0.019	-0.029	0.338	0.346	-0.127	-0.078	0.813	0.888	-2.062	-2.762	2.978	3.783
	M2	-0.018	-0.027	0.304	0.314	-0.112	-0.070	0.711	0.767	-1.802	-2.519	2.714	3.516
	M3	-0.011	-0.016	0.156	0.153	0.097	-0.117	3.794	4.297	-2.353	-4.775	10.405	14.289
	M4	-0.052	-0.011	0.495	0.176	-0.526	-0.165	4.499	5.111	-2.803	-5.088	11.389	15.642

Table 4: Summary statistics for $\{y_t\}$.

Mean	Median	Std.Dev.	Skewness	Kurtosis	Min	Max
0.00	0.03	1.9	-0.23	9.82	-16.04	13.2

Table 5: The fitted coefficients (coef) with standard errors (s.e.) of the ARMA(1, 1)-GARCH(1, 1) (upper), ALDAR(1, 2) (middle) and ESTAR(1)-GARCH(1, 1) (lower) models, where M1 and M2 represent the semi-parametric and parametric CQR methods, respectively.

		α_1	β_1	ω	γ_1	ν_1				
M1	coef	0.416	-0.531		0.116	0.834				
	s.e.	0.180	0.166		0.107	0.066				
M2	coef	0.406	-0.520	0.034	0.042	0.828				
	s.e.	0.180	0.167	0.012	0.010	0.031				
		ϕ_1	ω	α_1^+	α_2^+	α_1^-	α_2^-			
M1	coef	-0.076	0.091	0.125	0.265	0.256				
	s.e.	0.030	0.045	0.051	0.059	0.061				
M2	coef	-0.077	0.578	0.053	0.081	0.155	0.145			
	coef	0.029	0.047	0.024	0.027	0.030	0.030			
		α_{00}	α_{01}	α_{10}	α_{11}	γ	c	ω	a	b
M1	coef	-0.124	-0.175	0.145	0.034	1.025	0.068		0.406	0.986
	s.e.	0.159	0.177	0.240	0.027	1.960	0.976		0.194	0.001
M2	coef	0.045	0.018	-0.081	-0.812	0.007	0.085	0.029	0.042	0.839
	s.e.	0.044	1.272	0.434	0.852	0.012	0.707	0.011	0.010	0.028

Table 6: The methods for comparison.

Abbreviation	Method description
M1a	ARMA-GARCH model fitted by semi-parametric CQR
M1b	ARMA-GARCH model fitted by parametric CQR
M1c	FHS with ARMA-GARCH model fitted by GQMLE
M1d	FHS with ARMA-GARCH model fitted by EQMLE
M2a	ALDAR model fitted by semi-parametric CQR
M2b	ALDAR model fitted by parametric CQR
M2c	FHS with ALDAR model fitted by GQMLE
M2d	FHS with ALDAR model fitted by EQMLE
M3a	ESTAR-GARCH model fitted by semi-parametric CQR
M3b	ESTAR-GARCH model fitted by parametric CQR
M3c	FHS with ESTAR-GARCH model fitted by GQMLE
M3d	FHS with ESTAR-GARCH model fitted by EQMLE
M4	Risk Metrics
M5	CAViaR with the specification of indirect GARCH$(1,1)$

Table 7: Empirical coverage rate (%) and p-values of two VaR backtests and MCS at the 5%, 10%, 90% and 95% quantile levels, where methods M1a-M5 are summarized in Table 6. The ECRs closest to the nominal level for each model are marked in bold.

	$\tau = 5\%$				$\tau = 10\%$				$\tau = 90\%$				$\tau = 95\%$			
	ECR	CC	DQ	MCS	ECR	CC	DQ	MCS	ECR	CC	DQ	MCS	ECR	CC	DQ	MCS
M1a	5.12	0.94	0.18	0.53	11.81	0.41	0.01	0.51	**90.54**	0.70	0.65	1.00	**95.08**	0.97	0.99	1.00
M1b	5.91	0.53	0.09	0.80	12.60	0.15	0.01	0.71	91.14	0.59	0.62	1.00	96.06	0.51	0.95	0.94
M1c	5.31	0.85	0.94	1.00	**11.40**	0.20	0.30	0.71	88.76	0.57	0.67	1.00	94.09	0.65	0.46	1.00
M1d	4.72	0.95	0.78	1.00	11.60	0.17	0.29	1.00	89.37	0.76	0.29	0.99	94.09	0.53	0.82	1.00
M2a	5.91	0.10	0.05	0.75	**12.40**	0.21	0.03	0.22	88.19	0.38	0.55	0.75	94.29	0.65	0.70	1.00
M2b	6.69	0.02	0.01	0.75	13.19	0.07	0.00	0.26	89.17	0.55	0.27	1.00	**95.67**	0.29	0.65	1.00
M2c	**4.53**	0.30	0.44	0.01	12.83	0.10	0.02	0.04	86.22	0.01	0.00	0.00	92.91	0.01	0.01	0.01
M2d	6.30	0.30	0.00	0.45	13.58	0.03	0.00	0.34	**89.76**	0.20	0.00	0.92	**95.67**	0.29	0.00	0.85
M3a	**4.89**	0.95	0.22	0.15	12.02	0.75	0.02	0.92	**90.54**	0.70	0.75	0.46	**95.20**	0.79	0.94	1.00
M3b	6.50	0.22	0.01	0.43	12.60	0.17	0.02	0.21	91.73	0.29	0.42	0.61	95.87	0.11	0.30	0.17
M3c	5.12	0.94	0.24	0.78	**11.02**	0.64	0.17	0.71	91.14	0.59	0.54	0.80	95.28	0.71	0.81	1.00
M3d	4.72	0.95	0.50	0.78	11.42	0.56	0.06	0.71	90.65	0.96	0.12	0.83	93.70	0.43	0.54	0.63
M4	5.51	0.77	0.35	1.00	**9.84**	0.26	0.81	1.00	92.13	0.19	0.56	0.99	95.87	0.65	0.95	0.89
M5	4.53	0.30	0.40	0.19	11.02	0.44	0.08	0.33	91.34	0.32	0.39	1.00	**95.08**	0.97	0.92	1.00

Table 8: Empirical coverage rate (%) and p-values of two VaR backtests and MCS at the 0.1%, 0.5%, 99.5% and 99.9% quantile levels, where methods M1a-M5 are summarized in Table 6. The ECRs closest to the nominal level for each model are marked in bold.

	$\tau = 0.1\%$				$\tau = 0.5\%$				$\tau = 99.5\%$				$\tau = 99.9\%$			
	ECR	CC	DQ	MCS	ECR	CC	DQ	MCS	ECR	CC	DQ	MCS	ECR	CC	DQ	MCS
M1a	**0.11**	0.97	0.83	1.00	0.77	0.03	0.01	1.00	**99.40**	0.58	0.24	0.87	99.74	0.01	0.00	1.00
M1b	0.18	0.34	0.73	1.00	**0.53**	0.85	0.98	1.00	99.34	0.28	0.31	1.00	99.78	0.08	0.14	0.87
M1c	0.13	0.80	0.97	1.00	0.75	0.07	0.00	0.42	99.27	0.10	0.08	0.80	**99.80**	0.18	0.14	0.98
M1d	0.15	0.56	0.44	1.00	0.71	0.14	0.00	0.26	99.29	0.14	0.17	0.98	**99.80**	0.18	0.11	0.24
M2a	0.15	0.56	0.68	1.00	**0.49**	0.89	1.00	0.13	99.40	0.58	0.57	0.40	99.78	0.08	0.21	0.98
M2b	**0.13**	0.80	0.77	0.71	0.73	0.10	0.17	0.68	**99.45**	0.78	0.53	0.72	**99.80**	0.18	0.45	0.72
M2c	0.22	0.08	0.01	0.00	**0.49**	0.89	0.09	0.01	99.40	0.00	0.00	0.01	99.76	0.04	0.00	0.00
M2d	0.20	0.18	0.09	0.92	0.84	0.01	0.00	0.09	99.29	0.14	0.03	0.63	99.71	0.01	0.00	0.78
M3a	0.29	0.00	0.00	0.57	0.66	0.15	0.00	0.65	**99.45**	0.78	0.32	1.00	99.76	0.04	0.02	1.00
M3b	0.20	0.18	0.50	1.00	**0.62**	0.47	0.54	1.00	99.36	0.37	0.43	1.00	**99.78**	0.08	0.00	0.70
M3c	**0.13**	0.80	0.79	1.00	0.73	0.10	0.38	0.83	99.25	0.07	0.13	0.58	**99.78**	0.18	0.38	0.97
M3d	**0.13**	0.80	0.92	1.00	0.64	0.08	0.50	1.00	**99.45**	0.59	0.78	1.00	99.76	0.04	0.04	1.00
M4	0.71	0.00	0.00	0.06	1.28	0.00	0.00	0.09	98.70	0.00	0.00	0.03	99.32	0.00	0.00	0.00
M5	1.79	0.00	0.00	0.00	1.15	0.00	0.00	0.00	99.16	0.01	0.00	0.05	99.23	0.00	0.00	0.00