Supplementary Material for

"ExpectHill estimation, extreme risk and heavy tails"

Abdelaati Daouia[a], Stéphane Girard[b] and Gilles Stupfler[c]

[a] Toulouse School of Economics, University of Toulouse Capitole, France

[b] Univ. Grenoble Alpes, Inria, CNRS, Grenoble INP, LJK, France

[c] Univ Rennes, Ensai, CNRS, CREST - UMR 9194, F-35000 Rennes, France

Section A illustrates the properties of the *expectHill* estimator of the tail index with a real data example. Simulation results are discussed in Section B. The proofs of all theoretical results in the main paper and additional technical results are provided in Section C.

A On expectHill estimation

The aim of this section is to illustrate the behavior of the *expectHill* estimator with empirical data and to highlight some of the theoretical findings in Section 3 of the main article. First, the purely expectile-based estimator

$$\widetilde{\gamma}_{\tau_n} = \frac{1}{\lfloor n(1-\tau_n)\rfloor} \sum_{i=1}^{\lfloor n(1-\tau_n)\rfloor} \log\left(\frac{\widetilde{\xi}_{1-(i-1)/n}}{\widetilde{\xi}_{1-\lfloor n(1-\tau_n)\rfloor/n}}\right)$$

of the tail index γ has exactly the same form as the quantile-based Hill estimator

$$\widehat{\gamma}_{\tau_n} = \frac{1}{\lfloor n(1-\tau_n)\rfloor} \sum_{i=1}^{\lfloor n(1-\tau_n)\rfloor} \log\left(\frac{\widehat{q}_{1-(i-1)/n}}{\widehat{q}_{1-\lfloor n(1-\tau_n)\rfloor/n}}\right)$$

with the tail empirical quantile process $\widehat{q}$ replaced by its least squares analogue $\widetilde{\xi}$. Theorem 1 gives its asymptotic normality. As pointed out in Remark 1, the conditions involving the auxiliary function A in Theorem 1 are also required to derive the asymptotic normality of Hill's estimator $\widehat{\gamma}_{\tau_n}$. These conditions are, however, difficult to check in practice, which makes the choice of the intermediate sequence τ_n a hard problem. A usual practice for selecting a reasonable estimate $\widehat{\gamma}_{\tau_n}$ is to set $\tau_n = 1 - k/n$ for a sequence of integers k, then to plot the graph of $k \mapsto \widehat{\gamma}_{1-k/n}$, and finally to pick out a value of k corresponding to the first stable part of the plot (see Remark 2). Yet, the Hill plot may be so unstable that reasonable values of k (which would correspond to the true value of γ) may be hidden in the graph. The least squares analogue $\widetilde{\gamma}_{1-k/n}$ affords a smoother and more stable plot which counteracts the

volatility defect of the Hill plot. To illustrate this advantage of the expectHill estimator, we use trade data on the SPDR S&P 500 ETF (SPY), which is an exchange traded fund (ETF) that tracks the S&P 500 index. The dataset comprises 10 years of trade data on SPY starting from June 15th, 2004, to June 13th, 2014. First, we estimate the tail index of the daily open-to-close loss returns (*i.e.* minus returns) using the whole sample with 2,497 days of trade data. Figure 1(a) shows the paths $k \mapsto \widehat{\gamma}_{1-k/n}$ in red and $k \mapsto \widetilde{\gamma}_{1-k/n}$ in blue based on these data, for the selected range of values $1 \leqslant k \leqslant n/\log n$ (this restriction allows to reject large values of k). To determine a reasonable pointwise estimate, say $\widehat{\gamma}_{1-k/n}$, or equivalently an appropriate sample fraction k, we applied the same simple automatic data-driven device as El Methni and Stupfler (2017) and Daouia *et al.* (2020). Built on the idea of balancing the potential estimation bias and variance, this automatic selection consists first in computing the standard deviations of the estimator over a *moving window* large enough to cover around 5% of the possible values of k in the selected range $1 \leqslant k \leqslant n/\log n$. The first window over which the standard deviation has a local minimum, and is less than the average standard deviation across all windows, may be viewed as the first stable region of the plot. The value of k for which $\widehat{\gamma}_{1-k/n}$ is the median estimate within this window defines then the desired sample fraction. Appropriate values of k, or $\tau_n = 1 - k/n$, in all our proposed expectile-based estimators are chosen according to this selection procedure.

The resulting final estimates $\widehat{\gamma}_{1-k/n}$ and $\widetilde{\gamma}_{1-k/n}$ are reported in the second row of Table 1 (third and fifth columns), along with their associated sample fractions k and stable regions. These results confirm the model assumption $\gamma < \frac{1}{2}$. To check whether the important tail heaviness of the loss returns is driven by the crisis period, we also estimate γ in three subperiods: Before Crisis, from June 15th, 2004, to August 29th, 2008 (1,053 trading days); During Crisis, from September 2nd, 2008, to May 29th, 2009 (185 trading days), and After Crisis, from June 1st, 2009, to June 13th, 2014 (1,259 trading days). The plots of the estimates $k \mapsto \widehat{\gamma}_{1-k/n}$ and $k \mapsto \widetilde{\gamma}_{1-k/n}$ corresponding to these subperiods are depicted in Figure 1(b)-(d). The final estimates $\widehat{\gamma}_{1-k/n}$ and $\widetilde{\gamma}_{1-k/n}$, reported in Table 1 for each subperiod (from the third row), are chosen according to the automatic selection procedure described above (for the crisis period whose length is only 185 trading days, we used a moving window which covers around 20% of the possible values of k in $[1, n/\log n]$). The messages yielded by the two methods are broadly similar, giving evidence of a great variability in loss returns and a fat tail of their distribution in the three periods before, during and after crisis. When using weekly loss returns (corresponding to Wednesdays) in the same sample period, we obtain the plots in Figure 2 and the final estimates in Table 2. Given the length of the crisis period (38 trading weeks), we perform our extreme value estimation here only for the full period ($n = 516$), the pre-crisis period ($n = 219$) and the post-crisis period ($n = 259$).

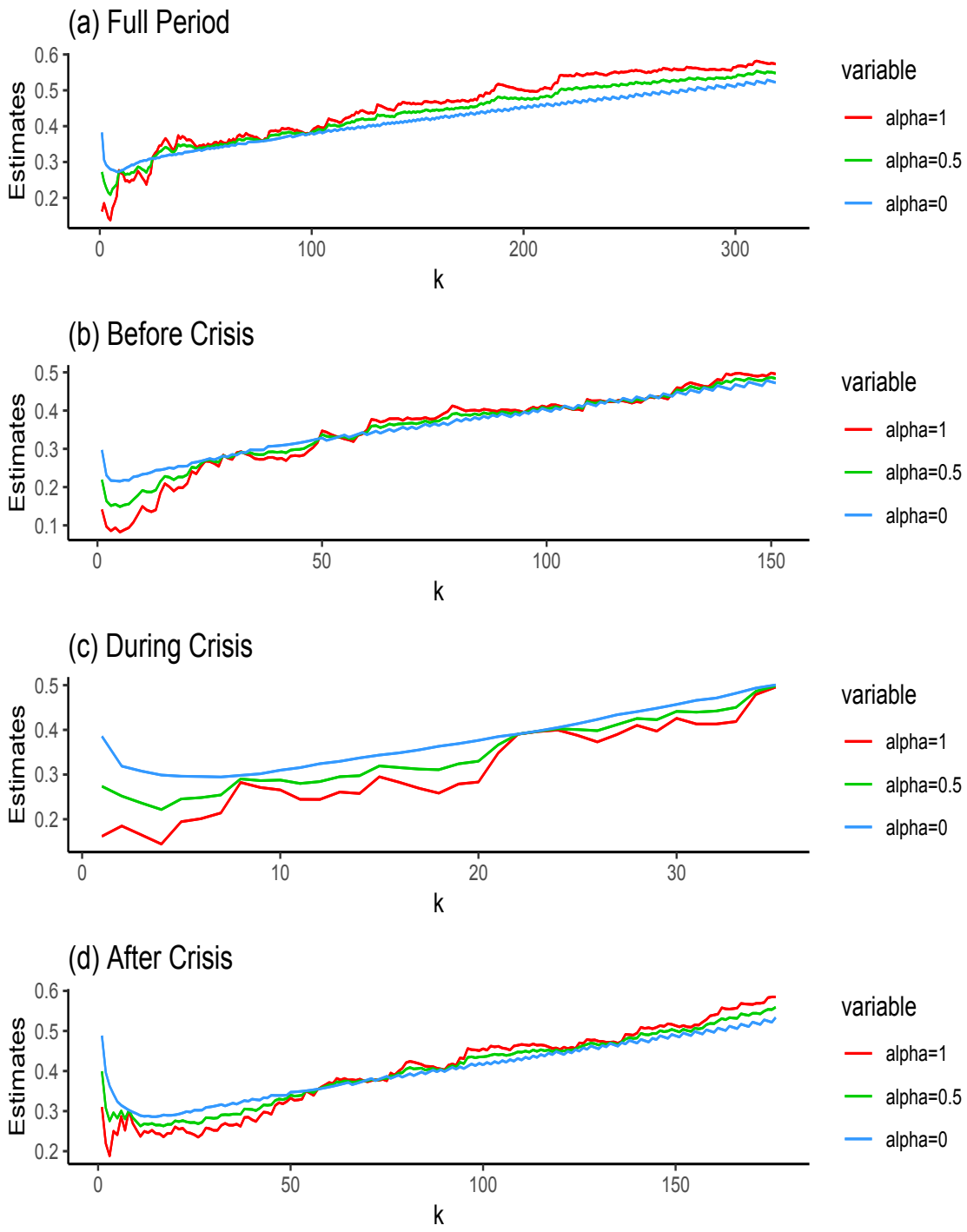


Figure 1: *Plots of* $\widehat{\gamma}_{1-k/n}$ *in red,* $\widetilde{\gamma}_{1-k/n}$ *in blue, and* $\overline{\gamma}_{1-k/n}(1/2)$ *in green, based on daily loss returns of the SPDR S&P 500 ETF (SPY), for the selected range of values* $1 \leqslant k \leqslant n/\log n$*. The estimates depicted in (a)-(d) correspond, respectively, to the full 10-years period (2004-2014) and the three sub-periods: Before Crisis (2004-2008), During Crisis (2008-2009) and After Crisis (2009-2014).*

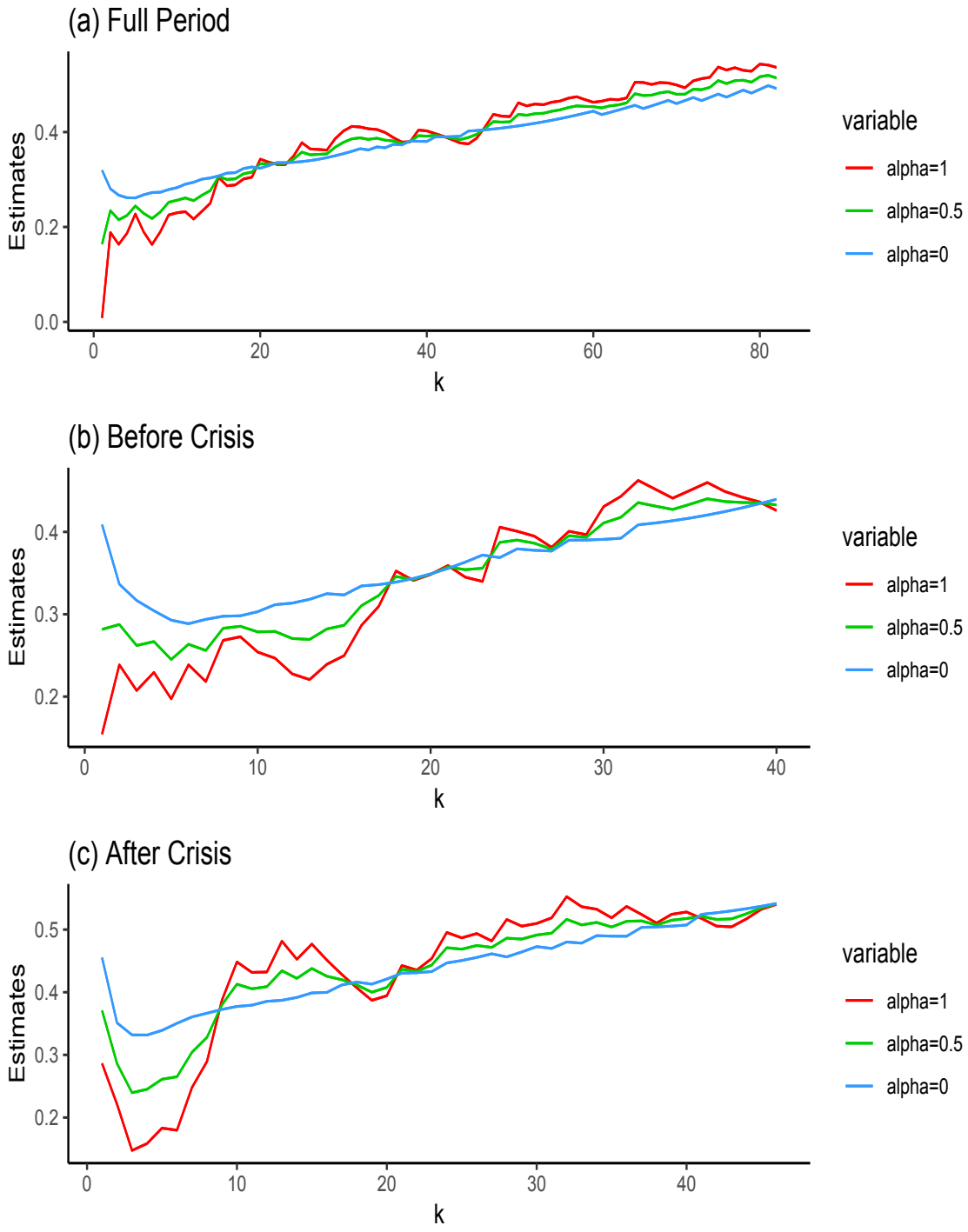


Figure 2: *Plots of* $\widehat{\gamma}_{1-k/n}$ *in red,* $\widetilde{\gamma}_{1-k/n}$ *in blue, and* $\overline{\gamma}_{1-k/n}(1/2)$ *in green, based on weekly loss returns of the SPDR S&P 500 ETF (SPY), for the selected range of values* $1 \leqslant k \leqslant n/\log n$. *The estimates depicted in (a)-(c) correspond, respectively, to the full 10-years period (2004-2014) and the two sub-periods: Before Crisis (2004-2008) and After Crisis (2009-2014).*

Period considered	n	$\widehat{\gamma}_{1-k/n}$ ($k \in$ window)	$\overline{\gamma}_{1-k/n}(\frac{1}{2})$ ($k \in$ window)	$\widetilde{\gamma}_{1-k/n}$ ($k \in$ window)
Full period	2,497	0.3585 ($57 \in [46, 60]$)	0.3437 ($53 \in [40, 54]$)	0.3344 ($47 \in [40, 54]$)
Before Crisis	1,053	0.2722 ($30 \in [29, 41]$)	0.2917 ($32 \in [28, 40]$)	0.3091 ($41 \in [34, 46]$)
During Crisis	185	0.2445 ($11 \in [9, 17]$)	0.2863 ($9 \in [8, 16]$)	0.2962 ($5 \in [3, 11]$)
After Crisis	1,259	0.2523 ($14 \in [12, 24]$)	0.2670 ($13 \in [12, 24]$)	0.2895 ($19 \in [10, 22]$)

Table 1: *Estimates of the tail index of the SPDR S&P 500 ETF (SPY), based on daily loss returns over the full 10-years period (2004-2014) and three sub-periods: Before Crisis (2004-2008), During Crisis (2008-2009) and After Crisis (2009-2014).*

Period considered	n	$\widehat{\gamma}_{1-k/n}$ ($k \in$ window)	$\overline{\gamma}_{1-k/n}(\frac{1}{2})$ ($k \in$ window)	$\widetilde{\gamma}_{1-k/n}$ ($k \in$ window)
Full period	516	0.3909 ($42 \in [30, 46]$)	0.3829 ($35 \in [30, 46]$)	0.3355 ($23 \in [15, 31]$)
Before Crisis	219	0.2182 ($7 \in [4, 12]$)	0.2705 ($12 \in [7, 15]$)	0.3031 ($10 \in [4, 12]$)
After Crisis	259	0.4316 ($11 \in [10, 26]$)	0.4199 ($17 \in [10, 26]$)	0.3997 ($16 \in [7, 23]$)

Table 2: *Estimates of the tail index based on weekly loss returns.*

In each considered period, the purely least asymmetrically weighted squares estimator $\widetilde{\gamma}_{1-k/n}$ seems to be beneficial in producing smoother and more stable and pleasing plots, but these plots may not be more revealing than Hill plots. Already in Figure 1(a)-(b), it may be seen that the smooth paths of $\widetilde{\gamma}_{1-k/n}$ can exhibit a sample-wise monotonic evolution with k. This may result in estimates with higher bias than the Hill estimates. One way to reduce this potential defect is by using a linear combination of $\widetilde{\gamma}$ and $\widehat{\gamma}$ for estimating γ. For $\alpha \in \mathbb{R}$ we have then defined the more general *expectHill* estimator

$$\overline{\gamma}_{\tau_n}(\alpha) = \alpha\widehat{\gamma}_{\tau_n} + (1-\alpha)\widetilde{\gamma}_{\tau_n}.$$

For example, as visualized in Figure 1, the simple mean $\overline{\gamma}_{\tau_n}(1/2)$ in green line would represent

a reasonable compromise between the use of large asymmetric least squares in $\widetilde{\gamma}_{\tau_n}$ and top order statistics in $\widehat{\gamma}_{\tau_n}$. Also, Figure 3 shows that the mean $\overline{\gamma}_{\tau_n}(1/2)$ affords a middle course between $\widehat{\gamma}_{\tau_n} \equiv \overline{\gamma}_{\tau_n}(1)$ and $\widetilde{\gamma}_{\tau_n} \equiv \overline{\gamma}_{\tau_n}(0)$ in terms of asymptotic variance.

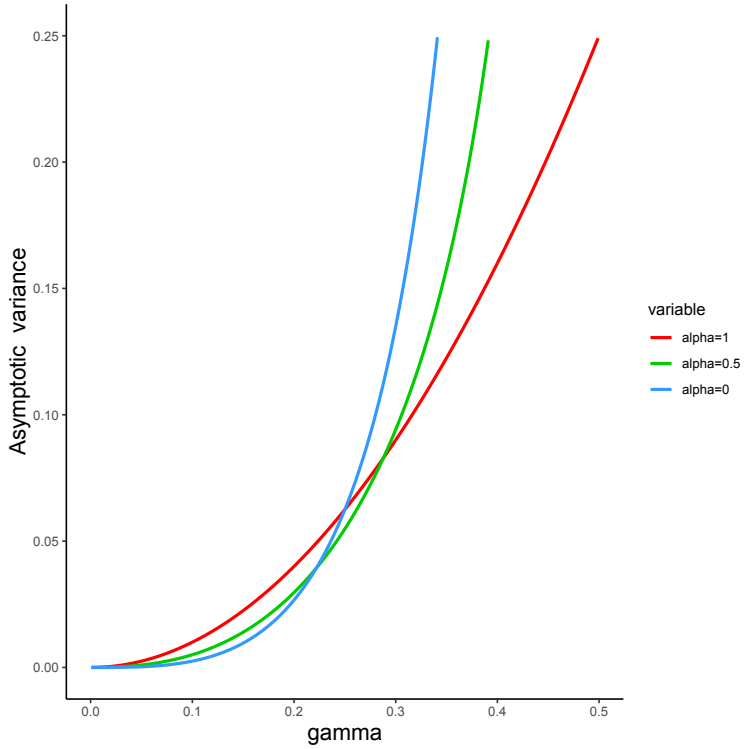


Figure 3: *Asymptotic variance* v_α *of* $\widehat{\gamma}_{\tau_n}$ *in red* ($\alpha = 1$), $\widetilde{\gamma}_{\tau_n}$ *in blue* ($\alpha = 0$), *and* $\overline{\gamma}_{\tau_n}(1/2)$ *in green* ($\alpha = 0.5$), *as functions of* $\gamma \in (0, 1/2)$.

The important question of how to select in practice the combination parameter α is discussed in Section 6 of the main paper and implemented in Section 8 through the same empirical data.

B Some simulation evidence

The aim of this section is to explore some features that were mentioned in Section 7 of the main article. We will illustrate the following points:

(B.1) Estimates of γ.

(B.2) Estimates of $\xi_{\tau'_n}$.

(B.3) Estimates of $\mathrm{XES}_{\tau'_n}$.

(B.4) Estimates of QES_{p_n}.

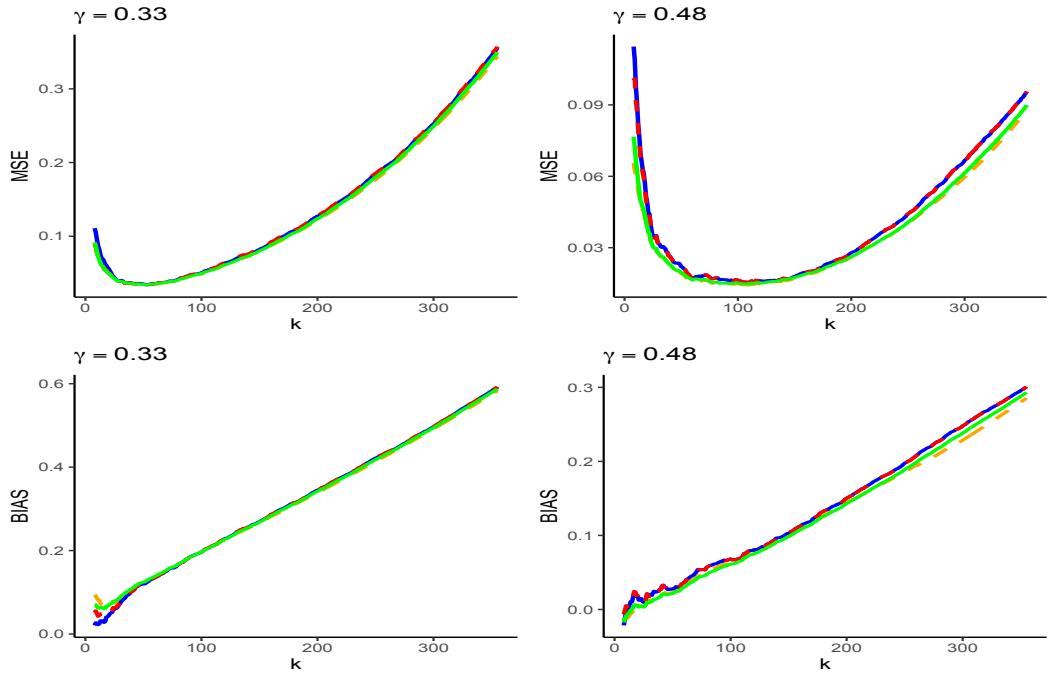
In order to illustrate the behavior of the presented estimation procedures, we use the same considerations as in Section 7 of the main paper. Namely, we consider the Student t-distribution with $1/\gamma$ degrees of freedom, the Fréchet distribution $F(x) = e^{-x^{-1/\gamma}},\ x > 0$, and the Pareto distribution $F(x) = 1 - x^{-1/\gamma},\ x > 1$. The finite-sample performance of the different estimators is evaluated through their relative Mean-Squared Error (MSE) and bias, computed over 200 replications. All the experiments have sample size $n = 2{,}500$ and true tail index $\gamma \in \{0.33, 0.48\}$. In our simulations we used the extreme levels $\tau'_n = p_n = 1 - \frac{1}{n}$ and the intermediate level $\tau_n = 1 - \frac{k}{n}$, where the integer k can be viewed as the effective sample size for tail extrapolation.

B.1 Estimation of the tail index

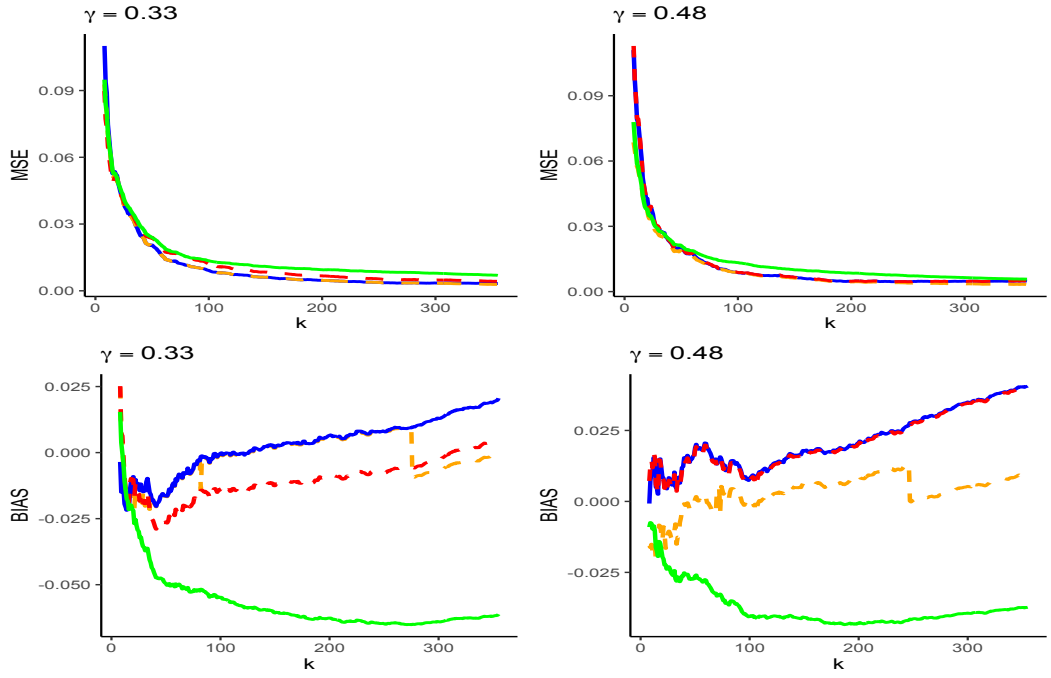
This section provides Monte Carlo evidence that the two-step expectHill estimator $\overline{\gamma}_{1-k/n}$, as introduced in (19), has a very similar accuracy to the optimal version $\overline{\gamma}_{1-k/n}(\alpha(\gamma))$. Both estimators have the common linear form described in (8), with the combination parameter being $\alpha = \overline{\alpha}_{1-k/n} := \alpha(\overline{\gamma}_{1-k/n}(\frac{1}{2}))$ in the former estimator and the true theoretical optimal weight $\alpha(\gamma)$ in the latter. These empirical and theoretical 'variance-optimal' expectHill estimators were compared with the 'oracle' expectHill estimator $\overline{\gamma}_{1-k/n}(\alpha_{\text{oracle}})$ obtained by selecting the value of α which minimizes its MSE, as well as the 'hybrid' version $\overline{\gamma}_{1-k/n}(1/2)$ obtained with the average weight $\alpha = 1/2$. In the case of Student distribution, Figure 4(a) gives the evolution of the MSE (in top panels) and the bias (in bottom panels) of $\overline{\gamma}_{1-k/n}/\gamma$ (dashed red line), $\overline{\gamma}_{1-k/n}(\alpha(\gamma))/\gamma$ (solid blue line), $\overline{\gamma}_{1-k/n}(\alpha_{\text{oracle}})/\gamma$ (dashed orange line) and $\overline{\gamma}_{1-k/n}(1/2)/\gamma$ (solid green line), as functions of the effective sample fraction k. It is remarkable that the three plots of the resulting 'variance-optimal' and 'hybrid' estimates (in red, blue and green) are overall almost identical in this case. Most importantly, they are extremely close to the benchmark plot (in orange) of the 'oracle' estimate in terms of both MSE and bias. The results obtained in the cases of Fréchet and Pareto distributions are displayed in Figures 4(b) and 5. Here also the MSE estimates are still very close for the 'variance-optimal' and 'oracle' estimators (in red, blue and orange), with a slight advantage in terms of bias for the 'oracle' estimator in the case $\gamma = 0.48$. However, the 'hybrid' version (in green), in contrast to the Student scenario, deteriorates slightly in terms of MSE and appreciably in terms of bias.

B.2 Estimates of $\xi_{\tau'_n}$

The simulation experiments undertaken in this section are concerned with the two-step estimator $\overline{\xi}^{\star}_{\tau'_n} := \overline{\xi}^{\star}_{\tau'_n}(\alpha, \beta)$ of the extreme expectile $\xi_{\tau'_n}$. It is defined in (21) by substituting in the



(a) Student distribution



(b) Fréchet distribution

Figure 4: *MSE estimates (top panels) and Bias estimates (bottom panels) of* $\overline{\gamma}_{1-k/n}/\gamma$ *(dashed red line),* $\overline{\gamma}_{1-k/n}(\alpha(\gamma))/\gamma$ *(solid blue line),* $\overline{\gamma}_{1-k/n}(\alpha_{oracle})/\gamma$ *(dashed orange line) and* $\overline{\gamma}_{1-k/n}(1/2)/\gamma$ *(solid green line), as functions of* k*, for* $\gamma = 0.33$ *(left) and* $\gamma = 0.48$ *(right).*

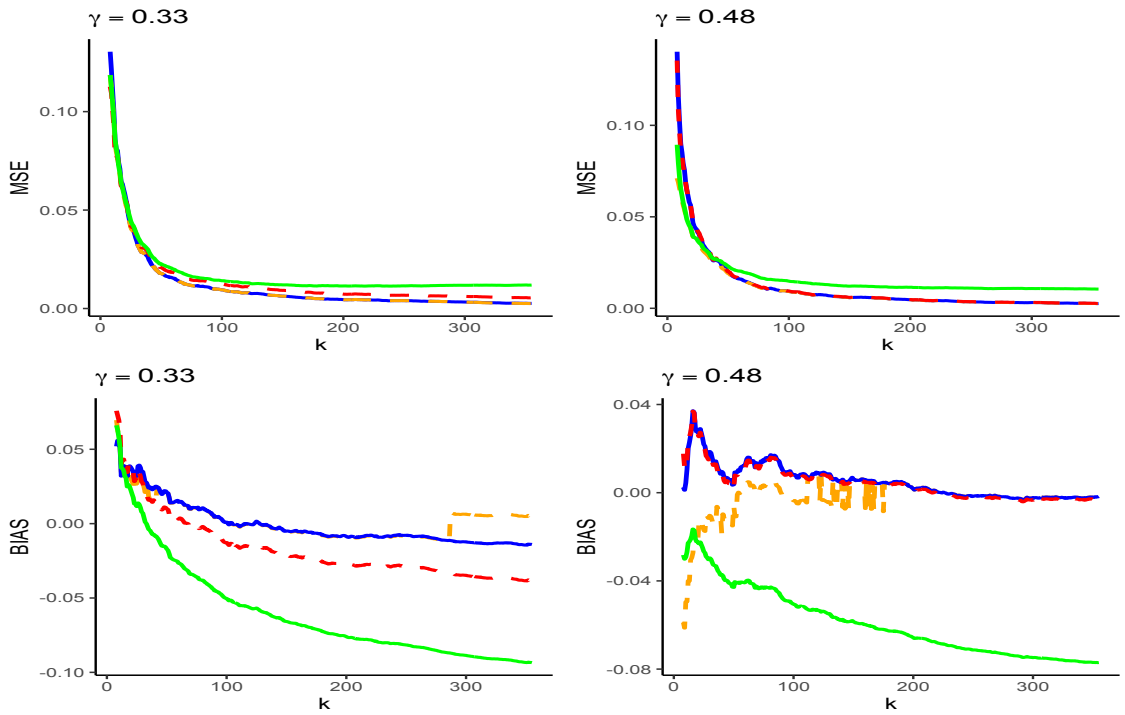


Figure 5: *MSE estimates (top panels) and Bias estimates (bottom panels) of* $\overline{\gamma}_{1-k/n}/\gamma$ *(dashed red line),* $\overline{\gamma}_{1-k/n}(\alpha(\gamma))/\gamma$ *(solid blue line),* $\overline{\gamma}_{1-k/n}(\alpha_{oracle})/\gamma$ *(dashed orange line) and* $\overline{\gamma}_{1-k/n}(1/2)/\gamma$ *(solid green line), as functions of* k*, for* $\gamma = 0.33$ *(left) and* $\gamma = 0.48$ *(right), in the case of a Pareto distribution.*

estimated values $\alpha = \overline{\alpha}_{\tau_n}$ and $\beta = \overline{\beta}_{\tau_n}$ of the theoretical optimal weights $\alpha(\gamma)$ and $\beta_{\alpha(\gamma)}(\gamma)$, as described in Section 6. Its accuracy is evaluated in comparison with the variance-optimal version $\overline{\xi}^{\star}_{\tau'_n}(\alpha(\gamma), \beta_{\alpha(\gamma)}(\gamma))$ that is obtained by replacing the combination parameters α and β with their true optimal values $\alpha(\gamma)$ and $\beta_{\alpha(\gamma)}(\gamma)$. We also considered the 'hybrid' version of $\overline{\xi}^{\star}_{\tau'_n}(\alpha, \beta)$ which corresponds to the average weights $\alpha = \beta = 1/2$, and the 'oracle' benchmark obtained by selecting the values of α and β which minimize the MSE of $\overline{\xi}^{\star}_{\tau'_n}(\alpha, \beta)$.

Figures 6 and 7 display the Monte Carlo estimates for $\overline{\xi}^{\star}_{\tau'_n}(\alpha(\gamma), \beta_{\alpha(\gamma)}(\gamma))/\xi_{\tau'_n}$ in blue, $\overline{\xi}^{\star}_{\tau'_n}/\xi_{\tau'_n}$ in red, $\overline{\xi}^{\star}_{\tau'_n}(1/2, 1/2)/\xi_{\tau'_n}$ in green, and $\overline{\xi}^{\star}_{\tau'_n}(\alpha_{\text{oracle}}, \beta_{\text{oracle}})/\xi_{\tau'_n}$ in orange. Clearly, the use of the estimated values $\alpha = \overline{\alpha}_{\tau_n}$ and $\beta = \overline{\beta}_{\tau_n}$ provides, in all cases, very similar results, in terms of both MSE and bias, to the true optimal weights $\alpha(\gamma)$ and $\beta_{\alpha(\gamma)}(\gamma)$ themselves. Most importantly, the MSE and bias estimates based on our variance-optimal selection of weights are also very similar to those based on the oracle selection device over the range of intermediate values, say, $k \leqslant 100$. Their behavior outside this range of k values remains quite good in comparison with the oracle estimates. As for the hybrid version $\overline{\xi}^{\star}_{\tau'_n}(1/2, 1/2)$, it exhibits a respectable evolution with k relative to both variance-optimal and oracle estimates, with a slightly better bias relative to the variance-optimal estimates.

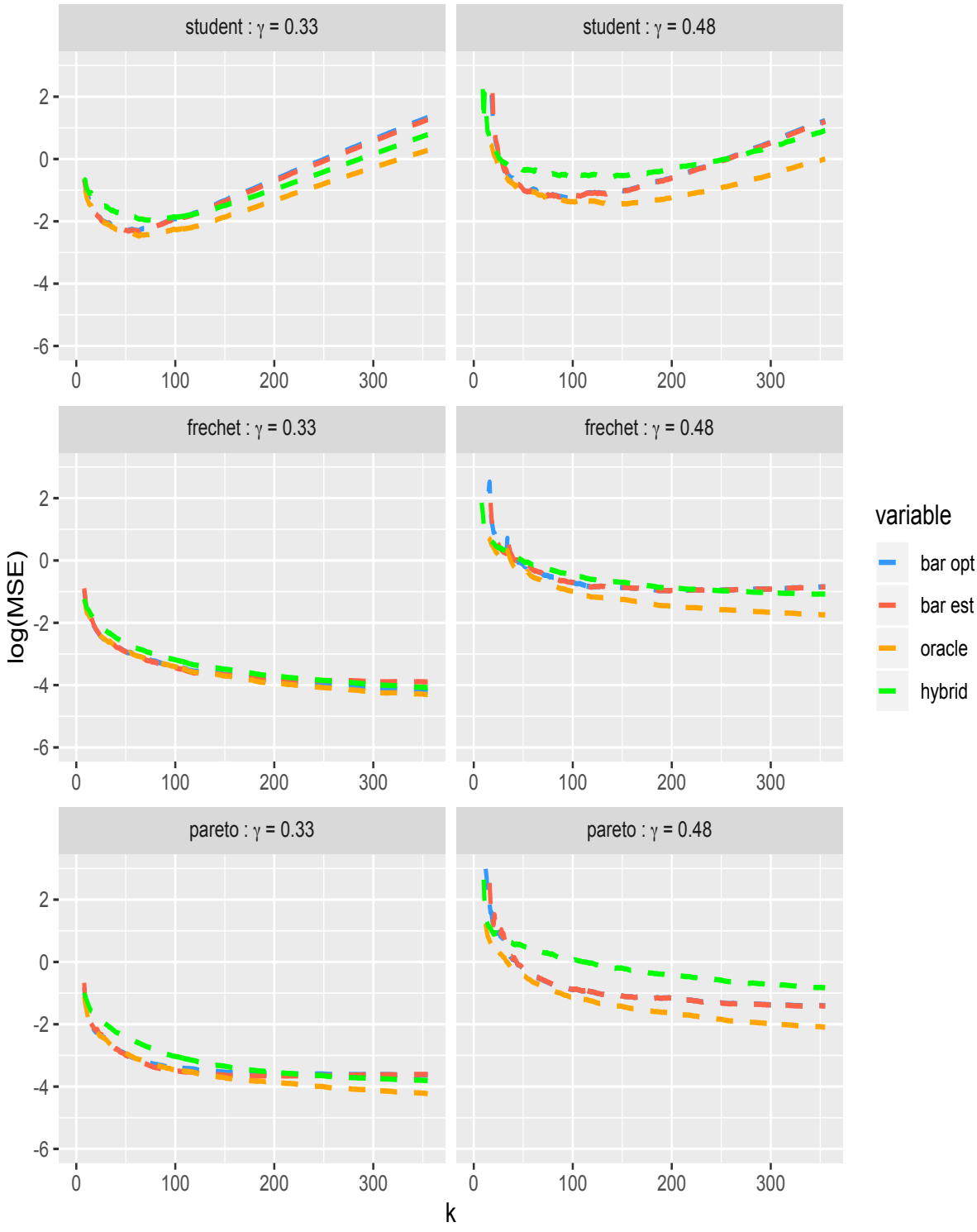


Figure 6: *MSE estimates (in log scale) against k, for Student (top), Fréchet (middle) and Pareto (bottom) distributions, with $\gamma = 0.33$ (left) and $\gamma = 0.48$ (right). Results for $\overline{\xi}^{\star}_{\tau'_n}(\alpha(\gamma), \beta_{\alpha(\gamma)}(\gamma))/\xi_{\tau'_n}$ in blue, $\overline{\xi}^{\star}_{\tau'_n}/\xi_{\tau'_n}$ in red, $\overline{\xi}^{\star}_{\tau'_n}(1/2, 1/2)/\xi_{\tau'_n}$ in green, and $\overline{\xi}^{\star}_{\tau'_n}(\alpha_{oracle}, \beta_{oracle})/\xi_{\tau'_n}$ in orange.*

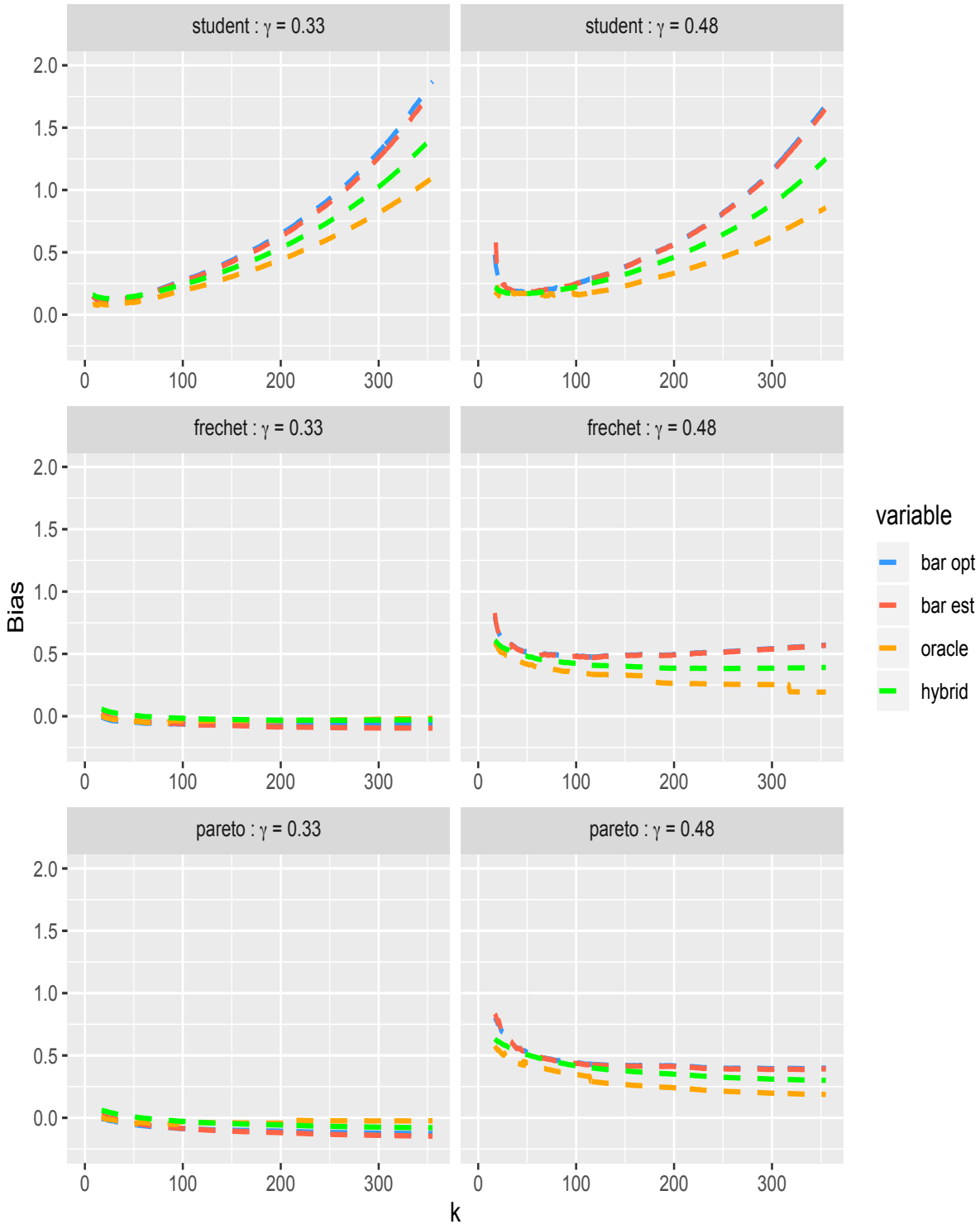


Figure 7: *Bias estimates in the same setting as in Figure 6.*

B.3 Estimates of $\mathrm{XES}_{\tau'_n}$

We have also investigated the accuracy of the two-step $\mathrm{XES}_{\tau'_n}$ estimators $\widetilde{\mathrm{XES}}^{\star}_{\tau'_n} := \widetilde{\mathrm{XES}}^{\star}_{\tau'_n}(\alpha)$ and $\overline{\mathrm{XES}}^{\star}_{\tau'_n} := \overline{\mathrm{XES}}^{\star}_{\tau'_n}(\alpha,\beta)$, introduced respectively in (22) and (23). They are computed by substituting in the estimated values $\alpha = \overline{\alpha}_{\tau_n}$ and $\beta = \overline{\beta}_{\tau_n}$ of the optimal weights $\alpha(\gamma)$ and $\beta_{\alpha(\gamma)}(\gamma)$. We also compared these competing estimators with their best versions $\widetilde{\mathrm{XES}}^{\star}_{\tau'_n}(\alpha(\gamma))$ and $\overline{\mathrm{XES}}^{\star}_{\tau'_n}(\alpha(\gamma),\beta_{\alpha(\gamma)}(\gamma))$ that are obtained by substituting in the true optimal weights $\alpha = \alpha(\gamma)$ and $\beta = \beta_{\alpha(\gamma)}(\gamma)$.

Figures 8 and 9 give the evolution of the MSE (in log scale) and bias estimates of $\overline{\mathrm{XES}}^{\star}_{\tau'_n}(\alpha(\gamma),\beta_{\alpha(\gamma)}(\gamma))/\mathrm{XES}_{\tau'_n}$ in blue, $\overline{\mathrm{XES}}^{\star}_{\tau'_n}/\mathrm{XES}_{\tau'_n}$ in red, $\widetilde{\mathrm{XES}}^{\star}_{\tau'_n}(\alpha(\gamma))/\mathrm{XES}_{\tau'_n}$ in violet, and $\widetilde{\mathrm{XES}}^{\star}_{\tau'_n}/\mathrm{XES}_{\tau'_n}$ in green. Here also, the two-step estimators (in red, respectively, green) have very similar results, in terms of both MSE and bias, to their optimal versions (in blue, respectively, violet) in all cases.

Moreover, in the case of Student distribution (top panels), it may be seen that the green and violet curves, which correspond to the estimates $\widetilde{\mathrm{XES}}^{\star}_{\tau'_n}$, perform better than the red and blue curves that correspond to the estimates $\overline{\mathrm{XES}}^{\star}_{\tau'_n}$. By contrast, in the cases of Fréchet and Pareto distributions (middle and bottom panels), the situation is reversed as the estimates $\overline{\mathrm{XES}}^{\star}_{\tau'_n}$ seem to be superior to their competitors $\widetilde{\mathrm{XES}}^{\star}_{\tau'_n}$ in terms of both MSE and bias.

B.4 Estimates of QES_{p_n}

We have also undertaken simulation experiments to evaluate the finite-sample performance of the composite versions $\widetilde{\mathrm{XES}}^{\star}_{\widehat{\tau}'_n(p_n)} := \widetilde{\mathrm{XES}}^{\star}_{\widehat{\tau}'_n(p_n)}(\alpha)$ in (24) and $\overline{\mathrm{XES}}^{\star}_{\widehat{\tau}'_n(p_n)} := \overline{\mathrm{XES}}^{\star}_{\widehat{\tau}'_n(p_n)}(\alpha,\beta)$ in (25), with $\alpha = \overline{\alpha}_{\tau_n}$ and $\beta = \overline{\beta}_{\tau_n}$, which estimate the same conventional expected shortfall QES_{p_n} as the purely quantile-based estimator $\widehat{\mathrm{QES}}^{\star}_{p_n}$ in (26).

Figures 10 and 11 give the MSE (in log scale) and bias estimates of $\overline{\mathrm{XES}}^{\star}_{\widehat{\tau}'_n(p_n)}/\mathrm{QES}_{p_n}$ in red, $\widetilde{\mathrm{XES}}^{\star}_{\widehat{\tau}'_n(p_n)}/\mathrm{QES}_{p_n}$ in green, and $\widehat{\mathrm{QES}}^{\star}_{p_n}/\mathrm{QES}_{p_n}$ in yellow. We superimpose in the same figures the MSE and bias estimates of the optimal versions $\overline{\mathrm{XES}}^{\star}_{\widehat{\tau}'_n(p_n)}(\alpha(\gamma),\beta_{\alpha(\gamma)}(\gamma))/\mathrm{QES}_{p_n}$ in blue and $\widetilde{\mathrm{XES}}^{\star}_{\widehat{\tau}'_n(p_n)}(\alpha(\gamma))/\mathrm{QES}_{p_n}$ in violet. The results indicate that both two-step estimators $\widetilde{\mathrm{XES}}^{\star}_{\widehat{\tau}'_n(p_n)}$ and $\overline{\mathrm{XES}}^{\star}_{\widehat{\tau}'_n(p_n)}$ are very accurate with respect to their optimal versions. They also suggest the superiority of the former estimator (green plots) in the case of Student distribution, and of the latter (red plots) in the other cases.

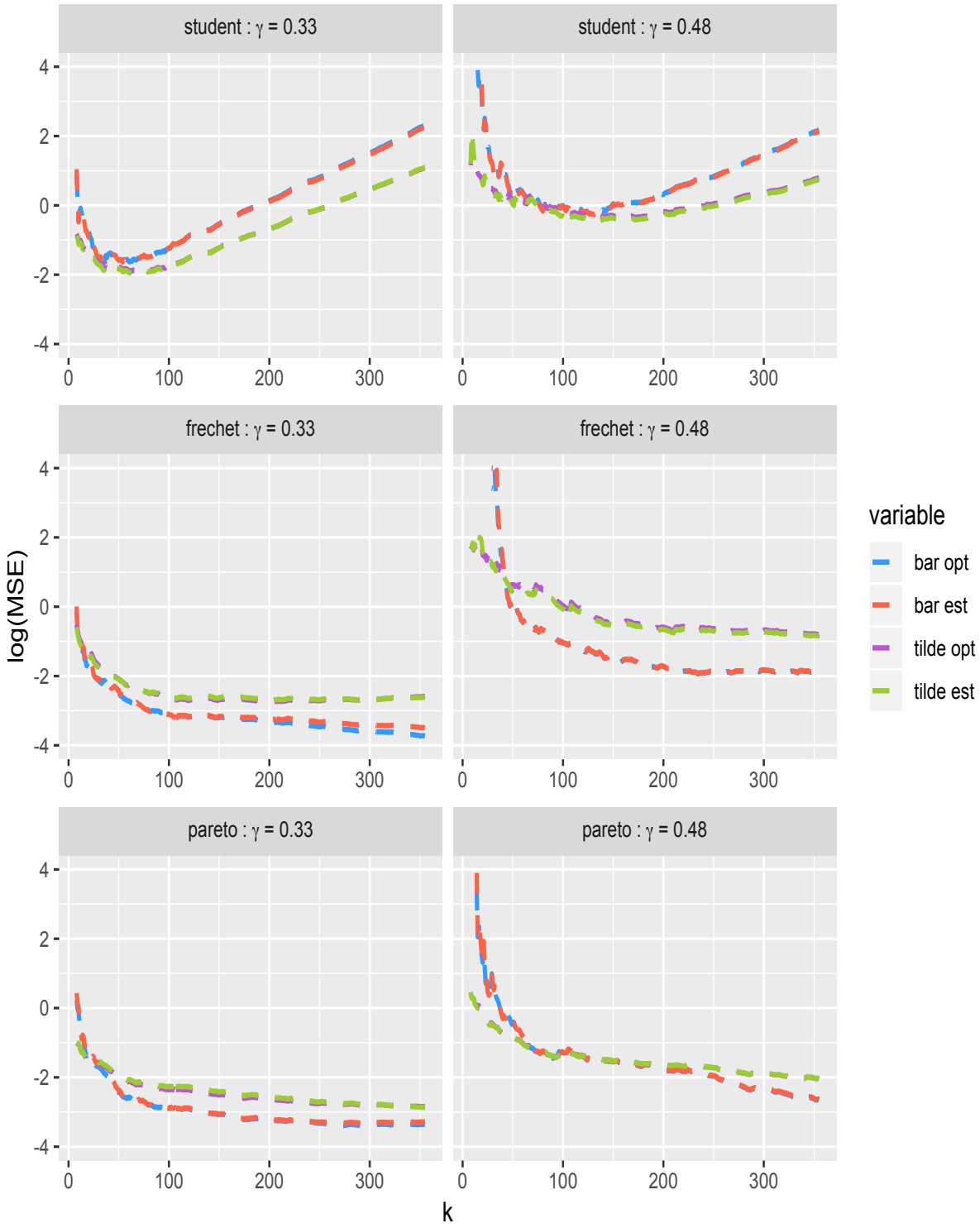


Figure 8: *MSE estimates (in log scale) against k, for Student (top), Fréchet (middle) and Pareto (bottom) distributions, with $\gamma = 0.33$ (left) and $\gamma = 0.48$ (right). Results for $\overline{XES}^{\star}_{\tau'_n}(\alpha(\gamma), \beta_{\alpha(\gamma)}(\gamma))$ in blue, $\overline{XES}^{\star}_{\tau'_n}$ in red, $\widetilde{XES}^{\star}_{\tau'_n}(\alpha(\gamma))$ in violet, and $\widetilde{XES}^{\star}_{\tau'_n}$ in green.*

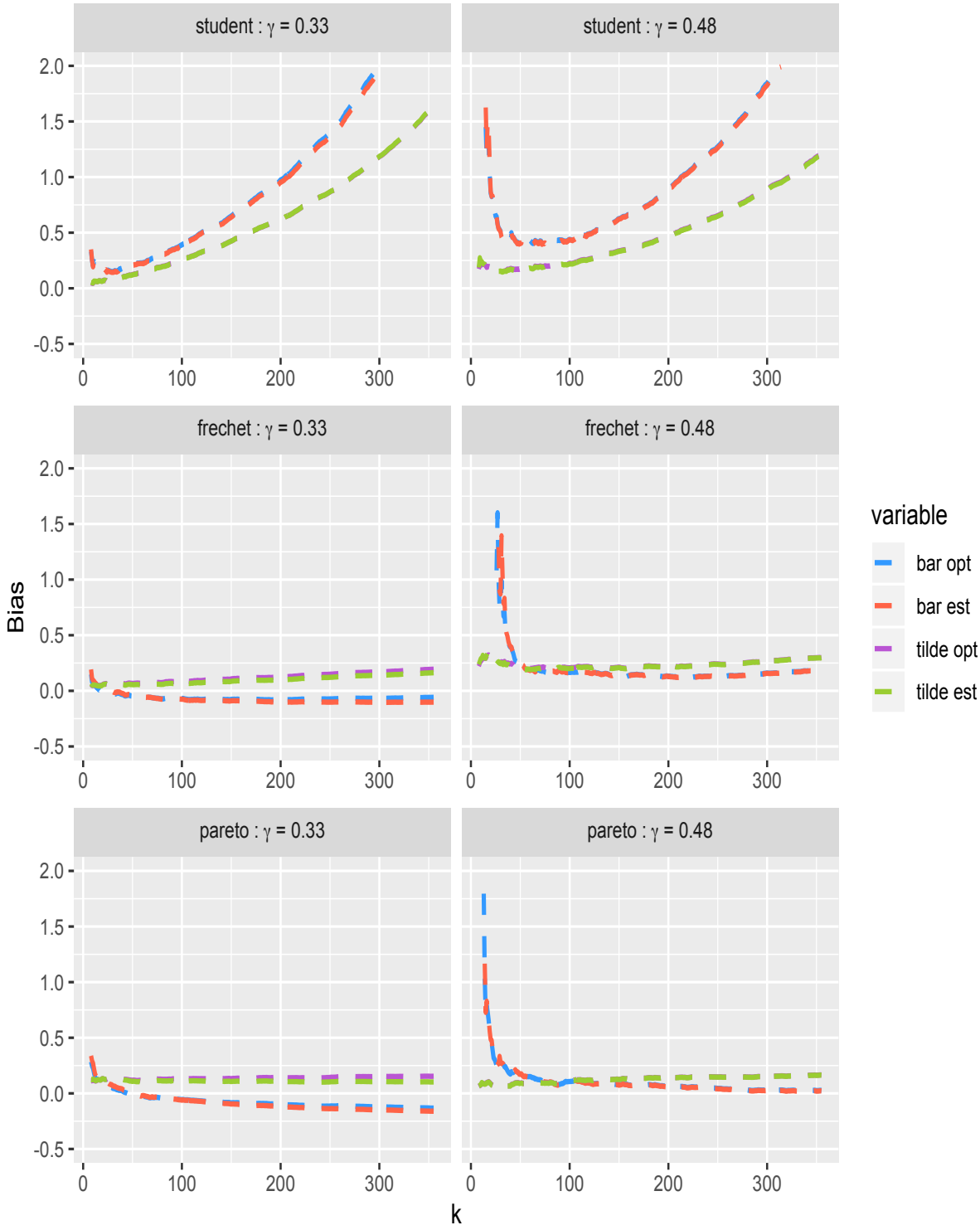


Figure 9: *Bias estimates in the same setting as in Figure 8.*

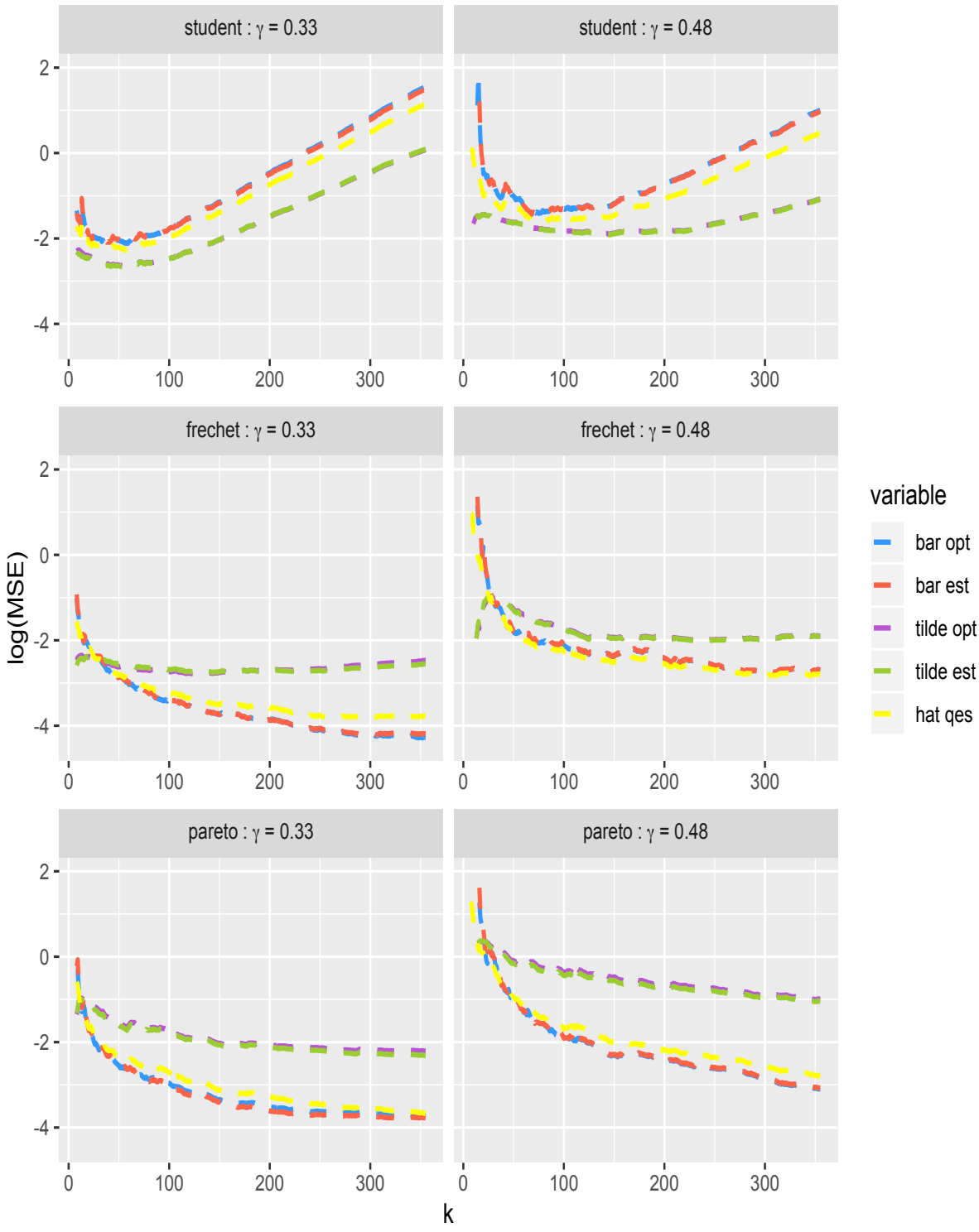


Figure 10: *MSE estimates (in log scale) against k, for Student (top), Fréchet (middle) and Pareto (bottom) distributions, with $\gamma = 0.33$ (left) and $\gamma = 0.48$ (right). Results for $\overline{\mathrm{XES}}^{\star}_{\widehat{\tau}'_n(p_n)}(\alpha(\gamma), \beta_{\alpha(\gamma)}(\gamma))$ in blue, $\overline{\mathrm{XES}}^{\star}_{\widehat{\tau}'_n(p_n)}$ in red, $\widetilde{\mathrm{XES}}^{\star}_{\widehat{\tau}'_n(p_n)}(\alpha(\gamma))$ in violet, $\widetilde{\mathrm{XES}}^{\star}_{\widehat{\tau}'_n(p_n)}$ in green, and $\widehat{QES}^{\star}_{p_n}$ in yellow.*

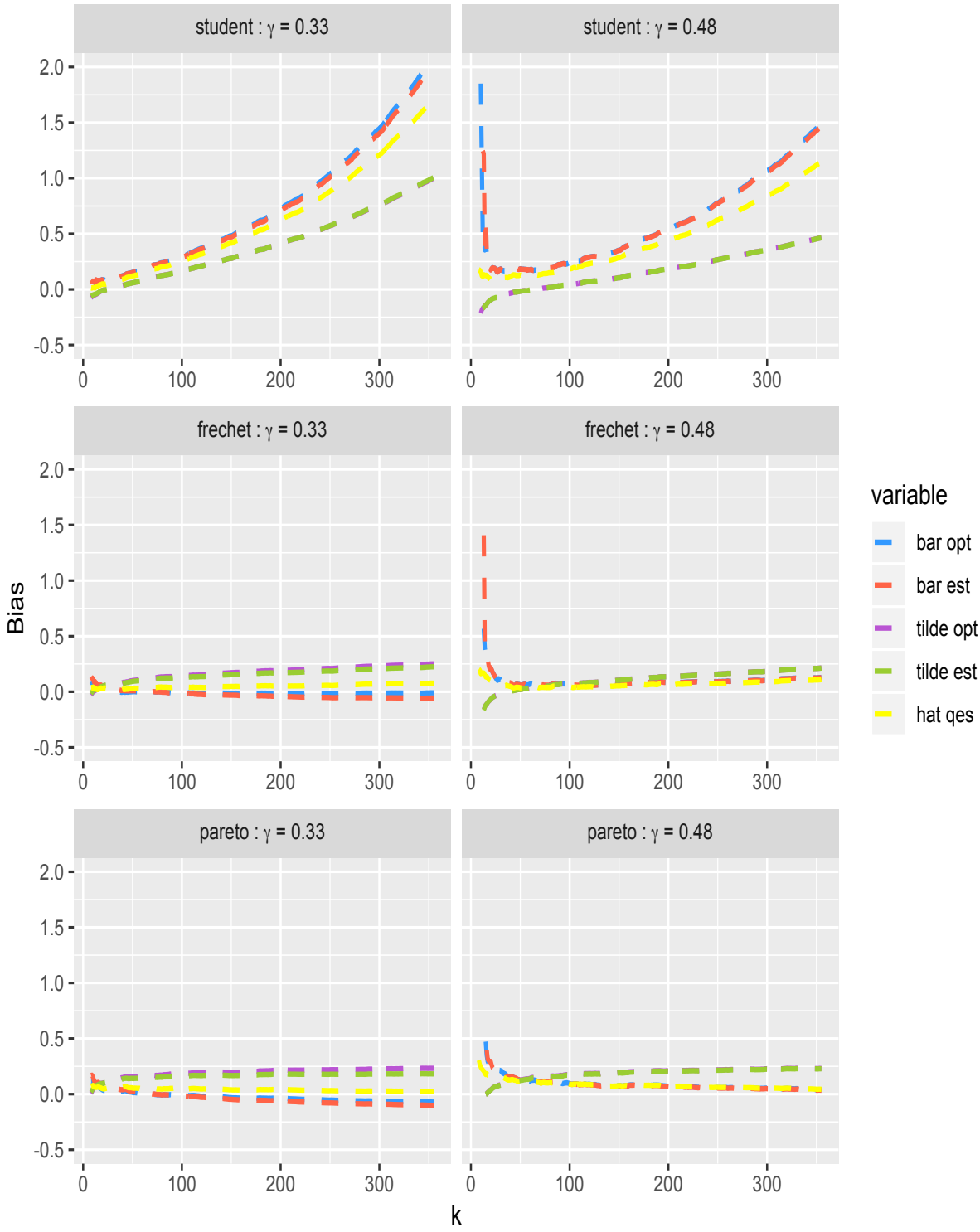


Figure 11: *Bias estimates in the same setting as in Figure 10.*

C Proofs

In all proofs, the sequence τ_n is replaced by the sequence $k = n(1-\tau_n)$.

Proof of Theorem 1. To show (i), the main idea is to combine Proposition 1 with a Taylor expansion of the logarithm function. This is not quite as straightforward as one might expect, because the error term in the approximation of the tail empirical expectile process given by Proposition 1 does not converge to 0 uniformly in s. The trick we use here is to split the integral defining $\breve{\gamma}_k$ in two parts, corresponding to "low" and "high" values of s respectively; we then show directly that the first part is asymptotically negligible, and we analyse the second part using the aforementioned Taylor expansion. A similar argument is used in *e.g.* page 113 of de Haan and Ferreira (2006) and El Methni and Stupfler (2017, 2018). Let us finally mention that to use Proposition 1, we should work with a suitable version of the tail expectile process that allows us to write its Gaussian approximation; we can of course do so since this operation leaves the distribution of the estimator $\breve{\gamma}_k$ unchanged. A similar idea will be used, without further mention, in the proof of Theorem 2.

Set then $s_n = k^{-(1-\varepsilon)/(1+2\varepsilon)}$ for some $\varepsilon > 0$ sufficiently small (and in particular less than $1/4$), and write

$$\breve{\gamma}_k = \int_0^{s_n} \log\left(\frac{\widetilde{\xi}_{1-ks/n}}{\widetilde{\xi}_{1-k/n}}\right) ds + \int_{s_n}^1 \log\left(\frac{\widetilde{\xi}_{1-ks/n}}{\widetilde{\xi}_{1-k/n}}\right) ds =: I_{n,1} + I_{n,2}. \tag{B.1}$$

We start by controlling directly $I_{n,1}$. This is done by writing

$$|I_{n,1}| \leqslant s_n \log\left(\frac{\widetilde{\xi}_1}{\widetilde{\xi}_{1-k/n}}\right).$$

Recall that $\widetilde{\xi}_1 = Y_{n,n}$ and use a combination of convergence $\xi_\tau/q_\tau \sim (\gamma^{-1}-1)^{-\gamma}$ as $\tau \to 1$ and Lemmas 2(i) and 3 in Daouia *et al.* (2020) to find that

$$\log\left(\frac{\widetilde{\xi}_1}{\widetilde{\xi}_{1-k/n}}\right) = \log\left(\frac{Y_{n,n}}{\widehat{q}_{1-k/n}}\right) + \mathrm{O}_{\mathbb{P}}(1).$$

Using further the heavy-tailed assumption on the distribution on Y, it follows from Theorem 1.1.6, Theorem 1.2.1 and Lemma 1.2.9 in de Haan and Ferreira (2006) that

$$\frac{Y_{n,n}}{U(n)} \xrightarrow{d} 1 + \gamma G_\gamma$$

where G_γ has distribution function $x \mapsto \exp(-(1+\gamma x)^{-1/\gamma})$, for $x > -1/\gamma$. It follows that the limiting variable $1+\gamma G_\gamma$ is positive and thus $\log(Y_{n,n}/U(n)) = \mathrm{O}_{\mathbb{P}}(1)$ by the continuous

mapping theorem. Besides, $\widehat{q}_{1-k/n}/U(n/k) = \widehat{q}_{1-k/n}/q_{1-k/n} \xrightarrow{\mathbb{P}} 1$, by Lemma 2(i) in Daouia *et al.* (2020) again. Therefore

$$\log\left(\frac{\widetilde{\xi}_1}{\widetilde{\xi}_{1-k/n}}\right) = \log\left(\frac{U(n)}{U(n/k)}\right) + \mathrm{O}_{\mathbb{P}}(1).$$

Potter bounds (see *e.g.* Proposition B.1.9.5 in de Haan and Ferreira, 2006) then yield

$$\log\left(\frac{\widetilde{\xi}_1}{\widetilde{\xi}_{1-k/n}}\right) = \mathrm{O}_{\mathbb{P}}(\log k). \tag{B.2}$$

Recalling that $s_n = k^{-(1-\varepsilon)/(1+2\varepsilon)}$ with $\varepsilon < 1/4$, it is now straightforward to get

$$\sqrt{k}|I_{n,1}| = \mathrm{O}_{\mathbb{P}}\left(s_n \times \sqrt{k}\log k\right) = \mathrm{o}_{\mathbb{P}}(1). \tag{B.3}$$

We now work on $I_{n,2}$. Note that for $s \in [s_n, 1]$, $s^{-1/2-\varepsilon}/\sqrt{k} \leqslant s_n^{-1/2-\varepsilon}/\sqrt{k} = k^{-\varepsilon/2} \to 0$; use then Proposition 1 and a Taylor expansion of the logarithm function to obtain

$$\begin{aligned}
I_{n,2} &= -\gamma\int_{s_n}^1 \log(s)\,ds \\
&+ \frac{1}{\sqrt{k}}\gamma^2\sqrt{\gamma^{-1}-1}\left(\int_{s_n}^1 s^{\gamma-1}\left[\int_0^s W_n(t)\,t^{-\gamma-1}\,dt\right]ds - (1-s_n)\int_0^1 W_n(t)\,t^{-\gamma-1}\,dt\right) \\
&+ \frac{\gamma(\gamma^{-1}-1)^{\gamma}}{q_{1-k/n}}\left(\mathbb{E}(Y)\int_{s_n}^1 (s^{\gamma}-1)\,ds + \mathrm{o}_{\mathbb{P}}(1)\right) \\
&+ \frac{(1-\gamma)(\gamma^{-1}-1)^{-\rho}}{1-\gamma-\rho}\left(\int_{s_n}^1 \frac{s^{-\rho}-1}{\rho}\,ds\right)A(n/k) + \mathrm{o}_{\mathbb{P}}\left(\frac{1}{\sqrt{k}}\right).
\end{aligned}$$

Since $s_n = k^{-(1-\varepsilon)/(1+2\varepsilon)}$, we find that

$$\sqrt{k}\left|\int_0^1 \log(s)\,ds - \int_{s_n}^1 \log(s)\,ds\right| = \mathrm{O}\left(k^{(-1/2+2\varepsilon)/(1+2\varepsilon)}\log k\right) \to 0.$$

Using again the fact that $s_n \to 0$, along with the conditions $1/q_{1-k/n} = \mathrm{O}_{\mathbb{P}}(1/\sqrt{k})$ and $A(n/k) = \mathrm{O}_{\mathbb{P}}(1/\sqrt{k})$, we get

$$\begin{aligned}
I_{n,2} &= \gamma + \frac{1}{\sqrt{k}}\gamma^2\sqrt{\gamma^{-1}-1}\left(\int_0^1 s^{\gamma-1}\left[\int_0^s W_n(t)\,t^{-\gamma-1}\,dt\right]ds - \int_0^1 W_n(t)\,t^{-\gamma-1}\,dt\right) \\
&- \mathbb{E}(Y)\frac{\gamma^2(\gamma^{-1}-1)^{\gamma}}{\gamma+1}\frac{1}{q_{1-k/n}} + \frac{(1-\gamma)(\gamma^{-1}-1)^{-\rho}}{(1-\rho)(1-\gamma-\rho)}A(n/k) + \mathrm{o}_{\mathbb{P}}\left(\frac{1}{\sqrt{k}}\right).
\end{aligned}$$

By an integration by parts (with the inner integral being differentiated as a function of s), we obtain

$$\int_0^1 s^{\gamma-1}\left[\int_0^s W_n(t)\,t^{-\gamma-1}\,dt\right]ds = \frac{1}{\gamma}\int_0^1 \frac{W_n(s)}{s}(s^{-\gamma}-1)ds$$

and therefore, denoting by

$$Z = \gamma\sqrt{\gamma^{-1}-1}\int_0^1 \frac{W(s)}{s}([1-\gamma]s^{-\gamma}-1)ds$$

where W is a standard Brownian motion, we find that

$$\sqrt{k}(I_{n,2}-\gamma) \xrightarrow{d} \frac{(1-\gamma)(\gamma^{-1}-1)^{-\rho}}{(1-\rho)(1-\gamma-\rho)}\lambda_1 - \mathbb{E}(Y)\frac{\gamma^2(\gamma^{-1}-1)^{\gamma}}{\gamma+1}\lambda_2 + Z.$$

It is now enough to compute the variance of Z, which is

$$\mathrm{Var}(Z) = \gamma(1-\gamma)\int_0^1\int_0^1 \frac{\min(s,t)}{st}([1-\gamma]s^{-\gamma}-1)([1-\gamma]t^{-\gamma}-1)\,ds\,dt.$$

It then follows from straightforward but lengthy computations that $\mathrm{Var}(Z) = 2\gamma^3/(1-2\gamma)$; we omit the details. Consequently

$$\sqrt{k}(I_{n,2}-\gamma) \xrightarrow{d} \mathcal{N}\left(\frac{(1-\gamma)(\gamma^{-1}-1)^{-\rho}}{(1-\rho)(1-\gamma-\rho)}\lambda_1 - \mathbb{E}(Y)\frac{\gamma^2(\gamma^{-1}-1)^{\gamma}}{\gamma+1}\lambda_2, \frac{2\gamma^3}{1-2\gamma}\right). \tag{B.4}$$

Combining (B.1), (B.3) and (B.4) completes the proof of (i).

To show (ii), it suffices to prove that

$$|\breve{\gamma}_k - \widetilde{\gamma}_{k,l}| = \mathrm{O}_{\mathbb{P}}\left(\frac{\log(k)}{l}\right). \tag{B.5}$$

Write then

$$|\breve{\gamma}_k - \widetilde{\gamma}_{k,l}| = \left|\sum_{i=1}^{l}\int_{(i-1)/l}^{i/l}\left[\log\left(\frac{\widetilde{\xi}_{1-ks/n}}{\widetilde{\xi}_{1-k/n}}\right) - \log\left(\frac{\widetilde{\xi}_{1-(i-1)k/(ln)}}{\widetilde{\xi}_{1-k/n}}\right)\right]ds\right|$$

and use the sample-wise monotonicity of the random function $s \mapsto \widetilde{\xi}_{1-ks/n}$ to get

$$|\breve{\gamma}_k - \widetilde{\gamma}_{k,l}| \leqslant \frac{1}{l}\sum_{i=1}^{l}\log\left(\frac{\widetilde{\xi}_{1-(i-1)k/(ln)}}{\widetilde{\xi}_{1-ik/(ln)}}\right) = \frac{1}{l}\log\left(\frac{\widetilde{\xi}_1}{\widetilde{\xi}_{1-k/n}}\right).$$

Conclude then using (B.2), which shows (B.5) and completes the proof. ∎

Proof of Theorem 2. Since

$$\overline{\gamma}_k(\alpha) = \alpha\widehat{\gamma}_k + (1-\alpha)\widetilde{\gamma}_k$$

it is sufficient to analyse the joint asymptotic behaviour of $(\widehat{\gamma}_k, \widetilde{\gamma}_k, \widehat{q}_{1-k/n}, \widetilde{\xi}_{1-k/n})$. Let us then start by remarking that

$$\widehat{\gamma}_k = \int_0^1 \log\left(\frac{\widehat{q}_{1-\lfloor k\rfloor s/n}}{q_{1-\lfloor k\rfloor/n}}\right)ds - \log\left(\frac{\widehat{q}_{1-\lfloor k\rfloor/n}}{q_{1-\lfloor k\rfloor/n}}\right).$$

Note that, in Proposition 1, the sequence of Brownian motions is left unchanged if k is changed into $\lfloor k \rfloor$ or $\lceil k \rceil$; this is indeed the fundamental argument behind the proof of Lemma 2(i) in Daouia *et al.* (2020). Arguing as in the proof of Theorem 1 (*i.e.* splitting the domain $s \in (0,1]$ into low and high values of s and using a Taylor expansion), and using the asymptotic equivalences $\sqrt{\lfloor k \rfloor} \sim \sqrt{k}$ and $A(n/\lfloor k \rfloor) \sim A(n/k)$ (the latter due to the regular variation of $|A|$), we get by Proposition 1 that:

$$\sqrt{k}(\widehat{\gamma}_k - \gamma) = \frac{\lambda_1}{1-\rho} + \gamma\sqrt{\gamma^{-1}-1}\left(\int_0^1 \frac{1}{s} W_n\left(\frac{s}{\gamma^{-1}-1}\right) ds - W_n\left(\frac{1}{\gamma^{-1}-1}\right)\right) + \mathrm{o}_{\mathbb{P}}(1). \quad \text{(B.6)}$$

Besides, an inspection of the proof of Theorem 1 shows that

$$\begin{aligned} \sqrt{k}(\widetilde{\gamma}_k - \gamma) &= \frac{(1-\gamma)(\gamma^{-1}-1)^{-\rho}}{(1-\rho)(1-\gamma-\rho)}\lambda_1 - \mathbb{E}(Y)\frac{\gamma^2(\gamma^{-1}-1)^{\gamma}}{\gamma+1}\lambda_2 \\ &+ \gamma\sqrt{\gamma^{-1}-1}\int_0^1 \frac{W_n(s)}{s}([1-\gamma]s^{-\gamma}-1)ds + \mathrm{o}_{\mathbb{P}}(1) \end{aligned} \quad \text{(B.7)}$$

where W_n is the sequence of Brownian motions appearing in Proposition 1. Using Proposition 1 twice more, we can also write

$$\sqrt{k}\left(\frac{\widehat{q}_{1-k/n}}{q_{1-k/n}} - 1\right) = \gamma\sqrt{\gamma^{-1}-1}\, W_n\left(\frac{1}{\gamma^{-1}-1}\right) + \mathrm{o}_{\mathbb{P}}(1) \quad \text{(B.8)}$$

as well as

$$\sqrt{k}\left(\frac{\widetilde{\xi}_{1-k/n}}{\xi_{1-k/n}} - 1\right) = \gamma^2\sqrt{\gamma^{-1}-1}\int_0^1 W_n(t)t^{-\gamma-1}dt + \mathrm{o}_{\mathbb{P}}(1). \quad \text{(B.9)}$$

As a consequence, the random vector

$$\sqrt{k}\left(\widehat{\gamma}_k - \gamma, \widetilde{\gamma}_k - \gamma, \frac{\widehat{q}_{1-k/n}}{q_{1-k/n}} - 1, \frac{\widetilde{\xi}_{1-k/n}}{\xi_{1-k/n}} - 1\right)$$

is asymptotically four-variate Gaussian, and as such

$$\sqrt{k}\left(\overline{\gamma}_k(\alpha) - \gamma, \frac{\widehat{q}_{1-k/n}}{q_{1-k/n}} - 1, \frac{\widetilde{\xi}_{1-k/n}}{\xi_{1-k/n}} - 1\right)$$

is asymptotically trivariate Gaussian. To complete the proof, we analyse the marginal asymptotic behaviour of each of the three components in this vector, as well as their pairwise asymptotic covariance structure.

Marginal asymptotic behaviour of $\overline{\gamma}_k(\alpha)$: This is determined by the joint convergence of $\sqrt{k}\,(\widehat{\gamma}_k - \gamma, \widetilde{\gamma}_k - \gamma)$, to what we already know to be a bivariate Gaussian distribution. We also know from Theorem 3.2.5 in de Haan and Ferreira (2006) that

$$\sqrt{k}(\widehat{\gamma}_k - \gamma) \xrightarrow{d} \mathcal{N}\left(\frac{\lambda_1}{1-\rho}, \gamma^2\right).$$

This is of course also a corollary of (B.6). Meanwhile, Theorem 1(ii) gives

$$\sqrt{k}(\widetilde{\gamma}_k - \gamma) \xrightarrow{d} \mathcal{N}\left(\frac{(1-\gamma)(\gamma^{-1}-1)^{-\rho}}{(1-\rho)(1-\gamma-\rho)}\lambda_1 - \mathbb{E}(Y)\frac{\gamma^2(\gamma^{-1}-1)^{\gamma}}{\gamma+1}\lambda_2, \frac{2\gamma^3}{1-2\gamma}\right).$$

It therefore only remains to calculate the limiting covariance of $\sqrt{k}(\widehat{\gamma}_k - \gamma, \widetilde{\gamma}_k - \gamma)$. This is obtained by computing the expectation of the product of the centred Gaussian terms appearing in the two asymptotic expansions (B.6) and (B.7). In other words, the limiting covariance is

$$\begin{aligned}
\text{cov} &= \text{cov}_1 - \text{cov}_2 \\
\text{with cov}_1 &:= \gamma(1-\gamma)\int_0^1\int_0^1 \frac{\min(s,(\gamma^{-1}-1)^{-1}t)}{st}([1-\gamma]s^{-\gamma}-1)ds\,dt \\
\text{and cov}_2 &:= \gamma(1-\gamma)\int_0^1 \frac{\min(s,(\gamma^{-1}-1)^{-1})}{s}([1-\gamma]s^{-\gamma}-1)ds.
\end{aligned}$$

Recalling that $\gamma^{-1}-1>1$, straightforward computations entail

$$\begin{aligned}
\text{cov}_1 &= (\gamma^{-1}-1)^{\gamma-1} - \gamma[\gamma+1+\gamma\log(\gamma^{-1}-1)] \\
\text{and cov}_2 &= \gamma[(\gamma^{-1}-1)^{\gamma} - 1 - \gamma\log(\gamma^{-1}-1)].
\end{aligned} \tag{B.10}$$

This results in

$$\text{cov} = \gamma[(\gamma^{-1}-1)^{\gamma-1} - \gamma] = \gamma^2\left(\frac{(\gamma^{-1}-1)^{\gamma}}{1-\gamma} - 1\right).$$

Wrapping up, we obtain

$$\sqrt{k}\,(\widehat{\gamma}_k - \gamma, \widetilde{\gamma}_k - \gamma) \xrightarrow{d} \mathcal{N}(\boldsymbol{m}, \boldsymbol{V}) \tag{B.11}$$

where $\boldsymbol{m}$ is the 2×1 vector

$$\boldsymbol{m} := \left(\frac{\lambda_1}{1-\rho}, \frac{(1-\gamma)(\gamma^{-1}-1)^{-\rho}}{(1-\rho)(1-\gamma-\rho)}\lambda_1 - \mathbb{E}(Y)\frac{\gamma^2(\gamma^{-1}-1)^{\gamma}}{\gamma+1}\lambda_2\right)$$

and $\boldsymbol{V}$ is the 2×2 matrix

$$\boldsymbol{V} := \begin{pmatrix} \gamma^2 & \gamma^2\left(\dfrac{(\gamma^{-1}-1)^{\gamma}}{1-\gamma} - 1\right) \\ \gamma^2\left(\dfrac{(\gamma^{-1}-1)^{\gamma}}{1-\gamma} - 1\right) & \dfrac{2\gamma^3}{1-2\gamma} \end{pmatrix}.$$

After some more straightforward computations, we conclude that

$$\sqrt{k}\,(\overline{\gamma}_k(\alpha) - \gamma) \xrightarrow{d} \mathcal{N}\,(b_\alpha, \mathfrak{V}_\alpha(1,1))$$

with the notation of the statement of Theorem 2.

Marginal asymptotic behaviour of $\widehat{q}_{1-k/n}$: It is a straightforward byproduct of Equation (B.8) that

$$\sqrt{k}\left(\frac{\widehat{q}_{1-k/n}}{q_{1-k/n}}-1\right)\xrightarrow{d}\mathcal{N}(0,\gamma^2).$$

Marginal asymptotic behaviour of $\widetilde{\xi}_{1-k/n}$: It is a direct consequence of Equation (B.9) that

$$\sqrt{k}\left(\frac{\widetilde{\xi}_{1-k/n}}{\xi_{1-k/n}}-1\right)\xrightarrow{d}\mathcal{N}\left(0,\frac{2\gamma^3}{1-2\gamma}\right).$$

See also the discussion below Theorem 1 in Daouia *et al.* (2020).

Asymptotic covariance structure of $(\overline{\gamma}_k(\alpha),\widehat{q}_{1-k/n})$: For this, we remark first that $\widehat{\gamma}_k-\gamma$ and $\widehat{q}_{1-k/n}/q_{1-k/n}-1$ are asymptotically independent: this is a consequence of the asymptotic representation of $\widehat{\gamma}_k-\gamma$ obtained in the proof of Theorem 3.2.5 in de Haan and Ferreira (2006) together with Lemma 3.2.3 therein. Besides, the limiting covariance structure of $\sqrt{k}(\widetilde{\gamma}_k-\gamma,\widehat{q}_{1-k/n}/q_{1-k/n}-1)$ is obtained by computing the expectation of the product of the centred Gaussian terms appearing in the asymptotic expansions (B.7) and (B.8). By (B.10) above, this limiting covariance is:

$$\mathrm{cov}_2=\gamma[(\gamma^{-1}-1)^{\gamma}-1-\gamma\log(\gamma^{-1}-1)]$$

with the notation of (B.10). The limiting covariance of $\sqrt{k}(\overline{\gamma}_k(\alpha)-\gamma,\widehat{q}_{1-k/n}/q_{1-k/n}-1)$ is then

$$(1-\alpha)\gamma[(\gamma^{-1}-1)^{\gamma}-1-\gamma\log(\gamma^{-1}-1)]=\mathfrak{V}_{\alpha}(1,2).$$

Asymptotic covariance structure of $(\overline{\gamma}_k(\alpha),\widetilde{\xi}_{1-k/n})$: It follows from Equations (B.6), (B.7) and (B.9) that the limiting covariance of $\sqrt{k}(\overline{\gamma}_k(\alpha)-\gamma,\widetilde{\xi}_{1-k/n}/\xi_{1-k/n}-1)$ is

$$\mathcal{COV}=\alpha\mathcal{COV}_1+(1-\alpha)\mathcal{COV}_2$$

with

$$\mathcal{COV}_1=\gamma^2(1-\gamma)\left[\int_0^1\int_0^1\frac{\min((\gamma^{-1}-1)^{-1}s,t)}{s}t^{-\gamma-1}ds\,dt-\int_0^1\min((\gamma^{-1}-1)^{-1},t)t^{-\gamma-1}dt\right]$$

and

$$\mathcal{COV}_2=\gamma^2(1-\gamma)\int_0^1\int_0^1\frac{\min(s,t)}{s}([1-\gamma]s^{-\gamma}-1)t^{-\gamma-1}ds\,dt.$$

Direct computations yield

$$\mathcal{COV}_1=\gamma^2\left(\frac{(\gamma^{-1}-1)^{\gamma}}{(1-\gamma)^2}-1\right)-\gamma^2\left(\frac{(\gamma^{-1}-1)^{\gamma}}{1-\gamma}-1\right)=\frac{\gamma^3(\gamma^{-1}-1)^{\gamma}}{(1-\gamma)^2}$$

and

$$\mathcal{COV}_2 = \frac{\gamma^3}{(1-\gamma)(1-2\gamma)}.$$

Consequently

$$\mathcal{COV} = \frac{\gamma^3}{(1-\gamma)^2}\left[\alpha(\gamma^{-1}-1)^\gamma + (1-\alpha)\frac{1-\gamma}{1-2\gamma}\right] = \mathfrak{V}_\alpha(1,3).$$

Asymptotic covariance structure of $(\widehat{q}_{1-k/n}, \widetilde{\xi}_{1-k/n})$*:* Combining Equations (B.8) and (B.9), we find that the limiting covariance of $\sqrt{k}(\widehat{q}_{1-k/n}/q_{1-k/n} - 1, \widetilde{\xi}_{1-k/n}/\xi_{1-k/n} - 1)$ is

$$\gamma^2(1-\gamma)\int_0^1 \min(t, (\gamma^{-1}-1)^{-1})t^{-\gamma-1}dt = \gamma^2\left(\frac{(\gamma^{-1}-1)^\gamma}{1-\gamma} - 1\right) = \mathfrak{V}_\alpha(2,3)$$

after some straightforward calculations.

Combining these arguments on marginal convergence and asymptotic covariance structure, we get

$$\sqrt{k}\left(\overline{\gamma}_k(\alpha) - \gamma, \frac{\widehat{q}_{1-k/n}}{q_{1-k/n}} - 1, \frac{\widetilde{\xi}_{1-k/n}}{\xi_{1-k/n}} - 1\right) \xrightarrow{d} \mathcal{N}(\mathfrak{m}_\alpha, \mathfrak{V}_\alpha) \tag{B.12}$$

with $\mathfrak{m}_\alpha$ and $\mathfrak{V}_\alpha$ as in the statement of Theorem 2. This concludes the proof. ∎

Proof of Theorem 3. Applying Theorem 2 and arguing as in the proof of Theorem 1 in Daouia *et al.* (2018), we get the joint convergence

$$\sqrt{k}\left(\frac{\widehat{\xi}_{1-k/n}(\alpha)}{\xi_{1-k/n}} - 1, \frac{\widetilde{\xi}_{1-k/n}}{\xi_{1-k/n}} - 1\right) \xrightarrow{d} \left([(1-\gamma)^{-1} - \log(\gamma^{-1}-1)]\Gamma_\alpha + \Theta - \lambda,\, \Xi\right)$$

where $(\Gamma_\alpha, \Theta, \Xi)$ is the limiting vector in Theorem 2, and

$$\lambda := \left(\frac{(\gamma^{-1}-1)^{-\rho}}{1-\gamma-\rho} + \frac{(\gamma^{-1}-1)^{-\rho} - 1}{\rho}\right)\lambda_1 + \gamma(\gamma^{-1}-1)^\gamma \mathbb{E}(Y)\lambda_2.$$

Then clearly

$$\sqrt{k}\left(\frac{\overline{\xi}_{1-k/n}(\alpha,\beta)}{\xi_{1-k/n}} - 1\right) \xrightarrow{d} [(1-\gamma)^{-1} - \log(\gamma^{-1}-1)]\beta\Gamma_\alpha + \beta\Theta + (1-\beta)\Xi - \beta\lambda.$$

Set $\Psi_\alpha = \Gamma_\alpha - b_\alpha$ and rearrange the bias component to complete the proof. ∎

Proof of Theorem 4. Define $p_n = 1 - \tau_n'$ and note that

$$\begin{aligned}\log\left(\frac{\overline{\xi}^\star_{1-p_n}(\alpha,\beta)}{\xi_{1-p_n}}\right) &= (\overline{\gamma}_{1-k/n}(\alpha) - \gamma)\log\left(\frac{k}{np_n}\right) + \log\left(\frac{\overline{\xi}_{1-k/n}(\alpha,\beta)}{\xi_{1-k/n}}\right) \\ &\quad - \log\left(\left[\frac{np_n}{k}\right]^\gamma \frac{\xi_{1-p_n}}{\xi_{1-k/n}}\right).\end{aligned}$$

The convergence $\log[k/(np_n)] \to \infty$ yields

$$\frac{\sqrt{k}}{\log[k/(np_n)]} \log\left(\frac{\overline{\xi}_{1-k/n}(\alpha,\beta)}{\xi_{1-k/n}}\right) = \mathrm{O}_{\mathbb{P}}\left(1/\log[k/(np_n)]\right) = \mathrm{o}_{\mathbb{P}}(1) \tag{B.13}$$

$$\begin{aligned}
\text{and} \quad & \frac{\sqrt{k}}{\log[k/(np_n)]} \log\left(\left[\frac{np_n}{k}\right]^{\gamma} \frac{\xi_{1-p_n}}{\xi_{1-k/n}}\right) \\
&= \frac{\sqrt{k}}{\log[k/(np_n)]} \left(\log\left(\frac{\xi_{1-p_n}}{q_{1-p_n}}\right) - \log\left(\frac{\xi_{1-k/n}}{q_{1-k/n}}\right) + \log\left(\left[\frac{np_n}{k}\right]^{\gamma} \frac{q_{1-p_n}}{q_{1-k/n}}\right)\right) \\
&= \mathrm{O}\left(\frac{\sqrt{k}}{\log[k/(np_n)]} \left[\frac{1}{q_{1-k/n}} + |A(n/k)| + \frac{1}{q_{1-p_n}} + |A(1/p_n)|\right]\right) \\
&= \mathrm{O}\left(\frac{\sqrt{k}}{\log[k/(np_n)]} \left[\frac{1}{q_{1-k/n}} + |A(n/k)|\right]\right) \\
&= \mathrm{o}(1).
\end{aligned} \tag{B.14}$$

Here, convergence (B.13) is a consequence of Theorem 3. Convergence (B.14) follows from a combination of Proposition 1 in Daouia *et al.* (2020), Theorem 2.3.9 in de Haan and Ferreira (2006) and the regular variation of $|A|$. Combining these convergences and using the delta-method leads to the desired result. ∎

Proof of Theorem 5. The proof of the convergence of $\widetilde{\mathrm{XES}}^{\star}_{\tau'_n}(\alpha)$ is entirely similar to that of Theorem 4 (applying Theorem 6 in Daouia *et al.* (2020) instead of Theorem 3, and Proposition 4 in Daouia *et al.* (2020) instead of their Proposition 1). We omit the details.

We proceed by examining the convergence of $\overline{\mathrm{XES}}^{\star}_{1-p_n}(\alpha,\beta)$. Define $p_n = 1-\tau'_n$ and write

$$\begin{aligned}
\log\left(\frac{\overline{\mathrm{XES}}^{\star}_{1-p_n}(\alpha,\beta)}{\mathrm{XES}_{1-p_n}}\right) &= \log\left(\frac{\overline{\xi}^{\star}_{1-p_n}(\alpha,\beta)}{\xi_{1-p_n}}\right) + \log\left(\frac{[1-\overline{\gamma}_{1-k/n}(\alpha)]^{-1}}{[1-\gamma]^{-1}}\right) \\
&\quad - \log\left(\frac{\mathrm{XES}_{1-p_n}}{[1-\gamma]^{-1}\xi_{1-p_n}}\right).
\end{aligned}$$

By Theorem 4 and the delta-method,

$$\frac{\sqrt{k}}{\log[k/(np_n)]} \log\left(\frac{\overline{\xi}^{\star}_{1-p_n}(\alpha,\beta)}{\xi_{1-p_n}}\right) \xrightarrow{d} \mathcal{N}(b_\alpha, v_\alpha). \tag{B.15}$$

Using then Theorem 2, the delta-method and the convergence $\log[k/(np_n)] \to \infty$, we get

$$\frac{\sqrt{k}}{\log[k/(np_n)]} \log\left(\frac{[1-\overline{\gamma}_{1-k/n}(\alpha)]^{-1}}{[1-\gamma]^{-1}}\right) \xrightarrow{\mathbb{P}} 0. \tag{B.16}$$

Using finally a combination of Propositions 1(i) and 4 in Daouia *et al.* (2020) and the regular variation of $|A|$ and $t \mapsto q_{1-t^{-1}}$, we obtain

$$\frac{\sqrt{k}}{\log[k/(np_n)]} \log\left(\frac{\mathrm{XES}_{1-p_n}}{[1-\gamma]^{-1}\xi_{1-p_n}}\right) \to 0. \tag{B.17}$$

Combining convergences (B.15), (B.16) and (B.17), it follows that

$$\frac{\sqrt{k}}{\log[k/(np_n)]} \log\left(\frac{\overline{\mathrm{XES}}^{\star}_{1-p_n}(\alpha,\beta)}{\mathrm{XES}_{1-p_n}}\right) \xrightarrow{d} \mathcal{N}(b_\alpha, v_\alpha).$$

Another use of the delta-method completes the proof of the convergence of $\overline{\mathrm{XES}}^{\star}_{1-p_n}(\alpha,\beta)$. ■

Proof of Theorem 6. We only show the result for $\widetilde{\mathrm{XES}}^{\star}_{\widehat{\tau}'_n(p_n)}(\alpha)$ as the proof of the other convergence is entirely similar. The key point is to write

$$\widetilde{\mathrm{XES}}^{\star}_{\widehat{\tau}'_n(p_n)}(\alpha) = \left(\frac{1-\widehat{\tau}'_n(p_n)}{1-\tau'_n(p_n)}\right)^{-\overline{\gamma}_{\tau_n}(\alpha)} \widetilde{\mathrm{XES}}^{\star}_{\tau'_n(p_n)}(\alpha). \tag{B.18}$$

It is, moreover, shown as part of the proof of Theorem 6 in Daouia *et al.* (2018) that

$$\frac{1-\widehat{\tau}'_n(p_n)}{1-\tau'_n(p_n)} = 1 + \mathrm{O}_{\mathbb{P}}\left(\frac{1}{\sqrt{n(1-\tau_n)}}\right).$$

[Combine (B.52), (B.53), (B.54) and (B.55) in the Supplementary Material document of Daouia *et al.* (2018), noting that the strict monotonicity of F_Y is not required thanks to Proposition 1(i) in Daouia *et al.* (2020); this also results in a corrected version of (B.51) in the former paper]. Therefore

$$\begin{aligned}
\left(\frac{1-\widehat{\tau}'_n(p_n)}{1-\tau'_n(p_n)}\right)^{-\overline{\gamma}_{\tau_n}(\alpha)} &= \exp\left(-\overline{\gamma}_{\tau_n}(\alpha)\log\left[\frac{1-\widehat{\tau}'_n(p_n)}{1-\tau'_n(p_n)}\right]\right) \\
&= \exp\left(-\left[\gamma + \mathrm{O}_{\mathbb{P}}\left(\frac{1}{\sqrt{n(1-\tau_n)}}\right)\right] \times \mathrm{O}_{\mathbb{P}}\left(\frac{1}{\sqrt{n(1-\tau_n)}}\right)\right) \\
&= 1 + \mathrm{O}_{\mathbb{P}}\left(\frac{1}{\sqrt{n(1-\tau_n)}}\right).
\end{aligned} \tag{B.19}$$

Furthermore, using the equivalent

$$1-\tau'_n(p_n) \sim (1-p_n)\,\frac{\gamma}{1-\gamma}, \tag{B.20}$$

we conclude that the conditions of Theorem 5 are satisfied if the parameter τ_n' there is set equal to $\tau_n'(p_n)$. By Theorem 5 then:

$$\frac{\sqrt{n(1-\tau_n)}}{\log[(1-\tau_n)/(1-\tau_n'(p_n))]}\left(\frac{\widetilde{\mathrm{XES}}^{\star}_{\tau_n'(p_n)}(\alpha)}{\mathrm{XES}_{\tau_n'(p_n)}}-1\right)\xrightarrow{d}\mathcal{N}(b_\alpha,v_\alpha).$$

Now

$$\log\left[\frac{1-\tau_n}{1-\tau_n'(p_n)}\right]=\log\left[\frac{1-\tau_n}{1-p_n}\right]+\log\left[\frac{1-p_n}{1-\tau_n'(p_n)}\right]$$

and in the right-hand side of this identity, the first term tends to infinity, while the second term converges to a finite constant in view of (B.20). As a conclusion

$$\log\left[\frac{1-\tau_n}{1-\tau_n'(p_n)}\right]\sim\log\left[\frac{1-\tau_n}{1-p_n}\right].$$

Hence the convergence

$$\frac{\sqrt{n(1-\tau_n)}}{\log[(1-\tau_n)/(1-p_n)]}\left(\frac{\widetilde{\mathrm{XES}}^{\star}_{\tau_n'(p_n)}(\alpha)}{\mathrm{XES}_{\tau_n'(p_n)}}-1\right)\xrightarrow{d}\mathcal{N}(b_\alpha,v_\alpha). \tag{B.21}$$

We conclude the proof by writing

$$\mathrm{XES}_{\tau_n'(p_n)}=\mathrm{QES}_{p_n}\times\left\{(1-\gamma)\frac{\mathrm{XES}_{\tau_n'(p_n)}}{\xi_{\tau_n'(p_n)}}\right\}\times\left\{(1-\gamma)\frac{\mathrm{QES}_{p_n}}{q_{p_n}}\right\}^{-1}$$

(since $\xi_{\tau_n'(p_n)}\equiv q_{p_n}$ by definition). By a combination of Proposition 4 in Daouia *et al.* (2020) with (B.20) and the regular variation of the functions $|A|$ and $t\mapsto q_{1-t^{-1}}$, one gets

$$(1-\gamma)\frac{\mathrm{XES}_{\tau_n'(p_n)}}{\xi_{\tau_n'(p_n)}}=1+\mathrm{o}\left(\frac{\log[(1-\tau_n)/(1-p_n)]}{\sqrt{n(1-\tau_n)}}\right).$$

Similarly and by Equation (B.20) in the Supplementary Material document of Daouia *et al.* (2020),

$$(1-\gamma)\frac{\mathrm{QES}_{p_n}}{q_{p_n}}=1+\mathrm{o}\left(\frac{\log[(1-\tau_n)/(1-p_n)]}{\sqrt{n(1-\tau_n)}}\right).$$

Therefore

$$\frac{\mathrm{XES}_{\tau_n'(p_n)}}{\mathrm{QES}_{p_n}}-1=\mathrm{o}\left(\frac{\log[(1-\tau_n)/(1-p_n)]}{\sqrt{n(1-\tau_n)}}\right).$$

Together with (B.21), this entails

$$\frac{\sqrt{n(1-\tau_n)}}{\log[(1-\tau_n)/(1-p_n)]}\left(\frac{\widetilde{\mathrm{XES}}^{\star}_{\tau_n'(p_n)}(\alpha)}{\mathrm{QES}_{p_n}}-1\right)\xrightarrow{d}\mathcal{N}(b_\alpha,v_\alpha). \tag{B.22}$$

Combining (B.18), (B.19) and (B.22) completes the proof. ■

References

Daouia, A., Girard, S. and Stupfler, G. (2018). Estimation of tail risk based on extreme expectiles, *Journal of the Royal Statistical Society: Series B* **80**: 263–292.

Daouia, A., Girard, S. and Stupfler, G. (2020). Tail expectile process and risk assessment, *Bernoulli* **26**: 531–556.

El Methni, J. and Stupfler, G. (2017). Extreme versions of Wang risk measures and their estimation for heavy-tailed distributions, *Statistica Sinica* **27**: 907–930.

El Methni, J. and Stupfler, G. (2018). Improved estimators of extreme Wang distortion risk measures for very heavy-tailed distributions, *Econometrics and Statistics* **6**: 129–148.

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