Optimization, Model Uncertainty, and Testing in Risk and Insurance

by

Zhanyi Jiao

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Doctor of Philosophy
in
Actuarial Science

Waterloo, Ontario, Canada, 2024

Examining Committee Membership

The following served on the Examining Committee for this thesis. The decision of the Examining Committee is by majority vote.

External Examiner:	Steven Vanduffel Professor, Solvay Business School, Vrije Universiteit Brussel
Supervisor:	Jun Cai Professor, Department of Statistics and Actuarial Science, University of Waterloo
Supervisor:	Ruodu Wang Professor, Department of Statistics and Actuarial Science, University of Waterloo
Internal Member:	Mario Ghossoub Associate Professor, Dept. of Statistics and Actuarial Science, University of Waterloo
Internal Member:	Fangda Liu Associate Professor, Dept. of Statistics and Actuarial Science, University of Waterloo
Internal-External Member:	Qi-Ming He Professor, Department of Management Science and Engineering, University of Waterloo

Author's Declaration

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

This thesis focuses on three important topics in quantitative risk management and actuarial science: risk optimization, risk sharing, and statistical hypothesis testing in risk.

For the risk optimization, we concentrate on risk optimization under model uncertainty where only partial information about the underlying distribution is available. One key highlight, detailed in Chapter 2, is the development of a novel formula named the reverse Expected Shortfall (ES) optimization formula. This formula is derived to better facilitate the calculation of the worst-case mean excess loss under two commonly used model uncertainty sets – moment-based and distance-based (Wasserstein) uncertainty sets. Further exploration reveals that the reverse ES optimization formula is closely related to the Fenchel-Legendre transforms, and our formulas are generalized from ES to optimized certainty equivalents, a popular class of convex risk measures. Chapter 3 considers a different approach to derive the closed-form worst-case target semi-variance by including distributional shape information, crucial for finance (symmetry) and insurance (non-negativity) applications. We demonstrate that all results are applicable to robust portfolio selection, where the closed-form formulas greatly simplify the calculations for optimal robust portfolio selections, either through explicit forms or via easily solvable optimization problems.

Risk sharing focuses on the redistribution of total risk among agents in a specific way. In contrast to the traditional risk sharing rules, Chapter 4 introduces a new risk sharing framework - anonymized risk sharing, which requires no information on preferences, identities, private operations, and realized losses from the individual agents. We establish an axiomatic theory based on four axioms of fairness and anonymity within the context of anonymized risk sharing. The development of this theory provides a solid foundation for further explorations on decentralized and digital economy including peer-to-peer (P2P) insurance, revenue sharing of digital contents and blockchain mining pools.

Hypothesis testing plays a vital role not only in statistical inference but also in risk management, particularly in the backtesting of risk measures. In Chapter 5, we address the problem of testing conditional mean and conditional variance for non-stationary data using the recent emerging concept of e-statistics. We build e-values and p-values for four

types of non-parametric composite hypotheses with specified mean and variance as well as other conditions on the shape of the data-generating distribution. These shape conditions include symmetry, unimodality, and their combination. Using the obtained e-values and p-values, we construct tests via e-processes, also known as testing by betting, as well as some tests based on combining p-values for comparison. To demonstrate the practical application of these methodologies, empirical studies using financial data are conducted under several settings.

Acknowledgments

First, I would like to express my sincere gratitude to my Ph.D. supervisors Dr. Jun Cai and Dr. Ruodu Wang. Dr. Cai is the kindest and most supportive professor I have ever met, always encouraging me without ever making me feel pressured, and his dedication to academia have greatly inspired me. Dr. Wang is the smartest and most knowledgeable, caring person I have ever met. He has provided invaluable insights into both academia and professional for me. His rigorous approach, meticulousness, and sincerity have shown me what it truly means to be a great scholar. I am always grateful and feel lucky that they were willing to take me on as a student four years ago, a decision that has changed my life. Without their guidance, I could never have experienced the wonderful journey I always dreamed of.

My sincere thanks also goes to the members of my thesis committee, Dr. Mario Ghossoub, Dr. Qi-Ming He, Dr. Fangda Liu and Dr. Steven Vanduffel for their valuable time and insightful comments.

I would like to thank Dr. Yuanying (Michelle) Guan, Dr. Steven Kou, Dr. Yang Liu, and Yixuan Fan for their insightful discussions on research. Special thanks go to Dr. Tiantian Mao and Dr. Qiuqi Wang for their supportive guidance on research, valuable advice for my academic career, and constant encouragement. I am also grateful to Dr. Lisa Gao, Dr. David Landriault, and my best academic brothers and sisters Dr. Yuyu Chen, Dr. Haiyan Liu and Dr. Yunran Wei for their valuable advice and encouragement during my job search. I would like to thank all the friends I met at the University of Waterloo, especially Meixi Chen, Liyuan Lin, Zhiqiao Song, Dr. Qinyu Wu, Dr. Mingren Yin, Jiayue Zhang, Yuanyuan Zhang, and Michael (Boyuan) Zhu, among others, for the unforgettable memories we shared. Additionally, I extend my gratitude to Ms. Mary Lou Dufton and Mr. Greg Preston for their indispensable supports with administrative issues during my Ph.D. studies.

On a personal note, I would like to thank my parents: Shuguang Jiao and Fen Liu, whose love and guidance are with me in whatever I pursue. They are the backbone of my success and gave me unconditional support throughout my life journey. Without their encouragement and understanding, it would have been impossible to complete this work.

Dedication

Dedicated to my family.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Risk measure

Risk is often described as "the potential for adverse consequences," but this definition only captures part of the risk. Risk is also opportunity particularly when it is effectively managed and exploited. As Peter Drucker, one of the most influential risk management thinkers, famously stated, "If you cannot measure the risk, you cannot manage the risk." This highlights that measuring risk is an essential aspect of risk management, serving as the foundational step towards understanding and controlling the potential outcomes, both negative and positive, associated with risk-taking activities.

Value-at-Risk (VaR) and Expected Shortfall (ES) (also known as Conditional Value-at-Risk (CVaR), Tail Value-at-Risk (TVaR) and Average Value-at-Risk (AVaR)) are the two most important risk measures in banking and insurance, and they are widely employed in regulatory capital computation, decision making, performance analysis, and risk management. In particular, ES is the standard risk measure in the current Basel Accords (see BCBS (2019)) as well as the Swiss Solvency Test, and VaR is standard in the insurance regulatory framework of Solvency II. We assume $\mathcal{X}$ to be the set of integrable random variables in a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. For $X \in \mathcal{X}$ with a probability level $\alpha \in [0, 1]$,[1]

[1]Throughout this thesis, positive values of the random variable X represent losses, while negative

VaR is defined as

$$\mathrm{VaR}_\alpha(X) = \inf\{x \in \mathbb{R} : \mathbb{P}(X \leqslant x) \geqslant \alpha\}. \tag{1.1}$$

ES is generally calculated as the average of the tail values that are worse than VaR at probability level α and is defined as

$$\mathrm{ES}_\alpha(X) = \frac{1}{1-\alpha}\int_\alpha^1 \mathrm{VaR}_\beta(X)\mathrm{d}\beta. \tag{1.2}$$

There has been extensive research on VaR, ES and their properties and applications in quantitative risk management. For comprehensive discussions and analyses of these two common risk measures, we refer McNeil et al. (2015) and Föllmer and Schied (2016), which provide detailed insights into their significance in the field.

Another important and commonly used risk measure in finance and actuarial science is the *partial moment*, also known as the *one-sided moment*. Partial moments are measures of risk that focus on either the lower or upper part of a distribution. Since, in this thesis, X represents the random loss, we are mainly interested in the upper partial moment, which measures the downside risk over a given threshold t. It is defined as follows:

$$\mathbb{E}[(X-t)_+^k] = \int_t^\infty (x-t)^k \mathrm{d}F(x). \tag{1.3}$$

where F is the distribution of X, and k represents the order of the upper partial moment. The first-order upper partial moment $\mathbb{E}[(X-t)_+]$ measures the expected loss above a threshold t. This measure is also referred to as the *expected regret* or *target shortfall* in finance (see, e.g., Testuri and Uryasev (2004)), and as the *mean excess function* in insurance, which is particularly useful in stop-loss insurance and reinsurance, where policies take effect after claims exceed a certain threshold (see, e.g., De Vylder and Goovaerts (1982)).[2] We provide a deeper insight of mean excess function and its application in insurance in Chapter 2.

The second-order upper partial moment $\mathbb{E}[(X-t)_+^2]$, also known as the *target semi-variance* (see, e.g., Rohatgi (2011)), quantifies the dispersion of all values that exceed a

values represent gains, unless otherwise stated.

[2]In this thesis, we distinguish between the term "mean excess function" as a function of t and "mean excess loss" as a function of X. More detailed will be discussed in the Section 2.3 of Chapter 2.

target t. When $t = \mathbb{E}[X]$, the target semi-variance reduces to the well-known semi-variance. The target semi-variance is particularly useful for assessing downside risk, serving as a suitable alternative to variance in the traditional mean-variance portfolio selection problem, where investors and risk managers are more concerned about the downside fluctuations. More details about the target semi-variance will be discussed in Chapter 3.

There is an interesting connection between VaR, ES and the first-order upper partial moment (mean excess function) first derived by Rockafellar and Uryasev (2000). For $\alpha \in (0,1)$ and $X \in \mathcal{X}$, we have

$$\mathrm{ES}_\alpha(X) = \mathrm{VaR}_\alpha(X) + \frac{1}{1-\alpha}\mathbb{E}[(X - \mathrm{VaR}_\alpha(X))_+]. \tag{1.4}$$

This formula indicates that ES at a fixed probability level can be calculated as a linear combination of VaR and a scaled mean excess function, with VaR serving as the threshold.[3] This relationship is foundational in risk management and financial engineering, particularly useful in ES optimization problems. For a comprehensive review, see Rockafellar and Uryasev (2013) and references therein. In Chapter 2, we derive another intriguing connection between the mean excess function and ES, where the mean excess function can be expressed through an optimization formula of ES.

$$\mathbb{E}[(X-t)_+] = \max_{\alpha\in[0,1]} \left\{(1-\alpha)\left(\mathrm{ES}_\alpha(X) - t\right)\right\}. \tag{1.5}$$

Furthermore, the above newly derived formula can be generalized from ES to optimized certainty equivalents (OCE), a widely used class of convex risk measures. In Chapter 2, we also show that the new formula (1.5) is particularly useful for risk optimization under model uncertainty.

1.2 Risk optimization under model uncertainty

A good and appropriate risk measure facilitates better risk management and decision-making processes. However, a risk measure alone is merely a tool for understanding and

[3] A more rigorous formula is provided in Section 2.1 of Chapter 2.

managing risk. Optimal decision-making arises when risk-taking is strategically optimized, enabling businesses and individuals to address emerging marketplace challenges effectively and enhance financial security. Risk optimization based on VaR, ES and upper partial moments is extensively studied in the contexts of optimal insurance/reinsurance and portfolio selection problems. In the insurance and reinsurance context, the focus primarily lies on minimizing the target risk measure of a loss function across a set of possible contracts (see e.g. Cai and Tan (2007), Chi and Tan (2011), Cai and Chi (2020), Ghossoub et al. (2023) and the references therein). In the portfolio selection context, the goal is to minimize the target risk measure of a loss function over a set of possible portfolio positions (see e.g., Rockafellar and Uryasev (2000, 2002), Alexander and Baptista (2002, 2004) and the references therein).

Most classical optimal insurance and portfolio selection models assume that the loss distribution is completely known. In reality, however, the loss distribution is estimated from historical data, expert opinions, past experiences, and prior beliefs. This leads to a gap between the true distribution and the estimated distribution due to insufficient data, prediction errors, or incorrect judgments. Considering such gap can be detrimental to the risk optimization and decision-making process, it drives us to consider model uncertainty where we only know the partial information of this underlying distribution and consider the worst-case scenario. The risk optimization over the uncertainty set $\mathcal{L}$ can be generally formulated as follows

$$\sup_{F\in\mathcal{L}} \rho^F(f(X)), \tag{1.6}$$

where $\rho : \mathcal{X} \to \mathbb{R}$ is risk measure, which can be VaR, ES or partial moments defined in Section 1.1, and f is a loss/cost function defined on $\mathbb{R}$. This modeling framework is commonly known as *distributionally robust optimization* (DRO). It has been widely applied in decision-making theory since the pioneering work of Ben-Tal and Nemirovski (1998). For applications in portfolio selection, refer to Ghaoui et al. (2003), Natarajan et al. (2008), Zhu and Fukushima (2009), Chen et al. (2011), and Blanchet et al. (2021). For applications in insurance and reinsurance, see Pflug et al. (2017), Liu and Mao (2022), Bernard et al. (2024) and Cai et al. (2024).

In a DRO problem, the uncertainty set offers a flexible approach for capturing model

uncertainty using information from various sources, such as historical data and expert insights. This information may include descriptive statistics, distances from a reference distribution, and structural properties. Specifically, we categorize commonly used uncertainty sets into the following types:

1. **Moment-based uncertainty set:** This set contains distributions characterized by their moments, such as the mean and variance, or other higher order moments.

2. **Distance-based uncertainty set:** This set contains distributions that are close to a reference distribution in terms of some discrepancy measures.

3. **Shape-preserving uncertainty set:** This set contains distributions with similar structural properties, such as symmetry, unimodality, non-negativity, or specific constraints.

There are other interesting uncertainty sets that combine the types of uncertainty sets mentioned above. For example, Bernard et al. (2023) and Cai et al. (2024) consider uncertainty sets that include both moment and distance information. It is also common to combine moment information with structural information, see e.g., Li et al. (2018) and Bernard et al. (2020). For more details on these uncertainty sets and the relevant literature, refer to Rahimian and Mehrotra (2022) for a comprehensive review of DRO problems. In this thesis, all the aforementioned types of uncertainty sets are discussed. In Chapter 2, we analyze the worst-case values of the mean excess function using the newly derived optimization formula

$$\sup_{F\in\mathcal{L}} \mathbb{E}^F[(f(X)-t)_+] = \max_{\alpha\in[0,1]} \left\{ (1-\alpha) \left(\sup_{F\in\mathcal{L}} \mathrm{ES}_\alpha(f(X)) - t \right) \right\}, \tag{1.7}$$

which provides a new methodology to calculate the worst-case mean excess function under two popular settings of model uncertainty: moment-based and distance-based uncertainty sets (primarily Wasserstein uncertainty set), since there are existing results on worst-case ES. In addition, its usefulness further demonstrated through an application using catastrophic insurance datasets, which is particularly appropriate for model uncertainty due to the data scarcity issue inherent in such datasets.

In Chapter 3, we focus on the worst-case target semi-variance under the shape-preserving uncertainty set. We derive derive the explicit and closed-form expressions for the worst-case target semi-variance when only the mean and variance of a loss are known and the loss is symmetric or non-negative. All the results are applicable to robust portfolio selection, where we can project multivariate uncertainty sets into one-dimensional uncertainty sets. This projection enables the use of previously derived closed-form formulas, simplifying the calculation of optimal portfolio selections through either explicit forms or straightforward and easy solvable optimization problems.

1.3 Risk sharing

Risk optimization assists decision makers in strategically optimizing their risk-taking behaviors to balance risk and return. In addition to managing risk optimally, it is important to understand how to appropriately reduce or share risk, leading to another crucial topic in risk management – risk sharing.

Risk sharing refers to pooling risks from several participants in a group and reallocating the total risk based on a specific way. Traditional insurance is the most common form of risk sharing, where the insurer collects the payments from the insureds and provides insurance coverage to all insureds. A risk sharing scheme also arises in other different forms, such as tontines, taxation, founders stock, investment profit sharing, and Bitcoin mining pools, to name a few. In these contexts, either wealth or losses, or both of them, may be shared among participants.

Specifically, the participants of a risk sharing scheme, such as individual investors, co-workers, financial institutions, policyholders and an insurer, peer-to-peer (P2P) insureds, and miners in a Bitcoin mining pool, are generally referred to as *agents*. Assume n agents have initial risk contributions $X_1, X_2, \dots, X_n \in \mathcal{X}$ (we denote the initial risk vector as $\boldsymbol{X} = (X_1, X_2, \dots, X_n)$), and it will be exchanged to a new position after a risk sharing scheme $\boldsymbol{A} : \mathcal{X}^n \to \mathcal{X}^n$:[4]

$$\boldsymbol{A}^{\boldsymbol{X}} = (A_1^{\boldsymbol{X}}, A_2^{\boldsymbol{X}}, \dots, A_n^{\boldsymbol{X}}), \tag{1.8}$$

[4]In Chapter 4, we will provide a more formal definition of risk sharing and some examples.

and satisfying the full-allocation property:

$$\sum_{i=1}^{n} A_i^{\boldsymbol{X}} = \sum_{i=1}^{n} X_i. \tag{1.9}$$

There are generally two types of setting in the risk sharing, including collaborative and competitive risk sharing. The key problem is to derive an equilibrium risk sharing rule. In the collaborative setting, we assume a central planner who knows preferences of all agents and solves the Pareto equilibrium. In the competitive setting, we require a trading mechanism (e.g., a market) and individual preferences such that each agent optimize their objectives individually. In both situations, concrete preference models are required. The common preference models include expected utility, mean-variance, dual utility, rank-dependent utility, cumulative prospect theory, risk measures (e.g., quantiles), robust/variational preferences, etc. Equilibrium risk sharing is studied in the classic work of Arrow and Debreu (1954) and Borch (1962) among a very rich literature, see the later work on risk sharing by Barrieu and El Karoui (2005) for convex risk measures, Carlier et al. (2012) for multivatiate stochastic dominance, Xia and Zhou (2016) for rank-dependent utilities, Cai et al. (2017) for reinsurance arrangements, and Embrechts et al. (2018) for quantile-based risk measures.

However, in practice, several challenges hinder the application of the preference framework to risk sharing problems. First, preferences are difficult to elicit or test, as most agents find it challenging to accurately articulate their preferences. Second, the allocation for one agent might depend on the preferences of other agents, which complicates the process. Additionally, providing false preferences might be advantageous, leading to potential manipulation.

These limitations drive us to develop a new risk sharing framework, called *anonymized risk sharing*, detailed in Chapter 4. The key feature of such framework is that agents do not need to disclose their preferences, identity, or wealth level, which fits perfectly with practical situations and the concept of decentralization. More precisely, the allocation to an agent is determined by the initial risk contributions of all agents, but not the specification of these agents. For this reason, anonymized risk sharing schemes are desirable in several

decentralized applications such as P2P insurance, Bitcoin mining pools and tontines. To better understand a suitable anonymized risk sharing rule, we put forward four simple and natural axioms: actuarial fairness, risk fairness, risk anonymity, and operational anonymity. We establish the remarkable fact that the four axioms characterize the conditional mean risk sharing rule, revealing the unique and prominent role of this popular risk sharing rule among all others within the decentralized system. Our characterization hence provides the first axiomatic foundation for conditional mean risk sharing and its applications in economic theory and decentralized finance and insurance. On the technical side, our work also lays a foundation for future research on the characterization of multi-dimensional and random-vector-valued objectives, such as risk sharing rule $\boldsymbol{A}$ in our work.

1.4 Hypothesis testing in risk management

Hypothesis testing plays a crucial role in risk management by enabling organizations to make informed decisions based on statistical evidence. By setting up hypotheses and testing them against real-world data, risk managers can assess the likelihood of various risk scenarios and evaluate their potential impact. In this thesis, we focus on testing the mean and variance in a sequential setting, which is common in financial practice. In the context of forecasting and financial risk assessment, hypothesis testing is vital for evaluating the mean and variance of financial metrics or predictive models. Testing the mean helps determine whether the expected return or outcome aligns with strategic goals or market benchmarks, while testing the variance assesses the stability and risk associated with these returns or outcomes. This process enables financial analysts and risk managers to validate or refute assumptions, ensuring that investment strategies or forecasts are robust and reliable.

Specifically, in this thesis, we are interested in testing the following null hypothesis H, where there are sequentially arriving data points $X_1, X_2, \ldots$, each drawn from unknown distributions:

$$H : \mathbb{E}[X_i|\mathcal{F}_{i-1}] \leqslant \mu_i \text{ and } \mathrm{Var}(X_i|\mathcal{F}_{i-1}) \leqslant \sigma_i^2, \text{ for each } i, \tag{1.10}$$

where $\mathcal{F}_i$ is generated by observations before X_i, and μ_i and σ_i can be data-dependent on

past observations. The main issue here is that the data points are not independent and identically distributed (iid), making it challenging to infer the distributions.

In Chapter 5, we address (1.10) using the emerging concept of e-tests or tests by betting (see e.g., Shafer (2021) and Vovk and Wang (2021)). E-tests offer several advantages over classic p-tests (which are based on p-values) when it comes to sequential inference, robustness to model misspecification, and dependence, among other benefits. For a detailed discussion of the advantages of e-tests over p-tests, see Wang and Ramdas (2022, Section 2). We construct e-values (and p-values) from a random variable associated with each data point for the null hypothesis. After obtaining these e-values, we typically combine them by forming an e-process to test the overall hypothesis. Alternatively, we also combine p-values using Fisher's method (Fisher (1925)) and Simes' method (Simes (1986)), but the power of these methods is generally weaker than that of e-combining methods.

Another crucial topic in hypothesis testing within risk management is backtesting risk measures. Backtesting risk measures is vital for financial regulators to evaluate the risk forecasts reported by financial institutions. For backtesting purposes, we denote the model space $\mathcal{M}$ as a set of distributions on $\mathbb{R}$. We define the functional $\psi = (\rho, \phi_1, \dots, \phi_{d-1})$: $\mathcal{M} \to \mathbb{R}^d$ as the collection of available statistical information. Here, ρ is the risk measure to be tested, and ϕ is the auxiliary statistics containing distributional information of the underlying random losses. Let $T \in \mathbb{N}$ be a finite time horizon, and $X_1, X_2, \dots, X_T$ be losses arriving sequentially up to time T. The risk measures $\rho(X_i|\mathcal{F}_{i-1})$ and $\phi(X_i|\mathcal{F}_{i-1})$ are predicted by $R_i : \Omega \to \mathbb{R}$ and $Z_i : \Omega \to \mathbb{R}^{d-1}$, respectively, which are $\mathcal{F}_{i-1}$-measurable for $i \in \{1, 2, \dots T\}$. Depending on different practical situations, the financial institution and the regulator may care about either the regulatory risk measure ρ or the statistic ϕ. For standard backtests, regulators are interested in the one-sided test for the following null hypothesis:

$$H : \ R_i \geqslant \rho(X_i|\mathcal{F}_{i-1}) \text{ and } Z_i \in \phi(X_i|\mathcal{F}_{i-1}), \quad \text{for each } i \in \{1, 2, \dots T\}. \tag{1.11}$$

This problem is also compatible with using e-tests due to the non-iid nature of the observations, the sequential settings, and the challenge of inferring the loss distributions. The recent work of Wang et al. (2022) discusses detailed model-free procedures for backtesting

VaR and ES using e-values and e-processes. Clearly, testing mean and variance in Chapter 5 can also be incorporated in such a framework by assuming the risk measure ρ to be variance and the auxiliary statistic ϕ to be the expectation with one-dimensional setting ($d = 2$).

Given that e-tests are still relatively new, there are many intriguing backtesting risk measure problems based on e-tests yet to be solved. This thesis proposes two potential future research directions and open questions regarding comparative e-backtests (discussed in Section 6.2.2 of Chapter 6) and e-backtests of Gini indices (discussed in Section 6.2.3 of Chapter 6).

To maintain each chapter's content as self-contained, important concepts such as different risk measures, model uncertainty sets, and risk-sharing rules, and some notations will be reintroduced in the following chapters, potentially with slightly different conventions.

Chapter 2

A reverse ES (CVaR) optimization formula

2.1 Introduction

An influential result on VaR and ES is an optimization formula obtained by Rockafellar and Uryasev (2000, 2002), which is the main motivation for this chapter. We officially give the definition of VaR and ES. Let $\mathcal{X}$ be the set of integrable random variables in a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. At a probability level $\alpha \in [0,1]$, VaR has two versions as the left- and right-quantiles. For $X \in \mathcal{X}$, define

$$\begin{aligned} \mathrm{VaR}_\alpha^-(X) &= \inf\{t \in \mathbb{R} : \mathbb{P}(X \leqslant t) \geqslant \alpha\}; \\ \mathrm{VaR}_\alpha^+(X) &= \inf\{t \in \mathbb{R} : \mathbb{P}(X \leqslant t) > \alpha\}. \end{aligned} \tag{2.1}$$

By definition, $\mathrm{VaR}_0^-(X) = -\infty$ and $\mathrm{VaR}_1^+(X) = \infty$. ES at probability level $\alpha \in [0,1]$ is defined as

$$\mathrm{ES}_\alpha(X) = \frac{1}{1-\alpha}\int_\alpha^1 \mathrm{VaR}_\beta^-(X)\mathrm{d}\beta, \quad X \in \mathcal{X},\ \alpha \in [0,1), \tag{2.2}$$

and $\mathrm{ES}_1(X) = \mathrm{VaR}_1^-(X)$. It is well known that an ES is a coherent risk measure (Artzner et al. (1999)) and a convex risk measure (Föllmer and Schied (2016)), and it admits an

axiomatization based on portfolio diversification (Wang and Zitikis (2021)). Below, we formally present the celebrated formula of Rockafellar and Uryasev (2002).

Theorem 2.1.1 (ES optimization formula)**.** *For $X \in \mathcal{X}$ and $\alpha \in (0,1)$, it holds*

$$\mathrm{ES}_\alpha(X) = \min_{t\in\mathbb{R}} \left\{ t + \frac{1}{1-\alpha}\mathbb{E}[(X-t)_+] \right\}, \tag{2.3}$$

and the set of minimizers for (2.3) is $[\mathrm{VaR}_\alpha^-(X), \mathrm{VaR}_\alpha^+(X)]$.

Theorem 2.1.1 has been a cornerstone of risk management and financial engineering since Rockafellar and Uryasev (2000, 2002) and Pflug (2000). This result has been tremendously useful in the optimization of ES (see Rockafellar and Uryasev (2013) for a review) and it has also been widely taught in actuarial science, see e.g., Denuit et al. (2005, Section 2.4.3) and Kaas et al. (2008, Section 5.6). Among other implications, this formula directly gives an elementary proof of subadditivity of ES; see Embrechts and Wang (2015) for a comparison with six other proofs.

In this chapter, we establish a new optimization formula based on ES, which can be seen as a reverse formula to Theorem 2.1.1. This formula reveals nice symmetries between the ES curve and the mean excess function, as we discuss in Section 2.3. The mean excess loss, also known as the stop-loss premium, has a deep root in actuarial science (e.g., De Vylder and Goovaerts (1982)). In Section 2.4, we apply the new formula to two popular settings of model uncertainty, one induced by information of mean and a higher moment and the other induced by a Wasserstein ball. In both settings, the worst-case ES admits an explicit formula in the recent literature (Pesenti et al. (2020); Liu et al. (2022)) whereas the worst-case mean excess function does not. Two insurance loss datasets are studied in Section 2.5 to illustrate the obtained results on the mean excess loss under model uncertainty induced by a Wasserstein ball. We present a few further technical results in Sections 2.6 and 2.7; more precisely, the reverse ES optimization formula are generalized to the class of optimized certainty equivalents introduced by Ben-Tal and Teboulle (2007) (Section 2.6), and two related formulas are obtained via Fenchel-Legendre transforms (Section 2.7). Section 2.8 concludes this chapter.

2.2 A reverse ES optimization formula

We start from the observation from Theorem 2.1.1 that $\mathrm{ES}_\alpha(X)$ for a fixed $\alpha \in (0,1)$ can be obtained through taking the minimum of a function involving $\mathbb{E}[(X-t)_+]$ over $t \in \mathbb{R}$. Having a mathematical symmetry in mind, a natural question is whether $\mathbb{E}[(X-t)_+]$ for a fixed $t \in \mathbb{R}$ can be obtained through taking the maximum of a function involving $\mathrm{ES}_\alpha(X)$ over $\alpha \in [0,1]$. This leads to the reverse ES optimization formula, the main result of this chapter. In what follows, we always use the convention $0 \times x = 0$ for $x \in [-\infty, \infty]$.

Theorem 2.2.1 (Reverse ES optimization formula)**.** *For $X \in \mathcal{X}$ and $t \in \mathbb{R}$, it holds*

$$\mathbb{E}[(X-t)_+] = \max_{\alpha \in [0,1]} \left\{ (1-\alpha)\left(\mathrm{ES}_\alpha(X) - t\right) \right\}, \tag{2.4}$$

and the set of maximizers for (2.4) is $[\mathbb{P}(X<t), \mathbb{P}(X \leqslant t)]$.

To prove Theorem 2.2.1, we first present a useful lemma which collects some standard properties of quantiles, which are known to specialists on quantiles. We provide a self-contained short proof since we could not find this precise formulation in the literature.

Lemma 2.2.1. *For $\alpha \in [0,1]$ and any random variable X, the following statements hold:*

(i) $\alpha > \mathbb{P}(X \leqslant t)$ if and only if $\mathrm{VaR}_\alpha^-(X) > t$;

(ii) $\alpha < \mathbb{P}(X < t)$ if and only if $\mathrm{VaR}_\alpha^+(X) < t$.

Remark 2.2.1. The statements in Lemma 2.2.1 can be equivalently stated in other forms, such as (i.a) $\mathbb{P}(X \leqslant t) \geqslant \alpha$ if and only if $\mathrm{VaR}_\alpha^-(X) \leqslant t$; (i.b) $\mathbb{P}(X > t) \leqslant 1-\alpha$ if and only if $\mathrm{VaR}_\alpha^-(X) \leqslant t$; (ii.a) $\mathbb{P}(X < t) \leqslant \alpha$ if and only if $\mathrm{VaR}_\alpha^+(X) \geqslant t$; (ii.b) $\mathbb{P}(X \geqslant t) \geqslant 1-\alpha$ if and only if $\mathrm{VaR}_\alpha^+(X) \geqslant t$.

Proof. To show (i), denote by $A_\alpha = \{t \in \mathbb{R} : \mathbb{P}(X \leqslant t) \geqslant \alpha\}$. Note that A_α is closed in $\mathbb{R}$ since $t \mapsto \mathbb{P}(X \leqslant t)$ is upper semicontinuous. This gives $\mathrm{VaR}_\alpha^-(X) = \min A_\alpha$. Hence, $\alpha > \mathbb{P}(X \leqslant t) \iff t \notin A_\alpha \iff \mathrm{VaR}_\alpha^-(X) > t$. To show (ii), denote by $B_\alpha = \{t \in \mathbb{R} : \mathbb{P}(X < t) \leqslant \alpha\}$ which is closed in $\mathbb{R}$ since $t \mapsto \mathbb{P}(X < t)$ is lower semicontinuous. This gives $\mathrm{VaR}_\alpha^+(X) = \max B_\alpha$. It follows that $\alpha < \mathbb{P}(X < t) \iff t \notin B_\alpha \iff \mathrm{VaR}_\alpha^+(X) < t$. □

Proof of Theorem 2.2.1. Let $g : [0,1] \to \mathbb{R}$, $\alpha \mapsto (1-\alpha)(\mathrm{ES}_\alpha(X) - t)$. Note that for any $\alpha, \alpha' \in [0,1]$,

$$g(\alpha) - g(\alpha') = \int_\alpha^{\alpha'} \left(\mathrm{VaR}_\beta^-(X) - t\right) \mathrm{d}\beta \tag{2.5}$$

$$= \int_\alpha^{\alpha'} \left(\mathrm{VaR}_\beta^+(X) - t\right) \mathrm{d}\beta. \tag{2.6}$$

Let $[c,d] = [\mathbb{P}(X < t), \mathbb{P}(X \leqslant t)]$. For $\alpha \leqslant d$, Lemma 2.2.1 (i) and (2.5) imply $g(\alpha) \leqslant g(d)$. For $\alpha < c$, Lemma 2.2.1 (ii) and (2.6) imply $g(\alpha) < g(c)$. For $\alpha \geqslant c$, Lemma 2.2.1 (ii) and (2.6) imply $g(\alpha) \leqslant g(c)$. For $\alpha > d$, Lemma 2.2.1 (i) and (2.5) imply $g(\alpha) < g(d)$. Summarizing the above inequalities, we obtain

$$g(\alpha_1) < g(c) = g(\alpha_2) = g(d) > g(\alpha_3) \quad \text{for any } \alpha_1 < c < \alpha_2 < d < \alpha_3.$$

Therefore, the set of maximizers for (2.4) is $[c,d]$. By using Lemma 2.2.1 (i) again,

$$g(d) = \int_{\mathbb{P}(X \leqslant t)}^1 \left(\mathrm{VaR}_\beta^-(X) - t\right) \mathrm{d}\beta = \int_0^1 \left(\mathrm{VaR}_\beta^-(X) - t\right)_+ \mathrm{d}\beta = \mathbb{E}[(X-t)_+],$$

thus showing (2.4). □

Remark 2.2.2. After the current chapter was published online as a journal paper, we became aware of earlier results very similar to reverse ES optimization formula (Theorem 2.2.1) in Ogryczak and Ruszczyński (2002, Theorem 3.1) and Rockafellar and Royset (2014, Theorem 2). Therefore, this formula should be attributed to them.

From the reverse ES optimization formula, instead of directly calculating $\mathbb{E}[(X-t)_+]$ for fixed $t \in \mathbb{R}$, we can maximize a quantile-based function $\alpha \mapsto (1-\alpha)(\mathrm{ES}_\alpha(X) - t)$ over $\alpha \in [0,1]$. Some implications of this result are discussed in Section 2.3.

The next corollary on a formula for $\mathbb{E}[X \wedge x]$ can be obtained from Theorem 2.2.1. To state this result, we define the left-ES for $\alpha \in [0,1]$ as

$$\mathrm{ES}_\alpha^-(X) = \frac{1}{\alpha} \int_0^\alpha \mathrm{VaR}_\beta^-(X) \mathrm{d}\beta, \quad X \in \mathcal{X},\ \alpha \in (0,1], \tag{2.7}$$

and $\mathrm{ES}_0^-(X) = \mathrm{VaR}_0^+(X)$.

Corollary 2.2.1. *For $t \in \mathbb{R}$ and $X \in \mathcal{X}$, it holds*

$$\mathbb{E}[X \wedge t] = \min_{\alpha \in [0,1]} \left\{ \alpha \mathrm{ES}_\alpha^-(X) + (1-\alpha)t \right\}, \tag{2.8}$$

and the set of minimizers for (2.8) is $[\mathbb{P}(X < t), \mathbb{P}(X \leqslant t)]$.

Proof. The formula (2.4) directly leads to

$$\begin{aligned}
\mathbb{E}[X \wedge t] &= \mathbb{E}[X] - \max_{\alpha \in [0,1]} \{(1-\alpha)(\mathrm{ES}_\alpha(X) - t)\} \\
&= \mathbb{E}[X] + \min_{\alpha \in [0,1]} \{(1-\alpha)(t - \mathrm{ES}_\alpha(X))\} \\
&= \min_{\alpha \in [0,1]} \{(1-\alpha)t + \mathbb{E}[X] - (1-\alpha)\mathrm{ES}_\alpha(X)\} \\
&= \min_{\alpha \in [0,1]} \left\{ (1-\alpha)t + \alpha \mathrm{ES}_\alpha^-(X) \right\}.
\end{aligned}$$

The corresponding statement on optimizers is the same as that in Theorem 2.2.1. □

2.3 Symmetries between ES optimization formula and reverse formula

The function $t \mapsto \mathbb{E}[(X-t)_+]$ is called the *mean excess function* of X according to McNeil et al. (2015), and the function $\alpha \mapsto \mathrm{ES}_\alpha(X)$ will be called the *ES profile* of X according to Burzoni et al. (2022). The ES profile also relates to the Lorenz curve (see e.g., Gastwirth (1971)), which can be written as $\alpha \mapsto \alpha \mathrm{ES}_\alpha^-(X)/\mathrm{ES}_0(X)$ for a non-negative random variable X representing the wealth distribution of a population. For clarity, we distinguish between the terms "mean excess function" (as a function of t) and "mean excess loss" (as a function of X), and analogously between the terms "ES profile" and "ES".

To appreciate Theorem 2.2.1 and contrast it with Theorem 2.1.1, we need to understand the roles of the mean excess function and the ES profile. The reason why Theorem 2.2.1 has not been explored in the literature is perhaps due to the perception that the ES profile is harder to obtain or to optimize than the mean excess function. Based on this reasoning,

it seems that using the mean excess function to compute ES is more natural than using the ES profile to compute the mean excess function. However, in theory, there is no such asymmetry: For a given random variable X, its mean excess function and its ES profile have perfectly symmetric roles, as we discuss below. Indeed, we shall see in Section 2.4 that in relevant applications, useful formulas for the mean excess function can be obtained from the ES profile via Theorems 2.2.1.

1. **Functional properties on $\mathcal{X}$.** Both the mean excess loss and ES have nice properties, symmetric to each other, as mappings on $\mathcal{X}$.

   (a) For a fixed $t \in \mathbb{R}$, the mapping $X \mapsto \mathbb{E}[(X-t)_+]$ is linear in the distribution of X and convex in the quantile of X. Indeed, this mapping satisfies the independence axiom of von Neumann and Morgenstern (1947).

   (b) For a fixed $\alpha \in (0,1)$, the mapping $X \mapsto \mathrm{ES}_\alpha(X)$ is linear in the quantile of X and concave in the distribution of X (e.g., Wang et al. (2020)). Indeed, this mapping satisfies the dual independence axiom of Yaari (1987).

2. **Optimization problems.** As for the optimization problems (2.3) and (2.4), we have the following symmetry.

   (a) In the minimization (2.3) over $t \in \mathbb{R}$, the function $t \mapsto t + \frac{1}{1-\alpha}\mathbb{E}[(X-t)_+]$ is convex in t.

   (b) In the maximization (2.4) over $\alpha \in [0,1]$, the function $\alpha \mapsto (1-\alpha)(\mathrm{ES}_\alpha(X) - t)$ is concave in α.

3. **Solutions to the optimization problems.** The optimizers to the optimization problems (2.3) and (2.4) also admit nice symmetry, as one is the quantile interval, and the other one is the probability interval.

4. **Parametric forms.** For commonly used distributions in risk management and actuarial science, if one of the mean excess loss and ES has an explicit formula, then so is the other one (e.g., Pareto distributions; see Example 2.4.1 below). Moreover, each of the two curves determines the whole distribution of the random variable.

To summarize, writing one as a minimum or maximum of the other as in Theorems 2.1.1 and 2.2.1 leads to the following implications for optimization:

(a) Theorem 2.1.1 allows one to transform the non-linear (in distribution) objective $\mathrm{ES}_\alpha(X)$ as the minimum over t of linear (in distribution) functions convex in t.

(b) Theorem 2.2.1 allows one to transform the non-linear (in quantile) objective $\mathbb{E}[(X-t)_+]$ as the maximum over α of linear (in quantile) functions concave in α.

Due to the above discussed symmetries, Theorem 2.2.1 serves as a natural dual formula to Theorem 2.1.1. Indeed, Theorems 2.1.1 and 2.2.1 are closely related to Fenchel-Legendre transformations, which we will discuss in Section 2.7.

2.4 Worst-case risk under model uncertainty

As discussed in Section 2.3, one of the greatest advantages of the ES optimization formula in Theorem 2.1.1 is that it allows us to translate optimization problems of ES to those of the mean excess function. More precisely, for a set of actions A and a loss function $f : A \times \mathbb{R}^d \to \mathbb{R}$, Theorem 2.1.1 implies

$$\min_{y\in A} \mathrm{ES}_\alpha(f(y,\mathbf{X})) = \min_{t\in\mathbb{R}} \left\{ t + \frac{1}{1-\alpha} \min_{y\in A} \mathbb{E}[(f(y,\mathbf{X})-t)_+] \right\},$$

and thus, for the minimization of ES, it suffices to minimize the mean excess loss $\mathbb{E}[(f(y,\mathbf{X})-t)_+]$ for each $t \in \mathbb{R}$, which is more convenient in many specific settings; see the review in Rockafellar and Uryasev (2013).

In contrast, Theorem 2.2.1 has a maximum operator in its formula (2.4), and it is useful in maximization problems. Moreover, even though risk often needs to be minimized, a maximum naturally appears in the presence of model uncertainty, which is often addressed via a worst-case approach. The worst-case risk evaluation under uncertainty appears in, e.g., Gilboa and Schmeidler (1989) and Maccheroni et al. (2006) in the context of decision making, Ghaoui et al. (2003) and Zhu and Fukushima (2009) in the context of optimization,

and Embrechts et al. (2013) in the context of risk aggregation. More precisely, suppose that there is uncertainty about a random vector $\mathbf{X}$, assumed to be in a set $\mathcal{U}$, and $f : \mathbb{R}^d \to \mathbb{R}$ is a loss function. Theorem 2.2.1 implies that the worst-case mean excess loss can be computed by (recall that the convention is $0 \times \infty = 0$), via exchanging the order of two suprema,

$$\sup_{\mathbf{X}\in\mathcal{U}} \mathbb{E}[(f(\mathbf{X}) - t)_+] = \max_{\alpha\in[0,1]} \left\{ (1-\alpha) \left(\sup_{\mathbf{X}\in\mathcal{U}} \mathrm{ES}_\alpha(f(\mathbf{X})) - t \right) \right\}, \tag{2.9}$$

which allows us to use rich existing results on worst-case ES.

Moreover, the maximization over $\alpha \in [0, 1]$ is attainable under a condition of uniform integrability, as we show below.

Proposition 2.4.1. *Let $\mathcal{Y}$ be a set of random variables and $t \in \mathbb{R}$. If $\mathcal{Y}$ is uniformly integrable, then*

$$\sup_{Y\in\mathcal{Y}} \mathbb{E}[(Y - t)_+] = \max_{\alpha\in[0,1]} \left\{ (1-\alpha) \left(\sup_{Y\in\mathcal{Y}} \mathrm{ES}_\alpha(Y) - t \right) \right\}. \tag{2.10}$$

In particular, if there exists $p > 1$ such that $\sup_{Y\in\mathcal{Y}} \mathbb{E}[|Y|^p] < \infty$, then $\mathcal{Y}$ is uniformly integrable and (2.10) holds.

Proof. We first show that uniform integrability of $\mathcal{Y}$ implies, for any $\alpha \in [0, 1]$,

$$\lim_{\alpha'\to\alpha} \sup_{Y\in\mathcal{Y}} \int_\alpha^{\alpha'} |\mathrm{VaR}_\beta(Y)| \mathrm{d}\beta \to 0, \tag{2.11}$$

where the limit is one-sided if $\alpha = 0$ or $\alpha = 1$. Suppose that (2.11) does not hold, and without loss of generality we consider $\alpha' \downarrow \alpha$ (in this case, $\alpha < 1$). It follows that there exists $m > 0$ such that, for any $\varepsilon \in (0, 1 - \alpha)$, there exists $Y_\varepsilon \in \mathcal{Y}$ satisfying $\int_\alpha^{\alpha+\varepsilon} |\mathrm{VaR}_\beta(Y_\varepsilon)| \mathrm{d}\beta > m$. Since VaR_α is monotone in α, we have

$$m < \int_\alpha^{\alpha+\varepsilon} |\mathrm{VaR}_\beta(Y_\varepsilon)| \mathrm{d}\beta \leqslant \int_0^{\varepsilon} |\mathrm{VaR}_\beta(Y_\varepsilon)| \mathrm{d}\beta + \int_{1-\varepsilon}^{1} |\mathrm{VaR}_\beta(Y_\varepsilon)| \mathrm{d}\beta.$$

For any $K > 0$, let $\varepsilon > 0$ be such that $4K\varepsilon < m$. It follows that

$$\begin{aligned}\mathbb{E}\left[|Y_\varepsilon|\mathbb{1}_{\{|Y_\varepsilon|>K\}}\right] &\geqslant \mathbb{E}[(|Y_\varepsilon| - K)_+] \\ &= \int_0^1 (|\mathrm{VaR}_\beta(Y_\varepsilon)| - K)_+ \mathrm{d}\beta \\ &\geqslant \int_0^\varepsilon (|\mathrm{VaR}_\beta(Y_\varepsilon)| - K)_+ \mathrm{d}\beta + \int_{1-\varepsilon}^1 (|\mathrm{VaR}_\beta(Y_\varepsilon)| - K)_+ \mathrm{d}\beta \\ &\geqslant \int_0^\varepsilon |\mathrm{VaR}_\beta(Y_\varepsilon)| \mathrm{d}\beta + \int_{1-\varepsilon}^1 |\mathrm{VaR}_\beta(Y_\varepsilon)| \mathrm{d}\beta - 2\varepsilon K > m/2.\end{aligned}$$

Hence, $\sup_{Y\in\mathcal{Y}} \mathbb{E}[|Y|\mathbb{1}_{\{|Y|>K\}}] > m/2$, contradicting uniform integrability. Therefore, (2.11) holds.

Let $g : [0,1] \to \mathbb{R}$, $\alpha \mapsto \sup_{Y\in\mathcal{Y}}(1-\alpha)(\mathrm{ES}_\alpha(Y) - t)$. Note that for any $\alpha, \alpha' \in [0,1]$, using (2.11),

$$\begin{aligned}|g(\alpha) - g(\alpha')| &= \left|\sup_{Y\in\mathcal{Y}} \int_0^{\alpha'} \left(\mathrm{VaR}_\beta^-(Y) - t\right) \mathrm{d}\beta - \sup_{Y\in\mathcal{Y}} \int_0^{\alpha} \left(\mathrm{VaR}_\beta^-(Y) - t\right) \mathrm{d}\beta\right| \\ &\leqslant \left|\sup_{Y\in\mathcal{Y}} \int_\alpha^{\alpha'} \left(\mathrm{VaR}_\beta^-(Y) - t\right) \mathrm{d}\beta\right| \leqslant \sup_{Y\in\mathcal{Y}} \int_\alpha^{\alpha'} |\mathrm{VaR}_\beta^-(Y)| \mathrm{d}\beta + |(\alpha' - \alpha)t|,\end{aligned}$$

which converges to 0 as $\alpha' \to \alpha$. This shows that g is continuous on $[0,1]$, and hence the maximum in (2.10) is attained. The last statement that boundedness of $\mathbb{E}[|Y|^p]$ implies uniformly integrability can be found in Exercise 5.5.1 of Durrett (2010). □

In two settings of uncertainty based on moment information and the Wasserstein metric which we study below, explicit formulas for the worst-case ES are available in Pesenti et al. (2020) and Liu et al. (2022), whereas the worst-case mean excess loss does not have an explicit formula. In the popular case that f is a portfolio loss function (i.e., $f(\mathbf{x}) = \mathbf{w}^\top \mathbf{x}$ for some $\mathbf{w} \in \mathbb{R}^d$), the multi-dimensional uncertainty sets reduce to one-dimensional sets of the same type; see Mao et al. (2022, Section 6). For this reason, we will focus on the one-dimensional uncertainty sets.

2.4.1 Uncertainty sets induced by moment information

We first study the uncertainty set induced by mean and a higher moment. For $p > 1$, $\mu \in \mathbb{R}$ and $\sigma \geqslant 0$, denote by

$$\mathcal{L}^p(\mu, \sigma) = \{X \in \mathcal{X} : \mathbb{E}[X] = \mu,\ \mathbb{E}[|X - \mu|^p] \leqslant \sigma^p\},$$

that is, the set of all random variables with given mean μ and a p-th centralized absolute moment at most σ^p. We are interested in the worst-case value of a functional over $\mathcal{L}^p(\mu, \sigma)$. The special case of this problem when $p = 2$, i.e., the setting with mean and variance information, has been the most popular; see e.g., Ghaoui et al. (2003), Li (2018) and Liu et al. (2020) on various risk measures.

Let $\rho : \mathcal{L}^p \to \mathbb{R}$ be a mapping where $\mathcal{L}^p$ is the set of all random variables with finite p-th moment. Note that the problem of $\sup_{X \in \mathcal{L}^p(\mu,\sigma)} \rho(X)$ is better suited for $\rho = \mathrm{ES}_\alpha$ or some other risk measures than for the mean excess loss $\rho : X \mapsto \mathbb{E}[(X-t)_+]$, because many risk measures, including VaR_α and ES_α, satisfy some simple properties which yield

$$\sup_{X \in \mathcal{L}^p(\mu,\sigma)} \rho(X) = \mu + \sigma \sup_{X \in \mathcal{L}^p(0,1)} \rho(X).$$

Therefore, we can convert the original problem to an optimization over $\mathcal{L}^p(0, 1)$. Such a relationship does not hold for the mean excess loss $\rho : X \mapsto \mathbb{E}[(X - t)_+]$.

The problem of the worst-case mean excess loss with moment conditions has a long history; see e.g., De Vylder and Goovaerts (1982) in actuarial science and Jagannathan (1977) in operations research. Pesenti et al. (2020, Corollary 1) obtained a closed-form formula for the worst-case ES_α over $\mathcal{L}^p(\mu, \sigma)$, that is,

$$\sup_{X \in \mathcal{L}^p(\mu,\sigma)} \mathrm{ES}_\alpha(X) = \mu + \sigma\alpha(\alpha^p(1 - \alpha) + (1 - \alpha)^p\alpha)^{-1/p}. \tag{2.12}$$

In particular, in case $p = 2$, it becomes $\sup_{X \in \mathcal{L}^2(\mu,\sigma)} \mathrm{ES}_\alpha(X) = \mu + \sigma(\frac{\alpha}{1-\alpha})^{1/2}$. By exchanging the order of two suprema, the problem of worst-case mean excess loss can be obtained by combining (2.12) and Theorem 2.2.1. In what follows, we use the convention that $1/0 = \infty$ and $1/\infty = 0$.

Proposition 2.4.2. *For $p > 1$, $\mu, t \in \mathbb{R}$ and $\sigma \geqslant 0$, we have*

$$\sup_{X \in \mathcal{L}^p(\mu,\sigma)} \mathbb{E}[(X-t)_+] = \max_{\alpha \in [0,1]} \left\{ (1-\alpha)(\mu - t) + \sigma \left((1-\alpha)^{1-p} + \alpha^{1-p} \right)^{-1/p} \right\}. \tag{2.13}$$

Proof. The proposition follows directly from putting together (2.10) and (2.12). □

In the most popular case $p = 2$, Proposition 2.4.2 gives

$$\sup_{X \in \mathcal{L}^2(\mu,\sigma)} \mathbb{E}[(X-t)_+] = \max_{\alpha \in [0,1]} \left\{ (1-\alpha)(\mu - t) + \sigma \sqrt{\alpha(1-\alpha)} \right\} = \frac{1}{2} \left(\mu - t + \sqrt{\sigma^2 + (\mu - t)^2} \right),$$

which coincides with Jagannathan (1977, Corollary 1.1). The maximum value in (2.13) for $p \neq 2$ can be computed numerically. We provide a numerical example below by simply taking $\mu = 0$ and $\sigma = 1$. Figure 2.1 shows value of worst-case mean excess loss with respect to different thresholds t under different moment conditions. We observe that a higher p leads to a lower value of the worst-case mean excess loss at any threshold level, and this is because the constraint $\mathbb{E}[|X - \mu|^p] \leqslant \sigma^p$ is more stringent with larger p. With Proposition 2.4.2, we can easily identify the worst-case values given a fixed threshold without knowing the exact distribution of loss.

2.4.2 Uncertainty sets induced by Wasserstein metrics

Next, we consider uncertainty sets induced by Wasserstein metrics (this setting of uncertainty will be called the Wasserstein uncertainty). Recall that the Wasserstein metric of order $p \geqslant 1$ between two distributions F and G on $\mathbb{R}$ is defined by

$$W_p(F,G) = \inf_{X \sim F,\ Y \sim G} (\mathbb{E}[|X-Y|^p])^{1/p} = \left(\int_0^1 |F^{-1}(x) - G^{-1}(x)|^p \mathrm{d}x \right)^{1/p},$$

where $X \sim F$ means that the distribution of X is F. For a benchmark loss $X \in \mathcal{X} = \mathcal{L}^1$ and an uncertainty level $\delta \geqslant 0$, the *Wasserstein ball* around X is $\{Y : W_p(F_X, F_Y) \leqslant \delta\}$, where F_X and F_Y are the distributions of X and Y, respectively. Note that $\delta = 0$ corresponds to

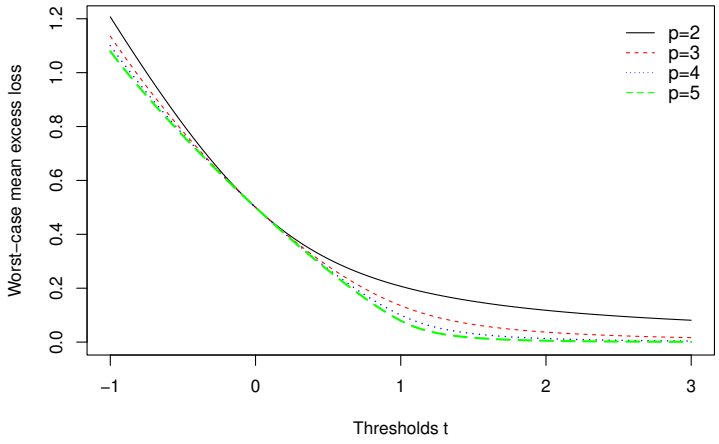


Figure 2.1: Worst-case mean excess loss with moment conditions in $\mathcal{L}^p(0,1)$

the case of no model uncertainty. The worst-case value of a risk measure $\rho : \mathcal{X} \to \mathbb{R}$ under the above uncertainty setting around X is

$$\sup\left\{\rho(Y) : W_p(F_X, F_Y) \leqslant \delta\right\}.$$

The worst-case ES_α under Wasserstein uncertainty is obtained by Liu et al. (2022, Proposition 4) with the closed-form formula

$$\sup\left\{\mathrm{ES}_\alpha(Y) : W_p(F_X, F_Y) \leqslant \delta\right\} = \mathrm{ES}_\alpha(X) + \frac{\delta}{(1-\alpha)^{1/p}}. \tag{2.14}$$

Based on (2.14) and Theorem 2.2.1, we can calculate the worst-case value of $\rho(X) = \mathbb{E}[(X-t)_+]$ for $t \in \mathbb{R}$, similarly to Proposition 2.4.2.

Proposition 2.4.3. *For $t \in \mathbb{R}$, $p \geqslant 1$, $\delta \geqslant 0$ and $X \in \mathcal{X}$, we have*

$$\sup\left\{\mathbb{E}[(Y-t)_+] : W_p(F_X, F_Y) \leqslant \delta\right\} = \max_{\alpha\in[0,1]} \left\{(1-\alpha)(\mathrm{ES}_\alpha(X) - t) + \delta(1-\alpha)^{1-1/p}\right\}. \tag{2.15}$$

Proof. The proposition follows directly from putting together (2.10) and (2.14). □

Comparing (2.15) with (2.4) in Theorem 2.2.1, to compute the function $\mathbb{E}[(Y-t)_+]$ based on ES, there is an extra term of $\delta(1-\alpha)^{1-1/p}$ in the maximization over $\alpha \in [0,1]$ to compensate for model uncertainty. As far as we are aware of, both formulas (2.13) and (2.15) in this section are new.

Example 2.4.1. Let the benchmark loss X follow a Pareto distribution with tail parameter $\theta > 1$, that is, $\mathbb{P}(X > x) = x^{-\theta}$ for $x \geqslant 1$. For simplicity we take $\theta = 2$ and consider the Wasserstein metric W_2. By straightforward calculation, $\mathrm{ES}_\alpha(X) = 2(1-\alpha)^{-1/2}$ for $\alpha \in [0,1)$. Using (2.15), we get,

$$\begin{aligned}\sup\left\{\mathbb{E}[(Y-t)_+] : W_2(F_X, F_Y) \leqslant \delta\right\} &= \max_{\alpha\in[0,1]} \left\{(2+\delta)\,(1-\alpha)^{1/2} - (1-\alpha)t\right\} \\ &= \frac{(1+\delta/2)^2}{t}\mathbb{1}_{\{t>1+\delta/2\}} + (2+\delta-t)\mathbb{1}_{\{t\leqslant 1+\delta/2\}}.\end{aligned} \tag{2.16}$$

Example 2.4.1 also illustrates how the level of model uncertainty affects the evaluation of the worst-case mean excess loss. Note that for the benchmark loss X,

$$\mathbb{E}[(X-t)_+] = \int_t^\infty \mathbb{P}(X>x)\mathrm{d}x = \int_{t\vee 1}^\infty x^{-2}\mathrm{d}x + (1-t)_+ = \frac{1}{t}\mathbb{1}_{\{t>1\}} + (2-t)\mathbb{1}_{\{t\leqslant 1\}}. \tag{2.17}$$

If $\delta = 0$, then there is no model uncertainty, and (2.16) and (2.17) coincide. If $\delta > 0$, then for $t > 1+\delta/2$, the worst-case value (2.16) of the mean excess loss increases from the benchmark value (2.17) by a factor of $(1+\delta/2)^2 > 1$; for $t \leqslant 1$, the worst-case value (2.16) increases from the benchmark value (2.17) by a constant $\delta > 0$. We observe from (2.14) that for ES_α with a fixed $\alpha \in [0,1)$, the level of model uncertainty δ always affects the worst-case risk evaluation linearly; this also holds for any coherent distortion risk measures as shown by Liu et al. (2022). In contrast, for the mean excess loss, the effect of δ is no longer linear in the interesting domain where t is large.

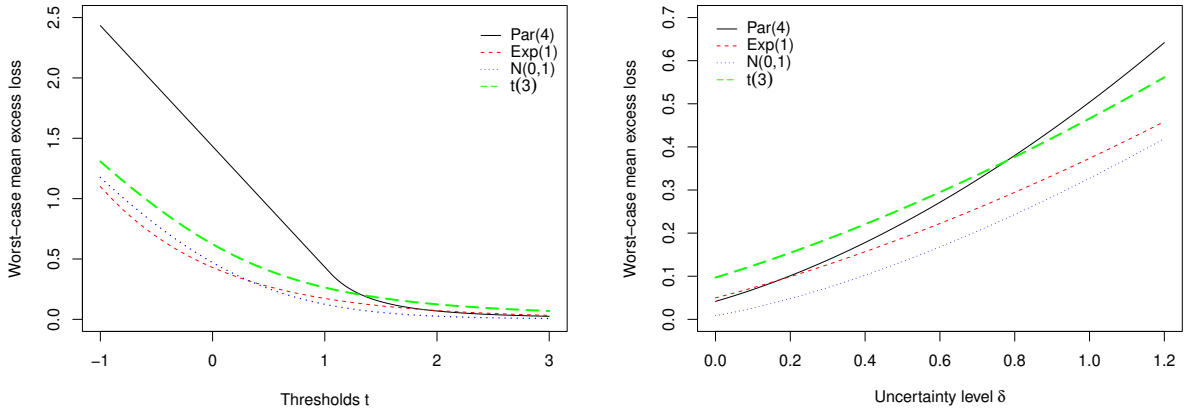
Example 2.4.2. Let the benchmark loss X follow a normal distribution with mean μ and standard deviation σ. We can calculate $\mathrm{ES}_\alpha(X) = \mu + \sigma\frac{\phi(\Phi^{-1}(\alpha))}{1-\alpha}$ for $\alpha \in [0,1)$, where

where ϕ and Φ^{-1} represent the density function and quantile function of the standard normal distribution, respectively. Using (2.15), we get

$$\sup\left\{\mathbb{E}[(Y-t)_+] : W_p(F_X, F_Y) \leqslant \delta\right\} = \max_{\alpha\in[0,1]} \left\{\sigma\phi(\Phi^{-1}(\alpha)) + (\mu - t)(1-\alpha) + \delta(1-\alpha)^{1-1/p}\right\}.$$

Although the above expression is not explicit, it can be easily computed numerically.

For a better understanding of Proposition 2.4.3, we provide another numerical example. In Figure 2.2 (a), by choosing the parameter $p = 2$ and the uncertainty level $\delta = 0.1$, we show how the worst-case values of the mean excess loss vary with the threshold t under different distributions, including Pareto, exponential, normal and t distributions. The obtained curves are similar to those in Figure 2.1. In Figure 2.2 (b), by taking $p = 2$ and $t = 2$, we report the worst-case values of the mean excess loss increases with the uncertainty level δ. As we can see, the effect of δ on the worst-case value of mean excess loss is non-linear, as we discussed in Example 2.4.1 for a Pareto distribution.



(a) Worst-case value changes with t (fixed $\delta = 0.1$) (b) Worst-case value changes with δ (fixed $t = 2$)

Figure 2.2: Worst-case mean excess loss with Wasserstein uncertainty

2.5 Empirical analysis for insurance data

In this section, we use insurance data to calculate the worst-case mean excess loss under uncertainty governed by the Wasserstein metric with $p = 2$. In addition, we check how the uncertainty level δ and the threshold level t may influence the value of worst-case mean excess loss compared to the one without uncertainty, and see their different performances in different datasets.

We choose two univariate datasets from the `R` package `CASdatasets`: normalized hurricane damages (`ushurricane`, 1900-2005) and normalized French commercial fire losses (`frecomfire`, 1982-1996) pooled by each month. Both datasets have around 180 observations and the distributions are highly right-skewed. We shall use `R` to fit the data with lognormal, Gamma and Weibull distributions as our benchmark distributions.

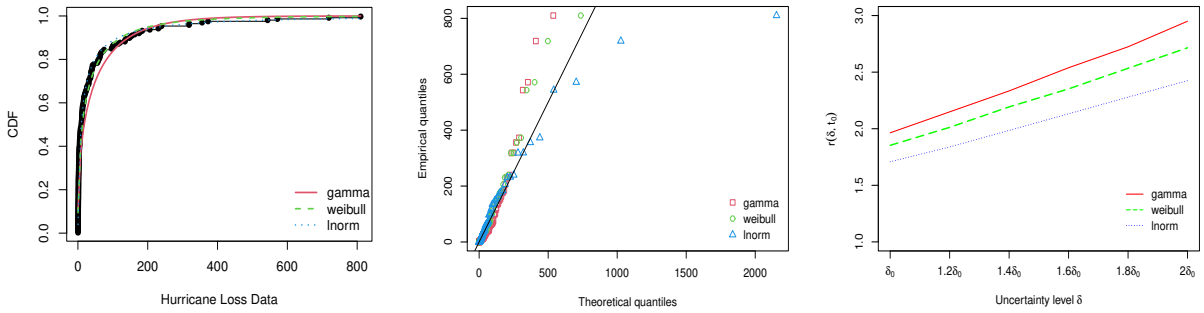
In the first part of the empirical analysis, we fix the threshold level t and let the uncertainty level δ vary to visualize how the worst-case mean excess loss varies. Since the two datasets are quite different, it is important to calibrate t and δ to make two datasets to be comparable. In particular, δ should be chosen in a statistically relevant range; see e.g., Blanchet et al. (2021) for a discussion on this point. Generally, if the uncertainty level δ is too large, then the data become less relevant; if δ is too small, we are not protected against model uncertainty, thus losing the desired robustness. For a meaningful comparison, we make the following heuristic choices. For each benchmark distribution, we let δ vary in $[\delta_0, 2\delta_0]$, where δ_0 is the Wasserstein distance (metric) between the fitted distribution and the empirical distribution. This choice ensures that the empirical distribution is inside the Wasserstein ball around the fitted distribution; intuitively, a poorly fitted distribution is associated with a larger δ_0, thus requiring a higher uncertainty level to be considered as robust. Moreover, δ_0 is of the order $n^{-1/2}$ if the estimation is $n^{-1/2}$-efficient, where n is the sample size. Table 2.1 shows the values of δ_0.

We are interested in the ratio $r(t, \delta)$ of the worst-case mean excess loss to that of the benchmark distribution, defined by

$$r(\delta, t) = \frac{\sup\{\mathbb{E}[(Y - t)_+] : W_2(F_X, F_Y) \leqslant \delta\}}{\mathbb{E}[(X - t)_+]},$$

	Lognormal	Weibull	Gamma
Hurricane loss	43.83	37.99	47.45
Fire loss	186.69	248.99	244.24

Table 2.1: Values of δ_0 for the lognormal, Weibull and Gamma distributions and for the hurricane loss and fire loss datasets. The level δ_0 is the Wasserstein metric with $p = 2$ between the empirical and the fitted distributions

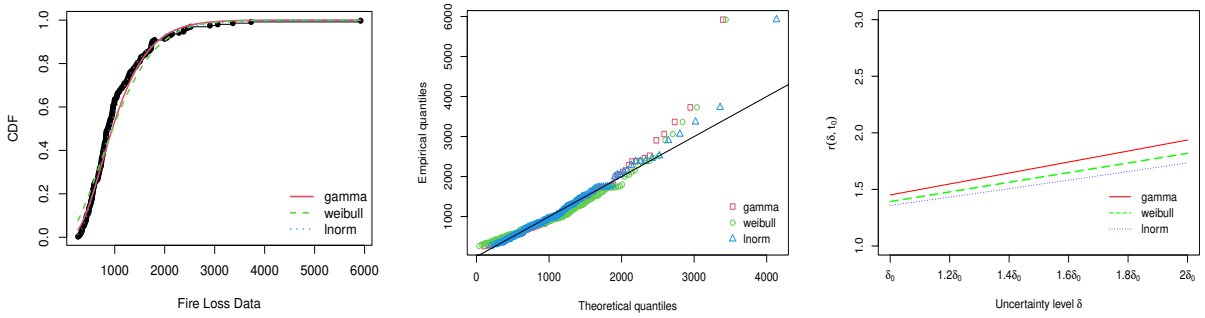


(a) Empirical and fitted CDF (b) Q-Q plot (c) The curve of $r(\delta, t_0)$

Figure 2.3: Empirical results on the hurricane loss data

where X follows one of the benchmark distributions (3 choices for each dataset). We first fix the threshold level t as the first quartile (25% quantile) t_0 of its corresponding benchmark distribution and let δ vary (Figures 2.3 and 2.4), and then we fix $\delta = \delta_0$ and let t vary (Figures 2.5 and 2.6).

Figure 2.3 (a) and (b) present goodness-of-fit plots for the fitted lognormal, Weibull and Gamma distributions to the hurricane data. We observe that the lognormal and Weibull distributions fit better to this dataset than the Gamma distribution. In Figure 2.3 (c), the Gamma model is penalized more for model uncertainty. We note that the curves $\delta \mapsto r(\delta, t_0)$ are almost linear in δ. The numerical values of $r(\delta, t_0)$ in Table 2.2 show that



(a) Empirical and fitted CDF (b) Q-Q plot (c) The curve of $r(\delta, t_0)$

Figure 2.4: Empirical results on the fire loss data

$r(\delta, t_0)$ is actually convex in δ, implying that the worst-case mean excess loss becomes more sensitive to δ for large values of δ, consistent with the numerical analysis in Section 2.4. Figure 2.4 based on the fire loss data exhibits similar patterns to the hurricane loss data. The lognormal distribution fits better to the fire loss than other two distributions so that the mean excess loss will be less affected by the Wasserstein uncertainty.

Comparing the curves $r(\delta, t_0)$ for two datasets, we can notice that the values of $r(\delta, t_0)$ for the hurricane data are much higher than the ones for the fire data, which means the hurricane loss is more severely affected by model uncertainty than fire loss. It may be explained by the fact that the hurricane losses are more catastrophic and right-skewed than fire losses so that more penalties should be added to hurricane case if model uncertainty is a concern.

In an insurance pricing context, the mean excess loss can be used to price the stop-loss premium where the threshold t can be seen as a deductible, and our method can be used to analyze the sensitivity of the stop-loss premium to the Wasserstein uncertainty. Taking the lognormal distribution as an example in Table 2.2, if we use $\delta = \delta_0$ and $t = t_0$ for pricing a hurricane insurance, the stop-loss premium will increase 70.8% compared to the one without considering model uncertainty. For the same choice $\delta = \delta_0$ and $t = t_0$

		δ_0	$1.2\delta_0$	$1.4\delta_0$	$1.6\delta_0$	$1.8\delta_0$	$2\delta_0$
Hurricane loss	Lognormal	1.708	1.839	1.985	2.132	2.279	2.425
	Weibull	1.853	2.012	2.193	2.352	2.534	2.715
	Gamma	1.964	2.149	2.334	2.539	2.724	2.950
Fire loss	Lognormal	1.358	1.431	1.505	1.582	1.657	1.735
	Weibull	1.400	1.481	1.564	1.649	1.733	1.819
	Gamma	1.456	1.548	1.644	1.740	1.837	1.937

Table 2.2: Values of $r(\delta, t_0)$ for the hurricane loss and the fire loss datasets. The threshold level t_0 is the first quartile of the benchmark distribution and the parameter δ_0 is the Wasserstein metric between the empirical and the fitted distributions

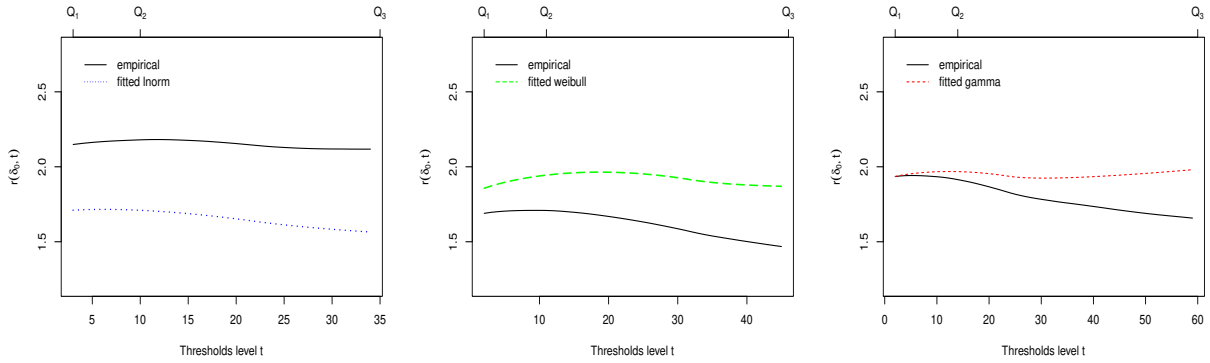
(although both t_0 and δ_0 depend on the dataset), the stop-loss premium will only increase 35.8% when pricing the fire insurance. It intuitively means that the hurricane insurance pricing is more sensitive to the Wasserstein uncertainty than the fire insurance pricing.

Next, we investigate how different threshold levels t may influence the mean excess loss with and without the Wasserstein uncertainty. The uncertainty level δ is fixed as δ_0 in this experiment and we look at $r(\delta_0, t)$ as t varies. Figures 2.5 and 2.6 report the ratio $r(\delta_0, t)$ in these settings, as well as the ratio

$$\hat{r}(\delta_0, t) = \frac{\sup\{\mathbb{E}[(Y-t)_+] : W_2(F_X, F_Y) \leqslant \delta_0\}}{\mathbb{E}[(\hat{X}-t)_+]},$$

where $\hat{X}$ follows the empirical distribution of the data. Note that $\hat{r}(\delta_0, t) \geqslant 1$ since δ_0 is chosen such that the empirical distribution is inside the Wasserstein ball. For both datasets, we let the threshold level t vary between the first quartile and the third quartile of the loss data. We observe from Figure 2.5 that the ratio $r(\delta_0, t)$ for the hurricane loss data is relatively stable with respect to the threshold level t, whereas Figure 2.6 shows that the ratio $r(\delta_0, t)$ for the fire loss data increases with the threshold level t in all selected benchmark distributions. Hence, compared to hurricane loss, the mean excess loss of fire loss data is more sensitive to model uncertainty with larger threshold levels. This

observation is less pronounced for the better fitted lognormal distribution in Figure 2.6 (a). The other ratio $\hat{r}(\delta_0, t)$ is quite stable for the fire loss data and it shows a decreasing trend in t for the hurricane loss data.

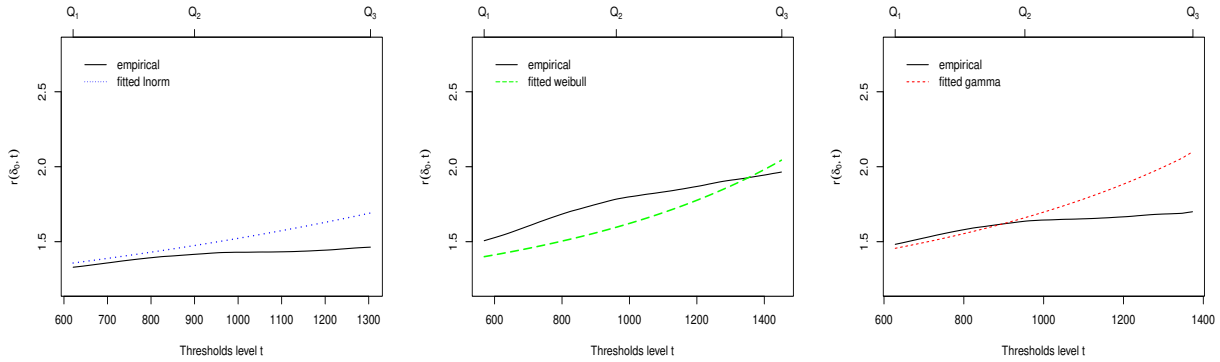


(a) Fitted to a lognormal distribution (b) Fitted to a Weibull distribution (c) Fitted to a Gamma distribution

Figure 2.5: Values of the ratios $r(\delta_0, t)$ and $\hat{r}(\delta_0, t)$ for the hurricane loss data, where Q_1, Q_2 and Q_3 represent the 1st, 2nd and 3rd quartiles of the data

2.6 Optimized certainty equivalents

We proceed to offer some more theoretical results and discussions on the reverse ES optimization formula. It would be interesting to see whether Theorem 2.2.1 can be generalized to other risk measures than the class ES_α. Note that ES_α belongs to the class of optimized certainty equivalents (OCE) of Ben-Tal and Teboulle (2007). The class of OCE includes ES and the entropic risk measures (Föllmer and Schied (2016)) as special cases. In this section, we work with the set $\mathcal{X}_B$ of essentially bounded random variables to avoid integrability issues. Let V be the set of increasing and convex functions $v : \mathbb{R} \to \mathbb{R}$ satisfying $v(0) = 0$, $\bar{v} \geqslant 1$ and $\lim_{t\to\infty} v'_+(-t) = 0$ where $\bar{v} = \sup_{x\in\mathbb{R}} v'_+(x)$ and v'_+ is the



(a) Fitted to a lognormal distribution

(b) Fitted to a Weibull distribution

(c) Fitted to a Gamma distribution

Figure 2.6: Values of the ratios $r(\delta_0, t)$ and $\hat{r}(\delta_0, t)$ for the fire loss data, where Q_1, Q_2 and Q_3 represent the 1st, 2nd and 3rd quartiles of the data

right derivative of v. An OCE is a risk measure R defined by

$$R(X) = \inf_{t \in \mathbb{R}} \{t + \mathbb{E}[v(X - t)]\}, \qquad X \in \mathcal{X}_B.$$

The finiteness of R is guaranteed if $v'_-(x) \leqslant 1 \leqslant v'_+(y)$ for some $x, y \in \mathbb{R}$ which is satisfied by $v \in V$ if $\bar{v} > 1$. If R is finite, then it is a convex risk measure in the sense of Föllmer and Schied (2016). In particular, if $v(x) = x_+/(1-\alpha)$ for $\alpha \in [0, 1)$, then R is ES_α as in Theorem 2.1.1. Moreover, under a continuity condition, ES_α is the only class of coherent risk measures in the class of OCE (Theorem 3.1 of Embrechts et al. (2021)).

Inspired by Theorem 2.2.1, we define a parametric family of OCE risk measures. For $v \in V$ and $\beta \in (0, \bar{v}]$, let

$$R^v_\beta(X) = \inf_{t \in \mathbb{R}} \left\{ t + \frac{1}{\beta} \mathbb{E}[v(X - t)] \right\}, \qquad X \in \mathcal{X}_B.$$

Here, the convention is $1/\infty = 0$. If $v(x) = x_+$, then $\bar{v} = 1$ and $R^v_\beta = \mathrm{ES}_{1-\beta}$ for $\beta \in (0, 1]$. The next result gives a reverse optimization formula for OCE, which includes the formula

(2.4) as a special case. This result is related to the Fenchel-Legendre transformation as we discuss in Section 2.7.

Theorem 2.6.1 (Reverse OCE optimization formula). *For $X \in \mathcal{X}_B$, $t \in \mathbb{R}$ and $v \in V$, it holds*

$$\mathbb{E}[v(X-t)] = \sup_{\beta \in (0,\bar{v}]} \left\{\beta(R^v_\beta(X) - t)\right\}. \tag{2.18}$$

Proof. Let $f : \mathbb{R} \to \mathbb{R}$ be defined by $f(t) = \mathbb{E}[v(X+t)]$, which is an increasing convex function on $\mathbb{R}$. As a convex function on $\mathbb{R}$, f is automatically continuous. Its Fenchel-Legendre transform f^*, called the conjugate function of f, is given by

$$f^*(\beta) = \sup_{t \in \mathbb{R}} \{t\beta - f(t)\}, \quad \beta \in \mathbb{R}, \tag{2.19}$$

which is not necessarily finite.

If $\beta < 0$, then letting $t \to -\infty$ gives $\sup_{t\in\mathbb{R}}\{t\beta - f(t)\} = \infty$ since f is increasing. On the other hand, if $\beta > \bar{v}$, then, since $v'_+(x) \leqslant \bar{v}$ for $x \in \mathbb{R}$, letting $t \to \infty$ gives $\sup_{t\in\mathbb{R}}\{t\beta - f(t)\} = \infty$.

Let $s \in \mathbb{R}$ be such that $f'_+(s) > \bar{v}/2$; such s exists since $\lim_{t\to\infty} f'_+(t) = \bar{v}$. For $\beta \in (0, \bar{v}/2]$, we have

$$f^*(\beta) = \sup_{t\in(-\infty,s]} \{t\beta - f(t)\} \leqslant s\beta + \sup_{t\in\mathbb{R}}\{-f(t)\} = s\beta + f^*(0).$$

Hence,

$$\limsup_{\beta\downarrow 0} f^*(\beta) \leqslant f^*(0). \tag{2.20}$$

Summarizing the above observations, for a fixed $t \in \mathbb{R}$,

$$\sup_{\beta\in\mathbb{R}} \{-t\beta - f^*(\beta)\} = \sup_{\beta\in[0,\bar{v}]} \{-t\beta - f^*(\beta)\} = \sup_{\beta\in(0,\bar{v}]} \{-t\beta - f^*(\beta)\}, \tag{2.21}$$

where the last equality is due to (2.20). For $\beta \in (0, \bar{v}]$,

$$\frac{-f^*(\beta)}{\beta} = \inf_{t\in\mathbb{R}} \left\{-t + \frac{1}{\beta} f(t)\right\} = \inf_{t\in\mathbb{R}} \left\{t + \frac{1}{\beta} f(-t)\right\} = R^v_\beta(X).$$

Thus, $f^*(\beta) = -\beta R^v_\beta(X)$. The Fenchel-Legendre theorem in the form of Proposition A.9 of Föllmer and Schied (2016) gives $f^{**} = f$. Therefore, using (2.21),

$$f(-t) = \sup_{\beta\in(0,\bar{v}]} \{-t\beta - f^*(\beta)\} = \sup_{\beta\in(0,\bar{v}]} \left\{-t\beta + \beta R^v_\beta(X)\right\} = \sup_{\beta\in(0,\bar{v}]} \left\{\beta(R^v_\beta(X) - t)\right\},$$

and hence (2.18) holds. □

As we can see from Theorem 2.6.1, the symmetry between the ES optimization formula (2.3) and the reverse formula (2.4) can be seen as a consequence of the Fenchel-Legendre transform mechanism.

2.7 Related Fenchel-Legendre transforms

As mentioned above, the reverse ES optimization formula and reverse OCE optimization formula is closely related to the Fenchel-Legendre transformation (e.g., Definition A.8 of Föllmer and Schied (2016)). In this section, we give two pairs of conjugate functions related to Theorem 2.2.1.

The Fenchel-Legendre transformation converts convex functions to their conjugate. For a convex function $f : \mathbb{R} \to \mathbb{R}$, its Fenchel-Legendre transform is the function f^* on $\mathbb{R}$ defined by

$$f^*(\beta) = \sup_{t\in\mathbb{R}} \{t\beta - f(t)\}, \qquad \beta \in \mathbb{R},$$

where β may be constrained to a subset of $\mathbb{R}$ such that f^* is real.

As we have seen in Theorem 2.6.1, Fenchel-Legendre transforms are closely related to our reverse ES optimization formula, as Theorems 2.1.1 and 2.2.1 can be expressed from each other via a Fenchel-Legendre transform. Below, we identify two other pairs of conjugate functions, one being quantile-based and one being expectation-based, analogously to the case of ES and the mean excess function.

Proposition 2.7.1. *Fix $X \in \mathcal{X}$.*

(i) The Fenchel-Legendre transform of the convex quantile-based function $f_1 : [0,1] \to \mathbb{R}$,

$$f_1(\alpha) = -(1-\alpha)\mathrm{ES}_\alpha(X),$$

is given by

$$f_1^*(t) = \max_{\alpha\in[0,1]} \{\alpha t - f_1(\alpha)\} = \mathbb{E}[X \vee t].$$

(ii) The Fenchel-Legendre transform of the convex quantile-based function $f_2 : [0,1] \to \mathbb{R}$,

$$f_2(\alpha) = \alpha\mathrm{ES}_\alpha^-(X),$$

is given by

$$f_2^*(t) = \max_{\alpha\in[0,1]} \{\alpha t - f_2(\alpha)\} = \mathbb{E}[(t-X)_+].$$

Moreover, the set of maximizers for both maximization problems is $[\mathbb{P}(X < t), \mathbb{P}(X \leqslant t)]$.

Proof. For the first statement, it is straightforward to identify that the quantile-based function $f_1 : \alpha \mapsto -(1-\alpha)\mathrm{ES}_\alpha(X)$ is convex by taking a derivative with respect to α. By definition of the Legendre-Fenchel transformation, we have

$$\begin{aligned} f_1^*(t) &= \sup_{\alpha\in[0,1]} \{\alpha t + (1-\alpha)\mathrm{ES}_\alpha(X)\} \\ &= \sup_{\alpha\in[0,1]} \{(\alpha-1)t + (1-\alpha)\mathrm{ES}_\alpha(X)\} + t \\ &= \sup_{\alpha\in[0,1]} \{(1-\alpha)(\mathrm{ES}_\alpha(X) - t)\} + t. \end{aligned}$$

By the reverse ES optimization formula in Theorem 2.2.1, we know that $\alpha \in [\mathbb{P}(X < t), \mathbb{P}(X \leqslant t)]$ is a maximizer of function $\alpha \mapsto (1 - \alpha(\mathrm{ES}_\alpha(X) - t))$, and hence the supremum above is attainable. Thus, we can conclude that

$$f_1^*(t) = \mathbb{E}[(X-t)_+] + t = \mathbb{E}[X \vee t].$$

The proof of the second Fenchel-Legendre transform follows the same routine. We apply the Fenchel-Legendre transform to the convex function $f_2(\alpha) = \alpha \mathrm{ES}_\alpha^-(X)$. Then we have

$$\begin{aligned} f_2^*(t) &= \sup_{\alpha\in[0,1]} \left\{\alpha t - \alpha \mathrm{ES}_\alpha^-(X)\right\} \\ &= \sup_{\alpha\in[0,1]} \left\{\alpha t - \mathbb{E}(X) + \int_0^1 \mathrm{VaR}_\beta^-(X)\mathrm{d}\beta - \int_0^\alpha \mathrm{VaR}_\beta^-(X)\mathrm{d}\beta\right\} \\ &= \sup_{\alpha\in[0,1]} \left\{(1-\alpha)(\mathrm{ES}_\alpha(X) - t)\right\} + t - \mathbb{E}[X]. \end{aligned}$$

By Theorem 2.2.1, we conclude that

$$f_2^*(t) = \mathbb{E}[(X-t)_+] + t - \mathbb{E}[X] = \mathbb{E}[(t-X)_+].$$

We can check that both $f_1^* : t \mapsto \mathbb{E}[X \vee t]$ and $f_2^* : t \mapsto \mathbb{E}[(t-X)_+]$ are convex functions. □

By applying Legendre-Fenchel transform mechanism to the convex functions $f_1^* : t \mapsto \mathbb{E}[X \vee t]$ and $f_2^* : t \mapsto \mathbb{E}[(t-X)_+]$, one obtains the corresponding quantile-based functions f_1 and f_2 in Proposition 2.7.1 as their conjugate functions.

2.8 Concluding remarks

The reverse ES optimization formula obtained in Theorem 2.2.1 serves as a dual formula to the celebrated ES optimization formula of Rockafellar and Uryasev (2000, 2002), and they are connected via the Fenchel-Legendre transforms. This new formula reveals profound symmetries between these two formulas regarding to their functional properties, parametric forms, optimization problems and the solutions to the optimization problems, and it can be generalized for the class of OCE of Ben-Tal and Teboulle (2007). The reverse ES optimization formula is particularly useful when directly calculating the mean excess loss is cumbersome, and this is illustrated by two settings of model uncertainty. The new formulas are applied to settings of model uncertainty and two insurance datasets.

The new formula may appear simple to risk experts, although we could not find it in the literature. The reason why such a natural formula has not been studied could partially

be explained by the fact that the need for utilizing existing ES results to compute the mean excess loss mainly arises in the recent years, when model uncertainty is actively studied in quantitative risk management, as we discuss in Sections 2.4 and 2.5.

The main purpose of this chapter is to introduce the new formula and discuss its direct implications. Given the importance of both the mean excess function and ES in actuarial science and risk management, we are optimistic about other potential applications of the formula, which will need future exploration.

Chapter 3

Worst-case values of target semi-variances with applications to robust portfolio selection

3.1 Introduction

Assume that X is a random variable denoting the loss of an investment portfolio. Hence, the larger the value of X is, the worse the portfolio is. The manager of an investment portfolio often has a target return $-t$ or equivalently a threshold loss t. Thus, the loss function $(X-t)_+$ represents the downside risk/loss of the portfolio, while $(-X-(-t))_+ = (X-t)_-$ denotes the excess profit of the portfolio over the target return. Here and throughout this chapter, $(x)_+ = \max\{x, 0\}$ and $(x)_- = \max\{-x, 0\}$ for any $x \in \mathbb{R} = (-\infty, \infty)$. Two important quantities of the downside risk $(X-t)_+$ are the first moment $\mathbb{E}[(X-t)_+]$ that measures the expected loss above the threshold loss t and the second moment $\mathbb{E}[(X-t)_+^2]$ that quantifies the dispersion of the loss that exceeds the threshold loss t. In the literature, the two moments are often called the first-order and second-order upper partial moments, respectively. In addition, the first moment $\mathbb{E}[(X-t)_+]$ is also referred to as the *expected regret* or *target shortfall* (see, e.g, Testuri and Uryasev (2004), Krokhmal et al. (2011)) while

the second moment $\mathbb{E}[(X-t)_+^2]$ is also referred to as the *target semi-variance* (see, e.g., Rohatgi (2011)). If the target return is equal to the expected return, namely, $t = \mathbb{E}[X]$, the target semi-variance $\mathbb{E}[(X-\mathbb{E}[X])_+^2]$ is called the semi-variance of the loss X. Both of the expected regret and the target semi-variance are important risk measures of the downside risk and have been extensively used in finance, insurance, operations research, and many other fields. It is well known that the expected regret is connected to the second-order stochastic dominance which coincides with all risk-aversion functions within expected utility framework. While the target semi-variance is consistent with third-order stochastic dominance, which aligns with all risk-averse and skewness-loving functions within the von Neumann-Morgenstern expected utility theory, and thus reflects the investor's actual risk perception.[1] For more detailed discussion of the relationship between the target semi-variance and stochastic dominance under the expected utility framework, see Bawa (1975), Porter (1974), Ogryczak and Ruszczyński (2001), and the references therein.

If the 'true' distribution of the loss X is known, the expected regret $\mathbb{E}[(X-t)_+]$ and the target semi-variance $\mathbb{E}[(X-t)_+^2]$ can be calculated analytically or numerically. However, in practice, the 'true' distribution of X is often unknown. A decision maker may have only partial information on X such as the mean and variance of X. If only partial information on X is available and the possible distributions of X belong to a distribution set $\mathcal{L}$, called an *uncertainty set* for X, a decision maker is often interested in $\sup_{F\in\mathcal{L}} \mathbb{E}^F[(X-t)_+]$ and $\sup_{F\in\mathcal{L}} \mathbb{E}^F[(X-t)_+^2]$, which are respectively called the *worst-case expected regret* and the *worst-case target semi-variance* over the uncertainty set $\mathcal{L}$. Here and throughout this chapter, for a function h defined on $\mathbb{R}$ and a risk measure ρ, such as expectation $\mathbb{E}$, variance Var, and conditional value-at-risk CVaR, $\rho^F[h(X)]$ means that the risk measure of $\rho(h(X))$ is calculated under the distribution F if the distribution of X is F. In the literature, for a random variable X and a loss/cost function h, when the 'true' distribution of X is unknown or uncertain but is assumed to be in an uncertainty set $\mathcal{L}$, the optimization problem of

$$\sup_{F\in\mathcal{L}} \rho^F[h(X)] \tag{3.1}$$

is called a distributionally robust optimization (DRO) problem and if there exists a dis-

[1] Risk-averse and skewness-loving refer to decreasing absolute risk averse among investors, which means they will choose to increase investments in risky assets as their wealth increases.

tribution $F^* \in \mathcal{L}$ such that $\sup_{F\in\mathcal{L}} \rho^F[h(X)] = \rho^{F^*}[h(X)]$, such a distribution is called a *worst-case distribution*. The DRO problem (3.1) and its applications have been extensively studied in the literature of finance, insurance, operations research, and many other fields. For instance, Jagannathan (1977) investigates problem (3.1) when $\rho = \mathbb{E}$, $h(x) = (x-t)_+$, $\mathcal{L}$ is a set containing distributions with the given first two moments, and X is an arbitrary or symmetric or non-negative random variable. Zuluaga et al. (2009) considers problem (3.1) when $\rho = \mathbb{E}$, $h(x) = (x-t)_+$, $\mathcal{L}$ is a set containing distributions with the given first three moments. Chen et al. (2011) studies problem (3.1) when $\rho = \mathbb{E}$, $h(x) = (x-t)_-^2$, and $\mathcal{L}$ is a set containing distributions with the given first two moments. Tang and Yang (2023) discusses problem (3.1) when $h(x) = x^m$ or $(x-t)_+^m$, $m = 1, 2, ...$, and $\mathcal{L}$ is a set containing distributions satisfying a distance constraint to a reference distribution. Cai et al. (2024) studies problem (3.1) when ρ is a distortion risk measure, $h(x) = (x-t)_+$, and $\mathcal{L}$ is a set containing distributions satisfying a distance constraint to a reference distribution and constraints on first two moments. For the studies and applications of the DRO problem (3.1) with other forms of the function h and the risk measure ρ, we refer to Ben-Tal and Nemirovski (1998), Bertsimas and Popescu (2002), Ghaoui et al. (2003), Hürlimann (2005), Natarajan et al. (2008), Zhu and Fukushima (2009), Zhu et al. (2009), Asimit et al. (2017), Pflug et al. (2017), Li (2018), Kang et al. (2019), Liu and Mao (2022), Bernard et al. (2023), Cai et al. (2023), and the references therein.

In many DRO problems, it is assumed that the mean and variance or second moment of a random variable X are the only known information on the distribution of X. This assumption is consistent with the fact that the mean and variance of a random variable X are two quantities that can be estimated easily from the observed data of X. In fact, the following uncertainty set

$$\begin{aligned}\mathcal{L}(\mu,\sigma) &= \left\{F \in \mathcal{F}(\mathbb{R}) : \int_{-\infty}^{\infty} x \,\mathrm{d}F(x) = \mu, \int_{-\infty}^{\infty} x^2 \,\mathrm{d}F(x) = \mu^2+\sigma^2\right\} \\ &= \left\{F \in \mathcal{F}(\mathbb{R}) : \mathbb{E}^F[X] = \mu, \mathbb{E}^F[X^2] = \mu^2+\sigma^2\right\}, \end{aligned} \tag{3.2}$$

is one of the popular uncertainty sets used in the study of DRO problems, where $\mathcal{F}(\mathbb{R})$ is the set of all the distributions defined on $\mathbb{R}$. In practice, a decision maker may have additional information on the distribution of X besides its mean and variance. In finance,

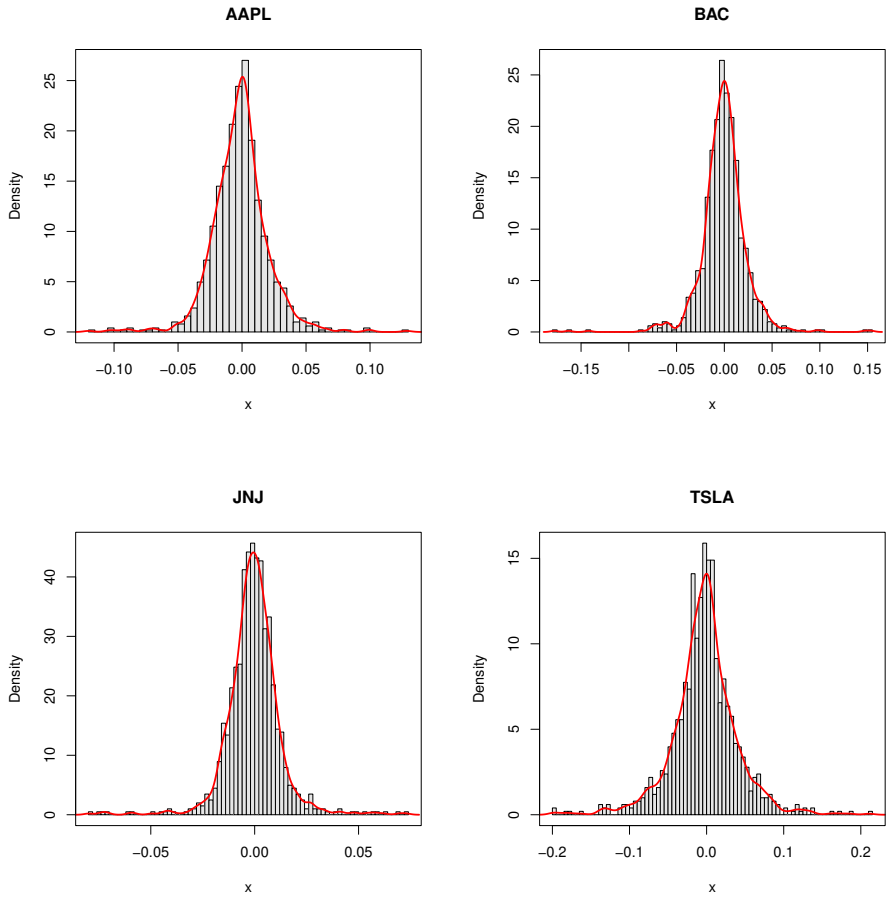


Figure 3.1: Histograms of daily losses of the stocks of Apple (AAPL), Bank of America (BAC), Johnson & Johnson (JNJ) and Tesla (TSLA). The data used for this figure covers a four-year period from January 2, 2019, to January 2, 2023, and includes 1007 observations of daily losses from Yahoo! Finance.

a decision maker may notice that the loss data have the symmetric features. For instance, Figure 3.1 displays the histograms of daily losses of the stocks of Apple, Netflix, Alphabet (Google) and eBay. The daily losses of these stocks exhibit a high degree of symmetry.

If fact, in many portfolio selection researches, the daily/monthly/quarterly losses of

the underlying assets are assumed to have multivariate symmetric distributions such as multivariate normal distributions, multivariate t-distributions, multivariate elliptical distributions, and so on, we refer Owen and Rabinovitch (1983), Buckley et al. (2008), Hu and Kercheval (2010), Fang (2018), and references therein.

In addition, in insurance, loss random variables often are the amounts and numbers of insurance claims that are non-negative random variables. Hence, the following two uncertainty sets

$$\mathcal{L}_S(\mu,\sigma)=\big\{F\in\mathcal{L}(\mu,\sigma):\ F \text{ is symmetric}\big\}, \tag{3.3}$$

$$\mathcal{L}^+(\mu,\sigma)=\big\{F\in\mathcal{L}(\mu,\sigma):\ F(0-)=0\big\}, \tag{3.4}$$

are also interesting in the study of DRO problems. In this chapter, the formal definitions of symmetric distributions are given in Definitions 3.2.1 and 3.4.1, and a non-negative distribution means that $F(0-)=\mathbb{P}\{X<0\}=0$ or F is a distribution of a non-negative random variable X.

The explicit and closed-form expressions for $\sup_{F\in\mathcal{L}}\mathbb{E}[(X-t)_+]$ have been derived in Jagannathan (1977) when $\mathcal{L}$ is one of the three uncertainty sets $\mathcal{L}(\mu,\sigma)$, $\mathcal{L}_S(\mu,\sigma)$, and $\mathcal{L}^+(\mu,\sigma)$. The explicit and closed-form expression for $\sup_{F\in\mathcal{L}}\mathbb{E}[(X-t)_-^2]$ has been obtained in Chen et al. (2011) when $\mathcal{L}=\mathcal{L}(\mu,\sigma)$. To the best of our knowledge, the worst-case values of $\mathbb{E}[(X-t)_+^2]$ over the uncertainty sets $\mathcal{L}_S(\mu,\sigma)$ and $\mathcal{L}^+(\mu,\sigma)$ have not been solved. As discussed in this chapter later, the methods and proofs used in Jagannathan (1977) and Chen et al. (2011) do not apply for the worst-case values of $\mathbb{E}[(X-t)_+^2]$ over the uncertainty sets $\mathcal{L}_S(\mu,\sigma)$ and $\mathcal{L}^+(\mu,\sigma)$.

In this chapter, first, we complement the study of Chen et al. (2011) on worst-case values of the target semi-variance and obtain the explicit and closed-form expressions for the worst-case values of the target semi-variance over the uncertainty sets $\mathcal{L}_S(\mu,\sigma)$ and $\mathcal{L}^+(\mu,\sigma)$. Second, motivated by the classical mean-variance (M-V) portfolio selection model, in this chapter, we discuss the applications of the worst-case target semi-variance in portfolio selection problems.

The uncertainty sets $\mathcal{L}_S(\mu,\sigma)$ and $\mathcal{L}^+(\mu,\sigma)$ have more constraints than $\mathcal{L}(\mu,\sigma)$. Finding the worst-case values of $\mathbb{E}[(X-t)_+^2]$ over the uncertainty sets $\mathcal{L}_S(\mu,\sigma)$ and $\mathcal{L}^+(\mu,\sigma)$ is a

challenging question, in particular, over the uncertainty sets $\mathcal{L}_S(\mu, \sigma)$. Our methods are different from those used in Jagannathan (1977) and Chen et al. (2011). The main method used in this chapter for finding these worst-case values is to reformulate these infinite-dimensional optimization problems to finite-dimensional optimization problems and then solve the finite-dimensional optimization problems to obtain the explicit and closed-form expressions for the worst-case values.

The rest of this chapter is structured as follows. In Section 3.2, we give the preliminaries of the worst-case values of the expected regret and target semi-variance and describe our motivation for studying the worst-case target semi-variance. In Section 3.3, we derive the explicit and close-form expressions for the worst-case target semi-variance over the uncertainty set $\mathcal{L}_S(\mu, \sigma)$. In Section 3.4, we propose robust portfolio selection models that minimize the target semi-variance under the different uncertainty sets discussed above. In Section 3.5, we use the real finance data to compare the investment performances of our portfolio selection methods with several existing portfolio selection models related to the models proposed in this chapter. Finally, in Section 3.6, we give concluding remarks.

3.2 Preliminary and motivation

Definition 3.2.1. The distribution F of a random variable X is said to be symmetric if there exists a constant a such that $\mathbb{P}(X - a > x) = \mathbb{P}(X - a < -x)$, under the distribution F, for all $x \in \mathbb{R}$. If such a constant a exists, random variable X or its distribution is said to be symmetric about a.

Intuitively, random variable X is symmetric about a if and only if $X - a$ is symmetric about the origin of $\mathbb{R}$. Examples of the continuous symmetric distributions include the Cauchy distribution, normal distributions, t-distributions, uniform distributions, logistic distributions, and many others. Examples of the discrete symmetric distributions include discrete uniform distributions, k-point symmetric distributions (where $k \geqslant 2$ is an integer), and many others. In addition, a degenerate distribution is also symmetric according to Definition 3.2.1.

To give a detailed review of the known results about the worst-case values $\sup_{F\in\mathcal{L}}\mathbb{E}^F[(X-t)_+]$ and $\sup_{F\in\mathcal{L}}\mathbb{E}^F[(X-t)_+^2]$ and illustrate our motivation for studying the worst-case target semi-variance, we state the results of Jagannathan (1977) and Chen et al. (2011) about these worst-case values and give remarks on these results and their proofs below.

Proposition 3.2.1. *(Jagannathan (1977)) For any $\mu, t \in \mathbb{R}$ and $\sigma \in \mathbb{R}^+ = (0,\infty)$, if the uncertainty set of random variable X is $\mathcal{L}(\mu,\sigma)$, then*

$$\sup_{F\in\mathcal{L}(\mu,\sigma)}\mathbb{E}^F[(X-t)_+] = \frac{1}{2}\left(\mu - t + \sqrt{\sigma^2+(\mu-t)^2}\right). \tag{3.5}$$

If the uncertainty set of X is $\mathcal{L}_S(\mu,\sigma)$, then

$$\sup_{F\in\mathcal{L}_S(\mu,\sigma)}\mathbb{E}^F[(X-t)_+] = \begin{cases} \frac{8(\mu-t)^2+\sigma^2}{8(\mu-t)}, & t < \mu - \frac{\sigma}{2}, \\ \frac{1}{2}(\mu+\sigma-t), & \mu-\frac{\sigma}{2} \leqslant t < \mu+\frac{\sigma}{2}, \\ \frac{\sigma^2}{8(t-\mu)}, & t \geqslant \mu+\frac{\sigma}{2}. \end{cases} \tag{3.6}$$

If the uncertainty set of X is $\mathcal{L}^+(\mu,\sigma)$ and $\mu > 0$, then

$$\sup_{F\in\mathcal{L}^+(\mu,\sigma)}\mathbb{E}^F[(X-t)_+] = \begin{cases} \mu - t, & t < 0, \\ \mu - \frac{\mu^2 t}{\sigma^2+\mu^2}, & 0 \leqslant t < \frac{\sigma^2+\mu^2}{2\mu}, \\ \frac{1}{2}\left(\mu - t + \sqrt{\sigma^2+(\mu-t)^2}\right), & t \geqslant \frac{\sigma^2+\mu^2}{2\mu}. \end{cases} \tag{3.7}$$

Remark 3.2.1. The main idea of Jagannathan (1977)'s proof for Proposition 3.2.1 is first to apply Cauchy-Schwarz's inequality for $\mathbb{E}[(X-t)_+]$ with

$$(\mathbb{E}[(X-t)_+])^2 = \left(\int_t^\infty (x-t)\,\mathrm{d}F(x)\right)^2 \leqslant \int_t^\infty \mathrm{d}F(x)\int_t^\infty (x-t)^2\,\mathrm{d}F(x)$$

and obtain the sharp upper bound for $\sup_{F\in\mathcal{L}}\int_t^\infty \mathrm{d}F(x)\int_t^\infty (x-t)^2\,\mathrm{d}F(x)$, and then verify that the upper bound is also the sharp bound for $\sup_{F\in\mathcal{L}}\mathbb{E}^F[(X-t)_+]$. We point out that the arguments and proofs used in Jagannathan (1977) for Proposition 3.2.1 do not work for the worst-case values of $\sup_{F\in\mathcal{L}}\mathbb{E}^F[(X-t)_+^2]$ when $\mathcal{L}$ is any of the uncertainty sets $\mathcal{L}(\mu,\sigma)$, $\mathcal{L}_S(\mu,\sigma)$, and $\mathcal{L}^+(\mu,\sigma)$. If fact, if one applies Cauchy-Schwarz's inequality to $(\mathbb{E}[(X-t)_+^2])^2$, one obtains

$$(\mathbb{E}[(X-t)_+^2])^2 = \left(\int_t^\infty (x-t)^2\,\mathrm{d}F(x)\right)^2 \leqslant \int_t^\infty \mathrm{d}F(x)\int_t^\infty (x-t)^4\,\mathrm{d}F(x).$$

However, the upper bound $\int_t^\infty \mathrm{d}F(x) \int_t^\infty (x-t)^4 \,\mathrm{d}F(x)$ does not provide any useful information for $(\mathbb{E}[(X-t)_+^2])^2$ since $\int_t^\infty (x-t)^4 \,\mathrm{d}F(x)$ may be equal to ∞ when F is in any of the sets $\mathcal{L}(\mu,\sigma)$, $\mathcal{L}_S(\mu,\sigma)$, and $\mathcal{L}^+(\mu,\sigma)$.

Remark 3.2.2. Indeed, Jagannathan (1977) reformulate the supremums of $\mathbb{E}^F[(X-t)_+]$ over $\mathcal{L}(\mu,\sigma)$, $\mathcal{L}_S(\mu,\sigma)$, and $\mathcal{L}^+(\mu,\sigma)$ to the supremums of $\mathbb{E}^F[(X-t)_+]$ over the corresponding uncertainty sets with finite point distributions (usually two-point or three-point distributions). However, Jagannathan (1977) does not explain why the supremums of $\mathbb{E}^F[(X-t)_+]$ over those infinite-dimensional uncertainty sets can be reduced to the finite-dimensional uncertainty sets. We provide detailed proofs of such technique in the Appendix 3.7.

Note that for any $x \in \mathbb{R}$, $(x)_+ - (x)_- = x$, $(x)_+^2 + (x)_-^2 = x^2$, and $(x)_+ = (-x)_-$. Hence, for any $t \in \mathbb{R}$, if $\mathcal{L}^*(\mu,\sigma) \subset \mathcal{F}(\mathbb{R})$ is a set of distributions satisfying that for any $F \in \mathcal{L}^*(\mu,\sigma)$, $\mathbb{E}^F[X] = \mu$ and $\mathbb{E}^F[X^2] = \mu^2 + \sigma^2$, then

$$\sup_{F\in\mathcal{L}^*(\mu,\sigma)} \mathbb{E}^F[(X-t)_+] = \sup_{F\in\mathcal{L}^*(\mu,\sigma)} \mathbb{E}^F[(X-t)_-] + \mu - t, \tag{3.8}$$

$$\inf_{F\in\mathcal{L}^*(\mu,\sigma)} \mathbb{E}^F[(X-t)_+] = \inf_{F\in\mathcal{L}^*(\mu,\sigma)} \mathbb{E}^F[(X-t)_-] + \mu - t, \tag{3.9}$$

$$\sup_{F\in\mathcal{L}^*(\mu,\sigma)} \mathbb{E}^F[(X-t)_+^2] = \sigma^2 + (t-\mu)^2 - \inf_{F\in\mathcal{L}^*(\mu,\sigma)} \mathbb{E}^F[(X-t)_-^2], \tag{3.10}$$

$$\inf_{F\in\mathcal{L}^*(\mu,\sigma)} \mathbb{E}^F[(X-t)_+^2] = \sigma^2 + (t-\mu)^2 - \sup_{F\in\mathcal{L}^*(\mu,\sigma)} \mathbb{E}^F[(X-t)_-^2]. \tag{3.11}$$

In addition, for random variable X, the conditions of $\mathbb{E}[X] = \mu$ and $\mathbb{E}[X^2] = \mu^2 + \sigma^2$ are equivalent to the conditions of $\mathbb{E}[-X] = -\mu$ and $\mathbb{E}[(-X)^2] = \mu^2 + \sigma^2$, and the condition that X is symmetric about μ is equivalent to the condition that $-X$ is symmetric about $-\mu$. Hence, if $\mathcal{L}_0(\mu,\sigma)$ is one of the sets $\mathcal{L}(\mu,\sigma)$ and $\mathcal{L}_S(\mu,\sigma)$, then for $k = 1, 2$,

$$\sup_{F\in\mathcal{L}_0(\mu,\sigma)} \mathbb{E}^F[(X-t)_+^k] = \sup_{F\in\mathcal{L}_0(-\mu,\sigma)} \mathbb{E}^F[(X-(-t))_-^k)], \tag{3.12}$$

$$\inf_{F\in\mathcal{L}_0(\mu,\sigma)} \mathbb{E}^F[(X-t)_+^k)] = \inf_{F\in\mathcal{L}_0(-\mu,\sigma)} \mathbb{E}^F[(X-(-t))_-^k)]. \tag{3.13}$$

We also point out that the downside risk of an investment portfolio is $(t-X)_+ = (X-t)_-$ if X represents the return or gain of the portfolio and t is the target return.

Proposition 3.2.2. *(Chen et al. (2011)) For any $\mu, t \in \mathbb{R}$ and $\sigma \in \mathbb{R}^+$, if the uncertainty set of random variable X is $\mathcal{L}(\mu, \sigma)$, then*

$$\sup_{F \in \mathcal{L}(\mu,\sigma)} \mathbb{E}^F[(X-t)_-^2] = \sigma^2 + (t-\mu)_+^2, \tag{3.14}$$

$$\sup_{F \in \mathcal{L}(\mu,\sigma)} \mathbb{E}^F[(X-t)_+^2] = \sigma^2 + (\mu-t)_+^2. \tag{3.15}$$

Remark 3.2.3. Formula (3.14) was proved by Chen et al. (2011) by the following idea: For $F \in \mathcal{L}(\mu, \sigma)$, (i) yield a lower bound for $\mathbb{E}[(X-t)_+^2]$ by using Jensen's inequality and then obtaining an upper bound for $\mathbb{E}[(X-t)_-^2]$ by using the equation $\mathbb{E}[(X-t)_-^2] = \mathbb{E}[(X-t)^2] - \mathbb{E}[(X-t)_+^2] = \sigma^2 + (\mu-t)^2 - \mathbb{E}[(X-t)_+^2]$, and (ii) verify the upper bound for $\mathbb{E}[(X-t)_-^2]$ is sharp for $\sup_{F \in \mathcal{L}(\mu,\sigma)} \mathbb{E}^F[(X-t)_-^2]$. Formula (3.15) follows directly by applying (3.12) to (3.14). Following the proof of Chen et al. (2011) for (3.14), we can show that $\sup_{F \in \mathcal{L}(\mu,\sigma)} \mathbb{E}^F[(X-t)_+^2] = \sup_{F \in \mathcal{L}^+(\mu,\sigma)} \mathbb{E}^F[(X-t)_-^2]$, which we present in the following corollary.

Corollary 3.2.1. *For $t \in \mathbb{R}$, $\mu > 0$, and $\sigma > 0$, we have*

$$\sup_{F \in \mathcal{L}^+(\mu,\sigma)} \mathbb{E}^F[(X-t)_+^2] \;=\; \sup_{F \in \mathcal{L}(\mu,\sigma)} \mathbb{E}^F[(X-t)_+^2] \;=\; \sigma^2 + (\mu-t)_+^2. \tag{3.16}$$

However, the proofs used in Chen et al. (2011) for the worst-case value $\sup_{F \in \mathcal{L}(\mu,\sigma)} \mathbb{E}^F[(X-t)_-^2]$ do not work for the worst-case values $\sup_{F \in \mathcal{L}_S(\mu,\sigma)} \mathbb{E}^F[(X-t)_+^2]$ since Jensen's inequality cannot yield tight upper bound for $\sup_{F \in \mathcal{L}_S(\mu,\sigma)} \mathbb{E}^F[(X-t)_+^2]$. In Theorem 3.3.1 of this chapter, we derive the closed-form expression for $\sup_{F \in \mathcal{L}_S(\mu,\sigma)} \mathbb{E}^F[(X-t)_+^2]$ by a method different from those used in Jagannathan (1977) and Chen et al. (2011).

3.3 Worst-case target semi-variances with symmetric distributions

In this section, we solve the worst-case target semi-variance over the uncertainty set (3.3) with symmetric distributions, which is the following optimization problem:

$$\sup_{F \in \mathcal{L}_S(\mu,\sigma)} \mathbb{E}^F[(X-t)_+^2]. \tag{3.17}$$

Problem (3.17) is an infinite-dimensional optimization problem. In this section, we first show problem (3.17) can be reformulated as a finite-dimensional optimization problem and then derive the explicit and closed-form expression for the worst-case target semi-variance. To do so, we introduce the definition and notation for a k-point (discrete) distribution.

Definition 3.3.1. Let $[x_1, p_1; \ldots; x_k, p_k]$ denote the probability function of a k-point random variable X, where $\mathbb{P}(X = x_i) = p_i$, $0 \leqslant p_i < 1$, $i = 1, \ldots, k$, $\sum_{i=1}^{k} p_k = 1$, $k \geqslant 2$, and there exists at least a pair (i, j) such that $0 < p_i < 1$, $0 < p_j < 1$, and $1 \leqslant i < j \leqslant k$, which means that a k-point distribution may be a l-point distribution in this chapter, where $2 \leqslant l < k$ and $k \geqslant 3$.

Furthermore, for $k = 2, 3, ...$, we define the subsets of $\mathcal{L}(\mu, \sigma)$, $\mathcal{L}_S(\mu, \sigma)$, $\mathcal{L}^+(\mu, \sigma)$ with k-point distributions as

$$\begin{aligned}
\mathcal{L}_k(\mu, \sigma) &= \{F \in \mathcal{L}(\mu, \sigma) : F \text{ is a } k\text{-point distribution}\}, \\
\mathcal{L}_{k,S}(\mu, \sigma) &= \big\{F \in \mathcal{L}_S(\mu, \sigma) : F \text{ is a } k\text{-point symmetric distribution}\}, \\
\mathcal{L}_k^+(\mu, \sigma) &= \big\{F \in \mathcal{L}^+(\mu, \sigma) : F \text{ is a } k\text{-point non-negative distribution}\}.
\end{aligned}$$

According to Definition 3.3.1, when $k \geqslant 3$, a k-point distribution may also be a l-point (discrete) distribution, where $2 \leqslant l \leqslant k$. Hence, $\mathcal{L}_k^*(\mu, \sigma) \subset \mathcal{L}_{k+1}^*(\mu, \sigma)$ for $k \geqslant 2$, where $\mathcal{L}_k^*(\mu, \sigma)$ is one of the sets $\mathcal{L}_k(\mu, \sigma)$, $\mathcal{L}_{k,S}(\mu, \sigma)$, $\mathcal{L}_k^+(\mu, \sigma)$. To solve problem (3.17), we need the following lemma:

Lemma 3.3.1. *For any $F \in \mathcal{L}(\mu, \sigma)$ with $\sigma > 0$, there exists a two-point distribution $F^* \in \mathcal{L}_2(\mu, \sigma)$ such that the support of F^* belongs to* [ess-inf F, ess-sup F].

Proof. Denote $\underline{F} = \text{ess-inf}\, F$ and $\overline{F} = \text{ess-sup}\, F$. We consider the following two cases:

Case (i): Assume that $\underline{F}$, $\overline{F} \in \mathbb{R}$. For any $F \in \mathcal{L}(\mu, \sigma)$, there exists $p = 1 - q \in (0, 1)$ such that $p\,\underline{F} + q\,\overline{F} = \mu$ as $\mu \in (\underline{F}, \overline{F})$. Let $G = [\underline{F}, p; \overline{F}, q]$ that is a two-point distribution. Clearly, $\mathbb{E}^G[X] = p\,\underline{F} + q\,\overline{F} = \mu$. In addition, note that the number of the sign changes of $F - G$ is one. By Theorem 3.A.44 of Shaked and Shanthikumar (2007), we have that

$F \leqslant_{\mathrm{cx}} G$ and thus, $\mathrm{Var}^G(X) \geqslant \sigma^2$. For any $\varepsilon \geqslant 0$, define the two-point distribution G_ε as $G_\varepsilon = [\underline{F} + \varepsilon q,\, p;\ \overline{F} - \varepsilon p,\, q]$. We have $\mathbb{E}^{G_\varepsilon}[X] = p\,(\underline{F} + \varepsilon q) + q\,(\overline{F} - \varepsilon p) = \mu$ and

$$\mathrm{Var}^{G_\varepsilon}(X) = p\,(\underline{F} + \varepsilon q)^2 + q\,(\overline{F} - \varepsilon p)^2 - \mu^2 = pq\varepsilon^2 - 2pq(\overline{F} - \underline{F})\varepsilon + p\underline{F}^2 + q\overline{F}^2 - \mu^2.$$

Thus, $\mathrm{Var}^{G_\varepsilon}(X)$ is a quadratic function of ε with $\mathrm{Var}^{G_0}(X) = p\underline{F}^2 + q\overline{F}^2 - \mu^2 = \mathrm{Var}^G(X) \geqslant \sigma^2$ and $\mathrm{Var}^{G_{\varepsilon_0}}(X) = 0$, where $\varepsilon_0 = -\frac{-2pq(\overline{F}-\underline{F})}{2pq} = \overline{F} - \underline{F}$. Since $\mathrm{Var}^{G_\varepsilon}(X)$ is continuous and decreasing in $\varepsilon \in [0, \varepsilon_0]$, there must exist $\varepsilon_\sigma \in [0, \varepsilon_0)$ such that $\mathrm{Var}^{G_{\varepsilon_\sigma}}(X) = \sigma^2$. Therefore, $G_{\varepsilon_\sigma} \in \mathcal{L}_2(\mu, \sigma)$ and the support of G_{ε_δ} belongs to $[\underline{F}, \overline{F}]$ as $\underline{F} \leqslant \underline{F} + \varepsilon_\sigma\, q \leqslant \underline{F} + \varepsilon_0 q = \underline{F} + (\overline{F} - \underline{F})q < \overline{F}$ and $\overline{F} \geqslant \overline{F} - \varepsilon_\sigma\, p \geqslant \overline{F} - \varepsilon_0\, p = \overline{F} - (\overline{F} - \underline{F})\, p > \underline{F}$.

Case (ii): Assume that $\underline{F} = -\infty$ or $\overline{F} = \infty$. In this case, it suffices to show that we can find a distribution $F_0 \in \mathcal{L}(\mu, \sigma)$ with bounded support satisfying $[\underline{F}_0, \overline{F}_0] \subseteq [\underline{F}, \overline{F}]$. To see it, we only give the proof of the case that $\underline{F} \in \mathbb{R}$ and $\overline{F} = \infty$ as the other cases can be proved easily by using the similar arguments for the case that $\underline{F} \in \mathbb{R}$ and $\overline{F} = \infty$. For the case $\underline{F} \in \mathbb{R}$ and $\overline{F} = \infty$, note that $\mu > \underline{F}$. For $\alpha \in (0, 1)$, define two-point random variable X_α with the following probability function:

$$\mathbb{P}\left(X_\alpha = \mu - \sigma\sqrt{\frac{1-\alpha}{\alpha}}\right) = \alpha = 1 - \mathbb{P}\left(X_\alpha = \mu + \sigma\sqrt{\frac{\alpha}{1-\alpha}}\right).$$

We have $\mathbb{E}[X_\alpha] = \mu$ and $\mathrm{Var}(X_\alpha) = \sigma^2$ for any $\alpha \in (0, 1)$. There exists $\alpha_0 \in (0, 1)$ such that $\mu - \sigma\sqrt{\frac{1-\alpha_0}{\alpha_0}} > \underline{F}$ as $\mu > \underline{F}$. Thus, the distribution of X_{α_0} is the desired distribution F_0. Therefore, we complete the proof of Lemma 3.3.1. □

Remark 3.3.1. For any $F \in \mathcal{L}_S(\mu, \sigma)$, if $\underline{F} = \text{ess-inf}\, F \in \mathbb{R}$ and $\overline{F} = \text{ess-sup}\, F \in \mathbb{R}$, it holds that $\overline{F} - \mu = \mu - \underline{F}$. In this case, in the proof of Lemma 3.3.1 for case (i), it holds that $p = q = 1/2$ and the two-point distribution $G_\varepsilon = [\underline{F} + \varepsilon q,\, p;\ \overline{F} - \varepsilon p,\, q]$ is symmetric about μ. Hence, there exists a two-point symmetric distribution $F^* \in \mathcal{L}_{2,S}(\mu, \sigma)$ such that the support of F^* belongs to $[\text{ess-inf}\, F,\, \text{ess-sup}\, F]$. Moreover, If $\mu = 0$, then $\overline{F} = -\underline{F}$ and $F^* = 0.5\,\delta_x + 0.5\,\delta_{-x}$ for some $x \in (0,\, \text{ess-sup}\, F]$, where δ_x means a degenerate distribution at x.

Remark 3.3.2. Let $\mathcal{L}^*(\mu, \sigma)$ be one of the sets $\mathcal{L}(\mu, \sigma)$, $\mathcal{L}_S(\mu, \sigma)$, and $\mathcal{L}_{\leqslant\lambda}(\mu, \sigma)$. Note that for any F with $X \sim F$, $\mathbb{E}[X] = \mu$, $\mathrm{Var}(X) = \sigma^2$, it holds that $\mathbb{E}[X - \mu] = 0$,

$\mathrm{Var}(X-\mu)=\sigma^2$, F is a symmetric distribution if and only if $X-\mu$ is symmetric about 0, and $X-t=X-\mu-(t-\mu)$. Hence, if

$$\sup_{F\in\mathcal{L}^*(0,\sigma)}\mathbb{E}^F[h(X-t)]=U(t,\sigma),\qquad \inf_{F\in\mathcal{L}^*(0,\sigma)}\mathbb{E}^F[h(X-t)]=L(t,\sigma),$$

then

$$\sup_{F\in\mathcal{L}^*(\mu,\sigma)}\mathbb{E}^F[h(X-t)]=U(t-\mu,\sigma),\qquad \inf_{F\in\mathcal{L}^*(\mu,\sigma)}\mathbb{E}^F[h(X-t)]=L(t-\mu,\sigma),\tag{3.18}$$

where h is a function defined on $\mathbb{R}$ and $t\in\mathbb{R}$. Therefore, without loss of generality, we can assume $\mu=0$ in the sets $\mathcal{L}(\mu,\sigma)$, $\mathcal{L}_S(\mu,\sigma)$, and $\mathcal{L}_{\leqslant\lambda}(\mu,\sigma)$ when solving optimization problems $\sup_{F\in\mathcal{L}^*(\mu,\sigma)}\mathbb{E}^F[h(X-t)]$ and $\inf_{F\in\mathcal{L}^*(\mu,\sigma)}\mathbb{E}^F[h(X-t)]$.

Theorem 3.3.1. *For $\mu,t\in\mathbb{R}$ and $\sigma\in\mathbb{R}^+$, we have*

$$\sup_{F\in\mathcal{L}_S(\mu,\sigma)}\mathbb{E}^F[(X-t)_+^2]=\begin{cases}\sigma^2+(t-\mu)^2, & t\leqslant\mu-\sigma,\\ \frac{1}{2}(\mu-t+\sigma)^2, & \mu-\sigma<t\leqslant\mu,\\ \frac{\sigma^2}{2}, & t>\mu.\end{cases}\tag{3.19}$$

Proof. According to Remark 3.3.2, we assume $\mu=0$ in the following proof. We consider the three cases that (i) $t\geqslant 0$; (ii) $t\leqslant-\sigma$; and (iii) $-\sigma<t<0$ below.

Case (i): Assume $t\geqslant 0$. In this case, it must hold that $0\leqslant\mathbb{P}(X>t)=\mathbb{E}^F[1_{\{X>t\}}]<1$ for any $F\in\mathcal{L}_S(0,\sigma)$. If $\mathbb{P}(X>t)=\mathbb{E}^F[1_{\{X>t\}}]=0$ for an $F\in\mathcal{L}_S(0,\sigma)$, then $\mathbb{E}^F[(X-t)_+^2]=0$ for this F. Hence, to determine $\sup_{F\in\mathcal{L}_S(0,\sigma)}\mathbb{E}^F[(X-t)_+^2]$, we only need to consider those distributions F in $\mathcal{L}_S(0,\sigma)$ satisfying $0<\mathbb{P}(X>t)=\mathbb{E}^F[1_{\{X>t\}}]<1$. For such a distribution F in $\mathcal{L}_S(0,\sigma)$ satisfying $0<\mathbb{P}(X>t)=\mathbb{E}^F[1_{\{X>t\}}]<1$, let $p=\mathbb{P}(X>t)=\mathbb{P}(X<-t)\in(0,1/2]$, and let G_t be the distribution of the (conditional) random variable $[X|X>t]$, then

$$G_t(x)=\mathbb{P}(X\leqslant x|X>t)=\begin{cases}0, & x\leqslant t,\\ \frac{F(x)-F(t)}{1-F(t)}, & x>t.\end{cases}$$

Note that ess-inf$G_t\geqslant t$ and ess-sup$G_t=$ ess-sup$F$. Denote the mean and variance of G_t by μ_t and σ_t^2, we have $\mu_t=\mathbb{E}[X|X>t]$ and $\mu_t^2+\sigma_t^2=\mathbb{E}[X^2|X>t]$. By applying

Lemma 3.3.1 to the distribution G_t, we know that there exists a two-point distribution $F_t \in \mathcal{L}_2(\mu_t, \sigma_t)$ such that $\mathbb{E}^{F_t}[X_t] = \mu_t$, $\mathbb{E}^{F_t}[X_t^2] = \mu_t^2 + \sigma_t^2$, and the support of F_t belongs to $[\text{ess-inf}G_t, \text{ess-sup}G_t] \subset [t, \text{ess-sup}F]$, where $X_t \sim F_t$. Note that $X_t \geqslant t \geqslant 0$.

Denote the probability function of X_t by $[x_{1,t}, p_{1,t}; x_{2,t}, p_{2,t}]$, where $0 < p_{i,t} < 1$, $i = 1, 2$, and $p_{1,t} + p_{2,t} = 1$. For any $\varepsilon \geqslant 0$, define a random variable X_ε^*, with distribution F_ε^*, as

$$X_\varepsilon^* = (X_t + \varepsilon)1_{\{U > 1-p\}} + 0\, 1_{\{p \leqslant U \leqslant 1-p\}} - (X_t + \varepsilon)1_{\{U < p\}}, \tag{3.20}$$

where $U \sim \mathrm{U}[0, 1]$ is a uniform random variable independent of X_t. Thus, X_ε^* is a five-point random variable valued on $\{-x_{2,t} - \varepsilon,\ -x_{1,t} - \varepsilon,\ 0,\ x_{1,t} + \varepsilon,\ x_{2,t} + \varepsilon\}$. For any $x \geqslant 0$, it holds that

$$\begin{aligned}
\mathbb{P}(X_\varepsilon^* > x) &= \mathbb{P}((X_t + \varepsilon)1_{\{U > 1-p\}} > x) = \mathbb{P}(X_t + \varepsilon > x,\ U > 1 - p) = p\,\mathbb{P}(X_t + \varepsilon > x) \\
&= p\,\mathbb{P}(-X_t - \varepsilon < -x) = \mathbb{P}(-(X_t + \varepsilon)1_{\{U < p\}} < -x) = \mathbb{P}(X_\varepsilon^* < -x),
\end{aligned}$$

where the third equality follows from the independence of X_t and U. Similarly, for any $x < 0$, it holds that

$$\begin{aligned}
\mathbb{P}(X_\varepsilon^* > x) &= \mathbb{P}((U > 1 - p) \cup (p \leqslant U \leqslant 1 - p) \cup (-(X_t + \varepsilon)1_{\{U < p\}} > x)) \\
&= p + 1 - 2p + \mathbb{P}(-(X_t + \varepsilon)1_{\{U < p\}} > x) \\
&= 1 - p + \mathbb{P}(X_t + \varepsilon < -x,\ U < p) = 1 - p + p\,\mathbb{P}(X_t + \varepsilon < -x) \\
&= \mathbb{P}((U < p) \cup (p \leqslant U \leqslant 1 - p) \cup ((X_t + \varepsilon)1_{\{U > 1-p\}} < -x)) = \mathbb{P}(X_\varepsilon^* < -x).
\end{aligned}$$

Therefore X_ε^* is a symmetric random variable at 0 and $\mathbb{E}[X_\varepsilon^*] = 0$. Moreover, note that $X_t \geqslant t \geqslant 0$. Thus, for any $\varepsilon \geqslant 0$, we have

$$\begin{aligned}
\mathbb{E}[(X_\varepsilon^* - t)_+^2] &= \mathbb{E}[\big((X_t + \varepsilon)1_{\{U > 1-p\}} - t\big)^2] = \mathbb{E}[(X_t + \varepsilon - t)^2 1_{\{U > 1-p\}}] \\
&= \mathbb{E}[(X_t + \varepsilon - t)^2]\,\mathbb{P}(U > 1 - p) \\
&\geqslant p\,\mathbb{E}[(X_t - t)^2] \\
&= \mathbb{P}(X > t)\left(\mathbb{E}[X_t^2] - 2t\mathbb{E}[X_t] + t^2\right) \\
&= \mathbb{P}(X > t)\left(\mathbb{E}[X^2 | X > t] - 2t\mathbb{E}[X | X > t] + t^2\right) \\
&= \mathbb{P}(X > t)\,\mathbb{E}^F[(X - t)^2 | X > t] = \mathbb{E}^F[(X - t)_+^2],
\end{aligned}$$

where the inequality follows from $X_t + \varepsilon - t \geqslant X_t - t \geqslant 0$. In addition,

$$\begin{aligned}\mathrm{Var}(X_0^*) &= \mathbb{E}[(X_0^*)^2] = \mathbb{E}[X_t^2\, 1_{\{U>1-p\}} + X_t^2\, 1_{\{U<p\}}] = 2p\,\mathbb{E}[X_t^2] = 2p\,\mathbb{E}[X^2|X>t] \\ &= 2\mathbb{E}[X^2 1_{\{X>t\}}] = \mathbb{E}[X^2 1_{\{X>t\}\cup\{X<-t\}}] \leqslant \mathbb{E}[X^2] = \sigma^2,\end{aligned}$$

where the second equality follows from the independence between X_t and U, the forth equality follows from $p = \mathbb{P}(X > t)$, and the fifth equality follows from that X is symmetric at 0.

Clearly, $\mathrm{Var}(X_\varepsilon^*) = \mathbb{E}[(X_t+\varepsilon)^2\, 1_{\{X>t\}} + (X_t+\varepsilon)^2\, 1_{\{X<-t\}}]$ is a quadratic function of ε with $\mathrm{Var}(X_\varepsilon^*) \to \infty$ as $\varepsilon \to \infty$. There exists $\varepsilon_\delta \geqslant 0$ such that $\mathrm{Var}(X_{\varepsilon_\delta}^*) = \sigma^2$. Hence, the distribution of $X_{\varepsilon_\delta}^*$ belongs to $\mathcal{L}_S(0,\sigma)$ and $\mathbb{E}^F[(X-t)_+^2] \leqslant \mathbb{E}^F[(X_{\varepsilon_\delta}^* - t)_+^2]$, where $X_{\varepsilon_\delta}^*$ has a five-point symmetric distribution about 0. Therefore, for $t > 0$, $\sup_{F\in\mathcal{L}_S(0,\sigma)} \mathbb{E}^F[(X-t)_+^2] = \sup_{F\in\mathcal{L}_{5,S}(0,\sigma)} \mathbb{E}^F[(X-t)_+^2]$. Note that the probability function of the five-point symmetric random variable $X_{\varepsilon_\delta}^*$ has the expression $[-x_2, p_2;\ -x_1, p_1;\ 0, p_0;\ x_1, p_1;\ x_2, p_2]$, where $0 \leqslant t \leqslant x_1 \leqslant x_2$, $0 < p_1 + p_2 \leqslant 1/2$, $0 \leqslant p_1 < 1/2$, $0 \leqslant p_2 < 1/2$, and $0 \leqslant p_0 < 1$. Thus, the problem $\sup_{F\in\mathcal{L}_S(0,\sigma)} \mathbb{E}^F[(X-t)_+^2]$ is equivalent to the problem

$$\begin{aligned}\sup_{(p_1,p_2,x_1,x_2)\in[0,\frac{1}{2}]^2\times\mathbb{R}_+^2} \quad & p_1(x_1-t)^2 + p_2(x_2-t)^2, \\ \text{s.t.} \quad & t \leqslant x_1 \leqslant x_2, \quad 0 < p_1+p_2 \leqslant 1/2, \quad p_1\, x_1^2 + p_2\, x_2^2 = \sigma^2/2.\end{aligned} \tag{3.21}$$

One can verify that the supremum of problem (3.21) is equal to $\sigma^2/2$. To see it, note that for any feasible solution of the problem (3.21), it holds that

$$\begin{aligned}p_1(x_1-t)^2 + p_2(x_2-t)^2 &= p_1x_1^2 + p_2x_2^2 + (p_1+p_2)t^2 - 2t(p_1x_1+p_2x_2) \\ &= \frac{\sigma^2}{2} + t\left[(p_1+p_2)t - 2(p_1x_1+p_2x_2)\right] \leqslant \frac{\sigma^2}{2},\end{aligned}$$

where the inequality follows from that $(p_1+p_2)t - 2(p_1x_1+p_2x_2) \leqslant 0$ as $x_1, x_2 \geqslant t$. On the other hand, for $\varepsilon > 0$ small enough, take $p_1 = 0$, $p_2 = \varepsilon$, $x_1 = t$, $x_2 = \sqrt{\frac{\sigma^2}{2\varepsilon}}$. We have the objective function in (3.21) is

$$p_1(x_1-t)^2 + p_2(x_2-t)^2 = \varepsilon\left(\sqrt{\frac{\sigma^2}{2\varepsilon}} - t\right)^2 = \left(\sqrt{\frac{\sigma^2}{2}} - \sqrt{\varepsilon}t\right)^2 \to \frac{\sigma^2}{2} \quad \text{as } \varepsilon \to 0.$$

Note that for any $F \in \mathcal{L}_S(0,\sigma)$,

$$\mathbb{E}^F[(X-t)_+^2] = \frac{\mathbb{E}^F[(X-t)_+^2] + \mathbb{E}^F[(X+t)_-^2]}{2} < \frac{\mathbb{E}[X^2]}{2} = \frac{\sigma^2}{2},$$

where the first equality follows from the symmetry of F at 0, and the inequality follows from $\mathbb{E}[X^2] = (\mathbb{E}[X_+^2] + \mathbb{E}[X_-^2])/2$ and $\mathbb{E}^F[(X-t)_+^2] + \mathbb{E}^F[(X+t)_-^2]$ is strictly decreasing in $t \geqslant 0$. The supremum $\sigma^2/2$ of problem (3.21) is the limit of $\mathbb{E}^{F\varepsilon}[(X-t)_+^2]$ as $\varepsilon \to 0$, where $F\varepsilon$ is the following three-point symmetric distribution:

$$\left[\mu - \sqrt{\frac{\sigma^2}{2\varepsilon}},\ \varepsilon;\ \ \mu,\ 1-2\varepsilon;\ \ \mu + \sqrt{\frac{\sigma^2}{2\varepsilon}},\ \varepsilon\right].$$

Case (ii): For $t \leqslant -\sigma$, on one hand, note that for any $F \in \mathcal{L}_S(0,\sigma)$, we have

$$\mathbb{E}^F[(X-t)_+^2] \leqslant \mathbb{E}^F[(X-t)_+^2] + \mathbb{E}^F[(X-t)_-^2] = \mathbb{E}^F[(X-t)^2] = \sigma^2 + t^2.$$

On the other hand, take $X \sim F$ as $\mathbb{P}(X=-\sigma) = \mathbb{P}(X=\sigma) = 1/2$. We have $X \geqslant t$ a.s., and $\mathbb{E}[(X-t)_+^2] = \frac{1}{2}(-\sigma-t)^2 + \frac{1}{2}(\sigma-t)^2 = \sigma^2+t^2$. Therefore, we have $\sup_{F\in\mathcal{L}_S(0,\sigma)} \mathbb{E}^F[(X-t)_+^2] = \sigma^2 + t^2$.

Case (iii): For $-\sigma < t < 0$, We first show that for any $F \in \mathcal{L}_S(0,\sigma)$, there exists a six-point distribution $G \in \mathcal{L}_S(0,\sigma)$ such that $\mathbb{E}^F[(X-t)_+^2] \leqslant \mathbb{E}^G[(X-t)_+^2]$. Note that in the case that $-\sigma < t < 0$, we must have $\mathbb{P}(t \leqslant X \leqslant -t) < 1$ as otherwise $\|X\| \leqslant -t$ a.s. which yields a contradiction with $\sigma > -t$. We then have $p := \mathbb{P}(X > -t) > 0$, and $\mathbb{P}(X < t) = p > 0$ by symmetry. Applying Lemma 3.3.1 to the distribution of $[X|X > -t]$, we see that there exists a two-point distribution F_t with support on $[-t,\infty)$, such that $\mathbb{E}^{F_t}[X_t] = \mathbb{E}^F[X|X > -t]$ and $\mathbb{E}^{F_t}[X_t^2] = \mathbb{E}^F[X^2|X > -t]$, where $X_t \sim F_t$ and $X_t \geqslant -t > 0$.

If $p < 1/2$, then $\mathbb{P}(t \leqslant X \leqslant -t) > 0$, and by applying Remark 3.3.1 to $[X|t \leqslant X \leqslant -t]$, we see that there exists $x \in (0,-t]$ such that $\mathbb{E}^{G_x}[Y_t^2] = \mathbb{E}[X^2|t \leqslant X \leqslant -t]$, where $G_x = \delta_x/2 + \delta_{-x}/2$ and $Y_t \sim G_x$. Define

$$X_\varepsilon^* = (X_t+\varepsilon)1_{\{U>1-p\}} + x1_{\{1/2<U\leqslant 1-p\}} - x1_{\{p\leqslant U\leqslant 1/2\}} - (X_t+\varepsilon)1_{\{U<p\}}, \tag{3.22}$$

where $x \in (0, -t]$, $\varepsilon \geqslant 0$, and $U \sim \mathrm{U}[0,1]$ is independent from X_t and $p = \mathbb{P}(X > -t) \in (0, 1/2]$. Otherwise, if $p = 1/2$, we still employ the definition of the random variable X_ε^* by (3.22), which reduces to

$$X_\varepsilon^* = (X_t + \varepsilon)1_{\{U > 1-p\}} - (X_t + \varepsilon)1_{\{U < p\}}.$$

In both cases, X_ε^* is a six-point random variable valued on $\{-x_{2,t}-\varepsilon,\ -x_{1,t}-\varepsilon,\ -x,\ x, x_{1,t}+\varepsilon,\ x_{2,t}+\varepsilon\}$, where $x \in (0, -t]$ and $-t \leqslant x_{1,t} < x_{2,t}$.

Similar to Case (i), we can verify that the distribution of X_ε^* is symmetric about 0 and that $\mathbb{E}[X_\varepsilon^*] = 0$ and $\mathbb{E}[(X_\varepsilon^* - t)_+^2] \geqslant \mathbb{E}^F[(X-t)_+^2]$ for any $\varepsilon \geqslant 0$, and $\mathrm{Var}[(X_0^*)] \leqslant \sigma^2$ for $\varepsilon = 0$. Moreover, one can verify that $\mathrm{Var}[(X_\varepsilon^*)]$ is a quadratic function of $\varepsilon \geqslant 0$. There exists $\varepsilon_\delta \geqslant 0$ such that the distribution of $X_{\varepsilon_\delta}^*$ belongs to $\mathcal{L}_S(0,\sigma)$. Denote the set $\mathcal{L}_{6,S}^*(0,\sigma)$ by

$$\begin{aligned}\mathcal{L}_{6,S}^*(0,\sigma) = \{&[-x_3, p_3;\ -x_2, p_2;\ -x_1, p_1;\ x_1, p_1; x_2, p_2;\ x_3, p_3] :\\ &0 < x_1 \leqslant -t \leqslant x_2 < x_3,\ p_1 + p_2 + p_3 = 1/2,\ 0 \leqslant p_i \leqslant 1/2,\ \text{for } i = 1,2,3\}.\end{aligned}$$

Then, $\mathcal{L}_{6,S}^*(0,\sigma) \subset \mathcal{L}_{6,S}(0,\sigma)$ and the distribution of $X_{\varepsilon_\delta}^*$ belongs to $\mathcal{L}_{6,S}^*(0,\sigma)$. Hence, it holds that $\sup_{F\in\mathcal{L}_S(0,\sigma)} \mathbb{E}^F[(X-t)_+^2] = \sup_{F\in\mathcal{L}_{6,S}^*(0,\sigma)} \mathbb{E}^F[(X-t)_+^2]$. Note that $\mathbb{E}^F[(X-t)_+^2] + \mathbb{E}[(X-t)_-^2] = \sigma^2 + t^2$ is fixed for any $F \in \mathcal{L}(0,\sigma)$. Thus, we have

$$\sup_{F\in\mathcal{L}_{6,S}^*(0,\sigma)} \mathbb{E}^F[(X-t)_+^2] = \sigma^2 + t^2 - \inf_{F\in\mathcal{L}_{6,S}^*(0,\sigma)} \mathbb{E}^F[(X-t)_-^2]. \tag{3.23}$$

Note that the problem $\inf_{F\in\mathcal{L}_{6,S}^*(0,\sigma)} \mathbb{E}^F[(X-t)_-^2]$ is equivalent to the problem

$$\begin{aligned}\inf_{(p_1,p_2,p_3,x_1,x_2,x_3)\in[0,1/2]^3\times\mathbb{R}_+^3} \quad & p_2(x_2+t)^2 + p_3(x_3+t)^2, \\ \text{s.t.} \quad & 0 < x_1 \leqslant -t \leqslant x_2 \leqslant x_3,\ p_1+p_2+p_3 = 1/2, \\ & p_1x_1^2 + p_2x_2^2 + p_3x_3^2 = \sigma^2/2,\end{aligned} \tag{3.24}$$

as $(-x)_- = (x)_+$, $x_2 + t \geqslant 0$, and $x_3 + t \geqslant 0$.

By $-t < \sigma$, we know that the constraints in problem (3.24) can not be satisfied at $x_2 = x_3 = -t$. Note that for any feasible solution of the problem (3.24), $(p_1, p_2, p_3, x_1, x_2, x_3)$, if

$x_1 < -t$, then take $\delta \in (0, -t - x_1)$ and $(p_1, p_2, p_3, x_1 + \delta, x_2 - \delta_1, x_3 - \delta_2)$, where $\delta_1, \delta_2 \geqslant 0$ satisfy $-t \leqslant x_2 - \delta_1 \leqslant x_3 - \delta_2$ and $p_1(x_1+\delta)^2 + p_2(x_2-\delta_1)^2 + p_3(x_3-\delta_2)^2 = \sigma^2/2$. It holds that the value of the objective function at the new feasible solution $(p_1, p_2, p_3, x_1 + \delta, x_2 - \delta_1, x_3 - \delta_2)$ is strictly smaller than that at $(p_1, p_2, p_3, x_1, x_2, x_3)$. Therefore, the infimum of problem (3.24) is attainable at $x_1 = -t$, which implies that problem (3.24) is equivalent to

$$\begin{aligned} \min_{(p_1,p_2,p_3,x_2,x_3)\in[0,1/2]^3\times\mathbb{R}^2_+} \quad & p_2(x_2+t)^2 + p_3(x_3+t)^2, \\ \text{s.t.} \quad & -t \leqslant x_2 \leqslant x_3,\ p_1+p_2+p_3 = 1/2, \\ & p_1t^2 + p_2x_2^2 + p_3x_3^2 = \sigma^2/2, \end{aligned} \tag{3.25}$$

which, together with $\mathbb{E}^F[(X-t)_+^2] + \mathbb{E}[(X-t)_-^2] = \sigma^2 + t^2$ and (3.18), implies that the problem $\sup_{F\in\mathcal{L}_S(\mu,\sigma)} \mathbb{E}^F[(X-t)_+^2]$ is equivalent to the following problem

$$\begin{aligned} \sigma^2 + (\mu - t)^2 - \min_{(p_1,p_2,p_3,x_2,x_3)\in[0,1)^3\times\mathbb{R}^2_+} \quad & p_2(x_2+t-\mu)^2 + p_3(x_3+t-\mu)^2, \\ \text{s.t.} \quad & -(t-\mu) \leqslant x_2 \leqslant x_3,\ p_1+p_2+p_3 = 1/2, \\ & p_1t^2 + p_2x_2^2 + p_3x_3^2 = \sigma^2/2, \end{aligned} \tag{3.26}$$

One can verify that for any feasible solution of (3.25), it holds that

$$\begin{aligned} p_2(x_2+t)^2 + p_3(x_3+t)^2 &= p_1(x_1+t)^2 + p_2(x_2+t)^2 + p_3(x_3+t)^2 \\ &= p_1x_1^2 + p_2x_2^2 + p_3x_3^2 + (p_1+p_2+p_3)t^2 + 2t(p_1x_1 + p_2x_2 + p_3x_3) \\ &= \frac{\sigma^2}{2} + \frac{t^2}{2} + 2t(p_1x_1 + p_2x_2 + p_3x_3). \end{aligned}$$

Noting that $t < 0$, the problem (3.25) is equivalent to

$$\begin{aligned} \max_{(p_1,p_2,p_3,x_2,x_3)\in[0,1/2]^3\times\mathbb{R}^2_+} \quad & 2p_1x_1 + 2p_2x_2 + 2p_3x_3, \\ \text{s.t.} \quad & -t \leqslant x_2 \leqslant x_3,\ 2p_1+2p_2+2p_3 = 1, \\ & 2p_1t^2 + 2p_2x_2^2 + 2p_3x_3^2 = \sigma^2, \end{aligned} \tag{3.27}$$

For any feasible solution, we have $2p_1x_1 + 2p_2x_2 + 2p_3x_3 \leqslant \sqrt{2p_1t^2 + 2p_2x_2^2 + 2p_3x_3^2} = \sigma$. On the other hand, take $p_1 = 0,\ p_2 = 0,\ x_1 = x_2 = -t,\ p_3 = 1/2,\ x_3 = \sigma > -t$. We have

$2p_1x_1 + 2p_2x_2 + 2p_3x_3 = \sigma$. Therefore, we have the supremum of problem (3.27) is σ, and thus, the infimum of problem (3.25) is $\frac{\sigma^2}{2} + \frac{t^2}{2} + t\sigma = \frac{(\sigma+t)^2}{2}$. It thus follows from (3.23) that the supremum of the objective function in problem (3.26) is $\sigma^2 + t^2 - \frac{(\sigma+t)^2}{2} = \frac{(\sigma-t)^2}{2}$. □

3.4 Applications to robust portfolio selection

In this section, we consider the applications of the worst-case target semi-variances derived in Sections 3.3 to robust portfolio selection problems.

Let $\boldsymbol{X}^\top = (X_1, ..., X_d) \in \mathbb{R}^d$ be a random vector representing the loss vector in an investment portfolio with X_i being the loss in investing in an asset i, $i = 1, ..., d$. The loss of the portfolio is $\boldsymbol{w}^\top \boldsymbol{X} = \sum_{i=1}^d w_i X_i$, where $\boldsymbol{w} = (w_1, ..., w_d) \in \mathbb{R}^d$ with w_i being the allocation/selection on asset i. Without loss of generality, we assume the total wealth of an investor to be 1 so that $\boldsymbol{w}^\top \boldsymbol{e} = 1$, where $\boldsymbol{e}^\top = (1, ..., 1)$ is a d-dimensional unit vector. Here we denote the set of portfolios that allows for short-selling as $\mathcal{W} = \{\boldsymbol{w} \in \mathbb{R}^d : \boldsymbol{w}^\top \boldsymbol{e} = 1\}$.

In the classical mean-variance (M-V) model, the mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$ of the loss vector $\boldsymbol{X}$ are assumed to be known, which essentially assume that the 'true' (joint) distribution G of loss vector $\boldsymbol{X}$ is unknown and only the mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$ of the loss vector $\boldsymbol{X}$ are available. In other words, the possible (joint) distribution G of loss vector $\boldsymbol{X}$ is assumed to belong to the following set of distributions:

$$\mathcal{M}(\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \big\{G \in \mathcal{F}(\mathbb{R}^d) : \mathbb{E}[\boldsymbol{X}] = \boldsymbol{\mu},\, \mathrm{Cov}[\boldsymbol{X}] = \boldsymbol{\Sigma}\big\}, \tag{3.28}$$

where $\mathcal{F}(\mathbb{R}^d)$ is the set of all d-dimensional distributions defined on $\mathbb{R}^d$. For any $G \in \mathcal{M}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, $\mathbb{E}^G[\boldsymbol{w}^\top \boldsymbol{X}] = \boldsymbol{w}^\top \boldsymbol{\mu}$ and $\mathrm{Var}^G(\boldsymbol{w}^\top \boldsymbol{X}) = \boldsymbol{w}^\top \boldsymbol{\Sigma} \boldsymbol{w}$ of the portfolio loss $X = \boldsymbol{w}^\top \boldsymbol{X}$ are deterministic functions of $\boldsymbol{w}$. The M-V portfolio selection model can be formulated as

$$\begin{aligned} &\min_{\boldsymbol{w} \in \mathcal{W}} \sup_{G \in \mathcal{M}(\boldsymbol{\mu}, \boldsymbol{\Sigma})} \mathrm{Var}^G(\boldsymbol{w}^\top \boldsymbol{X}) = \min_{\boldsymbol{w} \in \mathcal{W}} \boldsymbol{w}^\top \boldsymbol{\Sigma} \boldsymbol{w} \\ &\text{s.t.} \quad \mathbb{E}^G[\boldsymbol{w}^\top \boldsymbol{X}] = \boldsymbol{w}^\top \boldsymbol{\mu} \leqslant \nu, \end{aligned} \tag{3.29}$$

where $\mathbb{E}^G[\boldsymbol{w}^\top \boldsymbol{X}] = \boldsymbol{w}^\top \boldsymbol{\mu} \leqslant \nu$ is equivalent to $\mathbb{E}^G[-\boldsymbol{w}^\top \boldsymbol{X}] = -\boldsymbol{w}^\top \boldsymbol{\mu} \geqslant -\nu$ that represents a constraint on the expected return of the portfolio. However, if the distribution G of $\boldsymbol{X}$

is uncertain and belongs to $\mathcal{M}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, then the expected downside loss or expected regret $\mathbb{E}^G[(\boldsymbol{w}^\top \boldsymbol{X} - t)_+]$ and the target semi-variance $\mathbb{E}^G[(\boldsymbol{w}^\top \boldsymbol{X} - t)_+^2]$ are also uncertain. To incorporate the symmetric information of loss vectors and minimize the worst-case target semi-variance of the portfolio loss, we propose the following target semi-variance (TSV)-based robust portfolio selection models: The mean-target semi-variance with symmetric information (M-TSV-S) robust portfolio selection model, which is formulated as

$$\begin{aligned} &\min_{\boldsymbol{w} \in \mathcal{W}} \sup_{G \in \mathcal{M}_S(\boldsymbol{\mu}, \boldsymbol{\Sigma})} \mathbb{E}^G[(\boldsymbol{w}^\top \boldsymbol{X} - t)_+^2] \\ &\text{s.t.} \quad \mathbb{E}^G[\boldsymbol{w}^\top \boldsymbol{X}] = \boldsymbol{w}^\top \boldsymbol{\mu} \leqslant \nu, \end{aligned} \tag{3.30}$$

where

$$\mathcal{M}_S(\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \big\{ G \in \mathcal{M}(\boldsymbol{\mu}, \boldsymbol{\Sigma}) : G \text{ is symmetric} \big\} \tag{3.31}$$

and ν is a risk level or equivalently $-\nu$ is a desirable minimum level for the expected return.

We first give the definition of symmetric random vector or multivariate symmetric distribution. To do so, denote the set of all Borel measurable sets in $\mathbb{R}^d$ by $\mathfrak{B}(\mathbb{R}^d)$. For a set $A \subset \mathbb{R}^d$ and a vector $\boldsymbol{a} \in \mathbb{R}^d$, denote $-A$ by $-A = \{\boldsymbol{x} \in \mathbb{R}^d : -\boldsymbol{x} \in A\}$ and denote $A - \boldsymbol{a}$ by $A - \boldsymbol{a} = \{\boldsymbol{x} - \boldsymbol{a} \in \mathbb{R}^d : \boldsymbol{x} \in A\}$.

Definition 3.4.1. The joint distribution G of a random vector $\boldsymbol{X} \in \mathbb{R}^d$ is said to be symmetric if there exists a vector $\boldsymbol{a} \in \mathbb{R}^d$ such that $\mathbb{P}(\boldsymbol{X} - \boldsymbol{a} \in B) = \mathbb{P}(\boldsymbol{X} - \boldsymbol{a} \in -B)$ under the distribution G, for all $B \in \mathfrak{B}(\mathbb{R}^d)$. If such a vector $\boldsymbol{a}$ exists, random vector $\boldsymbol{X}$ or its distribution is said to be symmetric about $\boldsymbol{a}$.

Intuitively, random vector $\boldsymbol{X}$ is symmetric about $\boldsymbol{a}$ if and only if $\boldsymbol{X} - \boldsymbol{a}$ is symmetric about the origin of $\mathbb{R}^d$. Examples of the continuous multivariate symmetric distributions include multivariate normal distributions, multivariate t-distributions, multivariate elliptical distributions, and many others. In addition, a constant random vector is also symmetric according to Definition 3.4.1.

Lemma 3.4.1. *(i) If random vector $\boldsymbol{X} \in \mathbb{R}^d$ or its distribution G is symmetric about $\boldsymbol{a}$, then for any vector $\boldsymbol{w} \in \mathbb{R}^d$, the distribution of $\boldsymbol{w}^\top \boldsymbol{X}$ is symmetric about $\boldsymbol{w}^\top \boldsymbol{a}$.*

(ii) If d-dimensional random vectors $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ are independent and symmetric about $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$, respectively, then $\boldsymbol{X}_1 + \boldsymbol{X}_2$ is symmetric about $\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2$.

Proof. (i) For any $x \in \mathbb{R}$, we have $\mathbb{P}\{\boldsymbol{w}^\top \boldsymbol{X} - \boldsymbol{w}^\top \boldsymbol{a} > x\} = \mathbb{P}\{\boldsymbol{w}^\top(\boldsymbol{X} - \boldsymbol{a}) > x\} = \mathbb{P}\{\boldsymbol{X} \in B\} = \mathbb{P}\{\boldsymbol{X} - \boldsymbol{a} \in B_{\boldsymbol{a}}\}$, where $B = \{\boldsymbol{y} \in \mathbb{R}^d : \boldsymbol{w}^\top(\boldsymbol{y} - \boldsymbol{a}) > x\}$ and $B_{\boldsymbol{a}} = B - \boldsymbol{a} = \{\boldsymbol{y} - \boldsymbol{a} \in \mathbb{R}^d : \boldsymbol{w}^\top(\boldsymbol{y} - \boldsymbol{a}) > x\} = \{\boldsymbol{x} \in \mathbb{R}^d : \boldsymbol{w}^\top \boldsymbol{x} > x\}$. Since $\boldsymbol{X}$ is is symmetric around $\boldsymbol{a}$ and $B_{\boldsymbol{a}}$ is a Borel measurable set in $\mathbb{R}^d$, we have $\mathbb{P}\{\boldsymbol{w}^\top \boldsymbol{X} - \boldsymbol{w}^\top \boldsymbol{a} > x\} = \mathbb{P}\{\boldsymbol{X} - \boldsymbol{a} \in B_{\boldsymbol{a}}\} = \mathbb{P}\{\boldsymbol{X} - \boldsymbol{a} \in -B_{\boldsymbol{a}}\}$, where

$$-B_{\boldsymbol{a}} = \{\boldsymbol{x} \in \mathbb{R}^d : -\boldsymbol{w}^\top \boldsymbol{x} > x\} = \{\boldsymbol{x} \in \mathbb{R}^d : \boldsymbol{w}^\top \boldsymbol{x} < -x\} = \{\boldsymbol{y} - \boldsymbol{a} \in \mathbb{R}^d : \boldsymbol{w}^\top(\boldsymbol{y} - \boldsymbol{a}) < -x\}.$$

Therefore, $\mathbb{P}\{\boldsymbol{w}^\top \boldsymbol{X} - \boldsymbol{w}^\top \boldsymbol{a} > x\} = \mathbb{P}\{\boldsymbol{X} - \boldsymbol{a} \in -B_{\boldsymbol{a}}\} = \mathbb{P}\{\boldsymbol{w}^\top \boldsymbol{X} - \boldsymbol{w}^\top \boldsymbol{a} < -x\}$, which means that the distribution of $\boldsymbol{w}^\top \boldsymbol{X}$ is symmetric around $\boldsymbol{w}^\top \boldsymbol{a}$ by Definition 3.2.1.

(ii) Note that for a random variable $\boldsymbol{X}$ is symmetric about $\boldsymbol{\mu}$ if and only if $\boldsymbol{X} - \boldsymbol{\mu}$ is symmetric at $\mathbf{0}$. Without loss of generality, assume that $\boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 = \mathbf{0}$. Thus, for any measurable set $B \in \mathfrak{B}(\mathbb{R}^d)$, it holds that $\mathbb{P}(\boldsymbol{X}_1 + \boldsymbol{X}_2 \in B) = \mathbb{P}(-\boldsymbol{X}_1 - \boldsymbol{X}_2 \in B) = \mathbb{P}(\boldsymbol{X}_1 + \boldsymbol{X}_2 \in -B)$, where the first equality follows that $-\boldsymbol{X}_1 \overset{\mathrm{d}}{=} \boldsymbol{X}_1$, $-\boldsymbol{X}_2 \overset{\mathrm{d}}{=} \boldsymbol{X}_2$, and $-\boldsymbol{X}_1$ and $-\boldsymbol{X}_2$ are independent which imply that $\boldsymbol{X}_1 + \boldsymbol{X}_2 \overset{\mathrm{d}}{=} -\boldsymbol{X}_1 - \boldsymbol{X}_2$. □

Lemma 3.4.1 (i) demonstrates that if $\boldsymbol{X}$ is a random vector with a symmetric multivariate distribution, then any linear combination of $\boldsymbol{X}$ has a one-dimensional symmetric distribution as defined in Definition 3.2.1. In addition, Lemma 3.4.1 (ii) implies that symmetric properties of random vectors are preserved under the sum of independent random vectors.

We now denote the the multivariate mean-covariance uncertainty set with symmetric distributions by $\mathcal{M}_S(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ that is defined in (3.31). Moreover, for a given $\boldsymbol{w} \in \mathbb{R}^d$, define $\mathcal{L}_{\boldsymbol{w},S}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ as the one-dimensional distribution set generated from the distribution of $\boldsymbol{w}^\top \boldsymbol{X}$ when the joint distribution of $\boldsymbol{X}$ belongs to $\mathcal{M}_S(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, namely

$$\mathcal{L}_{\boldsymbol{w},S}(\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \{F_{\boldsymbol{w}^\top \boldsymbol{X}} \in \mathcal{F}(\mathbb{R}) : \text{The joint distribution } G \text{ of } \boldsymbol{X} \text{ belongs to } \mathcal{M}_S(\boldsymbol{\mu}, \boldsymbol{\Sigma})\}. \tag{3.32}$$

Lemma 3.4.2. *If the covariance matrix $\boldsymbol{\Sigma}$ of the loss random vector $\boldsymbol{X}$ is positive definite and $\boldsymbol{w} \neq \mathbf{0}$, then*

$$\mathcal{L}_{\boldsymbol{w},S}(\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \mathcal{L}_S\left(\boldsymbol{w}^\top \boldsymbol{\mu}, \sqrt{\boldsymbol{w}^\top \boldsymbol{\Sigma} \boldsymbol{w}}\right) \tag{3.33}$$

and

$$\sup_{G\in\mathcal{M}_S(\boldsymbol{\mu},\boldsymbol{\Sigma})}\mathbb{E}^G[(\boldsymbol{w}^\top\boldsymbol{X}-t)_+^2]=\sup_{F\in\mathcal{L}_S\left(\boldsymbol{w}^\top\boldsymbol{\mu},\sqrt{\boldsymbol{w}^\top\boldsymbol{\Sigma}\boldsymbol{w}}\right)}\mathbb{E}^F[(X-t)_+^2], \tag{3.34}$$

where set $\mathcal{L}_S(\mu,\sigma)$ *is defined in* (3.3) *for any* $\mu\in\mathbb{R}$ *and any* $\sigma\in\mathbb{R}^+$*, and* X *is a random variable with a distribution belonging to* $\mathcal{L}_S\left(\boldsymbol{w}^\top\boldsymbol{\mu},\sqrt{\boldsymbol{w}^\top\boldsymbol{\Sigma}\boldsymbol{w}}\right)$.

Proof. For any distribution $F_{\boldsymbol{w}^\top\boldsymbol{X}}\in\mathcal{L}_{\boldsymbol{w},S}(\boldsymbol{\mu},\boldsymbol{\Sigma})$ with the joint distribution G of $\boldsymbol{X}$ belonging to $\mathcal{M}_S(\boldsymbol{\mu},\boldsymbol{\Sigma})$, we have $\mathbb{E}^G[\boldsymbol{w}^\top\boldsymbol{X}]=\boldsymbol{w}^\top\boldsymbol{\mu}$ and $\mathrm{Cov}^G[\boldsymbol{X}]=\boldsymbol{w}^\top\boldsymbol{\Sigma}\boldsymbol{w}$. In addition, by Lemma 3.4.1(i), we see that $\boldsymbol{w}^\top\boldsymbol{X}$ is symmetric as $\boldsymbol{X}$ is symmetric. Hence, $F_{\boldsymbol{w}^\top\boldsymbol{X}}\in\mathcal{L}_S\left(\boldsymbol{w}^\top\boldsymbol{\mu},\sqrt{\boldsymbol{w}^\top\boldsymbol{\Sigma}\boldsymbol{w}}\right)$. Thus, $\mathcal{L}_{\boldsymbol{w},S}(\boldsymbol{\mu},\boldsymbol{\Sigma})\subseteq\mathcal{L}_S\left(\boldsymbol{w}^\top\boldsymbol{\mu},\sqrt{\boldsymbol{w}^\top\boldsymbol{\Sigma}\boldsymbol{w}}\right)$. Next, we prove $\mathcal{L}_S\left(\boldsymbol{w}^\top\boldsymbol{\mu},\sqrt{\boldsymbol{w}^\top\boldsymbol{\Sigma}\boldsymbol{w}}\right)\subseteq\mathcal{L}_{\boldsymbol{w},S}(\boldsymbol{\mu},\boldsymbol{\Sigma})$. Similar to the proof of Chen et al. (2011, Lemma 2.4), for $\boldsymbol{w}\neq 0\in\mathbb{R}^d$ and any distribution $F\in\mathcal{L}_S\left(\boldsymbol{w}^\top\boldsymbol{\mu},\sqrt{\boldsymbol{w}^\top\boldsymbol{\Sigma}\boldsymbol{w}}\right)$, we construct a d-dimensional random vector $\boldsymbol{X}^*$ as

$$\boldsymbol{X}^*=\frac{((\boldsymbol{w}^\top\boldsymbol{\Sigma}\boldsymbol{w})\boldsymbol{\Sigma}-\boldsymbol{\Sigma}ww^\top\boldsymbol{\Sigma})^{\frac{1}{2}}\boldsymbol{Z}}{\sqrt{\boldsymbol{w}^\top\boldsymbol{\Sigma}\boldsymbol{w}}}+\frac{(Y-\boldsymbol{w}^\top\boldsymbol{\mu})\boldsymbol{\Sigma}\boldsymbol{w}}{\boldsymbol{w}^\top\boldsymbol{\Sigma}\boldsymbol{w}}+\boldsymbol{\mu},$$

where Y is a random variable with the distribution F and $\boldsymbol{Z}$ is a d-dimensional standard normal random vector independent of Y. Then, $\boldsymbol{w}^\top\boldsymbol{X}^*=Y$, $\mathbb{E}[\boldsymbol{X}^*]=\boldsymbol{\mu}$, and $\mathrm{Cov}[\boldsymbol{X}^*]=\boldsymbol{\Sigma}$. In addition, by Lemma 3.4.1(i) and (ii), we see that $\boldsymbol{X}^*$ is symmetric and thus $\boldsymbol{w}^\top\boldsymbol{X}^*=Y$ is symmetric as well. Hence, the joint distribution of $\boldsymbol{X}^*$ belongs to $\mathcal{M}_S(\boldsymbol{\mu},\boldsymbol{\Sigma})$ and the distribution F of $\boldsymbol{w}^\top\boldsymbol{X}^*=Y$ belongs to $\mathcal{L}_{\boldsymbol{w},S}(\boldsymbol{\mu},\boldsymbol{\Sigma})$, which mean that $\mathcal{L}_S\left(\boldsymbol{w}^\top\boldsymbol{\mu},\sqrt{\boldsymbol{w}^\top\boldsymbol{\Sigma}\boldsymbol{w}}\right)\subseteq\mathcal{L}_{\boldsymbol{w},S}(\boldsymbol{\mu},\boldsymbol{\Sigma})$. Therefore, we conclude that $\mathcal{L}_{\boldsymbol{w},S}(\boldsymbol{\mu},\boldsymbol{\Sigma})=\mathcal{L}_S\left(\boldsymbol{w}^\top\boldsymbol{\mu},\sqrt{\boldsymbol{w}^\top\boldsymbol{\Sigma}\boldsymbol{w}}\right)$. It is obvious that

$$\sup_{G\in\mathcal{M}_S(\boldsymbol{\mu},\boldsymbol{\Sigma})}\mathbb{E}^G[(\boldsymbol{w}^\top\boldsymbol{X}-t)_+^2]=\sup_{F\in\mathcal{L}_{\boldsymbol{w},S}(\boldsymbol{\mu},\boldsymbol{\Sigma})}\mathbb{E}^F[(X-t)_+^2],$$

which, together with (3.33), implies (3.34). □

To better present the optimal portfolio selections derived in this chapter, we introduce notations u, v_0, v_1, v_2 as follows:

$$u=(\boldsymbol{e}^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{e})(\boldsymbol{\mu}^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu})-(\boldsymbol{e}^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu})^2,\ \ v_0=\frac{\boldsymbol{e}^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{e}}{u},\ \ v_1=\frac{\boldsymbol{e}^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}}{u},\ \ v_2=\frac{\boldsymbol{\mu}^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}}{u}, \tag{3.35}$$

where $\boldsymbol{w}, \boldsymbol{\mu} \in \mathbb{R}^d$ with $\boldsymbol{w}^\top \boldsymbol{e} = 1$, and $\boldsymbol{\Sigma}$ is a $d \times d$ positive definite matrix. Note that for any $\boldsymbol{\mu} \in \mathbb{R}^d$, it holds $u \geqslant 0$ since $\boldsymbol{\Sigma}$ is a positive definite matrix. However, to guarantee that the optimal solutions exist, we assume that $\boldsymbol{\mu}$ and $\boldsymbol{e}$ are not linearly dependent, or equivalently, assume that for any $c \in \mathbb{R}$, $\boldsymbol{\mu} \neq c\,\boldsymbol{e}$. This assumption implies $u > 0$ and is also used in the classical M-V portfolio selection model.

To present the optimal solution to problem (3.30), we define

$$h_{S,t}(\mu, \sigma) = \sup_{F \in \mathcal{L}_S(\mu,\sigma)} \mathbb{E}^F[(X-t)_+^2].$$

By Theorem 3.3.1, we can write $h_{S,t}(\mu, \sigma)$ as a function of σ with the following expression:

(i) If $\mu > t$, then

$$h_{S,t}(\mu, \sigma) = \begin{cases} \sigma^2 + (t-\mu)^2, & 0 < \sigma \leqslant \mu - t, \\ \frac{1}{2}(\mu - t + \sigma)^2, & \sigma > \mu - t. \end{cases} \tag{3.36}$$

(ii) If $\mu \leqslant t$, then

$$h_{S,t}(\mu, \sigma) = \frac{\sigma^2}{2}, \quad \sigma > 0. \tag{3.37}$$

In addition, define

$$\xi_1^* = \arg\min_{\xi \leqslant t} \; h_{S,t}\left(\xi, \sqrt{v_0\xi^2 - 2v_1\xi + v_2}\right), \tag{3.38}$$

$$h_{S,t}^1(\xi_1^*) = h_{S,t}\left(\xi_1^*, \sqrt{v_0(\xi_1^*)^2 - 2v_1\xi_1^* + v_2}\right) = \frac{1}{2}\big(v_0(\xi_1^*)^2 - 2v_1\xi_1^* + v_2\big), \tag{3.39}$$

$$\xi_2^* = \arg\min_{t \leqslant \xi \leqslant v} \; h_{S,t}\left(\xi, \sqrt{v_0\xi^2 - 2v_1\xi + v_2}\right), \tag{3.40}$$

$$h_{S,t}^2(\xi_2^*) = h_{S,t}\left(\xi_2^*, \sqrt{v_0(\xi_2^*)^2 - 2v_1\xi_2^* + v_2}\right). \tag{3.41}$$

Note that by (3.37),

$$\xi_1^* = \arg\min_{\xi \leqslant t} \frac{1}{2}\big(v_0\xi^2 - 2v_1\xi + v_2\big) = \min\left\{\frac{v_1}{v_0}, t\right\}. \tag{3.42}$$

By (3.36), if $t < \nu$, we see that $h_{S,t}\left(\xi, \sqrt{v_0\xi^2 - 2v_1\xi + v_2}\right)$ is a continuous function of ξ on $[t, \nu]$. Thus, there exists ξ_2^* such that

$$\min_{t \leqslant \xi \leqslant \nu} \; h_{S,t}\left(\xi, \sqrt{v_0\xi^2 - 2v_1\xi + v_2}\right) = h_{S,t}\left(\xi_2^*, \sqrt{v_0(\xi_2^*)^2 - 2v_1\xi_2^* + v_2}\right) = h_{S,t}^2(\xi_2^*).$$

Proposition 3.4.1. *Assume the covariance matrix $\boldsymbol{\Sigma}$ of the loss random vector $\boldsymbol{X}$ is positive definite. Then, problems (3.30) has a unique solution $\boldsymbol{w}^*_{S,\nu,t}$ that has the following expression:*

$$\boldsymbol{w}^*_{S,\nu,t} = (\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}, \quad \boldsymbol{\Sigma}^{-1}\boldsymbol{e}) \begin{pmatrix} v_0 & -v_1 \\ -v_1 & v_2 \end{pmatrix} \begin{pmatrix} \xi^* \\ 1 \end{pmatrix}. \tag{3.43}$$

Here, ξ^* in (3.43) has the following expressions:

(i) If $t \geqslant \nu$, then $\xi^* = \min\left\{\frac{v_1}{v_0}, \nu\right\}$.

(ii) If $t < \nu$ and $h^1_{S,t}(\xi^*_1) \leqslant h^2_{S,t}(\xi^*_2)$, then $\xi^* = \min\left\{\frac{v_1}{v_0}, t\right\}$.

(iii) If $t < \nu$ and $h^1_{S,t}(\xi^*_1) > h^2_{S,t}(\xi^*_2)$, then $\xi^* = \xi^*_2 = \arg\min_{t\leqslant\xi\leqslant\nu} h_{S,t}\left(\xi, \sqrt{v_0\xi^2 - 2v_1\xi + v_2}\right)$.

Proof. By Lemma 3.4.2, for the positive definite matrix $\boldsymbol{\Sigma}$ and $\boldsymbol{w} \neq \mathbf{0}$, the inner optimization problem of (3.30) reduces to the problem

$$\sup_{F\in\mathcal{L}_S\left(\boldsymbol{w}^\top\boldsymbol{\mu}, \sqrt{\boldsymbol{w}^\top\boldsymbol{\Sigma}\boldsymbol{w}}\right)} \mathbb{E}^F[(X-t)^2_+] = h_{S,t}\left(\boldsymbol{w}^\top\boldsymbol{\mu}, \sqrt{\boldsymbol{w}^\top\boldsymbol{\Sigma}\boldsymbol{w}}\right),$$

which has been solved in Theorem 3.3.1. Hence, problem (3.30) is equivalent to the following problem:

$$\begin{aligned} \min_{\boldsymbol{w}\in\mathcal{W}} \quad & h_{S,t}\left(\boldsymbol{w}^\top\boldsymbol{\mu}, \sqrt{\boldsymbol{w}^\top\boldsymbol{\Sigma}\boldsymbol{w}}\right), \\ \text{s.t.} \quad & \boldsymbol{w}^\top\boldsymbol{\mu} \leqslant \nu. \end{aligned} \tag{3.44}$$

Problem (3.44) is equivalent to

$$\begin{aligned} \min_{\xi\in\mathbb{R},\, \boldsymbol{w}\in\mathcal{W}} \quad & h_{S,t}\left(\xi, \sqrt{\boldsymbol{w}^\top\boldsymbol{\Sigma}\boldsymbol{w}}\right), \\ \text{s.t.} \quad & \boldsymbol{w}^\top\boldsymbol{\mu} = \xi \leqslant \nu, \end{aligned}$$

which can be expressed as the following problem:

$$\begin{aligned} \min_{\xi\in\mathbb{R}} \quad & \min_{\boldsymbol{w}\in\mathbb{R}^d,\, \boldsymbol{w}^\top\boldsymbol{e}=1,\, \boldsymbol{w}^\top\boldsymbol{\mu}=\xi} h_{S,t}\left(\xi, \sqrt{\boldsymbol{w}^\top\boldsymbol{\Sigma}\boldsymbol{w}}\right), \\ \text{s.t.} \quad & \xi \leqslant \nu. \end{aligned} \tag{3.45}$$

By (3.36) and (3.37), it is easy to see that for any $\xi \in \mathbb{R}$, $h_{S,t}(\xi, \sqrt{\sigma^2})$ is increasing in σ^2. Therefore,

$$\min_{\boldsymbol{w}\in\mathbb{R}^d,\, \boldsymbol{w}^\top \boldsymbol{e}=1,\, \boldsymbol{w}^\top \boldsymbol{\mu}=\xi} h_{S,t}\left(\xi,\, \sqrt{\boldsymbol{w}^\top \boldsymbol{\Sigma} \boldsymbol{w}}\right) = h_{S,t}\left(\xi, \sqrt{\min_{\boldsymbol{w}\in\mathbb{R}^d,\, \boldsymbol{w}^\top \boldsymbol{e}=1,\, \boldsymbol{w}^\top \boldsymbol{\mu}=\xi} \boldsymbol{w}^\top \boldsymbol{\Sigma} \boldsymbol{w}}\right), \quad (3.46)$$

which means that

$$\boldsymbol{w}_\xi^* = \operatorname*{arg\,min}_{\boldsymbol{w}\in\mathbb{R}^d,\, \boldsymbol{w}^\top \boldsymbol{e}=1,\, \boldsymbol{w}^\top \boldsymbol{\mu}=\xi} h_{S,t}\left(\xi, \sqrt{\boldsymbol{w}^\top \boldsymbol{\Sigma} \boldsymbol{w}}\right) = \operatorname*{arg\,min}_{\boldsymbol{w}\in\mathbb{R}^d,\, \boldsymbol{w}^\top \boldsymbol{e}=1,\, \boldsymbol{w}^\top \boldsymbol{\mu}=\xi} \boldsymbol{w}^\top \boldsymbol{\Sigma} \boldsymbol{w}.$$

It is well-known that

$$\boldsymbol{w}_\xi^* = \operatorname*{arg\,min}_{\boldsymbol{w}\in\mathbb{R}^d,\, \boldsymbol{w}^\top \boldsymbol{e}=1,\, \boldsymbol{w}^\top \boldsymbol{\mu}=\xi} \boldsymbol{w}^\top \boldsymbol{\Sigma} \boldsymbol{w} = (\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} \quad \boldsymbol{\Sigma}^{-1}\boldsymbol{e}) \begin{pmatrix} v_0 & -v_1 \\ -v_1 & v_2 \end{pmatrix} \begin{pmatrix} \xi \\ 1 \end{pmatrix}$$

and $(\boldsymbol{w}_\xi^*)^\top \boldsymbol{\Sigma} \boldsymbol{w}_\xi^* = v_0\xi^2 - 2v_1\xi + v_2$. Therefore, problem (3.45) is reduced to the following one-variance optimization problem:

$$\min_{\xi \leqslant \nu} \; h_{S,t}\left(\xi, \sqrt{v_0\xi^2 - 2v_1\xi + v_2}\right). \quad (3.47)$$

(a) If $t \geqslant \nu$, by (3.37), we have $\min_{\xi\leqslant\nu} \; h_{S,t}\left(\xi, \sqrt{v_0\xi^2 - 2v_1\xi + v_2}\right) = \min_{\xi\leqslant\nu} \frac{1}{2}\left(v_0\xi^2 - 2v_1\xi + v_2\right)$. Let $\xi^* = \arg\min_{\xi\leqslant\nu} \frac{1}{2}\left(v_0\xi^2 - 2v_1\xi + v_2\right)$. It is easy to see that $\xi^* = \min\left\{\frac{v_1}{v_0}, \nu\right\}$.

(b) If $t < \nu$, we have

$$\begin{aligned} &\min_{\xi<\nu} \; h_{S,t}\left(\xi, \sqrt{v_0\xi^2 - 2v_1\xi + v_2}\right) \\ &= \min\left\{\min_{\xi\leqslant t} \; h_{S,t}\left(\xi, \sqrt{v_0\xi^2 - 2v_1\xi + v_2}\right), \; \min_{t\leqslant\xi\leqslant\nu} \; h_{S,t}\left(\xi, \sqrt{v_0\xi^2 - 2v_1\xi + v_2}\right)\right\}, \end{aligned}$$

where by (3.37),

$$\min_{\xi\leqslant t} \; h_{S,t}\left(\xi, \sqrt{v_0\xi^2 - 2v_1\xi + v_2}\right) = \frac{1}{2}\min_{\xi\leqslant t} \; v_0\xi^2 - 2v_1\xi + v_2 = \frac{1}{2}\left(v_0(\xi_1^*)^2 - 2v_1\xi_1^* + v_2\right),$$

and $\xi_1^* = \arg\min_{\xi\leqslant t} v_0\xi^2 - 2v_1\xi + v_2 = \min\left\{\frac{v_1}{v_0}, t\right\}$. In addition, by (3.36), we can notice that $h_{S,t}\left(\xi, \sqrt{v_0\xi^2 - 2v_1\xi + v_2}\right)$ is a continuous function of ξ on $[t, \nu]$. Thus, there exists ξ_2^* such that $\min_{t\leqslant\xi\leqslant\nu} \; h_{S,t}\left(\xi, \sqrt{v_0\xi^2 - 2v_1\xi + v_2}\right) = h_{S,t}\left(\xi_2^*, \sqrt{v_0(\xi_2^*)^2 - 2v_1\xi_2^* + v_2}\right)$.

By combining cases (a) and (b), we complete the proof. □

Remark 3.4.1. The optimal portfolio selection $\boldsymbol{w}_{\nu}^{*}$ to the classical M-V problem (3.29) has the following expression:

$$\boldsymbol{w}_{\nu}^{*} = (\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}, \quad \boldsymbol{\Sigma}^{-1}\boldsymbol{e}) \begin{pmatrix} v_0 & -v_1 \\ -v_1 & v_2 \end{pmatrix} \begin{pmatrix} \min\left\{\frac{v_1}{v_0}, \nu\right\} \\ 1 \end{pmatrix}. \tag{3.48}$$

For the proof of above result, we refer Markowitz (1959) and Markowitz et al. (1993). Proposition 3.4.1 shows that if an investor is too conservative and sets a very low target return $-t$ or a very high threshold loss level t, say $t \geqslant \nu$, then $\xi^{*} = \min\left\{\frac{v_1}{v_0}, \nu\right\}$ and the optimal strategy $\boldsymbol{w}_{S,\nu,t}^{*}$ is the same as the optimal strategy (3.48) derived from the classical M-V model (3.29). However, if an investor has a higher return target $-t$ or a lower threshold loss level t, say $t < \min\{\frac{v_1}{v_0}, \nu\}$, then $\xi^{*} = \min\left\{\frac{v_1}{v_0}, t\right\} = t$ or $\xi^{*} = \xi_2^{*}$ and the optimal strategy $\boldsymbol{w}_{S,\nu,t}^{*}$ is different from the optimal strategy (3.48) derived from the classical M-V model (3.29). As illustrated in the numerical experiments given in Section 3.5 of this chapter, the portfolio performance with the strategy derived from (3.30) outperforms the portfolio performance with the strategy derived from the classical M-V model (3.29).

3.5 Numerical experiments with real financial data

In this section, we conduct a numerical study using real financial data to calculate the optimal portfolios derived in Section 3.4 and compare the investment performances of the optimal portfolios with several existing portfolio selection methods related to the models proposed in this chapter. For this study, we select 12 stocks from the four largest sectors (Technology, Health Care, Financials, and Consumer Discretionary) of the S&P 500, choosing three with the highest market capitalizations from each sector.[2] We use data from a four-year period starting from January 1, 2019, to January 1, 2023, which include 1008 observations of daily stock prices. The daily losses are expressed by percentage and calculated by $l_t = -(V_{t+1} - V_t)/V_t$, where V_t is the close price on trading day t. Note that the positive value represents the loss and negative value represents the gain.

[2]Selected stocks are AAPL, MSFT, GOOG, JPM, BAC, BRK-B, PFE, JNJ, UNH, HD, TSLA, AMZN

We aim to compare investment performance across several existing models related to the models proposed in this chapter, including:

(a) TSV model: Minimizing the target semi-variance of the portfolio loss and formulated as (see e.g., Chen et al. (2011))

$$\min_{\boldsymbol{w}\in\mathcal{W}} \sup_{G\in\mathcal{M}(\boldsymbol{\mu},\boldsymbol{\Sigma})} \mathbb{E}^G[(\boldsymbol{w}^\top \boldsymbol{X} - t)_+^2]. \tag{3.49}$$

(b) M-TSV-S model: Minimizing the target semi-variance of the portfolio loss, incorporating the symmetric information of loss vectors and the constraint on the expected portfolio loss, and proposed in (3.30) and solved in Proposition 3.4.1.

(c) M-V model: The classical mean-variance model formulated in (3.29).

(d) S&P 500 index: Investing all money into S&P 500 index, regarding as a passive investment strategy.

We construct portfolio rebalancing strategies by solving optimization problems with models (a)-(g) listed above. The experiment is set up as follows. The initial portfolio $\boldsymbol{w}_0^*$ is calculated on January 3, 2020 using the data from January 2, 2019 to January 2, 2020 as the in-sample dataset (252 trading days).[3] We compute the out-of-sample portfolio returns as $-\boldsymbol{w}_0^{*\top}\hat{\boldsymbol{l}}_0$, where $\hat{\boldsymbol{l}}_0$ represents the daily loss on January 3, 2021. We proceed to optimize the portfolio selections on a daily basis using a rolling window approach and subsequently rebalance the portfolio. This involves using the preceding 755 trading days to calculate the optimal portfolio $\boldsymbol{w}_t^*$ for trading day t, serving as an updated portfolio for each trading day starting from January 3, 2020. The resulting portfolio returns $-\boldsymbol{w}_t^{*\top}\hat{\boldsymbol{l}}_t$ for trading day t are obtained using the out-of-sample return vector $\hat{\boldsymbol{l}}_t$ and the rebalanced portfolio weights $\boldsymbol{w}_t^*$. In the TSV-based models (a) and (b), the parameters u, v_0, v_1, and v_2 defined in (3.35) are also updated as the data rolls forward. These parameters rely on sample mean and sample covariance, which evolve with rebalance process over time. To conduct the numerical experiment, parameters need to be chosen for models (a)-(c). We give the

[3]No trading data is available for January 1.

following general guidelines for selecting parameters: the target return $-t$ in models (a) and (b), the maximum expected loss level ν in models (b) and (c).

(1) Positive (negative) values of a portfolio loss random variable X represent losses (returns). Initially, it might seem logical for an investor to choose a higher target return $-t$ or a lower threshold loss level t to expect better investment performance. However, the expected excess profit $\mathbb{E}[(X-t)_-]$ increases with t, while the expected downside risk $\mathbb{E}[(X-t)_+]$ decreases with t. Thus, opting for a lower threshold loss level t results in higher expected downside risk. Consequently, a reasonable choice for the target return $-t$ is to set it slightly larger than the sample mean of the daily returns of the selected stocks. Equivalently, the threshold loss level t can be set slightly smaller than the sample mean of the daily losses of the selected stocks.

(2) Note that $\mathbb{E}[X] \leqslant \nu$ is equivalent to $\mathbb{E}[-X] \geqslant -\nu$; where $\mathbb{E}[-X]$ represents the expected return, and $-t$ is the target return. Thus, it is natural to require $-t \geqslant -\nu$ or equivalently $\nu \geqslant t$. Additionally, a high value of ν is not desirable. Therefore, a reasonable choice for the maximum expected loss level ν is to set it slightly larger than t.

According to the above guidelines, in this experiment, we choose a target return $-t = 0.003$ for all the TSV-based models (a)-(c); a maximum expected loss level $\nu = -0.001$ for models (b) and (c).

Figure 3.2 shows the cumulative wealth of a portfolio comprised of the four selected stocks under the strategies derived from models (a)-(d) listed above. It is evident that all the strategies, except for the TSV model, outperform the passive investment strategy of the S&P 500 index. We can also see from Figure 3.2 that the M-TSV-S model (d) outperforms all the other models listed above. The additional information regarding the symmetry of the loss distribution (as indicated Figure 3.1 for several stocks), greatly improves the practicality of the models proposed in this chapter. Therefore, incorporating symmetric information into portfolio selection models can improve investment performance when using the models proposed in this chapter.

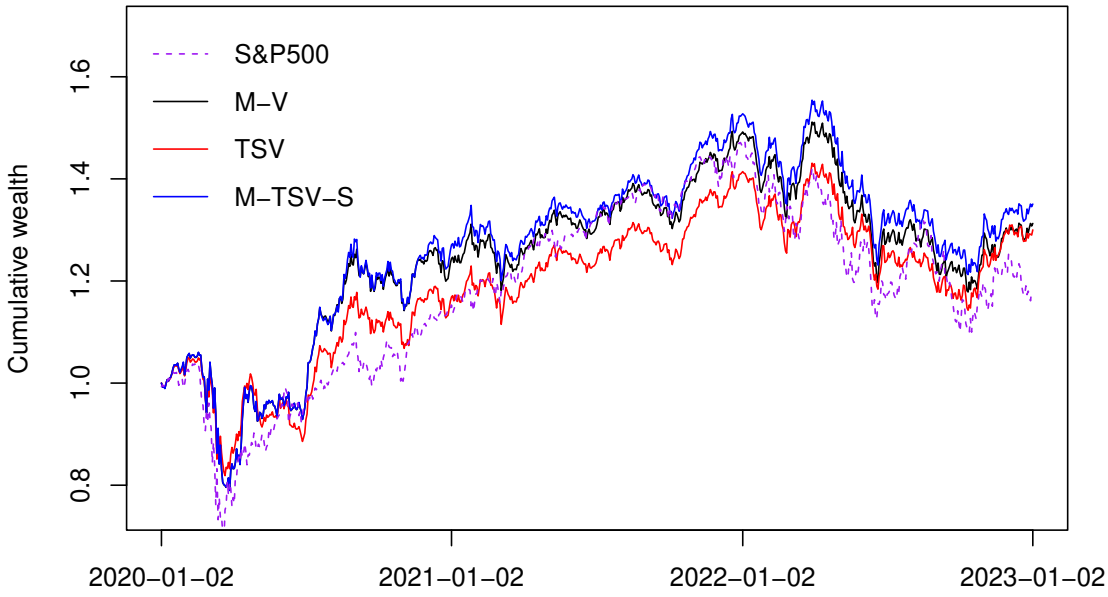


Figure 3.2: Cumulative wealth comparison across portfolio rebalancing strategies based on models (a)-(c). The target return $t = -0.003$ for all the TSV-based models; risk level $\nu = -0.001$ for the M-V and M-TSV-S models.

3.6 Concluding remarks

In this chapter, we explore the worst-case target semi-variance of a random loss within mean-variance uncertainty sets, considering additional distributional information such as symmetry and non-negativity of the random loss. The main contribution of this chapter is that we complement the study of Chen et al. (2011), where the worst-case target semi-variance was derived for an arbitrary random loss. We extend this by deriving the worst-case target semi-variances for symmetric or non-negative losses, thus obtaining results corresponding to the worst-case expected regrets investigated in Jagannathan (1977).

As illustrated in numerical experiments, the investment performance with the optimal strategies derived from the proposed models outperforms the classical mean-variance strategy and the passive investment strategy (investing S&P 500 index only). We believe the results and models developed in this chapter have more potential, and we will explore more applications in future research.

3.7 Appendix: Technical details

Lemma 3.7.1. *For any $F \in \mathcal{L}(0,\sigma)$ and Borel sets A, $B \subset \mathbb{R}$ satisfying that $A \cap B = \emptyset$, denote $p = \mathbb{E}^F[1_{\{X\in B\}}] = \mathbb{P}(X \in B)$, $q = \mathbb{E}^F[1_{\{X\in A\}}] = \mathbb{P}(X \in A)$. If $0 < p < 1$, $q = 1-p$, let F_ε be a two-point distribution of the two-point random variable X_ε that is defined as*

$$X_\varepsilon = (x_1 + \varepsilon p)1_{\{X\in A\}} + (x_2 - \varepsilon q)1_{\{X\in B\}}, \tag{3.50}$$

where $x_1 = \mathbb{E}[X|X \in A]$, $x_2 = \mathbb{E}[X|X \in B]$, and $\varepsilon \geqslant 0$, then there exists a constant $\varepsilon_\sigma \geqslant 0$ such that $F_{\varepsilon_\sigma} \in \mathcal{L}(0,\sigma)$.

Proof. For any $F \in \mathcal{L}(0,\sigma)$, it is easy to see that

$$\begin{aligned}
\mathbb{E}[X_\varepsilon] &= \mathbb{E}[X_\varepsilon|X \in A]\, q + \mathbb{E}[X_\varepsilon|X \in B]\, p = (x_1+\varepsilon p)q + (x_2 - \varepsilon q)p \\
&= x_1 q + x_2 p = \mathbb{E}^F[X|X \in A]\, q + \mathbb{E}^F[X|X \in B]\, p = \mathbb{E}^F[X] = 0, \\
\mathrm{Var}(X_\varepsilon) = \mathbb{E}[X_\varepsilon^2] &= \mathbb{E}[X_\varepsilon^2|X \in A]\, q + \mathbb{E}[X_\varepsilon^2|X \in B]\, p = (x_1+\varepsilon p)^2\, q + (x_2 - \varepsilon q)^2\, p \\
&= x_1^2\, q + x_2^2\, p + 2pq(x_1 - x_2)\varepsilon + pq\varepsilon^2 \triangleq g(\varepsilon).
\end{aligned}$$

Note that for any random variable Y with $\mathbb{E}[Y] = \mu$ and $\mathrm{Var(Y)} = \sigma^2$, if $0 < \mathbb{P}(Y \in A) < 1$, then

$$(\mathbb{E}[Y|Y \in A])^2 \leqslant \mathbb{E}[Y^2|Y \in A]. \tag{3.51}$$

Thus, by (3.51), we have $g(0) = x_1^2\, q + x_2^2\, p \leqslant \mathbb{E}^F[X^2|X \in A]\, q + \mathbb{E}^F[X^2|X \in B]\, p = \mathbb{E}^F[X^2] = \sigma^2$. Since $g(\varepsilon)$ is a quadratic function with $g(\infty) = \infty$, there exists an $\varepsilon_\sigma \geqslant 0$ such that $g(\varepsilon_\sigma) = \mathrm{Var(X}_{\varepsilon_\sigma}) = \sigma^2$. Therefore, $F_{\varepsilon_\sigma} \in \mathcal{L}(0,\sigma)$. □

Lemma 3.7.2. *For any* $t \in \mathbb{R}$,

$$\sup_{F \in \mathcal{L}(0,\sigma)} \mathbb{E}^F[(X-t)_+] = \sup_{F \in \mathcal{L}_2(0,\sigma)} \mathbb{E}^F[(X-t)_+]. \tag{3.52}$$

Proof. For any $F \in \mathcal{L}(0,\sigma)$, since $\mathbb{E}^F[X] = 0$ and $\mathrm{Var}^F(X) = \sigma^2 > 0$, there are the following three possible cases about $\mathbb{E}^F[1_{\{X \leqslant t\}}] = \mathbb{P}(X \leqslant t)$.

- **Case 1**: Assume $0 < \mathbb{P}(X \leqslant t) < 1$. Let $A = (X > t)$ and $B = (X \leqslant t)$. By Lemma 3.7.1 and its notations, $x_1 = \mathbb{E}^F[X|X > t] \geqslant t$, $x_2 = \mathbb{E}^F[X|X \leqslant t] \leqslant t$, $0 < p = \mathbb{P}(X \leqslant t) < 1$, $q = \mathbb{P}(X > t) = 1 - p$, and there exists a two-point distribution F_{ε_σ} such that $F_{\varepsilon_\sigma} \in \mathcal{L}(0,\sigma)$. With this distribution,

$$\begin{aligned} \mathbb{E}^{F_{\varepsilon_\sigma}}[(X_{\varepsilon_\sigma} - t)_+] &= (x_1 + \varepsilon_\sigma p - t)_+ \, q + (x_2 - \varepsilon_\sigma q - t)_+ \, p \\ &= (x_1 + \varepsilon_\sigma p - t) \, q \\ &\geqslant (x_1 - t) \, q = \mathbb{E}^F[(X-t)_+]. \end{aligned} \tag{3.53}$$

- **Case 2**: Assume $t > 0$ and $\mathbb{P}(X \leqslant t) = 1$. In this case, $\mathbb{E}^F[(X-t)_+] = 0$ and let $A = (X \leqslant 0)$ and $B = (0 < X \leqslant t)$. By Lemma 3.7.1 and its notations, $0 < p = \mathbb{P}(0 < X \leqslant t) < 1$, $q = \mathbb{P}(X \leqslant 0) = 1 - p$, and there exists a two-point distribution F_{ε_σ} such that $F_{\varepsilon_\sigma} \in \mathcal{L}(0,\sigma)$. With this distribution,

$$\mathbb{E}^{F_{\varepsilon_\sigma}}[(X_{\varepsilon_\sigma} - t)_+] \geqslant 0 = \mathbb{E}^F[(X-t)_+]. \tag{3.54}$$

- **Case 3**: Assume $t < 0$ and $\mathbb{P}(X \leqslant t) = 0$. In this case, $\mathbb{E}^F[(X-t)_+] = \mathbb{E}^F(X-t) = -t$ as $\mathbb{P}(X > t) = 1$ and let $A = (X > 0)$ and $B = (t < X \leqslant 0)$. By Lemma 3.7.1 and its notations, $0 < p = \mathbb{P}(t < X \leqslant 0) < 1$, $q = \mathbb{P}(X > 0) = 1 - p$, and there exists a two-point distribution F_{ε_σ} such that $F_{\varepsilon_\sigma} \in \mathcal{L}(0,\sigma)$. With this distribution and Jensen's inequality,

$$\mathbb{E}[(X_{\varepsilon_\sigma} - t)_+] \geqslant (\mathbb{E}[X_{\varepsilon_\sigma}] - t)_+ = (0-t)_+ = -t = \mathbb{E}^F[(X-t)_+]. \tag{3.55}$$

By combining all the three cases (3.53)-(3.55), we see that for any $F \in \mathcal{L}(0,\sigma)$, there exists a two-point distribution $F_{\varepsilon_\sigma} \in \mathcal{L}(0,\sigma)$ such that $\mathbb{E}[(X_{\varepsilon_\sigma}-t)_+] \geqslant \mathbb{E}[(X-t)_+]$, which implies that

$$\sup_{F\in\mathcal{L}_2(0,\sigma)} \mathbb{E}^F[(X-t)_+] \geqslant \sup_{F\in\mathcal{L}(0,\sigma)} \mathbb{E}^F[(X-t)_+], \tag{3.56}$$

Since $\mathcal{L}_2(0,\sigma) \subset \mathcal{L}(0,\sigma)$, we have $\sup_{F\in\mathcal{L}_2(0,\sigma)} \mathbb{E}^F[(X-t)_+] \leqslant \sup_{F\in\mathcal{L}(0,\sigma)} \mathbb{E}^F[(X-t)_+]$, which, together with (3.56), implies (3.52) holds. □

In this chapter, for a set $A \in \mathbb{R}$, $A > 0$ means $x > 0$ for any $x \in A$ and $A < 0$ means $x < 0$ for any $x \in A$. We point out that a two-point symmetric distribution about 0 can as viewed as a degenerated three-point distribution about 0 if the probability mass function at 0 is 0. In this chapter, a degenerated three-point symmetric distribution about 0 or a two-point symmetric distribution about 0 is also called a three-point symmetric distribution about 0.

Lemma 3.7.3. *For any $F \in \mathcal{L}_S(0,\sigma)$ and Borel sets A_1, A_2, $A_3 \subset \mathbb{R}$ satisfying that $A_1 > 0$, $A_3 < 0$, $A_1 = -A_3$, $0 \in A_2$, and A_2 is a symmetric set about 0, $A_i \cap A_j = \emptyset$ for any $1 \leqslant i < j \leqslant 3$, denote $p_i = \mathbb{E}^F[1_{\{X\in A_i\}}] = \mathbb{P}(X \in A_i)$, $i = 1,2,3$. If $p_1 + p_2 + p_3 = 1$, $0 < p_1 = p_3 < 1$, $0 \leqslant p_2 < 1$, let F_ε be the three-point symmetric distribution of the three-point random variable X_ε that is is defined as*

$$X_\varepsilon = (x_1+\varepsilon)1_{\{X\in A_1\}} + x_2 1_{\{X\in A_2\}} + (x_3-\varepsilon)1_{\{X\in A_3\}}, \tag{3.57}$$

where $x_i = \mathbb{E}[X|X \in A_i]$ for $i = 1,2,3$, and $\varepsilon \geqslant 0$, then there exists a constant $\varepsilon_\sigma \geqslant 0$ such that the three-point symmetric distribution $F_{\varepsilon_\sigma} \in \mathcal{L}_S(0,\sigma)$.

Proof. Under the assumptions of Lemma 3.7.3, if $0 < p_2 < 1$, then X_ε is a three-point symmetric random variable about 0; if $p_2 = 0$, then X_ε is reduced to $X_\varepsilon = (x_1+\varepsilon)1_{\{X\in A_1\}} + (x_3-\varepsilon)1_{\{X\in A_3\}}$, which is a two-point symmetric random variable about 0. In Lemma 3.7.3 and its proof, if $p_2 = 0$, $x_2 = \mathbb{E}^F[X|X \in A_2]$ is read as 0; if $p_2 > 0$, $x_2 = \mathbb{E}^F[X|X \in A_2] = 0$ as X is symmetric random variable about 0. In addition, $x_1 = -x_3$ and $p_1 = p_3$.

It is easy to see that for any $F \in \mathcal{L}_S(0,\sigma)$,

$$\begin{aligned}\mathbb{E}^F[X_\varepsilon] &= (x_1+\varepsilon)p_1 + x_2p_2 + (x_3-\varepsilon)p_3 = x_1p_1 + x_2p_2 + x_3p_3 = \mathbb{E}^F[X] = 0,\\ \mathrm{Var}(X_\varepsilon) &= \mathbb{E}^F[X_\varepsilon^2] = (x_1+\varepsilon)^2\,p_1 + x_2^2\,p_2 + (x_3-\varepsilon)^2\,p_3\\ &= x_1^2\,p_1 + x_2^2\,p_2 + x_3^2\,p_3 + 2(x_1p_1 - x_3p_3)\varepsilon + (p_1+p_3)\varepsilon^2 \triangleq g(\varepsilon).\end{aligned}$$

Thus, by (3.51), we have $g(0) = x_1^2\,p_1 + x_2^2\,p_2 + x_3^2\,p_3 \leqslant \mathbb{E}^F[X^2] = \sigma^2$. Since $g(\varepsilon)$ is a quadratic function with $g(\infty) = \infty$, there exists a constant $\varepsilon_\sigma \geqslant 0$ such that $g(\varepsilon_\sigma) = \mathrm{Var}(\mathrm{X}_{\varepsilon_\sigma}) = \sigma^2$. Therefore, $F_{\varepsilon_\sigma} \in \mathcal{L}(0,\sigma)$. □

Lemma 3.7.4. *For any $t \in \mathbb{R}$,*

$$\sup_{F\in\mathcal{L}_S(0,\sigma)} \mathbb{E}^F[(X-t)_+] = \sup_{F\in\mathcal{L}_{3,S}(0,\sigma)} \mathbb{E}^F[(X-t)_+]. \tag{3.58}$$

Proof. We consider the cases $t \geqslant 0$ and $t < 0$.

- **Case 1**: Assume that $t \geqslant 0$. For any $F \in \mathcal{L}_S(0,\sigma)$, let $X \sim F$, there are the following two possible cases about $\mathbb{P}(X > t)$, $\mathbb{P}(-t \leqslant X \leqslant t)$, $\mathbb{P}(X < -t)$.
  - **Case 1.1:** Assume $0 < \mathbb{P}(X > t) < 1$, $0 \leqslant \mathbb{P}(-t \leqslant X \leqslant t) < 1$, $0 < \mathbb{P}(X < -t) < 1$. Let $A_1 = (X > t)$ and $A_2 = (-t \leqslant X \leqslant t)$, $A_3 = (X < -t)$. Then by Lemma 3.7.3 and its notations, we know that $x_1 = \mathbb{E}^F[X|X > t] \geqslant t$, $x_2 = \mathbb{E}^F[X|-t \leqslant X \leqslant t] = 0$, $x_3 = \mathbb{E}^F[X|X < -t] \leqslant 0$, and there exists a three-point symmetric distribution F_{ε_σ} such that $F_{\varepsilon_\sigma} \in \mathcal{L}_S(0,\sigma)$. With this distribution F_{ε_σ},

$$\begin{aligned}\mathbb{E}^{F_{\varepsilon_\sigma}}[(X_{\varepsilon_\sigma}-t)_+] &= (x_1+\varepsilon_\sigma-t)_+\,p_1 + (x_2-t)_+\,p_2 + (x_3-\varepsilon_\sigma-t)_+\,p_3\\ &= (x_1+\varepsilon_\sigma-t)\,p_1\\ &\geqslant (x_1-t)\,p_1 = \mathbb{E}^F[(X-t)_+].\end{aligned} \tag{3.59}$$

  - **Case 1.2:** Assume $\mathbb{P}\{-t \leqslant X \leqslant t\} = 1$. In this case, $\mathbb{E}^F[(X-t)_+] = 0$. Let $A_1 = (0 < X \leqslant t)$, $A_2 = (X = 0)$, $A_3 = (-t \leqslant X < 0)$. Then by Lemma 3.7.3 and its notations, there exists a three-point symmetric distribution F_{ε_σ} such that $F_{\varepsilon_\sigma} \in \mathcal{L}_S(0,\sigma)$, With this distribution F_{ε_σ},

$$\mathbb{E}^{F_{\varepsilon_\sigma}}[(X_{\varepsilon_\sigma}-t)_+] \geqslant 0 = \mathbb{E}^F[(X-t)_+]. \tag{3.60}$$

- **Case 2**: Assume that $t < 0$. For any $F \in \mathcal{L}_S(0,\sigma)$, let $X \sim F$, there are the following two possible cases about $\mathbb{P}(X > -t)$, $\mathbb{P}(t \leqslant X \leqslant -t)$, $\mathbb{P}(X < t)$.

  - **Case 2.1:** Assume $0 < \mathbb{P}(X > -t) < 1$, $0 \leqslant \mathbb{P}(t \leqslant X \leqslant -t) < 1$, and $0 < \mathbb{P}(X < t) < 1$. Let $A_1 = (X > -t)$, $A_2 = (t \leqslant X \leqslant -t)$, $A_3 = (X < t)$. Then by Lemma 3.7.3 and its notations, we see that $x_1 = \mathbb{E}^F[X|X > -t] \geqslant -t > 0$, $x_2 = \mathbb{E}^F[X|t \leqslant X \leqslant -t] = 0$, $x_3 = \mathbb{E}^F[X|X < t] \leqslant t < 0$, and there exists a three-point symmetric distribution F_{ε_σ} such that $F_{\varepsilon_\sigma} \in \mathcal{L}_S(0,\sigma)$. With this F_{ε_σ} and noticing that $-t > 0$, we have

$$\begin{aligned}\mathbb{E}^{F_{\varepsilon_\sigma}}[(X_{\varepsilon_\sigma} - t)_+] &= (x_1 + \varepsilon_\sigma - t)_+ p_1 + (x_2 - t)_+ p_2 + (x_3 - \varepsilon_\sigma - t)_+ p_3 \\ &= (x_1 + \varepsilon_\sigma - t)\, p_1 + (x_2 - t)\, p_2 \\ &\geqslant (x_1 - t)\, p_1 + (x_2 - t)\, p_2 = \mathbb{E}^F[(X - t)_+]. \end{aligned} \tag{3.61}$$

  - **Case 2.2:** Assume $\mathbb{P}(t \leqslant X \leqslant -t) = 1$. In this case, $\mathbb{E}^F[(X-t)_+] = \mathbb{E}^F[X-t] = -t$. Let $A_1 = (0 < X \leqslant -t)$, $A_2 = (X = 0)$, $A_3 = (t \leqslant X < 0)$. Then by Lemma 3.7.3 and its notations, we see that $0 \leqslant x_1 = \mathbb{E}^F[X|0 < X \leqslant -t] \leqslant -t$, $x_2 = \mathbb{E}^F[X|X = 0] = 0$, $x_3 = \mathbb{E}^F[X| - t \leqslant X < 0] \leqslant 0$, and there exists a three-point symmetric distribution F_{ε_σ} such that $F_{\varepsilon_\sigma} \in \mathcal{L}_S(0,\sigma)$. With this distribution F_{ε_σ} and by Jensen's inequality, we have

$$\mathbb{E}^{F_{\varepsilon_\sigma}}[(X_{\varepsilon_\sigma} - t)_+] \geqslant [(\mathbb{E}[X_{\varepsilon_\sigma}] - t)_+] = (-t)_+ = -t = \mathbb{E}^F[(X - t)_+]. \tag{3.62}$$

By combing all the four cases (3.59)-(3.62), we see that for any $F \in \mathcal{L}_S(0,\sigma)$, there exists a three symmetric distribution F_{ε_σ} such that $\mathbb{E}[(X_{\varepsilon_\sigma} - t)_+] \geqslant \mathbb{E}[(X - t)_+]$, which implies that

$$\sup_{F \in \mathcal{L}_{3,S}(0,\sigma)} \mathbb{E}^F[(X - t)_+] \geqslant \sup_{F \in \mathcal{L}_S(0,\sigma)} \mathbb{E}^F[(X - t)_+]. \tag{3.63}$$

Since $\mathcal{L}_{3,S}(0,\sigma) \subset \mathcal{L}_S(0,\sigma)$, we have $\sup_{F \in \mathcal{L}_{3,S}(0,\sigma)} \mathbb{E}^F[(X-t)_+] \leqslant \sup_{F \in \mathcal{L}(0,\sigma)} \mathbb{E}^F[(X-t)_+]$, which, together with (3.63), implies that (3.58) holds.

□

Lemma 3.7.5. *For any $t \in \mathbb{R}_+$,*

$$\sup_{F\in\mathcal{L}_+(\mu,\sigma)} \mathbb{E}^F[(X-t)_+] = \sup_{F\in\mathcal{L}_2^+(\mu,\sigma)} \mathbb{E}^F[(X-t)_+]. \tag{3.64}$$

Proof. It suffices to show that for any $F \in \mathcal{L}^+(\mu,\sigma)$, there exists $F^* \in \mathcal{L}_2^+(\mu,\sigma)$ such that $\mathbb{E}^F[(X-t)_+] \leqslant \mathbb{E}^{F^*}[(X-t)_+]$. Note that $\mathbb{E}^F[(X-t)_+] \geqslant (\mu-t)_+$ for any $F \in \mathcal{L}^+(\mu,\sigma)$. We only need to consider the distribution $F \in \mathcal{L}_+(\mu,\sigma)$ with $\mathbb{E}^F[(X-t)_+] > (\mu-t)_+ \geqslant \max\{\mu-t, 0\}$, that means, $\mathbb{P}^F(X>t) \in (0,1)$.

Let $X \sim F$. Note that for any $F \in \mathcal{L}^+(\mu,\sigma)$, it holds that $\mu > 0$, and let $X \sim F$. Assume that $q = \mathbb{P}(X > t) \in (0,1)$ with $p = \mathbb{P}(X \leqslant t) = 1-q$. Define F^* as a two-point distribution of the random variable $\overline{X}$ that is defined as $\overline{X} = x_1 1_{\{X>t\}} + x_2 1_{\{X\leqslant t\}}$, where $x_1 = x_1^F = \mathbb{E}^F[X|X>t] > t$, $0 \leqslant x_2 = x_2^F = \mathbb{E}^F[X|X\leqslant t] \leqslant t$ as $F \in \mathcal{L}^+(\mu,\sigma)$, and $x_1 > x_2$.

Clearly, $\mathbb{E}^F[\overline{X}] = x_1\, q + x_2\, p = \mathbb{E}^F[X] = \mu$. In addition, by (3.51), we have

$$\begin{aligned}\mathrm{Var}(\overline{X}) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = x_1^2\, q + x_2^2\, p - \mu^2 &\leqslant \mathbb{E}^F[X^2|X>t]\, q + \mathbb{E}[X^2|X\leqslant t]\, p - \mu^2 \\ &= \mathbb{E}[X^2] - \mu^2 = \mathrm{Var}(X) = \sigma^2.\end{aligned}$$

Define the random variable $\overline{X}_\varepsilon$ as

$$\overline{X}_\varepsilon = (x_1 + p\varepsilon)1_{\{X>t\}} + (x_2 - q\varepsilon)1_{\{X\leqslant t\}}, \quad \varepsilon \in [0, x_2/p].$$

We have $\mathbb{E}[\overline{X}_\varepsilon] = \mu$, and

$$\mathrm{Var}(\overline{X}_\varepsilon) = (x_1+p\varepsilon)^2 q + (x_2 - q\varepsilon)^2 p - \mu^2 = \sigma_0^2 + 2pq(x_1-x_2)\varepsilon + p^2 q\varepsilon^2 + qp^2\varepsilon^2,$$

which is continuous and increasing in $\varepsilon \in [0, \bar{\varepsilon}]$ with $\bar{\varepsilon} = x_2/q$ and $\sigma_0^2 = x_1^2 q + x_2^2 p - \mu^2$. We next find the desired F^* by considering the following two cases.

(1) If $\mathrm{Var}(\overline{X}_{\bar{\varepsilon}}) \geqslant \sigma^2$, then there exists $\varepsilon_0 \in [0, \bar{\varepsilon}]$ such that $\mathrm{Var}(\overline{X}_{\varepsilon_0}) = \sigma^2$. In addition, noting that for any $\varepsilon \in [0, \bar{\varepsilon}]$, $x_1 + p\varepsilon \geqslant x_1 > t \geqslant x_2 \geqslant x_2 - q\varepsilon$, we have

$$\mathbb{E}[(X_\varepsilon - t)_+] = (x_1 + p\varepsilon - t)q \geqslant (x_1 - t)q = \mathbb{E}^F[(X-t)_+],$$

and thus, the distribution of $X_{\bar{\varepsilon}}$ is the desired distribution F^*.

(2) If $\mathrm{Var}(\overline{X}_{\overline{\varepsilon}}) < \sigma^2$, note that $\overline{X}_{\overline{\varepsilon}} = (x_1 + p\overline{\varepsilon})1_{\{X>t\}} + 0 \times 1_{\{X\leqslant t\}} =: \overline{x}1_{\{X>t\}} + 0 \times 1_{\{X\leqslant t\}}$, where $\overline{x} = x_1 + p\overline{\varepsilon}$. For $\delta \in [0, q)$, let $U \sim \mathrm{U}[0, 1]$, and define random variable $\overline{X}_{\overline{\varepsilon},\delta}$ as

$$\overline{X}_{\overline{\varepsilon},\delta} = \frac{q\,\overline{x}}{q-\delta} 1_{\{U\geqslant p+\delta\}}.$$

One can verify that $\mathbb{E}[\overline{X}_{\overline{\varepsilon},\delta}] = \mathbb{E}[\overline{X}_{\overline{\varepsilon}}] = \mu$, and

$$\mathrm{Var}(\overline{X}_{\overline{\varepsilon},\delta}) = \frac{(q\,\overline{x})^2}{q-\delta} - \mu^2,$$

which is continuous and increasing in $\delta \in [0, q)$ and satisfies $\mathrm{Var}(\overline{X}_{\overline{\varepsilon},0}) = \mathrm{Var}(\overline{X}_{\overline{\varepsilon}}) < \sigma^2$ and $\lim_{\delta\to q} \mathrm{Var}(\overline{X}_{\overline{\varepsilon},\delta}) = \infty$. There thus exists $\delta_0 \in (0, q)$ such that $\mathrm{Var}(\overline{X}_{\overline{\varepsilon},\delta_0}) = \sigma^2$, that is, the distribution of $\overline{X}_{\overline{\varepsilon},\delta_0}$ belongs to $\mathcal{L}_2^+(\mu, \sigma)$. In addition, note that

$$\begin{aligned}\mathbb{E}[(\overline{X}_{\overline{\varepsilon},\delta_0} - t)_+] &= \left(\frac{q\,\overline{x}}{q-\delta_0} - t\right)(q-\delta_0) \\ &= q\,\overline{x} - t(q-\delta_0) \geqslant q\,\overline{x} - tq \geqslant (x_1 - t)q = \mathbb{E}^F[(X-t)_+],\end{aligned}$$

where the two inequalities follow from $\delta_0 \geqslant 0$ and $\overline{x} \geqslant x_1$, respectively. Thus, the distribution of $X_{\overline{\varepsilon},\delta_0}$ is the desired distribution F^*.

Combining cases (1) and (2), we complete the proof. □

By combining Lemmas 3.7.2, 3.7.4, 3.7.5, we conclude that the proof of Jagannathan (1977) for Proposition 3.2.1 is correct and that the supremums of $\mathbb{E}^F[(X-t)_+]$ over $\mathcal{L}(\mu, \sigma)$, $\mathcal{L}_S(\mu, \sigma)$, and $\mathcal{L}^+(\mu, \sigma)$ can be reformulated as the supremums of $\mathbb{E}^F[(X-t)_+]$ over $\mathcal{L}_2(\mu, \sigma)$, $\mathcal{L}_{3,S}(\mu, \sigma)$, and $\mathcal{L}_2^+(\mu, \sigma)$, respectively.

Chapter 4

An axiomatic theory for anonymized risk sharing

4.1 Introduction

Risk sharing, as one of the classical risk management mechanisms, refers to pooling risks from several participants in a group and reallocating the total risk in a specific way. A risk sharing scheme arises in many different business practices, such as insurance, tontines, profit sharing contracts in investment, to name a few. In these contexts, either profits or losses, or both of them, may be shared among participants. The traditional approach to study risk sharing is via equilibrium, either Pareto or competitive equilibria, dating back to the classic work of Arrow and Debreu (1954) and Borch (1962); see, e.g., Starr (2011) for a general treatment. In either form of equilibria, information on the preferences of the agents is required to define and compute an equilibrium. Commonly used preferences include expected utility, rank-dependent utility, cumulative prospect theory, risk measures, and many more advanced models; see Wakker (2010) for decision models and Föllmer and Schied (2016) for risk measures. However, in many practical situations, one rarely has precise information on the preferences, since elicitation of preferences can be challenging and costly (e.g., Leonard (1983)), and preferences may be incomplete, ambiguous, or falsely

supplied (e.g., Delage and Li (2018)). This is especially relevant in digital economy, where anonymized digital platforms are used to implement risk sharing, such as P2P insurance, revenue sharing of digital music and movies, and blockchain mining pools.

In this chapter, we consider a framework of *anonymized risk sharing*, where no information on preferences is required or used. The key feature of this framework is that agents do not need to disclose their preferences, identity, or wealth level.[1] More precisely, the allocation to an agent is determined by the initial risk contributions of all agents, but not the specification of these agents. For this reason, anonymized risk sharing schemes are desirable in several application such as P2P insurance (e.g., Denuit (2019) and Abdikerimova and Feng (2022)), Bitcoin mining pools (e.g., Eyal and Sirer (2018) and Leshno and Strack (2020)), and tontines (e.g., Chen et al. (2019) and Hieber and Lucas (2022)), and revenue sharing of digital content (e.g., Meyn et al. (2023)). Anonymized risk sharing is also closely related to the concept of decentralized risk sharing, where some key features are that no capital reserve is needed and agents settle their allocations via a pre-specified contract and a payment network; see Feng (2023). Examples of decentralized insurance include decentralized P2P insurance, mutual aid, and catastrophic risk pooling; see Feng et al. (2022) for a summary of models for decentralized insurance.

Since no information on individual preferences is known in the setting of anonymized risk sharing, it is difficult to apply equilibrium approaches. To better understand a suitable anonymized risk sharing rule, we put forward four natural axioms, namely, actuarial fairness, risk fairness, risk anonymity, and operational anonymity. The interpretation and desirability of these axioms will be discussed in detail in Section 4.3. Quite remarkably, we show in Section 4.4 that these four axioms *uniquely* identify one risk sharing rule (Theorem 4.4.1), the *conditional mean risk sharing (CMRS)*. As far as we know, this chapter provides the first axiomatic result for any risk sharing rules.

As an important risk sharing rule in economic theory with many attractive properties, CMRS was used by Landsberger and Meilijson (1994) to study Pareto optimality

[1] We chose the term "anonymized risk sharing" over "anonymous risk sharing", as the former emphasizes that individual information is deliberately masked (but it could be available), and the latter stresses that such information is not known or supplied.

of comonotonic risk allocations, and its properties were studied in detail by Denuit and Dhaene (2012); see Denuit et al. (2022a) for a summary of these properties. Our characterization hence provides a first axiomatic foundation for CMRS and its applications in economic theory and decentralized finance and insurance.

On the technical side, the proof of Theorem 4.4.1 relies on a new characterization of the conditional expectation which we present in Theorem 4.4.2. We further show that the four axioms are independent (Proposition 4.4.1), and the characterization holds also on half spaces (Theorem 4.4.3), common in the context of loss sharing or profit sharing. Several other properties related to our axioms are studied in Section 4.5, including backtracking, universal improvement, comonotonicity, and symmetry. In particular, we show that CMRS is the unique risk sharing rule satisfying universal improvement, risk anonymity, and operational anonymity, complementing the characterization in Theorem 4.4.1. Section 4.7 further generalizes the above framework to include modeling of target information. The formulation of target information is motivated by the consideration that in some risk sharing contexts, participants may prefer settling the risk transfer via some other information in addition to the total incurred loss or profit. This formulation allows for a wider spectrum of practical applications. Applications discussed in Section 4.8 include blockchain mining pools, and revenue sharing of digital music and videos, highlighting the suitability of our axiomatic and the implication on risk sharing mechanisms. Proofs of all results are relegated to the Appendix 4.10.

Research on axiomatic approaches for decision models and risk measures has a long history. For a specimen, see the monographs by Gilboa (2009), Wakker (2010) and the extensive lists of references therein. Axiomatic studies on risk functionals have been prolific in decision theory (e.g., Yaari (1987), Schmeidler (1989), Maccheroni et al. (2006) and Gilboa et al. (2010)) and risk measures (e.g., Artzner et al. (1999), Föllmer and Schied (2002) and Wang and Zitikis (2021)). Gilboa et al. (2019) had a recent discussion on the usefulness of axiomatic approaches in modern economic theory. Despite the huge success of axiomatic theories for risk functionals, an axiomatic study for risk sharing rules is missing from the literature; our work fills in this gap. Our new framework imposes substantial technical challenges compared to the above literature, as the risk sharing rules

are multi-dimensional and random-vector-valued, as opposed to preference functionals or risk measures, which are typically real- or vector-valued.

4.2 Risk sharing rules: Definition and examples

We describe in this section the main object of the chapter, the risk sharing rules. For this, we first need to fix some notation. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $\mathcal{X}$ be a convex cone of random variables on this space, representing the set of possible random losses of interest. Assume $0 \in \mathcal{X}$. Positive values of random variables represent losses and negative values represent gains; this convention makes no difference in all mathematical results. We always treat almost surely (a.s.) equal random variables as identical, and we use $\sup X$ for the essential supremum of X, that is, $\sup X = \inf\{x \in \mathbb{R} : \mathbb{P}(X > x) = 0\}$.

In our framework, n economic agents share a total risk, where $n \geqslant 3$ is an integer. We assume $n \geqslant 3$ since the case $n = 2$ is technically different and practically limiting; see Example 4.5.1 in Section 4.5.1. For notation simplicity we write

$$[n] = \{1, \dots, n\}.$$

Each agent $i \in [n]$ faces an initial risk X_i, which is the risk contribution of agent i to the risk sharing pool. We use the term "risk" to reflect that the random variable X_i may be positive or negative, and sometimes we use the term "loss" to emphasize its positive side. For any random variable S, the set of all *allocations* of S is denoted by

$$\mathbb{A}_n(S) = \left\{ (Y_1, \dots, Y_n) \in \mathcal{X}^n : \sum_{i=1}^{n} Y_i = S \right\}.$$

Throughout, we write $\mathbf{X} = (X_1, \dots, X_n)$ for the initial risk (contribution) vector, and $S^{\mathbf{X}} = \sum_{i=1}^{n} X_i$ for the total risk.

A *risk sharing rule* is a mapping $\mathbf{A} : \mathcal{X}^n \to \mathcal{X}^n$ satisfying $\mathbf{A}^{\mathbf{X}} = (A_1^{\mathbf{X}}, \dots, A_n^{\mathbf{X}}) \in \mathbb{A}_n(S^{\mathbf{X}})$ for each $\mathbf{X} \in \mathcal{X}^n$. The requirement $\mathbf{A}^{\mathbf{X}} \in \mathbb{A}_n(S^{\mathbf{X}})$ means that $\mathbf{A}^{\mathbf{X}}$ sums up to the total risk. In other words, there is no external fund coming in or out of the risk sharing pool

except for the initial contributions of the agents, a most natural requirement for defining an allocation rule. Each component of $\mathbf{A}^{\mathbf{X}}$ represents the (random) allocation of risk to an agent. Through the rule $\mathbf{A}$, the initial risk vector $\mathbf{X}$ enters the sharing pool as an input, and the allocation vector $\mathbf{A}^{\mathbf{X}}$ comes out as the output. Given each scenario $\omega \in \Omega$, the actual payment is settled as the vector $\mathbf{A}^{\mathbf{X}}(\omega) \in \mathbb{R}^n$. A positive payment $A_i^{\mathbf{X}}(\omega) = x > 0$ means that agent i needs to pay the amount of x, because positive values represent losses. This simple procedure is illustrated in Figure 4.1.

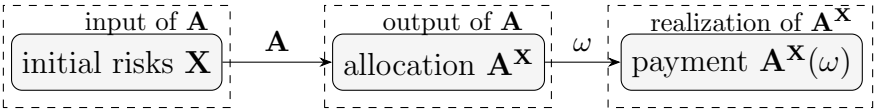


Figure 4.1: Risk sharing procedure

As a key feature of this framework, different from the large body of risk sharing problems studied in the literature, a risk sharing rule $\mathbf{A}$ does not require any information on the preferences of the agents, a risk exchange market, or subjective decisions of a central planner. The risk allocation will be determined completely through the mechanism design and the input risk vector.

We next provide several simple examples of risk sharing rules; see Denuit et al. (2022a) for a collection of risk sharing rules and their properties. Throughout, for $q \in [0, \infty)$, denote by $L^q = L^q(\Omega, \mathcal{F}, \mathbb{P})$ the set of all random variables with a finite q-th moment, and L^q_+ be the set of non-negative elements of L^q. We use the shorthand L^q, and we will write the full $L^q(\Omega, \mathcal{G}, Q)$ when we encounter another probability space $(\Omega, \mathcal{G}, Q)$. Some of the risk sharing rules below require $\mathcal{X}$ to be a subset of some specific spaces. We always use the convention $0/0 = 0$ which may appear in degenerate cases of (vi) and (vii).

(i) The identity risk sharing rule

$$\mathbf{A}^{\mathbf{X}}_{\mathrm{id}} = \mathbf{X} \quad \text{for } \mathbf{X} \in \mathcal{X}^n.$$

(ii) The all-in-one risk sharing rule

$$\mathbf{A}^{\mathbf{X}}_{\mathrm{all}} = \big(S^{\mathbf{X}}, 0, \dots, 0\big) \quad \text{for } \mathbf{X} \in \mathcal{X}^n.$$

(iii) The mean-adjusted all-in-one risk sharing rule

$$\mathbf{A}^{\mathbf{X}}_{\mathrm{ma}} = \big(S^{\mathbf{X}} - \mathbb{E}[S^{\mathbf{X}}], 0, \dots, 0\big) + \mathbb{E}[\mathbf{X}] \quad \text{for } \mathbf{X} \in \mathcal{X}^n \subseteq (L^1)^n.$$

(iv) The uniform risk sharing rule

$$\mathbf{A}^{\mathbf{X}}_{\mathrm{unif}} = S^{\mathbf{X}}\left(\frac{1}{n}, \dots, \frac{1}{n}\right) \quad \text{for } \mathbf{X} \in \mathcal{X}^n.$$

(v) The conditional mean risk sharing rule (CMRS)

$$\mathbf{A}^{\mathbf{X}}_{\mathrm{cm}} = \mathbb{E}\left[\mathbf{X}|S^{\mathbf{X}}\right] \quad \text{for } \mathbf{X} \in \mathcal{X}^n \subseteq (L^1)^n.$$

(vi) The mean proportional risk sharing rule

$$\mathbf{A}^{\mathbf{X}}_{\mathrm{prop}} = \frac{S^{\mathbf{X}}}{\mathbb{E}[S^{\mathbf{X}}]}\mathbb{E}[\mathbf{X}] \quad \text{for } \mathbf{X} \in \mathcal{X}^n \subseteq (L^1_+)^n.$$

(vii) The covariance risk sharing rule

$$\mathbf{A}^{\mathbf{X}}_{\mathrm{cov}} = \frac{S^{\mathbf{X}} - \mathbb{E}[S^{\mathbf{X}}]}{\mathrm{Var}(S^{\mathbf{X}})}\mathrm{Cov}(\mathbf{X}, S^{\mathbf{X}}) + \mathbb{E}[\mathbf{X}] \quad \text{for } \mathbf{X} \in \mathcal{X}^n \subseteq (L^2)^n.$$

These examples will be revisited repeatedly in the chapter. Among them, CMRS in (v) is the most important for our theory of anonymized risk sharing.

4.3 Four axioms for anonymized risk sharing

We next discuss desirable criteria for risk sharing rules by addressing the considerations of both *fairness* and *anonymity*. Fairness refers to the feature that an agent does not receive an absurd or unjustified allocation. Anonymity refers to the feature that agents do not need to disclose information on their identity, wealth, preferences, rank or ordering, and final realized losses. Given a risk sharing rule $\mathbf{A}$, the only information required to determine the risk allocation is the initial risk vector $\mathbf{X}$. Anonymity also guarantees that each agent will

not be treated differently and this reduces discrimination. As such, anonymity is closely related to fairness, although the two concepts have different motivations. To reflect these key features, we propose four natural axioms on a candidate risk sharing rule $\mathbf{A}$. Two of these axioms may be categorized as fairness axioms, and two may be categorized as anonymity axioms.

Axiom AF (Actuarial fairness). The expected value of each agent's allocation coincides with the expected value of the initial risk. That is, $\mathbb{E}[\mathbf{A}^{\mathbf{X}}] = \mathbb{E}[\mathbf{X}]$ for $\mathbf{X} \in \mathcal{X}^n$.

Axiom AF is one of the most ancient and formidable idea in risk management, which dates back to at least the 16th century; see Heras et al. (2020) for a history. AF serves as the basis for premium pricing in insurance, and this served as one of the earliest sources for studying probability and statistics.[2] Certainly, not all risk exchanges in practice are actuarially fair. In our framework, because of no information on the preferences or identities of the agents, it should not happen that one agent would receive an allocation with a higher expected value than her contribution, and some others receive allocations with lower expected values. Recall that the sum of these expected values is equal to the sum of the total risk, and hence agents on average receive the same expected value before and after risk sharing. Based on the above reasons, AF is a most natural requirement for anonymized risk sharing, and here we observe a joint effect of fairness and anonymity. The recent book Friedman (2020) has a comprehensive non-technical treatment on the historical importance of actuarial fairness and probability theory in insurance and social welfare.

Axiom AF can be alternatively formulated via incentives to join a risk sharing pool. AF means that the risk sharing rule does not exclude any risk-neutral agents, who would not join the risk sharing pool if the expected value of their risk increases after risk exchange. Since preferences are not revealed, a risk sharing rule should not exclude by design risk-neutral agents; we recall that a fundamental model of insurance (Arrow (1963)) involves a risk-neutral insurer to help share losses from risk-averse insureds.

Axiom RF (Risk fairness). The allocation to each agent should not exceed their maximum possible loss. That is, for $\mathbf{X} \in \mathcal{X}^n$ and $i \in [n]$, it holds that $A_i^{\mathbf{X}} \leqslant \sup X_i$.

[2] As we know, another important early source, roughly around the same time, is gambling, which motivated some work of Blaise Pascal, Pierre de Fermat, and Jacob Bernoulli.

Axiom RF reflects the idea that agents join the pool to share their risk, and they should not have to suffer more than their worst-case loss.[3] For each realization of the actual losses, the allocation satisfies the no rip-off condition in the insurance pricing literature (Deprez and Gerber (1985)), which says that an insured will never pay more premium than their maximum possible loss. If $A_i^{\mathbf{X}} \leqslant \sup X_i$ does not hold, then an agent with no risk of default may introduce positive probability of default after risk sharing, a clearly undesirable situation. For instance, using a power or logarithmic utility function, an agent's potential loss should never exceed her total wealth level (this may be called *bankruptcy aversion*), and Axiom RF says that there is no bankruptcy after the risk exchange if the initial risk is safe in this regard. Hence, formulated via incentives, this axiom means that not all bankruptcy-averse agents are excluded, which is arguably a weak requirement.

Two special implications of RF may be useful.[4] First, if the agent brings a pure surplus to the pool, i.e., $X_i \leqslant 0$, its allocation should also be a pure surplus; this is certainly true if $\mathcal{X}$ is contained in a half space such as the space of negative random variables. Second, in conjunction with Axiom AF, RF yields

$$\text{for } \mathbf{X} \in \mathcal{X}^n \text{ and } i \in [n], \text{ if } X_i = x \text{ is a constant, then } A_i^{\mathbf{X}} = x; \tag{4.1}$$

this follows from $A_i^{\mathbf{X}} \leqslant x$ and $\mathbb{E}[A_i^{\mathbf{X}}] = x$. That is, if the initial risk of an agent is a constant, then there is no risk exchange for this agent; this is quite intuitive since any risk-averse agent (in the sense of Rothschild and Stiglitz (1970)) would not trade a constant risk with a non-constant risk with the same mean. As a particular example, for a risk vector $(X, 0, \dots, 0)$, i.e., only the first agent has a non-zero initial risk, this property implies

$$A_1^{(X,0,\dots,0)} = X \text{ and } A_j^{(X,0,\dots,0)} = 0 \text{ for } j \neq 1, \tag{4.2}$$

which is arguably the only reasonable allocation in this particular case. On a point related to (4.1) and (4.2), our framework does not include the mechanism of side-payments, as

[3] We can alternatively formulate Axiom RF using $A_i^{\mathbf{X}} \geqslant \inf X_i$, where inf is the essential infimum. This has the same interpretation if we interpret positive values of random variables as gains instead of losses. With this formulation, mathematical results in the chapter remain the same due to symmetry.

[4] We can also check that Axiom RF is implied by these two simpler properties in conjunction with Axiom OA.

in e.g., selling insurance, because deciding side-payments usually requires the knowledge of specific identities or preferences (e.g., which agent is institutional, more risk averse, or with more bargaining power).

Axiom RA (Risk anonymity). The realized value of the allocation to each agent is determined by that of the total risk. That is, for $\mathbf{X} \in \mathcal{X}^n$, $\mathbf{A}^{\mathbf{X}}$ is $\sigma(S^{\mathbf{X}})$-measurable.

Axiom RA is central to the idea of designing a risk sharing mechanism. It means that the total realized allocation is determined only by the total loss suffered by the risk sharing pool, and not by specific losses from the individual participants. This resembles the earliest idea in insurance and risk sharing: Individuals get together to share their total future losses (in early years, these losses are typically caused by unexpected deaths, diseases or injuries), regardless of which one of them is the realized cause of the future loss. In other words, once an agent enters the pool, her own realized loss no longer matters, and only the realized loss of the pool matters. This reflects anonymity, as each agent does not need to disclose what is the realized loss; all individual losses are masked and only the total loss is revealed. The knowledge of the initial risk vector is only used for the design of the risk sharing mechanism, but not for the settlement of actual losses (see Figure 4.1).

Technically, RA holds for $\mathbf{X} \in \mathcal{X}^n$ satisfying that $\mathbf{A}^{\mathbf{X}}$ is comonotonic. As studied by Borch (1962) and Landsberger and Meilijson (1994), comonotonicity is closely related to Pareto optimality for risk-averse agents; see Section 4.5 for details.

Axiom OA (Operational anonymity). The allocation to one agent is not affected if risks of two other agents merge. That is, $A_k^{\mathbf{Y}} = A_k^{\mathbf{X}}$ for $k \neq i, j$ for $\mathbf{X} \in \mathcal{X}^n$, $i, j \in [n]$ and $\mathbf{Y} = \mathbf{X} + X_j \mathbf{e}_i - X_j \mathbf{e}_j$, where $\mathbf{e}_k = (0, \dots, 0, 1, 0, \dots, 0)$ is the unit vector along the k-th axis (the k-th component is 1).

In the definition of Axiom OA, the risk vector $\mathbf{Y}$ can be written by $Y_i = X_i + X_j$, $Y_j = 0$ and $Y_k = X_k$ for $k \neq i, j$. Axiom OA means that merging the risks of two agents will not affect the allocation components of uninvolved agents. This also implies that a redistribution of risks between agent i and j does not affect agent k for $k \neq i, j$. In an anonymized risk sharing framework, two agents may be two different accounts of the same

family, same organization, or even the same person. Their internal (private) operations do not need to be disclosed and should not affect the allocation to other agents. OA is called the property of fair merging by Denuit et al. (2022a), which clearly has a connection to fairness, although our motivation is different from the latter chapter. In the context of Bitcoin mining, Leshno and Strack (2020) formulated two axioms, called robustness to Sybil attacks and robustness to merging, which together reflect the same consideration as OA. This property is further explained in the following simple example.

Example 4.3.1. Assume $\mathbf{X} = (X_1, X_2, X_3)$ and $\mathbf{Y} = (X_1, X_2 + X_3, 0)$. In this setting, we have $A_1^{\mathbf{Y}} = A_1^{\mathbf{X}}$ if Axiom OA holds. Further, we have $A_3^{\mathbf{Y}} = 0$ from (4.1) implied by AF and RF, leading to $A_2^{\mathbf{Y}} = A_2^{\mathbf{X}} + A_3^{\mathbf{X}}$. Therefore, by merging risks from agents 2 and 3, agent 2 now takes up the total allocation to the two agents, and the allocation to agent 1 is unaffected by this operation.

Axiom OA can be alternatively formulated by another intuitive property that $A_i^{\mathbf{X}}$ is determined by $(X_i, S^{\mathbf{X}})$ for each i and $\mathbf{X}$. This latter property implies OA by definition. To see that OA implies this property, it suffices to observe, by repeatedly merging all agents except for agent 1, that

$$A_1^{\mathbf{X}} = A_1^{(X_1, S^{\mathbf{X}} - X_1, 0, \dots, 0)} \tag{4.3}$$

holds. We summarize the above observation in the following proposition, which is convenient to use for our later discussions.

Proposition 4.3.1. *A risk sharing rule* $\mathbf{A}$ *satisfies Axiom OA if and only if for all* $\mathbf{X} \in \mathcal{X}^n$ *and* $i \in [n]$, $A_i^{\mathbf{X}}$ *is determined by* $(X_i, S^{\mathbf{X}})$.

As discussed in this section, the four axioms are mathematically very simple and arguably natural in the framework of anonymized risk sharing. Quite remarkably, these four axioms uniquely pin down one risk sharing rule, which will be studied in the next section.

4.4 Two axiomatic characterizations of CMRS

4.4.1 An axiomatic characterization with four axioms

In this subsection, we show that Axioms AF, RF, RA and OA uniquely identify CMRS among all risk sharing rules. Recall that CMRS is defined as

$$\mathbf{A}^{\mathbf{X}} = \mathbb{E}\left[\mathbf{X}|S^{\mathbf{X}}\right] \quad \text{for } \mathbf{X} \in \mathcal{X}^n \subseteq (L^1)^n,$$

and equivalently, $A_i^{\mathbf{X}} = \mathbb{E}[X_i|S^{\mathbf{X}}]$ for $i \in [n]$. For the ease of presentation, we take $\mathcal{X} = L^1$ or L^1_+ in all our results; these results hold true for $\mathcal{X} = L^q$ and $\mathcal{X} = L^q_+$ with $q \in [1, \infty]$ following the same proof; see Remark 4.10.1.

We first briefly check that CMRS satisfies the four axioms by using properties of the conditional expectation $\mathbb{E}[X|S]$ for any $(X, S) \in \mathcal{X}^2$ which will be chosen as $(X_i, S^{\mathbf{X}})$ for $i \in [n]$. First, AF holds by the tower property $\mathbb{E}[\mathbb{E}[X|S]] = \mathbb{E}[X]$. Second, RF holds since $\mathbb{E}[X|S] \leqslant \sup X$. Third, RA holds by definition since $\mathbb{E}[X|S]$ is a function of S. Fourth, OA holds since $\mathbb{E}[X|S]$ is determined solely by (X, S). See also Denuit et al. (2022a) for these and other properties of CMRS.

Theorem 4.4.1. *Assume $\mathcal{X} = L^1$, i.e., the set of all integrable random variables. A risk sharing rule satisfies Axioms AF, RF, RA and OA if and only if it is CMRS.*

Theorem 4.4.1 is the main result of the chapter, showing that the four fairness and anonymity axioms allow for only one risk sharing rule. As far as we know, Theorem 4.4.1 is the first axiomatic characterization of risk sharing rules in the literature.

The "if" statement in Theorem 4.4.1, that CMRS satisfies the four axioms, has been checked above. The "only if" statement, which is the most important part of Theorem 4.4.1, requires a much more involved proof based some advanced results from functional analysis. Below we provide an intuitive sketch of the proof in the case that $(\Omega, \mathcal{F}, \mathbb{P})$ is discrete. A full proof is presented in Appendix 4.10.1.

For a discrete Ω, the main idea is to analyze each possible realized value $s \in \mathbb{R}$ of $S^{\mathbf{X}}$ one by one. There are at most countably many such s. Let $\mathbf{A}$ be a risk sharing rule satisfying

the four axioms. We focus on the allocation to agent 1, and aim to show $A_1^{\mathbf{X}} = \mathbb{E}[X_1|S^{\mathbf{X}}]$ for all $\mathbf{X} \in \mathcal{X}^n$; the allocations to the other agents are similar. Fix $S \in \mathcal{X}$, and we will first consider the risk vector $(X, S-X, 0, \dots, 0)$ by allowing X to vary within $\mathcal{X}$. Let V be the set of possible values taken by S. By RA, the value of the allocation $A_1^{(X,S-X,0,\dots,0)}$ to agent 1 is a determined by the realized value $s \in V$ of S and X. Denote this value by $h^{S,s}(X)$, that is, for fixed $S \in \mathcal{X}$ and $s \in V$,

$$h^{S,s}(X) = A_1^{(X,S-X,0,\dots,0)} \text{ given } S = s. \tag{4.4}$$

We can carefully check that $h^{S,s} : \mathcal{X} \to \mathbb{R}$ satisfies the following properties (in the last property we allow s to vary in V):

(a) normalization: $h^{S,s}(t) = t$ for all $t \in \mathbb{R}$; (by (4.1))

(b) additivity: $h^{S,s}(X+Y) = h^{S,s}(X) + h^{S,s}(Y)$ for $X, Y \in \mathcal{X}$; (by OA and (a))

(c) monotonicity: $h^{S,s}(Y) \geqslant h^{S,s}(X)$ if $Y \geqslant X$; (by RF and (b))

(d) $h^{S,s}(S) = s$; (by RF and (4.2))

(e) $\sum_{t\in V} h^{S,t}(X)\mathbb{P}(S=t) = \mathbb{E}[X]$ for $X \in \mathcal{X}$. (by AF)

The properties (a), (b) and (c) together guarantee that $h^{S,s}$ is linear, monotone, and normalized. Using a standard representation theorem (such as that of Riesz), there exists a probability measure $P_{S,s}$ such that

$$h^{S,s}(X) = \int X \mathrm{d}P_{S,s} \quad \text{for all } X \in \mathcal{X}. \tag{4.5}$$

The next task is to show that $P_{S,s}$ is precisely the conditional probability $\mathbb{P}(\cdot|S=s)$. Let $x \wedge y$ represent the minimum of $x, y \in \mathbb{R}$. Using (d) and taking $X = S \wedge x$ in (4.5), we arrive at

$$h^{S,s}(S \wedge x) \leqslant \int S \mathrm{d}P_{S,s} \wedge \int x \mathrm{d}P_{S,s} = s \wedge x.$$

This inequality further implies

$$\mathbb{E}[S \wedge x] = \sum_{t\in V} (t \wedge x)\mathbb{P}(S=t) \geqslant \sum_{t\in V} h^{S,t}(S \wedge x)\mathbb{P}(S=t) = \mathbb{E}[S \wedge x],$$

where the last equality is due to (e). Hence, we get $h^{S,s}(S \wedge x) = s \wedge x$ for each $s \in V$, and this gives, in particular, $\int (S \wedge s) \mathrm{d}P_{S,s} = s$. Therefore, $P_{S,s}(S \geqslant s) = 1$. Using symmetric arguments, we can show $P_{S,s}(S \leqslant s) = 1$. As a result, $P_{S,s}(S = s) = 1$. Using this equality and (e), for any $B \subseteq \{S = s\}$, we have

$$\mathbb{P}(B|S = s) = \frac{\mathbb{E}[\mathbb{1}_B]}{\mathbb{P}(S = s)} = \frac{\sum_{t \in V} h^{S,t}(\mathbb{1}_B)\mathbb{P}(S = t)}{\mathbb{P}(S = s)} = \frac{\sum_{t \in V} P_{S,t}(B)\mathbb{P}(S = t)}{\mathbb{P}(S = s)} = P_{S,s}(B).$$

Therefore, $P_{S,s}(\cdot) = \mathbb{P}(\cdot|S = s)$ and $h^{S,s}(X) = \mathbb{E}[X|S = s]$ for $X \in \mathcal{X}$. Based on this result, we can finally get $A_1^{\mathbf{X}} = \mathbb{E}[X_1|S^{\mathbf{X}}]$ for a general $\mathbf{X}$ using (4.3) guaranteed by OA. This concludes the proof of Theorem 4.4.1 in case Ω is discrete.

It is clear that the above proof sketch heavily relies on the assumption that $\mathbb{P}(S = s) > 0$ for $s \in V$, and it cannot be directly generalized to non-discrete spaces. For a proof of Theorem 4.4.1 on general probability spaces, we need a more refined representation result in functional analysis. We obtain such a result in Theorem 4.4.2 below, which may be of independent interest. In what follows, $\sigma(S)$ is the σ-field generated by S.

Theorem 4.4.2. *For a random variable S on $(\Omega, \mathcal{F}, \mathbb{P})$ and $\mathcal{G} = \sigma(S)$, consider the mapping $\phi : L^1(\Omega, \mathcal{F}, \mathbb{P}) \to L^1(\Omega, \mathcal{G}, \mathbb{P})$. The mapping ϕ is the functional induced by the conditional expectation, i.e., $\phi(X) = \mathbb{E}[X|S]$ for $X \in L^1(\Omega, \mathcal{F}, \mathbb{P})$, if and only if ϕ satisfies the following properties: (a) $\phi(t) = t$ for all $t \in \mathbb{R}$; (b) $\phi(X + Y) = \phi(X) + \phi(Y)$ for all X, Y; (c) $\phi(Y) \geqslant \phi(X)$ if $Y \geqslant X$; (d) $\phi(S) = S$, and (e) $\mathbb{E}[\phi(X)] = \mathbb{E}[X]$ for all X.*

The $\mathcal{G}$-conditional expectation as a mapping from $L^1(\Omega, \mathcal{F}, \mathbb{P})$ to $L^1(\Omega, \mathcal{G}, \mathbb{P})$ admits a few different sets of characterizations; Pfanzagl (1967) has a collection of several early results. For a more recent account, see Eisner et al. (2015, Chapter 13) in the context of Markov operators. Theorem 4.4.2 extends the above literature by offering a new characterization of the conditional expectation. There is a substantial gap between characterizing risk sharing rules, which is absent from the literature, and characterizing the conditional expectation, which has a rich history. We mention three notable differences. First, our framework concerns risk sharing rules, which are mappings from $\mathcal{X}^n$ to $\mathcal{X}^n$ with the additional property of full allocation, whereas the literature on the conditional expectation concerns mappings from $\mathcal{X}$ to $\mathcal{X}$. Second, the property $\phi \circ \phi = \phi$, called idempotentness

or projection (see e.g., Douglas (1965) and Pfanzagl (1967)), is essential to characterize the conditional expectation, whereas in our framework this property is implied as a consequence of the four axioms on the risk sharing rule. Third, in our framework, the output σ-field $\mathcal{G}$ depends on the input random vector $\mathbf{X}$, whereas for results on the conditional expectation, the σ-field $\mathcal{G}$ is pre-specified.

In the next proposition, we verify that the four axioms are independent, and thus none of them can be removed from Theorem 4.4.1.

Proposition 4.4.1. *Axioms AF, RF, RA and OA are independent. That is, any combination of three of Axioms AF, RF, RA and OA does not imply the remaining fourth axiom.*

For each axiom, we will provide in Example 4.4.1 a risk sharing rule that only satisfies three of them but not the fourth one; some of these examples have been listed in Section 4.2. The technical details of these claims are in Appendix 4.10.1.

Example 4.4.1. (i) The Q-CMRS $\mathbf{A}^{\mathbf{X}}_{Q\text{-cm}} = \mathbb{E}^Q[\mathbf{X}|S^{\mathbf{X}}]$ with $\mathcal{X} \subseteq L^1(\Omega, \mathcal{F}, Q)$ for a probability measure $Q \neq \mathbb{P}$ satisfies RF, RA and OA, but not AF.

(ii) The mean-adjusted all-in-one risk sharing rule

$$\mathbf{A}^{\mathbf{X}}_{\text{ma}} = \big(S^{\mathbf{X}} - \mathbb{E}[S^{\mathbf{X}}], 0, \dots, 0\big) + \mathbb{E}[\mathbf{X}]$$

with $\mathcal{X} \subseteq L^1$ satisfies AF, RA and OA, but not RF. As another example, the covariance risk sharing rule in Section 4.2 also satisfies AF, RA and OA but not RF.

(iii) The identity risk sharing rule $\mathbf{A}^{\mathbf{X}}_{\text{id}} = \mathbf{X}$ satisfies AF, RF and OA, but not RA.

(iv) A combination of $\mathbf{A}_{\text{all}}$ and $\mathbf{A}_{\text{cm}}$, defined by $\mathbf{A}^{\mathbf{X}} = \mathbf{A}^{\mathbf{X}}_{\text{all}} = (S^{\mathbf{X}}, 0, \dots, 0)$ if $\mathbf{X}$ is standard Gaussian, and $\mathbf{A}^{\mathbf{X}} = \mathbf{A}^{\mathbf{X}}_{\text{cm}} = \mathbb{E}[\mathbf{X}|S^{\mathbf{X}}]$ otherwise, satisfies AF, RF and RA, but not OA.

In some applications, risk allocation and risk contributions are restricted to being all positive or all negative, depending on the context. For instance, if agents are sharing P2P

insurance losses, then it may be sensible to assume $\mathbf{X}$ and $A^{\mathbf{X}}$ both take non-negative vector values; if agents are sharing profits from an investment, then it is the opposite; recall that positive values represent losses and negative values represent gains. A similar characterization to Theorem 4.4.1 for the case of positive random variables is presented below; the case of negative random variables holds by symmetry and is omitted.

Theorem 4.4.3. *Assume $\mathcal{X} = L^1_+$, i.e., the set of all non-negative integrable random variables. A risk sharing rule satisfies Axioms AF, RF, RA and OA if and only if it is CMRS.*

Technically, Theorem 4.4.3 requires a more sophisticated analysis than Theorem 4.4.1, as by restricting $\mathcal{X}$ to a half space, the mapping $h^{S,s}$ in (4.4) is no longer well-defined on a closed set under addition. Hence, some extension arguments are needed, and a full proof is put in Appendix 4.10.1.

By working with the positive half space, Axiom RF in Theorem 4.4.3 can be replaced by the simpler Property CP stated in (4.1). This is because CP and OA imply RF in case $\mathcal{X} = L^1_+$.

Property CP (Constant preserving). A constant initial risk results in constant allocation. That is, for $\mathbf{X} \in \mathcal{X}^n$ and $i \in [n]$, $X_i = x \in \mathbb{R}$ implies $A_i^{\mathbf{X}} = x$.

The restriction of $\mathcal{X}$ to a half space allows for the example of the mean proportional risk sharing rule

$$\mathbf{A}^{\mathbf{X}}_{\mathrm{prop}} = \frac{S^{\mathbf{X}}}{\mathbb{E}[S^{\mathbf{X}}]}\mathbb{E}[\mathbf{X}], \quad \mathbf{X} \in \mathcal{X}^n \subseteq (L^1_+)^n.$$

Similar to the covariance risk sharing rule, $\mathbf{A}_{\mathrm{prop}}$ satisfies AF, RA and OA but not RF or CP, since a constant X_i does not lead to a constant $\mathbf{A}_i^{\mathbf{X}}$.

4.4.2 Another axiomatic characterization

In this section, we provide another way to characterize CMRS in a discrete space (Theorem 4.4.4). We first introduce a property called law specification which is satisfied by the CMRS rule.

Property LS (Law Specification). The risk sharing rule $A_i^{\mathbf{X}}$ is determined by the conditional distribution of X_i on $S = s$ for $i = 1, \dots, n$ and s in the range of S.

Theorem 4.4.4. *Assume $(\Omega, \mathcal{F}, \mathbb{P})$ is discrete and $\mathcal{X} = L^1(\Omega, \mathcal{F}, \mathbb{P})$. A risk sharing rule satisfies Property LS and CP, Axioms RF, RA and OA if and only if it is CMRS.*

Compared to the Theorem 4.4.1, we substitute Axiom AF for Property LS and CP. Obviously, Theorem 4.4.4 is weaker than Theorem 4.4.1 since Property LS is a quite strong property. In practice, it is unlikely to know the information of conditional distribution of the individual risk.

It is more reasonable to assume that a risk sharing rule $A_i^{\mathbf{X}}$ is determined by the distribution of (X_i, S), or determined by the joint distribution of $\mathbf{X}$. However, we show in Proposition 4.3.1 that a risk sharing rule satisfies Axiom OA if and only if it is determined by (X_i, S), and Axiom RF, RA and OA cannot uniquely characterize CMRS (see Example 4.4.1(i)). Also, when OA holds, it is equivalent to assume $A_i^{\mathbf{X}}$ is determined by the distribution of (X_i, S) and determined by the joint distribution of $\mathbf{X}$. Thus, it is unlikely to loose the Property LS in this situation. We put this theorem here as a mathematical result for characterizing CMRS.

4.5 Other properties and their connection to the four axioms

In this section, we discuss several further properties that CMRS satisfies or does not satisfy. These properties are known and straightforward to check. The purpose of this section is to clarify their relationship with the four axioms in Section 4.3.

4.5.1 Universal improvement in terms of convex order

For two random variables X and Y, we say X is improved compared to Y in *convex order* if $\mathbb{E}[u(X)] \leqslant \mathbb{E}[u(Y)]$ for any convex function $u : \mathbb{R} \to \mathbb{R}$; this is denoted by $X \leqslant_{\mathrm{cx}} Y$.

The most appealing feature of CMRS, as argued by Landsberger and Meilijson (1994) and Denuit and Dhaene (2012), is that it universally improve the risk for a larger class of decision makers, via the following property.

Property UI (Universal improvement). The allocation improves the initial risk in convex order. That is, for any $\mathbf{X} \in \mathcal{X}^n$ and $i \in [n]$, it holds that $A_i^{\mathbf{X}} \leqslant_{\mathrm{cx}} X_i$.

CMRS satisfies UI as a direct result of conditional Jensen's inequality. Intuitively, UI means that the initial risk for each agent has larger variability than the allocation to that agent. As a consequence, risk-averse agents in the classic sense of Rothschild and Stiglitz (1970), i.e., those who prefer both an improvement of convex order and a sure gain, will prefer their UI allocations over their initial risks. In a similar spirit, Denuit and Robert (2020, Proposition 4.2) showed that if risks in the pool are independent then the CMRS allocation improves in convex order for each existing agent when the pool is enlarged.

To illustrate the important role of UI for CMRS, we note that UI implies both AF and RF, since $X \leqslant_{\mathrm{cx}} Y$ implies $X \leqslant \sup X \leqslant \sup Y$ and $\mathbb{E}[X] = \mathbb{E}[Y]$ for any random variables $X, Y \in L^1$. We summarize this observation in the following proposition.

Proposition 4.5.1. *Property UI implies Axioms AF and RF and Property CP.*

Combining Proposition 4.5.1 with Theorems 4.4.1 and 4.4.3, we immediately obtain in Corollary 4.5.1 another characterization of CMRS with AF and RF replaced by UI. Proposition 4.5.1 and Corollary 4.5.1 also illustrate that UI is a very strong property. Recall that our characterization in Theorem 4.4.1 relies on the weaker axioms of AF and RF, and thus the more important "only if" statement is stronger than that of Corollary 4.5.1.

Theorem 4.5.1. *Assume $\mathcal{X} = L^1$ or L^1_+. A risk sharing rule satisfies Axioms RA and OA and Property UI if and only if it is CMRS.*

Finally, we provide a subtle example, showing that the condition $n \geqslant 3$ which we assumed from the beginning is indispensable, and this remains true even if we further assume the stronger property of UI. The intuition is that in case $n = 2$, Axiom OA is

empty since no merging operation is possible when one agent's risk is fixed. As a result, we cannot obtain the additivity of $h^{S,s}$ in (4.4) which requires some "wiggle room" provided by the third dimension.

Example 4.5.1. Let $n = 2$. We design a risk sharing rule $\mathbf{A}$ which satisfies all of AF, RF, RA, OA and UI, but it is not CMRS. Let $\mathbf{A} = \mathbf{A}_{\mathrm{cm}}$ for all $\mathbf{X} \in \mathcal{X}^n$ except for a specific $\mathbf{Y} = (Y_1, Y_2)$, which is given by

$$A_1^{\mathbf{Y}} = \mathbb{E}[Y_1|S^{\mathbf{Y}}] + h(S^{\mathbf{Y}}) \text{ and } A_2^{\mathbf{Y}} = \mathbb{E}[Y_2|S^{\mathbf{Y}}] - h(S^{\mathbf{Y}}),$$

where h satisfies $\mathbb{E}[h(S^{\mathbf{Y}})] = 0$. The intuition is that, if h is sufficiently small and $\mathbb{E}[Y_i|S^{\mathbf{Y}}]$ is sufficiently different from Y_i, then $A_i^{\mathbf{Y}} \leqslant_{\mathrm{cx}} Y_i$ still holds, thus satisfying RA, OA and UI. To make the example explicit, let us take $Y_1 \sim \mathrm{N}(0,1)$ and $Y_2 \sim \mathrm{N}(0,2)$, and Y_1, Y_2 are independent. Let $h(s) = s/6$. Note that $S^{\mathbf{Y}} \sim \mathrm{N}(0,3)$. We can compute

$$A_1^{\mathbf{Y}} = \frac{1}{3}S^{\mathbf{Y}} + \frac{1}{6}S^{\mathbf{Y}} = \frac{1}{2}S^{\mathbf{Y}} \sim \mathrm{N}(0, 0.75) \quad \text{and} \quad A_2^{\mathbf{Y}} = \frac{2}{3}S^{\mathbf{Y}} - \frac{1}{6}S^{\mathbf{Y}} = \frac{1}{2}S^{\mathbf{Y}} \sim \mathrm{N}(0, 0.75).$$

Hence, for this particular $\mathbf{Y}$, everything in Axiom RA and Property UI (hence Axioms AF and RF) is satisfied. Axiom OA holds trivially as its statement is empty. Therefore, $\mathbf{A}$ is not CMRS but it satisfies the four axioms and Property UI.

4.5.2 Backtracking

The second property we discuss is the backtracking property, which means that, for any $i \in [n]$ if $S^{\mathbf{X}}$ is able to determine X_i, then $A_i^{\mathbf{X}} = X_i$, and thus there is no risk exchange involving agent i. It is straightforward to verify that CMRS satisfies this property.

Property BT (Backtracking). For each $\mathbf{X} \in \mathcal{X}^n$ and $i \in [n]$, if X_i is $\sigma(S^{\mathbf{X}})$-measurable, then $A_i^{\mathbf{X}} = X_i$.

Property BT is sometimes argued as an undesirable property; see Denuit et al. (2022b). We give a simple example below for the purpose of discussion.

Example 4.5.2. Suppose that $X_1 = \sigma_1 Y_1$, $X_2 = \sigma_2 Y_2$, and $X_3 = \sigma_3 Y_3$, where Y_1, Y_2, Y_3 are iid taking values in $\{0, \dots, 9\}$, and $\sigma_1 = 1001$, $\sigma_2 = 1010$ and $\sigma_3 = 1100$. Note that $S^{\mathbf{X}}$ uniquely determines the value of (X_1, X_2, X_3) since the last three digits of $S^{\mathbf{X}}$ are precisely Y_3, Y_2, Y_1. In this example, X_1, X_2, X_3 have similar distributions, and they are independent. Intuitively, some risk sharing effect is possible for such $\mathbf{X}$, but $\mathbf{A}^{\mathbf{X}}_{\mathrm{cm}} = \mathbf{X}$ due to the backtracking property.

Property BT intuitively means that there is no risk sharing effect if $S^{\mathbf{X}}$ is too informative compared to the individual contributions. As a consequence, there are some situations, although perhaps rare, in which CMRS discourages some participants to enter the risk sharing pool, even if they bring in risks independent of the other participants. Theorem 4.4.1 provides the additional insight that BT is unavoidable, given the four natural axioms of fairness and anonymity. If some applications demand BT to be avoided, then one has to relax some axioms. For this, one naturally wonders which of the four axioms are responsible for Property BT.

The axiom which involves $\sigma(S^{\mathbf{X}})$ is RA, and a first guess may be that RA is connected to BT. Somewhat surprisingly, this is not true. In the next result we establish that AF, RF and OA are sufficient for BT if $\mathbf{A}^{\mathbf{X}}$ is further assumed $\sigma(\mathbf{X})$-measurable; the last assumption holds in virtually all applications of risk sharing rules as the risk settlement should not involve extra randomness outside $\sigma(\mathbf{X})$.

Proposition 4.5.2. *Assume $\mathcal{X} = L^1$ or L^1_+. If a risk sharing rule $\mathbf{A}$ satisfies Axioms AF, RF and OA, and $\mathbf{A}^{\mathbf{X}}$ is $\sigma(\mathbf{X})$-measurable for all $\mathbf{X} \in \mathcal{X}^n$, then it satisfies Property BT.*

An example of a risk sharing rule satisfying AF, RF and OA but not RA is the mixture $\mathbf{A} = \lambda \mathbf{A}_{\mathrm{id}} + (1 - \lambda)\mathbf{A}_{\mathrm{cm}}$ for some $\lambda \in (0, 1]$; such a rule satisfies BT. On the other hand, the mean-adjusted all-in-one and covariance risk sharing rules in Example 4.4.1 satisfy AF, RA and OA, and it does not satisfy BT or RF.

4.5.3 Comonotonicity

Next, we discuss comonotonicity, an important concept in risk sharing. A random vector $(X_1, \dots, X_n)$ is *comonotonic* if there exists increasing (in the non-strict sense) functions $g_1, \dots, g_n$ and a random variable Z such that $X_i = g_i(Z)$ a.s. for $i \in [n]$.

Property CM (Comonotonicity). For each $\mathbf{X} \in \mathcal{X}^n$, $\mathbf{A}^{\mathbf{X}}$ is comonotonic.

Property CM implies Axiom RA since each component of a comonotonic random vector can be written as an increasing function of the sum; see Denneberg (1994). Therefore, Property CM can be equivalently formulated as that for $\mathbf{X} \in \mathcal{X}^n$ and $i \in [n]$, there exists an increasing function $g_i^{\mathbf{X}} : \mathbb{R} \to \mathbb{R}$ such that $A_i^{\mathbf{X}} = g_i^{\mathbf{X}}(S^{\mathbf{X}})$.

If Property CM holds, then the allocation to each agent increases as the total realized risk increases. Under belief homogeneity and mild assumptions, Property CM is also a necessary condition for a risk sharing rule to be Pareto optimal for risk-averse agents or for an exchange market with linear prices; see e.g., Borch (1962), Landsberger and Meilijson (1994) and Boonen et al. (2021). As notable exceptions, quantile-based risk sharing and belief heterogeneity both result in non-comonotonic Pareto-optimal allocations; see Embrechts et al. (2018, 2020).

CMRS does not generally satisfy Property CM, which may be seen as a drawback of CMRS when the preferences of the agents are specified and risk averse, as the allocation is suboptimal. In our context of anonymized risk sharing, optimality cannot be discussed this way, since agents' specific preferences are not relevant. Nevertheless, if $s \mapsto \mathbb{E}[X_i | S = s]$ is increasing for each $i \in [n]$, then CMRS is comonotonic. There are many specific models of $\mathbf{X}$ for which comonotonicity of CMRS holds; see Denuit et al. (2022a) and the references therein. In several contexts, such as those with risk-averse agents or moral hazard, comonotonicity is desirable. On this point, our Theorem 4.4.1 implies the negative result that the four axioms and Property CM conflict each other. We further strengthen this result by showing that OA, CM, and a weak version of CP cannot be satisfied by the same risk sharing rule. This weak version of CP is the following property.

Property ZP (Zero preserving). For $\mathbf{X} \in \mathcal{X}^n$ and $i \in [n]$, if $X_i = 0$, then $A_i^{\mathbf{X}} = 0$.

It might be useful to recall some logical relationship among some properties and axioms mentioned above, that is,

$$\text{UI} \Longrightarrow \text{AF} + \text{RF} \Longrightarrow \text{CP} \Longrightarrow \text{ZP}; \qquad \text{CM} \Longrightarrow \text{RA}.$$

Proposition 4.5.3. *Assume $\mathcal{X} = L^1$. There is no risk sharing rule satisfying Axiom OA and Properties CM and ZP.*

If Property CM is needed in a specific application, one may need to relax some of the axioms. In the following example, we provide two relaxations to show that it is possible to have both CM and OA or both CM and UI.

Example 4.5.3. (i) The mean-adjusted all-in-one risk sharing rule satisfies CM (implying RA), AF, and OA, but not RF or ZP, as we see from Example 4.4.1.

(ii) For each $\mathbf{X} \in \mathcal{X}^n \subseteq (L^1)^n$, the comonotonic improvement of Landsberger and Meilijson (1994) gives a comonotonic vector $\mathbf{X}'$ such that each component of $\mathbf{X}'$ is dominated by the corresponding component of $\mathbf{X}$ and $S^{\mathbf{X}'} = S^{\mathbf{X}}$; see also Rüschendorf (2013). The risk sharing rule given by $\mathbf{A}^{\mathbf{X}} = \mathbf{X}'$ satisfies CM (implying RA) and UI (implying AF and RF), but not OA.

Remark 4.5.1. Property CM is closely related to Pareto optimality for risk-averse decision makers with specified preferences. Equilibrium and optimality will be discussed in detail in Section 4.6.

4.5.4 Symmetry

Symmetry is another important property reflecting the spirit both fairness and anonymity. Let Π_n be the set of n-permutations, and we write $\mathbf{X}_\pi = (X_{\pi(1)}, \dots, X_{\pi(n)})$ for $\pi \in \Pi_n$ and $\mathbf{X} \in \mathcal{X}^n$.

Property SM (Symmetry). For each $\mathbf{X} \in \mathcal{X}^n$ and $\pi \in \Pi_n$, $(\mathbf{A}^{\mathbf{X}})_\pi = \mathbf{A}^{\mathbf{X}_\pi}$.

Property SM reflects that consideration that if agents i and j exchange their initial risk contributions, then they also exchange their allocations. Hence, their identities or positions in the risk sharing pool does not matter; this clearly relates to both fairness and anonymity. Property SM is called the reshuffling property by Denuit et al. (2022a), and a similar property is called anonymity by Leshno and Strack (2020) in the setting of Bitcoin reward sharing.

Property SM is not directly assumed among our axioms, and CMRS satisfies Property SM by definition. Therefore, SM must follow from some of the axioms we impose. Since SM is very intuitive for an anonymized risk sharing rule, we wonder which axioms yield SM. It turns out that OA and ZP are sufficient for SM.

Proposition 4.5.4. *Axiom OA and Property ZP imply Property SM.*

We can briefly verify that SM does not come from any one of the four axioms alone. The mean-adjusted all-in-one risk sharing rule in Example 4.4.1 satisfies AF, RA and OA, but not SM or ZP. The combination of $\mathbf{A}_{\mathrm{all}}$ and $\mathbf{A}_{\mathrm{cm}}$ in Example 4.4.1 satisfies AF, RF (hence ZP) and RA, but not SM or OA.

4.6 Equilibrium analysis and optimality

In this section, we focus on equilibrium analysis of CMRS. An existing result by Denuit and Dhaene (2012) shows that $\mathbf{g}^{\mathbf{X}}(S)$ is Pareto-optimal if each component of $\mathbf{g}^{\mathbf{X}}$ are comonotonic. However, a number of interesting questions are worth exploring. For example, is the conditional mean risk sharing a Pareto equilibrium for some models of preferences? Is it possible to show that, for some preferences, and for all $\mathbf{X}$, the conditional mean risk sharing always gives a Pareto equilibrium? We intend to answer these questions in this section.

We say that an agent with preference (complete order) $\preceq$ on $\mathcal{X}$, with its strict relation denoted by $\prec$ and equivalence relation denoted by $\simeq$, is

(i) *monotone* if $X \geqslant Y$ implies $X \succeq Y$;

(ii) *risk-averse* if it is monotone and $X \geqslant_{\mathrm{cx}} Y$ implies $X \preceq Y$;

(iii) *strictly risk-averse* if it is risk-averse and $X \geqslant_{\mathrm{cx}} Y$ and $X \overset{\mathrm{d}}{\neq} Y$ imply $X \prec Y$.

Given $S \in \mathcal{X}$, we define the set of *allocations* of S as

$$\mathbb{A}_n(S) = \left\{ (X_1, \dots, X_n) \in \mathcal{X}^n : \sum_{i=1}^{n} X_i = S \right\}. \tag{4.6}$$

An allocation $(Y_1, \dots, Y_n) \in \mathbb{A}_n(S)$ of S is *Pareto optimal* if for any $(Z_1, \dots, Z_n) \in \mathbb{A}_n(S)$ satisfying $Z_i \succeq_i Y_i$ for all $i = 1, \dots, n$, we have $Z_i \simeq_i Y_i$ for all $i = 1, \dots, n$. Intuitively, one cannot find another allocation $(Z_1, \dots, Z_n)$ such that all agents are better off, with at least one of them strictly better off.

For an initial risk vector $(X_1, \dots, X_n) \in \mathbb{A}_n(S)$, an allocation $(Y_1, \dots, Y_n) \in \mathbb{A}_n(S)$ is *individually rational* if $Y_i \succeq_i X_i$ for all $i = 1, \dots, n$. Intuitively, individual rationality means that the situation of each agent has either improved or remain indifferent. This condition is necessary for the participation of these agents in the risk sharing problem.

Proposition 4.6.1. *The conditional mean risk sharing rule is individually rational for any n agents with risk-averse preferences.*

Below we show a weak result on Pareto optimality where we need to assume the agents have identical preferences. Choquet utilities of Yaari (1987) include, for instance, those induced by an Expected Shortfall (e.g., Wang and Zitikis (2021)).

Proposition 4.6.2. *Let $\mathbf{g}^{\mathbf{X}}$ be the conditional mean risk sharing rule for some $\mathbf{X} \in \mathcal{X}^n$ with sum S. If each component of $\mathbf{g}^{\mathbf{X}}$ is increasing, then $\mathbf{g}^{\mathbf{X}}(S)$ is Pareto optimal for any n agents with identical and risk-averse preferences numerically represented by a finite Choquet utility on $\mathcal{X}$.*

4.7 Generalized risk sharing rules with target information

In some applications, more information than simply the realized value of the total risk is observable, and one may wish to allocate risks according to such information; see Sections 4.8.1 and 4.8.2 below for real-world examples. This leads us to propose a generalization of risk sharing rules.

Denote by Σ a set of sub-σ-fields of $\mathcal{F}$. A *generalized risk sharing rule* is a mapping $\widehat{\mathbf{A}} : (\mathcal{X}^n \times \Sigma) \to \mathcal{X}^n$ satisfying $\widehat{\mathbf{A}}^{\mathbf{X}|\mathcal{G}} = (\widehat{A}_1^{\mathbf{X}|\mathcal{G}}, \dots, \widehat{A}_n^{\mathbf{X}|\mathcal{G}}) \in \mathbb{A}_n(S^{\mathbf{X}})$ for each $\mathbf{X} \in \mathcal{X}^n$ and $\mathcal{G} \in \Sigma$. The input σ-field $\mathcal{G}$ represents the information used to determine the realized values of the allocation, called *target information*. Note that $\sum_{i=1}^n \widehat{A}_i^{\mathbf{X}|\mathcal{G}} = S^{\mathbf{X}}$ implies that the σ-field of $\widehat{\mathbf{A}}^{\mathbf{X}|\mathcal{G}}$ must contain $\sigma(S^{\mathbf{X}})$ regardless of the choice of $\mathcal{G}$. There may not exist $\widehat{\mathbf{A}}$ such that $\widehat{\mathbf{A}}^{\mathbf{X}|\mathcal{G}}$ is $\mathcal{G}$-measurable for every $\mathcal{G} \in \Sigma$ and every $\mathbf{X} \in \mathcal{X}^n$, because $\mathcal{G}$ may not contain the information of $S^{\mathbf{X}}$.

To address this issue, we merge the information in $\sigma(S^{\mathbf{X}})$ into $\mathcal{G}$, and denote by $\mathcal{G}^{\mathbf{X}} = \sigma(S^{\mathbf{X}}, \mathcal{G})$ the σ-field generated by $S^{\mathbf{X}}$ and $\mathcal{G}$. Below, we present two properties describing how the information modeled by $\mathcal{G}$ and $\mathcal{G}^{\mathbf{X}}$ is used for the generalized risk sharing rule $\widehat{\mathbf{A}}$.

Property IA (Information anonymity). For $\mathbf{X} \in \mathcal{X}^n$ and $\mathcal{G} \in \Sigma$, $\widehat{\mathbf{A}}^{\mathbf{X}|\mathcal{G}}$ is $\mathcal{G}^{\mathbf{X}}$-measurable.

Property IA reflects on the idea that the risk allocations may not be solely determined by the realized value of $S^{\mathbf{X}}$ but also depend on other information represented by $\mathcal{G}$. Property IA is a generalization of Axiom RA. Property IA gives $\mathcal{G}^{\mathbf{X}}$-measurability instead of $\mathcal{G}$-measurability.

Property IB (Information backtracking). For each $\mathbf{X} \in \mathcal{X}^n$ and $\mathcal{G} \in \Sigma$, if $\mathbf{X}$ is $\mathcal{G}^{\mathbf{X}}$-measurable, then $\widehat{\mathbf{A}}^{\mathbf{X}|\mathcal{G}} = \mathbf{X}$.

Property IB is a generalization of Property BT. It reflects on the consideration that

getting an allocation determined by $\mathcal{G}^{\mathbf{X}}$ is our target.[5] More precisely, no risk exchange happens if the initial risk is already determined by $\mathcal{G}^{\mathbf{X}}$.

For a generalized risk sharing rule $\widehat{\mathbf{A}}$, we say that it satisfies an axiom or property introduced for risk sharing rules, if the mapping $\mathbf{X} \mapsto \widehat{\mathbf{A}}^{\mathbf{X}|\mathcal{G}}$ satisfies the corresponding axiom or property for each $\mathcal{G} \in \Sigma$. The next result characterizes the *generalized CMRS*, defined by

$$\widehat{\mathbf{A}}_{\mathrm{cm}}^{\mathbf{X}|\mathcal{G}} = \mathbb{E}\left[\mathbf{X}|\mathcal{G}^{\mathbf{X}}\right] \quad \text{for } \mathbf{X} \in \mathcal{X}^n \text{ and } \mathcal{G} \in \Sigma, \tag{4.7}$$

among all generalized risk sharing rules.

Theorem 4.7.1. *Assume $\mathcal{X} = L^1$ or L^1_+. A generalized risk sharing rule satisfies Axioms AF, RF and OA and Properties IA and IB if and only if it is the generalized CMRS.*

Theorem 4.7.1 indicates that, assuming AF, RF, OA, IA and IB, the only generalized risk sharing rule with a given target information $\mathcal{G}^{\mathbf{X}}$ needs to be calculated based on the conditional expectation with respect to $\mathcal{G}^{\mathbf{X}}$. This interpretation is similar to the result of Theorem 4.4.1. The generalized CMRS characterized in Theorem 4.7.1 will be useful as many practical applications involve allocations that are not solely determined by the total risk. We discuss some of them in Section 4.8.

4.8 Applications

In this section, we discuss a few examples of risk or reward sharing, including blockchain mining rewards, the revenue sharing of digital content, and P2P health care insurance, as illustrative examples of our axiomatic theory for anonymized risk sharing.

[5] For a given risk sharing rule $\mathbf{A}$, we can define $\widehat{\mathbf{A}}$ by $\widehat{\mathbf{A}}^{\mathbf{X}|\mathcal{G}} = \mathbf{A}^{\mathbf{X}}$ for each $\mathbf{X} \in \mathcal{X}^n$ and $\mathcal{G} \in \Sigma$; that is, the information $\mathcal{G}$ is ignored. In this case, $\mathbf{A}$ satisfies RA if and only if $\widehat{\mathbf{A}}$ satisfies IA. For instance, the generalized risk sharing rule $\widehat{\mathbf{A}}$ defined by $\widehat{\mathbf{A}}^{\mathbf{X}|\mathcal{G}} = \mathbf{A}_{\mathrm{cm}}^{\mathbf{X}}$ for each $(\mathbf{X}, \mathcal{G})$ satisfies Property IA, but this case is not interesting for this section, as it ignores $\mathcal{G}$; in particular, Property IB excludes this case, since $\widehat{\mathbf{A}}^{\mathbf{X}|\sigma(\mathbf{X})} = \mathbf{X}$ needs to hold by Property IB.

4.8.1 Blockchain mining rewards

A single Bitcoin mining pool

By the design of the Bitcoin protocol (Nakamoto (2008)), when a computational puzzle is solved by a decentralized network of anonymous computers, which are commonly called miners. After a block in the Bitcoin blockchain is successfully mined by a randomly selected miner, a predetermined number of bitcoins is rewarded to that miner. For more background on Bitcoin mining, including criticisms on its environmental and economic impact and the conflict between mining pools and decentralization, we refer to Chiu and Koeppl (2022), Eyal and Sirer (2018) and Leshno and Strack (2020). It is not our intention to say whether mining pools are good or bad; taking their existence as given, our focus is the design of reward sharing mechanisms within mining pools.

Mining activities are risky, with a large monetary value of the reward and a small probability of success, for individual miners. For this reason, mining pools are formed by groups of miners to share the risk. Risk-averse miners always have incentives to join mining pools to improve their utility.

Suppose that n miners form a mining pool to share the possible reward from mining the next block. Let the random variable $P > 0$ represent the monetary value of the next block at the time of solving the block. The miners' initial contribution vector is $\mathbf{X} = (\mathbb{1}_{D_1}, \dots, \mathbb{1}_{D_n})$, where $D_i \subseteq \Omega$ is the event that miner i successfully issues the next block,[6] and the probability $\mathbb{P}(D_i)$ represents the computational contribution of the miner i, measured by the number of hashes tried, divided by that of all miners in the world mining the block. Since the reward has a non-negative monetary value, we interpret positive values of allocations as rewards in this section, a different sign convention from the rest of the chapter; this causes no technical problem as all our results are invariant with respect to their signs. The events $D_1, \dots, D_n$ are mutually exclusive because at most one miner can issue the next block. We assume that these events are independent of the bitcoin price P because P is determined by market activities and $D_1, \dots, D_n$ are determined by randomly

[6] All sets $D \subseteq \Omega$ that appear in this section are assumed measurable, i.e., $D \in \mathcal{F}$.

trying solutions. Denote by $D = \bigcup_{i=1}^{n} D_i$ the event that any miner from this pool issues the block.

Our model addresses miners in mining pools. This setting is distinct from the settings of Eyal and Sirer (2018) and Leshno and Strack (2020), which focus on the reward mechanism of home miners, i.e., those who do not participate in mining pools. Moreover, in the setting of their papers, the reward to an individual miner is a binary outcome, characterized by the probability of receiving 1 block, whereas in our framework of a mining pool, miners can receive non-binary outcomes, depending on the pool performance and theirs shares in the pool.[7] This distinction is also useful in our later analysis of application of mining in multiple blockchains in Section 4.8.1.

To make our analysis of reward sharing rigorous, let $P \in \mathcal{X}$ be a fixed positive random variable, and denote by $\mathcal{B}_n$ the set of all possible contribution vectors from the miners, that is,

$$\mathcal{B}_n = \{(\mathbb{1}_{D_1}, \dots, \mathbb{1}_{D_n}) : D_1, \dots, D_n \subseteq \Omega \text{ are disjoint and independent of } P\}.$$

We assume that the probability space is rich enough so that a continuously distributed random variable independent of P exists. A *reward sharing rule* is a mapping $\mathbf{A} : \mathcal{B}_n \to \mathcal{X}^n$ satisfying $\mathbf{A}^{\mathbf{X}} = \mathbb{A}_n(S^{\mathbf{X}})$ for each $\mathbf{X} \in \mathcal{B}_n$ and $A_i^{\mathbf{X}} = A_j^{\mathbf{X}}$ for $i, j \in [n]$ with $\mathbb{P}(D_i) = \mathbb{P}(D_j)$ where $\mathbf{X} = P(\mathbb{1}_{D_1}, \dots, \mathbb{1}_{D_n})$. The last requirement reflects that only the amount of computational contribution of each miner is supplied (instead of the specification of the events D_i and D_j), as assumed in Eyal and Sirer (2018) and Leshno and Strack (2020). The four axioms of fairness and anonymity have natural interpretations and desirability in this setting.

(i) Axiom AF means that no agent gets less (or gets more) than their initial contribution in expectation, a simple form of fairness among anonymous participants.

(ii) Axiom RF (with a sign flip, i.e., $A_i^{\mathbf{X}} \geqslant \inf X_i$ for $i \in [n]$) means that any miner has a non-negative reward, since $\inf(P\mathbb{1}_{D_i}) = 0$ if $\mathbb{P}(D_i) \in [0, 1)$. Note that the case

[7] All axioms in Leshno and Strack (2020) are formulated on probabilities, and therefore they are different from our setting, where axioms are formulated on monetary outcomes.

$\mathbb{P}(D_i) = 1$ is trivial since all other agents have 0 contribution and 0 reward (using AF), and agent i receives the whole reward P.

(iii) Axiom RA means that the reward does not depend on which miner issues the block. If the block is issued by the pool, the rewards to miners depend on their computational contributions and the bitcoin price, but not the actual issuing miner. This feature is central to the idea of creating a mining pool and joining computational resources.

(iv) Axiom OA means that the mechanism is safe against merging and Sybil attacks, i.e., creating multiple accounts of the same participant. Recall that miners are represented by computers and online accounts, and merging, splitting, or creating them is not disclosed to other miners. Hence, such operations between some miners should not affect the reward to an uninvolved miner.

Anticipated from Theorem 4.4.1, the only reward sharing rule satisfying Axioms RA, RF, AF and OA should be CMRS. This is indeed true, although a separate proof is needed, as the set $\mathcal{B}_n$ is much smaller than $\mathcal{X}^n$ and it is not closed under addition, preventing us from directly applying Theorem 4.4.1 or its proof.

Proposition 4.8.1. *Assume $P \in \mathcal{X} = L^1$ and $P > 0$. A reward sharing rule $\mathbf{A} : \mathcal{B}_n \to \mathcal{X}^n$ satisfies Axioms RA, RF, AF and OA if and only if it is specified by*

$$A_i^{\mathbf{X}} = \frac{\mathbb{P}(D_i)}{\mathbb{P}(D)} P \mathbb{1}_D, \quad i \in [n], \ \mathbf{X} = P(\mathbb{1}_{D_1}, \dots, \mathbb{1}_{D_n}) \in \mathcal{B}_n, \tag{4.8}$$

which is CMRS.

The allocation (4.8) is precisely the common practice in mining pools; see Eyal and Sirer (2018) and Leshno and Strack (2020) in case $P = 1$. The total value P is shared proportionally to the computational contribution of each miner if this mining pool successfully issues the block (i.e., $\mathbb{1}_D = 1$), and the rewards are 0 otherwise. This reward sharing rule is CMRS, since $\mathbb{E}[P\mathbb{1}_{D_i}|P\mathbb{1}_D] = P\mathbb{1}_D\mathbb{E}[\mathbb{1}_{D_i}|\mathbb{1}_D] = P\mathbb{1}_D\mathbb{P}(D_i)/\mathbb{P}(D)$.[8] An example is shown in Figure 4.2 to illustrate how (4.8) works for three miners. Before joining a mining

[8] In this simple setting, CMRS also coincides with the mean proportional risk sharing rule.

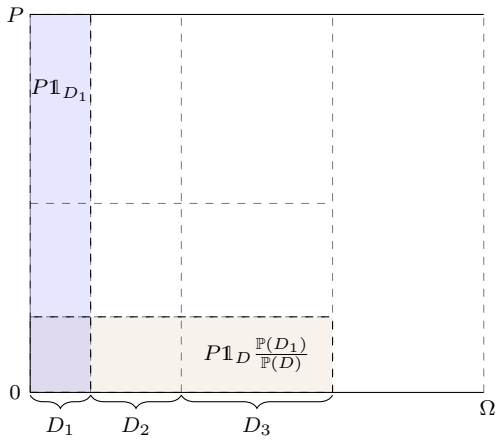


Figure 4.2: An illustration of a Bitcoin mining pool of 3 miners

pool, miner 1 gets the reward P if she issues the block (purple area in Figure 4.2) and otherwise she receives nothing. After miner 1 joins the group, the reward for her will be $P\mathbb{P}(D_1)/\mathbb{P}(D)$ if any of the three miners issues the block (brown area in Figure 4.2).

The new insight offered by Proposition 4.8.1 is that the reward sharing rule (4.8) is the unique possible mechanism if our four axioms are considered as desirable, and thus they fully rationalize the choice of this mechanism in practice.

Multiple mining pools

Next, suppose that miners can choose to participate in multiple pools by allocating their computational resources among these pools. We would like to pin down a suitable allocation rule in this setting with the help of the generalized CMRS and a specific choice $\mathcal{G}$ of target information.

There are m mining pools. Let $E_1, \ldots, E_m$ be mutually exclusive events where E_j represents the event that pool j successfully issues the block. Miners can choose to join one or several mining pools, and their initial risk vector is given by $\mathbf{X} = P(\mathbb{1}_{D_1}, \ldots, \mathbb{1}_{D_n})$ as in Section 4.8.1. Since the n miners form the m mining pools, we have the equality $\bigcup_{i=1}^{n} D_i =$

$\bigcup_{j=1}^m E_j$ and the decomposition $D_i = \sum_{j=1}^m \mathbb{1}_{D_i \cap E_j}$ where $D_i \cap E_j$ is the contribution of miner i to pool j.

Due to the separation of mining pools, we consider a generalized reward sharing rule with target information $\mathcal{G} = \sigma(P, \mathbb{1}_{E_1}, \ldots, \mathbb{1}_{E_m})$, that is, the information of the Bitcoin price P and the winning pool which successfully mines the block. For this choice of $\mathcal{G}$, the generalized CMRS $\widehat{\mathbf{A}}$ in (4.7) is given by

$$\widehat{A}_i^{\mathbf{X}|\mathcal{G}} = \sum_{j=1}^m \frac{\mathbb{P}(D_i \cap E_j)}{\mathbb{P}(E_j)} P \mathbb{1}_{E_j}. \tag{4.9}$$

Note that this rule can be easily implemented in practice as $\mathbb{P}(D_i \cap E_j)/\mathbb{P}(E_j)$ is the relative share of computational contribution of worker i to pool j. This rule can be equivalently explained by the mechanism in which all m pools are allocated separately and each of them uses CMRS.

Similarly to Proposition 4.8.1, we can show that (4.9) is the only generalized reward sharing rule with target information $\mathcal{G}$, and this rationalizes the practice of allocating rewards across multiple mining pools.

Multiple blockchains

We proceed to consider a pool of n miners with a collection of m cryptocurrencies (which we call coins) with random prices $P_1, \ldots, P_m$ in a pre-specified period of time. The computational contribution of miner i to coin j is fixed during this period of time. For simplicity, we assume that for each of these coins at most one block may be issued during this period of time. Denote by D_{ij} the event that miner i issues the block for coin j, and by $D^j = \bigcup_{i=1}^n D_{ij}$ is the event that coin j is successfully mined by the pool. The events D_{ij} are mutually exclusive across $i \in [n]$ for the same j. We further assume that D_{ij} is independent of $\{D_{k\ell} : k \in [n],\ \ell \in [m] \setminus \{j\}\}$, because issuing the block of one coin should not affect issuing the block of another one.

Similarly to Section 4.8.1, we assume that the prices $P_1, \ldots, P_m$ are independent to the issuance events. The initial risk vector is given by $\mathbf{X} = \sum_{j=1}^m P_j(\mathbb{1}_{D_{1j}}, \ldots, \mathbb{1}_{D_{nj}})$.

Finally, we consider the target information $\mathcal{G} = \sigma(P_1, \ldots, P_m, \mathbb{1}_{D^1}, \ldots, \mathbb{1}_{D^m})$, which is the information of the coin prices and the events of whether each of them is successfully mined.

For this choice of $\mathcal{G}$, the generalized CMRS $\widehat{\mathbf{A}}$ in (4.7) is given by

$$\widehat{A}_i^{\mathbf{X}|\mathcal{G}} = \sum_{j=1}^{m} \frac{\mathbb{P}(D_{ij})}{\mathbb{P}(D^j)} P_j \mathbb{1}_{D^j},$$

because $\mathbb{E}[P_j \mathbb{1}_{D_{ij}} | \mathcal{G}] = \mathbb{E}[P_j \mathbb{1}_{D_{ij}} | P_j \mathbb{1}_{D^j}] = P_j \mathbb{P}(D_{ij})/\mathbb{P}(D^j)$. In other words, each miner gets a proportion $\mathbb{P}(D_{ij})/\mathbb{P}(D^j)$ of each successfully mined coin, where the proportion is determined by its relative contribution to the pool for that particular coin.

4.8.2 Revenue sharing of digital content

Our next application concerns revenue sharing in subscription-based online platforms for digital content. The primary examples are music platforms such as Spotify, Deezer, or Apple Music; see Meyn et al. (2023) for a description of different revenue sharing schemes in subscription-based music platforms.

Suppose that there are n artists and m potential users in a specific month (many platforms collect subscription fees monthly). In this context, m is usually much larger than n. We assume that each user can subscribe to the platform because of one artist i, which is unobservable from the platform or the artist.

Let D_{ij} be the event that user j subscribes because of artist i, and D_{ij} are mutually singular across $i \in [n]$. Assume that the subscription events across different users are independent; i.e., D_{ij} is independent of $\{D_{k\ell} : k \in [n],\ \ell \in [m] \setminus \{j\}\}$ for each $i \in [n]$ and $j \in [m]$. Let $D^j = \bigcup_{i=1}^{n} D_{ij}$ be the event that user j subscribes to the platform, which is observable, and it generates a non-random revenue $q_j > 0$ (i.e., subscription fee, which may vary across users). If D^j does not occur, then user j does not subscribe to the platform during the considered month. Suppose that for $j \in [m]$, a proportion δ_j of q_j will be shared by the artists (the other proportion is kept by the platform or used to cover costs), and we denote by $p_j = \delta_j q_j$.

In this model, the initial risk vector is given by $\mathbf{X} = \sum_{j=1}^{m} p_j(\mathbb{1}_{D_{1j}}, \dots, \mathbb{1}_{D_{nj}})$, which is not observable to the platform. The target information is modelled by $\mathcal{G} = \sigma(\mathbb{1}_{D^1}, \dots, \mathbb{1}_{D^m})$, that is, the information based on the events of subscription. For this choice of $\mathcal{G}$, the generalized CMRS $\widehat{\mathbf{A}}$ in (4.7) is given by

$$\widehat{A}_i^{\mathbf{X}|\mathcal{G}} = \sum_{j=1}^{m} \frac{\mathbb{P}(D_{ij})}{\mathbb{P}(D^j)} p_j \mathbb{1}_{D^j}, \tag{4.10}$$

similarly to the model in Section 4.8.1.

Although $\mathbb{P}(D_{ij})$ and $\mathbb{P}(D^j)$ are not directly observable, their ratio $\mathbb{P}(D_{ij})/\mathbb{P}(D^j)$ can be estimated the ratio s_{ij} of the number of streams of user j using (e.g., listening to) the work of artist i to that of all streams of user j, and such data are available to the platform. Intuitively, the more user j uses the work of artist i, the more likely that user j subscribed because of artist i.

With the ratio $\mathbb{P}(D_{ij})/\mathbb{P}(D^j)$ estimated by s_{ij}, the revenue sharing mechanism (4.10) is exactly the *user-centric* remuneration model promoted by some platforms based on an argument of fairness.[9] We refer to Meyn et al. (2023) for a comparison of this revenue sharing rule with other rules. Thus, our framework provides a theoretical justification for the user-centric system of sharing digital content.

4.9 Concluding remarks

Decentralization in finance and insurance is getting increasing attention from both academia and the financial industry. As an important feature of decentralization, anonymity guarantees that agents are not distinguished by their preference, identity, private operations, and realized losses. Anonymized risk sharing is especially relevant in digital economy with applications including the revenue sharing of digital arts, and blockchain mining pools. Although there is extensive literature on axiomatic approaches in decision theory, so far there is no axiomatic theory for risk sharing.

[9]For instance, the platform `Deezer` is promoting the user-centric payment system; see https://www.deezer-blog.com/how-much-does-deezer-pay-artists/ (accessed April 2023).

This chapter presents an axiomatic theory of risk sharing in the context of anonymized risk sharing. Based on two axioms of fairness and two axioms of anonymity, we prove that the conditional mean risk sharing rule is the only risk sharing rule that satisfies the four axioms. Furthermore, the conditional mean risk sharing rule is only the rule that is compatible with the incentives of all risk-averse agents and satisfies the anonymity requirement. We do not see any general reasons to dispute any of the four axioms in the framework of anonymized risk sharing as a whole, although in certain specific applications the conditional mean risk sharing rule may be suboptimal for some individual agents. As such, our work serves as a theoretical support to the wide applications of the conditional mean risk sharing as a standard tool in many relevant applications in decentralized risk sharing.

As a potential limitation, the conditional mean risk sharing requires a full specification of the joint distribution of the risk contributions from the agents to compute. This is not a problem for the applications discussed in this chapter, due to the availability of the information. For some other applications, computational issues can be cumbersome for a large set of heterogeneous agents; for computing CMRS in some specific models, see Denuit (2019) and Denuit et al. (2022a) and the references therein.

4.10 Appendix: Technical details

4.10.1 Proofs in Section 4.4

We first prove Theorem 4.4.2, as it will be used in the proof of Theorem 4.4.1.

Proof of Theorem 4.4.2. First, we prove that ϕ is continuous. Suppose that $X_n \to X$ in $L^1(\Omega, \mathcal{F}, \mathbb{P})$. By using (b) and (c), we have

$$\phi(X_n) - \phi(X) = \phi(X_n - X) \leqslant \phi(|X_n - X|).$$

Similarly,

$$\phi(X_n) - \phi(X) = \phi(X_n - X) \geqslant \phi(-|X_n - X|) = -\phi(|X_n - X|).$$

Hence,

$$|\phi(X_n) - \phi(X)| \leqslant |\phi(|X_n - X|)| = \phi(|X_n - X|),$$

where the last equality is due to $\phi(|X_n - X|) \geqslant \phi(0) = 0$ by (a). Using (e), we have $\mathbb{E}[\phi(|X_n - X|)] = \mathbb{E}[|X_n - X|] \to 0$. Therefore, $\mathbb{E}[|\phi(X_n) - \phi(X)|] \leqslant \mathbb{E}[\phi(|X_n - X|)] \to 0$. This means $\phi(X_n) \to \phi(X)$ in $L^1(\Omega, \mathcal{G}, \mathbb{P})$, thus showing the continuity of $\phi : L^1(\Omega, \mathcal{F}, \mathbb{P}) \to L^1(\Omega, \mathcal{G}, \mathbb{P})$.

Next, we prove that ϕ is linear. Based on the fact that ϕ is (b) additive and (c) monotone, we have

$$\phi(X) \geqslant \phi(0) = 0 \qquad \text{for any } X \in L^1(\Omega, \mathcal{F}, \mathbb{P}) \text{ and } X \geqslant 0,$$

which implies that ϕ is linear (see, e.g., Theorem 1.10 of Aliprantis and Burkinshaw (2006)):

$$\phi(\alpha X + \beta Y) = \alpha\phi(X) + \beta\phi(Y) \qquad \text{for any } \alpha, \beta \in \mathbb{R} \text{ and } X, Y \in L^1(\Omega, \mathcal{F}, \mathbb{P}).$$

It further follows that ϕ is a positive operator on $\mathcal{X}$. Recall that a linear operator between two ordered vector spaces is a positive operator if it maps positive elements to positive elements.

Finally, we show that ϕ satisfies the following property

$$\phi(X) = X \text{ for all } \sigma(S)\text{-measurable } X. \tag{4.11}$$

For $t \in \mathbb{R}$, by (c) we have $\phi(S \vee t) \geqslant \phi(S) \vee \phi(t)$, and by (a) and (d) we get $\phi(S) \vee \phi(t) = S \vee t$. Since $\phi(S \vee t) \geqslant S \vee t$ and they have the same mean by (e), we know

$$\phi(S \vee t) = S \vee t \qquad \text{for all } t \in \mathbb{R}. \tag{4.12}$$

Write $T_{s,t} = \frac{1}{t-s}(S \vee t - S \vee s)$ for $t > s$. Note that for all $t > s$,

$$\mathbb{1}_{\{S \leqslant s\}} \leqslant T_{s,t} \leqslant \mathbb{1}_{\{S \leqslant t\}}. \tag{4.13}$$

The linearity of ϕ and (4.12) imply that $\phi(T_{s,t}) = T_{s,t}$. Using the above equality, (c) and (4.13), we have, for all $t > s$,

$$\phi\left(\mathbb{1}_{\{S \leqslant s\}}\right) \leqslant \phi(T_{s,t}) = T_{s,t} \leqslant \mathbb{1}_{\{S \leqslant t\}},$$

and

$$\phi\left(\mathbb{1}_{\{S\leqslant t\}}\right)\geqslant\phi\left(T_{s,t}\right)=T_{s,t}\geqslant\mathbb{1}_{\{S\leqslant s\}}.$$

It follows that for all $\varepsilon>0$,

$$\mathbb{1}_{\{S\leqslant t-\varepsilon\}}\leqslant\phi\left(\mathbb{1}_{\{S\leqslant t\}}\right)\leqslant\mathbb{1}_{\{S\leqslant t+\varepsilon\}}.$$

Hence, $\mathbb{1}_{\{S<t\}}\leqslant\phi\left(\mathbb{1}_{\{S\leqslant t\}}\right)\leqslant\mathbb{1}_{\{S\leqslant t\}}$ for all $t\in\mathbb{R}$. Using (e), we know $\phi\left(\mathbb{1}_{\{S\leqslant t\}}\right)=\mathbb{1}_{\{S\leqslant t\}}$. From this equality, using (a) and linearity of ϕ, it follows that $\phi\left(\mathbb{1}_{\{S>t\}}\right)=\mathbb{1}_{\{S>t\}}$ for all $t\in\mathbb{R}$.

Define the class

$$\mathcal{C}=\left\{C\in\mathcal{F}:\phi(\mathbb{1}_C)=\mathbb{1}_C\right\}.$$

Hence, $\{S\leqslant t\}\in\mathcal{C}$ for any $t\in\mathbb{R}$. We have $\Omega\in\mathcal{C}$. Using linearity of ϕ, we have that if $C\in\mathcal{C}$, then

$$\phi(\mathbb{1}_{C^c})=\phi(1-\mathbb{1}_C)=\phi(1)-\phi(\mathbb{1}_C)=1-\mathbb{1}_C=\mathbb{1}_{C^c},$$

which implies that the complement set $C^c\in\mathcal{C}$. Suppose that $\{C_i\}_{i\geqslant1}\subseteq\mathcal{C}$ are disjoint. We proceed to show that $\bigcup_{i=1}^{\infty}C_i\in\mathcal{C}$. Indeed, using monotonicity and additivity of ϕ, we have

$$\phi\left(\mathbb{1}_{\{\bigcup_{i=1}^{\infty}C_i\}}\right)\geqslant\phi\left(\mathbb{1}_{\{\bigcup_{i=1}^{m}C_i\}}\right)=\phi\left(\sum_{i=1}^{m}\mathbb{1}_{C_i}\right)=\sum_{i=1}^{m}\mathbb{1}_{C_i},\quad\text{for all } m\geqslant1.$$

Letting $m\to\infty$, we have

$$\phi\left(\mathbb{1}_{\{\bigcup_{i=1}^{\infty}C_i\}}\right)\geqslant\sum_{i=1}^{\infty}\mathbb{1}_{C_i}=\mathbb{1}_{\{\bigcup_{i=1}^{\infty}C_i\}}.$$

Based on (e), we have

$$\mathbb{E}\left[\phi\left(\mathbb{1}_{\{\bigcup_{i=1}^{\infty}C_i\}}\right)\right]=\mathbb{E}\left[\mathbb{1}_{\{\bigcup_{i=1}^{\infty}C_i\}}\right],$$

which implies $\phi\left(\mathbb{1}_{\{\bigcup_{i=1}^{\infty}C_i\}}\right)=\mathbb{1}_{\{\bigcup_{i=1}^{\infty}C_i\}}$ and $\bigcup_{i=1}^{\infty}C_i\in\mathcal{C}$. Hence, the class $\mathcal{C}$ is a σ-field and $\sigma(S)\subseteq\mathcal{C}$ based on the monotone class theorem. It follows that $\phi(\mathbb{1}_B)=\mathbb{1}_B$ for all $B\in\mathcal{G}$. Since any $\mathcal{G}$-measurable X can be upper and lower approximated by the summation of simple functions, using linearity and monotonicity we conclude that $\phi(X)=X$ for all $\mathcal{G}$-measurable X.

The conditions that ϕ is continuous, linear and monotone and satisfies (4.11) guarantee the representation of ϕ (see Proposition 2.6 of Filipović et al. (2012) or Theorem 1 of Pfanzagl (1967)), as

$$\phi(X) = \mathbb{E}[ZX|S] \qquad \text{for all } X \in L^1(\Omega, \mathcal{F}, \mathbb{P}), \tag{4.14}$$

for some $Z \geqslant 0$ satisfying $\mathbb{E}[Z|S] = 1$. Using (e), we get $1 = \mathbb{E}[Z] = \mathbb{E}[\phi(Z)] = \mathbb{E}[\mathbb{E}[Z^2|S]] = \mathbb{E}[Z^2]$. Since $\mathbb{E}[Z^2] = \mathbb{E}[Z] = 1$, we know $Z = 1$. Hence, we have $\phi(X) = \mathbb{E}[X|S]$ for $X \in \mathcal{X}$ and this completes the proof. □

Remark 4.10.1. The key step in the proof of Theorem 4.4.2 is to obtain the property (4.11); ϕ with such a property is sometimes called a projection. Several characterizations of the conditional expectation directly rely on this property; see Pfanzagl (1967) and Eisner et al. (2015). In particular, Theorem 1 of Pfanzagl (1967) holds for subspaces of $L^1(\Omega, \mathcal{F}, \mathbb{P})$, and hence our results in Theorems 4.4.1-4.4.3 hold for general $\mathcal{X} = L^q$ where $q \in [1, \infty]$.

Proof of Theorem 4.4.1. The "if" statement is checked in Section 4.4.1. We proceed to prove the "only if" statement. Let $\mathbf{A}$ be a risk sharing rule satisfying Axioms AF, RF, RA and OA. Fix any $S \in \mathcal{X}$. Define the mapping

$$h^S : \mathcal{X} \to L^1(\Omega, \sigma(S), \mathbb{P}), \ X \mapsto A_1^{(X, S-X, 0, \dots, 0)}.$$

Note that RA guarantees that h^S takes values in $L^1(\Omega, \sigma(S), \mathbb{P})$. We will verify that h^S satisfies the following properties on $\mathcal{X}$:

(a) constant preserving: $h^S(t) = t$ for all $t \in \mathbb{R}$;

(b) additivity: $h^S(X + Y) = h^S(X) + h^S(Y)$ for $X, Y \in \mathcal{X}$;

(c) monotonicity: $h^S(Y) \geqslant h^S(X)$ if $Y \geqslant X$;

(d) $h^S(S) = S$;

(e) $\mathbb{E}[h^S(X)] = \mathbb{E}[X]$ for $X \in \mathcal{X}$;

First, (a) follows directly from (4.1) and the definition of h^S. Next, we proceed to prove (b). By using Axiom OA, we have, for any $X, Y \in \mathcal{X}$ (note that here we use the fact that $n \geqslant 3$),

$$\begin{aligned} h^S(X+Y) &= A_1^{(X+Y,S-X-Y,0,\dots,0)} \\ &= A_1^{(X,S-X-Y,Y,0,\dots,0)} + A_3^{(X,S-X-Y,Y,0,\dots,0)} \\ &= A_1^{(X,S-X,0,\dots,0)} + A_3^{(0,S-Y,Y,0,\dots,0)} = h^S(X) + A_3^{(0,S-Y,Y,0,\dots,0)} \end{aligned} \tag{4.15}$$

where Axiom OA is used in the second and third equalities. In particular, by choosing $X = 0$ and using the fact that $h^S(0) = 0$ in (a), (4.15) implies $h^S(Y) = A_3^{(0,S-Y,Y,0,\dots,0)}$. Using this relationship and (4.15), we further have

$$h^S(X+Y) = h^S(X) + h^S(Y),$$

and hence (b) holds. Next, we show (c). Using Axiom RF, we have $h^S(X-Y) \leqslant 0$ if $X - Y \leqslant 0$. Hence, by (b), we obtain (c). Moreover, (d) is implied by the equality $A_1^{(S,0,\dots,0)} = S$ from (4.2). Finally, (e) follows from Axiom AF.

Using Theorem 4.4.2, (a)-(e) imply that h^S admits the representation

$$h^S(X) = \mathbb{E}[X|S] \text{ for all } X \in \mathcal{X}.$$

For any $\mathbf{X} \in \mathcal{X}^n$, let $S = S^{\mathbf{X}} = \sum_{i=1}^n X_i$. Using Axiom OA and the representation of h^S, we have

$$A_1^{\mathbf{X}} = A_1^{(X_1,S-X_1,0,\dots,0)} = h^S(X_1) = \mathbb{E}[X_1|S] = \mathbb{E}\left[X_1|S^{\mathbf{X}}\right].$$

Similarly, we have $A_j^{\mathbf{X}} = \mathbb{E}[X_j|S^{\mathbf{X}}]$ for any $j = 2, \dots, n$, which gives that $\mathbf{A}$ is CMRS. □

Proof of Proposition 4.4.1. (i) The Q-CMRS rule $\mathbf{A}^{\mathbf{X}}_{Q\text{-cm}} = \mathbb{E}^Q[\mathbf{X}|S^{\mathbf{X}}]$ satisfies Axioms RA, RF and OA with the same reasoning as the CMRS. Since $\mathbb{E}[\mathbf{A}^{\mathbf{X}}_{Q\text{-cm}}] = \mathbb{E}^Q[\mathbf{X}]$, AF does not hold as long as $Q \neq \mathbb{P}$.

(ii) For the mean-adjusted all-in-one risk sharing rule

$$\mathbf{A}^{\mathbf{X}}_{\text{ma}} = \left(S^{\mathbf{X}} - \mathbb{E}[S^{\mathbf{X}}], 0, \dots, 0\right) + \mathbb{E}[\mathbf{X}],$$

it is clear that Axioms RA and AF hold by definition. Axiom OA holds because the allocation to agent $i \in [n]$ is determined only by (X_i, S). Axiom RF does not hold because the allocation to agent 1 is not a constant if $S^{\mathbf{X}}$ is not a constant, regardless of whether X_1 is a constant, violating (4.1).

(iii) For the identity risk sharing rule $\mathbf{A}^{\mathbf{X}}_{\mathrm{id}} = \mathbf{X}$, it is clear that Axioms AF, RF and OA hold. Axiom RA does not hold because $\mathbf{A}^{\mathbf{X}}_{\mathrm{id}}$ is not necessarily a function of $S^{\mathbf{X}}$.

(iv) Consider a combination of $\mathbf{A}_{\mathrm{all}}$ and $\mathbf{A}_{\mathrm{cm}}$, defined by $\mathbf{A}^{\mathbf{X}} = \mathbf{A}^{\mathbf{X}}_{\mathrm{all}} = (S^{\mathbf{X}}, 0, \dots, 0)$ if $\mathbf{X}$ is standard Gaussian, and $\mathbf{A}^{\mathbf{X}} = \mathbf{A}^{\mathbf{X}}_{\mathrm{cm}} = \mathbb{E}[\mathbf{X}|S^{\mathbf{X}}]$ otherwise. Axioms AF, RF and RA and be checked separately for $\mathbf{A}_{\mathrm{all}}$ and $\mathbf{A}_{\mathrm{cm}}$, by noting that RF only needs to be checked for $\mathbf{A}_{\mathrm{cm}}$ since the standard Gaussian $\mathbf{X}$ is not included in the statement of RF.

To verify that OA does not hold, it suffices to consider $n = 3$. Let $\mathbf{X} = (X_1, X_2, X_3)$ follow a standard Gaussian distribution. By definition, $A_1^{\mathbf{X}} = S^{\mathbf{X}}$. However, for $\mathbf{Y} = (X_1, X_2 + X_3, 0)$, we have $A_1^{\mathbf{Y}} = \mathbb{E}[X_1|S^{\mathbf{X}}] = S^{\mathbf{X}}/3 \neq A_1^{\mathbf{X}}$, thus violating OA. □

Proof of Theorem 4.4.3. The "if" statement is checked in Section 4.4.1. We proceed to prove the "only if" statement. The gap between the proof of Theorem 4.4.1 and this result is that we need the following extension argument.

Let $\mathbf{A}$ be a risk sharing rule satisfying Axioms AF, RF, RA and OA. Fix any $S \in \mathcal{X} = L^1_+$. Denote by $B_S = \{X \in L^1_+ : X \leqslant S\}$, which is the set of random variables between 0 and S. Define the mapping as in the proof of Theorem 4.4.1,

$$h^S : B_S \to L^1(\Omega, \sigma(S), \mathbb{P}),\ X \mapsto A_1^{(X, S-X, 0, \dots, 0)}.$$

It is clear that h^S is well-defined on B_S and satisfies additivity

$$h^S(X+Y) = h^S(X) + h^S(Y) \text{ for } X, Y, X+Y \in B_S,$$

which can be checked by the same argument as in the proof of Theorem 4.4.1. Define $C_S = \{\lambda X : \lambda \in \mathbb{R}_+,\ X \in B_S\}$ which is the cone generated by B_S, and $L_S = \{\lambda X : \lambda \in$

$\mathbb{R},\ X \in B_S\}$ which is the linear space generated by B_S. According to Lemma 4.10.1 below, h^S can be uniquely extended on C_S and L_S and it is linear on L_S. This allows us to use the same arguments in Theorem 4.4.1 to get

$$h^S(X) = \mathbb{E}[X|S], \quad X \in L_S,$$

and following the rest of the steps for the proof of Theorem 4.4.1 yields that **A** is CMRS. □

Lemma 4.10.1. *Fix $S \in L^1_+$. Any additive function $\phi : B_S \to L^1_+$ has a unique additive extension on C_S and a unique linear extension on L_S.*

Proof. For $X \in C_S$, denote by $\gamma_X = \sup\{\gamma \in [0,1] : \gamma X \in B_S\}$. Note that there exists $\lambda_X \in \mathbb{R}_+$ and $Y \in B_S$ such that $\lambda_X Y = X$, and hence $\gamma_X \geqslant 1/\lambda_X > 0$. Moreover, we have $\gamma_X X \in B_S$ since B_S is closed. Define $\widehat{\phi}(X) = \phi(\gamma_X X)/\gamma_X$ for $X \in C_S$. It is clear that $\widehat{\phi} = \phi$ on B_S because $\lambda_X = 1$ for all $X \in B_S$. We next verify that $\widehat{\phi}$ is additive.

Take $m, k \in \mathbb{N}$ such that $m \leqslant k$. By additivity of ϕ on B_S, we have $\phi(mX/k) = m\phi(X/k)$ for $X \in B_S$. By taking $m = 1$, we get $\phi(X/k) = \phi(X)/k$, which in turn gives $\phi(mX/k) = m\phi(X)/k$. Since X is non-negative, positivity (monotonicity) of ϕ further gives $\phi(\lambda X) = \lambda\phi(X)$ for any real number $\lambda \in [0,1]$.

For any $X, Z \in C_S$ such that $Z \geqslant X$, since $\gamma_Z \leqslant \gamma_X$, we obtain, by choosing $\lambda = \gamma_Z/\gamma_X$,

$$\phi(\gamma_Z X) = \phi(\lambda\gamma_X X) = \lambda\phi(\gamma_X X) = \frac{\gamma_Z}{\gamma_X}\phi(\gamma_X X). \tag{4.16}$$

Take any $X, Y \in C_S$ and write $Z = X + Y$. Using (4.16) and additivity of ϕ on B_S,

$$\begin{aligned}\widehat{\phi}(X+Y) &= \frac{1}{\gamma_Z}\phi(\gamma_Z(X+Y)) \\ &= \frac{1}{\gamma_Z}\phi(\gamma_Z X) + \frac{1}{\gamma_Z}\phi(\gamma_Z Y) = \frac{1}{\gamma_X}\phi(\gamma_X X) + \frac{1}{\gamma_Y}\phi(\gamma_Y Y) = \widehat{\phi}(X) + \widehat{\phi}(Y).\end{aligned}$$

Therefore, $\widehat{\phi}$ is additive on C_S. The extension is unique because any two additive and monotone functions agreeing on B_S must agree on C_S. The unique linear extension to L_S follows from Theorem 1.10 of Aliprantis and Burkinshaw (2006). □

Proof of Theorem 4.4.4. Let $\mathbf{A}$ be a risk sharing rule satisfying Property LS, CP and Axioms RA, RF and OA, and fix any $S \in \mathcal{X}$. For each s in the range of S, define the mapping

$$h^{S,s} : \mathcal{X} \to \mathbb{R},\ X \mapsto A_1^{(X,S-X,0,\ldots,0)}(s).$$

we know that $h^{S,s}$ satisfies the following properties on $\mathcal{X}$:

(a) normalization: $h^{S,s}(t) = t$ for all $t \in \mathbb{R}$.

(b) additivity: $h^{S,s}(X+Y) = h^{S,s}(X) + h^{S,s}(Y)$ for $X, Y \in \mathcal{X}$.

(c) monotonicity: $h^{S,s}(Y) \geqslant h^{S,s}(X)$ if $Y \geqslant X$.

The arguments above have been proved in Theorem 4.4.1. Also, based on the fact that $h^{S,s}$ is additive and monotone, we have

$$h^{S,s}(X) \geqslant h^{S,s}(0) = 0 \qquad \text{for any } X \in L^1(\Omega, \mathcal{F}, \mathbb{P}) \text{ and } X \geqslant 0,$$

which implies that $h^{S,s}$ is linear (see e.g., Theorem 1.10 of Aliprantis and Burkinshaw (2006)):

$$h^{S,s}(\alpha X + \beta Y) = \alpha h^{S,s}(X) + \beta h^{S,s}(Y) \qquad \text{for any } \alpha, \beta \in \mathbb{R} \text{ and } X, Y \in L^1(\Omega, \mathcal{F}, \mathbb{P}).$$

Thus, $h^{S,s}$ is normalized, linear and monotone. By the Riesz representation theorem (see e.g., Theorem 27.10 of Aliprantis and Burkinshaw (1990)), we know that $h^{S,s}$ admits the representation

$$h^{S,s}(X) = \int X \mathrm{d}P_{S,s}, \qquad X \in \mathcal{X},$$

for some probability measure $P_{S,s}$.

Write the probability measure $\mathbb{P}_s(\cdot) = \mathbb{P}(\cdot|S=s)$. We claim that $P_{S,s} = \mathbb{P}_s$. Suppose that there exists $B \in \mathcal{F}$ such that $P_{S,s}(B) \neq \mathbb{P}_s(B)$. First, we consider the case $P_{S,s}(A) > \mathbb{P}_s(A)$. Let m be a positive integer such that

$$P_{S,s}(B) > \frac{1}{m} > \mathbb{P}_s(B).$$

Take disjoint events $B_1, \ldots, B_m$ each with probability $1/m$ under $\mathbb{P}_s$ such that $B_1 \supset B$. Such disjoint events exist since $\mathbb{P}_s$ is atomless. By Property LS, we have

$$P_{S,s}(B_j) = h^{S,s}(\mathbb{1}_{B_j}) = h^{S,s}(\mathbb{1}_{B_1}) = P_{S,s}(B_1) \geqslant P_{S,s}(B) > \frac{1}{m}$$

for $j = 2, \ldots, m$. Hence,

$$P_{S,s}(\Omega) = P_{S,s}\left(\bigcup_{i=1}^{m} B_i\right) \geqslant m P_{S,s}(B) > 1,$$

which contradicts the fact that $P_{S,s}$ is a probability measure. Next, we consider the case $P_{S,s}(B_1) < \mathbb{P}_s(B_1)$. Let m be a positive integer such that

$$P_{S,s}(B_1) < \frac{1}{m} < \mathbb{P}_s(B_1).$$

Take disjoint events $B_1, \ldots, B_m$ each with probability $1/m$ under $\mathbb{P}_s$ such that $B_1 \subset B$. By Property LS, we have

$$P_{S,s}(B_j) = h^{S,s}(\mathbb{1}_{B_j}) = h^{S,s}(\mathbb{1}_{B_1}) = P_{S,s}(B_1) \leqslant P_{S,s}(B_1) < \frac{1}{m}$$

for $j = 2, \ldots, m$. Hence,

$$P_{S,s}(\Omega) = P_{S,s}\left(\bigcup_{i=1}^{m} B_i\right) \leqslant m P_{S,s}(B_1) < 1,$$

which contradicts the fact that $P_{S,s}$ is a probability measure.

Hence, we have $P_{S,s}(\cdot) = \mathbb{P}_s = \mathbb{P}(\cdot|S = s)$. We now obtain

$$h^{S,s}(X) = \int X \mathrm{d}P_{S,s} = \mathbb{E}[X|S = s], \quad X \in \mathcal{X}.$$

Now, let us take an arbitrary $\mathbf{X} = (X_1, \ldots, X_n) \in \mathcal{X}^n$ with $S = \sum_{i=1}^n X_i$. Using Axiom OA again, we have, for each s in the range of S,

$$A_1^{\mathbf{X}}(s) = A_1^{(X_1, X_2 + \cdots + X_n, 0, \ldots, 0)}(s) = A_1^{(X_1, S - X_1, 0, \ldots, 0)}(s) = h^{S,s}(X_1) = \mathbb{E}[X_1|S = s].$$

Similarly, the same conclusion holds for all $A_i^{\mathbf{X}}$ for $i = 1, \ldots, n$. Therefore, we obtain that $\mathbf{A}$ is indeed the conditional mean risk sharing rule. □

4.10.2 Proofs in Section 4.5

Proof of Proposition 4.5.1. Since $X \leqslant_{\mathrm{cx}} Y$ implies $X \leqslant \sup X \leqslant \sup Y$ and $\mathbb{E}[X] = \mathbb{E}[Y]$, UI implies both RF and AF. Property CP follows from AF and RF as discussed in (4.1). □

Proof of Proposition 4.5.2. We only show the case that $\mathcal{X} = L^1$, as the case $\mathcal{X} = L^1_+$ is analogous. Let $\mathbf{A} : \mathcal{X}^n \to \mathcal{X}^n$ be a risk sharing rule satisfying Axioms AF, RF and OA. Fix $S \in \mathcal{X}$. For any $X \in L^1(\Omega, \sigma(S), \mathbb{P})$, $A_1^{(X,S-X,0,\dots,0)}$ is $\sigma(S)$-measurable, because it is $\sigma(X, S-X)$-measurable by assumption and $\sigma(S) = \sigma(X, S-X)$. Define the mapping

$$h^S : L^1(\Omega, \sigma(S), \mathbb{P}) \to L^1(\Omega, \sigma(S), \mathbb{P}),\ X \mapsto A_1^{(X,S-X,0,\dots,0)},$$

where we use the fact that $h^S(X)$ is $\sigma(S)$-measurable for $X \in L^1(\Omega, \sigma(S), \mathbb{P})$. The arguments in the proof of Theorem 4.4.1 yield that h^S satisfies the conditions in Theorem 4.4.2. By Theorem 4.4.2, h^S is the identity on $L^1(\Omega, \sigma(S), \mathbb{P})$. Using this and OA, for any $\mathbf{X} \in \mathbb{A}_n(S)$ and $X_1 \in L^1(\Omega, \sigma(S), \mathbb{P})$, we have

$$A_1^{\mathbf{X}} = A_1^{(X_1,S-X_1,0,\dots,0)} = h^S(X_1) = X_1.$$

The other case of $A_j^{\mathbf{X}}$ for $j \in [n]$ are similar. □

Proof of Corollary 4.5.1. The proof follows directly from Theorems 4.4.1 and 4.4.3, Proposition 4.5.1, and the fact that CMRS satisfies Property UI. □

Proof of Proposition 4.5.3. Fix a non-constant $S \in \mathcal{X}$, and write $h^S : \mathcal{X} \to \mathcal{X},\ X \mapsto A_1^{(X,S-X,0,\dots,0)}$. By ZP, we have $h^S(S) = S$. Using additivity (4.15) guaranteed by OA in the proof of Theorem 4.4.1, we have $h^S(2S) = 2h^S(S) = 2S$. This and additivity give $h^S(-S) = -S$, and therefore, $\mathbf{A}^{(-S,2S,0,\dots,0)} = (-S, 2S, 0, \dots, 0)$ is not comonotonic. □

Proof of Proposition 4.5.4. By Proposition 4.3.1, OA implies that $A_i^{\mathbf{X}}$ is determined by $(X_i, S^{\mathbf{X}})$ and $i \in [n]$. It suffices to show that $i \in [n]$ is also not relevant. Using ZP and OA, we have, for the pair $(1, 3)$ and any $X, S \in \mathcal{X}$,

$$A_1^{(X,S-X,0,\dots,0)} = S - A_2^{(X,S-X,0,\dots,0)} = S - A_2^{(0,S-X,X,0,\dots,0)} = A_3^{(0,S-X,X,0,\dots,0)}.$$

The other pairs (i, j) are similar. Therefore, $A_i^{\mathbf{X}}$ is determined by $(X_i, S^{\mathbf{X}})$ regardless of $i \in [n]$, showing that SM holds. □

4.10.3 Proofs in Section 4.6

Proof of Proposition 4.6.1 . Individual rationality is implied by Property UI and the fact that the preferences are risk-averse. □

Proof of Proposition 4.6.2. Let $\mathcal{U}$ be the Choquet utility representing the agents' preferences. Recall that a Choquet utility is comonotonic additive (Yaari (1987)). Therefore, since $g_1^{\mathbf{X}}, \dots, g_n^{\mathbf{X}}$ are increasing functions, we have

$$\sum_{i=1}^{n} \mathcal{U}(g_i^{\mathbf{X}}(S_n)) = \mathcal{U}\left(\sum_{i=1}^{n} g_i^{\mathbf{X}}(S_n)\right) = \mathcal{U}(S_n).$$

By the comonotone improvement theorem (e.g., Theorem 10.49 of Rüschendorf (2013)), we obtain, for any $(Y_1, \dots, Y_n) \in \mathbb{A}_n(S_n)$, there exists $(Y_1^*, \dots, Y_n^*)$ which is comonotonic, such that $Y_i^* \leqslant_{\mathrm{cx}} Y_i$ which implies $U(Y_i^*) \geqslant U(Y_i)$, $i = 1, \dots, n$. Hence,

$$\sum_{i=1}^{n} \mathcal{U}(Y_i) \leqslant \sum_{i=1}^{n} \mathcal{U}(Y_i^*) = \mathcal{U}\left(\sum_{i=1}^{n} Y_i^*\right) = \mathcal{U}(S_n) = \sum_{i=1}^{n} \mathcal{U}(g_i^{\mathbf{X}}(S_n)).$$

Therefore, $(Y_1, \dots, Y_n)$ cannot dominate $\mathbf{g}^{\mathbf{X}}(S_n)$ for these agents. Thus, the allocation $\mathbf{g}^{\mathbf{X}}(S_n)$ is Pareto optimal. □

4.10.4 Proofs in Section 4.8

Proof of Proposition 4.8.1. We have seen that CMRS satisfies the four axioms for mappings on general spaces, and hence also on $\mathcal{B}_n$, and it clearly satisfies the definition of a reward sharing rule. Below we show that the four axioms are sufficient for CMRS. Fix $D \subseteq \Omega$ independent of P and denote by

$$I_D = \{C \subseteq D : C \text{ is independent of } P\} \quad \text{and} \quad M_D = \{P\mathbb{1}_C : C \in I_D\}.$$

Define the mapping as in the proof of Theorem 4.4.1,

$$h^D : M_D \to L^1(\Omega, \sigma(P\mathbb{1}_D), \mathbb{P}), \ P\mathbb{1}_C \mapsto A_1^{(P\mathbb{1}_C, P(\mathbb{1}_D - \mathbb{1}_C), 0, \dots, 0)}.$$

We can check that h^D satisfies additivity

$$h^D(X+Y) = h^D(X) + h^D(Y) \text{ for } X, Y \in M_D \text{ with } X + Y \in M_D,$$

and monotonicity

$$h^D(X) \leqslant h^D(Y) \text{ for } X, Y \in M_D \text{ with } X \leqslant Y;$$

these statements can be shown by arguments using AF, RF and OA as in the proof of Theorem 4.4.1. For $m \in \mathbb{N}$, take $C_1, \dots, C_m \in I_D$ such that $\mathbb{P}(C_1) = \cdots = \mathbb{P}(C_m)$ and $\bigcup_{j=1}^m C_j = D$. Using $h^D(P\mathbb{1}_D) = P\mathbb{1}_D$ (guaranteed by AF and RF) and $h^D(P\mathbb{1}_{C_1}) = h^D(P\mathbb{1}_{C_j})$ for $j \in [m]$ (by the definition of a reward sharing rule), we get from additivity of h^D that

$$h^D(P\mathbb{1}_{C_1}) = \frac{P\mathbb{1}_D}{m} = P\mathbb{1}_D \frac{\mathbb{P}(C_1)}{\mathbb{P}(D)}.$$

Since C_1 is arbitrary, we get that, for any $C \in I_D$ with $\mathbb{P}(C) = \mathbb{P}(D)/m$,

$$h^D(P\mathbb{1}_C) = P\mathbb{1}_D \frac{\mathbb{P}(C)}{\mathbb{P}(D)} = P\mathbb{1}_D \mathbb{E}[\mathbb{1}_C | P\mathbb{1}_D] = \mathbb{E}[P\mathbb{1}_C | P\mathbb{1}_D]. \tag{4.17}$$

Using additivity again, we know that (4.17) holds for any $j \in [m]$ and $C \in I_D$ with $\mathbb{P}(C) = j\mathbb{P}(D)/m$, and finally, by monotonicity of h^D, we get (4.17) for all $C \in I_D$. Following the rest of the steps for the proof of Theorem 4.4.1 yields that $\mathbf{A}$ is CMRS on $\mathcal{B}_n$. □

Chapter 5

Testing mean and variance by e-processes

5.1 Introduction

Testing mean and variance in various settings is a classic problem in statistics. In parametric inference concerning testing the mean, well-known tests like Student's t-test and z-test, as well as tests related to variance such as the Chi-Squared test and the F-test, are commonly employed; see e.g., Lehmann et al. (1986). However, parametric tests always come with assumptions about the parameters of the population distribution from which samples are derived. Deviating from these assumptions can lead to significantly flawed results. For situations where these assumptions might be compromised, non-parametric methods provide a great alternative. Comprehensive and well-established methods of non-parametric techniques for testing means and variances can be found in e.g., Conover (1999) and Hollander et al. (2013). Different from the classic settings, we will consider the problem of testing composite hypotheses in which data are not stationary.

Suppose that a tester has sequentially arriving data points (possibly dependent) $X_1, X_2, \ldots$, each from an unknown distribution (possibly different). The tester is interested in testing

whether

$$\mathbb{E}[X_i|\mathcal{F}_{i-1}] \leqslant \mu_i \quad \text{and} \quad \mathrm{Var}(X_i|\mathcal{F}_{i-1}) \leqslant \sigma_i^2 \quad \text{for each } i, \tag{5.1}$$

where $\mathcal{F}_{i-1}$ is generated by observations before X_i, and μ_i and σ_i may also depend on past data; see Section 5.2 for a precise formulation. If independence is further assumed, then this problem reduces to the classic problem of testing mean and variance. Testing conditional mean and conditional variance is common in some contexts such as forecasting (e.g., Henzi and Ziegel (2022)) and financial risk assessment (e.g., Fissler and Ziegel (2016)).

Problem (5.1) can be interpreted in two different ways (we omit "conditional" here):

(A) testing both the mean and the variance;

(B) testing the mean under the knowledge of an upper bound on the variance.

The interpretation (A) is relevant when the tester is interested in whether a time-series has switched away from a given regime with specified mean and variance bounds. We mainly use interpretation (A), while keeping in mind that interpretation (B) is useful when comparing with the literature. Of course, one could also interpret it as testing the variance knowing the mean.

Clearly, problem (5.1) is a composition of many complicated, non-parametric, composite hypotheses on each observation. The key challenge in this setting is that the data points are not iid, and hence we cannot make inference of the distributions themselves.

This problem can be addressed with the following general methodology, called e-tests or tests by betting, a successful example being Waudby-Smith and Ramdas (2024). We first consider a simpler problem: constructing an e-value from one random variable from each data point with the corresponding hypothesis on its mean and variance (which corresponds to $n = 1$). For a general background on e-values in hypothesis testing, see Vovk and Wang (2021), Grünwald et al. (2023), and the review by Ramdas et al. (2023). After obtaining these e-values, we combine them, usually by forming an e-process, to construct a test for the overall hypothesis. Alternatively, we can construct p-values instead of e-values, but the power of such a strategy is usually quite weak, as seen from our experiments.

Section 5.2 formally describes the hypotheses and defines e-variables, e-processes, and p-variables. As mentioned above, we will first address the case of one data point, i.e., $n = 1$, presented in Section 5.3. We consider four types of composite hypotheses on mean, variance and the shape of the distribution: symmetry, unimodality and their combination. Our main results are ways that are *optimal*, in a natural sense, to constructions of p-values and e-values in this setting. Although our main methodology is based on e-processes, we present results also for p-values, which may be useful in multiple testing (not treated in this chapter). Considering a non-parametric composite hypothesis with a given mean and variance as the baseline case, assuming symmetry approximately improves the baseline p-variable by a multiplicative factor of $1/2$, unimodality by a factor of $4/9$, and both by a factor of $2/9$. Similarly, the corresponding baseline e-variable is improved by multiplicative factors of 2, 1, and 2, respectively, in these scenarios.

We propose in Section 5.4 several methods to test using multiple data points, thus addressing the main task of the tester. The main proposals are e-process based tests, which follow the idea of testing by betting in Shafer (2021), Wasserman et al. (2020) and Waudby-Smith and Ramdas (2024). Although we mainly focus on one-sided hypotheses, our methodology can be easily adapted to test the two-sided hypothesis on the mean, that is,

$$\mathbb{E}[X_i|\mathcal{F}_{i-1}] \in [\mu_i^L, \mu_i^U] \quad \text{and} \quad \mathrm{Var}(X_i|\mathcal{F}_{i-1}) \leqslant \sigma_i^2 \quad \text{for each } i,$$

where $[\mu_i^L, \mu_i^U]$ is an interval or a singleton for each i; this is discussed in Section 5.4.3.

The closest methodological work related to this chapter is Waudby-Smith and Ramdas (2024), where the authors test in a non-parametric setting the conditional mean of sequential data, which are assumed to be bounded within a pre-specified range, thus a generally smaller class of distributions. Our problem and methodology are different from Waudby-Smith and Ramdas (2024) in the sense that we assume a bounded variance instead of a bounded range. Since a bounded range implies bounded variance, the assumption needed to apply our methodology is weaker than in the setting of Waudby-Smith and Ramdas (2024), following interpretation (B) of the main testing problem. Moreover, we are able to utilize the additional information on the distributional shape to obtain better e-values than without such information. A great advantage of the tests of Waudby-Smith and Ramdas

(2024) is that their power adapts to the unknown true variance of the distribution if data come from an iid population. Our method based on the growth rate of empirical e-values has a similar feature, which uses a betting strategy similar to that of Waudby-Smith and Ramdas (2024). Another closely related methodology is Wang et al. (2022), where, other statistical functions are tested other than the mean. Once e-variables are constructed, we will build e-processes in a similar way to Wang et al. (2022). The methods of Howard et al. (2020, 2021) and Wang and Ramdas (2023) based on exponential test supermartingales, which are e-processes, can also be applied to test (5.1). These methods differ from ours as our e-process is obtained by combining individual e-variables.

Section 5.5 provides simulation studies for the proposed methods and compare them with the method of Waudby-Smith and Ramdas (2024) when the model has both bounded support and bounded variance and with methods based on exponential test supermartingale of Howard et al. (2021) and Wang and Ramdas (2023). Section 5.6 contains empirical studies using financial asset return data during the 2007–2008 financial crisis, further demonstrating the effectiveness of the e-process based methods. Section 5.7 concludes the chapter. All proofs in the chapter are provided in Section 5.8.

5.2 General setting

5.2.1 Hypotheses to test

We first describe our main testing problem. Let n be a positive integer or ∞, and denote by $[n] = \{1, \dots, n\}$. Throughout, fix a sample space. Suppose that data points $(X_i)_{i\in[n]}$ arrive sequentially, each possibly from a different distribution, and not necessarily independent. A hypothesis is a collection H of probability measures that govern $(X_i)_{i\in[n]}$. Denote by $\mathcal{F}_i$ the σ-field generated by $X_1, \dots, X_i$ for $i \in [n]$ with $\mathcal{F}_0$ being the trivial σ-field. The main hypotheses of interest are variations (by adding shape information) of the following hypothesis

$$H = \left\{Q : \mathbb{E}^Q[X_i|\mathcal{F}_{i-1}] \leqslant \mu_i \text{ and } \mathrm{Var}^Q(X_i|\mathcal{F}_{i-1}) \leqslant \sigma_i^2 \text{ for } i \in [n]\right\}, \tag{5.2}$$

where μ_i and σ_i are $\mathcal{F}_{i-1}$-measurable for each $i \in [n]$; that is, they can be data-dependent on past observations. A simple case is

$$H = \left\{ Q : \mathbb{E}^Q[X_i|\mathcal{F}_{i-1}] \leqslant \mu \text{ and } \mathrm{Var}^Q(X_i|\mathcal{F}_{i-1}) \leqslant \sigma^2 \text{ for } i \in [n] \right\}, \tag{5.3}$$

where μ and σ are two constants; that is, we would like to test whether data exhibit conditional mean and conditional variance in $(-\infty, \mu] \times [0, \sigma^2]$. Although (5.3) looks simpler, it is indeed equivalent to (5.2) by noting that μ_i and σ_i can be absorbed into X_i since they are $\mathcal{F}_{i-1}$ measurable. Therefore, we will focus on the formulation (5.3) for the rest of the chapter. If data are independent (but not necessarily identically distributed), then the problem is to test the unconditional mean and variance.

We will further consider hypotheses with additional shape information, by assuming that some (or all) of the distributions of $X_1, \dots, X_n$ are unimodal, symmetric, or both. A distribution on $\mathbb{R}$ is *unimodal* if there exists $x \in \mathbb{R}$ such that the distribution has an increasing density on $(-\infty, x)$ and a decreasing density on (x, ∞); it may have a point-mass at x. A distribution on $\mathbb{R}$ with mean μ is *symmetric* if for all $x \in \mathbb{R}$ it assigns equal probabilities to $(-\infty, \mu - x]$ and $[\mu + x, \infty)$. If a distribution with mean μ is both unimodal and symmetric, then its mode must be either μ or an interval centered at μ.

Remark 5.2.1. The main question in Waudby-Smith and Ramdas (2024) is to test the conditional mean m with data taking values in $[0, 1]$. We note that any random variable with mean at most m and range $[0, 1]$ has variance at most $1/4$ (if $m \geqslant 1/2$) or $m(1-m)$ (if $m < 1/2$), attained by a Bernoulli random variable. Therefore, our hypothesis with $\mu = m$ and $\sigma^2 = 1/4$ or $\sigma^2 = m(1-m)$ has less restrictive assumptions than their setting (except they formulated two-sided hypotheses; see Remark 5.2.2 below) and in particular, our setting can handle unbounded data.

Remark 5.2.2. Our hypotheses are formulated as one-sided on both μ and σ^2. Certainly, all validity results remain true for the two-sided hypotheses. Testing $\mathbb{E}^Q[X_i] \geqslant \mu$ is symmetric to testing $\mathbb{E}^Q[X_i] \leqslant \mu$, but such symmetry does not hold for testing the variance. Building e-processes to test the two-sided hypothesis on the mean is discussed in Section 5.4.3.

5.2.2 P-variables and e-variables

We formally define p-variables and e-variables, following Vovk and Wang (2021). A *p-variable* P for a hypothesis H is a random variable that satisfies $Q(P \leqslant \alpha) \leqslant \alpha$ for all $\alpha \in (0,1)$ and all $Q \in H$. In other words, a p-variable is stochastically larger than $\mathrm{U}[0,1]$, often truncated at 1. An *e-variable* E for a hypothesis H is a $[0,\infty]$-valued random variable satisfying $\mathbb{E}^Q[E] \leqslant 1$ for all $Q \in H$. E-variables are often obtained from stopping an *e-process* $(E_t)_{t\geqslant 0}$, which is a nonnegative stochastic process adapted to a pre-specified filtration, $(\mathcal{F}_i)_{i\in[n]}$ in our problem, such that $\mathbb{E}^Q[E_\tau] \leqslant 1$ for any stopping time τ and any $Q \in H$.

Some p-variables and e-variables are useless, like $P = 1$ or $E = 0$. A p-variable P for H is *precise* if $\sup_{Q\in H} Q(P \leqslant \alpha) = \alpha$ for each $\alpha \in (0,1)$, and an e-variable E for H is *precise* if $\sup_{Q\in H} \mathbb{E}^Q[E] = 1$. In other words, a p-variable or an e-variable being precise means that it is not wasteful in a natural sense. A p-variable P is *semi-precise* for H if $\sup_{Q\in H} Q(P \leqslant \alpha) = \alpha$ for each $\alpha \in (0,1/2]$. Semi-precise p-variables require the sharp probability bound $\sup_{Q\in H} Q(P \leqslant \alpha) = \alpha$ only for the case $\alpha \leqslant 1/2$ which is relevant for testing purposes. We will see that for some hypotheses, precise p-variables do not exist, but semi-precise ones do.

Realizations of p-variables and e-variables are referred to as p-values and e-values.

5.3 Best p- and e-variables for one data point

We begin by considering the simple setting where one data point X is available, from which we will build a p-variable or e-variable for the hypothesis. Although it may be unconventional to test based on one observation, there are several situations where this construction becomes useful.

1. Testing by betting: To construct an e-process, one needs to sequentially obtain one e-value from each observation (or a batch of observations). This is the main setting in the current chapter.

2. Testing multiple hypotheses: One observation is obtained for each hypothesis, and p-values or e-values for each of them are computed and fed into a multiple testing procedure such as that of Benjamini and Hochberg (1995); this setting is particularly relevant for e-values as in Wang and Ramdas (2022). Even if for some hypotheses there is only one data point, a p-value or e-value (even moderate, say $e = 0.8$ or $e = 1.2$) from this hypothesis may be useful for the overall testing problem; see Ignatiadis et al. (2023) where e-values are used as weights, so $e = 0.8$ or $e = 1.2$ matters.

3. Testing a global null: One may first obtain a p-value or e-value for each experiment and then combine them to test the global null, as in meta-analysis; see Vovk and Wang (2020, 2021) and the references therein.

E-values are relevant for all of the three contexts, and p-values are relevant for the second and the third contexts.

Throughout, we make the assumption that a larger value of X indicates stronger evidence against the null; this is intuitive because we are testing the mean less or equal to μ in (5.3). This means that a p-variable is a decreasing function of X and an e-variable is an increasing function of X (always in the non-strict sense).

Remark 5.3.1. In the contexts of multiple testing and sequential e-values, the dependence among several e-values or p-values obtained is preserved from the dependence among the data points, if the monotonicity assumption above holds. This will be helpful when applying statistical methods based on dependence assumptions; see Benjamini and Yekutieli (2001) for the BH (Benjamini and Hochberg (1995)) procedure with positive dependence and Chi et al. (2022) for BH with negative dependence. Both concepts of dependence are preserved under monotone transforms.

5.3.1 Two technical lemmas

The following lemma establishes that the minimum of p-variables based on the same data point X is still a p-variable. This result relies on our assumption that p-variables are decreasing functions of X.

Lemma 5.3.1. *For a given observation X and hypothesis H, the infimum of p-variables is a p-variable. As a consequence, there exists a smallest p-variable.*

Although the smallest p-variable for H exists, it may not be precise. Indeed, in Theorems 5.3.2 and 5.3.4 below we will see that there may not exist any precise p-variable for some hypotheses.

The following lemma allows us to convert conditions on distribution functions into conditions on the corresponding quantile functions. For a probability measure Q, denote by

$$T_Y^Q(\alpha) = \inf\{x \in \mathbb{R} : Q(Y \leqslant x) \geqslant \alpha\} \quad \text{for } \alpha \in (0,1);$$

that is, T_Y^Q is the left-quantile function of Y under Q.

Lemma 5.3.2. *For a random variable P and a hypothesis H,*

(i) P is a p-variable if and only if $\inf_{Q\in H} T_P^Q(\alpha) \geqslant \alpha$ for all $\alpha \in (0,1)$;

(ii) P is a precise p-variable if and only if $\inf_{Q\in H} T_P^Q(\alpha) = \alpha$ for all $\alpha \in (0,1)$;

(iii) P is a semi-precise p-variable if and only if $\inf_{Q\in H} T_P^Q(\alpha) = \alpha$ for all $\alpha \in (0,1/2)$ and $\inf_{Q\in H} T_P^Q(\alpha) \geqslant \alpha$ for $\alpha \in [1/2,1)$.

The proof of Lemma 5.3.2 is essentially identical to that of Lemma 1 of Vovk and Wang (2020), which gives the equivalence between probability statements and quantile statements for merging functions of p-values.

5.3.2 Main results

Recall that we have only one observation, denoted by X. We consider the following four classes of non-parametric composite hypotheses, where $\mu \in \mathbb{R}$ and $\sigma > 0$.

$$\begin{aligned}
H(\mu,\sigma) &= \left\{Q : \mathbb{E}^Q[X] \leqslant \mu \text{ and } \mathrm{Var}^Q(X) \leqslant \sigma^2\right\};\\
H_{\mathrm{S}}(\mu,\sigma) &= \{Q \in H(\mu,\sigma) : X \text{ is symmetrically distributed}\};\\
H_{\mathrm{U}}(\mu,\sigma) &= \{Q \in H(\mu,\sigma) : X \text{ is unimodally distributed}\};\\
H_{\mathrm{US}}(\mu,\sigma) &= H_{\mathrm{U}}(\mu,\sigma) \cap H_{\mathrm{S}}(\mu,\sigma).
\end{aligned}$$

For our main results on the "best" p-variables and e-variables, it will be clear from our proofs that the condition $\mathrm{Var}^Q(X) \leqslant \sigma^2$ in each hypothesis can be replaced by $\mathrm{Var}^Q(X) = \sigma^2$, and the condition $\mathbb{E}^Q[X] \leqslant \mu$ in each hypothesis can be replaced by $\mathbb{E}^Q[X] = \mu$. All results remain true with any combinations of the above alternatives. Possible improvement for the two-sided test is discussed in Section 5.4.3.

The above four sets of distributions are studied in a very different context by Li et al. (2018) to compute worst-case risk measures under model uncertainty in finance. Some of our techniques for constructing p-variables use results from Li et al. (2018) and Bernard et al. (2020) for finding bounds on quantiles (called the Value-at-Risk in finance).

In what follows, for $x \in \mathbb{R}$, we write $x_+ = \max\{x, 0\}$, $x_- = \max\{-x, 0\}$, $x_+^2 = (x_+)^2$, and $x_-^2 = (x_-)^2$. We first consider the simplest case of testing $H(\mu, \sigma)$.

Theorem 5.3.1. *A precise p-variable for $H(\mu, \sigma)$ is $P = (1 + (X - \mu)_+^2/\sigma^2)^{-1}$, and a precise e-variable for $H(\mu, \sigma)$ is $E = (X - \mu)_+^2/\sigma^2$.*

It may be interesting to compare P and $1/E$ obtained from Theorem 5.3.1. Note that any e-variable can be converted into a p-variable via the so-called calibrator $e \mapsto \min\{1/e, 1\}$ (Vovk and Wang (2021)). As $1/E$ is a p-variable for an e-variable E, we have $P \leqslant 1/E$. In Theorem 5.3.1, we obtain $1/P = 1 + E > E$, as expected.

In the subsequent analysis, we will compare p-variables and e-variables for other hypotheses with those in Theorem 5.3.1. For a concise presentation, we will always write

$$P_0 = (1 + (X - \mu)_+^2/\sigma^2)^{-1} \quad \text{and} \quad E_0 = (X - \mu)_+^2/\sigma^2, \tag{5.4}$$

which are the p-variable and e-variable in Theorem 5.3.1, and note the connection $P_0 = (1 + E_0)^{-1}$.

We next consider the hypothesis $H_{\mathrm{S}}(\mu, \sigma)$ of symmetric distributions.

Theorem 5.3.2. *A semi-precise p-variable for $H_{\mathrm{S}}(\mu, \sigma)$ is $P = \min\{(2E_0)^{-1}, P_0\}$, and a precise e-variable for $H_{\mathrm{S}}(\mu, \sigma)$ is $E = 2E_0$. Precise p-variables do not exist for $H_{\mathrm{S}}(\mu, \sigma)$.*

From Theorem 5.3.2, the e-variable for $H_S(\mu, \sigma^2)$, which we denote by E_{S} is improved by a factor of two from E_0 for $H(\mu, \sigma^2)$ due to the additional assumption of symmetry.

On the other hand, the p-variable in Theorem 5.3.2, denoted by P_{S}, is improved from P_0 by taking an extra minimum with $1/E_{\mathrm{S}}$. In the most relevant case that $P_0 \leqslant 1/2$, or equivalently, $E_0 \geqslant 1$ (i.e., there is some evidence against the null), we have $P_{\mathrm{S}} = 1/E_{\mathrm{S}}$.

Next, we will see that the hypothesis $H_{\mathrm{U}}(\mu, \sigma)$ of unimodal distributions admits the same precise e-variable but a quite improved p-variable, compared to P_0 and E_0. This class includes, for instance, the commonly used gamma, beta, and log-normal distributions.

Theorem 5.3.3. *A precise p-variable for $H_{\mathrm{U}}(\mu, \sigma)$ is*

$$P = \max\left\{\frac{4}{9}P_0, \frac{4P_0 - 1}{3}\right\},$$

and a precise e-variable for $H_{\mathrm{U}}(\mu, \sigma)$ is $E = E_0$.

We denote the p-variable in Theorem 5.3.3 by P_{U} and the e-variable by E_{U}. If P_0 is smaller than $3/8$, corresponding to $(X - \mu)/\sigma > \sqrt{5/3}$, then $P_{\mathrm{U}} = 4P_0/9$; that is, the unimodality assumption reduces the p-variable by a multiplicative factor of $4/9$ compared to $H(\mu, \sigma)$. On the other hand, the e-variable E_{U} does not get improved at all compared to E_0.

Finally, we consider the hypothesis $H_{\mathrm{US}}(\mu, \sigma)$ of unimodal-symmetric distributions. This class includes, for instance, the popular normal, t-, and Laplace distributions. To construct a semi-precise p-variable for this hypothesis, we will use the following lemma of quantile bounds within $H_{\mathrm{US}}(\mu, \sigma)$, which may be of independent interest. In what follows, $\mathbb{1}$ is the indicator function; that is, $\mathbb{1}_A(x) = 1$ if $x \in A$ and $\mathbb{1}_A(x) = 0$ otherwise.

Lemma 5.3.3. *For $\alpha \in (0, 1)$, it holds that*

$$\sup_{Q \in H_{\mathrm{US}}(0,1)} T_X^Q(1 - \alpha) = \sqrt{\frac{2}{9\alpha}}\mathbb{1}_{(0,1/6]}(\alpha) + \sqrt{3}(1 - 2\alpha)\mathbb{1}_{(1/6,1/2]}(\alpha).$$

The general formula for $H_{\mathrm{US}}(\mu, \sigma)$ can be easily obtained from Lemma 5.3.3 via

$$\sup_{Q \in H_{\mathrm{US}}(\mu,\sigma)} T_X^Q(1 - \alpha) = \mu + \sigma \sup_{Q \in H_{\mathrm{US}}(0,1)} T_X^Q(1 - \alpha).$$

Theorem 5.3.4. *A semi-precise p-variable for $H_{\mathrm{US}}(\mu,\sigma)$ is*

$$P = \frac{2}{9E_0}\mathbb{1}_{[4/3,\infty)}(E_0) + \frac{3-\sqrt{3E_0}}{6}\mathbb{1}_{(0,4/3)}(E_0) + \mathbb{1}_{\{0\}}(E_0).$$

and a precise e-variable for $H_{\mathrm{US}}(\mu,\sigma)$ is $E = 2E_0$. Precise p-variables do not exist for $H_{\mathrm{US}}(\mu,\sigma)$.

We denote the p-variable obtained from Theorem 5.3.4 by P_{US} and the e-variable by E_{US}. One may check that P_{US} is smaller than both P_{U} and P_{S} unless $X \leqslant \mu$ (in which case they are equal to 1). For $(X-\mu)/\sigma \geqslant \sqrt{5/3}$, or equivalently, $P_0 \leqslant 3/8$, we have the following simple relation:

$$P_{\mathrm{S}} = \frac{P_0}{2(1-P_0)}, \quad P_{\mathrm{U}} = \frac{4}{9}P_0, \quad \text{and} \quad P_{\mathrm{US}} = \frac{2P_0}{9(1-P_0)},$$

implying the order $P_0 > P_{\mathrm{S}} > P_{\mathrm{U}} > P_{\mathrm{US}}$ unless $P_0 = 0$. For instance, if we observe $(X-\mu)/\sigma = 3$, then the p-values are

$$P_0 = \frac{1}{10} = 0.1, \quad P_{\mathrm{S}} = \frac{1}{18} \approx 0.056, \quad P_{\mathrm{U}} = \frac{2}{45} \approx 0.044, \quad \text{and} \quad P_{\mathrm{US}} = \frac{2}{81} \approx 0.025.$$

On the other hand, the corresponding e-values are

$$E_0 = 9, \quad E_{\mathrm{S}} = 18, \quad E_{\mathrm{U}} = 9, \quad \text{and} \quad E_{\mathrm{US}} = 18.$$

For a comparison, if we are testing the simple parametric hypothesis $\mathrm{N}(0,1)$ against $\mathrm{N}(3,1)$ with one observation $X = 3$, then the corresponding (Neyman-Pearson) p-value is 0.00135 and the corresponding likelihood ratio e-value is 90.02. This is not surprising as generally p-values and e-values built for composite hypotheses are more conservative than those for simple hypotheses based on the same data.

We summarize our construction formulas for p-variables and e-variables in Table 5.1 by breaking them down using ranges of X. To obtain the formulas for a general (μ,σ) other than $(0,1)$, it suffices to replace X in Table 5.1 by $(X-\mu)/\sigma$.

We conclude the section by making two technical remarks on the obtained results.

Hypothesis	p-variable		e-variable
$H(0,1)$	$(1+X_+^2)^{-1}$		X_+^2
$H_{\mathrm{S}}(0,1)$	$\frac{1}{2}X^{-2}$	if $X \geqslant 1$	$2X_+^2$
	$(1+X_+^2)^{-1}$	if $X < 1$	
$H_{\mathrm{U}}(0,1)$	$\frac{4}{9}(1+X^2)^{-1}$	if $X \geqslant \sqrt{5/3}$	X_+^2
	$\frac{4}{3}(1+X_+^2)^{-1} - \frac{1}{3}$	if $X < \sqrt{5/3}$	
$H_{\mathrm{US}}(0,1)$	$\frac{2}{9}X^{-2}$	if $X \geqslant \sqrt{4/3}$	$2X_+^2$
	$\frac{1}{2} - \frac{\sqrt{3}}{6}X$	if $0 < X < \sqrt{4/3}$	
	1	if $X \leqslant 0$	

Table 5.1: Formulas for p-variables and e-variables

First, all results holds true if the conditions $\mathbb{E}^Q[X] \leqslant \mu$ and $\mathrm{Var}^Q(X) \leqslant \sigma^2$ in each hypothesis is replaced by $\mathbb{E}^Q[X] = \mu$ and $\mathrm{Var}^Q(X) = \sigma^2$, respectively. Such modifications narrow the hypotheses and hence all validity statements hold. The precision statements can be checked with similar arguments to our proofs, and we omit them. Therefore, knowing $\mathrm{Var}^Q(X) = \sigma^2$ on top of $\mathrm{Var}^Q(X) \leqslant \sigma^2$, or $\mathbb{E}^Q[X] = \mu$ on top of $\mathbb{E}^Q[X] \leqslant \mu$, does not lead to more powerful one-sided p-variables or e-variables.

Second, admissibility of the proposed p-variables and e-variables needs future research. For e-variables, admissibility is not difficult to establish, but the picture is different for p-variables. By Lemma 5.3.1, there always exists a smallest p-variable. It remains unclear whether the p-variables we obtained in Theorems 5.3.1-5.3.4 are the smallest ones for the four hypotheses, respectively.

5.4 Testing the null hypotheses

We build tests based on e-values and p-values in Section 5.3. Section 5.4.1 describes the main methodology based on e-processes for the one-sided testing problem; Section 5.4.2 describes a few other methods using our results in Section 5.3; and Section 5.4.3 discusses the two-sided testing problem on the mean with given variance.

5.4.1 Constructing e-processes

Let $\mu \in \mathbb{R}$ and $\sigma > 0$. We consider the following hypotheses by keeping the same notation as in Section 5.3:

$$\begin{aligned}
H(\mu,\sigma) &= \left\{Q : \mathbb{E}^Q[X_i|\mathcal{F}_{i-1}] \leqslant \mu \text{ and } \mathrm{Var}^Q(X_i|\mathcal{F}_{i-1}) \leqslant \sigma^2 \text{ for } i \in [n]\right\};\\
H_{\mathrm{S}}(\mu,\sigma) &= \{Q \in H(\mu,\sigma) : X_i|\mathcal{F}_{i-1} \text{ is symmetrically distributed for } i \in [n]\};\\
H_{\mathrm{U}}(\mu,\sigma) &= \{Q \in H(\mu,\sigma) : X_i|\mathcal{F}_{i-1} \text{ is unimodally distributed for } i \in [n]\};\\
H_{\mathrm{US}}(\mu,\sigma) &= H_{\mathrm{U}}(\mu,\sigma) \cap H_{\mathrm{S}}(\mu,\sigma).
\end{aligned}$$

Recall that it is without loss of generality to consider μ and σ^2 as constants. We can also test the hypotheses where some data are symmetric or unimodal and some are not, because we will build e-values from each of them separately. For simplicity, we only list the above four representative cases. Using a similar formulation, the hypothesis in Waudby-Smith and Ramdas (2024) is

$$H_{\mathrm{WSR}}(\mu) = \{Q \in H(\mu,1) : X_i|\mathcal{F}_{i-1} \text{ is supported in } [0,1] \text{ almost surely for } i \in [n]\}.$$

In the above formulation, the choice of $\sigma = 1$ is simply to remove the variance constraint; see Remark 5.2.1.

There are several simple ways to use results in Section 5.3 to construct an e-variable or p-variable for the above hypotheses; some of these methods are more useful than the others. In general, we can compute an e-variable E_i or p-variable P_i based on X_i for $i \in [n]$ using Theorems 5.3.1-5.3.4, and then combine them.

Our main proposal is to use e-processes. An e-process $M = (M_t)_{t\in[n]}$ can be constructed using

$$M_t = \prod_{i=1}^{t}(1 - \lambda_i + \lambda_i E_i), \tag{5.5}$$

where λ_i is $\mathcal{F}_{i-1}$-measurable and takes values in $[0, 1)$. This idea is the main methodology behind game-theoretic statistics; see Shafer (2021), Shafer and Vovk (2019), and Waudby-Smith and Ramdas (2024, Proposition 3). It has been used by Waudby-Smith and Ramdas (2024) for testing the mean and Wang et al. (2022) for testing risk measures. To find good choices of $\lambda = (\lambda_i)_{i\in[n]}$ is a non-trivial task. We propose to specify λ in two different ways.

(a) **E-mixture method**: We first take several $\lambda_i = \lambda \in [0, 1)$, which is a constant for each $i \in [n]$, and then average the resulting e-processes from (5.5) over these choices to get an e-process. An uninformative choice of the values of λ may be some points in $[0, 0.2]$. We avoid choosing λ close to 1 because our e-value may take the value 0 with substantial probability, leading a small value of $\mathbb{E}^Q[\log(1 - \lambda + \lambda E)]$. This quantity measures the growth rate of an e-process; see Grünwald et al. (2023) and Waudby-Smith and Ramdas (2024). In our simulation and empirical studies, we average over $\lambda = 0.01 \times \{1, \dots, 20\}$.

(b) **E-GREE method**: In the GREE (growth-rate for empirical e-statistics) method of Wang et al. (2022) for λ_i, $i \in [n]$ in (5.5), λ_i is determined by solving the following optimization problem:

$$\lambda_i = \left(\operatorname*{arg\,max}_{\lambda\in[0,1)} \frac{1}{i-1} \sum_{j=1}^{i-1} \log(1 - \lambda + \lambda E_j) \right) \wedge \frac{1}{2}. \tag{5.6}$$

To simplify the maximization in (5.6), a fast and approximate solution can be obtained using Taylor expansion as in Waudby-Smith and Ramdas (2024). This leads to the following simple formula

$$\lambda_i = \left(\frac{\sum_{j=1}^{i-1}(E_j - 1)}{\sum_{j=1}^{i-1}(E_j - 1)^2} \right)_+ \wedge \frac{1}{2}. \tag{5.7}$$

We will use (5.7) for all e-GREE related calculations for the following results.

When the hypothesis to test is $H_{\mathrm{WRS}}(\mu)$, the e-GREE method reduces to the method of Waudby-Smith and Ramdas (2024); see Section 5.5.2. An optimization procedure related to (5.6) is studied by Kumon et al. (2011).

For either the e-GREE or the e-mixture method, we fix $\alpha \in (0,1)$ and reject the null hypothesis if the e-process M goes beyond $1/\alpha$, that is, when $M_t \geqslant 1/\alpha$ for the first time. If we are only interested in testing with a fixed finite number n data points, then this is equivalent to using $\max_{t\in[n]} M_t$. The type-I error control is guaranteed by Ville's inequality (Ville (1939)) as $\mathbb{P}(\max_{t\in[n]} M_t \geqslant 1/\alpha) \leqslant \alpha$ for any positive integer n.

The result below clarifies consistency of the e-GREE method in the most idealistic setting.

Proposition 5.4.1. *Suppose that data are iid and generated from an alternative probability Q. The e-GREE method has asymptotic power approaching 1 as $n \to \infty$, that is, $Q(\sup_{t\in[n]} M_t \geqslant 1/\alpha) \to 1$ for any $\alpha \in (0,1)$ if and only if $\mathbb{E}^Q[E_1] > 1$.*

Although Proposition 5.4.1 requires an iid assumption, this assumption is not needed for consistency in practical situations; an example is in Section 5.5.1.

5.4.2 Some other methods

Below we list some other methods, where we assume that n is finite. They generally do not work well as shown by the simulation studies, but nevertheless we list them as they follow from our results in Section 5.3, and they are presented only for a comparison.

(c) **P-Fisher method**: Construct a p-variable P using the Fisher combination (Fisher (1925))
$$P = 1 - \chi_{2n}(-2(\log P_1 + \cdots + \log P_n)),$$
where χ_{2n} is the cdf of a chi-square distribution with $2n$ degrees of freedom.

(d) **P-Simes method**: Construct a p-variable P using the Simes combination (Simes (1986))
$$P = \min_{i\in[n]} \frac{n}{i} P_{(i)},$$

where $P_{(i)}$ is i-th order statistic of $P_1, \ldots, P_n$ from the smallest to the largest.

Although in general p-Fisher and p-Simes require independence among p-variables, they are valid in our setting since our p-variables are conditionally valid, and they can be treated as iid when we work with combination of p-values.

The next two methods use all data directly, and requires independence among $X_1, \ldots, X_n$. A most natural statistic is the sample mean $T = \sum_{i=1}^n X_i/n$. Under $H(\mu, \sigma)$, T has at most mean μ and variance at most σ^2/n. Moreover, symmetry of T follows from symmetry of $X_1, \ldots, X_n$. Nevertheless, T is not necessarily unimodal even if $X_1, \ldots, X_n$ are unimodal, and hence unimodality of T cannot be used. The following e-variables and p-variables are constructed by directly applying Theorems 5.3.1-5.3.4.

(e) **E-batch method**: An e-variable for $H(\mu, \sigma)$ or $H_{\mathrm{U}}(\mu, \sigma)$ is

$$E_0 = n(T - \mu)_+^2/\sigma^2,$$

an e-variable for $H_{\mathrm{S}}(\mu, \sigma)$ or $H_{\mathrm{US}}(\mu, \sigma)$ is

$$E_{\mathrm{S}} = 2n(T - \mu)_+^2/\sigma^2.$$

(f) **P-batch method**: A p-variable for $H(\mu, \sigma)$ or $H_{\mathrm{U}}(\mu, \sigma)$ is

$$P_0 = (1 + E_0)^{-1},$$

a p-variable for $H_{\mathrm{S}}(\mu, \sigma)$ or $H_{\mathrm{US}}(\mu, \sigma)$ is

$$P_{\mathrm{S}} = \min\{(2E_0)^{-1}, P_0\}.$$

All methods described in this section have Type-I error control under the null hypothesis and with finite sample (with methods (e) and (f) additionally requiring independence) without requiring that the data are identically distributed.

5.4.3 Two-sided e-values testing the mean given variance

We briefly discuss the two-sided mean testing problem, where the main hypothesis $H(\mu^L, \mu^U, \sigma)$ to test is

$$\left\{Q : \mathbb{E}^Q[X_i|\mathcal{F}_{i-1}] \in [\mu^L, \mu^U] \text{ and } \mathrm{Var}^Q(X_i|\mathcal{F}_{i-1}) \leqslant \sigma^2 \text{ for } i \in [n]\right\},$$

where $\mu^L \leqslant \mu^U$ are constants. The case $\mu^L = \mu^U$ corresponds to testing whether the mean is equal to a precise value.

Our methodology can be easily adapted to test this hypothesis. First, we note that the e-variable E given by

$$E = \frac{(X - \mu^U)_+^2 + (X - \mu^L)_-^2}{\sigma^2}, \tag{5.8}$$

is a precise e-variable for $H(\mu^L, \mu^U, \sigma)$ formulated on a single observation X. To see this, it suffices to note that for $Q \in H(\mu^L, \mu^U, \sigma)$,

$$\begin{aligned}\mathbb{E}^Q[E] &= \mathbb{E}^Q\left[\frac{(X - \mu^U)_+^2 + (X - \mu^L)_-^2}{\sigma^2}\right] \\ &\leqslant \mathbb{E}^Q\left[\frac{(X - \mathbb{E}^Q[X])_+^2 + (X - \mathbb{E}^Q[X])_-^2}{\sigma^2}\right] = \frac{\mathrm{Var}^Q(X)}{\sigma^2} \leqslant 1.\end{aligned}$$

The statement on its precision can be verified similarly to Theorem 5.3.1.

If $\mu^L = \mu^U = \mu$, then the e-variable in (5.8) is

$$E = (X - \mu)^2/\sigma^2.$$

This e-variable satisfies the property that $\mathbb{E}^Q[E] > 1$ if $\mathbb{E}^Q[X] \neq \mu$ and $\mathrm{Var}^Q(X) = \sigma^2$; this condition is useful to establish consistency in Proposition 5.4.1.

Following the same procedure in Section 5.4.1 using (5.8), we obtain e-processes for the two-sided problem $H(\mu^L, \mu^U, \sigma)$. Due to a smaller null hypothesis, this e-process is generally more powerful than the one in Section 5.4.1 testing the one-sided mean.

There are special, adversarial scenarios where such two-sided tests may not be powerful. For instance, if data are independent with $\mathbb{E}[X_i] < \mu$ and $\mathbb{E}[X_j] > \mu$ appearing in an

alternating sequence; this forms a dataset that looks like iid data with mean μ, thus very difficult to detect. The same challenge exists for other methods based on e-processes, such as that of Waudby-Smith and Ramdas (2024).

Remark 5.4.1. Under the additional information of symmetry, the e-variable in (5.8) can be used, but it cannot be multiplied by two as in Theorem 5.3.2. In this case, an alternative way to take advantage of symmetry is to build two e-processes in Section 5.4.1: one to test $\mathbb{E}[X_i|\mathcal{F}_{i-1}] \leqslant \mu^U$ and another one to test $\mathbb{E}[-X_i|\mathcal{F}_{i-1}] \leqslant -\mu^L$. Taking the average of these two e-processes yields a valid e-process for the null hypothesis. As long as one of the two e-processes has good power for the true data generating procedure, the average e-process has good power.

5.4.4 Power of the e-values with fixed mean and growing variance

In this section, we analyze the power of the e-variables. For a given e-variable E, its e-power, using the terminology of Vovk and Wang (2024), for an alternative probability Q is defined as $\mathbb{E}^Q[\log E]$; see Shafer (2021) and Grünwald et al. (2023) for using this quantity as a notion of power. Certainly, the power depends on the specific alternative Q. We are particularly interested in how the e-power changes as the variance in the alternative hypothesis grows.

For this purpose, we consider a simplistic, yet representative setting, where a class of simple alternatives $(Q_\sigma)_{\sigma>1}$ is indexed by $\sigma > 1$, such that our data point X under Q_σ is distributed as σZ, where Z has a fixed distribution with mean 0 and variance 1 satisfying the null hypothesis, which can be one of $H(0,1)$, $H_{\mathrm{S}}(0,1)$, $H_{\mathrm{U}}(0,1)$ and $H_{\mathrm{US}}(0,1)$. Note that in this setting, the mean of the data is always 0, and only its variance grows under the alternative. We denote by Q_0 a null probability. Below, we will show that the e-power of each e-variable grows at a rate of $\log \sigma$ as the alternative variance σ^2 grows, regardless of the distribution of Z.

Let E be the e-variable computed based on X as in Section 5.3. Due to the construction of the e-process M in (5.5), the e-power of relevance is defined as

$$\Pi^{Q_\sigma} = \sup_{\lambda\in[0,1]} \mathbb{E}^{Q_\sigma}[\log(1-\lambda+\lambda E)] = \sup_{\lambda\in[0,1]} \mathbb{E}^{Q_0}[\log(1-\lambda+\lambda\sigma^2 E)],$$

that is, the best-achievable e-power of in each multiplicative term in the e-process M.

Proposition 5.4.2. *Suppose* $p := Q_0(E \geqslant 1) > 0$. *For* $\sigma > 1$,

$$(2p \log \sigma - \log 2)_+ \leqslant \Pi^{Q_\sigma} \leqslant 2 \log \sigma. \tag{5.9}$$

Moreover, $0 \leqslant \Pi^{Q_\sigma} - \Pi^{Q_\delta} \leqslant 2(\log \sigma - \log \delta)$ *for* $\sigma > \delta > 1$.

Proposition 5.4.2 suggests that the growth rate of the e-process M is roughly a constant times $\log \sigma$ when the alternative variance σ^2 is larger than 1. An additional negative term $-\log 2$ in (5.9) is not surprising, because our conditions do not guarantee $\Pi^{Q_\sigma} > 0$ for σ very close to 1. Below, we give an example to illustrate the sharpness of bounds in (5.9).

Example 5.4.1. Suppose that $Q_0(E = 0) = Q_0(E = 2) = 1/2$. We can compute

$$\Pi^{Q_\sigma} = \sup_{\lambda \in [0,1]} \frac{1}{2} \left(\log(1 - \lambda) + \log(1 + \lambda(2\sigma^2 - 1)) \right) = \frac{1}{2} \log \frac{\sigma^4}{2\sigma^2 - 1}.$$

It is clear that Π^{Q_σ} is approximately equivalent to $\log \sigma$ for large σ, corresponding to the left side of (5.9) with $p = 1/2$.

5.5 Simulation studies

5.5.1 A comparison of different e-combining methods

In this section, we conduct simulation studies for the non-parametric hypotheses in Section 5.4. We set $\mu = 0$ and $\sigma = 1$ without loss of generality.

We first concentrate on the null hypothesis $H(0, 1)$, as the other four cases are similar. For all the methods stated in Section 5.4, we do not make the assumption that the data are identically distributed. Thus, we generate a sample of n independent data points (although independence is not needed for methods (a)-(d)), alternating from two different distributions: $X_1, X_3, \ldots$, follow a normal distribution, and $X_2, X_4, \ldots$, follow a Laplace

distribution, with the same mean ν and the same variance η^2.[1] We denote this data generating process as $\mathrm{NL}(\nu, \eta^2)$ with the null parameters being $(\nu, \eta^2) = (0, 1)$. We consider two alternatives: (1) Data generated from $\mathrm{NL}(0, \eta^2)$ where $\eta > 1$; (2) Data generated from $\mathrm{NL}(\nu, 1)$ where $\nu > 0$. In our setting, the tester does not know the alternating data generating mechanism. For each alternative model, we compute the rejection rate over 1000 runs using the thresholds of $E \geqslant 1/\alpha$ and $P \leqslant \alpha$, with $\alpha = 0.05$, for e-values and p-values, respectively.

For the e-mixture method, we experiment by averaging λ in the interval $[0.01, 0.20]$ with step size 0.01. The e-GREE method is similar to the e-mixture method, except that λ_i is dynamically updated with different $i \in [n]$ using the formula (5.6).

Figure 5.1 shows the rejection rates for all methods with data generated from $\mathrm{NL}(0, \eta^2)$, where $\eta \in [1, 4]$ (left panel), and from $\mathrm{NL}(\nu, 1)$ where $\nu \in [0, 1]$ (right panel). For alternative model $\mathrm{NL}(0, \eta^2)$, we see that the e-mixture and the e-GREE methods outperform the other methods, with the e-mixture method being the most powerful. For $\eta < 1.5$, the rejection rates of all methods are very low, making it challenging to distinguish their efficiency. As $\eta > 1.5$, both the e-mixture method and the e-GREE method exhibit significantly higher rejection rates compared to other methods, demonstrating their effectiveness in testing $H(0, 1)$. The other four methods have almost no power. For alternative model $\mathrm{NL}(\nu, 1)$, we observe that e-batch method and the p-batch method show significant high rejection rates, since they are quite sensitive to the sample mean. Recall that these methods rely on independence, so the central limit theorem kicks in.

Among all methods, only the e-process based methods satisfy anytime-validity, that is, decision can be made at any stopping time when data arrive sequentially. This situation is common in financial applications, where realized losses accumulate over time; see the empirical study in Section 5.6.

The testing procedures for H_{S}, H_{U} and H_{US} are the same as for testing H. We generate 100 data points from $\mathrm{NL}(0.5, 2)$ and calculated the rejection rates for testing H_{S}, H_{U} and H_{US} with null hypotheses $\mu = 0$ and $\sigma = 1$. Table 5.2 displays the rejection rates for all

[1] The assumption that the two distributions have the same mean and variance is not necessary when evaluating the power of the methods. We assume this only for simplicity.

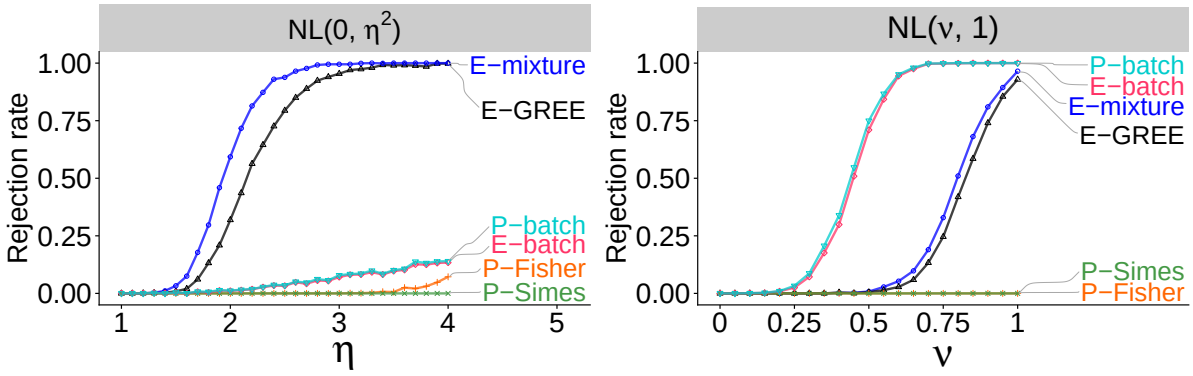


Figure 5.1: Rejection rates for all methods for testing $H(0,1)$ with sample size $n = 100$ over 1000 runs using the threshold 20.

hypotheses. It is clear that the extra information of symmetry improves the power.

	E-mixture	E-GREE	P-Fisher	P-Simes	E-batch	P-batch
H	0.419	0.315	0.000	0	0.639	0.664
H_{S}	0.998	0.882	0.000	0	0.900	0.900
H_{U}	0.419	0.315	0.006	0	0.639	0.664
H_{US}	0.998	0.882	0.763	0	0.900	0.900

Table 5.2: Rejection rates of testing H, H_{S}, H_{U} and H_{US} with $n = 100$ data generated from the model $\mathrm{NL}(0.5, 2)$.

5.5.2 A comparison with the GRAPA method

Recall that our model can also be interpreted as testing the mean under the knowledge of an upper bound on the variance. This allows us to compare our testing approach with the GRAPA (Growth Rate Adaptive to the Particular Alternative) method proposed by Waudby-Smith and Ramdas (2024). GRAPA is similar to the e-GREE method discussed in Section 5.4, but it requires the random variable to be bounded. The e-process $(M_t)_{t\in[n]}$

for the GRAPA method is constructed as follows:

$$M_t = \prod_{i=1}^{t}(1 + \lambda_i(X_i - \mu)), \tag{5.10}$$

where μ is the conditional mean being tested and λ_i is $\mathcal{F}_{i-1}$-measurable and takes value in $(-1/(1-\mu), 1/\mu)$, maximizing the growth of (5.10) in some sense similar to (5.6). It is clear that $1 + \lambda_i(X_i - \mu)$ is an e-variable for each $i \in [n]$. One way to obtain λ_i is to solve the following equation:

$$\frac{1}{i-1}\sum_{j=1}^{i-1}\frac{X_j - \mu}{1 + \lambda_j(X_j - \mu)} = 0. \tag{5.11}$$

Due to computational complexity, Waudby-Smith and Ramdas (2024) offered an alternative way to obtain λ_i by taking a Taylor expansion, which they call the approximate GRAPA method, so that the λ_i is computed by the following way

$$\lambda_i = -\frac{c}{1-\mu} \vee \frac{\widehat{\mu}_{i-1} - \mu}{\widehat{\sigma}^2_{i-1} + (\widehat{\mu}_{i-1} - \mu)^2} \wedge \frac{c}{\mu}, \tag{5.12}$$

where $\widehat{\mu}_i$ and $\widehat{\sigma}^2_i$ are empirical mean and variance of the observations $X_1 \ldots, X_i$, and $c \leqslant 1$ is fixed. From (5.12), it is clear that the GRAPA method is able to use the sample variance information adaptively; this is also true for (5.6) and (5.11) but implicitly. In this simulation, we use the approximate GRAPA method and choose $c = 1/2$.

We compare five methods for testing the mean under various conditions:

(a) GRAPA: The (approximate) GRAPA method with a bounded support $[0, 1]$.

(b) E-GREE: The e-GREE method with the variance upper bound σ^2.

(c) E-mixture: The e-mixture method with the variance upper bound σ^2.

(d) E-GREE-2s: The two-sided e-GREE method with the variance upper bound σ^2.

(e) E-mixture-2s: The two-sided e-mixture method with the variance upper bound σ^2.

The two-sided e-GREE and the two-sided e-mixture methods are abbreviated as e-GREE-2s and e-mixture-2s, respectively. We also note that GRAPA is also designed as a two-sided test, although it can easily be adjusted by restricting λ_i in (5.10) to be nonnegative.

Remark 5.5.1. We could also implement the e-GREE and e-mixture methods without an upper bounded variance but using the bounded support, as described in Remark 5.2.1. Although these methods are valid, they have poor power in our setting, because their assumption is strictly weaker than both bounded variance and bounded support. We omit their results.

We set $\mu = 0.2$ and apply both one-sided and two-sided tests on the same dataset. We generate a sample consisting of n independent data points from a beta distribution, denoted by Beta(ν, σ^2), where ν and σ^2 represent the mean and variance of the beta distribution.[2] Here, we use ν and σ^2 instead of the standard beta parameters α and β for the sake of convenience. Note that the parameters α and β can be easily recovered based on given mean ν and variance σ^2: $\alpha = \nu(\nu - \nu^2 - \sigma^2)/\sigma^2$ and $\beta = (\nu^2 + \sigma^2 - \nu)(\nu - 1)/\sigma^2$. Since the beta distribution has a bounded support $[0, 1]$, we can make meaningful comparisons between the GRAPA method and the e-GREE and e-mixture methods.

We first compare the rejection rates, using a threshold of 20 over 1000 runs, for all methods mentioned above under different ν with fixed σ^2. We consider $\nu \geqslant 0.35$ and $\sigma = 0.05$, $\sigma = 0.1$ and $\sigma = 0.3$. We use 20 data points for each run.

Figure 5.2 shows the performance of the three methods. First, the e-GREE method is always better than the e-mixture method. Second, the two-sided versions of both the e-GREE and e-mixture methods show a slight improvement over their respective one-sided methods, as expected. Third, in case $\sigma = 0.05$ and $\sigma = 0.1$, the e-GREE method outperforms the GRAPA method; in case $\sigma = 0.3$, the GRAPA method demonstrates superior performance compared to the other methods. This is intuitive, because the variance information is less useful for larger σ; recall that for any distribution supported in $[0, 1]$ with mean $\mu \leqslant 0.35$, the maximum possible variance is 0.2275 ($\sigma \approx 0.477$).

Figure 5.3 shows the average logarithmic e-processes for n up to 50 by using $\nu = \mu + \sigma$ for each alternative model. The relative rankings of these methods are consistent with their rejection rates, with e-GREE performing the best when σ is relatively small.

[2]None of the methods requires that the data follow identical distributions; we use a single distribution just for simplicity.

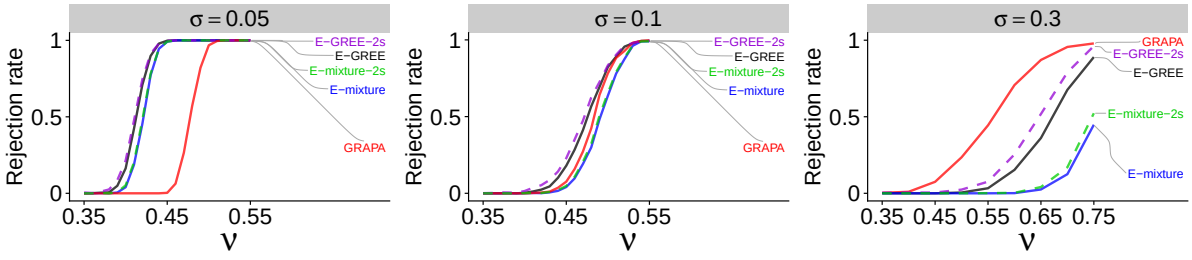


Figure 5.2: Rejection rates for the GRAPA, the e-GREE, the e-mixture and the two-sided e-GREE-2s and the e-mixture-2s methods over 1000 runs using the threshold 20 and $\mu = 0.35$. Data are generated from Beta(ν, σ^2) with sample size $n = 20$, where $\nu \geqslant 0.35$ and $\sigma \in \{0.05, 0.1, 0.3\}$.

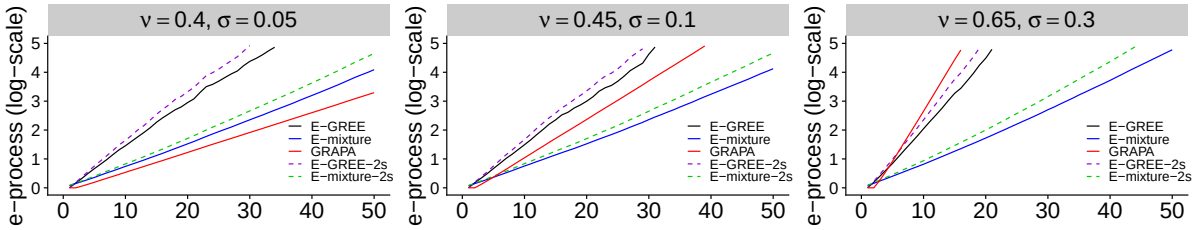


Figure 5.3: Average logarithmic e-processes for the GRAPA, the e-GREE, the e-mixture and the two-sided e-GREE-2s and the e-mixture-2s methods with varying sample size and $\mu = 0.35$. Data are generated from Beta(ν, σ^2) where $\sigma \in \{0.05, 0.1, 0.3\}$ and $\nu = \mu + \sigma$.

From the simulation results, our general recommendation is to use e-GREE to construct the e-process when the variance is relatively small, and to use GRAPA when the variance is relatively large compared to the bounded support.

5.5.3 A comparison with exponential test supermartingale

Next, we compare our methods with the exponential test supermartingale methods that directly construct e-processes, rather than using a betting strategy to combine sequential e-variables.

Wang and Ramdas (2023) extends the idea from Catoni (2012) to construct a non-negative test supermartingale called the Catoni supermartingale to test mean and variance in sequential settings. The test supermartingale is constructed as follows:

$$M_t^{\mathrm{C}} = \prod_{i=1}^{t} \exp\left(\phi(\lambda_i(X_i - \mu)) - \frac{\lambda_i^2 \sigma^2}{2}\right), \tag{5.13}$$

where ϕ is the influence function and $(\lambda_i)_{i\in[n]}$ is any predictable process. Following the recommendation of Wang and Ramdas (2023), we choose the influence function

$$\phi(x) = \begin{cases} \log(1 + x + x^2/2), & \text{if } x \geqslant 0; \\ -\log(1 - x + x^2/2), & \text{if } x < 0. \end{cases}$$

and $(\lambda_i)_{i\in[n]}$ as

$$\lambda_i = \left(\frac{2\log(1/\alpha)}{i(\sigma^2 + \eta_i^2)}\right)^{1/2} \quad \text{where} \quad \eta_i = \left(\frac{2\sigma^2 \log(1/\alpha)}{i - 2\log(1/\alpha)}\right)^{1/2}. \tag{5.14}$$

A different approach by Howard et al. (2021) is to use a framework for non-parametric confidence sequences based on the concept of exponential supermartingales. They introduce the concept of a "sub-ψ process" in Howard et al. (2021, Definition 1), where a pair of $\mathcal{F}_t$-adapted processes (S_t, V_t) is a so-called sub-ψ process. Here, S_t is the zero-mean deviation of the sample sum from its estimand at time t and V_t and ψ make the following process

$$M_t^{\psi} = \exp\{\lambda S_t - \psi(\lambda) V_t\} \tag{5.15}$$

a supermartingale for all λ in an interval $[0, \lambda_{\max})$. This framework allows for testing mean and variance under a wide variety of assumptions, including bounded supports, self-normalized bounds, and symmetric conditions. We refer to Howard et al. (2021, Appendix J, Table 3) for a collection of commonly used ψ functions and variance processes for $S_t = \sum_{i=1}^{t}(X_i - \mu)$ under various assumptions. We choose two special cases for comparison with our methods: the self-normalized bounds test supermartingale, denoted by $M_t^{\psi,\mathrm{SN}}$, and the symmetric condition test supermartingale, denoted by $M_t^{\psi,\mathrm{sym}}$. For $\lambda \in [0, \infty)$, these test supermartingales are constructed as follows:

$$M_t^{\psi,\mathrm{SN}} = \prod_{i=1}^{t} \exp\left(\lambda(X_i - \mu) - \frac{\lambda^2 (X_i - \mu)^2 + 2\sigma^2}{6}\right), \tag{5.16}$$

which also appears in Wang and Ramdas (2023, Section 5), and

$$M_t^{\psi,\mathrm{sym}} = \prod_{i=1}^{t} \exp\left(\lambda(X_i - \mu) - \frac{\lambda^2(X_i - \mu)^2}{2}\right). \tag{5.17}$$

We follow a simple method of choosing λ suggested by Howard et al. (2021, Section 3.2), that is, to use the mixture supermarginagle $\int \exp(\lambda S_t - \psi(\lambda)V_t)\,\mathrm{d}\Phi(\lambda)$ by assuming $\lambda \sim \Phi = \mathrm{N}(0,1)$. Now, we further compare the following methods:

(f) WR23-Catoni: The Catoni method with the variance upper bound σ^2.

(g) HRMS21-SN: The self-normalized method with the variance upper bound σ^2.

(h) HRMS21-sym: The sub-ψ method with symmetry, but without variance information.

(i) E-GREE-sym: The e-GREE method with the variance upper bound σ^2 and symmetry.

(j) E-mixture-sym: The e-mixture method with the variance upper bound σ^2 and symmetry.

We compare above five methods, along with the e-GREE and e-mixture methods that do not utilize symmetric information (methods (a) and (b) described in the previous section), in testing $H(0,1)$. Following the same data generating process as described in Section 5.5.2, we generate n independent data points alternating between the normal and Laplace distributions, denoted by $\mathrm{NL}(\nu,\eta^2)$. Figure 5.4 shows rejection rates for above methods with data generated from three cases: $\mathrm{NL}(\nu,1^2)$ for $\nu \in [0,1]$, $\mathrm{NL}(\nu,(1+\nu)^2)$ for $\nu \in [0,1]$, and $\mathrm{NL}(\nu/5,(1+\nu)^2)$ for $\nu \in [0,2]$.

For $\mathrm{NL}(\nu,1^2)$, the Catoni method outperforms other methods, while methods utilizing symmetric information generally perform well. For $\mathrm{NL}(\nu,(1+\nu))^2)$, where both the mean and variance of the data generating process change, the power of methods from Howard et al. (2021) reduces. In contrast, the power of our e-value based methods increases, as our construction of e-values is sensitive to the changes to variance. In the last case, $\mathrm{NL}(\nu/5,(1+\nu)^2)$, the impact of changes in mean is small and the variance effect is large,

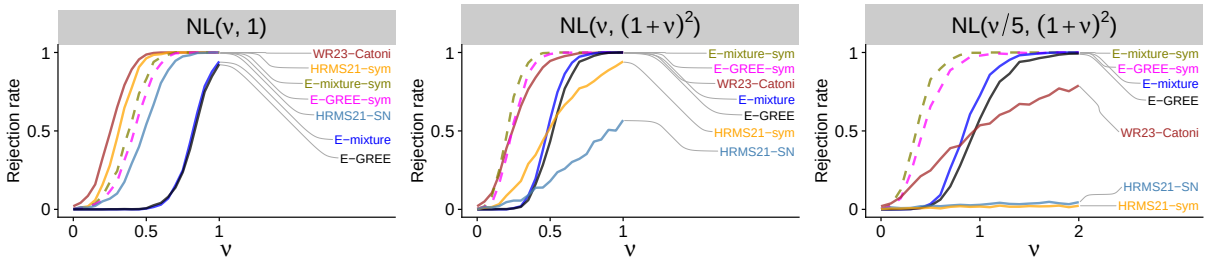


Figure 5.4: Rejection rates for methods (a), (b) and (f)-(j) for testing $H(0,1)$ with sample size $n = 100$ over 1000 runs using the threshold 20.

e-value based methods generally outperform others. Although method (h) benefits from not requiring information about variance or even the existence of variance, it demonstrates minimal power when testing mean with varying variance, due to its penalization term $-(X_i - \mu)^2$ in the exponential form of (5.16) and (5.17). In summary, our methods are comparatively more powerful when the alternative variance defers from the null.

5.6 Empirical study with financial data

In this section, we conduct an empirical study to test the hypothesis $H(\mu, \sigma)$ on the daily losses of financial assets. We aim to calculate the number of trading days required to detect evidence for rejecting the null hypothesis $H(\hat{\mu}, \hat{\sigma})$ during the 2007–2008 financial crisis period. Here, $\hat{\mu}$ and $\hat{\sigma}$ represent the sample mean and sample variance estimated from historical data prior to the testing period. That is, we are testing whether the historical estimations before the testing period are still valid. If the null hypothesis can be rejected at a reasonable thresholds level rather swiftly, this will serve as evidence of the effectiveness of e-process methods and could help investors switch strategies in a timely manner.

We choose 20 stocks from 10 different sectors of the S&P 500 list with the large market capitalization in each sector. Moreover, we include two companies with the largest market capitalization from the to-be Real Estate sector.[3] We first calculate the daily losses for

[3]Real Estate becomes the 11th sector of S&P500 in 2016.

each of the selected stocks from January 1, 2001 to December 31, 2010. The daily losses are expressed by percentage and calculate by $L_t = -(S_{t+1} - S_t)/S_t$, where S_t is the close price at day t. Note that the positive value represents a loss and negative value represents a gain. We could also use the log-loss data instead of the linear loss data, but the difference between the two is minor. We use the loss data from January 1, 2001 to December 31, 2006 to estimate the mean and variance for the null hypothesis. We compute the e-values using both the e-mixture method and the e-GREE method based on the construction of (5.5) as the daily loss from January 1, 2007 fed into the e-process.

Following a methodology similar to the simulation study in Section 5.5, we report the evidence against the null hypothesis when the e-process exceeds thresholds of 2, 5, 10, and 20.[4] E-values exceeding 5 or 10 provide substantial evidence to reject the null hypothesis, while a threshold of 20 offers strong evidence against the null hypothesis. It is important to note that, although a threshold of 2 may not be substantial enough to reject the null hypothesis, it can still serve as an early warning that the stock's performance may be different from its historical path.

To illustrate the e-process detection procedure, we first focus on a single stock as an example. Figure 5.5 reports the stock price for Simon Property (SPG) throughout the detection period and its corresponding e-process initiated on January 1, 2007. Observing from the e-process figure, it is evident that both the e-mixture method and the e-GREE method effectively reject the null hypothesis at thresholds of 2, 5, 10, and 20 before the financial crisis ends. Notably, the e-GREE method generally takes fewer trading days compared to the e-mixture method to achieve this rejection across various threshold levels. Also, the null hypothesis is rejected using e-GREE method prior to another significant decline in the stock price during February 2009 to June 2009, thus preventing potential larger losses and underscoring the effectiveness of e-process methods.

Compared to e-batch and other p-variable based methods stated in Section 5.4, e-

[4]In accordance with Jeffrey's rule of thumb about e-values (see Jeffreys (1961) and Vovk and Wang (2021)), if the e-value falls within the interval of $(10^{1/2}, 10)$, the evidence against the null hypothesis is considered substantial; If the e-value falls within the interval of $(10, 10^{3/2})$, the evidence against the null hypothesis is regarded as strong.

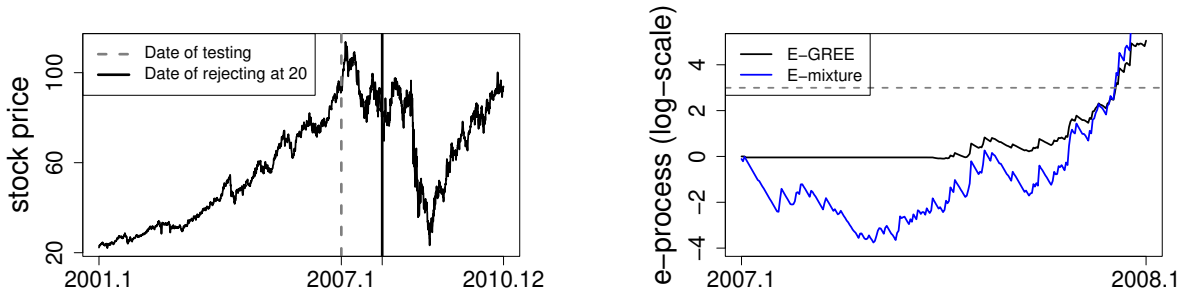


Figure 5.5: Sample path and logarithmic e-process using the e-GREE and the e-mixture methods testing of $H(\hat{\mu}, \hat{\sigma})$ for Simon Property (SPG) stock from January 2007 to January 2008, where $\hat{\mu} = -0.001028$ and $\hat{\sigma} = 0.012123$ are the sample mean and variance estimated from historical data for stock SPG from January 1, 2001 to December 31, 2006.

process based methods exhibit a unique advantage in sequential settings, particularly in financial applications where actual losses accumulate sequentially over time. In such scenarios, the e-process permits the early termination without a specified sampling period, potentially preventing further losses at an earlier stage. Table 5.3 displays the number of trading days required to reject the null hypothesis at various threshold levels for the selected 20 stocks from 10 different sectors and the two stocks in Real Estate. The table shows that stocks in sectors significantly impacted by the 2007–2008 subprime crisis, such as Financials and Consumer Discretionary, Energy (as evident in the top three stock price figures in Figure 5.6), could generally be detected using e-process based methods. In particular, the representative companies in Real Estate are rejected the earliest (see the last rows of Table 5.3). In contrast, for stocks in sectors less affected by the subprime crisis, such as Technology, Health Care, and Consumer Staples (as evident in the bottom three stock price figures in Figure 5.6), we are unable to reject the null hypothesis. This is intuitive, given that their prices and returns remain relatively stable or even increase during the financial crisis.

		E-GREE				E-mixture			
	Threshold	2	5	10	20	2	5	10	20
Financials	Bank of America	376	382	385	385	393	394	395	403
	Morgan Stanley	429	434	436	439	447	447	447	447
Utilities	The Southern	-	-	-	-	-	-	-	-
	Duke Energy	-	-	-	-	-	-	-	-
Communication Services	Verizon Comms.	-	-	-	-	-	-	-	-
	AT&T	-	-	-	-	-	-	-	-
Consumer Staples	Walmart	-	-	-	-	-	-	-	-
	PepsiCo	-	-	-	-	-	-	-	-
Consumer Discretionary	Ford Motor	457	476	483	491	546	594	594	594
	Las Vegas Sands	441	443	445	445	451	454	457	457
Energy	Texas Pacific Land	158	219	242	261	242	261	261	263
	Pioneer	486	515	539	548	-	-	-	-
Materials	Southern Copper	473	476	484	496	539	-	-	-
	Air Products	473	477	491	516	-	-	-	-
Health Care	Johnson & Johnson	-	-	-	-	-	-	-	-
	Pfizer	-	-	-	-	-	-	-	-
Technology	Int. Business Machines	-	-	-	-	-	-	-	-
	Microsoft	-	-	-	-	-	-	-	-
Industrials	General Electric	526	542	546	557	-	-	-	-
	United Parcel Service	457	476	488	491	542	604	-	-
Real Estate	Simon Property	165	223	238	250	223	239	250	253
	Prologis	264	270	271	271	270	271	271	275

Table 5.3: The number of trading days taken to detect evidence against $H(\hat{\mu}, \hat{\sigma})$ using the e-GREE method and the e-mixture method for different stocks from January 1, 2007 to December 31, 2010; "–" means no detection is observed till December 31, 2010.

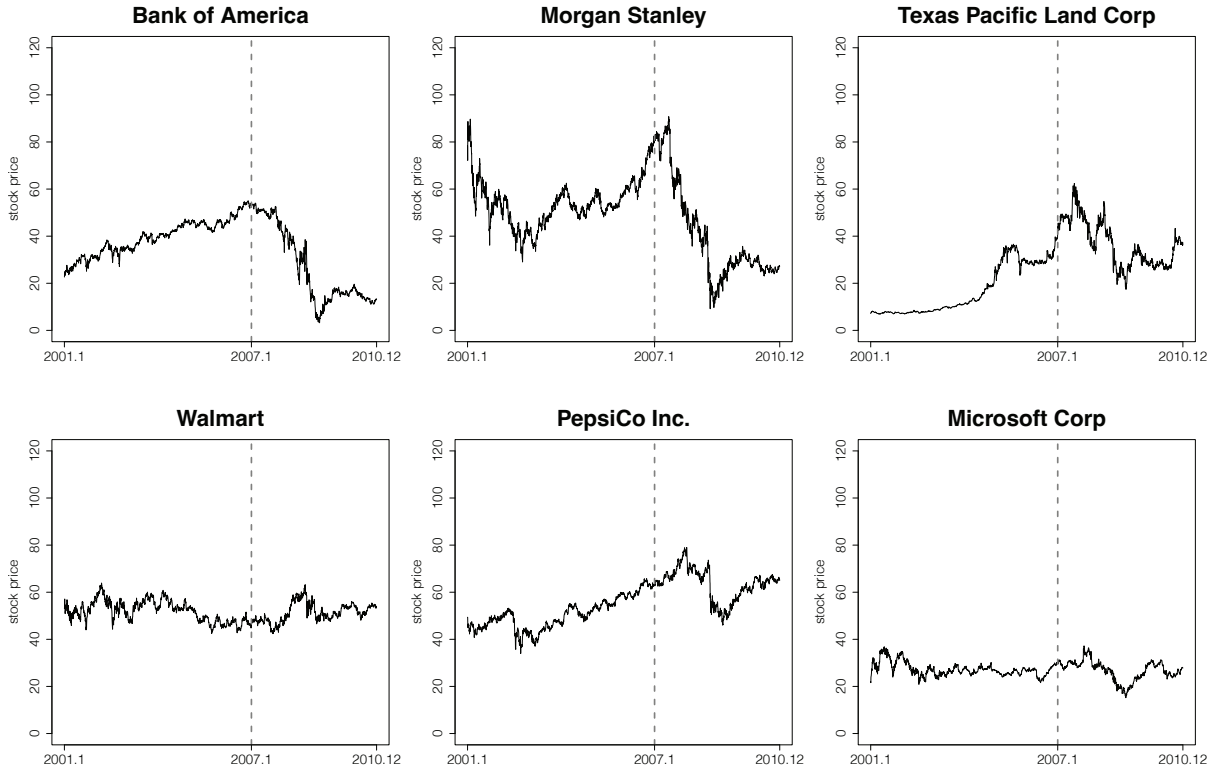


Figure 5.6: Sample paths for the stock price of Bank of America Corp. (BAC), Morgan Stanley (MS), Texas Pacific Land Corp. (TPL), Walmart Inc. (WMT), PepsiCo Inc. (PEP), Microsoft Corp. (MSFT) from January 1, 2001 to December 31, 2010.

5.7 Concluding remarks

This chapter proposes an e-process based approach for testing mean and variance from non-stationary data. We consider four classes of non-parametric composite hypotheses with specified mean and variance bound along with additional constraints of distribution, such as symmetry, unimodality, or a combination thereof. For this purpose, our main technical results give the best p-variables and e-variables in the simple setting where one summary data point is observed. The explicit formulas are summarized in Table 5.1. Using the obtained e-variables, we construct an e-process using either the e-mixture method or

the e-GREE method. Simulation studies and empirical analysis are conducted to show the performance of the proposed methods in comparison with GRAPA of Waudby-Smith and Ramdas (2024) and with the exponential supermartingale methods of Howard et al. (2020, 2021) and Wang and Ramdas (2023).

As mentioned in Section 5.3, our constructions of p-values and e-values are potentially useful for multiple testing, which is not addressed in this chapter. The literature on using e-values in multiple testing is growing recently. For instance, e-values are used for false discovery control in knockoffs; see Ren and Barber (2023) for derandomization, Ahn et al. (2023) for Bayesian linear models, and Gablenz and Sabatti (2023) for resolution-adaptive variable selection. Finally, the obtained e-variables may also be useful to build e-confidence regions (see Vovk and Wang (2023)) and e-posterior (see Grünwald (2023)) for (μ, σ^2), although we mainly consider a non-parametric setting.

5.8 Appendix: Technical details

We collect all proofs in the chapter in this section.

Proof of Lemma 5.3.1. Let $\mathcal{P}$ be any collection of p-variables for H. For $Q \in H$, using the fact that the elements of $\mathcal{P}$ are comonotonic, we have

$$Q(\inf\{P \in \mathcal{P}\} > \alpha) = Q\left(\bigcap_{P\in\mathcal{P}} \{P > \alpha\}\right) = \inf_{P\in\mathcal{P}} Q(P > \alpha) \geqslant 1 - \alpha.$$

This implies

$$Q(\inf\{P \in \mathcal{P}\} \leqslant \alpha) \leqslant \alpha.$$

Hence, the infimum of all p-variables for H is still a p-variable, which is the smallest one. □

For all theorems below, we will prove precision statements for the formulation of $\mathbb{E}^Q[X] = \mu$ instead of $\mathbb{E}^Q[X] \leqslant \mu$, making these statements stronger. For the validity statements, it is easy to verify that those p-variables and e-variables are valid under both formulations.

Proof of Theorem 5.3.1. Since the problem is invariant under location shift and scaling, it suffices to consider the normalized case of $(\mu, \sigma) = (0, 1)$.

It is clear that P is decreasing in X and E is increasing in X.

For $Q \in H(0,1)$, Cantelli's inequality implies $Q(X > x) \leqslant 1/(1+x^2)$ for $x > 0$, which implies, for each $\alpha \in (0,1)$,

$$Q(P \leqslant \alpha) = Q(1 + X_+^2 \geqslant 1/\alpha) = Q\left(X \geqslant \sqrt{(1-\alpha)/\alpha}\right) \leqslant \frac{1}{1 + 1/\alpha - 1} = \alpha.$$

The inequality above is an equality if Q is chosen such that

$$Q\left(X = \sqrt{(1-\alpha)/\alpha}\right) = \alpha = 1 - Q\left(X = \sqrt{\alpha/(1-\alpha)}\right), \tag{5.18}$$

and we can easily verify that $\mathbb{E}^Q[X] = 0$ and $\mathrm{Var}^Q(X) = 1$. This implies that $\sup_{Q \in H(0,1)} Q(P \leqslant \alpha) = \alpha$ for each $\alpha \in (0,1)$, and therefore $P = 1/(1+X_+^2)$ is a precise p-variable for $H(0,1)$.

For $Q \in H(0,1)$, we have $\mathbb{E}^Q[X_+^2] \leqslant \mathbb{E}^Q[X^2] \leqslant 1$. To show that E is precise, let Q be given by (5.18), which satisfies $\mathbb{E}^Q[X_+^2] = \alpha$. By taking $\alpha \uparrow 1$ we know $\sup_{Q \in H(0,1)} \mathbb{E}^Q[E] = 1$, and therefore $E = X_+^2$ is a precise e-variable for $H(0,1)$. □

Proof of Theorem 5.3.2. We first show the statement on the e-variable. Set $(\mu, \sigma) = (0,1)$ as in the proof of Theorem 5.3.1. For $Q \in H_{\mathrm{S}}(0,1)$, we have $2\mathbb{E}^Q[X_+^2] = \mathbb{E}^Q[X^2] \leqslant 1$, with equal sign holding if $\mathrm{Var}^Q(X) = 1$. Therefore, $E = 2X_+^2$ is a precise e-variable for $H_{\mathrm{S}}(0,1)$.

Since $E = 2X_+^2$ is an e-variable, by Markov's inequality, $1/E = (2X_+)^{-2}$ is a p-variable for $H_{\mathrm{S}}(0,1)$. In Theorem 5.3.1 we have seen that P_0 is a p-variable for $H(0,1)$, and hence also a p-variable for $H_{\mathrm{S}}(0,1) \subseteq H(0,1)$. Using Lemma 5.3.1, the minimum of P_0 and $(2E_0)^{-1}$ is a p-variable for $H_{\mathrm{S}}(0,1)$.

Next, we show that P is semi-precise. For $\alpha \in (0, 1/2]$, let Q be chosen such that

$$Q\left(X = (2\alpha)^{-1/2}\right) = \alpha = Q\left(X = (2\alpha)^{-1/2}\right) \text{ and } Q(X = 0) = 1 - 2\alpha.$$

We can verify that $\mathbb{E}^Q[X] = 0$, $\mathrm{Var}^Q(X) = 1$, and X is symmetrically distributed. It follows that $Q(P \leqslant \alpha) = Q(X = (2\alpha)^{-1/2}) = \alpha$. This implies that $\sup_{Q \in H_{\mathrm{S}}(0,1)} Q(P \leqslant \alpha) = \alpha$ for $\alpha \in (0, 1/2]$. Therefore, P is a semi-precise p-variable for $H_{\mathrm{S}}(0,1)$.

Finally, we show that there do not exist precise p-variables for $H_{\mathrm{S}}(0,1)$. Suppose that $P = f(X)$ is a precise p-variable, where f is a decreasing function. Note that $Q(X \leqslant 0) \geqslant 1/2$ for all $Q \in H_{\mathrm{S}}(0,1)$. It follows that $Q(P \geqslant f(0)) \geqslant 1/2$ and $Q(P < f(0)) \leqslant 1/2$. If $f(0) > 1/2$, then for $\alpha \in [1/2, f(0)]$, $Q(P \leqslant \alpha) \leqslant 1/2 < \alpha$, implying that P is not precise. If $f(0) \leqslant 1/2$, then, by taking Q as the point-mass at 0, we have $Q(P \leqslant 1/2) = 1$, implying that P is not a p-variable. Either way we have a contradiction, and hence does not exist a precise p-variable. □

Proof of Theorem 5.3.3. Set $(\mu, \sigma) = (0,1)$ as in the proof of Theorem 5.3.1. By Theorem 1 of Bernard et al. (2020),

$$\sup_{Q \in H_{\mathrm{U}}(0,1)} T_X^Q(1-\alpha) = \max\left\{\sqrt{\frac{4}{9\alpha} - 1}, \sqrt{\frac{3-3\alpha}{1+3\alpha}}\right\} \quad \text{for } \alpha \in (0,1). \tag{5.19}$$

Note that P is a decreasing function of X, and we denote this by $P = f(X)$ where

$$f(x) = \max\left\{\frac{4}{9}(1+x_+^2)^{-1}, \frac{4}{3}(1+x_+^2)^{-1} - \frac{1}{3}\right\}.$$

For $\alpha \in (0, 1/6]$, we have

$$\sup_{Q \in H_{\mathrm{U}}(0,1)} T_X^Q(1-\alpha) = \sqrt{\frac{4}{9\alpha} - 1},$$

and hence

$$\inf_{Q \in H_{\mathrm{U}}(0,1)} T_P^Q(\alpha) = f\left(\sup_{Q \in H_{\mathrm{U}}(0,1)} T_X^Q(1-\alpha)\right) = \frac{4}{9}\left(1 + \frac{4}{9\alpha} - 1\right)^{-1} = \alpha.$$

For $\alpha \in (1/6, 1)$, it is

$$\sup_{Q \in H_{\mathrm{U}}(0,1)} T_X^Q(1-\alpha) = \sqrt{\frac{3-3\alpha}{1+3\alpha}},$$

and hence

$$\inf_{Q \in H_{\mathrm{U}}(0,1)} T_P^Q(\alpha) = f\left(\sup_{Q \in H_{\mathrm{U}}(0,1)} T_X^Q(1-\alpha)\right) = \frac{4}{3}\left(1 + \frac{3-3\alpha}{1+3\alpha}\right)^{-1} - \frac{1}{3} = \alpha.$$

Using Lemma 5.3.2, we obtain that P is a precise p-variable for $H_{\mathrm{U}}(0,1)$.

As E is an e-variable for $H(0,1)$, it is also an e-variable for $H_{\mathrm{U}}(0,1)$. To show that it is precise, fix any $a \in (0,1)$, and let $p > 0$ and $b > 0$ satisfy

$$a^2 = \frac{3-3p}{3p+2-p^2} \quad \text{and} \quad b = \frac{1+p}{1-p}a.$$

Note that such p exists for any $a \in (0,1)$ since the range of $(3-3p)/(3p+2-p^2)$ covers $(0,1)$. Choose Q such that the distribution of X has a point-mass at $-a$ with probability p and a uniform density on $[-a,b]$. We can compute

$$\mathbb{E}^Q[X] = -ap + \frac{b-a}{2}(1-p) = -ap + ap = 0,$$

and

$$\mathbb{E}^Q[X^2] = a^2 p + \frac{a^2}{3}(1-p) + \frac{b^2}{3}(1-p) = \frac{a^2(3p+2-p^2)}{3(1-p)} = 1.$$

Therefore $Q \in H_{\mathrm{U}}(0,1)$. We also have

$$\mathbb{E}^Q[E] = \mathbb{E}^Q[X_+^2] = 1 - a^2 p - \frac{a^2}{3}(1-p) \geqslant 1 - a^2.$$

Since $a \in (0,1)$ is arbitrary, we get $\sup_{Q \in H_{\mathrm{U}}(0,1)} \mathbb{E}^Q[E] = 1$, and hence E is a precise e-variable. □

Proof of Lemma 5.3.3. For $\alpha \geqslant 1/2$, since $Q \in H_{\mathrm{US}}(0,1)$ is symmetric about 0, we have $T_X^Q(1-\alpha) \leqslant 0$, with $T_X^Q(1-\alpha) = 0$ if Q is the point-mass at 0. We assume $\alpha < 1/2$ below.

Take any $Q \in H_{\mathrm{US}}(0,1)$, and we will find another distribution R with smaller variance and the same α-quantile (we omit "left" because the quantile is unique for Q and R). Note that Q has a decreasing density on $(0,\infty)$ and possibly a point-mass at 0. Denote by $x_0 = T_X^Q(1-\alpha)$ and g the density function of Q on $(0,\infty)$. Consider a different distribution R symmetric with respect to 0 which has uniform density equal to $g(x_0)$ on $(0,b)$ for some $b > x_0$ and a point-mass at 0, such that $R([x_0,b)) = \alpha = Q([x_0,\infty)) = R([x_0,\infty))$. Denote by h the density function of R on $(0,\infty)$, and note that $h(x) = 0$ for $x > b$. Since Q has a

decreasing density g on $(0,\infty)$, $g \geqslant h$ on $(0,x_0)$ and $g \leqslant h$ on (x_0,b). The above conditions imply

$$\int_0^{x_0} x^2 g(x)\mathrm{d}x \geqslant \int_0^{x_0} x^2 h(x)\mathrm{d}x \quad \text{and} \quad \int_{x_0}^{\infty} x^2 g(x)\mathrm{d}x \geqslant \int_{x_0}^{\infty} x^2 h(x)\mathrm{d}x, \tag{5.20}$$

where the second inequality is due to $R([x_0,\infty)) = Q([x_0,\infty))$. Note that both inequalities in (5.20) are equalities if and only if $g = h$, and equivalently, $Q = R$. It follows that $\mathbb{E}^Q[X^2] \geqslant \mathbb{E}^R[X^2]$, and hence $R \in H_{\mathrm{US}}(0,1)$. Note that the condition $Q([x_0,\infty)) = \alpha = R([x_0,\infty))$ guarantees $T_X^Q(1-\alpha) = x_0 = T_X^R(1-\alpha)$; that is R has the same α-quantile as G.

The above argument shows that it suffices for us to consider distributions Q which can be represented by a mixture of point-mass at 0 and a uniform distribution on $[-b,b]$. We also assume that Q has variance 1; if the variance is less than 1, then a rescaled distribution from Q has variance 1 and a larger α-quantile. Let $p = Q((0,\infty)) \in (0,1/2]$. We can compute $\mathbb{E}^Q[X^2] = 2pb^2/3 = 1$, and hence $b = 3^{1/2}(2p)^{-1/2}$. This gives

$$T_X^Q(1-\alpha) = b(1-\alpha/p) = \sqrt{\frac{3}{p}}\left(1-\frac{\alpha}{p}\right).$$

Maximizing the above term over $p \in (0,1/2]$ gives $p = 3\alpha$ if $\alpha \leqslant 1/6$ and $p = 1/2$ if $\alpha \in (1/6,1/2]$, showing the desired supremum formula in the lemma. □

Proof of Theorem 5.3.4. Set $(\mu,\sigma) = (0,1)$ as in the proof of Theorem 5.3.1. By Theorem 5.3.3, $E = 2E_0$ is an e-variable for $H_{\mathrm{US}}(0,1)$. It is precise because $\mathbb{E}^Q[2X_+^2] = 1$ for any $Q \in H_{\mathrm{US}}(0,1)$ with $\mathrm{Var}^Q(X) = 1$.

The fact that precise p-variables do not exist for $H_{\mathrm{US}}(0,1)$ follows from the same argument as in the proof of the corresponding statement in Theorem 5.3.2.

It remains to show that P is a semi-precise p-variable for $H_{\mathrm{US}}(0,1)$. Write $P = f(X)$ where

$$f(x) = \frac{2}{9x^2}\mathbb{1}_{[4/3,\infty)}(x_+^2) + \frac{3-\sqrt{3}x}{6}\mathbb{1}_{(0,4/3)}(x_+^2) + \mathbb{1}_{(-\infty,0]}(x).$$

Using Lemma 5.3.3, for $\alpha \in (0,1/6]$, we have

$$\sup_{Q \in H_{\mathrm{US}}(0,1)} T_X^Q(1-\alpha) = \sqrt{\frac{2}{9\alpha}} \geqslant \sqrt{\frac{4}{3}}$$

and

$$\inf_{Q\in H_{\mathrm{US}}(0,1)} T_P^Q(\alpha) = f\left(\sup_{Q\in H_{\mathrm{US}}(0,1)} T_X^Q(1-\alpha)\right) = \frac{2}{9}\times\frac{9\alpha}{2} = \alpha.$$

Similarly, for $\alpha \in (1/6, 1/2)$, we have

$$\sup_{Q\in H_{\mathrm{US}}(0,1)} T_X^Q(1-\alpha) = \sqrt{3}(1-2\alpha) \in \left(0, \sqrt{4/3}\right)$$

and

$$\inf_{Q\in H_{\mathrm{US}}(0,1)} T_P^Q(\alpha) = f\left(\sup_{Q\in H_{\mathrm{US}}(0,1)} T_X^Q(1-\alpha)\right) = \frac{3-3(1-2\alpha)}{6} = \alpha.$$

Finally, for $\alpha \in [1/2, 1)$, we have $\inf_{Q\in H_{\mathrm{US}}(0,1)} T_P^Q(\alpha) = 1$ since $\mathbb{P}(X \leqslant 0) \geqslant 1/2$. Using Lemma 5.3.2, the above three cases together imply that P is a semi-precise p-variable for $H_{\mathrm{US}}(0,1)$. □

Proof of Proposition 5.4.1. The assumption that data are iid implies that $E_1, E_2, \dots$ are iid. The "only if" statement is trivial since $\mathbb{E}^Q[E_1] \leqslant 1$ implies that $(M_t)_{t\geqslant 1}$ is an e-process for Q, and hence $Q(\sup_{t\in[n]} M_t \geqslant 1/\alpha) \leqslant \alpha$ for all $n \in \mathbb{N}$. Next we show the "if" statement. For this, we use Theorem 3 of Wang et al. (2022), which states that, under the iid assumption,

$$\frac{1}{t}\left(\log M_T(\boldsymbol{\lambda}^{\mathrm{GREE}}) - \log M_t(\boldsymbol{\lambda}^{\mathrm{GRO}})\right) \xrightarrow{L^1(Q)} 0 \quad \text{as } t\to\infty,$$

where $M_t(\boldsymbol{\lambda}^{\mathrm{GREE}})$ is given by (5.5) with each λ_i computed form the e-GREE method, and $M_t(\boldsymbol{\lambda}^{\mathrm{GRO}})$ is given by (5.5) with each λ_i given by its theoretically growth-rate optimal value

$$\lambda^* = \operatorname*{arg\,max}_{\lambda\in(0,1]} \mathbb{E}^Q[\log(1-\lambda+\lambda E_1)],$$

and this gives

$$\frac{1}{t}\log M_t(\boldsymbol{\lambda}^{\mathrm{GRO}}) = \max_{\lambda\in(0,1]} \mathbb{E}^Q[\log(1-\lambda+\lambda E_1)].$$

Therefore, we have

$$\frac{1}{t}\log M_t \xrightarrow{Q} \max_{\lambda\in(0,1]} \mathbb{E}^Q[\log(1-\lambda+\lambda E_1)] \quad \text{as } t\to\infty.$$

It remains to verify $\max_{\lambda\in(0,1]} \mathbb{E}^Q[\log(1-\lambda+\lambda E_1)] > 1$. Note that $\mathbb{E}[E_1] > 1$ implies $\mathbb{E}[E_1\wedge K] > 1$ for some $K \geqslant 1$. We denote by $Y = E_1\wedge K$. Since $\mathbb{E}[(Y-1)_+]-\mathbb{E}[(Y-1)_-] = \mathbb{E}[Y-1] > 0$, there exists some $\varepsilon \in (0,1)$ such that

$$\frac{1}{1+\varepsilon}\mathbb{E}[(Y-1)_+] - \frac{1}{1-\varepsilon}\mathbb{E}[(Y-1)_-] > 0.$$

Note that $\log(1+x) \geqslant x/(1+\varepsilon)$ for $x \in [0,\varepsilon)$ and $\log(1+x) \geqslant x/(1-\varepsilon)$ for $x \in (-\varepsilon, 0)$, that is,

$$\log(1+x) \geqslant \frac{x_+}{1+\varepsilon} - \frac{x_-}{1-\varepsilon} \quad \text{for } x \in (-\varepsilon,\varepsilon).$$

Hence, for $\lambda \in (0, \varepsilon/K)$, implying $\lambda(Y-1) \in (-\varepsilon,\varepsilon)$, we have

$$\begin{aligned}\mathbb{E}[\log(1-\lambda+\lambda E_1)] &\geqslant \mathbb{E}[\log(1+\lambda(Y-1))] \\ &\geqslant \frac{1}{1+\varepsilon}\mathbb{E}[\lambda(Y-1)_+] - \frac{1}{1-\varepsilon}\mathbb{E}[\lambda(Y-1)_-] > 0,\end{aligned}$$

thus showing the desired inequality. □

Proof of Proposition 5.4.2. First, it is clear that $\Pi^{Q_\sigma} \geqslant 0$ by choosing $\lambda = 0$ in the supremum. Second, by Jensen's inequality, for $\sigma > 1$,

$$\mathbb{E}^{Q_0}[\log(1-\lambda+\lambda\sigma^2 E)] \leqslant \log(1-\lambda+\lambda\sigma^2\mathbb{E}^{Q_0}[E]) \leqslant \log(1-\lambda+\lambda\sigma^2) = 2\log\sigma.$$

We next show $\Pi^{Q_\sigma} \geqslant 2p\log\sigma - \log 2$. Note that

$$\mathbb{E}^{Q_0}[\log(1-\lambda+\lambda\sigma^2 E)] \geqslant (1-p)\log(1-\lambda) + p\log(1-\lambda+\lambda\sigma^2).$$

Maximizing the above term over $\lambda \in [0,1]$, the maximizer is $\lambda^* = (p\sigma^2-1)/(\sigma^2-1)$. The corresponding maximum value satisfies

$$\begin{aligned}(1-p)\log\frac{(1-p)\sigma^2}{\sigma^2-1} + p\log(p\sigma^2) &\geqslant (1-p)\log(1-p) + p\log p + p\log\sigma^2 \\ &\geqslant -\log 2 + p\log\sigma^2,\end{aligned}$$

where we used the fact that $x \log x + (1-x)\log(1-x)$ on $[0,1]$ is maximized at $x = 1/2$. This shows $\Pi^{Q_\sigma} \geqslant 2p \log \sigma - \log 2$, completing the proof of (5.9).

Finally, we prove the last statement $0 \leqslant \Pi^{Q_\sigma} - \Pi^{Q_\delta} \leqslant 2(\log \sigma - \log \delta)$ for $\sigma > \delta > 1$. For any $\lambda \in [0,1]$, let $\lambda' = \lambda\delta^2/\sigma^2 \in [0,1]$. We have

$$\Pi^{Q_\sigma} \geqslant \log(1 - \lambda' + \lambda'\sigma^2 E) \geqslant \log(1 - \lambda + \lambda\delta^2 E).$$

Taking a supremum over $\lambda \in [0,1]$ yields $\Pi^{Q_\sigma} \geqslant \Pi^{Q_\delta}$. To show the other inequality,

$$\begin{aligned}\Pi^{Q_\sigma} &\leqslant \sup_{\lambda\in[0,1]} \log\left(\frac{\sigma^2}{\delta^2}(1-\lambda) + \lambda\sigma^2 E\right) \\ &= \log\frac{\sigma^2}{\delta^2} + \sup_{\lambda\in[0,1]} \log(1 - \lambda + \lambda\delta^2 E) = 2\log\frac{\sigma}{\delta} + \Pi^{Q_\delta}.\end{aligned}$$

This gives $\Pi^{Q_\sigma} - \Pi^{Q_\delta} \leqslant 2(\log \sigma - \log \delta)$ and completes the proof. □

Chapter 6

Conclusions and future research

6.1 Conclusion

This thesis first studies risk optimization under distributional model uncertainty, focusing on three main types of uncertainty sets: moment-based, distance-based and shape-preserving uncertainty sets. Other than risk optimization problem, we also study an innovative risk sharing problem, where we establish a new risk sharing framework that does not require predefined risk preferences. In addition, this thesis also studies the hypothesis testing problem in the context of quantitative risk management, particularly focusing on testing mean and variance using the new concept of e-values.

In Chapter 2, we develop the reverse ES optimization formula compared to the well-known ES optimization formula in Rockafellar and Uryasev (2002). The reverse ES optimization formula reveals that a mean excess function at any fixed threshold is the maximum of an ES curve minus a linear function, which may appear simple to risk experts. Such reverse formula proves to be useful in calculating the worst-case mean excess loss under the moment-based uncertainty set and Wasserstein uncertainty set.

In Chapter 3, we derive the explicit and closed-form expressions for the worst-case target semi-variance when only the mean and variance of a loss are known and the loss

is symmetric or non-negative. The closed-form results are extended to multivariate cases with applications in robust portfolio selection.

Chapter 4 proposes a concept of anonymized risk sharing. In contrast to other traditional risk sharing rules, it requires no information on preference, identity, irrelevant operations and actual loss in the risk sharing pool. To depict the key features of anonymity in risk sharing, we thus propose four economic axioms: risk anonymity, risk fairness, actuarial fairness and operational anonymity. We establish the surprising fact that these four very basic axioms uniquely characterize the conditional mean risk sharing; no other risk sharing rules satisfy these properties. We show that the anonymized risk sharing is compatible with the concept of decentralized systems, serving as a theoretical support to the wide applications of the conditional mean risk sharing as a standard tool in many relevant applications in decentralized risk sharing, including Bitcoin mining, tontines, P2P insurance.

Chapter 5 introduces an e-test approach for testing mean and variance in a non-parametric setting. Starting from a simple scenario where only one data point is observed, we develop p-variables and e-variables for null hypotheses that specify mean and variance bound along with additional constraints of distribution, such as symmetry, unimodality, or a combination thereof. We then establish e-processes by using e-combining methods that integrate these e-variables derived from single data points. Simulation results demonstrate that e-test methods are markedly superior to p-test methods. Furthermore, e-values are particularly compatible with sequential testing common in financial applications, where realized losses accumulate over time.

In the following sections, some future research and open questions that emerge from the findings of this thesis will be discussed. It aims to provide a roadmap for extending the methodologies and insights developed here to the quantitative risk management field.

6.2 Future research and open questions

6.2.1 Several distributionally robust reinsurance problems

In Chapter 3, our discussion mainly centers on the worst-case target semi-variance within the framework of moment-based and shape-preserving uncertainty sets. The closed-form formulas are applied to robust portfolio selection problems. However, numerous intriguing problems remain, particularly in the application of those moment-based uncertainty sets, as introduced in Section 3.2 of Chapter 3, to the insurance and reinsurance applications.

Several literature explores the robust insurance applications with distance-based uncertainty sets (mainly Wasserstein distance introduced in Section 2.4.2 of Chapter 2). We refer to Birghila and Pflug (2019), Boonen and Jiang (2024) and Cai et al. (2024) for the robust general insurance/reinsurance applications, and Bernard et al. (2024) for the robust life insurance. In addition, Liu and Mao (2022) investigates distributionally robust reinsurance under the moment-based uncertainty set, but restricts risk measures to VaR and ES, and limits the insurance indemnity functions to stop-loss functions only. Thus, a natural extension arises as follows:

Question 1. Can we extend the risk measures ρ to more generalized cases, for example distortion risk measure, or set the general form of indemnity functions $I(X)$ (for example, increasing convex indemnity functions) instead of limiting it to be stop-loss only:

$$\min_{I\in\mathcal{I}} \sup_{F\in\mathcal{L}} \rho^F(X - I(X) + \pi(I(X))),$$

where $\pi(X)$ indicates the premium principle.

The problem outlined above can be challenging to solve under very general conditions. One way to simplify is to choose a specific indemnity function. A direct connection to the worst-case target semi-variance with a moment-based uncertainty set (discussed in the Chapter 3), but in the insurance context, is to consider the target semi-variance premium principle

$$\pi(X) = \mathbb{E}[X] + \theta\mathbb{E}[(X-t)_+^2].$$

Question 2. We link the results in Chapter 3 to address a more specific robust reinsurance problem:

$$\min_{d\geqslant 0} \sup_{F\in\mathcal{L}} \rho^F(X\wedge t+\mathbb{E}[(X-t)_+]+\theta\mathbb{E}[(X-t)_+^2]),$$

where moment-based with shape preserving uncertainty set $\mathcal{L}$ is defined in Introduction of Chapter 3.

The above problem is considered from the perspective of insurer in a two-party reinsurance contract. It would be interesting to explore this robust reinsurance problem from the reinsurer's point of view.

Question 3. From the perspective of the reinsurer, we aim to solve the following optimization problem:

$$\min_{I\in\mathcal{I}} \sup_{F\in\mathcal{L}} \rho^F(I(X)-\pi(I(X))).$$

In more specific and common reinsurance settings, with the stop-loss contract $I(X) = (X-t)_+$ and expected premium principle $\pi(X) = (1+\theta)\mathbb{E}[X]$ we aim to solve

$$\min_{d\geqslant 0} \sup_{F\in\mathcal{L}} \rho^F((X-d)_+-(1+\theta)\mathbb{E}[(X-t)_+]).$$

Questions 1, 2 and 3 are all natural and intuitive extensions of current insurance framework under the moment-based uncertainty sets introduced in Chapter 2 and 3. Furthermore, more sophisticated reinsurance problems involving multiple risks or parties, as well as negotiable premiums as decision variables (see Cai and Chi (2020) and the references therein), can also be studied within a distributionally robust insurance context.

6.2.2 Standard and comparative e-backtests based on elicitability

There are many unresolved questions concerning the application of e-values in quantitative risk management, particularly in the backtesting of risk measures like VaR, ES, and expectile. Inspired by the work in Wang et al. (2022), which introduces a model-free traditional backtesting procedure for VaR and ES using e-values and e-processes, it is natural

to extend the traditional/standard backtests to the comparative backtests (proposed by Nolde and Ziegel (2017)) that evaluate different forecasting methods for these risk measures based on e-values and e-processes introduced in the Chapter 5.

Standard backtests are crucial for detecting risk underestimation and model misspecification, which are key concerns in financial regulation. Essentially, they assess whether the risk predictions made by financial institutions deviate significantly from a correct or reference model, typically unknown in practical scenarios. However, when a financial institution employs multiple forecasting methods, standard backtests fall short in helping regulators or risk managers select the most accurate method among the alternatives. To address this gap, we can use the idea of comparative backtests. This innovative approach allows for a direct comparison between different forecasting methods, providing a more detailed analysis that can identify the most reliable method among various alternatives. More importantly, the comparative e-backtest can also be conducted in a model-free manner, unlike the original comparative backtests described in Nolde and Ziegel (2017), which are based on p-values and highly dependent on the model.

As introduced by Nolde and Ziegel (2017), the evaluation for comparing risk forecasts can be formulated by the so-called S-dominance defined below. Given a fixed finite time horizon $T \in \mathbb{N}$ with $[T] = \{1, \dots T\}$, and model space $\mathcal{M}$ as a set of distributions on $\mathbb{R}$, we have $X_1, X_2, \dots, X_T$ to be losses arriving sequentially up to time T. The risk measure $\psi : \mathcal{M} \to 2^{\mathbb{R}^d}$ is predicted by the internal process $\{R_i\}_{i\in[T]}$ and the alternative process $\{R_i^*\}_{i\in[T]}$ that are both predictable, where $R_i, R_i^* : \Omega \to \mathbb{R}^d$ for all $i \in [T]$. Instead of an unknown correct process, we regard the alternative process $\{R_i^*\}_{i\in[T]}$ as the standard or reference process. A comparative backtest should give a conclusion whether the internal process $\{R_i\}_{i\in[T]}$ passes or not compared with the reference $\{R_i^*\}_{i\in[T]}$.

Definition 6.2.1. Let $\psi : \mathcal{M} \to \mathbb{R}^d$ be elicitable with scoring function $S : \mathbb{R}^{d+1} \to \mathbb{R}$. We say $\{R_i\}_{i\in\mathbb{N}}$ *S-dominates* $\{R_i^*\}_{i\in\mathbb{N}}$ if $\mathbb{E}[S(X_i, R_i) - S(X_i, R_i^*)] \leqslant 0$ for all $i \in [T]$. We say $\{R_i\}_{i\in\mathbb{N}}$ *conditionally S-dominates* $\{R_i^*\}_{i\in\mathbb{N}}$ if $\mathbb{E}[S(X_i, R_i) - S(X_i, R_i^*)|\mathcal{F}_{i-1}] \leqslant 0$ for all $i \in [T]$.

We consider the following null hypotheses for our comparative backtests based on con-

ditional S-dominance:

$$H^-(\psi) : \{R_i\}_{i\in[T]} \text{ conditionally } S\text{-dominates } \{R_i^*\}_{i\in[T]}. \tag{6.1}$$

$$H^+(\psi) : \{R_i^*\}_{i\in[T]} \text{ conditionally } S\text{-dominates } \{R_i\}_{i\in[T]}. \tag{6.2}$$

In practice, merely considering one of the two tests above does not always give a safe conclusion. For instance, if hypothesis $H^-(\psi)$ is not rejected, we are not able to say predictions $\{R_i\}_{i\in[T]}$ pass the comparative backtest without further justification. Instead, another test should be conducted for hypothesis $H^+(\psi)$.

Question 4. Based on the Definition 6.2.1, how do we properly construct e-variables and e-processes for testing the null hypotheses in (6.1) and (6.2) with the comparative backtests between internal model $\{R_i\}_{i\in[T]}$ and reference model $\{R_i^*\}_{i\in[T]}$?

Once appropriate e-processes are established for comparative backtesting, several intriguing challenges remain. The e-variables and e-processes are constructed via scoring function of its corresponding risk measures. We know that the scoring function for a specific risk measure may not be unique, and different scoring functions may vary in their discrimination ability, leading to different e-values when using different scoring functions. This variability leads to the following critical question:

Question 5. How do we appropriately select scoring functions for comparative backtests in our settings, especially considering that different scoring functions can significantly vary in their discrimination abilities?

Furthermore, the derivation of these e-processes opens up a broad spectrum of potential simulation studies and empirical analyses. These could include time series data analysis and real financial data analysis.

6.2.3 Elicitability and e-test of Gini indices

Backtesting risk measures heavily relies on the property of elicibility, which refers to the ability of a risk measure to be accurately predicted or forecasted in a consistent valid way,

see details in He et al. (2022). VaR and expectiles are examples of elicitable risk measures, which means they can be directly assessed through backtesting because there exists a loss function for which these measures are the minimizers of expected loss. While variance and ES are not individually elicitable, they can still be effectively backtested when combined with other metrics. Variance, for example, is jointly elicitable with the expectation, and ES is jointly elicitable with VaR, refer to Gneiting (2011) and Fissler and Ziegel (2016). This joint elicibility enables effective assessment in both traditional and e-process-based backtesting frameworks.

Now, we are interested in to backtesting the Gini indices, specifically the Gini deviation (GD) and Gini coefficient (GC). These indices are crucial measures of dispersion, deviation, and inequality and are widely applied in economics and finance. Fix a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. For $q \in [1, \infty)$, let L^q be the set of random variables with a finite q-th moment, and $\mathcal{M}^q$ be the distributions of elements of L^q. The Gini deviation $\mathrm{GD} : L^1 \to \mathbb{R}$ is defined as

$$\mathrm{GD}(Y) = \frac{1}{2}\mathbb{E}[|Y - Y'|],$$

where Y' is an iid copy of Y. Alternatively, we can write GD as a function from $\mathcal{M}^1 \to \mathbb{R}$ as a signed Choquet integral

$$\mathrm{GD}(F) = \int_0^1 F^{-1}(t)(2t-1)\mathrm{d}t = \int_{\mathbb{R}} F(t)(1-F(t))\mathrm{d}t,$$

see e.g., Wang et al. (2020, Example 1).

Let $L^q_+ = \{Y \in L^q : Y \geqslant 0,\ Y \neq 0\}$ and $\mathcal{M}^q_+$ be the distributions of elements of L^q_+. The Gini coefficient $\mathrm{GC} : L^1_+ \to [0, 1]$ is defined as

$$\mathrm{GC}(Y) = \frac{\mathrm{GD}(Y)}{\mathbb{E}[Y]} = \frac{\int_0^\infty F(t)(1-F(t))\mathrm{d}t}{\int_0^\infty (1-F(t))\mathrm{d}t}.$$

It is known that no deviation measures are elicitable, as stated in Wang and Wei (2020, Proposition 2.4). Thus, none of GD or GC is elicitable. Thus, following the idea of joint elicitability, the subsequent question arises naturally:

Question 6. Is it possible to identify other functionals that, when paired with GD or GC, render them jointly elicitable? If not, can we demonstrate, GD or GC are not jointly elicitable with other functionals, using the terminology of Fissler and Ziegel (2016), that GD and GC are not k-th order elicitable for any $k \in \mathbb{N}$?

If it proves challenging or even impossible to identify auxiliary statistics that render the Gini deviation (GD) or Gini coefficient (GC) elicitable, we might consider an alternative approach to handle these non-elicitable risk measures. The concept of multi-observation elicitability, proposed by Casalaina-Martin et al. (2017) and further discussed in Frongillo et al. (2019), provides a viable pathway by allowing the scoring function to depend on multiple data points simultaneously. Below, we give the formal definition:

Definition 6.2.2. For $k \in \mathbb{N}$, a mapping $\rho : \mathcal{M} \rightrightarrows \mathbb{R}^k$ is called d-observation $\mathcal{M}$-*elicitable* if there exists a function $S : \mathbb{R}^{k+d} \to \mathbb{R}$ such that

$$\rho(F) = \operatorname*{arg\,min}_{\mathbf{x} \in \mathbb{R}^k} \int_{\mathbb{R}^d} S(\mathbf{x}, \mathbf{y}) \mathrm{d}F^d(\mathbf{y}), \quad F \in \mathcal{M},$$

where $F^d(y_1, \ldots, y_d) = \prod_{j=1}^d F(y_j)$.

This naturally leads us to the question as follows:

Question 7. Can we identify a scoring function for GD or GC that makes them elicitable under the framework in Definition 6.2.2?

Our preliminary analysis indicates that it is feasible to identify a multi-observation scoring function, specifically a 2-observation function, for the GD and GC. This discovery allows us to further develop model-free e-statistics for standard backtesting procedures for these risk measures.

However, it seems that a major problem is that we need independent and identically distributed multiple observations. For example, for risk measures that are 2-observation elicitable, two iid data points are required simultaneously. This makes its application much more restrictive. Suppose that we observe two series of iid data. The following two questions arise:

Question 8. What is a natural way of building an e-process based on the above model-free e-statistic in this setting?

Question 9. How does the testing method based the e-process compare to other testing methods in the literature?

Indeed, multi-observation elicibility necessitates the use of multiple data points simultaneously. However, in the most natural application domain, such as backtesting financial risk forecasts, it seems that assuming iid data is not realistic. In the future research, it is also important to find some good applications to justify multi-observation settings for GD and GC.

References

Abdikerimova, S. and Feng, R. (2022). Peer-to-Peer multi-risk insurance and mutual aid. *European Journal of Operational Research*, **299**(2), 735–749.

Ahn, T., Lin, L. and Mei, S. (2023). Near-optimal multiple testing in Bayesian linear models with finite-sample FDR control. *arXiv*: `2211.02778`.

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**(7–8), 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.

Aliprantis, C. D., and Burkinshaw, O. (1990). *Principles of Real Analysis*. Gulf Professional Publishing.

Aliprantis, C. D. and Burkinshaw, O. (2006). *Positive Operators*. Springer Dordrecht.

Arrow, K. J. (1963). Uncertainty and the welfare economics of medical care. *American Economic Review*, **53**(5), 941–973.

Arrow, K. J. and Debreu, G. (1954). Existence of an equilibrium for a competitive economy. *Econometrica*, **22**(3), 265–290.

Artzner, P., Delbaen, F., Eber, J.-M. and Heath, D. (1999). Coherent measures of risk. *Mathematical Finance*, **9**(3), 203–228.

Asimit, V., Bignozzi, V., Cheung, K., Hu, J. and Kim, E. (2017). Robust and Pareto optimality of insurance contracts. *European Journal of Operational Research*, **262**(2), 720-732.

Barrieu, P. and El Karoui, N. (2005). Inf-convolution of risk measures and optimal risk transfer. *Finance and Stochastics*, **9**(2), 269–298.

Bawa, V. S. (1975). Optimal rules for ordering uncertain prospects. *Journal of Financial Economics*, **2**(1), 95–121.

BCBS (2019). Basel Committee on Banking Supervision. *Minimum Capital Requirements for Market Risk*. Bank for International Settlements. https://www.bis.org/bcbs/publ/d457.pdf

Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society Series B*, **57**(1), 289–300.

Benjamini, Y. and Yekutieli, D. (2001). The control of the false discovery rate in multiple testing under dependency. *Annals of Statistics*, **29**(4), 1165–1188.

Ben-Tal, A. and Nemirovski, A. (1998). Robust convex optimization. *Mathematics of Operations Research*, **23**(4), 769–805.

Ben-Tal, A. and Teboulle, M. (2007). An old-new concept of convex risk measures: The optimized certainty equivalent. *Mathematical Finance*, **17**(3), 449–476.

Bernard, C., De Vecchi, C. and Vanduffel, S. (2024). Robust assessment of life insurance products. *Annals of Operations Research*, published online, https://doi.org/10.1007/s10479-024-05969-8.

Bernard, C., Kazzi, R. and Vanduffel, S. (2020). Range value-at-risk bounds for unimodal distributions under partial information. *Insurance: Mathematics and Economics*, **94**, 9–24.

Bernard, C., Pesenti, S. M. and Vanduffel, S. (2023). Robust distortion risk measures. *Mathematical Finance*, published online, https://doi.org/10.1111/mafi.12414.

Bertsimas, D. and Popescu, I. (2002). On the relation between option and stock prices: a convex optimization approach. *Operations Research*, **50**(2), 358–374.

Buckley, I., Saunders, D. and Seco, L. (2008). Portfolio optimization when asset returns have the Gaussian mixture distribution. *European Journal of Operational Research*, **185**(3): 1434–1461.

Birghila, C. and Pflug, G. C. (2019). Optimal XL-insurance under Wasserstein-type ambiguity. *Insurance: Mathematics and Economics*, **88**, 30–43.

Blanchet, J., Chen, L. and Zhou, X. Y. (2021). Distributionally robust mean-variance portfolio selection with Wasserstein distances. *Management Science*, **17**(3), 449–476.

Boonen, T. J. and Jiang, W. (2024). Robust insurance design with distortion risk measures. *European Journal of Operational Research*, **316**(2), 694–706

Boonen, T. J., Liu, F. and Wang, R. (2021). Competitive equilibria in a comonotone market. *Economic Theory*, **72**(4), 1217–1255.

Borch, K. (1962). Equilibrium in a reinsurance market. *Econometrica*, **30**(3), 424–444.

Burzoni, M., Munari, C. and Wang, R. (2022). Adjusted Expected Shortfall. *Journal of Banking and Finance*, **134**, 106297.

Carlier, G., Dana, R.-A. and Galichon, A. (2012). Pareto efficiency for the concave order and multivariate comonotonicity. *Journal of Economic Theory*, **147**(1), 207–229.

Cai, J. and Chi, Y. (2020). Optimal reinsurance designs based on risk measures: A review. *Statistical Theory and Related Fields*, **4**(1), 1–13.

Cai, J., Li, J. Y. and Mao, T. (2023). Distributionally robust optimization under distorted expectations. *Operations Research*, published online, https://doi.org/10.1287/opre.2020.0685.

Cai, J., Liu, H. and Wang, R. (2017). Pareto-optimal reinsurance arrangements under general model settings. *Insurance: Mathematics and Economics*, **77**, 24–37.

Cai, J., Liu, F. and Yin, M. (2024). Worst-case risk measures of stop-loss and limited loss random variables under distribution uncertainty with applications to robust reinsurance. *European Journal of Operational Research*, published online, https://doi.org/10.1016/j.ejor.2024.03.016.

Cai, J. and Tan, K. S. (2007). Optimal retention for a stop-loss reinsurance under the VaR and CTE risk measures. *ASTIN Bulletin*, **37**(1), 93–112.

Casalaina-Martin, S., Frongillo, R., Morgan, T. and Waggoner, B. (2017). Multi-observation elicitation. *Proceedings of the 2017 Conference on Learning Theory,* PMLR **65**, 449–464.

Catoni, O. (2012). Challenging the empirical mean and empirical variance: a deviation study. In *Annales de l'IHP Probabilités et statistiques*, Vol. 48, No. 4, pp. 1148–1185.

Chen, L., He, S. and Zhang, S. (2011). Tight bounds for some risk measures, with applications to robust portfolio selection. *Operations Research*, **59**(4), 847–865.

Chen, A., Hieber, P. and Klein, J. K. (2019). Tonuity: A novel individual-oriented retirement plan. *ASTIN Bulletin*, **49**(1), 5–30.

Chi, Y. and Tan, K. S. (2011). Optimal reinsurance under VaR and CVaR risk measures: a simplified approach. *ASTIN Bulletin*, **41**(2), 487–509.

Chi, Z., Ramdas, A. and Wang, R. (2022). Multiple testing under negative dependence. *arXiv*: `2212.09706`.

Chiu, J. and Koeppl, T. V. (2022). The economics of cryptocurrency: Bitcoin and beyond. *Canadian Journal of Economics/Revue canadienne d'économique*, **55**(4), 1762–1798.

Conover, W. J. (1999). *Practical Nonparametric Statistics*, John Wiley & Sons.

Delage, E. and Li, Y. (2018). Minimizing risk exposure when the choice of a risk measure is ambiguous. *Management Science*, **64**(1), 327–344.

Denneberg, D. (1994). *Non-additive Measure and Integral.* Springer Science & Business Media.

Denuit, M. (2019). Size-biased transform and conditional mean risk sharing, with application to P2P insurance and tontines. *ASTIN Bulletin*, **49**(3), 591–617.

Denuit, M. and Dhaene, J. (2012). Convex order and comonotonic conditional mean risk sharing. *Insurance: Mathematics and Economics*, **51**(2), 265–270.

Denuit, M., Dhaene, J., Goovaerts, M.J. and Kaas, R. (2005). *Actuarial Theory for Dependent Risks: Measures, Orders and Models.* John Wiley & Sons.

Denuit, M., Dhaene, J. and Robert, C. Y. (2022a). Risk-sharing rules and their properties, with applications to peer-to-peer insurance. *Journal of Risk and Insurance*, **89**(3), 615–667.

Denuit, M., Hieber, P. and Robert, C. Y. (2022b). Mortality credits within large survivor funds. *ASTIN Bulletin*, **52**(3), 813–834.

Denuit, M. and Robert, C. Y. (2020). Risk reduction by conditional mean risk sharing with application to collaborative insurance. Available at https://ideas.repec.org/p/aiz/louvad/2020024.html.

Deprez, O. and Gerber, H. U. (1985). On convex principles of premium calculation. *Insurance: Mathematics and Economics*, **4**(3), 179–189.

De Vylder, F. and Goovaerts, M. J. (1982). Analytical best upper bounds on stop-loss premiums. *Insurance: Mathematics and Economics*, **1**(3), 163–175.

Douglas, R. (1965). Contractive projections on an L_1 space. *Pacific Journal of Mathematics*, **15**(2), 443–462.

Durrett, R. (2010). *Probability: Theory and Examples*. Cambridge University Press. Fourth Edition.

Eisner, T., Farkas, B., Haase, M. and Nagel, R. (2015). *Operator Theoretic Aspects of Ergodic Theory*. Springer, Heidelberg.

Embrechts, P., Puccetti, G. and Rüschendorf, L. (2013). Model uncertainty and VaR aggregation. *Journal of Banking and Finance*, **37**(8), 2750–2764.

Embrechts, P., Liu, H. and Wang, R. (2018). Quantile-based risk sharing. *Operations Research*, **66**(4), 936–949.

Embrechts, P., Liu, H., Mao, T. and Wang, R. (2020). Quantile-based risk sharing with heterogeneous beliefs. *Mathematical Programming Series B*, **181**(2), 319–347.

Embrechts, P., Mao, T., Wang, Q. and Wang, R. (2021). Bayes risk, elicitability, and the Expected Shortfall. *Mathematical Finance*, **31**, 1190–1217.

Embrechts, P. and Wang, R. (2015). Seven proofs for the subadditivity of Expected Shortfall. *Dependence Modeling*, **3**, 126–140.

Eyal, I. and Sirer, E. G. (2018). Majority is not enough: Bitcoin mining is vulnerable. *Communications of the ACM*, **61**(7), 95–102.

Fang, K. W. (2018). *Symmetric Multivariate and Related Distributions*. Chapman and Hall/CRC.

Feng, R. (2023). *Decentralized Insurance*. Springer, New York, Heidelberg.

Feng, R., Liu, M. and Zhang, N. (2022). A unified theory of decentralized insurance. *SSRN*: 4013729.

Filipović, D., Kupper, M. and Vogelpoth, N. (2012). Approaches to conditional risk. *SIAM Journal on Financial Mathematics*, **3**(1), 402–432.

Fisher, R. A. (1925). *Statistical Methods for Research Workers*. Edinburgh, Oliver and Boyd.

Fissler, T. and Ziegel, J. F. (2016). Higher order elicitability and Osband's principle. *The Annals of Statistics*, **44** (4), 1680–1707.

Föllmer, H. and Schied, A. (2002). Convex measures of risk and trading constraints. *Finance and Stochastics*, **6**(4), 429–447.

Föllmer, H. and Schied, A. (2016). *Stochastic Finance. An Introduction in Discrete Time*. Fourth Edition. Berlin, Walter de Gruyter.

Friedman, R. Z. (2020). *Probable Justice: Risk, Insurance, and the Welfare State*. University of Chicago Press, Chicago.

Frongillo, R., Mehta, N. A., Morgan, T. and Waggoner, B. (2019). Multi-observation regression. *Proceedings of the 22nd International Conference on Artificial Intelligence and Statistics,* PMLR **89**, 2691–2700.

Gablenz, P. and Sabatti, C. (2023). Catch me if you can: Signal localization with knockoff e-values. *arXiv*: 2306.09976.

Gastwirth, J. L. (1971). A general definition of the Lorenz curve. *Econometrica*, **39**(6), 1037–1039.

Ghaoui, L. M. E., Oks, M. and Oustry, F. (2003). Worst-case value-at-risk and robust portfolio optimization. *Operations Research*, **51**(4), 543–556.

Ghossoub, M., Jiang, W. and Ren, J. (2023). Optimal insurance for a prudent decision maker under heterogeneous beliefs. *European Actuarial Journal*, **13**(2), 703–730.

Gilboa, I. (2009). *Theory of Decision under Uncertainty.* Cambridge University Press, Cambridge.

Gilboa, I. Maccheroni, F., Marinacci, M. and Schmeidler, D. (2010). Objective and subjective rationality in a multiple prior model. *Econometrica*, **78**(2), 755–770.

Gilboa, I., Postlewaite, A., Samuelson, L. and Schmeidler, D. (2019). What are axiomatizations good for? *Theory and Decision*, **86**(3), 339–359.

Gilboa, I. and Schmeidler, D. (1989). Maxmin expected utility with non-unique prior. *Journal of Mathematical Economics*, **18**(2), 141–153.

Gneiting, T. (2011). Making and evaluating point forecasts. *Journal of the American Statistical Association*, **106**(494), 746–762.

Grünwald, P., de Heide, R. and Koolen, W. M. (2023). Safe testing. *Journal of the Royal Statistical Society Series B*, published online, https://doi.org/10.1093/jrsssb/qkae011.

Grünwald, P. (2023). The e-posterior. *Philosophical Transactions of the Royal Society A*, **381**(2247), 20220146.

Haneveld, W. K. K. (2013). *Duality in stochastic linear and dynamic programming (Vol. 274)* Springer Science and Business Media.

He, X. D., Kou, S. and Peng, X. (2022). Risk measures: robustness, elicitability, and backtesting. *Annual Review of Statistics and Its Application*, **9**, 141–166.

Henzi, A. and Ziegel, J. F. (2022). Valid sequential inference on probability forecast performance. *Biometrika*, **109**(3), 647–663.

Heras, A. J., Pradier, P. C. and Teira, D. (2020). What was fair in actuarial fairness? *History of the Human Sciences*, **33**(2), 91–114.

Hieber, P. and Lucas, N. (2022). Modern life-care tontines. *ASTIN Bulletin*, **52**(2), 563–589.

Hollander, M., Wolfe, D. A. and Chicken, E. (2013). *Nonparametric Statistical Methods*, John Wiley & Sons.

Howard, S. R., Ramdas, A., McAuliffe, J., and Sekhon, J. (2020). Time-uniform Chernoff bounds via nonnegative supermartingales. *Probability Surveys*, **17**, 257–317.

Howard, S. R., Ramdas, A., McAuliffe, J. and Sekhon, J. (2021). Time-uniform, nonparametric, nonasymptotic confidence sequences. *The Annals of Statistics*, **49**(2), 1055–1080.

Hu, W. and Kercheval, A. N. (2010). Portfolio optimization for student t and skewed t returns. *Quantitative Finance*, **10**(1), 91–105.

Hürlimann, W. (2005). Excess of loss reinsurance with reinstatements revisited. *ASTIN Bulletin*, **35**(1), 211–238.

Ignatiadis, N., Wang, R. and Ramdas, A. (2024). E-values as unnormalized weights in multiple testing. *Biometrika*, **111**(4), 417–439.

Jagannathan, R. (1977). Technical note–Minimax procedure for a class of linear programs under uncertainty. *Operations Research*, **25**(1), 173-177.

Jeffreys, H. (1998). *The Theory of Probability*, 3rd Ed. Oxford University Press.

Kaas, R., Goovaerts, M., Dhaene, J. and Denuit, M. (2008). *Modern Actuarial Risk Theory: Using R*. Springer Science & Business Media.

Kang, Z., Li, X., Li, Z. and Zhu, S. (2019). Data-driven robust mean-CVaR portfolio selection under distribution ambiguity. *Quantitative Finance*, **19**(1), 105–121.

Krokhmal, P., Zabarankin, M. and Uryasev, S. (2011). Modeling and optimization of risk. *Surveys in Operations Research and Management Science*, **16**(2), 49–66.

Kumon, M., Takemura, A. and Takeuchi, K. (2011). Sequential optimizing strategy in multi-dimensional bounded forecasting games. *Stochastic Processes and their Applications*, **121**(1), 155–183.

Landsberger, M. and Meilijson, I. (1994). Co-monotone allocations, Bickel-Lehmann dispersion and the Arrow-Pratt measure of risk aversion. *Annals of Operations Research*, **52**(2), 97–106.

Lehmann, E. L., Romano, J. P. and Casella, G. (1986). *Testing Statistical Hypotheses*, New York: Springer.

Leonard, H. B. (1983). Elicitation of honest preferences for the assignment of individuals to positions. *Journal of Political Economy*, **91**(3), 461–479.

Leshno, J. D. and Strack, P. (2020). Bitcoin: An axiomatic approach and an impossibility theorem. *American Economic Review: Insights*, **2**(3), 269–286.

Li, Y. M. J. (2018). Technical note–Closed-form solutions for worst-case law invariant risk measures with application to robust portfolio optimization. *Operations Research*, **66**(6), 1533–1541.

Li, L., Shao, H., Wang, R. and Yang, J. (2018). Worst-case Range Value-at-Risk with partial information. *SIAM Journal on Financial Mathematics*, **9**(1), 190–218.

Liu, H. and Mao, T. (2022). Distributionally robust reinsurance with Value-at-Risk and Conditional Value-at-Risk. *Insurance: Mathematics and Economics*, **107**, 393–417.

Liu, F., Cai, J., Lemieux, C. and Wang, R. (2020). Convex risk functionals: Representation and applications. *Insurance: Mathematics and Economics*, **90**, 66–79.

Liu, F., Mao, T., Wang, R. and Wei, L. (2022). Inf-convolution, optimal allocations, and model uncertainty for tail risk measures. *Mathematics of Operations Research*, **47**(3), 2494–2519.

Maccheroni, F., Marinacci, M. and Rustichini, A. (2006). Ambiguity aversion, robustness, and the variational representation of preferences. *Econometrica*, **74**(6), 1447–1498.

Mao, T., Wang, R. and Wu, Q. (2022). Model aggregation for risk evaluation and robust optimization. *arXiv*: `2201.06370`.

Markowitz, H. (1959). *Portfolio Selection: Efficient Diversification of Investments*. John Wiley & Sons.

Markowitz, H., Todd, P., Xu, G. and Yamane, Y. (1993). Computation of mean-semi-variance efficient sets by the critical line algorithm. *Annals of Operations Research*, **45**, 307-317.

Meyn, J., Kandziora, M., Albers, S. and Clement, M. (2023). Consequences of platforms' remuneration models for digital content: initial evidence and a research agenda for streaming services. *Journal of the Academy of Marketing Science*, **51**(1), 114–131.

McNeil, A. J., Frey, R. and Embrechts, P. (2015). *Quantitative Risk Management: Concepts, Techniques and Tools*. Revised Edition. Princeton, Princeton University Press.

Nakamoto, S. (2008). Bitcoin: A peer-to-peer electronic cash system. *Decentralized Business Review*, 21260.

Natarajan, K., Pachamanova, D. and Sim, M. (2008). Incorporating asymmetric distributional information in robust value-at-risk optimization. *Management Science*, **54**(3), 573–585.

Nolde, N. and Ziegel, J. F. (2017). Elicitability and backtesting: Perspectives for banking regulation (with discussion). *Annals of Applied Statistics*, **11**(4), 1833–1874.

Ogryczak, W. and Ruszczyński, A. (2001). On consistency of stochastic dominance and mean–semideviation models. *Mathematical Programming*, **89**, 217–232.

Ogryczak, W. and Ruszczyński, A. (2002). Dual stochastic dominance and related mean-risk models. *SIAM Journal on Optimization*, **13**(1), 60–78.

Owen, J. and Rabinovitch, R. (1983). On the class of elliptical distributions and their applications to the theory of portfolio choice. *The Journal of Finance*, **38**(3), 745–752.

Pesenti, S., Wang, Q. and Wang, R. (2020). Optimizing distortion riskmetrics with distributional uncertainty. *arXiv*: 2011.04889.

Pfanzagl, J. (1967). Characterizations of conditional expectations. *Annals of Mathematical Statistics*, **38**(2), 415–421.

Pflug, G. C. (2000). Some remarks on the value-at-risk and the conditional value-at-risk. In *Probabilistic Constrained Optimization* (Ed. Uryasev, S.), pp. 272–281, Springer, Dordrecht.

Pflug, G. C., Timonina-Farkas, A. and Hochrainer-Stigler, S. (2017). Incorporating model uncertainty into optimal insurance contract design. *Insurance: Mathematics and Economics*, **73**, 68–74.

Porter, R. B. (1974). Semivariance and stochastic dominance: A comparison. *The American Economic Review*, **64**(1), 200–204.

Rahimian, H. and Mehrotra, S. (2022). Frameworks and results in distributionally robust optimization. *Open Journal of Mathematical Optimization*, **3**, 1–85.

Ramdas, A., Grünwald, P., Vovk, V. and Shafer, G. (2023). Game-theoretic statistics and safe anytime-valid inference. *Statistical Science*, **38**(4), 576–601.

Ren, Z. and Barber, R. F. (2023). Derandomized knockoffs: Leveraging e-values for false discovery rate control. *Journal of the Royal Statistical Society Series B*, **86**(1), 122-154.

Rockafellar, R. T. and Royset, J. O. (2014). Random variables, monotone relations, and convex analysis. *Mathematical Programming*, **148**, 297–331.

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 and Finance*, **26**(7), 1443–1471.

Rockafellar, R. T. and Uryasev, S. (2013). The fundamental risk quadrangle in risk management, optimization and statistical estimation. *Surveys in Operations Research and Management Science*, **18**(1-2), 33–53.

Rohatgi, V.K. (2011). Semi-Variance in Finance. In: Lovric, M. (eds). *International Encyclopedia of Statistical Science*, 1297-1298, Springer, Berlin.

Rothschild, M. and Stiglitz, J. E. (1970). Increasing risk: I. A definition. *Journal of Economic Theory*, **2**(3), 225–243.

Rüschendorf, L. (2013). *Mathematical Risk Analysis. Dependence, Risk Bounds, Optimal Allocations and Portfolios*. Springer, Heidelberg.

Simes, R. J. (1986). An improved Bonferroni procedure for multiple tests of significance. *Biometrika*, **73**(3), 751–754.

Shafer, G. (2021). The language of betting as a strategy for statistical and scientific communication. *Journal of the Royal Statistical Society, Series A*, **184**(2), 407–431.

Shafer, G. and Vovk, V. (2019). *Game-Theoretic Foundations for Probability and Finance*. New York, Wiley.

Shaked, M. and Shanthikumar, J. G. (2007). *Stochastic Orders*. Springer Series in Statistics.

Starr, R. M. (2011). *General Equilibrium Theory: An Introduction.* Second Edition, Cambridge University Press.

Schmeidler, D. (1989). Subjective probability and expected utility without additivity. *Econometrica*, **57**(3), 571–587.

Tang, Q. and Yang, Y. (2023). Worst-case moments under partial ambiguity. *ASTIN Bulletin*, **53**(2), 443–465.

Testuri, C. E. and Uryasev, S. (2004). On relation between expected regret and conditional value-at-risk. *Handbook of Computational and Numerical Methods in Finance*, 361–372, Birkhäuser, Boston, MA.

Ville, J. (1939). Étude critique de la notion de collectif. *Thèses de l'entre-deux-guerres*, 218.

Vovk, V. and Wang, R. (2020). Combining p-values via averaging. *Biometrika*, **107**(4), 791–808.

Vovk, V. and Wang, R. (2021). E-values: Calibration, combination, and applications. *Annals of Statistics*, **49**(3), 1736–1754.

Vovk, V. and Wang, R. (2023). Confidence and discoveries with e-values. *Statistical Science*, **38**(2), 329–354.

Vovk, V. and Wang, R. (2024). Nonparametric e-tests of symmetry. *New England Journal of Statistics in Data Science*, **2**(2), 261–270.

von Neumann, J. and Morgenstern, O. (1947). *Theory of Games and Economic Behavior*. Princeton University Press, Second Edition.

Wakker, P. P. (2010). *Prospect Theory: For Risk and Ambiguity*. Cambridge University Press.

Wang, R. and Ramdas, A. (2022). False discovery rate control with e-values. *Journal of the Royal Statistical Society Series B*, **84**(3), 822–852.

Wang, H. and Ramdas, A. (2023). Catoni-style confidence sequences for heavy-tailed mean estimation. *Stochastic Processes and their Applications*, **163**, 168–202.

Wang, Q., Wang, R. and Ziegel, J. (2022). E-backtesting. *arXiv*: 2209.00991.

Wang, R. and Wei, Y. (2020). Risk functionals with convex level sets. *Mathematical Finance*, **30**(4), 1337–1367.

Wang, R., Wei, Y. and Willmot, G. E. (2020). Characterization, robustness and aggregation of signed Choquet integrals. *Mathematics of Operations Research*, **45**(3), 993–1015.

Wang, R. and Zitikis, R. (2021). An axiomatic foundation for the Expected Shortfall. *Management Science*, **67**(3), 1413–1429.

Wasserman, L., Ramdas, A. and Balakrishnan, S. (2020). Universal inference. *Proceedings of the National Academy of Sciences*, **117**(29), 16880–16890.

Waudby-Smith, I. and Ramdas, A. (2023). Estimating means of bounded random variables by betting. *Journal of the Royal Statistical Society Series B*, **86**(1), 1–27.

Xia, J. and Zhou, X. Y. (2016). Arrow–Debreu equilibria for rank-dependent utilities. *Mathematical Finance*, **26**(3), 558–588.

Yaari, M. E. (1987). The dual theory of choice under risk. *Econometrica*, **55**(1), 95–115.

Zhu, S. and Fukushima, M. (2009). Worst-case conditional value-at-risk with application to robust portfolio management. *Operations Research*, **57**(5), 1155-1168.

Zhu, S., Li, D. and Wang, S. (2009). Robust portfolio selection under downside risk measures. *Quantitative Finance*, **9**(7), 869–885.

Zuluaga, L.F., Peña, J. and Du, D. (2009). Third-order extensions of Lo's semiparametric bound for European call options. *European Journal of Operational Research*, **198**(2), 557–570.