



HAL
open science

Sample Average Approximation for Portfolio Optimization under CVaR constraint in an (re)insurance context

Jérôme Lelong, Véronique Maume-Deschamps, William Thevenot

► **To cite this version:**

Jérôme Lelong, Véronique Maume-Deschamps, William Thevenot. Sample Average Approximation for Portfolio Optimization under CVaR constraint in an (re)insurance context. Computational Optimization and Applications, 2026, <10.1007/s10589-026-00767-3>. <hal-04733015v3>

HAL Id: hal-04733015

<https://hal.science/hal-04733015v3>

Submitted on 7 Feb 2026

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons CC BY-NC-ND 4.0 - Attribution - Non-commercial use - No Derivative Works - International License

Sample Average Approximation for Portfolio Optimization under CVaR constraint in a (re)insurance context

Jérôme Lelong * Véronique Maume-Deschamps †

William Thevenot ‡

February 7, 2026

Abstract

We consider optimal allocation problems with Conditional Value-At-Risk (CVaR) constraint. We prove, under very mild assumptions, the convergence of the Sample Average Approximation method (SAA) applied to this problem, and we also exhibit a convergence rate and discuss the uniqueness of the solution. These results provide (re)insurers with a practical approach to portfolio optimization under regulatory risk constraints.

Keywords: Value-At-Risk, Conditional Value-At-Risk, Expected shortfall, Sample average approximation, Portfolio optimization, Insurance, Reinsurance, Uniform strong large law of numbers, Central limit theorem.

1 Introduction

It is in the interest of (re)insurance companies to reduce risk through diversification in order to enhance their risk–return profile. The approach developed in this paper provides a quantitative framework to identify an ideal mix of risk exposures that balances expected profit with the Solvency Capital Requirement (SCR), which is represented in our model by the Conditional Value-at-Risk (CVaR). In the European market, insurers are subject to Solvency II regulations, which require that they hold an amount of eligible own funds at least equal to the SCR. This capital buffer is defined as the amount needed to ensure that the (re)insurance company can meet its obligations over the next 12 months with a probability of at least 99.5%. While the SCR is formally based on the Value-at-Risk (VaR) at the confidence level $\alpha = 0.995$, the Conditional Value-at-Risk (CVaR) is commonly used as a more tractable and conservative proxy. Accordingly, we adopt CVaR as the risk measure in our analysis.

The classic approach to portfolio optimization was introduced by Markowitz in 1952 [3]. It consists in maximizing the expected return of the portfolio subject to a constraint on its variance, or, equivalently, in minimizing the variance for a fixed level of expected return. This framework is known as the mean–variance optimization. The corresponding formulation with Conditional Value-at-Risk (CVaR) as risk measure is referred to as the mean–CVaR optimization.

*Univ. Grenoble Alpes, CNRS, Grenoble INP, LJK, 38000 Grenoble, France. jerome.lelong@univ-grenoble-alpes.fr

†Université Claude Bernard Lyon 1, CNRS, Ecole Centrale de Lyon, INSA Lyon, Université Jean Monnet, ICJ UMR5208, 69622 Villeurbanne, France. veronique.maume-deschamps@univ-lyon1.fr

‡Université Claude Bernard Lyon 1, CNRS, Ecole Centrale de Lyon, INSA Lyon, Université Jean Monnet, ICJ UMR5208, 69622 Villeurbanne, France. and Risk Knowledge team at SCOR SE, Paris, France thevenot.william@hotmail.com

We model the (re)insurance asset market with business lines represented by the random vector \mathbf{X} of asset returns, taking values in a subset $\mathcal{R}_{\mathbf{X}}$ of \mathbb{R}^d . Unlike the classical financial case, \mathbf{X} cannot take negative values because it represents the results of the business lines. We assume that $\mathbb{E}(|\mathbf{X}|) < +\infty$. A (re)insurance portfolio is defined by a vector $\boldsymbol{\gamma} \in \mathbb{R}^d$ representing the quantity held in each business line by the (re)insurer. In our business case, it cannot be interpreted in the same way as in the classical finance formulation. There is no initial wealth and $\boldsymbol{\gamma}$ therefore does not represent the repartition of it. Vector $\boldsymbol{\gamma}$ represents the (re)insurance company activity.

Let us fix some notations, with $\alpha \in]0, 1[$:

$$V_{\alpha}(\boldsymbol{\gamma}) = VaR_{\alpha}(-\boldsymbol{\gamma}^T \mathbf{X}) = \min \left\{ M \in \mathbb{R} : \mathbb{P} \left(-\boldsymbol{\gamma}^T \mathbf{X} \leq M \right) \geq \alpha \right\},$$

$$C_{\alpha}(\boldsymbol{\gamma}) = CVaR_{\alpha}(-\boldsymbol{\gamma}^T \mathbf{X}) = \frac{1}{1-\alpha} \int_{\alpha}^1 V_{\beta}(\boldsymbol{\gamma}) d\beta.$$

Our original goal is to solve the following problem with a fixed $\alpha \in]0, 1[$ and constraints on the weights, together with a capital-requirement limit $K > 0$. We consider a loss function L_0 , which commonly depends on $C_{\alpha}(\boldsymbol{\gamma})$. Even though $L_0(\boldsymbol{\gamma}, \mathbf{X})$ may already account for this dependency implicitly, it is convenient to make the dependence on $C_{\alpha}(\boldsymbol{\gamma})$ explicit for the purposes of future analysis.

$$\begin{aligned} v^* &:= \inf_{\boldsymbol{\gamma} \in \mathbb{R}_+^d} \mathbb{E}(L_0(\boldsymbol{\gamma}, C_{\alpha}(\boldsymbol{\gamma}), \mathbf{X})) \\ \text{s.t.} \quad &\gamma_i^{\text{low}} \leq \gamma_i \leq \gamma_i^{\text{up}} \quad \forall i \in \{1, \dots, d\} \\ \text{s.t.} \quad &C_{\alpha}(\boldsymbol{\gamma}) \leq K. \end{aligned} \tag{1.1}$$

Through reinsurance and other risk transfer solutions, (re)insurers can manage their exposure to individual risks and improve overall diversification, thereby increasing the return per unit of risk. Achieving this objective requires identifying an ideal risk composition, both at a given moment and over time through underwriting and portfolio management. Reinsurers, in particular, steer his portfolio by adjusting the reinsurance shares they offer during tenders, according to their risk appetite and renewal strategy. This strategic allocation process, which does not rely on the internal structure of each business line, is formalized in our model through the minimization of a loss function, or equivalently the maximization of a return function, under the CVaR constraints that reflect regulatory capital requirements. The allocation vector $\boldsymbol{\gamma}$ may also implicitly account for market size, legacy positions, and underwriting partnerships, while supporting the design of multi-period strategies aimed at gradually reaching a target risk exposure.

The standard formulation of CVaR is inconvenient because it requires computing VaR followed by a conditional expectation. An alternative approach was introduced by R.T. Rockafellar and S. Uryasev in 2000 [4], and later extended by P. Krokmal, J. Palmquist, and S. Uryasev in 2002 [2]. They proposed an embedding technique that reformulates the CVaR as a minimization problem involving an auxiliary parameter, which can be integrated into the global optimization framework. This reformulation circumvents the complexity of the original approach and significantly reduces computational burden.

When the distribution of the portfolio payoff is accessed through a finite sample, a common approach in the literature consists in constructing a finite-dimensional empirical optimization problem based on the available scenarios. In the presence of CVaR-type risk measures or constraints, this empirical formulation often admits a linear programming representation through the Rockafellar–Uryasev reformulation, which can then be addressed using standard LP solvers such as simplex or interior-point methods; see, e.g., [1]. More generally, depending on the structure of the objective function, the resulting finite-dimensional problem may also be solved using first-order optimization methods.

In this work, we adopt the sample average approximation (SAA) framework [5] as a modeling device to construct a finite-dimensional optimization problem from empirical data. It is important to emphasize that the sample average approximation is not an optimization algorithm, but rather a procedure for approximating the original stochastic problem by an empirical counterpart. While SAA is most often used as a purely numerical tool, the originality of our contribution lies in a detailed theoretical analysis of the optimal solutions generated by the SAA problem within a risk-constrained portfolio optimization framework.

For this formulation with explicit constraints, no results on convergence or convergence rates of the SAA method have been published, to the best of our knowledge. The closest work is [10], where, however, the function to be minimized does not depend on the data sample. Two other related contributions are [8] and [11], which employ a parametric approach: they first identify a data structure through parameterization and then minimize over the resulting parameters. Such methods can be advantageous with small sample sizes, but this setting does not apply to our case. Moreover, the formulations in these works differ from ours, leading to distinct challenges in solving the problem. Finally, we emphasize that SAA is non-parametric and does not require prior knowledge of the underlying distribution of \mathbf{X} , it only assumes access to samples from \mathbf{X} .

In a related direction, Trindade, Uryasev, Shapiro, and Zrazhevsky (2007) [9] studied the asymptotic behavior of empirical CVaR estimators and established the consistency of CVaR-constrained regression, showing that the limiting properties depend on whether the CVaR constraint is active.

Under assumptions of convexity, continuity, and integrability, which will be introduced progressively, we prove almost sure convergence and find a rate of convergence for the SAA version in the case where the function to be minimized depends on the data sample as do the constraint. Moreover, if CVaR appears in the function to be minimized, we show that, under a monotonicity assumption, it can be replaced by the auxiliary function introduced in [2] and [4]. We also provide a sufficient condition for the uniqueness of the solution.

The paper is organized as follows. Section 2 describes the context and states the problem. Section 2.2 introduces the SAA method and presents our main results on convergence (Theorem 3.8), uniqueness (Theorems 3.11 and 3.10) and convergence rate (Theorem 3.18). Section 4 is dedicated to some numerical studies on two different cases, where we show how these results can be used concretely in the context of a (re)insurer seeking to balance its various exposures.

2 Optimization with CVaR constraint

This section presents the transformation of the original problem and the application of the SAA method to its equivalent formulation.

For any real number x , we write $x^+ := \max(x, 0)$ for its positive part.

2.1 An equivalent setting

Note that the Conditional Value-At-Risk can be rewritten using the Value-At-Risk as follows:

$$C_\alpha(\boldsymbol{\gamma}) = V_\alpha(\boldsymbol{\gamma}) + (1 - \alpha)^{-1} \mathbb{E} \left((-\boldsymbol{\gamma}^T \mathbf{X} - V_\alpha(\boldsymbol{\gamma}))^+ \right).$$

We introduce $g(\boldsymbol{\gamma}, \zeta) = \zeta + (1 - \alpha)^{-1} \mathbb{E} \left((-\boldsymbol{\gamma}^T \mathbf{X} - \zeta)^+ \right)$ for any $\zeta \in \mathbb{R}$.

Rockafellar, R. T., Uryasev, S. (2000) [4] showed that $g(\boldsymbol{\gamma}, \cdot)$ is convex for any $\boldsymbol{\gamma}$, $C_\alpha(\boldsymbol{\gamma})$ can be determined by minimising $g(\boldsymbol{\gamma}, \cdot)$:

$$\begin{aligned} C_\alpha(\boldsymbol{\gamma}) &= \min_{\zeta \in \mathbb{R}} g(\boldsymbol{\gamma}, \zeta), \quad A_\alpha(\boldsymbol{\gamma}) = \operatorname{argmin}_{\zeta \in \mathbb{R}} g(\boldsymbol{\gamma}, \zeta), \\ V_\alpha(\boldsymbol{\gamma}) &= \min(\{x \in A_\alpha(\boldsymbol{\gamma})\}), \quad C_\alpha(\boldsymbol{\gamma}) = g(\boldsymbol{\gamma}, V_\alpha(\boldsymbol{\gamma})). \end{aligned}$$

Krokhmal P., Jonas Palmquist J., Uryasev S. (2002) [2] showed the equivalence of the following two optimization problems in the sense that their objective functions attain the same minimum value. (Proof in Annex 1)

Proposition 2.1. *For any function $R : \mathbb{R}^d \times \mathbb{R} \rightarrow \mathbb{R}$ such that for any $\boldsymbol{\gamma} \in \mathbb{R}^d$, $R(\boldsymbol{\gamma}, \cdot)$ is non-increasing then the following two problems are equivalent*

$$\inf_{\boldsymbol{\gamma} \in \Xi} -R(\boldsymbol{\gamma}, C_\alpha(\boldsymbol{\gamma})) \quad \text{s.t.} \quad C_\alpha(\boldsymbol{\gamma}) \leq K \quad (2.1)$$

and

$$\inf_{(\boldsymbol{\gamma}, \zeta) \in \Xi \times \mathbb{R}} -R(\boldsymbol{\gamma}, g(\boldsymbol{\gamma}, \zeta)) \quad \text{s.t.} \quad g(\boldsymbol{\gamma}, \zeta) \leq K. \quad (2.2)$$

Moreover, if the C_α constraint is active in (2.1), then $(\boldsymbol{\gamma}^*, \zeta^*)$ solves (2.2) if and only if $\boldsymbol{\gamma}^*$ achieves the minimum of (2.1) and $\zeta^* \in A_\alpha(\boldsymbol{\gamma}^*)$.

It is convenient to write $L(\boldsymbol{\gamma}, \zeta, \mathbf{X}) = L_0(\boldsymbol{\gamma}, g(\boldsymbol{\gamma}, \zeta), \mathbf{X})$ for future analysis.

Note that g is also continuous and convex.

This result is key to simplifying our problem: we no longer need to compute $C_\alpha(\boldsymbol{\gamma})$, which requires first evaluating $V_\alpha(\boldsymbol{\gamma})$ and is computationally expensive. Instead, we compute $g(\boldsymbol{\gamma}, \zeta)$ which does not involve solving an optimisation problem. Using the equivalence between (2.1) and (2.2), (1.1) may be written as:

$$\begin{aligned} v^* &:= \inf_{\boldsymbol{\gamma}, \zeta \in \mathbb{R}_+^d \times \mathbb{R}} \mathbb{E}(L(\boldsymbol{\gamma}, \zeta, \mathbf{X})) \\ &\text{s.t.} \quad \gamma_i^{\text{low}} \leq \gamma_i \leq \gamma_i^{\text{up}} \quad \forall i \\ &\text{s.t.} \quad \zeta^{\text{low}} \leq \zeta \leq \zeta^{\text{up}} \\ &\text{s.t.} \quad g(\boldsymbol{\gamma}, \zeta) \leq K \end{aligned} \quad (2.3)$$

where

$$\begin{aligned} \zeta^{\text{low}} &\leq \min_{\gamma_i^{\text{low}} \leq \gamma_i \leq \gamma_i^{\text{up}}, \forall i} \text{VaR}_\alpha(-\boldsymbol{\gamma}^\top \mathbf{X}), \\ \zeta^{\text{up}} &\geq \max_{\gamma_i^{\text{low}} \leq \gamma_i \leq \gamma_i^{\text{up}}, \forall i} \text{VaR}_\alpha(-\boldsymbol{\gamma}^\top \mathbf{X}). \end{aligned}$$

Remark 2.2. The most commonly used loss function is $L_0(\boldsymbol{\gamma}, \mathbf{X}) = -\boldsymbol{\gamma}^\top \mathbf{X}$. A more realistic formulation includes the cost of capital in insurance, it is usual to penalize the function with a term proportional to the capital requirement i.e. the CVaR exceeding a threshold $0 < C_0 < K$. So the loss could become $L_0(\boldsymbol{\gamma}, \mathbf{X}) = -\boldsymbol{\gamma}^\top \mathbf{X} + c \cdot (C_\alpha(\boldsymbol{\gamma}) - C_0)^+ = L_0(\boldsymbol{\gamma}, C_\alpha(\boldsymbol{\gamma}), \mathbf{X})$ with $c \in]0, 1[$ where c typically equals 0.05. In this formulation, the loss becomes $L(\boldsymbol{\gamma}, \zeta, \mathbf{X}) = -\boldsymbol{\gamma}^\top \mathbf{X} + c \cdot (g(\boldsymbol{\gamma}, \zeta) - C_0)^+$ as in Problem (2.3). The positive-part function could also be replaced by a more sophisticated nonlinear function.

2.2 Sample Average Approximation

Since $\mathbb{E}(L(\boldsymbol{\gamma}, \zeta, \mathbf{X}))$, is generally not available, we rely on a real or simulated sample of \mathbf{X} using the Sample Average Approximation (SAA) method [6]. Let $(\mathbf{X}^{(j)})_{j \geq 1}$ be a sequence of iid random variables with the same distribution as \mathbf{X} . We consider an approximation of Problem (2.3). For $N \geq 1$, consider

$$g_N(\boldsymbol{\gamma}, \zeta) := \zeta + \frac{(1-\alpha)^{-1}}{N} \sum_{j=1}^N (-\boldsymbol{\gamma}^\top \mathbf{X}^{(j)} - \zeta)^+$$

and

$$\ell_N(\boldsymbol{\gamma}, \zeta) := \frac{1}{N} \sum_{j=1}^N L(\boldsymbol{\gamma}, \zeta, \mathbf{X}^{(j)}).$$

The sample average approximation of (2.3) is given by

$$\begin{aligned}
v_N &:= \inf_{\boldsymbol{\gamma}, \zeta \in \mathbb{R}_+^d \times \mathbb{R}} \ell_N(\boldsymbol{\gamma}, \zeta) \\
\text{s.t.} \quad & 0 \leq \gamma_i^{\text{low}} \leq \gamma_i \leq \gamma_i^{\text{up}} \quad \forall i \\
& \zeta^{\text{low}} \leq \zeta \leq \zeta^{\text{up}} \\
& g_N(\boldsymbol{\gamma}, \zeta) \leq K.
\end{aligned} \tag{2.4}$$

Let us introduce $\Gamma := \{\boldsymbol{\gamma} : 0 \leq \gamma_i^{\text{low}} \leq \gamma_i \leq \gamma_i^{\text{up}} \forall i\}$ and $\mathcal{U} := \Gamma \times \{\zeta : \zeta^{\text{low}} \leq \zeta \leq \zeta^{\text{up}}\}$.

Define $\ell : (\boldsymbol{\gamma}, \zeta) \mapsto \mathbb{E}(L(\boldsymbol{\gamma}, \zeta, \mathbf{X}))$, we assume the following properties

(P1) $\mathbb{P}((\boldsymbol{\gamma}, \zeta) \rightarrow L(\boldsymbol{\gamma}, \zeta, \mathbf{X}) \text{ is continuous and convex}) = 1$.

(P2) $\mathbb{E} \left(\sup_{(\boldsymbol{\gamma}, \zeta) \in \mathcal{U}} |L(\boldsymbol{\gamma}, \zeta, \mathbf{X})| \right) < +\infty$.

The functions ℓ_N and ℓ are continuous and convex by (P1) and (P2). Note that g_N and g are convex and continuous.

Now we define the feasibility sets of the original problem and of its SAA counterpart; $\mathcal{U}^K := \{(\boldsymbol{\gamma}, \zeta) : g(\boldsymbol{\gamma}, \zeta) \leq K\} \cap \mathcal{U}$ and $\mathcal{U}_N^K := \{(\boldsymbol{\gamma}, \zeta) : g_N(\boldsymbol{\gamma}, \zeta) \leq K\} \cap \mathcal{U}$. We assume that $\mathcal{U}, \mathcal{U}^K$ have non-empty interiors.

We also define the solution sets of the original problem and of its SAA counterpart.

$$S := \underset{(\boldsymbol{\gamma}, \zeta) \in \mathcal{U}^K}{\operatorname{argmin}} \ell(\boldsymbol{\gamma}, \zeta) ; S_N := \underset{(\boldsymbol{\gamma}, \zeta) \in \mathcal{U}_N^K}{\operatorname{argmin}} \ell_N(\boldsymbol{\gamma}, \zeta),$$

Our goal is to prove that v_N converges to v^* a.s. and $\mathbb{D}(S_N, S) \xrightarrow[N \rightarrow \infty]{} 0$ a.s. with $\mathbb{D}(A, B) := \sup_{x \in A, x' \in B} \|x - x'\|$.

3 Main results

3.1 Convergence of the SAA method

We rely on two results from Shapiro et al. (2009) [7], to establish almost sure convergence of our SAA method (Theorem 3.4/ Theorem 3.8).

Theorem 3.1 (Theorem 7.48 p375 in [7]).

Consider a random function $H : \Xi \times \mathcal{R}_X \rightarrow \mathbb{R}$, where Ξ is a non-empty compact subset of \mathbb{R}^d . Define $h(\boldsymbol{\xi}) := \mathbb{E}(H(\boldsymbol{\xi}, \mathbf{X}))$ and $h_N(\boldsymbol{\xi}) := N^{-1} \sum_{j=1}^N H(\boldsymbol{\xi}, \mathbf{X}^{(j)})$. Suppose that:

(i) $\mathbb{P}(\boldsymbol{\xi} \rightarrow H(\boldsymbol{\xi}, \mathbf{X}) \text{ is continuous}) = 1$.

(ii) $\mathbb{E} \left(\sup_{\boldsymbol{\xi} \in \Xi} |H(\boldsymbol{\xi}, \mathbf{X})| \right) < +\infty$.

Then, the function h is continuous on Ξ , and $(h_N)_{N \in \mathbb{N}^*}$ converges to h a.s. uniformly on Ξ i.e.

$$\sup_{\boldsymbol{\xi} \in \Xi} |h_N(\boldsymbol{\xi}) - h(\boldsymbol{\xi})| \rightarrow 0 \text{ a.s. as } N \rightarrow \infty.$$

We use Theorem 3.1 with the functions L and ℓ_N on the compact set Γ using (P1) and (P2) to obtain the following proposition.

Proposition 3.2. *If (P1) and (P2) hold, then, the function ℓ is continuous and the sequence $(\ell_N)_{N \in \mathbb{N}^*}$ converges to ℓ a.s. uniformly on Γ .*

And we use Theorem 3.1 again with $G((\boldsymbol{\gamma}, \zeta), \mathbf{X}) = \zeta + (1 - \alpha)^{-1}(-\boldsymbol{\gamma}^T \mathbf{X} - \zeta)^+$ and g_N on the compact set \mathcal{U} to obtain the following proposition.

Proposition 3.3. *The sequence $(g_N)_{N \in \mathbb{N}^*}$ converges to g a.s. uniformly on \mathcal{U} .*

The set \mathcal{U} is compact and \mathcal{U}^K is non-empty by assumption, so neither is S . The function ℓ is continuous and convex, by (P1) and (P2), and $(\ell_N)_{N \in \mathbb{N}^*}$ converges to ℓ a.s. uniformly on \mathcal{U} from Proposition 3.2 with (P1) and (P2). Thanks to these results, we can state a modified version of Theorem 5.5 p160 in [7] adapted to our framework.

Theorem 3.4 (Adapted from Theorem 5.5 p160 in [7]).

If the following conditions hold:

- (a) *Let $(u_N)_N$ be a sequence taking values in \mathcal{U} and such that u_N converges to \bar{u} a.s. If $u_N \in \mathcal{U}_N^K$ for all N , then $\bar{u} \in \mathcal{U}^K$.*
- (b) *For $u \in S$, there exists a sequence $(u_N)_{N \in \mathbb{N}^*}$ such that $u_N \in \mathcal{U}_N^K$ for all N and $u_N \xrightarrow{N \rightarrow \infty} u$ a.s.*
- (c) *A.s. for N large enough the set S_N is nonempty and $S_N \subset \mathcal{U}$.*

Then $v_N \rightarrow v^$ a.s. and $\mathbb{D}(S_N, S) \rightarrow 0$ a.s. when $N \rightarrow \infty$.*

Proposition 3.5. *Let $(u_N)_N$ be a sequence taking values in \mathcal{U} and such that u_N converges to \bar{u} a.s. If $u_N \in \mathcal{U}_N^K$ for all N , then $\bar{u} \in \mathcal{U}^K$.*

Proof. Let $(u_N)_{N \in \mathbb{N}^*}$ be a sequence taking values in \mathcal{U} such that u_N converges to \bar{u} a.s. and such that $u_N \in \mathcal{U}_N^K$ for all N . We have

$$g_N(u_N) \leq K. \quad (3.1)$$

By Proposition 3.3, $(g_N)_{N \in \mathbb{N}^*}$ converges to g uniformly on \mathcal{U} , so $(g_N(u_N))_{N \in \mathbb{N}^*}$ converges to $g(\bar{u})$ a.s. and we deduce from (3.1), that $g(\bar{u}) \leq K$ so $(\bar{u}) \in \mathcal{U}^K$. ■

Lemma 3.6. *If the interior of \mathcal{U}^K is non-empty, there exists $\bar{u} \in \mathcal{U}^K$ and $N' \in \mathbb{N}^*$ such that $g_N(\bar{u}) < K$ for all $N > N'$.*

Proof. By assumption, \mathcal{U}^K has a non-empty interior, so that there exists \bar{u} such that $g(\bar{u}) < K$. We apply Proposition 3.3 to prove the a.s. uniform convergence of $(g_N)_{N \in \mathbb{N}^*}$ to g on \mathcal{U} . Then, for $\varepsilon = K - g(\bar{u}) > 0$, there exists N' such that $|g_N(\bar{u}) - g(\bar{u})| < \varepsilon$ for all $N > N'$. Therefore, $g_N(\bar{u}) < K$. ■

Proposition 3.7. *For any $u \in S$, there exists a sequence $(u_N)_{N \in \mathbb{N}^*}$ such that $(u_N) \in \mathcal{U}_N^K$ for all $N \in \mathbb{N}^*$ and $u_N \xrightarrow{N \rightarrow \infty} u$ a.s.*

Proof. It suffices to prove this for all $u \in \mathcal{U}^K$ since $S \subset \mathcal{U}^K$. Let $u^* \in \mathcal{U}^K$. We construct a sequence $u_N \in \mathcal{U}_N^K$ which converges to u^* a.s. Since the interior of \mathcal{U}^K is non-empty, by Lemma 3.6, there exists $\bar{u} = (\bar{\boldsymbol{\gamma}}, \bar{\zeta}) \in \bigcap_{N > N'} \mathcal{U}_N^K$ for some $N' > 0$. For $\lambda \in [0, 1]$, we define $\theta_N(\lambda) := g_N(\lambda \bar{u} + (1 - \lambda)u^*)$ and $\theta(\lambda) := g(\lambda \bar{u} + (1 - \lambda)u^*)$. Note that $\theta_N(1) = g_N(\bar{u}) < K$ and $g(u^*) \leq K$.

Let us define $\lambda_N = \min\{\lambda \in [0, 1]; \theta_N(\lambda) \leq K\}$ and $u_N = \lambda_N \bar{u} + (1 - \lambda_N)u^*$, $u_N \in \mathcal{U}_N^K$.

Case 1: $g(u^*) < K$: For any $\varepsilon > 0$, by Proposition 3.3 there exists N' such that for all $N > N'$ we have $\sup_{\mathcal{U}} |g_N - g| \leq \varepsilon$ so $\theta_N(0) = g_N(u^*) \leq K - \varepsilon$. We deduce that for N large enough $\lambda_N = 0$ i.e. $u_N = u^*$ so that $u_N \rightarrow u^*$ a.s.

Case 2: $g(u^*) = K$: Note that the function θ_N is convex because g_N is convex. So that $\theta_N(\lambda_N) \leq \frac{\theta_N(0) + \lambda_N(\theta_N(1) - \theta_N(0))}{\lambda_N}$. As λ_N is bounded, we may extract a convergent sub-sequence $\lambda_{\phi(N)} \rightarrow \bar{\lambda}$ a.s. We assume that $\bar{\lambda} > 0$ and we use a proof by contradiction. We have:

$$\theta_{\phi(N)}(\lambda_{\phi(N)}) \leq \theta_{\phi(N)}(0) + \lambda_N(\theta_{\phi(N)}(1) - \theta_{\phi(N)}(0))$$

taking the limit leads to

$$\theta(\bar{\lambda}) - \theta(0) \leq \bar{\lambda}(\theta(1) - \theta(0)).$$

Now $\theta(0) = K$ and $\theta(1) < K$. Therefore $\theta(\bar{\lambda}) - K < 0$ which we rewrite as $\theta(\bar{\lambda}) \leq K - 3\varepsilon$ for some $\varepsilon > 0$. By continuity of θ , there exists $\delta > 0$ such that $\theta(\bar{\lambda} - \delta) \leq K - 2\varepsilon$. Note that $\sup_{\lambda} |\theta_N(\lambda) - \theta(\lambda)| \xrightarrow{N \rightarrow \infty} 0$ a.s. because of Proposition 3.3. Then, for N large enough $|\theta_N(\lambda) - \theta(\lambda)| \leq \varepsilon$ for all $\lambda \in [0, 1]$, so that $\theta_N(\bar{\lambda} - \delta) \leq K - \varepsilon$. Thus, $\lambda_N \leq \bar{\lambda} - \delta$. This contradicts $\lambda_{\phi(N)} \rightarrow \bar{\lambda}$ a.s. Hence, $(\lambda_N)_{N \in \mathbb{N}^*}$ must converge to 0. We conclude that $\lambda_N \rightarrow 0$ a.s., hence $u_N \rightarrow u^*$ a.s. which completes the proof. ■

Theorem 3.8. *Let v_N define by (1.1) and v^* define by (2.4). If (P1) and (P2) hold, then*

$$\begin{aligned} v_N & \xrightarrow[N \rightarrow \infty]{a.s.} v^* \\ \mathbb{D}(S_N, S) & \xrightarrow[N \rightarrow \infty]{a.s.} 0. \end{aligned} \tag{3.2}$$

Proof. We want to apply Theorem 3.4.

Assumption (a) holds thanks to Proposition 3.5 and Assumption (b) holds thanks to Proposition 3.7.

\mathcal{U}^K has a non-empty interior. We can apply Lemma 3.6 and find $\bar{u} \in \mathcal{U}^K$ and $N' \in \mathbb{N}^*$ such that $g_N(\bar{u}) < K$ for all $N > N'$. So that for $N > N'$, $\bar{u} \in \mathcal{U}_N^K$, then \mathcal{U}_N^K is non-empty. Finally, v_N is well defined. Since \mathcal{U}_N^K is closed, we have $S_N \subset \mathcal{U}_N^K \subset \mathcal{U}$. Hence, assumption (c) holds.

We can therefore apply Theorem 3.4 and conclude the proof. ■

3.2 Uniqueness Condition

Let us focus on the uniqueness of the solution to (1.1) and therefore (2.3).

Let us define $\ell_0 : \gamma \mapsto \mathbb{E}(L_0(\gamma, \mathbf{X}))$. We shall make the following assumptions:

(P3) $\forall \gamma \in \mathbb{R}^d \setminus \{0\}$, $\gamma^T \mathbf{X}$ has a density on \mathbb{R}^d .

(P4) $C_\alpha(\gamma) > K, \forall \gamma \in \arg \min_{\Gamma} \ell_0$.

(P5) For all $\gamma \in \Gamma \setminus \{0_d\}$, $\text{supp}(\gamma^T \mathbf{X})$ is connected, i.e., an interval.

Remark 3.9. This setting is realistic, as capital naturally acts as a limiting constraint. Moreover, the support of $\gamma^T \mathbf{X}$ is typically an interval, since aggregated losses across business lines generally have connected supports. Rare risk types with disconnected support, such as those related to natural disasters, are usually smoothed out by other continuous components. Assumption (P3) is satisfied when \mathbf{X} admits a density, and (P4) ensures that the capital constraint is active at the optimum.

Note that we exclude the case where 0_d is an optimal solution because we choose a positive capital requirement limit K .

Assumption (P2) implies $\mathbb{E} \left(\sup_{\gamma \in \Gamma} |L_0(\gamma, \mathbf{X})| \right) < +\infty$ and assumption (P1) implies

$\mathbb{P}(\gamma \rightarrow L_0(\gamma, \mathbf{X}) \text{ is continuous and convex}) = 1$ since $L_0(\gamma, \mathbf{X}) = L(\gamma, \tilde{\zeta}, \mathbf{X})$ with $\tilde{\zeta} \in A_\alpha(\gamma)$. Therefore, function ℓ_0 is continuous and convex thanks to (P1) and (P2).

Rockafellar and Uryasev (2000) [4] showed that, under assumption (P3), the function $g(\boldsymbol{\gamma}, \cdot)$ is continuously differentiable for any $\boldsymbol{\gamma}$.

We are now in a position to state our main results. Theorem 3.10 establishes the uniqueness of the solution to the original problem (1.1), while Theorem 3.11 extends this result to the equivalent formulation (2.3).

Generally, the presence of a constraint does not guarantee uniqueness of the solution. However, in our setting, the strict convexity of the constraint function plays a key role in ensuring uniqueness. A crucial step in the analysis is to verify that the capital constraint is active at the optimum, allowing this strict convexity to be exploited. This is established through Lemma 3.12 and Proposition 3.13.

Rather than proving Theorems 3.10 and 3.11 directly, we first present and prove intermediate results from which they follow.

Theorem 3.10. *If (P1) to (P4) hold, then Problem (1.1) has a unique solution $\boldsymbol{\gamma}^*$.*

Theorem 3.11. *If (P1) to (P5) hold, then Problem (2.3) has a unique solution $(\boldsymbol{\gamma}^*, \zeta^*)$ and in addition $\boldsymbol{\gamma}^*$ is the unique solution to Problem (1.1).*

We now present the intermediate results used in the proofs.

Lemma 3.12. *Let $\boldsymbol{\gamma}^*$ be an optimal solution to (1.1). If assumptions (P1), (P2), and (P4) hold, then $C_\alpha(\boldsymbol{\gamma}^*) = K$.*

Proof. Suppose, for the sake of contradiction, that $C_\alpha(\boldsymbol{\gamma}^*) < K$. Let $\tilde{\boldsymbol{\gamma}} \in \arg \min_{\Gamma} \ell_0$.

Define

$$\tilde{\lambda} = \max \{ \lambda \in [0, 1] : C_\alpha(\lambda \tilde{\boldsymbol{\gamma}} + (1 - \lambda) \boldsymbol{\gamma}^*) \leq K \}.$$

By continuity of C_α , we have $\tilde{\lambda} < 1$, since $C_\alpha(\tilde{\boldsymbol{\gamma}}) > K$ by (P4), and $C_\alpha(\boldsymbol{\gamma}^*) < K$ by assumption.

Using the convexity of ℓ_0 (from assumptions (P1) and (P2)), we obtain:

$$\ell_0(\tilde{\lambda} \tilde{\boldsymbol{\gamma}} + (1 - \tilde{\lambda}) \boldsymbol{\gamma}^*) \leq \tilde{\lambda} \ell_0(\tilde{\boldsymbol{\gamma}}) + (1 - \tilde{\lambda}) \ell_0(\boldsymbol{\gamma}^*) < \ell_0(\boldsymbol{\gamma}^*),$$

where the strict inequality follows from the fact that $\boldsymbol{\gamma}^* \notin \arg \min_{\Gamma} \ell_0$ (by (P4)). This contradicts the optimality of $\boldsymbol{\gamma}^*$. Hence, we conclude that $C_\alpha(\boldsymbol{\gamma}^*) = K$. \blacksquare

Proposition 3.13. *Under assumption (P3), C_α is strictly convex on the unit sphere in \mathbb{R}^d , i.e., strictly convex up to positive scaling. In contrast, it is linear along rays.*

The proof of Proposition 3.13 is provided in the appendix.

We are now in a position to prove Theorem 3.10.

Proof of Theorem 3.10. Let $\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2$ be two optimal solutions of Problem (1.1). By Lemma 3.12, both satisfy the capital constraint with equality:

$$\ell_0(\boldsymbol{\gamma}_1) = \ell_0(\boldsymbol{\gamma}_2) = v^*, \quad C_\alpha(\boldsymbol{\gamma}_1) = C_\alpha(\boldsymbol{\gamma}_2) = K. \quad (3.3)$$

Let $\boldsymbol{\gamma}_\mu := \mu \boldsymbol{\gamma}_1 + (1 - \mu) \boldsymbol{\gamma}_2$ for some $\mu \in (0, 1)$. Since C_α is convex (by sub-additivity of CVaR), we have:

$$C_\alpha(\boldsymbol{\gamma}_\mu) \leq \mu C_\alpha(\boldsymbol{\gamma}_1) + (1 - \mu) C_\alpha(\boldsymbol{\gamma}_2) = K.$$

Similarly, convexity of ℓ_0 (by assumptions (P1) and (P2)) gives:

$$\ell_0(\boldsymbol{\gamma}_\mu) \leq \mu \ell_0(\boldsymbol{\gamma}_1) + (1 - \mu) \ell_0(\boldsymbol{\gamma}_2) = v^*.$$

If $\ell_0(\boldsymbol{\gamma}_\mu) < v^*$, this contradicts the optimality of $\boldsymbol{\gamma}_1$ and $\boldsymbol{\gamma}_2$. Thus, $\ell_0(\boldsymbol{\gamma}_\mu) = v^*$, and $\boldsymbol{\gamma}_\mu$ is optimal. Applying Lemma 3.12 again yields $C_\alpha(\boldsymbol{\gamma}_\mu) = K$.

Therefore, for all $\mu \in [0, 1]$, we have:

$$C_\alpha(\boldsymbol{\gamma}_\mu) = \mu C_\alpha(\boldsymbol{\gamma}_1) + (1 - \mu)C_\alpha(\boldsymbol{\gamma}_2),$$

which implies equality in the convexity inequality. By Proposition 3.13, C_α is strictly convex on the unit sphere. Hence, if both $\boldsymbol{\gamma}_1$ and $\boldsymbol{\gamma}_2$ are optimal, they must lie on the same ray, i.e., $\boldsymbol{\gamma}_1 = a \boldsymbol{\gamma}_2$ for some $a > 0$. Condition (3.3) then implies $a = 1$, so the solution to Problem (1.1) is unique. \blacksquare

Proposition 3.14. *For any fixed $\boldsymbol{\gamma} \neq 0_d$, if (P3) and (P5) holds, the function $\zeta \mapsto g(\boldsymbol{\gamma}, \zeta)$ is strictly convex on $\text{supp}(-\boldsymbol{\gamma}^T \mathbf{X})$.*

Proof. Consider the derivative $\frac{\partial g}{\partial \zeta}(\boldsymbol{\gamma}, \zeta) = 1 - (1 - \alpha)^{-1} \mathbb{E}(-\mathbb{1}_{\{-\boldsymbol{\gamma}^T \mathbf{X} - \zeta > 0\}}) = 1 - (1 - \alpha)^{-1} \mathbb{P}(-\boldsymbol{\gamma}^T \mathbf{X} > \zeta)$. Let us $\zeta_1 < \zeta_2 \in \text{supp}(-\boldsymbol{\gamma}^T \mathbf{X})$

$$\begin{aligned} \frac{\partial g}{\partial \zeta}(\boldsymbol{\gamma}, \zeta_2) - \frac{\partial g}{\partial \zeta}(\boldsymbol{\gamma}, \zeta_1) &= (1 - \alpha)^{-1} \left(\mathbb{P}(-\boldsymbol{\gamma}^T \mathbf{X} > \zeta_1) - \mathbb{P}(-\boldsymbol{\gamma}^T \mathbf{X} > \zeta_2) \right) \\ &= (1 - \alpha)^{-1} \mathbb{P}(\zeta_1 < -\boldsymbol{\gamma}^T \mathbf{X} < \zeta_2) \geq 0 \end{aligned}$$

this quantity is non negative on \mathbb{R} and positive on $\text{supp}(-\boldsymbol{\gamma}^T \mathbf{X})$ because $-\boldsymbol{\gamma}^T \mathbf{X}$ has a density by (P3) and $\text{supp}(-\boldsymbol{\gamma}^T \mathbf{X})$ is an interval thanks to (P5). \blacksquare

We now proceed with the proof of Theorem 3.11.

Proof of Theorem 3.11. Using Theorem 3.10, let $\boldsymbol{\gamma}^*$ be the unique solution to Problem (1.1). The equivalence between the Problems (2.1) and (2.2) implies that any $(\tilde{\boldsymbol{\gamma}}, \tilde{\zeta})$ solution to (2.3) satisfies $\tilde{\boldsymbol{\gamma}} = \boldsymbol{\gamma}^*$.

Moreover $C_\alpha(\boldsymbol{\gamma}^*) = \min_{\zeta \in \mathbb{R}} g(\boldsymbol{\gamma}^*, \zeta)$ and thanks to Proposition 3.14, $A_\alpha(\boldsymbol{\gamma}^*) = \{\zeta^*\}$. Therefore $(\boldsymbol{\gamma}^*, \zeta^*)$ is the unique solution of Problem (2.3). \blacksquare

3.3 Convergence rate

In Shapiro et al. (2009) [7] the rate of convergence is also obtained. We state a version (Theorem 3.16), adapted to our setting. It is used to get the convergence rate in Theorem 3.18.

Consider some additional notations and assumptions:

(P6) $\mathbb{E}(|\mathbf{X}|^2) < +\infty$.

(P7) $\mathbb{E} \left(\sup_{(\boldsymbol{\gamma}, \zeta) \in \mathcal{U}} |L(\boldsymbol{\gamma}, \zeta, \mathbf{X})|^2 \right) < +\infty$.

(P8) There exists a measurable function $C_L : X \rightarrow \mathbf{R}^+$ such that $\mathbb{E}(C_L(\mathbf{X})^2) < \infty$ and $\forall u, u' \in \mathcal{U}$, $|L(u, \mathbf{X}) - L(u', \mathbf{X})| \leq C_L(\mathbf{X})|u - u'|$, almost surely.

(P9) $(\boldsymbol{\gamma}, \zeta) \rightarrow \ell(\boldsymbol{\gamma}, \zeta)$ is differentiable.

Remark 3.15. For loss functions of the form $L(\boldsymbol{\gamma}, \zeta, \mathbf{X}) = \mathbb{E}(-\boldsymbol{\gamma}^T \mathbf{X} + c \cdot (\zeta + \frac{1}{1-\alpha}(-\boldsymbol{\gamma}^T \mathbf{X} - \zeta)^+))$, or $L(\boldsymbol{\gamma}, \mathbf{X}) = \mathbb{E}(-U(\boldsymbol{\gamma}^T \mathbf{X}))$ where U is a convex utility function, it is straightforward to establish the existence of the associated function C_L .

Theorem (5.11 p173 [7]) deals with problems of the following form:

$$\min_{u \in \mathcal{U}} \ell_N(u) \quad \text{s.t.} \quad g_{iN}(u) \leq 0, \quad i = 1, \dots, p. \quad (3.4)$$

In Problem (2.3) there are two types of constraints: one on the parameters $\boldsymbol{\gamma}$ and $\boldsymbol{\zeta}$ and another involving the function g . Consider $g_1 : (\boldsymbol{\gamma}, \boldsymbol{\zeta}) \mapsto g(\boldsymbol{\gamma}, \boldsymbol{\zeta}) - K$, $g_2 : \boldsymbol{\gamma} \mapsto \boldsymbol{\gamma} - \boldsymbol{\gamma}^{\text{up}}$ and $g_3 : \boldsymbol{\gamma} \mapsto \boldsymbol{\gamma}^{\text{low}} - \boldsymbol{\gamma}$. We have $G_1 : ((\boldsymbol{\gamma}, \boldsymbol{\zeta}), \mathbf{X}) \mapsto G((\boldsymbol{\gamma}, \boldsymbol{\zeta}), \mathbf{X}) - K = \boldsymbol{\zeta} + (1 - \alpha)^{-1}(-\boldsymbol{\gamma}^T \mathbf{X} - \boldsymbol{\zeta})^+ - K$, $G_2 = g_2$ and $G_3 = g_3$. They are defined such that for any $u \in \mathcal{U}$ we have $g_i(u) = \mathbb{E}(G_i(u, \mathbf{X}))$. We denote by g_{iN} the approximated version of g_i .

Let $u \in \mathcal{U}$, then the empirical mean estimator $\ell_N(u)$ of $\ell(u)$ is unbiased and has a variance of $\sigma^2(u) := \mathbb{V}ar(L(u, \mathbf{X}))$ which is assumed to be finite. Moreover the standard Central Limit Theorem gives:

$$N^{1/2} (\ell_N(u) - \ell(u)) \xrightarrow[N \rightarrow +\infty]{\mathcal{D}} \mathcal{N}(0, \sigma^2(u))$$

where $\xrightarrow{\mathcal{D}}$ denotes the convergence in distribution. Let $Y(u)$ be a random variable such that $Y(u) \sim \mathcal{N}(0, \mathbb{V}ar(L(u, \mathbf{X}))$. Let $Y_i(u)$ be a random variable such that $Y_i(u) \sim \mathcal{N}(0, \mathbb{V}ar(G_i(u, \mathbf{X})))$, $i = 1, 2, 3$.

We introduce the Lagrangian of Problem (3.4):

$$\mathcal{L}(u, \boldsymbol{\lambda}) := \ell(u) + \sum_{i=1}^p \lambda_i g_i(u),$$

which can be rewritten as

$$\mathcal{L}(\boldsymbol{\gamma}, \boldsymbol{\zeta}, (\boldsymbol{\lambda}, \bar{\boldsymbol{\mu}}, \underline{\boldsymbol{\mu}})) := \ell(\boldsymbol{\gamma}, \boldsymbol{\zeta}) + \boldsymbol{\lambda}(g(\boldsymbol{\gamma}, \boldsymbol{\zeta}) - K) + \bar{\boldsymbol{\mu}}^\top (\boldsymbol{\gamma} - \boldsymbol{\gamma}^{\text{up}}) - \underline{\boldsymbol{\mu}}^\top (\boldsymbol{\gamma} - \boldsymbol{\gamma}^{\text{low}}). \quad (3.5)$$

In our context, Problem (2.3) is convex because ℓ is convex. In addition the set S of optimal solutions to the original problem (1.1) is non-empty and bounded. Moreover, ℓ and g_i are bounded on \mathcal{U} and in a neighbourhood of S . Thanks to these results, we can state a modified version of Theorem 5.11, p173, in [7], adapted to our framework.

Theorem 3.16 (Restatement of Theorem 5.11, p173, in [7]). *Let v_N be the optimal value of the above problem (3.4). Assume that the following assumptions are satisfied:*

- (i) *There exists $\bar{u} \in \mathcal{U}$ such that $g_i(\bar{u}) < 0$, $i = 1, 2, 3$. (Slater's condition).*
- (ii) *The following two assumptions are satisfied for $H = L, G_i, i = 1, 2, 3$:*
 - (A₁) *There exists $\tilde{u} \in \mathcal{U}$ such that $\mathbb{E}(H(u, \mathbf{X})^2) < +\infty$.*
 - (A₂) *There exists a measurable function $C : X \rightarrow \mathbf{R}^+$ such that $\mathbb{E}(C(\mathbf{X})^2) < \infty$ and $\forall u, u' \in \mathcal{U}$, $|H(u, \mathbf{X}) - H(u', \mathbf{X})| \leq C(\mathbf{X})|u - u'|$ almost surely.*

Then

$$N^{1/2} (v_N - v^*) \xrightarrow{\mathcal{D}} \inf_{u \in S} \sup_{\boldsymbol{\lambda} \in \Lambda} \left[Y(u) + \sum_{i=1}^p \lambda_i Y_i(u) \right]. \quad (3.6)$$

A set Λ of Lagrange vectors $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_p)$ is associated to each optimal solution $u^* \in S$ which satisfy,

$$u^* \in \arg \min_{u \in \mathcal{U}} \mathcal{L}(u, \boldsymbol{\lambda}), \quad \lambda_i \leq 0 \text{ and } \lambda_i g_i(u^*) = 0, \quad i = 1, \dots, p.$$

The set Λ coincides with the set of optimal solutions to the dual of the original problem and is therefore the same for any optimal solution $u^* \in S$.

Furthermore, if $S = \{u^*\}$ and $\Lambda = \{\boldsymbol{\lambda}^*\}$ are singletons, then,

$$N^{1/2} (v_N - v^*) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma^2)$$

with

$$\sigma^2 := \text{Var} \left(L(u^*, \mathbf{X}) + \sum_{i=1}^P \lambda_i^* G_i(u^*, \mathbf{X}) \right).$$

Lemma 3.17. *Given a set of constraints $\boldsymbol{\gamma}^{\text{up}}, \boldsymbol{\gamma}^{\text{low}}$ and $\boldsymbol{\gamma} \in \Gamma$, if (P1), (P2) hold, then there exists at most a single $K \leq \min_{\boldsymbol{\gamma} \in \arg \min_{\Gamma} \ell_0} C_\alpha(\boldsymbol{\gamma})$ such that $\boldsymbol{\gamma}$ solves (1.1) with K as the capital limit.*

Proof. This is a corollary of Lemma 3.12, because if $\boldsymbol{\gamma}$ is solution of (1.1) with $K' \leq C_\alpha(\arg \min_{\Gamma} \ell_0)$, then $K' = C_\alpha(\boldsymbol{\gamma})$ and so there can only be one K' . \blacksquare

Theorem 3.18. *If (P1) to (P9) hold, denote by $u^* = (\boldsymbol{\gamma}^*, \zeta^*)$ the unique solution to Problem (2.3), then*

$$N^{1/2} (v_N - v^*) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma^2)$$

with

$$\sigma^2 := \text{Var} (L(u^*, \mathbf{X}) + \lambda^*(G(u^*, \mathbf{X}) - K))$$

and λ^* is the maximum value of the Lagrange multiplier λ associated with the capital constraint.

Proof of Theorem 3.18. The uniqueness of u^* is given by (P1) to (P5) using Theorem 3.11. We look at the conditions for applying Theorem 3.16. Condition (i) holds because we assumed that the interior of \mathcal{U}^K is non-empty which means that Slater's condition is satisfied for g_1, g_2, g_3 . The functions G_2, G_3 are deterministic and linear functions of $\boldsymbol{\gamma}$ so (A₁) and (A₂) hold. Conditions (A₁) and (A₂) hold for L by (P7) and (P8) and for G_1 it is easy to prove with (P6).

We can apply Theorem 3.16 to obtain (3.6). G_2 and G_3 are deterministic so $Y_2 = Y_3 = 0$, therefore $\sum_{i=1}^P \lambda_i Y_i((\boldsymbol{\gamma}, \zeta)) = \lambda_1 Y_1((\boldsymbol{\gamma}, \zeta))$. Thanks to the uniqueness of u^* , we only need the maximum value of λ_1 or the uniqueness of λ_1 to apply the second part of Theorem 3.16.

Consider the Lagrangian associated with Problem (3.4),

$$\mathcal{L}(\boldsymbol{\gamma}, \zeta, (\lambda, \bar{\boldsymbol{\mu}}, \underline{\boldsymbol{\mu}})) = \ell(\boldsymbol{\gamma}, \zeta) + \lambda(g(\boldsymbol{\gamma}, \zeta) - K) + \bar{\boldsymbol{\mu}}^\top (\boldsymbol{\gamma} - \boldsymbol{\gamma}^{\text{up}}) - \underline{\boldsymbol{\mu}}^\top (\boldsymbol{\gamma} - \boldsymbol{\gamma}^{\text{low}}),$$

where $g(\boldsymbol{\gamma}, \zeta) = \zeta + (1 - \alpha)^{-1} \mathbb{E}[(-\boldsymbol{\gamma}^\top \mathbf{X} - \zeta)^+]$. Let $(\boldsymbol{\gamma}^*, \zeta^*)$ be an optimal solution and $(\lambda, \bar{\boldsymbol{\mu}}, \underline{\boldsymbol{\mu}})$ a corresponding vector of Lagrange multipliers. Since $\zeta^* \in A_\alpha(\boldsymbol{\gamma}^*)$, one has $\mathbb{P}(-\boldsymbol{\gamma}^{*\top} \mathbf{X} - \zeta^* \geq 0) = 1 - \alpha$.

The stationarity condition with respect to ζ yields

$$\frac{\partial \mathcal{L}}{\partial \zeta}(\boldsymbol{\gamma}^*, \zeta^*) = \frac{\partial \ell}{\partial \zeta}(\boldsymbol{\gamma}^*, \zeta^*) + \lambda \left(1 - (1 - \alpha)^{-1} \mathbb{P}(-\boldsymbol{\gamma}^{*\top} \mathbf{X} - \zeta^* \geq 0) \right) = \frac{\partial \ell}{\partial \zeta}(\boldsymbol{\gamma}^*, \zeta^*),$$

so that

$$\frac{\partial \ell}{\partial \zeta}(\boldsymbol{\gamma}^*, \zeta^*) = 0. \tag{3.7}$$

The stationarity condition with respect to $\boldsymbol{\gamma}$ reads

$$\frac{\partial \ell}{\partial \boldsymbol{\gamma}}(\boldsymbol{\gamma}^*, \zeta^*) + \lambda \mathbb{E}(-\mathbf{X} \mid -\boldsymbol{\gamma}^{*\top} \mathbf{X} - \zeta^* \geq 0) + \bar{\boldsymbol{\mu}} - \underline{\boldsymbol{\mu}} = 0,$$

that is, for each $i = 1, \dots, d$,

$$\frac{\partial \ell}{\partial \gamma_i}(\boldsymbol{\gamma}^*, \zeta^*) + \lambda \mathbb{E}(-X_i \mid -\boldsymbol{\gamma}^{*\top} \mathbf{X} - \zeta^* \geq 0) + \bar{\mu}_i - \underline{\mu}_i = 0. \tag{3.8}$$

Since the CVaR constraint is active and ℓ is convex, there exists at least one index $j \in \{1, \dots, d\}$ such that $\frac{\partial \ell}{\partial \gamma_j}(\boldsymbol{\gamma}^*, \zeta^*) \neq 0$. This index cannot correspond to an active bound; otherwise, a feasible perturbation would decrease the objective, contradicting optimality. Hence $\bar{\mu}_j = \underline{\mu}_j = 0$.

For such a free coordinate j , (3.8) reduces to

$$\frac{\partial \ell}{\partial \gamma_j}(\boldsymbol{\gamma}^*, \zeta^*) + \lambda \mathbb{E}(-X_j \mid -\boldsymbol{\gamma}^{*\top} \mathbf{X} - \zeta^* \geq 0) = 0.$$

Under the non-degeneracy condition $\mathbb{E}(-X_j \mid -\boldsymbol{\gamma}^{*\top} \mathbf{X} - \zeta^* \geq 0) \neq 0$, this identity uniquely determines the Lagrange multiplier,

$$\lambda = -\frac{\frac{\partial \ell}{\partial \gamma_j}(\boldsymbol{\gamma}^*, \zeta^*)}{\mathbb{E}(-X_j \mid -\boldsymbol{\gamma}^{*\top} \mathbf{X} - \zeta^* \geq 0)}.$$

Therefore, the Lagrange multiplier λ associated with the CVaR constraint is unique.

Moreover, if we denote this Lagrange multiplier by λ^* , then $\text{Var}(L(u^*, \mathbf{X}) + \sum_{i=1}^p \lambda_i^* G_i(u^*, \mathbf{X}))$ becomes $\text{Var}(L(u^*, \mathbf{X}) + \lambda^*(G(u^*, \mathbf{X}) - K))$. \blacksquare

4 Numerical study

This section illustrates the results of section 3, focusing on convergence and convergence rates. When possible, we compare the results of the SAA method with explicit computational solutions. We first consider the Gaussian case for which an explicit solution is known for the mean-CVaR, where the asset returns do not all follow the same distribution. This first case will allow us to validate the algorithm by comparing it with theoretical solutions. Second, we consider a more realistic case where all the asset returns do not follow the same distribution. This last case is close to what can be found in (re)insurance.

All numerical examples are solved within a few seconds using first-order optimization methods implemented in standard Python libraries, primarily SciPy and NumPy. These computation times can be contrasted with those typically required by linear programming approaches, as discussed for instance in [1]. In such formulations, the dimensionality of the optimization problem grows linearly with the sample size, since each additional scenario introduces new variables and constraints. By contrast, first-order methods operate in the original decision space and evaluate the empirical objective and its subgradients through sample averages, without increasing the dimensionality of the optimization problem. As a result, linear programming methods may become computationally demanding when large data samples are considered, even though efficient and specialized solvers are available.

4.1 Context

In what follows, we consider $L_0(\boldsymbol{\gamma}, \mathbf{X}) = -\boldsymbol{\gamma}^T \mathbf{X}$ with a risk level $\alpha = 0.99$.

Note that L_0 is convex and continuous so (P1) holds.

We rewrite $L(\boldsymbol{\gamma}, \zeta, \mathbf{X}) = -\boldsymbol{\gamma}^T \mathbf{X}$ and $\ell_N(\boldsymbol{\gamma}, \zeta) := -\frac{\boldsymbol{\gamma}^T}{N} \sum_{j=1}^N \mathbf{X}^{(j)}$.

Once we have chosen the asset distributions and other parameters for the problem $(K, \boldsymbol{\gamma}^{\text{up}}, \boldsymbol{\gamma}^{\text{low}})$, for a given N , we solve the SAA problem (2.4) k times, using k independent i.i.d. samples of \mathbf{X} of size N to obtain a size- k sample of $\boldsymbol{\gamma}_N$ and v_N values.

4.2 Gaussian case

We consider the Gaussian case, where $\mathbf{X} \sim \mathcal{N}_d(\mathbf{1}_d, \Sigma)$, with Σ a positive definite matrix. Note that $\boldsymbol{\gamma}^T \mathbf{X} \sim \mathcal{N}(\boldsymbol{\gamma}^T \mathbf{1}_d, \boldsymbol{\gamma}^T \Sigma \boldsymbol{\gamma})$ and $\text{CVaR}_\alpha(-\boldsymbol{\gamma}^T \mathbf{X}) = -\boldsymbol{\gamma}^T \mathbf{1}_d + T_Z \sqrt{\boldsymbol{\gamma}^T \Sigma \boldsymbol{\gamma}}$, with $T_Z = \text{CVaR}_\alpha(\mathcal{N}(0, 1))$.

If $\tilde{\mathbf{X}} \sim \mathcal{N}(\tilde{\mu}, \tilde{\Sigma})$, then $\boldsymbol{\gamma}^T \tilde{\mathbf{X}} = (\boldsymbol{\gamma} \odot \tilde{\mu})^T \mathbf{X}$ and with $\mathbf{X} \sim \mathcal{N}(\mathbf{1}_d, \Sigma)$ and $\Sigma = \tilde{\mu}^T \tilde{\Sigma} \tilde{\mu}$. Therefore, assuming $\mathbb{E}(X) = \mathbf{1}_d$ is without loss of generality.

Denote $\boldsymbol{\gamma}^*$ the optimal solution to (1.1), if $\gamma_i^{\text{low}} < \gamma_i^* < \gamma_i^{\text{up}}$ for $i = 1, \dots, d$, then:

$$\boldsymbol{\gamma}^* = K \left(\frac{1}{T_Z - \sqrt{\mathbf{1}_d^T \Sigma^{-1} \mathbf{1}_d}} \right) \frac{\Sigma^{-1} \mathbf{1}_d}{\sqrt{\mathbf{1}_d^T \Sigma^{-1} \mathbf{1}_d}}, \quad (4.1)$$

See Appendix 2 for a proof.

Properties (P2), (P5) and (P6) hold immediately. We choose the parameter bounds for (P4) to hold: $C_\alpha(\boldsymbol{\gamma}) > K, \forall \boldsymbol{\gamma} \in \arg \min_{\boldsymbol{\gamma} \in \Gamma} \ell_0(\boldsymbol{\gamma})$. Otherwise, the capital limit would not be reached. Note also that (P7) through (P9) hold.

We set $d = 5$, $\alpha = 0.99$, $K = 100$, $k = 5000$ runs and a covariance matrix Σ picked at random.

$$\Sigma = \begin{pmatrix} 4.490 & -0.377 & 0.059 & 0.585 & -1.709 \\ -0.377 & 6.109 & -1.300 & 0.229 & 1.380 \\ 0.059 & -1.300 & 7.059 & -1.401 & 0.210 \\ 0.585 & 0.229 & -1.401 & 8.400 & -1.250 \\ -1.709 & 1.380 & 0.210 & -1.250 & 19.934 \end{pmatrix}.$$

4.2.1 Without bounds

This case is without bounds in the sense that we set $\boldsymbol{\gamma}^{\text{low}}, \boldsymbol{\gamma}^{\text{up}}$ to have $\gamma_i^{\text{low}} < \gamma_i^* < \gamma_i^{\text{up}}$, in other words in such a way that they are never reached. Therefore, we use equation (4.1) to compute $\boldsymbol{\gamma}^* = (15.113, 12.576, 12.610, 8.702, 3.958)$ and $v^* = -52.960$.

We also compute the standard estimator of $\sigma^2 = \mathbb{V}ar(-\boldsymbol{\gamma}^{*T} \mathbf{X} + \lambda^*(\zeta^* + \frac{1}{1-\alpha}(-\boldsymbol{\gamma}^{*T} \mathbf{X} - \zeta^*)^+ - K))$ from Theorem 3.18 and we obtain $\hat{\sigma} = 149.860$.

Figures 1a to 1e and 4a illustrate Theorem 3.8, showing that the solution to the SAA formulation (2.4) converges to the original solution found with (4.1). Indeed, the convergence of $\boldsymbol{\gamma}_N$ and v_N is observed and a sample size of 10^5 is sufficient to obtain an acceptable error, here a 2% on each component of $\boldsymbol{\gamma}^*$ and 1% on v^* .

Figure 4b illustrates a convergence rate of \sqrt{N} seen in Theorem 3.18 since the slope of each curve is -0.5 . In Figures 2a to 2e we observe the same empirical behavior.

The histogram of Figure 5 has a Gaussian shape with parameters $(\mu, \sigma) = (-9.15, 160.33)$. This illustrates Theorem 3.18 with $\hat{\sigma} = 149.860$.

While we do not claim a theoretical result for the weights, our experiments indicate that as the sample size increases, their distribution converges toward Gaussian. This motivates fitting a normal model and aligns with a law of large numbers intuition. Figure 3 illustrates this phenomenon, suggesting the asymptotic normality of the parameters.

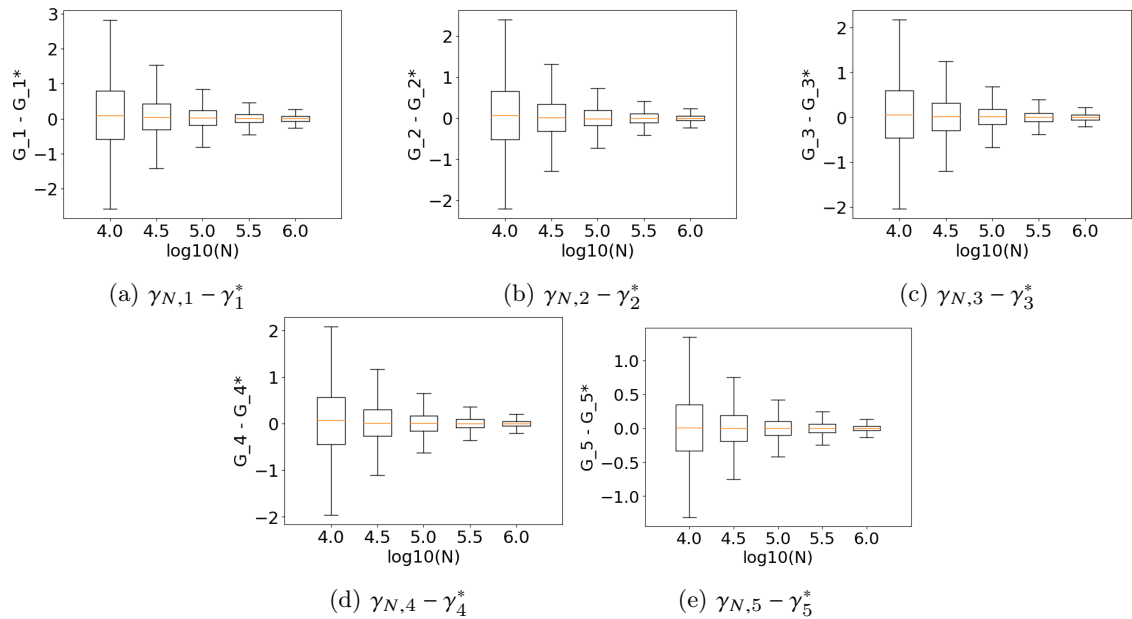


Figure 1: Component-wise Boxplots of $\gamma_N - \gamma^*$. (boxplot standard parameters: Q1-1.5IQR, Q1, median, Q3, Q3+1.5IQR)

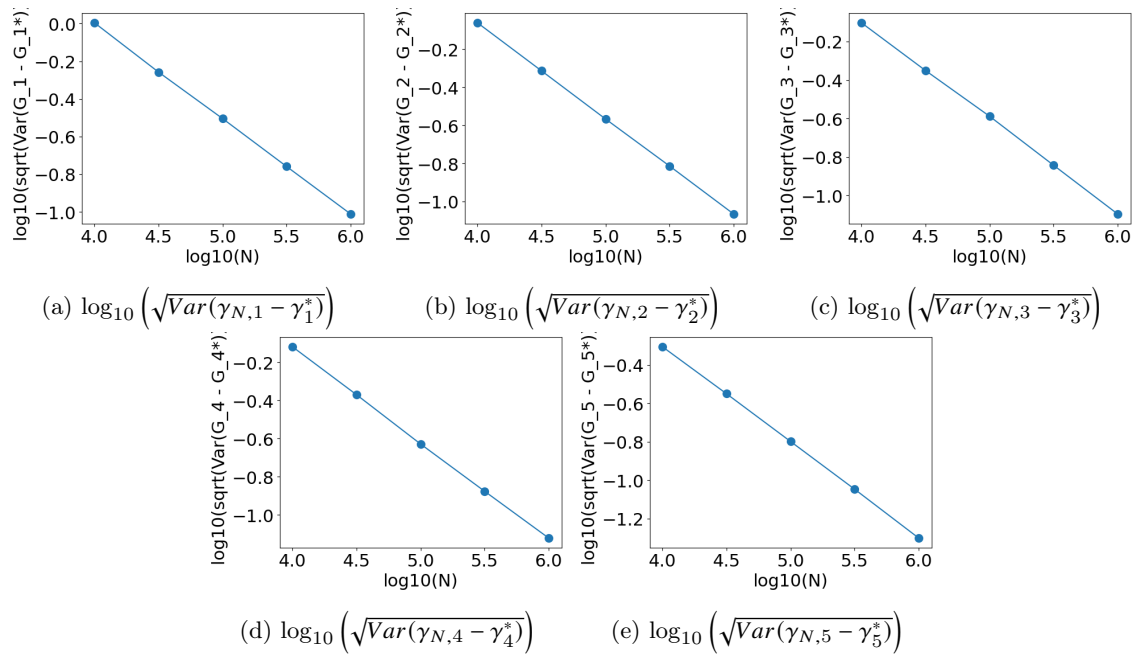


Figure 2: Log plot of $\log_{10}(\sqrt{\text{Var}(\gamma_N - \gamma^*)})$ components-wise.

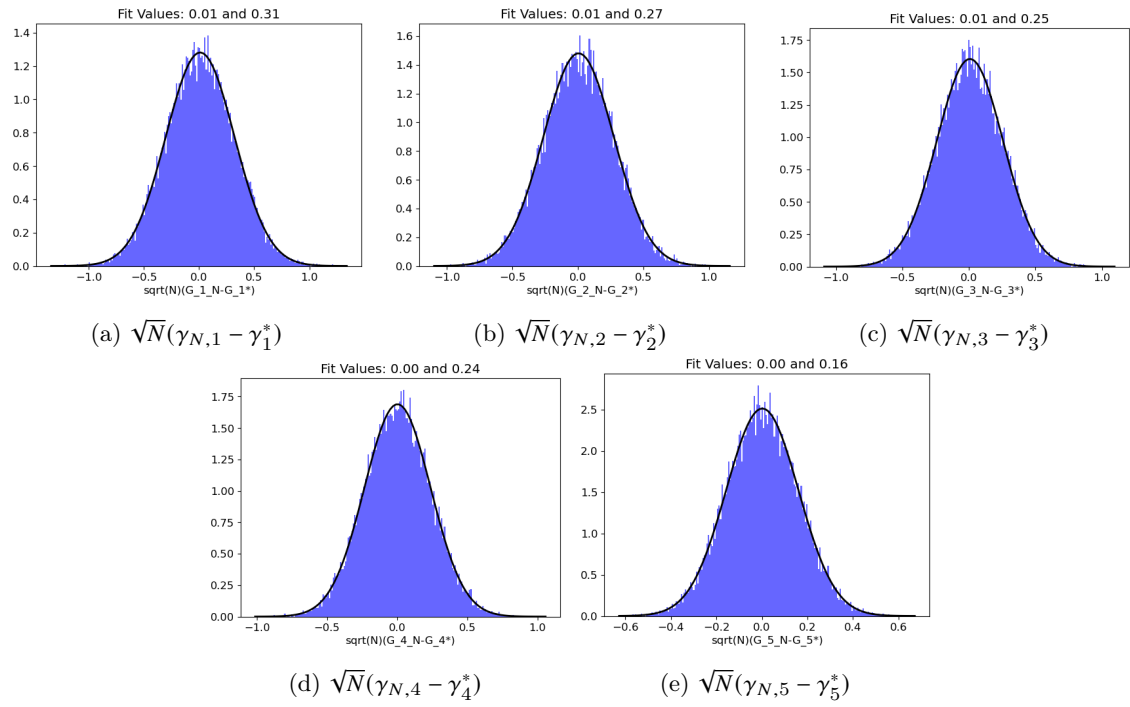


Figure 3: Components-wise histograms of $\sqrt{N}(\gamma_N - \gamma^*)$ for $N = 10^5$ and plot of a fitted normal distribution.

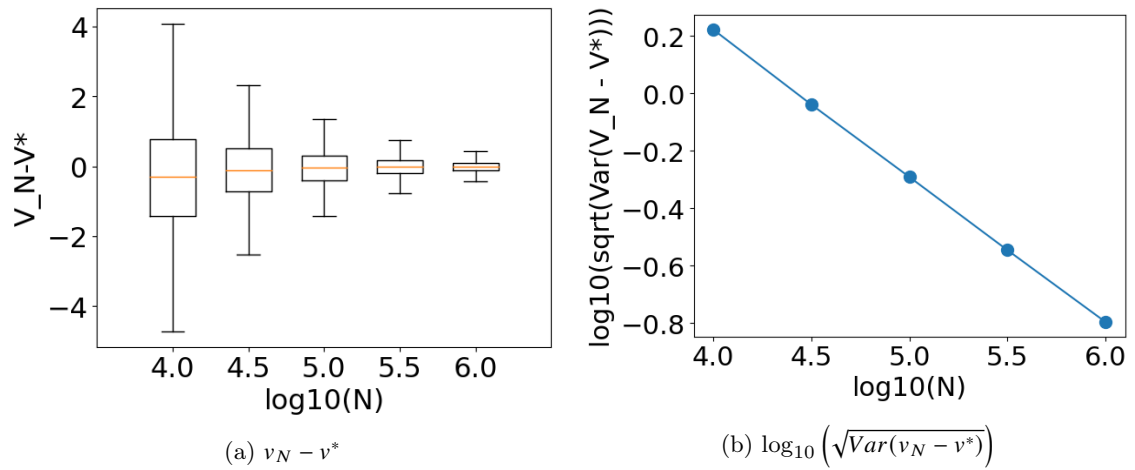


Figure 4: Boxplots of $v_N - v^*$ and \log_{10} plot of $\log_{10}(\sqrt{\text{Var}(v_N - v^*)})$.

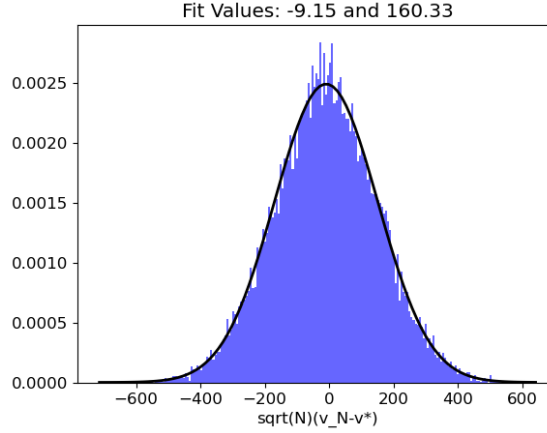


Figure 5: Histogram of $\sqrt{N}(v_N - v^*)$ for $N = 10^5$ and plot of a fitted normal distribution.

4.2.2 With bounds

Now, in the case where the optimal parameters reach at least one bound, we cannot compute the theoretical values for $\boldsymbol{\gamma}^*$ and v^* .

Based on the unconstrained optimal values (15.113, 12.576, 12.610, 8.702, 3.958), we set the bounds $\boldsymbol{\gamma}^{\text{low}} = (0, 0, 15, 0, 6)$ and $\boldsymbol{\gamma}^{\text{up}} = (10, 30, 30, 30, 30)$ to ensure that some parameters hit their limits. We expect parameter 1 to hit its higher limit and parameters 3 and 5 to hit their lower limits. These claims are confirmed by a simulation with $N = 10^6$, where $\boldsymbol{\gamma}10^6 = (10, 9.747, 15, 7.867, 6)$ and $v10^6 = -48.970$, which is higher as expected, than in the unconstrained case (-52.960). Figures 6 and 8 show convergence to a solution $\boldsymbol{\gamma}^*$ in which components 1, 3, and 5 are bounded (see Figures 6a, 6c, 6e), as expected. Figure 9 shows that $N^{1/2}(v_N - v^*) \sim \mathcal{N}(0, \sigma^2)$, in accordance with Theorem 3.18. We approximate v^* by $\mathbb{E}(v_N)$ and $\boldsymbol{\gamma}^*$ by $\mathbb{E}(\boldsymbol{\gamma}_N)$ over k runs. Since we do not have access to a theoretical σ value, we estimate it as $\tilde{\sigma} = 226.99$ using a Gaussian fit. Even though we could not prove any central limit theorem for the portfolio weights, we wanted to see how close to a Gaussian shape their distribution was. Figure 7a shows an approximately Gaussian asymptotic behavior, whereas Figure 7b deviates more noticeably from a Gaussian distribution. We believe that this deviation is partly explained by the fact that some weight components have reached their bounds, which obscures the emergence of a clear Gaussian distribution.

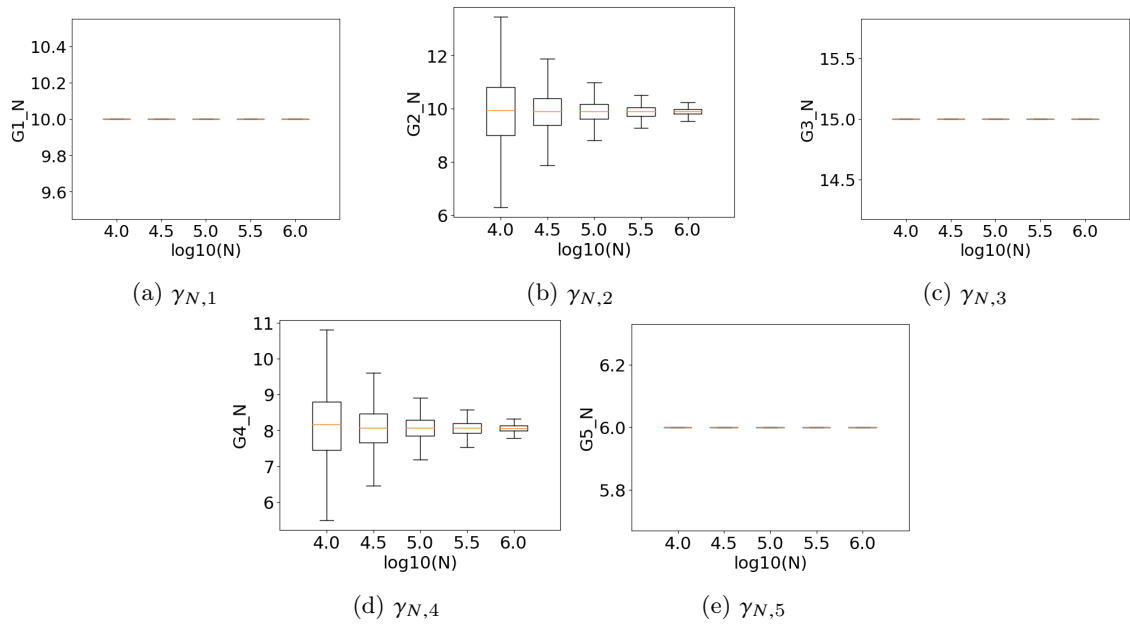


Figure 6: Component-wise Boxplots of γ_N . (boxplot standard parameters: Q1-1.5IQR, Q1, median, Q3, Q3+1.5IQR)

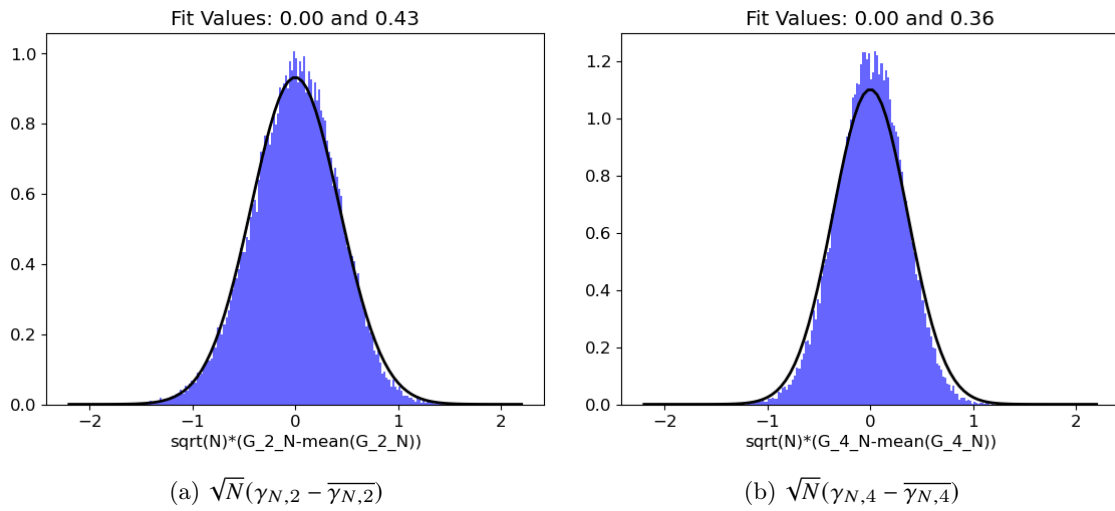


Figure 7: Histograms of $\sqrt{N}(\gamma_N - \gamma^*)$ for components 2 and 4, for $N = 10^5$ and plot of a fitted normal distribution on each.

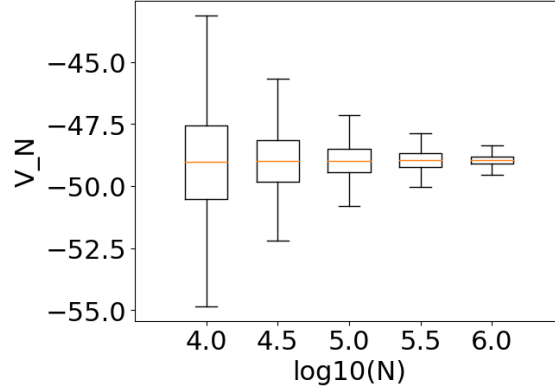


Figure 8: Boxplots of v_N . (boxplot standard parameters: Q1-1.5IQR, Q1, median, Q3, Q3+1.5IQR)

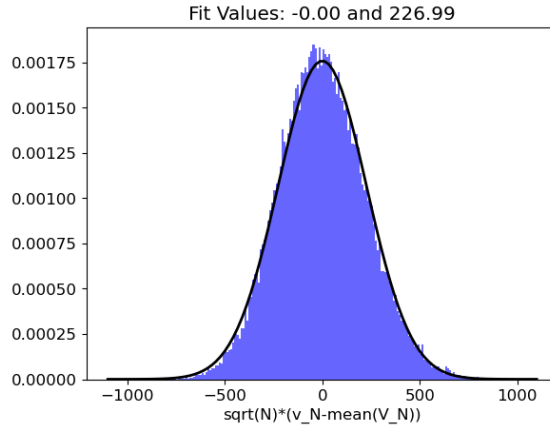


Figure 9: Histogram of $\sqrt{N}(v_N - \bar{v}_N)$ for $N = 10^5$ and plot of a fitted normal distribution.

4.3 Multi laws setting

In practice, we only have access to a sample of data for the returns of each Line of Business (LoB). This sample may result from complex modeling and might not follow any known probability distribution. Each LoB follows a distinct distribution and exhibits a markedly different risk profile.

We selected five LoBs and, for each, chose a distribution that we then shifted to have a mean of 1, as in the Gaussian case. Consider five random variables $\tilde{\mathbf{X}}_i$, for $i = 1, \dots, 5$. $\tilde{\mathbf{X}}_1$ and $\tilde{\mathbf{X}}_2$ follow generalized Pareto distributions with shape parameters $\alpha_{\text{pareto}} = 0.45$ and 0.25 , respectively, as defined in the SciPy Python package. $\tilde{\mathbf{X}}_3$ and $\tilde{\mathbf{X}}_4$ follow log-normal distributions with parameters $(\mu, \sigma) = (1.7, 1)$ and $(1.3, 1)$, respectively.

Finally, $\tilde{\mathbf{X}}_5$ follows a normal distribution with $(\mu, \sigma) = (2, 6)$.

We define $\mathbf{X}_i = 1 - (\tilde{\mathbf{X}}_i - \bar{\tilde{\mathbf{X}}}_i)$ for $i = 1, \dots, 5$. In such a setting, the standalone VaR and $CVaR$ can be calculated analytically. The resulting values are summarized in Table 1.

We choose three different dependency settings: a Gaussian copula, a clayton copula and the independent case. For the clayton case we use $\alpha_{\text{clayton}} = 2$ and in the Gaussian case we use the following correlation matrix picked at random.

i	$\mu(\mathbf{X}_i)$	$\sigma(\mathbf{X}_i)$	$VaR_{0.99}(-\mathbf{X}_i)$	$CVaR_{0.99}(-\mathbf{X}_i)$
1	1	5.75	12.61	27.05
2	1	1.89	6.32	10.53
3	1	17.49	46.94	80.55
4	1	4.89	17.25	53.73
5	1	6	12.96	15.99

Table 1: distribution description for the multi laws case

$$\Sigma = \begin{pmatrix} 1 & -0.1285 & 0.3979 & -0.4731 & 0.3879 \\ -0.1285 & 1 & -0.0574 & -0.2253 & -0.3532 \\ 0.3979 & -0.0574 & 1 & -0.5363 & 0.12 \\ -0.4731 & -0.2253 & -0.5363 & 1 & 0.0999 \\ 0.3879 & -0.3532 & 0.12 & 0.0999 & 1 \end{pmatrix}.$$

We run the optimization with $K = 100$ and $N = 10^6$, considering two cases: one without constraints, and one with bounds defined by $\boldsymbol{\gamma}^{\text{low}} = (2, 0, 0, 1, 0)$ and $\boldsymbol{\gamma}^{\text{up}} = (10, 5, 10, 10, 10)$. The corresponding optimization results are reported in Table 2.

$N = 10^6$	γ_1	γ_2	γ_3	γ_4	γ_5	v
without bound						
independent	1.407	7.596	0.101	1.196	4.571	-14.922
Gaussian	0.575	10.386	0.0628	1.131	5.498	-17.655
clayton	0.7521	6.535	0.0178	0.656	3.794	-11.763
with bounds						
independent	2	5	0.146	1.501	5.343	-13.987
Gaussian	2	5	0.0333	1.577	4.839	-13.404
clayton	2	5	2.968	1	2.969	-10.959

Table 2: multi laws results

At the optimum, and across all cases, we expect that the Lines of Business (LoBs) with the lowest $CVaR$ values, such as LoBs 2 and 5, are prioritized over those with higher $CVaR$, such as LoB 3. In the case without bounds, we expect LoBs 1 and 2 to reach their natural limits, which is indeed what we observe. As expected, the presence of bounds dampens the diversification effects, resulting in lower returns compared to the unconstrained case.

In the independent case, γ_i is approximately inversely proportional to $CVaR_{0.99}(-\mathbf{X}_i)$, except for the bounded LoBs. In the Gaussian dependency case without bounds, returns improve due to diversification enabled by the chosen correlation matrix. For instance, LoB 2 benefits from negative correlations with LoBs 4 and 5, as given by $\Sigma_{2,4} = -0.2253$ and $\Sigma_{2,5} = -0.3532$. However, this diversification benefit cannot be fully exploited under constraints, mainly because LoB 2 is limited to a maximum of 5. The Clayton copula introduces stronger tail dependence, which weakens the diversification effect on $CVaR$, leading to lower returns.

5 Appendix

This appendix has two parts: the first presents the proof of the Equivalent Theorem used in 2.1, and the second provides the derivation of the optimal value formula for the Gaussian case when bounds are not binding, as discussed in 4.2.

5.1 Equivalent Theorem proof

In Section 2.1, we use a result from Krokmal, Palmquist, and Uryasev (2002) [2] concerning the equivalence of two optimization problems. Since this specific property is not explicitly proven in the article, we provide a complete proof here.

The proof is based on the Karush-Kuhn-Tucker necessary and sufficient conditions.

Proposition 2.1. *For any function $R : \mathbb{R}^d \times \mathbb{R} \rightarrow \mathbb{R}$ such that for any $\boldsymbol{\gamma} \in \mathbb{R}^d$, $R(\boldsymbol{\gamma}, \cdot)$ is non-increasing then the following two problems are equivalent*

$$\inf_{\boldsymbol{\gamma} \in \Xi} -R(\boldsymbol{\gamma}, C_\alpha(\boldsymbol{\gamma})) \quad \text{s.t.} \quad C_\alpha(\boldsymbol{\gamma}) \leq K \quad (2.1)$$

and

$$\inf_{(\boldsymbol{\gamma}, \zeta) \in \Xi \times \mathbb{R}} -R(\boldsymbol{\gamma}, g(\boldsymbol{\gamma}, \zeta)) \quad \text{s.t.} \quad g(\boldsymbol{\gamma}, \zeta) \leq K. \quad (2.2)$$

Moreover, if the C_α constraint is active in (2.1), then $(\boldsymbol{\gamma}^*, \zeta^*)$ solves (2.2) if and only if $\boldsymbol{\gamma}^*$ achieves the minimum of (2.1) and $\zeta^* \in A_\alpha(\boldsymbol{\gamma}^*)$.

Proof. Let us write down the necessary and sufficient Karush-Kuhn-Tucker conditions for problems (2.1) and (2.2).

For $R(\cdot, C_\alpha(\cdot))$ to achieve its minimum at $\boldsymbol{\gamma}^* \in \Gamma$, it must exist a constant μ_1 such that, for all $\boldsymbol{\gamma} \in \Gamma$:

$$\begin{aligned} -R(\boldsymbol{\gamma}^*, C_\alpha(\boldsymbol{\gamma}^*)) + \mu_1 C_\alpha(\boldsymbol{\gamma}^*) &\leq -R(\boldsymbol{\gamma}, C_\alpha(\boldsymbol{\gamma})) + \mu_1 C_\alpha(\boldsymbol{\gamma}), \\ \mu_1(C_\alpha(\boldsymbol{\gamma}^*) - K) &= 0. \end{aligned} \quad (\text{KKT-(2.1)})$$

If $\mu_1 \geq 0$, then the conditions are also sufficient.

We have the same kind of conditions for (2.2). For $(\boldsymbol{\gamma}, \zeta) \in \mathcal{U}$:

$$\begin{aligned} -R(\boldsymbol{\gamma}^*, g(\boldsymbol{\gamma}^*, \zeta^*)) + \mu_2 g(\boldsymbol{\gamma}^*, \zeta^*) &\leq -R(\boldsymbol{\gamma}, g(\boldsymbol{\gamma}, \zeta)) + \mu_2 g(\boldsymbol{\gamma}, \zeta), \\ \mu_2(g(\boldsymbol{\gamma}^*, \zeta^*) - K) &= 0, \quad \mu_2 \geq 0. \end{aligned} \quad (\text{KKT-(2.2)})$$

First, suppose that $\boldsymbol{\gamma}^*$ is a solution to (2.1) and $\zeta^* \in A_\alpha(\boldsymbol{\gamma}^*)$. Let us show that $(\boldsymbol{\gamma}^*, \zeta^*)$ is a solution to (2.2). Using necessary and sufficient conditions (KKT-(2.1))

$$\begin{aligned} -R(\boldsymbol{\gamma}^*, g(\boldsymbol{\gamma}^*, \zeta^*)) + \mu_1 g(\boldsymbol{\gamma}^*, \zeta^*) &= -R(\boldsymbol{\gamma}^*, C_\alpha(\boldsymbol{\gamma}^*)) + \mu_2 C_\alpha(\boldsymbol{\gamma}^*) \\ &\leq -R(\boldsymbol{\gamma}, C_\alpha(\boldsymbol{\gamma})) + \mu_1 C_\alpha(\boldsymbol{\gamma}) = -R(\boldsymbol{\gamma}, \min_{\zeta} g(\boldsymbol{\gamma}, \zeta)) + \mu_1 \min_{\zeta} g(\boldsymbol{\gamma}, \zeta), \\ &\leq -R(\boldsymbol{\gamma}, g(\boldsymbol{\gamma}, \zeta)) + \mu_1 g(\boldsymbol{\gamma}, \zeta) \quad \text{because } R(\boldsymbol{\gamma}, \cdot) \text{ is not increasing,} \end{aligned}$$

and

$$\mu_2(g(\boldsymbol{\gamma}^*, \zeta^*) - K) = \mu_1(C_\alpha(\boldsymbol{\gamma}^*) - K)0, \quad \mu_1 \geq 0, (\boldsymbol{\gamma}, \zeta) \in \mathcal{U}.$$

Thus, (KKT-(2.2)) conditions are satisfied and $(\boldsymbol{\gamma}^*, \zeta^*)$ is a solution to (2.2).

Now, we suppose that $(\boldsymbol{\gamma}^*, \zeta^*)$ achieves the minimum of (2.2) and $\mu_2 > 0$. For a fixed $\boldsymbol{\gamma}^*$, ζ^* minimizes the function $\zeta \mapsto -R(\boldsymbol{\gamma}^*, g(\boldsymbol{\gamma}^*, \zeta)) + \mu_2 g(\boldsymbol{\gamma}^*, \zeta)$, and, consequently, the function $\zeta \mapsto g(\boldsymbol{\gamma}^*, \zeta)$ because $R(\boldsymbol{\gamma}, \cdot)$ is not increasing. Then, it implies that $\zeta^* \in A_\alpha(\boldsymbol{\gamma}^*)$. Thus

$$\begin{aligned} -R(\boldsymbol{\gamma}^*, C_\alpha(\boldsymbol{\gamma}^*)) + \mu_2 C_\alpha(\boldsymbol{\gamma}^*) &= -R(\boldsymbol{\gamma}^*, g(\boldsymbol{\gamma}^*, \zeta^*)) + \mu_2 g(\boldsymbol{\gamma}^*, \zeta^*) \\ &\leq -R(\boldsymbol{\gamma}, g(\boldsymbol{\gamma}, V_\alpha(\boldsymbol{\gamma}))) + \mu_2 g(\boldsymbol{\gamma}, V_\alpha(\boldsymbol{\gamma})) = -R(\boldsymbol{\gamma}, C_\alpha(\boldsymbol{\gamma})) + \mu_2 C_\alpha(\boldsymbol{\gamma}) \end{aligned}$$

and

$$\mu_2(C_\alpha(\boldsymbol{\gamma}^*) - K) = \mu_2(g(\boldsymbol{\gamma}^*, \zeta^*) - K) = 0, \quad \mu_2 \geq 0, \boldsymbol{\gamma} \in \Gamma.$$

We proved that (KKT-(2.2)) are satisfied, i.e. $\boldsymbol{\gamma}^*$ is a solution to (2.1) which completes the proof. \blacksquare

5.2 Proof of Proposition 3.13

This appendix presents two lemmas and provides the proof of Proposition 3.13

Lemma 5.1. *Let $\alpha \in]0, 1[$. For any real-valued random variable X with cumulative distribution function F_X , set $U_X := F_X(X)$ and $A_X := \mathbb{1}_{\{U_X \geq \alpha\}}$. Let Y and Z be two real-valued random variables with densities on \mathbb{R} . If $\mathbb{E}(YA_Y) + \mathbb{E}(ZA_Z) - \mathbb{E}((Y+Z)A_{Y+Z}) = 0$, then $\{U_Y \geq \alpha\} = \{U_{Y+Z} \geq \alpha\} = \{U_Z \geq \alpha\}$.*

Proof. Note that A_Y and A_{Y+Z} have the same distribution, thus, $\mathbb{E}(A_Y - A_{Y+Z}) = 0$. Then, for all $m \in \mathbb{R}$ we have $\mathbb{E}(Y(A_Y - A_{Y+Z})) = \mathbb{E}((Y - m)(A_Y - A_{Y+Z}))$. Take $m = F_Y^{-1}(\alpha)$, with F^{-1} the generalized inverse of the cdf. Then, $(Y - m)(A_Y - A_{Y+Z}) \geq 0$. As a consequence $\mathbb{E}(Y(A_Y - A_{Y+Z})) \geq 0$ and $\mathbb{E}(Z(A_Z - A_{Y+Z})) \geq 0$. The assumption can be rewritten as $\mathbb{E}(Y(A_Y - A_{Y+Z})) + \mathbb{E}(Z(A_Z - A_{Y+Z})) = 0$. We conclude that $\mathbb{E}(Y(A_Y - A_{Y+Z})) = \mathbb{E}(Z(A_Z - A_{Y+Z})) = 0$. In addition, $\mathbb{E}((Y - m)(A_Y - A_{Y+Z})) = 0$ if and only if $A_Y = A_{Y+Z}$ almost surely. Similarly, $A_Z = A_{Y+Z}$ almost surely, since $(Y - m) \neq 0$ almost surely on \mathbb{R} . Finally, $\{U_Y \geq \alpha\} = \{U_{Y+Z} \geq \alpha\} = \{U_Z \geq \alpha\}$. ■

Lemma 5.2. *Let \mathcal{R} be a d -dimensional connected subset of \mathbb{R}^d , denote by $\mathring{\mathcal{R}}$ its interior and let ϕ_1, ϕ_2 be two non-zero affine forms on \mathbb{R}^d . If $\phi_1\phi_2 \geq 0$ on \mathcal{R} , and $\mathring{\mathcal{R}} \cap \ker(\phi_i) \neq \emptyset$ for $i = 1, 2$, then $\phi_1 = a\phi_2$ for some $a > 0$.*

Proof. Suppose $\ker(\phi_1) \neq \ker(\phi_2)$, which are two $(d-1)$ -dimensional affine hyperplanes. There exists $x \in \mathring{\mathcal{R}} \cap (\ker(\phi_1) \setminus \ker(\phi_2))$, so $\phi_1(x) = 0$ and $\phi_2(x) \neq 0$. Assume $\phi_2(x) < 0$. Then, there exists an open neighbourhood $v(x) \subset \mathcal{R}$ such that for all $y \in v(x)$, $\phi_2(y) < 0$. Since $v(x)$ is open in \mathbb{R}^d and $\ker(\phi_1)$ is $(d-1)$ -dimensional, there exists $y' \in v(x)$ such that $\phi_1(y') > 0$. This implies $\phi_1(y')\phi_2(y') < 0$, contradicting the assumption. A symmetric argument applies if $\phi_2(x) > 0$. Therefore, $\ker(\phi_1) = \ker(\phi_2)$, and it follows that $\phi_1 = a\phi_2$ for some $a > 0$. ■

Proof of Proposition 3.13. The function $\boldsymbol{\gamma} \mapsto C_\alpha(\boldsymbol{\gamma})$ is convex on \mathbb{R}^d due to the sub-additivity property of the Conditional Value-at-Risk. We now aim to prove that this function is strictly convex on the unit sphere. To this end, let $\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2 \in \mathbb{R}^d \setminus \{0\}$ be such that, for all $\mu \in (0, 1)$, the following equality holds:

$$C_\alpha(\mu\boldsymbol{\gamma}_1) + C_\alpha((1-\mu)\boldsymbol{\gamma}_2) = C_\alpha(\mu\boldsymbol{\gamma}_1 + (1-\mu)\boldsymbol{\gamma}_2). \quad (5.1)$$

We adopt the notations from Lemma 5.1 and recall that

$$C_\alpha(\boldsymbol{\gamma}) = \frac{1}{1-\alpha} \mathbb{E} \left[-\boldsymbol{\gamma}^T \mathbf{X} \cdot \mathbb{1}_{\{U_{-\boldsymbol{\gamma}^T \mathbf{X}} \geq \alpha\}} \right].$$

Let $Y := -\mu\boldsymbol{\gamma}_1^T \mathbf{X}$ and $Z := -(1-\mu)\boldsymbol{\gamma}_2^T \mathbf{X}$. Then equality (5.1) can be rewritten as

$$\mathbb{E}(YA_Y) + \mathbb{E}(ZA_Z) = \mathbb{E}((Y+Z)A_{Y+Z}),$$

Therefore, by Lemma 5.1 and (P3), we have

$$\{U_Y \geq \alpha\} = \{U_Z \geq \alpha\} = \{U_{Y+Z} \geq \alpha\}.$$

These events can be rewritten as

$$\{U_Y \geq \alpha\} = \{Y \geq \text{VaR}_\alpha(Y)\} = \{-\boldsymbol{\gamma}_1^T \mathbf{X} \geq t_1\}, \quad \text{with } t_1 := \text{VaR}_\alpha(-\boldsymbol{\gamma}_1^T \mathbf{X}),$$

and similarly for Z , with threshold t_2 . Therefore,

$$\mathbb{P} \left((-\boldsymbol{\gamma}_1^T \mathbf{X} - t_1)(-\boldsymbol{\gamma}_2^T \mathbf{X} - t_2) \geq 0 \right) = 1.$$

Define $\phi_i(\mathbf{x}) := -\boldsymbol{\gamma}_i^T \mathbf{x} - t_i$ for $i = 1, 2$. Then $\phi_1 \phi_2 \geq 0$ almost surely on $\mathcal{R}_{\mathbf{X}}$. Since $t_i \in \text{supp}(-\boldsymbol{\gamma}_i^T \mathbf{X})$, the level sets $\ker(\phi_i)$ intersect the interior of the support. By Lemma 5.2, this implies $\phi_1 = a \phi_2$ for some $a > 0$, hence $\boldsymbol{\gamma}_1 = a \boldsymbol{\gamma}_2$. Thus the two candidates belong to the same ray, reflecting that C_α is linear along rays and only strictly convex once directions are normalized on the unit sphere. \blacksquare

5.3 The Gaussian case resolution

We provide here the proof of some theoretical results used in Section 4.2.

The Lagrangian of (1.1) in the Gaussian case can be written as:

$$\begin{aligned} \mathcal{L}(\boldsymbol{\gamma}, (\lambda, \bar{\boldsymbol{\mu}}, \underline{\boldsymbol{\mu}})) &:= -\boldsymbol{\gamma}^T \mathbf{1}_d + \lambda(-\boldsymbol{\gamma}^T \mathbf{1}_d + T_Z \sqrt{\boldsymbol{\gamma}^T \Sigma \boldsymbol{\gamma}} - K) \\ &\quad + \bar{\boldsymbol{\mu}}(\boldsymbol{\gamma} - \boldsymbol{\gamma}^{\text{up}}) - \underline{\boldsymbol{\mu}}(\boldsymbol{\gamma} - \boldsymbol{\gamma}^{\text{low}}). \end{aligned} \quad (5.2)$$

$\frac{\partial \mathcal{L}}{\partial \lambda} = 0$ implies

$$K = -\boldsymbol{\gamma}^T \mathbf{1}_d + T_Z \sqrt{\boldsymbol{\gamma}^T \Sigma \boldsymbol{\gamma}}, \quad (5.3)$$

and $\frac{\partial \mathcal{L}}{\partial \boldsymbol{\gamma}} = 0$ implies

$$\left(\frac{1 - (\bar{\boldsymbol{\mu}} - \underline{\boldsymbol{\mu}})}{\lambda} + 1 \right) \mathbf{1}_d = T_Z \frac{\Sigma \boldsymbol{\gamma}}{\sqrt{\boldsymbol{\gamma}^T \Sigma \boldsymbol{\gamma}}}. \quad (5.4)$$

If the optimal solution does not activate the bounds, then $\bar{\boldsymbol{\mu}} = \underline{\boldsymbol{\mu}} = 0$. Let $y = \frac{1}{\lambda} + 1$ and define $\sigma_S = \sqrt{\boldsymbol{\gamma}^T \Sigma \boldsymbol{\gamma}}$.

Multiplying equation (5.4) on the left by $\boldsymbol{\gamma}^T$, we obtain:

$$\frac{y}{T_Z} \boldsymbol{\gamma}^T \mathbf{1}_d = \frac{\boldsymbol{\gamma}^T \Sigma \boldsymbol{\gamma}}{\sqrt{\boldsymbol{\gamma}^T \Sigma \boldsymbol{\gamma}}},$$

so

$$\frac{y}{T_Z} \boldsymbol{\gamma}^T \mathbf{1}_d = \sigma_S. \quad (5.5)$$

Multiplying (5.4) on the left by $\mathbf{1}_d^T \Sigma^{-1}$, we obtain:

$$y(\mathbf{1}_d^T \Sigma^{-1} \mathbf{1}_d) = \frac{T_Z}{\sigma_S} \mathbf{1}_d^T \boldsymbol{\gamma}.$$

Now we replace $\mathbf{1}_d^T \boldsymbol{\gamma}$ by $\frac{T_Z}{y} \sigma_S$ with (5.5) and isolate y ,

$$y = \frac{T_Z}{\sqrt{\mathbf{1}_d^T \Sigma^{-1} \mathbf{1}_d}}. \quad (5.6)$$

In the same way, we replace $\mathbf{1}_d^T \boldsymbol{\gamma}$ by $\frac{T_Z}{y} \sigma_S$ with (5.5) in (5.3),

$$K = -\frac{T_Z}{y} \sigma_S + T_Z \sigma_S,$$

and replace $\frac{T_Z}{y}$ by $\sqrt{\mathbf{1}_d^T \Sigma^{-1} \mathbf{1}_d}$ with (5.6) and isolate σ_S ,

$$\sigma_S = K \left(\frac{1}{T_Z - \sqrt{\mathbf{1}_d^T \Sigma^{-1} \mathbf{1}_d}} \right). \quad (5.7)$$

We now substitute the expressions for y and σ_S into equation (5.4):

$$\frac{T_Z}{\sqrt{\mathbf{1}_d^T \Sigma^{-1} \mathbf{1}_d}} \mathbf{1}_d = T_Z \frac{\Sigma \boldsymbol{\gamma}}{K \left(\frac{1}{T_Z - \sqrt{\mathbf{1}_d^T \Sigma^{-1} \mathbf{1}_d}} \right)},$$

finally

$$\boldsymbol{\gamma} = K \left(\frac{1}{T_Z - \sqrt{\mathbf{1}_d^T \Sigma^{-1} \mathbf{1}_d}} \right) \frac{\Sigma^{-1} \mathbf{1}_d}{\sqrt{\mathbf{1}_d^T \Sigma^{-1} \mathbf{1}_d}}. \quad (5.8)$$

References

- [1] Siddharth Alexander, Thomas F. Coleman, and Yuying Li. “Minimizing CVaR and VaR for a Portfolio of Derivatives”. In: *Journal of Banking & Finance* 30.2 (2006), pp. 583–605.
- [2] Pavlo Krokmal, Jonas Palmquist, and Stanislav Uryasev. “Portfolio optimization with conditional value-at-risk objective and constraints”. In: *Journal of risk* 4 (2002), pp. 43–68.
- [3] Harry Markowitz. “Modern portfolio theory”. In: *Journal of Finance* 7.11 (1952), pp. 77–91.
- [4] R Tyrrell Rockafellar and Stanislav Uryasev. “Optimization of conditional value-at-risk”. In: *Journal of risk* 2 (2000), pp. 21–42.
- [5] Reuven Y Rubinstein and Alexander Shapiro. *Discrete event systems: sensitivity analysis and stochastic optimization by the score function method*. Vol. 13. Wiley, 1993.
- [6] Alexander Shapiro. “Monte Carlo sampling approach to stochastic programming”. In: *ESAIM: proceedings*. Vol. 13. EDP Sciences, 2003, pp. 65–73.
- [7] Alexander Shapiro, Darinka Dentcheva, and Andrzej Ruszczyński. *Lectures on stochastic programming: modeling and theory*. SIAM, 2021.
- [8] Haoze Sun, Chengguo Weng, and Yi Zhang. “Optimal multivariate quota-share reinsurance: A nonparametric mean-CVaR framework”. In: *Insurance: Mathematics and Economics* 72 (2017), pp. 197–214.
- [9] A Alexandre Trindade et al. “Financial prediction with constrained tail risk”. In: *Journal of Banking & Finance* 31.11 (2007), pp. 3524–3538.
- [10] Wei Wang and Shabbir Ahmed. “Sample average approximation of expected value constrained stochastic programs”. In: *Operations Research Letters* 36.5 (2008), pp. 515–519.
- [11] Haixiang Yao, Zhongfei Li, and Yongzeng Lai. “Mean-CVaR portfolio selection: A nonparametric estimation framework”. In: *Computers & Operations Research* 40.4 (2013), pp. 1014–1022.