Bias-Corrected Peaks-Over-Threshold Estimation of the CVaR (Supplementary Material)

Dylan Troop[1] **Frédéric Godin**[2] **Jia Yuan Yu**[3]

[1]Department of Computer Science and Software Engineering, Concordia University, Montreal, Canada
[2]Department of Mathematics and Statistics, Concordia University, Montreal, Canada
[3]Concordia Institute for Information Systems Engineering, Concordia University, Montreal, Canada

A PROOFS

We first recall the stochastic order notation (e.g., Vaart [1998, Section 2.2]), which will be used throughout subsequent proofs.

Definition A.1 (Stochastic o and O symbols)**.** *Let X_n, R_n denote sequences of random variables. Then,*

$$X_n = o_p(R_n) \quad \text{means} \quad \forall \varepsilon > 0,\ \lim_{n\to\infty} \mathbb{P}(|X_n/R_n| > \varepsilon) = 0,$$
$$X_n = O_p(R_n) \quad \text{means} \quad \forall \varepsilon > 0,\ \exists M, N > 0,\ \forall n > N,\ \mathbb{P}(|X_n/R_n| > M) < \varepsilon.$$

The often used notation $X_n = o_p(1)$ means that X_n converges to zero in probability, and $X_n = O_p(1)$ means that X_n is bounded in probability.

A.1 PROOF OF THEOREM 3.1

We first state the following lemma, which is equivalent to Beirlant et al. [2003, Proposition 1] with different notation.

Lemma A.1. *Suppose assumption 2.1 and assumption 2.2 hold. Then $\forall \varepsilon > 0,\ \exists t_0,\ \forall t, x$ such that $t \geq t_0$ and $tx \geq t_0$,*

$$(1-\varepsilon)e^{-\varepsilon|\log x|} \leq \left[\frac{U(tx)-U(t)}{a(t)} - \frac{x^\xi - 1}{\xi}\right] / \left[A\left(t\right) I_{\xi,\rho}(x)\right] \leq (1+\varepsilon)e^{\varepsilon|\log x|}, \tag{1}$$

where

$$A(t) = \frac{tU''(t)}{U'(t)} - \xi + 1 \quad \text{and} \quad I_{\xi,\rho}(x) = \begin{cases} \frac{1}{\rho}\left(\frac{x^{\xi+\rho}-1}{\xi+\rho} - \frac{x^\xi - 1}{\xi}\right), & \rho < 0,\ \xi+\rho \neq 0, \\ \frac{1}{\rho}\left(\log x - \frac{x^\xi-1}{\xi}\right), & \rho < 0,\ \xi+\rho = 0, \\ \frac{1}{\xi}\left(x^\xi \log x - \frac{x^\xi - 1}{\xi}\right), & \rho = 0. \end{cases}$$

Proof. In Beirlant et al. [2003], the statement is given as $\forall \varepsilon > 0,\ \exists t_0,\ \forall t, x$ such that $t \geq t_0$ and $t + x \geq t_0$,

$$(1-\varepsilon)e^{-\varepsilon|x|} \leq \left[\frac{V(t+x)-V(t)}{V'(t)} - \frac{e^{\xi x}-1}{\xi}\right] / \left[\tilde{A}\left(e^t\right)\tilde{I}_{\xi,\rho}(x)\right] \leq (1+\varepsilon)e^{\varepsilon|x|}, \tag{2}$$

where

$$V(t) = (\bar{F})^{-1}\left(e^{-t}\right), \qquad \tilde{A}(t) = \frac{V''(\log t)}{V'(\log t)} - \xi, \qquad \tilde{I}_{\xi,\rho}(x) = I_{\xi,\rho}(e^x).$$

Then, for $t \geq 1$,

$$V(\log t) = (\bar{F})^{-1}(1/t) = (1/\bar{F})^{-1}(t) = U(t),$$

and

$$V'(\log t) = tU'(t) = a(t), \quad V''(\log t) = t^2U''(t) + tU'(t) \quad \Rightarrow \quad \tilde{A}(t) = A(t).$$

Since $\log$ is strictly increasing, eq. (2) holds with $\log t$ and $\log x$ where $\log t \geq t_0$ and $\log tx \geq t_0$. Substituting expressions in eq. (2), we get eq. (1). □

The following corollary will also be used in the main proof of this section.

Corollary A.1. *An immediate consequence of lemma A.1 is for all $x > 0$,*

$$\lim_{t\to\infty} \frac{\frac{U(tx)-U(t)}{a(t)} - \frac{x^\xi - 1}{\xi}}{A(t)} = I_{\xi,\rho}(x). \tag{3}$$

corollary A.1 can also be found in De Haan and Ferreira [2006, Theorem 2.3.12]. Before proving our main result of this section, we first recall the dominated convergence theorem which will be needed later.

Theorem A.1 (Dominated convergence theorem)**.** *Let $\{f_n\}_{n=1}^\infty$ be a sequence of real-valued functions defined on $S \subset \mathbb{R}$ such that $\forall x \in S$, $\lim_{n\to\infty} f_n(x) \to f(x)$. If $\forall x \in S$, n,*

$$|f_n(x)| \leq g(x)$$

for some integrable (i.e., the integral is finite over S) function g, then

$$\lim_{n\to\infty} \int_S f_n(x)dx = \int_S \lim_{n\to\infty} f_n(x)dx = \int_S f(x)dx.$$

Proof of Theorem 3.1. We use corollary A.1 to derive a convergence result for the approximation error of the VaR, i.e, $q_\alpha - q_{u,\alpha}$. Then, using lemma A.1 and theorem A.1, we will be able to derive the convergence of $\epsilon_{u,\alpha}$.

For any $p \in (0, 1)$ and $y \in \operatorname{dom} F$ such that $F(y) = p$,

$$\left(\frac{1}{\bar{F}}\right)(y) = \frac{1}{\bar{F}(y)} = \frac{1}{1-p},$$

which implies that

$$U\left(\frac{1}{1-p}\right) = \left(\frac{1}{\bar{F}}\right)^{-1}(1/(1-p)) = y.$$

Hence,

$$U(1/(1-\alpha)) = U(\tau_u\beta) = q_\alpha \quad \text{and} \quad U(\tau_u) = u.$$

Then, from the definition of $q_{u,\alpha}$ we get

$$q_{u,\alpha} = u + \frac{\sigma(u)}{\xi}\left(\beta^\xi - 1\right) = U(\tau_u) + \frac{a(\tau_u)}{\xi}(\beta^\xi - 1).$$

Setting $D_u(\beta) = (q_\alpha - q_{u,\alpha})/(a(\tau_u)A(\tau_u))$, it then follows from the previous two equations and corollary A.1 with $t = \tau_u$, $x = \beta$ that

$$D_u(\beta) = \frac{U(\tau_u\beta) - U(\tau_u) - \frac{a(\tau_u)}{\xi}(\beta^\xi - 1)}{a(\tau_u)A(\tau_u)} = \frac{\frac{U(\tau_u\beta)-U(\tau_u)}{a(\tau_u)} - \frac{\beta^\xi-1}{\xi}}{A(\tau_u)} \to I_{\xi,\rho}(\beta) \quad \text{as } u \to \infty. \tag{4}$$

From the definition of the GPD approximation error and the CVaR, for a fixed u, α,

$$\frac{\epsilon_{u,\alpha}}{a(\tau_u)A(\tau_u)} = \frac{c_{u,\alpha} - c_\alpha}{a(\tau_u)A(\tau_u)} = -\frac{1}{1-\alpha}\int_\alpha^1 \frac{q_\gamma - q_{u,\gamma}}{a(\tau_u)A(\tau_u)}d\gamma = -\beta\int_\beta^\infty \frac{D_u(x)}{x^2}dx, \tag{5}$$

where $D_u(x) = (q_\gamma - q_{u,\gamma})/(a(\tau_u)A(\tau_u))$ and we have used the substitution $x = \bar{F}(u)/(1-\gamma)$. We now apply the dominated convergence theorem to get the limiting behaviour of eq. (5) as $u \to \infty$. From lemma A.1, $\forall \varepsilon > 0$, $\exists u_0$ such that $\forall u \geq u_0,\ x \in [\beta, \infty)$,

$$\left|\frac{D_u(x)}{x^2}\right| \leq (1+\varepsilon)x^{\varepsilon-2} I_{\xi,\rho}(x).$$

$(1+\varepsilon)x^{\varepsilon-2} I_{\xi,\rho}(x)$ is integrable over $[\beta, \infty)$ as long as $\varepsilon < 1-\xi$. Since $\xi < 1$, let $\varepsilon = (1-\xi)/2$. Then theorem A.1 can be applied to $D_u(x)/x^2$. Setting $K_{\xi,\rho}(\beta) = -\beta \int_\beta^\infty [I_{\xi,\rho}(x)/x^2]dx$, it follows that

$$\begin{aligned}
\lim_{u\to\infty} \frac{\epsilon_{u,\alpha}}{a(\tau_u)A(\tau_u)} &= \lim_{u\to\infty} -\beta \int_\beta^\infty \frac{D_u(x)}{x^2}dx \\
&= -\beta \int_\beta^\infty \lim_{u\to\infty} \frac{D_u(x)}{x^2}dx \\
&= -\beta \int_\beta^\infty \frac{I_{\xi,\rho}(x)}{x^2}dx \\
&= K_{\xi,\rho}(\beta),
\end{aligned}$$

where the last integral can be computed explicitly to obtain eq. (8). □

A.2 PROOF OF THEOREM 4.2

First recall Slutsky's lemma (see, for example, Vaart [1998, Lemma 2.8]).

Lemma A.2 (Slutsky). *Let X_n, X, Y_n be random vectors or variables. If $X_n \xrightarrow{d} X$ and $Y_n \xrightarrow{p} c$ for a constant c, then*

(i) $X_n + Y_n \xrightarrow{d} X + c$;

(ii) $X_n Y_n \xrightarrow{d} Xc$;

(iii) $X_n/Y_n \xrightarrow{d} X/c$ *provided* $c \neq 0$.

Proof of Theorem 4.2. First note that since $\hat{A}_n$ and $\hat{b}_n$ are consistent, i.e.,

$$\frac{\hat{A}_n}{A(n/k)} \xrightarrow{p} 1, \qquad \hat{b}_n \xrightarrow{p} b_{\xi,\rho},$$

then by lemma A.2, the fact that $\hat{\sigma}^{(n)}_{\text{MLE}}/a(n/k) \xrightarrow{p} 1$ (which follows from theorem 4.1), and the assumption of theorem 4.1 that $\lim_{n\to\infty} \sqrt{k}A(n/k) \to \lambda < \infty$,

$$\begin{aligned}
\sqrt{k}\hat{A}_n \left(\hat{b}^{(1)}_n,\ \frac{\hat{\sigma}^{(n)}_{\text{MLE}}}{a(n/k)}\hat{b}^{(2)}_n\right) &= \sqrt{k}A(n/k)\frac{\hat{A}_n}{A(n/k)} \left(\hat{b}^{(1)}_n,\ \frac{\hat{\sigma}^{(n)}_{\text{MLE}}}{a(n/k)}\hat{b}^{(2)}_n\right) \\
&\xrightarrow{p} \lambda \left(b^{(1)}_{\xi,\rho},\ b^{(2)}_{\xi,\rho}\right) \\
&= \lambda b_{\xi,\rho}.
\end{aligned}$$

Then, by expanding terms and applying lemma A.2 once again,

$$\begin{aligned}
\sqrt{k}(\hat{\xi}_n - \xi,\ \hat{\sigma}_n/a(n/k) - 1) &= \sqrt{k}(\hat{\xi}^{(n)}_{\text{MLE}} - \hat{A}_n\hat{b}^{(1)}_n - \xi,\ \hat{\sigma}^{(n)}_{\text{MLE}}(1 - \hat{A}_n\hat{b}^{(2)}_n)/a(n/k) - 1) \\
&= \sqrt{k}(\hat{\xi}^{(n)}_{\text{MLE}} - \xi,\ \hat{\sigma}^{(n)}_{\text{MLE}}/a(n/k) - 1) - \sqrt{k}\hat{A}_n \left(\hat{b}^{(1)}_n,\ \frac{\hat{\sigma}^{(n)}_{\text{MLE}}}{a(n/k)}\hat{b}^{(2)}_n\right) \\
&\xrightarrow{d} N(\lambda b_{\xi,\rho}, \Sigma) - \lambda b_{\xi,\rho} \\
&= N(0, \Sigma).
\end{aligned}$$

□

A.3 PROOF OF THEOREM 4.3

We first give the delta method, which can be found in, for example, Rémillard [2016, Appendix B.3.4.1].

Theorem A.2 (Delta method)**.** *Let $\hat{\theta}_n \in \mathbb{R}^m$ be a random vector based on a sample of size n. Suppose that $h : \mathbb{R}^m \mapsto \mathbb{R}$ is such that for $i = 1, \ldots, m$, $\frac{\partial h}{\partial \theta_i}$ exists and is continuous in a neighborhood of θ. If $\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{d} N(0, V)$, then*

$$\sqrt{n}(h(\hat{\theta}_n) - h(\theta)) \xrightarrow{d} N(0, \nabla h(\theta)^\top V \nabla h(\theta)),$$

where $\nabla h(\theta)$ is the gradient of h evaluated at θ.

Next, we prove some useful lemmas which will be used in the proof of theorem 5.1.

Lemma A.3. *Let $X_1, \ldots, X_n$ be an i.i.d. sample with common cdf F, and suppose $k = k_n \to \infty$ and $k/n \to 0$ as $n \to \infty$. With $u = X_{(n-k,n)}$ and $\xi \in \mathbb{R}$,*

$$\sqrt{k}\left((k\tau_u/n)^\xi - 1\right) \xrightarrow{d} N(0, \xi^2).$$

Proof. Letting $h_\xi(x) = x^{-\xi}$,

$$\sqrt{k}\left((k\tau_u/n)^\xi - 1\right) = \sqrt{k}\left((n\bar{F}(u)/k)^{-\xi} - 1\right) = \sqrt{k}\left(h_\xi(n\bar{F}(u)/k) - h_\xi(1)\right).$$

From Beirlant et al. [2009, Theorem 3.1], we know that $\sqrt{k}(n\bar{F}(u)/k - 1) \xrightarrow{d} N(0, 1)$. Hence, by theorem A.2,

$$\sqrt{k}\left((k\tau_u/n)^\xi - 1\right) \xrightarrow{d} N(0, h'_\xi(1) \cdot 1 \cdot h'_\xi(1)) = N(0, \xi^2).$$

□

Corollary A.2. *Let $\alpha = \alpha_n = 1 - (1/\beta)k/n$ where $\beta > 1$ is a constant not depending on n. Then,*

$$\sqrt{k}(s_{u,\alpha}^\xi - \beta^\xi) \xrightarrow{d} N(0, \xi^2 \beta^{2\xi}).$$

Proof.

$$s_{u,\alpha} = \bar{F}(u)\left(\frac{n}{k}\right)\frac{k}{n(1-\alpha)} = \beta n \bar{F}(u)/k = \frac{\beta n}{k\tau_u},$$

and so

$$\sqrt{k}(s_{u,\alpha}^\xi - \beta^\xi) = \beta^\xi \sqrt{k}((k\tau_u/n)^{-\xi} - 1) = -\beta^\xi (n\bar{F}(u)/k)^\xi \sqrt{k}((k\tau_u/n)^\xi - 1). \tag{6}$$

Beirlant et al. [2009, Theorem 3.1] implies that $n\bar{F}(u)/k \xrightarrow{p} 1$. Hence, lemma A.3 with eq. (6) implies

$$\sqrt{k}(s_{u,\alpha}^\xi - \beta^\xi) \xrightarrow{d} N(0, \xi^2 \beta^{2\xi}).$$

□

Lemma A.4. *Suppose that the assumptions of theorem 4.1 hold. Then as $n \to \infty$,*

$$\sqrt{k}\left(\frac{a(\tau_u)}{a(n/k)} - (k\tau_u/n)^\xi\right) = o_p(1). \tag{7}$$

Proof. Under assumption 2.1 and assumption 2.2, the following uniform inequality from De Haan and Ferreira [2006, Theorem 2.3.6] holds: for any $\varepsilon, \delta > 0$ there exists $t_0 = t_0(\varepsilon, \delta)$ such that for all $t, tx \geq t_0$,

$$\left|\frac{\frac{a(tx)}{a(t)} - x^\xi}{A(t)} - x^\xi \frac{x^\rho - 1}{\rho}\right| \leq \varepsilon x^{\xi+\rho} \max\left(x^\delta, x^{-\delta}\right). \tag{8}$$

Hence, with $t = n/k$ and $x = k\tau_u/n$, for any $\varepsilon, \delta > 0$ and with large enough n,

$$\begin{aligned}
\sqrt{k}\left(\frac{a(\tau_u)}{a(n/k)} - (k\tau_u/n)^\xi\right) &= \sqrt{k}A(n/k)\left[\frac{\frac{a(\tau_u)}{a(n/k)} - (k\tau_u/n)^\xi}{A(n/k)} - (k\tau_u/n)^\xi \frac{(k\tau_u/n)^\rho - 1}{\rho}\right] \\
&\quad + \sqrt{k}A(n/k)(k\tau_u/n)^\xi \frac{(k\tau_u/n)^\rho - 1}{\rho} \\
&\leq \sqrt{k}A(n/k)\left|\frac{\frac{a(\tau_u)}{a(n/k)} - (k\tau_u/n)^\xi}{A(n/k)} - (k\tau_u/n)^\xi \frac{(k\tau_u/n)^\rho - 1}{\rho}\right| \\
&\quad + \sqrt{k}A(n/k)(k\tau_u/n)^\xi \frac{(k\tau_u/n)^\rho - 1}{\rho} \\
&\leq \sqrt{k}A(n/k)\varepsilon(k\tau_u/n)^{\xi+\rho} \max\left((k\tau_u/n)^\delta, (k\tau_u/n)^{-\delta}\right) \\
&\quad + \sqrt{k}A(n/k)(k\tau_u/n)^\xi \frac{(k\tau_u/n)^\rho - 1}{\rho}.
\end{aligned}$$

Since $k\tau_u/n \xrightarrow{p} 1$ and $\sqrt{k}A(n/k) \to \lambda < \infty$ as $n \to \infty$ (by the assumption of theorem 4.1), and since ε can be made arbitrarily small as $n \to \infty$, both terms tend to 0 in probability as $n \to \infty$, hence eq. (7) follows. □

The following corollary is an immediate result by combining lemma A.3 and lemma A.4.

Corollary A.3. *Suppose that the assumptions of theorem 4.1 hold. Then,*

$$\sqrt{k}\left(\frac{a(\tau_u)}{a(n/k)} - 1\right) \xrightarrow{d} N(0, \xi^2).$$

Proof of Theorem 4.3. With $\alpha = 1 - (1/\beta)k/n$ and $u = X_{(n-k,n)}$,

$$\frac{\hat{c}_\alpha^{(n)}}{a(n/k)} = \frac{\hat{\sigma}_n/a(n/k)}{1-\hat{\xi}_n}\left(1 + \frac{\beta^{\hat{\xi}_n} - 1}{\hat{\xi}_n}\right) + \frac{u}{a(n/k)} = d_\beta(\hat{\xi}_n, \hat{\sigma}_n/a(n/k)) + \frac{u}{a(n/k)},$$

and recalling that $s_{u,\alpha} = \bar{F}(u)/(1-\alpha)$,

$$\begin{aligned}
\frac{c_{u,\alpha}}{a(n/k)} &= \frac{\sigma(u)/a(n/k)}{1-\xi}\left(1 + \frac{s_{u,\alpha}^\xi - 1}{\xi}\right) + \frac{u}{a(n/k)} \\
&= \frac{1}{1-\xi}\left(1 + \frac{s_{u,\alpha}^\xi - 1}{\xi}\right) + \frac{\sigma(u)/a(n/k) - 1}{1-\xi}\left(1 + \frac{s_{u,\alpha}^\xi - 1}{\xi}\right) + \frac{u}{a(n/k)}.
\end{aligned}$$

Then for the first term,

$$\begin{aligned}
\frac{1}{1-\xi}\left(1 + \frac{s_{u,\alpha}^\xi - 1}{\xi}\right) &= \frac{1}{1-\xi}\left(1 + \frac{s_{u,\alpha}^\xi - 1 + \beta^\xi - \beta^\xi}{\xi}\right) \\
&= \frac{1}{1-\xi}\left(1 + \frac{\beta^\xi - 1}{\xi} + \frac{s_{u,\alpha}^\xi - \beta^\xi}{\xi}\right) \\
&= d_\beta(\xi, 1) + \frac{s_{u,\alpha}^\xi - \beta^\xi}{\xi(1-\xi)}.
\end{aligned}$$

Recalling that $\sigma(u) = a(\tau_u)$ and combining the previous expressions,

$$\begin{aligned}
\frac{\sqrt{k}}{a(n/k)}\left(\hat{c}_\alpha^{(n)} - c_{u,\alpha}\right) &= \sqrt{k}\left(d_\beta(\hat{\xi}_n, \hat{\sigma}_n/a(n/k)) - d_\beta(\xi, 1)\right) \\
&\quad - \sqrt{k}\left(\frac{a(\tau_u)/a(n/k) - 1}{1-\xi}\left(1 + \frac{s_{u,\alpha}^\xi - 1}{\xi}\right) + \frac{s_{u,\alpha}^\xi - \beta^\xi}{\xi(1-\xi)}\right) = S - R, \quad (9)
\end{aligned}$$

where we have denoted each term by S and R, respectively. By the delta method (theorem A.2),

$$S \xrightarrow{d} N(0, \nabla d_\beta(\xi, 1)^\top \Sigma \nabla d_\beta(\xi, 1)).$$

For the second term, let

$$P = \sqrt{k}\left(\frac{a(\tau_u)}{a(n/k)} - (k\tau_u/n)^\xi\right) \quad \text{and} \quad Q = \sqrt{k}\left((k\tau_u/n)^\xi - 1\right).$$

Then $\sqrt{k}(a(\tau_u)/a(n/k) - 1) = P + Q$ and from eq. (6), $\sqrt{k}(s_{u,\alpha}^\xi - \beta^\xi) = -s_{u,\alpha}^\xi Q$. Hence,

$$R = \frac{1}{\xi(1-\xi)}\left[(P+Q)(\xi + s_{u,\alpha}^\xi - 1) - s_{u,\alpha}^\xi Q\right] = \frac{(\xi + s_{u,\alpha}^\xi - 1)}{\xi(1-\xi)}P - \frac{Q}{\xi}.$$

corollary A.2 implies that $s_{u,\alpha}^\xi = \beta^\xi + o_p(1)$ as $n \to \infty$ and from lemma A.4 we know that $P = o_p(1)$ as $n \to \infty$. By lemma A.3, $Q \xrightarrow{d} N(0, \xi^2)$, and so $R \xrightarrow{d} N(0, 1)$. Compared to S, the asymptotic variance of R is approximately 0, and we ignore its contribution in further calculations. Then, eq. (9) implies

$$\frac{\sqrt{k}}{a(n/k)}\left(\hat{c}_\alpha^{(n)} - c_{u,\alpha}\right) \xrightarrow{d} N\left(0, \nabla d_\beta(\xi, 1)^\top \Sigma \nabla d_\beta(\xi, 1)\right) = N(0, V),$$

where

$$\nabla d_\beta(\xi, 1) = \left[\frac{\partial d_\beta}{\partial x}(\xi, 1), \frac{\partial d_\beta}{\partial y}(\xi, 1)\right]^\top,$$

with

$$\frac{\partial d_\beta}{\partial x}(\xi, 1) = \frac{\beta^\xi(2\xi + \xi(1-\xi)\log\beta - 1)}{\xi^2(1-\xi)^2} + \frac{1}{\xi^2}, \quad \frac{\partial d_\beta}{\partial y}(\xi, 1) = \frac{\beta^\xi + \xi - 1}{\xi(1-\xi)}.$$

□

A.4 PROOF OF THEOREM 5.1

We first prove the following lemma, which shows that when τ_u is replaced by n/k in theorem 3.1, the asymptotic behaviour of $\epsilon_{u,\alpha}$ is the same (in probability).

Lemma A.5. *Suppose that the assumptions of theorem 4.1 hold. Let $\alpha = 1 - (1/\beta)k/n$ where $\beta > 1$ is a constant not depending on n. Then,*

$$\frac{\epsilon_{u,\alpha}}{a(n/k)A(n/k)K_{\xi,\rho}(\beta)} \xrightarrow{p} 1.$$

Proof. We follow the same line of reasoning as in the proof of theorem 3.1. First, we have

$$\begin{aligned}
q_{u,\alpha} &= u + \frac{\sigma(u)}{\xi}(s_{u,\alpha}^\xi - 1) \\
&= U(\tau_u) + \frac{a(\tau_u)}{\xi}(s_{u,\alpha}^\xi - 1) \\
&= U(n/k) + \frac{a(n/k)}{\xi}(\beta^\xi - 1) + U(\tau_u) - U(n/k) + \frac{1}{\xi}\left[a(\tau_u)(s_{u,\alpha}^\xi - 1) - a(n/k)(\beta^\xi - 1)\right] \\
&= U(n/k) + \frac{a(n/k)}{\xi}(\beta^\xi - 1) + U(\tau_u) - U(n/k) \\
&\qquad + \frac{1}{\xi}\left[a(\tau_u)(s_{u,\alpha}^\xi - \beta^\xi) + (a(\tau_u) - a(n/k))(\beta^\xi - 1)\right].
\end{aligned}$$

Given that $q_\alpha = U(1/(1-\alpha)) = U(n\beta/k)$,

$$\frac{q_\alpha - q_{u,\alpha}}{a(n/k)A(n/k)} = \frac{\frac{U(n\beta/k)-U(n/k)}{a(n/k)} - \frac{\beta^\xi - 1}{\xi}}{A(n/k)}$$
$$- \frac{U(\tau_u) - U(n/k)}{a(n/k)A(n/k)} - \frac{a(\tau_u)(s_{u,\alpha}^\xi - \beta^\xi)}{\xi a(n/k)A(n/k)} - \left(\frac{a(\tau_u)}{a(n/k)} - 1\right)\frac{\beta^\xi - 1}{\xi A(n/k)}$$
$$= I - II - III - IV.$$

In what follows, terms I-IV will be analyzed separately then finally combined.

I: By corollary A.1, with $t = n/k$ and $x = \beta$ we know that term I tends to $I_{\xi,\rho}(\beta)$ as $n \to \infty$.

II: Under assumption 2.1 and assumption 2.2, the following uniform inequality from De Haan and Ferreira [2006, Theorem 2.3.6] holds: for any $\varepsilon, \delta > 0$ there exists $t_0 = t_0(\varepsilon, \delta)$ such that for all $t, tx \geq t_0$,

$$\left| \frac{\frac{U(tx)-U(t)}{a(t)} - \frac{x^\xi - 1}{\xi}}{A(t)} - \frac{x^{\xi+\rho} - 1}{\xi + \rho} \right| \leq \varepsilon x^{\xi+\rho} \max\left(x^\delta, x^{-\delta}\right). \tag{10}$$

We can write II as

$$II = \frac{\frac{U(\tau_u)-U(n/k)}{a(n/k)} - \frac{(k\tau_u/n)^\xi - 1}{\xi}}{A(n/k)} + \frac{(k\tau_u/n)^\xi - 1}{\xi A(n/k)}.$$

Hence, with $t = n/k$ and $x = k\tau_u/n$, the first term tends to 0 in probability using eq. (10) and essentially the same arguments as in the proof of lemma A.4. So, by the assumption of theorem 4.1 that $\sqrt{k}A(n/k) \to \lambda < \infty$ as $n \to \infty$ and lemma A.3,

$$II = \frac{Q}{\xi\sqrt{k}A(n/k)} + o_p(1) = \frac{Z}{\lambda} + o_p(1), \quad n \to \infty,$$

where $Q = \sqrt{k}((k\tau_u/n)^\xi - 1)$ and Z denotes a standard normal random variable.

III: corollary A.3 implies that $a(\tau_u)/a(n/k) \xrightarrow{p} 1$ and by eq. (6), $s_{u,\alpha}^\xi - \beta^\xi = -s_{u,\alpha}^\xi Q/\sqrt{k}$. Corollary A.2 implies that $s_{u,\alpha}^\xi \xrightarrow{p} \beta^\xi$, and so

$$III = \frac{a(\tau_u)(s_{u,\alpha}^\xi - \beta^\xi)}{\xi a(n/k)A(n/k)} = \frac{-s_{u,\alpha}^\xi Q}{\xi\sqrt{k}A(n/k)}(1 + o_p(1)) = -\frac{\beta^\xi}{\lambda}Z + o_p(1), \quad n \to \infty.$$

IV: With $P = \sqrt{k}\left(\frac{a(\tau_u)}{a(n/k)} - (k\tau_u/n)^\xi\right)$ and applying lemma A.4,

$$IV = \left(\frac{a(\tau_u)}{a(n/k)} - 1\right)\frac{\beta^\xi - 1}{\xi A(n/k)} = (P + Q)\frac{\beta^\xi - 1}{\xi\sqrt{k}A(n/k)} = \frac{\beta^\xi - 1}{\lambda}Z + o_p(1), \quad n \to \infty.$$

Now combining all terms, as $n \to \infty$,

$$\frac{q_\alpha - q_{u,\alpha}}{a(n/k)A(n/k)} = I - II - III - IV$$
$$= I_{\xi,\rho}(\beta) - \frac{Z}{\lambda} + \frac{\beta^\xi}{\lambda}Z - \frac{\beta^\xi - 1}{\lambda}Z + o_p(1) = I_{\xi,\rho}(\beta) + o_p(1), \quad n \to \infty.$$

Hence, following the same reasoning as in the proof of theorem 3.1,

$$\frac{\epsilon_{u,\alpha}}{a(n/k)A(n/k)} = \frac{c_{u,\alpha} - c_\alpha}{a(n/k)A(n/k)}$$
$$= -\frac{1}{1-\alpha}\int_\alpha^1 \frac{q_\gamma - q_{u,\gamma}}{a(n/k)A(n/k)}d\gamma \xrightarrow{p} -\beta\int_\beta^\infty \frac{I_{\xi,\rho}(x)}{x^2}dx = K_{\xi,\rho}(\beta).$$

□

Proof of Theorem 5.1. First,

$$\hat{c}_{\epsilon,\alpha}^{(n)} - c_\alpha = \hat{c}_\alpha^{(n)} - \hat{\epsilon}_\alpha^{(n)} - c_\alpha = \hat{c}_\alpha^{(n)} - c_{u,\alpha} - \hat{\epsilon}_\alpha^{(n)} + c_{u,\alpha} - c_\alpha = \hat{c}_\alpha^{(n)} - c_{u,\alpha} - \hat{\epsilon}_\alpha^{(n)} + \epsilon_{u,\alpha}.$$

Hence,

$$\frac{\sqrt{k}(\hat{c}_{\epsilon,\alpha}^{(n)} - c_\alpha)}{\hat{\sigma}_n} = \frac{\sqrt{k}(\hat{c}_\alpha^{(n)} - c_{u,\alpha})}{\hat{\sigma}_n} - \frac{\sqrt{k}(\hat{\epsilon}_\alpha^{(n)} - \epsilon_{u,\alpha})}{\hat{\sigma}_n}. \tag{11}$$

For the first term on the right-hand side of eq. (11),

$$\frac{\sqrt{k}(\hat{c}_\alpha^{(n)} - c_{u,\alpha})}{\hat{\sigma}_n} = \frac{a(n/k)}{\hat{\sigma}_n} \frac{\sqrt{k}(\hat{c}_\alpha^{(n)} - c_{u,\alpha})}{a(n/k)} \xrightarrow{d} N(0, V), \tag{12}$$

which follows from theorem 5.1 and applying lemma A.2 with the fact that $\hat{\sigma}_n / a(n/k) \xrightarrow{p} 1$.

For the second term, first recall that

$$\frac{\hat{\epsilon}_\alpha^{(n)}}{a(n/k)A(n/k)} = \frac{\hat{\sigma}_n \hat{A}_n \hat{K}_n}{a(n/k)A(n/k)} \xrightarrow{p} K_{\xi,\rho}(\beta),$$

which follows from lemma A.2 and the continuous mapping theorem. Then, under the assumption that $\sqrt{k}A(n/k) \to \lambda < \infty$ $(n \to \infty)$, it follows from lemma A.2 and lemma A.5 that

$$\begin{aligned} \frac{\sqrt{k}(\hat{\epsilon}_\alpha^{(n)} - \epsilon_{u,\alpha})}{\hat{\sigma}_n} &= \frac{a(n/k)\sqrt{k}A(n/k)}{\hat{\sigma}_n} \left(\frac{\hat{\epsilon}_\alpha^{(n)} - \epsilon_{u,\alpha}}{a(n/k)A(n/k)} \right) \\ &= \lambda(1 + o_p(1)) \left(K_{\xi,\rho}(\beta) - K_{\xi,\rho}(\beta) + o_p(1)\right) = o_p(1), \quad n \to \infty. \end{aligned} \tag{13}$$

Combining the convergence in eq. (12) and eq. (13) with eq. (11), it follows that

$$\frac{\sqrt{k}(\hat{c}_{\epsilon,\alpha}^{(n)} - c_\alpha)}{\hat{\sigma}_n} \xrightarrow{d} N(0, V),$$

and hence,

$$\frac{\sqrt{k}(\hat{c}_{\epsilon,\alpha}^{(n)} - c_\alpha)}{\hat{\sigma}_n \sqrt{\hat{V}_n}} = \frac{\sqrt{k}(\hat{c}_{\epsilon,\alpha}^{(n)} - c_\alpha)}{\hat{\sigma}_n \sqrt{V}} \frac{\sqrt{V}}{\sqrt{\hat{V}_n}} \xrightarrow{d} N(0, 1),$$

which follows from the fact that $\hat{V}_n \xrightarrow{p} V$ (from the continuous mapping theorem) and lemma A.2. □

A.5 CONSISTENCY OF *A(n/k)* ESTIMATOR

In Haouas et al. [2018], an estimator for $A_0(n/k)$ is given,[1] where the function A_0 satisfies the second-order condition of De Haan and Ferreira [2006, Theorem 2.3.9], where for all $x > 0$,

$$\lim_{t \to \infty} \frac{\frac{U(tx)}{U(t)} - x^\xi}{A_0(t)} = x^\xi \frac{x^\rho - 1}{\rho}. \tag{14}$$

Note that under assumption 2.1 and assumption 2.2, eq. (14) is satisfied. The relation between the function A defined in eq. (6) and A_0 is given in De Haan and Ferreira [2006, Table 3.1], where

$$A = \frac{\xi + \rho}{\xi} A_0. \tag{15}$$

We shall use this relation and an estimator for $A_0(n/k)$ to derive an estimator for $A(n/k)$. To prove consistency of the forthcoming estimator, we start with the following relation from Haouas et al. [2018]:

$$\lim_{t \to \infty} \frac{A_0(t)}{R(t)} = 1, \tag{16}$$

[1] The results of Haouas et al. [2018] are presented in the truncated data setting, where for a sample (X_i, Y_i), $i = 1, \ldots, n$ from a couple of independent random variables (X, Y), X_i is only observed when $X_i \leq Y_i$. Their results can be adapted to the non-truncation setting by assuming that $\mathbb{P}(X \leq Y) = 1$.

where

$$R(t) = \frac{(1-\rho^2)(M^{(2)}(t) - 2(M^{(1)}(t))^2)}{2\rho M^{(1)}(t)}, \quad M^{(j)}(t) = t\int_{U(t)}^{\infty} \log^j (x/U(t))\, dF(x).$$

This leads to an estimator for $A_0(n/k)$ Haouas et al. [2018, p. 7],

$$\hat{A}_0^{(n)} = \frac{(1-\hat{\rho}_n^2)(\hat{M}_n^{(2)} - 2(\hat{M}_n^{(1)})^2)}{2\hat{\rho}_n \hat{M}_n^{(1)}},$$

where $\hat{M}_n^{(j)}$ is an estimator for $M^{(j)}(n/k)$, given by

$$\hat{M}_n^{(j)} = \frac{1}{k}\sum_{i=1}^{k}[\log X_{(n-i+1,k)} - \log X_{(n-k,n)}]^j,$$

which is also given in section 4. Note that $\hat{M}_n^{(1)}$ is the well-known Hill estimator of ξ. $\hat{M}_n^{(j)}$ is consistent for $j = 1, 2$ under the conditions of the following lemma.

Lemma A.6. *Suppose that assumption 2.1 holds. If $k = k_n \to \infty$, $k/n \to 0$ as $n \to \infty$, then*

$$\frac{\hat{M}_n^{(j)}}{M^{(j)}(n/k)} \xrightarrow{p} 1, \qquad j = 1, 2.$$

Proof. From Haouas et al. [2018, equation 1.9],

$$M^{(1)}(t) \to \xi \quad \text{and} \quad M^{(2)}(t) \to 2\xi^2, \quad t \to \infty.$$

By De Haan and Ferreira [2006, Theorem 3.2.2], $\hat{M}_n^{(1)} \xrightarrow{p} \xi$, and by De Haan and Ferreira [2006, Equation 3.5.7], $\hat{M}_n^{(2)} \xrightarrow{p} 2\xi^2$. Hence, by lemma A.2,

$$\frac{\hat{M}_n^{(1)}}{M^{(1)}(n/k)} \to 1 \quad \text{and} \quad \frac{\hat{M}_n^{(2)}}{M^{(2)}(n/k)} \to 1, \quad n \to \infty.$$

□

From eq. (15), an estimator for $A(n/k)$ is

$$\hat{A}_n \triangleq \frac{(\hat{\xi}_{\text{MLE}}^{(n)} + \hat{\rho}_n)(1-\hat{\rho}_n)^2(\hat{M}_n^{(2)} - 2(\hat{M}_n^{(1)})^2)}{2\hat{\xi}_{\text{MLE}}^{(n)}\hat{\rho}_n \hat{M}_n^{(1)}},$$

which is consistent under the conditions of the following lemma.

Lemma A.7. *Suppose that the assumptions of theorem 4.1 hold. If $\hat{\rho}_n \xrightarrow{p} \rho$,*

$$\frac{\hat{A}_n}{A(n/k)} \xrightarrow{p} 1.$$

Proof. By theorem 4.1, $\hat{\xi}_{\text{MLE}}^{(n)} \xrightarrow{p} \xi$, and so by eq. (16) and eq. (15),

$$\frac{\hat{A}_n}{A(n/k)} = \frac{\hat{\xi}_{\text{MLE}}^{(n)} + \hat{\rho}_n}{\hat{\xi}_{\text{MLE}}^{(n)}} \cdot \frac{\xi}{\xi+\rho} \cdot \frac{\hat{A}_0^{(n)}}{A_0(n/k)} = (1+o_p(1))\frac{\hat{A}_0^{(n)} R(n/k)}{R(n/k)A_0(n/k)} = (1+o_p(1))\frac{\hat{A}_0^{(n)}}{R(n/k)},$$

as $n \to \infty$. By lemma A.6,

$$\begin{aligned}\frac{\hat{A}_0^{(n)}}{R(n/k)} &= \frac{\rho(1-\hat{\rho}_n^2)M^{(1)}(n/k)}{\hat{\rho}_n(1-\rho^2)\hat{M}_n^{(1)}} \cdot \frac{\hat{M}_n^{(2)} - 2(\hat{M}_n^{(1)})^2}{M^{(2)}(n/k) - 2(M^{(1)}(n/k))^2} \\ &= (1+o_p(1))\frac{\hat{M}_n^{(2)} - 2(\hat{M}_n^{(1)})^2}{M^{(2)}(n/k) - 2(M^{(1)}(n/k))^2},\end{aligned}$$

as $n \to \infty$. From Gomes et al. [2002, p. 389],

$$\frac{\hat{M}_n^{(2)} - 2(\hat{M}_n^{(1)})^2}{A_0(n/k)} \xrightarrow{p} \frac{2\xi\rho}{(1-\rho)^2}.$$

Hence,

$$\frac{\hat{M}_n^{(2)} - 2(\hat{M}_n^{(1)})^2}{M^{(2)}(n/k) - 2(M^{(1)}(n/k))^2} = (1 + o_p(1))\frac{2\xi\rho}{(1-\rho)^2} \cdot \frac{A_0(n/k)}{M^{(2)}(n/k) - 2(M^{(1)}(n/k))^2}, \quad n \to \infty.$$

Finally, combining eq. (16) with the fact that $M^{(1)}(n/k) \to \xi$ as $n \to \infty$,

$$\frac{A_0(n/k)}{M^{(2)}(n/k) - 2(M^{(1)}(n/k))^2} = (1 + o(1))\frac{(1-\rho)^2}{2\rho M^{(1)}(n/k)} = (1 + o(1))\frac{(1-\rho)^2}{2\rho\xi}, \quad n \to \infty,$$

and thus

$$\frac{\hat{M}_n^{(2)} - 2(\hat{M}_n^{(1)})^2}{M^{(2)}(n/k) - 2(M^{(1)}(n/k))^2} \xrightarrow{p} 1 \qquad \text{which implies} \qquad \frac{\hat{A}_n}{A(n/k)} \xrightarrow{p} 1.$$

□

B ESTIMATION ALGORITHMS

B.1 ADAPTIVE ρ ESTIMATION

The ρ estimator given in section 4,

$$\hat{\rho}_n = \frac{3(T_n^{(\tau)}(m) - 1)}{T_n^{(\tau)}(m) - 3},$$

requires the choice of two parameters: a sample fraction m, and tuning parameter τ. Depending on the underlying distribution, the reliability of $\hat{\rho}_n$ can be very sensitive to the choice m and τ. The adaptive algorithm of [Caeiro and Gomes, 2015, Section 4.1] provides an automated way to select these parameters. We present a slightly modified version of their algorithm here, which we use in our experiments.

Algorithm 1: Adaptive algorithm for ρ estimation (ADARHO)

Input: An i.i.d. sample $X_1, .., X_n$, test parameters $\tau_1, \dots \tau_q$, test sample fractions $m_1, \dots, m_r$, precision p.
Output: $\hat{\rho}_n$

for $i = 1, \dots q$ **do**
  for $j = 1, \dots r$ **do**
    Compute $\hat{\rho}_n^{(\tau_i)}(m_j)$ using eq. (12), rounded to p decimal places
  end
  Set $m_{\text{min}}^{(\tau_i)}$, $m_{\text{max}}^{(\tau_i)}$ to be the minimum and maximum m values associated with the longest run of consecutive equal $\hat{\rho}_n^{(\tau_i)}$ values
  Set $l^{(\tau_i)} = m_{\text{max}}^{(\tau_i)} - m_{\text{min}}^{(\tau_i)} + 1$, the length of the largest run
end
Set $k = \arg\max_{i=1,\dots,q} l^{(\tau_i)}$
Set $\hat{\rho}_n$ to the median of $\hat{\rho}_n^{(\tau_k)}(m_{\text{min}}^{(\tau_k)}), \hat{\rho}_n^{(\tau_k)}(m_{\text{min}}^{(\tau_k)} + 1), \dots, \hat{\rho}_n^{(\tau_k)}(m_{\text{max}}^{(\tau_k)})$

B.2 AUTOMATED THRESHOLD SELECTION

The method of Bader et al. [2018] is as follows. Consider a fixed set of thresholds $u_1 < \ldots < u_l$, where for each u_i we have k_i excesses. The sequence of null hypotheses for each respective test i, $i = 1, \dots, l$, is given by

$$H_0^{(i)}: \quad \text{The distribution of the } k_i \text{ excesses above } u_i \text{ follows the GPD.}$$

For each threshold u_i, let $\hat{\theta}_i = (\hat{\xi}^{(n)}_{u_i}, \hat{\sigma}^{(n)}_{u_i})$ denote the MLEs computed from the k_i excesses above u_i. The Anderson-Darling (AD) test statistic comparing the empirical threshold excesses distribution with the GPD is then calculated. Let $y_{(1)} < ... < y_{(k_i)}$ denote the ordered threshold excesses for test i, and apply the transformation $z_{(j)} = G_{\hat{\theta}_i}(y_{(j)})$, $j = 1, \dots k_i$, where G denotes the cdf of the GPD. The AD statistic for test i is then

$$A_i^2 = -k_i - \frac{1}{k_i} \sum_{j=1}^{k_i} (2j-1) \left[\log\left(z_{(j)}\right) + \log\left(1 - z_{(k_i+1-j)}\right) \right]. \tag{17}$$

Corresponding p-values for each test statistic can then be found by referring to a lookup table (e.g., Choulakian and Stephens [2001]) or computed on-the-fly. Using the p-values $p_1, \dots, p_l$ calculated for each test, the ForwardStop rule of G'Sell et al. [2016] is used to choose the threshold. This is done by calculating

$$\hat{w}_F = \max \left\{ w \in I \,\middle|\, -\frac{1}{w} \sum_{i=1}^{w} \log\left(1 - p_i\right) \leq \gamma \right\}, \tag{18}$$

where γ is a chosen significance parameter and $I \subseteq \{1, \dots, l\}$, $I \neq \emptyset$. Under this rule, the threshold u_v is chosen, where $v = \min\{w \in I \,|\, w > \hat{w}_F\}$. If no $\hat{w}_F$ exists, then no rejection is made and $u_{\min(I)}$ is chosen. If $\hat{w}_F = \max(I)$, then $u_{\max(I)}$ is chosen. The overall procedure is summarized in Algorithm 2.

Remark B.1. *In the threshold selection procedure of Bader et al. [2018], $\hat{w}_F$ is given with $I = \{1, \dots, l\}$, but we make the modification that I is an arbitrary index set in view of CVaR estimation: since $c_{u,\alpha}$ tends to infinity when ξ tends to 1, in order to ensure reasonable estimates of the CVaR we use a cutoff parameter $\xi_{max} < 1$, where the MLE $\hat{\xi}^{(n)}_{u_i}$ and corresponding threshold u_i are discarded if $\hat{\xi}^{(n)}_{u_i} > \xi_{max}$.*

Remark B.2. *Instead of choosing the candidate thresholds $u_1, \dots, u_l$ directly, it is usually more convenient to choose threshold percentiles $q_1, \dots, q_l$ and compute u values via the empirical quantile function, i.e., $u_i = \hat{F}_n^{-1}(q_i)$.*

Algorithm 2: Automated threshold selection (AUTOTHRESH)

Input: An i.i.d. sample $X_1, .., X_n$, significance parameter γ, threshold percentiles $0 < q_1, \dots, q_l < 1$, cutoff $\xi_{max} < 1$.
Output: $(\hat{\xi}^{(n)}_{\text{MLE}}, \hat{\sigma}^{(n)}_{\text{MLE}})$, u if $I \neq \emptyset$, otherwise return NaN

$I \leftarrow \emptyset$
for $i = 1, \dots l$ **do**
    Set $u_i = \hat{F}_n^{-1}(q_i)$
    Compute $(\hat{\xi}^{(n)}_{u_i}, \hat{\sigma}^{(n)}_{u_i})$ from k_i threshold excesses using maximum likelihood
    if $\hat{\xi}^{(n)}_{u_i} \leq \xi_{max}$ **then**
        Compute A_i^2 using eq. (17)
        Set p_i to p-value for A_i^2 using lookup table
        $I \leftarrow I \cup \{i\}$
    end
end
if $I \neq \emptyset$ **then**
    Set $W = \{w \in I \,|\, -\frac{1}{w} \sum_{i=1}^{w} \log\left(1 - p_i\right) \leq \gamma\}$
    if $W \neq \emptyset$ **then**
        Compute $\hat{w}_F$ using eq. (18)
        if $\hat{w}_F = \max(I)$ **then**
            $v \leftarrow \max(I)$
        else
            $v \leftarrow \min\{w \in I \,|\, w > \hat{w}_F\}$
        end
    else
        $v \leftarrow \min(I)$
    end
    $u \leftarrow u_v$
    $(\hat{\xi}^{(n)}_{\text{MLE}}, \hat{\sigma}^{(n)}_{\text{MLE}}) \leftarrow (\hat{\xi}^{(n)}_{u_v}, \hat{\sigma}^{(n)}_{u_v})$
end

B.3 ALGORITHM TO COMPUTE THE UNBIASED POT ESTIMATOR

This section provides the algorithm used to compute UPOT in its entirety, which makes use of both algorithm 1 and algorithm 2. In our experiments, we set $\tau_1 = -1.5, \tau_2 = -1.25, \ldots, \tau_{13} = 1.5$, $m_1 = 100, m_2 = 200, \ldots, m_r = n - 1$, $p = 1$ in algorithm 1, and $\gamma = 0.1$, $q_1 = 0.79, q_2 = 0.80, \ldots, q_{20} = 0.98$, $\xi_{max} = 0.9$ in algorithm 2. Assume these choices of values in the following algorithm.

Algorithm 3: Unbiased peaks-over-threshold CVaR estimator (UPOT)

Input: An i.i.d. sample $X_1, .., X_n$, confidence level α

Output: $\hat{c}_{\epsilon,\alpha}^{(n)}$

$\mathbf{x} \leftarrow$ AUTOTHRESH$(X_1, .., X_n)$

if $\mathbf{x}$ is not NaN **then**

  $(\hat{\xi}_{\text{MLE}}^{(n)}, \hat{\sigma}_{\text{MLE}}^{(n)}), u \leftarrow \mathbf{x}$

  $\hat{\rho_n} \leftarrow$ ADARHO$(X_1, .., X_n)$

  Compute $\hat{b}_n$ using eq. (14)

  Compute $\hat{A}_n$ using eq. (13) and the k threshold excesses above u

  $(\hat{\xi}_n, \hat{\sigma}_n) \leftarrow (\hat{\xi}_{\text{MLE}}^{(n)} - \hat{A}_n \hat{b}_n^{(1)},\ \hat{\sigma}_{\text{MLE}}^{(n)}(1 - \hat{A}_n \hat{b}_n^{(2)}))$

  Compute $\hat{\epsilon}_{\alpha}^{(n)}$ using eq. (18)

  Compute $\hat{c}_{\alpha}^{(n)}$ using eq. (16)

  $\hat{c}_{\epsilon,\alpha}^{(n)} \leftarrow \hat{c}_{\alpha}^{(n)} - \hat{\epsilon}_{\alpha}^{(n)}$

end

Remark B.3. *It may happen that Algorithm 3 fails if AUTOTHRESH returns NaN, in which case no suitable estimates of ξ are found. This is an indication that the underlying data distribution does not satisfy the condition $\xi < 1$ and the CVaR does not exist. To make Algorithm 3 robust, the sample average estimate is used as a fallback when the latter occurs. We report the failure rate of UPOT during experiments in table 1.*

C EXAMPLES OF HEAVY-TAILED DISTRIBUTIONS

C.1 BURR

The Burr distribution with parameters c, d has cdf given by

$$F_{c,d}(x) = 1 - (1 + x^c)^{-d}, \quad c, d, x > 0.$$

The CVaR for the Burr distribution can be derived from its expression for the conditional moment given in Kumar [2017, Section 2.2]. If $X \sim \text{Burr}(c, d)$,

$$\text{CVaR}_\alpha(X) = \frac{d[(1/q_\alpha)^c]^{d-1/c}}{(1-\alpha)(d-1/c)}\, {}_2F_1\left(d - \frac{1}{c}, 1 + d, d - \frac{1}{c} + 1; -\frac{1}{q_\alpha}\right), \quad cd > 1, \tag{19}$$

where ${}_2F_1$ denotes the hypergeometric function and $q_\alpha = F_{c,d}^{-1}(\alpha)$. Values of ξ, ρ and functions a and A are given by

$$\xi = \frac{1}{cd}, \quad \rho = -\frac{1}{d}, \quad a(t) = \frac{t^{1/d}}{cd}\left(t^{1/d} - 1\right)^{1/c-1}, \quad A(t) = \frac{1-c}{cd(t^{1/d} - 1)},$$

where a and A are defined for $t \geq 1$.

C.2 FRÉCHET

The Fréchet distribution with parameter γ has cdf given by

$$F_\gamma(x) = e^{-x^{-\gamma}}, \quad \gamma, x > 0.$$

If $X \sim$ Fréchet(γ),

$$\text{CVaR}_\alpha(X) = (1-\alpha)^{-1}\left[\Gamma\left(\gamma - 1/\gamma\right) - \Gamma\left(\gamma - 1/\gamma, -\log(\alpha)\right)\right], \quad \gamma > 1, \tag{20}$$

where $\Gamma(\cdot)$ and $\Gamma(\cdot,\cdot)$ denote the gamma and upper incomplete gamma functions, respectively. Values of ξ, ρ and functions a and A are given by

$$\xi = \frac{1}{\gamma}, \quad \rho = -1, \quad a(t) = \frac{\log\left(\frac{t}{t-1}\right)^{-1-\frac{1}{\gamma}}}{\gamma(t-1)}, \quad A(t) = -\frac{1+\gamma+\gamma t\log(1-\frac{1}{t})}{\gamma(1-t)\log(1-\frac{1}{t})} - \frac{1}{\gamma},$$

where a and A are defined for $t \geq 1$.

C.2.1 Asymptotic variance of SA estimator for Fréchet distribution

An expression for the asymptotic variance (AVAR) of the SA estimator is given in [Trindade et al., 2007, Theorem 2]. Let Z be a continuous random variable such that $\mathbb{E}[Z^2]$ is finite. Then, for a confidence level α,

$$\sqrt{n}\left(\text{CVaR}_\alpha(Z) - \widehat{\text{CVaR}}_{n,\alpha}(Z)\right) \xrightarrow{d} N(0,\theta^2),$$

where $\widehat{\text{CVaR}}_{n,\alpha}(Z)$ is the SA estimator given in eq. (3) and

$$\theta^2 = \frac{\text{Var}\left([Z-q_\alpha]^+\right)}{(1-\alpha)^2},$$

and $[x]^+ = \max\{0,x\}$. If $Z \sim$ Fréchet(γ), the condition that $\mathbb{E}[Z^2]$ is finite is equivalent to $\gamma > 2$. By the law of total expectation,

$$\mathbb{E}[[Z-q_\alpha]^+] = \mathbb{P}(Z \leq q_\alpha)\mathbb{E}[0] + \mathbb{P}(Z > q_\alpha)\mathbb{E}[Z-q_\alpha | Z > q_\alpha] = (1-e^{-q_\alpha^{-\gamma}})\mathbb{E}[Z-q_\alpha|Z>q_\alpha].$$

The distribution of the conditional random variable on the right hand side has the same form as the excess distribution function, given in definition 2.4. Let

$$F_{\alpha,\gamma}(z) = \mathbb{P}(Z-q_\alpha \leq z | Z > q_\alpha) = \frac{F_\gamma(z+q_\alpha) - F_\gamma(q_\alpha)}{1-F_\gamma(q_\alpha)} = \frac{e^{-(z+q_\alpha)^{-\gamma}} - e^{-q_\alpha^{-\gamma}}}{1-e^{-q_\alpha^{-\gamma}}},$$

$$f_{\alpha,\gamma}(z) = F'_{\alpha,\gamma}(z) = \frac{\gamma(z+q_\alpha)^{-\gamma-1}e^{-(z+q_\alpha)^{-\gamma}}}{1-e^{-q_\alpha^{-\gamma}}}, \quad z > 0.$$

Hence,

$$\mathbb{E}[[Z-q_\alpha]^+] = \int_0^\infty \gamma(z+q_\alpha)^{-\gamma-1} z e^{-(z+q_\alpha)^{-\gamma}} dz = \int_{q_\alpha}^\infty t^{-1/\gamma}e^{-t}dt = \Gamma(1-1/\gamma, q_\alpha), \quad \gamma > 1,$$

where $t = (z+q_\alpha)^{-\gamma}$ and $\Gamma(\cdot,\cdot)$ denotes the upper incomplete gamma function. With a similar calculation, the second moment is

$$\mathbb{E}[([Z-q_\alpha]^+)^2] = \Gamma(1-2/\gamma, q_\alpha), \quad \gamma > 2.$$

Finally, we can compute the AVAR of the SA estimator for the Fréchet distribution, which is

$$\frac{\theta^2}{n} = \frac{\mathbb{E}[([Z-q_\alpha]^+)^2] - \mathbb{E}[[Z-q_\alpha]^+]^2}{n(1-\alpha)^2} = \frac{\Gamma(1-2/\gamma, q_\alpha) - \Gamma(1-1/\gamma, q_\alpha)^2}{n(1-\alpha)^2}, \quad \gamma > 2.$$

C.3 HALF-T

If X follows the t distribution with ν degrees of freedom, then $|X|$ follows the half-t distribution, which has cdf given by

$$F_\nu(x) = 2 - \mathcal{I}_{t(x)}\left(\frac{\nu}{2}, \frac{1}{2}\right), \quad \nu > 0, x \geq 0,$$

where $t(x) = \frac{\nu}{x^2+\nu}$ and $\mathcal{I}_t(a,b)$ is the regularized incomplete Beta function. The CVaR for the half-t distribution can be derived from the expression for the CVaR of the t-distribution given in [Norton et al., 2019, Proposition 12]. If $X \sim$ half-$t(\nu)$, then

$$\text{CVaR}_\alpha(X) = 2\frac{\nu + q_\alpha}{(\nu-1)(1-\alpha)} g_\nu(q_\alpha), \quad \nu > 1,$$

where g_ν is the probability density function of the standardized t-distribution, and $q_\alpha = T^{-1}\left(\frac{\alpha+1}{2}\right)$ where T^{-1} is the inverse of the cdf of standardized t-distribution. The half-t distribution is in MDA(H_ξ) with $\xi = 1/\nu$, and has $\rho = -2/\nu$ (Caeiro and Gomes [2015, Remark 2.1]). It does not seem possible to compute closed-form expressions for the functions a and A for the half-t distribution.

D NUMERICAL RESULTS

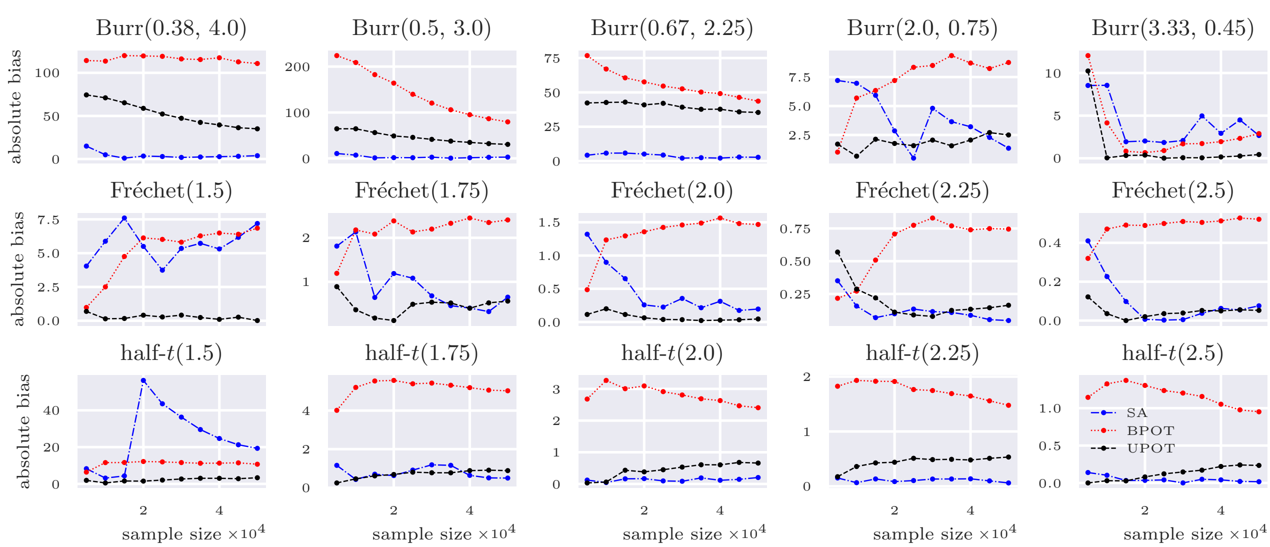

Figure 1: Absolute bias of estimating $\text{CVaR}_{0.998}$ using **UPOT** (black), **BPOT** (red), and **SA** (blue).

Table 1: Data for all distributions used in experiments. CVaR_α denotes the exact CVaR value for $\alpha = 0.998$. Given at a sample size $n = 50000$, UPOT, BPOT, and SA denote the average estimated CVaR values across $N = 1000$ independent runs, and TP denotes the average threshold percentile chosen by algorithm 2. FR denotes the failure rate, the number of independent runs where algorithm 2 returned NaN, i.e., where no suitable estimate of ξ could be obtained and no CVaR estimate could be produced by the POT methods. This value is given at a sample size of $n = 5000$ since very few failures occurred beyond this sample size. CP denotes the coverage probability achieved by our confidence interval at a sample size $n = 50000$.

	CVaR_α	UPOT	BPOT	SA	TP	FR	CP
Burr(0.38, 4.0)	124.87	89.83	235.70	121.03	0.96	2	0.73
Burr(0.5, 3.0)	166.18	135.62	245.74	163.39	0.92	1	0.87
Burr(0.67, 2.25)	175.93	140.53	219.55	173.19	0.84	0	0.88
Burr(2.0, 0.75)	188.98	191.48	180.22	190.34	0.80	0	0.94
Burr(3.33, 0.45)	190.15	189.71	187.26	192.83	0.80	4	0.95
Fréchet(1.5)	188.96	188.94	182.11	181.76	0.80	4	0.89
Fréchet(1.75)	81.32	81.88	78.91	81.97	0.80	2	0.93
Fréchet(2.0)	44.71	44.76	43.25	44.52	0.80	1	0.94
Fréchet(2.25)	28.49	28.66	27.75	28.45	0.80	2	0.95
Fréchet(2.5)	20.02	20.07	19.49	19.94	0.80	1	0.95
half-t(1.5)	156.58	159.92	145.89	175.91	0.81	0	0.94
half-t(1.75)	74.52	75.40	69.49	74.02	0.82	2	0.94
half-t(2.0)	44.70	45.36	42.29	44.50	0.83	0	0.94
half-t(2.25)	30.74	31.27	29.26	30.68	0.84	1	0.95
half-t(2.5)	23.10	23.34	22.15	23.12	0.85	1	0.94

Table 2: Error values at sample size $n = 50000$.

	RMSE			Bias		
	UPOT	BPOT	SA	UPOT	BPOT	SA
Burr(0.38, 4.0)	48.56	134.15	64.04	-35.03	110.83	-3.84
Burr(0.5, 3.0)	47.71	121.18	124.71	-30.56	79.56	-2.78
Burr(0.67, 2.25)	48.88	58.97	81.34	-35.41	43.62	-2.75
Burr(2.0, 0.75)	17.48	22.27	88.48	2.50	-8.76	1.36
Burr(3.33, 0.45)	13.83	19.40	128.88	-0.44	-2.89	2.67
Fréchet(1.5)	19.47	21.31	69.35	-0.02	-6.85	-7.19
Fréchet(1.75)	6.10	7.07	24.25	0.56	-2.41	0.65
Fréchet(2.0)	2.71	3.36	7.45	0.05	-1.47	-0.20
Fréchet(2.25)	1.50	1.90	3.21	0.16	-0.75	-0.05
Fréchet(2.5)	0.92	1.18	1.69	0.05	-0.52	-0.08
half-t(1.5)	16.78	22.68	765.05	3.34	-10.69	19.33
half-t(1.75)	6.11	8.72	16.40	0.89	-5.03	-0.50
half-t(2.0)	3.58	4.92	7.62	0.66	-2.41	-0.20
half-t(2.25)	2.07	2.78	3.49	0.53	-1.48	-0.06
half-t(2.5)	1.44	1.88	2.06	0.23	-0.95	0.02

References

Brian Bader, Jun Yan, Xuebin Zhang, et al. Automated threshold selection for extreme value analysis via ordered goodness-of-fit tests with adjustment for false discovery rate. *The Annals of Applied Statistics*, 12(1):310–329, 2018.

Jan Beirlant, J. P. Raoult, and Rym Worms. On the relative approximation error of the generalized Pareto approximation for a high quantile. *Extremes*, 6:335–360, 2003.

Jan Beirlant, Elisabeth Joossens, and Johan Segers. Second-order refined peaks-over-threshold modelling for heavy-tailed distributions. *Journal of Statistical Planning and Inference*, 139(8):2800 – 2815, 2009. ISSN 0378-3758. doi: https://doi.org/10.1016/j.jspi.2009.01.006. URL http://www.sciencedirect.com/science/article/pii/S0378375809000068.

Frederico Caeiro and M. Ivette Gomes. Bias reduction in the estimation of a shape second-order parameter of a heavy-tailed model. *Journal of Statistical Computation and Simulation*, 85(17):3405–3419, 2015. doi: 10.1080/00949655.2014.975707.

V Choulakian and M. A Stephens. Goodness-of-fit tests for the generalized Pareto distribution. *Technometrics*, 43(4): 478–484, 2001. doi: 10.1198/00401700152672573. URL https://doi.org/10.1198/00401700152672573.

Laurens De Haan and Ana Ferreira. *Extreme Value Theory: An Introduction*. Springer-Verlag New York, 2006. doi: 10.1007/0-387-34471-3.

M. Gomes, L. Haan, and L. Peng. Semi-parametric estimation of the second order parameter in statistics of extremes. *Extremes*, 5:387–414, 2002.

Max Grazier G'Sell, Stefan Wager, Alexandra Chouldechova, and Robert Tibshirani. Sequential selection procedures and false discovery rate control. *Journal of the Royal Statistical Society Series B*, 78(2):423–444, March 2016. URL https://ideas.repec.org/a/bla/jorssb/v78y2016i2p423-444.html.

Nawel Haouas, Abdelhakim Necir, and Brahim Brahimi. Estimating the second-order parameter of regular variation and bias reduction in tail index estimation under random truncation. *Journal of Statistical Theory and Practice*, 13, 10 2018.

Devendra Kumar. The Singh–Maddala distribution: properties and estimation. *International Journal of System Assurance Engineering and Management*, 8, 03 2017. doi: 10.1007/s13198-017-0600-1.

Matthew Norton, Valentyn Khokhlov, and Stan Uryasev. Calculating CVaR and bPOE for Common Probability Distributions With Application to Portfolio Optimization and Density Estimation. *Annals of Operations Research*, 2019. URL https://doi.org/10.1007/s10479-019-03373-1.

Bruno Rémillard. *Statistical methods for financial engineering*. Chapman and Hall/CRC, 2016.

A. Alexandre Trindade, Stan Uryasev, Alexander Shapiro, and Grigory Zrazhevsky. Financial prediction with constrained tail risk. *Journal of Banking & Finance*, 31(11):3524 – 3538, 2007. ISSN 0378-4266. doi: https://doi.org/10.1016/j.jbankfin.2007.04.014. URL http://www.sciencedirect.com/science/article/pii/S0378426607001392. Risk Management and Quantitative Approaches in Finance.

A. W. van der Vaart. *Asymptotic Statistics*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 1998. doi: 10.1017/CBO9780511802256.