Supplementary Material for
"**Estimation of Tail Risk based on Extreme Expectiles**"

Abdelaati Daouia, Stéphane Girard and Gilles Stupfler

Further simulation results are discussed in Section A. The proofs of all theoretical results in the main paper and additional technical results are provided in Section B.

A Additional simulations

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

(A.1) Bias and MSE estimates.

(A.2) Forecast verification and validation.

(A.3) Quality of asymptotic approximations.

Let us first comment on some implementation details. We used in all our simulations the Hill estimator of γ, the extreme level $\tau_n' = 0.995$ for $n = 100$ and $\tau_n' = 0.9994$ for $n = 1000$. The corresponding true extreme expectiles $\xi_{\tau_n'}$ can be calculated by the existing function "et(τ_n', df)" in the R package *'expectreg'*. In what concerns the intermediate levels τ_n involved in both estimators $\widetilde{\xi}^{\star}_{\tau_n'} \equiv \widetilde{\xi}^{\star}_{\tau_n'}(\tau_n)$ and $\widehat{\xi}^{\star}_{\tau_n'} \equiv \widehat{\xi}^{\star}_{\tau_n'}(\tau_n)$, we used the same considerations as in Ferreira et al. (2003). Namely, they always considered $\tau_n = 1 - \frac{k}{n}$ with the range of intermediate integers k, say, from $\log(n^{1-\varepsilon})$ to $n/\log(n^{1-\varepsilon})$, where $\varepsilon = 0.1$ [this restriction allows to reject too small values or those very near $n^{1-\varepsilon}$]. The value k can actually be viewed as the effective sample size for tail extrapolation. A larger k leads to estimators with more bias, while smaller k results in higher variance.

A.1 Bias and MSE estimates

Figures 1 and 2 (respectively, Figures 3 and 4) give the root-MSE estimates computed over $10,000$ replications for samples of size 100 and 1000 simulated from the Student (respectively, positive Student) t-models, while Figures 5 and 6 (respectively, Figures 7 and 8) give the bias estimates for the same models. Each figure displays the evolution of the obtained Monte-Carlo results, for the two normalized estimators $\widetilde{\xi}^{\star}_{\tau_n'}(k)/\xi_{\tau_n'}$ and $\widehat{\xi}^{\star}_{\tau_n'}(k)/\xi_{\tau_n'}$, as functions of the sample fraction k. Tables 1 and 2 report the root-MSE and bias estimates obtained by using for each estimator the optimal value of k minimizing its MSE.

As regards the Student distributions which correspond to real-valued profit-loss variables, our tentative conclusion from Figures 1-2 and Figures 5-6 is that the indirect estimator $\widehat{\xi}^{\star}_{\tau_n'}$ has a harder time with small samples, and this can be compensated by taking larger samples. Indeed, for $n = 100$, the direct estimator $\widetilde{\xi}^{\star}_{\tau_n'}$ performs better than $\widehat{\xi}^{\star}_{\tau_n'}$ in terms of both MSE and bias, whatever the thickness of the tails. Also, in contrast to the direct estimator's plot, the indirect one exhibits more volatility. In

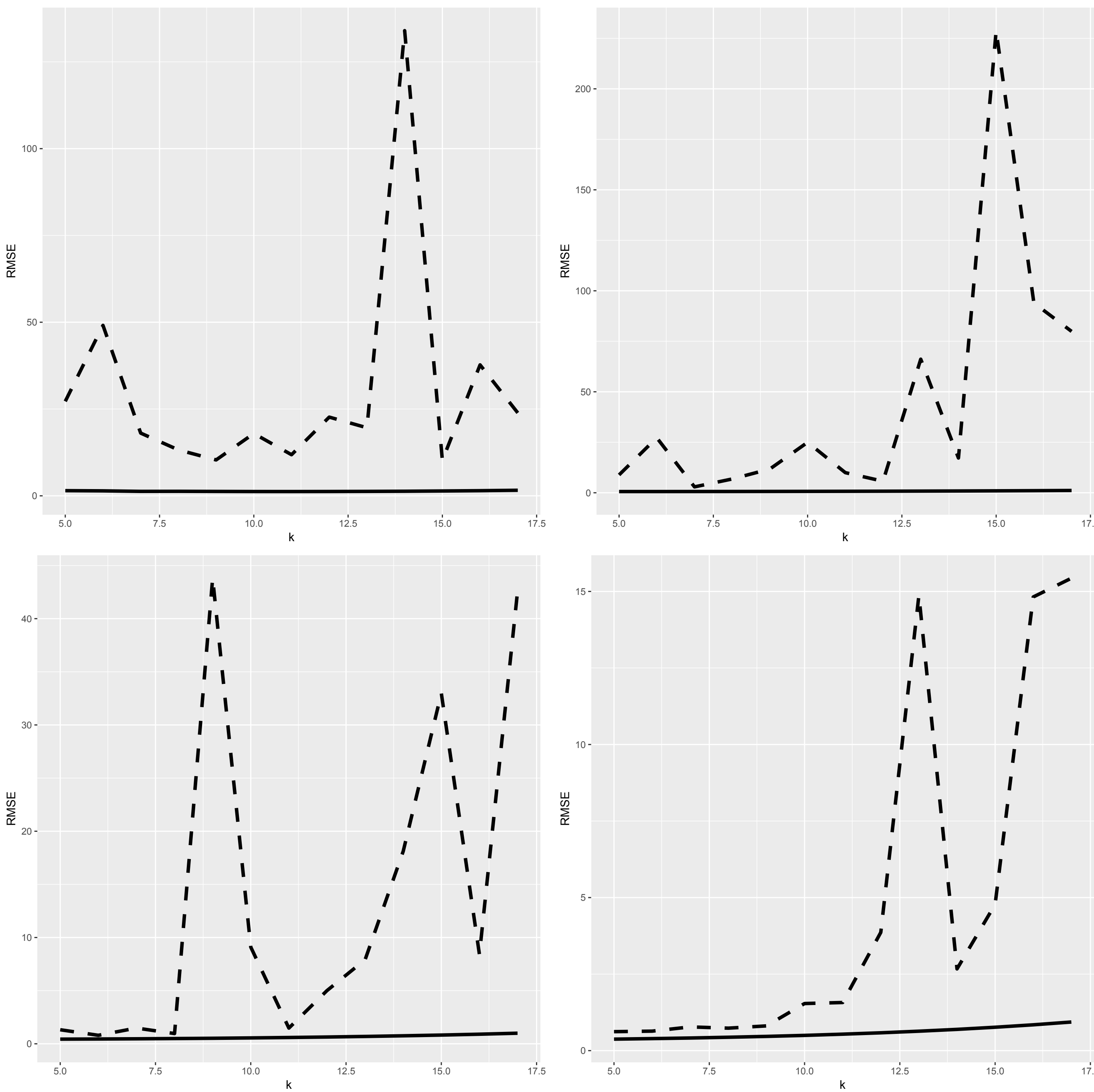

Figure 1: *Root MSE estimates of* $\widetilde{\xi}^{\star}_{\tau_n'}(k)/\xi_{\tau_n'}$ *(solid line) and* $\widehat{\xi}^{\star}_{\tau_n'}(k)/\xi_{\tau_n'}$ *(dashed line), as functions of* k*, for the* t_3*,* t_5*,* t_7 *and* t_9*-distributions, respectively, from top to bottom and from left to right. Results for the sample size* $n = 100$*.*

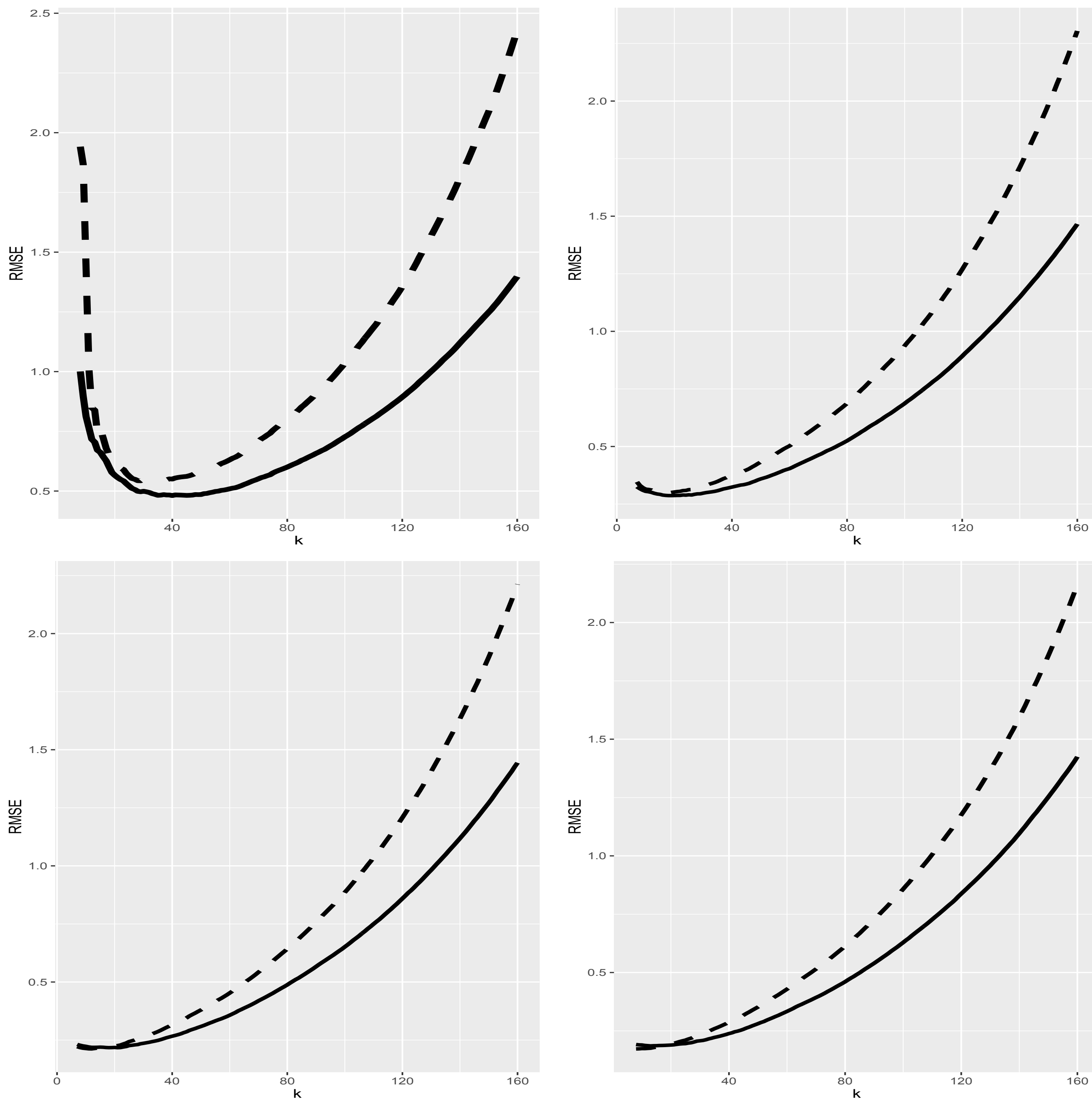

Figure 2: *As before—Results for the sample size* $n = 1000$.

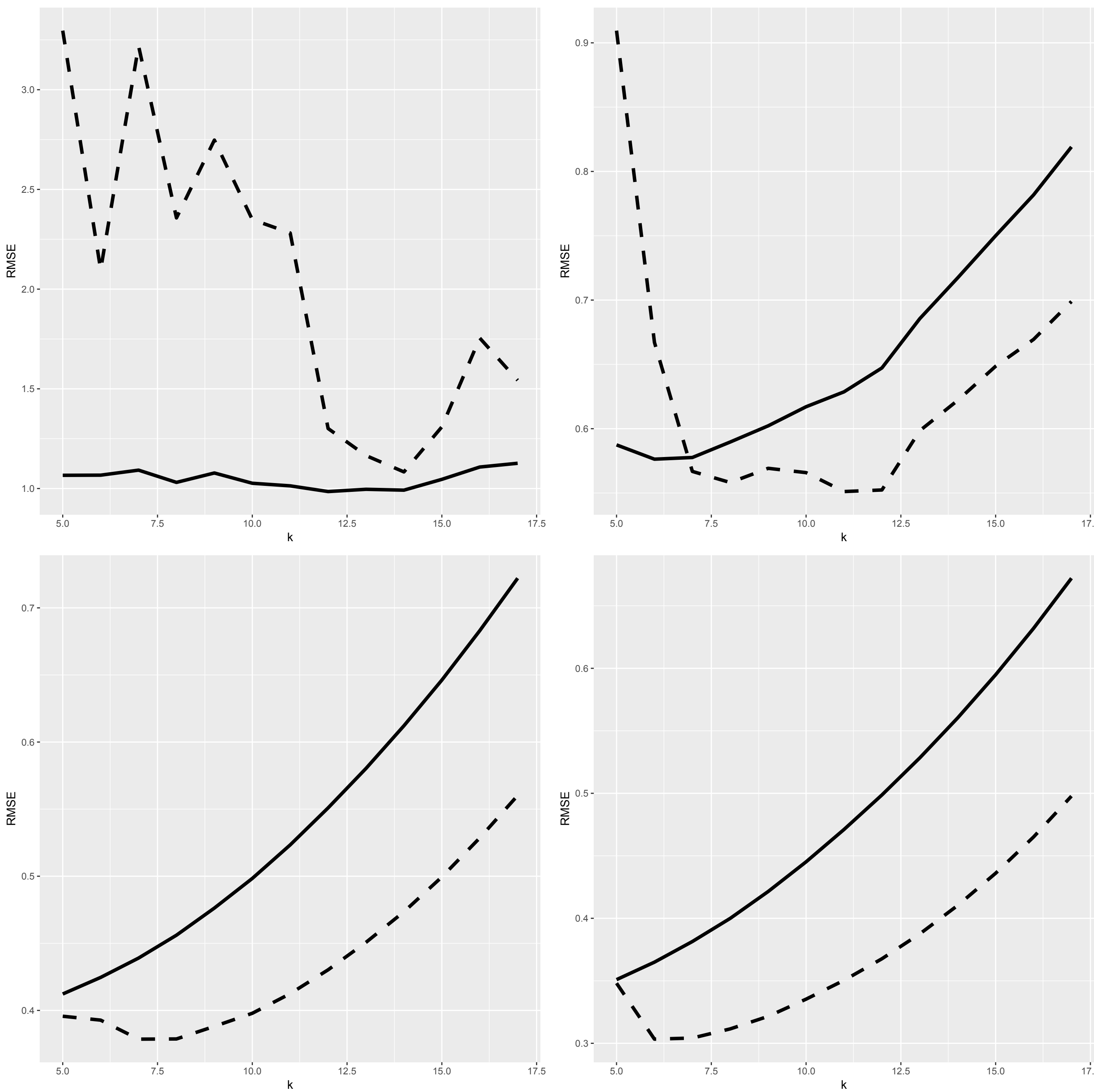

Figure 3: *Root MSE estimates of* $\widetilde{\xi}^{\star}_{\tau_n'}(k)/\xi_{\tau_n'}$ *(solid line) and* $\widehat{\xi}^{\star}_{\tau_n'}(k)/\xi_{\tau_n'}$ *(dashed line), as functions of* k*, for the positive Student* t_3*,* t_5*,* t_7 *and* t_9*-distributions, respectively, from top to bottom and from left to right. Results for the sample size* $n = 100$*.*

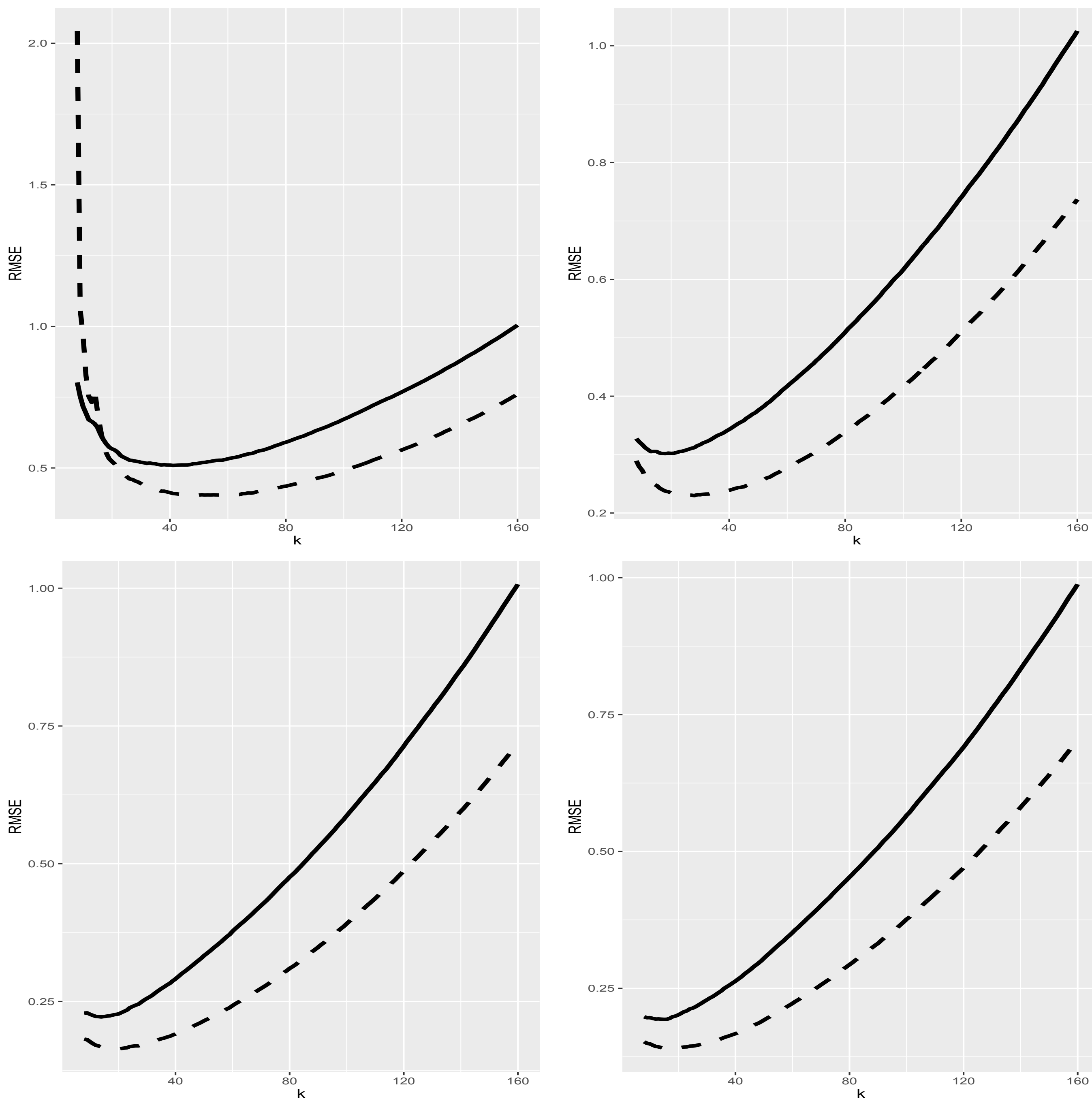

Figure 4: *As before—Results for the sample size* $n = 1000$.

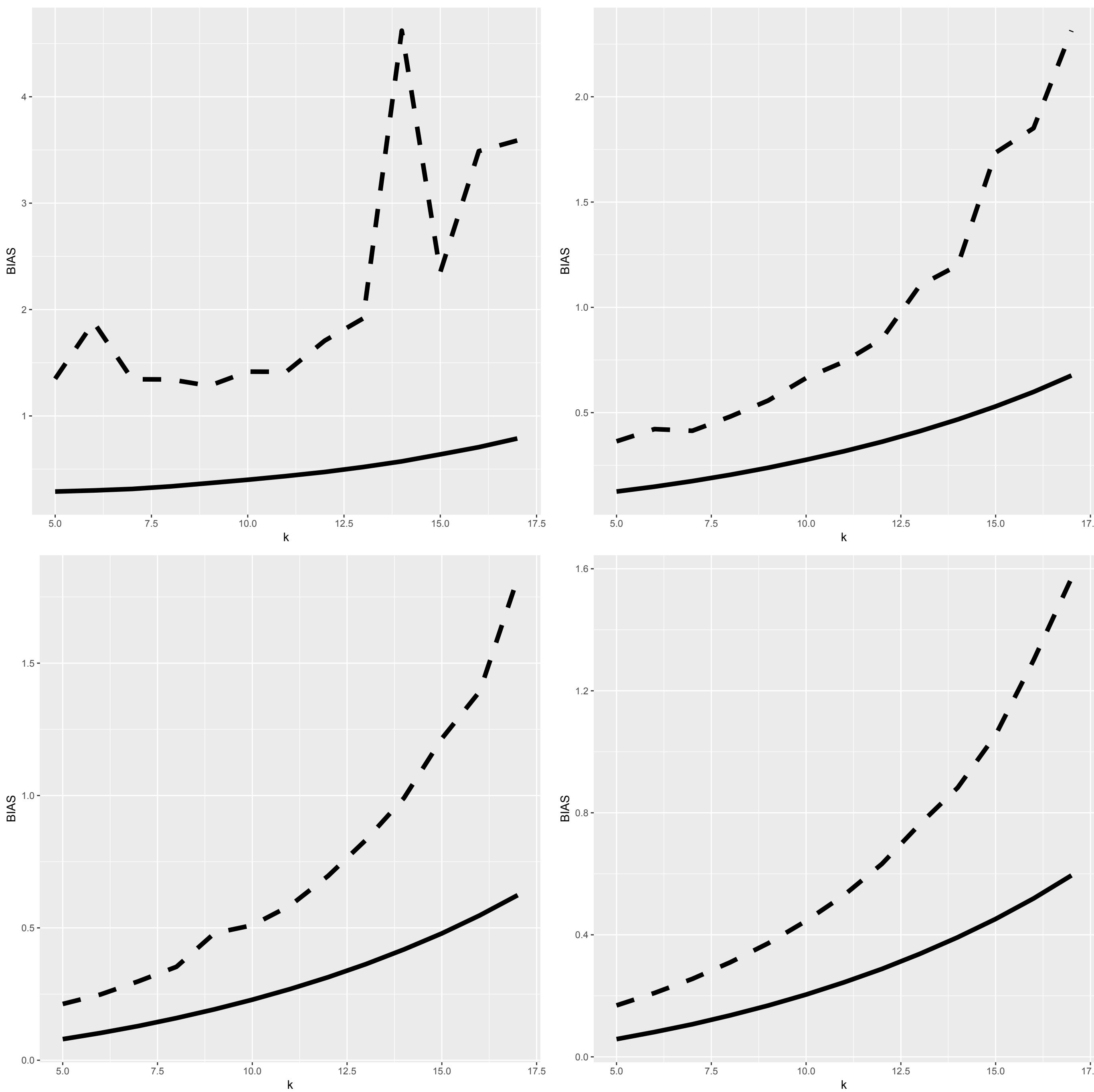

Figure 5: *Bias estimates of* $\widetilde{\xi}^{\star}_{\tau_n'}(k)/\xi_{\tau_n'}$ *(solid line) and* $\widehat{\xi}^{\star}_{\tau_n'}(k)/\xi_{\tau_n'}$ *(dashed line), as functions of* k*, for the* t_3*,* t_5*,* t_7 *and* t_9*-distributions, respectively, from top to bottom and from left to right. Results for the sample size* $n = 100$*.*

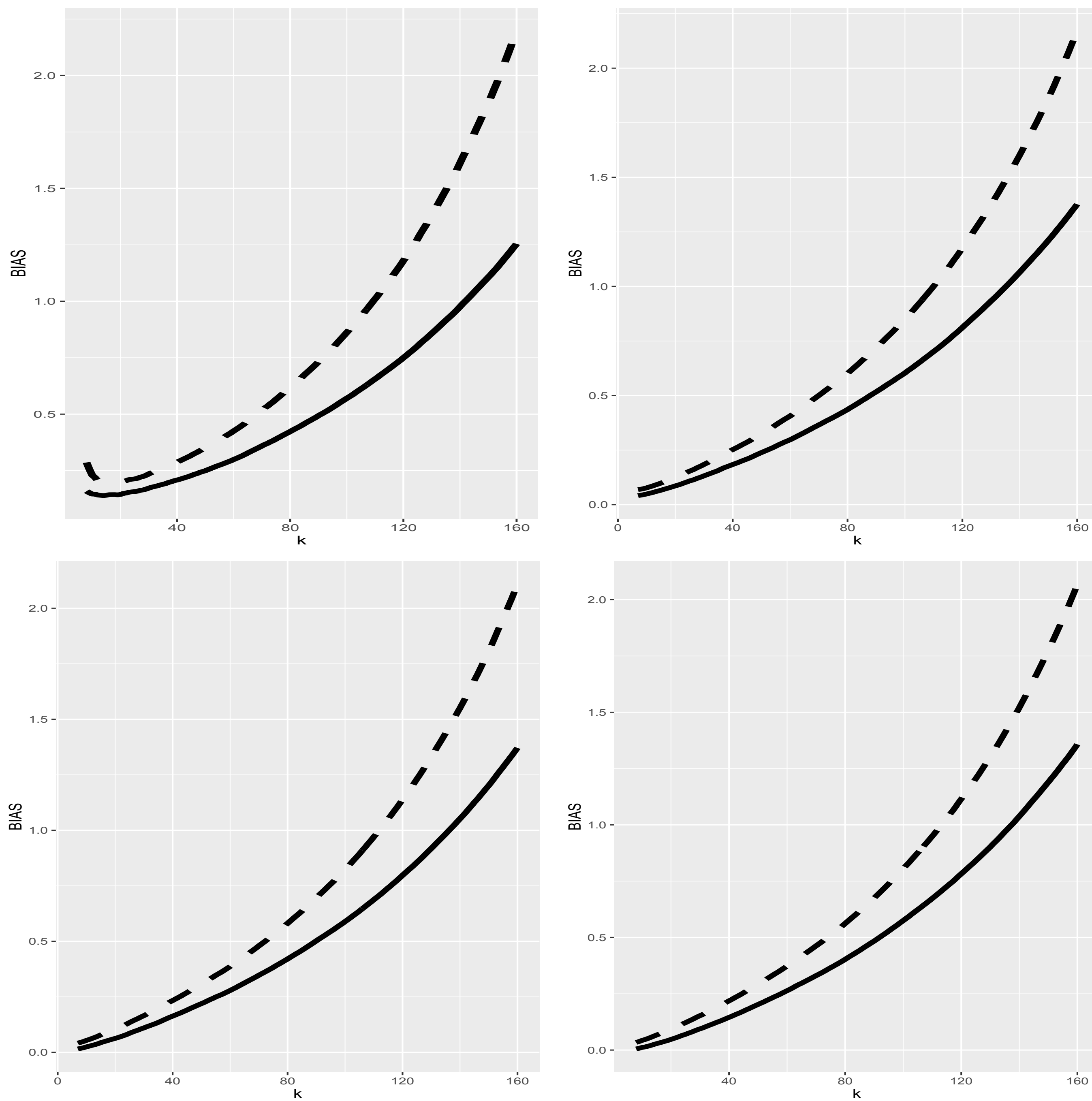

Figure 6: *As before—Results for the sample size* $n = 1000$.

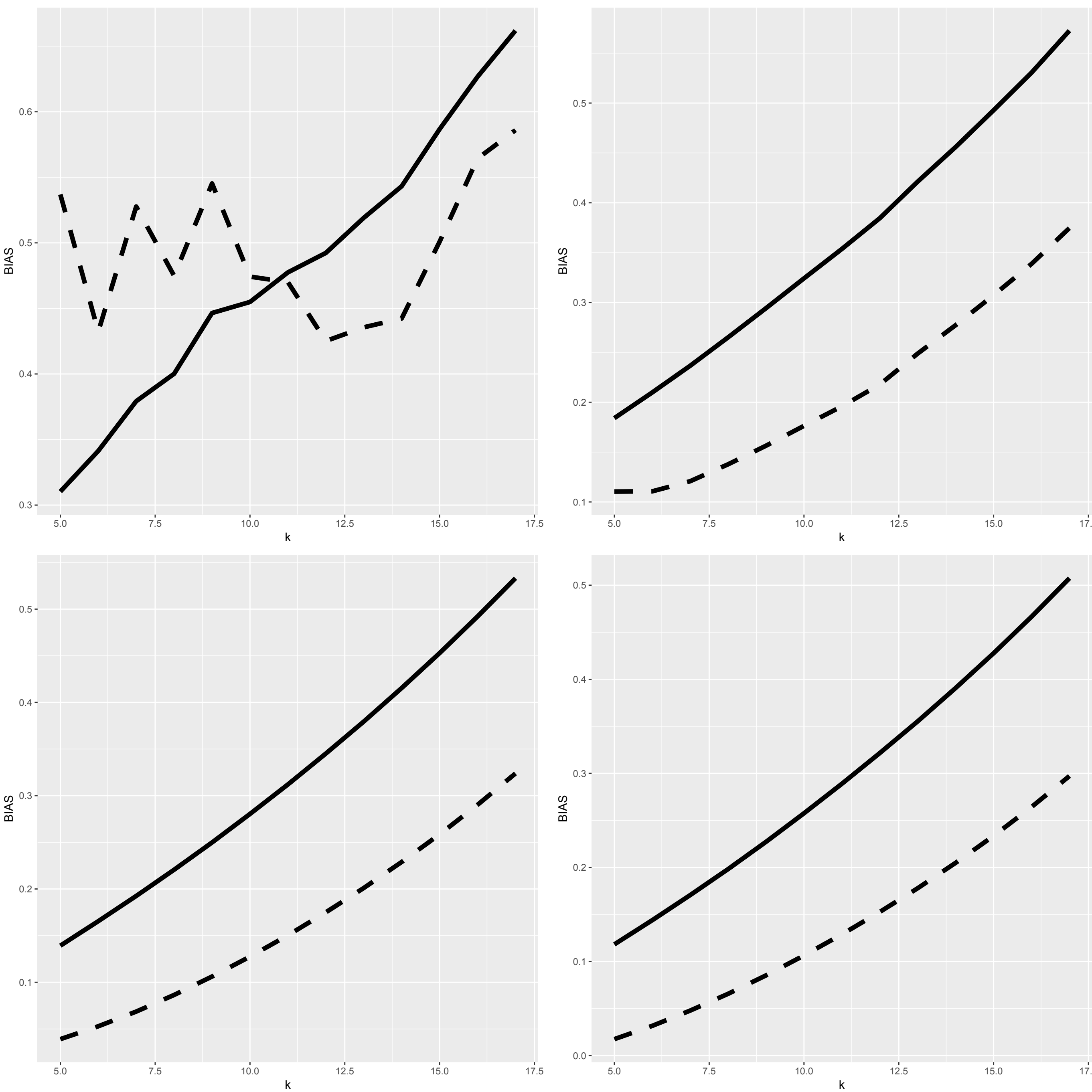

Figure 7: *Bias estimates of* $\widetilde{\xi}^{\star}_{\tau'_n}(k)/\xi_{\tau'_n}$ *(solid line) and* $\widehat{\xi}^{\star}_{\tau'_n}(k)/\xi_{\tau'_n}$ *(dashed line), as functions of* k*, for the positive Student* t_3*,* t_5*,* t_7 *and* t_9*-distributions, respectively, from top to bottom and from left to right. Results for the sample size* $n = 100$*.*

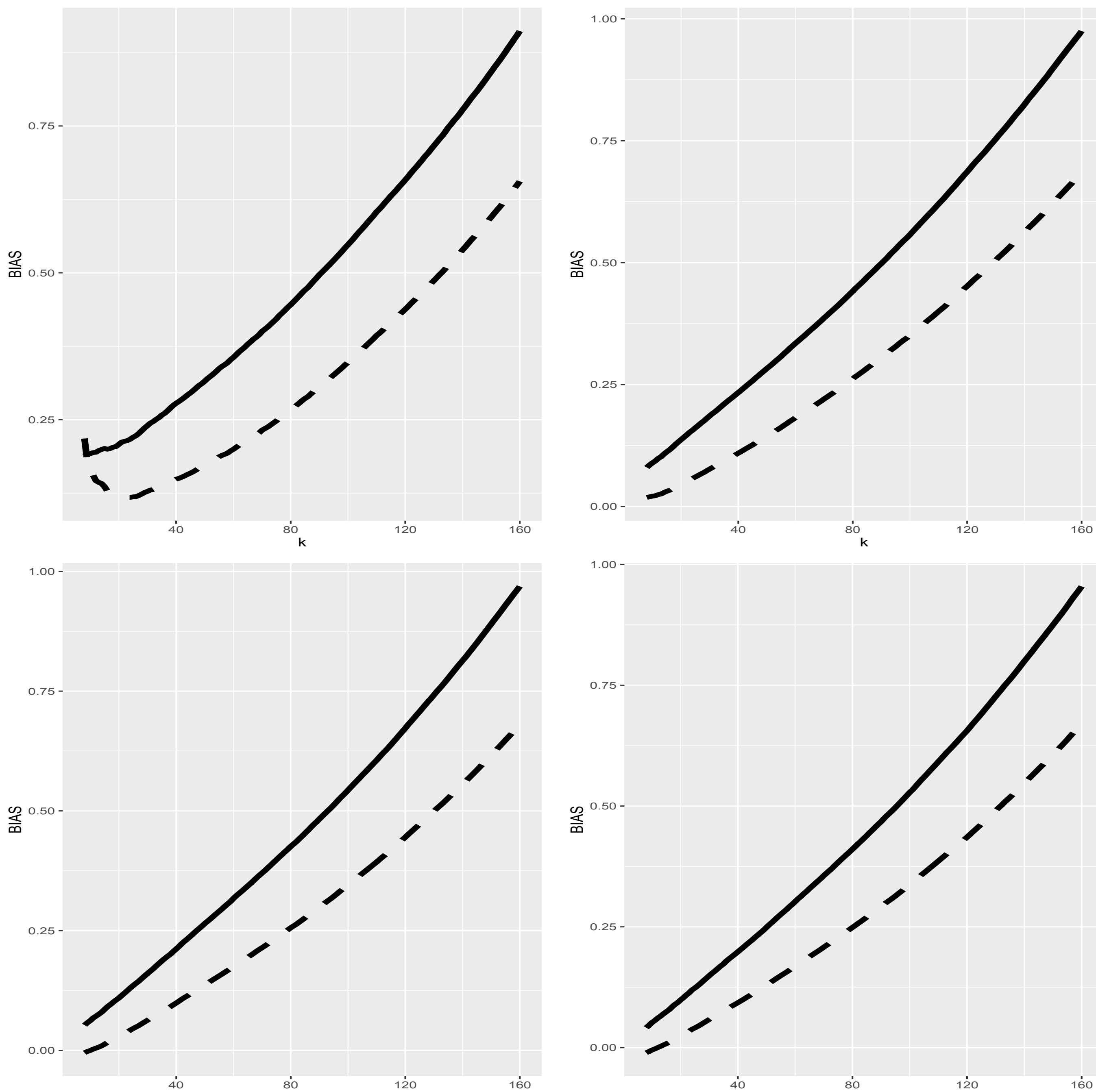

Figure 8: *As before—Results for the sample size* $n = 1000$.

what concerns $n = 1000$, it seems that $\widehat{\xi}^{\star}_{\tau'_n}$ is superior to $\widetilde{\xi}^{\star}_{\tau'_n}$ only in terms of MSE for slightly heavy tails (*i.e.* $df = 7, 9$), whereas the accuracy of $\widetilde{\xi}^{\star}_{\tau'_n}$ is more respectable for heavier tails (*i.e.* $df = 3, 5$), as can be seen from Table 1. It should be, however, clear that even in the favorable case to $\widehat{\xi}^{\star}_{\tau'_n}$, where $n = 1000$ and $df \in \{7, 9\}$, the estimator $\widetilde{\xi}^{\star}_{\tau'_n}$ has actually almost overall a smaller MSE except for a very small zone of values of k, as can be seen from Figure 2 (bottom panels). Due to the tightness of that zone, the detection of the optimal k which minimizes the MSE of $\widehat{\xi}^{\star}_{\tau'_n}$ is hard to manage in practice.

By contrast, in the case of the positive Student distributions which correspond to non-negative loss variables, it can be seen from Figures 3-4 and Figures 7-8 as well as Table 2 that the indirect estimator $\widehat{\xi}^{\star}_{\tau'_n}$ is superior to the direct estimator $\widetilde{\xi}^{\star}_{\tau'_n}$ in all scenarios except for the single case $n = 100$ and $df = 3$. We repeated this kind of exercise with the Fréchet distribution $F(y) = e^{-y^{-1/\gamma}}$, $y > 0$, and Pareto distribution $F(y) = 1 - y^{-1/\gamma}$, $y > 1$, and arrived at the same tentative conclusion.

It may also be seen that most of the error is due to variance, the squared bias being much smaller in all cases. It is interesting that in almost all cases the bias was positive. This may be explained by the sensitivity of high expectiles to the magnitude of heavy tails, since they are based on "squared" error loss minimization.

$n = 100$

	RMSE		BIAS	
df	$\widetilde{\xi}^{\star}_{\tau'_n}$	$\widehat{\xi}^{\star}_{\tau'_n}$	$\widetilde{\xi}^{\star}_{\tau'_n}$	$\widehat{\xi}^{\star}_{\tau'_n}$
3	1.5010	47.9486	0.4888	1.7107
5	0.5963	2.9132	0.1253	0.4139
7	0.4385	0.8001	0.0797	0.2486
9	0.3753	0.6200	0.0579	0.1685

$n = 1000$

	RMSE		BIAS	
df	$\widetilde{\xi}^{\star}_{\tau'_n}$	$\widehat{\xi}^{\star}_{\tau'_n}$	$\widetilde{\xi}^{\star}_{\tau'_n}$	$\widehat{\xi}^{\star}_{\tau'_n}$
3	0.4809	0.5403	0.2080	0.2599
5	0.2867	0.2981	0.0816	0.1088
7	0.2172	0.2119	0.0666	0.0629
9	0.1908	0.1781	0.0271	0.0440

Table 1: *Monte-Carlo results obtained for the Student t_3, t_5, t_7 and t_9-distributions, using the optimal sample fraction k minimizing the MSE of each estimator.*

$n = 100$

	RMSE		BIAS	
df	$\widetilde{\xi}^{\star}_{\tau'_n}$	$\widehat{\xi}^{\star}_{\tau'_n}$	$\widetilde{\xi}^{\star}_{\tau'_n}$	$\widehat{\xi}^{\star}_{\tau'_n}$
3	0.9848	1.0833	0.4923	0.4423
5	0.5762	0.5511	0.2098	0.1959
7	0.4122	0.3786	0.1392	0.0685
9	0.3509	0.3033	0.1181	0.0315

$n = 1000$

	RMSE		BIAS	
df	$\widetilde{\xi}^{\star}_{\tau'_n}$	$\widehat{\xi}^{\star}_{\tau'_n}$	$\widetilde{\xi}^{\star}_{\tau'_n}$	$\widehat{\xi}^{\star}_{\tau'_n}$
3	0.5089	0.4023	0.2810	0.1932
5	0.3016	0.2297	0.1269	0.0697
7	0.2219	0.1639	0.0800	0.0280
9	0.1934	0.1393	0.0742	0.0107

Table 2: *Monte-Carlo results obtained for the positive Student t_3, t_5, t_7 and t_9-distributions, using the optimal sample fraction k minimizing the MSE of each estimator.*

A.2 Forecast verification and validation

Another way of validating the presented estimation procedures for the extreme risk measure $\xi_{\tau'_n}$ on historical data is by using the elicitability property of expectiles as pointed out in Section 1. Following the ideas of Gneiting (2011), the competing estimates $\widehat{\xi}^{\star}_{\tau'_n}$ and $\widetilde{\xi}^{\star}_{\tau'_n}$ can be compared from a forecasting perspective by means of the consistent loss function

$$L_{\tau'_n} : (\xi, r) \mapsto L_{\tau'_n}(\xi, r) = \eta_{\tau'_n}(r - \xi)$$

which represents the penalty when the point forecast $\xi \in \mathbb{R}$ is issued and the observation $r \in \mathbb{R}$ materializes, with $\eta_{\tau'_n}(y) = |\tau'_n - \mathbb{I}\{y \leqslant 0\}|y^2$ being the expectile check function. For a given simulated series of size $N = 1500$, the estimates $\widehat{\xi}^{\star}_{\tau'_n}(k)$ and $\widetilde{\xi}^{\star}_{\tau'_n}(k)$ are computed on rolling windows of length $n = 1000$, for each sample fraction k. This corresponds to $T = N - n = 500$ forecast cases with corresponding point forecasts $\left(\xi_1^{(m)}(k), \ldots, \xi_T^{(m)}(k)\right)_{m=1,2}$ and realizing observations $(r_1, \ldots, r_T)$, where $\xi_t^{(1)}(k) = \widehat{\xi}^{\star}_{\tau'_n}(k)$ and $\xi_t^{(2)}(k) = \widetilde{\xi}^{\star}_{\tau'_n}(k)$ for each $t = 1, \ldots, T$. The two competing forecast procedures can be ranked by computing their realized losses

$$\mathcal{L}^{(m)}_{\tau'_n}(k) = \frac{1}{T}\sum_{t=1}^{T} L_{\tau'_n}\left(\xi_t^{(m)}(k), r_t\right),$$

for each $m = 1, 2$, and each sample fraction k (the lower the better). Figures 9 and 10 display the averages of the two realized losses computed over 200 simulated series from, respectively, the Student and positive Student t_3, t_5, t_7 and t_9-distributions. In Figure 11 we considered 200 simulated series from a Garch(1,1) model with Student t innovations [more sophisticated econometric models for expectiles have been pursued in Taylor (2008), Kuan *et al.* (2009) and De Rossi and Harvey (2009)]. The resulting average values of the realized loss seem to favor the direct forecaster $\widetilde{\xi}^{\star}_{\tau'_n}$ in the case of Student t-distributions (Figures 9 and 11), while they tend to prefer the rival indirect forecaster $\widehat{\xi}^{\star}_{\tau'_n}$ in the case of positive Student t-distributions (Figure 10).

A.3 Quality of asymptotic approximations

We first investigate the normality of the estimators $\widehat{\xi}^{\star}_{\tau'_n}$ and $\widetilde{\xi}^{\star}_{\tau'_n}$. The asymptotic normality of $\widehat{\xi}^{\star}_{\tau'_n}/\xi_{\tau'_n}$ in Corollary 3 can be expressed as $r_n \log(\widehat{\xi}^{\star}_{\tau'_n}/\xi_{\tau'_n}) \xrightarrow{d} \Gamma$, with $r_n = \frac{\sqrt{k}}{\log[k/(n(1-\tau'_n))]}$. Likewise, the asymptotic normality of $\widetilde{\xi}^{\star}_{\tau'_n}/\xi_{\tau'_n}$ in Corollary 4 can be expressed as $r_n \log(\widetilde{\xi}^{\star}_{\tau'_n}/\xi_{\tau'_n}) \xrightarrow{d} \Gamma$. The limit distribution Γ of the Hill estimator is $\mathcal{N}(\lambda_2/(1-\rho), \gamma^2)$, as pointed out below Theorem 1. It can be shown that the Student t_ν distributions satisfy the conditions of the two aforementioned corollaries, with $\gamma = 1/\nu$, $\rho = -2/\nu$ and

$$A(t) \sim \frac{\nu+1}{\nu+2}(c_\nu t)^{-2/\nu}, \quad c_\nu = \frac{2\Gamma((\nu+1)/2)\nu^{(\nu-1)/2}}{\sqrt{\nu\pi}\Gamma(\nu/2)}.$$

Hence, we can compare the distributions of

$$\widehat{W}_n := \left[r_n \log(\widehat{\xi}^{\star}_{\tau'_n}/\xi_{\tau'_n}) - \lambda_2/(1-\rho)\right]/\gamma \quad \text{and} \quad \widetilde{W}_n := \left[r_n \log(\widetilde{\xi}^{\star}_{\tau'_n}/\xi_{\tau'_n}) - \lambda_2/(1-\rho)\right]/\gamma$$

with the limit distribution $\mathcal{N}(0,1)$, with $\lambda_2 = \sqrt{k}A(\frac{n}{k})$. The Q–Q-plots in Figures 12 and 13 present, respectively, the sample quantiles of $\widehat{W}_n$ and $\widetilde{W}_n$, based on $10,000$ simulated samples of size $n = 1000$, versus the theoretical standard normal quantiles. For each estimator, we used the optimal value of k that minimizes its MSE as in Table 1. It may be seen that the scatters for the Student t_ν distributions, with $\nu = 3, 5, 7, 9$ displayed respectively from top to bottom and from left to right, are quite encouraging especially for the LAWS estimator $\widetilde{\xi}^{\star}_{\tau'_n}$ (Figure 13). Likewise, we conclude from the scatters for the positive Student t_ν distributions, displayed in Figures 14 and 15, that the limit Theorem 3 and its Corollaries 3 and 4 provide adequate approximations for finite sample sizes.

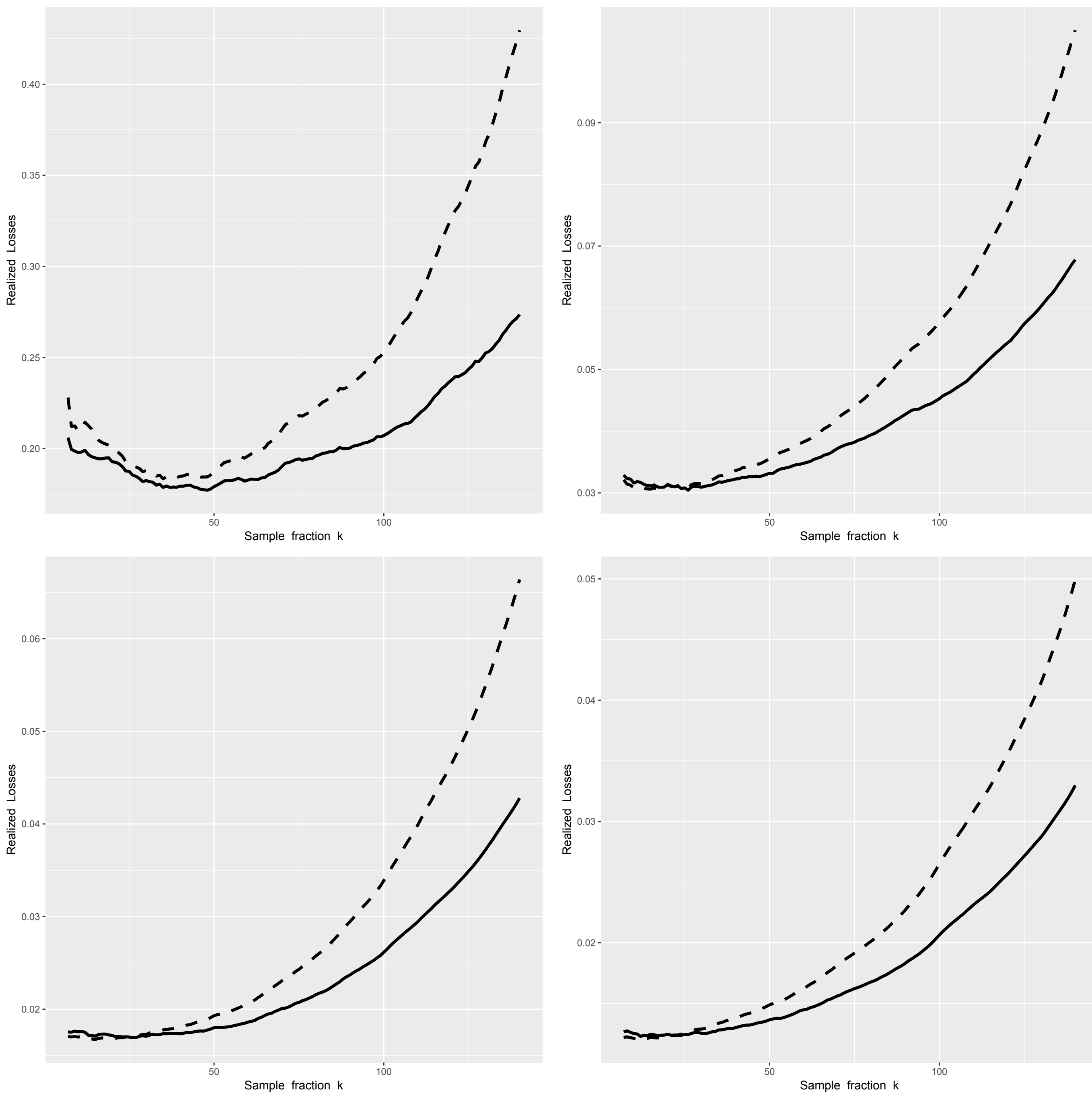

Figure 9: *The average values of the realized loss* $\mathcal{L}^{(m)}_{\tau'_n}(k)$ *for both estimators* $\widehat{\xi}^{\star}_{\tau'_n}(k)$ *in dashed line and* $\widetilde{\xi}^{\star}_{\tau'_n}(k)$ *in solid line, as functions of* k. *From top to bottom and from left to right, the* t_3, t_5, t_7 *and* t_9*-distributions, respectively.*

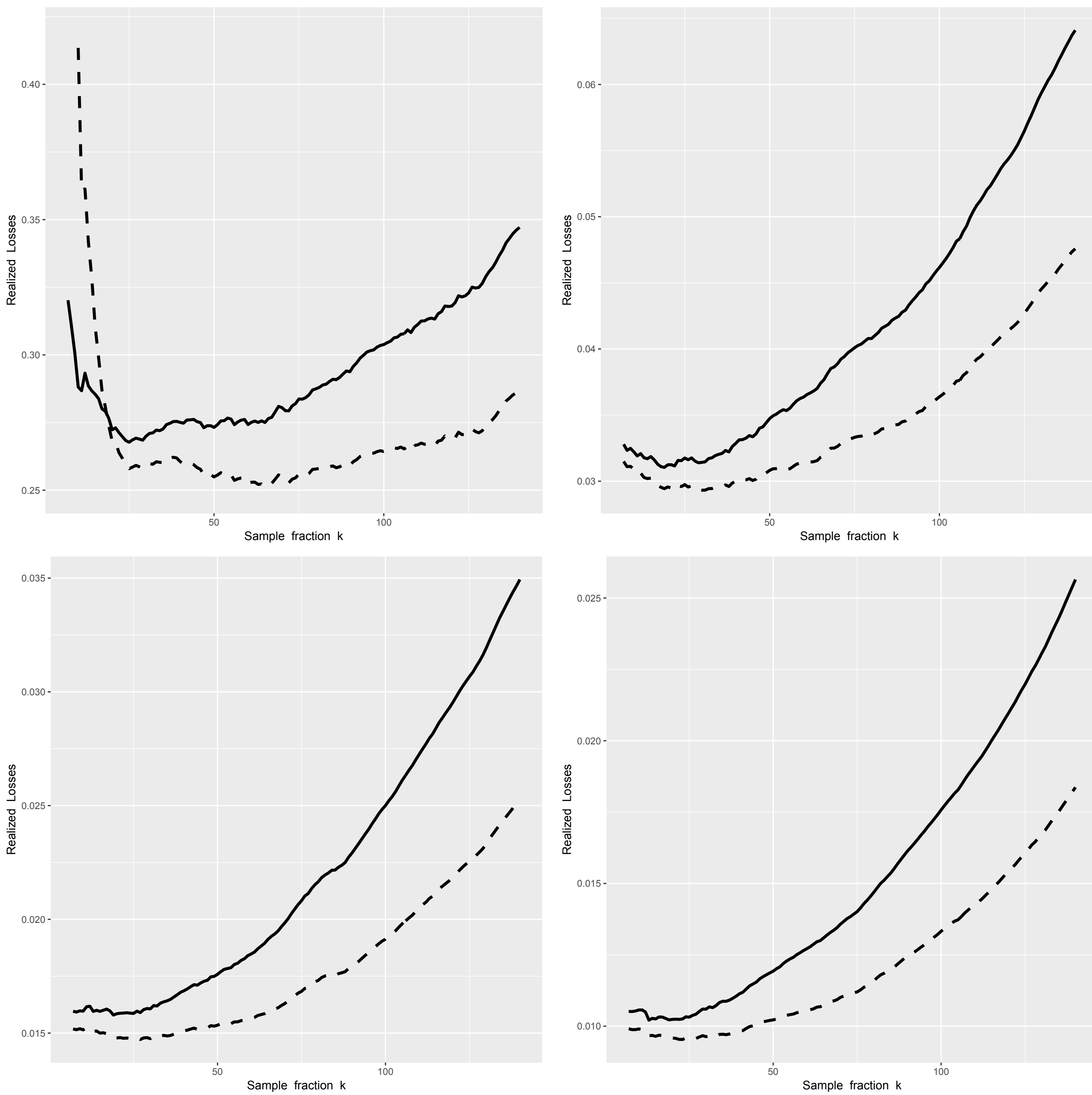

Figure 10: *The average values of the realized loss* $\mathcal{L}^{(m)}_{\tau'_n}(k)$ *for both estimators* $\widehat{\xi}^{\star}_{\tau'_n}(k)$ *in dashed line and* $\widetilde{\xi}^{\star}_{\tau'_n}(k)$ *in solid line, as functions of* k*. From top to bottom and from left to right, the positive Student* t_3*,* t_5*,* t_7 *and* t_9*-distributions, respectively.*

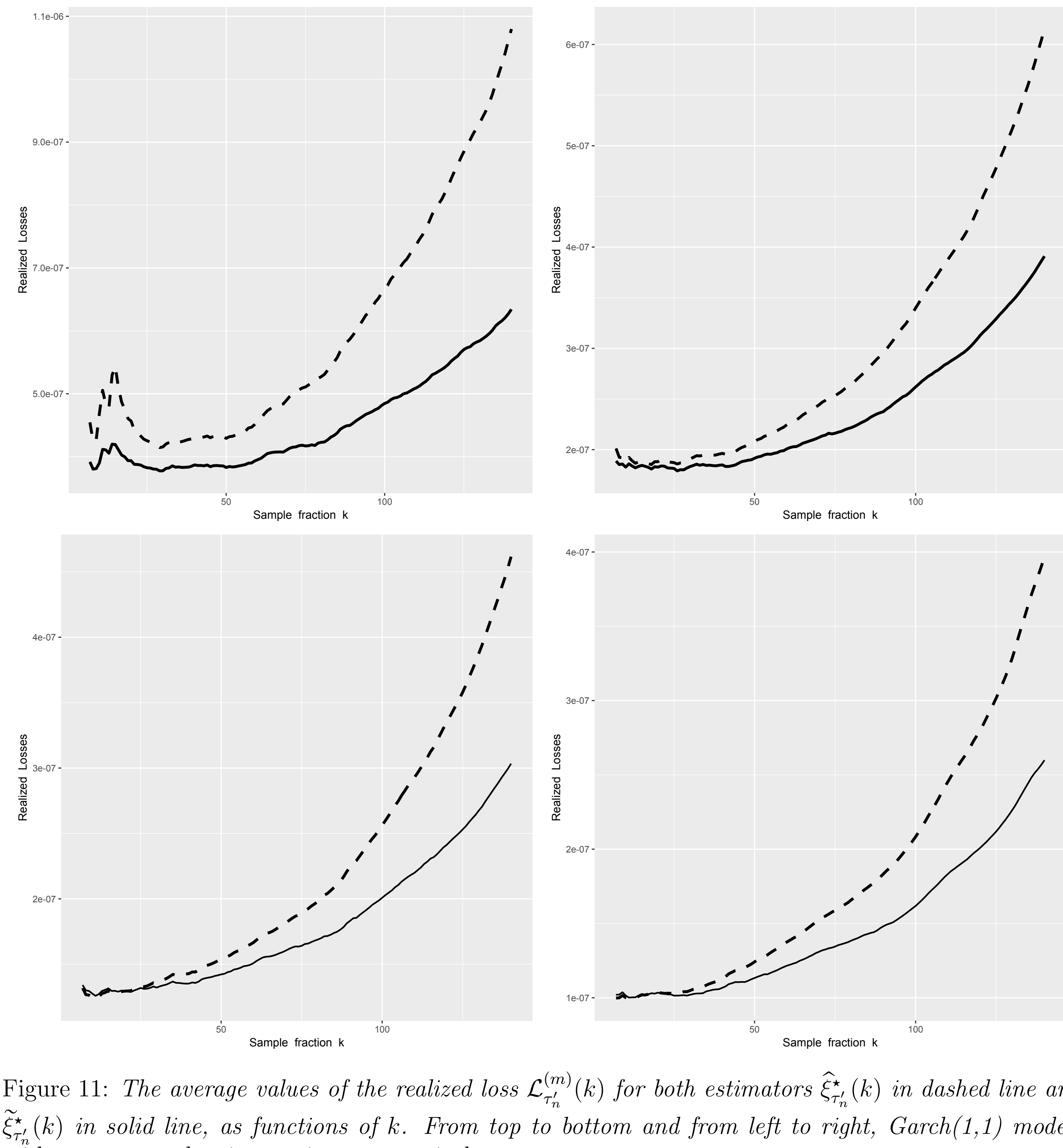

Figure 11: *The average values of the realized loss* $\mathcal{L}^{(m)}_{\tau_n'}(k)$ *for both estimators* $\widehat{\xi}^{\star}_{\tau_n'}(k)$ *in dashed line and* $\widetilde{\xi}^{\star}_{\tau_n'}(k)$ *in solid line, as functions of* k*. From top to bottom and from left to right, Garch(1,1) models with* t_3*,* t_5*,* t_7 *and* t_9 *innovations, respectively.*

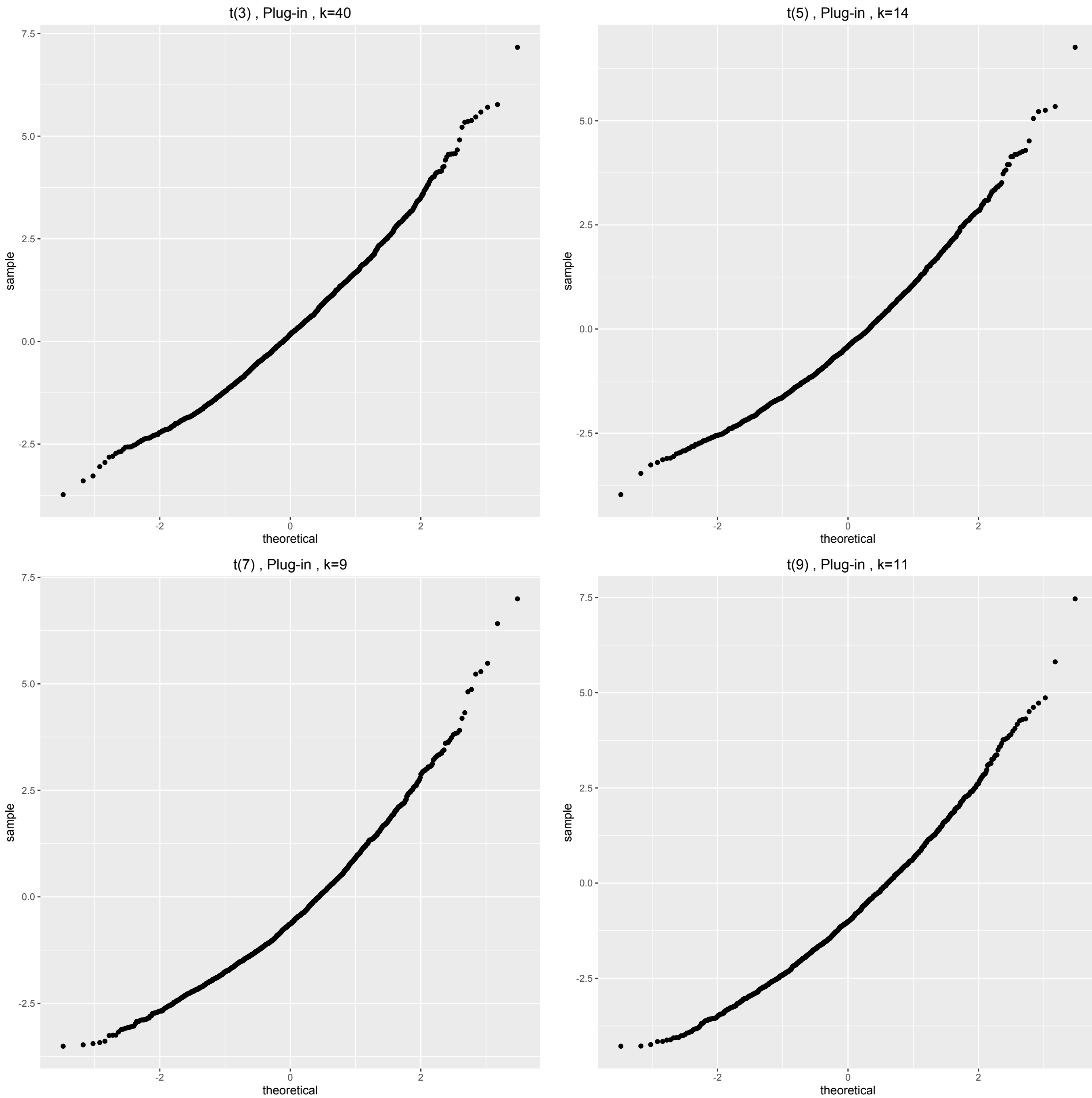

Figure 12: *Q–Q-plots on quality of asymptotic approximations. Each plot shows the sample quantiles of* $\widehat{W}_n$ *versus the theoretical standard normal quantiles, based on* $10,000$ *samples of size* $n = 1000$*. Data are simulated from the Student* t_ν *with* $\nu = 3, 5, 7, 9$*, respectively, from top to bottom and from left to right.*

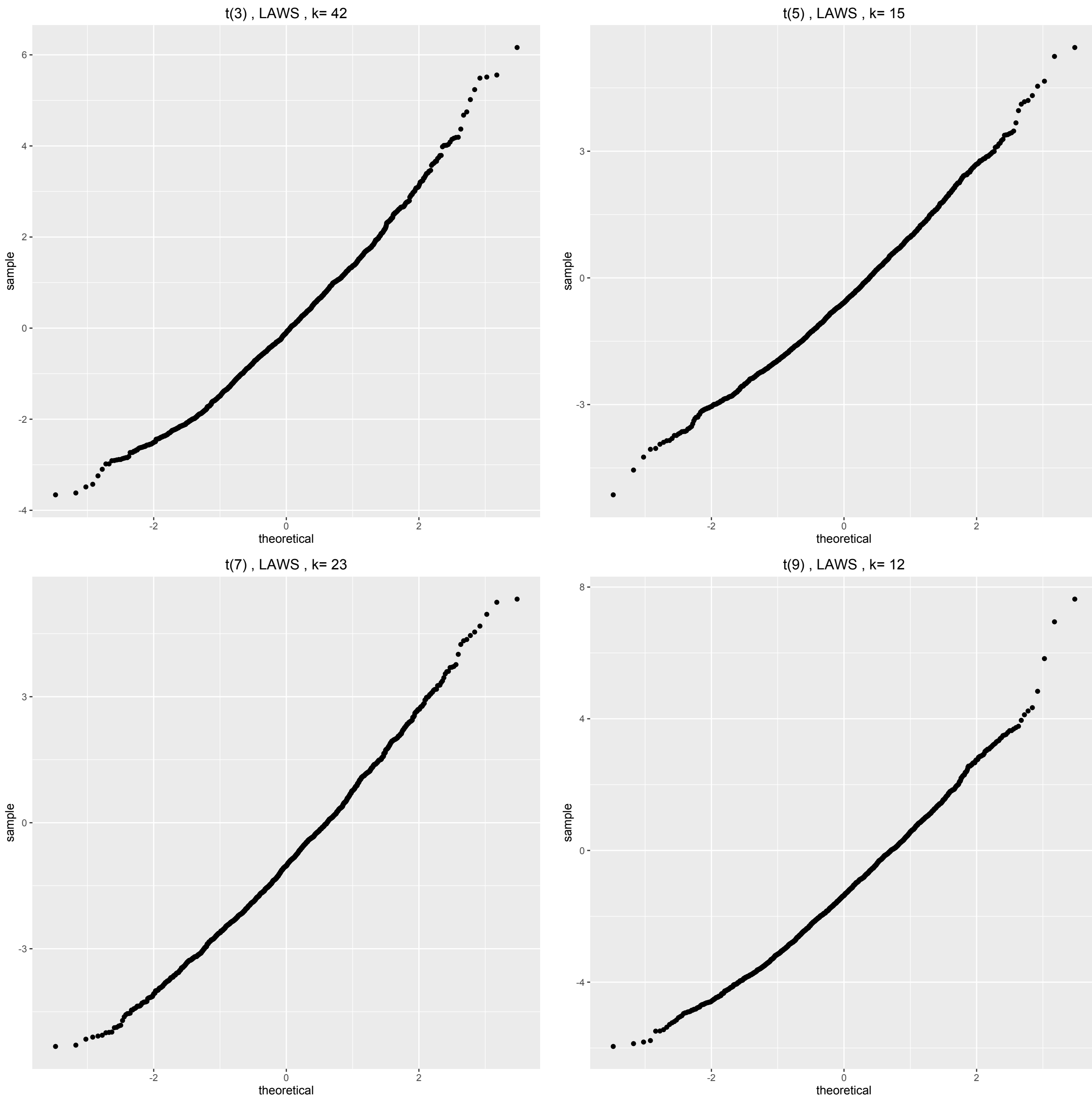

Figure 13: *As before—Scatters for* $\widetilde{W}_n$.

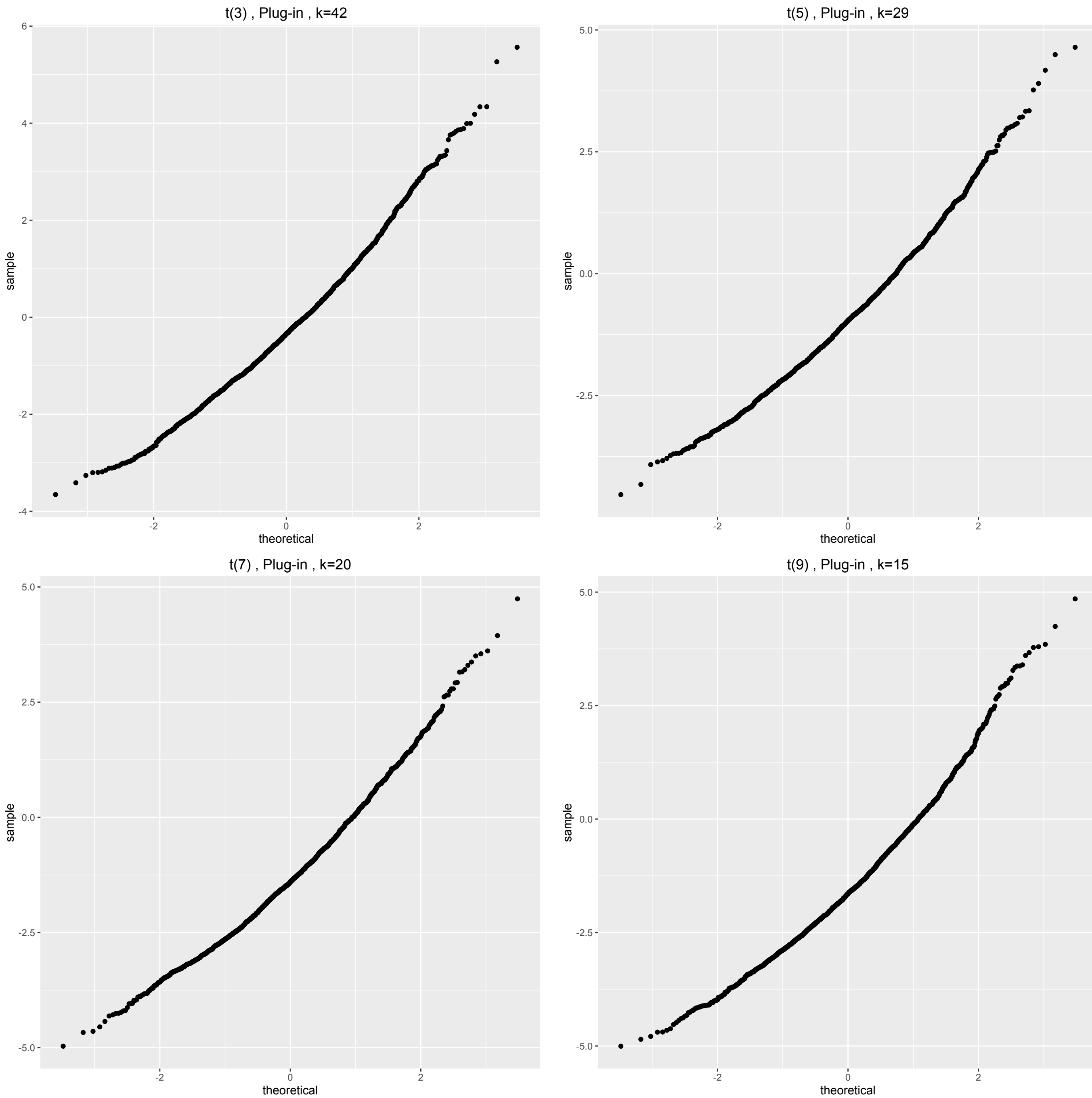

Figure 14: *Q–Q-plots on quality of asymptotic approximations. Each plot shows the sample quantiles of* $\widehat{W}_n$ *versus the theoretical standard normal quantiles, based on* $10,000$ *samples of size* $n = 1000$. *Data are simulated from the positive Student* t_ν *with* $\nu = 3, 5, 7, 9$, *respectively, from top to bottom and from left to right.*

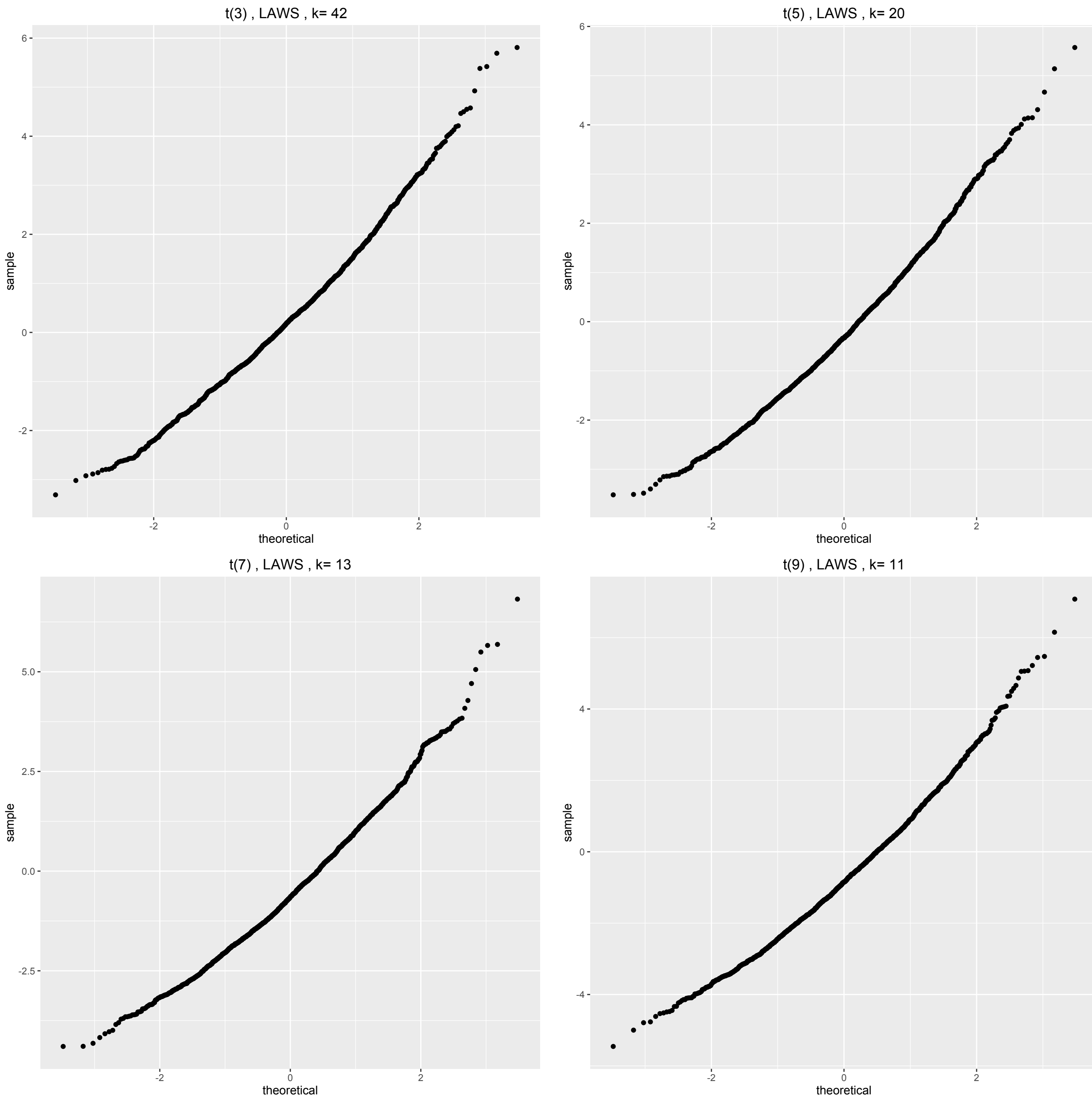

Figure 15: *As before—Scatters for* $\widetilde{W}_n$.

Next, we investigate the normality of the estimators $\widetilde{\text{XMES}}^{\star}(\tau_n')$ and $\widehat{\text{XMES}}^{\star}(\tau_n')$ by comparing the distributions of

$$\widetilde{W}_n = \left[r_n \log\left(\frac{\widetilde{\text{XMES}}^{\star}}{\text{XMES}}(\tau_n') \right) - \frac{\lambda_2}{(1-\rho_X)} \right] / \gamma_X \quad \text{and} \quad \widehat{W}_n = \left[r_n \log\left(\frac{\widehat{\text{XMES}}^{\star}}{\text{XMES}}(\tau_n') \right) - \frac{\lambda_2}{(1-\rho_X)} \right] / \gamma_X$$

with the limit distribution $\mathcal{N}(0,1)$, where $r_n = \sqrt{k}/\log[k/(n(1-\tau_n'))]$ and $\lambda_2 = \sqrt{k}A_X(n/k)$. The scatters in Figures 16 and 17 present, respectively, the sample quantiles of $\widehat{W}_n$ and $\widetilde{W}_n$, based on $10,000$ simulated samples of size $n = 1000$, versus the theoretical standard normal quantiles. For each estimator, we used the optimal value of k that minimizes its MSE. The obtained Q–Q-plots for the Student t_ν-distributions on $(0,\infty)^2$, with $\nu = 3, 5, 7, 9$, indicate that the limit Theorems 4 and 5 provide adequate approximations for finite sample sizes, with a slight advantage for the estimator $\widehat{\text{XMES}}^{\star}(\tau_n')$ in Figure 16.

B Proofs

For notational simplicity, let $\overline{F} = \overline{F}_Y$ be the survival function of Y. It is a consequence of Theorem 2.3.9 in de Haan and Ferreira (2006, p.48) that condition $\mathcal{C}_2(\gamma,\rho,A)$ entails the following second-order condition for the related survival function $\overline{F}$:

$$\forall x > 0,\ \lim_{t\to\infty} \frac{1}{A(1/\overline{F}(t))} \left[\frac{\overline{F}(tx)}{\overline{F}(t)} - x^{-1/\gamma} \right] = x^{-1/\gamma} \frac{x^{\rho/\gamma} - 1}{\gamma\rho}. \tag{B.1}$$

Proof of Proposition 1. We start by noticing that the equation

$$\xi_\tau - \mathbb{E}(Y) = \frac{2\tau - 1}{1-\tau} \mathbb{E}\left[(Y - \xi_\tau)_+\right] \tag{B.2}$$

entails, for τ sufficiently large so that $\xi_\tau > 0$,

$$1 - \frac{\mathbb{E}(Y)}{\xi_\tau} = \frac{2\tau - 1}{1-\tau} \mathbb{E}\left(\left[\frac{Y}{\xi_\tau} - 1 \right] 1\!\!1\{Y/\xi_\tau \geqslant 1\} \right). \tag{B.3}$$

An integration by parts yields

$$\begin{aligned} \mathbb{E}\left(\left[\frac{Y}{\xi_\tau} - 1 \right] 1\!\!1\{Y/\xi_\tau \geqslant 1\} \right) &= \int_1^\infty \overline{F}(\xi_\tau x)dx \\ &= \overline{F}(\xi_\tau) \left(\frac{\gamma}{1-\gamma} + \int_1^{+\infty} \left[\frac{\overline{F}(\xi_\tau x)}{\overline{F}(\xi_\tau)} - x^{-1/\gamma} \right] dx \right). \end{aligned}$$

Recall that since Y has an infinite right endpoint, $\xi_\tau \to \infty$ as $\tau \uparrow 1$; using together equation (B.1), Theorem 2.3.9 in de Haan and Ferreira (2006) and a uniform inequality such as Theorem B.3.10 in de Haan and Ferreira (2006) applied to the function $\overline{F}$, we get after some easy computations

$$\mathbb{E}\left(\left[\frac{Y}{\xi_\tau} - 1 \right] 1\!\!1\{Y/\xi_\tau \geqslant 1\} \right) = \overline{F}(\xi_\tau) \left(\frac{\gamma}{1-\gamma} + A\left(\frac{1}{\overline{F}(\xi_\tau)} \right) \frac{1 + \text{o}(1)}{(1-\gamma)(1-\rho-\gamma)} \right). \tag{B.4}$$

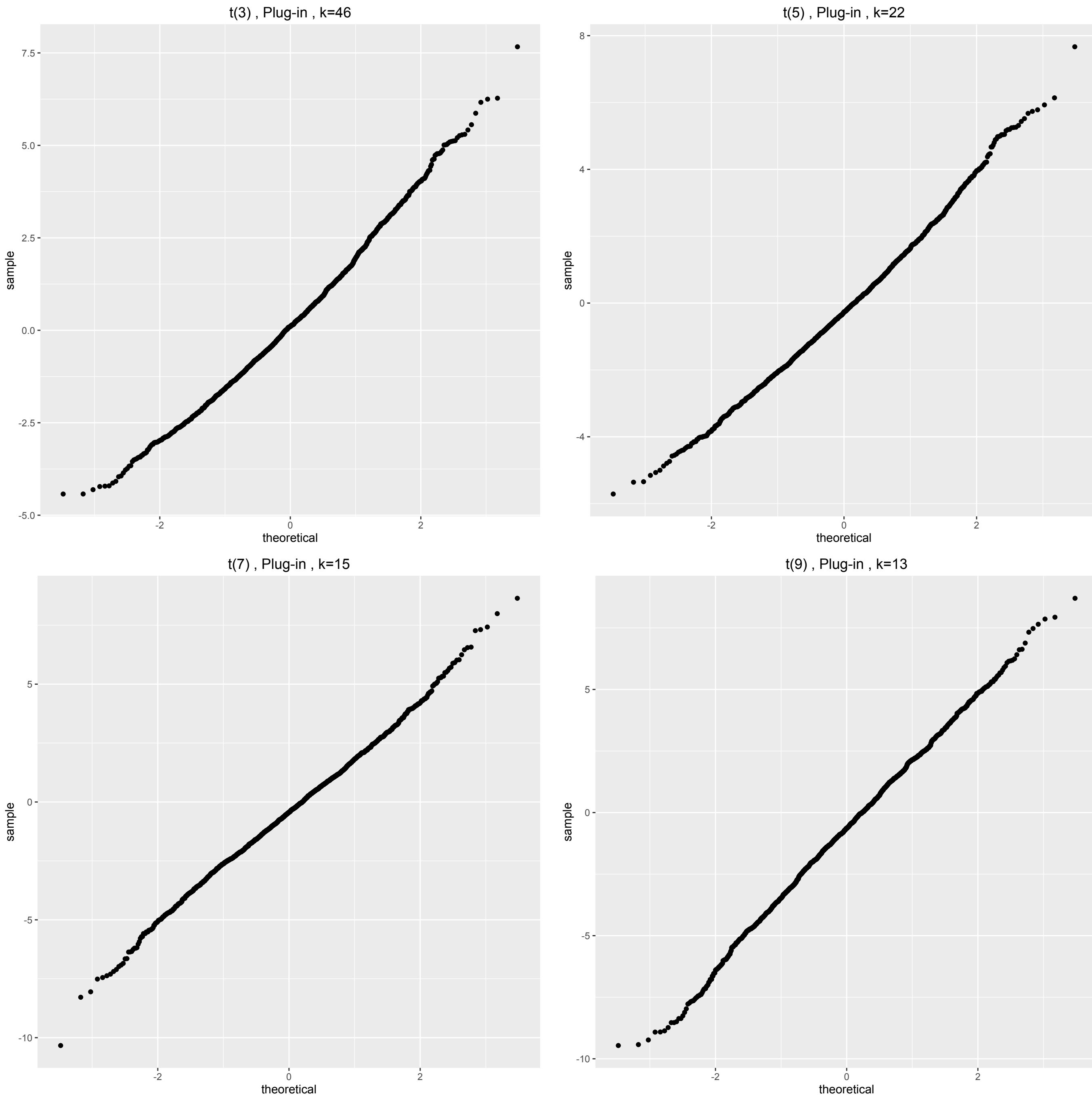

Figure 16: *Q–Q-plots of the sample quantiles of* $\widehat{W}_n$ *versus the theoretical standard normal quantiles, based on* $10,000$ *samples of size* $n = 1000$. *Data are simulated from the Student* t_ν*-distribution on* $(0,\infty)^2$ *with* $\nu = 3, 5, 7, 9$*, respectively, from top to bottom and from left to right.*

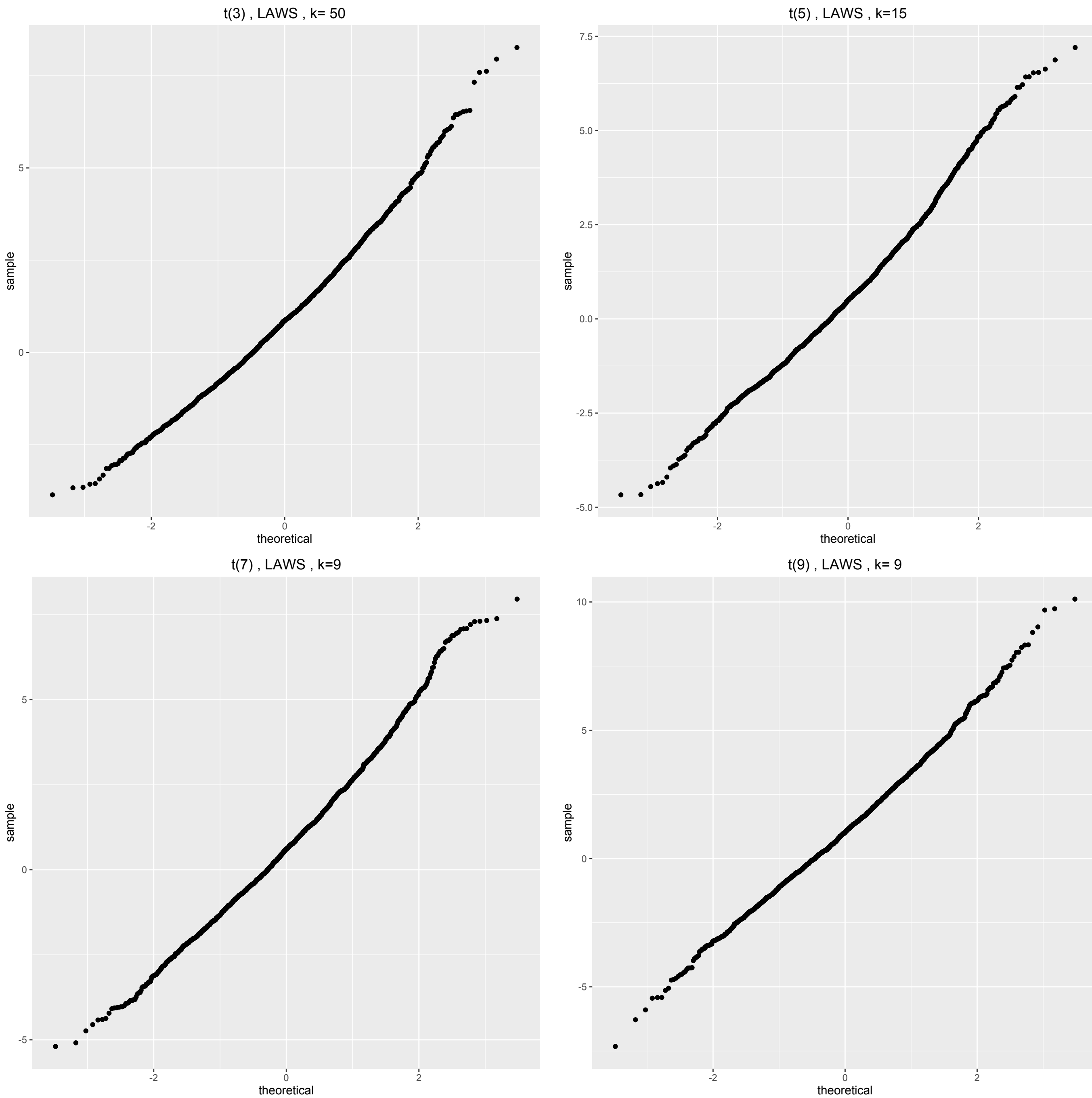

Figure 17: *As before—Scatters for* $\widetilde{W}_n$.

Plugging this equality into (B.3), we thus get

$$\frac{\overline{F}(\xi_\tau)}{1-\tau} = (\gamma^{-1}-1)\left(1-\frac{\mathbb{E}(Y)}{\xi_\tau}\right)\frac{1}{2\tau-1}\left(1+A\left(\frac{1}{\overline{F}(\xi_\tau)}\right)\frac{1}{\gamma(1-\rho-\gamma)}(1+\mathrm{o}(1))\right)^{-1}$$

and therefore

$$\begin{aligned}\frac{\overline{F}(\xi_\tau)}{1-\tau} &= (\gamma^{-1}-1)\left(1-\frac{\mathbb{E}(Y)}{\xi_\tau}(1+\mathrm{o}(1))+2(1-\tau)(1+\mathrm{o}(1))\right.\\ &\qquad \left. -A\left(\frac{1}{\overline{F}(\xi_\tau)}\right)\frac{1}{\gamma(1-\rho-\gamma)}(1+\mathrm{o}(1))\right).\end{aligned}$$

In particular, as noted in Bellini *et al.* (2014):

$$\frac{\overline{F}(\xi_\tau)}{1-\tau} \to (\gamma^{-1}-1) \ \text{ and thus } \ \xi_\tau = \left(\gamma^{-1}-1\right)^{-\gamma} q_\tau(1+\mathrm{o}(1)) \tag{B.5}$$

as $\tau \uparrow 1$. Because $\gamma < 1$, a consequence of this is that $(1-\tau)\xi_\tau = \mathrm{O}((1-\tau)q_\tau) \to 0$ as $\tau \uparrow 1$ and so

$$\begin{aligned}\frac{\overline{F}(\xi_\tau)}{1-\tau} &= (\gamma^{-1}-1)\left(1-\frac{(\gamma^{-1}-1)^{\gamma}\mathbb{E}(Y)}{q_\tau}(1+\mathrm{o}(1))\right.\\ &\quad \left. - \frac{(\gamma^{-1}-1)^{-\rho}}{\gamma(1-\rho-\gamma)}A((1-\tau)^{-1})(1+\mathrm{o}(1))\right)\end{aligned}$$

where the regular variation property of $|A|$ was used. This completes the proof. ■

The key element in the proof of Corollary 1 is to apply Proposition 1 in conjunction with the following generic result.

Lemma 1. *Assume that v, V are such that $v(\tau) \uparrow \infty$ and $V(\tau) \downarrow 0$, as $\tau \uparrow 1$, and there exists $B > 0$ such that*

$$\frac{V(\tau)}{\overline{F}(v(\tau))} = B(1+e(\tau))$$

where $e(\tau) \to 0$ as $\tau \uparrow 1$. If condition $\mathcal{C}_2(\gamma,\rho,A)$ holds, with $\gamma > 0$ and F strictly increasing, then

$$\frac{v(\tau)}{U(1/V(\tau))} = B^{\gamma}\left(1+\gamma e(\tau)(1+\mathrm{o}(1))+A(1/V(\tau))\left[\frac{B^{\rho}-1}{\rho}+\mathrm{o}(1)\right]\right) \quad \textit{as} \quad \tau \uparrow 1.$$

Proof of Lemma 1. Apply the function U to get

$$\frac{v(\tau)}{U(1/V(\tau))} - B^{\gamma} = \frac{U(B[1+e(\tau)]/V(\tau))}{U(1/V(\tau))} - B^{\gamma}.$$

By Theorem 2.3.9 in de Haan and Ferreira (2006), we may find a function A_0, equivalent to A at infinity, such that for any $\varepsilon > 0$, there is $t_0(\varepsilon) > 1$ such that for t, $tx \geqslant t_0(\varepsilon)$,

$$\left|\frac{1}{A_0(t)}\left(\frac{U(tx)}{U(t)}-x^{\gamma}\right)-x^{\gamma}\frac{x^{\rho}-1}{\rho}\right| \leqslant \frac{\varepsilon}{[(2B)^{\gamma+\rho}+(B/2)^{\gamma+\rho}][(2B)^{\varepsilon}+(B/2)^{-\varepsilon}]}x^{\gamma+\rho}\max(x^{\varepsilon},x^{-\varepsilon}).$$

Thus, for τ sufficiently close to 1, using this inequality with $t = 1/V(\tau)$ and $x = B[1+e(\tau)]$ gives that

$$\left| \frac{1}{A_0(1/V(\tau))} \left(\frac{U(B[1+e(\tau)]/V(\tau))}{U(1/V(\tau))} - B^\gamma (1+e(\tau))^\gamma \right) \right.$$
$$\left. - \quad B^\gamma (1+e(\tau))^\gamma \frac{B^\rho (1+e(\tau))^\rho - 1}{\rho} \right| \leqslant \varepsilon$$

and therefore

$$\frac{1}{A_0(1/V(\tau))} \left(\frac{U(B[1+e(\tau)]/V(\tau))}{U(1/V(\tau))} - B^\gamma (1+e(\tau))^\gamma \right) \to B^\gamma \frac{B^\rho - 1}{\rho} \quad \text{as} \quad \tau \uparrow 1.$$

The desired result follows by a simple first-order Taylor expansion. ■

Proof of Corollary 1. We have in view of Proposition 1 that

$$\frac{1-\tau}{\overline{F}(\xi_\tau)} = (\gamma^{-1} - 1)^{-1}(1 + e(\tau))$$

with

$$e(\tau) = \frac{(\gamma^{-1}-1)^\gamma \mathbb{E}(Y)}{q_\tau}(1 + \mathrm{o}(1)) + \frac{(\gamma^{-1}-1)^{-\rho}}{\gamma(1-\rho-\gamma)} A((1-\tau)^{-1})(1+\mathrm{o}(1)) \quad \text{as} \quad \tau \uparrow 1.$$

Using Lemma 1 and recalling that $U(1/(1-\tau)) = q_\tau$ gives the result. ■

Proof of Theorem 1. The consistency statement is an immediate consequence of the convergence

$$\frac{Y_{n-\lfloor n(1-\tau_n)\rfloor,n}}{q_{\tau_n}} = \frac{Y_{n-\lfloor n(1-\tau_n)\rfloor,n}}{U((1-\tau_n)^{-1})} = \frac{Y_{n-\lfloor n(1-\tau_n)\rfloor,n}}{U(n/\lfloor n(1-\tau_n)\rfloor)}(1+\mathrm{o}(1)) \xrightarrow{\mathbb{P}} 1$$

which follows from the regular variation of U and Corollary 2.2.2 in de Haan and Ferreira (2006, p.41). The asymptotic distribution is obtained by writing

$$\frac{\widehat{\xi}_{\tau_n}}{\xi_{\tau_n}} - 1 = \left(\frac{(\widehat{\gamma}^{-1}-1)^{-\widehat{\gamma}}}{(\gamma^{-1}-1)^{-\gamma}} - 1 \right) + \left(\frac{\widehat{q}_{\tau_n}}{q_{\tau_n}} - 1 \right)(1 + \mathrm{o}_{\mathbb{P}}(1)) - r(\tau_n)(1+\mathrm{o}_{\mathbb{P}}(1)),$$

where $\sqrt{n(1-\tau_n)}\, r(\tau_n) \to \lambda$ in view of Corollary 1. Since

$$\forall x \in (0,1), \; \frac{d}{dx}\left((x^{-1}-1)^{-x} \right) = (x^{-1}-1)^{-x} \left\{ (1-x)^{-1} - \log(x^{-1}-1) \right\},$$

the delta-method entails

$$\sqrt{n(1-\tau_n)} \left(\frac{(\widehat{\gamma}^{-1}-1)^{-\widehat{\gamma}}}{(\gamma^{-1}-1)^{-\gamma}} - 1 \right) \xrightarrow{d} [(1-\gamma)^{-1} - \log(\gamma^{-1}-1)]\Gamma = m(\gamma)\Gamma, \tag{B.6}$$

from which the result easily follows. ■

Before moving to the proof of Theorem 2, we shall show a couple of useful preliminary results. The next two lemmas are entirely based on non-probabilistic arguments. In the first one, we use the fact that $\eta_\tau(y)/2$ is continuously differentiable with derivative

$$\varphi_\tau(y) := |\tau - \mathbb{1}\{y \leqslant 0\}|y.$$

Lemma 2. *For all $x, y \in \mathbb{R}$ and $\tau \in (0,1)$,*

$$\frac{1}{2}(\eta_\tau(x-y) - \eta_\tau(x)) = -y\varphi_\tau(x) - \int_0^y (\varphi_\tau(x-t) - \varphi_\tau(x))dt.$$

Proof of Lemma 2. The result is a simple consequence of the equality

$$\frac{1}{2}(\eta_\tau(x-y) - \eta_\tau(x)) = \int_x^{x-y} \varphi_\tau(s)ds = -\int_0^y \varphi_\tau(x-t)dt$$

obtained by the change of variables $s = x - t$. ∎

The next result gives a Lipschitz property for the derivative φ_τ.

Lemma 3. *For all $x, h \in \mathbb{R}$ and $\tau \in (0,1)$, we have*

$$\varphi_\tau(x-h) - \varphi_\tau(x) = -h|\tau - \mathbb{1}\{x \leqslant 0\}| + (1-2\tau)(x-h)(\mathbb{1}\{x \leqslant h\} - \mathbb{1}\{x \leqslant 0\}),$$

and in particular $|\varphi_\tau(x-h) - \varphi_\tau(x)| \leqslant |h|(1 - \tau + 2\mathbb{1}\{x > \min(h,0)\})$.

Proof of Lemma 3. Write

$$\varphi_\tau(x-h) - \varphi_\tau(x) = -h|\tau - \mathbb{1}\{x \leqslant 0\}| + (x-h)(|\tau - \mathbb{1}\{x \leqslant h\}| - |\tau - \mathbb{1}\{x \leqslant 0\}|).$$

Besides,

$$\begin{aligned} & |\tau - \mathbb{1}\{x \leqslant h\}| - |\tau - \mathbb{1}\{x \leqslant 0\}| \\ = \ & (1-\tau)(\mathbb{1}\{x \leqslant h\} - \mathbb{1}\{x \leqslant 0\}) + \tau(\mathbb{1}\{x > h\} - \mathbb{1}\{x > 0\}) \\ = \ & (1-2\tau)(\mathbb{1}\{x \leqslant h\} - \mathbb{1}\{x \leqslant 0\}), \end{aligned}$$

from which the desired equality follows. The required bound on $|\varphi_\tau(x-h) - \varphi_\tau(x)|$ is then obtained by noting that

$$|\tau - \mathbb{1}\{x \leqslant 0\}| = \tau\mathbb{1}\{x > 0\} + (1-\tau)\mathbb{1}\{x \leqslant 0\} \leqslant 1 - \tau + \mathbb{1}\{x > 0\} \tag{B.7}$$

and

$$|x-h||\mathbb{1}\{x \leqslant h\} - \mathbb{1}\{x \leqslant 0\}| \leqslant |h||\mathbb{1}\{x \leqslant h\} - \mathbb{1}\{x \leqslant 0\}| \leqslant |h|\mathbb{1}\{x > \min(h,0)\}. \tag{B.8}$$

Combining (B.7) and (B.8) completes the proof. ∎

The last result will be useful to derive the limit distribution of the objective function $\psi_n(u)$ described in (8).

Lemma 4. *Pick $a > 1$ and assume that $\mathbb{E}(Y_a^-)$ and $0 < \gamma < 1/a$. Then*

$$\mathbb{E}(|\varphi_\tau(Y - \xi_\tau)|^a) = a\xi_\tau^a(1-\tau)(\gamma^{-1}-1)B(a, \gamma^{-1}-a)(1+\mathrm{o}(1)) \quad \text{as} \quad \tau \uparrow 1,$$

where $B(s,t) = \int_0^1 u^{s-1}(1-u)^{t-1}du$ is the Beta function evaluated at (s,t).

Proof of Lemma 4. As a first step, write

$$\mathbb{E}(|\varphi_\tau(Y - \xi_\tau)|^a) = (1-\tau)^a\mathbb{E}([\xi_\tau - Y]^a 1\!\!1\{Y \leqslant \xi_\tau\}) + \tau^a\mathbb{E}([Y - \xi_\tau]^a 1\!\!1\{Y > \xi_\tau\}). \tag{B.9}$$

Furthermore, for any x, y such that $x < y$, $(y-x)^a \leqslant 2^{a-1}(|x|^a + |y|^a)$ by Hölder's inequality, so that

$$\mathbb{E}([\xi_\tau - Y]^a 1\!\!1\{Y \leqslant \xi_\tau\}) \leqslant 2^{a-1}\mathbb{E}([|\xi_\tau|^a + |Y|^a]1\!\!1\{Y \leqslant \xi_\tau\}).$$

The condition $\gamma < 1/a$ ensures that $\mathbb{E}|Y|^a < \infty$. Recall that $\xi_\tau \uparrow \infty$ as $\tau \uparrow 1$ and use the dominated convergence theorem to get

$$\mathbb{E}([\xi_\tau - Y]^a 1\!\!1\{Y \leqslant \xi_\tau\}) = \mathrm{O}(\xi_\tau^a) \quad \text{as} \quad \tau \uparrow 1. \tag{B.10}$$

Besides, an integration by parts and a change of variables entail

$$\begin{aligned}
\mathbb{E}([Y - \xi_\tau]^a 1\!\!1\{Y > \xi_\tau\}) &= a\xi_\tau^{a-1}\int_{\xi_\tau}^\infty \left(\frac{x}{\xi_\tau} - 1\right)^{a-1}\overline{F}(x)dx \\
&= a\xi_\tau^a\overline{F}(\xi_\tau)\int_1^\infty (v-1)^{a-1}\frac{\overline{F}(\xi_\tau v)}{\overline{F}(\xi_\tau)}dv.
\end{aligned}$$

Using a uniform convergence theorem such as Proposition B.1.10 in de Haan and Ferreira (2006, p.360) gives

$$\mathbb{E}([Y - \xi_\tau]^a 1\!\!1\{Y > \xi_\tau\}) = a\xi_\tau^a\overline{F}(\xi_\tau)\int_1^\infty (v-1)^{a-1}v^{-1/\gamma}dv(1+\mathrm{o}(1)) \quad \text{as} \quad \tau \uparrow 1.$$

Combining this equality with (B.5) yields

$$\mathbb{E}([Y - \xi_\tau]^a 1\!\!1\{Y > \xi_\tau\}) = a\xi_\tau^a(1-\tau)(\gamma^{-1}-1)\int_1^\infty (v-1)^{a-1}v^{-1/\gamma}dv(1+\mathrm{o}(1)) \quad \text{as} \quad \tau \uparrow 1. \tag{B.11}$$

Combining (B.9), (B.10), (B.11) and using the change of variables $u = 1 - v^{-1}$ gives the desired result. ■

Proof of Theorem 2. Use Lemma 2 to write, for any u,

$$\begin{aligned}
\psi_n(u) &= -uT_{1,n} + T_{2,n}(u) \qquad \text{(B.12)} \\
\text{with} \quad T_{1,n} &:= \frac{1}{\sqrt{n(1-\tau_n)}}\sum_{i=1}^n \frac{1}{\xi_{\tau_n}}\varphi_{\tau_n}(Y_i - \xi_{\tau_n}) =: \sum_{i=1}^n S_{n,i} \\
\text{and} \quad T_{2,n}(u) &:= -\frac{1}{\xi_{\tau_n}^2}\sum_{i=1}^n \int_0^{u\xi_{\tau_n}/\sqrt{n(1-\tau_n)}} (\varphi_{\tau_n}(Y_i - \xi_{\tau_n} - t) - \varphi_{\tau_n}(Y_i - \xi_{\tau_n}))dt.
\end{aligned}$$

The random variables $S_{n,i}$ are independent, identically distributed, and centered since

$$\xi_{\tau_n} = \mathrm{argmin}_{u\in\mathbb{R}}\mathbb{E}(\eta_{\tau_n}(Y_i - u) - \eta_{\tau_n}(Y_i)) \Rightarrow \mathbb{E}(\varphi_{\tau_n}(Y_i - \xi_{\tau_n})) = 0$$

(where a differentiation under the expectation sign was used). We shall prove that

$$\frac{T_{1,n}}{\sqrt{\mathrm{Var}(T_{1,n})}} \xrightarrow{d} \mathcal{N}(0,1) \tag{B.13}$$

for which it is sufficient to show that for some $\delta > 0$,

$$\frac{n\mathbb{E}|S_{n,1}|^{2+\delta}}{[n\,\mathrm{Var}(S_{n,1})]^{1+\delta/2}} \to 0 \quad \text{as} \quad n \to \infty$$

and use Lyapunov's criterion. Choose $\delta > 0$ small enough so that $\gamma < 1/(2+\delta)$ and apply Lemma 4 to get

$$\frac{n\mathbb{E}|S_{n,1}|^{2+\delta}}{[n\,\mathrm{Var}(S_{n,1})]^{1+\delta/2}} = \mathrm{O}([n(1-\tau_n)]^{-\delta/2}) \to 0 \quad \text{as} \quad n \to \infty.$$

Convergence (B.13) follows and, especially, Lemma 4 entails

$$T_{1,n} \xrightarrow{d} \mathcal{N}\left(0, \frac{2\gamma}{1-2\gamma}\right). \tag{B.14}$$

We now turn to the control of the second term $T_{2,n}(u)$. Write

$$T_{2,n}(u) = T_{3,n}(u) - \frac{n}{\xi_{\tau_n}^2}\int_0^{u\xi_{\tau_n}/\sqrt{n(1-\tau_n)}} [\mathbb{E}(\varphi_{\tau_n}(Y - \xi_{\tau_n} - t)) - \mathbb{E}(\varphi_{\tau_n}(Y - \xi_{\tau_n}))]dt. \tag{B.15}$$

The random term $T_{3,n}(u)$ is a sum of independent, identically distributed and centered random variables, which we shall examine after having controlled first the nonrandom term on the right-hand side of (B.15). By Lemma 3, we obtain

$$\begin{aligned} & \mathbb{E}(\varphi_{\tau_n}(Y - \xi_{\tau_n} - t)) - \mathbb{E}(\varphi_{\tau_n}(Y - \xi_{\tau_n})) \\ = \; & (1-2\tau_n)\mathbb{E}((Y - \xi_{\tau_n} - t)(1\!\!1\{Y \leqslant \xi_{\tau_n} + t\} - 1\!\!1\{Y \leqslant \xi_{\tau_n}\})) \\ - \; & t\mathbb{E}(|\tau_n - 1\!\!1\{Y \leqslant \xi_{\tau_n}\}|). \end{aligned} \tag{B.16}$$

Clearly

$$\mathbb{E}(|\tau_n - 1\!\!1\{Y \leqslant \xi_{\tau_n}\}|) = \tau_n\overline{F}(\xi_{\tau_n}) + (1-\tau_n)F(\xi_{\tau_n}).$$

It therefore follows from (B.5) that

$$\mathbb{E}(|\tau_n - 1\!\!1\{Y \leqslant \xi_{\tau_n}\}|) = \gamma^{-1}(1-\tau_n)(1+\mathrm{o}(1)) \tag{B.17}$$

as $n \to \infty$. Let further $\psi(t) := \mathbb{E}((Y-t)1\!\!1\{Y > t\})$ and observe that

$$\begin{aligned} & \mathbb{E}((Y - \xi_{\tau_n} - t)(1\!\!1\{Y \leqslant \xi_{\tau_n} + t\} - 1\!\!1\{Y \leqslant \xi_{\tau_n}\})) \\ = \; & \mathbb{E}((Y - \xi_{\tau_n} - t)(1\!\!1\{Y > \xi_{\tau_n}\} - 1\!\!1\{Y > \xi_{\tau_n} + t\})) \\ = \; & \psi(\xi_{\tau_n}) - \psi(\xi_{\tau_n} + t) - t\overline{F}(\xi_{\tau_n}). \end{aligned}$$

Integrating by parts entails

$$\psi(\xi_{\tau_n}) - \psi(\xi_{\tau_n} + t) = \int_{\xi_{\tau_n}}^{\xi_{\tau_n}+t} \overline{F}(x)dx = \xi_{\tau_n}\overline{F}(\xi_{\tau_n}) \int_1^{1+t/\xi_{\tau_n}} \frac{\overline{F}(\xi_{\tau_n} v)}{\overline{F}(\xi_{\tau_n})} dv$$

from which we deduce that

$$\begin{aligned} & \mathbb{E}((Y - \xi_{\tau_n} - t)(1\!\!1\{Y \leqslant \xi_{\tau_n} + t\} - 1\!\!1\{Y \leqslant \xi_{\tau_n}\})) \\ = \; & t\overline{F}(\xi_{\tau_n}) \left(\frac{\xi_{\tau_n}}{t} \int_1^{1+t/\xi_{\tau_n}} \frac{\overline{F}(\xi_{\tau_n} v)}{\overline{F}(\xi_{\tau_n})} dv - 1 \right). \end{aligned}$$

We now bound the term into brackets as follows: let $I_n(u) = [0, |u|\xi_{\tau_n}/\sqrt{n(1-\tau_n)}]$ and write

$$\begin{aligned} & \sup_{|t| \in I_n(u)} \left| \frac{\xi_{\tau_n}}{t} \int_1^{1+t/\xi_{\tau_n}} \frac{\overline{F}(\xi_{\tau_n} v)}{\overline{F}(\xi_{\tau_n})} dv - 1 \right| \\ \leqslant \; & \sup_{|t| \in I_n(u)} \frac{\xi_{\tau_n}}{|t|} \left| \int_1^{1+t/\xi_{\tau_n}} \left[\frac{\overline{F}(\xi_{\tau_n} v)}{\overline{F}(\xi_{\tau_n})} - v^{-1/\gamma} \right] dv \right| + \mathrm{o}(1) \\ = \; & \mathrm{o}(1) \end{aligned}$$

by the uniform convergence theorem for regularly varying functions [see Theorem 1.5.2 in Bingham *et al.* (1987), p.22], the continuity of $v \mapsto v^{-1/\gamma}$ at 1 and the convergence $n(1-\tau_n) \to \infty$. As a consequence, by (B.5), the equality

$$\mathbb{E}((Y - \xi_{\tau_n} - t)(1\!\!1\{Y \leqslant \xi_{\tau_n} + t\} - 1\!\!1\{Y \leqslant \xi_{\tau_n}\})) = t(1-\tau_n) r_n(t) \tag{B.18}$$

holds with $r_n(t) \to 0$ uniformly in t such that $|t| \in I_n(u)$. Combine (B.15), (B.16), (B.17) and (B.18) to get

$$\begin{aligned} T_{2,n}(u) &= \frac{u^2}{2\gamma}(1 + \mathrm{o}(1)) + T_{3,n}(u), \qquad \text{(B.19)} \\ \text{with } \; T_{3,n}(u) &:= -\frac{1}{\xi_{\tau_n}^2} \sum_{i=1}^n \int_0^{u\xi_{\tau_n}/\sqrt{n(1-\tau_n)}} [\mathcal{S}_{n,i}(\xi_{\tau_n} + t) - \mathcal{S}_{n,i}(\xi_{\tau_n})] dt \end{aligned}$$

where the $\mathcal{S}_{n,i}(v) := \varphi_{\tau_n}(Y_i - v) - \mathbb{E}(\varphi_{\tau_n}(Y_i - v))$ are independent copies of $\mathcal{S}_n(v) := \varphi_{\tau_n}(Y - v) - \mathbb{E}(\varphi_{\tau_n}(Y - v))$. Thus

$$\mathrm{Var}(T_{3,n}(u)) = \frac{n}{\xi_{\tau_n}^4} \mathrm{Var}\left(\int_0^{u\xi_{\tau_n}/\sqrt{n(1-\tau_n)}} [\mathcal{S}_n(\xi_{\tau_n} + t) - \mathcal{S}_n(\xi_{\tau_n})] dt \right).$$

We now notice that for any v, $\mathcal{S}_n(v)$ is centered and thus

$$\mathrm{Var}(T_{3,n}(u)) = \frac{n}{\xi_{\tau_n}^4} \int_{[0,\, u\xi_{\tau_n}/\sqrt{n(1-\tau_n)}]^2} \mathbb{E}([\mathcal{S}_n(\xi_{\tau_n} + s) - \mathcal{S}_n(\xi_{\tau_n})][\mathcal{S}_n(\xi_{\tau_n} + t) - \mathcal{S}_n(\xi_{\tau_n})]) ds\, dt$$

(where the integrability properties of Y were used to switch integrals and expectation). By the Cauchy-Schwarz inequality,

$$\operatorname{Var}(T_{3,n}(u)) \leqslant \frac{n}{\xi_{\tau_n}^4}\left(\int_0^{u\xi_{\tau_n}/\sqrt{n(1-\tau_n)}} \sqrt{\mathbb{E}(|\mathcal{S}_n(\xi_{\tau_n}+t)-\mathcal{S}_n(\xi_{\tau_n})|^2)}\, dt\right)^2. \tag{B.20}$$

Applying Lemma 3, we get for any t

$$|\mathcal{S}_n(\xi_{\tau_n}+t)-\mathcal{S}_n(\xi_{\tau_n})| \leqslant 2|t|[1-\tau_n+\mathbb{1}\{Y>\xi_{\tau_n}+\min(t,0)\}+\overline{F}(\xi_{\tau_n}+\min(t,0))].$$

Using the inequality $|a+b+c|^2 \leqslant 3(a^2+b^2+c^2)$ yields

$$\mathbb{E}(|\mathcal{S}_n(\xi_{\tau_n}+t)-\mathcal{S}_n(\xi_{\tau_n})|^2) \leqslant 12t^2[(1-\tau_n)^2+\overline{F}(\xi_{\tau_n}+\min(t,0))(1+\overline{F}(\xi_{\tau_n}+\min(t,0)))]. \tag{B.21}$$

Finally, using again the regular variation property of $\overline{F}$ and the convergence $n(1-\tau_n)\to\infty$,

$$\sup_{|s|\in I_n(u)} |\overline{F}(\xi_{\tau_n}+s)-\overline{F}(\xi_{\tau_n})| = \overline{F}(\xi_{\tau_n}) \sup_{|s|\in I_n(u)} \left|\frac{\overline{F}(\xi_{\tau_n}+s)}{\overline{F}(\xi_{\tau_n})}-1\right| = \mathrm{o}(\overline{F}(\xi_{\tau_n})) = \mathrm{o}(1-\tau_n) \tag{B.22}$$

in view of (B.5). Using (B.5) once again and combining (B.20), (B.21) and (B.22) yields

$$\operatorname{Var}(T_{3,n}(u)) = \mathrm{O}\left(\frac{n}{\xi_{\tau_n}^4}(1-\tau_n)\left|\int_0^{u\xi_{\tau_n}/\sqrt{n(1-\tau_n)}} |t|\, dt\right|^2\right) = \mathrm{O}\left(\frac{1}{n(1-\tau_n)}\right) \to 0$$

as $n\to\infty$. Whence the convergence $T_{3,n}(u) \xrightarrow{\mathbb{P}} 0$; combining (B.12), (B.14) and (B.19) entails

$$\psi_n(u) \xrightarrow{d} -uZ\sqrt{\frac{2\gamma}{1-2\gamma}} + \frac{u^2}{2\gamma} \quad \text{as} \quad n\to\infty$$

(with Z being standard Gaussian) in the sense of finite-dimensional convergence. As a function of u, this limit is almost surely finite and defines a convex function which has a unique minimum at

$$u^\star = \gamma\sqrt{\frac{2\gamma}{1-2\gamma}}Z \stackrel{d}{=} \mathcal{N}\left(0, \gamma^2\frac{2\gamma}{1-2\gamma}\right).$$

Applying the convexity lemma of Geyer (1996) completes the proof. ■

Proof of Theorem 3. By the equalities (9) and (10), we have

$$\log\left(\frac{\overline{\xi}^\star_{\tau'_n}}{\xi_{\tau'_n}}\right) = \log\left(\frac{\widehat{q}^\star_{\tau'_n}}{q_{\tau'_n}}\right) + \log\left(\frac{\overline{\xi}_{\tau_n}}{\xi_{\tau_n}}\right) - \log\left(\frac{\widehat{q}_{\tau_n}}{q_{\tau_n}}\right) + \log\left(\frac{\xi_{\tau_n}}{q_{\tau_n}}\right) - \log\left(\frac{\xi_{\tau'_n}}{q_{\tau'_n}}\right).$$

Furthermore, the convergence $\log[(1-\tau_n)/(1-\tau_n')] \to \infty$ entails

$$\frac{\sqrt{n(1-\tau_n)}}{\log[(1-\tau_n)/(1-\tau_n')]} \log\left(\frac{\widehat{q}^{\star}_{\tau_n'}}{q_{\tau_n'}}\right) \xrightarrow{d} \Gamma, \tag{B.23}$$

$$\begin{aligned}\frac{\sqrt{n(1-\tau_n)}}{\log[(1-\tau_n)/(1-\tau_n')]} \log\left(\frac{\overline{\xi}_{\tau_n}}{\xi_{\tau_n}}\right) &= \mathrm{O}_{\mathbb{P}}\left(1/\log[(1-\tau_n)/(1-\tau_n')]\right) \\ &= \mathrm{o}_{\mathbb{P}}(1),\end{aligned} \tag{B.24}$$

$$\begin{aligned}\frac{\sqrt{n(1-\tau_n)}}{\log[(1-\tau_n)/(1-\tau_n')]} \log\left(\frac{\widehat{q}_{\tau_n}}{q_{\tau_n}}\right) &= \mathrm{O}_{\mathbb{P}}\left(1/\log[(1-\tau_n)/(1-\tau_n')]\right) \\ &= \mathrm{o}_{\mathbb{P}}(1),\end{aligned} \tag{B.25}$$

$$\begin{aligned}\frac{\sqrt{n(1-\tau_n)}}{\log[(1-\tau_n)/(1-\tau_n')]}\left[\log\left(\frac{\xi_{\tau_n}}{q_{\tau_n}}\right) - \log\left(\frac{\xi_{\tau_n'}}{q_{\tau_n'}}\right)\right] &= \mathrm{O}\left(\frac{\sqrt{n(1-\tau_n)}[r(\tau_n)+r(\tau_n')]}{\log[(1-\tau_n)/(1-\tau_n')]}\right) \\ &= \mathrm{O}\left(\frac{\sqrt{n(1-\tau_n)}r(\tau_n)}{\log[(1-\tau_n)/(1-\tau_n')]}\right) \\ &= \mathrm{o}(1).\end{aligned} \tag{B.26}$$

Here, Theorem 4.3.8 in de Haan and Ferreira (2006, p.138) was used to show (B.23), while (B.24) and (B.25) follow from the distributional convergence assumption on $\overline{\xi}_{\tau_n}$ and from Theorem 2.4.1 in de Haan and Ferreira (2006, p.50), respectively. Convergence (B.26) is a consequence of Corollary 1 and, in what concerns the relationship $r(\tau_n') = \mathrm{O}(r(\tau_n))$, of the regular variation of $s \mapsto q_{1-s^{-1}}$ and $|A|$. A combination of these convergence results and a use of the delta-method give the desired conclusion. ■

Proof of Proposition 2. By Corollary 1,

$$\mathrm{XES}(\tau) = \frac{1}{1-\tau}\int_\tau^1 \xi_\alpha d\alpha = (\gamma^{-1}-1)^{-\gamma}\left\{\frac{1}{1-\tau}\int_\tau^1 q_\alpha(1+r(\alpha))d\alpha\right\}$$

where $r(\alpha) \to 0$ as $\alpha \to 1$. It is then clear that

$$\mathrm{XES}(\tau) \sim (\gamma^{-1}-1)^{-\gamma}\left\{\frac{1}{1-\tau}\int_\tau^1 q_\alpha d\alpha\right\} = (\gamma^{-1}-1)^{-\gamma}\mathrm{QES}(\tau) \quad \text{as} \quad \tau \to 1.$$

This proves that

$$\frac{\mathrm{XES}(\tau)}{\mathrm{QES}(\tau)} \sim (\gamma^{-1}-1)^{-\gamma} \sim \frac{\xi_\tau}{q_\tau} \quad \text{as} \quad \tau \to 1,$$

by applying Corollary 1 again. Besides, the equality $q_\alpha = U((1-\alpha)^{-1})$ and a change of variables entail

$$\frac{\mathrm{QES}(\tau)}{q_\tau} = \frac{1}{1-\tau}\int_\tau^1 \frac{q_\alpha}{q_\tau} d\alpha = \int_1^\infty y^{-1}\frac{U((1-\tau)^{-1}y)}{U((1-\tau)^{-1})}\frac{dy}{y}.$$

The condition $\gamma < 1$ and a uniform convergence theorem such as Proposition B.1.10 in de Haan and Ferreira (2006, p.360) entail

$$\frac{\mathrm{QES}(\tau)}{q_\tau} \to \int_1^\infty y^{\gamma-2}dy = \frac{1}{1-\gamma} \quad \text{as} \quad \tau \to 1.$$

Consequently

$$\frac{\mathrm{XES}(\tau)}{\xi_\tau} \sim \frac{\mathrm{QES}(\tau)}{q_\tau} \to \frac{1}{1-\gamma} \quad \text{as} \quad \tau \to 1.$$

Let us now turn to the terms $\mathrm{XTCE}(\tau)/\mathrm{QTCE}(\tau)$ and $\mathrm{XTCE}(\tau)/\xi_\tau$. On the one hand, we have

$$\mathrm{XTCE}(\tau) = \frac{\mathbb{E}\left[Y\mathbb{I}(Y > \xi_\tau)\right]}{\overline{F}(\xi_\tau)} = \frac{\mathbb{E}\left[(Y - \xi_\tau)_+\right]}{\overline{F}(\xi_\tau)} + \xi_\tau,$$

where $y_+ = \max(y, 0)$. On the other hand, it follows from the proof of Theorem 11 in Bellini *et al.* (2014) that

$$\frac{\mathbb{E}\left[(Y - \xi_\tau)_+\right]}{\overline{F}(\xi_\tau)} \sim \frac{\xi_\tau}{\gamma^{-1} - 1} \quad \text{as} \quad \tau \to 1.$$

Therefore $\dfrac{\mathrm{XTCE}(\tau)}{\xi_\tau} \sim \dfrac{1}{1-\gamma}$ as $\tau \to 1$. Likewise, we have

$$\mathrm{QTCE}(\tau) = \frac{\mathbb{E}\left[Y\mathbb{I}(Y > q_\tau)\right]}{\overline{F}(q_\tau)} = \frac{\mathbb{E}\left[(Y - q_\tau)_+\right]}{\overline{F}(q_\tau)} + q_\tau,$$

with

$$\frac{\mathbb{E}\left[(Y - q_\tau)_+\right]}{\overline{F}(q_\tau)} \sim \frac{q_\tau}{\gamma^{-1} - 1} \quad \text{as} \quad \tau \to 1.$$

Then $\dfrac{\mathrm{QTCE}(\tau)}{q_\tau} \sim \dfrac{1}{1-\gamma}$ as $\tau \to 1$. Whence $\dfrac{\mathrm{XTCE}(\tau)}{\mathrm{QTCE}(\tau)} \sim \dfrac{\xi_\tau}{q_\tau}$ as $\tau \to 1$, which completes the proof. ■

Proof of Proposition 3. The starting point is Corollary 1, which yields

$$\begin{aligned}
\mathrm{XES}(\tau) &= \frac{1}{1-\tau}\int_\tau^1 \xi_\alpha d\alpha \\
&= (\gamma^{-1} - 1)^{-\gamma}\Bigg(\mathrm{QES}(\tau) + \gamma(\gamma^{-1} - 1)^{\gamma}\mathbb{E}(Y)(1 + \mathrm{o}(1)) \\
&+ \left\{\frac{(\gamma^{-1} - 1)^{-\rho}}{1 - \rho - \gamma} + \frac{(\gamma^{-1} - 1)^{-\rho} - 1}{\rho} + \mathrm{o}(1)\right\}\frac{1}{1-\tau}\int_\tau^1 q_\alpha A((1-\alpha)^{-1})d\alpha\Bigg).
\end{aligned}$$

Use a change of variables to get

$$\frac{1}{1-\tau}\int_\tau^1 q_\alpha A((1-\alpha)^{-1})d\alpha = U((1-\tau)^{-1})A((1-\tau)^{-1})\int_1^\infty y^{-1}\frac{U((1-\tau)^{-1}y)A((1-\tau)^{-1}y)}{U((1-\tau)^{-1})A((1-\tau)^{-1})}\frac{dy}{y}.$$

This entails, using a uniform convergence theorem such as Proposition B.1.10 in de Haan and Ferreira (2006, p.360), that

$$\begin{aligned}
\frac{1}{1-\tau}\int_\tau^1 q_\alpha A((1-\alpha)^{-1})d\alpha &\sim U((1-\tau)^{-1})A((1-\tau)^{-1})\int_1^\infty y^{\gamma+\rho-2}dy \quad \text{as} \quad \tau\to 1 \\
&= \frac{q_\tau A((1-\tau)^{-1})}{1-\rho-\gamma}.
\end{aligned}$$

Since $\text{QES}(\tau)\sim q_\tau/(1-\gamma)$, our earlier expansion yields

$$\begin{aligned}
\frac{\text{XES}(\tau)}{\text{QES}(\tau)} &= (\gamma^{-1}-1)^{-\gamma}\Bigg(1+\frac{\gamma(1-\gamma)(\gamma^{-1}-1)^{\gamma}\mathbb{E}(Y)}{q_\tau}(1+\text{o}(1)) \\
&+ \left\{\frac{(\gamma^{-1}-1)^{-\rho}}{1-\rho-\gamma}+\frac{(\gamma^{-1}-1)^{-\rho}-1}{\rho}+\text{o}(1)\right\}\frac{1-\gamma}{1-\rho-\gamma}A((1-\tau)^{-1})\Bigg). \qquad \text{(B.27)}
\end{aligned}$$

Furthermore, it is a consequence of a uniform inequality such as Theorem B.3.10 in de Haan and Ferreira (2006) applied to the function U that

$$\begin{aligned}
\frac{\text{QES}(\tau)}{q_\tau} &= \int_1^\infty y^{-1}\frac{U((1-\tau)^{-1}y)}{U((1-\tau)^{-1})}\frac{dy}{y} \\
&= \int_1^\infty y^{-1}\left(y^\gamma + A((1-\tau)^{-1})y^\gamma\frac{y^\rho-1}{\rho}(1+\text{o}(1))\right)\frac{dy}{y} \\
&= \int_1^\infty y^{\gamma-2}dy + \frac{A((1-\tau)^{-1})}{\rho}\int_1^\infty \left(y^{\gamma+\rho-2}-y^{\gamma-2}\right)dy(1+\text{o}(1)) \\
&= \frac{1}{1-\gamma}\left(1+\frac{1}{1-\rho-\gamma}A((1-\tau)^{-1})(1+\text{o}(1))\right). \qquad \text{(B.28)}
\end{aligned}$$

Finally, Corollary 1 reads

$$\begin{aligned}
\frac{q_\tau}{\xi_\tau} &= (\gamma^{-1}-1)^{\gamma}\Bigg(1-\frac{\gamma(\gamma^{-1}-1)^{\gamma}\mathbb{E}(Y)}{q_\tau}(1+\text{o}(1)) \\
&- \left(\frac{(\gamma^{-1}-1)^{-\rho}}{1-\rho-\gamma}+\frac{(\gamma^{-1}-1)^{-\rho}-1}{\rho}+\text{o}(1)\right)A((1-\tau)^{-1})\Bigg). \qquad \text{(B.29)}
\end{aligned}$$

A use of the identity

$$\frac{\text{XES}(\tau)}{\xi_\tau} = \frac{\text{XES}(\tau)}{\text{QES}(\tau)}\times\frac{\text{QES}(\tau)}{q_\tau}\times\frac{q_\tau}{\xi_\tau}$$

and a combination of (B.27), (B.28) and (B.29) complete the proof of the first part after some straightforward computations.

Let us now turn to the second part of the Proposition. The starting point is equation (B.2), which is equivalent to

$$\frac{\text{XTCE}(\tau)}{\xi_\tau} = 1+\frac{1-\tau}{\overline{F}(\xi_\tau)}\frac{1}{2\tau-1}\left(1-\frac{\mathbb{E}(Y)}{\xi_\tau}\right).$$

We have by Proposition 1 and (B.5), with the notation therein, that

$$\frac{1-\tau}{\overline{F}(\xi_\tau)} = \frac{\gamma}{1-\gamma}[1-\varepsilon(\tau)(1+\mathrm{o}(1))] \ \text{ and } \ \frac{1}{\xi_\tau} = \frac{(\gamma^{-1}-1)^\gamma}{q_\tau}(1+\mathrm{o}(1)),$$

where the o(·) terms have to be understood in the asymptotic sense as $\tau \uparrow 1$. Using a Taylor expansion thus yields:

$$\begin{aligned}\frac{\mathrm{XTCE}(\tau)}{\xi_\tau} &= \frac{1}{1-\gamma} + \frac{\gamma}{1-\gamma}\Bigg[2(1-\tau)(1+\mathrm{o}(1)) - \varepsilon(\tau)(1+\mathrm{o}(1)) \\ &- \frac{(\gamma^{-1}-1)^\gamma \mathbb{E}(Y)}{q_\tau}(1+\mathrm{o}(1))\Bigg].\end{aligned}$$

The condition $\gamma < 1$ entails $(1-\tau)q_\tau \to 0$ as $\tau \uparrow 1$, so that

$$\frac{\mathrm{XTCE}(\tau)}{\xi_\tau} = \frac{1}{1-\gamma} - \frac{\gamma}{1-\gamma}\left[\varepsilon(\tau)(1+\mathrm{o}(1)) + \frac{(\gamma^{-1}-1)^\gamma \mathbb{E}(Y)}{q_\tau}(1+\mathrm{o}(1))\right].$$

Using once again Proposition 1 gives

$$\varepsilon(\tau) + \frac{(\gamma^{-1}-1)^\gamma \mathbb{E}(Y)}{q_\tau} = -\frac{(\gamma^{-1}-1)^{-\rho}}{\gamma(1-\rho-\gamma)}A((1-\tau)^{-1})(1+\mathrm{o}(1)) + \mathrm{o}(q_\tau^{-1}),$$

whence

$$\frac{\mathrm{XTCE}(\tau)}{\xi_\tau} = \frac{1}{1-\gamma}\left[1 + \frac{(\gamma^{-1}-1)^{-\rho}}{1-\rho-\gamma}A((1-\tau)^{-1})(1+\mathrm{o}(1)) + \mathrm{o}(q_\tau^{-1})\right].$$

■

Proof of Proposition 4. As indicated in the main paper, the coherence of $\mathrm{XES}(\tau)$ is a straightforward consequence of the coherence of the expectile-based VaR. Here, we focus on the translation invariance and positive homogeneity of the Tail Conditional Expectation

$$\mathrm{XTCE}(\tau; X) := \mathbb{E}(X | X > \xi_\tau(X))$$

where $\xi_\tau(X)$ denotes the $\tau-$th expectile of the random variable X. Recall that the expectile is itself a coherent risk measure, and hence satisfies these properties.

1. To prove translation invariance, let $c \in \mathbb{R}$ and write

$$\begin{aligned}\mathrm{XTCE}(\tau; X+c) = \mathbb{E}(X+c | X+c > \xi_\tau(X+c)) &= \mathbb{E}(X | X+c > \xi_\tau(X+c)) + c \\ &= \mathbb{E}(X | X+c > \xi_\tau(X)+c) + c \\ &= \mathbb{E}(X | X > \xi_\tau(X)) + c \\ &= \mathrm{XTCE}(\tau; X) + c.\end{aligned}$$

2. Positive homogeneity is shown in the same way: for any $a \geqslant 0$,

$$\begin{aligned}\mathrm{XTCE}(\tau; aX) = \mathbb{E}(aX | aX > \xi_\tau(aX)) &= a\mathbb{E}(X | aX > \xi_\tau(aX)) \\ &= a\mathbb{E}(X | aX > a\xi_\tau(X)) \\ &= a\mathbb{E}(X | X > \xi_\tau(X)) \\ &= a\mathrm{XTCE}(\tau; X).\end{aligned}$$

This completes the proof of the proposition. ■

Proof of Proposition 5. We shall actually prove the more general statement

$$\lim_{t\to\infty} \frac{\mathbb{E}(X|Y>t)}{U_X(1/\overline{F}_Y(t))} = \int_0^\infty R(x^{-1/\gamma_X},1)dx \tag{B.30}$$

which contains both Proposition 1 in Cai *et al.* (2015) and our desired result, because $\xi_{Y,\tau} \to \infty$ as $\tau \uparrow 1$. For any $t>0$,

$$\begin{aligned}
\frac{\mathbb{E}(X|Y>t)}{U_X(1/\overline{F}_Y(t))} &= \frac{1}{U_X(1/\overline{F}_Y(t))} \int_0^\infty \frac{\mathbb{P}(X>s, Y>t)}{\overline{F}_Y(t)} ds \\
&= \frac{1}{U_X(1/\overline{F}_Y(t))} \int_0^\infty \frac{\mathbb{P}(\overline{F}_X(X) \leqslant \overline{F}_X(s), \overline{F}_Y(Y) \leqslant \overline{F}_Y(t))}{\overline{F}_Y(t)} ds \\
&= \int_0^\infty \frac{\mathbb{P}(\overline{F}_X(X) \leqslant \overline{F}_X(U_X(1/\overline{F}_Y(t))x), \overline{F}_Y(Y) \leqslant \overline{F}_Y(t))}{\overline{F}_Y(t)} dx.
\end{aligned} \tag{B.31}$$

Note now that because X is heavy-tailed, $\overline{F}_X(Tx) \sim x^{-1/\gamma_X} \overline{F}_X(T)$ as $T \to \infty$ and we have that:

$$\forall x>0,\ \overline{F}_X(U_X(1/\overline{F}_Y(t))x) \sim x^{-1/\gamma_X} \overline{F}_Y(t) \ \text{ as } \ t \to \infty.$$

Thus, by condition $\mathscr{JC}(R)$:

$$\forall x>0,\ \lim_{t\to\infty} \frac{\mathbb{P}(\overline{F}_X(X) \leqslant \overline{F}_X(U_X(1/\overline{F}_Y(t))x), \overline{F}_Y(Y) \leqslant \overline{F}_Y(t))}{\overline{F}_Y(t)} = R(x^{-1/\gamma_X},1). \tag{B.32}$$

It only remains to show that the integral in (B.31) and the limit in (B.32) can be interchanged, and this can be done exactly in the same way as in the proof of Proposition 1 of Cai *et al.* (2015), so we omit the details.

To show the second convergence result (20), we apply (B.30) to $t = \xi_{Y,\tau}$ and $t = q_{Y,\tau}$ in conjunction with (4) to get

$$\lim_{\tau\uparrow 1} \frac{\text{XMES}(\tau)}{\text{QMES}(\tau)} = \lim_{\tau\uparrow 1} \frac{U_X(1/\overline{F}_Y(\xi_{Y,\tau}))}{U_X(1/\overline{F}_Y(q_{Y,\tau}))} = \lim_{\tau\uparrow 1} \left(\frac{\overline{F}_Y(q_{Y,\tau})}{\overline{F}_Y(\xi_{Y,\tau})}\right)^{\gamma_X} = \left(\gamma_Y^{-1} - 1\right)^{-\gamma_X}.$$

■

Proof of Theorem 4. We start by the case when $X>0$ almost surely. In this situation,

$$\widetilde{\text{XMES}}(\tau_n) = \frac{\sum_{i=1}^n X_i \mathbb{I}\{Y_i > \widetilde{\xi}_{Y,\tau_n}\}}{\sum_{i=1}^n \mathbb{I}\{Y_i > \widetilde{\xi}_{Y,\tau_n}\}}.$$

Write then

$$\begin{aligned}
\log\left(\frac{\widetilde{\text{XMES}}^\star(\tau_n')}{\text{XMES}(\tau_n')}\right) &= \log\left(\frac{\widetilde{\text{XMES}}(\tau_n)}{\text{XMES}(\tau_n)}\right) + \log\left(\frac{\text{XMES}(\tau_n)}{\text{XMES}(\tau_n')}\left(\frac{1-\tau_n'}{1-\tau_n}\right)^{-\gamma_X}\right) \\
&+ (\widehat{\gamma}_X - \gamma_X)\log\left(\frac{1-\tau_n}{1-\tau_n'}\right).
\end{aligned}$$

Using the delta-method, the proof shall then be complete provided that

$$\sqrt{n(1-\tau_n)}\left(\frac{\widetilde{\text{XMES}}(\tau_n)}{\text{XMES}(\tau_n)}-1\right)=\mathrm{O}_{\mathbb{P}}(1) \tag{B.33}$$

and

$$\sqrt{n(1-\tau_n)}\left(\frac{\text{XMES}(\tau_n)}{\text{XMES}(\tau_n')}\left(\frac{1-\tau_n'}{1-\tau_n}\right)^{-\gamma_X}-1\right)=\mathrm{O}(1). \tag{B.34}$$

To show (B.33), write

$$\frac{\widetilde{\text{XMES}}(\tau_n)}{\text{XMES}(\tau_n)}=\frac{\mathbb{E}(\mathbb{I}_{\{Y>\xi_{Y,\tau_n}\}})}{\frac{1}{n}\sum_{i=1}^n \mathbb{I}_{\{Y_i>\xi_{Y,\tau_n}\}}}\frac{\frac{1}{n}\sum_{i=1}^n X_i\mathbb{I}_{\{Y_i>\xi_{Y,\tau_n}\}}}{\mathbb{E}(X\mathbb{I}_{\{Y>\xi_{Y,\tau_n}\}})}\frac{\sum_{i=1}^n \mathbb{I}_{\{Y_i>\xi_{Y,\tau_n}\}}}{\sum_{i=1}^n \mathbb{I}_{\{Y_i>\tilde{\xi}_{Y,\tau_n}\}}}\frac{\sum_{i=1}^n X_i\mathbb{I}_{\{Y_i>\tilde{\xi}_{Y,\tau_n}\}}}{\sum_{i=1}^n X_i\mathbb{I}_{\{Y_i>\xi_{Y,\tau_n}\}}}. \tag{B.35}$$

Firstly,

$$\sqrt{n(1-\tau_n)}\left(\frac{\frac{1}{n}\sum_{i=1}^n \mathbb{I}_{\{Y_i>\xi_{Y,\tau_n}\}}}{\mathbb{E}(\mathbb{I}_{\{Y>\xi_{Y,\tau_n}\}})}-1\right)=\mathrm{O}_{\mathbb{P}}(1) \tag{B.36}$$

because the variance of the term on the left-hand side is bounded in view of Proposition 1. Secondly,

$$\begin{aligned}\operatorname{Var}\left[\sqrt{n(1-\tau_n)}\left(\frac{\frac{1}{n}\sum_{i=1}^n X_i\mathbb{I}_{\{Y_i>\xi_{Y,\tau_n}\}}}{\mathbb{E}(X\mathbb{I}_{\{Y>\xi_{Y,\tau_n}\}})}-1\right)\right] &\leqslant \frac{(1-\tau_n)\mathbb{E}(X^2\mathbb{I}_{\{Y>\xi_{Y,\tau_n}\}})}{[\mathbb{E}(X\mathbb{I}_{\{Y>\xi_{Y,\tau_n}\}})]^2}\\ &= \frac{(1-\tau_n)}{\mathbb{P}(Y>\xi_{Y,\tau_n})}\frac{\mathbb{E}(X^2|Y>\xi_{Y,\tau_n})}{[\mathbb{E}(X|Y>\xi_{Y,\tau_n})]^2}.\end{aligned}$$

Applying Proposition 1 and then Proposition 5,

$$\operatorname{Var}\left[\sqrt{n(1-\tau_n)}\left(\frac{\frac{1}{n}\sum_{i=1}^n X_i\mathbb{I}_{\{Y_i>\xi_{Y,\tau_n}\}}}{\mathbb{E}(X\mathbb{I}_{\{Y>\xi_{Y,\tau_n}\}})}-1\right)\right]=\mathrm{O}\left(\frac{\mathbb{E}(X^2|Y>\xi_{Y,\tau_n})}{[U_X(1/\overline{F}_Y(\xi_{Y,\tau_n}))]^2}\right).$$

Notice then that condition $\mathcal{JC}(R)$ is equivalent, for all x and y which are not both infinite, to

$$\lim_{t\to\infty} t\mathbb{P}(X\geqslant U_X(t/x), Y\geqslant U_Y(t/y))=R(x,y).$$

Since $(U_X)^2=U_{X^2}$ (because $X>0$), this entails

$$\lim_{t\to\infty} t\mathbb{P}(X^2\geqslant U_{X^2}(t/x), Y\geqslant U_Y(t/y))=R(x,y).$$

Hence, (X^2,Y) also satisfies condition $\mathcal{JC}(R)$. Thus, by Proposition 5,

$$\frac{\mathbb{E}(X^2|Y>\xi_{Y,\tau_n})}{[U_X(1/\overline{F}_Y(\xi_{Y,\tau_n}))]^2}=\frac{\mathbb{E}(X^2|Y>\xi_{Y,\tau_n})}{U_{X^2}(1/\overline{F}_Y(\xi_{Y,\tau_n}))}=\mathrm{O}(1)$$

which entails

$$\operatorname{Var}\left[\sqrt{n(1-\tau_n)}\left(\frac{\frac{1}{n}\sum_{i=1}^n X_i\mathbb{I}_{\{Y_i>\xi_{Y,\tau_n}\}}}{\mathbb{E}(X\mathbb{I}_{\{Y>\xi_{Y,\tau_n}\}})}-1\right)\right]=\mathrm{O}(1)$$

and therefore

$$\sqrt{n(1-\tau_n)}\left(\frac{\frac{1}{n}\sum_{i=1}^n X_i \mathbb{I}_{\{Y_i>\xi_{Y,\tau_n}\}}}{\mathbb{E}(X\mathbb{I}_{\{Y>\xi_{Y,\tau_n}\}})}-1\right)=\mathrm{O}_{\mathbb{P}}(1). \tag{B.37}$$

Thirdly, by Theorem 2, $\widetilde{\xi}_{Y,\tau_n}$ is $\sqrt{n(1-\tau_n)}$-relatively consistent, so that for any $\varepsilon>0$, we may find $K>0$ such that

$$\left|\frac{\widetilde{\xi}_{Y,\tau_n}}{\xi_{Y,\tau_n}}-1\right|\leqslant\frac{K}{\sqrt{n(1-\tau_n)}}$$

with probability larger than $1-\varepsilon$ eventually. In what follows we assume that K is chosen so that this is the case. With probability larger than $1-\varepsilon$ eventually, we then have

$$\begin{aligned}&\left|\frac{\sum_{i=1}^n \mathbb{I}_{\{Y_i>\widetilde{\xi}_{Y,\tau_n}\}}}{\sum_{i=1}^n \mathbb{I}_{\{Y_i>\xi_{Y,\tau_n}\}}}-1\right|\\ \leqslant\ &\max\left(\left|\frac{\sum_{i=1}^n \mathbb{I}_{\{Y_i>\xi_{Y,\tau_n}(1+K/\sqrt{n(1-\tau_n)})\}}}{\sum_{i=1}^n \mathbb{I}_{\{Y_i>\xi_{Y,\tau_n}\}}}-1\right|,\left|\frac{\sum_{i=1}^n \mathbb{I}_{\{Y_i>\xi_{Y,\tau_n}(1-K/\sqrt{n(1-\tau_n)})\}}}{\sum_{i=1}^n \mathbb{I}_{\{Y_i>\xi_{Y,\tau_n}\}}}-1\right|\right).\end{aligned}$$

By straightforward variance calculations,

$$\begin{aligned}&\sqrt{n(1-\tau_n)}\left(\frac{\sum_{i=1}^n \mathbb{I}_{\{Y_i>\xi_{Y,\tau_n}(1\pm K/\sqrt{n(1-\tau_n)})\}}}{\sum_{i=1}^n \mathbb{I}_{\{Y_i>\xi_{Y,\tau_n}\}}}-1\right)\\ =\ &\mathrm{O}_{\mathbb{P}}\left(\sqrt{n(1-\tau_n)}\left|\frac{\mathbb{P}(Y>\xi_{Y,\tau_n}(1\pm K/\sqrt{n(1-\tau_n)}))}{\mathbb{P}(Y>\xi_{Y,\tau_n})}-1\right|\right).\end{aligned}$$

By a uniform inequality such as Theorem B.3.10 in de Haan and Ferreira (2006) applied to the function $\overline{F}_Y$, we get

$$\frac{\mathbb{P}(Y>\xi_{Y,\tau_n}(1\pm K/\sqrt{n(1-\tau_n)}))}{\mathbb{P}(Y>\xi_{Y,\tau_n})}-1=\mathrm{O}\left(\frac{1}{\sqrt{n(1-\tau_n)}}\right) \tag{B.38}$$

and therefore

$$\sqrt{n(1-\tau_n)}\left(\frac{\sum_{i=1}^n \mathbb{I}_{\{Y_i>\widetilde{\xi}_{Y,\tau_n}\}}}{\sum_{i=1}^n \mathbb{I}_{\{Y_i>\xi_{Y,\tau_n}\}}}-1\right)=\mathrm{O}_{\mathbb{P}}(1). \tag{B.39}$$

Lastly, write with probability larger than $1-\varepsilon$ eventually:

$$\begin{aligned}&\left|\frac{\sum_{i=1}^n X_i\mathbb{I}_{\{Y_i>\widetilde{\xi}_{Y,\tau_n}\}}}{\sum_{i=1}^n X_i\mathbb{I}_{\{Y_i>\xi_{Y,\tau_n}\}}}-1\right|\\ \leqslant\ &\max\left(\left|\frac{\sum_{i=1}^n X_i\mathbb{I}_{\{Y_i>\xi_{Y,\tau_n}(1+K/\sqrt{n(1-\tau_n)})\}}}{\sum_{i=1}^n X_i\mathbb{I}_{\{Y_i>\xi_{Y,\tau_n}\}}}-1\right|,\left|\frac{\sum_{i=1}^n X_i\mathbb{I}_{\{Y_i>\xi_{Y,\tau_n}(1-K/\sqrt{n(1-\tau_n)})\}}}{\sum_{i=1}^n X_i\mathbb{I}_{\{Y_i>\xi_{Y,\tau_n}\}}}-1\right|\right).\end{aligned}$$

By a straightforward modification of (B.39),

$$\begin{aligned}&\sqrt{n(1-\tau_n)}\left(\frac{\sum_{i=1}^n X_i\mathbb{I}_{\{Y_i>\xi_{Y,\tau_n}(1\pm K/\sqrt{n(1-\tau_n)})\}}}{\sum_{i=1}^n X_i\mathbb{I}_{\{Y_i>\xi_{Y,\tau_n}\}}}-1\right)\\ =\ &\mathrm{O}_{\mathbb{P}}\left(\sqrt{n(1-\tau_n)}\left|\frac{\mathbb{E}(X\mathbb{I}_{\{Y>\xi_{Y,\tau_n}(1\pm K/\sqrt{n(1-\tau_n)})\}})}{\mathbb{E}(X\mathbb{I}_{\{Y>\xi_{Y,\tau_n}\}})}-1\right|\right).\end{aligned}$$

Applying (B.38) and Proposition 5, we obtain

$$
\begin{aligned}
&\sqrt{n(1-\tau_n)}\left(\frac{\sum_{i=1}^n X_i \mathbb{I}_{\{Y_i>\xi_{Y,\tau_n}(1\pm K/\sqrt{n(1-\tau_n)})\}}}{\sum_{i=1}^n X_i \mathbb{I}_{\{Y_i>\xi_{Y,\tau_n}\}}}-1\right)\\
&= \mathrm{O}_{\mathbb{P}}\left(\sqrt{n(1-\tau_n)}\left|\frac{\mathbb{E}(X|Y>\xi_{Y,\tau_n}(1\pm K/\sqrt{n(1-\tau_n)}))}{\mathbb{E}(X|Y>\xi_{Y,\tau_n})}-1\right|\right)\\
&= \mathrm{O}_{\mathbb{P}}\left(\sqrt{n(1-\tau_n)}\left|\frac{\mathbb{E}(X|Y>\xi_{Y,\tau_n}(1\pm K/\sqrt{n(1-\tau_n)}))-\mathbb{E}(X|Y>\xi_{Y,\tau_n})}{U_X(1/\overline{F}_Y(\xi_{Y,\tau_n}))}\right|\right).
\end{aligned}
$$

It is therefore enough to show that

$$
\frac{|\mathbb{E}(X|Y>\xi_{Y,\tau_n}(1\pm K/\sqrt{n(1-\tau_n)}))-\mathbb{E}(X|Y>\xi_{Y,\tau_n})|}{U_X(1/\overline{F}_Y(\xi_{Y,\tau_n}))}=\mathrm{O}_{\mathbb{P}}\left(\frac{1}{\sqrt{n(1-\tau_n)}}\right).
$$

Because expectiles and quantiles are asymptotically proportional in view of Corollary 1:

$$
\xi_{Y,\tau_n}=(\gamma_Y^{-1}-1)^{-\gamma_Y}q_{Y,\tau_n}\left(1+\mathrm{O}\left(\frac{1}{\sqrt{n(1-\tau_n)}}\right)\right),
$$

this can be achieved by using condition $\mathcal{JC}_2(R,\beta,\kappa)$ in the same way followed by Cai *et al.* (2015) to examine the convergence of the term J_2 that they introduce in the proof of their Proposition 3, see pp.438-439 therein. We then get

$$
\sqrt{n(1-\tau_n)}\left(\frac{\sum_{i=1}^n X_i \mathbb{I}_{\{Y_i>\tilde{\xi}_{Y,\tau_n}\}}}{\sum_{i=1}^n X_i \mathbb{I}_{\{Y_i>\xi_{Y,\tau_n}\}}}-1\right)=\mathrm{O}_{\mathbb{P}}(1). \tag{B.40}
$$

Combining (B.35), (B.36), (B.37), (B.39) and (B.40) concludes the proof of (B.33).

Now, to prove (B.34), notice first that

$$
\begin{aligned}
\sqrt{n(1-\tau_n)}\left(\frac{\mathrm{XMES}(\tau_n)}{U_X(1/\overline{F}_Y(\xi_{Y,\tau_n}))}-\int_0^\infty R(x^{-1/\gamma_X},1)dx\right) &= \mathrm{O}(1) \qquad \text{(B.41)}\\
\text{and}\quad \sqrt{n(1-\tau_n)}\left(\frac{\mathrm{XMES}(\tau_n')}{U_X(1/\overline{F}_Y(\xi_{Y,\tau_n'}))}-\int_0^\infty R(x^{-1/\gamma_X},1)dx\right) &= \mathrm{O}(1); \qquad \text{(B.42)}
\end{aligned}
$$

this can be verified using condition $\mathcal{JC}_2(R,\beta,\kappa)$ along the lines of proof of Lemma 3 and (28) in Cai *et al.* (2015), because expectiles and quantiles are asymptotically proportional. Besides, by Proposition 1,

$$
\begin{aligned}
\sqrt{n(1-\tau_n)}\left(\frac{\overline{F}_Y(\xi_{Y,\tau_n})}{1-\tau_n}-(\gamma_Y^{-1}-1)\right) &= \mathrm{O}(1)\\
\text{and}\quad \sqrt{n(1-\tau_n)}\left(\frac{\overline{F}_Y(\xi_{Y,\tau_n'})}{1-\tau_n'}-(\gamma_Y^{-1}-1)\right) &= \mathrm{O}(1)
\end{aligned}
$$

so that by condition $\mathcal{C}_2(\gamma_X, \rho_X, A_X)$ and convergence $\sqrt{n(1-\tau_n)}A_X((1-\tau_n)^{-1}) \to \lambda_2 \in \mathbb{R}$,

$$\sqrt{n(1-\tau_n)}\left(\frac{U_X(1/\overline{F}_Y(\xi_{Y,\tau_n}))}{U_X(1/\overline{F}_Y(\xi_{Y,\tau_n'}))}\left(\frac{1-\tau_n'}{1-\tau_n}\right)^{-\gamma_X}-1\right) = \mathrm{O}(1). \tag{B.43}$$

Combining (B.41), (B.42) and (B.43) completes the proof of (B.34).

We now show how the condition that $X > 0$ almost surely can be dropped in our framework. Define $X_+ = \max(X, 0)$ and

$$\mathrm{XMES}^+(\tau_n') := \mathbb{E}(X_+|Y > \xi_{Y,\tau_n'}),$$

i.e. XMES^+ is the marginal expected shortfall of the positive part of X, and write

$$\frac{\widetilde{\mathrm{XMES}}^\star(\tau_n')}{\mathrm{XMES}(\tau_n')} = \frac{\widetilde{\mathrm{XMES}}^\star(\tau_n')}{\mathrm{XMES}^+(\tau_n')}\frac{\mathrm{XMES}^+(\tau_n')}{\mathrm{XMES}(\tau_n')}. \tag{B.44}$$

The first part of the proof of Theorem 2 in Cai *et al.* (2015), see pp.440–441 and in particular condition (35) there, shows that X_+ satisfies condition $\mathcal{JC}_2(R, \beta, \kappa)$. As a consequence, we may apply the result we have just shown to the random variable X_+ to get

$$\frac{\sqrt{n(1-\tau_n)}}{\log[(1-\tau_n)/(1-\tau_n')]}\left(\frac{\widetilde{\mathrm{XMES}}^\star(\tau_n')}{\mathrm{XMES}^+(\tau_n')}-1\right) \xrightarrow{d} \Gamma. \tag{B.45}$$

Let now $X_- = X - X_+$ so that

$$\frac{\mathrm{XMES}(\tau_n')}{\mathrm{XMES}^+(\tau_n')} = 1 + \frac{\mathbb{E}(X_-|Y > \xi_{Y,\tau_n'})}{\mathrm{XMES}^+(\tau_n')}. \tag{B.46}$$

Since $\xi_{Y,\tau_n'} \uparrow \infty$, we have for n large enough that

$$\frac{\mathrm{XMES}^+(\tau_n')}{U_X(1/\overline{F}_Y(\xi_{Y,\tau_n'}))} = \frac{\mathrm{XMES}^+(\tau_n')}{U_{X^+}(1/\overline{F}_Y(\xi_{Y,\tau_n'}))}$$

so that

$$\frac{\mathrm{XMES}^+(\tau_n')}{U_X(1/\overline{F}_Y(\xi_{Y,\tau_n'}))} \to \int_0^\infty R(x^{-1/\gamma_X}, 1)dx$$

as $n \to \infty$ and therefore

$$\frac{\mathrm{XMES}(\tau_n')}{\mathrm{XMES}^+(\tau_n')} = 1 + \mathrm{O}\left(\frac{\mathbb{E}(X_-|Y > \xi_{Y,\tau_n'})}{U_X(1/\overline{F}_Y(\xi_{Y,\tau_n'}))}\right). \tag{B.47}$$

Since extreme expectiles and extreme quantiles are asymptotically proportional, we have as in Cai *et al.* (2015) that

$$|\mathbb{E}(X_-|Y > \xi_{Y,\tau_n'})| = \mathrm{O}\left((1-\tau_n')^{-1+(1-\kappa)(1-\gamma_X)}\right) \text{ and } \frac{1}{U_X(1/\overline{F}_Y(\xi_{Y,\tau_n'}))} = \mathrm{O}\left((1-\tau_n')^{\gamma_X}\right). \tag{B.48}$$

Plugging this into (B.47) entails

$$\frac{\mathrm{XMES}(\tau_n')}{\mathrm{XMES}^+(\tau_n')} = 1 + \mathrm{O}\left((1-\tau_n')^{-\kappa(1-\gamma_X)}\right) = 1 + \mathrm{o}\left(\frac{1}{\sqrt{n(1-\tau_n)}}\right). \tag{B.49}$$

Plugging (B.49) and (B.45) into (B.44) concludes the proof. ■

Proof of Theorem 5. Write

$$\log\left(\frac{\widehat{\text{XMES}}^{\star}(\tau_n')}{\text{XMES}(\tau_n')}\right) = \log\left(\frac{\widehat{\text{QMES}}^{\star}(\tau_n')}{\text{QMES}(\tau_n')}\right) + \log\left(\frac{(\widehat{\gamma}_Y^{-1}-1)^{-\widehat{\gamma}_X}}{(\gamma_Y^{-1}-1)^{-\gamma_X}}\right) - \log\left((\gamma_Y^{-1}-1)^{\gamma_X}\frac{\text{XMES}(\tau_n')}{\text{QMES}(\tau_n')}\right).$$

Firstly, by Theorem 2 in Cai *et al.* (2015),

$$\frac{\sqrt{n(1-\tau_n)}}{\log[(1-\tau_n)/(1-\tau_n')]}\log\left(\frac{\widehat{\text{QMES}}^{\star}(\tau_n')}{\text{QMES}(\tau_n')}\right) \xrightarrow{d} \Gamma. \tag{B.50}$$

Secondly, writing

$$\log\left(\frac{(\widehat{\gamma}_Y^{-1}-1)^{-\widehat{\gamma}_X}}{(\gamma_Y^{-1}-1)^{-\gamma_X}}\right) = -\left[(\widehat{\gamma}_X-\gamma_X)\log(\widehat{\gamma}_Y^{-1}-1) + \gamma_X\left(\log(\widehat{\gamma}_Y^{-1}-1) - \log(\gamma_Y^{-1}-1)\right)\right]$$

we get

$$\frac{\sqrt{n(1-\tau_n)}}{\log[(1-\tau_n)/(1-\tau_n')]}\log\left(\frac{(\widehat{\gamma}_Y^{-1}-1)^{-\widehat{\gamma}_X}}{(\gamma_Y^{-1}-1)^{-\gamma_X}}\right) \xrightarrow{\mathbb{P}} 0. \tag{B.51}$$

Thirdly, by (B.42), which is

$$\sqrt{n(1-\tau_n)}\left(\frac{\text{XMES}^{+}(\tau_n')}{U_{X^+}(1/\overline{F}_Y(\xi_{Y,\tau_n'}))} - \int_0^\infty R(x^{-1/\gamma_X},1)dx\right) = \text{O}(1)$$

together with (B.49) and the equality $U_{X^+}=U_X$ in a neighborhood of infinity, we obtain

$$\sqrt{n(1-\tau_n)}\left(\frac{\text{XMES}(\tau_n')}{U_X(1/\overline{F}_Y(\xi_{Y,\tau_n'}))} - \int_0^\infty R(x^{-1/\gamma_X},1)dx\right) = \text{O}(1).$$

The similar relationship (Cai *et al.*, 2015)

$$\sqrt{n(1-\tau_n)}\left(\frac{\text{QMES}(\tau_n')}{U_X(1/\overline{F}_Y(q_{Y,\tau_n'}))} - \int_0^\infty R(x^{-1/\gamma_X},1)dx\right) = \text{O}(1)$$

then yields

$$\log\left((\gamma_Y^{-1}-1)^{\gamma_X}\frac{\text{XMES}(\tau_n')}{\text{QMES}(\tau_n')}\right) = \log\left((\gamma_Y^{-1}-1)^{\gamma_X}\frac{U_X(1/\overline{F}_Y(\xi_{Y,\tau_n'}))}{U_X(1/\overline{F}_Y(q_{Y,\tau_n'}))}\right) + \text{O}\left(\frac{1}{\sqrt{n(1-\tau_n)}}\right). \tag{B.52}$$

Now, by Proposition 1,

$$\sqrt{n(1-\tau_n)}\left(\frac{\overline{F}_Y(\xi_{Y,\tau_n'})}{1-\tau_n'} - (\gamma_Y^{-1}-1)\right) = \text{O}(1)$$

and $\overline{F}_Y(q_{Y,\tau_n'}) = 1-\tau_n'$ by continuity of F_Y, so that by condition $\mathcal{C}_2(\gamma_X,\rho_X,A_X)$ and convergence $\sqrt{n(1-\tau_n)}A_X((1-\tau_n)^{-1}) \to 0$,

$$\sqrt{n(1-\tau_n)}\left((\gamma_Y^{-1}-1)^{\gamma_X}\frac{U_X(1/\overline{F}_Y(\xi_{Y,\tau_n'}))}{U_X(1/\overline{F}_Y(q_{Y,\tau_n'}))} - 1\right) = \text{O}(1).$$

In conjunction with (B.52), this entails

$$\frac{\sqrt{n(1-\tau_n)}}{\log[(1-\tau_n)/(1-\tau_n')]} \log\left((\gamma_Y^{-1}-1)^{\gamma_X} \frac{\mathrm{XMES}(\tau_n')}{\mathrm{QMES}(\tau_n')} \right) \to 0. \tag{B.53}$$

A combination of (B.50), (B.51), (B.53) and the delta-method completes the proof. ■

Proof of Proposition 6. We start by obtaining an equivalent for the numerator of $1-\tau_n'(\alpha_n)$, which is equal to

$$q_{\alpha_n} \mathbb{E}\left(\left[\frac{Y}{q_{\alpha_n}} - 1 \right] 1\!\!1\{Y/q_{\alpha_n} > 1\} \right).$$

Just as in the proof of Proposition 1, we integrate by parts to obtain

$$\mathbb{E}\left(\left[\frac{Y}{q_{\alpha_n}} - 1 \right] 1\!\!1\{Y/q_{\alpha_n} > 1\} \right) = \overline{F}(q_{\alpha_n}) \left(\frac{\gamma}{1-\gamma} + \int_1^{+\infty} \left[\frac{\overline{F}(q_{\alpha_n} x)}{\overline{F}(q_{\alpha_n})} - x^{-1/\gamma} \right] dx \right).$$

Since $q_{\alpha_n} \to \infty$ as $n \to \infty$, we can apply Proposition B.1.10 in de Haan and Ferreira (2006) to the function $\overline{F}$ to get

$$\mathbb{E}\left(\left[\frac{Y}{q_{\alpha_n}} - 1 \right] 1\!\!1\{Y/q_{\alpha_n} > 1\} \right) = \overline{F}(q_{\alpha_n}) \left(\frac{\gamma}{1-\gamma} + \mathrm{o}(1) \right) = (1-\alpha_n)\frac{\gamma}{1-\gamma}(1+\mathrm{o}(1)).$$

To obtain an equivalent of the denominator, we note that

$$\mathbb{E}\,|Y - q_{\alpha_n}| = q_{\alpha_n} \mathbb{E} \left| \frac{Y}{q_{\alpha_n}} - 1 \right| = q_{\alpha_n}(1+\mathrm{o}(1))$$

where we used the dominated convergence theorem together with the fact that $q_{\alpha_n} \to \infty$. Wrapping up, we obtain

$$\frac{\mathbb{E}\left(|Y - q_{\alpha_n}|\, 1\!\!1\{Y > q_{\alpha_n}\} \right)}{\mathbb{E}\,|Y - q_{\alpha_n}|} = (1-\alpha_n)\frac{\gamma}{1-\gamma}(1+\mathrm{o}(1))$$

which is the desired result. ■

Proof of Theorem 6. Our first goal is to show that

$$\frac{1-\widehat{\tau}_n'(\alpha_n)}{1-\tau_n'(\alpha_n)} - 1 = \mathrm{O}_{\mathbb{P}}(1). \tag{B.54}$$

To this end, we write

$$\frac{1-\widehat{\tau}_n'(\alpha_n)}{1-\tau_n'(\alpha_n)} - 1 = \frac{\widehat{\gamma}}{\gamma} \times \frac{1-\gamma}{1-\widehat{\gamma}} \times \frac{(1-\alpha_n)\dfrac{\gamma}{1-\gamma}}{1-\tau_n'(\alpha_n)} - 1. \tag{B.55}$$

The delta-method yields

$$\sqrt{n(1-\tau_n)} \left(\frac{\gamma}{\widehat{\gamma}} \times \frac{1-\widehat{\gamma}}{1-\gamma} - 1 \right) = \mathrm{O}_{\mathbb{P}}(1). \tag{B.56}$$

Recall now (B.4) in the proof of Proposition 1 which here translates into

$$\frac{(1-\alpha_n)\dfrac{\gamma}{1-\gamma}}{\mathbb{E}\left[\left|\dfrac{Y}{q_{\alpha_n}}-1\right| \mathbb{1}_{\{Y>q_{\alpha_n}\}}\right]} - 1 = \mathrm{O}[A(1/\overline{F}(q_{\alpha_n}))] = \mathrm{O}[A((1-\alpha_n)^{-1})] = \mathrm{O}(1/\sqrt{n(1-\tau_n)}), \tag{B.57}$$

because $q_{\alpha_n} = \xi_{\tau_n'(\alpha_n)}$ and using the regular variation property of A. Write further

$$\begin{aligned}
\mathbb{E}\left|\frac{Y}{q_{\alpha_n}}-1\right| - 1 &= \mathbb{E}\left[\left|\frac{Y}{q_{\alpha_n}}-1\right| \mathbb{1}_{\{Y>q_{\alpha_n}\}}\right] + \mathbb{E}\left[\left(1-\frac{Y}{q_{\alpha_n}}\right) \mathbb{1}_{\{Y\leqslant q_{\alpha_n}\}}\right] - 1 \\
&= \mathbb{E}\left[\left|\frac{Y}{q_{\alpha_n}}-1\right| \mathbb{1}_{\{Y>q_{\alpha_n}\}}\right] - \frac{\mathbb{E}(Y\,\mathbb{1}_{\{Y\leqslant q_{\alpha_n}\}})}{q_{\alpha_n}} - \overline{F}(q_{\alpha_n}) \\
&= \mathrm{O}\left(\max\left\{1-\alpha_n, 1/q_{\alpha_n}\right\}\right) = \mathrm{O}\left(1/q_{\alpha_n}\right) = \mathrm{o}(1/\sqrt{n(1-\tau_n)})
\end{aligned} \tag{B.58}$$

where we successively used (B.57), the dominated convergence theorem, the relationship $1-\alpha_n = \mathrm{o}(1/q_{\alpha_n})$ valid because $0<\gamma<1$, and the regular variation property of $t \mapsto q_{1-t^{-1}}$. Combining (B.55), (B.56), (B.57) and (B.58) with the definition

$$1-\tau_n'(\alpha_n) = \frac{\mathbb{E}\left\{|Y-q_{\alpha_n}|\,\mathbb{1}(Y>q_{\alpha_n})\right\}}{\mathbb{E}\,|Y-q_{\alpha_n}|}$$

results in (B.54).

The idea to prove (i) is now to write

$$\widehat{\xi}^{\star}_{\widehat{\tau}_n'(\alpha_n)} = \left(\frac{1-\widehat{\tau}_n'(\alpha_n)}{1-\tau_n}\right)^{-\widehat{\gamma}} \widehat{\xi}_{\tau_n} = \left(\frac{1-\widehat{\tau}_n'(\alpha_n)}{1-\tau_n'(\alpha_n)}\right)^{-\widehat{\gamma}} \times \left\{\left(\frac{1-\tau_n'(\alpha_n)}{1-\tau_n}\right)^{-\widehat{\gamma}} \widehat{\xi}_{\tau_n}\right\}. \tag{B.59}$$

We have

$$\begin{aligned}
\left(\frac{1-\widehat{\tau}_n'(\alpha_n)}{1-\tau_n'(\alpha_n)}\right)^{-\widehat{\gamma}} &= \exp\left(-\widehat{\gamma}\log\left[\frac{1-\widehat{\tau}_n'(\alpha_n)}{1-\tau_n'(\alpha_n)}\right]\right) \\
&= \exp\left(-\left[\gamma + \mathrm{O}_{\mathbb{P}}\left(\frac{1}{\sqrt{n(1-\tau_n)}}\right)\right] \times \mathrm{O}_{\mathbb{P}}\left(\frac{1}{\sqrt{n(1-\tau_n)}}\right)\right) \\
&= 1 + \mathrm{O}_{\mathbb{P}}\left(\frac{1}{\sqrt{n(1-\tau_n)}}\right)
\end{aligned} \tag{B.60}$$

by a Taylor expansion. Furthermore

$$\left(\frac{1-\tau_n'(\alpha_n)}{1-\tau_n}\right)^{-\widehat{\gamma}} \widehat{\xi}_{\tau_n} = \widehat{\xi}^{\star}_{\tau_n'(\alpha_n)}$$

by definition of the class of estimators $\widehat{\xi}^\star$. From Proposition 6, we conclude that the conditions of Corollary 3 are satisfied if the parameter τ_n' there is replaced by $\tau_n'(\alpha_n)$. By Corollary 3 then:

$$\frac{\sqrt{n(1-\tau_n)}}{\log[(1-\tau_n)/(1-\tau_n'(\alpha_n))]}\left(\frac{\widehat{\xi}^\star_{\tau_n'(\alpha_n)}}{\xi_{\tau_n'(\alpha_n)}}-1\right)\xrightarrow{d}\Gamma.$$

Finally

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

and the first term above tends to infinity, while the second term converges to a finite constant in view of Proposition 6. Consequently

$$\log\left[\frac{1-\tau_n}{1-\tau_n'(\alpha_n)}\right]=\log\left[\frac{1-\tau_n}{1-\alpha_n}\right](1+\mathrm{o}(1)).$$

Together with the equality $\xi_{\tau_n'(\alpha_n)}=q_{\alpha_n}$ which is true by definition of $\tau_n'(\alpha_n)$, this entails

$$\frac{\sqrt{n(1-\tau_n)}}{\log[(1-\tau_n)/(1-\alpha_n)]}\left(\frac{\widehat{\xi}^\star_{\tau_n'(\alpha_n)}}{q_{\alpha_n}}-1\right)\xrightarrow{d}\Gamma. \tag{B.61}$$

Combining (B.59), (B.60) and (B.61) completes the proof of (i). The proof of (ii) is similar (just apply Corollary 4 instead of Corollary 3 when needed) and is therefore omitted. ■

Proof of Theorem 8. Again, we only show how to prove (i), the proof of (ii) being similar. Write

$$\begin{aligned}\widetilde{\mathrm{XMES}}^\star(\widehat{\tau}_n'(\alpha_n)) &= \left(\frac{1-\widehat{\tau}_n'(\alpha_n)}{1-\tau_n}\right)^{-\widehat{\gamma}}\widetilde{\mathrm{XMES}}(\tau_n)\\ &= \left(\frac{1-\widehat{\tau}_n'(\alpha_n)}{1-\tau_n'(\alpha_n)}\right)^{-\widehat{\gamma}}\times\left\{\left(\frac{1-\tau_n'(\alpha_n)}{1-\tau_n}\right)^{-\widehat{\gamma}}\widetilde{\mathrm{XMES}}(\tau_n)\right\}.\end{aligned}$$

The first term is controlled by using (B.54), and the second one is handled by arguing just as in the proof of Theorem 6, with $\widehat{\xi}$ replaced by $\widetilde{\mathrm{XMES}}$ throughout and by applying Theorem 4 instead of Corollary 3. We omit the details. ■

References

[1] Bellini, F., Klar, B., Müller, A. and Gianina, E.R. (2014). Generalized quantiles as risk measures, *Insurance: Mathematics and Economics*, **54**, 41–48.

[2] Bingham, N.H., Goldie, C.M. and Teugels, J.L. (1987). *Regular Variation*, Cambridge University Press.

[3] Cai, J., Einmahl, 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.

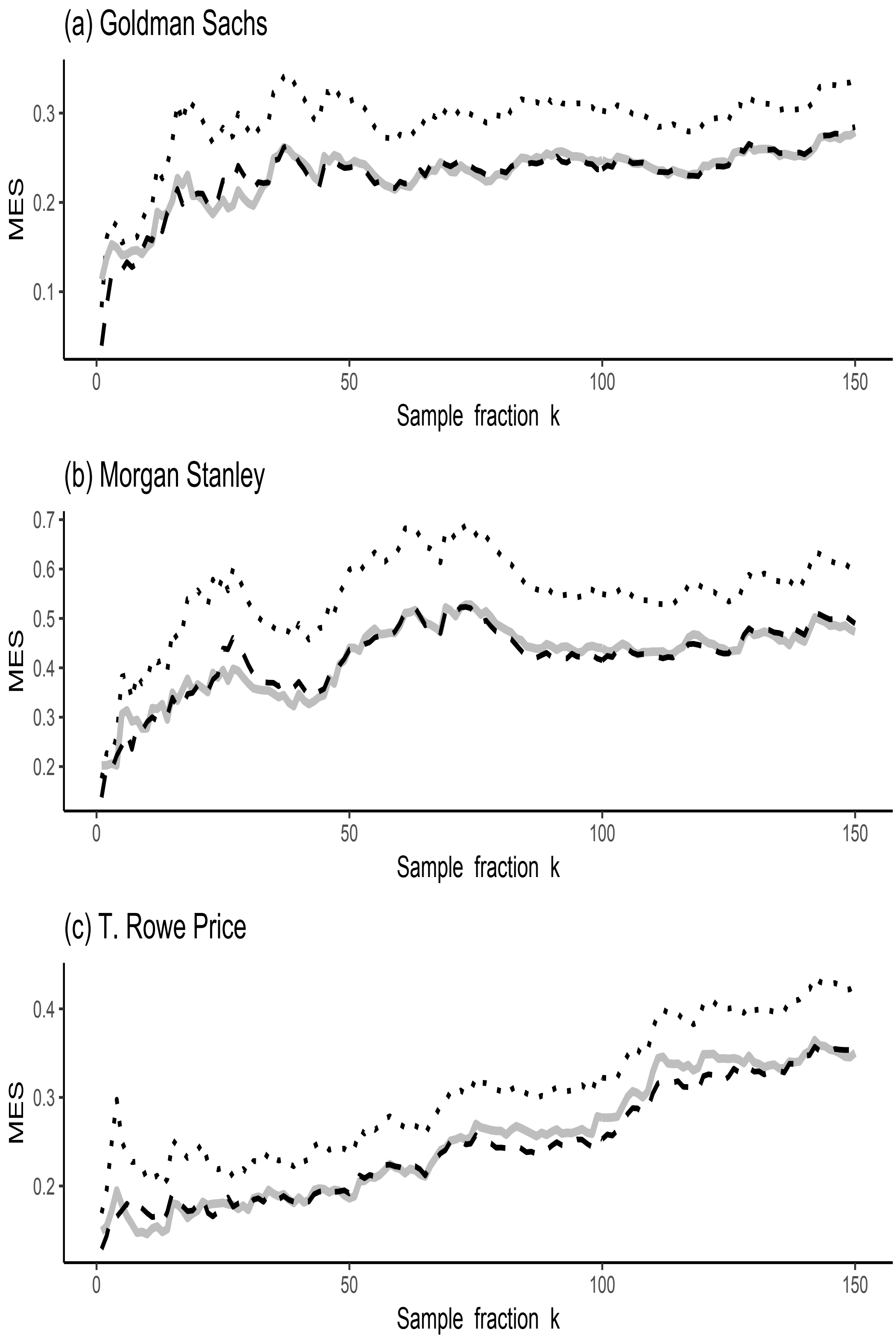

Figure 18: *The estimates* $\widehat{XMES}^{\star}$ *(dashed),* $\widetilde{XMES}^{\star}$ *(solid) and* $\widehat{QMES}^{\star}$ *(dotted) for the three banks.*

[4] de Haan, L. and Ferreira, A. (2006). *Extreme Value Theory: An Introduction*. Springer-Verlag, New York.

[5] De Rossi, G. and Harvey, H. (2009). Quantiles, expectiles and splines, *Journal of Econometrics*, **152**, 179–185.

[6] Geyer, C.J. (1996). On the asymptotics of convex stochastic optimization, unpublished manuscript.

[7] Gneiting, T. (2011). Making and evaluating point forecasts. *Journal of the American Statistical Association*, **106**, 746–762.

[8] Kuan, C-M., Yeh, J-H. and Hsu, Y-C. (2009). Assessing value at risk with CARE, the Conditional Autoregressive Expectile models, *Journal of Econometrics*, **2**, 261–270.

[9] Taylor, J. (2008). Estimating value at risk and expected shortfall using expectiles, *Journal of Financial Econometrics*, **6**, 231–252.