Non-concave portfolio optimization with Average Value-at-Risk

Fangyuan Zhang

Data Science Department, Eurecom, Campus SophiaTech, Biot, 06410, France.

E-mail: fangyuan.zhang@eurecom.fr;

Abstract

Average Value-at-Risk (AVaR) is a potential alternative to Value-at-Risk in the financial regulation of banking and insurance institutions. To understand how AVaR influences a company's investment behavior, we study portfolio optimization under the AVaR constraint. Our main contribution is to derive analytical solutions for non-concave portfolio optimization problems under the AVaR constraint in a complete financial market by quantile formulation and the decomposition method, where the non-concavity arises from assuming that the company is surplus-driven. Given the AVaR constraint, the company takes three investment strategies depending on its initial budget constraint. Under each investment strategy, we derive the fair return for the company's debt holders fulfilling the risk-neutral pricing constraint in closed form. Further, we illustrate the above analytical results in a Black-Scholes market. We find that the fair return varies drastically, e.g., from 4.99% to 37.2% in different situations, implying that the company's strategy intimately determines the default risk faced by its debt holders. Our analysis and numerical experiment show that the AVaR constraint cannot eliminate the company's default risk but can reduce it compared with the benchmark portfolio. However, the protection for the debt holders is poor if the company has a low initial budget.

Keywords: Average Value-at-Risk, Non-concave portfolio optimization, Risk-neutral pricing constraint, Quantile formulation

1 Introduction

Average Value-at-Risk (AVaR), a.k.a. Tail Value-at-Risk, or Conditional Value-at-Risk, is an alternative risk measure to Value-at-Risk (VaR) for market risk in banking regulation (see the report of Basel Committee on Banking Supervision [7]).[1] Meanwhile, it is also a significant risk measure in insurance regulation (see Solvency II and the Swiss Solvency Test). Mathematically, AVaR measures the average quantile of a random variable in its left tail, which can be interpreted as a portfolio's expected loss in the worst economic states. AVaR has several competent properties, e.g., it is a coherent risk measure [1, 4], and it encourages portfolio diversification and punishes risk concentration [34], which make AVaR a preferable risk measure in the current framework of financial regulation.

To understand the impact of financial regulation on a company's investment behavior, there is growing literature studying portfolio optimization problems using the risk measure (VaR or AVaR) as a constraint. A popular framework is to consider an expected utility maximization problem under one or multiple risk constraints. For example, since the seminal work by Basak and Shapiro [6], portfolio optimization problems under the VaR constraint have been extensively studied ([12, 13, 16]).[2] However, compared to VaR, the number of relevant studies on companies' portfolio choices under AVaR is much less. A recent work by Wei [36] studies portfolio optimization under the AVaR constraint in the standard framework of expected utility theory.

This paper contributes to the above stream of literature by providing analytical solutions for portfolio optimization problems under the AVaR constraint. Different from [36], we consider a non-concave utility maximization problem under the AVaR constraint. Non-concave settings offer the standard expected utility framework more flexibility to incorporate behavioral observations or explore contracts with more complicated payoff designs. For instance, one can consider an S-shape utility function from cumulative prospect theory [33] to model loss aversion [3, 8, 18], introduce a benchmark to asset allocation problems [5], or construct non-linear contingent payoff functions [10, 11]. In the literature, a non-concave utility maximization problem without risk constraints has been tackled by different methods, e.g., the concavification method in [10] and [30], the decomposition method in [25] and [17], and the Lagrangian approach in [8], [5] and [11]. However, regarding incorporating risk constraints, only a few recent studies investigate the non-concave optimization under the VaR constraint [18, 29]. The models in [17] and [3] analyze the impact of AVaR on non-concave optimization in restrictive settings. The model in [3] illustrates that the AVaR constraint is *useless* in constraining

[1] These different terminologies lead to the same mathematical concept although some researchers may prefer one term over the others depending on specific problems. This paper considers a continuous probability space, where these terminologies define the same function. We choose the name AVaR to be consistent with Föllmer and Schied [21].

[2] For instance, [13, 16] extend the portfolio optimization problem to a multi-period setting with multiple VaR constraints. [12] considers the portfolio optimization problem under a combined VaR and portfolio insurance constraint.

the company's investment behavior, assuming that the company can incur an arbitrarily large loss.[3] In contrast, we consider a company with limited liability and non-negative wealth. We show that AVaR *can reduce the company's default probability* compared with the benchmark portfolio. The analysis in [17] restricts the AVaR constraint to the affine or the convex part of the problem. To incorporate all possible solutions, we allow for an arbitrary AVaR constraint that can intervene in the affine and the concave part of the problem. Our study compensates for the existing works on non-concave portfolio optimization under the AVaR constraint by providing full explicit treatment.

Non-concavity and the AVaR constraint make it nontrivial to solve the optimization problem explicitly. The major challenge is that the AVaR constraint as a quantile-based function is defined on the real interval $[0, 1]$, while the expected (non-concave) utility function to be maximized is defined on the physical probability space. Hence, to tackle the constrained optimization problem, we should either construct an equivalent objective function on the real interval $[0, 1]$ or transfer the AVaR constraint to equivalent risk constraints on the physical probability space. A recent study by Chen, Stadje, and Zhang [14] follows the second method and transfers the AVaR constraint to equivalent Expected Shortfall constraints, also called Limited Expected Loss constraints, on the physical probability space.[4] Then, by solving the optimization problems under the equivalent ES constraints first, they derive the set of optimal solutions that contain the one under the AVaR constraint. Different from their method, we construct an equivalent objective function on the real interval $[0, 1]$ by quantile formulation and solve the non-concave optimization problem under the AVaR constraint by the decomposition method. We compare and discuss the connection between the two methods in Section 4.

Quantile formulation, formally introduced by He and Zhou [23] and Xu [37], attempts to solve complicated portfolio optimization problems with law-invariant objectives explicitly. The idea is to construct an equivalent optimization problem on $[0, 1]$ based on the quantile functions of the portfolio. It has been applied to study Yaari's dual problem, rank-dependent utility maximization, cumulative prospect theory, and utility maximization under weighted VaR; for details, see [23, 32, 35, 36].

The decomposition method is used in [25], and [17] to address the non-concavity along with an S-shape utility function or a non-linear contingent payoff function. The idea is to decompose the choice variable (e.g., the terminal wealth of the company) into two random variables, X and Y, such that the non-concave objective function is strictly concave in X, and affine or convex in Y. Thus, after solving two sub-optimization problems (i.e., maximizing the concave and minimizing the convex), determining the global maximum is equivalent to finding the optimal decomposition. However, due to the AVaR constraint, the optimal decomposition involves several cases depending on the

[3] The company's wealth can be negative in this model, but its economic interpretation is unclear.
[4] Under some conditions, the AVaR constraint is equivalent to the Expected Shortfall constraint. See Lemma 4.1 in [14].

initial budget constraint and hence is complicated to determine. We provide detailed proofs in the appendix.

We motivate our model by considering optimal investment choices for a surplus-driven financial company with limited liability under the AVaR constraint. The company's surplus, i.e., the positive difference between its asset and its liability, is non-linear in the terminal asset. Thus, assuming the company is surplus-driven, the corresponding utility maximization problem is non-concave. This setting arises from the fact that most managerial boards of financial companies represent their shareholders' benefits, which are positive if there are positive surplus. By explicitly distinguishing the debt holders' and the shareholders' benefits, it is straightforward to analyze whether and how the financial regulation can protect the debt holders' benefits. This setting is in line with Chen, Stadje and Zhang [14].

We find that the surplus-driven company adopts three different investment strategies depending on its initial budget constraint. The multiple structures of optimal solutions in the non-concave setting are different from the unique structure of optimal solutions in the standard concave setting; See, e.g., [14] and [18]. Further, each optimal solution has a set of zero assets in its most left tail, which implies that the company will default in the worst financial states. This observation shows two important information: a) The AVaR constraint alone is not enough to prevent the company's default in the worst scenarios; b) Since the set of zero assets depends on the initial budget constraint, the default risk faced by the debt holders varies in different solutions.

According to these findings, a natural question is how the AVaR constraint protects the debt holders in different cases. Following [11], we derive the fair return for the debt holders fulfilling the risk-neutral pricing constraint. The fair return reflects the magnitude of the default risk faced by the debt holders, that is, the higher the fair return, the more significant the default risk. We conduct numerical experiments in a Black-Scholes market. Our results show that the AVaR constraint can reduce the company's default probability with each investment strategy compared with the benchmark portfolio. However, if the company has a low initial budget, the AVaR constraint provides poor protection for the debt holders.

There exist abundant studies comparing VaR and AVaR as risk measures in financial regulation. However, most studies compare them from the viewpoint of statistical properties. For instance, AVaR obeys sub-additivity (i.e., the risk of combining portfolios is smaller than not combining), and hence is a coherent risk measure, which is superior to VaR, see [1, 2, 4]. In contrast, VaR has elicitability (i.e., estimating VaR is equivalent to minimizing the objective forecasting function [27]), and hence can be back-tested, which is more advantageous than AVaR. For a more comprehensive comparison of VaR and AVaR, see, for example, [19, 20] and the references therein. This paper adds one more degree to comparing VaR and AVaR as standard risk measures in financial regulation. Our results confirm the finding in Chen, Stadje and Zhang [14] that VaR and AVaR have similar regulatory effects in the non-concave portfolio optimization problem, which is to reduce the company's default probability but cannot fully illuminate it. Further, our numerical examples show that the AVaR constraint provides poor protection for the debt holders if the company has a low initial budget.

The paper proceeds as follows. Section 2 introduces the definition of the AVaR constraint and the constrained portfolio optimization problem. Section 3 contains the optimal solution for the constrained optimization problem. Especially, this section demonstrates how to apply quantile formulation and the decomposition to solve non-concave optimization. Section 4 discusses the connection between the terminal wealth method in Chen, Stadje and Zhang [14] and our method. Section 5 provides numerical illustrations in a Black-Scholes market, including the pre-horizon wealth and the optimal investment strategies. The same section also discusses the fair return for the debt holders under each optimal solution. Section 6 concludes. Technical proofs are put in the appendix.

2 Model Setup

2.1 The financial market

We assume a *complete* financial market without transaction costs in continuous time that contains one traded risk-free asset S_0 (the bank account) and m traded risky assets denoted by the stochastic processes $\boldsymbol{S} = (S_1, \cdots, S_m)'$.[5] We fix a filtered probability space $(\Omega, \mathcal{F} = (\mathcal{F}_t)_{t\in[0,T]}, \mathbb{P})$, $T < \infty$. The unique local martingale measure is denoted by $\mathbb{Q}$. The state price density process is defined by $\xi_T := \frac{S_0(0)\mathrm{d}\mathbb{Q}}{S_0(T)\mathrm{d}\mathbb{P}}$.[6] The financial institution endowed with an initial capital x_0 chooses an investment strategy that we describe by $\pi_i(t)$, the units of ith risky asset in the portfolio at time t. We assume that $\boldsymbol{\pi(t)} = (\pi_1(t), \cdots, \pi_m(t))$ is adaptive with respect to the filtration $\mathcal{F} = (\mathcal{F}_t)_{t\in[0,T]}$. The strategy is self-financing (i.e., no intermediate income) such that

$$X_t^\pi = X_0 + \sum_{i=0}^{m} \int_0^t \pi_i(s)\mathrm{d}S_i(s) = X_0 + \int_0^t \boldsymbol{\pi(s)}\mathrm{d}\boldsymbol{S(s)}, \quad X_0 = x_0 > 0.$$

In addition, the set of attainable terminal wealth is defined by

$$\mathcal{X} := \{X_T^\pi \geq 0, \text{is } \mathcal{F}_T\text{-measurable, replicable, and} \quad \mathbb{E}[\xi_T X_T^\pi] = x_0\}.^7$$

[5] $\boldsymbol{S}$ is an m-dimensional vector and $'$ denotes the transposed sign.

[6] The state price density ξ_T is defined in this way such that the discounted asset process is a local martingale under the risk-neutral probability measure $\mathbb{Q}$.

[7] $\mathbb{E}[]$ denotes the expectation under the physical probability $\mathbb{P}$.

Notice that in a complete financial market, it is sufficient to determine the optimal terminal wealth from the set of attainable wealth [15, 26], and the corresponding investment strategy can be obtained by the martingale approach.[8] Hence, from now on we omit the dependence of X_T on π and focus on finding the optimal terminal wealth. Without ambiguity, we use ξ instead of ξ_T to denote the state price density at time T.

2.2 The optimal asset allocation problem with the AVaR constraint

We consider a surplus-driven financial institution operating in $[0, T]$, $T < \infty$. At time 0, the company's asset consists of the equity, E_0, from the shareholders and the liability, D_0, from the debt holders. Hence, at time zero, the company's asset value is given by $x_0 = E_0 + D_0$. We assume that the company has limited liability, i.e., the terminal payoff to the debt holders at horizon T is given by

$$\varphi_L(X_T) = \min(D_T, X_T).$$

The debt holders are fully paid back if the terminal asset of the company can cover the promised payoff D_T to them. Otherwise, the debt holders have the priority to claim the residual of the company's asset. In contrary, the equity holders obtain the positive difference between the company's asset and the payoff to the debt holders, i.e.,

$$\varphi_E(X_T) = X_T - \varphi_L(X_T) = \max(X_T - D_T, 0) =: (X_T - D_T)^+.$$

We assume that the debt holders obtain a deterministic return, i.e., $D_T = D_0 \exp(gT), g > 0$. Since the company has limited liability, the debt holders face the default risk of the company. Therefore, the return g has to be larger than the risk-free return in the financial market, i.e., $g > r > 0$, due to the no-arbitrage assumption. Intuitively, the return to the debt holders should reflect the default risk faced by them, which leads to the so-called fairness contract problem [11]. We will discuss how to determine the fair return to the debt holders in Section 5.

In line with Chen, Stadje and Zhang [14], we assume that the surplus-driven financial company makes investment decision maximizing the expected utility of the surplus, i.e., the positive difference between the company's asset and its debt. This is a reasonable assumption since, in reality, the board members of a company usually represent the benefits of its shareholders. Considering the AVaR constraint in the financial regulation, we introduce the following optimization problem,

[8] After determining the optimal terminal wealth, one can calculate the optimal investment strategy for a constant relative risk aversion (CRRA) investor through a standard procedure via Itô's Lemma in a Black-Scholes market, see the examples in [5, 6, 10, 12].

Problem AVaR

$$\max_{X_T \in \mathcal{X}} \quad \mathbb{E}[U((X_T - D_T)^+)], \text{ subject to } \frac{1}{\alpha}\int_0^\alpha VaR_{X_T}(\beta)\mathrm{d}\beta \geq L, \ \mathbb{E}[X_T\xi]{=}x_0, \tag{1}$$

$$\text{with} \quad VaR_{X_T}(\beta) := \inf\{x|\mathbb{P}(X_T \leq x) \geq \beta\},$$

where $U(\cdot)$ is the utility function of the company's managerial board, which is strictly increasing, strictly concave and twice continuously differentiable. The value $L > 0$ is the regulatory threshold for the average quantile of the terminal wealth in the left tail. Thus, the optimal terminal asset X_T^* (if it exists) maximizes the expected utility of the company's surplus, and at the same time, has a minimum average quantile L in the worst financial scenarios.

2.3 Technical Assumptions

Before proceeding to the optimal solution for the Problem AVaR (1), we present several technical assumptions for the optimization problem.

Assumption 1: The utility function satisfies the Inada and Asymptotic Elasticity (AE) conditions:

$$\textbf{Inada:} \quad U'(0) = \lim_{x\to 0} U'(x) = \infty, \quad U'(\infty) = \lim_{x\to\infty} U'(x) = 0, \tag{2}$$

$$\textbf{AE:} \quad \lim_{x\to\infty} \sup \frac{xU'(x)}{U(x)} < 1, \tag{3}$$

where $U'(\cdot)$ denotes the first derivative of the utility function, also called the marginal utility.

Assumption 2: Letting $I(\cdot) = (U')^{-1}(\cdot)$ denote the inverse function of the marginal utility, we assume that

$$\mathbb{E}[U(I(\lambda\xi)] < \infty, \quad \mathbb{E}[\xi I(\lambda\xi)] < \infty,$$

where $\lambda > 0$ is a positive real number.

Assumption 3: The state price density ξ is atomless. In addition,

$$\text{ess inf } \xi := \inf\{c \in \mathbb{R}|\mathbb{P}(\xi < c) > 0\} = 0, \tag{4}$$

$$\text{ess sup } \xi := \sup\{c \in \mathbb{R}|\mathbb{P}(\xi > c) > 0\} = +\infty. \tag{5}$$

Let $F_\xi(\cdot)$ and $F_\xi^{-1}(\cdot)$ denote the cumulative distribution function and its inverse function (i.e., the quantile function of ξ), respectively. We assume that $F_\xi^{-1}(\cdot)$ is continuous. Since ξ is assumed to be atomless, we have that $F_\xi^{-1}(\cdot)$

is strictly increasing. Moreover,

$$F_\xi^{-1}(0) = 0, \quad F_\xi^{-1}(1) = +\infty. \tag{6}$$

These are standard technical assumptions when dealing with a utility maximization problem in a complete financial market; See, for instance, [23, 36]. Later we will illustrate the optimal solutions in a Black Scholes financial market (Section 5), where the state price density ξ is log-normal distributed and satisfies the above assumptions.

Assumption 4: The utility function is bounded from below, so we assume $U(0) = 0$ without loss of generality. A bounded utility function means that the company feels " finite pain" by incurring default in the worst financial scenarios, indicating that default risk exists. In addition, the Problem AVaR (1) is law invariant. Thus, adding a finite constant to the utility function will not change the optimal solution.

For the Problem AVaR (1), an unbounded utility function ($U(0) = -\infty$) will lead to a trivial solution. The decision-maker will take a portfolio insurance strategy in the worst financial scenarios in the first place, implying that the regulation is redundant. See Chen, Stadje and Zhang [14].

3 The optimal solution for the Problem (1)

The major challenge in solving the Problem AVaR (1) is that the maximization problem and the AVaR constraint are defined in different sets. To be concrete, the expected utility to be maximized is a function of the terminal wealth X_T, defined on the probability space Ω, while the AVaR constraint is a quantile function defined on the real interval $[0, 1]$. To proceed with the Problem AVaR (1), we can either transfer the quantile-based AVaR constraint to an equivalent terminal wealth-based constraint defined on the probability space Ω, or transfer the terminal wealth-based maximization problem to an equivalent quantile-based maximization problem for functions defined on $[0, 1]$. Chen, Stadje and Zhang [14] study the first method. With this method, they derive a set of solutions that include the optimal solution for the Problem AVaR (1). However, this method is not enough to determine the exact optimal solution. Our work applies the second method to provide the unique optimal solution for the Problem AVaR (1). Later, we also discuss the connection between the two methods in Section 4.

3.1 Quantile formulation

Note that the expected utility is law-invariant, i.e., different random variables with the same distribution will give the same expected utility. Since the quantile function of a random variable always has the same distribution as the original random variable, the optimization problem based on the quantile function of the terminal wealth is equivalent to the original problem. After obtaining the optimal quantile function, we can determine the *optimal terminal wealth* via its unique relationship with the *optimal quantile function* in a complete financial market. This procedure is called quantile formulation ([23, 37]). Moreover, Proposition C.1 in Jin and Zhou [25] gives that if $Q_X^*(\cdot)$ is the optimal solution to the equivalent quantile-based optimization problem, then the optimal terminal wealth to the original optimization problem is given by

$$X_T^* = Q_X^*(1 - F_\xi(z)), \quad z \in [0,1]. \tag{7}$$

A general case involving the distortion of the probability measure can be found in Xu [37]. Equation (7) indicates that the optimal terminal wealth is decreasing with the state price density, i.e., the higher the state price density, the lower the terminal wealth. This decreasing relationship is reasonable because the state price density represents the price for the Arrow-Debreu security.[9] A higher price of the Arrow-Debreu security implies a lower return for this state and vice versa.

Quantile formulation provides a different perspective on the terminal wealth-based expected utility maximization problem. It has been applied in solving, e.g., a rank-dependent utility maximization problem [9] or a *concave* utility maximization problem under the AVaR constraint [36]. In this work, we apply quantile formulation to solve the *non-concave* utility maximization problem under the AVaR constraint.

3.2 The benchmark problem

We first study the unconstrained non-concave utility maximization problem, i.e., the benchmark problem. The benchmark problem is given by

$$\max_{X_T \in \mathcal{X}} \mathbb{E}[U((X_T - D_T)^+)], \quad \text{subject to} \quad \mathbb{E}[X_T \xi] = x_0. \tag{8}$$

The optimal solution to the benchmark problem (8) is given in the following proposition. Let $\widetilde{F}(z) = F_\xi^{-1}(1-z), z \in [0,1]$.

Proposition 1 *Let $\mathbb{1}_{\mathcal{A}}(\cdot)$ be an indicator function on a set $\mathcal{A}$ such that*

$$\mathbb{1}_{\mathcal{A}}(x) = \begin{cases} 1, & \text{if} \quad x \in \mathcal{A}; \\ 0, & \quad\text{otherwise.} \end{cases}$$

Let $\lambda_B > 0$ and $z_D^* \in (0,1)$ satisfy the following two equations:

$$-U(I(\lambda_B \widetilde{F}(z_D^*))) + \lambda_B I(\lambda_B \widetilde{F}(z_D^*))\widetilde{F}(z_D^*) + \lambda_B \widetilde{F}(z_D^*) D_T = 0, \tag{9}$$

$$\int_{z_D^*}^1 (I(\lambda_B \widetilde{F}(z)) + D_T) \mathrm{d}z = x_0. \tag{10}$$

[9]The Arrow-Debreu security, a.k.a state-price security, is a security that pays one unit of a numeraire (e.g., dollar) if a particular state occurs and pays zero if other states occur.

Then, the optimal solution to the benchmark problem (8) *is given by*

$$X_T^B = (I(\lambda_B \xi) + D_T)\mathbb{1}_{\xi < \xi_D^*}, \tag{11}$$

where $\xi_D^* = \widetilde{F}(z_D^*)$.

Note that the benchmark problem has been extensively studied in the literature by different methods, e.g., the concavification [10, 30] or the Lagrangian approach [5, 8, 14]. The optimal solution (Proposition 1) is not new to the literature. However, we are particularly interested in obtaining this solution by quantile formulation. The reasons are the following:

1. Quantile formulation alone is not enough to solve the Problem AVaR (1) and the benchmark problem (8) because the utility of surplus $U((X_T - D_T)^+)$ is not strictly concave. For this, we will apply the decomposition method. Different from previous works, we apply the idea of decomposition to the set of quantile functions. Moreover, we find the optimal decomposition through the first-order condition, which has an intuitive interpretation. The benchmark problem is much simpler than the Problem AVaR (1). Hence, we will showcase the technique using the benchmark problem as an example.
2. In addition to the technical challenge, non-concave utility maximization also brings new economic implications. Assuming that the utility function is bounded from below, a *surplus-driven* company will inevitably default in the worst financial scenarios. Having the benchmark behavior in mind is necessary to understand the economic impact of the AVaR constraint on the surplus-driven company.

In the next section, we demonstrate solving the benchmark problem (8) by quantile formulation and the decomposition method.

3.3 The equivalent quantile-based benchmark problem

Let $\mathcal{Q}$ denote the set of all quantile functions $Q_X(z)$ of the attainable terminal wealth (Section 2.1), i.e.,

$$Q_X(z) := \inf\{x | \mathbb{P}(X \le x) \ge z,\}, z \in [0,1]; \quad \mathcal{Q} := \{Q_X(z), z \in [0,1] | X \in \mathcal{X}\}.$$

The quantile-based benchmark problem is given by

$$\max_{Q_X \in \mathcal{Q}} \int_0^1 U((Q_X(z) - D_T)^+)\mathrm{d}z, \text{ subject to } \int_0^1 Q_X(z)\widetilde{F}(z)\mathrm{d}z = x_0. \tag{12}$$

Observe that

$$(Q_X(z) - D_T)^+ = \begin{cases} Q_X(z) - D_T > 0, & \text{if} \quad Q_X(z) > D_T, \\ 0, & \text{if} \quad Q_X(z) \le D_T. \end{cases}$$

Given a terminal debt D_T, we define the inverse of a quantile function at D_T as follows:

$$z(D_T) \equiv Q_X^{-1}(D_T) := \max\{z \in (0,1) | Q_X(z) \leq D_T\}.$$

Since quantile functions are non-decreasing, we have that $Q_X(z) > D_T$ if $z > z(D_T)$. Let us consider an arbitrary probability $z_D \in [0,1]$ and define a subset of quantile functions in the following way:

$$\mathcal{Q}(z_D) := \{Q_X(z) \in \mathcal{Q} | z(D_T) \leq z_D\}, \; \mathcal{Q} = \bigcup_{z_D \in [0,1]} \mathcal{Q}(z_D).$$

Note that the quantile functions in $\mathcal{Q}(z_D)$ are larger than D_T on $z > z_D$. Moreover, the union of the sets $\mathcal{Q}(z_D), z_D \in [0,1]$ is the set of quantile functions of the attainable terminal wealth.

In addition, we express the quantile functions in set $\mathcal{Q}(z_D)$ in the following way:

$$Q_X(z) = Q_X^1(z)\mathbb{1}_{z_D < z \leq 1} + Q_X^2(z)\mathbb{1}_{0 \leq z \leq z_D}, \quad Q_X(z) \in \mathcal{Q}(z_D), \tag{13}$$

where $Q_X^1(\cdot)$ and $Q_X^2(\cdot)$ are both non-decreasing functions. We have that $Q_X^1(z_D) \geq D_T \geq Q_X^2(z_D)$. Writing the quantile function in this way enables us to decompose the original problem (12) into the following three sub problems:
P1:

$$\max_{Q_X(z) \in \mathcal{Q}_D} \int_{z_D}^1 U(Q_X(z) - D_T)\mathrm{d}z, \text{ subject to } \int_{z_D}^1 Q_X(z)\widetilde{F}(z)\mathrm{d}z = x_0^+.$$

P2:

$$\min_{Q_X(z) \in \mathcal{Q}_D} \int_0^{z_D} Q_X(z)\widetilde{F}(z)\mathrm{d}z = x_0 - x_0^+.$$

P3:

$$\max_{z_D \in [0,1]} \max_{Q_X(z) \in \mathcal{Q}_D} \int_{z_D}^1 U(Q_X(z) - D_T)\mathrm{d}z, \text{ subject to } \int_0^1 Q_X(z)\widetilde{F}(z)\mathrm{d}z = x_0.$$

Thus, the original non-concave optimization problem (12) has been decomposed into a concave maximization problem on $(z_D, 1]$ (***P1***), and a minimization problem on $[0, z_D]$ (***P2***). In ***P3***, we determine the global optimal solution by finding the optimal $z_D \in [0,1]$, which is a key component in the quantile function (13).

Remark 1:

The decomposition method is discussed in [25] and [17], where the terminal wealth is decomposed to two random variables, X and Y, such that the utility function is strictly concave in X, and convex in Y. Our idea of decomposition is similar. While [25] and [17] focus on the terminal wealth, we focus on decomposing the quantile function into two non-decreasing functions on disjoint sets.

Since ***P1*** is concave maximization, it can be solved by the standard Lagrangian approach. ***P2*** is trivial that $Q_X^2(z) = 0, z \in [0, z_D]$ minimizes the cost on $[0, z_D]$. Thus, given a $z_D \in [0,1]$, the argmax to ***P1*** and ***P2*** is given by

$$Q_X^*(z_D) = (I(\lambda_B \widetilde{F}(z)) + D_T)\mathbb{1}_{z_D < z \leq 1}. \tag{14}$$

Recall that $\widetilde{F}(z) = F_\xi^{-1}(1-z),\ z \in [0,1]$. In ***P3***, we determine the optimal z_D to obtain the global argmax. Because of (14), ***P3*** reduces to a concave maximization problem

$$\max_{z_D \in [0,1]} \int_{z_D}^1 U(Q_X^*(z_D) - D_T)\mathrm{d}z, \text{ subject to } \int_{z_D}^1 Q_X^*(z_D)\widetilde{F}(z)\mathrm{d}z = x_0.$$

Given a $\lambda > 0$, the Lagrangian of ***P3*** is given by the following,

$$G(z_D) = \int_{z_D}^1 U(Q_X^*(z_D) - D_T)\mathrm{d}z - \lambda \int_{z_D}^1 Q_X^*(z_D)\widetilde{F}(z)\mathrm{d}z. \tag{15}$$

The first order condition is given by

$$\begin{aligned} G'(z_D) &= -U(Q_X^*(z_D) - D_T) + \lambda Q_X^*(z_D)\widetilde{F}(z_D) \\ &= -U(I(\lambda \widetilde{F}(z_D)) + \lambda \widetilde{F}(z_D) I(\lambda \widetilde{F}(z_D)) + \lambda \widetilde{F}(z_D) D_T = 0. \end{aligned} \tag{16}$$

The following lemma is useful to discuss (16). The proof of Lemma 3.1 is given in Appendix A.

Lemma 3.1 *The utility function satisfying the Inada* (2) *and AE* (3) *conditions has the following features:*

1. *Given a constant $d > 0$, the function, $H(x) = U(x-d) - U'(x-d)x,\ x > d$, has a unique zero root in x denoted by x^*. Moreover, the zero root is larger than the constant d, i.e., $x^* > d > 0$.*
2. *The function $H(x)$ is increasing in x.*

Let $\lambda > 0$ and $\widehat{D}_T$ satisfy $U(\widehat{D}_T - D_T) - U'(\widehat{D}_T - D_T)\widehat{D}_T = 0$. Thus, $G'(z_D^*) = 0$ with $z_D^* = 1 - F\left(\frac{U'(\widehat{D}_T - D_T)}{\lambda}\right)$.

Lemma 3.1 gives that $G'(z_D^*) = 0$ (16) with $z_D^* = 1 - F\left(U'(\widehat{D}_T - D_T)/\lambda\right)$. Thus, the argmax to ***P3*** is given by

$$Q_X^*(z) = (I(\lambda \widetilde{F}(z)) + D_T)\mathbb{1}_{z_D^* < z \leq 1}. \tag{17}$$

Lemma A.1 shows that the Lagrangian multiplier λ in (17) exists. The proof of Lemma A.1 is given in Appendix A. By Equation (7), the optimal terminal wealth (15) is given by

$$X_T^B = (I(\lambda_B \xi) + D_T)\mathbb{1}_{\xi < \xi_D^*},\ \xi_D^* = \widetilde{F}(z_D^*).$$

Then, it can be shown that X_T^B is the optimal benchmark solution, following a routine argument. See, for example, Appendix B.1 in Chen, Stadje, and Zhang [14].

We have illustrated how to solve the benchmark problem (8) by quantile formulation and the decomposition method. Following a similar procedure, we solve the Problem AVaR (1). Before moving to the optimal solution under the AVaR constraint, we give a remark on the benchmark solution.

Remark 2

1. Lemma 3.1 provides a way to determine the optimal decomposition (the critical probability z_D) in the benchmark problem. The other methods, e.g., the concavification [10] and the pointwise Lagrangian [8, 14], apply a similar idea to find the *tangent point* for the utility function, which is the key step in deriving the optimal solution. Applying quantile formulation and the decomposition method, the critical probability z_D^* is obtained via the first-order condition, which gives a different perspective than the tangent point in the concavification.

2. The benchmark solution gives that the optimal terminal wealth will end with zero in the worst financial scenarios given a bounded utility function. The intuition is that if the punishment for inducing the minimum wealth in the worst scenarios is not large enough, the surplus-driven company will *default* in the worst financial scenarios. Assuming a classic concave utility maximization problem cannot model this behavior. We illustrate the optimal solution for the benchmark problem in Figure 1. We use the red dashed line to denote the left tail of the portfolio with $\alpha = 0.1$, meaning that the regulator is concerned about the company's performance in the worst 10% scenarios. In our toy example, the probability of default is much higher than 10%, justifying the necessity for financial regulation.

3.4 The optimal solution for the Problem AVaR (1)

Now we consider the optimization problem with the AVaR constraint. We first formulate the equivalent quantile-based optimization problem to the Problem

AVaR (1). Then, we apply the decomposition method to find the optimal quantile function. In the end, we recover the optimal terminal wealth via Equation (7).

Quantile-based Problem AVaR

The equivalent quantile-based optimization problem for the Problem AVaR(1) is given by

$$\begin{aligned} &\max_{Q_X\in\mathcal{Q}} \int_0^1 U((Q_X(z)-D_T)^+)\mathrm{d}z, \\ &\text{subject to} \quad \frac{1}{\alpha}\int_0^\alpha Q_X(z)\mathrm{d}z \ge L,\ \int_0^1 Q_X(z)\widetilde{F}(z)\mathrm{d}z = x_0. \end{aligned} \tag{18}$$

Following the same idea as we did in solving the benchmark problem, we decompose the set of quantile functions in the following way:

$$\mathcal{Q}(z_D) := \{Q_X(z)\in\mathcal{Q}|z(D_T)\le z_D\},\ \mathcal{Q} = \bigcup_{z_D\in[0,1]}\mathcal{Q}(z_D).$$

Then, the quantile-based Problem AVaR (18) can be decomposed into the following three sub optimization problems:
P1:

$$\begin{aligned} &\max_{Q_X(z)\in\mathcal{Q}_D} \int_{z_D}^1 U(Q_X(z)-D_T)\mathrm{d}z, \\ &\text{subject to } \frac{1}{\alpha}\int_0^\alpha Q_X(z)\mathrm{d}z \ge L,\ \int_{z_D}^1 Q_X(z)\widetilde{F}(z)\mathrm{d}z = x_0^+. \end{aligned}$$

P2:

$$\min_{Q_X(z)\in\mathcal{Q}_D} \int_0^{z_D} Q_X(z)\widetilde{F}(z)\mathrm{d}z = x_0 - x_0^+,\ \frac{1}{\alpha}\int_0^\alpha Q_X(z)\mathrm{d}z \ge L.$$

P3:

$$\begin{aligned} &\max_{z_D\in[0,1]}\max_{Q_X(z)\in\mathcal{Q}_D} \int_{z_D}^1 U(Q_X(z)-D_T)\mathrm{d}z, \\ &\qquad\text{subject to } \frac{1}{\alpha}\int_0^\alpha Q_X(z)\mathrm{d}z \ge L,\ \int_0^1 Q_X(z)\widetilde{F}(z)\mathrm{d}z = x_0. \end{aligned}$$

Note that if $0<\alpha\le z_D$, we only need to consider the AVaR constraint in ***P2***, which is a relatively simple case. If $\alpha > z_D$, we need to consider the AVaR constraint in both ***P1*** and ***P2***, which is a more complicated case. Before proceeding with the optimization problem under the AVaR constraint, we define the effective risk constraint.

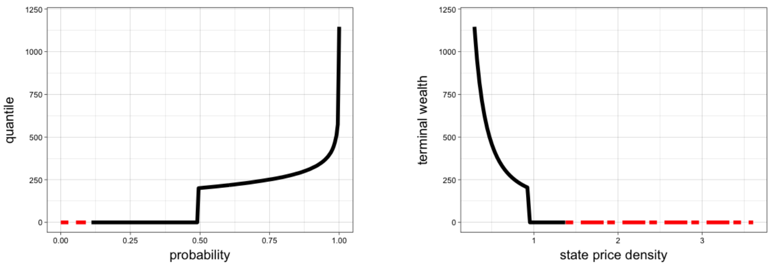

Fig. 1: This figure illustrates the optimal quantile function and the corresponding optimal terminal wealth for the benchmark problems (12) and (8), respectively. The parameters for the underlying financial market and the utility function are consistent with the numerical experiment in Section 5 (Table 1). The left graph plots the optimal quantile function (17), where the red dashed line denotes the AVaR of the quantile function with $\alpha = 0.1$. Correspondingly, the right graph plots the optimal terminal wealth (11), where the red dashed line denotes the worst scenario payoff with probability 0.1.

Effective risk constraint

If the optimal solution for the Problem AVaR (1) is *different* from the benchmark solution given the same initial wealth, we say the AVaR constraint is effective. If the optimal solution for the Problem AVaR (1) is the *same* as the benchmark solution, we say the AVaR constraint is redundant.

The benchmark solution (Proposition 1) tells that the surplus-driven company will inevitably default in the worst economic states. However, assuming a deterministic debt level D_T, the probability of default varies with the initial budget. We illustrate this fact in Figure 2. We can see that the default probability increases as the initial budget decreases. Hence, given the AVaR constraint with fixed parameters α and L, we can imagine that there will be multiple solutions depending on the initial wealth.

Proposition 2 provides the optimal solution for the Problem AVaR (1). The proof for Proposition 2 is given in Appendix B.

Proposition 2 *The optimal solutions can be divided into the following cases.*
Case 1:
The threshold in the AVaR constraint is smaller than the debt level: $0 < L < D_T$.

1. *Let* λ_1 *satisfy* $\mathbb{E}[\xi X_T^{AVaR}] = x_0$ *and* $\widehat{L}^* = \frac{\alpha L}{\alpha-\beta^*}$. *If* $L < \widehat{L}^* < D_T$, *the optimal solution for the Problem AVaR (1) is given by*

$$X_T^{AVaR} = (I(\lambda_1\xi) + D_T)\mathbb{1}_{\xi<\xi_1} + \widehat{L}^*\mathbb{1}_{\xi_1\leq\xi<\xi_2}, \tag{19}$$

where $\xi_1 = \widetilde{F}(z_D^*)$, $\xi_2 = \widetilde{F}(\beta^*)$, z_D^* *and* β^* *satisfy the following two equations:*

$$-U(I(\lambda_1\widetilde{F}(z_D^*)) + \lambda_1\widetilde{F}(z_D^*)I(\lambda_1\widetilde{F}(z_D^*) + \lambda_1\widetilde{F}(z_D^*)\left(D_T - \frac{\alpha L}{\alpha-\beta^*}\right) = 0, \tag{20}$$

$$\int_{\beta^*}^{z_D^*} \widetilde{F}(z)\mathrm{d}z - (\alpha-\beta^*)\widetilde{F}(\beta^*) = 0. \tag{21}$$

(The existence of β^* *is shown in Lemma B.1 and the existence of* z_D^* *is given by Eq (B9) and Lemma 3.1.)*

2. *Let* $\widehat{L}^*$ *satisfy* $\mathbb{G}'(\widehat{L}^*) = 0$ (B14). *(The existence of* $\widehat{L}^*$ *is shown by Lemma B.3). If* $D_T < \widehat{L}^* < \widehat{D}_T$, *the optimal solution for the problem AVaR (1) is given by*

$$X_T^{AVaR} = (I(\lambda_1\xi) + D_T)\mathbb{1}_{\xi<\xi_1} + \widehat{L}^*\mathbb{1}_{\xi_1\leq\xi<\xi_2}, \tag{22}$$

where $\xi_1 = \frac{U'(\widehat{L}^*-D_T)}{\lambda_1}$, *and* $\xi_2 = \widetilde{F}\left(\frac{\alpha(\widehat{L}^*-L)}{\widehat{L}^*}\right)$.

3. *Let* λ_2 *satisfy* $\frac{1}{\alpha}\int_0^\alpha Q_X^{AVaR}(z)\mathrm{d}z = L$. *If* $\widehat{L}^* > \widehat{D}_T$, *the optimal solution for the problem AVaR (1) is given by*

$$X_T^{AVaR} = \begin{cases} I(\lambda_1\xi) + D_T, & \text{if} \quad \xi < \xi_1, \\ \widehat{L}^*, & \text{if} \quad \xi_1 < \xi < \xi_2, \\ I\left(\lambda_1\xi - \frac{\lambda_1\lambda_2}{\alpha}\right) + D_T, & \text{if} \quad \xi_2 < \xi < \xi_3, \\ 0, & \text{if} \quad \xi > \xi_3, \end{cases} \tag{23}$$

where $\xi_1 = \frac{U'(\widehat{L}^*-D_T)}{\lambda_1}$, $\xi_2 = \frac{U'(\widehat{L}^*-D_T)}{\lambda_1} + \frac{\lambda_2}{\alpha}$, $\xi_3 = \frac{U'(\widehat{D}_T-D_T)}{\lambda_1} + \frac{\lambda_2}{\alpha}$, *and* $Q_X^{AVaR}(\cdot)$ *is the quantile function of the optimal wealth* X_T^{AVaR}.

Case 2:
The threshold in the AVaR constraint is higher than the debt level: $L \geq D_T$. *Let* $\widehat{L}^*$ *satisfy* $\mathbb{G}'(\widehat{L}^*) = 0$ (B14).

1. *If* $D_T < \widehat{L}^* < \widehat{D}_T$, *the optimal solution for the problem AVaR (1) is the same as (22).*
2. *If* $\widehat{L}^* > \widehat{D}_T$, *the optimal solution for the problem AVaR (1) is the same as (23).*

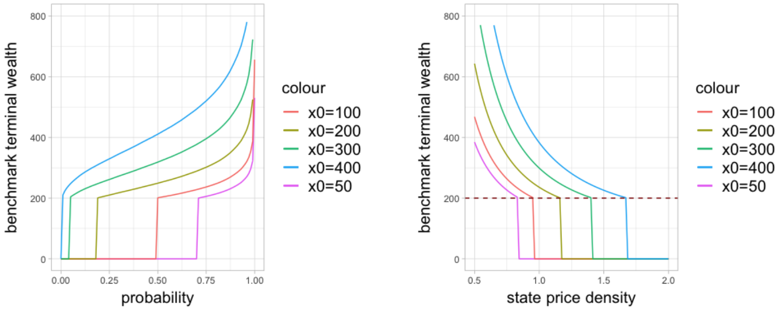

Fig. 2: This figure illustrates the optimal quantile function and the optimal terminal wealth for Proposition 1 with the same terminal debt D_T and different initial wealth. We can see the default probability varies with the initial budget. In addition, the higher the initial budget, the smaller the default probability. Note that the benchmark wealth is either higher than the tangent point $\widehat{D}_T$ (Lemma 3.1), or is zero.

We use a numerical example to illustrate the different structures of the optimal solutions for the Problem AVaR (1) (Proposition 2) in the Black Scholes market with the parameters in Table 1. Suppose that the AVaR constraint is given by $\frac{1}{0.1}\int_0^{0.1} VaR_X(\beta)\mathrm{d}\beta \geq 40\% D_T$. Its economic interpretation is that the average quantile value of the company's portfolio in the worst 10% scenarios has to be at least 40% of its debt. The surplus-driven company will adopt different strategies according to its initial wealth. We consider three values of the initial budget: $x_0 = 132$, $x_0 = 154$, and $x_0 = 218$. Let the terminal debt level $D_T = 100$ be fixed. The three optimal wealth correspond to Eq (19), (22), and (23), respectively. In comparison, we plot the benchmark wealth (i.e., without the AVaR constraint) under the corresponding initial budget in Figure 3. Each line in Figure 3 jumps to zero if the state price density ξ_T is large enough. This implies that the surplus-driven company will default in the worst financial states. Nevertheless, under each initial budget, the default probability (the length of the flat region) of the wealth with the AVaR constraint is smaller than the benchmark wealth, demonstrating the effectiveness of the regulation. However, the default probability is not zero in any case.

We give a remark for Proposition 2.

Remark 3

Given the AVaR constraint, the regulatory effect depends on the company's initial budget.[10] If the initial budget is large enough, the AVaR constraint is redundant. On the other hand, if the initial wealth is too low ($x_0 < Le^{-rT}$), the optimal solution does not exist. In Proposition 2, we use $\widehat{L}^*$ to distinguish different cases of the optimal solutions. We remark that there is a one-to-one relationship between $\widehat{L}^*$ and the Lagrangian multiplier λ_1, which

[10] It is also possible to fix the initial budget and analyse the regulatory effect with varying α.

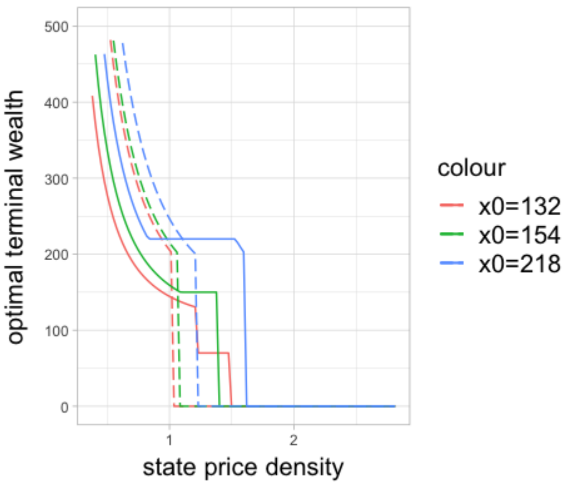

Fig. 3: This figure illustrates Propositions 1 and 2 by plotting the optimal terminal wealth with different initial budgets. The solid lines are the wealth under the AVaR constraint (Proposition 2), while the dashed lines are the corresponding benchmark wealth (Proposition 1). The red line corresponds to Eq (19) with $\widehat{L}^* = 70 < D_T$, the green line corresponds to Eq (22) with $D_T < \widehat{L}^* = 120 < \widehat{D}_T$ and the blue line corresponds to Eq (23) with $\widehat{L}^* = 220 > \widehat{D}_T$.

is uniquely determined by the initial budget x_0 through the budget constraint $\mathbb{E}[\xi X_T^{AVaR}] = x_0$ (Lemma A.1).[11] Hence, all possible cases are covered by Proposition 2. We explain how to determine the range of the initial wealth for each solution in Appendix B.2.

4 The connection between the terminal-wealth method and the quantile-formulation method

The analysis in Chen, Stadje, and Zhang [14] for the non-concave optimization problem under the AVaR constraint applies the terminal-wealth method. They establish equivalent Expected Shortfall (ES) constraints to the AVaR constraint on the physical probability space, and solve the constrained optimization problem on the physical probability space. To compare the terminal-wealth method and our quantile-formulation method, we briefly summarize the optimal solution for the non-concave optimization problem under the equivalent ES constraint. We first define the ES constraint.

[11] For a comprehensive discussion of the budget constraint as a function of the Lagrangian multiplier, we recommend [24].

Definition of the Expected Shortfall constraint

Given a regulatory threshold $L^{ES} > 0$ and $\epsilon > 0$, the ES constraint of a portfolio is defined by

$$\text{ES} := \mathbb{E}[(L^{ES} - X_T)^+] \leq \epsilon, \quad (L^{ES} - X_T)^+ := \max(L^{ES} - X_T, 0). \quad (24)$$

The ES constraint restricts the average loss measured by $(L^{ES} - X_T)$ of the portfolio in the worst financial scenarios to be below the constant ϵ. Different from the AVaR constraint, the worst scenarios under the ES constraint are not described by the probability but by the given threshold L^{ES}, i.e., the ES constraint concerns the loss of the portfolio when the portfolio is below L^{ES} (see the example in Figure 4). Thus, if the threshold L^{ES} equals to the α-quantile of the portfolio, the ES and the AVaR constraints agree on the worst scenarios. Consequently, the two constraints are equivalent. For a formal discussion on the equivalence between the AVaR constraint and the ES constraint, we refer to Theorems 1 and 2 in Rockafellar and Uryasev [31] and Proposition 4.51 in Föllmer and Schied [21]. For the equivalence of the optimization problems on the physical probability space between the two constraints, we refer to Lemma 2.7 and Theorem 2.9 in Gandy [22] and Proposition 6 in Chen, Stadje, and Zhang [14].

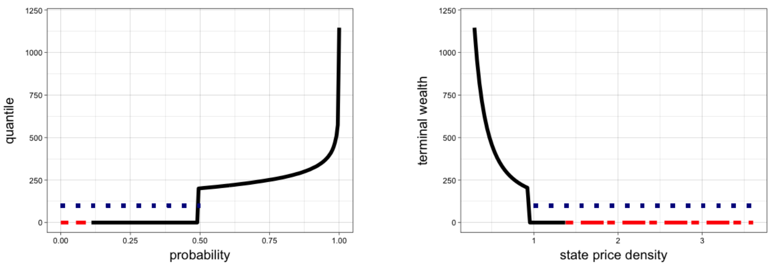

Fig. 4: This figure illustrates the worst financial scenarios defined by the terminal-wealth-based ES constraint. The black curves denote the optimal quantile function and the optimal terminal wealth for the benchmark problem (Figure 1). We consider a toy example where the dark blue dotted line denotes the given regulatory threshold in the ES constraint, e.g., $L^{ES} = 100$. Thus, the region $X_T^B < 100$ represents the financial scenarios concerned by the regulation. Compared to Figure 1, the ES constraint concerns a larger region than the AVaR constraint with $\alpha = 0.1$ in this example.

The optimal solution by the terminal wealth method

The non-concave optimization problem under the ES constraint is given by

Problem ES.

$$\max_{X_T \in \mathcal{X}} \mathbb{E}[U((X_T - D_T)^+)], \text{ s.t. } \mathbb{E}[(L^{ES} - X_T)^+] \leq \alpha(L^{ES} - L), \ \mathbb{E}[X_T \xi] \leq x_0, \tag{25}$$

where α and L are the parameters in the AVaR constraint (1).

The following proposition provides the optimal solution for the Problem ES (25). The proofs can be found in Appendix B in Chen, Stadje and Zhang [14].

Proposition 3 *The optimal solution for the Problem ES (25) is given by the following.*

1. *If $L^{ES} \leq D_T$, the optimal wealth is:*

$$X_T^{ES} = \left(I(\lambda_1^{ES}\xi) + D_T\right) \mathbb{1}_{\xi<\xi_b} \quad \text{if } \xi_b \leq \xi_a, \tag{26}$$
$$X_T^{ES} = \left(I(\lambda_1^{ES}\xi) + D_T\right) \mathbb{1}_{\xi<\xi_a} + L^{ES}\mathbb{1}_{\xi_a \leq \xi < \xi_b} \quad \text{if } \xi_a < \xi_b, \tag{27}$$

where $\xi_a = U'(L' - (D_T - L^{ES}))/\lambda_1^{ES}$, L' satisfies $U((L' - (D_T - L^{ES})) - L'U'(L' - (D_T - L^{ES})) = 0$, ξ_b is defined through $\mathbb{E}[L^{ES}\mathbb{1}_{\xi \geq \xi_b}] = \alpha(L^{ES} - L)$, and λ_1^{ES} satisfies $\mathbb{E}[\xi X_T^{ES}] = x_0$.

2. *If $D_T < L^{ES} \leq \widehat{D}_T$, the optimal wealth is:*

$$X_T^{ES} = \left(I(\lambda_1^{ES}\xi) + D_T\right) \mathbb{1}_{\xi<\xi_d} \quad \text{if } \xi_d < \xi_c, \tag{28}$$
$$X_T^{ES} = \left(I(\lambda_1^{ES}\xi) + D_T\right) \mathbb{1}_{\xi<\xi_c} + L^{ES}\mathbb{1}_{\xi_c \leq \xi < \xi_d} \quad \text{if } \xi_c \leq \xi_d, \tag{29}$$

where $\xi_c = U'(L^{ES} - D_T)/\lambda_1^{ES}$ and ξ_d is defined through $\mathbb{E}[L^{ES}\mathbb{1}_{\xi \geq \xi_d}] = \alpha(L^{ES} - L)$, and λ_1^{ES} satisfies $\mathbb{E}[\xi X_T^{ES}] = x_0$.

3. *If $L^{ES} > \widehat{D}_T$, the optimal wealth is:*

$$X_T^{ES} = \begin{cases} \left(I(\lambda_1^{ES}\xi) + D_T\right), & \text{if } \ \xi < \xi_e \\ L^{ES}, & \text{if } \ \xi_e \leq \xi < \xi_f, \\ \left(I(\lambda_1^{ES}\xi - \lambda_2^{ES}) + D_T\right), & \text{if } \ \xi_f \leq \xi < \xi_g, \\ 0, & \text{if } \ \xi > \xi_g, \end{cases} \tag{30}$$

where $\xi_e = U'(L^{ES} - D_T)/\lambda_1^{ES}$, $\xi_f = U'(L^{ES} - D_T) + \lambda_2^{ES}/\lambda_1^{ES}$, $\xi_g = U'(\widehat{D}_T - D_T) + \lambda_2^{ES}/\lambda_1^{ES}$, λ_1^{ES} and λ_2^{ES} satisfy $\mathbb{E}[\xi X_T^{ES}] = x_0$ and $\mathbb{E}[(L^{ES} - X_T^{ES})^+] = \alpha(L^{ES} - L)$.

For an arbitrary $L^{ES} > L$, the optimal wealth in Proposition 3 satisfies the AVaR constraint, and thus is a feasible solution for the Problem AVaR (1). Proposition 6 in Chen, Stadje, and Zhang [14] shows that the union of the optimal terminal wealth for the Problems ES (25) with all such L^{ES} contains the optimal solution for the Problem AVaR (1). However, it is difficult to further determine the exact optimal solution for the Problem AVaR (1) by the terminal-wealth method. In Proposition 4, we give the connection between the optimization problems under the AVaR and the ES constraints. The proof is given in Appendix C.

Proposition 4 *We make the following statements.*

1. *Given an AVaR constraint ($\frac{1}{\alpha}\int_0^\alpha VaR_{X_T}(\beta)\mathrm{d}\beta \geq L$), if the regulatory threshold in the ES constraint ($\mathbb{E}[(L^{ES} - X_T)^+] \leq \alpha(L^{ES} - L)$) is equal to $\widehat{L}^*$ defined in (19), (23) or (22), i.e., $L^{ES} = \widehat{L}^*$, then the optimal solution under the ES constraint is the **same** as the optimal solution under the AVaR constraint.*

2. *Consider the optimal wealth satisfying the ES constraint, $\mathbb{E}[(L^{ES} - X_T^{ES})^+] = \alpha(L^{ES} - L)$. If the optimal terminal wealth is (27), (29) or (30), the AVaR constraint is binding, i.e., $\frac{1}{\alpha}\int_0^\alpha VaR_{X_T^{ES}}(\beta)\mathrm{d}\beta = L$. If the optimal terminal wealth is (26) or (28), the AVaR constraint is not binding, i.e., $\frac{1}{\alpha}\int_0^\alpha VaR_{X_T^{ES}}(\beta)\mathrm{d}\beta > L$.*

We give a remark for Proposition 4.

Remark 4

Proposition 4 reveals the connection between the quantile formulation method and the terminal wealth method in solving the Problem AVaR (1). Given the optimal solution in Proposition 2 by quantile formulation, we can directly calculate the equivalent ES constraint. However, given the optimal solution in Proposition 3 by the terminal wealth method, we can calculate the *lower bound* for the AVaR of the portfolio. Nevertheless, by comparing Propositions 2 and 3, we find that there are many cases where the optimal solutions are *equivalent*. Moreover, the Problem ES (25) is much easier to solve.

5 Numerical examples and the fairness of the contract

In this section, we illustrate the optimal solution (Proposition 2) for the Problem AVaR (1) in a Black Scholes financial market. Previously, the underlying financial market is assumed to be *complete* and *atomless* to obtain the closed form solution. A Black Scholes market is an example of such a financial market, where we can explicitly calculate the optimal terminal wealth and the corresponding optimal investment strategies.

For simplicity, we assume that there is a risky asset $S(t)$ and a risk-free asset $B(t)$ in the financial market and the dynamics are given by

$$\mathrm{d}S(t) = \mu S(t)\mathrm{d}t + \sigma S(t)\mathrm{d}W_t, \quad S(0) = s_0, \quad \mathrm{d}B(t) = rB(t)\mathrm{d}t, \quad B(0) = b_0,$$

where $\mu > 0$ is the drift, and $\sigma > 0$ is the volatility of the risky asset. The risk-free rate is given by $r > 0$. Note that the market price of risk in the Black Scholes market is given by $\theta := (\mu - r)/\sigma$. Moreover, we assume that the company has a power utility function defined by $U(x) = \frac{x^{1-\gamma}}{1-\gamma}, \gamma \neq 1$. The basic parameters are given in Table 1.

Table 1: Basic Parameters

drift μ	risk-free rate r	volatility σ	horizon T	risk aversion γ
0.08	0.03	0.2	1	0.5

In section 5.1, we show the pre-horizon optimal wealth and the corresponding investment strategies for the Problem AVaR (1) in the Black Scholes market. In section 5.2, we discuss the fair contract for the debt holders considering the default probability of the surplus-driven company.

5.1 Pre-horizon optimal wealth

Let $T = 1$ be the terminal time of the investment. We consider the optimal wealth at $t: \ 0 < t < T$. We also compute the proportion of the optimal wealth invested into the risky asset in the Black Scholes market. The results are given in Proposition 5. The proof is given in Appendix D.

Proposition 5 *For the benchmark portfolio (Proposition 1):*

1. *Let* $k(t) := \exp(-(r + 0.5\theta^2)(T - t)(1 - 1/\gamma) + 0.5\theta^2(1 - 1/\gamma)^2(T - t))$. *Let* ξ_t *denote the state price density at time* t, λ_B *and* ξ_D^*, *be defined in Proposition 1. The optimal benchmark wealth at time* t *is given by*

$$\begin{aligned} X_t^B =&(\lambda_B \xi_t)^{-1/\gamma} k(t) \Phi(j(\xi_D^*/\xi_t) + 1/\gamma\theta\sqrt{T-t}) \\ &+ D_T \exp(-r(T-t))\Phi(j(\xi_D^*/\xi_t)), \end{aligned} \tag{31}$$

where $j(\cdot) = \frac{\ln(\cdot)+(r+0.5\theta^2)(T-t)}{\theta\sqrt{T-t}} - \theta\sqrt{T-t}$, *and* $\Phi(\cdot)$ *is the cumulative distribution function of the standard normal distribution.*

2. *Given the optimal wealth* X_t^B, *the company invests* $\pi_t^B X_t^B$ *in the stock and* $(1 - \pi_t^B)X_t^B$ *into the bank account. The fraction of the benchmark portfolio*

invested into the risky asset is given by

$$\pi_t^B = \frac{\theta}{X_t^B \sigma\gamma}(\lambda_B \xi_t)^{-1/\gamma} k(t)\Phi(j(\xi_D^*/\xi_t) + \frac{\exp(-r(T-t))}{X_t^B}\frac{\widehat{D}_T \Phi'(j(\xi_D^*/\xi_t))}{\sigma\sqrt{T-t}}, \tag{32}$$

where $\Phi'(\cdot)$ *is the probability density function of the standard normal distribution.*

For the optimal AVaR portfolio (Proposition 2),

1. *If the solution* X_T^{AVaR} *is Eq (19), let* λ_1 *satisfy* $\mathbb{E}[\xi X_T^{AVaR}] = x_0$*. Let* $\xi_1 = \widetilde{F}(z_D^*)$ *and* $\xi_2 = \widetilde{F}(\beta^*)$ *(defined in Eq (20) and (21)). The optimal wealth at time* $t < T$ *is given by*

$$\begin{aligned} X_t^{AVaR} =& (\lambda_1 \xi_t)^{-1/\gamma} k(t)\Phi(j(\xi_1/\xi_t) + 1/\gamma\theta\sqrt{T-t}) \\ &+ D_T \exp(-r(T-t))\Phi(j(\xi_1/\xi_t)) \\ &+ \widehat{L}^* \exp(-r(T-t))(\Phi(j(\xi_2/\xi_t)) - \Phi(j(\xi_1/\xi_t))), \end{aligned} \tag{33}$$

where $\widehat{L}^* = \frac{\alpha L}{\alpha - \beta^*}$*.*
The optimal investment strategy at time $t < T$ *is given by*

$$\begin{aligned} \pi_t^{AVaR} =& \frac{k(t)}{\sigma X_t^{AVaR}}(\lambda_1 \xi_t)^{-1/\gamma}\left(\frac{\theta}{\gamma}\Phi\left(j(\xi_1/\xi_t) + 1/\gamma\theta\sqrt{T-t}\right)\right) \\ &+ \frac{(\lambda_1 \xi_1)^{-1/\gamma}}{\sigma X_t^{AVaR}} \exp(-r(T-t)) \frac{\Phi'\left(j(\xi_1/\xi_t)\right)}{\sqrt{T-t}} & (34) \\ &+ \frac{\widehat{L}^*}{\sigma X_t^{AVaR}} \exp(-r(T-t)) \frac{\Phi'\left(j(\xi_2/\xi_t)\right)}{\sqrt{T-t}}. & (35) \end{aligned}$$

2. *If the solution* X_T^{AVaR} *is Eq (22), let* $\widehat{L}^*$ *satisfy* $\mathbb{G}'(\widehat{L}^*) = 0$ *(B14),* $\xi_1 = U'(\widehat{L}^* - D_T)/\lambda_1$ *and* $\xi_2 = \widetilde{F}(\alpha(\widehat{L}^* - L)/\widehat{L}^*)$*. The optimal wealth at time* $t < T$ *is given by*

$$\begin{aligned} X_t^{AVaR} =& (\lambda_1 \xi_t)^{-1/\gamma} k(t)\Phi(j(\xi_1/\xi_t) + 1/\gamma\theta\sqrt{T-t}) \\ &+ D_T \exp(-r(T-t))\Phi(j(\xi_1/\xi_t)) \\ &+ \widehat{L}^* \exp(-r(T-t))(\Phi(j(\xi_2/\xi_t)) - \Phi(j(\xi_1/\xi_t))). \end{aligned} \tag{36}$$

The optimal investment strategy at $t < T$ *is given by*

$$\begin{aligned} \pi_t^{AVaR} =& \frac{k(t)}{\sigma X_t^{AVaR}}(\lambda_1 \xi_t)^{-1/\gamma}\left(\frac{\theta}{\gamma}\Phi\left(j(\xi_1/\xi_t) + 1/\gamma\theta\sqrt{T-t}\right)\right) \\ &+ \frac{\widehat{L}^*}{\sigma X_t^{AVaR}} \exp(-r(T-t)) \frac{\Phi'\left(j(\xi_2/\xi_t)\right)}{\sqrt{T-t}}. \end{aligned} \tag{37}$$

3. *If the solution X_T^{AVaR} is Eq (23) with Q_X^{AVaR} being its quantile function, let λ_2 satisfy $\frac{1}{\alpha}\int_0^\alpha Q_X^{AVaR}(z)\mathrm{d}z = L$. Let $\xi_1 = \frac{U'(\widehat{L}^* - D_T)}{\lambda_1}$, $\xi_2 = \frac{U'(\widehat{L}^* - D_T)}{\lambda_1} + \frac{\lambda_2}{\alpha}$ and $\xi_3 = \frac{U'(\widehat{D}_T - D_T)}{\lambda_1} + \frac{\lambda_2}{\alpha}$. The optimal wealth at $t < T$ is given by*

$$\begin{aligned} X_t^{AVaR} =& (\lambda_1\xi_t)^{-1/\gamma} k(t)\Phi(j(\xi_1/\xi_t) + 1/\gamma\theta\sqrt{T-t}) \\ &+ D_T \exp(-r(T-t))\Phi(j(\xi_1/\xi_t)) \\ &+ \widehat{L}^* \exp(-r(T-t))(\Phi(j(\xi_2/\xi_t)) - \Phi(j(\xi_1/\xi_t))) \\ &+ \exp(-r(T-t))\int_{j(\xi_2/\xi_t)}^{j(\xi_3/\xi_t)} (\lambda_1\xi_t m(z) - \lambda_1\lambda_2/\alpha)^{-1/\gamma}\Phi'(z)\mathrm{d}z \\ &+ D_T \exp(-r(T-t))(\Phi(j(\xi_3/\xi_t))) - \Phi(j(\xi_2/\xi_t)))), \end{aligned} \tag{38}$$

where $m(z) = \exp(-(r + 0.5\theta^2)(T-t) - \theta\sqrt{T-t}z)$. The optimal investment strategy at $t < T$ is given by

$$\begin{aligned} \pi_t^{AVaR} =& \frac{k(t)}{\sigma X_t^{AVaR}} (\lambda_1\xi_t)^{-1/\gamma} \frac{\theta}{\gamma} \Phi\left(j(\xi_1/\xi_t) + 1/\gamma\theta\sqrt{T-t}\right) \\ &+ \frac{\widehat{D}}{\sigma X_t^{AVaR}} \exp(-r(T-t)) \frac{\Phi'\left(j(\xi_3/\xi_t)\right)}{\sqrt{T-t}}. \end{aligned} \tag{39}$$

To illustrate Proposition 5, we plot the optimal wealth under the AVaR constraint at time $t = 0.5$ under the different initial budgets, and compare it with the corresponding benchmark wealth in Figure 5. We observe that if the economic states are good (with small state price densities ξ), the benchmark wealth is larger than the constrained wealth. If the economic states are not good, the constrained wealth is larger than the benchmark. However, both benchmark and constrained wealth tend to be zero in the worst economic states.

In Figure 6, we plot the proportion of the pre-horizon wealth invested in the risky asset, π_t^{AVaR} and π_t^B, and the ratio π_t^{AVaR}/π_t^B, under the different initial budgets, respectively. We can see that the proportions π_t^B and π_t^{AVaR} increases as the state price density increases, implying that the company takes riskier strategies as the financial states worsens. Unlike in the concave utility maximization, where the benchmark strategy π_t^* is a constant in all states (e.g., the Merton constant [28]), the benchmark strategy π_t^B in the non-concave utility maximization is a stochastic process. Moreover, the proportion π_t^{AVaR} under the AVaR constraint is a also a stochastic process. Hence, it is more difficult to compare the two strategies explicitly. Nevertheless, our numerical examples indicate that the constrained strategy π_t^{AVaR} is less risky than the benchmark strategy especially if the initial budget is not too large. Moreover, Figure 3 shows that the default probability of the wealth under the AVaR constraint is always smaller than the corresponding benchmark case. We conclude that the AVaR constraint is *effective* in reducing the surplus-driven company's investment risk. This contrasts with the conclusion in the standard concave utility maximization, where the constrained strategies by the AVaR constraint is even riskier than the benchmark strategy [14, 36].

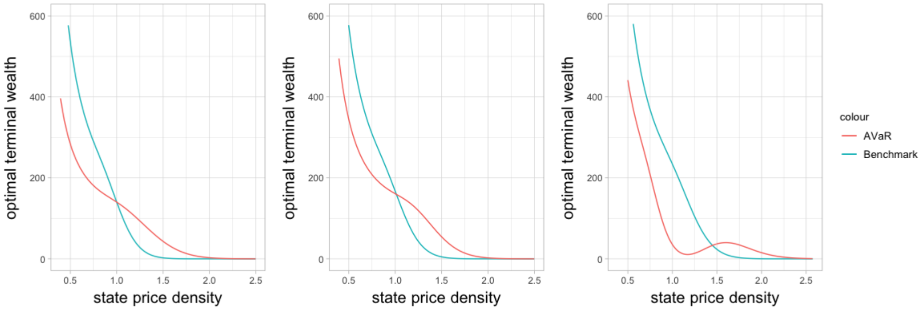

Fig. 5: This figure plots the optimal wealth at time $t = 0.5$ with the AVaR constraint (the red line), and without the AVaR constraint (the blue line), respectively. The parameters are the same as the numerical examples in Figure 3 and Table 1. The left figure illustrates Eq (33). The middle figure illustrates the case Eq (36). The right figure denotes the case Eq (38).

5.2 The fair return for the debt holders of a surplus-driven company

The numerical examples in previous sections indicate that given the AVaR constraint, the company will choose different investment strategies depending on its initial budget. Different strategies imply different default probability; hence, the debt holders are exposed to various levels of risk. In line with [11], we introduce the following definition of a fair contract.

Definition of a fair contract.

Let $\mathbb{Q}$ be the risk neutral measure and $r > 0$ be the risk free return. If the payoff to the debt holders, $\varphi_L(X_T)$, fulfills the equation

$$\mathbb{E}_{\mathbb{Q}}[e^{-rT}\varphi_L(X_T)] = D_0, \tag{40}$$

then the contract is regarded as a fair contract.

The equation (40) is reasonable for a contingent payoff in a complete financial market. The initial debt D_0 can be considered as the risk-neutral price of the contingent payoff D_T for the debt holders. Figure 7 plots the payoff to the debt holders as a consequence of the optimal solutions for the company with the AVaR constraint, under different initial budgets, respectively. The debt holders face the default risk in each case, but the probability of default differs. Intuitively, the return to the debt holders should be different in various

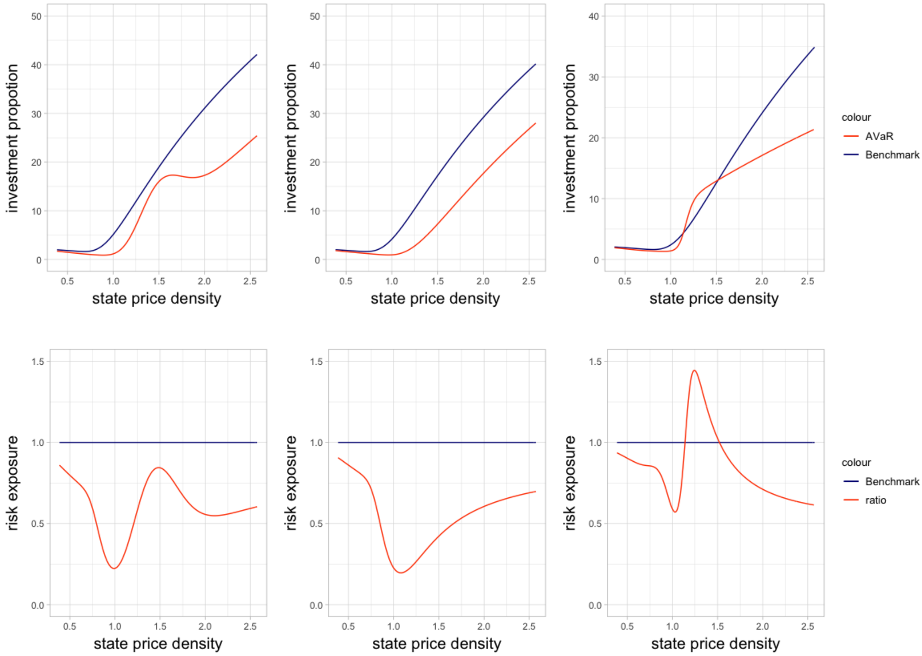

Fig. 6: We plot the proportion of the pre-horizon wealth in the risky asset (the upper three figures), π_t^{AVaR} and π_t^B, and the ratio (the lower three figures), π_t^{AVaR}/π_t^B, under the different initial budgets, respectively. The red line denotes π_t^{AVaR} and the blue line denotes π_t^B. From the left to the right, the three figures correspond to the case $\widehat{L}^* < D_T$ (35), $D_T < \widehat{L}^* < \widehat{D}_T$ (37), and $\widehat{L}^* > \widehat{D}_T$ (39), respectively. The parameters are consistent with the numerical examples in Figure 3 and Table 1.

situations. The following proposition introduces how to determine the return to the debt holders in a fair contract.

Proposition 6 *Suppose that the return to the debt holders is constant, i.e., $D_T = D_0 \exp(-gT)$, where $g > r$ due to the no-arbitrage condition.*

1. *Let $\widehat{L}^* = \frac{\alpha L}{\alpha - \beta^*}$, $\xi_1 = \widetilde{F}(z_D^*)$ and $\xi_2 = \widetilde{F}(\beta^*)$ determined by Eqs (20) and (21). Suppose the company has the optimal terminal wealth (19). The fair return to its debt holders is given by*

$$g^* = -\ln\left(\mathbb{E}[\xi \mathbb{1}_{\xi<\xi_1}] + \frac{\widehat{L}^*}{D_T}\mathbb{E}[\xi \mathbb{1}_{\xi_1 \le \xi < \xi_2}]\right)/T. \tag{41}$$

2. *Let $\widehat{L}^*$ satisfy $\mathbb{G}'(\widehat{L}^*) = 0$ (B14) and $\xi_2 = \widetilde{F}\left(\frac{\alpha(\widehat{L}^* - L)}{\widehat{L}^*}\right)$. Suppose the company has the optimal terminal wealth (22). The fair return to its debt holders*

is given by

$$g^* = -\ln\left(\mathbb{E}[\xi \mathbb{1}_{\xi<\xi_2}]\right)/T. \tag{42}$$

3. *Suppose the company has the optimal terminal wealth* (23) *and* λ_1 *satisfy* $\mathbb{E}[\xi X_T^{AVaR}] = x_0$. *Let* λ_2 *satisfy* $\frac{1}{\alpha}\int_0^\alpha Q_X^{AVaR}(z)\mathrm{d}z = L$, *where* $Q_X^{AVaR}(\cdot)$ *is the quantile function of* X_T^{AVaR}, *and* $\xi_3 = \frac{U'(\widehat{D}_T - D_T)+\lambda_2/\alpha}{\lambda_1}$. *The fair return to its debt holders is given by*

$$g^* = -\ln\left(\mathbb{E}[\xi \mathbb{1}_{\xi<\xi_3}]\right)/T. \tag{43}$$

Proof The payoff function to the debt holders of a company with limited liability is given by $\varphi_L(X_T) = \min(D_T, X_T)$. Then, Proposition 6 is a direct application of the fair contract condition (40) on the optimal solution under the AVaR constraint (Proposition 2). □

To illustrate the fair return to the debt holders in different cases, we construct the numerical examples in the Black Scholes market. The results are reported in Table 2. In the numerical experiment, we fix the terminal debt level in each case, i.e., $D_T = 100$. There exist three potential optimal investment strategies depending on the initial wealth. If the company has a low initial budget, the fair return to the debt holders is almost 40%. As the company's initial budget increases, it changes its optimal investment strategies. Consequently, the default probability decreases, and the fair return to the debt holders also decreases. Our main observation is, although the AVaR constraint can reduce the default probability compared with the benchmark portfolio, it cannot eliminate it completely; See Figure 3. In addition, if the company has a low initial budget, the debt holders will require a very high return even if the company fulfills the AVaR-based regulation.

Table 2: Fair return in different cases.

$\alpha = 0.1$	D_T	D_0	g^*	D_0/x_0	$\widehat{L}^*$
Case 1:	100	68.95	0.37180	52.1%	70
Case 2:	100	86.25	0.1479	55.72%	120
Case 3:	100	95.14	0.0499	43.5%	220

This table provides the numerical illustrations for the fair return g^* to the debt holders in different cases. The parameters for the initial budget constraint, the AVaR constraint and the Black Scholes market are consistent with the numerical examples in Figure 3 and Table 1. The fair return to the debt holders in each optimal solution is calculated according to Proposition 6. Case 1 is the case $\widehat{L}^* < D_T$ (41), Case 2 is $D_T < \widehat{L}^* < \widehat{D}_T$ (42), and Case 3 is $\widehat{L}^* > \widehat{D}_T$ (43).

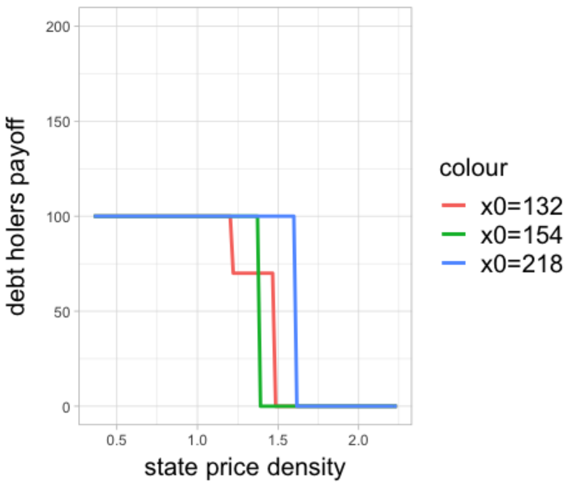

Fig. 7: This figure plots the payoff to the debt holders under the same AVaR constraint with different initial budget constraints. The parameters are consistent with the numerical examples in Figure 3 and Table 1. The red line denotes the case with a low initial wealth. The green line denotes the case with a intermediate level of initial wealth. The blue line denotes the case with a high initial wealth. We can see that the default probability varies in different cases.

6 Conclusion

This paper studies the non-concave portfolio optimization problem under the AVaR constraint, where the non-concavity arises from assuming that the company is surplus-driven. We solve the optimization problem explicitly by quantile formulation and the decomposition method. There are three optimal solutions for the optimization problem, implying that the company takes three different portfolio choices depending on the initial budget constraint. We provide numerical illustrations of the pre-horizon wealth and the optimal investment strategies in the Black Scholes financial market. In addition, we derive the fair return for the company's debt holders under each optimal solution fulfilling the risk-neutral pricing constraint.

Our contributions are twofold. First, we provide analytical solutions for non-concave portfolio optimization problems under the AVaR constraint, which is essential for researchers and practitioners to understand the impact of the AVaR-based financial regulation on companies' investment behavior. Second, we investigate whether and how the AVaR-based financial regulation can protect the debt holders' benefits in a surplus-driven financial company with limited liability. We find that the AVaR constraint can reduce the company's probability default but cannot fully illuminate it. Further, the numerical examples in a Black-Scholes market indicate that the the AVaR constraint provides poor protection for the debt holders if the company has a low initial budget. This finding gives one more degree of the comparison of VaR and AVaR as risk measures in financial regulation, where most comparisons are from the viewpoint of statistical properties.

There are multiple directions to proceed with this topic. Within the current model, one can study how to protect debt holders of the surplus-driven companies, especially if the company has a low initial budget. One can also calibrate the optimal solutions in this work to real data, and investigate the optimal regulatory parameters. With the technique in this paper, one can also study the AVaR constraint in other (non-concave) behavioral models, e.g., the cumulative prospect theory or rank-dependent utility maximization. Due to the limited space, we leave these topics for future study.

Acknowledgement. I would like to express my gratitude to the editor, the associate editor, and two anonymous reviewers, for their time and insightful comments, which helped to improve the paper significantly. I am also grateful to An Chen and Mitja Stadje for their helpful suggestions on an earlier version of the paper. All potential typos and errors are mine.

References

[1] Carlo Acerbi, Claudio Nordio, and Carlo Sirtori. Expected shortfall as a tool for financial risk management. *arXiv preprint cond-mat/0102304*, 2001.

[2] Carlo Acerbi and Dirk Tasche. On the coherence of expected shortfall. *Journal of Banking and Finance*, 26(7):1487–1503, 2002.

[3] John Armstrong and Damiano Brigo. Risk managing tail-risk seekers: Var and expected shortfall vs s-shaped utility. *Journal of Banking and Finance*, 101:122–135, 2019.

[4] Philippe Artzner, Freddy Delbaen, Jean-Marc Eber, and David Heath. Coherent measures of risk. *Mathematical Finance*, 9(3):203–228, 1999.

[5] Suleyman Basak, Anna Pavlova, and Alexander Shapiro. Optimal asset allocation and risk shifting in money management. *The Review of Financial Studies*, 20(5):1583–1621, 2007.

[6] Suleyman Basak and Alexander Shapiro. Value-at-risk-based risk management: optimal policies and asset prices. *The Review of Financial Studies*, 14(2):371–405, 2001.

[7] BCBS. *Minimum capital requirements for market risk: February 2019.* Basel Committee on Banking Supervision, Basel, 2019.

[8] Arjan B Berkelaar, Roy Kouwenberg, and Thierry Post. Optimal portfolio choice under loss aversion. *Review of Economics and Statistics*, 86(4):973–987, 2004.

[9] Carole Bernard, Xuedong He, Jia-An Yan, and Xun Yu Zhou. Optimal insurance design under rank-dependent expected utility. *Mathematical Finance*, 25(1):154–186, 2015.

[10] Jennifer N Carpenter. Does option compensation increase managerial risk appetite? *The Journal of Finance*, 55(5):2311–2331, 2000.

[11] An Chen, Peter Hieber, and Thai Nguyen. Constrained non-concave utility maximization: An application to life insurance contracts with guarantees. *European Journal of Operational Research*, 273(3):1119–1135, 2019.

[12] An Chen, Thai Nguyen, and Mitja Stadje. Optimal investment under VaR-regulation and minimum insurance. *Insurance: Mathematics and Economics*, 79:194–209, 2018.

[13] An Chen, Thai Nguyen, and Mitja Stadje. Risk management with multiple var constraints. *Mathematical Methods of Operations Research*, 88(2):297–337, 2018.

[14] An Chen, Mitja Stadje, and Fangyuan Zhang. On the equivalence between value-at-risk and expected shortfall in non-concave optimization. *Available at SSRN 3533948*, 2022.

[15] John C Cox and Chi-fu Huang. Optimal consumption and portfolio policies when asset prices follow a diffusion process. *Journal of Economic Theory*, 49(1):33–83, 1989.

[16] Domenico Cuoco, Hua He, and Sergei Isaenko. Optimal dynamic trading strategies with risk limits. *Operations Research*, 56(2):358–368, 2008.

[17] Carmine De Franco and Peter Tankov. Portfolio insurance under a risk-measure constraint. *Insurance: Mathematics and Economics*, 49(3):361–370, 2011.

[18] Yinghui Dong and Harry Zheng. Optimal investment with S-shaped utility and trading and Value at Risk constraints: An application to defined contribution pension plan. *European Journal of Operational Research*, 281(2):341–356, 2020.

[19] Paul Embrechts, Giovanni Puccetti, Ludger Rüschendorf, Ruodu Wang, and Antonela Beleraj. An academic response to basel 3.5. *Risks*, 2(1):25–48, 2014.

[20] Susanne Emmer, Dirk Tasche, and Marie Kratz. What is the best risk measure in practice? a comparison of standard measures. *Journal of Risk*, 18(2):31–60, 2015.

[21] Hans Föllmer and Alexander Schied. *Stochastic Finance*. de Gruyter, Boston, 2016.

[22] Ralf Gandy. *Portfolio optimization with risk constraints*. PhD thesis, Universität Ulm, 2005.

[23] Xue Dong He and Xun Yu Zhou. Portfolio choice via quantiles. *Mathematical Finance*, 21(2):203–231, 2011.

[24] Hanqing Jin, Zuo Quan Xu, and Xun Yu Zhou. A convex stochastic optimization problem arising from portfolio selection. *Mathematical Finance*, 18(1):171–183, 2008.

[25] Hanqing Jin and Xun Yu Zhou. Behavioral portfolio selection in continuous time. *Mathematical Finance*, 18(3):385–426, 2008.

[26] Ioannis Karatzas and Steven E Shreve. *Methods of Mathematical Finance*, volume 39. Springer, New York, 1998.

[27] Steven Kou and Xianhua Peng. On the measurement of economic tail risk. *Operations Research*, 64(5):1056–1072, 2016.

[28] Robert C Merton. Lifetime portfolio selection under uncertainty: The continuous-time case. *The Review of Economics and Statistics*, 51(1969):247–257, 1969.

[29] Thai Nguyen and Mitja Stadje. Nonconcave optimal investment with value-at-risk constraint: An application to life insurance contracts. *SIAM Journal on Control and Optimization*, 58(2):895–936, 2020.

[30] Christian Reichlin. Utility maximization with a given pricing measure when the utility is not necessarily concave. *Mathematics and Financial Economics*, 7(4):531–556, 2013.

[31] R Tyrrell Rockafellar and Stanislav Uryasev. Optimization of conditional value-at-risk. *Journal of Risk*, 2:21–42, 2000.

[32] Ludger Rüschendorf and Steven Vanduffel. On the construction of optimal payoffs. *Decisions in Economics and Finance*, 43(1):129–153, 2020.

[33] Amos Tversky and Daniel Kahneman. Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty*, 5(4):297–323, 1992.

[34] Ruodu Wang and Ričardas Zitikis. An axiomatic foundation for the expected shortfall. *Management Science*, 67(3):1413–1429, 2021.

[35] Pengyu Wei. Risk management with weighted var. *Mathematical Finance*, 28(4):1020–1060, 2018.

[36] Pengyu Wei. Risk management with expected shortfall. *Mathematics and Financial Economics*, pages 1–37, 2021.

[37] Zuo Quan Xu. A note on the quantile formulation. *Mathematical Finance*, 26(3):589–601, 2016.

Appendix A Proof of Lemma 3.1

Proof

Letting $y = x - d$, because of the Inada (2) and AE (3) conditions, we have that

$$H(x) = U(x-d) - U'(x-d)x = U(y)\left(1 - \frac{U'(y)}{U(y)}\right) - U'(y)d, \quad y > 0,$$
$$\longrightarrow \begin{cases} \lim_{y\to\infty} U(y)\left(1 - \frac{U'(y)}{U(y)}\right) - U'(y)d = \infty \\ \lim_{y\to 0} U(y)\left(1 - \frac{U'(y)}{U(y)}\right) - U'(y)d = -\infty. \end{cases}$$

In addition, because the utility function is a concave function, we have that

$$H'(x) = -U''(x)x > 0.$$

Therefore, $H(x)$ is an increasing and continuous function, and there exits a unique zero root for the equation $H(x) = 0$.[12] *Moreover, the zero root $x^* > d$ and does not depend on the financial market.*

Let $y = I(\lambda \widetilde{F}(z_D))$. Then, $G'(z_D)$ (16) can be rewritten as

$$G'(z_D) = -U(y) + U'(y)y + U'(y)D_T. \tag{A1}$$

Then Eq (A1) has a unique zero root in y, which is $y = \widehat{D}_T - D_T$. Hence, given $\lambda > 0$, $G'(z_D^*) = 0$ with $z_D^* = 1 - F\left(\frac{\widehat{D}_T - D_T}{\lambda}\right)$. □

Lemma A.1 *Given an initial wealth $x_0 > 0$ and $z_D(\lambda) := z_D^* \equiv 1 - F\left(\frac{U'(\widehat{D}_T - D_T)}{\lambda}\right)$, the function*

$$f(\lambda) := \int_{z_D(\lambda)}^{1} (I(\lambda \widetilde{F}(z)) + D_T)\widetilde{F}(z)\mathrm{d}z - x_0 \tag{A2}$$

has a unique zero root λ_B, i.e., $f(\lambda_B) = 0$.

Proof:

The function $f(\lambda)$ is continuous in λ. Since $\widehat{D}_T$ is a constant for a given utility function (Lemma 3.1), we have that $z_D(\lambda)$ is comonotonic with λ, i.e., $z_D(\lambda)$ increases as λ increases. Then, we have the following

$$\lim_{\lambda\to 0} f(\lambda) = \int_0^1 (I(\lambda \widetilde{F}(z)) + D_T)\widetilde{F}(z)\mathrm{d}z - x_0 = +\infty,$$

[12] In concavification, the unique zero root x^* is called the tangent point of the function $U(x-d)$, see for instance [10, 14, 30].

$$\lim_{\lambda\to+\infty} f(\lambda) = \int_1^1 (I(\lambda\widetilde{F}(z)) + D_T)\widetilde{F}(z)\mathrm{d}z - x_0 = -x_0.$$

Without loss of generality, we assume that $\lambda_1 > \lambda_2 > 0$ such that

$$f(\lambda_1) = \int_{z_D(\lambda_1)}^1 (I(\lambda_1\widetilde{F}(z)) + D_T)\widetilde{F}(z)\mathrm{d}z - x_0,$$
$$f(\lambda_2) = \int_{z_D(\lambda_2)}^1 (I(\lambda_2\widetilde{F}(z)) + D_T)\widetilde{F}(z)\mathrm{d}z - x_0.$$

Because $z_D(\lambda_1) > z_D(\lambda_2)$ and $I(\lambda_1\widetilde{F}(z)) < I(\lambda_2\widetilde{F}(z))$, we conclude that $f(\lambda_1) < f(\lambda_2)$. Therefore, $f(\lambda)$ is a strictly decreasing function in λ. Thus, $-x_0 < f(\lambda) < +\infty$ and $f(\lambda)$ has a unique zero root, i.e., $f(\lambda_B) = 0$. □

Appendix B Proof of Proposition 2

B.1 Proof of Proposition 2

Poof:

As in the benchmark problem, for an arbitrary probability $z_D \in [0,1]$, we define the set of quantile functions such that

$$\mathcal{Q}(z_D) = \{Q_X(z)|Q_X^{-1}(D_T) \leq z_D\}.$$

The quantile functions in $\mathcal{Q}(z_D)$ are larger than D_T if $z > z_D$ and are smaller than D_T if $z < z_D$. Given $\alpha > 0$ in the AVaR constraint, we consider two situations, $z_D > \alpha$ and $z_D < \alpha$.

Case 1: $z_D > \alpha$.

Note that if $z_D > \alpha$, the AVaR constraint is only effective on the region $0 \leq z \leq \alpha < z_D$. In this case, writing the quantile function from $\mathcal{Q}(z_D)$ as $Q_X(z) = Q_X^1(z)\mathbb{1}_{z_D<z\leq 1} + Q_X^2(z)\mathbb{1}_{0\leq z\leq z_D}$, we decompose the quantile-based problem AVaR into the following three sub-problems.

P1:

$$\max_{Q_X(z)\in\mathcal{Q}_D} \int_{z_D}^1 U(Q_X(z) - D_T)\mathrm{d}z, \quad \text{subject to} \quad \int_{z_D}^1 Q_X(z)\widetilde{F}(z)\mathrm{d}z = x_0^+.$$

P2:

$$\min_{Q_X(z)\in\mathcal{Q}(z_D)} \int_0^{z_D} Q_X(z)\widetilde{F}(z)\mathrm{d}z = x_0 - x_0^+, \quad \text{subject to} \quad \frac{1}{\alpha}\int_0^\alpha Q_X(z)\mathrm{d}z \geq L.$$

P3:

$$\max_{z_D \in (\alpha,1]} \max_{Q_X(z)\in\mathcal{Q}(z_D)} \int_{z_D}^1 U(Q_X(z)-D_T)\mathrm{d}z,$$
$$\text{subject to} \quad \int_0^1 Q_X(z)\widetilde{F}(z)\mathrm{d}z = x_0, \quad \frac{1}{\alpha}\int_0^\alpha Q_X(z)\mathrm{d}z \geq L.$$

In the Problem ***P1***, we maximize the expected utility on the region $z > z_D$ under the budget constraint. In the Problem ***P2***, we minimize the cost on the region $0 \leq z \leq z_D$, while satisfying the AVaR constraint. In the Problem ***P3***, we find the global maximum over $z_D \in (\alpha, 1]$.

Let $\lambda_1 > 0$ be the Lagrangian multiplier for the budget constraint. The optimal quantile function for the Problem ***P1*** is given by

$$Q_X^{1*}(z) = I(\lambda_1 \widetilde{F}(z)) + D_T.$$

Let $\beta \in [0,\alpha]$. The quantile function on $[0, z_D]$ satisfying the AVaR constraint is given by $Q_X^2(z) = l\mathbb{1}_{\beta<z\leq z_D}$. Lemma B.1 gives the minimum cost on $\beta \in [0, z_D]$.

Lemma B.1 *The cost function on $z \in [0, z_D]$ given by*

$$g(\beta) := \int_\beta^{z_D} \frac{\alpha L}{\alpha-\beta} F_\xi^{-1}(1-z)\mathrm{d}z \tag{B3}$$

attains its minimum at β^*, where β^* satisfies the following equation

$$g'(\beta^*) = \frac{\alpha L}{(\alpha-\beta^*)^2}\left(\int_{\beta^*}^{z_D} F_\xi^{-1}(1-z)\mathrm{d}z - (\alpha-\beta^*)F_\xi^{-1}(1-\beta^*)\right) = 0. \tag{B4}$$

Proof:

The derivative of the cost function (B3) is

$$\begin{aligned} g'(\beta) &= \frac{\alpha L}{(\alpha-\beta)^2}\int_\beta^{z_D} F_\xi^{-1}(1-z)\mathrm{d}z - \frac{\alpha L}{\alpha-\beta}F_\xi^{-1}(1-\beta) \\ &= \frac{\alpha L}{(\alpha-\beta)^2}\left(\int_\beta^{z_D} F_\xi^{-1}(1-z)\mathrm{d}z - (\alpha-\beta)F_\xi^{-1}(1-\beta)\right). \end{aligned} \tag{B5}$$

Note that $g'(\beta) \to -\infty$ as $\beta \to 0$. In addition, as $\beta \to \alpha$, both upper and lower bounds of the cost function tend to infinity. I.e.,

$$\lim_{\beta\to\alpha} g'(\beta) < \lim_{\beta\to\alpha} \overline{g'(\beta)} := \lim_{\beta\to\alpha} \frac{\alpha L}{(\alpha-\beta)^2}((z_D-\alpha)F_\xi^{-1}(1-\beta)) = \infty.$$
$$\lim_{\beta\to\alpha} g'(\beta) > \lim_{\beta\to\alpha} \underline{g'(\beta)} := \lim_{\beta\to\alpha} \frac{\alpha L}{(\alpha-\beta)^2}(z_D-\beta)F_\xi^{-1}(1-z_D) = \infty.$$

Therefore we conclude that $g'(\beta)$ tends to ∞ as β tends to α. Without loss of generality, assuming that $0<\beta_1<\beta_2<\alpha$,

$$
\begin{aligned}
& g'(\beta_1)-g'(\beta_2) \\
=& \frac{\alpha L}{(\alpha-\beta_1)^2}\left(\int_{\beta_1}^{z_D} F_\xi^{-1}(1-z)\mathrm{d}z-(\alpha-\beta_1)F_\xi^{-1}(1-\beta_1)\right) \\
-& \frac{\alpha L}{(\alpha-\beta_2)^2}\left(\int_{\beta_2}^{z_D} F_\xi^{-1}(1-z)\mathrm{d}z-(\alpha-\beta_2)F_\xi^{-1}(1-\beta_2)\right) \\
<& \frac{\alpha L}{(\alpha-\beta_2)^2}\left(\int_{\beta_1}^{z_D} F_\xi^{-1}(1-z)\mathrm{d}z-(\alpha-\beta_1)F_\xi^{-1}(1-\beta_1)\right. \\
-& \left.\int_{\beta_2}^{z_D} F_\xi^{-1}(1-z)\mathrm{d}z+(\alpha-\beta_2)F_\xi^{-1}(1-\beta_2)\right) \\
=& \frac{\alpha L}{(\alpha-\beta_2)^2}\left(\int_{\beta_1}^{\beta_2} F_\xi^{-1}(1-z)\mathrm{d}z+(\alpha-\beta_2)F_\xi^{-1}(1-\beta_2)-(\alpha-\beta_1)F_\xi^{-1}(1-\beta_1)\right) \\
<& \frac{\alpha L}{(\alpha-\beta_2)^2}\left((\beta_2-\beta_1)F_\xi^{-1}(1-\beta_1)+(\alpha-\beta_2)F_\xi^{-1}(1-\beta_2)-(\alpha-\beta_1)F_\xi^{-1}(1-\beta_1)\right) \\
=& \frac{\alpha L}{(\alpha-\beta_2)^2}\left((\alpha-\beta_2)\underbrace{\left(F_\xi^{-1}(1-\beta_2)-F_\xi^{-1}(1-\beta_1)\right)}_{<0 \quad \text{because} \quad \beta_1<\beta_2}\right)<0.
\end{aligned}
$$

Hence, $g'(\beta)$ is an increasing function. We conclude that there is a unique zero root of the function (B5), which is denoted by β^*.□

Note that the cost function decreases on $\beta<\beta^*$ $(g'(\beta)<0)$, and increases on $\beta>\beta^*$ $(g'(\beta)>0)$. Thus, $g(\beta^*)$ attains the minimum cost. Correspondingly, the quantile function on the region $z<z_D$ is $Q_X^2(z)=\frac{\alpha L}{\alpha-\beta^*}\mathbb{1}_{\beta^*<z\le z_D}$.

Together with the quantile function on $z>z_D$, the optimal quantile function on the entire set is given by

$$
Q_X^*(z)=(I(\lambda_1\widetilde{F}(z)+D_T)\mathbb{1}_{z_D<z\le 1}+\frac{\alpha L}{\alpha-\beta^*}\mathbb{1}_{\beta^*<z\le z_D}. \tag{B6}
$$

Note that $z_D>\alpha$ implies $D_T\ge\frac{\alpha L}{\alpha-\beta^*}$. Hence, the quantile function (B6) is an increasing function. Now we find the global maximum of the problem over $z_D\in(\alpha,1]$ (***P3***). The Lagrangian for ***P3*** is given by

$$
\begin{aligned}
& \mathbb{G}(z_D) \qquad \text{(B7)} \\
=& \int_{z_D}^1 U(Q_X^1(z)-D_T)\mathrm{d}z-\lambda_1\int_{z_D}^1 Q_X^1(z)\widetilde{F}(z)\mathrm{d}z-\lambda_1\left(\int_0^{z_D} Q_X^2(z)\widetilde{F}(z)\mathrm{d}z\right).
\end{aligned}
$$

The derivative of $\mathbb{G}(z_D)$ (B7) with respect to z_D is given by

$$\begin{aligned} &\mathbb{G}'(z_D) && \text{(B8)} \\ =& - U(Q_X^1(z_D) - D_T) + \lambda_1 Q_X^1(z_D)\widetilde{F}(z_D) - \lambda_1 Q_X^2(z_D)\widetilde{F}(z_D) \\ =& - U(I(\lambda_1 \widetilde{F}(z_D))) + \lambda_1 \widetilde{F}(z_D) I(\lambda_1 \widetilde{F}(z_D)) + \lambda_1 \widetilde{F}(z_D) D_T - l\lambda_1 \widetilde{F}(z_D) \\ =& - U(I(\lambda_1 \widetilde{F}(z_D))) + \lambda_1 \widetilde{F}(z_D)(I(\lambda_1 \widetilde{F}(z_D) + D_T - l), && \text{(B9)} \end{aligned}$$

where $l = \frac{\alpha L}{\alpha - \beta^*}$. Letting $x = I(\lambda_1 F_\xi^{-1}(1 - z_D)) + D_T - l$, Eq (B9) can be written as

$$\mathbb{G}'(z_D) = -U(x - (D_T - l)) + U'(x - (D_T - l))x.$$

Thus, by Lemma 3.1, Eq (B9) has a unique zero root, which is denoted by z_D^*. Further, if $z_D^* > \alpha$, the globally optimal quantile function is given by

$$Q_X^*(z) = (I(\lambda_1 \widetilde{F}(z) + D_T)\mathbb{1}_{z_D^* < z \leq 1} + \frac{\alpha L}{\alpha - \beta^*}\mathbb{1}_{\beta^* < z \leq z_D^*}. \tag{B10}$$

However, if $z_D^* \leq \alpha$, we go to the second case $z_D < \alpha$.

Case 2: $z_D < \alpha$.

In this case, the AVaR constraint is effective on both regions $\alpha > z > z_D$ and $z < z_D$. Let us first consider the extreme case that $z_D = 0$, which implies that the quantile functions in $\mathcal{Q}(z_D)$ are all above D_T. The optimization problem becomes:

Problem P0:

$$\max_{Q_X(z) \in \mathcal{Q}_D} \int_0^1 U(Q_X(z) - D_T)\mathrm{d}z, \text{ s.t. } \int_0^1 Q_X(z)\widetilde{F}(z)\mathrm{d}z = x_0, \ \frac{1}{\alpha}\int_0^\alpha Q_X(z)\mathrm{d}z \geq L.$$

Note that the ***Problem P0*** is a concave optimization problem. Wei [36] provides the analytical solution to a similar problem by the concavification technique. In this study, we apply the point-wise Lagrangian approach to solve the problem.

We consider the quantile function $Q_X(z) = Q_X^1(z)\mathbb{1}_{\alpha < z \leq 1} + Q_X^2(z)\mathbb{1}_{0 \leq z \leq \alpha}$ and the following Lagrangian

$$\mathbb{G}(\alpha) = \begin{cases} \int_\alpha^1 U(Q_X^1(z) - D_T)\mathrm{d}z - \lambda_1 \int_\alpha^1 Q_X^1(z)\widetilde{F}(z)\mathrm{d}z, \\ \int_0^\alpha U(Q_X^2(z) - D_T)\mathrm{d}z - \lambda_1 \int_0^\alpha Q_X^2(z)\widetilde{F}(z)\mathrm{d}z + \lambda_1 \frac{\lambda_2}{\alpha} \int_0^\alpha Q_X^2(z)\mathrm{d}z. \end{cases} \tag{B11}$$

The first order condition gives that

$$Q^*_X(z) = \begin{cases} I(\lambda_1 F_\xi^{-1}(1-z)) + D_T, & \alpha < z \le 1 \\ \left(I\left(\lambda_1 F_\xi^{-1}(1-z) - \frac{\lambda_1 \lambda_2}{\alpha} \right) + D_T \right), & 0 \le z \le \alpha. \end{cases} \tag{B12}$$

Note that there are infinite pairs of $z \in (\alpha, 1]$ and $s \in [0, \alpha]$ such that

$$F_\xi^{-1}(1-z) = F_\xi^{-1}(1-s) - \frac{\lambda_2}{\alpha}. \tag{B13}$$

And for each pair (z_i, s_i) satisfying (B13), we have the following:

$$\begin{cases} \text{if} \quad \alpha < z < z_i : I(\lambda_1 \widetilde{F}(z)) < I(\lambda_1 \widetilde{F}(z_i)) = I\left(\lambda_1 \widetilde{F}(s_i) - \frac{\lambda_1 \lambda_2}{\alpha} \right), \\ \text{if} \quad \alpha \ge s > s_i : I\left(\lambda_1 \widetilde{F}(s)) - \frac{\lambda_1 \lambda_2}{\alpha} \right) > I\left(\lambda_1 \widetilde{F}(s_i) - \frac{\lambda_1 \lambda_2}{\alpha} \right) = I(\lambda_1 \widetilde{F}(z_i)). \end{cases}$$

Therefore, the optimal quantile function should take the following form:

$$Q^*_X(z) = (I(\lambda_1 \widetilde{F}(z)) + D_T) \mathbb{1}_{z_i < z \le 1} + \widehat{L}_i \mathbb{1}_{s_i < z \le z_i} + \left(I\left(\lambda_1 \widetilde{F}(z) - \frac{\lambda_1 \lambda_2}{\alpha} \right) + D_T \right) \mathbb{1}_{0 \le z \le s_i}, \tag{B14}$$

where $\widehat{L}_i = I(\lambda_1 F_\xi^{-1}(1-z_i)) + D_T = I\left(\lambda_1 F_\xi^{-1}(1-s_i) - \frac{\lambda_1 \lambda_2}{\alpha} \right) + D_T$ and (z_i, s_i) satisfy the function (B13).

Lemma B.2 *Given a positive Lagrangian multiplier $\lambda_2 > 0$, we have that*

1. $\alpha < z_i \le 1 - F\left(F_\xi^{-1}(1-\alpha) - \frac{\lambda_2}{\alpha} \right) =: \bar{z}$; $\quad \bar{s} := 1 - F\left(F_\xi^{-1}(1-\alpha) + \frac{\lambda_2}{\alpha} \right) \le s_i \le \alpha$.
2. *The constant quantile function $\widehat{L}_i$ falls in the bound*

$$\widehat{L}^b := I(\lambda_1 F_\xi^{-1}(1-\alpha)) + D_T \le \widehat{L}_i \le \widehat{L}^u =: I\left(\lambda_1 F_\xi^{-1}(1-\alpha) - \frac{\lambda_1 \lambda_2}{\alpha} \right) + D_T.$$

3. *If $\widehat{L}_1 = I(\lambda_1 F_\xi^{-1}(1-z_1)) + D_T < \widehat{L}_2 = I(\lambda_1 F_\xi^{-1}(1-z_2)) + D_T$, then $s_1 < s_2 < z_1 < z_2$.*

Proof:

Note that $z_i = 1 - F(F^{-1}(1-s_i) - \frac{\lambda_2}{\alpha})$ and $s_i = 1 - F(F_\xi^{-1}(1-z_i) + \frac{\lambda_2}{\alpha})$ by (B13). Because $s_i \leq \alpha$, we know that

$$\alpha < z_i \leq 1 - F\bigg(F^{-1}(1-\alpha) - \frac{\lambda_2}{\alpha}\bigg) = \bar{z}.$$

Similarly, since $z_i > \alpha$, we have the bounds for s_i

$$\bar{s} = 1 - F(F_\xi^{-1}(1-\alpha) + \frac{\lambda_2}{\alpha}) < s \leq \alpha.$$

Moreover, as $\widehat{L}_i = I(\lambda_1 F_\xi^{-1}(1-z_i)) + D_T$, we derive the bounds for $\widehat{L}_i$

$$\widehat{L}^b = I(\lambda_1 F_\xi^{-1}(1-\alpha)) + D_T \leq \widehat{L}_i \leq \widehat{L}^u = I\left(\lambda_1 F_\xi^{-1}(1-\alpha) - \frac{\lambda_1\lambda_2}{\alpha}\right) + D_T.$$

Further, if $\widehat{L}_1 < \widehat{L}_2$, we have that $I(\lambda_1 F_\xi^{-1}(1-z_1)) < I(\lambda_1 F_\xi^{-1}(1-z_2))$. Since $I(\cdot)$ is an increasing function of z, we have that $z_1 < z_2$. With a similar argument, we obtain that $s_1 < s_2$.□

Now, we regard the Lagrangian (B11) plugged in with the optimal quantile function (B14) as a function of $\widehat{L}_i$. Then, we find the global maximum of the Lagrangian (B11) over the domain of $\widehat{L}_i$. Recall that for each given $\widehat{L}_i$, (z_i, s_i) can also be expressed as functions of $\widehat{L}_i$, i.e.,

$$z_i = f(\widehat{L}_i) := 1 - F\left(\frac{U'(\widehat{L}_i - D_T)}{\lambda_1}\right);\ s_i = g(\widehat{L}_i) := 1 - F\left(\frac{U'(\widehat{L}_i - D_T)}{\lambda_1} + \frac{\lambda_2}{\alpha}\right).$$

Lemma B.3 *The Lagrangian* (B11) *plugged in with the quantile function* (B14) *is given by*

$$\begin{aligned}\mathbb{G}(\alpha, \widehat{L}_i) = &\int_{f(\widehat{L}_i)}^1 U(Q_1(z) - D_T)\mathrm{d}z + \int_{g(\widehat{L}_i)}^{f(\widehat{L}_i)} U(\widehat{L}_i - D_T)\mathrm{d}z + \int_0^{g(\widehat{L}_i)} U(Q_2(z) - D_T)\mathrm{d}z \\ &-\lambda_1 \int_{f(\widehat{L}_i)}^1 Q_1(z)\widetilde{F}(z)\mathrm{d}z - \lambda_1 \int_{g(\widehat{L}_i)}^{f(\widehat{L}_i)} \widehat{L}_i \widetilde{F}(z)\mathrm{d}z - \lambda_1 \int_0^{g(\widehat{L}_i)} Q_2(z)\widetilde{F}(z)\mathrm{d}z \\ &+\frac{\lambda_1\lambda_2}{\alpha}\int_{g(\widehat{L}_i)}^{\alpha} \widehat{L}_i \mathrm{d}z + \frac{\lambda_1\lambda_2}{\alpha}\int_0^{g(\widehat{L}_i)} Q_2(z)\mathrm{d}z, \end{aligned} \tag{B15}$$

where $Q_1(z) = I(\lambda_1 F_\xi^{-1}(1-z)) + D_T$ *if* $z > f(\widehat{L}_i)$ *and* $Q_2(z) = I\bigg(\lambda_1 F_\xi^{-1}(1-z) - \frac{\lambda_1\lambda_2}{\alpha}\bigg) + D_T$ *if* $z < g(\widehat{L}_i)$*. The Lagrangian function* (B15) *is increasing on* $\widehat{L}_i < \widehat{L}^*$ *and is decreasing on* $\widehat{L}_i > \widehat{L}^*$*, and attains its maximum at* $\widehat{L}_i = \widehat{L}^*$*, where* $\widehat{L}^*$ *is the unique zero root of the function* $G'(\widehat{L}_i) = 0$*.*

Proof:

If we can show that the derivative of the Lagrangian function (B15) has a unique zero root, i.e, $\mathbb{G}'(\widehat{L}^*) = 0$, and $\mathbb{G}'(\widehat{L}_i) > 0$ if $\widehat{L}_i < \widehat{L}^*$ and $\mathbb{G}'(\widehat{L}_i) < 0$ if $\widehat{L}_i > \widehat{L}^*$, then Lemma B.3 is proved.

The derivative of the Lagrangian with respect to $\widehat{L}_i$ is given by[13]

$$\begin{aligned}
&\mathbb{G}'(\widehat{L}_i)\\
&= -f'(\widehat{L}_i)U(Q_1(f(\widehat{L}_i) - D_T) + g'(\widehat{L}_i)U(Q_2(g(\widehat{L}_i) - D_T) + (f'(\widehat{L}_i) - g'(\widehat{L}_i))U(\widehat{L}_i - D_T)\\
&+\lambda_1 f'(\widehat{L}_i)\widetilde{F}(f(\widehat{L}_i))Q_1(f(\widehat{L}_i)) - \lambda_1 g'(\widehat{L}_i)\widetilde{F}(g(\widehat{L}_i))Q_2(g(\widehat{L}_i)) + \int_{g(\widehat{L}_i)}^{f(\widehat{L}_i)} U'(\widehat{L}_i - D_T)\mathrm{d}z\\
&-\lambda_1 f'(\widehat{L}_i)\widetilde{F}(f(\widehat{L}_i))\widehat{L}_i + \lambda_1 g'(\widehat{L}_i)\widetilde{F}(g(\widehat{L}_i))\widehat{L}_i - \lambda_1 \int_{g(\widehat{L}_i)}^{f(\widehat{L}_i)} \widetilde{F}(z)\mathrm{d}z\\
&-\frac{\lambda_1\lambda_2}{\alpha} g'(\widehat{L}_i)\widehat{L}_i + \frac{\lambda_1\lambda_2}{\alpha} g'(\widehat{L}_i)Q_2(g(\widehat{L}_i)) + \frac{\lambda_1\lambda_2}{\alpha}(\alpha - g(\widehat{L}_i)).
\end{aligned}$$

Since

$$Q_1(f(\widehat{L}_i)) := I(\lambda_1 \widetilde{F}(z_i)) + D_T = I\big(\lambda_1 \widetilde{F}(s_i) - \frac{\lambda_1\lambda_2}{\alpha}\big) + D_T =: Q_2(g(\widehat{L}_i)) = \widehat{L}_i$$

the derivative function simplifies to

$$\begin{aligned}
&\mathbb{G}'(\widehat{L}_i)\\
&= \int_{g(\widehat{L}_i)}^{f(\widehat{L}_i)} U'(\widehat{L}_i - D_T)\mathrm{d}z - \lambda_1 \int_{g(\widehat{L}_i)}^{f(\widehat{L}_i)} \widetilde{F}(z)\mathrm{d}z + \frac{\lambda_1\lambda_2}{\alpha}(\alpha - g(\widehat{L}_i))\\
&\quad = \underbrace{\lambda_1 \widetilde{F}(f(\widehat{L}_i))}_{=U'(\widehat{L}_i - D_T)}(f(\widehat{L}_i) - g(\widehat{L}_i)) - \lambda_1 \int_{g(\widehat{L}_i)}^{f(\widehat{L}_i)} \widetilde{F}(z)\mathrm{d}z + \frac{\lambda_1\lambda_2}{\alpha}(\alpha - g(\widehat{L}_i)).
\end{aligned} \tag{B16}$$

Recalling that $\widehat{L}^b \leq \widehat{L}_i \leq \widehat{L}^u$, we derive the bounds for the derivative of $\mathbb{G}'(\widehat{L}_i)$ (B16). Note that $\widehat{L}_i = \widehat{L}^u$ implies that $f(\widehat{L}^u) = \bar{z}$ and $g(\widehat{L}^u) = \alpha$. Thus, we have that

$$\begin{aligned}
G'(\widehat{L}^u) &= \lambda_1 \widetilde{F}(f(\widehat{L}^u))(f(\widehat{L}^u) - g(\widehat{L}^u)) - \lambda_1 \int_{g(\widehat{L}^u)}^{f(\widehat{L}^u)} \widetilde{F}(z)\mathrm{d}z + \frac{\lambda_1\lambda_2}{\alpha}(\alpha - g(\widehat{L}^u))\\
&= \lambda_1 \widetilde{F}(\bar{z})(\bar{z} - \alpha) - \lambda_1 \int_{\alpha}^{\bar{z}} \widetilde{F}(z)\mathrm{d}z < \lambda_1 \widetilde{F}(\bar{z})(\bar{z} - \alpha) - \lambda_1(\bar{z} - \alpha)\widetilde{F}(\bar{z}) = 0.
\end{aligned}$$

[13] We omit the dependence of the Lagrangian on α because α is a constant.

Similarly, we have that $G'(\widehat{L}^b) > 0$,

$$
\begin{aligned}
& G'(\widehat{L}^b) \\
=& \lambda_1 \widetilde{F}(f(\widehat{L}^b))(f(\widehat{L}^u) - g(\widehat{L}^b)) - \lambda_1 \int_{g(\widehat{L}^u)}^{f(\widehat{L}^b)} \widetilde{F}(z)\mathrm{d}z + \frac{\lambda_1\lambda_2}{\alpha}(\alpha - g(\widehat{L}^b)) \\
&= \lambda_1 \widetilde{F}(\alpha)(\alpha - \bar{s}) - \lambda_1 \int_{\bar{s}}^{\alpha} \widetilde{F}(z)\mathrm{d}z + \frac{\lambda_1\lambda_2}{\alpha}(\alpha - \bar{s}) \\
&> \lambda_1 \widetilde{F}(\alpha)(\alpha - \bar{s}) - \lambda_1(\alpha - \bar{s})\widetilde{F}(\bar{s}) + \frac{\lambda_1\lambda_2}{\alpha}(\alpha - \bar{s}) \\
&= \lambda_1 \widetilde{F}(\alpha)(\alpha - \bar{s}) - \lambda_1(\alpha - \bar{s}) \underbrace{\left(\widetilde{F}(\bar{s}) - \frac{\lambda_2}{\alpha}\right)}_{=\widetilde{F}(f(\widehat{L}^b))=\widetilde{F}(\alpha)} = 0.
\end{aligned}
$$

Moreover, since $\mathbb{G}'(\widehat{L}_i)$ is continuous, there is at least one $\widehat{L}_i$ such that $\mathbb{G}'(\widehat{L}_i) = 0$.

Next, we show that $\mathbb{G}'(\widehat{L}_i)$ (B16) is a decreasing function. Without loss of generosity, we assume that $\widehat{L}_1 < \widehat{L}_2$, which implies that $g(\widehat{L}_1) < g(\widehat{L}_2) < f(\widehat{L}_1) < f(\widehat{L}_2)$. Now, we compare $\mathbb{G}'(\widehat{L}_1)$ and $\mathbb{G}'(\widehat{L}_2)$, i.e.,

$$
\begin{aligned}
& \mathbb{G}'(\widehat{L}_1) - \mathbb{G}'(\widehat{L}_2) \\
=& \lambda_1 \widetilde{F}(f(\widehat{L}_1))(f(\widehat{L}_1) - g(\widehat{L}_1)) - \lambda_1 \int_{g(\widehat{L}_1)}^{f(\widehat{L}_1)} \widetilde{F}(z)\mathrm{d}z + \lambda_1 \frac{\lambda_2}{\alpha}(\alpha - g(\widehat{L}_1)) \\
& - \lambda_1 \widetilde{F}(f(\widehat{L}_2))(f(\widehat{L}_2) - g(\widehat{L}_2)) + \lambda_1 \int_{g(\widehat{L}_2)}^{f(\widehat{L}_2)} \widetilde{F}(z)\mathrm{d}z - \lambda_1 \frac{\lambda_2}{\alpha}(\alpha - g(\widehat{L}_2)) \\
&= \lambda_1 F_{\xi}^{-1}(1 - f(\widehat{L}_1))(f(\widehat{L}_1) - g(\widehat{L}_1)) - \lambda_1 F_{\xi}^{-1}(1 - f(\widehat{L}_2))(f(\widehat{L}_2) - g(\widehat{L}_2)) \\
& + \lambda_1 \frac{\lambda_2}{\alpha}(g(\widehat{L}_2) - g(\widehat{L}_1)) + \lambda_1 \int_{f(\widehat{L}_1)}^{f(\widehat{L}_2)} \widetilde{F}(z)\mathrm{d}z + \lambda_1 \int_{g(\widehat{L}_2)}^{f(\widehat{L}_1)} \widetilde{F}(z)\mathrm{d}z \\
& - \lambda_1 \int_{g(\widehat{L}_1)}^{g(\widehat{L}_2)} \widetilde{F}(z)\mathrm{d}z - \lambda_1 \int_{g(\widehat{L}_2)}^{f(\widehat{L}_1)} \widetilde{F}(z)\mathrm{d}z \\
&= \lambda_1 \underbrace{\int_{f(\widehat{L}_1)}^{f(\widehat{L}_2)} \widetilde{F}(z)\mathrm{d}z}_{A} - \lambda_1 \underbrace{\int_{g(\widehat{L}_1)}^{g(\widehat{L}_2)} \widetilde{F}(z)\mathrm{d}z}_{B} + \lambda_1 \underbrace{\frac{\lambda_2}{\alpha}(g(\widehat{L}_2) - g(\widehat{L}_1))}_{C} \\
& + \lambda_1 \underbrace{\widetilde{F}(f(\widehat{L}_1))(f(\widehat{L}_1) - g(\widehat{L}_1))}_{D} - \lambda_1 \underbrace{\widetilde{F}(f(\widehat{L}_2))(f(\widehat{L}_2) - g(\widehat{L}_2))}_{E}.
\end{aligned}
$$

We make use of integration by parts to deal with the terms A and B, i.e,

$$
\begin{aligned}
\int_{f(\widehat{L}_1)}^{f(\widehat{L}_2)} \widetilde{F}(z)\mathrm{d}z =& \widetilde{F}(z)z|_{f(\widehat{L}_1)}^{f(\widehat{L}_2)} - \int_{f(\widehat{L}_1)}^{f(\widehat{L}_2)} z\mathrm{d}\widetilde{F}(z) \\
=& \widetilde{F}(f(\widehat{L}_2))f(\widehat{L}_2) - \widetilde{F}(f(\widehat{L}_1))f(\widehat{L}_1) - \int_{f(\widehat{L}_1)}^{f(\widehat{L}_2)} z\mathrm{d}\widetilde{F}(z) \\
\int_{g(\widehat{L}_1)}^{g(\widehat{L}_2)} \widetilde{F}(z)\mathrm{d}z =& \widetilde{F}(z)z|_{g(\widehat{L}_1)}^{g(\widehat{L}_2)} - \int_{g(\widehat{L}_1)}^{g(\widehat{L}_2)} z\mathrm{d}\widetilde{F}(z) \\
=& \widetilde{F}(g(\widehat{L}_2))g(\widehat{L}_2) - \widetilde{F}(g(\widehat{L}_1))g(\widehat{L}_1) - \int_{g(\widehat{L}_1)}^{g(\widehat{L}_2)} z\mathrm{d}\widetilde{F}(z).
\end{aligned}
$$

Note that the terms $D - E + C$ is

$$
\begin{aligned}
& D - E + C \\
=& \widetilde{F}(f(\widehat{L}_1))f(\widehat{L}_1) - \widetilde{F}(f(\widehat{L}_2))f(\widehat{L}_2) - (\widetilde{F}(f(\widehat{L}_1)) + \frac{\lambda_2}{\alpha})g(\widehat{L}_1) \\
+& (\widetilde{F}(f(\widehat{L}_2)) + \frac{\lambda_2}{\alpha})g(\widehat{L}_2) \\
=& \widetilde{F}(f(\widehat{L}_1))f(\widehat{L}_1) - \widetilde{F}(f(\widehat{L}_2))f(\widehat{L}_2) - \widetilde{F}(g(\widehat{L}_1))g(\widehat{L}_1) + \widetilde{F}(g(\widehat{L}_2))g(\widehat{L}_2).
\end{aligned}
$$

Therefore, $\mathbb{G}^{'}(\widehat{L}_1) - \mathbb{G}^{'}(\widehat{L}_2)$ simplifies to

$$
\begin{aligned}
\mathbb{G}^{'}(\widehat{L}_1) - \mathbb{G}^{'}(\widehat{L}_2) =& \lambda_1 \left(\int_{g(\widehat{L}_1)}^{g(\widehat{L}_2)} z\mathrm{d}{F_\xi}^{-1}(1-z) - \int_{f(\widehat{L}_1)}^{f(\widehat{L}_2)} z\mathrm{d}{F_\xi}^{-1}(1-z) \right) \\
=& \lambda_1 \int_{f(\widehat{L}_1)}^{f(\widehat{L}_2)} (g(\widehat{L}_i) - z)\mathrm{d}{F_\xi}^{-1}(1-z) \\
=& \lambda_1 \int_{f(\widehat{L}_1)}^{f(\widehat{L}_2)} \underbrace{(g(\widehat{L}_i) - z)}_{<0} \underbrace{({F_\xi}^{-1})^{'}(1-z)}_{<0} \mathrm{d}z > 0.
\end{aligned}
$$

Hence, we conclude that $\mathbb{G}'(\widehat{L}_i)$ is a decreasing function, and there is a unique zero root to the function $\mathbb{G}'(\widehat{L}_i) = 0$. We use $\widehat{L}^*$ to denote the unique zero root of the function $\mathbb{G}'(\widehat{L}_i) = 0$ and the correspondingly

$$
f(\widehat{L}^*) = 1 - F\left(\frac{U'(\widehat{L}^* - D_T)}{\lambda_1}\right) =: z^*, \quad g(\widehat{L}^*) = 1 - F\left(\frac{U'(\widehat{L}^* - D_T)}{\lambda_1} + \frac{\lambda_2}{\alpha}\right) =: s^*.
$$

In addition, the Lagrangian $\mathbb{G}(\widehat{L}_i)$ increases on $\widehat{L}_i < \widehat{L}^*$ ($\mathbb{G}'(\widehat{L}_i) > 0$) and decreases on $\widehat{L}^* < \widehat{L}_i$ ($\mathbb{G}'(\widehat{L}_i) < 0$), and attains its maximum at $\widehat{L}^*$.□

Therefore, the optimal quantile function for Problem P0 is given by

$$Q_X^0 = (I(\lambda_1 \widetilde{F}(z)) + D_T)\mathbb{1}_{z^* < z \le 1} + \widehat{L}^* \mathbb{1}_{s^* < z \le z^*} + \left(I\left(\lambda_1 \widetilde{F}(z) - \frac{\lambda_1 \lambda_2}{\alpha}\right) + D_T \right) \mathbb{1}_{0 \le z \le s^*}. \tag{B17}$$

Now let us consider the general case, i.e., $0 < z_D < \alpha$. Lemma B.3 is still valid except that the function (B15) becomes

$$\mathbb{G}(z_D, \widehat{L}_i) \tag{B18}$$

$$\begin{aligned}
&= \int_{f(\widehat{L}_i)}^{1} U(Q_1(z) - D_T)\mathrm{d}z + \int_{g(\widehat{L}_i)}^{f(\widehat{L}_i)} U(\widehat{L}_i - D_T)\mathrm{d}z + \int_{z_D}^{g(\widehat{L}_i)} U(Q_2(z) - D_T)\mathrm{d}z \\
&-\lambda_1 \int_{f(\widehat{L}_i)}^{1} Q_1(z)\widetilde{F}(z)\mathrm{d}z - \lambda_1 \int_{g(\widehat{L}_i)}^{f(\widehat{L}_i)} \widehat{L}_i \widetilde{F}(z)\mathrm{d}z \\
&-\lambda_1 \int_{z_D}^{g(\widehat{L}_i)} Q_2(z)\widetilde{F}(z)\mathrm{d}z + \frac{\lambda_1\lambda_2}{\alpha}\int_{g(\widehat{L}_i)}^{\alpha} \widehat{L}_i \mathrm{d}z + \frac{\lambda_1\lambda_2}{\alpha}\int_{z_D}^{g(\widehat{L}_i)} Q_2(z)\mathrm{d}z,
\end{aligned} \tag{B19}$$

where $Q_1(z) = I(\lambda_1 \widetilde{F}(z)) + D_T$ and $Q_2(z) = I(\lambda_1 \widetilde{F}(z) - \frac{\lambda_1\lambda_2}{\alpha}) + D_T$. Note that the partial derivative of $\mathbb{G}(z_D, \widehat{L}_i)$ with respect to $\widehat{L}_i$ is the same as $\mathbb{G}'(\widehat{L}_i)$ (B16), and is not affected by z_D. Therefore, by Lemma B.3, there is a unique zero root for $\partial\mathbb{G}(z_D, \widehat{L}_i)/\partial\widehat{L}_i = 0$ denoted by $\widehat{L}^*$. Correspondingly, we have that $f(\widehat{L}^*) = 1 - F_\xi\left(\frac{U'(\widehat{L}^* - D_T)}{\lambda_1}\right) =: z^*$ and $g(\widehat{L}^*) = 1 - F_\xi\left(\frac{U'(\widehat{L}^* - D_T)}{\lambda_1} + \frac{\lambda_2}{\alpha}\right) =: s^*$. Next, we find the global maximum of $\mathbb{G}(z_D, \widehat{L}_i)$ over $z_D < \alpha$.

Lemma B.4 *For a fixed $z_D < \alpha$, the maximum of the function (B19) is given by*

$$\mathbb{G}(z_D, \widehat{L}^*) \tag{B20}$$

$$\begin{aligned}
&= \int_{z^*}^{1} U(Q_1(z) - D_T)\mathrm{d}z + \int_{s^*}^{z^*} U(\widehat{L}^* - D_T)\mathrm{d}z + \int_{z_D}^{s^*} U(Q_2(z) - D_T)\mathrm{d}z \\
&-\lambda_1 \int_{z^*}^{1} Q_1(z)\widetilde{F}(z)\mathrm{d}z - \lambda_1 \int_{s^*}^{z^*} \widehat{L}_i \widetilde{F}(z)\mathrm{d}z - \lambda_1 \int_{z_D}^{s^*} Q_2(z)\widetilde{F}(z)\mathrm{d}z \\
&+\frac{\lambda_1\lambda_2}{\alpha}\int_{s^*}^{\alpha} \widehat{L}_i \mathrm{d}z + \frac{\lambda_1\lambda_2}{\alpha}\int_{z_D}^{s^*} Q_2(z)\mathrm{d}z,
\end{aligned} \tag{B21}$$

where $Q_1(z) = I(\lambda_1 \widetilde{F}(z)) + D_T$ and $Q_2(z) = I(\lambda_1 \widetilde{F}(z) - \frac{\lambda_1\lambda_2}{\alpha}) + D_T$. The partial derivative of the function (B21) has a unique zero root, i.e., $\partial\mathbb{G}(z_D, \widehat{L}^*)/\partial z_D(z_D^*) = 0$.

1. *If $\widehat{L}^* > \widehat{D}_T$, the Lagrangian (B21) attains its maximum at $z_D = z_D^*$;*

2. If $L \le \widehat{L}^* \le \widehat{D}_T$, the Lagrangian (B21) attains its maximum at $z_D = \alpha(\widehat{L}^* - L)/\widehat{L}^*$.

Proof:
The partial derivative of $\mathbb{G}(z_D, \widehat{L}^*)$ with respect to z_D is given by

$$
\begin{aligned}
&\frac{\partial \mathbb{G}(z_D, \widehat{L}^*)}{\partial z_D} \\
&= -U(Q_2(z_D) - D_T) + \lambda_1 Q_2(z_D)\widetilde{F}(z_D) - \frac{\lambda_1\lambda_2}{\alpha} Q_2(z_D) \\
&= -U\left(I\left(\lambda_1 \widetilde{F}(z) - \frac{\lambda_1\lambda_2}{\alpha}\right)\right) + \lambda_1 \widetilde{F}(z_D)\left(I\left(\lambda_1 \widetilde{F}(z_D) - \frac{\lambda_1\lambda_2}{\alpha}\right) + D_T\right) \\
&- \frac{\lambda_1\lambda_2}{\alpha}\left(I\left(\lambda_1 \widetilde{F}(z_D) - \frac{\lambda_1\lambda_2}{\alpha}\right) + D_T\right) \\
&= -U\left(I\left(\lambda_1 \widetilde{F}(z) - \frac{\lambda_1\lambda_2}{\alpha}\right)\right) + \left(\lambda_1 \widetilde{F}(z_D) - \frac{\lambda_1\lambda_2}{\alpha}\right)\left(I\left(\lambda_1 \widetilde{F}(z_D) - \frac{\lambda_1\lambda_2}{\alpha}\right) + D_T\right).
\end{aligned}
\tag{B22}
$$

Letting $x = I\left(\lambda_1 \widetilde{F}(z_D) - \frac{\lambda_1\lambda_2}{\alpha}\right) + D_T$, the partial derivative (B22) can be writte as

$$
\frac{\partial \mathbb{G}(z_D, \widehat{L}^*)}{\partial z_D} = -U(x - D_T) + U^{'}(x - D_T)x.
$$

Thus, Lemma 3.1 provides that there is a unique root for the function

$$
\frac{\partial \mathbb{G}(z_D, \widehat{L}^*)}{\partial z_D} = 0,
$$

which is $z_D^* = 1 - F_\xi(\frac{U'(\widehat{D}_T - D_T)}{\lambda_1} + \frac{\lambda_2}{\alpha})$. Moreover, Lemma 3.1 also gives that $\partial \mathbb{G}(z_D, \widehat{L}^*)/\partial z_D > 0$ if $z_D < z_D^*$ and $\partial \mathbb{G}(z_D, \widehat{L}^*)/\partial z_D < 0$ if $z_D > z_D^*$. Thus, $\mathbb{G}(z_D^*, \widehat{L}^*)$ is the maximum of the Lagrangian.

If $\widehat{L}^* > \widehat{D}_T$, we have that

$$
z_D^* = 1 - F_\xi\left(\frac{U'(\widehat{D}_T - D_T)}{\lambda_1} + \frac{\lambda_2}{\alpha}\right) < s^* = 1 - F_\xi\left(\frac{U'(\widehat{L}^* - D_T)}{\lambda_1} + \frac{\lambda_2}{\alpha}\right),
$$

because $U'(\cdot)$ is a decreasing function. Thus, $\mathbb{G}(z_D, \widehat{L}^*)$ indeed attains its maximum at $z_D = z_D^*$.

However, if $\widehat{L}^* \le \widehat{D}_T$, we know that $z_D^* > s^*$. In this case, the Lagrangian $\mathbb{G}(z_D, \widehat{L}^*)$ becomes

$$\mathbb{G}(z_D, \widehat{L}^*) = \int_{z^*}^{1} U(Q_1(z) - D_T)\mathrm{d}z + \int_{z_D}^{z^*} U(\widehat{L}^* - D_T)\mathrm{d}z - \lambda_1 \int_{z^*}^{1} Q_1(z)\widetilde{F}(z)\mathrm{d}z$$
$$-\lambda_1 \int_{z_D}^{z^*} \widehat{L}^* \widetilde{F}(z)\mathrm{d}z + \frac{\lambda_1\lambda_2}{\alpha} \int_{z_D}^{\alpha} \widehat{L}^*\mathrm{d}z. \tag{B23}$$

The partial derivative of (B23) is given by

$$\begin{aligned}\frac{\partial \mathbb{G}(z_D, \widehat{L}^*)}{\partial z_D} &= -U(\widehat{L}^* - D_T) + \lambda_1 \widehat{L}^* \widetilde{F}(z_D) - \frac{\lambda_1\lambda_2}{\alpha}\widehat{L}^* \\ &= -U(\widehat{L}^* - D_T) + \lambda_1 \widehat{L}^*(\widetilde{F}(z_D) - \frac{\lambda_2}{\alpha}).\end{aligned}$$

Note that in this case z_D takes value in (s^*, α). Then, we have that

$$\begin{aligned}\lim_{z_D \to s^*} &= \frac{\partial \mathbb{G}(z_D, \widehat{L}^*)}{\partial z_D} \\ &= -U(\widehat{L}^* - D_T) + \lambda_1 \widehat{L}^*(\widetilde{F}(s^*) - \frac{\lambda_2}{\alpha}) = \widehat{L}^* U'(\widehat{L}^* - D_T) - U(\widehat{L}^* - D_T) > 0 \\ \lim_{z_D \to \alpha} &= \frac{\partial \mathbb{G}(z_D, \widehat{L}^*)}{\partial z_D} \\ &= -U(\widehat{L}^* - D_T) + \lambda_1 \widehat{L}^*(\widetilde{F}(\alpha) - \frac{\lambda_2}{\alpha}) \\ &> -U(\widehat{L}^* - D_T) + \lambda_1 \widehat{L}^*(\widetilde{F}(z^*)) = \widehat{L}^* U'(\widehat{L}^* - D_T) - U(\widehat{L}^* - D_T) > 0.\end{aligned}$$

Obviously, $\partial \mathbb{G}(z_D, \widehat{L}^*)/\partial z_D$ is a monotone (decreasing) function of z_D. Hence, the function $\mathbb{G}(z_D, \widehat{L}^*)$ is increasing in z_D. However, to satisfy the AVaR constraint, z_D is at most $z_D^{**} = \alpha(\widehat{L}^* - L)/\widehat{L}^*$.□ □

B.2 Discussion of the budget constraint

In the previous section, we have shown the optimal solutions for the Problem AVaR (1) in different situations assuming they exist. In this section, we give a short discussion of the existence of the optimal solutions.

Note that we always assume that the AVaR constraint is given, i.e., α and L are fixed, and the debt level is given, i.e., D_T is fixed. Correspondingly, the tangent point $\widehat{D}_T$ is a constant (Lemma 3.1). Recall that the benchmark solution is either above the tangent point $\widehat{D}_T$ or is zero. Hence, given the AVaR constraint $\frac{1}{\alpha}\int_0^{\alpha} Q_X^{AVaR}(z)\mathrm{d}z \ge L$, the trivial solution is given by $X_T^{AVaR} = L$. Therefore, if the budget is below Le^{-rT}, the optimal solution does not exist.

Let X_T^B denote the benchmark solution (Proposition 1) with the budget constraint $\mathbb{E}[X_T^B\xi] = x_0^B$. If $\frac{1}{\alpha}\int_0^\alpha Q_B(z)\mathrm{d}z \geq L$, where $Q_B(z)$ is the quantile function of X_T^B, then the AVaR constraint is redundant. It means that the benchmark solution is also the optimal solution under the AVaR constraint. Therefore, the optimal solutions in Proposition 2 exist if the initial budget is $Le^{-rT} < x_0 < x_0^B$.

The optimal solution (19) depends on β^* (Lemma B.1) and z_D^* (Eq B9 and Lemma 3.1). Given the AVaR constraint, we have that $0 < \beta^* < \bar{\beta} := \alpha(D_T - L)/D_T$. Therefore, given a $\beta^* \in (0, \bar{\beta})$, z_D^* can be written as a function of β^* and λ_1 by Eq B9 and Lemma 3.1. The constraint $z_D^* > \alpha$ can be transferred to a constraint on λ_1, which implies a bound of the budget. Let $x_0(\alpha)$ denote the budget if $z_D^* = \alpha$. Then, if $Le^{-rT} < x_0 < x_0(\alpha)$, Eq 19 is the optimal solution for the Problem AVaR (1).

Similarly, Lemma B.3 tells that $\widehat{L}^*$ is a function of λ_1 and λ_2. Let $x_0(\widehat{D}_T)$ denote the budget when $\widehat{L}^* = \widehat{D}_T$. Then, if $x_0(\alpha) < x_0 < x_0(\widehat{D}_T)$, Eq (22) is the optimal solution for the Problem AVaR (1); Otherwise, the optimal solution is Eq (23).

Note that if $L > D_T$, Eq (19) does not exist.

Appendix C Proof of Proposition 4

Proof:

The optimal solution for Problem AVaR (1) has three potential structures depending on the initial budgets.

Case 1: $L^{ES} = \widehat{L}^* < D_T$.

If the equations $\widetilde{F}(z_D^*) = \xi_a,\ \widetilde{F}(\beta^*) = \xi_b$ hold, then the two solutions (27) and (19) are the same. Note that $\xi_b = \widetilde{F}(\alpha(L^{ES} - L)/L^{ES})$ because it is defined by $P(\xi \geq \xi_b) = \alpha(L^{ES} - L)/L^{ES}$. Note that $\beta^* = \alpha(\widehat{L}^* - L)/\widehat{L}^*$. Thus, if $L^{ES} = \widehat{L}^*$, then $\widetilde{F}(\beta^*) = \xi_b$.

Similarly, we have that $\widetilde{F}(z_D^*) = U'(\widetilde{L} - (D_T - \widehat{L}^*))/\lambda_1$ as $\widetilde{L}$ satisfies $U(\widetilde{L} - (D_T - \widehat{L}^*)) - U'(\widetilde{L} - (D_T - \widehat{L}^*))\widetilde{L} = 0$. Notice that $\xi_a = U'(\widetilde{L} - (D_T - L^{ES}))/\lambda_1^{ES}$, with $\widetilde{L}$ satisfying $U((\widetilde{L} - (D_T - L^{ES})) - \widetilde{L}U'(\widetilde{L} - (D_T - L^{ES})) = 0$. Given that $L^{ES} = \widehat{L}^*$, we have $\widetilde{F}(z_D^*) = \xi_a$.

Lemma A.1 gives that the budget function is a monotone function. Hence, $\lambda_1 = \lambda_1^{ES}$.

Case 2: $D_T < L^{ES} = \widehat{L}^* < \widehat{D}_T$.

The solutions (29) and (22) are the same if $\xi_1 = \xi_c,\ \xi_2 = \xi_d$.

Recall that $\xi_2 = \widetilde{F}(\alpha(\widehat{L}^* - D_T)/\widehat{L}^*)$. In addition, we have $\xi_d = \widetilde{F}(\alpha(L^{ES} - L)/L^{ES})$ because ξ_d is defined through $P(\xi \geq \xi_d) = \alpha(L^{ES} - L)/L^{ES}$. Thus, if $\widehat{L}^* = L^{ES}$, then $\xi_2 = \xi_d$.

Moreover, $\xi_1 = U'(\widehat{L}^* - D_T)/\lambda_1$ and $\xi_c = U'(L^{ES} - D_T)/\lambda_1^{ES}$. Hence, if $L^{ES} = \widehat{L}^*$, $\xi_1 = \xi_c$

Case 3: $L^{ES} = \widehat{L}^* > \widehat{D}_T$.

Similar to Case 2, we have $\xi_1 = U'(\widehat{L}^* - D_T)/\lambda_1$, $\xi_2 = (U'(\widehat{L}^* - D_T) + \lambda_2/\alpha)/\lambda_1$ and $\xi_3 = (U'(\widehat{D}_T - D_T) + \lambda_2/\alpha)/\lambda_1$. In addition, we know that $\xi_e = U'(L^{ES} - D_T)/\lambda_1^{ES}$, $\xi_f = (U'(L^{ES} - D_T) + \lambda_2^{ES})/\lambda_1^{ES}$ and $\xi_g = (U'(\widehat{D}_T - D_T) + \lambda_2^{ES})/\lambda_1^{ES}$. Hence, if $L^{ES} = \widehat{L}^*$, the solution (30) is the same as (23).

The above arguments also work if we start from the solutions (27), (29) and (30). It means that given the optimal wealth (27), (29) and (30) under the ES constraint $\mathbb{E}[(L^{ES} - X_T)^+] \leq \alpha(L^{ES} - L)$, we can compute the AVaR of the portfolio that is $\frac{1}{\alpha}\int_0^\alpha VaR_{X_T}(\beta)\mathrm{d}\beta = L$. In another case, if the optimal wealth is (26) or (28), the AVaR of the portfolio is $\frac{1}{\alpha}\int_0^\alpha VaR_{X_T}(\beta)\mathrm{d}\beta > L$ by Lemma 4.1 in Chen, Stadje and Zhang [14]. □

Appendix D Proof of Proposition 5

In the Black Scholes market with two financial assets, the unique state price density ξ_T follows a log normal distribution, i.e.,

$$\xi_T/\xi_0 \sim \mathcal{LN}(-(r + 0.5\theta^2)T, \theta\sqrt{T}), \quad \theta = (\mu - r)/\sigma, \quad \xi_0 = 1. \tag{D24}$$

In addition, the process $X_T^{AVaR}\xi_T$ is a martingale, i.e.,

$$X_t^{AVaR}\xi_t = \mathbb{E}\left[X_T^{AVaR}\xi_T \middle| \mathcal{F}_t\right]. \tag{D25}$$

Based on these two facts, we can express the pre-horizon wealth X_t^{AVaR} as a function of ξ_t.

Next, we explain how to obtain the investment strategy π_t^{AVaR}. Given the investment strategy π_t^{AVaR}, the dynamics of the optimal portfolio can be expressed as

$$\begin{aligned} \mathrm{d}X_t^{AVaR} &= \pi_t^{AVaR}\frac{X_t^{AVaR}}{S_t}\mathrm{d}S_t + (1 - \pi_t^{AVaR})\frac{X_t^{AVaR}}{B_t}\mathrm{d}B_t \\ &= (r + \pi_t^{AVaR}(\mu - r))X_t^{AVaR}\mathrm{d}t + \pi_t^{AVaR}\sigma X_t^{AVaR}\mathrm{d}W_t. \end{aligned} \tag{D26}$$

Moreover, the portfolio X_t^{AVaR} is a Itô process in the Black Scholes market. After we obtain X_t^{AVaR} from Eq (D25), we can compute its derivatives by Itô's lemma. Then, the optimal investment strategies are obtained by comparing the stochastic term of the derivative of X_t^{AVaR} with Eq (D26).