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] Université Grenoble Alpes, Inria, CNRS, Grenoble INP, LJK, France

[c] School of Mathematical Sciences, University of Nottingham, United Kingdom

Section A illustrates the behavior of the *expectHill* estimator of the tail index with data examples. Section B plots and comments on the asymptotic variance of the *expectHill* estimator. Simulation results are discussed in Section C. The proofs of all theoretical results in the main paper and additional technical results are provided in Section D.

A Examples of tail index estimation

The aim of this section is to illustrate the behavior of the *expectHill* estimator with data examples 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 2 gives its asymptotic normality. As pointed out in Remark 1, the conditions involving the auxiliary function A in Theorem 2 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. This is illustrated in the following two examples.

Example 1. Our first motivating example is concerned with data from three large US financial institutions. We consider the same investment banks as in the study of Cai *et al.* (2015), namely Goldman Sachs, Morgan Stanley and T. Rowe Price. The dataset consists of the loss returns (*i.e.* minus log-returns) on their equity prices at a daily frequency from July 3rd, 2000, to June 30th, 2010. Figure 1(a)-(c) 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. The chosen stable regime in Cai *et al.* (2015) is $k \in [70, 100]$ for the three Hill plots. To gain stability in the estimates, they took the average of the estimates $\widehat{\gamma}_{1-k/n}$ over this region. The results are reported in the second column of Table 1. As regards the asymmetric least squares estimator $\widetilde{\gamma}_{1-k/n}$, we applied a very simple technique which consists in computing its standard deviations over a moving window of 30 successive values of k [same length as the chosen interval in Cai *et al.* (2015)]; this corresponds to a window large enough to cover around 20% of the possible values of k in the selected range $1 \leqslant k \leqslant 150$. The value of k where the standard deviation (and hence the variation) of the estimates is minimal defines the desired sample fraction $\widetilde{k}$. We found $\widetilde{k} = 72$ in the window $[54, 84]$ for Goldman Sachs, $\widetilde{k} = 80 \in [62, 92]$ for Morgan Stanley, and $\widetilde{k} = 88 \in [68, 98]$ for T. Rowe Price. The final estimates $\widetilde{\gamma}_{1-\widetilde{k}/n}$ are reported in the third column of Table 1. The messages yielded by the two methods are broadly similar, indicating particularly that Morgan Stanley displays a greater variability in loss returns and a much heavier tail of their distribution than do Goldman Sachs and T. Rowe Price.

Bank	$\widehat{\gamma}_{1-k/n}$	$\widetilde{\gamma}_{1-k/n}$
Goldman Sachs	0.388	0.372
Morgan Stanley	0.465	0.422
T. Rowe Price	0.378	0.365

Table 1: *Tail index estimates based on daily loss returns* ($n = 2513$).

Example 2. Our second example is concerned with large medical insurance claims from the SOA (Society of Actuaries) group. We consider the same database as in Beirlant *et al.* (2004) that contains 75,789 claim amounts exceeding 25,000 USD, collected over the year 1991 from 26 insurers. Figure 1(d) shows the paths $k \mapsto \widehat{\gamma}_{1-k/n}$ and $k \mapsto \widetilde{\gamma}_{1-k/n}$ for these data. The minimal standard deviation of the estimates indicates a pointwise estimate $\widehat{\gamma}_{1-k/n}$ around 0.36 achieved over the window $[119, 259]$, and an estimate $\widetilde{\gamma}_{1-k/n}$ around 0.35 attained over $[331, 471]$. Here also the standard deviations were computed over a moving window large enough to cover 20% of the possible values of k in the selected range $1 \leqslant k \leqslant 700$. It is remarkable that, on this large dataset, the estimator $k \mapsto \widetilde{\gamma}_{1-k/n}$ appears to stabilize very quickly around the estimated value.

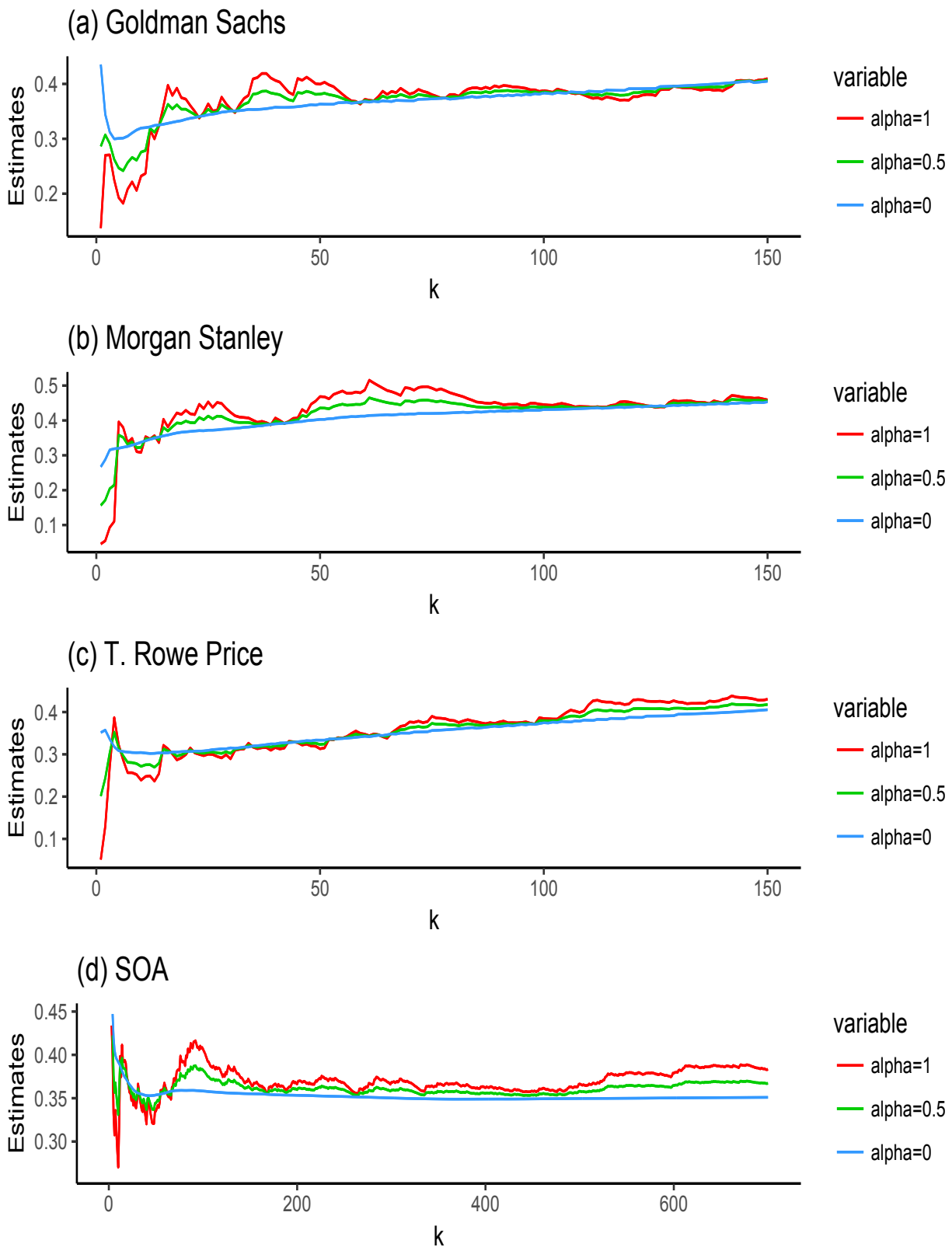


Figure 1: *Plots of* $\widehat{\gamma}_{1-k/n}$ *in red,* $\widetilde{\gamma}_{1-k/n}$ *in blue, and* $\overline{\gamma}_{1-k/n}(1/2)$ *in green, for the three banks in (a)-(c) and for the SOA group medical insurance large claims in (d).*

In these examples, the purely least asymmetrically weighted squares estimator $\widetilde{\gamma}_{1-k/n}$ seems to be beneficial in producing smoother and more pleasing plots, but these plots may not be more revealing than Hill plots. Already in Figure 1(a)-(c), 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

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

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 $\widehat{\gamma}_{\tau_n}$.

B Asymptotic variance of the *expectHill* estimator

The optimal value of the weighting coefficient α in (A.1), which minimizes the asymptotic variance v_α of $\overline{\gamma}_{\tau_n}(\alpha)$, only depends on the tail index γ and has the explicit expression

$$\alpha(\gamma) = \frac{(1-\gamma) - (1-2\gamma)(\gamma^{-1}-1)^{\gamma}}{(1-\gamma)(3-4\gamma) - 2(1-2\gamma)(\gamma^{-1}-1)^{\gamma}}.$$

Its plot against $\gamma \in (0, 1/2)$ is given in Figure 2(a). Interestingly, the optimal α is negative for small values of γ, say $\gamma \leqslant 0.2$. By contrast, for large values of γ (close to $1/2$), the optimal α tends to one, favoring thus the robustness of order statistics over the tail sensitivity of asymmetric least squares. It can also be seen that the simple mean $\overline{\gamma}_{\tau_n}(1/2)$ of $\widehat{\gamma}_{\tau_n}$ and $\widetilde{\gamma}_{\tau_n}$, with $\alpha = 1/2$, is optimal for $\gamma = 1/4$. This is unsurprising since both $\widehat{\gamma}_{\tau_n}$ and $\widetilde{\gamma}_{\tau_n}$ have the same asymptotic variance in this case, as illustrated in Figure 2(b). This Figure also shows that the mean $\overline{\gamma}_{\tau_n}(1/2)$ of $\widehat{\gamma}_{\tau_n}$ and $\widetilde{\gamma}_{\tau_n}$ 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. In terms of smoothness, $\overline{\gamma}_{\tau_n}(1/2)$ offers a middle course as well, as shown above in Figure 1, where the plot of $\overline{\gamma}_{\tau_n}(1/2)$ is superimposed in green line with the plots of $\widehat{\gamma}_{\tau_n}$ and $\widetilde{\gamma}_{\tau_n}$.

C Some simulation evidence

The aim of this section is to explore some features that were mentioned in Section 6 of the main article. We will illustrate the following points:

(C.1) Estimates of γ.

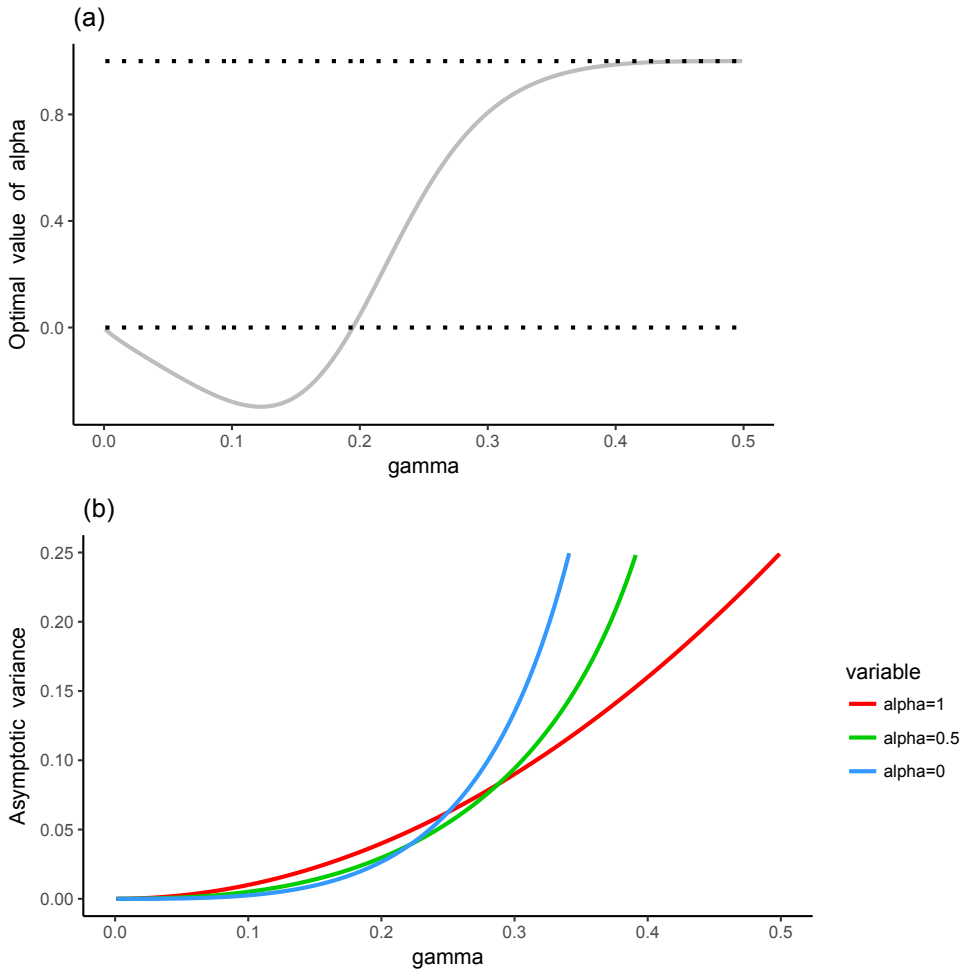


Figure 2: *(a) – Evolution of the optimal value* $\alpha(\gamma)$ *against* $\gamma \in (0, 1/2)$. *The dotted lines represent the values* $\alpha = 0$ *and* $\alpha = 1$. *(b) – Asymptotic variance* v_α *of* $\widehat{\gamma}_{\tau_n}$ *in red* ($\alpha = 1$), $\widetilde{\gamma}_{\tau_n}$ *in blue* ($\alpha = 0$), *and* $\overline{\gamma}_{\tau_n}(1/2)$ *in green* ($\alpha = 0.5$), *as functions of* $\gamma \in (0, 1/2)$.

(C.2) Estimates of $\mathrm{XES}_{\tau'_n}$.

(C.3) Estimates of QES_{p_n}.

(C.4) Confidence intervals for QES_{p_n}.

In order to illustrate the behavior of the presented estimation procedures, we use the same considerations as in Section 6 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 $n = 500$ and true tail index $\gamma \in \{0.35, 0.45\}$. 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.

C.1 Estimation of the tail index

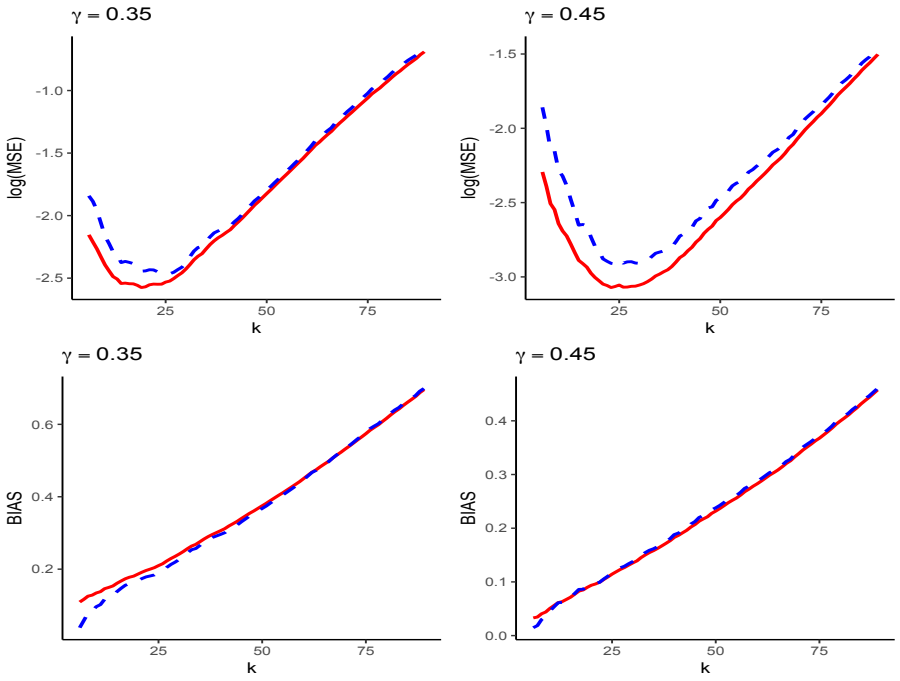
This section provides Monte-Carlo evidence that the *expectHill* estimator $\overline{\gamma}_{1-k/n}(\alpha)$, introduced in (A.1) with the weight $\alpha = 1/2$, is more efficient relative to the standard Hill estimator $\widehat{\gamma}_{1-k/n}$, for both Student and Fréchet distributions. In the case of the real-valued Student distribution, Figure 3(a) gives the evolution of the MSE (in top panels) and the bias (in bottom panels) of $\overline{\gamma}_{1-k/n}(\frac{1}{2})/\gamma$ and $\widehat{\gamma}_{1-k/n}/\gamma$, as functions of the effective sample fraction k. It may be seen that $\overline{\gamma}_{1-k/n}(\frac{1}{2})$ performs better than $\widehat{\gamma}_{1-k/n}$ in terms of MSE, for all values of k, without sacrificing too much quality in terms of bias, especially for the larger value of γ. We arrive at the same tentative conclusion in the case of the Fréchet distribution as may be seen from Figure 3(b). By contrast, in the special case of the Pareto distribution, the Hill estimator $\widehat{\gamma}_{1-k/n}$ is exactly the maximum likelihood estimator of γ and is unbiased, whereas the *expectHill* estimator $\overline{\gamma}_{1-k/n}(\frac{1}{2}) = \frac{1}{2}(\widehat{\gamma}_{1-k/n} + \widetilde{\gamma}_{1-k/n})$ is biased in this case. Unsurprisingly, the Monte Carlo results obtained in Figure 4 indicate that $\widehat{\gamma}_{1-k/n}$ is as expected the winner in this case.

C.2 Estimates of $\mathrm{XES}_{\tau'_n}$

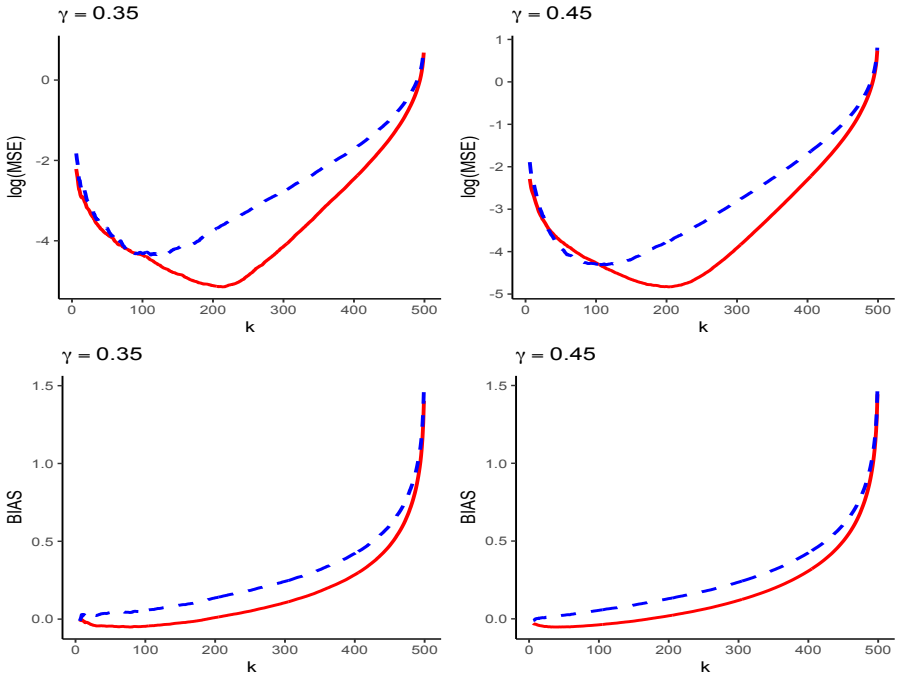
Before comparing the performance of $\overline{\mathrm{XES}}^{\star}_{\tau'_n}(\alpha,\beta)$, $\widehat{\mathrm{XES}}^{\star}_{\tau'_n}(\alpha,\beta)$ and $\widetilde{\mathrm{XES}}^{\star}_{\tau'_n}(\alpha)$ as estimators of $\mathrm{XES}_{\tau'_n}$, we first investigated the accuracy of each estimator in terms of the associated weights α and β.

Figures 5 and 6 give the evolution of the MSE (in log scale) and bias estimates of $\widetilde{\mathrm{XES}}^{\star}_{\tau'_n}(\alpha)/\mathrm{XES}_{\tau'_n}$, as functions of the sample fraction k, for $\alpha \in \{0, 0.25, 0.5, 0.75, 1\}$. In the case of Student distribution, it may be seen that the red curve ($\alpha = 1$) gives the best estimates in terms of both MSE and Bias. In the case of Fréchet distribution, it may be seen that the purple curve ($\alpha = 0.5$) performs quite well in terms of MSE for both values of γ, but the blue curve ($\alpha = 0$) performs clearly better in terms of Bias. In the case of Pareto distribution, it may be seen that the purple ($\alpha = 0.5$) and red ($\alpha = 1$) curves have, respectively, a quite respectable accuracy in terms of MSE for $\gamma = 0.35$ and $\gamma = 0.45$, while the purple curve ($\alpha = 0.5$) behaves clearly better in terms of Bias for both values of γ.

For $(\alpha,\beta) \in \{(0,0),(0,0.5),(0,1),(0.5,0),(0.5,0.5),(0.5,1),(1,0),(1,0.5),(1,1)\}$, Figures 7 and 8 give, respectively, the MSE (in log scale) and Bias estimates of $\overline{\mathrm{XES}}^{\star}_{\tau'_n}(\alpha,\beta)/\mathrm{XES}_{\tau'_n}$, against k. It may be seen that the winner is the red curve ($\alpha = 0.5, \beta = 0$) in the case of



(a) Student distribution



(b) Fréchet distribution

Figure 3: *MSE estimates in log scale (top panels) and Bias estimates (bottom panels) of* $\overline{\gamma}_{1-k/n}(1/2)/\gamma$ *(solid red line) and* $\widehat{\gamma}_{1-k/n}/\gamma$ *(dashed blue line), as functions of* k*, for* $\gamma = 0.35$ *(left) and* $\gamma = 0.45$ *(right).*

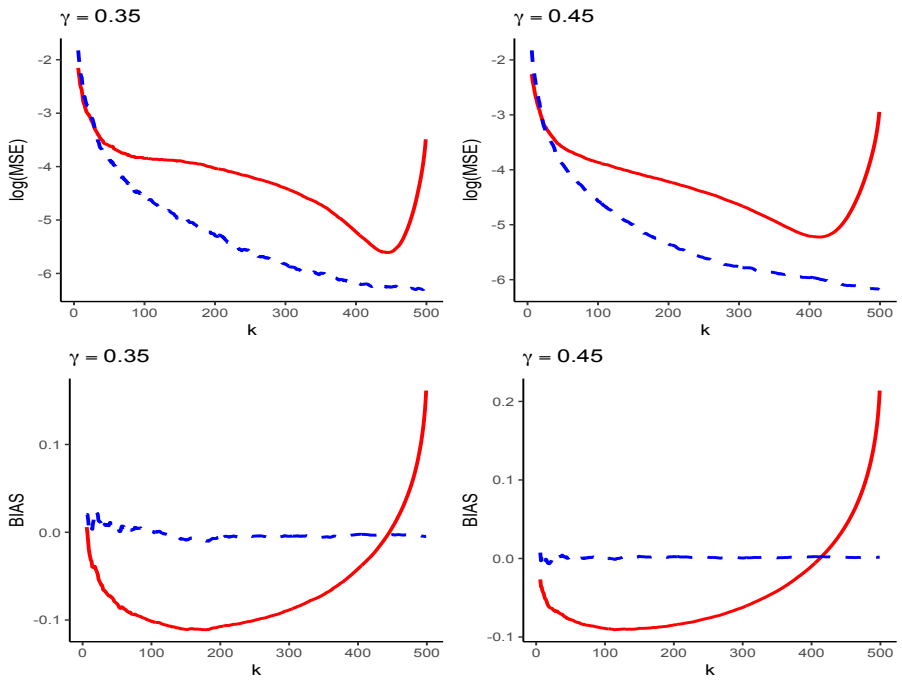


Figure 4: *MSE estimates (top panels) and Bias estimates (bottom panels) of* $\overline{\gamma}_k(1/2)/\gamma$ *(solid red line) and* $\widehat{\gamma}_k/\gamma$ *(dashed blue line), as functions of* k*, for* $\gamma = 0.35$ *(left) and* $\gamma = 0.45$ *(right), in the case of a Pareto distribution.*

Student distribution, the blue curve ($\alpha = 0.5, \beta = 1$) in the case of Fréchet distribution, and the black curve ($\alpha = \beta = 1$) in the case of Pareto distribution.

The Monte Carlo estimates for $\widehat{\mathrm{XES}}^{\star}_{\tau'_n}(\alpha,\beta)/\mathrm{XES}_{\tau'_n}$ are displayed in Figures 9 and 10. It may be seen that the winner is the orange curve ($\alpha = 1, \beta = 0$) in the case of Student distribution, the blue curve ($\alpha = 0.5, \beta = 1$) in the case of Fréchet distribution, and the black curve ($\alpha = \beta = 1$) in the case of Pareto distribution.

Finally, when comparing the three estimators $\widetilde{\mathrm{XES}}^{\star}_{\tau'_n}(\alpha)$, $\overline{\mathrm{XES}}^{\star}_{\tau'_n}(\alpha,\beta)$ and $\widehat{\mathrm{XES}}^{\star}_{\tau'_n}(\alpha,\beta)$ with each other, we arrive at the following tentative conclusions:

- In the case of the Student distribution, the best $\widetilde{\mathrm{XES}}^{\star}_{\tau'_n}(\alpha)$, achieved at $\alpha = 1$, is superior to the best $\widehat{\mathrm{XES}}^{\star}_{\tau'_n}(\alpha,\beta)$, achieved at $\alpha = 1$ and $\beta = 0$, which in turn is superior to the best $\overline{\mathrm{XES}}^{\star}_{\tau'_n}(\alpha,\beta)$, achieved at $\alpha = 0.5$ and $\beta = 0$.

- In the case of the Fréchet distribution, the best $\widehat{\mathrm{XES}}^{\star}_{\tau'_n}(\alpha,\beta)$, achieved at $\alpha = 0.5$ and $\beta = 1$, is superior but not by much to the best $\overline{\mathrm{XES}}^{\star}_{\tau'_n}(\alpha,\beta)$, achieved at $\alpha = 0.5$ and $\beta = 1$, which in turn is superior to the best $\widetilde{\mathrm{XES}}^{\star}_{\tau'_n}(\alpha)$ achieved at, say, $\alpha \in \{0, 0.5\}$.

- In the case of the Pareto distribution, the best $\widehat{\mathrm{XES}}^{\star}_{\tau'_n}(\alpha,\beta)$, achieved at $\alpha = \beta = 1$, is superior but not by much to the best $\overline{\mathrm{XES}}^{\star}_{\tau'_n}(\alpha,\beta)$, achieved at $\alpha = \beta = 1$, which in

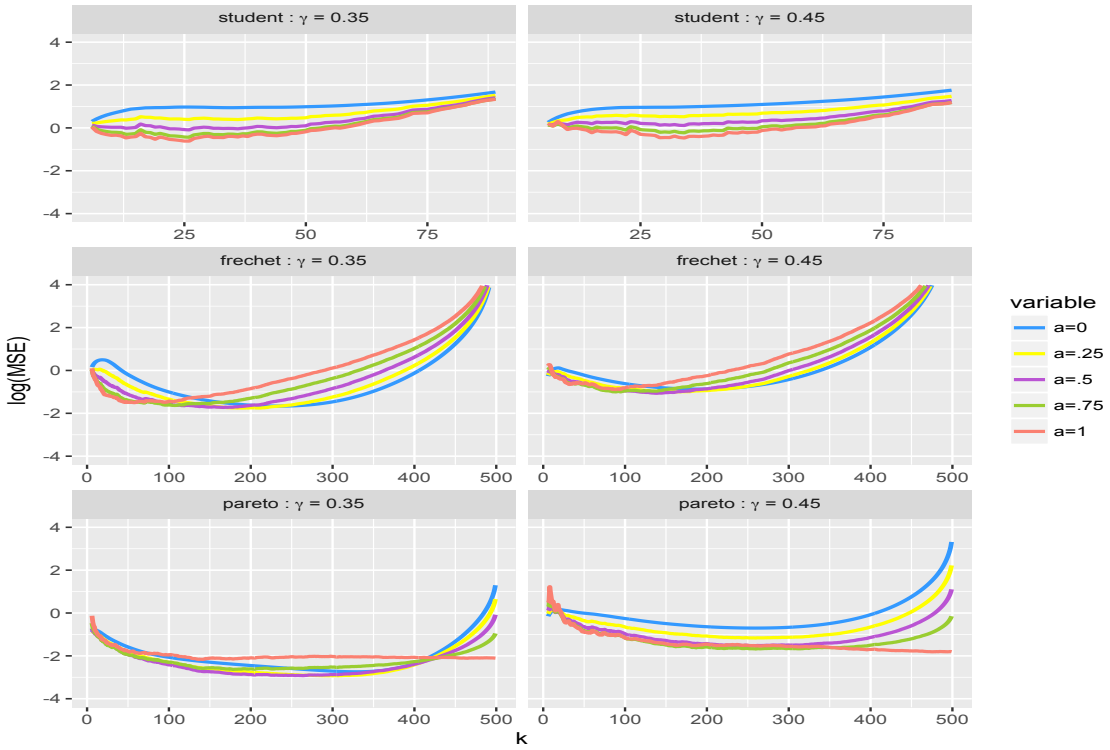


Figure 5: *MSE estimates (in log scale) of* $\widetilde{XES}^{\star}_{\tau'_n}(\alpha)/XES_{\tau'_n}$*, against* k*, for Student (top), Fréchet (middle) and Pareto (bottom) distributions, with* $\gamma = 0.35$ *(left) and* $\gamma = 0.45$ *(right).*

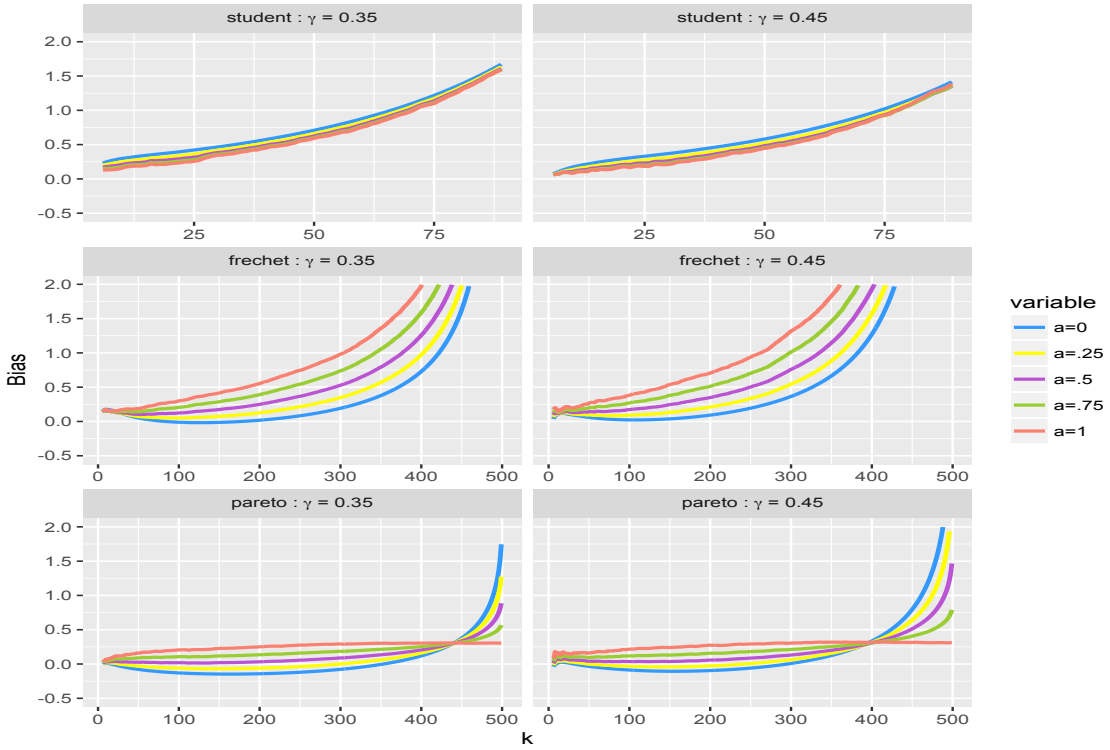


Figure 6: *Bias estimates of* $\widetilde{XES}^{\star}_{\tau'_n}(\alpha)/XES_{\tau'_n}$*.*

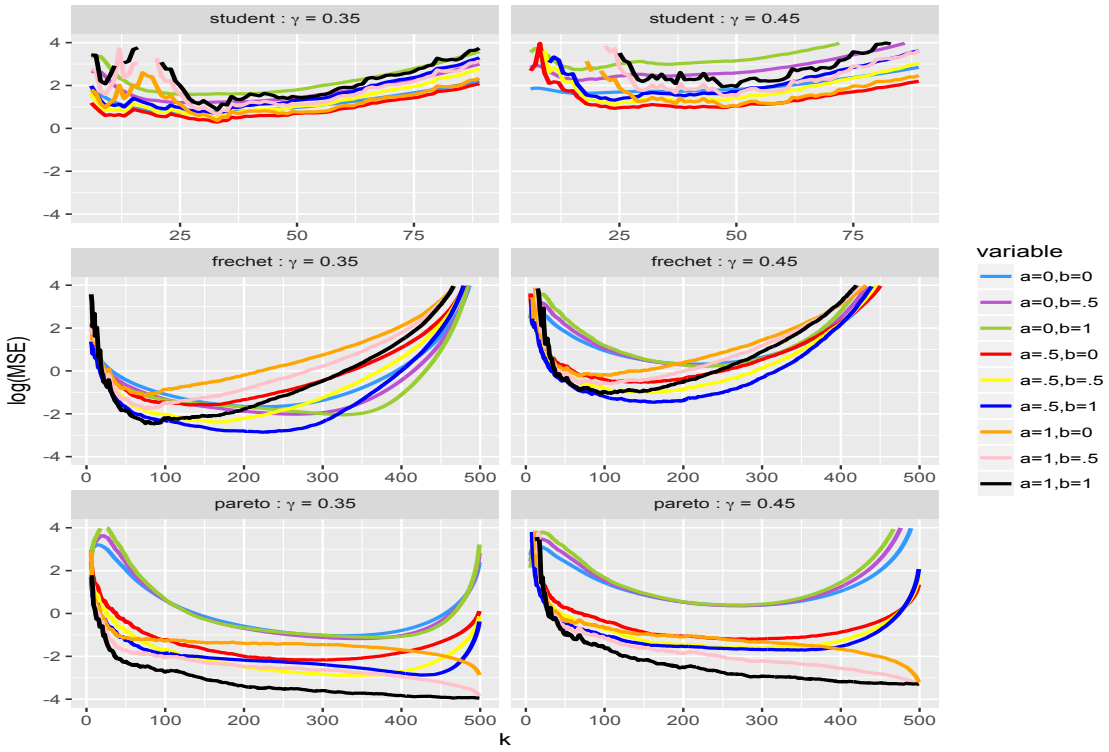


Figure 7: *MSE estimates (in log scale) of* $\overline{XES}^{\star}_{\tau'_n}(\alpha,\beta)/XES_{\tau'_n}$.

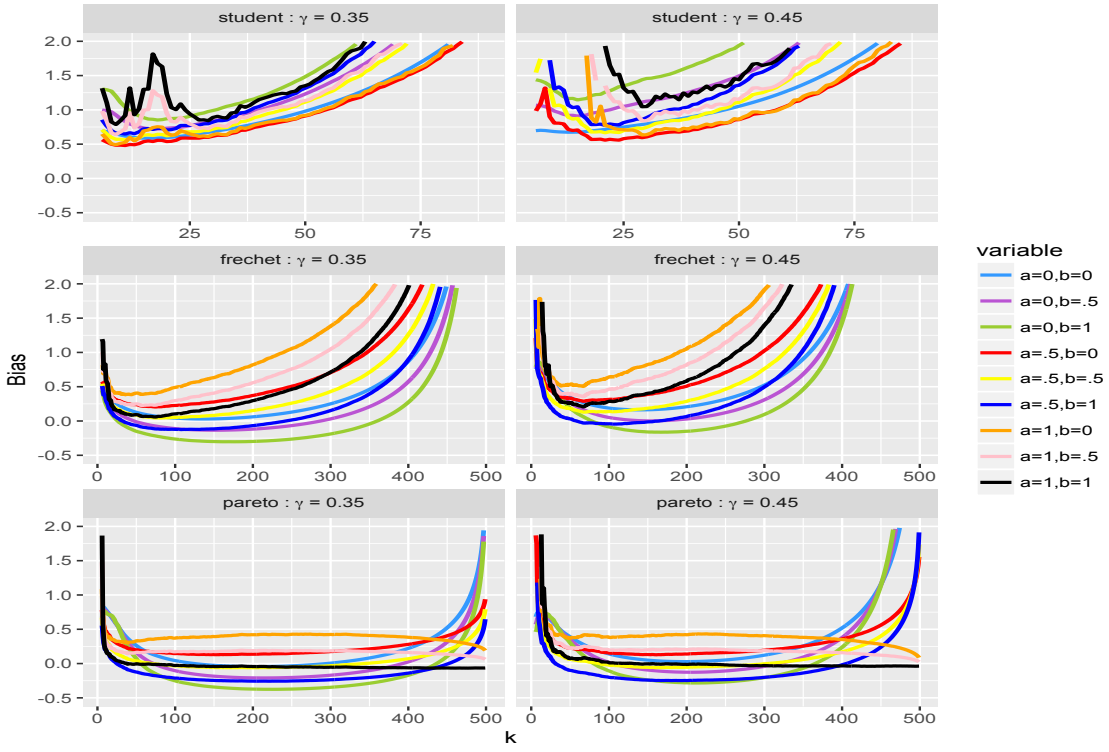


Figure 8: *Bias estimates of* $\overline{XES}^{\star}_{\tau'_n}(\alpha,\beta)/XES_{\tau'_n}$.

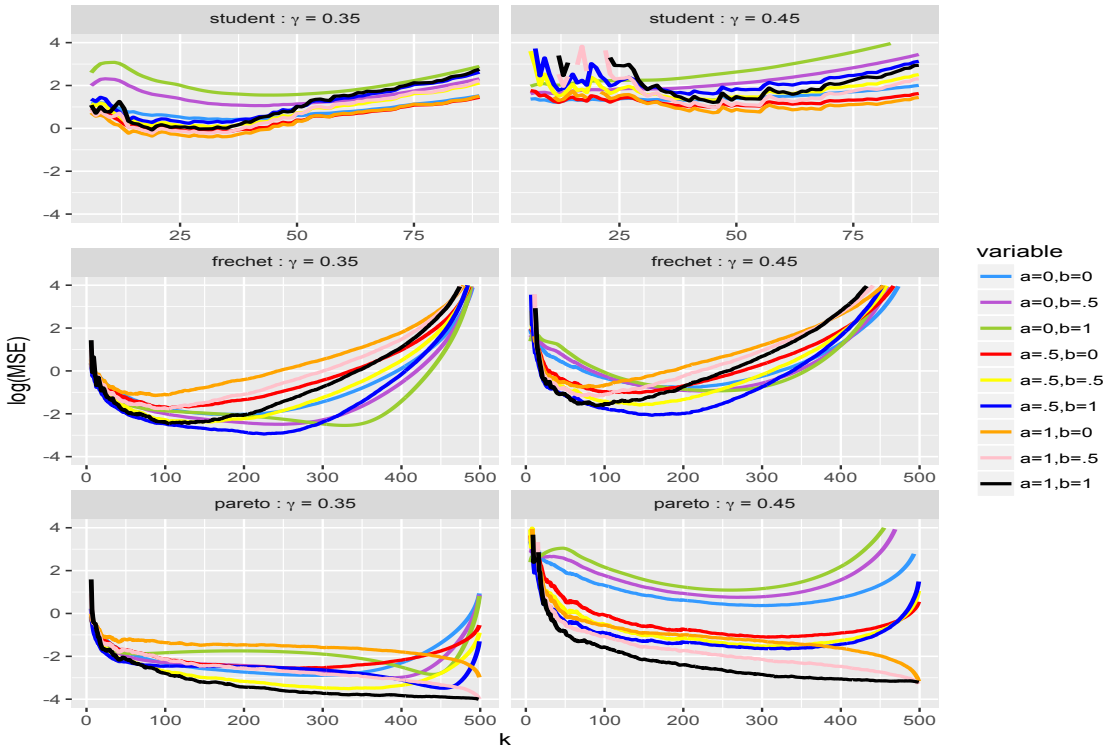


Figure 9: *MSE estimates (in log scale) of* $\widehat{XES}^{\star}_{\tau'_n}(\alpha,\beta)/XES_{\tau'_n}$.

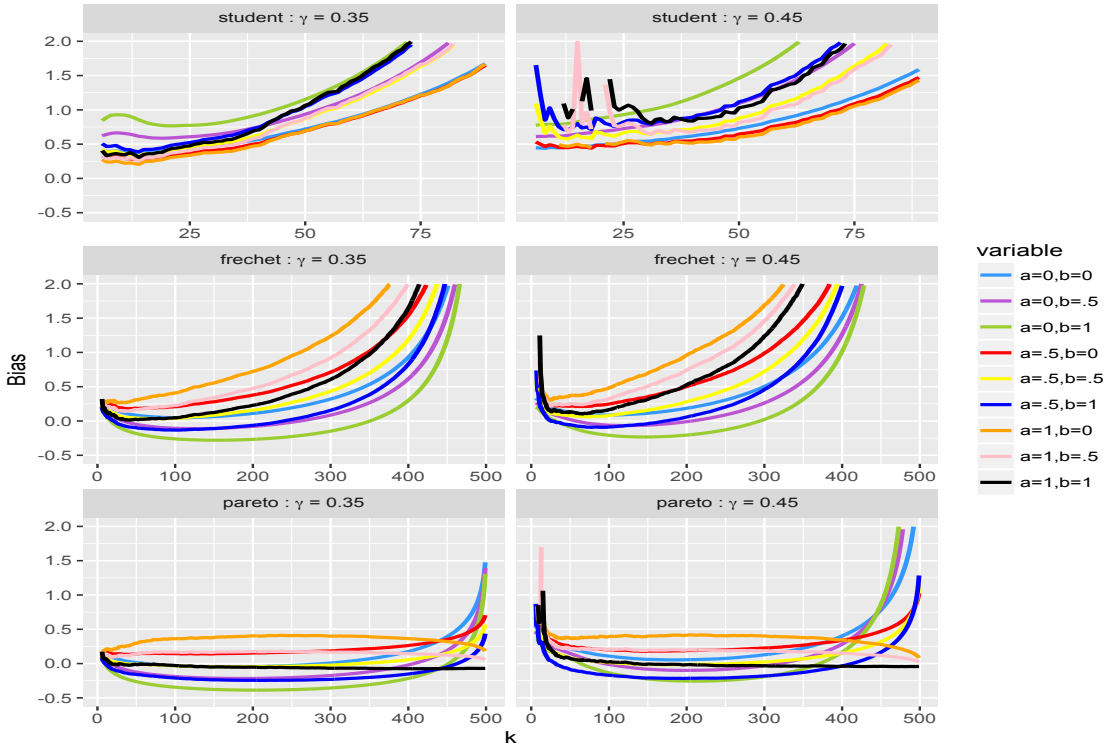


Figure 10: *Bias estimates of* $\widehat{XES}^{\star}_{\tau'_n}(\alpha,\beta)/XES_{\tau'_n}$.

turn is superior to the best $\widetilde{\mathrm{XES}}^{\star}_{\tau'_n}(\alpha)$ achieved at, say, $\alpha \in \{0.5, 1\}$.

In particular, it seems that $\widetilde{\mathrm{XES}}^{\star}_{\tau'_n}(\alpha)$ is the winner in the case of the real-valued profit-loss Student distribution for $\alpha = 1$, while $\widehat{\mathrm{XES}}^{\star}_{\tau'_n}(\alpha, \beta)$ is most efficient in the case of the non-negative Fréchet and Pareto loss distributions, for $\alpha \in \{0.5, 1\}$ and $\beta = 1$.

C.3 Estimates of QES_{p_n}

We have also undertaken simulation experiments to evaluate the finite-sample performance of the composite versions $\overline{\mathrm{XES}}^{\star}_{\widehat{\tau}'_n(p_n)}(\alpha, \beta)$, $\widehat{\mathrm{XES}}^{\star}_{\widehat{\tau}'_n(p_n)}(\alpha, \beta)$ and $\widetilde{\mathrm{XES}}^{\star}_{\widehat{\tau}'_n(p_n)}(\alpha)$ studied in Theorem 7. These composite expectile-based estimators estimate the same conventional expected shortfall QES_{p_n} as the direct quantile-based estimator $\widehat{\mathrm{QES}}^{\star}_{p_n}(\alpha) \equiv \widehat{\mathrm{XES}}^{\star}_{\widehat{\tau}'_n(p_n)}(\alpha, 1)$. We first examined the accuracy of each estimator for various values of α and β.

Figures 11 and 12 give the MSE (in log scale) and Bias estimates of $\widehat{\mathrm{QES}}^{\star}_{p_n}(\alpha)/\mathrm{QES}_{p_n}$. The results suggest the choice of $\alpha = 1$ (red curve) for Student and Pareto distributions, and $\alpha = 0.5$ (purple curve) for Fréchet distribution.

Figures 13 and 14 give the Monte Carlo estimates of $\widetilde{\mathrm{XES}}^{\star}_{\widehat{\tau}'_n(p_n)}(\alpha)/\mathrm{QES}_{p_n}$. It may be seen that the choice of $\alpha = 0$ (blue curve) globally provides quite respectable behavior for the three distributions.

The Monte Carlo estimates of $\overline{\mathrm{XES}}^{\star}_{\widehat{\tau}'_n(p_n)}(\alpha, \beta)/\mathrm{QES}_{p_n}$ are displayed in Figures 15 and 16. It may be seen that the appropriate choice of (α, β) is, respectively, $(1, 0)$ for Student distribution (orange curve), $(0.5, 1)$ for Fréchet distribution (blue curve), and $(1, 1)$ for Pareto distribution (black curve).

The results for $\widehat{\mathrm{XES}}^{\star}_{\widehat{\tau}'_n(p_n)}(\alpha, \beta)/\mathrm{QES}_{p_n}$ are graphed in Figures 17 and 18. Here, it may be seen that the appropriate choice of (α, β) is, respectively, $(0, 0)$ for Student distribution (light blue), $(0.5, 1)$ for Fréchet distribution (dark blue), and $(1, 1)$ for Pareto distribution (black curve).

Finally, when comparing the four estimators $\widehat{\mathrm{QES}}^{\star}_{p_n}(\alpha)$, $\widetilde{\mathrm{XES}}^{\star}_{\widehat{\tau}'_n(p_n)}(\alpha)$, $\overline{\mathrm{XES}}^{\star}_{\widehat{\tau}'_n(p_n)}(\alpha, \beta)$ and $\widehat{\mathrm{XES}}^{\star}_{\widehat{\tau}'_n(p_n)}(\alpha, \beta)$ with each other, we arrive at the following tentative conclusions:

- In the case of the real-valued profit-loss Student distribution, the best estimator seems to be $\widetilde{\mathrm{XES}}^{\star}_{\widehat{\tau}'_n(p_n)}(\alpha = 0)$;
- In the case of the non-negative Fréchet and Pareto loss distributions, the best estimators seem to be $\overline{\mathrm{XES}}^{\star}_{\widehat{\tau}'_n(p_n)}(\alpha = 0.5, \beta = 1)$ and $\widehat{\mathrm{QES}}^{\star}_{p_n}(\alpha = 1) \equiv \widehat{\mathrm{XES}}^{\star}_{\widehat{\tau}'_n(p_n)}(\alpha = 1, \beta = 1)$, respectively.

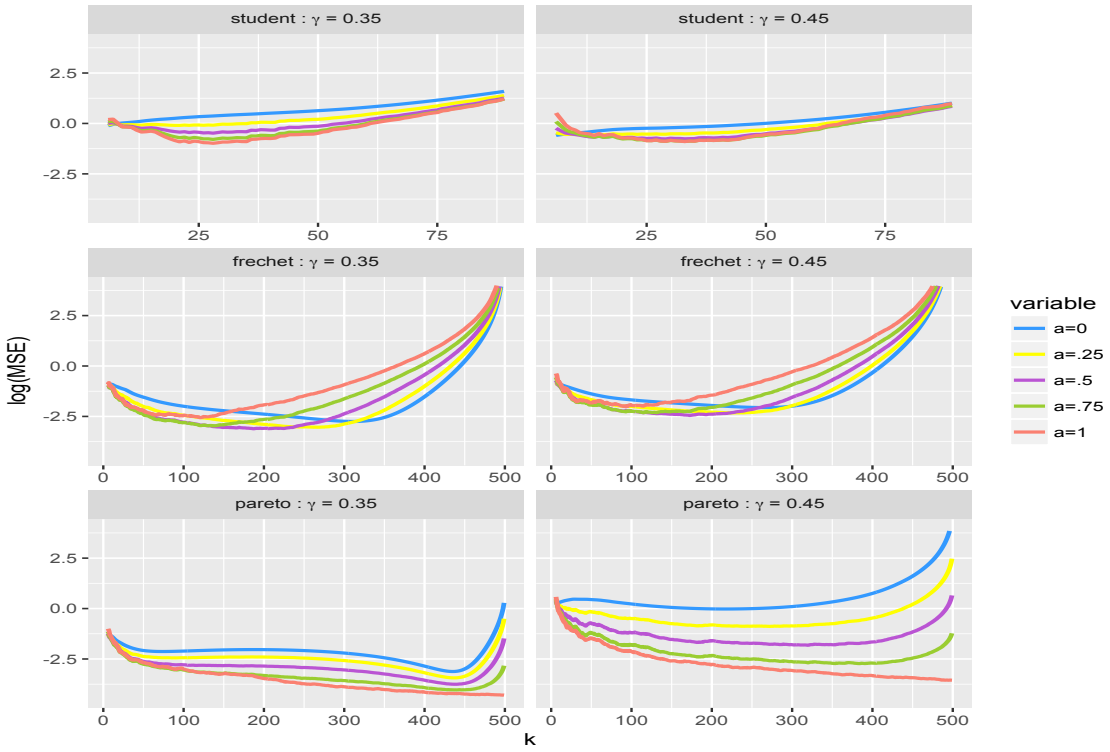


Figure 11: *MSE estimates (in log scale) of* $\widehat{QES}^{\star}_{p_n}(\alpha)/QES_{p_n}$*, against* k*, for Student (top), Fréchet (middle) and Pareto (bottom) distributions, with* $\gamma = 0.35$ *(left) and* $\gamma = 0.45$ *(right).*

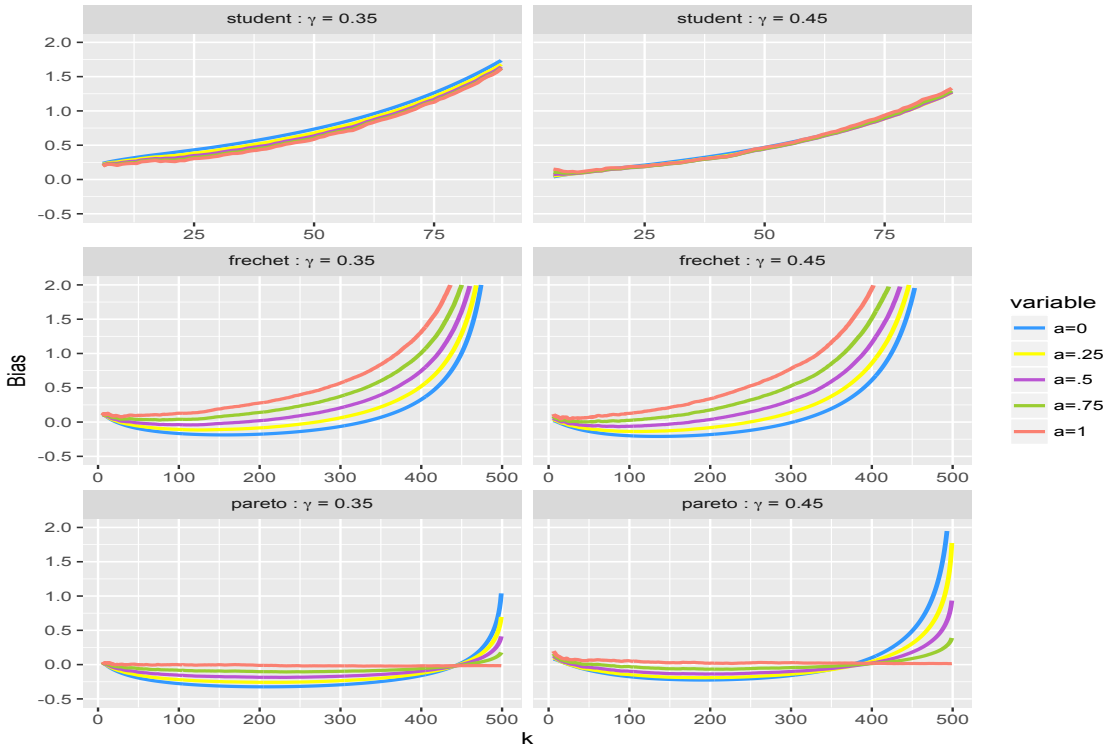


Figure 12: *Bias estimates of* $\widehat{QES}^{\star}_{p_n}(\alpha)/QES_{p_n}$.

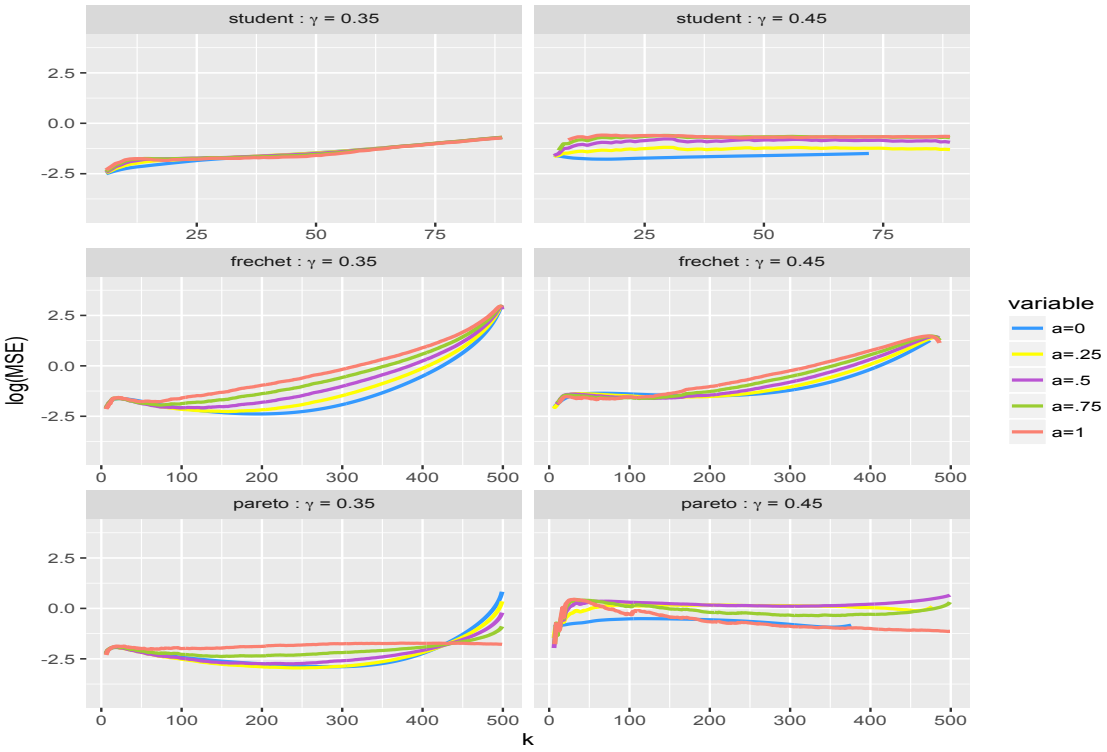


Figure 13: *MSE estimates (in log scale) of* $\widetilde{XES}^{\star}_{\widehat{\tau}'_n(p_n)}(\alpha)/QES_{p_n}$.

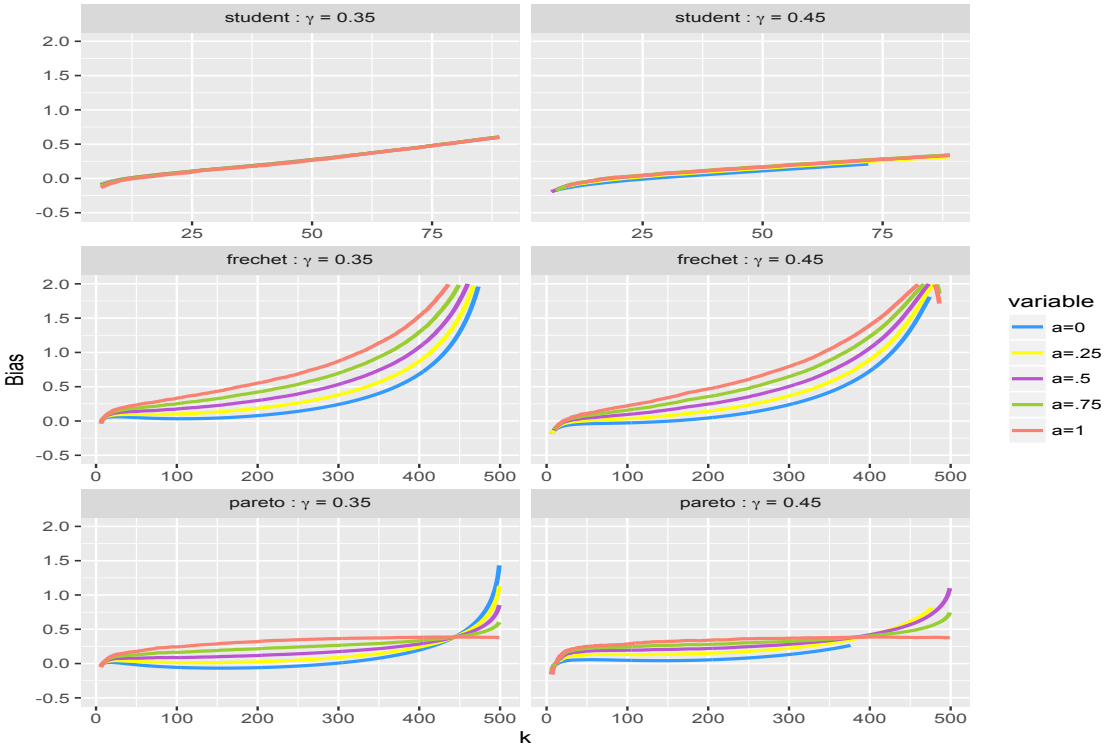


Figure 14: *Bias estimates of* $\widetilde{XES}^{\star}_{\widehat{\tau}'_n(p_n)}(\alpha)/QES_{p_n}$.

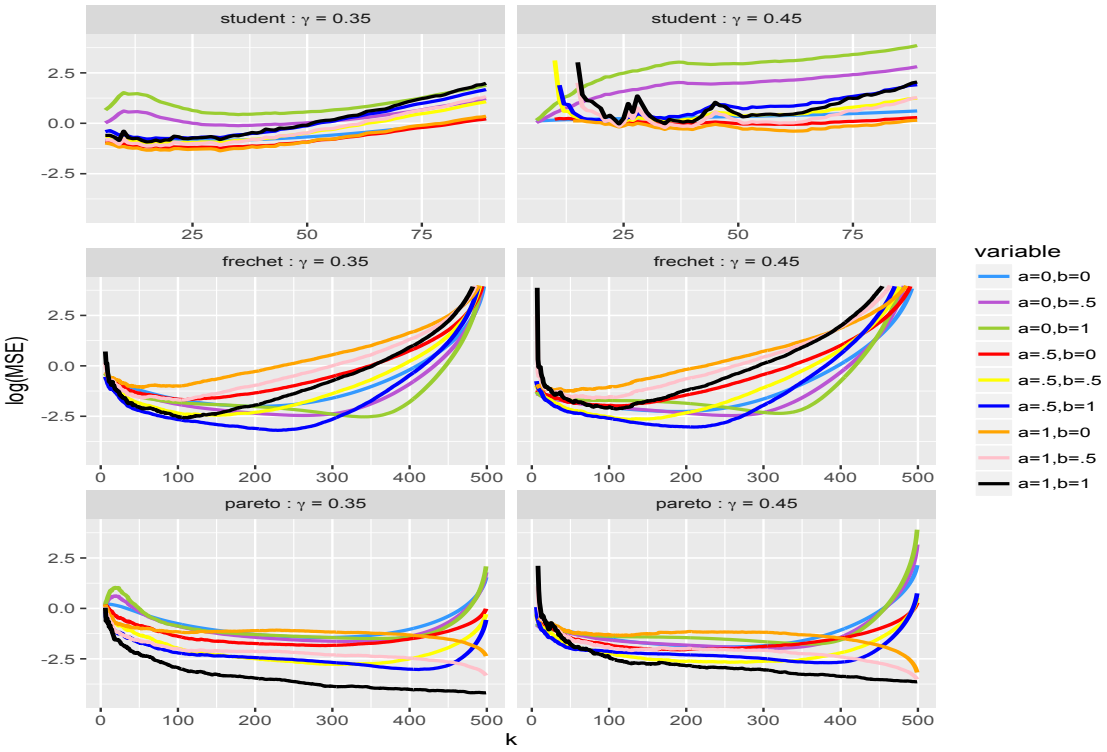


Figure 15: *MSE estimates (in log scale) of* $\overline{XES}^{\star}_{\widehat{\tau}'_n(p_n)}(\alpha,\beta)/QES_{p_n}$.

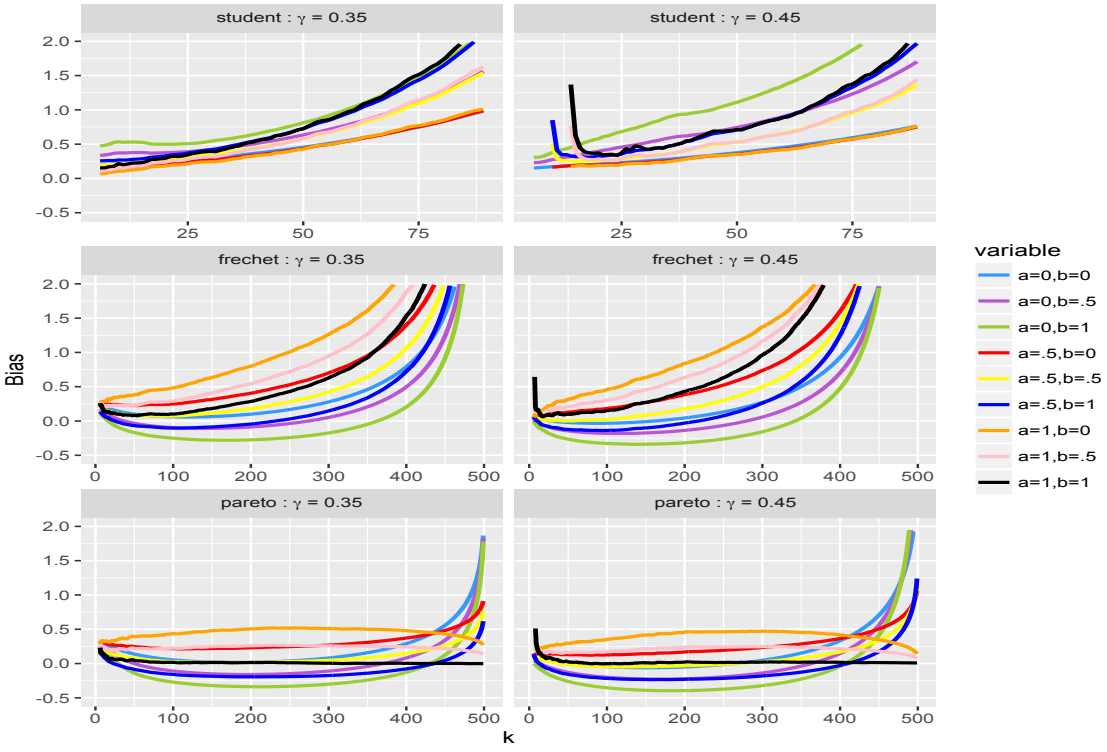


Figure 16: *Bias estimates of* $\overline{XES}^{\star}_{\widehat{\tau}'_n(p_n)}(\alpha,\beta)/QES_{p_n}$.

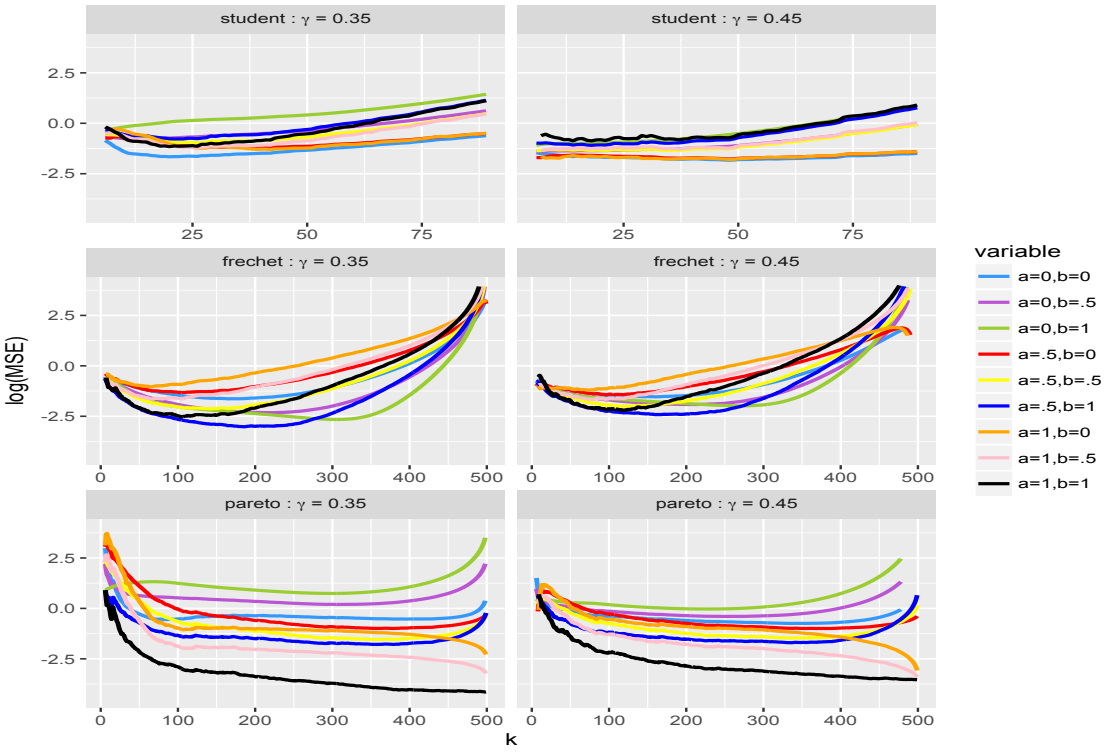


Figure 17: *MSE estimates (in log scale) of* $\widehat{XES}^{\star}_{\widehat{\tau}'_n(p_n)}(\alpha,\beta)/QES_{p_n}$.

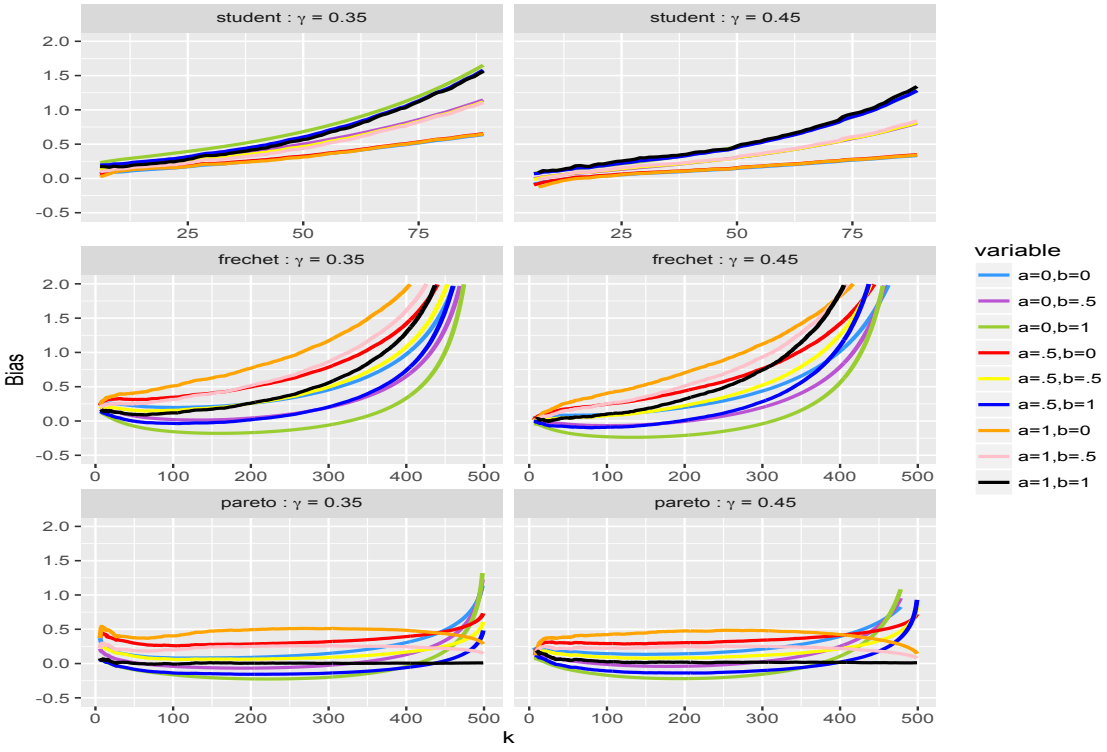


Figure 18: *Bias estimates of* $\widehat{XES}^{\star}_{\widehat{\tau}'_n(p_n)}(\alpha,\beta)/QES_{p_n}$.

C.4 Confidence intervals for QES_{p_n}

We also investigated the performance of the three asymptotic confidence intervals described in Section 6.2.3, namely $\overline{\mathrm{CI}}_{0.95}(k)$, $\widehat{\mathrm{CI}}_{0.95}(k)$ and $\widetilde{\mathrm{CI}}_{0.95}(k)$. For the classical 95% confidence level, we take the value $z_{0.975} \approx 1.960$. Our experiments indicate that,

- in the case of the real-valued Student profit-loss distribution, while $\widetilde{\mathrm{XES}}^{\star}_{\widehat{\tau}'_n(p_n)}(\alpha)$ provides the best estimates in terms of MSE and bias for $\alpha = 0$, as seen before, its corresponding 95% asymptotic confidence intervals $\widetilde{\mathrm{CI}}_{0.95}(k)$ perform better in terms of average lengths and achieved coverages for the different weight $\alpha = 1$. This is visualized in Figures 19 and 20, where the average lengths and achieved coverages of $\widetilde{\mathrm{CI}}_{0.95}(k)$ are graphed against k in blue for $\alpha = 0$ and in red for $\alpha = 1$. The choice of $\alpha = 0$ is not without disadvantage as can be seen from the average lengths, where good results require a large sample size of the order of several thousands. For the sake of clarity, we do not report here the other comparisons with $\overline{\mathrm{CI}}_{0.95}(k)$ and $\widehat{\mathrm{CI}}_{0.95}(k)$, for different values of α and β;

- in the case of the non-negative Fréchet and Pareto loss distributions, $\overline{\mathrm{XES}}^{\star}_{\widehat{\tau}'_n(p_n)}(\alpha, \beta)$ and $\widehat{\mathrm{XES}}^{\star}_{\widehat{\tau}'_n(p_n)}(\alpha, \beta)$ provide the best estimates in terms of MSE and bias for the weights $(\alpha = 0.5, \beta = 1)$ and $(\alpha = 1, \beta = 1)$, respectively, as seen before. Their corresponding asymptotic confidence intervals $\overline{\mathrm{CI}}_{0.95}(k)$ and $\widehat{\mathrm{CI}}_{0.95}(k)$ seem also to be superior in terms of average lengths and achieved coverages, but the best choice of weights for $\overline{\mathrm{CI}}_{0.95}(k)$ is $(\alpha = 1, \beta = 0.5)$ instead of $(\alpha = 0.5, \beta = 1)$. Here also, we report in Figures 21 and 22 only the results for $\overline{\mathrm{CI}}_{0.95}(k)$, with $(\alpha = 1, \beta = 0.5)$ in blue lines and $(\alpha = 0.5, \beta = 1)$ in red lines, and for $\widehat{\mathrm{CI}}_{0.95}(k)$ with $(\alpha = 1, \beta = 1)$ in black lines. It may be seen that the asymptotic confidence interval $\overline{\mathrm{CI}}_{0.95}(k)$, for $(\alpha = 0.5, \beta = 1)$ in red, disappoints by its average lengths for $\gamma = 0.45$ and by its achieved coverages for both values of γ. By contrast, both $\overline{\mathrm{CI}}_{0.95}(k)$, for $(\alpha = 1, \beta = 0.5)$ in blue, and $\widehat{\mathrm{CI}}_{0.95}(k)$, for $(\alpha = 1, \beta = 1)$ in black, perform overall quite well.

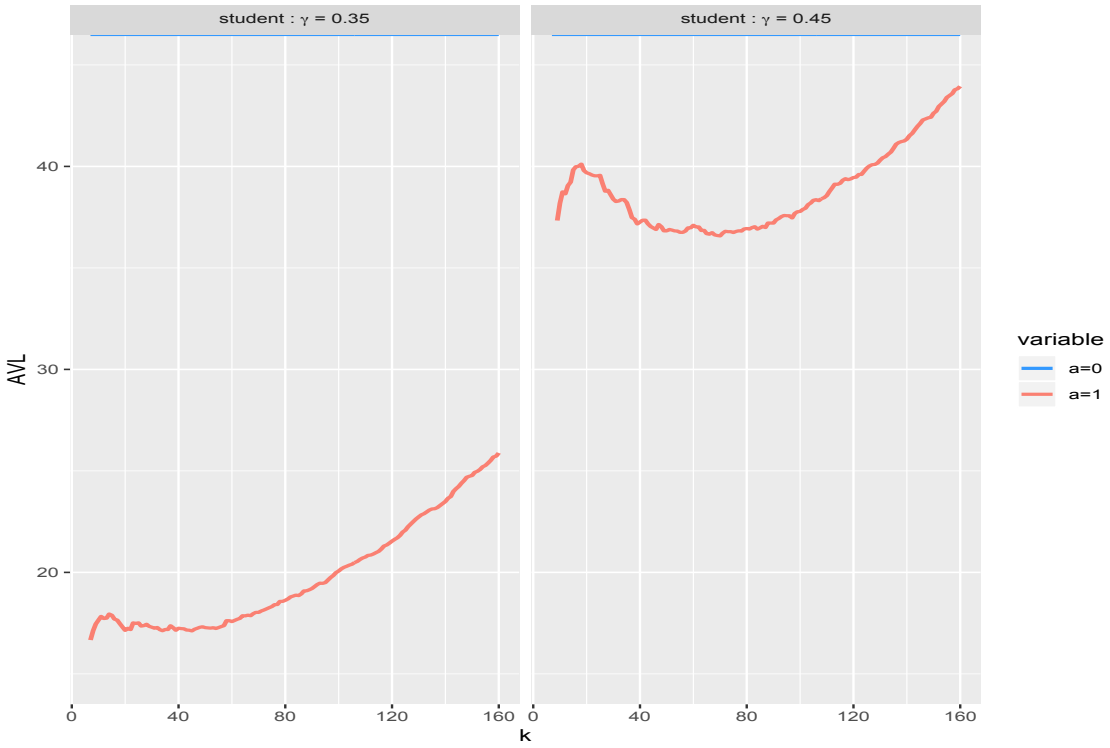


Figure 19: *Case of Student distribution—Average lengths of* $\widetilde{CI}_{0.95}(k)$ *against* k*, for* $\alpha = 0$ *in blue and* $\alpha = 1$ *in red. From left to right,* $\gamma = 0.35, 0.45$.

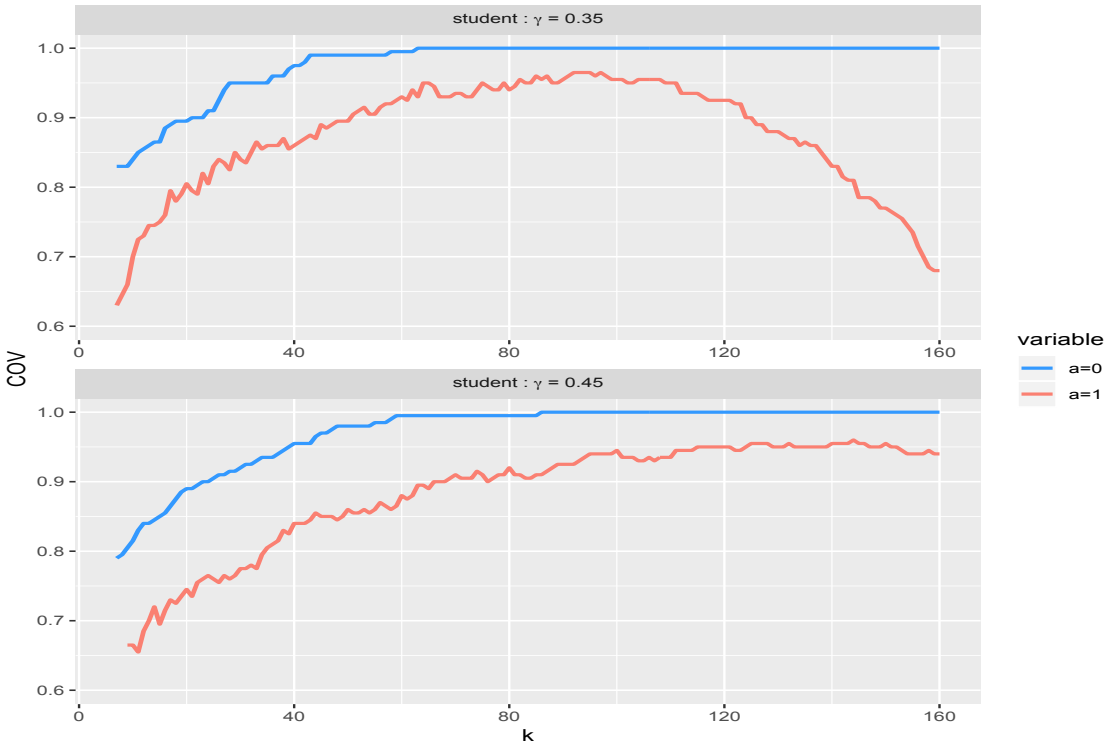


Figure 20: *Case of Student distribution—Achieved coverages of* $\widetilde{CI}_{0.95}(k)$ *against* k*, for* $\alpha = 0$ *in blue and* $\alpha = 1$ *in red. From top to bottom,* $\gamma = 0.35, 0.45$.

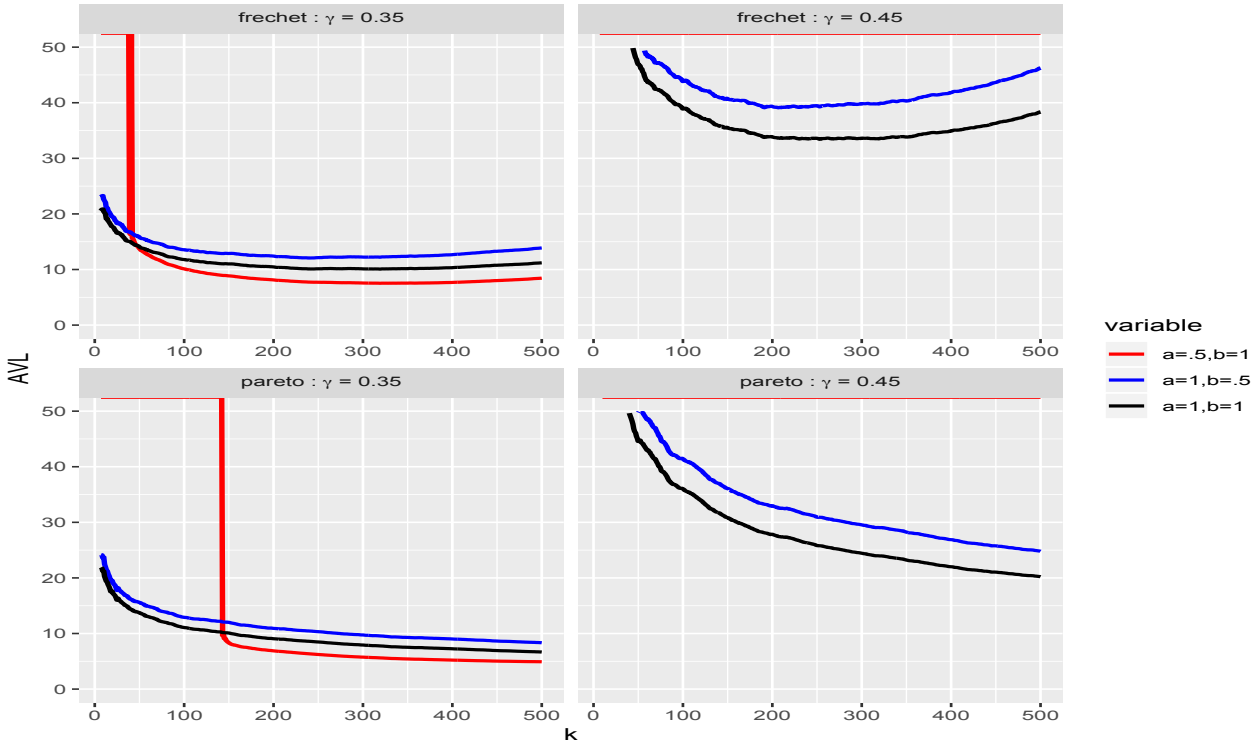


Figure 21: *Case of Fréchet and Pareto distributions—Average lengths of* $\overline{CI}_{0.95}(k)$ *against* k*, for* $(\alpha = 1, \beta = 0.5)$ *in blue and* $(\alpha = 0.5, \beta = 1)$ *in red, and of* $\widehat{CI}_{0.95}(k)$ *for* $(\alpha = 1, \beta = 1)$ *in black.*

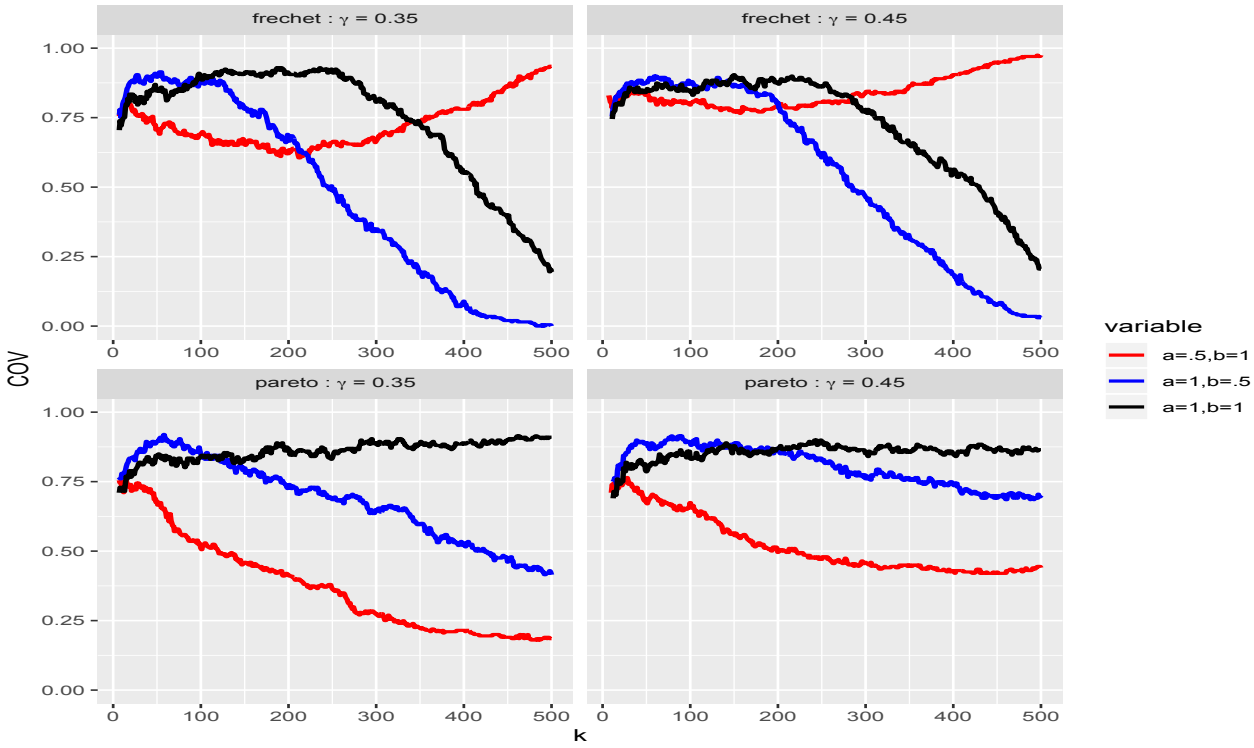


Figure 22: *Case of Fréchet and Pareto distributions—Achieved coverages against* k*.*

D Proofs

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

Proof of Theorem 2. To show (i), the main idea is to combine Theorem 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 Theorem 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 Theorem 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 3.

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

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

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.* (2018b) 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.* (2018b) 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

$$\log\left(\frac{\widetilde{\xi}_1}{\widetilde{\xi}_{1-k/n}}\right) = \mathrm{O}_{\mathbb{P}}(\log k). \tag{C.3}$$

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

$$\sqrt{k}|I_{n,1}| = \mathrm{O}_{\mathbb{P}}\left(s_n \times \sqrt{k}\log k\right) = \mathrm{o}_{\mathbb{P}}(1). \tag{C.4}$$

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 Theorem 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

$$\sqrt{k}(I_{n,2}-\gamma) \xrightarrow{d} \mathcal{N}\left(\frac{(1-\gamma)(\gamma^{-1}-1)^{-\rho}}{(1-\rho)(1-\gamma-\rho)}\lambda_1 - \mathbb{E}(Y)\frac{\gamma^2(\gamma^{-1}-1)^{\gamma}}{\gamma+1}\lambda_2, \frac{2\gamma^3}{1-2\gamma}\right). \tag{C.5}$$

Combining (C.2), (C.4) and (C.5) completes the proof of (i).

To show (ii), it suffices to prove that

$$|\breve{\gamma}_k - \widetilde{\gamma}_{k,l}| = \mathrm{O}_{\mathbb{P}}\left(\frac{\log(k)}{l}\right). \tag{C.6}$$

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 (C.3), which shows (C.6) and completes the proof. ∎

Proof of Theorem 3. 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 Theorem 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.* (2018b). Arguing as in the proof of Theorem 2 (*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 Theorem 1 that:

$$\sqrt{k}(\widehat{\gamma}_k - \gamma) = \frac{\lambda_1}{1-\rho} + \gamma\sqrt{\gamma^{-1}-1}\left(\int_0^1 \frac{1}{s} W_n\left(\frac{s}{\gamma^{-1}-1}\right) ds - W_n\left(\frac{1}{\gamma^{-1}-1}\right)\right) + \mathrm{o}_{\mathbb{P}}(1). \quad \text{(C.7)}$$

Besides, an inspection of the proof of Theorem 2 shows that

$$\begin{aligned} \sqrt{k}(\widetilde{\gamma}_k - \gamma) &= \frac{(1-\gamma)(\gamma^{-1}-1)^{-\rho}}{(1-\rho)(1-\gamma-\rho)}\lambda_1 - \mathbb{E}(Y)\frac{\gamma^2(\gamma^{-1}-1)^{\gamma}}{\gamma+1}\lambda_2 \\ &+ \gamma\sqrt{\gamma^{-1}-1}\int_0^1 \frac{W_n(s)}{s}([1-\gamma]s^{-\gamma}-1)ds + \mathrm{o}_{\mathbb{P}}(1) \end{aligned} \quad \text{(C.8)}$$

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

$$\sqrt{k}\left(\frac{\widehat{q}_{1-k/n}}{q_{1-k/n}} - 1\right) = \gamma\sqrt{\gamma^{-1}-1}\, W_n\left(\frac{1}{\gamma^{-1}-1}\right) + \mathrm{o}_{\mathbb{P}}(1) \quad \text{(C.9)}$$

as well as

$$\sqrt{k}\left(\frac{\widetilde{\xi}_{1-k/n}}{\xi_{1-k/n}} - 1\right) = \gamma^2\sqrt{\gamma^{-1}-1}\int_0^1 W_n(t)t^{-\gamma-1}dt + \mathrm{o}_{\mathbb{P}}(1). \quad \text{(C.10)}$$

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 (C.7). Meanwhile, Theorem 2(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 (C.7) and (C.8). 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

$$\begin{aligned}
\text{cov}_1 &= (\gamma^{-1}-1)^{\gamma-1} - \gamma[\gamma+1+\gamma\log(\gamma^{-1}-1)] \\
\text{and cov}_2 &= \gamma[(\gamma^{-1}-1)^{\gamma} - 1 - \gamma\log(\gamma^{-1}-1)].
\end{aligned} \tag{C.11}$$

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

$$\sqrt{k}\,(\widehat{\gamma}_k - \gamma, \widetilde{\gamma}_k - \gamma) \xrightarrow{d} \mathcal{N}(\boldsymbol{m}, \boldsymbol{V}) \tag{C.12}$$

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 3.

Marginal asymptotic behaviour of $\widehat{q}_{1-k/n}$: It is a straightforward byproduct of Equation (C.9) 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 (C.10) 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.* (2018b).

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 (C.8) and (C.9). By (C.11) above, this limiting covariance is:

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

with the notation of (C.11). 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 (C.7), (C.8) and (C.10) 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 (C.9) and (C.10), 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

$$\sqrt{k}\left(\overline{\gamma}_k(\alpha) - \gamma, \frac{\widehat{q}_{1-k/n}}{q_{1-k/n}} - 1, \frac{\widetilde{\xi}_{1-k/n}}{\xi_{1-k/n}} - 1\right) \xrightarrow{d} \mathcal{N}(\mathfrak{m}_\alpha, \mathfrak{V}_\alpha) \tag{C.13}$$

with $\mathfrak{m}_\alpha$ and $\mathfrak{V}_\alpha$ as in the statement of Theorem 3. This concludes the proof. ∎

Proof of Theorem 4. Applying Theorem 3 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 3, 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 5. 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

$$\frac{\sqrt{k}}{\log[k/(np_n)]} \log\left(\frac{\overline{\xi}_{1-k/n}(\alpha,\beta)}{\xi_{1-k/n}}\right) = \mathrm{O}_{\mathbb{P}}\left(1/\log[k/(np_n)]\right) = \mathrm{o}_{\mathbb{P}}(1) \tag{C.14}$$

$$\begin{aligned}
\text{and} \quad & \frac{\sqrt{k}}{\log[k/(np_n)]} \log\left(\left[\frac{np_n}{k}\right]^{\gamma} \frac{\xi_{1-p_n}}{\xi_{1-k/n}}\right) \\
&= \frac{\sqrt{k}}{\log[k/(np_n)]} \left(\log\left(\frac{\xi_{1-p_n}}{q_{1-p_n}}\right) - \log\left(\frac{\xi_{1-k/n}}{q_{1-k/n}}\right) + \log\left(\left[\frac{np_n}{k}\right]^{\gamma} \frac{q_{1-p_n}}{q_{1-k/n}}\right)\right) \\
&= \mathrm{O}\left(\frac{\sqrt{k}}{\log[k/(np_n)]} \left[\frac{1}{q_{1-k/n}} + |A(n/k)| + \frac{1}{q_{1-p_n}} + |A(1/p_n)|\right]\right) \\
&= \mathrm{O}\left(\frac{\sqrt{k}}{\log[k/(np_n)]} \left[\frac{1}{q_{1-k/n}} + |A(n/k)|\right]\right) \\
&= \mathrm{o}(1).
\end{aligned} \tag{C.15}$$

Here, convergence (C.14) is a consequence of Theorem 4. Convergence (C.15) follows from a combination of Proposition 1 in Daouia *et al.* (2018b), 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 6. The proof of the convergence of $\widetilde{\mathrm{XES}}^{\star}_{\tau'_n}(\alpha)$ is entirely similar to that of Theorem 5 (applying Theorem 6 in Daouia *et al.* (2018b) instead of Theorem 4, and Proposition 4 in Daouia *et al.* (2018b) instead of their Proposition 1). We omit the details.

We proceed by examining first 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 5 and the delta-method,

$$\frac{\sqrt{k}}{\log[k/(np_n)]} \log\left(\frac{\overline{\xi}^{\star}_{1-p_n}(\alpha,\beta)}{\xi_{1-p_n}}\right) \xrightarrow{d} \mathcal{N}(b_\alpha, v_\alpha). \tag{C.16}$$

Using then Theorem 3, the delta-method and the convergence $\log[k/(np_n)] \to \infty$, we get

$$\frac{\sqrt{k}}{\log[k/(np_n)]} \log\left(\frac{[1-\overline{\gamma}_{1-k/n}(\alpha)]^{-1}}{[1-\gamma]^{-1}}\right) \xrightarrow{\mathbb{P}} 0. \tag{C.17}$$

Using finally a combination of Propositions 1(i) and 4 in Daouia *et al.* (2018b) and the regular variation of $|A|$ and $t \mapsto q_{1-t^{-1}}$, we obtain

$$\frac{\sqrt{k}}{\log[k/(np_n)]} \log\left(\frac{\mathrm{XES}_{1-p_n}}{[1-\gamma]^{-1}\xi_{1-p_n}}\right) \to 0. \tag{C.18}$$

Combining convergences (C.16), (C.17) and (C.18), 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 $\overline{\mathrm{XES}}^{\star}_{1-p_n}(\alpha,\beta)$.

We now show the convergence of $\widehat{\mathrm{XES}}^{\star}_{1-p_n}(\alpha,\beta)$. For this we write

$$\begin{aligned}\log\left(\frac{\widehat{\mathrm{XES}}^{\star}_{1-p_n}(\alpha,\beta)}{\mathrm{XES}_{1-p_n}}\right) &= \log\left(\frac{\overline{\xi}^{\star}_{1-p_n}(\alpha,\beta)}{\xi_{1-p_n}}\right) + \log\left(\frac{\widehat{\mathrm{QES}}_{1-k/n}}{\widehat{q}_{1-k/n}} \cdot \frac{q_{1-k/n}}{\mathrm{QES}_{1-k/n}}\right) \\ &\quad + \log\left(\frac{\mathrm{QES}_{1-k/n}}{q_{1-k/n}}\right) - \log\left(\frac{\mathrm{XES}_{1-p_n}}{\xi_{1-p_n}}\right)\end{aligned}$$

where we set

$$\widehat{\mathrm{QES}}_{1-k/n} := \frac{1}{\lfloor k \rfloor} \sum_{i=1}^{\lfloor k \rfloor} Y_{n-i+1,n} = \int_0^1 \widehat{q}_{1-\lfloor k \rfloor s/n}\, ds.$$

Remark now that, since $\widehat{q}_{1-\lfloor k \rfloor/n} = Y_{n-\lfloor k \rfloor,n} = \widehat{q}_{1-k/n}$, we have

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

Combine then Theorem 1, the delta-method, and Equation (B.20) in the Supplementary Material document of Daouia *et al.* (2018b) together with a Taylor expansion to obtain

$$\frac{\sqrt{k}}{\log[k/(np_n)]} \log\left(\frac{\widehat{\mathrm{QES}}_{1-k/n}}{\widehat{q}_{1-k/n}} \cdot \frac{q_{1-k/n}}{\mathrm{QES}_{1-k/n}}\right) = \mathrm{O}_{\mathbb{P}}(1/\log[k/(np_n)]) = \mathrm{o}_{\mathbb{P}}(1). \tag{C.19}$$

Besides, a combination of Equation (B.20) in the Supplementary Material document of Daouia *et al.* (2018b) and Proposition 4 in Daouia *et al.* (2018b) with a Taylor expansion yields

$$\begin{aligned}&\frac{\sqrt{k}}{\log[k/(np_n)]} \left[\log\left(\frac{\mathrm{QES}_{1-k/n}}{q_{1-k/n}}\right) - \log\left(\frac{\mathrm{XES}_{1-p_n}}{\xi_{1-p_n}}\right)\right] \\ &= \mathrm{O}\left(\frac{\sqrt{k}}{\log[k/(np_n)]} \left[\frac{1}{q_{1-k/n}} + |A(n/k)| + \frac{1}{q_{1-p_n}} + |A(1/p_n)|\right]\right) \\ &= \mathrm{O}\left(\frac{\sqrt{k}}{\log[k/(np_n)]} \left[\frac{1}{q_{1-k/n}} + |A(n/k)|\right]\right) \\ &= \mathrm{o}(1).\end{aligned} \tag{C.20}$$

Finally, use together (C.16), (C.19) and (C.20) and the delta-method to complete the proof. ■

Proof of Theorem 7. We only show the result for $\widetilde{\mathrm{XES}}^{\star}_{\widehat{\tau}'_n(p_n)}(\alpha)$ as the proofs of the other convergences are entirely similar. The key point is to write

$$\widetilde{\mathrm{XES}}^{\star}_{\widehat{\tau}'_n(p_n)}(\alpha) = \left(\frac{1-\widehat{\tau}'_n(p_n)}{1-\tau'_n(p_n)}\right)^{-\overline{\gamma}_{\tau_n}(\alpha)} \widetilde{\mathrm{XES}}^{\star}_{\tau'_n(p_n)}(\alpha). \tag{C.21}$$

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.* (2018b); this also results in a corrected version of (B.51) in the former paper]. Therefore

$$\begin{aligned}
\left(\frac{1-\widehat{\tau}'_n(p_n)}{1-\tau'_n(p_n)}\right)^{-\overline{\gamma}_{\tau_n}(\alpha)} &= \exp\left(-\overline{\gamma}_{\tau_n}(\alpha)\log\left[\frac{1-\widehat{\tau}'_n(p_n)}{1-\tau'_n(p_n)}\right]\right) \\
&= \exp\left(-\left[\gamma + \mathrm{O}_{\mathbb{P}}\left(\frac{1}{\sqrt{n(1-\tau_n)}}\right)\right] \times \mathrm{O}_{\mathbb{P}}\left(\frac{1}{\sqrt{n(1-\tau_n)}}\right)\right) \\
&= 1 + \mathrm{O}_{\mathbb{P}}\left(\frac{1}{\sqrt{n(1-\tau_n)}}\right).
\end{aligned} \tag{C.22}$$

Furthermore, using the equivalent

$$1-\tau'_n(p_n) \sim (1-p_n)\,\frac{\gamma}{1-\gamma}, \tag{C.23}$$

we conclude that the conditions of Theorem 6 are satisfied if the parameter τ'_n there is set equal to $\tau'_n(p_n)$. By Theorem 6 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 (C.23). 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

$$\frac{\sqrt{n(1-\tau_n)}}{\log[(1-\tau_n)/(1-p_n)]}\left(\frac{\widetilde{\mathrm{XES}}^{\star}_{\tau'_n(p_n)}(\alpha)}{\mathrm{XES}_{\tau'_n(p_n)}}-1\right)\xrightarrow{d}\mathcal{N}(b_\alpha,v_\alpha). \tag{C.24}$$

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.* (2018b) with (C.23) 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.* (2018b),

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

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 (C.24), this entails

$$\frac{\sqrt{n(1-\tau_n)}}{\log[(1-\tau_n)/(1-p_n)]}\left(\frac{\widetilde{\mathrm{XES}}^{\star}_{\tau'_n(p_n)}(\alpha)}{\mathrm{QES}_{p_n}}-1\right)\xrightarrow{d}\mathcal{N}(b_\alpha,v_\alpha). \tag{C.25}$$

Combining (C.21), (C.22) and (C.25) completes the proof. ■

References

Beirlant, J., Goegebeur, Y., Segers, J. and Teugels, J. (2004). *Statistics of Extremes: Theory and Applications*, Wiley.

Cai, J.J., Einmahl, J.H.J., de Haan, L. and Zhou, C. (2015). Estimation of the marginal expected shortfall: the mean when a related variable is extreme, *Journal of the Royal Statistical Society: Series B* **77**: 417–442.

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. (2018b). Tail expectile process and risk assessment, preprint. Available at `https://hal.archives-ouvertes.fr/hal-01744505`.

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.