

# Calibrating distribution models from PELVE

Hirbod Assa\*      Liyuan Lin†      Ruodu Wang‡

May 1, 2023

## Abstract

The Value-at-Risk (VaR) and the Expected Shortfall (ES) are the two most popular risk measures in banking and insurance regulation. To bridge between the two regulatory risk measures, the Probability Equivalent Level of VaR-ES (PELVE) was recently proposed to convert a level of VaR to that of ES. It is straightforward to compute the value of PELVE for a given distribution model. In this paper, we study the converse problem of PELVE calibration, that is, to find a distribution model that yields a given PELVE, which may either be obtained from data or from expert opinion. We discuss separately the cases when one-point, two-point,  $n$ -point and curve constraints are given. In the most complicated case of a curve constraint, we convert the calibration problem to that of an advanced differential equation. We apply the model calibration techniques to estimation and simulation for datasets used in insurance. We further study some technical properties of PELVE by offering a few new results on monotonicity and convergence.

**Keywords:** Value-at-Risk, Expected Shortfall, risk measures, heavy tails, advanced differential equation.

## 1 Introduction

Value-at-Risk (VaR) and Expected Shortfall (ES, also known as TVaR and CVaR) are the most widely used risk measures for regulation in finance and insurance. The former has gained its popularity due to its simplistic approach toward risk as the risk quantile, and the second one is perceived to be useful as a modification of VaR with more appealing properties, such as tail-sensitivity and subadditivity, as studied in the seminal work of [Artzner et al. \(1999\)](#).

---

\*Kent Business School, UK. [h.assa@kent.ac.uk](mailto:h.assa@kent.ac.uk).

†Department of Statistics and Actuarial Science, University of Waterloo, Canada. [189lin@uwaterloo.ca](mailto:189lin@uwaterloo.ca).

‡Department of Statistics and Actuarial Science, University of Waterloo, Canada. [wang@uwaterloo.ca](mailto:wang@uwaterloo.ca).

We refer the reader to the online version for color figures.

In the Fundamental Review of the Trading Book (FRTB), the Basel Committee on Banking Supervision (BCBS (2019)) proposed to replace VaR at 1% confidence with ES with a 2.5% confidence interval for the internal model-based approach.<sup>1</sup> The main reason, as mentioned in the FRTB, was that ES can better capture tail risk; see Embrechts et al. (2018) for a concrete risk sharing model where tail risk is captured by ES and ignored by VaR. On the other hand, VaR also has advantages that ES does not have, such as elicibility (e.g., Gneiting (2011) and Kou and Peng (2016)) or backtesting tractability (e.g., Acerbi and Székely (2014)), and the two risk measures admit different axiomatic foundations (see Chambers (2009) and Wang and Zitikis (2021)). We refer to the reviews of Embrechts et al. (2014) and Emmer et al. (2015) for general discussions on VaR and ES, and McNeil et al. (2015) for a standard treatment on risk management including the use of VaR and ES. The technical contrasts of the two risk measures and their co-existence in regulatory practice give rise to great interest from both researchers and practitioners to explore the relationship between them.

To understand the balancing point of VaR and ES during the transition in the FRTB, Li and Wang (2022) proposed the Probability Equivalent Level of VaR-ES (PELVE). The value of PELVE is the multiplier to the tail probability when replacing VaR with ES such that the capital calculation stays unchanged. More precisely, the PELVE of  $X$  at level  $\varepsilon$  is the multiplier  $c$  such that  $ES_{c\varepsilon}(X) = VaR_{\varepsilon}(X)$ ; such  $c$  uniquely exists under mild conditions. For instance, if  $VaR_{1\%}(X) = ES_{2.5\%}(X)$  for a future portfolio loss  $X$ , then PELVE of  $X$  at probability level 0.01 is the multiplier 2.5. In this case, replacing  $VaR_{1\%}$  with  $ES_{2.5\%}$  in FRTB does not have much effect on the capital requirement for the bank bearing the loss  $X$ . Instead, if  $ES_{2.5\%}(X) > VaR_{1\%}(X)$ , then the bank has a larger capital requirement under the new regulatory risk measure; this is often the case for financial assets and portfolios as shown by the empirical studies in Li and Wang (2022). The PELVE enjoys many convenient properties, and it has been extended in a few ways. In particular, Fiori and Gianin (2022) defined generalized PELVE by replacing VaR and ES with another pair of monotone risk measures  $(\rho, \tilde{\rho})$ , and Barczy et al. (2022) extended PELVE by replacing ES with a higher-order ES.

For a given distribution model or a data set, its PELVE can be computed or estimated in a straightforward manner. As argued by Li and Wang (2022), the PELVE for a small  $\varepsilon$  may be seen as a summarizing index measuring tail heaviness in a non-limit sense. As such, one may like to generate models for a given PELVE, in a way similar to constructing models for other given statistical information; see e.g., Embrechts et al. (2002, 2016) for constructing multivariate models with a given correlation or tail-dependence matrix. Such statistical information may be obtained either from data or from expert opinion, but there is no a priori guarantee that a corresponding model exists. Since PELVE involves a parameter  $\varepsilon \in (0, 1)$ , its information is represented by a curve. The calibration problem, that is, to find a distribution model for given PELVE values or a given PELVE curve, turns

---

<sup>1</sup>In this paper, we use the “small  $\alpha$ ” convention for VaR and ES. Hence, “VaR at 1% confidence” and “ES at 2.5% confidence” correspond to  $VaR_{99\%}$  and  $ES_{97.5\%}$  in BCBS (2019), respectively.

out to be highly non-trivial, and it is the main objective of the current paper.

From now on, suppose that we receive some information on the PELVE of a certain random loss from an expert opinion, and we aim to build a distribution model consistent with the supplied information. Since PELVE is location-scale invariant, such a distribution, if it exists, is not unique.

The calibration problem is trivial if we are supplied with only one point on the PELVE curve. As the PELVE curve of the generalized Pareto distribution is a constant when the PELVE is well defined, we can use the generalized Pareto distribution to match the given PELVE value, which has a tail index implied from the expert opinion. The calibration problem becomes more involved if we are supplied with two points on the PELVE curve, because the value of the PELVE at two different probability levels interact with each other. The situation becomes more complicated as the number of points increases, and we further turn to the problem of calibration from a fully specified PELVE curve. Calibrating distribution from the PELVE curve can be reformulated as solving for a function  $f$  via the integral equation  $\int_0^y f(s) ds = yf(z(y)y)$ , where the curve  $z$  is computed from the PELVE curve. This integral equation can be further converted to an advanced differential equation (see [Bellman and Cooke \(1963\)](#)). For the case that  $z$  is a constant curve, we can explicitly obtain all solutions for  $f$ . We find other distributions that also have constant PELVE curves other than the simple ones with a Pareto or exponential distribution. As a consequence, a PELVE curve does not characterize a unique location-scale family of distributions; this provides a negative answer to a question posed by [Li and Wang \(2022, Section 7, Question \(iv\)\)](#). For general function  $z$ , we develop a numerical method to compute  $f$ .

The calibrated distribution can be used to estimate the value of other risk measures such as VaR and ES at different levels. We illustrate by an empirical example that two points of PELVE give a quite good summary of the tail distribution of risk. Daily log-losses (negative log-returns) of AAPL from Yahoo Finance are collected for the period from January 3, 2012 to December 31, 2021 within total of 2518 observations. We calculate the empirical PELVE at levels 0.01 and 0.05 using the empirical PELVE estimator provided by [Li and Wang \(2022, Section 5\)](#) with a moving window of 500 trading days. For each pair of two points of PELVE at levels 0.01 and 0.05, we produce a quantile curve from the two empirical PELVE points by our calibration model in [Section 3.2](#), which is scaled such that  $\text{VaR}_{0.01}$  and  $\text{VaR}_{0.05}$  are equal to their empirical values.<sup>2</sup> [Figure 1](#) presents the empirical and calibrated quantile curves on December 31, 2021 using 500 trading days prior to that date. The two quantile curves are close to each other, with our calibrated curve being more smooth. We also report the values of  $\text{ES}_{0.025}$  of the calibrated distribution, which we call the calibrated  $\text{ES}_{0.025}$ , and compare it with empirical  $\text{ES}_{0.025}$ . The left panel of [Figure 2](#) shows the curves of empirical and calibrated  $\text{ES}_{0.025}$ . In the right panel of [Figure 2](#), we create a scatter plot using empirical and calibrated  $\text{ES}_{0.025}$ . Both

---

<sup>2</sup>Recall that PELVE is location-scale free, and hence we need to pick two free parameters to specify a distribution calibrated from PELVE.

figures show that the empirical and calibrated  $ES_{0.025}$  curves are quite close.

Figure 1: Empirical VaR and calibrated VaR

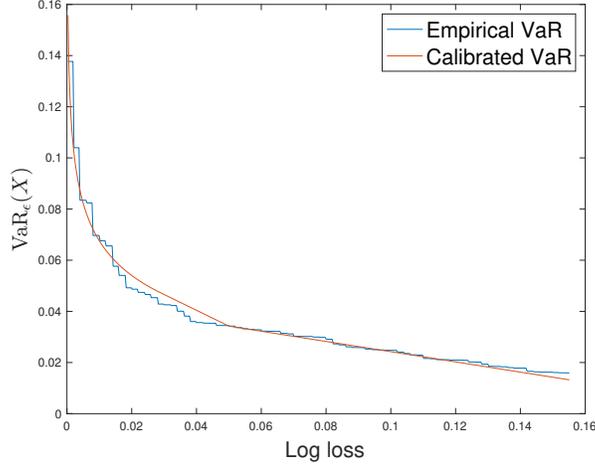
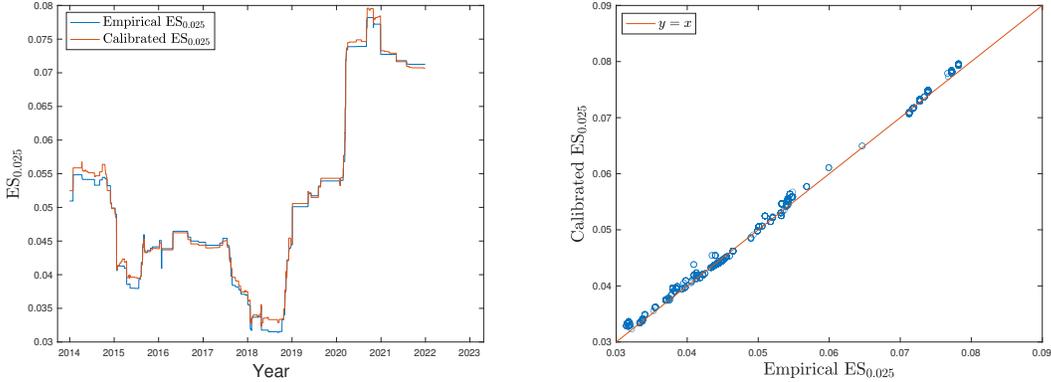


Figure 2: Empirical  $ES_{0.05}$  and calibrated  $ES_{0.05}$



To further enrich the theory of PELVE, we study a few technical properties of PELVE, such as monotonicity and convergence as the probability level goes to 0. A decreasing PELVE indicates a relatively larger impact of ES in risk assessment than VaR moving towards the tail. As we will see, while for the most known parametric distributions the PELVE is decreasing, there exist some examples at some risk levels it is not decreasing. This means that for those examples VaR becomes a stricter risk measure when moving towards the tail. To obtain conditions for monotonicity, we define the dual PELVE by moving the multiplier  $c$  from the ES side to the VaR side. PELVE can be seen as a functional measure of tail heaviness in the sense that a heavier-tailed distribution has a higher PELVE curve (Li and Wang (2022, Theorem 1)). The hazard rate, on the other hand, is another functional measure of tail heaviness. We show that the PELVE is decreasing (increasing) if the inverse of the hazard rate

is convex (concave). Monotonicity also leads to conditions for the PELVE to have a limit at the tail, which from the risk management perspective, identifies the ultimate relative positions of ES and VaR in the tail region. From a mathematical perspective, the limit of PELVE at 0 allows us to extend the domain of PELVE to include 0 as a measure of tail heaviness.

The rest of this paper is organized as follows. Section 2 introduces the background and examples of the PELVE. In Section 3 we calibrate a distribution from finitely many points in the PELVE curve. Section 4 calibrates the distribution from given PELVE curves, where we give a class of explicit solutions for constant PELVE functions and numerical solutions for general PELVE functions. In Section 5, we study the monotonicity and convergence of the PELVE. Section 6 presents two examples of the model calibration techniques applied to datasets used in insurance. A conclusion is given in Section 7. Some technical proofs of results in Sections 3, 4 and 5 are provided in the Appendices.

## 2 Definitions and background

Let us consider an atomless probability space  $(\Omega, \mathcal{F}, \mathbb{P})$ , where  $\mathcal{F}$  is the set of the measurable sets and  $\mathbb{P}$  is the probability measure. Let  $L^1$  be the set of integrable random variables, i.e.,  $L^1 = \{X : \mathbb{E}[|X|] < \infty\}$ , where  $\mathbb{E}$  is the expectation with respect to  $\mathbb{P}$ .

We first define VaR and ES in  $L^1$ , the two most popular risk measures. The VaR and at probability level  $p \in (0, 1)$  is defined as

$$\text{VaR}_p(X) = \inf\{x \in \mathbb{R} : \mathbb{P}(X \leq x) \geq 1 - p\} = F^{-1}(1 - p), \quad X \in L^1, \quad (1)$$

where  $F$  is the distribution of  $X$ . The ES at probability level  $p \in [0, 1)$  is defined as

$$\text{ES}_p(X) = \frac{1}{p} \int_0^p \text{VaR}_q(X) dq, \quad X \in L^1.$$

Note that we use the “small  $\alpha$ ” convention for VaR and ES, which is different from [Liu and Wang \(2021\)](#). Let  $\text{VaR}_0(X) = \text{ES}_0(X) = \text{ess-sup}(X)$  and  $\text{VaR}_1(X) = \text{ess-inf}(X)$ . We have that  $\text{ES}_1(X)$  is the mean of  $X$ . We will also call  $p \mapsto \text{VaR}_p(X)$  the quantile function of  $X$ , keeping in mind that in our convention this function is decreasing.<sup>3</sup>

For  $\varepsilon \in (0, 1)$ , the PELVE at level  $\varepsilon$ , proposed by [Li and Wang \(2022\)](#), is defined as

$$\Pi_X(\varepsilon) = \inf\{c \in [1, 1/\varepsilon] : \text{ES}_{c\varepsilon}(X) \leq \text{VaR}_\varepsilon(X)\}, \quad X \in L^1,$$

where  $\inf(\emptyset) = \infty$ . [Li and Wang \(2022\)](#) used  $\Pi_\varepsilon(X)$  for our  $\Pi_X(\varepsilon)$ , and our choice of notation is due

---

<sup>3</sup>Throughout the paper, all terms like “increasing” and “decreasing” are in the non-strict sense.

to the fact that the curve  $\varepsilon \rightarrow \Pi_X(\varepsilon)$  is the main quantity of interest in this paper.

The PELVE of  $X$  is finite if and only if  $\text{VaR}_\varepsilon(X) \geq \mathbb{E}[X]$ . The value of the PELVE is the multiplier  $c$  such that  $\text{ES}_{c\varepsilon}(X) = \text{VaR}_\varepsilon(X)$ . If  $\text{VaR}_p(X)$  is not a constant for  $p \in (0, \varepsilon]$ , then the PELVE is the unique solution for the multiplier. By Theorem 1 in [Li and Wang \(2022\)](#), the PELVE is location-scale invariant. The distribution with a heavy tail will have a higher PELVE value.

If  $X$  is a normal distributed random variable and  $\varepsilon = 1\%$ , we have  $\Pi_X(\varepsilon) \approx 2.5$ . It means that  $\text{ES}_{2.5\%}(X) \approx \text{VaR}_{1\%}(X)$ . That is, the replacement suggested by BCBS is fair for normally distributed risks. In other words, a higher PELVE will result in a higher capital requirement after the replacement.

In this paper, we are generally interested in the question of which distributions have a specified or partially specified PELVE curve. We first look at a few simple examples.

**Example 1** (Constant PELVE). We first list some distributions that have constant PELVE curves. From the definition of the PELVE, we know that the PELVE should be larger than 1. As we can see from [Table 1](#), the PELVE for the generalized Pareto distribution takes values on  $(1, \infty)$ . For  $X \sim \text{GPD}(\xi)$ , we have  $1 < \Pi_X(\varepsilon) < e$  when  $\xi < 0$ ,  $\Pi_X(\varepsilon) = e$  when  $\xi = 0$  and  $\Pi_X(\varepsilon) > e$  when  $\xi > 0$ . Furthermore, if  $X$  follows the point-mass distribution  $\delta_c$  or the Bernoulli distribution, we have  $\Pi_X(\varepsilon) = 1$ .

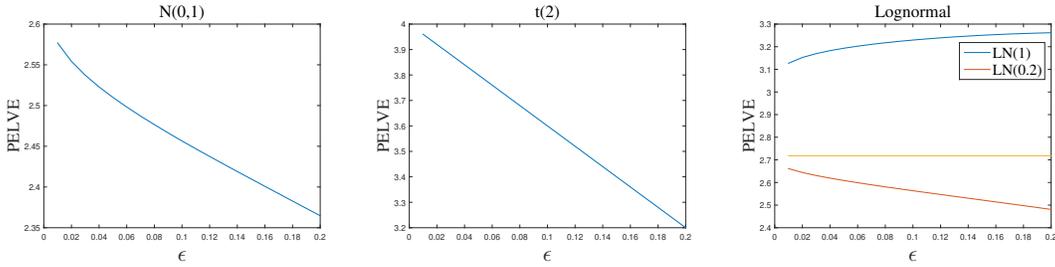
Table 1: Example of constant PELVE

Distribution	Distribution or probability function of $X$	$\Pi_X(\varepsilon)$
$\delta_c$	$\mathbb{P}(X = c) = 1$	$\Pi_X(\varepsilon) = 1$ for $\varepsilon \in (0, 1)$
$B(1, p)$	$\mathbb{P}(X = 1) = p$ and $\mathbb{P}(X = 0) = 1 - p$	$\Pi_X(\varepsilon) = 1$ for $\varepsilon \in (0, p)$
$U(0, 1)$	$F(t) = t$ for $t \in (0, 1)$	$\Pi_X(\varepsilon) = 2$ for $0 < \varepsilon < 1/2$
$\text{EXP}(\lambda)$	$F(t) = 1 - \exp(-\lambda t)$ , $\lambda > 0$	$\Pi_X(\varepsilon) = e$ for $0 < \varepsilon < 1/e$
$\text{GPD}(\xi)^1$	$F(x) = \begin{cases} 1 - (1 + \xi x)^{-\frac{1}{\xi}} & \xi \neq 0 \\ 1 - \exp(-x) & \xi = 0 \end{cases}$	$\Pi_X(\varepsilon) = (1 - \xi)^{-\frac{1}{\xi}}$ for $0 < \varepsilon < (1 - \xi)^{\frac{1}{\xi}}$

<sup>1</sup> The distribution  $\text{GPD}(\xi)$  is called the standard generalized Pareto distribution. As  $\mathbb{E}[X] < \infty$  when  $\xi < 1$ , the PELVE exists only when  $\xi < 1$ . The support of  $\text{GPD}(\xi)$  is  $[0, \infty)$  when  $\xi > 0$  and  $[0, -\frac{1}{\xi}]$  when  $\xi < 0$ . When  $\xi = 0$ , the  $\text{GPD}(\xi)$  is exactly exponential distribution with  $\lambda = 1/\sigma$ . There is a three-parameter  $\text{GPD}(\mu, \sigma, \xi)$ , which is a location-scale transform of standard GPD. Therefore,  $\text{GPD}(\mu, \sigma, \xi)$  has the same PELVE as  $\text{GPD}(\xi)$ .

**Example 2.** Here we present some non-constant PELVE examples. We write  $t(v)$  for the t-distribution with parameter  $(0, 1, v)$ , and  $\text{LN}(\sigma)$  for the log-normal distribution with parameter  $(0, \sigma^2)$ . As we can see, for normal distribution and t-distribution, the PELVE curve is decreasing as  $\varepsilon$  increasing. The monotonicity of the PELVE of the lognormal distribution depends on the value of  $\sigma$ . The monotonicity

Figure 3: PELVE for normal distribution, t-distribution and lognormal distribution



of the PELVE will be further discussed in Section 5. For more PELVE examples, see [Li and Wang \(2022\)](#).

### 3 Calibration from finite-point constraints

In this section, we discuss the calibration problem when some points of the PELVE are given. We will focus on the case where one point or two points on the PELVE curve are specified, for which we can explicitly construct a corresponding quantile function.

We first note that the calibrated distribution is not unique. For example, if we are given  $\Pi_X(0.01) = 2.5$ , we can assume the distribution of  $X$  is the Normal distribution or the generalized Pareto distribution with tail parameter  $\xi$  satisfying  $(1 - \xi)^{-1/\xi} = 2.5$  from Table 3. Therefore, the distributions obtained in our results are only some possible choices, which we choose to have a generalized Pareto tail, as Pareto tails are standard in risk management applications.

#### 3.1 Calibration from a one-point constraint

Based on Table 1, we can calibrate the distribution for  $X$  from one given PELVE point  $(\varepsilon_1, c_1)$  such that  $\Pi_X(\varepsilon_1) = c_1$ . A simple idea is to take the generalized Pareto distribution when  $c_1 > 1$  and  $\delta_c$  when  $c_1 = 1$ . We summarize the idea in the following Proposition.

**Proposition 1.** *Let  $\varepsilon_1 \in (0, 1)$  and  $c_1 \in [1, \infty)$  such that  $c_1 \varepsilon_1 \leq 1$ . If  $c_1 > 1$ , let  $\xi \in \mathbb{R}$  such that  $(1 - \xi)^{-\frac{1}{\xi}} = c_1$ . Then,  $X \sim \text{GPD}(\xi)$  has  $\Pi_X(\varepsilon_1) = c_1$ . If  $c_1 = 1$ , then  $X = k$  for some constant  $k \in \mathbb{R}$  has  $\Pi_X(\varepsilon_1) = c_1$ .*

The proof can be directly derived from Table 1 and it is omitted. By Proposition 1, if we have the value of PELVE at point  $\varepsilon_1$ , we can find a distribution of  $X$  which has the same PELVE value at  $\varepsilon_1$ . If we also have the value of VaR at  $\varepsilon_1$ , we can determine the scale parameter ( $\sigma$ ) for the GPD distribution or the value of  $k$  to match the value of VaR. For Table 1, we can see that the calibrated generalized Pareto distribution can also serve as a solution for a more prudent condition  $\Pi_X(\varepsilon) \geq c_1$  when  $\varepsilon \in (0, \varepsilon_1)$ .

### 3.2 Calibration from a two-point constraint

The calibration problem would be much more difficult when we are given two points of the PELVE curve. Given two points  $(\varepsilon_1, c_1)$  and  $(\varepsilon_2, c_2)$  such that  $\varepsilon_1 < \varepsilon_2$ , we want to find a distribution for  $X \in L^1$  such that  $\Pi_X(\varepsilon_1) = c_1$  and  $\Pi_X(\varepsilon_2) = c_2$ . Nevertheless, the choices of  $(\varepsilon_1, c_1)$  and  $(\varepsilon_2, c_2)$  are not arbitrary. First, we need  $1 \leq c_1 \leq 1/\varepsilon_1$  and  $1 \leq c_2 \leq 1/\varepsilon_2$  by the definition of the PELVE. Then, we will show that the value of  $c_2$  will be restricted if  $(\varepsilon_1, c_1)$  and  $\varepsilon_2$  are given.

**Lemma 1.** *For any  $X \in L^1$ , let  $\varepsilon_1, \varepsilon_2 \in (0, 1)$  be such that  $\mathbb{E}[X] \leq \text{VaR}_{\varepsilon_2}(X)$  and  $\varepsilon_1 < \varepsilon_2$ . Then, we have  $\varepsilon_1 \Pi_X(\varepsilon_1) \leq \varepsilon_2 \Pi_X(\varepsilon_2)$ .*

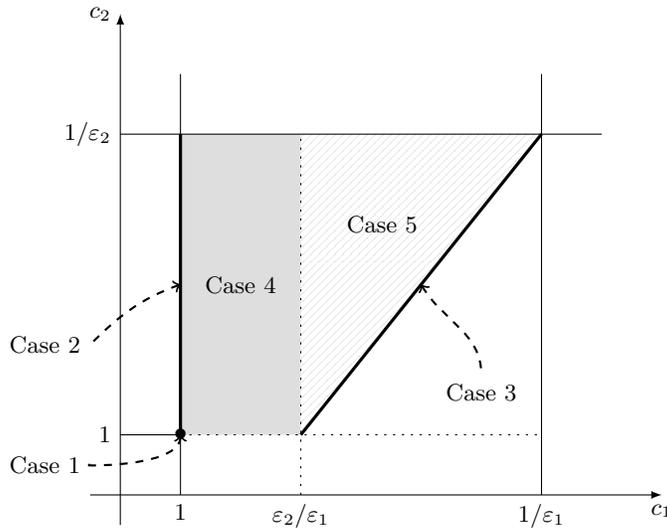
By Lemma 1, for given  $\varepsilon_1, \varepsilon_2$  and  $c_1$ , the value of  $c_2$  is bounded below by both 1 and  $c_1 \varepsilon_1 / \varepsilon_2$ . We also note that if  $c_2 = 1$ , then  $p \mapsto \text{VaR}_p(X)$  is constant on  $(0, \varepsilon_2)$ , which implies  $c_1 = 1$ . In Appendix A, Proposition 6 shows that the above lower bound is achieved if and only if  $\text{VaR}_{\varepsilon_1}(X) = \text{VaR}_{\varepsilon_2}(X)$ .

From the definition of the PELVE and Lemma 1, for  $\varepsilon_1 < \varepsilon_2$ , the possible choices of  $(\varepsilon_1, c_1)$  and  $(\varepsilon_2, c_2)$  should satisfy  $1 \leq c_1 \leq 1/\varepsilon_1$ ,  $1 \leq c_2 \leq 1/\varepsilon_2$  and  $c_1 \varepsilon_1 \leq c_2 \varepsilon_2$ . We denote by  $\Delta$  the admissible set for  $(\varepsilon_1, c_1, \varepsilon_2, c_2)$ , that is,

$$\Delta = \{(\varepsilon_1, c_1, \varepsilon_2, c_2) \in ((0, 1) \times [1, \infty))^2 : \varepsilon_1 < \varepsilon_2, c_1 \varepsilon_1 \leq 1, c_2 \varepsilon_2 \leq 1, c_1 \varepsilon_1 \leq c_2 \varepsilon_2\}.$$

We illustrate the possible region of  $(c_1, c_2)$  with given  $\varepsilon_1$  and  $\varepsilon_2$  in Figure 4. We divide the region into 5 cases and calibrate the distribution for each case.

Figure 4: Admissible region of  $(c_1, c_2)$



The calibration process is to construct a continuous and decreasing quantile function that can

satisfy two equivalent conditions between VaR and ES, which are

$$\text{ES}_{c_1\varepsilon_1}(X) = \text{VaR}_{\varepsilon_1}(X) \quad \text{and} \quad \text{ES}_{c_2\varepsilon_2}(X) = \text{VaR}_{\varepsilon_2}(X). \quad (2)$$

As we can see, only the values of  $\text{VaR}_\varepsilon(X)$  for  $\varepsilon \in (0, c_2\varepsilon_2]$  matters for the equivalent condition (2). Therefore, we focus on constructing  $\text{VaR}_\varepsilon(X)$  for  $\varepsilon \in (0, c_2\varepsilon_2]$ . In addition, we want a continuous calibrated quantile function.

The case  $c_1 = 1$  or  $c_2 = 1$  is special, which means that  $\text{VaR}_\varepsilon(X)$  is a constant on the tail part. If  $c_1 > 1$ , we can set the tail distribution as the generalized Pareto distribution from Table 3 such that  $\Pi_X(\varepsilon_1) = c_1$ .

For  $\mathbf{z} = (\varepsilon_1, c_1, \varepsilon_2, c_2) \in \Delta$ , we will construct a class of functions, denoted by  $G_{\mathbf{z}}$ , in five different cases according to Figure 4. The function  $t \mapsto G_{\mathbf{z}}(t)$  will be our desired quantile function. If  $c_1 = 1$ , let  $\hat{k}, \tilde{k} \in \mathbb{R}$  be any two constants satisfying  $\tilde{k} < \hat{k}$ . If  $c_1 > 1$ , let  $\xi \in (-\infty, 1)$  be such that  $(1 - \xi)^{-1/\xi} = c_1$ ,

$$k(\varepsilon) = \begin{cases} \frac{1}{\xi}(\varepsilon^{-\xi} - 1), & \xi \neq 0, \\ -\log(\varepsilon), & \xi = 0, \end{cases}$$

and  $k = \int_0^{\varepsilon_1} k(\varepsilon)d\varepsilon$ . We first claim that the function  $G_{\mathbf{z}}$  can be any arbitrary continuous and decreasing function on  $[c_2\varepsilon_2, 1)$  since the values of  $\text{VaR}_t(X)$  for  $t \in [c_2\varepsilon_2, 1)$  do not affect its PELVE at  $\varepsilon_1$  and  $\varepsilon_2$ . The value of  $G_{\mathbf{z}}$  on  $(0, c_2\varepsilon_2]$  is given by

(i) Case 1,  $c_2 = 1$  (which implies  $c_1 = 1$ ):  $G_{\mathbf{z}}(\varepsilon) = \hat{k}$ ;

(ii) Case 2,  $c_1 = 1$  and  $1 < c_2 \leq 1/\varepsilon_2$ :

$$G_{\mathbf{z}}(\varepsilon) = \begin{cases} \hat{k}, & \varepsilon \in (0, \varepsilon_1), \\ a_1\varepsilon + b_1, & \varepsilon \in [\varepsilon_1, \varepsilon_2), \\ a_2\varepsilon + b_2, & \varepsilon \in [\varepsilon_2, c_2\varepsilon_2], \end{cases} \quad \text{where} \quad \begin{cases} a_1 = \frac{\tilde{k} - \hat{k}}{\varepsilon_2 - \varepsilon_1}, \\ b_1 = \hat{k} - a_1\varepsilon_1, \\ a_2 = \frac{(\tilde{k} - \hat{k})(\varepsilon_1 + \varepsilon_2)}{(c_2\varepsilon_2 - \varepsilon_2)^2}, \\ b_2 = \tilde{k} - a_2\varepsilon_2; \end{cases}$$

(iii) Case 3,  $\varepsilon_2/\varepsilon_1 < c_1 \leq 1/\varepsilon_1$  and  $c_2 = c_1\varepsilon_1/\varepsilon_2$ :

$$G_{\mathbf{z}}(\varepsilon) = \begin{cases} k(\varepsilon), & \varepsilon \in (0, \varepsilon_1), \\ k(\varepsilon_1), & \varepsilon \in [\varepsilon_1, \varepsilon_2), \\ a\varepsilon + b, & \varepsilon \in [\varepsilon_2, c_2\varepsilon_2], \end{cases} \quad \text{where} \quad \begin{cases} a = \frac{2(k(\varepsilon_1)\varepsilon_1 - k)}{(c_2\varepsilon_2 - \varepsilon_2)^2}, \\ b = k(\varepsilon_1) - a\varepsilon_2; \end{cases}$$

(iv) Case 4,  $1 < c_1 \leq \varepsilon_2/\varepsilon_1$  and  $1 < c_2 \leq 1/\varepsilon_2$ :

$$G_{\mathbf{z}}(\varepsilon) = \begin{cases} k(\varepsilon), & \varepsilon \in (0, c_1\varepsilon_1), \\ a_1\varepsilon + b_1, & \varepsilon \in [c_1\varepsilon_1, \varepsilon_2), \\ a_2\varepsilon + b_2, & \varepsilon \in [\varepsilon_2, c_2\varepsilon_2], \end{cases} \quad \text{where} \quad \begin{cases} a_1 = -(c_1\varepsilon_1)^{-\xi-1}, \\ b_1 = k(c_1\varepsilon_1) - a_1c_1\varepsilon_1, \\ a_2 = \frac{a_1(\varepsilon_2^2 - (c_1\varepsilon_1)^2) + 2(k(c_1\varepsilon_1) - k(\varepsilon_1))c_1\varepsilon_1}{(c_2\varepsilon_2 - \varepsilon_2)^2}, \\ b_2 = a_1\varepsilon_2 + b_1 - a_2\varepsilon_2; \end{cases}$$

(v) Case 5,  $\varepsilon_2/\varepsilon_1 < c_1 \leq 1/\varepsilon_1$  and  $c_1\varepsilon_1/\varepsilon_2 < c_2 \leq 1/\varepsilon_2$ :

$$G_{\mathbf{z}}(\varepsilon) = \begin{cases} k(\varepsilon), & \varepsilon \in (0, \varepsilon_1), \\ a_1\varepsilon + b_1, & \varepsilon \in [\varepsilon_1, \varepsilon_2), \\ a_1\varepsilon_2 + b_1, & \varepsilon \in [\varepsilon_2, c_1\varepsilon_1), \\ a_2\varepsilon + b_2, & \varepsilon \in [c_1\varepsilon_1, c_2\varepsilon_2], \end{cases} \quad \text{where} \quad \begin{cases} a_1 = \frac{k(\varepsilon_1)\varepsilon_1 - k}{(\varepsilon_2 - \varepsilon_1)(c_1\varepsilon_1 - 1/2(\varepsilon_1 + \varepsilon_2))}, \\ b_1 = k(\varepsilon_1) - a_1\varepsilon_1, \\ a_2 = \frac{2c_1\varepsilon_1(a_1\varepsilon_2 + b_1 - k(\varepsilon_1))}{(c_1\varepsilon_1 - c_2\varepsilon_2)^2}, \\ b_2 = a_1\varepsilon_2 + b_1 - a_2c_1\varepsilon_1. \end{cases}$$

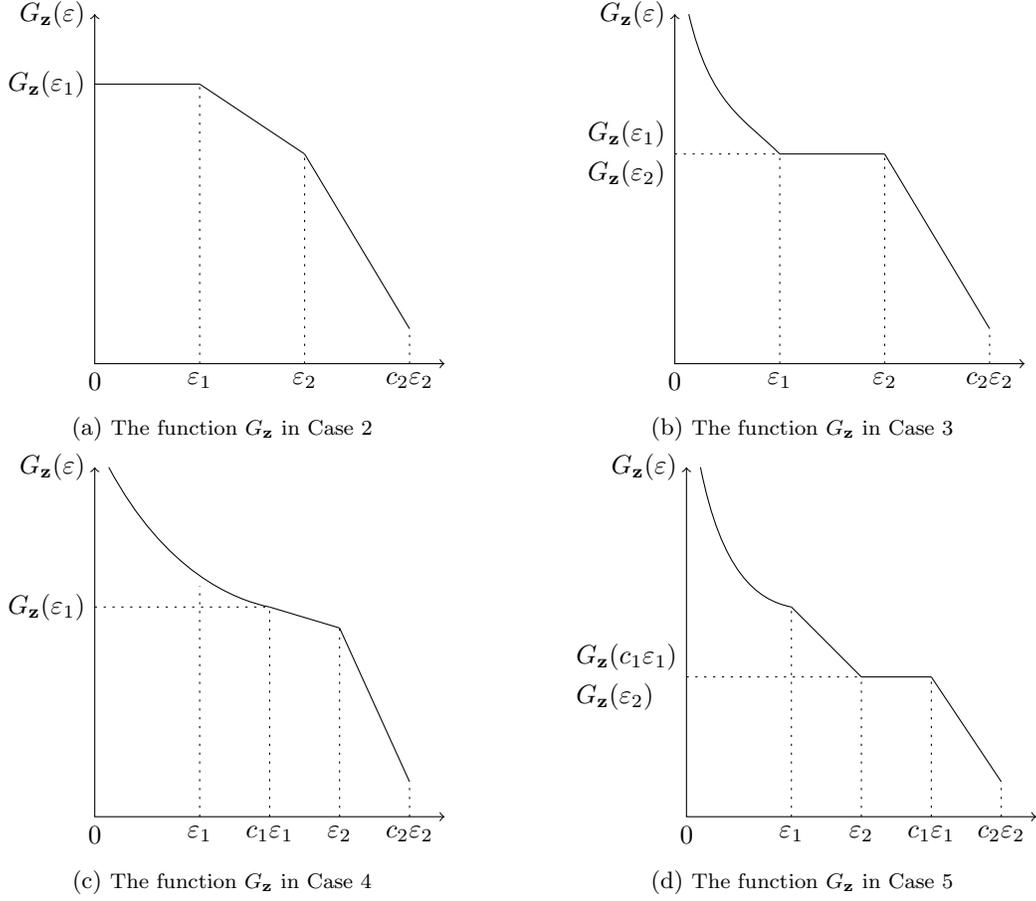
An illustration of the functions  $G_{\mathbf{z}}$  on  $[0, c_2\varepsilon_2]$  in Case 2 to Case 5 is presented in Figure 5, and we omit Case 1 in which  $G_{\mathbf{z}}$  is a constant function on  $[0, c_2\varepsilon_2]$ .

**Theorem 1.** For  $\mathbf{z} = (\varepsilon_1, c_1, \varepsilon_2, c_2) \in \Delta$ , the random variable  $X$  with a continuous quantile function given by  $t \mapsto \text{VaR}_t(X) = G_{\mathbf{z}}(t)$  satisfies  $\Pi_X(\varepsilon_1) = c_1$  and  $\Pi_X(\varepsilon_2) = c_2$ .

*Remark 1.* As we can see from Figure 5, some parts of the calibrated quantile function may be flat, corresponding to the existence of atoms in the distribution. This may be considered as undesirable from a modeling perspective, and indeed it is forced by the boundary cases of  $(\varepsilon_1, c_1, \varepsilon_2, c_2) \in \Delta$  in Figure 4. The flat parts in Cases 1 to 3 are necessary due to Propositions 6. On the other hand, the flat part in Case 5 can be replaced by a strictly decreasing function. For instance, we can replace the flat part with a strictly decreasing linear segment as long as  $c_2$  satisfies the bounds shown in Propositions 7 in Appendix A. Another way is to set  $\text{VaR}_\varepsilon(X)$  as  $k(\varepsilon)$  for  $\varepsilon \in (0, c_1\varepsilon_1)$  if  $c_2 \leq \left( c_1\varepsilon_1(\varepsilon_1^{-\xi} - (c_1\varepsilon_1)^{-\xi}) \right) / \left( \varepsilon_2^{-\xi} - (c_1\varepsilon_1)^{-\xi} \right)$ , and this choice is applied in the numerical examples in the Introduction and Section 6. The interested reader can see Propositions 6 and 7 in Appendix A, where we show that a strictly decreasing quantile function cannot attain the boundary cases  $(\varepsilon_1, c_1, \varepsilon_2, c_2)$ , and hence the flat parts are necessary to include and unify these cases.

We can easily get the distribution of  $X$  from  $\text{VaR}_\varepsilon(X)$ . As the PELVE is scale-location invariant, we can scale or move the distribution we get to match more information. For example, if  $\text{VaR}_{\varepsilon_1}(X)$  and  $\text{VaR}_{\varepsilon_2}(X)$  are given, we can choose two constants  $\lambda$  and  $\mu$  such that  $\lambda X + \mu$  matches the specified VaR values. In a similar spirit, the calibration problem can be extended to calibrate the distributions

Figure 5: An illustration of  $G_{\mathbf{z}}$  in cases 2 to 5



from some given ES and VaR values. The two points calibration problem can be regarded as given two ES and VaR values. Calibrating from only ES or VaR would be easy. However, the choices of ES values will also be limited by VaR values if we consider them at the same time, which is the same as the choice of  $c_1, c_2$  as we discussed in this section.

### 3.3 Calibration from an n-point constraint

As we see above, the PELVE calibration problem is quite technical even when only two points on the PELVE curve are given. By extending the constraint to more than two points, the problem will in general become much more complicated. We briefly discuss this problem in this section.

For the  $n$ -point constraint problem, we first need to figure out the admissible set for  $(\varepsilon_i, c_i)_{i=1, \dots, n}$ . By Lemma 1, the admissible set for the  $n$ -point calibration problem is a subset of

$$\{(\varepsilon_i, c_i)_{i=1, \dots, n} : 0 < \varepsilon_1 < \dots < \varepsilon_n < 1, c_1, \dots, c_n \geq 1, 0 < c_1\varepsilon_1 \leq \dots \leq c_n\varepsilon_n \leq 1\}.$$

However, it is not clear whether each point in the above set is admissible. There are other constraints for the admissible points such as Proposition 7. Once the admissible set is determined, we need to divide the admissible set according to the position of  $\varepsilon_i$  and  $c_i\varepsilon_i$ ,  $i = 1, \dots, n$ . Furthermore, the case  $c_i = 1$  and  $c_i\varepsilon_i = c_j\varepsilon_j$  for  $i, j = 1, \dots, n$  need special attention as Cases 1, 2 and 3 in the two-point constraint problem. For instance, in the three-point constraint problem, we need to discuss over 10 separate cases.

Below, we only discuss some special cases of  $(\varepsilon_i, c_i)_{i=1, \dots, n}$ . First, if  $c_n = 1$ , then the problem becomes trivial, as the calibrated quantile functions satisfy  $\text{VaR}_t(X) = \hat{k}$  for some  $\hat{k} \in \mathbb{R}$  in  $[0, c_n\varepsilon_n]$ .

For the case  $c_k\varepsilon_k > \varepsilon_k \geq c_{k-1}\varepsilon_{k-1}$  for  $k = 3, \dots, n$ , we can set the calibrated quantile function in  $(0, c_n\varepsilon_n]$  recursively. This is because such a configuration of  $(\varepsilon_i, c_i)_{i=1, \dots, n}$  allows for separation of the constraints, in the sense that we can adjust the values of  $\text{VaR}_t$  for  $t \in [\varepsilon_k, c_k\varepsilon_k]$  to match PELVE at  $\varepsilon_k$  without disturbing  $\text{VaR}_t$  for  $t \leq c_{k-1}\varepsilon_{k-1}$ . Let  $\text{VaR}_t^k(X)$  be the calibrated quantile function from the  $k$ -point constraint problem for  $k = 2, \dots, n$  where  $\text{VaR}_t^2(X)$  follows Theorem 1. The calibrated quantile function for the  $n$ -point constraint problem is

$$\text{VaR}_t^k(X) = \begin{cases} \text{VaR}_t^{k-1}(X), & t \in [0, c_{k-1}\varepsilon_{k-1}], \\ a_{k-1}t + b_{k-1}, & t \in (c_{k-1}\varepsilon_{k-1}, \varepsilon_k], \\ a_k t + b_k, & t \in (\varepsilon_k, c_k\varepsilon_k], \end{cases} \quad \text{where} \quad \begin{cases} a_k = \frac{a_{k-1}(\varepsilon_k^2 + c_{k-1}^2\varepsilon_{k-1}^2 - 2c_{k-1}\varepsilon_{k-1}^2)}{(c_k\varepsilon_k - \varepsilon_k)^2}, \\ b_k = a_{k-1}\varepsilon_k + b_{k-1} - a_k\varepsilon_k. \end{cases}$$

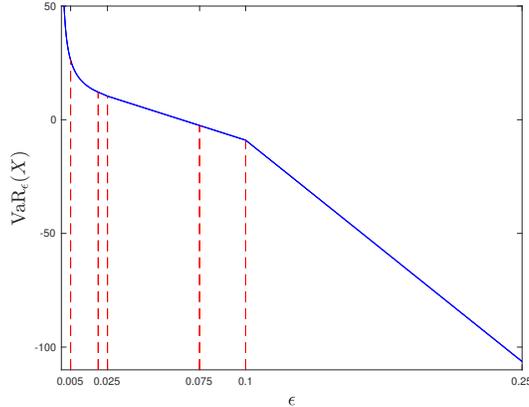
In particular, for  $n = 3$ , and assuming  $c_3\varepsilon_3 > \varepsilon_3 \geq c_2\varepsilon_2$ , the calibrated function is given by, with  $\mathbf{z} = (\varepsilon_1, c_1\varepsilon_1, \varepsilon_2, c_2\varepsilon_2) \in \Delta$ ,

$$\text{VaR}_t(X) = \begin{cases} G_{\mathbf{z}}(t), & t \in [0, c_2\varepsilon_2], \\ a_2 t + b_2, & t \in (c_2\varepsilon_2, \varepsilon_3], \\ a_3 t + b_3, & t \in (\varepsilon_3, c_3\varepsilon_3], \end{cases} \quad \text{where} \quad \begin{cases} a_2 = \frac{a_1(\varepsilon_2^2 - (c_1\varepsilon_1)^2) + 2(k(c_1\varepsilon_1) - k(\varepsilon_1))c_1\varepsilon_1}{(c_2\varepsilon_2 - \varepsilon_2)^2}, \\ b_2 = -(c_1\varepsilon_1)^{-\xi-1}(\varepsilon_2 - c_1\varepsilon_1) - k(c_1\varepsilon_1) \\ a_3 = \frac{a_2(\varepsilon_3^2 + c_2^2\varepsilon_2^2 - 2c_2\varepsilon_2^2)}{(c_3\varepsilon_3 - \varepsilon_3)^2}, \\ b_3 = a_2\varepsilon_3 + b_2 - a_3\varepsilon_3. \end{cases}$$

In Figure 6, we show the calibrated quantile function for the case  $(\varepsilon_1, \varepsilon_2, \varepsilon_3) = (0.005, 0.025, 0.1)$  and  $(c_1, c_2, c_3) = (4, 3, 2.5)$ . Note that the condition  $c_2\varepsilon_2 \leq \varepsilon_3$  is needed here.

Although we cannot solve the  $n$ -point constraint problem in general, we can instead discuss calibration from a given PELVE curve, which is the problem addressed in the next section.

Figure 6: Calibrated quantile function when  $(\varepsilon_1, \varepsilon_2, \varepsilon_3) = (0.005, 0.025, 0.1)$  and  $(c_1, c_2, c_3) = (4, 3, 2.5)$



## 4 Calibration from a curve constraint

By the location-scale invariance properties of the PELVE, we know that the solution cannot be unique. Conversely, it would be interesting to ask whether all solutions can be linearly transformed from a particular solution; that is, for a given function  $\varepsilon \mapsto \Pi(\varepsilon)$ , whether the set  $\{X \in \mathcal{X} : \Pi_X = \Pi\}$  is a location-scale class. This question, as well as identifying  $X$  satisfying  $\Pi_X = \Pi$ , is the main objective of this section.

### 4.1 PELVE and dual PELVE

First, we note that calibrated distributions from an entire PELVE curve  $\varepsilon \mapsto \Pi(\varepsilon)$  on  $(0, 1)$  would be unnatural, because the existence of the PELVE requires  $\mathbb{E}[X] \leq \text{VaR}_\varepsilon(X)$  which may not hold for  $\varepsilon$  not very small. Thus, the PELVE curve  $\Pi_X$  does not behave well on some parts of  $(0, 1)$ . To address this issue, we introduce a new notion called the dual PELVE and an integral equation which can help us to calibrate the distribution by differential equations. The dual PELVE is defined by moving the multiplier in PELVE from the ES side to the VaR side.

**Definition 1.** For  $X \in L^1$ , the dual PELVE function of  $X$  at level  $\varepsilon \in (0, 1]$  is defined as

$$\pi_X(\varepsilon) = \inf \{d \geq 1 : \text{ES}_\varepsilon(X) \leq \text{VaR}_{\varepsilon/d}(X)\}, \quad \varepsilon \in (0, 1].$$

The existence and uniqueness of  $\pi_X(\varepsilon)$  can be shown in the same way as the existence and uniqueness of the PELVE. There are advantages and disadvantages of working with both notions; see [Li and Wang \(2022, Remark 2\)](#). In our context, the main advantage of using the dual PELVE is that  $\pi_X(\varepsilon)$  is finite for all  $\varepsilon \in (0, 1]$ , while  $\Pi_X(\varepsilon)$  is finite only when  $\mathbb{E}[X] \leq \text{VaR}_\varepsilon(X)$ .

Note that for  $X$  with a discontinuous quantile function, there may not exist  $d$  such that  $\text{ES}_\varepsilon(X) =$

$\text{VaR}_{\varepsilon/d}(X)$ . In order to guarantee the above equivalence, we make the following assumption for the quantile function, represented by general function  $f$ .

**Assumption 1.** The function  $f$  is strictly decreasing and continuous, and  $\int_0^1 |f(s)| ds < \infty$ .

Let  $\mathcal{X}$  be the set of  $X \in L^1$  with quantile function satisfying Assumption 1. The requirement that the quantile function of  $X$  is continuous and strictly decreasing is equivalent to that the distribution function is continuous and strictly increasing in  $(\text{ess-inf}(X), \text{ess-sup}(X))$ ; see [Embrechts and Hofert \(2013\)](#). We limit our discussion to random variables  $X \in \mathcal{X}$ , which include the most common models in risk management.

**Proposition 2.** For  $X$  with quantile function satisfying Assumption 1 and  $\varepsilon \in (0, 1)$ , we have  $\Pi_X(\varepsilon/\pi_X(\varepsilon)) = \pi_X(\varepsilon)$  and  $\pi_X(\Pi_X(\varepsilon)\varepsilon) = \Pi_X(\varepsilon)$  if  $\mathbb{E}[X] \leq \text{VaR}_\varepsilon(X)$ . Furthermore,  $\pi_X(\varepsilon)$  is the unique solution  $d \geq 1$  to the equation

$$\text{ES}_\varepsilon(X) = \text{VaR}_{\varepsilon/d}(X).$$

It is straightforward to verify Proposition 2. By Proposition 2, we can calibrate the distribution functions from dual PELVE instead of PELVE, and the calibrated distributions should satisfy the equation  $\text{ES}_\varepsilon(X) = \text{VaR}_{\varepsilon/d}(X)$ .

## 4.2 An integral equation associated with dual PELVE

In order to calibrate distributions from the dual PELVE, we can equivalently focus on quantile functions. Let us consider  $X \in \mathcal{X}$  and  $f(s) = \text{VaR}_s(X)$ . Then, solving  $\pi_X(\varepsilon)$  is the same as solving  $z$  in following equation:

$$\int_0^y f(s) ds = yf(zy) \quad (3)$$

for  $y = \varepsilon$ . The solution is  $z = 1/\pi_X(y)$ . As  $f(s) = \text{VaR}_s(X)$ ,  $f$  satisfies Assumption 1. Denote by  $\mathcal{C}$  the set of all  $f$  satisfying Assumption 1. For any  $f \in \mathcal{C}$ , the existence of the solution  $z$  is guaranteed by the mean-value theorem and its uniqueness is obvious. For  $y \in (0, 1]$ , let  $z_f(y)$  be the solution to (3) associated with  $f$ . Clearly,  $z_f(y) \leq 1$  and  $y \mapsto yz_f(y)$  is strictly increasing. This is similar to Lemma 1 for the two-point case. Obviously,  $z_f(y)$  is also location-scale invariant under linear transformation on  $f \in \mathcal{C}$ . That is,  $z_{\lambda f + b} = z_f$  for  $\lambda > 0$  and  $b \in \mathbb{R}$ . Furthermore,  $z_f$  is continuous as  $f$  is continuous and strictly decreasing. The next proposition is a simple connection between  $z_f$  and  $\pi_X$ .

**Proposition 3.** For any  $f$  satisfying Assumption 1,  $X = f(U)$  for some  $U \sim U(0, 1)$  has the dual PELVE  $\pi_X(y) = 1/z_f(y)$  for all  $y \in (0, 1)$  where  $z_f$  is solution to (3). For  $X$  with quantile function satisfying Assumption 1, there exists  $f$  satisfying Assumption 1 such that  $X = f(U)$  for some  $U \sim U(0, 1)$  and the solution to (3) is  $z_f(y) = 1/\pi_X(y)$  for all  $y \in (0, 1)$ .

*Proof.* For any  $f$  satisfying Assumption 1, let  $F(x) = 1 - f^{-1}(x)$ . Hence,  $F$  is a continuous and strictly increasing distribution function and  $F^{-1}(s) = f(1 - s)$  for  $s \in (0, 1)$ . Let  $U \sim \text{U}(0, 1)$  and  $X = F^{-1}(U) = f(1 - U)$ . Then  $X \in \mathcal{X}$  and  $X \sim F$ . As  $F^{-1}(1 - s) = f(s)$ , we have  $\pi_X(y) = 1/z_f(y)$ . Take  $U' = 1 - U$ . We have  $X = f(U')$  and  $U' \sim \text{U}(0, 1)$ .

For  $X \in \mathcal{X}$ , let  $f(s) = \text{VaR}_s(X)$ . Then, we have  $z_f(y) = 1/\pi_X(y)$  for  $y \in (0, 1]$ . Furthermore, we have  $F^{-1}(s) = f(1 - s)$ . Therefore, there exists  $U \sim \text{U}(0, 1)$  such that  $X = f(1 - U)$ . Let  $U' = 1 - U$ . Then, we have  $X = f(U')$  and  $U' \sim \text{U}(0, 1)$ .  $\square$

Proposition 3 allows us to study  $z$  instead of  $\pi$  for the calibration problem. The integral equation (3) can be very helpful in characterizing the distribution from the dual PELVE.

Some examples of  $\pi_X$  and  $z_f$  are listed in Table 2, which is corresponding to the PELVE presented in Table 1.

Table 2: Example of  $\pi_X$  and  $z_f$

$X$	$\pi_X(\varepsilon)$	$f$	$z_f$
$\text{U}(0, 1)$	$\pi_X(\varepsilon) = 2$	$f(x) = 1 - x$	$z_f(y) = 1/2$
$\text{Exp}(\lambda)$	$\pi_X(\varepsilon) = e$	$f(x) = -\log(x)/\lambda$	$z_f(y) = 1/e$
$\text{GPD}(\xi)$	$\pi_X(\varepsilon) = (1 - \xi)^{-\frac{1}{\xi}}$	$f(x) = \begin{cases} 1/\xi (x^{-\xi} - 1) & \xi \neq 0 \\ -\log(x) & \xi = 0 \end{cases}$	$z_f = (1 - \xi)^{\frac{1}{\xi}}$

For a given dual PELVE curve  $\pi$ , we find the solution to the integral equation by the following steps.

1. Let  $z(y) = \frac{1}{\pi(y)}$  for all  $y \in (0, 1]$ .
2. Find  $f \in \mathcal{C}$  that satisfies  $\int_0^y f(s)ds = yf(z(y)y)$  for all  $y \in (0, 1]$ .
3. By Proposition 3,  $X = f(U)$  for some  $U \sim \text{U}(0, 1)$  will have the given dual PELVE  $\pi$ .

Therefore, we will focus on characterizing  $f$  from a given  $z : (0, 1] \rightarrow (0, 1]$  below. Generally, it is hard to characterize  $f$  explicitly. We first formulate the problem as an advanced differential equation, which helps us to find solutions.

### 4.3 Advanced differential equations

In this section, we show that the main objective (3) can be represented by a differential equation. The use of differential equations in computing risk measures has not been actively developed. The only paper we know is Balbás et al. (2020) which addresses a different problem.

Let us recall the integral equation (3) from Section 4.2. For a function  $f \in \mathcal{C}$ , we solve the function  $z_f : (0, 1) \rightarrow \mathbb{R}$  from (3). We represent (3) by an advanced differential equation using the following steps.

1. Let  $\omega_f(y) = yz_f(y)$ . It is easy to see that  $z_f(y) \leq 1$ . Hence,  $\omega_f$  is strictly increasing and continuous on  $(0, 1]$  and  $\omega_f(y) \leq y$ .
2. Let  $u_f$  be the inverse function of  $\omega_f$ . We have that  $u_f : (0, z_f(1)] \mapsto (0, 1]$  is a continuous and strictly increasing function and  $u_f(w) \geq w$ .
3. Replacing  $y$  with  $u_f(w)$  in (3), we have  $f(w) = \int_0^{u_f(w)} f(s) ds / u_f(w)$ .
4. Assume  $u_f$  is continuously differentiable. It is clear that  $f$  is continuously differentiable on  $(0, z_f(1))$ . Hence, we can represent (3) by the following advanced differential equation

$$f'(w) + \frac{u_f'(w)}{u_f(w)} (f(w) - f(u_f(w))) = 0.$$

For a given function  $z : (0, 1] \rightarrow \mathbb{R}$ , let  $u = \omega^{-1}$  such that  $\omega(y) = yz(y)$  for  $y \in (0, 1]$ . Then, we solve the function  $f$  by the following differential equation

$$f'(w) + \frac{u'(w)}{u(w)} f(w) - \frac{u'(w)}{u(w)} f(u(w)) = 0. \quad (4)$$

If  $z = 1/\pi_X$  for some  $X \in \mathcal{X}$ , then  $u$  is a strictly increasing and continuous function such that  $u(w) \geq w$ . Furthermore, if  $z$  is continuously differentiable, then we can characterize all  $X \in \mathcal{X}$  with  $\pi_X = 1/z$  by (4). As  $u'(w)/u(w) \geq 0$  and  $u(w) \geq w$ , (4) is a linear advanced differential equation which is well studied in the literature. In [Berezansky and Braverman \(2011\)](#), it is shown that there exists a non-oscillatory solution for (4).

#### 4.4 The constant PELVE curve

We first solve the case that  $z(y) = c$  for all  $y \in (0, 1]$  and some constant  $c \in (0, 1)$ . As we can see from Table 2, the power function and logarithm function have constant  $z_f$ . If  $f(x) = \lambda x^\alpha + b$  for  $\alpha > -1$ , we can see that  $(\alpha + 1)^{-1/\alpha} = c$ . In this section, we can characterize all the other solutions which can not be expressed as a linear transformation of the power function. That is, we will see that the set

$$\{f \in \mathcal{C} : z_f(y) = z(y), y \in (0, 1]\}$$

is not a location-scale class. Hence, we can answer the question at the beginning of the section; that is, in the case the PELVE is a constant, the set  $\{X \in \mathcal{X} : \Pi_X = c\}$  is not a location-scale class.

**Theorem 2.** For  $c \in (0, 1)$ , any  $X$  with quantile function satisfying Assumption 1 and  $\pi_X(\varepsilon) = 1/c$  for  $\varepsilon \in (0, 1)$  can be written as  $X = f(U)$  for some  $U \sim U(0, 1)$  and  $f$  satisfying Assumption 1. Furthermore, such  $f$  has the form

$$f(y) = C_1 + C_2 y^\alpha + O(y^\zeta),$$

where  $\alpha$  is the root of  $(\alpha + 1)^{-1/\alpha} = c$ ,  $\zeta > \max\{0, \alpha\}$ ,  $C_1, C_2 \in \mathbb{R}$ ,  $C_2\alpha < 0$  and  $O(y^\zeta)$  is a function such that  $\limsup_{y \rightarrow 0} O(y^\zeta)/y^\zeta$  is a constant.

The proof of Theorem 2 is provided in Appendix B. As we can see, Theorem 2 characterizes all  $X \in \mathcal{X}$  such that  $\pi_X(\varepsilon) = 1/c$ . If  $c \in (0, 1/e)$ ,  $\alpha$  is negative. As  $\zeta > 0$ , we can see that  $X = f(U)$  is regularly varying of index  $\alpha$ . Hence, one can then consider the Pareto distribution with survival function  $S(x) = x^\alpha$  as a representative solution for the tail behavior. An open question is that, in the general case that the PELVE is not necessarily constant, whether all the solutions behave similarly regarding their tail behavior.

Another interesting implication of the theorem and its proof is that one can give a non-trivial solution for  $z$  is a constant.

**Example 3.** For  $c \in (0, 1)$ , let  $(\theta, \eta)$  be a solution of

$$\begin{cases} c \log c = -\frac{\eta \exp\left(-\frac{\eta}{\tan(\eta)}\right)}{\sin(\eta)}, \\ \theta = -\frac{\eta}{\tan(\eta)}. \end{cases}$$

Then, the function  $f$ , given by

$$f(y) = C_1 + C_2 y^\alpha + C_3 y^\zeta \sin(-\sigma \log(y)), \quad 0 < y < 1, \quad (5)$$

satisfies  $\int_0^y f(s) ds = yf(cy)$  and Assumption 1, where  $\alpha$  solves  $(\alpha + 1)^{-1/\alpha} = c$ ,  $\zeta = \theta/\log c - 1$ ,  $\sigma = -\eta/\log c$ ,  $C_2$  is a constant such that  $C_2\alpha < 0$  and  $0 < C_3 < -C_2\alpha/(\zeta + |\sigma|)$ .

If we take  $C_3 = 0$ , we get the simplest power function for  $z(x) = c$ . If  $C_3 \neq 0$ , the solution (5) is not a linear transformation of the power function solution.

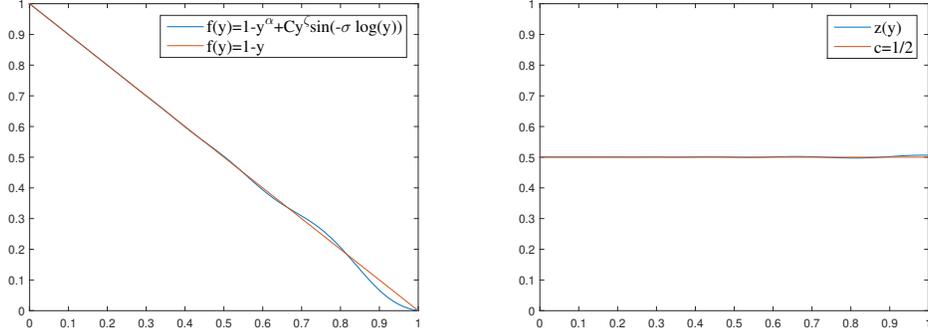
Let us look at the example where  $\pi(\varepsilon) = 2$  for all  $\varepsilon \in (0, 1]$ , which means  $z(y) = 1/2$  for  $y \in (0, 1]$ . As we have seen in Table 2,  $f(y) = 1 - y$  can be a solution that leads to  $X \sim U(0, 1)$ . Furthermore, according to Example 3, we can have another solution

$$f(y) = 1 - y^\alpha + C y^\zeta \sin(-\sigma \log(y)),$$

where  $\alpha = 1$ ,  $C = 0.05096$ ,  $\zeta = 4.0184$  and  $\sigma = -15.4090$ . In the left of Figure 7, we have depicted

the two solutions for  $f$ . We can see they are quite different when  $y$  goes to 1. In the right of Figure 7, we numerically calculate  $z_f$  for  $f(y) = 1 - y^\alpha + Cy^\zeta \sin(-\sigma \log(y))$ . We can see its numerical value is almost  $1/2$  and the discrepancy is due to limited computational accuracy.

Figure 7: Non-unique calibrated functions for  $z(y) = 1/2$ .



By letting  $X = f(U)$ , we get  $\pi_X(\varepsilon) = 2$  for all  $\varepsilon \in (0, 1]$  and such  $X$  does not follow the uniform distribution.

## 4.5 A numerical method

In general, it is hard to get an explicit solution to (4). Here we present a numerical method to solve (4). Let us introduce the following process.

1. Let  $a_0 = 1, a_1 = a, \dots, a_n = u^{-1}(a_{n-1})$ .
2. For  $a \in (0, 1)$ , let  $\xi$  be the solution to  $(1 - \xi)^{\frac{1}{\xi}} = a$ . Let

$$f_0(x) = \begin{cases} \frac{1}{\xi} (x^{-\xi} - 1), & \xi \neq 0, \\ -\log(x), & \xi = 0, \end{cases} \quad (6)$$

on  $[a, 1]$ .

3. We can solve the following ODE on  $[a_2, a_1]$ :

$$f_1'(w) + \frac{u'(w)}{u(w)} f_1(w) = \frac{u'(w)}{u(w)} f_0(u(w)), \quad w \in [a_2, a_1].$$

4. Now we can repeat step 3 by induction on  $[a_{n+1}, a_n]$  for  $n > 1$  by solving

$$f_n'(w) + \frac{u'(w)}{u(w)} f_n(w) = \frac{u'(w)}{u(w)} f_{n-1}(u(w)), \quad w \in [a_{n+1}, a_n].$$

5. In general, the solution for differential equation  $\frac{dy}{dx} + P(x)y = Q(x)$  is

$$y = e^{-\int^x P(\lambda) d\lambda} \left[ \int^x e^{\int^\lambda P(\varepsilon) d\varepsilon} Q(\lambda) d\lambda + C \right].$$

So, we get the following solution for  $f_n$ :

$$f_n(w) = e^{\int_w^{a_n} \frac{u'(\lambda)}{u(\lambda)} d\lambda} \left[ f_{n-1}(a_n) - \int_w^{a_n} e^{-\int_\lambda^{a_n} \frac{u'(\varepsilon)}{u(\varepsilon)} d\varepsilon} \frac{u'(\lambda)}{u(\lambda)} f_{n-1}(u(\lambda)) d\lambda \right], \quad w \in [a_{n+1}, a_n].$$

6. Finally, let  $f = f_n$  on  $[a_{n+1}, a_n]$ .

Note that since we start with a strictly decreasing function, then from equation (4) we have

$$f'(w) = \frac{u'(w)}{u(w)} (f(u(w)) - f(w)) < 0,$$

so  $f$  remains strictly decreasing.

The solution produced by the numerical method heavily relies on  $f_0$ . The equation (4) does not have a unique solution, but the solution from the above process is unique. We set  $f_0$  as (6) by assuming  $z$  can be extended from  $(0, 1]$  to  $\mathbb{R}^+$  and set  $z(y) = a$  for all  $y > 1$ . We use this assumption for simplification as we can know that (6) satisfies (4) for a constant  $z$  from Section 4.4. This choice of  $f_0$  is the same as the choice of  $k(\varepsilon)$  in the two-point calibration problem, and this reflects our subjective view of the importance of the Pareto distribution in risk management. Especially, when  $z(y) = c$  for some constant  $c$ , we have  $u(x) = x/c$ . Therefore, (5) gives

$$f_n(w) = \frac{a_n}{w} \left[ f_{n-1}(a_n) - \int_w^{a_n} \frac{1}{a_n} f_{n-1} \left( \frac{\lambda}{c} \right) d\lambda \right].$$

If we set  $f_0$  as (6), we can have  $f_1$  also in the form of (6). Then, it is obvious that  $f_n$  is also in the form of (6). Therefore, the numerical method gives the simplest power function or logarithm function when  $z(y)$  is a constant on  $(0, 1]$  as Table 2, which leads to the generalized Pareto distribution for  $X$ .

## 4.6 Numerical calibrated quantile function

Now let us explore the method in Section 4.5 with simulation. Here we present the results for a few cases. In Figures 8 to 11, we compare the solution from the numerical method with the standard formula in Table 2 in the left panel, and compare  $\int_0^y f(s) ds$  with  $yf(z(y)y)$  to validate the equation (3) in the right panel.

We first try some examples where  $z$  is constant as shown in Table 2, i.e.  $z(x) = 1/2$  (Figure 8),  $z(x) = 1/e$  (Figure 9) and  $z(x) = 0.9^{10}$  (Figure 10). For Figure 8 to 10, we can see that the numerical method provides exactly the same function  $f$  as Table 2.

In Figure 11, we check the case  $z(x) = \log(x/(1 - e^{-x}))/x$ . The function  $f(x) = e^{-x}$  satisfies (3). We can see that the solution from the numerical method is close to a function of the form  $f(x) = \lambda e^{-x} + b$ , which is known to satisfy the integral equation.

Figure 8: Calibrated function and validation for  $z(x) = 1/2$

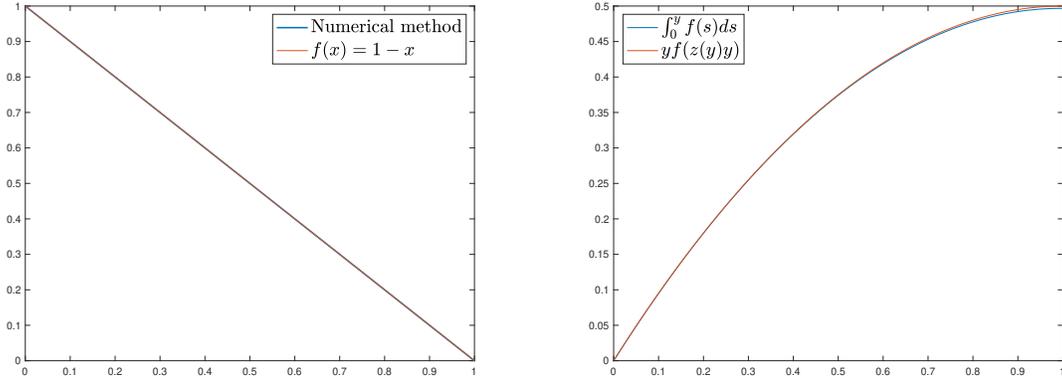
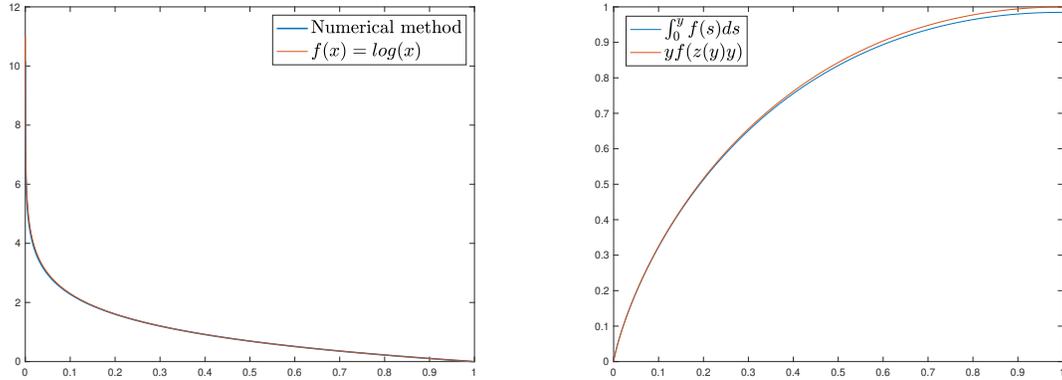


Figure 9: Calibrated function and validation for  $z(x) = 1/e$



## 5 Technical properties of the PELVE

We now take a turn to study several additional properties of PELVE. In particular, we will obtain results on the monotonicity and convergence of the dual PELVE as well as the PELVE.

### 5.1 Basic properties of dual PELVE

The following proposition that shows the PELVE and dual PELVE share some basic properties such as monotonicity (i), location-scale invariance (ii) and shape relevance (iii)-(iv) below.

Figure 10: Calibrated function and validation for  $z(x) = 0.9^{10}$

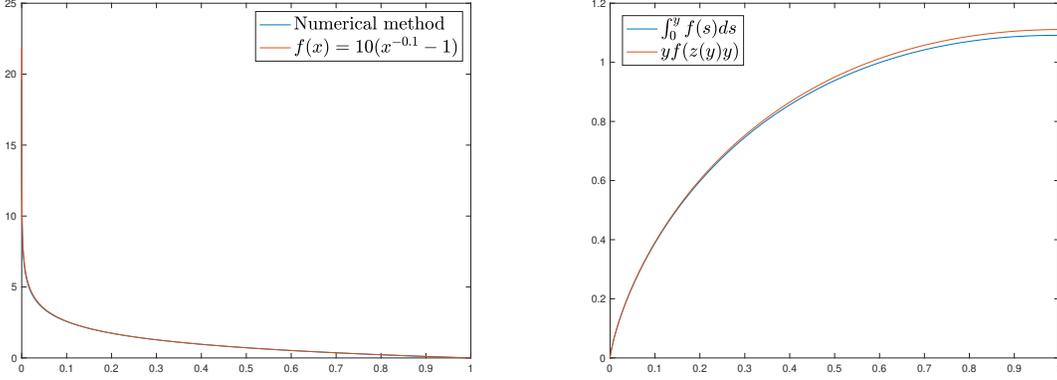
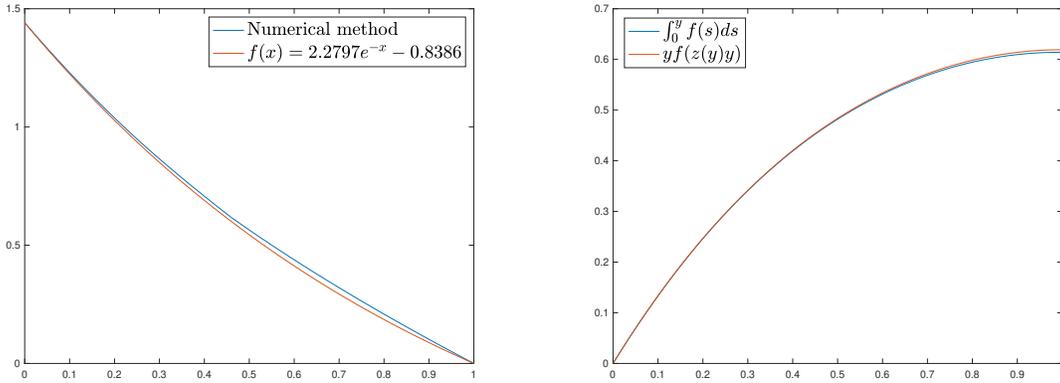


Figure 11: Calibrated function and validation for  $z(x) = \log(x/(1 - e^{-x}))/x$



**Proposition 4.** Suppose the quantile function of  $X$  satisfies Assumption 1 and  $\varepsilon \in (0, 1]$ .

- (i)  $\Pi_X(\varepsilon)$  is increasing (decreasing) in  $\varepsilon$  if and only if so is  $\pi_X(\varepsilon)$ .
- (ii) For all  $\lambda > 0$  and  $a \in \mathbb{R}$ ,  $\pi_{\lambda X+a}(\varepsilon) = \pi_X(\varepsilon)$ .
- (iii)  $\pi_{f(X)}(\varepsilon) \leq \pi_X(\varepsilon)$  for all strictly increasing concave functions:  $f: \mathbb{R} \rightarrow \mathbb{R}$  with  $f(X) \in \mathcal{X}$ .
- (iv)  $\pi_{g(X)}(\varepsilon) \geq \pi_X(\varepsilon)$  for all strictly increasing convex functions:  $g: \mathbb{R} \rightarrow \mathbb{R}$  with  $g(X) \in \mathcal{X}$ .

The statements (ii)-(iv) are parallel to the corresponding statements in Theorem 1 of Li and Wang (2022) on PELVE. The proof of Proposition 4 is put in Appendix C. Proposition 4 allows us to study the monotonicity and convergence of the PELVE by analyzing the corresponding properties of the dual PELVE, which is more convenient in many cases. In the following sections, we focus on finding the conditions which make the dual PELVE monotone and convergent at 0. By Proposition 4, those conditions can also apply to the PELVE.

## 5.2 Non-monotone and non-convergent examples

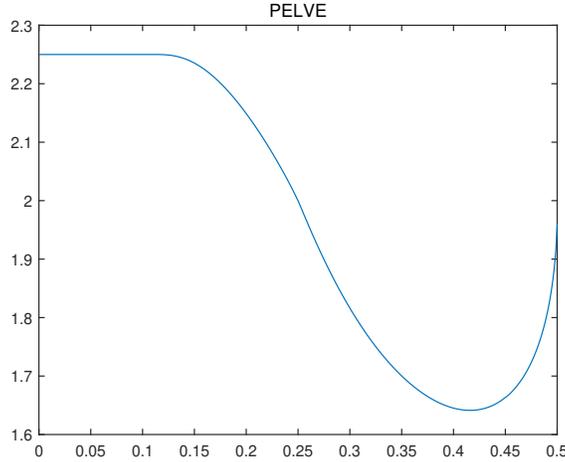
In this section, we study the monotonicity and convergence of dual PELVE. For monotonicity, we have shown some well-known distributions such as normal distribution, t-distribution and lognormal distribution have monotone PELVE curves in Example 3. However, the PELVE is not monotone for all  $X \in \mathcal{X}$ . Below we provide an example.

**Example 4** (Non-monotone PELVE). Let us consider the following density function  $g$  on  $[-2, 2]$ ,

$$g(x) = \frac{1}{2} \left( (x+2)\mathbb{1}_{\{x \in [-2, -1]\}} - x\mathbb{1}_{\{x \in (-1, 0]\}} + x\mathbb{1}_{\{x \in (0, 1]\}} + (2-x)\mathbb{1}_{\{x \in (1, 2]\}} \right).$$

For  $X$  with density function  $g$ , Figure 12 presents the value of  $\Pi_X(\varepsilon)$  for  $\varepsilon \in (0, 0.5)$ . As one can see, the PELVE is not necessarily decreasing, and so is the dual PELVE.

Figure 12: PELVE for  $X$  with density  $g$



For the convergence, it is clear that  $\pi_X(\varepsilon)$  is continuous in  $(0, 1)$  for  $X \in \mathcal{X}$ . Therefore,  $\lim_{\varepsilon \rightarrow p} \pi_X(\varepsilon)$  exists for all  $p \in (0, 1)$ . However, both  $\Pi_X(\varepsilon)$  and  $\pi_X(\varepsilon)$  are not well defined at  $\varepsilon = 0$ . If  $\lim_{\varepsilon \rightarrow 0} \pi_X(\varepsilon)$  exists, we can define  $\pi_X(0)$  as the limit, and  $\Pi_X(0)$  similarly. However, the following example shows that the limit does not exist for some  $X \in \mathcal{X}$ .

**Example 5** (No limit at 0). We can construct a random variable  $X \in \mathcal{X}$  such that  $\lim_{\varepsilon \rightarrow 0} \pi_X(\varepsilon)$  does not exist from the integral equation (3) in Section 4.2. Equivalently, we will find a continuous and strictly decreasing function  $f \in \mathcal{C}$  such that  $\lim_{y \rightarrow 0} z_f(y)$  does not exist. Let  $c$  be the Cantor ternary function on  $[0, 1]$ . Note that  $x \mapsto c(x)$  is continuous and increasing on  $(0, 1)$  and  $c(x/3) = c(x)/2$ . Let

$f(x) = -c(x) - x^{\log 2 / \log 3}$ . It is clear that  $f \in \mathcal{C}$  and  $f(x/3) = f(x)/2$ . For each  $y \in (0, 1]$ , we have

$$\begin{aligned} yf(z_f(y)y) &= \int_0^y f(x)dx \\ &= 2 \int_0^y f\left(\frac{1}{3}x\right)dx = 6 \int_0^{\frac{1}{3}y} f(x)dx = 2yf\left(\frac{1}{3}yz_f\left(\frac{1}{3}y\right)\right) = yf\left(yz_f\left(\frac{1}{3}y\right)\right). \end{aligned}$$

Since  $f$  is strictly decreasing,  $z_f(y) = z_f(y/3)$  for  $y \in (0, 1]$ . It means that  $z_f(y)$  is a constant on  $(0, 1]$  if  $\lim_{y \rightarrow 0} z_f(y)$  exists. Now, let us look at two particular points of  $z_f(y)$ . We can show that  $z_f(1) \neq z_f(4/9)$ . Let  $z = (\log 2 / \log 3 + 1)^{-(\log 3 / \log 2)}$ . Then, we have  $1/3 < z \approx 0.46 < 1/2$ . For  $y = 1$ , we have  $\int_0^1 c(s)ds = c(z) = 1/2$  and  $\int_0^c s^{\log 2 / \log 3} ds = z^{\log 2 / \log 3}$ . Therefore, we get  $z_f(1) = z < 1/2$ . For  $y = 4/9$ , we have

$$\begin{aligned} f\left(\frac{4}{9}z_f\left(\frac{4}{9}\right)\right) &= \frac{9}{4} \int_0^{4/9} f(s)ds \\ &= -\frac{9}{4} \left( \frac{1}{\frac{\log 2}{\log 3} + 1} \left(\frac{4}{9}\right)^{\frac{\log 2}{\log 3} + 1} + \frac{1}{12} + \frac{1}{2} \left(\frac{4}{9} - \frac{1}{3}\right) \right) < -0.68 < f\left(\frac{2}{9}\right) \approx -0.64. \end{aligned}$$

As  $f$  is strictly decreasing, we have  $(4/9)z_f(4/9) > 2/9$  which implies  $z_f(4/9) > 1/2 > z_f(1)$ . As a result,  $\lim_{y \rightarrow 0} z_f(y)$  does not exist. Therefore, we have a continuous and strictly decreasing  $f$  such that  $\lim_{y \rightarrow 0} z_f(y)$  does not exist.

### 5.3 Sufficient condition for monotonicity and convergence

In risk management applications, for a random variable  $X$  modeling a random loss, the behavior of its tail is the most important. Let  $F^{[p,1]}$  be the upper  $p$ -tail distribution of  $F$  (see e.g., [Liu and Wang \(2021\)](#)), namely

$$F^{[p,1]}(x) = \frac{(F(x) - p)_+}{1 - p}, \quad x \in \mathbb{R}.$$

We will see that the dual PELVE of  $F^{[p,1]}$  is a part of the dual PELVE of  $F$ .

**Lemma 2.** *Let  $F$  be the distribution function of  $X$  with quantile function satisfying Assumption 1. For  $p \in (0, 1)$  and  $X' \sim F^{[p,1]}$ , it holds*

$$\pi_{X'}(\varepsilon) = \pi_X(\varepsilon(1 - p)).$$

*Proof.* It is clear that  $\text{VaR}_\varepsilon(X') = \text{VaR}_{\varepsilon(1-p)}(X)$  and  $\text{ES}_\varepsilon(X') = \text{ES}_{\varepsilon(1-p)}(X)$ . Therefore,

$$\begin{aligned} \pi_{X'}(\varepsilon) &= \inf\{d \geq 1 : \text{ES}_\varepsilon(X') \leq \text{VaR}_{\varepsilon/d}(X')\} \\ &= \inf\{d \geq 1 : \text{ES}_{\varepsilon(1-p)}(X) \leq \text{VaR}_{\varepsilon(1-p)/d}(X)\} = \pi_X(\varepsilon(1 - p)). \end{aligned}$$

Thus, we have the desired result.  $\square$

The tail distribution can provide a condition to check whether the dual PELVE is decreasing.

**Proposition 5.** *Let  $F$  be the distribution function of  $X$  with quantile function satisfying Assumption 1. If  $x \mapsto F^{-1}((1-p)F(x) + p)$  is convex (concave) for all  $p \in (0, 1)$ , then  $\pi_X$  and  $\Pi_X$  are decreasing (increasing).*

*Proof.* For any  $p \in (0, 1)$ , let  $X' \sim F^{[p,1]}$ . By Lemma 2, we have  $\pi_{X'}(\varepsilon) = \pi_X(\varepsilon(1-p))$ . Furthermore, we have

$$\left(F^{[p,1]}\right)^{-1}(t) = F^{-1}((1-p)t + p) = F^{-1}\left((1-p)F(F^{-1}(t)) + p\right), \quad t \in [0, 1].$$

Let  $U \sim U(0, 1)$ ,  $X = F^{-1}(U)$  and  $X' = (F^{[p,1]})^{-1}(U)$ .

We assume that  $x \mapsto F^{-1}((1-p)F(x) + p)$  is a convex function on  $(\text{ess-inf}(X), \text{ess-sup}(X))$  first. Let  $f : \mathbb{R} \rightarrow \mathbb{R}$  be a strictly increasing convex function such that  $f(x) = F^{-1}((1-p)F(x) + p)$  for  $x \in (\text{ess-inf}(X), \text{ess-sup}(X))$ . Then, we have  $X' = f(X)$ . By Proposition 4, we get  $\pi_{X'}(\varepsilon) \geq \pi_X(\varepsilon)$ . As  $\pi_{X'}(\varepsilon) = \pi_X(\varepsilon(1-p))$ , we have  $\pi_X(\varepsilon(1-p)) \geq \pi_X(\varepsilon)$  for all  $p \in (0, 1)$ . Thus,  $\pi_X$  is decreasing. By Proposition 4, we have  $\Pi_X$  is also decreasing.

On the other hand, if  $x \mapsto F^{-1}((1-p)F(x) + p)$  is concave, we have  $\pi_X(\varepsilon(1-p)) \leq \pi_X(\varepsilon)$  for all  $p \in (0, 1)$  and  $\pi_X$  is increasing. So is  $\Pi_X$ .  $\square$

The condition  $x \mapsto F^{-1}((1-p)F(x) + p)$  is convex (concave) for all  $p \in (0, 1)$  is generally hard to check. Intuitively, this condition means that  $F^{[p,1]}$  has a less heavy tail compared to  $F$ . We can further simplify this condition by using the hazard rate function. For  $X \in \mathcal{X}$  with distribution function  $F$  and density function  $f$ , let  $S = 1 - F$  be the survival function and  $\eta = f/S$  be the hazard rate function. As  $F$  is continuous and strictly increasing,  $S$  is continuous and strictly decreasing.

**Theorem 3.** *For  $X$  with quantile function satisfying Assumption 1, let  $\eta$  be the hazard rate function of  $X$ . If  $1/\eta$  is second-order differentiable and convex (concave), then  $\pi_X$  and  $\Pi_X$  are decreasing (increasing).*

The proof of Theorem 3 is provided in Appendix C.

**Example 6.** For the normal distribution, we can give a short proof of the convexity of  $1/\eta$ . Let  $S$  be the survival function of the standard normal distribution and  $f$  its density. Let  $I(x) = 1/\eta(x) = S(x)/f(x) = \exp(x^2/2) \int_{-\infty}^{-x} \exp(-s^2/2) ds$ . One can easily see that

$$I'(x) = xI(x) - 1 \tag{7}$$

which gives  $I''(x) = xI'(x) + I(x)$ . This with (7) implies that

$$I''(x) = (1 + x^2)I(x) - x. \quad (8)$$

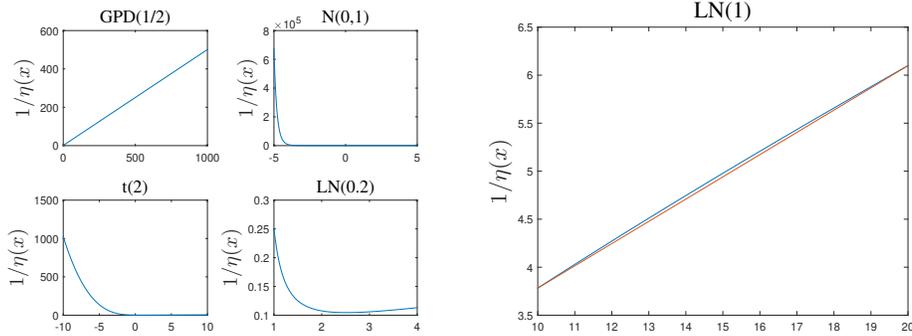
First, consider the negative line i.e.,  $x < 0$ . In this case (7) and (8) imply  $I'(x) = xI(x) - 1 < 0$ , and  $I''(x) = (1 + x^2)I(x) + (-x) > 0$ . The implication of the two relations is that  $I$  is a convex and decreasing function on negative line. Now we consider the case  $x > 0$ . In this case, let  $i(x) = I'(-x)$ . From what we have proved it is clear that  $i$  is an increasing function on  $x > 0$ . On the other hand, we have  $I(x) + I(-x) = 1/f(x) = \sqrt{2\pi} \exp(x^2/2)$ . This combined with (7) gives us

$$I'(x) = x(I(x) + I(-x)) + i(x) = x\sqrt{2\pi} \exp(x^2/2) + i(x), x > 0.$$

This means  $I'$  is an increasing function on  $x > 0$  as it is a summation of two other increasing functions, so  $I$  is convex on the positive line as well.

Figure 13 presents the curve  $1/\eta$  for the generalized Pareto distribution, the Normal distribution, the t-distribution and the Lognormal distribution. For distributions GPD(1/2), N(0, 1) and t(2), we can see that the curves  $1/\eta$  are convex, and this coincides with decreasing PELVE shown in Example 2. For the Lognormal distribution, the shape of  $1/\eta$  depends on  $\sigma$ . As shown in Example 2, the PELVE for LN( $\sigma$ ) is visibly decreasing for  $\sigma^2 = 0.04$  and increasing for  $\sigma = 1$ . Corresponding to the above observations, we see that  $1/\eta$  is convex for  $\sigma^2 = 0.04$  and concave for  $\sigma^2 = 1$ .

Figure 13:  $1/\eta$  for GPD(1/2), N(0, 1), t(2), LN(0, 2) and LN(1) in blue curves; in the right panel, the red curve is linear



**Corollary 1.** *If the hazard rate of a random variable  $X$  is second-order differentiable and concave, then  $\pi_X$  and  $\Pi_X$  are decreasing.*

*Proof.* Just note that if  $\eta$  is concave, then  $\eta\eta''$  is non-positive. It follows that

$$\left(\frac{1}{\eta}\right)'' = \left(-\frac{\eta'}{\eta^2}\right)' = \frac{2(\eta')^2 - \eta\eta''}{\eta^3} \geq 0.$$

Thus,  $1/\eta$  is convex, and the desired statement follows from Theorem 3.  $\square$

The corollary above is a result of the fact that the concavity of  $\eta$  implies convexity of  $1/\eta$ . Therefore, concave  $\eta$  always leads to decreasing PELVE. For example, the Gamma distribution  $G(\alpha, \lambda)$  with density  $f(x) = \lambda^\alpha t^{\alpha-1} e^{-\lambda t} / \Gamma(\alpha)$  has concave hazard rate function when  $\alpha > 1$ . Furthermore, by Theorem 3, we can easily find more well-known distributions that have decreasing  $\pi_X$ .

As the tail distribution determines  $\pi_X$  around 0, we can focus on the tail distribution to discuss the convergence of  $\pi_X$  at 0. Note that if the survival distribution function is regularly varying, then its tail parameter one-to-one corresponds to the limit of  $\Pi_X$  at 0 as shown by Theorem 3 of Li and Wang (2022). Hence, the limit of  $\Pi_X$ , if it exists, can be useful as a measure of tail heaviness, and it is well defined even for distributions that do not have a heavy tail. By the monotone convergence theorem, we have  $\lim_{\varepsilon \rightarrow 0} \pi_X(\varepsilon)$  exists if  $\pi_X$  is monotone. The limit may be finite or infinite.

**Corollary 2.** *For  $X$  with quantile function satisfying Assumption 1, let  $\eta$  be the hazard rate of  $X$ . If  $1/\eta(x)$  is second-order differentiable and convex (concave) in  $(F^{-1}(\delta), \text{ess-sup}(X))$  for some  $\delta \in (0, 1)$ , then  $\lim_{\varepsilon \rightarrow 0} \pi_X(\varepsilon)$  exists. In particular, this is true if  $\eta$  is second-order differentiable and concave on  $(F^{-1}(\delta), \text{ess-sup}(X))$ .*

*Proof.* Let  $X' \sim F^{[\delta, 1]}$ . Then, the survival function for  $X'$  is  $S_{X'}(x) = S(x)/(1-p)$  for  $x \geq F^{-1}(\delta)$ . The density function is  $f_{X'}(x) = f(x)/(1-p)$  for  $x \geq F^{-1}(\delta)$ . Therefore, the hazard rate function is  $\eta_{X'}(x) = f(x)/S(x) = \eta(x)$  for  $x \geq F^{-1}(\delta)$ .

As  $1/\eta(x)$  is convex (concave) when  $x > F^{-1}(\delta)$ , we have  $1/\eta_{X'}(x)$  is convex (concave). By Theorem 3, we have  $\pi_{X'}(\varepsilon)$  is decreasing (increasing) on  $(0, 1)$ . As a result, we have  $\pi_X(\varepsilon)$  is decreasing (increasing) on  $(0, \delta)$  and  $\lim_{\varepsilon \rightarrow 0} \pi_X(\varepsilon)$  exists.

By Corollary 1, if  $\eta$  is concave on  $(F^{-1}(\delta), \text{ess-sup}(X))$ ,  $1/\eta$  is convex on  $(F^{-1}(\delta), \text{ess-sup}(X))$  and  $\lim_{\varepsilon \rightarrow 0} \pi_X(\varepsilon)$  also exists.  $\square$

**Example 7.** If  $\lim_{\varepsilon \rightarrow 0} \pi_X(\varepsilon)$  is a constant, we have  $\lim_{\varepsilon \rightarrow 0} \Pi_\varepsilon(X) = \lim_{\varepsilon \rightarrow 0} \pi_X(\varepsilon)$  as  $\pi_X(\Pi_X(\varepsilon)\varepsilon) = \Pi_X(\varepsilon)$ . We give the numerical values of  $\Pi_X(\varepsilon)$  at very small probability levels  $\varepsilon$  for normal, t, and log-normal distributions. These distributions do not have a constant PELVE curve, and using Corollary 2 we can check that their PELVE have limits. As we can see from Table 3, PELVE can still distinguish the heaviness of the tail even when  $\varepsilon$  is very small. The heavier tailed distributions report a higher PELVE value. For the normal distribution and the log-normal distribution with  $\sigma = 0.2$ , the value of PELVE is close to  $e \approx 2.7183$  as  $\varepsilon \downarrow 0$ . From the numerical values, it is unclear whether  $\Pi_X(\varepsilon) \rightarrow e$

for all log-normal distributions, but there is no practical relevance to compute  $\Pi_X(\varepsilon)$  for  $\varepsilon < 10^{-11}$  in applications.

Table 3: Values of  $\Pi_X(\varepsilon)$

Distribution	N	LN(1)	LN(0.5)	LN(0.2)	t(2)	t(3)
$\varepsilon = 10^{-10}$	2.6884	2.9167	2.7944	2.7290	4.0000	3.3750
$\varepsilon = 10^{-11}$	2.6909	2.9077	2.7920	2.7287	4.0000	3.3750

## 6 Applications to datasets used in insurance

In this section, we apply the PEVLE calibration techniques to datasets used in insurance to show how to use the calibrated distribution in estimating risk measure values and simulation.

### 6.1 Dental expenditure data

In this example, we apply the calibration model to the 6494 complete household component's total dental expenditure data from Medical Expenditure Panel Survey for 2020. An earlier version of the same dataset is used by [Behan et al. \(2010\)](#) to study the relationship between worker absenteeism and overweight or obesity. The main purpose of this experiment is to construct tractable models, with continuous and simple quantile functions, which have similar risk measure values as the original dataset, and the same PELVE at certain levels. We present in [Figure 14](#) two quantile functions calibrated from  $\Pi_X(\varepsilon_1)$  and  $\Pi_X(\varepsilon_2)$ , with  $(\varepsilon_1, \varepsilon_2) = (0.01, 0.05)$  and  $(\varepsilon_1, \varepsilon_2) = (0.05, 0.1)$ , respectively. The two calibrated quantile functions are scaled up according to the empirical  $\text{VaR}_{\varepsilon_1}(X)$  and  $\text{VaR}_{\varepsilon_2}(X)$ . By [Theorem 1](#), we can calibrate the quantile functions from Case 4 when  $(\varepsilon_1, \varepsilon_2) = (0.01, 0.05)$ , and from Case 5 when  $(\varepsilon_1, \varepsilon_2) = (0.05, 0.1)$ . As mentioned before, for  $(\varepsilon_1, \varepsilon_2) = (0.05, 0.1)$ , we set the calibrated quantile function in  $(0, c_1\varepsilon_1)$  as the Pareto quantile function. Hence, there is no flat part in the two calibrated quantile functions shown in [Figure 14](#). As we can see, both the two calibrated quantile functions fit the empirical quantile functions well. The calibrated quantile function can be regarded as a special parameterized model for tail distribution, which can fit the value of VaR and ES at specified levels. With the parameterized calibrated model, we can estimate the value of tail risk measures (see [Liu and Wang \(2021\)](#)) such as ES, VaR, and Range-VaR (RVaR), amongst others. In [Tables 4 and 5](#), we compute the values of ES and RVaR for the calibrated model and compare them with empirical ES and RVaR values, respectively, where the risk measure RVaR is defined as

$$\text{RVaR}_{\alpha, \beta}(X) = \frac{1}{\beta - \alpha} \int_{\beta}^{\alpha} \text{VaR}_{\gamma}(X) d\gamma$$

for  $0 \leq \alpha < \beta < 1$ ; see [Cont et al. \(2010\)](#) and [Embrechts et al. \(2018\)](#). As we scale the calibrated quantile function to empirical  $\text{VaR}_{\varepsilon_1}(X)$  and  $\text{VaR}_{\varepsilon_2}(X)$ , the calibrated ES and empirical ES are identical at levels  $\varepsilon_1 \Pi_X(\varepsilon_1)$  and  $\varepsilon_2 \Pi_X(\varepsilon_2)$  by the definition of PELVE. For other probability levels, the calibrated ES and RVaR in [Tables 4 and 5](#) are close to their empirical counterparts. When  $(\varepsilon_1, \varepsilon_2) = (0.01, 0.05)$ , it may only be useful to compute calibrated  $\text{ES}_p(X)$  for  $p < 0.05 \Pi_X(0.05) = 0.11591$  because the calibrated quantile function is arbitrary beyond the level 0.11591. If we need to estimate ES or RVaR for a larger probability level, we can choose a higher  $\varepsilon_2$  as long as  $\mathbb{E}[X] \leq \text{VaR}_{\varepsilon_2}(X)$  is satisfied. For this dataset, the highest  $\varepsilon_2$  we can use is 0.1983.

Using the methods in [Section 3](#), for quantile levels between  $(0, \varepsilon_1)$ , the distribution calibrated from one point  $(\varepsilon_1, c_1)$  is the same as the one calibrated from two points  $(\varepsilon_1, c_1)$  and  $(\varepsilon_2, c_2)$ . Hence, the results for  $\text{ES}_p$  of the one-point calibrated function are also shown in [Tables 4 and 5](#) in the cells  $p \leq \varepsilon_1$ .

Figure 14: Empirical and calibrated  $\text{VaR}_\varepsilon$  for the dental expenditure data

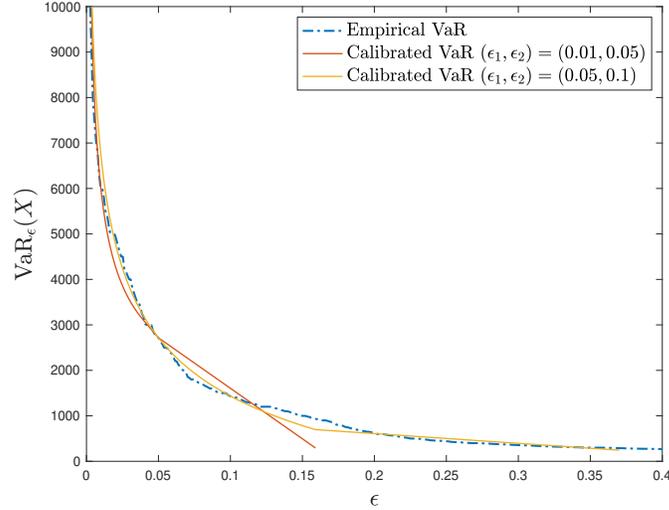


Table 4: Empirical ES and calibrated ES for the dental expenditure data

$p$	0.01	0.05	0.1	0.2	0.3
Empirical $\text{ES}_p$	10073.1	5361.7	3624.8	2317.9	1696.7
Calibrated $\text{ES}_p$ from $(\varepsilon_1, \varepsilon_2) = (0.01, 0.05)$	11703.9	5357.7	3759.1	-	-
Calibrated $\text{ES}_p$ from $(\varepsilon_1, \varepsilon_2) = (0.05, 0.1)$	10878.1	5439.6	3711.3	2293.7	1696.4

Table 5: Empirical RVaR and calibrated RVaR for the dental expenditure data

$(\alpha, \beta)$	(0.01, 0.02)	(0.02, 0.05)	(0.05, 0.1)
Empirical $\text{RVaR}_{\alpha, \beta}$	5748.5	3662.4	1887.9
Calibrated $\text{RVaR}_{\alpha, \beta}$ from $(\varepsilon_1, \varepsilon_2) = (0.01, 0.05)$	5003.7	3360.2	2160.6
Calibrated $\text{RVaR}_{\alpha, \beta}$ from $(\varepsilon_1, \varepsilon_2) = (0.05, 0.1)$	5634.6	3561.8	1983.1

## 6.2 Hospital costs data

In this example, we apply the calibration process to the Hospital Costs data of [Frees \(2009\)](#) which were originally from the Nationwide Inpatient Sample of the Healthcare Cost and Utilization Project (NIS-HCUP). The data contains 500 hospital costs observations with 244 males and 256 females which can be regarded as the losses of the health insurance policies. Using the calibration model of the two-point constraint problem, we calibrate quantile functions for females and males from PELVE at probability levels  $\varepsilon_1 = 0.05$  and  $\varepsilon_2 = 0.1$ , which are shown in [Figure 15](#). Except for estimating the risk measure, the calibrated distribution is useful in simulation. Assume the insurance company wants to know the top 10% hospital costs; that is  $X|X > \text{VaR}_{0.1}(X)$  where  $X$  is the hospital costs. There are only 24 available data for males and 25 available data for females, which would be not enough for making statistically solid decisions. To generate more pseudo-data points, we can simulate data from the calibrated distribution; that is, we simulate data from  $F^{[p,1]}$  where  $F$  is the calibrated distribution in [Figure 15](#). Taking  $p = 0.9$ , we have  $F^{[p,1]}(t) = \text{VaR}_{(1-p)(1-t)}(X)$  with  $\text{VaR}_t(X)$  from [Figure 15](#). We simulate 1000 data from the calibrated distributions based on PELVE at  $\varepsilon_1 = 0.05$  and  $\varepsilon_2 = 0.1$ . In [Figure 16](#), we present two QQ plots of simulated data against empirical data for females and males respectively. As we can see, the simulated data has a similar distribution as the empirical data. Those simulated pseudo-data points can be used for estimating risk measures or making other decisions. For example, the simulated hospital cost can be used to design health insurance contrasts or set the premium in complex systems, where sometimes methods based on simulated data are more convenient to work with than methods relying on distribution functions. This may be seen as an alternative, smoothed, version of bootstrap; recall that the classic bootstrap sample can only take the values represented in the dataset. Furthermore, we compare the simulated data of hospital costs for females and males in [Figure 17](#), which shows that the distribution of the hospital costs for females has a heavier tail than that for males.

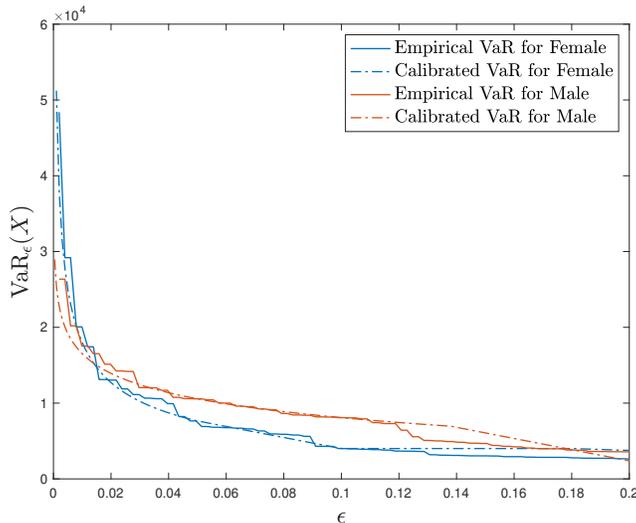


Figure 15: Empirical and calibrated  $\text{VaR}_\epsilon$  for the hospital costs data

## 7 Conclusion

In this paper, we offer several contributions to the calibration problem and properties of the PELVE. The calibration problem concerns, with some given values from a PELVE curve, how one can build a distribution that has this PELVE. We solve a few settings of calibration based on a one-point constraint, a two-point constraint, or the entire curve constraint. In particular, the calibration for a given PELVE curve involves solving an integral equation  $\int_0^y f(s) ds = yf(z(y)y)$  for a given function  $z$ , and this requires some advanced analysis and a numerical method in differential equations. For the case that  $z$  is a constant curve, we can identify all solutions, which are surprisingly complicated. In addition, we see that if  $\pi_X$  is a constant larger than  $e$ , which is observed from typical values in financial return data (Li and Wang (2022)),  $X$  share the same tail behavior with the corresponding Pareto solution. We also applied our calibration techniques to two datasets used in insurance.

On the technical side, we study whether the PELVE is monotone and whether it converges at 0. We show that the monotonicity of the PELVE is associated with the shape of the hazard rate. If the inverse of the hazard rate is convex (concave), the PELVE is decreasing (increasing). The monotonicity at the tail part of the PELVE leads to conditions to check the convergence of the PELVE at 0. If the inverse of the hazard rate is convex (concave) at the tail of the distribution, the limit of the PELVE at 0 exists.

There are several open questions related to PELVE that we still do not fully understand. One particular such question is whether the tail behavior, e.g., tail index, of a distribution is completely determined by its PELVE. We have seen that this holds true in the case of a constant PELVE (see

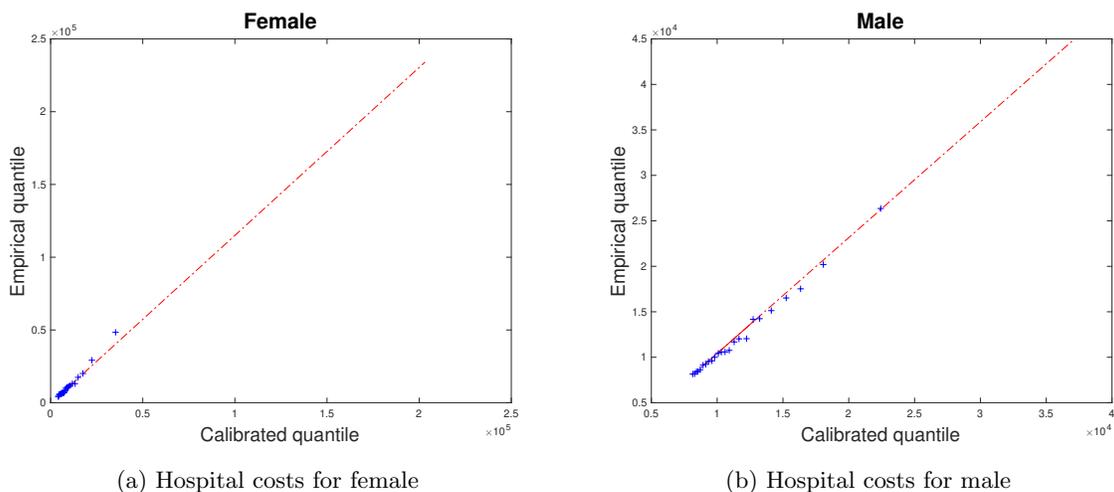


Figure 16: QQ plot: simulated data VS empirical data

Theorem 2), but we do not have a general conclusion. In the case of regularly varying survival functions, Li and Wang (2022, Theorem 3) showed that the limit of PELVE determines its tail parameter, but it is unclear whether this can be generalized to other distributions. Another challenging task is, for a specified curve  $\pi$  on  $[0, 1]$ , to determine whether there exists a model  $X$  with  $\pi_X = \pi$ . The case of  $n$ -point constraints for large  $n$  may require a new design of verification algorithms. This question concerns the compatibility of given information with statistical models, which has been studied, in other applications of risk management, by Embrechts et al. (2002, 2016) and Krause et al. (2018).

## Acknowledgements

The authors thank Xiyue Han for many helpful comments. Ruodu Wang is supported by the Natural Sciences and Engineering Research Council of Canada (RGPIN-2018-03823, RGPAS-2018-522590).

## References

- Acerbi, C. and Székely, B. (2014). Backtesting expected shortfall. *Risk Magazine*, 17 Nov 2014.
- Artzner, P., Delbaen, F., Eber, J.-M. and Heath, D. (1999). Coherent measures of risk. *Mathematical Finance*, **9**(3), 203–228.
- Balbás, A., Balbás, B. and Balbás, R. (2017). Differential equations connecting VaR and CVaR. *Journal of Computational and Applied Mathematics*, **326**, 247–267.
- Barczy, M., Nedényi, F. K. and Sütő, L. (2022). Probability equivalent level of Value at Risk and higher-order Expected Shortfalls. *Insurance: Mathematics and Economics*, **108**, 107–128.

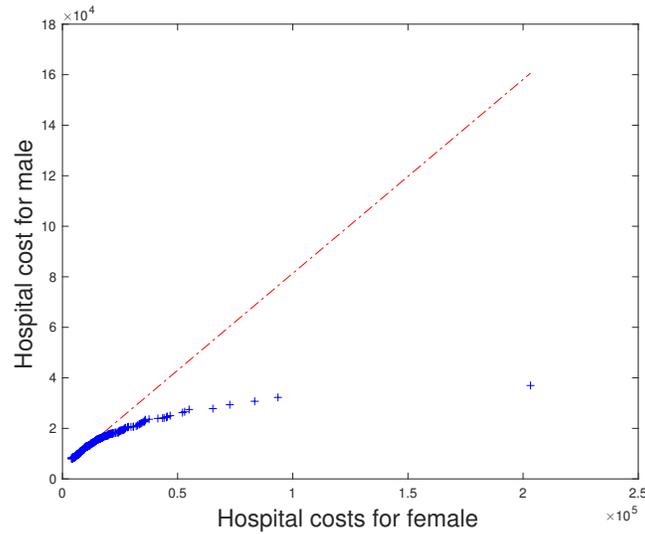


Figure 17: QQ plot of simulated data of hospital costs: female VS male

- BCBS (2019). *Minimum Capital Requirements for Market Risk. February 2019*. Basel Committee on Banking Supervision. Basel: Bank for International Settlements. BIS online publication No. bcbs457.
- Behan, D. F., Cox, S. H., Lin, Y., Pai, J., Pedersen, H. W. and Yi, M. (2010). Obesity and its relation to mortality and morbidity costs. *Society of Actuaries Research Projects – Life Insurance*, December 2010.
- Bellman, R. and Cooke, K. L. (1963). *Differential-difference Equations*. The RAND Corporation.
- Berezansky, L. and Braverman, E. (2011). On nonoscillation of advanced differential equations with several terms. *Abstract and Applied Analysis*, **2011**, 637142.
- Chambers, C. P. (2009). An axiomatization of quantiles on the domain of distribution functions. *Mathematical Finance*, **19**(2), 335–342.
- Cont, R., Deguest, R. and Scandolo, G. (2010). Robustness and sensitivity analysis of risk measurement procedures. *Quantitative Finance*, **10**(6), 593–606.
- Embrechts, P. and Hofert, M. (2013). A note on generalized inverses. *Mathematical Methods of Operations Research*, **77**(3), 423–432.
- Embrechts, P., Hofert, M. and Wang, R. (2016). Bernoulli and tail-dependence compatibility. *Annals of Applied Probability*, **26**(3), 1636–1658.
- Embrechts, P., Liu, H. and Wang, R. (2018). Quantile-based risk sharing. *Operations Research*, **66**(4), 936–949.
- Embrechts, P., McNeil, A. and Straumann, D. (2002). Correlation and dependence in risk management: properties and pitfalls. In *Risk Management: Value at Risk and Beyond* (Ed. Dempster, M. A. H.), 176–223, Cambridge University Press.

- Embrechts, P., Puccetti, G., Ruschendorf, L., Wang, R. and Beleraj, A. (2014). An academic response to Basel 3.5. *Risks*, **2**(1), 25–48.
- 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.
- Fiori, A. M. and Gianin, E. R. (2022). Generalized PELVE and applications to risk measures. *European Actuarial Journal*, 1–33.
- Frees, E. W. (2009). *Regression Modeling with Actuarial and Financial Applications*. Cambridge University Press.
- Gneiting, T. (2011). Making and evaluating point forecasts. *Journal of the American Statistical Association*, **106**(494), 746–762.
- Krause, D., Scherer, M., Schwinn, J. and Werner, R. (2018). Membership testing for Bernoulli and tail-dependence matrices. *Journal of Multivariate Analysis*, **168**, 240–260.
- Kou, S. and Peng, X. (2016). On the measurement of economic tail risk. *Operations Research*, **64**(5), 1056–1072.
- Li, H. and Wang, R. (2022). PELVE: Probability equivalent level of VaR and ES. *Journal of Econometrics*, **234**(1), 353–370.
- Liu, F. and Wang, R. (2021). A theory for measures of tail risk. *Mathematics of Operations Research*, **46**(3), 1109–1128.
- McNeil, A. J., Frey, R. and Embrechts, P. (2015). *Quantitative Risk Management: Concepts, Techniques and Tools*. Revised Edition. Princeton, NJ: Princeton University Press.
- Norton, M., Khokhlov, V. and Uryasev, S. (2021). Calculating CVaR and bPOE for common probability distributions with application to portfolio optimization and density estimation. *Annals of Operations Research*, **299**(1), 1281–1315.
- Rudin, W. (1987). *Real and Complex Analysis*. International Series in Pure and Applied Mathematics (3rd Ed.). McGraw–Hill.
- Siewert, C. E. and Burniston, E. E. (1973). Exact analytical solutions of  $ze^z = a$ . *Journal of Mathematical Analysis and Applications*, **43**(3), 626–632.
- Wang, R. and Zitikis, R. (2021). An axiomatic foundation for the Expected Shortfall. *Management Science*, **67**, 1413–1429.

## A Omitted proofs in Section 3

*Proof of Lemma 1.* As  $\mathbb{E}[X] \leq \text{VaR}_{\varepsilon_2}(X)$  and  $\varepsilon_1 < \varepsilon_2$ ,  $\mathbb{E}[X] \leq \text{VaR}_{\varepsilon_2}(X) \leq \text{VaR}_{\varepsilon_1}(X)$ . By Proposition 1 in Li and Wang (2022),  $\Pi_X(\varepsilon_1) < \infty$  and  $\Pi_X(\varepsilon_2) < \infty$ .

For any  $\varepsilon \in (0, 1)$  satisfying  $\mathbb{E}[X] \leq \text{VaR}_\varepsilon(X)$ ,

$$\begin{aligned}\varepsilon \Pi_X(\varepsilon) &= \varepsilon \inf\{c \in [1, 1/\varepsilon] : \text{ES}_{c\varepsilon}(X) \leq \text{VaR}_\varepsilon(X)\} \\ &= \inf\{c \in [\varepsilon, 1] : \text{ES}_{c\varepsilon}(X) \leq \text{VaR}_\varepsilon(X)\} \\ &= \inf\{k \in [\varepsilon, 1] : \text{ES}_k(X) \leq \text{VaR}_\varepsilon(X)\}.\end{aligned}$$

Let  $A(\varepsilon) = \{k \in [\varepsilon, 1] : \text{ES}_k(X) \leq \text{VaR}_\varepsilon(X)\}$ . For any  $k \in A(\varepsilon_2)$ , we have  $1 \geq k \geq \varepsilon_2 > \varepsilon_1$  and  $\text{ES}_k(X) \leq \text{VaR}_{\varepsilon_2}(X) \leq \text{VaR}_{\varepsilon_1}(X)$ . Hence,  $k \in A(\varepsilon_1)$  and this gives  $A(\varepsilon_2) \subseteq A(\varepsilon_1)$ . Therefore,  $\varepsilon_2 \Pi_X(\varepsilon_2) = \inf A(\varepsilon_2) \geq \inf A(\varepsilon_1) = \varepsilon_1 \Pi_X(\varepsilon_1)$ .  $\square$

*Proof of Theorem 1.* We will check the equivalent condition (2) between VaR and ES. Note that if  $t \mapsto \text{VaR}_t(X)$  is a constant on  $(0, \varepsilon)$ , then  $\Pi_X(\varepsilon) = 1$ . If  $t \mapsto \text{VaR}_t(X)$  is not a constant on  $(0, \varepsilon)$ , then  $\Pi_X(\varepsilon)$  is the unique solution that satisfies  $\text{ES}_{\varepsilon \Pi_X(\varepsilon)}(X) = \text{VaR}_\varepsilon(X)$ .

- (i) Case 1,  $c_2 = 1$ . It is clear that  $\text{VaR}_t(X)$  is a constant for  $t \in (0, c_2 \varepsilon_2]$  and (2) is satisfied. Hence,  $\Pi_X(\varepsilon_2) = 1$ . Moreover,  $\text{VaR}_t(X)$  is also a constant for  $t \in (0, c_1 \varepsilon_1]$ , which implies  $\Pi_X(\varepsilon_1) = 1$ .
- (ii) Case 2,  $c_1 = 1$  and  $1 < c_2 \leq 1/\varepsilon_2$ . For  $t \in (0, \varepsilon_1)$ ,  $\text{VaR}_t(X) = G_{\mathbf{z}}(t)$  is a constant for  $t \in (0, c_1 \varepsilon_1)$ . Hence,  $\Pi_X(\varepsilon_1) = 1$ . Next, we check whether  $\text{ES}_{c_2 \varepsilon_2}(X) = \text{VaR}_{\varepsilon_2}(X)$ . The value of  $\text{ES}_{c_2 \varepsilon_2}(X)$  is

$$\begin{aligned}\text{ES}_{c_2 \varepsilon_2}(X) &= \frac{1}{c_2 \varepsilon_2} \left( \int_0^{\varepsilon_1} \hat{k} d\varepsilon + \int_{\varepsilon_1}^{\varepsilon_2} (a_1 \varepsilon + b_1) d\varepsilon + \int_{\varepsilon_2}^{c_2 \varepsilon_2} (a_2 \varepsilon + b_2) d\varepsilon \right) \\ &= \frac{1}{c_2 \varepsilon_2} \left( \varepsilon_1 \hat{k} + \frac{1}{2} a_1 (\varepsilon_2^2 - \varepsilon_1^2) + b_1 (\varepsilon_2 - \varepsilon_1) + \frac{1}{2} a_2 (c_2^2 \varepsilon_2^2 - \varepsilon_2^2) + b_2 (c_2 \varepsilon_2 - \varepsilon_2) \right) \\ &= \frac{1}{c_2 \varepsilon_2} \left( \frac{1}{2} a_1 (\varepsilon_2 - \varepsilon_1)^2 + \hat{k} \varepsilon_2 + \frac{1}{2} a_2 (c_2 \varepsilon_2 - \varepsilon_2)^2 + \tilde{k} (c_2 \varepsilon_2 - \varepsilon_2) \right) \\ &= \frac{1}{c_2 \varepsilon_2} \left( \frac{1}{2} (\tilde{k} - \hat{k}) (\varepsilon_2 - \varepsilon_1) + \hat{k} \varepsilon_2 + \frac{1}{2} (\tilde{k} - \hat{k}) (\varepsilon_1 + \varepsilon_2) + \tilde{k} (c_2 \varepsilon_2 - \varepsilon_2) \right) = \tilde{k}\end{aligned}$$

The value of  $\text{VaR}_{\varepsilon_2}(X)$  is  $a_2 \varepsilon_2 + b_2 = \tilde{k}$ . Thus, (2) is satisfied. As  $\text{VaR}_t(X)$  is not a constant for  $t \in (0, c_2 \varepsilon_2)$ , we have  $\Pi_X(\varepsilon_2) = c_2$ .

- (iii) Case 3,  $1 < c_1 \leq 1/\varepsilon_1$  and  $c_2 = \frac{c_1 \varepsilon_1}{\varepsilon_2}$ . In this case, we have

$$\text{VaR}_{\varepsilon_1}(X) = G_{\mathbf{z}}(\varepsilon_1) = k(\varepsilon_1) = a \varepsilon_2 + b = G_{\mathbf{z}}(\varepsilon_2) = \text{VaR}_{\varepsilon_2}(X)$$

and  $\text{ES}_{c_1 \varepsilon_1}(X) = \text{ES}_{c_2 \varepsilon_2}(X)$  as  $c_1 \varepsilon_1 = c_2 \varepsilon_2$ . Thus, we only need to check whether  $\text{ES}_{c_2 \varepsilon_2}(X) =$

$\text{VaR}_{\varepsilon_2}(X)$ . The value of  $\text{ES}_{c_2\varepsilon_2}(X)$  is

$$\begin{aligned}
\text{ES}_{c_2\varepsilon_2}(X) &= \frac{1}{c_2\varepsilon_2} \left( \int_0^{\varepsilon_1} k(\varepsilon) d\varepsilon + \int_{\varepsilon_1}^{\varepsilon_2} k(\varepsilon_1) d\varepsilon + \int_{\varepsilon_2}^{c_2\varepsilon_2} a\varepsilon + b d\varepsilon \right) \\
&= \frac{1}{c_2\varepsilon_2} \left( k + k(\varepsilon_1)(\varepsilon_2 - \varepsilon_1) + \frac{1}{2}a(c_2^2\varepsilon_2^2 - \varepsilon_2^2) + (k(\varepsilon_1) - a\varepsilon_2)(c_2\varepsilon_2 - \varepsilon_2) \right) \\
&= \frac{1}{c_2\varepsilon_2} \left( k + k(\varepsilon_1)(c_2\varepsilon_2 - \varepsilon_1) + \frac{1}{2}a(c_2\varepsilon_2 - \varepsilon_2)^2 \right) \\
&= \frac{1}{c_2\varepsilon_2} (k + k(\varepsilon_1)(c_2\varepsilon_2 - \varepsilon_1) + k(\varepsilon_1)\varepsilon_1 - k) = k(\varepsilon_1).
\end{aligned}$$

The value of  $\text{VaR}_{\varepsilon_2}(X)$  is also  $k(\varepsilon_1)$ . Hence, (2) is satisfied and  $\Pi_X(\varepsilon_1) = c_1$ ,  $\Pi_X(\varepsilon_2) = c_2$  because  $t \mapsto \text{VaR}_t(X)$  is not a constant on  $(0, \varepsilon_1)$ .

- (iv) Case 4,  $1 < c_1 \leq \varepsilon_2/\varepsilon_1$  and  $1 < c_2 \leq 1/\varepsilon_2$ . The first equivalent condition of (2) for  $\text{VaR}_{\varepsilon_1}(X)$  and  $\text{ES}_{c_1\varepsilon_1}(X)$  is satisfied because  $\text{VaR}_t(X) = k(t)$  is the quantile function for  $\text{GPD}(\xi)$  with PELVE  $c_1$  and  $t \in (0, c_1\varepsilon_1)$ . Hence, we have  $\Pi_X(\varepsilon_1) = c_1$ . Moreover,  $\text{ES}_{c_1\varepsilon_1}(X) = \text{VaR}_{\varepsilon_1}(X) = k(\varepsilon_1)$ . We choose  $a_1 = k'(c_1\varepsilon_1)$  and  $b_1$  such that  $a_1c_1\varepsilon_1 + b_1 = k(c_1\varepsilon_1)$ . For the equivalent condition between  $\text{ES}_{c_2\varepsilon_2}(X)$  and  $\text{VaR}_{\varepsilon_2}(X)$ , we can verify

$$\begin{aligned}
\text{ES}_{c_2\varepsilon_2}(X) &= \frac{1}{c_2\varepsilon_2} \left( \int_0^{c_1\varepsilon_1} k(\varepsilon) d\varepsilon + \int_{c_1\varepsilon_1}^{\varepsilon_2} a_1\varepsilon + b_1 d\varepsilon + \int_{\varepsilon_2}^{c_2\varepsilon_2} (a_2\varepsilon + b_2) d\varepsilon \right) \\
&= \frac{1}{c_2\varepsilon_2} \left( c_1\varepsilon_1 k(\varepsilon_1) + \frac{1}{2}a_1(\varepsilon_2^2 - c_1^2\varepsilon_1^2) + b_1(\varepsilon_2 - c_1\varepsilon_1) + \frac{1}{2}a_2(c_2^2\varepsilon_2^2 - \varepsilon_2^2) + b_2(c_2\varepsilon_2 - \varepsilon_2) \right) \\
&= \frac{1}{c_2\varepsilon_2} \left( c_1\varepsilon_1 k(\varepsilon_1) + \frac{1}{2}a_1(2c_2\varepsilon_2^2 - \varepsilon_2^2 - c_1^2\varepsilon_1^2) + b_1(c_2\varepsilon_2 - c_1\varepsilon_1) + \frac{1}{2}a_2(c_2\varepsilon_2 - \varepsilon_2)^2 \right) \\
&= \frac{1}{c_2\varepsilon_2} (a_1c_2\varepsilon_2^2 + b_1c_2\varepsilon_2) = a_1\varepsilon_2 + b_1 = \text{VaR}_{\varepsilon_2}(X).
\end{aligned}$$

Thus, (2) is satisfied and we have  $\Pi_X(\varepsilon_2) = c_2$ .

- (v) Case 5,  $\varepsilon_2/\varepsilon_1 < c_1 \leq 1/\varepsilon_1$  and  $\frac{c_1\varepsilon_1}{\varepsilon_2} < c_2 \leq 1/\varepsilon_2$ . The equality between  $\text{VaR}_{\varepsilon_1}(X)$  and  $\text{ES}_{c_1\varepsilon_1}(X)$  can be checked by

$$\begin{aligned}
\text{ES}_{c_1\varepsilon_1}(X) &= \frac{1}{c_1\varepsilon_1} \left( \int_0^{\varepsilon_1} k(\varepsilon) d\varepsilon + \int_{\varepsilon_1}^{\varepsilon_2} (a_1\varepsilon + b_1) d\varepsilon + (a_1\varepsilon_2 + b_1)(c_1\varepsilon_1 - \varepsilon_2) \right) \\
&= \frac{1}{c_1\varepsilon_1} \left( k + \frac{1}{2}a_1(\varepsilon_2^2 - \varepsilon_1^2) + (k(\varepsilon_1) - a_1\varepsilon_1)(\varepsilon_2 - \varepsilon_1) + (a_1\varepsilon_2 + b_1)(c_1\varepsilon_1 - \varepsilon_2) \right) \\
&= \frac{1}{c_1\varepsilon_1} (k + a_1(\varepsilon_2 - \varepsilon_1)(c_1\varepsilon_1 - 1/2(\varepsilon_2 + \varepsilon_1)) + k(\varepsilon_1)(c_1\varepsilon_1 - \varepsilon_1)) \\
&= \frac{1}{c_1\varepsilon_1} (k + k(\varepsilon_1)\varepsilon_1 - k + k(\varepsilon_1)(c_1\varepsilon_1 - \varepsilon_1)) = k(\varepsilon_1) = G_{\mathbf{z}}(\varepsilon_1) = \text{VaR}_{\varepsilon_1}(X).
\end{aligned}$$

The equality between  $\text{VaR}_{\varepsilon_2}(X)$  and  $\text{ES}_{c_2\varepsilon_2}(X)$  can be checked by

$$\begin{aligned}
\text{ES}_{c_2\varepsilon_2}(X) &= \frac{1}{c_2\varepsilon_2} \left( \int_0^{c_1\varepsilon_1} k(\varepsilon)d\varepsilon + \int_{c_1\varepsilon_1}^{c_2\varepsilon_2} (a_2\varepsilon + b_2)d\varepsilon \right) \\
&= \frac{1}{c_2\varepsilon_2} \left( c_1\varepsilon_1 k(\varepsilon_1) + \frac{1}{2}a_2(c_2^2\varepsilon_2^2 - c_1^2\varepsilon_1^2) + b_2(c_2\varepsilon_2 - c_1\varepsilon_1) \right) \\
&= \frac{1}{c_2\varepsilon_2} \left( c_1\varepsilon_1 k(\varepsilon_1) + \frac{1}{2}a_2(c_2\varepsilon_2 - c_1\varepsilon_1)^2 + (a_1\varepsilon_2 + b_1)(c_2\varepsilon_2 - c_1\varepsilon_1) \right) \\
&= \frac{1}{c_2\varepsilon_2} (c_1\varepsilon_1 k(\varepsilon_1) + c_1\varepsilon_1(a_1\varepsilon_2 + b_1 - k(\varepsilon_1)) + (a_1\varepsilon_2 + b_1)(c_2\varepsilon_2 - c_1\varepsilon_1)) \\
&= a_1\varepsilon_2 + b_1 = G_{\mathbf{z}}(\varepsilon_2) = \text{VaR}_{\varepsilon_2}(X)
\end{aligned}$$

Hence, (2) is satisfied, and  $\Pi_X(\varepsilon_1) = c_1$  and  $\Pi_X(\varepsilon_2) = c_2$ .

Therefore, it is checked that  $X$  satisfies  $\Pi_X(\varepsilon_1) = c_1$  and  $\Pi_X(\varepsilon_2) = c_2$  for all five cases.  $\square$

The following propositions address the issue discussed in Remark 1 by showing that the boundary cases of  $(\varepsilon_1, c_1, \varepsilon_2, c_2)$  cannot be achieved by strictly decreasing quantile functions, and hence our construction of quantiles with a flat region in Figure 5 are needed.

**Proposition 6.** *For any  $X \in L^1$ , let  $\varepsilon_1, \varepsilon_2 \in (0, 1)$  be such that  $\mathbb{E}[X] \leq \text{VaR}_{\varepsilon_2}(X)$  and  $\varepsilon_1 < \varepsilon_2$ . Then,  $\Pi_X(\varepsilon_2) = \max\{1, \Pi_X(\varepsilon_1)\varepsilon_1/\varepsilon_2\}$  if and only if  $\text{VaR}_{\varepsilon_1}(X) = \text{VaR}_{\varepsilon_2}(X)$ .*

*Proof.* Using the same logic as in Lemma 1, we have that  $\Pi_X(\varepsilon_1)$  and  $\Pi_X(\varepsilon_2)$  are finite.

We first show the “if” statement. Assume  $\text{VaR}_{\varepsilon_1}(X) = \text{VaR}_{\varepsilon_2}(X)$ . As  $\text{VaR}_{\varepsilon}(X)$  is decreasing, we know that  $\text{VaR}_{\varepsilon}(X)$  is a constant on  $[\varepsilon_1, \varepsilon_2]$ .

If  $\text{VaR}_{\varepsilon}(X) = \text{VaR}_{\varepsilon_1}(X)$  for  $\varepsilon \in (0, \varepsilon_1)$ , then  $\text{VaR}_{\varepsilon}(X)$  is a constant on  $(0, \varepsilon_2]$ . Therefore, we can get  $\Pi_X(\varepsilon_1) = \Pi_X(\varepsilon_2) = 1$ . Note that  $\Pi_X(\varepsilon_1)\varepsilon_1/\varepsilon_2 = \varepsilon_1/\varepsilon_2 < 1$ . Thus, we obtain  $\Pi_X(\varepsilon_2) = \max\{1, \Pi_X(\varepsilon_1)\varepsilon_1/\varepsilon_2\}$ .

If there exists  $\varepsilon \in (0, \varepsilon_1)$  such that  $\text{VaR}_{\varepsilon}(X) > \text{VaR}_{\varepsilon_1}(X)$ , then  $\text{ES}_{\varepsilon}(X)$  is strictly decreasing on  $[\varepsilon_1, 1]$ . By the equivalent condition between VaR and ES,  $\text{VaR}_{\varepsilon_1}(X) = \text{VaR}_{\varepsilon_2}(X)$  implies  $\text{ES}_{\varepsilon_1\Pi_X(\varepsilon_1)}(X) = \text{ES}_{\varepsilon_2\Pi_X(\varepsilon_2)}(X)$ . Thus,  $\varepsilon_1\Pi_X(\varepsilon_1) = \varepsilon_2\Pi_X(\varepsilon_2)$ . Furthermore, we have

$$\text{VaR}_{\varepsilon_1\Pi_X(\varepsilon_1)}(X) < \text{ES}_{\varepsilon_1\Pi_X(\varepsilon_1)}(X) = \text{VaR}_{\varepsilon_1}(X) = \text{VaR}_{\varepsilon_2}(X).$$

Thus,  $\varepsilon_1\Pi_X(\varepsilon_1) > \varepsilon_2$  and we get  $\Pi_X(\varepsilon_2) = \max\{1, \Pi_X(\varepsilon_1)\varepsilon_1/\varepsilon_2\}$ .

Next, we show the “only if” statement. Assume  $\Pi_X(\varepsilon_2) = \max\{1, \Pi_X(\varepsilon_1)\varepsilon_1/\varepsilon_2\}$ .

If  $\Pi_X(\varepsilon_2) = 1$ , then  $\text{VaR}_{\varepsilon_2}(X) = \text{ES}_{\varepsilon_2}(X)$ . This implies that  $\text{VaR}_{\varepsilon}(X)$  is a constant on  $(0, \varepsilon_2]$ , which gives  $\text{VaR}_{\varepsilon_1}(X) = \text{VaR}_{\varepsilon_2}(X)$ .

If  $\Pi_X(\varepsilon_2) = \Pi_X(\varepsilon_1)\varepsilon_1/\varepsilon_2$ , then  $\varepsilon_2\Pi_X(\varepsilon_2) = \varepsilon_1\Pi_X(\varepsilon_1)$ . Hence, we have

$$\text{VaR}_{\varepsilon_1}(X) = \text{ES}_{\varepsilon_1\Pi_X(\varepsilon_1)}(X) = \text{ES}_{\varepsilon_2\Pi_X(\varepsilon_2)}(X) = \text{VaR}_{\varepsilon_2}(X).$$

Thus, we complete the proof.  $\square$

**Proposition 7.** For any  $X \in L^1$ , let  $\varepsilon_1, \varepsilon_2 \in (0, 1)$  be such that  $\mathbb{E}[X] \leq \text{VaR}_{\varepsilon_2}(X)$  and  $\varepsilon_1 < \varepsilon_2$ . Let  $c_1 = \Pi_X(\varepsilon_1)$  and  $c_2 = \Pi_X(\varepsilon_2)$ . If  $\text{VaR}_{\varepsilon_1}(X) > \text{VaR}_{\varepsilon_2}(X)$ , then

$$\hat{c} \leq c_2 \leq \begin{cases} \min \left\{ \frac{1}{\varepsilon_2}, \frac{c_1\varepsilon_1}{\varepsilon_2} \left( \frac{\text{VaR}_{\varepsilon_1}(X) - \text{VaR}_{c_1\varepsilon_1}(X)}{\text{VaR}_{\varepsilon_2}(X) - \text{VaR}_{c_1\varepsilon_1}(X)} \right) \right\}, & \text{VaR}_{c_1\varepsilon_1}(X) < \text{VaR}_{\varepsilon_2}(X), \\ \frac{1}{\varepsilon_2}, & \text{VaR}_{c_1\varepsilon_1}(X) \geq \text{VaR}_{\varepsilon_2}(X), \end{cases}$$

where

$$\hat{c} = \inf \left\{ t \in (1, 1/\varepsilon_2] : \frac{(t\varepsilon_2 - c_1\varepsilon_1)(\text{VaR}_{\varepsilon_2}(X) - \text{VaR}_{t\varepsilon_2}(X))}{c_1\varepsilon_1(\text{VaR}_{\varepsilon_1}(X) - \text{VaR}_{\varepsilon_2}(X))} \geq 1 \right\}.$$

Moreover,  $\hat{c} \geq \max\{1, c_1\varepsilon_1/\varepsilon_2\}$ .

*Proof.* As  $\mathbb{E}[X] \leq \text{VaR}_{\varepsilon_2}(X)$ , we have  $c_1 < \infty$  and  $c_2 < \infty$ . By definition,  $c_2 \leq 1/\varepsilon_2$ . From Lemma 6, we get  $c_2 > \max\{1, c_1\varepsilon_1/\varepsilon_2\}$ . Thus, the value of  $c_2$  should be in  $(\max\{1, c_1\varepsilon_1/\varepsilon_2\}, 1/\varepsilon_2]$ .

Note that  $c_1, \varepsilon_1, c_2, \varepsilon_2$  satisfy the equivalent condition (2). We can rewrite (2) as

$$\int_0^{c_1\varepsilon_1} \text{VaR}_\varepsilon(X) d\varepsilon = c_1\varepsilon_1 \text{VaR}_{\varepsilon_1}(X) \quad \text{and} \quad \int_0^{c_2\varepsilon_2} \text{VaR}_\varepsilon(X) d\varepsilon = c_2\varepsilon_2 \text{VaR}_{\varepsilon_2}(X).$$

Therefore, we have

$$\int_{c_1\varepsilon_1}^{c_2\varepsilon_2} \text{VaR}_\varepsilon(X) d\varepsilon = c_2\varepsilon_2 \text{VaR}_{\varepsilon_2}(X) - c_1\varepsilon_1 \text{VaR}_{\varepsilon_1}(X).$$

Furthermore, by the monotonicity of VaR, we have

$$(c_2\varepsilon_2 - c_1\varepsilon_1) \text{VaR}_{c_2\varepsilon_2}(X) \leq \int_{c_1\varepsilon_1}^{c_2\varepsilon_2} \text{VaR}_\varepsilon(X) d\varepsilon \leq (c_2\varepsilon_2 - c_1\varepsilon_1) \text{VaR}_{c_1\varepsilon_1}(X).$$

The two inequality will provide an upper bound and a lower bound for  $c_2$ .

An upper bound on  $c_2$ . Using  $c_2\varepsilon_2 \text{VaR}_{\varepsilon_2}(X) - c_1\varepsilon_1 \text{VaR}_{\varepsilon_1}(X) \leq (c_2\varepsilon_2 - c_1\varepsilon_1) \text{VaR}_{c_1\varepsilon_1}(X)$ , we have

$$c_2\varepsilon_2 (\text{VaR}_{\varepsilon_2}(X) - \text{VaR}_{c_1\varepsilon_1}(X)) \leq c_1\varepsilon_1 (\text{VaR}_{\varepsilon_1}(X) - \text{VaR}_{c_1\varepsilon_1}(X)). \quad (9)$$

If  $\text{VaR}_{c_1\varepsilon_1}(X) \geq \text{VaR}_{\varepsilon_2}(X)$ , the left side of (9) is less or equal to 0 and the right side of (9) is larger or equal to 0 because  $\text{VaR}_{\varepsilon_1}(X) \geq \text{VaR}_{c_1\varepsilon_1}(X)$ . Therefore, (9) is satisfied for any  $c_2 \in (\max\{1, c_1\varepsilon_1/\varepsilon_2\}, 1/\varepsilon_2]$ . The upper bound for  $c_2$  is unchanged.

On the other hand, if  $\text{VaR}_{c_1\varepsilon_1}(X) < \text{VaR}_{\varepsilon_2}(X)$ , we have

$$c_2 \leq \frac{c_1\varepsilon_1}{\varepsilon_2} \left( \frac{\text{VaR}_{\varepsilon_1}(X) - \text{VaR}_{c_1\varepsilon_1}(X)}{\text{VaR}_{\varepsilon_2}(X) - \text{VaR}_{c_1\varepsilon_1}(X)} \right).$$

Thus, an upper bound for  $c_2$  is  $\min \left\{ \frac{1}{\varepsilon_2}, \frac{c_1\varepsilon_1}{\varepsilon_2} \left( \frac{\text{VaR}_{\varepsilon_1}(X) - \text{VaR}_{c_1\varepsilon_1}(X)}{\text{VaR}_{\varepsilon_2}(X) - \text{VaR}_{c_1\varepsilon_1}(X)} \right) \right\}$ .

A lower bound on  $c_2$ . It holds that

$$(c_2\varepsilon_2 - c_1\varepsilon_1)\text{VaR}_{c_2\varepsilon_2}(X) \leq c_2\varepsilon_2\text{VaR}_{\varepsilon_2}(X) - c_1\varepsilon_1\text{VaR}_{\varepsilon_1}(X).$$

Subtracting  $(c_2\varepsilon_2 - c_1\varepsilon_1)\text{VaR}_{\varepsilon_2}(X)$  from both sides, we get

$$(c_2\varepsilon_2 - c_1\varepsilon_1)(\text{VaR}_{\varepsilon_2}(X) - \text{VaR}_{c_2\varepsilon_2}(X)) \geq c_1\varepsilon_1(\text{VaR}_{\varepsilon_1}(X) - \text{VaR}_{\varepsilon_2}(X)). \quad (10)$$

For  $t \in (0, 1/\varepsilon_2)$ , let

$$f(t) = (t\varepsilon_2 - c_1\varepsilon_1)(\text{VaR}_{\varepsilon_2}(X) - \text{VaR}_{t\varepsilon_2}(X)).$$

As we can see,  $f(1) = 0$ ,  $f(c_1\varepsilon_1/\varepsilon_2) = 0$  and  $f(t) \leq 0$  if  $t \in [\min\{1, c_1\varepsilon_1/\varepsilon_2\}, \max\{1, c_1\varepsilon_1/\varepsilon_2\}]$ . The  $f$  is increasing in the interval  $(\max\{1, c_1\varepsilon_1/\varepsilon_2\}, 1/\varepsilon_2)$ , decreasing in  $(0, \min\{1, c_1\varepsilon_1/\varepsilon_2\})$ . Hence, by (10), the lower bound for  $c_2$  is

$$\hat{c} = \inf \left\{ t \in (1, 1/\varepsilon_2] : \frac{(t\varepsilon_2 - c_1\varepsilon_1)(\text{VaR}_{\varepsilon_2}(X) - \text{VaR}_{t\varepsilon_2}(X))}{c_1\varepsilon_1(\text{VaR}_{\varepsilon_1}(X) - \text{VaR}_{\varepsilon_2}(X))} \geq 1 \right\}.$$

As  $c_1\varepsilon_1(\text{VaR}_{\varepsilon_1}(X) - \text{VaR}_{\varepsilon_2}(X)) > 0$ , we have  $\hat{c} \geq \max\{1, c_1\varepsilon_1/\varepsilon_2\}$ .  $\square$

## B Omitted proofs in Section 4.4

*Proof of Theorem 2.* By Proposition 3, for any  $X \in \mathcal{X}$ , we can find  $f \in \mathcal{C}$  satisfying (3) such that  $z_f(y) = 1/\pi_X(y) = c$  and  $X = f(U)$ . As  $z(y) = c$  is a continuously differentiable function, we know that all such  $f$  is characterized by the advanced differential equation (4). First, we show for any strictly decreasing solution  $f$  to (4) can be represented as

$$f(y) = C_0 + C_1y^\alpha + O(y^\zeta).$$

Let us start with (4). If  $z(y) = c$ , we need to solve  $f$  from

$$f(y) = f(cy) + cyf'(cy), \quad y \in (0, 1].$$

Even though  $f$  in the first place is considered on  $(0, 1]$ , given that  $c < 1$ , and this final equation, one

can expand it to the whole positive line:

$$f(y) = f(cy) + cyf'(cy), \quad y > 0.$$

Next, let  $x(t) = e^{-t}f(e^{-t})$  for  $t \in \mathbb{R}$  and  $a = -\log(c) > 0$ . This is equivalent to say that  $f(y) = x(-\log(y))/y$ . This changing variable simply gives the following delayed differential equation:

$$x'(t) = -e^{-a}x(t-a), \quad t \in \mathbb{R}.$$

Since we have assumed that  $f$  is strictly decreasing, i.e.,  $f' < 0$ , we have an extra restriction on  $x$ . Note that

$$x'(t) = -e^{-t}f(e^{-t}) - e^{-2t}f'(e^t) = -x(t) - e^{-2t}f'(e^t).$$

Thus, we have  $f' < 0 \Leftrightarrow x' + x > 0$ . Therefore, we are looking for a solution to the following delay differential equation (DDE):

$$\begin{cases} x'(t) = -e^{-a}x(t-a), \\ x'(t) + x(t) > 0, \end{cases} \quad t \in \mathbb{R}. \quad (11)$$

A standard approach of finding the solutions is to assume that they are in the form of a characteristic function  $t \mapsto e^{mt}$ . Putting this solution inside the equation, we get

$$me^{mt} = -e^{-a}e^{m(t-a)} \implies ame^{am} = (-a)e^{(-a)}.$$

This means any solution is given by  $x(t) = e^{mt}$  where  $m$  solves the characteristic equation

$$l(ma) = l(-a), \quad (12)$$

where  $l(x) = xe^x$ . Let  $b = l(-a)$ . As  $a > 0$ , we have  $b \in [-1/e, 0)$ . This equation has one obvious real solution at  $m_1 = -1$ . To find  $m$ , we need to know about the inverse of  $l$ . The inverse of the function  $l$  is known as the Lambert  $W$  function and plays an essential role in solving delayed and advanced differential equations.

From the Lambert  $W$  function, we know that  $l(z) = ze^z = b$  has two real solutions when  $b \in (-1/e, 0)$  and one real solution when  $b = -1/e$ . As illustrated by Figure 18, if  $0 < c < 1/e$ , the two real solutions are  $z_1 = -a < -1$  and  $z_2 = m_2a > -1$ ; thus,  $-1 < m_2 < 0$ . If  $0 < c < 1/e$ , the two real solutions are  $-1 < z_1 = -a < 0$  and  $z_2 = m_2a < -1$ ; thus,  $m_2 < -1$ . If  $c = 1/e$ , there is only one real solution  $z_1 = z_2 = -1$ ; thus  $m_2 = m_1 = -1$ .

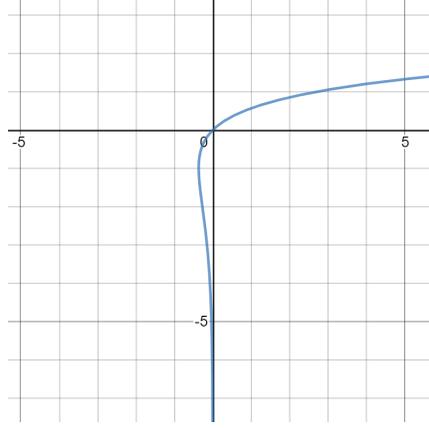


Figure 18: Lambert  $W$  function.

It is important to note that in general an equation like  $l(z) = ze^z = b$  has infinite complex roots. Let  $z = \theta + i\eta$ , and  $b \in [-1/e, 0)$ . In that regard, we have

$$\begin{aligned} b &= ze^z = (\theta + i\eta) e^{\theta+i\eta} \\ &= (\theta + i\eta) (\cos(\eta) + i \sin(\eta)) \\ &= (\theta \cos(\eta) - \eta \sin(\eta)) + i(\theta \sin(\eta) + \eta \cos(\eta)). \end{aligned}$$

This implies that  $\theta \sin(\eta) + \eta \cos(\eta) = 0$ , and  $b = e^\theta (\theta \cos(\eta) - \eta \sin(\eta))$ , leading to

$$\eta = 0, b = \theta e^\theta \quad \text{or} \quad \theta = -\frac{\eta}{\tan(\eta)}, b = -\frac{\eta \exp\left(-\frac{\eta}{\tan(\eta)}\right)}{\sin(\eta)}.$$

We plot the curves  $b = \theta e^\theta$  and  $\left(-\frac{\eta \exp\left(-\frac{\eta}{\tan(\eta)}\right)}{\sin(\eta)}, -\frac{\eta}{\tan(\eta)}\right)$  to find out the relation between  $b$  and the real part of the solution in Figure 19. The  $x$ -axis is  $b$  and the  $y$ -axis is  $\theta$ . The blue curve is associated with  $b = \theta e^\theta$ , which is essentially the principle branch of the Lambert  $W$  function. For any  $b$ , one can find the real values of the roots by fixing  $b$ . For instance, the green dashed line is associated with  $b = -0.12$ . As one can see, the curves intersect this line in infinite negative values. For  $b \in [-1/e, 0)$ , we can see that the real roots are greater than the real part of the complex roots. For more explanation of this, see [Siewert and Burniston \(1973\)](#).

Now assume that all the complex solutions for  $(am)e^{am} = (-a)e^{(-a)}$  are  $m_k = \lambda_k + \sigma_k i$  for  $k = 1, 2, 3, \dots$ , where  $(\lambda_1, \sigma_1) = (-1, 0)$  and  $(\lambda_2, \sigma_2) = (m_2, 0)$ . Based on the above discussions, we have

$$\lambda_1 = -1 > \lambda_2 = m_2 > \lambda_3 > \lambda_4 > \dots \quad \text{when } c \in (1/e, 1),$$

$$0 > \lambda_2 = m_2 > \lambda_1 = -1 > \lambda_3 > \lambda_4 > \dots \quad \text{when } c \in (0, 1/e),$$

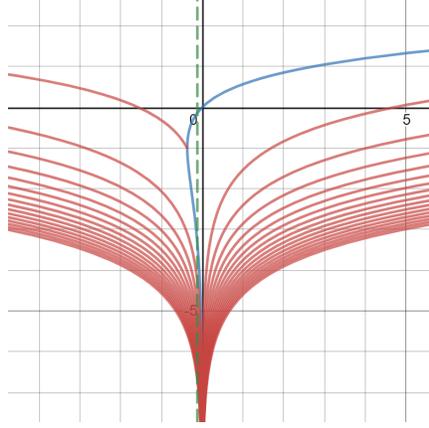


Figure 19: The real part of the Lambert  $W$  roots.

and

$$\lambda_1 = \lambda_2 = -1 > \lambda_3 > \lambda_4 > \dots \text{ when } c = 1/e.$$

Let

$$\mathbb{C} = \left\{ \begin{pmatrix} \lambda & -\sigma \\ \sigma & \lambda \end{pmatrix} = \lambda + \sigma i \mid \lambda, \sigma \in \mathbb{R} \right\}$$

be the set of all complex numbers. Then,

$$x_{\mathbb{C}}(t) = \exp \left\{ \begin{pmatrix} \lambda & -\sigma \\ \sigma & \lambda \end{pmatrix} t \right\}$$

is a complex solution that solves (11). It is clear that (11) still holds for the linear transform of  $x(t)$ . Therefore, for any two  $2 \times 1$  vector  $A$  and  $B$ ,

$$x(t) = A' x_{\mathbb{C}}(t) B = e^{\lambda t} (C_1 \cos(\sigma t) + C_2 \sin(\sigma t))$$

is also a solution to (11).

In [Bellman and Cooke \(1963\)](#), it is shown that all complex solutions to (11) can be represented as follows:

$$x_{\mathbb{C}}(t) = \sum_{k=1}^{\infty} C_k e^{m_k t}.$$

Putting it in the real-valued context, we have that all real-valued solutions are in the following form:

$$x(t) = C_1 e^{-t} + C_2 e^{m_2 t} + \sum_{k=3}^{\infty} e^{\lambda_k t} (C_{k,1} \cos(\sigma_k t) + C_{k,2} \sin(\sigma_k t)).$$

Now let us check  $x' + x > 0$ . This means that  $C_2$ ,  $\{C_{k,1}\}_{k \geq 3}$  and  $\{C_{k,2}\}_{k \geq 3}$  for all  $t$  must satisfy

$$x'(t) + x(t) = (1 + m_2)C_2e^{m_2t} + \sum_{k=3}^{\infty} e^{\lambda_k t} ((\lambda_k C_{k,1} + \sigma_k C_{k,2} + C_{k,1}) \cos(\sigma_k t) + (\lambda_k C_{k,2} - \sigma_k C_{k,1} + C_{k,2}) \sin(\sigma_k t)) > 0.$$

As  $\lambda_k < m_2$  for  $k \geq 3$ , we have  $\lim_{t \rightarrow \infty} (x'(t) + x(t)) / e^{m_2 t} = (1 + m_2)C_2$ . Therefore, we need  $C_2 < 0$  if  $m_2 < -1$  or  $C_2 > 0$  if  $m_2 > -1$ . That is  $C_2(1 + m_2) > 0$ . Then the solution can be written as

$$x(t) = C_1 e^{-t} + C_2 e^{m_2 t} + O(e^{\lambda_3 t}).$$

By a change of variable, we get

$$f(y) = C_1 + C_2 y^\alpha + O(y^\zeta),$$

where  $\alpha = -(1 + m_2)$  and  $\zeta = -(1 + \lambda_3)$ . As  $\lambda_3 < \min\{-1, m_2\}$  and  $C_2(1 + m_2) > 0$ , we have  $\zeta > \max\{0, \alpha\}$  and  $C_2 \alpha < 0$ . Also note that since  $m_2$  solves  $(am)e^{am} = (-a)e^{(-a)}$ , by replacing  $m_2 = -1 - \alpha$  and  $a = -\log(c)$ , we get  $(\alpha + 1)^{-1/\alpha} = c$ .  $\square$

## C Omitted proofs in Section 5

*Proof of Proposition 4.* Before proving the statements in the proposition, we introduce the following linking function, for  $X \in \mathcal{X}$ ,

$$\Gamma_X(\varepsilon) = 1 - F_X(\text{ES}_\varepsilon(X)), \quad \varepsilon \in [0, 1].$$

As  $X \in X$  it is easy to check that  $\Gamma_X$  satisfies Assumption 1 for any  $X \in \mathcal{X}$ . The domain of  $\Gamma_X(\varepsilon)$  is  $[0, 1]$  and its range is  $[0, 1 - F_X(\mathbb{E}[X])]$ .

As  $\text{ES}_\varepsilon(X) = \text{VaR}_{\Gamma_X(\varepsilon)}(X)$ , we have

$$\text{VaR}_{\Gamma_X(\varepsilon)}(X) = \text{ES}_\varepsilon(X) = \text{VaR}_{\varepsilon/\pi_X(\varepsilon)}(X) \quad \text{for } \varepsilon \in (0, 1].$$

Hence we have the simple relationship  $\Gamma_X(\varepsilon) = \varepsilon/\pi_X(\varepsilon)$ . Therefore,  $\Pi_X(\varepsilon) = \pi_X(\Gamma_X^{-1}(\varepsilon))$  and  $\pi_X(\varepsilon) = \Pi_X(\Gamma_X(\varepsilon))$ . The function  $\Gamma_X$  yields an association between a point on the PELVE on  $(0, 1 - F_X(\mathbb{E}[X])]$  and a point on the dual PELVE curve on  $(0, 1]$  with the same value. Furthermore, we have  $\pi_X$  is continuous on  $(0, 1)$  as  $\pi_X(\varepsilon) = \varepsilon/\Gamma_X(\varepsilon)$  and  $\Gamma_X$  is continuous.

Next, we show the statements (i)-(iv). The equivalence (i) of monotonicity of  $\Pi_X$  and that of  $\pi_X(\cdot)$  follows from  $\Pi_X(\varepsilon) = \pi_X(\Gamma_X^{-1}(\varepsilon))$ ,  $\pi_X(\varepsilon) = \Pi_X(\Gamma_X(\varepsilon))$  and that  $\Gamma_X$  is increasing.

For (ii), (iii) and (iv), we first show that  $\Gamma_X$  is location-scale invariant and shape relevant (in the sense of (13)). Assume that  $f : \mathbb{R} \rightarrow \mathbb{R}$  is a strictly increasing concave function such that  $f(X) \in \mathcal{X}$ . By Jensen's inequality and the dual representation of  $\text{ES}_p$ , we have

$$\text{ES}_p(f(X)) \leq f(\text{ES}_p(X))$$

for all  $p \in (0, 1)$ . This statement can be found in Appendices A.2 in Li and Wang (2022). Therefore,

$$\begin{aligned} \Gamma_{f(X)}(\varepsilon) &= 1 - F_{f(X)}(\text{ES}_\varepsilon(f(X))) \\ &\geq 1 - F_{f(X)}(f(\text{ES}_\varepsilon(X))) = 1 - F_X(\text{ES}_\varepsilon(X)) = \Gamma_X(\varepsilon). \end{aligned} \quad (13)$$

Then, we have  $\Gamma_{f(X)}(\varepsilon) \geq \Gamma_X(\varepsilon)$  for all strictly increasing concave functions:  $f : \mathbb{R} \rightarrow \mathbb{R}$  with  $f(X) \in \mathcal{X}$ .

For any strictly increasing convex function  $g : \mathbb{R} \rightarrow \mathbb{R}$  with  $g(X) \in \mathcal{X}$ , we can take  $f(x) = g^{-1}(X)$ , which is a strictly increasing concave function. Therefore, we have  $\Gamma_{g(X)}(\varepsilon) \leq \Gamma_X(\varepsilon)$  for all strictly increasing convex functions  $g$ .

For  $\lambda > 0$  and  $a \in \mathbb{R}$ , we have that  $f(x) = \lambda X + a$  is both convex and concave. Therefore,  $\Gamma_{\lambda X + a}(\varepsilon) = \Gamma_X(\varepsilon)$  for all  $\varepsilon \in [0, 1]$ . In conclusion, we have the following results for  $\Gamma$ .

- (1) For all  $\lambda > 0$  and  $a \in \mathbb{R}$ ,  $\Gamma_{\lambda X + a}(\varepsilon) = \Gamma_X(\varepsilon)$ .
- (2)  $\Gamma_{f(X)}(\varepsilon) \geq \Gamma_X(\varepsilon)$  for all strictly increasing concave functions:  $f : \mathbb{R} \rightarrow \mathbb{R}$  with  $f(X) \in \mathcal{X}$ .
- (3)  $\Gamma_{g(X)}(\varepsilon) \leq \Gamma_X(\varepsilon)$  for all strictly increasing convex functions:  $g : \mathbb{R} \rightarrow \mathbb{R}$  with  $g(X) \in \mathcal{X}$ .

Then, we have (ii), (iii) and (iv) from  $\pi_X(\varepsilon) = \varepsilon/\Gamma_X(\varepsilon)$ . □

*Proof of Theorem 3.* The idea is to prove that if  $1/\eta$  is convex (concave), then  $x \mapsto F^{-1}((1-p)F(x)+p)$  is convex (concave) for all  $p \in (0, 1)$ . Then, we can get the desired result by Proposition 5. We will use the following steps to show this statement.

Step 1. Let  $s(x) = \log(1 - F(x))$  for  $x \in (\text{ess-inf}(X), \text{ess-sup}(X))$ . Then,  $s$  is a continuous and strictly decreasing function and  $s(x) < 0$ . Let  $s^{-1}$  be the inverse function of  $s$ . Now, we have

$$F(x) = 1 - e^{s(x)}, \quad x \in (\text{ess-inf}(X), \text{ess-sup}(X))$$

and

$$F^{-1}(t) = s^{-1}(\log(1 - t)), \quad t \in (0, 1).$$

Therefore,

$$\begin{aligned}
F^{-1}((1-p)F(x)+p) &= s^{-1}(\log(1-(1-p)F(x)-p)) \\
&= s^{-1}(\log(1-F(x))+\log(1-p)) \\
&= s^{-1}(\log(e^{s(x)})+\log(1-p)) \\
&= s^{-1}(s(x)+\log(1-p)).
\end{aligned}$$

Let  $\theta = \log(1-p)$ . It follows that the statement that  $x \mapsto F^{-1}((1-p)F(x)+p)$  is convex (concave) for all  $p \in (0, 1)$  is equivalent to the statement that  $x \mapsto s^{-1}(s(x)+\theta)$  is convex (concave) for all  $\theta < 0$ .

Step 2. Let  $g(x) := -s^{-1}(x)$ . Then,  $g$  is strictly increasing. We will show that if  $1/\eta$  is convex (concave),  $\log(g'(x))$  is convex (concave).

As  $g(x) = -s^{-1}(x) = -S^{-1}(e^x)$ , we have

$$g'(x) = \frac{e^x}{f(S^{-1}(e^x))}.$$

Let  $H(x) := \log(g'(x)) = x - \log(f(S^{-1}(e^x)))$ . We have

$$H(\log(S(x))) = \log S(x) - \log f(x) = -\log \eta(x). \quad (14)$$

Then, taking the derivative on both sides of (14), we get

$$\begin{aligned}
-H'(\log(S(x)))\eta(x) &= -\frac{\eta'(x)}{\eta(x)} \\
\iff H'(\log(S(x))) &= \frac{\eta'(x)}{\eta^2(x)} = -\frac{d}{dx} \left( \frac{1}{\eta(x)} \right).
\end{aligned}$$

Taking a derivative in both sides again, we get

$$-H''(\log(S(x)))\eta(x) = -\frac{d^2}{dx^2} \left( \frac{1}{\eta(x)} \right).$$

Then,  $1/\eta$  is a convex (concave) function means  $H''(x) \geq 0$  ( $H''(x) \leq 0$ ), which gives that  $\log g'(x)$  is convex (concave).

Step 3. For  $\theta < 0$ , let  $G_\theta(x) := s^{-1}(s(x)+\theta)$ . We are going to show

$$\lim_{z \rightarrow 0} \frac{G_\theta(x+z) - G_\theta(x)}{z} \leq \lim_{z \rightarrow 0} \frac{G_\theta(x'+z) - G_\theta(x')}{z} \quad (15)$$

for all  $x < x'$ .

We take  $z > 0$  first. As  $s$  is strictly decreasing,  $s^{-1}$  is also strictly decreasing. Then,  $G_\theta$  is a continuous and strictly increasing function. As  $\theta < 0$ , we also have  $G_\theta(x) > x$ . Take arbitrary  $x, x', y$

and  $z$  such that  $x < x'$ ,  $x < y$  and  $z > 0$ . Let  $\theta = s(y) - s(x)$ . Then, we have  $G_\theta(x) = y$ . Define

$$h = G_\theta(x + z) - y, \quad y' = G_\theta(x') \quad \text{and} \quad h' = G_\theta(x' + z) - y'.$$

By the definition of  $G_\theta$ , we have  $s(y + h) = s(x + z) + \theta$ ,  $s(y') = s(x') + \theta$  and  $s(y' + h') = s(x' + z) + \theta$ .

As a result, we have

$$s(y + h) - s(y) = s(x + z) - s(x) \quad \text{and} \quad s(y' + h') - s(y') = s(x' + z) - s(x').$$

By the mean-value theorem, there exists  $\zeta \in (y, y + h)$ ,  $\zeta' \in (y', y' + h')$ ,  $\xi \in (x, x + z)$  and  $\xi' \in (x', x' + z)$  such that

$$s'(\zeta)h = s'(\xi)z \quad \text{and} \quad s'(\zeta')h' = s'(\xi')z.$$

Furthermore,  $x, x', y$  and  $y'$  satisfy  $x < x' < y'$  and  $x < y < y'$ . If  $z$  is small enough, then  $h$  and  $h'$  will also be small enough as  $G_\theta$  is continuous. Therefore, we have  $\xi < \xi' < \zeta'$  and  $\xi < \zeta < \zeta'$  when  $z$  is small enough.

In Step 2, we have that  $\log g'(x)$  is convex when  $1/\eta$  is convex. Therefore, we get

$$\log(g'(a)) + \log(g'(b)) \geq \log(g'(a')) + \log(g'(b')),$$

for all  $a < a' < b$  and  $a < b' < b$ , which means

$$g'(a)g'(b) \geq g'(a')g'(b').$$

As  $g(x) = -s^{-1}(x)$ , we have

$$\frac{1}{s'(s^{-1}(a))s'(s^{-1}(b))} \geq \frac{1}{s'(s^{-1}(a'))s'(s^{-1}(b'))}.$$

As  $s^{-1}(x)$  is strictly decreasing, it means that

$$s'(\alpha)s'(\beta) \leq s'(\alpha')s'(\beta')$$

for  $\alpha > \alpha' > \beta$  and  $\alpha > \beta' > \beta$ . Therefore, we have  $s'(\xi)s'(\zeta') \leq s'(\xi')s'(\zeta)$  as  $\zeta' > \zeta > \xi$  and  $\zeta' > \xi' > \xi$ . That is,

$$h = \frac{s'(\xi)z}{s'(\zeta)} \leq \frac{s'(\xi')z}{s'(\zeta')} = h'.$$

On the other hand,  $h = G_\theta(x + z) - G_\theta(x)$  and  $h' = G_\theta(x' + z) - G_\theta(x')$ . Therefore,

$$G_\theta(x + z) - G_\theta(x) \leq G_\theta(x' + z) - G_\theta(x')$$

when  $z$  is small enough. If  $z < 0$ , we can also get (15) by an analogous argument.

Hence, the second-order derivative of  $G_\theta$  is increasing for each  $\theta < 0$ , which means that  $x \mapsto F^{-1}((1 - p)F(x) + p)$  is convex for all  $p \in (0, 1)$  if  $1/\eta$  is convex.

An analogous argument yields that  $x \mapsto F^{-1}((1 - p)F(x) + p)$  is concave for all  $p \in (0, 1)$  when  $1/\eta$  is concave. □