

Essays on Risk Management



Pengyu Wei
Somerville College
University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Trinity 2017

This thesis is dedicated to
my parents
for their love and support throughout my life

Acknowledgements

First and foremost, I would like to thank my parents for their endless love and support.

I would like to thank my supervisor, Prof. Xunyu Zhou, for introducing me into the field of stochastic control and mathematical finance, and guiding and supporting me over the years. Xunyu set an example of excellence as a scholar, mentor and role model. I would also like to thank my co-supervisor, Prof. Hanqing Jin. I enjoyed numerous discussion with Hanqing on convex analysis.

Special thanks to Prof. Ken Seng Tan, who invited me to visit the University of Waterloo and introduced me to the field of actuarial science. I also appreciate the help from Prof. Phelim Boyle.

During my research, I have benefited from fruitful and helpful discussions with many professors, colleagues, and friends. My sincere thanks go to Prof. Suleyman Basak, Prof. Carole Bernard, Prof. Samuel Cohen, Dr. Siyuan Li, Prof. Michael Monoyios, Prof. Xianming Sun, Dr. Wei Wei, Xiaole Xue, Jinggong Zhang, Prof. Shengchao Zhuang, Dr. Ning Wang, seminar participants at OMI, Oxford, Waterloo, and conference participants at the 10th Actuarial and Financial Mathematics Conference in Brussels.

I want to thank the financial support from the Mathematical and Computational Finance Group, Mathematical Institute, Oxford-Man Institute of Quantitative Finance, and Oxford-Nie Financial Big Data Laboratory.

Finally, I would like to thank my beloved Joy.

Abstract

This thesis consists of three parts. The first part studies the optimal portfolio selection of expected utility maximizing investors who must also manage their market-risk exposures. The risk is measured by a so-called weighted Value-at-Risk (WVaR) risk measure, which is a generalization of both Value-at-Risk (VaR) and Expected Shortfall (ES). The feasibility, well-posedness, and existence of the optimal solution are examined. We obtain the optimal solution (when it exists) and show how risk measures change asset allocation patterns.

The second part analyses the impact of ES-based market-risk regulation on portfolio choice and asset prices. We study the optimal, dynamic portfolio and wealth/consumption policies of expected utility maximizing investors who must also manage market-risk exposure which is measured by Expected Shortfall (ES). We find that ES managers can incur larger losses when losses occur, compared to both VaR and benchmark managers. A general-equilibrium analysis reveals that the presence of ES managers increases the market volatility during periods of significant financial market stress, in both pure-exchange and production economies.

The third part studies the optimal dynamic reinsurance policy for an insurance company whose surplus is modeled by the diffusion approximation of the classical Cramér-Lundberg model. We assume the reinsurance premium is calculated according to the Mean-CVaR premium principle which generalizes Denneberg's absolute deviation principle and expected value principle. Moreover, we require that both the ceded loss and retention functions are non-decreasing to rule out the moral hazard. Under the objective of minimizing the ruin probability, we obtain the optimal reinsurance policy explicitly, which is more complicated than the contracts widely studied in the dynamic reinsurance literature.

Contents

1	Introduction	1
2	Risk Management with Weighted VaR	8
2.1	Risk Measures	9
2.1.1	VaR, ES, and WVaR	10
2.1.2	Spectral Risk Measures	11
2.1.3	Distortion Risk Measures	13
2.2	Model	14
2.2.1	Market	14
2.2.2	Benchmark Agent	16
2.2.3	Value-at-Risk-based Risk Management	16
2.2.4	Risk Management with Weighted VaR	17
2.3	Solution	17
2.3.1	Quantile Formulation	18
2.3.2	Feasibility and Well-Posedness	20
2.3.3	Optimal Solution	22
2.3.4	Proof of Theorem 2.3.1	25
2.3.4.1	Lagrange Approach	25
2.3.4.2	Lagrangian Dual Problem	27
2.3.4.3	Existence of Lagrange Multipliers	30
2.3.4.4	Optimal Solution	37
2.3.5	An Example	38
2.4	Properties	41
2.4.1	Existence of Optimal Solutions	41
2.4.2	Impacts on Asset Allocation	43
2.5	Concluding Remarks	50

3	Equilibrium Analysis of Expected Shortfall	51
3.1	Model	52
3.2	Partial Equilibrium	52
3.2.1	Optimal Solution	53
3.2.2	Proof of Proposition 3.2.1	54
3.2.3	Properties	57
3.2.3.1	Terminal Wealth	57
3.2.3.2	Time- t Wealth and Portfolio Weights	60
3.3	General Equilibrium Analysis	64
3.3.1	Equilibrium in a Production Economy	64
3.3.2	Equilibrium in a Pure-Exchange Economy	67
3.4	Concluding Remarks	76
4	Optimal Dynamic Reinsurance Policies Under Mean-CVaR Principle	77
4.1	Denneberg's Absolute Deviation Principle	78
4.2	Dynamic Reinsurance Model	80
4.3	Solution	83
4.3.1	HJB Equation	83
4.3.2	Relaxed Solutions	83
4.3.3	Modified Solutions	85
4.3.4	Verification Theorem	88
4.4	Concluding Remarks	92
5	Conclusion	94
A	Appendix to Chapter 2	96
A.1	Boundary Solution	96
	Bibliography	98

List of Figures

2.1	$\varphi_{\lambda_2}(\cdot)$ and its concave envelope $\delta_{\lambda_2}(\cdot)$	39
2.2	Optimal terminal wealth under ES	40
2.3	Optimal terminal wealth under the Wang (2000) risk measure	49
3.1	Optimal horizon wealth	58
3.2	Optimal time-t wealth	63
3.3	Time-t relative risk exposure	64
3.4	Time-t market price	74
3.5	Time-t relative risk exposure	75
4.1	Excess-of-loss over a cap	92
4.2	Excess-of-loss over a layer	93

Chapter 1

Introduction

Risk management is a constant challenge to financial institutions. The objectives and tools for risk management vary across different financial institutions. An investment institution (investor) faces primarily the financial risk and allocates its assets to seek a balance between return and risk that best suits its preference, which is typically described by a utility function in the rational choice theory, whereas an insurer manages its insurance risk by means of reinsurance to maintain its business viability, which can be measured by the ruin probability. This thesis studies risk management for financial institutions from the perspectives of asset allocation and reinsurance, respectively. Chapter 2 and 3 focus on the impact of risk measures on portfolio selection and assets prices, and Chapter 4 studies optimal reinsurance for insurers.

In Chapter 2, we address the optimal portfolio selection of expected utility maximizing investors who must also manage their market-risk exposures in a continuous-time, complete market. The risk is quantified by the weighted Value-at-Risk (WVaR) proposed in He et al. (2015), which is a generalization of both Value-at-Risk (VaR) and Expected Shortfall (ES) and covers spectral risk measures, distortion risk measures and many law-invariant coherent risk measures. This is motivated by the recent advancement in risk measurement.

VaR has long been an industry standard, whether by choice or by regulation (Jorion (1997); SEC (1997); Dowd (1998); Saunders (2000); Jorion (2002); BCBS (2011)). However, since its introduction in approximately 1994, VaR has been criticized in both academia and industry, for its weaknesses as a benchmark. VaR fails to capture “tail risk” and it is not subadditive, defying the notion of diversification. Recognizing the shortcomings of VaR, Artzner et al. (1999) argue that a good risk measure should satisfy a set of reasonable axioms, leading to the so-called coherent risk measures. Recently there has been a movement in both academia (Artzner

et al. (1999); Rockafellar and Uryasev (2000); Acerbi and Tasche (2002); Rockafellar and Uryasev (2002); Alexander and Baptista (2006); Embrechts et al. (2014)) and industry (BCBS (2016)) to replace VaR with ES.¹ Simultaneously, alternative risk measures, such as spectral risk measures and distortion risk measures, have arisen in the portfolio selection literature (Acerbi and Simonetti (2002); Adam et al. (2008); Sereda et al. (2010)). However, despite rich research on risk measurement, little is known about the effects of these risk measures on portfolio selection, especially in the expected utility maximization framework. Chapter 2 contributes to filling that gap.

We first solve the problem completely, with the help of the so-called quantile formulation, which was developed recently mainly in the context of behavioral portfolio selection. Feasibility, well-posedness, and attainability are examined in greater detail. These issues have been more or less overlooked in the related literature. We propose the notion of *risk reduction per cost* (RRPC), which depends on only the risk measure and the market, to measure the tradeoff between reducing the risk and incurring the cost. We find that if a risk measure's RRPC for extreme gains is infinite, then the model is unattainable whenever the constraint is non-redundant, i.e., the optimal value is finite but is not achievable by any portfolios, indicating that the model is misformulated.

We then study how different risk measures can change optimal trading patterns when the optimal solution exists. In particular, we characterize two classes of risk measures. On the one hand, if a risk measure's RRPC for extreme losses is infinite, then the agent will follow a trading strategy that creates endogenous portfolio insurance, shielding himself from large losses as it is the most efficient way to meet the requirement. This could be of particular interest to regulators. In general, portfolio insurance is costly, and it is highly unlikely that the agent will utilize such a strategy. Regulators can encourage economic agents to use the portfolio insurance strategy by imposing a constraint within this type, such as the Wang (2000) risk measure and the beta family of distortion risk measures. On the other hand, if a risk measure's RRPC for extreme losses is 0, then the agent will ignore losses in the bad states and leave himself completely uninsured because it is costly and inefficient to insure against such losses. Moreover, a further inspection reveals that these risk measures allow economic agents to engage in the "regulatory capital arbitrage" defined by Jones (2000), in the sense that if a large loss occurs, it is likely to be even larger than it

¹For example, as stated in BCBS (2016), one of the key enhancements in the revised market risk framework is a *shift from Value-at-Risk (VaR) to an Expected Shortfall (ES) measure of risk under stress. Use of ES will help to ensure a more prudent capture of "tail risk" and capital adequacy during periods of significant financial market stress.*

would have been in the absence of the risk constraint. This could be a source of concern for both regulators and real-world risk managers. Risk measures are viewed by many as a tool to protect economic agents from large losses, which could cause credit and solvency problems. However, risk measures within this type, such as VaR and ES, actually defeat the purpose of such regulations. Basak and Shapiro (2001) obtain similar results for VaR, but to the best of our knowledge, we are the first to characterize risk measures that can create endogenous portfolio insurance or lead to “regulatory capital arbitrage” within a relatively general class of risk measures.

Chapter 2 contributes to the literature on utility maximization with risk constraints. Basak and Shapiro (2001) consider utility maximization in a continuous-time complete market, assuming that agents must limit their risks as measured by VaR. Basak et al. (2006) study the portfolio choice problem in which VaR is evaluated relative to a benchmark. Gabih et al. (2005, 2009) consider the problem when the risk is measured by either VaR or the expected loss. Gundel and Weber (2008) generalize VaR to a class of convex risk measures. Rogers (2009) considers law-invariant coherent risk measures. Cahuich and Hernández-Hernández (2013) study rank-dependent utility maximization with a risk constraint. However, all of the above papers, except for those by Basak and Shapiro (2001) and Basak et al. (2006), are more of a mathematical treatment of the problem with little analysis of its economic implications. In particular, less attention has been paid to how various risk measures can alter portfolio choice patterns.

A different approach in portfolio selection, pioneered by Markowitz (1952), is to use the mean and variance to measure the return and risk, respectively, and the economic agent chooses among the portfolios that yield a pre-specified level of expected return while minimizing the variance of the portfolio’s return. In addition to variance, researchers have also considered alternative risk measures (Kast et al. (1999); Rockafellar and Uryasev (2000); Campbell et al. (2001); Rockafellar and Uryasev (2002); Acerbi and Simonetti (2002); Alexander and Baptista (2002, 2004); Adam et al. (2008)). These researchers all study single-period mean-risk portfolio selection problems. There have also been extensions of the mean-risk model from the single-period setting to the dynamic, continuous-time setting, including but not limited to Bielecki et al. (2005); Jin et al. (2005); Basak and Chabakauri (2010), etc. In particular, He et al. (2015) recently considered continuous-time mean-risk portfolio choice problems with WVaR. They found that the model is prone to being ill-posed, especially when bankruptcy is allowed, leading to extreme risk-taking behaviors.

Chapter 2 makes two main contributions. First, by completely solving the corresponding constrained continuous-time utility maximization problem, we can characterize the optimal terminal wealth when the risk is measured by various risk measures. Our model offers a variety of attractive features compared to its mean-WVaR counterpart (He et al. (2015)). In particular, we provide a critique of the current risk management practices and especially the new Basel Accord (BCBS (2016)).

Second, the technical analysis performed in Chapter 2 contributes to the mathematical aspects of the portfolio selection literature. Typically, the continuous-time utility maximization problem is solved by the Lagrange dual method. We find that, with an additional constraint on the risk, the dual method can fail under some circumstances; in other words, the optimal solution to the original problem can exist but might not be given by the Lagrange dual problem. After establishing the relationship between the original problem and the Lagrange dual problem, we solve the dual problem completely. Although the main techniques for solving the dual problem, i.e., quantile formulation and the concave envelope relaxation, have been employed in the literature (Rogers (2009); Xia and Zhou (2016); Xu (2016)), the existence of optimal solutions requires a more nuanced analysis. For example, in finding the Lagrange multipliers, we characterize the monotonicity and continuity of a function's concave envelope's right derivative with respect to the original function's parameter.

In a partial equilibrium setting, Chapter 2 shows that the purpose of the regulation incentives can be defeated not only if risk constraints are given in terms of the often criticized VaR but also in case of the recently recommended ES. We continue this line of research in Chapter 3 by performing a detailed analysis of ES-based risk management. We first study dynamic portfolio selection of expected utility maximizing investors who must also manage market-risk exposure measured by ES. Based on the general solution in Chapter 2, we derive the optimal terminal wealth under the ES constraint explicitly. Consistent with our previous findings, the ES manager optimally chooses to incur larger losses when losses occur. Under the additional assumptions that the state price density is lognormally distributed and the preference is characterized by the CRRA utility functions, we also derive the optimal portfolio weights in closed-form. When the state price density is large enough, the ES manager invests a larger portion of his wealth in the risky assets compared to a benchmark agent who does not need to meet the ES constraint.

We then extend our analysis to a general equilibrium setting. In particular, we study how ES affects asset prices in the general equilibrium. Abundant work has

been done on the impact of different risk measures on portfolio selection in the partial equilibrium. However, general equilibrium analyses of risk measures on market dynamics are limited, except Basak and Shapiro (2001); Leippold et al. (2006) for VaR. To our best knowledge, we are the first to study the impact of ES on market dynamics. We develop a production and a pure-exchange general equilibrium model featuring ES risk managers, respectively. We find that ES managers optimally choose larger risk exposure and thus increase the market volatility during periods of significant financial market stress in both economies. These findings are similar to those of VaR. One common criticism on VaR is that it fails to take the magnitude of losses into account. Although ES takes the sizes of losses into account, it does not suffice to alleviate risk-taking behaviors.

The problem studied in Chapter 2 is closely related to Basak and Shapiro (2001); Leippold et al. (2006). Basak and Shapiro (2001) study the effects of VaR on optimal wealth and consumption policies of risk managers. Leippold et al. (2006) study the asset-pricing implications of dynamic VaR regulation in incomplete continuous-time economies with stochastic opportunity set and heterogeneous attitudes to risk. In contrast, we analyze the asset pricing implications of ES, in both pure-exchange and production economies.

While portfolio selection is important to financial institutions, reinsurance is the most common tool for insurance companies to manage their risk exposures. Chapter 4 considers the risk management for an insurer from the perspective of optimal reinsurance. Reinsurance is a contract between an insurer (the “cedent”) and a reinsurer. The insurer diverts a share of its incoming claims to the reinsurer. In return, the reinsurer is paid a premium which is calculated according to some premium principles. There is a voluminous optimal reinsurance literature examining how an insurance company should choose a reinsurance strategy to minimize its ruin probability, which is a measure that allows the insurer to indicate whether the collected premiums are sufficient to pay the claims. A fundamental assumption in the bulk of this line of research is that the expected value principle is used in calculating the premium and the insurer chooses either the proportional reinsurance (Schmidli (2001, 2002); Taksar and Markussen (2003); Promislow and Young (2005); Luo et al. (2008); Zhang et al. (2016)) or the excess-of-loss reinsurance (Hipp and Vogt (2003); Dickson and Waters (2006); Zhang et al. (2016)). However, their settings do not allow for the determination of the optimal shape of the reinsurance contract. Meng and Zhang (2010) and Hipp and Taksar (2010) show that the excess-of-loss reinsurance treaty is optimal among the class of plausible reinsurance treaties under the expected value

principle. Hipp and Taksar (2010) find that the proportional reinsurance treaty is optimal under the variance premium principle. Nevertheless, it is unknown whether the excess-of-loss reinsurance or the proportional reinsurance are still optimal under premium principles other than the expected value principle or variance premium principle. Chapter 4 is partly to fill that gap.

Premium principle should satisfy a set of elementary and plausible requirements: (1) the safety loading (premium minus expected value of the loss) is nonnegative, (2) the premium cannot exceed the maximal loss, (3) the premium for $X + a$ should be the premium for X increased by that fixed amount a , and (4) the premium for doubling a risk is twice the premium of the single risk. Denneberg (1990) argues that one should replace the standard deviation with average absolute deviation in calculating the premium, as the standard deviation principle violate the second requirement that the premium cannot exceed the maximum loss, i.e. no *rip-off*.² However, the Denneberg's absolute deviation principle can recover the widely used expected value principle only if the probability of no claim is larger than one half. Inspired by Denneberg's absolute deviation principle, we propose the Mean-CVaR premium principle which is a generalization of Denneberg's absolute deviation principle and expected value principle. Specifically, the Mean-CVaR premium is a weighted average of the risk's mean and CVaR at the confidence level α . One advantage of this premium principle is its flexibility. The confidence level α measures the degree to which the reinsurer cares about the tail risk. The *premium loading* θ is analogous to the safety loading in the classical expected value principle. The *risk loading* β measures the relative importance of the tail risk to the mean. Each combination of these parameters reflects a specific reinsurer's preference.

We consider the reinsurance design problem for an insurer in a dynamic setting. The surplus of the insurer is modeled by the diffusion approximation of the classical Cramér-Lundberg model. To minimize the reinsurance risk, i.e., the ruin probability, the insurer can reinsure part of each claim. In return, the insurer must pay premiums to the reinsurer and we assume the premiums are calculated according to the Mean-CVaR premium principle. In addition, we impose the monotonicity constraint that both the insurer and reinsurer need to pay more for a larger loss. This constraint completely rules out the *moral hazard* in the sense that the reinsurance contract creates the opportunity for the insurer to mis-report its actual losses to the reinsurer; yet

²In fact, the variance premium principle also suffers from the problem of rip-off.

mathematically it gives rise to substantial difficulty in solving the optimization problem. As far as we know, this constraint has not appeared in the dynamic reinsurance literature.

We use dynamic programming to solve the problem and obtain the optimal contract and value function explicitly. We find that, under the Mean-CVaR premium principles, the optimal reinsurance contracts have at most two different shapes depending on the relationship between the premium loading θ and risk loading β . In both cases, the insurer prefers to reinsure against large losses which can lead to insolvency, because the objective of the insurer is to minimize the ruin probability. If $\theta \leq \beta$, the reinsurer cares more about the tail risk and the optimal contract, *excess-of-loss over a cap*, resembles a cap reinsurance contract for small losses and then an excess-of-loss treaty for large losses. Under this contract, small claims are fully covered. If $\theta > \beta$, i.e., it is costly to insure against small claims, the optimal coverage shifts from small losses to medium losses. The optimal contract, *excess-of-loss over a layer*, has a layer above which it resembles an excess-of-loss arrangement. These results are completely new in the dynamic reinsurance literature.

The main contributions of this chapter are as follows. First, we propose the Mean-CVaR premium principle which generalizes the expected value principle and Denneberg's absolute deviation principle. The new premium principle is flexible enough to reflect the reinsurer's risk attitude. Second, we study dynamic reinsurance design beyond the standard expected value principle and variance principle. Moreover, we require the insurer and reinsurer are obligated to pay more for a larger loss to rule out moral hazard. These new features can provide new directions in the future dynamic reinsurance research. Third, we derive two types of the optimal contracts explicitly which are completely new in the dynamic reinsurance literature. The forms of these contracts are interesting in their own rights.

Chapter 5 concludes the thesis.

Chapter 2

Risk Management with Weighted VaR¹

This chapter addresses the portfolio selection problem of utility maximizing investors under a joint budget and risk constraint in the framework of a complete continuous-time financial market model. The risk is quantified by the weighted Value-at-Risk (WVaR) proposed in He et al. (2015), which is a generalization of both VaR and ES and covers spectral risk measures, distortion risk measures, and many law-invariant coherent risk measures.

The constrained optimization problem is solved by the Lagrange dual method with the help of the quantile formulation and the concave envelope relaxation. The technical contribution of this chapter is a detailed analysis of feasibility and existence of optimal solutions and the related Lagrange multipliers.

In addition to the mathematical treatment of the portfolio selection problem, we also present a study of the economic implications of different risk measures through the notion of *risk reduction per cost* (RRPC). The main contribution of this chapter is the characterization of three classes of risk measures: if a risk measure's RRPC for extreme gains is infinite, then the model is unattainable whenever the constraint is non-redundant, i.e., the optimal value is finite but is not achievable by any portfolios; if a risk measure's RRPC for extreme losses is infinite, then the agent will follow a trading strategy that creates endogenous portfolio insurance, shielding himself from large losses; if a risk measure's RRPC for extreme losses is 0, then the agent will ignore losses in the bad states and engage in the “regulatory capital arbitrage”, in the sense that if a large loss occurs, it is likely to be even larger than it would have been in the absence of the risk constraint.

¹This chapter is based on my paper Wei (2017b) forthcoming in *Mathematical Finance*.

The remainder of this chapter is organized as follows: In Section 2.1, we introduce WVaR. In Section 2.2, we formulate the risk management with weighted VaR (WVaR-RM) problem, which is then solved completely in Section 2.3. Impacts on portfolio choice are analyzed in Section 2.4. Section 2.5 concludes this chapter.

2.1 Risk Measures

In this section, we introduce the WVaR risk measure. Consider a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, and denote by $L^\infty := L^\infty(\Omega, \mathcal{F}, \mathbb{P})$ the set of all \mathbb{P} -essentially bounded random variables; in addition, $LB := LB(\Omega, \mathcal{F}, \mathbb{P})$, the set of all lower-bounded finite-valued random variables (but different random variables could have different lower bounds). Let $X \in LB$ represent the profit and loss (P&L) of an investment.

Let \mathbb{F} be the set of cumulative distribution functions (CDFs hereafter) of all lower bounded random variables that take values in \mathbb{R} , in other words,

$$\mathbb{F} = \{F(\cdot) : \mathbb{R} \rightarrow [0, 1], \text{ nondecreasing, right continuous, } \\ F(a) = 0 \text{ for some } a \in \mathbb{R} \text{ and } F(+\infty) = 1\}.$$

The lower boundedness above corresponds to the required tameness of the portfolios (to be discussed in the next section). For any $F(\cdot) \in \mathbb{F}$, denote by $F^{-1}(\cdot)$ its right-inverse, in other words,

$$F^{-1}(t) = \inf\{x \in \mathbb{R} : F(x) > t\} = \sup\{x \in \mathbb{R} : F(x) \leq t\}, \quad t \in [0, 1).$$

Let $\mathbb{G} := \{F^{-1}(\cdot)(t) : F(\cdot) \in \mathbb{F}\}$ be the corresponding set of quantile functions, or

$$\mathbb{G} = \{G(\cdot) : [0, 1) \rightarrow \mathbb{R}, \text{ nondecreasing, right continuous with left limits (RCLL)}\},$$

where $G(1) := G(1-)$. Throughout the chapter, we will use $F_X(\cdot)$ and $G_X(\cdot)$ to denote a random variable X 's CDF and quantile function respectively.

Consider a functional over the set of random P&Ls, $\rho : X \rightarrow \rho(X) \in \mathbb{R}$. It may fulfill some of the following axioms:

- A1 *Monotonicity*: $\rho(X) \geq \rho(Y)$ for any $X, Y \in LB$ such that $X \leq Y$;
- A2 *Translation-invariance*: $\rho(X + a) = \rho(X) - a$ for any $X \in LB$ and $a \in \mathbb{R}$;
- A3 *Truncation continuity*: $\rho(X) = \lim_{n \rightarrow +\infty} \rho(X \wedge n)$ for any $X \in LB$;
- A4 *Positive homogeneity*: $\rho(\lambda X) = \lambda \rho(X)$ for any $X \in LB$ and $\lambda > 0$;

A5 *Sub-additivity*: $\rho(X + Y) \leq \rho(X) + \rho(Y)$ for any $X, Y \in LB$;

A6 *Convexity*:² $\rho(\alpha X + (1 - \alpha)Y) \leq \alpha\rho(X) + (1 - \alpha)\rho(Y)$ for any $X, Y \in LB$ and $\alpha \in (0, 1)$;

A7 *Law-invariance*: $\rho(X) = \rho(Y)$ for any $X, Y \in LB$ with the same distribution function;

A8 *Comonotonic additivity*: $\rho(X + Y) = \rho(X) + \rho(Y)$ for any $X, Y \in LB$ such that X and Y are comonotonic, i.e.,

$$(X(\omega) - X(\omega'))(Y(\omega) - Y(\omega')) \geq 0 \text{ for all } (\omega, \omega') \in \Omega \times \Omega.$$

ρ is called a risk measure if it satisfies A1-A3³ and it is a coherent risk measure if it satisfies A1-A5, as introduced in Artzner et al. (1999).

Based on the above axioms, we introduce WVaR as a generalization of VaR and ES, two popular regulatory risk measures.

2.1.1 VaR, ES, and WVaR

VaR is one of the most important measures (if not the most used measure) of risk in finance. VaR describes the loss that can occur over a given period, at a given confidence level. It has long been an industry standard, whether by choice or by regulation (Jorion (1997); SEC (1997); Dowd (1998); Saunders (2000); BCBS (2011)). The VaR at a specified threshold $\alpha \in (0, 1)$ is defined as

$$\text{VaR}_\alpha(X) = -G_X(\alpha).$$

Since its introduction in approximately 1994, VaR has been criticized in both academia and industry, for its weaknesses as a benchmark. VaR fails to capture “tail risk” and it is not subadditive, which means that the risk of a portfolio can be larger than the sum of the stand-alone risks of its components when measured by VaR (Artzner et al. (1999)).

Recognizing the shortcomings of VaR, there has been a movement to replace VaR with ES, also known as Average Value-at-Risk (AVaR) and Conditional Value-at-Risk

²A4 and A5 imply A6.

³In the literature, risk measures are often defined on L^∞ and are assumed to satisfy only A1-A2. In this chapter, we consider risk measures on LB as in He et al. (2015). The truncation continuity is thus imposed to guarantee that the risk of any unbounded P&L can be computed through its truncations. To verify that ρ satisfies some of the aforementioned axioms, it suffices to show that these axioms are satisfied when ρ is restricted on L^∞ and the truncation continuity is satisfied.

(CVaR), in both academia (Artzner et al. (1999); Rockafellar and Uryasev (2000); Acerbi and Tasche (2002); Rockafellar and Uryasev (2002); Alexander and Baptista (2006); Embrechts et al. (2014)) and industry (BCBS (2016)).⁴ ES measures the riskiness of a position by considering both the size and the likelihood of losses above a certain confidence level. The ES at level $\alpha \in (0, 1)$ is defined as

$$\text{ES}_\alpha(X) = -\frac{1}{\alpha} \int_0^\alpha G_X(z) dz.$$

ES satisfies A1-A8, and it is the smallest coherent, comonotonic additive and law-invariant risk measure that dominates VaR, according to Dhaene et al. (2004). Moreover, ES is in accordance with second stochastic dominance, according to Bertsimas et al. (2004) and Leitner (2005).

In this chapter, we focus on the weighted VaR (WVaR) risk measures⁵ introduced in He et al. (2015), which take the following form

$$\rho_\Phi(X) = - \int_{[0,1]} G_X(z) \Phi(dz), \quad (2.1)$$

where $\Phi \in P[0, 1]$, and $P[0, 1]$ is the set of all probability measures on $[0, 1]$. This is a generalization of VaR and ES: when Φ is a Dirac measure, it becomes VaR; when $\Phi([0, z]) = \frac{z}{\alpha} \wedge 1, z \in [0, 1]$, it becomes ES_α . WVaR is a law-invariant comonotonic additive risk measure, and it covers a large class of law-invariant coherent risk measures and all law-invariant risk measures that are both convex and comonotonic additive. See He et al. (2015) for a more detailed discussion.

We now use examples to show how WVaR generalizes many well-known risk measures that are widely used in finance and actuarial sciences. The generality of WVaR allows us to solve the risk-constrained utility maximization problem in a unified framework. Readers who are already familiar with risk measures can skip the remainder of this section.

2.1.2 Spectral Risk Measures

Spectral risk measures, proposed in Acerbi (2002), cover a large class of coherent risk measures. One distinctive feature of such risk measures is that they map any rational investor's subjective risk aversion onto a coherent measure and vice-versa.

⁴If the underlying distribution of X is a continuous distribution then ES is equivalent to Tail Conditional Expectation which is defined as $\text{TCE}_\alpha(X) = -\mathbb{E}[X|X \leq -\text{VaR}_\alpha(X)]$.

⁵Cherny (2006) also proposed a version of weighted VaR risk measures, which is a special case of He et al. (2015).

Definition 2.1.1. An element $\phi \in L^1([0, 1])$ is said to be an admissible risk spectrum if

1. ϕ is non-negative;
2. ϕ is non-increasing;
3. $\|\phi\| := \int_0^1 \phi(z) dz = 1$.

The spectral risk measures are defined as

$$M_\phi(X) = - \int_0^1 F_X^{-1}(z) \phi(z) dz = - \int_0^1 G_X(z) \phi(z) dz$$

for each admissible risk spectrum ϕ , and ϕ is also called a risk-aversion function. It is easy to see that spectral risk measures are special examples of WVaR. Spectral risk measures are law-invariant, convex, and comonotonic additive.

Examples of spectral risk measures include:

1. The ES_α has the spectrum given by

$$\phi(z) = \begin{cases} \frac{1}{\alpha}, & z \in [0, \alpha]; \\ 0, & z \in (\alpha, 1]. \end{cases}$$

2. The exponential spectral risk measures, proposed in Cotter and Dowd (2006) and Dowd et al. (2008), have spectrums given by

$$\phi(z) = \frac{Re^{-Rz}}{1 - e^{-R}},$$

where $R > 0$ is the investor's absolute risk aversion coefficient.

3. The power spectral risk measures, proposed in Dowd et al. (2008), have spectrums given by

$$\phi(z) = \begin{cases} \gamma z^{\gamma-1}, & 0 < \gamma < 1; \\ \gamma(1-z)^{\gamma-1}, & \gamma > 1, \end{cases}$$

where γ is the investor's relative risk aversion coefficient.

2.1.3 Distortion Risk Measures

Distortion risk measures originated from Yaari's dual theory of choice under risk (Yaari (1987)). Yaari's idea consists of measuring risk by applying a distortion function g on F_X ,

$$\rho_g(X) = - \int_{-\infty}^{\infty} x d(g(F_X(x))) = - \int_0^1 G_X(z) dg(z),$$

where the distortion function g is continuous, non-decreasing and satisfies $g(0) = 0$ and $g(1) = 1$. It is a concave distortion risk measure if g is further concave. A distortion risk measure is coherent if and only if it is a concave distortion risk measure, as in Sereda et al. (2010). Distortion risk measures are also special cases of WVaR.

Distortion risk measures are widely used in insurance and actuarial sciences. For a comprehensive review, please refer to Wirch and Hardy (1999) and Sereda et al. (2010). Some well-known examples are presented below:

1. The (negative) expectation risk measure uses $g(z) = z$.
2. The Wang (2000) risk measure uses

$$g(z) = \Phi_N(\Phi_N^{-1}(z) - \Phi_N^{-1}(q)),$$

where $0 < q \leq 0.5$ is a given parameter, and Φ_N is the standard normal distribution function. The Wang (2000) risk measure is a concave distortion risk measure.

3. The beta family of distortion risk measures, proposed in Wirch and Hardy (1999), uses the distribution function of the beta distribution

$$g(z) = \int_0^z \frac{1}{\beta(a, b)} t^{a-1} (1-t)^{b-1} dt,$$

where $\beta(a, b)$ is the beta function with parameters $a > 0$ and $b > 0$. It is concave if and only if $a \leq 1$ and $b \geq 1$; and it is strictly concave if $a < 1$ and $b > 1$.

4. The proportional hazard (PH) risk measure, discussed in Wang (1995, 1996) and Wang et al. (1997), uses

$$g(z) = z^{\frac{1}{\gamma}}, \quad \gamma \geq 1.$$

This is an example of the beta family of distortion risk measures, with $a = \frac{1}{\gamma}$, $b = 1$.

5. The dual power risk measure uses

$$g(z) = 1 - (1 - z)^\kappa, \quad \kappa \geq 1.$$

It is also an example of the beta family of distortion risk measures, with $a = 1, b = \kappa$.

We see that power spectral risk measures, although they arise from a different context, are in fact proportional hazard/dual power risk measures.

2.2 Model

In this section, we formulate our risk management with weighted VaR (WVaR-RM) problem.

2.2.1 Market

Let $T > 0$ be a given terminal time, and let $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \in [0, T]}, \mathbb{P})$ be a filtered probability space on which we define a standard $(\mathcal{F}_t)_{t \in [0, T]}$ -adapted n -dimensional Brownian motion $W(t) \equiv (W^1(t), \dots, W^n(t))^\top$ with $W(0) = 0$, and hence, the probability space is atomless. It is assumed that $\mathcal{F}_t = \sigma\{W(s) : 0 \leq s \leq t\}$, augmented by all the \mathbb{P} -null sets in \mathcal{F} . Here and henceforth, A^\top denotes the transpose of a matrix A . We define a continuous-time financial market following Karatzas and Shreve (1998). In the market, $n + 1$ assets are being traded continuously. One of the assets is a bank account whose price process $S_0(t)$ is subject to the following equation

$$dS_0(t) = r(t)S_0(t)dt, \quad t \in [0, T]; \quad S_0(0) = s_0 > 0, \quad (2.2)$$

where the interest rate $r(\cdot)$ is a uniformly bounded, $(\mathcal{F}_t)_{t \in [0, T]}$ -progressively measurable, scalar-valued stochastic process. The other n assets are risky securities whose price processes $S_i(t)$, $i = 1, \dots, n$, satisfy the following stochastic differential equation (SDE)

$$dS_i(t) = S_i(t)[b_i(t)dt + \sum_{j=1}^n \sigma_{ij}(t)dW^j(t)], \quad t \in [0, T]; \quad S_i(0) = s_i > 0, \quad (2.3)$$

where $b_i(\cdot)$ and $\sigma_{ij}(\cdot)$, the appreciation and volatility rates respectively, are scalar-valued, $(\mathcal{F}_t)_{t \in [0, T]}$ -progressively measurable stochastic processes with

$$\int_0^T \left[\sum_{i=1}^n |b_i(t)| + \sum_{i=1}^n \sum_{j=1}^n |\sigma_{ij}(t)|^2 \right] ds < \infty, \quad a.s.$$

Set the excess rate of return process as

$$B(t) := (b_1(t) - r(t), \dots, b_n(t) - r(t))^\top,$$

and define the volatility matrix process $\sigma(t) := (\sigma_{ij}(t))_{n \times n}$. The basic assumptions imposed on the market parameters throughout this chapter are summarized as follows,

Assumption 2.2.1. *There is a unique, $(\mathcal{F}_t)_{t \in [0, T]}$ -progressively measurable, \mathbb{R}^n -valued process $\theta(t)$ with $\mathbb{E} e^{\frac{1}{2} \int_0^T |\theta(t)|^2 dt} < \infty$ such that*

$$\sigma(t)\theta(t) = B(t), \text{ a.s., a.e. } t \in [0, T].$$

Consequently, we have a complete model of a securities market.

Consider an economic agent, with an initial endowment $x > 0$ and an investment horizon $[0, T]$, whose total wealth at time $t \geq 0$ is denoted by $X(t)$. Assume that the trading of shares occurs continuously in a self-financing fashion and there are no transaction costs. Then, $X(\cdot)$ satisfies

$$dX(t) = [r(t)X(t) + B(t)^\top \pi(t)]dt + \pi(t)^\top \sigma(t)dW(t), \quad t \in [0, T]; \quad X(0) = x, \quad (2.4)$$

where $\pi_i(t)$ denotes the total market value of the agent's wealth in stock i at time t . The process $\pi(\cdot) \equiv (\pi_1(\cdot), \dots, \pi_n(\cdot))$ is called an admissible portfolio if it is $(\mathcal{F}_t)_{t \in [0, T]}$ -progressively measurable with

$$\int_0^T |\sigma(t)^\top \pi(t)|^2 dt < \infty \text{ and } \int_0^T |B(t)^\top \pi(t)| dt < \infty, \text{ a.s.}$$

and is tame, i.e., the corresponding wealth process $X(\cdot)$ is almost surely bounded from below, although the bound could depend on $\pi(\cdot)$. It is standard in the continuous-time portfolio choice literature for a portfolio to be required to be tame, enabling, among other things, the exclusion of the doubling strategy.

With the complete market assumption, we can define the pricing kernel or state price density process as

$$\xi(t) := \exp \left\{ - \int_0^t [r(s) + \frac{1}{2} |\theta(s)|^2] ds - \int_0^t \theta(s)^\top dW(s) \right\}.$$

Let $\xi := \xi(T)$. It is clear that under Assumption 2.2.1 and the uniform boundedness of $r(\cdot)$, $0 < \xi < +\infty$ a.s. and $0 < \mathbb{E}\xi < +\infty$. Then, in view of the standard martingale approach (Pliska (1986); Cox and Huang (1989); Karatzas and Shreve (1998)), finding the optimal portfolio in this economy is equivalent to finding the optimal terminal wealth.

2.2.2 Benchmark Agent

Let $X := X(T)$. The benchmark model is the standard utility maximization problem,

$$\begin{aligned} \max_X \quad & \mathbb{E}[u(X)] \\ \text{subject to} \quad & \mathbb{E}[\xi X] \leq x, \end{aligned} \tag{2.5}$$

where u is a utility function with the following assumption as in most of the literature,

Assumption 2.2.2. *$u(\cdot)$ is twice continuously differentiable, strictly increasing and strictly concave. Furthermore, $u'(\cdot)$ satisfies the Inada condition, i.e., $u'(0+) = +\infty$ and $u'(+\infty) = 0$.*

By convention, we set $u(x) = -\infty$ for $x < 0$.

We also impose the following integrability assumption throughout the chapter.

Assumption 2.2.3. *$\mathbb{E}[(u')^{-1}(\lambda_0 \xi) \xi] < \infty$ for all $\lambda_0 > 0$.*

This integrability assumption is standard in expected utility maximization problems under which we have,

Proposition 2.2.1. *The optimal solution to (2.5) is given by*

$$X_0^* = (u')^{-1}(\lambda_0^* \xi), \tag{2.6}$$

where λ_0^* solves $\mathbb{E}[\xi X_0^*] = x$.

To avoid unnecessary technical details, we assume throughout the chapter that (2.5) is well-posed, in other words,

Assumption 2.2.4. *The optimal value of (2.5) is finite for all $x > 0$.*

This assumption can be weakened, such as in Kramkov and Schachermayer (1999) and Jin et al. (2008). In fact, in a more general market setting, the optimal expected utility is finite under certain technical conditions (such as reasonable asymptotic elasticity of the utility function).

2.2.3 Value-at-Risk-based Risk Management

We motivate our new portfolio choice model by the Value-at-Risk-based risk management (VaR-RM) proposed by Basak and Shapiro (2001). Recognizing that risk management is typically not an economic agent's primary objective, those scholars

focus on portfolio choice within the familiar (continuous-time) complete markets setting, with the assumption that agents must limit their risks, as measured by VaR, while maximizing the expected utility.

The VaR-RM model is given by

$$\begin{aligned} & \max_X \mathbb{E}[u(X)] \\ & \text{subject to } \mathbb{E}[\xi X] \leq x, \\ & \mathbb{P}(X \geq \underline{x}) \geq 1 - \alpha. \end{aligned} \tag{2.7}$$

where $\mathbb{P}(X \geq \underline{x}) \geq 1 - \alpha$ is the VaR constraint and \underline{x} is the “floor” terminal wealth specified exogenously.

2.2.4 Risk Management with Weighted VaR

We follow Basak and Shapiro (2001) to embed the risk management with weighted VaR (WVaR-RM) into the standard utility maximization. We assume that an economic agent uses the expected utility model when managing his trading portfolio and that as a consequence of either the internal risk management or the external regulation such as by the SEC (1997); BCBS (2011, 2016), he has decided to manage the portfolio’s risk by imposing a constraint on the portfolio’s WVaR. Consequently, the WVaR-RM model is

$$\begin{aligned} & \max_X \mathbb{E}[u(X)] \\ & \text{subject to } \mathbb{E}[\xi X] \leq x, \\ & \rho_\Phi(X) \leq -\underline{x}, \end{aligned} \tag{2.8}$$

where ρ_Φ is the weighted VaR risk measure given by (2.1).

Note that $\rho_\Phi(X) \leq -\underline{x}$ is equivalent to $\rho_\Phi(X - x) \leq x - \underline{x}$ and therefore is consistent with the literature. Rogers (2009) considers a similar problem with law-invariant coherent risk measures, which is a proper subset of the WVaR risk measure. Moreover, both the feasibility and existence of the optimal solution and the impact on the portfolio choice are absent from his paper.

2.3 Solution

In this section, we present the solution to the WVaR-RM (2.8). The second constraint in (2.8) is based on the quantile function instead of the terminal wealth itself; thus, the standard convex duality method that is employed to solve the expected utility

maximization problem cannot be applied directly. To overcome this difficulty, we employ the so-called quantile formulation. To this end, we impose the following assumption on ξ ,

Assumption 2.3.1. ξ is atomless.

This assumption is satisfied when the investment opportunity set, i.e., the triplet $(r(\cdot), b(\cdot), \sigma(\cdot))$, is deterministic, $\int_0^T |\theta(t)|^2 dt \neq 0$, and ξ is lognormally distributed (which is the case with a Black-Scholes market).

Remark 2.3.1. The problem with an atomic pricing kernel can be solved by following the method in Xu (2014).

2.3.1 Quantile Formulation

Quantile formulation, developed in a series of papers including Schied (2004); Carlier and Dana (2006); Jin and Zhou (2008); He and Zhou (2011); Carlier and Dana (2011); Xia and Zhou (2016); Xu (2016), is a technique of solving optimal terminal payoff in portfolio choice problems. This technique can be applied once the objective function and constraints, except for the initial budget constraint, in a portfolio choice problem are law-invariant and the objective function is improved with a higher level of the terminal wealth (i.e., the more the better). The basic idea of the quantile formulation is to choose quantile functions as the decision variables. The advantage of this formulation is that the quantile-based risk constraint can be directly embedded into the optimization, and hence traditional optimization techniques become applicable. Furthermore, there is a simple connection between the optimal solution to the portfolio choice problem and its quantile formulation.

We first show that in our problem, the objective function is improved with a higher level of the terminal wealth.

Lemma 2.3.1. *If X^* is optimal to (2.8), then $\mathbb{E}[\xi X^*] = x$.*

Proof of Lemma 2.3.1. If $E[\xi X^*] < x$, then let $X = X^* + \frac{x - E[\xi X^*]}{E[\xi]} > X^*$. We have $E[\xi X] = x$, $\rho_{\Phi}(X) \geq -x$, and thus, X satisfies both constraints. Because $u(\cdot)$ is strictly increasing, $\mathbb{E}[u(X)] > \mathbb{E}[u(X^*)]$, which is a contradiction. \square

From Lemma 2.3.1, we also see that the optimal value of (2.8), if it exists, is strictly increasing in x .

Let $F_{\xi}(\cdot)$ denote the CDF of ξ , and let $F_{\xi}^{-1}(\cdot)$ denote the quantile function of ξ , which is strictly increasing as ξ is atomless. Let us introduce the following assumption,

Assumption 2.3.2. $\text{ess inf } \xi = 0$, $\text{ess sup } \xi = +\infty$, i.e., $F_\xi^{-1}(0) = 0$ and $F_\xi^{-1}(1) = +\infty$.

This assumption stipulates that, for any given positive value, there is a state of nature in which the market offers a return that is greater (less) than that value. In particular, this assumption is valid when the investment opportunity set, i.e., the triplet $(r(\cdot), b(\cdot), \sigma(\cdot))$, is deterministic, $\int_0^T |\theta(t)|^2 dt \neq 0$, and ξ is lognormally distributed (which is the case with a Black-Scholes market).

We also need the following lemma from Jin and Zhou (2008),

Lemma 2.3.2 (Jin and Zhou (2008)). $\mathbb{E}[\xi G(1 - F_\xi(\xi))] \leq \mathbb{E}[\xi X]$ for any lower bounded random variable X whose quantile function is G . Furthermore, if $\mathbb{E}[\xi G(1 - F_\xi(\xi))] < \infty$, then the inequality becomes equality if and only if $X = G(1 - F_\xi(\xi))$, a.s.

In view of Lemmas 2.3.2 and 2.3.1, we can consider the quantile formulation of (2.8). Define

$$U(G(\cdot)) := \int_{[0,1]} u(G(z)) dz, \quad G(\cdot) \in \mathbb{G}.$$

We consider the following problem

$$\begin{aligned} V(x, \underline{x}) &:= \max_{G(\cdot) \in \mathbb{G}} U(G(\cdot)) \\ \text{subject to } &\int_{[0,1]} F_\xi^{-1}(1 - z) G(z) dz = x, \\ &\int_{[0,1]} G(z) \Phi(dz) \geq \underline{x}. \end{aligned} \quad (2.9)$$

The following theorem verifies the equivalence of the portfolio selection problem (2.8) and the quantile formulation (2.9) in terms of the feasibility, well-posedness, existence and uniqueness of the optimal solution. The proof is similar to that in He and Zhou (2011).

Theorem 2.3.1. *We have the following assertions.*

1. *Problem (2.8) is feasible (well-posed) if and only if problem (2.9) is feasible (well-posed). Furthermore, they have the same optimal value.*
2. *The existence (uniqueness) of optimal solutions to (2.8) is equivalent to the existence (uniqueness) of optimal solutions to (2.9).*
3. *If X^* is optimal to (2.8), then $G^*(\cdot)$ is optimal to (2.9). If $G^*(\cdot)$ is optimal to (2.9), then $X^* = G^*(1 - F_\xi(\xi))$ is optimal to (2.8).*

Remark 2.3.2. In general, the quantile formulation is only applicable in the complete market. It is tacitly assumed that imposing the WVaR constraint does not induce market incompleteness and the quantile formulation can still be applied. In fact, with reasonable parameters, all terminal wealth satisfying the budget constraint and WVaR constraint jointly can be replicated.

Before we proceed, let us introduce the following set function

$$\kappa_{\Phi}(A) = \frac{\Phi(A)}{\int_A F_{\xi}^{-1}(1-z)dz}, \forall A \in \mathcal{B}[0,1],$$

where $\mathcal{B}[0,1]$ denotes all Borel-measurable sets in $[0,1]$. κ_{Φ} turns out to be closely related to the feasibility, existence, and properties of the optimal solution.

κ_{Φ} measures the *risk reduction per cost* (RRPC) across different states of the market at time T . Consider an economic agent who optimally chooses a terminal wealth X starting from an initial wealth x . We have demonstrated that we can consider only the terminal wealth of the form $X = G_X(1 - F_{\xi}(\xi))$, where G_X is the quantile function of X . The risk of X is $\rho_{\Phi}(X) = -\int_{[0,1]} G_X(z)\Phi(dz)$. Consider a contingent claim whose payoff at time T is $\varepsilon\mathbf{1}_{1-F_{\xi}(\xi) \in [a,b]}$, $\varepsilon > 0, 0 \leq a < b \leq 1$. Its cost at time 0 is given by $\mathbb{E}[\xi\varepsilon\mathbf{1}_{1-F_{\xi}(\xi) \in [a,b]}] = \varepsilon \int_{[a,b]} F_{\xi}^{-1}(1-z)dz$. If the agent purchases this contingent claim with additional initial wealth $\varepsilon \int_{[a,b]} F_{\xi}^{-1}(1-z)dz$ at time 0, then his terminal wealth becomes $X_{\varepsilon,a,b} := X + \varepsilon\mathbf{1}_{1-F_{\xi}(\xi) \in [a,b]}$. A rough approximation for the quantile function of $X_{\varepsilon,a,b}$ is $G_X(z) + \varepsilon\mathbf{1}_{z \in [a,b]}$, when ε and $b-a$ are sufficiently small. We then have $\rho_{\Phi}(X_{\varepsilon,a,b}) \approx -\int_{[0,1]} (G_X(z) + \varepsilon\mathbf{1}_{z \in [a,b]})\Phi(dz) = -\int_{[0,1]} G_X(z)\Phi(dz) - \varepsilon\Phi([a,b])$. Thus, the agent can reduce the risk of his terminal wealth (approximately) by $\varepsilon\Phi([a,b])$ at an extra cost of $\varepsilon \int_{[a,b]} F_{\xi}^{-1}(1-z)dz$. $\kappa_{\Phi}([a,b])$ is the ratio between the claim's risk reduction and cost and thus measures the tradeoff between reducing the risk and incurring the cost by investing in the future state $\{1 - F_{\xi}(\xi) \in [a,b]\}$. High a and b correspond to good states of the market, because these states are associated with low ξ .

2.3.2 Feasibility and Well-Posedness

An optimization problem is feasible if it admits at least one feasible solution (i.e., a solution that satisfies all the constraints involved), and it is well-posed if it has a finite optimal value. A feasible solution is optimal if it achieves the finite optimal value. See Jin et al. (2008) for a detailed discussion of these terminologies in the context of portfolio selection.

The feasibility of (2.8) depends on the solution to the following problem

$$\begin{aligned} C_\Phi(x) &:= \max_X -\rho_\Phi(X) \\ &\text{subject to } \mathbb{E}[\xi X] \leq x. \end{aligned} \tag{2.10}$$

$-C_\Phi(x)$ is the minimal risk of the terminal wealth X that an investor can achieve with an initial wealth x .

Proposition 2.3.1. *We have the following assertions:*

1. If $\sup_{0 < c < 1} \kappa_\Phi((c, 1]) > \frac{1}{\mathbb{E}\xi}$, then $C_\Phi(x) = +\infty$.
2. If $\sup_{0 < c < 1} \kappa_\Phi((c, 1]) \leq \frac{1}{\mathbb{E}\xi}$, then $C_\Phi(x) = \frac{x}{\mathbb{E}\xi}$.

Proof of Proposition 2.3.1. In view of the previous analysis, it is equivalent to consider the following problem

$$\begin{aligned} C_\Phi(x) &= \max_{G(\cdot) \in \mathbb{G}} \int_{[0,1]} G(z) \Phi(dz) \\ &\text{subject to } \int_{[0,1]} F_\xi^{-1}(1-z) G(z) dz = x. \end{aligned}$$

1. If $\sup_{0 < c < 1} \kappa_\Phi((c, 1]) > \frac{1}{\mathbb{E}\xi}$, then there exists $c \in (0, 1)$ such that $\kappa_\Phi((c, 1]) > \frac{1}{\mathbb{E}\xi} + \varepsilon$ for some $\varepsilon > 0$. Consider

$$G_n(z) = \frac{x - n \int_{[c,1]} F_\xi^{-1}(1-z) dz}{\mathbb{E}\xi} + n \mathbf{1}_{c \leq z \leq 1},$$

we have

$$\begin{aligned} \int_{[0,1]} G_n(z) \Phi(dz) &= \frac{x - n \int_{[c,1]} F_\xi^{-1}(1-z) dz}{\mathbb{E}\xi} + n \Phi([c, 1]) \\ &> x + \varepsilon n \int_{[c,1]} F_\xi^{-1}(1-z) dz, \end{aligned}$$

and the claim follows easily.

2. If $\sup_{0 < c < 1} \kappa_\Phi((c, 1]) \leq \frac{1}{\mathbb{E}\xi}$, then for any $G(\cdot) \in \mathbb{G}$, we have

$$\begin{aligned} \int_{[0,1]} G(z) \Phi(dz) &= \int_{[0,1]} \Phi([z, 1]) dG(z) + G(0-) \\ &\leq \frac{1}{\mathbb{E}\xi} \int_{[0,1]} F_\xi^{-1}(1-z) G(z) dz \\ &= \frac{x}{\mathbb{E}\xi}, \end{aligned}$$

and the equality holds when $G(z) = \frac{x}{\mathbb{E}\xi}$.

□

Define

$$\Delta_1 := \{(x, \underline{x}) : x > 0, 0 < \underline{x} < C_\Phi(x)\}. \quad (2.11)$$

Because of Proposition 2.3.1, (2.8) is feasible if $(x, \underline{x}) \in \Delta_1$, and it is infeasible if $\underline{x} > C_\Phi(x)$. Accordingly, from now on, we focus only on models with $(x, \underline{x}) \in \Delta_1$. We do not include the boundary $\underline{x} = C_\Phi(x) = \frac{x}{\mathbb{E}\xi}$, because the Lagrange method could fail on the boundary; see Remark 2.3.7.

Finally, as the optimal value of the unconstrained problem (2.5) is always larger than or equal to that of the constrained problem (2.8), we know that the optimal value of (2.8) is finite, provided it is feasible and (2.5) is well-posed.

2.3.3 Optimal Solution

Before we present the optimal solution, let us discuss when the risk constraint is non-redundant. Recall that X_0^* is the optimal solution to (2.5). If $\rho_\Phi(X_0^*) \leq -\underline{x}$, then X_0^* is already optimal to (2.8). In other words, the risk constraint is non-redundant if and only if $\rho_\Phi(X_0^*) > -\underline{x}$. Define

$$R_\Phi(x) := -\rho_\Phi((u')^{-1}(\lambda_0^*\xi)) \text{ where } \lambda_0^* \text{ solves } \mathbb{E}[\xi(u')^{-1}(\lambda_0^*\xi)] = x, \quad (2.12)$$

and

$$\Delta_2 := \{(x, \underline{x}) : \underline{x} > R_\Phi(x)\} \cap \Delta_1. \quad (2.13)$$

$-R_\Phi(x)$ is the risk of the optimal terminal wealth X_0^* of an investor with an initial wealth x , in the absence of the risk constraint. To exclude trivial cases, we further make the following assumption on Φ ,

Assumption 2.3.3. $\Phi(\{1\}) = 0$ and $\kappa_\Phi(A)$ is not constant for all $A \in \mathcal{B}[0, 1]$.

Remark 2.3.3.

1. If $\Phi(\{1\}) > 0$, then $R_\Phi(x) = +\infty$, and the risk constraint is always satisfied by X_0^* . Thus, the risk constraint is redundant.
2. If $\kappa_\Phi(A) = c$ for all $A \in \mathcal{B}[0, 1]$ where c is a positive constant, then $\Phi([0, z]) = c \int_{[0, z]} F_\xi^{-1}(1-s) ds$, $z \in [0, 1]$. It is illustrated that we can consider only the terminal wealth of the form $X = G_X(1 - F_\xi(\xi))$, where G_X is the quantile function of X . We then have $\rho_\Phi(X) = -\int_{[0, 1]} G_X(z) \Phi(dz) = -c \int_{[0, 1]} G_X(z) F_\xi^{-1}(1-z) dz = -c \mathbb{E}[\xi X]$, and the risk constraint becomes a multiple of the budget constraint. In this case, one of the constraints is redundant.

Let us introduce several notations.

Define

$$\varphi_{\lambda_2}(z) := - \int_{[z,1)} F_{\xi}^{-1}(1-s)ds + \lambda_2 \Phi((z, 1]), \quad z \in [0, 1),$$

with $\varphi_{\lambda_2}(1) := 0$, and its left-continuous version

$$\varphi_{\lambda_2}(z-) := - \int_{[z,1)} F_{\xi}^{-1}(1-s)ds + \lambda_2 \Phi([z, 1]), \quad z \in [0, 1),$$

with $\varphi_{\lambda_2}(1-) := 0$, where we have used $\Phi(\{1\}) = 0$.

Denote the concave envelope of $\varphi_{\lambda_2}(\cdot-)$ by $\delta_{\lambda_2}(\cdot)$, in other words

$$\delta_{\lambda_2}(z) := \sup_{0 \leq a \leq z \leq b \leq 1} \frac{(b-z)\varphi_{\lambda_2}(a-) + (z-a)\varphi_{\lambda_2}(b-)}{b-a}, \quad z \in [0, 1]. \quad (2.14)$$

Let

$$\begin{aligned} \mathcal{A}_{\Phi} &:= \{\lambda : \lambda > 0, \delta'_{\lambda}(z+) > 0, \forall z \in [0, 1)\} \\ &= \{\lambda : \lambda > 0, \varphi_{\lambda}(z-) < 0, \forall z \in [0, 1)\} \end{aligned}$$

where $\delta'_{\lambda}(\cdot+)$ is the right derivative of δ_{λ} .

For any given $\lambda_2 \in \{0\} \cup \mathcal{A}_{\Phi}$, define

$$f_{\lambda_2}(\lambda_1) := \int_{[0,1)} (u')^{-1}(\lambda_1 \delta'_{\lambda_2}(z+)) F_{\xi}^{-1}(1-z) dz, \quad \lambda_1 > 0,$$

and

$$g(\lambda_2, x) := f_{\lambda_2}^{-1}(x), \quad x > 0.$$

For any $\lambda_2 \in \{0\} \cup \mathcal{A}_{\Phi}$, $x > 0$, define

$$h(\lambda_2, x) := \int_{[0,1)} (u')^{-1}(g(\lambda_2, x) \delta'_{\lambda_2}(z+)) \Phi(dz),$$

and

$$S_{\Phi}(x) = \sup_{\lambda_2 \in \mathcal{A}_{\Phi}} h(\lambda_2, x). \quad (2.15)$$

We can now characterize the solutions to (2.8).

Theorem 2.3.1. *With C_{Φ}, R_{Φ} , and S_{Φ} defined by (2.12), (2.10), and (2.15), respectively, we have the following:*

1. *When $\limsup_{z \uparrow 1} \kappa_{\Phi}([z, 1)) = \infty$:*

(a) *If $\underline{x} \leq R_{\Phi}(x)$, then the optimal solution to (2.8) is given by (2.6).*

(b) *If $\underline{x} > R_{\Phi}(x)$, then there is no optimal solution.*

2. When $\limsup_{z \uparrow 1} \kappa_{\Phi}([z, 1]) < \infty$:

(a) If $\underline{x} \leq R_{\Phi}(x)$, the optimal solution is given by (2.6).

If, in addition,

$$\mathbb{E}[(u')^{-1}(\lambda_1 \delta'_{\lambda_2}((1 - F_{\xi}(\xi))_+))\xi] < \infty, \quad (2.16)$$

for all $\lambda_1 > 0$ and $\lambda_2 \in \mathcal{A}_{\phi}$, we have the following:

(b) If $R_{\Phi}(x) < \underline{x} < S_{\Phi}(x)$, the optimal solution is given by

$$X^* = (u')^{-1}(\lambda_1^* \delta'_{\lambda_2^*}((1 - F_{\xi}(\xi))_+)), \quad (2.17)$$

where $\delta'_{\lambda_2}(\cdot)_+$ is the right derivative of $\delta_{\lambda_2}(\cdot)$, and λ_1^* and λ_2^* solve

$$\begin{aligned} \mathbb{E}[\xi X^*] &= x, \\ \rho_{\Phi}(X^*) &= -\underline{x}. \end{aligned} \quad (2.18)$$

(c) If $S_{\Phi}(x) < \underline{x} < C_{\Phi}(x)$, then there is no optimal solution.

Remark 2.3.4.

1. If $F_{\xi}^{-1}(\cdot)$ is differentiable, $\limsup_{z \uparrow 1} \kappa_{\Phi}([z, 1]) = 0$, and there exists $z_{\Phi} \in [0, 1)$ such that $\Phi([z, 1])$ (as a function of z) is twice differentiable on $(z_{\Phi}, 1)$, then (2.16) holds. The differentiability of $F_{\xi}^{-1}(\cdot)$ is satisfied with a lognormal ξ . The twice differentiability of $\Phi([z, 1])$ for z near 1 is satisfied with all of the aforementioned risk measures. Moreover, as seen later, $0 < \liminf_{z \uparrow 1} \kappa_{\Phi}([z, 1]) < \infty$ holds only on rare occasions.
2. $S_{\Phi}(x)$ is the supremum of \underline{x} that can ensure the existence of the optimal solution to (2.8). For $\underline{x} = S_{\Phi}(x)$, the existence of the optimal solution depends on the attainability of $S_{\Phi}(x)$. In general, $S_{\Phi}(x)$ and $C_{\Phi}(x)$ can be different. $C_{\Phi}(x)$ is related to the feasibility of (2.8) and does not depend on the agent's preference, whereas $S_{\Phi}(x)$ is related to the existence of the optimal solution and does depend on the agent's preference.
3. The Lagrange multipliers, i.e., the solutions to (2.18), might not be unique. However, they must result in the same X^* , as the optimal solution (if it exists) must be unique due to the strict concavity of u . We can take any λ_1^* and λ_2^* that solve (2.18), and they will always result in the same X^* .

Remark 2.3.5. It is shown in Bernard et al. (2015a) that, under certain technical assumptions, the optimal portfolio for an agent with any law-invariant and non-decreasing preference can be rationalized by the expected utility setting, i.e., it is the optimal portfolio for an expected utility maximizer. We have similar results in our setting since the preference in 2.8 is law-invariant and increasing. Following the method in Bernard et al. (2015a), we can construct a concave utility function such that X^* (should it exist) is the optimal portfolio for an expected utility maximizer with this new utility function (but without the risk constraint).

2.3.4 Proof of Theorem 2.3.1

We apply the Lagrange dual method to solve (2.9). The proof of Theorem 2.3.1 is decomposed into four steps: first, establish the relationship between the Lagrangian dual problem and the original problem; second, solve the Lagrangian dual problem; third, prove the existence of Lagrange multipliers; and fourth, solve (2.9) by recalling Theorem 2.3.1.

2.3.4.1 Lagrange Approach

Define the Lagrangian

$$U_{\lambda_1, \lambda_2}(G(\cdot)) := \int_{[0,1]} u(G(z)) dz - \lambda_1 \int_{[0,1]} G(z) F_\xi^{-1}(1-z) dz + \lambda_1 \lambda_2 \int_{[0,1]} G(z) \Phi(dz),$$

and consider the problem

$$\bar{V}(\lambda_1, \lambda_2) = \max_{G(\cdot) \in \mathbb{G}} U_{\lambda_1, \lambda_2}(G(\cdot)). \quad (2.19)$$

Remark 2.3.6. In contrast to the convention, we define the second Lagrange multiplier in the form $\lambda'_2 := \lambda_1 \lambda_2$. This simplifies the notation in the analysis below.

We now attempt to link (2.19) with (2.9).

Before we present the result, let us prove two useful lemmas.

Lemma 2.3.3. *If $(x, \underline{x}) \in \Delta_2$, and X^* is optimal to (2.8), then*

$$\rho_\Phi(X^*) = -\underline{x}.$$

Proof of Lemma 2.3.3. If $\rho_\Phi(X^*) < -\underline{x}$, then define

$$\varepsilon := \frac{-\underline{x} - \rho_\Phi(X^*)}{\rho_\Phi(X_0^*) - \rho_\Phi(X^*)} \in (0, 1).$$

Let $X = G_X(1 - F_\xi(\xi))$, where $G_X(z) = (\varepsilon G_{X_0^*} + (1 - \varepsilon)G_{X^*})(z) \in \mathbb{G}$. We have $\mathbb{E}[\xi X] = x$ and $\rho_\Phi(X) = -\underline{x}$, and thus, X satisfies both constraints. In view of the strict concavity of u and the fact that the optimal value of (2.5) is no less than that of (2.8), we have $\mathbb{E}[u(X)] > \mathbb{E}[u(X^*)]$, which contradicts the optimality of X^* . \square

Lemma 2.3.4. $V(\mathbf{x}) := V(x, \underline{x})$, $\mathbf{x} = (x, \underline{x}) \in \Delta_1$ is a concave function, with a non-empty superdifferential $\partial V(\mathbf{x})$ at every $\mathbf{x} \in \Delta_1$. $V(x, \underline{x})$ is strictly increasing in x , decreasing in \underline{x} , and the monotonicity is strict whenever $\mathbf{x} \in \Delta_2$.

Proof of Lemma 2.3.4. In view of Lemmas 2.3.1 and 2.3.3, the monotonicity is obvious. Δ_1 is a convex, open set and $V(\mathbf{x})$, $\mathbf{x} \in \Delta_1$ is a concave function due to the strict concavity of $u(\cdot)$. Thus, the superdifferential of $V(\cdot)$ at any $\mathbf{x} \in \Delta_1$, $\partial V(\mathbf{x})$, is non-empty. \square

The following proposition links (2.19) with (2.9).

Proposition 2.3.2. For $(x, \underline{x}) \in \Delta_2$,

If (2.9) admits an optimal solution $G^*(\cdot)$, then there exists $\lambda_1 > 0$ and $\lambda_2 > 0$ such that $G^*(\cdot)$ is also optimal for (2.19) and $U_{\lambda_1, \lambda_2}(G^*(\cdot)) < \infty$;

Conversely, if $G^*(\cdot)$ solves (2.19) for some λ_1 and λ_2 satisfying

$$\begin{cases} \int_{[0,1]} G^*(z) F_\xi^{-1}(1-z) dz = x, \\ \int_{[0,1]} G^*(z) \Phi(dz) = \underline{x}, \end{cases} \quad (2.20)$$

then $G^*(\cdot)$ also solves (2.9) and $U(G^*(\cdot)) < \infty$.

Proof of Proposition 2.3.2. Let $G^*(\cdot)$ solve (2.9) with $\mathbf{x} = (x, \underline{x}) \in \Delta_2$ and $U(G^*(\cdot)) < \infty$. For any $(\lambda_1, -\lambda_3) \in \partial V(\mathbf{x})$, i.e., a supergradient of V at \mathbf{x} , and any $\mathbf{y} \in \Delta_1$, we have $V(\mathbf{y}) \leq V(\mathbf{x}) + (\lambda_1, -\lambda_3)^\top (\mathbf{y} - \mathbf{x})$, or equivalently $V(\mathbf{y}) - (\lambda_1, -\lambda_3)^\top \mathbf{y} \leq V(\mathbf{x}) - (\lambda_1, -\lambda_3)^\top \mathbf{x}$. Because $V(\cdot)$ is strictly concave, strictly increasing in x , and strictly decreasing in \underline{x} on Δ_2 , $\lambda_1 > 0$ and (there is at least one) $\lambda_3 > 0$, and we can choose $\lambda_2 := \frac{\lambda_3}{\lambda_1}$. Next for any $G \in \mathbb{G}$, let $\mathbf{y} = (\int_{[0,1]} G(z) F_\xi^{-1}(1-z) dz, \int_{[0,1]} G(z) \Phi(dz))$. We

have

$$\begin{aligned}
& U_{\lambda_1, \lambda_2}(G(\cdot)) \\
&= \int_{[0,1]} u(G(z))dz - \lambda_1 \int_{[0,1]} G(z)F_\xi^{-1}(1-z)dz + \lambda_1\lambda_2 \int_{[0,1]} G(z)\Phi(dz) \\
&\leq V(\mathbf{y}) - (\lambda_1, -\lambda_3)^\top \mathbf{y} \\
&\leq V(\mathbf{x}) - (\lambda_1, -\lambda_3)^\top \mathbf{x} \\
&= \int_{[0,1]} u(G^*(z))dz - \lambda_1 x + \lambda_1\lambda_2 \underline{x} \\
&= U_{\lambda_1, \lambda_2}(G^*(\cdot)) < \infty,
\end{aligned}$$

which implies that $G^*(\cdot)$ is also optimal for (2.19). Conversely, if $G^*(\cdot)$ solves (2.19) for some λ_1 and λ_2 satisfying (2.20), then for any $G(\cdot) \in \mathbb{G}$ that satisfies all the constraints in (2.9),

$$\begin{aligned}
& U(G(\cdot)) - \lambda_1 x + \lambda_1\lambda_2 \underline{x} \\
&\leq U_{\lambda_1, \lambda_2}(G(\cdot)) \\
&\leq U_{\lambda_1, \lambda_2}(G^*(\cdot)) \\
&= U(G^*(\cdot)) - \lambda_1 x + \lambda_1\lambda_2 \underline{x},
\end{aligned}$$

which thereby proves the desired result. \square

Remark 2.3.7. If $\underline{x} = C_\Phi(x) = \frac{x}{\mathbb{E}\xi}$, then the superdifferential of $V(\cdot)$ at $(x, \frac{x}{\mathbb{E}\xi})$ could be empty and the Lagrange multipliers may not exist. This fact has been neglected by many authors, including Rogers (2009) and Cahuich and Hernández-Hernández (2013). In the continuous-time utility maximization literature, the Lagrange dual method is often employed to solve (2.5), and it is assumed *a priori* that the optimal solution exists which can be found by solving the dual problem and finding a suitable Lagrange multiplier that corresponds to the budget constraint. For (2.8) or (2.9), if $\underline{x} = C_\Phi(x) = \frac{x}{\mathbb{E}\xi}$, then the Lagrange multiplier that corresponds to the risk constraint might not exist, but the optimal solution to the original problem might exist. In this case, the optimal solution to (2.8) or (2.9) is no longer given by the dual problem. We discuss this case in Appendix A.1.

2.3.4.2 Lagrangian Dual Problem

We now solve (2.19). The technique used in this section is similar to Rogers (2009), who focuses on (2.19) when Φ admits a density.

Recall that

$$\varphi_{\lambda_2}(z) = - \int_{[z,1]} F_\xi^{-1}(1-s)ds + \lambda_2\Phi((z, 1]), \quad z \in [0, 1),$$

with $\varphi_{\lambda_2}(1) = 0$, and

$$\begin{aligned}\varphi_{\lambda_2}(z-) &= - \int_{[z,1)} F_{\xi}^{-1}(1-s)ds + \lambda_2 \Phi([z,1]) \\ &= (\lambda_2 \kappa_{\Phi}([z,1)) - 1) \int_{[z,1)} F_{\xi}^{-1}(1-s)ds, \quad z \in [0,1),\end{aligned}\tag{2.21}$$

with $\varphi_{\lambda_2}(1-) = 0$.

The Lagrangian is now given by

$$U_{\lambda_1, \lambda_2}(G(\cdot)) = \int_{[0,1)} u(G(z))dz - \lambda_1 \int_{[0,1)} G(z)d\varphi_{\lambda_2}(z),\tag{2.22}$$

and we consider the problem

$$\bar{V}(\lambda_1, \lambda_2) = \max_{G(\cdot) \in \mathbb{G}} \int_{[0,1)} u(G(z))dz - \lambda_1 \int_{[0,1)} G(z)d\varphi_{\lambda_2}(z), \quad \lambda_1 > 0, \quad \lambda_2 > 0.\tag{2.23}$$

Note that $\varphi'_{\lambda_2}(z)$ may not exist, and it is not necessarily decreasing or non-negative (provided it exists), and thus, point-wise optimization fails. Inspired by Rogers (2009), Xia and Zhou (2016), and Xu (2016), we replace $\varphi_{\lambda_2}(\cdot)$ with $\delta_{\lambda_2}(\cdot)$, the concave envelope of $\varphi_{\lambda_2}(\cdot-)$.

Recall that

$$\mathcal{A}_{\Phi} = \{\lambda : \lambda > 0, \delta'_{\lambda}(z+) > 0, \quad z \in [0,1)\}.$$

We now present the solution to (2.23).

Proposition 2.3.3.

1. If $\lambda_2 \in \mathcal{A}_{\Phi}$, the optimal solution to (2.23) is given by $G^*(\cdot) := (u')^{-1}(\lambda_1 \delta'_{\lambda_2}(\cdot+))$.
2. If $\lambda_2 > 0$ and $\lambda_2 \notin \mathcal{A}_{\Phi}$, $\bar{V}(\lambda_1, \lambda_2) = \infty$, $\forall \lambda_1 > 0$.

Proof of Proposition 2.3.3. $\delta_{\lambda_2}(\cdot)$ is concave and $\delta_{\lambda_2}(0) = \varphi_{\lambda_2}(0-)$, $\delta_{\lambda_2}(1) = \varphi_{\lambda_2}(1-) = 0$. For any $G(\cdot) \in \mathbb{G}$,

$$\int_{[0,1)} (\varphi_{\lambda_2}(z-) - \delta_{\lambda_2}(z))dG(z) \leq 0,$$

and applying Fubini's theorem, we have

$$\begin{aligned}& \int_{[0,1)} u(G(z))dz - \lambda_1 \int_{[0,1)} G(z)d\varphi_{\lambda_2}(z) \\ & \leq \int_{[0,1)} u(G(z))dz - \lambda_1 \int_{[0,1)} G(z)d\delta_{\lambda_2}(z) \\ & = \int_{[0,1)} u(G(z))dz - \lambda_1 \int_{[0,1)} G(z)\delta'_{\lambda_2}(z+)dz.\end{aligned}$$

For $\lambda_2 \in \mathcal{A}_\phi$, we have

$$\int_{[0,1)} [u(G(z)) - \lambda_1 \delta'_{\lambda_2}(z+)G(z)]dz \leq \int_{[0,1)} [u(\bar{G}(z)) - \lambda_1 \delta'_{\lambda_2}(z+)\bar{G}(z)]dz,$$

where $\bar{G}(z) := (u')^{-1}(\lambda_1 \delta'_{\lambda_2}(z+))$ is given by point-wise optimization.

It is now sufficient to show that

$$\int_{[0,1)} u(\bar{G}(z))dz - \lambda_1 \int_{[0,1)} \bar{G}(z)d\varphi_{\lambda_2}(z) = \int_{[0,1)} [u(\bar{G}(z)) - \lambda_1 \delta'_{\lambda_2}(z+)\bar{G}(z)]dz,$$

or equivalently

$$\int_{[0,1)} (u')^{-1}(\lambda_1 \delta'_{\lambda_2}(z+))d\varphi_{\lambda_2}(z) - \int_{[0,1)} (u')^{-1}(\lambda_1 \delta'_{\lambda_2}(z+))d\delta_{\lambda_2}(z) = 0.$$

Applying Fubini's theorem again, the above identity is equivalent to

$$\int_{[0,1)} [\delta_{\lambda_2}(z) - \varphi_{\lambda_2}(z-)] \frac{1}{u''((u')^{-1}(\lambda_1 \delta'_{\lambda_2}(z+)))} d\delta'_{\lambda_2}(z+) = 0.$$

Because $\delta_{\lambda_2}(\cdot)$ dominates $\varphi_{\lambda_2}(\cdot-)$ on $[0, 1]$, $u''(\cdot) < 0$, and $\delta'_{\lambda_2}(\cdot+)$ is constant on any sub-interval of $\{z \in (0, 1) : \delta_{\lambda_2}(z) > \varphi_{\lambda_2}(z-)\}$, the above identity holds.

If $\lambda_2 \notin \mathcal{A}_\phi$, then there exists $0 \leq c < 1$ such that $\delta'_{\lambda_2}(z+) \leq 0$, $z \in [c, 1]$. Let

$$\bar{G}_n(z) = (u')^{-1}(\lambda_1(\delta'_{\lambda_2}(z+) \vee \frac{1}{n})),$$

and

$$\bar{V}_n = \int_{[0,1)} u(\bar{G}_n(z))dz - \lambda_1 \int_{[0,1)} \bar{G}_n(z)d\varphi_{\lambda_2}(z).$$

Similarly, we can show that

$$\begin{aligned} \bar{V}_n &= \int_{[0,1)} u(\bar{G}_n(z))dz - \lambda_1 \int_{[0,1)} \bar{G}_n(z)d\varphi_{\lambda_2}(z) \\ &= \int_{[0,1)} [u(\bar{G}_n(z)) - \lambda_1 \delta'_{\lambda_2}(z+)\bar{G}_n(z)]dz. \end{aligned}$$

Because $u'(\cdot)$ is strictly decreasing, $\bar{V}_n \leq \bar{V}_{n+1}$ and there exists $0 \leq c < 1$ such that $\delta'_{\lambda_2}(z) \leq 0$, $z \in [c, 1]$, we have

$$\lim_{n \rightarrow \infty} \bar{V}_n \geq \lim_{n \rightarrow \infty} \int_{[c,1)} [u(\bar{G}_n(z)) - \lambda_1 \delta'_{\lambda_2}(z+)\bar{G}_n(z)]dz = \infty.$$

□

In view of Proposition (2.3.2), the optimal solution, if it exists, must be given by $(u')^{-1}(\lambda_1 \delta'_{\lambda_2}(\cdot+))$, where $\lambda_2 \in \mathcal{A}_\Phi$. However, \mathcal{A}_Φ is not easy to obtain. We provide an equivalent characterization.

Lemma 2.3.5.

$$\mathcal{A}_\Phi = \{\lambda : \lambda > 0, \varphi_\lambda(z-) < 0, \forall z \in [0, 1)\}.$$

Proof of Lemma 2.3.5. $\forall \lambda \in \mathcal{A}_\Phi$, $\delta'_\lambda(z+) > 0$, $z \in [0, 1)$ implies $\varphi_\lambda(z-) \leq \delta_\lambda(z) < \delta_\lambda(1) \leq 0$, $z \in [0, 1)$.

$\forall \lambda \in \{\lambda : \lambda > 0, \varphi_\lambda(z-) < 0, \forall z \in [0, 1)\}$, as $\varphi_\lambda(z-) \geq \varphi_\lambda(z+)$, $z \in (0, 1)$, $\varphi_\lambda(z-)$ is upper semi-continuous. We then have $\sup_{z \in [0, c)} \varphi_\lambda(z-) < 0$, $\forall c \in (0, 1)$. From (2.14), we know $\delta_\lambda(z) < 0$, $z \in [0, 1)$ and $\delta'_\lambda(1-) > 0$. Thus $\delta'_\lambda(z+) \geq \delta'_\lambda(1-) > 0$, $z \in [0, 1)$. \square

Proposition 2.3.4.

1. If $\limsup_{z \uparrow 1} \kappa_\Phi([z, 1)) = \infty$, then $\mathcal{A}_\Phi = \emptyset$.
2. If $\limsup_{z \uparrow 1} \kappa_\Phi([z, 1)) < \infty$, then $\mathcal{A}_\Phi \neq \emptyset$.

Proof of Proposition 2.3.4.

1. $\limsup_{z \uparrow 1} \kappa_\Phi([z, 1)) = \infty$ implies that for any $\lambda_2 > 0$, there exists $z \in (0, 1)$ such that $\kappa_\Phi([z, 1)) > \frac{1}{\lambda_2}$ and $\varphi_{\lambda_2}(z-) \geq 0$. Thus, $\mathcal{A}_\Phi = \emptyset$.
2. $\limsup_{z \uparrow 1} \kappa_\Phi([z, 1)) < \infty$ implies that there exists $a > 0$ such that $\kappa_\Phi([z, 1)) < a$, $z \in [0, 1)$. Then $\forall \lambda_2 \leq \frac{1}{a}$, $\varphi_{\lambda_2}(z-) < 0$, $z \in [0, 1)$. Thus $\mathcal{A}_\Phi \neq \emptyset$.

\square

2.3.4.3 Existence of Lagrange Multipliers

We now show the existence of Lagrange multipliers, which is the solution to the following system of equations

$$\begin{cases} \int_{[0,1)} (u')^{-1}(\lambda_1 \delta'_{\lambda_2}(z+)) F_\xi^{-1}(1-z) dz = x, \\ \int_{[0,1)} (u')^{-1}(\lambda_1 \delta'_{\lambda_2}(z+)) \Phi(dz) = \underline{x}. \end{cases} \quad (2.24)$$

We impose the following integrability assumption.

Assumption 2.3.4.

$$\mathbb{E}[(u')^{-1}(\lambda_1 \delta'_{\lambda_2}((1 - F_\xi(\xi))_+))\xi] = \int_{[0,1)} (u')^{-1}(\lambda_1 \delta'_{\lambda_2}(z_+)) F_\xi^{-1}(1 - z) dz < \infty,$$

for all $\lambda_1 > 0$ and $\lambda_2 \in \mathcal{A}_\phi$.

The following lemma provides a sufficient condition for Assumption 2.3.4.

Lemma 2.3.6. *If $F_\xi^{-1}(\cdot)$ is differentiable, $\limsup_{z \uparrow 1} \kappa_\Phi([z, 1]) = 0$, and there exists $z_\Phi \in [0, 1)$ such that $\Phi([z, 1])$ (as a function of z) is twice differentiable on $(z_\Phi, 1)$, then*

$$\int_{[0,1)} (u')^{-1}(\lambda_1 \delta'_{\lambda_2}(z_+)) F_\xi^{-1}(1 - z) dz < \infty,$$

for all $\lambda_1 > 0$ and $\lambda_2 \in \mathcal{A}_\phi$.

Proof of Lemma 2.3.6. Fix $\lambda_1 > 0$ and $\lambda_2 \in \mathcal{A}_\phi$. Because $\Phi([z, 1])$ (as a function of z) is twice differentiable on $(z_\Phi, 1)$, we have $\Phi([z, 1]) = \int_{[z,1)} \phi(s) ds$ for $z \in (z_\Phi, 1)$ and some $\phi(\cdot)$. $\limsup_{z \uparrow 1} \kappa_\Phi([z, 1]) = 0$ implies $\lim_{z \uparrow 1} \kappa_\Phi([z, 1]) = 0$. According to L'Hopital's rule,

$$\lim_{z \uparrow 1} \kappa_\Phi([z, 1]) = \lim_{z \uparrow 1} \frac{-\phi'(z)}{(F_\xi^{-1})'(1 - z)} = 0.$$

There exists $b \in (z_\Phi, 1)$ such that $\lambda_2 |\phi'(z)| < (F_\xi^{-1})'(1 - z)$, $z \in (b, 1)$. Consequently, $\varphi_{\lambda_2}(\cdot -)$ is strictly concave on $(b, 1)$. We now show that there exists $c \in (b, 1)$ such that $\delta_{\lambda_2}(z) = \varphi_{\lambda_2}(z)$, $z \in [c, 1]$. Otherwise, we can find $z_1 \in (b, 1)$ such that $\delta_{\lambda_2}(z_1) > \varphi_{\lambda_2}(z_1)$. Let $z_2 := \inf\{z > z_1 | \delta_{\lambda_2}(z) = \varphi_{\lambda_2}(z)\}$. Because

$$\begin{aligned} 0 &\leq \delta'_{\lambda_2}(1-) \leq \lim_{z \uparrow 1} \frac{\varphi_{\lambda_2}(1-) - \varphi_{\lambda_2}(z-)}{1 - z} \\ &= \lim_{z \uparrow 1} \frac{(1 - \lambda_2 \kappa_\Phi([z, 1])) \int_{[z,1)} F_\xi^{-1}(1 - s) ds}{1 - z} \\ &= \lim_{z \uparrow 1} F_\xi^{-1}(1 - z) \\ &= 0, \end{aligned}$$

we must have $z_2 < 1$. We can then find $z_3 \in (z_2, 1)$ such that $\delta_{\lambda_2}(z_3) > \varphi_{\lambda_2}(z_3)$. We have $z_4 := \sup\{z < z_3 | \delta_{\lambda_2}(z) = \varphi_{\lambda_2}(z)\} \in (z_2, z_3)$ and $z_5 := \inf\{z > z_3 | \delta_{\lambda_2}(z) = \varphi_{\lambda_2}(z)\} \in (z_3, 1)$. Then, $\varphi'_{\lambda_2}(z_4) \leq \delta'_{\lambda_2}(z_4+) = \delta'_{\lambda_2}(z_5-) \leq \varphi'_{\lambda_2}(z_5)$, which is a contradiction.

Next, we can find $d \in (c, 1)$ such that $\delta'_{\lambda_2}(z+) = F_\xi^{-1}(1-z) - \lambda_2\phi(z) > \frac{1}{2}F_\xi^{-1}(1-z)$, $z \in [d, 1)$. We then have

$$\begin{aligned}
& \int_{[0,1)} (u')^{-1}(\lambda_1\delta'_{\lambda_2}(z+))F_\xi^{-1}(1-z)dz \\
&= \int_{[0,d)} (u')^{-1}(\lambda_1\delta'_{\lambda_2}(z+))F_\xi^{-1}(1-z)dz + \int_{[d,1)} (u')^{-1}(\lambda_1\delta'_{\lambda_2}(z+))F_\xi^{-1}(1-z)dz \\
&< (u')^{-1}(\lambda_1\delta'_{\lambda_2}(d+)) \int_{[0,d)} F_\xi^{-1}(1-z)dz + \int_{[d,1)} (u')^{-1}\left(\frac{\lambda_1}{2}F_\xi^{-1}(1-z)\right)F_\xi^{-1}(1-z)dz \\
&\leq (u')^{-1}(\lambda_1\delta'_{\lambda_2}(d+))\mathbb{E}\xi + \mathbb{E}[(u')^{-1}\left(\frac{\lambda_1}{2}\xi\right)\xi] < \infty,
\end{aligned}$$

and the claim follows easily. \square

Under Assumption 2.3.4, we can show the integrability of the second equation of (2.24).

Lemma 2.3.7. *If $\limsup_{z \uparrow 1} \kappa_\Phi([z, 1)) < \infty$ and Assumption 2.3.4 holds, then*

$$\int_{[0,1)} (u')^{-1}(\lambda_1\delta'_{\lambda_2}(z+))\Phi(dz) < \infty,$$

for all $\lambda_1 > 0$ and $\lambda_2 \in \mathcal{A}_\phi$.

Proof of Lemma 2.3.7. For any $K > \limsup_{z \uparrow 1} \kappa_\Phi([z, 1))$, there exists $0 < b < 1$ such that $\kappa_\Phi([z, 1)) < K$, $z \in [b, 1)$. We then have

$$\begin{aligned}
& \int_{[0,1)} (u')^{-1}(\lambda_1\delta'_{\lambda_2}(z+))\Phi(dz) \\
&\leq (u')^{-1}(\lambda_1\delta'_{\lambda_2}(b+)) + \int_{(b,1)} (u')^{-1}(\lambda_1\delta'_{\lambda_2}(z+))\Phi(dz) \\
&= (u')^{-1}(\lambda_1\delta'_{\lambda_2}(b+)) + (u')^{-1}(\lambda_1\delta'_{\lambda_2}(b+))\Phi((b, 1)) \\
&\quad + \int_{(b,1)} \Phi([z, 1))d((u')^{-1}(\lambda_1\delta'_{\lambda_2}(z+))) \\
&< (u')^{-1}(\lambda_1\delta'_{\lambda_2}(b+)) + K(u')^{-1}(\lambda_1\delta'_{\lambda_2}(b+)) \int_{(b,1)} F_\xi^{-1}(1-z)dz \\
&\quad + K \int_{(b,1)} \left(\int_{[z,1)} F_\xi^{-1}(1-s)ds \right) d((u')^{-1}(\lambda_1\delta'_{\lambda_2}(z+))) \\
&= (u')^{-1}(\lambda_1\delta'_{\lambda_2}(b+)) + K \int_{(b,1)} (u')^{-1}(\lambda_1\delta'_{\lambda_2}(z+))F_\xi^{-1}(1-z)dz < \infty.
\end{aligned}$$

for all $\lambda_1 > 0$ and $\lambda_2 \in \mathcal{A}_\phi$, and the claim follows easily. \square

Before we proceed, let us show a very useful lemma.

Lemma 2.3.8.

1. Denote the convex conjugate of $-\varphi_\lambda(z-)$ by

$$\hat{\varphi}_\lambda(x) := \sup_{z \in [0,1]} \{xz + \varphi_\lambda(z-)\}, \quad x \in R,$$

and let $A(x, \lambda) := \{z : \hat{\varphi}_\lambda(x) = xz + \varphi_\lambda(z-)\}$. The right (left) derivatives of $\hat{\varphi}_\lambda(x)$ (with respect to x) are given by

$$\begin{aligned} \hat{\varphi}'_\lambda(x+) &= \max_{z \in A(x, \lambda)} z, \\ \hat{\varphi}'_\lambda(x-) &= \min_{z \in A(x, \lambda)} z, \end{aligned}$$

respectively.

2. For $\lambda_1 < \lambda_2$,

$$\begin{aligned} \hat{\varphi}'_{\lambda_2}(x+) &\leq \hat{\varphi}'_{\lambda_1}(x+), \\ \hat{\varphi}'_{\lambda_2}(x-) &\leq \hat{\varphi}'_{\lambda_1}(x-), \quad \forall x \in R. \end{aligned}$$

Moreover, $\forall \varepsilon > 0, \exists \delta > 0$, such that whenever $|\lambda_1 - \lambda_2| < \delta$, we have

$$\begin{aligned} |\hat{\varphi}'_{\lambda_2}(x+) - \hat{\varphi}'_{\lambda_1}(x+)| &< \varepsilon, \\ |\hat{\varphi}'_{\lambda_2}(x-) - \hat{\varphi}'_{\lambda_1}(x-)| &< \varepsilon, \quad \forall x \in R. \end{aligned}$$

3.

$$\delta_\lambda(z) = -\sup_{x \in R} \{xz - \hat{\varphi}_\lambda(x)\}, \quad z \in [0, 1]$$

and the set $B(z, \lambda) := \{x : \delta_\lambda(z) = \hat{\varphi}_\lambda(x) - xz\}$ is non-empty for $z \in (0, 1)$.

4. The right (left) derivatives of $\delta_\lambda(z)$ with respect to $z \in (0, 1)$ are given by

$$\begin{aligned} \delta'_\lambda(z+) &= -\max_{x \in B(z, \lambda)} x, \\ \delta'_\lambda(z-) &= -\min_{x \in B(z, \lambda)} x, \end{aligned}$$

respectively. For $z \in (0, 1)$, $\delta'_\lambda(z+)$ is continuous and non-increasing in λ .

Proof of Lemma 2.3.8.

1. Note that $\varphi_\lambda(\cdot-)$ is upper semi-continuous, the proof is similar to that of Corollary 4 in Milgrom and Segal (2002).

2. Now for $\lambda_1 < \lambda_2$, let

$$\begin{aligned} z_1 &= \max_{z \in A(x, \lambda_1)} z = \hat{\varphi}'_{\lambda_1}(x+), \\ z_2 &= \max_{z \in A(x, \lambda_2)} z = \hat{\varphi}'_{\lambda_2}(x+). \end{aligned}$$

$\forall z > z_1$,

$$\begin{aligned} & xz + \varphi_{\lambda_2}(z-) \\ &= xz + \varphi_{\lambda_1}(z-) + (\lambda_2 - \lambda_1)\Phi([z, 1]) \\ &< xz_1 + \varphi_{\lambda_1}(z_1-) + (\lambda_2 - \lambda_1)\Phi([z_1, 1]) \\ &= xz_1 + \varphi_{\lambda_2}(z_1-), \end{aligned}$$

Thus, $z_2 \leq z_1$ and $\hat{\varphi}'_{\lambda_2}(x+) \leq \hat{\varphi}'_{\lambda_1}(x+)$. Similarly, $\hat{\varphi}'_{\lambda_2}(x-) \leq \hat{\varphi}'_{\lambda_1}(x-)$.

Without loss of generality, let us assume $\lambda_1 < \lambda_2$. $\forall z > z_2$, $xz + \varphi_{\lambda_2}(z) < xz_2 + \varphi_{\lambda_2}(z_2)$, and $\forall \varepsilon > 0$, $\exists \delta > 0$ such as whenever $z - z_2 > \varepsilon$, we have

$$xz + \varphi_{\lambda_2}(z-) < xz_2 + \varphi_{\lambda_2}(z_2-) - \delta.$$

Because $\Phi([z_2, z]) \leq 1$, we have for $z - z_2 > \varepsilon$

$$\begin{aligned} & xz + \varphi_{\lambda_1}(z-) - xz_2 - \varphi_{\lambda_1}(z_2-) \\ &= xz + \varphi_{\lambda_2}(z-) - xz_2 - \varphi_{\lambda_2}(z_2-) - (\lambda_1 - \lambda_2)\Phi([z_2, z]) \\ &< -\delta + (\lambda_2 - \lambda_1)\Phi([z_2, z]) \\ &< -\frac{1}{2}\delta, \end{aligned}$$

whenever $|\lambda_1 - \lambda_2| < \frac{1}{2}\delta$. Thus, $0 \leq z_1 - z_2 < \varepsilon$, i.e., $0 \leq \hat{\varphi}'_{\lambda_1}(x+) - \hat{\varphi}'_{\lambda_2}(x+) < \varepsilon$. Similarly, we can show that for the left derivatives.

3. First,

$$\begin{aligned} \underline{\delta}_\lambda(z) &:= -\sup_{x \in R} \{xz - \hat{\varphi}_\lambda(x)\} \\ &= -\sup_{x \in R} \{xz - \sup_{y \in [0,1]} \{xy + \varphi_\lambda(y-)\}\} \\ &\geq -\sup_{x \in R} \{xz - xz - \varphi_\lambda(z-)\} \\ &= \varphi_\lambda(z-), \quad z \in [0, 1]. \end{aligned}$$

Because $\hat{\varphi}_\lambda(x) \geq x$, and $\hat{\varphi}_\lambda(x) = x$ for $x > 0$, $\underline{\delta}_\lambda(1) = -\sup_{x \in R} \{x - \hat{\varphi}_\lambda(x)\} = 0 = \varphi_\lambda(1-)$.

$$\begin{aligned} \underline{\delta}_\lambda(0) &= \inf_{x \in R} \{ \sup_{z \in [0,1]} \{xz + \varphi_\lambda(z-)\} \} \\ &\leq \inf_n \{ \sup_{z \in [0,1]} \{-nz + \varphi_\lambda(z-)\} \} \\ &= \inf_n \{-nz_n + \varphi_\lambda(z_n-)\}, \end{aligned}$$

where z_n can be any z that achieves the supremum above. Now $\{z_n\}$ has a subsequence $z_{n_k} \rightarrow z_\infty$. If $z_\infty > 0$, then $\delta_\lambda(0) \leq \liminf_{n \rightarrow \infty} \{-nz_n + \varphi_\lambda(z_n -)\} = -\infty$, which is a contradiction. Thus $z_\infty = 0$ and $\underline{\delta}_\lambda(0) \leq \lim_{k \rightarrow \infty} \{-n_k z_{n_k} + \varphi_\lambda(z_{n_k} -)\} \leq \varphi_\lambda(0-)$, and we have $\underline{\delta}_\lambda(0) = \varphi_\lambda(0-)$. It is then a simple exercise to show that $\underline{\delta}_\lambda$ defined in this way is indeed the concave envelope of $\varphi_\lambda(\cdot-)$.

We show that $x = +\infty$ or $-\infty$ is never optimal for $xz - \hat{\varphi}_\lambda(x)$, when $z \in (0, 1)$. First, as $\hat{\varphi}_\lambda(x) = x$ for $x > 0$, $x = +\infty$ is not optimal when $z < 1$. Because $\hat{\varphi}_\lambda(x) \geq \varphi_\lambda(0-)$ for all x , $x = -\infty$ is not optimal for $z > 0$.

4. Fix $z \in (0, 1)$. Because $\hat{\varphi}_\lambda(x) = x$ for $x > 0$,

$$\begin{aligned} \sup_{x>0} \{xz - \hat{\varphi}_\lambda(x)\} &\leq \sup_{x>0} \{x - \hat{\varphi}_\lambda(x)\} \\ &= 0 \\ &= \hat{\varphi}_\lambda(0) \\ &\leq \sup_{x \leq 0} \{xz - \hat{\varphi}_\lambda(x)\}, \end{aligned}$$

and $\delta_\lambda(z) = -\sup_{x \leq 0} \{xz - \hat{\varphi}_\lambda(x)\}$.

Let $x_\lambda := \max_{x \in B(z, \lambda)} x$ and $x'_\lambda := \min_{x \in B(z, \lambda)} x$. Because $\hat{\varphi}_\lambda(\cdot)$ is convex, x_λ and x'_λ exist. Moreover, $-\infty < x'_\lambda \leq x_\lambda \leq 0$.

We have

$$\hat{\varphi}'_\lambda(x_\lambda -) \leq z \leq \hat{\varphi}'_\lambda(x_\lambda +),$$

and $\forall \varepsilon > 0$,

$$z < \hat{\varphi}'_\lambda((x_\lambda + \varepsilon)-).$$

Now for $\lambda_1 < \lambda_2$, if $x_{\lambda_2} < x_{\lambda_1}$,

$$z < \hat{\varphi}'_{\lambda_2}(x_{\lambda_1} -) \leq \hat{\varphi}'_{\lambda_1}(x_{\lambda_1} -),$$

which is a contradiction. Thus, $x_{\lambda_2} \geq x_{\lambda_1}$.

Without loss of generality, let us assume $\lambda_1 < \lambda_2$. Suppose $\exists \varepsilon_0 > 0$ such that for all $\delta > 0$ there exists λ_1, λ_2 satisfying $0 < \lambda_2 - \lambda_1 < \delta$ and $x_{\lambda_2} > x_{\lambda_1} + \varepsilon_0$. We have $z < \hat{\varphi}'_{\lambda_1}(x_{\lambda_2} -)$. From 2 we know, for $\varepsilon_1 = \frac{1}{2}(\hat{\varphi}'_{\lambda_1}(x_{\lambda_2} -) - z)$, $\exists \delta_1 > 0$ such that whenever $0 < \lambda_2 - \lambda_1 < \delta_1$,

$$z < \hat{\varphi}'_{\lambda_1}(x_{\lambda_2} -) - \varepsilon_1 < \hat{\varphi}'_{\lambda_1}(x_{\lambda_1} -),$$

which is a contradiction. Thus, x_λ is continuous in λ .

Similarly, we can show that x'_λ is non-decreasing and continuous in λ . Therefore, $B(z, \lambda)$ is compact. The expression for the right (left) derivatives follows from a similar argument as in 1.

Summarizing the above results, we have, for $z \in (0, 1)$, $\delta'_\lambda(z+)$ is continuous and non-increasing in λ .

□

Recall that

$$f_{\lambda_2}(\lambda_1) = \int_{[0,1)} (u')^{-1}(\lambda_1 \delta'_{\lambda_2}(z+)) F_\xi^{-1}(1-z) dz, \quad \lambda_1 > 0.$$

For any given $\lambda_2 \in \{0\} \cup \mathcal{A}_\Phi$, $f_{\lambda_2}(\cdot)$ is continuous, strictly decreasing on $(0, \infty)$, and

$$\lim_{\lambda_1 \rightarrow 0^+} f_{\lambda_2}(\lambda_1) = \infty, \quad \lim_{\lambda_1 \rightarrow +\infty} f_{\lambda_2}(\lambda_1) = 0.$$

For any $x > 0$, there exists a unique $\lambda_1^* > 0$ such that $f_{\lambda_2}(\lambda_1^*) = x$. Therefore, $g(\lambda_2, x) = f_{\lambda_2}^{-1}(x)$ is well-defined, and we know that $g(\cdot, x)$ is continuous, non-decreasing for all $x > 0$.

Recall that

$$h(\lambda_2, x) = \int_{[0,1)} (u')^{-1}(g(\lambda_2, x) \delta'_{\lambda_2}(z+)) \Phi(dz).$$

As will become clear in Lemma 2.4.1, $\delta'_{\lambda_2}(0+) = +\infty$ if $\Phi(\{0\}) > 0$, we have

$$h(\lambda_2, x) = \int_{(0,1)} (u')^{-1}(g(\lambda_2, x) \delta'_{\lambda_2}(z+)) \Phi(dz).$$

We now have

Lemma 2.3.9. *For all $x > 0$, $h(\cdot, x)$ is continuous on $\{0\} \cup \mathcal{A}_\Phi$.*

Proof of Lemma 2.3.9. We want to show that $\forall \lambda \in \{0\} \cup \mathcal{A}_\Phi$ and $\lambda_n \rightarrow \lambda$, $\lim_{n \rightarrow \infty} h(\lambda_n, x) = h(\lambda, x)$. By Fatou's lemma, we have $h(\lambda, x) \leq \liminf_{n \rightarrow \infty} h(\lambda_n, x)$. Next, let

$$h_0(\lambda_1, \lambda_2, x) := \int_{(0,1)} (u')^{-1}(g(\lambda_1, x) \delta'_{\lambda_2}(z+)) \Phi(dz).$$

We have that $h(\lambda, x) = h_0(\lambda, \lambda, x)$ and $h_0(\lambda_1, \lambda_2, x)$ is non-increasing in λ_1 and non-decreasing in λ_2 . For $\lambda_n \uparrow \lambda$ (except for $\lambda = 0$),

$$\begin{aligned} h_0(\lambda_n, \lambda_n, x) &= h_0(\lambda_n, \lambda, x) + h_0(\lambda_n, \lambda_n, x) - h_0(\lambda_n, \lambda, x) \\ &\leq h_0(\lambda_n, \lambda, x) \rightarrow h_0(\lambda, \lambda, x), \quad n \rightarrow \infty, \end{aligned}$$

where the convergence follows from the dominated convergence theorem. Thus, $h(\lambda, x) \geq \limsup_{n \rightarrow \infty} h(\lambda_n, x)$ and $h(\lambda, x) = \lim_{n \rightarrow \infty} h(\lambda_n, x)$.

For $\lambda_n \downarrow \lambda$ (except for $\lambda = \sup \mathcal{A}_\Phi$),

$$\begin{aligned} h_0(\lambda_n, \lambda_n, x) &= h_0(\lambda, \lambda_n, x) + h_0(\lambda_n, \lambda_n, x) - h_0(\lambda, \lambda_n, x) \\ &\leq h_0(\lambda, \lambda_n, x) \rightarrow h_0(\lambda, \lambda, x), \quad n \rightarrow \infty. \end{aligned}$$

Thus, $h(\lambda, x) \geq \limsup_{n \rightarrow \infty} h(\lambda_n, x)$ and $h(\lambda, x) = \lim_{n \rightarrow \infty} h(\lambda_n, x)$, which proves the lemma. \square

Note that for all $x > 0$, $h(\cdot, x)$ is continuous and $h(0, x) = R_\Phi(x)$. Because

$$S_\Phi(x) = \sup_{\lambda_2 \in \mathcal{A}_\Phi} h(\lambda_2, x),$$

we conclude that for any $\underline{x} \in (R_\Phi(x), S_\Phi(x))$ there exists at least one $\lambda_2^* > 0$ such that $h(\lambda_2^*, x) = \underline{x}$. Moreover, there is no solution if $\underline{x} > S_\Phi(x)$.

Define

$$\begin{aligned} \Delta_3 &:= \{(x, \underline{x}) : \underline{x} < S_\Phi(x)\} \cap \Delta_2, \\ \Delta_4 &:= \{(x, \underline{x}) : \underline{x} > S_\Phi(x)\} \cap \Delta_1. \end{aligned}$$

We now have

Proposition 2.3.5. $\forall (x, \underline{x}) \in \Delta_3$, there exists at least one pair of $\lambda_1^* > 0, \lambda_2^* > 0$ that solves (2.24). $\forall (x, \underline{x}) \in \Delta_4$, (2.24) admits no solution.

Proof of Proposition 2.3.5. In view of the above analysis, there exists at least one $\lambda_2^* > 0$ such that $h(\lambda_2^*, x) = \underline{x}$. Now $(g(\lambda_2^*, x), \lambda_2^*)$ solves (2.24). \square

Remark 2.3.8. For $\underline{x} = S_\Phi(x)$, there exists a solution if and only if there exists $\lambda_2^* \in \mathcal{A}_\Phi$ such that $h(\lambda_2^*, x) = \sup_{\lambda_2 \in \mathcal{A}_\Phi} h(\lambda_2, x)$.

2.3.4.4 Optimal Solution

In view of previous analysis, we can now characterize the optimal solution to (2.8).

Proof of Theorem 2.3.1. The claim follows from Proposition 2.3.4, Proposition 2.3.3, Proposition 2.3.5 and Theorem 2.3.1. \square

2.3.5 An Example

Before we close this section, we use an example to illustrate Theorem 2.3.1. We solve (2.8) when the risk constraint is given by $\text{ES}_\alpha(X) \leq -\underline{x}$. We assume that the ES constraint is binding and graphically show how the optimal wealth can be obtained. A more rigorous proof can be found in the next chapter in which we present a detailed study of ES-based risk management.

Example 2.3.1. For the constraint $\text{ES}_\alpha(X) \leq -\underline{x}$, we have $\Phi([z, 1]) = (1 - \frac{z}{\alpha}) \vee 0$ and

$$\varphi_{\lambda_2}(z-) = \begin{cases} -\int_{[z,1)} F_\xi^{-1}(1-s)ds + \lambda_2(1 - \frac{z}{\alpha}), & z \in [0, \alpha]; \\ -\int_{[z,1)} F_\xi^{-1}(1-s)ds, & z \in (\alpha, 1]. \end{cases}$$

The solid line in Figure 2.1 plots $\varphi_{\lambda_2}(\cdot-)$. Its concave envelope $\delta_{\lambda_2}(\cdot)$ replaces part of $\varphi_{\lambda_2}(\cdot-)$ with a chord (dashed line) between z_1 and z_2 , at which the slopes of $\varphi_{\lambda_2}(\cdot-)$ equal the slope of the chord, in other words,

$$\delta_{\lambda_2}(z) = \begin{cases} \varphi_{\lambda_2}(z-), & z \in [0, z_1]; \\ F_\xi^{-1}(1 - z_2)(z - z_1) + \varphi_{\lambda_2}(z_1-), & z \in (z_1, z_2]; \\ \varphi_{\lambda_2}(z-), & z \in (z_2, 1], \end{cases}$$

where $z_1 < \alpha$ and $z_2 > \alpha$ satisfy the following condition

$$\begin{cases} \delta_{\lambda_2}(z_2) = \varphi_{\lambda_2}(z_2-), \\ F_\xi^{-1}(1 - z_1) - \frac{\lambda_2}{\alpha} = F_\xi^{-1}(1 - z_2). \end{cases}$$

The optimal terminal wealth is then given by

$$X^* = \begin{cases} (u')^{-1}(\lambda_1^*(\xi - \frac{\lambda_2^*}{\alpha})), & \xi \in (F_\xi^{-1}(1 - z_1^*), +\infty); \\ (u')^{-1}(\lambda_1^*F_\xi^{-1}(1 - z_2^*)), & \xi \in (F_\xi^{-1}(1 - z_2^*), F_\xi^{-1}(1 - z_1^*]); \\ (u')^{-1}(\lambda_1^*\xi), & \xi \in (0, F_\xi^{-1}(1 - z_2^*]), \end{cases}$$

where $z_1^* < \alpha$, $z_2^* > \alpha$, $\lambda_1^* > 0$, and $\lambda_2^* > 0$ satisfy the following condition

$$\begin{cases} \delta_{\lambda_2^*}(z_2^*) = \varphi_{\lambda_2^*}(z_2^*-), \\ F_\xi^{-1}(1 - z_1^*) - \frac{\lambda_2^*}{\alpha} = F_\xi^{-1}(1 - z_2^*), \\ \mathbb{E}[\xi X^*] = x, \\ \text{ES}_\alpha(X^*) = -\underline{x}. \end{cases}$$

The solid line in Figure 2.2 plots the optimal terminal wealth under the ES constraint. The dashed line represents the optimal terminal wealth of a Benchmark agent who solves (2.5).

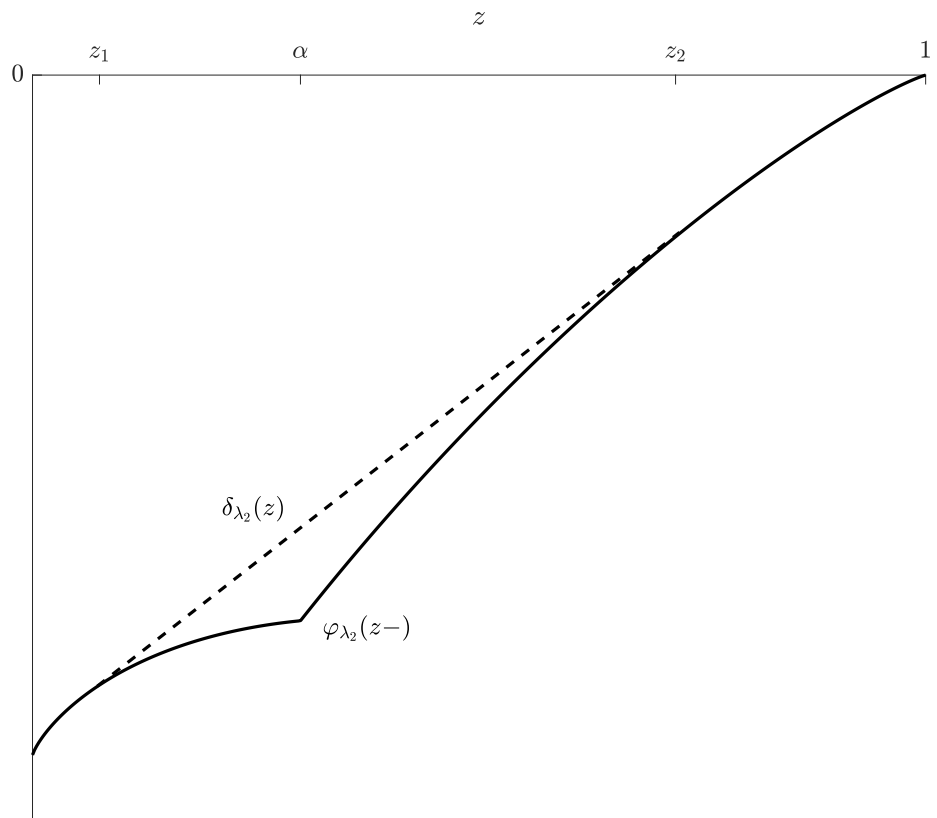


Figure 2.1: $\varphi_{\lambda_2}(\cdot-)$ and its concave envelope $\delta_{\lambda_2}(\cdot)$

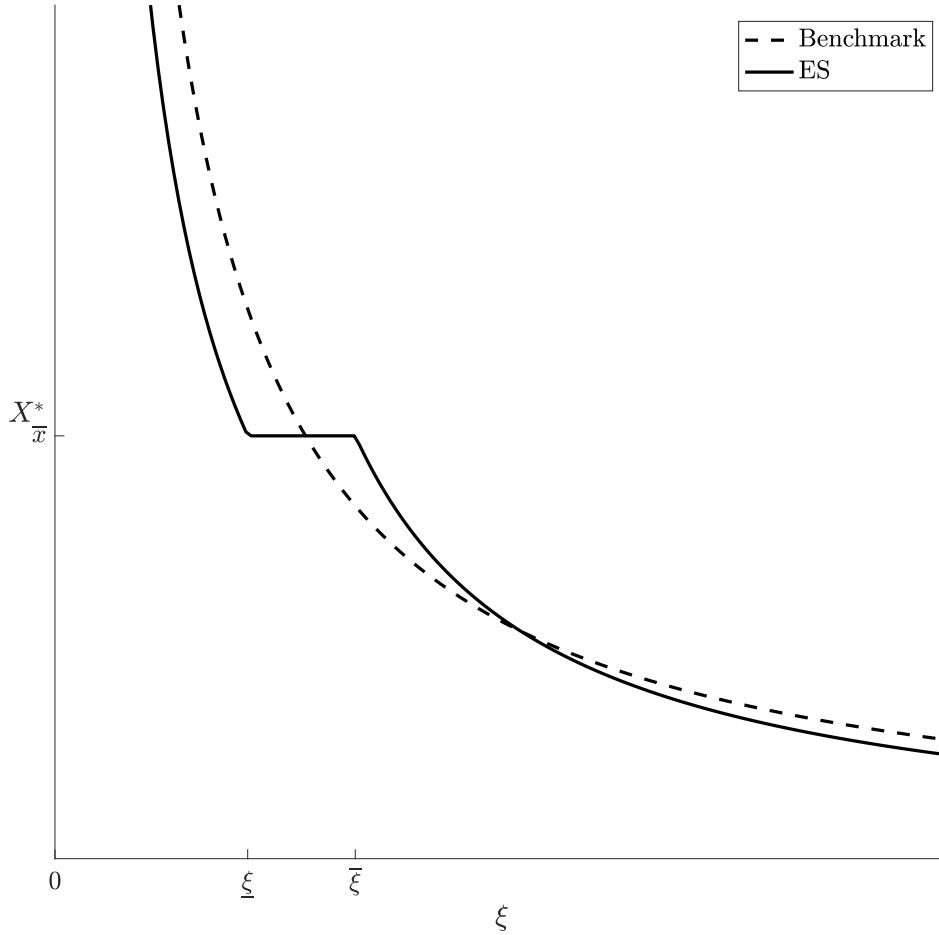


Figure 2.2: Optimal terminal wealth under ES

The figure plots the optimal terminal wealth of a benchmark agent (without the risk constraint, dashed line) and an ES agent (with an ES constraint, solid line), with the same initial wealth x , as functions of the horizon state price density ξ . Here, $\underline{\xi} := F_{\xi}^{-1}(1 - z_2^*)$, $\bar{\xi} := F_{\xi}^{-1}(1 - z_1^*)$, and $\bar{x} := (u')^{-1}(\lambda_1^* F_{\xi}^{-1}(1 - z_2^*)) = (u')^{-1}(\lambda_1^* \underline{\xi})$. When ξ is sufficiently large, the terminal wealth of the ES agent is lower than that of the benchmark agent. We emphasize that the optimal terminal wealth under the ES constraint is fundamentally different from that under the LEL constraint where the present value of the agent's losses are constrained. Under the LEL constraint, the optimal terminal wealth is strictly larger than that of a benchmark agent for ξ large enough (Figure 6 in Basak and Shapiro (2001)).

2.4 Properties

In this section, we perform a detailed analysis of trading behaviors under different risk measures. WVaR-RM exhibits a much richer variety of investment behaviors than its mean-risk counterpart (He et al. (2015)). He et al. (2015) find that the mean-WVaR is likely to be ill-posed, and the asymptotically optimal strategy is binary, which is to bank most of the money and invest the remainder in an extremely risky but highly rewarding lottery. In the WVaR-RM, the presence of the utility offers a variety of interesting features. We first study the existence of optimal solutions and then characterize risk-taking behaviors.

2.4.1 Existence of Optimal Solutions

Because the optimal value of (2.8) is always finite, we claim that if the optimal solution does not exist, the model is unattainable, i.e., the optimality of (2.8) cannot be achievable by any admissible portfolio.

When the risk constraint is active, Theorem 2.3.1 characterizes a class of risk measures that will lead to unattainability. When $\limsup_{z \uparrow 1} \kappa_{\Phi}([z, 1]) = \infty$, i.e., the RRPC is infinity in the extremely good states as the risk measure places too many weights, the optimal solution does not exist for all non-trivial levels of risk constraints. When the risk is measured by this type of risk measure, investments in the extremely good states are not only the cheapest but also the most efficient way to reduce the risk. The agent is thus incentivized to meet the risk constraint by assuming the greatest possible risk exposure. However, the risk aversion (implied by the strictly concave utility) prevents the agent from taking extremely risky positions. Such a conflict will result in an unattainable model.

We now examine this circumstance for the risk measures discussed in Section 2.1.

Proposition 2.4.1. *We have the following assertions:*

1. *For the negative expectation, $\lim_{z \uparrow 1} \kappa_{\Phi}([z, 1]) = \infty$.*
2. *For the VaR_{α} , $0 < \alpha < 1$, $\lim_{z \uparrow 1} \kappa_{\Phi}([z, 1]) = 0$.*
3. *For the ES_{α} , $0 < \alpha < 1$, $\lim_{z \uparrow 1} \kappa_{\Phi}([z, 1]) = 0$.*
4. *For the exponential spectral risk measures, $\lim_{z \uparrow 1} \kappa_{\Phi}([z, 1]) = \infty$.*
5. *For the power spectral risk measures with $0 < \gamma < 1$, $\lim_{z \uparrow 1} \kappa_{\Phi}([z, 1]) = \infty$.*

If we further assume that ξ is log-normally distributed, i.e., $F_\xi(x) = \Phi_N\left(\frac{\ln x - \mu_\xi}{\sigma_\xi}\right)$, for some μ_ξ and $\sigma_\xi > 0$, we have the following:

6. For the power spectral risk measures with $\gamma > 1$, $\lim_{z \uparrow 1} \kappa_\Phi([z, 1]) = 0$.

7. For the Wang (2000) risk measure,

$$\lim_{z \uparrow 1} \kappa_\Phi([z, 1]) = \begin{cases} 0, & q < \Phi_N(-\sigma_\xi); \\ e^{-\frac{1}{2}(\Phi^{-1}(q))^2 - \mu_\xi}, & q = \Phi_N(-\sigma_\xi); \\ +\infty, & q > \Phi_N(-\sigma_\xi). \end{cases}$$

8. For the beta family of distortion risk measures,

$$\lim_{z \uparrow 1} \kappa_\Phi([z, 1]) = \begin{cases} 0, & b > 1; \\ +\infty, & 0 < b \leq 1. \end{cases}$$

Proof of Proposition 2.4.1. 1-5 is trivial. Now assume that $F_\xi(x) = \Phi_N\left(\frac{\ln x - \mu_\xi}{\sigma_\xi}\right)$. It is a simple exercise to show that $\lim_{z \downarrow 0} \frac{\Phi_N^{-1}(z)}{\ln z} = 0$. We then have

6. For the power spectral risk measures with $\gamma > 1$,

$$\lim_{z \uparrow 1} \kappa_\Phi([z, 1]) = \lim_{z \downarrow 0} \gamma e^{(\gamma-1-\sigma_\xi \frac{\Phi_N^{-1}(z)}{\ln z}) \ln z - \mu_\xi} = 0.$$

7. For the Wang (2000) risk measure,

$$\begin{aligned} \lim_{z \uparrow 1} \kappa_\Phi([z, 1]) &= \lim_{z \uparrow 1} \frac{e^{\Phi_N^{-1}(q)\Phi_N^{-1}(z) - \frac{1}{2}(\Phi_N^{-1}(q))^2}}{e^{-\sigma_\xi \Phi_N^{-1}(z) + \mu_\xi}} \\ &= \lim_{z \uparrow 1} e^{(\sigma_\xi + \Phi_N^{-1}(q)\Phi_N^{-1}(z) - \frac{1}{2}(\Phi_N^{-1}(q))^2 - \mu_\xi)} \\ &= \begin{cases} 0, & q < \Phi_N(-\sigma_\xi); \\ e^{-\frac{1}{2}(\Phi_N^{-1}(q))^2 - \mu_\xi}, & q = \Phi_N(-\sigma_\xi); \\ \infty, & q > \Phi_N(-\sigma_\xi). \end{cases} \end{aligned}$$

8. For the beta family of distortion risk measures,

$$\begin{aligned} \lim_{z \uparrow 1} \kappa_\Phi([z, 1]) &= \lim_{z \uparrow 1} \frac{\frac{1}{\beta(a,b)} z^{a-1} (1-z)^{b-1}}{e^{\sigma_\xi \Phi_N^{-1}(1-z) + \mu_\xi}} \\ &= \lim_{z \downarrow 0} \frac{1}{\beta(a,b)} (1-z)^{a-1} e^{(b-1) \ln z - \sigma_\xi \Phi_N^{-1}(z) - \mu_\xi} \\ &= \begin{cases} 0, & b > 1; \\ +\infty, & 0 < b \leq 1. \end{cases} \end{aligned}$$

□

Coherent risk measures are believed to be better alternatives to traditional risk measures. However, the above analysis reveals that even law-invariant, coherent, comonotonic additive risk measures can still be inappropriate in the context of portfolio selection. In contrast, distortion risk measures that are widely used in the actuarial sciences, depending on the parameters, might or might not be appropriate for the purpose of risk management. Economic agents should use caution when adopting these risk measures.

Overall, we suggest that a “good” risk measure, for the purpose of risk management in asset allocation, should only focus on the downside risk.

2.4.2 Impacts on Asset Allocation

We now give a detailed analysis of how risk measures affect asset allocation, assuming that the risk constraint is active and the optimal solution exists. Throughout this section, we assume that $F_\xi^{-1}(\cdot)$ is differentiable. We focus on risk measures such that $\limsup_{z \uparrow 1} \kappa_\Phi([z, 1]) = 0$ and $\Phi([z, 1])$ (as a function of z) is twice differentiable on $(z_\Phi, 1)$ for some $z_\Phi \in [0, 1)$, and we assume that $R_\Phi(x) < \underline{x} < S_\Phi(x)$. Consequently, X^* , the optimal solution to (2.8), is given by (2.17).

First, we focus on the potential gains from the stock market.

Proposition 2.4.2. $\text{ess sup } X^* = +\infty$.

Proof of Proposition 2.4.2. From the proof of Lemma 2.3.6, we have $\delta'_{\lambda_2}(1-) = 0$.

Thus

$$\begin{aligned} \lim_{\xi \downarrow 0} X^* &= \lim_{\xi \downarrow 0} (u')^{-1}(\lambda_1^* \delta'_{\lambda_2^*}((1 - F_\xi(\xi)) +)) \\ &= \lim_{z \uparrow 1} (u')^{-1}(\lambda_1^* \delta'_{\lambda_2^*}(z +)) \\ &\geq \lim_{z \uparrow 1} (u')^{-1}(\lambda_1^* \delta'_{\lambda_2^*}(z -)) \\ &= +\infty. \end{aligned}$$

□

In the mean-WVaR (He et al. (2015)), although the investors can benefit from the stock market, the reward is capped and fixed. The investors receive the same amount of reward when the market is sufficiently good, regardless of how good it is, which makes such strategies less appealing. In contrast, under the WVVaR-RM, the potential gains from the stock market are unbounded. This could be attractive to some investors: although risk management is costly, they can still participate in the potential unlimited gains.

Gains and losses are always associated. Although the gains under WVaR-RM are unbounded irrespective of the risk measures, the losses are in a more complex situation. We first characterize a class of risk measures that can give rise to endogenous portfolio insurance. A portfolio insurance trading strategy is defined as one which guarantees a minimum level of wealth at some specified horizon, yet also participates in the potential gains of some reference portfolio (Luskin (1988); Grossman and Vila (1989)). Under portfolio insurance, the agent's downside risk is significantly reduced because all losses are capped at a prescribed level.

Proposition 2.4.3. *If $\liminf_{z \downarrow 0} \kappa_{\Phi}([0, z]) = +\infty$, then $\text{ess inf } X^* > 0$.*

To prove Proposition 2.4.3, we first show a lemma.

Lemma 2.4.1.

1. *If $\limsup_{z \downarrow 0} \frac{\varphi_{\lambda_2^*}(z-) - \varphi_{\lambda_2^*}(0-)}{z} < +\infty$, then $\delta'_{\lambda_2^*}(0+) < +\infty$ and $\text{ess inf } X^* > 0$.*
2. *If $\limsup_{z \downarrow 0} \frac{\varphi_{\lambda_2^*}(z-) - \varphi_{\lambda_2^*}(0-)}{z} = +\infty$, then $\delta'_{\lambda_2^*}(0+) = +\infty$ and $\text{ess inf } X^* = 0$.*

Proof of Lemma 2.4.1.

1. If $\limsup_{z \downarrow 0} \frac{\varphi_{\lambda_2^*}(z-) - \varphi_{\lambda_2^*}(0-)}{z} < +\infty$, we claim $\delta'_{\lambda_2^*}(0+) < +\infty$. Otherwise, $\delta'_{\lambda_2^*}(0+) = +\infty$. For any $a > \limsup_{z \downarrow 0} \frac{\varphi_{\lambda_2^*}(z-) - \varphi_{\lambda_2^*}(0-)}{z}$, there exists $b \in (0, 1]$ such that $\frac{\varphi_{\lambda_2^*}(z-) - \varphi_{\lambda_2^*}(0-)}{z} < a < \frac{\delta_{\lambda_2^*}(z) - \delta_{\lambda_2^*}(0)}{z}$, $z \in (0, b]$. Thus, on $(0, b)$, $\delta_{\lambda_2^*}(z) > \varphi_{\lambda_2^*}(z-)$, $\delta_{\lambda_2^*}(z)$ is affine, but $\delta'_{\lambda_2^*}(z+) = +\infty$, which is a contradiction. Thus, we have

$$\begin{aligned} \lim_{\xi \uparrow \infty} X^* &= \lim_{\xi \uparrow \infty} (u')^{-1}(\lambda_1^* \delta'_{\lambda_2^*}((1 - F_{\xi}(\xi)) +)) \\ &= \lim_{z \downarrow 0} (u')^{-1}(\lambda_1^* \delta'_{\lambda_2^*}(z+)) \\ &> 0. \end{aligned}$$

2. If $\limsup_{z \downarrow 0} \frac{\varphi_{\lambda_2^*}(z-) - \varphi_{\lambda_2^*}(0-)}{z} = +\infty$,

$$\begin{aligned} \delta'_{\lambda_2^*}(0+) &= \limsup_{z \downarrow 0} \frac{\delta_{\lambda_2^*}(z) - \delta_{\lambda_2^*}(0)}{z} \\ &\geq \limsup_{z \downarrow 0} \frac{\varphi_{\lambda_2^*}(z-) - \varphi_{\lambda_2^*}(0-)}{z} \\ &= \infty, \end{aligned}$$

and

$$\begin{aligned} \lim_{\xi \uparrow +\infty} X^* &= \lim_{\xi \uparrow +\infty} (u')^{-1}(\lambda_1^* \delta'_{\lambda_2^*}((1 - F_{\xi}(\xi)) +)) \\ &= \lim_{z \downarrow 0} (u')^{-1}(\lambda_1^* \delta'_{\lambda_2^*}(z+)) \\ &= 0. \end{aligned}$$

□

Proof of Proposition 2.4.3. We have

$$\begin{aligned}
& \limsup_{z \downarrow 0} \frac{\varphi_{\lambda_2^*}(z-) - \varphi_{\lambda_2^*}(0-)}{z} \\
&= \limsup_{z \downarrow 0} \frac{(1 - \lambda_2^* \kappa_{\Phi}([0, z])) \int_{[0, z]} F_{\xi}^{-1}(1 - s) ds}{z} \\
&\leq \limsup_{z \downarrow 0} F_{\xi}^{-1}(1 - z) \cdot (1 - \liminf_{z \downarrow 0} \lambda_2^* \kappa_{\Phi}([0, z])) \\
&< 0.
\end{aligned}$$

The claim then follows from Lemma 2.4.1. □

This proposition says that if a risk measure's RRPC in the extremely bad states (such as a catastrophic loss) is infinity, then the agent will insure against these states endogenously. Even though these states are the most expensive states to insure against, the risk constraint incentivizes the agent to follow the portfolio insurance strategy because it is the most efficient way to satisfy the requirement. This could be of interest to regulators, because portfolio insurance is, in general, costly. As noted in Leland (1980) and Benninga and Blume (1985), it is highly unlikely that an investor would utilize such a strategy in a complete market. Regulators can encourage economic agents to do so by imposing a risk constraint of this type.

We also provide additional evidence of why people buy portfolio insurance. Although much work has been conducted on the effects on prices of the presence of portfolio insurance in the economy (Grossman and Vila (1989); Basak (1995); Grossman and Zhou (1996); Basak (2002)), in which investors are assumed to be portfolio insurers, justifications for the necessity of portfolio insurance are limited, especially in complete markets. To the best of our knowledge, the only work on why investors use portfolio insurance in complete markets is He and Zhou (2016), who find in a rank-dependent portfolio choice model that a sufficiently high level of fear endogenously necessitates portfolio insurance. We suggest that people can use portfolio insurance strategies as a means to manage their market-risk exposure.

However, not all risk measures within WVaR can lead to portfolio insurance. The following theorem characterizes another class of risk measures that can increase risk exposure because they result in larger losses when losses occur.

Proposition 2.4.4. *If $\liminf_{z \downarrow 0} \kappa_{\Phi}([0, z]) = 0$, then*

1. $\text{ess inf } X^* = 0$.

2. If, in addition, $\limsup_{z \downarrow 0} \kappa_\Phi([0, z]) = 0$ and there exists $z_\Phi \in (0, 1]$ such that $\Phi([0, z])$ (as a function of z) is twice differentiable on $(0, z_\Phi)$, then there exists $\bar{\xi}$ such that $X^* < X_0^*$ when $\xi > \bar{\xi}$, where X_0^* is the benchmark agent's optimal wealth given by (2.6).

Proof of Proposition 2.4.4.

1. Because $\liminf_{z \downarrow 0} \kappa_\Phi([0, z]) = 0$, we can find a sequence $z_n \downarrow 0$ such that $\lim_{n \rightarrow \infty} \kappa_\Phi([0, z_n]) = 0$. We then have

$$\begin{aligned}
& \limsup_{z \downarrow 0} \frac{\varphi_{\lambda_2^*}(z-) - \varphi_{\lambda_2^*}(0-)}{z} \\
&= \limsup_{z \downarrow 0} \frac{(1 - \lambda_2^* \kappa_\Phi([0, z])) \int_{[0, z]} F_\xi^{-1}(1-s) ds}{z} \\
&\geq \lim_{n \rightarrow \infty} \frac{(1 - \lambda_2^* \kappa_\Phi([0, z_n])) \int_{[0, z_n]} F_\xi^{-1}(1-s) ds}{z_n} \\
&> \frac{1}{2} \lim_{z \downarrow 0} F_\xi^{-1}(1-z) \\
&= +\infty,
\end{aligned}$$

and the first claim follows from Lemma 2.4.1.

2. From Lemma 2.3.8, we have $\lambda_1^* \geq \lambda_0^*$. If $\lambda_1^* = \lambda_0^*$, then the budget/risk constraint cannot be satisfied simultaneously. Thus, $\lambda_1^* > \lambda_0^*$.

From Lemma 2.4.1, $\delta'_{\lambda_2^*}(0+) = +\infty$. Because $\Phi([0, z])$ (as a function of z) is twice differentiable on $(0, z_\Phi)$, we have $\Phi([0, z]) = \int_{[0, z]} \phi(s) ds$ for $z \in (0, z_\Phi)$ and some $\phi(\cdot)$. Similar to the proof of Lemma 2.3.6, we can show that there exists $b \in (0, 1]$ such that $\delta_{\lambda_2^*}(z) = \varphi_{\lambda_2^*}(z-)$, $z \in (0, b)$. Next, we can find $c \in (0, b)$ such that $\frac{\phi(z)}{F_\xi^{-1}(1-z)} < (1 - \frac{\lambda_0^*}{\lambda_1^*}) \frac{1}{\lambda_2^*}$, $z \in (0, c)$. For $z \in (0, c)$, $\delta'_{\lambda_2^*}(z+) = \varphi'_{\lambda_2^*}(z) = F_\xi^{-1}(1-z) - \lambda_2^* \phi(z) > \frac{\lambda_0^*}{\lambda_1^*} F_\xi^{-1}(1-z)$. Thus, for $\xi > \bar{\xi} := F_\xi^{-1}(1-c)$, $X^* < X_0^*$.

□

Remark 2.4.1. If $\lim_{z \downarrow 0} \kappa_\Phi([0, z]) = a \in (0, +\infty)$, whether there is portfolio insurance depends on λ_2^* . Note that

$$\begin{aligned}
\lim_{z \downarrow 0} \frac{\varphi_{\lambda_2^*}(z-) - \varphi_{\lambda_2^*}(0-)}{z} &= \lim_{z \downarrow 0} \frac{(1 - \lambda_2^* \kappa_\Phi([0, z])) \int_{[0, z]} F_\xi^{-1}(1-s) ds}{z} \\
&= \begin{cases} \infty, & \lambda_2^* < \frac{1}{a}; \\ -\infty, & \lambda_2^* > \frac{1}{a}. \end{cases}
\end{aligned}$$

If a risk measure's RRPC for catastrophic losses is 0, then the agent will ignore these losses and leave himself completely uninsured, incurring all losses, because it is costly and inefficient to insure against these losses. Moreover, in the bad states ($\xi > \bar{\xi}$), the terminal wealth is typically lower than it would have been in the absence of the risk constraint. In other words, under such regulations, if a large loss occurs, then it is likely to be an even larger loss compared to the benchmark agent and consequently, the probability of extreme losses is higher. The economic agent exploits differences between a portfolio's true economic risks and the measurements of risk. This perverse consequence is often referred to as "regulatory capital arbitrage" (Jones (2000)) that banks can reduce substantially their regulatory measures of risk, with little or no corresponding reduction in their overall economic risks.

This could be a source of concern for regulators and real-world risk managers. Risk measures are viewed by many as a tool to shield economic agents from large losses which could drive them out of business. However, many risk measures, although they have some desirable properties such as law-invariance, coherence, and comonotonic-additivity, could backfire, and thus would be more likely to lead to credit and solvency problems, defeating the purpose of such regulations. Such an undesirable property has been observed in Basak and Shapiro (2001) for VaR, but to the best of our knowledge, we are the first to characterize "regulatory capital arbitrage" for general risk measures.

Based on the above characterizations, we now examine the risk measures discussed in Section 2.1. It is easy to see that there exists $z_\Phi \in (0, 1]$ such that $\Phi([0, z])$ (as a function of z) is twice differentiable on $(0, z_\Phi)$ for all of the aforementioned risk measures.

Proposition 2.4.5. *We have the following assertions:*

1. For the VaR_α , $0 < \alpha < 1$, $\liminf_{z \downarrow 0} \kappa_\Phi([0, z]) = 0$.
2. For the ES_α , $0 < \alpha < 1$, $\liminf_{z \downarrow 0} \kappa_\Phi([0, z]) = 0$.
3. For the power spectral risk measures with $\gamma > 1$, $\liminf_{z \downarrow 0} \kappa_\Phi([0, z]) = 0$.

If we further assume that ξ is log-normally distributed, i.e. $F_\xi(x) = \Phi_N(\frac{\ln x - \mu_\xi}{\sigma_\xi})$, for some μ_ξ and $\sigma_\xi > 0$, then we have the following:

4. For the Wang (2000) risk measure with $q < \Phi_N(-\sigma_\xi)$, $\liminf_{z \downarrow 0} \kappa_\Phi([0, z]) = +\infty$.

5. For the beta family of distortion risk measures with $b > 1$,

$$\liminf_{z \downarrow 0} \kappa_{\Phi}([0, z]) = \begin{cases} 0, & a \geq 1; \\ +\infty, & 0 < a < 1. \end{cases}$$

Proof of Proposition 2.4.5. 1-3 is trivial. Now assume that $F_{\xi}(x) = \Phi_N(\frac{\ln x - \mu_{\xi}}{\sigma_{\xi}})$. Recall that $\lim_{z \downarrow 0} \frac{\Phi_N^{-1}(z)}{\ln z} = 0$. We have the following:

4. For the Wang (2000) risk measure with $q < \Phi_N(-\sigma_{\xi})$,

$$\liminf_{z \downarrow 0} \kappa_{\Phi}([0, z]) = \lim_{z \downarrow 0} \frac{e^{\Phi_N^{-1}(q)\Phi_N^{-1}(z) - \frac{1}{2}(\Phi_N^{-1}(q))^2}}{e^{-\sigma_{\xi}\Phi_N^{-1}(z) + \mu_{\xi}}} = +\infty.$$

5. For the beta family of distortion risk measures with $b > 1$,

$$\begin{aligned} \liminf_{z \downarrow 0} \kappa_{\Phi}([0, z]) &= \lim_{z \downarrow 0} \frac{\frac{1}{\beta(a, b)} z^{a-1} (1-z)^{b-1}}{e^{-\sigma_{\xi}\Phi_N^{-1}(z) + \mu_{\xi}}} \\ &= \lim_{z \downarrow 0} \frac{1}{\beta(a, b)} (1-z)^{b-1} e^{(a-1)\ln z + \sigma_{\xi}\Phi_N^{-1}(z) - \mu_{\xi}} \\ &= \begin{cases} 0, & a \geq 1; \\ +\infty, & 0 < a < 1. \end{cases} \end{aligned}$$

□

To our surprise, distortion risk measures such as the Wang (2000) risk measure and the beta family of distortion risk measures, although originally designed as premium principles and capital adequacy principles, can give rise to endogenous portfolio insurance and thus reduce the magnitude of losses (with a suitable choice of parameters) because they account for the severity of extreme losses. Figure 2.3 plots the optimal terminal wealth under the Wang (2000) risk measure. In the bad states of the market, the agent behaves like a portfolio insurer and his terminal wealth is higher than that of a benchmark agent when ξ is sufficiently large.

However, VaR, ES, and some well-known spectral risk measures actually increase risk exposure in the bad states. The case of VaR is consistent with Basak and Shapiro (2001) and the empirical evidence provided by Berkowitz and O'Brien (2002), who document that when a bank suffers losses, such losses are often substantially larger than the bank's reported VaR. ES is proposed as an effective alternative to VaR in financial risk management (Artzner et al. (1999); Acerbi et al. (2001)). It is believed that ES will help to ensure a more prudent capture of "tail risk" and capital adequacy during periods of significant financial market stress (BCBS (2016)). However, we

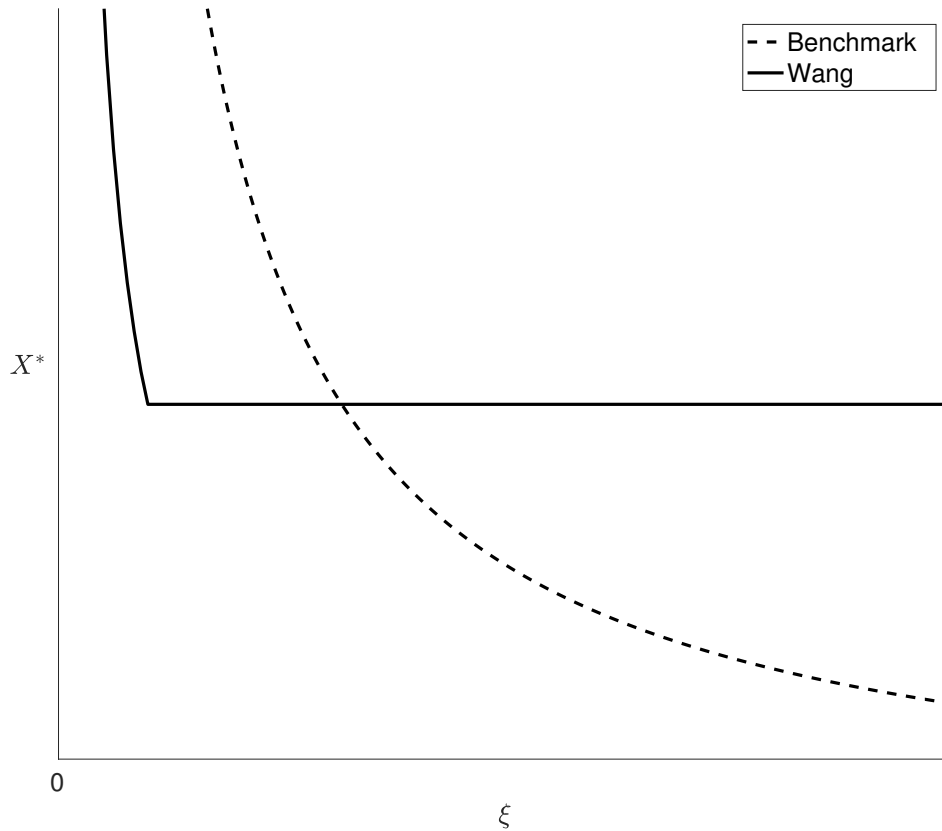


Figure 2.3: Optimal terminal wealth under the Wang (2000) risk measure. The figure plots the optimal terminal wealth of a benchmark agent (without the risk constraint, dashed line) and a Wang agent (with a risk constraint and the risk is measured by the Wang (2000) risk measure, solid line), with the same initial wealth x , as functions of the horizon state price density ξ . We assume that $F_\xi(x) = \Phi_N\left(\frac{\ln x - \mu\xi}{\sigma_\xi}\right)$ and $q < \Phi_N(-\sigma_\xi)$. When ξ is sufficiently large, the terminal wealth of the Wang agent is higher than that of the benchmark agent.

show that this objective might not be achieved because ES actually increases the magnitude of the loss when a loss occurs instead of reducing it. Figure 2.2 plots the optimal terminal wealth under the ES constraint (solid line). When the market is sufficiently bad, the ES agent suffers from larger losses than a benchmark agent. One common criticism of VaR is that it fails to account for the magnitude of losses. Our results reveal that even though ES accounts for the sizes of the losses, it is far from adequate. ES places equal weights for all levels of losses that exceed a certain threshold. However, the costs to insure against these losses differ, and it is costlier to insure against a larger loss. Thus, the agent finds it inefficient to insure against catastrophic losses when the risk is measured by ES.

Overall, to mitigate or even preclude “regulatory capital arbitrage” in asset management, we suggest that the regulatory risk measure’s sensitivity to losses should be relevant to the severity of the losses. Distortion risk measures, such as the Wang (2000) risk measure and the beta family of distortion risk measures should be preferred over current regulatory risk measures such as VaR and ES.

2.5 Concluding Remarks

Although we have shown how various risk measures can alter asset allocation patterns, our analysis is in a partial equilibrium setting. It is equally or even more important to study how they can affect the market price dynamics. In the next chapter, we will study both the partial equilibrium and the general equilibrium of an economy that features agents who must manage their ES-measured risks.

Chapter 3

Equilibrium Analysis of Expected Shortfall¹

This chapter continues the line of research on risk management from the portfolio selection's perspective. We study the impact of Expected Shortfall (ES) on portfolio selection and assets prices. We have obtained the optimal terminal wealth under the WVaR constraint in the last chapter. Although ES is a special case of WVaR, the optimal terminal wealth under the ES constraint is determined up to the concave envelope of a function which was derived graphically in the last chapter. In this chapter, we rigorously derive the explicit expression of the required concave envelope and corresponding optimal terminal wealth of an expected utility maximizer with the ES constraint. We then perform a detailed analysis of the portfolio weights under the ES-based constraint.

The second part of this chapter focuses on general equilibrium analysis of ES. In particular, we explore how the introduction of ES-based constraints can affect the market volatility. Abundant work has been done on the impact of different risk measures on portfolio selection. In particular, we have demonstrated in the last section that many risk measures, including VaR and ES, can lead to larger losses when losses occur, defeating the purpose of risk measures. However, general equilibrium analyses of risk measures on market dynamics are limited, except Basak and Shapiro (2001); Leippold et al. (2006) for the case of VaR. Therefore, we complement the literature by studying how ES affect risk-taking behaviors in general equilibrium. To our best knowledge, we are the first to study the impact of ES on market dynamics. We develop a production and a pure-exchange general equilibrium model featuring ES risk managers respectively. We find that the presence of the ES constraint increases

¹This chapter is based on my working paper Wei (2017a) available online at SSRN.

the market volatility during periods of significant financial market stress in both economies. These findings are similar to those of VaR.

The rest of the chapter is organized as follow: In Section 3.1, we formulate the Expected-Shortfall risk management (ES-RM), which is then analyzed in Section 3.2. Section 3.3 carries out general equilibrium analysis. Section 3.4 concludes.

3.1 Model

In this section, we formulate our ES-based risk management (ES-RM) problem. We consider a financial market identical to that described in the last chapter and follow the notations accordingly, except that we denote by $\pi(t)$ the fraction of wealth invested in stocks.

We also consider the Benchmark agent who solves the following problem,

$$\begin{aligned} \max_{X(T)} \quad & \mathbb{E}[u(X(T))] \\ \text{subject to} \quad & \mathbb{E}[\xi X(T)] \leq X(0), \end{aligned} \tag{3.1}$$

and the VaR agent who solves

$$\begin{aligned} \max_{X(T)} \quad & \mathbb{E}[u(X(T))] \\ \text{subject to} \quad & \mathbb{E}[\xi X(T)] \leq X(0), \\ & \mathbb{P}(X(T) \geq \underline{X}) \geq 1 - \alpha. \end{aligned} \tag{3.2}$$

In addition, we consider the ES agent who solves

$$\begin{aligned} \max_{X(T)} \quad & \mathbb{E}[u(X(T))] \\ \text{subject to} \quad & \mathbb{E}[\xi X(T)] \leq X(0), \\ & ES_\alpha(X(T)) \leq -\underline{X}. \end{aligned} \tag{3.3}$$

If $\alpha = 0$, then the agent is a portfolio insurer; if $\alpha = 1$, then the optimal solution does not exist, as shown in the last chapter. We have graphically derived the optimal solution to (3.3) without a proof. In the next section, we make our analysis concrete and rigorous.

3.2 Partial Equilibrium

In this section, we perform partial equilibrium analysis for the benchmark agent problem (3.1), the VaR agent problem (3.2) and the ES agent problem (3.3). We

follow the general scheme developed in the last chapter to solve the problem. To this end, we impose the assumptions on ξ as in the last chapter.

We will not discuss feasibility and well-posedness of the problem, as these issues have been examined thoroughly in the last chapter for general risk measures. We assume all parameters are within reasonable ranges so that optimal solutions exist.

3.2.1 Optimal Solution

The following proposition characterizes the optimal solutions to (3.1), (3.2), and (3.3).

Proposition 3.2.1. *We have the following assertions:*

1. *The optimal time- T wealth of the Benchmark agent (3.1) is*

$$X_B(T) = I(\lambda_B \xi),$$

where $I(\cdot) = (u')^{-1}(\cdot)$ and λ_B solves $\mathbb{E}[\xi X_B(T)] = X(0)$.

2. *The optimal time- T wealth of the VaR agent (3.2) is*

$$X_{VaR}(T) = \begin{cases} I(\lambda_{VaR} \xi) & \xi < \underline{\xi}_{VaR}, \\ \underline{X} & \underline{\xi}_{VaR} \leq \xi < \bar{\xi}_{VaR}, \\ I(\lambda_{VaR} \xi) & \bar{\xi}_{VaR} \leq \xi, \end{cases}$$

where $\underline{\xi}_{VaR} \equiv u'(\underline{X})/\lambda_{VaR}$, $\bar{\xi}_{VaR}$ is such that $\mathbb{P}(\xi > \bar{\xi}_{VaR}) \equiv \alpha$, and $\lambda_{VaR} \geq 0$ solves $\mathbb{E}[\xi X_{VaR}(T)] = X(0)$. The VaR constraint in (3.2) is binding if, and only if, $\underline{\xi}_{VaR} < \bar{\xi}_{VaR}$.

3. *The optimal time- T wealth of the ES agent (3.3) is*

$$X_{ES}(T) = \begin{cases} I(\lambda_{ES} \xi) & \xi < \underline{\xi}_{ES}, \\ I(\lambda_{ES} \underline{\xi}_{ES}) & \underline{\xi}_{ES} \leq \xi < \bar{\xi}_{ES}, \\ I(\lambda_{ES} \xi - \frac{1}{\alpha} \lambda_{ES} \mu_{ES}) & \bar{\xi}_{ES} \leq \xi, \end{cases}$$

where $\underline{\xi}_{ES} > 0$ solves $h_{\mu_{ES}}(1 - F_{\xi}(\underline{\xi}_{ES})) = 0$, $\bar{\xi}_{ES} = \underline{\xi}_{ES} + \frac{1}{\alpha} \mu_{ES}$, and $\lambda_{ES} > 0, \mu_{ES} \geq 0$ solve

$$\mathbb{E}[\xi X_{ES}(T)] = X(0)$$

$$ES_{\alpha}(X_{ES}(T)) \leq -\underline{X}.$$

$h_{\mu}(\cdot)$ is define by (3.5). The ES constraint is binding, if and only if $ES_{\alpha}(X_B(T)) = -\underline{X}$ (or $\mu_{ES} > 0$). Moreover, if the ES constraint is binding, $\mathbb{P}(\xi > \underline{\xi}_{ES}) > \alpha, \mathbb{P}(\xi > \bar{\xi}_{ES}) < \alpha$.

Remark 3.2.1. We emphasize that the optimal terminal wealth under the ES constraint is different from that under the LEL constraint in Basak and Shapiro (2001). LEL measures the present value (under the risk-neutral measure) of the agent's loss while ES measures the expected loss which is not deflated by state prices. Therefore, the third result in Proposition 3.2.1 does not follow from Proposition 4 in Basak and Shapiro (2001).

3.2.2 Proof of Proposition 3.2.1

The first claim is standard. The second can be found in Basak and Shapiro (2001). We now follow the arguments in the last chapter to solve the ES agent problem.

As illustrated in the last chapter, we have

$$\varphi_{\lambda_2}(z-) = \varphi_{\lambda_2}(z) = \begin{cases} -\int_{[z,1]} F_{\xi}^{-1}(1-s)ds + \lambda_2(1 - \frac{z}{\alpha}), & z \in [0, \alpha]; \\ -\int_{[z,1]} F_{\xi}^{-1}(1-s)ds, & z \in (\alpha, 1]. \end{cases}$$

The solid line in Figure 2.1 plots $\varphi_{\lambda_2}(\cdot-)$. We expect its concave envelope $\delta_{\lambda_2}(\cdot)$ replaces part of $\varphi_{\lambda_2}(\cdot-)$ with a chord (dashed line) between z_1 and z_2 , at which the slopes of $\varphi_{\lambda_2}(\cdot-)$ equal the slope of the chord, in other words,

$$\delta_{\lambda_2}(z) = \begin{cases} \varphi_{\lambda_2}(z-), & z \in [0, z_1]; \\ F_{\xi}^{-1}(1-z_2)(z-z_1) + \varphi_{\lambda_2}(z_1-), & z \in (z_1, z_2]; \\ \varphi_{\lambda_2}(z-), & z \in (z_2, 1], \end{cases}$$

where $z_1 < \alpha$ and $z_2 > \alpha$ satisfy the following condition

$$\begin{cases} \delta_{\lambda_2}(z_2) = \varphi_{\lambda_2}(z_2-), \\ F_{\xi}^{-1}(1-z_1) - \frac{\lambda_2}{\alpha} = F_{\xi}^{-1}(1-z_2). \end{cases} \quad (3.4)$$

We now show the existence of the solution to (3.4). First, we derive $\mathcal{A}_{ES} := \{\lambda_2 : \lambda_2 > 0, \delta'_{\lambda_2}(z+) > 0, z \in [0, 1)\}$.

Lemma 3.2.1.

$$\mathcal{A}_{ES} = \{\lambda_2 : \lambda_2 > 0 \text{ and } \varphi_{\lambda_2}(1 - F_{\xi}(\frac{\lambda_2}{\alpha})) < 0\}$$

Proof of Lemma 3.2.1. First, if $\lambda_2 \leq \alpha F_{\xi}^{-1}(1-\alpha)$, then $\varphi'_{\lambda_2}(z+) > 0, z \in [0, 1)$ and consequently $\varphi_{\lambda_2}(z) < 0, z \in [0, 1)$.

Next, for any $\lambda_2 > \alpha F_{\xi}^{-1}(1-\alpha)$, $\varphi'_{\lambda_2}(z) > 0, z \in [0, 1 - F_{\xi}(\frac{\lambda_2}{\alpha})]$, $\varphi'_{\lambda_2}(1 - F_{\xi}(\frac{\lambda_2}{\alpha})) = 0$, $\varphi'_{\lambda_2}(z) < 0, z \in (1 - F_{\xi}(\frac{\lambda_2}{\alpha}), \alpha)$, and $\varphi'_{\lambda_2}(z) > 0, z \in (\alpha, 1)$. Therefore, $\varphi_{\lambda_2}(z) < 0, z \in [0, 1)$ if and only if $\varphi_{\lambda_2}(1 - F_{\xi}(\frac{\lambda_2}{\alpha})) < 0$. The rest of the proof follows from Lemma 2.3.5 in Chapter 2. \square

For $\alpha < z \leq \bar{z} := 1 - F_\xi((F_\xi^{-1}(1 - \alpha) - \frac{\lambda_2}{\alpha}) \vee 0)$, define

$$s_{\lambda_2}(z) = 1 - F_\xi(F_\xi^{-1}(1 - z) + \frac{\lambda_2}{\alpha}),$$

and we have $\varphi'_{\lambda_2}(s_{\lambda_2}(z)) = F_\xi^{-1}(1 - z) = \varphi'_{\lambda_2}(z)$, $0 < s_{\lambda_2}(z) < \alpha < z \leq \bar{z}$. Define

$$h_{\lambda_2}(z) = \varphi_{\lambda_2}(z) - \varphi_{\lambda_2}(s_{\lambda_2}(z)) - \varphi'_{\lambda_2}(z)(z - s_{\lambda_2}(z)). \quad (3.5)$$

The following lemma shows the existence of z_2 .

Lemma 3.2.2. *For $\lambda_2 \in \mathcal{A}_{ES}$, there exist $\alpha < z_2 < 1$ such that $h_{\lambda_2}(z) < 0$ for $\alpha < z < z_2$, $h_{\lambda_2}(z_2) = 0$, and $h_{\lambda_2}(z) > 0$, $z_2 < z < 1$.*

Proof of Lemma 3.2.2. Note

$$\begin{aligned} h_{\lambda_2}(z) &= \varphi_{\lambda_2}(z) - \varphi_{\lambda_2}(s_{\lambda_2}(z)) - \varphi'_{\lambda_2}(z)(z - s_{\lambda_2}(z)) \\ &= \int_{s_{\lambda_2}(z)}^z F_\xi^{-1}(1 - t) dt - \lambda_2(1 - \frac{s_{\lambda_2}(z)}{\alpha}) - F_\xi^{-1}(1 - z)(z - s_{\lambda_2}(z)), \end{aligned}$$

we have

$$\begin{aligned} h_{\lambda_2}(z) &> F_\xi^{-1}(1 - z)(z - s_{\lambda_2}(z)) - \lambda_2(1 - \frac{s_{\lambda_2}(z)}{\alpha}) - F_\xi^{-1}(1 - z)(z - s_{\lambda_2}(z)) \\ &= -\lambda_2(1 - \frac{s_{\lambda_2}(z)}{\alpha}), \end{aligned}$$

and

$$\begin{aligned} h_{\lambda_2}(z) &< F_\xi^{-1}(1 - s_{\lambda_2}(z))(z - s_{\lambda_2}(z)) - \lambda_2(1 - \frac{s_{\lambda_2}(z)}{\alpha}) - (F_\xi^{-1}(1 - s_{\lambda_2}(z)) - \frac{\lambda_2}{\alpha})(z - s_{\lambda_2}(z)) \\ &= \frac{\lambda_2}{\alpha}(z - \alpha). \end{aligned}$$

Thus, $h_{\lambda_2}(\alpha+) < 0$. If $\alpha F_\xi^{-1}(1 - \alpha) \geq \lambda_2$, $\bar{z} = 1 - F_\xi(F_\xi^{-1}(1 - \alpha) - \frac{\lambda_2}{\alpha})$, $s_{\lambda_2}(\bar{z}) = \alpha$, and we have $h_{\lambda_2}(\bar{z}) > 0$. If $\alpha F_\xi^{-1}(1 - \alpha) < \lambda_2$, $\bar{z} = 1$, $s_{\lambda_2}(\bar{z}) = 1 - F_\xi(\frac{\lambda_2}{\alpha})$. By Lemma (3.2.1), we have

$$\begin{aligned} h_{\lambda_2}(1) &= \varphi_{\lambda_2}(1) - \varphi_{\lambda_2}(1 - F_\xi(\frac{\lambda_2}{\alpha})) - \varphi'_{\lambda_2}(1)(1 - 1 + F_\xi(\frac{\lambda_2}{\alpha})) \\ &= 0 - \varphi_{\lambda_2}(1 - F_\xi(\frac{\lambda_2}{\alpha})) \\ &> 0. \end{aligned}$$

Thus, $h_{\lambda_2}(\bar{z}) > 0$. Since $h(\cdot)$ is continuous, there exists at least one z_2 such that $h_{\lambda_2}(z_2) = 0$.

Next, for $\alpha < t_1 < t_2 \leq \bar{z}$,

$$\begin{aligned}
& h_{\lambda_2}(t_1) - h_{\lambda_2}(t_2) \\
&= [\varphi_{\lambda_2}(t_1) - \varphi_{\lambda_2}(t_2)] - [\varphi_{\lambda_2}(s_{\lambda_2}(t_1)) - \varphi_{\lambda_2}(s_{\lambda_2}(t_2))] \\
&\quad - [\varphi'_{\lambda_2}(t_1)t_1 - \varphi'_{\lambda_2}(t_2)t_2] + [\varphi'_{\lambda_2}(s_{\lambda_2}(t_1))s_{\lambda_2}(t_1) - \varphi'_{\lambda_2}(s_{\lambda_2}(t_2))s_{\lambda_2}(t_2)] \\
&= \int_{t_2}^{t_1} \varphi'_{\lambda_2}(z)dz - \int_{s_{\lambda_2}(t_2)}^{s_{\lambda_2}(t_1)} \varphi'_{\lambda_2}(z)dz - \left[\int_{t_2}^{t_1} \varphi'_{\lambda_2}(z)dz + \int_{t_2}^{t_1} z d\varphi'_{\lambda_2}(z) \right] \\
&\quad + \left[\int_{s_{\lambda_2}(t_2)}^{s_{\lambda_2}(t_1)} \varphi'_{\lambda_2}(z)dz + \int_{s_{\lambda_2}(t_2)}^{s_{\lambda_2}(t_1)} z d\varphi'_{\lambda_2}(z) \right] \\
&= \int_{s_{\lambda_2}(t_2)}^{s_{\lambda_2}(t_1)} z d\varphi'_{\lambda_2}(z) - \int_{t_2}^{t_1} z d\varphi'_{\lambda_2}(z) \\
&= \int_{t_2}^{t_1} s_{\lambda_2}(z) d\varphi'_{\lambda_2}(s_{\lambda_2}(z)) - \int_{t_2}^{t_1} z d\varphi'_{\lambda_2}(z) \\
&= - \int_{t_1}^{t_2} [s_{\lambda_2}(z) - z] d\varphi'_{\lambda_2}(z) \\
&< 0.
\end{aligned}$$

Thus, $h(\cdot)$ is strictly increasing. This completes the proof. \square

We can now character $\delta_{\lambda_2}(\cdot)$.

Proposition 3.2.2. For $\lambda_2 \in \mathcal{A}_{ES}$,

$$\delta_{\lambda_2}(z) = \begin{cases} \varphi_{\lambda_2}(z-), & z \in [0, z_1]; \\ F_{\xi}^{-1}(1 - z_2)(z - z_1) + \varphi_{\lambda_2}(z_1-), & z \in (z_1, z_2]; \\ \varphi_{\lambda_2}(z-), & z \in (z_2, 1], \end{cases}$$

and

$$\delta'_{\lambda_2}(z) = \begin{cases} F_{\xi}^{-1}(1 - z) - \frac{\lambda_2}{\alpha} & z \in [0, z_1], \\ F_{\xi}^{-1}(1 - z_2) & z \in (z_1, z_2], \\ F_{\xi}^{-1}(1 - z) & z \in (z_2, 1], \end{cases}$$

where $\alpha < z_2 < 1$ is the unique root of $h_{\lambda_2}(z) = 0$, and $z_1 = s_{\lambda_2}(z_2) \in (0, \alpha)$.

Proof of Proposition 3.2.2. $\delta_{\lambda_2}(\cdot)$ is obviously concave. Note $\delta_{\lambda_2}(z) = \varphi_{\lambda_2}(z)$, $z \in [0, z_1] \cup [z_2, 1]$, $\delta'_{\lambda_2}(z) > \varphi'_{\lambda_2}(z)$, $z \in (z_1, \alpha)$, and $\delta'_{\lambda_2}(z) < \varphi'_{\lambda_2}(z)$, $z \in (\alpha, z_2)$, we have $\delta_{\lambda_2}(z) > \varphi_{\lambda_2}(z)$, $z \in (z_1, z_2)$. Moreover, since $\delta'_{\lambda_2}(\cdot)$ is constant on (z_1, z_2) , we conclude δ_{λ_2} is the concave envelope of φ_{λ_2} . \square

Proof of Proposition 3.2.1. The claim follows from Theorem 2.3.1 in Chapter 2. \square

3.2.3 Properties

In this section we study trading behaviors under the ES-RM. We first compare optimal terminal wealth and then study portfolio weights under the additional assumptions of CRRA preferences and lognormal state prices.

3.2.3.1 Terminal Wealth

If $\underline{X} \leq I(\lambda_B F_\xi^{-1}(1 - \alpha))$, then the VaR agent becomes the Benchmark agent. Moreover, since $ES_\alpha(X_B(T)) = -\frac{1}{\alpha} \int_0^\alpha I(\lambda_B F_\xi^{-1}(1 - z)) dz > -I(\lambda_B F_\xi^{-1}(1 - \alpha))$, when the VaR constraint is binding, the ES constraint is also binding. To compare optimal wealth of different agents, we assume $\underline{X} > I(\lambda_B F_\xi^{-1}(1 - \alpha))$.

Proposition 3.2.3. *We have the following assertions:*

1. $I(\lambda_{ES} \underline{\xi}_{ES}) > \underline{X}$;
2. $\lambda_{ES} > \lambda_{VaR} > \lambda_B$;
3. $0 < \underline{\xi}_{ES} < \underline{\xi}_{VaR} < \bar{\xi}_{VaR} < \bar{\xi}_{ES}$;
4. $X_{VaR}(T) < X_B(T)$ when $\xi \geq \bar{\xi}_{VaR}$;
5. $X_{ES}(T) < X_B(T)$ when $\xi > \bar{\xi}_1 := \frac{\lambda_{ES} \mu_{ES}}{\alpha(\lambda_{ES} - \lambda_B)}$;
6. $X_{ES}(T) < X_{VaR}(T)$ when $\xi > \bar{\xi}_2 := \frac{\lambda_{ES} \mu_{ES}}{\alpha(\lambda_{ES} - \lambda_{VaR})}$. Moreover, $\bar{\xi}_2 > \bar{\xi}_1$.

Proof of Proposition 3.2.3. First, since $\mathbb{P}(\xi > \underline{\xi}_{ES}) > \alpha$, $\mathbb{P}(\xi > \bar{\xi}_{ES}) < \alpha$ and $\mathbb{P}(\xi > \bar{\xi}_{VaR}) = \alpha$, we have $\underline{\xi}_{ES} < \bar{\xi}_{VaR} < \bar{\xi}_{ES}$. Next, if $I(\lambda_{ES} \underline{\xi}_{ES}) \leq \underline{X}$, then the ES constraint cannot be satisfied as $X_{ES}(T) < I(\lambda_{ES} \underline{\xi}_{ES}) \leq \underline{X}$ when $\xi > \bar{\xi}_{ES}$. Thus, $I(\lambda_{ES} \underline{\xi}_{ES}) > \underline{X}$.

$\lambda_{VaR} > \lambda_B$ is due to Basak and Shapiro (2001). Suppose $\lambda_{ES} \leq \lambda_{VaR}$. Since $I(\lambda_{ES} \xi) \geq I(\lambda_{VaR} \xi)$, $I(\lambda_{ES} \xi - \frac{1}{\alpha} \lambda_{ES} \mu_{ES}) > I(\lambda_{VaR} \xi)$, we have $X_{ES}(T) \geq X_{VaR}(T)$, *a.s.*, and the inequality is strict when $\xi > \underline{\xi}_{VaR}$, violating the budget constraint. Therefore, $\lambda_{ES} > \lambda_{VaR}$. Moreover, since $I(\lambda_{VaR} \bar{\xi}_{ES}) > I(\lambda_{ES} \bar{\xi}_{ES}) > \underline{X} = I(\lambda_{VaR} \bar{\xi}_{VaR})$, we must have $\underline{\xi}_{ES} < \underline{\xi}_{VaR}$. The rest of the proof follows from direct comparison. \square

Figure 3.1 depicts the optimal terminal wealth of a benchmark agent, a VaR agent and an ES agent, with identical \underline{X} and $0 < \alpha < 1$. Here, $\underline{X}_{VaR} := I(\lambda_{VaR} \bar{\xi}_{VaR})$ and $\underline{X}_{ES} := I(\lambda_{ES} \underline{\xi}_{ES})$.

Similar to the VaR agent, the ES agent endogenously classify the scenarios of the future into three states, but his economic behaviours in these states are quite

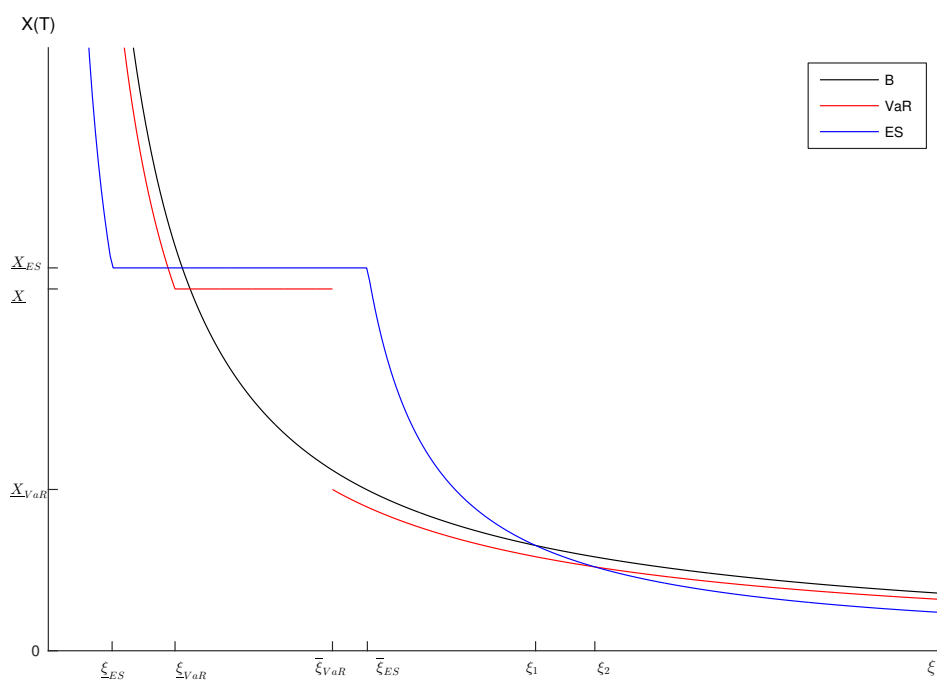


Figure 3.1: Optimal horizon wealth

The figure plots the optimal terminal wealth of a Benchmark agent (black line), a VaR agent (red line) and an ES agent (blue line), with identical \underline{X} and $0 < \alpha < 1$, as functions of the horizon state price density ξ .

different. In the good states $[\xi < \underline{\xi}_{ES}]$, the ES agent behaves like the Benchmark and the VaR agent but he is willing to accept a lower wealth level than the Benchmark and the VaR agent, i.e. $X_{ES}(T) < X_{VaR}(T) < X_B(T)$. In return, he fully insures against the intermediate states $[\underline{\xi}_{ES} \leq \xi < \bar{\xi}_{ES}]$. Moreover, both the scope and the wealth level of the insured states are larger than that of the VaR agent, i.e. $\underline{\xi}_{ES} < \underline{\xi}_{VaR} < \bar{\xi}_{VaR} < \bar{\xi}_{ES}$ and $\underline{X}_{ES} > \underline{X}$. He accepts losses in the bad states $[\xi \geq \bar{\xi}_{ES}]$, because these are the most expensive states to insure against. However, in case of a significant financial market stress $[\xi \geq \bar{\xi}_1]$, the ES agent will incur larger losses than the Benchmark agent, and even larger than the VaR agent in the worst states $[\xi \geq \bar{\xi}_2]$.

This highlights an discouraging feature of the ES-RM: ES only helps to insure against losses in the intermediate states, at the expense of larger extreme losses. Though the scope of the bad states is smaller, the wealth in the extreme bad states is below the benchmark and even the VaR wealth, i.e. the ES-RM reduces the probability of the loss but increases the magnitude of extreme losses. Use of ES is believed to be able to help to ensure a more prudent capture of "tail risk" and capital adequacy during periods of significant financial market stress (BCBS (2016)), but it is shown that the largest losses are more severe than without the ES-RM, defeating the very purpose of ES. One common criticism on VaR is that it fails to take the magnitude of losses into account. Our results reveal that even though ES takes the sizes of losses into account, it is far from enough. ES places equal weights for all levels of losses that exceed a certain threshold. In other words, the marginal riskiness associated with a 1 million dollar loss is equal to that associated with a 2 million dollar loss, or even a 1 billion dollar loss. However, the costs to insure against these losses differ, and it is costlier to insure against a larger loss.

We also note that the ES agent chooses both $\underline{\xi}_{ES}$ and $\bar{\xi}_{ES}$ endogenously, which depend on the agent's preferences and endowment, the market ξ and the ES constraint α and \underline{X} . In contrast, $\bar{\xi}_{VaR}$ solely depends on α and the market. Moreover, the terminal wealth of the VaR agent has a discontinuity at $\bar{\xi}_{VaR}$, whereas the wealth of the ES agent is continuous across all states.

Of course, VaR and ES at the same level α are not comparable. For example, as stated in BCBS (2016), in calculating the ES, a 97.5th percentile, one-tailed confidence level is to be used, while 99th percentile is used in calculating the VaR. For this reason, we will from now on compare the ES agent only to the Benchmark agent.

3.2.3.2 Time- t Wealth and Portfolio Weights

For analytical tractability, we specialize the setting to CRRA preferences,

$$u(X) = \begin{cases} \frac{X^{1-\gamma}}{1-\gamma} & \gamma > 0 \text{ and } \gamma \neq 1, \\ \ln X & \gamma = 1, \end{cases} \quad (3.6)$$

and to lognormal state prices with constant interest and market price of risk.

The following proposition presents explicit expressions for the Benchmark's and the ES agent's optimal wealth and portfolio strategies before the horizon.

Proposition 3.2.4. *Assume u is given by (3.6), and r , μ , and σ are constant.*

For the Benchmark agent:

1. *The optimal time- t wealth is*

$$X_B(t) = \frac{e^{\Gamma(t)}}{(\lambda_B \xi(t))^{\frac{1}{\gamma}}},$$

where $\Gamma(t) := \frac{1-\gamma}{\gamma}(r + \frac{\|\theta\|^2}{2\gamma})(T-t)$ and λ_B is as in Proposition 3.2.1.

2. *The fraction of wealth invested in stocks is*

$$\pi_B(t) = \frac{1}{\gamma}(\sigma')^{-1}\theta, \quad (3.7)$$

For the ES agent:

1. *The optimal time- t wealth is*

$$\begin{aligned} X_{ES}(t) = & \frac{e^{\Gamma(t)}}{(\lambda_{ES}\xi(t))^{\frac{1}{\gamma}}} - \frac{e^{\Gamma(t)}}{(\lambda_{ES}\xi(t))^{\frac{1}{\gamma}}} N(-d_1(\underline{\xi}_{ES})) \\ & + \underline{X}_{ES} e^{-r(T-t)} [N(-d_2(\underline{\xi}_{ES})) - N(-d_2(\bar{\xi}_{ES}))] \\ & + e^{-r(T-t)} G(\bar{\xi}_{ES}, \gamma). \end{aligned}$$

where $\underline{\xi}_{ES}$, λ_{ES} and μ_{ES} are as in Proposition 3.2.1, and

$$\begin{aligned} d_2(x) &:= \frac{\ln \frac{x}{\xi(t)} + (r - \frac{\|\theta\|^2}{2})(T-t)}{\|\theta\| \sqrt{T-t}}, \\ d_1(x) &:= d_2(x) + \frac{1}{\gamma} \|\theta\| \sqrt{T-t}, \\ \bar{\xi}_{ES} &:= \underline{\xi}_{ES} + \frac{1}{\alpha} \mu_{ES}, \\ G(x, \gamma) &:= \int_{d_2(x)}^{+\infty} \phi(z) \frac{1}{(\lambda_{ES}\xi(t) e^{\|\theta\| \sqrt{T-t} z - (r - \frac{\|\theta\|^2}{2})(T-t)} - \frac{1}{\alpha} \lambda_{ES} \mu_{ES})^{\frac{1}{\gamma}}} dz, \end{aligned}$$

$N(\cdot)$ is the standard normal distribution function.

2. The fraction of wealth invested in stocks is

$$\pi_{ES}(t) = q_{ES}(t)\pi_B(t),$$

where

$$q_{ES}(t) := 1 - \frac{\underline{X}_{ES}e^{-r(T-t)}(N(-d_2(\underline{\xi}_{ES})) - N(-d_2(\bar{\xi}_{ES})))}{X_{ES}(t)} + \frac{1}{\alpha}\lambda_{ES}\mu_{ES}\frac{e^{-r(T-t)}}{X_{ES}(t)}G(\bar{\xi}_{ES}, \frac{\gamma}{1+\gamma}). \quad (3.8)$$

3. The exposure to risky assets relative to the benchmark is bounded below: $q_{ES}(t) \geq 0$, and

$$\lim_{\xi(t) \rightarrow 0} q_{ES}(t) = \lim_{\xi(t) \rightarrow \infty} q_{ES}(t) = 1.$$

4. When the ES constraint is binding, $q_{ES}(t) > 1$ if $\xi(t) > \xi_{ES}(t)$ for some deterministic $\xi_{ES}(t)$.

Proof of Proposition 3.2.4. We only show the ES agent part.

1. It is well-known that $\xi(t)X_{ES}(t)$ is a martingale

$$\xi(t)X_{ES}(t) = \mathbb{E}[\xi(T)X_{ES}(T)|\mathcal{F}_t].$$

When r , σ and θ are constant, conditional on \mathcal{F}_t , $\ln \xi(T)$ is normally distributed with mean $\ln \xi(t) - (r + \frac{\|\theta\|^2}{2})(T-t)$ and variance $\|\theta\|^2(T-t)$. Evaluating the conditional expectation gives $X_{ES}(t)$.

2. Applying Ito's lemma to the expression of $X_{ES}(t)$ and comparing the coefficient of the $dW(t)$ term with that of (2.4), we obtain the expression of $\pi_{ES}(t)$. Dividing it by $\pi_B(t)$ given by (3.7) yields

$$q_{ES}(t) = \frac{1}{X_{ES}(t)} \left[\frac{e^{\Gamma(t)}}{(\lambda_{ES}\xi(t))^{\frac{1}{\gamma}}} - \frac{e^{\Gamma(t)}}{(\lambda_{ES}\xi(t))^{\frac{1}{\gamma}}} N(-d_1(\underline{\xi}_{ES})) + e^{-r(T-t)} G(\bar{\xi}_{ES}, \frac{\gamma}{1+\gamma}) \right]. \quad (3.9)$$

Rearranging (3.9) gives (3.8).

3. (3.9) reveals that it is non-negative. The limits are straightforward to verify.

4. Define

$$F(\xi(t)) := -\underline{X}_{ES}(N(-d_2(\underline{\xi}_{ES})) - N(-d_2(\bar{\xi}_{ES}))) + \frac{1}{\alpha}\lambda_{ES}\mu_{ES}G(\bar{\xi}_{ES}, \frac{\gamma}{1+\gamma}),$$

it suffices to show $F'(\xi(t)) < 0$ for $\xi(t)$ large enough. We have

$$\begin{aligned}
F'(\xi(t)) &= -\underline{X}_{ES}(\phi(d_2(\underline{\xi}_{ES})) - \phi(d_2(\bar{\xi}_{ES}))) \frac{1}{\|\theta\| \sqrt{T-t} \xi(t)} \\
&\quad + \lambda_{ES}(\bar{\xi}_{ES} - \underline{\xi}_{ES}) \frac{\phi(d_2(\bar{\xi}_{ES}))}{(\lambda_{ES} \underline{\xi}_{ES})^{1+\frac{1}{\gamma}}} \frac{1}{\|\theta\| \sqrt{T-t} \xi(t)} \\
&\quad - \lambda_{ES}(\bar{\xi}_{ES} - \underline{\xi}_{ES}) \int_{d_2(\bar{\xi}_{ES})}^{+\infty} \phi(z) \\
&\quad \cdot \frac{(1 + \frac{1}{\gamma}) \lambda_{ES} e^{\|\theta\| \sqrt{T-t} z - (r - \frac{\|\theta\|^2}{2})(T-t)}}{(\lambda_{ES} \xi(t) e^{\|\theta\| \sqrt{T-t} z - (r - \frac{\|\theta\|^2}{2})(T-t)} - \frac{1}{\alpha} \lambda_{ES} \mu_{ES})^{2+\frac{1}{\gamma}}} dz \\
&= \frac{\phi(d_2(\bar{\xi}_{ES}))}{(\lambda_{ES} \underline{\xi}_{ES})^{\frac{1}{\gamma}} \|\theta\| \sqrt{T-t} \xi(t)} g(d_2(\bar{\xi}_{ES})),
\end{aligned} \tag{3.10}$$

where

$$\begin{aligned}
g(x) &:= a - e^{-\frac{(\ln a)^2}{2\|\theta\|^2(T-t)}} a^{\frac{x}{\|\theta\| \sqrt{T-t}}} - \frac{(1 + \frac{1}{\gamma}) (\lambda_{ES} \underline{\xi}_{ES})^{2+\frac{1}{\gamma}} \|\theta\| \sqrt{T-t} a(a-1)}{\phi(x)} \\
&\quad \cdot \int_x^{+\infty} \phi(z) \frac{e^{\|\theta\| \sqrt{T-t}(z-x)}}{(\lambda_{ES} \bar{\xi}_{ES} e^{\|\theta\| \sqrt{T-t}(z-x)} - \frac{1}{\alpha} \lambda_{ES} \mu_{ES})^{2+\frac{1}{\gamma}}} dz, \\
a &:= \frac{\bar{\xi}_{ES}}{\underline{\xi}_{ES}},
\end{aligned}$$

and $\phi(\cdot)$ is the standard normal probability density function. It is a simple exercise to show $\lim_{x \rightarrow -\infty} g(x) = -\infty$, thereby completing the proof. \square

We find that when the state price density process is sufficiently low, the ES agent has to invest more in stocks to meet the ES constraint. The ES constraint increases the risk exposure (represented by the fraction of wealth invested in risky assets) in the extremely bad states, which is consistent with our previous findings.

Figure 3.2 depicts the optimal time-t wealth of a benchmark agent and an ES agent. The time-t wealth of the ES agent is similar to its time-T behaviors. The ES agent behaves like a Benchmark agent in the good and bad states, whereas in the intermediate states, his wealth flattens as time approaches the horizon. In the good and extremely bad states, his wealth is below the Benchmark agent's wealth.

Figure 3.3 depicts the ES agent's time-t relative risk exposure $q_{ES}(t)$, which is S-shaped. We may classify it into five states. In the two extremes of $\xi(t)$, the B behavior dominates. In the good states, his relative risk exposure is low and increases

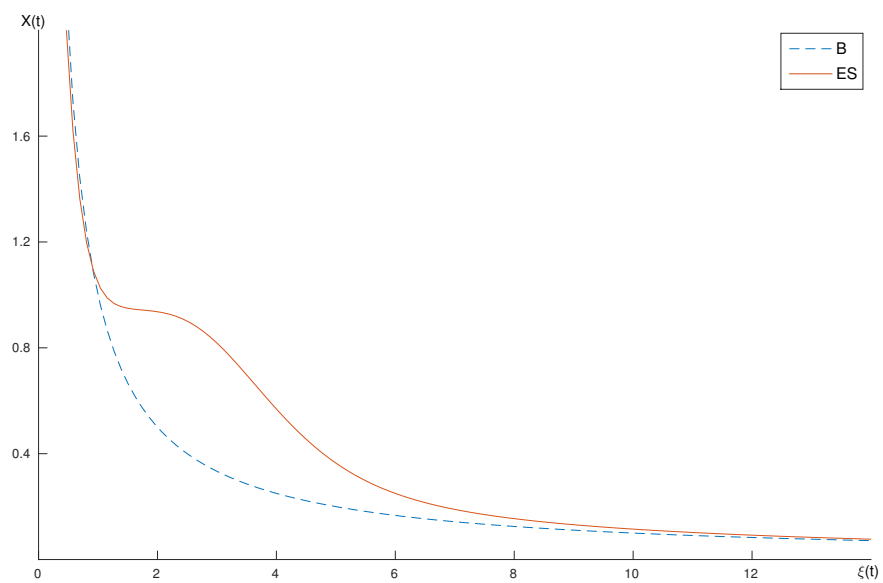


Figure 3.2: Optimal time-t wealth

The figure plots the optimal time-t wealth of a Benchmark agent (blue, dotted line) and an ES agent (orange line), as functions of the time-t state price density $\xi(t)$. We assume CRRA preferences and lognormal state price density. The parameters are $T = 1, t = 0.5, r = 0.05, \theta = 0.4, X(0) = 1, \underline{X} = 0.8, \alpha = 0.05, \gamma = 1$.

as the market transits into bad states. His risk exposure is high in the bad states but decreases as the market continues to deteriorate.

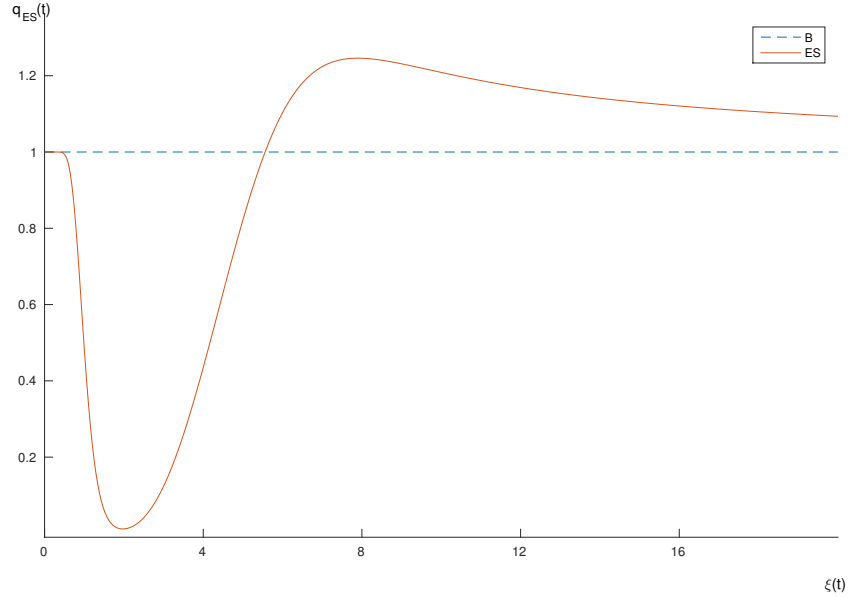


Figure 3.3: Time-t relative risk exposure

The figure plots the ES agent's time-t exposure to risky assets relative to the Benchmark agent, as a function of the time-t state price density $\xi(t)$. We assume CRRA preferences and lognormal state price density. The parameters are $T = 1, t = 0.5, r = 0.05, \theta = 0.4, X(0) = 1, \underline{X} = 0.8, \alpha = 0.05, \gamma = 1$.

3.3 General Equilibrium Analysis

3.3.1 Equilibrium in a Production Economy

In Section 3.2.1, we illustrated that, the ES agent optimally shifts wealth from good and bad states to intermediate states to meet the requirement on ES. This motivates us to consider a production economy in which aggregate consumption/ wealth can be postponed or shifted and aggregate nonzero holdings in a riskless investment are allowed.

We consider a continuous-time, finite horizon $[0, T]$ production economy similar to that in Basak (2002); Basak and Shapiro (2005), which is a variation on the Cox et al. (1985) economy. In this economy, the investment opportunities available are constant-returns-to-scale production technologies, using the single consumption good as their only input and producing the consumption good as output. The production

technologies have perfectly elastic supplies, and net returns given by (2.2) and (2.3), where the (exogenously specified) parameters r , μ , and σ are assumed constant. The first production technology is riskless and the others risky. The economy is populated by two types of agents whose initial wealth is exogenously specified as units of consumption good:

1. The Benchmark agent, who solves

$$\begin{aligned} & \max_{X(T)} \mathbb{E}[u(X_B(T))] \\ & \text{subject to } \mathbb{E}[\xi X_B(T)] \leq X_B(0); \end{aligned}$$

2. the ES agent, who solves

$$\begin{aligned} & \max_{X_{ES}(T)} \mathbb{E}[u(X_{ES}(T))] \\ & \text{subject to } \mathbb{E}[\xi X_{ES}(T)] \leq X_{ES}(0), \\ & \quad ES_\alpha(X_{ES}(T)) \leq -\underline{X}, \end{aligned}$$

where u is given by (3.6) and \underline{X} is a given positive constant.

We refer to this economy in which the first agent is the Benchmark agent and the second agent is the ES agent as the ES economy, and the economy in which both agents are Benchmark agents as the Benchmark economy.

Equilibrium in this production economy requires both agents to act optimally, and for all wealth to be invested in the production technologies. The optimal trading strategies in the partial equilibrium setting (Section 3.2.3) persist. Our goal is to compare equilibrium in the presence of the ES constraint with equilibrium in the benchmark economy without the constraint. In particular, we are interested in the impact of the ES constraint on the market value dynamics. The price of the market portfolio, X_M , is defined as the aggregate wealth invested in the production technologies, which equals both agents' net worth:

$$X_M(t) := X_B(t) + X_{ES}(t).$$

The equilibrium market-price dynamics can be represented by

$$dX_M(t) = X_M(t)[\mu_M(t)dt + \sum_{j=1}^n \sigma_{M,j}(t)dW^j(t)],$$

where $\mu_M(t)$ is the market drift and $\|\sigma_M\| := \sqrt{\sum_{j=1}^n (\sigma_{M,j}(t))^2}$ is the market volatility.

The following proposition presents the equilibrium market price, volatility, and risk premium and contrasts those with the Benchmark economy.

Proposition 3.3.1. *The equilibrium market price, volatility, and risk premium in the Benchmark economy are given by*

$$\begin{aligned} X_M^B(t) &= \frac{X_M^B(0)}{(\xi(t))^{\frac{1}{\gamma}}} e^{\frac{\gamma-1}{\gamma}(r+\frac{\|\theta\|^2}{2\gamma})t}, \\ \|\sigma_M^B(t)\| &= \frac{1}{\gamma} \|\theta\|, \\ \mu_M^B(t) - r &= \frac{1}{\gamma} \|\theta\|^2. \end{aligned}$$

The equilibrium market price, volatility, and risk premium in the ES economy are given by

$$\begin{aligned} X_M^{ES}(t) &= \frac{e^{\Gamma(t)}}{(\lambda_B \xi(t))^{\frac{1}{\gamma}}} + \frac{e^{\Gamma(t)}}{(\lambda_{ES} \xi(t))^{\frac{1}{\gamma}}} - \frac{e^{\Gamma(t)}}{(\lambda_{ES} \xi(t))^{\frac{1}{\gamma}}} N(-d_1(\xi_{ES})) \\ &\quad + \underline{X}_{ES} e^{-r(T-t)} [N(-d_2(\xi_{ES})) - N(-d_2(\bar{\xi}_{ES}))], \\ \|\sigma_M^{ES}(t)\| &= \frac{1}{\gamma} \left(1 - (1 - q_{ES}(t)) \frac{X_{ES}(t)}{X_M^{ES}(t)}\right) \|\theta\|, \\ \mu_M^{ES}(t) - r &= \frac{1}{\gamma} \left(1 - (1 - q_{ES}(t)) \frac{X_{ES}(t)}{X_M^{ES}(t)}\right) \|\theta\|^2. \end{aligned}$$

Thus, $\|\sigma_M^{ES}(t)\| > \|\sigma_M^B(t)\|$ and $\mu_M^{ES}(t) - r > \mu_M^B(t) - r$, if $\xi(t) > \xi_{ES}(t)$.

Proof of Proposition 3.3.1. Summing over the two Benchmark agents' time- t wealth in Proposition 3.2.4, and expressing the Lagrange multipliers in terms of the initial wealth gives the Benchmark economy's time- t market value. Summing over the Benchmark agent's and the ES agent's time- t wealth gives the ES economy's time- t market value. In view of (2.4) and Proposition 3.2.4, we obtain the expressions for the equilibrium market volatility and risk premiums in both economies. The last claim is due to the property of $q_{ES}(t)$. \square

Proposition 3.3.1 reveals that the equilibrium market volatility and risk premium are increased by the presence of the ES constraint, in the bad states of the world ($\xi(t) > \xi_{ES}(t)$). This is because, as reflected in Proposition 3.2.4, in the bad states, the ES agent has a higher demand for risky investment opportunities than the Benchmark agent. In these scenarios, the ES economy's aggregate investment in the risky technologies is higher than that in the Benchmark economy, leading to higher market volatility and risk premium. This may be a source of concern for policy-makers: ES increases rather than decreases the market volatility in the most adverse states of the world.

3.3.2 Equilibrium in a Pure-Exchange Economy

In the production economy with exogenous investment technologies, we show that the market volatility and risk premium is increased by the presence of ES. We re-evaluate this in an economy in which all quantities except the aggregate consumption process are determined endogenously. We follow Basak (1995); Basak and Shapiro (2001) to develop a pure-exchange general equilibrium model featuring ES risk managers.

We assume that the economy is populated by two types of agents, the Benchmark and the ES agent, who derive utility from intertemporal (continuous) consumption over their lifetime $[0, T']$. As opposed to the Benchmark agent, the ES agent is subject to the additional ES constraint as in (3.3) over time- T wealth, where $T < T'$. We refer to the economy in which the first agent is the Benchmark agent and the second agent is the ES agent as the ES economy, and the economy in which both agents are Benchmark agents as the Benchmark economy. For simplicity, we assume there is a single consumption good serving as the numeraire and $n = 1$, i.e., all the uncertainty is represented by the one-dimensional Brownian motion W .

There are two investment opportunities: one instantaneously riskless and the other risky. The riskless investment is a bond in zero net supply; the risky investment is a stock in constant net supply of 1 and paying out a dividend stream at rate δ . We assume the (exogenously given) dividend process to follow a geometric Brownian motion:

$$d\delta(t) = \delta(t)[\mu_\delta dt + \sigma_\delta dW(t)],$$

with constant $\mu_\delta, \sigma_\delta$ and $\delta(0) > 0$.

In light of Basak (1995); Basak and Shapiro (2001), we anticipate that the constraint applied at the ES horizon T may result in jumps in the equilibrium security and state prices. The price processes of the riskless asset $S_0(t)$ and the risky asset $S(t)$ are subject to the following equations:

$$\begin{aligned} dS_0(t) &= S_0(t)[r(t)dt + \eta dA(t)], \\ dS(t) + \delta(t)dt &= S(t)[\mu(t)dt + \sigma(t)dW(t) + \eta dA(t)], \end{aligned}$$

where r, b and σ are endogenous and are determined in equilibrium, $\eta dA(t)$ is the changes of security prices at time T . $A(t)$ is a right-continuous step function defined by $A(t) := \mathbf{1}_{t \geq T}$, and the jump size η is an $\{\mathcal{F}_T\}$ -measurable random variable. To prevent arbitrage opportunities, the jump size q in all security prices must be equal, see Basak (1995); Basak and Shapiro (2001) for a detailed discussion. The state price

density process ξ is given by

$$\xi(t) := \exp \left\{ - \int_0^t [r(s) + \frac{1}{2} \|\theta(s)\|^2] ds - \int_0^t \theta(s)^\top dW(s) - \eta A(t) \right\}. \quad (3.11)$$

where θ is the unique bounded, $\{\mathcal{F}_t\}$ -progressively measurable market price of risk, given by $\theta(t) := \frac{\mu(t) - r(t)}{\sigma(t)}$.

At time 0, each agent is endowed with $e_i, i = B, ES$, units (exogenous) of the risky asset, that is, an initial wealth of $X_i(0) = S(0)e_i$ and we assume $e_B + e_{ES} = 1$. Assume that the trading of shares takes place continuously in a self-financing fashion and there are no transaction costs. Then the wealth process of agent i , $X_i(\cdot)$ satisfies

$$dX_i(t) = X_i(t)r(t)dt - c_i(t)dt + X_i(t)[\mu(t) - r(t)]\pi_i(t)dt + X_i(t)\pi_i(t)\sigma(t)dW(t) + X_i(t)\eta dA(t),$$

where $\pi_i(t)$ and $c_i(t)$ denote the fractions of the agent's wealth in the risky asset and consumption at time t , respectively. We require $\pi_i(t)$ and $c_i(t)$ to be \mathcal{F}_t -progressively measurable with

$$\begin{aligned} \int_0^T |X_i(t)\sigma\pi_i(t)|^2 dt &< \infty, a.s., \\ \int_0^T |X_i(t)[\mu(t) - r(t)]\pi_i(t)|^2 dt &< \infty, a.s., \\ \int_0^T |c_i(t)|^2 dt &< \infty, a.s., \end{aligned}$$

and $X_i(T') \geq 0, a.s..$

Given the dynamics of the state price density (3.11), we have the following result.

Lemma 3.3.1.

$$X_i(t) = \frac{1}{\xi(t)} \mathbb{E} \left[\int_t^{T'} \xi(s)c_i(s)ds + \xi(T')X(T') \mid \mathcal{F}_t \right], t \in [0, T']. \quad (3.12)$$

For tractability, we assume that all agents have logarithmic utility. The Benchmark agent solves the following problem:

$$\begin{aligned} \max_{c_B} \mathbb{E} \left[\int_0^{T'} \ln(c_B(t)) dt \right] \\ \text{subject to } \mathbb{E} \left[\int_0^{T'} \xi(t)c_B(t) dt \right] \leq X_B(0). \end{aligned} \quad (3.13)$$

The ES agent solves the following problem:

$$\begin{aligned}
& \max_{c_{ES}, X_{ES}(T-)} \mathbb{E} \left[\int_0^{T'} \ln(c_{ES}(t)) dt \right] \\
& \text{subject to } \mathbb{E} \left[\int_0^T \xi(t) c_{ES}(t) dt + \xi(T-) X_{ES}(T-) \right] \leq X_{ES}(0), \\
& \mathbb{E} \left[\int_T^{T'} \xi(t) c_{ES}(t) dt \right] \leq \xi(T-) X_{ES}(T-), \text{ a.s.}, \\
& ES_\alpha(X_{ES}(T-)) \leq -\underline{X}.
\end{aligned} \tag{3.14}$$

The following proposition characterizes each agent's optimal policy, under Assumptions 2.3.1 and 2.3.2². We will verify later that in the equilibrium the state price density process indeed satisfies these assumptions.

Proposition 3.3.2. *We have the following assertions:*

1. *The optimal consumption policies and time- T wealth of the Benchmark agent (3.13) is*

$$\begin{aligned}
c_B(T) &= \frac{1}{\lambda_B \xi(t)}, t \in [0, T'], \\
X_B(T-) &= \frac{T' - T}{\lambda_B \xi(T-)}.
\end{aligned}$$

where $\lambda_B = \frac{T'}{X_B(0)}$.

2. *The optimal consumption policies and time- T wealth of the ES agent (3.14) is*

$$\begin{aligned}
c_{ES}(t) &= \begin{cases} \frac{1}{\lambda_{ES,1} \xi(t)} & t \in [0, T), \\ \frac{1}{\lambda_{ES,2} \xi(t)} & t \in [T, T'], \end{cases} \\
X_{ES}(T-) &= \begin{cases} \frac{T' - T}{\lambda_{ES,1} \xi(T-)} & \xi(T-) < \underline{\xi}_{ES}, \\ \frac{T' - T}{\lambda_{ES,1} \underline{\xi}_{ES}} & \underline{\xi}_{ES} \leq \xi(T-) < \bar{\xi}_{ES}, \\ \frac{T' - T}{\lambda_{ES,1} \xi(T-) - \frac{1}{\alpha} \lambda_{ES,1} \mu_{ES}} & \bar{\xi}_{ES} \leq \xi(T-), \end{cases}
\end{aligned}$$

where $\underline{\xi}_{ES} > 0$ solves $h_{\mu_{ES}}(1 - F_\xi(\underline{\xi}_{ES})) = 0$, $\bar{\xi}_{ES} = \underline{\xi}_{ES} + \frac{1}{\alpha} \mu_{ES}$, $\lambda_{ES,2} = \frac{T' - T}{\xi(T-) X_{ES}(T-)}$, and $\lambda_{ES,1} > 0, \mu_{ES} \geq 0$ solve

$$\begin{aligned}
\frac{T}{\lambda_{ES,1}} + \mathbb{E}[\xi(T-) X_{ES}(T-)] &= X_B(0), \\
ES_\alpha(X_{ES}(T-)) &= -\underline{X}.
\end{aligned}$$

$h_\mu(\cdot)$ is defined by (3.5). Here, F_ξ is the quantile function of $\xi(T-)$.

²From now on we denote by F_ξ the quantile function of $\xi(T-)$

Proof of Proposition 3.3.3. The Benchmark agent's optimal consumption and time- T wealth are standard in the literature. For the ES agent, we first consider the following problem:

$$\begin{aligned} & \max_{c_{ES}(t), t \in [T, T']} \mathbb{E} \left[\int_T^{T'} \ln(c_{ES}(t)) dt \middle| \mathcal{F}_T \right] \\ & \text{subject to } \mathbb{E} \left[\int_T^{T'} \xi(t) c_{ES}(t) dt \middle| \mathcal{F}_T \right] \leq \xi(T-) X_{ES}(T-), a.s.. \end{aligned}$$

The optimal consumption is

$$c_{ES}(t) = \frac{1}{\lambda_{ES,2} \xi(t)}, t \in [T, T'],$$

where

$$\lambda_{ES,2} = \frac{T' - T}{\xi(T-) X_{ES}(T-)},$$

and the optimal value is

$$(T' - T) \ln(X_{ES}(T-)) + \mathbb{E} \left[\int_T^{T'} \ln\left(\frac{\xi(T-)}{(T' - T)\xi(t)}\right) dt \middle| \mathcal{F}_T \right].$$

Consequently, we can consider the following problem:

$$\begin{aligned} & \max_{c_{ES}(t), t \in [0, T], X_{ES}(T-)} \mathbb{E} \left[\int_0^T \ln(c_{ES}(t)) dt + (T' - T) \ln(X_{ES}(T-)) \right] \\ & \text{subject to } \mathbb{E} \left[\int_0^T \xi(t) c_{ES}(t) dt + \xi(T-) X_{ES}(T-) \right] \leq X_{ES}(0), \\ & \quad ES_\alpha(X_{ES}(T-)) \leq -\underline{X}. \end{aligned}$$

The rest of the proof is a straightforward extension of Proposition 3.2.1 \square

Note that when $\mu_{ES} = 0$ we recover the Benchmark agent. We next define and then characterize the equilibrium in our setting.

Definition 3.3.1. An equilibrium is a collection of (r, b, σ) and optimal $(c_B, c_{ES}, \pi_B, \pi_{ES})$, such that the good, stock, and bond markets clear, that is, $\forall t \in [0, T']$,

$$\begin{aligned} c_B(t) + c_{ES}(t) &= \delta(t), \\ \pi_B(t) X_B(t) + \pi_{ES}(t) X_{ES}(t) &= S(t), \\ X_B(t) + X_{ES}(t) &= S(t). \end{aligned}$$

The following proposition characterizes the equilibrium state price density, interest rate and market price of risk explicitly.

Proposition 3.3.3. *We have the following assertions:*

1. *The equilibrium state price density is given by*

$$\xi(t) = \begin{cases} \left(\frac{1}{\lambda_B} + \frac{1}{\lambda_{ES,1}}\right) \frac{1}{\delta(t)} & t \in [0, T), \\ \left(\frac{1}{\lambda_B} + \frac{1}{\lambda_{ES,2}}\right) \frac{1}{\delta(t)} & t \in [T, T'], \end{cases} \quad (3.15)$$

where λ_B , $\lambda_{ES,1}$, and $\lambda_{ES,2}$ are given in Proposition 3.3.2 with (3.15) substituted in. (3.15) satisfies Assumptions 2.3.1 and 2.3.2.

2. *The equilibrium interest rate and market price of risk are constants, at all $t \in [0, T']$, given by*

$$\begin{aligned} r &= \mu_\delta - \|\sigma_\delta\|^2, \\ \theta &= \|\sigma_\delta\|. \end{aligned}$$

3. *The jump size is*

$$\eta = \ln\left(\frac{1}{\lambda_B} + \frac{1}{\lambda_{ES,2}}\right) - \ln\left(\frac{1}{\lambda_B} + \frac{1}{\lambda_{ES,1}}\right).$$

Proof of Proposition 3.3.3. Clearing the consumption good market gives (3.15). The proof that clearing the good market implies all other markets are cleared appears in Basak (1995). Moreover, it is straightforward to verify (3.15) satisfies Assumption 2.3.1 and 2.3.2. r and θ are determined by applying Ito's lemma to (3.15). \square

The price of the equity market portfolio, $X_M(t)$, is defined as the aggregate optimally invested wealth in the risky asset, i.e.,

$$X_M(t) := \pi_B(t)X_B(t) + \pi_{ES}(t)X_{ES}(t) = X_B(t) + X_{ES}(t) = S(t).$$

The equity market value is always equal to the stock price since there is only one stock in the market ³.

We represent the equilibrium market dynamics as

$$dX_M(t) + \delta(t)dt = X_M(t)[\mu_M(t)dt + \sigma_M(t)dW(t)],$$

where μ_M is the equity market drift, $\|\sigma_M\|$ is the equity market volatility and $\mu_M - r$ is the equity market risk premium.

The following proposition presents these quantities in equilibrium and contrasts them with the Benchmark economy.

³If there are multiple stocks in the market, the equity market value will be the sum of values of all stocks

Proposition 3.3.4. *We have the following assertions:*

1. *The equilibrium market price, volatility, and risk premium in the Benchmark economy are given by*

$$\begin{aligned} X_M^B(t) &= (T' - t)\delta(t), \\ \|\sigma_M^B(t)\| &= \|\sigma_\delta\|, \\ \mu_M^B(t) - r &= \|\sigma_\delta\|^2. \end{aligned} \tag{3.16}$$

2. *Before the ES horizon, the equilibrium market price, volatility, and risk premium in the ES economy are given by*

$$\begin{aligned} X_M^{ES}(t) &= (T' - t)\delta(t) - \frac{\lambda_B(T' - T)}{\lambda_B + \lambda_{ES,1}}\delta(t)N(-\hat{d}_1(\bar{\delta})) \\ &\quad + \frac{\lambda_B(T' - T)}{\lambda_B + \lambda_{ES}}\bar{\delta}e^{-(\mu_\delta - \|\sigma_\delta\|^2)(T-t)}[N(-\hat{d}_2(\bar{\delta})) - N(-\hat{d}_2(\underline{\delta}))] \\ &\quad + (T' - T)e^{-(\mu_\delta - \|\sigma_\delta\|^2)(T-t)}\hat{G}(\underline{\delta}, 1), \end{aligned} \tag{3.17}$$

$$\begin{aligned} \|\sigma_M^{ES}(t)\| &= \hat{q}(t)\|\sigma_\delta\|, \\ \mu_M^{ES}(t) - r &= \hat{q}(t)\|\sigma_\delta\|^2, \end{aligned}$$

where

$$\begin{aligned} \bar{\delta} &:= \frac{\frac{1}{\lambda_B} + \frac{1}{\lambda_{ES,1}}}{\xi_{ES}}, \\ \underline{\delta} &:= \frac{\frac{1}{\lambda_B} + \frac{1}{\lambda_{ES,1}}}{\bar{\xi}_{ES}}, \\ \hat{d}_1(x) &:= \frac{\ln \frac{\delta(t)}{x} + (\mu_\delta - \frac{1}{2}\|\sigma_\delta\|^2)(T-t)}{\|\sigma_\delta\|\sqrt{T-t}}, \\ \hat{d}_2(x) &:= \hat{d}_1(x) - \|\sigma_\delta\|\sqrt{T-t}, \\ \hat{G}(x, y) &:= \int_{\hat{d}_2(x)}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \frac{1}{\left(\frac{\lambda_B + \lambda_{ES,1}}{\lambda_B \delta(t)} e^{\|\sigma_\delta\|\sqrt{T-t}z - (\mu_\delta - \frac{3}{2}\|\sigma_\delta\|^2)(T-t)} - \frac{1}{\alpha}\lambda_{ES,1}\mu_{ES}\right)^{\lambda_2}} dz, \\ \hat{q}(t) &:= 1 - \frac{\lambda_B(T' - T)}{(\lambda_B + \lambda_{ES,1})X_M^{ES}(t)}\bar{\delta}e^{-(\mu_\delta - \|\sigma_\delta\|^2)(T-t)}[N(-\hat{d}_2(\bar{\delta})) - N(-\hat{d}_2(\underline{\delta}))] \\ &\quad + \frac{1}{\alpha}\lambda_{ES,1}\mu_{ES}\frac{(T' - T)e^{-(\mu_\delta - \|\sigma_\delta\|^2)(T-t)}}{X_M^{ES}(t)}\hat{G}(\underline{\delta}, 2). \end{aligned}$$

After the ES horizon, market prices, volatility, and risk premiums in both economies are identical.

3. For $t \in [0, T)$, $X_M^{ES}(t) > X_M^B(t)$ and $\frac{X_M^{ES}(t)}{\delta(t)} > \frac{X_M^B(t)}{\delta(t)}$.

4. For $t \in [0, T)$, $\|\sigma_M^{ES}(t)\| > \|\sigma_M^B(t)\|$ and $\mu_M^{ES}(t) > \mu_M^B(t)$ if $\delta(t) < \delta^*(t)$, for some deterministic $\delta^*(t)$.

Proof of Proposition 3.3.4.

1. We only prove the ES part, since the Benchmark economy is a special case of the ES economy when the ES constraint is not binding, i.e., $\mu_{ES} = 0$.
2. By Lemma 3.3.1, we have

$$\begin{aligned} X_M(t) &= X_B(t) + X_{ES}(t) \\ &= \frac{1}{\xi(t)} \mathbb{E} \left[\int_t^T \xi(s) (c_B(s) + c_{ES}(s)) ds \middle| \mathcal{F}_t \right] + \\ &\quad \frac{1}{\xi(t)} \mathbb{E} [\xi(T-)(X_B(T-) + X_{ES}(T-)) \middle| \mathcal{F}_t], t \in [0, T). \end{aligned}$$

By Proposition 3.2.4, 3.3.2 and 3.3.3, we arrive at (3.17). Applying Ito's lemma to $X_M(t)$ yields the expression for $\|\sigma_M^{ES}(t)\|$ and $\mu_M^{ES}(t)$.

3. Note that when $X_{ES}(T-) \geq \frac{T'-T}{\lambda_{ES,1}\xi_{ES}}$ then $X_B(T-) + X_{ES}(T-) = \delta(T-)$, and when $X_{ES}(T-) < \frac{T'-T}{\lambda_{ES,1}\xi_{ES}}$ then $X_B(T-) + X_{ES}(T-) > \delta(T-)$. Hence, $X_M^{ES}(T-) \geq X_M^B(T-)$ and the inequality is strict with a non-zero probability, proving the result.
4. The proof is as of Proposition 3.2.4.

□

The ES economy is qualitatively the same as the VaR economy in Basak and Shapiro (2001). We find that the prehorizon market price in the ES economy is always higher than in the Benchmark economy. This is because the ES agent values the dividend at T more than the prehorizon consumption to meet the ES constraint. The equity market value is then pushed up as it is the claim against future dividends. The price-dividend ratio is thus increased. When the market state is bad, that is, the output $\delta(t)$ is low and the state price density $\xi(t)$ is high, the market volatility in the ES economy is amplified compared to the Benchmark economy. The interest rate and the market price of risk are constant in the equilibrium. When the output is low, the ES agent will need to invest more in the risky asset. The market volatility is then increased and the market risk premium must also increase accordingly to keep the market price of risk constant. This may be a source of concern for policy-makers: ES increases the market volatility during periods of significant financial market stress.

Figure 3.4 depicts the time- t market prices of the Benchmark economy and the ES economy. We may classify it into three states. In the two extremes of $\delta(t)$, the market price of the ES economy evolves like the B economy, whereas in the intermediate states it is significantly higher.

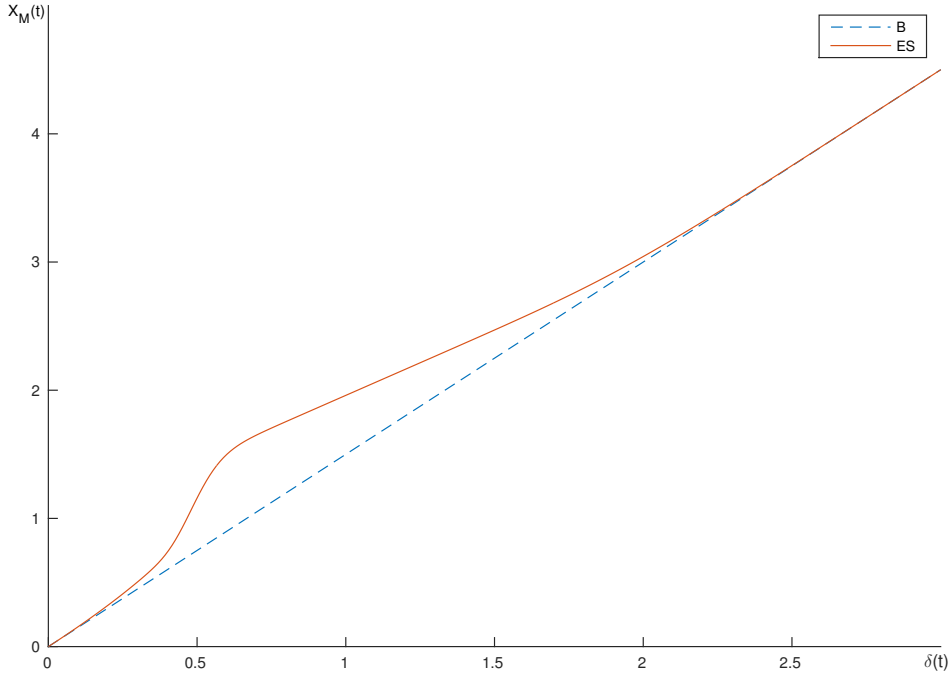


Figure 3.4: Time- t market price

The figure plots the time- t market prices of the Benchmark economy and the ES economy, as functions of the time- t dividend $\delta(t)$. The parameters are $\mu_\delta = 0.1$, $\sigma_\delta = 0.2$, $\lambda_B = 1$, $\lambda_{ES,1} = 1.07$, $\mu_{ES} = 0.02$.

Figure 3.5 depicts the ES economy's time- t market volatility and risk premium relative to the Benchmark economy, given by $\hat{q}_{ES}(t) = \frac{\|\sigma_M^{ES}(t)\|}{\|\sigma_M^B(t)\|} = \frac{\mu_M^{ES}(t) - r}{\mu_M^B(t) - r}$. In the two extremes of $\delta(t)$, the ES economy is similar to the Benchmark economy. As the output increases, the ES market volatility and risk premium relative to the Benchmark economy first increase from one, then decrease, and finally increase to one. When the output is high, these quantities in the ES economy is significantly lower than in the Benchmark economy. However, the ES market is more volatile and the risk premium is higher during periods of market stress, that is, when the output is low.

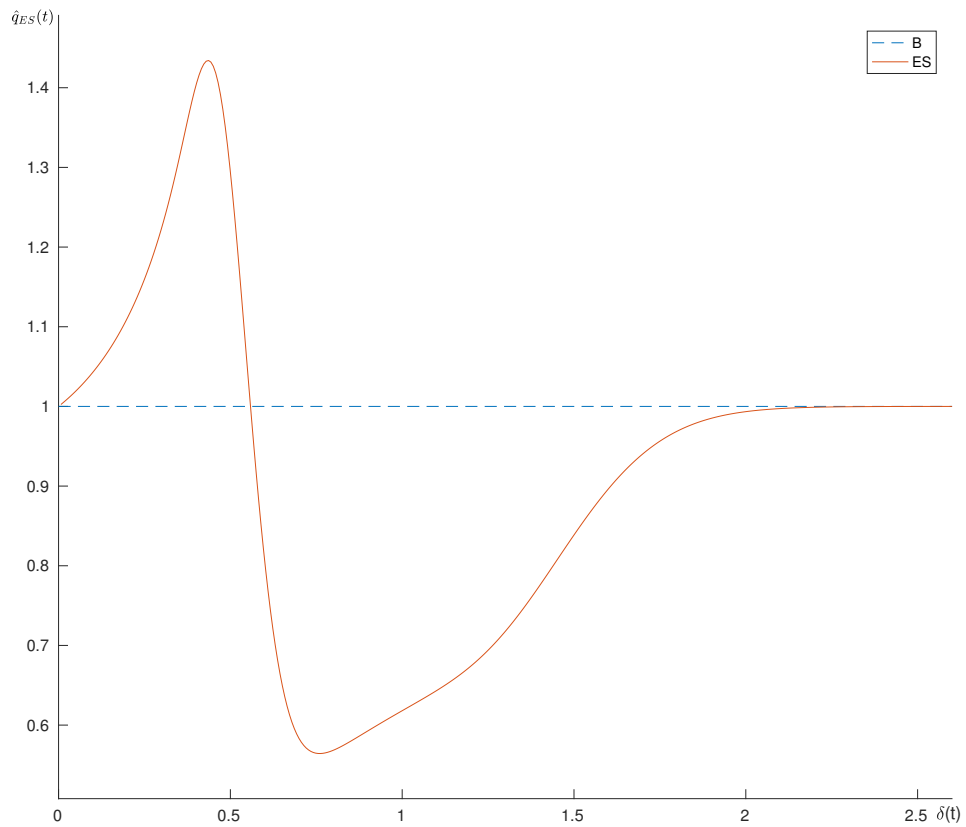


Figure 3.5: Time-t relative risk exposure

The figure plots the ES economy's time-t market volatility and risk premium relative to the Benchmark economy, as a function of the time-t dividend $\delta(t)$. The parameters are $T' = 2, T = 1, t = 0.5, \mu_\delta = 0.1, \sigma_\delta = 0.2, \lambda_B = 1, \lambda_{ES,1} = 1.07, \mu_{ES} = 0.02$.

3.4 Concluding Remarks

Finally, let us comment on the setting of our models in Chapter 2 and Chapter 3. Following Basak and Shapiro (2001); Basak et al. (2006), we measure the investment risk by applying a risk measure on the terminal wealth. The risk is evaluated at the beginning of the investment and the agent needs to commit himself to comply with the constraint in all future dates. There are papers, e.g., Yiu (2004); Cuoco and Liu (2006); Leippold et al. (2006); Cuoco et al. (2008), that apply the VaR dynamically. Generally, closed-form solutions are unavailable and one must resort to numerical solutions or asymptotic solutions. Alternatively, one can apply dynamic time-consistent risk measures. However, Kupper and Schachermayer (2009) show that the only dynamic risk measure that is both law-invariant and time consistent is the entropic risk measure, which is too restrictive to accommodate VaR and ES, two popular regulatory risk measures. To maintain analytical tractability and a relatively general framework, we choose our current setting. Our setting allows us to calculate relevant quantities in the general equilibrium models explicitly. Moreover, for reporting purposes the time horizon for calculating VaR or ES is typically one to ten days. In practice, it is almost impossible to measure a portfolio's risk continuously as in Yiu (2004); Cuoco and Liu (2006); Leippold et al. (2006); Cuoco et al. (2008) and the risk is usually reported on a daily basis. The time horizon T in our models can be chosen to be small enough to match the industry practices.

Chapter 4

Optimal Dynamic Reinsurance Policies Under Mean-CVaR Principle

This chapter addresses the risk management for insurers from the optimal reinsurance's perspective. An active line of research in optimal reinsurance focuses on the question how an insurance company should choose a reinsurance strategy to minimize its risk. A particular stability measure in the dynamic reinsurance literature for insurers is the *ruin probability*, which is an indication of the long-term viability of the business. While the majority of the optimal reinsurance literature, especially in the dynamic setting, focuses on ruin probability minimization based on the assumption that the insurer can only use proportional or excess-of-loss reinsurance arrangements, and/or the assumption that the reinsurance premium is calculated according to the expected value or variance principle, this chapter goes a bit further by relaxing these assumptions.

We consider the dynamic reinsurance design problem for an insurer whose surplus is modeled by the diffusion approximation of the classical Cramér-Lundberg model. The insurer can reinsure part of each claim to minimize the ruin probability and the reinsurance premiums paid to the reinsurer are calculated according to the Mean-CVaR premium principle which is a generalization of the expected value principle and Denneberg's absolute deviation principle. We use dynamic programming to solve the problem and obtain the optimal contract and the value function explicitly.

One of the contributions is that we impose a monotonicity constraint on the contracts to rule out the moral hazard. We overcome the difficulties arising from this constraint by means of the relaxation method and derive two forms of contracts that have not appeared in the dynamic reinsurance literature before.

The rest of this chapter is organized as follows. Section 4.1 introduces the Denneberg's absolute deviation principle and Mean-CVaR premium principle. Section 4.2 formulates the dynamic reinsurance model. In Section 4.3, we derive the HJB equation and characterize the optimal reinsurance strategy explicitly. Section 4.4 gives some concluding remarks.

4.1 Denneberg's Absolute Deviation Principle

Reinsurance premium calculation is the key element to actuarial science. A reinsurance risk X is often defined as a non-negative loss random variable, Mathematically, a premium calculation principle is a mapping from X to the nonnegative real line. Among premium principles, the expected value principle is probably the most widely studied principle in the insurance literature, e.g. Cheung (2010); Chi and Tan (2010); Meng and Zhang (2010); Meng and Siu (2011); Bernard et al. (2015b); Xu et al. (2015). One focus of this chapter is the Denneberg's absolute deviation principle (Denneberg (1990)), which will be introduced in the following. First of all, let us define Value-at-Risk (VaR) and Conditional Value-at-Risk (CVaR).

Definition 4.1.1. The VaR and CVaR of a random variable Z at a confidence level α (with $0 < \alpha < 1$) are respectively defined as

$$VaR_\alpha(Z) \triangleq \inf\{z \in \mathbb{R} : F_Z(z) \geq \alpha\},$$

and

$$CVaR_\alpha(Z) \triangleq \frac{1}{1-\alpha} \int_\alpha^1 VaR_s(Z) ds,$$

provided that the integral exists.

Given the distribution function F_Z , the generalized inverse F_Z^{-1} is called the *quantile function* of F_Z . Formally, the quantile function of Z is defined as

$$F_Z^{-1}(t) = \inf\{z \in \mathbb{R} : F_Z(z) \geq t\}.$$

It follows readily that

$$VaR_\alpha(Z) = F_Z^{-1}(\alpha)$$

and

$$CVaR_\alpha(Z) = \frac{1}{1-\alpha} \int_\alpha^1 F_Z^{-1}(s) ds.$$

We emphasize that the definitions in this chapter are different from those in the previous chapters. This is because the VaR and CVaR in Chapter 2 and 3 are defined over the profits while in this chapter the risk measures are defined over losses.

The definition of the Denneberg's absolute deviation principle (Denneberg (1990)) is given as follows:

Definition 4.1.2. The Denneberg's absolute deviation principle is

$$\pi(Z) = \mathbb{E}(Z) + \rho\tau(Z). \quad (4.1)$$

where $0 \leq \rho < 1$ and $\tau(Z)$ is the average absolute deviation from the median, i.e. $VaR_{0.5}(Z)$.

As discussed in Denneberg (1990), this premium principle coincides with the expected value principle

$$\pi(X) = (1 + \rho)\mathbb{E}(Z).$$

for some special distributions satisfying $F_Z(0) \geq 0.5$. Moreover, it can be shown that

$$\begin{aligned} \pi(Z) &= \mathbb{E}(Z) + \rho\tau(Z) \\ &= \int_0^1 F_Z^{-1}(t)dt + \rho \int_0^1 |F_Z^{-1}(t) - F_Z^{-1}(0.5)|dt \\ &= \int_0^{\frac{1}{2}} F_Z^{-1}(t)(1 - \rho)dt + \int_{\frac{1}{2}}^1 F_Z^{-1}(t)(1 + \rho)dt \\ &= (1 - \rho) \int_0^1 F_Z^{-1}(t)dt + 2\rho \int_{\frac{1}{2}}^1 F_Z^{-1}(t)dt \\ &= (1 - \rho)\mathbb{E}(Z) + \rho CVaR_{\frac{1}{2}}(Z). \end{aligned}$$

Essentially, the premium of a risk under Denneberg's absolute deviation principle is a weighted average of the risk's expectation and CVaR at the confidence level 0.5. Inspired by the Denneberg's absolute deviation principle, we propose the following Mean-CVaR premium principles:

Definition 4.1.3. The Mean-CVaR premium principle is defined as

$$\pi(Z) = \frac{1 + \theta}{1 + \beta} \left[\mathbb{E}(Z) + \beta CVaR_{\alpha}(Z) \right], \quad (4.2)$$

where $\theta, \beta \geq 0$ and $\alpha \in [0, 1]$.

The Mean-CVaR premium is a weighted average of the risk's expectation and CVaR at the confidence level α . This specification greatly adds flexibility to the premium principle. α measures the degree to which the reinsurer cares about the tail risk. θ is analogous to the safety loading in the classical expected value principle. For this reason, we refer to it as the *premium loading*. β , which is similar to the weighting parameter ρ in the Denneberg's absolute deviation principle, measures the relative importance of the tail risk. Therefore, we refer to it as the *risk loading*.

Our premium principle generalizes the expected value premium principle and Denneberg's absolute deviation principle. When $\beta = 0$ or $\alpha = 0$, the Mean-CVaR premium principle reduces to the expected value premium principle. When $\theta = 0$, $\beta = \frac{\rho}{1-\rho}$, and $\alpha = 0.5$, the Mean-CVaR premium principle reduces to the Denneberg's absolute deviation principle. When $\theta = 0$ and $\beta = \frac{\rho}{1-\rho}$, the Mean-CVaR premium principle can be regarded as an extension of Denneberg's absolute deviation principle in which $\tau(X)$ is replaced with the average absolute deviation from the $VaR_\alpha(Z)$. It is easy to show that the Mean-CVaR premium principle satisfy several desirable properties of premium principles, such as independence, risk loading, maximal loss, translation invariance, scale invariance, subadditivity, comonotonic additivity, monotonicity, preservation of the first stochastic dominance ordering, preservation of the stop-loss ordering and continuity. See Young (2004) for a more detailed discussion.

4.2 Dynamic Reinsurance Model

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space with filtration $\{\mathcal{F}_t\}$. Consider an insurer whose surplus process is described by the classical Cramér-Lundberg (CL) model:

$$R_t = R_0 + p - \sum_{i=1}^{N_t} Z_i, \quad (4.3)$$

where $p > 0$ is the premium rate, R_0 is the initial reserve, N_t is the Poisson process of the incoming claims and Z_i is a sequence of i.i.d. random variables, representing the sizes of the successive claims. Without loss of generality, we assume that the intensity of the process N_t is 1. It is also assumed that Z_i has the cumulative distribution function $F_Z(\cdot)$, quantile function $F_Z^{-1}(\cdot)$ and $p > E[Z_i]$.

Suppose the insurer can purchase reinsurance to control his or her insurance business risk. Recall that Z_i is the loss insured by the insurer in the absence of reinsurance. Mathematically, we target to obtain an optimal partition on Z_i into $I(Z_i)$ and $H(Z_i)$ so that $Z_i = I(Z_i) + H(Z_i)$, where $I : [0, \infty] \rightarrow [0, \infty]$ is a measurable function

and so is H . In the reinsurance setting, $I(X)$ represents the portion of loss that is ceded to a reinsurer and $H(X)$ is the residual loss retained by the insurer. In the context of optimal reinsurance, the ceded loss function I is often restricted to the set \mathcal{I} as given below (Chi and Tan (2010); Cai et al. (2016); Chi and Lin (2014); Boonen et al. (2016)):

$$\mathcal{I} := \left\{ I : [0, \infty] \rightarrow [0, \infty] \mid 0 \leq I(x) - I(y) \leq x - y \ \forall x \leq y \right\}. \quad (4.4)$$

Following the standard in the literature, we require that $I \in \mathcal{I}$. The increasing assumption on the ceded loss function I and the retained loss function H is imposed to reduce the moral hazard risk. If the ceded loss function I is strictly decreasing over certain interval(s), the insurer can claim less loss to get additional compensations. If the retained function H is strictly decreasing over certain interval(s), the insurer can encourage the policyholders to claim more so as to reduce its own retained loss but increase the ceded loss on the reinsurer. Therefore, in order to exclude the moral hazard, we assume that both the insurer and reinsurer are obligated to pay more for a larger loss X . In fact, without this constraint, the optimal reinsurance contract can be strictly decreasing over some interval(s) in the single period reinsurance model (Weng and Zhuang (2016)).

Corresponding to a chosen reinsurance policy I , we assume the reinsurance premium rate $\bar{p}(I)$ payable to the reinsurer is given by the Mean-CVaR premium principle, i.e.

$$\bar{p}(I) = \pi(I(Z_i)).$$

By using change of variable, we can get

$$\begin{aligned} \pi(I(Z_i)) &= \frac{1+\theta}{1+\beta} \left[\mathbb{E}[I(Z_i)] + \beta CVaR_\alpha(I(Z_i)) \right] \\ &= \frac{1+\theta}{1+\beta} \left[\int_0^1 F_{I(Z_i)}^{-1}(t) dt + \frac{\beta}{1-\alpha} \int_\alpha^1 F_{I(Z_i)}^{-1}(t) dt \right] \\ &= \frac{1+\theta}{1+\beta} \left[\int_0^1 I(F_Z^{-1}(t)) dt + \frac{\beta}{1-\alpha} \int_\alpha^1 I(F_Z^{-1}(t)) dt \right] \\ &= \frac{1+\theta}{1+\beta} \left[\int_0^\infty I(z) dF_Z(z) + \frac{\beta}{1-\alpha} \int_{F_Z^{-1}(\alpha)}^\infty I(z) dF_Z(z) \right] \\ &= \int_0^{+\infty} I(z) g(z) dF_Z(z) \end{aligned}$$

where

$$g(z) = \begin{cases} \frac{1+\theta}{1+\beta}, & 0 \leq z \leq F_Z^{-1}(\alpha), \\ \frac{(1+\theta)(1-\alpha+\beta)}{(1-\alpha)(1+\beta)}, & F_Z^{-1}(\alpha) \leq z. \end{cases}$$

Here, we use the notation $g(\cdot)$ to simplify $\pi(I(Z))$.

We impose the following assumption.

Assumption 4.2.1. $\mathbb{E}[Z_i] < p < \pi(Z_i) = \int_0^{+\infty} zg(z)dF_Z(z)$.

This assumption stipulates that the insurer has a positive safety loading; otherwise the ruin is certain. Moreover, the reinsurance is *noncheap*. If the reinsurance is *cheap*, then the insurer can always choose to fully reinsure all claims and the corresponding ruin probability is 0.

Thus, the insurer's reserve process under reinsurance is given by

$$R_t^I = R_0 + (p - \bar{p}(I))t - \sum_{i=1}^{N_t} (Z_i - I(Z_i)).$$

According to Grandell (2012), the reserve process can be approximated by a diffusion process of the following form

$$dR_t^I = (p - \bar{p}(I) - \mathbb{E}[Z_i - I(Z_i)])dt + \sqrt{\mathbb{E}[(Z_i - I(Z_i))^2]}dB_t,$$

where B_t is a standard Brownian motion.

Define

$$\begin{aligned} c &:= p - \mathbb{E}[Z_i], \\ \mu(I) &:= -\bar{p}(I) + \mathbb{E}[I(Z_i)] = -\int_0^{+\infty} I(z)d(z)dF_Z(z), \\ \sigma^2(I) &:= \mathbb{E}[(Z_i - I(Z_i))^2] = \int_0^{+\infty} (z - I(z))^2 dF_Z(z), \end{aligned}$$

where

$$d(z) := g(z) - 1 = \begin{cases} \frac{1+\theta}{1+\beta} - 1, & 0 \leq z < F_Z^{-1}(\alpha), \\ \frac{(1+\theta)(1-\alpha+\beta)}{(1-\alpha)(1+\beta)} - 1, & F_Z^{-1}(\alpha) \leq z. \end{cases}$$

Suppose the insurer chooses the reinsurance policy I_t at time t . An admissible reinsurance policy I is described by a $(\mathcal{F}_t)_{t \geq 0}$ -adapted stochastic process $I_t(Z)$, where $I_t \in \mathcal{I}, \forall t \geq 0$. Also, we denote by R_t^I the reserve process corresponding to an admissible policy I . Consequently, the dynamics of the reserve process can be written as

$$dR_t^I = (c + \mu(I_t))dt + \sigma(I_t)dB_t. \quad (4.5)$$

The ruin time is defined as

$$\tau^I = \inf\{t > 0 | R_t^I \leq 0\},$$

and the insurer's objective is to choose a reinsurance policy I to minimize its ruin probability

$$V^I(x) = \mathbb{P}(\tau^I < \infty | R_t^I = x).$$

4.3 Solution

4.3.1 HJB Equation

We use the dynamic programming approach to solve the above problem. Let $V(x)$ be the minimal ruin probability, i.e., $V(x) = \inf_I V^I(x)$. V solves the following Hamilton-Jacobi-Bellman (HJB) equation

$$0 = \min_{I \in \mathcal{I}} \left[(c + \mu(I))V'(x) + \frac{1}{2}\sigma^2(I)V''(x) \right]. \quad (4.6)$$

In what follows we will seek a twice continuously differentiable solution to (4.6) subject to the following boundary conditions

$$\begin{aligned} V(0) &= 1, \\ V(\infty) &= 0. \end{aligned}$$

These boundary conditions reflect the fact that in a typical case the probability of ruin approaches 1 when the surplus approaches 0 and the probability of ruin approaches 0, when the surplus approaches infinity.

We make the ansatz $V(x) = e^{-ax}$ for some positive constant a to be determined. The inner optimization in (4.6) then becomes

$$h(a) := \min_{I \in \mathcal{I}} J(I; a), \quad (4.7)$$

where

$$J(I; a) := -\mu(I) + \frac{a}{2}\sigma^2(I) = \int_0^{+\infty} \left\{ I(z)d(z) + \frac{a}{2}(z - I(z))^2 \right\} dF_Z(z). \quad (4.8)$$

In what follows we will first solve (4.7) for a fixed $a > 0$, and then determine a by solving $h(a) = c$.

4.3.2 Relaxed Solutions

As we require $I \in \mathcal{I}$, it is difficult to solve (4.7) directly. Instead, we relax (4.7) by changing the feasible set (for I) from \mathcal{I} to $\{0 \leq I(z) \leq z\}$ and consider the following problem

$$\min_{0 \leq I(z) \leq z} J(I; a) = \min_{0 \leq I(z) \leq z} \int_0^{+\infty} \left\{ I(z)d(z) + \frac{a}{2}(z - I(z))^2 \right\} dF_Z(z). \quad (4.9)$$

Note that

$$I(z)d(z) + \frac{a}{2}(z - I(z))^2$$

is strictly convex in $I(z)$. Thus, pointwise optimization immediately gives the optimal solution

$$\tilde{I}(z; a) = \min \left\{ z, \max \left\{ 0, z - \frac{d(z)}{a} \right\} \right\}. \quad (4.10)$$

For ease of exposition, we define

$$k_1 := \frac{(1 + \theta)(1 - \alpha + \beta)}{(1 - \alpha)(1 + \beta)} - 1,$$

and

$$k_2 := \frac{1 + \theta}{1 + \beta} - 1 = \frac{\theta - \beta}{1 + \beta}.$$

It follows easily that $k_2 \leq k_1$. There are three cases:

Case 4.3.1. $k_1 \leq 0$.

We have $k_2 \leq 0$ and

$$\begin{aligned} & \int_0^{+\infty} z g(z) dF_Z(z) - \int_0^{+\infty} z dF_Z(z) \\ &= \int_0^{+\infty} z d(z) dF_Z(z) \leq 0, \end{aligned}$$

violating Assumption 4.2.1. Thus, $k_1 > 0$.

Case 4.3.2. $k_1 > 0$ and $\theta \leq \beta$.

We have $k_2 \leq 0$ and $\tilde{I}(z; a)$ in (4.10) reduces to

$$\tilde{I}(z; a) = \begin{cases} z, & 0 \leq z < F_Z^{-1}(\alpha), \\ \max\{0, z - \frac{k_1}{a}\}, & F_Z^{-1}(\alpha) \leq z. \end{cases} \quad (4.11)$$

Case 4.3.3. $k_1 > 0$ and $\theta > \beta$.

We have $k_2 > 0$ and $\tilde{I}(z; a)$ in (4.10) reduces to

$$\tilde{I}(z; a) = \begin{cases} \max\{0, z - \frac{k_2}{a}\}, & 0 \leq z < F_Z^{-1}(\alpha), \\ \max\{0, z - \frac{k_1}{a}\}, & F_Z^{-1}(\alpha) \leq z. \end{cases} \quad (4.12)$$

We see that (4.11) and (4.12) might not satisfy the requirement that $I \in \mathcal{I}$ as they are not necessarily monotone at $F_Z^{-1}(\alpha)$. In the following, we will modify these two solutions to satisfy the monotonicity constraint.

4.3.3 Modified Solutions

In the following, we modify (4.11) and (4.12) so that they can satisfy the requirement that $I \in \mathcal{I}$.

Lemma 4.3.1. *In Case 4.3.2, for any given feasible contract $I \in \mathcal{I}$, there exists a contract given by*

$$I^*(z) = \begin{cases} z, & 0 \leq z \leq m, \\ m, & m < z \leq m + \frac{k_1}{a}, \\ z - \frac{k_1}{a}, & m + \frac{k_1}{a} < z, \end{cases} \quad (4.13)$$

where $m \leq F_Z^{-1}(\alpha)$, such that $J(I^*; a) \leq J(I; a)$.

Proof of Lemma 4.3.1. For any given feasible contract $I \in \mathcal{I}$, we denote

$$m := \max \left\{ I(F_Z^{-1}(\alpha)), F_Z^{-1}(\alpha) - \frac{k_1}{a} \right\}$$

and construct the contract (4.13) according to the value of m . Then it follows that

$$\begin{aligned} J(I^*; a) - J(I; a) &\leq \int_m^{\frac{k_1}{a}+m} \left\{ I^*(z)d(z) + \frac{a}{2}(z - I^*(z))^2 \right\} dF_Z(z) \\ &\quad - \int_m^{\frac{k_1}{a}+m} \left\{ I(z)d(z) + \frac{a}{2}(z - I(z))^2 \right\} dF_Z(z) \\ &\leq 0, \end{aligned}$$

where the last inequality is due to the fact that

$$yd(z) + \frac{a}{2}(z - y)^2$$

is strictly convex in y and

$$\begin{aligned} \tilde{I}(z; a) &\geq I^*(z) \geq I(z), & m \leq z \leq F_Z^{-1}(\alpha), \\ \tilde{I}(z; a) &\leq I^*(z) \leq I(z), & F_Z^{-1}(\alpha) < z \leq m. \end{aligned}$$

□

Lemma 4.3.2. *In Case 4.3.3, for any given feasible contract $I \in \mathcal{I}$, there exists a contract given by:*

$$I^*(z) = \begin{cases} 0, & 0 \leq z \leq \frac{k_2}{a}, \\ z - \frac{k_2}{a}, & \frac{k_2}{a} < z \leq m + \frac{k_2}{a}, \\ m, & m + \frac{k_2}{a} < z \leq m + \frac{k_1}{a}, \\ z - \frac{k_1}{a}, & m + \frac{k_1}{a} < z, \end{cases} \quad (4.14)$$

where $m \leq \max\{0, F_Z^{-1}(\alpha) - \frac{k_2}{a}\}$, such that $J(I^*; a) \leq J(I; a)$.

Proof of Lemma 4.3.2. For any given feasible contract $I \in \mathcal{I}$, we denote

$$m := \min \left\{ \max \left\{ 0, F_Z^{-1}(\alpha) - \frac{k_2}{a} \right\}, \max \left\{ I(F_Z^{-1}(\alpha)), F_Z^{-1}(\alpha) - \frac{k_1}{a} \right\} \right\}$$

and construct the contract (4.14) according to the value of m . If $m > 0$, then

$$\begin{aligned} J(I^*; a) - J(I; a) &\leq \int_{\frac{k_2}{a}+m}^{\frac{k_1}{a}+m} \left\{ I^*(z)d(z) + \frac{a}{2}(z - I^*(z))^2 \right\} dF_Z(z) \\ &\quad - \int_{\frac{k_2}{a}+m}^{\frac{k_1}{a}+m} \left\{ I(z)d(z) + \frac{a}{2}(z - I(z))^2 \right\} dF_Z(z) \\ &\leq 0, \end{aligned}$$

where the last inequality is due to the fact that

$$yd(z) + \frac{a}{2}(z - y)^2$$

is strictly convex on y and

$$\begin{aligned} \tilde{I}(z; a) &\geq I^*(z) \geq I(z), & \frac{k_2}{a} + m \leq z \leq F_Z^{-1}(\alpha), \\ \tilde{I}(z; a) &\leq I^*(z) \leq I(z), & F_Z^{-1}(\alpha) < z \leq \frac{k_1}{a} + m. \end{aligned}$$

Otherwise, $m = 0$ and $I^*(z) = \max\{0, z - \frac{k_1}{a}\} = \tilde{I}(z; a)$ which is optimal. \square

Consequently, we can consider the following one-dimensional optimization problem

$$\bar{J}(a) := \min_{m \leq F_Z^{-1}(\alpha)} \int_0^{+\infty} \left\{ I(z; a, m)d(z) + \frac{a}{2}(z - I(z; a, m))^2 \right\} dF_Z(z), \quad (4.15)$$

where

$$I(z; a, m) = \begin{cases} z, & 0 \leq z \leq m, \\ m, & m < z \leq m + \frac{k_1}{a}, \\ z - \frac{k_1}{a}, & m + \frac{k_1}{a} < z, \end{cases}$$

if $\theta \leq \beta$, and

$$I(z; a, m) = \begin{cases} 0, & 0 \leq z \leq \frac{k_2}{a}, \\ z - \frac{k_2}{a}, & \frac{k_2}{a} < z \leq m + \frac{k_2}{a}, \\ m, & m + \frac{k_2}{a} < z \leq m + \frac{k_1}{a}, \\ z - \frac{k_1}{a}, & m + \frac{k_1}{a} < z, \end{cases}$$

otherwise.

Lemma 4.3.3. *For any $a > 0$, there exists a unique $I^*(z; a) \in \mathcal{I}$ such that $I^*(z; a)$ solves (4.7), i.e. $\min_{I \in \mathcal{I}} J(I; a)$. Moreover, $I^*(z; a) = I(z; a, m^*)$ where m^* is the unique minimizer of (4.15).*

Proof of Lemma 4.3.3. From Lemma 4.3.1 and Lemma 4.3.2, (4.7) is equivalent to (4.15). The existence of the optimal m^* can be proved by the Weierstrass extreme value theorem.

Suppose that both m_1 and m_2 are optimal to (4.15), then both $I^1 := I(z; a, m_1)$ and $I^2 := I(z; a, m_2)$ are optimal to (4.7). Define

$$\bar{I} := \frac{1}{2}I^1 + \frac{1}{2}I^2,$$

which is still a feasible solution. Because $J(I; a)$ is strictly convex in I , we have

$$J(\bar{I}; a) < \frac{1}{2}J(I^1; a) + \frac{1}{2}J(I^2; a) = J(I^1; a) = J(I^2; a),$$

which contradicts the optimality of I^1 and I^2 . \square

We now determine a that solves (4.6), namely $h(a) = \bar{J}(a) = c$.

Lemma 4.3.4. *There is a unique positive a^* such that $h(a^*) = c$.*

Proof of Lemma 4.3.4. In view of the envelope theorem, h is continuous and strictly increasing. Because $I(z) \equiv 0$ is a feasible solution, we have

$$h(a) \leq \int_0^{+\infty} \frac{a}{2} z^2 dF_Z(z).$$

Thus,

$$\lim_{a \downarrow 0} h(a) \leq \lim_{a \downarrow 0} \int_0^{+\infty} \frac{a}{2} z^2 dF_Z(z) = 0.$$

Next,

$$\begin{aligned} \lim_{a \uparrow +\infty} h(a) &\geq \lim_{a \uparrow +\infty} \int_0^{+\infty} \left\{ \tilde{I}(z; a) d(z) + \frac{a}{2} (z - \tilde{I}(z; a))^2 \right\} dF_Z(z) \\ &= \lim_{a \uparrow +\infty} \int_0^{+\infty} \left\{ zd(z) \mathbf{1}_{d(z) \leq 0} + \left[zd(z) - \frac{(d(z))^2}{2a} \right] \mathbf{1}_{0 < \frac{d(z)}{a} < z} + \frac{a}{2} z^2 \mathbf{1}_{\frac{d(z)}{a} \geq z} \right\} dF_Z(z) \\ &\geq \lim_{a \uparrow +\infty} \int_0^{+\infty} \left\{ zd(z) \mathbf{1}_{d(z) \leq 0} + \left[zd(z) - \frac{(d(z))^2}{2a} \right] \mathbf{1}_{0 < \frac{d(z)}{a} < z} \right\} dF_Z(z) \\ &= \int_0^{+\infty} zd(z) dF_Z(z), \end{aligned}$$

where the convergence follows from the monotone convergence theorem. From Assumption 4.2.1, it follows that

$$\lim_{a \uparrow \infty} h(a) > c$$

and hence there is a unique positive a^* such that $h(a^*) = c$. \square

Summarizing the above results, we can find the optimal solution in three steps:

1. Fix a , find the minimizer $m^*(a)$ of (4.15). Here, $m^*(a)$ is a function of a .
2. Find a^* such that $h(a^*) = 0$.
3. Find the optimal contract I^* .

If $\theta \leq \beta$,

$$I^*(z) = \begin{cases} z, & 0 \leq z \leq m^*(a^*), \\ m^*(a^*), & m^*(a^*) < z \leq m^*(a^*) + \frac{k_1}{a^*}, \\ z - \frac{k_1}{a^*}, & m^*(a^*) + \frac{k_1}{a^*} < z. \end{cases} \quad (4.16)$$

If $\theta > \beta$,

$$I^*(z) = \begin{cases} 0, & 0 \leq z \leq \frac{k_2}{a^*}, \\ z - \frac{k_2}{a^*}, & \frac{k_2}{a^*} < z \leq m^*(a^*) + \frac{k_2}{a^*}, \\ m^*(a^*), & m^*(a^*) + \frac{k_2}{a^*} < z \leq m^*(a^*) + \frac{k_1}{a^*}, \\ z - \frac{k_1}{a^*}, & m^*(a^*) + \frac{k_1}{a^*} < z. \end{cases} \quad (4.17)$$

4.3.4 Verification Theorem

We verify that (4.16) and (4.17) are indeed optimal. We first show a useful lemma.

Lemma 4.3.5. *For any policy I and any $N > 0$,*

$$\tau_N^I < \infty \quad (4.18)$$

almost surely, where $\tau_N^I = \inf\{t \geq 0 : R_t^I \notin (0, N)\}$ with R_t^I given by (4.5).

Proof of Lemma 4.3.5. Define

$$\begin{aligned} \psi(t) &= \int_0^t \sigma^2(I_s) ds, \\ \gamma(t) &= \int_0^t (c + \mu(I_s)) ds, \\ \Delta &= \int_0^{+\infty} z g(z) dF_Z(z) - p. \end{aligned}$$

Assumption 4.2.1 implies that $\Delta > 0$. Define

$$\eta_0 = 0, \eta_{k+1} = \inf\{t > \eta_k : \psi(t) - \psi(\eta_k) = 1\}. \quad (4.19)$$

We next give an upper bound of $\gamma(t)$ on $\{\psi(\eta_{k+1}) - \psi(\eta_k) \leq 1\}$.

$$\begin{aligned}
\gamma(t\eta_{k+1}) - \gamma(t\eta_k) &= \int_{\eta_k}^{\eta_{k+1}} (p - \int_0^\infty I_s(z)g(z)dF_Z(z) + \int_0^\infty (I_s(z) - z)dF_Z(z))ds \\
&= \int_{\eta_k}^{\eta_{k+1}} (p - \int_0^\infty zg(z)dF_Z(z) + \int_0^\infty (z - I_s(z))g(z)dF_Z(z) \\
&\quad + \int_0^\infty (I_s(z) - z)dF_Z(z))ds \\
&< \int_{\eta_k}^{\eta_{k+1}} (\int_0^\infty (z - I_s(z))g(z)dF_Z(z) + \int_0^\infty (I_s(z) - z)dF_Z(z))ds \\
&\leq \int_{\eta_k}^{\eta_{k+1}} (\sqrt{\int_0^\infty (z - I_s(z))^2dF_Z(z)}\sqrt{\int_0^\infty g(z)^2dF_Z(z)} \\
&\quad + \sqrt{\int_0^\infty (I_s(z) - z)^2dF_Z(z)})ds \\
&\leq \left(1 + \sqrt{\int_0^\infty g(z)^2dF_Z(z)}\right).
\end{aligned}$$

Suppose that $\mathbb{P}(\eta_k < \infty) = 1, \forall k$. Following the arguments in Taksar and Markussen (2003), we have that

$$\mathbb{P}(\tau_N^I > \eta_m) \leq (1 - \delta_0)^n,$$

with

$$\delta_0 = \Phi\left(-N - (1 + \sqrt{\int_0^\infty g(z)^2dF_Z(z)})\right).$$

Here Φ is the cumulative distribution function of a standard normal random variable.

Letting $n \rightarrow \infty$, we then have that

$$P(\tau_N^I < \infty) = 1.$$

Next we suppose that $\mathbb{P}(\eta_{m+1} = \infty) = \delta$ for some $m \geq 0$. Then similar to the above derivation we have. for $t > \eta_m$,

$$\int_t^\infty (c + \mu(I_s))ds \leq \int_t^\infty -\Delta ds + \left(1 + \sqrt{\int_0^\infty g(z)^2dF_Z(z)}\right) = -\infty$$

Define

$$\begin{aligned}
\xi_0 &= \eta_m, \\
\eta'_{k+1} &= \inf\{s : s > \xi_k, \psi(s) - \psi(\eta_k) = 1\}, \\
\eta''_{k+1} &= \inf\{s : s > \xi_k, \gamma(s) - \gamma(\xi_k) < -K\}, \\
\xi_{k+1} &= \min\{\eta'_{k+1}, \eta''_{k+1}\},
\end{aligned}$$

where K is chosen in a way such that

$$\frac{1}{(N - K)^2} < \frac{\delta}{2}.$$

Then the arguments in Taksar and Markussen (2003) yield that

$$\mathbb{P}(\tau_N^I > \xi_n) < (1 - \frac{\delta}{2})^n.$$

This completes the proof. \square

We are now able to prove the verification theorem.

Theorem 4.3.1 (Verification theorem). *The optimal reinsurance policy is given by (4.16) or (4.17), and the minimal ruin probability is*

$$V(x) = e^{-a^*x}, \quad (4.20)$$

where a^* solves $h(a^*) = c$.

Proof of Theorem 4.3.1. It is not difficult to see that $V(x)$ given by (4.20) is

$$\mathbb{P}(\tau^{I^*} < \infty | R_0^{I^*} = x),$$

where I^* is given by (4.16) or (4.17). It suffices to show that I^* is optimal.

For any admissible policy I such that the SDE (4.5) admits a unique solution for $R_0 = x, \forall x \geq 0$, by Dynkin's formula we have that

$$\mathbb{E}[V(R_{\tau^I \wedge t}^I) | R_0^I = x] = V(x) + \mathbb{E} \int_0^{\tau^I \wedge t} (c + \mu(I_s)V'(R_s^I) + \frac{1}{2}\sigma^2(I_s)V''(R_s^I))ds | R_0 = x]$$

The the HJB equation (4.6) yields that

$$\mathbb{E}[V(R_{\tau^I \wedge t}^I) | R_0^I = x] \geq V(x)$$

Since V is continuous and bounded by 1, we then have that

$$\begin{aligned} \lim_{t \rightarrow \infty} \mathbb{E}[V(R_{\tau^I \wedge t}^I) | R_0^I = x] &= \lim_{t \rightarrow \infty} \mathbb{E}[V(R_t^I)I_{\tau^I = \infty} | R_0^I = x] + \mathbb{E}[V(R_{\tau^I}^I)I_{\tau^I < \infty} | R_0^I = x] \\ &= \mathbb{E}[\lim_{t \rightarrow \infty} e^{-a^*R_t^I} I_{\tau^I = \infty} | R_0^I = x] + \mathbb{E}[e^{-a^*R_{\tau^I}^I} I_{\tau^I < \infty} | R_0^I = x] \end{aligned}$$

It follows from Lemma 4.3.5 that either $\tau^I < \infty$ or $\lim_{t \rightarrow \infty} R_t^I = \infty$. Consequently,

$$\mathbb{P}(\tau^I < \infty | R_0^I = x) = \lim_{t \rightarrow \infty} \mathbb{E}V(R_{\tau^I \wedge t}^I | R_0^I = x) \geq V(x).$$

\square

Remark 4.3.1. If $\beta = 0$, then the Mean-CVaR premium principle reduces to the expected value principle. We have

$$k_1 = k_2 = \theta,$$

and

$$I^*(z) = \max\{0, z - \frac{\theta}{a^*}\},$$

where a^* satisfies $h(a^*) = c$.

If $\alpha = 0$, then the optimal m in (4.13) and (4.14) is equal to 0. Then the optimal reinsurance strategy in both cases will reduce to the excess-of-loss treaty, which is consistent with the results in Meng and Zhang (2010); Hipp and Taksar (2010).

Remark 4.3.2. When $\theta = 0$, $\beta = \frac{\rho}{1-\rho}$, and $\alpha = 0.5$, i.e., the Mean-CVaR premium principle reduces to the Denneberg's absolute deviation principle, we have $\theta < \beta$ and the optimal reinsurance contract is given in the form of (4.16).

Theorem 4.3.1 adds to various results in the literature. It is shown in Meng and Zhang (2010) and Hipp and Taksar (2010) that the excess-of-loss reinsurance is optimal when the premium is calculated according to the expected value principle, and in Hipp and Taksar (2010) that the proportional reinsurance is optimal when the premium is calculated according to the variance premium principle. We find that, under the Mean-CVaR premium principle the optimal reinsurance contract is likely to involve a layer. To the best of our knowledge, we are the first to show the optimality of such contracts in a dynamic setting. In the literature, similar contracts have been derived in the static settings under different objectives, e.g. Cai et al. (2016); Cheung et al. (2012).

The optimal contract can have different shapes depending on the relationship between the premium loading θ and risk loading β . In all cases, the optimal contract covers large losses, which can lead to insolvency and drive the insurer out of the business. Because the insurer's objective is to minimize the ruin probability, it optimally chooses to reinsure against all large losses.

If $m^*(a^*) = 0$, the optimal contract degenerates to an excess-of-loss contract. Otherwise, there are two cases. If $\theta \leq \beta$, i.e., the reinsurer charges more for the tail risk, the insurer optimally chooses to fully insure against small losses and the optimal contract is given by (4.16). Figure 4.1 plots (4.16). The contract resembles a cap reinsurance contract first and then an excess-of-loss treaty. Therefore, we term such contract the *excess-of-loss over a cap*. If $\theta > \beta$, i.e., it is costly to insure against small losses, the optimal coverage shifts from small losses to medium losses and it is given

by (4.17). Figure 4.2 plots (4.17). The contract has a layer above which it resembles an excess-of-loss arrangement. Thus, we term such contract the *excess-of-loss over a layer*. In both contracts, α determine the boundary ($F_Z^{-1}(\alpha)$) between the cap (or layer) and excess-of-loss.

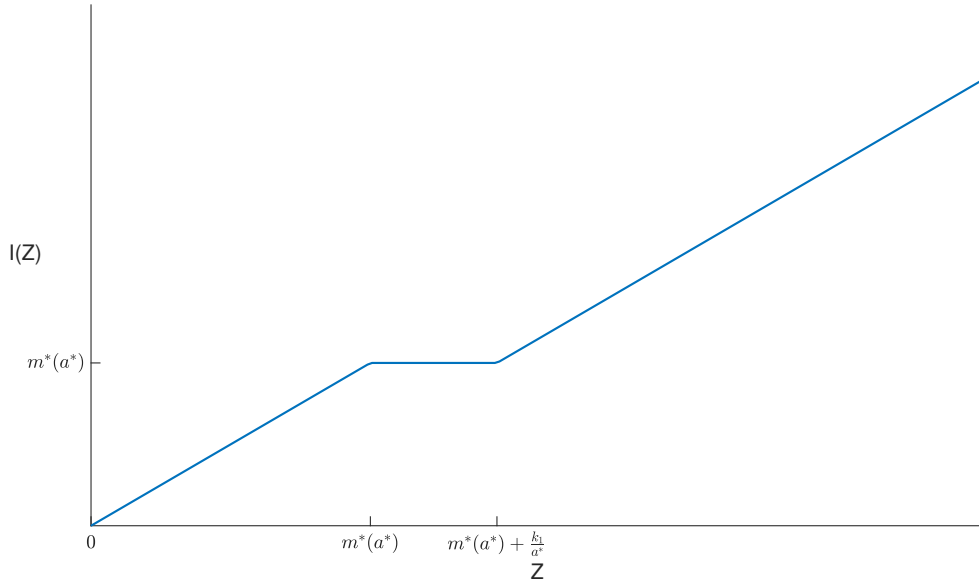


Figure 4.1: Excess-of-loss over a cap

This figure plots the optimal (non-degenerate) contract when $\theta \leq \beta$. For the first layer of the loss ($Z < m^*(a^*)$), the reinsurer is responsible for the amount of losses up to the limit $m^*(a^*)$. For losses that fall in the second layer ($m^*(a^*) \leq Z \leq m^*(a^*) + \frac{k_1}{a^*}$), the reinsurer indemnifies a constant cap amount of $m^*(a^*)$. Finally for the spill-over layer of loss ($Z > m^*(a^*) + \frac{k_1}{a^*}$), the reinsurer is responsible for any loss in excess of $m^*(a^*) + \frac{k_1}{a^*}$. The contract resembles a cap contract capped at $m^*(a^*)$ for $Z < F_Z^{-1}(\alpha)$, and an excess-of-loss contract with retention $m^*(a^*) + \frac{k_1}{a^*}$ for $Z \geq F_Z^{-1}(\alpha)$.

4.4 Concluding Remarks

The models and problems studied in this chapter can be explored further in different ways. For example, it would be interesting to generalize our results to a more general class of premium principles, such as the Wang's premium principle (Wang (1995, 1996); Wang et al. (1997)). In addition, we can consider alternative objectives such as the *expected discounted dividends* objective as in Taksar and Zhou (1998); Jgaard and Taksar (1999); Asmussen et al. (2000); Choulli et al. (2003); He and Liang (2009).

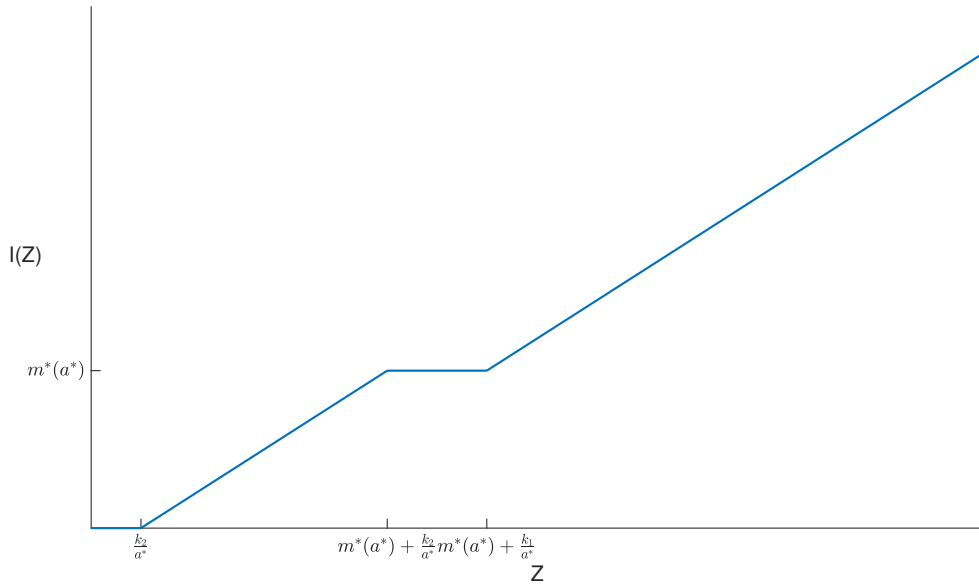


Figure 4.2: Excess-of-loss over a layer

This figure plots the optimal (non-degenerate) contract when $\theta > \beta$. The loss in the first layer ($Z < \frac{k_2}{a^*}$) is retained by the insurer. For losses that fall in the second layer ($\frac{k_2}{a^*} \leq Z \leq m^*(a^*) + \frac{k_2}{a^*}$), the reinsurer is responsible for the amount of losses in excess of $\frac{k_2}{a^*}$ up to the limit $m^*(a^*) + \frac{k_2}{a^*}$. For losses that fall in the third layer ($m^*(a^*) + \frac{k_2}{a^*} \leq Z \leq m^*(a^*) + \frac{k_1}{a^*}$), the reinsurer indemnifies a constant cap amount of $m^*(a^*)$. Finally for the spill-over layer of loss ($Z > m^*(a^*) + \frac{k_1}{a^*}$), the reinsurer is responsible for any loss in excess of $m^*(a^*) + \frac{k_1}{a^*}$. The contract resembles a layer contract with retention $\frac{k_2}{a^*}$ and cap $m^*(a^*)$ for $Z < F_Z^{-1}(\alpha)$, and an excess-of-loss contract with retention $m^*(a^*) + \frac{k_1}{a^*}$ for $Z \geq F_Z^{-1}(\alpha)$.

Chapter 5

Conclusion

This thesis considered the risk management for investment and insurance institutions respectively. In Chapter 2, we studied the effects of the WVaR-based risk management on the portfolio choice of expected utility maximizers, who derive utility from wealth at some horizon and must comply with a WVaR constraint imposed at that horizon. The feasibility, well-posedness, and existence of optimal solutions were discussed. When the optimal solution exists, we revealed several interesting effects. We characterized a class of risk measures that allows economic agents to engage in “regulatory capital arbitrage.” In particular, VaR and ES, two popular regulatory risk measures, often incur even larger losses in the most adverse states. This provides a critique of the current risk management practices and the Basel Committee’s plan to replace VaR with ES for calculating market risk capital requirements. On the other hand, we found conditions on risk measures that can lead to endogenous portfolio insurance and thus mitigate “regulatory capital arbitrage.” These findings may be of potential interest to regulators.

In Chapter 3, we studied the effects of Expected Shortfall on portfolio choice of expected utility maximizers, who derive utility from wealth at some horizon and must comply with a ES constraint imposed at that horizon. In the partial equilibrium analysis, we found that ES agents will insure against only intermediate states and incur larger losses than both VaR and Benchmark agents. We then extended our analysis to general equilibrium asset pricing models featuring ES agents. We showed that the market volatility and risk premium in the ES economy are larger than in the Benchmark economy when the output is low, in both production and pure-exchange models.

We then considered risk management for insurers in Chapter 3. We derived explicitly the optimal reinsurance contract under the Mean-CVaR premium principle, which is a generalization of Denneberg’s absolute deviation principle and classical expected

value principle. We imposed a monotonicity constraint on the reinsurance contract to eliminate the moral hazard risk, and overcame the difficulties arising from this constraint. The method in deriving the optimal contract is interesting in its own right. We found that there are two types of the optimal reinsurance contract, depending on the premium loading θ and risk loading β . Moreover, the optimal contracts are likely to involve a layer. These contracts are more complicated than the contracts widely studied in the dynamic reinsurance literature. We hope our findings can provide new directions in the future optimal reinsurance study.

Appendix A

Appendix to Chapter 2

A.1 Boundary Solution

We now study the optimal solution to (2.8) when (x, \underline{x}) is on the boundary of the feasible set. Suppose that $\sup_{0 < c < 1} \kappa_\Phi((c, 1]) = \frac{1}{\mathbb{E}\xi}$ and let $C := \{c \in (0, 1) \mid \kappa_\Phi((c, 1]) = \frac{1}{\mathbb{E}\xi}\}$. There are two cases:

1. C is empty, then $X = \frac{x}{\mathbb{E}\xi}$ is the unique terminal wealth that satisfies both $\mathbb{E}[\xi X] = x$ and $\rho_\Phi(X) = -\frac{x}{\mathbb{E}\xi}$, which is also optimal to (2.8). Because of Proposition 2.4.2, we cannot find this optimal solution by solving the Lagrange dual problem.
2. C is non-empty. From the proof of Proposition 2.10, the set of all feasible quantile functions is given by

$$\mathbb{S} := \{G(\cdot) \in \mathbb{G} \mid \int_{[0,1)} \mathbf{1}_{z \notin C} dG(z) = 0 \text{ and } \int_{[0,1)} F_\xi^{-1}(1-z)G(z)dz = x\}.$$

If C is finite, then it is easy to see that

$$\begin{aligned} \mathbb{S} = \{G(\cdot) \in \mathbb{G} \mid G(z) = a + \sum_{c_i \in C} b_i \mathbf{1}_{c_i \leq z \leq 1}, \quad a \in \mathbb{R}, b_i \geq 0, \\ \text{and } \int_{[0,1)} F_\xi^{-1}(1-z)G(z)dz = x\}. \end{aligned}$$

It is then a finite-dimensional optimization problem to find the optimal solution to (2.8). If the optimal solution exists, then it cannot be given by solving the Lagrange dual problem due to Proposition 2.4.2.

If C is infinite, then the problem is subtler. Let \overline{C} be the closure of C . Define

$$\mathbb{S}(C) := \{G(\cdot) \in \mathbb{G} \mid G(z) = \frac{x - b \int_{(c,1)} F_\xi^{-1}(1-z)dz}{\mathbb{E}\xi} + b \mathbf{1}_{c \leq z \leq 1}, \quad b \geq 0, c \in C\},$$

and let $\overline{\text{conv}}(\mathbb{S}(\overline{C}))$ be the closed convex hull of $\mathbb{S}(\overline{C})$. It is not difficult to show that

$$\mathbb{S}(C) \subset \mathbb{S} \subset \overline{\text{conv}}(\mathbb{S}(\overline{C})).$$

We can search over $\overline{\text{conv}}(\mathbb{S}(\overline{C}))$ to find the optimal solution to (2.8), which remains an interesting topic for future research.

Bibliography

- Acerbi, C. (2002). Spectral measures of risk: a coherent representation of subjective risk aversion. *Journal of Banking & Finance*, 26(7):1505–1518.
- Acerbi, C., Nordio, C., and Sirtori, C. (2001). Expected shortfall as a tool for financial risk management. *arXiv preprint cond-mat/0102304*.
- Acerbi, C. and Simonetti, P. (2002). Portfolio optimization with spectral measures of risk. *preprint*.
- Acerbi, C. and Tasche, D. (2002). On the coherence of expected shortfall. *Journal of Banking & Finance*, 26(7):1487–1503.
- Adam, A., Houkari, M., and Laurent, J.-P. (2008). Spectral risk measures and portfolio selection. *Journal of Banking & Finance*, 32(9):1870–1882.
- Alexander, G. J. and Baptista, A. M. (2002). Economic implications of using a mean-VaR model for portfolio selection: A comparison with mean-variance analysis. *Journal of Economic Dynamics and Control*, 26(7):1159–1193.
- Alexander, G. J. and Baptista, A. M. (2004). A comparison of VaR and CVaR constraints on portfolio selection with the mean-variance model. *Management Science*, 50(9):1261–1273.
- Alexander, G. J. and Baptista, A. M. (2006). Does the Basle Capital Accord reduce bank fragility? An assessment of the value-at-risk approach. *Journal of Monetary Economics*, 53(7):1631–1660.
- Artzner, P., Delbaen, F., Eber, J.-M., and Heath, D. (1999). Coherent measures of risk. *Mathematical Finance*, 9:203–228.
- Asmussen, S., Højgaard, B., and Taksar, M. (2000). Optimal risk control and dividend distribution policies. example of excess-of loss reinsurance for an insurance corporation. *Finance and Stochastics*, 4(3):299–324.

- Basak, S. (1995). A general equilibrium model of portfolio insurance. *Review of Financial Studies*, 8(4):1059–1090.
- Basak, S. (2002). A comparative study of portfolio insurance. *Journal of Economic Dynamics and Control*, 26(7):1217–1241.
- Basak, S. and Chabakauri, G. (2010). Dynamic mean-variance asset allocation. *Review of Financial Studies*, 23(8):2970–3016.
- Basak, S. and Shapiro, A. (2001). Value-at-risk-based risk management: optimal policies and asset prices. *Review of Financial Studies*, 14(2):371–405.
- Basak, S. and Shapiro, A. (2005). A model of credit risk, optimal policies, and asset prices. *The Journal of Business*, 78(4):1215–1266.
- Basak, S., Shapiro, A., and Teplá, L. (2006). Risk management with benchmarking. *Management Science*, 52(4):542–557.
- BCBS (2011). *Revisions to the Basel II market risk framework*. Basel Committee on Banking Supervision, Bank for International Settlements: Basel, Switzerland.
- BCBS (2016). *Standards January 2016. Minimum capital requirements for market risk*. Basel Committee on Banking Supervision, Bank for International Settlements: Basel, Switzerland.
- Benninga, S. and Blume, M. (1985). On the optimality of portfolio insurance. *Journal of Finance*, 40(5):1341–1352.
- Berkowitz, J. and O’Brien, J. (2002). How accurate are value-at-risk models at commercial banks? *Journal of Finance*, 57(3):1093–1111.
- Bernard, C., Chen, J. S., and Vanduffel, S. (2015a). Rationalizing investors choices. *Journal of Mathematical Economics*, 59:10–23.
- Bernard, C., He, X., Yan, J.-A., and Zhou, X. Y. (2015b). Optimal insurance design under rank-dependent expected utility. *Mathematical Finance*, 25(1):154–186.
- Bertsimas, D., Lauprete, G. J., and Samarov, A. (2004). Shortfall as a risk measure: properties, optimization and applications. *Journal of Economic Dynamics and Control*, 28(7):1353–1381.

- Bielecki, T. R., Jin, H., Pliska, S. R., and Zhou, X. Y. (2005). Continuous-time mean-variance portfolio selection with bankruptcy prohibition. *Mathematical Finance*, 15(2):213–244.
- Boonen, T. J., Tan, K. S., and Zhuang, S. C. (2016). Pricing in reinsurance bargaining with comonotonic additive utility functions. *Astin Bulletin*, 46(02):507–530.
- Cahuich, L. D. and Hernández-Hernández, D. (2013). Quantile portfolio optimization under risk measure constraints. *Applied Mathematics & Optimization*, 68(2):157–179.
- Cai, J., Lemieux, C., and Liu, F. (2016). Optimal reinsurance from the perspectives of both an insurer and a reinsurer. *ASTIN Bulletin: The Journal of the IAA*, 46(3):815–849.
- Campbell, R., Huisman, R., and Koedijk, K. (2001). Optimal portfolio selection in a value-at-risk framework. *Journal of Banking & Finance*, 25(9):1789–1804.
- Carrier, G. and Dana, R.-A. (2006). Law invariant concave utility functions and optimization problems with monotonicity and comonotonicity constraints. *Statistics & Decisions*, 24(1/2006):127–152.
- Carrier, G. and Dana, R.-A. (2011). Optimal demand for contingent claims when agents have law invariant utilities. *Mathematical Finance*, 21(2):169–201.
- Cherny, A. S. (2006). Weighted V@R and its properties. *Finance and Stochastics*, 10(3):367–393.
- Cheung, K., Liu, F., and Yam, S. (2012). Average value-at-risk minimizing reinsurance under wang’s premium principle with constraints. *Astin Bulletin*, 42(02):575–600.
- Cheung, K. C. (2010). Optimal reinsurance revisited—a geometric approach. *Astin Bulletin*, 40(01):221–239.
- Chi, Y. and Lin, X. S. (2014). Optimal reinsurance with limited ceded risk: A stochastic dominance approach. *Astin Bulletin*, 44(01):103–126.
- Chi, Y. and Tan, K. S. (2010). Optimal reinsurance under var and cvar risk measures: a simplified approach. *Astin Bulletin*, 41(2):487–509.

- Choulli, T., Taksar, M., and Zhou, X. Y. (2003). A diffusion model for optimal dividend distribution for a company with constraints on risk control. *SIAM Journal on Control and Optimization*, 41(6):1946–1979.
- Cotter, J. and Dowd, K. (2006). Extreme spectral risk measures: an application to futures clearinghouse margin requirements. *Journal of Banking & Finance*, 30(12):3469–3485.
- Cox, J. C. and Huang, C.-f. (1989). Optimal consumption and portfolio policies when asset prices follow a diffusion process. *Journal of Economic Theory*, 49(1):33–83.
- Cox, J. C., Ingersoll Jr, J. E., and Ross, S. A. (1985). An intertemporal general equilibrium model of asset prices. *Econometrica: Journal of the Econometric Society*, pages 363–384.
- Cuoco, D., He, H., and Isaenko, S. (2008). Optimal dynamic trading strategies with risk limits. *Operations Research*, 56(2):358–368.
- Cuoco, D. and Liu, H. (2006). An analysis of VaR-based capital requirements. *Journal of Financial Intermediation*, 15(3):362–394.
- Denneberg, D. (1990). Premium calculation: why standard deviation should be replaced by absolute deviation. *Astin Bulletin*, 20(02):181–190.
- Dhaene, J., Vanduffel, S., Tang, Q., Goovaerts, M. J., Kaas, R., and Vyncke, D. (2004). Capital requirements, risk measures and comonotonicity. *Belgian Actuarial Bulletin*, 4(1):53–61.
- Dickson, D. C. and Waters, H. R. (2006). Optimal dynamic reinsurance. *Astin Bulletin*, 36(02):415–432.
- Dowd, K. (1998). *Beyond value at risk: the new science of risk management*. John Wiley & Sons.
- Dowd, K., Cotter, J., and Sorwar, G. (2008). Spectral risk measures: properties and limitations. *Journal of Financial Services Research*, 34(1):61–75.
- Embrechts, P., Puccetti, G., Rüschendorf, L., Wang, R., and Beleraj, A. (2014). An academic response to Basel 3.5. *Risks*, 2(1):25–48.
- Gabih, A., Grecksch, W., and Wunderlich, R. (2005). Dynamic portfolio optimization with bounded shortfall risks. *Stochastic Analysis and Applications*, 23(3):579–594.

- Gabih, A., Sass, J., and Wunderlich, R. (2009). Utility maximization under bounded expected loss. *Stochastic Models*, 25(3):375–407.
- Grandell, J. (2012). *Aspects of risk theory*. Springer Science & Business Media.
- Grossman, S. J. and Vila, J.-L. (1989). Portfolio insurance in complete markets: A note. *Journal of Business*, 62(4):473–476.
- Grossman, S. J. and Zhou, Z. (1996). Equilibrium analysis of portfolio insurance. *Journal of Finance*, 51(4):1379–1403.
- Gundel, A. and Weber, S. (2008). Utility maximization under a shortfall risk constraint. *Journal of Mathematical Economics*, 44(11):1126–1151.
- He, L. and Liang, Z. (2009). Optimal financing and dividend control of the insurance company with fixed and proportional transaction costs. *Insurance: Mathematics and Economics*, 44(1):88–94.
- He, X. D., Jin, H., and Zhou, X. Y. (2015). Dynamic portfolio choice when risk is measured by Weighted VaR. *Mathematics of Operations Research*, 40(3):773–796.
- He, X. D. and Zhou, X. Y. (2011). Portfolio choice via quantiles. *Mathematical Finance*, 21(2):203–231.
- He, X. D. and Zhou, X. Y. (2016). Hope, fear, and aspirations. *Mathematical Finance*, 26(1):3–50.
- Hipp, C. and Taksar, M. (2010). Optimal non-proportional reinsurance control. *Insurance: Mathematics and Economics*, 47(2):246–254.
- Hipp, C. and Vogt, M. (2003). Optimal dynamic xl reinsurance. *Astin Bulletin*, 33(02):193–207.
- Jgaard, B. H. and Taksar, M. (1999). Controlling risk exposure and dividends payout schemes: insurance company example. *Mathematical Finance*, 9(2):153–182.
- Jin, H., Xu, Z. Q., and Zhou, X. Y. (2008). A convex stochastic optimization problem arising from portfolio selection. *Mathematical Finance*, 18(1):171–183.
- Jin, H., Yan, J.-A., and Zhou, X. Y. (2005). Continuous-time mean-risk portfolio selection. *Annales de l'IHP Probabilités et statistiques*, 41(3):559–580.

- Jin, H. and Zhou, X. Y. (2008). Behavioral portfolio selection in continuous time. *Mathematical Finance*, 18(3):385–426.
- Jones, D. (2000). Emerging problems with the Basel Capital Accord: Regulatory capital arbitrage and related issues. *Journal of Banking & Finance*, 24(1):35–58.
- Jorion, P. (1997). *Value at risk: the new benchmark for controlling market risk*. Irwin Professional Pub.
- Jorion, P. (2002). How informative are value-at-risk disclosures? *The Accounting Review*, 77(4):911–931.
- Karatzas, I. and Shreve, S. E. (1998). *Methods of mathematical finance*, volume 39. Springer Science & Business Media.
- Kast, R., Luciano, E., and Peccati, L. (1999). Value-at-risk as a decision criterion. In *EURO Working Group on Financial Modeling Meeting, Paris (May 1998)*.
- Kramkov, D. and Schachermayer, W. (1999). The asymptotic elasticity of utility functions and optimal investment in incomplete markets. *Annals of Applied Probability*, 9(3):904–950.
- Kupper, M. and Schachermayer, W. (2009). Representation results for law invariant time consistent functions. *Mathematics and Financial Economics*, 2(3):189–210.
- Leippold, M., Trojani, F., and Vanini, P. (2006). Equilibrium impact of value-at-risk regulation. *Journal of Economic Dynamics and Control*, 30(8):1277–1313.
- Leitner, J. (2005). A short note on second-order stochastic dominance preserving coherent risk measures. *Mathematical Finance*, 15(4):649–651.
- Leland, H. E. (1980). Who should buy portfolio insurance? *Journal of Finance*, 35(2):581–594.
- Luo, S., Taksar, M., and Tsoi, A. (2008). On reinsurance and investment for large insurance portfolios. *Insurance: Mathematics and Economics*, 42(1):434–444.
- Luskin, D. L. (1988). *Portfolio Insurance: a guide to dynamic hedging*. Wiley.
- Markowitz, H. (1952). Portfolio selection. *Journal of Finance*, 7(1):77–91.

- Meng, H. and Siu, T. K. (2011). Optimal mixed impulse-equity insurance control problem with reinsurance. *SIAM Journal on Control and Optimization*, 49(1):254–279.
- Meng, H. and Zhang, X. (2010). Optimal risk control for the excess of loss reinsurance policies. *Astin Bulletin*, 40(01):179–197.
- Milgrom, P. and Segal, I. (2002). Envelope theorems for arbitrary choice sets. *Econometrica*, 70(2):583–601.
- Pliska, S. R. (1986). A stochastic calculus model of continuous trading: optimal portfolios. *Mathematics of Operations Research*, 11(2):371–382.
- Promislow, D. S. and Young, V. R. (2005). Minimizing the probability of ruin when claims follow brownian motion with drift. *North American Actuarial Journal*, 9(3):110–128.
- Rockafellar, R. T. and Uryasev, S. (2000). Optimization of conditional value-at-risk. *Journal of Risk*, 2:21–42.
- Rockafellar, R. T. and Uryasev, S. (2002). Conditional value-at-risk for general loss distributions. *Journal of Banking & Finance*, 26(7):1443–1471.
- Rogers, L. C. G. (2009). Optimal and robust contracts for a risk-constrained principal. *Mathematics and Financial Economics*, 2(3):151–171.
- Saunders, A. (2000). *Financial institutions management: a modern perspective*. McGraw-Hill College.
- Schied, A. (2004). On the Neyman–Pearson problem for law-invariant risk measures and robust utility functionals. *Annals of Applied Probability*, 14(3):1398–1423.
- Schmidli, H. (2001). Optimal proportional reinsurance policies in a dynamic setting. *Scandinavian Actuarial Journal*, 2001(1):55–68.
- Schmidli, H. (2002). On minimizing the ruin probability by investment and reinsurance. *Annals of Applied Probability*, pages 890–907.
- SEC (1997). Disclosure of accounting policies for derivative financial instruments and derivative commodity instruments and disclosure of quantitative and qualitative information about market risk inherent in derivative financial instruments, other

- financial instruments, and derivative commodity instruments. *Other Financial Instruments, and Derivative Commodity Instruments, SEC, Washington, DC.*
- Sereda, E. N., Bronshtein, E. M., Rachev, S. T., Fabozzi, F. J., Sun, W., and Stoyanov, S. V. (2010). Distortion risk measures in portfolio optimization. In *Handbook of Portfolio Construction*, pages 649–673. Springer.
- Taksar, M. I. and Markussen, C. (2003). Optimal dynamic reinsurance policies for large insurance portfolios. *Finance and Stochastics*, 7(1):97–121.
- Taksar, M. I. and Zhou, X. Y. (1998). Optimal risk and dividend control for a company with a debt liability. *Insurance: Mathematics and Economics*, 22(1):105–122.
- Wang, S. (1995). Insurance pricing and increased limits ratemaking by proportional hazards transforms. *Insurance: Mathematics and Economics*, 17(1):43–54.
- Wang, S. (1996). Premium calculation by transforming the layer premium density. *Astin Bulletin*, 26(01):71–92.
- Wang, S. S. (2000). A class of distortion operators for pricing financial and insurance risks. *Journal of Risk and Insurance*, 67(1):15–36.
- Wang, S. S., Young, V. R., and Panjer, H. H. (1997). Axiomatic characterization of insurance prices. *Insurance: Mathematics and economics*, 21(2):173–183.
- Wei, P. (2017a). Equilibrium analysis of expected shortfall. Available at SSRN: <https://ssrn.com/abstract=2956970>.
- Wei, P. (2017b). Risk management with weighted VaR. *Mathematical Finance*, forthcoming.
- Weng, C. and Zhuang, S. C. (2016). Cdf formulation for solving an optimal reinsurance problem. *Scandinavian Actuarial Journal*, forthcoming.
- Wirch, J. L. and Hardy, M. R. (1999). A synthesis of risk measures for capital adequacy. *Insurance: mathematics and Economics*, 25(3):337–347.
- Xia, J. and Zhou, X. Y. (2016). Arrow–Debreu equilibria for rank-dependent utilities. *Mathematical Finance*, 26(3):558–588.
- Xu, Z. Q. (2014). A new characterization of comonotonicity and its application in behavioral finance. *Journal of Mathematical Analysis and Applications*, 418(2):612–625.

- Xu, Z. Q. (2016). A note on the quantile formulation. *Mathematical Finance*, 26(3):589–601.
- Xu, Z. Q., Zhou, X. Y., and Zhuang, S. C. (2015). Optimal insurance with rank-dependent utility and increasing indemnities.
- Yaari, M. E. (1987). The dual theory of choice under risk. *Econometrica*, 55(1):95–115.
- Yiu, K.-F. C. (2004). Optimal portfolios under a value-at-risk constraint. *Journal of Economic Dynamics and Control*, 28(7):1317–1334.
- Young, V. R. (2004). Premium principles. *Encyclopedia of actuarial science*.
- Zhang, N., Jin, Z., Li, S., and Chen, P. (2016). Optimal reinsurance under dynamic VaR constraint. *Insurance: Mathematics and Economics*, 71:232–243.