

Fix- k Asymptotically Unbiased Estimation of Tail Properties with Complete, Censored, or Truncated Data

Yulong Wang

Syracuse University

Department of Economics and Center of Policy Research

Oct. 1st 2017

Abstract

This paper considers estimating tail properties such as high quantiles and tail conditional expectations. We provide new asymptotically (quantile) unbiased estimators that are applicable to (i) complete data; (ii) tail censored (top-coded) data with known or unknown censoring value; and (iii) tail truncated data with known and unknown truncation value. The new method relies on the sole assumption that the largest k observations satisfy the extreme value theory, for a given and fixed k . This asymptotics leads to excellent small sample bias and risk properties as shown by Monte Carlo simulations, and the empirical relevance is illustrated by estimating the high quantiles of the U.S. hurricane damage. In addition to i.i.d. data, the new method is generalized to accommodate stochastic volatility models by proving that the residuals of fitting a correctly specified AR-GARCH model satisfy our assumption.

Keywords: bias, extreme quantiles, tail conditional expectations, extreme value theory, fat-tailed distribution, unbiasedness, censoring, truncation

1 Introduction

Estimating tail properties such as high quantile and tail conditional expectation (TCE) has been an important issue in finance (Engle and Manganelli (2004), Jorion (2007), McNeil and Frey (2000), Fissler and Ziegel (2016), and Patton, Ziegel, and Chen (2017)), physics (Barkai, Metzler, and Klafter (2000), Meerschaert, Benson, Scheffler, and Becker-Kern (2002)), linguistics (Zipf (1936), Baek, Bernhardsson, and Minnhagen (2011)), hydrology and seismology (Anderson and Meerschaert (1998), Benson, Wheatcraft, and Meerschaert (2000), Pacheco, Scholtz, and Sykes (1992)), and many other fields.

When the data are complete, a large number of estimators have been developed based on the extreme value theory and tail regularity conditions. See Embrechts, Klupperberg, and Mikosch (1997), Reiss and Thomas (2007), Resnick (2007), and de Haan and Ferreira (2007), Beirlant, Caeiro, and Gomes (2012), Gomes and Guillou (2015) for reviews and references. Given the assumption that the underlying distribution F is in the domain of attraction, its tail can be well approximated by a generalized Pareto distribution (cf. Pickands (1975)), and then common tail properties of interest, including high quantile and TCE, are functions of three parameters only, at least approximately. These parameters include the tail index ξ which is the exponential component, the scale, and the location. Along this line, numerous suggestions have been made about estimating these parameters, and the corresponding estimators of high quantile and TCE can be constructed by plugging in estimators of these parameters.

One concern of the above mentioned methods is that they rely on the "increasing- k " asymptotics, which, however, may lead to a poor small sample approximation when the sample size is only moderately large, say 500. More specifically, the consistent estimate of ξ requires the asymptotics under $k_n \rightarrow \infty$ and $k_n/n \rightarrow 0$ where n is the sample size. So given a certain sample, the choice of k can be difficult to keep the delicate balance that (i) k has to be large enough for the asymptotic normality to hold on estimating ξ ; and (ii) k has to be so small relative to n that the largest k observations satisfy the extreme value theorem. Therefore, in many empirical applications such as financial daily data collected from one year, there could be no choice of k that results in a satisfactory small sample performance (cf. Kuester, Mittnik, and Paolella (2006)).

In addition to the concern about choosing k , the above mentioned methods cannot be applied to incomplete data due to censoring or truncation, which are common in empirical applications. In particular, data censoring usually exists in surveys about earnings and

wealth such as the Current Population Survey, and the problems subject to data truncation can be found in finance, hydrology, fire ecology, and seismology (see, for example, Groisman, Knight, Karl, Easterling, Sun, and Lawrimore (2004) and Malamud, Morein, and Turcotte (1998)). Estimating tail properties with such incomplete data is even more difficult, since the largest observations are very informative about tail but they are unfortunately unavailable. Existing literature typically makes parametric assumptions on the whole distribution (cf. Aban, Meerschaert, and Panorska (2006) and Jenkins, Burkhauser, Feng, and Larrimore (2010)) but this approach may suffer severe misspecification.

To overcome the above issues, this paper develops a unified framework to accommodate all three type of data and considers the "fixed- k " asymptotic embedding developed by Müller and Wang (2016) under the sole assumption that the k largest order statistics jointly converge to k jointly extreme value distributed variables, for fixed and given k . This means we only require a fixed number of tail observations are approximately stemming from a (generalized) Pareto while leaving the main body of the underlying distribution unspecified. Consequently, it is asymptotically equivalent to deal with a small sample problem where we are estimating a quantity as a function of the underlying distribution based on k observations. Then, we set up the Lagrangian problem to find the optimal estimator that minimizes a weighted average risk criteria and satisfies some unbiased constraint, and solve it by the generic algorithm suggested by Müller and Wang (2015). The unbiasedness is important since it is naturally embedded in the definition of quantile and TCE. In particular, in estimating the quantile of the underlying distribution F , it makes sense to require the estimator, $\hat{Q}(p)$, to satisfy the quantile unbiasedness: $P\left(Y_i > \hat{Q}(p)\right) = 1 - p$ where Y_i is another independent draw from F . For TCE, we show that the mean unbiasedness is equivalent to some average quantile unbiasedness that measures both the size and the likelihood of tail above a certain confidence level. This is especially important in finance as it is the exact reason why Basel III suggests switching from value at risk (VaR) to expected shortfall (ES) for measuring financial risk. See Basel Committee on Banking Supervision (2013) for more details.

More specifically, we focus in this paper on the construction of estimators for the $1 - h/n$ quantile, for given h , and the corresponding tail conditional expectation. This captures the empirical situation where only few observations can be considered relevant for tail properties. Regarding incomplete data, we model censoring as that in the sample, the largest m observations are unobserved, for a known m , and model truncation as that the data are generated from a truncated F with either known or unknown truncation value. We start with an i.i.d. sample and show by Monte Carlo simulations that the new estimators have

excellent small sample unbiasedness and risk properties for moderately large n , such as 250. Then, we extend the results to stochastic volatility models by establishing that the innovations of fitting a GARCH model still satisfy the joint extreme value theory. We illustrate the application of the new estimators with U.S. hurricane data.

The rest of the paper is organized as follows. Section 2 contains the details of the new approach, including deriving the "fixed- k " asymptotics, setting up and solving the Lagrangian problems, and extending the results to stochastic volatility models. Section 3 reports Monte Carlo simulations and explicitly derives the bias properties of the empirical estimator. Section 5 concludes.

2 Derivation of the Estimator for Complete Data

2.1 High Quantile

We start with estimating a high quantile based a random sample Y_1, Y_2, \dots, Y_n drawn from a certain cumulative distribution function (CDF) F . The i.i.d. setup is extended to accommodate the stochastic volatility model in Section 6. To capture the fact that we only have limited information about the tail, we focus on estimating the $1 - h/n$ th quantile of F , denoted by $Q(F, 1 - h/n)$, for a given and fixed h , indicating that the object of interest is of the same order of magnitude as the sample maximum. Typical choices of h can be 0.1, 1, 5, and 10, corresponding to the quantile at levels 99.98%, 99.8%, 99%, and 98% for a sample of 500. Notice that there is a naturally embedded quantile unbiasedness constraint on the estimator \hat{Q} such that $E \left[P \left(Y_i > \hat{Q} \right) \right] = h/n$ for an independent draw Y_i from F . Hence our objective is to construct the optimal \hat{Q} that satisfies such quantile unbiasedness restriction, at least asymptotically.

To avoid assuming an increasing k that may lead to a poor finite sample approximation (see Section 5), we follow Müller and Wang (2016) to consider the fixed- k asymptotics. In particular, we use only the largest k order statistics as our effective sample, denoted as $\mathbf{Y} = (Y_{n:n}, Y_{n:n-1}, \dots, Y_{n:n-k+1})'$ where $Y_{n:1} \leq Y_{n:2} \leq \dots \leq Y_{n:n}$ denote the order statistics.

Our approach relies on the extreme value theory (see, for example, de Haan and Ferreira (2007)) that if there exist sequences a_n and b_n such that

$$\frac{Y_{n:n} - b_n}{a_n} \Rightarrow X_1 \tag{1}$$

for some nondegenerate random variable X_1 , then there exists constants a and b such that

the distribution of $aX_1 + b$ has the following CDF

$$G_\xi(x) = \begin{cases} \exp[-(1 + \xi x)^{-1/\xi}], & 1 + \xi x \geq 0, \text{ for } \xi \neq 0 \\ \exp[-e^{-x}], & x \in \mathbb{R}, \xi = 0 \end{cases} \quad (2)$$

where ξ is referred to as the tail index, the parameter measuring the decay rate of the tail. This distribution is referred to as the generalized extreme value distribution, and the cases with $\xi < 0$, $\xi = 0$ and $\xi > 0$ correspond to Weibull, Gumbel and Fréchet type, respectively. Without loss of generality, assume that the CDF of X_1 in (1) is exactly G_ξ by subsuming a and b in a_n and b_n . In addition to the sample maximum, the extreme value theory also extends to the first k order statistics such that if (1) holds, then for any fixed k ,

$$\left(\begin{array}{c} \frac{Y_{n:n} - b_n}{a_n} \\ \vdots \\ \frac{Y_{n:n-k+1} - b_n}{a_n} \end{array} \right) \Rightarrow \mathbf{X} = \left(\begin{array}{c} X_1 \\ \vdots \\ X_k \end{array} \right) \quad (3)$$

where the joint probability density function (PDF) of \mathbf{X} is given by $f_{\mathbf{X}}(x_1, \dots, x_k) = G_\xi(x_k) \prod_{i=1}^k g_\xi(x_i)/G_\xi(x_i)$ on $0 \leq x_k \leq x_{k-1} \leq \dots \leq x_1$, where $g_\xi(x) = dG_\xi(x)/dx$.

The normalizing constants a_n and b_n depend on the underlying distribution F , and are typically difficult to estimate. Suppose they were known, the limiting problem then only involves a k -dimensional draw \mathbf{X} whose distribution is fully characterized by the scalar parameter ξ , and we seek to construct an estimator that satisfies the asymptotic quantile unbiasedness whenever (3) holds. To implement this idea, we next impose some equivariance on the estimator to avoid estimating a_n and b_n and derive the explicit form of the limiting problem.

First, notice that $P(Y_{n:n} > Q(F, 1 - h/n)) = (1 - h/n)^n \rightarrow e^{-h}$. Thus, under (3), $(Q(F, 1 - h/n) - b_n)/a_n$ converges to the e^{-h} quantile of X_1 , denoted as $q(\xi, h)$ in the following. Some calculation shows $q(\xi, h) = (h^{-\xi} - 1)/\xi$ for $\xi \neq 0$ and $q(0, h) = -\log(h)$. If a_n and b_n were known, the asymptotic problem then becomes estimation about $q(\xi, h)$ based on the $k \times 1$ vector of observations \mathbf{X} .

Next, we impose location and scale equivariance on the estimator \hat{Q} (see, for example, Lehmann and Romano (2005)), such that for any constants $a \neq 0$ and b ,

$$\hat{Q}(a\mathbf{Y} + b) = a\hat{Q}(\mathbf{Y}) + b. \quad (4)$$

Such equivariance can be implemented by constructing $\hat{Q}(\mathbf{Y}) = (Y_{n:n} - Y_{n:n-k+1})\hat{Q}(\mathbf{Y}^s) + Y_{n:n-k+1}$ where

$$\mathbf{Y}^s = \left(\frac{Y_{n:n} - Y_{n:n-k+1}}{Y_{n:n} - Y_{n:n-k+1}}, \frac{Y_{n:n-1} - Y_{n:n-k+1}}{Y_{n:n} - Y_{n:n-k+1}}, \dots, \frac{Y_{n:n-k+1} - Y_{n:n-k+1}}{Y_{n:n} - Y_{n:n-k+1}} \right)$$

is a maximal invariant to the linear transformations. The continuous mapping theorem and (3) imply

$$\mathbf{Y}^s \Rightarrow \mathbf{X}^s \equiv \left(\frac{X_1 - X_k}{X_1 - X_k}, \frac{X_2 - X_k}{X_1 - X_k}, \dots, \frac{X_k - X_k}{X_1 - X_k} \right)$$

whose distribution then depends on ξ only. Therefore asymptotically, the original problem amounts to determining \hat{Q} on the $k - 2$ dimensional subset of \mathbb{R}^k where \mathbf{X}^s has first element equal to 1 and last element equal to zero.

Under the equivariance constraint, we can derive quantile bias as follows

$$\begin{aligned} & nP \left(Y_i > \hat{Q}(\mathbf{Y}) \mid \mathbf{Y} \right) \\ &= n \left(1 - F \left((Y_{n:n} - Y_{n:n-k+1}) \hat{Q}(\mathbf{Y}^s) + Y_{n:n-k+1} \right) \right) \\ &= n \left(1 - F \left(a_n \left(\frac{Y_{n:n} - Y_{n:n-k+1}}{a_n} \hat{Q}(\mathbf{Y}^s) + \frac{Y_{n:n-k+1} - b_n}{a_n} \right) + b_n \right) \right) \\ &\approx n \left(1 - F \left(a_n \left((X_1 - X_k) \hat{Q}(\mathbf{X}^s) + X_k \right) + b_n \right) \right). \end{aligned}$$

Assume F is in the domain of attraction, which then implies that $n(1 - F(a_n y + b_n)) \rightarrow (1 + \xi y)^{-1/\xi}$ for all y such that $1 + \xi y > 0$. Hence the above expression converges to $\left(1 + \xi \left((X_1 - X_k) \hat{Q}(\mathbf{X}^s) + X_k \right) \right)^{-1/\xi}$, and the quantile bias is asymptotically of the following form

$$nP \left(Y_i > \hat{Q}(\mathbf{Y}) \right) \rightarrow E_\xi \left[\left(1 + \xi \left((X_1 - X_k) \hat{Q}(\mathbf{X}^s) + X_k \right) \right)^{-1/\xi} \right]$$

where the expectation is taken w.r.t. the vector $(X_1 - X_k, X_k, \mathbf{X}^s)$ whose distribution can be derived from the PDF of \mathbf{X} via change of variables.

Note that ξ cannot be consistently estimated as we only have a fixed k number of observations. Alternatively, we impose the asymptotically quantile unbiasedness for all the values of ξ in an empirically relevant set $\Xi \subset \mathbb{R}$. The asymptotic problem then is the construction of \hat{Q} that satisfies

$$E_\xi \left[\left(1 + \xi \left((X_1 - X_k) \hat{Q}(\mathbf{X}^s) + X_k \right) \right)^{-1/\xi} \right] = h \text{ for all } \xi \in \Xi. \quad (5)$$

To construct the optimal \hat{Q} among those satisfying (5), we focus on the one that minimizes the mean absolute deviation (MAD) criterion

$$\int E_\xi \left[\left| (X_1 - X_k) \hat{Q}(\mathbf{X}^s) + X_k - q(\xi, h) \right| \right] dW(\xi) \quad (6)$$

where W is a positive measure with support on Ξ .¹ Thus combining the asymptotic versions of the constraint (5) and the objective (6), the limiting problem can be formulated as

$$\begin{aligned} & \min_{\hat{Q}(\cdot)} \int E_{\xi} \left[\left| (X_1 - X_k) \hat{Q}(\mathbf{X}^s) + X_k - q(\xi, h) \right| \right] dW(\xi) \\ & -\varepsilon \leq E_{\xi} \left[\left(1 + \xi \left((X_1 - X_k) \hat{Q}(\mathbf{X}^s) + X_k \right) \right)^{-1/\xi} \right] - h \leq \varepsilon \text{ for all } \xi \in \Xi \end{aligned} \quad (7)$$

where ε is some small tolerance for numerical reason, say 0.01.

By writing the expectations in (7) in terms of the densities $f_{\mathbf{X}^s|\xi}$ of \mathbf{X}^s , the above problem can be written in a Lagrangian form

$$\begin{aligned} & \min_{\hat{Q}(\cdot)} \int_{\Xi} E_{\xi} \left[\left| (X_1 - X_k) \hat{Q}(\mathbf{X}^s) + X_k - q(\xi, h) \right| \mid \mathbf{X}^s \right] f_{\mathbf{X}^s|\xi}(\mathbf{X}^s) dW(\xi) \\ & + \int_{\Xi} \lambda_1(\xi) E_{\xi} \left[\left(1 + \xi \left((X_1 - X_k) \hat{Q}(\mathbf{X}^s) + X_k \right) \right)^{-1/\xi} \mid \mathbf{X}^s \right] f_{\mathbf{X}^s|\xi}(\mathbf{X}^s) d\xi \\ & - \int_{\Xi} \lambda_2(\xi) E_{\xi} \left[\left(1 + \xi \left((X_1 - X_k) \hat{Q}(\mathbf{X}^s) + X_k \right) \right)^{-1/\xi} \mid \mathbf{X}^s \right] f_{\mathbf{X}^s|\xi}(\mathbf{X}^s) d\xi \end{aligned} \quad (8)$$

where the functions $\lambda_1(\xi)$ and $\lambda_2(\xi)$ are Lagrangian multipliers, and the expectations in the above expression can be numerically computed by Gaussian quadrature. Therefore, the limiting problem can be treated as estimating $q(\xi, h)$, a function of the scalar parameter $\xi \in \Xi$, based on a single observation \mathbf{X}^s from a parametric distribution indexed only by ξ . The only remaining challenge is thus to identify suitable Lagrangian multipliers. To this end, we resort to the numerical algorithm developed in Müller and Wang (2015). Further details are provided in the appendix.

2.2 Tail Conditional Expectation

Now consider the problem of constructing an asymptotically valid fixed- k estimator for the TCE: $T_n = E[Y_i | Y_i \geq Q(F, 1 - h/n)]$, for given h . Assume F is in the domain of attraction with tail index $\xi < 1$ (otherwise, the tail conditional expectation does not exist). We then impose the mean unbiasedness restriction on the estimator of TCE, denoted as $\hat{T}(\mathbf{Y})$, that is,

$$E \left[\hat{T}(\mathbf{Y}) \right] - T_n = 0. \quad (9)$$

Recall that for a positive random variable Z with CDF F_Z , $E[Z] = \int (1 - F_Z(z)) dz$. Denote $\hat{F}^{TCE}(\cdot)$ as the CDF that leads to the TCE as $\hat{T}(\mathbf{Y})$, then the constraint (9) is equivalent

¹The mean squared error criterion might not be well-defined for $\xi \geq 1/2$.

to

$$\int_{Q(F, 1-h/n)}^{\infty} \left(\hat{F}^{TCE}(z) - F(z) \right) dz = 0$$

which can be interpreted as an average quantile unbiasedness above the true quantile. In finance, the high quantile and tail conditional expectation can be interpreted as the VaR and the ES, respectively. The above expression then suggests that the expected shortfall captures the unbiasedness at all the confidence levels above a certain VaR, which, however, measures the risk at only one particular level. As pointed out by Basel Committee on Banking Supervision (2013): "*A number of weaknesses have been identified with using VaR for determining regulatory capital requirements, including its inability to capture "tail risk". For this reason, the Committee proposed in May 2012 to replace VaR with ES. ES measures the riskiness of a position by considering both the size and the likelihood of losses above a certain confidence level.*"

To impose the constraint (9), we have

$$\begin{aligned} & a_n^{-1} E_{\xi} \left[\hat{T}(\mathbf{Y}) - T_n \right] \\ = & E_{\xi} \left[\frac{\hat{T}(\mathbf{Y}) - b_n}{a_n} - \frac{T_n - b_n}{a_n} \right] \\ \rightarrow & E \left[(X_1 - X_k) \hat{T}(\mathbf{X}^s) + X_k - \tau(\xi, h) \right] \end{aligned}$$

where $\tau(\xi, h) = q(\xi, h) / (1 - \xi) - 1/\xi$ (cf. page 7 of Müller and Wang (2016)). Similar as before, we minimize a weighted average MAD criteria in estimating the TCE. After some calculation, the fixed- k asymptotic equivariant estimation problem about the tail conditional expectation can be written as

$$\begin{aligned} & \min_{\hat{Q}(\cdot)} \int_{\Xi} E_{\xi} \left[\left| (X_1 - X_k) \hat{Q}(\mathbf{X}^s) + X_k - \tau(\xi, h) \right| | \mathbf{X}^s \right] f_{\mathbf{X}^s | \xi}(\mathbf{X}^s) dW(\xi) \quad (10) \\ & + \int_{\Xi} \tilde{\lambda}_1(\xi) E_{\xi} \left[(X_1 - X_k) \hat{T}(\mathbf{X}^s) + X_k - \tau(\xi, h) | \mathbf{X}^s \right] f_{\mathbf{X}^s | \xi}(\mathbf{X}^s) d\xi \\ & - \int_{\Xi} \tilde{\lambda}_2(\xi) E_{\xi} \left[(X_1 - X_k) \hat{T}(\mathbf{X}^s) + X_k - \tau(\xi, h) | \mathbf{X}^s \right] f_{\mathbf{X}^s | \xi}(\mathbf{X}^s) d\xi \end{aligned}$$

where $(\tilde{\lambda}_1, \tilde{\lambda}_2)$ is another set of Lagrangian multipliers to be numerically determined by the generic algorithm developed in Müller and Wang (2015).

3 Derivation for Censored and Truncated Data

3.1 Censoring

Now consider data censoring such that in an i.i.d. sample of n observations, the largest m are censored with a known m . We first discuss the case in which the censoring value is unavailable. The case with a known censoring value follows from similar derivation, and is postponed to the Appendix. To reduce notation, we introduce estimation of the high quantile only since the TCE follows a similar argument.

Note that for fixed m and k , the extreme value theory (1) applies to the largest $m+k$ order statistics. Then, to implement the previously introduced approach, we modify the definition of \mathbf{X}^s as

$$\mathbf{X}_m^s = \left(\frac{X_{m+1} - X_{m+k}}{X_{m+1} - X_{m+k}}, \frac{X_{m+2} - X_{m+k}}{X_{m+1} - X_{m+k}}, \dots, \frac{X_{m+k} - X_{m+k}}{X_{m+1} - X_{m+k}} \right)$$

which is invariant to location and scale transformation, and construct the estimator of $Q(1-h/n)$ as $\hat{Q}(\mathbf{X}_m) = (X_{m+1} - X_{m+k})\hat{Q}(\mathbf{X}_m^s) + X_{m+k}$ where $\mathbf{X}_m = (X_{m+1}, \dots, X_{m+k})$. Then the density of \mathbf{X}_m^s as well as asymptotic quantile bias (5) and risk (6) can be adjusted accordingly. Therefore, the asymptotic Lagrangian problem can be written as

$$\begin{aligned} & \min_{\hat{Q}(\cdot)} \int E_\xi \left[\left| (X_{m+1} - X_{m+k})\hat{Q}(\mathbf{X}_m^s) + X_{m+k} - q(\xi, h) \right| \mathbf{1}_{\mathbf{X}_m^s} f_{\mathbf{X}_m^s|\xi}(\mathbf{X}_m^s) dW(\xi) \right] \quad (11) \\ & + \int \lambda_1(\xi) E_\xi \left[\left(1 + \xi \left((X_{m+1} - X_{m+k})\hat{Q}(\mathbf{X}_m^s) + X_{m+k} \right) \right)^{-1/\xi} \mathbf{1}_{\mathbf{X}_m^s} \right] f_{\mathbf{X}_m^s|\xi}(\mathbf{X}_m^s) d\xi \\ & - \int \lambda_2(\xi) E_\xi \left[\left(1 + \xi \left((X_{m+1} - X_{m+k})\hat{Q}(\mathbf{X}_m^s) + X_{m+k} \right) \right)^{-1/\xi} \mathbf{1}_{\mathbf{X}_m^s} \right] f_{\mathbf{X}_m^s|\xi}(\mathbf{X}_m^s) d\xi. \end{aligned}$$

3.2 Truncation

Truncated data exist when values outside a certain range are automatically eliminated, i.e., the observed data are generated from a truncated F . To capture the effect that only tail is truncated, we assume the truncation value is $Q(1-\tilde{h}/n)$ for some unknown \tilde{h} , and focus on the case in which this quantity is unknown (cf. Aban, Meerschaert, and Panorska (2006)). Similar derivation applies to the situation where the truncation value, $Q(1-\tilde{h}/n)$, is observed (\tilde{h} still unobserved). See Appendix for more details.

We still consider the largest k observations, whose limiting distribution is stated in the following lemma.

Lemma 1 *Suppose data are i.i.d. and generated from a top truncated CDF F at $q \left(1 - \tilde{h}/n\right)$ with some unknown fixed value $\tilde{h} \geq 0$ and F belongs to some domain of attraction. Then*

$$\frac{Y_{n:n} - b_n}{a_n}, \dots, \frac{Y_{n:n-k+1} - b_n}{a_n} \Rightarrow \tilde{\mathbf{X}} = \left(\tilde{X}_1, \dots, \tilde{X}_k\right)$$

where the PDF of $\tilde{\mathbf{X}}$ is $G_{\xi, \tilde{h}}(\tilde{x}_k) \prod_{i=1}^k g_{\xi, \tilde{h}}(\tilde{x}_i) / G_{\xi, \tilde{h}}(\tilde{x}_i)$ with $G_{\xi, \tilde{h}}(x) = \exp(\tilde{h}) G_{\xi}(x)$ and $g_{\xi, \tilde{h}}(x) = \exp(\tilde{h}) g_{\xi}(x)$.

The density of the maximal invariant

$$\tilde{\mathbf{X}}^s = \frac{\tilde{X}_1 - \tilde{X}_k}{\tilde{X}_1 - \tilde{X}_k}, \dots, \frac{\tilde{X}_k - \tilde{X}_k}{\tilde{X}_1 - \tilde{X}_k}$$

can be derived with more tedious algebra, as well as the asymptotic bias (5) and risk (6). In particular, the estimator of $Q(1 - h/n)$ is constructed as $\hat{Q}(\tilde{\mathbf{X}}) = (\tilde{X}_1 - \tilde{X}_k) \hat{Q}(\tilde{\mathbf{X}}^s) + \tilde{X}_k$, and \tilde{h} shows up in the limiting problem as an additional nuisance parameter. Given a two-dimensional weight $W(\xi, \tilde{h})$ and the set of Lagrangian multipliers $\lambda_1(\xi, \tilde{h})$ and $\lambda_2(\xi, \tilde{h})$ defined on $\Xi \times H$ for $H = [0, \bar{h}]$ with some pre-specified \bar{h} , the Lagrangian problem can be rewritten as

$$\begin{aligned} & \min_{\hat{Q}(\cdot)} \int_{\Xi \times H} E_{\xi, \tilde{h}} \left[\left| (\tilde{X}_1 - \tilde{X}_k) \hat{Q}(\tilde{\mathbf{X}}^s) + \tilde{X}_k - q(\xi, h) \right| |\tilde{\mathbf{X}}^s| f_{\tilde{\mathbf{X}}^s | \xi, \tilde{h}}(\tilde{\mathbf{X}}^s) \right] dW(\xi, \tilde{h}) \quad (12) \\ & + \int_{\Xi \times H} \lambda_1(\xi, \tilde{h}) E_{\xi, \tilde{h}} \left[\left(1 + \xi \left((\tilde{X}_1 - \tilde{X}_k) \hat{Q}(\tilde{\mathbf{X}}^s) + \tilde{X}_k \right) \right)^{-1/\xi} |\tilde{\mathbf{X}}^s| \right] f_{\tilde{\mathbf{X}}^s | \xi, \tilde{h}}(\tilde{\mathbf{X}}^s) d\xi d\tilde{h} \\ & - \int_{\Xi \times H} \lambda_2(\xi, \tilde{h}) E_{\xi, \tilde{h}} \left[\left(1 + \xi \left((\tilde{X}_1 - \tilde{X}_k) \hat{Q}(\tilde{\mathbf{X}}^s) + \tilde{X}_k \right) \right)^{-1/\xi} |\tilde{\mathbf{X}}^s| \right] f_{\tilde{\mathbf{X}}^s | \xi, \tilde{h}}(\tilde{\mathbf{X}}^s) d\xi d\tilde{h}. \end{aligned}$$

Note that our approach can be easily adapted to estimate the $1 - h/n$ quantile of the truncated distribution, which equals the $1 - (h + \tilde{h})/n + h\tilde{h}/n^2$ quantile of the original distribution. Therefore, the same asymptotic problem can be set up with h replaced by $h + \tilde{h}$ since the term $h\tilde{h}/n^2$ is asymptotically negligible.

4 Monte Carlo Simulations

This section reports some small sample results for $h = 0.5$ and 5 and $n = 250$, corresponding to confidence levels at 99.8% and 98%. For simplicity, we only report the results for $k = 20$. We consider six data generating processes: A Pareto law with tail index equal to $\xi = 0.25$,

a standard normal distribution, a standard lognormal distribution, a Student-t distribution with 3 degrees of freedom, and the empirical distributions the GARCH(1,1) residuals of S&P500 and Nasdaq daily returns from 02/08/1990 to 04/17/2017. For the censored data model, the largest 5 observations are censored. For the truncated data model, we generate the data from those distributions truncated at $Q(F, 1 - \tilde{h}/n)$ with $\tilde{h} = \{0, 1, 2\}$, and impose the unbiasedness for the data truncated at up to $Q(F, 1 - 2/n)$, i.e., $\bar{h} = 2$.

Tables 1 and 2 present the (quantile) bias and the mean absolute deviation of four methods: (i) the suggested procedure in previous sections (fixed- k), (ii) the procedure implemented by fitting the exceedances, $Y_{n:n} - Y_{n:n-k+1}, \dots, Y_{n:n-k} - Y_{n:n-k+1}$ with a generalized Pareto distribution (GPD) as described in McNeil and Frey (2000), (iii) the estimators described in Chapter 4 of de Haan and Ferreira (2007) textbook (dH-F): $\hat{Q}_{dHF} = Y_{n:n-k} + \hat{a}(n/k) ((h/k)^{-\hat{\xi}^M} - 1) / \hat{\xi}^M$ and $\hat{T}_{dHF} = Y_{n:n-k} + \hat{a}(n/k) ((h/k)^{-\hat{\xi}^M} - 1 + \hat{\xi}^M) / (\hat{\xi}^M (1 - \hat{\xi}^M))$ where $\hat{a}(n/k)$ and $\hat{\xi}^M$ moment estimators of the scale and the tail index, correspondingly, (see also Dekkers and de Haan (1989) and de Haan and Rootzén (1993)); and (iv) the classic Weissman (1978) estimator (W-H): $\hat{Q}_{WH} = Y_{n:n-k} (h/k)^{-\hat{\xi}^H}$ and $\hat{T}_{WH} = (Y_{n:n-k} (h/k)^{-\hat{\xi}^H}) / (1 - \hat{\xi}^H)$ where $\hat{\xi}^H$ denotes the classic Hill (1975) estimator. For all three $k_n \rightarrow \infty$ methods we impose the same parameter space restriction $\Xi = [-1/2, 1/2]$ on the tail index that we chose in the implementation of the fixed- k method.

The bias for quantile is reported as $100 \left(E \left[P \left(Y_i > \hat{Q}(\mathbf{Y}) \mid \mathbf{Y} \right) \right] - h/n \right)$, that is, the probability measured in percentage that an additional random draw from F is larger than the quantile estimator minus the target tail probability. For TCE, we report both the mean bias $E \left[\hat{T}(\mathbf{Y}) - E \left[Y_i \mid Y_i > \hat{Q}(\mathbf{Y}) \right] \right]$ and the bias measured in h . More precisely, the h bias is defined as $\hat{h} - h$ where \hat{h} is value of h that $E \left[\hat{T}(\mathbf{Y}) \right]$ corresponds to in the underlying distribution. Linear interpolation is implemented for the GARCH(1,1) residuals. We find that the new method has much smaller risk and very accurate coverage across all h , in contrasts to the other three methods for very small h . In particular, the quantile bias of the GPD estimator is approximately 0.2% at $h = 0.5$, which means the GPD method approximately delivers the 99.6% level quantile while the true target is 99.8%th quantile. The other two estimators exhibit small sample biases that differ a lot across distributions, indicating that $k = 20$ is still too small for their increasing- k asymptotics to perform satisfactorily. For relatively large h such as 5, the unbiasedness restriction is much easier to impose as reflected by substantially less MADs. This is because the quantity is more close to the central part of the distribution and therefore more observations can be collected from the right side.

Table 1: Small Sample Properties for the $1 - h/n$ Quantile

Quantile	$h = 0.5$							
	fixed- k		GPD		dH-F		W-H	
	Bias	MAD	Bias	MAD	Bias	MAD	Bias	MAD
Pareto	0.00	2.89	0.23	1.11	0.21	0.94	0.08	0.83
Normal	0.02	0.58	0.24	0.30	0.18	0.30	-0.15	0.88
Lognormal	-0.01	13.6	0.22	5.25	0.18	4.71	-0.02	4.48
Student-t	0.01	6.29	0.23	2.51	0.20	2.19	-0.01	2.00
SP500	0.01	0.84	0.24	0.42	0.19	0.40	-0.13	0.93
Nasdaq	0.02	0.84	0.24	0.39	0.17	0.37	-0.12	0.79
	$h = 5$							
	Bias	MAD	Bias	MAD	Bias	MAD	Bias	MAD
Pareto	-0.01	0.22	0.02	0.21	0.16	0.20	0.21	0.20
Normal	0.02	0.13	0.17	0.13	0.21	0.13	0.37	0.13
Lognormal	-0.01	1.08	0.17	1.01	0.21	0.94	0.40	0.82
Student-t	-0.04	0.48	0.14	0.44	0.19	0.42	0.35	0.37
SP500	-0.00	0.15	0.15	0.15	0.25	0.15	0.37	0.15
Nasdaq	-0.01	0.13	0.15	0.13	0.22	0.13	0.35	0.13

Note: Entries are quantile biases and mean absolute deviations of estimators in a sample of size $n = 250$ about the $1 - h/n$ quantile of the underlying distribution F , based on the largest 20 order statistics. See the main text for a description of the four types of estimators. Based on 5,000 Monte Carlo simulations.

Some unreported results suggest that the empirical quantile works very well in terms of both bias and MAD properties, as long as it is well defined (h is an integer and larger than 1). Following Arnold, Balakrishnan, and Nagaraja (1992), it is easy to show that

$$E[1 - F(Y_{n:n-h+1})] = -\frac{h}{n(n+1)}.$$

This result shows that the empirical quantile is nearly quantile unbiased when n is reasonably large. But the bias of the empirical TCE estimator, i.e., taking average of the largest $h - 1$ order statistics, depends on the underlying distribution F and can be much larger compared with the fixed- k method.

For most F and h combinations, choosing a larger k does not degrade bias by much, and the MAD decreases. This is because that the distributions considered in the experiment are benign in the sense that a relatively large fraction of the data can be well approximated by a Pareto tail. But as demonstrated Müller and Wang (2016), it is easy to construct underlying distributions whose tail behavior is so worse behaved that any choice of a moderately large

Table 2: Small Sample Properties for TCE above the $1 - h/n$ Quantile

TCE	$h = 0.5$											
	fixed- k			GPD			dH-F			W-H		
	Bias	h Bias	MAD	Bias	h Bias	MAD	Bias	h Bias	MAD	Bias	h Bias	MAD
Pareto	-0.07	0.02	2.29	0.86	-0.20	3.39	0.06	-0.02	1.95	-0.16	0.06	1.53
Normal	-0.04	0.07	0.44	-0.15	0.33	0.48	0.01	-0.01	0.47	-1.20	14.9	2.02
Lognormal	0.40	-0.03	10.3	4.82	-0.24	15.6	-2.13	0.18	10.6	-9.45	1.90	14.5
Student-t	-0.53	0.07	5.53	1.74	-0.16	8.32	0.02	-0.00	5.09	-2.83	0.59	5.70
SP500	-0.05	0.06	0.67	-0.12	0.16	0.79	0.04	-0.05	0.67	-1.07	5.20	2.23
Nasdaq	-0.08	0.09	0.63	-0.20	0.28	0.75	0.11	-0.10	0.67	-0.72	2.25	1.74

	$h = 5$											
	Bias	h Bias	MAD	Bias	h Bias	MAD	Bias	h Bias	MAD	Bias	h Bias	MAD
Pareto	0.02	-0.11	0.52	0.06	-0.32	0.62	0.10	-0.52	0.49	-0.04	0.24	0.48
Normal	0.00	0.00	0.18	-0.05	0.72	0.18	0.02	-0.23	0.18	-0.38	7.97	0.38
Lognormal	0.08	-0.09	2.45	0.28	-0.32	2.94	-0.08	0.10	2.47	-1.90	3.03	2.43
Student-t	0.03	-0.07	1.21	0.13	-0.31	1.43	0.11	-0.27	1.16	-0.72	2.34	1.09
SP500	0.01	-0.11	0.22	-0.05	0.61	0.23	0.02	-0.20	0.23	-0.40	6.97	0.42
Nasdaq	0.01	-0.12	0.20	-0.04	0.53	0.21	0.02	-0.22	0.20	-0.33	6.09	0.35

Note: Entries are biases and mean absolute deviations of estimators in a sample of size $n = 250$ about the tail conditional expectation above the $1 - h/n$ quantile of the underlying distribution F , based on the largest 20 order statistics. See the main text for a description of the three types of estimators and the definitions of biases. Based on 5,000 Monte Carlo simulations.

Table 3: Small Sample Properties for the $1 - h/n$ Quantile with Censored Data

Quantile	$h = 0.5$				$h = 5$			
	fixed- k		GPD		fixed- k		GPD	
	Bias	MAD	Bias	MAD	Bias	MAD	Bias	MAD
Pareto	0.01	5.19	2.20	2.12	0.00	0.28	2.32	0.44
Normal	0.04	3.12	2.10	0.86	0.00	0.16	2.50	0.35
Lognormal	-0.02	25.6	2.18	10.2	-0.02	1.38	2.29	2.10
Student-t	0.04	10.7	2.17	4.67	-0.02	0.59	2.30	0.91
SP500	0.06	3.67	2.16	1.07	0.02	0.19	2.47	0.39
Nasdaq	0.06	3.29	2.15	0.92	0.02	0.17	2.47	0.34

Note: Entries are quantile biases and mean absolute deviations of estimators in a sample of size $n = 250$ about the $1 - h/n$ quantile of the underlying distribution F , with the largest 5 observations censored. See the main text for a description of the two types of estimators. Based on 5,000 Monte Carlo simulations.

k could lead to poor finite sample approximations of the extreme value theory.

Table 3 lists the small sample bias of the fixed- k and GPD methods for censored data, since the other two estimators are not applicable. These results suggest that the GPD method always substantially underestimate the high quantile. In particular, the GPD estimator is approximately the 97.8% quantile while the true target is 99.8%. Table 4 depicts the performance of the fixed- k method with data truncation. As far as we know, this is the only method that is applicable to truncated data and imposes no parametric assumptions on F . These numbers suggest that the new approach has an excellent small sample bias. The substantial difference in MAD across models indicates that the largest observation is very informative about the extreme quantile. Hence if it is unobserved due to either censoring or truncation, the unbiasedness has to be satisfied at a much larger cost in terms of MAD.

5 Application to Hurricane Damage

To illustrate empirical relevance, this session applies the new approach to estimate high quantiles of the U.S. Hurricanes damage. Müller and Wang (2016) construct confidence intervals for the high quantiles of hurricane damage based on the historical data collected from 1995-2010, while this paper completes their study by providing a point estimate and using more updated data. In particular, we collect the damage estimate of 26 costliest U.S. hurricanes in the period 1995-2014 from <http://www.icatdamageestimator.com> (see also Pielke, Gratz, Landsea, Collins, Saunders, and Musulin (2008)). These data are damage

Table 4: Small Sample Properties for the $1 - h/n$ Quantile with Truncated Data

Quantile	$h = 0.5$						$h = 5$					
	$\tilde{h} = 0$		$\tilde{h} = 1$		$\tilde{h} = 2$		$\tilde{h} = 0$		$\tilde{h} = 1$		$\tilde{h} = 2$	
Truncation	Bias	MAD	Bias	MAD	Bias	MAD	Bias	MAD	Bias	MAD	Bias	MAD
Pareto	0.05	9.98	-0.01	5.85	0.01	3.27	0.06	1.73	-0.01	1.48	0.05	1.09
Normal	0.07	5.19	-0.01	2.64	0.03	1.70	0.06	1.02	-0.05	0.84	0.05	0.69
Lognormal	0.03	48.3	-0.00	28.6	-0.00	16.1	0.00	8.35	-0.02	7.24	0.03	5.28
Student-t	0.07	20.5	0.02	12.3	0.03	6.57	0.13	3.49	0.12	3.06	0.15	2.23
SP500	0.07	6.39	0.00	3.19	0.07	2.02	0.05	1.18	-0.11	0.96	-0.01	0.78
Nasdaq	0.10	5.72	0.03	2.85	0.04	1.89	0.04	1.06	-0.01	0.87	0.06	0.69

Note: Entries are quantile biases and mean absolute deviations of estimators in a sample of size $n = 250$ about the $1 - h/n$ quantile of the underlying distribution F , with data generated from the truncated F at $Q(F, 1 - \tilde{h}/n)$. Based on 5,000 Monte Carlo simulations.

estimates measured in 2017 US dollars and adjusted for inflation, wealth per capita and affected county population, Panel A in Table 5 replicates these data points for convenience.

Note that in this example, the number n of total tropical cyclones is not known to us. Nevertheless, under the assumption that hurricane damage is i.i.d. and hurricane arrival is stationary, the $1 - h/n$ quantiles can be naturally interpreted as follows: a hurricane causing at least that amount of damage is expected every $20/h$ years (cf.).

Panel B in Table 5 provides estimates for $Q(1 - h/n)$ for $h \in \{0.5, 5\}$ using the fixed- k approach developed in previous sections for three models of data. The "complete" data model means that we simply use the largest k order statistics. For the "censored" data model, we only use the 2nd until the 21st largest data with the largest one (Katrina in 2005) being treated as censored ($m = 1$). In the last "truncated" data, we drop the largest observation as well and consider the 2nd to the 21st largest damage, with the assumption that at most $1/n$ tail probability of the underlying distribution cannot be recorded ($\bar{h} = 1$). These two incomplete data experiments are artificial but can be helpful to eliminate the measure error problem that the exact damage of the extremely costly hurricane is hard to estimate (cf. Downton and Pielke (2005) and Pielke, Gratz, Landsea, Collins, Saunders, and Musulin (2008)). All the estimates for $h = 0.5$ are much larger than the sample maximum, suggesting that the distribution of hurricane damage exhibits a heavy tail. In addition, the two incomplete data approaches deliver larger estimates than the complete data one for $h = 0.5$. The intuition is that the cost of the most severe hurricane in history could be much

Table 5: Empirical Results on Damage of U.S. Mainland Hurricanes

Panel A: 26 Costliest Mainland United State Hurricanes, 1995-2014

Rank	1	2	3	4	5	6	7	8	9	10	11	12	13
Damage	91.13	54.66	31.02	26.04	24.4	22.32	15.68	12.96	11.67	10.9	10.21	9.69	9.08
Rank	14	15	16	17	18	19	20	21	22	23	24	25	26
Damage	7.93	5.90	5.02	3.28	3.08	2.5	2.03	1.7	1.44	1.34	1.13	1.03	0.99

Panel B: Estimate of $1 - h/n$ Quantile of the United State Hurricane Damage

Quantile	$h = 0.5$			$h = 5$		
Data Type	complete	censored	truncated	complete	censored	truncated
Estimate	139.2	170.6	189.7	23.4	23.8	21.1

Note: Data source: <http://www.icatdamageestimator.com>. Measured in 2017 U.S.\$ billions. See the main text for a description of the three data type. Based on the $k = 20$ order statistics from Panel A.

larger than 91.13 given that the 2nd largest one (Sandy 2012) is already so devastating. Hence treating Katrina as the observed sample maximum leads to smaller estimates than if the largest one is left as unobserved. But such difference is small when estimating the quantile $Q(1 - 5/n)$, which is relatively less affected by the sample maximum. Finally, all three estimates are substantial amounts from macroeconomic perspective and hence indicate a strong need of insurance.

6 Generalization to Stochastic Volatility Models with Complete Data

In financial applications, the i.i.d. assumption is usually violated since data may exhibit time series correlation and heteroskedasticity. To overcome this difficulty, we can resort to the stochastic volatility models with i.i.d. driving innovations. In particular, we assume an $AR(\tilde{p})$ -GARCH(p, q) model (cf. McNeil and Frey (2000)). The conditional quantile of a one-step ahead forecast then simply becomes the product of the square root of the conditional heteroskedasticity function and the estimated quantile or TCE of the driving innovations. We show that estimation error of the AR and GARCH parameters is negligible for our asymptotic theory when the data is complete, so that we can apply our estimators to the estimated innovations.²

²It is unclear about how to obtain a consistent estimator of the GARCH parameters with incomplete data. If such consistent estimates were available, the approach introduced before can still be applied.

More specifically, let Z_t denote the real data which is assumed to be the following stationary time series

$$\begin{aligned} Z_t &= \mu_t + \sigma_t Y_t \\ \sigma_t^2 &= \alpha_0 + \alpha_1 Y_{t-1}^2 + \dots + \alpha_q Y_{t-q}^2 + \beta_1 \sigma_{t-1}^2 + \dots + \beta_p \sigma_{t-p}^2 \\ \mu_t &= \bar{\mu} + \phi_1 Z_{t-1} + \dots + \phi_{\tilde{p}} Z_{t-\tilde{p}}. \end{aligned}$$

where the innovation Y_t is i.i.d. with CDF F , as specified in the previous discussion. As standard in the literature, we assume that μ_t and σ_t are measurable with respect to \mathcal{F}_{t-1} , the information available up to time $t - 1$.

To estimate the unknown coefficients, we can apply the pseudo maximum likelihood (PML) estimator, which maximizes the likelihood under the assumption of standard Gaussian innovations. Given the PML estimator, we can back out the estimated conditional mean and standard deviation series, denoted as $\{\hat{\mu}_t\}$ and $\{\hat{\sigma}_t\}$, respectively. Then, the residuals can be calculated as

$$\hat{Y}_t = \frac{Z_t - \hat{\mu}_t}{\hat{\sigma}_t}$$

which can be used as i.i.d. data for estimating VaR and TCE. The following theorem shows that the error in fitting the AR-GARCH type models is asymptotically negligible if the estimator of the coefficients is consistent.

Theorem 1 *Suppose there exists a consistent estimator of the AR(\tilde{p})-GARCH(p, q) coefficients for some known positive integers (\tilde{p}, p, q) , then the estimated innovations $\{\hat{Y}_t\}$ satisfy the extreme value theorem, i.e., the largest k innovations with a fixed k , $(\hat{Y}_{n:n}, \dots, \hat{Y}_{n:n-k+1})$, satisfy (3).*

The proof is in the appendix. This theorem validates the weak convergence (3) for the ordered estimated innovations $(\hat{Y}_{n:n}, \dots, \hat{Y}_{n:n-k+1})$, and hence the previously suggested approach is applicable again. As a summary, our estimator can be implemented by the following steps:

Step 1 For time $t > n$, fit the data $\{Z_t, Z_{t-1}, \dots, Z_{t-n+1}\}$ with an AR-GARCH type model and obtain the standardized innovations $\{\hat{Y}_t\}$

Step 2 Compute the empirical estimators or the fixed- k estimators \hat{Q} and \hat{T} by using the largest k standardized innovations, denoted by $\hat{\mathbf{Y}}_t$, and solving the Lagrangian problems (8) and (10).

Step 3 Plug in the conditional mean and standard deviation at time t to construct the one-step prediction of the VaR and TCE, that is, $\hat{Q}(\hat{\mathbf{Y}}_t) \hat{\sigma}_t + \hat{\mu}_t$ and $\hat{T}(\hat{\mathbf{Y}}_t) \hat{\sigma}_t + \hat{\mu}_t$.

Appendix

A.1 Computational Details

The estimators defined in (8) and (10) require evaluation of the following items. Use the expression for f_X below (2), and define $\Gamma(\cdot)$ as the Gamma function and $b_0(\xi) = -1/\xi$ for $\xi < 0$, and $b_0(\xi) = \infty$ otherwise. Also define $e(\mathbf{x}^s, s) = \exp\left(- (1 + 1/\xi) \sum_{i=1}^k \log(1 + \xi x_i^s s)\right)$. Then for a positive $\hat{Q}(\mathbf{X}^s)$, some calculations yield the following expressions.

1. For the complete data case, the asymptotic bias terms read

$$\begin{aligned} & E_\xi \left[\left| (X_1 - X_k) \hat{Q}(\mathbf{X}^s) + X_k - q(\xi, h) \right| \middle| \mathbf{X}^s \right] f_{\mathbf{X}^s | \xi}(\mathbf{X}^s) \\ &= \Gamma(k+1) \int_0^{b_0(\xi)} (1 + \xi s q(\mathbf{X}^s))^{-1/\xi} s^{k-2} e(\mathbf{x}^s, s) ds \\ & E_\xi \left[(X_1 - X_k) \hat{T}(\mathbf{X}^s) + X_k - \tau(\xi, h) \middle| \mathbf{X}^s \right] f_{\mathbf{X}^s | \xi}(\mathbf{X}^s) \\ &= \hat{T}(\mathbf{X}^s) \Gamma(k - \xi) \int_0^{b_0(\xi)} s^{k-1} e(\mathbf{x}^s, s) ds \\ & \quad + \left(\frac{\Gamma(k - \xi) - \Gamma(k)}{\xi} - \tau(\xi, h) \Gamma(k) \right) \int_0^{b_0(\xi)} s^{k-2} e(\mathbf{x}^s, s) ds \end{aligned}$$

and the risk terms read

$$\begin{aligned} & E_\xi \left[\left| (X_1 - X_k) \hat{Q}(\mathbf{X}^s) + X_k - q(\xi, h) \right| \middle| \mathbf{X}^s \right] f_{\mathbf{X}^s | \xi}(\mathbf{X}^s) \\ &= |\xi|^{-1} \int_0^{b_0(\xi)} g(s) s^{k-2} e(\mathbf{x}^s, s) ds \end{aligned}$$

where for $a(s) = 1 + s\xi\hat{Q}(\mathbf{X}^s)$

$$g(s) = \begin{cases} \begin{pmatrix} -h^{-\xi}(\Gamma[k] - 2\Gamma[k, a(s)^{1/\xi} h]) \\ +a(s)(\Gamma[k - \xi] - 2\Gamma[k - \xi, a(s)^{1/\xi} h]) \end{pmatrix} & \text{if } a(s) > 0 \\ (h^{-\xi}\Gamma[k] - a(s)\Gamma[k - \xi]) & \text{otherwise,} \end{cases}$$

and

$$\begin{aligned} & E_\xi \left[\left| (X_1 - X_k) \hat{T}(\mathbf{X}^s) + X_k - \tau(\xi, h) \right| \middle| \mathbf{X}^s \right] f_{\mathbf{X}^s} \\ &= |\xi|^{-1} \int_0^{b_0(\xi)} \tilde{g}(s) s^{k-2} e(\mathbf{x}^s, s) ds \end{aligned}$$

where for $a(s) = 1 + s\xi\hat{T}(\mathbf{X}^s)$

$$\tilde{g}(s) = \begin{cases} \begin{pmatrix} -\frac{h^{-\xi}}{1-\xi}(\Gamma[k] - 2\Gamma[k, a(s)^{1/\xi} h(1-\xi)^{1/\xi}]) \\ +a(s)(\Gamma[k-\xi] - 2\Gamma[k-\xi, a(s)^{1/\xi} h(1-\xi)^{1/\xi}]) \end{pmatrix} & \text{if } a(s) > 0 \\ (h^{-\xi}\Gamma[k] - a(s)\Gamma[k-\xi]) & \text{otherwise.} \end{cases}$$

2. For the censored data case, the asymptotic bias term reads

$$\begin{aligned} & E_\xi \left[\left(1 + \xi \left((X_{m+1} - X_{m+k}) \hat{Q}(\mathbf{X}_m^s) + X_{m+k} \right) \right)^{-1/\xi} \middle| \mathbf{X}_m^s \right] f_{\mathbf{X}_m^s | \xi}(\mathbf{X}_m^s) \\ &= \frac{\Gamma(m+k+1)}{m!} \int_0^{b_0(\xi)} (1 + \xi s q(\mathbf{X}^s))^{-1/\xi} s^{k-2} \exp\left(-\frac{m}{\xi} \log(1 + \xi s)\right) e(\mathbf{x}^s, s) ds \end{aligned}$$

and the asymptotic risk term reads

$$\begin{aligned} & E_\xi \left[\left| (X_{m+1} - X_{m+k}) \hat{Q}(\mathbf{X}_m^s) + X_{m+k} - q(\xi, h) \right| \middle| \mathbf{X}_m^s \right] f_{\mathbf{X}_m^s | \xi}(\mathbf{X}_m^s) \\ &= \frac{1}{m! |\xi|} \int g(k, h, \xi, s) \exp\left(-\frac{m}{\xi} \log(1 + \xi s)\right) e(\mathbf{x}_m^s, s) s^{k-2} ds \end{aligned}$$

where for $a(s) = (1 + s\xi\hat{Q}(\mathbf{X}^s))$,

$$g(s) = \begin{cases} \begin{pmatrix} -h^{-\xi}(\Gamma[k+m] - 2\Gamma[k+m, a(s)^{1/\xi} h]) \\ +a(s)(\Gamma[k+m-\xi] - 2\Gamma[k+m-\xi, a(s)^{1/\xi} h]) \end{pmatrix} & \text{if } a(s) > 0 \\ (h^{-\xi}\Gamma[k+m] - a(s)\Gamma[k+m-\xi]) & \text{otherwise.} \end{cases}$$

3. for the truncated data case, the asymptotic bias term reads

$$\begin{aligned} & E_{\xi, \tilde{h}} \left[\left| (\tilde{X}_1 - \tilde{X}_k) \hat{Q}(\tilde{\mathbf{X}}) + \tilde{X}_k - q(\xi, h) \right| \middle| \tilde{\mathbf{X}}^s \right] f_{\tilde{\mathbf{X}}^s | \xi, \tilde{h}}(\tilde{\mathbf{X}}^s) \\ &= \exp(\tilde{h}) \int_0^{b_0(\xi)} \Gamma(k+1, \tilde{h}(1+\xi s)^{1/\xi}) \left(1 + \xi s q(\tilde{\mathbf{X}}^s) \right)^{-1/\xi} s^{k-2} e(\tilde{\mathbf{x}}_m^s, s) ds \end{aligned}$$

and the asymptotic risk term reads

$$\begin{aligned} & E_{\xi, \tilde{h}} \left[\left(1 + \xi \left((\tilde{X}_1 - \tilde{X}_k) \hat{Q}(\tilde{\mathbf{X}}^s) + \tilde{X}_k \right) \right)^{-1/\xi} \middle| \tilde{\mathbf{X}}^s \right] f_{\tilde{\mathbf{X}}^s | \xi, \tilde{h}}(\tilde{\mathbf{X}}^s) \\ &= \exp(\tilde{h}) |\xi|^{-2} \int_0^{b_0(\xi)} g(k, h, \xi, s) e(\tilde{\mathbf{x}}_m^s, s) s^{k-2} ds \end{aligned}$$

where for $a(s) = 1 + s\xi\hat{Q}(\tilde{\mathbf{X}}^s)$

$$g(k, h, \xi, s)$$

$$= \begin{cases} \xi \left(-h^{-\xi}(\Gamma[k, \tilde{h}(1+\xi s)^{1/\xi}] - 2\Gamma[k, a(s)^{1/\xi} h]) + a(s) (\Gamma[k - \xi, \tilde{h}(1+\xi s)^{1/\xi}] - 2\Gamma[k - \xi, a(s)^{1/\xi} h]) \right) \\ \quad \text{if } \xi > 0, a(s) > h^{-\xi}(1+\xi s)/\tilde{h}^{-\xi} \text{ or } \xi < 0, 0 < a(s) < h^{-\xi}(1+\xi s)/\tilde{h}^{-\xi} \\ \quad \xi \left(h^{-\xi}\Gamma[k, \tilde{h}(1+\xi s)^{1/\xi}] - a(s) (\Gamma[k - \xi, \tilde{h}(1+\xi s)^{1/\xi}]) \right) \\ \quad \text{if } \xi > 0, 0 < a(s) < h^{-\xi}(1+\xi s)/\tilde{h}^{-\xi} \text{ or } \xi < 0, a(s) < 0 \text{ or } \xi < 0, a(s) > h^{-\xi}(1+\xi s)/\tilde{h}^{-\xi}. \end{cases}$$

We evaluate these by numerical quadrature.

To determine the suitable Lagrangian multipliers λ and $\tilde{\lambda}$, we follow the algorithm suggested by Müller and Wang (2015). In particular, for the complete and censored data case, we restrict λ_i and $\tilde{\lambda}_i$ for $i = 1$ and 2 to be discrete distributions with support on $\Xi = \{-1/2, -1/2 + 1/19, \dots, 1/2\}$, and determine the 20 point masses by fixed-point iterations based on importance sample Monte Carlo estimates of bias. In particular, we simulate the biases with 5,000 i.i.d. draws from a proposal with ξ randomly drawn from Ξ , and iteratively increase or decrease the 20 point masses on Ξ as a function of whether the (estimated) bias given that value of ξ is larger or smaller than zero. After 4000 iterations, the resulting discrete distribution is a candidate for the Lagrangian multiplier. Regarding the truncated data, we take $\Xi \times H = \{-1/2, -1/2 + 1/9, \dots, 1/2\} \times \{0, 0.5, 1.0, 1.5, 2.0\}$ and compute the Lagrangian multipliers on this 10×5 grids. For the weighting function W , we simply use a uniform weight on Ξ and an exponential weight on H for the truncated data. Note that the choice of $\xi \leq 1/2$ covers all the distributions with a finite second moment. Our approach can be easily extended to cover larger range of ξ .

For any given k and h , the Lagrangian multipliers λ and $\tilde{\lambda}$ only need to be determined once. Conditional on λ and $\tilde{\lambda}$, the estimator is readily computed from (8), (10), (11), and (12). The tables of λ and $\tilde{\lambda}$ and corresponding Matlab code are provided on the website: <https://sites.google.com/site/yulongwanghome/>.

A.2 Data with Known Censoring or Truncation value

In the censored data case, if the censoring value is also observed, we may still consider maximal invariant introduced in Section 3.1. Denote the censoring value as c , we have

$$\frac{Y_{n:n} - Y_{n:n-(m+k)+1}}{Y_{n:n-m} - Y_{n:n-(m+k)+1}}, \dots, \frac{Y_{n:n-(m+k)+1} - Y_{n:n-(m+k)+1}}{Y_{n:n-m} - Y_{n:n-(m+k)+1}} \Rightarrow \mathbf{X}^s$$

where $x_i^s > t = \frac{c - Y_{(m+k)}}{Y_{(m+1)} - Y_{(m+k)}} > 1$ for $i \leq m$ and $x_i^s \in [0, 1]$ for $i > m$. The density of $\mathbf{X}_m^s = (X_{m+1}^s, \dots, X_{m+k}^s)$ can be derived as follows

$$f_{\mathbf{X}_m^s} = \Gamma(m+k) \int_{1+\xi x_i^s > 0 \text{ for } i \leq m+k} s^{m+k-2} \exp\left(-\left(1 + \frac{1}{\xi}\right) \sum_{i=1}^{m+k} \log(1 + \xi x_i^s s)\right) ds$$

$$f_{\mathbf{X}_m^s | \xi}(\mathbf{X}_m^s) = \int \dots \int_{x_1^s \geq x_2^s \geq \dots \geq x_m^s > t} f_{\mathbf{X}^s | \xi}(\mathbf{x}^s) (dx_1^s \dots dx_m^s)$$

$$= \Gamma(m+k) \int_0^{b_0(\xi)} \frac{s^{k-2}}{m!} (1+\xi ts)^{-m/\xi} \exp\left(-\left(1+\frac{1}{\xi}\right) \sum_{i=m+1}^{m+k} \log(1+\xi x_i^s s)\right) ds,$$

and then the asymptotic bias and risk terms can be derived in a similar fashion.

In the truncated data case with a known truncation value c , we can still model $c = Q(1 - \tilde{h}/n)$ for some unknown \tilde{h} since the quantile function $Q(\cdot)$ is unknown (and not easy to estimate). Consider

$$\begin{aligned} & \frac{Y_{n:n} - Y_{n:n-k+1}}{c - Y_{n:n-k+1}}, \dots, \frac{Y_{n:n-k+1} - Y_{n:n-k+1}}{c - Y_{n:n-k+1}} \\ \Rightarrow \tilde{\mathbf{X}}^s(c) & \equiv \frac{\tilde{X}_1 - \tilde{X}_k}{q_\xi(\tilde{h}) - \tilde{X}_k}, \dots, \frac{\tilde{X}_k - \tilde{X}_k}{q_\xi(\tilde{h}) - \tilde{X}_k}. \end{aligned}$$

Given the density of $\tilde{\mathbf{X}}$ remains the same as stated in Lemma 4, the density of $\tilde{\mathbf{X}}^s(c)$ can be derived in a similar fashion as

$$f_{\tilde{\mathbf{X}}^s|\xi, \tilde{h}}(\tilde{\mathbf{X}}^s) = \exp(\tilde{h}) \int_0^{b_1(\xi)} \Gamma(k, \tilde{h}(1+\xi s)^{1/\xi}) s^{k-2} e(\tilde{\mathbf{x}}^s, s) ds$$

where $b_1(\xi) = \infty$ if $\xi > 0$ and $-1/(\xi \tilde{x}_1^s(c))$ if $\xi < 0$, and similarly for the asymptotic bias and risk terms as well.

A.3 Proof

Proof of Lemma 1. Given the CDF (2), it is equivalent to show that

$$\frac{Y_{(1)} - b_n, \dots, Y_{(k)} - b_n}{a_n} \Rightarrow \left(\frac{(h^c + E_1^*)^{-\xi} - 1}{\xi}, \frac{(h^c + E_1^* + E_2^*)^{-\xi} - 1}{\xi}, \dots, \frac{(h^c + E_1^* + E_2^* + \dots + E_k^*)^{-\xi} - 1}{\xi} \right)$$

where E_1^*, \dots, E_k^* are i.i.d. standard exponentials. To show this, denote $F^c(\cdot)$ as the truncated CDF by $Q(1 - h^c/n)$, that is, $F^c(\cdot) = F(\cdot)/(1 - h^c/n)$. Define $U(t) = F^{-1}(1 - 1/t)$ and similarly for $U^c(t)$. Then

$$Y_{(1)}, \dots, Y_{(k)} \stackrel{d}{=} U^c\left(\frac{1}{1 - e^{-E_{1,n}}}\right), U^c\left(\frac{1}{1 - e^{-E_{2,n}}}\right), \dots, U^c\left(\frac{1}{1 - e^{-E_{k,n}}}\right)$$

where $E_{1,n}, \dots, E_{k,n}$ are order statistics of n i.i.d. standard exponentials. Next, note that $U^c\left(\frac{1}{1 - e^{-E_{1,n}}}\right) = U\left(\frac{1}{1 - (1 - h^c/n)e^{-E_{1,n}}}\right)$. Then the proof follows from the same argument of Theorem 2.1.1 of de Haan and Ferreira (2007) and the fact that $n(1 - (1 - h^c/n)\exp(-x/n)) \rightarrow x + h^c$. ■

Proof of Theorem 1. For notational ease, we prove the theorem without the autoregression part, i.e., assuming $\phi = \bar{\mu} = 0$. The proof with it follows the same logic with more tedious algebra,

and is available upon request. We start with the simplest GARCH(1,1) case, i.e., $\sigma_t^2 = \alpha_0 + \alpha_1 y_{t-1}^2 + \beta \sigma_{t-1}^2$. By iteration, we have $\sigma_t^2 = \sum_{l=0}^{t-1} \beta^l (\alpha_0 + \alpha_1 y_{t-1}^2)$ and plugging in the PML estimator, denoted as $(\hat{\alpha}_0, \hat{\alpha}_1, \hat{\beta})$, of the coefficients leads to an estimator of $\hat{\sigma}_t^2$, that is, $\hat{h}_t = \sum_{l=0}^{t-1} \hat{\beta}^l (\hat{\alpha}_0 + \hat{\alpha}_1 y_{t-1}^2)$.

Note that

$$\begin{aligned} \sup_{0 \leq w} \left| \frac{\hat{a} + w\hat{b}}{a + wb} - 1 \right| &\leq \frac{\max(|a - \hat{a}|, |b - \hat{b}|)}{\min(a, b)} \\ \sup_{0 \leq w} \left| \frac{\hat{a}\hat{c} + w\hat{b}}{ac + wb} - 1 \right| &\leq \frac{\max(|ac - \hat{a}\hat{c}|, |b - \hat{b}|)}{\min(ac, b)} \\ &\leq \frac{\max(a|c - \hat{c}|, c|a - \hat{a}|, |c - \hat{c}| \cdot |a - \hat{a}|, |b - \hat{b}|)}{\min(ac, b)} \end{aligned}$$

since

$$\begin{aligned} ac - \hat{a}\hat{c} &= a(c - \hat{c} + \hat{c}) - \hat{a}\hat{c} \\ &= a(c - \hat{c}) + \hat{c}(a - \hat{a}) \\ &= a(c - \hat{c}) + (\hat{c} - c + c)(a - \hat{a}) \\ &= a(c - \hat{c}) + c(a - \hat{a}) + (\hat{c} - c)(a - \hat{a}). \end{aligned}$$

Thus, by repeated applications of these inequalities, we have

$$\sup_{\frac{y_{t-1}^2}{\sigma_t^2}} \left| \frac{\hat{\sigma}_t^2}{\sigma_t^2} - 1 \right| \leq \frac{\max(\beta|\alpha_0 - \hat{\alpha}_0|, \beta|\alpha_1 - \hat{\alpha}_1|, \max(\alpha_1, \alpha_0) \sup_l |\beta^l - \hat{\beta}^l|)}{\alpha_0}$$

which converges to zero in probability by consistency of the PML estimator of the GARCH coefficients and $\alpha_0 > 0$.

Thus, $\sup_t |\hat{\sigma}_t^2/\sigma_t^2 - 1| \xrightarrow{p} 0$, and also $\sup_t |\hat{\sigma}_t/\sigma_t - 1| \xrightarrow{p} 0$. Let $Y_t = Z_t/\sigma_t$ and $\hat{Y}_t = Z_t/\hat{\sigma}_t$, so that $\hat{Y}_t = Y_t \sigma_t/\hat{\sigma}_t$. Then these results also imply $\sup_t |\hat{Y}_t/Y_t - 1| \xrightarrow{p} 0$. Now suppose Y_t satisfies (3), that is,

$$\begin{pmatrix} \frac{Y_{n:n} - b_n}{a_n} \\ \vdots \\ \frac{Y_{n:n-k+1} - b_n}{a_n} \end{pmatrix} \Rightarrow \mathbf{X}$$

where \mathbf{X} is jointly extreme value distributed as below (3). Let $I = (I_1, \dots, I_k) \in \{1, \dots, T\}^k$ be the k random indices such that $Y_{n:n-j+1} = Y_{I_j}$, $j = 1, \dots, k$, and let \hat{I} be the corresponding indices such that $\hat{Y}_{n:n-j+1} = \hat{Y}_{\hat{I}_j}$. We claim that $I - \hat{I} \xrightarrow{p} 0$. Suppose otherwise, then (3) implies that $\sup_t |\hat{Y}_t/Y_t - 1|$ is not $o_p(a_n)$. This contradicts $\sup_t |\hat{Y}_t/Y_t - 1| \xrightarrow{p} 0$ (since $a_n \rightarrow \infty$).

Thus,

$$\begin{aligned}
\begin{pmatrix} \frac{\hat{Y}_{I_1} - b_n}{a_n} \\ \vdots \\ \frac{\hat{Y}_{I_k} - b_n}{a_n} \end{pmatrix} &= \begin{pmatrix} \frac{\hat{Y}_{I_1} - b_n}{a_n} \\ \vdots \\ \frac{\hat{Y}_{I_k} - b_n}{a_n} \end{pmatrix} + o_p(1) \\
&= \text{diag}\left(\frac{\sigma_{I_1}}{\hat{\sigma}_{I_1}}, \dots, \frac{\sigma_{I_k}}{\hat{\sigma}_{I_k}}\right) \begin{pmatrix} \frac{Y_{I_1} \sigma_{I_1} / \hat{\sigma}_{I_1} - b_n}{a_n} \\ \vdots \\ \frac{Y_{I_k} \sigma_{I_k} / \hat{\sigma}_{I_k} - b_n}{a_n} \end{pmatrix} + o_p(1) \\
&\Rightarrow \mathbf{X}
\end{aligned}$$

by the Slutsky's theorem.

Now for GARCH(p, q) model, we have $\sigma_t^2 = \alpha_0 + \sum_{i=1}^q \alpha_i y_{t-i}^2 + \sum_{i=1}^p \beta_i \sigma_{t-i}^2$ with $\alpha_0 > 0$, $\alpha_i \geq 0$, $\beta_i \geq 0$ and $\sum_{i=1}^p \beta_i < 1$.

Let $B(x) = 1 - \beta_1 x - \beta_2 x^2 - \dots - \beta_p x^p$ and $A(x) = \alpha_1 x + \alpha_2 x^2 + \dots + \alpha_q x^q$, then we have $B(L)\sigma_t^2 = \alpha_0 + A(L)y_t^2$

$$\begin{aligned}
\sup_{y_{t-1}^2, \dots, y_{t-p}^2} \left| \frac{\hat{\sigma}_t^2}{\sigma_t^2} - 1 \right| &\leq \sup_{y_{t-1}^2, \dots, y_{t-p}^2} \frac{1}{a_0} \left| \hat{\alpha}_0 \hat{B}^{-1}(1) - \alpha_0 B^{-1}(1) + \left(\hat{A}(L) \hat{B}^{-1}(L) - A(L) B^{-1}(L) \right) y_t^2 \right| \\
&\leq \frac{1}{a_0} \left| \hat{\alpha}_0 \hat{B}^{-1}(1) - \alpha_0 B^{-1}(1) \right| + \sup_{y_{t-1}^2, \dots, y_{t-p}^2} \frac{1}{a_0} \left| \left(\hat{A}(L) \hat{B}^{-1}(L) - A(L) B^{-1}(L) \right) y_t^2 \right| \\
&\leq \left| \hat{B}^{-1}(1) - B^{-1}(1) \right| + \frac{\hat{B}^{-1}(1)}{\alpha_0} |\hat{\alpha}_0 - \alpha_0| \\
&\quad + \sup_{y_{t-1}^2, \dots, y_{t-p}^2} \frac{1}{a_0} \left| \left(\hat{A}(L) \hat{B}^{-1}(L) - A(L) B^{-1}(L) \right) y_t^2 \right| \\
&\leq o_p(1) + \frac{(\max_i \hat{\alpha}_i)}{a_0} \sup_{y_{t-1}^2, \dots, y_{t-p}^2} \left| \left(\hat{B}^{-1}(L) - B^{-1}(L) \right) y_t^2 \right| \\
&\quad + \frac{(\max B^{-1}(L))}{a_0} \sup_{y_{t-1}^2, \dots, y_{t-p}^2} \left| \left(\hat{A}(L) - A(L) \right) y_t^2 \right| \\
&= o_p(1)
\end{aligned}$$

where $B^{-1}(L) = \frac{1}{B(L)} = \sum_{j=1}^{\infty} b_j L^j$ with coefficients b_j decaying exponentially fast and $\max B^{-1}(L)$ denotes the maximum of $\{b_1, b_2, \dots\}$. In the last inequality, we implicitly use the fact that the consistency of \hat{B} implies the consistency of \hat{B}^{-1} . Then the rest of proof is the same as in the GARCH(1,1) case. ■

References

- ABAN, I. B., M. M. MEERSCHAERT, AND A. K. PANORSKA (2006): “Parameter Estimation for the Truncated Pareto Distribution,” *Journal of the American Statistical Association*, 101(473), 270–277.
- ANDERSON, P., AND M. M. MEERSCHAERT (1998): “Modeling River Flows With Heavy Tails,” *Water Resources Research*, 34, 2271–2280.
- ARNOLD, B. C., N. BALAKRISHNAN, AND H. H. N. NAGARAJA (1992): *A First Course in Order Statistics*. Siam.
- BAEK, S. K., S. BERNHARDSSON, AND P. MINNHAGEN (2011): “Zipf’s Law Unzipped,” *New Journal of Physics*, 13.
- BARKAI, E., R. METZLER, AND J. KLAFTER (2000): “From Continuous Time Random Walks to the Fractional Fokker-Planck Equation,” *Physics Review*, 61, 132–138.
- BASEL COMMITTEE ON BANKING SUPERVISION (2013): “Fundamental Review of the Trading Book: A Revised Market Risk Framework,” *Consultative Document*.
- BEIRLANT, J., F. CAEIRO, AND M. I. GOMES (2012): “An Overview and Open Research Topics in Statistics of Univariate Extremes,” *Revstat*, 10(1), 1–31.
- BENSON, D., S. WHEATCRAFT, AND M. M. MEERSCHAERT (2000): “Application of a Fractional Advection-Dispersion Equation,” *Water Resources Research*, 36, 1403–1412.
- DE HAAN, L., AND A. FERREIRA (2007): *Extreme Value Theory: An Introduction*. Springer Science & Business Media, New York.
- DE HAAN, L., AND H. ROOTZÉN (1993): “On the estimation of high quantiles,” *Journal of Statistical Planning and Inference*, 35, 1–13.
- DEKKERS, A. L. M., AND L. DE HAAN (1989): “On the estimation of the extreme-value index and large quantile estimation,” *Annals of Statistics*, 17, 1795–1832.
- DOWNTON, M., AND R. A. PIELKE (2005): “How accurate are disaster loss data? The case of U.S. flood damage,” *Natural Hazards*, 35(2), 211–228.
- EMBRECHTS, P., C. KLUPPERBERG, AND T. MIKOSCH (1997): *Modelling extremal events for insurance and finance*. Springer, New York.
- ENGLE, R. F., AND S. MANGANELLI (2004): “CAViAR: Conditional Autoregressive Value at Risk by Regression Quantiles,” *Journal of Business & Economic Statistics*, 22(4),

367–381.

- FISSLER, T., AND J. F. ZIEGEL (2016): “Higher order elicibility and Osband’s principle,” *Annals of Statistics*.
- GOMES, M. I., AND A. GUILLOU (2015): “Extreme Value Theory and Statistics of Univariate Extreme: a Review,” *International Economic Review*, 83, 263–292.
- GROISMAN, P. Y., R. W. KNIGHT, T. R. KARL, D. R. EASTERLING, B. SUN, AND J. H. LAWRIE (2004): “Contemporary Changes of the Hydrological Cycle Over the Contiguous United States: Trends Derived From in Situ Observations,” *Journal of Hydrometeorology*, 5, 64–85.
- HILL, B. M. (1975): “A Simple General Approach to Inference about the Tail of a Distribution,” *Annals of Statistics*, 3(5), 1163–1174.
- JENKINS, S. P., R. V. BURKHAUSER, S. FENG, AND J. LARRIMORE (2010): “Measuring Inequality Using Censored Data: A Multiple-imputation Approach to Estimation and Inference,” *Journal of the Royal Statistical Society Series A*, 174(1), 63–81.
- JORION, P. (2007): *Value at Risk - The New Benchmark for Managing Financial Risk*. McGraw-Hill.
- KUESTER, K., S. MITTNIK, AND M. S. PAOLELLA (2006): “Value-at-Risk Prediction: A Comparison of Alternative Strategies,” *Journal of Financial Econometrics*, 4(1), 53–89.
- LEHMANN, E. L., AND J. P. ROMANO (2005): *Testing Statistical Hypothesis*. Springer, New York.
- MALAMUD, B. D., G. MOREIN, AND D. L. TURCOTTE (1998): “Forest Fire: An Example of Self-Organized Critical Behavior,” *Science*, 281, 1840–1842.
- MCNEIL, A., 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.
- MEERSCHAERT, M. M., D. A. BENSON, H. P. SCHEFFLER, AND P. BECKER-KERN (2002): “Governing Equations and Solutions of Anomalous Random Walk Limits,” *Physics Review*, 66, 102–105.
- MÜLLER, U. K., AND Y. WANG (2015): “Nearly Weighted Risk Minimal Unbiased Estimation,” *Working Paper*.

- (2016): “Fixed-k Asymptotic Inference about Tail Properties,” *Accepted at the Journal of the American Statistical Association*.
- PACHECO, J., C. H. SCHOLTZ, AND L. R. SYKES (1992): “Changes in Frequency-Size Relationship from Small to Large Earthquakes,” *Nature*, 355, 71–73.
- PATTON, A. J., J. ZIEGEL, AND R. CHEN (2017): “Dynamic Semiparametric Models for Expected Shortfall (and Value-at-Risk),” working paper.
- PICKANDS, III, J. (1975): “Statistical inference using extreme order statistics,” *Annals of Statistics*, 3(1), 119–131.
- PIELKE, R. A., J. GRATZ, C. W. LANDSEA, D. COLLINS, M. A. SAUNDERS, AND R. MUSULIN (2008): “Normalized Hurricane Damage in the United States: 1900–2005,” *Natural Hazards Review*, 9(1), 29–42.
- REISS, R. D., AND M. THOMAS (2007): *Statistical Analysis of Extreme Values*. Birkhäuser Verlag, Basel.
- RESNICK, S. (2007): *Heavy-tail phenomena: probabilistic and statistical modeling*. Springer Science & Business Media.
- WEISSMAN, I. (1978): “Estimation of parameters and large quantiles based on the k largest observations,” *Journal of the American Statistical Association*, 73, 812–815.
- ZIPF, G. K. (1936): *The Psycho-Biology of Language*. Routledge, London.