



HAL
open science

Utility Maximization Problem with Uncertainty and Stochastic Algorithms for Systemic Risk Measures

Achraf Tamtalini

► **To cite this version:**

Achraf Tamtalini. Utility Maximization Problem with Uncertainty and Stochastic Algorithms for Systemic Risk Measures. Optimization and Control [math.OC]. Le Mans Université, 2022. English. NNT : 2022LEMA1030 . tel-04378581

HAL Id: tel-04378581

<https://theses.hal.science/tel-04378581>

Submitted on 8 Jan 2024

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.

THESE DE DOCTORAT DE

LE MANS UNIVERSITE

ECOLE DOCTORALE N° 601
Mathématiques et Sciences et Technologies
de l'Information et de la Communication
Spécialité : *Mathématiques*

Par :

Achraf Tamtalini

« Utility Maximization Problem With Uncertainty and Stochastic Algorithms For Systemic Risk Measures »

Thèse présentée et soutenue à Le Mans le 02 décembre 2022
Unité de recherche : Laboratoire Manceau de Mathématiques
Thèse N° : 2022LEMA1030

Rapporteurs avant soutenance :

Huyên PHAM
Jianfeng ZHANG

Professeur (Université Paris Cité)
Professeur (University of Southern California)

Composition du Jury :

Examineurs : Gilles PAGES
Caroline HILLAIRET
Nizar TOUZI

Professeur (Sorbonne Université)
Professeur (ENSAE)
Professeur (Ecole Polytechnique)

Dir. de thèse : Anis MATOUSSI
Co-dir. de thèse : Sarah KAAKAI

Professeur (Le Mans Université)
Maître de conférences (Le Mans Université)

Remerciements

Il est temps pour moi de clôturer la rédaction de cette thèse en écrivant quelques mots de remerciements. Evidemment, cela passera par remercier tout d'abord mon directeur de thèse, Anis Matoussi, qui m'a accordé sa confiance en acceptant de superviser ce travail. Il a toujours soutenu mes différentes initiatives et su m'encourager durant les moments difficiles de cette thèse. Ses qualités humaines et scientifiques m'ont été d'un grand recours. Donc, j'aimerais à travers ces quelques mots lui exprimer ma gratitude.

Je ne manquerai pas de remercier ma co-directrice de thèse Sarah Kaakai, qui a été présente durant ces trois années par ses remarques et ses idées pertinentes. Je voudrais ici lui exprimer ma reconnaissance pour le temps qu'elle m'a consacré.

Je voudrais ensuite remercier Huyên Pham et Jianfeng Zhang de m'avoir fait l'honneur de rapporter cette thèse, malgré leurs emplois du temps très chargés. Je tiens aussi à remercier Gilles Pagès et Caroline Hillairet qui ont accepté de faire partie du jury de la thèse en tant qu'examineurs. Une pensée particulière à Nizar Touzi, sans qui cette grande aventure de thèse n'existerait pas, car c'est lui qui m'a mis en contact avec Anis Matoussi.

Je pense aussi à tous les membres du laboratoire Manceau de Mathématiques et de l'institut du risque et de l'assurance avec qui j'ai passé ces trois années (ou presque, crise sanitaire l'oblige!). Je suis reconnaissant pour leur accueil et leur sympathie.

Pour finir, et un peu plus personnellement, je ne saurai remercier mes proches pour leur soutien et leurs encouragements durant tout mon parcours académique. Je remercie mon père et ma belle-mère pour leur soutien et leur amour inconditionnel. L'accomplissement de ce travail est en grande partie dû à eux. Je remercie également mon frère Mehdi, qui malgré la distance qui nous sépare, est toujours présent dans ma vie. Puis, je voudrais dire merci à mon petit frère Saad, qui égayait mes jours dans les moments difficiles et qui était toujours soucieux de l'état d'avancement de ma thèse malgré son jeune âge. J'aimerais aussi remercier tous mes amis avec qui j'ai passé des moments agréables: Amine, Mohamed et Zakaria. Un dernier mot pour ma copine Salma, sur qui je peux toujours compter, et qui a su m'encourager et me supporter durant ces trois années. Je garde pour la fin la personne à qui je dédie cette thèse: ma mère. Je sais que de là où tu es, tu seras fière de moi !

TABLE OF CONTENTS

| | | |
|----------|---|-----------|
| 1 | Introduction | 7 |
| 1.1 | Robust Utility Maximization Problems | 8 |
| 1.1.1 | Utility maximization theory | 8 |
| 1.1.2 | Robust utility maximization problems | 10 |
| 1.1.3 | Main contributions | 12 |
| 1.2 | Systemic Risk Measures and their Estimation using Stochastic Algorithms | 19 |
| 1.2.1 | Univariate Monetary Risk Measures | 19 |
| 1.2.2 | Multivariate risk measures | 22 |
| 1.2.3 | Numerical approximations for risk measures | 25 |
| 1.2.4 | Main contributions | 27 |
| 1.3 | Multivariate Optimized Certainty Equivalent Risk Measures | 30 |
| 1.3.1 | Univariate Optimized Certainty Equivalent Risk Measures | 30 |
| 1.3.2 | Main contributions | 34 |
| 1.4 | Future research perspectives | 39 |
| 2 | Robust Utility Maximization Problem in a Jump Setting | 41 |
| 2.1 | Introduction | 41 |
| 2.2 | Framework of the optimization problem | 43 |
| 2.2.1 | Setting and notations | 43 |
| 2.2.2 | The optimization problem | 45 |
| 2.3 | Some helpful estimates and existence of optimal probability | 48 |
| 2.3.1 | Auxiliary estimates | 48 |
| 2.3.2 | Existence of optimal probability | 51 |
| 2.4 | Related BSDE with jumps | 57 |
| 2.5 | Appendix | 67 |
| 3 | Estimation of Systemic Shortfall Risk Measures using Stochastic Algorithms | 71 |
| 3.1 | Introduction | 71 |
| 3.2 | About Multivariate Risk Measures | 74 |
| 3.3 | Multivariate Systemic Risk Measures and Stochastic Algorithms | 78 |
| 3.3.1 | Properties of * | 79 |
| 3.3.2 | Almost Surely Convergence | 81 |

| | | |
|-------|--|-----|
| 3.3.3 | Asymptotic normality..... | 81 |
| 3.3.4 | Polyak-Ruppert Averaging principle | 83 |
| 3.3.5 | Estimator of asymptotic variance | 84 |
| 3.4 | Numerical examples | 86 |
| 3.4.1 | Toy example | 86 |
| 3.4.2 | Second example..... | 91 |
| 3.5 | Appendix | 98 |
| 3.5.1 | ODE method and related concepts..... | 98 |
| 3.5.2 | Closed Formulas for the first example..... | 106 |

| | | |
|----------|--|------------|
| 4 | Multivariate Optimized Certainty Equivalent Risk Measures and their Numerical Computation | 107 |
| 4.1 | Introduction | 107 |
| 4.2 | Multivariate OCE..... | 110 |
| 4.3 | Computational aspects..... | 122 |
| 4.4 | Numerical Analysis and Examples..... | 128 |
| 4.4.1 | A first example | 129 |
| 4.4.2 | Second example..... | 131 |
| 4.5 | Appendix: Estimation of MNIG parameters..... | 135 |
| 4.5.1 | Computational aspects..... | 135 |
| 4.5.2 | Numerical aspects..... | 138 |

INTRODUCTION

This thesis deals with two but closely related topics: Robust utility maximization problem and multivariate risk measures. More precisely, in the first part, we address a utility maximization problem under model uncertainty in a jump setting. In the second part, we are interested in systemic risk measures and their numerical approximation using stochastic algorithms.

In the first section of this introduction, we start by giving a general background on the utility maximization theory and explain how these problems can be solved with the help of BSDEs. We then expose the extension of these problems into robust ones where the uncertainty in choosing the pricing probability is taken into account. Our main contribution consists in extending the work of Bordigoni et al. [18] in which they work under continuous filtration, into a model that allows for jumps.

The second section is concerned with the problem of numerical approximation of systemic risk measures. More precisely, we are interested in the numerical approximation of the multivariate shortfall risk measures (MSRM) that were introduced in Armenti et al. [4]. Being inspired by the work of Bardou et al. [6] on the approximation of VaR and CVaR, we show how stochastic algorithms can be an efficient tool to estimate MSRM and propose estimators for the latter with the corresponding central limit theorems.

Lastly, we propose in the third section an extension into a multivariate setting of Optimized Certainty Equivalent (OCE) risk measures. The framework of OCE risk measure allows to retrieve most of the classical risk measures such as CVaR, entropic risk measure, etc. We then study the properties of this new type of multivariate risk measures and propose stochastic algorithms to approximate them.

1.1 Robust Utility Maximization Problems

1.1.1 Utility maximization theory

In a complete market, under assumptions of no arbitrage, there exists a unique probability measure, often called the risk neutral probability or the equivalent martingale measure (see Harrison and Pliska [64] and Delbaen and Schachermayer [37]), such that the price of any claim at a future time T is the expectation under this unique probability of its discounted payoff. Moreover, every claim can be hedged perfectly using an auto-financing portfolio composed of tradeable assets available in the market. However, in the case of constrained markets, where for instance, the agent cannot have a negative number of shares or the agent's investment in risky assets is not allowed to exceed a certain threshold, perfect hedging is no longer possible. Therefore, one cannot define in a proper way the price of a given contingent claim. Superhedging is one of the possible solutions in this case. The superreplicating price of a claim is defined as the minimal investment that an agent has to put in order to find a strategy that dominates the payoff of the claim. El Karoui and Quenez [45] characterized this superreplicating price as the essential supremum over the set of equivalent martingale measures of the expectation of the discounted payoff. In Cvitanić and Karatzas [34], it has been proved that the superreplicating price is actually bounded between two values h^{low} and h^{up} and that any price in the open interval (h^{low}, h^{up}) is arbitrage free. However, superreplicating price approach suffers from giving high prices. Another approach that provides a satisfactory answer to this problem was proposed in Hodges and Neuberger [66] and Clewlow and Hodges [31] and is based on maximizing the expected utility of the agent. Since then, many authors have used this approach usually with the exponential utility function (see for instance, Rouge and El Karoui [116], Delbaen et al. [36], Becherer [9], Musiela and Zariphopoulou [102] and Mania and Schweizer [93]). In their approach, an agent who has sold a vanilla option $C(S_T)$ is trying to replicate the outcomes from this claim at a future date T . Therefore, the agent is trying to solve the following optimization problem:

$$V(x) := \sup_{\pi \in \text{ffi}} E [U(X_T^{\pi,x} - C(S_T))], \quad (1.1)$$

where $X^{\pi,x}$ is an auto-financing hedging portfolio having the initial value x . The expectation is taken under the reference/historical probability denoted P , ffi is the set of admissible strategies and U is a utility function. They also defined two types of prices in their setting: the selling price and the buying price. The selling price is defined as the smallest price required to provide the same expected utility as not selling the option. In the same way, the buying price is defined as the maximum price worth paying to buy the option. These prices are called *indifference* prices and were studied in depth in Carmona [24]. The value function of their problem is obtained by the application of the dynamic programming principle which lead them to solving a partial

derivative equation (PDE). In the literature, we can find also another type of problems related to maximizing expected utility from terminal wealth and consumption. These problems are by now well known, understood and take in general the following form:

$$V(x) = \sup_{(\pi, c) \in \text{ffi}} E \int_0^T U_1(t, c(t)) dt + E[U_2(X_T^{\pi, c, x})], \quad (1.2)$$

where $U_2(\cdot)$, $U_1(t, \cdot)$ are utility functions for each $t \in [0, T]$, $c(\cdot)$ is the consumption rate and (X_t) is the wealth process. The first term reflects the utility coming from the consumption and the second one from the terminal wealth. For the case where the portfolio is constrained to take values in a given closed convex set, we refer to the work of Cvitanić and Karatzas [33] and for the unconstrained one, see Karatzas et al. [76], Karatzas [75], Cox and Huang [32] and Harrison and Pliska [63].

Utility maximization problems have been extensively studied by numerous authors. In the framework of a continuous-time model, they were studied for the first time by R. Merton in two seminal papers Merton [96] and Merton [97] in complete markets. Using techniques of stochastic control, a nonlinear PDE/Bellman equation for the value function was derived and closed formulas were given for special cases of utility function such as the power, logarithm, and the exponential case. The PDE method relies mainly on the assumption of Markov state processes. However, when this is not the case, there exists two approaches to solve utility maximization problems in the literature. The first one goes back to Bismut [17] and is based on duality and convex optimization (see also Deelstra et al. [35], Karatzas et al. [77], Schied and Wu [120], Källblad et al. [74] or Žitković [133]). The second approach is based on stochastic control and the powerful tool of BSDEs. BSDEs were introduced the first time in Bismut [17] for the case with linear generator, but the real starting point of the theory as it is known today is the article of Pardoux and Peng [106]. According to these authors, a solution of a BSDE consists in a pair of adapted processes (Y, Z) which satisfies:

$$dY_t = -f(t, Y_t, Z_t)dt + Z_t dW_t, \quad Y_T = \xi, \quad (1.3)$$

where f is called the generator and ξ the terminal condition. The pair (f, ξ) is usually called the data of the BSDE (for an overview on BSDEs, see Zhang [132]). This type of equations naturally appears in many problems in financial mathematics. One of the first applications of BSDE is the hedging problem of a contingent claim ξ at future time T . In a complete market, it is possible to construct a replication portfolio Y whose final value is ξ such that the dynamics of Y are given by a BSDE with a linear generator f and the price of the claim at time t is given by Y_t by non-arbitrage conditions. In incomplete markets, as said before, it is not always possible to construct a portfolio who have a the same value as the payoff ξ . El Karoui and Quenez [46]

considered superhedging strategies and defined the upper price as the smallest investment that allows to superhedge the claim ξ . Although the upper price itself does not verify a BSDE, they showed that it can be written as the increasing limit of solutions of nonlinear BSDEs.

When it comes to utility maximization problems, BSDEs can be of a great help. In fact, duality methods relies on the assumption that the trading strategies take values in a convex set, so if we do not want to restrict our strategies to convex sets, one might use BSDEs techniques and *Bellman Optimality Principle* to solve the problem. Duffie and Epstein [40] presented a stochastic differential formulation of recursive utility as the objective function of a problem of investment-consumption. Recursive utility is an extension of standard utility where the instantaneous utility does not depend only on the consumption rate (c_t) but also on the future utility. Besides, problems with standard utilities correspond to linear BSDEs whereas recursive ones correspond to nonlinear BSDEs. For the particular case of exponential utility function, Rouge and El Karoui [116] addressed the problem of hedging for an investor who is selling a claim while performing maximization of his expected utility and applied dynamic control techniques to obtain explicit representation of the optimal wealth and portfolios through solutions to a BSDE with quadratic growth. More recently, by assuming the trading strategies to take values only in a closed set, Hu et al. [68] used rather a simpler approach (not relying on *Bellman Optimality Principle*): In order to obtain the value function and optimal strategy, they constructed a process R^π such that for each π , R^π is a supermartingale with terminal value equals to the utility of the trader's terminal wealth. They also showed that there exists at least one particular strategy π^* such that R^{π^*} is a martingale and of course this particular strategy must be the optimal one. Their work was taken under the Brownian setting and therefore they used BSDEs to construct the optimal strategy π^* . As to problems with jumps, we can cite for instance Morlais [100], Morlais [98] and Becherer [8] who solved the standard utility maximization problem in a jump setting using dynamic programming. However, one should note that these works assumed a final wealth that is bounded.

1.1.2 Robust utility maximization problems

All the papers mentioned above considered a model where beliefs are represented by a single probability, the reference probability P . This supposes the perfect knowledge of the probability distribution governing the market evolution which is obviously something not true in practice. In fact, agents have different models, each of them being only an approximation of the "true" underlying one. Therefore, we need a novel approach that accounts for the possibility that a specific probabilistic model may not be correct but only an approximation. In these cases, before solving his utility maximization problem, the agent has to solve an intermediate problem of choosing the pricing probability for his expected utility. In other words, the agent tries to take into account all possible scenarios he might face and among these, he maximizes his utility

under the “worst” one. The problems where the agent is facing some uncertainty in choosing his pricing probability are called *Robust Utility Maximization Problems* or *Utility Maximization Problems with Ambiguity*. The optimization problem the agent need to solve becomes:

$$\sup_{\pi} \inf_{Q \in \mathcal{Q}} E_Q [U(X_T^\pi - F)], \quad (1.4)$$

where π runs through a set of strategies and $Q \in \mathcal{Q}$ through a set of appropriate models. The case of the non robust optimization problem in (1.1) is obtained by taking the set $\mathcal{Q} = \{P\}$. In Quenez [109], they studied the sup-inf problem above, and showed that, under some suitable conditions, there exists a saddle point for this problem using a duality approach. A complete solution in a complete market was given in Schied [119] under the condition that the set \mathcal{Q} admits the so-called “least-favorable” measure Q_0 and that the probabilities in \mathcal{Q} are all equivalent to P . In this setting, the sup-inf problem becomes then equivalent to the standard utility maximization problem under Q_0 . Schied and Wu [120] solved the problem without assuming that the priors are equivalent to the reference probability P which allows to handle many robust utility functional such as AVaR. However, this comes with a price: dropping the equivalence of prior models may lead to arbitrage strategies under the “least favorable” model. For more references on utility maximization problems of type (1.4), see Talay and Zheng [128], Föllmer and Gundel [49] and Hernández-Hernández and Schied [65] among many others.

Robust decision makers might want their decisions to be insensitive to specification errors, whereas in the optimization problem in (1.4), we are not taking them into account. Motivated by this, Maccheroni et al. [91] suggested modeling agent preferences by adding a relative entropy penalty function γ to the problem in (1.4), that is,

$$\sup_{\pi} \inf_{Q \in \mathcal{Q}} E_Q [U(X_T^\pi - F)] + \gamma(Q). \quad (1.5)$$

More generally, the penalty term $\gamma(Q)$ measures a “distance” between the probability Q and the reference probability P and will penalize certain scenarios against others. In Anderson et al. [3] and Hansen et al. [62], discounted entropic penalty was used in a Markovian setting and hence Hamilton-Jacobi-Bellman (HJB) equations were derived to solve the problem. In Bordigoni et al. [18], they addressed the robust utility maximization problem in a continuous setting which is non Markovian by using stochastic control techniques and BSDEs. More precisely, they studied the following problem:

$$\sup_{\pi} \inf_{Q \in \mathcal{Q}} U(\pi, Q), \quad (1.6)$$

where again π runs through a set of strategies and $Q \in \mathcal{Q}$ through a set of appropriate models, and the process U is defined as follows:

$$U(\pi, Q) := E_Q[U_{0,T} + \beta R_{0,T}(Q)],$$

where the first term in the expectation corresponds to the utility term coming from consumption and final wealth and the second term is the entropic penalty term. The factor β might be seen as a parameter that measures the degree of confidence of the reference probability P : The higher β is, the more confident we are in P . The main novelty in the work of Bordigoni et al. [18] is that they only assumed the utility term to be only exponentially integrable and not bounded as it is assumed usually. In Faidi et al. [47], they generalized the work of Bordigoni et al. [18] to other types of penalty terms by using f-divergence penalties and time-consistent penalties in a Brownian filtration.

1.1.3 Main contributions

Motivation. There are few works studying robust utility maximization problems in a jump setting. First, there is the work of Jeanblanc et al. [72] in which they solve the same problem as in Bordigoni et al. [18] in a discontinuous filtration with a one-point jump distribution and an entropic penalty. Also, Laeven and Stadje [84] addressed a robust utility maximization problem in a jump setting with general penalty but they assumed the payoffs to be bounded. The boundedness condition might be very restrictive. Indeed, payoffs as simple as $(W_T)^+$ or $|W_T|$, where W is a Brownian motion and T is a final time, do not verify this condition. In our first work, in order to estimate the value function from (see [HYPERLINK \l ""_bookmark165" 4.20](#)) to have a good approximation of the value function to be bounded but to only be exponentially integrable. This obviously comes with a price: before proving the existence of an optimal pricing probability, we first need to establish some estimates. This is crucial to prove the well-posedness of the problem and to show that some key quantities of the problem are well controlled (see section 2.3 for more details).

Now, we start by introducing the setting of our work and some notations. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a filtered probability space with finite time horizon $T < \infty$ and a filtration \mathbb{F} satisfying the usual conditions of right continuity and completeness, in which all semimartingales are taken to have right continuous paths with left limits. We assume that \mathcal{F}_0 is trivial and $\mathbb{F} = \mathbb{F}_T$. On this stochastic basis, let W be a d -dimensional standard Brownian motion and let $\mu(dt, dx) = (\mu(w, dt, dx) | w \in \Omega)$ denote an integer-valued random measure on $([0, T] \times E, \mathcal{B}([0, T]) \otimes E)$ with $E := \mathbb{R}^d \setminus \{0\}$ equipped with its Borel σ -field $\mathcal{E} := \mathcal{B}(E)$. Let $\tilde{\Omega} := \Omega \times [0, T] \times E$ and $\tilde{\mathcal{P}} := \mathcal{P} \otimes \mathcal{E}$ where \mathcal{P} is the predictable σ -field on $\Omega \times [0, T]$. We say that a function on $\tilde{\Omega}$ is predictable if it is $\tilde{\mathcal{P}}$ -measurable. We denote by $\nu := \nu^p(w, dt, dx)$ the compensator of μ under \mathbb{P} . We will assume that the compensator is absolutely continuous with respect to $\lambda \times dt$ with a density ξ . More

precisely, we assume that,

$$v(w, dt, dx) = \xi_t(w, x)\lambda(dx)dt, \quad (1.7)$$

where λ is a σ -finite measure on (E, \mathcal{E}) , that satisfies the following condition: $\int_E 1 \wedge |x|^2 \lambda(dx) < \infty$ and the density ξ is positive and bounded:

$$0 < \xi_t(w, x) \leq C_v < \infty, \quad P \times \lambda(dx)dt - a.e. \text{ for some constant } C_v. \quad (1.8)$$

Let μ denote the compensated measure of μ :

$$\tilde{\mu} := \mu(w, dt, dx) - v(w, dt, dx). \quad (1.9)$$

To alleviate the notation, the stochastic integrals with respect to W , $\tilde{\mu}$ and v will be denoted as

follows:

$$\left(\int_0^t \eta_s dW_s \right) := \int_0^t \eta_s dW_s, \quad \left(\int_0^t \psi_s \tilde{\mu}_s \right) := \int_0^t \int_E \psi_s(x) \tilde{\mu}(ds, dx), \quad \left(\int_0^t \psi_s v_s \right) := \int_0^t \int_E \psi_s(x) v(ds, dx),$$

for η progressively measurable process and ψ a predictable process verifying $\int_0^T |\eta_s|^2 ds < \infty$ and $(|\psi|^2 \otimes v)_T < \infty, P - a.s.$ We will also assume that W and $\tilde{\mu}$ satisfy the following weak representation property with respect to P and \mathcal{F} : Every local martingale M with respect to (\mathcal{F}, P) admits the following decomposition:

$$M_t = M_0 + (\eta \cdot W)_t + (\psi \otimes \tilde{\mu})_t, \quad \forall t \geq 0, P - a.s. \quad (1.10)$$

As mentioned previously, we will work in a more flexible framework under which we assume that the final wealth is only exponentially integrable and not necessarily bounded. We introduce the following spaces:

- L^{exp} is the space of all \mathcal{F}_T -measurable random variables X such that:

$$E[\exp(\gamma|X|)] < \infty, \quad \forall \gamma > 0.$$

- D^{exp} is the space of progressively measurable processes $(X_t)_{0 \leq t \leq T}$ with

$$E \exp \gamma \text{ess sup}_{0 \leq t \leq T} |X_t| < \infty, \quad \forall \gamma > 0.$$

- $D_{\mathcal{F}}^{\text{exp}}$ is the space of progressively measurable processes $(X_t)_{0 \leq t \leq T}$ with

$$E \exp \gamma \int_0^T |X_s| ds < \infty, \quad \forall \gamma > 0.$$

- $H_\lambda^{2,p}$ is the space of predictable processes ψ such that

$$E \int_0^T |\psi_{s,\lambda}^p|^2 ds < \infty,$$

where

$$|\psi_{s,\lambda}^p| := \int_E |\psi_s(x)|^2 \xi_s(x) \lambda(dx).$$

- $H^{2,p}$ is the set of all predictable processes η such that

$$E \int_0^T |\eta_s|^2 ds < \infty.$$

Now that we have introduced all the notations, we can state our robust optimization problem. Given a probability $Q \ll P$ on F_T , we denote by $D = (D_t)_{0 \leq t \leq T}$ its Radon-Nikodym density with respect to P , that is,

$$D_t = E \frac{dQ}{dP} \cdot F_t, \quad t \geq 0.$$

We will identify the probability $Q \ll P$ with the density process (D_t) . (D_t) is càdlàg nonnegative P -martingale and we can show that it can be represented as exponential martingale, i.e., there exists predictable processes η and ψ such that,

$$D_t = E((\eta \cdot W) + (\psi \cdot \mu))_t, \quad t \geq 0, Q - a.s. \tag{1.11}$$

The case $\eta \equiv 0$ and $\psi \equiv 0$ corresponds to the reference probability P . We will be penalizing the probability Q through the processes η and ψ , that is we consider a time-consistent penalty function of the following form:

$$\gamma_t^Q := E_Q \int_t^T r(s, w, \eta_s, \psi_s) ds \cdot F_t, \tag{1.12}$$

where $r : [0, T] \times \Omega \times \mathbb{R}^d \times L^2(E, \lambda; \mathbb{R}) \rightarrow [0, \infty]$ is a suitable function that is convex and lower-semicontinuous in (η, ψ) and such that $r(t, 0, 0) = 0$. The latter condition is natural since the “distance” of Q from P is supposed to be zero. In the following, we will work with probabilities that have a finite penalty, i.e., we consider $Q \in Q_f$ where

$$Q_f := \{Q \ll P, \gamma_0(Q) < \infty\}. \tag{1.13}$$

In order to solve the optimization problem with the help of BSDE, we need to introduce some growth condition on the penalty function. In a Brownian setting, Faidi et al. [47] assumed the penalty function to be bounded from below by the relative entropy. We will keep this condition and will assume that there exists $\bar{K}_1, \bar{K}_2 > 0$ such that,

$$\gamma_0(Q) \geq -\bar{K}_2 + \bar{K}_1 H(Q|P),$$

where $H(Q|P) = E_Q \left[\int_0^T \log \frac{dQ}{dP} \right]$ is the entropy of Q with respect to P . A sufficient condition on the function r that guarantees that the above inequality is verified is the following: $\exists K_1, K_2 > 0$ such that for all $w \in \Omega, t \in [0, T], \eta \in \mathbb{R}^d$ and $\psi \in L^2(E, \lambda; \mathbb{R})$, we have,

(ff_r) There exists $K_1, K_2 > 0$ such that for all $w \in \Omega, t \in [0, T], \eta \in \mathbb{R}^d$ and $\psi \in L^2(E, \lambda; \mathbb{R})$, we have,

$$r(t, w, \eta, \psi) \geq -K_2 + K_1 \left(\frac{|\eta|^2}{2} + \int_E f(\psi(x)) \xi_t(w, x) \lambda(dx) \right)$$

where f is the function defined as follow:

$$f(x) = \begin{cases} (1 + x) \log(1 + x) - x, & \text{if } x \geq -1, \\ -\infty & \text{otherwise.} \end{cases} \tag{1.14}$$

With a particular choice of the function r , we can retrieve the case of relative entropy penalty. In fact, letting $r(t, \eta, \psi) = \frac{1}{2} |\eta|^2 + \int_E f(\psi(x)) \xi_t(x) \lambda(dx)$, we show in Proposition 2.5.2 in Chapter 2, that with this choice, the penalty γ_0 corresponds to the entropy penalty, i.e.,

$$\gamma_0(Q) = E_Q \left[\int_0^T \frac{|\eta_t|^2}{2} + \int_E f(\psi_t(x)) \xi_t(w, x) \lambda(dx) dt \right] = H_0(Q|P).$$

Now that we have introduced all the notations, we introduce the optimization problem we solve in a jump setting. Given a positive adapted process δ , we define the discounting process:

$$S_t^\delta := \exp \left(- \int_0^t \delta_s ds \right), \quad 0 \leq t \leq T,$$

and the quantities:

$$U_{t,T}^\delta := \alpha \int_t^T \frac{S_s^\delta}{S_t^\delta} U_s ds + \alpha \frac{S_T^\delta}{S_t^\delta} U_T, \quad 0 \leq t \leq T, \quad \alpha, \alpha \geq 0$$

$$R_{t,T}^\delta(Q) := \int_t^T \frac{S_s^\delta}{S_t^\delta} r(s, \eta_s, \psi_s) ds, \quad 0 \leq t \leq T,$$

We consider the cost functional c defined as:

$$c(w, Q) := U_{0,T}^\delta(w) + \beta R_{0,T}(Q)(w), \quad (1.15)$$

which consists of two terms. The first one is a utility term that itself consists of the sum of the discounted utility over $[0, T]$ with a utility rate (U_s) and a final discounted utility. The second term of the cost functional is simply a penalty term. Our goal is to solve the following optimization problem:

$$\text{Minimize the functional } Q \mapsto \Gamma(Q) := E_Q[c(\cdot, Q)], \quad (1.16)$$

over the set $Q_{\mathcal{F}}$. As said before, we will need some exponential integrability assumptions on the utility rate process (U_s) and the final utility \bar{U}_T . Namely, we will assume that:

- (ffi_u)
- i. The discounting process is bounded by some constant $\|\delta\|_\infty$;
 - ii. The process U belongs to D_1^{exp} ;
 - iii. The terminal utility \bar{U}_T belongs to L^{exp} .

Existence of an optimal probability: In order to show the existence of an optimal probability $Q^* \in Q_{\mathcal{F}}$, the starting point is to consider a minimizing sequence of probabilities Q^n and the associated densities D^n . Since each D^n is nonnegative, using a Komlòs-type argument, we can construct a sequence \bar{Q}^n that is a convex combination of the minimizing sequence (Q^n) such that \bar{D}_T^n -convergence almost surely to some \bar{D}_T^∞ . Next, by viewing the cost functional $Q \mapsto \Gamma(Q)$ as a function g of the whole density process (D_t), it is easy to see that this function g is convex. If it was moreover lower-semicontinuous with respect to P -almost surely convergence, then the optimality of \bar{Q}^∞ follows immediately. However, the lower-semicontinuity of g is not obvious at all and we must use some estimates of the functional Γ to show the optimality of \bar{Q}^∞ . This is done in the section 2.3 in chapter 2.

Link with BSDEs: We now explicit the BSDE with jumps that is verified by the dynamic value process associated with our optimization problem. By denoting S the set of stopping times with values in $[0, T]$, we define the value of the control problem started at time $\tau \in S$ instead of 0 and assuming one has used the model Q up to time τ :

$$V(Q, \tau) := Q - \operatorname{ess\,inf}_{Q \in D(Q, \tau)} \tilde{V}(Q, \tau),$$

$$\tilde{V}(Q, \tau) := E_Q [U_{\tau,T}^\delta | \mathcal{F}_\tau] + \beta E_Q [R_{\tau,T}^\delta(Q) | \mathcal{F}_\tau],$$

where $D(Q, \tau) := \{Q \in Q_{\mathcal{F}}, Q = Q \text{ on } \mathcal{F}_\tau\}$. The key point to derive the BSDE verified by the dynamic value process is to use *Bellman Optimality Principle* from stochastic control. More

precisely, we define as in Karoui [78] the minimal conditional cost at time τ by

$$J(Q, \tau) := Q - \operatorname{ess\,inf}_{Q \in \mathcal{D}(Q, \tau)} \Gamma(Q, \tau),$$

where,

$$\Gamma(Q, \tau) := E_Q[c(\cdot, Q) | \mathcal{F}_\tau].$$

Then, the *Bellman Optimality Principle* states the following:

1. The family $\{J(Q, \tau) | \tau \in S, Q \in \mathcal{Q}_f\}$ is a submartingale system, that is for any $Q \in \mathcal{Q}_f$ and stopping times $\sigma \leq \tau$, we have,

$$E_Q[J(Q, \tau) | \mathcal{F}_\sigma] \geq J(Q, \sigma) \quad Q - a.s.$$

2. $Q \in \mathcal{Q}_f$ is optimal if and only if the family $\{J(Q, \tau) | \tau \in S\}$ is a Q -martingale system which means that for any stopping times $\sigma \leq \tau$

$$E_Q[J(Q, \tau) | \mathcal{F}_\sigma] = J(Q, \sigma) \quad Q - a.s.$$

3. For each $Q \in \mathcal{Q}_f$, there exists an adapted RCLL process $J^Q = (J_t^Q)_{t \in [0, T]}$ which is a right closed Q -submartingale such that for every stopping time τ

$$J_\tau^Q = J(Q, \tau) \quad Q - a.s.$$

Then, we use the relationship between the minimal conditional cost and the process V . Indeed, by Bayes' formula and the definition of $R_{\tau, T}(Q)$, it is easy to see that $V(Q, \tau)$ depends only on the values of the density process D on $[\tau, T]$ and is therefore independent of Q . Hence, we can denote $V(Q, \tau)$ by $V(\tau)$. From the third point of martingale optimality principle, there exists an adapted RCLL process denoted $(J_t^Q)_{t \in [0, T]}$ such that $J(Q, \tau) = J_\tau^Q$, $Q - a.s$ for every stopping time τ . Comparing $V(\tau)$ and J_τ^Q for $Q \in \mathcal{Q}_f^e := \{Q \in \mathcal{Q}_f, Q \sim \tilde{P}\}$ leads us to the following relation:

$$J_\tau^Q = S_\tau^\delta V(\tau) + \alpha \int_0^\tau S_t^\delta U_t dt + \beta \int_0^\tau S_t^\delta r(t, \eta_t, \psi_t) dt, \quad dt \times dP - a.e. \quad (1.17)$$

Choosing $Q = P$ in the above equation, from the Bellman optimality principle, J^P is a P -submartingale and thus we deduce that V is a P -special martingale, i.e. its canonical decomposition can be written as:

$$V = V_0 + M^V + A^V,$$

where M^V is a local martingale and A^V is a predictable finite variation process. By the weak representation assumption, the local martingale part can be expressed as two integrals with

respect to the Brownian motion and the compensated random measure. The two first point of the Bellman optimality principle and (1.17) can tell us more about the finite variation part. This is shown in the next theorem.

Theorem 1.1.1. *Assume assumptions (ffi_u) and (ffi_r) hold. If the optimal probability \bar{Q}^∞ is equivalent to P , then (V, Z, \tilde{Z}) is solution in $D_0^{\text{exp}} \times H^{2,p} \times H_\lambda^{2,p}$ of the following BSDE:*

$$\begin{cases} \cdot dV_t = \delta_t V_t - \alpha U_t + \beta r^*(t, \frac{Z_t}{\beta}, \xi_t \frac{\tilde{Z}_t}{\beta}) dt - Z_t dW_t - \int_E \tilde{Z}_t(x) \tilde{\mu}(dx, dt), \\ \cdot V_T = \alpha U_T \end{cases} \quad (1.18)$$

where r^* is the Fenchel conjugate of r defined as:

$$r^*(t, \eta, \psi) = \sup_{\eta \in \mathbb{R}^d, \psi \in L^2(\lambda)} \left(\eta + \int_E (x)\psi(x)\lambda(dx) - r(t, \eta, \psi) \right).$$

Note that thanks to (ffi_r), denoting $f(t, v, \eta, \psi)$ the driver of the above BSDE, we have,

$$|f(t, v, \eta, \psi)| \leq \|\delta\|_\infty |v| + \alpha |U_t| + \beta K_2 + \frac{1}{2\beta K_1} |\eta|^2 + \beta K_1 \int_E f^*(\eta, \psi)(x) \xi_t(x) \lambda(dx),$$

where f^* the Fenchel conjugate of f is given by, $f^*(\eta, \psi) = e^{-\eta - \int_E (x)\psi(x)\lambda(dx)} - 1$. This is why this type of BSDE is called Quadratic-Exponential BSDEs with jumps. In Karoui et al. [79] existence of solution of this type of BSDEs is shown by constructing a sequence of exponential semimartingales which converges and using stability results to prove the convergence of the finite variation and martingale parts. The unicity in their work was proved only in the case of bounded terminal condition. We close this section by stating a comparison result which is needed to prove unicity (see Proposition 2.4.4).

Proposition 1.1.2. *Assume that for $k = 1, 2$, (V^k, Z^k, \tilde{Z}^k) is a solution of the BSDE (??) in $D_0^{\text{exp}} \times H^{2,p} \times H_\lambda^{2,p}$ associated with (U^k, U^k) . If $(U^1, U^1) \leq (U^2, U^2)$, then,*

$$\forall t \in [0, T], V_t^1 \leq V_t^2 \quad P - a.s.$$

Comparison results are usually obtained through an estimate of the quantity $((V^1 - V^2)^+)^2$ and use Itô-Tanaka's formula. In contrast, in our case, even though the terminal condition is not bounded, we were able to show the unicity of the solution of our BSDE by making use of the convexity property of the generator and estimating the quantity $V^1 - \vartheta V^2$ for $\vartheta \in (0, 1)$. Our result is an extension of the work of Briand and Hu [20] in a discontinuous setting.

1.2 Systemic Risk Measures and their Estimation using Stochastic Algorithms

Univariate risk measures are widely used in the finance and insurance industry and the theory behind is now well known and developed. Risk measures can be seen as a mapping that links each risk factor X (a random variable) to a real number representing the risk associated with this risk factor. Now, from the point of view of a risk manager of an insurance company, who observes the losses of different contracts (or the losses of different lines of business), how the global and individual risks can be assessed? In fact, the risk factor X , which represents the losses, is no longer a single random variable but a random vector. In this case, what could be the risk of it? Should it be a single amount? or a real vector? How can we compute it? In this section, we try to give answers to these questions. We first start by recalling the basics of univariate risk measures.

1.2.1 Univariate Monetary Risk Measures

Since the seminal paper of Hodges and Neuberger [66] in 1989, pricing using utility criterion has been a popular approach to price claims in incomplete markets. Let's consider the simple case of an agent with a utility function u who wants to buy a claim X . The agent is willing to pay the maximal amount $\pi(X)$ such that the expected utility if he enters into the transaction remains the same as if does not, that is:

$$E[u(X - \pi(X))] = u(0).$$

The price $\pi(X)$, which is called the *indifference price* (see Carmona [24] or Munk [101]), is not a transaction price but only an upper bound to the price of this claim. Let's be more explicit and suppose that the agent's utility is an exponential one with the following form: $u(x) = -\gamma \exp(-\frac{1}{\gamma}x)$ where γ is the risk tolerance of the agent. In order to decide whether to buy the claim or not, the agent will try to find the indifference price $\pi(X)$ by solving the above equation and will obtain that,

$$\pi(X) = -\gamma \ln E \exp \left[-\frac{1}{\gamma} X \right].$$

This indifference price $\pi(X)$ actually verifies some desirable and intuitive properties: it is increasing, convex and translation invariant in the sense that $\pi(X + m) = \pi(X) + m$. These properties are known in the literature as axioms and constitute the basis of the modern theory of risk measure.

This axiomatic approach of pricing rule was first introduced in insurance under the name of

convex premium principle (see Deprez and Gerber [38]) and was then developed in the finance context (see El Karoui and Quenez [45]). To summarize, when adopting the exponential utility framework, we end up with a functional π , which is nothing but the well known *entropic risk measure* (see for instance Föllmer and Knispel [50]). It is therefore possible to adopt a pricing approach based on risk measures which provides a suitable framework with a desirable set of properties mentioned above. This method consists in replacing the maximization utility criterion by minimizing risk exposure. The risk measure approach is the natural extension of the hedging and pricing problem in complete markets. Indeed, in a complete market, in the Black-Scholes world, there is a unique risk-neutral measure, such that with delta-hedging, the agent ends with no risk and no profit. In an incomplete market, an extension of this idea would be that the agent charges a price such that with dynamic hedging, he minimizes his risk exposure. Therefore, given a vanilla option $C(S_T)$, the problem in (1.1) is replaced by the following one:

$$V(x) := \inf_{\pi \in \text{ffi}} \eta(C(S_T) - X_T^{\pi, x}), \quad (1.19)$$

where η is a risk measure. Note that the utility maximization problem in (1.1) can also be retrieved with the risk minimizing approach. Indeed, for a financial loss X , taking $\eta(X) = E[-U(-X)]$, then it is easy to see that,

$$\text{minimizing } \eta(C(S_T) - X_T^{\pi, x}) \Leftrightarrow \text{maximizing } E[U(X_T^{\pi, x} - C(S_T))].$$

Univariate monetary risk measures have been introduced for the first time in the seminal paper of Artzner et al. [5] and have been the subject of intensive developments since then (see Frittelli and Gianin [54], Frittelli and Gianin [53], Jouini et al. [73], Acharya et al. [1] among many others). In their paper, risk measures are introduced through the *Acceptance Sets* that verifies certain axioms: Denoting $L^0(\mathbb{R})$ the space of random variables, a monetary risk measure is a map $\eta : L^0(\mathbb{R}) \rightarrow \mathbb{R}$ that is defined as the minimal cash amount that needed to be added to a financial position $X \in L^0(\mathbb{R})$ to make the resulting payoff “acceptable”, i.e,

$$\eta(X) := \inf\{m \in \mathbb{R}, X + m \in A\}, \quad (1.20)$$

where $A \subset L^0(\mathbb{R})$ is the acceptance set, is assumed to be monotone decreasing. One characterizing feature of the risk measure defined in (1.20) is the cash additivity property:

$$\eta(X + m) = \eta(X) - m.$$

In addition, the risk measure in (1.20) is said to be *positively homogeneous* if it verifies the following property:

$$\eta(\lambda X) = \lambda \eta(X).$$

Subadditivity is also one important desirable property as it reflects the fact that the risk of a diversified portfolio $X+Y$ is always less or equal to the sum of the individual risks. Risk measures verifying all the properties above are called *coherent risk measure*. In many situations, the risk of a financial position might increase in a nonlinear way (think of a liquidity risk that might arise if a position is multiplied by a large factor). This led Föllmer and Schied [57] to relax the subadditivity and positive homogeneity conditions and to require the weaker condition of convexity which reflects the effect of diversification on the risk. The corresponding risk measures are called *convex risk measures*. In Frittelli and Maggis [56], the mathematical interpretation of “diversification should not increase the risk” is translated by the weaker *quasiconvexity* property:

$$\eta(\lambda X + (1 - \lambda)Y) \leq \max(\eta(X), \eta(Y)).$$

Therefore, they introduced the *quasiconvex* risk measures by only assuming monotonicity and quasiconvexity.

Risk measures are widely used in the financial industry for risk management purposes. For instance, insurance underwriters use risk measures to compute (extra) capital allocations for their different lines of business. On the other hand, regulators use risk measures to set capital requirements for all market participants in a financial system. If there was one risk measure to cite, it would be definitely Value-at-Risk (VaR shortly) and its convex version Conditional Value-at-Risk (CVaR). This risk measure that finds its origins in the insurance industry, was introduced in the financial markets in the late 80s. Its popularity is due to its simplicity and its easy interpretation: It corresponds to the maximal loss that a portfolio can have with a fixed probability in a given time horizon.

Other risk measures used by practitioners include utility-based shortfall risk measures (SR shortly) introduced by Föllmer and Schied [57] or Optimized Certainty Equivalent (OCE) risk measures introduced by Ben-Tal and Teboulle [10]. Given a loss function and a threshold, the shortfall risk of a financial position is defined as the minimal capital to add to the position such that the expected loss of the new position is below the threshold. It can be actually seen as a risk measure with the following acceptance set:

$$A := \{X \in L^\infty, E[l(-X)] \leq \lambda\},$$

where L^∞ is the set of bounded variables, λ is a threshold and l is a loss function. In order to define a proper convex risk measure, the loss function l must be increasing, convex and not identically constant. Moreover, The restriction to bounded variable is not necessary and we can allow for more general space. We will see in chapter 3 that Orlicz spaces are the natural spaces to work under. The framework of SR is very flexible because one can choose his own loss function according to his risk aversion. As for the OCE risk measures, we will discuss them

and their multivariate extension in the third section of this introduction.

1.2.2 Multivariate risk measures

When it comes to a system of financial institutions or portfolios, the question about the global risk carried by the system as a whole entity arises. For instance, the risk manager of a diversified investment firm has to assess and control the collective risk of all the desks within the firm. An insurance company who provides risk transfers for its clients, needs to evaluate the risks of its contracts in order to set up an adequate risk sharing mechanism. From a regulatory point of view, the supervising authority of a financial market wishes to have a global view of the risk and the distribution of profits and losses of all market participants. Following the 2008 financial crisis, the traditional approach of measuring risk, that consists in considering each financial institution as a single entity isolated from other institutions, has shown its limits. This has brought awareness of the need of new approach that first models the complex structure of a financial system and provides a metric of the global risk carried by the system. Many researchers have studied the structure of financial networks and the analysis of the contagion and the spread of a potential exogenous shock into the system. Among the many contributions on this subject, we mention the contagion model in Eisenberg and Noe [44] (see also Ait-Sahalia et al. [2]) and the liquidity cascade model in Hurd et al. [69] and Lee [86]. Much of the existing literature on the systemic risk has been focused on the study of risk of firms in isolation from each other and without taking into account any dependence structure. Indeed, to assess the risk of a system of portfolios $X = (X_1, \dots, X_d)$, the traditional approach consisted in applying a univariate risk measure to each component, that is, the systemic risk measure $R(X)$ can be written as:

$$R(X) = \sum_{i=1}^d \eta_i(X_i), \quad (1.21)$$

where η_i is a univariate risk measure. Then, Chen et al. [27] proposed an approach that is very close in spirit to the axiomatic framework initiated by Artzner et al. [5]. They showed that any systemic risk measure verifying their axioms is the composition of a univariate risk measure η and an aggregation function Λ . More precisely, if $X = (X_1, \dots, X_d)$ is a random vector of losses/profits, then any function $R : L^0(\mathbb{R}^d) \rightarrow \mathbb{R}$ verifying their axioms is the composition of a univariate risk measure η and an aggregation function Λ , i.e.,

$$R(X) = (\eta \circ \Lambda)(X) = \inf\{m \in \mathbb{R}, \Lambda(X) + m \in A\}. \quad (1.22)$$

The representation in (1.22) is known as the ‘‘Aggregate then Add Cash’’ approach as it consists to first, to aggregate the outcomes X_1, \dots, X_d through the aggregation function Λ and to apply a univariate risk measure. One of the most common ways to aggregate multivariate risk is

to simply sum the risk factors, that is to take $\Lambda(x) = \sum_{i=1}^d x_i$. It is worth noticing that, while summing up profit and losses might seem reasonable from the point of view of a portfolio manager because portfolios profits and losses compensate each other, this aggregation rule seems inadequate from the point of view of a regulator where cross-subsidization between institution is not realistic since no institution will be willing to pay for losses of another one. The same argument is also valid in insurance where the principle of mutualization is flawed in the presence of heterogeneous clients. Moreover, this approach does not give the individual risk contribution of each institution in the systemic risk, preventing the regulator from taking action on the level of risky institutions.

Motivated by these considerations, Biagini et al. [15] proposed another approach to measure the systemic risk. They first considered the systemic risk as the minimal capital that secures the system by injecting capital into the single institutions, *before aggregating the individual risks*:

$$R(X) := \inf_{m_i} \left\{ \sum_{i=1}^d m_i, \Lambda(X + m) \in \mathbb{A} \right\}. \quad (1.23)$$

This approach, known as “Add Cash then Aggregate” consists in adding the amount m_i to the financial position X_i before the corresponding total loss $\Lambda(X + m)$ is computed. The systemic risk is then measured as the minimal total amount $\sum_{i=1}^d m_i$ injected into the institutions to make it acceptable. With this approach, a joint measure of total risk as well as individuals risk contributions to systemic risk is obtained. If $m^* = (m_1^*, \dots, m_d^*)$ is an optimum, that is $R(X) = \sum_{i=1}^d m_i^*$ and $\Lambda(X + m^*) \in \mathbb{A}$, one could order the m_i^* and hence be able to say that institution i requires more cash allocation or is riskier than institution j if $m_i^* \geq m_j^*$. Another main novelty in Biagini et al. [15] is that they allow adding to X not only a vector of cash m but a random vector $Y \in \mathbb{C} \subset L^0(\mathbb{R}^d)$, which will represent admissible asset with random payoffs.

Motivated by the use of risk measures in practice, Armenti et al. [4] suggested an extension of univariate SR risk measures by taking the aggregation function in (1.23) $\Lambda(x) = I_S(-x)$ where I_S is a multivariate loss function (see below) and the acceptance set $\mathbb{A} = \{X \in L^0(\mathbb{R}), E[X] \leq 0\}$. In the multidimensional setting, a multivariate loss function is defined in Armenti et al. [4] as follow:

Definition 1.2.1. A function $I_S : \mathbb{R}^d \rightarrow (-\infty, \infty]$ is called a loss function if

1. I_S is increasing, that is $I_S(x) \geq I_S(y)$ if $x \geq y$;
2. I_S is convex, lower semicontinuous with $\inf I_S < 0$;
3. $I_S(x) \geq \sum x_i - c$ for some constant c .

A loss function I_S is permutation invariant if $I_S(x) = I_S(\pi(x))$ for every permutation π of its components.

The first and second properties express the classical properties of preference and diversification in risk measure. The third one expresses a form of risk aversion: the loss function must put more weights on losses than a risk neutral evaluation. The permutation invariance property expresses the fact that the loss function should treat each component in the same way, the risk components being of the same type (think of banks or portfolios in a trading floor). One simple example of exponential loss function is the following:

$$I_S(x_1, \dots, x_d) = \frac{1}{1 + \alpha} \sum_{i=1}^d e^{\beta x_i} + \alpha e^{\beta \sum_{i=1}^d x_i} - \frac{\alpha + d}{\alpha + 1}, \quad (1.24)$$

where $\beta > 0$ is a risk aversion coefficient. α is what we call the systemic weight since by setting it to 0, the loss function becomes just a sum of individual losses.

For integrability reasons, Armenti et al. [4] worked under the Orlicz heart, that is they considered loss vectors X (that is negative values of X corresponds to actual profits) that belong to:

$$M^\vartheta := \{X \in L^0(\mathbb{R}^d) : E[\vartheta(\lambda X)] < \infty, \forall \lambda \in \mathbb{R}^+\},$$

where $\vartheta(x) = I_S(|x|)$, $x \in \mathbb{R}^d$. Now, for $X \in M^\vartheta$ a loss vector, multivariate shortfall risk measures were defined as:

Definition 1.2.2. *The multivariate shortfall risk of $X \in M^\vartheta$ is*

$$R(X) := \inf_{\substack{m_i \\ i=1}}^{\sum^d} m_i : E[I_S(X - m)] \leq 0. \quad (1.25)$$

Armenti et al. [4] also showed that the risk measure in (1.25) verifies all the classical properties of the univariate shortfall risk measure including its dual representation. Besides, they proved that optimal allocations exists and characterized them through the first order conditions as shown in the next theorem.

Theorem 1.2.3. *[Theorem 3.4 in Armenti et al. [4]]. If I_S is a permutation invariant loss function, then for every $X \in M^\vartheta$, risk allocations m^* exist. They are characterized by the first order conditions:*

$$1 \in \lambda^* E[OI_S(X - m^*)] \text{ and } E[I_S(X - m^*)] = 0, \quad (1.26)$$

where λ^* is a Lagrange multiplier. Moreover, if $I_S(x + \cdot)$ is strictly convex along zero-sums allocations for every x with $I_S(x) \geq 0$, then the risk allocation is unique.

In the next subsection, we discuss the numerical approximations of risk measures. In particular, we are interested in the computational aspects of multivariate shortfall risk measures.

1.2.3 Numerical approximations for risk measures

Motivations. In the financial industry, measuring financial risk relies on the availability of efficient algorithms for the estimation of risk measures and is one of the key issues for financial institutions and regulatory authorities. More precisely, we would like to approximate the infimum $m^* = (m_1^*, \dots, m_d^*)$. The vector m^* could have different interpretations: In insurance, it could for example serve as provisions/capital allocations for each line of business. We can also see m^* as the vector of new premiums that the insurance would require for its clients in case of a renewal after observing the losses.

In Armenti et al. [4], to obtain the optimal allocations in (1.26), they used a two-steps algorithm:

1. For each $m \in \mathbb{R}^d$, given a sample of X , approximate $E[l_S(X - m)]$ using Monte Carlo, Fourier or Chebyshev methods.
2. Use a deterministic algorithm to find the infimum.

The deterministic algorithm used to find the infimum is the sequential least squares programming algorithm (SLSQP). Even though this approach has the merit to be fast sometimes, it still has some disadvantages. First, the deterministic search algorithm for the minimum is sensitive to the starting point and one needs to be careful when setting it. Second, with this methods, we have no control over the error in the approximations and no confidence intervals are available to us.

We propose here an alternative to the above method that is based on *Stochastic Algorithms* (SA) by taking advantage of the characterizing equations in (1.26). Stochastic algorithms, which we discuss briefly in the next paragraph, are efficient tools whenever we are looking for the zeros of a function that can be expressed as an expectation, see Duflo [42] and Shapiro et al. [122].

Many practical and theoretical problems in diverse areas can be reduced to finding the zeros of a certain function. Indeed, it suffices to notice that solving many problems consists at the end in optimizing some function $F(\cdot)$. If $F(\cdot)$ is differentiable, the optimization problem reduces to finding the root of $h(\cdot) = \frac{dF(\cdot)}{d\theta}$. In the case where the function F or its derivatives can be observed directly without any errors, one could use many numerical methods for solving the problem such as deterministic gradient descent. In contrast, stochastic algorithms were motivated by problems when the function h is unknown explicitly but can be estimated. The basic paradigm in its simplest form is the following stochastic differential difference equation:

$$Z_{n+1} = Z_n + \gamma_n Y_n, \quad (1.27)$$

where Y_n is the noisy observable variable and γ_n is the step size sequence that must be chosen such as $\sum \gamma_n = \infty$ and $\sum \gamma_n^2 < \infty$. In the particular case, where $h(\cdot) = E[H(X, \cdot)]$, the

stochastic algorithms take the following form:

$$Z_{n+1} = Z_n + \gamma_n Y_n = Z_n + \gamma_n H(X_{n+1}, Z_n),$$

where (X_n) is a sequence of i.i.d random variables having the same law as X . There exists many results on the almost surely convergence of (Z_n) towards the zero $*$ of h under various conditions. Robbins and Monro [111] were the first ones to establish almost surely convergence results of stochastic algorithms, hence the name Robbins-Monro (RM) algorithm for the equation (1.27). In many cases, the analysis of these algorithms uses the so-called Ordinary Differential Equation (ODE) method introduced by Ljung [90]. The main idea is to show that in the long run, the noise is eliminated. Asymptotically, the behavior of the algorithm is the same as that of the “mean” ODE:

$$\dot{z} = h(z).$$

However, the algorithm can show bad behavior as soon as the function h grows rapidly. More precisely, the almost surely convergence result (see for instance Duflo [42]) relies on the assumption that h has a sub-linear growth reducing drastically the scope of application of these algorithms. To fix this, Kushner and Sanvicente [81] introduced projection techniques to prevent the algorithm from explosion. This consists in projecting the sequence Z_n into a compact set K that contains the zeros of h each time it goes outside of K . An excellent survey on projection techniques and their links with ODE can be found in Kushner and Yin [83].

Once the almost surely convergence is established, we are interested in the rate of convergence of the sequence $(Z_n - *)$ in order to establish Central Limit Theorems (CLT). It has been showed under different assumptions that the sequence $(Z_n - *)/\sqrt{\gamma_n}$ converges in distribution to a normal distribution (see Kushner and Yin [83], Ljung [90] or Ruppert [117]). In fact, with the choice of $\gamma_n = c/n$ with c sufficiently large, a convergence rate of $\sqrt{c/n}$ can be obtained. But, this has a price: choosing c too large may lead to slower convergence as the effects of large noises early in the procedure might be hard to overcome later. To avoid this, one can think of moving average techniques as they allow to have smoother behavior and also to “forget” the first terms of the sequence if the average window is chosen appropriately. Averaging algorithms were introduced in Polyak and Juditsky [108] and Ruppert [117] and then widely investigated by many researchers. Kushner and Yang [82] and Kushner and Yin [83] studied averaged algorithms in combination with projection and proved a CLT for them.

In the context of risk measures, the stochastic algorithms framework can be very useful. Bardou et al. [6] use SA to estimate both VaR and CVaR. Their idea to compute both VaR and CVaR comes from the fact that they appear as the solutions and the value of the same optimization problem as pointed out in Rockafellar and Uryasev [113]. Moreover, both the objective function of the optimization problem and its gradient read as an expectation. This led them to define

consistent and asymptotically normal estimators for both quantities by using stochastic algorithms. Since VaR and CVaR are closely related to the simulation of rare events, they propose a parallel recursive variance reduction method based on importance sampling. In Dunkel and Weber [43], stochastic approximation algorithms are presented for the estimation of SR in the univariate case. The authors show that the estimators are consistent and approximately normal, and they provide formulas for their rate convergence and their asymptotic variance. The performance of the proposed stochastic algorithms was investigated for various loss functions and P&L distributions.

1.2.4 Main contributions

Being motivated by the usefulness of stochastic algorithms, the results in Chapter 3 show how stochastic algorithms can be used to approximate multivariate shortfall risk measures introduced recently in Armenti et al. [4]. We first introduce consistent estimators and we then give central limit theorems for the study of their rate of convergence. This allow us to have confidence intervals, which was not the case with the two-steps algorithms proposed in Armenti et al. [4].

More specifically, let l_S be a multivariate loss function and a vector of losses $X \in M^\theta$. Under the assumptions of Theorem 1.2.3, there exists a unique risk allocation m^* such that $\lambda^* = (m^*, \lambda^*)$ is the unique root of the function $h(\lambda) = E[H(X, \lambda)]$, where,

$$H(X, \lambda) = \frac{\lambda \partial_m l_S(X - m) - 1}{l(X - m)_S}, \quad X \in M^\theta. \quad (1.28)$$

As pointed out before, it is not possible to use the classical Robbins-Monro algorithm, as it requires the sub-linearity of the function h . Instead, we use the projected Robbins-Monro algorithm:

$$Z_{n+1} = \Pi_K[Z_n + \gamma_n H(X_{n+1}, Z_n)], \quad Z_0 = z_0 \in K. \quad (1.29)$$

We assume that K is a hyperrectangle such that λ^* is in the interior of K and that $\gamma_n = \frac{c}{n^\nu}$ where c is a positive constant and $\nu \in (\frac{1}{2}, 1]$. To prove the convergence of Z_n toward λ^* , we use the ODE approach. This approach consists in showing first that λ^* is an asymptotically stable equilibrium point for the associated ODE with algorithm (3.5). This is done in Proposition 3.3.3 in Chapter 3. Next, we introduce the following quantities:

$$\begin{aligned} \cdot \quad \sigma^2(\lambda) &= E[\|H(X, \lambda) - h(\lambda)\|^2]; \\ \cdot \quad m^{2+p}(\lambda) &= E[\|H(X, \lambda) - h(\lambda)\|^{2+p}]; \\ \cdot \quad \Sigma(\lambda) &= E[(H(X, \lambda) - h(\lambda))(H(X, \lambda) - h(\lambda))^T], \end{aligned}$$

and the following assumptions:

- (ffi_{a.s.})
- i. h is continuous on K ;
 - ii. $\sup_{\theta \in K} \sigma^2(\theta) < \infty$.

In the next theorem, we state the consistency result of the estimator in (3.5).

Theorem 1.2.4. Assume that the sequence (Z_n) is defined by algorithm (3.5) and that assumptions (ffi_{a.s.}) holds. Then, $Z_n \rightarrow \theta^*$ P -a.s. as $n \rightarrow \infty$.

Under the following stronger conditions (ffi_{a.n.}), Theorem 3.3.5 in Chapter 3 shows that the sequence (Z_n) is asymptotically normal.

- (ffi_{a.n.})
- i. $m \mapsto E[O_{\gamma}(X - m)]$ is continuously differentiable. Let $A := Dh(\theta^*)$;
 - ii. $(Y_n \mathbf{1}_{|Z_n - \theta^*| \leq \rho})$ is uniformly integrable for small $\rho > 0$;
 - iii. For some $\rho > 0$ and $\rho > 0$, $\sup_{|\theta - \theta^*| \leq \rho} m^{2+\rho}(\theta) < \infty$;
 - iv. $\Sigma(\cdot)$ is continuous at θ^* . Let $\Sigma^* := \Sigma(\theta^*)$.

Theorem 1.2.5. Assume that $\gamma \in (\frac{1}{2}, 1)$ and that assumptions (ffi_{a.s.}) and (ffi_{a.n.}) hold. Then,

$$\sqrt{\frac{c}{n}}(Z_n - \theta^*) \rightarrow N\left(0, c^2 \int_0^{\infty} e^{cAt} \Sigma^* e^{cA^T t} dt\right).$$

If furthermore, $cA + \frac{1}{2}I$ is a Hurwitz matrix and $cI - P$ is positive definite with P solution to the Lyapunov's equation: $A^T P + PA = -I$, then,

$$\sqrt{\frac{c}{n}}(Z_n - \theta^*) \rightarrow N\left(0, c^2 \int_0^{\infty} e^{(cA + \frac{1}{2}I)t} \Sigma^* e^{(cA^T + \frac{1}{2}I)t} dt\right).$$

From the above theorem, we can see the key role played by the constant c . In order to have the best rate of convergence, c must be sufficiently large but as said before, this can lead to slower convergence (note that c appears also in the asymptotic covariance matrix). In order to circumvent this problem, we are led to modify our algorithm by using averaging procedures. The following theorem describes the Polyak-Ruppert (PR) algorithm that we have used to approximate optimal allocations.

Theorem 1.2.6. Assume $\gamma \in (\frac{1}{2}, 1)$ and that assumptions (ffi_{a.s.}) and (ffi_{a.n.}) hold. For any arbitrary $t > 0$, let

$$\bar{Z}_n = \frac{1}{t} \sum_{i=n}^{n+t-1} Z_i \tag{1.30}$$

where any upper summation index $u \in \mathbb{R}^+$ is interpreted as its integer part. If Σ^* is positive definite, then we have the following CLT:

$$\frac{1}{\sqrt{n}} (\bar{Z}_n - Z^*) \rightarrow N(0, V + o(\frac{1}{n})), \quad (1.31)$$

where $V = A^{-1}\Sigma^*(A^{-1})^\top$.

The previous CLT asserts that under some conditions, the PR estimator is asymptotically normal. The asymptotic variance V depends on the quantities Σ^* and A . In practice, these quantities are unknown and need also to be approximated in order to derive confidence intervals for our estimators. In Proposition 3.3.9 in Chapter 3, we provide consistent estimators of these two quantities. The proof relies mainly on the Martingale Convergence Theorem.

Proposition 1.2.7. Assume (ffi_{a.s.}) and (ffi_{a.n.}) hold.

If $\mathbb{E}[|H(X, Z)|^4]$ is locally bounded around Z^* , then,

$$\frac{1}{n} \sum_{k=1}^n H(X_k, Z_{k-1})^\top H(X_k, Z_{k-1}) \rightarrow \Sigma^* \text{ a.s.} \quad (1.32)$$

Let A_n^ϵ the matrix whose elements $A_n^\epsilon(i, j)$ for $i, j \in \{1, \dots, d+1\}$ are defined as follows:

$$A_n^\epsilon(i, j) := \frac{1}{\epsilon n} \sum_{k=1}^n H_i(X_k, Z_{k-1} + \epsilon e_j) - H_i(X_k, Z_{k-1}),$$

then,

$$\lim_{\epsilon \rightarrow 0} \lim_{n \rightarrow \infty} A_n^\epsilon = A \text{ a.s.} \quad (1.33)$$

By using the estimators for the matrices Σ^* and A above, we are able to derive confidence intervals in one run.

In Chapter 3, we test the RM and PR algorithms on different loss functions and different loss distributions. We started by a constructive and simple example in the case of the exponential loss function in (1.24) and a normal loss distribution. In the case $d = 2$, we can solve explicitly, in the same way as in Armenti et al. [4], the first order conditions and obtain closed formulas for optimal allocations:

$$m_i^* = \begin{cases} \frac{\beta \sigma_i^2}{2}, & \text{if } \alpha = 0, \\ \frac{\beta \sigma_i^2}{2} + \frac{1}{\beta} \text{SRC}(\rho, \sigma_1, \sigma_2, \alpha, \beta), & \text{if } \alpha > 0, \end{cases}$$

where σ_1, σ_2 are the standard deviation of X_1 and X_2 and ρ their correlation. The *Systemic Risk Contribution* (SRC) is given by the following formula:

$$SRC(\rho, \sigma_1, \sigma_2, \alpha, \theta) = \ln \frac{\alpha e^{\rho \theta^2 \sigma_1 \sigma_2}}{\frac{1}{1 + \alpha(\alpha + 2)} e^{\frac{\rho \theta^2 \sigma_1^2 \sigma_2^2}{\sigma_1^2 + \sigma_2^2}}}$$

This shows that the optimal allocations are the sum of two components: an individual contribution σ_j^2 and a systemic contribution SRC. To push the analysis further, we consider a second loss function that is given by:

$$I(x_1, \dots, x_d) = \sum_{i=1}^d x_i + \frac{1}{d} \sum_{i=1}^d x_i^2 + \alpha \sum_{i < j} x_i x_j,$$

and because we want to model the losses of insurance contracts, we consider a multivariate Compound Poisson distribution for the loss vector distribution with $d = 10$. We implement the estimators and provide confidence intervals using Python. Detailed results can be found in section 3.4.2 of Chapter 3).

1.3 Multivariate Optimized Certainty Equivalent Risk Measures

In this section, we present a new class of multivariate risk measures inspired from Optimized Certainty Equivalent (OCE).

1.3.1 Univariate Optimized Certainty Equivalent Risk Measures

In their classical work, Von Neumann and Morgenstern [131] have developed a set of axioms concerning preferences over random payoffs. Under these axioms, a random variable X is preferred against a random variable Y if and only if there exists a utility function u , unique up to a monotone increasing affine transformation, such that $E[u(X)] \geq E[u(Y)]$. Moreover, the *Certainty Equivalent* (CE) of a random variable X is defined by:

$$C_u(X) := u^{-1}(E[u(X)]). \tag{1.34}$$

It is the sure amount for which the decision maker is indifferent to the random payoff X , that is, $u(C_u(X)) = E[u(X)]$. When u is strictly increasing, the preference order in terms of expected utility is equivalent to,

$$X \geq Y \text{ if and only if } C_u(X) \geq C_u(Y).$$

Bühlmann [23] introduced another certainty equivalent called *u-mean*. It is defined for any random variable X by $M_u(X)$ satisfying

$$E[u(X - M_u(X))] = 0. \quad (1.35)$$

The last equation is also known as the *principle of zero utility*.

For arbitrary utility functions, the certainty equivalent does not have the additivity property, except for the exponential case. This has led the authors in Ben-Tal and Teboulle [11] to suggest a new certainty equivalent that possesses many of the properties that the classical certainty equivalent in (1.34) possesses only for the exponential utilities. They considered a decision maker expecting to receive a payoff X in the future and has the possibility to consume part of it at present. Let u denotes his utility. If he chooses to consume η amount of cash, the resulting “present value” of X is then $\eta + E[u(X - \eta)]$. Here the utility function can be interpreted as a discounting function. Hence, the sure present value of X denoted $S_u(X)$, which is called the *Optimized Certainty Equivalent* (OCE) in Ben-Tal and Teboulle [11], is the result of an optimal allocation of X between present and future consumption, i.e.,

$$S_u(X) := \sup_{\eta \in \mathbb{R}} \{\eta + E[u(X - \eta)]\}. \quad (1.36)$$

There exists a strong relationship between certainty equivalents and risk measures. In fact, any certainty equivalent inducing a preference order on random variables generates a corresponding risk measure ρ , and vice versa:

$$\rho(X) = -CE(X),$$

where $CE(X)$ is any certainty equivalent in (1.34) or (1.35). One important example is given by the relationship between shortfall risk measures and the *u-mean* certainty equivalents. Indeed, if the utility function is strictly increasing, it is easy to see that,

$$\rho(X) := \inf\{\eta, E[u(X - \eta)] \geq 0\} = -M_u(X).$$

Inspired by this, Ben-Tal and Teboulle [10] re-examined the OCE, extended its main properties and put it in perspective of the recent theory of risk measure. They showed that,

$$\rho_u(X) = -S_u(X), \quad (1.37)$$

is a convex risk measure under some reasonable conditions on the utility function u . More precisely, let \mathcal{U} denote the set of utility functions $u : \mathbb{R} \rightarrow [-\infty, \infty)$ that are proper, closed, concave and nondecreasing functions. Furthermore, assume that $u(0) = 0$ and $1 \in \partial u(0)$.

Then, it is easy to show that ρ_u is a convex risk measure.

The OCE provides also a simple way to generate classical risk measures via particular choices of a utility function. We give here some interesting choices of utility functions that allow to retrieve some of the classical risk measures.

Example 1.3.1. (Exponential Utility Function)

Let $u(x) = 1 - e^{-x}$. By solving (1.36), we get,

$$S_u(X) = -\log E[e^{-X}].$$

By taking the negative of S_u , we obtain the classical entropy risk measure.

Example 1.3.2. (Quadratic Utility)

Let

$$u(t) = \begin{cases} x - \frac{1}{2}x^2 & \text{if } x < 1, \\ \frac{1}{2} & \text{otherwise.} \end{cases}$$

Let X be a bounded random variable. A direct computation from (1.36) yields,

$$S_u(X) = \mu - \frac{1}{2}\sigma^2,$$

where $\mu := E[X]$ and σ^2 is the variance. The corresponding risk measure $\rho(X) = -S_u(X)$ is known as the mean-variance risk measure.

In the next example, we show that the risk measures in Pflug and Ruszczynski [107] and the popular risk measure CVaR is just a special case of the OCE via piecewise linear utility functions.

Example 1.3.3. (Piecewise Linear Utility)

Let

$$u(t) = \begin{cases} \gamma_2 x & \text{if } x \leq 0, \\ \gamma_1 x & \text{if } x > 0, \end{cases}$$

where $0 \leq \gamma_1 < 1 < \gamma_2$. By rewriting the utility function u as:

$$u(x) = \gamma_1 x^+ - \gamma_2 x^-,$$

we therefore obtain the following convex risk measure:

$$\rho(X) = \inf_{\eta \in \mathbb{R}} \{ \eta - \gamma_2 E[(-\eta - X)_+] - \gamma_1 E[(\eta + X)_+] \}.$$

With the special case of $\gamma_1 = 0$ and $\alpha := 1/\gamma_2$, we retrieve the case of CVaR_α (see Rockafellar

and Uryasev [113]), i.e.,

$$CVaR = \inf_{\eta \in \mathbb{R}} \left\{ \eta + \frac{1}{\alpha} E[(\eta + X)^-] \right\}$$

More recently, Drapeau et al. [39] studied OCE risk measures from the perspective of loss functions and under Orlicz spaces (instead of L^∞). In their work, a function $I: \mathbb{R} \rightarrow \mathbb{R}$ is called a loss function if

1. I is increasing and convex;
2. $I(0) = 0$ and $I(x) \geq x$;
3. $I(x) > x$ for all x such that $|x|$ is large enough.

In the following, we give a summary of the properties of OCE risk measures in the univariate case. First, let (Ω, \mathcal{F}, P) be probability space and I^* denote the convex conjugate of I , i.e., $I^*(y) = \sup_{x \in \mathbb{R}} \{xy - I(x)\}$, a financial position X is assumed to belong to the Orlicz heart,

$$X_I := \{X \in L^0, E[I(c|X|)] < \infty \text{ for all } c > 0\},$$

and the corresponding Orlicz space is given by

$$X_I^* := \{Y \in L^0, E[I^*(c|Y|)] < \infty \text{ for some } c > 0\}.$$

For $X \in X_I$, the OCE risk measure is then obtained from (1.36) by taking $u(x) = -I(-x)$:

$$\rho(X) = \inf_{\eta \in \mathbb{R}} \{ \eta + E[I(-X - \eta)] \}. \quad (1.38)$$

They showed that ρ is a well defined risk measure taking values in \mathbb{R} and derived a dual representation. We summarize this in the following proposition:

Proposition 1.3.1. [Proposition 1.3 in Drapeau et al. [39]] *Let I be a loss function. Then the Optimized Certainty Equivalent on X_I is a lower semicontinuous cash additive risk measure taking values in \mathbb{R} .*

In addition, for any $X \in X_I$, there exists an optimal allocation $\eta^ = \eta^*(X) \in \mathbb{R}$ such that,*

$$\rho(X) = \eta^* + E[I(-X - \eta^*)],$$

and this optimal allocation η^ satisfies*

$$E[I_-(-X - \eta^*)] \leq 1 \leq E[I_+(-X - \eta^*)]. \quad (1.39)$$

Finally, the OCE has the representation,

$$\rho(X) = \max_{Q \in M_{1, P^*}} E_Q[-X] - E_P \int^* \frac{dQ}{dP}, \quad X \in X_I,$$

where M_{1, P^*} is the set of probabilities measures which are absolutely continuous with respect to P and whose densities are in X_I^* .

1.3.2 Main contributions

From a practical point of view, the approach of Armenti et al. [4] on systemic risk measures based on multivariate loss function constitutes one of the major works that can find applications in the real world. Nevertheless, the choice of the multivariate loss function and its interpretation remains one of main challenges in practice. Univariate OCE risk measures does not have this flaw. In fact, even though it is also based on loss functions, with the OCE framework, one can retrieve the classical risk measures such as CVaR, mean-variance, entropic risk measure, etc. This allows the practitioner to have a financial interpretation when using OCE risk measures. Moreover, as stated in Theorem 1.2.3, the existence of optimal allocations in Armenti et al. [4] requires the permutation invariance of the multivariate loss function. We shall see that, with our new class of multivariate risk measures, this is no longer needed allowing us to deal with heterogeneous risk factors.

Motivated by this, the objective of Chapter 4 is twofold: First, we extend the OCE risk measure to a multivariate setting where the system is allowed to be heterogeneous. Second, we use stochastic algorithms to approximate the multivariate OCE risk measures. To the best of our knowledge, there are only few works on the numerical approximation of univariate OCE risk measures (Drapeau et al. [39]). Since the optimal allocations are characterized as a zero of a function that can be expressed as an expectation, Drapeau et al. [39] proposed to use a deterministic root finding algorithm combined with Fourier method for the computation of the expectations. As mentioned in the previous section, this type of methods does not provide any control over the error of the estimations (no confidence intervals) and remains sensitive to the initial point of the deterministic root finding algorithm used. This is why we propose to use stochastic algorithms to circumvent these type of problems. We start by extending the definition of an OCE loss function in the multivariate case which is slightly different from the one introduced in Armenti et al. [4] and the multivariate OCE risk measure.

Definition 1.3.2. A function $l_C : \mathbb{R}^d \rightarrow (-\infty, \infty]$ is called a multivariate OCE loss function, if it satisfies the following properties:

1. l_C is nondecreasing, that is if $x \leq y$ componentwise, then $l_C(x) \leq l_C(y)$;
2. l_C is lower-semicontinuous and convex;

3. $l_C(0) = 0$ and $l_C(x) > \sum_{i=1}^d x_i, \forall x \neq 0$.

Note that in our definition of the loss function l_C , we do not require permutation invariant property as it was the case in the previous section with the loss function l_S . Next, we give examples of loss functions l_C that are used in the Section 4.3 in Chapter 4:

$$l_C(x) = \sum_{i=1}^d \frac{e^{\lambda_i x_i} - 1}{\lambda_i} + \alpha e^{-\sum_{i=1}^d \lambda_i x_i}, \lambda_i > 0, \alpha \geq 0, \quad (1.40)$$

$$l_C(x) = \sum_{i=1}^d \frac{([1 + x_i]^{\vartheta_i}) - 1}{\vartheta_i} + \alpha \sum_{i < j} \frac{([1 + x_i]^{\vartheta_i}) ([1 + x_j]^{\vartheta_j})}{\vartheta_i \vartheta_j}, \vartheta_i > 1, \alpha \geq 0. \quad (1.41)$$

For integrability reasons, we will work in the multivariate Orlicz heart M^ϑ that was introduced previously in (1.2.2). On this space, we define the Luxembourg norm as:

$$\|X\|_\vartheta := \inf_{\lambda > 0} \lambda^{-1} E[\vartheta(\lambda X)] \leq 1.$$

Under the Luxembourg norm, M^ϑ is a Banach lattice and its dual with respect to this norm is given by the Orlicz space L^{ϑ^*} :

$$L^{\vartheta^*} := \{X \in L^0, E[\vartheta^*(\lambda X)] < \infty, \text{ for some } \lambda > 0\}.$$

We also introduce the set of d -dimensional measure densities in L^{ϑ^*} , that is:

$$Q^{\vartheta^*} := \frac{dQ}{dP} := (Z_1, \dots, Z_d), Z \in L^{\vartheta^*}, Z_k \geq 0 \text{ and } E[Z_k] = 1.$$

Note that for $Q \ll P$ a vector of probabilities such that $\frac{dQ}{dP} \in Q^{\vartheta^*}$ and a random vector $X \in M^\vartheta$, $\frac{dQ}{dP} \cdot X \in L^1$ thanks to Fenchel inequality and for the sake of simplicity, we will write $E_Q[X] := E[\frac{dQ}{dP} \cdot X]$.

Definition 1.3.3. Assume l_C is a multivariate OCE loss function. The multivariate OCE is defined for every $X \in M^\vartheta$ as:

$$R(X) = \inf_{w \in \mathbb{R}^d} \sum_{i=1}^d w_i + E[l_C(-X - w)]. \quad (1.42)$$

In the next theorem, we show that the multivariate OCE is a convex risk measure as defined in Föllmer and Schied [51] and gives its dual representation. The proof uses results of risk measures on Orlicz spaces in Cheridito and Li [29] and convex analysis in Rockafellar [114]. The univariate version of this theorem can be found in Drapeau et al. [39].

Theorem 1.3.4. The function R in (1.42) is real valued, convex risk, monotone and cash

invariant¹, and in particular, is continuous and subdifferentiable. If l_c is positive homogeneous, then R is too. Furthermore, it admits the following representation:

$$R(X) = \max_{Q \in D^{c*}} \{E_Q[-X] - \alpha(Q)\}, \tag{1.43}$$

where the penalty function α is defined for $Q = (Q_1, \dots, Q_d) \ll P$ by: $\alpha(Q) = E \left[l_{OCE}^* \left(\frac{dQ}{dP} \right) \right]$ and $D^{c*} = \{Q \ll P, \alpha(Q) < \infty\} := \text{dom}(\alpha)$.

Definition 1.3.5. A risk allocation is any minimizer of (1.42). When it is uniquely determined, we denote it $RA(X)$.

The next theorem states that the set of risk allocations is non empty and gives its characterization.

Theorem 1.3.6. Let l_c be a multivariate OCE loss function. Then, for every $X \in M^\theta$, the set of risk allocations is non empty and bounded. Furthermore, risk allocations are characterized by the following first order condition:

$$1 \in E[\partial l_c(-X - m^*)]. \tag{1.44}$$

Moreover, the supremum in (1.43) is attained for D^* such that $D^* \in \partial l_c(-X - m^*)$ a.s. and $E[D^*] = 1$.

In Chapter 4, sensitivity results are also obtained for our multivariate risk measure. We first give the definition of the marginal risk contribution of a random vector $Y \in M^\theta$ to $X \in M^\theta$.

Definition 1.3.7. For $X, Y \in M^\theta$, we define the marginal risk contribution of Y to X as the sensitivity of the risk associated to X when an impact Y is applied as

$$R(X, Y) := \limsup_{\epsilon \searrow 0} \frac{R(X + \epsilon Y) - R(X)}{\epsilon}. \tag{1.45}$$

If $R(X + \epsilon Y)$ admits a unique risk allocation $RA(X + \epsilon Y)$ for small enough $\epsilon \geq 0$, then we define the risk allocation martingales of X with respect to the impact of Y as:

$$RA_i(X; Y) := \limsup_{\epsilon \searrow 0} \frac{RA_i(X + \epsilon Y) - RA_i(X)}{\epsilon}, \quad i = 1, \dots, d. \tag{1.46}$$

Theorem 1.3.8. Let $X, Y \in M^\theta$ and assume that l is differentiable. Then,

$$R(X, Y) = -E \left[\sum_{n=1}^d \frac{\partial l}{\partial m_n} (X - m^*) Y^n \right] = \sum_{n=1}^d \frac{\partial R}{\partial m_n} [Y^n], \tag{1.47}$$

1. In the following sens: $R(X + m) = R(X) - \sum_{i=1}^d m_i$

where m^* is such that, $E[Ol_C(-X - m^*)] = 1$, i.e. an infimum for (4.1) and $d_{QP}^{Q^*} := Ol_C(-X - m^*)$.

If furthermore, l_C is twice differentiable and such that we can interchange the differentiation and expectation of $m \mapsto E[Ol_C(-X - m)]$ and $M := E[O^2l_C(-X - m^*)]$ is invertible, then,

- There exists a unique m_ϵ optimum of $R(X + \epsilon Y)$ for small enough $\epsilon \geq 0$.
- As a function of ϵ , m_ϵ is differentiable and we have

$$RA(X, Y) = M^{-1}V, \quad V := -E[O^2l_C(-X - m^*)Y]. \quad (1.48)$$

Finally, we obtain in Section 4.3 estimators for the optimal allocations as well as the risk measure $R(X)$. Because we do not assume growth conditions on the multivariate OCE loss function l_C , we use projection techniques as in the previous section of this introduction. More precisely, let l_C be a multivariate OCE loss function and let K be a hyperrectangle such that $m^* \in K$. We define for $X \in M^\vartheta$ and $m \in \mathbb{R}^d$:

$$\begin{aligned} & \cdot H_1(\mathbf{X}, m) := Ol_C(-\mathbf{X} - m) - 1, \\ & \cdot h_1(m) := E[H_1(\mathbf{X}, m)], \\ & \cdot \sigma^2(m) := E[\|H_1(\mathbf{X}, m) - h_1(m)\|^2], \\ & \cdot m^{2+p}(m) := E[\|H_1(\mathbf{X}, m) - h_1(m)\|^{2+p}], \quad p > 0, \\ & \cdot \Sigma(m) := E[(H_1(\mathbf{X}, m) - h_1(m))(H_1(\mathbf{X}, m) - h_1(m))^T]. \end{aligned}$$

We introduce the following set of assumptions:

- (ffi_{a.s.})
- $\sum_{n \geq 0} \gamma_n = +\infty$ and $\sum_{n \geq 0} \gamma_n^2 < \infty$. Namely, $\gamma_n = c/n^\gamma$, where $c > 0$ and $\gamma \in (1/2, 1]$.
 - h_1 is continuous on K ;
 - $\sup_{m \in K} \sigma^2(m) < \infty$.

Theorem 1.3.9. Let (X_n) a sequence of random variables having the same law as $X \in M^\vartheta$ and define the sequence (m_n) as follows:

$$m_{n+1} = \Pi_K [m_n + \gamma_n H_1(X_{n+1}, m_n)], \quad m_0 \in L^0, \quad (1.49)$$

where Π_K is the projection into K . Under (ffi_{a.s.}) we have, $m_n \rightarrow m^*$ a.s. as $n \rightarrow \infty$.

Once an estimator of m^* is obtained, it comes the question of estimating the multivariate OCE $R(\mathbf{X}) = \sum_{i=1}^d m_i^* + E[l_C(-\mathbf{X} - m^*)]$. A naive way consists in estimating $R(\mathbf{X})$ in a two steps procedure:

- Step 1: Use the estimate m_n from (1.49) to have a good approximation of m^* .

- Step 2: Use another sample of X to approximate $R(X)$ using Monte Carlo:

$$R(X) \approx \sum_{i=1}^n m^* + \frac{1}{n} \sum_{k=1}^n I_C(-X_k - m^*). \quad (1.50)$$

An alternative way to avoid this two steps procedure is to use a companion procedure (CP) of the algorithm (1.49) and to replace the quantity m^* in (1.50) by its estimate at step $k - 1$, that is,

$$R_n = \frac{1}{n} \sum_{k=1}^n \left(\sum_{i=1}^k m_{k-1}^i \right) + I_C(-X_k - m_{k-1}).$$

Note that R_n is a sequence of empirical means of non i.i.d. random variables that can be written also as:

$$R_{n+1} = R_n - \frac{1}{n+1} H_2(X_{n+1}, R_n, m_n), \quad n \geq 0, \quad R_0 \in L^0, \quad (1.51)$$

where

$$H_2(x, R, m) := R - F(x, m) := R - \sum_{i=1}^n m^i + I_C(-x - m).$$

In Theorem 4.3.2, we prove the consistency of R_n using the step sequence (γ_n) instead of $(1/(n + 1))_n$:

$$R_{n+1} = R_n - \gamma_n H_2(X_{n+1}, R_n, m_n), \quad n \geq 0, \quad R_0 \in L^0. \quad (1.52)$$

We need the following assumption:

- (ffi_{CP}) i. $m \rightarrow E[I_C(-X - m)]$ is continuous on K ;
- ii. $\forall m \in K \quad I_C(-X - m) \in L^2$ and $m \rightarrow E[|I_C(-X - m)|^2]$ is bounded around m^* .

Theorem 1.3.10. Assume that both assumptions (ffi_{a.s.}) and (ffi_{CP}) hold and let (m_n) be given by (1.49) and (R_n) by (1.52). Then $R_n \rightarrow R(X)$ a.s.

Finally, we give the CLT theorem using averaging techniques to avoid problems related to the specifications of the constant c . We introduce the following assumptions:

- (ffi_{a.n.}) i. h_1 is continuously differentiable and let $A := Dh_1(m^*)$;
- ii. $(H_1(X_{n+1}, m_n) \mathbf{1}_{|m_n - m^*| \leq \rho})$ is uniformly integrable for small $\rho > 0$;
- iii. For some $\rho > 0$ and $\rho > 0$, $\sup_{|m - m^*| \leq \rho} m^{2+\rho}(m) < \infty$;
- iv. $\Sigma(\cdot)$ is continuous at m^* and $\Sigma^* := \Sigma(m^*)$ is positive definite.

Theorem 1.3.11. Assume $\gamma \in (\frac{1}{2}, 1)$ and that assumptions (ffi_{a.s.}) and (ffi_{a.n.}) hold. For any arbitrary $t > 0$, we define the (PR) sequence (m_n) as:

$$m_n := \frac{\gamma_n}{t} \sum_{i=n}^{n+t\gamma_n-1} m_i, \quad (1.53)$$

where any upper summation index is interpreted as its integer part. Then, we have

$$\frac{\mathbf{s}}{y_n} (\bar{m}_n - m^*) \rightarrow \mathcal{N} \left(0, A^{-1} \Sigma^* (A^{-1})^\top + \mathcal{O} \left(\frac{1}{t} \right) \right). \quad (1.54)$$

Note that in order to be able to derive confidence intervals for the estimators, we need to estimate the matrices A and Σ^* . This is done using the same method used in the previous section of the introduction. In section 4.3, we analyze and test the numerical methods above for the estimation of optimal allocations and the risk measure. We start by a bidimensional case with exponential loss function and a Gaussian distribution for which closed solutions are available. We then test our algorithms with a polynomial loss function and Multivariate Normal Inverse Gaussian distribution in a trivariate case. Finally, we compare our approach with the Monte Carlo method in terms of precision and computational time (see section 4.4 for more details).

1.4 Future research perspectives

We present here some future research perspectives regarding the three works in this thesis.

- In the first part of this thesis (Chapter 2), we derive the BSDE characterizing the value process associated to our optimization problem under the assumption that the optimal probability Q^* is equivalent to the reference probability P . Therefore, one possible direction of future research is to see if this assumption always holds. This was proven in Bordigoni et al. [18] for the special case of an entropic penalty. Moreover, we are also exploring the numerical approximation of the solution of the QBS-DEJ in (1.18) as an extension of the work of Chassagneux and Richou [26] in the continuous case.
- In the second work, we proposed stochastic algorithms to approximate multivariate Short-fall risk measures. This can find applications in the insurance industry: Current work is in progress to apply this framework to the construction insurance within the “Chair Risques Émergents en Assurance” under the aegis of Fondation du Risque, a joint initiative by Le Mans Université and Covéa. Another potential application of multivariate risk measures is in cyber insurance by using compound multivariate Hawkes processes for the vector of losses (see Bessy-Roland et al. [14]). Finally, a future direction of research concerns including variance reduction techniques in parallel with our stochastic algorithms.
- In the last work, we introduced the multivariate extension of the OCE risk measures and proposed stochastic algorithms to approximate them. In Buehler et al. [22], univariate

OCE risk measures are used to solve a hedging problem by using deep learning (deep hedging). Now, suppose that a vector of future payoffs need to be hedged. Instead of hedging each single portfolio, an alternative would be to use using multivariate OCE risk measures to take into account the dependence structure between the payoffs. This might give better hedging errors for the vector of payoffs.

ROBUST UTILITY MAXIMIZATION PROBLEM IN A JUMP SETTING

Abstract

We study a robust utility maximization problem in the unbounded case with a general penalty term and information including jumps. We focus on time consistent penalties and we prove that there exists an optimal probability measure solution of the robust problem. Then, we characterize the dynamic value process of our stochastic control problem as the unique solution of a Quadratic-Exponential BSDE.

2.1 Introduction

One of the major problems in asset pricing is the valuation in incomplete markets. In such markets, the decision maker/agent could use the well known utility maximization approach and the literature is particularly rich on the subject (see for example Rouge and El Karoui [116], Hu et al. [68] Morlais [99] and Carmona [24] among many others). However, in many cases, the decision maker does not know the probability distribution (also called prior or model) governing the stochastic nature of the problem she/he is facing. Thus, before solving the utility maximization problem, the decision maker is faced with an intermediate problem of choosing an “optimal” probability. This type of problems are called robust utility maximization problems or utility maximization problems under model uncertainty. In the mathematical finance literature, we can find two types of approaches to solve robust utility maximization problems. The first one relies on convex duality methods which are presented in Quenez [109], Gundel [61], Schied [118] and Schied and Wu [120]. The second one, which we will follow in this article, is based on a stochastic control approach and the powerful tool of BSDEs.

In this article, this uncertainty is captured by considering a set of plausible probability measures that will be penalized through a penalty functional. This penalty functional will measure the distance between any plausible probability Q and the reference/historical one denoted P . In Anderson et al. [3] and Hansen et al. [62] for example, a hedging problem was addressed by using the classical entropic penalty under a Markovian setting and hence Hamilton-Jacobi-Bellman

(HJB) equations were derived in order to characterize the optimal strategies. The authors in Skiadas [126] have followed the same point of view and obtained a BSDE that coincides with the one describing a stochastic differential utility (see also Duffie and Epstein [40] and Duffie and Skiadas [41] for more about stochastic differential utilities).

More recently, Bordigoni et al. [18] addressed a robust problem in a more general setting which is non Markovian by using stochastic control techniques. More precisely, they studied the following robust maximization problem:

$$\sup_{\pi} \inf_{Q \in \mathcal{Q}} U(\pi, Q) \tag{2.1}$$

where π runs through a set of strategies and $Q \in \mathcal{Q}$ through a set of models. The simplest case corresponds to the case where the set \mathcal{Q} is the singleton $\{P\}$ and $U(\pi, P)$ is simply the P -expected utility from a non bounded terminal wealth and consumption/investment portfolio. The term $U(\pi, Q)$ is the sum of Q -expected utility and an entropic penalization term. The set \mathcal{Q} is assumed to have certain properties and usually does not need to be specified in any detail. Their work is cast in the case of a continuous filtration and the first minimization problem was solved by proving the existence of a unique optimal probability Q^* . They also characterized the value process of the stochastic control problem as the unique solution of a Quadratic BSDE (QBSDE). In the same spirit, Faidi et al. [47] studied the same problem using two type of penalties: the first one is the f -divergence penalties in the general framework of a continuous filtration and the second one is the time-consistent penalties studied in the context of a Brownian filtration. For the latter, they also characterized the value process as the unique solution of a QBSDE.

In this paper, we study the first minimization problem in (2.1) in the case of discontinuous filtration (where the information includes jumps) using time consistent penalties. Note that in our framework, the relative entropic penalty, as we will see further, is a special case of the class of time-consistent penalties. We first start by showing that the minimization problem in (2.1) is well posed and we prove the existence of an optimal probability Q^* using a Komlós-type argument. Second, we prove that the value process of the minimization problem is described by a class of Quadratic-Exponential BSDE with jumps (QEBSDEJ) with unbounded terminal condition. We stress that for a given unbounded terminal condition, the study of Quadratic BSDEs is a difficult problem, see for instance Briand and Hu [19], Briand and Hu [20] and Barrieu and El Karoui [7] in the continuous framework and we emphasize that adding jumps to our optimization problem involves significant difficulties in solving the related BSDEs. Karoui et al. [79] have obtained existence result for this new class of BSDEs with jumps with unbounded terminal condition. However, they have showed uniqueness only in the bounded case. In this paper, we use the convexity of the generator of our BSDE to show the uniqueness of solution of the BSDE by

extending the work of Briand and Hu [20] in the Brownian setting.

The paper is structured as follows. Section 2.2 establishes the general framework, in which we assume the existence of a stochastic basis carrying a Brownian motion and a compensated integer-valued random measure that possesses a weak predictable representation property. In section 2.3, we give a number of estimates for subsequent use. We then prove with the help of Komlòs theorem that there exists an optimal probability Q^* . Finally, in section 2.4, we treat our optimization problem from a stochastic control point of view, and show, thanks to Bellman Optimality Principle, that the corresponding value process is the unique solution of a QEBSDEJ.

2.2 Framework of the optimization problem

2.2.1 Setting and notations

This section sets out the notation and the assumptions that will be assumed to hold in the sequel. Let $(\Omega, \mathcal{F}, \mathbb{F}, P)$ be a filtered probability space with a finite time horizon $T < \infty$ and a filtration $\mathbb{F} = (\mathcal{F}_t)_{t \in [0, T]}$ satisfying the usual conditions of right continuity and completeness, in which all semimartingales are taken to have right continuous paths with left limits. We assume that \mathcal{F}_0 is trivial and $\mathcal{F} = \mathcal{F}_T$. On this stochastic basis, let W be a d -dimensional standard Brownian motion and let $\mu(dt, dx) = (\mu(w, dt, dx) | w \in \Omega)$ denote an integer-random valued measure on $([0, T] \times E, \mathcal{B}([0, T]) \otimes E)$ with compensator $\nu := \nu^P(w, dt, dx)$ under P , where $E := \mathbb{R}^d \setminus \{0\}$ is equipped with its Borel σ -field $\mathcal{E} := \mathcal{B}(E)$.

On $(\tilde{\Omega}, \tilde{\mathcal{F}}) := (\Omega \times [0, T] \times E, \mathcal{P} \otimes \mathcal{B}([0, T]) \otimes \mathcal{E})$, we define the measure $P \times \nu$ by

$$P \times \nu(\tilde{B}) = E \int_{[0, T] \times E} \mathbb{1}_{\tilde{B}}(w, t, x) \nu(w, dt, dx), \quad \tilde{B} \in \tilde{\mathcal{F}}. \quad (2.2)$$

We denote by $\tilde{\mathcal{P}} := \mathcal{P} \otimes \mathcal{E}$ where \mathcal{P} is the predictable σ -field on $\Omega \times [0, T]$. We say that a function on $\tilde{\Omega}$ is *predictable* if it is $\tilde{\mathcal{P}}$ -measurable.

We will assume that the compensator ν is absolutely continuous with respect to the $\lambda \otimes dt$ with a density ξ :

$$\nu(w, dt, dx) = \xi_t(w, x) \lambda(dx) dt, \quad (2.3)$$

where λ is a σ -finite measure on (E, \mathcal{E}) , that satisfies the following condition: $\int_E 1 \wedge |x|^2 \lambda(dx) < \infty$ and the density ξ is $\tilde{\mathcal{P}}$ -measurable, positive and bounded:

$$0 < \xi_t(w, x) \leq C_\nu < \infty, \quad P \times \lambda(dx) \times dt - a.e. \quad \text{for some constant } C_\nu. \quad (2.4)$$

Note that, thanks to (2.3), we have that $v(\{t\} \times E) = 0$ for all t , and $v([0, T] \times E) \leq C_v T \lambda(E)$. For ψ a predictable function on $\tilde{\Omega}$, we define its integral process with respect to μ as:

$$(\psi \boxtimes \mu)_t := \begin{cases} \int_{[0, t] \times E} \psi_t(x) \mu(w, ds, dx) & \text{if } \int_{[0, t] \times E} |\psi_t(w, x)| \mu(w, ds, dx) < \infty, \\ +\infty & \text{otherwise.} \end{cases} \quad (2.5)$$

In the same way, we define the integral process with respect to v . Let μ be the compensated measure of μ

$$\mu(w, dt, dx) = \mu(w, dt, dx) - v(w, dt, dx). \quad (2.6)$$

To alleviate the notations, we will omit the dependence on w in the different stochastic quantities. In the following, we recall some properties that can be found in Becherer [8] or Jacod and Shiryaev [70]. First, for any predictable function ψ , the process $\psi \boxtimes v$ is a predictable process whereas $\psi \boxtimes \mu$ is an optional process. We recall that $E[|\psi| \boxtimes \mu_T] = E[|\psi| \boxtimes v_T]$. If $(|\psi|^2 \boxtimes \mu)^{1/2}$ is locally integrable, then ψ is integrable with respect to μ and $\psi \boxtimes \mu$ is defined as the purely discontinuous local martingale (under P) with jump process $\int_E U\mu(\{t\}, dx)$. If the increasing process $|\psi| \boxtimes \mu$ (or equivalently, $|\psi| \boxtimes v$) is locally integrable, then again, ψ is integrable with respect to μ and is the purely discontinuous local martingale as in the first case and we have

$\psi \boxtimes \mu = \psi \boxtimes \mu - \psi \boxtimes v$. Finally, if the process $|\psi|^2 \boxtimes v$ is integrable, then U is integrable with respect to μ and $Z \boxtimes \mu$ is a square integrable martingale, purely discontinuous, with predictable quadratic variation $\langle \psi \boxtimes \mu \rangle = |\psi|^2 \boxtimes v$. These properties and their proofs can be found in Section II.1.d of Jacod and Shiryaev [70].

We will assume that W and μ satisfy the following weak representation property with respect to P and F : Every local martingale M with respect to (F, P) admits the following decomposition:

$$M_t = M_0 + (\eta \cdot W)_t + (\psi \boxtimes \mu)_t := M_0 + \int_0^t \eta_s dW_s + (\psi \boxtimes \mu)_t \quad \forall t \geq 0, P - a.s. \quad (2.7)$$

where η is a progressively measurable process and ψ a predictable process such that

$$\int_0^T |\eta_s|^2 ds < \infty, \quad (|\psi|^2 \boxtimes v)_T < \infty, \quad P - a.s. \quad (2.8)$$

We introduce the following spaces:

- L^{exp} is the space of all F_T -measurable random variables X such that:

$$E[\exp(\gamma|X|)] < \infty, \quad \forall \gamma > 0.$$

- $D_{\mathbb{F}}^{\text{exp}}$ is the space of progressively measurable processes $(X_t)_{0 \leq t \leq T}$ with

$$E \exp \gamma \operatorname{ess\,sup}_{0 \leq t \leq T} |X_t| < \infty, \quad \forall \gamma > 0.$$

- $D_{\mathbb{F}}^{\text{exp}}$ is the space of progressively measurable processes $(X_t)_{0 \leq t \leq T}$ with

$$E \exp \gamma \int_0^T |X_s| ds < \infty, \quad \forall \gamma > 0.$$

- $H_{\lambda}^{2,p}$ is the space of predictable processes ψ such that

$$E \int_0^T |\psi_{s,\lambda}^p|^2 ds < \infty,$$

where

$$|\psi_{s,\lambda}^p|^2 := \int_E |\psi_s(x)|^2 \xi_s(x) \lambda(dx).$$

- $H^{2,p}$ is the set of all predictable processes η such that

$$E \int_0^T |\eta_s|^2 ds < \infty.$$

2.2.2 The optimization problem

For every probability $Q \ll P$ on \mathbb{F}_T , we denote by $D = (D_t)_{0 \leq t \leq T}$ its Radon-Nikodym density with respect to P , that is,

$$D_t = E \frac{dQ}{dP} \cdot \mathbb{F}_t, \quad t \geq 0.$$

D is a càdlàg nonnegative P -martingale. Let $\tau_n := \inf\{t \geq 0, D_t \leq 1/n\}$ and consider the local martingale $M^n = \int_0^{t \wedge \tau_n} D^{-1} dD_s$. Thanks to the weak representation property, there exist two predictable processes $(\eta^n)_s$ and $(\psi^n)_s$, $s \leq \tau_n$, such that, $\int_0^{t \wedge \tau_n} |\eta^n|^2 ds < \infty$ and $(|\psi^n|^2 \otimes \nu)_{t \wedge \tau_n} < \infty$ and

$$D_{t \wedge \tau_n} = E((\eta^n \cdot W) + (\psi^n \otimes \mu))_{t \wedge \tau_n}, \quad t \geq 0, \quad P - a.s.$$

Consistency requires that we should have ${}^t \eta^n = {}^t \eta^{n+1}$ $dt \otimes dP$ -a.e and ${}^t \psi^n(x) = {}^t \psi^{n+1}(x) \nu(dx) \otimes dt \otimes dP$ -a.e on $\{t \leq \tau_n \wedge T\}$. By the fact that $\tau_n \nearrow \infty$ Q -a.s., we obtain the existence of Q -a.s. defined predictable processes η and ψ such that,

$$D_t = E((\eta \cdot W) + (\psi \otimes \mu))_t, \quad t \geq 0, \quad Q - a.s. \tag{2.9}$$

where $\int_0^T |\eta_s|^2 ds < \infty$ and $(|\psi|^2 \boxtimes \nu)_T < \infty$ $Q - a.s.$. Note that since for all $t \in [0, T]$, $D_t > 0$ $Q - a.s.$, then we must have for every $t \in [0, T]$, $\psi_t(x) > -1$ $dQ \times \nu(dx, dt)$ -a.e. and we can rewrite (D_t) as in the following:

$$D_t = \exp \left((\eta \cdot W)_t + (\psi \boxtimes \mu)_t - \frac{1}{2} \int_0^t |\eta_s|_2^2 ds + ((\ln(1 + \psi) - \psi) \boxtimes \mu)_t \right) \quad Q - a.s. \quad (2.10)$$

We now introduce the following time consistent penalty for a probability $Q \ll P$ on F_T :

$$\gamma_t^Q := E_Q \left[\int_t^T r(s, w, \eta_s, \psi_s) ds \mid \mathcal{F}_t \right] \quad (2.11)$$

where $r : [0, T] \times \Omega \times \mathbb{R}^d \times L^2(E, \lambda) \rightarrow [0, +\infty]$ is a suitable measurable function that is convex and lower-semicontinuous in (η, ψ) and such that $r(t, 0, 0) = 0$. Note that, since r is non-negative, r is minimal at $\eta = 0$ and $\psi = 0$ and this corresponds to the probabilistic model P . Therefore, the reference probability has the highest plausibility. In the following, we will consider probabilities $Q \in \mathcal{Q}^f$ where

$$\mathcal{Q}^f = \{Q \ll P, \gamma_0(Q) < \infty\}. \quad (2.12)$$

In order to solve the stochastic control problem with BSDEs, we need to impose some regularity and growth conditions on the penalty function. In a Brownian setting, Faidi et al. [47] assumed the penalty function to be bounded from below by the relative entropy. In the same way, we will assume that there exists $\tilde{K}_2, \tilde{K}_1 > 0$ such that,

$$\gamma_0(Q) \geq -\tilde{K}_2 + \tilde{K}_1 H(Q|P).$$

Let f be the function defined as follow:

$$f(x) = \begin{cases} (1+x) \log(1+x) - x, & \text{if } x \geq -1; \\ \infty & \text{otherwise;} \end{cases} \quad (2.13)$$

For the latter inequality to be verified, a sufficient condition on r will be the following:

(ffi_r) There exists $K_1, K_2 > 0$ such that for all $w \in \Omega$, $t \in [0, T]$, $\eta \in \mathbb{R}^d$ and $\psi \in L^2(E, \lambda; \mathbb{R})$, we have,

$$r(t, w, \eta, \psi) \geq -K_2 + K_1 \left(\frac{|\eta|^2}{2} + \int_E f(\psi(x)) \xi_t(w, x) \lambda(dx) \right).$$

The following proposition shows that the entropic penalty can be retrieved with a special choice for r . A detailed proof is given in the Appendix 2.5.

Proposition 2.2.1. Let $r(t, \eta, \psi) = \frac{1}{2} |\eta|^2 + \int_E f(\psi(x)) \xi_t(x) \lambda(dx)$ and $Q \in \mathcal{Q}^f$. Then, the penalty

function corresponds to the relative entropy, that is

$$V_0(Q) = E_Q \left[\int_0^T \frac{|\eta_t|^2}{2} + \int_E f(\psi_t(x)) \xi_t(x) \lambda(dx) \right] dt = H_0(Q|P),$$

where,

$$H(Q|P) = E_Q \log \frac{dQ}{dP}. \quad (2.14)$$

Moreover, we have for a general r verifying (ffi_r),

$$H(Q|P) \leq \frac{V_0(Q)}{K_1} + \frac{TK_2}{K_1}. \quad (2.15)$$

In particular, $H(Q|P)$ is finite for all $Q \in Q_f$.

Remark 2.2.2. Let $r^*(t, \cdot, \cdot) = \sup_{\eta \in \mathbb{R}^d, \psi \in L^2(\lambda)} (\cdot \eta + \int_E \tilde{\psi}(x) \psi(x) \lambda(dx) - r(t, \eta, \psi))$ the Fenchel conjugate r . Assumption (ffi_r) implies that, for $w \in \Omega$, $\tilde{\psi} \in \mathbb{R}^d$, $t \in [0, T]$ and $\tilde{\psi} \in L^2(E, \lambda; \mathbb{R})$,

$$r^*(t, w, \tilde{\psi}, \tilde{\psi}) \leq K_2 + \frac{|\tilde{\psi}|^2}{2K_1} + K_1 \int_E f \left(\frac{\tilde{\psi}(x)}{K_1 \xi_t(x)} \right) \xi_t(x) \lambda(dx), \quad (2.16)$$

where $f^*(x) := e^x - x - 1$ is the Fenchel conjugate of the function f .

Now, given a positive adapted process δ , we define the discounting process:

$$\delta := \exp \left(- \int_0^t \delta_s ds \right), \quad 0 \leq t \leq T,$$

and the auxiliary quantities,

$$U_{t,T}^\delta := \alpha \int_t^T \frac{S_s^\delta}{S_t^\delta} U_s ds + \alpha \frac{S_T^\delta}{S_t^\delta} U_T, \quad 0 \leq t \leq T, \quad \alpha, \alpha \geq 0,$$

$$R_{t,T}^\delta(Q) := \int_t^T \frac{S_s^\delta}{S_t^\delta} r(s, \eta_s, \psi_s) ds, \quad 0 \leq t \leq T.$$

Now we consider the cost functional

$$c(w, Q) := U_{0,T}^\delta(w) + \beta R_{0,T}^\delta(Q)(w), \quad (2.17)$$

which consists of two terms. The first one is a discounted utility term that is the sum of a final utility \bar{U}_T and a continuous utility with utility rate (U_s) . For instance, (U_s) can be seen as the utility coming from investing/consuming and \bar{U}_T as the utility coming from the terminal wealth. The second term is simply the penalty term and measure the “distance” between the probability Q and the reference probability P . The parameter β might be viewed as measuring the degree

of confidence of the reference probability P . The higher β is, the more confident we are in P , with the limiting case $\beta \uparrow \infty$ (respectively $\beta \downarrow 0$) corresponding to full degree of confidence (respectively distrust).

Our objective is to solve the following optimization problem:

$$\text{Minimize the functional } Q \mapsto \Gamma(Q) := E_Q[c(\cdot, Q)], \quad (2.18)$$

over the set $Q_{\mathcal{F}}$. To guarantee the well-posedness of the problem, we will assume the following:

(ffi_U) i. The discounting process is bounded by some constant $\|\delta\|_{\infty}$;

ii. The process U belongs to D_1^{exp} ;

iii. The terminal utility \bar{U}_T belongs to L^{exp} .

Remark 2.2.3. Under assumption (ffi_U), we have

$$E \exp \lambda \int_0^T |U_s| ds + \mu |\bar{U}_T| < \infty, \quad \forall \lambda, \mu \in \mathbb{R}^+. \quad (2.19)$$

Indeed, using the convexity of the exponential function, we get,

$$\begin{aligned} E \exp \lambda \int_0^T |U_s| ds + \mu \bar{U}_T &= E \exp \frac{1}{2} \times 2\lambda \int_0^T |U_s| ds + \frac{1}{2} \times 2\mu \bar{U}_T \\ &\leq E \frac{1}{2} \exp 2\lambda \int_0^T |U_s| ds + \frac{1}{2} \exp 2\mu |\bar{U}_T| \\ &= \frac{1}{2} E \exp 2\lambda \int_0^T |U_s| ds + \frac{1}{2} E \exp 2\mu |\bar{U}_T| < \infty, \end{aligned}$$

where the finiteness of the two last expectations is due to assumption (ffi_U).

2.3 Some helpful estimates and existence of optimal probability

2.3.1 Auxiliary estimates

The main objective of this section is to prove the existence of an optimal probability Q^* that minimizes the functional Γ . To achieve this, we start by proving some useful auxiliary estimates. We will adapt the steps in Bordigoni et al. [18] and the inequalities therein into our setting.

Proposition 2.3.1. Under assumption (ffi_r) and (ffi_U), there exists a constant $C \in (0, \infty)$ which depends only on $\bar{\alpha}, \alpha, \beta, \delta, T, \bar{U}, \bar{U}_T$ such that

$$\Gamma(Q) \leq E_Q[[c(\cdot, Q)]] \leq C(1 + \gamma_0(Q)), \quad \text{for all } Q \in Q_{\mathcal{F}}. \quad (2.20)$$

In particular, this shows that $\Gamma(Q)$ is well defined and finite for every $Q \in Q_f$.

Proof. The first inequality is obvious. As for the second, denoting $U := \alpha \int_0^T |U_t| dt + \alpha |U_T|$, we

have for $Q \in Q_f$, using the fact that, $0 \leq S^{\sharp} \leq 1$,

$$E_Q[|c(\cdot, Q)|] \leq E_Q \left[\alpha \int_0^T |U_t| dt + \bar{\alpha} |U_T| \right] + \beta E_Q \int_0^T r(t, \eta_t, \psi_t) dt \quad (2.21)$$

$$= E_Q[U] + \beta \gamma_0(Q). \quad (2.22)$$

Fenchel inequality applied to $x \mapsto x \log(x)$, gives

$$xy \leq \frac{1}{\lambda} (x \log(x) + e^{\lambda y - 1}), \quad \forall (x, y, \lambda) \in \mathbb{R}_+^* \times \mathbb{R}_+^* \times \mathbb{R}^*. \quad (2.23)$$

Therefore, using this inequality with $\lambda = 1$, we get,

$$\begin{aligned} E_Q[U] &= E \frac{dQ}{dP} U \\ &\leq E \frac{dQ}{dP} \log \frac{dQ}{dP} + E[e^{U-1}] \\ &= H(Q|P) + E[e^{U-1}] \\ &\leq \frac{\gamma_0(Q)}{K_1} + \frac{TK_2}{K_1} + e^{-1} E[e^U], \end{aligned}$$

where we used (2.15) in the last inequality. Going back to (2.21), we obtain,

$$E_Q[|c(\cdot, Q)|] \leq \beta + \frac{1}{K_1} \gamma_0(Q) + \frac{TK_2}{K_1} + e^{-1} E[e^U],$$

where the term $E[e^U]$ is finite as pointed out in remark 2.2.3. We then conclude by setting $C := \max \left\{ \beta + \frac{1}{K_1}, \frac{TK_2}{K_1} + e^{-1} E[e^U] \right\}$. \square

The next result shows that Γ is bounded from below by $\gamma_0(Q)$. This will be very useful for proving the existence of an optimal probability.

Proposition 2.3.2. *Assume (ffi_r) and (ffi_u) hold. Then, there exists $C \in (0, \infty)$ depending on $\alpha, \bar{\alpha}, \beta, \delta, T, U, \bar{U}_T$ such that for all $Q \in Q_f$*

$$\gamma_0(Q) \leq C(1 + \Gamma(Q)). \quad (2.24)$$

In particular, we have $\inf_{Q \in Q_f} \Gamma(Q) > -\infty$

Proof. Using the same notation as in the proof of the previous proposition, we have,

$$E_Q[U_{0,T}^\delta] \geq -E_Q[U].$$

For every $\lambda \in \mathbb{R}^*$, using the inequality (2.23), we get,

$$\begin{aligned} E_Q[U] &\leq \frac{1}{\lambda} H(Q|P) + \frac{1}{\lambda} E[e^{\lambda U-1}] \\ &\leq \frac{1}{\lambda} \frac{\nu_0(Q)}{K_1} + \frac{TK_2}{K_1} + \frac{e^{-1}}{\lambda} E[e^{\lambda U}], \end{aligned}$$

where we used (2.15) in the last inequality. On the other hand, since the discounting process is bounded from above, we have

$$E_Q[R_{0,T}] \stackrel{(\Delta)}{=} E_Q \left[\int_0^T S_t^\delta r(t, \eta_t, \psi_t) dt \right] \geq e^{-\|\delta\|_\infty T} E_Q \left[\int_0^T r(t, \eta_t, \psi_t) dt \right] = e^{-\|\delta\|_\infty T} \nu_0(Q).$$

Combining the two previous inequalities leads to the following,

$$\Gamma(Q) = E_Q[U_{0,T}^\delta + \beta R_{0,T}^\delta] \geq \beta e^{-\|\delta\|_\infty T} - \frac{1}{\lambda K_1} \nu_0(Q) - \frac{TK_2}{\lambda K_1} - \frac{e^{-1}}{\lambda} E[e^{\lambda U}]$$

Choosing λ large enough such that $\mu := \beta e^{-\|\delta\|_\infty T} - \frac{1}{\lambda K_1} > 0$, we get the desired result by setting $C := \frac{1}{\mu} \max \left\{ 1, \frac{TK_2}{\lambda K_1} + \frac{e^{-1}}{\lambda} E[e^{\lambda U}] \right\}$. \square

The following is a direct consequence of the previous proposition and inequality (2.15).

Corollary 2.3.3. *Under assumptions (ffi_r) and (ffi_u), there exists $K \in (0, \infty)$ such that for every $Q \in Q_f$, we have the following*

$$H(Q|P) \leq K(1 + \Gamma(Q)). \tag{2.25}$$

In the same spirit of the proof of the above proposition, we have the following estimate that is crucial in proving the existence of an optimal probability $Q^* \in Q_f$.

Lemma 2.3.4. *For any $\lambda > 0$ and any measurable set $A \in F_T$, we have for every $Q \in Q_f$*

$$E_Q[|U_{0,T}^\delta|_{1A}] \leq \frac{\nu_0(Q)}{\lambda K_1} + \frac{TK_2}{\lambda K_1} + \frac{e^{-1}}{\lambda} E_{1A} \exp(\lambda \alpha \int_0^T |U_s| ds + \lambda \bar{\alpha} |U_T|). \tag{2.26}$$

Proof. Using inequality (2.23), we have for every $\lambda > 0$ and $Q \in Q_f$,

$$\frac{dQ}{dP} |U_{0,T}^\delta|_{1A} \leq \frac{1}{\lambda} \frac{dQ}{dP} \log \frac{dQ}{dP} + e^{-1} e^{\lambda U} |_{1A}.$$

Taking the expectation under P and using (2.15), we consequently get, $E_Q[|U_{0,T}^\delta|_{1A}] \leq \frac{\nu_0(Q)}{\lambda K_1} + \frac{TK_2}{\lambda K_1} + \frac{e^{-1}}{\lambda} E_{1A} \exp(\lambda U)$. \square

2.3.2 Existence of optimal probability

In this subsection, we prove the existence of an optimal probability $Q^* \in \mathcal{Q}_{\mathcal{F}}$ using a standard Komlòs-type argument, but before let us show two important properties of the functionals Γ and γ_0 . We will introduce the following Linderberg condition on sequences of martingales converging almost surely to 0. This technical assumption is needed to prove the lower-semicontinuity of γ_0 :

(ffi_L) Every sequence (M_n) of locally square integrable martingales with the representation $dM_t^n = \eta_t^n dW_t + \int_E \psi_t^n(x) \mu(dt, dx)$ converging P -a.s. to 0 for each $t \in [0, T]$, verifies the following Linderberg condition:

$$\forall \epsilon \in (0, 1], \int_0^T \int_E |\psi_t^n(x)|^2 1_{|\psi_t^n(x)| \geq \epsilon} \nu(dt, dx) \xrightarrow{n \rightarrow \infty} 0, \text{ in } P - \text{Probability}$$

Proposition 2.3.5. *Under the assumption (ffi_L), we have the following:*

1. $\mathcal{Q}_{\mathcal{F}}$ is a convex set and the functional $Q \in \mathcal{Q}_{\mathcal{F}} \mapsto \Gamma(Q)$ is convex.
2. γ_0 is lower-semicontinuous for $L^1(P)$ convergence.

Proof. 1. Let $\lambda \in (0, 1)$, $Q, \tilde{Q} \in \mathcal{Q}_{\mathcal{F}}$ and $Q^\lambda := \lambda Q + (1 - \lambda)\tilde{Q}$. Let D and \tilde{D} denote the corresponding density processes and $(\eta, \psi), (\tilde{\eta}, \tilde{\psi})$ the associated processes via (2.9).

Consider the following processes:

$$\eta_t^\lambda := \frac{\lambda D_{t-} \eta_t + (1 - \lambda) \tilde{D}_{t-} \tilde{\eta}_t}{D_{t-}^\lambda} 1_{D_{t-}^\lambda > 0},$$

$$\psi_t^\lambda := \frac{\lambda D_{t-} \psi_t + (1 - \lambda) \tilde{D}_{t-} \tilde{\psi}_t}{D_{t-}^\lambda} 1_{D_{t-}^\lambda > 0},$$

where $D^\lambda := \lambda D + (1 - \lambda)\tilde{D}$ is the density process of Q^λ with respect to P . It is easy to see that the density D^λ satisfies the following SDE:

$$dD_t^\lambda = D_{t-}^\lambda \left(\eta_t^\lambda dW_t + \int_E \psi_t^\lambda(x) \tilde{\mu}(dt, dx) \right), t \in [0, T], Q^\lambda - a.s.$$

Hence, using the convexity assumption of r , we have,

$$\begin{aligned} \gamma_0(Q^\lambda) &= E_{Q^\lambda} \int_0^T r(t, \eta_t^\lambda, \psi_t^\lambda) dt \\ &\leq E_{Q^\lambda} \int_0^T \left(\lambda \frac{D_{t-}^\lambda}{D_{t-}^\lambda} r(t, \eta_t, \psi_t) + (1-\lambda) \frac{\tilde{D}_{t-}}{D_{t-}^\lambda} r(t, \eta_t, \tilde{\psi}_t) \right) dt \\ &= E \int_0^T \left(\lambda \frac{D_{t-}^\lambda}{D_{t-}^\lambda} r(t, \eta_t, \psi_t) + (1-\lambda) \frac{\tilde{D}_{t-}}{D_{t-}^\lambda} r(t, \eta_t, \tilde{\psi}_t) \right) dt \end{aligned}$$

Using Fubini's Theorem to interchange integral and expectation followed by conditioning on F_t and the martingale property of the density process D^λ , yields,

$$\gamma_0(Q^\lambda) \leq E \int_0^T \left(\lambda \frac{D_{t-}^\lambda}{D_{t-}^\lambda} r(t, \eta_t, \psi_t) + (1-\lambda) \frac{\tilde{D}_{t-}}{D_{t-}^\lambda} r(t, \eta_t, \tilde{\psi}_t) \right) dt$$

Since D^λ is right continuous, the set $\{t \in [0, T], D_t^\lambda \neq D_{t-}^\lambda\}$ is countable. Therefore, we have,

$$\begin{aligned} \gamma_0(Q^\lambda) &= E \int_0^T \left(\lambda \frac{D_{t-}^\lambda}{D_{t-}^\lambda} r(t, \eta_t, \psi_t) + (1-\lambda) \frac{\tilde{D}_{t-}}{D_{t-}^\lambda} r(t, \eta_t, \tilde{\psi}_t) \right) dt \\ &\leq E \int_0^T \left(\lambda D_{t-} r(t, \eta_t, \psi_t) - \lambda D_{t-} r(t, \eta_t, \tilde{\psi}_t) \right) dt \\ &= \lambda \gamma_0(Q) + (1-\lambda) \gamma_0(\tilde{Q}) < \infty. \end{aligned}$$

We have showed then that Q_F is convex. The convexity of the functional Γ follows readily by using the same arguments used above.

- Let (Q^n) be a sequence of probability measures that converges to Q in $L^1(\Omega, P)$, i.e., $D_T^n \rightarrow D_T$ in $L^1(\Omega, P)$ where D^n and D the corresponding densities processes. Let (η^n, ψ^n) and (η, ψ) (resp.) be the processes given by (2.9) of D^n and D (resp.). Since we know that D_T^n converges to D_T in $L^1(P)$, the maximal Doob's inequality

$$P\left(\sup_{0 \leq t \leq T} |D_t^n - D_t| \geq \epsilon\right) \leq \frac{1}{\epsilon} E[|D_T^n - D_T|], \quad \forall \epsilon > 0,$$

implies that $(\sup_{0 \leq t \leq T} |D_t^n - D_t|)$ converges to 0 in P -probability. By passing by a subsequence, we can assume that $(\sup_{0 \leq t \leq T} |D_t^n - D_t|)$ converges to 0 P -a.s.

We denote $M_t^n := \sup_{0 \leq s \leq t} |D_s^n - D_s|$ and introduce the following stopping time $\tau_n := \inf\{t \in [0, T], M_t^n \geq 1\} \wedge T$. We have $M_{\tau_n}^n \leq M_{\tau_n}^n + |D_{\tau_n}^n - D_{\tau_n}|$ and by taking expectation in the

latter we get

$$E[M_{\tau_n}^n] \leq E[M_{\tau_n}^n] + E[|D_{\tau_n}^n - D_{\tau_n}|]. \quad (2.27)$$

Recall that $M_t^n \rightarrow 0$ and since $(M_t^n)_t$ is a nondecreasing process we have $M_{\tau_n}^n \leq M_T^n$ so that $M_{\tau_n}^n \rightarrow 0$. We also have by the definition of the stopping time τ_n that $M_{\tau_n}^n \leq 1$. Hence, by the dominated convergence theorem, we obtain that

$$E[M_{\tau_n}^n] \rightarrow 0 \text{ as } n \rightarrow \infty. \quad (2.28)$$

Furthermore,

$$\begin{aligned} E[|D_{\tau_n}^n - D_{\tau_n}|] &= E[|E[D_T^n | \mathcal{F}_{\tau_n}^n] - E[D_T | \mathcal{F}_{\tau_n}^n]|] = E[|E[D_T^n - D_T | \mathcal{F}_{\tau_n}^n]|] \\ &\leq E[E[|D_T^n - D_T| | \mathcal{F}_{\tau_n}^n]] = E[|D_T^n - D_T|] \rightarrow 0. \end{aligned} \quad (2.29)$$

Combining (2.27), (2.28) and (2.29), we deduce that $M_{\tau_n}^n$ converges to 0 in $L^1(P)$. Then, by Burkholder-Davis-Gundy's inequality, we get that $[D^n - D]_{\tau_n}^2$ converges to 0 in $L^1(P)$ and a fortiori in P -probability. Now, as $[D^n - D]_T = [D^n - D]_{\tau_n \wedge \tau_n = T} + [D^n - D]_{T \wedge \tau_n < T}$, then for every $\epsilon > 0$,

$$\begin{aligned} P([D^n - D]_T \geq \epsilon) &\leq P([D^n - D]_{\tau_n \wedge \tau_n = T} \geq \epsilon) + P([D^n - D]_{T \wedge \tau_n < T} \geq \epsilon) \\ &\leq P([D^n - D]_{\tau_n} \geq \epsilon) + P(\tau_n < T), \end{aligned}$$

and

$$P(\tau_n < T) = P(\exists t \in [0, T] \text{ s.t. } M_t^n \geq 1) \leq P(M_T^n \geq 1) \rightarrow 0.$$

So, we get that $[D^n - D]_T$ converges to 0 in P -probability. On the other hand, since $D_t^n - D_t \rightarrow 0$, thanks to the assumption (ffil), we get from Corollary 1 in Shiriyayev [125] that $\langle D^n - D \rangle_T$ converges to 0 in P -probability and by passing to a subsequence while keeping the same notation, we may say that $\langle D^n - D \rangle_T$ converges to 0 P -a.s.. But, we know that,

$$\langle D^n - D \rangle_T = \int_0^T |D_t^n - D_t|^2 dt + \int_0^T \int_E |D_t^n \psi_t^n(x) - D_t \psi_t(x)|^2 \nu(dt, dx).$$

Therefore, we immediately obtain that $D_t^n \psi_t^n(x) \rightarrow D_t \psi_t(x)$ in $L^2(E, \lambda)$. Next, we will show that $\gamma_0(Q) \leq \liminf_{n \rightarrow \infty} \gamma_0(Q^n)$. Assume by way of contradiction that $\gamma_0(Q) > l := \liminf_{n \rightarrow \infty} \gamma_0(Q^n)$. By passing to a subsequence, we may assume that $\gamma_0(Q) > l$. Let $\zeta := \inf\{t \in [0, T], D_t = 0\}$. Since $D_t^n = 0$ on $\{t > \zeta_n\}$, we must have $\zeta \leq \liminf_{n \rightarrow \infty} \zeta_n$.

Hence, for $\epsilon := \frac{\gamma_0(Q) - l}{2}$, there is $k \in \mathbb{N}$ such that for $T_k := \zeta \wedge \{\zeta_k, \zeta_{k+1}, \dots\}$, we have

$$\begin{aligned} \gamma_0(Q) &= E_Q \int_0^{T_k} r(t, \eta_t, \psi_t) dt = E \int_0^{\zeta} D_{t-} r(t, \eta_t, \psi_t) dt \\ &\leq E \int_0^{T_k} D_{t-} r(t, \eta_t, \psi_t) dt + \epsilon \\ &= E \int_0^{T_k} g(t, D_{t-}, D_{t-} \eta_t, D_{t-} \psi_t) dt + \epsilon, \end{aligned}$$

where $g(t, x, y, z) := xr(t, \frac{y}{x}, \frac{z}{x})$. Clearly, since r is lower-semicontinuous in (η, ψ) , we get that also g is also lower-semicontinuous. Hence, by Fatou's lemma, we obtain

$$\begin{aligned} E \int_0^{T_k} g(t, D_{t-}, D_{t-} \eta_t, D_{t-} \psi_t) dt &\leq \liminf_{n \geq k} E \int_0^{T_k} g(t, D_{t-}^n, D_{t-}^n \eta_t, D_{t-}^n \psi_t) dt \\ &\leq \liminf_{n \geq k} E \int_0^{\zeta_k} g(t, D_{t-}^n, D_{t-}^n \eta_t, D_{t-}^n \psi_t) dt \\ &= \liminf_{n \geq k} \gamma_0(Q^n) = l, \end{aligned}$$

so that we have $\gamma_0(Q) \leq l + \epsilon < \gamma_0(Q)$ which is a contradiction. □

In the next theorem, we show the existence of an optimal probability $Q^* \in Q_{\mathcal{F}}$.

Theorem 2.3.6. Assume (ffi_r), (ffi_u) and (ffi_L) hold. Then there exists a probability measure Q^* minimizing $Q \mapsto \Gamma(Q)$ over $Q_{\mathcal{F}}$.

Proof. Let Q^n a minimizing sequence in $Q_{\mathcal{F}}$ such that

$$\Gamma(Q^n) \searrow \inf_{Q \in Q_{\mathcal{F}}} \Gamma(Q),$$

and we denote by D^n the corresponding density process. Since we have $D_T^n \geq 0$, it follows from Komlòs' lemma that there exists a sequence denoted \overline{D}^n such that $\overline{D}^n \in \text{conv}(\mathcal{D}^n, \mathcal{D}^{n+1}, \dots)$ for each $n \in \mathbb{N}$ and (\overline{D}_T^n) converges $P - a.s.$ to a random variable \overline{D}_T^∞ . Now, we will show that \overline{D}_T^∞ is associated with a probability measure Q^∞ . First, we have \overline{D}_T^∞ is nonnegative as the $P - a.s.$ limit of the nonnegative sequence $(\overline{D}_T^n)_n$. Second, since $Q_{\mathcal{F}}$ is convex, each \overline{D}_T^n is associated with a probability measure $\overline{Q}^n \in Q_{\mathcal{F}}$. Now, thanks to the convexity of Γ and the fact that $(\Gamma(Q^n))_n$ is decreasing, we have the following,

$$\Gamma(\overline{Q}^n) \leq \sup_{m \geq n} \Gamma(Q^m) = \Gamma(Q^n) \leq \Gamma(Q^1). \quad (2.30)$$

Consequently, using (2.25), we get,

$$\sup_{n \in \mathbb{N}} E\left[\frac{D^n}{T} \log\left(\frac{D^n}{T}\right)\right] = \sup_{n \in \mathbb{N}} H(Q^n | P) \leq K(1 + \sup_{n \in \mathbb{N}} \Gamma(\frac{Q^n}{Q})) \leq K(1 + \Gamma(Q^1)) < \infty.$$

By Vallée-Poussin's criterion, the sequence $(\frac{D^n}{T})$ is P -uniformly integrable and therefore converges to $\frac{D^\infty}{T}$ in $L^1(P)$. Hence, we have, $E[\frac{D^\infty}{T}] = \lim_{n \rightarrow \infty} E[\frac{D^n}{T}] = 1$ since $E[\frac{D^n}{T}] = 1$ for all $n \in \mathbb{N}$. This shows that $\frac{D^\infty}{T}$ can be associated with a probability Q^∞ on F_T such that $dQ^\infty = \frac{D^\infty}{T} dP$. Our next step is to prove that this probability Q^∞ belongs to $Q_{\mathcal{F}}$. By Proposition 2.3.5, we know that γ_0 is lower-semicontinuous with respect to $L^1(\Omega, P)$ convergence. Therefore, we get since $\frac{D^n}{T} \xrightarrow{L^1} \frac{D^\infty}{T}$,

$$\gamma_0(Q^\infty) \leq \liminf_{n \rightarrow \infty} \gamma_0(Q^n).$$

But, thanks to (2.24), we know that $\gamma_0(Q) \leq C(1 + \Gamma(Q))$. Consequently, we obtain that,

$$\liminf_{n \rightarrow \infty} \gamma_0(Q^n) \leq C(1 + \sup_{n \in \mathbb{N}} \Gamma(\frac{Q^n}{Q})).$$

The RHS of the last inequality is finite thanks to (2.30). We then conclude that $\gamma_0(Q^\infty) < \infty$, i.e., $Q^\infty \in Q_{\mathcal{F}}$. It remains to show that Q^∞ is optimal. Note that using the same arguments in the proof of Proposition 2.3.5, the function $Q \mapsto E_Q[R_{0,T}^\delta(Q)] = E_Q[\int_0^T S^\delta(t, \eta_t, \psi_t) dt]$ is lower-semicontinuous for $L^1(\Omega, P)$ convergence and therefore we get immediately that,

$$E_{Q^\infty}^h R_{0,T}^\delta(Q^\infty) \leq \liminf_{n \rightarrow \infty} E_{Q^n}^h R_{0,T}^\delta(Q^n).$$

We denote $Y^n := \frac{D^n}{T} U^\delta$ and $Y^\infty := \frac{D^\infty}{T} U^\delta$. If we prove that we also have

$$E[Y^\infty] \leq \liminf_{n \rightarrow \infty} E[Y^n], \quad (2.31)$$

then we would have

$$\begin{aligned} \Gamma(Q^\infty) &= E[Y^\infty] + E_{Q^\infty}^h [R_{0,T}^\delta(Q^\infty)] \\ &\leq \liminf_{n \rightarrow \infty} E[Y^n] + \liminf_{n \rightarrow \infty} E_{Q^n}^h R_{0,T}^\delta(Q^n) \\ &\leq \liminf_{n \rightarrow \infty} E[Y^n] + E_{Q^n}^h R_{0,T}^\delta(Q^n) \\ &= \liminf_{n \rightarrow \infty} \Gamma(Q^n) = \inf_{Q \in Q_{\mathcal{F}}} \Gamma(Q), \end{aligned}$$

which proves that indeed Q^∞ is optimal. Although Y^n is linear in $\frac{D^n}{T}$ we cannot use Fatou's

lemma since the term $U_{0,T}^\delta$ has no lower bound. To remediate this, we introduce the following:

$$\tilde{R}_m := U_{0,T}^\delta \mathbb{1}_{U_{0,T}^\delta \geq -m}, \quad m \in \mathbb{N}.$$

Hence, we have for $n \in \mathbb{N} \cup \{\infty\}$,

$$\bar{Y}^n = \bar{D}_T^n U_{0,T}^\delta = \bar{D}_T^n \tilde{R}_m + \bar{D}_T^n U_{0,T}^\delta \mathbb{1}_{U_{0,T}^\delta < -m}.$$

Because now \tilde{R}_m is bounded below by $-m$, we can use Fatou's lemma to get,

$$E[\bar{D}^\infty \tilde{R}_m] \leq \liminf_{n \rightarrow \infty} E[\bar{D}^n \tilde{R}_m].$$

Consequently, by adding and subtracting $E[\bar{D}_T^n U_{0,T}^\delta \mathbb{1}_{U_{0,T}^\delta < -m}]$, we obtain,

$$\begin{aligned} E[\bar{Y}^\infty] &\leq \liminf_{n \rightarrow \infty} E[\bar{D}^n \tilde{R}_m] + E[\bar{D}_T^n U_{0,T}^\delta \mathbb{1}_{U_{0,T}^\delta < -m}] \\ &\leq \liminf_{n \rightarrow \infty} E[\bar{Y}^n] + 2 \sup_{n \in \mathbb{N} \cup \{\infty\}} E[\bar{D}_T^n |U_{0,T}^\delta| \mathbb{1}_{U_{0,T}^\delta < -m}]. \end{aligned}$$

The desired inequality (2.31) will follow once we prove that

$$\lim_{m \rightarrow \infty} \sup_{n \in \mathbb{N} \cup \{\infty\}} E[\bar{D}_T^n |U_{0,T}^\delta| \mathbb{1}_{U_{0,T}^\delta < -m}] = 0,$$

and this is where we use Lemma 2.3.4. Indeed, thanks to this lemma, we have,

$$\begin{aligned} E[\bar{D}_T^n |U_{0,T}^\delta| \mathbb{1}_{U_{0,T}^\delta < -m}] &= E[\bar{D}_T^n [|U_{0,T}^\delta| \mathbb{1}_{U_{0,T}^\delta < -m}]] \leq \frac{\nu_0(Q^n)}{\lambda K_1} + \frac{TK_2}{\lambda K_1} + \frac{e^{-1} E[\exp(\lambda U) \mathbb{1}_{U_{0,T}^\delta < -m}]}{\lambda} \\ &\leq \frac{C(1 + \Gamma(Q^n))}{\lambda K} + \frac{TK_2}{\lambda K} + \frac{\bar{e}^{-1} E[\exp(\lambda U) \mathbb{1}_{U_{0,T}^\delta < -m}]}{\lambda}. \end{aligned}$$

Using (2.30), we deduce that

$$\sup_{n \in \mathbb{N} \cup \{\infty\}} E[\bar{D}_T^n |U_{0,T}^\delta| \mathbb{1}_{U_{0,T}^\delta < -m}] \leq \frac{C(1 + \max(\Gamma(Q^1), \Gamma(Q^\infty)))}{\lambda K} + \frac{TK_2}{\lambda K} + \frac{e^{-1}}{\lambda} E[\exp(\lambda U) \mathbb{1}_{U_{0,T}^\delta < -m}].$$

By the dominated convergence theorem, the third term in the RHS of the previous inequality goes to 0 as $m \rightarrow \infty$. Hence, we for all $\lambda > 0$, we have

$$\lim_{m \rightarrow \infty} \sup_{n \in \mathbb{N} \cup \{\infty\}} E[\bar{D}_T^n |U_{0,T}^\delta| \mathbb{1}_{U_{0,T}^\delta < -m}] \leq \frac{C(1 + \max(\Gamma(Q^1), \Gamma(Q^\infty)))}{\lambda K} + \frac{TK_2}{\lambda K}.$$

Sending λ to ∞ , we finally obtain the desired result. \square

2.4 Related BSDE with jumps

This section is devoted to the study of the dynamic value process V associated to the optimization problem (2.18) using stochastic control techniques. More precisely, we prove that the dynamic process is the unique solution of a certain QEBSDEJ. This extends the previous work by Schroder and Skiadas [121], Skiadas [126] and Lazrak and Quenez [85].

We first introduce some notations that we will use below. Let S denote the set of all stopping time τ with values in $[0, T]$ and D the space of all density processes D^Q with $Q \in \mathcal{Q}_f$. We also define,

$$\begin{aligned} D(Q, \tau) &:= \{Q \in \mathcal{Q}_f, Q = Q \text{ on } \mathcal{F}_\tau\}, \\ \Gamma(Q, \tau) &:= E_Q[c(\cdot, Q)|\mathcal{F}_\tau]. \end{aligned}$$

As in Karoui [78], we define the minimal conditional cost at time τ by

$$J(Q, \tau) := Q - \operatorname{ess\,inf}_{Q \in D(Q, \tau)} \Gamma(Q, \tau).$$

For $Q \in \mathcal{Q}_f$ and $\tau \in S$, we now define the value of the control problem starting at time τ instead of 0 and assuming one has used the model Q up to time τ ,

$$\begin{aligned} \tilde{V}(Q, \tau) &:= E_Q[U_{\tau, T}^\delta | \mathcal{F}_\tau] + \beta E_Q[R_{\tau, T}^\delta(Q) | \mathcal{F}_\tau], \\ V(Q, \tau) &:= Q - \operatorname{ess\,inf}_{Q \in D(Q, \tau)} \tilde{V}(Q, \tau). \end{aligned}$$

The following martingale optimality principle is a consequence of Theorems 1.15, 1.17 and 1.21 in Karoui [78]. It is the analogue of Proposition 3.4 in Faidi et al. [47] in a Brownian setting but the proofs also hold in our setting with obvious modifications.

Proposition 2.4.1. *Under (ffi_u) and (ffi_r), we have:*

- *The family $\{J(Q, \tau) | \tau \in S, Q \in \mathcal{Q}_f\}$ is a submartingale system, that is for any $Q \in \mathcal{Q}_f$ and stopping times $\sigma \leq \tau$, we have,*

$$E_Q[J(Q, \tau) | \mathcal{F}_\sigma] \geq J(Q, \sigma) \quad Q - a.s. \quad (2.32)$$

- *$Q \in \mathcal{Q}_f$ is optimal if and only if the family $\{J(Q, \tau) | \tau \in S\}$ is a Q -martingale system which means that for any stopping times $\sigma \leq \tau$*

$$E_Q[J(Q, \tau) | \mathcal{F}_\sigma] = J(Q, \sigma) \quad Q - a.s.$$

- *For each $Q \in \mathcal{Q}_f$, there exists an adapted RCLL process $J^Q = (J_t^Q)_{t \in [0, T]}$ which is a right*

closed Q -submartingale such that for every stopping time τ

$$J_\tau^Q = J(Q, \tau) \quad Q - a.s.$$

Before stating the BSDE verified by the value process V , we will need to define a strong order relation on the set of increasing processes defined below.

Definition 2.4.2. Let X and Y two increasing processes. We say that $X \leq Y$ if the process $Y - X$ is increasing.

Theorem 2.4.3. Assume assumptions (ffi_r), (ffi_u) and (ffi_L) hold. If the optimal probability \bar{Q}^∞ in Theorem 2.3.6 is equivalent to P , then there exists Z and \tilde{Z} such that (V, Z, \tilde{Z}) is solution in $D_0^{\text{exp}} \times H^{2,p} \times H^{2,p}_\lambda$ of the following BSDE:

$$\begin{cases} dV_t = \delta V_t - \alpha U_t + \beta r^* - \frac{Z_t}{\beta} \xi - \frac{\tilde{Z}_t}{\beta} \xi \quad dt - Z \cdot dW_t - \int_E \tilde{Z}(x) \tilde{\mu}(dx, dt), \\ V_T = \alpha \bar{U}_T. \end{cases} \quad (2.33)$$

Proof. We will split the proof into three steps: First, we will prove that the value process V is a P -special martingale, that is it can be decomposed as $V = V_0 + M^V + A^V$, where M^V is a local martingale that can be written as $M^V = (Z \cdot W) + (\tilde{Z} \boxtimes \tilde{\mu})$ and A^V a predictable finite variation process. Then, we will show that (V, Z, \tilde{Z}) is a solution of the BSDE. Finally, we will prove that (V, Z, \tilde{Z}) is in the required spaces.

Step 1: First, note that since we assumed that $\bar{Q}^\infty \sim P$, then,

$$\inf_{Q \in Q_f} \Gamma(Q) = \inf_{Q \in Q_f^e} \Gamma(Q),$$

where $Q_f^e := \{Q \in Q_f, Q \sim P\}$ and we define $D^e(Q, \tau)$ accordingly. Hence, we will restrict our attention to probabilities $Q \in Q_f^e$ and all essential infimums can be taken with respect to P in the expression of $V(Q, \tau)$ and $J(Q, \tau)$, i.e.,

$$\begin{aligned} J(Q, \tau) &= P - \text{ess inf}_{Q \in D^e(Q, \tau)} \Gamma(Q, \tau), \\ V(Q, \tau) &= P - \text{ess inf}_{Q \in D^e(Q, \tau)} \tilde{V}(Q, \tau). \end{aligned}$$

By Bayes' formula and the definition of $R_{\tau, T}(Q)$, it is easy to see that $\tilde{V}(Q, \tau)$ depends only on the values of the density process D of Q on $[\tau, T]$ and is therefore independent of Q . Hence,

we can denote $V(Q, \tau)$ by $V(\tau)$. From the definition of $R_{0,\tau}^\delta(Q)$ and $U_{0,\tau}^\delta$, we have

$$\begin{aligned} R_{0,\tau}^\delta(Q) &= \int_0^\tau S_t^\delta r(t, q_t, \psi_t) dt + S_\tau^\delta R_{\tau,T}^\delta(Q), \\ U_{0,\tau}^\delta &= \alpha \int_0^\tau S_t^\delta U_t dt + U_{\tau,T}^\delta. \end{aligned}$$

Comparing $V(\tau)$ and $J(Q, \tau)$ yields for $Q \in \mathcal{Q}_f^e$ with density process $D^Q = E((\eta \cdot W) + (\psi \boxtimes \mu))$,

$$J(Q, \tau) = S_\tau^\delta V(\tau) + \alpha \int_0^\tau S_t^\delta U_t dt + \beta \int_0^\tau S_t^\delta r(t, \eta_t, \psi_t) dt, \quad P - a.s. \quad (2.34)$$

From the martingale optimality principle in Proposition 2.4.1, there exists an adapted RCLL process denoted $J^Q = (J_t^Q)_{t \in [0, T]}$ such that $J_\tau^Q = J(Q, \tau)$, $Q - a.s.$ From (2.34), we deduce that we can choose an adapted RCLL process $(V_t)_{t \in [0, T]}$ such that $V_\tau = V(\tau)$, $P - a.s.$ for all $\tau \in S$. We can then rewrite (2.34) for every $Q \in \mathcal{Q}_f^e$ as,

$$J_t^Q = S_t^\delta V_t + \alpha \int_0^t S_s^\delta U_s ds + \beta \int_0^t S_s^\delta r(s, \eta_s, \psi_s) ds, \quad dt \times dP - a.e. \quad (2.35)$$

As the probability $P \in \mathcal{Q}_f^e$ corresponds to $\eta = 0$ and $\psi = 0$ and $r(t, 0, 0) = 0$, we get in particular for $Q = P$ in (2.35) that $J^P = S^\delta V + \alpha \int_0^\cdot S^\delta U_s ds$. By Proposition 2.4.1, J^P is a P -submartingale and thus we deduce that V is a P -special semimartingale, i.e. its canonical decomposition can be written as

$$V = V_0 + M^V + A^V, \quad (2.36)$$

where M^V is a local martingale and A^V is a predictable finite variation process. By the weak representation assumption, the local martingale M^V can be written as:

$$M^V = -(Z \cdot W) - (\tilde{Z} \boxtimes \tilde{\mu}).$$

Step 2: We now prove that (V, Z, \tilde{Z}) is a solution of QEBSDEJ in (2.33). Plugging (2.36) into (2.35) yields

$$dJ_t^Q = -\delta_t S_t^\delta V_t dt + \alpha S_t^\delta U_t dt + S_t^\delta \beta r(t, \eta_t, \psi_t) dt - Z_t dW_t - \int_E \tilde{Z}_t(x) \tilde{\mu}(dx, dt) + dA_t^V.$$

For each $Q \in \mathcal{Q}_f^e$, we have, $D^Q = E((\eta \cdot W) + (\psi \boxtimes \mu))$, $P - a.s.$ and by Girsanov's theorem, we have, $dW_t^Q = dW_t - \eta_t dt$ is a Q Brownian motion and $\nu^Q(dx, dt) = (1 + \psi_t(x))\nu(dt, dx)$ is the

compensation of μ under Q . Rewriting the dynamic of J^Q , we obtain,

$$dJ_t^Q = -\delta_t S_t^\delta V_t dt + \alpha S_t^\delta U_t dt + S_t^\delta \int_{\mathcal{E}} \beta r(t, \eta_t, \psi_t) dt - Z_t \eta_t dt - \int_{\mathcal{E}} \tilde{Z}_t(x) \psi_t(x) v(dx, dt) + dA_t^V - S_t^\delta \int_{\mathcal{E}} Z_t dW_t^Q + \int_{\mathcal{E}} \tilde{Z}_t(x) \tilde{\mu}^Q(dx, dt) . \quad (2.37)$$

But, we know thanks to Proposition 2.4.1, that for every $Q \in \mathcal{Q}_f^e$, J^Q is a Q -submartingale and $J^{\bar{Q}^\infty}$ is a \bar{Q}^∞ -martingale. This means that we should have,

$$dA_t^V \geq Z_t \eta_t dt + \int_{\mathcal{E}} \tilde{Z}_t(x) \psi_t(x) v(dx, dt) - \beta r(t, \eta_t, \psi_t) dt + \delta_t V_t dt - \alpha S_t^\delta U_t dt, \quad dt \times dQ - a.e.$$

$$dA_t^V = Z_t \eta_t^\infty dt + \int_{\mathcal{E}} \tilde{Z}_t(x) \bar{\psi}_t^\infty(x) v(dx, dt) - \beta r(t, \bar{\eta}_t^\infty, \bar{\psi}_t^\infty) + \delta_t V_t dt - \alpha S_t^\delta U_t dt, \quad dt \times d\bar{Q}^\infty - a.e.$$

Note that the above inequality and equality are verified $dt \times dP - a.e.$ since $Q \in \mathcal{Q}_f^e$ and by the assumption that $\bar{Q}^\infty \in \mathcal{Q}_f^e$, in which case they become equivalent to,

$$dA_t^V \geq \operatorname{ess\,sup}_{\eta_t \in \mathbb{R}^d, \psi_t \in L^2(\lambda)} Z_t \eta_t dt + \int_{\mathcal{E}} \tilde{Z}_t(x) \psi_t(x) v(dx, dt) - \beta r(t, \eta_t, \psi_t) dt + \delta_t V_t dt - \alpha S_t^\delta U_t dt, \quad dt \times dP - a.e. \quad (2.38)$$

$$dA_t^V = Z_t \eta_t^* dt + \int_{\mathcal{E}} \tilde{Z}_t(x) \bar{\psi}_t^\infty(x) v(dx, dt) - \beta r(t, \bar{\eta}_t^\infty, \bar{\psi}_t^\infty) + \delta_t V_t dt - \alpha S_t^\delta U_t dt, \quad dt \times dP - a.e. \quad (2.39)$$

By denoting

$$r^*(t, \cdot, \cdot) = \sup_{\eta \in \mathbb{R}^d, \psi \in L^2(\lambda)} \left(\int_{\mathcal{E}} \tilde{Z}_t(x) \psi(x) \lambda(dx) - r(t, \eta, \psi) \right),$$

the Fenchel conjugate of r , equation (2.38) implies that $dt \times dP - a.e.$,

$$dA_t^V = \operatorname{ess\,sup}_{\eta_t \in \mathbb{R}^d, \psi_t \in L^2(\lambda)} Z_t \eta_t dt + \int_{\mathcal{E}} \tilde{Z}_t(x) \psi_t(x) v(dx, dt) - \beta r(t, \eta_t, \psi_t) dt + \delta_t V_t dt - \alpha S_t^\delta U_t dt$$

$$= \beta r^*(t, \frac{1}{\beta} Z_t, \frac{1}{\beta} \xi_t \tilde{Z}_t) dt + \delta_t V_t dt - \alpha S_t^\delta U_t dt$$

$$= Z_t \eta_t^\infty dt + \int_{\mathcal{E}} \tilde{Z}_t(x) \bar{\psi}_t^\infty(x) v(dx, dt) - \beta r(t, \bar{\eta}_t^\infty, \bar{\psi}_t^\infty) + \delta_t V_t dt - \alpha S_t^\delta U_t dt. \quad (2.40)$$

This shows in particular that

$$\frac{Z_t}{\beta}, \frac{\tilde{Z}_t}{\beta} \xi_t \in \partial r(t, \bar{\eta}_t^\infty, \bar{\psi}_t^\infty), \quad dt \times dP - a.e. \quad (2.41)$$

Going back to equation (2.36) and replacing the finite variation process A^V by its expression in (2.40), it follows that (V, Z, \tilde{Z}) is solution of the following equation,

$$\begin{aligned} dV_t &= \delta_t V_t - \alpha U_t + \beta r^* \left(t, \frac{Z_t}{\beta}, \xi_t \frac{\tilde{Z}_t}{\beta} \right) dt - Z dW_t - \int_E \tilde{Z}(x) \tilde{\mu}(dx, dt), \\ V_T &= \bar{\alpha} U_T. \end{aligned}$$

Step 3: In this step, we show that the $(V, Z, \tilde{Z}) \in D_0^{\exp} \times H_{\lambda}^{2,p} \times H^{2,2}$. $V \in D_0^{\exp}$ follows as in Faidi et al. [47]. As for Z and \tilde{Z} , the proof will lean on some exponential transform. We introduce the following processes defined for $t \in [0, T]$ as:

$$\begin{aligned} Y_t^- &= -CV_t + C \int_0^t (\alpha |U_s| + K_2 \beta) ds + C \int_0^t \delta_s |V_s| ds, \\ Y_t^+ &= CV_t + C \int_0^t (\alpha |U_s| + K_2 \beta) ds + C \int_0^t \delta_s |V_s| ds, \\ K_t^- &= \exp(Y_t^-), \quad K_t^+ = \exp(Y_t^+), \end{aligned}$$

where $C = \frac{1}{K_1 \beta}$. For any $p \geq 1$, we have

$$\begin{aligned} \sup_{t \in [0, T]} (K_t^\pm)^p &= \sup_{t \in [0, T]} \exp(p Y_t^\pm) \leq \sup_{t \in [0, T]} \exp\left(pC |V_t| + pC \int_0^t (\alpha |U_s| + K_2 \beta) ds + pC \int_0^t \delta_s |V_s| ds \right) \\ &\leq \exp(pC \sup_{t \in [0, T]} |V_t| + pC \alpha \int_0^T |U_s| ds + pCK_2 \beta T + pC \|\delta\|_{\infty} \sup_{t \in [0, T]} |V_t|). \end{aligned}$$

Since $V \in D_0^{\exp}$ and $U \in D_1^{\exp}$, from the above inequality we deduce that $\sup_{t \in [0, T]} K_t^\pm \in L^p(\Omega)$.

We turn our attention to the process Y^- . Using (2.33), the process Y^- verifies:

$$\begin{aligned} dY_t^- &= -CdV_t + C(\alpha |U_t| + K_2 \beta) dt + C\delta_t |V_t| dt \\ &= C \left(\delta_t (|V_t| - V_t) + \alpha (|U_t| + U_t) + K_2 \beta - \beta r^* \left(t, \frac{Z_t}{\beta}, \xi_t \frac{\tilde{Z}_t}{\beta} \right) \right) dt + CZ_t dW_t + \int_E C\tilde{Z}_t(x) \tilde{\mu}(dt, dx) \\ &= CK_2 \beta - C\beta r^* \left(t, \frac{Z_t}{\beta}, \xi_t \frac{\tilde{Z}_t}{\beta} \right) + \frac{|CZ_t|^2}{2} + \int_E f^*(C\tilde{Z}_t(x)) \xi_t(x) \lambda(dx) dt + C\delta_t (|V_t| - V_t) dt \\ &\quad + \left(\alpha |U_t| + K_2 \beta - \beta r^* \left(t, \frac{Z_t}{\beta}, \xi_t \frac{\tilde{Z}_t}{\beta} \right) \right) dt + \int_E C\tilde{Z}_t(x) \tilde{\mu}(dt, dx) \\ &= dL_t^- + dL_t^-, \end{aligned}$$

where

$$dL_t^- = -C\delta_t(|V_t| - V_t) + C\alpha(|U_t| + U_t) + CK_2\beta + CBr^* \left(\frac{Z_t}{\beta} + \xi_t \frac{\tilde{Z}_t}{\beta} + \frac{|CZ_t|^2}{2} \right) dt + \int_E f^*(C\tilde{Z}_t(x))\xi_t(x)\lambda(dx)dt$$

$$dL^- = CZ dW - \frac{|CZ_t|^2}{2} dt + \int_E C\tilde{Z}_t(x)\tilde{\mu}(dt, dx) - \int_E f^*(C\tilde{Z}_t(x))\xi_t(x)\lambda(dx)dt.$$

Thanks to inequality given in (2.16), we have the following:

$$-CBr^* \left(\frac{Z_t}{\beta} + \xi_t \frac{\tilde{Z}_t}{\beta} + \frac{|CZ_t|^2}{2} \right) + CK_2\beta + \int_E f^*(C\tilde{Z}_t(x))\xi_t(x)\lambda(dx) \geq 0, \quad dt \times dP - a.e.$$

It is also easy to see, by the definition of Doléans-Dade's exponential, that $\exp(L_t^-) = E(M^-)_t$ where, $dM_t^- = CZ_t dW_t + \frac{e^{CZ_t(x)} - 1}{e^{CZ_t(x)}} \tilde{\mu}(dt, dx)$. Therefore, we obtain $K_t^- = \frac{1}{\exp(L_t^-)} = \frac{1}{\exp(V_0) \exp(L_t^-)} E(M^-)_t$.

Using the integration by part formula, we get, $dK_t^- = K_t^- (dM_t^- + dL_t^-)$, which implies, that the predictable quadratic variation of K^- verifies, $d\langle K^- \rangle_t = (K_t^-)^2 d\langle M^- \rangle_t$ and as a consequence, $d\langle M^- \rangle_t = \frac{1}{(K_t^-)^2} d\langle K^- \rangle_t$.

$$\langle M^- \rangle_T \leq \sup_{t \in [0, T]} \frac{1}{(K_t^-)^2} \times \langle K^- \rangle_T. \tag{2.42}$$

Now, we need to have an estimate for $\langle K^- \rangle$ in order to get one for $\langle M^- \rangle$. Itô's formula yields,

$$d(K_t^-)^2 = 2K_t^- dK_t^- + d[K^-]_t = 2(K_t^-)^2 (dM_t^- + dL_t^-) + d[K^-]_t.$$

Taking a sequence of stopping times (T_n) such that for each $n \in \mathbb{N}$, $(\int_0^{T \wedge T_n} 2(K_s^-)^2 dM_s^-)_t$ is a uniformly integrable martingale and integrating the above equation between a stopping time $\sigma \leq T$ and $T \wedge T_n$, we get,

$$[K^-]_{T \wedge T_n} - [K^-]_\sigma = (K_{T \wedge T_n}^-)^2 - (K_\sigma^-)^2 - 2 \int_\sigma^{T \wedge T_n} (K_t^-)^2 (dM_t^- + dL_t^-).$$

Since $\int_0^{T_n} (K_t^-)^2 dL_t^- \geq 0$, by taking conditional expectations on both sides, we obtain,

$$E[\langle K^- \rangle_{T \wedge T_n} - \langle K^- \rangle_\sigma | \mathcal{F}_\sigma] = E[[K^-]_{T \wedge T_n} - [K^-]_\sigma | \mathcal{F}_\sigma] \leq E[(K_{T \wedge T_n}^-)^2 | \mathcal{F}_\sigma] \leq E[\sup_{t \in [0, T]} (K_t^-)^2 | \mathcal{F}_\sigma].$$

Finally, passing to the limit as $n \rightarrow +\infty$ and using the Monotone Convergence theorem, we have,

$$E[\langle K^- \rangle_T - \langle K^- \rangle_0 | \mathcal{F}_0] \leq E[\sup_{t \in [0, T]} (K_t^-)^2 | \mathcal{F}_0].$$

Now, since for every $p \geq 1$, $\sup_{t \in [0, T]} K_t^- \in L^p(\Omega)$, it follows from Garcia and Neveu Lemma (see for example Lemma 4.3 in Barrieu and El Karoui [7] or Neveu [104]) that

$$E[\langle K^- \rangle_T^p] < \infty, \quad \forall p \geq 1. \tag{2.43}$$

With the same arguments used to show that $\sup_{t \in [0, T]} K_t^- \in L^p(\Omega)$, we have also that $\sup_{t \in [0, T]} \frac{1}{K_t^-} \in L^p(\Omega)$ for any $p \geq 1$. From (2.42) and (2.43) together with Cauchy-Schwartz inequality, we deduce that

$$E[\langle M^- \rangle_T^p] < \infty, \quad \forall p \geq 1. \tag{2.44}$$

As for the process Y^+ , it verifies, $dY^+ = dI^+ + dL^+$ where,

$$dI_t^+ = C\delta_t(|V_t| + V_t) + C\alpha(|U_t| - U_t) + CK_2\beta + C\beta r^* + \frac{Z_t}{\beta} \xi_t \frac{\tilde{Z}_t}{\beta} + \frac{|CZ_t|^2}{2} dt + \int f^*(-C\tilde{Z}_t(x))\xi_t(x)\lambda(dx)dt,$$

$$dL^+ = -CZ dW - \int_E |CZ_t|^2 \tilde{\mu}(dt, dx) - \int_E f^*(-C\tilde{Z}_t(x))\xi_t(x)\lambda(dx)dt.$$

As r^* and f^* are positive functions, the process I^+ is increasing and as previously, by easy calculations, we can see that $\exp(L^+) = E(M^+)_t$ where dM^+

$$E[\langle M^- \rangle_T] < \infty, \quad \forall p \geq 1. \tag{2.45}$$

But, expressing the expression of predictable quadratic variation of M^+ and M^- , we get,

$$E[\langle M^- \rangle_T] = \int_0^T \int_E |CZ_t|^2 dt + \int_0^T \int_E (e^{-CZ_t x} - 1) \nu(dt, dx) < \infty, \quad p \geq 1.$$

$\|\cdot\|_p$

Robust Utility Maximization Problem in a Jump Setting

() 2

$$\| \int_T \dots \int_T \int \dots \|_p$$

() 2

This implies from one hand that,

$$E \int_0^T |Z_t|^2 dt \|_p < \infty, \quad p \geq 1,$$

and from the other hand, using the fact that $|y|^2 \leq 2(|e^y - 1|^2 + |e^{-y} - 1|^2)$, $y \in \mathbb{R}$, we get that,

$$E \int_0^T \int_E |\tilde{Z}_t(x)|^2 v(dt, dx) < \infty, \quad p \geq 1.$$

□

In conclusion, we have showed that $Z \in H^{2,p}$ and $\tilde{Z} \in H_\lambda^{2,p}$.

In the next proposition, we establish a comparison theorem for the class of BSDEs in (2.33). For two random variables, we write $A \leq B$ if $A \leq B$ $P - a.s.$ and for two processes X and Y , we write $X \leq Y$ if $\forall t \in [0, T], X_t \leq Y_t$ $P - a.s.$ Finally, we write $(A, X) \leq (B, Y)$ if $A \leq B$ and $X \leq Y$.

Proposition 2.4.4. *Assume that for $k = 1, 2$, (V^k, Z^k, \tilde{Z}^k) is a solution of the BSDE (2.33) in $D^{\text{exp}} \times H^{2,p} \times H^{2,p}$ associated with (U^k, U^-) . If $(U^1, U^-) \leq (U^2, U^-)$, then,*

$$\forall t \in [0, T], V_t^1 \leq V_t^2 \quad P - a.s.$$

Proof. In general, establishing comparison theorems for BSDEs is obtained through an estimate of the quantity $((V^1 - V^2)^+)^2$. Here, in order to take advantage of the convexity of the finite variation part of the BSDE, we will rather estimate $V^1 - \vartheta V^2$ for each $\vartheta \in (0, 1)$. Similar idea was used in Briand and Hu [20] for the continuous case.

Let $\vartheta \in (0, 1)$ and $V^\vartheta = V^1 - \vartheta V^2$. We define accordingly $Z^\vartheta, \tilde{Z}^\vartheta, U^\vartheta$ and \bar{U}_T^ϑ . From (2.33), the dynamics of the process V^ϑ discounted are given by

$$\begin{aligned} dS_t^\delta V_t^\vartheta &= S_t^\delta (-\alpha U_t^\vartheta + \beta r^*(t, \frac{Z_t^1}{\beta}, \xi_t \frac{\tilde{Z}_t^1}{\beta}) - \vartheta r^*(t, \frac{Z_t^2}{\beta}, \xi_t \frac{\tilde{Z}_t^2}{\beta})) dt \\ &\quad - S_t^\delta Z_t^\vartheta dW_t - \int_E S_t^\delta \tilde{Z}_t^\vartheta(x) \tilde{\mu}(dx, dt) \\ &= S_t^\delta (-\alpha U_t^\vartheta + \beta(r^{*,1} - \vartheta r^{*,2})) dt - S_t^\delta Z_t^\vartheta dW_t - \int_E S_t^\delta \tilde{Z}_t^\vartheta(x) \tilde{\mu}(dx, dt), \end{aligned}$$

where, to alleviate the notations, we have denoted, $r^{*,i} = r^*(t, \frac{Z_t^i}{\beta}, \xi_t \frac{\tilde{Z}_t^i}{\beta})$. Now, since r^* is convex, the term $r^{*,1} - \vartheta r^{*,2}$ can be bounded from above. Indeed,

$$\begin{aligned} r^{*,1} &= r^*(t, \frac{Z_t^1}{\beta}, \xi_t \frac{\tilde{Z}_t^1}{\beta}) \leq \vartheta r^*(t, \frac{Z_t^2}{\beta}, \xi_t \frac{\tilde{Z}_t^2}{\beta}) + (1 - \vartheta) r^*(t, \frac{Z_t^\vartheta}{\beta(1 - \vartheta)}, \xi_t \frac{\tilde{Z}_t^\vartheta}{\beta(1 - \vartheta)}) \\ &:= \vartheta r^{*,2} + (1 - \vartheta) r^{*,\vartheta}. \end{aligned} \quad (2.46)$$

Moreover, thanks to (2.16), we have

$$r^{*,\vartheta} \leq K_2 + \frac{|Z_t^\vartheta|^2}{2K_1\beta^2(1 - \vartheta)^2} + K_1 \int_E f^*(\frac{\tilde{Z}_t^\vartheta(x)}{K_1\beta(1 - \vartheta)} \xi_t(x) \lambda(dx).$$

Using this last inequality in (2.46), we get that,

$$\beta(r^{*1} - \vartheta r^{*2}) \leq K_2 \beta(1 - \vartheta) + \frac{|Z_t^\vartheta|^2}{2 K_1 \beta(1 - \vartheta)} + K_1 \beta(1 - \vartheta) \int_E f^* \left(\frac{\tilde{Z}_t^\vartheta(x)}{K_1 \beta(1 - \vartheta)} \right) \xi_t(x) \lambda(dx). \quad (2.47)$$

To get rid of the quadratic and exponential terms in the inequality above, we will use an exponential change of variables. More precisely, let c be a negative constant (to be specified later), and set $P_t = \exp(cS_t^\delta V_t^\vartheta)$, $Q_t = cS_t^\delta P_t - Z_t^\vartheta$ and $\tilde{Q}_t = cS_t^\delta P_t - \tilde{Z}_t^\vartheta$. Using Itô formula, we deduce,

$$\begin{aligned} dP_t &= P_t \left[cd(S_t^\delta V_t^\vartheta) + \frac{c^2}{2} d\langle S^\delta V^\vartheta \rangle_t + \int_E f^*(-cS_t^\delta \tilde{Z}_t^\vartheta(x)) \mu(dx, dt) \right] \\ &= cP_t \left[S_t^\delta (-\alpha U_t^\vartheta + \beta(r^{*1} - \vartheta r^{*2})) dt - S_t^\delta Z_t^\vartheta dW_t - \int_E S_t^\delta \tilde{Z}_t^\vartheta(x) \tilde{\mu}(dx, dt) \right. \\ &\quad \left. + \frac{c}{2} |S_t^\delta Z_t^\vartheta|^2 dt + \frac{1}{c} \int_E f^*(-cS_t^\delta \tilde{Z}_t^\vartheta(x)) \mu(dx, dt) \right] \\ &= cS_t^\delta P_t \left[-\alpha U_t^\vartheta + \beta(r^{*1} - \vartheta r^{*2}) + \frac{c}{2} S_t^\delta |Z_t^\vartheta|^2 + \frac{1}{cS_t^\delta} \int_E f^*(-cS_t^\delta \tilde{Z}_t^\vartheta(x)) \xi_t(x) \lambda(dx) \right] dt \\ &\quad - Q_t dW_t - \int_E \tilde{Q}_t(x) \tilde{\mu}(dx, dt) + P_t \int_E f^*(-cS_t^\delta \tilde{Z}_t^\vartheta(x)) \tilde{\mu}(dx, dt) \\ &:= G_t dt - Q_t dW_t - \int_E \tilde{Q}_t(x) \tilde{\mu}(dx, dt) + P_t \int_E f^*(-cS_t^\delta \tilde{Z}_t^\vartheta(x)) \tilde{\mu}(dx, dt). \end{aligned}$$

Thanks to equation (2.47), the G_t term is bounded from above,

$$\begin{aligned} G_t &\leq cS_t^\delta P_t \left[-\alpha U_t^\vartheta + K_2 \beta(1 - \vartheta) + \frac{|Z_t^\vartheta|^2}{2 K_1 \beta(1 - \vartheta)} + cS_t^\delta \int_E f^* \left(\frac{\tilde{Z}_t^\vartheta(x)}{K_1 \beta(1 - \vartheta)} \right) \xi_t(x) \lambda(dx) \right] \\ &:= cS_t^\delta P_t \left[-\alpha U_t^\vartheta + K_2 \beta(1 - \vartheta) + \frac{|Z_t^\vartheta|^2}{2 K_1 \beta(1 - \vartheta)} + cS_t^\delta \int_E h(K_1 \beta(1 - \vartheta), \tilde{Z}_t^\vartheta(x)) \xi_t(x) \lambda(dx) \right], \end{aligned} \quad (2.48)$$

where $h : \mathbb{R} \times \mathbb{R}^d \rightarrow \mathbb{R}$ defined as $h(x, z) := xf^*(z/x) = xe^{z/x} - x - z$. We need to choose c such that the term next to $|Z_t^\vartheta|^2$ is negative, that choose c such that,

$$\frac{1}{K_1 \beta(1 - \vartheta)} \leq -cS_t^\delta.$$

Since $S_t^\delta \geq e^{|\delta| \| \infty^T$, it is sufficient to set $c(\vartheta) := -\frac{e^{-|\delta| \| \infty^T}{K \beta(1 - \vartheta)}$. Computing the derivative of h with respect to x , we get, $\partial_x h(x, z) = e^{z/x} - (z/x)e^{z/x} - 1$. Studying the sign of the function $x \rightarrow e^x - xe^x - 1$ by calculating its derivative, we obtain that $e^x - xe^x - 1 \leq 0, \forall x \in \mathbb{R}$. Therefore,

we deduce that $\partial_x h(x, \cdot) \leq 0, \forall x \in \mathbb{R}$, that h is decreasing. Hence, going back to (2.48), we get that,

$$G_t \leq c(\vartheta) S_t^\delta P_t (-\alpha U_t^\vartheta + K_2 \beta (1 - \vartheta)) \leq S_t^\delta P_t e^{-\|\delta\|_\infty T} \left(\frac{\alpha}{K_1 \beta} U_t^1 - \frac{K_2}{K_1} \right), \quad (2.49)$$

where we have used, in the second inequality, the fact that,

$$U_t^\vartheta = U_t^1 - \vartheta U_t^2 = \vartheta(U_t^1 - U_t^2) + (1 - \vartheta)U_t^1 \leq (1 - \vartheta)U_t^1.$$

Finally, denoting $D_t = \exp \left(-\int_0^t S_s^\delta \frac{\alpha}{K_1 \beta} U_s^1 - \frac{K_2}{K_1} ds \right)$, and introducing $P_t^D := D_t P_t$, $Q_t^D := Q_t D_t$ and $\tilde{Q}_t^D := \tilde{Q}_t D_t$. Using again Itô's formula, for any stopping time $0 \leq t \leq \tau \leq T$,

$$P_t \geq P_\tau + \int_t^\tau Q_s dW_s + \int_t^\tau Q_s(x) \mu(dx, ds) - \int_t^\tau \int E P_{s-} D_s f(-c S_s Z_s(x)) \mu(dx, ds).$$

Considering a localizing sequence of stopping time τ_n , such that the local martingales, in the above inequality, stopped in τ_n are martingales, we obtain,

$$P_t \geq E \left[P_{\tau_n} \exp \left(-\int_t^{\tau_n} S_s \left(\frac{\alpha}{K_1 \beta} U_s^1 - \frac{K_2}{K_1} \right) ds \right) \mid \mathcal{F}_t \right].$$

In view of the integrability assumptions on U^1 and on V , by the dominated convergence theorem, we can deduce that,

$$P_t \geq E \left[P_T \exp \left(-\int_t^T S_s \left(\frac{\alpha}{K_1 \beta} U_s^1 - \frac{K_2}{K_1} \right) ds \right) \mid \mathcal{F}_t \right].$$

But by definition of P , $P_T = \exp(c(\vartheta) S_T^\delta V_T^\vartheta) = \exp(c(\vartheta) S_T^\delta (U_T^1 - \vartheta U_T^2))$, and because $U_T^1 \leq U_T^2$ and $c(\vartheta)$ is negative, we get,

$$c(\vartheta) S_T^\delta V_T \geq - \frac{e^{-\|\delta\|_\infty T}}{K_1 \beta} S_T U_T.$$

Therefore, we have,

$$P_t \geq E \left[\exp \left(-\int_t^T S_s \left(\frac{\alpha}{K_1 \beta} U_s^1 - \frac{K_2}{K_1} \right) ds \right) \mid \mathcal{F}_t \right],$$

which implies that,

$$V_t \leq - \frac{K_1 \beta (1 - \vartheta) e^{\|\delta\|_\infty T}}{S_t^\vartheta} \ln E \left[\exp \left(-\int_t^T S_s \left(\frac{\alpha}{K_1 \beta} U_s^1 - \frac{K_2}{K_1} \right) ds \right) \mid \mathcal{F}_t \right].$$

Taking the limit when $\vartheta \nearrow 1$, we finally get,

$$V_t^1 \leq V_t^2.$$

□

The following corollary is a direct consequence of the comparison result above.

Corollary 2.4.5. *Under assumptions (ffi_r) and (ffi_w), the BSDE (2.33) has a unique solution $(V, Z, \tilde{Z}) \in D_0^{\text{exp}} \times H^{2,p} \times H_\lambda^{2,p}$.*

2.5 Appendix

Lemma 2.5.1. *Let $r(t, \eta, \psi) = \frac{1}{2}|\eta|^2 + \int_E f(\psi(x))\xi_t(x)\lambda(dx)$ and $Q \in \mathcal{Q}_f$. Then, the following processes,*

$$M_t = \int_0^t \eta_s dW_s^Q, \quad M_t = \int_0^t \int_E \log(1 + \psi(x)) \mu^Q(ds, dx),$$

are Q -martingales.

Proof. Since $Q \in \mathcal{Q}_f$, we have,

$$E_Q \int_0^T r(t, \eta_t, \psi_t) dt = E_Q \int_0^T \left[\frac{1}{2} |\eta_t|^2 + \int_E f(\psi_t(x)) \xi_t(x) \lambda(dx) \right] dt < \infty. \quad (2.50)$$

In particular, $E_Q \int_0^T \frac{1}{2} |\eta_t|^2 dt < \infty$, which implies that M is Q -martingale. Now, we prove that M is also a Q -martingale. First, note that,

$$f(x) = (1+x) \log(1+x) - x \geq \frac{1}{6} (1+x) \log^2(1+x) \geq 0, \quad \text{for } -1 \leq x \leq e^2 - 1.$$

Hence, as the RHS of (2.50) is finite, we get that

$$\log(1 + \psi_s(x)) \mathbb{1}_{\psi_s(x) \leq e^2 - 1} \in L^2(dQ \times \nu^Q(ds, dx)). \quad (2.51)$$

Moreover, for $x > e^2 - 1$, we have,

$$(1+x) \log(1+x) \leq 2((1+x) \log(1+x) - x).$$

Again, as the RHS of (2.50) is finite, we get that,

$$\log(1 + \psi_s(x)) \mathbb{1}_{\psi_s(x) > e^2 - 1} \in L^1(dQ \times \nu^Q(ds, dx)). \quad (2.52)$$

From (2.51) and (2.52), and using Theorem 1.8(i) in Jacod and Shiryaev [70], we obtain that M is Q -martingale. \square

Proposition 2.5.2. Let $r(t, \eta, \psi) = \frac{1}{2}|\eta|^2 + \int_E f(\psi(x))\xi_t(x)\lambda(dx)$ and $Q \in \mathcal{Q}_f$. Then, we have,

$$H(Q|P) = E_Q \int_0^T r(t, \eta_t, \psi_t) dt = E_Q \int_0^T \left[\frac{1}{2}|\eta_t|^2 + \int_E f(\psi_t(x))\xi_t(x)\lambda(dx) \right] dt. \quad (2.53)$$

Proof. Let $Q \in \mathcal{Q}_f$ with corresponding (η, ψ) . We introduce the following sequence of processes $(\psi_m)_{m \in \mathbb{N}^*}$ defined as:

$$\psi_{m,s}(x) = \psi_s(x) \mathbb{1}_{\psi_s(x) \leq m} \mathbb{1}_{|x| \geq 1/m}.$$

It is clear that $\psi_m \in L^2(dQ \times \nu^Q(ds, dx))$. Developing the logarithm of Radon-Nikodym derivative of Q w.r.t P gives Q -a.s.:

$$\begin{aligned} \log \frac{dQ}{dP} &= \lim_{m \rightarrow \infty} \log E \left[\exp \left(\int_0^T \eta_s dW_s + \int_0^T \int_E \psi_{m,s}(x) \mu(ds, dx) \right) \right] \\ &= \lim_{m \rightarrow \infty} \left(\int_0^T \eta_s dW_s - \frac{1}{2} \int_0^T |\eta_s|^2 ds + \int_0^T \int_E \psi_{m,s}(x) \mu(ds, dx) \right. \\ &\quad \left. + \int_0^T \int_E (\log(1 + \psi_{m,s}(x)) - \psi_{m,s}(x)) \mu(ds, dx) \right) \\ &= \lim_{m \rightarrow \infty} \left(\int_0^T \eta_s dW_s + \frac{1}{2} \int_0^T |\eta_s|^2 ds + \int_0^T \int_E \psi_{m,s}(x) \mu(ds, dx) \right. \\ &\quad \left. + \int_0^T \int_E (\log(1 + \psi_{m,s}(x)) - \psi_{m,s}(x)) \mu(ds, dx) \right) \\ &= \lim_{m \rightarrow \infty} \left(\int_0^T \eta_s dW_s + \frac{1}{2} \int_0^T |\eta_s|^2 ds + \int_0^T \int_E \psi_{m,s}(x) \mu(ds, dx) \right. \\ &\quad \left. + \int_0^T \int_E (\log(1 + \psi_{m,s}(x)) - \psi_{m,s}(x)) \mu(ds, dx) \right) \\ &= \lim_{m \rightarrow \infty} \left(\int_0^T \eta_s dW_s + \frac{1}{2} \int_0^T |\eta_s|^2 ds + \int_0^T \int_E \psi_{m,s}(x) \mu(ds, dx) \right. \\ &\quad \left. + \int_0^T \int_E (\log(1 + \psi_{m,s}(x)) - \psi_{m,s}(x)) \mu(ds, dx) \right) \end{aligned} \quad (2.54)$$

where we used from the second to the third inequality that, by the definition of the process ψ_m , $(1 + \psi_s)(\log(1 + \psi_{m,s}) - \psi_{m,s}) \in L^1(\nu(ds, dx))$ and $\psi_m \psi \in L^1(\nu(ds, dx))$. In particular, $\int_0^t \int_E (\log(1 + \psi_{m,s}(x)) - \psi_{m,s}(x)) \mu(ds, dx)$ is a well defined. Lemma 2.5.1 above insures that the following processes:

$$M_t := \int_0^t \eta_s dW_s + \frac{1}{2} \int_0^t |\eta_s|^2 ds + \int_0^t \int_E \psi_{m,s}(x) \mu(ds, dx), \quad M^{m,t} := \int_0^t \int_E \log(1 + \psi_{m,s}(x)) \mu(ds, dx),$$

are Q -martingales. Moreover, we also have that $M_{m,T}$ converges to M_T in $L^1(Q)$. Indeed,

decomposing M and M_m as in Lemma 2.5.1 in the following way,

$$M_{m,t} = \hat{M}_{m,t} + \tilde{M}_{m,t}$$

where,

$$\hat{M}_{m,t} := \int_0^t \int_E \log(1 + \psi_{s,m}(x)) \mathbb{1}_{\psi_{s,m}(x) \leq e^{-1}} \nu(ds, dx) \quad \tilde{M}_{m,t} := \int_0^t \int_E \log(1 + \psi_{s,m}(x)) \mathbb{1}_{\psi_{s,m}(x) > e^{-1}} \nu(ds, dx)$$

We have the positive (resp. negative) part of $\log(1 + \psi_{m,s}(x)) \mathbb{1}_{\psi_{s,m}(x) \leq e^{-1}}$ increases (resp. decreases) to the positive (resp. negative) part of $\log(1 + \psi_s(x)) \mathbb{1}_{\psi_s(x) \leq e^{-1}}$ as m goes to infinity

and by the monotone convergence we deduce that $M_{m,T}$ converges to \hat{M}_T in $L^1(Q)$. Using the same argument, we have also $\tilde{M}_{m,T}$ converges to \tilde{M}_T in $L^1(Q)$ as m goes to infinity. Hence, we obtain that $M_{m,T}$ converges to M_T as $m \rightarrow \infty$. By passing to a subsequence, we may assume that $M_{m,T}$ converges to M_T Q -a.s. Finally, by Monotone convergence theorem, the last term in (2.54) also converges to $\int_0^T \int_E f(\psi_s(x)) \nu(ds, dx)$. Consequently, (2.54) becomes,

$$\log \frac{dQ}{dP} = \int_0^T \int_E \eta_s dW_s^Q + \frac{1}{2} \int_0^T \int_E |\eta_s|^2 ds + \int_0^T \int_E \log(1 + \psi_s(x)) \nu(ds, dx) + \int_0^T \int_E f(\psi_s(x)) \nu(ds, dx).$$

Taking the expectation under Q and using the fact that M and M_m are martingales yields (2.53). □

ESTIMATION OF SYSTEMIC SHORTFALL RISK MEASURES USING STOCHASTIC ALGORITHMS

Abstract

Systemic risk measures were introduced to capture the global risk and the corresponding contagion effects that is generated by an interconnected system of financial institutions. To this purpose, two approaches were suggested. In the first one, systemic risk measures can be interpreted as the minimal amount of cash needed to secure a system *after aggregating* individual risks. In the second approach, systemic risk measures can be interpreted as the minimal amount of cash that secures a system by allocating capital to each single institution *before aggregating* individual risks. Although the theory behind these risk measures has been well investigated by several authors, the numerical part has been neglected so far. In this paper, we use stochastic algorithms schemes in estimating MSRM and prove that the resulting estimators are consistent and asymptotically normal. We also test numerically the performance of these algorithms on several examples.

3.1 Introduction

The axiomatic theory of risk measures, first initiated by the seminal paper of Artzner et al. [5], has been widely studied during the last years. Value-at-Risk (VaR) is one of the most known and common risk measures used by practitioners and regulation authorities. However, VaR lacks one important property: it does not take into account the diversification effect. To circumvent this problem, the VaR was replaced by the Conditional Value-at-Risk (CVaR) and a more general framework of improved risk measures has been introduced: *Utility-based Shortfall Risk* (SR). Nevertheless, when it comes to a system of financial institutions or portfolios, the question about how to assess the global risk as well as individual risks arise. Following the 2008 crisis, the traditional approach of measuring systemic risk that consists in considering each institution

as a single entity isolated from other institutions, has shown its own limits. Indeed, with this approach, the risk associated to a vector of positions $X = (X_1, \dots, X_d)$ can be written as:

$$R(X) := \sum_{i=1}^d \eta_i(X_i),$$

where each η_i is a univariate risk measure. Then, Chen et al. [27] proposed an approach that is very close in spirit to the axiomatic framework initiated by Artzner et al. [5]. They showed that any systemic risk measure verifying certain axioms is the composition of a univariate risk measure η and an aggregation Λ , i.e.,

$$R(X) = \eta(\Lambda(X)).$$

The previous representation is known as the “Aggregate then Add Cash” approach as it consists first in aggregating the positions X_1, \dots, X_d through the aggregation function Λ and then to apply a univariate risk measure. One of the most common ways to aggregate the outcomes X_i is to simply take the sum, that is to consider, $\Lambda(x) = \sum_{i=1}^d x_i$. It is worth noticing that, while summing up profit and losses might seem reasonable from the point of view of a portfolio manager because portfolios profits and losses compensate each other, this aggregation rule seems inadequate from the point of view of a regulator where cross-subsidization between institutions is not realistic since no institution will be willing to pay for the losses of another one. Motivated by these considerations, Biagini et al. [15] proposed another approach to measure the systemic risk. They first considered the systemic risk as the minimal capital that secures the system by injecting capital into the single institutions, *before aggregating the individual risks*:

$$R(X) := \inf_{i=1}^{\sum d} \{ m_i, \Lambda(X + m) \in \mathbf{A} \}, \quad (3.1)$$

where \mathbf{A} is an acceptance set. This approach, known as “Add Cash then Aggregate” consists in adding the amount m_i to the financial position X_i before the corresponding total loss $\Lambda(X + m)$ is computed. The systemic risk is then measured as the minimal total amount $\sum_{i=1}^d m_i$ injected into the institutions to make it acceptable. With this approach, a joint measure of total risk as well as individuals risk contributions to systemic risk is obtained. If $m^* = (m_1^*, \dots, m_d^*)$ is an opti-

mum, that is $R(X) = \sum_{i=1}^d m_i^*$ and $\Lambda(X + m^*) \in \mathbf{A}$, one could order the m_i^* s and hence be able to say that institution i requires more cash allocation or is riskier than institution j if $m_i^* \geq m_j^*$.

In this article, we are interested in the numerical approximation of the multivariate shortfall risk measure (MSRM) that was introduced in Armenti et al. [4]. They are an extension of univariate SR and can be obtained by taking the aggregation function $\Lambda(x) = I_S(-x)$ where I_S is a multivariate loss function (see Section 3.2) and the acceptance set $\mathbf{A} = \{X \in L^0(\mathbb{R}), E[X] \leq 0\}$.

To meet the regulatory requirements, financial institutions need to develop a reliable risk management framework to face all kind of financial risks associated to their portfolios. Most of the time, financial institutions use the standard VaR and CVaR although it suffers from some deficiencies. The most common method used to compute VaR is the inversion of the simulated empirical P&L distribution function using Monte Carlo or historical simulation tools (see Glasserman [58] and Glasserman et al. [59]). Another idea to compute VaR and CVaR comes from the fact that they are solutions and the value of the same convex optimization problem as pointed out in Rockafellar and Uryasev [113]. Moreover, as they can be expressed as an expectation, this led Bardou et al. [6] to define consistent and asymptotically normal estimators of both quantities using a classical Robbins-Monro (RM) procedure. Since VaR and CVaR are both related to the simulation of rare events, they also introduced a recursive and adaptive variance method based on importance sampling paradigm.

RM algorithms have been the subject of an enormous literature, both theoretical and applied. The basic paradigm in its simplest form is the following stochastic difference equation: $Z_{n+1} = Z_n + \gamma_n Y_n$, where Z_n takes its values in some Euclidean space, Y_n is a noisy observable variable, and $\gamma_n > 0$ is the step size that goes to zero as $n \rightarrow \infty$. The original work was motivated by the classic problem of finding a root of a continuous function $\theta \rightarrow g(\theta)$, which is unknown but such that, we are able to take only “noisy” measurements at any desired value θ . This is the case when the function g can be expressed as an expectation, that is $g(\theta) = E[G(X, \theta)]$, where X is some random variable. In such situation, the noisy observation variable is simply $Y_n = G(X_{n+1}, Z_n)$, where (X_n) is a sequence of i.i.d random variables with the same law as X . If moreover, the random variable X is not directly simulatable, but can only be approximated by another easily simulatable random variable, Frikha [52] recently extended the scope of multi-level Monte Carlo to the framework of stochastic algorithms and proved central limit theorems. In many cases, the analysis of these algorithms uses the so-called ODE (Ordinary Differential Equation) method introduced by Ljung [90]. The main idea is to show that, in the long run, the noise is eliminated so that, asymptotically, the behaviour of the algorithm is determined by that of the “mean” ODE: $\dot{\theta} = g(\theta)$. An introductory approach to RM algorithms and their convergence rate can be found in Duflo [42] and Benveniste et al. [12]. To ensure the convergence of RM algorithms to the root of the function g , it does not require too restrictive assumptions except for one: the sub-linear growth of the function g . One way to deal with restrictive assumption is to use projection techniques. This consists in using the projection into a compact K each time the sequence Z_n goes out of K . This procedure was first introduced by Kushner and Sanvicente [81] in order to deal with problems of convex optimization with constraints. Another way to deal with this constraint in the framework of variance reduction using importance sampling method was proposed in Lemaire and Pagès [87]. They have showed that under some regularity assumption on the density of the law of X , we can obtain almost-surely convergence result

and central limit theorems. In this paper, for the sake of simplicity, we will rather use projection techniques. An excellent survey on projection techniques, their links with ordinary differential equation (ODE) and stochastic algorithms can be found in Kushner and Yin [83].

SR can be characterized as the unique root of a function $g : \mathbb{R} \rightarrow \mathbb{R}$ that is expressed as an expectation. Therefore, a straightforward approach for estimating SR consists in, first, using a deterministic root finding algorithm that would converge to the root, and second, designing an efficient Monte Carlo procedure that estimates $g(s)$ at each given argument $s \in \mathbb{R}$. One could also use variance reduction techniques in order to accelerate the estimation of the function g at each argument $s \in \mathbb{R}$. This idea is very close to sample average methods in stochastic programming. For more details, see, for example, Kleywegt et al. [80], Linderoth et al. [89], Mak et al. [92], Shapiro and Nemirovski [123], Verweij et al. [129] and Verweij et al. [130]. An alternative to this combination of Monte Carlo method and deterministic root finding schemes is to use stochastic algorithm as presented in Dunkel and Weber [43]. In their work, they did not assume the sub-linear growth of the function g , and therefore used projection techniques to prevent the algorithm from explosion.

In this paper, we will see that the optimal allocations of multivariate shortfall risk measures can also be characterized as the root of a function that is expressed as an expectation. More precisely, the optimal allocations are characterized as the solution of the first order condition of the Lagrangian associated to the multivariate risk measure. Again, because we do not want to reduce drastically the scope of application, we will use stochastic algorithms with projection to approximate the optimal allocations.

The paper is organized as follows. The next section, is dedicated to MSRM and the definitions related to them. The main theorem that characterizes the optimal allocations for MSRM is presented. In Section 3.3, we explain the ODE method and recall some stability results that we will use later to establish convergence results. Finally, section 3.4 is devoted to some numerical experiments of our procedures. We present a first testing example with an exponential loss function, where we have a closed formula for optimal allocations. We also give a second example using a loss function with a mixture of positive part and quadratic functions.

3.2 About Multivariate Risk Measures

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, and denote by $L^0(\mathbb{R}^d)$ the space of \mathcal{F} -measurable d -variate random variables on this space with $d \geq 2$. For $x, y \in \mathbb{R}^d$, we say that that $x \geq y$ ($x > y$ resp.) if $x_k \geq y_k$ ($x_k > y_k$ resp.) for every $1 \leq k \leq d$. We denote by $\|\cdot\|$ the Euclidean norm, and $x \cdot y = \sum x_k y_k$. For a function $f : \mathbb{R}^d \rightarrow [-\infty, \infty]$, we denote by $f^*(y) = \sup_x \{x \cdot y - f(x)\}$ the convex conjugate of f . The space $L^0(\mathbb{R}^d)$ inherits the lattice structure of \mathbb{R}^d and therefore, we can use the classical notations in \mathbb{R}^d in a \mathbb{P} -almost-sure sens. We say, for example, for

$X, Y \in L^0(\mathbb{R}^d)$, that $X \geq Y$ (or $X > Y$ resp.) if $P(X \geq Y) = 1$ (or $P(X > Y) = 1$ resp.). To simplify the notation, we will simply write L^0 instead of $L^0(\mathbb{R}^d)$. Now, let $X = (X_1, \dots, X_d) \in L^0$ be a random vector of financial losses, i.e., negative values of X_k represents actually profits. We want to assess the systemic risk of the whole system and to determine a monetary risk measure, which will be denoted $R(X)$, as well as a risk allocation $RA_k(X)$, $k = 1, \dots, d$ among the d risk components. Inspired by the univariate case introduced in Föllmer and Schied [51], Armenti et al. [4] introduced a multivariate extension of shortfall risk measures by the means of loss functions and sets of acceptable monetary risk allocations.

Definition 3.2.1. A function $l : \mathbb{R}^d \rightarrow (-\infty, \infty]$ is called a loss function if:

- (A1) l is increasing, that is $l(x) \geq l(y)$ if $x \geq y$;
- (A2) l is convex and lower-semicontinuous with $\inf l < 0$;
- (A3) $l(x) \geq \sum x_k - c$ for some constant c .

Furthermore, a loss function l is said to be permutation invariant if $l(x) = l(\pi(x))$ for every permutation π of its components.

Comment: The property (A1) expresses the normative fact about the risk, that is, the more losses we have, the riskier is our system. As for (A2), it expresses the desired property of diversification. Finally, (A3) says that the loss function put more weight on high losses than a risk neutral evaluation.

Example 3.2.1. Let $h : \mathbb{R} \rightarrow \mathbb{R}$ be one dimensional loss function satisfying condition (A1), (A2) and (A3). We could build a multivariate loss function using this one dimensional loss function in the following way:

- (C1) $l(x) = h(\sum x_k)$;
- (C2) $l(x) = \sum h(x_k)$;
- (C3) $l(x) = \alpha h(\sum x_k) + (1 - \alpha) \sum h(x_k)$ for $0 \leq \alpha \leq 1$.

More specifically, in (C1), we are aggregating losses before evaluating the risk, whereas in (C2), we evaluate individual risks before aggregating. The loss function in (C3) is a convex combination of those in (C1) and (C2).

One of the main examples we will be studying in this paper are the two following ones:

$$h(x) = \frac{1}{\alpha + 1} \left(\sum_i e^{\beta x_i} + \alpha e^{\beta \sum x_i} \right) - \frac{\alpha + d}{\alpha + 1}, \quad h(x) = \sum_i x_i + \frac{1}{2} \sum_i (x_i^+)^2 + \alpha \sum_{i < j} x_i^+ x_j^+,$$

where the coefficient $\alpha > 0$ is called the systemic weight and $\beta > 0$ is a risk aversion coefficient. In the following, we will consider multivariate risk measures defined on Orlicz spaces (see Rao and Ren [110] for further details on the theory of Orlicz spaces). This has several advantages. From a mathematical point of view, it is a more general setting than L^∞ , and in the same time, it simplifies the analysis especially for utility maximization problems. Therefore, we will consider loss vectors in the following multivariate Orlicz heart:

$$M^\vartheta = \{X \in L^0 : E[\vartheta(\lambda X)] < \infty, \forall \lambda > 0\},$$

where $\vartheta(x) = I(|x|)$, $x \in \mathbb{R}^d$. See Appendix in Armenti et al. [4] for more details about Orlicz spaces.

Next, we give the definition of multivariate shortfall risk measures as it was introduced in Armenti et al. [4].

Definition 3.2.2. Let I be a multivariate loss function and $X \in M^\vartheta$, we define the acceptance set $\text{ffi}(X)$ by:

$$\text{ffi}(X) := \{m \in \mathbb{R}^d : E[I(X - m)] \leq 0\}.$$

The multivariate shortfall risk of $X \in M^\vartheta$ is defined as:

$$R(X) := \inf \left\{ \sum m_k : m \in \text{ffi}(X) \right\} = \inf \left\{ \sum m_k : E[I(X - m)] \leq 0 \right\}. \quad (3.2)$$

Remark 3.2.3. When $d = 1$, the above definition corresponds exactly to the univariate shortfall risk measure in Föllmer and Schied [57].

The following theorem from Armenti et al. [4] shows that the multivariate shortfall risk measure has the desired properties and admits a dual representation as in the case of univariate shortfall risk measure. We introduce Q^{ϑ^*} the set of measure densities in L^{ϑ^*} , the dual space of M^ϑ :

$$Q^{\vartheta^*} := \frac{dQ}{dP} := (Z_1, \dots, Z_d), \quad Z \in L^{\vartheta^*}, \quad Z_k \geq 0 \text{ and } E[Z_k] = 1 \text{ for every } k.$$

Theorem 3.2.4. [Theorem 2.10 in Armenti et al. [4]] The function

$$R(X) := \inf \left\{ \sum m_k : m \in \text{ffi}(X) \right\},$$

is real-valued, convex, monotone and translation invariant. Moreover, it admits the dual representation:

$$R(X) = \max_{Q \in Q^{\vartheta^*}} \{E_Q[X] - \alpha(Q)\}, \quad X \in M^\vartheta,$$

where the penalty function is given by

$$\alpha(Q) = \inf_{\lambda > 0} E \lambda I^* \frac{dQ}{\lambda dP}, \quad Q \in Q^{\vartheta*}.$$

Now, we address the question of existence and uniqueness of a risk allocation which are not straightforward in the multivariate case. Armenti et al. [4] showed that if the loss function is permutation invariant, then risk allocations exist and they are characterized by Kuhn-Tucker conditions. We denote by $Z = \{m \in \mathbb{R}^d, \sum \bar{m}_i = 0\}$ the *zero-sum allocations set*.

Definition 3.2.5. A risk allocation is an acceptable monetary risk allocation $m \in \text{ffi}(X)$ such that $R(X) = \sum \bar{m}_k$. When a risk allocation is uniquely determined, we denote it by $RA(X)$.

We make the following assumption on the loss function I and the vector of losses $X \in M^{\vartheta}$:

- (ffi_i) i. For every m_0 , $m \mapsto I(X - m)$ is differentiable at m_0 a.s.;
 ii. I is permutation invariant.

Theorem 3.2.6. [Theorem 3.4 in Armenti et al. [4]] Let I be a loss function and $X \in M^{\vartheta}$ such that assumption (ffi_i) holds. Then, risk allocations $m^* \in \mathbb{R}^d$ exists and they are characterized by the first order conditions:

$$1 = \lambda^* E[OI(X - m^*)], \quad E[I(X - m^*)] = 0,$$

where $\lambda^* \geq 0$ is a Lagrange multiplier. If moreover $I(x + \cdot)$ is strictly convex along zero sum allocations for every x such that $I(x) \geq 0$, the risk allocation is unique.

Comment: Let $f_0(m) = \sum_{i=1}^d m_i$ and $f_1(m) := E[I(X - m)]$, for $m \in \mathbb{R}^d$ and $X \in M^{\vartheta}$. The assumption (ffi_i)-i. together with the convexity of the function $m \mapsto I(X - m)$, we have that, by Theorem 7.46 in Shapiro et al. [124], f_1 is differentiable at every $m \in \mathbb{R}^d$ and that,

$$Of_1(m) = -E[OI(X - m)], \quad m \in \mathbb{R}^d.$$

Therefore, the first order conditions given in the above theorem are equivalent to :

$$\begin{aligned} Of_0(m^*) + \lambda^* Of_1(m^*) &= 0, \\ \lambda^* f_1(m^*) &= 0. \end{aligned}$$

Furthermore, we also know, thanks to Theorem 28.3 in Rockafellar [115], that the above conditions are equivalent to saying that (m^*, λ^*) is a saddle point of the Lagrangian associated to the problem in (3.2), i.e.,

$$L(m, \lambda) := f_0(m) + \lambda f_1(m) = \sum_{i=1}^d m_i + \lambda E[I(X - m)]. \quad (3.3)$$

Under the assumptions of the above theorem $m^* := (m^*, \lambda^*)$ is the unique solution of $h(m, \lambda) = 0$, where:

$$h(m, \lambda) := \frac{\lambda E[O(X - m)] - 1}{E[I(X - m)]}, \quad (m, \lambda) \in \mathbb{R}^d \times [0, \infty[.$$

Thus, in order to find the unique risk allocation m^* , we can look for the zeros of the function h . We suggest here to use stochastic algorithms as they present the advantage of being incremental, less sensitive to dimension, and offer a flexible framework that can be conveniently combined with features such as importance sampling (see Dunkel and Weber [43]) and model uncertainty.

3.3 Multivariate Systemic Risk Measures and Stochastic Algorithms

Let I be a loss function satisfying assumption (ffi) and a vector of losses $X \in M^d$. We recall that in order to have the uniqueness of risk allocations, we need to add the convexity condition:

- (ffi₁) i. For every $m_0, m \mapsto I(X - m)$ is differentiable at m_0 a.s.;
- ii. I is permutation invariant;
- iii. $m \mapsto E[I(X - m)]$ is strictly convex.

Under (ffi₁), Theorem 3.2.6 ensures that there exists a unique risk allocation m^* such that $m^* = (m^*, \lambda^*)$ is the unique root of the function $h(m, \lambda) := E[H(X, m, \lambda)]$, where we set

$$H(X, m, \lambda) = \frac{\lambda O_m I(X - m) - 1}{I(X - m)}, \quad X \in M^d. \tag{3.4}$$

In all the following, we will work under the assumption (ffi₁). The aim of this section is to construct an algorithm that converges to the root $m^* = (m^*, \lambda^*)$ under some suitable assumptions. As pointed out in the introduction, we will not use a regular Robbins-Monro algorithm as it requires the sublinearity of the function h , and consequently will not offer a general framework that is flexible enough to cover a wide range of loss functions. In order to be able to use the ODE method (see Section 3.5.1 for more details), we suggest instead the projected Robbins-Monro (RM) Algorithm:

$$\begin{aligned} Z_{n+1} &= \Pi_K[Z_n + \gamma_n H(X_{n+1}, Z_n)], \quad Z_0 = z_0 \in K \\ &= \Pi_K[Z_n + \gamma_n h(Z_n) + \gamma_n \delta M_n], \end{aligned} \tag{3.5}$$

where $\delta M_n = H(X_{n+1}, Z_n) - h(Z_n)$. In the sequel, we denote $F_n = \sigma(Z_0, X_i, i \leq n)$. δM_n is a martingale difference sequence with respect to the filtration $F = (F_n)$. We assume that K is hyperrectangle such that m^* is in the interior of K : $K = \{m \in \mathbb{R}^d, a_i \leq m_i \leq b_i\} \times [0, A]$. $(X_n)_{n \geq 1}$

is an i.i.d sequence of random variables with the same distribution as X , independent of Z_0 and $(\gamma_n)_{n \geq 1}$ is a deterministic step sequence decreasing to zero and satisfying:

$$\sum_{n \geq 1} \gamma_n = +\infty \text{ and } \sum_{n \geq 1} \gamma_n^2 < +\infty. \quad (3.6)$$

In the sequel, we will take $\gamma_n = \frac{c}{n^\nu}$ where c is a positive constant and $\nu \in]\frac{1}{2}, 1]$.

3.3.1 Properties of z^*

Before giving the results about the almost surely convergence, let us give some properties of z^* . From paragraph 3.5.1 in Section 3.5, we know that (3.5) is associated with the following ODE:

$$\dot{z} = h(z) + C(z), \quad C(z) \in -C(z), \quad (3.7)$$

where $C(z)$ is the convex cone determined by the outer normals to the faces that need to be truncated at z and $C(z)$ is the minimum force needed to bring back z to K (For more details about concepts related to the ODE method and stability results, see Section 3.5). Now, since z^* is interior to K and $h(z^*) = 0$, z^* is an equilibrium point for the projected ODE 3.7. In order to study the asymptotic stability of the equilibrium z^* , one needs to find some convenient Lyapunov function V . A natural and classical choice for this type of problems is $V(z) = \|z - z^*\|^2$. It is obvious that V is positive definite. The following proposition shows that its derivative along any state trajectory is negative semi-definite on K .

Proposition 3.3.1. *The function $V(z) = \|z - z^*\|^2$ is such that $\dot{V}(z) = \langle \nabla V(z), h(z) + C(z) \rangle$ is negative semi-definite on K with the respect to the ODE in (3.7).*

Proof. First, let $z = (m, \lambda) \in \text{int}(K)$ so that $\dot{V}(z) = \langle \nabla V(z), h(z) \rangle = 2\langle z - z^*, h(z) \rangle$, and define L the Lagrangian as defined in (3.3). We have:

$$\begin{aligned} \langle z - z^*, h(z) \rangle &= \langle m - m^*, \lambda E[OI(X - m)] - \mathbf{1} \rangle + (\lambda - \lambda^*) E[I(X - m)] \\ &= -\langle m - m^*, \partial_m L(m, \lambda) \rangle + (\lambda - \lambda^*) \partial_\lambda L(m, \lambda). \end{aligned}$$

Now, thanks to the convexity of L with respect to m , we have: $L(m^*, \lambda) \geq L(m, \lambda) + \langle m^* - m, \partial_m L(m, \lambda) \rangle$. This in turn implies that

$$L(m^*, \lambda) - L(m, \lambda) + (\lambda - \lambda^*) \partial_\lambda L(m, \lambda) \geq \langle m^* - m, \partial_m L(m, \lambda) \rangle + (\lambda - \lambda^*) \partial_\lambda L(m, \lambda).$$

But, we also have,

$$\begin{aligned} L(m, \lambda) &= \sum m_i + \lambda E[J(\mathbf{X} - m)] = \sum m_i + \lambda^* E[J(\mathbf{X} - m)] + (\lambda - \lambda^*) E[J(\mathbf{X} - m)] \\ &= L(m, \lambda^*) + (\lambda - \lambda^*) \partial_\lambda L(m, \lambda). \end{aligned}$$

The previous inequality becomes then

$$\langle m - m^*, \partial_m L(m, \lambda) \rangle - (\lambda - \lambda^*) \partial_\lambda L(m, \lambda) \geq L(m, \lambda^*) - L(m^*, \lambda)$$

The RHS of the last inequality is non-negative, because, (m^*, λ^*) is a saddle point, that is $L(m^*, \lambda) \leq L(m^*, \lambda^*) \leq L(m, \lambda^*)$. Moreover, because L is strictly convex with respect to m , it is also negative if $m \neq m^*$. Therefore, we get that,

$$\langle m - m^*, h(m) \rangle \leq 0. \quad (3.8)$$

Note that this is true irrespective of whether $m \in \text{int}(K)$ or not. Now, if $m_i = b_i$ and $h_i(m) > 0$ for some i , then $C_i = -h_i(m)$, and hence $\langle m - m^*, C \rangle \leq 0$. This shows that in this case, $\dot{V}(m)$ is less than the LHS of 3.8 and it is in turn negative. This can be easily generalized for all other boundary cases. As a conclusion, we have shown that \dot{V} is negative semi-definite on K . \square

Remark 3.3.2. We cannot conclude that \dot{V} is negative definite on K , because $m \neq m^*$ does not imply that $m = m^*$. Besides, if $m = (m^*, \lambda)$ such that $\lambda \neq \lambda^*$, we have $\dot{V}(m) = 0$ and $m \neq m^*$.

Proposition 3.3.3. The equilibrium point m^* of the ODE (3.7) is asymptotically stable.

Proof. A direct application of Theorem 3.5.8, allows us to conclude that m^* is stable. Still, due to the previous remark, we cannot say that it is asymptotically stable. This is where the use of the invariant set Theorem 3.5.11 and its Corollary 3.5.12 come in. Indeed, by taking $\Omega = K$ in Corollary 3.5.11, we deduce that, provided that the largest invariant set M in $R = \{ m \in K, \dot{V}(m) = 0 \}$ is the singleton $\{ m^* \}$, every trajectory originating in K converges to m^* and hence the asymptotic stability of m^* . Now, we need to explore the set R and find the largest invariant set M in R . Let $m = (m, \lambda) \in M \subseteq R \subseteq K$. As discussed in the proof of Proposition 3.3.1, if $m = (m, \lambda) \in K$ such that $\dot{V}(m) = 0$, then necessarily $m = m^*$, that is $R \subseteq I := \{ m = (m, \lambda) \in K, m = m^* \}$. Since M is an invariant set, every trajectory originating in M should remain in M for all future times, and therefore in I . In other words, if $m(0) = (m^*, \lambda)$ for some $\lambda \geq 0$, then $m(t) = (m(t), \lambda(t)) = (m^*, \lambda(t))$ for all $t \geq 0$. Furthermore, $m(\cdot)$ is solution of the following ODE,

$$\begin{aligned} \frac{dm(t)}{dt} &= \lambda(t) E[O|(\mathbf{X} - m(t))] - \mathbf{1} + C(m(t)), \quad t \geq 0, \\ \frac{d\lambda(t)}{dt} &= E[J(\mathbf{X} - m(t))] + C(\lambda(t)), \quad t \geq 0. \end{aligned} \quad (3.9)$$

Now, since $\forall t \geq 0, m(t) = m^*$ and $m^* \in \text{int}(K)$, we get that, $C(m(t)) = 0$ and $\frac{dm(t)}{dt} = 0, \forall t \geq 0$. Moreover, we have $\forall t \geq 0, E[l(X - m(t))] = E[l(X - m^*)] = 0$ (recall that $h(m^*) = 0$), we obtain again that $C(\lambda(t)) = 0$ and $\frac{d\lambda(t)}{dt} = 0$ and consequently $t \rightarrow \lambda(t)$ is a constant function, i.e., $\lambda(t) = \lambda, \forall t \geq 0$. But we also know that, $\frac{dm(t)}{dt} = 0, \forall t \geq 0$ which implies that the right hand side of the first equation in (3.9) is 0, i.e. $\lambda E[O_m l(X - m^*)] - \mathbf{1} = 0$. Finally, we deduce that $\lambda = \lambda^*$ given that (m^*, λ^*) is the unique solution such that $h(m^*) = 0$.

We have then showed that the largest invariant set is simply $\{(m^*, \lambda^*)\}$ and therefore (m^*, λ^*) is asymptotically stable equilibrium for the ODE (3.7). □

3.3.2 Almost Surely Convergence

In the current section, we prove consistency of the algorithm (3.5). Let $\sigma^2(\cdot), \Sigma(\cdot)$ and $m^{2+p}(\cdot)$, for $p > 0$, be defined as follows:

$$\begin{aligned} \sigma^2(\cdot) &= E[||H(X, \cdot) - h(\cdot)||^2]; \\ m^{2+p}(\cdot) &= E[||H(X, \cdot) - h(\cdot)||^{2+p}]; \\ \Sigma(\cdot) &= E[(H(X, \cdot) - h(\cdot))(H(X, \cdot) - h(\cdot))^\top]. \end{aligned}$$

We make the following assumption:

- (ffi_{a.s.})
- i. h is continuous on K ;
 - ii. $\sup_{\theta \in K} \sigma^2(\theta) < \infty$.

Theorem 3.3.4. *Assume that the sequence (Z_n) is defined by the algorithm (3.5) and that assumptions (ffi₁) and (ffi_{a.s.}) hold. Then, $Z_n \rightarrow (m^*, \lambda^*)$ P- almost surely as $n \rightarrow \infty$.*

Proof. We already know that, because (m^*, λ^*) is asymptotically stable, the trajectory given by the ODE (3.7) converges to (m^*, λ^*) . Thus, (m^*, λ^*) is the only limiting for the ODE. Theorem 5.2.1 in Kushner and Yin [83] implies that $Z_n \rightarrow (m^*, \lambda^*)$ as $n \rightarrow \infty$ if we can verify their conditions (A2.1)-(A2.5). (A2.1) is guaranteed by the second assumption in (ffi_{a.s.}). (A2.2), (A2.3), (A2.4) and (A2.5) are verified thanks to the first point in (ffi_{a.s.}) and (3.6). □

3.3.3 Asymptotic normality

- (ffi_{a.n.})
- i. $m \mapsto E[O_l(X - m)]$ is continuously differentiable. Let $A := Dh(m^*)$ (Jacobian matrix of h at m^*);
 - ii. $(Y_n \mathbf{1}_{|Z_n - (m^*, \lambda^*)| \leq \rho})$ is uniformly integrable for small $\rho > 0$;
 - iii. For some $\rho > 0$ and $p > 0$, $\sup_{|m - m^*| \leq \rho} m^{2+p}(\cdot) < \infty$;
 - iv. $\Sigma(\cdot)$ is continuous at (m^*, λ^*) . Let $\Sigma^* := \Sigma(m^*, \lambda^*)$.

Theorem 3.3.5. Assume that $\gamma \in (\frac{1}{2}, 1)$ and that assumptions (ffi_l), (ffi_{a.s.}) and (ffi_{a.n.}) hold. Then,

$$\sqrt{\frac{1}{n}}(Z_n - m^*) \rightarrow N\left(0, c^2 \int_0^\infty e^{cAt} \Sigma^* e^{cA^T t} dt\right).$$

If furthermore, $cA + \frac{1}{2}I$ is a Hurwitz matrix and $cl - P$ is positive definite with P solution to the Lyapunov's equation: $A^T P + PA = -I$, then,

$$\sqrt{\frac{1}{n}}(Z_n - m^*) \rightarrow N\left(0, c^2 \int_0^\infty e^{(cA + \frac{1}{2}I)t} \Sigma^* e^{(cA + \frac{1}{2}I)^T t} dt\right).$$

Proof. We will verify that the assumptions (A2.0)-(A2.7) in Theorem 10.2.1 in Kushner and Yin [83] hold. First, let us start with the case $\gamma \in (\frac{1}{2}, 1)$. Assumption (A2.0) is automatically verified. (A2.1) is satisfied by assumption (ffi_{a.n.})-ii.. (A2.2) is a consequence of Theorem 3.3.4 and the fact that m^* is stable as shown in Section 3.3.1. (A2.4) follows from Taylor's expansion and (ffi_{a.n.})-i.. (A2.5) follows from the fact that $h(m^*) = 0$. The first and second parts of (A2.7) are guaranteed thanks to (ffi_{a.n.})-iii. and (ffi_{a.n.})-iv.. (A2.3) follows easily from Theorem 10.4.1 of Kushner and Yin [83] since all their assumptions (A4.1)-(A4.5) are satisfied. It remains to show that (A2.6) hold, that is the matrix A is a Hurwitz matrix. In fact, we have:

$$A = \begin{array}{c|c} \lambda^* DE[OI(X - m^*)] & E[OI(X - m^*)] \\ \hline -E[OI(X - m^*)] & \hat{A} \end{array} = \begin{array}{c|c} 0 & \hat{A} \\ \hline 0 & \lambda^* \end{array}$$

where $\hat{A} := -\lambda^* DE[OI(X - m^*)]$ corresponds to the second derivative of the Lagrangian L with respect to m . Note that L is strictly convex with respect to m due to the strict convexity of $m \mapsto E[I(X - m)]$. This implies that \hat{A} is positive definite matrix. Thanks to Theorem 3.6 in Benzi et al. [13], we deduce that A is a Hurwitz matrix.

For the case $\gamma = 1$, we need to verify some extra conditions related to assumptions (A2.3) and (A2.6). Indeed, the additional condition that appears in (A2.6) is satisfied since we assumed that $cA + \frac{1}{2}I$ is a Hurwitz matrix. The condition $cl - P$ is positive definite guarantees that the condition (A4.5) in Theorem 10.4.1 in Kushner and Yin [83] is satisfied so that the assumption (A2.3) is still verified in this case. □

Remark 3.3.6.

1. Note that, for convex optimization problems, where the matrix A is symmetric negative definite, the two additional conditions reduce to the classical condition $cA + \frac{1}{2}I$ is negative

definite. Indeed, in this case, the solution of the Lyapunov's equation $A^T P + PA = -I$ is simply $P = -A^{-1}/2$ and the condition $cI - P$ is positive definite, becomes equivalent to

Estimation of Systemic Shortfall Risk Measures using Stochastic Algorithms

$cA + \frac{1}{2}I$ is negative definite.

2. From a formal point of view, the choice $\gamma = 1$ gives the best rate of convergence. The asymptotic variance in this case depends on the constant c . We need to choose it such that $cA + \frac{1}{2}I$ is a Hurwitz matrix and $cI - P$ is positive definite. Setting c too small may lead to no convergence at all, while setting it too large, may lead to slower convergence as the effects of large noises early in the procedure might be hard to overcome in a reasonable period of time.
3. The choice of the constant c is a burning issue. One way to bypass this problem is to premultiply A by a conditioning matrix Γ , nonsingular, that will make A close to a constant times the identity. This can be done by considering $\gamma_n = \Gamma/n$ and we can draw the same conclusions as in Theorem 3.3.5 as soon as $\Gamma A + \frac{1}{2}I$ is a Hurwitz matrix. This will lead to the following asymptotic behaviour:

$$\sqrt{n}(Z_n - *) \rightarrow N \left(0, \int_0^{\infty} e^{(\Gamma A + \frac{1}{2}I)t} \Gamma \Sigma^* \Gamma^T e^{(A^T \Gamma^T + \frac{1}{2}I)t} dt \right).$$

The optimal choice of the conditioning matrix Γ , which is also called the gain matrix, is the one that will minimize the trace of asymptotic covariance:

$$\int_0^{\infty} e^{(\Gamma A + \frac{1}{2}I)t} \Gamma \Sigma^* \Gamma^T e^{(A^T \Gamma^T + \frac{1}{2}I)t} dt.$$

This is done by taking $\Gamma = -A^{-1}$ which yields the asymptomatic optimal covariance: $A^{-1} \Sigma^* (A^{-1})^T$.

4. The optimal choice of Γ depends on the function h and the equilibrium point $*$ which are unknown to us. Adaptive procedures that choose the matrix Γ dynamically by estimating $Dh(*)$ adaptively have been suggested in the literature (see for example Ruppert [117]), but are generally not as efficient as the Polyak-Ruppert averaging estimators discussed in the following section.

3.3.4 Polyak-Ruppert Averaging principle

In order to ease the tuning of the step parameter which known to monitor the numerical efficiency of RM algorithms, we are led to modify our algorithm and to use an averaging procedure. Averaging algorithms were introduced by Ruppert (see Ruppert [117]) and Polyak (see Polyak and Juditsky [108]) and then widely investigated by many authors. Kushner and Yin [83] and Kushner and Yang [82] studied these algorithms in combination with projection and proved a Central Limit Theorem (CLT) for averaging constrained algorithms.

The following theorem describes the Polyak-Ruppert algorithm for MSRM and states its asymptotic normality. It is a direct consequence of theorem 11.1.1 in Kushner and Yin [83].

Theorem 3.3.7. *Assume $\gamma \in (\frac{1}{2}, 1)$ and that assumptions (ffi_l), (ffi_{a.s.}) and (ffi_{a.n.}) hold. For any arbitrary $t > 0$, we define*

$$\bar{Z}_n = \frac{\gamma_n}{t} \sum_{i=n}^{n+t\gamma_n-1} Z_i \tag{3.10}$$

where any upper summation index $u \in \mathbb{R}^+$ is interpreted as its integer part. If Σ^* is positive definite, then we have the following CLT:

$$\frac{t}{\gamma_n} \bar{Z}_n \xrightarrow{d} N(0, V + O(\frac{1}{t})), \tag{3.11}$$

where $V = A^{-1}\Sigma^*(A^{-1})^T$.

Remark 3.3.8.

1. In (3.10), the window of averaging is t/γ_n for any arbitrary real $t > 0$. Equivalently, $\gamma_n \times$ (size of window) does not go to infinity as $n \rightarrow \infty$, hence the name “minimal window” of averaging. In contrast, the “maximal window” of averaging allow to take a window size $q_n t$ such that $\gamma_n q_n \rightarrow \infty$. A natural and a classical choice is taking $\gamma_n = c/n^\gamma$ and $q_n = n$. In the case of maximal window of averaging, under some extra conditions, we are able to achieve the optimal asymptotic variance without an extra term $O(1/t)$ (see Theorem 11.3.1 in Kushner and Yin [83]).
2. Two sided averages can also be used instead of the one-sided average in (3.10).

3.3.5 Estimator of asymptotic variance

The previous CLT theorems assert that, under some suitable conditions, our RM and PR algorithms converge to the root $*$ with a corresponding rate. More specifically, in Theorem 3.3.7, the asymptotic variance V depends on Σ^* and A . In practice, these two quantities are unknown and need to be approximated in order to derive confidence intervals for our estimators. In Theorem 3.3.5, in both cases, $\gamma = 1$ and $\gamma \in (\frac{1}{2}, 1)$, the asymptotic variance is expressed as an infinite integral that involves Σ^* and A . The numerical evaluation of these integrals is a non-trivial exercise even when Σ^* and A are known. In Hsieh and Glynn [67], they described an approach that produces confidence regions and that avoids the necessity of having to explicitly estimate these integrals.

In the following proposition, we provide consistent estimators of these two quantities. The proof relies mainly on the Martingale Convergence Theorem.

Proposition 3.3.9. Assume (ffi_i), (ffi_{a.s.}) and (ffi_{a.n.}) hold.

If $\mathbb{E}[||H(X, Z)||^4]$ is locally bounded around Σ^* , then,

$$\frac{1}{n} \sum_{k=1}^n H(X_k, Z_{k-1})^\top H(X_k, Z_{k-1}) \rightarrow \Sigma^* \text{ a.s.} \quad (3.12)$$

Let A_n^ϵ the matrix whose elements $A_n^\epsilon(i, j)$ for $i, j \in \{1, \dots, d+1\}$ are defined as follows:

$$A_n^\epsilon(i, j) := \frac{1}{\epsilon n} \sum_{k=1}^n H_i(X_k, Z_{k-1} + \epsilon e_j) - H_i(X_k, Z_{k-1}),$$

then,

$$\lim_{\epsilon \rightarrow 0} \lim_{n \rightarrow \infty} A_n^\epsilon = A \text{ a.s.} \quad (3.13)$$

Proof. Let $(S_n)_{n \in \mathbb{N}^*}$ be the sequence defined as:

$$S_n = H(X_n, Z_{n-1})^\top H(X_n, Z_{n-1}) - \Sigma(Z_{n-1}) - h(Z_{n-1})^\top h(Z_{n-1}), \quad n \geq 1$$

$(S_n)_{n \in \mathbb{N}^*}$ is a martingale difference sequence adapted to \mathbb{F} and consequently the following sequence $(M_n)_{n \in \mathbb{N}^*}$ defined as:

$$M_n = \sum_{i=1}^n S_i, \quad n \geq 1,$$

is a \mathbb{F} -martingale. Moreover, the boundedness of $\mathbb{E}[||H(X, Z)||^4]$ around Σ^* and assumptions (ffi_{a.s.})-i. and (ffi_{a.n.})-iv. imply that:

$$\sup_{n \geq 1} \mathbb{E}[||S_n||^2 | \mathbb{F}_{n-1}] < \infty \text{ a.s.}$$

Thus, the martingale convergence theorem ensures the existence of a finite random variable M_∞ such that $M_n \rightarrow M_\infty$ a.s. Kronecker's lemma then guarantees that $\frac{1}{n} \sum_{i=1}^n S_i \rightarrow 0$. Now,

since,

$$\Sigma_n = \frac{1}{n} \sum_{i=1}^n S_i + \frac{1}{n} \sum_{i=1}^n \Sigma(Z_{i-1}) + \frac{1}{n} \sum_{i=1}^n h(Z_{i-1})^\top h(Z_{i-1}),$$

we deduce that $\Sigma_n \rightarrow \Sigma^*$.

The proof of (3.13) follows using the same arguments above. □

Remark 3.3.10.

1. Instead of averaging on all observations, one could modify the estimators above and average only on recent ones. This might improve the behaviour of these estimators.
2. If we denote $V_n := A_n^{-1} \Sigma_n (A_n^{-1})^\top$, then we obtain an approximate confidence interval for

PR estimator with a confidence of $1 - \alpha$ in the following form:

$$\bar{Z}_{j,n} - \frac{s}{tn^\gamma} q_\alpha \leq \bar{Z}_{j,n} + \frac{s}{tn^\gamma} q_\alpha, j \in \{1 \dots d\}, \gamma \in (0, 1), \quad (3.14)$$

where q_α is the $1 - \frac{\alpha}{2}$ quantile of a standard normal random variable. Note that this confidence interval has the advantage of being obtained with one simulation run. For RM estimators, confidence intervals could be estimated empirically.

3.4 Numerical examples

In this section, we test the performance of the proposed stochastic algorithms schemes for MSRM. In Armenti et al. [4], the optimal allocations were estimated by using a combination of Monte Carlo/Fourier method to estimate the expectation in (3.2) and deterministic built-in search algorithm in Python to find the optimal allocations. Although their method provides good approximations, it does not provide any rate of convergence and therefore one cannot say anything about the confidence interval of their estimations. In this section, we will first test the consistency properties of the different estimators and then their normal asymptotic behaviour with and without averaging. Two examples are considered. In the first one, we consider a loss function of an exponential type coupled with a normal distribution. This example is relevant for our numerical analysis as we can explicitly express the optimal allocations in a closed form. In the second example, we consider a loss function that involves positive part function with a Gaussian and a compound Poisson distributions.

In the following, n will denote the number of steps in one simulation run and N the number of simulations. We introduce the following sequences:

$$\bar{D}_n := \sqrt{\frac{s}{tn^\gamma}} (\bar{Z}_n - \mu), \gamma \in (\frac{1}{2}, 1), \quad (3.15)$$

$$D_n := \sqrt{\frac{s}{n^\gamma}} (Z_n - \mu), \gamma \in (\frac{1}{2}, 1]. \quad (3.16)$$

3.4.1 Toy example

As a first simple example, we will consider a exponential loss function of the following form:

$$l(x_1, \dots, x_d) = \frac{1}{1 + \alpha} \sum_{i=1}^d e^{\beta x_i} + \alpha e^{\beta \sum_{i=1}^d x_i} - \frac{\alpha + d}{\alpha + 1} \quad (3.17)$$

We will set $d = 2$ and consider a bivariate normal vector $X = (X_1, X_2) \sim N(0, M)$ with $M = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$. α is a systemic weight parameter taken to be non negative and $\beta > 0$ is the risk aversion coefficient. In this case, we can explicitly solve the first order conditions and derive closed formulas for optimal allocations (see Section 3.5.2). This will be useful to test our algorithms:

$$m_i^* = \begin{cases} \frac{\beta\sigma_i^2}{2}, & \text{if } \alpha = 0, \\ \frac{\beta\sigma_i^2}{2} + \frac{1}{\beta} SRC(\rho, \sigma_1, \sigma_2, \alpha, \beta), & \text{if } \alpha > 0. \end{cases}$$

This shows that, in the case $\alpha > 0$, the risk allocations are disentangled into two components: an individual contribution $\frac{\beta\sigma_i^2}{2}$ and a *Systemic Risk Contribution* (SRC) given by:

$$SRC(\rho, \sigma_1, \sigma_2, \alpha, \beta) = \ln \frac{\alpha e^{\rho\beta^2\sigma_1\sigma_2}}{1 + \frac{\alpha(\alpha+2)}{\rho\beta^2\sigma_1\sigma_2}}$$

Note that taking $\alpha \rightarrow 0$ makes the SRC null as expected because, the systemic weight α is responsible of the systemic contribution in the loss function l . One can also show, by easy calculations, that the SRC is increasing with respect to ρ : the higher the correlation is, the more costly the acceptable monetary allocations are. This could be explained by the fact that, with a higher correlation between the two components, the losses of one will induce the loss of the other and consequently the system will become riskier. Note also that we could also express in a closed form the Jacobian matrix A and Σ^* .

In all this example, we fix $\alpha = 1$, $\beta = 1$ and $\sigma_1 = \sigma_2 = 1$. With $\rho \in \{-0.5, 0, 0.5\}$, we obtain the exact values in the table below. Note that since we have $X_1 \sim X_2 \sim N(0, 1)$ and l is permutation invariant, it follows that $m_1^* = m_2^*$.

| ρ | $m_1^* = m_2^*$ |
|--------|-----------------|
| -0.5 | 0.3868 |
| 0 | 0.5 |
| 0.5 | 0.6364 |

Table 3.1: Exact optimal risk allocations.

For RM/PR algorithms, we used a number of steps $n = 10^5$. As for the compact K , we took $K = [0, 2]^3$ and Z_0 was taken uniformly on K . We run the different algorithms for $\gamma = 1$ and $\gamma = 0.7$. We chose an averaging parameter $t = 10$ and we set $c = 2$ in a first step. Figure 3.1 shows that, for different values of $\rho \in \{-0.5, 0, 0.5\}$, our RM algorithm with $\gamma = 1$ converges relatively quickly to the optimal allocations, whereas when $\gamma = 0.7$, noise is still persisting. This

is due to the step parameter c as discussed in the previous section. In order to get a smoother numerical behaviour, two solutions are available to us: either we use PR averaging (c.f. Figure 3.1), or we reduce the value of the parameter c . This is shown in Figure 3.2.

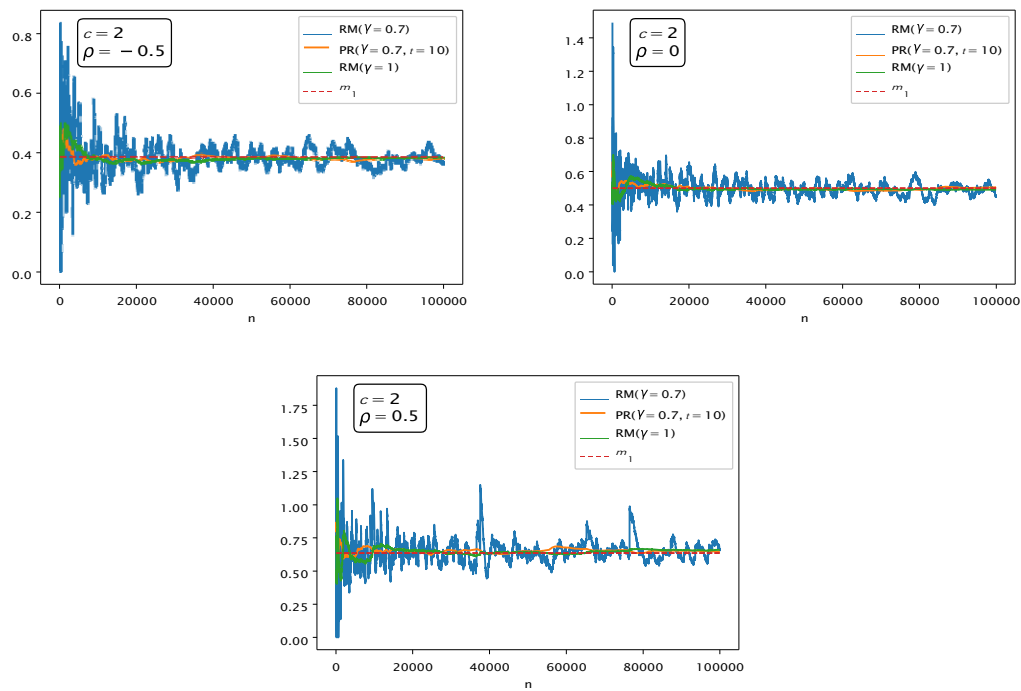


Figure 3.1: Consistency of RM/PR estimators with for different values of ρ .

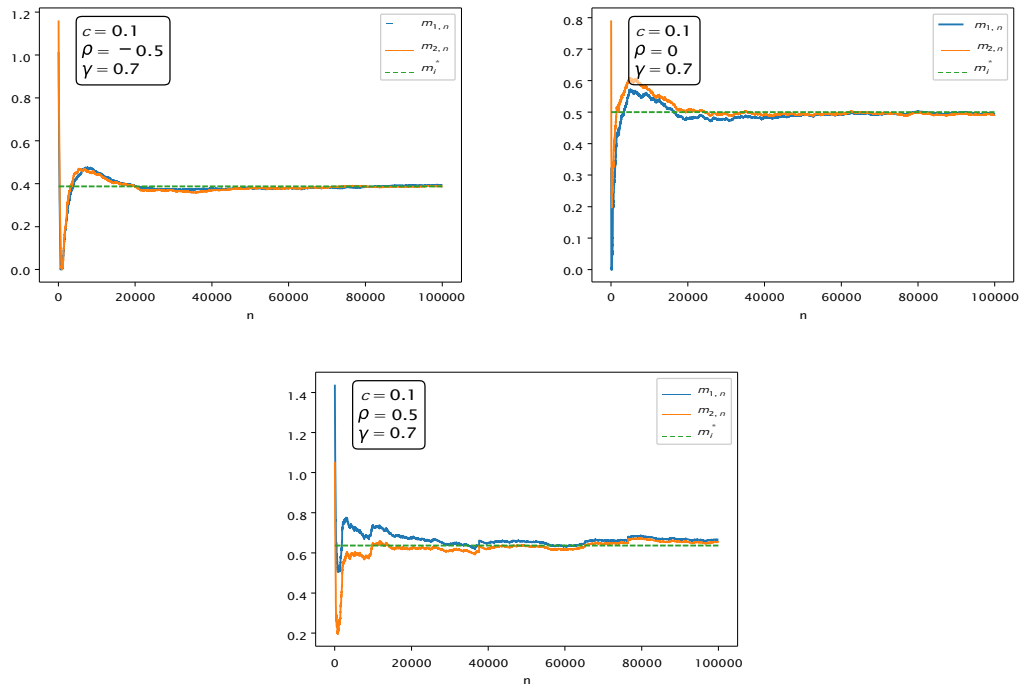


Figure 3.2: Consistency of RM estimators with $c = 0.1$ for different values of ρ .

Note that we can easily verify that all conditions in (ffi_{a.s.}) and (ffi_{a.n.}) hold. We can also verify thanks to the exact formula of Σ^* , that this matrix is positive definite for the different values of ρ used. This is a condition needed in Theorem 3.3.7.

For any random estimator, constructing confidence intervals is important to assess the error in the estimation. For PR estimator, confidence interval can be obtained in one simulation run after estimating matrices Σ^* and A and hence the asymptotic variance matrix V . Figure 3.3 shows the convergence, in the case $\rho = 0$, of the estimator of $V_n = A_n^{-1}\Sigma_n(A_n^{-1})^\top$ where A_n and Σ_n are as introduced in Proposition 3.3.9.

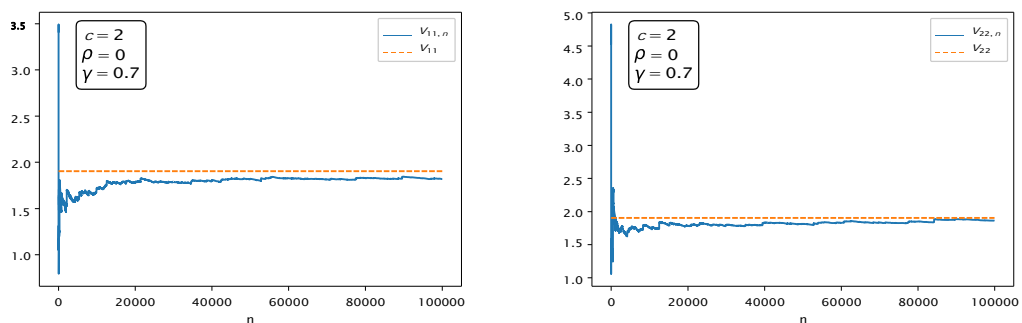


Figure 3.3: Convergence of the estimator V_n .

In the following table, we give the estimated confidence interval for PR estimator with a confidence coefficient of 95%: As for RM estimators, it is difficult to estimate the asymptotic

| ρ | CI for m_1^* | CI for m_2^* |
|--------|------------------|------------------|
| -0.5 | [0.3772, 0.4047] | [0.3679, 0.3949] |
| 0 | [0.4962, 0.5259] | [0.4912, 0.5213] |
| 0.5 | [0.6194, 0.6629] | [0.6203, 0.6665] |

Table 3.2: Confidence intervals for PR estimators.

covariance matrix due to its complexity. In order to visualize the normal behaviour of these estimators, we give the empirical probability density function (EPDF) in both cases $\gamma = 1$ and $\gamma = 0.7$. To this end, we use again a number of steps $n = 100000$ and we repeat the procedure $N = 10000$ times. We restrict our attention to the case $\rho = 0$. Figure 3.4 shows that $D_{n,i} = \sqrt{n\gamma}(m_{n,i} - m_i^*)$, $i \in \{1, 2\}$ are very close to a normal distribution.

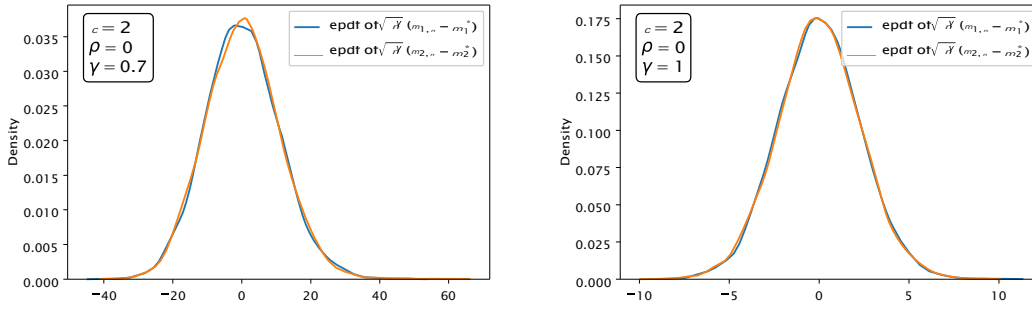


Figure 3.4: Empirical cumulative density function of $m_n - m^*$.

In order to appreciate the quality of convergence of RM estimators, we also give the empirical cumulative density function (ECDF) of the error $m_n - m^*$.

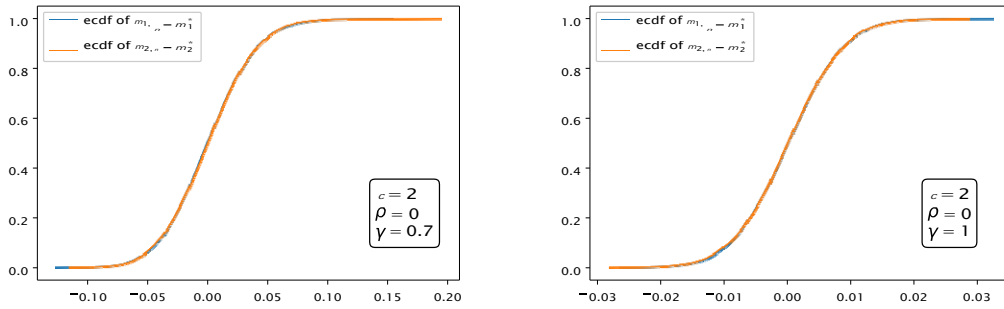


Figure 3.5: Empirical cumulative density function of $m_n - m^*$.

From the two figures above, the width of the 90% confidence interval of the RM estimator for the case $\gamma = 0.7$ is approximately 8% and for the case $\gamma = 1$ is roughly 2%.

3.4.2 Second example

As a second example, we will consider consider the following loss function used in Armenti et al. [4]:

$$l(x_1, \dots, x_d) = \sum_{i=1}^d x_i + \frac{1}{2} \sum_{i=1}^d (x_i^+)^2 + \alpha \sum_{i < j} x_i^+ x_j^+.$$

First case: Gaussian distribution and $d = 2$

We start by a simple case where we fix $d = 2$ and use standard two dimensional Gaussian distribution for the loss vector X . We take $K = [0, 2]^3$, $n = 10^5$, $t = 10$, $\alpha = 1$ and $c = 6$. Again, we compare RM and PR estimators for different values of ρ . The following figure 3.6 allows us to draw the same conclusions as in the previous example: RM estimator with $\gamma = 1$ and PR estimator are better than RM estimator with $\gamma = 0.7$. RM estimator with $\gamma = 0.7$ is noisy and one can remediate to this by choosing a smaller value of c as we did in the first example.

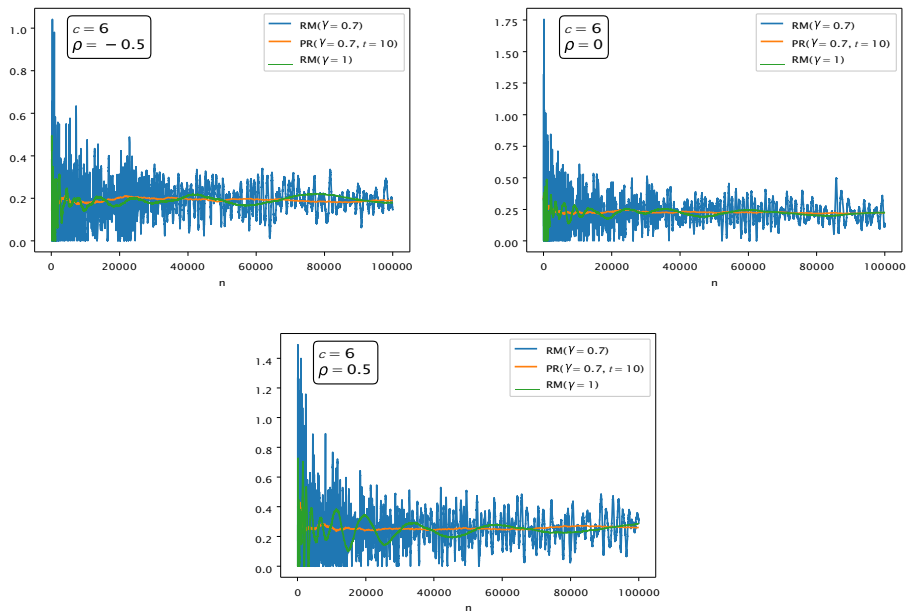


Figure 3.6: Consistency of RM/PR estimators with for different values of ρ .

In order to assess the accuracy of our PR estimator, we give the confidence interval with a 95% confidence coefficient, using the estimators of Proposition 3.3.9. For RM estimator, we

| ρ | $m_1^* = m_2^*$ | CI for m_1^* | CI for m_2^* |
|--------|-----------------|------------------|------------------|
| -0.5 | 0.188 | [0.1790, 0.2089] | [0.1746, 0.2045] |
| 0 | 0.21 | [0.1963, 0.2303] | [0.2044, 0.2385] |
| 0.5 | 0.25 | [0.2415, 0.2769] | [0.2424, 0.2777] |

Table 3.3: Confidence intervals for PR estimators.

plotted the EPDF of $D_{n,i}$, $i \in \{1, 2\}$ as well as the ECDF of the error $m_n - m^*$ for the case $\rho = 0$ and $\gamma = 1$. These figures shows that the length of the confidence interval of 90% in the case $\gamma = 0.7$ is much higher that in the case $\gamma = 1$ (approx 0.2 against 0.04).

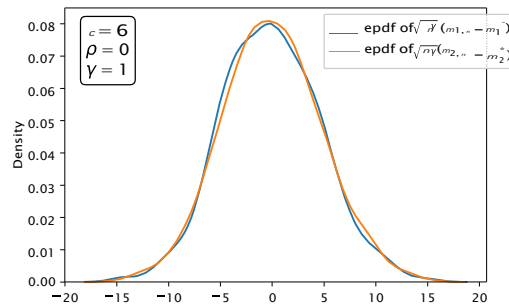


Figure 3.7: Empirical cumulative density function of $m_n - m^*$.

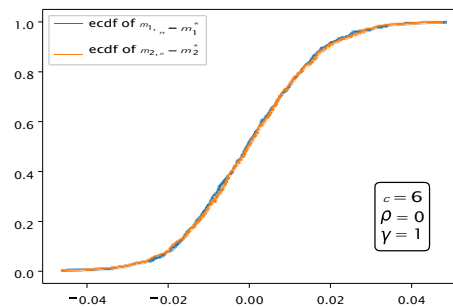


Figure 3.8: Empirical cumulative density function of $m_n - m^*$.

Second case: Compound Poisson Distribution and higher dimensions

In this section, we propose to use compound Poisson processes to model the loss vector X . The scope of application of compound Poisson processes is very wide. It ranges from statistical physics and biology to financial mathematics. In biology, they are used to study dynamics of populations. In the modern financial modeling, compound Poisson processes are used to describe dynamics of risk factors such as interest rates (see for instance Li et al. [88]), foreign exchange rates and option pricing (see Jaimungal and Wang [71]). In actuarial science, compound processes are extensively used to model claims sizes and to compute the ruin probability, i.e. the probability that the initial reserves increased by premiums received from clients and decreased by their claims, drops below zero.

More precisely, given a final time T , we consider a multivariate Poisson random vector $N^T = (N_1^T, \dots, N^T)$, where each $N^T \sim P(\lambda_i T)$ and the loss corresponding to the i^{th} component is $X = \sum_i N_i^T G^k$ and (G^k) is an i.i.d sequence representing the jump sizes and independent of

N^T . We will take two examples for the distribution of the jumps sizes: One with a Gaussian distribution and another one with an exponential one. The correlation between the different compo-

nents of X will be done through the correlations between components of N^T . In what follows, we detail the method of generating a multivariate Poisson random vector, $N = (N_1, \dots, N_d)$ with a vector of corresponding intensities $(\lambda_1, \dots, \lambda_d)$. To do so, we will use a method that is based on the Gaussian vectors. More precisely, denote $\eta = (\eta_1, \dots, \eta_d)$ to be a Gaussian random vector having a centered normal distribution with correlation matrix $R = (\rho_{kl})$ and Φ to be the standard normal cdf. Then, the random vector $\xi = (\Phi(\eta_1), \dots, \Phi(\eta_d))$ has a multivariate distribution with standard uniform marginal distributions. Let $P_\lambda(x) = \sum_{j=0}^{[x]} (\lambda^j / j!) e^{-\lambda}$ be the cdf of the Poisson distribution with parameter λ . Now, consider the vector $\zeta = (\zeta_1, \dots, \zeta_d)$ where $\zeta_k = P_{\lambda_k}^{-1}(\Phi(\eta_k))$, $k = 1, \dots, d$. ζ has therefore Poisson marginal distributions with intensities $(\lambda_1, \dots, \lambda_d)$. We can express the correlation coefficient $\rho_{kl}^* = \text{corr}(\zeta_k, \zeta_l)$ as a function of $\rho_{kl} = \text{corr}(\eta_k, \eta_l)$:

$$\begin{aligned} \rho_{kl}^* &= \frac{E(\zeta_k \zeta_l) - E(\zeta_k)E(\zeta_l)}{\sqrt{\sigma(\zeta_k)\sigma(\zeta_l)}} \\ &= \frac{E(\zeta_k \zeta_l) - \lambda_k \lambda_l}{\sqrt{\lambda_k \lambda_l}}. \end{aligned}$$

We need to express the expectation $E(\zeta_k \zeta_l)$ as a function of ρ_{kl} . We have:

$$\begin{aligned} E(\zeta_k \zeta_l) &= E \left[P_{\lambda_k}^{-1}(\Phi(\eta_k)) P_{\lambda_l}^{-1}(\Phi(\eta_l)) \right] \\ &= \sum_{m=0}^{\infty} \sum_{n=0}^{\infty} mn P(\zeta_k = m, \zeta_l = n) \\ &= \sum_{m=0}^{\infty} \sum_{n=0}^{\infty} mn P(u_{m-1}^k \leq \Phi(\eta_k) \leq u_m^k, u_{n-1}^l \leq \Phi(\eta_l) \leq u_n^l), \end{aligned}$$

where $u_j^i = P_\lambda(j)$, $i, j = 1, \dots, d$. It remains to explicit the probabilities in the last equality. If we denote $\Phi_2(\cdot, \cdot, \rho_{kl})$ the bivariate Normal distribution function, we get finally that,

$$\begin{aligned} Z_{mn}(\rho_{kl}) &:= P(u_{m-1}^k \leq \Phi(\eta_k) \leq u_m^k, u_{n-1}^l \leq \Phi(\eta_l) \leq u_n^l) \\ &= \Phi_2(A_m^k, B_n^l, \rho_{kl}) - \Phi_2(A_{m-1}^k, B_n^l, \rho_{kl}) - \Phi_2(A_m^k, B_{n-1}^l, \rho_{kl}) + \Phi_2(A_{m-1}^k, B_{n-1}^l, \rho_{kl}) \end{aligned}$$

where $A_m^k = \Phi^{-1}(P_{\lambda_k}(m))$ and $B_n^l = \Phi^{-1}(P_{\lambda_l}(n))$. As a conclusion, we obtain,

$$\sum_{m=1}^{\infty} \sum_{n=1}^{\infty} mn Z_{mn}(\rho_{kl}) = \lambda_k \lambda_l + \rho_{kl}^* \sqrt{\lambda_k \lambda_l}. \quad (3.18)$$

The equation (3.18) gives an implicit relation between ρ_{kl}^* and ρ_{kl} . It also involves two infinite sums which makes it hard to solve. In practice, one needs to truncate this sum and choose some appropriate upper-limits M^* and N^* . We are then able to compute the elements of the correlation matrix ρ of the Gaussian vector given the correlation matrix ρ^* of the vector N .

However, there is a problem of sufficient conditions for a given positive semi-definite matrix to be a correlation matrix of a multivariate Poisson random vector. This issue is tackled in Griffiths et al. [60] where it is shown that each ρ_{kl}^* has to be in a certain range,

$$-1 < \rho_{kl}^{\min} \leq \rho_{kl}^* \leq \rho_{kl}^{\max} \leq 1. \quad (3.19)$$

Algorithm 1: Algorithm for generating a sample of X with Compound Poisson Distribution

Input: $(\lambda_i)_{i=1,\dots,d}$, $(\sigma_i)_{i=1,\dots,d}$, T , and ρ^* correlation matrix of N_T ;

Ensure: For each $k > l$, ρ_{kl}^* verifies the inequality in (3.19);

- 1 Solve the equation (3.18) to find ρ the correlation of the Gaussian vector $\eta = (\eta_1, \dots, \eta_d)$;
- 2 Generate a sample of Gaussian vector η with correlation matrix ρ and for $i = 1, \dots, d$;
- 3 **for** $i = 1, \dots, d$ **do**
- 4 **Set** $N_i^T = P^{-1}(\Phi(\eta_i))$;
- 5 Generate a i.i.d sample of G_i^k of size N_i^T ;
- 6 **Set** $X_i = \sum_{k=1}^T G_i^k$.

Output: X ;

The following figure shows the covariance matrix for the loss vector X of dimension $d = 10$ obtained by generating a random correlation matrix (ρ_{kl}) and using a Gaussian distribution for the jump sizes, i.e. $G_i^k \sim N(1, 1)$. The intensity vector was taken uniformly in $[1, 3]^{10}$.

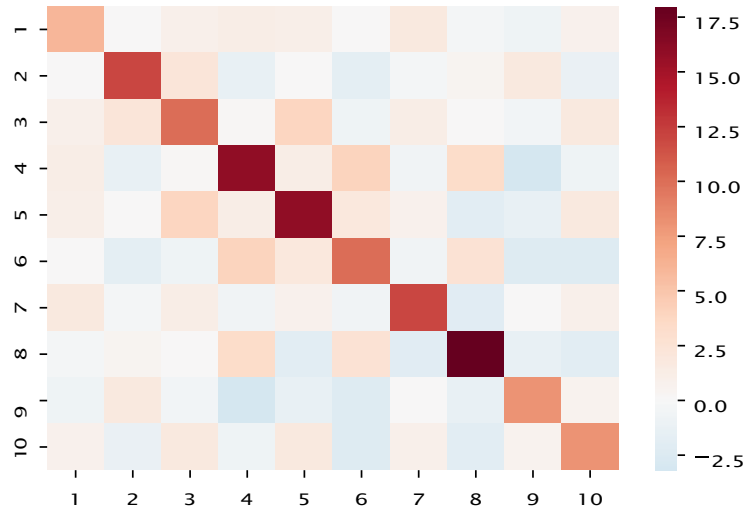


Figure 3.9: Correlation matrix of the vector loss X with Gaussian jump sizes with means and variances equal to 1.

Setting $K = [-20, 20]^{10} \times [0, 20]$, the averaging parameter $t = 10$ and $c = 6$, $\gamma = 0.7$ and the number of steps $n = 100000$, we obtain the following optimal allocations for both cases $\alpha = 0$ and $\alpha = 1$.

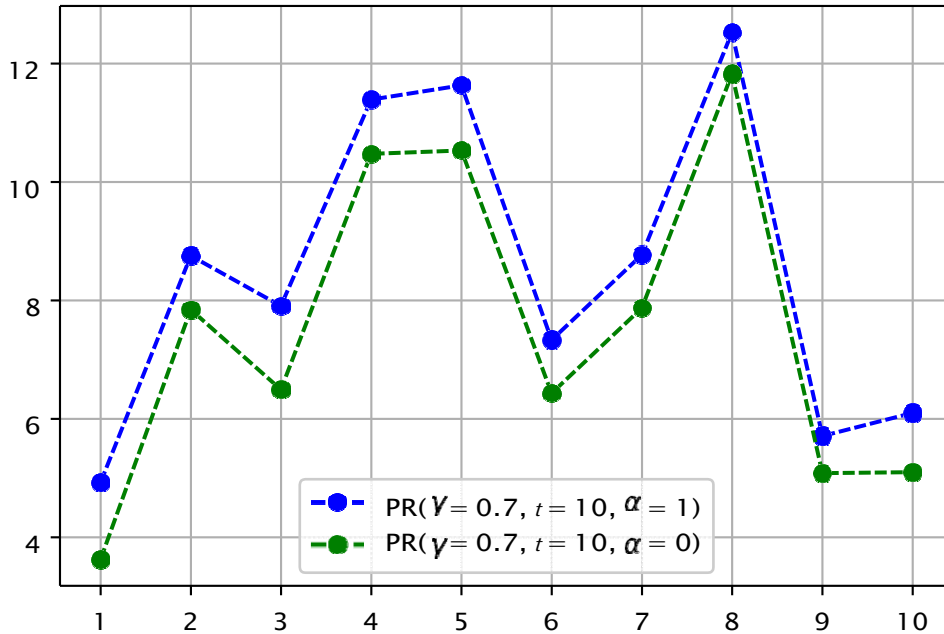


Figure 3.10: PR estimators for optimal allocations.

The above figure shows that there are components with the same optimal allocations for the case $\alpha = 0$. This is something we expect to see, since with $\alpha = 0$, correlations between components are not involved, so components with the same variance should have the same optimal allocations. This is the case for components 4 – 5 and components 9 – 10. However, once α is taken non null, we see that the same components have no longer the same optimal allocations. For instance, component 10 has higher optimal allocation than 9 when $\alpha = 1$. This could be explained by the fact that component 10 is more correlated with other components that have high variances, such as components 4 and 5, than component 9. We now consider an exponential distribution for jump sizes as a second example, i.e. $G_i^k \sim E(a_i)$. The parameters a_i were generated randomly in $[0.2, 1.2]$. As for the other parameters in this example, we took again $K = [-20, 20]^{10} \times [0, 20]$, $c = 6$, $t = 10$ and $\gamma = 0.7$. Covariance matrix of the loss vector X in this case and estimators of the optimal allocations obtained through PR algorithms with a number of steps $n = 100000$ together with corresponding confidence intervals are given in the following figures.

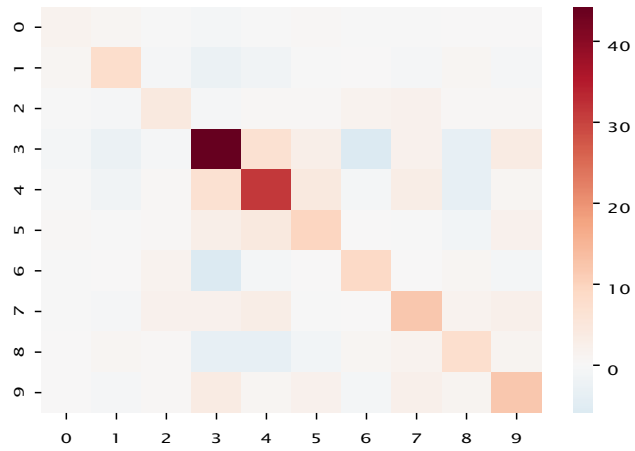


Figure 3.11: Correlation matrix of the vector loss X with exponential distribution for the law of jump sizes .

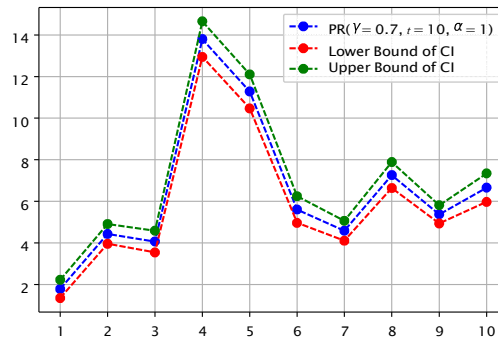
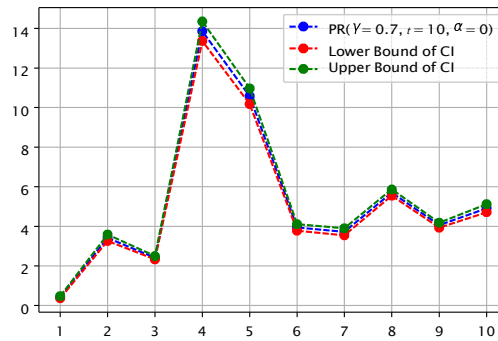


Figure 3.12: PR estimators for optimal allocations together with bounds of CI.

3.5 Appendix

3.5.1 ODE method and related concepts

Suppose we want to find the zeros of a function h . If we had a closed formula for h , under some classical conditions, we could use the following algorithm that ensures that at each step, we are going in the right direction: $Z_{n+1} = Z_n \pm \gamma_n h(Z_n)$, where (γ_n) could be a constant sequence or decreasing toward 0. However, if we do not have access to h , but only to random estimates Y_n that are close to h on average, then we could replace $h(Z_n)$ by Y_n : $Z_{n+1} = Z_n + \gamma_n Y_n$. This is typically the case when h is expressed as an expectation: $h(\cdot) = E[H(X, \cdot)]$ with X is a random variable. An estimate of h at step $n + 1$, given all the $(Z_i)_{i=0, \dots, n}$, is $Y_n = H(X_{n+1}, Z_n)$, where X_{n+1} is a random variable that has the same law as X . Then we could write the algorithm as:

$$Z_{n+1} = Z_n + \gamma_n H(X_{n+1}, Z_n). \tag{3.20}$$

If we denote by (F_n) the following filtration:

$$F_n = \sigma(Z_0, X_i, i \leq n),$$

and rewrite:

$$Y_n = h(Z_n) + \delta M_n,$$

where $\delta M_n = Y_n - h(Z_n)$. Observe that $h(Z_n) = E[Y_n | F_n]$ implies that δM_n is a martingale difference sequence. Therefore, another way to write the (3.22) is as the following:

$$Z_{n+1} = Z_n + \gamma_n h(Z_n) + \gamma_n \delta M_n.$$

The algorithm (3.20) is the regular Robbins-Monro (RM) procedure with mean function h . In order, to obtain a.s. convergence of the algorithm toward the \ast , one crucial condition among others, is the sublinearity of h , which is very constraining on the type of functions h we can use. Consequently, we will drop the classical version of RM and will adopt the ordinary differential equation (ODE) point of view which offers more flexibility. The ODE method has its own drawbacks: it requires the sequence $(Z_n)_{n \geq 1}$ to be in a compact set K for non-explosion reasons. Still, this is not very constraining: In fact, each time Z_n goes out of K , we will replace it by the closest point to Z_n in K , using projection.

To get the intuition behind the ODE method, we will assume that δM_n have bounded variances. First, note that since $\gamma_n \rightarrow 0$, for large n the values of Z_n change slowly. For small Δ , define m_n^Δ such that:

$$\sum_{i=n}^{n+m_n^\Delta-1} \gamma_i \approx \Delta.$$

Then,

$$Z_{n+m_n^\Delta} - Z_n \approx \Delta h(Z_n) + \text{''error''} \quad (3.21)$$

where the error term is given by

$$\sum_{i=n}^{n+m_n^\Delta-1} \gamma_i \delta M_i.$$

Now, since (δM_n) is a martingale difference sequence, this implies that $E[\delta M_i \delta M_j] = 0$ for $i \neq j$. Thus, the variance of the error term is

$$E \left[\sum_{i=n}^{n+m_n^\Delta-1} \gamma_i \delta M_i \right]^2 = \sum_{i=n}^{n+m_n^\Delta-1} E[\gamma_i^2 \delta M_i^2] = \sum_{i=n}^{n+m_n^\Delta-1} O(\gamma_i^2) = O(\Delta) \gamma_n.$$

This bound and equation (3.21) show that, over the time interval $[n, n + m_n^\Delta]$ for small Δ and large n , the change in the value of Z_n is actually due to the mean term more than the error term or noise term. Then, at least formally, the equation (3.21) suggests that the asymptotic behaviour of the algorithm can be approximated by the asymptotic behaviour of the solution to the ODE:

$$\dot{z} = h(z).$$

As said before, in order to be able to use the ODE method, we need to make sure that the terms Z_n are in some compact set K . This will be done using projection on a fixed compact set K . We will assume that K is a hyperrectangle, i.e. $K = \{z, a_i \leq z_i \leq b_i\}$ (more general compact sets have been studied in Kushner and Yin [83]), and we consider the following algorithm:

$$Z_{n+1} = \Pi_K[Z_n + \gamma_n Y_n], \quad Z_0 \in K, \quad (3.22)$$

where $(X_n)_{n \geq 1}$ is an i.i.d sequence of random variables with the same distribution as X , independent of Z_0 and $(\gamma_n)_{n \geq 1}$ is a deterministic step sequence. The previous algorithm in (3.22) could be rewritten using the correction term C_n defined as:

$$Z_{n+1} = Z_n + \gamma_n Y_n + \gamma_n C_n. \quad (3.23)$$

Thus $\gamma_n C_n = Z_{n+1} - Z_n - \gamma_n Y_n$; it is the vector of shortest Euclidean length needed to take $Z_n + \gamma_n Y_n$ back to the hyperrectangle K if it is not in K .

To get a geometric feeling for the correction term C_n , refer to the figures 3.13 and 3.14. In situations such as Figure 3.13, where only one component is being truncated, C_n points inward and is orthogonal to the boundary. If more than one component needs to be truncated, as in figure 3.14, C_n again points inward but toward the corner, and it is proportional to a convex combination of the inward normals at the faces that border that corner. In both cases, we have

$C_n \in -C(Z_{n+1})$, where $C(Z)$ is the convex cone determined by the outer normals to the faces that need to be truncated at Z .

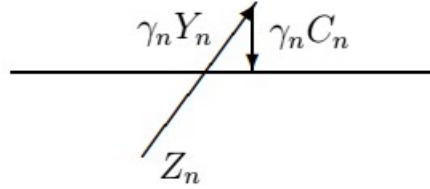


Figure 3.13: A projection with one violated constraint.

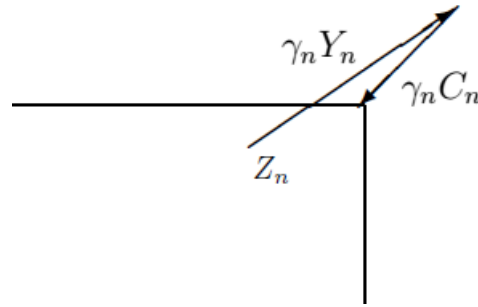


Figure 3.14: A projection with two violated constraints.

Constant-wise Interpolation

The ODE method uses a continuous-time interpolation of the sequence (Z_n) . More precisely, define $t_0 = 0$ and $t_n = \sum_{i=0}^n \gamma_i$. For $t \geq 0$, let $m(t)$ be the unique value of n such that $t_n \leq t < t_{n+1}$. Let $Z^0(t)$ be the continuous time interpolation defined for $t \geq 0$ by:

$$Z^0(t) = Z_n, \text{ for } t_n \leq t < t_{n+1}.$$

Since the step sequence decreases to 0, the process $Z^0(\cdot)$ becomes more regular, so it is natural to introduce the *shifted version* $Z^n(t)$ defined by:

$$Z^n(t) = Z^0(t + t_n), \quad t \geq 0.$$

For the correction term, we define $C^0(t)$ as:

$$C^0(t) = \sum_{i=0}^{m(t)-1} \gamma_i C_i, \quad t \geq 0,$$

and

$$C^n(t) = C^o(t_n + t) - C^o(t) = \sum_{i=n}^{m(t_n+t)-1} \gamma_i C_i, \quad t \geq .0$$

Define $\delta M^n(t)$ in the same way as $C^n(t)$. By all the definitions above:

$$\begin{aligned} Z^n(t) &= Z_n + \sum_{i=n}^{m(t_n+t)-1} \gamma_i [Y_i + C_i] = Z_n + \sum_{i=n}^{m(t_n+t)-1} \gamma_i [h(Z_i)] + \delta M_i + C_i \\ &= Z_n + \int_t^{m(t_n+t)-1} h(Z^n(s)) ds + \delta M^n(t) + C^n(t) + \rho^n(t), \quad t \geq 0, \end{aligned} \quad (3.24)$$

where, $\rho^n(t)$ is due to the replacement of the first sum by an integral. $\rho^n(\cdot)$ converges to 0 uniformly in t when $n \rightarrow \infty$. Formally, the tail behaviour of the sequence Z_n is equivalent to that of the process $Z^n(\cdot)$ under some broad conditions. Thanks to the previous equation, we see that the process $Z^n(\cdot)$ is close to the solution of the ODE:

$$\dot{z} = h(z) + C, \quad \text{where } C(t) \in -C(z(t)), \quad (3.25)$$

where $C(\cdot)$ is again the minimum force needed to keep (\cdot) in K . More closely, if (\cdot) is in $\text{int}(K)$ on some time interval, then $C(\cdot)$ is zero on that interval. If, for some i , $i(t) = a_i$ (or $i(t) = b_i$ resp.) and $h_i(i(t)) > 0$ (or $h_i(i(t)) < 0$ resp.), then $C_i(t) = -h_i(i(t))$. More generally, if (t) is on an edge or a corner of K , with $h(i(t))$ pointing “out” of K , then $C(t)$ point inward and takes values in the convex cone generated by the inward normals on the faces impinging on the edge or corner, i.e. $C(t)$ takes its values in $-C(i(t))$.

Concepts of stability of an ODE

As we have seen in the previous section, to study the behaviour of the sequence (Z_n) , we need to study the behaviour of the associated ODE. In this section, we recall some key concepts of the stability of an ODE $\dot{z} = h(z)$. We start by giving the definition of an equilibrium point for the ODE.

Definition 3.5.1. *A state z^* is an equilibrium of the ODE if $h(z^*) = 0$. In other words, this means that once (t) is equal to z^* it remains equal to z^* for all future times.*

To describe the behaviour of the system around the equilibrium, a number of stability concepts are needed. Let us first introduce the basic concepts of stability. To alleviate the notations, we will take 0 as an equilibrium state.

Definition 3.5.2. *The equilibrium $z^* = 0$ is said to be stable, if for any $R > 0$, there exists $r > 0$ such that if $\|z(0)\| < r$, then $\|z(t)\| < R$ for all $t \geq 0$. Otherwise, the equilibrium is unstable.*

Essentially, this means, the system can be kept arbitrarily close to the origin by starting sufficiently close to it. This is also known as *Lyapunov stability*. In some applications, Lyapunov stability is not enough: we not only want the system to remain in a certain range but we also want it to converge to the equilibrium. This behaviour is captured by the concept of asymptotic stability.

Definition 3.5.3. *An equilibrium point $x^* = 0$ is asymptotically stable if it is stable, and if in addition, there exists some $r > 0$ such that $\|x(0)\| < r$ implies that $x(t) \rightarrow 0$ as $t \rightarrow \infty$. The ball B_r is called a domain of attraction of the equilibrium point.*

The above definitions are formulated to characterize the local behaviour of the system, i.e., how the state evolves after starting near the equilibrium point. Local properties tell little about how the system will behave when the initial state is some distance away from the equilibrium. Global concepts are required for this purpose.

Definition 3.5.4. *If asymptotic stability holds for all initial states, the equilibrium is said to be globally asymptotically stable.*

In the case of a linear system, i.e. described by $\dot{x} = Ax$, where A is a $\mathbb{R}^d \times \mathbb{R}^d$ non-singular matrix, the solution is given by: $\forall t \geq 0, x(t) = x(0) \exp(tA)$. Therefore, the stability behaviour of the equilibrium point $x^* = 0$ is stated by the eigenvalues of A . More precisely, the equilibrium point $x^* = 0$ is globally asymptotically stable if and only if all eigenvalues of A have negative real parts. Moreover, if at least one eigenvalue of A has positive real part, then the equilibrium is unstable.

For nonlinear systems, Lyapunov's linearization method states that a nonlinear system should behave similarly to its linearized approximation locally around the equilibrium. For instance, consider the system $\dot{x} = h(x)$ where $h : \mathbb{R}^d \rightarrow \mathbb{R}^d$ is supposed to be continuously differentiable. Then, the system dynamics can be rewritten as:

$$\dot{x} = \frac{\partial h}{\partial x} \Big|_{x=0} x + h_{h.o.t}(x),$$

where $h_{h.o.t}$ stands for higher-order terms in x . Let us denote A the Jacobian matrix of h at 0 , $A = \frac{\partial h}{\partial x} \Big|_{x=0}$. Then, the system $\dot{x} = Ax$ is called the linearization of the original system at the equilibrium point 0 . The following result (see Theorem 3.1 in Slotine and Li [127]) establishes the relationship between the stability of the linear system and that of the original nonlinear system.

Theorem 3.5.5 (Theorem 3.1 in Slotine and Li [127]).

- *If all eigenvalues of A , the Jacobian matrix at 0 , have negative real parts, then the equilibrium point is asymptotically stable for actual nonlinear system.*

- *If at least one eigenvalue of A has positive real part, then the equilibrium is unstable for the nonlinear system.*

The linearization method tells little about the global behaviour of stability of nonlinear systems. This motivates a deeper approach, known as Lyapunov's direct method.

Lyapunov's Direct Method

The intuition behind Lyapunov's direct method is a mathematical extension of a fundamental physical observation: if the total energy of a mechanical or electrical system is continuously dissipated, then the system must eventually settle down to an equilibrium point. The basic procedure of Lyapunov is to generate an *energy-like* scalar function for the system and examine the time variation of that scalar function. This way, we may draw conclusions on the stability of differential equations without using the difficult stability definitions or requiring explicit knowledge of solutions.

The first property that need to be verified by this scalar function is *positive definiteness*.

Definition 3.5.6. *A scalar continuous function $V(\cdot)$ is said to be locally positive definite if $V(0) = 0$ and in around 0, we have, $\forall \epsilon > 0 \Rightarrow V(\cdot) > 0$.*

If the above property holds over the whole state space, then $V(\cdot)$ is said to be globally positive definite.

The above definition implies that the function V has a unique minimum at the origin 0. Actually, given any function having a unique minimum point in a certain ball, we can construct a locally positive definite function by simply adding a constant to that function.

Next, we define the "derivative of V " with respect to time along the system trajectory. Assuming that V is differentiable, this derivative is defined as:

$$\dot{V}(\cdot) = \frac{dV(\cdot)}{dt} = \nabla V \cdot \dot{\cdot} = \nabla V \cdot h(\cdot).$$

Definition 3.5.7. *Let V be a positive definite function and continuously differentiable. If its time derivative along any state trajectory is negative semi-definite, i.e.,*

$$\dot{V}(\cdot) = \nabla V \cdot h(\cdot) \leq 0, \forall \cdot,$$

then V is said to be a Lyapunov function for the system.

Equilibrium Point Theorems

The relations between Lyapunov functions and the stability of systems are made precise in a number of theorems in Lyapunov's direct method. Such theorems usually have local and global

versions. The local versions are concerned with stability properties in the neighborhood of equilibrium point and usually a locally positive definite function. The next theorem (see Theorem 3.2 in Slotine and Li [127]) gives a precise relation between Lyapunov function and stability.

Theorem 3.5.8 (Theorem 3.2 in Slotine and Li [127]). *If, around 0, there exists a scalar function V with continuous derivative such that:*

- V is locally positive definite;
- \dot{V} is locally negative semi-definite.

Then, the equilibrium point 0 is stable. Moreover, if \dot{V} is locally negative definite, then the stability is asymptotic;

The above theorem applies to the local analysis of stability. In order to assess the *global asymptotic stability* of a system, one might expect naturally that the local conditions in the above theorem has to be expanded to the whole state space. This is indeed necessary but not enough. An additional condition on the function V has to be satisfied: V must be coercive. We give more details in the following theorem (See Theorem 3.3 in Slotine and Li [127]).

Theorem 3.5.9 (Theorem 3.3 in Slotine and Li [127]). *Assume that there exists a scalar function V continuously differentiable such that:*

- V is positive definite;
- \dot{V} is negative definite;
- $V(\cdot) \rightarrow \infty$ when $\|\cdot\| \rightarrow \infty$.

Then, the equilibrium at origin is globally asymptotically stable.

Note that the coercive condition along with the negative definiteness of \dot{V} , implies, that given any initial condition x_0 , the trajectories remain in the bounded region defined by $V(\cdot) \leq V(x_0)$.

Invariant Set Theorems

It is important to realize that the theorems in Lyapunov analysis are all sufficiency theorems. If for a particular choice of Lyapunov function candidate V , one of the conditions is not met, one cannot draw any conclusions on the stability of the system. In this kind of situations, fortunately, it is still possible to draw conclusions on asymptotic stability, with the help of the invariant set theorems introduced by La Salle. The central concept in these theorems is that of invariant set, a generalization of the concept of equilibrium point.

Definition 3.5.10. Let (\cdot) be a solution of some ODE. A set G is said to be an invariant set for this ODE if $(0) \in G$ implies that $(t) \in G, \forall t \geq 0$.

For instance, the singleton $\{*\}$ where $*$ is an equilibrium point is an invariant set. Its domain of attraction is also an invariant set. One other trivial invariant set is the whole state-space, $\cup_0\{(t), t > 0, (0) = o\}$. We first discuss the local version of the invariant set theorems as follows (see Theorem 3.4 of Slotine and Li [127]).

Theorem 3.5.11 (see Theorem 3.4 of Slotine and Li [127]). Consider the following ODE: $\dot{x} = h(x)$ and assume that h is continuous. Let V be a scalar function continuously differentiable such that:

- For some $l > 0$, the region $\Omega_l := \{x, V(x) < l\}$ is bounded.
- $\dot{V}(x) \leq 0$ for all $x \in \Omega_l$.

Let R be the set of all points within Ω_l where $\dot{V}(x) = 0$ and M be the largest invariant set in R . Then, every solution (\cdot) originating in Ω_l tends to M as $t \rightarrow \infty$.

Note that La Salle's invariance theorem is only about convergence and not stability. The stability will be guaranteed once the condition of positive definiteness of V is satisfied. However, La Salle's theorem allow us to draw conclusions about the asymptotic behaviour of the system when Lyapunov's direct method cannot be applied.

Corollary 3.5.12. Let V be a scalar function continuously differentiable and assume that in a certain neighborhood Ω of the origin:

- V is locally positive definite;
- \dot{V} is negative semi-definite;
- The largest invariant set in $R := \{x \in \Omega, \dot{V} = 0\}$ is reduced to $\{0\}$.

Then, the equilibrium point 0 is asymptotically stable.

The above corollary replaces the negative definiteness condition on \dot{V} in Lyapunov's local asymptotic stability theorem by a negative semi-definiteness condition on \dot{V} , combined with a third condition on the trajectories within R .

The above invariant set theorem and its corollary can be easily extended to a global result by requiring again the radial unboundedness of the scalar function V .

3.5.2 Closed Formulas for the first example

In this subsection, we give the closed formulas obtained for the optimal risk allocations in the first example when $\alpha > 0$. Recall that $d = 2$ and

$$I(x_1, x_2) = \frac{1}{1 + \alpha} \left(e^{\beta x_1} + e^{\beta x_2} + \alpha e^{\beta(x_1 + x_2)} \right) - \frac{\alpha + 2}{\alpha + 1}$$

and the loss vector X is taken to follow a centered normal distribution with a covariance matrix

$M = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$. The optimal risk allocations m_i^* are characterized by the first order conditions given in Theorem 3.2.6, i.e.

$$\begin{aligned} & \frac{\lambda^*}{1 + \alpha} E \left[e^{\beta(X_i - m_i^*)} + \alpha e^{\beta(X_1 + X_2 - m_1^* - m_2^*)} \right] = 1, \quad i = 1, 2, \\ & \frac{1}{1 + \alpha} E \left[e^{\beta(X_1 - m_1^*)} + e^{\beta(X_2 - m_2^*)} + \alpha e^{\beta(X_1 + X_2 - m_1^* - m_2^*)} \right] - \frac{\alpha + 2}{\alpha + 1} = 0. \end{aligned}$$

The two first equations implies that $E[e^{\beta(X_1 - m_1^*)}] = E[e^{\beta(X_2 - m_2^*)}]$, which in turn gives that,

$$\frac{\beta\sigma_1^2}{2} - m_1^* = \frac{\beta\sigma_2^2}{2} - m_2^*. \tag{3.26}$$

The third equation gives $e^{-\frac{\beta\sigma_1^2}{2} - \beta m_1^*} + e^{-\frac{\beta\sigma_2^2}{2} - \beta m_2^*} + \alpha e^{\beta\left(\frac{\sigma_1\sigma_2}{2} + \frac{\sigma_1^2 + \sigma_2^2}{2} - \beta(m_1 + m_2)\right)} = 2 + \alpha$. Thanks to (3.26) and denoting $Q = e^{-\frac{\beta\sigma_1^2}{2} - \beta m_1^*}$, we get that, $\alpha e^{\beta\left(\frac{\sigma_1\sigma_2}{2} + \frac{\sigma_1^2 + \sigma_2^2}{2} - \beta(m_1 + m_2)\right)} = 2 + \alpha - Q$. Taking the positive solution of the last equation gives $Q = \frac{1 + \alpha(2 + \alpha) - \sqrt{\alpha^2(2 + \alpha)^2 - 4\alpha e^{\beta\left(\frac{\sigma_1\sigma_2}{2} + \frac{\sigma_1^2 + \sigma_2^2}{2} - \beta(m_1 + m_2)\right)}}{2}$. Now, denoting by

$\text{SRC} = -\log(Q)$, we obtain that $m_i^* = \frac{\beta\sigma_i^2}{2} + \frac{1}{\beta} \text{SRC}$.

MULTIVARIATE OPTIMIZED CERTAINTY EQUIVALENT RISK MEASURES AND THEIR NUMERICAL COMPUTATION

Abstract

We present a framework for constructing multivariate risk measures that is inspired from univariate Optimized Certainty Equivalent (OCE) risk measures. We show that this new class of risk measures verifies the desirable properties such as convexity, monotonicity and cash invariance. We also address numerical aspects of their computations using stochastic algorithms instead of using Monte Carlo or Fourier methods that do not provide any error of the estimation.

4.1 Introduction

One of the major concerns in finance is how to assess or quantify the risk associated with a random cashflow in the future. Starting with the pioneering work of Markowitz [94], the risk associated with a random outcome was quantified by its variance. Then, Artzner et al. [5] published their famous seminal paper in which they introduce the theory of risk measures. In their paper, risk measures were defined as a map verifying certain properties, which are called “axioms”, namely: Subadditivity, translation invariance, monotonicity and positive homogeneity. Such risk measures are called *coherent* risk measures. Many extensions have been proposed and studied in the literature after the introduction of the axiomatic approach. One important extension is the notion of *convex risk measure* developed by Föllmer and Schied [57] and Frittelli and Gianin [55] where the subadditivity and positive homogeneity properties were replaced by the weaker property of convexity. The latter reflects the fact that diversification decreases the risk. In the banking industry, one of the most popular risk measures is the Value at Risk (VaR in short). This is due first, to its financial interpretation and second, to its easy and fast implementation. Indeed, VaR is defined as the minimal cash amount that needed to be added to a financial position in order to have a probability of losses below a certain threshold. Its

computation amounts to the calculation of a quantile of the portfolio distribution. Nevertheless, VaR suffers from one drawback: it does not verify the convexity property. This has prompted the search for new examples of risk measures, the most prominent being the Conditional Value at Risk (CVaR), the entropic risk measure and the utility based risk measure (also known as shortfall risk measure).

Some decision making problem based on utility functions are closely related to risk measures. One can cite the optimized certainty equivalent (OCE) that was first introduced by Ben-Tal and Teboulle [11]. The idea behind the definition of OCE is as follows: Assume that a decision maker, with some normalized utility function u (i.e., $u(0) = 0$ and $u'(0) = 1$), is expecting a random income X in the future and can consume a part of it at present. If he chooses to consume m dollars, the resulting present value of X is then $P(X, m) := m + E[u(X - m)]$. This might be very surprising in a first glance because $P(X, m)$ is the sum of cash term and a utility term. However, in Ben-Tal and Teboulle [11], they defined OCE for normalized utilities functions that are twice continuously differentiable. More precisely, in this case, we can show that u can be rewritten as,

$$u(x) = \int_0^x e^{-\int_0^y r(u) du} dy,$$

where $r(x) := -\frac{u''(x)}{u'(x)}$ is the *measure of local risk aversion*. Thanks to this formula, u can be interpreted as a discount function. Moreover, we can see that we no longer have this problem of units since the u is the “sum” (integral) of “discounted” (the exponential term) of infinitesimal wealth dy .

We can define the sure present value of X (i.e., its certainty equivalent) as the result of an optimal allocation of X between present and future consumption, that is the decision maker will try to find m that maximizes $P(X, m)$. The main properties of the OCE were studied in Ben-Tal and Teboulle [10] where it is showed that the opposite of the OCE provides a wide family of risk measures that verifies the axiomatic formalism of convex risk measures. They also proved that several risk measures, such as CVaR and the entropic risk measure, can be derived as special cases of the OCE by using particular utility functions (see also Cherny and Kupper [30]).

From a systemic point of view, the financial crisis of 2008 has demonstrated the need for novel approaches that capture the risk of a system of financial institutions. More precisely, given a network/system of $d \in \mathbb{N}$ different but dependent portfolios $X := (X_1, \dots, X_d)$, we are interested in measuring/quantifying the risk carried by this system of portfolios. A classical approach consists in first aggregating the portfolios using some aggregation function $\Lambda : \mathbb{R}^d \rightarrow \mathbb{R}$ and then apply some univariate risk measure applied to the aggregated portfolio. In practice, most of the times the aggregation function is just the sum of the components, i.e., $\Lambda(x) = \sum_{i=1}^d x_i$. This will result in having a systemic risk measure of the form: $R(X) = \eta(\Lambda(X)) = \eta(\sum_{i=1}^d X_i)$, where η is a univariate risk measure, such as the VaR, CVaR, entropic risk measure, etc. The mech-

anism behind this approach is also known as “Aggregate then Inject Cash” mechanism (see Biagini et al. [15]). However, this approach suffers from one major drawback: While it quantifies the systemic risk carried by the whole system, it does not provide risk levels of each portfolio, and thus, one could not have a ranking of portfolios in terms of their systemic riskiness. One way to remediate to this, is to consider the reverse mechanism, that is to “Inject Cash then Aggregate”. This consists in associating to each portfolio a risk measure and summing up the resulting risk levels. This results in considering systemic risk measures $R(X)$ of the following form: $R(X) = \sum_{i=1}^d \eta_i(X_i)$, where η_i 's are the univariate risk measures associated to each portfolio. Obviously, one could use the same univariate for all portfolios, that is $\eta_i = \eta, \forall i \in \{1, \dots, d\}$. However, by doing so, we are assuming that the system is made of “isolated” portfolios with no interdependence structure, and hence, we might be overestimating or underestimating the systemic risk. This led several authors to look for approaches that address simultaneously the design of an overall risk measure and the allocation of this risk measure among the different components of the system. In this spirit, an extension of shortfall risk measures, introduced in Föllmer and Schied [51], has been studied in Armenti et al. [4] based on multivariate loss functions. However, one should note that, to ensure the existence of optimal allocation problem, these loss functions must verify a key property: permutation invariance. In other words, each component of the system is treated as if it has the same risk profile as all the other components and thus one cannot discriminate a particular component against one another. Moreover, classical risk measures such that the CVaR and the entropic risk measure cannot be recovered using multivariate shortfall risk measures, which limit their use in practice. We will see that, with our multivariate extension of OCE risk measure, the permutation invariance condition is no longer needed and by choosing the appropriate loss functions, we can retrieve most of the classical risk measures.

One of the major issues that arises when studying risk measures is their numerical approximation. The standard VaR can be computed by inverting the simulated empirical distribution of the financial position using Monte Carlo (see Glasserman [58] and Glasserman et al. [59]). An alternative method for computing VaR and CVaR is to use stochastic algorithms (SA). The rational idea behind this perspective comes from the fact that both VaR and CVaR are the solutions and the value of the same convex optimization problem as pointed out in Rockafellar and Uryasev [112] and the fact that the objective function is expressed as an expectation. This was done in Bardou et al. [6], where they prove the consistency and the asymptotic normality of the estimators. In the same direction, in Chapter 3, we extended the work of Dunkel and Weber [43] to approximate multivariate shortfall risk measures using stochastic algorithms. In Neufeld [103], they developed numerical schemes for the computations of univariate OCE using Fourier transform methods.

The outline of this paper is as follows: in section 4.2, we give the definition of multivariate OCE

by introducing first the class of appropriate loss functions. Then, we show that this class of risk measures verifies the desirable properties. We also characterize the optimal solutions, give a dual representation and study the sensitivity with respect to external shocks. Finally, section 4.3 treats the computational aspects of approximating multivariate OCE using a deterministic scheme and a stochastic one.

4.2 Multivariate OCE

Let (Ω, \mathcal{F}, P) a probability space and we denote by $L^0(\mathbb{R}^d)$ the space of \mathcal{F} -measurable random vectors taking values in \mathbb{R}^d . For x, y in \mathbb{R}^d , we denote by $\|\cdot\|$ the Euclidean norm and $x \cdot y = \sum_{i=1}^d x_i y_i$. For a function $f : \mathbb{R}^d \rightarrow [-\infty, \infty]$, we define f^* the convex conjugate of f as $f^*(y) = \sup_x \{x \cdot y - f(x)\}$. The space $L^0(\mathbb{R}^d)$ inherits the lattice structure of \mathbb{R}^d and hence, we can use the classical notations in \mathbb{R}^d in a P -almost-surely sens. We will say for example, for $X, Y \in L^0(\mathbb{R}^d)$ that $X \geq Y$ if $P(X \geq Y) = 1$. To alleviate the notations, we will drop the reference to \mathbb{R}^d in $L^0(\mathbb{R}^d)$ whenever it is unnecessary. For $Q = (Q_1, \dots, Q_d)$ a vector of probabilities, we will write $Q \ll P$ if for all $i = 1, \dots, d$, we have $Q_i \ll P$. In this section, we introduce the notion of multivariate Optimized Certainty Equivalent (OCE) and give its main properties. The latter is an extension of univariate OCE that was introduced and studied in details in Ben-Tal and Teboulle [10]. First, we start by giving the definition of a multivariate loss function that will be used in the rest of the paper. For the rest of the paper, the random vector $X = (X_1, \dots, X_d) \in L^0$ represents profits and losses of d portfolios.

Definition 4.2.1. A function $l : \mathbb{R}^d \rightarrow (-\infty, \infty]$ is called a loss function, if it satisfies the following properties:

1. l is nondecreasing, that is if $x \leq y$ componentwise, then $l(x) \leq l(y)$.
2. l is lower-semicontinuous and convex.
3. $l(0) = 0$ and $l(x) > \sum_{i=1}^d x_i, \forall x \neq 0$.

For integrability reasons, we will work in the multivariate Orlicz heart defined as:

$$M^\vartheta := \{X \in L^0 : E[\vartheta(\lambda X)] < \infty, \forall \lambda > 0\},$$

where $\vartheta(x) = l(|x|), x \in \mathbb{R}^d$. On this space, we define the Luxembourg norm as:

$$\|X\|_\vartheta := \inf_{\lambda > 0} \lambda^{-1} E[\vartheta(\lambda X)] \leq 1.$$

Under the Luxembourg norm, M^ϑ is a Banach lattice and its dual with respect to this norm is given by the Orlicz space L^{ϑ^*} :

$$L^{\vartheta^*} := \{X \in L^0, E[\vartheta^*(\lambda X)] < \infty, \text{ for some } \lambda > 0\}.$$

We also introduce the set of d -dimensional measure densities in L^{ϑ^*} , that is:

$$Q^{\vartheta^*} := \frac{dQ}{dP} := (Z_1, \dots, Z_d), Z \in L^{\vartheta^*}, Z_k \geq 0 \text{ and } E[Z_k] = 1 \quad .$$

Note that for $Q \in Q^{\vartheta^*}$ and $X \in M^\vartheta$, $\frac{dQ}{dP} \cdot X \in L^1$, thanks to Fenchel inequality and for the sake of simplicity, we will write $E_Q[X] := E[dQ/dP \cdot X]$. We refer to Appendix B in Armenti et al. [4] for more details about multivariate Orlicz spaces.

Definition 4.2.2. Assume l is a loss function. The multivariate OCE risk measure is defined for every $X \in M^\vartheta$ as:

$$R(X) = \inf_{w \in \mathbb{R}^d} \sum_{i=1}^d w_i + E[l(-X - w)] \quad . \quad (4.1)$$

Example 4.2.1. When $d = 1$, we can recover some important convex risk measures such CVaR (also called Expected Shortfall or Average Value at Risk) and Entropic risk measure.

1. CVaR: Let $\alpha \in (0, 1)$ and take $l(x) = \frac{1}{1-\alpha}x^+$, then the associated risk measure is the CVaR (see Rockafellar and Uryasev [112]).
2. Polynomial loss function: For an integer $\gamma > 1$, the polynomial loss function is defined by: $l(x) = \frac{([1+x]^+)^{\gamma} - 1}{\gamma}$. When $\gamma = 2$, the corresponding risk measure is the Monotone Mean-Variance (see Černý et al. [25]).
3. Entropic risk measure: Fix $\lambda > 0$ and let $l(x) := \frac{\exp(\lambda x) - 1}{\lambda}$. Then, the problem in (4.1) can be explicitly solved and the optimal w^* and $R(X)$ are given by:

$$w^* = \frac{1}{\lambda} \log(E[e^{-\lambda X}]), \quad R(X) = w^* = \frac{1}{\lambda} \log(E[e^{-\lambda X}]).$$

Using univariate loss functions, we can construct multivariate loss functions in the following way: Given l_1, \dots, l_d univariate loss functions and a nonnegative, convex and lower-semicontinuous function $\Lambda : \mathbb{R} \rightarrow \mathbb{R}^+$ with $\Lambda(0) = 0$, one can define a multivariate loss function as follows:

$$l(x) := \sum_{i=1}^d l_i(x_i) + \Lambda(x). \quad (4.2)$$

It is easy to see that l verifies all the conditions in the definition 4.2.1. Note that by taking Λ the null function, the corresponding multivariate OCE boils down to a sum of univariate OCE.

It is in this function Λ where the dependence between the different components in the system is taken into account. In this paper, we will focus on the following multivariate loss functions inspired from the univariate risk measures above:

$$I(x) = \sum_{i=1}^d \frac{e^{\lambda_i x_i} - 1}{\lambda_i} + \alpha e^{-\sum_{i=1}^d \lambda_i x_i}, \quad \lambda_i > 0, \quad \alpha \geq 0, \quad (4.3)$$

$$I(x) = \sum_{i=1}^d \frac{([1 + x_i]^{\vartheta_i} - 1)}{\vartheta_i} + \alpha \sum_{i < j} \frac{([1 + x_i]^{\vartheta_i} - 1)([1 + x_j]^{\vartheta_j} - 1)}{\vartheta_i \vartheta_j}, \quad \vartheta_i > 1, \quad \alpha \geq 0, \quad (4.4)$$

$$I(x) = \sum_{i=1}^d \frac{x_i^+}{1 - \beta_i} + \alpha \sum_{i < j} \frac{x_i^+ x_j^+}{1 - \beta_i 1 - \beta_j}, \quad 0 < \beta_i < 1, \quad \alpha \geq 0. \quad (4.5)$$

In the next theorem, we show that the multivariate OCE is a convex risk measure as defined in Föllmer and Schied [51].

Theorem 4.2.3. *The function R in (4.1) is real valued, convex, monotone and cash invariant¹ risk measure. In particular, it is continuous and subdifferentiable. If I is positive homogeneous, then R is too. Furthermore, it admits the following representation:*

$$R(X) = \max_{Q \in \mathcal{D}^{I^*}} \{E_Q[-X] - \alpha(Q)\}, \quad (4.6)$$

where the penalty function α is defined for $Q = (Q_1, \dots, Q_d) \ll P$ by: $\alpha(Q) = E^h \int^i \frac{dQ}{dP}$ and $\mathcal{D}^{I^*} = \{Q \ll P, \alpha(Q) < \infty\} := \text{dom}(\alpha)$.

Proof.

- $R(X) \in \mathbb{R}$ for all $X \in M^\vartheta$: Since $M^\vartheta \subseteq L^1$, by the third property of loss functions, we have for every $X \in M^\vartheta$ and $w \in \mathbb{R}^d$: $\sum_{i=1}^d w_i + E[I(-X - w)] \geq E[-\text{sum}_{i=1}^d X_i] > -\infty$. $R(X) < +\infty$ since for $w = 0$, we have $E[I(-X)] < \infty$.
- **Monotonicity:** Let $X, Y \in M^\vartheta$ such that $X \leq Y$. Since I is non-decreasing, then $E[I(-X - w)] \geq E[I(-Y - w)]$ for every $w \in \mathbb{R}^d$, which in turn implies $R(X) \geq R(Y)$.

1. In the following sens: $R(X + m) = R(X) - \sum_{i=1}^d m_i$

- Convexity: Let $X, Y \in M^\vartheta$ and $\lambda \in [0, 1]$. We have thanks to the convexity of I :

$$\begin{aligned}
 R(\lambda X + (1 - \lambda)Y) &= \inf_{w \in \mathbb{R}^d} \left\{ \sum_{i=1}^d w_i + E[I(-X - w)] \right\} \\
 &= \inf_{w^1, w^2 \in \mathbb{R}^d} \left(\sum_{i=1}^d \lambda w_i^1 + (1 - \lambda) w_i^2 + E[I(\lambda(-X - w^1) + (1 - \lambda)(-Y - w^2))] \right) \\
 &\leq \inf_{w^1 \in \mathbb{R}^d} \inf_{w^2 \in \mathbb{R}^d} \left(\lambda \left(\sum_{i=1}^d w_i^1 + E[I((-X - w^1)] \right) + (1 - \lambda) \left(\sum_{i=1}^d w_i^2 + E[I((-Y - w^2)] \right) \right) \\
 &= \lambda R(X) + (1 - \lambda)R(Y).
 \end{aligned}$$

- Cash Invariance: Let $m \in \mathbb{R}^d$, we have:

$$\begin{aligned}
 R(X + m) &= \inf_{w \in \mathbb{R}^d} \left(\sum_{i=1}^d w_i + E[I(-X - m - w)] \right) \\
 &= \inf_{w \in \mathbb{R}^d} \left(\sum_{i=1}^d (w_i + m_i) - \sum_{i=1}^d m_i + E[I(-X - m - w)] \right) \\
 &= R(X) - \sum_{i=1}^d m_i
 \end{aligned}$$

- Continuity and subdifferentiability: Since $(M^\vartheta, \|\cdot\|_\vartheta)$ is a Banach space, this is a direct consequence of Theorem 4.1 in Cheridito and Li [29] or Theorem 1 in Biagini and Frittelli [16].
- Positive homogeneity: If I is positive homogeneous, then by the definition of $R(X)$, we have for $\lambda > 0$:

$$\begin{aligned}
 R(\lambda X) &= \inf_{w \in \mathbb{R}^d} \left(\sum_{i=1}^d w_i + E[I(-\lambda X - w)] \right) \\
 &= \inf_{w \in \mathbb{R}^d} \left(\sum_{i=1}^d w_i + \lambda E[I(-X - \frac{w}{\lambda})] \right) \\
 &= \lambda \inf_{w \in \mathbb{R}^d} \left(\sum_{i=1}^d \frac{w_i}{\lambda} + E[I(-X - \frac{w}{\lambda})] \right) \\
 &= \lambda \inf_{w \in \mathbb{R}^d} \left(\sum_{i=1}^d w_i + E[I(-X - w)] \right) = \lambda R(X).
 \end{aligned}$$

- Representation: First, because R is convex and continuous, Fenchel-Moreau theorem

implies that:

$$R(X) = \sup_{Y \in L^{\theta^*}} \{E[X \cdot Y] - R^*(Y)\} = \max_{Y \in \mathcal{D}^{\theta^*}} \{E[X \cdot Y] - R^*(Y)\} \quad (4.7)$$

where $R^*(Y) = \sup\{E[X \cdot Y] - R(X), X \in M^{\theta}\}$, $Y \in L^{\theta^*}$. Now, if $Y \not\leq 0$, then by the bipolar theorem, there exists $X_1 \in M^{\theta}$ such that $X_1 \geq 0$ and $E[X_1 \cdot Y] > 0$. Using the definition of $R^*(Y)$, we get the following:

$$\begin{aligned} R^*(Y) &= \sup_{X \in M^{\theta}} \{E[X \cdot Y] - R(X)\} \\ &\geq \sup_{\lambda > 0} \{\lambda E[X_1 Y] - R(\lambda X_1)\} \\ &\geq \sup_{\lambda > 0} \{\lambda E[X_1 Y]\} - R(0) = +\infty, \end{aligned}$$

where the last inequality is due to the monotonicity of R . Therefore, the maximum can be taken over $Y \leq 0$. For $k \in \{1, \dots, d\}$, let $X = (0, \dots, x, \dots)$ and $x > 0$. By the translation invariance property, we have $R(X_k) = R(0) - x$. Consequently,

$$\begin{aligned} R^*(Y) &= \sup_{X \in M^{\theta}} \{E[X \cdot Y] - R(X)\} \\ &\geq xE[Y_k] - R(0) + x = x(E[Y_k] + 1) - R(0). \end{aligned}$$

If $E[Y_k] < -1$, then by sending x to infinity, we get that $R^*(Y) = \infty$. Finally, this shows that the maximum in (4.7) could be taken over \mathcal{D}^{θ^*} , i.e., $R(X) = \max_{Q \in \mathcal{D}^{\theta^*}} \{E[-\frac{dQ}{dP} \cdot X] -$

$R^*(-Q)\} = \max_{Q \in \mathcal{D}^{\theta^*}} \{E_Q[-X] - R^*(-Q)\}$. Let us now explicit more the expression of $R^*(-Q)$

for $Q \in D^{\vartheta^*}$:

$$\begin{aligned}
 R^*(-Q) &= \sup_{X \in M^{\vartheta}} E \left[-\frac{dQ}{dP} \cdot X \right] - R(X) \\
 &= \sup_{X \in M^{\vartheta}} E \left[-\frac{dQ}{dP} \cdot X \right] - \inf_{m \in \mathbb{R}^d} \sum_{i=1}^d m_i + E[I(-X - m)] \\
 &= \sup_{X \in M^{\vartheta}} E \left[-\frac{dQ}{dP} \cdot X \right] + \sup_{m \in \mathbb{R}^d} \left[-\sum_{i=1}^d m_i - E[I(-X - m)] \right] \\
 &= \sup_{X \in M^{\vartheta}} \sup_{m \in \mathbb{R}^d} E \left[-\frac{dQ}{dP} \cdot X \right] - \sum_{i=1}^d m_i - E[I(-X - m)] \\
 &= \sup_{m \in \mathbb{R}^d} \left[-\sum_{i=1}^d m_i + \sup_{X \in M^{\vartheta}} E \left[-\frac{dQ}{dP} \cdot X \right] - E[I(-X - m)] \right] \\
 &= \sup_{m \in \mathbb{R}^d} \left[-\sum_{i=1}^d m_i + \sup_{W \in M^{\vartheta}} E \left[-\frac{dQ}{dP} \cdot (W - m) \right] - E[I(-W)] \right] \\
 &= \sup_{m \in \mathbb{R}^d} \left[-\sum_{i=1}^d m_i + E \left[\frac{dQ}{dP} \cdot m \right] + \sup_{W \in M^{\vartheta}} E \left[-\frac{dQ}{dP} \cdot W \right] - E[I(-W)] \right] \\
 &= \sup_{m \in \mathbb{R}^d} \left[-\sum_{i=1}^d m_i + m_i E \left[\frac{dQ_i}{dP} \right] + \sup_{W \in M^{\vartheta}} E \left[\frac{dQ}{dP} \cdot W \right] - E[I(W)] \right] \\
 &= \sup_{m \in \mathbb{R}^d} \left[0 + \sup_{W \in M^{\vartheta}} E \left[\frac{dQ}{dP} \cdot W \right] - I(W) \right] \\
 &= \sup_{W \in M^{\vartheta}} E \left[\frac{dQ}{dP} \cdot W \right] - I(W)
 \end{aligned}$$

Note that, for $W \in M^{\vartheta}$, we have for $Q \in D^{\vartheta^*}$, $\frac{dQ}{dP} \cdot W \in L^1$, thanks to Fenchel inequality. Furthermore, since $\sum W_i \leq I(W) \leq \vartheta(W)$ and both $\vartheta(W)$ and $\sum W_i$ are in L^1 , we have $I(W) \in L^1$. This allows us to write in the lines above $E \left[\frac{dQ}{dP} \cdot W \right] - E[I(W)] = E \left[\frac{dQ}{dP} \cdot W \right] - I(W)$.

Now, we would like to interchange the expectation with the supremum. To this end, we use Corollary on page 534 of Rockafellar [114] with $L = M^{\vartheta}$, $L^* = L^{\vartheta^*}$ and $F(x) = I(x)$. Note that I is a lower-semicontinuous proper convex function, and it is easy to verify that M^{ϑ} and L^{ϑ^*} are decomposable in their sens, so that all the conditions needed to apply this Corollary are satisfied. We get finally that,

$$R^*(-Q) = E \left[I^* \left(\frac{dQ}{dP} \right) \right] := \alpha(Q).$$

Finally, since $R(X)$ is finite, then the maximum can be taken over D^{I^*} instead of D^{θ^*} .

□

Definition 4.2.4. A risk allocation is any minimizer of (4.1). When it is uniquely determined, we denote it $RA(X)$.

Theorem 4.2.5. Let l be a loss function. Then, for every $X \in M^\theta$, the set of risk allocations is non empty and bounded. Furthermore, risk allocations are characterized by the following first order condition:

$$1 \in E[\partial l(-X - m^*)]. \quad (4.8)$$

Moreover, the supremum in (4.6) is attained for Z^* such that $Z^* \in \partial l(-X - m^*)$ a.s. and $E[Z^*] = 1$.

Proof. The arguments used in this proof are an extension of the univariate case. To prove that the set of risk allocations is non empty and bounded, it is sufficient to show that the objective function has no direction of recession thanks to Theorem 27.1(d) in Rockafellar [115]. Let $w \neq 0$ and let $f(w) := \sum_{i=1}^d w_i + E[l(-X - w)]$. We have,

$$\begin{aligned} f_0^+(w) &= \lim_{r \rightarrow \infty} \frac{f(m + rw) - f(m)}{r} \\ &= \lim_{r \rightarrow \infty} \frac{\sum m_i + r \sum w_i + E[l(-X - m - rw)] - \sum m_i - E[l(-X - m)]}{r} \\ &= \sum w_i + \lim_{r \rightarrow \infty} \frac{E[l(-X - m - rw)] - E[l(-X - m)]}{r} \\ &= \sum w_i + \lim_{r \rightarrow \infty} \frac{E[l(-X - m - rw)]}{r}. \end{aligned}$$

Now, since l is convex and $l(0) = 0$, for $\lambda > 1$ we have $\frac{1}{\lambda}l(\lambda x) \geq l(x)$. This implies, together with Lebesgue's dominated convergence theorem and lower-semicontinuity of l

$$\begin{aligned} f_0^+(w) &\geq \sum w_i + \lim_{r \rightarrow \infty} E \left[l \left(\frac{-X - m}{r} - w \right) \right] \\ &= \sum w_i + E \left[\liminf_{r \rightarrow \infty} l \left(\frac{-X - m}{r} - w \right) \right] \\ &\geq \sum w_i + l(-w) > 0. \end{aligned}$$

The last strict inequality is a consequence of the third property of l . So we have shown that for every $w \neq 0$, $f_0^+(w) > 0$, i.e., f has no direction of recession. We conclude that the set of minimizers is non empty bounded set. Moreover, we have $m^* \in \operatorname{argmin} f$ if and only if m^* satisfies $0 \in \partial f(m^*)$. Using Theorem 4.47 in Shapiro et al. [124], we can interchange the partial

operator and the expectation sign leading to the following characterization of minimizers:

$$m^* \text{ is a minimizer of } f \Leftrightarrow 1 \in E[\partial l(-X - m^*)].$$

In the following, we prove that the maximum in (4.6) is attained for $Z^* \in \partial l(-X - m^*)$ a.s. and $E[Z^*] = 1$. We start by proving the existence of such Z^* . Let m^* be such that $1 \in E[\partial l(-X - m^*)]$. Note that, for each $v \in \partial l(\cdot)$, if $v \in \partial l(\cdot)$, then v is nonnegative. In fact, by definition, we have, $l(x) \geq l(\cdot) + \sum_{i=1}^d v_i(x_i - \cdot_i), \forall x \in \mathbb{R}^d$. So, if for some $k \in 1, \dots, d, v_k < 0$, then choosing $x = \cdot - ne_k < \cdot$ where e_k is the k -th standard unit vector, we get that $-nv_k \leq l(x) - l(\cdot) \leq 0$. By sending n to $+\infty$, we get a contradiction. Therefore, since $1 \in E[\partial l(-X - m^*)]$, there exists a random variable Z^* such that $Z^* \geq 0$ and $Z^* \in \partial l(-X - m^*)$ a.s. and $E[Z^*] = 1$.

Next, we will show that $Z^* \in D^{\theta^*}$, that is $E[l^*(Z^*)] < \infty$. Note that since $Z^* \in \partial l(-X - m^*)$, we have that,

$$l^*(Z^*) = Z^* \cdot (-X - m^*) - l(-X - m^*), \text{ a.s.} \tag{4.9}$$

First, we will start by proving that $Z^* \cdot X \in L^1$. Thanks to (4.9), we have $X \cdot Z^* + l^*(Z^*) = -m^* \cdot Z - l(-X - m^*)$. Because $X \in M^\theta$, the right term of the previous equality is in L^1 . So, this shows that $X \cdot Z^* + l^*(Z^*) \in L^1$. Recall that $l^*(\cdot) \geq 0$ for all $\cdot \in \mathbb{R}^d$ so that we have $(X \cdot Z^*)^+ \in L^1$. It remains to show that $(X \cdot Z^*)^- \in L^1$. Using the convexity of l , we have the following inequality:

$$l(2(-X - m^*)) \geq l(-X - m^*) + Z^* \cdot (-X - m^*), \text{ a.s.}$$

This in turn implies that $X \cdot Z \geq l(-X - m^*) - l(2(-X - m^*)) - Z^* \cdot m^*$. The RHS of this inequality is in L^1 as $X \in M^\theta$. Hence, we get that $(Z^* \cdot X)^- \in L^1$. We are now able to say that all the terms in the RHS of (4.9) are in L^1 . We conclude that $l^*(Z^*) \in L^1$. Moreover, we have,

$$\begin{aligned} E[-X \cdot Z^*] - E[l^*(Z^*)] &= E[-X \cdot Z^*] - E[Z^* \cdot (-X - m^*) - l(-X - m^*)] \\ &= E[-X \cdot Z^* - Z^* \cdot (-X - m^*) + l(-X - m^*)] = E[Z^* \cdot m^* + l(-X - m^*)] \\ &= R(X), \end{aligned}$$

where we used the optimality of m^* in the last equality. This completes the proof. □

Example 4.2.2. *The following example with a bidimensional loss function of exponential type as in Example 4.2.1, that is:*

$$l(x_1, x_2) = \frac{e^{\lambda_1 x_1} - 1}{\lambda_1} + \frac{e^{\lambda_2 x_2} - 1}{\lambda_2} + \alpha e^{\lambda_1 x_1 + \lambda_2 x_2}, \text{ where } \lambda_1 > 0, \lambda_2 > 0, \alpha \geq 0.$$

If $X \sim N(0, \Sigma)$ with $\Sigma = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$, then we can solve explicitly the optimal risk allocations in (4.8) and to obtain

$$m_j^* = \begin{cases} \frac{\lambda_j \sigma_j^2}{2}, & \text{if } \alpha = 0, \\ \frac{\lambda_j \sigma_j^2}{2} - \frac{1}{\lambda_j} \ln(SC_{ij}), & j \neq i, \text{ if } \alpha > 0, \end{cases} \quad (4.10)$$

where the term SC_{ij} , $i \neq j$, is the positive solution to the following second order equation:

$$\alpha \lambda_j \exp(\rho \sigma_i \sigma_j \lambda_i \lambda_j) X^2 + (1 + \alpha(\lambda_i - \lambda_j) \exp(\rho \sigma_i \sigma_j \lambda_i \lambda_j)) X - 1 = 0$$

The risk measure could also be derived in explicit form:

$$R(X) = m_1^* + m_2^* + \frac{2-\alpha}{\lambda_1} (SC_{12} - 1) = m_1^* + m_2^* + \frac{2-\alpha}{\lambda_2} (SC_{21} - 1), \quad (4.11)$$

Remark 4.2.6. The formula obtained in (4.10) is close to the one in Example 3.12 in Armenti et al. [4]. It shows that the optimal allocations are disentangled into two components: the first one is an individual contribution which takes the form of the entropic risk measure of X_i and the second one is a systemic contribution which involves correlations between the two components of the system. This formula shows also an interesting feature: the partial differential of SRC with respect to ρ is always positive. This can be interpreted in the following way: the more correlated the system is, the riskier is. Note that this is not true in general and depends on the loss function l used.

Corollary 4.2.7. Let l a strictly convex loss function. Then,

$$RA(X + r) = RA(X) - \sum_{i=1}^d r_i, \text{ for every } X \in M^\vartheta \text{ and } r \in \mathbb{R}^d.$$

If l is additionally positive homogeneous, then

$$RA(\lambda X) = \lambda RA(X), \text{ for every } X \in M^\vartheta \text{ and } \lambda > 0.$$

Proof. Let $X \in M^\vartheta$ and $r \in \mathbb{R}^d$. $m : RA(X+r)$ is the unique solution of $1 \in E[\partial l(-X-r-m)]$. Setting $w = r+m$, we obtain that w satisfies $1 \in E[\partial l(-X-w)]$, which by uniqueness implies

that $w = RA(X)$, that is $RA(X + r) = RA(X) - r$. Let $\lambda > 0$, we have,

$$\begin{aligned}
 RA(\lambda X) &= \operatorname{argmin}_w \left(\sum_{i=1}^d w_i + E[I(-\lambda X - w)] \right) \\
 &= \operatorname{argmin}_w \left(\sum_{i=1}^d w_i + \lambda E \left(I \left(-X - \frac{w}{\lambda} \right) \right) \right) \\
 &= \operatorname{argmin}_w \left(\sum_{i=1}^d \frac{w_i}{\lambda} + \lambda E \left(I \left(-X - \frac{w}{\lambda} \right) \right) \right) \\
 &= \lambda \operatorname{argmin}_w \left(\sum_{i=1}^d w_i + E[I(-X - w)] \right) = \lambda RA(X).
 \end{aligned}$$

□

Now, we focus on the study of the sensitivity of our multivariate risk measure. We first give the definition of the marginal risk contribution of $Y \in M^\vartheta$ to $X \in M^\vartheta$.

Definition 4.2.8. For $X, Y \in M^\vartheta$, we define the marginal risk contribution of Y to X as the sensitivity of the risk associated to X when an impact Y is applied as

$$R(X, Y) := \limsup_{\epsilon \searrow 0} \frac{R(X + \epsilon Y) - R(X)}{\epsilon}. \tag{4.12}$$

If $R(X + \epsilon Y)$ admits a unique risk allocation $RA(X + \epsilon Y)$ for small enough $\epsilon \geq 0$, then we define the risk allocation marginals of X with respect to the impact of Y as:

$$RA_i(X; Y) := \limsup_{\epsilon \searrow 0} \frac{RA_i(X + \epsilon Y) - RA_i(X)}{\epsilon}, \quad i = 1, \dots, d. \tag{4.13}$$

Theorem 4.2.9. Let $X, Y \in M^\vartheta$ and assume that I is differentiable. Then,

$$R(X, Y) = \sum_{i=1}^d \frac{\partial}{\partial p_i} \left(-E[O(-X - m^*)] \right) [Y^i], \tag{4.14}$$

where m^* is such that, $E[OI(-X - m^*)] = 1$, i.e. an infimum for (4.1) and $\frac{dO^*}{dp} := OI(-X - m^*)$. If furthermore, I is twice differentiable such that we can interchange the differentiation and expectation of $m \mapsto E[OI(-X - m)]$ and $M := E[O^2I(-X - m^*)]$ is invertible, then we have,

- There exists a unique m_ϵ optimum of $R(X + \epsilon Y)$ for small enough $\epsilon \geq 0$.
- As a function of ϵ , m_ϵ is differentiable and we have

$$RA(X, Y) = M^{-1}V, \quad V := -E[O^2I(-X - m^*)Y]. \tag{4.15}$$

Proof. Take $X, Y \in M^\theta$ and let m^* be an infimum for $R(X)$. We have $R(X) = \sum_{i=1}^d m_i^* + E[l(-X - m^*)]$ and $E[Ol(-X - m^*)] = 1$. By the definition of $R(X + \epsilon Y)$, we have

$$\begin{aligned} \frac{R(X + \epsilon Y) - R(X)}{\epsilon} &\leq \frac{\sum m_i^* + E[l(-X - \epsilon Y - m^*)] - \sum m_i^* - E[l(-X - m^*)]}{\epsilon} \\ &= E \frac{l(-X - \epsilon Y - m^*) - l(-X - m^*)}{\epsilon} . \end{aligned}$$

Using the convexity, monotonicity of l and the fact that $-l(x) \geq -\sum_{i=1}^d x_i$, $x \in R^d$, for $0 < \epsilon < \frac{1}{2}$, we get that,

$$\begin{aligned} \frac{l(-X - \epsilon Y - m^*) - l(-X - m^*)}{\epsilon} &\leq \frac{l(-X - m^* - (1 - \epsilon)Y) - l(-X - m^*)}{1 - \epsilon} \\ &\leq \frac{l(|X| + |m^*| + (1 - \epsilon)|Y|) - l(-X - m^*)}{1 - \epsilon} \\ &\leq 2 \cdot l(|X| + |m^*| + |Y|) + \sum_{i=1}^d X_i + m_i . \end{aligned}$$

Since X and Y are in M^θ , the last term is bounded from above by a random variable which is in L^1 . Therefore, using Fatou's lemma, we obtain that,

$$\limsup_{\epsilon \searrow 0} \frac{R(X + \epsilon Y) - R(X)}{\epsilon} \leq E[-Y \cdot Ol(-X - m^*)].$$

Now, using the representation given in Theorem 4.2.5 $R(X + \epsilon Y) = \max_{Q \in D^{r^*}} E_Q[-(X + \epsilon Y)] - E[l^*(\frac{dQ}{dP})]$, and that $R(X) = E_{Q^*}[-X] - E[l^*(\frac{dQ^*}{dP})]$ with $\frac{dQ^*}{dP} = Ol(-X - m^*)$, we get,

$$R(X + \epsilon Y) \geq E[-(X + \epsilon Y) \cdot \frac{dQ^*}{dP} - l^*(\frac{dQ^*}{dP})] = R(X) - \epsilon E[Y \cdot Ol(-X - m^*)],$$

Consequently, the other inequality follows:

$$\limsup_{\epsilon \searrow 0} \frac{R(X + \epsilon Y) - R(X)}{\epsilon} \geq -E[Y \cdot Ol(-X - m^*)].$$

Second assertion is a direct application of Theorem 6 pp 34 in Fiacco and McCormick [48]. \square

In the following Corollary, we explicit the impact of an independent exogenous shock in the case X and Y are independent.

Corollary 4.2.10. *If X and Y are independent, then under assumptions of Theorem 4.2.9, we have,*

$$R(X, Y) = -\sum_{i=1}^d E[Y_i], \quad RA(X, Y) = -E[Y]. \tag{4.16}$$

Remark 4.2.11.

1. The equations in (4.14) and (4.15) are very interesting and show the relevance of the dual optimizer Q_* . More precisely, (4.14) shows that the marginal risk contribution can be quantified thanks to the optimal probability Q_* .
2. If only the value of portfolio i changes by a cash amount, that is $Y^i = c^i$ and $Y^j = 0$ for $j \neq i$, then the marginal risk contribution $R(X, Y) = -c^i$ is exactly covered by the marginal risk allocation $RA_i(X, Y) = -c^i$ of portfolio i , whereas marginal risk allocations of other portfolios remain unchanged, i.e. $RA_j(X, Y) = 0$ for $j \neq i$. This property of full responsibility for one's own changes in financial position is known as causal responsibility (see Brunnermeier and Cheridito [21]). In general, this is no longer true if Y^i is a random variable, but in the particular case when Y^i is independent of X , this property remains true as suggested by equation (4.16).
3. Equation (4.16) shows an interesting feature: Assume that two institutions $i \neq j$ change their positions in opposite direction, that is $Y^i = -Y^j$, then the marginal risk contribution is zero, as if the portfolios compensate each other and a risk sharing mechanism take place.

Example 4.2.3. In this example, we illustrate the impact of an exogenous shock that may depend on X . More specifically, we consider a system with two portfolios $X = (X_1, X_2)$, an exogenous shock $Y = (Y_1, 0)$ impacting the first component only and a loss function of exponential type as in (4.3):

$$l(x_1, x_2) = \frac{e^{\lambda_1 x_1} - 1}{\lambda_1} + \frac{e^{\lambda_2 x_2} - 1}{\lambda_2} + \alpha e^{\lambda_1 x_1 + \lambda_2 x_2}.$$

As per Theorem 4.2.5, there exists a unique risk allocations m^* . To alleviate the expressions, we denote the following:

$$\begin{aligned} C_{X_1} &:= E[e^{\lambda_1(-X_1 - m_1^*)}], & C_{X_2} &:= E[e^{\lambda_2(-X_2 - m_2^*)}], \\ C_X &:= E[e^{\lambda_1(-X_1 - m_1^*) + \lambda_2(-X_2 - m_2^*)}], \\ C_{X_1 Y} &:= E[Y_1 e^{\lambda_1(-X_1 - m_1^*)}], & C_{X_2 Y} &:= E[Y_1 e^{\lambda_2(-X_2 - m_2^*)}], \\ &\vdots \\ C_{XY} &:= E[Y_1 e^{\lambda_1(-X_1 - m_1^*) + \lambda_2(-X_2 - m_2^*)}]. \end{aligned}$$

The matrix M and vector V in Theorem 4.2.9 can be expressed thanks to the quantities above after some simple but lengthy computations (omitted here):

$$M = \begin{pmatrix} \lambda_1 C_{X_1} + \alpha \lambda_1^2 C_X & \alpha \lambda_1 \lambda_2 C_X \\ \alpha \lambda_1 \lambda_2 C_X & \lambda_2 C_{X_2} + \alpha \lambda_2^2 C_X \end{pmatrix}, \quad V = \begin{pmatrix} -\lambda_1 C_{X_1 Y} - \alpha \lambda_1^2 C_{XY} \\ -\alpha \lambda_1 \lambda_2 C_{XY} \end{pmatrix}.$$

The risk contribution marginal and risk allocations marginals follows:

$$R(X, Y) = -C_{X_1 Y} - \alpha \lambda_1 C_{XY}, \tag{4.17}$$

$$\begin{aligned} (\quad) = \text{Common Factor} & \quad -C_{X_2 C_{X_1 Y}} \frac{\alpha (\lambda_1 C_{X_2 C_{XY}} + \lambda_2 C_X C_{X_1 Y})}{RA_{X, Y}} \\ & \quad -\alpha (\lambda_1 C_{X_1 C_{XY}} - \lambda_1 C_X C_{X_1 Y}). \end{aligned} \tag{4.18}$$

We notice the following:

- $R(X, Y)$ is disentangled into two components. The first one is the contribution of the first component in the risk contribution marginal and the second is a systemic contribution that is proportional to α . This same remark holds for $RA_1(X, Y)$.
- The asymmetry of the shock on X_1 can be seen in the systemic contribution in $RA_1(X, Y)$ and $RA_2(X, Y)$. Indeed, we notice that both components are impacted by the shock and this is reflected by the term $-\alpha \lambda_1 C_{X_2} C_X Y$ for the first component and $-\alpha \lambda_1 C_{X_1} C_X Y$ for the second. However, there is a correction term proportional to λ_2 that is subtracted from the first component whereas another correction term proportional to λ_1 is added to the second component.
- In the case $\alpha = 0$, i.e. without a systemic component, the risk marginal of the second portfolio is zero. This something we would expect as we applied a shock only on the first component. In other words, the first component takes full responsibility in this case.

In the rest of the paper, for every $X \in M^\theta$, we will assume the following:

- (ffi) i. For every $m_0 \in \mathbb{R}^d$, $m \mapsto I(-X - m)$ is differentiable at m_0 a.s.;
- ii. $m \mapsto E[I(-X - m)]$ is strictly convex.

Under assumption (ffi), there exists a unique risk allocation m^* that is characterized through the following equation:

$$1 = E[OI(-X - m^*)].$$

4.3 Computational aspects

In this section, we develop numerical schemes to compute the optimal risk allocations m^* and $R(X)$ using stochastic algorithms (SA). This is because the optimal allocations are solutions of a convex optimization problem whose objective function can be expressed as an expectation. Stochastic algorithms are generally used to find zeros of a certain function h that is unknown but could be approximated using some estimate. More specifically, SA algorithms take the following form: $Z_{n+1} = Z_n \pm \gamma_n Y_n$, where Y_n is a noisy estimate of $h(Z_n)$ and (γ_n) is a step

sequence decreasing toward 0. This algorithm is known as Robbins-Monro algorithm (RM). For an overview of SA algorithms, we refer to Duflo [42]. However, in order to be able to use classical convergence results of SA, we need a sub-linear growth over the function h (see for example condition (8) of Theorem 2.2 in Bardou et al. [6]), which in our case, considerably limits the choice of loss functions. To circumvent this condition, we will use a “constrained” variant where we force the iterations of the (RM) algorithm to remain in a certain compact K set that contains the optimal allocations. One could also use the well-known projection “à la Chen” algorithm based on reinitializations of the algorithm and taking larger compact sets each time the iteration goes out of the compact set (cf. Chen and Zhu [28]). For the sake of simplicity, we will use the classical “constrained” version with a fixed compact set K as it has the same asymptotic behaviour as the one with projection “à la Chen”. In Armenti et al. [4], numerical schemes were developed to find optimal allocations for multivariate shortfall risk measures. They first estimated the different expectations using Monte Carlo/Fourier methods and then a root finding algorithm was used to find the optimum. Although this method shows good results of convergence and is quite fast, it has several drawbacks: It is sensitive to the starting point of the root finding algorithm and one has no control over the error of estimation. With SA, there is one major advantage over the former method: One could derive Central Limits Theorems (CLT) for the estimation and therefore obtain confidence intervals could be obtained for the estimators.

We will study the behaviour of SA algorithms for the different loss functions in example 4.2.1. Recall that, for $X \in M^\vartheta$ and under the assumption (ffi), there exists a unique risk allocation m^* solution of $1 = E[OI(-X - m^*)]$. We fix K a hyperrectangle such that $m^* \in \text{int}(K)$ and we define for $X \in M^\vartheta$ and $m \in \mathbb{R}^d$:

$$\begin{aligned}
 & H_1(X, m) := OI(-X - m) - 1, \\
 & \quad h_1(m) := E[H_1(X, m)], \\
 & \quad \sigma^2(m) := E[||H_1(X, m) - h_1(m)||^2], \\
 & \quad m^{2+p}(m) := E[||H_1(X, m) - h_1(m)||^{2+p}], \quad p > 0, \\
 & \quad \Sigma(m) := E[(H_1(X, m) - h_1(m))(H_1(X, m) - h_1(m))^T].
 \end{aligned} \tag{4.19}$$

We introduce the following set of assumptions:

$$\begin{aligned}
 (\text{ffi}_{\text{a.s.}}) \quad & \text{i. } \sum_{n \geq 0} \gamma_n = +\infty \text{ and } \sum_{n \geq 0} \gamma_n^2 < \infty; \\
 & \text{ii. } h_1 \text{ is continuous on } K; \\
 & \text{iii. } \sup_{m \in K} \sigma^2(m) < \infty.
 \end{aligned}$$

Theorem 4.3.1. *Let (X_n) a sequence of random variables having the same law as $X \in M^\vartheta$*

and define the sequence (m_n) as follows:

$$m_{n+1} = \Pi_K [m_n + \gamma_n H_1(X_{n+1}, m_n)], \quad m_0 \in L^0, \quad (4.20)$$

where Π_K is the projection into K . Under (ffi_l) and (ffi_{a.s.}) we have, $m_n \rightarrow m^*$ a.s. as $n \rightarrow \infty$.

Proof. Following the same arguments of Theorem 3.3.4 in chapter 3, the only limit point of the projected ODE associated to the algorithm in (4.20) is m^* . Thus, we can use Theorem 2.1 in Kushner and Yin [83] that argues that m_n will converge to the limit point m^* if we can verify their assumptions (A2.1)-(A2.5). Indeed, (A2.1) is guaranteed thanks to the assumption (ffi_{a.s.})-iii.. The other assumptions are verified thanks to (ffi_{a.s.})-ii.. \square

Once we have an estimator of m^* , it comes the question of estimating the multivariate OCE $R(X) = \sum_{i=1}^d m_i^* + E[l(-X - m^*)]$. A naive way consists in estimating $R(X)$ in a two steps procedure:

- Step 1: Use the estimate m_n from (4.20) to have a good approximation of m^* .
- Step 2: Use another sample of X to approximate $R(X)$ using Monte Carlo:

$$R(X) \approx \sum_{i=1}^d m_i^* + \frac{1}{n} \sum_{k=1}^n l(-X_k - m^*). \quad (4.21)$$

A natural way to avoid this two steps procedure is to use a companion procedure (CP) of the algorithm (4.20) and to replace the quantity m^* in (4.21) by its estimate at step $k - 1$, that is,

$$R_n = \frac{1}{n} \sum_{k=1}^n \left(\sum_{i=1}^d m_{k-1}^i + l(-X_k - m_{k-1}) \right).$$

Note that R_n is a sequence of empirical means of non i.i.d. random variables that can be written also as:

$$R_{n+1} = R_n - \frac{1}{n+1} H_2(X_{n+1}, R_n, m_n), \quad n \geq 0, \quad R_0 \in L^0, \quad (4.22)$$

where

$$H_2(x, R, m) := R - F(x, m) := R - \sum_{i=1}^d m^i + l(-x - m).$$

We are now facing two procedures with different time steps: one for the estimation of m^* and the other one for the estimation of $R(X)$. In the following theorem, we prove the consistency of the second procedure using the same time step as the first one (γ_n), namely,

$$R_{n+1} = R_n - \gamma_n H_2(X_{n+1}, R_n, m_n), \quad n \geq 0, \quad R_0 \in L^0. \quad (4.23)$$

To this purpose we need the following assumption:

(ffi_{CP}) $\forall m \in K$, $I(-X - m) \in L^2$ and $m \rightarrow E[|I(-X - m)|^2]$ is bounded around m^* .

Theorem 4.3.2. Assume that assumptions (ffi_i), (ffi_{a.s.}) and (ffi_{CP}) hold and let (m_n) be given by (4.20) and (R_n) by (4.23). Then $R_n \rightarrow R(X)$ a.s.

Proof. For $n \in \mathbb{N}$, define the sequence (S_n) as:

$$S_n = \frac{1}{\prod_{k=0}^{n-1} (1 - \gamma_k)}, \quad S_0 = 1.$$

We have,

$$S_{n+1} = \frac{S_n}{1 - \gamma_n} = S_n \left(1 + \frac{\gamma_n}{1 - \gamma_n} \right) = S_n + \gamma_n S_{n+1}. \quad (4.24)$$

Therefore using (4.23), we have,

$$\begin{aligned} S_{n+1}R_{n+1} &= S_{n+1}(R_n - \gamma_n H_2(X_{n+1}, R_n, m_n)) \\ &= S_n R_n + \gamma_n S_{n+1} R_n - \gamma_n S_{n+1} H_2(X_{n+1}, R_n, m_n) \\ &= S_n R_n + \gamma_n S_{n+1} R_n - \gamma_n S_{n+1} R_n + \gamma_n S_{n+1} F(X_{n+1}, m_n) \\ &= S_n R_n + \gamma_n S_{n+1} F(X_{n+1}, m_n). \end{aligned}$$

This implies for $n \in \mathbb{N}^*$,

$$R_n = \frac{1}{S_n} R_0 + \frac{1}{S_n} \sum_{k=0}^{n-1} \gamma_k S_{k+1} F(X_{k+1}, m_k). \quad (4.25)$$

First, we have

$$\log(S_n) = - \sum_{k=0}^{n-1} \log(1 - \gamma_k) \geq \sum_{k=0}^{n-1} \gamma_k$$

and since by assumption, the RHS of the last inequality goes to ∞ as $n \rightarrow \infty$, we deduce that $S_n \rightarrow \infty$ as $n \rightarrow \infty$ and we get immediately that the first term of the RHS of (4.25) goes to 0 as n goes to ∞ . Rewriting (4.25) by introducing $f(m) := E[F(X, m)]$ and the martingale difference sequence $\delta M_{k+1} = F(X_{k+1}, m_k) - f(m_k)$ with respect to the natural filtration $\mathcal{F}_k := \sigma(m_0, X_1, \dots, X_k)$, we obtain,

$$R_n = \frac{1}{S_n} R_0 + \frac{1}{S_n} \sum_{k=0}^{n-1} \gamma_k S_{k+1} \delta M_{k+1} + \frac{1}{S_n} \sum_{k=0}^{n-1} \gamma_k S_{k+1} f(m_k). \quad (4.26)$$

Thanks to (4.24), we have $\sum_{k=0}^{n-1} \gamma_k S_{k+1} = S_n$. Because f is convex (assumption (ffi_i)-ii.) and therefore continuous at m^* , Cesaro's Lemma implies that the third term in the previous equality

converges to $f(m^*) = R(X)$. The a.s. convergence of R_n will follow from the a.s. convergence of the second term toward 0. Indeed, let us denote,

$$M_n^y = \sum_{k=1}^n \delta M_k^{k=1} \gamma$$

Note that (M_n^y) is a F -martingale such that

$$\sum_{k=1}^{\infty} E[|\delta M_k|^2 | \mathcal{F}_{\infty}^{n=0}] = \gamma$$

But we also have,

$$E[|\delta M_n|^2 | \mathcal{F}_{n-1}] \leq E[|(-X - m)|^2]_{|m=m_{n-1}},$$

and assumption (ffi_{CP})-ii. implies that

$$\sup_{n \geq 1} E[|\delta M_n|^2 | \mathcal{F}_{n-1}] < \infty, \text{ a.s.}$$

Using the martingale convergence theorem, we get that (M_n^y) converges to some random variable. Finally, by Kronecker's Lemma we deduce that the second term of (4.26) converges to 0. This completes the proof. \square

The step sequence in (ffi_{a.s.})-i. is typically of the following form $\gamma_n = \frac{c}{n^\nu}$, where $\nu \in (\frac{1}{2}, 1]$ and c is a positive constant. The choice of c plays a key role in the rate of convergence of SA algorithms. In order to circumvent problems related to the specification of the constant c , which are classical, we will use "averaging" techniques introduced by Ruppert [117] and Polyak and Juditsky [108]. We introduce the following assumptions:

- (ffi_{a.n.}) i. h_1 is continuously differentiable and let $A := Dh_1(m^*)$;
- ii. $(H_1(X_{n+1}, m_n) \mathbf{1}_{|m_n - m^*| \leq \rho})$ is uniformly integrable for small $\rho > 0$;
- iii. For some $\rho > 0$ and $\nu > 0$, $\sup_{|m - m^*| \leq \rho} m^{2+\nu} < \infty$;
- iv. $\Sigma(\cdot)$ is continuous at m^* and $\Sigma^* := \Sigma(m^*)$ is positive definite.

The next theorem states the rate convergence of the average of the iterates of (RM) algorithm.

Theorem 4.3.3. Assume $\nu \in (\frac{1}{2}, 1)$ and that assumptions (ffi_i), (ffi_{a.s.}) and (ffi_{a.n.}) hold. For any arbitrary $t > 0$, we define the (PR) sequence (m_n) as:

$$m_n := \frac{\gamma_n}{t} \sum_{i=n}^{n+t\gamma_n-1} m_i, \tag{4.27}$$

where any upper summation index is interpreted as its integer part. Then, we have

$$\frac{1}{\gamma_n} \sum_{k=0}^{\lfloor \frac{t}{\gamma_n} \rfloor} (\bar{m}_n - m^*) \rightarrow N \left(0, A^{-1} \Sigma^* (A^{-1})^\top + O \left(\frac{1}{t} \right) \right). \quad (4.28)$$

Proof. This is a consequence of Theorem 1.1 chapter 11 page 377 in Kushner and Yin [83] if we could verify their assumption (A1.1). Thanks to Theorem 2.1 of chapter 10 in Kushner and Yin [83], the condition (A1.1) is satisfied as soon as their conditions (A2.0)-(A2.7) hold. Assumption (A2.0) is automatically verified. (A2.1) is satisfied by (ffi_{a.n.})-ii.. (A2.2) is a consequence of Theorem 4.3.1. (A2.4) follows from Taylor's expansions and (ffi_{a.n.})-i.. (A2.5) follows from the fact that $h_1(m^*) = 0$. (A2.6) is satisfied since m^* is the optimum of a convex optimization problem. The first part and second parts of (A2.7) are guaranteed thanks to (ffi_{a.n.})-iii. and (ffi_{a.n.})-iv.. Finally, (A2.3) follows easily from Theorem 4.1 chapter 10 page 341 in Kushner and Yin [83] since all their assumptions (A4.1)-(A4.5) are satisfied. \square

Remark 4.3.4. *The previous CLT theorem states that under suitable conditions our average sequence is asymptotically normal with a corresponding covariance matrix that depends on Σ^* and A . These quantities are unknown to us because, first, in general, they cannot be expressed in a closed form and second they depend on the optimum m^* . So, in practice, these two quantities need to be approximated in order to derive confidence intervals. In the following proposition, we provide consistent estimators of these two quantities.*

Proposition 4.3.5. *Suppose (ffi), (ffi_{a.s.}) and (ffi_{a.n.}) hold. If $m \rightarrow E[||H_1(X, m)||^4]$ is bounded around m^* , then,*

$$\Sigma_n := \frac{1}{n} \sum_{k=0}^{\lfloor \frac{t}{\gamma_n} \rfloor} H_1(X_{k+1}, m_k) (H_1(X_{k+1}, m_k) - \Sigma(m_k))^* \text{ a.s. as } n \rightarrow \infty. \quad (4.29)$$

Let A_n^ϵ be the matrix whose elements $A_n^\epsilon(i, j)$ for $i, j \in \{1, \dots, d\}$ are defined as follows:

$$A_n^\epsilon(i, j) := \frac{1}{\epsilon n} \sum_{k=0}^{n-1} H_1^i(X_{k+1}, m_k + \epsilon e_j) - H_1^i(X_{k+1}, m_k),$$

then,

$$\lim_{\epsilon \rightarrow 0} \lim_{n \rightarrow \infty} A_n^\epsilon = A \text{ a.s.} \quad (4.30)$$

Proof. The proof of this proposition relies mainly on the martingale convergence theorem. Let $(\delta M_k)_{k \in \mathbb{N}}$ be the sequence defined as:

$$\delta M_k := H_1(X_{k+1}, m_k)^\top H_1(X_{k+1}, m_k) - \Sigma(m_k) - h_1(m_k)^\top h_1(m_k), \quad k \geq 0.$$

$(\delta M_k)_{k \geq 0}$ is a martingale difference sequence adapted to F and therefore the following sequence $(M_k)_{k \in \mathbb{N}^*}$ defined as:

$$M_k := \sum_{i=1}^k \frac{\delta M_i}{i}, \quad k \geq 1,$$

is a F -martingale. Furthermore, the boundedness of $m \rightarrow E[|H_1(X, m)|^4]$ around m^* , assumptions (ffi_{a.s.})-ii. and (ffi_{a.n.})-iv. imply that:

$$\sup_{k \geq 1} E[|\delta M_n|^2 | F_n] < \infty \text{ a.s.}$$

Consequently, the martingale convergence theorem implies the existence of a finite random variable M_∞ such that $M_n \rightarrow M_\infty$. We then apply Kronecker's Lemma to get that $\frac{1}{n} \sum_{k=0}^{n-1} \delta M_{k+1} \rightarrow 0$. Since,

$$\Sigma_n = \frac{\sum_{k=0}^{n-1} \delta M_k}{n} = \frac{\sum_{k=0}^{n-1} m_k}{n} = \frac{1}{n} \sum_{k=0}^{n-1} h_1(m_k) = (h_1)m_n,$$

we deduce that $\Sigma_n \rightarrow \Sigma^*$. The proof of (4.30) follows using the same arguments above. □

Remark 4.3.6.

1. Instead of averaging on all observations for the estimators above, we could average using only recent ones. This might improve the behaviour of these estimators.
2. If we denote $V_n^\epsilon = (A_n^\epsilon)^{-1} \Sigma_n ((A_n^\epsilon)^{-1})^\top$, then we can obtain an approximate confidence interval for m^* with a confidence level of $1 - \alpha$ in the following form:

$$m_n - \frac{s_j}{tcn^\nu} q_\alpha \leq m_n \leq m_n + \frac{s_j}{tcn^\nu} q_\alpha, \quad j \in \{1, \dots, d\}, \nu \in (\frac{1}{2}, 1), \quad (4.31)$$

where q_α is the $1 - \frac{\alpha}{2}$ quantile of a standard random variable.

4.4 Numerical Analysis and Examples

In this section, we analyze and test the numerical methods developed in the previous section for the estimation of optimal allocations given by (4.27) and risk measures given by (4.23). The implementation was done on a standard computer using Python and we write CT for computational time. All the computations were run on a standard laptop with a processor Intel(R) Core(TM) i7-9850H CPU @ 2.60GHz. The common parameters used in the computations are summarized in the following table:

| Parameters | Values |
|------------|-----------|
| n | 500000 |
| γ | 0.8 |
| t | 10 |
| c | 1 |
| ϵ | 10^{-6} |

Table 4.1: Set of common parameters.

4.4.1 A first example

We start here by estimating optimal allocations and multivariate OCE for the first loss function in (4.3), that is:

$$I(x) = \sum_{i=1}^d \frac{e^{\lambda_i x_i} - 1}{\lambda_i} + \alpha e^{-\sum_{i=1}^d \lambda_i x_i}, \quad \lambda_i > 0, \quad \alpha \geq 0.$$

We denote by λ the vector of $\lambda_i, i \in \{1, \dots, d\}$. First, we test our algorithms in the case $d = 2$ and the vector X having a Gaussian distribution, as optimal allocations are expressed in a closed form in this case (see (4.10)). This will allow us to test the efficiency of our algorithms. Three cases are considered: In the first case, we take $\alpha = 0$ and $\lambda = (1, 2)$, which as previously mentioned, corresponds to the entropic risk measure, a second one with $\alpha = 1$ and $\lambda = (1, 1)$, and finally a third one with $\alpha = 1, \lambda = (1, 2)$. As for the parameters for the normal distribution of X , we fix $\sigma_1 = \sigma_2 = 1$ and we take $\rho \in \{-0.9, -0.5, 0, 0.5, 0.9\}$ for each of the three cases. The compact set K was set to $[0, 3]^2$ and the initial term $m_0 = (0, 0)$.

| ρ | R_n | \bar{m}_n^1 | \bar{m}_n^2 | CI1 | CI2 | $R(X)$ | m_*^1 | m_*^2 | CT(s) |
|--------|--------|---------------|---------------|------------------|------------------|--------|---------|---------|---------|
| -0.9 | 1.5133 | 0.4987 | 0.9983 | [0.4945, 0.5030] | [0.9874, 1.0093] | 1.5 | 0.5 | 1 | 68.8087 |
| -0.5 | 1.5220 | 0.4964 | 1.0010 | [0.4922, 0.5007] | [0.9908, 1.0112] | 1.5 | 0.5 | 1 | 68.5388 |
| 0 | 1.5054 | 0.4999 | 1.0022 | [0.4956, 0.5042] | [0.9888, 1.0156] | 1.5 | 0.5 | 1 | 69.1064 |
| 0.5 | 1.5147 | 0.5049 | 0.9906 | [0.5006, 0.5092] | [0.9803, 1.0009] | 1.5 | 0.5 | 1 | 70.0251 |
| 0.9 | 1.5264 | 0.5031 | 0.9970 | [0.4988, 0.5074] | [0.9867, 1.0073] | 1.5 | 0.5 | 1 | 70.8549 |

Table 4.2: Numerical results: $\alpha = 0$ and $\lambda = (1, 2)$.

The table above summarizes the numerical results for the first case. The two columns CI1 and CI2 represent the confidence intervals of the (PR) estimators with a confidence level of 95%. Since $\alpha = 0$, the exact optimal allocations do not depend on the correlation coefficient ρ .

This explains why we obtain the same values for m^* for different values of ρ . The same remark goes for the estimators \bar{m}_n . Since $\lambda_2 > \lambda_1$, we expect as per formula (4.10) that $m_*^2 > m_*^1$. These numerical results suggest that the (PR) estimators \bar{m}_n as well as the (RM) estimator R_n approximate very well the exact optimal allocations m^* and the risk measure $R(X)$. The width of the first confidence intervals (resp. second confidence intervals) is approximately 0.008 (resp. 0.02) which gives an accuracy of 1.5% (resp. 2%) for the first estimator \bar{m}_n^1 (resp. \bar{m}_n^2).

| ρ | R_n | \bar{m}_n^1 | \bar{m}_n^2 | CI1 | CI2 | $R(X)$ | m_*^1 | m_*^2 | CT(s) |
|--------|--------|---------------|---------------|------------------|------------------|--------|---------|---------|---------|
| -0.9 | 1.3139 | 0.7689 | 0.7704 | [0.7651, 0.7728] | [0.7665, 0.7742] | 1.3036 | 0.7702 | 0.7702 | 69.1816 |
| -0.5 | 1.4198 | 0.8511 | 0.8552 | [0.8471, 0.8550] | [0.8512, 0.8592] | 1.4105 | 0.8545 | 0.8545 | 73.6128 |
| 0 | 1.5897 | 0.9827 | 0.9801 | [0.9782, 0.9873] | [0.9755, 0.9846] | 1.5804 | 0.9812 | 0.9812 | 69.8510 |
| 0.5 | 1.8171 | 1.1368 | 1.1280 | [1.1307, 1.1430] | [1.1220, 1.1339] | 1.7928 | 1.1301 | 1.1301 | 71.4592 |
| 0.9 | 2.0305 | 1.2697 | 1.2651 | [1.2612, 1.2782] | [1.2568, 1.2734] | 1.9932 | 1.2636 | 1.2636 | 73.0060 |

Table 4.3: Numerical results: $\alpha = 1$ and $\lambda = (1, 1)$.

When taking the same values for λ_1 and λ_2 , the system becomes symmetric and we obtain the same optimal allocations for the first and second component. We also notice that optimal allocations and their estimators increase with the correlation coefficient ρ as it was expected from remark 4.2.6. Again, the estimators approximate well the optimal allocations and the risk measure. The accuracy of all confidence intervals is around $\approx 1\%$.

| ρ | R_n | \bar{m}_n^1 | \bar{m}_n^2 | CI1 | CI2 | $R(X)$ | m_*^1 | m_*^2 | CT(s) |
|--------|--------|---------------|---------------|------------------|------------------|--------|---------|---------|---------|
| -0.9 | 1.6477 | 0.6194 | 1.1275 | [0.6152, 0.6237] | [1.1184, 1.1366] | 1.6354 | 0.6202 | 1.1285 | 79.6036 |
| -0.5 | 1.7734 | 0.7045 | 1.2366 | [0.7001, 0.7089] | [1.2280, 1.2452] | 1.7544 | 0.7071 | 1.2344 | 81.1432 |
| 0 | 2.0148 | 0.8479 | 1.4449 | [0.8421, 0.8538] | [1.4309, 1.4588] | 1.9943 | 0.8465 | 1.4406 | 76.7199 |
| 0.5 | 2.3749 | 0.9922 | 1.7260 | [0.9844, 1.0001] | [1.7044, 1.7476] | 2.3354 | 0.9859 | 1.7344 | 80.4979 |
| 0.9 | 2.7790 | 1.0812 | 2.0416 | [1.0710, 1.0914] | [1.9981, 2.0850] | 2.6652 | 1.0728 | 2.0285 | 81.1827 |

Table 4.4: Numerical results: $\alpha = 1$ and $\lambda = (1, 2)$.

In this final case, we take different values for λ_1 and λ_2 . Table 4.4 shows that the optimal allocations can be well approximated by the estimator in (4.27). This is also the case for the estimator R_n . Again, the optimal allocations as well as the risk measure increase with the correlation coefficient (see Figure 4.1). All confidence intervals have an accuracy between 1% and 2% except the second confidence interval in the case $\rho = 0.9$ where the accuracy is a bit higher and is approximately around 4%.

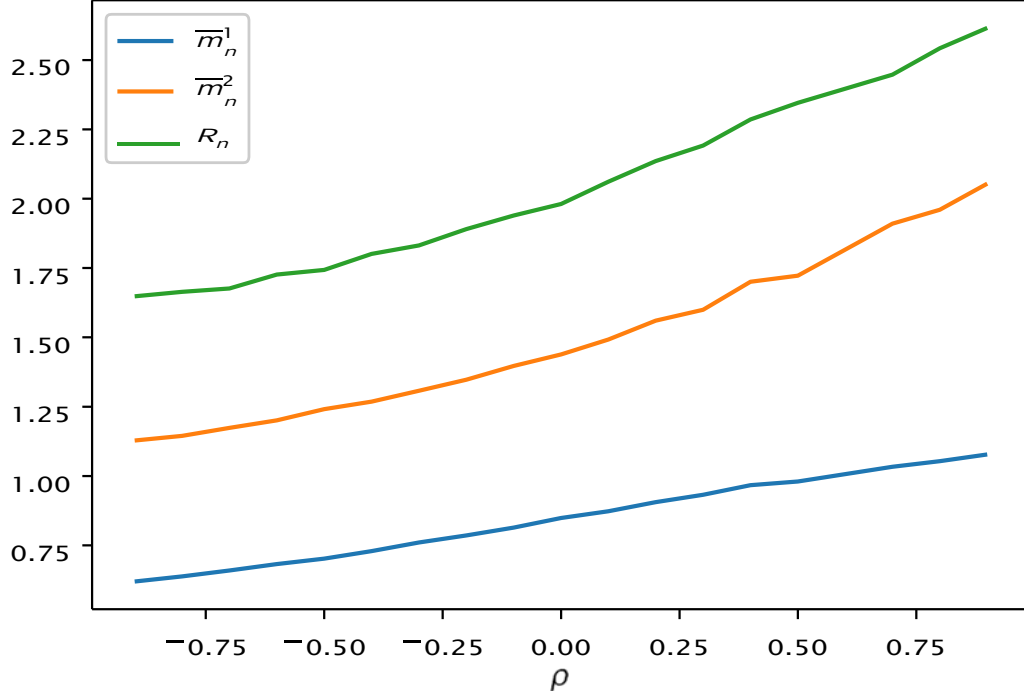


Figure 4.1: R_n , \bar{m}_n^1 and \bar{m}_n^2 as a function of ρ .

4.4.2 Second example

Simulated data

In this example, we will be working with a Multivariate Normal Inverse Gaussian (MNIG) distribution for the vector X instead of a Gaussian distribution. The MNIG distribution yields a more flexible family of distributions that can be skewed and have fatter tails than the Gaussian distribution. For a fixed d , a MNIG distributed random variable is a variance-mean mixture of a d -Gaussian random variable Y with a univariate inverse Gaussian distributed mixing variable Z . The MNIG distribution has five parameters $\alpha_{MNIG} > 0$, $\beta \in \mathbb{R}^d$, $\delta > 0$, $\mu \in \mathbb{R}^d$ and $\Gamma \in \mathbb{R}^{d \times d}$ and can be constructed as follows:

$$X = \mu + Z\Gamma\beta + \sqrt{\frac{\alpha_{MNIG}}{Z}}\Gamma^{1/2}Y, \quad (4.32)$$

where $Z \sim IG(\delta^2, \alpha_{MNIG}^2 - \beta^\top \Gamma \beta)$ and $IG(\chi, \psi)$ denotes the Inverse Gaussian distribution with parameters $\chi, \psi > 0$ and $Y \sim N(0, I_d)$. Note that the random variable $X|Z \sim N(\mu + Z\Gamma\beta, Z\Gamma)$, hence the name variance-mean mixture. The parameters of the MNIG distribution have natural

interpretations. The parameter α_{MNIG} is a shape parameter and controls the heaviness of the tails. The parameter β is a skewness vector parameter, δ is a scale parameter and μ is a vector translation parameter. Finally, the matrix Γ is assumed to be a positive semidefinite symmetric matrix and controls the degree of correlations between components and assumed to be such that $\det(\Gamma) = 1$. In order for the MNIG to exist, the inequality $\alpha_{MNIG}^2 > \beta^T \Gamma \beta$ must be satisfied. The cumulant generating function of the MNIG could be derived easily in a closed form of the parameters:

$$\Phi_X(t) = \delta \frac{q}{\alpha_{MNIG}^2 - \beta^T \Gamma \beta} - \frac{q}{\alpha_{MNIG}^2 - (\beta + it)^T \Gamma (\beta + it)} + it^T \mu.$$

This shows that the MNIG is infinitely divisible. Thus, we can easily evaluate the moments of this distribution. The mean vector and the covariance matrix Σ of X are given in the following:

$$E[X] = \mu + \frac{\delta \Gamma \beta}{\alpha_{MNIG}^2 - \beta^T \Gamma \beta}, \tag{4.33}$$

$$\Sigma = \delta \frac{\alpha_{MNIG}^2 - \beta^T \Gamma \beta}{\alpha_{MNIG}^2 - \beta^T \Gamma \beta}^{-1/2} \Gamma + \frac{\alpha_{MNIG}^2 - \beta^T \Gamma \beta}{\alpha_{MNIG}^2 - \beta^T \Gamma \beta}^{-1} \Gamma \beta \beta^T \Gamma. \tag{4.34}$$

Note that due to the parameter β , even when $\mu = 0$ (and $\Gamma = I_d$ resp.), the mean of the MNIG distribution is not null (the covariance matrix is not diagonal resp.). For more details about MNIG, we refer to Oigård et al. [105].

In order to make the numerical analysis more realistic, we fitted, the parameters of the MNIG distribution on the daily log-return of three European indices: CAC 40, BEL 20 and AEX. The estimated parameters obtained using the Expectation Maximization (EM), explained in details in Section 4.5, are summarized in the first column of the following table 4.5.

| Parameters | MNIG |
|-----------------|---|
| α_{MNIG} | 365.78 |
| δ | 0.00373 |
| β | (-64.28, 41.45, 7.35) |
| μ | (0.00084, 0.00024, 0.00055) |
| | 2.338 1.796 2.080 |
| Γ | : 1.796 2.327 2.088; 2.080 2.088 2.555 |

Table 4.5: Parameters sets for the MNIG.

The covariance matrix obtained from the MNIG calibrated distribution is given in the follow-

ing:

$$\Sigma = \begin{pmatrix} 2.45 \times 10^{-5} & 1.86 \times 10^{-5} & 2.16 \times 10^{-5} \\ 1.86 \times 10^{-5} & 2.40 \times 10^{-5} & 2.16 \times 10^{-5} \\ 2.16 \times 10^{-5} & 2.16 \times 10^{-5} & 2.65 \times 10^{-5} \end{pmatrix}$$

This shows that the log-returns of the three indices over the period considered are barely correlated. The following figure shows also that they almost have the same distribution:

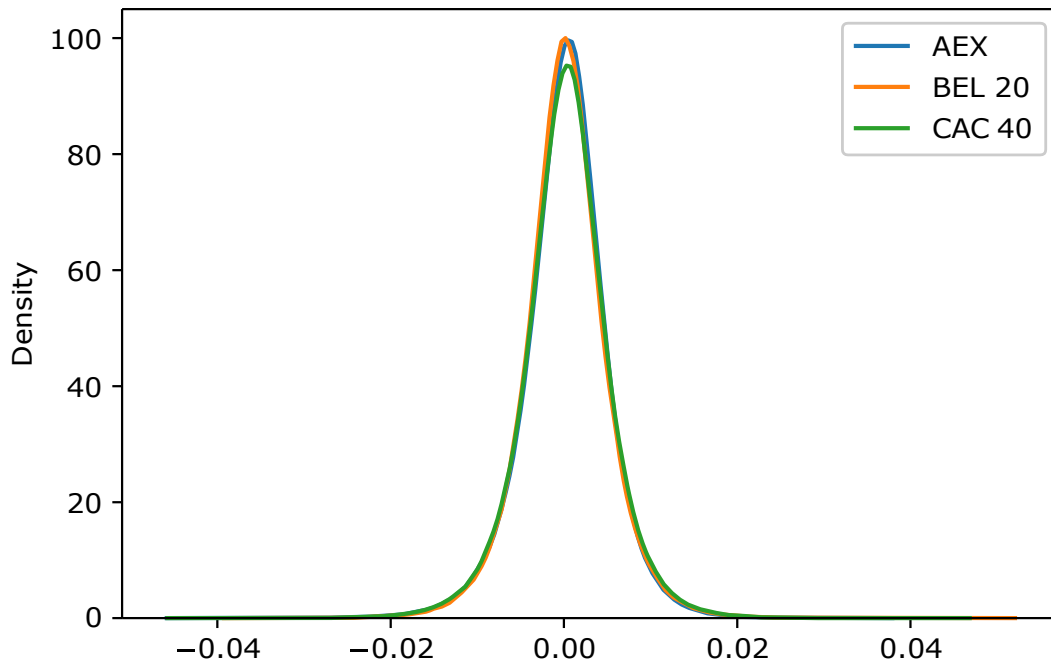


Figure 4.2: Densities of the log-returns of the three indices.

Numerical Results

We will test our SA algorithms with a trivariate MNIG distribution for the polynomial loss functions. We recall that the polynomial loss function is given by:

$$l(x) = \sum_{i=1}^d \frac{([1+x_i])^{+\vartheta_i} - 1}{\vartheta_i} + \alpha \sum_{i < j} \frac{([1+x_i])^{+\vartheta_i} ([1+x_j])^{+\vartheta_j}}{\vartheta_i \vartheta_j}, \vartheta_i > 1, \alpha \geq 0.$$

Since no closed formula is available to us in this case, we decided to use a Monte Carlo scheme as a benchmark to the SA method. This scheme consists in approximating the expectation in

(4.11) by the corresponding Monte Carlo estimator and then to use Nelder-Mead algorithm as a minimization algorithm to find the optimal allocations. The compact set K for the SA method was set to $[0, 2]^3$ and α was taken to be equal to 1. First, we compare both methods in the case where the parameter ϑ was taken to be equal to $\vartheta = (2, 2, 2)$. Then, in a second case, we test both algorithms with the parameter $\vartheta = (1, 2, 3)$.

| | SA | CI-SA | Monte Carlo |
|---------|---------|--------------------|-------------|
| m^*_1 | 0.31747 | [0.31746, 0.31749] | 0.31748 |
| m^*_2 | 0.31748 | [0.31746, 0.31750] | 0.31745 |
| m^*_3 | 0.31742 | [0.31740, 0.31743] | 0.31737 |
| $R(X)$ | | 0.31336 | 0.31332 |
| CT(s) | | 141.20 | 28.07 |

Table 4.6: Numerical results: Polynomial loss function with $\vartheta = (2, 2, 2)$ and MNIG distribution.

The table 4.6 show that both methods give approximately the same values for the optimal allocations m^* as well as the risk measure $R(X)$. The values of the optimal allocations are approximately the same among the three components. This could be explained by the fact that the three components have almost the same distribution as shown in the figure 4.2 and the fact that we have taken $\vartheta = (2, 2, 2)$, so that the system becomes nearly symmetric. The Monte Carlo method is seven times faster that the SA method. However, with the Monte Carlo method, we do not have any confidence intervals and hence no control over the error of estimation. Moreover, since in the Monte Carlo method, we are using a deterministic minimization algorithm, it is sensitive to the initial values (Recall that we do not have this problem with the SA method). We do not have this problem with the SA method.

| | SA | CI-SA | Monte Carlo |
|---------|---------|--------------------|-------------|
| m^*_1 | 0.21996 | [0.21995, 0.21998] | 0.21994 |
| m^*_2 | 0.25127 | [0.25125, 0.25129] | 0.25130 |
| m^*_3 | 0.29929 | [0.29927, 0.29931] | 0.29926 |
| $R(X)$ | | 0.37532 | 0.37529 |
| CT(s) | | 98.48 | 9.85 |

Table 4.7: Numerical results: Polynomial loss function with $\vartheta = (1, 2, 3)$ and MNIG distribution.

4.5 Appendix: Estimation of MNIG parameters

4.5.1 Computational aspects

In this section, we give more details about the estimation of the MNIG parameters. The most conventional way to estimate the latter is the maximum likelihood estimation method. However, in the case of MNIG, this method shows slow convergence due to the complexity of the likelihood. We therefore, propose here to use the Expectation Maximization (EM) algorithm which is known to be fast and accurate. The EM algorithm is a powerful tool that is used for maximum likelihood estimation for data containing “missing” values. This is suitable for distributions arising as mixtures which is the case of MNIG distributions where the mixing variable Z is unobserved. The EM algorithm is an iterative algorithm that consists of two steps at each iteration. Denoting $\vartheta = (\delta, \mu, \beta, \alpha_{MNIG}, \Gamma)$, $\mathbf{X} = (X_1, \dots, X_N)$ the observed data and $Z = (Z_1, \dots, Z_N)$ the unobserved one, $L(\mathbf{X}, Z, \vartheta) = \log(P(\mathbf{X}, Z|\vartheta))$ the complete data likelihood and ϑ^n the estimate of ϑ at step n , we repeat the two following steps until some convergence criteria is verified:

- E-step : Compute $Q(\vartheta|\vartheta^n) := E_{Z|\mathbf{X}, \vartheta^n}[L(\mathbf{X}, Z, \vartheta^n)]$.
- M-step : choose $\vartheta^{n+1} = \underset{\vartheta}{\operatorname{argmax}} Q(\vartheta|\vartheta^n)$.

Next, we explicit the calculations of $Q(\vartheta|\vartheta^n)$ in the E-step for the MNIG distribution. We have, by taking the constants away,

$$\begin{aligned} L(\mathbf{X}, Z, \vartheta) &= \log(P(\mathbf{X}, Z|\vartheta)) \\ &= \log(P(\mathbf{X}|Z, \vartheta)) + \log(P(Z|\vartheta)) \\ &= -\frac{d}{2} \sum_{i=1}^N \log(Z_i) - \frac{N}{2} \log(\det(\Gamma)) - \frac{1}{2} \sum_{i=1}^N \frac{1}{Z_i} (\mathbf{X}_i - \mu - Z_i \Gamma \beta)^\top \Gamma^{-1} (\mathbf{X}_i - \mu - Z_i \Gamma \beta) + \\ &\quad N \frac{q}{\delta^2 (\alpha_{MNIG}^2 - \beta^\top \Gamma \beta)} + N \log(\delta) - \frac{3}{2} \sum_{i=1}^N \log(Z_i) - \frac{1}{2} \sum_{i=1}^N \delta^2 \frac{1}{Z_i} + (\alpha_{MNIG}^2 - \beta^\top \Gamma \beta) Z_i . \end{aligned}$$

Taking the conditional expectation on the both sides and denoting $\zeta^n = (\zeta_i^n)_{i=1, \dots, N}$ and $\phi^n = (\phi_i^n)_{i=1, \dots, N}$, where $\zeta_i^n := E_{Z|\mathbf{X}, \vartheta^n}[Z_i]$ and $\phi_i^n := E_{Z|\mathbf{X}, \vartheta^n}[\frac{1}{Z_i}]$, we get, again by removing the quantities that does not depend on ϑ ,

$$\begin{aligned} Q(\vartheta|\vartheta^n) &= -\frac{N}{2} \log(\det(\Gamma)) - \frac{1}{2} \sum_{i=1}^N \phi_i^n (\mathbf{X}_i - \mu)^\top \Gamma^{-1} (\mathbf{X}_i - \mu) + \zeta_i^n \beta^\top \Gamma \beta - 2(\mathbf{X}_i - \mu)^\top \beta \\ &\quad + N \frac{q}{\delta^2 (\alpha_{MNIG}^2 - \beta^\top \Gamma \beta)} + N \log(\delta) - \frac{1}{2} \sum_{i=1}^N \phi_i^n \delta^2 + \zeta_i^n (\alpha_{MNIG}^2 - \beta^\top \Gamma \beta) . \end{aligned}$$

The quantities ϕ_i^n and ζ_i^n can be derived from the fact that $Z|X, \vartheta$ follows a Generalized Inverse Gaussian distribution, i.e., $Z|X, \vartheta \sim GIG_{-\frac{d+1}{2}, q(X), \alpha_{MNIG}}$, where q is given as:

$$q(x) = \frac{q}{\delta^2 + (x - \mu)^\top \Gamma^{-1} (x - \mu)}. \tag{4.35}$$

More precisely, they are given by,

$$\zeta_i := E_{Z_i|X_i, \vartheta} [Z_i] = \frac{q(X_i) K_{(d-1)/2}(\alpha q(X_i))}{\alpha K_{(d+1)/2}(\alpha q(X_i))}, \tag{4.36}$$

$$\phi_i := E_{Z_i|X_i, \vartheta} [1/Z_i] = \frac{\alpha K_{(d+1)/2}(\alpha q(X_i))}{q(X_i) K_{(d+3)/2}(\alpha q(X_i))}, \tag{4.37}$$

K_ν is the modified Bessel function of the second kind with index $\nu \in \mathbb{R}$. Having calculated $Q(\vartheta|\vartheta^n)$, we now need to calculate the next term $\vartheta^{n+1} := \underset{\vartheta}{\operatorname{argmax}} Q(\vartheta|\vartheta^n)$. This will be done by first calculating the gradient of Q .

$$\begin{aligned} \frac{\partial Q}{\partial \delta} &= N \frac{q}{\alpha_{MNIG}^2 - \beta^\top \Gamma \beta} + \frac{N}{\delta} - \delta \sum_{i=1}^n \phi_i^n, \\ \frac{\partial Q}{\partial \alpha_{MNIG}} &= \frac{N \delta \alpha}{\alpha^2} - \sum_{i=1}^n \zeta_i^n, \\ \frac{\partial Q}{\partial \beta} &= \Gamma^{-1} \sum_{i=1}^n \phi_i^n (X_i - \mu) - \frac{N \beta}{\alpha_{MNIG}^2 - \beta^\top \Gamma \beta}, \\ \frac{\partial Q}{\partial \Gamma} &= \frac{1}{2} \Gamma^{-1} \sum_{i=1}^n \phi_i^n (X_i - \mu) (X_i - \mu)^\top \Gamma^{-1} - \frac{N \delta \beta \beta^\top}{\alpha_{MNIG}^2 - \beta^\top \Gamma \beta}. \end{aligned}$$

To alleviate the expressions, we will denote $\bar{\phi}^n := \frac{1}{N} \sum_{i=1}^N \phi_i^n$, $\bar{\zeta}^n := \frac{1}{N} \sum_{i=1}^N \zeta_i^n$, $\bar{X}\phi^n := \frac{1}{N} \sum_{i=1}^N \phi_i^n X_i$ and $\bar{X} := \frac{1}{N} \sum_{i=1}^N X_i$. Setting the previous set of equations to 0, we obtain,

$$1 - \frac{1}{\bar{\phi}^n} + \frac{\alpha^2}{\delta} - \frac{\beta^\top \Gamma \beta}{\alpha_{MNIG}^2} = 0, \quad (4.38)$$

$$\frac{\delta}{\alpha_{MNIG}^2} - \frac{\beta^\top \Gamma \beta}{\alpha_{MNIG}^2} - \frac{\zeta}{\bar{\phi}^n} = 0, \quad (4.39)$$

$$\Gamma^{-1} \bar{X}\phi^n - \bar{\phi}^n \Gamma^{-1} \mu - \beta = 0, \quad (4.40)$$

$$\bar{X} - \mu - \frac{\delta \Gamma \beta}{\alpha_{MNIG}^2 - \beta^\top \Gamma \beta} = 0, \quad (4.41)$$

$$\frac{1}{N} \sum_{i=1}^N \phi_i (X_i - \mu)(X_i - \mu)^\top - \Gamma - \frac{\delta \Gamma \beta \beta^\top \Gamma}{\alpha_{MNIG}^2 - \beta^\top \Gamma \beta} = 0. \quad (4.42)$$

From the second equation we deduce that,

$$\frac{\delta}{\alpha_{MNIG}^2 - \beta^\top \Gamma \beta} = \frac{\zeta}{\bar{\phi}^n}. \quad (4.43)$$

Plugging this into the first equation gives us,

$$\delta = \frac{1}{\frac{\alpha_{MNIG}^2}{\bar{\phi}^n} - \frac{\beta^\top \Gamma \beta}{\zeta}}. \quad (4.44)$$

Thanks to the third equation, we have,

$$\Gamma \beta = \bar{X}\phi^n - \mu \bar{\phi}^n. \quad (4.45)$$

Using this in the fourth equation, we obtain,

$$\mu = \frac{\bar{X} - \bar{\zeta}^n \bar{X}\phi^n}{1 - \bar{\zeta}^n \bar{\phi}^n}. \quad (4.46)$$

Now that μ and $\Gamma \beta$ are explicitly known, denoting $R := \frac{1}{N} \sum_{i=1}^N \phi_i (X_i - \mu)(X_i - \mu)^\top - \bar{\zeta}^n \Gamma \beta (\Gamma \beta)^\top$, we get from the fifth equation,

$$\Gamma = \frac{R}{\det(R)^{1/d}}. \quad (4.47)$$

Going back to (4.45), we get,

$$\beta = \Gamma^{-1} (\bar{X}\phi^n - \bar{\phi}^n \mu). \quad (4.48)$$

Finally, using (4.43), α_{MNIG} can be deduced as:

$$\alpha = \frac{\delta^2}{\zeta^n} + \beta + \Gamma\beta. \quad (4.49)$$

In the following, we summarize the EM algorithm for the parameters estimation of MNIG distribution: The convergence properties of the EM algorithm are discussed in details in McLachlan

Algorithm 2: EM algorithm for parameters estimation of MNIG distribution

Input: Observations (X_1, \dots, X_N) , initial value ϑ^0 , tolerance tol and M number of iterations;

- 1 Set $\vartheta^n = \vartheta^0$ and compute ϕ^n, ζ^n with (4.36) and (4.37);
- 2 Compute ϑ^{n+1} using in order (4.44), (4.46), (4.45), (4.47), (4.48) and (4.49);
- 3 $n = 0$;
- 4 **while** $\|\vartheta^{n+1} - \vartheta^n\| \geq \text{tol}$ **and** $n < M$ **do**
- 5 E-step: $\vartheta^n \leftarrow \vartheta^{n+1}$ and compute the new ϕ^n and ζ^n with (4.36) and (4.37);
- 6 M-step: Compute ϑ^{n+1} using in order (4.44), (4.46), (4.45), (4.47), (4.48) and (4.49);
- 7 $n \leftarrow n + 1$;

Output: Estimated parameters $\hat{\vartheta}$;

and Krishnan [95]. However, to avoid getting stuck in a local maximum, we will need to run the algorithm from several starting points to ensure that the obtained maximum is the global one. We can also combine the algorithm with other numerical methods, such as Newton-Raphson algorithm, to speed up the convergence.

4.5.2 Numerical aspects

We applied the EM algorithm described in the above subsection to a data set of daily log return of three European stock indices (CAC 40, BEL 20 and AEX) for a period from 12/05/2020 to 10/05/2022 obtained from the website of Euronext. The data set consisted of 514 observations. In order to test the behavior of the algorithm, several initial values were considered. Note that the conditional expectations in the E-step do not involve the parameters β and hence the convergence of the algorithm will not depend on the initial value of β . We fixed $\beta = (0, 0, 0)$ and we stopped the iterations when $\|\vartheta^{n+1} - \vartheta^n\| < \text{tol}$ for $\text{tol} = 10^{-5}$ and $\text{tol} = 10^{-10}$. The values of the

estimates of estimates for initial values were the same and are given in the following:

$$\begin{aligned} \hat{\alpha} &= 365.78, \\ \hat{\delta} &= 0.00373, \\ \hat{\beta} &= (-64.28, 41.45, 7.35), \\ \hat{\mu} &= (0.00084, 0.00024, 0.00055), \\ \hat{\Gamma} &= \begin{pmatrix} 2.338 & 1.796 & 2.080 \\ 1.796 & 2.327 & 2.088 \\ 2.080 & 2.088 & 2.555 \end{pmatrix}. \end{aligned}$$

The number of iterations needed until convergence along with the computational (CT) time can be seen in 4.8.

| Initial Values | | | | tol = 10 ⁻⁵ | | tol = 10 ⁻¹⁰ | |
|----------------|----------|-----------|----------|------------------------|--------|-------------------------|--------|
| α | δ | μ | Γ | Iterations | CT(ms) | Iterations | CT(ms) |
| 1 | 0 | (0, 0, 0) | I_3 | 93 | 1267 | 164 | 2175 |
| 20 | 0 | (0, 0, 0) | I_3 | 93 | 1174 | 164 | 2139 |
| 1 | 2 | (0, 0, 0) | I_3 | 194 | 2428 | 265 | 3377 |
| 20 | 2 | (0, 0, 0) | I_3 | 453 | 5543 | 524 | 6471 |
| 1 | 0 | (1, 1, 1) | I_3 | 189 | 2425 | 260 | 3391 |
| 20 | 0 | (1, 1, 1) | I_3 | 410 | 5058 | 481 | 5898 |
| 1 | 2 | (1, 1, 1) | I_3 | 206 | 2563 | 217 | 3393 |
| 20 | 2 | (1, 1, 1) | I_3 | 558 | 6658 | 629 | 7644 |
| 1 | 0 | (0, 0, 0) | $2I_3$ | 94 | 1237 | 165 | 2078 |
| 20 | 0 | (0, 0, 0) | $2I_3$ | 94 | 1154 | 165 | 2100 |
| 1 | 2 | (0, 0, 0) | $2I_3$ | 194 | 2383 | 265 | 3315 |
| 20 | 2 | (0, 0, 0) | $2I_3$ | 453 | 5536 | 524 | 6558 |
| 1 | 0 | (1, 1, 1) | $2I_3$ | 177 | 2160 | 248 | 3032 |
| 20 | 0 | (1, 1, 1) | $2I_3$ | 328 | 3947 | 399 | 4876 |
| 1 | 2 | (1, 1, 1) | $2I_3$ | 201 | 2468 | 272 | 3282 |
| 20 | 2 | (1, 1, 1) | $2I_3$ | 509 | 6208 | 580 | 7118 |

Table 4.8: Number of iterations and computational time for various initial values and stopping criteria.

REFERENCES

- [1] V. Acharya, R. Engle, and M. Richardson. "Capital shortfall: A new approach to ranking and regulating systemic risks". In: *American Economic Review* 102.3 (2012), pp. 59-64.
- [2] Y. Aït-Sahalia, J. Cacho-Diaz, and R. J. Laeven. "Modeling financial contagion using mutually exciting jump processes". In: *Journal of Financial Economics* 117.3 (2015), pp. 585-606.
- [3] E. W. Anderson, L. P. Hansen, and T. J. Sargent. "A quartet of semigroups for model specification, robustness, prices of risk, and model detection". In: *Journal of the European Economic Association* 1.1 (2003), pp. 68-123.
- [4] Y. Armenti, S. Crépey, S. Drapeau, and A. Papantoleon. "Multivariate Shortfall Risk Allocation and Systemic Risk". In: *SIAM J. FINANCIAL MATH.* 9.1 (2018), pp. 90-126.
- [5] P. Artzner, F. Delbaen, J.-M. Eber, and D. Heath. "Coherent measures of risk". In: *Mathematical Finance* 9.3 (1999), pp. 203-228.
- [6] O. Bardou, N. Frikha, and G. Pagès. "Computing VaR and CVaR using stochastic approximation and adaptive unconstrained importance sampling". In: *Monte Carlo Methods and Applications* 15.3 (2009), pp. 173-210.
- [7] P. Barrieu and N. El Karoui. "Monotone stability of quadratic semimartingales with applications to unbounded general quadratic BSDEs". In: *The Annals of Probability* 41.3B (2013), pp. 1831-1863.
- [8] D. Becherer. "Bounded solutions to backward SDEs with jumps for utility optimization and indifference hedging". In: *The Annals of Applied Probability* 16.4 (2006), pp. 2027-2054.
- [9] D. Becherer. "Rational hedging and valuation with utility-based preferences". In: (2001).
- [10] A. Ben-Tal and M. Teboulle. "An old-new concept of convex risk measures: the optimized certainty equivalent". In: *Mathematical Finance* 17.3 (2007), pp. 449-476.
- [11] A. Ben-Tal and M. Teboulle. "Expected utility, penalty functions, and duality in stochastic nonlinear programming". In: *Management Science* 32.11 (1986), pp. 1445-1466.
- [12] A. Benveniste, M. Metivier, and P. Priouret. *Adaptive Algorithms and Stochastic Approximations*. 1st ed. Springer-Verlag, Berlin, 1990.
- [13] M. Benzi, G. H. Golub, and J. Liesen. "Numerical solution of saddle point problems". In: *Acta Numerica* 14 (2005), 1-137.

REFERENCES

- [14] Y. Bessy-Roland, A. Boumezoued, and C. Hillairet. "Multivariate Hawkes process for cyber insurance". In: *Annals of Actuarial Science* 15.1 (2021), pp. 14-39.
- [15] F. Biagini, J. Fouque, M. Frittelli, and T. Meyer-Brandis. "A Unified Approach to Systemic Risk Measures Via Acceptance Sets". In: *Mathematical Finance* 29.1 (2019), pp. 329-367.
- [16] S. Biagini and M. Frittelli. "On the extension of the Namioka-Klee theorem and on the Fatou property for risk measures". In: *Optimality and risk-modern trends in mathematical finance*. Springer, 2009, pp. 1-28.
- [17] J.-M. Bismut. "Conjugate convex functions in optimal stochastic control". In: *Journal of Mathematical Analysis and Applications* 44.2 (1973), pp. 384-404.
- [18] G. Bordigoni, A. Matoussi, and M. Schweizer. "A stochastic control approach to a robust utility maximization problem". In: *Stochastic analysis and applications*. Springer, 2007, pp. 125-151.
- [19] P. Briand and Y. Hu. "BSDE with quadratic growth and unbounded terminal value". In: *Probability Theory and Related Fields* 136.4 (2006), pp. 604-618.
- [20] P. Briand and Y. Hu. "Quadratic BSDEs with convex generators and unbounded terminal conditions". In: *Probability Theory and Related Fields* 141.3 (2008), pp. 543-567.
- [21] M. K. Brunnermeier and P. Cheridito. "Measuring and allocating systemic risk". In: *Risks* 7.2 (2019), p. 46.
- [22] H. Buehler, L. Gonon, J. Teichmann, and B. Wood. "Deep hedging". In: *Quantitative Finance* 19.8 (2019), pp. 1271-1291.
- [23] H. Bühlmann. *Mathematical methods in risk theory*. Vol. 172. Springer Science & Business Media, 2007.
- [24] R. Carmona. *Indifference pricing: theory and applications*. Princeton University Press, 2008.
- [25] A. Černý, F. Maccheroni, M. Marinacci, and A. Rustichini. "On the computation of optimal monotone mean-variance portfolios via truncated quadratic utility". In: *Journal of Mathematical Economics* 48.6 (2012), pp. 386-395.
- [26] J.-F. Chassagneux and A. Richou. "Numerical simulation of quadratic BSDEs". In: *The Annals of Applied Probability* 26.1 (2016), pp. 262-304.
- [27] C. Chen, G. Iyengar, and C. C. Moallemi. "An axiomatic approach to systemic risk". In: *Management Science* 59.6 (2013), pp. 1373-1388.
- [28] H. Chen and Y. Zhu. "Stochastic approximation procedures with randomly varying truncations". In: *Science in China, Ser. A* (1986).

-
- [29] P. Cheridito and T. Li. "Risk measures on Orlicz hearts". In: *Mathematical Finance: An International Journal of Mathematics, Statistics and Financial Economics* 19.2 (2009), pp. 189-214.
- [30] A. S. Cherny and M. Kupper. "Divergence utilities". In: *Available at SSRN 1023525* (2007).
- [31] L. Clewlow and S. Hodges. "Optimal delta-hedging under transactions costs". In: *Journal of Economic Dynamics and Control* 21.8-9 (1997), pp. 1353-1376.
- [32] J. C. Cox and C.-f. Huang. "Optimal consumption and portfolio policies when asset prices follow a diffusion process". In: *Journal of economic theory* 49.1 (1989), pp. 33-83.
- [33] J. Cvitanić and I. Karatzas. "Convex duality in constrained portfolio optimization". In: *The Annals of Applied Probability* (1992), pp. 767-818.
- [34] J. Cvitanić and I. Karatzas. "Hedging contingent claims with constrained portfolios". In: *The Annals of Applied Probability* (1993), pp. 652-681.
- [35] G. Deelstra, H. Pham, and N. Touzi. "Dual formulation of the utility maximization problem under transaction costs". In: *Annals of Applied Probability* (2001), pp. 1353-1383.
- [36] F. Delbaen, P. Grandits, T. Rheinländer, D. Samperi, M. Schweizer, and C. Stricker. "Exponential hedging and entropic penalties". In: *Mathematical finance* 12.2 (2002), pp. 99-123.
- [37] F. Delbaen and W. Schachermayer. "A general version of the fundamental theorem of asset pricing". In: *Mathematische annalen* 300.1 (1994), pp. 463-520.
- [38] O. Deprez and H. U. Gerber. "On convex principles of premium calculation". In: *Insurance: Mathematics and Economics* 4.3 (1985), pp. 179-189.
- [39] S. Drapeau, M. Kupper, and A. Papapantoleon. "A Fourier approach to the computation of CVaR and optimized certainty equivalents". In: *arXiv preprint arXiv:1212.6732* (2012).
- [40] D. Duffie and L. G. Epstein. "Stochastic differential utility". In: *Econometrica: Journal of the Econometric Society* (1992), pp. 353-394.
- [41] D. Duffie and C. Skiadas. "Continuous-time security pricing: A utility gradient approach". In: *Journal of Mathematical Economics* 23.2 (1994), pp. 107-131.
- [42] M. Duflo. *Algorithmes stochastiques*. Springer Berlin Heidelberg, 1996.
- [43] J. Dunkel and S. Weber. "Stochastic root finding and efficient estimation of convex risk measures". In: *Operations Research* 58.5 (2010), pp. 1505-1521.
- [44] L. Eisenberg and T. H. Noe. "Systemic risk in financial systems". In: *Management Science* 47.2 (2001), pp. 236-249.

REFERENCES

- [45] N El Karoui and M. Quenez. "Non-linear pricing theory and backward stochastic differential equations". In: *Financial mathematics* (1997), pp. 191-246.
- [46] N. El Karoui and M.-C. Quenez. "Dynamic programming and pricing of contingent claims in an incomplete market". In: *SIAM journal on Control and Optimization* 33.1 (1995), pp. 29-66.
- [47] W. Faidi, A. Matoussi, and M. Mni. "Robust utility maximization problem with a general penalty term". In: *Preprint* (2013).
- [48] A. V. Fiacco and G. P. McCormick. *Nonlinear programming: sequential unconstrained minimization techniques*. SIAM, 1990.
- [49] H. Föllmer and A. Gundel. "Robust projections in the class of martingale measures". In: *Illinois Journal of Mathematics* 50.1-4 (2006), pp. 439-472.
- [50] H. Föllmer and T. Knispel. "Entropic risk measures: Coherence vs. convexity, model ambiguity and robust large deviations". In: *Stochastics and Dynamics* 11.02n03 (2011), pp. 333-351.
- [51] H. Föllmer and A. Schied. "Convex measures of risk and trading constraints". In: *Finance and stochastics* 6.4 (2002), pp. 429-447.
- [52] N. Frikha. "Multi-level stochastic approximation algorithms". In: *The Annals of Applied Probability* 26.2 (2016), pp. 933-985.
- [53] M. Frittelli and E. R. Gianin. "Dynamic convex risk measures". In: *Risk measures for the 21st century* (2004), pp. 227-248.
- [54] M. Frittelli and E. R. Gianin. "Law invariant convex risk measures". In: *Advances in mathematical economics*. Springer, 2005, pp. 33-46.
- [55] M. Frittelli and E. R. Gianin. "Putting order in risk measures". In: *Journal of Banking & Finance* 26.7 (2002), pp. 1473-1486.
- [56] M. Frittelli and M. Maggis. "Complete duality for quasiconvex dynamic risk measures on modules of the L_p -type". In: *Statistics & Risk Modeling* 31.1 (2014), pp. 103-128.
- [57] H. Föllmer and A. Schied. "Convex measures of risk and trading constraints". In: *Finance and Stochastics* 6 (2002), pp. 429-447.
- [58] P. Glasserman. *Monte Carlo Methods in Financial Engineering*. Springer, New York, NY, 2004.
- [59] P. Glasserman, W. Kang, and P. Shahabuddin. "Fast simulation of multifactor portfolio credit risk". In: *Operations Research* 56.5 (), pp. 1200-1217.
- [60] R. Griffiths, R. Milne, and R Wood. "Aspects of correlation in bivariate Poisson distributions and processes". In: *Australian Journal of Statistics* 21.3 (1979), pp. 238-255.

-
- [61] A. Gundel. "Robust utility maximization for complete and incomplete market models". In: *Finance and Stochastics* 9.2 (2005), pp. 151-176.
- [62] L. P. Hansen, T. J. Sargent, G. Turmuhambetova, and N. Williams. "Robust control and model misspecification". In: *Journal of Economic Theory* 128.1 (2006), pp. 45-90.
- [63] J. M. Harrison and S. R. Pliska. "A stochastic calculus model of continuous trading: complete markets". In: *Stochastic processes and their applications* 15.3 (1983), pp. 313-316.
- [64] J. M. Harrison and S. R. Pliska. "Martingales and stochastic integrals in the theory of continuous trading". In: *Stochastic processes and their applications* 11.3 (1981), pp. 215-260.
- [65] D. Hernández-Hernández and A. Schied. "Robust utility maximization in a stochastic factor model". In: *Statistics & Risk Modeling* 24.1 (2006), pp. 109-125.
- [66] S. Hodges and A. Neuberger. "Optimal replication of contingent claims under transaction costs". In: *Review Futures Market* 8 (1989), pp. 222-239.
- [67] M.-h. Hsieh and P. W. Glynn. "Confidence regions for stochastic approximation algorithms". In: *Proceedings of the Winter Simulation Conference*. Vol. 1. IEEE. 2002, pp. 370-376.
- [68] Y. Hu, P. Imkeller, and M. Müller. "Utility maximization in incomplete markets". In: *The Annals of Applied Probability* 15.3 (2005), pp. 1691-1712.
- [69] T. R. Hurd, D. Cellai, S. Melnik, and Q. Shao. "Illiquidity and insolvency: a double cascade model of financial crises". In: *Available at SSRN 2424877* (2014).
- [70] J. Jacod and A. Shiryaev. *Limit theorems for stochastic processes*. Vol. 288. Springer Science & Business Media, 2013.
- [71] S. Jaimungal and T. Wang. "Catastrophe options with stochastic interest rates and compound Poisson losses". In: *Insurance: Mathematics and Economics* 38.3 (2006), pp. 469-483.
- [72] M. Jeanblanc, A. Matoussi, and A. Ngoupeyou. "Robust utility maximization problem in a discontinuous filtration". In: *arXiv preprint arXiv:1201.2690* (2013).
- [73] E. Jouini, M. Meddeb, and N. Touzi. "Vector-valued coherent risk measures". In: *Finance and stochastics* 8.4 (2004), pp. 531-552.
- [74] S. Källblad, J. Obłój, and T. Zariphopoulou. "Dynamically consistent investment under model uncertainty: the robust forward criteria". In: *Finance and Stochastics* 22.4 (2018), pp. 879-918.

REFERENCES

- [75] I. Karatzas. "Optimization problems in the theory of continuous trading". In: *SIAM Journal on Control and Optimization* 27.6 (1989), pp. 1221-1259.
- [76] I. Karatzas, J. P. Lehoczky, and S. E. Shreve. "Optimal portfolio and consumption decisions for a "small investor" on a finite horizon". In: *SIAM journal on control and optimization* 25.6 (1987), pp. 1557-1586.
- [77] I. Karatzas, J. P. Lehoczky, S. E. Shreve, and G.-L. Xu. "Martingale and duality methods for utility maximization in an incomplete market". In: *SIAM Journal on Control and optimization* 29.3 (1991), pp. 702-730.
- [78] N. E. Karoui. "Les aspects probabilistes du contrôle stochastique". In: *École d'été de Probabilités de Saint-Flour IX-1979*. Springer, 1981, pp. 73-238.
- [79] N. E. Karoui, A. Matoussi, and A. Ngoupeyou. "Quadratic exponential semimartingales and application to BSDEs with jumps". In: *arXiv preprint arXiv:1603.06191* (2016).
- [80] A. J. Kleywegt, a. Shapiro, and T. Homem de Mello. "The sample average method for stochastic discrete optimization". In: *SIAM Journal on Optimization* 12 (2001), pp. 479-502.
- [81] H. J. Kushner and E. Sanvicente. "Stochastic approximation for constrained systems with observation noise on the system and constraint". In: *Automatica* 11 (1975), pp. 375-380.
- [82] H. J. Kushner and J. Yang. "Stochastic approximation with averaging and feedback: Faster convergence". In: *Adaptive Control, Filtering, and Signal Processing*. Ed. by K. Aström, G. Goodwin, and P. Kumar. Springer, New York, 1995, pp. 205-228.
- [83] H. J. Kushner and G. Yin. *Stochastic approximation and recursive algorithms and applications*. 2nd ed. Springer, New York, NY, 2003.
- [84] R. J. Laeven and M. Stadjé. "Robust portfolio choice and indifference valuation". In: *Mathematics of Operations Research* 39.4 (2014), pp. 1109-1141.
- [85] A. Lazrak and M. C. Quenez. "A generalized stochastic differential utility". In: *Mathematics of operations research* 28.1 (2003), pp. 154-180.
- [86] S. H. Lee. "Systemic liquidity shortages and interbank network structures". In: *Journal of Financial Stability* 9.1 (2013), pp. 1-12.
- [87] V. Lemaire and G. Pagès. "Unconstrained recursive importance sampling". In: *The Annals of Applied Probability* 20.3 (2010), pp. 1029-1067.
- [88] S. Li, C. Yin, X. Zhao, and H. Dai. "Stochastic Interest Model Based on Compound Poisson Process and Applications in Actuarial Science". In: *Mathematical Problems in Engineering* 2017 (2017).

-
- [89] J. T. Linderoth, a. Shapiro, and S. Wright. "The empirical behavior of sampling methods for stochastic programming". In: *Annals of Operations Research* 142 (2006), pp. 215-241.
- [90] L. Ljung. "Analysis of recursive stochastic algorithms". In: *IEEE Transactions on Automatic Control* 22.4 (1977), pp. 551-575.
- [91] F. Maccheroni, M. Marinacci, and A. Rustichini. "Ambiguity aversion, robustness, and the variational representation of preferences". In: *Econometrica* 74.6 (2006), pp. 1447-1498.
- [92] W.-K. Mak, D. P. Morton, and R. K. Wood. "Monte Carlo bounding techniques for determining solution quality in stochastic programs". In: *Operations Research Letters* 24 (1999), pp. 47-56.
- [93] M. Mania and M. Schweizer. "Dynamic exponential utility indifference valuation". In: *The Annals of Applied Probability* 15.3 (2005), pp. 2113-2143.
- [94] H. Markowitz. "Portfolio selection". In: *The Journal of Finance* 7.1 (1952), pp. 77-91.
- [95] G. J. McLachlan and T. Krishnan. *The EM algorithm and extensions*. Vol. 382. John Wiley & Sons, 2007.
- [96] R. C. Merton. "Lifetime portfolio selection under uncertainty: The continuous-time case". In: *The review of Economics and Statistics* (1969), pp. 247-257.
- [97] R. C. Merton. "Optimum consumption and portfolio rules in a continuous-time model". In: *Stochastic optimization models in finance*. Elsevier, 1975, pp. 621-661.
- [98] M.-A. Morlais. "A new existence result for quadratic BSDEs with jumps with application to the utility maximization problem". In: *Stochastic processes and their applications* 120.10 (2010), pp. 1966-1995.
- [99] M.-A. Morlais. "Quadratic BSDEs driven by a continuous martingale and applications to the utility maximization problem". In: *Finance and Stochastics* 13.1 (2009), pp. 121-150.
- [100] M.-A. Morlais. "Utility maximization in a jump market model". In: *Stochastics: An International Journal of Probability and Stochastic Processes* 81.1 (2009), pp. 1-27.
- [101] C. Munk. "The valuation of contingent claims under portfolio constraints: reservation buying and selling prices". In: *Review of Finance* 3.3 (1999), pp. 347-388.
- [102] M. Musiela and T. Zariphopoulou. "A valuation algorithm for indifference prices in incomplete markets". In: *Finance and Stochastics* 8.3 (2004), pp. 399-414.
- [103] A Neufeld. "Antonis Papapantoleon". In: *Finance and Stochastics* 12 (2008), pp. 265-292.
- [104] J. Neveu. *Martingales à temps discret*. Vol. 1. 1. Masson Paris, 1972.

REFERENCES

- [105] T. A. Oigård, A. Hanssen, and R. E. Hansen. "The multivariate normal inverse Gaussian distribution: EM-estimation and analysis of synthetic aperture sonar data". In: *2004 12th European Signal Processing Conference*. IEEE. 2004, pp. 1433-1436.
- [106] E. Pardoux and S. Peng. "Adapted solution of a backward stochastic differential equation". In: *Systems & control letters* 14.1 (1990), pp. 55-61.
- [107] G. C. Pflug and A. Ruszczyński. "A risk measure for income processes". In: (2004).
- [108] B. T. Polyak and A. B. Juditsky. "Acceleration of Stochastic Approximation by Averaging". In: *SIAM Journal on Control and Optimization* 30.4 (1992), pp. 838-855.
- [109] M.-C. Quenez. "Optimal portfolio in a multiple-priors model". In: *Seminar on Stochastic Analysis, Random Fields and Applications IV*. Springer. 2004, pp. 291-321.
- [110] M. M. Rao and Z. D. Ren. *Theory of Orlicz Spaces*. Marcel Dekker Inc. N.Y., 1991.
- [111] H. Robbins and S. Monro. "A stochastic approximation method". In: *The annals of mathematical statistics* (1951), pp. 400-407.
- [112] R. T. Rockafellar and S. Uryasev. "Conditional value-at-risk for general loss distributions". In: *Journal of banking & finance* 26.7 (2002), pp. 1443-1471.
- [113] R. T. Rockafellar and S. Uryasev. "Optimization of conditional value-at-risk". In: *Journal of Risk* 2.3 (2000), pp. 21-41.
- [114] R. Rockafellar. "Integrals which are convex functionals". In: *Pacific journal of mathematics* 24.3 (1968), pp. 525-539.
- [115] R. T. Rockafellar. *Convex Analysis*. Princeton Math. Ser.28. Princeton University Press, Princeton, NJ, 1970.
- [116] R. Rouge and N. El Karoui. "Pricing via utility maximization and entropy". In: *Mathematical Finance* 10.2 (2000), pp. 259-276.
- [117] D. Ruppert. "Stochastic Approximation". In: *Handbook of Sequential Analysis*. Ed. by B. K. Ghosh and P. Sen. Dekker, New York, 1991, pp. 503-529.
- [118] A. Schied. "Optimal investments for risk-and ambiguity-averse preferences: a duality approach". In: *Finance and Stochastics* 11.1 (2007), pp. 107-129.
- [119] A. Schied. "Optimal investments for robust utility functionals in complete market models". In: *Mathematics of Operations Research* 30.3 (2005), pp. 750-764.
- [120] A. Schied and C.-T. Wu. "Duality theory for optimal investments under model uncertainty". In: (2005).
- [121] M. Schroder and C. Skiadas. "Optimal consumption and portfolio selection with stochastic differential utility". In: *Journal of Economic Theory* 89.1 (1999), pp. 68-126.

-
- [122] A. Shapiro, D. Dentcheva, and A. Ruszczyński. *Lectures on stochastic programming: modeling and theory*. SIAM, 2021.
- [123] A. Shapiro and A. Nemirovski. “On complexity of stochastic programming problems”. In: *Continuous Optimization. Current Trends and Applications*. Ed. by V. Jeyakumar and A. Rubinov. Vol. 99. Springer, Boston, MA, 2005, pp. 47-56.
- [124] A. Shapiro, D. Dentcheva, and A. Ruszczyński. *Lectures on Stochastic Programming: Modeling and Theory*. SIAM, 2009.
- [125] A. N. Shiryaev. “Martingales: recent developments, results and applications”. In: *International Statistical Review/Revue Internationale de Statistique* (1981), pp. 199-233.
- [126] C. Skiadas. “Robust control and recursive utility”. In: *Finance and Stochastics* 7.4 (2003), pp. 475-489.
- [127] J.-J. E. Slotine and W. Li. *Applied Nonlinear Control*. Vol. 199. 1. Prentice hall Englewood Cliffs, NJ, 1991.
- [128] D. Talay and Z. Zheng. “Worst case model risk management”. In: *Finance and Stochastics* 6.4 (2002), pp. 517-537.
- [129] B. Verweij, S. Ahmed, A. J. Kleywegt, G. Nemhauser, and A. Shapiro. “The sample average approximation method applied to stochastic routing problems: a computational study”. In: *Computational Optimization and Applications* 24 (2003), pp. 289-333.
- [130] B. Verweij, S. Ahmed, A. J. Kleywegt, G. Nemhauser, and A. Shapiro. “The sample average approximation method applied to stochastic routing problems: a computational study”. In: *Computational Optimization and Applications* 24 (2003), pp. 289-333.
- [131] J. Von Neumann and O. Morgenstern. “Theory of games and economic behavior, 2nd rev”. In: (1947).
- [132] J. Zhang. “Backward stochastic differential equations”. In: *Backward Stochastic Differential Equations*. Springer, 2017, pp. 79-99.
- [133] G. Žitković. “A dual characterization of self-generation and exponential forward performances”. In: *The Annals of Applied Probability* 19.6 (2009), pp. 2176-2210.

Titre : Problème de maximisation d'utilité avec incertitude et algorithmes stochastiques pour les mesures de risque systémiques.

Mots clés : BSDE Quadratiques avec sauts, Maximisation d'utilité, Incertitude, Mesures de Risque Systémiques, Equivalent Certain Optimisé, Algorithmes Stochastiques.

Résumé : Cette thèse est constituée de trois chapitres portant sur la maximisation d'utilité en présence d'incertitude, ainsi que l'étude et l'approximation numérique des mesures de risque systémiques.

Le premier chapitre est consacré à l'étude d'un problème de maximisation d'utilité avec incertitude, en considérant une filtration discontinue et une condition terminale non bornée. L'incertitude sur le modèle est prise en compte dans le problème d'optimisation par l'ajout d'une pénalité générale incluant le cas entropique. Nous montrons l'existence d'un modèle optimal et ensuite nous caractérisons le processus valeur comme la solution unique d'une Equation Différentielle Stochastique Rétrograde Quadratique-Exponentielle avec sauts.

Dans le deuxième chapitre, nous étudions

l'approximation numérique des mesures de risque systémiques de type Shortfall par des algorithmes stochastiques. Nous montrons dans un premier temps la convergence presque-sûre des estimateurs proposés. Des théorèmes centrales limites sont également obtenus permettant d'avoir des intervalles de confiance pour les estimateurs. Enfin, nous implémentons les différents schémas à travers plusieurs exemples.

Enfin dans le dernier chapitre, une nouvelle classe de mesures de risque est introduite, généralisant dans le cadre multivarié les mesures de risque dites *Optimized Certainty Equivalent*. Nous proposons des algorithmes stochastiques pour l'approximation de ces mesures de risque et nous comparons nos résultats avec ceux obtenus par la méthode Monte Carlo.

Title : Utility Maximization Problem with Uncertainty and Stochastic Algorithms for Systemic Risk Measures.

Keywords : Quadratic BSDE with jumps, Robust Utility Maximization problem, Systemic Risk Measures, Optimized Certainty Equivalent, Stochastic Algorithms.

Abstract : This thesis is split into three parts and deals with a robust utility maximization problem, systemic risk measures and their numerical approximation.

In the first part, a utility maximization problem under model uncertainty is addressed in a discontinuous filtration and unbounded terminal condition. The model uncertainty is taken into account in our optimization problem through a general penalty that includes the entropic one. We show the existence of an optimal model and we characterize the value process as the unique solution of an Exponential-Quadratic Backward Stochastic Differential Equation.

In the second part, we are interested by the

numerical approximation of multivariate Shortfall risk measures with stochastic algorithms. We first start by showing the consistency of the estimators. Central limits theorems are also obtained providing us with confidence intervals for the estimators. Lastly, we implement and test our numerical schemes with different examples.

In the last chapter, a new class of risk measures is introduced, generalizing in the multivariate case the *Optimized Certainty Equivalent* risk measures. We propose stochastic algorithms for the approximation of these risk measures and compare our results with those obtained by the Monte Carlo method.