

SIMULTANEOUS CONFIDENCE BANDS FOR CONDITIONAL VALUE-AT-RISK AND EXPECTED SHORTFALL

SHUO LI

Tianjin University of Finance and Economics

LIUHUA PENG

The University of Melbourne

XIAOJUN SONG

Peking University

Conditional value-at-risk (CVaR) and conditional expected shortfall (CES) are widely adopted risk measures which help monitor potential tail risk while adapting to evolving market information. In this paper, we propose an approach to constructing simultaneous confidence bands (SCBs) for tail risk as measured by CVaR and CES, with the confidence bands uniformly valid for a set of tail levels. We consider one-sided tail risk (downside or upside tail risk) as well as relative tail risk (the ratio of upside to downside tail risk). A general class of location-scale models with heavy-tailed innovations is employed to filter out the return dynamics. Then, CVaR and CES are estimated with the aid of extreme value theory. In the asymptotic theory, we consider two scenarios: (i) the extreme scenario that allows for extrapolation beyond the range of the available data and (ii) the intermediate scenario that works exclusively in the case where the available data are adequate relative to the tail level. For finite-sample implementation, we propose a novel bootstrap procedure to circumvent the slow convergence rates of the SCBs as well as infeasibility of approximating the limiting distributions. A series of Monte Carlo simulations confirm that our approach works well in finite samples.

1. INTRODUCTION

When investing in a risky asset, investors assuming a long position are exposed to downside tail risk due to a sharp fall of the asset price, and investors assuming a short position are exposed to upside tail risk due to a sharp rise of the asset price.

We thank the Editor (Professor Peter Phillips), the Co-Editor (Professor Dennis Kristensen), and two anonymous referees for constructive comments which led to great improvements of the paper. Li acknowledges financial support from the National Natural Science Foundation of China (Grant No. 11801399). Song acknowledges financial support from the National Science Foundation of China (Grant No. 71973005). Li dedicates this paper to his beloved wife Zhang Xian and their new son Li Guxin. Address correspondence to Liuhua Peng, School of Mathematics and Statistics, The University of Melbourne, Parkville, Victoria 3010, Australia; e-mail: liuhua.peng@unimelb.edu.au.

© The Author(s), 2022. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted re-use, distribution, and reproduction in any medium, provided the original work is properly cited.

1009

In the practice of risk management and investment evaluation, it is important to monitor potential tail risk while adapting to evolving market information, particularly the volatility dynamics (McNeil and Frey, 2000, pp. 272–273). A leading vehicle that aids in fulfilling this task is the conditional tail risk measure, among whose subordinates the most prominent are conditional value-at-risk (CVaR) and conditional expected shortfall (CES).¹ Let $\{R_t\}$ be a sequence of returns on a certain risky asset, and let \mathcal{I}_t be the historical information available up to time t . Based on observed data from time 1 through n , at a tail level τ which is close to 0, the one-period ahead upside CVaR and CES are, respectively, defined as

$$U\text{-CVaR}(\tau) = Q_{R_{n+1}}(1 - \tau | \mathcal{I}_n) \text{ and } U\text{-CES}(\tau) = E[R_{n+1} | R_{n+1} > U\text{-CVaR}(\tau), \mathcal{I}_n],$$

where $Q_{R_{n+1}}(\cdot | \mathcal{I}_n)$ is the conditional quantile function of R_{n+1} given \mathcal{I}_n . Similarly, the one-period ahead downside CVaR and CES at the tail level τ are, respectively, defined as

$$D\text{-CVaR}(\tau) = Q_{-R_{n+1}}(1 - \tau | \mathcal{I}_n) \text{ and } D\text{-CES}(\tau) = E[-R_{n+1} | -R_{n+1} > D\text{-CVaR}(\tau), \mathcal{I}_n],$$

where $Q_{-R_{n+1}}(\cdot | \mathcal{I}_n)$ is the conditional quantile function of $-R_{n+1}$ given \mathcal{I}_n . In addition, to facilitate a comparison of downside and upside tail risk, we define, respectively, the one-period ahead relative CVaR and CES at the tail level τ as

$$R\text{-CVaR}(\tau) = \frac{U\text{-CVaR}(\tau)}{D\text{-CVaR}(\tau)} \text{ and } R\text{-CES}(\tau) = \frac{U\text{-CES}(\tau)}{D\text{-CES}(\tau)}.$$

In this paper, we propose an approach to constructing simultaneous confidence bands (SCBs) for upside, downside, and relative CVaR and CES, with the SCBs uniformly valid for a set of tail levels. The construction of confidence intervals for one-sided CVaR and CES at a single tail level has been studied in several works (e.g., Chan et al., 2007; Martins-Filho, Yao, and Torero, 2018; Hoga, 2019a). Compared with the confidence interval at a single tail level, simultaneous inference based on the SCB is more appealing both practically and theoretically. First, risk measurement at a single tail level gives a limited view of the potential loss and hence can be misleading, particularly during turbulent periods when the market risk increases (e.g., Francq and Zakoïan, 2016). Therefore, it is desirable to use multiple tail levels for robust analysis and overall control of the tail risk. For instance, although it is common for financial institutions to specify their trading limits as the 1% daily CVaR or CES of their trading books, jointly considering the tail levels in [0.5%, 1%] helps understand the potential loss amid more extreme market downturns. Second, from the perspective of statistical inference, one cannot base joint inference for multiple tail levels on a direct combination of separate

¹In this paper, we focus on the *conditional* tail risk measure given historical information, in contrast to the *unconditional* tail risk measure. A main distinction between conditional and unconditional risk measures lies in that conditional risk measures are mainly for daily monitoring of market risk and unconditional risk measures are suitable for longer-term risk management (see, for example, (McNeil and Frey, 2000, p. 272) and Hoga (2019b, p. 590)).

inference at each tail level, because this incurs the multiplicity effect that leads to deflated coverage rates of confidence intervals (e.g., Romano, Shaikh, and Wolf, 2018). On the contrary, the SCB is immune to the multiplicity effect and asymptotically targets the nominal confidence level.

We model the dynamics of R_t via a semiparametric location-scale model, in which the conditional mean and variance are parametrically specified, but the innovation distribution is unspecified. The key assumption concerning the innovation distribution is that it has Pareto-type heavy tails. We first fit the location-scale model to obtain standardized residuals, and then employ extreme value theory (EVT) to estimate the risk measures. This type of two-stage approach is pioneered by McNeil and Frey (2000) and has now been standard in the field of conditional risk measure estimation (e.g., Chan et al., 2007; Martins-Filho et al., 2018; Hoga, 2019a, 2022).

When deriving the asymptotic results, we work under the double asymptotics, namely $n \rightarrow \infty$ and $\tau \rightarrow 0$, which is common to EVT-based asymptotic theory (e.g., de Haan and Ferreira, 2006). In our asymptotic theory, we consider two scenarios: (i) the extreme scenario that allows for extrapolation beyond the range of the available data and (ii) the intermediate scenario that works exclusively in the case where the data are adequate relative to the tail level. Both of the two scenarios have their relative merits, and when to use which depends on whether extrapolation is needed. Theoretically, when the data are adequate relative to the tail level (that is, when extrapolation is not needed), the intermediate scenario can provide a more accurate finite-sample approximation, because some terms that are neglected in the extreme scenario are recalled and hence the approximation error is reduced. This is confirmed by a set of Monte Carlo simulations in Section 5. Nonetheless, our simulations also reveal that when extrapolation is indeed necessary (that is, when the tail lies out of the range of the available data), the extreme scenario works reasonably well.

This paper makes several contributions. First, this paper provides a unifying asymptotic theory for simultaneous inference for tail risk. The literature has witnessed continued efforts to develop asymptotic theories for conditional risk measure estimators. Chan et al. (2007) and Hoga (2019a) derive asymptotic properties of estimators of CVaR and CES in the autoregressive moving average (ARMA)–generalized autoregressive conditional heteroskedastic (GARCH) model, whereas Martins-Filho et al. (2018) do so in a nonparametric location-scale model. The above papers confine their attention to one-sided tail risk (downside or upside only) and only consider tail risk measures at a single tail level. Hoga (2022) establishes a limiting theory for the distortion risk measure and expectile in a general location-scale model and considers SCBs. Again, only one-sided tail risk is considered in Hoga (2022). In addition, each of the existing studies only considers one of the two scenarios of asymptotic theory in this paper. Chan et al. (2007) and Hoga (2019a, 2022) derive their asymptotic theory in the extreme scenario, and the asymptotic theory of Martins-Filho et al. (2018) is developed under a condition that is essentially equivalent to the intermediate scenario. By incorporating the two

scenarios in a unifying framework, our theory substantially generalizes the existing studies.

Second, this paper provides a limiting theory for the relative risk measure, allowing risk managers to monitor tail risk not only in the absolute sense, but also in the relative sense. Inference for relative tail risk is relevant because it is useful to both expected utility and nonexpected utility investors in making investment decisions. The cumulative prospect theory (CPT) of Tversky and Kahneman (1992) has confirmed that nonexpected utility investors are generally risk-seeking for tail gains, but are risk-averse for tail losses. A CPT investor generally overweights the tails of the return distribution and hence prefers lottery-like, or positively skewed, assets. Therefore, our approach possesses practical relevance in that it can help CPT investors to identify assets with more tail gains than tail losses (that is, assets with more upside risk than downside risk), which can make their overall wealth more lottery-like. On the other hand, Barberis and Huang (2008) show that preference for a lottery-like asset can make the asset overpriced, and hence creates a chance for expected utility investors who can try to exploit the overpricing by assuming short positions. Nonetheless, as Barberis and Huang (2008) argue, when shorting the lottery-like asset, expected utility investors face the risk of poor short-term performance. Then, our approach would help expected utility investors to monitor the short selling strategy by quantifying the loss relative to the gain.

Third, we propose a novel bootstrap procedure for finite-sample implementation. Two implementational issues arise in employing our established theory to construct SCBs. The first issue is about the slow convergence rates of the theoretical SCBs in both the extreme and intermediate scenarios, and the other issue is due to infeasibility of approximating the limiting distributions in the intermediate scenario. To circumvent these issues, we suggest an easy-to-implement bootstrap procedure. The bootstrap procedure is based on uniform expansions of the risk measure estimators and employs the idea of multiplier bootstrap. Extensive simulations confirm that the bootstrap procedure delivers favorable finite-sample performance.

Finally, we give a formal treatment to the issue of information truncation under general conditions. Information truncation arises frequently due to the fact that, although the dynamics usually involve infinite historical information, a feasible forecast of tail risk is based on finite observations. In addition, our framework allows for flexible dynamic models. Existing studies are usually confined to linear models such as the ARMA–GARCH model (McNeil and Frey, 2000; Chan et al., 2007; Hoga, 2019a). On the contrary, our framework facilitates flexible specifications, particularly those depicting nonlinear and nonsmooth dynamics (e.g., the threshold GARCH [TGARCH] model). The flexibility is of much practical importance, as accounting for nonlinearity is often critical to the success of financial time series modeling (e.g., Tsay, 2010, Chaps. 3 and 4).

The rest of this paper is organized as follows. Section 2 presents the framework and illustrates how to estimate CVaR and CES. Section 3 establishes the asymptotic

theory. Section 4 describes the bootstrap method and justifies its validity. Section 5 contains results of a series of Monte Carlo simulations. Section 6 concludes the paper and describes future work. All mathematical proofs of main results are relegated to the Appendix. Auxiliary lemmas along with their proofs can be found in the Supplementary Material.

2. THE FRAMEWORK AND ESTIMATION OF CVAR AND CES

2.1. The Framework

Assumption 2.1. R_t follows the location-scale model:

$$R_t = m(\mathcal{I}_{t-1}, \theta_0) + \sigma(\mathcal{I}_{t-1}, \theta_0)\varepsilon_t, \quad (2.1)$$

where $m(\mathcal{I}_{t-1}, \theta_0)$ and $\sigma^2(\mathcal{I}_{t-1}, \theta_0)$, both known up to an unknown parameter θ_0 , are, respectively, the conditional mean and conditional variance of R_t given \mathcal{I}_{t-1} . The innovation ε_t is independent of \mathcal{I}_{t-1} and forms a sequence of independent and identically distributed (i.i.d.) continuous random variables with zero mean and unit variance. The information set at time t is $\mathcal{I}_{t-1} = \sigma(\varepsilon_{t-1}, \varepsilon_{t-2}, \dots; \eta_{t-1}, \eta_{t-2}, \dots)$, the σ -algebra generated by $\{\varepsilon_{t-i}\}_{i=1}^{\infty}$ and possibly additional random vectors $\{\eta_{t-i}\}_{i=1}^{\infty}$. $\{\eta_t\}_{t=-\infty}^{\infty}$ is a sequence of random vectors independent of the innovation sequence $\{\varepsilon_t\}_{t=-\infty}^{\infty}$. In addition, $m(\mathcal{I}_{t-1}, \theta_0)$ and $\sigma(\mathcal{I}_{t-1}, \theta_0)$ are measurable with respect to \mathcal{I}_{t-1} .

Our framework allows for flexible specifications. Model (2.1) naturally nests many commonly used models, such as the ARMA–GARCH model that has been widely adopted in estimation of conditional tail risk measures (e.g., Chan et al., 2007; Hoga, 2019a). In addition, our setting facilitates specifications that are nonlinear (even nonsmooth) in the information set. A typical example is the TGARCH of Glosten, Jagannathan, and Runkle (1993). Finally, our framework subsumes the popular class of volatility-in-mean models (e.g., Engle, Lilien, and Robins, 1987).

Remark 2.1. In an alternative framework, we may work with nonparametric conditional mean and variance functions. Inspired by Martins-Filho et al. (2018), suppose \mathbf{X}_t is a d -dimensional random vector which may include lagged returns $\{R_{t-\ell}\}_{\ell=1}^p$ for positive integers d and p , we may model R_t nonparametrically by

$$R_t = m(\mathbf{X}_t) + \sigma(\mathbf{X}_t)\varepsilon_t, \quad (2.2)$$

where $m(\cdot)$ and $\sigma(\cdot)$ are unknown functions, and the innovation ε_t is independent of \mathbf{X}_t . The advantage of the nonparametric model (2.2) is that it can avoid the risk of misspecification that emerges from parametric functional forms for $m(\cdot)$ and $\sigma(\cdot)$. In this paper, we focus on the parametric model in Assumption 2.1, but our theory can be extended to the nonparametric model (2.2) with appropriate technical adjustments. We refer the reader to Remark 3.1 and Lemma S2.1 in the Supplementary Material for more detailed discussions.

We also assume that the distribution of ε_t has Pareto-type heavy tails. Let $F_\varepsilon(\cdot) = \Pr(\varepsilon_t \leq \cdot)$ and $F_{-\varepsilon}(\cdot) = \Pr(-\varepsilon_t \leq \cdot)$ be, respectively, the distribution functions of ε_t and $-\varepsilon_t$. Then, assume that $1 - F_\varepsilon$ and $1 - F_{-\varepsilon}$ are both regularly varying at infinity in the sense that

$$\lim_{x \rightarrow \infty} [1 - F_\varepsilon(xy)]/[1 - F_\varepsilon(x)] = y^{-\gamma_R^{-1}}, \text{ for all } y > 0, \tag{2.3}$$

and

$$\lim_{x \rightarrow \infty} [1 - F_{-\varepsilon}(xy)]/[1 - F_{-\varepsilon}(x)] = y^{-\gamma_L^{-1}}, \text{ for all } y > 0, \tag{2.4}$$

where $\gamma_R > 0$ and $\gamma_L > 0$ are the extreme value indices associated with the right and left tails, respectively. It is worth mentioning that, under Assumption 2.1, the innovation ε_t is assumed to have unit variance. This implicitly implies that $\gamma_R < 1/2$ and $\gamma_L < 1/2$.

The motivation for assuming heavy-tailed innovations is twofold. First, in the context of financial time series modeling, models with the classical normal innovation are frequently found to be inadequate (e.g., Hamilton, 1994, Chap. 21) and the literature has witnessed increasing popularity of models with heavy-tailed innovations such as the Student’s t (e.g., Bollerslev, 1987) and the skewed t (e.g., Theodossiou and Savva, 2015) variables. Second, in the context of conditional risk measure estimation, McNeil and Frey (2000) show by backtesting that procedures based on heavy-tailed innovations deliver better risk measure estimates than methods that ignore the heavy-tailed feature. Since McNeil and Frey (2000), heavy-tailed innovations have been pervasive in subsequent studies (e.g., Chan et al., 2007; Martins-Filho et al., 2018; Hoga, 2019a).

Based on the model structure (2.1), the conditional tail risk measures become

$$\text{U-CVaR}(\tau) = m(\mathcal{I}_n, \theta_0) + \sigma(\mathcal{I}_n, \theta_0)Q_\varepsilon(1 - \tau), \tag{2.5}$$

$$\text{U-CES}(\tau) = m(\mathcal{I}_n, \theta_0) + \sigma(\mathcal{I}_n, \theta_0)E[\varepsilon | \varepsilon > Q_\varepsilon(1 - \tau)], \tag{2.6}$$

$$\text{D-CVaR}(\tau) = -m(\mathcal{I}_n, \theta_0) - \sigma(\mathcal{I}_n, \theta_0)Q_\varepsilon(\tau), \tag{2.7}$$

and

$$\text{D-CES}(\tau) = -m(\mathcal{I}_n, \theta_0) - \sigma(\mathcal{I}_n, \theta_0)E[\varepsilon | \varepsilon < Q_\varepsilon(\tau)], \tag{2.8}$$

where $Q_\varepsilon(\cdot)$ is the quantile function of ε_t .

Suppose, for type $\in \{\text{U, D, R}\}$, that we have estimators $\widehat{\text{type-CVaR}}$ and $\widehat{\text{type-CES}}$ for, respectively, type-CVaR and type-CES. For a tail region $[\tau_l, \tau_u]$ with $0 < \tau_l \leq \tau_u$, we consider the maximum absolute log-ratios

$$\sup_{\tau \in [\tau_l, \tau_u]} \left| \log \left[\frac{\widehat{\text{type-CVaR}}(\tau)}{\text{type-CVaR}(\tau)} \right] \right| \text{ and } \sup_{\tau \in [\tau_l, \tau_u]} \left| \log \left[\frac{\widehat{\text{type-CES}}(\tau)}{\text{type-CES}(\tau)} \right] \right|. \tag{2.9}$$

It is the task of Section 3 to establish the asymptotic distributions of the maximum absolute log-ratios, which enable us to construct SCBs for the risk measures.

2.2. Estimation of CVaR and CES

In this section, we demonstrate how to employ EVT to estimate CVaR and CES for a given tail level, based on observed data from time 1 to n . The key step is to obtain estimators of the quantiles $Q_\varepsilon(1 - \tau)$ and $Q_\varepsilon(\tau)$ as well as estimators of the conditional tail means $E[\varepsilon|\varepsilon > Q_\varepsilon(1 - \tau)]$ and $E[\varepsilon|\varepsilon < Q_\varepsilon(\tau)]$. We note that simple nonparametric estimators of the above quantities based on the empirical distribution function have also been considered, and their asymptotic properties have been studied (see Gao and Song, 2008).

Suppose that there is a consistent estimator $\hat{\theta}$ of θ_0 . Let $\tilde{\mathcal{I}}_{t-1}$ be the truncated information set, which is generated by feasible information up to time $t - 1$. The truncation is necessary when the information set relies on infinite past observations. For example, when $\mathcal{I}_{t-1} = \{R_{t-1}, R_{t-2}, \dots, R_1, R_0, R_{-1}, \dots\}$, the truncated information set is $\tilde{\mathcal{I}}_{t-1} = \{R_{t-1}, R_{t-2}, \dots, R_1\}$. We obtain the standardized residuals $\hat{\varepsilon}_t = (R_t - m(\tilde{\mathcal{I}}_{t-1}, \hat{\theta})) / \sigma(\tilde{\mathcal{I}}_{t-1}, \hat{\theta})$, for $t = 1, \dots, n$. For $d_n < n$ such that $d_n \rightarrow \infty$ as $n \rightarrow \infty$, we discard the residuals for $t < d_n$ and work with $\hat{\varepsilon}_t$ for $t = d_n, \dots, n$. The discarding eliminates the effect of information truncation. Similar discarding is conducted in Chan et al. (2007) and Hoga (2019a) for the ARMA-GARCH model. Here, we generalize their treatments to general models.

The following estimation procedure is standard in EVT-based tail estimation (see Hoga (2019a) for a recent example and de Haan and Ferreira (2006) for a comprehensive documentation). Denote by F_ε^\leftarrow the left continuous inverse of F_ε . Then, for $x > 0$, the $(1 - 1/x)$ -quantile of F_ε is $Q_\varepsilon(1 - 1/x) \equiv U_\varepsilon(x) = F_\varepsilon^\leftarrow(1 - 1/x)$. By Theorem 1.2.1 and Corollary 1.2.10 of de Haan and Ferreira (2006), (2.3) is equivalent to

$$\lim_{x \rightarrow \infty} \frac{U_\varepsilon(xy)}{U_\varepsilon(x)} = y^{\gamma_R}, \text{ for all } y > 0. \tag{2.10}$$

Similarly, let $F_{-\varepsilon}^\leftarrow$ be the left continuous inverse of $F_{-\varepsilon}$ and $U_{-\varepsilon}(x) \equiv F_{-\varepsilon}^\leftarrow(1 - 1/x)$, for $x > 0$, and (2.4) is equivalent to

$$\lim_{x \rightarrow \infty} \frac{U_{-\varepsilon}(xy)}{U_{-\varepsilon}(x)} = y^{\gamma_L}, \text{ for all } y > 0. \tag{2.11}$$

We first consider estimation of the right tail, namely $Q_\varepsilon(1 - \tau)$ and $E[\varepsilon|\varepsilon > Q_\varepsilon(1 - \tau)]$. For this purpose, we employ the integer sequence $k_1 \equiv k_{1,n} \rightarrow \infty$ with $1 \leq k_1 \leq n$ and $k_1/n \rightarrow 0$ as $n \rightarrow \infty$. Note that, for a small τ , (2.10) implies the following approximation:

$$Q_\varepsilon(1 - \tau) = U_\varepsilon\left(\frac{1}{\tau}\right) \approx U_\varepsilon\left(\frac{n}{k_1}\right) \left(\frac{n\tau}{k_1}\right)^{-\gamma_R}. \tag{2.12}$$

Then, an estimator of $Q_\varepsilon(1 - \tau)$ can be obtained based on estimators of $U_\varepsilon(n/k_1)$ and γ_R . Let $\hat{\varepsilon}_{(k)}$ be the $(k + 1)$ -largest value of $\{\hat{\varepsilon}_i\}_{i=d_n}^n$, for $k = 0, 1, \dots, n - d_n$. A suitable estimator of $U_\varepsilon(n/k_1)$ is $\hat{\varepsilon}_{(k_1)}$, which is usually known as the intermediate order statistic (de Haan and Ferreira, 2006, Chap. 2). For estimation of γ_R , we concentrate on the prevalent Hill's (1975) estimator

$$\hat{\gamma}_R = \frac{1}{k_1} \sum_{i=0}^{k_1} \log \left(\frac{\hat{\varepsilon}_{(i)}}{\hat{\varepsilon}_{(k_1)}} \right).$$

Plugging $\hat{\varepsilon}_{(k_1)}$ and $\hat{\gamma}_R$ into (2.12) delivers the following estimator of $Q_\varepsilon(1 - \tau)$:

$$\hat{Q}_\varepsilon(1 - \tau) = \hat{\varepsilon}_{(k_1)} \left(\frac{n\tau}{k_1} \right)^{-\hat{\gamma}_R}. \tag{2.13}$$

To estimate $E[\varepsilon | \varepsilon > Q_\varepsilon(1 - \tau)]$, we invoke Proposition 4.1 of Pan, Leng, and Hu (2013), which implies that $\lim_{\tau \rightarrow 0} \frac{E[\varepsilon | \varepsilon > Q_\varepsilon(1 - \tau)]}{Q_\varepsilon(1 - \tau)} = \frac{1}{1 - \hat{\gamma}_R}$. Since τ is close to 0, the above equation motivates the following estimator of $E[\varepsilon | \varepsilon > Q_\varepsilon(1 - \tau)]$:

$$\hat{E}[\varepsilon | \varepsilon > Q_\varepsilon(1 - \tau)] = \frac{\hat{Q}_\varepsilon(1 - \tau)}{1 - \hat{\gamma}_R}. \tag{2.14}$$

Estimation of the left tail proceeds in the same way as what is done for the right tail. Note that $U_{-\varepsilon}(1/\tau) = -Q_\varepsilon(\tau)$. Then, estimation of $Q_\varepsilon(\tau)$ is equivalent to estimation of $-U_{-\varepsilon}(1/\tau)$. For another integer sequence $k_2 \equiv k_{2,n} \rightarrow \infty$ with $1 \leq k_2 \leq n$ and $k_2/n \rightarrow 0$ as $n \rightarrow \infty$, following similar arguments leading to (2.13) and (2.14), we have the estimator $\hat{Q}_\varepsilon(\tau) = \hat{\varepsilon}_{(n-k_2-d_n)} \left(\frac{n\tau}{k_2} \right)^{-\hat{\gamma}_L}$ of $Q_\varepsilon(\tau)$ and the estimator $\hat{E}[\varepsilon | \varepsilon < Q_\varepsilon(\tau)] = \frac{\hat{Q}_\varepsilon(\tau)}{1 - \hat{\gamma}_L}$ of $E[\varepsilon | \varepsilon < Q_\varepsilon(\tau)]$, where $\hat{\gamma}_L$ is the Hill estimator of γ_L based on $\{-\hat{\varepsilon}_i\}$ and is written as $\hat{\gamma}_L = \frac{1}{k_2} \sum_{i=0}^{k_2} \log \left(\frac{-\hat{\varepsilon}_{(n-i-d_n)}}{-\hat{\varepsilon}_{(n-k_2-d_n)}} \right) = \frac{1}{k_2} \sum_{i=0}^{k_2} \log \left(\frac{\hat{\varepsilon}_{(n-i-d_n)}}{\hat{\varepsilon}_{(n-k_2-d_n)}} \right)$.

Plugging the above estimators into (2.5)–(2.8), we have estimators of the risk measures

$$\widehat{\text{U-CVaR}}(\tau) = m(\tilde{\mathcal{I}}_n, \hat{\theta}) + \sigma(\tilde{\mathcal{I}}_n, \hat{\theta}) \hat{Q}_\varepsilon(1 - \tau),$$

$$\widehat{\text{U-CES}}(\tau) = m(\tilde{\mathcal{I}}_n, \hat{\theta}) + \sigma(\tilde{\mathcal{I}}_n, \hat{\theta}) \hat{E}[\varepsilon | \varepsilon > Q_\varepsilon(1 - \tau)],$$

$$\widehat{\text{D-CVaR}}(\tau) = -m(\tilde{\mathcal{I}}_n, \hat{\theta}) - \sigma(\tilde{\mathcal{I}}_n, \hat{\theta}) \hat{Q}_\varepsilon(\tau),$$

and

$$\widehat{\text{D-CES}}(\tau) = -m(\tilde{\mathcal{I}}_n, \hat{\theta}) - \sigma(\tilde{\mathcal{I}}_n, \hat{\theta}) \hat{E}[\varepsilon | \varepsilon < Q_\varepsilon(\tau)].$$

And apparently, we have estimators of the relative risk measures

$$\widehat{\text{R-CVaR}}(\tau) = \frac{\widehat{\text{U-CVaR}}(\tau)}{\widehat{\text{D-CVaR}}(\tau)} \text{ and } \widehat{\text{R-CES}}(\tau) = \frac{\widehat{\text{U-CES}}(\tau)}{\widehat{\text{D-CES}}(\tau)}.$$

3. ASYMPTOTIC THEORY

3.1. Assumptions

Assumption 3.1. The estimator $\hat{\theta}$ of θ_0 satisfies $\|\hat{\theta} - \theta_0\| = O_p(n^{-\nu_0/2})$ for some positive ν_0 .

Assumption 3.1 states the convergence rate of $\hat{\theta}$. It is standard to assume $n^{1/2}$ -consistency (Bai and Ng, 2001; Hoga, 2019a), which is nested by Assumption 3.1 with $\nu_0 = 1$. In addition, Assumption 3.1 allows for estimators with slower convergence rates, thereby providing more practical flexibility. Take the GARCH model for example. For the conventional Gaussian quasi-maximum likelihood estimator (QMLE), a standard condition to ensure $n^{1/2}$ -consistency is the moment condition $E(\varepsilon_t^4) < \infty$. When the moment condition is suspicious, Assumption 3.1 can still be satisfied by, for example, the estimator of Hill (2015), which is $n^{\nu_0/2}$ -consistent with ν_0 as close to 1 as desired.

Assumption 3.2. Let Θ_0 be a neighborhood of θ_0 .

- (i) $\sup_{\theta \in \Theta_0} m(\mathcal{I}_n, \theta) = O_p(1)$.
- (ii) $\inf_{\theta \in \Theta_0} \sigma(\mathcal{I}_{t-1}, \theta) > c$ for some $c > 0$ for all $t = d_n, \dots, n+1$.
- (iii) Both $m(\mathcal{I}_{t-1}, \theta)$ and $\sigma(\mathcal{I}_{t-1}, \theta)$ are differentiable with respect to θ in Θ_0 for all $t = d_n, \dots, n+1$.
- (iv) $E[\sup_{\theta \in \Theta_0} \|\partial m(\mathcal{I}_{t-1}, \theta)/\partial \theta\|]^{v_1} < \infty$ and $E[\sup_{\theta \in \Theta_0} \|\partial \sigma(\mathcal{I}_{t-1}, \theta)/\partial \theta\|]^{v_1} < \infty$ for some $v_1 > 2/\nu_0$ for all $t = d_n, \dots, n+1$, where ν_0 is defined in Assumption 3.1.

Assumption 3.2(i) and (ii) is standard in estimation of heteroskedastic time series models. Assumption 3.2(iii) imposes common differentiability conditions. Assumption 3.2(iv) is similar to Assumptions A2 and A3 of Bai and Ng (2001) and is satisfied by some commonly used models with suitable moment conditions on ε_t . For example, Assumption 3.2(iv) is easily shown to hold for the ARMA(1, 1) model when $E|\varepsilon_t|^{v_1} < \infty$ for some $v_1 > 2/\nu_0$.

Assumption 3.3. (i) For each $\theta \in \Theta_0$, $m(\tilde{\mathcal{I}}_{t-1}, \theta)$ and $\sigma(\tilde{\mathcal{I}}_{t-1}, \theta)$ are measurable with respect to \mathcal{I}_{t-1} .

(ii) $\sup_{\theta \in \Theta_0} |m(\tilde{\mathcal{I}}_n, \theta) - m(\mathcal{I}_n, \theta)| = O_p(n^{-\nu_0/2})$ and $\sup_{\theta \in \Theta_0} |\sigma(\tilde{\mathcal{I}}_n, \theta) - \sigma(\mathcal{I}_n, \theta)| = O_p(n^{-\nu_0/2})$, where ν_0 is defined in Assumption 3.1.

(iii) $\sum_{t=d_n}^n E[\sup_{\theta \in \Theta_0} |m(\tilde{\mathcal{I}}_{t-1}, \theta) - m(\mathcal{I}_{t-1}, \theta)|] = o(1)$ and $\sum_{t=d_n}^n E[\sup_{\theta \in \Theta_0} |\sigma(\tilde{\mathcal{I}}_{t-1}, \theta) - \sigma(\mathcal{I}_{t-1}, \theta)|] = o(1)$.

Assumption 3.3(i) is common. Assumption 3.3(ii) and (iii) is used to control the effect of information truncation so that the truncation does not affect the asymptotic theory (see Bai and Ng (2001) for similar assumptions). In particular, Assumptions 3.3(iii) means that the cumulative effect of information truncation is asymptotically negligible. We note that the effect of discarding the residuals for $t < d_n$ matters. If there is no discarding (that is, the cumulative effect starts from $t = 1$ rather than $t = d_n$), we generally have $\sum_{t=1}^n E[\sup_{\theta \in \Theta_0} |m(\tilde{\mathcal{I}}_{t-1}, \theta) - m(\mathcal{I}_{t-1}, \theta)|] = O(1)$

and $\sum_{t=1}^n E[\sup_{\theta \in \Theta_0} |\sigma(\tilde{\mathcal{I}}_{t-1}, \theta) - \sigma(\mathcal{I}_{t-1}, \theta)|] = O(1)$, and thus the cumulative effect of truncation is no longer negligible. Assumptions 3.2 and 3.3 are high-level conditions, and it takes effort to verify them for specific models. Appendix A.3 collects primitive conditions for them in the ARMA–GARCH, the TGARCH, and the GARCH-in-mean models.

Remark 3.1. As discussed in Remark 2.1, it is possible to extend our work to the nonparametric model (2.2). Let $\hat{m}(\cdot)$ and $\hat{\sigma}(\cdot)$ be, respectively, estimators of $m(\cdot)$ and $\sigma(\cdot)$, and let $\hat{\varepsilon}_t^* = (R_t - \hat{m}(\mathbf{X}_t))/\hat{\sigma}(\mathbf{X}_t)$, for $t = 1, \dots, n$, be the corresponding standardized nonparametric residuals. The $\hat{\varepsilon}_t^*$ are employed to estimate CVaR and CES in the same manner as in Section 2.2. Inspecting the technical derivations of our theoretical results, we find that the only adjustment needed to obtain the theoretical results for nonparametric models is to prove that the results of Lemma S1.7 in the Supplementary Material hold for $\hat{\varepsilon}_t^*$. This lemma quantifies the difference between the tail empirical distribution of $\{\hat{\varepsilon}_t^*\}$ and that of $\{\varepsilon_t\}$. This can be achieved by assuming some uniform convergence conditions on the estimators $\hat{m}(\cdot)$ and $\hat{\sigma}(\cdot)$, for example, $\sup_{\mathbf{x} \in \mathcal{G}} |\hat{m}(\mathbf{x}) - m(\mathbf{x})| = O_p(n^{-\nu_0/2})$ and $\sup_{\mathbf{x} \in \mathcal{G}} |\hat{\sigma}(\mathbf{x}) - \sigma(\mathbf{x})| = O_p(n^{-\nu_0/2})$, for some $\nu_0 > 0$, where \mathcal{G} is the support of \mathbf{X}_t . Fortunately, these conditions have been justified by Martins-Filho et al. (2018, Lems. 3 and 4 and Cor. 1), where $m(\cdot)$ and $\sigma(\cdot)$ are estimated by local linear smoothing. For theoretical details, we refer the reader to Section S2 of the Supplementary Material.

Assumption 3.4. (i) There exist $\rho_R < 0$ and a function $A_R(\cdot)$ which is eventually positive or negative with $\lim_{x \rightarrow \infty} A_R(x) = 0$ such that

$$\lim_{x \rightarrow \infty} \frac{\frac{U_\varepsilon(xy)}{U_\varepsilon(x)} - y^{\rho_R}}{A_R(x)} = y^{\rho_R} \frac{y^{\rho_R} - 1}{\rho_R}, \text{ for all } y > 0,$$

where $k_1^{1/2} A_R(n/k_1) \rightarrow 0$ as $n \rightarrow \infty$.

(ii) There exist $\rho_L < 0$ and a function $A_L(\cdot)$ which is eventually positive or negative with $\lim_{x \rightarrow \infty} A_L(x) = 0$ such that

$$\lim_{x \rightarrow \infty} \frac{\frac{U_{-\varepsilon}(xy)}{U_{-\varepsilon}(x)} - y^{\rho_L}}{A_L(x)} = y^{\rho_L} \frac{y^{\rho_L} - 1}{\rho_L}, \text{ for all } y > 0,$$

where $k_2^{1/2} A_L(n/k_2) \rightarrow 0$ as $n \rightarrow \infty$.

Assumption 3.4(i) and (ii) provides second-order controls over the quality of the approximations in (2.10) and (2.11), respectively. This type of second-order condition has been commonly employed to develop asymptotic results in tail estimation (see, e.g., de Haan and Ferreira, 2006; Chan et al., 2007; Einmahl, de Haan, and Zhou, 2016; Hoga, 2019a). A commonly employed class of distributions that satisfies Assumption 3.4 is the Hall class of heavy-tailed distribution (Hall, 1982). For the Hall class, the distribution function $F(x)$ can be expanded as

$$1 - F(x) = Cx^{-\beta_1} [1 + Dx^{-\beta_2} + o(x^{-\beta_2})], \text{ as } x \rightarrow \infty,$$

where $C > 0$, $\beta_1 > 0$, $\beta_2 > 0$, and D is a real number. Then, Assumption 3.4 holds with $\gamma = 1/\beta_1$ and $\rho = -\beta_2/\beta_1$. A typical member of the Hall class is the Student's t distribution (sstd). The t distribution with the degree of freedom ν belongs to the Hall class with $\beta_1 = \nu$ and $\beta_2 = 2$, which implies that both of its left and right tails satisfy Assumption 3.4 with $\gamma_R = \gamma_L = 1/\nu$ and $\rho_R = \rho_L = -2/\nu$.

Assumption 3.5. (i) $\tau_l \leq \tau_u$ satisfies $\tau_l \rightarrow 0$ as $n \rightarrow \infty$ and $\tau_u = \Delta \tau_l$ for some constant $\Delta \geq 1$. (ii) There exist two constants $c_0 > 0$ and $c_1 \geq 0$ such that $\lim_{n \rightarrow \infty} k_1/k_2 = c_0$ and $\lim_{n \rightarrow \infty} n\tau_l/k_1 = c_1$. (iii) $\lim_{n \rightarrow \infty} k_1^{-1/2} \log(n\tau_l/k_1) = 0$ and $k_1 = o(n^{\delta_0})$ for some $0 < \delta_0 < \nu_0$, where ν_0 is defined in Assumption 3.1. (iv) $d_n = o(k_1^{1/2})$.

Assumption 3.5 states the orders of the effective sample sizes k_1 and k_2 as well as the order of the discarded sample size d_n . In particular, the conditions $\lim_{n \rightarrow \infty} k_1^{-1/2} \log(n\tau_l/k_1) = 0$ and $k_1 = o(n^{\delta_0})$, for some $0 < \delta_0 < \nu_0$, ensure that the convergence rate of $\hat{\theta}$ is faster than that of the estimators of CVaR and CES, so that the estimation uncertainty of $\hat{\theta}$ does not affect the limiting behaviors of the estimators of CVaR and CES. In addition, it is necessary to make more discussions on the condition $k_1 = o(n^{\delta_0})$, which requires that only a limited portion of the sample can be used for estimation so as to ignore the estimation error contained in $\hat{\theta}$. We note that this condition is mild when $\hat{\theta}$ is $n^{1/2}$ -consistent (that is, $\nu_0 = 1$), because EVT always requires that $k_1/n \rightarrow 0$. When ν_0 is small, however, a potential issue is that fewer data can be employed and the estimation error of $\hat{\theta}$ is large. We investigate the severity of this issue through simulations in Section 5. The simulation results indicate that, for a reasonable range of k_1 and k_2 , the estimation error of $\hat{\theta}$ does not have evident influence on the finite-sample performance of our approach.

3.2. Two Scenarios of Asymptotics

The following three propositions serve as the basis of our theoretical findings. In Propositions 3.1 and 3.2, we only present the results for upside CVaR and CES. The results for downside CVaR and CES are similar and can be found in Propositions A.1 and A.2 in Appendix A.1.

PROPOSITION 3.1. *Under Assumptions 2.1 and 3.1–3.5, we have, uniformly in $\tau \in [\tau_l, \tau_u]$,*

$$\frac{\widehat{U-CVaR}(\tau)}{U-CVaR(\tau)} - 1 = \frac{\hat{Q}_\varepsilon(1-\tau)}{Q_\varepsilon(1-\tau)} - 1 + o_p(k_1^{-1/2} \log(n\tau/k_1)), \text{ and}$$

$$\frac{\widehat{U-CES}(\tau)}{U-CES(\tau)} - 1 = \frac{\hat{E}[\varepsilon | \varepsilon > Q_\varepsilon(1-\tau)]}{E[\varepsilon | \varepsilon > Q_\varepsilon(1-\tau)]} - 1 + o_p(k_1^{-1/2} \log(n\tau/k_1)).$$

PROPOSITION 3.2. *Under Assumptions 2.1 and 3.1–3.5, we have, uniformly in $\tau \in [\tau_l, \tau_u]$,*

$$\frac{\hat{Q}_\varepsilon(1-\tau)}{Q_\varepsilon(1-\tau)} - 1 = \left[\frac{\hat{\varepsilon}_{(k_1)}}{U_\varepsilon(n/k_1)} - 1 \right] + \log\left(\frac{k_1}{n\tau}\right) (\hat{\gamma}_R - \gamma_R) + o_p(k_1^{-1/2} \log(n\tau/k_1)), \text{ and}$$

$$\frac{\hat{E}[\varepsilon|\varepsilon > Q_\varepsilon(1-\tau)]}{E[\varepsilon|\varepsilon > Q_\varepsilon(1-\tau)]} - 1 = \left[\frac{\varepsilon_{(k_1)}}{U_\varepsilon(n/k_1)} - 1 \right] + \left[\log\left(\frac{k_1}{n\tau}\right) + \frac{1}{1-\gamma_R} \right] (\hat{\gamma}_R - \gamma_R) + o_p(k_1^{-1/2} \log(n\tau/k_1)).$$

PROPOSITION 3.3. *Under Assumptions 2.1 and 3.1–3.5, as $n \rightarrow \infty$,*

$$\left(\begin{array}{c} k_1^{1/2} \gamma_R^{-1} [\hat{\varepsilon}_{(k_1)}/U_\varepsilon(n/k_1) - 1] \\ k_1^{1/2} \gamma_R^{-1} (\hat{\gamma}_R - \gamma_R) \\ k_2^{1/2} \gamma_L^{-1} [-\hat{\varepsilon}_{(n-k_2-d_n)}/U_{-\varepsilon}(n/k_2) - 1] \\ k_2^{1/2} \gamma_L^{-1} (\hat{\gamma}_L - \gamma_L) \end{array} \right)$$

is asymptotically four-dimensional standard normal.

Proposition 3.1 indicates that the estimation uncertainty of the risk measures are dominated by the estimation uncertainty of the quantiles and conditional tail means of ε_t . In particular, the estimation effect of the model parameter does not affect the limiting behaviors of risk measure estimators up to the dominating order. Proposition 3.1 generalizes similar results established by Chan et al. (2007) and Hoga (2019a) in ARMA–GARCH models to general dynamic models and also enhances them to obtain uniform results. Proposition 3.2 moves on to decompose the estimation uncertainty of the quantiles and conditional tail means of ε_t into two parts, with one part concerning the order statistics and the other concerning the extreme value index estimators. Finally, Proposition 3.3 indicates that the order statistics and the extreme value index estimators have the same convergence rates and are asymptotically independent.

Contemplating the decompositions in Proposition 3.2, we observe that $\hat{\gamma}_R - \gamma_R$ is scaled by $\log(k_1/(n\tau))$. Consequently, the order of $\log(k_1/(n\tau))$ plays a key role in the limiting behaviors. Take the decomposition of $\frac{\hat{Q}_\varepsilon(1-\tau)}{Q_\varepsilon(1-\tau)} - 1$ for example. Proposition 3.3 indicates that the convergence rates of $\frac{\hat{\varepsilon}_{(k_1)}}{U_\varepsilon(n/k_1)} - 1$ and $\hat{\gamma}_R - \gamma_R$ are the same. If $c_1 = 0$, (that is, $\lim_{n \rightarrow \infty} k_1/(n\tau) = \infty$), then the limiting behavior of $\frac{\hat{Q}_\varepsilon(1-\tau)}{Q_\varepsilon(1-\tau)} - 1$ is dominated by $\log\left(\frac{k_1}{n\tau}\right) (\hat{\gamma}_R - \gamma_R)$ because $\frac{\hat{\varepsilon}_{(k_1)}}{U_\varepsilon(n/k_1)} - 1$ is asymptotically negligible. If instead $c_1 > 0$ (that is, $\lim_{n \rightarrow \infty} k_1/(n\tau) < \infty$), then $\frac{\hat{\varepsilon}_{(k_1)}}{U_\varepsilon(n/k_1)} - 1$ is no longer asymptotically negligible and the limiting behavior of $\frac{\hat{Q}_\varepsilon(1-\tau)}{Q_\varepsilon(1-\tau)} - 1$ is dominated by both $\frac{\hat{\varepsilon}_{(k_1)}}{U_\varepsilon(n/k_1)} - 1$ and $\log\left(\frac{k_1}{n\tau}\right) (\hat{\gamma}_R - \gamma_R)$.

Based on the above discussion, we distinguish our asymptotic theory between two scenarios, depending on the magnitude of the effective sample sizes k_1 and k_2 relative to $n\tau$. The first one, *the extreme scenario*, corresponds to the case when $c_1 = 0$. In this scenario, τ is of smaller order than k_1/n , meaning that the tail to be estimated lies out of the range of the available data. The second one, *the intermediate scenario*, corresponds to the case when $c_1 > 0$. In this scenario, τ is of the same order as k_1/n , meaning that the tail level is comparable to the size of the available data.

We argue that both the extreme and intermediate scenarios bear their own appeal and when to use which depends on the specific problem. In the extreme scenario, the intuition behind the condition $c_1 = 0$ is that the starting point of extrapolation (that is, k_1/n and k_2/n) is more central relative to the level at which the tail is to be estimated (that is, τ). Accordingly, this scenario is capable of estimating very deep tails. For example, when one needs to estimate the tail risk at $\tau = 0.01\%$ but the available data are only of size 1,000, the tail to be estimated lies out of the range of the data. In this case, the extreme scenario applies, but the intermediate scenario cannot work because no reasonable k_1 is comparable to $n\tau = 0.1$.

To appreciate the relative merit of the intermediate scenario, consider the case where the sample size is 1,000 and the tail level is 1%. In this case, the size of available data is comparable to the tail level. Though theoretically the extreme scenario still applies to this case, the intermediate scenario seems more appealing. The reason is that the terms that are negligible in the extreme scenario (that is, $\frac{\hat{\varepsilon}(k_1)}{U_\varepsilon(n/k_1)} - 1$ and $\frac{-\hat{\varepsilon}(n-k_2-d_n)}{U_{-\varepsilon}(n/k_2)} - 1$) may result in approximation errors when the associated limiting distributions are used to construct SCBs. On the contrary, the neglected terms are recalled in the intermediate scenario and hence can lead to more reliable finite-sample performance. Indeed, our simulation results in Section 5 confirm that, when the available data are adequate relative to the tail level, the intermediate scenario produces more accurate coverage rates than the extreme scenario.

3.3. Main Results

The following theorem establishes uniform convergences of the maximum absolute log-ratios in the two scenarios.

THEOREM 3.1. *Suppose that Assumptions 2.1 and 3.1–3.5 hold.*

(i) *(extreme scenario, $c_1 = 0$) As $n \rightarrow \infty$,*

$$\sup_{u \in [1, \Delta]} \left| \frac{k_1^{1/2}}{\log(k_1/(n\tau))} \left\{ \log \left[\frac{\widehat{\text{type-CVaR}}(u\tau)}{\text{type-CVaR}(u\tau)} \right] \right\} \right| \rightarrow |Z_{\text{type}}| \text{ and}$$

$$\sup_{u \in [1, \Delta]} \left| \frac{k_1^{1/2}}{\log(k_1/(n\tau))} \left\{ \log \left[\frac{\widehat{\text{type-CES}}(u\tau)}{\text{type-CES}(u\tau)} \right] \right\} \right| \rightarrow |Z_{\text{type}}|$$

in distribution, where

$$Z_{type} \sim \begin{cases} \mathcal{N}(0, \gamma_R^2), & \text{if type} = U, \\ \mathcal{N}(0, c_0 \gamma_L^2), & \text{if type} = D, \\ \mathcal{N}(0, \gamma_R^2 + c_0 \gamma_L^2), & \text{if type} = R. \end{cases}$$

(ii) (intermediate scenario, $c_1 > 0$) As $n \rightarrow \infty$,

$$\sup_{u \in [1, \Delta]} \left| k_1^{1/2} \left\{ \log \left[\frac{\widehat{\text{type-CVaR}}(u\tau_l)}{\text{type-CVaR}(u\tau_l)} \right] \right\} \right| \rightarrow \sup_{u \in [1, \Delta]} |G_{1,type}(u)| \text{ and}$$

$$\sup_{u \in [1, \Delta]} \left| k_1^{1/2} \left\{ \log \left[\frac{\widehat{\text{type-CES}}(u\tau_l)}{\text{type-CES}(u\tau_l)} \right] \right\} \right| \rightarrow \sup_{u \in [1, \Delta]} |G_{2,type}(u)|$$

in distribution, where $G_{1,type}(u)$ and $G_{2,type}(u)$ are centered Gaussian processes with variance–covariance functions $\gamma_{1,type}(u_1, u_2) = \text{Cov}(G_{1,type}(u_1), G_{1,type}(u_2))$ and $\gamma_{2,type}(u_1, u_2) = \text{Cov}(G_{2,type}(u_1), G_{2,type}(u_2))$ satisfying that

$$\gamma_{1,type}(u_1, u_2) = \begin{cases} \gamma_R^2 [1 + \log(c_1 u_1) \log(c_1 u_2)], & \text{if type} = U, \\ c_0 \gamma_L^2 [1 + \log(c_0 c_1 u_1) \log(c_0 c_1 u_2)], & \text{if type} = D, \\ \gamma_R^2 [1 + \log(c_1 u_1) \log(c_1 u_2)] \\ \quad + c_0 \gamma_L^2 [1 + \log(c_0 c_1 u_1) \log(c_0 c_1 u_2)], & \text{if type} = R, \end{cases}$$

and

$$\gamma_{2,type}(u_1, u_2) = \begin{cases} \gamma_R^2 \{1 + [(1 - \gamma_R)^{-1} - \log(c_1 u_1)][(1 - \gamma_R)^{-1} - \log(c_1 u_2)]\}, & \text{if type} = U, \\ c_0 \gamma_L^2 \{1 + [(1 - \gamma_L)^{-1} - \log(c_0 c_1 u_1)][(1 - \gamma_L)^{-1} - \log(c_0 c_1 u_2)]\}, & \text{if type} = D, \\ \gamma_R^2 \{1 + [(1 - \gamma_R)^{-1} - \log(c_1 u_1)][(1 - \gamma_R)^{-1} - \log(c_1 u_2)]\} \\ \quad + c_0 \gamma_L^2 \{1 + [(1 - \gamma_L)^{-1} - \log(c_0 c_1 u_1)][(1 - \gamma_L)^{-1} - \log(c_0 c_1 u_2)]\}, & \text{if type} = R. \end{cases}$$

Remark 3.2. In the extreme scenario, CVaR and CES share the same limiting distributions for each type. This is because $\frac{\hat{\varepsilon}(k_1)}{U_\varepsilon(n/k_1)} - 1$ and $\frac{-\hat{\varepsilon}(n-k_2-dn)}{U_{-\varepsilon}(n/k_2)} - 1$ are asymptotically negligible and hence CVaR and CES have the same asymptotic expansion. Similar results are also found in Hoga (2019a).

Remark 3.3. According to Theorem 3.1(i), we can construct SCBs for the risk measures in the extreme scenario based on consistent estimators of the limiting variances. Take the relative risk measures for example. A consistent estimator of the limiting variance is $\hat{\gamma}_R^2 + k_1 \hat{\gamma}_L^2/k_2$. Hence, the asymptotic $(1 - \alpha)$ -SCBs for R-CVaR(τ) and R-CES(τ) are, respectively,

$$\begin{aligned} \widehat{\text{R-CVaR}}(\tau) \exp \left[-\Phi^{-1}(1 - \alpha/2) k_1^{-1/2} \log(k_1/(n\tau)) \sqrt{\hat{\gamma}_R^2 + k_1 \hat{\gamma}_L^2/k_2} \right] &\leq \text{R-CVaR}(\tau) \\ &\leq \widehat{\text{R-CVaR}}(\tau) \exp \left[\Phi^{-1}(1 - \alpha/2) k_1^{-1/2} \log(k_1/(n\tau)) \sqrt{\hat{\gamma}_R^2 + k_1 \hat{\gamma}_L^2/k_2} \right], \end{aligned}$$

and

$$\widehat{\text{R-CES}}(\tau) \exp \left[-\Phi^{-1}(1 - \alpha/2) k_1^{-1/2} \log(k_1/(n\tau)) \sqrt{\hat{\gamma}_R^2 + k_1 \hat{\gamma}_L^2/k_2} \right] \leq \text{R-CES}(\tau) \leq \widehat{\text{R-CES}}(\tau) \exp \left[\Phi^{-1}(1 - \alpha/2) k_1^{-1/2} \log(k_1/(n\tau)) \sqrt{\hat{\gamma}_R^2 + k_1 \hat{\gamma}_L^2/k_2} \right],$$

where $\Phi^{-1}(\cdot)$ is the quantile function of the standard normal distribution.

By incorporating the two scenarios as well as relative risk measures in a unifying theory, Theorem 3.1 substantially generalizes and unifies some existing studies. The extreme scenario is conventional in EVT-based tail estimation (e.g., de Haan and Ferreira, 2006), particularly in estimation of conditional risk measures. Chan et al. (2007) and Hoga (2019a) derive the asymptotic properties of estimators of CVaR and CES in the ARMA–GARCH model. The intermediate scenario is less popular, but is still employed in several places. Remark 4.3.7 of de Haan and Ferreira (2006) mentions a similar result in a discussion of quantile estimation. Recently, Martins-Filho et al. (2018) estimate the CVaR and CES by fitting the generalized Pareto distribution to the residuals of an estimated heteroskedastic nonparametric regression. Their asymptotic theory is developed under a condition that is essentially equivalent to the intermediate scenario of this paper. The above works confine their attention to one-sided tail risk (downside or upside only) and only consider risk measures at a single tail level.

4. BOOTSTRAP IMPLEMENTATION

Although Theorem 3.1 provides limiting distributions for the maximum absolute log-ratios, two implementational issues arise when employing the theorem to construct SCBs. The first issue is due to the slow convergence rates of the maximum absolute ratios because k_1 and k_2 are small compared to n and τ is close to 0. Consequently, even if Remark 3.3 gives the SCBs in the extreme scenario, our preliminary simulation results indicate that the coverage of such theoretical SCBs is seriously distorted. The second issue arises because it is quite hard to approximate the limiting distributions in the intermediate scenario, which are constructed from Gaussian processes that depend on the underlying model in a very complicated manner. To circumvent the two issues, we propose an easy-to-implement bootstrap procedure for finite-sample implementation.

We first discuss the bootstrap procedure for the extreme scenario. For any real number x , let $\log^> x$ take value $\log x$ if $x > 1$ and 0 otherwise. Under Assumptions 2.1 and 3.1–3.5 and for type $\in \{U, D, R\}$, the proof of Theorem 3.1 implies that

$$\frac{k_1^{1/2}}{\log(k_1/(n\tau))} \log \left[\frac{\widehat{\text{type-CVaR}}(\tau)}{\text{type-CVaR}(\tau)} \right] = \mathbb{I}(\text{type} \in \{U, R\}) k_1^{1/2} (\hat{\gamma}_R - \gamma_R) - \mathbb{I}(\text{type} \in \{D, R\}) (k_1/k_2)^{1/2} k_2^{1/2} (\hat{\gamma}_L - \gamma_L) + o_p(1),$$

and

$$\frac{k_1^{1/2}}{\log(k_1/(n\tau))} \log \left[\frac{\widehat{\text{type-CES}}(\tau)}{\text{type-CES}(\tau)} \right]$$

$$= \mathbb{I}(\text{type} \in \{U, R\}) k_1^{1/2} (\hat{\gamma}_R - \gamma_R) \mathbb{I}(\text{type} \in \{D, R\}) (k_1/k_2)^{1/2} k_2^{1/2} (\hat{\gamma}_L - \gamma_L) + o_p(1),$$

uniformly in $\tau \in [\tau_l, \tau_u]$, where $\mathbb{I}(A)$ is the indicator function of the event A which equals 1 if A occurs and 0 otherwise. It is worth mentioning that, in the extreme scenario, CVaR and CES share the same asymptotic expansions in terms of the log-ratio, although they are very different quantities. The reason is as follows. As judged from (2.14), the tail mean estimator is constructed by dividing the quantile estimator by the factor $1 - \hat{\gamma}_R$. The estimation uncertainty of $\hat{\gamma}_R$ is of order $k_1^{1/2}$, which is smaller than $k_1^{-1/2} \log(k_1/(n\tau))$ (the order of the leading term in the extreme scenario). Therefore, the estimation uncertainty of $\hat{\gamma}_R$ does not affect the limiting property up to the leading order.

PROPOSITION 4.1. *Under Assumptions 2.1 and 3.1–3.5,*

$$\hat{\gamma}_R - \gamma_R = M_R + o_p(k_1^{-1/2}) \quad \text{and} \quad \hat{\gamma}_L - \gamma_L = M_L + o_p(k_2^{-1/2}),$$

with

$$M_R = \frac{1}{k_1} \sum_{t=d_n}^n \mathbb{I}(\varepsilon_t > \varepsilon_{(k_1)}) \left(\log^> \frac{\varepsilon_t}{\varepsilon_{(k_1)}} - \gamma_R \right), \tag{4.1}$$

and

$$M_L = \frac{1}{k_2} \sum_{t=d_n}^n \mathbb{I}(\varepsilon_t < \varepsilon_{(n-k_2-d_n)}) \left(\log^> \frac{\varepsilon_t}{\varepsilon_{(n-k_2-d_n)}} - \gamma_L \right). \tag{4.2}$$

Proposition 4.1 gives linear expansions of the tail index estimators. In addition, Proposition 3.3 implies that the tail index estimators $\hat{\gamma}_R$ and $\hat{\gamma}_L$ are asymptotically independent. Our proposal is to employ the idea of the multiplier bootstrap to approximate the uncertainty characterized by components $\hat{\gamma}_R - \gamma_R$ and $\hat{\gamma}_L - \gamma_L$, by exploiting the linear expansions as well as the asymptotic independence. Specifically, suppose we have two independent sequences of multipliers $\{\xi_t\}_{t=d_n}^n$ and $\{\nu_t\}_{t=d_n}^n$, which consist of i.i.d. random variables with zero mean, unit variance, and bounded support, and are both independent of the original data. Then, multiplying the summands in (4.1) and (4.2), respectively, with $\{\xi_t\}_{t=d_n}^n$ and $\{\nu_t\}_{t=d_n}^n$, we obtain

$$M_R^* = \frac{1}{k_1} \sum_{t=d_n}^n \mathbb{I}(\hat{\varepsilon}_t > \hat{\varepsilon}_{(k_1)}) \left(\log^> \frac{\hat{\varepsilon}_t}{\hat{\varepsilon}_{(k_1)}} - \hat{\gamma}_R \right) \xi_t, \tag{4.3}$$

and

$$M_L^* = \frac{1}{k_2} \sum_{t=d_n}^n \mathbb{I}(\hat{\epsilon}_t < \hat{\epsilon}_{(n-k_2-d_n)}) \left(\log^> \frac{\hat{\epsilon}_t}{\hat{\epsilon}_{(n-k_2-d_n)}} - \hat{\gamma}_L \right) v_t, \tag{4.4}$$

which are used to mimic M_R and M_L , respectively. Then, define for type $\in \{U, D, R\}$

$$T_{\text{type}}^{\text{ex}*} = \mathbb{I}(\text{type} \in \{U, R\}) k_1^{1/2} M_R^* - \mathbb{I}(\text{type} \in \{D, R\}) (k_1/k_2)^{1/2} k_2^{1/2} M_L^*,$$

which is used to approximate $\frac{k_1^{1/2}}{\log(k_1/(n\tau))} \log \left[\frac{\widehat{\text{type-CVaR}}(\tau)}{\text{type-CVaR}(\tau)} \right]$ and $\frac{k_1^{1/2}}{\log(k_1/(n\tau))} \log \left[\frac{\widehat{\text{type-CES}}(\tau)}{\text{type-CES}(\tau)} \right]$.

Bootstrap Algorithm for the Extreme Scenario

1. Specify a proper dynamic model and estimate it to obtain the residuals $\{\hat{\epsilon}_t\}$. Compute the tail index estimates $\hat{\gamma}_R$ and $\hat{\gamma}_L$ as well as the maximum absolute log-ratios as defined in (2.9).
2. Generate i.i.d. random variables $\{\xi_t\}$ and $\{v_t\}$ from a certain distribution with zero mean and unit variance, then compute M_R^* and M_L^* to obtain $T_{\text{type}}^{\text{ex}*}$.
3. For a large integer B , repeat Step 2 B times to obtain $\{T_{\text{type}, b}^{\text{ex}*}\}_{b=1}^B$. Then, for $\alpha \in (0, 1)$, calculate $z_{\text{type}, 1-\alpha}^{\text{ex}*}$, which is the $(1 - \alpha)$ -sample quantile of $\{|T_{\text{type}, b}^{\text{ex}*}|\}_{b=1}^B$.
4. For type $\in \{U, D, R\}$, asymptotic $(1 - \alpha)$ -SCBs for type-CVaR(τ) and type-CES(τ) are, respectively,

$$\left[\widehat{\text{type-CVaR}}(\tau) \exp \left[-z_{\text{type}, 1-\alpha}^{\text{ex}*} k_1^{-1/2} \log(k_1/(n\tau)) \right], \right. \\ \left. \widehat{\text{type-CVaR}}(\tau) \exp \left[z_{\text{type}, 1-\alpha}^{\text{ex}*} k_1^{-1/2} \log(k_1/(n\tau)) \right] \right],$$

and

$$\left[\widehat{\text{type-CES}}(\tau) \exp \left[-z_{\text{type}, 1-\alpha}^{\text{ex}*} k_1^{-1/2} \log(k_1/(n\tau)) \right], \right. \\ \left. \widehat{\text{type-CES}}(\tau) \exp \left[z_{\text{type}, 1-\alpha}^{\text{ex}*} k_1^{-1/2} \log(k_1/(n\tau)) \right] \right].$$

Now, we move on to propose the bootstrap procedure for the intermediate scenario. Under Assumptions 2.1 and 3.1–3.5 and for type $\in \{U, D, R\}$, the proof of Theorem 3.1 implies that

$$k_1^{1/2} \log \left[\frac{\widehat{\text{type-CVaR}}(\tau)}{\text{type-CVaR}(\tau)} \right] \\ = \mathbb{I}(\text{type} \in \{U, R\}) \left[\log(k_1/(n\tau)) k_1^{1/2} (\hat{\gamma}_R - \gamma_R) + \mathcal{X}_R \right] \\ - \mathbb{I}(\text{type} \in \{D, R\}) \left\{ (k_1/k_2)^{1/2} \left[\log(k_2/(n\tau)) k_2^{1/2} (\hat{\gamma}_L - \gamma_L) + \mathcal{X}_L \right] \right\} + o_p(1),$$

and

$$\begin{aligned}
 & k_1^{1/2} \log \left[\frac{\widehat{\text{type-CES}}(\tau)}{\text{type-CES}(\tau)} \right] \\
 &= \mathbb{I}(\text{type} \in \{U, R\}) \left\{ \left[\log(k_1/(n\tau)) + (1 - \gamma_R)^{-1} \right] k_1^{1/2} (\hat{\gamma}_R - \gamma_R) + \mathcal{X}_R \right\} \\
 &\quad - \mathbb{I}(\text{type} \in \{D, R\}) \left\{ (k_1/k_2)^{1/2} \left[\log(k_2/(n\tau)) + (1 - \gamma_L)^{-1} \right] k_2^{1/2} (\hat{\gamma}_L - \gamma_L) + \mathcal{X}_L \right\} + o_p(1),
 \end{aligned}$$

uniformly in $\tau \in [\tau_l, \tau_u]$, where \mathcal{X}_R and \mathcal{X}_L are two independent normal random variables with zero mean and respective variances γ_R^2 and γ_L^2 , and they are both independent of $\hat{\gamma}_R - \gamma_R$ and $\hat{\gamma}_L - \gamma_L$.

Unlike the extreme scenario, we need to account for the contributions of the components \mathcal{X}_R and \mathcal{X}_L apart from those of the tail index estimators. To this end, we randomly draw observations from \mathcal{X}_R^* and \mathcal{X}_L^* , which are independent normal random distributions with zero mean and respective variances $\hat{\gamma}_R^2$ and $\hat{\gamma}_L^2$, and are independent of $\{\xi_t\}_{t=d_n}^n$ and $\{v_t\}_{t=d_n}^n$. For CVaR, define for type $\in \{U, D, R\}$ and $\tau \in [\tau_l, \tau_u]$

$$\begin{aligned}
 T_{\text{type-CVaR}}^{\text{int}*}(\tau) &= \mathbb{I}(\text{type} \in \{U, R\}) \left[\log(k_1/(n\tau)) k_1^{1/2} M_R^* + \mathcal{X}_R^* \right] \\
 &\quad - \mathbb{I}(\text{type} \in \{D, R\}) \left\{ (k_1/k_2)^{1/2} \left[\log(k_2/(n\tau)) k_2^{1/2} M_L^* + \mathcal{X}_L^* \right] \right\},
 \end{aligned} \tag{4.5}$$

where M_R^* and M_L^* are defined in (4.3) and (4.4), respectively.

Then, for type $\in \{U, D, R\}$, $T_{\text{type-CVaR}}^{\text{int}*}(\tau)$ is used to mimic $k_1^{1/2} \left\{ \log \left[\widehat{\text{type-CVaR}}(\tau) / \text{type-CVaR}(\tau) \right] \right\}$. Similarly, for CES, define for type $\in \{U, D, R\}$ and $\tau \in [\tau_l, \tau_u]$

$$\begin{aligned}
 & T_{\text{type-CES}}^{\text{int}*}(\tau) \\
 &= \mathbb{I}(\text{type} \in \{U, R\}) \left\{ \left[\log(k_1/(n\tau)) + (1 - \hat{\gamma}_R)^{-1} \right] k_1^{1/2} M_R^* + \mathcal{X}_R^* \right\} \\
 &\quad - \mathbb{I}(\text{type} \in \{D, R\}) (k_1/k_2)^{1/2} \left\{ \left[\log(k_2/(n\tau)) + (1 - \hat{\gamma}_L)^{-1} \right] k_2^{1/2} M_L^* + \mathcal{X}_L^* \right\}.
 \end{aligned} \tag{4.6}$$

Then, for type $\in \{U, D, R\}$, $T_{\text{type-CES}}^{\text{int}*}(\tau)$ is used to mimic $k_1^{1/2} \left\{ \log \left[\widehat{\text{type-CES}}(\tau) / \text{type-CES}(\tau) \right] \right\}$.

Bootstrap Algorithm for the Intermediate Scenario

1. Specify a proper dynamic model and estimate it to obtain the residuals $\{\hat{\varepsilon}_t\}$. Compute the tail index estimates $\hat{\gamma}_R$ and $\hat{\gamma}_L$ as well as the maximum absolute log-ratios as defined in (2.9).

2. Generate i.i.d. random variables $\{\xi_t\}$ and $\{v_t\}$ from a certain distribution with zero mean and unit variance, then compute M_R^* and M_L^* . Generate \mathcal{X}_R^* and \mathcal{X}_L^* from normal distributions with zero mean and respective variances $\hat{\gamma}_R^2$ and $\hat{\gamma}_L^2$, then compute $T_{\text{type-CVaR}}^{\text{int}*}(\tau)$ and $T_{\text{type-CES}}^{\text{int}*}(\tau)$ for $\tau \in [\tau_l, \tau_u]$.
3. For a large integer B , repeat Step 2 B times to obtain $\{T_{\text{type-CVaR}}^{\text{int}*}(\tau)\}_{b=1}^B$ and $\{T_{\text{type-CES}}^{\text{int}*}(\tau)\}_{b=1}^B$. Then, for $\alpha \in (0, 1)$, calculate $z_{1, \text{type}, 1-\alpha}^{\text{int}*}$ and $z_{2, \text{type}, 1-\alpha}^{\text{int}*}$, which are, respectively, the $(1 - \alpha)$ -sample quantiles of $\{\sup_{\tau \in [\tau_l, \tau_u]} |T_{\text{type-CVaR}}^{\text{int}*}(\tau)|\}_{b=1}^B$ and $\{\sup_{\tau \in [\tau_l, \tau_u]} |T_{\text{type-CES}}^{\text{int}*}(\tau)|\}_{b=1}^B$.
4. For $\text{type} \in \{U, D, R\}$, asymptotic $(1 - \alpha)$ -SCBs for $\text{type-CVaR}(\tau)$ and $\text{type-CES}(\tau)$ are, respectively,

$$\left[\widehat{\text{type-CVaR}}(\tau) \exp\left(-z_{1, \text{type}, 1-\alpha}^{\text{int}*} k_1^{-1/2}\right), \widehat{\text{type-CVaR}}(\tau) \exp\left(z_{1, \text{type}, 1-\alpha}^{\text{int}*} k_1^{-1/2}\right) \right],$$

and

$$\left[\widehat{\text{type-CES}}(\tau) \exp\left(-z_{2, \text{type}, 1-\alpha}^{\text{int}*} k_1^{-1/2}\right), \widehat{\text{type-CES}}(\tau) \exp\left(z_{2, \text{type}, 1-\alpha}^{\text{int}*} k_1^{-1/2}\right) \right].$$

The validity of the proposed bootstrap algorithms are formally justified by the following theorem. By showing that the bootstrap statistics converge to the same limiting distributions as their original versions, Theorem 4.1 guarantees that the coverage rates of the bootstrap SCBs converge to the nominal confidence level.

THEOREM 4.1. (i) Under the assumptions of Theorem 3.1(i), for $\text{type} \in \{U, D, R\}$, $\sup_{u \in [1, \Delta]} |T_{\text{type}}^{\text{ex}*}|$ converges in distribution to $|Z_{\text{type}}|$ conditional on the original data in probability, where Z_{type} is defined in Theorem 3.1(i).

(ii) Under the assumptions of Theorem 3.1(ii), for $\text{type} \in \{U, D, R\}$, $\sup_{u \in [1, \Delta]} |T_{\text{type-CVaR}}^{\text{int}*}(u\tau_l)|$ converges in distribution to $\sup_{u \in [1, \Delta]} |G_{1, \text{type}}(u)|$ and $\sup_{u \in [1, \Delta]} |T_{\text{type-CES}}^{\text{int}*}(u\tau_l)|$ converges in distribution to $\sup_{u \in [1, \Delta]} |G_{2, \text{type}}(u)|$ conditional on the original data in probability, where $G_{1, \text{type}}(u)$ and $G_{2, \text{type}}(u)$ are defined in Theorem 3.1 (ii).

5. NUMERICAL STUDY

In this section, we first provide guidelines on how to choose the effective sample sizes k_1 and k_2 to implement the proposed approach, and then proceed to show the results of a Monte Carlo study.

5.1. Data-Driven Selection of k_1 and k_2

The implementation of the proposed approach requires selection of the effective sample sizes k_1 and k_2 . Previous studies on EVT-based risk measure estimation have found that the choice of the effective sample size is crucial (e.g., Chan et al., 2007; Martins-Filho et al., 2018). Various authors have proposed different empirical suggestions. For example, when constructing pointwise confidence

intervals for downside CVaR, Chan et al. (2007) suggest $\lfloor 1.5(\log n)^2 \rfloor$ for $n = 1,000$, where $\lfloor \cdot \rfloor$ is the integer part of a real number. Nonetheless, their suggestion is based on their simulations for $n = 1,000$, and Spierdijk (2016) finds that it is no longer adequate for larger sample sizes. Compared to the ad hoc suggestions, a data-driven rule is desirable because it selects k_1 and k_2 that are tailored to the data at hand.

Below we suggest a data-driven method for selecting k_1 and k_2 . Our estimation relies on the Pareto approximation to the tails of the innovation distribution. Then, k_1 and k_2 actually determine where the approximation starts for the right and left tails, respectively. Therefore, it is reasonable to choose the effective sample size that leads to the best Pareto approximation. To this end, we follow the wisdom of Danielsson et al. (2019), who propose to measure the Pareto approximation by the maximum distance between the fitted Pareto-type tail and the empirical quantile. Then, the optimal effective sample size is chosen as the minimizer of the measure. Since estimating CVaR essentially involves estimating quantiles, the quantile-based method is suitable for CVaR estimation.

Following the proposal of Danielsson et al. (2019), for the estimation of upside CVaR, we use k_1 equal to

$$k_1^* = \arg \min_{l=k_{min}, \dots, k_{max}} \left\{ \max_{j=1, \dots, k_{max}} \left[\left| \hat{\varepsilon}_{(j)} - \hat{\varepsilon}_{(l)} (j/l)^{-\hat{\gamma}_R(l)} \right| \right] \right\}, \tag{5.1}$$

where $1 \leq k_{min} < k_{max} \leq n$ and $\hat{\gamma}_R(l)$ is the estimator of γ_R with $k_1 = l$. In (5.1), $\hat{\varepsilon}_{(j)}$ is the empirical quantile and $\hat{\varepsilon}_{(l)} (j/l)^{-\hat{\gamma}_R(l)}$ is the quantile estimator given by the Pareto approximation. Similarly, when estimating downside CVaR, k_2 is chosen as k_2^* , which is obtained by applying the above criterion to $\{-\hat{\varepsilon}_t\}$.

For a conceptually suitable criterion for CES estimation, we modify (5.1) by employing its tail mean-based analog. Specifically, when estimating upside CES, we use k_1 equal to

$$k_1^* = \arg \min_{l=k_{min}, \dots, k_{max}} \left\{ \max_{j=1, \dots, k_{max}} \left[\left| \frac{\sum_{t=d_n}^n \mathbb{I}(\hat{\varepsilon}_t > \hat{\varepsilon}_{(j)}) \hat{\varepsilon}_t}{\sum_{t=d_n}^n \mathbb{I}(\hat{\varepsilon}_t > \hat{\varepsilon}_{(j)})} - \frac{\hat{\varepsilon}_{(l)} (j/l)^{-\hat{\gamma}_R(l)}}{1 - \hat{\gamma}_R(l)} \right| \right] \right\}, \tag{5.2}$$

where $\frac{\sum_{t=d_n}^n \mathbb{I}(\hat{\varepsilon}_t > \hat{\varepsilon}_{(j)}) \hat{\varepsilon}_t}{\sum_{t=d_n}^n \mathbb{I}(\hat{\varepsilon}_t > \hat{\varepsilon}_{(j)})}$ is the empirical estimator of the tail mean over the $(1 - j/n)$ -quantile and $\frac{\hat{\varepsilon}_{(l)} (j/l)^{-\hat{\gamma}_R(l)}}{1 - \hat{\gamma}_R(l)}$ is the estimator given by the Pareto approximation. Again, when estimating downside CES, k_2 is chosen as k_2^* , which is obtained by applying the above criterion to $\{-\hat{\varepsilon}_t\}$.

Finally, we note that Assumption 3.5(ii) and (iii) sets conditions for k_1 and k_2 . Below we provide suggestions on how these conditions can be satisfied. First, we choose k_1 and k_2 from the same $[k_{min}, k_{max}]$ so that k_1 and k_2 are comparable as required by Assumption 3.5(ii). Second, for the intermediate scenario (corresponding to Assumption 3.5(ii) with $c_1 > 0$), k_1 and k_2 should be comparable to $n\tau$. Accordingly, we can set $k_{min} = a n \tau_u$ and $k_{max} = b n \tau_u$ with $0 < a < b$.

For example, if the tail region is $[0.5\%, 1\%]$, we can set $k_{min} = 2\% \times n$ and $k_{max} = 15\% \times n$, corresponding to the setting that $a = 2$ and $b = 15$. In addition, the choice of k_{max} means that at most 15% of the sample are used, meeting the requirement of Assumption 3.5(iii) that k_{max} should be small relative to n^{ν_0} , where ν_0 is usually equal to, or at least close to, 1. Third, for the extreme scenario (corresponding to Assumption 3.5(ii) with $c_1 = 0$), k_1 and k_2 should be larger relative to $n\tau$. Accordingly, we should set k_{min} to be larger than $n\tau_u$. For example, if the tail region is $[0.05\%, 0.1\%]$, we can set $k_{min} = 1\% \times n$ and $k_{max} = 10\% \times n$. Again, the choice of k_{max} meets the requirement of Assumption 3.5(iii). Lastly, given the above two rules, $k_1^{1/2}$ would generally be larger than $|\log(n\tau_l/k_1)|$ and hence the condition $\lim_{n \rightarrow \infty} k_1^{-1/2} \log(n\tau_l/k_1) = 0$ in Assumption 3.5(iii) is satisfied.

5.2. Monte Carlo Simulations

In this section, we conduct a series of Monte Carlo simulations to evaluate the finite-sample performance of the proposed approach. In the bootstrap implementation, the number of bootstrap replications is $B = 500$ and the multipliers are generated from the two-point distribution with the probability mass function $P(\xi = 1) = P(\xi = -1) = 1/2$. The confidence level is 95%. The number of discarded observations is $d_n = \lfloor 5 \times (n\tau_l)^{1/3} \rfloor$, which should satisfy $d_n = o(k_1^{1/2})$ to satisfy Assumption 3.5. Since $k_1 \sim n\tau_l$ under the intermediate scenario and $k_1 = o(n\tau_l)$ under the extreme scenario, setting d_n to be $\lfloor 5 \times (n\tau_l)^{1/3} \rfloor$ meets the required order of d_n for both the intermediate and extreme scenarios. All the reported simulation results are obtained based on 1,000 replications.

The data are generated from the GARCH(1,1) model

$$R_t = \sigma_t \varepsilon_t, \quad \sigma_t^2 = \omega + \omega_1 R_{t-1}^2 + \omega_2 \sigma_{t-1}^2, \quad (5.3)$$

where $\omega = 0.001$, $\omega_1 = 0.04$, and $\omega_2 = 0.85$. The innovation term ε_t follows the standardized skewed sstd with the shape parameter ν and the skewness parameter λ (in short, sstd(ν, λ); see Fernández and Steel, 1998). The sstd is symmetric if $\lambda = 1$ and asymmetric otherwise, being skewed to the left if $\lambda < 1$ and to the right if $\lambda > 1$. In our simulations, we consider sstd(5, λ) where $\lambda \in \{0.95, 1\}$. The conclusions are qualitatively unchanged for other models such as the TGARCH model.

In the first part, we compare the finite-sample performance of the SCBs based on the two scenarios of asymptotics. The tail region is $[0.5\%, 1\%]$, and the sample size $n \in \{500, 1,000, 1,500, 5,000\}$. Note that the tail region is not beyond the range of the sample. Hence, theoretically, both the extreme and intermediate scenarios apply. The effective sample sizes k_1 and k_2 are selected using the data-driven method in Section 5.1 with $k_{min} = 2\% \times n$ and $k_{max} = 15\% \times n$ for the intermediate scenario and $k_{min} = 5\% \times n$ and $k_{max} = 15\% \times n$ for the extreme scenario. Apart from the data-driven (k_1^*, k_2^*) , we also use $(k_1^* + n/100, k_2^* + n/100)$ and $(k_1^* + n/50, k_2^* + n/50)$. This means that the user may want to employ more data, but at the

cost of less accurate Pareto approximations. The model parameters are estimated with the Gaussian QMLE method. We only report the results for the relative risk measures to save space.

The empirical coverage rates and relative lengths are collected in [Tables 1](#) and [2](#). Since our proposed SCBs are constructed in a multiplication-division manner, it is more informative to consider the relative length defined as the ratio of the SCB's upper bound to its lower bound. Several observations are in the sequel. First, the SCB based on the intermediate scenario has satisfactory coverage rates, but the SCB based on the extreme scenario suffers from evident underestimation. This confirms our discussion in [Section 3.2](#) that, when data are adequate so that no extrapolation is needed, the intermediate scenario provides more accurate approximations. Second, the more accurate coverage of the intermediate scenario is accompanied by a larger length. This means that the risk measure estimators under the intermediate scenario has more variability, the consequence of recalling the neglected terms in the extreme scenario. In addition, the coverage of the extreme scenario does not improve even if n increases to 5,000. This indicates that the impact of the neglected terms persist even for a very large sample size. Third, for $n = 500$, (k_1^*, k_2^*) leads to better coverage than $(k_1^* + n/100, k_2^* + n/100)$ and $(k_1^* + n/50, k_2^* + n/50)$ under the intermediate scenario. However, as n increases, the performance becomes more robust. Finally, the results are basically robust to different innovation distributions.

To summarize, we find that when the data are adequate so that extrapolation is not needed, the intermediate scenario produces SCBs with more accurate coverage rates than the extreme scenario, but accompanied by larger lengths. In addition, the data-driven methods yield the best finite-sample performance when $n = 500$ and the performance tends to be robust as n increases.

In the second part, we evaluate the performance of the SCB based on the extreme scenario when the tail region is indeed beyond the range of the available data, that is, when extrapolation is needed. We consider the tail region $[0.05\%, 0.1\%]$. To ensure extrapolation, the sample size should be no larger than 1,000, and hence we consider $n \in \{500, 1,000\}$. The data are still generated from model [\(5.3\)](#). The effective sample sizes k_1 and k_2 are selected using the data-driven method in [Section 5.1](#) with $k_{min} = 1\% \times n$ and $k_{max} = 10\% \times n$. Moreover, in addition to the data-driven (k_1^*, k_2^*) , we also use $(k_1^* + n/100, k_2^* + n/100)$ and $(k_1^* + n/50, k_2^* + n/50)$ to check robustness. The simulation results are collected in [Table 3](#). We observe that the coverage rates of the SCBs for CVaR are very close to the nominal level, but the coverage rates of the SCBs for CES are underestimated. Hence, it seems more difficult to estimate CES than CVaR under the extreme scenario. The same issue is also found in the simulation study of Martins-Filho et al. (201). Finally, we observe that the SCBs based on (k_1^*, k_2^*) have more accurate coverage rates and larger lengths than the SCBs based on $(k_1^* + n/100, k_2^* + n/100)$ and $(k_1^* + n/50, k_2^* + n/50)$.

A potential explanation for the undercoverage for CES under the extreme scenario is as follows. As judged from [\(2.14\)](#), the tail mean estimator is constructed

TABLE 1. Comparison of the two scenarios when extrapolation is not needed: $\tau \in [0.5\%, 1\%]$ and $\varepsilon_t \sim \text{sstd}(5,1)$

| <i>n</i> | Scenario | (k_1^*, k_2^*) | | | | $(k_1^* + n/100, k_2^* + n/100)$ | | | | $(k_1^* + n/50, k_2^* + n/50)$ | | | |
|----------|--------------|------------------|-------|-----------------|-------|----------------------------------|-------|-----------------|-------|--------------------------------|-------|-----------------|-------|
| | | Coverage | | Relative length | | Coverage | | Relative length | | Coverage | | Relative length | |
| | | CVaR | CES | CVaR | CES | CVaR | CES | CVaR | CES | CVaR | CES | CVaR | CES |
| 500 | Intermediate | 0.956 | 0.955 | 2.031 | 3.025 | 0.942 | 0.926 | 2.086 | 3.125 | 0.954 | 0.934 | 2.112 | 3.120 |
| | Extreme | 0.899 | 0.739 | 1.784 | 1.784 | 0.908 | 0.710 | 1.797 | 1.797 | 0.930 | 0.716 | 1.835 | 1.835 |
| 1,000 | Intermediate | 0.954 | 0.959 | 1.668 | 2.268 | 0.947 | 0.953 | 1.702 | 2.296 | 0.954 | 0.941 | 1.714 | 2.290 |
| | Extreme | 0.917 | 0.732 | 1.510 | 1.510 | 0.906 | 0.686 | 1.529 | 1.529 | 0.901 | 0.688 | 1.544 | 1.544 |
| 1,500 | Intermediate | 0.948 | 0.949 | 1.523 | 1.971 | 0.948 | 0.946 | 1.552 | 1.995 | 0.946 | 0.949 | 1.560 | 1.987 |
| | Extreme | 0.905 | 0.710 | 1.405 | 1.405 | 0.896 | 0.690 | 1.419 | 1.419 | 0.922 | 0.716 | 1.427 | 1.427 |
| 5,000 | Intermediate | 0.949 | 0.952 | 1.267 | 1.477 | 0.952 | 0.949 | 1.276 | 1.472 | 0.952 | 0.946 | 1.280 | 1.466 |
| | Extreme | 0.902 | 0.676 | 1.204 | 1.204 | 0.923 | 0.721 | 1.212 | 1.212 | 0.912 | 0.706 | 1.216 | 1.216 |

Notes: The results are for the relative risk measures under the confidence level 95%. k_1^* and k_2^* are selected by the data-driven methods in Section 5.1. The relative length is defined as the ratio of the SCB's upper bound to its lower bound. We note that under the extreme scenario, the SCBs for CVaR and CES have the same relative length.

TABLE 2. Comparison of the two scenarios when extrapolation is not needed: $\tau \in [0.5\%, 1\%]$ and $\varepsilon_t \sim \text{sstd}(5, 0.95)$

| n | Scenario | (k_1^*, k_2^*) | | | | $(k_1^* + n/100, k_2^* + n/100)$ | | | | $(k_1^* + n/50, k_2^* + n/50)$ | | | |
|-------|--------------|------------------|-------|-----------------|-------|----------------------------------|-------|-----------------|-------|--------------------------------|-------|-----------------|-------|
| | | Coverage | | Relative length | | Coverage | | Relative length | | Coverage | | Relative length | |
| | | CVaR | CES | CVaR | CES | CVaR | CES | CVaR | CES | CVaR | CES | CVaR | CES |
| 500 | Intermediate | 0.959 | 0.961 | 2.025 | 2.999 | 0.959 | 0.943 | 2.083 | 3.094 | 0.953 | 0.938 | 2.117 | 3.141 |
| | Extreme | 0.876 | 0.719 | 1.780 | 1.780 | 0.888 | 0.683 | 1.805 | 1.805 | 0.905 | 0.692 | 1.834 | 1.834 |
| 1,000 | Intermediate | 0.944 | 0.957 | 1.659 | 2.245 | 0.957 | 0.945 | 1.703 | 2.293 | 0.948 | 0.936 | 1.719 | 2.303 |
| | Extreme | 0.897 | 0.708 | 1.513 | 1.513 | 0.908 | 0.685 | 1.530 | 1.530 | 0.913 | 0.713 | 1.537 | 1.537 |
| 1,500 | Intermediate | 0.949 | 0.952 | 1.527 | 1.982 | 0.947 | 0.947 | 1.550 | 1.988 | 0.951 | 0.953 | 1.557 | 1.981 |
| | Extreme | 0.904 | 0.711 | 1.400 | 1.400 | 0.914 | 0.693 | 1.415 | 1.415 | 0.912 | 0.706 | 1.430 | 1.430 |
| 5,000 | Intermediate | 0.950 | 0.952 | 1.267 | 1.476 | 0.945 | 0.942 | 1.275 | 1.471 | 0.952 | 0.949 | 1.279 | 1.464 |
| | Extreme | 0.902 | 0.669 | 1.204 | 1.204 | 0.897 | 0.665 | 1.210 | 1.210 | 0.906 | 0.701 | 1.216 | 1.216 |

Notes: The results are for the relative risk measures under the confidence level 95%. k_1^* and k_2^* are selected by the data-driven methods in Section 5.1. The relative length is defined as the ratio of the SCB's upper bound to its lower bound. We note that under the extreme scenario, the SCBs for CVaR and CES have the same relative length.

TABLE 3. Performance of the extreme scenario under extrapolation: $\tau \in [0.05\%, 0.1\%]$ and $\varepsilon_t \sim \text{sstd}(5, \lambda)$

| λ | n | (k_1^*, k_2^*) | | | $(k_1^* + n/100, k_2^* + n/100)$ | | | $(k_1^* + n/50, k_2^* + n/50)$ | | |
|-----------|-------|------------------|-------|-----------------|----------------------------------|-------|-----------------|--------------------------------|-------|-----------------|
| | | Coverage | | Relative length | Coverage | | Relative length | Coverage | | Relative length |
| | | CVaR | CES | CVaR and CES | CVaR | CES | CVaR and CES | CVaR | CES | CVaR and CES |
| 1 | 500 | 0.942 | 0.841 | 3.369 | 0.924 | 0.822 | 3.338 | 0.929 | 0.812 | 3.331 |
| | 1,000 | 0.946 | 0.862 | 2.405 | 0.936 | 0.843 | 2.377 | 0.935 | 0.844 | 2.373 |
| 0.95 | 500 | 0.957 | 0.849 | 3.380 | 0.937 | 0.816 | 3.323 | 0.944 | 0.837 | 3.321 |
| | 1,000 | 0.944 | 0.859 | 2.405 | 0.917 | 0.810 | 2.375 | 0.932 | 0.846 | 2.357 |

Notes: The results are for the relative risk measures under the confidence level 95%. k_1^* and k_2^* are selected by the data-driven methods in Section 5.1. The relative length is defined as the ratio of the SCB's upper bound to its lower bound. We note that under the extreme scenario, the SCBs for CVaR and CES have the same relative length.

by dividing the quantile estimator by the factor $1 - \hat{\gamma}_R$, thereby introducing more variability into the CES estimator than the CVaR estimator. Nonetheless, Theorem 3.1 indicates that CVaR and CES share the same limiting distributions under the extreme scenario, meaning that the variability of the dividing factor is of smaller order. In finite samples, however, the variability of the dividing factor increases estimation uncertainty. Since the multiplier bootstrap only mimics the first-order stochastic behavior, the resulting SCBs tend to be too tight.

In the third part, we investigate the impact of the convergence rate of $\hat{\theta}$ on the performance of the proposed approach. We employ model (5.3) with $\text{sstd}(3,1)$ innovations, and the sample size is $n = 1,000$. We consider $\tau = 1\%$ for the intermediate scenario and $\tau = 0.1\%$ for the extreme scenario. For the estimation of the model parameters, two methods are employed: the Gaussian QMLE and the Laplace QMLE. When the innovation is $\text{sstd}(3,1)$, the convergence rate of the Laplace QMLE is $n^{-1/2}$ (Francq and Zakoian, 2010, Exam. 9.3), but the convergence rate of the Gaussian QMLE is roughly $n^{-1/3}$ (Mikosch and Straumann, 2006, Thm. 4.4). We also consider setting the parameter estimates to be the true parameter to investigate the performance when there is no estimation uncertainty. Figure 1 shows the coverage rates and relative lengths of the confidence intervals for R-CVaR for $k_1 = k_2 = k$ with $k \in \{20, 25, \dots, 195, 200\}$. It appears that there is no visible difference among the performances based on different estimators and the true parameter. Hence, for a reasonable range of k_1 and k_2 (at most 20% of the sample here), the first-stage estimation uncertainty does not seem to affect the performance of the proposed approach.

In the final part, we compare our EVT-based approach with the purely nonparametric approach of Gao and Song (2008). We employ model (5.3) with $\text{sstd}(5,1)$ innovations, and the sample size is $n = 1,000$. We apply our intermediate scenario and the approach of Gao and Song (2008) to construct confidence intervals for downside CVaR and CES for $\tau \in \{0.1\%, 0.2\%, \dots, 1\%\}$. For our approach, k_1 and k_2 are selected by the data-driven methods in Section 5.1 with $k_{\min} = \tau + 3\% \times n$ and $k_{\max} = \tau + 10\% \times n$. Figure 2 shows the coverage rates and lengths for different τ . For a direct comparison of the interval length, we use the absolute length defined as the difference between the confidence interval's upper bound and its lower bound.

We first discuss the results for CVaR. When $0.4\% \leq \tau \leq 1\%$, the coverage rates are basically comparable, but the nonparametric approach has larger lengths. When τ becomes more extreme (that is, when $\tau < 0.4\%$), the coverage rate of the nonparametric approach declines sharply and the length of the nonparametric approach is much smaller than the length of the EVT-based approach. Turning to the results for CES, we observe that the EVT-based approach has better coverage and, again, the gap in coverage becomes very evident when $\tau < 0.7\%$. In addition, the length of the EVT-based approach is larger than that of the nonparametric approach. Combining the above observations, we find that our approach performs more reliably for small tail levels, thereby highlighting the merit of EVT in estimating deep tails.

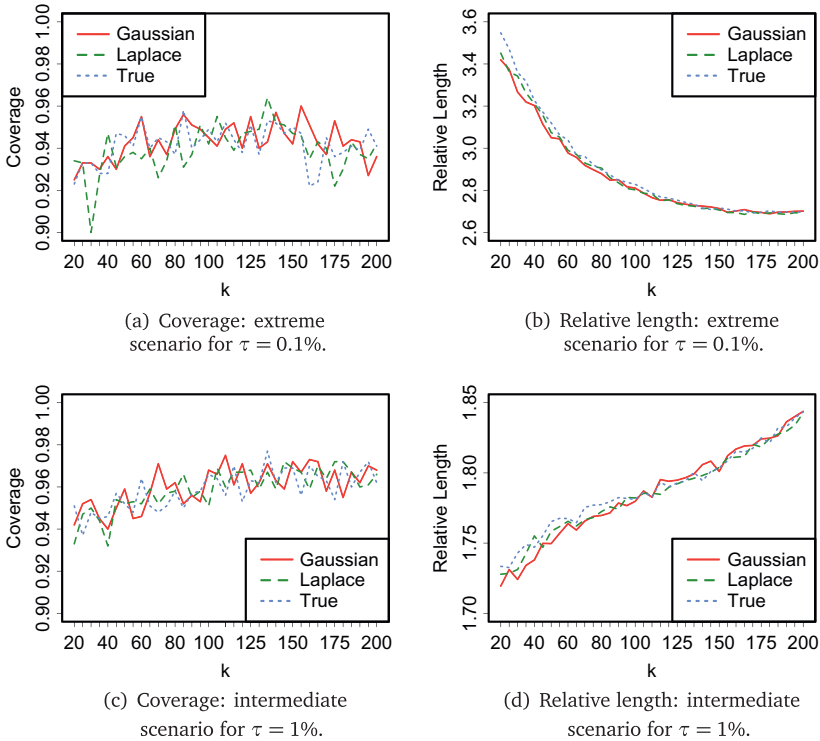


FIGURE 1. Comparison of coverage rates and relative lengths of the confidence intervals for R-CVaR based on different estimators of θ_0 . The confidence level is 95%. “Gaussian” means the Gaussian QMLE, “Laplace” means the Laplace QMLE, and “True” means the true parameter. The relative length is defined as the ratio of the confidence interval’s upper bound to its lower bound.

6. CONCLUSION

This paper proposes an approach to construct SCBs for CVaR and CES. We consider the one-sided risk measures as well as the relative risk measures. In the asymptotic theory, we consider two scenarios: (i) the extreme scenario that allows for extrapolation beyond the range of the available data and (ii) the intermediate scenario that works exclusively in the case where the available data are adequate relative to the tail level. For finite-sample implementation, we propose a novel bootstrap procedure to circumvent the slow convergence rates of the SCBs as well as infeasibility of approximating the limiting distributions. Simulations confirm that the proposed approach has favorable finite-sample performance. For future research, our theoretical results and the bootstrap procedure can be potentially extended to other risk measures such as the expectile (e.g., Daouia, Girard, and Stupfler, 2017; Hoga, 2022).

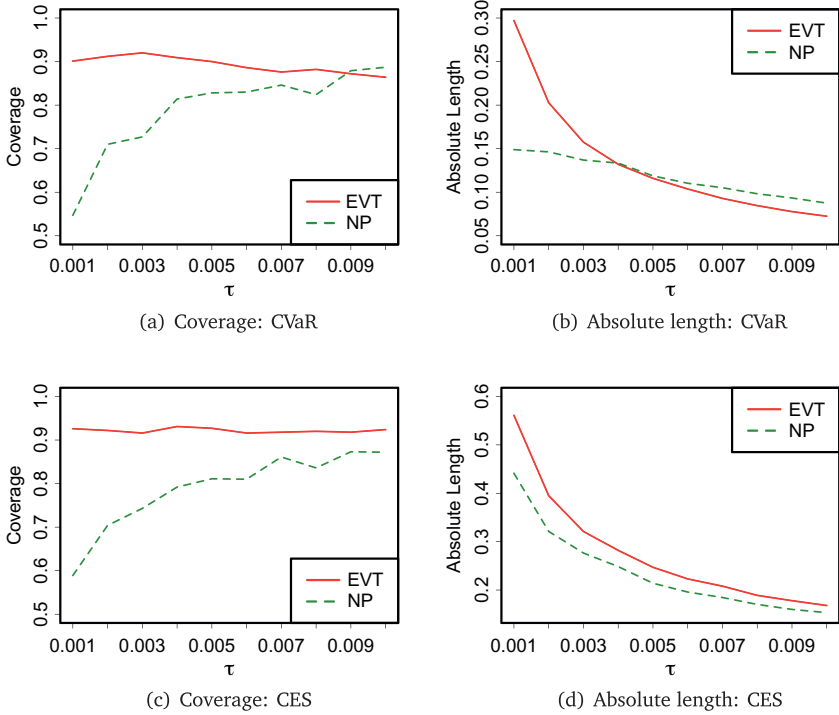


FIGURE 2. Coverage and absolute length of the confidence intervals. The confidence level is 95%. “EVT” denotes our proposed EVT-based approach, and “NP” denotes the nonparametric approach of Gao and Song (2008). The absolute length is defined as the difference between the confidence interval’s upper bound and its lower bound.

APPENDIX

The appendix consists of three parts. The first part provides full versions of Propositions 3.1 and 3.2 and their proofs. The second part contains the proofs of the main results. The last part collects primitive conditions for Assumptions 3.2 and 3.3 in the ARMA–GARCH, TGARCH, and GARCH-in-mean models.

A. Full Versions of Propositions 3.1 and 3.2

PROPOSITION A.1. *Under Assumptions 2.1 and 3.1–3.5, we have, uniformly in $\tau \in [\tau_l, \tau_u]$,*

$$\frac{\widehat{U-CVaR}(\tau)}{U-CVaR(\tau)} - 1 = \frac{\hat{Q}_\varepsilon(1-\tau)}{Q_\varepsilon(1-\tau)} - 1 + o_p(k_1^{-1/2} \log(n\tau/k_1)),$$

$$\frac{\widehat{D-CVaR}(\tau)}{D-CVaR(\tau)} - 1 = \frac{\hat{Q}_\varepsilon(\tau)}{Q_\varepsilon(\tau)} - 1 + o_p(k_1^{-1/2} \log(n\tau/k_1)),$$

$$\frac{\widehat{U-CES}(\tau)}{U-CES(\tau)} - 1 = \frac{\hat{E}[\varepsilon|\varepsilon > Q_\varepsilon(1-\tau)]}{E[\varepsilon|\varepsilon > Q_\varepsilon(1-\tau)]} - 1 + o_p(k_1^{-1/2} \log(n\tau/k_1)),$$

$$\frac{\widehat{D-CES}(\tau)}{D-CES(\tau)} - 1 = \frac{\hat{E}[\varepsilon|\varepsilon < Q_\varepsilon(\tau)]}{E[\varepsilon|\varepsilon < Q_\varepsilon(\tau)]} - 1 + o_p(k_1^{-1/2} \log(n\tau/k_1)).$$

Proof. By simple algebra,

$$\frac{\widehat{U-CVaR}(\tau)}{U-CVaR(\tau)} - 1 = \frac{\frac{m(\tilde{\mathcal{I}}_n, \hat{\theta}) - m(\mathcal{I}_n, \theta_0)}{\sigma(\mathcal{I}_n, \theta_0)Q_\varepsilon(1-\tau)} + \frac{\sigma(\tilde{\mathcal{I}}_n, \hat{\theta})}{\sigma(\mathcal{I}_n, \theta_0)} \left[\frac{\hat{Q}_\varepsilon(1-\tau)}{Q_\varepsilon(1-\tau)} - 1 \right] + \left[\frac{\sigma(\tilde{\mathcal{I}}_n, \hat{\theta})}{\sigma(\mathcal{I}_n, \theta_0)} - 1 \right]}{1 + \frac{m(\mathcal{I}_n, \theta_0)}{\sigma(\mathcal{I}_n, \theta_0)Q_\varepsilon(1-\tau)}}.$$

First, Lemma S1.16 in the Supplementary Material indicates that $\sigma(\tilde{\mathcal{I}}_n, \hat{\theta})/\sigma(\mathcal{I}_n, \theta_0) - 1 = O_p(n^{-\nu_0/2})$. In addition, as $Q_\varepsilon(1-\tau)$ is decreasing in τ and $Q_\varepsilon(1-\tau_u) \rightarrow \infty$ as $n \rightarrow \infty$ (de Haan and Ferreira, 2006, Prop. B.1.9), $Q_\varepsilon(1-\tau) \rightarrow \infty$ uniformly in $\tau \in [\tau_l, \tau_u]$. It follows that $\frac{m(\mathcal{I}_n, \theta_0)}{\sigma(\mathcal{I}_n, \theta_0)Q_\varepsilon(1-\tau)} = o_p(1)$ and $\frac{m(\tilde{\mathcal{I}}_n, \hat{\theta}) - m(\mathcal{I}_n, \theta_0)}{\sigma(\mathcal{I}_n, \theta_0)Q_\varepsilon(1-\tau)} = o_p(n^{-\nu_0/2})$ uniformly in $\tau \in [\tau_l, \tau_u]$. Thus,

$$\begin{aligned} \frac{\widehat{U-CVaR}(\tau)}{U-CVaR(\tau)} - 1 &= \left\{ \left[\frac{\hat{Q}_\varepsilon(1-\tau)}{Q_\varepsilon(1-\tau)} - 1 \right] \left[1 + O_p(n^{-\nu_0/2}) \right] + O_p(n^{-\nu_0/2}) \right\} \left[1 + o_p(1) \right] \\ &= \frac{\hat{Q}_\varepsilon(1-\tau)}{Q_\varepsilon(1-\tau)} - 1 + o_p(k_1^{-1/2} \log(n\tau/k_1)) \end{aligned}$$

uniformly in $\tau \in [\tau_l, \tau_u]$, where the last equality is due to Lemma S1.10(ii) in the Supplementary Material and Assumption 3.5 that $k_1 = o(n^{\delta_0})$ with $\delta_0 < \nu_0$. The proofs of other terms are similar and are thus omitted here. \square

PROPOSITION A.2. *Under Assumptions 2.1 and 3.1–3.5, we have, uniformly in $\tau \in [\tau_l, \tau_u]$,*

$$\begin{aligned} \frac{\hat{Q}_\varepsilon(1-\tau)}{Q_\varepsilon(1-\tau)} - 1 &= \left[\frac{\hat{\varepsilon}(k_1)}{U_\varepsilon(n/k_1)} - 1 \right] \\ &\quad + \log\left(\frac{k_1}{n\tau}\right) (\hat{\gamma}_R - \gamma_R) + o_p(k_1^{-1/2} \log(n\tau/k_1)), \\ \frac{\hat{Q}_\varepsilon(\tau)}{Q_\varepsilon(\tau)} - 1 &= \left[\frac{-\hat{\varepsilon}(n-k_2-d_n)}{U_{-\varepsilon}(n/k_2)} - 1 \right] \\ &\quad + \log\left(\frac{k_2}{n\tau}\right) (\hat{\gamma}_L - \gamma_L) + o_p(k_1^{-1/2} \log(n\tau/k_1)), \\ \frac{\hat{E}[\varepsilon|\varepsilon > Q_\varepsilon(1-\tau)]}{E[\varepsilon|\varepsilon > Q_\varepsilon(1-\tau)]} - 1 &= \left[\frac{\varepsilon(k_1)}{U_\varepsilon(n/k_1)} - 1 \right] \\ &\quad + \left[\log\left(\frac{k_1}{n\tau}\right) + \frac{1}{1-\gamma_R} \right] (\hat{\gamma}_R - \gamma_R) + o_p(k_1^{-1/2} \log(n\tau/k_1)), \\ \frac{\hat{E}[\varepsilon|\varepsilon < Q_\varepsilon(\tau)]}{E[\varepsilon|\varepsilon < Q_\varepsilon(\tau)]} - 1 &= \left[\frac{-\varepsilon(n-k_2-d_n)}{U_{-\varepsilon}(n/k_2)} - 1 \right] \\ &\quad + \left[\log\left(\frac{k_2}{n\tau}\right) + \frac{1}{1-\gamma_L} \right] (\hat{\gamma}_L - \gamma_L) + o_p(k_1^{-1/2} \log(n\tau/k_1)). \end{aligned}$$

Proof. This proposition follows immediately from Lemmas S1.10(ii) and S1.13(ii) in the Supplementary Material. \square

B. Proofs of the Main Results

Proof of Theorem 3.1. Theorem 3.1 follows from Propositions A.1 and A.2 and Lemmas S1.12 and S1.15 in the Supplementary Material, combined with the facts that

$$\begin{aligned} \log \left[\frac{\widehat{R-CVaR}(\tau)}{\widehat{R-CVaR}(\tau)} \right] &= \log \left[\frac{\widehat{U-CVaR}(\tau)}{\widehat{U-CVaR}(\tau)} \right] - \log \left[\frac{\widehat{D-CVaR}(\tau)}{\widehat{D-CVaR}(\tau)} \right] \\ &= \left[\frac{\hat{Q}_\varepsilon(1-\tau)}{\hat{Q}_\varepsilon(1-\tau)} - 1 + o_p(k_1^{-1/2} \log(n\tau/k_1)) \right] [1 + o_p(1)] \\ &\quad - \left[\frac{\hat{Q}_\varepsilon(\tau)}{\hat{Q}_\varepsilon(\tau)} - 1 + o_p(k_1^{-1/2} \log(n\tau/k_1)) \right] [1 + o_p(1)] \\ &= \left[\frac{\hat{Q}_\varepsilon(1-\tau)}{\hat{Q}_\varepsilon(1-\tau)} - 1 \right] - \left[\frac{\hat{Q}_\varepsilon(\tau)}{\hat{Q}_\varepsilon(\tau)} - 1 \right] + o_p(k_1^{-1/2} \log(n\tau/k_1)) \end{aligned}$$

and

$$\begin{aligned} \log \left[\frac{\widehat{R-CES}(\tau)}{\widehat{R-CES}(\tau)} \right] &= \log \left[\frac{\widehat{U-CES}(\tau)}{\widehat{U-CES}(\tau)} \right] - \log \left[\frac{\widehat{D-CES}(\tau)}{\widehat{D-CES}(\tau)} \right] \\ &= \left[\frac{\hat{E}[\varepsilon|\varepsilon > Q_\varepsilon(1-\tau)]}{E[\varepsilon|\varepsilon > Q_\varepsilon(1-\tau)]} - 1 \right] - \left[\frac{\hat{E}[\varepsilon|\varepsilon < Q_\varepsilon(\tau)]}{E[\varepsilon|\varepsilon < Q_\varepsilon(\tau)]} - 1 \right] \\ &\quad + o_p(k_1^{-1/2} \log(n\tau/k_1)). \end{aligned} \quad \square$$

Proof of Proposition 4.1. According to Lemma S1.9(ii) in the Supplementary Material, $\hat{\varepsilon}_{(k_1)}$, $\hat{\varepsilon}_{(n-k_2-d_n)}$, $\hat{\gamma}_R - \gamma_R$, and $\hat{\gamma}_L - \gamma_L$ are asymptotically independent and converge to a four-dimensional normal distribution jointly. The results of this proposition then follow immediately from Lemmas S1.10 and S1.13 and the proof of Lemma S1.9(ii) in the Supplementary Material, and Theorem 3.1. \square

Proof of Theorem 4.1. (i) It is clear that, for type $\in \{U, D, R\}$, $T_{\text{type}}^{\text{ex}*}$ is free of τ . In addition, M_R^* and M_L^* are independent conditional on the original data. Thus, it suffices to show that $k_1^{1/2} \gamma_R^{-1} M_R^*$ and $k_2^{1/2} \gamma_L^{-1} M_L^*$ both converge to standard normal distribution conditional on the original data.

According to Lemma S1.8 in the Supplementary Material and $k_1^{-1} \sum_{t=d_n}^n \mathbb{I}(\hat{\varepsilon}_t > \hat{\varepsilon}_{(k_1)}) \left(\log > \frac{\hat{\varepsilon}_t}{\hat{\varepsilon}_{(k_1)}} \right)^2 = \int_1^\infty 2\hat{F}^{(R)}(s) \log(s) \frac{ds}{s}$, we have

$$k_1^{-1/2} \sum_{t=d_n}^n \mathbb{I}(\hat{\varepsilon}_t > \hat{\varepsilon}_{(k_1)}) \left(\log > \frac{\hat{\varepsilon}_t}{\hat{\varepsilon}_{(k_1)}} \right)^2 = k_1^{-1/2} \sum_{t=d_n}^n \mathbb{I}(\varepsilon_t > \varepsilon_{(k_1)}) \left(\log > \frac{\varepsilon_t}{\varepsilon_{(k_1)}} \right)^2 + o_p(1),$$

where

$$k_1^{-1} \sum_{t=d_n}^n \mathbb{I}(\varepsilon_t > \varepsilon_{(k_1)}) \left(\log > \frac{\varepsilon_t}{\varepsilon_{(k_1)}} \right)^2 = 2\gamma_R^2 + o_p(1)$$

is due to Lemma 3.5.1 of de Haan and Ferreira (2006).

We first show that conditional on the original data, $k_1^{1/2} \gamma_R^{-1} M_R^*$ converges to standard normal distribution. It suffices to check the Lindeberg–Feller condition for $\{\Xi_{R,t}^*\}_{t=d_n}^n$, with $\Xi_{R,t}^* = k_1^{-1/2} \gamma_R^{-1} \mathbb{I}(\hat{\varepsilon}_t > \hat{\varepsilon}_{(k_1)}) \left(\log > \frac{\hat{\varepsilon}_t}{\hat{\varepsilon}_{(k_1)}} - \hat{\gamma}_R \right) \xi_t$. Denote $E_*(\cdot)$ as the expectation conditional on the original data. It is clear that conditional on the original data, $\Xi_{R,d_n}^*, \dots, \Xi_{R,n}^*$ are independent with mean zero. Then, $E_*(\Xi_{R,t}^*) = 0$ and

$$E_*(\Xi_{R,t}^{*2}) = k_1^{-1} \gamma_R^{-2} \mathbb{1}(\hat{\varepsilon}_t > \hat{\varepsilon}_{(k_1)}) \left(\log > \frac{\hat{\varepsilon}_t}{\hat{\varepsilon}_{(k_1)}} - \hat{\gamma}_R \right)^2,$$

for $t = d_n, \dots, n$. Thus,

$$\begin{aligned} \sum_{t=d_n}^n E_*(\Xi_{R,t}^{*2}) &= k_1^{-1} \gamma_R^{-2} \sum_{t=d_n}^n \mathbb{1}(\hat{\varepsilon}_t > \hat{\varepsilon}_{(k_1)}) \left(\log > \frac{\hat{\varepsilon}_t}{\hat{\varepsilon}_{(k_1)}} - \hat{\gamma}_R \right)^2 \\ &= k_1^{-1} \gamma_R^{-2} \sum_{t=d_n}^n \mathbb{1}(\hat{\varepsilon}_t > \hat{\varepsilon}_{(k_1)}) \left(\log > \frac{\hat{\varepsilon}_t}{\hat{\varepsilon}_{(k_1)}} \right)^2 \\ &\quad - 2k_1^{-1} \gamma_R^{-2} \hat{\gamma}_R \sum_{t=d_n}^n \mathbb{1}(\hat{\varepsilon}_t > \hat{\varepsilon}_{(k_1)}) \left(\log > \frac{\hat{\varepsilon}_t}{\hat{\varepsilon}_{(k_1)}} \right) + k_1^{-1} \gamma_R^{-2} \hat{\gamma}_R^2 \sum_{t=d_n}^n \mathbb{I}(\hat{\varepsilon}_t > \hat{\varepsilon}_{(k_1)}), \end{aligned}$$

where the first term in the last equation is $k_1^{-1} \gamma_R^{-2} \sum_{t=d_n}^n \mathbb{1}(\hat{\varepsilon}_t > \hat{\varepsilon}_{(k_1)}) \left(\log > \frac{\hat{\varepsilon}_t}{\hat{\varepsilon}_{(k_1)}} \right)^2 = 2 + o_p(1)$. For the second term and the last term, we have

$$\begin{aligned} -2k_1^{-1} \gamma_R^{-2} \hat{\gamma}_R \sum_{t=d_n}^n \mathbb{1}(\hat{\varepsilon}_t > \hat{\varepsilon}_{(k_1)}) \left(\log > \frac{\hat{\varepsilon}_t}{\hat{\varepsilon}_{(k_1)}} \right) &= -2\gamma_R^{-2} \hat{\gamma}_R^2 \text{ and} \\ k_1^{-1} \gamma_R^{-2} \hat{\gamma}_R^2 \sum_{t=d_n}^n \mathbb{I}(\hat{\varepsilon}_t > \hat{\varepsilon}_{(k_1)}) &= \gamma_R^{-2} \hat{\gamma}_R^2. \end{aligned}$$

Thus, we obtain that

$$\sum_{t=d_n}^n E_*(\Xi_{R,t}^{*2}) = 2 + o_p(1) - \gamma_R^{-2} \hat{\gamma}_R^2 = 1 + o_p(1)$$

as $\hat{\gamma}_R$ is a consistent estimator of γ_R .

For $t = d_n, \dots, n$, denote $\Lambda_{R,1,t} = k_1^{-1/2} \gamma_R^{-1} \mathbb{I}(\hat{\varepsilon}_t > \hat{\varepsilon}_{(k_1)}) \left(\log^> \frac{\hat{\varepsilon}_t}{\hat{\varepsilon}_{(k_1)}} - \hat{\gamma}_R \right)$ and $\Lambda_{R,2,t} = k_1^{-1} \gamma_R^{-2} \mathbb{I}(\hat{\varepsilon}_t > \hat{\varepsilon}_{(k_1)}) \left(\log^> \frac{\hat{\varepsilon}_t}{\hat{\varepsilon}_{(k_1)}} - \hat{\gamma}_R \right)^2$. Define

$$\Lambda_{R,1} = \max \left\{ \left| \Lambda_{R,1,t} / \sqrt{\sum_{t=d_n}^n E_*(\Xi_{R,t}^{*2})} \right| : t = d_n, \dots, n \right\}.$$

Note that, for $t = d_n, \dots, n$,

$$\begin{aligned} \min \left(0, k_1^{-1/2} \gamma_R^{-1} \mathbb{I}(\hat{\varepsilon}_{(k_1-1)} > \hat{\varepsilon}_{(k_1)}) \left(\log^> \frac{\hat{\varepsilon}_{(k_1-1)}}{\hat{\varepsilon}_{(k_1)}} - \hat{\gamma}_R \right) \right) \\ \leq \Lambda_{R,1,t} \leq k_1^{-1/2} \gamma_R^{-1} \mathbb{I}(\hat{\varepsilon}_{(0)} > \hat{\varepsilon}_{(k_1)}) \left(\log^> \frac{\hat{\varepsilon}_{(0)}}{\hat{\varepsilon}_{(k_1)}} - \hat{\gamma}_R \right). \end{aligned}$$

According to Corollary 1.2.4 of de Haan and Ferreira (2006), $\varepsilon_{(0)}/U_R(n) = O_p(1)$. In addition, as $\varepsilon_{(k_1)}/U(n/k_1) = 1 + o_p(1)$ and $U_R(n)/U_R(n/k_1) = O(k_1^{\gamma_R})$, we have that $\varepsilon_{(0)}/\varepsilon_{(k_1)} = O(k_1^{\gamma_R})$. According to Lemma S1.16 in the Supplementary Material, $m(\tilde{\mathcal{I}}_n, \hat{\theta}) - m(\mathcal{I}_n, \theta_0) = O_p(n^{-\nu_0/2})$ and $\sigma(\tilde{\mathcal{I}}_n, \hat{\theta}) - \sigma(\mathcal{I}_n, \theta_0) = O_p(n^{-\nu_0/2})$. As $\sigma(\mathcal{I}_n, \theta_0) > c$ for some $c > 0$, simple algebra yields that

$$\begin{aligned} \frac{\hat{\varepsilon}_t}{\varepsilon_t} &= \frac{R_t - m(\tilde{\mathcal{I}}_n, \hat{\theta})}{\frac{R_t - m(\tilde{\mathcal{I}}_n, \hat{\theta})}{\sigma(\tilde{\mathcal{I}}_n, \hat{\theta})}} = \frac{R_t - m(\tilde{\mathcal{I}}_n, \hat{\theta})}{R_t - m(\mathcal{I}_n, \theta_0)} \frac{\sigma(\mathcal{I}_n, \theta_0)}{\sigma(\tilde{\mathcal{I}}_n, \hat{\theta})} \\ &= \left[1 - \frac{m(\tilde{\mathcal{I}}_n, \hat{\theta}) - m(\mathcal{I}_n, \theta_0)}{\sigma(\mathcal{I}_n, \theta_0) \varepsilon_t} \right] \left[1 + \frac{\sigma(\tilde{\mathcal{I}}_n, \hat{\theta}) - \sigma(\mathcal{I}_n, \theta_0)}{\sigma(\mathcal{I}_n, \theta_0)} \right]^{-1} = 1 + o_p(1). \end{aligned}$$

Thus, we have $\hat{\varepsilon}_{(0)}/\hat{\varepsilon}_{(k_1)} = O(k_1^{\gamma_R})$. In addition, according to Lemma S1.9 in the Supplementary Material, $\hat{\varepsilon}_{(k_1-1)}/U_\varepsilon(n/(k_1 - 1)) = 1 + O_p(k_1^{-1/2})$ and $\hat{\varepsilon}_{(k_1)}/U_\varepsilon(n/k_1) = 1 + O_p(k_1^{-1/2})$. Furthermore, it is easy to show that $U_\varepsilon(n/(k_1 - 1))/U_\varepsilon(n/k_1) = 1 + o(1)$. So we have $\hat{\varepsilon}_{(k_1-1)}/\hat{\varepsilon}_{(k_1)} = 1 + o_p(1)$, and this implies $\Lambda_{R,1} = o_p(1)$.

Note that $\Xi_{R,t}^* = \Lambda_{R,1,t} \xi_t$ and $\Xi_{R,t}^{*2} = \Lambda_{R,2,t} \xi_t^2$. For any $\epsilon > 0$,

$$\begin{aligned} &\left[\sum_{t=d_n}^n E_*(\Xi_{R,t}^{*2}) \right]^{-1} \sum_{t=d_n}^n E_* \left[\Xi_{R,t}^{*2} \mathbb{I} \left(|\Xi_{R,t}^*| > \sqrt{\sum_{t=d_n}^n E_*(\Xi_{R,t}^{*2})} \right) \right] \\ &= \left[\sum_{t=d_n}^n E_*(\Xi_{R,t}^{*2}) \right]^{-1} \sum_{t=d_n}^n \Lambda_{R,2,t} E_* \left[\xi_t^2 \mathbb{I} \left(|\Lambda_{R,1,t} \xi_t| > \sqrt{\sum_{t=d_n}^n E_*(\Xi_{R,t}^{*2})} \right) \right] \\ &= \left[\sum_{t=d_n}^n E_*(\Xi_{R,t}^{*2}) \right]^{-1} \sum_{t=d_n}^n \Lambda_{R,2,t} E_* \left[\xi_t^2 \mathbb{I} \left(\left| \Lambda_{R,1,t} \xi_t / \sqrt{\sum_{t=d_n}^n E_*(\Xi_{R,t}^{*2})} \right| > \epsilon \right) \right] \end{aligned}$$

$$\begin{aligned}
 &\leq \left[\sum_{t=d_n}^n E_*(\Xi_{R,t}^{*2}) \right]^{-1} \sum_{t=d_n}^n \Lambda_{R,2,t} E_* \left[\xi_t^2 \mathbb{I}(\Lambda_{R,1} |\xi_t| > \epsilon) \right] \\
 &= \left[\sum_{t=d_n}^n E_*(\Xi_{R,t}^{*2}) \right]^{-1} \sum_{t=d_n}^n \Lambda_{R,2,t} E_* \left[\xi_1^2 \mathbb{I}(\Lambda_{R,1} |\xi_1| > \epsilon) \right] \\
 &= E_* \left[\xi_1^2 \mathbb{I}(\Lambda_{R,1} |\xi_1| > \epsilon) \right] \\
 &\leq [E_*(\xi_1^4)]^{1/2} [\Pr_*(\Lambda_{R,1} |\xi_1| > \epsilon)]^{1/2} \\
 &= o_p(1),
 \end{aligned}$$

where the second last equality is due to $\sum_{t=d_n}^n \Lambda_{R,2,t} = \sum_{t=d_n}^n E_*(\Xi_{R,t}^{*2})$, the last inequality is from Cauchy–Schwarz inequality, and the last equality is because of $\Lambda_{R,1} = o_p(1)$. Thus, $\{\Xi_{R,t}^*\}_{t=d_n}^n$ satisfies the Lindeberg–Feller condition, and hence $k_1^{1/2} \gamma_R^{-1} M_R^*$ converges to standard normal distribution. Similarly, we can show that $k_2^{1/2} \gamma_L^{-1} M_L^*$ also converges to standard normal distribution conditional on the original data and thus the first part of this theorem.

(ii) Regarding the formulations of $T_{\text{type-CVaR}}^{\text{int}*}$ and $T_{\text{type-CES}}^{\text{int}*}$, along with $k_1^{1/2} \gamma_R^{-1} M_R^*$ and $k_2^{1/2} \gamma_L^{-1} M_L^*$ both converge to standard normal distribution conditional on the original data as shown in part (i), the proof of this part of this theorem can be proved in the same manner as part (ii) of Lemma S1.11 in the Supplementary Material and thus is omitted here. Then, we finish the proof of this theorem. □

C. Primitive Conditions for Assumptions 3.2 and 3.3.

This section provides primitive conditions for Assumptions 3.2 and 3.3 in the ARMA–GARCH, the TGARCH, and the GARCH-in-mean models. In the following discussions, we restrict our attention to the simplest forms of these models to reduce notational complication.

Example C.1 (ARMA–GARCH).

$$R_t = \mu + \phi R_{t-1} + e_t + \psi e_{t-1}, \quad e_t = \sigma_t \varepsilon_t, \quad \sigma_t^2 = \omega + \omega_1 e_{t-1}^2 + \omega_2 \sigma_{t-1}^2,$$

where $\omega > 0$, $\omega_1 > 0$, and $\omega_2 > 0$. Here, $\theta_0 = (\mu, \phi, \psi, \omega, \omega_1, \omega_2)$, $\mathcal{I}_{t-1} = \sigma \{ \varepsilon_{t-1}, \varepsilon_{t-2}, \dots \}$, $m(\mathcal{I}_{t-1}, \theta_0) = \mu + \phi R_{t-1} + \psi e_{t-1}$, and $\sigma^2(\mathcal{I}_{t-1}, \theta_0) = \omega + \omega_1 e_{t-1}^2 + \omega_2 \sigma_{t-1}^2$.

Example C.2 (TGARCH).

$$R_t = \mu + \sigma_t \varepsilon_t, \quad \sigma_t^2 = \omega + (\omega_0 N_{t-1} + \omega_1) \sigma_{t-1}^2 \varepsilon_{t-1}^2 + \omega_2 \sigma_{t-1}^2,$$

where N_t equals 1 if $\varepsilon_t < 0$ and 0 otherwise, $\omega > 0$, $\omega_0 + \omega_1 > 0$, $\omega_1 > 0$, and $\omega_2 > 0$. Here, $\theta_0 = (\mu, \omega, \omega_0, \omega_1, \omega_2)$, $\mathcal{I}_{t-1} = \sigma \{ \varepsilon_{t-1}, \varepsilon_{t-2}, \dots \}$, $m(\mathcal{I}_{t-1}, \theta_0) = \mu$, and $\sigma^2(\mathcal{I}_{t-1}, \theta_0) = \omega + (\omega_0 N_{t-1} + \omega_1) \sigma_{t-1}^2 \varepsilon_{t-1}^2 + \omega_2 \sigma_{t-1}^2$.

Example C.3 (GARCH-in-mean).

$$R_t = \mu + \phi X_{t-1} + \psi \sigma_{t-1}^2 + \sigma_t \varepsilon_t, \quad \sigma_t^2 = \omega + \omega_1 \sigma_{t-1}^2 \varepsilon_{t-1}^2 + \omega_2 \sigma_{t-1}^2,$$

where $\omega > 0$, $\omega_1 > 0$, and $\omega_2 > 0$. X_t is an exogenous variable generated by $\{\eta_t\}_{t=-\infty}^{\infty}$, a sequence of i.i.d. random vectors independent of the innovation sequence $\{\varepsilon_t\}_{t=-\infty}^{\infty}$. Here, $\theta_0 = (\mu, \phi, \psi, \omega, \omega_1, \omega_2)$, $\mathcal{I}_{t-1} = \sigma\{\varepsilon_{t-1}, \varepsilon_{t-2}, \dots; \eta_{t-1}, \eta_{t-2}, \dots\}$, $m(\mathcal{I}_{t-1}, \theta_0) = \mu + \phi X_{t-1} + \psi \sigma_{t-1}$, and $\sigma^2(\mathcal{I}_{t-1}, \theta_0) = \omega + \omega_1 \sigma_{t-1}^2 \varepsilon_{t-1}^2 + \omega_2 \sigma_{t-1}^2$.

PROPOSITION C.1. *Under the condition that $E(|R_t|^{v_1}) < \infty$, for some $v_1 > 2/v_0$, Assumptions 3.2 and 3.3 are satisfied:*

- (i) *by Example A.1, if $|\phi| < 1$, $\omega > 0$, $0 < \omega_1 < 1$, $0 < \omega_2 < 1$, and $0 < \omega_1 + \omega_2 < 1$;*
- (ii) *by Example A.2, if $\omega > 0$, $0 < \omega_0 + \omega_1 < 1$, $0 < \omega_1 < 1$, $0 < \omega_2 < 1$, $0 < \omega_0 + \omega_1 + \omega_2 < 1$; and*
- (iii) *by Example A.3, if $X_t = \psi_1 X_{t-1} + \eta_t$ with $E(|\eta_t|^{v_1}) < \infty$, $|\psi_1| < 1$, $\omega > 0$, $0 < \omega_1 < 1$, $0 < \omega_2 < 1$, and $0 < \omega_1 + \omega_2 < 1$.*

The proof of Proposition C.3 can be found in Section S3 of the Supplementary Material.

SUPPLEMENTARY MATERIAL

To view the online supplementary material for this article, please visit:
<https://doi.org/10.1017/S0266466622000275>

REFERENCES

- Bai, J. and S. Ng (2001) A consistent test for conditional symmetry in time series models. *Journal of Econometrics* 103, 225–258.
- Barberis, N. and M. Huang (2008) Stocks as lotteries: The implications of probability weighting for security prices. *American Economic Review* 98, 2066–2100.
- Bollerslev, T. (1987) A conditionally heteroskedastic time series model for speculative prices and rates of return. *Review of Economics and Statistics* 69, 542–547.
- Chan, N.H., S.-J. Deng, L. Peng, and Z. Xia (2007) Interval estimation of value-at-risk based on GARCH models with heavy-tailed innovations. *Journal of Econometrics* 137, 556–576.
- Danielsson, J., L.M. Ergun, L. de Haan, and C.G. de Vries (2019) Tail Index Estimation: Quantile Driven Threshold Selection. Discussion paper, Bank of Canada.
- Daouia, A., S. Girard, and G. Stupfler (2017) Estimation of tail risk based on extreme Expectiles. *Journal of the Royal Statistical Society Series B* 80, 263–292.
- de Haan, L. and A. Ferreira (2006) *Extreme Value Theory: An Introduction*. Springer Series in Operations Research and Financial Engineering. Springer.
- Einmahl, J.H.J., L. de Haan, and C. Zhou (2016) Statistics of heteroscedastic extremes. *Journal of the Royal Statistical Society Series B (Methodological)* 78, 31–51.
- Engle, R.F., D.M. Lilien, and R.P. Robins (1987) Estimating time varying risk premia in the term structure: The arch-M model. *Econometrica* 55, 391–407.
- Fernández, C. and M.F.J. Steel (1998) On Bayesian modeling of fat tails and skewness. *Journal of the American Statistical Association* 93(441), 359–371.
- Franq, C. and J.-M. Zakoïan (2010) *GARCH Models: Structure, Statistical Inference and Financial Applications*. Wiley.
- Franq, C. and J.-M. Zakoïan (2016) Looking for efficient QML estimation of conditional VaRs at multiple risk levels. *Annals of Economics and Statistics* 123, 9–28.
- Gao, F. and F. Song (2008) Estimation risk in GARCH VaR and ES estimates. *Econometric Theory* 24(5), 1404–1424.

- 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. *Journal of Finance* 48, 1779–1801.
- Hall, P. (1982) On some simple estimates of an exponent of regular variation. *Journal of the Royal Statistical Society. Series B* 44, 37–42.
- Hamilton, J. (1994) *Time Series Analysis*. Princeton University Press.
- Hill, B.M. (1975) A simple general approach to inference about the tail of a distribution. *Annals of Statistics* 3, 1163–1174.
- Hill, J.B. (2015) Robust estimation and inference for heavy tailed GARCH. *Bernoulli* 21, 1629–1669.
- Hoga, Y. (2019a) Confidence intervals for conditional tail risk measures in ARMA–GARCH models. *Journal of Business & Economic Statistics* 37(4), 613–624.
- Hoga, Y. (2019b) Extreme conditional tail moment estimation under serial dependence. *Journal of Financial Econometrics* 17, 587–615.
- Hoga, Y. (2022) Limit theory for forecasts of extreme distortion risk measures and expectiles. *Journal of Financial Econometrics* 20, 18–44.
- 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(1), 23–67.
- McNeil, A.J. and R. Frey (2000) Estimation of tail-related risk measures for heteroscedastic financial time series: An extreme value approach. *Journal of Empirical Finance* 7, 271–300.
- Mikosch, T. and D. Straumann (2006) Stable limits of martingale transforms with application to the estimation of GARCH parameters. *Annals of Statistics* 34, 493–592.
- Pan, X., X. Leng, and T. Hu (2013) The second-order version of Karamata’s theorem with applications. *Statistics & Probability Letters* 83, 1397–1403.
- Romano, J., A. Shaikh, and M. Wolf (2018) Multiple testing. In S. Durlauf and L. Blume (eds.), *The New Palgrave Dictionary of Economics, 3rd Edition*, pp. 9185–9189. Palgrave Macmillan.
- Spierdijk, L. (2016) Confidence intervals for ARMA–GARCH value-at-risk: The case of heavy tails and skewness. *Computational Statistics and Data Analysis* 100, 545–559.
- Theodossiou, P. and C.S. Savva (2015) Skewness and the relation between risk and return. *Management Science* 62(6), 1598–1609.
- Tsay, R.S. (2010) *Analysis of Financial Time Series Analysis*. Wiley Series in Probability and Statistics. Wiley.
- Tversky, A. and D. Kahneman (1992) Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty* 5, 297–323.