Appendix B Additional Simulations

The aim of this section is twofold. First, we present additional insight into when our method is likely to work well in terms of bias, RMSE and coverage. Second, we rerun the simulations for the models of Section 3 and three additional models used in this Appendix with an additional estimator of γ to further corroborate our main conclusions. We do so using the estimator of Csörgő and Viharos (1998), given by

$$\widehat{\gamma}_{CV_\zeta}(t) = \int_1^\infty \left[\int_0^{\widehat{F}_n(t,y)} J_\zeta(s)\mathrm{d}s \right] \frac{\mathrm{d}y}{y}$$

with $J_\zeta(s) = \zeta^{-1}(\zeta+1)[1-(\zeta+1)s^\zeta]$, $s \in [0,1]$, $\zeta > 0$. We remark that Theorem 1 holds verbatim for this estimator with $\widehat{\sigma}^2_{\widehat{\gamma},\gamma} = (2\zeta+2)/(2\zeta+1)\widehat{\gamma}^2_{CV_\zeta}(1)$. We refer to Hoga (2017, Example 3) for more detail. In the following, we set $\zeta = 1$ and term the corresponding estimator CV_1. Using other values of ζ (e.g., $\zeta = 10$) did not change the results much.

We consider the following three additional models. For simplicity, we use the GARCH(1,1) specification in (27), yet we use three different distributions for the innovations U_i now. First, we use the (standardized) $st_3(0)$-distribution instead of $st_3(5)$ to see if coverage improves for a non-skewed distribution with infinite fourth moment, due to $\gamma = 1/3$. As a second and third distribution, we take equally heavy-tailed Burr distributions, Burr(β, λ, τ), with d.f.

$$F(x) = 1 - \left(\frac{\beta}{\beta + x^\tau}\right)^\lambda, \qquad x > 0, \quad \beta,\ \tau,\ \lambda > 0.$$

The extreme value index is given by $\gamma = 1/(\tau\lambda)$. The function $A(\cdot)$ in Assumption 5 can be chosen as $A(x) = 1/(\tau\lambda)x^{-1/\lambda}$. Hence, the smaller $\lambda > 0$ the faster $A(x)$ tends to zero, as $x \to \infty$. By (15), this means that the smaller λ the better the approximation to true Pareto behavior, where identically $U(xy)/U(x) = y^\gamma$. We set the parameters (β, λ, τ) equal to $(1, 1.5, 2)$ and $(1, 1, 3)$ for i.i.d. r.v.s ξ_i to study the effect of a more accurate approximation to the Pareto tail with smaller λ, while keeping the extreme value index equal to $\gamma = 1/3$ in both cases by the appropriate choice of τ. Since the ξ_i have positive support, we use the symmetrized innovations $U_i = B_i\xi_i/\sqrt{\mathrm{E}[\xi_1^2]}$, where B_i are i.i.d. zero-mean binary random variables on $\{-1, 1\}$, independent of the ξ_i. Note that $\mathrm{E}[U_i] = 0$ and $\mathrm{E}[U_i^2] = 1$, as required. The three resulting models will simply be termed $st_3(0)$, Burr(1.5) (for $(\beta, \lambda, \tau) = (1, 1.5, 2)$) and Burr(1) (for $(\beta, \lambda, \tau) = (1, 1, 3)$). As model (27) in the main paper, they will be estimated using Laplace QMLE.

Table 4 shows the simulation results for these models. We draw the following conclusions:

1. Using the symmetric $st_3(0)$-distribution, instead of the highly skewed $st_3(5)$-distribution as in

model (27), leads to improved coverage in all cases. This is particularly true for the Hill estimator and self-normalized intervals, where coverage is very close to the nominal level now. This may be explained by the better approximation to the Pareto shape of the left tail (that we consider here) of the $st_3(0)$-distribution. In contrast, the $st_3(5)$-distribution, while theoretically possessing two tails with the same power-like tail decay (Aas and Haff, 2006), has a left tail that displays marked deviations from the Pareto shape at finite levels due to the significant skewness. This result serves to highlight the importance of empirically checking the regular variation assumption of the innovations using some of the methods discussed in Remark 3 (c). By doing so, one can possibly identify cases in advance, where coverage may not be accurate.

2. Comparing models Burr(1.5) and Burr(1), we find that for Burr(1) – where the Pareto approximation is more accurate – the estimators work better. While coverage is similar for both models, bias (in most cases) and RMSE (in all cases) are reduced and, as it should be, estimation uncertainty – as measured by the width of confidence intervals – is reduced as well. Additionally, and also as it should be, our method of choosing k^* leads to more upper order statistics being used in the Burr(1) case, as the Pareto shape can be exploited for more tail observations.

   It is interesting to note that if the innovation distributions are close to Pareto, as is the case for the Burr models, the Hill estimator performs consistently better in terms of RMSE than the MR estimator. This is not the case for $st_3(0)$ and the other models considered in Section 3. This may be explained by the fact that the Hill estimator is the maximum likelihood estimate of the extreme value index of true Pareto-distributed r.v.s (Hill, 1975) and thus possesses the usual optimality properties in this distribution class. So the closer the distributional tail of the innovations is to true Pareto behavior, the better the relative performance of the Hill estimator can be expected to be.

Tables 2 and 4 reveal that self-normalization quite generally leads to more accurate coverage of CVaR and CES. Yet, for the MR estimator and extreme CVaR, coverage deteriorated. So Table 5 provides additional simulation evidence in favor of self-normalization. It displays simulation results for the CV_1 estimator for all models considered in Tables 2 and 4. For this particular estimator we find coverage to be much improved using $I_{\text{sn}}^{0.95}$. Indeed, coverage is even better than for the Hill estimator and quite often within 1% of the nominal level. The RMSE of the CV_1 estimator is roughly comparable with that of the other two. However, the self-normalized intervals tend to be somewhat wider.

Model	Estimator	k^*	z	α	Bias	RMSE	Coverage		Int. length	
							$I_{\text{na}}^{0.95}$	$I_{\text{sn}}^{0.95}$	$I_{\text{na}}^{0.95}$	$I_{\text{sn}}^{0.95}$
$st_3(0)$	Hill	63	CVaR	2.5%	0.18	1.15	66.5	88.3	1.5	3.1
				1%	0.08	1.99	87.5	91.5	4.5	6.7
				0.5%	−0.31	2.74	92.3	94.1	8.3	11.7
			CES	2.5%	−0.57	2.85	46.3	94.2	2.6	11.1
				1%	−1.74	5.83	68.1	95.0	7.8	22.5
				0.5%	−3.21	8.19	73.8	95.4	14.3	36.0
	MR	70	CVaR	2.5%	0.64	1.27	63.1	82.2	1.9	3.5
				1%	0.95	2.09	84.5	82.3	5.7	6.9
				0.5%	1.00	2.84	91.2	84.4	10.3	11.7
			CES	2.5%	0.57	2.77	51.3	85.0	3.3	10.9
				1%	0.29	5.35	73.7	86.4	9.6	20.7
				0.5%	−0.33	7.56	81.5	88.0	17.3	34.0
Burr(1.5)	Hill	69	CVaR	2.5%	0.10	0.94	70.2	89.4	1.5	3.0
				1%	0.09	1.62	87.1	91.6	4.1	6.2
				0.5%	−0.17	2.77	91.0	93.4	7.5	10.8
			CES	2.5%	−0.36	2.68	48.0	93.6	2.4	9.8
				1%	−1.00	4.43	68.3	94.8	6.8	19.2
				0.5%	−2.02	8.04	75.5	94.7	12.4	30.5
	MR	74	CVaR	2.5%	0.47	1.05	70.1	85.6	1.9	3.5
				1%	0.74	1.78	86.0	84.1	5.4	6.8
				0.5%	0.79	3.15	91.5	85.3	9.7	11.6
			CES	2.5%	0.44	2.76	51.3	86.1	3.2	10.7
				1%	0.34	4.61	72.0	86.3	8.8	20.0
				0.5%	−0.11	9.63	80.3	87.5	15.9	32.4
Burr(1)	Hill	79	CVaR	2.5%	0.07	0.70	74.4	90.5	1.3	2.6
				1%	0.09	1.26	87.4	91.7	3.5	5.3
				0.5%	0.01	2.21	89.3	92.7	6.2	9.1
			CES	2.5%	−0.01	1.78	49.9	92.4	2.1	7.8
				1%	−0.23	3.27	66.5	92.8	5.5	14.6
				0.5%	−0.57	5.25	73.3	93.2	9.7	23.0
	MR	81	CVaR	2.5%	0.32	0.79	76.3	88.5	1.8	3.1
				1%	0.46	1.46	88.8	86.9	4.8	6.2
				0.5%	0.53	2.54	91.5	86.4	8.5	10.5
			CES	2.5%	0.44	2.05	53.2	86.3	2.8	9.0
				1%	0.42	3.72	71.1	86.9	7.5	17.0
				0.5%	0.33	6.01	77.6	86.6	13.3	27.0

Table 4: Average of k^*, bias, RMSE, coverage probabilities in % (nominal coverage: 95%) and average interval lengths for Hill and MR estimator for models $st_3(0)$, Burr(1.5) and Burr(1).

Model	Estimator	k^*	z	α	Bias	RMSE	Coverage		Int. length	
							$I_{\text{na}}^{0.95}$	$I_{\text{sn}}^{0.95}$	$I_{\text{na}}^{0.95}$	$I_{\text{sn}}^{0.95}$
(27)	CV_1	81	CVaR	2.5%	0.01	0.74	56.6	82.9	0.33	0.76
				1%	−0.07	0.90	76.4	87.5	0.70	1.19
				0.5%	−0.14	1.15	82.0	90.2	1.04	1.65
			CES	2.5%	−0.06	1.00	46.5	88.1	0.39	1.35
				1%	−0.12	1.16	64.6	89.9	0.82	2.01
				0.5%	−0.11	1.39	71.6	91.1	1.23	2.67
(28)	CV_1	62	CVaR	2.5%	0.95	2.26	59.7	89.0	3.3	8.5
				1%	0.72	3.69	87.0	92.8	9.7	15.6
				0.5%	−0.44	5.02	94.1	94.6	17.2	24.9
			CES	2.5%	−0.64	4.34	51.6	95.2	5.2	21.8
				1%	−3.16	8.32	71.8	95.4	15.0	40.1
				0.5%	−6.65	13.8	76.8	94.6	26.6	61.4
(29)	CV_1	60	CVaR	2.5%	0.54	1.17	53.9	84.8	1.4	3.4
				1%	0.45	1.53	84.7	90.6	3.9	5.8
				0.5%	−0.00	2.13	93.2	94.1	6.6	9.1
			CES	2.5%	−0.01	1.63	52.4	93.8	2.0	7.4
				1%	−0.86	2.83	74.6	95.6	5.3	12.5
				0.5%	−2.03	4.50	80.0	94.5	9.1	18.7
$st_3(0)$	CV_1	68	CVaR	2.5%	0.45	1.19	66.3	90.5	1.8	4.1
				1%	0.32	2.04	89.1	93.7	5.3	8.4
				0.5%	−0.14	2.84	94.1	95.0	9.8	14.4
			CES	2.5%	−0.46	2.98	51.2	95.4	3.2	14.2
				1%	−1.86	6.47	71.4	95.3	9.4	28.1
				0.5%	−3.65	9.15	76.5	95.1	17.2	45.4
Burr(1.5)	CV_1	76	CVaR	2.5%	0.31	0.96	72.6	92.9	1.8	3.9
				1%	0.21	1.68	89.6	94.1	4.9	7.9
				0.5%	−0.19	3.03	93.8	95.4	9.0	13.7
			CES	2.5%	−0.41	2.82	53.3	95.4	3.0	12.9
				1%	−1.35	5.00	70.7	95.3	8.4	24.5
				0.5%	−2.78	9.69	78.1	95.6	15.3	40.1
Burr(1)	CV_1	87	CVaR	2.5%	0.19	0.74	77.8	94.4	1.6	3.4
				1%	0.07	1.35	91.1	95.1	4.3	7.0
				0.5%	−0.18	2.41	93.1	95.0	7.6	11.6
			CES	2.5%	−0.16	1.99	55.4	95.4	2.6	10.3
				1%	−0.74	3.73	71.8	95.4	6.9	19.5
				0.5%	−1.49	6.07	77.8	95.0	12.2	30.8

Table 5: Average of k^*, bias, RMSE, coverage probabilities in % (nominal coverage: 95%) and average interval lengths for CV_1 estimator for models (27), (28), (29), $st_3(0)$, Burr(1.5) and Burr(1). For models (28) and (29), values of bias, RMSE and interval lengths are premultiplied with 10^3.

References

Aas, K. and I.H. Haff (2006). "The Generalized Hyperbolic Skew Student's t-Distribution". *Journal of Financial Econometrics* 4, pp. 275–309.

Csörgő, S. and L. Viharos (1998). "Estimating the Tail Index". *Asymptotic Methods in Probability and Statistics*. Ed. by B. Szyszkowicz. Amsterdam: Elsevier, pp. 833–881.

Hill, B. (1975). "A Simple General Approach to Inference About the Tail of a Distribution". *The Annals of Statistics* 3, pp. 1163–1174.

Hoga, Y. (2017). "Change Point Tests for the Tail Index of β-Mixing Random Variables". *Econometric Theory* 33, pp. 915–954.