PELVE: Probability Equivalent Level of VaR and ES

Hengxin Li*   Ruodu Wang†

November 6, 2021

Abstract

In the recent Fundamental Review of the Trading Book (FRTB), the Basel Committee on Banking Supervision proposed the shift from the 99% Value-at-Risk (VaR) to the 97.5% Expected Shortfall (ES) for internal models in market risk assessment. Inspired by the above transition, we introduce a new distributional index, the probability equivalence level of VaR and ES (PELVE), which identifies the balancing point for the equivalence between VaR and ES. PELVE enjoys many desirable theoretical properties and it distinguishes empirically heavy-tailed distributions from light-tailed ones via a threshold of 2.72. Convergence properties and asymptotic normality of the empirical PELVE estimators are established. Applying PELVE to financial asset and portfolio data leads to interesting observations that are not captured by VaR or ES alone. We find that, in general, the transition from VaR to ES in the FRTB yields an increase in risk capital for single-asset portfolios, but for well-diversified portfolios, the capital requirement remains almost unchanged. This leads to both a theoretical justification and an empirical evidence for the conclusion that the use of ES rewards portfolio diversification more than the use of VaR.

Key-words: Value-at-Risk; Expected Shortfall; regulatory capital; heavy tails; portfolio diversification

*Haas School of Business, University of California Berkeley, USA. ✉ hanson.li@berkeley.edu

†Department of Statistics and Actuarial Science, University of Waterloo, Canada. ✉ wang@uwaterloo.ca

1 Introduction

The *Value-at-Risk* (VaR) and the *Expected Shortfall* (ES, also known as CVaR, TVaR and AVaR)[1] are the most popular risk measures used in banking and insurance, and they are widely used for regulatory capital calculation, decision making, performance analysis, and risk management (see Section 2 for precise definitions). In particular, both VaR and ES are currently employed in the banking regulation framework of Basel III/IV and the insurance regulation frameworks of Solvency II and the Swiss Solvency Test. In the above context, the main interpretation of the values of these risk measures is the capital requirement for a potential loss faced by a financial institution.

In the recent Fundamental Review of the Trading Book (FRTB), the Basel Committee on Banking Supervision confirmed to replace a VaR with an ES as the standard risk measure for market risk; see BCBS (2019). Nevertheless, the presence of VaR remains in Basel III/IV for backtesting risk models, and it is still the dominating risk measure in Solvency II. Reflecting on this matter, there has been extensive research in the literature on the comparative advantages of VaR, ES and alternative risk measures; see Embrechts et al. (2018), Chen et al. (2019), Patton et al. (2019) and the references therein. As a general consensus, VaR and ES both have advantages from different perspectives, such as estimation, forecasts and backtests, optimization, economic implications, and numerical calculation. Moreover, the two risk measures will likely co-exist in risk management practice for a long period of time.

In this paper, we look at the issue of VaR and ES from a novel angle. When a VaR is to be replaced by an ES, a proper probability level needs to be specified. Quoting BCBS (2013), Page 22, discussing the potential transition:

"*... using an ES model, the Committee believes that moving to a confidence level of 97.5% (relative to the 99th percentile confidence level for the current VaR measure) is appropriate.*"

The current FRTB specifies that banks should use $\mathrm{ES}_{0.975}$ to replace $\mathrm{VaR}_{0.99}$, regardless of the asset class that they invest in. If $\mathrm{ES}_{0.975} \approx \mathrm{VaR}_{0.99}$, then the bank's portfolio capital requirement is almost unchanged in this VaR-ES transition, whereas, if $\mathrm{ES}_{0.975} > \mathrm{VaR}_{0.99}$, then the bank's portfolio capital charge is increased in the transition. One of the main reasons

[1]We use the name ES to be consistent with the terminology used by the Basel Committee on Banking Supervision, which reflects our main motivation.

for the Basel Committee to replace VaR by ES is that ES "better captures tail risk" as explained in BCBS (2016). The implementation of ES, which facilitates a better assessment of tail risk,[2] will lead to changes in the overall capital requirement of banks. On the other hand, it is undesirable and impractical if the regulator increases capital charges by too much; for instance, this would have been the case if $\mathrm{VaR}_{0.99}$ was to be replaced by $\mathrm{ES}_{0.99}$, which will be clearly unacceptable for the banks. Therefore, a balance needs to be reached, and a natural question arises from this transition:

For various portfolios, asset classes and risk models, what is an equivalent probability level to choose when an ES is to replace a VaR?

To answer this question, we introduce the *Probability Equivalent Level of VaR-ES (PELVE)*, which is the ratio of the ES confidence level to that of VaR with an equivalent risk value. More precisely, for a chosen level ε close to 0 and a loss random variable X, the PELVE is a constant multiplier c such that $\mathrm{ES}_{1-c\varepsilon}(X) = \mathrm{VaR}_{1-\varepsilon}(X)$; a formal definition is given in Section 2. The value of PELVE addresses the issue of the regulatory capital change if one replaces a VaR by an ES. In particular, for $\varepsilon = 0.01$, if PELVE is larger than 2.5, then $\mathrm{ES}_{0.975} > \mathrm{ES}_{1-0.01c} = \mathrm{VaR}_{0.99}$, that is, replacing $\mathrm{VaR}_{0.99}$ by $\mathrm{ES}_{0.975}$ will lead to an increase of the risk value. On the other hand, if PELVE is smaller than 2.5, then replacing $\mathrm{VaR}_{0.99}$ by $\mathrm{ES}_{0.975}$ will lead to a decrease of the risk value. We will see in this paper that for commonly used parametric distributions in risk management and real asset return data, their PELVE values are typically larger than 2.5. This partially explains (among other changes) the survey finding of BCBS (2015, p.3), where two thirds of the banks reported an increase of capital charges under the proposed new market risk framework. Moreover, a larger value of PELVE indicates a more drastic increase of capital during the VaR-ES transition.

Motivated by the above practical question, we formally analyze properties of PELVE as a statistical quantity of loss distributions. By simply looking at PELVE for classic distribution families, we observe that PELVE can be seen as a measure of distributional variability because a higher value is reported for a loss distribution with a heavier tail. We discover several desirable properties of PELVE. First, PELVE is invariant under increasing[3] linear transformations, a convenient feature in portfolio risk analysis, which is often scale-free. Second, PELVE

[2]See Embrechts et al. (2018) for an explanation of ES better capturing tail risk in the context of risk sharing and optimization.

[3]In this paper, terms "increasing" and "decreasing" are in the non-strict sense.

generally increases (decreases) if an increasing convex (concave) transform is applied to the loss random variable. In particular, this implies that purchasing a put or call option on an asset reduces the PELVE of a long-only portfolio, a put option more so, consistent with the general intuition that a put option (or an insurance) protects the investor. Third, PELVE is both quasi-convex and quasi-concave for comonotonic risks, leading to advantages in both risk comparison and optimization. These properties lead to a profound implication: a well diversified (more Gaussian-like) portfolio has a smaller PELVE than individual assets, and the use of ES rewards portfolio diversification more than the use of VaR, a hidden feature of the aforementioned regulatory transition. As far as we know, our theory is the first in the literature to explain the above advantage of ES in portfolio diversification. This advantage cannot be explained by Extreme Value Theory (EVT)[4] or the coherence of ES.[5] See Section 3 for a theoretical treatment leading to the above claims.

Financial data typically exhibit heavy tails (see e.g., Jansen and De Vries (1991) and Cont (2001)). We analyze PELVE within the class of regularly varying distributions, which is an important class of heavy-tailed distributions. In particular, we found that the value of $e \approx 2.72$ can be seen as a threshold for PELVE separating (empirically) heavy- and light-tailed distributions, and hence the Basel proposal of using a multiplier of 2.5 will lead to an increase of capital requirement for heavy-tailed loss distributions. See Section 4 for an analysis of PELVE for commonly used distributions in risk management, which gives practical interpretation of PELVE values obtained from various models and data sets. For non-parametric estimation of PELVE, we show that in the iid and the α-mixing cases, the empirical PELVE estimator and its smoothed version are both consistent and have asymptotic normality. The asymptotic variance is also obtained explicitly; see Section 5 for the above results.

We conduct a few empirical studies, including PELVE curves for different S&P 500 sectors, and an analysis of PELVE for well-diversified portfolios in Section 6. We find that the $1/N$ portfolio exhibits a much smaller PELVE value on average ($\approx e$) compared to individual stocks (≈ 3), thus showing a novel advantage of the $1/N$ portfolio. According to the com-

[4]Diversification does not reduce the tail index in EVT, even for a portfolio of iid risks. See Section 4.2 for the definition and discussions.

[5]Coherence of ES only justifies that ES does not penalize diversification. However, VaR has diversification benefit for common portfolio data, and coherence of ES does not address which of ES and VaR has a larger effect of diversification; note that the mean is also a coherent risk measure but it does not have any diversification benefit. See also our discussions in Section 3.3.

mon interpretation, a proper diversification would reduce both risk measures VaR and ES simultaneously. The above result further suggests that the reduction in ES is more significant empirically, complementing the theoretical results in Section 3. Proofs of all technical results, as well as several additional examples, simulation results, and figures, are relegated to the Appendices.

From the main results of this paper, PELVE enjoys several practical advantages. For instance, it is well defined for both bounded and unbounded distributions, its values are easily interpretable, and it directly relates to the FRTB transition between VaR and ES. The definition of PELVE involves minimal model assumptions (essentially, finite first moment),[6] in contrast to other measures of variability (which typically require higher moments) and measures of tail heaviness (which typically require asymptotics of the distribution function). Furthermore, its estimation procedure is quite standard when working with data. From the empirical studies, we can see that PELVE provides a powerful complement to classic risk measures such as VaR and ES, and it captures features in the financial data which ES or VaR alone does not detect.

The work is related to various studies on VaR and ES and their relationship. For recent developments on these risk measures within the FRTB, we refer to Embrechts et al. (2018), Li and Xing (2019), Wang and Zitikis (2021) and the references therein. Heavy-tailed distributions are a popular topic in EVT; a classic reference to EVT is de Haan and Ferreira (2006) and its applications in finance and insurance are found in Embrechts et al. (1997). The derivation of the asymptotic normality of PELVE estimators is based on recent results in Asimit et al. (2019), where a capital allocation method is built on the relationship between VaR and ES. For a comparison of VaR and ES estimators and issues related to their backtests, see the recent papers Daníelsson and Zhou (2016), Fissler and Ziegel (2016) and Du and Escanciano (2017). We also refer to the book McNeil et al. (2015) for a comprehensive treatment on VaR and ES in quantitative risk management.

[6]Although there are debates on whether short-term (daily or more frequent) financial returns have finite variance (see e.g., Grabchak and Samorodnitsky (2010)), it is commonly agreed that they have finite mean, and thus PELVE is applicable.

2 Definition and basic properties of PELVE

2.1 VaR, ES and PELVE

Let L^1 be the set of random variables with finite mean in an atomless probability space $(\Omega, \mathcal{F}, \mathbb{P})$, which will be the set of random losses that we are interested in. Throughout, for a random variable X, denote by $F_X(x) = \mathbb{P}(X \leqslant x)$, $x \in \mathbb{R}$, its distribution function, and $F_X^{-1}(p) = \inf\{x \in \mathbb{R} : F_X(x) \geqslant p\}$, $p \in [0,1]$, its quantile function.

Below we define VaR and ES, the two most popular risk measures. The VaR at confidence level (or probability level) $p \in (0,1)$ is defined as

$$\mathrm{VaR}_p(X) = \inf\{x \in \mathbb{R} : \mathbb{P}(X \leqslant x) \geqslant p\} = F_X^{-1}(p), \qquad X \in L^1,$$

and the ES at confidence level $p \in [0,1)$ is defined as

$$\mathrm{ES}_p(X) = \frac{1}{1-p}\int_p^1 \mathrm{VaR}_q(X)\mathrm{d}q, \qquad X \in L^1.$$

Note that ES_0 corresponds to the mean of a random variable.

With the definitions of VaR and ES above, we introduce our key concept in this paper, the *Probability Equivalent Level of VaR-ES (PELVE)*. For a level $\varepsilon \in (0,1)$, the PELVE of a random variable X, denoted by $\Pi_\varepsilon(X)$, is a constant multiplier $c \in [1, 1/\varepsilon]$ such that ES at level $1 - c\varepsilon$ and VaR at level $1-\varepsilon$ become equal, that is,

$$\mathrm{ES}_{1-c\varepsilon}(X) = \mathrm{VaR}_{1-\varepsilon}(X). \tag{1}$$

The typical value of ε of interest is close to 0, such as $\varepsilon = 0.01$, which corresponds to $\mathrm{VaR}_{0.99}$ in Basel II.

Later we will see that (1) usually admits a unique solution $c \in [1, 1/\varepsilon]$, but this requires some weak conditions. To formulate PELVE rigorously for all random variables in L^1, we formally define $\Pi_\varepsilon : L^1 \to \mathbb{R}$ as

$$\Pi_\varepsilon(X) = \inf\left\{c \in [1, 1/\varepsilon] : \mathrm{ES}_{1-c\varepsilon}(X) \leqslant \mathrm{VaR}_{1-\varepsilon}(X)\right\}, \qquad X \in L^1. \tag{2}$$

Here, we use the convention $\inf(\varnothing) = \infty$, which may happen if ε is large. We will sometimes say "the PELVE of a distribution", which of course means the PELVE of a random variable with that distribution. Note that (2) is defined in terms of an infimum instead of a unique point. That is precisely because in some rare situations, there are possibly multiple values, or even no value of $c \in [1, 1/\varepsilon]$ such that (1) holds. In Section 2.2 below we will investigate the issue of existence and uniqueness of c for (1). Before that, to facilitate a better understanding of numeric PELVE values, we look at a few typical examples of commonly used distributions in Table 1. Formulas for these PELVE values as well as more numerical illustrations are presented in Section 4.

ε	Dirac	U	N	Exp	LN(σ^2)			t(ν)			Pareto(α)		
					0.04	0.25	1	2	10	30	2	4	10
0.100	1.00	2.00	2.46	2.72	2.56	2.76	3.23	3.60	2.58	2.49	4.00	3.16	2.87
0.050			2.51		2.61	2.79	3.19	3.80	2.65	2.55			
0.010			2.58		2.66	2.81	3.13	3.96	2.74	2.63			
0.005			2.59		2.67	2.81	3.10	3.98	2.77	2.65			

Table 1: PELVE for some commonly used distributions and $\varepsilon = 0.1, 0.05, 0.01, 0.005$

From Table 1, we see that the PELVE values are relatively stable across different ε for the same distribution; in particular, they are constant in ε for degenerate (Dirac), uniform (U), exponential (Exp) and Pareto distributions. On the other hand, the PELVE values vary significantly across distributions with different levels of tail heaviness (see Section 4.2 for a rigorous treatment of tail heaviness). PELVE generally reports a higher value if the tail of the distribution is heavier, and it reports a smaller value if the distribution is lighter-tailed. The smallest possible value of $\Pi_\varepsilon(X) = 1$ is attained by the degenerate distribution, and the largest value in Table 1 is attained by a Pareto distribution with shape parameter 2, whose variance is infinite. As such, PELVE can be seen as an index of variability or tail-heaviness of the loss distribution. We also note that PELVE is close to 2.5 for normal (N) distributions, which suggests $\mathrm{VaR}_{0.99} \approx \mathrm{ES}_{0.975}$ (the latter is slightly larger) for normal distributions in the context of the FRTB.

2.2 Existence and uniqueness

Below, we analyze issues on existence and uniqueness of $c \in [1, 1/\varepsilon]$ such that (1) holds. Fortunately, there are very simple conditions to guarantee both existence and uniqueness, and they are satisfied for almost all practically relevant choices of X and ε.

Proposition 1 (Existence of PELVE). *For $X \in L^1$ and $\varepsilon \in (0,1)$, the following statements are equivalent.*

(i) There exists $c \in [1, 1/\varepsilon]$ such that (1) holds.

(ii) $\Pi_\varepsilon(X) \in [1, 1/\varepsilon]$ and (1) holds for $c = \Pi_\varepsilon(X)$.

(iii) $\mathbb{E}[X] \leqslant \mathrm{VaR}_{1-\varepsilon}(X)$.

(iv) $\Pi_\varepsilon(X) < \infty$.

By Proposition 1, the PELVE value $\Pi_\varepsilon(X)$ is in $[1, 1/\varepsilon]$ if and only if $\mathbb{E}[X] \leqslant \mathrm{VaR}_{1-\varepsilon}(X)$. The existence is guaranteed because $\mathbb{E}[X] = \mathrm{ES}_0(X) \leqslant \mathrm{VaR}_{1-\varepsilon}(X) \leqslant \mathrm{ES}_{1-\varepsilon}(X)$ and $p \mapsto \mathrm{ES}_p(X)$ is continuous, implying that there exists $c \in [1, 1/\varepsilon]$ such that $\mathrm{ES}_{1-c\varepsilon}(X) = \mathrm{VaR}_{1-\varepsilon}(X)$. Note that ε is typically very small, and hence this condition usually holds, especially for portfolio losses, which are the main concern in the FRTB. Concerning the uniqueness of $c \in [1, 1/\varepsilon]$ for (1), we have the following condition.

Proposition 2 (Uniqueness of PELVE). *Suppose $\mathbb{E}[X] \leqslant \mathrm{VaR}_{1-\varepsilon}(X)$ and $\mathrm{VaR}_p(X)$ is not a constant for $p \in [1-\varepsilon, 1)$. Then, there exists a unique $c \in [1, 1/\varepsilon]$ such that (1) holds.*

To utilize the results in Propositions 1-2, we state the following simple assumption, which will be used repeatedly in the rest of the paper.

Assumption 1. $\mathbb{E}[X] < \mathrm{VaR}_{1-\varepsilon}(X)$ and $\mathrm{VaR}_p(X)$ is not a constant for $p \in [1-\varepsilon, 1)$.

Note that Assumption 1 is satisfied by most meaningful models for portfolio losses used in practice, noting that ε is typically very small. In particular, if X is continuously distributed, then $\mathrm{VaR}_p(X)$ is strictly increasing in p. Under Assumption 1, $\Pi_\varepsilon(X)$ is the unique solution of $c \in (1, 1/\varepsilon)$ to the equation (1). This fact will be convenient for the discussions in the next few sections.

Remark 1. For the asymptotic results in Section 5, we will need to require $\mathbb{E}[X] < \mathrm{VaR}_{1-\varepsilon}(X)$ instead of $\mathbb{E}[X] \leqslant \mathrm{VaR}_{1-\varepsilon}(X)$ (as in Propositions 1 and 2) to avoid hitting the boundary case of $c = 1/\varepsilon$ in (2). For this purpose, we use the current formulation of Assumption 1.

Remark 2. An alternative way of formulating an object similar to PELVE is to define, for $\delta \in (0,1)$ which is the ES confidence level of interest,

$$\Pi'_\delta(X) = \inf\{d \geqslant 1 : \mathrm{ES}_{1-\delta}(X) \leqslant \mathrm{VaR}_{1-\delta/d}(X)\}, \quad X \in L^1. \tag{3}$$

We can compare the formulation (3) with the formulation (2). The advantage of using (3) is that we do not need to assume that δ is small; indeed, $\Pi'_\delta(X)$ gives a finite value for all $\delta \in (0,1)$. The disadvantage, on the other hand, is that the equation

$$\mathrm{ES}_{1-\delta}(X) = \mathrm{VaR}_{1-\delta/d}(X) \tag{4}$$

is not always satisfied for $d = \Pi'_\delta(X)$ or any $d > 0$,[7] whereas $\mathrm{ES}_{1-c\varepsilon}(X) = \mathrm{VaR}_{1-\varepsilon}(X)$ always holds for $c = \Pi_\varepsilon(X)$ if $\Pi_\varepsilon(X) < \infty$, as implied by Proposition 1. To balance comparative advantages, and to reflect our main motivation of the FRTB in the introduction, we decided to work with the formulation (2). Notably, for distributions with a constant PELVE, such as Pareto, uniform or exponential distributions (see Section 4), both formulations (2) and (3) give identical values for Π_ε and Π'_δ.

Remark 3. In the recent engineering and optimization literature, Mafusalov and Uryasev (2018) introduced the buffered Probability of Exceedance (bPOE), defined as the number $c_x(X) = c$ which solves $\mathrm{ES}_{1-c}(X) = x$, if it exists. For a continuously distributed X, PELVE is connected to the ratio $\theta_x(X) = c_x(X)/\mathbb{P}(X > x)$ via the following argument: if $\mathbb{P}(X > x) = \varepsilon$, then $\Pi_\varepsilon(X) = c_x(X)/\varepsilon = \theta_x(X)$. Our notion of PELVE is different from the ratio $\theta_x(X)$ in two ways. First, $\Pi_\varepsilon(X)$ has a parameter $\varepsilon \in (0,1)$ whereas $\theta_x(X)$ has a parameter $x \in \mathbb{R}$. The mappings Π_ε and θ_x exhibit very different theoretical properties; for instance, Π_ε is location-scale invariant (see Theorem 1) whereas θ_x is not. Second, PELVE is designed for and motivated by financial regulation and capital assessment (thus, ε has an economic meaning), whereas θ_x was motivated by buffering probability of failure in reliability engineering.

3 Theoretical features of PELVE

3.1 Invariance and ordering properties

In this section, we discuss a few theoretical features of PELVE: location-scale invariance, shape-relevance and quasi-convexity/concavity for comonotonic risks. As we have seen in Table 1, PELVE can be seen as a measure of variability. A key feature of PELVE, in contrast to commonly used measures of variability, such as variance or standard deviation, is location-

[7] For instance, (4) has a solution d only for a few values of $\delta \in (0,1)$ if X has a discrete distribution.

scale invariance. That means, adding a constant loss or gain, or multiplying a portfolio by a constant, does not change its PELVE value. Hence, PELVE captures the shape of the distribution instead of its location or scale. This is a particular advantage for portfolio risk assessment, as one often needs to compare portfolios in a scale-free way, such as using the Sharpe ratio.

Another attractive feature of PELVE is that it reports a higher value for a convex transform of the risk. Generally, for an increasing and convex function f and a random variable X, $f(X)$ has a heavier (right-)tail than X. For instance, if X is normally distributed, then e^X is log-normally distributed, which has a heavier tail than normal; if X has a Pareto(α) distribution, then X^2 has a Pareto($\alpha/2$) distribution, which has a heavier tail than Pareto(α). Therefore, with the property of reporting a higher value for $f(X)$ than for X, PELVE can be seen as a measure of tail-heaviness. Moreover, this property is an axiom for measures of (right) skewness of Oja (1981).

Finally, PELVE is both quasi-convex and quasi-concave for comonotonic risks.[8] The property of satisfying both quasi-convexity and quasi-concavity is known as *betweenness* in decision theory (e.g., Chew (1983)). The economic interpretation of this property is that combining two non-hedging risks does not yield a PELVE value beyond that of the worse risk, or one below that of the better risk, noting that comonotonic risks do not hedge each other; see e.g., Yaari (1987). Quasi-convexity and quasi-concavity are also important for optimization due to techniques in quasi-convex programming (e.g., Kiwiel (2001)). The above three features are summarized in the following theorem.

Theorem 1. *Suppose that $X \in L^1$, $\varepsilon \in (0,1)$ and $\mathbb{E}[X] \leqslant \mathrm{VaR}_{1-\varepsilon}(X)$.*

(i) For all $\lambda > 0$ and $a \in \mathbb{R}$, $\Pi_\varepsilon(\lambda X + a) = \Pi_\varepsilon(X)$.

(ii) $\Pi_\varepsilon(f(X)) \leqslant \Pi_\varepsilon(X)$ for all increasing concave functions $f : \mathbb{R} \to \mathbb{R}$ with $f(X) \in L^1$.

(iii) $\Pi_\varepsilon(g(X)) \geqslant \Pi_\varepsilon(X)$ for all strictly increasing convex functions $g : \mathbb{R} \to \mathbb{R}$ with $g(X) \in L^1$.

[8] Two random variables X and Y are said to be *comonotonic* if there exists a random variable Z and increasing functions f and g such that $X = f(Z)$ and $Y = g(Z)$ almost surely. A real function f on a set D is *quasi-convex* if $f(\lambda x + (1-\lambda)y) \leqslant \max\{f(x), f(y)\}$ for all $x, y \in D$, and it is *quasi-concave* if $f(\lambda x + (1-\lambda)y) \geqslant \min\{f(x), f(y)\}$ for all $x, y \in D$. Quasi-convexity (quasi-concavity) is obviously weaker than convexity (concavity).

(iv) *For all $\lambda \in [0,1]$ and increasing functions $f, g : \mathbb{R} \to \mathbb{R}$ with $f(X), g(X) \in L^1$, the value $\Pi_\varepsilon(\lambda f(X) + (1-\lambda)g(X))$ is between $\Pi_\varepsilon(f(X))$ and $\Pi_\varepsilon(g(X))$.*

The next few examples illustrate important implications of Theorem 1 in risk management. The first two examples address portfolio losses from an asset and simple derivatives on the same asset, and the third example is a comparison between linear losses and log-losses. The last example puts PELVE in the setting of risk forecast using time-series models.

Example 1 (Reducing PELVE with options)**.** Consider portfolios which consist of an asset and a European call or put option on the asset (with the same maturity). Via Theorem 1, we can compare the PELVE of the time-t loss of the following portfolios: (a) long one unit of the asset, with loss X_A; (b) long a call option, with loss X_C; (c) long one unit of the asset and a call option, with loss X_{AC}; (d) long one unit of the asset and a put option, with loss X_{AP}. By design, the losses from the above portfolios have some general relationships via concave or convex functions. More precisely, all of X_C, X_{AC}, and X_{AP} are increasing concave functions of X_A, and X_{AP} is an increasing concave function of X_{AC}; details of the derivation are given in Appendix A.2. Using Theorem 1 (ii) and the put-call parity, which implies $\Pi_\varepsilon(X_C) = \Pi_\varepsilon(X_{AP})$, the above relationships lead to

$$\Pi_\varepsilon(X_C) = \Pi_\varepsilon(X_{AP}) \leqslant \Pi_\varepsilon(X_{AC}) \leqslant \Pi_\varepsilon(X_A). \tag{5}$$

The conclusion in (5) is consistent with the natural intuition that purchasing options reduces the portfolio risk (measured by PELVE), and the protective put portfolio (d) is less risky than the long-asset-long-call portfolio (c). We remark that the relationship (5) does not depend on ε or any model assumptions on the financial asset. Moreover, the weights in the options and the asset in the above portfolios are irrelevant since they do not affect the relationship via concave or convex transforms.

Example 2 (Comparison of PELVE using quasi-convexity/concavity)**.** We continue in the setting of Example 1. Note that the loss from the asset X_A and that from the call option X_C are comonotonic. By the quasi-convexity/concavity in Theorem 1 (iv), we get

$$\min\{\Pi_\varepsilon(X_C), \Pi_\varepsilon(X_A)\} \leqslant \Pi_\varepsilon(X_{AC}) \leqslant \max\{\Pi_\varepsilon(X_C), \Pi_\varepsilon(X_A)\}. \tag{6}$$

If we know $\Pi_\varepsilon(X_C) \leqslant \Pi_\varepsilon(X_A)$, e.g., by using Theorem 1 (ii), then with the put-call parity we also arrive at (5). Note that the relationship (6) still holds true if the call option in X_C and X_{AC} is replaced by any other payoffs, as long as it is an increasing function of the asset price (not necessarily convex or concave).

Example 3 (Linear loss and log-loss)**.** For an asset with price X_t at time $t = 0, 1, \dots$ (e.g. daily prices), the one-period linear return at time t is defined as $R_t = X_t/X_{t-1} - 1$, and the log-return is $r_t = \log(X_t/X_{t-1})$. Note that risk measures are applied to losses instead of gains. Hence, relevant quantities are the linear loss (negative return) $-R_t = 1 - X_t/X_{t-1}$ and the log-loss (negative log-return) $-r_t = -\log(X_t/X_{t-1})$, which we shall analyze in the empirical studies in Section 6. Since $y \mapsto -\log(1-y)$ is a strictly increasing convex function, by Theorem 1 (iii), $\Pi_\varepsilon(-r_t) \geqslant \Pi_\varepsilon(-R_t)$, i.e. using log-loss will produce a (slightly) higher PELVE value than the linear loss; see Figure 9 in the Appendix for a comparison.

Example 4 (PELVE for time-series models)**.** In risk management practice, risk measure forecast (such as VaR and ES forecasts) is usually done via conditional models; see, e.g., Chapter 4 of McNeil et al. (2015) and Patton et al. (2019). Consider a time-series model for risk factors (e.g., GARCH)

$$X_t = \mu_t + \sigma_t Z_t, \quad t \in \mathbb{Z},$$

where μ_t and σ_t are the conditional mean and standard deviation given a σ-field $\mathcal{F}_{t-1}$, and Z_t is a random error independent of $\mathcal{F}_{t-1}$. Using Theorem 1 (i), we have $\Pi_\varepsilon(X_t|\mathcal{F}_{t-1}) = \Pi_\varepsilon(Z_t)$. Hence, the conditional PELVE is determined by that of the error term. Note that by using PELVE, we are interested in the shape of the loss distribution, instead of its location and scale, which are precisely the information contained in $\mathcal{F}_{t-1}$.

Remark 4. The increasing convex function g in Theorem 1 (iii) is required to be strictly increasing, and this requirement is essential. For instance, if g is a constant function, then $\Pi_\varepsilon(g(X)) = 1$ for all X since $g(X)$ is degenerate, which clearly violates $\Pi_\varepsilon(g(X)) \geqslant \Pi_\varepsilon(X)$ if $\Pi_\varepsilon(X) > 1$. Note, however, that for the increasing concave function f in Theorem 1 (ii), it does not have to be strictly increasing.

Remark 5. Generally, Π_ε is not quasi-convex or quasi-concave on non-comonotonic random variables, and this is because VaR_p is neither quasi-convex nor quasi-concave, although ES_p is convex.

3.2 Convergence properties

Next, we discuss convergence properties of PELVE, which are useful for statistical inference of PELVE (in particular, consistency). Let $\{X_n\}_{n\in\mathbb{N}}$ be a sequence of random variables such that $X_n \to X$ in distribution as $n \to \infty$. The following two common assumptions are used to guarantee the convergence of VaR and ES. More precisely, Assumption 2 below guarantees $\mathrm{VaR}_{1-\varepsilon}(X_n) \to \mathrm{VaR}_{1-\varepsilon}(X)$, and Assumption 3 below guarantees $\mathrm{ES}_q(X_n) \to \mathrm{ES}_q(X)$ for all $q \in (0,1)$. These claims are standard because, assuming convergence in distribution, the quantile function converges at continuous points of its limit, and ES_q is continuous with respect to L^1 convergence which is guaranteed by uniform integrability.

Assumption 2 (Continuous quantile)**.** $F_X^{-1}(q)$ is continuous at $q = 1-\varepsilon$.

Assumption 3 (Uniform integrability)**.** $\{X_n\}_{n\in\mathbb{N}}$ is uniformly integrable.

With the above assumptions, we can establish a convergence result of PELVE in the following theorem. Throughout, $\xrightarrow{\mathrm{d}}$ represents convergence in distribution.

Theorem 2. *Suppose that $\{X_n\}_{n\in\mathbb{N}} \subset L^1$, $X \in L^1$ and $\varepsilon \in (0,1)$ satisfy Assumptions 1-3 and $X_n \xrightarrow{\mathrm{d}} X$ as $n \to \infty$. Then $\Pi_\varepsilon(X_n) \to \Pi_\varepsilon(X)$ as $n \to \infty$.*

Note that for iid or α-mixing (see Section 5.2) data, the empirical distribution converges to the true distribution. Moreover, for stationary observations, Assumption 3 is satisfied automatically, where X_n follows from the empirical distribution based on n observations. Therefore, Theorem 2 implies that the empirical PELVE estimator is consistent under Assumptions 1-2; see Corollary 1 in Section 5.

3.3 Effect of diversification

Diversification techniques are widely used in portfolio risk management. A proper diversification should, in theory, reduce the volatility of the portfolio, and this is often reflected in a reduction of standard deviation, VaR, ES, or other risk measures of the portfolio loss compared to the the sum of the individual assets.[9] Below we will discuss the impact on PELVE of portfolio diversification through results in Theorems 1 and 2.

[9]Here the term "risk measures" is used in a broad sense, which includes non-monotone functionals such as standard deviation and variance.

Suppose that $X_1, \dots, X_n$ are losses from individual sources with finite variances. The portfolio loss S_n is defined as $S_n = \sum_{i=1}^n w_i X_i$ for some weights $w_1, \dots, w_n \geqslant 0$, and we will allow n to grow. Assume that as n increases, $(S_n - a_n)/b_n$ converges in distribution to a normal distribution for some constants $a_n \in \mathbb{R}$ and $b_n > 0$. This assumption is satisfied as soon as any form of the central limit theorem (CLT) applies (Bilingsley (1999) has many such conditions). By Theorems 1 and 2,

$$\Pi_\varepsilon(S_n) = \Pi_\varepsilon\left(\frac{S_n - a_n}{b_n}\right) \to \Pi_\varepsilon(N), \tag{7}$$

where N is a standard normal random variable. For instance, $\Pi_{0.01}(N) = 2.58$ from Table 1. Most common distributions in risk management have higher PELVE values than the normal distribution; this will be clear from the examples in Section 4 and the empirical results in Section 6. Therefore, if CLT conditions hold, then the loss from a well-diversified portfolio has asymptotically smaller PELVE than individual assets. Nevertheless, we need to keep in mind that both individual assets and portfolio loss data are empirically not very close to normality due to the presence of systematic risk; see e.g., Chapter 3 of McNeil et al. (2015). Thus, even if there is a tendency to normality, the convergence is slow. Hence, we do not claim $\Pi_\varepsilon(S_n) \approx \Pi_\varepsilon(N)$; instead, $\Pi_\varepsilon(S_n)$ is likely closer to $\Pi_\varepsilon(N)$ than $\Pi_\varepsilon(X_1), \dots, \Pi_\varepsilon(X_n)$ are. This statement will be verified in the empirical analysis in Section 6.

Diversification techniques can simultaneously reduce both VaR and ES of a portfolio, and it is not obvious which one is reduced more significantly.[10] PELVE provides an answer to the above question: assuming that the individual losses have a higher PELVE than normal, the portfolio ES is reduced more significantly than VaR by a proper diversification.

Remark 6. The result in this section should not be confused with, and is not a consequence of, the coherence of ES in Artzner et al. (1999). Indeed, the conclusion needs to be flipped if the individual assets have $\Pi_\varepsilon(X_i) < \Pi_\varepsilon(N)$ (e.g., uniform distributions), although such a case is unlikely observed in asset loss data.

[10] For instance, for a vector $(X_1, \dots, X_n)$ following an elliptical distribution, diversification ratios for VaR and ES are identical; see Theorem 8.28 of McNeil et al. (2015).

4 PELVE for various distributions and its limit

In this section, we present formulas and curves of Π_ε for common distributions which are used in Table 1, and discuss the interpretation of these values. We proceed to analyze the limit of PELVE as $\varepsilon \downarrow 0$ for heavy-tailed distributions, noting that in risk management practice ε is typically quite small.

4.1 Formulas and numerical values for commonly used distributions

As shown in Theorem 1, location and scale parameters are irrelevant for the calculation of Π_ε. We thus only need to consider a representative of each location-scale family of distributions. Furthermore, we are interested in the cases that ε is close to 0, and hence all results are stated for small values of ε so that Assumption 1 holds. All calculations are based on analytical formulas for VaR and ES and solving (1), whose details are given in Appendix A.3.

Example 5 (Uniform/Pareto/Exponential distributions)**.** We first illustrate three distributions that have a constant PELVE.

(i) If X is uniformly distributed on some interval, then $\Pi_\varepsilon(X) = 2$ for $\varepsilon \leqslant 1/2$.

(ii) If X is exponentially distributed, then $\Pi_\varepsilon(X) = e$ for $\varepsilon \leqslant 1/e$.

(iii) If X has a Pareto(α) distribution with shape parameter $\alpha > 1$, that is, $\mathbb{P}(X > x) = x^{-\alpha},\ x \geqslant 1$,[11] then $\Pi_\varepsilon(X) = (\frac{\alpha}{\alpha-1})^\alpha$ for $\varepsilon \leqslant (\frac{\alpha}{\alpha-1})^{-\alpha}$.

Figure 1 panel (a) reports the value of $\Pi_\varepsilon(X)$ for Pareto(α) over α. By elementary calculus, the function $\alpha \mapsto (\frac{\alpha}{\alpha-1})^\alpha$ is decreasing in α, and

$$\Pi_\varepsilon(X) \geqslant \lim_{\alpha\to\infty} \left(\frac{\alpha}{\alpha-1}\right)^\alpha = e \approx 2.718. \tag{8}$$

As its shape parameter increases, the PELVE of a Pareto distribution decreases to e, which is that of an exponential distribution. Hence, the value e can be seen as a threshold of PELVE which distinguishes empirically between power-like distributions and light-tailed distributions. This observation will be discussed further in Section 4.2.

[11] Any other parametrization of the Pareto distribution, such as $\mathbb{P}(X > x) = (\theta/(x+\theta))^{-\alpha},\ x \geqslant 0$, where $\theta > 0$, has the same $\Pi_\varepsilon(X)$ because of location-scale invariance.

Example 6 (Normal/t/log-normal distributions)**.** For normal, t and log-normal distributions, PELVE does not admit an analytical formula, but it can be easily calculated from VaR and ES formulas. We write $t(\nu)$ for the t-distribution with parameter $(0, 1, \nu)$, where $\nu > 1$ is the parameter of degrees of freedom, and $\mathrm{LN}(\sigma^2)$ for the log-normal distributions with parameter $(0, \sigma^2)$. Curves of PELVE for the above distributions are reported in Figure 1 panels (b)-(f).

Most PELVE values of the normal distribution are close to 2.5, consistent with the fact that the normal distribution is a typical example of a light-tailed distribution. The $\Pi_{0.01}$ curve of t-distributions with $\nu > 1$ is very similar to that of Pareto distributions with $\alpha > 1$, and as $\nu \to \infty$, $\Pi_{0.01}$ of $t(\nu)$ approaches the corresponding value of a normal distribution (2.58), as the t-distribution converges to a normal distribution.

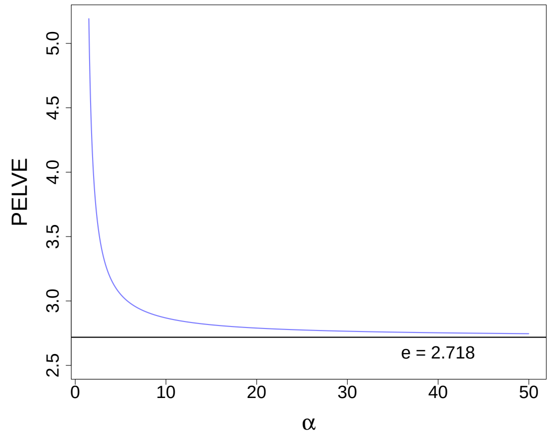
The value $\Pi_{0.01}(X) = 3.13$ for $X \sim \mathrm{LN}(1)$ is relatively large, which is roughly equivalent to that of Pareto(4). Using PELVE and the threshold $e \approx 2.718$, for a practical value of $\varepsilon = 0.01$, the log-normal distribution with $\sigma^2 = 1$ behaves (practically) like a heavy-tailed distribution, whereas the log-normal distribution with $\sigma^2 = 0.04$ behaves like a light-tailed distribution. The log-normal distributions, regardless of their parameters, are not categorized as heavy-tailed distributions in theory; for instance, in McNeil et al. (2015), heavy-tailed distributions have power-like survival functions. This suggests that PELVE can practically (instead of asymptotically) distinguish distributions with moderate tails better than methods based on the tail-index or moments; this point will be continued in Remark 7 below.

4.2 Limit of PELVE for heavy-tailed distributions

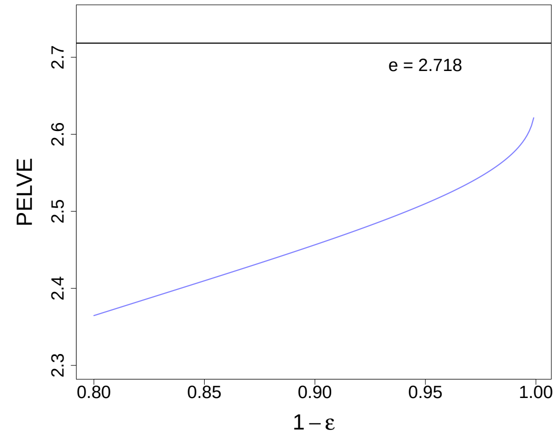
It is well known that financial return data show heavy tails (see e.g. Cont (2001); Andersen et al. (2021)), and hence it is common to model financial returns and portfolio losses using heavy-tailed distributions. These heavy-tailed distributions are typically defined with a *regularly varying* tail, a classic notion in Extreme Value Theory (EVT); for a general treatment, we refer to Embrechts et al. (1997). In this section, we illustrate the role of PELVE in the analysis of tail heaviness.

An eventually non-negative (that is, $f(x) \geqslant 0$ for x large enough) measurable function f is said to be *regularly varying* (RV) with a regularity index $\gamma \in \mathbb{R}$, if

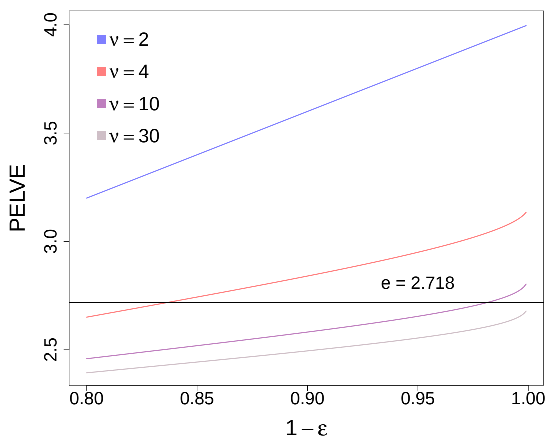
$$\lim_{x \to \infty} \frac{f(tx)}{f(x)} = t^{\gamma}, \quad \text{for all } t > 0.$$



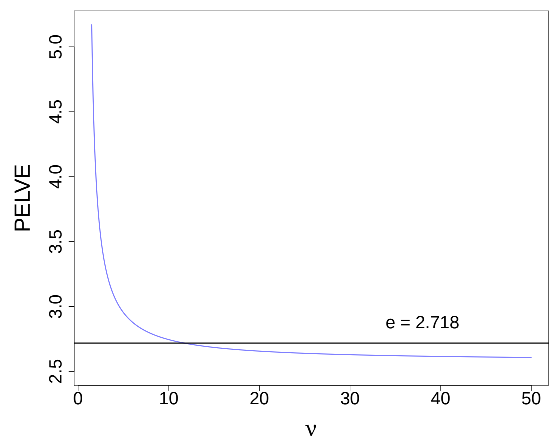
(a) Pareto(α), $\alpha \in [1.5, 50]$



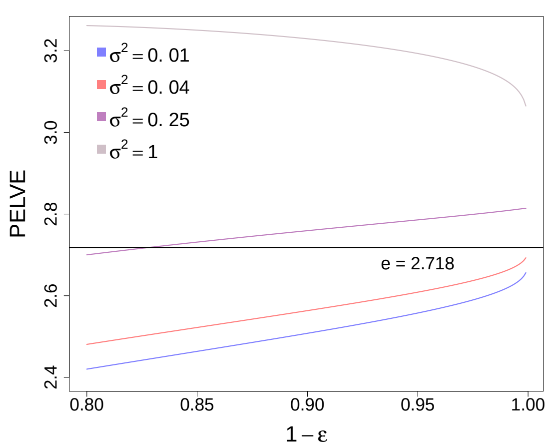
(b) N(μ, σ^2), $\varepsilon \in (0, 0.2]$



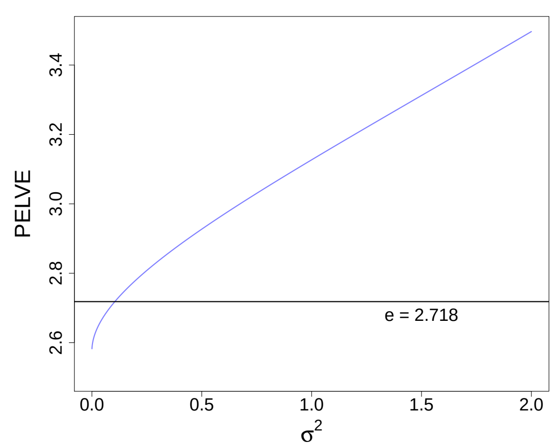
(c) t(ν), $\varepsilon \in (0, 0.2]$, $\nu = 2, 4, 10, 30$



(d) t(ν), $\varepsilon = 0.01$ $\nu \in [1.5, 50]$



(e) LN(σ^2), $\varepsilon \in (0, 0.2]$, $\sigma^2 = 0.01, 0.04, 0.25, 1$



(f) LN(σ^2), $\varepsilon = 0.01$, $\sigma^2 \in (0, 2]$

Figure 1: Values of Π_ε for various common distributions

Denote this by $f \in \mathrm{RV}_\gamma$. For a distribution function F, let $\overline{F}(\cdot) = 1 - F(\cdot)$ denote its survival function. Many heavy tailed distributions has an RV survival function, such as the t-distribution or the Pareto distribution. The *tail-index* of a distribution F with an RV survival function, is the constant $\xi > 0$ such that $\overline{F} \in \mathrm{RV}_{-\xi}$. For instance, a Pareto distribution with shape parameter α has a tail-index α, and a t-distribution with ν degrees of freedom has a tail-index ν. A tail-index larger than 1 indicates a finite mean, and a tail-index larger than 2 indicates a finite variance.

Below we investigate the PELVE of distributions with a tail index $\alpha > 1$. The condition $\alpha > 1$ is assumed to guarantee the existence of the mean; recall that Π_ε is defined on L^1. Note that an RV function only concerns its behaviour close to infinity, and as such we analyze the limit of Π_ε as $\varepsilon \downarrow 0$.

Theorem 3. *Suppose that the function $\overline{F}(x) = \mathbb{P}(X > x)$ is $\mathrm{RV}_{-\alpha}$, $\alpha > 1$. Then*

$$\lim_{\varepsilon \downarrow 0} \Pi_\varepsilon(X) = \left(\frac{\alpha}{\alpha - 1}\right)^\alpha.$$

The limiting value in Theorem 3 coincides with the PELVE of Pareto distributions for fixed $\varepsilon > 0$, which is of course not a surprise. Moreover, Theorem 3 suggests a monotonic relationship between the limit of PELVE and the tail index. Indeed, the heavier the tail (i.e., smaller α), the larger the limit of the PELVE value. For heavy-tailed distributions, the limit of their PELVE is larger than $e \approx 2.718$, even if the tail index is very large; see Figure 1 (a) and (8). This further supports that the value $e \approx 2.718$ can be seen as a threshold for PELVE to distinguish empirically heavy-tailed phenomenon and light-tailed phenomenon, as discussed in Example 5. Recall that $\Pi_{0.01}(X) = 2.58$ for a normally distributed X, showing that normal distributions are indeed light-tailed, which is consistent to our common interpretation. On the other hand, $\Pi_{0.01}(X) = 3.13$ for a log-normally distributed X with parameter $\sigma^2 = 1$ as we have seen in Example 6.

Although the limit of PELVE and the tail index are linked by Theorem 3, they are conceptually quite different, and PELVE has several practical advantages. First, the tail index relies on the asymptotic behaviour of the distribution, as well as the technical assumption of regular variation, whereas PELVE is well defined for all distributions with a finite mean, regardless of the tail being regularly varying or not. Second, PELVE is well defined for random variables with a bounded support. Third, estimating the tail index typically involves an ad-hoc choice of the threshold (such as in the Hill estimator; see Embrechts et al. (1997)), whereas PELVE for a fixed ε can be estimated directly from empirical data with minimal model assumptions (Section 5).[12] Fourth, PELVE has the advantage of capturing diversification (see the empirical results in Section 6) whereas the tail index remains the same for the average of iid heavy tailed risks. Fifth, $\Pi_{0.01}$ has a concrete meaning for banking regulation, which is the probability multiplier to change from VaR to ES while maintaining the same capital level.

Remark 7. As PELVE depends on a small but fixed probability level ε, the heavy-tailed phenomenon in this section should be interpreted as "empirically or practically heavy-tailed", meaning that the observed data have a similar pattern as those from a heavy-tailed distribution, at a probability level of risk management relevance. The purpose of PELVE is not to theoretically detect tail-heaviness in the classic sense (an asymptotic property), which is typically defined via a regular varying tail or the existence of a moment generating function.

Theorem 3 has a practical implication for the choice of risk measure in the FRTB, where the Basel Committee used the specific pair of $q = 0.975$ and $p = 0.99$, leading to a constant multiplier of 2.5. Note that the limit of Π_ε as $\varepsilon \downarrow 0$ is larger than e for heavy tailed distributions. This means, by changing from $\mathrm{VaR}_{0.99}$ to $\mathrm{ES}_{0.975}$, the capital requirement will be increased for heavy-tailed risks. To make the above statements precise, we give an approximation of $\mathrm{ES}_{0.975}(X)/\mathrm{VaR}_{0.99}(X)$ for heavy-tailed risk X.

Proposition 3. *Suppose that the function $\overline{F}(x) = \mathbb{P}(X > x)$ is $\mathrm{RV}_{-\alpha}$, $\alpha > 1$. Then, for $c > 1$,*

$$\lim_{\varepsilon \downarrow 0} \frac{\mathrm{ES}_{1-c\varepsilon}(X)}{\mathrm{VaR}_{1-\varepsilon}(X)} = \frac{\alpha}{\alpha - 1} c^{-1/\alpha} = \lim_{\varepsilon \downarrow 0} \left(\frac{\Pi_\varepsilon(X)}{c} \right)^{1/\alpha}.$$

In particular,

$$\lim_{\varepsilon \downarrow 0} \frac{\mathrm{ES}_{1-2.5\varepsilon}(X)}{\mathrm{VaR}_{1-\varepsilon}(X)} = \frac{\alpha}{\alpha - 1} 2.5^{-1/\alpha} = \lim_{\varepsilon \downarrow 0} \left(\frac{\Pi_\varepsilon(X)}{2.5} \right)^{1/\alpha} > 1.$$

The ratio $\frac{\alpha}{\alpha-1} 2.5^{-1/\alpha} \approx (\frac{\Pi_\varepsilon(X)}{2.5})^{1/\alpha}$ in Proposition 3 approximates the ratio between the capital requirements of the same risk under $\mathrm{ES}_{0.975}$ and $\mathrm{VaR}_{0.99}$. This ratio is typically not

[12] Recall that in the context of our main motivation, ε is specified exogenously by a regulatory accord. On the other hand, if one is interested in the asymptotic behaviour (e.g., tail heaviness), then the parameter ε may be seen as a counterpart to the ad-hoc threshold in estimating the tail index.

very far away from 1; for instance, it is 1.02 for $\alpha = 8$, 1.06 for $\alpha = 4$, and 1.26 for $\alpha = 2$. Note however that even the ratio is close to 1, $\Pi_\varepsilon(X)$ may not be close to 2.5; instead, $\lim_{\varepsilon \downarrow 0} \Pi_\varepsilon(X)$ goes to e as $\alpha \to \infty$ by Theorem 3. For t-distributions, the value $\varepsilon = 0.01$ used by the FTRB is sufficiently small for the approximation in Proposition 3 to be useful.

5 Non-parametric estimation of PELVE

5.1 Empirical PELVE estimator for iid data

Suppose that $X_1, X_2, \dots$ are iid sample observations for $X \in L^1$ with distribution F satisfying Assumption 1, and denote by F_n the empirical distribution of $X_1, \dots, X_n$. Here we assume iid for tractability of the asymptotic variance formula; the more general case of α-mixing sequences is analyzed in Section 5.2.

By construction, the estimation of PELVE involves estimating VaR and ES. The most straightforward way to estimate VaR and ES is to directly use those of the empirical distribution. Then, one can define the *empirical PELVE estimator* $\widehat{\Pi}_\varepsilon(n)$ by solving the equation

$$\widehat{\mathrm{ES}}_{1-c\varepsilon} = \widehat{\mathrm{VaR}}_{1-\varepsilon} \quad \text{for } c \in [1, 1/\varepsilon], \tag{9}$$

where $\widehat{\mathrm{ES}}$ and $\widehat{\mathrm{VaR}}$ are the empirical ES and VaR, respectively, i.e., those of the empirical distribution F_n from the data. Recall that the empirical VaR is the sample quantile given by

$$\widehat{\mathrm{VaR}}_p = X_{[i]} \ \text{ for } \ p \in \left(\frac{i-1}{n}, \frac{i}{n}\right], \ i = 1, \dots, n,$$

where $X_{[1]} \leqslant \dots \leqslant X_{[n]}$ are the order statistics of the data $X_1, \dots, X_n$, and the empirical ES is given by

$$\widehat{\mathrm{ES}}_p = \frac{1}{1-p} \int_p^1 \widehat{\mathrm{VaR}}_q \mathrm{d}q, \quad p \in (0, 1).$$

We implicitly assume that c solving (9) is unique since Assumption 1 is almost always satisfied for the empirical distribution. Consistency of $\widehat{\Pi}_\varepsilon(n)$ is a direct consequence of Theorem 2.

Corollary 1. *Suppose that $X, X_1, X_2, \dots \in L^1$ are iid, $\varepsilon \in (0, 1)$, and Assumptions 1-2 hold. Then $\widehat{\Pi}_\varepsilon(n) \to \Pi_\varepsilon(X)$ as $n \to \infty$.*

Remark 8. As an alternative to the empirical quantile, one may implement a smoothed quantile estimator, such as the standard linearly interpolated VaR in McNeil et al. (2015, Section 9.2.6), to compute both VaR and ES estimates. This leads to the smoothed empirical PELVE estimator which we treat in Appendix B. From the simulation results (see Appendix C), the smoothed estimator does not improve the overall estimation quality and leads to a visible bias. As such, we stick to the empirical estimator $\widehat{\Pi}_\varepsilon(n)$ in our main paper.

Next, we establish the asymptotic normality of $\widehat{\Pi}_\varepsilon(n)$ under a stronger condition than Assumptions 1-2. For the consistency result in Corollary 1, it suffices if X has the first moment, and hence, it works for Pareto(α) for $\alpha > 1$; for the asymptotic normality result in Theorem 4, we need slightly more than the second moment (Assumption 4), and as such, for Pareto(α) it is valid only for $\alpha > 2$.

Assumption 4 (Regularity)**.** $X \sim F$ has a density function f which is positive and continuous at $F^{-1}(1-\varepsilon)$, and $\mathbb{E}[|X|^{2+\delta}] < \infty$ for some $\delta > 0$.

In the next theorem, $\widetilde{\Pi}(n)$ is the smoothed empirical PELVE estimator explained in Remark 8 and formally defined in Appendix B.

Theorem 4 (Asymptotic normality)**.** *Suppose that $X, X_1, X_2, \dots \in L^1$ are iid, $\varepsilon \in (0,1)$, and Assumptions 1 and 4 hold. Let $c = \Pi_\varepsilon(X)$ and $\hat{c}_n = \widehat{\Pi}(n)$ or $\hat{c}_n = \widetilde{\Pi}(n)$. Then, as $n \to \infty$*

$$\sqrt{n}(\hat{c}_n - c) \xrightarrow{\mathrm{d}} \frac{1}{F^{-1}(p) - F^{-1}(q)} \left(\int_q^1 \frac{W(t)}{\varepsilon f(F^{-1}(t))} \mathrm{d}t - \frac{cW(p)}{f\big(F^{-1}(p)\big)} \right),$$

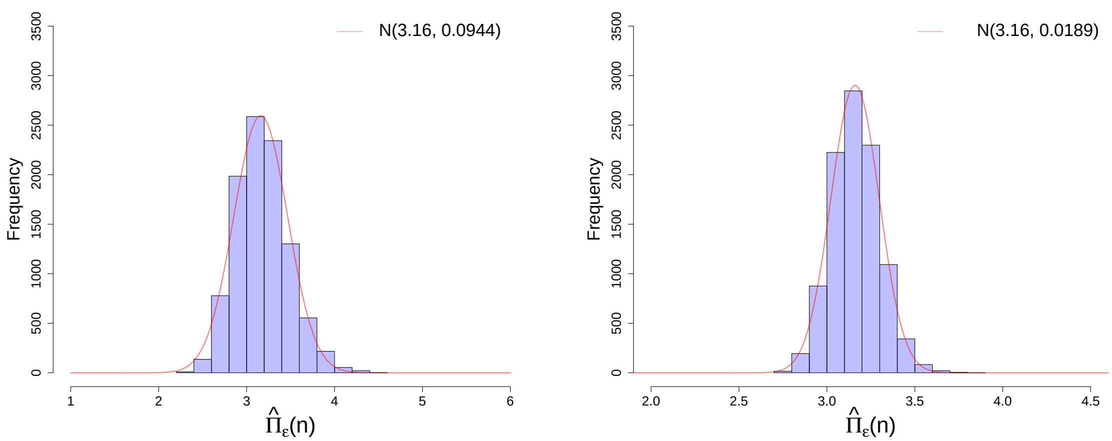
where $p = 1-\varepsilon$, $q = 1 - c\varepsilon$, and W is a standard Brownian bridge on $[0,1]$. In particular, $\sqrt{n}(\hat{c}_n - c) \xrightarrow{\mathrm{d}} \mathrm{N}(0, \sigma^2)$ and σ^2 can be computed as

$$\sigma^2 = \frac{1}{b^2} \left(a^2(\varepsilon - \varepsilon^2) + \frac{2}{\varepsilon^2} \int_{F^{-1}(q)}^{\infty} E_{F(x)} F(x) \mathrm{d}x - \frac{2a}{\varepsilon} E_p + 2a(E_q - b) \right), \tag{10}$$

where $E_t = \mathbb{E}[(X - F^{-1}(t))_+]$ for $t \in (0,1)$, $a = c/f(F^{-1}(p))$ and $b = F^{-1}(p) - F^{-1}(q)$.

Two examples of σ^2 for uniform and Pareto distributions are presented in Appendix A.5, where we see that σ^2 depends on ε and explodes at the rate of $O(\varepsilon^{-1})$ as $\varepsilon \downarrow 0$. This is because the non-parametric estimators for $\mathrm{ES}_{1-c\varepsilon}$ and $\mathrm{VaR}_{1-\varepsilon}$ effectively use $cn\varepsilon$ and $n\varepsilon$ data points, respectively. In practice, a very small value of ε may lead to large estimation errors of $\widehat{\Pi}_\varepsilon$, similar to the empirical estimation of VaR and ES. Fortunately, as shown in Table 1, Π_ε is typically quite stable across values of ε, and hence one can get useful information on PELVE without choosing ε too small.

Figure 2 below is produced with sample sizes $n = 1000$ and $n = 5000$ with losses simulated from Pareto(4), and the procedure is repeated 10,000 times. We observe that the PELVE empirical estimates match quite well with the density functions of $\mathrm{N}(3.16, 0.0944)$ and $\mathrm{N}(3.16, 0.0189)$, the respective theoretical limiting distributions. More simulation results are put in Appendix C.



(a) Histogram of $\widehat{\Pi}_\varepsilon(n)$, $n = 1000$

(b) Histogram of $\widehat{\Pi}_\varepsilon(n)$, $n = 5000$

Figure 2: Empirical PELVE estimators for Pareto(4), $\varepsilon = 0.05$

5.2 Non-parametric estimation for dependent data

Next, we analyze the empirical PELVE estimators for a dependent sequence of data. We follow the notation in Section 5.1, without assuming that our observations comes from an iid sample. Instead, we assume that the sequence[13] $\{X_n\}_{n\in\mathbb{Z}}$ is strictly stationary and α-mixing, that is,

$$\alpha(k) := \sup\left\{\left|\mathbb{P}(A\cap B) - \mathbb{P}(A)\mathbb{P}(B)\right| : A \in \mathcal{F}^i_{-\infty}, B \in \mathcal{F}^\infty_{i+k}, -\infty < i < \infty\right\} \to 0$$

[13]Here, $\mathbb{Z}$ is the set of integers, and we include X_n for negative n in the sequence to formally define an α-mixing sequence. Only X_n for $n \in \mathbb{N}$ will be used for estimation.

as $k \to \infty$, where $\mathcal{F}_a^b$ denotes the σ-field generated by $\{X_n\}_{a \leqslant n \leqslant b}$. Certainly, an iid sequence is a special case of α-mixing sequences, with $\alpha(k) = 0$ for $k \geqslant 1$.

Assumption 5 (α-mixing observations)**.** The sequence $\{X_n\}_{n\in\mathbb{Z}}$ is strictly stationary and α-mixing, and $\alpha(k) = O(a^k)$ as $k \to \infty$ for some $a \in (0, 1)$.

The definition of the empirical PELVE estimator $\widehat{\Pi}(n)$ is the same as in Section 5.1, given by (9). Recall that the data used for $\widehat{\Pi}(n)$ are $X_1, \dots, X_n$. The consistency of the estimator $\widehat{\Pi}(n)$ is guaranteed by Theorem 2, noting that the empirical distribution converges to the true distribution for an α-mixing sequence of observations. Below, we establish the asymptotic normality, which completes our theoretical analysis.

Proposition 4 (Asymptotic normality for α-mixing data)**.** *Suppose that Assumptions 1, 4 and 5 hold. Let $c = \Pi_\varepsilon(X_0)$ and $\hat{c}_n = \widehat{\Pi}(n)$ or $\hat{c}_n = \widetilde{\Pi}(n)$. Then, as $n \to \infty$,*

$$\sqrt{n}(\hat{c}_n - c) \xrightarrow{\mathrm{d}} \frac{1}{F^{-1}(p) - F^{-1}(q)} \left(\int_q^1 \frac{W(t)}{\varepsilon f(F^{-1}(t))} \mathrm{d}t - \frac{cW(p)}{f\big(F^{-1}(p)\big)} \right), \tag{11}$$

where $p = 1 - \varepsilon$, $q = 1 - c\varepsilon$, and W is a Gaussian process on $[0, 1]$ with mean 0 and covariance

$$\mathrm{Cov}(W(t), W(s)) = \sum_{i\in\mathbb{Z}} \Big(\mathbb{P}(\{F(X_1) \leqslant s\} \cap \{F(X_{1+i}) \leqslant t\}) - ts \Big). \tag{12}$$

In particular, $\sqrt{n}(\hat{c}_n - c) \xrightarrow{\mathrm{d}} \mathrm{N}(0, \sigma^2)$ where σ^2 is the variance of the right-hand side of (11).

Unlike in Theorem 4, the asymptotic variance formula of the empirical PELVE estimator for the α-mixing data is not explicit. Therefore, confidence intervals are not straightforward to construct, and one may need to rely on other variance estimation techniques such as bootstrap. In Section 6 below, we produce bootstrap confidence bands for our estimated PELVE curves.

5.3 Asymptotic variance estimation

To estimate the asymptotic variance σ^2 of the estimator $\widehat{\Pi}(n)$ in the setting of α-mixing data, we implement a non-overlapping block-bootstrap method of Carlstein (1986). Other common dependent bootstrap methods work similarly; see Lahiri (2003) for a comprehensive treatment.

Given a sample $X_1, \ldots, X_n$, we break it into k blocks of the form $(X_{im+1}, \ldots, X_{(i+1)m})$ for $i = 0, 1, \ldots, k-1$, where $k = n/m$ (which may hold only approximately). We assume that m and k both tend to ∞ as $n \to \infty$, as in the usual setting of block bootstrap; for instance, one may choose $m = O(n^{1/3})$ as suggested by Hall et al. (1995), which has an asymptotically optimal efficiency for the block-bootstrap variance estimator. A bootstrap sequence is constructed by sampling k blocks with replacement, and pasting them together to form a sequence of length n.

Let F_n^b be the empirical distribution of the bootstrapped series for $b = 1, \ldots, B$, and define the bootstrap error $S_b = \sqrt{n}(T(F_n^b) - T(F_n))$. The block-bootstrap variance estimator $\hat{\sigma}_B^2$ is the empirical variance of S_b, $b = 1, \ldots, B$. Letting $B \to \infty$, $\hat{\sigma}_B^2$ converges to the bootstrap variance $\hat{\sigma}_*^2$ conditional on the data. The following simple result shows that the bootstrap variance converges to the asymptotic variance of the empirical PELVE estimator.

Proposition 5. *Under the assumptions of Proposition 4, we have $\hat{\sigma}_*^2 \to \sigma^2$ as $m, n \to \infty$ and $m = O(n^\delta)$ for some $\delta \in (0, 1/2)$.*

The consistency result in Proposition 5 is obtained from classic results on the bootstrap empirical process (Bühlmann (1994)) and the weak convergence in (11). A small simulation example on the block-bootstrap variance estimator is reported in Table 7 in Appendix C.

6 Empirical analysis for financial data

With the theoretical attractions of PELVE discussed in the previous sections, we next illustrate, using real data, the advantages and additional information provided by PELVE beyond VaR and ES, by means of two empirical studies. To reflect the distributional feature of realized losses and avoid imposing any model specification, we mainly focus on PELVE of the empirical distributions of financial losses.

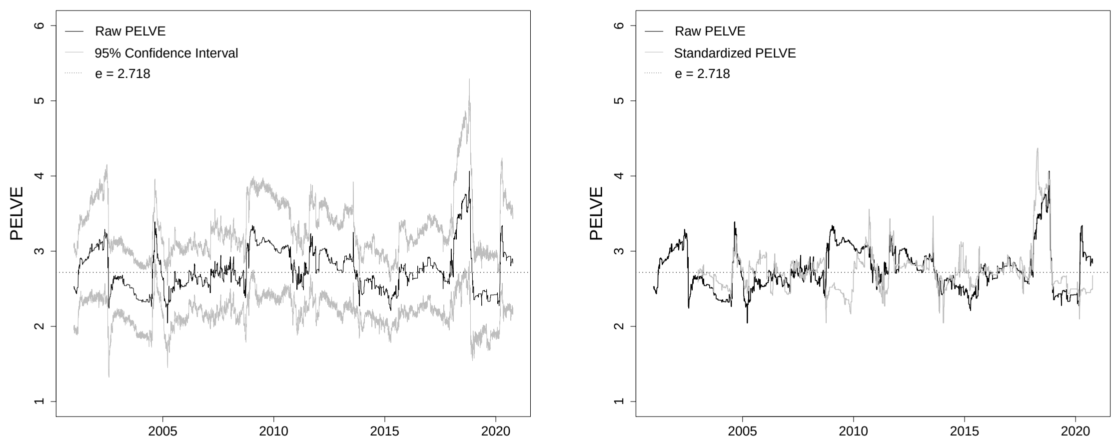
6.1 PELVE for asset return data

We gather the historical price movements, ranging from Jan 4, 1999 to Oct 9, 2020, of the S&P 500 index and its constituents.[14] For the index and each individual constituent, we

[14] The price data are obtained from Yahoo Finance.

use its daily log-loss (negative log-return) data[15] over the observation period with a moving window of 500 days for daily estimations of VaR, ES and PELVE. For instance, an estimation of PELVE on 01/02/2001 will use log-loss data roughly from Jan 1999 to Dec 2000.

We report in Figure 3 (a) the daily 5% empirical PELVE[16] of log-losses from the index, using the empirical estimator defined in Section 5. We also report 95% bootstrap confidence bands, which are constructed by fitting an AR(1)-GARCH(1,1) model with t-innovations to each window and resampling the residuals (see e.g., Section 4 of Asimit et al. (2019)).[17] The main observation here is that the S&P 500 index has a PELVE around $e \approx 2.72$ over the entire period (with mean 2.76).



(a) Bootstrap confidence bands for 5% PELVE

(b) Raw vs standardized 5% PELVE

Figure 3: 5% PELVE of the S&P 500 index

Next, we calculate 5% empirical PELVE to examine tail-heaviness of different sectors of S&P 500 from Jan 2001 to Oct 2020. We take the average PELVE value of stocks in each sector of S&P 500 to obtain the sector-wise PELVE values on each observation day. In Figure 4, we report the 5% PELVE values of the Financials and IT sectors, as well as their average 99% VaR and 97.5% ES.[18] For the remaining 9 sectors, see Figures 10 and 11 in Appendix

[15] We use log-loss instead of linear loss to be consistent with most studies on financial asset return data. In Example 3, we have seen that using log-loss will likely increase the PELVE value compared to using linear loss. We report the difference between the PELVE value of log-loss and that of linear loss for S&P 500 in Figure 9 in the Appendix. We can see that the difference is negligible. This is an advantage of PELVE over the tail index, since a log transform will change the tail index drastically (at least in theory).

[16] We choose the level 5% for statistically reliable estimates of PELVE. For results on 1% and 2.5% PELVE, which have more statistical errors, see Appendix D.3.

[17] We use 200 bootstrap samples in each daily estimation.

[18] The sector-wise bootstrap sample are constructed by fitting an AR(1)-GARCH(1,1) model with t-

D. We make several observations by comparing the PELVE values across different sectors and time periods.

1. We report an overall average estimated PELVE of 2.98 across all 11 sectors. Most point estimates of sector-average PELVE curves lie between 2.8 and 3.5 prior to the COVID-19 pandemic outbreak in the US (i.e., early 2020), and they are almost always above $e \approx 2.72$. These values are consistent with the common view that the daily log-loss distributions are empirically heavy-tailed with a tail index typically between 3 and 5 (see e.g. Jansen and De Vries (1991); Cont (2001)), which suggests a PELVE value between 3.05 and 3.38 according to Theorem 3 (at least for small ε).[19] In contrast, the PELVE values of the S&P 500 index are around $e \approx 2.72$, which does not indicate a heavy tail.

2. Spikes in PELVE are located around the 2008 subprime crisis and the COVID-19 period for all sectors, and the same shock heightened the sector-wise tail distribution differently. Overall, PELVE values are quite stable during the past 20 years except around 2008 and 2020.

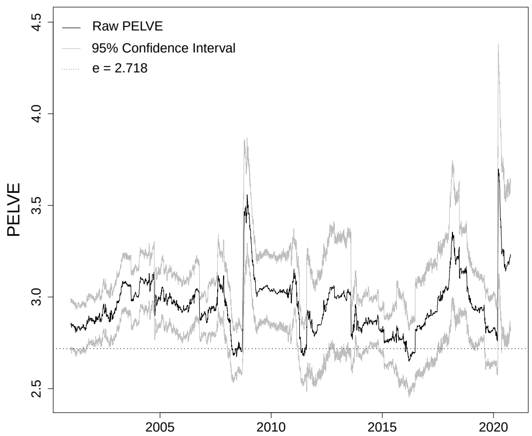
3. By looking at the 20-year VaR/ES reported in Figure 4 (c) and (d) alone, the dot-com bubble seems to be much more severe than both the 2008 subprime crisis and COVID-19 in the IT sector, and the subprime crisis is much more severe than COVID-19 in the Financials sector. In sharp contrast, empirical PELVE curves exhibit unprecedented peaks during the COVID-19 period for almost all sectors of S&P 500, indicating extreme turbulence in the financial market.

4. Comparing panels (a) and (b) in Figure 4, we observe that the sector-wise PELVE of Financials fluctuates more than that of IT over time and with wider confidence bands. This could possibly be explained by the fact that stocks in Financials are more correlated.[20]

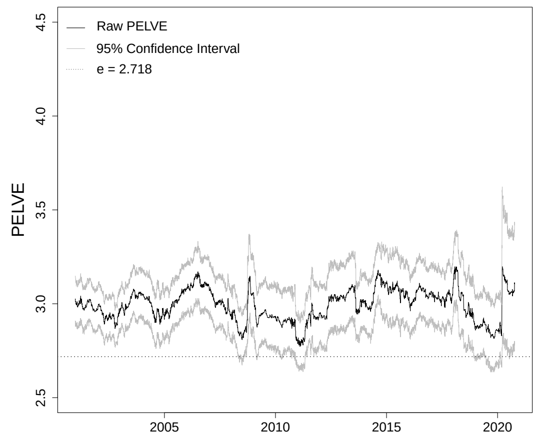
innovations to each stock and resampling the vectors of residuals, so that the dependence among stocks within the sector is reflected.

[19] Note however that the estimation of the tail index requires regular variation which may not be satisfied by financial return distribution, whereas our PELVE does not require any extra assumption except for a finite mean. This shows another technical advantage of using PELVE over the tail index.

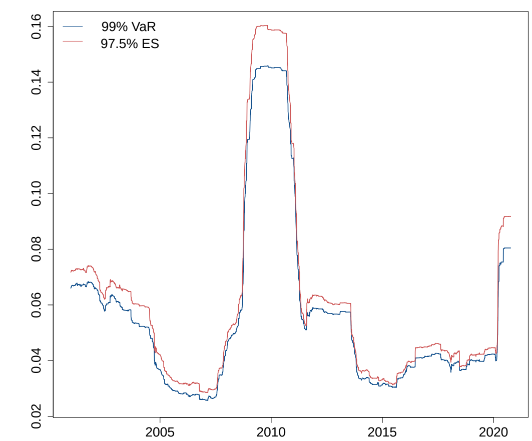
[20] The average 20-year log-return correlation within S&P 500 Financials is 0.4905, and the average log-return correlation within S&P 500 IT is 0.3616.



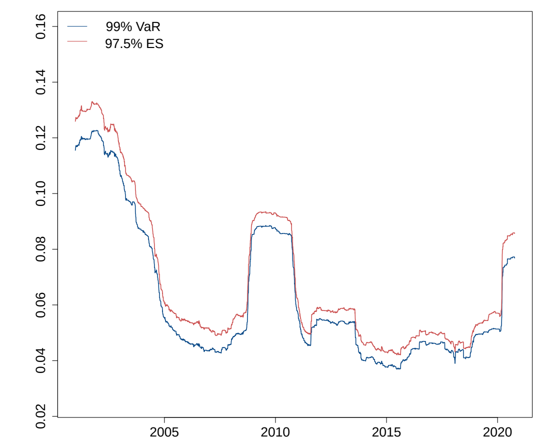
(a) Average PELVE of S&P 500 Financials



(b) Average PELVE of S&P 500 IT



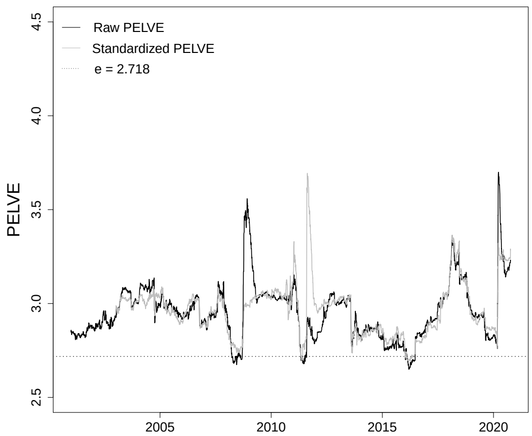
(c) Average VaR and ES of S&P 500 Financials



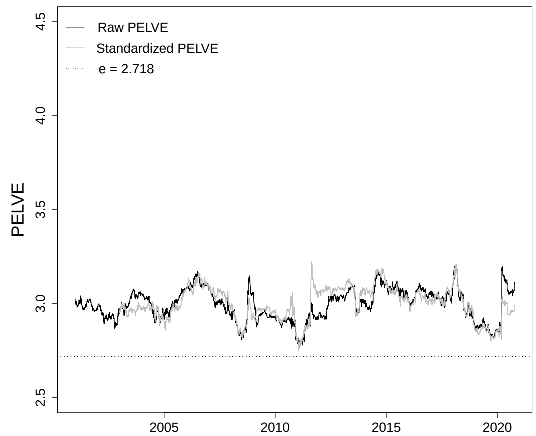
(d) Average VaR and ES of S&P 500 IT

Figure 4: Average 5% PELVE, 99% VaR and 97.5% ES

Raw log-loss data are used for the analysis above. To account for time-varying volatility commonly observed in financial data, one can use return series standardized by realized volatility, as studied by Andersen et al. (2007) and Todorov and Tauchen (2014). We present in Figures 3 (b) and 5 the S&P 500 PELVE estimates and sector-average PELVE estimates computed using daily returns standardized by realized volatility from a moving window of 500 days (for other sectors, see Figure 10 in Appendix D). At time t, we compute standardized return as $r_t' = r_t/\sigma\{[r(\tau)]_{t-500\leqslant\tau\leqslant t-1}\}$, where r_t is the unstandardized log-return at time t. Thanks to the scale-invariance property of PELVE, we can directly compare PELVE of r_t with that of r_t'. Both the standardized PELVE estimates for Financials and IT are very close to the raw counterparts. As the standardized PELVE estimates still fall under the heavy-tailed region, our main qualitative observations remain unchanged.



(a) Average PELVE of S&P 500 Financials



(b) Average PELVE of S&P 500 IT

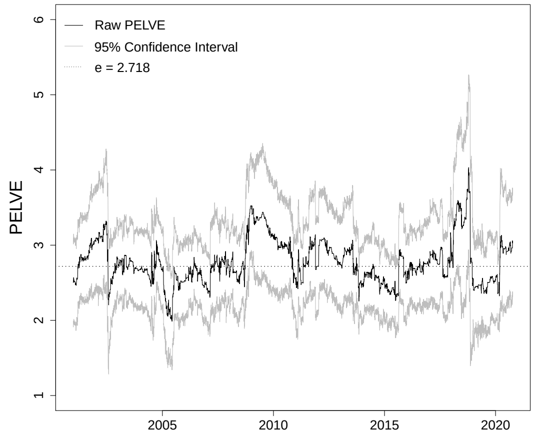
Figure 5: Comparison of standardized and raw 5% PELVE

Finally, we report in Figure 6 the PELVE of residuals from the fitted AR(1)-GARCH(1,1) model with t-innovations. Similarly to PELVE of the empirical log-losses, the residuals from the S&P 500 index have PELVE values around $e \approx 2.72$, whereas the residuals from the individual sectors have PELVE values mostly in the heavy-tailed region.

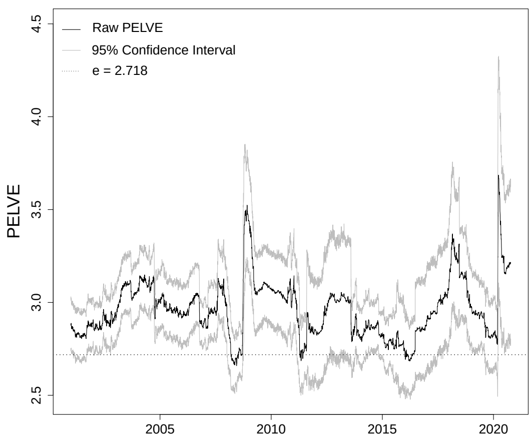
6.2 PELVE for well-diversified portfolios

Next, we analyze PELVE of well-diversified portfolios instead of individual assets, as discussed briefly in Section 3.3. A classic method of constructing a well-diversified portfolio is to use equal weights on a collection of N assets, which is known as the $1/N$ portfolio (e.g. DeMiguel et al. (2009)). Using this idea, we construct a monthly rebalanced equally weighted portfolio invested in $N = 500$ constituents of S&P 500 from Jan 4, 1999 to Oct 9, 2020.[21] More precisely, the portfolio with value process $(V_t)_{t=0,1,\dots}$ is constructed by investing the amount V_0/N in each stock at month 0 and rebalance the portfolio weights to V_t/N at the end of each month t. Since the constituents of S&P 500 change over time, we consider two types of $1/N$ portfolios. Portfolio A replaces delisted constituents with new constituents, and thus every month it rebalances among all current constituents ($N = 500$ always). Portfolio B rebalances only within constituents remained in the index since Jan 1999 (N decreases over

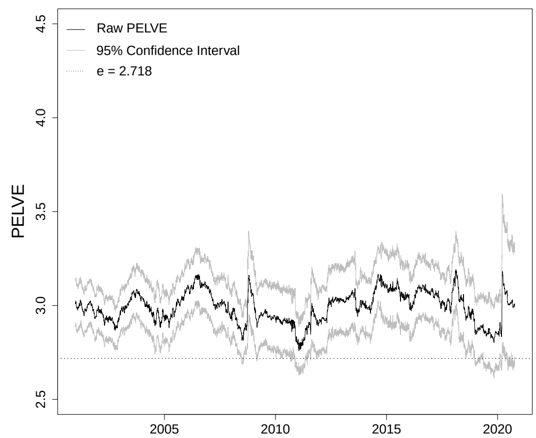
[21] The dynamic S&P constituents list is obtained from Bloomberg.



(a) PELVE of the S&P 500 index residuals



(b) Average PELVE of S&P 500 Financials residuals



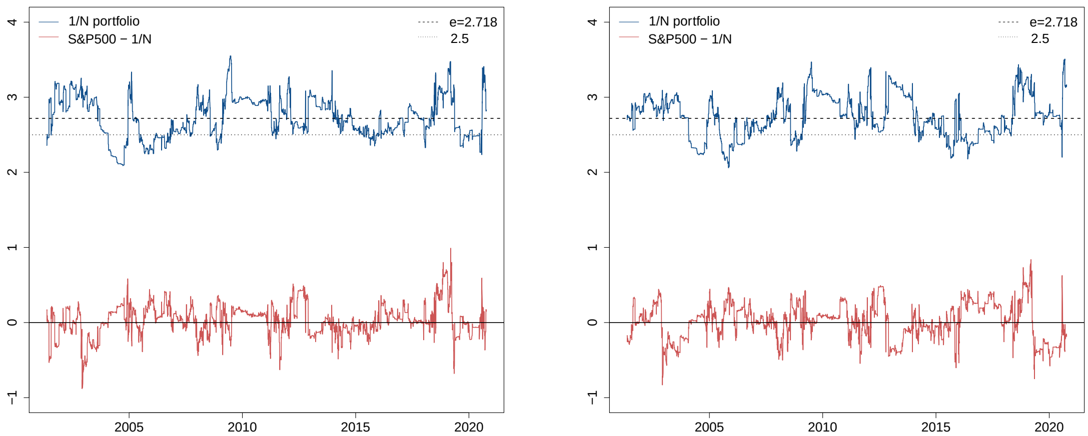
(c) Average PELVE of S&P 500 IT residuals

Figure 6: Bootstrap confidence bands for 5% PELVE of AR(1)-GARCH(1,1) residuals

time to 183). Clearly, both portfolios are valid portfolio strategies without survival bias. Note that the S&P 500 index itself can also be seen as a diversified portfolio but with survival bias.

Using the same method for asset return data in Section 6.1, we report in Figure 7 the PELVE values of the two $1/N$ portfolios during the period 2001 - 2020 and the difference between these PELVE values and those of the S&P 500 index. A summary of data is reported in Table 2. We discover that the PELVE curves of the $1/N$ portfolios fluctuate around $e \approx 2.72$, similar to that of S&P 500. Recall that PELVE values for individual assets are usually between 2.8 and 3.5 with an average of 2.98 across all S&P 500 constituents. This shows that well-diversified portfolios have a much smaller PELVE in general, often quite close to e.[22]

[22]From Table 2, we see that both $1/N$ portfolios have much higher return than S&P during the past 20 years. The combination of high return and low PELVE of the $1/N$ portfolios is consistent with its excellent



(a) $1/N$ portfolio with replacement (A) (b) $1/N$ portfolio without replacement (B)

Figure 7: 5% PELVE of $1/N$ portfolios (Jan 02, 2001 - October 9, 2020)

	mean PELVE	mean log-return	# stocks
S&P 500	2.76	4.06%	500 (always)
$1/N$ with repl. (A)	2.73	8.13%	500 (always)
$1/N$ without repl. (B)	2.75	8.96%	183 (Oct 2020)

Table 2: 20-year summary of the $1/N$ portfolios and S&P 500

The above observation has a profound implication for portfolio risk management. For a "good" (in the sense of diversification) portfolio, switching from 99% VaR to 97.5% ES is not going to cause a drastic change in its regulatory capital requirement, whereas the same statement does not hold for an arbitrary (or "bad") portfolio or an individual asset. In other words, if a well-diversified portfolio and a non-diversified one have the same $\mathrm{VaR}_{0.99}$, then the well-diversified one will have a smaller $\mathrm{ES}_{0.975}$. Thus, the use of ES encourages and rewards portfolio diversification more than the use of VaR, a latent feature of the current regulatory transition from VaR to ES.

As explained in Section 3.3, a theoretical explanation of the above observation may be that a well-diversified portfolio is more Gaussian-like, and hence its PELVE is closer to the corresponding Gaussian value (≈ 2.5 for $\varepsilon = 0.05$). Note, however, that typically daily log-returns of individual assets are highly correlated due to systematic risk (see e.g., McNeil et al. (2015, Chapter 3)), and normality cannot be expected for daily returns of either individual

performance well documented in the literature (e.g. DeMiguel et al. (2009) and Guo et al. (2019)).

assets or well-diversified portfolios.[23]

7 Concluding remarks

The new risk management tool PELVE that we develop in this paper is motivated by the recent transition from VaR to ES instructed by the Basel FRTB. According to our analysis of PELVE for various loss distributions, the new capital requirement based on 97.5% ES will increase for heavy-tailed returns where it will remain almost the same if returns are normally distributed or have lighter tails. Applied to financial asset and portfolio data, we discovered that the PELVE of individual stock returns in all sectors of S&P 500 during the period 2001 - 2020 lies above $e \approx 2.72$, the threshold which distinguishes heavy tails and the light ones. This suggests that the underlying loss distribution of US equity market appears to be heavy-tailed, and hence the capital requirement for single-asset portfolios under a 97.5% ES will be higher comparing to that under a 99% VaR. Intriguingly, the loss distribution of an equally weighted $(1/N)$ portfolio has a much smaller PELVE (close to e) compared all other financial data sets that we considered. This observation suggests that for a well-diversified portfolio, changing from 99% VaR to 97.5% ES does not lead to a significantly higher capital requirement. Therefore, the implementation of 97.5% ES rewards portfolio diversification more than 99% VaR which was used in the previous regulatory framework.

Our contributions in this paper expand far beyond the context of banking regulation, as risk measures are widely used in many quantitative areas of economics and management. PELVE enjoys many nice theoretical properties, and in particular, it is location-scale invariant, decreasing under increasing concave transformation of the loss, and quasi-convex/concave for comonotonic risks. Proper diversification techniques often reduce PELVE, and there is a monotonic relationship between the limit of PELVE and the tail index of a regularly varying distribution. Unlike measures of tail heaviness such as the tail-index, PELVE is well-defined for all commonly used statistical models with finite mean, including bounded distributions, without requiring asymptotic properties of the distribution functions, which can be difficult to verify. Empirical estimation of PELVE is standard and as simple as estimating VaR and ES,

[23] On a related note, the sum of iid heavy tailed losses with tail index α has the same tail index α, whereas its PELVE (at a fixed ε) will decrease due to the fact that the sum is more Gaussian-like (for $\alpha > 2$, which is the case of financial return data). This shows that PELVE is able to capture portfolio diversification whereas the tail index cannot.

and the asymptotic normality of the estimators is established under minimal assumptions.

There are several relevant questions of PELVE that lead to future research. We outline a few directions below.[24]

(i) **Implications of PELVE in backtesting ES.** Since VaR is simpler to backtest than ES, in practice one often backtests the quality of VaR forecasts from a predictive model instead of directly backtesting ES; see e.g., Emmer et al. (2015, Section 5.1). If relatively accurate and time-stable daily ε-PELVE estimates are available, then one can backtest $\mathrm{VaR}_{1-\varepsilon}$ daily forecasts to approximately backtest $\mathrm{ES}_{1-c\varepsilon}$ daily forecasts, where c is the estimated ε-PELVE. This approach may have advantages over backtesting via averaging VaR forecasts; a thorough study is needed for concrete conclusions.

(ii) **PELVE for residuals.** Asymptotic normality in Theorem 4 and Proposition 4 is obtained for raw data. For financial data, one usually implements time-series models. Hence, it is important to understand the asymptotic distribution of PELVE for residuals. In our empirical study, we observe from Figure 6 that the PELVE values of residuals and those of the unconditional distributions are quite similar. A comprehensive theory on PELVE of residuals is left for future studies.

(iii) **Measures of tail heaviness and skewness.** As illustrated by Theorem 1 (ii)-(iii), PELVE satisfies an axiom for measures of (right) skewness of Oja (1981). As such, PELVE can be seen as either a measure of tail heaviness or a measure of skewness. A thorough comparison of PELVE with other measures of tail heaviness and measures of skewness may reveal features of PELVE relative to the existing measures.

(iv) **Characterization of a distribution family.** We wonder whether the PELVE curve $\varepsilon \mapsto \Pi_\varepsilon(X)$ characterizes the distribution of X up to location-scale transforms. For instance, if $\Pi_\varepsilon(X) = e$ for all $\varepsilon \in [0, 1/e]$, does X necessarily follow an exponential distribution up to location-scale transforms? A related question is how to generate a distribution model from a given PELVE curve. These questions turn out to be very challenging, and an answer would help to understand the fundamental role of PELVE in distribution models.

[24]Directions (i)-(iii) are inspired by comments from anonymous referees.

Acknowledgements

The authors thank Hirbod Assa, Paul Embrechts, Paul Glasserman, Mary Hardy, Liyuan Lin, Tiantian Mao, Stephen Mildenhall, Alfred Müller, Liang Peng, Greg Rice, Andreas Tsanakas, Viktor Todorov, Dan Wang, an Editor, an Associate Editor, and two anonymous referees for helpful comments and discussions on an earlier version of the paper. Wang acknowledges financial support from the Natural Sciences and Engineering Research Council of Canada (RGPIN-2018-03823, RGPAS-2018-522590).

References

Andersen, T., Bollerslev, T. and Dobrev, D. (2007). No-arbitrage semi-martingale restrictions for continuous-time volatility models subject to leverage effects, jumps and i.i.d. noise: Theory and testable distributional implications. *Journal of Econometrics*, **138**(1), 125–180.

Andersen, T., Todorov, V. and Ubukata, M. (2021). Tail risk and return predictability for the Japanese equity market. *Journal of Econometrics*, **222**(1), 344–363.

Artzner, P., Delbaen, F., Eber, J.-M. and Heath, D. (1999). Coherent measures of risk. *Mathematical Finance*, **9**(3), 203–228.

Asimit, V., Peng, L., Wang, R. and Yu, A. (2019). An efficient approach to quantile capital allocation and sensitivity analysis. *Mathematical Finance*, **29**(4), 1131–1156.

BCBS (2013). *Consultative Document October 2013. Fundamental review of the trading book: A revised market risk framework.* Basel Committee on Banking Supervision. Basel: Bank for International Settlements. BIS online publication No. bcbs265.

BCBS (2015). *Fundamental review of the trading book – interim impact analysis. November 2015.* Basel Committee on Banking Supervision. Basel: Bank for International Settlements. BIS online publication No. d346.

BCBS (2016). *Minimum Capital Requirements for Market Risk. January 2016.* Basel Committee on Banking Supervision. Basel: Bank for International Settlements. BIS online publication No. d352.

BCBS (2019). *Minimum Capital Requirements for Market Risk. February 2019.* Basel Committee on Banking Supervision. Basel: Bank for International Settlements. BIS online publication No. d457.

Berghaus, B., Bücher, A. and Volgushev, S. (2017). Weak convergence of the empirical copula process with respect to weighted metrics. *Bernoulli*, **23**(1), 743–772.

Billingsley, P. (1999). *Convergence of Probability Measures*, Second Edition, Wiley.

Bühlmann, P. (1994). Blockwise bootstrapped empirical process for stationary sequences. *Annals of Statistics*, **22**(2), 995–1012.

Carlstein, E. (1986). The use of subseries values for estimating the variance of a general statistic from a stationary sequence. *Annals of Statistics*, **14**(3), 1171–1179.

Chen, Y., Wang, Z. and Zhang, Z. (2019). Mark to market value at risk. *Journal of Econometrics*, **208**(1), 299–321.

Chew, H. S. (1983). A generalization of the quasilinear mean with applications to the measurement of income inequality and decision theory resolving the Allais paradox. *Econometrica*, **51**(4), 1065–1092.

Cont, R. (2001). Empirical properties of asset returns: stylized facts and statistical issues. *Quantitative Finance*, **1**, 223–236.

Daníelsson, J., and Zhou, C. (2016). Why risk is so hard to measure. `SSRN:2597563`.

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

DeMiguel, V., Garlappi, L. and Uppal, R. (2009). Optimal versus naive diversification: How inefficient is the 1/N portfolio strategy? *The Review of Financial Studies*, **22**(5), 1915–1953.

Du, Z. and Escanciano, J. C. (2017). Backtesting expected shortfall: Accounting for tail risk. *Management Science*, **63**(4), 940–958.

Embrechts, P., Klüppelberg, C. and Mikosch, T. (1997). *Modelling Extremal Events for Insurance and Finance*. Springer, Heidelberg.

Embrechts, P., Liu, H. and Wang, R. (2018). Quantile-based risk sharing. *Operations Research*, **66**(4), 936–949.

Embrechts, P., Wang, B. and Wang, R. (2015). Aggregation-robustness and model uncertainty of regulatory risk measures. *Finance and Stochastics*, **19**(4), 763–790.

Embrechts, P. and Wang, R. (2015). Seven proofs for the subadditivity of Expected Shortfall. *Dependence Modeling*, **3**, 126–140.

Fissler, T. and Ziegel, J. F. (2016). Higher order elicitability and Osband's principle. *Annals of Statistics*, **44**(4), 1680–1707.

Emmer, S., Kratz, M. and Tasche, D. (2015). What is the best risk measure in practice? A comparison of standard measures. *Journal of Risk*, **18**(2), 31–60.

Grabchak, M. and Samorodnitsky, G. (2010). Do financial returns have finite or infinite variance? A paradox and an explanation. *Quantitative Finance*, **10**(8), 883–893.

Guo, D., Boyle, P. P., Weng, C. and Wirjanto, T. S. (2019). Age matters. `SSRN:3363360`.

Hall, P., Horowitz, J. and Jing, B. (1995). On blocking rules for the bootstrap with dependent data. *Biometrika*, **82**(3), 561–574.

Jansen, D. W. and De Vries, C. G. (1991). On the frequency of large stock returns: Putting booms and busts into perspective. *Review of Economics and Statistics*, **73**(1), 18–24.

Kiwiel, K. C. (2001). Convergence and efficiency of subgradient methods for quasiconvex minimization. *Mathematical Programming Series A*, **90**(1), 1–25.

Lahiri, S. N. (2003). *Resampling Methods for Dependent Data*. Springer-Verlag, New York.

Li, L. and Xing, H. (2019). Capital allocation under the Fundamental Review of Trading Book. *Risk Magazine*, 06 Jun 2019. `arXiv:1801.07358v1`.

Mafusalov, A. and Uryasev, S. (2018). Buffered probability of exceedance: Mathematical properties and optimization. *SIAM Journal on Optimization*, **28**(2), 1077–1103.

McNeil, A. J., Frey, R. and Embrechts, P. (2015). *Quantitative Risk Management: Concepts, Techniques and Tools*. Revised Edition. Princeton, NJ: Princeton University Press.

Oja, H. (1981). On location, scale, skewness and kurtosis of univariate distributions. *Scandinavian Journal of Statistics*, **8**(3), 154–168.

Patton, A., Ziegel, J. F. and Chen, R. (2019). Dynamic semiparametric models for Expected Shortfall (and Value-at-Risk). *Journal of Econometrics*, **211**, 388–413.

Shao, Q. M. and Yu, H. (1996). Weak convergence for weighted empirical processes of dependent sequences. *Annals of Probability*, **24**(4), 2098–2127.

Todorov, V. and Tauchen, G. (2014). Limit theorems for the empirical distribution function of scaled increments of Itô semimartingales at high frequencies. *Annals of Applied Probability*, **24**(5), 1850–1888.

Wang, R. and Zitikis, R. (2021). An axiomatic foundation for the Expected Shortfall. *Management Science*, **67**, 1413–1429.

Yaari, M. E. (1987). The dual theory of choice under risk. *Econometrica*, **55**(1), 95–115.

Appendices

A Technical details

A.1 Proofs for results in Section 2

Proof of Proposition 1. We use the simple fact that, for any random variable $X \in L^1$, $\mathrm{ES}_t(X)$ is continuous and increasing in $t \in (0,1)$. This fact can be directly checked by definition of ES. We will prove the proposition in the order (i)⇒(ii)⇒(iii)⇒(iv)⇒(i).

1. (i)⇒(ii): As mentioned above, $\mathrm{ES}_t(X)$ is continuous and increasing in $t \in (0,1)$. Therefore, the infimum of c satisfying $\mathrm{ES}_{1-c\varepsilon}(X) \leqslant \mathrm{VaR}_{1-\varepsilon}(X)$, assuming such c exists, satisfies $\mathrm{ES}_{1-c\varepsilon}(X) = \mathrm{VaR}_{1-\varepsilon}(X)$.

2. (ii)⇒(iii): From the fact that $\mathrm{ES}_t(X)$ is increasing in $t \in (0,1)$, we have $\mathbb{E}[X] = \inf_{c\in[1,1/\varepsilon]} \mathrm{ES}_{1-c\varepsilon}(X)$. If $\mathbb{E}[X] > \mathrm{VaR}_{1-\varepsilon}(X)$, then $\inf_{c\in[1,1/\varepsilon]} \mathrm{ES}_{1-c\varepsilon}(X) > \mathrm{VaR}_{1-\varepsilon}(X)$, making $\Pi_\varepsilon(X) = \infty$. Therefore, $\Pi_\varepsilon(X) < \infty$ implies $\mathbb{E}[X] \leqslant \mathrm{VaR}_{1-\varepsilon}(X)$.

3. (iii)⇒(iv): $\mathbb{E}[X] \leqslant \mathrm{VaR}_{1-\varepsilon}(X)$ implies that $\{c \in [1,1/\varepsilon] : \mathrm{ES}_{1-c\varepsilon}(X) \leqslant \mathrm{VaR}_{1-\varepsilon}(X)\}$ is non-empty, and hence $\Pi_\varepsilon(X) < \infty$.

4. (iv)⇒(i): Since $\mathrm{ES}_{1-\varepsilon}(X) \geqslant \mathrm{VaR}_{1-\varepsilon}(X)$ and $\mathrm{ES}_0(X) \leqslant \mathrm{VaR}_{1-\varepsilon}(X)$, by the Intermediate Value Theorem, there exists $c \in [1,1/\varepsilon]$ such that (1) holds. □

Proof of Proposition 2. Existence of c is given by Proposition 1. For the uniqueness of c, it suffices to notice that $\mathrm{ES}_t(X)$, as the average of $F_X^{-1}(p)$ for $p \in [t,1)$, is a strictly increasing function of $t \in [0,1-\varepsilon)$ if $F_X^{-1}(p)$ is not a constant on $(1-\varepsilon,1)$. □

A.2 Proofs for results in Section 3

Proof of Theorem 1. We first show that (i) is implied by (ii). Let $f:\mathbb{R}\to\mathbb{R}$, $x \mapsto \lambda x + a$ and $g:\mathbb{R}\to\mathbb{R}$, $y \mapsto (y-a)/\lambda$, which is the inverse of f. Both f and g are increasing linear functions, which are concave. Hence, if (ii) holds, we have

$$\Pi_\varepsilon(f(X)) \leqslant \Pi_\varepsilon(X) = \Pi_\varepsilon(g \circ f(X)) \leqslant \Pi_\varepsilon(f(X)).$$

Therefore, $\Pi_\varepsilon(f(X)) = \Pi_\varepsilon(X)$ and (i) holds.

Next, we show (ii), which implies (i) by the above argument. Note that $\mathrm{VaR}_p(h(Y)) = h(\mathrm{VaR}_p(Y))$ for all continuous and increasing function h, random variable Y and $p \in (0,1)$, and concave functions on $\mathbb{R}$ are continuous. Hence,

$$\mathrm{VaR}_p(f(Y)) = f(\mathrm{VaR}_p(Y)). \tag{A.1}$$

Write $c_1 = \Pi_\varepsilon(X)$ and $c_2 = \Pi_\varepsilon(f(X))$. If $c_1 = \infty$ then there is nothing to show. Next, assume c_1 is finite, which, by Proposition 1, implies

$$\mathrm{VaR}_{1-\varepsilon}(X) = \mathrm{ES}_{1-c_1\varepsilon}(X). \tag{A.2}$$

Using Jensen's inequality, (A.1) and (A.2), we have

$$\mathbb{E}[f(X)] \leqslant f(\mathbb{E}[X]) \leqslant f(\mathrm{ES}_{1-c_1\varepsilon}(X)) = f(\mathrm{VaR}_{1-\varepsilon}(X)) = \mathrm{VaR}_{1-\varepsilon}(f(X)).$$

Using Proposition 1 again, we obtain that c_2 is finite.

For any $p \in (0,1)$, we can write ES_p as its dual representation (see e.g., Lemma 3.14 of Embrechts and Wang (2015)), that is, for some set of probability measures $\mathcal{Q}$,

$$\mathrm{ES}_p(Y) = \sup_{Q \in \mathcal{Q}} \mathbb{E}^Q[Y], \quad Y \in L^1.$$

By Jensen's inequality, we have, for all $p \in (0,1)$,

$$\mathrm{ES}_p(f(X)) = \sup_{Q \in \mathcal{Q}} \mathbb{E}^Q[f(X)] \leqslant \sup_{Q \in \mathcal{Q}} f(\mathbb{E}^Q[X]) = f\left(\sup_{Q \in \mathcal{Q}} \mathbb{E}^Q[X]\right) = f(\mathrm{ES}_p(X)).$$

Using the above inequality, (A.1) and (A.2), we get

$$\mathrm{ES}_{1-c_1\varepsilon}(f(X)) \leqslant f(\mathrm{ES}_{1-c_1\varepsilon}(X)) = f(\mathrm{VaR}_{1-\varepsilon}(X)) = \mathrm{VaR}_{1-\varepsilon}(f(X)). \tag{A.3}$$

By definition of PELVE, c_2 is the smallest value c such that

$$\mathrm{ES}_{1-c\varepsilon}(f(X)) \leqslant \mathrm{VaR}_{1-\varepsilon}(f(X)).$$

Hence, (A.3) implies $c_2 \leqslant c_1$, as desired.

For a proof of (iii), it suffices to note that, for a strictly increasing convex function g, its inverse f is increasing and concave. Applying the result in (ii), we get

$$\Pi_\varepsilon(X) = \Pi_\varepsilon(f \circ g(X)) \leqslant \Pi_\varepsilon(g(X)).$$

Finally, we prove (iv). Note that for any $c, d \in [1, 1/\varepsilon]$ and any random variable Y, by definition and noting that ES_p is continuous in p, we have

$$d < \Pi_\varepsilon(Y) \leqslant c \iff \mathrm{ES}_{1-c\varepsilon}(Y) \leqslant \mathrm{VaR}_{1-\varepsilon}(Y) < \mathrm{ES}_{1-d\varepsilon}(Y). \tag{A.4}$$

Moreover, $\lambda f(X)$ and $(1-\lambda)g(X)$ are comonotonic, and VaR is comonotonic-additive (see e.g., Section 8.2 of McNeil et al. (2015)), implying

$$\mathrm{VaR}_{1-\varepsilon}(\lambda f(X)+(1-\lambda)g(X))=\lambda\mathrm{VaR}_{1-\varepsilon}(f(X))+(1-\lambda)\mathrm{VaR}_{1-\varepsilon}(g(X)). \tag{A.5}$$

Let $c_1=\Pi_\varepsilon(f(X))$ and $c_2=\Pi_\varepsilon(g(X))$. First, if $c_1=c_2=\infty$, then it suffices to prove

$$\Pi_\varepsilon(\lambda f(X)+(1-\lambda)g(X))=\infty.$$

By Proposition 1, this is equivalent to

$$\mathbb{E}[\lambda f(X)+(1-\lambda)g(X)]>\mathrm{VaR}_{1-\varepsilon}(\lambda f(X)+(1-\lambda)g(X)),$$

which is guaranteed by (A.5) and the fact (obtained also from Proposition 1) that

$$\mathbb{E}[f(X)]>\mathrm{VaR}_{1-\varepsilon}(f(X)) \quad \text{and} \quad \mathbb{E}[g(X)]>\mathrm{VaR}_{1-\varepsilon}(g(X)).$$

Next, suppose that at least one of c_1 and c_2 is finite. For any $d<\min\{c_1,c_2\}$, by (A.4), we have

$$\mathrm{VaR}_{1-\varepsilon}(f(X))<\mathrm{ES}_{1-d\varepsilon}(f(X)), \quad \text{and} \quad \mathrm{VaR}_{1-\varepsilon}(g(X))<\mathrm{ES}_{1-d\varepsilon}(g(X)). \tag{A.6}$$

Together with (A.5) and noting that ES is also comonotonic-additive, (A.6) implies

$$\mathrm{VaR}_{1-\varepsilon}(\lambda f(X)+(1-\lambda)g(X))<\mathrm{ES}_{1-d\varepsilon}(\lambda f(X)+(1-\lambda)g(X)).$$

Using (A.4) again, we arrive at

$$\Pi_\varepsilon(\lambda f(X)+(1-\lambda)g(X))>d.$$

Since $d<\min\{c_1,c_2\}$ is arbitrary, we obtain

$$\Pi_\varepsilon(\lambda f(X)+(1-\lambda)g(X))>\min\{c_1,c_2\}. \tag{A.7}$$

If $\max\{c_1,c_2\}=\infty$, then (A.7) is sufficient for the conclusion in (iv). If $\max\{c_1,c_2\}<\infty$, by taking $c=\max\{c_1,c_2\}$, using a similar argument via (A.4), we have

$$\mathrm{ES}_{1-c\varepsilon}(f(X))\leqslant\mathrm{VaR}_{1-\varepsilon}(f(X)) \quad \text{and} \quad \mathrm{ES}_{1-c\varepsilon}(g(X))\leqslant\mathrm{VaR}_{1-\varepsilon}(g(X)).$$

As a consequence,

$$\mathrm{ES}_{1-c\varepsilon}(\lambda f(X)+(1-\lambda)g(X))\leqslant\mathrm{VaR}_{1-\varepsilon}(\lambda f(X)+(1-\lambda)g(X)).$$

Using (A.4) again, we obtain

$$\Pi_\varepsilon(\lambda f(X)+(1-\lambda)g(X)) \leqslant c = \max\{c_1,c_2\}. \tag{A.8}$$

Thus we obtain (iv). $\square$

Details of the assertion in Example 1. Here we give a detailed derivation of the assertion in Example 1 that X_C, X_{AC}, X_{AP} are all increasing concave functions of X_A, and X_{AP} is an increasing concave function of X_{AC}, regardless of any model assumptions.

Let S_t be the price of an asset S at time t in the future, S_0 be its current price, and $X_A = S_0 - S_t$ represents the loss random variable from investment in one unit of S. Consider a European put option with current price P_0, time-t payoff $(K-S_t)_+$ and some strike price K, and a European call option with current price C_0, time-t payoff $(S_t-K)_+$ and the same strike. We analyze the following four portfolios in Example 1.

(a) Long one unit of the asset, with loss X_A.

(b) Long a call option, with loss $X_C = C_0 - (S_t-K)_+ = C_0 - (S_0-K-X_A)_+$.

(c) Long one unit of the asset and a call option, with loss $X_{AC} = X_A + C_0 - (S_0-K-X_A)_+$.

(d) Long one unit of the asset and a put option, with loss $X_{AP} = X_A + P_0 - (K-S_0+X_A)_+$.

First, X_C is an increasing concave functions of X_A. This implies $X_{AC} = X_A + X_C$ is also an increasing concave functions of X_A. Also note that

$$X_{AP} = X_A + P_0 - (K-S_0+X_A)_+ = \min\{S_0-K, X_A\} + P_0,$$

which is again an increasing concave function of X. Finally, let $c = S_0-K+C_0$, $d = S_0-K+P_0$, and

$$f(x) = \frac{1}{2}\min\{x,c\} + d - \frac{c}{2}, \quad x\in\mathbb{R},$$

and it is clear that f is an increasing concave function. We can check

$$\begin{aligned} f(X_{AC}) &= \frac{1}{2}\min\{X_{AC},c\} + d - \frac{c}{2} \\ &= \frac{c}{2}\mathbb{1}_{\{X_A\geqslant S_0-K\}} + \frac{2X_A+2C_0-c}{2}\mathbb{1}_{\{X_A<S_0-K\}} + d - \frac{c}{2} \\ &= d\mathbb{1}_{\{X_A\geqslant S_0-K\}} + (X+d+C_0-c)\mathbb{1}_{\{X_A<S_0-K\}} \\ &= d + (X_A+K-S_0)\mathbb{1}_{\{X_A<S_0-K\}} \\ &= \min\{S_0-K, X_A\} + P_0 = X_{AP}, \end{aligned}$$

which shows that X_{AP} is an increasing concave function of X_{AC}. $\square$

Proof of Theorem 2. By the convergence in distribution of $\{X_n\}_{n\in\mathbb{N}}$ to X, we know that $\mathrm{VaR}_t(X_n) \to \mathrm{VaR}_t(X)$ for all $t \in (0,1)$ at which the quantile function of X is continuous. Because $\{X_n\}_{n\in\mathbb{N}}$ is uniformly integrable, $\mathrm{ES}_t(X_n) \to \mathrm{ES}_t(X)$ as $n \to \infty$ for all $t \in [0,1)$. As a consequence, the functions $f_n : [0,1) \to \mathbb{R}$, $t \mapsto \mathrm{ES}_t(X_n) - \mathrm{VaR}_{1-\varepsilon}(X_n)$ converge to the function $f : t \mapsto \mathrm{ES}_t(X) - \mathrm{VaR}_{1-\varepsilon}(X)$ as $n \to \infty$ for each $t \in (0,1)$. Note that both f_n and f are increasing and continuous functions, and f is strictly increasing. This means that on the compact interval $[0, 1-\varepsilon]$, $f_n \to f$ uniformly.

By Assumption 1, we know that for n sufficiently large, $\mathbb{E}[X_n] < \mathrm{VaR}_{1-\varepsilon}(X_n)$, and hence $\Pi_\varepsilon(X_n) \in [1, 1/\varepsilon]$ by Proposition 1. Thus, $\Pi_\varepsilon(X_n)$ is the left-most root c of the equation $f_n(1-c\varepsilon) = 0$, and $\Pi_\varepsilon(X)$ is the unique root c of the equation $f(1-c\varepsilon) = 0$. Since $f_n \to f$ uniformly, elementary analysis yields that $\Pi_\varepsilon(X_n) \to \Pi_\varepsilon(X)$ as $n \to \infty$. □

A.3 Proofs for results in Section 4

VaR and ES formulas in Examples 5 and 6. Here we collect some VaR_{1-t} and ES_{1-t} formulas for $t \in (0,1)$, which, plugged in (1), give PELVE formulas and numerical values in Examples 5 and 6. These formulas are available by direct calculation, or one may see Section 2.3 of McNeil et al. (2015).

1. $\underline{X \sim \mathrm{U}[0,1]}$: $\mathrm{VaR}_{1-t}(X) = 1 - t$ and $\mathrm{ES}_{1-t}(X) = (2-t)/2$.
2. $\underline{X \sim \mathrm{Pareto}(\alpha),\ \alpha > 1}$: $\mathrm{VaR}_{1-t}(X) = t^{-\alpha}$ and $\mathrm{ES}_{1-t}(X) = \frac{\alpha}{\alpha-1} t^{-\alpha}$.
3. $\underline{X \sim \mathrm{Exp}(1)}$: $\mathrm{VaR}_{1-t}(X) = -\log t$ and $\mathrm{ES}_{1-t}(X) = -\log t + 1$.
4. $\underline{X \sim \mathrm{N}(0,1)}$: $\mathrm{VaR}_{1-t}(X) = \Phi^{-1}(1-t)$ and $\mathrm{ES}_{1-t}(X) = \frac{1}{t}\phi(\Phi^{-1}(1-t))$, where ϕ and Φ are the pdf and cdf of a standard normal distribution, respectively.
5. $\underline{X \sim \mathrm{t}(\nu),\ \nu > 1}$: $\mathrm{VaR}_{1-t}(X) = G_\nu^{-1}(1-t)$ and $\mathrm{ES}_{1-t}(X) = \frac{1}{t} g_\nu(G_\nu^{-1}(1-t))(\frac{\nu + (G_\nu^{-1}(1-t))^2}{\nu-1})$, where g_ν and G_ν are the pdf and cdf of $\mathrm{t}(\nu)$, respectively.
6. $\underline{X \sim \mathrm{LN}(\sigma),\ \sigma > 0}$: $\mathrm{VaR}_{1-t}(X) = e^{\sigma \Phi^{-1}(1-t)}$ and $\mathrm{ES}_{1-t}(X) = e^{\frac{\sigma^2}{2}} \frac{1}{t}\Phi(\sigma - \Phi^{-1}(1-t))$.

Since PELVE is invariant in a location-scale family, the above values also give PELVE for other distributions in the above families. □

Proof of Theorem 3. We shall use some classic results in EVT, which can be found in Embrechts et al. (1997). First, the inverse of an RV function with index $\xi \neq 0$ is again RV, and has index $1/\xi$. Thus, for $t > 0$,

$$\lim_{\varepsilon \downarrow 0} \frac{\mathrm{VaR}_{1-t\varepsilon}(X)}{\mathrm{VaR}_{1-\varepsilon}(X)} = t^{-1/\alpha}. \tag{A.9}$$

The Karamata Theorem for VaR/ES gives

$$\lim_{\varepsilon \downarrow 0} \frac{\mathrm{ES}_{1-\varepsilon}(X)}{\mathrm{VaR}_{1-\varepsilon}(X)} = \frac{\alpha}{\alpha - 1}. \tag{A.10}$$

Combining (A.9) and (A.10), we obtain, for $t > 0$,

$$\lim_{\varepsilon \downarrow 0} \frac{\mathrm{ES}_{1-t\varepsilon}(X)}{\mathrm{VaR}_{1-\varepsilon}(X)} = \frac{\alpha}{\alpha - 1} t^{-1/\alpha}. \tag{A.11}$$

Note that the right-hand side of (A.11), $\frac{\alpha}{\alpha-1}t^{-1/\alpha}$, is strictly decreasing in t, and $\frac{\alpha}{\alpha-1}t^{-1/\alpha} = 1$ if and only if $t = (\frac{\alpha}{\alpha-1})^\alpha$. Take arbitrary $t_1 < (\frac{\alpha}{\alpha-1})^\alpha$ and $t_2 > (\frac{\alpha}{\alpha-1})^\alpha$. By (A.11), we have, for ε close enough to 0,

$$\mathrm{ES}_{1-t_1\varepsilon}(X) > \mathrm{VaR}_{1-\varepsilon}(X) > \mathrm{ES}_{1-t_2\varepsilon}(X). \tag{A.12}$$

Note since the mean of X is finite, and $\mathrm{VaR}_{1-\varepsilon}(X) \to \infty$ as $\varepsilon \downarrow 0$, we know that Assumption 1 holds for ε close to 0. In this case, by Proposition 2, we know that $c = \Pi_\varepsilon(X) \in [1, 1/\varepsilon]$ satisfies (1). By (A.12), $\Pi_\varepsilon(X) \in (t_1, t_2)$. Since t_1 and t_2 are arbitrarily close to $(\frac{\alpha}{\alpha-1})^\alpha$, we know that $\Pi_\varepsilon(X) \to (\frac{\alpha}{\alpha-1})^\alpha$ as $\varepsilon \downarrow 0$. This completes the proof of the theorem. □

Proof of Proposition 3. This proposition is a direct consequence of (A.11) and Theorem 3. □

A.4 Proofs for results in Section 5

We will show Proposition 4 before showing Theorem 4, because the first statement of Theorem 4 is a special case of Proposition 4, by noting that the iid model is a special case of the α-mixing model, where W in Proposition 4 becomes a standard Brownian bridge.

Proof of Proposition 4. The proof is similar to Theorem 3.1 of Asimit et al. (2019). We only consider the case $\hat{c}_n = \widehat{\Pi}_\varepsilon(n)$ as the case $\hat{c}_n = \widetilde{\Pi}_\varepsilon(n)$ is analogous. We will use results on weighted empirical processes. Let $U_n(t) = F_n(F^{-1}(t))$ and $A_n(t) = \sqrt{n}(U_n(t) - t)$ for $t \in [0, 1]$ as in Shao and Yu (1996). Let W be a Gaussian process with mean 0 and covariance in (12). Under our assumptions, it follows from e.g., Theorem 2.2 of Shao and Yu (1996) that

$$A_n/\beta \text{ converges weakly in } \ell^\infty[0, 1] \text{ to } W/\beta \text{ as } n \to \infty \tag{A.13}$$

for any weight function β with $\beta(t) \geqslant C(t(1-t))^\delta$ for some $C > 0$ and $\delta \in (0, 1/2)$. Moreover, for any $\delta \in (0, 1/2)$ and $\lambda \in (0, 1)$,

$$\sup_{0<t<1} \frac{|A_n(t)|}{t^\delta(1-t)^\delta} = O_p(1), \qquad \sup_{1/n^\lambda < t \leqslant 1 - 1/n^\lambda} \frac{|\sqrt{n}(U_n^{-1}(t) - t)|}{t^\delta(1-t)^\delta} = O_p(1); \tag{A.14}$$

see e.g., Proposition 4.4 of Berghaus et al. (2017).

Recall that $p = 1 - \varepsilon$, $q = 1 - c\varepsilon$, and we write $q_n = 1 - \hat{c}_n\varepsilon$. By definition of PELVE, we have

$$F^{-1}(p) = \frac{1}{1-q}\int_q^1 F^{-1}(s)\mathrm{d}s = F^{-1}(q) + \frac{1}{1-q}\int_{F^{-1}(q)}^{\infty}(1-F(x))\mathrm{d}x. \tag{A.15}$$

and its empirical version

$$F_n^{-1}(p) = \frac{1}{1-q_n}\int_{q_n}^1 F_n^{-1}(s)\mathrm{d}s = F_n^{-1}(q) + \frac{1}{1-q_n}\int_{F_n^{-1}(q_n)}^{\infty}(1-F_n(x))\mathrm{d}x. \tag{A.16}$$

Using (A.15) and (A.16), we get

$$\begin{aligned}
0 &= \frac{1}{1-q_n}\int_{q_n}^1 F_n^{-1}(s)\mathrm{d}s - F_n^{-1}(p)\\
&= \underbrace{\frac{1}{1-q_n}\int_{q_n}^1 F_n^{-1}(s)\mathrm{d}s - \frac{1}{1-q}\int_q^1 F_n^{-1}(s)\mathrm{d}s}_{I_1} + \frac{1}{1-q}\int_q^1 F_n^{-1}(s)\mathrm{d}s - F_n^{-1}(p)\\
&= I_1 + \underbrace{\frac{1}{1-q}\int_{F_n^{-1}(q)}^{F^{-1}(q)}\big(1-F_n(x)\big)\mathrm{d}x}_{I_2} + \frac{1}{1-q}\int_{F^{-1}(q)}^{\infty}\big(1-F_n(x)\big)\mathrm{d}x + F_n^{-1}(q) - F_n^{-1}(p)\\
&= I_1 + I_2 + \frac{1}{1-q}\int_{F^{-1}(q)}^{\infty}\big(F(x)-F_n(x)\big)\mathrm{d}x + F^{-1}(p) - F^{-1}(q) + F_n^{-1}(q) - F_n^{-1}(p)\\
&= I_1 + I_2 + \underbrace{\frac{1}{1-q}\int_q^1\big(s-U_n(s)\big)\mathrm{d}F^{-1}(s)}_{I_3} + \underbrace{\big(F_n^{-1}(q)-F^{-1}(q)\big)}_{I_4} + \underbrace{\big(F^{-1}(p)-F_n^{-1}(p)\big)}_{I_5}.
\end{aligned}$$

Moreover, the mean-value theorem gives

$$I_1 = (q_n - q)\frac{-(1-\tilde{q})F_n^{-1}(\tilde{q}) + \int_{\tilde{q}}^1 F_n^{-1}(s)\mathrm{d}s}{(1-\tilde{q})^2}$$

for some $\tilde{q}$ between q and q_n. By (A.13) and (A.14), all terms $I_2, \ldots, I_5$ converge to 0 in probability as $n \to \infty$. We have $I_1 \to 0$ as well, since

$$I_1 + I_2 + I_3 + I_4 + I_5 = 0. \tag{A.17}$$

Hence, $q_n \xrightarrow{\mathrm{p}} q$ as $n \to \infty$. With the regularity condition in Assumption 4, this leads to

$$\begin{aligned}\frac{-(1-\tilde{q})F_n^{-1}(\tilde{q}) + \int_{\tilde{q}}^1 F_n^{-1}(s)\mathrm{d}s}{(1-\tilde{q})^2} &= \frac{-(1-q)F^{-1}(q) + \int_q^1 F^{-1}(s)\mathrm{d}s}{(1-q)^2} + o_p(1) \\ &= \frac{F^{-1}(p) - F^{-1}(q)}{1-q} + o_p(1). \end{aligned} \tag{A.18}$$

Using (A.14) and Assumption 4, we have

$$\sqrt{n}(I_2 + I_4) = \frac{\sqrt{n}}{1-q} \int_{F^{-1}(U_n^{-1}(q))}^{F^{-1}(q)} (1 - F(s) - (1-q))\mathrm{d}s = o_p(1). \tag{A.19}$$

Assumption 4 and (A.13) guarantee the (joint) convergence

$$\sqrt{n} I_3 \xrightarrow{\mathrm{d}} \frac{1}{1-q} \int_q^1 W(s)\mathrm{d}F^{-1}(s) \quad \text{and} \quad \sqrt{n} I_5 \xrightarrow{\mathrm{d}} -\frac{W(p)}{f(F^{-1}(p))}$$

where $\xrightarrow{\mathrm{d}}$ is convergence in distribution, and as a consequence,

$$\sqrt{n}(I_3 + I_5) \xrightarrow{\mathrm{d}} \frac{1}{1-q} \int_q^1 W(s)\mathrm{d}F^{-1}(s) - \frac{W(p)}{f(F^{-1}(p))}. \tag{A.20}$$

Putting (A.17), (A.18), (A.19) and (A.20) together, we obtain

$$\sqrt{n}(q_n - q) \xrightarrow{\mathrm{d}} \frac{1}{F^{-1}(p) - F^{-1}(q)} \left(\int_q^1 \frac{W(s)}{f(F^{-1}(s))}\mathrm{d}s - \frac{(1-q)W(p)}{f(F^{-1}(p))} \right),$$

and thus (11) holds. $\square$

Proof of Theorem 4. It suffices to show the second statement, that is, the formula (10) for the asymptotic variance σ^2. Recall that we set $a = c/f(F^{-1}(p))$, $b = F^{-1}(p) - F^{-1}(q)$ and $E_t = \int_{F^{-1}(t)}^{\infty} (1 - F(x))\mathrm{d}x = \mathbb{E}[(X - F^{-1}(t))_+]$ for $t \in (0,1)$. Using the covariance property of a Brownian bridge, that is, $\mathbb{E}[W(s)W(t)] = \min(s,t) - st$, $s, t \in [0,1]$, we get

$$\begin{aligned} b^2\sigma^2 &= \mathbb{E}\left[\left(\int_q^1 \frac{W(t)}{\varepsilon f(F^{-1}(t))}\mathrm{d}t - aW(p) \right)^2 \right] \\ &= \mathbb{E}\left[\left(\frac{1}{\varepsilon} \int_{F^{-1}(q)}^{\infty} W(F(x))\mathrm{d}x - aW(p) \right)^2 \right] \\ &= a^2(\varepsilon - \varepsilon^2) + \frac{1}{\varepsilon^2}\mathbb{E}\left[\left(\int_{F^{-1}(q)}^{\infty} W(F(x))\mathrm{d}x \right)^2 \right] - \frac{2a}{\varepsilon} \int_{F^{-1}(q)}^{\infty} \mathbb{E}[W(p)W(F(x))]\mathrm{d}x. \end{aligned}$$

Direct computation gives

$$\mathbb{E}\left[\left(\int_{F^{-1}(q)}^{\infty} W(F(x))\mathrm{d}x\right)^2\right] = 2\int_{F^{-1}(q)}^{\infty}\int_{x}^{\infty}(1-F(y))F(x)\mathrm{d}y\mathrm{d}x = 2\int_{F^{-1}(q)}^{\infty} E_{F(x)}F(x)\mathrm{d}x$$

and

$$\begin{aligned}\int_{F^{-1}(q)}^{\infty}\mathbb{E}[W(p)W(F(x))]\mathrm{d}x &= p\int_{F^{-1}(p)}^{\infty}(1-F(x))\mathrm{d}x - \varepsilon\int_{F^{-1}(q)}^{F^{-1}(p)} F(x)\mathrm{d}x \\ &= E_p + \varepsilon(E_q - b).\end{aligned}$$

Therefore,

$$b^2\sigma^2 = a^2(\varepsilon-\varepsilon^2) + \frac{2}{\varepsilon^2}\int_{F^{-1}(q)}^{\infty} E_{F(x)}F(x)\mathrm{d}x - \frac{2a}{\varepsilon}E_p + 2a(E_q - b), \tag{A.21}$$

and we obtain the σ^2 formula (10). □

Proof of Proposition 5. In the proof of Proposition 4, we have seen that the convergence of the empirical process $A_n \to W$ in (A.13) leads to the asymptotic result (11). By Theorem 4.3 of Lahiri (2003), the bootstrap empirical process, denoted by A_n^*, also converges to W weakly. Therefore, using the same arguments as in the proof of Proposition 4 gives that the bootstrap empirical estimator has the same asymptotic variance σ^2. □

A.5 Two examples of the asymptotic variance in the iid case

We first look at a very simple example, $X \sim \mathrm{U}[0,1]$. In this case, $a = c = 2$, $F^{-1}(t) = t$, and $E_t = (1-t)^2/2$, for $t \in (0,1)$. Using the σ^2 formula (A.21), we get

$$\begin{aligned}\sigma^2 &= \frac{1}{(p-q)^2}\left(4\varepsilon(1-\varepsilon) + \frac{1}{\varepsilon^2}\int_q^1 x(1-x)^2\mathrm{d}x - \frac{2}{\varepsilon}(1-p)^2 + 4(2\varepsilon^2-\varepsilon)\right) \\ &= \frac{1}{\varepsilon^2}\left(4\varepsilon - 4\varepsilon^2 + \frac{1}{\varepsilon^2}\frac{(1-q)^3(3q+1)}{12} - 2\varepsilon + 8\varepsilon^2 - 4\varepsilon\right) \\ &= \frac{1}{\varepsilon^2}\left(-2\varepsilon + 4\varepsilon^2 + \frac{8\varepsilon(4-6\varepsilon)}{12}\right) = \frac{2}{3\varepsilon}.\end{aligned}$$

We observe that, as $\varepsilon \downarrow 0$, the asymptotic variance σ^2 explodes at a rate of $O(\varepsilon^{-1})$.

As a second example, we look at $X \sim \mathrm{Pareto}(\alpha)$ for $\alpha > 2$, as required by Assumption 4. In this case, we study the asymptotic behaviour as $\varepsilon \downarrow 0$, which means $p, q \uparrow 1$. We can calculate $c = (\frac{\alpha}{\alpha-1})^\alpha$, $f(x) = \alpha x^{-\alpha-1}$ for $x \geqslant 1$, $F^{-1}(p) = \varepsilon^{-1/\alpha}$, $F^{-1}(q) = (c\varepsilon)^{-1/\alpha} = \frac{\alpha-1}{\alpha}\varepsilon^{-1/\alpha}$, $a = c/f(F^{-1}(p)) = c\varepsilon^{-(\alpha+1)/\alpha}/\alpha$, $b = \varepsilon^{-1/\alpha}/\alpha$, $E_t = \frac{1}{\alpha-1}(1-t)^{(\alpha-1)/\alpha}$, $t \in (0,1)$, $E_p = \frac{1}{\alpha-1}\varepsilon^{(\alpha-1)/\alpha}$ and $E_q = \frac{1}{\alpha-1}\varepsilon^{(\alpha-1)/\alpha}(\frac{\alpha}{\alpha-1})^{\alpha-1}$. We analyze each term in the right-hand-side of (A.21) as $\varepsilon \downarrow 0$. The first term is

$$a^2(\varepsilon - \varepsilon^2) \sim a^2\varepsilon = \left(\frac{c}{\alpha}\right)^2 \varepsilon^{-2/\alpha-1};$$

the second term is

$$\begin{aligned}
\frac{2}{\varepsilon^2}\int_{F^{-1}(q)}^{\infty} E_{F(x)}F(x)\mathrm{d}x &= \frac{2}{\varepsilon^2}\int_{F^{-1}(q)}^{\infty} \frac{1}{\alpha-1}x^{1-\alpha}(1-x^{-\alpha})\mathrm{d}x \\
&\sim \frac{2}{\varepsilon^2}\int_{F^{-1}(q)}^{\infty} \frac{1}{\alpha-1}x^{1-\alpha}\mathrm{d}x \\
&= \frac{2}{\varepsilon^2}\frac{1}{(\alpha-1)(\alpha-2)}(F^{-1}(q))^{2-\alpha} \\
&= \frac{2}{\varepsilon^2}\frac{1}{(\alpha-1)(\alpha-2)}c^{1-2/\alpha}\varepsilon^{(\alpha-2)/\alpha} \\
&= \frac{2}{(\alpha-1)(\alpha-2)}c^{1-2/\alpha}\varepsilon^{-2/\alpha-1};
\end{aligned}$$

the third term is

$$-\frac{2a}{\varepsilon}E_p = -\frac{2a}{\alpha-1}\varepsilon^{-1/\alpha} = -\frac{2c}{(\alpha-1)\alpha}\varepsilon^{-2/\alpha-1},$$

and finally, the fourth term is

$$2a(E_q - b) \sim -2ab = -\frac{2c}{\alpha^2}\varepsilon^{-2/\alpha-1}.$$

Using the above four equations and (A.21), and noting that $\varepsilon^{-2/\alpha-1}/b^2 = \alpha^2\varepsilon^{-1}$, we get

$$\begin{aligned}
\sigma^2 &\sim \frac{\alpha^2}{\varepsilon}\left(\left(\frac{c}{\alpha}\right)^2 + \frac{2}{(\alpha-1)(\alpha-2)}c^{1-2/\alpha} - \frac{2c}{(\alpha-1)\alpha} - \frac{2c}{\alpha^2}\right) \\
&= \frac{1}{\varepsilon}\left(c^2 + \frac{2c(\alpha-1)}{\alpha-2} - \frac{2\alpha c}{\alpha-1} - 2c\right) \\
&= \frac{c}{\varepsilon}\left(c + \frac{2}{\alpha-2} - \frac{2\alpha}{\alpha-1}\right). \qquad \text{(A.22)}
\end{aligned}$$

Once again, as $\varepsilon \to 0$, the asymptotic variance explodes at a rate of $O(\varepsilon^{-1})$. In particular, if $\alpha = 4$, we have $c = 3.16$, and (A.22) gives

$$\sigma^2 \approx \frac{3.16}{\varepsilon}\left(3.16 + 1 - \frac{8}{3}\right) = 4.71 \times \varepsilon^{-1}.$$

B The smoothed empirical PELVE estimator

The empirical ES function $\widehat{\mathrm{ES}}_p$ in Section 5 is continuous in p, whereas the sample quantile $\widehat{\mathrm{VaR}}_p$ is a step function and it has jumps at $p = i/n$ for integers i. This may create some non-smoothness in the empirical PELVE estimator, as discussed in Remark 8. To overcome this issue, one can use the *smoothed empirical PELVE estimator* denote by $\widetilde{\Pi}_\varepsilon(n)$, based on the standard linearly interpolated quantile. The standard linear interpolated estimator for VaR in Section 9.2.6 of McNeil et al. (2015) is defined as

$$\widetilde{\mathrm{VaR}}_p = \lambda_{p,k,n} X_{[n-k]} + (1 - \lambda_{p,k,n}) X_{[n-k+1]}, \text{ for } p \in \left(\frac{i-1}{n}, \frac{i}{n} \right], \ i = 1, \dots, n,$$

where $k = \lceil (n-1)(1-p) \rceil$ and $\lambda_{p,k,n} = (n-k) - (n-1)p$. The corresponding smoothed empirical ES estimator is given by

$$\widetilde{\mathrm{ES}}_p = \frac{1}{1-p} \int_p^1 \widetilde{\mathrm{VaR}}_q \mathrm{d}q, \quad p \in (0,1).$$

We then define the smoothed empirical PELVE estimator $\widetilde{\Pi}_\varepsilon(n)$ as the solution $c \in [1, 1/\varepsilon]$ to the equation

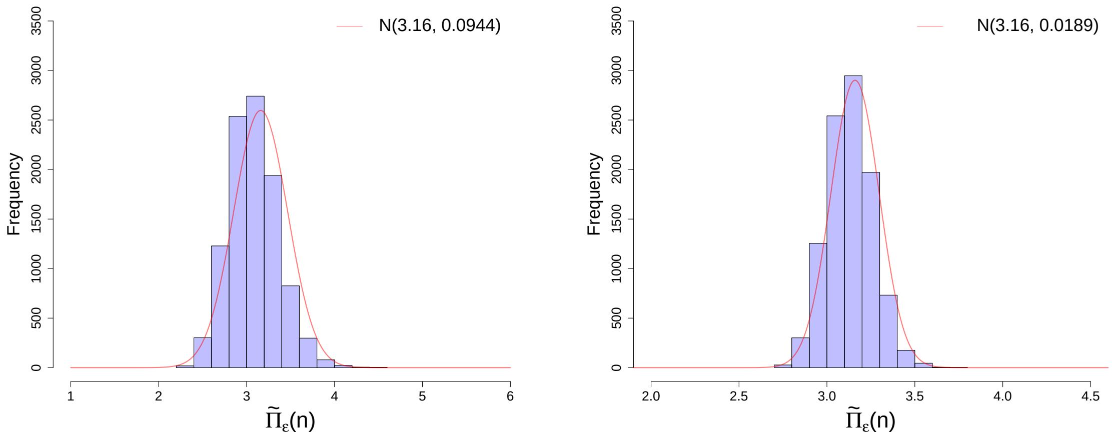
$$\widetilde{\mathrm{ES}}_{1-c\varepsilon} = \widetilde{\mathrm{VaR}}_{1-\varepsilon}.$$

The consistency and asymptotic normality results in Section 5 remain valid for the smoothed empirical PELVE estimator $\widetilde{\Pi}_\varepsilon(n)$. Since a smoothed version of quantile differs from the empirical quantile in an order of $1/n$, $|\widetilde{\Pi}_\varepsilon(n) - \widehat{\Pi}_\varepsilon(n)|$ is also of order $1/n$. In particular, $\sqrt{n}(\widetilde{\Pi}_\varepsilon(n) - c)$ shares the same asymptotic limit as $\sqrt{n}(\widehat{\Pi}_\varepsilon(n) - c)$ in the proof of Theorem 4.

In Figure 8 we report histograms of $\widetilde{\Pi}_\varepsilon(n)$ in the setting of Figure 2. The smoothed empirical PELVE estimator $\widetilde{\Pi}_\varepsilon(n)$ tends to slightly underestimate although its sample variance is also slightly smaller. The tables in Appendix C contain more numerical results with similar observations.

C Simulation of empirical PELVE estimators

In Tables 3-6 below we report simulation results for a numerical comparison of the empirical PELVE estimators $\widehat{\Pi}_\varepsilon$ (standard) and $\widetilde{\Pi}_\varepsilon$ (smoothed) and their theoretical limit for normal, uniform and Pareto loss distributions. In the following, s^2 denotes the sample variance of a PELVE estimator $\widehat{\Pi}_\varepsilon$ or $\widetilde{\Pi}_\varepsilon$, and $\sigma_n^2 = \sigma^2/n$ denotes the corresponding theoretical limit derived from (A.21). As shown by Theorem 4, both $\widehat{\Pi}_\varepsilon$ and $\widetilde{\Pi}_\varepsilon$ have a normal limit with mean $\Pi_\varepsilon(X)$ and variance σ_n^2. We observe from the tables below that the sample variance of $\widehat{\Pi}_\varepsilon$ or $\widetilde{\Pi}_\varepsilon$ converges to the asymptotic formula as n increases. Note again that the effective



(a) Histogram of $\widetilde{\Pi}_\varepsilon(n)$, $n = 1000$

(b) Histogram of $\widetilde{\Pi}_\varepsilon(n)$, $n = 5000$

Figure 8: Smoothed empirical PELVE estimators for Pareto(4), $\varepsilon = 0.05$

sample size is roughly $n\varepsilon$ for $\mathrm{VaR}_{1-\varepsilon}$, and hence the simulation results for $\varepsilon = 0.01$ are not very close to the theoretical limit. We also find that the smoothed PELVE estimator has a significant bias for Pareto(4) although its sample variance is smaller; a possible explanation is that the smoothed ES estimator has a significant negative bias.

ε	n	Π_ε (theoretical)		$\widetilde{\Pi}_\varepsilon$ (smoothed)		$\widehat{\Pi}_\varepsilon$ (standard)	
		value	σ_n^2	mean	s^2	mean	s^2
0.10	500	2.46	0.0242	2.44	0.0236	2.48	0.0242
	1000	2.46	0.0121	2.45	0.0120	2.47	0.0121
	5000	2.46	0.0024	2.45	0.0024	2.46	0.0024
0.05	500	2.51	0.0538	2.48	0.0522	2.56	0.0550
	1000	2.51	0.0269	2.49	0.0263	2.53	0.0271
	5000	2.51	0.0053	2.50	0.0052	2.51	0.0052
0.01	500	2.58	0.3028	2.50	0.2625	2.82	0.3239
	1000	2.58	0.1514	2.51	0.1399	2.69	0.1575
	5000	2.58	0.0302	2.55	0.0294	2.60	0.0304

Table 3: Empirical PELVE estimators for $\mathrm{N}(0, 1)$

In Table 7, we consider an AR(1) model with Gaussian innovation, that is,

$$X_{t+1} = aX_t + Z_t, \quad t \in \mathbb{Z},$$

where Z_t, $t \in \mathbb{Z}$, are standard Gaussian noises, and $a \in \{-0.1, 0.3, 0.9\}$. For this model, the stationary distribution of X_t is Gaussian, and we know the true value of PELVE of X_t from Table 3. We simulate a sample path from this AR(1) model of length $10,000$. We provide a point estimate $\widehat{\Pi}_\varepsilon$ and a block-bootstrap standard deviation estimator described

ε	n	Π_ε (theoretical)		$\widetilde{\Pi}_\varepsilon$ (smoothed)		$\widehat{\Pi}_\varepsilon$ (standard)	
		value	σ_n^2	mean	s^2	mean	s^2
0.10	500	2.00	0.0133	2.00	0.0136	2.03	0.0139
	1000	2.00	0.0067	2.00	0.0067	2.02	0.0067
	5000	2.00	0.0013	2.00	0.0013	2.00	0.0013
0.05	500	2.00	0.0266	2.00	0.0258	2.04	0.0265
	1000	2.00	0.0133	2.00	0.0130	2.02	0.0132
	5000	2.00	0.0027	2.00	0.0027	2.00	0.0027
0.01	500	2.00	0.1333	2.03	0.1273	2.23	0.1421
	1000	2.00	0.0667	2.01	0.0631	2.11	0.0667
	5000	2.00	0.0133	2.00	0.0131	2.02	0.0132

Table 4: Empirical PELVE estimators for U$[0, 1]$

ε	n	Π_ε (theoretical)		$\widetilde{\Pi}_\varepsilon$ (smoothed)		$\widehat{\Pi}_\varepsilon$ (standard)	
		value	σ_n^2	mean	s^2	mean	s^2
0.10	500	3.16	0.0944	3.08	0.0785	3.18	0.0928
	1000	3.16	0.0472	3.11	0.0415	3.17	0.0473
	5000	3.16	0.0094	3.15	0.0092	3.15	0.0097
0.05	500	3.16	0.1888	3.04	0.1441	3.20	0.1811
	1000	3.16	0.0944	3.08	0.0763	3.18	0.0911
	5000	3.16	0.0189	3.13	0.0174	3.17	0.0191
0.01	500	3.16	0.9443	2.91	0.5858	3.38	0.9012
	1000	3.16	0.4721	2.97	0.3287	3.27	0.4557
	5000	3.16	0.0944	3.09	0.0783	3.19	0.0933

Table 5: Empirical PELVE estimators for Pareto(4)

in Section 5.3 based on parameters $m = 20$, $k = 500$ and $B = 100,000$. We also include the average of estimates and the standard deviation obtained from Monte Carlo simulation of 10,000 independent paths, which can be approximately treated as the true PELVE and the true standard deviation of $\widehat{\Pi}_\varepsilon$. As we can see from Table 7, the block-bootstrap standard deviation is reasonably close to the Monte Carlo standard deviation, consistent with the result in Proposition 5.

D Additional figures

D.1 PELVE(log-loss) vs PELVE(loss)

In Figures 9, we compare the empirical PELVE values for S&P 500 log-loss and linear loss. As implied by Example 3, PELVE of the log-loss is generally larger than PELVE of the linear loss, but we can see that this difference is negligible (the average of the difference is 0.011, reported in the red curve in Figure 9).

ε	n	Π_ε (theoretical)		$\widetilde{\Pi}_\varepsilon$ (smoothed)		$\widehat{\Pi}_\varepsilon$ (standard)	
		value	σ_n^2	mean	s^2	mean	s^2
0.10	500	2.87	0.0515	2.83	0.0475	2.90	0.0511
	1000	2.87	0.0257	2.84	0.0251	2.88	0.0263
	5000	2.87	0.0051	2.86	0.0051	2.87	0.0051
0.05	500	2.87	0.1030	2.80	0.0951	2.92	0.1069
	1000	2.87	0.0515	2.83	0.0496	2.89	0.0536
	5000	2.87	0.0103	2.85	0.0099	2.87	0.0101
0.01	500	2.87	0.5153	2.71	0.3805	3.10	0.5201
	1000	2.87	0.2576	2.75	0.2110	2.99	0.2594
	5000	2.87	0.0515	2.82	0.0472	2.89	0.0508

Table 6: Empirical PELVE estimators for Pareto(10)

ε	a	Π_ε value	$\widehat{\Pi}_\varepsilon$ est.	BB std est.	MC mean	MC std
0.10	-0.1	2.46	2.426	0.0392	2.458	0.0346
	0.3	2.46	2.438	0.0327	2.458	0.0349
	0.9	2.46	2.435	0.0523	2.456	0.0461
0.05	-0.1	2.51	2.561	0.0470	2.513	0.0516
	0.3	2.51	2.467	0.0524	2.513	0.0519
	0.9	2.51	2.427	0.0763	2.509	0.0659
0.01	-0.1	2.58	2.499	0.1104	2.590	0.1229
	0.3	2.58	2.517	0.1371	2.589	0.1242
	0.9	2.58	2.599	0.1248	2.578	0.1436

Table 7: Block-bootstrap (BB) std estimates versus Monte Carlo (MC) std in an AR(1) simulation

D.2 Empirical analysis for other asset classes

In Figure 10, we report the empirical average PELVE values for the other 9 sectors of S&P 500, complementing the discussion in Section 6 where empirical average PELVE values for Financials and IT are presented. We observe that the PELVE estimates of each sector are almost always above $e \approx 2.72$, indicating the presence of heavy tails empirically. Moreover, most PELVE curves exhibit unprecedented peaks around the COVID-19 outbreak. We also report the average 99% VaR and 97.5% ES of each sector in Figure 11.

D.3 Empirical PELVE estimates at levels 1% and 2.5%

Choosing a smaller ε close to 0 leads to more noisy estimation of PELVE since effectively less data are used for a smaller level ε. We present 1%, 2.5% and 5% empirical PELVE point estimates in Figure 12. We observe that 2.5% PELVE hovers around 5% PELVE estimates computed using a fixed lookback window of size 500; on the other hand, 1% PELVE is more noisy as expected. Nevertheless, the main qualitative conclusions in Section 6.1 that the

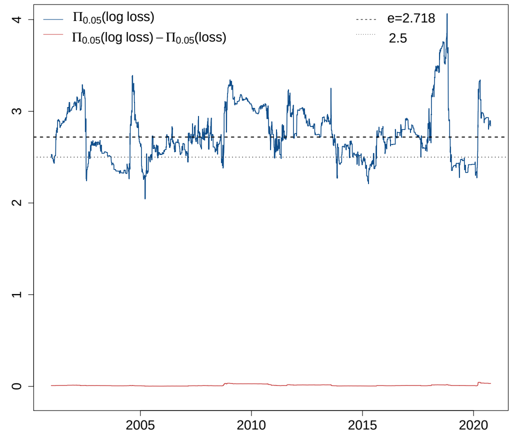
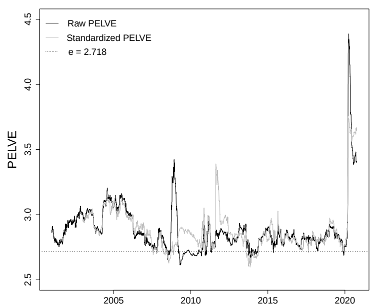
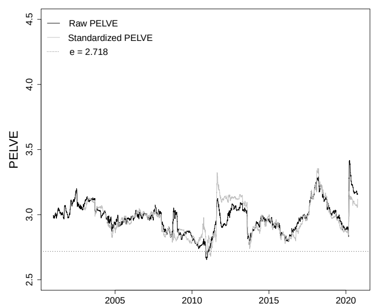


Figure 9: 5% PELVE of S&P 500 log-loss vs linear loss (Jan 02, 2001 - October 9, 2020)

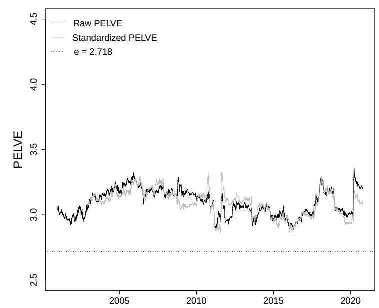
individual assets are heavier-tailed remain unchanged for PELVE with $\varepsilon = 1\%, 2.5\%, 5\%$.



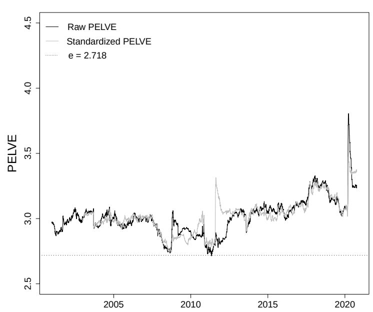
(a) S&P 500 Real Estate



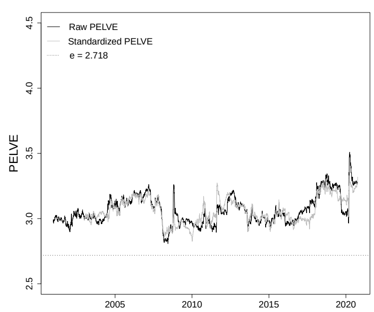
(b) S&P 500 Industrials



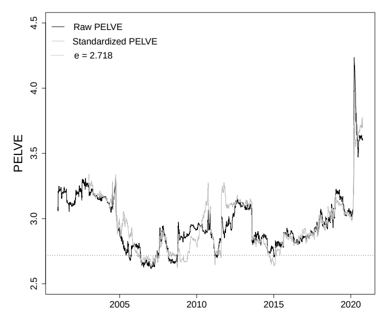
(c) S&P 500 Health Care



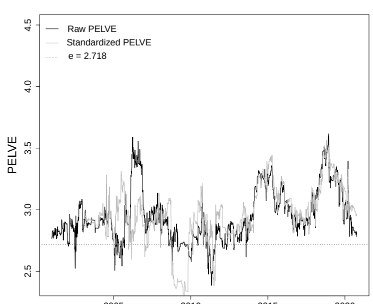
(d) S&P 500 Consumer Discretionary



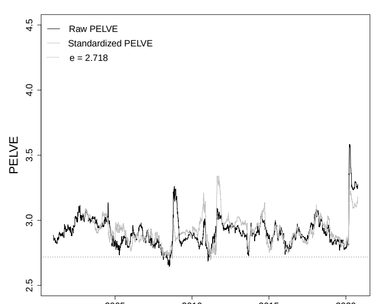
(e) S&P 500 Consumer Staples



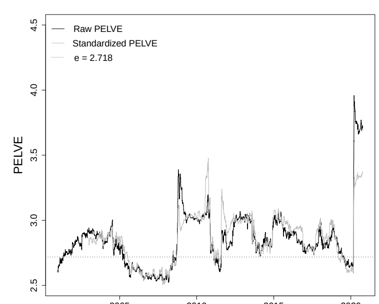
(f) S&P 500 Utilities



(g) S&P 500 Communication Services

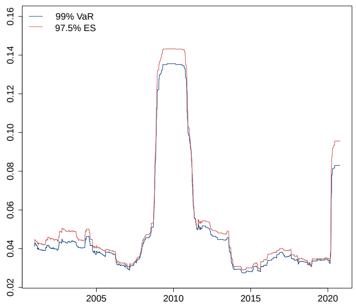


(h) S&P 500 Materials

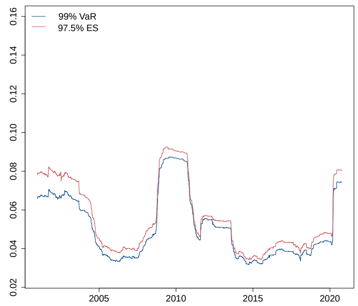


(i) S&P 500 Energy

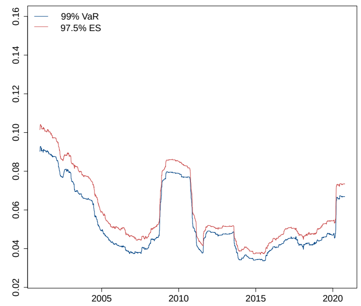
Figure 10: Raw and standardized sector-wise 5% PELVE (Jan 02, 2001 - October 9, 2020)



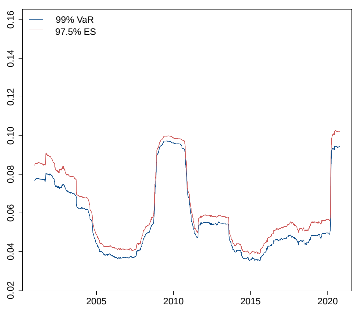
(a) S&P 500 Real Estate



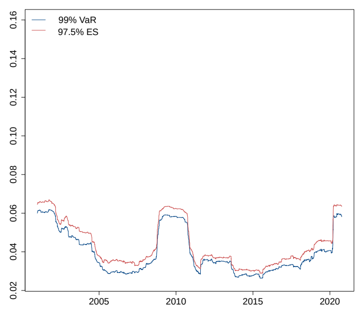
(b) S&P 500 Industrials



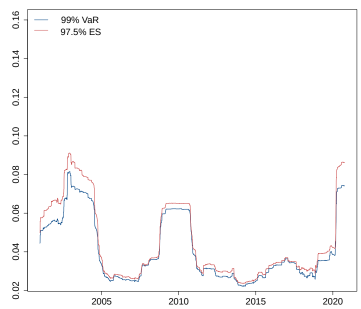
(c) S&P 500 Health Care



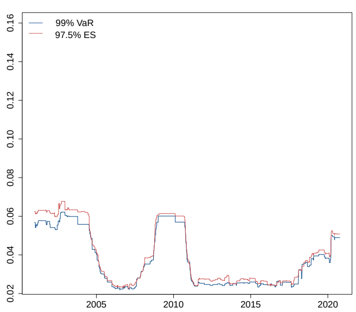
(d) S&P 500 Consumer Discretionary



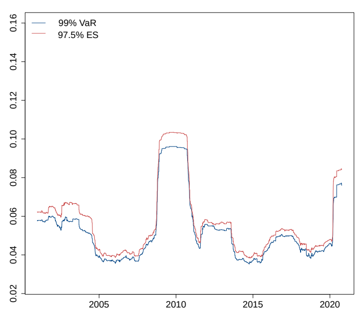
(e) S&P 500 Consumer Staples



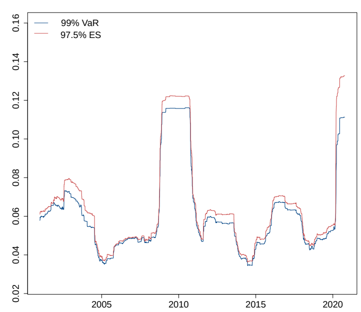
(f) S&P 500 Utilities



(g) S&P 500 Communication Services

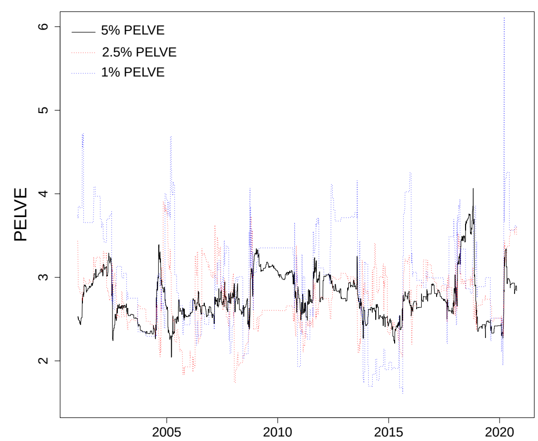


(h) S&P 500 Materials

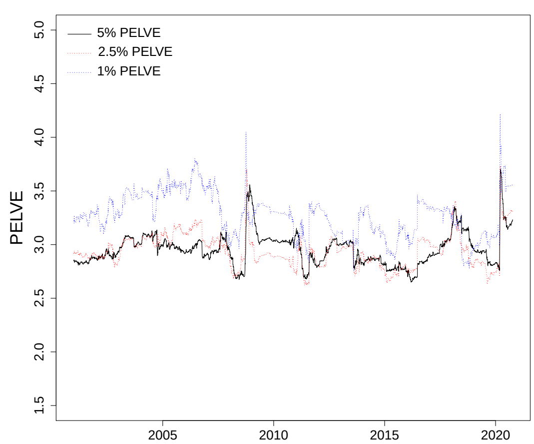


(i) S&P 500 Energy

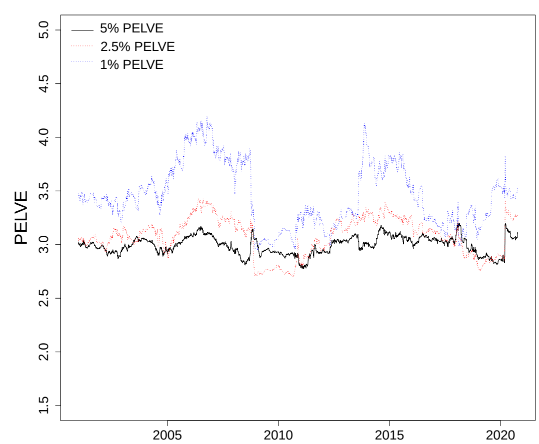
Figure 11: Sector-wise 99% VaR and 97.5% ES (Jan 02, 2001 - October 9, 2020)



(a) PELVE for the S&P index



(b) Average PELVE for S&P Financials



(c) Average PELVE for S&P IT

Figure 12: 1%, 2.5% and 5% empirical PELVE estimates