Commun. Math. Res.
doi: 10.4208/cmr.2022-0071

Vol. **40**, No. 1, pp. 102-124
February 2024

Uncertainty Comparison Between Value-at-Risk and Expected Shortfall

Qing Liu[1], Weimin Liu[2], Liang Peng[3,*] and Gengsheng Qin[2]

[1] *School of Statistics, Jiangxi University of Finance and Economics, Nanchang, Jiangxi 330013, China.*
[2] *Department of Mathematics and Statistics, Georgia State University, Atlanta, GA 30303, USA.*
[3] *Maurice R. Greenberg School of Risk Science, Georgia State University, Atlanta, GA 20303, USA.*

Received 26 December 2022; Accepted 2 June 2023

Abstract. Value-at-Risk (VaR) and expected shortfall (ES) are two key risk measures in financial risk management. Comparing these two measures has been a hot debate, and most discussions focus on risk measure properties. This paper uses independent data and autoregressive models with normal or t-distribution to examine the effect of the heavy tail and dependence on comparing the nonparametric inference uncertainty of these two risk measures. Theoretical and numerical analyses suggest that VaR at 99% level is better than ES at 97.5% level for distributions with heavier tails.

AMS subject classifications: 62P05, 62E20

Key words: α-mixing, asymptotic variance, expected shortfall, Value-at-Risk.

---

1 Introduction

Using economically meaningful risk measures is vital in market regulation, portfolio management, and the banking and insurance industry. Two popular risk measures are Value-at-Risk (VaR) and expected shortfall (ES). The Value-at-Risk

*Corresponding author. *Email address:* `lpeng@gsu.edu` (L. Peng)

has been adopted for measuring market risk in trading portfolios since 1990. Because of its lack of subadditivity and insensitivity to extreme losses, Artzner *et al.* [3] advocate the coherent expected shortfall risk measure. In 2016, the minimum capital requirement for market risk in the recent revision by the Basel Committee on Banking Supervision (BCBS) had moved from Value-at-Risk at 99% level to expected shortfall at 97.5% level to capture more extreme risks (Danielsson and Zhou [9]). The reason to use different risk levels is that the difference between these two risk measures is tiny when the loss has the standard normal distribution.

Comparing these two risk measures has been hot and intensive in the literature. Emmer *et al.* [11] compare the pros and cons of Value-at-Risk and expected shortfall and argue that expected shortfall is better in practice, despite some shortcomings regarding its estimation backtesting. Embrechts *et al.* [10] discuss from risk aggregation and model uncertainty viewpoint and provide a broadly accessible critical assessment of the Value-at-Risk and expected shortfall debate triggered by Basel III. Because Cont *et al.* [8] argue that robustness is as vital as the coherence properties, Kou *et al.* [16] compare these two risk measures using robustness related to model misspecification and tiny changes in data. Krätschmer *et al.* [17] compare a list of risk measures, including Value-at-Risk and expected shortfall by the index of qualitative robustness. Gneiting [14] shows that ES is not elicitable, while Fissler and Ziegel [13] show that ES is jointly elicitable with VaR.

In this paper, we theoretically and empirically examine the effect of heavy tails and serial dependence on comparing the nonparametric inference efficiency of the Value-at-Risk at 99% level and the expected shortfall at 97.5% level. A related but different study is Barnard *et al.* [4], where they compare the nonparametric inference efficiency using independent observations with exponential power distributions. Our main conclusion is that using VaR at 99% level is better than ES at 97.5% level in terms of nonparametric inference efficiency when the underlying loss distribution has a heavier tail. This conflicts with the preference of using ES as it is argued that ES takes more extremes into account.

We organize the paper as follows. Section 2 presents our theoretical and numerical comparison results. Section 3 is a simulation study to confirm our findings in Section 2. Section 4 analyzes two insurance datasets. Section 5 concludes.

2 Theoretical and numerical comparisons

For a random variable X representing the loss of a financial institution or risk variable, the Value-at-Risk and expected shortfall at risk level $p \in (0,1)$ are defined as

$$VaR(p;X)=\sup\{x:P(X\le x)\le p\},\quad ES(p;X)=E\{X\,|\,X>VaR(p;X)\},$$

respectively. When we observe losses $X_1,\cdots,X_n$ from X, the nonparametric estimators for Value-at-Risk and expected shortfall at level p are

$$\widehat{VaR}(p;X)=F_n^-(p),\quad \widehat{ES}(p;X)=\frac{1}{1-p}\left(\frac{1}{n}\sum_{i=1}^n X_i I\big(X_i>\widehat{VaR}(p;X)\big)\right),$$

respectively, where

$$F_n(x)=\frac{1}{n+1}\sum_{i=1}^n I(X_i\le x),$$

F_n^- denotes the generalized inverse function of F_n, and $I(\cdot)$ is the indicator function. The study of nonparametric estimation for VaR and ES and conditional risk measures includes [1,2,5–7,15,18–20].

Suppose we observe $X_1,\cdots,X_n$ from an α-mixing sequence $\{X_t\}$ with mixing coefficient α_n and distribution function F satisfying

C(i) $\alpha_k\le C\rho^k$ for some $\rho\in(0,1)$ and $C>0$.

C(ii) F is continuous and has a continuous second derivative in a neighborhood of $F^-(p)$.

C(iii) $E|X_t|^{2+\delta}\le C$ for some $\delta>0$.

Under the above regulation conditions, it follows from [6,7] that

$$\begin{aligned}&\sqrt{n}\big\{\widehat{VaR}(p;X_1)-VaR(p;X_1)\big\}\ \xrightarrow{d}\ N\big(0,\sigma_1^2(p)\big),\\ &\sqrt{n}\big\{\widehat{ES}(p;X_1)-ES(p;X_1)\big\}\ \xrightarrow{d}\ N\big(0,\sigma_2^2(p)\big)\end{aligned}$$

as $n\to\infty$, where

$$\begin{aligned}\sigma_1^2(p)&=\frac{1}{\big(F'(F^-(p))\big)^2}\sum_{k=-\infty}^{\infty}\mathrm{Cov}\big[I(X_1\le F^-(p)),I(X_{1+k}\le F^-(p))\big],\\ \sigma_2^2(p)&=\frac{1}{(1-p)^2}\sum_{k=-\infty}^{\infty}\mathrm{Cov}\big[(X_1-F^-(p))I(X_1>F^-(p)),\\ &\qquad\qquad (X_{1+k}-F^-(p))I(X_{1+k}>F^-(p))\big]\end{aligned}$$

with Cov and $N(\mu,\sigma^2)$ denoting the covariance and the normal distribution with mean μ and variance σ^2.

Using the asymptotic results above, this paper examines the effect of the degree of freedom in t-distributions and serial dependence in an autoregressive (AR) model on the ratio of $\sigma_1(0.99)/\sigma_2(0.975)$. Although normal distribution is the limit of t-distribution with an infinite degree of freedom, we study them separately for comparison. More specifically, we look at various cases according to normal or t-distribution and independent or dependent observations. When this ratio is smaller (larger) than one, nonparametric inference for VaR at the 99% level is more efficient (inefficient) than ES at the 97.5% level. When both risk measures are allowed by regulations, one may prefer the risk measure to be estimated efficiently.

Case 1. X_i are independent and identically distributed (iid) random variables with $N(0,\sigma^2)$. Because

$$
\begin{aligned}
\sigma_1^2(p) &= \frac{p(1-p)}{\big(F'\big(F^-(p)\big)\big)^2} = \frac{p(1-p)\sigma^2}{\big(\Phi'\big(\Phi^-(p)\big)\big)^2} = 2\pi\sigma^2 p(1-p)e^{\Phi^-(p)^2}, \\
\sigma_2^2(p) &= \frac{1}{(1-p)^2} Var\big[\big(X_1 - F^-(p)\big) I\big(X_1 > F^-(p)\big)\big] \\
&= \frac{1}{(1-p)^2}\bigg(\sigma^2(1-p)\big(1+p\Phi^-(p)^2\big) - \frac{\sigma^2}{2\pi}e^{-\Phi^-(p)^2} \\
&\qquad\qquad + (1-2p)\frac{\sigma^2}{\sqrt{2\pi}}\Phi^-(p)e^{-\frac{\Phi^-(p)^2}{2}}\bigg),
\end{aligned}
$$

$\Phi^-(0.99) = 2.326348$, and $\Phi^-(0.975) = 1.959964$, we have

$$
\sigma_1^2(0.99) = 13.937061\sigma^2, \quad \sigma_2^2(0.975) = 10.235226\sigma^2,
$$

implying that $\sigma_1^2(0.99) > \sigma_2^2(0.975)$. That is, it is theoretically better to employ ES at 97.5% than VaR at 99% level when nonparametric inference efficiency is the concern. This preference for ES differs from the argument that ES takes more extremes into account.

Case 2. X_i are iid random variables with $t(\nu)$ and $\nu > 2$ (t-distribution with ν degrees of freedom). The probability density function (PDF) of $t(\nu)$ is

$$
f_\nu(t) = \frac{1}{\sqrt{\nu}B(\nu/2,1/2)}\left(1+\frac{t^2}{\nu}\right)^{-\frac{\nu+1}{2}},
$$

where $B(a,b)$ is the Beta function defined as

$$B(a,b)=\int_0^1 x^{a-1}(1-x)^{b-1}dx.$$

Note that

$$\begin{aligned}\sigma_1^2(p):=\sigma_1^2(p,\nu)&=\frac{p(1-p)}{\big(F'\big(F^-(p)\big)\big)^2}\\&=p(1-p)\nu\left[B\left(\frac{\nu}{2},\frac{1}{2}\right)\right]^2\left(1+\frac{[t_p^-(\nu)]^2}{\nu}\right)^{\nu+1}.\end{aligned}$$

It is straightforward to compute that

$$\begin{aligned}&E\big[X_1 I\big(X_1>t_p^-(\nu)\big)\big]\\&=\int_{t_p^-(\nu)}^{\infty}\frac{t}{\sqrt{\nu}B(\nu/2,1/2)}\left(1+\frac{t^2}{\nu}\right)^{-\frac{\nu+1}{2}}dt\\&=\frac{\sqrt{\nu}}{(\nu-1)B(\nu/2,1/2)}\left(1+\frac{[t_p^-(\nu)]^2}{\nu}\right)^{-\frac{\nu-1}{2}},\\&E\big[X_1^2 I\big(X_1>t_p^-(\nu)\big)\big]\\&=\int_{t_p^-(\nu)}^{\infty}\frac{t^2}{\sqrt{\nu}B(\nu/2,1/2)}\left(1+\frac{t^2}{\nu}\right)^{-\frac{\nu+1}{2}}dt\\&=\frac{\sqrt{\nu}t_p^-(\nu)}{(\nu-1)B(\nu/2,1/2)}\left(1+\frac{[t_p^-(\nu)]^2}{\nu}\right)^{-\frac{\nu-1}{2}}\\&\quad+\frac{\sqrt{\nu}}{(\nu-1)B(\nu/2,1/2)}\int_{t_p^-(\nu)}^{\infty}\left(1+\frac{t^2}{\nu}\right)^{-\frac{\nu-1}{2}}dt\\&=\frac{\sqrt{\nu}t_p^-(\nu)}{(\nu-1)B(\nu/2,1/2)}\left(1+\frac{[t_p^-(\nu)]^2}{\nu}\right)^{-\frac{\nu-1}{2}}\\&\quad+\frac{1}{2}\frac{\nu}{\nu-1}\frac{1}{B(\nu/2,1/2)}\int_0^{\frac{\nu}{\nu+[t_p^-(\nu)]^2}}x^{\frac{\nu}{2}-2}(1-x)^{-\frac{1}{2}}dx\\&=\frac{\sqrt{\nu}t_p^-(\nu)}{(\nu-1)B(\nu/2,1/2)}\left(1+\frac{[t_p^-(\nu)]^2}{\nu}\right)^{-\frac{\nu-1}{2}}\end{aligned}$$

$$+\frac{1}{2}\frac{\nu}{\nu-1}\frac{1}{B(\nu/2,1/2)}B\left(\frac{\nu}{\nu+[t_p^-(\nu)]^2};\frac{\nu}{2}-1,\frac{1}{2}\right),$$

where the third equality follows from the transform $x=(1+t^2/\nu)^{-1}$, and $B(y;a,b)$ is the incomplete Beta function defined as

$$B(y;a,b)=\int_0^y x^{a-1}(1-x)^{b-1}dx.$$

Hence, we have

$$\sigma_2^2(p):=\sigma_2^2(p,\nu)=\frac{g(p,\nu)}{(1-p)^2},$$

where

$$\begin{aligned}g(p,\nu)=&\frac{(1-2p)\sqrt{\nu}t_p^-(\nu)}{(\nu-1)B(\nu/2,1/2)}\left(1+\frac{[t_p^-(\nu)]^2}{\nu}\right)^{-\frac{\nu-1}{2}}\\&+\frac{1}{2}\frac{\nu}{\nu-1}\frac{1}{B(\nu/2,1/2)}B\left(\frac{\nu}{\nu+[t_p^-(\nu)]^2};\frac{\nu}{2}-1,\frac{1}{2}\right)\\&-\frac{1}{[B(\nu/2,1/2)]^2}\frac{\nu}{(\nu-1)^2}\left(1+\frac{[t_p^-(\nu)]^2}{\nu}\right)^{-(\nu-1)}+p(1-p)[t_p^-(\nu)]^2.\end{aligned}$$

It is challenging to compare $\sigma_1^2(0.99,\nu)$ with $\sigma_2^2(0.975,\nu)$ theoretically, although we conjecture that there exists ν_0 such that $\sigma_1^2(0.99,\nu)/\sigma_2^2(0.975,\nu) <$ or $=$ or > 1 as ν is small or equal to or larger than ν_0. To support this conjecture numerically, we plot the curves of $((\nu-2)/\nu)\sigma_1^2(0.99,\nu)$ and $((\nu-2)/\nu)\sigma_2^2(0.975,\nu)$ as a function of ν in Fig. 1. Using the factor $(\nu-2)/\nu$ is equivalent to fixing the variance of t-distributions at one. As we can see from the figure, both functions decrease with ν, but the latter decreases faster. The cross point is around $\nu_0=5.657$, which conjectures that $\sigma_1^2(0.99,\nu)<\sigma_2^2(0.975,\nu)$ when $\nu<\nu_0$, and $\sigma_1^2(0.99,\nu)>\sigma_2^2(0.975,\nu)$ when $\nu>\nu_0$. When $\nu\to\infty$, the two functions go to 13.937061 and 10.235226, respectively, consistent with Case 1.

Case 3. Stationary AR(s) model

$$X_t=\sum_{i=1}^s\phi_iX_{t-i}+\varepsilon_t,$$

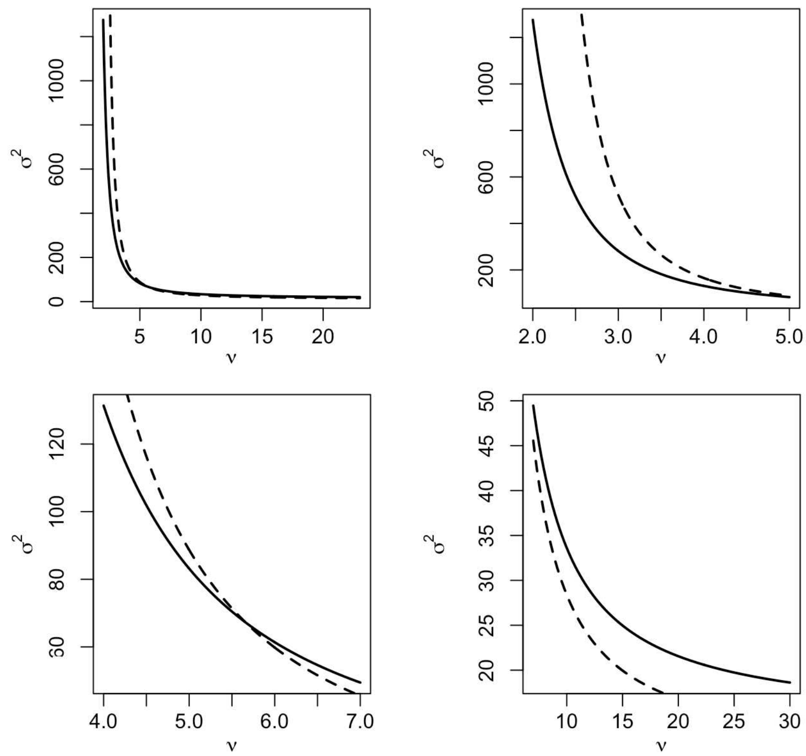

Figure 1: We plot $((\nu-2)/\nu)\sigma_1^2(0.99,\nu)$ (solid line) and $((\nu-2)/\nu)\sigma_2^2(0.975,\nu)$ (dashed line) against ν for Case 2 of Section 2.

where ε_t are iid random variables with $N(0,\sigma^2)$. Define $\gamma_k = \mathrm{Cov}(X_1, X_{1+k})$ and $\rho_k = \gamma_k/\gamma_0$. Then, the joint PDF of (X_1, X_{1+k}) is

$$f(x,y) = \frac{1}{2\pi\gamma_0\sqrt{1-\rho_k^2}} \exp\left\{-\frac{x^2+y^2-2\rho_k xy}{2\gamma_0(1-\rho_k^2)}\right\}.$$

For $k \geq 1$, using the transforms

$$u = \frac{(x-\rho_k y)}{\sqrt{(1-\rho_k^2)\gamma_0}}, \quad v = \frac{y}{\sqrt{\gamma_0}},$$

we have

$$\begin{aligned}
&E\big[I\big(X_1 < F^-(p)\big)I\big(X_{1+k} < F^-(p)\big)\big] \\
&= P\big(X_1 < F^-(p), X_{1+k} < F^-(p)\big)
\end{aligned}$$

$$
\begin{aligned}
&=\int_{-\infty}^{F^-(p)}\int_{-\infty}^{F^-(p)}\frac{1}{2\pi\gamma_0\sqrt{1-\rho_k^2}}\exp\left\{-\frac{x^2+y^2-2\rho_k xy}{2\gamma_0(1-\rho_k^2)}\right\}dxdy\\
&=\frac{1}{2\pi}\int_{-\infty}^{\frac{F^-(p)}{\sqrt{\gamma_0}}}\int_{-\infty}^{\frac{F^-(p)-\rho_k v\sqrt{\gamma_0}}{\sqrt{\gamma_0(1-\rho_k^2)}}}e^{-\frac{u^2+v^2}{2}}dudv\\
&=\frac{1}{2\pi}\int_{-\infty}^{\Phi^-(p)}\int_{-\infty}^{\frac{\Phi^-(p)-\rho_k v}{\sqrt{1-\rho_k^2}}}e^{-\frac{u^2+v^2}{2}}dudv,
\end{aligned}
$$

implying that

$$
\begin{aligned}
&\frac{1}{\big(F'\big(F^-(p)\big)\big)^2}\mathrm{Cov}\big[I\big(X_1\le F^-(p)\big),I\big(X_{1+k}\le F^-(p)\big)\big]\\
&=\gamma_0 e^{\Phi^-(p)^2}\left\{\int_{-\infty}^{\Phi^-(p)}\int_{-\infty}^{\frac{\Phi^-(p)-\rho_k v}{\sqrt{1-\rho_k^2}}}e^{-\frac{u^2+v^2}{2}}dudv-2\pi p^2\right\}.
\end{aligned}
$$

Similarly, we can show that

$$
P\big(X_1>F^-(p),X_{1+k}>F^-(p)\big)=\frac{1}{2\pi}\int_{\Phi^-(p)}^{\infty}\int_{\frac{\Phi^-(p)-\rho_k v}{\sqrt{1-\rho_k^2}}}^{\infty}e^{-\frac{u^2+v^2}{2}}dudv
$$

and

$$
\begin{aligned}
&E\big[X_1X_{1+k}I\big(X_1>F^-(p)\big)I\big(X_{1+k}>F^-(p)\big)\big]\\
&=\frac{\gamma_0\sqrt{1-\rho_k^2}}{2\pi}\int_{\Phi^-(p)}^{\infty}\int_{\frac{\Phi^-(p)-\rho_k v}{\sqrt{1-\rho_k^2}}}^{\infty}uve^{-\frac{u^2+v^2}{2}}dudv\\
&\quad+\frac{\gamma_0\rho_k}{2\pi}\int_{\Phi^-(p)}^{\infty}\int_{\frac{\Phi^-(p)-\rho_k v}{\sqrt{1-\rho_k^2}}}^{\infty}v^2e^{-\frac{u^2+v^2}{2}}dudv,
\end{aligned}
$$

implying that

$$
\begin{aligned}
&\mathrm{Cov}\big[\big(X_1-F^-(p)\big)I\big(X_1>F^-(p)\big),\big(X_{k+1}-F^-(p)\big)I\big(X_{k+1}>F^-(p)\big)\big]\\
&=E\big[X_1X_{1+k}I\big(X_1>F^-(p)\big)I\big(X_{1+k}>F^-(p)\big)\big]\\
&\quad+[F^-(p)]^2P\big(X_1>F^-(p),X_{1+k}>F^-(p)\big)\\
&\quad-2F^-(p)E\big[X_{1+k}I\big(X_1>F^-(p)\big)I\big(X_{1+k}>F^-(p)\big)\big]
\end{aligned}
$$

$$
\begin{aligned}
&-\left\{E\left[X_1 I\left(X_1>F^{-}(p)\right)\right]-(1-p)F^{-}(p)\right\}^2\\
=&\frac{\gamma_0\sqrt{1-\rho_k^2}}{2\pi}\int_{\Phi^{-}(p)}^{\infty}\int_{\frac{\Phi^{-}(p)-\rho_k v}{\sqrt{1-\rho_k^2}}}^{\infty} uve^{-\frac{u^2+v^2}{2}}dudv\\
&+\frac{\gamma_0\rho_k}{2\pi}\int_{\Phi^{-}(p)}^{\infty}\int_{\frac{\Phi^{-}(p)-\rho_k v}{\sqrt{1-\rho_k^2}}}^{\infty} v^2e^{-\frac{u^2+v^2}{2}}dudv\\
&+\frac{\gamma_0\Phi^{-}(p)^2}{2\pi}\int_{\Phi^{-}(p)}^{\infty}\int_{\frac{\Phi^{-}(p)-\rho_k v}{\sqrt{1-\rho_k^2}}}^{\infty} e^{-\frac{u^2+v^2}{2}}dudv\\
&-\frac{\gamma_0\Phi^{-}(p)}{\pi}\int_{\Phi^{-}(p)}^{\infty}\int_{\frac{\Phi^{-}(p)-\rho_k v}{\sqrt{1-\rho_k^2}}}^{\infty} ve^{-\frac{u^2+v^2}{2}}dudv\\
&-\gamma_0\left\{\frac{1}{\sqrt{2\pi}}e^{-\frac{\Phi^{-}(p)^2}{2}}-\Phi^{-}(p)(1-p)\right\}^2.
\end{aligned}
$$

Next, we focus on the comparison for the simplest case of $s=1$, i.e.

$$X_t=\phi X_{t-1}+\varepsilon_t,\quad \{\varepsilon_t\}\ iid\sim N(0,\sigma^2).$$

In this case, we have $\gamma_0=\sigma^2/(1-\phi^2)$, $\gamma_k=\phi^k\gamma_0$, and $\rho_k=\phi^k$ for $k\geq 1$. Define

$$g_1(\phi)=13.937061+2e^{\Phi^{-}(0.99)^2}\sum_{k=1}^{\infty}\left\{\int_{-\infty}^{\Phi^{-}(0.99)}\int_{-\infty}^{\frac{(\Phi^{-}(0.99)-\phi^k v)}{\sqrt{1-\phi^{2k}}}} e^{-\frac{u^2+v^2}{2}}dudv-2\pi 0.99^2\right\},$$

$$
\begin{aligned}
g_2(\phi)=10.235226+\frac{2}{0.000625}\sum_{k=1}^{\infty}\Bigg\{&\frac{\sqrt{1-\phi^{2k}}}{2\pi}\int_{\Phi^{-}(0.975)}^{\infty}\int_{\frac{\Phi^{-}(0.975)-\phi^k v}{\sqrt{1-\phi^{2k}}}}^{\infty} uve^{-\frac{u^2+v^2}{2}}dudv\\
&+\frac{\phi^k}{2\pi}\int_{\Phi^{-}(0.975)}^{\infty}\int_{\frac{\Phi^{-}(0.975)-\phi^k v}{\sqrt{1-\phi^{2k}}}}^{\infty} v^2e^{-\frac{u^2+v^2}{2}}dudv\\
&+\frac{\Phi^{-}(0.975)^2}{2\pi}\int_{\Phi^{-}(0.975)}^{\infty}\int_{\frac{\Phi^{-}(0.975)-\phi^k v}{\sqrt{1-\phi^{2k}}}}^{\infty} e^{-\frac{u^2+v^2}{2}}dudv\\
&-\frac{\Phi^{-}(0.975)}{\pi}\int_{\Phi^{-}(0.975)}^{\infty}\int_{\frac{\Phi^{-}(0.975)-\phi^k v}{\sqrt{1-\phi^{2k}}}}^{\infty} ve^{-\frac{u^2+v^2}{2}}dudv
\end{aligned}
$$

$$-\left(\frac{1}{\sqrt{2\pi}}e^{-\frac{\Phi^-(0.975)^2}{2}}-\Phi^-(0.975)\times 0.025\right)^2\Bigg\}.$$

Then

$$\sigma_1^2(0.99,\phi)=\frac{\sigma^2}{1-\phi^2}g_1(\phi),\quad \sigma_2^2(0.975,\phi)=\frac{\sigma^2}{1-\phi^2}g_2(\phi).$$

Because $g_1(\phi)$ and $g_2(\phi)$ are difficult to compute, we approximate $\sum_{k=1}^{\infty}$ in g_1 and g_2 by $\sum_{k=1}^{2000}$. That is, we study

$$h_1(\phi)=13.937061+2e^{\Phi^-(0.99)^2}\sum_{k=1}^{2000}\left\{\int_{-\infty}^{\Phi^-(0.99)}\int_{-\infty}^{\frac{\Phi^-(0.99)-\phi^k v}{\sqrt{1-\phi^{2k}}}}e^{-\frac{u^2+v^2}{2}}dudv-2\pi 0.99^2\right\},$$

$$\begin{aligned}h_2(\phi)=10.235226+\frac{2}{0.000625}\sum_{k=1}^{2000}\Bigg\{&\frac{\sqrt{1-\phi^{2k}}}{2\pi}\int_{\Phi^-(0.975)}^{\infty}\int_{\frac{\Phi^-(0.975)-\phi^k v}{\sqrt{1-\phi^{2k}}}}^{\infty}uve^{-\frac{u^2+v^2}{2}}dudv\\&+\frac{\phi^k}{2\pi}\int_{\Phi^-(0.975)}^{\infty}\int_{\frac{\Phi^-(0.975)-\phi^k v}{\sqrt{1-\phi^{2k}}}}^{\infty}v^2e^{-\frac{u^2+v^2}{2}}dudv\\&+\frac{\Phi^-(0.975)^2}{2\pi}\int_{\Phi^-(0.975)}^{\infty}\int_{\frac{\Phi^-(0.975)-\phi^k v}{\sqrt{1-\phi^{2k}}}}^{\infty}e^{-\frac{u^2+v^2}{2}}dudv\\&-\frac{\Phi^-(0.975)}{\pi}\int_{\Phi^-(0.975)}^{\infty}\int_{\frac{\Phi^-(0.975)-\phi^k v}{\sqrt{1-\phi^{2k}}}}^{\infty}ve^{-\frac{u^2+v^2}{2}}dudv\\&-\left(\frac{1}{\sqrt{2\pi}}e^{-\frac{\Phi^-(0.975)^2}{2}}-\Phi^-(0.975)\times 0.025\right)^2\Bigg\}.\end{aligned}$$

Table 1 shows that $h_2(\phi)$ is always smaller than $h_1(\phi)$, implying that $\sigma_1^2(0.99,\phi)>\sigma_2^2(0.975,\phi)$. Note that the case of $\phi=0$ becomes Case 1. Our numerical results conjecture that using ES at 97.5% level is better than VaR at 99% level regardless of the strength of serial dependence when the underlying process is stationary with normal errors.

Case 4. Stationary AR(1) model with t-distributed errors

$$X_t=\phi X_{t-1}+\varepsilon_t,\quad \{\varepsilon_t\}\ iid\sim t(\nu),\quad |\phi|<1.$$

Table 1: Values of $h_1(\phi)$ and $h_2(\phi)$ defined in Case 3 of Section 2.

ϕ	-0.9	-0.8	-0.7	-0.6	-0.5	-0.4	-0.3	-0.2	-0.1	0
$h_1(\phi)$	39.8613	22.6312	17.6158	15.5423	14.5787	14.1162	13.8994	13.819	13.8407	13.9741
$h_2(\phi)$	36.9212	19.2272	14.0085	11.849	10.8523	10.3789	10.16	10.0801	10.1022	10.2354
ϕ	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.99
$h_1(\phi)$	14.2577	14.7562	15.5692	16.8589	18.9139	22.3128	28.4138	41.3179	81.6488	826.772
$h_2(\phi)$	10.5202	11.0256	11.8584	13.1903	15.3225	18.849	25.1498	38.3603	79.1593	823.213

The equation above has a stationary solution

$$X_t=\sum_{i=-\infty}^{t}\phi^{t-i}\varepsilon_i=\sum_{i=0}^{\infty}\phi^i\varepsilon_{t-i}.$$

Define $Y_k=\varepsilon_{k+1}+\phi\varepsilon_k+\cdots+\phi^{k-1}\varepsilon_2$, then

$$X_{1+k}=Y_k+\phi^k\sum_{i=0}^{\infty}\phi^i\varepsilon_{1-i}=Y_k+\phi^kX_1.$$

Since the distribution of a linear combination of independent Student's t random variables is very difficult to obtain, we have to compute the covariances approximately.

There are two methods in the literature to approximate the distribution of weighted sums of Student's t random variables, see [21]. We follow the idea of [12] to approximate the distribution of Y_k with a single random variable c_kT_k, where T_k is a Student's t random variable. By setting Y_k and c_kT_k have the same variance and kurtosis, we have

$$\nu_k=4+\left(\frac{1}{\nu-4}\frac{(1+\phi^{2k})(1-\phi^2)}{(1-\phi^{2k})(1+\phi^2)}+\frac{2}{3}\frac{\phi^{2k}-\phi^2}{(1-\phi^{2k})(1+\phi^2)}\right)^{-1}, \tag{2.1}$$

$$c_k=\sqrt{\frac{\nu_k-2}{\nu_k}\frac{\nu}{\nu-2}\frac{1-\phi^{2k}}{1-\phi^2}}, \tag{2.2}$$

where the degree ν is required to be larger than 4. To approximate random series $X_1=\sum_{i=0}^{\infty}\phi^i\varepsilon_{1-i}$, we use $\sum_{i=0}^{n_0}\phi^i\varepsilon_{1-i}$ with a large n_0, which can be further approximated by $c_{n_0}T_{n_0}$ as above. Then $F^{-}(p)$ in the formula of asymptotic variance can be approximated by the p-th quantile of $c_{n_0}T_{n_0}$, which is $c_{n_0}t^{-}_{\nu_{n_0}}(p)$.

For $k\geq 1$, we have

$$
\begin{aligned}
&E\big[I\big(X_1\leq F^{-}(p)\big)I\big(X_{1+k}\leq F^{-}(p)\big)\big]\\
=&P\big(X_1\leq F^{-}(p),X_{1+k}\leq F^{-}(p)\big)\\
=&P\big(X_1\leq F^{-}(p),Y_k+\phi^k X_1\leq F^{-}(p)\big)\\
\approx&P\big(c_{n_0}T_{n_0}\leq c_{n_0}t^{-}_{\nu_{n_0}}(p),c_kT_k+\phi^k c_{n_0}T_{n_0}\leq c_{n_0}t^{-}_{\nu_{n_0}}(p)\big)\\
=&\int_{-\infty}^{t^{-}_{\nu_{n_0}}(p)}p_{n_0}(x)\left(\int_{-\infty}^{(t^{-}_{\nu_{n_0}}(p)-\phi^k x)c_{n_0}/c_k}p_k(t)dt\right)dx,
\end{aligned}
$$

where

$$
p_k(t)=\frac{1}{\sqrt{\nu_k}B(\nu_k/2,1/2)}\left(1+\frac{t^2}{\nu_k}\right)^{-\frac{\nu_k+1}{2}}.
$$

Hence,

$$
\begin{aligned}
\sigma_1^2(p)=&\frac{1}{\big(F'\big(F^{-}(p)\big)\big)^2}\sum_{k=-\infty}^{\infty}\mathrm{Cov}\big[I\big(X_1\leq F^{-}(p)\big),I\big(X_{1+k}\leq F^{-}(p)\big)\big]\\
=&\frac{p(1-p)}{\big(F'\big(F^{-}(p)\big)\big)^2}+\frac{2}{\big(F'\big(F^{-}(p)\big)\big)^2}\sum_{k=1}^{\infty}\mathrm{Cov}\big[I\big(X_1\leq F^{-}(p)\big),I\big(X_{1+k}\leq F^{-}(p)\big)\big]\\
\approx&\, p(1-p)c_{n_0}^2\nu_{n_0}\left[B\left(\frac{\nu_{n_0}}{2},\frac{1}{2}\right)\right]^2\left(1+\frac{[t^{-}_{\nu_{n_0}}(p)]^2}{\nu_{n_0}}\right)^{\nu_{n_0}+1}\\
&+2c_{n_0}^2\nu_{n_0}\left[B\left(\frac{\nu_{n_0}}{2},\frac{1}{2}\right)\right]^2\left(1+\frac{[t^{-}_{\nu_{n_0}}(p)]^2}{\nu_{n_0}}\right)^{\nu_{n_0}+1}\\
&\times\sum_{k=1}^{\infty}\left\{-p\int_{t^{-}_{\nu_{n_0}}(p)}^{\infty}p_{n_0}(x)\left(\int_{-\infty}^{(t^{-}_{\nu_{n_0}}(p)-\phi^k x)c_{n_0}/c_k}p_k(t)dt\right)dx\right.\\
&\qquad\left.+(1-p)\int_{-\infty}^{t^{-}_{\nu_{n_0}}(p)}p_{n_0}(x)\left(\int_{-\infty}^{(t^{-}_{\nu_{n_0}}(p)-\phi^k x)c_{n_0}/c_k}p_k(t)dt\right)dx\right\}.
\end{aligned}
$$

Similarly, we have

$$
E\big[\big(X_1-F^{-}(p)\big)I\big(X_1>F^{-}(p)\big)\big]\approx c_{n_0}\int_{t^{-}_{\nu_{n_0}}(p)}^{\infty}\big(x-t^{-}_{\nu_{n_0}}(p)\big)p_{n_0}(x)dx,
$$

$$
\begin{aligned}
&E\Big[\big(X_{1+k}-F^{-}(p)\big)I\big(X_{1+k}>F^{-}(p)\big)\Big]\\
\approx\,& c_{n_0}\int_{-\infty}^{\infty}p_{n_0}(x)\left(\int_{(t^{-}_{\nu_{n_0}}(p)-\phi^k x)c_{n_0}/c_k}^{\infty}\left(\frac{c_k}{c_{n_0}}t+\phi^k x-t^{-}_{\nu_{n_0}}(p)\right)p_k(t)dt\right)dx,\\
&E\big[\big(X_1-F^{-}(p)\big)\big(X_{1+k}-F^{-}(p)\big)I\big(X_1>F^{-}(p)\big)I\big(X_{1+k}>F^{-}(p)\big)\big]\\
\approx\,& c_{n_0}^2\int_{t^{-}_{\nu_{n_0}}(p)}^{\infty}\big(x-t^{-}_{\nu_{n_0}}(p)\big)p_{n_0}(x)\\
&\times\left(\int_{(t^{-}_{\nu_{n_0}}(p)-\phi^k x)c_{n_0}/c_k}^{\infty}\left(\frac{c_k}{c_{n_0}}t+\phi^k x-t^{-}_{\nu_{n_0}}(p)\right)p_k(t)dt\right)dx.
\end{aligned}
$$

Finally,

$$
\begin{aligned}
\sigma_2^2(p)=&\frac{1}{(1-p)^2}\sum_{k=-\infty}^{\infty}\mathrm{Cov}\big[\big(X_1-F^{-}(p)\big)I\big(X_1>F^{-}(p)\big),\\
&\qquad\big(X_{1+k}-F^{-}(p)\big)I\big(X_{1+k}>F^{-}(p)\big)\big]\\
=&\frac{1}{(1-p)^2}\Big\{E\big[\big(X_1-F^{-}(p)\big)^2 I\big(X_1>F^{-}(p)\big)\big]\\
&\qquad-\Big(E\big[\big(X_1-F^{-}(p)\big)I\big(X_1>F^{-}(p)\big)\big]\Big)^2\Big\}\\
&+\frac{2}{(1-p)^2}\sum_{k=1}^{\infty}\mathrm{Cov}\big[\big(X_1-F^{-}(p)\big)I\big(X_1>F^{-}(p)\big),\\
&\qquad\big(X_{1+k}-F^{-}(p)\big)I\big(X_{1+k}>F^{-}(p)\big)\big]\\
\approx&\frac{c_{n_0}^2}{(1-p)^2}\left(\int_{t^{-}_{\nu_{n_0}}(p)}^{\infty}\big(x-t^{-}_{\nu_{n_0}}(p)\big)^2 p_{n_0}(x)dx\right.\\
&\qquad\left.-\left\{\int_{t^{-}_{\nu_{n_0}}(p)}^{\infty}\big(x-t^{-}_{\nu_{n_0}}(p)\big)p_{n_0}(x)dx\right\}^2\right)\\
&+\frac{2c_{n_0}^2}{(1-p)^2}\sum_{k=1}^{\infty}\left\{\int_{t^{-}_{\nu_{n_0}}(p)}^{\infty}\big(x-t^{-}_{\nu_{n_0}}(p)\big)p_{n_0}(x)\right.\\
&\qquad\times\left(\int_{(t^{-}_{\nu_{n_0}}(p)-\phi^k x)c_{n_0}/c_k}^{\infty}\mu(x,t)dt\right)dx\\
&\qquad-\int_{t^{-}_{\nu_{n_0}}(p)}^{\infty}\big(x-t^{-}_{\nu_{n_0}}(p)\big)p_{n_0}(x)dx
\end{aligned}
$$

$$\times \int_{-\infty}^{\infty} p_{n_0}(x) \left(\int_{(t_{\nu_{n_0}}^{-}(p)-\phi^k x)c_{n_0}/c_k}^{\infty} \mu(x,t)dt \right) dx \Bigg\},$$

where

$$\mu(t,x) = \left(c_k t/c_{n_0} + \phi^k x - t_{\nu_{n_0}}^{-}(p)\right) p_k(t).$$

To compare $\sigma_1^2(0.99)$ with $\sigma_2^2(0.975)$, we have to compute the integrals above numerically. Since some integrals converge rather slowly, we truncate integrals' lower and upper limits to -50 and 50, respectively. We present the numerical comparison of $\sigma_1^2(0.99)$ and $\sigma_2^2(0.975)$ for the case of $\phi = 0.5$ in Table 2, where we set $n_0 = 1{,}000$ and take the sum of the first 100 items in the series of covariances. These results indicate that, as the degree of freedom ν increases from 4.5, both $\sigma_1^2(0.99)$ and $\sigma_2^2(0.975)$ decrease. Moreover, $\sigma_1^2(0.99) < \sigma_2^2(0.975)$ when $4 < \nu \le 5.3$, $\sigma_1^2(0.99) > \sigma_2^2(0.975)$ when $\nu \ge 5.4$, and the cross point falls in (5.3,5.4). Although the value of the cross point may not be accurate due to several approximations, the results above indicate that VaR at 99% level is preferred when the underlying distribution has a heavier tail.

Table 2: Values of $\sigma_1^2(0.99)$ and $\sigma_2^2(0.975)$ for $\phi = 0.5$ and Case 4 of Section 2.

ν	4.5	4.8	5	5.2	5.3	5.4
$\sigma_1^2(0.99)$	168.053	136.355	119.985	106.578	100.769	95.3964
$\sigma_2^2(0.975)$	204.858	153.397	129.058	110.175	102.308	95.2992
ν	5.5	5.6	5.8	6	7	8
$\sigma_1^2(0.99)$	90.2913	85.856	77.9886	71.2308	48.4993	36.1985
$\sigma_2^2(0.975)$	89.0311	83.4062	73.7671	65.8558	41.6872	29.9891

In summary, our theoretical and numerical analyses above suggest that VaR at 99% level is better than ES at 97.5% level when the loss distribution has a heavier tail, and the concern is statistical efficiency. This conflicts with the preference of using ES as it is argued that ES takes more extremes into account. We conclude that employing more extremes in measuring risk may lead to an inefficient nonparametric inference when the loss distribution has a heavier tail.

3 Simulation study

Section 2 compares VaR at 99% level with ES at 97.5% in terms of the asymptotic variance of the nonparametric estimation. This section will report results from a simulation study designed to evaluate the finite sample performance of the nonparametric estimation for VaR at 99% level and ES at 97.5% level. We consider various scenarios below.

- Model 1. Independent observations with normal distribution: $X_i \overset{iid}{\sim} N(0,1)$.
- Model 2. Independent observations with t-distribution: $X_i \overset{iid}{\sim} t(v)$.
- Model 3. AR(1) model with normally distributed errors: $X_t = 0.5X_{t-1} + \varepsilon_t$, $\varepsilon_t \overset{iid}{\sim} N(0,1)$.
- Model 4. AR(2) model with normally distributed errors: $X_t = 0.9X_{t-1} - 0.2X_{t-2} + \varepsilon_t, \varepsilon_t \overset{iid}{\sim} N(0,1)$.
- Model 5. MA(2) model (moving average with order two) with normally distributed errors: $X_t = \varepsilon_t + 0.65\varepsilon_{t-1} + 0.24\varepsilon_{t-2}, \varepsilon_t \overset{iid}{\sim} N(0,1)$.
- Model 6. AR(1) model with t-distributed errors: $X_t = 0.5X_{t-1} + \varepsilon_t, \varepsilon_t \overset{iid}{\sim} t(v)$.
- Model 7. MA(1) model with t-distributed errors: $X_t = \varepsilon_t + 0.5\varepsilon_{t-1}, \varepsilon_t \overset{iid}{\sim} t(v)$.
- Model 8. MA(2) model with t-distributed errors: $X_t = \varepsilon_t + 0.65\varepsilon_{t-1} + 0.24\varepsilon_{t-2}$, $\varepsilon_t \overset{iid}{\sim} t(v)$.

We generate 5,000 random samples for each model above using package stats in the R software. The sample size ranges from 125 to 2,000 to 5,000, representing six months to eight years to twenty years. We compute the true risk measures of VaR at 99% level and ES at 97.5% level for Models 1 and 2 using `qnorm`, `qt`, `ESnorm`, and `ESst` functions in the R software. To obtain the true risk measures of VaR at 99% level and ES at 97.5% level in Models 3-8, we approximate them by the averages of estimators based on 200,000 repetitions with a sample size of 100,000. The simulation results are reported in Tables 3-10, including bias, standard deviation (SD), and root of mean squared error (RMSE).

We conclude from the above tables that, for models with normal distributions, the standard deviation of VaR at 99% level is always greater than that of ES at 97.5% level, which is in line with findings in Section 2. For models with t-distributions, the difference of the standard errors of the nonparametric estimator for VaR at 99% level and ES at 97.5% level changes sign as the degree of freedom and sample size change. When the sample size is smaller, $\sigma_1(0.99) < \sigma_2(0.975)$ under different degrees of freedom. However, when the sample size increases,

Table 3: Nonparametric estimates for VaR at 99% level and ES at 97.5% for Model 1 of Section 3, where the corresponding true risk measures are 2.326 and 2.338.

N	VaR			ES		
	Bias	SD	RMSE	Bias	SD	RMSE
125	−0.139552	0.276416	0.309622	0.464425	0.324709	0.566662
250	−0.078575	0.213509	0.227489	0.192338	0.219466	0.291804
500	−0.040661	0.159185	0.164281	0.059472	0.144663	0.156397
1000	−0.020869	0.115238	0.117101	−0.008487	0.100648	0.100995
2000	−0.009727	0.084555	0.085105	−0.003447	0.073491	0.073565
5000	−0.003711	0.051956	0.052083	−0.002232	0.045138	0.045188

Table 4: Nonparametric estimates for VaR at 99% level and ES at 97.5% for Model 2 of Section 3.

N	df	VaR			ES		
		Bias	SD	RMSE	Bias	SD	RMSE
125	5	−0.237249	0.669247	0.709992	0.5766088	0.881272	1.053073
	6	−0.232140	0.590901	0.634810	0.538394	0.743998	0.918308
	7	−0.205614	0.522319	0.561284	0.546558	0.649491	0.848811
250	5	−0.1242035	0.524350	0.538808	0.254494	0.623335	0.673228
	6	−0.109883	0.449115	0.462318	0.240566	0.506711	0.560872
	7	−0.102595	0.403137	0.415948	0.235267	0.451725	0.509279
500	5	−0.077497	0.390841	0.398411	0.065558	0.421042	0.426073
	6	−0.070065	0.332313	0.339586	0.067042	0.347884	0.354251
	7	−0.060380	0.306190	0.312056	0.068385	0.304289	0.311849
1000	5	−0.035039	0.283916	0.286042	−0.017435	0.299563	0.300040
	6	−0.034714	0.237461	0.239961	−0.017136	0.240302	0.240889
	7	−0.035169	0.215366	0.218197	−0.017258	0.209577	0.210266
2000	5	−0.012556	0.204709	0.205074	−0.002282	0.213002	0.212993
	6	−0.022766	0.175191	0.176646	−0.007530	0.175790	0.175933
	7	−0.019036	0.152541	0.153709	−0.006684	0.148601	0.148736
5000	5	−0.004520	0.130911	0.130976	−0.001842	0.134486	0.134485
	6	−0.007967	0.108652	0.108933	−0.005533	0.108762	0.108892
	7	−0.006228	0.099708	0.099892	−0.001588	0.095667	0.095671

Table 5: Nonparametric estimates for VaR at 99% level and ES at 97.5% for Model 3 of Section 3, where the corresponding true risk measures are 2.686 and 2.699.

N	VaR			ES		
	Bias	SD	RMSE	Bias	SD	RMSE
125	−0.188089	0.380002	0.423970	0.496774	0.455696	0.674093
250	−0.100794	0.287636	0.304758	0.207522	0.300532	0.365194
500	−0.052604	0.215710	0.222010	0.062923	0.205286	0.214693
1000	−0.026012	0.156756	0.158884	−0.013247	0.142963	0.143561
2000	−0.013938	0.109880	0.110749	−0.006863	0.100179	0.100404
5000	−0.003876	0.071951	0.072048	−0.001710	0.064796	0.064812

Table 6: Nonparametric estimates for VaR at 99% level and ES at 97.5% for Model 4 of Section 3, where the corresponding true risk measures are 3.589 and 3.607.

N	VaR			ES		
	Bias	SD	RMSE	Bias	SD	RMSE
125	−0.300552	0.610722	0.680616	0.573992	0.741722	0.937821
250	−0.167875	0.466000	0.495272	0.229342	0.493971	0.544567
500	−0.084923	0.353010	0.363047	0.059402	0.345601	0.350635
1000	−0.042575	0.251710	0.255260	−0.029489	0.236234	0.238044
2000	−0.019674	0.177976	0.179042	−0.012105	0.168573	0.168990
5000	−0.004442	0.114627	0.114702	−0.002853	0.108621	0.108647

Table 7: Nonparametric estimates for VaR at 99% level and ES at 97.5% for Model 5 of Section 3, where the corresponding true risk measures are 2.830 and 2.844.

N	VaR			ES		
	Bias	SD	RMSE	Bias	SD	RMSE
125	−0.185919	0.402933	0.443721	0.537040	0.481291	0.721115
250	−0.104865	0.308878	0.326164	0.219996	0.321131	0.389234
500	−0.055459	0.222439	0.229227	0.065244	0.211636	0.221445
1000	−0.027949	0.162173	0.164548	−0.015750	0.149494	0.150306
2000	−0.013661	0.113985	0.114789	−0.006886	0.103537	0.103756
5000	−0.004400	0.073039	0.073164	−0.002088	0.066252	0.066279

Table 8: Nonparametric estimates for VaR at 99% level and ES at 97.5% for Model 6 of Section 3, where the corresponding true risk measures of VaR are 2.912 (df=5), 2.874 (df=6), 2.846 (df=7) and the corresponding true risk measures of ES are 3.025 (df=5), 2.959 (df=6), 2.913 (df=7).

N	df	VaR			ES		
		Bias	SD	RMSE	Bias	SD	RMSE
125	5	−0.228424	0.638608	0.678172	0.458566	0.823759	0.942722
	6	−0.216832	0.585460	0.624269	0.474249	0.738349	0.877476
	7	−0.221829	0.541025	0.584686	0.479191	0.676520	0.828982
250	5	−0.122225	0.482702	0.497889	0.190024	0.584571	0.614625
	6	−0.118548	0.426595	0.442720	0.193092	0.496332	0.532523
	7	−0.111046	0.402305	0.417311	0.206490	0.458142	0.502484
500	5	−0.073441	0.358464	0.365875	0.046205	0.400397	0.403014
	6	−0.071989	0.319986	0.327952	0.048871	0.342263	0.345700
	7	−0.066271	0.298616	0.305852	0.051722	0.312041	0.316268
1000	5	−0.035051	0.250517	0.252933	−0.021506	0.276190	0.276999
	6	−0.034133	0.232010	0.234485	−0.023574	0.240107	0.241238
	7	−0.030519	0.217403	0.219514	−0.015984	0.218812	0.219373
2000	5	−0.020921	0.184896	0.186058	−0.011881	0.201663	0.201992
	6	−0.019362	0.164056	0.165178	−0.012998	0.169946	0.170426
	7	−0.020564	0.156072	0.157405	−0.013357	0.155064	0.155622
5000	5	−0.007811	0.115766	0.116018	−0.004477	0.126095	0.126162
	6	−0.008566	0.104952	0.105290	−0.006908	0.106802	0.107014
	7	−0.006934	0.097940	0.098175	−0.003105	0.097173	0.097213

$\sigma_1(0.99) > \sigma_2(0.975)$ for $t(7)$, supporting our conjecture that VaR at 99% level is more efficient than ES at 97.5% level for heavier tails.

4 Empirical study

This section compares VaR at 99% level with ES at 97.5% for two insurance datasets, in which asymptotic variances are estimated using the bootstrap method.

Table 9: Nonparametric estimates for VaR at 99% level and ES at 97.5% for Model 7 of Section 3, where the corresponding true risk measures of VaR are 2.844 (df=5), 2.804 (df=6), 2.774 (df=7) and the corresponding true risk measures of ES are 2.959 (df=5), 2.890 (df=6), 2.843 (df=7).

N	df	VaR			ES		
		Bias	SD	RMSE	Bias	SD	RMSE
125	5	−0.208311	0.610577	0.645076	0.456630	0.782493	0.905916
	6	−0.193162	0.549093	0.582027	0.490074	0.690228	0.846458
	7	−0.198238	0.495273	0.533427	0.481318	0.617209	0.782648
250	5	−0.116698	0.450089	0.464928	0.189179	0.540787	0.572871
	6	−0.108587	0.402984	0.417318	0.200123	0.469609	0.510429
	7	−0.102911	0.374642	0.388483	0.207938	0.427419	0.475278
500	5	−0.067903	0.339257	0.345952	0.049885	0.381576	0.384786
	6	−0.062915	0.300237	0.306729	0.056284	0.319666	0.324552
	7	−0.059551	0.278040	0.284319	0.059328	0.287280	0.293314
1000	5	−0.032620	0.245560	0.247693	−0.017177	0.267102	0.267628
	6	−0.025250	0.219425	0.220851	−0.007900	0.227431	0.227545
	7	−0.030569	0.204917	0.207164	−0.013997	0.204881	0.205338
2000	5	−0.014379	0.175032	0.175604	−0.006470	0.192913	0.193003
	6	−0.015883	0.156326	0.157115	−0.011777	0.161254	0.161668
	7	−0.008719	0.146591	0.146835	−0.003618	0.145882	0.145913
5000	5	−0.006191	0.110273	0.110435	−0.004504	0.119546	0.119618
	6	−0.007714	0.099971	0.100258	−0.002680	0.102189	0.102213
	7	−0.004051	0.092636	0.092715	−0.002109	0.092049	0.092064

4.1 Data analysis on Danish fire losses

The first dataset is Danish fire losses, including 2,167 Danish fire loss records from January 1980 through December. The mean and standard deviation are 3.385088 and 8.507452, respectively. The minimum loss is 1, and the maximum loss is 263.2504.

After standardizing the losses (i.e. minus mean and divided by standard deviation), we compute the nonparametric estimators for VaR at 99% level and ES at 97.5% level. Then we use the bootstrap method to get the standard errors of

Table 10: Nonparametric estimates for VaR at 99% level and ES at 97.5% for Model 8 of Section 3, where the corresponding true risk measures of VaR are 3.063 (df=5), 3.021 (df=6), 2.991 (df=7) and the corresponding true risk measures of ES are 3.175 (df=5), 3.105 (df=6), 3.057 (df=7).

N	df	VaR			ES		
		Bias	SD	RMSE	Bias	SD	RMSE
125	5	−0.197257	0.725665	0.751927	0.506359	0.884032	1.018703
	6	−0.21698	0.624433	0.660998	0.491469	0.766491	0.910457
	7	−0.209216	0.583419	0.619743	0.510358	0.704984	0.870269
250	5	−0.108069	0.518758	0.529844	0.208089	0.613358	0.647637
	6	−0.100246	0.461206	0.471930	0.228431	0.536682	0.583224
	7	−0.119081	0.421002	0.437479	0.204927	0.472534	0.515014
500	5	−0.067563	0.372968	0.379001	0.051132	0.412241	0.415359
	6	−0.066322	0.342134	0.348469	0.060447	0.366892	0.371802
	7	−0.069283	0.321774	0.329117	0.0559418	0.338578	0.343135
1000	5	−0.033550	0.274321	0.276338	−0.022785	0.302044	0.302872
	6	−0.033444	0.247684	0.249908	−0.020491	0.257882	0.258670
	7	−0.031736	0.226929	0.229115	−0.017844	0.228250	0.228924
2000	5	−0.016413	0.194878	0.195548	−0.006453	0.215618	0.215693
	6	−0.013529	0.174801	0.175306	−0.007450	0.182360	0.182494
	7	−0.015090	0.164075	0.164751	−0.009030	0.163444	0.163677
5000	5	−0.002412	0.121419	0.121431	−0.001871	0.132954	0.132954
	6	−0.006953	0.109131	0.109342	−0.005452	0.112882	0.113003
	7	−0.005499	0.104008	0.104143	−0.003330	0.103391	0.103434

these two estimators. More specifically, after drawing a random sample from the original data with the same sample size, we compute the nonparametric estimators for VaR at 99% level using the bootstrap sample. We repeat the procedure 500,000 times to get 500,000 bootstrap estimators for VaR and compute the sample standard deviation of these bootstrap estimators to get the standard error of the nonparametric estimator for VaR at 99% level. Similarly, we compute the standard error of the nonparametric estimator for ES at 97.5% level. We also compute the skewness and kurtosis of the data.

The VaR at 99% level and ES at 97.5% level for the standardized losses are 2.663246 and 3.829128, respectively. The bootstrap method with 500,000 repetitions estimates the standard errors as 0.2956586 for VaR at 99% level and 0.7136469 for ES at 97.5%, implying that the nonparametric estimator of VaR at 99% level is more efficient than that of ES at 97.5% level for this dataset. The kurtosis of these losses is 482.198, suggesting the losses have a heavier tail as found in the literature. So, the conclusion of preferring VaR at 99% level to ES at 97.5% level is in line with our analyses in Sections 2 and 3 when the efficiency of nonparametric inference is concerned.

4.2 Data analysis on unemployment insurance initial claims

The second data set is the unemployment insurance initial claims statewide from 1971 to April 2021, including 604 monthly counts of initial claims for regular unemployment insurance benefits. Initial claims include new claims as well as subsequent additional claims filed. The mean and standard deviation are 114,121.3 and 65,678.67, respectively. The minimum is 49,263, and the maximum is 1,168 and 446.

As before, we find that VaR at 99% level and ES at 97.5% level are 2.745115 and 3.994838 for the standardized data. The bootstrap method estimates the standard errors as 0.9180034 for Var at 99% level and 1.136914 for ES at 97.5% level. The kurtosis of data is 119.7269, suggesting heavier tails. Hence, when the concern is the nonparametric inference efficiency, one should prefer the use of VaR at 99% level to ES at 97.5% level according to our theoretical and numerical analyses in Sections 2 and 3, respectively.

5 Conclusions

It has been a hot debate comparing these two popular risk measures of Value-at-Risk and expected shortfall. Because VaR at 99% level and ES at 97.5% level are very close when the underlying distribution is the standard normal distribution, this paper compares VaR at 99% level with ES at 97.5% level in terms of nonparametric inference efficiency. We theoretically and numerically examine the effect of heavy tails and serial dependence on the comparison of estimating these two risk measures nonparametrically. We find that VaR at 99% is better than ES at 97.5% when the loss distribution has a heavier tail, which conflicts with the preference of ES in the literature as it is argued that ES takes more extremes into account. A simulation study supports our theoretical and numerical findings. Applications to two insurance datasets align with our conjecture that VaR at 99% is better than ES at 97.5% in statistical efficiency for losses with a heavier tail. Therefore, we conclude that employing more extremes in measuring risk may lead to an inefficient nonparametric inference when the loss distribution has a heavier tail.

References

[1] R. Alemany, C. Bolancé, and M. Guillén, *Nonparametric estimation of Value-at-Risk*, XARXA de Referència en Economia Aplicada XREAP 2012–19, (2012).

[2] R. Alemany, C. Bolancé, and M. Guillén, *A nonparametric approach to calculating value-at-risk*, Insur. Math. Econ. 52(2) (2013), 255–262.

[3] P. Artzner, F. Delbaen, J. M. Eber, and D. Heath, *Thinking coherently*, Risk 10(11) (1999), 68–71.

[4] R. W. Barnard, K. Pearce, and A. A. Trindade, *When is tail mean estimation more efficient than tail median? Answers and implications for quantitative risk management*, Ann. Oper. Res. 262 (2018), 47–65.

[5] Z. Cai and X. Wang, *Nonparametric estimation of conditional VaR and expected shortfall*, J. Econometrics 147 (2008), 120130.

[6] S. X. Chen, *Nonparametric estimation of expected shortfall*, J. Financial Econ. 6 (2008), 87–107.

[7] S. X. Chen and C. Y. Tang, *Nonparametric inference of value-at-risk for dependent financial returns*, J. Financial Econ. 3 (2005), 227–255.

[8] R. Cont, R. Deguest, and G. Scandolo, *Robustness and sensitivity analysis of risk measurement procedures*, Quant. Finance 20 (2010), 593–606.

[9] J. Danielsson and C. Zhou, *Why risk is so hard to measure?* De Nederlandsche Bank Working Paper 494 (2016).

[10] P. Embrechts, G. Puccetti, L. Rüschendorf, and R. Wang, *An academic response to Basel 3.5*, Risks 2 (2014), 25–48.

[11] E. S. Emmer, M. Kratz, and D. Tasche, *What is the best risk measure in practice? A comparison of standard measures*, J. Risk 18 (2015), 31–60.

[12] W. R. Fairweather, *A method of obtaining an exact confidence interval for the common mean of several normal populations*, Applied Statistics 21(3) (1972), 229–233.

[13] T. Fissler and J. F. Ziegel, *Higher order elicitability and Osband's principle*, Ann. Statist. 44 (2016), 1680–1707.

[14] T. Gneiting, *Making and evaluating point forecasts*, J. Amer. Statist. Assoc. 106 (2011), 746–762.

[15] S. O. Jeong and K. H. Kang, *Nonparametric estimation of value-at-risk*, J. Appl. Stat. 36 (2008), 1225–1238.

[16] S. Kou, X. Peng, and C. C. Heyde, *External risk measures and Basel Accords*, Math. Oper. Res. 38(3) (2013), 393–616.

[17] V. Krätschmer, A. Schied, and H. Zähle, *Comparative and qualitative robustness for law-invariant risk measures*, Finance Stoch. 18 (2014), 271–295.

[18] C. Martins-Filho and F. Yao, *Estimation of Value-at-Risk and expected shortfall based on nonlinear models of return dynamics and extreme value theory*, Stud. Nonlinear Dyn. Econ. 10(2) (2006), 1–43.

[19] C. Martins-Filho, F. Yao, and M. Torero, *Nonparametric estimation of conditional value-at-risk and expected shortfall based on extreme value theory*, Econometric Theory 34 (2018), 23–67.

[20] O. Scaillet, *Nonparametric estimation and sensitivity analysis of expected shortfall*, Math. Finance 14 (2004), 115–129.

[21] V. Witkovský, *Matlab algorithm TDIST: The distribution of a linear combination of Student's t random variables*, in: COMPSTAT 2004 Symposium, Physica-Verlag/Springer, 2004.