Risk Analysis via Generalized Pareto Distributions

Yi He
University of Amsterdam

Liang Peng
Georgia State University

Dabao Zhang
Purdue University

Zifeng Zhao
University of Notre Dame

October 9, 2020

Abstract

We compute the value-at-risk of financial losses by fitting a generalized Pareto distribution to exceedances over a threshold. Following the common practice of setting the threshold as high sample quantiles, we show that, for both independent observations and time-series data, the asymptotic variance for the maximum likelihood estimation depends on the choice of threshold unlike the existing study of using a divergent threshold. We also propose a random weighted bootstrap method for interval estimation of VaR, with critical values computed by the empirical distribution of the absolute differences between the bootstrapped estimators and the maximum likelihood estimator. While our asymptotic results unify the cases of fixed or divergent thresholds, the finite sample studies via simulation and application to real data show that the derived confidence intervals well cover true VaR in insurance and finance.

Keywords: ARMA-GARCH models; Generalized Pareto distribution; Random weighted bootstrap; Value-at-risk; Weighted empirical process

1 Introduction

Measuring risk and quantifying its uncertainty is crucial in insurance and finance. A well-studied and widely employed risk measure is the so-called Value-at-Risk (VaR) at level $1-p \in (0,1)$, which is defined as the quantile of the distribution function of a risk variable or a portfolio; see Duffie and Pan (1997) and Jorion (2006) for an overview of VaR. Given n identically distributed observations, the VaR at level $1-p$ can be well estimated nonparametrically by the sample quantile when $n(1-p)$ is neither close to n nor zero. Quantifying the inference uncertainty can be done via direct estimation of the asymptotic variance or resampling methods such as the bootstrap and the empirical likelihood in Owen (2001).

In practice, the level $1-p$ of VaR is often set to be close to one by regulators such as 99% and 99.9%. Therefore, when the sample size is not particularly large, the nonparametric VaR estimation is inefficient and may seriously underestimate the risk. An obvious way to improve inference efficiency is to fit a parametric distribution family to the risk variable. It is known that efficient likelihood based inference mainly utilizes the information around the center of data. As $1-p$ is close to one, the information in the upper tail of the distribution becomes more crucial to the study of VaR. Therefore, one may build a parametric model for observations above a threshold to ensure that the upper tail's fitting is accurate and robust. This raises an interesting question on how to model the excess distribution above a threshold given by

$$F_u(x) = \mathbb{P}(X - u \le x | X > u) = \frac{F(x+u) - F(u)}{1 - F(u)} \quad \text{for} \quad 0 \le x < x_F - u,$$

where x_F is the right endpoint of the distribution function $F(x) = \mathbb{P}(X \le x)$, i.e., $x_F = \sup\{x : F(x) < 1\}$.

As stated in the Extreme Value Theory, see Resnick (1987) and Embrechts et al. (1997) for an overview, when F is in the domain of attraction of extreme value distribution, there exists a function $\beta(u) > 0$ such that

$$\lim_{u \to x_F} \sup_{0 \le x < x_F - u} |F_u(x) - G_{\gamma,\beta(u)}(x)| = 0, \tag{1.1}$$

where $G_{\gamma,\beta(u)}(x) = 1 - (1 + \gamma x/\beta(u))^{-1/\gamma}$ for $1 + \gamma x/\beta(u) > 0$ is the cumulative distribution function of the generalized Pareto distribution with the shape parameter γ and scale parameter $\beta(u)$; see Balkema and de Haan (1974). Fitting a generalized Pareto distribution to exceedances over a high threshold has been studied in the literature. For example, Smith (1987) and Drees et al. (2004) have studied the maximum likelihood estimation when a deterministic divergent threshold and a random divergent threshold are chosen, respectively; see also Davison and Smith (1990). The choice of the threshold depends on the approximation errors in (1.1), which generally is defined as a second order regular variation. Typically, a large threshold gives a big variance, and a small threshold leads to large estimation bias. Given the difficulty in choosing this divergent threshold, researchers often advise practitioners to plot estimators against various thresholds and find a relatively stable region. In this case, the estimator has a non-negligible bias, which complicates interval estimation.

Nevertheless, as a rule of thumb, practitioners often ignore the asymptotic bias and pick up 90% or 95% sample quantile as a threshold; see the discussions in Section 13.6.1 of Hull (2018). This is especially the case when modeling the so-called dynamic tail risk by some critical economic variables. Some applications of the generalized Pareto distribution include Rootzén and Tajvidi (1997) and Brodin and Rootzén (2009) for wind storm losses, Barro and Jin (2011) for economic disasters, and McNeil and Frey (2000), Chavez-Demoulin and Embrechts (2004), Bollerslev and Todorov (2011), and Allen et al. (2012) for financial time series. For dynamically modeling the generalized Pareto distribution, we refer to Chavez-Demoulin et al. (2005), Kelly and Jiang (2014), Chavez-Demoulin et al. (2014), and Massacci (2017) for financial returns and Hall and Tajvidi (2000) for climate data.

In reality, practitioners often choose the threshold as 90% or 95% sample quantile and ignore the estimation bias caused by the model approximation error. Hence, it becomes natural to model exceedances over an (unknown) fixed threshold by a generalized Pareto distribution. In other words, instead of fitting a parametric model, it is good to fit the exceedances over a threshold by the Generalized Pareto distribution and model the data below the threshold nonparametrically like Smith (1987) and Drees et al. (2004) for independent data, McNeil and Frey (2000) for an ARMA-GARCH model, and Martins-Filho et al. (2018) for nonparametric regression. Under such a model assumption, when the threshold is chosen as a sample quantile, inference for parameters and VaR will indeed depend on the random threshold selected, which is in stark contrast with the existing study of using a divergent threshold. A particular semi-parametric model we focus on is

$$F(x) = \begin{cases} \frac{\theta G(x)}{G(x_0)} & \text{if } x \le x_0 \\ 1 - (1-\theta)\left(1 + \frac{\gamma(x-x_0)}{\sigma}\right)^{-1/\gamma} & \text{if } x > x_0, \end{cases} \tag{1.2}$$

where $\theta \in (0,1)$, and G is a distribution function.

This paper aims to provide a comprehensive inference for such a model based on independent observations and time series data. We focus on VaR, but the developed methodologies can be extended/applied to other tail-related risk measures such as expected shortfall and expectile. As it is arguably reasonable to assume that insurance losses are independent, we first derive the asymptotic distribution for parameters and risk estimation based on independent data. We develop a unified inference theory for a universal threshold statistic, which can be a deterministic threshold based on prior knowledge, an order statistic based on the observations, or a more sophisticated quantile estimator. To quantify the inference uncertainty, we investigate the naive bootstrap method and the random weighted bootstrap method.

For dependent data such as financial time series, we propose considering conditional VaR by combining an ARMA-GARCH model and the semiparametric model for the residuals. To ensure the normality of VaR estimation for the ARMA-GARCH model with heavy-tailed errors, we propose a two-step self-weighted procedure to estimate the ARMA-GARCH model before fitting the residual distribution semiparametrically. We first estimate the ARMA parameters by a self-weighted least-squares method. Then, the GARCH parameters using the self-weighted exponential quasi-likelihood in Zhu and Ling (2011) with the least-squares ARMA residuals. Our approach maintains the natural condition that the ARMA errors have a zero mean, rather than a zero median in the previous paper, when relaxing the kurtosis condition on GARCH errors. To quantify the uncertainty of the conditional VaR estimation, we employ the random weighted bootstrap method, which is much less computationally intensive than the residual based bootstrap method.

The existing methods of using a divergent threshold face the severe difficulty of choosing the threshold. When one concerns interval estimation, the efficient way is to choose a larger threshold such that the estimation bias is negligible. This essentially assumes the exceedance follows an exact Generalized Pareto distribution. In other words, when the exceedance has an approximate Generalized Pareto distribution, our proposed point estimation and interval estimation are still valid when we choose a divergent threshold larger enough such that the model approximation error is a smaller order of the estimation error.

We organize this paper as follows. Sections 2 and 3 present our methodologies and asymptotic results for independent observations and an ARMA-GARCH model, respectively. Sections 4 and 5 contain simulation study and data analysis. Section 6 concludes. The detailed proofs of the theorems are available in the supplement. We denote by A^T the transpose of a matrix or vector A. Throughout this paper, we denote $\xrightarrow{d}$ as convergence in distribution and $\xrightarrow{\mathbb{P}}$ as convergence in probability. All the asymptotic results hold as the sample size $n \to \infty$.

2 Methodologies and Asymptotic Results for Independent Data

Consider a random variable $X \in \mathbb{R}$ with distribution function F and quantile function $Q(\cdot) = F^{\leftarrow}(\cdot)$. For a threshold $u_0 = F^{\leftarrow}(1-\alpha_0)$ with exceeding probability $\alpha_0 \in (0,1)$, we make the following assumption for the exceedance $X - u_0|X > u_0$.

Assumption 1 (Generalized Pareto Model). There exist a shape parameter $\gamma_0 \in \mathbb{R}$ and a scale parameter $\sigma_{\alpha_0} > 0$ such that

$$\mathbb{P}\left(X > u_0 + x | X > u_0\right) = \begin{cases} \left(1 + \frac{\gamma_0 x}{\sigma_{\alpha_0}}\right)^{-1/\gamma_0}, & \gamma_0 \neq 0, \\ \exp\left(-\frac{x}{\sigma_{\alpha_0}}\right), & \gamma_0 = 0. \end{cases}$$

where $\alpha_0 = \mathbb{P}(X > u_0)$ and we require $1 + \gamma_0 x/\sigma_{\alpha_0} > 0$ for $\gamma_0 \neq 0$ and $x > 0$ for $\gamma_0 = 0$.

The shape parameter γ_0 is called the extreme value index for the exceedance $X - u_0|X > u_0$. When $\gamma_0 < 0$, there is a finite right endpoint $u^* = u_0 - \frac{\sigma_{\alpha_0}}{\gamma_0}$ in the support of the distribution of X, i.e. $F(x) = 1$ for all $x \geq u^*$. When $\gamma = 0$, $X - u_0|X > u_0$ has an exponential distribution with mean σ_{α_0}. When $\gamma_0 > 0$, $X - u_0|X > u_0$ has a heavy tail with up to $\frac{1}{\gamma_0}$-th finite moments. Note that we write σ_{α_0} instead of $\sigma(u_0)$.

Observe that for any higher threshold $u > u_0$, the exceedance $X - u|X > u$ again follows the generalized Pareto distribution with the same shape parameter γ_0 but a different scale parameter $\sigma_\alpha = \left(\frac{\alpha_0}{\alpha}\right)^{\gamma_0} \sigma_{\alpha_0}$, where $\alpha = 1 - F(u)$ is the exceeding probability. Specifically,

$$\mathbb{P}\left(X > u + x | X > u\right) = \begin{cases} \left(1 + \frac{\gamma_0 x}{\sigma_\alpha}\right)^{-1/\gamma_0}, & \gamma_0 \neq 0, \\ \exp\left(-\frac{x}{\sigma_\alpha}\right), & \gamma_0 = 0. \end{cases}$$

A direct calculation yields the $(1-p)$-th quantile of X, i.e., the VaR at level $1-p$ takes the form

$$\mathrm{VaR}_X(1-p) = Q(1-p) = \begin{cases} u + \frac{\sigma_\alpha}{\gamma_0}\left(\left(\frac{\alpha}{p}\right)^{\gamma_0} - 1\right) & \text{if} \quad \gamma_0 \neq 0, \\ u + \sigma_\alpha \log\left(\frac{\alpha}{p}\right) & \text{if} \quad \gamma_0 = 0, \end{cases} \tag{2.1}$$

for all given $p \in (0, \alpha_0)$.

As Assumption 1 above does not model the distribution below the threshold parametrically, computing VaR$(1-p)$ based on (2.1) is a semiparametric method and achieves a good balance between robustness and efficiency. It is easy to check that model (1.2) satisfies Assumption 1. Unlike the existing studies on fitting GPD to exceedances over a divergent threshold, we investigate the inference based on a non-divergent threshold. In this case, the threshold may play a role in quantifying the inference uncertainty of VaR in (2.1). On the other hand, if the threshold diverges fast enough such that the estimation bias is negligible, then the model approximation error is negligible. Hence, the developed method for fitting an exact Generalized Pareto distribution is valid for using a larger divergent threshold under the setting that the exceedance has an approximate Generalized Pareto distribution.

Suppose we have a random sample $X_1, \ldots, X_n$ from F satisfying Assumption 1. Let the order statistics be $X_{1:n} \leq \ldots \leq X_{n:n}$. Take a large threshold, say, u_n, either deterministic or random, corresponding to the sample exceeding probability

$$\widehat{\alpha}_n = \frac{1}{n}\sum_{i=1}^{n} \delta(X_i - u_n), \tag{2.2}$$

where $\delta(x) := \mathbb{1}(x > 0)$ denotes the step function taking value 1 on the positive line and value 0 otherwise. Denote the adaptive exceeding probability $\alpha_n = 1 - F(u_n)$, which may be either deterministic or random depending on our choice of the threshold u_n.

Given an exceedance $X_i - u_n = x > 0$, the log-likelihood function for the Pareto parameters $\boldsymbol{\nu} := (\gamma, \log\sigma)^T \in \mathbb{R}^2$ is given by

$$l(\boldsymbol{\nu}|x) = -\left\{\frac{1+\gamma}{\gamma}\log\left(1+\frac{\gamma x}{\sigma}\right) + \log\sigma\right\}.$$

Note that the above function is well defined for $\gamma = 0$ by continuity as

$$l\left((0, \log\sigma)^T|x\right) = -\frac{x}{\sigma} - \log\sigma.$$

Thus, the full log-likelihood function given $X_1 - u_n, \ldots, X_n - u_n$ becomes

$$\sum_{i=1}^n \delta(X_i - u_n) l(\boldsymbol{\nu}|X_i - u_n),$$

resulting in the score equations

$$\sum_{i=1}^n \delta(X_i - u_n)\frac{\partial l(\boldsymbol{\nu}|X_i - u_n)}{\partial\gamma} = \sum_{i=1}^n \delta(X_i - u_n) s_1(\boldsymbol{\nu}|X_i - u_n) = 0, \tag{2.3}$$

$$\sum_{i=1}^n \delta(X_i - u_n)\frac{\partial l(\boldsymbol{\nu}|X_i - u_n)}{\partial\log\sigma} = \sum_{i=1}^n \delta(X_i - u_n) s_2(\boldsymbol{\nu}|X_i - u_n) = 0, \tag{2.4}$$

where

$$s_1(\gamma, \log\sigma|x) = \frac{1}{\gamma^2}\left(\log\left(1+\gamma\frac{x}{\sigma}\right) - \frac{\gamma x/\sigma}{1+\gamma x/\sigma}\right) - \frac{x/\sigma}{1+\gamma x/\sigma},$$
$$s_2(\gamma, \log\sigma|x) = -1 + (1+\gamma)\frac{x/\sigma}{1+\gamma x/\sigma},$$

and for $\gamma = 0$ the above equations take the form

$$s_1(0, \log\sigma|x) = \frac{1}{2}\left(\frac{x}{\sigma}\right)^2 - \frac{x}{\sigma}, \quad s_2(0, \log\sigma|x) = -1 + \frac{x}{\sigma}.$$

In this paper, we only consider the regular case, i.e., $\gamma_0 > -\frac{1}{2}$ as in Davison and Smith (1990) and Drees et al. (2004), and it is often the case of $\gamma_0 > 0$ regarding heavy-tailed losses in insurance and finance; see also Bücher and Segers (2017) for more discussions. For dealing with the irregular case, i.e., $\gamma_0 \leq -1/2$, we refer to Smith (1985), Zhou (2009), Zhou (2010), and Peng and Qi (2009). Davison and Smith (1990) disregard the randomness of threshold while Drees et al. (2004) obtain the asymptotic normality of MLE for a divergent random threshold (i.e., $\bar{\alpha} = \bar{\alpha}(n) \to 0$ as $n \to \infty$) under (1.1), which holds for Assumption 1. Here, we present a universal asymptotic normality result under Assumption 1 in the sense of unifying the cases of using either a deterministic threshold or a random threshold:

Assumption 2 (Universal threshold statistic)*.* The threshold $u_n = u_n(X_1, \ldots, X_n)$ is an arbitrary measurable statistic such that $u_n \xrightarrow{\mathbb{P}} Q(1-\bar{\alpha})$ for some $\bar{\alpha} \in (0, \alpha_0)$.

The assumption above allows a flexible choice of the threshold u_n for the practitioners, who may choose a deterministic threshold based on prior knowledge, an order statistic based on the observations, or an even more sophisticated quantile estimator.

Normalizing the estimators with the adaptive values $\boldsymbol{\theta}_0^{(n)} = (\gamma_0, \log \sigma_{\alpha_n}, \log \alpha_n)$ rather than its limit $\boldsymbol{\theta}_0 = (\gamma_0, \log \sigma_{\bar{\alpha}}, \log \bar{\alpha})$, we have a unified inference procedure for a general threshold statistic u_n:

Theorem 1 (Universal inference for generalized Pareto parameters)**.** *Suppose that Assumption 1 holds with a true parameter $\gamma_0 > -\frac{1}{2}$ and the choice of sequence u_n satisfies Assumption 2.*

(i) With probability tending to 1, there exists a maximum likelihood estimator $\widehat{\boldsymbol{\theta}}_n = (\widehat{\gamma}_n, \log \widehat{\sigma}_n, \log \widehat{\alpha}_n)$, solving the score equations (2.2)– (2.4) simultaneously, in the local parameter space

$$\bar{\Theta}_n^{\varepsilon} = \left\{ \boldsymbol{\theta} \in \mathbb{R}^3 : \left\| \boldsymbol{\theta} - \boldsymbol{\theta}_0^{(n)} \right\| < n^{-1/2+\varepsilon} \right\}, \tag{2.5}$$

for any $\varepsilon \in (0, \min\{\gamma_0 + 1/2, 1/2\})$, where $\boldsymbol{\theta}_0^{(n)} = (\gamma_0, \log \sigma_{\alpha_n}, \log \alpha_n)$ denotes the adaptive true values.

(ii) Any maximum likelihood estimator sequence from part (i) is asymptotically normal in such a way that

$$\sqrt{n\bar{\alpha}} \left(\widehat{\gamma}_n - \gamma_0, \frac{\widehat{\sigma}_n}{\sigma_{\alpha_n}} - 1, \frac{\widehat{\alpha}_n}{\alpha_n} - 1 \right) \xrightarrow{d} \mathcal{N} \left(\boldsymbol{0}, \begin{bmatrix} \mathcal{I}^{-1} & 0 \\ 0 & 1 - \bar{\alpha} \end{bmatrix} \right)$$

where the inverse Fisher information matrix

$$\mathcal{I}^{-1} = \left(E \left(\frac{\partial l(\boldsymbol{\nu}_0|Z)}{\partial \boldsymbol{\nu}} \frac{\partial l(\boldsymbol{\nu}_0|Z)}{\partial \boldsymbol{\nu}^T} \middle| Z > 0 \right) \right)^{-1} = \begin{bmatrix} (1+\gamma_0)^2 & -(1+\gamma_0) \\ -(1+\gamma_0) & 2(1+\gamma_0) \end{bmatrix}$$

with $Z = X - Q(1-\bar{\alpha})$.

In practice, it is common to fix a proportion of data, say, $\bar{\alpha} \in (0, 1)$ and use the $[n\bar{\alpha}]$th largest observation $u_n = X_{n-[n\bar{\alpha}]:n}$ as the threshold. It is then easy to deduce the following corollary.

Corollary 1. *Under the conditions of Theorem 1 with $u_n = X_{n-[n\bar{\alpha}]:n}$, as $n \to \infty$,*

$$\sqrt{n\bar{\alpha}} \begin{pmatrix} \widehat{\gamma}_n - \gamma_0 \\ \frac{\widehat{\sigma}_n}{\sigma_{\bar{\alpha}}} - 1 \\ \frac{\widehat{\alpha}_n}{\bar{\alpha}} - 1 \end{pmatrix} \xrightarrow{d} \mathcal{N}\left(\boldsymbol{0}, \begin{bmatrix} (1+\gamma_0)^2 & -(1+\gamma_0) & 0 \\ -(1+\gamma_0) & 2(1+\gamma_0)+\gamma_0^2(1-\bar{\alpha}) & -\gamma_0(1-\bar{\alpha}) \\ 0 & -\gamma_0(1-\bar{\alpha}) & 1-\bar{\alpha} \end{bmatrix} \right).$$

Remark 1. Our asymptotic variance formula is unified for the threshold being finite or divergent and deterministic or random when the exceedance follows a generalized Pareto distribution. In the supplement, we deduce that the results remain true if $\bar{\alpha} = \bar{\alpha}_n$ is an intermediate sequence such that $\bar{\alpha} \to 0$ and $n\bar{\alpha} \to \infty$. If we rewrite Assumption 2 as $u_n/Q(1-\bar{\alpha}) \xrightarrow{\mathbb{P}} 1$, one may approximate $\bar{\alpha}$ by its limit 0 if necessary. When $\bar{\alpha}$ is vanishing, it is easy to allow the true threshold α_0 to vanish as well as long as $\bar{\alpha}/\alpha_0$ is bounded strictly below 1. It is clear from the proof that our results remain true if the approximation error between the exceedance distribution and a generalized Pareto distribution is a smaller order than the parametric rate. More specifically, suppose our observations $(X_1^{(n)}, \dots, X_n^{(n)})$ come from an triangular array of i.i.d. random variables and denote their common distribution as $F^{(n)}$. Our inference remains valid if our generalized Pareto model is approximately true, that is,

$$\sup_{x \geq u_0} \left| \frac{1 - F_{u_0}^{(n)}(x)}{1 - G_{\gamma,\sigma_{\alpha_0}}(x)} - 1 \right| = o((n\bar{\alpha})^{-1/2}), \tag{2.6}$$

where $F_{u_0}^{(n)}(x) = \frac{F^{(n)}(x) - F^{(n)}(u_0)}{\alpha_0}$ denotes the excess distribution function, the exceeding probability $\alpha_0 = 1 - F^{(n)}(u_0)$ may be either fixed or vanishing, and $G_{\gamma,\sigma_{\alpha_0}}$ denotes the generalized Pareto distribution with the shape parameter γ and scale parameter σ_{α_0}. For example, consider the universal approximation (1.1) and take a sufficiently large sequence $u = u(n) \to \infty$ as $n \to \infty$, our results remain true for the array data

$$X_i^{(n)} = \frac{X_i - u}{\beta(u)}, \; i = 1, \dots, n,$$

which satisfies condition (2.6) under the high-order regular variation conditions for X_i as shown in, e.g., Drees et al. (2004). In summary, our fixed-$\bar{\alpha}$ approach in the above theorem is robust and covers more practical applications.

Remark 2. As argued by Dombry (2015), there is no guarantee that the global maximum likelihood estimator is unique. Even if a global MLE is attainable, the classical regularity conditions in Cramér (1946) are not fulfilled, and it requires a detailed verification of the local asymptotic normality (LAN) conditions in Bücher and Segers (2017). Also, the global estimation theory in the above paper does not apply as our 'true' values $\boldsymbol{\theta}_0^{(n)}$ are a sequence of adaptive values depending on the (random) threshold statistic rather than a fixed point. Therefore, we consider a local maximum likelihood estimator and leave the global estimation theory for future research. Note that this challenge remains for a divergent threshold, as the asymptotic normality results in, e.g., Drees et al. (2004) are not guaranteed to hold for an arbitrary global estimator sequence; see, e.g., Zhou (2009) and Zhou (2010) for comments.

Plugging the estimator $(\widehat{\gamma}_n, \widehat{\sigma}_n, \widehat{\alpha}_n)$ from Theorem 1 in VaR formula (2.1), the MLE of $\mathrm{VaR}_X(1-p)$ is given by

$$\widehat{\mathrm{VaR}}_X(1-p) = u_n + \frac{\widehat{\sigma}_n}{\widehat{\gamma}_n}\left(\left(\frac{\widehat{\alpha}_n}{p}\right)^{\widehat{\gamma}_n} - 1\right). \tag{2.7}$$

It should be interpreted as $\widehat{\mathrm{VaR}}_X(1-p) = u_n + \widehat{\sigma}_n \log\left(\frac{\widehat{\alpha}_n}{p}\right)$ by continuity if $\widehat{\gamma}_n = 0$. The asymptotic normality of the quantile estimator (2.7) then follows directly from the continuous mapping theorem, since we can expand the true quantile in (2.1) similarly by

$$\mathrm{VaR}_X(1-p) = u_n + \frac{\sigma_{\alpha_n}}{\gamma_0}\left(\left(\frac{\alpha_n}{p}\right)^{\widehat{\gamma}_n} - 1\right), \tag{2.8}$$

even with a random adaptive exceeding probability α_n, conditional on the event $\alpha_n > \alpha_0$, which occurs with probability tending to 1. Again, our quantile inference is asymptotically correct for a universal threshold statistic.

Theorem 2 (Universal inference for high quantile)**.** *Under the conditions of Theorem 1, for every $p \in (0, \alpha_0)$,*

$$\frac{\sqrt{n\bar{\alpha}}}{\sigma_p}\left(\widehat{\mathrm{VaR}}_X(1-p) - \mathrm{VaR}_X(1-p)\right) \xrightarrow{d} \mathcal{N}\left(0, q\left(\frac{\bar{\alpha}}{p}\right)^T \mathcal{I}^{-1} q\left(\frac{\bar{\alpha}}{p}\right) + 1 - \bar{\alpha}\right)$$

as $n \to \infty$, where for $\gamma_0 \neq 0$, the vector function

$$q(t) = \left(\int_1^t \left(\frac{s}{t}\right)^{\gamma_0} \frac{\log s}{s}\, ds,\ \frac{1 - t^{-\gamma_0}}{\gamma_0}\right)^T, \quad t > 0,$$

and it should be interpreted by continuity as $\left(\frac{1}{2}(\log t)^2, \log t\right)^T$ when $\gamma_0 = 0$.

Remark 3. Ignoring all common factors, one may search for the best threshold as $u_n = X_{n-[n\lambda p]:n}$ with λ in a neighborhood of 1 minimizing the asymptotic variance

$$\frac{1}{\lambda}\left(\widehat{q}(\lambda)^T\widehat{\mathcal{I}}^{-1}\widehat{q}(\lambda)+1\right),$$

where $\widehat{\mathcal{I}}^{-1}$ and $\widehat{q}$ may be constructed using some preliminary estimate $\widehat{\gamma}$ of the extreme value index γ_0 as given below. If necessary, one may update $\widehat{\gamma}$ with the new choice of λ until convergence. On the other hand, it is important to develop a distribution-free goodness-of-fit test for fitting a generalized Pareto distribution to exceedances over a threshold. It is challenging to extend the existing parametric testing methods in, e.g., Koul and Ling (2006) to our semi-parametric models, which will be our future research.

It is straightforward to quantify the uncertainty of $\widehat{\mathrm{VaR}}_X(1-p)$ based on the normal approximation. More specifically, we estimate $\bar{\alpha}$ by $\widehat{\alpha}_n$ (if $\bar{\alpha}$ is unknown), the scale parameter σ_p by

$$\widehat{\sigma}_p = \widehat{\sigma}_n\left(\widehat{\alpha}_n/p\right)^{\widehat{\gamma}_n},$$

and the limiting variance by

$$\widehat{\tau}_n^2 := q\left(\frac{\widehat{\alpha}_n}{p}\right)^T\widehat{\mathcal{I}}^{-1}q\left(\frac{\widehat{\alpha}_n}{p}\right)+1-\widehat{\alpha}_n$$

with

$$\widehat{\mathcal{I}}^{-1} = \begin{bmatrix} (1+\widehat{\gamma}_n)^2 & -(1+\widehat{\gamma}_n) \\ -(1+\widehat{\gamma}_n) & 2(1+\widehat{\gamma}_n)\end{bmatrix}.$$

Hence, a normal approximation confidence interval of $\mathrm{VaR}_X(1-p)$ with level a is

$$I_{NA}(a) = \left[\widehat{\mathrm{VaR}}_X(1-p) - \frac{z_{(1+a)/2}}{\sqrt{n\widehat{\alpha}_n}}\widehat{\sigma}_n\left(\frac{\widehat{\alpha}_n}{p}\right)^{\widehat{\gamma}_n}\widehat{\tau}_n,\ \widehat{\mathrm{VaR}}_X(1-p) + \frac{z_{(1+a)/2}}{\sqrt{n\widehat{\alpha}_n}}\widehat{\sigma}_n\left(\frac{\widehat{\alpha}_n}{p}\right)^{\widehat{\gamma}_n}\widehat{\tau}_n\right],$$

where $z_{(1+a)/2}$ is the $\frac{1+a}{2}$-quantile of the standard normal distribution. Unfortunately, our simulation study below shows that this interval has a poor coverage probability in small samples, which calls for more efficient methods.

To improve finite-sample coverage, we propose a resampling method called the random weighted bootstrap procedure. The random weighted bootstrap method is less computationally intensive than the naive bootstrap method when we estimate risk based on a time series model (in the next section). Zhu (2016, 2019) recently applies this method to conduct a Portmanteau test and infer autoregressive models.

- *Step B1)* Draw a random sample with sample size n from a distribution function with mean one and variance one such as the standard exponential distribution, say $\xi_1^{(b)}, \cdots, \xi_n^{(b)}$.

- *Step B2)* Choose a threshold statistic $u_n^{(b)}$, possibly dependent on $\xi_1^{(b)}, \ldots, \xi_n^{(b)}$. Solve the following random weighted score equations for $\bar{\alpha}$, γ, and $\log \sigma$:

$$\sum_{i=1}^{n} \xi_i \left(\delta(X_i - u_n^{(b)}) - \bar{\alpha} \right) = 0 \tag{2.9}$$

$$\sum_{i=1}^{n} \xi_i \delta(X_i - u_n^{(b)}) s_1(\boldsymbol{\nu} | X_i - u_n^{(b)}) = 0, \tag{2.10}$$

$$\sum_{i=1}^{n} \xi_i \delta(X_i - u_n^{(b)}) s_2(\boldsymbol{\nu} | X_i - u_n^{(b)}) = 0. \tag{2.11}$$

  Denote these estimators by $\widehat{\alpha}_n^{(b)}$, $\widehat{\gamma}_n^{(b)}$, and $\widehat{\sigma}_n^{(b)}$, we have

$$\widehat{\mathrm{VaR}}_X^{(b)}(1-p) = u_n^{(b)} + \frac{\widehat{\sigma}_n^{(b)}}{\widehat{\gamma}_n^{(b)}}((\widehat{\alpha}_n^{(b)}/p)^{\widehat{\gamma}_n^{(b)}} - 1).$$

- *Step B3)* Repeat the above two steps B times to obtain $\{\widehat{\mathrm{VaR}}_X^{(b)}(1-p)\}_{b=1}^{B}$. Let $\bar{D}_{1:B} \le \cdots \le \bar{D}_{B:B}$ denote the order statistics of

$$\log\left(\frac{\widehat{\mathrm{VaR}}_X^{(b)}(1-p)}{\widehat{\mathrm{VaR}}_X(1-p)} \right), \; b = 1, \ldots, B,$$

  and let $\bar{D}_{(1)} \le \cdots \le \bar{D}_{(B)}$ denote the order statistics of

$$\left| \log\left(\frac{\widehat{\mathrm{VaR}}_X^{(b)}(1-p)}{\widehat{\mathrm{VaR}}_X(1-p)} \right) \right|, \; b = 1, \ldots, B,$$

  Hence, the confidence intervals with level a for $\log(\mathrm{VaR}_X(1-p))$ are

$$I_{RWB,1}(a) = \left[\log(\widehat{\mathrm{VaR}}_X(1-p)) - \bar{D}_{[\frac{B+Ba}{2}]:B}, \quad \log(\widehat{\mathrm{VaR}}_X(1-p)) - \bar{D}_{[\frac{B-Ba}{2}]:B} \right]$$

  and

$$I_{RWB,2}(a) = \left[\log(\widehat{\mathrm{VaR}}_X(1-p)) - \bar{D}_{(Ba)}, \quad \log(\widehat{\mathrm{VaR}}_X(1-p)) + \bar{D}_{(Ba)} \right].$$

The following theorem establishes the validity of our random weighted bootstrap method.

Theorem 3 (Random weighted bootstrap)**.** *Suppose the conditions of Theorem 1 hold. For an arbitrary bootstrap threshold statistic* $u_n^{(b)} = u_n + o_{\mathbb{P}}(1)$ *and let* $\alpha_n^{(b)} = 1 - F(u_n^{(b)})$*:*

(i) With probability tending to 1, there exists a random weighted maximum likelihood estimator $\widehat{\boldsymbol{\theta}}_n^{(b)} = (\widehat{\gamma}_n^{(b)}, \log\widehat{\sigma}_n^{(b)}, \log\widehat{\alpha}_n^{(b)})$*, solving the score equations* (2.9)– (2.11) *simultaneously in the local parameter space*

$$\Theta_{\varepsilon}^{(b)} = \left\{\boldsymbol{\theta} \in \mathbb{R}^3 : \left\|\boldsymbol{\theta} - \boldsymbol{\theta}_0^{(b)}\right\| < n^{-1/2+\varepsilon}\right\}, \tag{2.12}$$

for any $\varepsilon \in (0, \min\{\gamma_0 + 1/2, 1/2\})$*, where* $\boldsymbol{\theta}_0^{(b)} = \left(\gamma_0, \log\sigma_{\alpha_n^{(b)}}, \log\alpha_n^{(b)}\right)$ *denotes the adaptive true values.*

(ii) For each probability level $a \in (0,1)$*,*

$$\mathbb{P}\left(\widehat{\mathrm{VaR}}_X(1-p) - \mathrm{VaR}_X(1-p) \le c_n(a)\right) \to 1 - a,$$

where

$$c_n(a) = \inf\left\{x : \mathbb{P}\left(\widehat{\mathrm{VaR}}^{(b)}(1-p) - \widehat{\mathrm{VaR}}_X(1-p) \le x | X_1, \ldots, X_n\right) > 1 - a\right\}.$$

The result remains true if $\mathrm{VaR}_X(1-p)$*,* $\widehat{\mathrm{VaR}}_X(1-p)$*, and* $\widehat{\mathrm{VaR}}^{(b)}(1-p)$ *are substituted by their logarithms, provided that* $\mathrm{VaR}_X(1-p) > 0$*.*

Remark 4. The random weighted bootstrap intervals for the extreme value index γ_0 is also valid. For each probability level $a \in (0,1)$, $\mathbb{P}\left(\widehat{\gamma}_n - \gamma_0 \le c_{n,\gamma}(a)\right) \to 1 - a$, where

$$c_{n,\gamma}(a) = \inf\left\{x : \mathbb{P}\left(\widehat{\gamma}_n^{(b)} - \widehat{\gamma}_n \le x | X_1, \ldots, X_n\right) > 1 - a\right\}.$$

The result remains true if we substitute γ_0, $\widehat{\gamma}_n$ and $\widehat{\gamma}_n^{(b)}$ by their logarithms, provided that $\gamma_0 > 0$. The random weighted bootstrap intervals for the adaptive scale parameter σ_{α_n} and the adaptive exceeding probability α_n, are asymptotically correct if the difference between the bootstrap threshold and the original threshold is asymptotically negligible in the sense that $u_n^{(b)} = u_n + o_{\mathbb{P}}((n\bar{\alpha})^{-1/2})$.

Remark 5 (Naive bootstrap). In the supplement, we show that Theorem 3 and Remark 4 remain true if replacing the random weighted bootstrap statistics with the naive bootstrap statistics. In simulations, we observe comparable performance between these two methods for independent data.

3 Methodologies and Asymptotic Results for ARMA-GARCH Models

It is reasonable to assume that insurance losses happen independently, but financial losses often exhibit some stylized facts such that time dependence, heavy tail, skewness, and persistent volatility; see, e.g., the survey Cont (2001). Since the seminal work of Engle (1982) and Bollerslev (1986), it has become a common practice to model a financial time series by an ARMA-GARCH model given by

$$\begin{cases} Y_t = \mu + \sum_{i=1}^{q_1} \phi_i Y_{t-i} + \sum_{j=1}^{q_2} \psi_j \varepsilon_{t-j} + \varepsilon_t \\ \varepsilon_t = \sqrt{\underline{h}_t}\eta_t, \quad \underline{h}_t = \underline{\omega} + \sum_{i=1}^{r} \underline{a}_i \varepsilon_{t-i}^2 + \sum_{j=1}^{s} \underline{b}_j \underline{h}_{t-j}, \end{cases} \tag{3.1}$$

where $\underline{\omega} > 0, \underline{a}_i \geq 0, \underline{b}_j \geq 0$, and $\{\eta_t\}$ is a sequence of i.i.d. random variables with zero mean and variance one. In this case, the so-called one-step ahead conditional VaR is more useful in forecasting risk and is defined as the conditional quantile of Y_{n+1} given the past information up to time n, i.e., $\mathcal{F}_n = \sigma(\dots, Y_{n-1}, Y_n)$. Hence, the one-step ahead conditional VaR is

$$\mathrm{VaR}_{Y,n}(1-p) = \mu + \sum_{i=1}^{q_1} \phi_i Y_{n+1-i} + \sum_{j=1}^{q_2} \psi_j \varepsilon_{n+1-j} + \sqrt{\underline{h}_{n+1}} \mathrm{VaR}_\eta(1-p), \tag{3.2}$$

and note that $\underline{h}_{n+1}$ is $\mathcal{F}_n-$measurable.

We remark that McNeil and Frey (2000) consider the model above, and Martins-Filho et al. (2018) study the nonparametric regression, which covers AR-ARCH models but not ARMA-GARCH models. Both papers only consider the case of divergent risk level, i.e., $p = p(n) \to 0$ as $n \to \infty$. In this case, the estimation for the ARMA-GARCH model in McNeil and Frey (2000) and the kernel smoothing estimation for the conditional mean and conditional standard deviation in Martins-Filho et al. (2018) do not play a role in the asymptotic variance of the VaR estimation. Unlike these two papers, we aim to allow both fixed and divergent risk levels and consider the uncertainties in fitting both the ARMA-GARCH model with fewer finite moments and the GPD to residuals.

As aforementioned, regulators often set p close to one, making it useful to model η_t over a high threshold by a GPD parametrically. To infer the above conditional VaR, we need to estimate the unknown parameters in (3.1) and (2.1). An obvious inference method for model (3.1) is the so-called quasi maximum likelihood estimation. The asymptotic normality of the quasi-Gaussian maximum likelihood estimator is available in Francq and Zakoïan (2004), which requires finite fourth moments of both ε_t and η_t. However, in practice, it is quite often that $\sum_{i=1}^r a_i + \sum_{j=1}^s b_j$ is close to one, making it problematic to assume $\mathbb{E}\varepsilon_t^4 < \infty$. When $\mathbb{E}\eta_t^4 < \infty$, Ling (2007) proposes a weighted quasi-maximum likelihood estimation to allow $\mathbb{E}\varepsilon_t^4 = \infty$ for having a normal limiting distribution. However, the asymptotic normality of these estimators may be lost when $\mathbb{E}\eta_t^4 = \infty$, see, e.g., Hall and Yao (2003). To further allow both $\mathbb{E}\eta_t^4 = \infty$ and $\mathbb{E}\varepsilon_t^4 = \infty$, Zhu and Ling (2011) proposed a self-weighted exponential likelihood estimation method, which has a normal limiting distribution and requires $E|\eta_t| = 1$ and zero median of η_t. Changing $\mathbb{E}\eta_t^2 = 1$ to $\mathbb{E}|\eta_t| = 1$ requires a scale transformation of h_t, which does not affect the inference of the conditional VaR, however, changing zero mean of η_t to zero median involves a shift transformation, which makes the inference of the conditional VaR infeasible.

Here, we instead propose a three-step inference of the conditional VaR (3.2) under model (3.1), which allows both $\mathbb{E}\varepsilon_t^4 = \infty$ and $\mathbb{E}\eta_t^4 = \infty$. This is important in estimating $\mathrm{VaR}_{Y,n}(1-p)$ when p is treated as a fixed number rather than a number converging to zero as $n \to \infty$.

We assume that $\mathbb{E}|\eta_t| = d > 0$ unknown and put $X_t = \eta_t/d$, $h_t = d^2\underline{h}_t$, $\omega = \underline{\omega}d^2$, $a_i = \underline{a}_i d^2$, and $b_j = \underline{b}_j$. Then (3.1) is equivalent to

$$\begin{cases} Y_t = \mu + \sum_{i=1}^{q_1} \phi_i Y_{t-i} + \sum_{j=1}^{q_2} \psi_j \varepsilon_{t-j} + \varepsilon_t \\ \varepsilon_t = \sqrt{h_t} X_t, \quad h_t = \omega + \sum_{i=1}^r a_i \varepsilon_{t-i}^2 + \sum_{j=1}^s b_j h_{t-j}, \end{cases} \tag{3.3}$$

where $\mathbb{E}|X_t| = \mathbb{E}|\eta_t|/d = 1$. The coefficients remain the same in the ARMA model for Y_t, and we can rewrite (3.2) as

$$\mathrm{VaR}_{Y,n}(1-p) = \mu + \sum_{i=1}^{q_1} \phi_i Y_{n+1-i} + \sum_{j=1}^{q_2} \psi_j \varepsilon_{n+1-j} + \sqrt{h_{n+1}}\mathrm{VaR}_X(1-p).$$

Observe that (3.3) is the model studied in Zhu and Ling (2011), but here we maintain the zero mean condition on X_t as required by the original model (3.1).

Let $\boldsymbol{\psi} = (\boldsymbol{\phi}^T, \boldsymbol{\phi}_h^T)^T$ denote the parameters in (3.3) with $\boldsymbol{\phi} = (\mu, \phi_1, \ldots, \phi_{q_1}, \psi_1, \ldots, \psi_{q_2})^T$ and $\boldsymbol{\phi}_h = (\omega, a_1, \ldots, a_r, b_1, \ldots, b_s)^T$. Before moving on to the quantile inference, we first develop a two-step estimator of $\boldsymbol{\psi}$ that is asymptotic normal without requiring any fourth moment condition.

Given the observations $Y_1, \dots, Y_n$ and the initial value $\bar{Y}_0 = \{Y_t : t \le 0\}$ generated by model (3.1), we can write the parametric model (3.3) as

$$\varepsilon_t(\boldsymbol{\phi}) = Y_t - \mu - \sum_{i=1}^{q_1} \phi_i Y_{t-i} - \sum_{j=1}^{q_2} \psi_j \varepsilon_{t-j}(\boldsymbol{\phi}),\ \ h_t(\boldsymbol{\psi}) = \omega + \sum_{i=1}^{r} a_i \varepsilon_{t-i}^2(\boldsymbol{\phi}) + \sum_{j=1}^{s} b_j h_{t-j}(\boldsymbol{\psi}),$$

$$X_t(\boldsymbol{\psi}) = \varepsilon_t(\boldsymbol{\phi})/\sqrt{h_t(\boldsymbol{\psi})}.$$

Obviously, $\varepsilon_t = \varepsilon_t(\boldsymbol{\phi}_0)$, $h_t = h_t(\boldsymbol{\psi}_0)$, and $X_t = X_t(\boldsymbol{\psi}_0)$, where $\boldsymbol{\phi}_0$ and $\boldsymbol{\psi}_0 = (\boldsymbol{\phi}_0^T, \boldsymbol{\phi}_{h0}^T)^T$ denote the true values of the parameters. In practice, however, we do not have the initial values $\bar{Y}_0 = \{Y_t : t \le 0\}$, which makes the calculation of $\varepsilon_t(\boldsymbol{\phi})$, $h_t(\boldsymbol{\psi})$ and $X_t(\boldsymbol{\psi})$ infeasible. To make the estimation feasible, in what follows, we replace $\bar{Y}_0 = \{Y_t : t \le 0\}$ by zeros like Ling (2007) and Zhu and Ling (2011) and instead define the feasible parametric model $\widetilde{\varepsilon}_t(\boldsymbol{\phi})$, $\widetilde{h}_t(\boldsymbol{\psi})$, $\widetilde{X}_t(\boldsymbol{\psi})$ based on the new initial values.

First, we estimate $\boldsymbol{\phi}$ by the following self-weighted least squares estimator

$$\widehat{\boldsymbol{\phi}} = \arg\min_{\boldsymbol{\phi}} \sum_{t=1}^{n} \widetilde{w}_t^2 \widetilde{\varepsilon}_t^2(\boldsymbol{\phi}), \tag{3.4}$$

where $\{\widetilde{w}_t\}$ are some proper weights designed to reduce the moment effect of $\{h_t\}$, and $\widetilde{\varepsilon}_t(\boldsymbol{\phi})$ is the feasible parametric model as defined above. The key idea in constructing such a weight function w_t is to bound $\sum_{i=1}^{\infty} \rho^i |Y_{t-i}|$ for some $\rho \in (0,1)$ and ensure that it is well-defined for all $t \ge 1$, where ρ depends on the underlying ARMA-GARCH model. Same as in He et al. (forthcoming), we use the feasible weight

$$\widetilde{w}_t = \left\{ \max\left(1, \sum_{i=1}^{t-1} e^{-\log^2(i+1)} |Y_{t-i}|\right) \right\}^{-1},$$

which is a truncated version of the oracle weight

$$w_t = \left\{ \max\left(1, \sum_{i=1}^{\infty} e^{-\log^2(i+1)} |Y_{t-i}|\right) \right\}^{-1}.$$

Second, we define the self-weighted estimator $\widehat{\boldsymbol{\phi}}_h$ of $\boldsymbol{\phi}_h$, which minimizes the self-weighted negative log quasi-exponential-likelihood

$$\sum_{t=1}^{n} \widetilde{w}_t^4 \widetilde{l}_t\left(\boldsymbol{\phi}_h | \widehat{\boldsymbol{\phi}}\right), \quad \widetilde{l}_t(\boldsymbol{\phi}_h|\widehat{\boldsymbol{\phi}}) = \log \sqrt{\widetilde{h}_t\left(\widehat{\boldsymbol{\phi}}, \boldsymbol{\phi}_h\right)} + \frac{\left|\widetilde{\varepsilon}_t(\widehat{\boldsymbol{\phi}})\right|}{\sqrt{\widetilde{h}_t\left(\widehat{\boldsymbol{\phi}}, \boldsymbol{\phi}_h\right)}}, \tag{3.5}$$

where $\widehat{\boldsymbol{\phi}}$ and $\widetilde{w}_t$ are the least-squares estimator and self-weights from the first step, respectively, and $\widetilde{h}_t(\boldsymbol{\phi}, \boldsymbol{\phi}_h)$ is the feasible parametric model.

To establish the joint asymptotic normality $\widehat{\boldsymbol{\psi}} = \left(\widehat{\boldsymbol{\phi}}^T, \widehat{\boldsymbol{\phi}}_h^T\right)^T$, we need the following additional regularity conditions.

A1. Let $\Theta^{\boldsymbol{\psi}} = \Theta^{\boldsymbol{\phi}} \times \Theta^{\boldsymbol{\phi}_h} \subset \mathbb{R}^{q_1+q_2+1} \times [0, \infty)^{r+s+1}$ denote the parameter space for $\boldsymbol{\psi} = \left(\boldsymbol{\phi}^T, \boldsymbol{\phi}_h^T\right)^T$. Assume that $\Theta^{\boldsymbol{\psi}}$ is compact and the true value of $\boldsymbol{\psi}$ is an interior point.

A2. For each $\boldsymbol{\phi} \in \Theta^{\boldsymbol{\phi}}$, $1 - \sum_{i=1}^{q_1} \phi_i z^i \neq 0$ and $1 + \sum_{j=1}^{q_2} \psi_j z^j \neq 0$ when $|z| \leq 1$, and $1 - \sum_{i=1}^{q_1} \phi_i z^i = 0$ and $1 + \sum_{j=1}^{q_2} \psi_j z^j = 0$ have no common root with $\phi_{q_1} \neq 0$ and $\psi_{q_2} \neq 0$.

A3. For each $\boldsymbol{\phi}_h \in \Theta^{\boldsymbol{\phi}_h}$, there is no common root for equations $\sum_{i=1}^{r} a_i z^i = 0$ and $\sum_{j=1}^{s} b_j z^j = 0$. Further, $\sum_{i=1}^{r} a_i \neq 0$, $a_r + b_s \neq 0$, and $\sum_{j=1}^{s} b_j < 1$.

A4. $\mathbb{E}\varepsilon_t^2 < \infty$.

A5. $X_t = \eta_t / \mathbb{E}|\eta_t|$ and $\{\eta_t\}_{t=1}^n$ is a sequence of independent and identically distributed random variables with mean zero, variance one, and continuous density function f such that $f(0) > 0$ and $\sup_{x\in\mathbb{R}} f(x) < \infty$.

Conditions A1–A3 are standard stationarity, invertibility, and identification conditions for ARMA-GARCH model (3.3) as in, e.g., Ling (2007). Condition A4 is equivalent to requiring that there is no unit root in the underlying GARCH process (3.1), that is,

$$\sum_{i=1}^{r} \underline{a}_i + \sum_{j=1}^{s} \underline{b}_j < 1.$$

By carefully checking our proofs, it can be seen that we may further relax the condition down to the first moment, that is, $\mathbb{E}|\varepsilon_t| < \infty$. This means that our results extend to IGARCH model with $\sum_{i=1}^{r} \underline{a}_i + \sum_{j=1}^{s} \underline{b}_j = 1$ under suitable conditions; see, e.g., part (iii) of Theorem 2.1 in Ling (2007). The second moment condition simplifies our later inference for the generalized Pareto model, and therefore we keep it throughout for simplicity. Condition A5 is similar to Assumption 2.6 in Zhu and Ling (2011), but we maintain the natural condition that X_t has a zero mean rather than a zero median.

Theorem 4. *Assume conditions A1–A5 hold.*

(i) The self-weighted estimator is consistent, that is, $\widehat{\boldsymbol{\psi}} := \left(\widehat{\boldsymbol{\phi}}^T, \widehat{\boldsymbol{\phi}}_h^T\right)^T \xrightarrow{\mathbb{P}} \boldsymbol{\psi}_0$.

(ii) The self-weighted estimator is asymptotic normal in such a way that

$$\sqrt{n}\left(\widehat{\boldsymbol{\psi}} - \boldsymbol{\psi}_0\right) \xrightarrow{d} \mathcal{N}\left(0, \Sigma^{-1}\Omega(\Sigma^{-1})^T\right),$$

with

$$\Sigma = \begin{bmatrix} \Sigma_1 & 0 \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}, \quad \Omega = \begin{bmatrix} \Omega_{11} & \Omega_{21}^T \\ \Omega_{21} & \Omega_{22} \end{bmatrix},$$

where the sub-matrices

$$\Sigma_1 = \mathbb{E}\left(w_t^2 \frac{\partial \varepsilon_t}{\partial \boldsymbol{\phi}} \frac{\partial \varepsilon_t}{\partial \boldsymbol{\phi}^T}\right), \ \Sigma_{21} = \frac{1}{8}\mathbb{E}\left(\frac{w_t^4}{h_t^2} \frac{\partial h_t}{\partial \boldsymbol{\phi}_h} \frac{\partial h_t}{\partial \boldsymbol{\phi}^T}\right), \Sigma_{22} = \frac{1}{8}\mathbb{E}\left(\frac{w_t^4}{h_t^2} \frac{\partial h_t}{\partial \boldsymbol{\phi}_h} \frac{\partial h_t}{\partial \boldsymbol{\phi}_h^T}\right),$$

$$\Omega_{11} = \mathbb{E}X_t^2 \cdot \mathbb{E}\left[w_t^4 h_t \frac{\partial \varepsilon_t}{\partial \boldsymbol{\phi}} \frac{\partial \varepsilon_t}{\partial \boldsymbol{\phi}^T}\right], \Omega_{22} = \frac{\mathbb{E}X_t^2 - 1}{4} \cdot \mathbb{E}\left[\frac{w_t^8}{h_t^2} \frac{\partial h_t}{\partial \boldsymbol{\phi}_h} \frac{\partial h_t}{\partial \boldsymbol{\phi}_h^T}\right], \ \textit{and}$$

$$\Omega_{21} = \mathbb{E}\left[X_t^2 \left(\mathbb{1}(X_t > 0) - \mathbb{1}(X_t < 0)\right)\right] \cdot \mathbb{E}\left[\frac{w_t^6}{2\sqrt{h_t}} \frac{\partial h_t}{\partial \boldsymbol{\phi}_h} \frac{\partial \varepsilon_t}{\partial \boldsymbol{\phi}^T}\right].$$

Next, we estimate the high quantile of X_t under the generalized Pareto model with the additional assumption

A6. $X_t \equiv \eta_t / \mathbb{E}|\eta_t|$ satisfies Assumption 1 with $\gamma_0 \in (0, \frac{1}{2})$ and scale parameter $\sigma_{\alpha_0} > 0$.

Note that $\gamma_0 < 1/2$ above ensures $\mathbb{E}\eta_t^2 < \infty$. Let $\widehat{X}_{1:n} \leq \ldots \leq \widehat{X}_{n:n}$ denote the order statistics of residuals $\{\widehat{X}_t := \widetilde{X}_t(\widehat{\boldsymbol{\psi}}) : t = 1, \ldots, n\}$, where $\widetilde{X}_t(\cdot)$ is the feasible parametric model as defined above. We then choose a threshold statistic such as

$$u_n = \widehat{X}_{n-[n\bar{\alpha}]:n}, \tag{3.6}$$

corresponding to an adaptive tail probability level $\alpha_n = 1 - F(u_n)$, where F denotes the distribution function of X_t. Under the conditions of Theorem 4, we show that the threshold estimator (3.6) is consistent, that is,

$$u_n \xrightarrow{\mathbb{P}} Q(1 - \bar{\alpha}), \text{ and equivalently } \alpha_n \xrightarrow{\mathbb{P}} \bar{\alpha}, \tag{3.7}$$

where $Q(\cdot) = F^{\leftarrow}(\cdot)$ denotes the quantile function of X_t. In general, our theory allows an arbitrary threshold statistic u_n that satisfies our Assumption 2 above. With a general threshold statistic, we estimate the adaptive exceeding probability α_n, the shape parameter γ_0 and the scale parameter σ_{α_n} by solving equations (2.2)–(2.4) with X_i therein replaced by the residual $\widehat{X}_t$ here. Denote the estimators by $\widehat{\alpha}$, $\widehat{\gamma}$, and $\widehat{\sigma}$ respectively, which gives the quantile estimator

$$\widehat{\mathrm{VaR}}_X(1-p) = u_n + \frac{\widehat{\sigma}}{\widehat{\gamma}}\left(\left(\frac{\widehat{\alpha}}{p}\right)^{\widehat{\gamma}} - 1\right),$$

Thus the estimator for $\mathrm{VaR}_{Y,n}(1-p)$ is given by

$$\widehat{\mathrm{VaR}}_{Y,n}(1-p) = \widehat{\mu} + \sum_{i=1}^{q_1} \widehat{\phi}_i Y_{n+1-i} + \sum_{j=1}^{q_2} \widehat{\psi}_j \widetilde{\varepsilon}_{n+1-j}(\widehat{\boldsymbol{\phi}}) + \sqrt{\widetilde{h}_t\left(\widehat{\boldsymbol{\phi}}, \widehat{\boldsymbol{\phi}}_h\right)}\widehat{\mathrm{VaR}}_X(1-p). \quad (3.8)$$

Note that $u_n = u_n(\widehat{\boldsymbol{\psi}})$, $\widehat{\gamma} = \widehat{\gamma}(\widehat{\boldsymbol{\psi}})$ and $\widehat{\sigma} = \widehat{\sigma}(\widehat{\boldsymbol{\psi}})$ all depend on the self-weighted estimator $\widehat{\boldsymbol{\psi}}$, whose effects do not fade away for any finite $p \in (0,1)$.

Theorem 5. *Assume conditions A1-A6 hold.*

(i) With probability tending to 1, there exists a maximum likelihood estimator $\widehat{\boldsymbol{\theta}} = (\widehat{\gamma}, \log\widehat{\sigma}, \log\widehat{\alpha})$ *solving the score equations* (2.2)– (2.4) *simultaneously for* $\{\widehat{X}_t\}$ *in the local parameter space*

$$\bar{\Theta}_n^{\varepsilon} = \left\{\boldsymbol{\theta} \in \mathbb{R}^3 : \left\|\boldsymbol{\theta} - \boldsymbol{\theta}_0^{(n)}\right\| < n^{-1/2+\varepsilon}\right\},$$

for any $\varepsilon \in (0, \min\{\gamma_0 + 1/2, 1/2\})$, *where* $\boldsymbol{\theta}_0^{(n)} = (\gamma_0, \log\sigma_{\alpha_n}, \log\alpha_n)$ *and* $\alpha_n = 1 - F(u_n)$ *denote the adaptive true values.*

(ii) Any maximum likelihood estimator sequence from part (i) is jointly asymptotic normal, in such a way that

$$\sqrt{n\bar{\alpha}}\begin{bmatrix}\widehat{\boldsymbol{\psi}} - \boldsymbol{\psi}_0 \\ \widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0^{(n)}\end{bmatrix} \xrightarrow{d} \mathcal{N}\left(\boldsymbol{0}, \left(\widetilde{\Sigma}^{-1}\right)\widetilde{\Omega}\left(\widetilde{\Sigma}^{-1}\right)^T\right)$$

where

$$\widetilde{\Sigma} = \begin{bmatrix}\Sigma & 0 & 0 \\ -\Gamma_1^T\Sigma & \mathcal{I} & 0 \\ -\frac{1}{1-\bar{\alpha}}\Gamma_2^T\Sigma & 0 & \frac{1}{1-\bar{\alpha}}\end{bmatrix}, \quad \widetilde{\Omega} = \begin{bmatrix}\bar{\alpha}\Omega & \bar{\alpha}\sigma_{\bar{\alpha}}\Gamma_3 & \frac{\bar{\alpha}\sigma_{\bar{\alpha}}}{1-\bar{\alpha}}\Gamma_4 \\ \bar{\alpha}\sigma_{\bar{\alpha}}\Gamma_3^T & \mathcal{I} & 0 \\ \frac{\bar{\alpha}\sigma_{\bar{\alpha}}}{1-\bar{\alpha}}\Gamma_4^T & 0 & \frac{1}{1-\bar{\alpha}}\end{bmatrix}$$

with

$$\Gamma_1 = \mathbb{E}\left[\frac{1}{2h_t}\frac{\partial h_t}{\partial \boldsymbol{\psi}}\right]\begin{bmatrix}\frac{1}{(1+\gamma_0)(1+2\gamma_0)}\\ \frac{1}{1+2\gamma_0}\end{bmatrix}^T + \Gamma_2\begin{bmatrix}-\frac{\gamma_0}{(1+\gamma_0)(1+2\gamma_0)}\\ \frac{1+\gamma_0}{1+2\gamma_0}\end{bmatrix}^T,$$

$$\Gamma_2 = \frac{1}{\sigma_{\bar{\alpha}}}\left\{Q(1-\bar{\alpha})\mathbb{E}\left[\frac{1}{2h_t}\frac{\partial h_t}{\partial \boldsymbol{\psi}}\right] - \mathbb{E}\left[\frac{1}{\sqrt{h_t}}\frac{\partial \varepsilon_t}{\partial \boldsymbol{\psi}}\right]\right\},$$

$$\Gamma_3 = \begin{bmatrix}\mathbb{E}\left[w_t^2\frac{\partial \varepsilon_t}{\partial \boldsymbol{\phi}}\sqrt{h_t}\right]\\ \mathbb{E}\left[\frac{w_t^4}{2h_t}\frac{\partial h_t}{\partial \boldsymbol{\phi}_h}\right]\end{bmatrix}\begin{bmatrix}\frac{1}{(1-\gamma_0)^2}\\ \frac{1}{1-\gamma_0}\end{bmatrix}^T, \quad \Gamma_4 = \begin{bmatrix}\mathbb{E}\left[w_t^2\frac{\partial \varepsilon_t}{\partial \boldsymbol{\phi}}\sqrt{h_t}\right]\left(\frac{Q(1-\bar{\alpha})}{\sigma_{\bar{\alpha}}} + \frac{1}{1-\gamma_0}\right)\\ \mathbb{E}\left[\frac{w_t^4}{2h_t}\frac{\partial h_t}{\partial \boldsymbol{\phi}_h}\right]\left(\frac{Q(1-\bar{\alpha})}{\sigma_{\bar{\alpha}}} + \frac{1}{1-\gamma_0} - \frac{1}{\sigma_{\bar{\alpha}}}\right)\end{bmatrix},$$

and $\mathcal{I}$ defined in Theorem 1.

Remark 6. Again, our inference is asymptotically correct regardless of the threshold being finite or divergent and deterministic or random if we rewrite Assumption 2 slightly as in Remark 1. By fixing $\bar{\alpha}$, we can effectively quantify the influence from the ARMA-GARCH model estimation errors for our generalized Pareto parameter inference based on residuals rather than the true errors. When $\bar{\alpha} = \bar{\alpha}_n \to 0$ is an intermediate sequence such that $n\bar{\alpha}/n^{\kappa} \to \infty$ for some $\kappa > 0$ as in, e.g, McNeil and Frey (2000), Martins-Filho et al. (2018), and Hoga (2019), we deduce in the supplement that the estimation error from the ARMA-GARCH model indeed becomes asymptotically negligible as

$$\sqrt{n\bar{\alpha}_n}\left(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0^{(n)}\right) \xrightarrow{d} \mathcal{N}\left(\mathbf{0}, \begin{bmatrix}\mathcal{I}^{-1} & 0\\ 0 & 1\end{bmatrix}\right),$$

where the asymptotic variance is the same as using the true errors rather than the residuals, and coincides with that in the theorem above by approximating $\bar{\alpha}$ to its limit 0. In other words, our approach unifies the inference for both non-divergent and divergent thresholds. Following Remark 1, it is natural to expect that our methods remain asymptotically correct when the true errors are array data that could be sufficiently well modeled by the generalized Pareto distribution.

From the theorem above, we can quantify the impact of the ARMA-GARCH model estimation errors to our inference of the Generalized Pareto parameters using residuals rather than the true errors. In particular, observe that

$$\sqrt{n\bar{\alpha}}\left(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0^{(n)}\right) \xrightarrow{d} \mathcal{N}\left(\mathbf{0}, \mathcal{I}_{\bar{\alpha}}^{-1} + \bar{\alpha}V_{\bar{\alpha}}\right),$$

where $\mathcal{I}_{\bar{\alpha}}^{-1} = \begin{bmatrix} \mathcal{I}^{-1} & 0 \\ 0 & 1-\bar{\alpha} \end{bmatrix}$ is the asymptotic covariance matrix in Theorem 1, and we have an additional variance term depending on the ARMA-GARCH model given by

$$V_{\bar{\alpha}} = \mathcal{I}_{\bar{\alpha}}^{-1} \left(A\Omega A^T + vA^T + Av^T \right) \mathcal{I}_{\bar{\alpha}}^{-1}, \; A = \begin{bmatrix} \Gamma_1^T \\ \frac{1}{1-\bar{\alpha}} \Gamma_2^T \end{bmatrix}, \; v = \begin{bmatrix} \sigma_{\bar{\alpha}} \Gamma_3^T \\ \frac{\sigma_{\bar{\alpha}}}{1-\bar{\alpha}} \Gamma_4^T \end{bmatrix}. \tag{3.9}$$

Now recall the quantile formula (2.8). The following quantile inference theorem follows by continuous mapping theorem.

Theorem 6. *Under the conditions of Theorem 5, for any* $p \in (0, \alpha_0)$

$$\frac{\sqrt{n\bar{\alpha}}}{\sigma_p} \left(\widehat{\mathrm{VaR}}_X(1-p) - \mathrm{VaR}_X(1-p) \right) \xrightarrow{d} N(0, \tau^2(\bar{\alpha}, p)),$$

where the variance

$$\tau^2(\bar{\alpha}, p) = q\left(\frac{\bar{\alpha}}{p}\right)^T \mathcal{I}^{-1} q\left(\frac{\bar{\alpha}}{p}\right) + 1 - \bar{\alpha} + \bar{\alpha} \begin{bmatrix} q(\bar{\alpha}/p) \\ 1 \end{bmatrix} V_{\bar{\alpha}} \begin{bmatrix} q(\bar{\alpha}/p) \\ 1 \end{bmatrix}^T,$$

with $\mathcal{I}^{-1}$ *defined in Theorem 1 and the additional variance term* $V_{\bar{\alpha}}$ *given in* (3.9)*, compared to that in Theorem 2.*

We omit the proof as it is completely analogous to that of Theorem 2. Now, with $\widehat{\sigma}_p = \widehat{\sigma}_n(\widehat{\alpha}/p)^{\widehat{\gamma}_n}$ and a consistent estimator $\widehat{\tau}^2(\bar{\alpha}, p)$ (e.g., replacing the moments by their sample versions, $\bar{\alpha}$ by $\widehat{\alpha}$, γ_0 by $\widehat{\gamma}$, $\sigma_{\bar{\alpha}}$ by $\widehat{\sigma}_n$, and $Q(1-\bar{\alpha})$ by u_n), a confidence interval with level a of $\mathrm{VaR}_X(1-p)$ is given by

$$\left[\widehat{\mathrm{VaR}}_X(1-p) - \frac{z_{(1+a)/2}}{\sqrt{n\bar{\alpha}}} \widehat{\sigma}_n(\widehat{\alpha}/p)^{\widehat{\gamma}_n} \widehat{\tau}(\bar{\alpha}, p), \widehat{\mathrm{VaR}}_X(1-p) + \frac{z_{(1+a)/2}}{\sqrt{n\bar{\alpha}}} \widehat{\sigma}_n(\widehat{\alpha}/p)^{\widehat{\gamma}_n} \widehat{\tau}(\bar{\alpha}, p) \right].$$

Substituting $\widehat{\mathrm{VaR}}_X(1-p)$ in (3.8) by the values in the above interval, we can construct a prediction interval for $\widehat{\mathrm{VaR}}_{Y,n}(1-p)$. Similar to the case of independent data, such an interval has a poor coverage probability in small samples. It is computationally intensive to employ the residual based bootstrap method. Here, to bypass the daunting task of estimating the asymptotic variance of the quantile estimator, we suggest a random weighted bootstrap procedure as follows.

- *Step C1)* Draw a random sample with sample size n from a distribution function with mean one and variance one, say $\xi_1^{(b)}, \cdots, \xi_n^{(b)}$.

- *Step C2)* First, we estimate $\boldsymbol{\phi}$ by

$$\widehat{\boldsymbol{\phi}}^{(b)} = \arg\min_{\boldsymbol{\phi}} \sum_{t=1}^{n} \xi_t^{(b)} \widetilde{w}_t^2 \widetilde{\varepsilon}_t^2(\boldsymbol{\phi}).$$

Second, we estimate $\boldsymbol{\phi}_h$ by maximizing

$$\sum_{t=1}^{n} \xi_t^{(b)} \widetilde{w}_t^4 \widetilde{l}_t \left(\boldsymbol{\phi}_h | \widehat{\boldsymbol{\phi}}^{(b)}\right)$$

and denote the estimator by $\widehat{\boldsymbol{\phi}}_h^{(b)}$. Define $\widehat{X}_t^{(b)} = \widetilde{\varepsilon}_t(\widehat{\boldsymbol{\phi}}^{(b)}) / \sqrt{\widetilde{h}_t(\widehat{\boldsymbol{\phi}}^{(b)}, \widehat{\boldsymbol{\phi}}_h^{(b)})}$ for $t = 1, \cdots, n$, $\widehat{u}_n^{(b)} = \widehat{X}_{n-[n\bar{\alpha}]:n}^{(b)}$, and estimate γ_0 and σ_{α_n} by solving

$$\sum_{t=1}^{n} \xi_t^{(b)} \delta(\widehat{X}_t^{(b)} - \widehat{u}_n^{(b)}) s_1(\boldsymbol{\nu} | \widehat{X}_t^{(b)} - \widehat{u}_n^{(b)}) = 0, \quad \sum_{t=1}^{n} \xi_t^{(b)} \delta(\widehat{X}_t^{(b)} - \widehat{u}_n^{(b)}) s_2(\boldsymbol{\nu} | \widehat{X}_t^{(b)} - \widehat{u}_n^{(b)}) = 0.$$

Denote the estimators by $\widehat{\gamma}^{(b)}$ and $\widehat{\sigma}^{(b)}$, which gives

$$\widehat{\mathrm{VaR}}_X^{(b)}(1-p) = \widehat{u}_n^{(b)} + \frac{\widehat{\sigma}^{(b)}}{\widehat{\gamma}^{(b)}}((\bar{\alpha}/p)^{\widehat{\gamma}^{(b)}} - 1).$$

- *Step C3)* Repeat the above two steps B times to obtain $\left\{\widehat{\mathrm{VaR}}_X^{(b)}(1-p)\right\}_{b=1}^{B}$. Let $\widetilde{D}_{1:B} \leq \cdots \leq \widetilde{D}_{B:B}$ denote the order statistics of

$$\log\left(\frac{\widehat{\mathrm{VaR}}_X^{(b)}(1-p)}{\widehat{\mathrm{VaR}}_X(1-p)}\right), \; b = 1, \ldots, B,$$

and let $\widetilde{D}_{(1)} \leq \cdots \leq \widetilde{D}_{(B)}$ denote the order statistics of

$$\left|\log\left(\frac{\widehat{\mathrm{VaR}}_X^{(b)}(1-p)}{\widehat{\mathrm{VaR}}_X(1-p)}\right)\right|, \; b = 1, \ldots, B.$$

Hence, the confidence intervals with level a for $\log(\mathrm{VaR}_{\bar{X}}(1-p))$ are

$$\widetilde{I}_{RWB,1}(a) = [\log(\widehat{\mathrm{VaR}}_{\bar{X}}(1-p)) - \widetilde{D}_{[\frac{B+Ba}{2}]:B}, \quad \log(\widehat{\mathrm{VaR}}_{\bar{X}}(1-p)) - \widetilde{D}_{[\frac{B-Ba}{2}]:B}]$$

and

$$\widetilde{I}_{RWB,2}(a) = [\log(\widehat{\mathrm{VaR}}_{\bar{X}}(1-p)) - \widetilde{D}_{(Ba)}, \quad \log(\widehat{\mathrm{VaR}}_{\bar{X}}(1-p)) + \widetilde{D}_{(Ba)}].$$

Again, substituting $\widehat{\mathrm{VaR}}_X(1-p)$ in (3.8) by the values in each interval above, we can construct the corresponding prediction interval for $\widehat{\mathrm{VaR}}_{Y,n}(1-p)$. The simulation study below shows that the above procedure provides a good finite-sample coverage performance. The asymptotic theory for the random weighted bootstrap method can be derived with rather tedious calculations and thus is skipped.

4 Simulation Study

4.1 Independent data

This subsection carries out a simulation study to evaluate the finite-sample behavior of the proposed method for estimating VaR based on independent observations.

We draw 10000 random samples with sample size $n = 500$ or 1200 or 2500 from (1.2) with $\gamma = 3$ or $1/3$, $\sigma = 1$, G being the standard normal distribution, $\theta = 0.9$. We use $\bar{\alpha} = 0.05$, $p = 0.01$ or 0.001, and $B = 10000$ in the naive bootstrap method and the random weighted bootstrap method. We use the `nlm` function in the R statistical software to minimize the likelihood function with the following initial values for γ and $\sigma_{\bar{\alpha}}$.

Let $Y_i = X_{(n-i+1):n} - X_{(n-[n\bar{\alpha}]):n}$ for $i = 1, \ldots, m$ with $m = [n\bar{\alpha}]$. As we consider positive index γ, we use the initial values

$$\gamma^{ini} = \frac{1}{\log 2}\left|\log \frac{Y_{[m(1-3/8)]:m} - Y_{[m(1-3/16)]:m}}{Y_{[m(1-3/4)]:m]} - Y_{[m(1-3/8)]:m}}\right| \quad \text{and } \sigma_{\bar{\alpha}}^{ini} = \frac{Y_{[m(1-3/8)]:m}\gamma^{ini}}{(3/8)^{-\gamma^{ini}} - 1}.$$

Here γ^{ini} is the Pickands tail index estimation in Pickands (1975).

The coverage probabilities of the proposed intervals with levels $a = 90\%$ and 95% are reported in Tables 1 and 2, which show that i) the normal approximation method is the worst, and ii) it is much better to use the naive bootstrap method and the random weighted bootstrap method with critical values computed from the empirical distribution of the absolute differences between the bootstrapped estimators and the maximum likelihood VaR estimator. Further, the normal Q-Q plots in Figures B.1 and B.2 of the supplement show that the distribution of the VaR estimator is away from a normal distribution, especially when p is very small. Hence we prefer $I_{Boot,2}(a)$ and $I_{RWB,2}(a)$ to $I_{Boot,1}(a)$ and $I_{RWB,1}(a)$ in risk analysis.

4.2 ARMA-GARCH sequence

This subsection carries out a simulation study to evaluate the finite-sample behavior of the proposed method for estimating VaR based on an AR-GARCH sequence.

Due to the computation burden of the random weighted bootstrap method, we draw 1000 random samples with sample size $n = 1200$ and 2500 from the following AR-GARCH

(n,p,γ)	$I_{NA}(0.90)$	$I_{Boot,1}(0.90)$	$I_{Boot,2}(0.90)$	$I_{RWB,1}(0.90)$	$I_{RWB,2}(0.90)$
(500,0.01,3)	0.7671	0.8447	0.9042	0.8516	0.9009
(500,0.001,3)	0.6634	0.8267	0.9012	0.8089	0.8936
(1200,0.01,3)	0.8247	0.8723	0.9005	0.8697	0.9012
(1200,0.001,3)	0.7392	0.8607	0.8971	0.8535	0.8984
(2500,0.01,3)	0.8573	0.8901	0.8987	0.8869	0.9007
(2500,0.001,3)	0.7837	0.8812	0.8957	0.8706	0.8972
(500,0.01,1/3)	0.8569	0.8571	0.8990	0.8591	0.8966
(500,0.001,1/3)	0.7453	0.7053	0.9318	0.6791	0.9210
(1200,0.01,1/3)	0.8803	0.8815	0.9034	0.8799	0.9036
(1200,0.001,1/3)	0.8027	0.7840	0.9145	0.7635	0.9136
(2500,0.01,1/3)	0.8928	0.8898	0.8985	0.8893	0.8975
(2500,0.001,1/3)	0.8494	0.8446	0.9029	0.8205	0.9048

Table 1: *Confidence intervals with level $a = 90\%$.* Empirical coverage probabilities are reported for the normal approximation confidence interval $I_{NA}(a)$, the naive bootstrap intervals $I_{Boot,1}(a)$ and $I_{Boot,2}(a)$, and the random weighted bootstrap intervals $I_{RWB,1}(a)$ and $I_{RWB,2}(a)$. We take $\gamma = 3$ or $1/3$, $\sigma = 1$, $G \sim N(0,1)$ and $\theta = 0.9$ in (1.2).

model:

$$\begin{aligned} Y_t =& 0.0337 - 0.0620 Y_{t-1} + \varepsilon_t \\ \varepsilon_t =& \sqrt{h_t} X_t, \quad h_t = 0.0123 + 0.0883\varepsilon_{t-1} + 0.8310 h_{t-1} \end{aligned}$$

where $X_t = (e_t - \mathbb{E}e_t)/\mathbb{E}|e_t|$, $e_t = \delta e_{t,1} - (1-\delta)e_{t,2}$, and $e_{t,1}$ and $e_{t,2}$ are independent GPD random variables with CDF $F(x) = 1-(1+\gamma x)^{-1/\gamma}$. The parameters are calibrated from the daily returns on the S&P500 index between 2012 and 2016. We consider $\gamma = 1/3$ and $1/6$ to ensure $\mathbb{E}X_t^2 < \infty$. We take $\delta = 0.5$ and use the random weighted bootstrap method with $B = 10000$. The coverage probabilities of the proposed intervals with levels $a = 90\%$ and 95% are reported in Table 3, which show that $\widetilde{I}_{RWB,2}(a)$ is again better than $\widetilde{I}_{RWB,1}(a)$ and performs well except under the case $(n,p) = (1200, 0.001)$, where over-coverage is observed.

In summary, for quantifying the inference uncertainty of VaR estimation, we prefer

(n, p, γ)	$I_{NA}(0.95)$	$I_{Boot,1}(0.95)$	$I_{Boot,2}(0.95)$	$I_{RWB,1}(0.95)$	$I_{RWB,2}(0.95)$
(500,0.01,3)	0.7931	0.8928	0.9523	0.8806	0.9508
(500,0.001,3)	0.6817	0.8718	0.9512	0.8393	0.9479
(1200,0.01,3)	0.8538	0.9184	0.9491	0.9130	0.9489
(1200,0.001,3)	0.7626	0.9075	0.9480	0.8932	0.9486
(2500,0.01,3)	0.8908	0.9359	0.9480	0.9306	0.9486
(2500,0.001,3)	0.8102	0.9316	0.9478	0.9173	0.9483
(500,0.01,1/3)	0.9008	0.9102	0.9490	0.9126	0.9483
(500,0.001,1/3)	0.7852	0.7631	0.9655	0.7162	0.9607
(1200,0.01,1/3)	0.9235	0.9301	0.9520	0.9274	0.9524
(1200,0.001,1/3)	0.8396	0.8253	0.9572	0.7986	0.9567
(2500,0.01,1/3)	0.9375	0.9408	0.9485	0.9399	0.9491
(2500,0.001,1/3)	0.8838	0.8843	0.9519	0.8608	0.9532

Table 2: *Confidence intervals with level* $a = 95\%$. Empirical coverage probabilities are reported for the normal approximation confidence interval $I_{NA}(a)$, the naive bootstrap intervals $I_{Boot,1}(a)$ and $I_{Boot,2}(a)$, and the random weighted bootstrap intervals $I_{RWB,1}(a)$ and $I_{RWB,2}(a)$. We take $\gamma = 3$ or $1/3$, $\sigma = 1$, $G \sim N(0, 1)$ and $\theta = 0.9$ in (1.2).

the random weighted bootstrap method with critical values computed from the empirical distribution of the absolute differences between the bootstrapped risk estimators and the risk estimator, which works well for independent data and dependent data.

5 Data Analysis

5.1 Danish fire insurance losses

This subsection analyzes the Danish fire insurance data[1] in McNeil (1997) using the proposed semi-parametric GPD model in (1.2) by treating the fire losses as independent data. The dataset consists of 2167 large fire insurance claims (i.e., losses) in Denmark from

[1]The dataset is publicly available via R package `evir`.

(n,p,γ)	$\widetilde{I}_{RWB,1}(0.90)$	$\widetilde{I}_{RWB,2}(0.90)$	$\widetilde{I}_{RWB,1}(0.95)$	$\widetilde{I}_{RWB,2}(0.95)$
(1200,0.01,1/3)	0.843	0.908	0.901	0.961
(1200,0.001,1/3)	0.629	0.940	0.689	0.968
(2500,0.01,1/3)	0.878	0.894	0.928	0.949
(2500,0.001,1/3)	0.700	0.906	0.763	0.952
(1200,0.01,1/6)	0.857	0.898	0.918	0.951
(1200,0.001,1/6)	0.703	0.936	0.764	0.970
(2500,0.01,1/6)	0.868	0.903	0.930	0.943
(2500,0.001,1/6)	0.757	0.896	0.811	0.943

Table 3: *Confidence intervals for AR-GARCH models.* Empirical coverage probabilities are reported for the random weighted bootstrap confidence intervals $\widetilde{I}_{RWB,1}(a)$ and $\widetilde{I}_{RWB,2}(a)$ with $a = 0.90$ and 0.95.

January 1980 until December 1990.

We first perform a sensitivity analysis of the proposed method w.r.t. the choice of threshold level $\bar{\alpha}$. Specifically, we calculate the $(1-p)\times 100\%$ VaR at level $p = 0.01, 0.005, 0.001$ under varying threshold levels $\bar{\alpha} = (0.05, 0.1)$. The confidence interval (C.I.) for VaR is calculated at $a = 90\%$ and 95% level. To construct the C.I., we implement the normal approximation (NA), the naive bootstrap method (Boot1 and Boot2), and the random weighted bootstrap method (RWB1 and RWB2). For comparison, we further conduct a naive non-parametric bootstrap (Naive), where we simply bootstrap the Danish fire insurance data and use sample quantile to estimate the VaR and its C.I.

The result is given in Figure 1. The performance of semi-parametric GPD is fairly stable w.r.t. $\bar{\alpha}$ for $p = 0.01, 0.005$ and has some variation for $p = 0.001$. Note that the naive non-parametric bootstrap (Naive) gives a very wide (and thus non-informative) C.I. for extreme VaR ($p = 0.001$), which highlights the value/necessity of the proposed semiparametric C.I. construction approach. We also report the Q-Q plots of $\log(\widehat{\mathrm{VaR}}_X^{(b)}(1-p)/\widehat{\mathrm{VaR}}_X(1-p))$, $b = 1, 2, \cdots, 10000$, for Boot and RWB, respectively, in Figures B.3 and B.4 of the supplement. These figures show that the distribution is generally skewed, especially for $p = 0.005$ and 0.001.

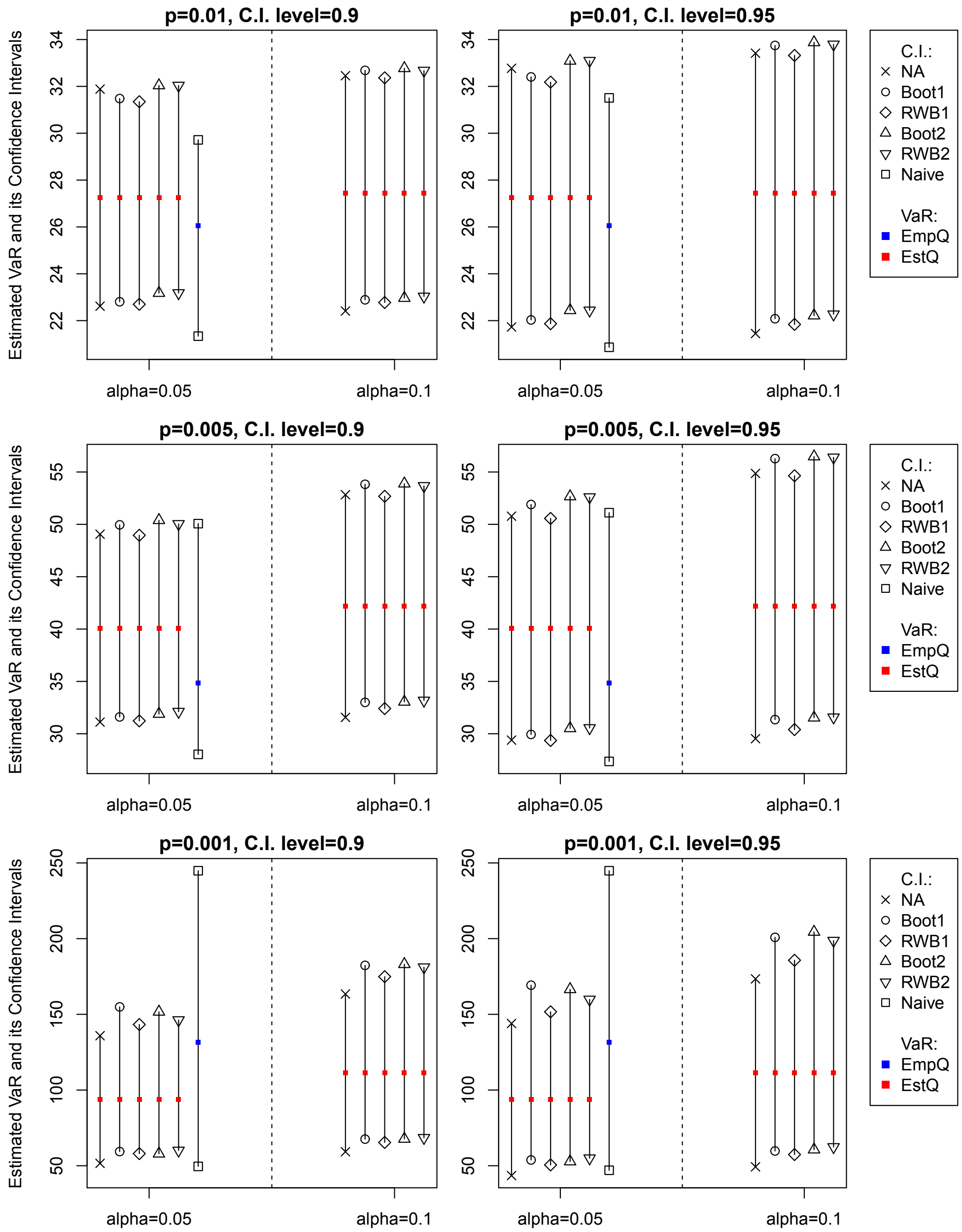

Figure 1: *Sensitivity analysis of constructed confidence intervals (C.I.) for the* $(1-p)\times 100\%$ *VaR at level* $p = 0.01, 0.005, 0.001$. *The y-axis is in the unit of 1 million Danish Krone. EmpQ stands for VaR estimated by the sample quantile, and EstQ stands for VaR estimated by the semi-parametric GPD. The result of the naive approach (Naive) does not depend on* $\bar{\alpha}$*, thus is only plotted under* $\bar{\alpha} = 0.05$ *to avoid confusion.*

We further conduct a Leave-One-Out Validation (LOOV) for the proposed semi-parametric GPD model and compare it with the naive nonparametric sample quantile approach (Naive). Specifically, for each observation X_i with $i = 1, 2, \cdots, 2167$, we use the leave-one-out sample X_{-i} to estimate the $(1 - p) \times 100\%$ VaR by either semi-parametric GPD or sample quantile (Naive). For evaluation, we use the empirical coverage rate, which is defined as the proportion of experiments where the left-out loss X_i is covered by (i.e. lower than) the estimated VaR based on the leave-one-out sample X_{-i} for $i = 1, \cdots, 2167$.

p	$\bar{\alpha}$	Emp. rate(GPD)	Emp. rate(Naive)	p-value(GPD)	p-value(Naive)
0.010	0.050	0.990	0.989	1.000	0.745
0.010	0.100	0.990	-	1.000	-
0.005	0.050	0.995	0.994	1.000	0.647
0.005	0.100	0.995	-	1.000	-
0.001	0.050	0.999	0.999	0.483	0.483
0.001	0.100	0.999	-	0.483	-

Table 4: *The empirical coverage rate (Emp. rate) of the estimated* $(1 - p) \times 100\%$ *VaR across the 2167 experiments at level* $p = 0.01, 0.005, 0.001$. *The result of the naive approach does not depend on* $\bar{\alpha}$, *thus is only reported under* $\bar{\alpha} = 0.05$ *to avoid confusion.*

Table 4 reports the empirical coverage rate of the estimated VaR by the two approaches across the 2167 experiments and further gives the corresponding p-values from the binomial tests[2] for the null hypothesis that the coverage probability of the estimated $(1-p) \times 100\%$ VaR is indeed the target level $1 - p$. As can be seen, both approaches give a satisfactory result with GPD providing a perfect performance. Moreover, note that the performance of the GPD approach is insensitive to the threshold level of $\bar{\alpha}$, indicating the statistical stability of the proposed approach.

[2]Under the null hypothesis, the number of coverage by the estimated $(1 - p) \times 100\%$ VaR across n experiments should follow a binomial distribution with parameters $(n, 1 - p)$. See Kratz et al. (2018) for more details of the binomial test.

5.2 Losses of the S&P500 index

This subsection analyzes the daily negative log-returns (i.e., losses) of the S&P500 index using the proposed semi-parametric GPD method with an ARMA-GARCH model. Precisely, on each day t, based on the past 2500 historical observations $(y_{t-2499}, y_{t-2498}, \cdots, y_t)$, we fit an AR(1)-GARCH(1,1) model using the proposed two-step self-weighted estimation method. We then calculate the one-day ahead $(1-p) \times 100\%$ conditional VaR by the semi-parametric GPD method with a threshold $\bar{\alpha}$ and construct the corresponding 90% or 95% C.I. by RWB.

For comparison, we also conduct the analysis using a traditional nonparametric approach (Trad). That is, we fit an AR(1)-GARCH(1,1) model by MLE and use the sample quantile of the fitted residuals to calculate the one-day ahead conditional VaR and bootstrap the residuals to construct the corresponding C.I. of VaR.

We let t start from 11/01/2007, which is roughly the start of the financial crisis, and we make the end date to be 10/20/2011, which roughly marks the end of the crisis. In other words, we aim to test the ability of the proposed GPD method for monitoring a financial system under stress.

p	$\bar{\alpha}$	Emp. rate(GPD)	Emp. rate(Trad)	p-value(GPD)	p-value(Trad)
0.010	0.050	0.985	0.979	0.111	0.002
0.010	0.100	0.985	-	0.111	-
0.005	0.050	0.995	0.989	1.000	0.020
0.005	0.100	0.995	-	1.000	-
0.001	0.050	1.000	0.998	0.632	0.264
0.001	0.100	1.000	-	0.632	-

Table 5: *The empirical coverage rate (Emp. rate) of the estimated* $(1-p) \times 100\%$ *VaR across the 1000 predictions at level* $p = 0.01, 0.005, 0.001$. *The result of the traditional approach does not depend on* $\bar{\alpha}$, *thus is only reported under* $\bar{\alpha} = 0.05$ *to avoid confusion.*

There are 1000 predictions of one-day ahead conditional VaR given by the semi-parametric GPD approach and the traditional nonparametric approach (Trad). We vary $p = 0.01, 0.005, 0.001$

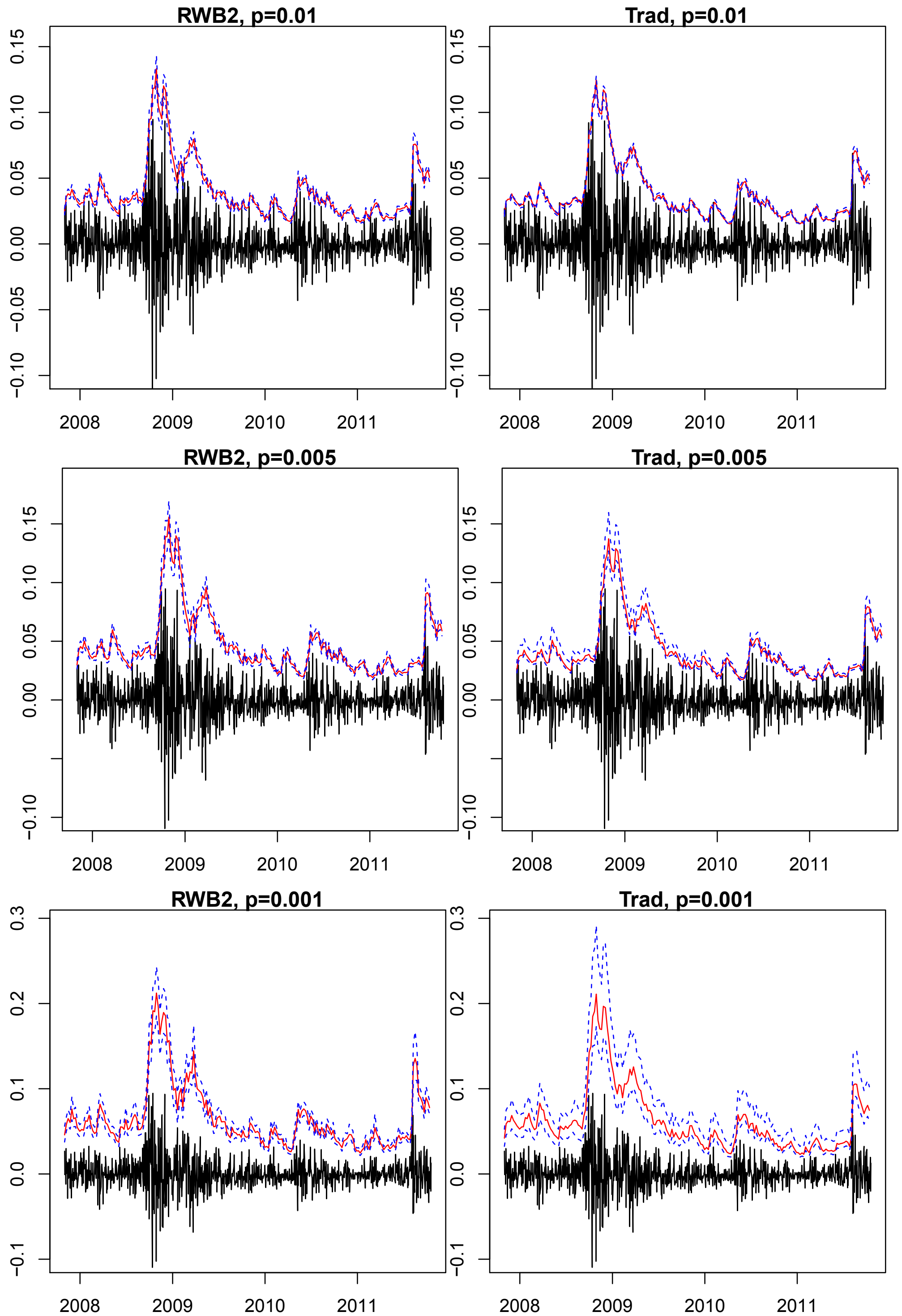

Figure 2: *Estimated one-day ahead conditional VaR (red line) and its 90% C.I. (blue dashed lines) by the random weight bootstrap (RWB2) and the traditional nonparametric approach (Trad). The black line denotes the negative daily log returns of the S&P500 index.*

and set the confidence level of the C.I. to be 90% or 95%. For the semi-parametric GPD approach, we further vary $\bar{\alpha} = 0.05, 0.1$. Table 5 reports the empirical coverage rate of the estimated VaR by the two approaches across the 1000 predictions, which is defined as the proportion of predictions where the observed loss is lower than the estimated one-day ahead conditional VaR. Table 5 also gives the corresponding p-values from the binomial test. As can be seen, the traditional nonparametric approach tends to underestimate the true conditional VaR, and thus imposes serious under-reserve risk. On the other hand, the semi-parametric GPD gives satisfactory prediction performance and passes all the binomial tests. Moreover, note that the performance of the GPD approach is again insensitive to the threshold level of $\bar{\alpha}$, indicating the statistical stability of the proposed approach.

For illustration, Figure 2 plots the estimated VaR and the corresponding C.I. given by RWB2 and the traditional nonparametric approach. For the plots, we set $\bar{\alpha} = 0.05$, the confidence level of C.I. $a = 90\%$ and vary $p = 0.01, 0.005, 0.001$. The result for RWB1 and the result for $\bar{\alpha} = 0.1$ and the confidence level $a = 95\%$ are similar and thus are omitted. Note that compared to RWB2, the C.I. given by the traditional nonparametric method is narrower for $p = 0.01$, which may be possibly due to the fact that the C.I. by the traditional nonparametric approach does not incorporate the estimation uncertainty of the AR(1)-GARCH(1,1) model. On the other hand, the nonparametric approach gives much wider C.I. for extreme quantiles $p = 0.005, 0.001$, indicating that a naive nonparametric bootstrap cannot construct an informative C.I. for extreme quantiles. This phenomenon is observed in the Danish insurance data analysis as well.

6 Conclusions

Given that regulators often set a high VaR level in risk management, fitting distribution in the tail is essential. This paper infers a semi-parametric model, which only models exceedances over a non-divergent threshold by the generalized Pareto distribution. Asymptotic results for parameters and VaR estimation are first derived for independent data. For financial data modeled by an ARMA-GARCH process, a three-step weighted estimation procedure is proposed to ensure a normal limit for estimating parameters and conditional VaR with heavy tailed observations. For efficiently quantifying the uncertainty of risk forecast, a random weighted bootstrap method is proposed and shown to be consistent. A simulation study and real data analysis confirm the advantages of the proposed methodologies. It is crucial to develop a distribution free goodness-of-fit test and the asymptotic theory for dynamic modeling of the generalized Pareto distribution, which will be our future research plan.

SUPPLEMENTARY MATERIAL

In this supplement, we provide the asymptotic theory for the naive bootstrap method for independent data (Remark 5), report some additional Q-Q plots discussed in Sections 4.1 and 5.1, prove Theorems 1–5, and deduce the results for divergent thresholds from Remarks 1 and 6 in details.

References

Allen, L., Bali, T.G., and Tang, Y. (2012). Does Systemic Risk in the Financial Sector Predict Future Economic Downturns? *The Review of Financial Studies*, 25, 3000–3036.

Balkema, A., and de Haan, L. (1974). Residual Life Time at Great Age. *The Annals of Probability*, 2, 792–804.

Barro, R.J., and Jin, T. (2011). On the Size Distribution of Macroeconomic Disasters. *Econometrica*, 79, 1567–1589.

Bollerslev, T. (1986). Generalized Autoregressive Conditional Heteroskedasticity. *Journal of Econometrics*, 31, 307–327.

Bollerslev, T., and Todorov, V. (2011). Estimation of Jump Tails. *Econometrica*, 79, 1727–1783.

Brodin, E., and Rootzén, H. (2009). Univariate and Bivariate GPD methods for Predicting Extreme Wind Storm Losses. *Insurance: Mathematics and Economics*, 44, 345–356.

Bücher, A., and Segers, J. (2017). On the Maximum Likelihood Estimator for the Generalized Extreme-Value Distribution. *Extremes*, 20, 839–872.

Chavez-Demoulin, V., Davison, A.C., and McNeil,A.J. (2005). Estimating Value-at-Risk: a Point Process Approach. *Quantitative Finance*, 5, 227–234.

Chavez-Demoulin, V., and Embrechts, P. (2004). Smooth Extremal Models in Finance and Insurance. *The Journal of Risk and Insurance* 71, 183–199.

Chavez-Demoulin, V., Embrechts, P., and Sardy, S. (2014). Extreme-quantile Tracking for Financial Time Series. *Journal of Econometrics*, 181, 44–52.

Cont, R. (2001). Empirical Properties of Asset Returns: Stylized Facts and Statistical Issues. *Quantitative Finance*, 1, 223–236.

Cramér, H. (1946). *Mathematical Methods of Statistics*, Princeton University Press.

Davison, A.C., and Smith, R.L. (1990). Models for Exceedances over High Thresholds. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 52, 393–442.

de Haan, L., and Ferreira, A. (2006). *Extreme Value Theory: an Introduction*, *New York: Springer*.

Dombry, C. (2015). Existence and Consistency of the Maximum Likelihood Estimators for the Extreme Value Index within the block maxima framework.", Bernoulli, 21, 420–436.

Drees, H., Ferreira, A., and de Haan, L. (2004). On Maximum Likelihood Estimation of the Extreme Value Index. *The Annals of Applied Probability*, 14, 1179–1201.

Duffie, D., and Pan, J. (1997). An Overview of Value at Risk. *The Journal of Derivatives*, 4, 7–49.

Embrechts, P., Klüppelberg, C., and Mikosch, T. (1997). *Modelling Extremal Events for Insurance and Finance, Berlin: Springer*.

Engle, R.F. (1982). Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of U.K. Inflation. *Econometrica*, 50, 987–1008.

Francq, C., and Zakoïan, J. (2004). Maximum Likelihood Estimation of Pure GARCH and ARMA-GARCH processes. *Bernoulli*, 10, 605–637.

Hall, P., and Tajvidi, N. (2000). Nonparametric Analysis of Temporal Trend when Fitting Parametric Models to Extreme-Value Data. *Statistical Science*, 15, 153–167.

Hall, P., and Yao, Q. (2003). Inference in ARCH and GARCH Models with Heavy Tailed Errors. *Econometrica*, 71, 285–317.

He, Y., Hou, Y., Peng, L., and Shen, H. (forthcoming). Inference for Conditional Value-at-Risk of a Predictive Regression. *The Annals of Statistics*.

Hoga, Y. (2019). Confidence Intervals for Conditional Tail Risk Measures in ARMA–GARCH Models. *Journal of Business & Economic Statistics*, 37, 613–624.

Hull, J. C. (2018). *Risk Management and Financial Institutions*, Fifth ed., Hoboken: Wiley.

Jorion, P. (2006). *Value at Risk: the New Benchmark for Measuring Financial Risk,* Third ed., New York:McGraw-Hill.

Kelly, B., and Jiang, H. (2014). Tail Risk and Asset Prices. *The Review of Financial Studies*, 27, 2841–2871.

Koul, H. L., and Ling, S. (2006). Fitting an Error Distribution in Some Heteroscedastic Time Series Models. *The Annals of Statistics*, 34, 994-1012.

Kratz, M., Lok, Y.H., and McNeil, A. (2018). Multinomial VaR backtests: A simple implicit approach to backtesting expected shortfall. *Journal of Banking & Finance*, 88, 393–407.

Ling, S. (2007). Self-Weighted and Local Quasi-Maximum Likelihood Estimators for ARMA-GARCH/IGARCH Models. *Journal of Econometrics*, 140, 849–873.

Martins-Filho, C., Yao, F., and Torero, M. (2018). Nonparametric Estimation of Conditional Value-at-risk and Expected Shortfall based on Extreme Value Theory. *Econometric Theory*, 34, 23–67.

Massacci, D. (2017). Tail Risk Dynamics in Stock Returns: Links to the Macroeconomy and Global Markets Connectedness. *Management Science*, 63, 3072–3089.

McNeil, A.J. (1997). Estimating the Tails of Loss Severity Distributions Using Extreme Value Theory. *ASTIN Bulletin*, 27, 117–137.

McNeil, A.J., and Frey, R. (2000). Estimation of Tail-Related Risk Measures for Heteroscedastic Financial Time Series: an Extreme Value Approach. *Journal of Empirical Finance*, 7, 271–300.

Owen, A. (2001). *Empirical Likelihood*, New York: Chapman & Hall.

Peng, L., and Qi, Y. (2009). Maximum Likelihood Estimation of Extreme Value Index for Irregular Cases. *Journal of Statistical Planning and Inference*, 139, 3361–3376.

Pickands, J. (1975). Statistical Inference Using Extreme Order Statistics. *The Annals of Statistics*, 3, 119–131.

Resnick, S.I. (1987). *Extreme Values, Regular Variation and Point Processes,* New York: Springer.

Rootzén, H., and Tajvidi, N. (1997). Extreme Value Statistics and Wind Storm Losses: a Case Study. *Scandinavian Actuarial Journal* 1, 70–94.

Smith, R.L. (1985). Maximum likelihood estimation in a class of nonregular case. *Biometrika*, 72, 67–90.

Smith, R.L. (1987). Estimating Tails of Probability Distributions. *The Annals of Statistics*, 15, 1174–1207.

Zhou, C. (2009). Existence and Consistency of the Maximum Likelihood Estimator for the Extreme Value Index. *Journal of Multivariate Analysis*, 100, 794–815.

Zhou, C. (2010). The Extent of the Maximum Likelihood Estimator for the Extreme Value Index. *Journal of Multivariate Analysis*, 101, 971–983.

Zhu, K., and Ling, S. (2011). Global Self-Weighted and Local Quasi-Maximum Exponential Likelihood Estimators for ARMA-GARCH/IGARCH Models. *The Annals of Statistics*, 39, 2131–2163.

Zhu, K. (2016). Bootstrapping the Portmanteau Tests in Weak Auto-Regressive Moving Average Models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 78, 463–485.

Zhu, K. (2019). Statistical Inference for Autoregressive Models Under Heteroscedasticity of Unknown Form. *The Annals of Statistics*, 6, 3185–3215.